

PROYECTO FINAL DE CURSO DE DATA SCIENCE DE IT ACADEMY

INTRODUCCIÓN

Uno de los principales motivos por los que he optado a especializarme en Ciencia de los Datos es mi convicción de que esta disciplina puede contribuir significativamente a mejorar la sociedad en diferentes ámbitos. Con buenos estudios y de datos y la creación de modelos se puede ayudar a tomar medidas y/o aplicar procesos que mejoren la calidad de vida de las personas.

Entre estos desafíos sociales he elegido realizar mi trabajo sobre el Bullying. Ya que es un problema que afecta a muchos niños en todo el mundo. Y más concretamente el Ciberbullying que está en constante aumento. Las consecuencias de estos fenómenos pueden ir desde algún trastorno psicológico como problemas sociales o de autoestima hasta el suicidio en el peor de los casos. Es un tema serio que merece toda nuestra atención.

El **bullying** es un fenómeno grupal de agresión intencional de uno o unos sobre otro u otros de forma reiterada y mantenida en el tiempo, en la que existe desequilibrio de poder entre quien agrede y es agredido, la víctima. El acoso escolar es también una conducta de transgresión moral, en la medida en que tanto el agresor como los observadores directos del fenómeno lo reconocen como una conducta inmoral e injusta.

Con la aparición de las nuevas tecnologías, el acceso a las redes y la disponibilidad de teléfonos móviles entre los más jóvenes ha aparecido un nuevo concepto relacionado con el bullying. El **ciberbullying**, también conocido como acoso cibernético. Es una forma de comportamiento hostil, intimidatorio o amenazador que ocurre a través de medios digitales, como internet y dispositivos electrónicos. Este fenómeno implica el uso de la tecnología para acosar, molestar, difamar, humillar o intimidar a otra persona.

Las formas de ciberbullying pueden variar y pueden incluir:

Acoso en línea: Envío repetido de mensajes amenazadores, hostiles o inapropiados a través de plataformas de redes sociales, correos electrónicos o mensajes de texto.

Suplantación de identidad: Crear perfiles falsos o utilizar información falsa para hacerse pasar por otra persona con el objetivo de difamar, ridiculizar o acosar.

Difusión de rumores: Compartir información falsa o malintencionada sobre alguien con el propósito de dañar su reputación.

Amenazas en línea: Enviar mensajes que contienen amenazas de violencia, daño físico o emocional.

Exclusión en línea: Excluir a alguien de grupos en línea, conversaciones o actividades, con el objetivo de hacerle sentir aislado.

El ciberbullying puede tener graves consecuencias emocionales y psicológicas para las personas afectadas. Puede ocurrir en diversas plataformas en línea, como redes sociales, foros, mensajería instantánea y correos electrónicos.

OBJETIVO DEL PROYECTO

El objetivo principal de este proyecto ha sido aplicar técnicas de Machine Learning de manera que puedan contribuir a la reducción de este problema, que está en constante aumento.

Antes de empezar a trabajar en el proyecto, estuve considerando de qué manera podían las técnicas de Machine Learning contribuir a aportar soluciones para reducir estas conductas entre los más jóvenes.

Los objetivos iniciales eran tres:

- 1) crear una predicción con un a regresión para poder predecir su evolución.
- 2) elaborar una clasificación de los grupos de personas en tres o cuatro (1 los que sufren bullying, 2 los acosadores, 3 los que ayudan a los acosadores, 4 los que no hacen nada).
- 3) crear un modelo en el que cuando se le introdujeran unos datos en concreto te dijera que probabilidades tienes de ser de 1 de estos 4 grupos.

Una vez establecidos los objetivos, el siguiente era es buscar conjuntos de datos que pudieran ayudar a conseguir estos objetivos.

