v. 2, n. 10, 2022

International Journal of Human Sciences Research

APLICACIONES DE MINERÍA DE DATOS EN DATOS DE INGRESO DE ESTUDIANTES DE INGENIERÍA

Danice D. Cano Barrón
Dpto. de Ingeniería en Sistemas Computacionales, Instituto Tecnológico Superior de Motul, Carretera Mérida Motul, México

Humberto J. Centurión Careña
Dpto. de Ingeniería Electromecánica, Instituto Tecnológico Superior de Motul Carretera Mérida
Motul, México

Resumen: El estudio exploratorio realizado en el Departamento de Ingeniería en Sistemas Computacionales del ITS Motul utiliza minería de datos para correlacionar el desempeño académico de los estudiantes a través del promedio de calificaciones, con la prueba de ingreso a nivel superior, a través del instrumento EXANI II, con el fin de identificar patrones acerca de los principales indicadores académicos que presentan los estudiantes de nuevo ingreso y que ayuden a predecir su comportamiento posterior. Además de clasificarlos utilizando el algoritmo de árbol J48 para identificar las variables principales que determinen su potencial rendimiento académico, resultando ser la comprensión lectora y el índice de pensamiento matemático. Se utiliza en el proyecto Weka, software que implementa técnicas de machine learning, para el análisis automático de los datos, dejando la interpretación de la información resultante en manos del investigador. Los indicadores elegidos están asociados únicamente al desempeño académico y los resultados preliminares del estudio son una pauta para realizar estudios más profundos con mayor cantidad de datos y número de indicadores que, a largo plazo, permitan una caracterización formal de las necesidades de los estudiantes de nuevo ingreso.
Palabras clave: minería de datos, educación superior, J48, Weka, EXANI II.

INTRODUCCIÓN

El Instituto Tecnológico Superior de Motul inició sus operaciones el 18 de septiembre de 2000, con 2 carreras, 67 alumnos y 7 profesores, en la actualidad se cuentan con 5 programas educativos, cerca de 1000 estudiantes y una planta docente de más de 50 profesores que atienden las necesidades formativas de manera permanente. Debido al crecimiento de su matrícula y a la diversidad de necesidades de los estudiantes de nuevo ingreso se han implementado una serie de actividades de apoyo para facilitar no sólo su ingreso sino principalmente su egreso y titulación como parte de las funciones sustantivas de la organización, buscando mantener e inclusive mejorar sus indicadores de calidad.

En este sentido se han establecido una serie de programas de apoyo a su formación, cursos de formación inicial, asesorías académicas para los que lo requieran, atención a estudiantes en situación de riesgo, etc., sin embargo se ha detectado la necesidad de identificar desde el ingreso aquellos elementos que puedan ayudar a identificar a aquellos estudiantes que requieran de apoyos específicos como un medio de trabajar inicialmente con los profesores y tutores en conjunto para hacer un seguimiento oportuno de los casos y más adelante poder predecir el comportamiento durante su proceso formativo para tomar acciones preventivas más que reactivas como se acostumbra.

El programa educativo de Ingeniería en Sistemas Computacionales pierde en promedio a 14 estudiantes entre su ingreso y su egreso (véase la Fig. 1), las razones son variadas desde aspectos socio-económicos, familiares y/o académicos, ya que los estudiantes de nivel superior no sólo dependen de los conocimientos académicos en su formación [1]. Son las últimas razones las que interesa a la institución resolver de manera temprana de manera que el progreso de los estudiantes sea exitoso.

