

Un análisis crítico sobre las evaluaciones internas del SNI de México a partir de distintas técnicas de análisis de datos

Área de investigación: Informática administrativa

Gerardo Reyes Ruiz

Facultad de Economía y Empresa
Universidad de Barcelona
España
ger_07@hotmail.com

Jordi Suriñach Caralt

Facultad de Economía y Empresa
Universidad de Barcelona
España
ger_07@hotmail.com



Octubre 3, 4 y 5 de 2012
Ciudad Universitaria
México, D.F.

XVII CONGRESO INTERNACIONAL DE CONTADURÍA ADMINISTRACIÓN E INFORMÁTICA

División FCA, Facultad de Economía y Empresa, UNAM - Fotografía: Rutilo López-Chavez

<http://congreso.investiga.fca.unam.mx>

informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax

52 (55) 5616.03.08



ANFECA
Asociación Nacional de Facultades y
Escuelas de Contaduría y Administración

División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

Un análisis crítico sobre las evaluaciones internas del SNI de México a partir de distintas técnicas de análisis de datos.

Resumen

En este artículo se aprovechan cuatro técnicas de análisis y agrupamiento de datos para profundizar sobre los criterios seguidos por el SNI en la evaluación de investigadores. Tres de ellas hacen referencia a métodos estadísticos y econométricos (Análisis de regresión Lineal Múltiple, Análisis Discriminante Lineal y C Means,) y una de ellas se basa en las denominadas de inteligencia artificial (Fuzzy C Means). A partir de dichas técnicas, no se justifica la actual clasificación del SNI en cuatro agrupaciones ni las evaluaciones obtenidas, quizá porque los evaluadores del SNI utilizan información no recogida en las variables reportadas en las solicitudes de evaluación. Además, se demuestra la necesidad de mejorar la información estadística utilizada como base para la evaluación, se señala las diferencias de clasificaciones que existen en las clasificaciones entre áreas de conocimiento y se propone utilizar estas técnicas como complemento a las evaluaciones por pares realizadas actualmente, siempre que se mejore la cantidad y calidad de la información disponible. Al final, ello debe servir para hacer más eficiente la selección de proyectos de investigación y desarrollo en un programa concerniente a la política de investigación en México.

Palabras clave: política de investigación, evaluación de proyectos, algoritmos de agrupamiento de datos, técnicas de pronóstico y distancia Hamming.

JEL Classification: C38, C51 y C67.

Octubre 3, 4 y 5 de 2012
Ciudad Universitaria
México, D.F.

<http://congreso.investiga.fca.unam.mx>
informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90
52 (55) 5622.84.80

Fax 52 (55) 5616.03.08



División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

Un análisis crítico sobre las evaluaciones internas del SNI de México a partir de distintas técnicas de análisis de datos.

1. Introducción.

La selección adecuada de proyectos requiere del manejo de una gran cantidad de información y de hacer suposiciones inteligentes (Prasanta, 2006). Tal como comentan Kan y Zhou (2007) la selección de las inversiones implica la toma de las decisiones acerca de qué proyectos debe acometer una organización dentro de sus limitaciones de capital y de acuerdo con su contribución al objetivo general de maximizar el valor actual neto o la riqueza de los accionistas. En la práctica, este objetivo se articula en cierto número de criterios y métodos de selección cuyo uso depende del entorno de decisión y de las características de las inversiones consideradas. Desde esta perspectiva, la selección de un proyecto de investigación bien puede verse como una inversión, y la elección de varios de ellos como la integración de un portafolio de inversión.

Por su parte, la elección de proyectos conlleva obligadamente a la selección entre recursos humanos. En el contexto de México, y muy en particular en el Sistema Nacional de Investigadores¹ (SNI), se realiza año tras año precisamente la selección de capital humano. Esta selección se hace considerando principalmente, entre otras, la producción científica realizada por un investigador/académico durante al menos sus últimos tres años. El SNI (inversor) finalmente decidirá, mediante la valoración de pares, qué solicitudes aceptará (portafolio) para con ello apoyarlas económicamente (inversión) durante cierto periodo de tiempo. En principio, fruto de esta selección y del conocer su repetición periódica, el investigador aumentará su producción científica (rendimiento).

La inversión en los proyectos de investigación que ofrecen estos investigadores, está más que justificada, ya que la creación y transferencia de nueva ciencia y tecnología no sería posible si un país no posee un nivel de capital humano intelectual adecuado para derivar todo ese posible beneficio (Schultz, 1961; Krueger et al, 1989). La logística que asume el SNI anualmente para la selección de proyectos y/o capital humano intelectual conlleva una gran inversión, de tiempo y recursos tanto humanos como materiales, pero sobre todo de recursos económicos. Es por ello que debe analizarse si el criterio de selección aplicado es adecuado o no.

La selección de una solicitud por parte del SNI indudablemente depende del factor humano ya que la valoración cualitativa, sin lugar a dudas, depende y seguirá dependiendo de un evaluador y/o seleccionador. Sin embargo, ¿cómo ha sido esta selección de solicitudes?, ¿existe en realidad una correspondencia entre la información presentada al SNI por cada investigador aprobado y el nombramiento que se le otorga dentro de dicho sistema?, ¿los

¹ El SNI es un subprograma del Programa de Fomento a la Investigación Científica, establecida por el Gobierno Federal, cuya conducción tiene por objeto promover y fortalecer, a través de una evaluación, la calidad de la investigación científica y tecnológica, y la innovación que se produce en el país.

resultados (*outputs* de investigación) de los investigadores evaluados justifican las resoluciones adoptadas? Con este estudio se pretende avanzar en este conocimiento ayudándonos de distintas técnicas de análisis de datos.

2. Objetivo.

El objetivo de este trabajo es mostrar que diferentes técnicas de análisis y agrupamiento de datos sirven de apoyo y soporte técnico para entender mejor el procedimiento actual de evaluación llevado a cabo por el SNI, y hacer más eficiente la selección de proyectos de investigación y desarrollo en un programa concerniente a la política de investigación en México.

3. Metodología.

En este trabajo se utilizan cuatro herramientas de análisis de datos; tres de ellas hacen referencia a Métodos Estadísticos y Econométricos (Análisis de Regresión Lineal Múltiple, Análisis Discriminante Lineal y *C Means*) y una de ellas se basa en la denominada Inteligencia Artificial (*Fuzzy C Means*). Se consideran estos algoritmos de datos porque permiten detectar tanto el nivel de asociación como la importancia de las variables involucradas en cada uno de ellos. Además, el algoritmo utilizado en dos de estas cuatro técnicas de análisis de datos (*C Means* y *Fuzzy C means*) está considerado entre los mejores diez algoritmos para la clasificación de datos (Wu X. et al, 2008). Por su parte, el Análisis Discriminante Lineal y el Análisis de Regresión Lineal Múltiple son dos técnicas que al transcurso del tiempo se han consolidado por sus eficientes resultados en cuanto a la clasificación y pronóstico de datos se refiere.

El insumo para estas cuatro técnicas de análisis de datos es, por una parte, la producción científica reportada al SNI por cada investigador que solicitó el ingreso/permanencia a dicho sistema de investigación, durante el periodo 1996-2003². Por otra parte, para detectar el total de trabajos por investigador en el ISI, y reportados al SNI durante el periodo 1996-2003, se utilizaron las bases de datos denominadas *Science Citation Index* (SCI) y *Social Science Citation Index* (SSCI), ambas ubicadas en el apartado *ISI Web of Knowledge*.

