

## **Análisis del logro académico de estudiantes en el nivel medio superior a través de minería de datos centrada en el usuario**

Gabriel Maldonado<sup>1</sup>, Guillermo Molero-Castillo<sup>2</sup>, José Rojano-Cáceres<sup>3</sup>,  
Alejandro Velázquez-Mena<sup>4</sup>

<sup>1</sup>Universidad Veracruzana, MSICU, Facultad de Estadística e Informática, México

<sup>2</sup>Universidad Veracruzana, CONACYT, México

<sup>3</sup>Universidad Veracruzana, Facultad de Estadística e Informática, México

<sup>4</sup>Universidad Nacional Autónoma de México, Facultad de Ingeniería, México

`gabriel.amhx@gmail.com`, `ggmoleroca@conacyt.mx`,  
`rrojano@uv.mx`, `mena@fi-b.unam.mx`

**Resumen.** Existe la necesidad natural de buscar nuevas formas de analizar y procesar datos de diferentes fuentes. Una de estas formas es mediante una minería de datos centrada en el usuario. Se analizó el logro académico de estudiantes en el nivel medio superior del país a través de un algoritmo de agrupamiento particional. Se observó variados niveles de dominio, destacando Insuficiente y Elemental en más del 70% de la población evaluada, mientras que Bueno y Excelente fueron alcanzados por un número reducido de escuelas, 20%. Esto contrasta una notable diferencia entre los niveles alcanzados por los estudiantes, trayendo como consecuencia que éstos retrasen o detengan sus estudios universitarios debido a que obtienen un certificado sin tener los conocimientos para aprobar los exámenes de ingreso en las universidades.

**Palabras Clave:** DCU, logro académico, minería de datos, PLANEA.

### **Analysis of the Academic Achievement of Students at the High School Level through User-Centered Data Mining**

**Abstract.** There is a natural need to find new ways to analyze and process data from different sources. One of these ways is through a user-centered data mining. The academic achievement of students at the high school level of the country was analyzed through a partitional clustering algorithm. Various

proficiency levels were observed, highlighting Inadequate and Elemental more than 70% of the study population, while Good and Excellent were hit by a small number of schools, 20%. This contrasts a notable difference between the levels achieved by students, bringing as a result they delay or stop their university studies because they receive a certificate without having the knowledge required to pass the entrance exams at universities.

**Keywords.** UCD, academic achievement, data mining, PLANEA.

## **1. Introducción**

En la actualidad, debido al crecimiento de la recolección de datos y la evolución del poder de cómputo, se tiene información almacenada en diferentes fuentes. Esto permite contar con datos históricos útiles para explicar el pasado, entender el presente y predecir situaciones futuras [1]. Por tanto, es cada vez mayor la necesidad de buscar nuevas formas de analizar y procesar las fuentes de datos existentes para obtener información y conocimiento útil. Sin embargo, el volumen de información que alcanzan estas fuentes, es con frecuencia una limitante para el análisis de forma manual, por lo que se han desarrollado tecnologías especializadas que permiten procesar y obtener información de interés con el propósito de servir de apoyo en el proceso de la toma de decisiones [2, 3]. Precisamente, una de estas tecnologías es la minería de datos, la cual es un área de la ciencia de la computación que permite realizar el análisis inteligente de datos, que tiene como propósito resolver dos grandes retos [4]: a) trabajar con conjuntos de datos para extraer y descubrir información de interés, y b) usar técnicas adecuadas para analizar e identificar tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos que ocurren en el entorno y sirvan de ayuda en el proceso de la toma de decisiones.

Para hacer minería de datos en la actualidad se tienen variados procesos que guían la planeación y desarrollo de los proyectos. Sin embargo, estos procesos, a pesar de su amplia variedad, tienen una reducida participación del usuario en cada una de sus etapas, trayendo como consecuencia desarrollos de minería de datos con limitaciones de usabilidad y accesibilidad [5]. Oportunamente, en los últimos años el diseño centrado en el usuario logra un importante posicionamiento en el desarrollo tecnológico a través de métodos que se centran en el diseño y la participación de los usuarios [6]. Algunas de las características del diseño centrado en el usuario son [7]: a) facilidad de aprendizaje, que mide qué tan fácil le resulta al usuario hacer una tarea; b) eficiencia, que mide qué tanto ha aprendido el usuario acerca del funcionamiento básico del software; c) cualidad de ser recordado, que mide con qué facilidad recuerdan los usuarios el uso del sistema; d) errores, que mide con qué facilidad los usuarios cometen errores y con qué facilidad los pueden resolver; y e) satisfacción, que mide qué tan sencillo le resultó al usuario utilizar el sistema. Además, el diseño centrado en el usuario es ampliamente utilizado en distintas aplicaciones, como: diseño de equipos médicos, creación de vehículos, productos de consumo, entre otros. Recientemente también se está proyectando su uso en la minería de datos, logrando con esto resultados con mayor satisfacción en los usuarios [8].

