

Descubrimiento basado en *clustering* de la influencia de rasgos psicológicos sobre el desempeño académico

Graciela L. Meza Lovón

*Graciela Meza obtuvo su MSc en la Universidade Federal de Mato Grosso do Sul (Brasil),
donde realizó su investigación en sistemas neurodifusos.
Posteriormente, trabajó en la Universidad Católica San Pablo de Arequipa.
Actualmente es candidata a doctora por la Concordia University de Montreal (Canadá).*

Cristian J. López del Álamo

*Cristian López del Álamo es magíster en Ingeniería de Software con experiencia
en docencia y desarrollo de software para toma de decisiones gerenciales.
Sus áreas de investigación son data mining, estructura de datos,
análisis y diseño de algoritmos e inteligencia artificial.
Desde el 2003 se desempeña como docente en la Universidad Católica San Pablo de Arequipa.
Ha trabajado en el sistema híbrido de apoyo en el diagnóstico en el tratamiento de pacientes
con trasplante renal, y actualmente trabaja en en data mining aplicada en métricas de software.*



Descubrimiento basado en *clustering* de la influencia de rasgos psicológicos sobre el desempeño académico

RESUMEN

Esta investigación aplica técnicas de *clustering* para descubrir patrones en datos psicológicos y académicos de los alumnos de Ingeniería Informática de la Universidad Católica San Pablo de Arequipa. Con esto se busca encontrar el conjunto de habilidades intelectuales y características psicológicas de un estudiante exitoso; de este modo es posible desarrollar estas habilidades y características en nuevos alumnos. Para la evaluación de los resultados se utilizó el índice de rand ajustado y bootstrapping. El artículo concluye mostrando que los factores: energía, significado de palabras y tesón guardan relación con la calificación de los alumnos.

PALABRA CLAVE

clustering, k-means, update-online, bootstrapping, rendimiento académico

INTRODUCCIÓN

En las últimas décadas, el avance de la tecnología ha crecido vertiginosamente posibilitando el almacenamiento de gran cantidad de información, la que puede ser analizada para encontrar patrones, y finalmente nuevo conocimiento. Se utilizarn muchas técnicas

con este propósito. Una de ellas es conocida como *clustering* y forma, a partir de un conjunto de datos, grupos cuyos elementos tienen similitud entre sí. Luego esta similitud es analizada para entender cuáles son los factores que hacen que estos elementos pertenezcan a un mismo grupo.

El objetivo de esta investigación es descubrir los rasgos de personalidad y madurez mental más influyentes en el rendimiento académico, utilizando técnicas de agrupamiento sobre los datos de los alumnos de Ingeniería Informática, obtenidos por medio de los tests de Madurez Mental de California, BF5 de Personalidad y de los registros de notas de la Universidad Católica San Pablo de Arequipa (UCSP). Con esto se busca orientar tanto a docentes como alumnos para así maximizar las posibilidades de aprendizaje, teniendo en cuenta las habilidades requeridas que llevan al éxito académico. El proyecto fue dividido en las siguientes etapas: preprocesamiento, *clustering*, evaluación, y análisis de resultados.

En las siguientes secciones se abordan las técnicas de *clustering* utilizadas en este proyecto y los conceptos de «rand index ajustado y bootstrapping» sobre los que se basa la metodología de evaluación. Luego se pasa revista a la aplicación de *clustering* en los datos de los alumnos de la UCSP. Finalmente, en las últimas secciones se muestran los resultados y las conclusiones.

ALGORITMOS DE CLUSTERING

K-Means

El algoritmo *K-Means* [9][6] particiona las entradas en k conjuntos iniciales. Luego, para cada conjunto se calcula su media o centroide. Se construye una nueva partición asociando cada punto al centroide más cercano y los centroides son recalculados. El algoritmo itera mientras no se cumpla una condición de convergencia, es decir, cuando alcanza un estado en el que los centroides no cambian. Para asignar el punto más próximo a un centroide, generalmente, se usa la distancia euclídea. Para recalcular el centroide de cada *cluster* es necesario, además de una medida de proximidad, una función que minimice la distancia cuadrática de cada punto a su centroide, usándose el error cuadrático

$$EC = \sum \sum d(c_i, x)^2$$

donde, x es un objeto del conjunto de datos; C_i es i ésimo *cluster*, el centroide del *cluster*; C_i ; k , el número de *cluster*, y d , la función distancia. El centroide del i ésimo *cluster* es $c_i = \frac{1}{m_i} \sum x$ donde m_i es el número de objetos en el i ésimo *cluster*.

