Predictive Models for Cyberbullying in Middle and High School Students

Rick Hsu

U BSS 6

**Abstract.** This study aims to build a predictive model of cyber bullying among middle and high school students in the United States, which would be useful in social studies and intervention efforts. Data from qualitative surveys conducted by the National Crime Victimization Survey were used. Predictive models were developed and tested using Stepwise Logistic regression analysis and Artificial Neural Network (ANN). Potential predictors tested include demographic characteristics and student’s participation in different school activities. From Stepwise Logistic regression analysis, five variables were kept in the predictive model, including gender, race, participating in community service, and spirit groups. The Area Under the Curve (AUC) for the Receiver Operating Characteristics was 79% which indicates good performance of the model. The Artificial Neural Network model also had good performance with an AUC of 77%. The findings of this study have shown clear relevancies of certain student characteristics with probability of being victimized of cyber bullying. These findings will be helpful in identifying and providing necessary assistance to the vulnerable students in middle and high schools.

**Rationale.** In recent decades, with the rising popularity of the internet and mobile devices, there has been a new form of bullying called cyberbullying. Cyberbullying or cyber harassment is defined as bullying or harassment through the use of electronics. These most commonly occur in social media, SMS, email, or means of electronic communication. 37% of students between the ages of 12 and 17 had been victims. The most highly preached ways of reducing cyberbullying is for the victim, other students, or the parents to be cautious about it and report it. With all these assemblies, posters, and media that students will see telling them to report cyberbullying, only one in ten victims report it. A way to identify those at risk the most may help utilize the limited resources efficiently in intervention and prevention efforts against cyberbullying.

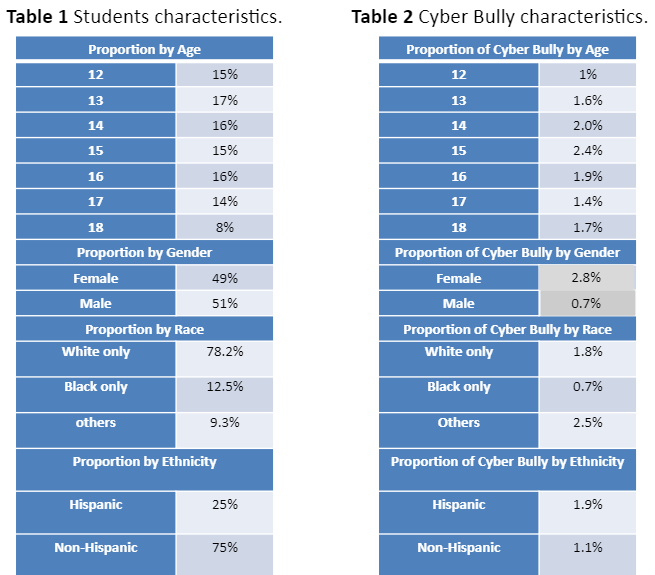
**Alternate Hypothesis.** There are traits/characteristics of significance that victims share.

