

## Minería de datos: Impacto de Actividades Cotidianas en el Rendimiento Estudiantil

### [ Data Mining: Impact of Daily Activities on Student Performance ]

*Huerta Luis, Ruiz Juan, Cabrera Nubia, Montiel Luis, Benítez Felipe, and Ramírez Víctor*

Department of Informatic, University of Istmo, Ciudad Ixtepec, Oaxaca, México

Copyright © 2016 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** The student performance has been affected for different factors, many of them are unobvious. The habits or daily activities undoubtedly have a deep effects on the student performance. In this work, the study of student daily activities, and the relationship with his academic performance, using Data Mining techniques was done. In the attribute selection phase, 5-13 attributes from the 35 total were selected. The students were classified in four classes related with their academic performance: low, regular, good and high; the classification accuracy was near to 90%, using algorithms like MLP, KNN and tree algorithms like Random Forest, Random Tree and J48. The activities and factors presented for low and high performance students, also the tendency of activities and factors in the four classes, are reported.

**KEYWORDS:** Student Performance, Classify, Attribute Selection.

#### 1 INTRODUCCIÓN

En México, existe un porcentaje considerable de estudiantes con debilidades académicas, así lo muestra la Evaluación Nacional de Logro Académico en Centros Educativos [1] en la categoría de estudiantes de bachillerato. En esta prueba se reporta que el 50% de estudiantes tiene habilidades lectoras y de comunicación en nivel insuficiente y elemental, así mismo, en habilidades matemáticas este porcentaje incrementa a 60.7%. Otro problema relacionado con los estudiantes de este país provocado por un bajo rendimiento académico es la deserción, de acuerdo a [2] el 43% de estudiantes entre 15 y 19 años que deberían cursar el nivel medio superior, bachillerato o carrera técnica no lo hacen. Aunque en los últimos años, el rendimiento académico ha ido mejorando, es necesario hacer un estudio profundo de los factores o actividades influyentes en tal rendimiento, con el fin de fomentar aquellos factores que beneficien y minimizar aquellos que lo afecten.

Entender los factores que contribuyen positiva y negativamente al rendimiento académico, es un tema del que se han ocupado disciplinas como la pedagogía, psicología, sociología y en los últimos años la minería de datos, y se ha vuelto de interés para las políticas públicas, directivos y docentes de los diferentes niveles educativos, dado que entender éstos factores permite tomar mejores decisiones para el diseño de estrategias que ayuden proveer una educación de calidad y así evitar la deserción escolar. La Minería de Datos como un área de investigación enfocada al descubrimiento de conocimientos en bases de datos, se ha aplicado con éxito en numerosas investigaciones en el ámbito educativo [3]. Sin embargo, hay aspectos de la vida cotidiana como el acoso escolar, el pasatiempo de los estudiantes, el número de horas que ven televisión, el deporte que practican y el uso del celular, por mencionar algunos, que han sido poco estudiados para determinar la relación que tienen con el bajo o alto rendimiento escolar de los estudiantes. El presente trabajo tiene como objetivo determinar, mediante técnicas de minería de datos, el impacto que tienen las actividades cotidianas de los estudiantes en su rendimiento escolar. Para ello se aplicaron encuestas a 208 estudiantes de nivel medio superior, 112 hombres y 96 mujeres. Se utilizó la herramienta WEKA 7.9 para el proceso de selección de atributos y clasificación de estudiantes de acuerdo a su rendimiento.

El artículo se encuentra organizado de la siguiente forma: en el capítulo 2 se hace una revisión de la aplicación de Minería de Datos para analizar factores que afectan el rendimiento de estudiantes y la deserción escolar, el capítulo 3 describen la muestra, los materiales y métodos que fueron utilizados en el estudio, el capítulo 4 describe el experimento, el capítulo 5 presenta los resultados obtenidos y finalmente el capítulo 6 presenta las conclusiones alcanzadas con este trabajo.