En este trabajo se presenta los resultados de la proyección de una minería de datos centrada en el usuario para analizar el logro académico de los estudiantes del nivel medio superior en el país. Para esto se utilizó como fuente de datos los registros del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA), específicamente datos de escuelas del nivel medio superior, de carácter público, federal y estatal, y planteles particulares reconocidos por la Secretaría de Educación Pública (SEP). PLANEA tiene como propósito conocer en qué medida los estudiantes logran dominar aprendizajes esenciales en Lenguaje y Comunicación (comprensión lectora) y Matemáticas al término de la educación media superior. Los patrones de datos obtenidos son útiles como instrumento de información para padres de familia, estudiantes, maestros, directivos, autoridades educativas y sociedad en general.

## **2. Procesos actuales de minería de datos**

En la actualidad, dada la amplia variedad de técnicas y herramientas existentes para llevar a cabo el análisis de datos, se requiere de procesos específicos para el desarrollo de proyectos de minería de datos [9]. Uno de los procesos utilizados es el KDD (Knowledge Discovery in Databases), que proporciona fases estructuradas de manera secuencial e iterativa. Sin embargo, este proceso no provee una descripción ampliada de las actividades que deben realizarse en cada una de sus fases, dejando la definición de cada una de éstas a criterio del equipo de trabajo [10]. Otro de los procesos es SEMMA (Sample, Explore, Modify, Model, and Assess) que se enfoca en aspectos técnicos del proyecto dejando de lado la fase de análisis y comprensión del negocio [11], teniendo como objetivos [12]: segmentar grupos de clientes, personalizar la gestión de relaciones, identificar clientes rentables e identificar clientes que se van con la competencia. SEMMA al igual que KDD no describen las actividades específicas que se deben realizar en cada una de éstas [10, 13]. En la Tabla 1 se muestra una síntesis de las principales características de éstos y otros procesos descritos.

Otro proceso es CRISP-DM (Cross Industry Standard Process for Data Mining) que es la guía actual más utilizada en el desarrollo de proyectos de minería de datos [14]. Este proceso sirve para guiar el desarrollo de un proyecto de a través de una secuencia dividida en fases [15]. Estas fases siguen un proceso iterativo, que es útil para realizar cambios en etapas anteriores. Sin embargo, no incluye tareas de control y monitoreo del plan de trabajo [11]. Por su parte, Microsoft propone soluciones para usuarios empresariales y no para usuarios finales, propiciando con esto que en ocasiones se obtenga aplicaciones no ergonómicas y con dificultad de uso [16]. Además, está vinculada a herramientas comerciales como SQL Server. Otros de los procesos utilizados es DWEP (Data warehouse engineering process), que basa su funcionamiento en el proceso unificado de desarrollo de software, abarcando desde la recolección de requerimientos hasta la fase de pruebas del producto final. Sin embargo, DWEP al ser un proceso orientado a ingeniería de software no comparte todas las fases que requiere la minería de datos, salvo en el análisis de datos y del problema [17].

Así, a pesar de la variedad de estos procesos, que han tenido que adaptarse a las necesidades de los proyectos y usuarios, éstos presentan algunas debilidades sobre todo en la fase de análisis del problema o del negocio, donde por lo general los requisitos del proyecto y necesidades de los usuarios no son recuperados adecuadamente, trayendo como consecuencia desarrollos de minería de datos con limitaciones de usabilidad y accesibilidad [18]. Estas necesidades se traducen en puntos de vista, preferencias, estrategias y decisiones de los usuarios [5].

**Tabla 1.** Principales características de los procesos de la minería de datos.

	KDD	CRISP-DM	SEMMA	DWEP	Microsoft
<b>Autor</b>	Fayyad <i>et al.</i> (1996)	SPSS, Daimler Chrysler y NCR (1996)	SAS Inc.	Lujan-Morán (2005)	Microsoft
<b>Área de aplicación</b>	Academia / Industria	Academia / Industria	Industria	Industria	Industria
<b>Objetivos</b>	Se centra en la comprensión del dominio de aplicación	Se centra en los objetivos empresariales del proyecto	Se centra en las características técnicas del desarrollo del proceso	Se centra en los flujos de trabajo para la integración de datos	Se centra en los objetivos empresariales del proyecto
<b>Estructura</b>	Fases	Fases y niveles	Fases	Fases y flujos de trabajo	Fases
<b>Número de fases</b>	5	6	5	4	6
<b>Fases</b>	Integración y recopilación de datos Selección, limpieza y transformación Minería de datos Evaluación e interpretación Uso del conocimiento	Comprensión del negocio Comprensión de los datos Preparación de los datos Modelado Evaluación Despliegue	Muestreo Exploración Modificación Modelado Evaluación	Inicio Elaboración Construcción Transición	Definir el problema Preparar los datos Explorar los datos Generar modelos Validar modelos Implementar y actualizar modelos
<b>Herramientas</b>	Libres y comerciales	Libres y comerciales	SAS	Libres y comerciales	Microsoft
<b>Iteración</b>	Si	Si	No	Si	Si