Neural Gas

La *Neural Gas* [4] es una red neuronal con un método de aprendizaje competitivo suave. Este método permite realizar la agrupación de un conjunto de puntos de datos en un espacio enedimensional con el conjunto más pequeño de vectores referencia o vectores centro conocidos también como prototipos C^k , donde $k = 1, 2, \dots, N$. Cada vector de datos es representado por el vector referencia más cercano. Estos vectores referencia son determinados de tal manera que la distancia euclidea esperada entre todos los vectores de datos y su correspondiente vector referencia sea mínima.

On-line Update

El algoritmo *On-line Update* [7] necesita que sean especificados el número de *clusters* que se desean formar, creándose para cada uno, una unidad a con un vector de referencia w de dimensión igual al número i de características de las entradas (objetos), $w_a = w_{a1}, w_{a2}, \dots, w_{ai}$. Para determinar la unidad ganadora, s , se utiliza un criterio de similaridad, $s = \operatorname{argmin}_x |x - w_a|$. *On-line Update*, usa un valor e , $0 \leq e \leq 1$ que es conocido como tasa de aprendizaje y que determina el grado en que la unidad ganadora se adapta al objeto de entrada. Valores cercanos a 1 provocan que el vector de referencia se adapte más rápidamente a x . Los *clusters* se van formando adaptando los vectores de referencia a los objetos más similares, mediante $Vw_s = e(x - w_s)$.

EVALUACIÓN: ÍNDICE RAND AJUSTADO Y BOOTSTRAPPING

Para evaluar una configuración de *cluster* sobre la base de un criterio externo es necesaria una medida de concordancia. Para el criterio externo se usó *bootstrapping*, y para la medida de concordancia se usó el Índice de Rand Ajustado (IRA) [5][10].

Sea S un conjunto de objetos, $U = u_1, u_2, \dots, u_R$ y $V = v_1, v_2, \dots, v_C$ dos particiones de objetos en S , tal que $U_{i=1}^R u_i = U_{j=1}^C v_j = S$ y $u_i \cap u_{i'} = v_j \cap v_{j'}$ para $1 \leq i \neq i' \leq R$ y $1 \leq j \neq j' \leq C$. Sea n_{ij} el número de objetos que están tanto en u_i como v_j , $n_i = \sum_{j=1}^C n_{ij}$ y $n_j = \sum_{i=1}^R n_{ij}$. El IRA se define como:

$$\frac{\sum C_2^{ij} - \left[\sum C_2^{n_i} \sum C_2^{n_j} \right] / C_2^n}{\frac{1}{2} \left[\sum C_2^{n_i} + \sum C_2^{n_j} \right] - \left[\sum C_2^{n_i} \sum C_2^{n_j} \right] / C_2^n}$$

Tanto V como U (actúa como criterio externo) representan dos configuraciones con el mismo número de *clusters*, obtenidas aplicando una técnica de *cluster* particular. El IRA

está definido en el rango de $[0,1]$, donde los valores más próximos a 1 indican una mayor concordancia, es decir, una mejor configuración de *cluster*.

Supóngase que $S = X'_1, X'_2, \dots, X'_n$ es una muestra aleatoria simple o una muestra aleatoria independiente de $P = x_1, x_2, \dots, x_N$, donde n es considerablemente menor de N . Supóngase que se está interesado en un estadístico $T = t(S)$ como un estimado de un parámetro $\theta = t(P)$ de la población P . *Bootstrapping* [2] consiste en construir R muestras, $S_1, S_2, \dots, S_b, \dots, S_R, S_b = X_1, X_2, \dots, X_N$ de tamaño N de P aplicando un muestreo con reemplazo. Cada muestra es tratada como un estimado de la población P , donde cada elemento de S_b es seleccionado con una probabilidad $\frac{1}{N}$. Luego, se calcula el estadístico T de cada muestra, i.e., $T_b = t(S_b)$. La distribución de T_b sobre el estimado original $T = t(S)$ es análogo a la distribución en el muestreo de el estimador T sobre un parámetro de la población $\theta = t(P)$ [3].