## **2 TRABAJOS RELACIONADOS**

Los siguientes trabajos relacionados tienen como fin determinar los factores que impactan en el rendimiento académico o predecir la deserción escolar. A diferencia del presente trabajo, ninguno hace un análisis de la manera en que factores cotidianos afectan el rendimiento académico. Los trabajos relacionados más recientes se detallan.

### **2.1 ANÁLISIS DE FACTORES QUE AFECTAN EL DESEMPEÑO ACADÉMICO DE LOS ESTUDIANTES, UN ENFOQUE DE MINERÍA DE DATOS**

El trabajo reportado en [4] hacen un estudio de factores relacionados con los cursos y los instructores. Uno de los objetivos es aplicar Minería de Datos, mediante la herramienta WEKA, para descubrir patrones entre los estudiantes que aprueban o reprueban un curso. El algoritmo de clasificación REPTree presentó los mejores resultados, con aproximadamente un 84% de exactitud. Este árbol de decisión determinó que el asistir continuamente a los cursos genera un mejor entendimiento y por lo tanto el estudiante aprobará con tomar el curso una sola vez. Si el estudiante considera las lecciones de un nivel simple o intermedio aprobará el curso tomándolo una sola vez. Si el estudiante percibe que el instructor está moderada o completamente preparado para impartir las clases, si además el instructor responde a las preguntas dentro y fuera del curso, y el sistema de evaluación del instructor se realiza de acuerdo con los objetivos del curso y no algún otro, entonces el estudiante aprobará el curso del instructor tomándolo una sola vez.

Básicamente, este trabajo relacionado toma como importante las asistencias, la preparación, forma de evaluar y resolución de dudas por parte del instructor, garantizan que el estudiante no repetirá el curso. A excepción de las asistencias, implícitamente el estudiante al dar una respuesta de cómo percibe la calidad del curso y del instructor indirectamente está respondiendo si pasó o no el curso, haciendo parcialmente evidente el patrón, de tal forma que para quienes reprobaron el curso y el instructor no fueron de buena calidad.

### **2.2 ESTUDIO DE ANÁLISIS DE FACTORES QUE AFECTAN EL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE PREGRADO EN PROGRAMAS INTERNACIONALES**

El estudio realizado en [5] tiene como objetivo identificar factores que influyen en el rendimiento académico del estudiante. Se hacen uso de clasificadores de árbol y una red neuronal; los primeros obtienen una precisión de 85.188%, superiores a la red en 1.313%. Se emplearon datos como: sexo, estado civil, ocupación del padre y madre, horas de trabajo, si cuenta con inglés pre-Universitario o no, si es hablante nativo del inglés, créditos tomados, entre otros datos. En las pruebas se emplea la herramienta de minería de datos WEKA.

Encontraron que los estudiantes que no tienen riesgo de un bajo rendimiento académico son aquellos que nunca llevaron un curso adicional de inglés, debido a tener sólidos conocimientos en el idioma antes de ingresar a la Universidad. Por otro lado, los estudiantes que tienen riesgo de un bajo rendimiento académico son los que llevan un curso adicional de inglés; algunos de estos estudiantes son casados o divorciados; algunos trabajan de un moderado a un alto número de horas por semestre; algunos de estos estudiantes tienen de 12 a 15 créditos tomados por semestre. Generalmente en la Universidad bajo estudio no se les permite tomar más de 12 créditos si su rendimiento académico es bajo ni más de 15 si no es alto. El atributo relacionado con el número de créditos da información, obvia de alguna manera para determinar el nivel de rendimiento académico del estudiante, puesto que el número de créditos que el estudiante cursa está condicionado por su rendimiento. En otras palabras, este atributo debió omitirse en los experimentos. Por otro lado, el estudio encuentra que tomar un curso adicional de inglés, el estado civil y las horas de trabajo afectan el rendimiento académico del estudiante.