Después de dedicar muchas horas a investigar en Internet y ponerme en contacto con organizaciones y fundaciones centradas en el Bullying y el Ciberbullying, me di cuenta de que los objetivos que había establecido para este proyecto resultaban bastante difíciles de alcanzar. No se trataba tanto de problemas técnicos directamente relacionados con el manejo de los datos, sino más bien de la complejidad para obtener acceso a conjuntos de datos adecuados.

Tengo constancia de que existen muchos datos sobre el tema provenientes de diferentes fuentes, he tenido dificultades para acceder a conjuntos de datos específicos. He tenido acceso a muchos estudios sobre el tema donde analizan sus datos y exponen sus conclusiones, pero lamentablemente no he tenido acceso a conjuntos de datos. No hay acceso a los datos. Estos organismos no tienen problemas a la hora de compartir sus informes y las conclusiones a las que llegan, pero si que son muy reticentes a la hora de compartir los conjuntos de datos de sus informes.

Finalmente, tuve la suerte de poder acceder a un dataset muy completo de la Comunidad de Andalucía creado por Universidad de Sevilla del Departamento de Psicología Evolutiva y de la Educación. Sus autores son: Esperanza del Rocío Espino Peñate y Rosario del Rey Alamillo.
<https://idus.us.es/handle/11441/130195>

PRESENTACIÓN DEL CONJUNTO DE DATOS:

Nombre del Dataset: Exploratory survey (2021) on bullying and cyberbullying, and its associated socio-emotional and familiar variables: [dataset]

Autores: Esperanza del Rocío Espino Peñate y Rosario del Rey Alamillo.

Publicado por: Universidad de Sevilla

Fecha de publicación: 2022

Resumen: Datos, anonimizados, de la encuesta administrada a jóvenes de entre 11 y 18 años de Andalucía, durante el año 2020. Este conjunto de datos incluye: variables sociodemográficas (género, edad, peso, altura, país de nacimiento, país de nacimiento de padre/madre, orientación sexual y uso de redes sociales virtuales), subescalas de victimización y agresión de la versión española del European Bullying Intervention Project Questionnaire (EBIPQ; Ortega-Ruiz et al., 2016) y subescalas de cibervictimización y ciberagresión de la versión española del European Cyberbullying Intervention Project Questionnaire (ECIPQ; Ortega-Ruiz et al., 2016); versión española de la Perceived Social Competence Scale II (Romera et al., 2016); Risky Online Behaviour Scale (Sasson & Mesch, 2016); Cyberbullying Victimization Emotional Impact Scale (CVEIS; Elipe et al., 2017); Coping Strategies Scale (Nacimiento & Mora-Merchán, 2014); Anger Regulation Scale (Bar-On & Parker, 2006); Socio-Emotional Competence Questionnaire (SECQ; Zhou & Ee, 2012); Multidimensional Scale of Perceived Social Support (MSPSS; Zimet et al., 1988, 2016); Self-Compassion Scale (Raes et al., 2011); Parental Control of Internet Use by Adolescents Questionnaire (Álvarez-García et al., 2019); and Need for Popularity Scale (Santor et al., 2000).

El alcance de este estudio es la Comunidad Autónoma de Andalucía y las franjas de edad que comprende es de los 11 a los 18 años.

El dataset consta de 174 variables y 1276 registros. El nivel de registros es bastante bueno, ya que proporciona bastantes valores con los que trabajar. Y nos podrá ayudar a tener unas predicciones mejores.

Se podría considerar que dentro del dataset hay varios “subdatasets”. Algunas variables están agrupadas como resultados de diferentes tipos de encuestas. Estas las explicaré brevemente cuando a medida que comente las variables del dataset.

A pesar de ser un dataset muy completo, no había apenas información sobre el significado de las variables. Esto me llevó a dedicar muchas horas en internet buscando información sobre estas. Para poder entender su significado y valor dentro del dataset.

DEFINICIÓN DE LAS VARIABLES

Como he comentado anteriormente, las variables están agrupadas por diferentes tipos de encuesta. Aquí las explicaré por agrupación.