En el contexto de aprendizaje máquina, *bootstrapping* es una técnica empleada para evaluar la estabilidad de los resultados de un modelo de clasificación o de *clustering*. La idea subyacente tras el *bootstrapping* es la misma que la explicada anteriormente, sólo que en este caso el estadístico se refiere a alguna métrica para evaluar los resultados del modelo en cuestión; en este caso, el IRA explicado líneas arriba.

CLUSTERING APLICADO A DATOS DE ESTUDIANTES

Obtención de los datos

Los datos como notas parciales, notas de evaluación permanente, entre otros, fueron proporcionados por el área de Informática de la UCSP. Los datos concernientes a cualidades intelectuales y de personalidad se obtuvieron aplicando el Test Personalidad BF5 y el Test de de Madurez Mental de California.

El Test BF5 de personalidad (TBF5) [1] es un modelo donde la personalidad es explicada por 5 dimensiones (cada una dividida en dos subdimensiones): 1) Energía (E), sus subdimensiones son Dinamismo (Di) y Dominancia (Do); 2) Afabilidad (A), Cooperación (Cp) y Cordialidad (Co); 3) Tesón (T), Escrupulosidad (Es) y Perseverancia (Pe); 4) Estabilidad Emocional (Ee), Control de Emociones (Ce) y Control de Impulsos (Ci); y finalmente, 5) Apertura Mental (Am), Apertura a la Cultura (Ac) y Apertura a la Experiencia (Ae).

El Test de Madurez Mental de California (TMMC) [8] reúne los componentes de la inteligencia en dos grupos: Verbales (V) y No Verbales (NV). El test se respalda en la teo-

ría multifactorial en la que la inteligencia está constituida por la concurrencia de factores mentales como relaciones espaciales, razonamiento lógico y numérico, conceptos verbales y memoria inmediata y duradera. El test consta de 7 sub tests en los que se evalúa los siguientes factores mentales: 1) Relaciones espaciales. Sub-test n.º 1 y 2: Sentido derecha-izquierda (NV) y Manipulación de áreas (NV); 2) Razonamiento lógico. Sub-test n.º 3 y 4: Analogías (NV) e Inferencias (V); 3) Razonamiento numérico. Sub-test n.º 5 y 6: Series Numéricas (NV) y Problemas de Aritmética (V); y, finalmente, 4) Conceptos verbales. Sub-test n.º 7: Vocabulario (V).

Preprocesamiento de los datos

Se seleccionaron los datos de los alumnos que tuvieron resultados válidos en ambos tests. Se seleccionaron los datos de los cursos de Álgebra Lineal, representando al curso de Matemática, el curso de Matemática Discreta representado a los cursos de Matemática para la Computación, el curso de Programación I, representando a los cursos de Programación y el curso de Metodología del Estudio representando a los cursos de letras.

Creación del vector de características

El vector de características está formado por 27 atributos detallados en el cuadro n.º 1. Es importante mencionar que no se realizó ningún proceso de reducción de dimensionalidad del vector de características; pero tampoco se usaron todos los atributos en todos los experimentos. En la sección siguiente se describe cada experimento y los atributos que fueron considerados en cada uno.

Cabe resaltar que los datos han sido normalizados antes de ser usados en los algoritmos.

Cuadro n.º 1: Descripción de atributos

N.º	Atributo	Descripción	Valores
1	Sexo	Sexo del alumno	F, M
2	De_Iz	Factor medido por el test Derecha-Izquierda del TMMC.	[0,20]
3	Areas	Factor medido por el test Manipulación de áreas del TMMC.	[0,15]
4	Anlg	Factor medido por el test Analogías del TMMC.	[0,15]
5	Infe	Factor medido por el test Inferencias del TMMC.	[0,15]
6	Se_Nu	Factor medido por el test Series numéricas del TMMC.	[0,15]
7	Prob	Factor medido por el test Problemas de aritmética del TMMC.	[0,15]
8	Si_Pa	Factor medido por el test Vocabulario del TMMC.	[0,50]
9	NV	Suma de las capacidades no verbales del TMMC.	[0,65]
10	V	Suma de las capacidades verbales del TMMC.	[0,80]