### 3. Minería de datos centrada en el usuario

La participación del usuario no sólo beneficia la obtención de mejores requerimientos, sino también la creación de proyectos de minería de datos personalizados, es decir, hechos a la medida [18, 19]. Además, centrarse en el usuario no significa enfocarse en uno solo, sino se debe tomar en cuenta a todos los involucrados en el proyecto, considerando las edades, capacidades, rasgos, diferencias y otras características de interés [20]. Precisamente, una de las disciplinas encargadas de recopilar las necesidades del usuario e incorporarlas en el producto final es el Diseño Centrado en el Usuario (DCU), que es ampliamente utilizado para el diseño y desarrollo de proyectos tecnológicos [6]. Un enfoque actual del DCU es el Modelo de Proceso de la Ingeniería y de la Accesibilidad (MPIu+a) definido por [21]. Este proceso no sólo incluye la usabilidad, sino también agrega el concepto de accesibilidad, permitiendo con esto el acceso a distintos tipos de usuario. Como pilares de este proceso destacan: a) la ingeniería de software, que sigue un ciclo de desarrollo en cascada (análisis, diseño, desarrollo y transición); b) el prototipado,

utilizado para explorar posibles mejoras acerca del diseño; c) la evaluación, que engloba y categoriza técnicas de evaluación existentes, como pruebas de caja blanca o caja negra; y d) usuario, que es el principal actor de todo el proceso y que interviene en cada una de sus etapas.

Una característica de MPIu+a es la usabilidad, que es la facilidad de uso de una aplicación interactiva [22]. Sin embargo, ésta sólo es una definición operativa, puesto que se logra a través de varias condiciones [7]: a) facilidad de aprendizaje, b) eficiencia, c) cualidad de ser recordado, d) errores, y e) satisfacción.

En consecuencia, existe la natural necesidad de definir un proceso de minería de datos que incluya el diseño centrado en el usuario. Mediante este tipo de procesos se diseñan e implementan aplicaciones personalizadas capaces de realizar el reconocimiento de patrones de datos. En este trabajo, se incluye como base procesos ampliamente utilizados por la comunidad científica: a) CRISP-DM y b) MPIu+a. Ambos aportan fundamentos significativos que de forma conjunta permiten definir etapas y acciones concretas para llevar a cabo proyectos de reconocimiento de patrones a través de la minería de datos centrada en el usuario. Como proyección general en la Fig. 1 se presenta la minería de datos centrada en el usuario, estructurada en cuatro etapas: a) análisis del problema, b) análisis y preparación de datos, c) adquisición de patrones, y d) presentación de patrones. Además, como parte de las etapas intervienen los usuarios, el prototipado y la evaluación.

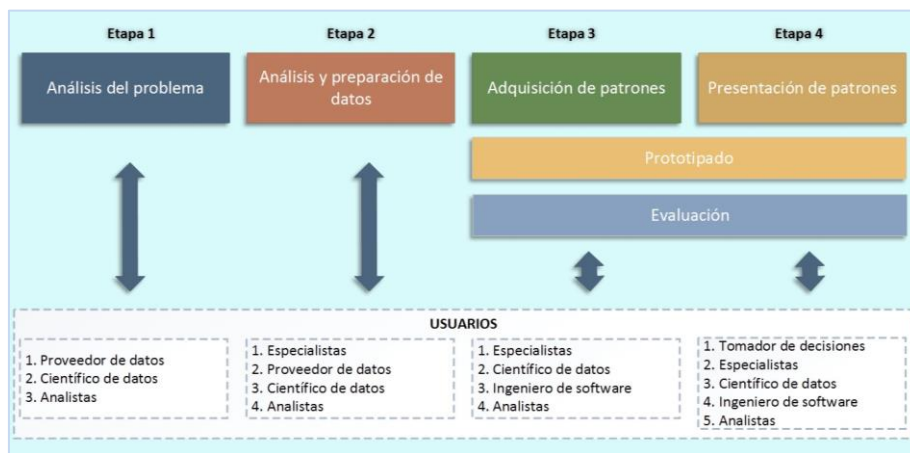


Fig. 1. Proyección general del proceso de minería de datos centrado en el usuario.