PRIMERA PARTE del Dataset:

Variables de tipo general:

- **'Sex'**: es el género. Se escribe como 0 o 1.
- **'Sex_other'**: en el dataset no hay ningún valor. De manera que eliminaremos esta columna. Cuesta creer que esté vacío en todo el dataset. Pero así estaban los datos. De manera que no aporta nada y por dicho motivo lo eliminaremos.
- **'Age'**: Edad. Este dataset va desde los 11 a los 18. Los registros en blanco los sustituiremos por la moda de los valores agrupada por cursos.
- **'Course'**: Curso escolar. Va del 1 al 4 (que serían de 1 de ESO a 4 de Eso)
- **'Height'**: Altura del alumno en m.
- **'Weight'**: Peso del alumno en kgs.
- **'Country_birth'**: País de Nacimiento. 1= España, 2= otro **446 registros en blanco** casi 1/3 parte del dataset
- **'Country_family'**: País de origen de la familia. 1= España, 2= otro **455 registros en blanco**
- **'Country_father'**: País de origen del padre. Vacías = España , otros países
- **'Country_mother'**: País de origen de la madre. Vacías = España , otros países
- **'Physical_attraction'**: Género por el que se sienten atraídos. 1= hombres, 2= mujeres y 3=? Entiendo que 3 será ambos. **124 registros en blanco**
- **'Physical_attraction_i_dont_know'**: los que no saben. Valores= Asexual, Bisexual, Both, Questioning y vacías

El primer problema con el que nos encontramos en el dataset fue determinar el sexo de los alumnos. Lo describe como 0 e 1 pero no podemos indentificar cual es cada uno. Al tener el peso y la altura de los alumnos podremos hacer un describe groupby Sex.

SEGUNDA PARTE del Dataset

En esta parte del dataset se puede en cuantas redes sociales tienen cuenta los objetos encuestados. Esto nos dará una idea de su nivel de conexión y relación con las redes.

- **'Account_in_WhatsApp'**: tienen cuenta de WhatsApp. 1= Si/True, 0= No/False
- **'Account_in_Instagram'**: tienen cuenta de Instagram. 1= Si/True, 0= No/False
- **'Account_in_Twitter'**: tienen cuenta de Twitter. 1= Si/True, 0= No/False
- **'Account_in_Facebook'**: tienen cuenta de Facebook. 1= Si/True, 0= No/False
- **'Account_in_TikTok'**: tienen cuenta de TikTok. 1= Si/True, 0= No/False
- **'Account_in_Other'**: Tienen cuenta en otras redes. 1= Si/True, 0= No/False
- **'Other_social_network'**: Listado de diferentes redes

TERCERA PARTE del Dataset

Esta parte está relacionada con el comportamiento prosocial (acciones y conductas que tienen la intención de beneficiar a los demás o contribuir al bienestar de la sociedad). s.

Perceived Social Competence Scale (PSCS)

Es una herramienta que evalúa las habilidades de competencia social y el comportamiento prosocial en niños y jóvenes. Se centra en medir cómo los jóvenes perciben su propia competencia social, incluyendo comportamientos como ayudar a otros y mostrar interés. Las respuestas se dan en una escala de 0 a 4.

- 'Prosociality_1': "Ayudo a otras personas"
- 'Prosociality_2': ""Pregunto a otros si puedo ser de ayuda"
- 'Prosociality_3': "Muestro preocupación por los demás"
- 'Prosociality_4': "Muestro cuidado por los demás".
- 'Prosociality_5': "Brindo apoyo a los demás"

CUARTA PARTE del Dataset

- 'Peers_injunctive_norms_1',
- 'Peers_injunctive_norms_2', '
- Peers_injunctive_norms_3',

No he podido encontrar información sobre esta parte del dataset. Al no saber a qué hace relación y no poder explicar las variables he decidido eliminarlas del conjunto a trabajar.

