

Visión del conocimiento: análisis de eficiencia terminal en una institución de educación superior como una herramienta de toma de decisiones

Vision of Knowledge: terminal efficiency analysis in an institution of higher education as a tool for decision-making

Gabriela Xicoténcatl Ramírez*

Elieth Velázquez Chávez**

Rosa María Romero González***

Fecha de recepción: 11 de marzo de 2013

Fecha de aprobación: 30 de abril de 2013

Resumen

Dentro de las Instituciones de Educación Superior (IES) las problemáticas como los índices de eficiencia terminal, la tasa de titulación, tasa de retención de estudiantes, tasa de permanencia e índices de reprobación son parámetros que miden la calidad educativa en el criterio de estudiantes. Analizar la información que se generan en los diversos sistemas y concentrarla es todo un reto que además apoyaría de manera importante la toma de decisiones.

Técnicas como minería de datos, análisis de regresión y correlación son las herramientas que permiten en el presente trabajo proponer una metodología que ayude a la visión del conocimiento y mediante ella la toma de decisiones que permitan detectar los agentes de cambio necesarios.

* Facultad de Informática, UAQ, Juriquilla, Querétaro México, gabyxico@uaq.mx

** Facultad de Informática, UAQ, Juriquilla, Querétaro México, elieth.velazquez@uaq.mx

*** Facultad de Informática, UAQ, Juriquilla, Querétaro México, rossyrg04@yahoo.com.mx

Palabras clave

Visión del conocimiento, tecnologías de información, eficiencia terminal, estrategia competitiva, institución de educación superior (ies), minería de datos (dm), clustering, soporte a las decisiones.

Abstract

In the Superior Education Institutions there are problems like terminal efficiency, graduation rate, retention rate, student's retention rate and the number of non-approved students, all of them are parameters that measure the quality of education in the students approach. Analyze that kind of information from different systems and get it is a big challenge. This challenge gives an important decision supports to get the competitiveness.

Some techniques like data mining, regression analysis and correlation are tools that help us to propose in this paper a methodology that support the knowledge management and therefore decision support and identify the change agents needed.

Keywords

Data mining, decision support systems, clustering, knowledge management.

1. Introducción

En la administración del conocimiento un elemento clave, es la visión del conocimiento, [19] Krogh, Ichijo y Nonaka (2000) afirman que una visión no solamente es la previsión de un estado futuro, también es necesaria una visión de la situación presente. Una visión del conocimiento se relaciona firmemente con una estrategia de avance, en la que se ponen de relieve el desempeño y el éxito futuros de una compañía u organización, la visión del conocimiento proporciona a los planificadores corporativos un mapa conceptual de tres dominios relacionados entre sí: el estado presente, una imagen del futuro que desean, y el conocimiento que

deben buscar y crear para transitar del presente y alcanzar cierto estado futuro. El caso de estudio es una Institución de Educación Superior, donde se analizarán los tres dominios de la visión del conocimiento.

En el ámbito de las Instituciones de Educación Superior (IES) se busca elevar la competitividad de cada IES en los ámbitos nacional e internacional, siguiendo los lineamientos establecidos desde 2001, en el Programa Nacional de Educación Superior (Pronaes) [1] y en el de PRONAES 2007-2012 [2], así como en los Planes Institucionales de Desarrollo (PIDE) de cada IES, en donde uno de sus principales objetivos es lograr el mejoramiento de la calidad educativa.

La Asociación Nacional de Universidades e Instituciones de Educación Superior (Anuies) retomó este objetivo estratégico y lo han plasmado en diversos programas que han ayudado a las IES a emprender acciones encaminadas al mejoramiento y aseguramiento de su calidad educativa. Las IES han tomado conciencia de que no basta con tener planes de estudio diseñados y actualizados de acuerdo a los lineamientos de la Secretaría de Educación Superior (SES), ni contar con una infraestructura moderna, si no se lleva a cabo el proceso enseñanza-aprendizaje entre los dos actores centrales que son los profesores y los estudiantes. En los últimos años se ha transitado hacia modelos educativos centrados en el aprendizaje del estudiante y ya se tienen avances en este sentido en gran parte de las IES a nivel nacional, según [3].

En el Pronaes [2] se retoma el apoyo integral a los estudiantes, que fue un programa de iniciativa de la Anuies en el 2000, donde se establece que acciones prioritarias como los programas institucionales de tutorías, los estudios de seguimiento de trayectoria académica y la implantación de nuevos métodos educativos, mejorarán la calidad educativa.

2. Visión del conocimiento

En este proyecto se hablará de la creación del conocimiento, Takeuchi (1999) dice que existen dos formas de crear el conocimiento, la primera se presenta en las personas que realizan funciones propias en base a conocimiento tácito, lo acumulan y lo crean o generan con forma de habilidades personales, basadas en la experiencia; y la segunda forma se presenta en los especialistas, que también acumulan, crean o generan y actualizan el conocimiento pero lo movilizan a través de formas estructuradas, de datos técnicos, científicos, y cuantificables.

Juntarung y Ussahawanitchakit [4] comentan que el conocimiento permite que la organización pronostique lo más cercano a la naturaleza y al potencial comercial de cam-

bios en el entorno y propicie la conveniencia de acciones estratégicas y tácticas. El conocimiento después de creado se transfiere y la velocidad efectiva a la que se transfiere el conocimiento dentro de una organización puede afectar significativamente la competitividad y funcionamiento de la organización. Es muy importante la transferencia del conocimiento.

Ahora es importante definir la Administración del Conocimiento, Knowledge Management (KM) por sus siglas en inglés, [4] menciona que es el proceso de administrar el capital intelectual, identificar y explotar los activos intangibles, con la finalidad de desarrollar nuevas oportunidades. [21] Nieves, Artiles y Goñi (2008) comentan que en la actualidad existe la convicción que la administración del conocimiento es una actividad necesaria para mantener y mejorar la competitividad.

