

Sistemas de soporte a las decisiones como herramienta de ventaja competitiva

M. S. I. ELIETH VELÁZQUEZ CHÁVEZ¹

L.I. GABRIELA XICOTÉNCATL RAMÍREZ²

M.I.S.D. CARLOS OLMOS TREJO³

Resumen

En las empresas, instituciones y organizaciones ha comenzado a ser importante el análisis de la información que se ha venido recopilando durante el tiempo que los sistemas de información han existido. Dicha actividad implica tiempo para realizar análisis con la suficiente profundidad sin garantizar que tome la información encontrada nos lleve a tomar la mejor decisión. Los Sistemas de Soporte a las Decisiones proveen a la organización de los elementos que apoyaran a los tomadores de decisiones organizacionales en la toma de decisión respondiendo con efectividad, precisión y rapidez a los nuevos modelos de negocios y comerciales de competencia.

Palabras Clave: Sistemas de Soporte a las Decisiones, Minería de datos, Clustering).

Abstract

In enterprises, institutions and organizations have begun to be important analysis of the information that has been compiled during the time that the information systems have existed. Such activity involves time for analysis with sufficient depth without ensuring that take the information found lead us to make the best decision. Support Systems to the Decisions provide to the organization of the elements that support to policyholder's organizational decisions in the decision-making responding effectively, precision and speed to the new business models and commercial competition.

Key words: decision support systems, clustering, data mining

1. Sistemas de soporte a las decisiones (SSD) como ventaja competitiva

El tomar una decisión implica seleccionar una alternativa de un conjunto de ellas. La teoría administrativa de la toma de decisiones describe el proceso en tres pasos: 1) enlistar todas las estrategias posibles alternativas, 2) determinar todas las consecuencias que tendría el tomar cada

¹ Profesor investigador de la Universidad Autónoma de Querétaro. Facultad de Informática, Campus Juriquilla.

² Universidad Autónoma de Querétaro. Facultad de Informática, Campus Juriquilla.

³ Universidad Autónoma de Querétaro. Facultad de Informática, Campus Juriquilla

una de las alternativas posibles, 3) evaluar comparativamente estos grupos de consecuencias. La actividad de tomar decisiones es un proceso que actualmente recae en los tomadores de decisiones apoyados por los expertos y analistas de decisiones. Los estrategas son vulnerables de retirarse con la gama de experiencia y conocimiento ganado en los años de colaboración, sin embargo, la empresa como organismo debe buscar mecanismos que le permitan dejar esa memoria grabada en la organización para solventar en el futuro situaciones similares que pudieran presentarse.

Un SSD es un sistema interactivo basado en computadoras, que asiste a los tomadores de decisiones, utilizando datos y modelos. El objetivo es resolver problemas semiestructurados y no estructurados dentro de una organización. El concepto no es necesariamente nuevo, de hecho se habla de ello desde 1960 especialmente sobre las áreas de finanzas y operaciones. Sin embargo, es hasta 1970 que Scott Morton publica el primer concepto bajo el término de Sistemas de Decisión Administrativa. Posteriormente Keen y Scott Morton (1978) citan que “Los sistemas de soporte a la decisión unen los recursos intelectuales del individuo con las capacidades de la computadora para mejorar la calidad de las decisiones” (Turban, 2005). Los SSD se apoyan de técnicas avanzadas de análisis de datos para apoyar la toma de decisiones con información analizada y que permite generar modelos predictivos o bien encontrar patrones en los datos.

Entre los beneficios de los SSD que dan ventajas competitivas a una empresa u organización están: mejora en el desempeño para análisis de datos, permiten hacer predicciones del comportamiento de los datos, permiten análisis multidimensional. Un SSD ideal debería permitir la administración del conocimiento, el modelado gráfico, debería ser fácil de utilizar, entre otras características que permitirían lograr eficiencia y efectividad en la toma de decisiones.