QUINTA PARTE del Dataset

Esta parte es una encuesta separada en cuatro partes en la que se valora si han recibido acciones relacionadas con el bullying (1), si han llevado a cabo acciones relacionadas con el bullying (2), si han recibido acciones relacionadas con el ciberbullying (3) o si han llevado a cabo acciones relacionadas con el Ciberbullying (4).

Los valores en nuestro Dataset van de 0 a 4. Dónde 0 es nunca y 4 siempre.

BULLYING

Víctima:

- 'EBIPQ_V_1': Alguien me ha golpeado, me ha pateado o me ha empujado
- 'EBIPQ_V_2': Alguien me ha insultado
- 'EBIPQ_V_3': Alguien le ha dicho a otras personas palabras malsonantes sobre mi
- 'EBIPQ_V_4': Alguien me ha amenazado
- 'EBIPQ_V_5': Alguien me ha robado o roto mis cosas
- 'EBIPQ_V_6': He sido excluido o ignorado por otras personas
- 'EBIPQ_V_7': Alguien ha difundido rumores sobre mi

Acosador:

- 'EBIPQ_A_8': He golpeado, pateado o empujado a alguien
- 'EBIPQ_A_9': He insultado y he dicho palabras malsonantes a alguien
- 'EBIPQ_A_10': He dicho a otras personas palabras malsonantes sobre alguien
- 'EBIPQ_A_11': He amenazado a alguien
- 'EBIPQ_A_12': He robado o he estropeado algo de alguien

- 'EBIPQ_A_13': He excluido o ignorado a alguien
- 'EBIPQ_A_14': He difundido rumores sobre alguien

CIBERBULLYING

Víctima

- 'ECIPQ_V_1': Alguien me ha dicho palabras malsonantes o me ha insultado usando el email o SMS.
- 'ECIPQ_V_2': Alguien ha dicho a otros palabras malsonantes sobre mí usando internet o SMS.
- 'ECIPQ_V_3': Alguien me ha amenazado a través de mensajes en internet o SMS.
- 'ECIPQ_V_4': Alguien ha pirateado mi cuenta de correo y ha sacado mi información personal.
- 'ECIPQ_V_5': Alguien ha pirateado mi cuenta y se ha hecho pasar por mí
- 'ECIPQ_V_6': Alguien ha creado una cuenta falsa para hacerse pasar por mí.
- 'ECIPQ_V_7': Alguien ha colgado información personal sobre mí en internet.
- 'ECIPQ_V_8': Alguien ha colgado videos o fotos comprometidas mías en internet.
- 'ECIPQ_V_9': Alguien ha retocado fotos mías que yo había colgado en internet.
- 'ECIPQ_V_10': He sido excluido o ignorado de una red social o de chat.
- 'ECIPQ_V_11': Alguien ha difundido rumores sobre mí por internet.

Acosador:

- 'ECIPQ_A_12': He dicho palabras malsonantes a alguien o le he insultado usando SMS o mensajes en internet.
- 'ECIPQ_A_13': He dicho palabras malsonantes sobre alguien a otras personas en mensajes por internet o por SMS
- 'ECIPQ_A_14': He amenazado a alguien a través de SMS o mensajes en internet.
- 'ECIPQ_A_15': He pirateado la cuenta de correo de alguien y he robado su información personal.
- 'ECIPQ_A_16': He pirateado la cuenta de alguien y me he hecho pasar por él/ella.
- 'ECIPQ_A_17': He creado una cuenta falsa para hacerme pasar por otra persona.
- 'ECIPQ_A_18': He colgado información personal de alguien en internet.
- 'ECIPQ_A_19': He colgado videos o fotos comprometidas de alguien en internet.
- 'ECIPQ_A_20': He retocado fotos o videos de alguien que estaban colgados en internet.
- 'ECIPQ_A_21': He excluido o ignorado a alguien en una red social o chat.
- 'ECIPQ_A_22': He difundido rumores sobre alguien en internet.