En este siglo donde el acceso a grandes volúmenes que se almacenan en bases de datos centralizadas y distribuidas en diversos dominios, como por ejemplo librerías digitales, archivos de imágenes, bioinformática, cuidados médicos, finanzas e inversión, fabricación y producción, negocios y marketing, redes de telecomunicación, etc.,

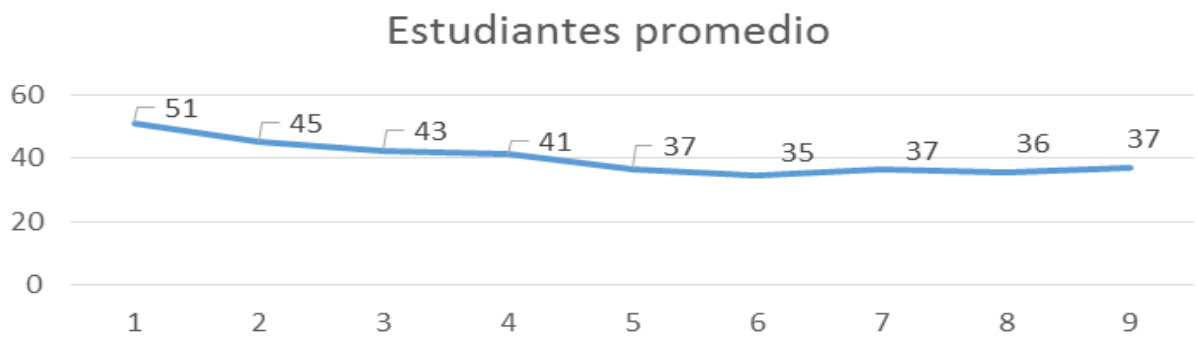

Fig. 1. Número de estudiantes promedio por semestre de las últimas 4 generaciones.

serán de gran importancia para interpretar la información y el conocimiento de los datos distribuidos por todo el mundo [2]. De aquí que las estrategias orientadas no sólo al almacén eficiente sino a la búsqueda de patrones relacionados con los datos dentro de ellas se han convertido en un área importante de desarrollo tecnológico.

La predicción del éxito de los estudiantes resulta crucial para las instituciones de educación superior debido a que la calidad del proceso de enseñanza y aprendizaje está fuertemente relacionada con la habilidad de responder a las necesidades de formación de los estudiantes [3]. En este sentido, existen datos e información que son almacenados de manera regular y permiten a las autoridades pertinentes entender en qué medida los estándares de calidad son conseguidos y en qué cambios deberán de hacerse en caso de ser requeridos. Es por eso que estudios de esta naturaleza que empiezan a tratar de dar sentido a la cantidad de información con la que se cuenta son importantes para dar sentido a lo que históricamente ha ocurrido con los estudiantes.

Con el ingreso asociado a pruebas estandarizadas como los Exámenes Nacionales de Ingreso (EXANI), se cuenta con datos relacionados no sólo con las habilidades académicas y la formación del nivel inmediato anterior de los estudiantes, sino con datos relacionados con características socioeconómicas, lo que debería facilita la toma de decisiones asociadas a su formación académica con base en los resultados de generaciones anteriores.

A través de técnicas de minería de datos aplicadas a los datos históricos almacenados en las bases de datos de una IES, es posible tratar de identificar las características de ingreso de los estudiantes que tienen una alta probabilidad de abandonar sus estudios. Identificar estos atributos permitiría predecir qué estudiantes son los más vulnerables y tomar acciones anticipadas [4].

MINERÍA DE DATOS Y EDUCACIÓN SUPERIOR

La minería de datos en la educación superior es un campo de investigación reciente que está ganando terreno debido al potencial impacto positivo que pudiera tener en las instituciones de educación superior [3]. Y en este sentido su principal fortaleza radica en que forma parte del proceso de descubrimiento de conocimiento a través de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles [5]. Traduciendo la cantidad de información obtenida del proceso educativo y de su resultados en estructuras de toma de

decisión que facilite a la administración y profesores la tarea de formar eficientemente a los estudiantes.

Desde hace muchos años, pedagogos y psicólogos de la educación se han preocupado por las condiciones socioeconómicas y educativas de los estudiantes y su influencia en el rendimiento académico [6]. Es por ello que instrumentos como el EXANI II que provee información del contexto, además de los resultados académicos revisten importancia para determinar un espectro más amplio de los factores que pueden estar afectando los resultados académicos de los postulantes.