Rodríguez y León (2006) afirman que en la administración del conocimiento se administran los activos no materiales de la organización; se genera, busca, almacena y transfiere el conocimiento con el propósito de aumentar la productividad y competitividad de las organizaciones. Las ventajas competitivas que produce una adecuada administración del conocimiento no dependen de la cantidad de conocimiento que se consiga reunir y almacenar sino del uso que se haga de ellos; por ello, es necesario adoptar una cultura corporativa que fomente el intercambio y la colaboración entre los miembros de una organización.

Esta es una buena razón para justificar la propuesta de una metodología basada en un estudio actual, ya que se enfocará al análisis de las bases de datos académicas de alumnos en una Institución de Educación Superior, debido a que muchas ocasiones las IES planean y pocas veces llegan a cumplir esos planes, los pronósticos propician nuevas estrategias y tácticas, generan cambios para mejorar la calidad educativa, una IES que pronostica creará conocimiento y la velocidad con que transfiere ese conocimiento

puede afectar significativamente su competitividad según [4].

Se ha hablado de lograr una adecuada administración del conocimiento, y para lograrla es indispensable la utilización de las tecnologías como herramientas fundamentales para la rápida y adecuada transmisión, generación y difusión del conocimiento como lo comentan Rodríguez y León (2006).

Colle (2005) comenta que existen algunas aplicaciones directamente destinadas a la administración del conocimiento, conocidas como herramientas de transferencia y extracción del conocimiento, de las cuales menciona:

- Manuales de organización y métodos.
- Aplicaciones de trabajo colaborativo (Groupware)
- Bases de datos y sistemas documentales avanzados
- Minería de datos (DM)
- Árboles de conocimiento o gestión de competencias
- Sistemas de aprendizaje
- Sistemas expertos
- Sistemas de apoyo a la toma de decisiones
- Internet / Intranet

3. Minería de datos

Clark y Boswell (2000) afirman que minería de datos es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Por otro lado, Hernández, Ramírez y Ferri (2004) comentan que la minería de datos tiene dos retos: por un lado trabajar con grandes volúmenes de datos, procedentes mayoritariamente de sistemas de información,

con los problemas que ello conlleva como ruido, datos ausentes, intratabilidad, volatilidad, etc., y por el otro lado usar técnicas adecuadas para analizarlos y extraer conocimiento novedoso y útil. Ochoa et al. (2009) mencionan que en los modelos inteligentes que usan minería de datos, se basan principalmente en árboles y reglas de decisión, reglas de asociación, redes neuronales, redes bayesianas, conjuntos aproximados (*rough sets*), algoritmos de agrupación (*clustering*), máquinas de soporte vectorial, algoritmos genéticos y lógica difusa.

Jing (2004) menciona que la minería de datos utiliza una combinación de una base de conocimiento explícita, de habilidades analíticas sofisticadas, y de un conocimiento del dominio para descubrir tendencias y patrones ocultos. Estas tendencias y patrones forman la base de los modelos de pronósticos que permiten a analistas producir nuevas observaciones de datos existentes. Los modelos contienen los pasos, los módulos, y los recursos del proceso de la minería de datos. Algunos modelos de minería de datos de ciertos datos incluyen todo el proceso para un propósito particular, sea agrupación (*clustering*) o de pronóstico (*forecasting*). Un modelo es, sin embargo, diferente de un algoritmo. Un algoritmo es una función matemática que conduce a la minería de datos, tal como una red neuronal, un árbol de clasificación y regresión (C&RT), o K-means.

3.1 Clustering

La agrupación de datos (*clustering*), también conocida como análisis de grupos, análisis de segmentación, análisis de taxonomía, o clasificación sin supervisión, es un método para la construcción de clústers o grupos de objetos, de modo tal que cada par de objetos en un clúster son similares y cada par de objetos, cada uno en diferente clúster son disimilares. El *clustering* de datos frecuentemente se confunde con la clasificación donde los objetos son asignados a clases predefinidas. *En el clustering, las clases son creadas al mismo tiempo que los objetos son clasificados* [5].

Formalmente el problema de clustering consiste en clasificar un conjunto de objetos en grupos homogéneos. Matemáticamente, un grupo de datos dados en un conjunto D puede ser representado por la función:

$$f: D \rightarrow [0,1]^k, x \rightarrow (x)$$

definida como sigue

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{pmatrix} \dots \quad (2)$$

donde $f_i(x) \in [0,1]$ para $i = 1,2,3,\dots,k$ y $x \in D$ y

$$\sum_{i=1}^k f_i(x) = 1 \forall x \in D \quad (2)$$

El clustering constituye una componente importante de los procesos de minería de datos, y consiste en un proceso de exploración y análisis de grandes cantidades de datos con el fin de descubrir información útil.

Clustering está situado como una técnica de DM indirecta donde la minería se hace sin tener un objetivo definido y la meta es descubrir algunas relaciones entre todas las variables, mientras en la minería directa, algunas variables son los objetivos de salida. En clustering de los datos no estamos seguros de lo que se obtendrá [5].

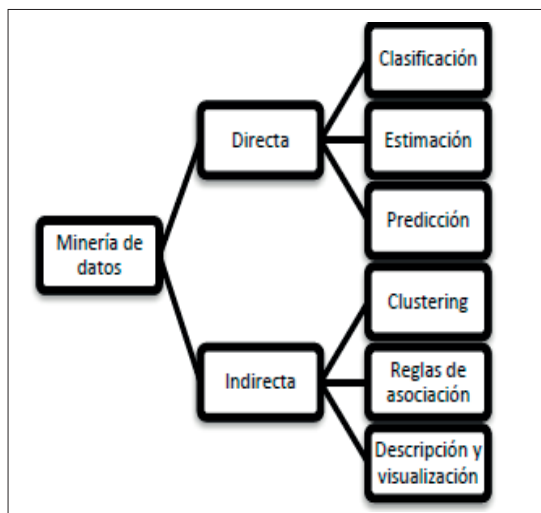
En la figura 1 se presentan las tareas asociadas a los dos tipos de DM.

Definición 1 Un algoritmo de clustering toma como entrada un conjunto de puntos en un espacio multidimensional métrico y como resultado un conjunto de grupos (clústers), $C = \{C_1 C_k\}$.

