

## **ESTUDIO DE LAS POSIBLES CAUSAS DEL BULLYING**

Lucía Carro Arcaya, Lidia Moreno Marín y Amparo Gálvez Vilar

El objetivo de este informe es ver la relación de distintas variables a la hora de predecir si un niño sufre bullying o no, ver cuáles de ellas influyen realmente en que esto pase.

Para ello se desarrollará un modelo que permita predecir si un niño sufre bullying a partir de las variables del dataset “Bullying\_2018.csv”. Se realizarán dos modelos, el primer modelo consideraremos las variables más influyentes intuitivamente y el segundo será el que nos surgiera la función step analizando todas las variables.

En ambos la variable bullying será la variable respuesta de tipo factor, la cual indica si los niños sufren o no bullying; y todas las demás las covariables.

Partiendo de las variables empleadas en los modelos extraeremos las conclusiones por las cuales hemos generado los modelos.

### **INTRODUCCIÓN**

Encuesta Mundial de Salud Escolar (GSHS), una encuesta escolar para obtener datos sobre la salud de los jóvenes y los factores de protección relacionados con las principales causas de mortalidad.

En la encuesta participaron 56981 estudiantes en 2018 en Argentina. Se obtuvo el 86% en la tasa de respuesta escolar, el 74% en tasa de respuesta de los estudiantes y el 63% como tasa de respuesta general.

El GSHS contaba con 18 variables, de las cuales se seleccionan las 12 variables más relacionadas con el acoso escolar para el desarrollo de la predicción del bullying.

Algunos factores de influencia han sido el sentimiento de soledad, la falta de amigos cercanos, una mala comunicación con los padres, además del peso de los jóvenes.

Para poder trabajar adecuadamente hemos realizado preprocesamiento de los datos. En él hemos modificado el tipo de las variables a tipo factor, excepto la edad que se ha puesto en tipo numérica. Esto se debe a que todas estas estaban en tipo carácter y no podíamos trabajar con ellas si estaban en ese patrón. Cabe destacar la variable peso, la cual es un factor de tres niveles y hemos calculado a partir de otras tres de los datos originales.

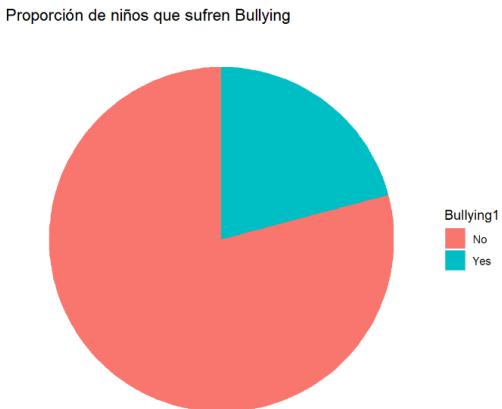
Por otra parte, hemos decidido descartar los NAs seguimos quedándonos con un gran número de observaciones y el trabajo sería más sencillo.

El objetivo de este estudio consiste en predecir el acoso escolar en función de las variables más influyentes y a partir de este tomar las medidas necesarias para prevenirlo.

Para conseguir estas predicciones hemos decidido, sabiendo que nuestra variable respuesta es categórica, utilizar el modelo de regresión logística múltiple, dado que esta modela la probabilidad de que nuestra variable pertenezca a una categoría en particular.

## RESULTADOS

Una buena aproximación antes de iniciar el modelo es visualizar de una manera gráfica el número de niños que sufren bullying de nuestro banco de datos, esto lo tendremos en cuenta para analizar los resultados obtenidos.



En cuestión de porcentajes observamos que el 20.9% de los niños sufren bullying y el 79.1% restante no.

- Observamos que unos 6000 de los 30000 niños de la encuesta sufren bullying, el número de registros disminuye por la inexactitud o falta de datos de alguno de ellos.

Previamente a implementar cualquier modelo hay una selección de variables y filtrado de datos en el cual ajustamos las variables para facilitar su uso en los modelos.