### **2.3 PREDICCIÓN DEL FRACASO ESCOLAR MEDIANTE TÉCNICAS DE MINERÍA DE DATOS**

La investigación realizada en [6] propone la utilización de técnicas de minería de datos para detectar, cuáles son los factores que más influyen para que los estudiantes fracasen; se realizaron un conjunto de experimentos con el objetivo de predecir con un buen grado de exactitud el estado académico de los estudiantes del primer semestre de la Unidad Académica Preparatoria de la Universidad Autónoma de Zacatecas (UAPUAZ); los datos de estudio fueron obtenidos de tres fuentes diferentes: de la encuesta aplicada a los alumnos de nuevo ingreso, del estudio socioeconómico del Centro Nacional de Evaluación (CENEVAL) y de la información proporcionada por el departamento de servicios Escolares, por lo que en total

fueron 77 atributos a estudiar, debido a la gran cantidad de atributos recopilados, realizaron un estudio de selección de atributos para seleccionar las más influyentes, quedando solamente 15 atributos de 670 alumnos: 610 aprobados y 60 suspendidos. Se realizaron cuatro experimentos usando 10 algoritmos de clasificación de tipo “caja-blanca” basada en reglas y en árboles de decisión con el objetivo de obtener la máxima exactitud de clasificación. Finalmente el proyecto de investigación demostró la utilidad de las técnicas de selección de características cuando se dispone de muchos atributos, consiguiendo mejorar la clasificación de los algoritmos al utilizar un conjunto reducido de 15 atributos de entre los 77 disponibles inicialmente y los resultados arrojados por los modelos obtenidos indican que los factores que más influyen en el fracaso escolar son: notas deficientes en Física I, Matemáticas I, Humanidades I e inglés; también la edad mayor a 15 años, el grupo al que asistió, el nivel regular de motivación, el no presentar el Taller de lectura y redacción, el vivir en una ciudad grande y el considerar que la asignatura más difícil sea matemáticas.

### 3 BASE DE DATOS EXPERIMENTAL

La información de la base de datos se obtuvo mediante una encuesta basada en la escala MEDMAR [7]. Esta escala contiene reactivos que solicitan datos como: edad, sexo, existencia de beca, número de hijo en la familia, entre otros datos. En general, el instrumento MEDMAR abarca indicadores familiares, escolares, de personalidad y perfil escolar, teniendo como fin identificar factores que afectan el rendimiento escolar. Uno de los objetivos de esta investigación, fue determinar el impacto de actividades cotidianas del estudiante en su rendimiento. Por tal razón, el instrumento de este trabajo fue complementado con reactivos como: horas dedicadas a la T.V., pasatiempos (hobby), deporte que practica, uso del celular, horas en internet, red social que frecuenta, horas dedicadas a la pareja, acoso escolar, importe gastado en pasajes, libros leídos por año, existencia de disciplina, entre los más importantes.

Se aplicaron encuestas a 208 jóvenes estudiantes, 112 hombres y 96 mujeres, en un rango de edad entre 16 y 23 años, que cursan en bachilleratos y de nuevo ingreso en la Universidad del Istmo de la Ciudad Ixtepec Oaxaca, México. A partir de esta información se creó la base de datos que contiene 35 atributos con datos como: número de integrantes en la familia, problemas familiares, pasatiempo, existe en el salón acoso escolar, dedica tiempo a las redes sociales, usa celular, realiza ejercicio, grado de estudios de los padres, gastos en transporte, tiempo dedicado a la tv, entre otros. Se crearon cuatro clases de estudiantes basados en sus promedios generales en la escala del 1 al 10; estudiantes con promedios iguales o mayores a 9.0 pertenecen al *alto* rendimiento o de clase 1, los de promedio igual o mayor a 8.0 pero menores a 9.0 pertenecen al *buen* rendimiento o de clase 2, los de promedio igual o mayor a 7.0 pero menores a 8.0 pertenecen al rendimiento *regular* o de clase 3 y debajo del 7.0 pertenecen al rendimiento *bajo* o de clase 4.

### 4 EXPERIMENTOS

En los experimentos se busca identificar la influencia que tienen factores cotidianos en el rendimiento académico de los estudiantes. En este proceso de identificación, se hace uso de la herramienta Weka 7.9, para efectuar actividades como: Selección de Atributos y Clasificación. Para contar con un referente de la importancia de la selección de atributos, se realizaron experimentos de clasificación usando todos los atributos derivados de las encuestas aplicadas a estudiantes, excluyendo únicamente el promedio que determina directamente la clase; se cuentan con 34 atributos por estudiante más el atributo clase.