Esta definición no describe como se forman los clústers o que criterio se considera para crearlos. Esto es porque hay diferentes formas o criterios para la definición de los clústers.

Formalmente, asumimos que la entrada es un conjunto de datos en un espacio métrico, con una función de distancia asociada. Denotaremos esta distancia como d , entonces la

Figura 1. Tipos de minería de datos

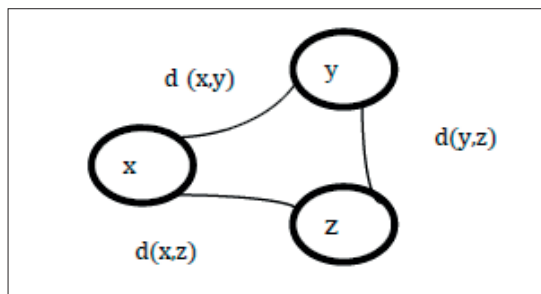


Fuente: [6].

distancia entre dos puntos x y y es dada por $d(x,y)$. Si los puntos están en un espacio métrico entonces se tienen tres requerimientos:

1. **Identidad:** $d(x,x) = 0$; esto es la distancia de cualquier punto así mismo es cero.
2. **Simetría:** $d(x,y) = d(y,x) > 0$ esto es, la distancia entre uno o dos puntos es la misma en ambas direcciones y es no negativa.
3. **Desigualdad Triangular:** $d(x,z) \leq d(x,y) + d(y,z)$; esto es, la longitud entre los puntos jamás superará la longitud total entre dos puntos a través de una ruta que incluye un tercer punto intermedio (ver figura 2).

Figura 2. Desigualdad Triangular



Fuente: elaboración propia..

Las distancias entre los clústers son funciones de las distancias entre datos, y hay varias formas de definir las: Sean y dos clústers.

1. Vecino más cercano: $d(A, B) = \text{mínimo} \{d(i, j)\}$ donde $i \in A, j \in B$. El vecino más cercano tiende a formar clústers más alargados.
2. Vecino más lejano: $d(A, B) = \text{máximo} \{d(i, j)\}$ donde $i \in A, j \in B$. El vecino más lejano forma clústers más esféricos.
3. Centroides: $d(A, B) = d(\bar{A}, \bar{B})$ en que \bar{A} y \bar{B} son los respectivos centroides de los conglomerados A y B.
4. Medoides: es la distancia entre los medoides de los grupos. Es el punto tal que sus coordenadas son las medianas de las variables respectivas.

Definir una función apropiada para la distancia es un reto. Si todos los datos son numéricos se puede utilizar la función de distancia Euclidiana (distancia de la línea recta) o la distancia (distancia máxima para cualquier coordenada). Considerando que los datos reales raramente se comportan así.

Formalmente, el agrupamiento de datos es un problema que consiste en la separación y asignación de datos, definidos como n-adas o vectores en espacios multidimensionales, a grupos o clústers, de acuerdo a algún criterio de similitud. Un clúster se define como un conjunto de objetos similares. El criterio de similitud se establece como la métrica Euclidiana consistente en la función $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, que asigna a cualesquier par de vectores del espacio Euclidiano n-dimensional $x=(x_1, x_2, \dots, x_n)$ y $y=(y_1, y_2, \dots, y_n)$, el número $d = \sqrt{\sum_{1 \leq i \leq n} (x_i - y_i)^2}$, produciendo así la distancia estándar en el espacio \mathbb{R}^n .

Sin embargo, el problema de agrupar un conjunto de n vectores en k clústers bajo funciones objetivo es NP-duro, aún cuando los puntos a ser agrupados se restringen al espacio euclidiano bidimensional [7].

2.1.1 K-Centros

Definición 2 El algoritmo de clustering k-center también conocido como el problema de k-center (kCP), genera un conjunto de k puntos $C = \{c_1, \dots, c_k\}$ (centros). Cada dato de entrada es asociado con un punto en C que sea más cercano a él. La calidad de los grupos está determinada por la distancia máxima de un punto a su centro más cercano, es decir,

$$\max_x \{ \min_i \{d(x, C_i)\} \} \quad (\text{B.Xiao 2001})$$

2.1.2 K.medias

Definición 3 El algoritmo de clustering k-media da como resultado un conjunto C de k puntos (medias) en un espacio métrico que se definen los clústers: cada punto de entrada es asociado con el punto C que están muy cerca de él. La calidad de los clústers está determinada por el promedio de la distancia de los puntos más cercanos al centro, esto es $\frac{1}{n} \sum_x \min_i d(x_i - y_i)^2$.

2.2 Regresión y correlación

El análisis de regresión trata con la descripción y evaluación de la relación entre una variable determinada y una o más variables adicionales.

En el caso de la correlación lineal es relación que presentan dos variables a partir de un coeficiente de correlación. En este trabajo se está utilizando el coeficiente de Pearson.

3. Casos de estudio

Dentro del presente proyecto se está proponiendo generar la visión del conocimiento a partir de dos importantes análisis. El primero es un análisis de las características de los aspirantes a ser alumnos de la Institución de Educación Superior apoyándonos de Clustering una técnica de minería de datos. El segundo es un análisis de Trayectoria Académica a través de técnicas de regresión lineal y correlación utilizando el coeficiente de Pearson.

3.1 Primer análisis: clustering

Considere el caso de estudio consistente en un conjunto de datos que corresponden a un vector de tres atributos obtenidos de la base de datos de los alumnos que han presentado examen de admisión para ingresar a una institución educación superior, correspondiente a los periodos 2003-2008.

La herramienta seleccionada es Mathematica, ya que provee un lenguaje que será utilizado por sus ventajas gráficas y cualidades de lenguaje basado en la programación funcional (programación declarativa basada en la utilización de funciones matemáticas), que facilitan un rápido y eficiente proceso de construcción de prototipos.