Realizamos un primer modelo en el que introducimos las variables que a primera vista nos resultan más determinantes a la hora de ver si un niño sufre bullying. Estas son la edad, el peso y el número de amigos cercanos.

```
modelo1 <- glm(Bullying ~ Custom_Age + Peso + Close_friends,  
                 data = bullying, family = binomial)
```

Al realizar las predicciones de este modelo y ver la matriz de confusión vemos que no predice que ningún niño sufra bullying.

modelo_entren.pred	
held.out.B	0_NoBullying
0	7554
1	1946

Esto se debe a que las probabilidades predichas en función a nuestro modelo son todas más pequeñas que 0.5, que es el rango que hemos definido (a partir del cual, si es un valor mayor el modelo predeciría que el niño sufre de bullying).

Por lo tanto, vemos que las variables elegidas no influyen lo suficiente para que el modelo sea bueno.

Para generar el siguiente modelo necesitamos hacerlo a partir de todas las variables para poder implementar la función step, y así ver cuáles son realmente influyentes en la determinación de si un niño sufre bullying.

El resultado que se obtiene nos genera muchas variables explicativas, por tanto, vamos a seleccionar las 4 con el menor AIC, ya que un valor AIC más bajo indica un mejor ajuste.

```
modelo3 <- glm(Bullying ~ Other_students_kind_and_helpful + Physically_attacked +  
                  Felt_Lonely + Cyber_bullied_in_past_12_months,  
                  data = bullying, family = binomial)
```

Los parámetros más significativos de este algoritmo son la ayuda de los compañeros cuando es nunca, la ayuda de los compañeros cuando es raramente y si ha sufrido ciberbullying los pasados 12 meses cuando sí que ocurre.

Tras calcular las predicciones y ver la matriz de confusión del modelo, obtenemos que 7298 niños sufren bullying y 307 no, en total 7605 predicciones correctas.

held.out.B	0_NoBullying	1_YesBullying
0	7240	298
1	1587	375

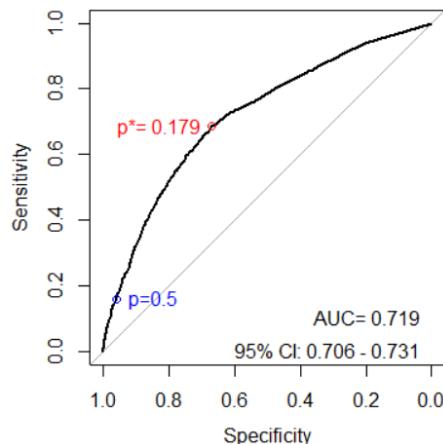
La sensibilidad de este modelo es del 19% y la especificidad del 97%.

Una sensibilidad del 19% significa que el modelo clasifica correctamente el 19% de los casos positivos reales y una especificidad del 97% significa que el modelo clasifica correctamente el 97% de los casos negativos reales.

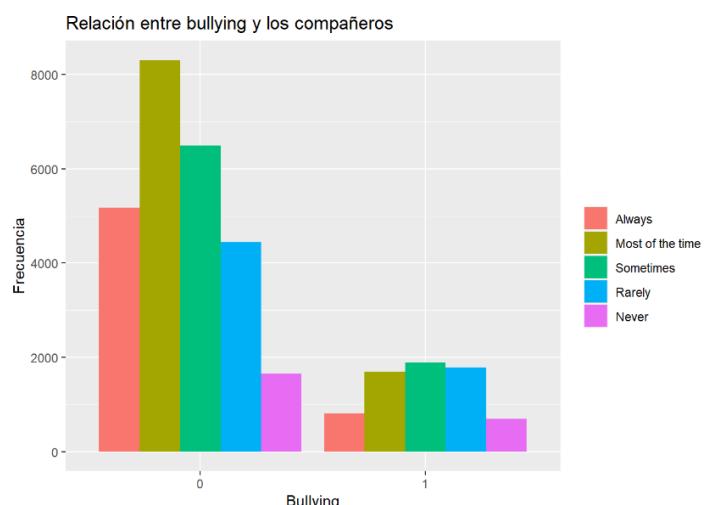
Este modelo es más eficiente en la clasificación de casos negativos que de casos positivos, pero hay un equilibrio ya que el error de predicción gira en torno al 20%.

