

# Clasificador de personalidades basado en clustering para prevención de ciber-acoso escolar

Isabel Prado<sup>1</sup>, Cristina Puentes<sup>1</sup>, Rafael Palacios<sup>1</sup>, José Ángel Olivas<sup>2</sup>

<sup>1</sup> Dpto. De telemática y Computación, ICAI, Universidad Pontificia Comillas, Madrid, España. { ipradosanz@gmail.com; cristina.puentes@comillas.edu; rafael.palacios@iit.comillas.edu }

<sup>2</sup> Dpto. de Tecnologías y Sistemas de la Información, Universidad de Castilla La-Mancha, Ciudad Real. joseangel.olivas@uclm.es

**Resumen.** El acoso escolar puede prevenirse analizando la personalidad de los alumnos y detectando qué parejas de alumnos tienen mayor o menor riesgo para desarrollar conductas de acoso. El trabajo que se muestra a continuación consiste en desarrollar un clasificador de personalidades lo más automático posible, para que sea una herramienta ágil y útil en los colegios. Se ha llevado a cabo una aplicación web para recoger y almacenar encuestas de alumnos, un clasificador de personalidades basado en técnicas de clustering, un analizador de parejas de niños basado en reglas de decisión, y una herramienta web de consulta de resultados para el profesor, tutor o psicopedagogo. Los resultados del clasificador han sido validados en un colegio obteniendo unas tasas de acierto del 80%.

**Palabras Clave:** Clustering, Kmeans, Personalidad, Bullying, Ciber-acoso

## 1 Introducción

El acoso escolar es uno de los problemas que más preocupa en los centros educativos, especialmente porque desde que se ha extendido el uso de las tecnologías de la comunicación y de las redes sociales por parte de los menores, el acoso (ahora ciberacoso) ha crecido enormemente. Según datos de la organización *Save the Children* sobre bullying en España [1][2], 9,3% de los niños entre 12 y 16 años sufren acoso.

Una manera de prevenir el acoso es clasificar a los niños de acuerdo a su tipo de personalidad, que puede ser: **agresivo, pasivo o asertivo** [3]. Un análisis de personalidad completo de toda la clase puede utilizarse para detectar aquellas parejas de alumnos que puedan tener mayor o menor riesgo a desarrollar una situación de acoso. La clasificación de personalidad la pueden realizar psicólogos mediante test especiales, y es un proceso que suele requerir mucho tiempo y mucho trabajo. El presente estudio consiste en el desarrollo de unas herramientas telemáticas online, cuyos principales objetivos son el análisis de las personalidades de los alumnos de un aula en un colegio, la detección de posibles casos de exclusión entre los alumnos de la misma y la creación de recomendaciones a los tutores para un mejor clima en el aula, siendo así posible minimizar los casos de bullying, acoso escolar o exclusión.

El funcionamiento del proyecto es el siguiente que se desglosa a continuación: Los alumnos contestan al test de personalidad y socioemocionograma a la vez durante las tutorías con su profesor principal o tutor. La mayoría de las preguntas son de tipo multirespuesta, otras son preguntas relacionadas con la identificación o selección de alumnos del mismo grupo o curso escolar.

La información recibida sobre los alumnos es analizada por un algoritmo de personalidad que determina y clasifica al alumno según el tipo de personalidad del mismo, el grado de empatía y su habilidad para trabajar en equipo, y los compañeros por los que es aceptado y rechazado.

Dicho algoritmo de personalidad está basado en un algoritmo de clusterización utilizando el método Kmeans, buscando tres clusters de personalidad, basados en los tres tipos de personalidad agresivo, pasivo y asertivo.

El coordinador o tutor recibe un informe detallado sobre la clase y sobre cada alumno en la plataforma web diseñada para profesores. La información que se extrae es, en el informe general de la clase, los alumnos que pueden estar en peligro de exclusión y las parejas que se recomiendan no formar para evitar conflictos en la clase. Por otro lado, la información recibida por alumnos hace relación a la aceptación y rechazo del alumno, tanto académicamente como emocionalmente.