Observando los datos vemos que hay 7 registros que tienen todos los valores 999 (vacíos) de la encuesta de Bullying y Cyberbullying. De manera que estos registros no nos interesan, ya que no nos ayudarán a decidir los factores que inciden en la detección del bullying. Habría dos opciones una darle a todo valor 0 que indicaría que no han sufrido ningún tipo de bullying o de cyberbullying. En este caso opté por eliminarlos.

En la métrica de Bullying hay dos mediciones la de víctima y la de acosador. Esto me permite extraer una puntuación de cada una de ellas. Pero lo primero que haremos será reemplazar los valores que siguen quedando vacíos en los registros, entre todas las columnas de bullying y cyberbullying son 94 registros con alguno de los valores en blanco. En este caso los sustituiremos por la moda de las columnas.

No todas las variables tendrían el mismo peso a la hora de computar, ya que no es lo mismo “haber dicho palabras malsonantes a alguien” que “piratear una cuenta de correo”. De manera que aplico una ponderación dependiendo de su “gravedad”.

SEXTA PARTE del Dataset

Cyberbullying Victimization Emotional Impact Scale (CVEIS; Elipe et al., 2017);

Hacen referencia al impacto emocional de las personas que sufren cyberbullying.

Estados de ánimo: (valorados del 0 al 4)

- **'Tense_nervous'**: Tenso/nervioso
- **'Animated'**: Animado
- **'Guilty'**: Culpable / Sentimiento de culpabilidad
- **'Energetic'**: Enérgico
- **'Scared_afraid'**: Asustado / tímido
- **'Angry_annoyed'**: Enojado / Molesto
- **'Satisfied_proud'**: Satisfecho / Orgullosos
- **'Lonely'**: Solitario
- **'Irritable_in_a_bad_mood'**: Irritable / De mal humor
- **'Ready_clear_headed'**: Despejado mentalmente
- **'Choleric_enraged'**: Colérico / Enfurecido
- **'Ashamed'**: Avergonzado
- **'Defenseless_helpless'**: Desprotegido / Impotente
- **'Determined_daring'**: Decidido/Audaz
- **'Depressed_sad'**: Deprimido / Triste
- **'Fed_up'**: Harto / Cansado de algo
- **'Active_alert'**: Activo / Alerta
- **'Jittery_worried'**: Nervioso / Preocupado

SÉPTIMA PARTE del Dataset

Validación de una escala para medir afrontamiento ante ciberagresiones

Esta parte de la encuesta sirve para medir las acciones que hace o podría hacer un joven frente a las ciberagresiones. Estaría segmentada en dos grandes grupos. 1) Búsqueda de apoyo social (5 primeras preguntas) y 2) Solución técnica (8 últimas preguntas). Dentro del segundo grupo se divide en dos grupos: 1) Solución técnica evasiva (preguntas 6,7 y 8) y 2) Solución técnica proactiva (9-13). La escala va de 0 (nunca) a 4 (siempre).

- **'Coping_1'**: Cuento con alguien que escucha mi problema
- **'Coping_2'**: Dejo que mis amigos me echen la mano
- **'Coping_3'**: Hablo con una persona de confianza sobre mi problema
- **'Coping_4'**: Paso tiempo con mis amigos
- **'Coping_5'**: Pido consejo a una persona de mi respeto
- **'Coping_6'**: Bloqueo al agresor que entra en contacto conmigo
- **'Coping_7'**: Borro el mensaje que me daña
- **'Coping_8'**: Borro o pido al administrador del sitio web que borre la publicación que me daña
- **'Coping_9'**: Cambio contraseña, nombre de usuario, cuenta de correo
- **'Coping_10'**: Cambio mi número de teléfono
- **'Coping_11'**: Dejo de usar la aplicación o tecnología por medio de la cual me perjudican
- **'Coping_12'**: Elimino mensajes anónimos sin leerlos
- **'Coping_13'**: Elimino mi cuenta (red social, correo electrónico, etc.)