2. La minería de datos y la inteligencia de negocios

A lo largo de la historia se ha manifestado interés por la recolección y análisis de datos. El calendario maya, los dibujos de la cueva de Árdales, la escritura cuneiforme de las tablillas babilónicas, y la piedra Rosetta son ejemplos de la necesidad humana de trascendencia a fin de dar sentido a la vida en el mundo a través del registro de datos mediante el uso de símbolos (Barry, 2001). Se puede decir también que los primeros indicios de búsqueda de información en los datos han sucedido al estudiar los símbolos de las antiguas civilizaciones con el fin de descifrarlos. Así ha sido posible darnos una idea de la manera en que vivían, cazaban, convivían, y al final poder aprender de las culturas pasadas para comprender un poco nuestro presente.

Actualmente este mismo proceso de registrar y analizar datos a fin de encontrar información relevante que sea la base para que las personas u organizaciones tomen mejores decisiones se ha dado en llamar *minería de datos* (DM, *Data Mining*). La DM tiene sus raíces entonces en una de las más antiguas actividades humanas: el deseo de dejar plasmada la experiencia en alguna forma

(numérica o simbólica), para describirla y preservarla. Tan pronto como describimos y preservamos la experiencia a través de los datos, comenzamos el proceso inevitable de interpretarlos, utilizando alguna técnica de DM.

En la época actual, después de más de 60 años de cómputo comienza a ser importante el análisis de la información que se ha venido generando. Ahora en México por ejemplo, podemos recuperar información y obtener actas de nacimiento de manera ágil, procedimientos que antes tomaban varios días, semanas y en ocasiones hasta meses. La iniciativa de e-gobierno que es parte del Plan Nacional de Desarrollo del período 2001-2003 (Presidencia de la República, 2001) constituye un momento importante que ha motivado la digitalización de todos los libros de registro de nacimiento que se tenían almacenados en los archivos del Registro Civil. Este suceso lleva entonces a un paso importante donde se comienza a analizar la información que se tiene, y así empezar a entender fenómenos sociales, naturales, planear mejor nuestros proyectos de ciudades, analizar riesgos, entre otras muchas oportunidades que puede dar el estudiar a fondo la información almacenada.

La DM se puede definir como el proceso que facilita la extracción no trivial de información que reside de manera implícita en los datos. Dicha información es previamente desconocida y podrá resultar útil para algún otro proceso. En otras palabras, la DM prepara, sondea y explora los datos para obtener la información oculta en ellos. Es importante señalar que DM no es una técnica, sino que ese nombre engloba a todo un conjunto de técnicas. El proceso de DM está fuertemente ligado a la supervisión de procesos industriales, ya que resulta muy útil cuando los datos almacenados en las bases de datos son procesados.

Los fundamentos de la DM se encuentran en las áreas de las ciencias computacionales y el análisis estadístico. Mediante los modelos basados en algoritmos de DM es posible generar la solución a problemas de predicción, clasificación y segmentación. El algoritmo de DM es el mecanismo que crea un modelo de minería de datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos y luego busca patrones y tendencias específicos. El algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de DM. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas. El DM crea un algoritmo que puede tomar diversas formas, incluyendo:

1. Un conjunto de reglas que describen cómo se agrupan los productos en una transacción.
2. Un árbol de decisión que predice si un cliente determinado comprará un producto.
3. Un modelo matemático que predice las ventas.

4. Un conjunto de clusters que describen cómo se relacionan los casos de un conjunto de datos.

Algunos de los tipos de algoritmos de DM (Figura 1) se pueden clasificar como: *algoritmos de clasificación*, que predicen una o más variables discretas, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de clasificación es el de árboles de decisión; *algoritmos de regresión*, que predicen una o más variables continuas, como las pérdidas o los beneficios, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de regresión es el de serie temporal; *algoritmos de segmentación*, que dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares. Un ejemplo de algoritmo de segmentación es el de clustering; y *algoritmos de asociación*, que buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación, que pueden utilizarse en un análisis de compras en sistemas del tipo comercio electrónico.