De esta manera, y soportado por la robustez de la arquitectura, los pedagogos y los tutores de los alumnos, a través de cualquier dispositivo (ordenador, tablet o móvil) podrán obtener una información detallada y accesible en cualquier momento sobre la situación de un grupo de estudiantes, ya que al ser una herramienta online y gratuita, los tests pueden ser realizados por el profesor siempre y cuando se estime oportuno, de manera mucho más rápida y eficiente que las soluciones actuales.

A nivel empresarial, plantea un innovador método de gestión de alumnos, trabajadores y equipos, motivado por la personalidad del encuestado, su aceptación dentro del grupo y su capacidad tanto de trabajar en equipo como su empatía, lo que resulta altamente beneficioso para centros escolares con muchos alumnos o, extrapoliándolo a la empresa, con equipos de trabajo.

## **2 Desarrollo del proyecto**

Ha sido necesario el desarrollo de una plataforma o aplicación online tanto como para la realización de los tests a los alumnos como para obtener la información por parte de los profesores y mostrar los resultados del tratamiento de los datos obtenidos y las recomendaciones sugeridas por los diversos algoritmos utilizados.

### **2.1 Recogida y muestra de la información**

La recogida de la información se hace a través de un test compuesto por 28 preguntas

a cada uno de los alumnos. Se ha tratado de utilizar un entorno amigable y de cercanía con los alumnos, es por esto que en el test se ha cuidado el detalle de los dibujos de las preguntas, que el interfaz sea intuitivo y que el lenguaje sea claro y adecuado a la edad de los encuestados. Para ello se ha requerido la ayuda de varios expertos en la materia, profesores, psicólogos y pedagogos de varios centros educativos de la Comunidad de Madrid.



**Fig. 1.** Interfaz para los tests de los alumnos.

El contenido de las respuestas mide diversos parámetros, desde la empatía y el trabajo en equipo que han sido obtenidas mediante un algoritmo basado en reglas de intuición, como las preguntas utilizadas para la personalidad, a las que se les aplica la distancia euclídea a los centroides de los tres distintos clusters como se detallará más adelante.

Se ha programado una aplicación web para el profesor para facilitar la muestra de los datos obtenidos, conclusiones y recomendaciones. En ella se detalla tanto alumno por alumno como en el conjunto de la clase las conclusiones sacadas tras el procesamiento de los datos, es decir, si la personalidad es agresiva, pasiva o asertiva; si el grado de empatía es alto, medio o bajo y si la capacidad de trabajo en equipo es alta, media o baja. Además, se cuantifica el porcentaje de aceptación y rechazo que tiene el alumno en la clase, tanto a nivel académico como emocionalmente. Los datos mostrados a nivel general de la clase son las alertas obtenidas por exclusión, y la recomendación de parejas que no deberían hacerse entre el conjunto de alumnos de la clase, basándose en los resultados del cluster de personalidad Kmeans y en reglas basadas en el conocimiento de varios expertos entrevistados en la materia.

Luna	Resumen de la clase
1. Fecha última modificación 2017-04-21	Ojalí Riesgo de exclusión del siguiente alumno: Gonzalo
2. Personalidad con rasgos assertivo	Ojalí Riesgo de exclusión del siguiente alumno: Sofía
3. Grado de empatía alto	No se recomienda la pareja entre Gonzalo y Alvaro
4. Capacidad de trabajo en equipo medio	No se recomienda la pareja entre Gonzalo y Lucas
5. Es rechazado emocionalmente por 0 compañeros, el 0% de la clase	No se recomienda la pareja entre Gonzalo y Alba
6. Es rechazado a nivel académico por 2 compañeros, el 12.5% de la clase Rechazado académicamente por: Alvaro Sergio	No se recomienda la pareja entre Gonzalo y Carlota
7. Es aceptado emocionalmente por 6 compañeros, el 37.5% de la clase Elegido emocionalmente por: Sara	No se recomienda la pareja entre Gonzalo y Alejandro
	No se recomienda la pareja entre Gonzalo y Carlos
	No se recomienda la pareja entre Gonzalo y Paula

**Fig. 2.** Interfaz del profesor, ejemplo del resumen por alumno e Interfaz del profesor, ejemplo del resumen por clase.