Debido a las capacidades de análisis matemático y los algoritmos que provee la herramienta fue seleccionada para los siguientes análisis experimentales. Los análisis experimentales, que se presentan a continuación, serán el principio de la minería de datos. Con el objetivo de buscar información que pudiera apoyar a la toma de decisiones. Sin embargo, es posible que algunos de los experimentos nos arrojen información que podría parecer poco relevante.

Para este análisis experimental se comienza con obtener los datos de entrada que estarán definidos en un espacio tridimensional. Las coordenadas estarán dadas por los mismos datos. El vector se presenta a continuación, aclarando que es una parte debido a que es un vector de 5000 registros.

Para este análisis experimental se comienza con obtener los datos de entrada que estarán definidos en un espacio tridimensional. Las coordenadas estarán dadas por los mismos datos. El vector se presenta a continuación, aclarando que es una parte debido a que es un vector de 5000 registros.

$\{ \{22.,29,177.25\}, \{22.,5,0.\}, \{22.,94,98.\}, \{9.,509,75.5\}, \{11.,511,9$

$5.\}, \{11.,511,95.\}, \{22.,94,135.\}, \{22.,94,135.\}, \{22.,46,76.75\}, \{22.$

$,27,92.25\}, \{22.,88,129.\}, \{22.,88,129.\}, \dots, \{22.,25,117.25\}, \{11.,4$

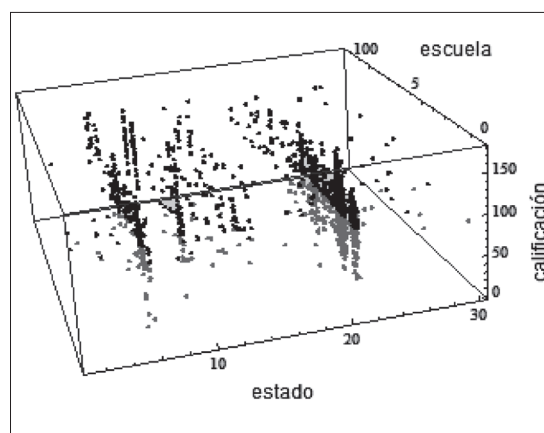
$11,83.\}, \{11.,411,83.\}, \{22.,25,0.\}, \{22.,13,0.\}, \{24.,27,0.\}, \{22.,68,$

$78.75\}, \{9.,30,0.\}, \{22.,85,0.\}, \{22.,75,116.\}, \{22.,75,116.\} \}$

Una solución al problema usando funciones avanzadas de Mathematica consiste en la identificación de clústers como lo ilustra la figura 3. La información que en este caso se está utilizando es el total de alumnos que han hecho solicitud los últimos seis años y han presentado examen para ingresar a la facultad, se puede ver que la población es de diversas partes del país, de diferentes escuelas y las calificaciones son desde cero hasta 150 puntos.

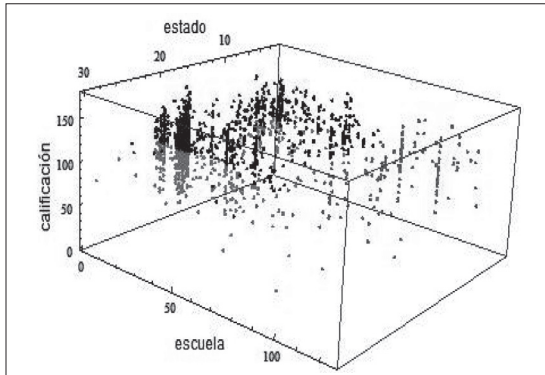
Utilizando un filtro para analizar únicamente la información de los alumnos aceptados en estos años obtenemos los resultados que se muestran en la figura 3. Donde se pueden ver claramente que los grupos se forman con respecto a las calificaciones, las más altas (por arriba de 70 puntos) se localizan con el color púrpura y las que están por debajo con

Figura 3. Gráfica con estado, calificación y escuela de procedencia con $k = 3$



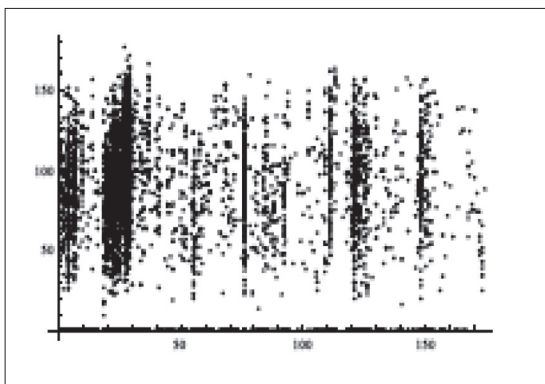
Fuente: elaboración propia.

Figura 4. Gráfica con estado, calificación y escuela de procedencia con k = 4



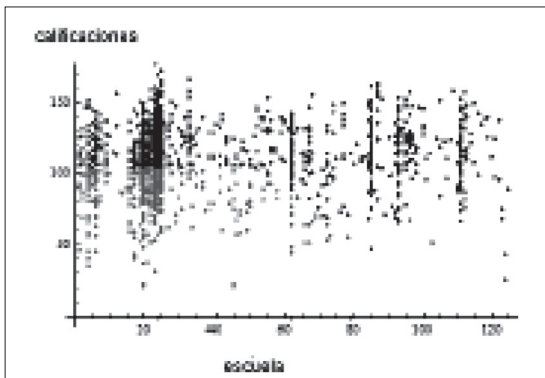
Fuente: elaboración propia.

Figura 5 Gráfica todos los a aspirantes



Fuente: elaboración propia.

Figura 6 Gráfica aspirantes aceptados por escuela y calificación, k=3



Fuente: elaboración propia.

azul. Además de que se están concentrando principalmente en el estado 22 que corresponde a Querétaro.

No podemos identificar solo una escuela como la que ha obtenido las mejores puntuaciones, pero las escuelas que han obtenido más de 150 puntos son las que tienen claves 030, 024, 029, es decir Prepa Norte UAQ, CETIS 16, CBTIS 118, respectivamente.

Sin embargo, se podría imponer un número determinado de clústers k para la separación de los datos de acuerdo a los requerimientos del usuario. Las figuras 3 y 4 muestra los casos para k=3 y k=4, respectivamente.