Para este estudio exploratorio se trabajó únicamente con indicadores académico y el promedio general de los estudiantes al cierre del semestre lectivo 2016 B, para estudiar en qué medida los resultados de las diferentes habilidades pudieran permitir predecir el rendimiento académico que presentan. Con ello en mente se pensó en un proceso de minería de datos a través del cual se pudiera hacer un estudio preliminar de los datos con los que se cuenta y hacer un proceso previo de análisis.

DESERCIÓN Y EDUCACIÓN SUPERIOR

La deserción de estudiantes universitarios es un problema particularmente serio en instituciones educativas públicas y privadas latinoamericanas, europeas y norteamericanas [7]. Esta realidad impacta no sólo a nivel interno de la instituciones por cuestiones presupuestales, sino a nivel gobierno que invierte importantes sumas de dinero en la formación de la población y necesita responder con buenos indicadores al uso de los recursos que son usados. Sin importar la nacionalidad este fenómeno requiere ser estudiado debido a que los índices cada vez son más altos y se requiere de disminuirlos [8].

Estudios recientes identifican como elementos explicativos la falta de personalidad y madurez intelectual de los estudiantes, así como la falta de conocimientos y habilidades previas necesarias para realizar estudios superiores, y combinar aspectos personales con aspectos académicos [9].

Sin embargo al analizar estudios que usan técnicas de estadística clásicas para determinan sus resultados, sin indagar más en posibles patrones ocultos en los datos que aporten una perspectiva diferente al problema de la deserción, establecen que los determinantes de la retención universitaria son: reclutamiento y admisión, servicios académicos, currículo e instrucción, servicios estudiantiles y ayudas financieras [10].

Independientemente de las razones, es importante el poder analizar cómo el fenómeno sucede en los diversos contextos y cómo éstos ayudarán a entender a nuevas generaciones.

EXAMEN NACIONAL DE INGRESO A LA EDUCACIÓN SUPERIOR (EXANI II)

El EXANI-II es un instrumento estándar utilizado en México para identificar el desarrollo académico de los aspirantes a una Institución Educativa de nivel Superior. El examen está compuesto en dos partes: una evalúa los conocimientos y la otra se trata de un cuestionario de contexto.

El examen de conocimientos se divide en uno de selección y uno de diagnóstico. El primero evalúa los conocimientos del sustentante en áreas de español, matemáticas, tecnologías de la información y comunicación, así como habilidades lógico-matemáticas como verbales. El segundo se aplica de acuerdo con la licenciatura a la que desea ingresar. Finalmente, el cuestionario de contexto sólo recaba información socioeconómica de los aspirantes, datos generales, escolares,