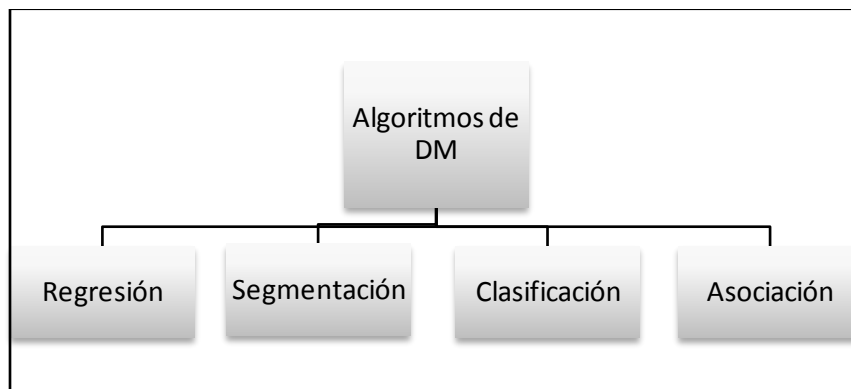


Figura 0. Algoritmos de minería de datos. Fuente: Elaboración propia con base en Barry (pp. 1-54, 151, 2001).

El concepto de Inteligencia de Negocios (*Business Intelligence, BI*) involucra conceptos de organización de datos en varias dimensiones de análisis a efecto de poder visualizar los datos. A manera de ejemplo, consideremos una base de datos con información de las ventas que se están teniendo en la empresa. Es posible organizar los datos a fin de poder visualizarlos desde el punto de vista de ventas por región, pero quizás sería también interesante analizarla desde la perspectiva de los productos vendidos. Entre más dimensiones se incluyan en el análisis es mejor la visualización de cómo se comporta el negocio, y así fundamentar la toma de decisiones. Es de llamar la atención que alrededor del término DM diferentes autores presentan también los términos: *Data Warehousing, BI, DSS*. La bibliografía varía, desde la que está dirigida para los empresarios, tratando de convencerlos acerca de la importancia de implementar dichas técnicas en sus empresas, hasta la que está dirigida a los arquitectos de software.

Actualmente en el mercado existen herramientas comerciales y también herramientas de software libre y arquitectura abierta que incluyen funciones que efectúan procesos de minería de datos.

Entre las herramientas comerciales se encuentran las propias de manejadores de bases de datos como Oracle con su *OBI Suite* e IBM *Cognos Business Intelligence and Financial Performance Management*. En cuanto a software libre se cuenta por ejemplo, con *Pentaho BI Suite* entre otros. A menudo, encontramos en el mercado de software de sistemas que las empresas venden servicios de DW, BI y SSD como la solución a cualquier problema empresarial.

La minería de datos, y en particular datos en bases de datos espaciales, revelan patrones o asociaciones que usualmente eran desconocidas. Este tipo de bases de datos con cantidades grandes de información (terabytes) son por ejemplo obtenidas de imágenes satelitales, equipos médicos, entre otros (Ng & Han, 1994). A estas técnicas se les conoce como Técnicas de Descubrimiento de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés que significan *Knowledge Discovery in Databases*) (Holsheimer & Siebes., 1994). El descubrimiento de conocimiento en bases de datos es el proceso de identificación de patrones válidos, potencialmente útiles y comprensibles, en los datos (Fayyad 1996). El objetivo es la extracción de conocimiento de los datos especialmente cuando las bases de datos son de gran tamaño. KDD es un proceso multidisciplinario que involucra las áreas de representación del conocimiento, aprendizaje, bases de datos, estadística, sistemas expertos y representación gráfica, entre otros.

Uno de los primeros retos que hace interesante el problema y justifica el estudio de técnicas de reconocimiento de patrones en datos en SSD, es el manejo de grandes cantidades de datos. Para ello, se utilizan técnicas de agrupamiento de datos (*clustering*), las cuales se han aplicado en los últimos 30 años a muchas áreas. En la medicina, por ejemplo, en la clasificación de enfermedades; en la química en el agrupamiento de compuestos equivalentes; en los estudios sociales para la determinación de grupos o redes sociales con base en datos estadísticos, entre otros.

3. Técnicas de agrupamiento (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 clusters o grupos de objetos, de modo tal que cada par de objetos en un cluster son similares y cada par de objetos, cada uno en diferente cluster son disimilares. El *clustering* de datos es 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.* (Gan, Chaoqun, & Jianhong, 2007)

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. (Gan, Chaoqun, & Jianhong, 2007). En la figura 1:5 se presentan las tareas asociadas a los dos tipos de DM.