**Tabla 1. Desempeño de algoritmos para clasificar tipos de estudiantes.**

Clasificador	% Efectividad
Random Forest	90.86
J48	90.86
Random Tree	90.86
Naive Bayes	81.25
Knn-1	90.86
Perceptron Multicapa	89.42
SMO	89.42

El proceso de clasificación se realizó usando validación cruzada de 10 dobleces, reportando los algoritmos que presentaron mejores resultados. Los resultados iniciales se muestran en la Tabla 1, se obtiene una clasificación aceptable

para las cuatro clases de estudiantes empleando 34 atributos; los clasificadores aquí presentados fueron los que obtuvieron mejor efectividad con respecto a otros disponibles en WEKA.

#### 4.1 SELECCIÓN DE ATRIBUTOS

La selección de atributos es una de las actividades realizadas en la fase de pre-proceso de la minería de datos, que consiste en identificar los atributos que mayor aporte tienen en la identificación de las instancias. En general, con la reducción de atributos se reduce la cantidad de tiempo de procesamiento y memoria requerida por los algoritmos de minería de datos, se permite visualizar más fácilmente los datos y se ayuda a eliminar ruido en los mismos [8]. Los algoritmos evaluadores tienen como función determinar en qué porcentaje el atributo es importante para la clasificación. En la Tabla 2 se muestran resultados de algoritmos evaluadores de atributos, se consideran aquellos atributos con una importancia del 90% o superior. En posteriores experimentos, se identificará a los atributos sugeridos por los algoritmos evaluadores: *Best First* como BF, *Rank Search* como RS y *Genetic Search* como GS.

Tabla 2. Algoritmos evaluadores y atributos seleccionados.

Algoritmo Evaluador	Atributos Propuestos
Best First	Edad, tv, gdomama, libros, pasaje, probfam, desercion
Rank Search	Edad, hobby, tv,gdomama, numhijo, pasaje, probfam, desercion
Genetic Search	Edad, hobby, tv, pasaje, desercion

Es importante hacer mención, que el atributo deserción se refiere a los deseos de desertar del estudiante, puede tomar tres valores: 1. *Si deseo desertar, es mejor conseguirme un trabajo*, 2. *Ocasionalmente deseo desertar* y 3. *No deseo desertar*, aún puedo continuar. Podría considerarse que este atributo provee información que directamente decide el rendimiento académico del estudiante, sin embargo no es así, puesto que el 56.52% de estudiantes de *alto rendimiento escolar* han seleccionado el valor 2, así como el 36.06% de estudiantes de *rendimiento regular*; solo un 26% de los estudiantes de *bajo rendimiento* han seleccionado el valor 1. Se puede observar que este atributo no se encuentra directamente relacionado con el rendimiento del estudiante.

#### 4.2 CLASIFICACIÓN DE ESTUDIANTES

En este apartado, se experimentará con la clasificación de los estudiantes aplicando los algoritmos empleados en la Tabla 1Tabla 1. Se experimenta con los atributos propuestos por BF, esto es, con siete de los 35 atributos existentes; en la segunda columna de la Tabla 3 se observa un desempeño aceptable para clasificadores de árbol como *Random Forest*, y *Random Tree*, así como para el *Knn*. El desempeño obtenido es igual al obtenido usando los 35 atributos. Los algoritmos *J48* y *Perceptrón Multicapa* presentan ligeramente menor efectividad. El resto de los clasificadores presentan una efectividad mucho menor.

Tabla 3. Desempeño de la Clasificación usando atributos BF, RS y GS.