En la figura 4 encontramos de color amarillo el clúster que identifica las calificaciones más altas. Apreciándose de manera más clara lo que comentábamos arriba.

Una variación del caso de estudio con dos variables nos permite visualizar más claramente los resultados. El siguiente experimento se ha realizado en un espacio de dos dimensiones, las cuales son clave de escuela de procedencia y puntuación obtenida en el examen de admisión.

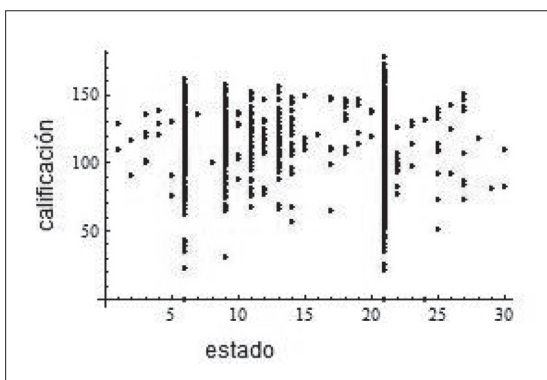
La gráfica sin filtrar a los alumnos aceptados es la 5 y la gráfica 6 representa el resultado con filtro de alumnos aceptados, además de normalizadas las escuelas de procedencia.

Se notan homogéneos los resultados es decir, no hay un clúster solo de puntuaciones mayores a 100. También podemos ver que la mayoría de los aspirantes con calificaciones altas no recaen solo en una escuela, pero si la mayoría están en un grupo que van desde la 20 a la 30. Siendo estas Prepa Sur UAQ, Prepa Norte UAQ, CBTIS 116 y algunas escuelas particulares como el Salesiano.

Otro experimento que complementa la información se ha realizado en un espacio de dos dimensiones.

En la figura 7 el estado 22 que corresponde a Querétaro es el que tiene mayor población

Figura 7 Calificaciones más bajas y más altas $k=4$



Fuente: elaboración propia.

se observa la variedad de puntuaciones obtenidas.

Se puede concluir que, el estado 6 que corresponde a Colima, y el 22 a Querétaro son los que predominan en puntuaciones altas y bajas. El 9 que es el DF también tiene aspirantes con puntuaciones altas y un dato por debajo de los 50 puntos. Guanajuato que corresponde al 11 tiene la mayoría de sus aspirantes con puntuaciones por arriba de 100. Se han mencionado hasta ahora aquellos estados que tienen una cantidad mayor a diez aspirantes y se observa una dominancia de estados circunvecinos a Querétaro. Sin embargo, es interesante mencionar que los aspirantes que vienen de estados lejanos a nuestro estado, como es el caso del estado 2 Baja California Sur y el 1 Aguascalientes, de los cuales se han tenido pocos aspirantes (menos de 10 en los periodos analizados) han sido aspirantes con puntuaciones mayores a 80 puntos.

En cuanto a los resultados del análisis podríamos comentar que aún cuando la base de datos con la que se trabajó es relativamente pequeña y no tiene una estructura relacional, nos permitió encontrar datos interesantes como las escuelas que a lo largo de estos años han traído a la Institución de estudio la mayor cantidad de aspirantes con resultados altos en su examen de admisión

con dicha información se han implementado estrategias para promocionar los planes de estudio en las escuelas cuyos estudiantes históricamente han obtenido las puntuaciones más altas en el examen de admisión, con ello mejorando la competitividad de la Institución. Los análisis anteriores han permitido a la institución posicionarse como una opción altamente competitiva, ya que se hace promoción completamente focalizada en las escuelas de bachillerato que históricamente han adquirido mejores puntuaciones y en los estados que más alumnos aportan. Maximizando de esta manera la influencia de la institución y obteniendo alumnos con mejores promedios y con mayores probabilidades de terminar los estudios de licenciatura elevando de esta manera indicadores de eficiencia terminal.

3.2 Segundo análisis trayectoria académica

En este análisis interesa mejorar la calidad de la trayectoria académica, la cual se refiere al tiempo que el estudiante permanece en una IES, desde que ingresa hasta que egresa de un programa de licenciatura. La trayectoria académica forma parte del rubro que tiene relación con Estudiantes, y tiene indicadores que evalúan la calidad académica en este rubro, entre ellos podemos mencionar los que establecen los Comités Interinstitucionales para la Evaluación de la Educación Superior [8]:

- a) Duración promedio de los estudios: número de años que tardan los estudiantes en finalizar sus estudios respecto del tiempo consignado en el plan de estudios;
- b) Tasa de retención en el primer año: proporción de estudiantes de la misma generación que se matriculan al año siguiente;
- c) Índice de rezago por ciclo escolar: proporción de estudiantes rezagados;
- d) Índice de aprobación: proporción de estudiantes aprobados en todas las asignaturas;

- e) Índice de abandono: proporción de estudiantes que abandonan sus estudios;
- f) Tasa de rendimiento: proporción de estudiantes que concluyen con éxito un ciclo escolar;
- g) Calificación promedio de las asignaturas (últimos cinco años).