11	Di	Subdimen. Dinamismo del TBF5.	[12,60]
12	Do	Subdimen. Domimancia del TBF5.	[12,60]
13	Cp	Subdimen. Cooperación del TBF5.	[12,60]
14	Co	Subdimen. Coordialidad del TBF5.	[12,60]
15	Es	Subdimen. Escrupulosidad del TBF5.	[12,60]
16	Pe	Subdimen. Perseverancia del TBF5.	[12,60]
17	Ce	Subdimen. Control de Emociones del TBF5.	[12,60]
18	Ci	Subdimen. Control de Impulsos del TBF5.	[12,60]
19	Ac	Subdimen. Apertura a la Cultura del TBF5.	[12,60]
20	Ae	Subdimen. Apertura a la Experiencia del TBF5.	[12,60]
21	E	Dimensión Energía del TBF5.	[24,120]
22	T	Dimensión Tesón del TBF5.	[24,120]
23	A	Dimensión Afabilidad del TBF5.	[24,120]
24	Ee	Dimensión Estabilidad Emocional del TBF5.	[24,120]
25	Am	Dimensión Apertura Mental del TBF5.	[24,120]
26	NOTA	Valor numérico de nota.	[0,20]

Metodología para la evaluación de *clusters*

Consiste en ejecutar iterativamente un algoritmo de *cluster* dado, sobre un conjunto de datos en particular para diferentes números de *clusters* usando *bootstrapping*. Los pasos son:

- Se establece el número de muestras k , para el *bootstrapping*.
- Se crean $2k$ muestras del conjunto de datos de acuerdo al *bootstrapping*.
- Se aplica un algoritmo de *clustering* a cada una de las $2k$ muestras, obteniéndose $2k$ configuraciones, todas con el mismo número de *cluster*.
- Se aplica IRA a cada dos configuraciones, esto es, a la configuración i y a la configuración $i+1$, $i=1,3,5,\dots,2k-1$, obteniéndose k valores para IRA.
- Se calcula el valor IRA final para una configuración con un número de *cluster* dado. El IRA final es la media de los IRA obtenidos anteriormente.
- Se realiza los pasos anteriores obteniéndose IRA finales para cada número diferente de *cluster*.
- Se obtiene el mayor valor de los IRA finales.

Nótese que el valor IRA final no es de una configuración particular, sino de una configuración de la que sólo se conoce el número de *cluster*, es decir, no se detalla qué objeto pertenece a qué *cluster*. La idea de esta metodología es determinar para una técnica en particular, con qué número de *cluster* se obtiene el mejor desempeño (mayor IRA final).

El proceso anterior es aplicado utilizando las tres técnicas de *clustering* en tres experimentos distintos. De este modo, para cada técnica de *clustering* aplicada a tres conjuntos de datos, obtenemos el valor medio del IRA, el cual nos indica el mejor número de *cluster* para esa técnica y para ese conjunto de datos. Finalmente, una vez determinado el número adecuado de *cluster*, se aplica nuevamente los tres algoritmos a los tres conjuntos de datos, y posteriormente se pasa a la fase de análisis de resultados.

PRUEBAS Y RESULTADOS

Descripción de los experimentos

Para determinar los atributos que influyeron en las notas de los alumnos, se realizaron tres experimentos:

1. Experimento 1: Se tomó en cuenta los atributos de inteligencia (atributos del 3 al 8 del cuadro 1) y Nota.
2. Experimento 2: Se tomó en cuenta los atributos de subdimensiones de personalidad (atributos del 11 al 20 del cuadro 1) y Nota.
3. Experimento 3: Se tomó en cuenta los atributos de dimensiones de personalidad (atributos del 21 al 25 del cuadro 1) y Nota.

Los resultados del *bootstrapping* para los cursos de Metodología del Estudio, Programación, Álgebra Lineal y Matemática Discreta, estimaron que el número óptimo de *cluster* es 3.