En la Fig. 2 se muestra las etapas y tareas generales del proceso. Tiene como característica principal la captura de requerimientos y necesidades de los usuarios para la construcción de interfaces personalizadas. Las primeras tres etapas corresponden al descubrimiento de patrones, mediante las cuales se hace un análisis del problema, análisis y preparación de los datos, así como la extracción de patrones significativos. Esta adquisición de patrones (Etapa 3) junto con la presentación (Etapa 4) representan la columna vertebral del diseño centrado en el usuario, en las cuales se definen

acciones específicas para orientar el desarrollo de aplicaciones personalizadas de minería de datos (software Ad hoc). Entre los tipos de usuarios identificados destacan: el proveedor de información, los especialistas del negocio, el tomador de decisiones, así como el equipo de desarrollo (analistas, científico de datos, ingeniero de software, entre otros). Además, como parte de proceso se incluyen también las fases de prototipado y evaluación, siendo la primera utilizada para crear representaciones del producto final, mientras que la segunda permite evaluar la calidad del producto obtenido.

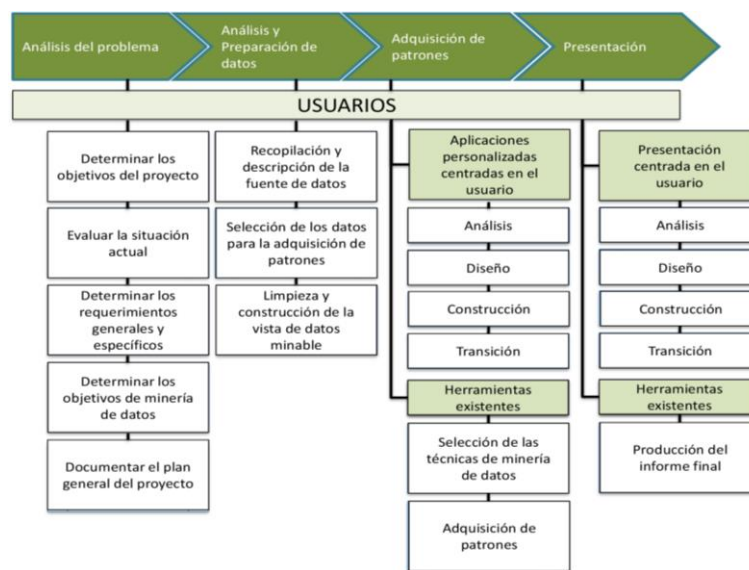


Fig. 2. Etapas y tareas del proceso de minería de datos centrado en el usuario.

Este proceso fue tomado como base para el análisis de grupos asociado al logro académico de estudiantes en el nivel medio superior. La fuente de datos utilizada corresponde al Plan Nacional para la Evaluación de los Aprendizajes (PLANEA).

#### 4. Logro académico en el nivel medio superior

En la actualidad, la educación es uno de los pilares para el desarrollo social y económico de un país. Para obtener resultados satisfactorios se necesita de una educación de calidad; la cual se logra a través de los sistemas educativos que juegan un papel decisivo en el mejoramiento de la calidad educativa [23]. Así, al ser el logro académico un parámetro importante de medición sobre la calidad de la enseñanza, que proporcionan los sistemas educativos, surge el interés de conocer en qué medida los estudiantes alcanzan aprendizajes esenciales en diferentes dominios al término de cada nivel educativo, con el propósito de hacer un diagnóstico del desempeño y de los conocimientos alcanzados por los estudiantes.

Actualmente uno de estos diagnósticos se hace a través del Plan Nacional para la Evaluación de los Aprendizajes en el nivel medio superior (PLANEA-MS) de la Secretaría de Educación Pública [24, 25]. Por tanto, hacer el análisis del logro académico de estudiantes de nivel medio superior sirve de apoyo en la toma de decisiones del ámbito educativo para mejorar la calidad del desempeño académico en los estudiantes [26]. En el caso de la educación del nivel medio superior existen casos de alumnos que al término de sus estudios no obtienen los conocimientos necesarios para aprobar los exámenes de ingreso a las universidades del país, trayendo como consecuencia que retrasen o detengan sus estudios universitarios.

La evaluación que realiza PLANEA-MS está dirigido a los alumnos de toda la República Mexicana que cursan el último ciclo escolar, inscritos a algún plantel educativo, ya sea autónomo, estatal, federal o particular. Las áreas de competencia que evalúa son Lenguaje y Matemáticas. La primera evaluación de PLANEA-MS se realizó en marzo de 2015, participaron más de un millón de alumnos de educación media superior, de un total de 14548 instituciones, entre escuelas públicas, autónomas y privadas. Los logros académicos están comprendidos en cuatro niveles de dominio: a) I, insuficiente; b) II, elemental; c) III, bueno; y d) IV, excelente.