## 2.2 Tratamiento de los datos

Los datos obtenidos del test se han recogido en una matriz de 28 preguntas por 45 muestras. De las 28 preguntas se han extraído los resultados contestados por las primeras 24 preguntas, ya que las últimas 4 preguntas tienen como respuesta nombres de personas en la clase que cada encuestado acepta o rechaza.

Se ha analizado la varianza en las contestaciones de cada uno de los encuestados y se han descartado las preguntas con baja varianza, de manera que los datos de las preguntas que se analizan no sean redundantes entre sí.

Las primeras 15 preguntas han sido destinadas al estudio de la empatía y la capacidad de trabajo en equipo, las preguntas recogidas entre la 16 y la 24 se han destinado al análisis de las personalidades mediante el algoritmo de Kmeans, ya que esperamos que los datos puedan ser agrupados en tres clusters.

Por último, las respuestas restantes recogen los alumnos a los que cada uno de los preguntados en el test rechaza o acepta dentro de un aula tanto emocional como académicamente. Estos datos serán tenidos en cuenta para analizar el porcentaje de la clase que acepta o rechaza a dicha persona y alertar al profesor en caso de un alto rechazo de un alumno por sus compañeros.

### 2.2.1 Algoritmo de clusterización

Se ha introducido una matriz de 6 preguntas por 45 muestras en el algoritmo Kmeans implementado en Matlab [4][5]. Para ello las columnas de la matriz introducida

corresponden a las variables o preguntas del test, mientras que las filas corresponden a las 45 muestras extraídas del test, La muestra de respuestas, representada a continuación, tiene como equivalente el número 1 como haber contestado “a”, 2 a “b”, 3 a “c”, etc. De manera que el algoritmo pueda analizar las respuestas aunque éstas sean letras y no números, como se muestra a continuación.

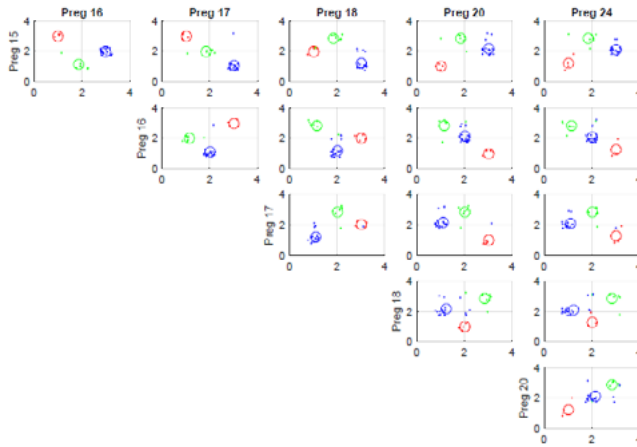
	preg15	preg16	preg17	preg18	preg19	preg20	preg24
3	2	1	1	1	2	2	2
3	2	1	1	2	2	2	2
2	1	2	3	3	3	3	3
2	1	2	3	2	3	3	3
3	2	1	1	2	2	2	2
1	3	3	2	2	1	1	1
1	3	3	2	2	1	1	1
3	2	1	1	2	2	2	2
3	2	1	1	2	2	2	2
2	1	2	3	2	3	3	3
3	2	1	2	2	2	2	2
2	1	2	3	2	3	3	3
3	2	1	1	2	2	2	2
3	2	1	1	2	2	2	2
3	2	1	1	2	2	2	2
1	3	3	2	2	1	1	1
3	2	1	1	2	2	2	2
2	1	2	3	2	2	2	2
3	2	1	1	2	2	2	2
3	2	1	1	2	2	2	2
3	2	1	1	2	2	2	2
2	1	2	3	2	3	3	3
1	2	2	2	2	3	3	3
1	3	3	2	1	1	2	2
3	2	1	2	2	3	3	3
3	2	3	2	2	2	2	2
3	2	1	1	2	3	2	2
3	2	1	2	2	2	3	3
3	2	1	1	2	3	2	2

