

## DESATANDO LA ESTADÍSTICA: APLICACIONES DE LEARNING ANALYTICS EN OPTIMIZACIÓN DE GRUPOS DE CAPACITACIÓN

---

Luis Medina Gual

Facultad de Educación, Centro de Alta Dirección en Ingeniería y Tecnología (CADIT),  
Universidad Anáhuac

ADRIÁN GARCÍA RUÍZ

Centro de Alta Dirección en Ingeniería y Tecnología (CADIT), Universidad Anáhuac

**RESUMEN:** El presente estudio tiene como propósito el discutir sobre la analítica del aprendizaje como una nueva tendencia en análisis de datos coherentes con el tipo de información que es posible recolectar en los albores del Siglo XXI. Para mostrar las posibilidades de este tipo de análisis se desarrolló un ejercicio de clusterización o elaboración de perfil de capacitación de directivos de instituciones de educación básica mexicanas a partir de un

instrumento aplicado a una muestra de n=2962 sujetos. El análisis develó la existencia de cinco segmentos que permiten proponer diferentes tipos de capacitaciones para los directivos encuestados.

**PALABRAS CLAVE:** investigación educativa, investigación cuantitativa, estadística, formación profesional, directores.

### Introducción

En el año 2013, el reporte Horizon (Johnson, et al., 2013) señaló como una de las tendencias que habrán de adoptarse en dos o tres años para la educación al “learning analytics” o “analítica del aprendizaje”. Esta nueva tendencia del análisis de datos en educación es relativamente nueva y se ha desarrollado a partir del crecimiento desarrollo de paquetes estadísticos especializados como SAS.

El origen de la “analítica del aprendizaje” tiene sus raíces en el “*business analytics*” (“analítica de los negocios”) que comenzó a emerger durante el inicio del siglo XXI. Quizá uno de los primeros autores en tratar sobre este enfoque fue Davenport (2006), quien argumentaba que cada vez era mucho más difícil diferenciarse de otras compañías por el simple hecho de poseer un “buen” producto. Para revertir esta situación, Davenport (2006) invitaba a usar tecnología de recolección y análisis de información sofisticada con el fin de determinar no únicamente qué es lo que los consumidores deseaban comprar sino

también averiguar cuándo, dónde, para qué y a qué precio lo querían comprar y, sobretodo, qué es lo que haría de un consumidor un cliente fiel.

En el caso del ámbito nacional habría que mencionar que la analítica de negocios no ha sido del todo aceptada y conocida. En noviembre de 2012, Mares escribió en el periódico “El Economista” un artículo titulado “¡Es la inteligencia analítica, estúpidos!” donde discutía la paradoja que en México ya existen programas de posgrado que preparan a los estudiantes en estos avatares y cómo algunas veces los egresados mexicanos no eran tan solicitados en empresas nacionales por el desconocimiento de los empresarios sobre esta tendencia. Sin embargo, advierte seriamente de las implicaciones de ignorarla en mercados globalizados e interconectados.

Pero en este punto valdría preguntar: ¿Qué es la analítica? ¿Qué diferencia existe entre la analítica y la estadística “tradicional”? y ¿Cómo se puede aplicar en y para contextos educativos?

Para comenzar a hablar sobre qué es la analítica se podría poner en contexto el hecho de que hace sólo una década, los datos recolectados por las computadoras fluían bit por bit. Hoy por hoy nos movemos en megabytes por megabytes y parece ser que no ha sido posible encontrar un límite a este acelerado ritmo de captura y sistematización de datos (KnowledgeAdvisors, 2012).

Muy a pesar de lo anterior, sería posible afirmar la existencia una paradoja: en la era del conocimiento, los datos sólo han servido para brindan información. Bajo la premisa anterior es que surge la idea de la analítica: una disciplina que permite no sólo usar los datos para describir la realidad sino que la buscar entender y comprender para lograr modelarla y así lograr optimizar los procesos y resultados insertos en esta.

A este respecto, el artículo How to Get a Seat at the Boardroom Table (Fraser, 2008) comienza con dos preguntas muy interesantes y que para quienes se dedican a trabajar con datos todos los días, suenan muy familiares: ¿La información o datos que proporcionamos normalmente son ignorados? Y ¿Se han tomado malas decisiones por una confianza desmedida en los datos?