Resultados del experimento 1

En el cuadro n.º 2 se muestra las medias de los factores de inteligencia para cada *cluster*. El cuadro está dividido por cursos (Programación y Metodología del Estudio) y técnica de *clustering* usada. Las siete primeras columnas hacen referencia a los factores de inteligencia, y la última, al atributo Nota. Se dará una breve explicación de la manera como está organizado el cuadro. Para el curso de Programación usando *K-means*, la fila Clus. 2 corresponde a las medias de los factores de inteligencia de los elementos del *cluster* 2. En los siguientes experimentos se presentará cuadros que deben ser leídos y analizados de la misma manera.

En el cuadro n.º 2, se puede observar que utilizando tanto *K-means* como *Neural Gas*, el manejo de áreas (Áreas) y el sentido derecha izquierda (De_Iz) influyen en la Nota. La influencia de estos factores es un resultado esperado pues generalmente, profesores y alumnos asocian este curso con la matemática y las aptitudes no verbales. Sin embargo,

la tarea de programar requiere en principio entender un problema y luego formular una resolución usando un lenguaje de programación. Para entender un problema específico, más que capacidades matemáticas, son necesarias capacidades verbales; por eso, los atributos Inferencia y Significado de Palabras tienen una influencia importante.

Cuadro n.º 2: Medias de los factores de inteligencia para los mejores resultados obtenidos del experimento 1 (Cursos: Programación y Metodología)

Programación - Inteligencia (K-means - 3 Clusters)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Nota
Clus 1.	0,81	0,65	0,59	0,80	0,88	0,75	0,80	0,78
Clus 2.	0,72	0,56	0,38	0,52	0,68	0,21	0,62	0,54
Clus 3.	0,48	0,26	0,42	0,62	0,66	0,34	0,59	0,44
Programación - Inteligencia (Neural Gas - 3 Clusters)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Nota
Clus 1.	0,53	0,32	0,40	0,55	0,64	0,25	0,57	0,41
Clus 2.	0,81	0,65	0,59	0,87	0,90	0,81	0,82	0,78
Clus 3.	0,77	0,60	0,46	0,61	0,76	0,40	0,69	0,69
Metodol. Estudio - Inteligencia (K-means - 3 Clusters)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Nota
Clus 1.	0,79	0,66	0,59	0,83	0,84	0,75	0,81	0,80
Clus 2.	0,71	0,48	0,43	0,37	0,66	0,20	0,45	0,61
Clus 3.	0,52	0,3	0,38	0,73	0,67	0,34	0,70	0,72
Metodol. Estudio - Inteligencia (On-line Update - 3 Clusters)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Nota
Clus 1.	0,43	0,17	0,33	0,75	0,80	0,32	0,68	0,71
Clus 2.	0,79	0,65	0,57	0,83	0,82	0,73	0,81	0,80
Clus 3.	0,69	0,48	0,44	0,45	0,60	0,23	0,52	0,64

Por otra parte, para el curso de Metodología del Estudio se observa que las capacidades no verbales dejan de ser influyentes, pues las medias de De_Iz y Áreas no guardan relación con el atributo Nota. El atributo Si_Pa continua siendo importante para la Nota, así como también la capacidad de Inferencia (Infe).

En el cuadro n.º 3 en el curso de Álgebra y Geometría con los resultados de *Neural Gas* se observa que Se_Nu, Prob, Si_Pa y Areas influyen en la Nota.

El Significado de Palabras es un atributo que pertenece a capacidades verbales. Se podría pensar: ¿cómo es posible que las habilidades verbales estén presentes de manera tan marcada en cursos donde aparentemente se requeriría mayores cualidades no verbales?. Esto se explica debido a que la capacidad de entender los conceptos es más importante que el aplicarlos de manera mecánica; en otras palabras, personas que tienen mayores habilidades para comprender e interpretar los conceptos tienen mayores posibilidades de aplicar de manera correcta estos conceptos para la resolución de nuevos problemas. Estas habilidades sumadas a capacidades no verbales como la manipulación de áreas permiten a los alumnos tener una mayor posibilidad de éxito en este tipo de cursos.

