

# Análisis Cluster y Sistemas de Decisión para la mejora del Rendimiento Académico en la Enseñanza Obligatoria

Antonio Muñoz-Ledesma<sup>1</sup> and José M. Cadenas<sup>2</sup>

<sup>1</sup> Instituto de Educación Secundaria “SANJE”  
Alcantarilla. Murcia, España

Dpto. Didáctica de las Ciencias Matemáticas y Sociales  
Facultad Educación. Universidad de Murcia. Murcia, España

<sup>2</sup> Dpto. Ingeniería de la Información y las Comunicaciones  
Universidad de Murcia. Murcia, España

{amledesma, jcadenas}@um.es

**Resumen** Principalmente, en la enseñanza obligatoria interesa identificar las acciones adecuadas al estudiante para que mejore su rendimiento académico. Este nivel de enseñanza es crucial en su desarrollo personal y en su futuro académico. Mediante un análisis cluster y de componentes principales exploramos los datos que nos proporcionan dos centros de Educación Secundaria para conocer los posibles estados y relaciones con respecto a las acciones a tomar para una mejora del rendimiento. A partir de esta exploración y mediante una técnica predictiva se construye un sistema de decisión para que los departamentos docentes puedan proporcionar la acción que se debe aplicar al estudiante.

**Keywords:** Rendimiento Académico, Minería de Datos, Clustering, Reglas para la Toma de Decisiones.

## 1 Introduction

En el área de la Educación encontramos diversas aplicaciones de las técnicas de Minería de Datos (MD) al análisis del Rendimiento Académico (RA) de los estudiantes. La mayoría de estos trabajos se han centrado en el primer curso de universidad en los que se trata de medir el riesgo al fracaso. Unos trabajos hacen uso de los exámenes de ingreso, tests de inteligencia, de motivación, notas del preuniversitario, factores de tipo socioeconómico, características personales, etc., y suelen utilizar técnicas de MD como: árboles de clasificación/regresión (combinándolos en algunos casos con análisis cluster) para buscar predictores, para identificar posibles abandonos y a los que necesitan una atención especial, [2,3,16]. En otros trabajos se utiliza el análisis cluster para analizar las características del perfil de aprendizaje de los estudiantes y así poder identificar el estilo de aprendizaje dominante y adecuar sus estrategias de enseñanza, [4].

En estos trabajos de la Enseñanza Universitaria se observa una preocupación:

- localizar predictores para la predicción del RA, tratar de detectar si un estudiante eligió correctamente la carrera adecuada y así evitar abandonos;
- detectar perfiles de aprendizaje en el estudiante y así mejorar el proceso de aprendizaje.

Mientras que en la Enseñanza Secundaria Obligatoria (ESO) existen escasos estudios dirigidos al RA, y principalmente lo que interesa es identificar las acciones adecuadas a cada estudiante para que mejore su RA y así conseguir éxito en la terminación de sus estudios. Este nivel de enseñanza es crucial en el desarrollo personal del estudiante y en su futuro académico, y es donde se centran algunos de estos estudios. En [15] realizan un estudio del perfil del estudiante repetidor y no repetidor y su incidencia en los resultados académicos. En [11,12] analizan la importancia e influencia de las asignaturas de Lengua y Matemáticas en el RA, además, construyen modelos predictivos (árboles de decisión) que indican la influencia de las asignaturas de Matemáticas y Lengua sobre las demás asignaturas. En [13] inician un estudio para buscar relaciones dependientes de las acciones a realizar con el estudiante.

Como objetivo principal de este trabajo hemos pretendido analizar los datos de un conjunto de estudiantes de ESO de Murcia (España) mediante un análisis de componentes principales y un análisis cluster e identificar patrones de correlación y homogeneidad entre atributos y estudiantes. A partir de este análisis se construirá un sistema basado en reglas que informe, para cada estudiante, la acción concreta que se le debe aplicar. El fin es construir un sistema de ayuda a la toma de decisiones por los deptos. implicados.