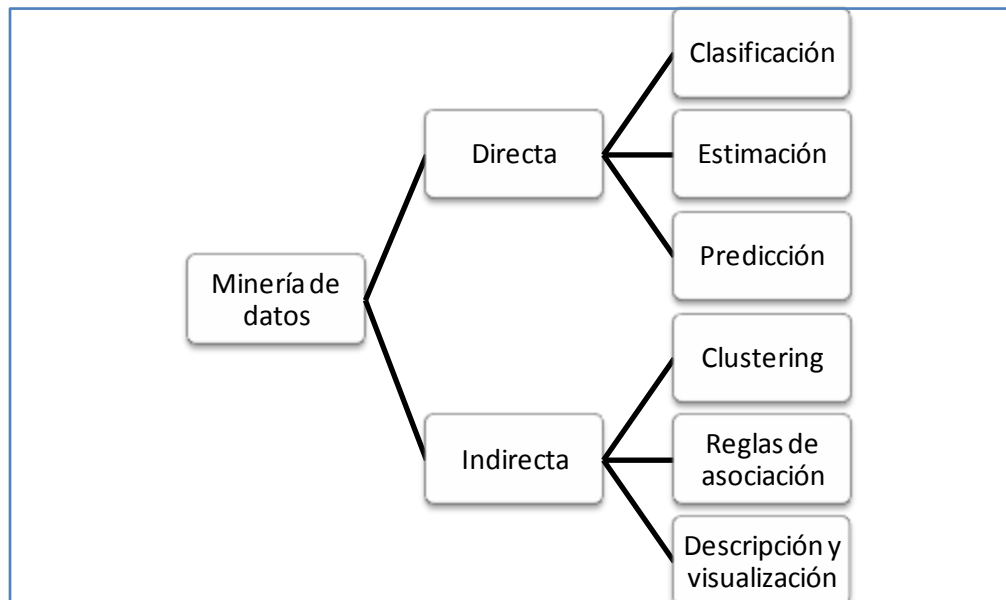


Figura 2 . Tipos de minería de datos. Fuente: Holsheimer y Siebes (1994)

La ventaja principal al usar las técnicas de agrupamiento de datos es la posibilidad de descubrir estructuras o categorías bien definidas y relevantes directamente en los datos, sin tener que contar con el conocimiento total del contexto. Ya que el concepto de agrupamiento se define de acuerdo al criterio de similitud y disimilitud, y a las diferentes formas (geométricas) de los clusters o patrones que tales criterios inducen, se han desarrollado una gran cantidad de algoritmos. Dichos algoritmos, en términos generales, se pueden clasificar de acuerdo a los tipos de clusters en: esféricos, lineales e irregulares. A continuación se presentan ejemplos de diferentes tipos de clusters encontrados utilizando un método jerárquico de identificación de clusters conocido como método L (Salvador & Chan). En la Figura 3 se presentan 1) clusters del tipo esférico (4000 pts), 2) nueve clusters cuadrados conectados por sus esquinas (9,000 pts), 3) diez clusters esféricos (5,200 pts), 4) Diez clusters esféricos separados y de diferentes tamaños (5,000 pts), 5) y 6) muestran clusters con diferentes formas (~9,100 pts) y. (~7,600 pts) respectivamente.

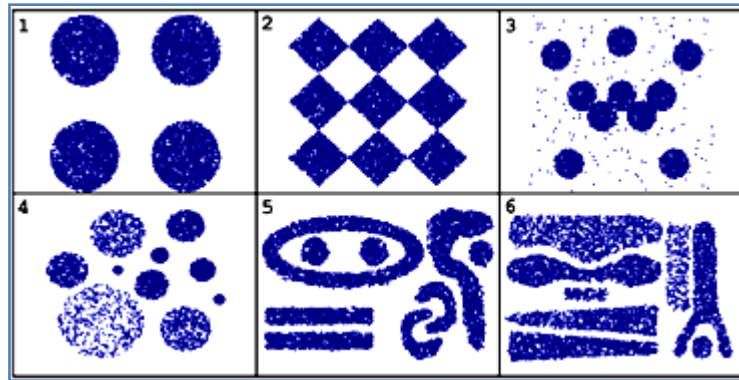


Figura 3. Diferentes tipos de clusters. Fuente: Salvador & Chan (1994).