Cuadro n.º 3: Medias de los factores de inteligencia para los mejores resultados obtenidos del experimento 1 (Cursos: Álgebra Lineal y Matemática Discreta)

Álgebra y Geometría - Inteligencia (Neural Gas - 3 Cluster)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Notas
Clus 1.	0,79	0,67	0,60	0,78	0,90	0,58	0,83	0,79
Clus 2.	0,52	0,31	0,40	0,51	0,60	0,14	0,57	0,42
Clus 3.	0,70	0,56	0,42	0,73	0,75	0,31	0,68	0,54
Matemática Discreta - Inteligencia (Neural Gas - 3 Cluster)								
	De_Iz	Areas	Anlg	Infe	Se_Nu	Prob	Si_Pa	Nota
Clus 1.	0,80	0,68	0,54	0,82	0,89	0,79	0,80	0,69
Clus 2.	0,64	0,36	0,27	0,75	0,75	0,42	0,68	0,40
Clus 3.	0,55	0,48	0,44	0,47	0,46	0,2	0,79	0,45

Por otro lado, en el curso de Matemática Discreta usando *Neural Gas*, se puede ver una mayor influencia de las habilidades de Significado de Palabras, Capacidad para realizar Analogías y habilidades de orientación espacial representadas por el atributo *De_Iz*. En ambos casos, (Álgebra y Geometría y Matemática Discreta) se observa una presencia muy marcada de una habilidad no verbal (*Si_Pa*), su interpretación es similar a la ya explicada.

Resultados del experimento n.º 2

En el cuadro n.º 4, se muestran las medias de los elementos que conforman cada *cluster* para cada atributo. Independientemente del curso, se ve que Energía (E) y Tesón (T) influyen en la Nota. Las puntuaciones altas en Energía describen a personas activas, enérgicas, dominantes, proactivas, con facilidad de palabra; es decir, personas extrovertidas. El Tesón está relacionado con la perseverancia y meticulosidad, lo que significa que las personas más perseverantes y ordenadas tienden a tener mejores calificaciones.

Este cuadro puede ser corroborado con el cuadro n.º 5 del experimento 3, donde se ve una mayor influencia de Di (Dinamismo), Do (Dominancia), Es (Escrupulosidad), Pe (Perseverancia). Recuérdese que Energía y Tesón son las dimensiones que agrupan a las subdimensiones de Dominancia-Dinamismo y Escrupulosidad-Perseverancia, respectivamente. También se muestran los resultados de las dimensiones de personalidad y Nota en los cursos de Álgebra y Geometría y Matemática Discreta. Se puede ver que en ambos casos existe una mayor influencia de Energía y Tesón, de la misma manera que en el caso anterior, lo cual implica que para estos cursos, alumnos muy dinámicos, activos, enérgicos, dominantes, reflexivos y perseverantes tiene mayores posibilidades de éxito frente a aquellos que carecen de estas habilidades.

Resultados del experimento 3

En el cuadro n.º 5, se muestran las medias de los elementos que conforman cada *cluster* para cada atributo. Se ve una mayor influencia de Di (Dinamismo), Do (Dominancia), Es (Escrupulosidad), Pe (Perseverancia). Este resultado corrobora lo obtenido en el cuadro n.º 4, donde Energía y Tesón tenían una mayor influencia en la Nota, y como se aprecia, las subdimensiones de Energía (Di y Do) y las subdimensiones de Tesón (Es y Pe) también presentan una influencia marcada. Se observa la influencia entre las subdimensiones de personalidad y la Nota en los cursos de Álgebra y Geometría y Matemática Discreta; aquí se aprecia que Dinamismo, Dominancia, Escrupulosidad y Perseverancia tienen influencia sobre la Nota. Es importante notar que a pesar que los *clusters* formados en el experimento 2 para estos mismos cursos y los *clusters* del cuadro n.º 5 son diferentes, se sigue manteniendo la influencia de las subdimensiones de Energía (Di y Do) y Tesón (Es y Pe) sobre la Nota, tal como sucedía en el experimento 2 con Energía y Tesón.