En este sentido, la analítica diferencia diversos grados de uso de los datos (Baker, 2007):

- a. Información (information): dar significado a los datos en un contexto.
- b. Conocimiento (knowledge): analizar y sintetizar la información generada.
- c. Sabiduría (wisdom): usar el conocimiento para el establecimiento/diseño de logros o metas.

Para lograr lo anterior, los académicos que han trabajado en la analítica definieron un modelo de análisis conocido como CRISP-DM que considera 7 pasos para el análisis de datos: entendimiento del negocio (o de lo que se pretende realizar), entendimiento y exploración de los datos (estadísticos descriptivos), preparación de datos, modelación, evaluación de modelos e implementación.

En general, el centro o corazón de la analítica se encuentra en el diseño de modelos que permitan dar cuenta de la realidad, actores y procesos. El desarrollo de modelos se divide en dos tipos (Berberena, 2009):

- a. Modelación supervisada: donde se cuenta con una variable dependiente o “target” que se desea predecir.
- b. Modelación no supervisada: donde no se cuenta con una variable dependiente y lo que se desea es segmentar a los sujetos con base en algunas variables.

Para no perder costumbre, los desarrollos de los negocios en analítica han comenzado a permear el ámbito educativo. Ello quizá facilitado por el reciente incremento de cursos en línea en LMS y de los ahora famosos Massive Online Open Courses (MOOCs) (que pueden tener más de un millar de alumnos a la vez). A través de éstos, se ha potenciado la recolección de información sobre los usuarios (Greller y Hendrick, 2012). Dichos datos son de muy diversa índole, desde el lugar de origen de la conexión de Internet hasta el tiempo transcurrido en algún objeto de aprendizaje, el número de veces que se intenta una test o el tiempo que se tarda en contestar una tarea o consigna.

Esta abundancia de datos ha llegado a provocar la elaboración de proyectos de investigación que tienen como objetivo de mejorar la experiencia de los estudiantes en línea (Chatti, Dyckhoff, Schroeder y Thüs, 2012). Así, en el caso específico de la analítica del aprendizaje, Elias (2011) sugiere que en la educación, el ciclo de analítica debería de considerar cuatro fases: la captura de datos, el procesamiento de la información, la aplicación de los modelos y su difusión entre los actores (*stakeholders*) interesados. Por tanto, y como sugiere Greller y Drachsler (2012), la analítica del aprendizaje sería un proceso paralelo a la evaluación de cualquier ciclo instruccional tradicional.

En resumen, se sugiere que la analítica puede tener las siguientes aplicaciones en educación (Bienkowski, Feng y Means, 2012):

- a. Modelación del conocimiento: conocer a profundidad el nivel de logro de los estudiantes.
- b. Modelación del comportamiento de los estudiantes: caracterización de las acciones del estudiante para determinar el tipo de situaciones o condiciones que lo motivan.
- c. Modelación de la experiencia o vivencia de los estudiantes: para modelar el gusto o satisfacción del estudiante para con el curso y/o sus objetos de aprendizaje
- d. Clusterización de estudiantes o elaboración de perfil: para la segmentación de los estudiantes según su nivel de logro, datos sociodemográficos, motivacionales, trayectoria en el curso, perfil de aprendizaje, etc.
- e. Modelación del conocimiento o mapas de conocimiento: con el fin de detectar conceptos, habilidades o actitudes claves que sirven como prerrequisito indispensable para continuar aprendiendo.
- f. Análisis de principios del aprendizaje y principios instruccionales: estudio de los principios instruccionales o tipos de prácticas instruccionales en diferentes momentos del curso y con diferentes tipos de estudiantes en pos de la eficacia, eficiencia y pertinencia de éstos.
- g. Adaptación y personalización de las experiencias del aprendizaje: con el fin de definir ritmos, rutas y preferencias instruccionales para cada estudiante.