Clasificador	% Efectividad con atributos BF	% Efectividad con atributos RS	% Efectividad con atributos GS
Random Forest	90.86	90.86	90.86
J48	89.42	89.42	89.42
Random Tree	90.86	90.86	90.86
Naive Bayes	77.40	74.03	87.01
Knn	90.86	90.86	90.86
Perceptron Multicapa	88.46	89.90	84.13
SMO	74.03	75.96	61.05

En la tercer columna de la Tabla 3 se observa la efectividad de los clasificadores usando los ocho atributos propuestos por RS. Hasta el momento se ha observado que el impacto negativo de usar una menor cantidad de atributos, propuestos por los selectores, ha sido casi nulo para la mayoría de los clasificadores experimentados. El selector GS propone a cinco atributos para clasificar el tipo de estudiante, presentando en la cuarta columna de la Tabla 3 la efectividad de los clasificadores

usando estos atributos, donde nuevamente los algoritmos *Random Forest*, *Random Tree* y *KNN* tienen un desempeño similar a usar los 35 atributos. La efectividad para clasificar usando cinco atributos, es ligeramente menor a la presentada en la Tabla 1 donde se usan los 35 atributos. El único clasificador afectado considerablemente fue SMO de un 89.42% a un 61.05% de efectividad. En general, se observa que los atributos seleccionados por los algoritmos evaluadores mantienen una efectividad similar en la clasificación, con respecto al uso de todos los atributos de la base de datos. Los algoritmos *Random Tree*, *Random Forest*, *J48* y *el KNN* son los que mejor estabilidad presentan en las pruebas con los atributos seleccionados por diversos algoritmos evaluadores.

**Tabla 4. Efectividad de la clasificación con atributos BF, RS y GS omitiendo el atributo edad.**

Clasificador	% Efectividad BF	% Efectividad RS	% Efectividad GS
Random Forest	90.86	90.86	89.42
J48	89.42	89.90	89.42
Random Tree	90.86	90.86	89.42
Knn-1	90.86	90.86	89.42

Con el fin de verificar la importancia en la clasificación de los atributos de la Tabla 2, se realizaron pruebas de clasificación omitiendo alguno de ellos. Se encontró que únicamente al eliminar el atributo edad el impacto en la efectividad de clasificación fue mínimo, ver la Tabla 4. Se observa que al clasificar con los atributos BF y RS sin el atributo edad el impacto fue nulo. La clasificación que usa los cinco atributos GS es afectada de manera mínima; puesto que la cantidad de atributos es pequeña la eliminación de algún atributo puede afectar negativamente el desempeño de los clasificadores; en este caso impacta negativa y aproximadamente en 1.44%.

## 5 TENDENCIAS DE LOS ATRIBUTOS EN LA BASE DE DATOS

En esta sección se hace una revisión de las tendencias observadas en la base de datos. Se analizaron las cualidades compartidas por los que tienen la nota mínima de 6. Se analizaron tendencias de factores entre las cuatro clases de estudiantes, así como la comparativa de información perteneciente a dos grupos de estudiantes, aquellos que tienen nota de 8 o superior y aquellos que tienen nota inferior al 8.

### 5.1 ESTUDIANTES DE RENDIMIENTO MÍNIMO

Existen únicamente 8 de 208 encuestados (3.84%) que tienen la nota mínima de 6.0, teniendo las siguientes características en común:

- Consideran que es mejor conseguir un trabajo que seguir estudiando, esta forma de pensar está correlacionada con la nota mínima 6.0.
- Reciben acoso escolar (bullying): es conocido que el acoso escolar o bullying es uno de los factores que afecta el rendimiento escolar. Estudios al respecto afirman que existen mayores probabilidades de obtener bajos logros entre los estudiantes que han sido víctimas de bullying [9]; otros estudios afirman que la relación entre bullying y rendimiento escolar a través de un análisis multinivel, denota que aquellos estudiantes que reportaban ser víctimas de abuso por parte de sus compañeros obtenían inferiores rendimientos en matemáticas y lectura [10].
- Ninguno practica deporte: Se ha demostrado que los deportes y las actividades físicas elevan los niveles de alerta física y mental, aumentando la función intelectual de los participantes [11]. Por otro lado, en [12] se menciona que “La actividad física reduce el estrés, la ansiedad y la depresión, e incrementa el aprendizaje y la memoria; todos estos factores conducen a un estado óptimo de salud y a un mayor rendimiento académico”.
- Carencia de facilidad de palabra, en promedio las habilidades del habla están relacionadas con las intelectuales. Los estudiantes con nota mínima no cuentan con facilidad de palabra. Este hecho ha sido estudiado en [13], en este trabajo se afirma que “los déficits del procesamiento auditivo deben estar relacionados a dificultades académicas con la comprensión auditiva. Los estudiantes deben tener dificultades con la interpretación de lecturas, pobre entendimiento de instrucciones orales y del aprendizaje de una lengua foránea”; el sistema de procesamiento auditivo procesa la capacidad de percibir y analizar sonidos, así como producirlos oralmente.
- Ninguno tiene hermanos, son hijos únicos. En lo que respecta a esta tendencia de la base de datos experimental, no se han encontrado estudios relacionados.