**Fig. 3.** Matriz de resultados a las preguntas incluidas en el algoritmo Kmeans, donde 1 corresponde a haber contestado a, 2 a b y así sucesivamente.

El resultado del algoritmo Kmeans es el vector de 45x1 que contiene los índices de los clusters de cada observación, por otro lado, devuelve una matriz de 3x6 que forma los centroides de los tres clusters en seis dimensiones.

A partir de este momento, cualquier nueva respuesta al test, realizará la distancia euclídea a cada uno de los centros de los tres clusters. El centro que esté a menor distancia del nuevo punto será el centro del cluster al que pertenezca la nueva muestra o respuesta del test. Tampoco se descarta la posibilidad de volver a correr el algoritmo Kmeans con un mayor conjunto de muestras para ajustar mejor las posiciones de los centroides.

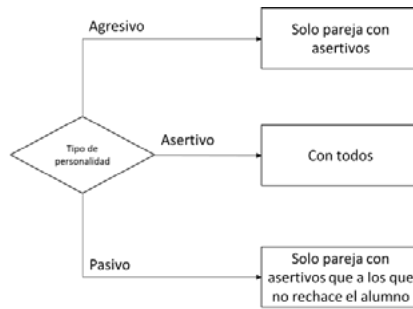
A continuación, se ha representado gráficamente la dispersión de las respuestas de los alumnos en proyecciones de cada pareja de preguntas. Se han coloreado las clases (personalidades) a las que pertenece cada muestra y se han representado con círculos los centroides de cada clase. Se puede observar que ninguna de las gráficas permite separar mediante reglas sencillas las muestras de las 3 clases, lo que demuestra la necesidad de utilizar herramientas como el clustering en una dimensión 6 para el análisis.



**Fig. 4.** Representación gráfica de las variables y sus centroides por colores, para la comprobación de la dispersión de los datos recogidos.

### 2.2.2 Clasificador basado en reglas

Con el conocimiento obtenido mediante las entrevistas a expertos en psicología, pedagogía y tutores de alumnos, se ha recogido una base de reglas con las que agrupar las parejas de los grupos de alumnos, basándolas en la personalidad de los alumnos y la afinidad o rechazo de los mismos.



**Fig. 5.** Árbol de decisión pareja.

Se ha construido el anterior árbol de decisión según las entrevistas de expertos, donde la primera recomendación es la utilización como “comodín” o pareja perfecta a una persona asertiva, haya rechazo entre las personas de la pareja o no. Por otro lado, la siguiente regla extraída de las recomendaciones de los profesionales entrevistados en la materia se resume en no emparejar personalidades pasivas y agresivas, o pasivas con asertivos que son en el test rechazados por el propio alumno con personalidad pasiva.

Por último, se ha decidido que los alumnos que recibieran más de un 50% de rechazos por los demás alumnos del aula debían ser notificados al profesor para obtener un trato más vigilado y hacer un seguimiento de las causas y la progresión del alumno.

### 3 Resultados

Los resultados recogidos tras las realizaciones de los diversos test y el posterior tratamiento de los datos según los algoritmos de clusterización y clasificación por reglas han sido muy positivos.