Con estas técnicas y algoritmos lo que se pretende es, y sumando el factor humano, obtener unos dictámenes más robustos y eficientes por parte del SNI. Este análisis tiene sentido, ya que gran parte de las variables utilizadas por las técnicas de agrupamiento y análisis de datos utilizadas en este trabajo son cuantitativas o cualitativas. Como consecuencia, en la medida de que se disponga de mayor (pero sobre todo mejor) información por parte del SNI, entonces se obtendrán evaluaciones más robustas, las cuales a su vez conllevarán a

² No ha podido considerarse un período temporal más amplio debido a que el SNI sólo proporcionó la información hasta 2003. Ello es una limitación, pero entendemos que no anula el interés del trabajo, puesto que permite ver igualmente las potencialidades de las cuatro técnicas aplicadas, y, además, permite valorar la racionalidad de los criterios de evaluación aplicados por el SNI, similares a los utilizados en un período más reciente.

tener un panorama más claro del potencial de los investigadores mexicanos que integran al SNI.

4. Resultados.

En este apartado se supone que los dictámenes emitidos por las Comisiones Evaluadoras del SNI son dados, es decir, congruentes o no, irrefutables e irrepetibles, y que las resoluciones, correctas a no, son perfectibles.

4.1 Algoritmos para estimar el nombramiento de un investigador aprobado por el SNI, de 1996 a 2003.

4.1.1 Análisis de Regresión Lineal Múltiple.

La búsqueda de relaciones causales es el principal objetivo en un análisis econométrico. El Análisis de Regresión Lineal Múltiple (ARLM), en términos muy generales, opera con variables explicativas que tienen un especial peso o contribución en la configuración de la variable explicada (véase Greene, 2008). Para los fines de este apartado la relación funcional se define como:

$$Y_i = \beta_K X_{ki} + \varepsilon_i$$

donde:

$$Y_i = \begin{cases} 0 & \text{Candidato a Investigador.} \\ 1 & \text{Investigador Nacional Nivel I.} \\ 2 & \text{Investigador Nacional Nivel II.} \\ 3 & \text{Investigador Nacional Nivel III.} \end{cases}$$

X_{ki} variables explicativas.

ε_i variable aleatoria, asociada al error de la estimación.

$i = 1, 2, 3, \dots, n$ es el total de observaciones en cada una de las áreas de conocimiento del SNI, no necesariamente son iguales.

$k = 1, 2, 3, \dots, 28$ es el número de parámetros estimados en cada regresión.



Octubre 3, 4 y 5 de 2012
Ciudad Universitaria
México, D.F.

<http://congreso.investiga.fca.unam.mx>
informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90
52 (55) 5622.84.80

Fax 52 (55) 5616.03.08



ANFECA
Asociación Nacional de Facultades y
Escuelas de Contaduría y Administración

División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

A su vez, estos resultados tiene dos implicaciones: primero que, la información integrada por el SNI no está en correspondencia con cada categoría, es decir, que la información solicitada por este sistema debería de ser diferente para cada nombramiento. En vista de lo cual se obtiene la segunda implicación, es decir, los criterios científicos recogidos en el modelo de regresión sirven para explicar las evaluaciones de los investigadores mexicanos con un nombramiento de Nivel I, pero para los otros niveles existen otros criterios que se utilizan para decidir la valoración. Por su parte, en el cuadro 2 se muestran las variables explicativas que resultaron ser estadísticamente significativas mediante el ARLM.

Cuadro 2. Agrupamiento de variables explicativas estadísticamente significativas mediante ARLM, 1996-2003.

Variable	Area I: Físico- Matemáticas y Ciencias de la Tierra	Area II: Biología y Química	Area III: Medicina y Ciencias de la Salud	Area IV: Humanidades y Ciencias de la Conducta	Area V: Sociales	Area VI: Biotecnología y Ciencias Agropecuarias	Area VII: Ingeniería y Tecnología
C	X	X	X	X	X	X	X
EDAD	X	X	X	X	X	X	X
GRADO	X	X	X	X	X	X	X
SIT	X	X	X	X	X	X	X
SEXO	X	X	X	X	X		X
CITAS	X	X	X		X	X	X
EST_INV	X	X	X	X		X	X
ART_SNI	X	X	X			X	X
CAP_LIB		X		X	X	X	X
PUBLIC_ISI	X	X	X			X	X
TESIS	X		X	X		X	X
UBIC_MEX	X	X		X		X	
LIB_ÉDIT				X	X		X
DISC_SNI					X		X
DES_TEC			X		X		
DISTIN		X				X	
INVITA	X				X		
LIB_TRAD	X			X			
PATENTES	X	X					
RESENAS				X	X		
GRU_INV							X
POSDOC	X						
DOCENCIA							
LIBROS							
MEMORIAS							

Elaboración propia a partir de resultados obtenidos mediante Eviews y datos históricos del SNI.

De este agrupamiento de variables se puede deducir que: existen evidencias estadísticas de que en cada una de las áreas de conocimiento del SNI, se utilizan diferentes criterios de valoración para aprobar una solicitud. En el cuadro 3 se recogen las clasificaciones que hubieran surgido de aplicar la técnica del ARLM, a partir de los indicadores disponibles, indicando las reasignaciones que se hubieran producido. Si se comparan ascensos y descensos, destaca que el modelo de regresión sobre todo ascendería a más Candidatos al

Nivel I⁵ y descendería a investigadores de Niveles II y III⁶. No obstante, el flujo total de las estimaciones indica que las áreas del SNI se hubiesen inclinado hacia un nombramiento de Investigador Nacional Nivel I.

Cuadro 3. Estimaciones mediante ARLM por tipo de movimiento y área del SNI, 1996-2003.

Tipo de movimiento	Area I: Físico Matemáticas y Ciencias de la Tierra	Area II: Biología y Química	Area III: Medicina y Ciencias de la Salud	Area IV: Humanidades y Ciencias de la Conducta	Area V: Sociales	Area VI: Biotecnología y Ciencias Agropecuarias	Area VII: Ingeniería y Tecnología	Total
Ascienden								
Candidato a Nivel I	425	489	242	247	257	260	315	2235
Candidato a Nivel II	0	0	0	1	0	0	0	1
Candidato a Nivel III	1	1	0	0	0	0	1	3
Nivel I a Nivel II	346	234	145	352	202	96	117	1492
Nivel I a Nivel III	1	1	1	1	3	1	1	9
Nivel II a Nivel III	23	16	8	6	3	4	12	72
Subtotal	796	741	396	607	465	361	446	3812
% estimaciones ascienden	18.5	17.2	17.0	16.6	18.7	16.7	16.1	17.3
Descienden								
Nivel I a Candidato	149	143	131	85	58	96	134	796
Nivel II a Nivel I	393	306	174	371	274	170	182	1870
Nivel II a Candidato	1	2	1	3	1	2	1	11
Nivel III a Nivel II	180	110	65	169	80	43	60	707
Nivel III a Nivel I	59	40	22	25	27	18	39	230
Nivel III a Candidato	0	0	0	0	0	1	0	1
Subtotal	782	601	393	653	440	330	416	3615
% estimaciones decienden	18.1	14.0	16.9	17.8	17.6	15.3	15.0	16.4
No Especificado	38	46	18	34	33	39	70	278
% no especificado	0.9	1.1	0.8	0.9	1.3	1.8	2.5	1.3
Estimaciones correctas	2697	2915	1520	2369	1555	1429	1840	14325
% estimaciones correctas	62.5	67.7	65.3	64.7	62.4	66.2	66.4	65.0
Total	4313	4303	2327	3663	2493	2159	2772	22030

Elaboración propia a partir de resultados obtenidos mediante Eviews y datos históricos del SNI.