En general, se observan factores que se encuentran relacionados o reflejan aspectos emocionales negativos del estudiante tales como: Acoso escolar, carencia de ejercicio y facilidad de palabra.

### 5.2 TENDENCIAS ENTRE CLASES DE ESTUDIANTES

En la base de datos experimental, se hallaron tendencias entre las cuatro clases de estudiantes, en atributos como: hobby, horas de tv y acoso escolar. En la Figura 1 se observa que la tendencia por el gusto a bailar está relacionada con la nota del estudiante. El rendimiento del estudiante es directamente proporcional al gusto por el hobby de bailar. La tesis de Spaeth sobre cultura escolar sostiene que “un modo de vida encaminado hacia la cultura desarrolla habilidades cognitivas, capacidades y gustos aumentan el rendimiento escolar, por tal motivo, animan y permiten al joven mejorar en su desarrollo escolar” [14]. Concretamente, en [15] se sostiene que “el aprendizaje y la retención de la información es más efectiva cuando niños están activos, escuchando y expresando ideas físicamente en el espacio”, en general se menciona la importancia de la danza en las escuelas.

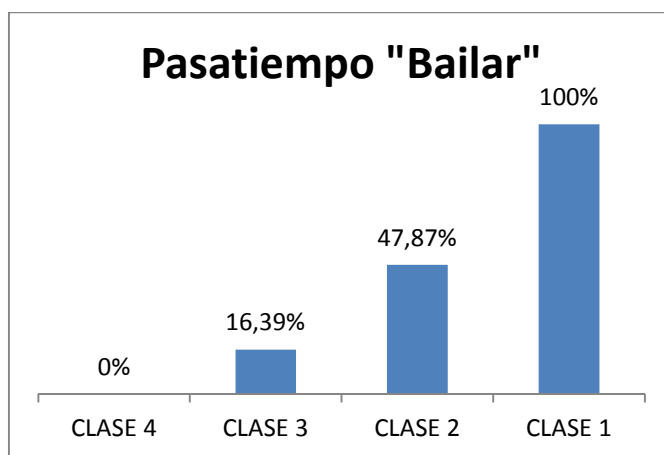


Figura 1. Tendencia del gusto por bailar

Los estudiantes de peor rendimiento ven en mayor porcentaje más de dos horas diarias TV, en contraste con los de mejor rendimiento, quienes reportaron ver diariamente menos de 2 horas de TV al día. El rendimiento del estudiante es inversamente proporcional al tiempo gastado en ver la TV, ver Figura 2. En la Figura 3 se observa la tendencia del acoso escolar entre las clases de estudiantes. El rendimiento académico es inversamente proporcional al acoso escolar. Estudiantes que tienen alto rendimiento tienen baja incidencia de acoso escolar, sin embargo, todos los estudiantes de bajo rendimiento reportaron recibir acoso escolar.