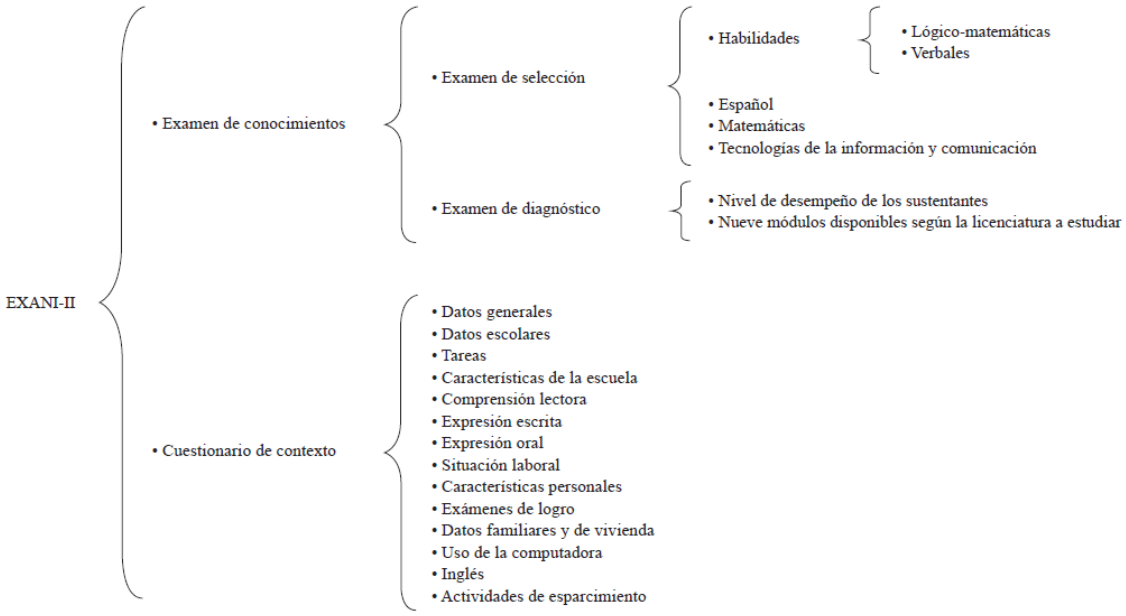

Fig. 2. Composición del EXANI III de CENEVAL.

situación laboral, características personales, datos familiares, de vivienda, entre otros. La composición general del examen se puede observar en la Fig. 2 [11].

Para este estudio, se consideraron únicamente los aspectos relacionados con las habilidades lógico-matemáticas y verbales, así como las secciones de español y matemáticas como indicadores importantes del perfil de los estudiantes, debido a que se trata de un primer intento de identificar elementos que pudieran funcionar como predictores del comportamiento de los estudiantes durante su proceso formativo.

Es una realidad que los resultados de la evaluación del CENEVAL debieran ser utilizadas, tanto por las autoridades responsables de la política educativa como por las propias instituciones de donde proceden los estudiantes, para corregir las causas del bajo aprovechamiento [12], de manera que este estudio pretende responder a esta área de oportunidad que no muchas instituciones utilizan de manera puntual para comprender la realidad que viven sus estudiantes de nuevo ingreso.

METODOLOGÍA

Este estudio parte de la gran cantidad de factores recabados por el cuestionario de contexto del EXANI II y la información que proporciona control escolar de los estudiantes sobre su situación escolar. Se consideraron para este trabajo dos generaciones de ingreso al programa de Ingeniería en Sistemas Computacionales 2015 y 2016 haciendo un total de 71 estudiantes. Cabe mencionar que este número de estudiantes no son todos los ingresos que se dieron durante ese tiempo debido a que se cuenta con la opción de ingresar con los resultados de la prueba aun cuando hayan presentado en otra institución, lo que limita el número potencial de ítems a estudiar pues no se tiene acceso a los datos de contexto.

Una vez identificados a los estudiantes y sus datos de contexto se procedió a limitar la base de datos para poder trabajar únicamente

Indicador	Escala	Rango	Categoría
IPMAT, IPAN, IELE, ICLE e ICNE.	700 - 1300	Índice < 1000 1000≤ Índice ≤1150 Índice>1150	Elemental Satisfactorio Sobresaliente
Promedio	0 - 100	Prom < 70 70 ≤ Prom < 80 80 ≤ Prom < 90 90 ≤ Prom	Insuficiente Suficiente Satisfactorio Sobresaliente

Tabla 1. Indicadores utilizados en el estudio y su categorización.

con 5 elementos básicos relacionados con su comportamiento académico: pensamiento matemático (IPMAT), pensamiento analítico (IPAN), estructura del lenguaje (IELE), comprensión lectora (ICLE) y el índice general de CENEVAL (ICNE).

Una vez elegidas los aspectos a considerar para el estudio exploratorio, se construyeron escalas para facilitar su análisis, como se puede observar en la Tabla 1. Para los indicadores de CENEVAL utilizaron las categorías que marcan en su sitio y para el caso del promedio general se siguió una métrica general semejante.