## 2 Construyendo un conjunto de datos de alumnos

De las siete Inteligencias Múltiples que se plantean en [7], centramos el estudio en dos de ellas que afectan de forma decisiva a dos asignaturas que son fundamentales para el buen rendimiento de los estudiantes de la ESO: la Lengua y las Matemáticas, [1].

Siguiendo las indicaciones en [10] medimos y extraemos información relevante para nuestro objetivo. A estudiantes de dos Institutos de Educación Secundaria (IES) de la región de Murcia (España) se les miden habilidades implícitas a cada una de esas inteligencias, lingüística y lógico-matemática, complementando la información que aportan las primeras notas del estudiante en Lengua y en Matemáticas en su curso actual y las notas medias de los dos cursos anteriores. La medición de las habilidades, autoconcepto y su grado de adaptación se realizan mediante los test psicométrico BADyG-renovado, AF5 y TAMAI. Por tanto, se utilizan tres fuentes de información:

La primera (DT), con resultados de los tres tests psicométricos.

- Capacidades del estudiante (BADyG-renovado), [17].
- Su autoconcepto (AF5), [6].
- Su adaptación a diferentes ámbitos (TAMAI), [9].

La segunda (DA), con resultados académicos de tres cursos académicos.

Y la tercera (DP), con datos personales extraídos del centro.

De estas tres fuentes de información utilizamos los siguientes ítems:

Datos Académicos (DA)		Datos Personales (DP)	
nlcC1	N. media en Len., C1	nem	Nivel estudios madre
nmtC1	N. media en Mat., C1	nh	Número hermanos
nmC1	N. media C1	pm	Profesión madre
nlcC2	N. media en Len., C2	anm	Año nacimiento madre
nmtC2	N. media en Mat., C2	anp	Año nacimiento padre
nmC2	N. media C2	nep	Nivel estudios padre
lc1C3	N. media 1ª ev. Len., C3	pp	Profesión padre
mt1C3	N. media 1ª ev. Mat., C3		
nm1C3	N. media 1ª ev. CA3		
Datos Test (DT)			
RE	Matrices de figuras	AL	Autoconcepto Académico-Laboral
EF	Eficiencia	EM	Autoconcepto Emocional
SN	Problemas Numéricos	H	Grado Inadaptación con hermanos
RL	Razonamiento Lógico	DE	Atención: Disc rimirar diferencias
RN	Series Numéricas	MV	Memoria Visual Ortográfica
CI	Cociente Intelectual	RAP	Rapidez
SE	Orientación Espacial	P	Adaptación Personal
MA	Memoria de Relato Oral	S	Adaptación Social
RV	Analogía Verbales	E	Adaptación Escolar
SV	Completar oraciones	Pa	Adaptación al padre
SOC	Autoconcepto Social	M	Adaptación a la madre
FAM	Autoconcepto Familiar	F	Adaptación Familiar
FIS	Autoconcepto Físico		

Adicionalmente, se utilizan los parámetros PLEN y PMAT como predictores del rendimiento en Lengua y Matemáticas [11], definidos como:.

PLEN = media aritmética de (RV, MV, MA, DE, SV)

PMAT = media aritmética de (RL, RN, SN, MV, MA, DE)

Además, disponemos de la información correspondiente a las acciones que toman los deptos. de Orientación, Lengua y Matemáticas, que denominaremos AcL y AcM, con cuatro posibles valores:

- Realizar apoyo en Lengua o Matemáticas (RAL, RAM)
- Realizar refuerzo en Lengua o Matemáticas (RRL, RRM)
- No tomar decisión alguna (X)
- Cuestionar rendimiento en Lengua o Matemáticas (CL, CM)

Por tanto, construimos instancias de 35 atributos (los ítems de DP, DA y los ítems de DT menos los utilizados en PLEN y PMAT, más los parámetros PLEN y PMAT). Todos los atributos son numéricos normalizados mediante la función  $f(x) = \frac{1}{1 + \exp(-(\frac{x-\mu}{\sigma}))}$ . El atributo 36 corresponde a la clasificación propuesta por los deptos. En total, el conjunto de datos consta de 552 instancias.