Es importante resaltar que las evaluaciones de las Comisiones Evaluadoras del SNI no pueden o no tienen por qué coincidir con los resultados del modelo de regresión. Porque al considerar el aspecto subjetivo de un evaluador implica que dichas evaluaciones son perfectibles. Es decir, existen otros factores no cuantitativos que son considerados por estos evaluadores para la asignación de un nombramiento, sobre todo en los dos niveles superiores (Nivel II y Nivel III) del SNI.

Como conclusión, señalar que el modelo nos indica que las comisiones utilizan otros factores distintos a los introducidos en la base de datos (Cuadro I del Anexo) para realizar sus evaluaciones. Ello sobre todo sucede en los niveles II y III. Como vía de mejora se

⁵ 2235 pasarían de Candidatos a Nivel I y sólo 796 de Nivel I a Candidato.

⁶ 707 investigadores pasarían del Nivel III al II, 230 del Nivel III al I, frente a ascensos de 72 (II al III) o 9 (I al III).

propone incorporar a estos factores ahora no explícitos (como por ejemplo la calidad de las publicaciones y/o factor de impacto), para facilitar el proceso de evaluación y evitar así críticas de arbitrariedad en el proceso de selección.

5.1.2 Análisis Discriminante Lineal.

El Análisis Discriminante Lineal (ADL) es un algoritmo supervisado que busca discriminar un subespacio (Shih-Wei et al, 2009), en el que los patrones de pertenencia de las clases estén tan estrechamente agrupados tanto como sea posible, mientras que los patrones pertenecientes a las otras clases estén lo más ampliamente separados (véase Fisher R.A., 1936).

Los resultados obtenidos para los estadísticos de prueba denominados *Lambda de Wilks*, *M de Box* y el contraste de la igualdad de medias⁷ mostraron conjuntamente que la discriminación es posible y tiene sentido. De los resultados obtenidos para los denominados autovalores se pudo afirmar que, para todas las áreas del SNI, la discriminación entre los nombramientos de Candidato y Nivel I se realiza de manera correcta, con un promedio para cada una de las siete áreas definidas por el SNI del 89.1% de la varianza total explicada. Esta discriminación para los nombramientos de Nivel I y Nivel II no fue tan clara, ya que se obtuvo un promedio para cada una de las áreas del SNI de 7.9%, con respecto al total de la varianza total explicada. La discriminación entre los denominados Nivel II y Nivel III fue menos clara aún, debido a que la tercera función discriminante absorbió tan solo el 2.9%, en promedio para cada una de las siete áreas del SNI, con respecto al total de variabilidad explicada. El resumen de este análisis para el total de estimaciones correctas se muestra en el cuadro 4.

Cuadro 4. Resumen de las estimaciones correctas mediante ADL para la variable dependiente NIVEL, 1996-2003.

Nivel del SNI	Total de solicitudes aprobadas por el SNI	Total de estimaciones a partir de ADL	% de estimaciones correctas
Candidato	3967	3265	82.3
Nivel I	12362	6671	54.0
Nivel II	3857	2017	52.3
Nivel III	1074	597	55.6
Total	21260	12550	59.0

Elaboración propia a partir de datos obtenidos mediante SPSS y datos históricos del SNI.

No se contabilizan 770 casos omitidos por el análisis.

Es claro que el nombramiento denominado Candidato a Investigador fue estimado mejor mediante el ADL, mientras que los demás nombramientos presentaron un porcentaje de

⁷ Los resultados son demasiado extensos para mostrarlos en el Anexo, sin embargo, la información está disponible para todo lector interesado en dichos estadísticos de prueba.

aciertos menor al 60%. Este resultado implica que para la categoría de Investigador Nacional (en sus tres niveles) se necesita una mayor, pero sobre todo mejor, información para poder ser diferenciados entre sí, puesto que con la información ahora disponible no se justifican las asignaciones realizadas⁸. Por su parte, el cuadro 5 muestra la reasignación de los posibles movimientos de nivel, durante el periodo 1996-2003 y por área de conocimiento del SNI.

El ADL muestra que el 57.0% del total de solicitudes aprobadas por las siete áreas de conocimiento del SNI fueron estimadas correctamente. Por su parte, el 20.4% del total de dichas solicitudes aprobadas mostraron una producción científica que bien las hacían meritorias a un nivel superior del SNI (sobresaliendo el movimiento de Nivel I a Nivel II). En contraparte, el 19.1% del total de solicitudes aprobadas por el SNI en dicho periodo de estudio, presentaron una producción científica que no correspondía al nivel asignado por las Comisiones Evaluadoras de dicho sistema, es decir, bien pudieron estar en un nivel inferior al asignado (resaltando el movimiento de Nivel I a Candidato). Finalmente, el 3.5% del total de dichas solicitudes aprobadas no pudieron ser agrupadas mediante un análisis ADL.

Cuadro 5. Resultados de las estimaciones mediante ADL para la variable NIVEL, por área del SNI.

Tipo de movimiento	Area I: Físico-Matemáticas y Ciencias de la Tierra	Area II: Biología y Química	Area III: Medicina y Ciencias de la Salud	Area IV: Humanidades y Ciencias de la Conducta	Area V: Sociales	Area VI: Biotecnología y Ciencias Agropecuarias	Area VII: Ingeniería y Tecnología	Total
Ascienden								
Candidato a Nivel I	154	160	85	96	38	82	81	696
Candidato a Nivel II	0	0	0	0	1	0	0	1
Candidato a Nivel III	1	2	0	1	0	0	1	5
Nivel I a Nivel II	451	434	235	480	368	248	202	2418
Nivel I a Nivel III	95	78	44	140	90	40	78	565
Nivel II a Nivel III	210	138	81	169	75	52	88	813
Subtotal	911	812	445	886	572	422	450	4498
% estimaciones ascienden	21.1	18.9	19.1	24.2	22.9	19.5	16.2	20.4
Descienden								
Nivel I a Candidato	496	580	344	267	294	302	425	2708
Nivel II a Nivel I	245	164	90	194	111	58	94	956
Nivel II a Candidato	20	7	1	13	15	9	6	71
Nivel III a Nivel II	83	61	32	66	43	28	27	340
Nivel III a Nivel I	39	20	13	12	11	6	25	126
Nivel III a Candidato	5	0	0	1	1	1	3	11
Subtotal	888	832	480	553	475	404	580	4212
% estimaciones descenden	20.6	19.3	20.6	15.1	19.1	18.7	20.9	19.1
No Especificado	153	226	70	75	61	70	115	770
% no especificado	3.5	5.3	3.0	2.0	2.4	3.2	4.1	3.5
Estimaciones correctas	2361	2433	1332	2149	1385	1263	1627	12550
% estimaciones correctas	54.7	56.5	57.2	58.7	55.6	58.5	58.7	57.0
Total	4313	4303	2327	3663	2493	2159	2772	22030

Elaboración propia a partir de datos obtenidos mediante SPSS y datos históricos del SNI.