Para este estudio se utilizó Weka para realizar el estudio preliminar de las variables de estudio, ya que contiene una colección de algoritmos de aprendizaje para actividades de minería de datos, e incluye herramientas para el preprocesado de datos, clasificación, regresión y reglas de asociación [13]. Una vez categorizados los datos se utilizó el explorador del sistema para los primeros análisis de las variables de estudio.

RESULTADOS OBTENIDOS

En la Fig.3 se puede observar la distribución de las categorías del promedio, donde la mayor cantidad de los estudiantes presentan desempeño en la categoría de satisfactorio y la menor cantidad presentan un desempeño en insuficiente, siendo las categorías de insuficiente y de sobresaliente semejantes en cantidad.

El comportamiento del índice de pensamiento matemático (IPMAT) se pude observar en la Fig. 4 que un alto número de los estudiantes con promedio insuficiente presentan alto índice de pensamiento matemático, mientras que una gran cantidad de los de nivel satisfactorio presentan un nivel elemental.

En cuanto al índice de pensamiento analítico (IPAN) se pude observar en la Fig. 5 que una cantidad mediana de estudiantes con desempeño insuficiente sacan calificaciones sobresalientes en el EXANI II mientras que los demás niveles de desempeño de calificaciones (insuficiente, suficiente y satisfactorio) se dividen en proporciones semejantes en los tres niveles del índice.

El índice relacionado con la estructura del lenguaje (IELE) se puede observar en la Fig. 6 que si bien presenta el mismo comportamiento de los casos anteriores, las proporciones son menores para el caso de los insuficientes que presenten niveles sobresalientes de desempeño, lo cual llevaría a pensar que este indicador sería una fuente importante de detección de necesidades de los estudiantes de nuevo ingreso.

El último de los indicadores del EXANI II relacionado con la habilidad de comprender textos es el único que permea a los

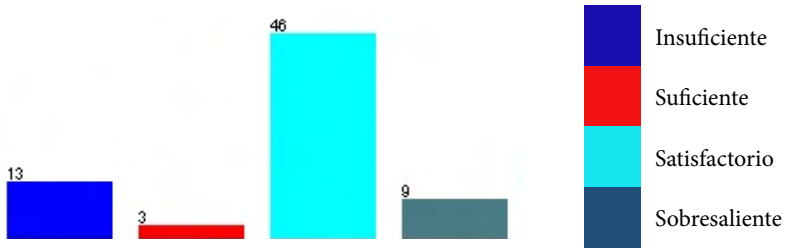

Fig. 3. Categorización de los estudiantes de acuerdo con su promedio general.

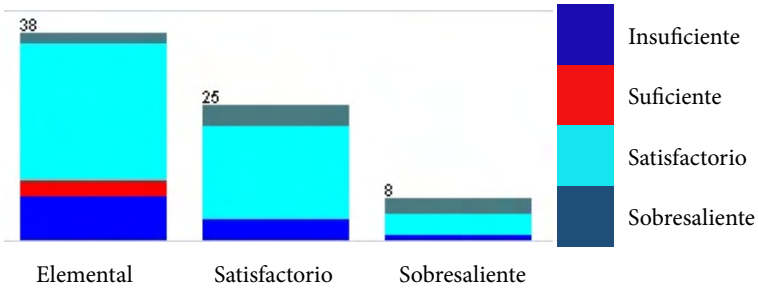

Fig. 4. Categorización de los estudiantes de acuerdo con el índice de Pensamiento matemático.

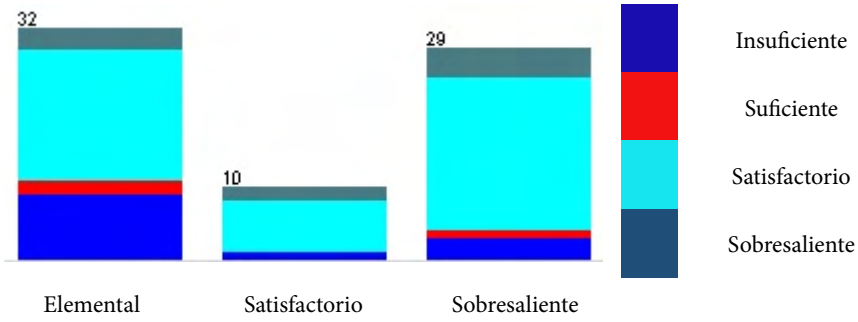

Fig. 5. Categorización de los estudiantes de acuerdo con el índice de pensamiento analítico