### 3 Relación entre los atributos

Inicialmente, queremos identificar los factores que pueden ser considerados más importantes por existir altas correlaciones entre los atributos. Para ello utilizamos el conjunto de datos sin el atributo de las acciones AcL y AcM. Utilizamos la técnica de Análisis de Componentes Principales (ACP) (técnica multivariante que permite identificar un grupo de variables que se denominan factores, [5], los cuales son una combinación lineal de los atributos originales e independientes entre sí) que nos proporciona la herramienta SPSS, [14].

Para comprobar la conveniencia de un ACP calculamos el estadístico de adecuación muestral de Kaiser-Meyer-Olkin (KMO) y aplicamos el test de esfericidad de Bartlett (EB), mostrando estos resultados: KMO=0.783, y EB=(Chi-cuadrado=17788.67, Sig.=0.00, df=595). Se comprueba que la adecuación muestral no solamente es deseable sino que presenta una buena adecuación a los datos, ya que el valor de KMO es aceptable para este tipo de análisis. Mediante la prueba EB podemos asegurar que si el nivel crítico es superior a 0.05 entonces no podremos rechazar la hipótesis nula de esfericidad. Como podemos comprobar en nuestro análisis la significación es perfecta ya que obtiene el valor 0.00. En definitiva, se confirma que existe interdependencia entre los atributos.

El criterio analítico adoptado para determinar el número de factores ha sido el criterio de la media aritmética, según el cual se seleccionan aquellos factores cuyo autovalor excede al de la media de los autovalores (al estar los atributos tipificados, los factores seleccionados serían aquellos que presenten un autovalor mayor que la unidad). En la Tabla 1 figuran todos los factores seleccionados acompañados de sus autovalores, la varianza, así como la varianza acumulada según el número de factores considerados.

**Tabla 1.** Factores seleccionados y rotación ortogonal

Factor	Valores iniciales			Valores tras rotación ortogonal		
	Autovalores	%Var	%Acumulado	Total	%Var	%Acumulado
1	8.920	25.512	25.512	8.312	23.749	25.512
2	2.738	7.822	33.334	2.853	8.152	31.901
3	2.614	7.469	40.803	2.671	7.632	39.532
4	2.043	5.836	46.639	1.999	5.710	45.243
5	2.020	5.770	52.410	1.718	4.910	50.133
6	1.678	4.794	57.203	1.679	4.797	54.949
7	1.396	3.988	61.191	1.467	4.192	59.141
8	1.135	3.243	64.435	1.456	4.159	63.300
9	1.118	3.194	67.628	1.359	3.882	67.182
10	1.024	2.927	70.555	1.181	3.373	70.555
11	0.972	2.776	73.332			
..	....	....	....			

Puede observarse que a partir del factor 10, el autovalor comienza a ser inferior a la unidad y además el porcentaje de la varianza explicada acumulada es casi un 71 %, lo que es un porcentaje más que aceptable. Para interpretar estos factores utilizamos la Rotación Factorial (método Varimax con normalización Kaiser) que transforma la matriz factorial inicial en otra denominada matriz factorial rotada más fácil de interpretar. Los factores rotados tratan de conseguir que los atributos originales tenga una correlación lo más próxima posible a uno de los factores, y correlaciones próximas a cero con los restantes (consiguiendo correlaciones altas con un grupo de atributos y bajas con el resto). La matriz de factores rotados se muestra en la Tabla 2, la cual muestra los valores por encima de 0, marcando en negrita las correlaciones más altas asociadas a cada factor.