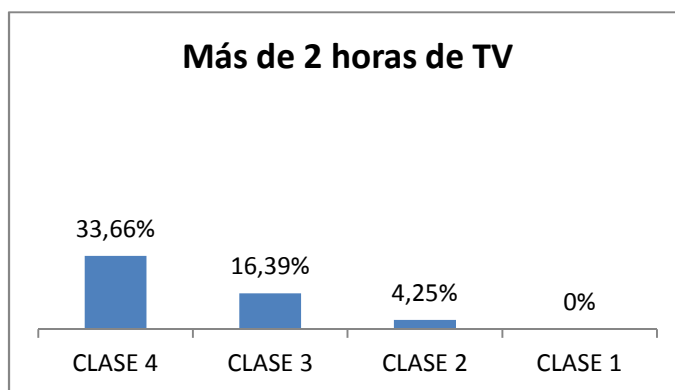


Figura 2. Tendencia de los estudiantes que ven más de do horas de TV

En el estudio realizado por Raychauduri se menciona que, “factores socioeconómicos como la asistencia a clases, ingreso familiar, nivel de educación de los padres, cercanía al profesor, profesores capacitados, sexo de los estudiantes y la distancia a la escuela, afectan el rendimiento de los estudiantes” [16]. En los resultados del presente trabajo se han encontrado

factores en común al estudio de *Raychaudhuri* como el costo para transportarse a la escuela y el nivel de educación de la madre, tal como se muestra en la Tabla 2.

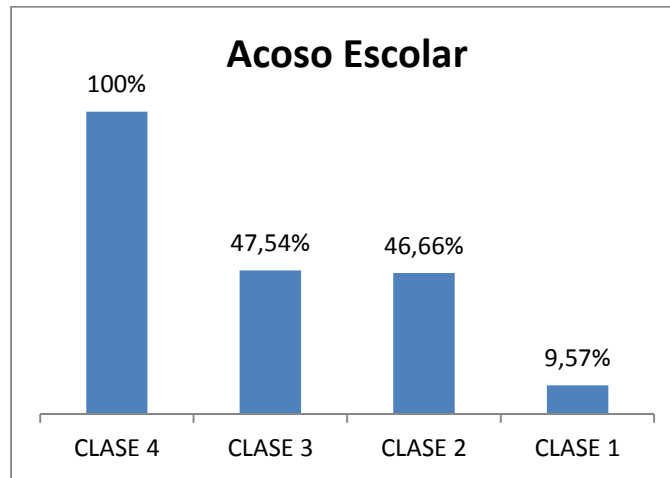


Figura 3. Tendencia del acoso escolar

5.3 TENDENCIAS ENTRE ESTUDIANTES DE ALTO Y BAJO RENDIMIENTO

En este análisis y con el fin de visualizar de una manera abstracta las tendencias entre los estudiantes, se discretizaron las cuatro clases existentes en sólo 2. Los estudiantes con promedio de 8.0 o superior integran en la clase “Alto rendimiento escolar”, y el resto en la clase “Bajo rendimiento escolar”. Las tendencias se observan gráficamente en la Figura 4, donde los porcentajes indicados no son necesariamente complementarios.