Derivado del análisis de datos de PLANEA, se obtuvo una vista de datos minable. La principal consideración fue determinar cuántas y cuáles son las variables apropiadas para el estudio. En la Tabla 2 se presenta las variables significativas de la vista de datos minable.

**Tabla 2.** Variables que conforman la vista de datos minable.

No.	Variable	Descripción	Tipo
1	Escuela	Nombre de la escuela	Nominal
2	Turno	Turno	Continuo
3	Entidad	Entidad	Continuo
4	Subsistema	Subsistema	Continuo
5	Sostenimiento	Sostenimiento	Continuo
6	Alumnos_evaluados	Alumnos evaluados	Continuo
7	Contestaron_50+_lenguaje	50% o más de preguntas contestadas en lenguaje	Continuo
8	Contestaron_50+_matematicas	50% o más de preguntas contestadas en matemáticas	Continuo
9	P1_lenguaje_dominio_I	% de alumnos evaluados en lenguaje (Nivel I)	Continuo
10	P1_lenguaje_dominio_II	% de alumnos evaluados en lenguaje (Nivel II)	Continuo
11	P1_lenguaje_dominio_III	% de alumnos evaluados en lenguaje (Nivel III)	Continuo
12	P1_lenguaje_dominio_IV	% de alumnos evaluados en lenguaje (Nivel IV)	Continuo
13	P1_matematicas_dominio_I	% de alumnos evaluados en matemáticas (Nivel I)	Continuo
14	P1_matematicas_dominio_II	% de alumnos evaluados en matemáticas (Nivel II)	Continuo
15	P1_matematicas_dominio_III	% de alumnos evaluados en matemáticas (Nivel III)	Continuo
16	P1_matematicas_dominio_IV	% de alumnos evaluados en matemáticas (Nivel IV)	Continuo

Los requerimientos principales para la construcción de la aplicación fueron: a) tener una sección para generar la vista de datos minable, b) incluir un algoritmo particional tipo K-means, c) incluir el método del codo (elbow method) para analizar el número deseado de grupos, d) permitir al usuario crear un proyecto de minería de datos que sea dinámico y de fácil uso, y e) tener opciones de ayuda para retroalimentar al usuario sobre el funcionamiento de la herramienta. Como resultado en la Fig. 3 se muestra la pantalla principal de la aplicación para la adquisición de patrones centrada en el usuario. Los módulos implementados son: a) origen de datos, b) análisis de datos, c) selección de variables, d) minería de datos, y e) validación. El

algoritmo K-means tiene como particularidad establecer a priori el número de grupos de entrada; por lo que se implementó el método del codo (elbow method) para obtener el número deseado de grupos. Este método permite identificar los grupos a través de una representación visual [27].

Para el análisis del logro académico se identificó 6 como número deseado de grupos (Fig. 4). Esta consideración fue con base en la variada información de los subsistemas educativos, sostenibilidad y los niveles de dominio del logro académico.

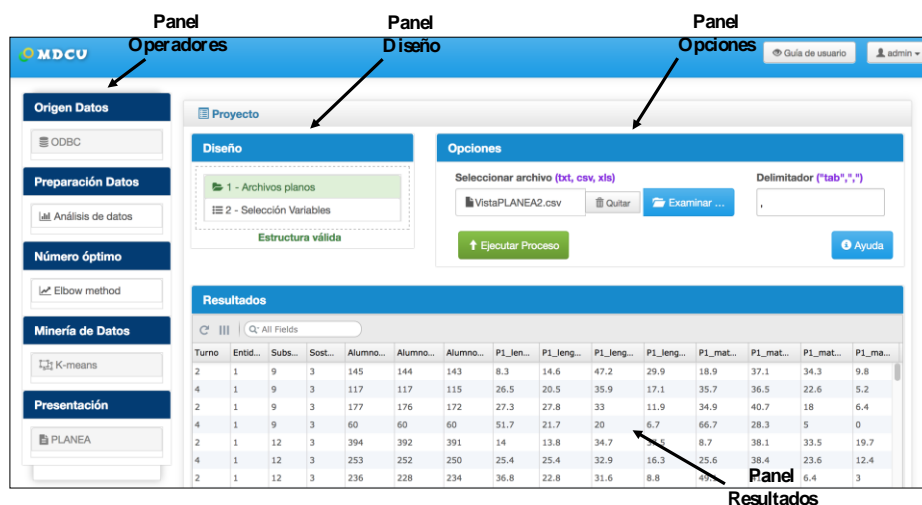


Fig. 3. Pantalla principal de la aplicación.