**Tabla 2.** Factores rotados. Correlaciones de los atributos a los distintos factores

Atributos	Factores									
	1	2	3	4	5	6	7	8	9	10
nm1C3	<b>0.968</b>									
nmC1	<b>0.968</b>									
nmC2	<b>0.967</b>									
n1c1C3	<b>0.946</b>									
n1cC1	<b>0.944</b>									
n1cC2	<b>0.936</b>									
nmt1C3	<b>0.920</b>	0.143								
nmtC1	<b>0.918</b>	0.154								
nmtC2	<b>0.906</b>	0.127								
PMAT	0.195	<b>0.821</b>							0.265	
CI	0.246	<b>0.737</b>								
PLEN	0.160	<b>0.736</b>							0.334	0.122
EF	0.113	<b>0.638</b>								
RE		<b>0.621</b>						0.121		
SE	0.148	<b>0.334</b>				0.223	0.208	0.187	0.298	
ne-m			<b>0.839</b>					0.200		
nh			<b>0.831</b>						0.142	
H			<b>0.821</b>						0.122	
pm			<b>0.726</b>							
an-p				<b>0.979</b>						
an-m				<b>0.978</b>						
pp					<b>0.910</b>					
ne-p					<b>0.905</b>					
S					0.106	<b>0.679</b>				
E						<b>0.666</b>		0.100		0.137
P						<b>0.531</b>				
F						<b>0.433</b>				0.365
SOC							<b>0.693</b>			0.150
FAM	0.135						<b>0.632</b>			
FIS							<b>0.504</b>	0.210		
AL	0.252	0.169			0.135		<b>0.381</b>		0.121	
M							0.140	<b>0.768</b>		
Pa								<b>0.733</b>	0.135	
RAP		0.139							<b>0.829</b>	
EM							0.115			<b>0.796</b>

Como observamos en la Tabla 2, las asignaciones de atributos a factores presentan una correlación positiva. Además, como hemos observado en la Tabla 1, los 5 primeros factores explican casi un 53 % de la varianza (explicando el factor 1 un 25.512 %). Interpretemos estos factores en función de sus atributos:

- Factor 1. Factor cuantitativo (F1) puesto que los atributos correlacionados miden las notas en lengua, matemáticas y medias de los cursos (9 atributos).
- Factor 2. Factor Lógico-Matemático-Lingüístico (F2), puesto que los atributos correlacionados tienen un fuerte contenido lógico/matemático y lingüístico (6 atributos).
- Factor 3. Factor Materno-Filial (F3), ya que los atributos correlacionados con este factor están relacionados con la madre y los hijos (4 atributos).
- Factor 4. Factor Fechas de nacimiento de los Padres (F4). Como su nombre indica, los atributos correlacionados son anm y anp (2 atributos).
- Factor 5. Factor Paterno (F5), ya que los dos atributos correlacionados son nep y pp (2 atributos).
- Factor 6. Factor Capacidad de Adaptación (F6). Los atributos correlacionados son P, S, F y E (4 atributos).
- Factor 7. Factor Autoconcepto (F7). Los atributos correlacionados son SOC, FAM, FIS y AL (4 atributos).
- Factor 8. Factor Inadaptación a los Padres (F8). Los atributos correlacionados son Pa y M (2 atributos).
- Factor 9. Factor Rapidez (F9). El atributo correlacionado es RAP (1 atributo).
- Factor 10. Factor Autoconcepto Emocional (F10). El atributo correlacionado es EM (1 atributo).

Para complementar el ACP aplicamos un análisis cluster con el método “average linkage entre grupos” (utilizamos la herramienta WEKA [8]) para agrupar los atributos. El mejor cluster es de tamaño 5 y lo interpretamos como sigue:

- C1 recoge los factores F1, F2 y F9 más el atributo AL del factor F7 (agrupa los ítems vinculados con el ambiente escolar y las capacidades asociadas).
- C2 coincide con el factor F10 (agrupa los ítems vinculados con el autoconcepto emocional).
- C3 recoge los factores F5 y F8 más los atributos FIS, SOC y FAM de F7 (agrupa los ítems vinculados con el factor paterno, inadaptación a los padres y el autoconcepto físico, social y familiar).
- C4 recoge los factores F3 y F6 (agrupa los ítems vinculados con el factor maternal y la capacidad de adaptación).
- C5 coincide con el factor F4 (agrupa los ítems vinculados con las fechas de nacimiento de los padres (F4)).