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 clusters, de acuerdo a algún criterio de similaridad. Un *clúster* se define como un conjunto de objetos similares. El criterio de similaridad se establece como la métrica Euclidiana consistente en la función $d: \mathcal{R}^n \times \mathcal{R}^n \rightarrow \mathcal{R}$, que asigna a cualquier par de vectores del espacio Euclidiano n -dimensional $\mathbf{x}=(x_1, x_2, \dots, x_n)$ y $\mathbf{y}=(y_1, y_2, \dots, y_n)$, el número $d = \sqrt{\sum_{1 \leq i \leq n} (x_i - y_i)^2}$, produciendo así la distancia estándar en el espacio \mathcal{R}^n . Sin embargo, el problema de agrupar un conjunto de n vectores en k clusters bajo funciones objetivo es NP-hard, aún cuando los puntos a ser agrupados se restringen al espacio euclidiano bidimensional (Gonzalez, 2006). Así que su complejidad computacional lo hace un problema equivalente, computacionalmente hablando, a aquel bien conocido problema llamado el problema de máximo-cut, el cual es también NP-hard o problema intratable.

Esto nos lleva a concentrarnos en los algoritmos de aproximación eficientes, aunque bajo ciertas condiciones existen heurísticas (formas de trabajo y que apoyan la realización consciente de actividades mentales exigentes) basadas en algoritmos eficientes. Tales algoritmos para resolver el problema de agrupamiento se pueden clasificar de acuerdo a sus métodos representativos como algoritmos basados en: árboles extendidos de costo mínimo, maximum-cut, k-medias (Hartigan, 1975), entre otros.

En este trabajo se aborda un caso de estudio que consiste en analizar la información de una base de datos de 5500 aspirantes a ingresar, en el período de 2003 a 2008, a las diferentes carreras ofrecidas por una Institución y que incluye: fecha de nacimiento del aspirante, periodo en el cual hace solicitud para ingresar, escuela de procedencia, resultado oficial obtenido o puntuación, estado de nacimiento, sexo, fecha de nacimiento, si fue o no aceptado y carrera a la que hizo solicitud. A partir de los resultados del análisis de información de los alumnos aspirantes a estudiar en la Institución analizada, el departamento de Planeación y Gestión Académica podrá

tomar mejores decisiones sustentadas en los resultados obtenidos del análisis de la información histórica.

Los algoritmos de agrupamiento se ajustan perfectamente a este tipo de proceso de DM, así como al proceso de agrupamiento (clustering) de datos como una actividad exploratoria. Resulta interesante demostrar que una solución a una instancia del problema de minería de datos, es también una solución al problema de agrupamiento; tanto para una solución factible como óptima (Jain 2000). Este tipo de esquemas de acoplamiento son pasos críticos para lograr las correspondencias semánticas de los atributos a través de fuentes heterogéneas de datos (Yu, 2004).

4. Metodología

Concretamente el método abordara los siguientes puntos relevantes: modelación de la información de acuerdo a los datos provenientes de las bases de datos; análisis de los resultados; y elaboración de conclusiones. Es necesario normalizar la base de datos a fin prepararla para la minería. La base de datos utilizada en el caso de estudio que se presentan en este trabajo, consiste en una Tabla que incluye 19 registros. De esta Tabla se seleccionaron aquellos parámetros que potencialmente representan un alto factor de correlación. Los datos seleccionados son los indicados en la Tabla 1

Tabla 0

Selección de atributos de la base de datos para los casos de estudio

Atributo	Descripción
SEXO	Femenino o masculino.
CVECARR	Clave correspondiente a la carrera que aspira el alumno.
RESUOFICIA	Resultado obtenido en el examen de admisión.
CVEESCORI	Clave de la escuela de procedencia del aspirante.
EDONAC	Estado de nacimiento del aspirante
FECNAC	Fecha de nacimiento del aspirante.

La cantidad de datos que se está manipulando es de un poco más de 5000 registros. De los cuales se estarán utilizando con diversos filtros para así, implementar el análisis de la información. Con filtros significan que durante el análisis de la información se detectó la necesidad de normalizar los datos. En la siguiente sección se presentan dichas normalizaciones.