La regresión logística previamente calculada predice correctamente sufrir o no sufrir bullying el 80.05% de los datos de prueba.

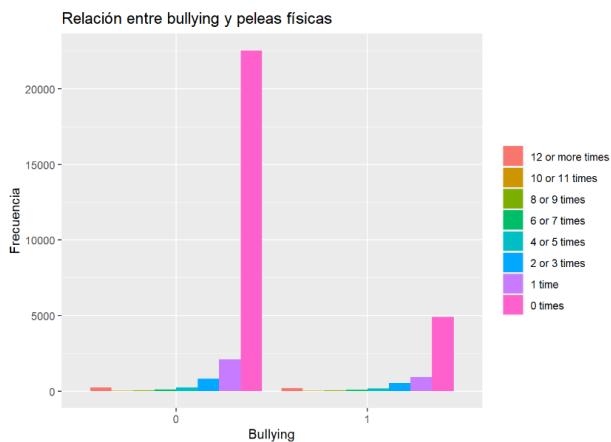
Vemos que el modelo calculado es un buen modelo gracias a el área de debajo de la curva ROC (AUC) que tiene un valor de 0.73. El intervalo de confianza del 95% tiene un valor de [0.717-0.742]. El error de predicción que se obtiene es del 20.26%.



La ayuda y simpatía de los compañeros se encuentra muy relacionada con el sufrir bullying o no. Podemos ver en la gráfica que la mayoría de los que no han sufrido bullying no sea han sentido solos la mayoría del tiempo, en cambio si nos fijamos en los niños que si lo han sufrido vemos que la mayoría a veces si que se han sentido que los compañeros han sido amables.

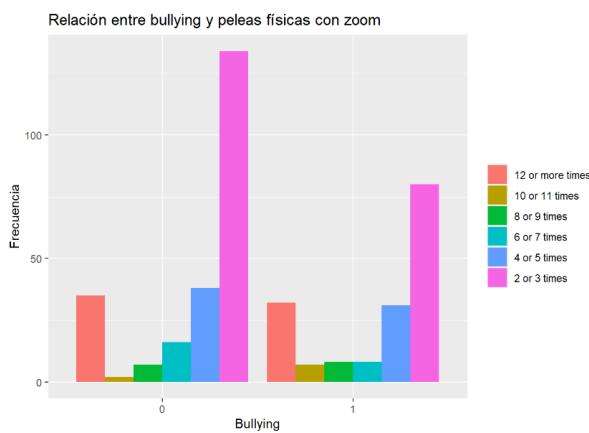


En el caso de las peleas físicas vemos un gran pico en que no haya sufrido bullying y no haya sufrido ninguna pelea nunca. Por ello vamos a realizar un zoom y centrarnos en otros niveles, vamos a eliminar también para el gráfico que hayan sufrido una pelea física, ya que también genera un gran pico en relación con el resto de valores.

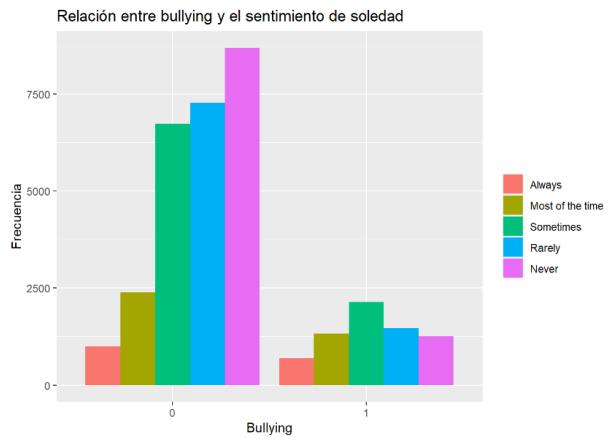


Quitando ese factor vemos con mayor claridad que los niños que han sufrido alguna pelea hay un mayor porcentaje que no ha sufrido bullying. Llama la atención que el número de peleas sea 12 o más en ambas gráficas es prácticamente el mismo, hay que tener en cuenta que el número de niños que sufren bullying es de 6000 y que los que no son en torno a 26000.