Cada cluster ha agrupado de manera coherente los distintos factores por su similitud y homogeneidad, a diferencia del factor F7 que es dividido en dos. Pero esto tiene sentido, dada la fuerte correlación del autoconcepto académico-laboral con el entorno en el que se desarrolla. El análisis cluster extrae de F7 el atributo AL y lo asigna a C1 ya que contiene los atributos más similares con él.

#### 4 Caracterizar el proceso de aplicación de las acciones

Ahora aplicamos el análisis cluster con el método “average linkage entre grupos” sobre el conjunto de datos para agrupar las instancias (estudiantes) que tengan

un comportamiento similar con respecto al atributo de las acciones aplicadas (tanto para Lengua - AcL-, como para Matemáticas - AcM).

#### 4.1 Comportamiento cuando aplicamos acciones AcL

Realizando un análisis cluster sobre el conjunto de datos de 552 instancias obtenemos un comportamiento de los distintos cluster con respecto a la suma de las distancias intragrupalas. El cluster de tamaño 5 es donde se empieza a estabilizar la medida. Sobre este cluster de tamaño 5, obtenemos el comportamiento del atributo AcL (Tabla 3).

**Tabla 3.** Distribución de las instancias según la acción asignada para AcL

AcL	C1	C2	C3	C4	C5
	17.9%	5.1%	23.6%	41.1%	12.3%
CL	4.3%	1.1%	2.7%	1.1%	2.0%
RAL	0.2%	-	-	40.0%	0.5%
X	13.4%	-	20.9%	-	9.8%
RRL	-	4.0%	-	-	-

#### 4.2 Comportamiento cuando aplicamos acciones AcM

Aplicando ahora el análisis cluster sobre el conjunto de datos de 552 instancias cuando aplicamos las acciones AcM obtenemos un cluster de tamaño 7. En la Tabla 4 mostramos la distribución de las instancias sobre cada grupo en función de los valores del atributo AcM.

**Tabla 4.** Distribución de las instancias según la acción asignada para AcM

AcM	C1	C2	C3	C4	C5	C6	C7
	13.6%	4.9%	21.0%	18.3%	10.9%	10.1%	21.2%
CM	-	-	0.4%	0.2%	-	10.1%	-
RAM	-	-	-	18.1%	-	-	21.2%
X	13.6%	-	20.5%	-	10.9%	-	-
RRM	-	4.9%	0.2%	-	-	-	-

## 5 Análisis de resultados y sistema de predicción

Mediante el ACP y análisis cluster sobre los atributos hemos destacado la importancia de los atributos medidos a los estudiantes. Destacamos la importancia de los factores Factor cuantitativo (F1), Factor Lógico-Matemático-Lingüístico (F2) y Factor Materno-Filial (F3) que recogen casi el 41 % de la varianza), seguidos del Factor Fechas de nacimiento de los Padres (F4), Factor Paterno (F5) y Factor Capacidad de Adaptación (F6) que recogen casi el 17 % de la varianza.

Al analizar los clusters respecto de sus atributos concluimos que un grupo lo forman los factores relacionados con el ambiente escolar (resultados académicos, rapidez, atención o esfuerzo) y sus capacidades asociadas a lengua y matemáticas y el autoconcepto académico-laboral (C1).

También destacar el grupo C4 que muestra la relación entre el factor materno-filial F3 (nivel de estudios y profesión de la madre, número de hermanos y su adaptación a ellos), con las capacidades de adaptación del estudiante en sus facetas física, escolar, social y personal (F6).

