Preventing Childhood Bullying

Julie Campbell

Introduction

Bullying in school aged children (6-18 years old) threatens a students' physical and emotional safety at educational institutions. Their peers can subject them to verbal, social, and/or physical bullying. The National Center for Education Statistics calculates that 1 out of 5 children (20.2%) are subjected to bullying (National Center for Education Statistics, 2019, pp. T-36). Children who participate in bullying others are likely to engage in violent and other risk behaviors into adulthood (Stop Bullying, 2021).

To measure the factors involved with children who participate in bullying, data was collected from the National Survey of Children's Health (United States Census Bureau, 2021). The dataset contains 461 features to examine the physical and emotional health of children and their parents. One of the features involves rating the frequency of bullying on a 1-5 scale.

Business Opportunity

The research conducted is intended to assist educational institutions to determine the risk of a child participating in frequent bullying (3+ on a 5-point scale, 1 = never, 5 = daily). Intervening after bullying takes place does not prevent the trauma to the child being bullied and the bully also suffers in not receiving the support needed that led to the behavior. According to APA's Tolerance Task Force, the zero-tolerance programs mandate a severe punishment such as long-term suspension or expulsion (American Psychological Association, 2015). This does not improve the school climate and doesn't address the prejudice that students hold.

The opportunity involved with predicting bullying behavior before it begins is to provide an intervention program for children at risk. This program could provide resources to the student and/or parent(s)/guardian(s) to provide support to focus on educational goals rather than negative interaction with their peers. This will indirectly have a positive effect on the school's average grades and testing scores. It will also provide the opportunity to redistribute resources to direct improvement of student's lives. Utilizing the bullying scale from the dataset, a model could be utilized to flag a student for participation in the intervention program.

Analytical Framework

The features collected in the National Survey of Children's Health are limited to the physical and emotional health of children in the United States. The survey is distributed to households with children that the screener identifies. The questionnaire is then completed by an adult that

is familiar with the child's health care. The 2021 dataset consists of a sample of approximately 300,000 addresses from the Census Master Address File (MAF) (United States Census Bureau, 2021). The survey had three versions distributed; however, only version 2 and 3 included the question about the child's participation in bullying. The resulting records include 3.4% children that were flagged as frequent bullying (3+ on a 5-point scale). Performance will be measured on how well the model can accurately predict students who participate in frequent bullying (at least once per month).

Results

The research was conducted in three stages consisting of data exploration, wrangling, and model building to determine the risk of a school aged child becoming a bully.

In the first stage, data was explored to determine the correlation between key characteristics of all children recorded in the survey. Child race was evaluated in the survey sample, and it was noted that most of the children documented were white. The age and gender of the children were documented to be approximately evenly distributed across the sample. Family size was considered when comparing the bullying and bullied frequency of children recorded in the survey. Figure 1 shows that children who bully others daily are often not bullied and consists of the highest average family size.