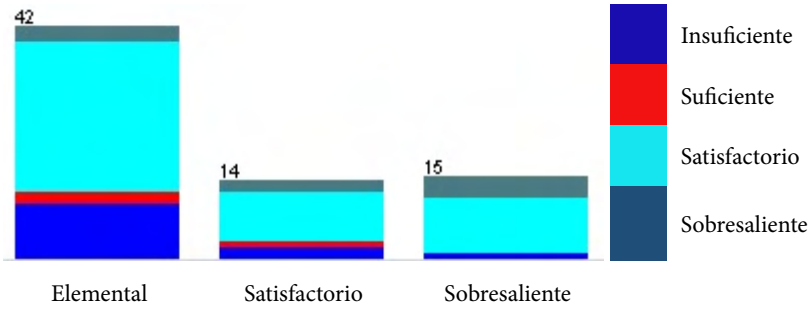

Fig. 6. Categorización de los estudiantes de acuerdo con el índice de estructura del lenguaje.

estudiantes que tienen niveles insuficientes de desempeño, ya que únicamente llegan a los niveles satisfactorios y no al sobresaliente como se observa en la Fig. 7. Considerando esto, sería importante que se tomara en cuenta este indicador prioritariamente al ingresar los estudiantes y trabajar con ello durante el curso propedéutico para mejorar el proceso de egreso.

Finalmente, en cuanto al índice general de CENEVAL (ICNE) se puede observar en la Fig. 8 que una alta proporción de los estudiantes presenta niveles elementales al ingresar al Instituto, lo que de entrada requiere de un proceso de nivelación. En cuanto a los demás niveles de desempeño, el nivel satisfactorio contiene a todo tipo de estudiantes, mientras que el sobresaliente no tiene elementos cuyos promedios estén en un nivel suficiente.

Este recorrido por los principales índices que arroja la prueba de CENEVAL permite empezar el proceso de correlación entre los resultados obtenidos por los aspirantes de nuevo ingreso a la carrera de Ingeniería en Sistemas Computacionales.

Se utilizó el algoritmo J48 para determinar si existía un modelo para clasificar a los estudiantes de acuerdo con sus principales indicadores de rendimiento académico propuestos por la prueba de CENEVAL.

En la Fig.8 se puede observar que el principal factor para clasificar a los estudiantes es su comprensión lectora, seguido por el pensamiento analítico y, dependiendo de la generación a la que pertenecen, el índice de pensamiento matemático.

Analizando a profundidad el árbol es importante destacar que los estudiantes con niveles sobresalientes de comprensión lectora y pensamiento matemático tiende a ser sobresalientes, pero únicamente si pertenecen a la generación 2016.

Esto parece coincidir cuando se evalúan los atributos para determinar su correlación con los resultados de su desempeño académico cuando se corren los algoritmos Gain Ratio e Info Gain como se puede observar en la Tabla 2.

CONCLUSIONES Y TRABAJOS FUTUROS

Este trabajo ha permitido tener una idea general de cómo los principales indicadores de la prueba EXANI II pueden ayudar a comprender el rendimiento académico de los estudiantes de la carrera. Si bien es una muestra muy básica de los indicadores que incluye la prueba, permite entender el proceso de minería de datos.

Entre los principales resultados se identificaron tres índices muy específicos que parecen estar relacionados con el desempeño de los estudiantes: ICLE, IPMAT e IELE que filtran en mayor o menor medida lo que ocurre con los estudiantes a lo largo de su proceso formativo.