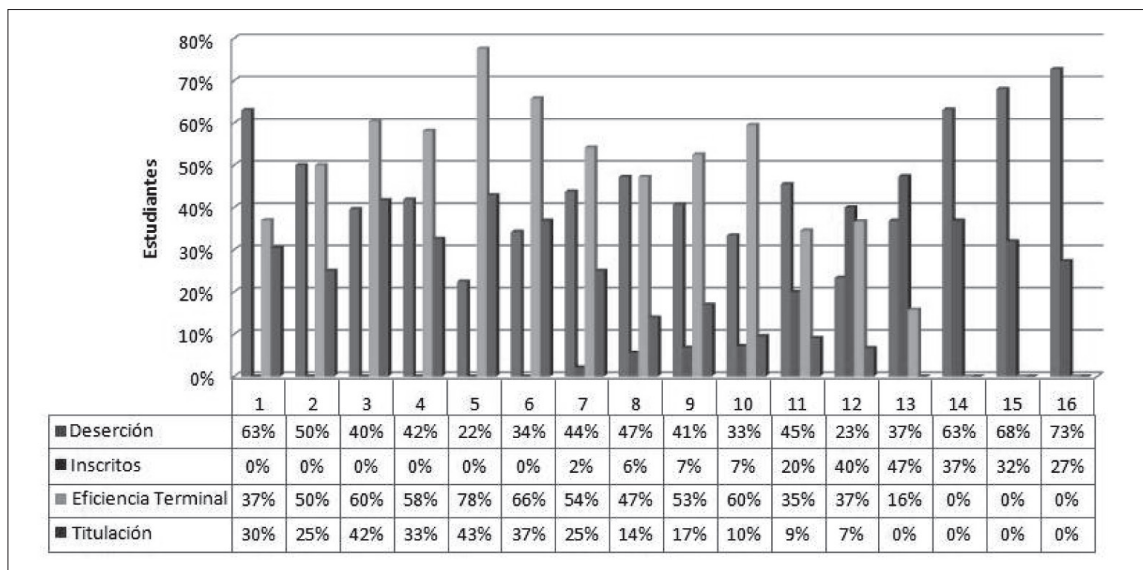
En una IES todos estos indicadores influyen en el rubro de Trascendencia del Programa, específicamente en los indicadores de eficiencia terminal y de titulación, si los estudiantes no tuvieron un buen desempeño durante su trayectoria académica, es difícil que terminen y se titulen en los tiempos y porcentajes mínimos que define la SES, y la realidad nos dice según las estadísticas del Plan Nacional de Desarrollo 2007-2012 (Poder Ejecutivo Federal, 2007b) que la eficiencia terminal en educación superior oscila entre 53 % y 63 %, según el tipo de programa. En todos los planes de desarrollo de las IES se definen políticas y estrategias para mejorar y consolidar la calidad educativa, y se establecen acciones para alcanzarlas, pero muchas veces la realidad de la IES no permite que sean alcanzados todos estos planes, lo que representa un gran problema.

3.2.1 Ingeniería en computación

Se realizó un análisis de los indicadores de resultados académicos del plan de estudios, que se muestra en la Figura, a nivel nacional la Eficiencia Terminal (ET) en Educación Superior oscila entre 53 % y 63 % según el tipo de programa [9], y el comportamiento de las generaciones del plan de estudios 1 es de la siguiente forma, la generación 5 alcanzó una ET de 78 %, de las 6 primeras generaciones, 4 generaciones tienen una ET entre 50 % y 60 %, y solo una generación tuvo una ET menor a 40 %, las generaciones de la 7 a la 13 ya tienen egresados pero aún tienen estudiantes inscritos, indica que existe rezago y aún se tiene la posibilidad de incrementar la ET, las generaciones 14, 15 y 16 tienen un 37 %, 32 % y 27 % de estudiantes inscritos, lo que indica que esos serán los porcentajes de ET máximo, muy por debajo de la media nacional de eficiencia terminal en educación superior.

En un análisis de correlación se mide el grado de asociación entre dos variables. Se calcularon los coeficientes de correlación y se obtuvo un valor *p* menor de 0,01 requerido para la significancia. El análisis muestra que seis de 28 variables (*Ingreso-No Inscritos*,

Figura 8. Trayectoria Académica de la carrera de Ingeniería en Computación (Plan de Estudios 1)



Fuente: elaboración propia.

Ingreso-Deserción, No Inscritos-Deserción, Baja Voluntaria-Deserción, Baja Voluntaria-Inscritos, Egreso-Titulados), fueron estadísticamente significativas y fueron mayores o iguales a .637. La correlación de Ingreso con las bajas por reglamento, bajas voluntarias tienden a ser débiles. La correlación de Ingreso con Egreso y Titulados tienden a ser débiles y no significantes. En general los resultados sugieren que son más los estudiantes que desertan que los que egresan.

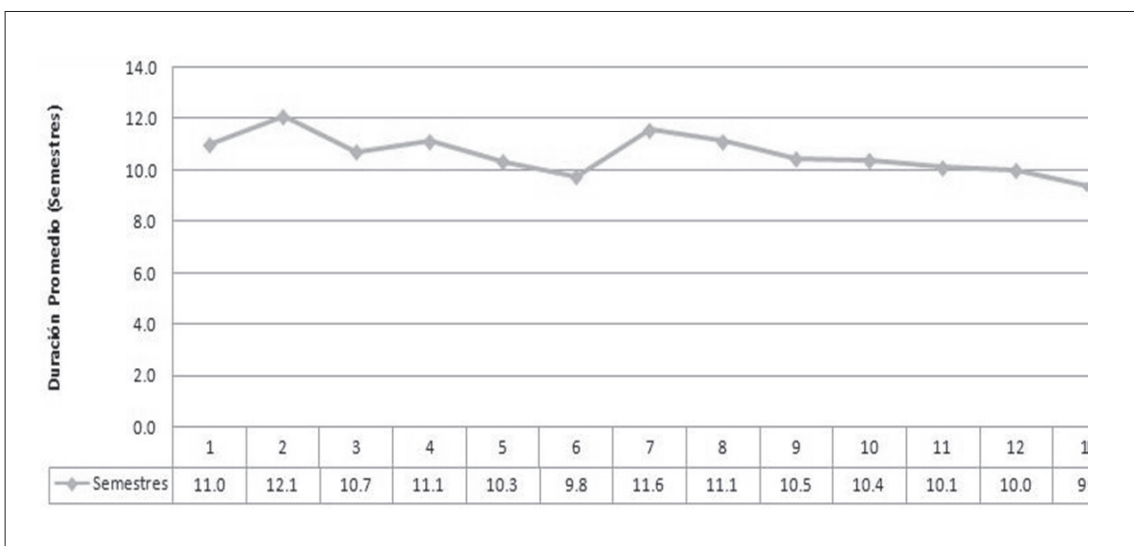
Cada una de las constantes está relacionada con la trayectoria académica de los estudiantes. La constante deserción se refiere a todos los estudiantes que abandonaron sus estudios, pero las razones pueden ser las constantes de Baja por reglamento, baja voluntaria o dejó de inscribirse sin notificar a la administración de la institución, que sería el caso de la constante No inscritos. La constante Egreso se refiere a los estudiantes que finalizaron sus estudios de cada generación, la constante inscritos son el total de estudiantes que ingresaron en cada generación y la constante titulados, son los estudiantes que después de egresar logran obtener su título profesional.