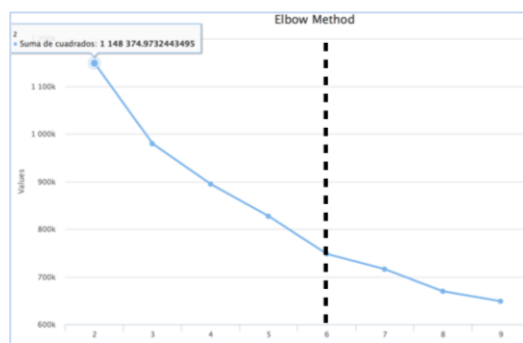


Fig. 4. Pantalla de la obtención del número deseado de grupos.

En la Fig. 5 se presenta un resumen de los grupos obtenidos por K-means. Como resultado se observó que el Grupo 1 alcanzó un mejor logro académico en Lenguaje en el nivel III (Bueno, 32%) y II en Matemáticas (Elemental, 39%). El Grupo 2 obtuvo un mayor logro académico para Lenguaje en el nivel I (insuficiente, 41%), mientras que en Matemáticas fue en el nivel I (insuficiente, 49%). El Grupo 3 destacó



por tener logros bajos, nivel I, insuficiente en Lenguaje (50%) e insuficiente en Matemáticas (59%). El Grupo 4 refleja el logro académico más bajo, con niveles Insuficientes en Lenguaje y Matemáticas con 64% y 78%, respectivamente. El Grupo 5 destacó por alcanzar altos porcentajes de insuficiencia en Lenguaje (39%) y Matemáticas (45%). Finalmente, el Grupo 6 obtuvo en Lenguaje niveles Bueno (30%) e Insuficiente (29%), mientras que en Matemáticas resaltan los niveles Insuficiente (35%) y Elemental (34%). Estos resultados contrastan que existen alumnos que al finalizar sus estudios obtienen el certificado sin tener los conocimientos necesarios para posteriormente aprobar los exámenes de ingreso en las universidades del país. Trayendo como consecuencia que retrasen o detengan sus estudios universitarios.

Grupo	Resumen	Grupo	Resumen
1	No. Instituciones: 3375 Promedio de alumnos evaluados: 26 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 27% (911 de 3375) Nivel II (Elemental): 23% (776 de 3375) Nivel III (Bueno): 32% (1080 de 3375) Nivel IV (Excelente): 18% (608 de 3375) Dominio Matemáticas: Nivel I (Insuficiente): 32% (1080 de 3375) Nivel II (Elemental): 39% (1316 de 3375) Nivel III (Bueno): 19% (641 de 3375) Nivel IV (Excelente): 10% (338 de 3375)	2	No. Instituciones: 1727 Promedio de alumnos evaluados: 192 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 41% (708 de 1727) Nivel II (Elemental): 21% (363 de 1727) Nivel III (Bueno): 26% (449 de 1727) Nivel IV (Excelente): 12% (207 de 1727) Dominio Matemáticas: Nivel I (Insuficiente): 49% (846 de 1727) Nivel II (Elemental): 32% (553 de 1727) Nivel III (Bueno): 13% (225 de 1727) Nivel IV (Excelente): 6% (103 de 1727)
3	No. Instituciones: 3132 Promedio de alumnos evaluados: 93 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 50% (1566 de 3132) Nivel II (Elemental): 20% (626 de 3132) Nivel III (Bueno): 21% (658 de 3132) Nivel IV (Excelente): 9% (282 de 3132) Dominio Matemáticas: Nivel I (Insuficiente): 59% (1848 de 3132) Nivel II (Elemental): 27% (846 de 3132) Nivel III (Bueno): 10% (313 de 3132) Nivel IV (Excelente): 4% (125 de 3132)	4	No. Instituciones: 5816 Promedio de alumnos evaluados: 23 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 64% (3722 de 5816) Nivel II (Elemental): 18% (1047 de 5816) Nivel III (Bueno): 14% (814 de 5816) Nivel IV (Excelente): 4% (233 de 5816) Dominio Matemáticas: Nivel I (Insuficiente): 78% (4537 de 5816) Nivel II (Elemental): 18% (1047 de 5816) Nivel III (Bueno): 3% (174 de 5816) Nivel IV (Excelente): 1% (58 de 5816)
5	No. Instituciones: 30 Promedio de alumnos evaluados: 969 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 39% (12 de 30) Nivel II (Elemental): 21% (6 de 30) Nivel III (Bueno): 26% (8 de 30) Nivel IV (Excelente): 14% (4 de 30) Dominio Matemáticas: Nivel I (Insuficiente): 45% (14 de 30) Nivel II (Elemental): 34% (10 de 30) Nivel III (Bueno): 14% (4 de 30) Nivel IV (Excelente): 7% (2 de 30)	6	No. Instituciones: 459 Promedio de alumnos evaluados: 356 Dominio Lenguaje y Comunicación: Nivel I (Insuficiente): 29% (133 de 459) Nivel II (Elemental): 21% (96 de 459) Nivel III (Bueno): 30% (138 de 459) Nivel IV (Excelente): 20% (92 de 459) Dominio Matemáticas: Nivel I (Insuficiente): 35% (161 de 459) Nivel II (Elemental): 34% (156 de 459) Nivel III (Bueno): 19% (87 de 459) Nivel IV (Excelente): 12% (55 de 459)

Fig. 5. Resumen de los grupos obtenidos por el algoritmo K-means.