OCTAVA PARTE del Dataset

No he encontrado los valores de esta encuesta. Pero a lo largo de mi trabajo con este dataset, he aprendido a ver como trabaja la encuesta. De manera que me atrevo a decir que va de menos a mas.

Pero al no saber las preguntas, también podría ser que algunas fueran de carácter positivo y otras de carácter negativo. De manera que mientras no podamos encontrar más información al respecto no las tendremos en cuenta.

- 'Anger_regulation_1',
- 'Anger_regulation_2',
- 'Anger_regulation_3',
- 'Anger_regulation_4',
- 'Anger_regulation_5',
- 'Anger_regulation_6',
- 'Anger_regulation_7',
- 'Anger_regulation_8',

NOVENA PARTE del Dataset

Development and Validation of the Social Emotional Competence - Questionnaire (SECQ)

Sirve para evaluar su desarrollo socioemocional y proporcionar intervenciones apropiadas en el desarrollo infantil y adolescente. Los valores para esta parte de la encuesta van de 0 a 4 donde (0= nunca y 4= siempre)

Conciencia Personal: reconocer e identificar las propias fortalezas y debilidades, sentimientos y emociones, y comprender cómo pueden afectar el rendimiento personal.

- 'SECQ_1': Sé lo que estoy pensando y haciendo.
- 'SECQ_2': Comprendo por qué hago lo que hago.
- 'SECQ_3': Comprendo mis estados de ánimo y sentimientos.
- 'SECQ_4': Sé cuándo estoy de mal humor.
- 'SECQ_5': Puedo leer las expresiones faciales de las personas cuando están enojadas.

Conciencia Social: la capacidad de interpretar las señales de otras personas y comprender y responder adecuadamente a sus sentimientos. Esto está estrechamente vinculado con la empatía, la capacidad de compartir el estado emocional de otra persona y, por lo tanto, relacionarse mejor con ella.

- 'SECQ_6': Reconozco cómo se sienten las personas al mirar sus expresiones faciales.
- 'SECQ_7': Es fácil para mí entender por qué las personas sienten lo que sienten.
- 'SECQ_8': Si alguien está triste, enojado o feliz, creo que sé en qué están pensando.
- 'SECQ_9': Comprendo por qué las personas reaccionan de la manera en que lo hacen.
- 'SECQ_10': Si un amigo está molesto, tengo una buena idea de por qué.

Autogestión: la capacidad de gestionar los impulsos y emociones propios. La autorregulación de las emociones es importante para desarrollar relaciones cercanas, tener éxito en el trabajo y mantener la salud física.

- 'SECQ_11': Puedo mantener la calma en situaciones estresantes.
- 'SECQ_12': Mantengo la calma y supero la ansiedad en situaciones nuevas o cambiantes.
- 'SECQ_13': Mantengo la calma cuando las cosas salen mal.
- 'SECQ_14': Puedo controlar la forma en que me siento cuando algo malo sucede.
- 'SECQ_15': Cuando estoy molesto con alguien, esperaré hasta que me haya calmado antes de discutir el problema.

Gestión de Relaciones:

- 'SECQ_16': Siempre me disculpo cuando lastimo a mi amigo involuntariamente.
- 'SECQ_17': Siempre trato de consolar a mis amigos cuando están tristes.
- 'SECQ_18': Trato de no criticar a mi amigo cuando discutimos.
- 'SECQ_19': Soy tolerante con los errores de mis amigos.
- 'SECQ_20': Defiendo mis propios intereses sin menospreciar a los demás.

Toma de Decisiones Responsable: la capacidad de considerar factores éticos, de seguridad y sociales al tomar decisiones, de modo que las personas puedan lidiar responsablemente con situaciones académicas y sociales cotidianas y contribuir al bienestar de su escuela y comunidad.