Los resultados finales de la clasificación por áreas de conocimiento (véase cuadro 6) muestran que el algoritmo de ADL pudo clasificar el 96.5% de los casos totales, mientras que el restante 3.5% no pudo ser clasificado. Se aprecia que la clasificación fue muy homogénea entre las áreas de conocimiento del SNI, obteniéndose un promedio de casos clasificados correctamente, para cada una de ellas, de casi el 60%.

Cuadro 6. Resumen de la clasificación.

Concepto	Area I: Físico- Matemáticas y Ciencias de la Tierra	Area II: Biología y Química	Area III: Medicina y Ciencias de la Salud	Area IV: Humanidades y Ciencias de la Conducta	Area V: Ciencias Sociales	Area VI: Biotecnología y Ciencias Agropecuarias	Area VII: Ingeniería y Tecnología	Total
Casos totales	4313	4303	2327	3663	2493	2159	2772	22030
No clasificados	153	226	70	75	61	70	115	770
Clasificados	4160	4077	2257	3588	2432	2089	2657	21260
% Clasificados correctamente	56.8%	59.7%	59.0%	59.9%	56.9%	60.5%	61.2%	59.1%

Elaboración propia a partir de datos obtenidos mediante SPSS y datos históricos del SNI.

Por tanto, y en síntesis, el ADL identifica únicamente dos grupos de investigadores o nombramientos, con base en la información reportada al SNI. Este resultado se desprende debido a que por una parte, la función discriminante lineal entre los niveles de Candidato y Nivel I está bien definida. Sin embargo, la categoría de Investigador Nacional, en sus tres niveles (Nivel I, Nivel II y Nivel III) no presenta discriminación alguna (a partir de los datos considerados en la base de datos), por lo que estos tres últimos niveles bien podrían fusionarse tan solo en uno, teniendo por consecuencia únicamente dos categorías en el SNI: aquellos investigadores, relativamente jóvenes, que comienzan su vida en la investigación y los investigadores que ya tienen una línea de investigación propia.

5.2 Algoritmos para detectar las características predominantes de los investigadores aprobados por el SNI, de 1996 a 2003.

5.2.1 Algoritmo *C Means*.

El algoritmo de *C Means* (MacQueen, 1967) es uno de los métodos más comúnmente utilizados para la clasificación y es un algoritmo exclusivo de agrupamiento no jerárquico, donde si un determinado punto de datos pertenece a un grupo definido entonces no puede pertenecer a otro grupo simultáneamente. Sin embargo, uno de los principales problemas de este algoritmo es seleccionar el mejor valor de *C*, es decir, el número de clases o grupos. No obstante, se ha señalado que para estos métodos no jerárquicos se puede tener mayor precisión si el punto de partida y el número de las agrupaciones son preestablecidos (Kuo et al, 2002).

El número total de *clusters* definidos para este algoritmo fueron cuatro (A, B, C y D), a semejanza del SNI. Lo que se desea es establecer una correspondencia entre los nombramientos otorgados por dicho sistema de investigación y los grupos pronosticados mediante este algoritmo. La similitud entre cada grupo pronosticado y el correspondiente nombramiento del SNI estará en función de la producción científica reportada a dicho sistema por cada uno de estos cuatro nombramientos internos.

Los promedios reales⁹ para una solicitud aprobada por el SNI durante el periodo 1996-2003, por nivel y concepto, se muestran en el cuadro 7, mientras que los promedios estimados son presentados en el cuadro 8.

Cuadro 7. Promedios reales para una solicitud aprobada en el SNI, por concepto y nivel 1996-2003.

Concepto	Candidato	Investigador Nacional		
		Nivel I	Nivel II	Nivel III
Artículos	3.7	8.8	15.6	23.0
Publicaciones en el ISI	0.3	0.7	1.4	2.5
Capítulos de libros	0.6	1.8	3.2	5.3
Citas realizadas	2.1	14.1	46.0	84.6
Desarrollos tecnológicos	0.2	0.5	0.5	0.9
Distinciones recibidas	1.9	2.9	4.6	6.1
Cursos académicos impartidos	0.2	0.2	0.1	0.7
Estancias de investigación	0.2	0.1	0.1	0.1
Estancias posdoctorales	0.4	0.6	1.0	1.1
Grupos de investigación	0.2	0.6	1.0	1.2
Invitaciones a congresos	5.7	9.9	13.2	20.5
Libros	0.3	0.8	1.3	1.8
Libros editados	0.1	0.2	0.5	0.9
Libros traducidos	0.0	0.1	0.1	0.2
Memorias en congresos	1.6	3.0	4.2	5.4
Patentes	0.0	0.1	0.2	0.3
Reseñas	0.1	0.3	0.5	0.7
Tesis dirigidas	1.9	5.2	8.1	9.8

Elaboración propia con datos históricos del SNI.

De la aplicación del algoritmo C Means, el 87.1% de las observaciones fueron clasificadas en el conglomerado B del cuadro 8. El mencionado conglomerado B captó el mayor número de solicitudes aprobadas durante dicho periodo (véase gráfica 1), ya que de los investigadores con un nombramiento de Candidato a Investigador el 91.7% fue clasificado en dicho conglomerado B; para los Investigadores Nacionales Nivel I esta clasificación fue del 90.3%; para los Investigadores Nacionales Nivel II fue del 77.2% y finalmente para los Investigadores Nacionales Nivel III fue de 66.7%. Cabe mencionar que el 3.5% del total de solicitudes aprobadas por el SNI durante este periodo de estudio no fueron clasificadas en ningún conglomerado.

⁹ El promedio real hace referencia a la media aritmética obtenida en cada concepto evaluado (trabajos, publicaciones en el ISI, capítulos de libros, etc.) por todas las Comisiones Evaluadoras del SNI.

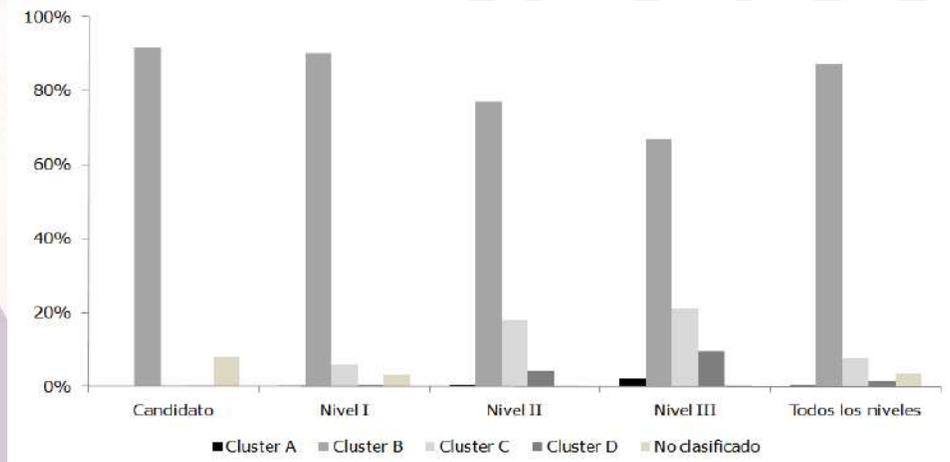
Cuadro 8. Promedios estimados para una solicitud aprobada en el SNI por concepto, 1996-2003.