El trabajo a futuro se da en dos líneas, la primera tiene que ver con la detección e inclusión de datos de contextos que permita identificar las necesidades de los estudiantes desde el primer ingreso y la segunda el poder incluir más estudiantes conforme ingresen más generaciones y verificar si el comportamiento es algo cambiante o permanece estable.

El siguiente proceso, será el poder determinar cómo los factores socioeconómicos pueden hacer el proceso de selección más certero y eficiente. Así mismo, será interesante el comparar las diversas carreras que se imparten para determinar su existen perfiles específicos para cada una de ellas.

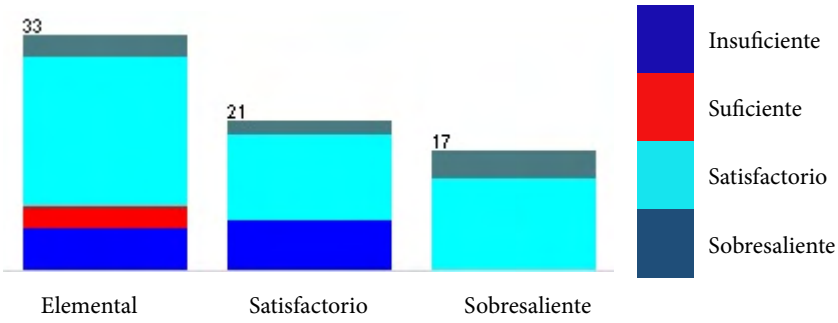

Fig. 7. Categorización de los estudiantes de acuerdo con el índice de comprensión lectora.

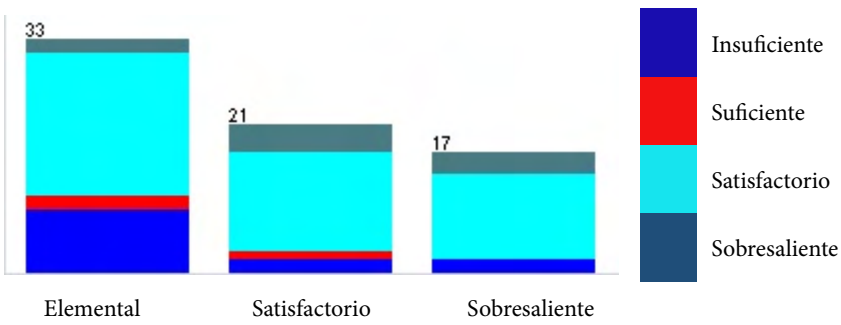

Fig. 8. Categorización de los estudiantes de acuerdo con el índice general de CENEVAL.

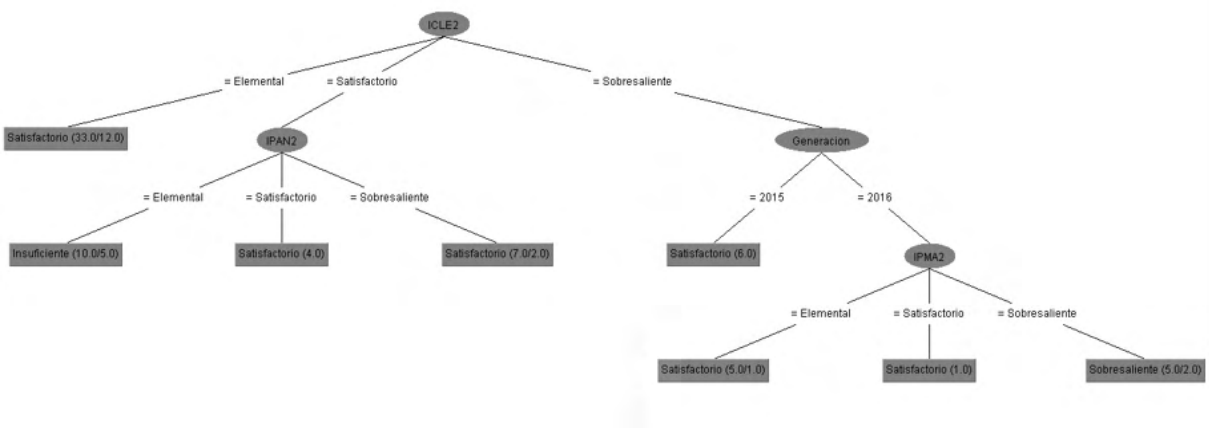

Fig. 8. Árbol de clasificación J48 de los datos.