A pesar de la existencia de documentos claves sobre la analítica del aprendizaje, no fue sino hasta años muy recientes que los investigadores educativos a nivel mundial han comenzado a organizarse con el fin de estudiar las posibles aplicaciones de este tipo de análisis. Ejemplo de lo anterior es la *Society for Learning Analytics Research* (SOLAR, <http://www.solaresearch.org>) organizada por diversas universidades americanas entre las que destacan Athabasca University, Stanford University, the University of Queensland entre otras. Hoy por hoy ofertan diversos cursos en formato MOOC que permiten a los interesados el estudiar la analítica del aprendizaje. De igual forma sería importante mencionar la fundación –por parte de este grupo- de la primera revista arbitrada sobre la analítica del aprendizaje durante este año: *Journal of Learning Analytics*.

Por lo anterior, es posible develar un creciente interés en este tipo de análisis por parte de los investigadores educativos a nivel internacional. Así pues, el presente trabajo tiene como propósito el mostrar una posible aplicación de la analítica al campo del aprendizaje. En específico, se planteó como ejercicio la definición de clústers o perfiles de

personas para la conformación de grupos de capacitación con el fin de optimizar los beneficios que se deriven de un posible curso diferenciado (modelación no supervisada, sin variable dependiente o *target*).

## Método

El estudio a realizar es de corte cuantitativo, no experimental y exploratorio. El software utilizado como herramienta de analítica fue SAS. En el caso de los estadígrafos descriptivos se empleó el paquete estadístico SPSS. Para realizar el ejercicio de clusterización se empleó el modelo de análisis conocido como CRISP-DM al utilizar una base de datos provista por Servicios Educativos Anáhuac (SEA) de  $n=2926$  directivos de instituciones de educación básica mexicana. A estos directivos les fue aplicado un breve cuestionario de contexto que tenía como propósito el indagar variables sociodemográficas como sexo, edad, tiempo en el cargo, etc. De igual forma, respondieron un inventario de competencias que tiene como propósito el analizar su perfil directivo. Dicho inventario fue elaborado por Cameron (1997) en inglés y traducido y calibrado por la Editorial Ramón Areces. Una vez respondido el instrumento, los ítems se agrupan en 20 dimensiones que el autor denomina competencias directivas entre las que se encuentran: iniciativa, asunción de riesgos, innovación, flexibilidad/adaptabilidad, etc.

## Resultados

### *Descripción de la muestra*

La muestra final estuvo compuesta por un total de  $n=2926$  directivos. De los cuáles el 32.7% había cursado la Normal o alguna Licenciatura y el 43.3% contaban con maestría. De igual forma es posible mencionar que el rango de edades típico se encontraba entre los 31 y 55 años. El 39.9% eran de sexo masculino y el 60.1% de sexo femenino. En cuanto a la experiencia, más de la mitad de la muestra mencionó encontrarse en los 5 primeros años de experiencia ejerciendo una función directiva (57.5%). Así mismo, cerca de una tercera parte eran directivos de educación inicial, otra tercera parte de preescolar y un último tercio eran directivos de educación primaria. Sobre el tipo de institución destaca el hecho que el 37.2% eran directivos de escuelas urbanas, el 42.8% rurales, 1.7% escuelas comunitarias y 16.8% indígenas. Para finalizar habría que mencionar que los directivos se distribuían de forma similar en todo el territorio nacional.

*Estadígrafos descriptivos (exploración de datos)*

A continuación se mostrarán los estadígrafos descriptivos de las diferentes dimensiones o competencias directivas medidas por el instrumento. Es importante mencionar que estos puntajes son brutos y no reflejan la calificación baremada sugerida por el manual del instrumento. Esta decisión se tomó debido a que la calibración fue realizada en España y, al convertir los puntajes, la varianza de los datos se veía comprometida:

Tabla 1: Estadígrafos descriptivos de las competencias directivas (n=2926)