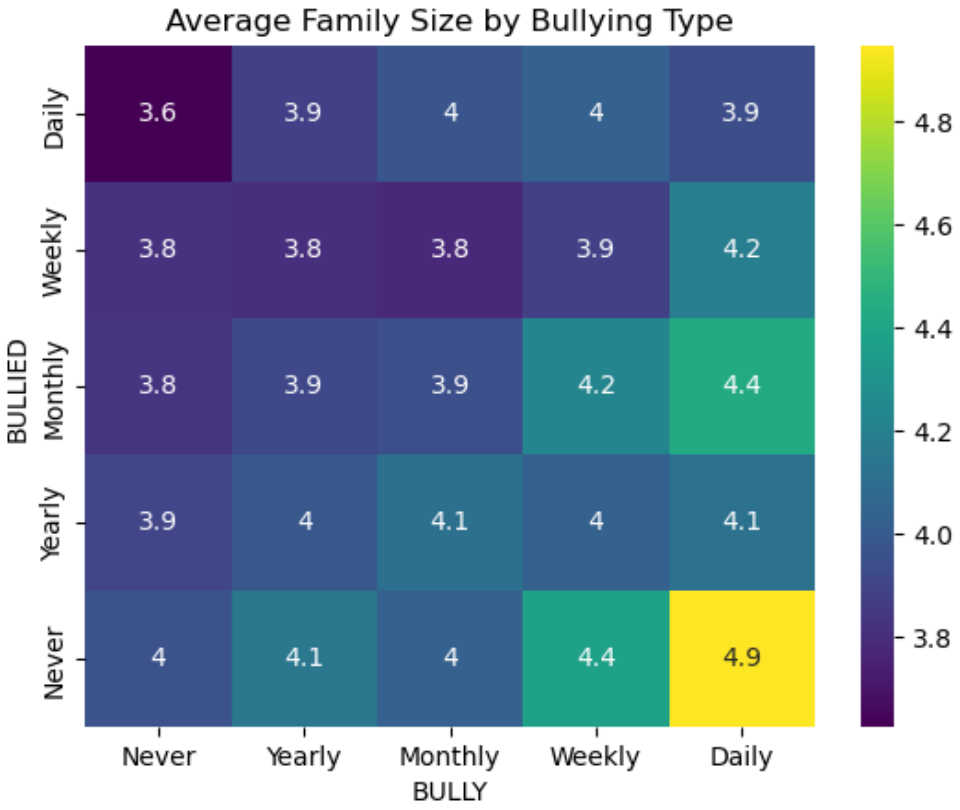

FIGURE 1: AVERAGE FAMILY SIZE BY BULLYING TYPE

By averaging the bullying rates by age, Figure 2 shows that the highest ratings were concentrated between ages 6-12 in 2021. Although most bullying cases happens in 9th grade or higher, the frequency of a bullying incident is slightly higher in the younger ages. According to an article by Boston Childrens Hospital, bullying in the elementary school age can take form of name-calling, excluding, or pushing, shoving, or tripping another child. In contrast, middle and high school students typically bully through spreading rumors and gossiping on school property or online (Rizzitano, 2021).

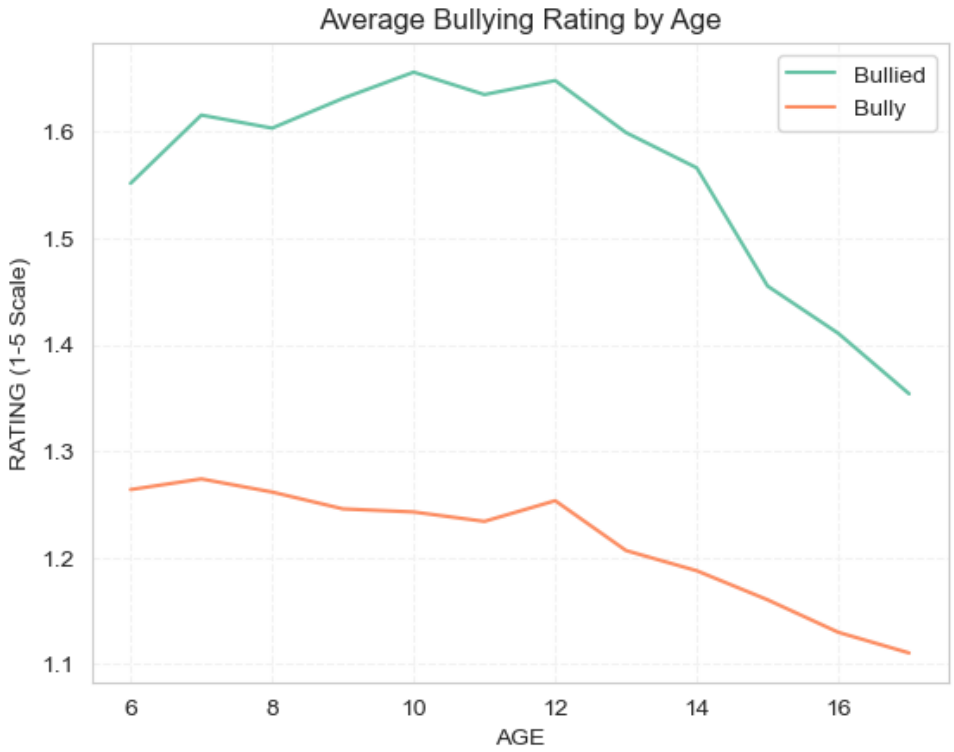

FIGURE 2: AVERAGE BULLYING RATE BY AGE

The parent's/guardian's ability to provide necessities (food/housing) in a household was also measured by bullying type. Figure 3 shows that the frequency of children who are bullied increases as the parent(s)/guardian(s) record that they can cover necessities. There is a slight increase in bullying behavior as well.