Gain Ratio		**Info Gain**	
0.1015	7 ICLE2	0.155	7 ICLE2
0.0863	2 Sexo	0.0948	4 IPMA2
0.0777	1 Generación	0.0758	1 Generación
0.0693	4 IPMA2	0.0706	3 ICNE2
0.0497	6 IELE2	0.0687	6 IELE2
0.0462	3 ICNE2	0.0685	2 Sexo
0.0395	5 IPAN2	0.057	5 IPAN2

Tabla 2. Indicadores utilizados en el estudio y su categorización.

REFERENCIAS

1. Pontón, M.: Factores que afectan el desempeño de los alumnos mexicanos en edad de educación secundaria. Un estudio dentro de la corriente de eficacia escolar. *Revista electrónica Iberoamericana sobre Calidad, eficacia y cambio en Educación*, 4 (3) pp. 30-53. (2006)

2. Riquelme, J., Ruiz, R.; Gilbert, K.: Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, pp. 11-18 (2006)

3. Al-Twijri, M.; Noamanb, A.: *A New Data Mining Model Adopted for Higher Institutions. Procedia Computer Science*, pp. 836 – 84 (2015)

4. Timarán, R.: Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos. *Revista científica Guillermo de Ockham*, pp. 121-130 (2010)

5. Fayyad, U.: Data Mining and Knowledge Discovery: Making Sense out of Data. IEEE Intelligent Systems, Vol. 11, No. 5, USA, ISSN: 0885-9000 (1996)

6. González, E.: Estudio sobre factores contexto en estudiantes universitarios para conocer por qué unos tienen éxito mientras otros fracasan. *Revista Intercontinental de Psicología y Educación*, vol. 15, núm. 2, julio-diciembre, Universidad Intercontinental. Distrito Federal, México. pp. 135-154, (2013)

7. Cruz, E., Gática, L., García, P.; García, J.: Academic Performance, School Desertion And Emotional Paradigm In University Students. Contemporary Issues In Education Research, pp. 25-35 (2010)

8. Arce, M., Crespo, B.; Míguez-Álvarez, C.: Higher Education Drop-Out in Spain—Particular Case of Universities in Galicia. International Education Studies, pp. 247-265 (2015)

9. Romo, A.; Fresán, M.: Los Factores Curriculares y Académicos Relacionados con el Abandono y el Rezago. http://proyectodeintervencion-ca2.wikispaces.com/file/view/docto+3.pdf. Accedido el 19 de Octubre de 2017

10. Miranda, M.; Guzmán, J.: Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos. *Formación Universitaria*, pp. 61-68 (2017)

11. García, G.; García-Hernández, R.; Ledeneva, Y.: Reglas que describen la deserción y permanencia en los estudiantes de la UAP Tianguistenco de la UAEM. *Revista Ciencia ergo-sum*, vol. 21, 2, julio-octubre 2014. Universidad Autónoma del Estado de México, Toluca, México, pp 121-132 (2013)

12. Vidal, R.: Evaluación y calidad. La ruta para el fortalecimiento de las ies. Ponencia presentada en la Jornadas Internacionales para la Gestión de la Calidad Educativa, Cancún, México, www.upn.mx/index.php/comunidad/carpeta/.../14-enero-2012. Accedido el 9 de abril de 2017

13. Machine Learning at the University of Waikato.: Weka 3: Data mining software in java. *University of Waikato*. http://www.cs.waikato.ac.nz/ml/weka/. Accedido el 19 de Abril de 2017