Concepto	Conglomerado estimado mediante <i>C Means</i>			
	A	B	C	D
Artículos	63.6	8.0	24.3	40.2
Publicaciones en el ISI	4.4	0.7	2.2	3.3
Capítulos de libros	6.7	1.8	3.6	5.8
Citas realizadas	825.4	6.2	109.1	344.5
Desarrollos tecnológicos	0.1	0.5	0.7	0.5
Distinciones recibidas	11.2	2.9	6.0	8.5
Cursos académicos impartidos	0.3	0.2	0.3	0.4
Estancias de investigación	0.2	0.1	0.1	0.1
Estancias posdoctorales	1.5	0.6	1.3	1.8
Grupos de investigación	1.0	0.6	1.0	1.2
Invitaciones a congresos	40.1	9.2	19.0	30.0
Libros	1.9	0.8	1.1	1.3
Libros editados	0.9	0.3	0.5	0.6
Libros traducidos	1.2	0.1	0.1	0.1
Memorias en congresos	7.0	2.7	5.9	9.0
Patentes	0.1	0.1	0.3	0.3
Reseñas	0.2	0.3	0.3	0.9
Tesis dirigidas	15.1	4.9	10.0	11.7

Elaboración propia a partir de datos históricos del SNI agrupados mediante SPSS.

De la mencionada gráfica 1 se desprenden tres comentarios muy importantes: el primero de ellos es que los nombramientos asignados por el SNI durante el periodo 1996-2003, convergen hacia un solo conglomerado, y en particular al *cluster B* de este análisis. El segundo se refiere a que los niveles del SNI denominados Candidato y Nivel I, presentan más atracción hacia dicho *cluster B*, mientras que los dos niveles superiores del SNI (Nivel II y Nivel III) lo hacen también, pero con una menor intensidad. Por último, además del *cluster B*, no se apreció otro *cluster* estimado mediante este algoritmo con una participación significativa. De estos resultados se puede deducir que la producción científica reportada al SNI, por todos los investigadores mexicanos aprobados, no justifica la existencia de cuatro niveles sino que sólo se justifica uno o como máximo dos ([Candidato] o bien [Candidato-Nivel I y Nivel II-Nivel III]).

Gráfica 1. Distribución de los nombramientos del SNI en los conglomerados obtenidos mediante *C Means*, 1996-2003.



Elaboración propia a partir de datos obtenidos mediante SPSS y datos históricos del SNI.

La información obtenida para los promedios reales mediante el algoritmo de *C Means* (cuadro 7) y los promedios estimados (cuadro 8) de una solicitud aprobada por el SNI, durante el periodo 1996-2003, permite llevar a cabo un comparativo mediante la distancia de *Hamming*¹⁰. Esta distancia se define de la siguiente manera:

$$\delta[\mu_{A(x)}, \mu_{B(y)}] = \frac{1}{n} \sum_{k=0}^n |x_k - y_k|$$

$A(x)$ es el vector de los promedios reales en cada nivel del SNI.

$B(y)$ es el vector de los promedios estimados en cada nivel del SNI.

$\mu_{A(x)}$ define a los atributos del conjunto $A(x)$.

$\mu_{B(y)}$ define a los atributos del conjunto $B(y)$.

X_k es el k -ésimo atributo del conjunto $A(x)$.

n es el total de atributos.

Octubre 3, 4 y 5 de 2012

Ciudad Universitaria
México, D.F.

donde:

<http://congreso.investigacion.fca.unam.mx>

informacion.congreso@fca.unam.mx

Teléfono: 52 (55) 5622.84.90

52 (55) 5622.84.80

Fax: 52 (55) 5616.03.08

¹⁰ Se hace uso de la distancia *Hamming* (Hamming, R.W., 1950) ya que los reactivos considerados bien pueden ser considerados como atributos de un perfil deseado.

La finalidad de utilizar esta distancia es detectar la similitud existente entre los vectores reales de la producción asociada a cada nivel del SNI y los vectores estimados mediante el algoritmo *C Means*. Es decir, si no existiera el criterio subjetivo en el proceso de evaluación del SNI entonces los nombramientos definitivamente tendrían una distribución muy diferente a la observada en el periodo de estudio. La matriz de distancias *Hamming* se muestra en el cuadro 9.

Cuadro 9. Matriz de distancias Hamming para los promedios reales y estimados mediante *C Means*.

Nivel / Cluster	A	B	C	D
Candidato	53.4	1.1	9.3	24.5
Nivel I	51.8	0.6	7.6	22.8
Nivel II	49.0	3.4	4.7	19.9
Nivel III	45.6	7.0	1.8	16.6

Elaboración propia a partir de datos históricos del SNI.

Con base en los resultados de la matriz de distancias del cuadro 9 se deduce que el total de solicitudes aprobadas por el SNI, durante el periodo 1996-2003, presentaron un claro agrupamiento hacia dos conglomerados: por su parte, tres de los cuatro nombramientos definidos en el SNI (Candidato, Nivel I y II) mostraron una clara convergencia hacia un solo conglomerado, a decir, el *cluster* B, mientras que los investigadores Nivel III convergen al clúster C. Asimismo, el clúster B presentó mucha similitud (menor distancia de *Hamming*) con el vector real referente a un Investigador Nacional Nivel I del SNI. Este último resultado implica que casi un 90% de los investigadores mexicanos aprobados por el SNI, durante el periodo 1996-2003, tuvieron el perfil productivo de un Investigador Nacional Nivel I.

Para validar la coherencia de los resultados obtenidos en el cuadro 9, se calculó también la matriz de distancias *Hamming* para los promedios reales de los criterios evaluados a los investigadores mexicanos aprobados por el SNI, durante el periodo 1996-2003. La mencionada matriz de distancias se muestra en el cuadro 10. De dicha matriz, destaca que el perfil productivo de un Candidato es muy similar al perfil productivo de un Nivel I¹¹.

Estos resultados obtenidos para el total de solicitudes aprobadas por el SNI durante el periodo 1996-2003, se obtuvieron también haciendo el análisis para la gran mayoría de las áreas del conocimiento definidas por dicho sistema. Es decir, en seis de las siete áreas del

SNI se identificaron, a lo más, dos conglomerados en donde se concentraron una gran proporción de dichas solicitudes aprobadas¹².

Cuadro 10. Matriz de distancias *Hamming* para los promedios reales de los criterios evaluados en el SNI, por nivel 1996-2003.

Nivel del SNI	Candidato	Nivel I	Nivel II	Nivel III
Candidato	0.0	1.7	4.6	8.1
Nivel I	1.7	0.0	2.9	6.4
Nivel II	4.6	2.9	0.0	3.5
Nivel III	8.1	6.4	3.5	0.0

Elaboración propia con datos históricos del SNI.

Otro resultado obtenido fue que en todas las áreas definidas por el SNI se utilizaron criterios internos de evaluación diferentes, ya que al diferir el conglomerado para cada nivel del SNI entonces bien se puede decir que cada área valoró criterios científicos diferentes al aprobar una solicitud, al menos durante el periodo 1996-2003. Además, sobresalió el hecho de que en la gran mayoría de estas áreas de conocimiento el nivel superior del SNI, denominado Investigador Nacional Nivel III, se diferencia claramente de los otros nombramientos.

Como los conglomerados estimados no son los mismos, o no deberían de ser los mismos, en cada una de las áreas de conocimiento del SNI, entonces no puede hacerse un análisis respecto al conglomerado más representativo de todas las áreas. No obstante, se pudo hacer una discriminación para los nombramientos otorgados en cada área del SNI, durante el periodo 1996-2003 (véase cuadro 11).

Cuadro 11. Matriz de nombramientos otorgados por área de conocimiento del SNI, 1996-2003.