Cuadro n.º 4: Medias de las dimensiones de personalidad para los mejores resultados obtenidos del experimento 1 (Cursos: Álgebra Lineal y Matemática Discreta)

Program. - Person.(5) (On-line Update - 3 Clus)						
	E	T	A	Ee	Am	Nota
Clus 1.	0,42	0,68	0,83	0,44	0,63	0,65
Clus 2.	0,65	0,75	0,66	0,77	0,71	0,66
Clus 3.	0,30	0,37	0,51	0,62	0,41	0,58
Metodol. Estudio - Person.(5) (Neural Gas - 3 Clus)						
	E	T	A	Ee	Am	Nota
Clus 1.	0,28	0,24	0,40	0,56	0,31	0,69
Clus 2.	0,58	0,74	0,58	0,56	0,65	0,78
Clus 3.	0,35	0,44	0,73	0,76	0,52	0,70

Álg. y Geom. - Person.(5) (K-means - 3 Clus)						
	E	T	A	Ee	Am	Notas
Clus 1.	0,46	0,60	0,81	0,66	0,60	0,60
Clus 2.	0,57	0,73	0,36	0,58	0,60	0,74
Clus 3.	0,24	0,30	0,50	0,61	0,39	0,50
Mat. Disc. - Person.(5) (K-means - 3 Clus)						
	E	T	A	Ee	Am	Nota
Clus 1.	0,28	0,34	0,48	0,60	0,40	0,43
Clus 2.	0,45	0,60	0,78	0,64	0,58	0,60
Clus 3.	0,63	0,76	0,29	0,49	0,69	0,69
Mat. Disc. - Person.(5) (Neural Gas - 3 Clus)						
	E	T	A	Ee	Am	Nota
Clus 1.	0,55	0,75	0,64	0,51	0,68	0,63
Clus 2.	0,34	0,45	0,75	0,72	0,51	0,53
Clus 3.	0,32	0,34	0,36	0,55	0,37	0,45
Mat. Disc. - Person.(5) (On-line Update - 3 Clus)						
	E	T	A	Ee	Am	Nota
Clus 1.	0,35	0,48	0,65	0,62	0,55	0,52
Clus 2.	0,25	0,21	0,39	0,57	0,24	0,45
Clus 3.	0,65	0,80	0,58	0,59	0,65	0,63

Cuadro n.º 5: Medias de las subdimensiones de personalidad para los mejores resultados obtenidos del experimento 1 (Cursos: Programación y Metodología)

Programación - Personalidad(10) (K-means - 3 Clusters)											
	Di	Do	Cp	Co	Es	Pe	Ce	Ci	Ac	Ae	Nota
Clus 1.	0,62	0,66	0,55	0,40	0,63	0,84	0,62	0,38	0,64	0,67	0,57
Clus 2.	0,32	0,24	0,41	0,19	0,28	0,21	0,54	0,41	0,35	0,34	0,73
Clus 3.	0,41	0,42	0,57	0,48	0,47	0,49	0,64	0,59	0,51	0,46	0,57
Metodol. Estudio - Personalidad(10) (K-means - 3 Clusters)											
	Di	Do	Cp	Co	Es	Pe	Ce	Ci	Λc	Λe	Nota
Clus 1.	0,58	0,69	0,42	0,30	0,69	0,83	0,66	0,40	0,63	0,57	0,79
Clus 2.	0,49	0,44	0,68	0,60	0,45	0,60	0,65	0,59	0,58	0,58	0,73
Clus 3.	0,32	0,28	0,44	0,26	0,37	0,27	0,56	0,46	0,38	0,38	0,68