5. Caso de estudio

Mathematica 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 y con ello dando competitividad a la Institución.

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 educativa, correspondiente a los periodos 2003-2008. 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 5,000 registros.

Short[lista]

```
{{22.,29,177.25},{22.,5,0.},{22.,94,98.},{9.,509,75.5},{11.,511,95.},{11.,511,95.},{22.,94,135.},{22.,94,135.},{22.,46,76.75},{22.,27,92.25},{22.,88,129.},{22.,88,129.},{22.,25,117.25},{11.,411,83.},{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 clusters como lo ilustra la Figura 4. 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.

6. Resultados

Utilizando un filtro para analizar únicamente la información de los alumnos aceptados en estos años obtenemos los resultados que se muestran en la Figura 4. 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 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. Se aplica la función de Mathematica especializada aplicar la **c1=findClusters**, la cual permite encontrar los clusters, posteriormente se utiliza un modelo gráfico que nos permita visualizar los clusters. **ListPointPlot3D[c1]**.

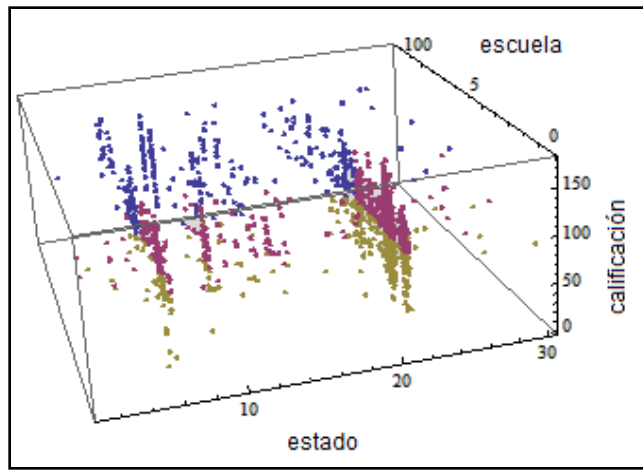


Figura 4. Gráfica con estado, calificación y escuela de procedencia con k = 3

Sin embargo, se podría imponer un número determinado de clusters k para la separación de los datos de acuerdo a los requerimientos del usuario. Las Figuras 4 y 5 muestra los casos para k=3 y k=4, respectivamente.

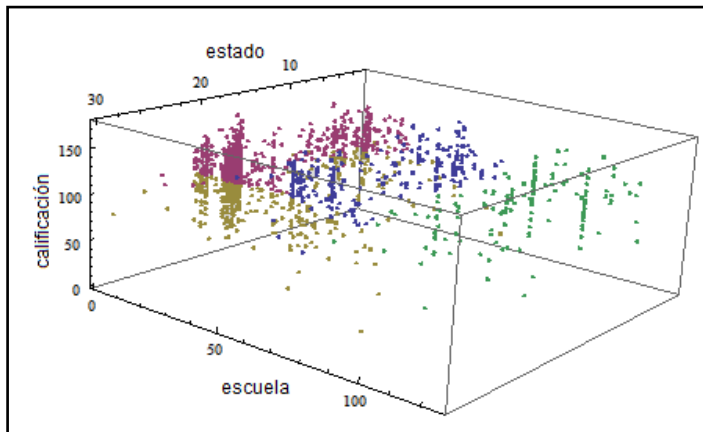


Figura 5. Gráfica con estado, calificación y escuela de procedencia con k = 4

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. Quedando una parte del vector de información de la siguiente manera.