El grupo C3 que relaciona el autoconcepto del estudiante en sus facetas física, social y familiar (F7), con el factor paterno (nivel de estudios y profesión del padre, F5) y la inadaptación con los padres (F8). La presencia física en la adolescencia tiene una repercusión social importante que debe ser potenciada desde el ámbito familiar. Vemos agrupada la influencia del padre, el grado de inadaptación al padre y a la madre, y el autoconcepto del estudiante.

Toda esta información puede ser relevante para los depts. (sobre todo al de Orientación) para que puedan tomar las decisiones adecuadas debido a que valores bajos de algunos de los factores en los estudiantes afectarán a los otros factores relacionados.

Además, a partir de los dos clusters obtenidos de tamaño 5 y 7 para Lengua y Matemáticas respectivamente construimos un sistema de toma de decisiones sobre las acciones concretas que afectan a las dos asignaturas. Utilizamos el método PART (que proporciona el herramienta Weka [8]) para obtener reglas para AcL y para AcM, que nos darán información de las acciones y de las capacidades valoradas en cada grupo de estudiantes.

Sobre el cluster de tamaño 5 para AcL y el de tamaño 7 para AcM, obtenemos las siguientes reglas PART con una precisión media de (utilizando un 5-fold cross validation) 97.17% (Tabla 5) y 99.49% (Tabla 6), respectivamente.

**Tabla 5.** Reglas PART para la acción AcL

C1	IF PLEN>0.394494 THEN AcL= X
	ELSE IF nlcC1>0.649611 THEN AcL=CL ELSE AcL=RAL
C2	IF nlcC1>0.610153 THEN AcL= RRL
	ELSE AcL= CL
C3	IF PLEN≤ 0.79208 THEN AcL=X
	ELSE AcL=CL
C4	IF PLEN≤0.752841 THEN AcL=RAL
	ELSE IF nlcC2≤0.13868 THEN AcL=CL
	ELSE AcL=RAL
C5	IF 0.369487<PLEN≤0.77409 THEN AcL=X
	ELSE IF EM>0.464619 THEN AcL=CL
	ELSE IF nlcC1≤0.687123 THEN AcL=RAL
	ELSE AcL=CL

**Tabla 6.** Reglas PART para la acción AcM

C1	AcM=X
C2	AcM=RRM
C3	IF PMAT $\leq$ 0.818546 THEN AcM=X ELSE IF F $>$ 0.377913 THEN AcM=CM ELSE AcM=RRM
C4	IF S $\leq$ 0.882653 THEN AcM=RAM ELSE IF pp $>$ 0.278186 THEN AcM=RAM ELSE AcM=CM
C5	AcM=X
C6	AcM=CM
C7	AcM=RAM

A partir de los clusters y de las reglas obtenidas sobre ellos, el sistema para la toma de decisiones es el siguiente:

#### Protocolo de asignación de una acción a un estudiante

1. Construir el vector de 35 atributos (“InfNA”) con los valores correspondientes a los antecedentes académicos del estudiante NA, la información disponible en el centro y los resultados de los distintos tests.
  - a. En el depto. de Lengua  
Calcular las distancias de InfNA a los centroides del cluster para AcL. Elegir el cluster que obtiene la mínima distancia (CAcLNA). Evaluar las reglas correspondientes al CAcLNA de la Tabla 5. La acción de salida se proporciona al depto. de Lengua como una acción que debería ser aplicada al estudiante NA en Lengua.
  - b. En el depto. de Matemáticas  
Calcular las distancias de InfNA a los centroides del cluster para AcM. Elegir el cluster que obtiene la mínima distancia (CAcMNA). Evaluar las reglas correspondientes al CAcMNA de la Tabla 6. La acción de salida se proporciona al depto. de Matemáticas como una acción que debería ser aplicada a al estudiante NA en Matemáticas.

## 6 Conclusiones

En el entorno académico en el que se desenvuelve este protocolo de actuación, observamos que el ACP y el análisis cluster realizado sobre los atributos nos muestra la relación que guardan entre sí los atributos agrupados en cada cluster (C1-C5) y por factores (F1-F10). Esta información permitirá inicialmente al departamento de Orientación de un IES analizar el entorno académico de cada estudiante y ver qué factores afectan a su rendimiento.