3.2.1.1 Análisis de indicadores de trayectoria académica. Plan de estudios

Uno de los indicadores de trayectoria académica es la duración promedio de los estudios, que es el tiempo promedio que los estudiantes utilizan para cursar todas las materias del programa académico, para el caso del plan de estudios 1 debería ser de nueve semestres y cada generación está tardando en promedio 10.6 semestres, esto implica que existe un rezago en la mayoría de los estudiantes. En la Figura 4.3 se aprecia que la generación 2, tardó en promedio más de 12 semestres, es la que tuvo la mayor duración promedio de los estudios. Además se observa una tendencia a la baja en la duración promedio y se mantiene en 10 semestres.

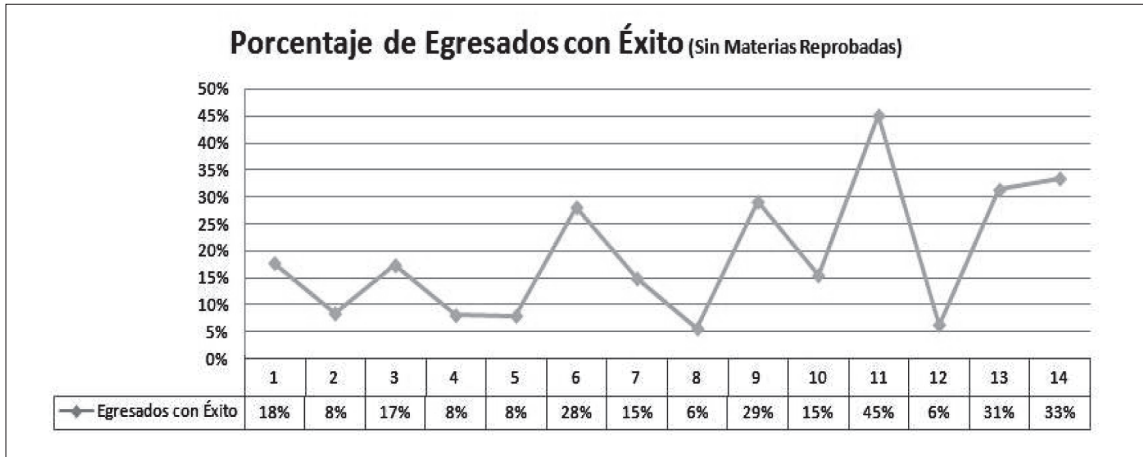
La *tasa de rendimiento*, otro indicador de la trayectoria académica, se refiere a la proporción de estudiantes que concluyen con éxito sus estudios; es decir, son aquellos estudiantes que terminaron sin haber reprobado ninguna materia durante su permanencia en el plan de estudios.

Figura 9. Duración promedio de los estudios de la carrera de ingeniería en computación (plan de estudios 1).



Fuente: elaboración propia.

Figura 10. Porcentaje de egresados con éxito (Sin materias reprobadas) de la carrera en Ingeniería en Computación Plan de estudios 1



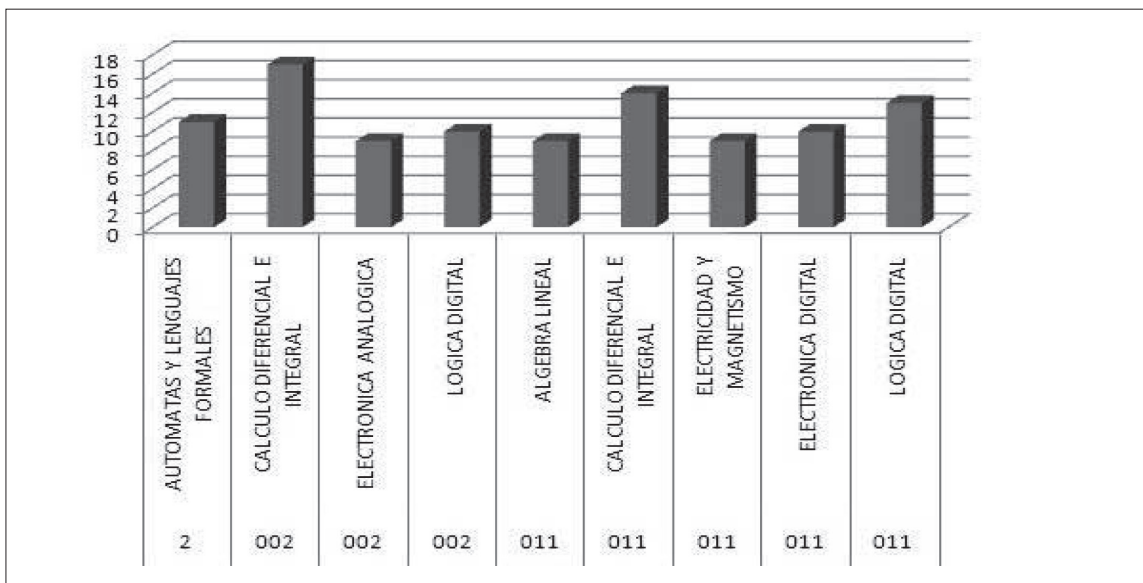
Fuente: elaboración propia.

En la figura 9 se observa que la generación 11 es la que tiene el mayor porcentaje de estudiantes (45 %) que concluyen con éxito sus estudios. Se observa que este indicador es muy irregular, ya que las generaciones con mayor porcentaje de estudiantes con alta tasa de rendimiento son las generaciones

impares que son las que tuvieron ingreso en el mes de Julio, pero no es una regla ya que la generación 6 y 14 también tienen alta tasa de rendimiento y tienen ingreso en enero.