Álg. y Geom. - Person.(10) (K-means - 3 Clus)											
	Di	Do	Cp	Co	Es	Pe	Ce	Ci	Ac	Ae	Nota
Clus 1.	0,36	0,35	0,51	0,38	0,38	0,36	0,60	0,54	0,44	0,41	0,52
Clus 2.	0,62	0,68	0,45	0,38	0,61	0,82	0,79	0,45	0,64	0,59	0,67
Clus 3.	0,50	0,47	0,70	0,52	0,64	0,68	0,45	0,42	0,65	0,66	0,64
Álg. y Geom. - Person.(10) (Neural Gas - 3 Clus)											
	Di	Do	Cp	Co	Es	Pe	Ce	Ci	Ac	Ae	Nota
Clus 1.	0,38	0,40	0,62	0,51	0,46	0,47	0,68	0,62	0,51	0,49	0,54
Clus 2.	0,59	0,61	0,54	0,42	0,63	0,79	0,61	0,41	0,65	0,63	0,65
Clus 3.	0,35	0,29	0,40	0,25	0,33	0,25	0,53	0,42	0,39	0,33	0,55
Matem. Disc. - Person.(10) (On-line Update - 3 Clus)											
	Di	Do	Cp	Co	Es	Pe	Ce	Ci	Ac	Ae	Nota
Clus 1.	0,56	0,61	0,54	0,46	0,59	0,79	0,64	0,37	0,66	0,62	0,62
Clus 2.	0,41	0,36	0,59	0,54	0,44	0,45	0,69	0,63	0,50	0,47	0,48
Clus 3.	0,31	0,23	0,37	0,25	0,16	0,15	0,52	0,32	0,30	0,35	0,46

DISCUSIÓN

Un punto importante que deberá ser tomado en cuenta en futuras investigaciones es el hecho de que las muestras de alumnos que estudian Ciencias de la Computación deben pertenecer a distintas universidades.

También deberán ser tomadas en cuenta otras variables como la inteligencia emocional, para analizar la influencia de las emociones, intelecto y la personalidad sobre el desempeño académico.

CONCLUSIONES

Los resultados muestran que la dimensión Energía se encuentra presente en todos los experimentos como un atributo influyente sobre la Nota, Teniendo en cuenta que las puntuaciones altas en Energía definen a personas dinámicas, activas, enérgicas, dominantes y locuaces, como lo indica [1]. Subjetivamente se suele etiquetar a los alumnos que estudian informática como alumnos retraídos, pasivos, solitarios; sin embargo, nuestros resultados muestran lo contrario. Para corroborar la generalización de estos resultados sería conveniente reproducir el estudio con muestras más grandes que incluyan alumnos de informática de otras universidades.

Los otros resultados encontrados tienen que ver con el factor no verbal, específicamente significado de palabras. Aunque se hubiera podido intuir que cursos como Matemática Discreta, Álgebra y Geometría y Programación requerirían factores intelectuales no verbales como manipulación de áreas, sentido derecha/izquierda, series numéricas, entre otros, la presencia notoria del factor significado de palabras pone de manifiesto la importancia del mismo en la solución de nuevos problemas relacionados con estos cursos.

Por otro lado, en la mayoría de los experimentos se notó una relación entre la dimensión de Tesón o sus correspondientes subdimensiones (Escrupulosidad y Perseverancia) con la Nota. Según [1], personas con Tesón elevado son personas reflexivas, ordenadas, y perseverantes. La característica de orden no debe ser entendida únicamente como orden físico, sino también como orden y disciplina en el estudio lo cual, sumado a otras habilidades, conlleva generalmente el éxito académico.

REFERENCIAS

- [1] Caprara, C., G. Barbaranelli y L. Borbogni, *Bfq-cuestionario big 5*, 1995.
- [2] B. Efron y R. J. Tibshirani. *An Introduction to the Bootstrap*, New York, Chapman and Hall, 1993.
- [3] Fox, J., *Bootstrapping regression models, appendix to an r and s-plus companion to applied regression*, enero 2002.
- [4] Fritzke, B. “A growing neural gas network learns topologies”, en *Advances in Neural Information Processing Systems 7*, MIT Press, 1995, pp. 625–632.
- [5] Hubert L., P. Arabie, *Comparing partitions. Journal of Classification*, 2:193–218, 1985.
- [6] V. Kumar, *An introduction to cluster analysis for data mining*, Technical report, CS Dept, University of Minnesota, 2000.
- [7] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- [8] T. Sullivan, E., Willis W. Clark, y E. W. Tiegs, *Prueba de madurez mental de california*, Forma abreviada, serie intermedia, 50-s, 1950.
- [9] Tan, P. N., M. Steinbach, V. Kumar, *Introduction to Data Mining*, Adison Wesley, 2005.
- [10] Yeung, K. Y., W. L. Ruzzo, “An empirical study on principal component analysis for clustering gene expression data”, en *Bioinformatics*, 17:763–774, 2001.