Area / Nivel	Candidato	Nivel I	Nivel II	Nivel III
AREA I: Físico-Matemáticas y Ciencias de la Tierra.		X		X
AREA II: Biología y Química.	X		X	
AREA III: Medicina y Ciencias de la Salud.	X		X	
AREA IV: Humanidades y Ciencias de la Salud.	X			X
AREA V: Sociales.		X		X
AREA VI: Biotecnología y Ciencias Agropecuarias.	X			X
AREA VII: Ingeniería y Tecnología.		X		
Todas las áreas		X	X	

Elaboración propia a partir de datos obtenidos mediante SPSS y datos históricos del SNI.

<http://congreso.investigacion.unam.mx>

informacongreso@fca.unam.mx

Teléfono: ¹² En el área de tecnologías sólo se detectó un conglomerado.

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax: 52 (55) 5616.03.08

Por tanto, y para cada una de las áreas de conocimiento del SNI se discriminarían los investigadores en base a los grupos definidos en el cuadro 11. Además y debido a que las áreas de conocimiento del SNI utilizan diferentes criterios de evaluación, los nombramientos emitidos por cada una de ellas no tienen por qué ser homogéneos. Es decir, el nombramiento de Candidato a Investigador, desde el punto de vista cualitativo, no es el mismo en cada una de estas áreas del conocimiento del SNI.

Como conclusión de la aplicación de esta técnica de clasificación, señalar que la producción científica, y más aún, la información solicitada por el SNI no estuvo, al menos durante el periodo 1996-2003, en correspondencia con el nombramiento otorgado, independientemente del área de conocimiento puesto que se deberían agrupar a todos los investigadores, a lo mucho, en dos conglomerados.

5.2.2 Algoritmo Fuzzy C Means.

El algoritmo *Fuzzy C Means* (Bezdek, 1981) identifica grupos relativamente homogéneos basándose en características previamente seleccionadas, como el centroide si se conoce de antemano. El algoritmo *C Means* es el método fundamental de agrupación, mientras que *Fuzzy C Means* (FCM) es el método de agrupación borroso que bien puede ser visto como la mejora de *C Means*, desde el punto de vista de la agrupación de datos (Shian-Chang Huang et al, 2009). La filosofía de agrupación de la lógica difusa es que cada punto tiene un grado de pertenencia a ciertas agrupaciones en lugar de pertenecer completamente a un solo grupo o clase. Además, el algoritmo FCM proporciona la probabilidad máxima de pertenencia a un determinado grupo o clase.

En este apartado, de nueva cuenta se utilizan los datos mostrados en el cuadro I del Anexo y las estimaciones fueron realizadas con el paquete FuzMe¹³ en su versión 3.5c. Debido a que las agrupaciones realizadas mediante el algoritmo FCM dependen del exponente borroso, en este apartado únicamente se muestran las cuatro particiones obtenidas mediante un exponente borroso de 1.75 y una distancia euclidiana. Asimismo, las cuatro clases estimadas mediante este algoritmo no fueron calculadas con un área homogénea. Así, el agrupamiento en las cuatro clases estimadas mediante un algoritmo de FCM para todos los nombramientos asignados por las Comisiones Evaluadoras del SNI, durante el periodo 1996-2003, se presenta en el cuadro 12.

Cuadro 12. Agrupamiento de todas las solicitudes aprobadas por el SNI en las clases estimadas mediante FCM, 1996-2003.

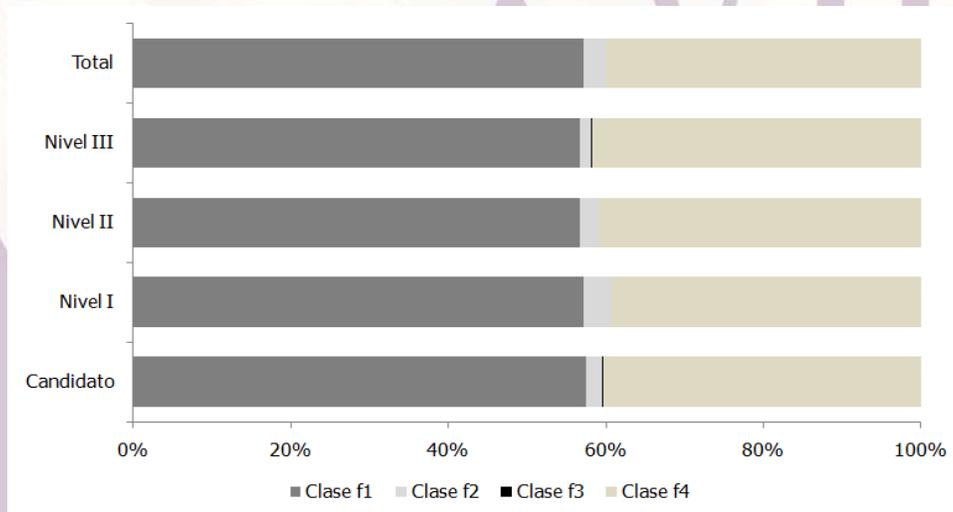
Nivel del SNI	Clase				Total solicitudes agrupadas	Total solicitudes no agrupadas	Total solicitudes aprobadas por el SNI	% Total solicitudes agrupadas
	f1	f2	f3	f4				
Candidato	2349	86	1	1651	4087	229	4316	94.7
Nivel I	7081	415	8	4872	12376	390	12766	96.9
Nivel II	2160	87	2	1560	3809	60	3869	98.4
Nivel III	601	15	1	442	1059	20	1079	98.1
Total	12191	603	12	8525	21331	699	22030	96.8

Elaboración propia a partir de resultados obtenidos mediante FuzMe y datos históricos del SNI.

De los resultados mostrados en el cuadro 12 se aprecia que el agrupamiento para cada nombramiento del SNI fue muy representativo, y de manera global se puede decir que el algoritmo de FCM presentó un nivel de agrupamiento bastante aceptable, el cual fue de casi el 97%. Por su parte, la participación porcentual de las cuatro clases estimadas, respecto al total de solicitudes aprobadas que fueron agrupadas mediante FCM, se presenta en la gráfica 2.

Mediante un agrupamiento de FCM se detectaron tres clases. Sin embargo, dicho agrupamiento se concentró en dos de ellas principalmente, a decir, la clase f1 y la clase f4. La clase f1 concentró el 57.2% del total de solicitudes clasificadas y la clase f4 agrupó al 40.0% respecto al total de solicitudes aprobadas clasificadas. Es decir, el 97.2% de las solicitudes aprobadas por el SNI fueron clasificadas en tan solo dos clases. Este resultado confirma el resultado ya obtenido en técnicas anteriores de que las solicitudes aprobadas por el SNI, durante el periodo 1996-2003, presentaron en su gran mayoría dos perfiles productivos, y no cuatro como se definen actualmente.