Para la presentación de los resultados obtenidos se realizaron sesiones de trabajo con especialistas en educación. Entre los requerimientos levantados destacan: a) los resultados deben mostrarse a nivel federal y estatal, b) el usuario debe pasar de un nivel a otro sin restricciones, y c) incluir gráficas para visualizar las variables representativas de PLANEA. Así, se incluyó una variedad de gráficas para tener un

mayor entendimiento sobre el logro académico registrado en PLANEA 2015. En la Fig. 6 se presenta un extracto de la presentación de patrones en el nivel estatal.



Fig. 6. Módulo de presentación de patrones sobre la fuente de datos PLANEA 2015.

Al dar clic en el mapa de la república mexicana se despliega una pantalla emergente con información de los grupos obtenidos mediante gráficas, dividido por variables, como: sostenimiento, subsistema y logro académico (lenguaje y comunicación, y matemáticas). En la Fig. 7 se visualiza también que el mayor número de escuelas en Veracruz pertenecen al subsistema de Tele Bachillerato, con 1022 instituciones. Con respecto al sostenimiento educativo, se observa que la mayor cantidad de instituciones son del tipo estatal, con 1212 escuelas, mientras que otro menor número están distribuidos entre particulares y federales.



Fig. 7. Distribución de las escuelas de acuerdo al tipo de subsistema educativo.

Las pruebas de usabilidad sobre la aplicación permitieron detectar mejoras en la interfaz para optimizar la experiencia del usuario. Estas pruebas se dividieron en dos etapas: a) con ocho usuarios con conocimientos de minería de datos para analizar la aplicación de adquisición de patrones, y b) con cuatro usuarios especialistas en educación para analizar el módulo de presentación de patrones. En la Fig. 8 se muestran los resultados obtenidos.

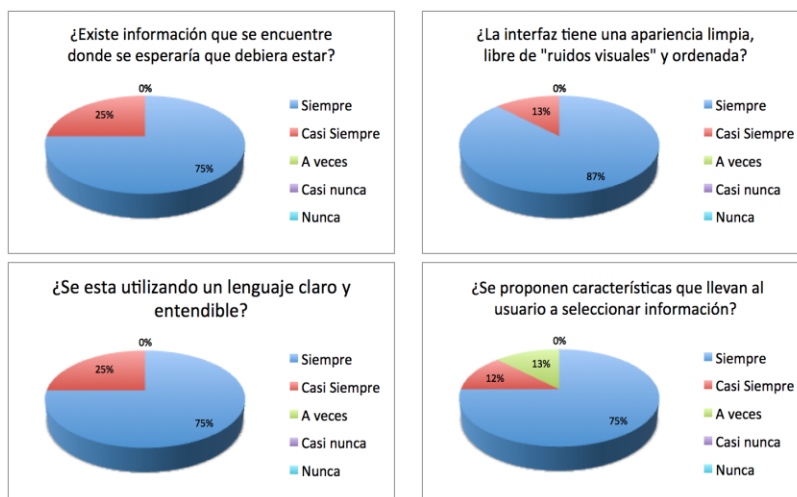


Fig. 8. Resumen resultados pruebas de usabilidad.

En general, las pruebas fueron favorecedoras, demostrando que la mayoría de los usuarios encontraron las cosas donde ellos esperaban, también mostraron agrado en el lenguaje en el que se les presentó la información, así mismo la mayoría coincidió en que la apariencia de la interfaz de usuario fue limpia y libre de ruidos visuales.

## 5. Conclusiones

Se realizó un análisis del logro académico de estudiantes en el nivel medio superior, tomando como base una minería de datos centrado en el usuario. Se utilizó como fuente de información la base de datos del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA).

Resultado de los grupos obtenidos tras la aplicación del algoritmo K-means, se observó que existe una deficiencia en el logro académico alcanzado por los estudiantes y por tanto por las instituciones educativas, trayendo como consecuencia que estos no adquieran los conocimientos necesarios para continuar sus estudios superiores.

Se comprobó que el método del codo fue útil para detectar el número deseado de grupos en el algoritmo K-means. Seis fueron los grupos obtenidos, destacando los niveles insuficiente y elemental, tanto para matemáticas y lenguaje.