El *índice de reprobación*, es otro indicador de la trayectoria académica, en este indicador

Figura 11. Materias con mayor índice de reprobación de la carrera en Ingeniería en Computación Plan de estudios 1 generación 2000-2 y 2001-1



Fuente: elaboración propia con base en el SIIA.

se encontró que en promedio el 34 % de los estudiantes reprueba al menos una materia por periodo de una población promedio de 380 estudiantes.

La pregunta posterior a este análisis fue encontrar cuales eran las materias con mayor índice de reprobación, este análisis se realizó por generación por lo que se pueden observar materias que aparecen más de una vez, es importante mencionar que se consideraron aquellas materias con un número igual o mayor a 9 estudiantes reprobados.

En la Figura 4.5 se presentan de la generación 2000-2 y 2001-1. La materia de Calculo Diferencial e Integral en ambas generaciones es una de las más reprobadas. En la Figura 4.6 se demuestra que también para las generaciones 2001-2 y 2002-1 aparece Cálculo Diferencial e Integral como materia con un índice importante de reprobación. También materias como Algebra lineal, Autómatas y Lenguajes Formales, Electricidad y Magnetismo, son materias que tienen un índice importante de reprobación y se presentan en más de una generación. Por lo que las áreas del conocimiento que requieren de mayor atención son las de *Matemáticas y Arquitectura de las Computadoras*.

4. Conclusiones

Al presente trabajo le falta desarrollar la fase de pronósticos que también se basa en un modelo matemático que a partir de la minería de datos presentada permitirá dar soporte a la toma de decisiones.

La técnica propuesta es mucho más barata que implementar que alguna de las herramientas de mercado conocidas como herramientas de Inteligencia de Negocios, ya que además de tener altos costos las licencias, la implementación, adecuación a la empresa y el mantenimiento hace realmente complicado el que empresas pequeñas y medianas puedan aprovechar las ventajas del análisis de datos.

Se presentó en esta investigación el primer dominio de la visión del conocimiento que es el estado presente de la IES. Analizando desde las características de sus aspirantes como la trayectoria escolar, específicamente de la carrera de Ingeniería en Computación. Donde se destaca que el alto índice de reprobación que afecta la deserción y rezago de los estudiantes, por consecuencia existe baja eficiencia terminal para el plan de estudios 1.

Podemos concluir que el análisis de la información es el arma estratégica para lograr la competitividad.

5. Referencias

- [1] Poder Ejecutivo Federal, Programa Nacional de Educación 2001-2006, México. 2001.
- [2] Poder Ejecutivo Federal, Programa Nacional de Educación 2007-2012, México. 2007a.
- [3] ANUIES, Consolidación y avance de la Educación Superior en México: Elementos de Diagnóstico y propuestas. Asociación Nacional de Universidades e Instituciones de Educación Superior. 2007.
- [4] N. Juntarung y P. Ussahawanitchakit, Knowledge management capability, market intelligence, and performance: an empirical investigation of electronic businesses in Thailand. *International Journal of business research*, vol. 8, núm. 3. 2008.
- [5] G. Gan, C. Ma And J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM. 2007.
- [6] E. Turban E, *Decision Support Systems and Intelligent Systems 6th edition* (2005). Prentice Hall, Upper Saddle River, NJ. 2005.
- [7] T.F. González, *On the computational complexity of clustering and related problems*. Springer Berlin: Heidelberg, 2006.

- [8] CIEES, Comités Interinstitucionales para la Evaluación de la Educación Superior, (2005). Metodología general CIEES para la evaluación de programas educativos. Documento de trabajo D.R. 2005 CIEES.
- [9] Poder Ejecutivo Federal, Plan Nacional de Desarrollo 2007-2012, México. Oficina de la Presidencia de la República Mexicana. 2007b.
- [10] B. De Ville, *Microsoft Data Mining: Integrated Business Intelligence for e-Commerce*. Woburn, MA: Digital Press. 2001.
- [11] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press. 1996.
- [12] J. A. Hartigan, *Clustering Algorithms*. New York: John Wiley & Sons. 1975.
- [13] M. Holsheimer and A. Siebes, *Data mining: The search for knowledge in databases*. Amsterdam, The Netherlands: In CWI Technical Report CS-R9406. 1994.
- [14] A. K. Jain, M. N. Murty and P. J. Flynn, Data Clustering, a review. *ACM Computing Surveys*, vol. 31, núm. 3. 1999.
- [15] *Presidencia de la República*, Plan Nacional de Desarrollo del período 2001-2003. [En línea] disponible en <http://pnd.fox.presidencia.gob.mx>. 2001.
- [16] S. Salvado and P. Chan, *Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms*. 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004), p. 576. Disponible en la base de datos IEEE.
- [17] R. Sprague and E. Carlson, *Building Effective Decision Support Systems*. PrenticeHall. 1982.
- [18] T. Raymond and H. Jiawei H. *Efficient and Effective Clustering Methods for Spatial Data Mining. Proceedings of the 20th International Conference on Very Large Data Bases*, (pp. 144-155). Morgan Kaufmann Publishers Inc. 1994.
- [19] C. C. Yu, A Web-based Consumer-Oriented Intelligent Decision Support System for Personalized E-Services. Sixth International Conference on Electronic Commerce (ICEC'04), (pp. 429-437). Obtenido el 19 de febrero de 2010 de la base de datos ACM. 2004.
- [20] C. C. Book: Enabling Knowledge Creation (May 2000) by Georg Von Krogh, Kazuo Ichijo, Ikujiro Nonaka www.gurteen.com/gurteen/gurteen.nsf/id/X00021C4A/.
- [21] C. C. MEMORIAS IBERGECYT 2008, la habana, noviembre, 2008. 14. Goñi, I. Artiles, S. y Nieves Sistema Automatizado para el
- [22] C. C. NOHRA LEON RODRIGUEZ, "Mecanismos para la regulación ambiental" En: Colombia2006. ed:Prisma Asociados Ltda ISBN: 9587016718 v. 1 pags. 119