```
{ {85,135.25}, {84,150.75}, {85,117.25}, {24,144.25}, {24,99.5}, {85,149.75}, {24,146.25}, {85,46.75}, {85,102.}, {23,141.75}, {25,137.75}, {17,108.5}, {20,105.75}, {96,141.}, {85,117.}, {85,115.75}, {85,111.75}, {85,118.25}, {113,117.75}, {39,112.25}, {20,22.25}, {85,75.75}, {23,137.5}, {16,108.}, {33,158.25}, {84,148.}, {17,105.75}, {24,95.5}, {93,153.25}, {25,141.5}, {103,51.5}, {110,107.}, {24,108.75}, {50,90.5}, {25,74.25}, {85,116.25}, {93,105.}, {23,100.5}, {37,101.5}, {23,139.}, {85,114.25}, {24,85.}, {92,106.75}, {85,108.5}, {20,108.5}, {98,101.}, {71,91.75}, {33,106.}, {24,132.25}, {123,74.25}, {93,122.5}, {47,
```

71.25},{17,107.5},{85,122.25},{22,122.25},{22,118.25},{24,128.25},{25,99.},{22,128.25},{20,120.25},{96,118.75},{85,119.25},{20,95.5}

La Figura 6 muestra los resultados sin filtrar a los alumnos aceptados y en la Figura 7 se presenta el resultado con filtro de alumnos aceptados, además de normalizadas las escuelas de procedencia.

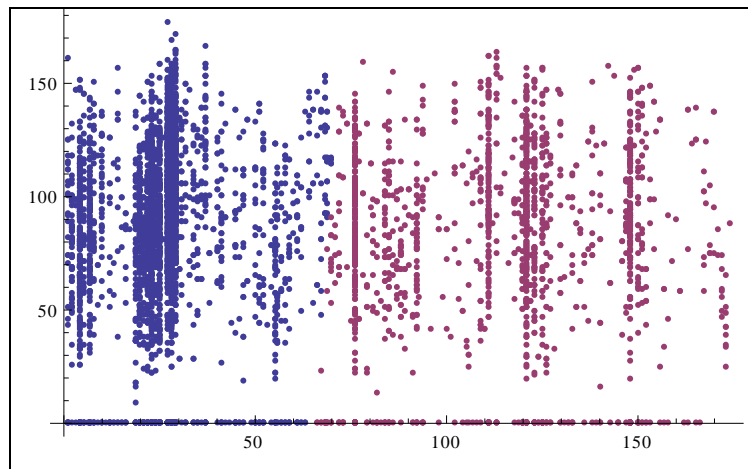


Figura 6. Gráfica todos los aspirantes

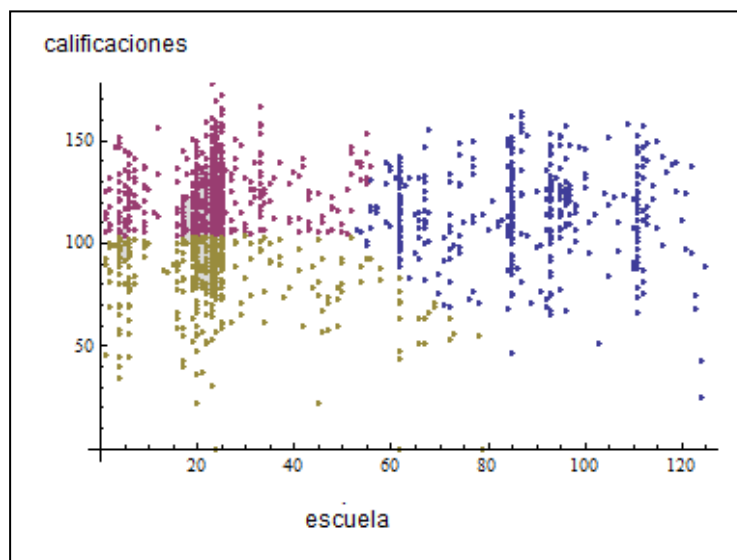


Figura 7. Gráfica aspirantes aceptados por escuela y calificación, k=3

Se notan homogéneos los resultados es decir, no hay un cluster 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. Quedando el vector segmentado:

{7,135.25},{6,150.75},{21,117.25},{6,144.25},{9,99.5},{6,149.75},{21,146.25},{21,46.75},{21,102.},{18,141.75},{4,137.75},{21,108.5},{6,105.75},{6,141.},{21,117.},{21,115.75},{21,111.75},{6,118.25},{6,117.75},{21,112.25},{6,22.25},{21,75.75},{21,137.5},{6,108.},{21,158.25},{17,148.},{21,105.75},{21,95.5},{6,153.25},{6,141.5},{25,51.5},{6,107.},{21,108.75},{21,90.5},{21,74.25},{21,116.25},{9,105.},{21,100.5},{21,101.5},{20,139.},{21,114.25},{6,85.},{9,106.75},{6,108.5},{21,108.5},{21,101.},{25,91.75},{21,106.},{21,132.25},{6,74.25},{21,122.5},{21,71.25},{21,107.5},{6,122.25},{21,122.25},{21,118.25},{21,128.25},{21,99.},{21,128.25},{21,120.25},{6,118.75},{6,119.25},{21,95.5},{3,102.},{21,47.75},{13,127.25},{21,129.75},{19,146.75},{21,90.},{9,124.75},{11,99.75},{21,111.25},{13,106.},{15,118.25},{6,123.5},{21,138.75},{21,102.},{21,96.75},{21,96.5},{21,100}

En la Figura 8 el estado 22 que corresponde a Querétaro es el que tiene mayor población se observa la variedad de puntuaciones obtenidas.

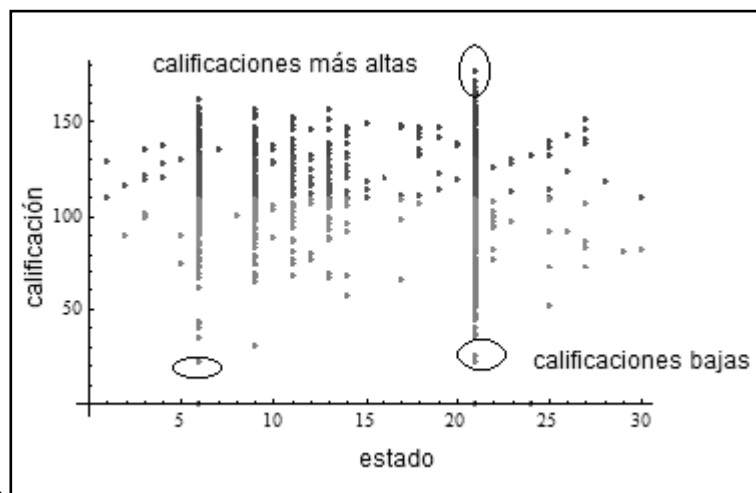