Las pruebas de usabilidad realizadas permitieron identificar mejoras sobre en la aplicación, destacando que la interfaz tiene una apariencia limpia y libre de ruidos visuales y que el lenguaje utilizado es claro y entendible.

Respecto a la satisfacción por parte de los usuarios que participaron en las pruebas de usabilidad, reflejaron resultados positivos, mostrando que la intervención del usuario en todo momento del proyecto ayuda a entender mejor el problema.

**Agradecimiento.** Este trabajo forma parte del proyecto “Infraestructura para agilizar el desarrollo de sistemas centrados en el usuario” financiado por el Consejo Nacional de Ciencia y Tecnología, en el marco de Cátedras CONACYT (Ref. 3053).

## Referencias

1. López, C.: Minería de datos: técnicas y herramientas. Editorial Paraninfo (2007)
2. Hernández J., Ramírez M. J., Ferri C.: Introducción a la Minería de Datos. Pearson Educación, Editorial Pearson Prentice Hall, Madrid, España (2004)
3. Witten, I., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers (2005)
4. Molero G.: Clasificador bayesiano para el pronóstico de la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen hispano (Tesis doctoral). Universidad de Guadalajara, México (2014)
5. Zhao, Y., Chen, Y., Yao, Y.: User-centered interactive data mining. In: Cognitive Informatics, 5th IEEE International Conference, 457–466 (2006)
6. Abras, C.; Maloney-Krichmar, D.; Preece, J.: User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications (2004)
7. Nielsen, J.: Usability engineering. Academic Press Limited, Massachusetts, Estados Unidos (1993)
8. Horberry, T., Burgess-Limerick, R., Steiner, L.: Human Centred Design for Mining Equipment and New Technology. In: Proceedings 19th Triennial Congress of the IEE, 9, 14 (2015)
9. Sumathi, S., Sivanandam, S.: Introduction to Data Mining and its Applications. Studies in Computational Intelligence, 29, editado por Springer-Verlag, Heidelberg, Alemania (2006)
10. Moine, J., Gordillo, S., y Haedo, A.: Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. In: Presentado en el VIII Workshop Bases de Datos y Minería de Datos, 931–938 (2011)
11. Peralta, F.: Elementos para un mapa de actividades para proyectos de explotación de información. Facultad Regional Buenos Aires, Argentina (2013)
12. Vanrell, J., Bertone, R.: Modelo de Proceso de Operación para Proyectos de Explotación de Información. In: XVI Congreso Argentino de Ciencias de la Computación, 674-682 (2010)
13. Moine, J.: Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo (Tesis doctoral). Universidad Nacional de la Plata, Buenos Aires, Argentina (2013)
14. KDnuggets.: Data Mining, Analytics, Big Data, and Data Science. [www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html](http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html)
15. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step Data Mining Guide (2000)
16. Flores, H.: Detección de Patrones de Daños y Averías en la Industria Automotriz (Tesis de Maestría). Universidad Tecnológica Nacional (2009)
17. Luján-Mora, S.: Data warehouse design with UML (Tesis Doctoral). Universidad de Alicante, España (2005)
18. Ho, T., Nguyen, T., Nguyen, D.: Visualization support for a user-centered KDD process. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Japan Advanced Institute of Science and Technology Tatsunokuchi, Ishikawa, 519–524 (2002)

19. Ho, T., Nguyen, T., Nguyen, D.: A User-Centered Visual Approach to Data Mining, The system D2MS, Estados Unidos, 213–224 (2002)
20. Martín, A.: MPIu+a Ágil: El modelo de proceso centrado en el usuario como metodología ágil, Universidad de Lleida (2010)
21. Granollers, T.: MPIu+ a. Una metodología que integra la Ingeniería del Software, la Interacción Persona-Ordenador y la Accesibilidad en el contexto de equipos de desarrollo multidisciplinares, Universidad de Lleida (2004)
22. Hassan, Y., Ortega, S.: Informe APEI sobre usabilidad, España (2009)
23. Arnaut, A., Giorguli, S.: Los grandes problemas de México, El colegio de México, Ciudad de México (2010)
24. SEP.: Plan nacional para la evaluación de los aprendizajes (PLANEA), [www.inee.edu.mx/index.php/planea](http://www.inee.edu.mx/index.php/planea)
25. PLANEA-MS: Manual para usuarios, Plan Nacional para las Evaluaciones de los Aprendizajes en el nivel medio superior (2015)
26. De Ibarrola, M.: Los grandes problemas del sistema educativo mexicano, Perfiles educativos, 34 (SPE), 16–28 (2012)
27. Kodinariya, T., Makwana, P.: Review on determining number of Cluster in K-Means Clustering, International Journal, 1(6), 90–95 (2013)