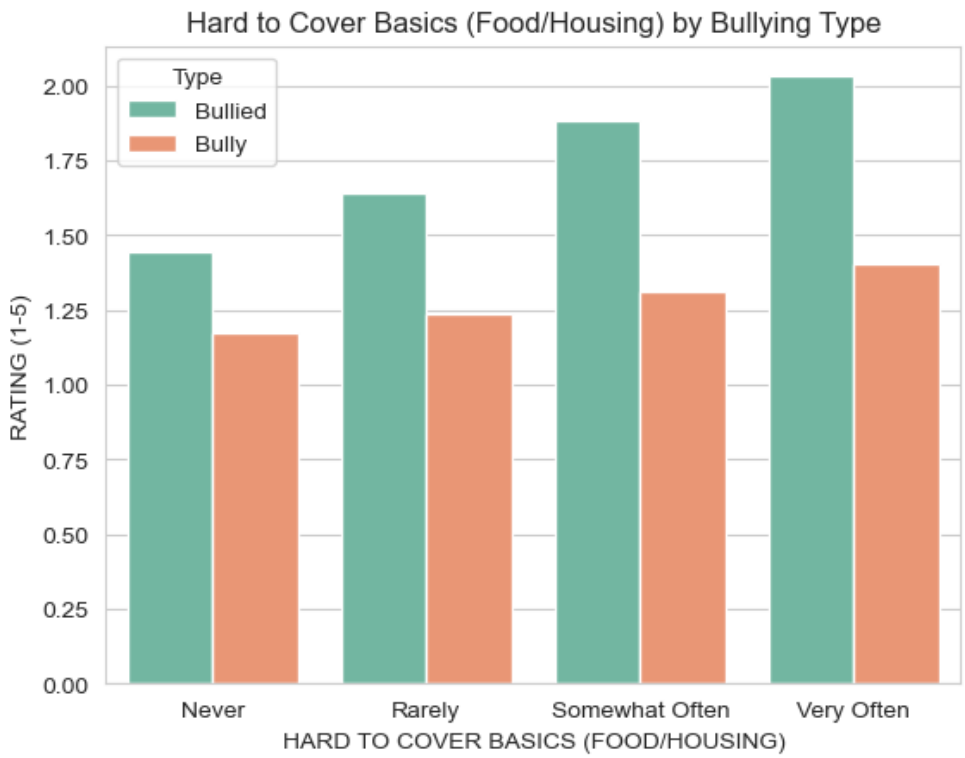

FIGURE 3: COVERING NECESSITIES BY BULLYING TYPE

In the next stage, data wrangling was performed to include only the features that provided direct value in relation to the bullying feature. All records from version 1 of the survey were removed as they had no data related to bullying as it targeted children under the age of 6. The 68 features that directly related to version 1 were also removed as they only consisted of null values. The unique household ID, form type, and operational sampling were removed as they were only used for survey documentation purposes. Features that had a significant correlation with another feature were also removed. For example, if a child was noted to have autism there were further features to discuss the medication and treatment plan implemented.

When reviewing the question-and-answer format of the surveys, it was found that most questions allowed for only two answers (1 or 2). The features that met these requirements were filtered out to observe a limited set of features that needed to be monitored for potential outliers. 24 features were found to meet these criteria, but no outliers were found. A new column was created to hold a value of 0 (False) or 1 (True) if a child was documented to engage in bullying at least once a month (3+ on a 5-point scale). The original bully feature was dropped in response to the creation of this new column.

All categorical columns had dummies created to prepare for model building. Any leftover values were filled with a 0 value. Data was split into an 80/20 framework (train/test). Each consisted of 3.4% records containing true bully values.

In the last stage, the training set was balanced utilizing the SMOTE methodology. The test set was not included in the resampling. A pipeline was created to be utilized in a grid search to find the best hyperparameters for the model. A min-max scaler was used for the features. After comparing two different model types, it was found that the model parameters that gave the best accuracy was the following:

```
LogisticRegression(C=1, penalty='l1', solver='liblinear')
```

The logistic regression model gave an accuracy of 96.8% and an R2 score of -0.0029. Figure 4 shows the confusion matrix for the model. Even though this model provided the best accuracy, we can see that 156 children were predicted not to be at risk of participating in bullying behavior even though they were reported true in the test dataset.