- 'SECQ_21': Al tomar decisiones, tengo en cuenta las consecuencias de mis acciones.
- 'SECQ_22': Aseguro que haya más resultados positivos al tomar una elección.
- 'SECQ_23': Evalúo las fortalezas de la situación antes de decidir mi acción.
- 'SECQ_24': Considero los criterios elegidos antes de hacer una recomendación.
- 'SECQ_25': Pienso en las fortalezas y debilidades de la estrategia antes de decidir usarla.

DÉCIMA PARTE del Dataset

Multidimensional Scale of Perceived Social Support+

Esta parte cuantifica si se sienten apoyados por otras personas, ya sea familia, amigos y otros.

0= totalmente en desacuerdo y 4= totalmente de acuerdo

- 'ApoyoSF_1_others': Hay una persona especial que está cerca cuando lo necesito.
- 'ApoyoSF_2_others': Hay una persona especial con la que puedo compartir alegrías y tristezas.
- 'ApoyoSF_3_family': Mi familia realmente intenta ayudarme.
- 'ApoyoSF_4_family': Recibo la ayuda y el apoyo emocional que necesito de mi familia.
- 'ApoyoSF_5_others': Tengo una persona especial que es una verdadera fuente de consuelo para mí.
- 'ApoyoSF_6_friends': Mis amigos realmente intentan ayudarme.
- 'ApoyoSF_7_friends': Puedo contar con mis amigos cuando las cosas van mal.
- 'ApoyoSF_8_family': Puedo hablar sobre mis problemas con mi familia.
- 'ApoyoSF_9_friends': Tengo amigos con los que puedo compartir mis alegrías y tristezas.
- 'ApoyoSF_10_others': Hay una persona especial en mi vida que se preocupa por mis sentimientos.
- 'ApoyoSF_11_family': Mi familia está dispuesta a ayudarme a tomar decisiones.
- 'ApoyoSF_12_friends': Puedo hablar sobre mis problemas con mis amigos.

UNDÉCIMA PARTE del Dataset

Test de Autocompasión

Este test permite ver el grado de autocompasión que una persona tiene hacia ella misma. A mejor autocompasión más dificultad de poder ser víctima de Bullying o Ciberbullying.

0= nunca y 4= siempre

- **'Selfcompassion_1'**: Cuando fallo en algo importante para mí, me veo consumido por sentimientos de incompetencia. (-)
- **'Selfcompassion_2'**: Intento ser comprensivo y paciente con aquellos aspectos de mi personalidad que no me gustan. (+)
- **'Selfcompassion_3'**: Cuando algo doloroso sucede, trato de tener una visión equilibrada de la situación. (+)
- **'Selfcompassion_4'**: Cuando me siento deprimido, tiendo a pensar que la mayoría de las personas probablemente son más felices que yo. (-)
- **'Selfcompassion_5'**: Intento ver mis fallas como parte de la condición humana. (+)
- **'Selfcompassion_6'**: Cuando estoy pasando por un momento muy difícil, me doy el cuidado y la ternura que necesito. (+)
- **'Selfcompassion_7'**: Cuando algo me perturba, trato de mantener mis emociones en equilibrio. (+)
- **'Selfcompassion_8'**: Cuando fallo en algo que es importante para mí, tiendo a sentirme solo en mi fracaso. (-)
- **'Selfcompassion_9'**: Cuando me siento deprimido, tiendo a obsesionarme y enfocarme en todo lo que está mal. (-)
- **'Selfcompassion_10'**: Cuando me siento incompetente de alguna manera, trato de recordarme a mí mismo que los sentimientos de incompetencia son compartidos por la mayoría de las personas. (+)
- **'Selfcompassion_11'**: Desapruebo y juzgo mis propias fallas e incompetencias. (-)
- **'Selfcompassion_12'**: Soy intolerante e impaciente con aquellos aspectos de mi personalidad que no me gustan. (-)