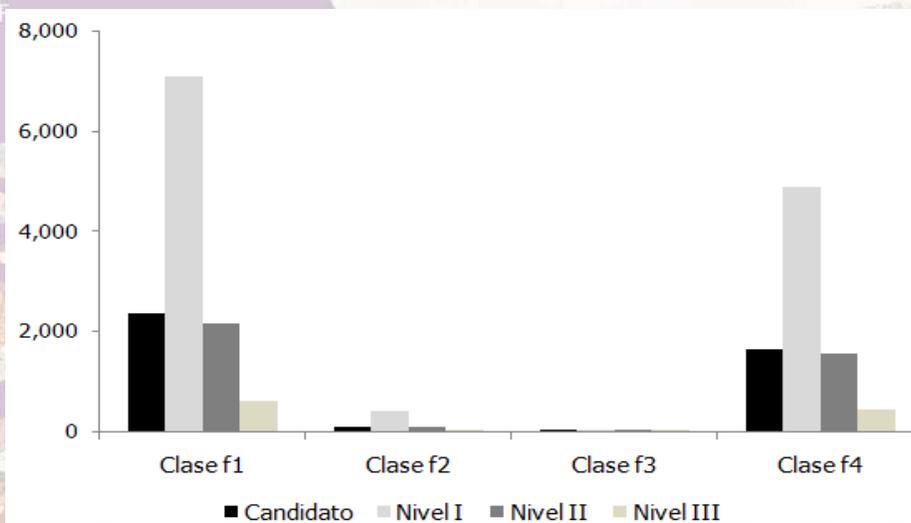
Gráfica 2. Participación de las clases estimadas mediante FCM por nivel del SNI, 1996-2003.



Elaboración propia a partir de resultados obtenidos mediante FuzMe y datos históricos del SNI.

Por su parte, la distribución de los nombramientos clasificados en las cuatro clases estimadas mediante FCM se muestra en la gráfica 3.

Gráfica 3. Distribución de los nombramientos otorgados por el SNI en las clases estimadas mediante FCM, 1996-2003.



Elaboración propia a partir de resultados obtenidos mediante FuzMe y datos históricos del SNI.

En dicha gráfica 3 se puede apreciar que la clase f1 tiene una distribución muy similar a la de la clase f4, en cuanto a los nombramientos del SNI se refiere. Sin embargo, sobresale el hecho de que el algoritmo FCM separa muy bien las dos clases identificadas. También es importante el resultado referente a que en las clases f1 y f4 el nombramiento predominante haya sido el de Investigador Nacional Nivel I. Este último resultado implica por una parte, que los niveles productivos en las clases f1 y f4 giran en torno al nivel productivo de un Investigador Nacional Nivel I del SNI y por otra parte, que el nombramiento definido como Nivel I es el punto de intersección predominante entre las clases f1 y f4, y que bien podría ser el nivel distintivo entre los niveles superiores del SNI. Estas implicaciones refuerzan el resultado de que la información solicitada por dicho sistema de investigación mexicano debería de ser diferente para cada nivel, lo cual a su vez podría derivar en una discriminación más eficiente entre las solicitudes que son presentadas a dicho círculo de investigación.

La información mostrada en la gráfica 3 valida el resultado de que la producción científica reportada al SNI, por los investigadores mexicanos aprobados, no justifica la existencia de cuatro niveles sino que a lo mucho dos. No obstante, el nombramiento definido como Nivel I sería mayoritario para estas dos clases o grupos estimados. Mediante FCM, y tomando en consideración las siete áreas del conocimiento definidas por el SNI, se obtuvo que ya no fueron dos las clases que se definieron predominantemente en la gran mayoría de ellas. Es decir, las áreas de conocimiento del SNI mostraron más diversidad en cuanto a otorgar un nombramiento se refiere. Sin embargo, sobresalió el hecho de que en ninguna de estas áreas del SNI se contabilizaran más de tres clases significativas. Esta misma diversidad se pudo apreciar para el agrupamiento realizado por nivel del SNI, en donde resaltaron los dos niveles superiores de dicho sistema. Por lo que con base en la producción científica reportada al SNI no es tan claro realizar una diferenciación entre estos dos niveles superiores del SNI.

La agrupación realizada mediante un algoritmo FCM presentó más diversidad en las cuatro clases estimadas. Sin embargo, fue claro que en la gran mayoría de estas clases estimadas el nombramiento del SNI con más participación relativa fue el denominado Nivel I. Este último resultado conllevó a los mostrados en el cuadro 13.

<http://congreso.investiga.fca.unam.mx>
informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90
52 (55) 5622.84.80

Fax

52 (55) 5616.03.08



División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

Cuadro 13. Nombramiento predominante para las clases estimadas, mediante el algoritmo FCM, con mayor participación en cada área del SNI, 1996-2003.

Area del SNI / Clase estimada por FCM	f1	f2	f3	f4
Area I: Físico-Matemáticas y Ciencias de la Tierra.			Nivel I	
Area II: Biología y Química.		Nivel I		
Area III: Medicina y Ciencias de la Salud.	Nivel I	Nivel I		
Area IV: Humanidades y Ciencias de la Conducta.	Nivel I	Nivel I		Nivel I
Area V: Sociales.	Nivel I	Nivel I		Nivel I
Area VI: Biotecnología y Ciencias Agropecuarias.		Nivel I	Nivel I	Nivel I
Area VII: Ingeniería y Tecnología.		Nivel I	Nivel I	

Elaboración propia a partir de resultados obtenidos mediante FuzMe.

La implicación inmediata de la información mostrada en dicho cuadro es trascendental, ya que confirma el hecho de que independientemente de la clase estimada y del área de conocimiento definida por el SNI, el perfil productivo predominante será el de un Investigador Nacional Nivel I.

5. Comentarios a modo de conclusiones.

Los algoritmos utilizados en este trabajo muestran que el SNI, visto como un todo, presenta un perfil productivo similar al de un Investigador Nacional Nivel I. Este resultado implica que la producción científica reportada por los investigadores mexicanos aprobados en este sistema de investigación, no se corresponde con el nombramiento otorgado por los evaluadores que conformaron cada una de las Comisiones Evaluadoras del SNI. En consecuencia, la información solicitada por el SNI debe ser diferente para cada nombramiento, y más aún, debe de ser diferente en cada una de las siete áreas de conocimiento definidas por dicho círculo de investigación mexicano. Ello debido a que los Criterios Internos a evaluar en cada una de estas áreas son diferentes y poco homogéneos.

Los resultados obtenidos indican que no se justifica la existencia de cuatro niveles diferenciados de investigadores (Candidato, Nivel I, II y III), sino que como máximo se justifican dos niveles. Este hecho, junto con el bajo nivel de ajuste obtenido en los modelos de regresión, hace pensar que los evaluadores del SNI deben utilizar información adicional a la reportada en la base de datos del SNI, que debería ser integrada, a ser posible, en la misma. Dado que el mayor desajuste se produce en los niveles II y III, es probable que tuvieran que incluirse indicadores sobre la calidad de la investigación realizada (adicional a la cantidad). En este sentido, se sugiere la incorporación de variables como el número de

<http://cc>

informa

Teléfonos

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax

52 (55) 5616.03.08

Asesoría Nacional de Facultades y Escuelas de Contaduría y Administración

División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

trabajos y/o citas publicadas en revistas del ISI-JCR. Los dos grupos que surgen de las técnicas aplicadas no son claros, y en algunos casos, son contradictorios entre ellas. Quizá los dos bloques que parecen ser los más claros son los investigadores mexicanos con registro en el SNI que están definiendo una línea de investigación propia y por otra parte, a los investigadores mexicanos del SNI que ya cuentan con experiencia en investigación y/o tienen una línea de investigación consolidada, al menos en el ámbito nacional. Con base en la producción científica reportada a este círculo de investigación mexicano en cada solicitud aprobada, el nombramiento definido como Investigador Nacional Nivel I, bien podría ser el nombramiento que separe a estos dos grupos de investigadores.