Statistics											
	Iniciativa	Asunción de riesgos	Innovación	Flexibilidad/Adaptabilidad	Pensamiento analítico	Toma de decisiones	Planificación	Orientación a la calidad	Comunicación oral	Sensibilidad	Relaciones
Mean	25.24	26.28	27.68	26.75	27.18	26.76	26.75	25.88	24.38	24.75	26.26
Std. Deviation	3.424	3.282	2.992	3.275	3.812	3.809	3.537	3.921	3.410	3.369	3.434
Skewness	.096	.037	-.095	-.041	-.003	-.044	.052	.017	.262	.219	-.047
Kurtosis	.342	.444	.751	.158	-.206	-.289	.176	.041	.703	.517	.367
Minimum	10	11	11	11	9	14	14	12	9	14	11
Maximum	39	39	39	38	39	38	38	40	39	40	39

	Logro	Centrado en el usuario	Conciencia de negocio	Orientación al aprendizaje	Autoridad/ Presencia	Motivación a los demás	Desarrollo profesional	Resistencia
Mean	26.29	28.16	28.01	26.79	25.86	25.53	25.46	25.97
Std. Deviation	3.256	3.787	3.804	3.477	3.108	3.148	3.176	3.234
Skewness	.172	-.339	-.106	.095	.246	.209	.219	.160
Kurtosis	.237	.009	-.118	.198	.901	.945	1.119	.468
Minimum	15	13	11	12	11	9	9	12

	Logro	Centrado en el usuario	Conciencia de negocio	Orientación al aprendizaje	Autoridad/ Presencia	Motivación a los demás	Desarrollo profesional	Resistencia
Mean	26.29	28.16	28.01	26.79	25.86	25.53	25.46	25.97
Std. Deviation	3.256	3.787	3.804	3.477	3.108	3.148	3.176	3.234
Skewness	.172	-.339	-.106	.095	.246	.209	.219	.160
Kurtosis	.237	.009	-.118	.198	.901	.945	1.119	.468
Minimum	15	13	11	12	11	9	9	12
Maximum	39	40	40	39	40	39	39	37

Como es posible observar en la tabla anterior, las medias y desviaciones estándares son relativamente similares. Es importante mencionar que se corrió la prueba de bondad de ajuste a la normal Kolmogorov-Smirnov Z. En ningún caso se encontró que los ítems se distribuyeran como la normal ( $p < 0.000$ ).

*Modelación: elaboración de perfiles para capacitación*

Para la fase de modelación se incluyeron en el análisis todas las variables de contexto que se contaban así como las diferentes dimensiones o competencias directivas medidas por el instrumento.

En este sentido, durante dicha fase (creación de grupos óptimos para capacitación – modelación no supervisada), se procedieron a correr diferentes análisis de clusterización empleando distintos tipos de estandarizaciones, métodos y fijando diferentes valores de cortes en el número máximo de segmentos a producir. El total de ejercicios de clusterización producidos fue de 14. Estos modelos se encuentran descritos en la siguiente tabla:

**Tabla 2: Descripción de los ejercicios de clusterización**

Cluster	Estandarización	Método	Máximos
1	Ninguno	Ward	Automáticos

2	Estandarización		
3	Rango		
4	Ninguna		
5	Estandarización	Promedio	
6	Rango		
7	Ninguna		
8	Estandarización	Centroide	
9	Rango		
10	Ninguna		
11		Default	2
12			4
13			6
14			8
			10

De igual forma, se procedió a realizar nueve análisis de redes neuronales no supervisados (*self organized maps*).

#### *Evaluación de los modelos*

A diferencia de los modelos supervisados, que cuentan con diferentes formas de evaluación como la tasa de mal clasificados (en variables nominales) o el error cuadrado promedio (en variables intervalares), se dice que los modelos no supervisados, al no contar con una variable dependiente o *target*, deberían ser valorados según la pertinencia y utilidad de los modelos para segmentar a la muestra. En este sentido, una clusterización que segmentara a la muestra en demasiados grupos con pocos casos dentro de cada subgrupo sería de muy poca utilidad. De igual forma, una clusterización que, al momento de ser interpretada no fuera lógica, no sería considerada como un modelo óptimo.