Figura 8 Calificaciones más bajas y más altas $k=4$

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.

7. Conclusiones

Una vez que los modelos de DM se tienen listos, se puede llevar a cabo diferentes tareas. Algunas tareas comunes que se pueden realizar: son el modelo para crear predicciones que puedan utilizarse para tomar decisiones, incrustar la funcionalidad de minería de datos directamente en una aplicación y crear un informe que permita a los usuarios realizar consultas directamente en un modelo de minería de datos existente. La actualización del modelo forma parte de la estrategia de implementación. A medida que la organización recibe más datos, debe volver a procesar los modelos para mejorar así su eficacia.

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.

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

Referencias

- De Ville, B. (2001). *Microsoft Data Mining: Integrated Business Intelligence for e-Commerce*. Woburn, MA: Digital Press.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.
- Gan, G., Ma. C., Wu J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM.
- Gonzalez, T. F. (2006). *On the computational complexity of clustering and related problems*. Springer Berlin / Heidelberg, 2006.
- Hartigan, J. A. (1975). *Clustering Algorithms*. New York: John Wiley & Sons.

- Holsheimer, M., Siebes, A. (1994). *Data mining: The search for knowledge in databases*. Amsterdam, The Netherlands: In CWI Technical Report CS-R9406
- Jain, A. K., Murty, M.N., Flynn, P. J. (1999). Data Clustering, a review. *ACM Computing Surveys* 31(3).
- Presidencia de la República (2001). Plan Nacional de Desarrollo del período 2001-2003. Obtenido el 18 de septiembre del 2009 desde <http://pnd.fox.presidencia.gob.mx>.
- Salvador, S., Chan, P. *Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms*. 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004), p. 576. Obtenido el 19 de febrero de 2010 de la base de datos IEEE.
- Sprague, R., Carlson, E. (1982). *Building Effective Decision Support Systems*. PrenticeHall.
- Raymond, T., Jiawei H. (1994). *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.
- Turban Efraim, *Decision Support Systems and Intelligent Systems* 6th edition (2005). Prentice Hall, Upper Saddle River, NJ
- Yu, C. C. (2004). *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.