Del análisis efectuado se desprende la necesidad de utilizar información estadística de calidad para proceder a la evaluación. Al ser presentada con mayor responsabilidad por parte del solicitante mejor será la información integrada por el SNI, y en consecuencia serán mejores y más representativos los nombramientos emitidos por cada una de las Comisiones Evaluadoras del SNI. No obstante, la valoración final de una solicitud, seguirá dependiendo del criterio subjetivo individual de los evaluadores que integran dichas Comisiones Evaluadoras del SNI.

En este trabajo se presentan algunas técnicas para hacer más eficiente el proceso de selección en el SNI con el objetivo de, primero, captar información de calidad y de primera mano; segundo, hacer más eficiente la recepción de la información; tercero, reducir el tiempo de respuesta; cuarto, obtener resultados más confiables, los cuales a su vez, se transforman en nombramientos más representativos; quinto y tal vez lo más importante, las técnicas aplicadas, además, deben ofrecerse como buenos instrumentos para complementar la evaluación del SNI por pares, siempre que mejoren la información cualitativa que ahora parecen utilizar los evaluadores, y que no queda recogida en las variables actuales.

<http://congreso.investiga.fca.unam.mx>

informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax

52 (55) 5616.03.08



División de Investigación, Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510

Referencias bibliográficas.

- Bezdek, J.C. (1981), Pattern recognition with Fuzzy Objective Function Algorithms. Ed. Plenum Press, New York.
- Fisher, R.A. (1936), The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*, 7: 179--188.
- Greene, W. (2008), *Econometric Analysis*, 6th Edition, Prentice Hall, N.Y.U.
- Hamming, R.W. (1950), Error detecting and error correcting codes; *The Bell System Technical Journal*, 26 (2): 147-160.
- Kan, R. y Zhou, G. (2007), Optimal Portfolio Choice with Parameter Uncertainty, *Journal of Financial and Quantitative Analysis*, 42 (3): 621-656.
- Krueger, A. y Ruttan V. (1989), Development thought and development assistance. En A. Krueger, Michalopoulos, C. and W. Ruttan, V. et al (Eds), *Aid and Development* (pp. 13-28). The Johns Hopkins University Press, Baltimore, USA.
- Kuo, R.J., Ho, L.M., y Hu, C.M. (2002), Integration of self-organizing feature map and K-means algorithm for market segmentation, *Computers & Operations Research*, 29 (11): 1475-1493.
- MacQueen, J.B. (1967), Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1: 281-297.
- Prasanta, K.D. (2006), Integrated project evaluation and selection using multiple-attribute decision-making technique, *International Journal Production Economics*, 103: 90-103.
- Schultz, T. (1961), Investment in human capital, *American Economic Review*, 51: 1-17.
- Shian-Chang, H., En-Chi, C. y Hsin-Hung, W. (2009), A case study of applying data mining techniques in an outfitter's customer value analysis, *Expert Systems with Applications*, 36: 5909-5915.
- Shih-Wei, L. y Shih-Chieh, C. (2009), A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis, *Applied Soft Computing*, 9: 1008-1015.
- Wu, X., Kumar, V., Quinlan, J.R. et al (2008), Top 10 algorithms in data mining, *Knowl Inf Syst*, 14: 1-37.

<http://congreso.investiga.fca.unam.mx>

informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax

52 (55) 5616.03.08



A n e x o.

Cuadro I. Descripción de las variables explicativas.

- 1 SEXO (sexo del solicitante). Hombre=1 y Mujer=0
- 2 EDAD (edad del investigador al momento de presentar su solicitud al SNI).
- 3 GRADO (último grado académico reportado al SNI por el solicitante). Doctorado=3; Maestría=2; Especialidad=1 y Licenciatura=0
- 4 NA (nivel anterior del investigador, únicamente para reingresos vigentes). Nivel III=3; Nivel II=2; Nivel I=1 y Candidato=0
- 5 NIVEL (nivel asignado por el SNI al momento de ingreso). Nivel III=3; Nivel II=2; Nivel I=1 y Candidato=0
- 6 AREA (área de conocimiento definida por el SNI). Físico-Matemáticas y ciencias de la tierra=1; Biología y Química=2; Medicina y ciencias de la salud=3; Humanidades y Ciencias de la conducta=4; Sociales=5; Biotecnología y ciencias Agropecuarias=6; Ingeniería y Tecnología=7.
- 7 DISC_SNI (disciplina asociada al área de conocimiento definida por el SNI). Clave del SNI asignada a cada disciplina del conocimiento.
- 8 SIT (situación del solicitante; Reingreso Vigente, Nuevo Ingreso, Reingreso No Vigente). Reingreso Vigente=2; Nuevo Ingreso=1 y Reingreso No Vigente=0
- 9 INST (institución de adscripción en México del solicitante). Clave del SNI asignada a cada institución de adscripción.
- 10 UBIC_MEX (ubicación geográfica en México de la institución de adscripción del solicitante). Aguascalientes=1;.....Distrito Federal=9;.....Zacatecas=31 y NE=0
- 11 ART_SNI (artículos reportados al SNI por el solicitante ya sean estos publicados, aceptados ó enviados). Número de artículos=0,1,2,.....n
- 12 CAP_LIB (capítulos de libros reportados por el solicitante). Número de capítulos de libros=0,1,2,.....n
- 13 CITAS (citas recibidas a los trabajos del solicitante). Total de citas=0,1,2,.....n
- 14 DES_TEC (desarrollos tecnológicos realizados por el solicitante). Número de desarrollos tecnológicos=0,1,2,.....n
- 15 DISTIN (distinciones recibidas por el solicitante). Número de distinciones=0,1,2,.....n
- 16 DOCENCIA (total de cursos académicos impartidos por el solicitante). Número de cursos impartidos=0,1,2,.....n
- 17 EST_INV (estancias de investigación realizadas por el solicitante). Número de estancias de investigación=0,1,2,.....n
- 18 POSDOC (posdoctorados realizados por el solicitante). Número de posdoctorados=0,1,2,.....n
- 19 GRU_INV (grupos de investigación a los que pertenece el solicitante). Número de grupos de investigación=0,1,2,.....n
- 20 INVITA (invitaciones a congresos nacionales o internacionales). Número de congresos=0,1,2,.....n
- 21 LIBROS (libros reportados por el solicitante). Número de libros=0,1,2,.....n
- 22 LIBEDIT (libros editados reportados por el solicitante). Número de libros editados=0,1,2,.....n
- 23 LIB_TRAD (libros traducidos reportados por el solicitante). Numero de libros traducidos=0,1,2,.....n
- 24 MEMORIAS (memorias de congresos reportados por el solicitante). Número de memorias=0,1,2,.....n
- 25 PATENTES (patentes registradas reportados por el solicitante). Número de patentes=0,1,2,.....n
- 26 RESENAS (reseñas reportadas por el solicitante). Número de reseñas=0,1,2,.....n
- 27 TESIS (tesis dirigidas reportadas por el solicitante). Número de tesis=0,1,2,.....n
- 28 PUBIC_ISI (total de publicaciones del investigador con registro SNI en el ISI). Número de publicaciones en el ISI=0,1,2,.....n

México, D.F.

<http://congreso.investiga.fca.unam.mx>

informacongreso@fca.unam.mx

Teléfonos

52 (55) 5622.84.90

52 (55) 5622.84.80

Fax

52 (55) 5616.03.08



ANFECA
Asociación Nacional de Facultades y
Escuelas de Contaduría y Administración

División de Investigación. Facultad de Contaduría y Administración, UNAM
Circuito Exterior s/n, Ciudad Universitaria, México, D.F., C.P. 04510