**Null Hypothesis.** There are no traits/characteristics of significance that victims share.

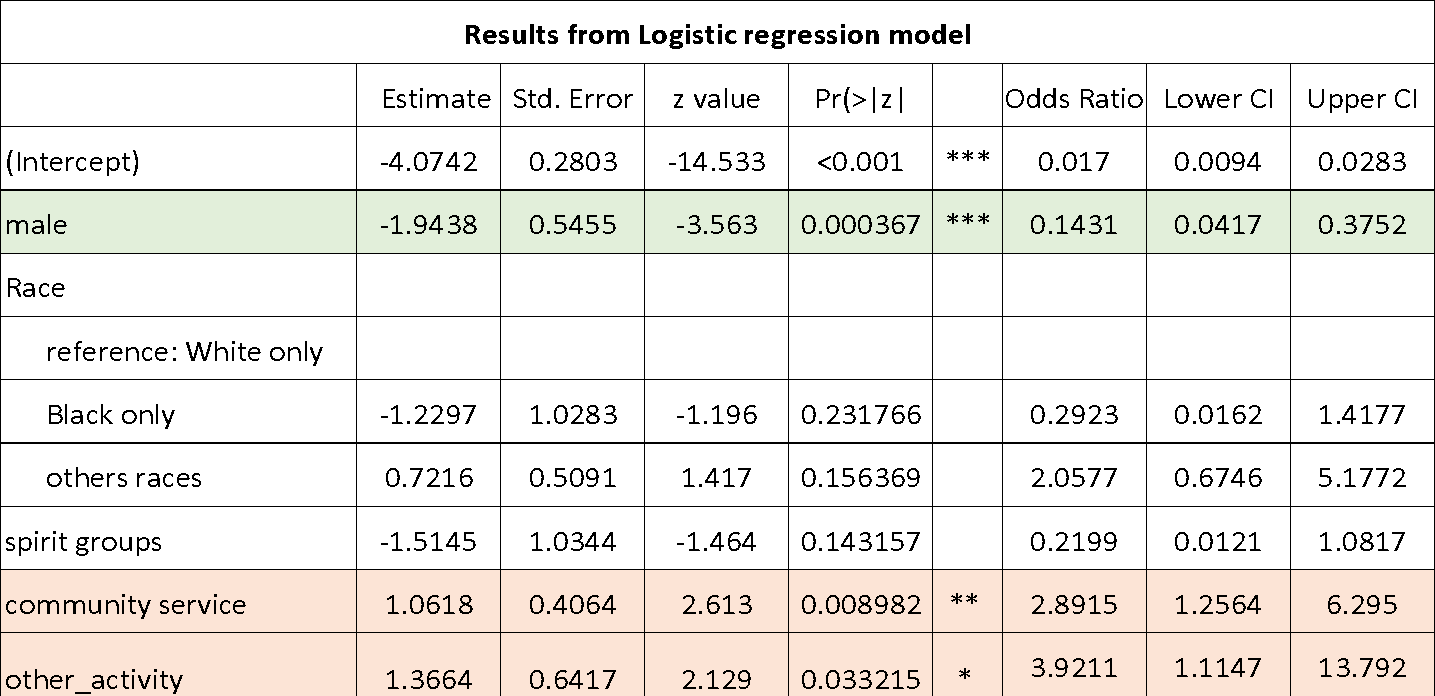
**Variables:** Race, Gender, participation in spirit groups, participation in sports, participation in other extracurricular.

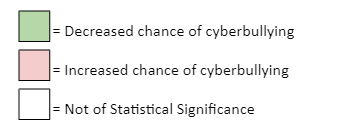
**Control**: The intercept variable.

**Trials:** Two trials/models. The logistic regression is a reliable and well-used statistical model that has good conclusions. The Artificial Neural Network is supplemental.