Así pues, de todos ejercicios de segmentación elaborados, el ejercicio de clusterización 7 fue el que mostró dividir la muestra en un número de segmentos con un número de sujetos razonables y con posibilidad de descripción lógica de los clústers. A continuación se muestra una tabla con la distribución de la muestra en los cinco modelos sugeridos por el ejercicio de clusterización 7:

**Tabla 3: Porcentaje de casos por segmento**

<b>Segmento</b>	<b>Porcentaje</b>
<b>2</b>	25.85%
<b>4</b>	25.17%
<b>5</b>	19.04%
<b>3</b>	15.64%
<b>1</b>	14.28%

Para finalizar el ejercicio sería importante mencionar que las variables de mayor importancia para esta clusterización permiten definir el tipo de capacitación que cada uno de los segmentos requeriría. Como dato interesante valdría mencionar que las variables sociodemográficas no parecieron contribuir a la segmentación:

**Tabla 4: Descripción de los segmentos de capacitación**

<b>Segmento</b>	<b>Tipo de curso sugerido</b>	<b>Variables de mayor importancia – perfil del directivo</b>
1	Entrepreneurship	Puntajes mayores a 28.29 en la escala de iniciativa, 29.95 en la escala de innovación y 29.40 en la escala de flexibilidad.

2	Planeación y calidad	Puntajes cercanos a 30.56 en centrado en el usuario, 21.46 en orientación a la calidad, 21.19 en pensamiento analítico, 24.65 en trabajo en equipo y 24.56 en conciencia de negocio.
3	Liderazgo y trabajo en equipo	Puntajes mayores a 31.82 en conciencia de negocio, 28.57 en trabajo en equipo y 28.56 en toma de decisiones.
4	Planeación para el logro de objetivos	Puntajes cercanos a 27.94 en planificación, 29.42 en pensamiento analítico y 25.00 en relaciones interpersonales.
5	Toma de decisiones y medición del riesgo	Puntajes cercanos a 23.63 en asunción de riesgos, 23.43 en centrados en el usuario, 27.69 en orientación a la calidad y 27.13 en logro.

## Conclusiones

El presente trabajo tuvo como propósito el mostrar un ejercicio de utilización de la analítica del aprendizaje para la elaboración de grupos de capacitación según el perfil de los directivos de educación básica de escuelas mexicanas. Al día de hoy, Servicios Educativos Anáhuac (SEA), emplea este tipo de análisis para la elaboración de

segmentos o grupos de capacitación “optimizados”. En este sentido, el SEA ha reportado una mayor satisfacción por parte de los directivos que han sido capacitados.

Por lo tanto, se podría pensar en la analítica es una herramienta que, si bien utiliza análisis estadísticos que han sido utilizados en el pasado, implican una diferente forma de aproximarse a los fenómenos (como el uso del modelo CRISP-DM. Sería importante reflexionar que la importancia de la analítica ha comenzado a permear el discurso de investigadores y diseñadores de políticas públicas fuera del país. Por tanto, este trabajo busca ser una invitación para los académicos mexicanos para explorar las potencialidades de este tipo de análisis. Este es un perfecto momento para que los investigadores y tomadores de decisiones nacionales se sumen a esta tendencia que comienza a mostrar diferentes aplicaciones en el ámbito educativo.

## Bibliografía

- Baker, B. (2007). A conceptual framework for making knowledge actionable through capital formation. D.Mgt. dissertation, University of Maryland University College, United States – Maryland
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics. Department of Education's, Office of Educational Technology,
- Berberena, V.H. (2009). Redes neuronales y su aplicación en modelos de cobranza. México: 5ta Convención Nacional de Profesionales de Cobranza.
- Cameron, A. (1997). Perfil del directivo: inventario de competencias. Madrid, España: Editorial Ramón Areces – HRD Press.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International

Journal of Technology Enhanced Learning,4(5), 318-331.

Davenport, T. H. (2006). Competing on analytics. Harvard Business Review, 84(1), 98.

Elias, T. (2011). Learning analytics: Definitions, processes, and potential. Unpublished manuscript.

Fraser, N. (2008). How to get a seat at the boardroom table. Open Options Co., Data Mining Conference.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. Educational Technology & Society, 15(3), 42-57.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). NMC horizon report: 2013 higher education edition.

KnowledgeAdvisors. (2012). The role and function of a learning analytics leader. Unpublished manuscript.

Mares, M. A. (2012, 23 de noviembre). ¡Es la inteligencia analítica, estúpidos! El Economista. pp. 26.