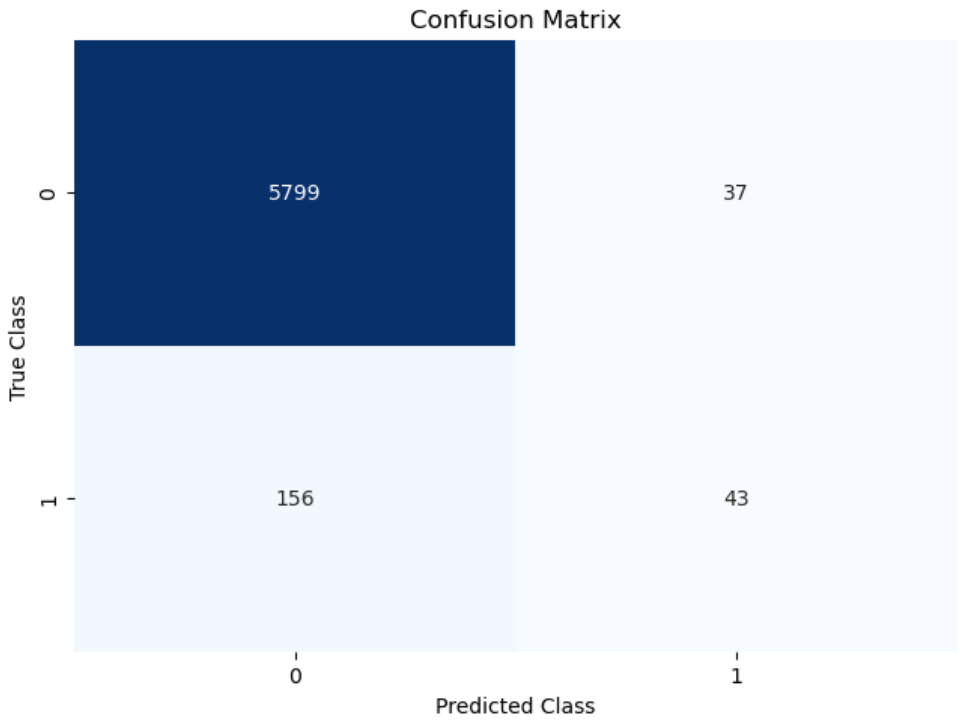

FIGURE 4: CONFUSION MATRIX

To further estimate the performance of the model, a classification report was created with the results shown in Figure 5. The recall and f1-score were particularly low for predicting the risk level of a child engaging in bullying behavior.

	precision	recall	f1-score	support
0	0.97	0.99	0.98	5836
1	0.54	0.22	0.31	199
accuracy			0.97	6035
macro avg	0.76	0.60	0.65	6035
weighted avg	0.96	0.97	0.96	6035

FIGURE 5: CLASSIFICATION REPORT

Next Steps

In conclusion, the logistic regression model built with the childhood health data had relatively low performance compared to its accuracy of 96.8%. Additional research would need to be conducted to determine what variables may be better predictors of childhood bullying. If possible, collecting a dataset that is responded to by the child would be preferrable. Children may not directly tell their parents all the behaviors that may happen at school that are not documented. This model could be utilized to begin screening students for resources, but it should not be used as the only source of determining risk.

References

American Psychological Association. (2015). *Do U.S. laws go far enough to prevent bullying at school?* Retrieved from https://www.apa.org/monitor/2016/02/ce-corner

National Center for Education Statistics. (2019). Retrieved from https://nces.ed.gov/pubs2019/2019054.pdf

Rizzitano, E. (2021, November 1). *Bullying at all ages: How to spot the signs and support your child.* Retrieved from Boston Children's Hospital: https://answers.childrenshospital.org/bullying/

Stop Bullying. (2021). *https://www.stopbullying.gov/bullying/effects.*

United States Census Bureau. (2021). *NSCH Datasets.* Retrieved from 2021 Topical Data and Input Files: https://www.census.gov/programs-surveys/nsch/data/datasets.html