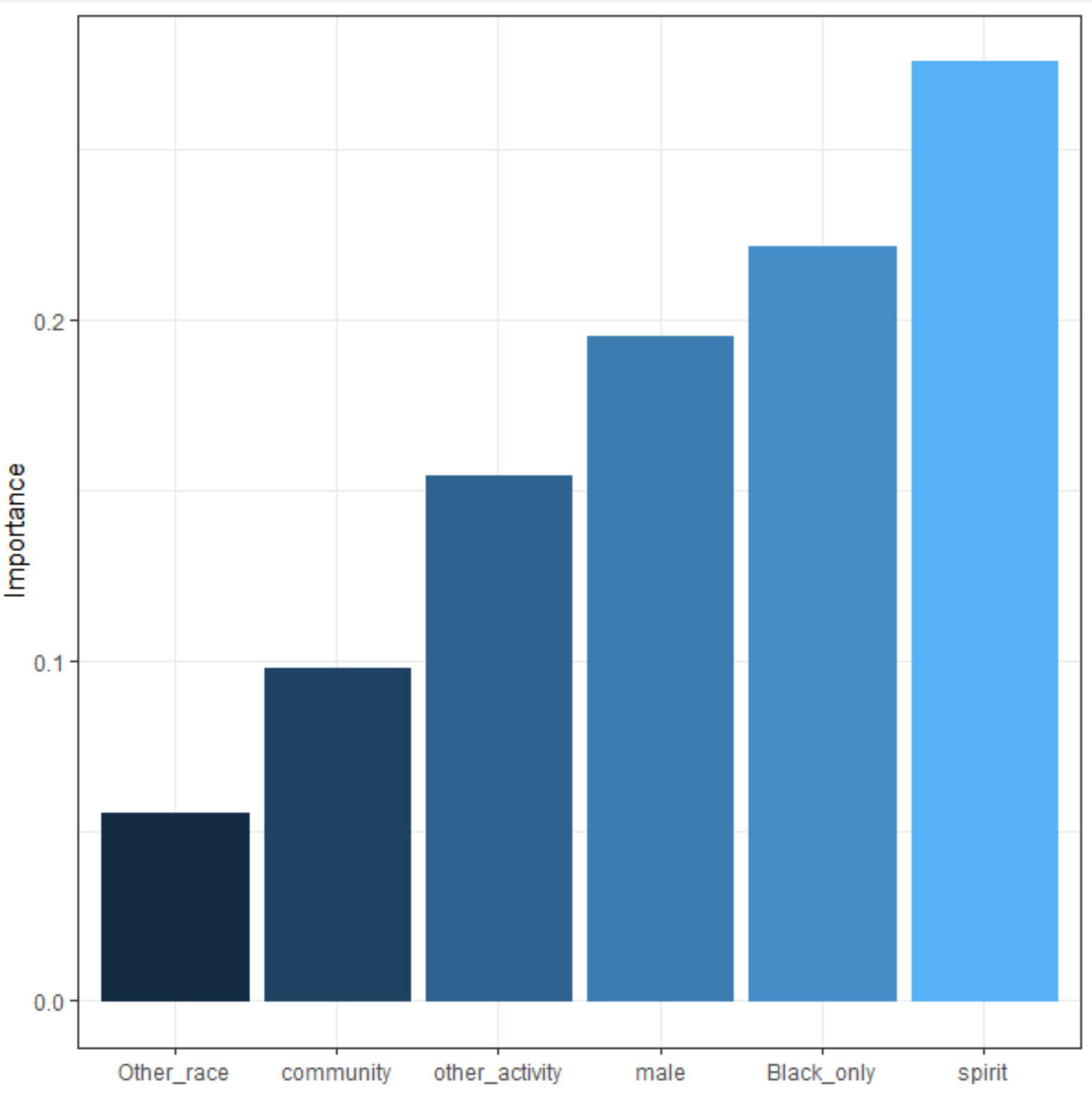
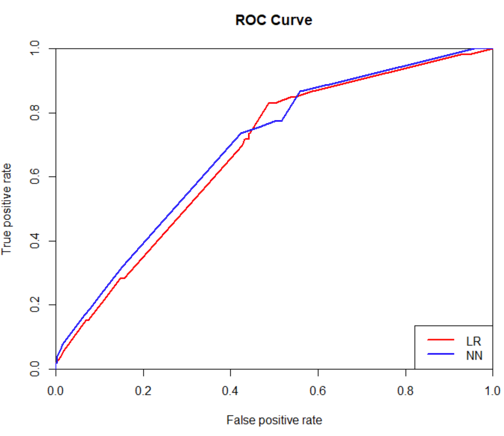


Data derived from NCVS:SCS survey. Compiled in the R coding language.





Logistic Regression Model. Row Variables are student traits/characteristics. Column variables are explained in more detail in conclusion.

Conclusions

Characteristics such as gender, race, and extracurricular activities showed a correlation to cyberbullying. **Odds Ratio**- decimal interpretation chance of victimization shows that : 0.14 times less for males, 2.89 times more for participation in community service. 3.9 times more “other activities. **Confidence Interval**- 95% confident that the true parameter falls into the estimated odds ratio. P values for three variables are less than 0.05, indicating that the null hypothesis can be rejected if less than 0.05 **Estimate**(population slope coefficient)- similar to odds ratio. Negative indicates decreased chance and vice versa.

**Neural Network**. Participation in spirit groups deemed as most important, then being black, being a male… Note: High Importance != High chances of victimization

Two models had similar and different results. Spirit group and race was not statistically significant in logistic regression but were the two most important factors in the neural network model.

**Further Study.** Variable “School Type” (private or public) was not included in the models due to insufficient reliable data. School type is likely strongly related with students’ experience at school and would improve the model. The question of cyber bullying could be framed not as a single question, but rather as a set of questions that leads to a tally of total score which is used in turn to decide whether the behaviors qualify as cyber bullying action or not. A scale of 1-10 could then be derived representing the degree of cyberbullying.

Work Cited

1. Refer to National Crime Victimization Survey (NCVS): School Crime Supplement (SCS) <https://nces.ed.gov/programs/crime/index.asp> for statics information

2. Evaluation of Predictive Models. Decision Systems Group, Brigham and Women’s Hospital Harvard Medical School.

3. Garson, G. D. Interpreting neural network connection weights. Artif. Intell. Expert 6, 46–51 (1991).

4. Ben-Joseph, E. P. (Ed.). (2018, April). Cyber bullying (for Teens) - Nemours KidsHealth. Retrieved from https://kidshealth.org/en/teens/cyberbullying.html