Una vez reconocido ese entorno, cada departamento didáctico (ya sea Lengua, ya sea Matemáticas), aplicará unas reglas a cada estudiante con arreglo al cluster donde se ha clasificado, pero inducido por la acción que queremos aplicar (AcL ó AcM). Ahora el comportamiento es distinto según la acción.

## Acknowledgement

Trabajo soportado por los proyectos TIN2011-27696-C02-02 y TIN2014-52099-R del Ministerio de Economía y Competitividad de España.

## Referencias

1. Andrade, M., Miranda, C., Freixas, G.: Predicción del rendimiento académico lingüístico y lógico-matemático por medio de las variables modificables de las inteligencias múltiples y del hogar. *Revista Digital de Investigación y Nuevas Tecnologías*, vol. 17, pp. 1–18 (2001)
2. Bacallao, C. J., Parapar, J. M., Roque, M., Bacallao, J.: Árboles de regresión y otras opciones metodológicas aplicadas a la predicción del rendimiento académico. *Revista Cubana Educación Médica Superior*, vol. 18(3) (2004).
3. Baradwaj, B. K., Pal, S.: Mining Educational Data to Analyze Students' Performance. *Int Jour of Advanced Computer Science and Applications*, vol. 2(6), pp. 63–69 (2011).
4. Durán, E., Costaguta, R.: Minería de datos para descubrir estilos de aprendizaje. *Revista Iberoamericana de Educación*, vol. 42(2), pp. 1–10 (2007).
5. Francis, P. J., Wills, B. J.: Introduction to Principal Components Analysis, ASP Conference Series, vol. 162, pp. 363–372 (1999).
6. García, F., Musitu, G.: AF5 - Autoconcepto Forma 5. Ed. TEA, Madrid (1999)
7. Gardner, H.: Estructura de la mente. La teoría de las inteligencias múltiples (2a Edición). Fondo de Cultura Económica, México (1994)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reumtemann, P., Witten, I. H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, vol. 11(1), pp. 10–18 (2009).
9. Hernández, P.: TAMAI. Test Autoevaluativo Multifactorial de Adaptación Infantil. Ed. TEA, Madrid (1983)
10. Jornet-Melía, J. M., Gonzalez-Such, J.: Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, vol. 16, pp. 103–123 (2009)
11. Muñoz, A., Cadenas, J.M.: Un modelo para estimar las acciones a aplicar a alumnos de ESO. *Revista de Investigación Educativa*, vol. 26(2), pp. 385–407 (2008).
12. Muñoz, A., Cadenas, J. M.: Validación de los predictores del rendimiento académico obtenidos mediante minería de datos . *Proc. XXXI Cong Nac de Estadística e Investigación Operativa* (pp. 1–8). Murcia, España (2009)
13. Muñoz, A., Cadenas, J. M.: Using clustering to identify actions-attributes. *Proc. I Cong Int de Formación del Profesorado* (pp. 141–150). Murcia, Spain (2011)
14. Pardo, A., Ruiz, M. A.: SPSS 11. Guía para el análisis de datos. McGraw-Hill, Madrid (2002)
15. Santos-Rego, M. A., Godás-Otero, A., Lorenzo-Moledo, M.: El perfil del alumnado repetidor y no repetidor en una muestra de estudiantes españoles y latinoamericanos: un estudio sobre los determinantes de sus logros académicos. *Estudios sobre Educación*, vol. 23, pp. 141–155 (2012)
16. Shovon, H. I., Haque, M.: An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree. *Int Jour of Advanced Computer Science and Applications*, vol. 3(8), pp. 146–149 (2012)
17. Yuste, C., Martínez, R., Galve J. L.: Manual Técnico de la batería de aptitudes diferenciales y generales (BADyG). Ed. CEPE, Madrid (2002)