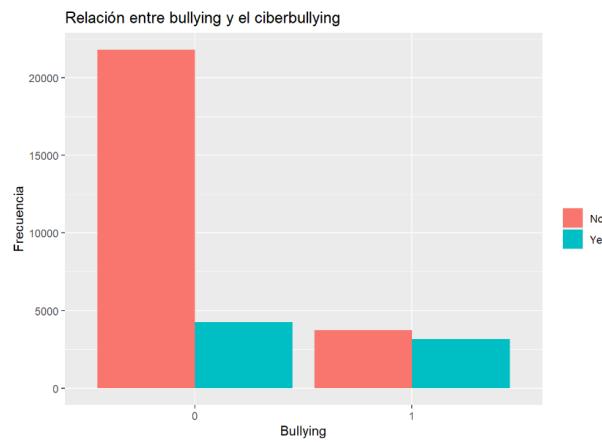
En proporción el número de peleas que tienen los niños que sufren bullying es mucho mayor, y vemos que es una variable muy influyente a la hora de determinar si sufren o no bullying.



Respecto a la variable acerca del sentimiento de soledad volvemos a apreciar algo parecido a lo que pasaba con los compañeros, ya que estas variables pueden estar relacionadas.



Por último, la variable de sufrir ciberbullying a primera vista no representa muy bien una posible relación. Ya que no sufrir ciberbullying sufriendo bullying tiene mayor porcentaje, pero no sufrir ni ciberbullying ni bullying tiene el mayor porcentaje. Al tener en cuenta que el número de datos de no sufrir bullying casi triplica a los que si sufren, vemos que los niños que sí sufren ciberbullying tanto si sufren o no bullying es prácticamente el mismo, por lo que es mucho mayor la proporción de niños que sufriendo bullying a su vez sufre ciberbullying que los que no.



## **CONCLUSIONES**

Tras realizar el estudio del modelo y el cálculo de las predicciones podemos concluir que los factores que están relacionados en cómo los niños se relacionan entre ellos con sufrir bullying. Antes de realizar el trabajo pensamos que el peso, el sexo o la situación familiar eran unos de los motivos más relevantes a la hora de predecir. Sin embargo, los resultados que hemos obtenido no han sido esos, de hecho, esas variables eran las que mayor AIC tenían, es decir menos influían.

Las variables que guardan mayor relación con el bullying son si los niños se han sentido solos, si los compañeros han sido amables, han sufrido una pelea física y si han sufrido cyberbullying.

Los niños que experimentan soledad pueden ser más propensos a convertirse en víctimas de bullying, ya que pueden ser percibidos como objetivos fáciles o vulnerables por los acosadores.

Además, la amabilidad de los compañeros juega un papel crucial en la prevención del bullying. Cuando los compañeros son amables y solidarios, se crea un ambiente escolar positivo en el que el acoso tiene menos probabilidades de ocurrir.

Los niños que han experimentado o participado en peleas físicas tienen una mayor probabilidad de exhibir comportamientos agresivos o intimidantes hacia otros, lo que aumenta su propensión a convertirse en acosadores.

Por último, en la era digital en la que vivimos, el cyberbullying también ha surgido como un importante factor relacionado con el bullying. El sufrimiento de cyberbullying puede tener un impacto significativo en la vida de los niños, y aquellos que han experimentado esta forma de acoso están más propensos a participar en el bullying tradicional.

En conclusión, para abordar eficazmente este problema, es fundamental promover la inclusión social, fomentar la empatía y la amabilidad, enseñar estrategias de resolución pacífica de conflictos y proporcionar apoyo y recursos para hacer frente al acoso en todas sus formas. Al trabajar juntos, podemos crear entornos seguros y respetuosos donde todos los niños puedan crecer y desarrollarse sin miedo al bullying.