Para la evaluación de los resultados se ha entrevistado al tutor de los alumnos de la clase y se le ha pedido clasificar personalidades y afinidad entre los alumnos, así como la capacidad de trabajo en equipo. Para la evaluación por parte del profesor de las personalidades se ha designado los adjetivos agresivo, asertivo o pasivo. Así mismo, para las parejas que NO se recomendarían en el grupo, el profesor ha rellenado una matriz en blanco con todos los alumnos como filas y como columnas para rellenar las parejas. Por último, tanto el grado de empatía como de trabajo en equipo, se han utilizado para clasificarlos por el profesor los mismos que se muestran en los resultados del algoritmo, es decir, alto, medio y bajo.

- *Resultados de las personalidades:* Los resultados de las personalidades tras la entrevista con el profesor han recibido un acierto de un 80%, de los restantes 20%, tras la entrevista con el profesor se ha estimado que un 12% son semi-verdaderas, es decir, si se hubiera buscado una personalidad dominante y una tendencia como segunda personalidad, se podrían haber dado como válidas. Es decir, una persona puede ser asertiva con tendencia agresiva según el experto.
- *Resultados de reglas trabajo en equipo y empatía:* Con respecto a las reglas de empatía y trabajo en equipo, se han obtenido resultados parejos de 87% de aciertos.
- *Resultado reglas parejas NO recomendadas:* Las parejas NO recomendadas han sumado las que más éxitos en la balanza de aciertos de resultados, obteniendo un 87.5% de aciertos, el más elevado de los tres resultados de los algoritmos.

### 4 Conclusiones

Se han desarrollado un conjunto de aplicaciones web y una metodología de análisis mediante técnicas de clustering que permite analizar rápidamente una clase de un colegio. El sistema desarrollado ha mostrado unos resultados muy buenos en las pruebas realizadas y está listo para ser aplicado a gran escala de manera sencilla.

Existen varias mejoras pendientes, especialmente en la parte de evaluación de resultados, por la complejidad relativa a penalizar errores de diferente manera ya que no es lo mismo confundir un perfil pasivo con agresivo que con asertivo.

Los resultados obtenidos de los algoritmos implican que, pese a la posible escasez de muestras, el rumbo de los mismos es positivo y es por esto que un futuro afinamiento de los datos y por concluyente de las reglas y algoritmos garantizará un éxito seguro en el objetivo principal que se había planteado al elaborar el consiguiente proyecto.

### **Agradecimientos.**

Este trabajo ha sido parcialmente financiado por la Agencia Estatal de Investigación (AEI) y el Fondo Europeo de Desarrollo Regional (FEDER) dentro del proyecto MERINET:TIN2016-76843-C4-2-R (AEI/FEDER, UE).

### **Referencias**

- [1] A. Sastre, J. Calmaestra, A. Escorial, P. García, C. del Moral, C. Perazzo, T. Ubrich, “Yo a eso no juego. Bullying y Cyberbullying en la infancia”, Save the Children. Feb. 2016. [https://www.savethechildren.es/sites/default/files/imce/docs/yo\\_a\\_eso\\_no\\_juego.pdf](https://www.savethechildren.es/sites/default/files/imce/docs/yo_a_eso_no_juego.pdf)
- [2] “Uno de cada 10 alumnos asegura que ha sufrido acoso escolar”, El País, 2016. [http://politica.elpais.com/politica/2016/02/18/actualidad/1455799539\\_145548.html?rel=mas](http://politica.elpais.com/politica/2016/02/18/actualidad/1455799539_145548.html?rel=mas)
- [3] O. Castanyer, *La asertividad, expresión de una sana autoestima*. Desclee de Brouwer, S.A. (1997)
- [4] G. Brassard, P. Bratley: *Fundamentos de Algoritmia*, Prentice Hall (1997)
- [5] D. Arthur, S. Vassilvitskii. “K-means++: The Advantages of Careful Seeding.” *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. (2007), pp. 1027–1035.