DUODÉCIMA PARTE del Dataset

Parental Control of Internet Use by Adolescents

Mide la restricción y la supervisión 0=Totalmente falso; 1=Más bien falso; 2=Más bien cierto; 3=Totalmente cierto

- **'Parental_control_1_M'**: Cuando accedo a Internet en mi tiempo libre, mis padres me vigilan y echan un vistazo a la pantalla
- **'Parental_control_2_M'**: Mis padres conocen las claves de acceso a mis cuentas de correo electrónico, redes sociales o programas de mensajería
- **'Parental_control_3_R'**: En casa me han puesto algunas normas sobre lo que puedo o no puedo hacer en Internet.
- **'Parental_control_4_R'**: En casa me limitan los contenidos a los que puedo acceder en Internet, mediante filtros en el ordenador.
- **'Parental_control_5_R'**: Mis padres me limitan las horas de uso de Internet (ya sea de palabra o configurando el ordenador).
- **'Parental_control_6_M'**: Mis padres revisan mis perfiles en las redes sociales (Tuenti, Facebook, Twitter, Instagram, por ejemplo).
- **'Parental_control_7_M'**: Mis padres conocen mis listas de contactos.

DECIMOTERCERA PARTE del Dataset

- **'Behavioural_control_1'**,
- **'Behavioural_control_2'**,
- **'Behavioural_control_3'**,
- **'Behavioural_control_4'**,

No he podido encontrar información sobre esta parte del dataset. Al no saber a qué hace relación y no poder explicar las variables he decidido eliminarlas del conjunto a trabajar.

DECIMOCUARTA PARTE del Dataset

Need for Popularity Scale (Santor et al., 2000)

- **'Popularity_1'**: He hecho cosas para volverme más popular, incluso cuando significaba hacer algo que normalmente no haría.
- **'Popularity_2'**: He descuidado a algunos amigos por lo que otros podrían pensar.
- **'Popularity_3'**: En ocasiones, he ignorado a algunas personas para ser más popular con otras.
- **'Popularity_4'**: Haría casi cualquier cosa para evitar que me vean como un 'perdedor'.
- **'Popularity_5'**: Es importante que la gente piense que soy popular.
- **'Popularity_6'**: En ocasiones, he salido con personas solo porque eran populares.
- **'Popularity_7'**: He comprado cosas porque estaban de moda.
- **'Popularity_8'**: A veces, he cambiado la forma en que me visto para ser más popular.
- **'Popularity_9'**: He sido amigo de algunas personas solo porque a otros les caían bien.
- **'Popularity_10'**: He ido a fiestas solo para ser parte de la multitud.
- **'Popularity_11'**: A menudo hago cosas solo para ser popular entre las personas de la escuela.
- **'Popularity_12'**: En ocasiones, he salido con algunas personas para que otros no pensarán que era impopular.

CONCLUSIÓN

Una vez analizadas todas las variables de las que he conseguido información y el dataset en su conjunto, soy consciente de que el objetivo de crear una predicción con un a regresión. Pero los objetivos relacionados con la clasificación y la clusterización si que se podrán llevar a cabo sin problemas.

Este análisis me ha hecho cambiar de perspectiva sobre mi proyecto. En un inicio mi intención era la de centrarme en el Bullying en general. Pero los datos que tengo están más relacionados con el Ciberbullying, de manera que tiene más sentido focalizarme en el Ciberbullying. Pero será interesante ver la relación existente entre el Bullying y el Ciberbullying.

Para poder hacer un análisis más a fondo, duplicaremos el dataset y crearemos totales de cada una de las partes del Dataset para poder entender mejor nuestros valores.

Para el modelado, usaremos el dataset como está sin las columnas que no entendemos. Luego lo limpiaremos e imputaremos valores a todos aquellos registros que contengan valores vacíos.

Habrà que ver de qué manera reemplazamos valores sin que comprometa a nuestro modelo.