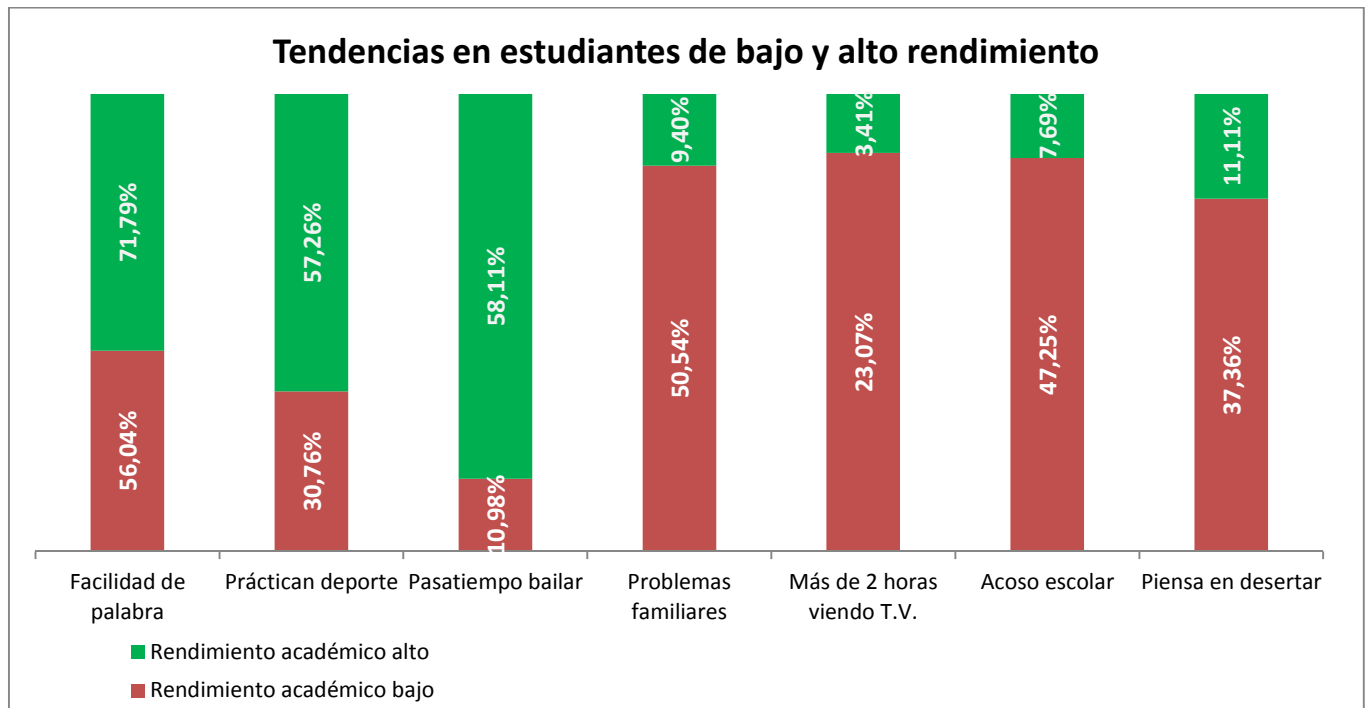


Figura 4. Tendencia del alto y bajo rendimiento académico.

La fluencia o facilidad de palabra, la práctica de deporte y el gusto por el baile son factores que predominan en los estudiantes de alto rendimiento. Por otro lado, los problemas familiares, el acoso escolar, ver la TV por más de 2 horas y deseos de dejar de estudiar son factores que predominan en los estudiantes de bajo rendimiento. En La barra correspondiente a *Piensa en desertar*, se ha considerado únicamente la opción 1. *Si deseo desertar, es mejor conseguirme un trabajo.*

## 6 DISCUSIÓN Y TRABAJO FUTURO

De acuerdo los resultados obtenidos en esta investigación y de acuerdo a la información disponible en la base de datos experimental, se mencionan los factores y patrones observados. En este trabajo, fue posible determinar el rendimiento escolar basándose en información como: el hobby, horas diarias de TV, el importe que pagan de pasajes, los deseos de desertar, problemas familiares, grado de estudios de la mamá, el número de hijo en la familia y el número de libros que lee al año. La precisión de la clasificación fue similar con cada una de las tres combinaciones de atributos, mostradas en la Tabla 2Tabla 2, alcanzando un máximo de 90.86% de precisión. Es posible realizar una clasificación cercana a 90% usando únicamente el hobby, horas diarias de TV, el importe que pagan de pasajes y los deseos de desertar. Los atributos que presentan claramente tendencias entre las cuatro clases fueron: El acoso escolar (bullying), más de dos horas de ver la TV y el gusto por el pasatiempo (hobby) de bailar. Factores presentes en estudiante de bajo nivel, citados de mayor a menor impacto fueron: Más de dos horas diarias de ver TV, acoso escolar, problemas familiares y deseos de desertar. Existe una correlación entre el bajo rendimiento, el acoso escolar, los deseos de desertar y la falta de facilidad de palabra, esto de acuerdo a los estudios realizados sobre los estudiantes con el mínimo rendimiento. Los factores presentes en estudiantes de alto nivel fueron: La práctica de deporte, practicar el pasatiempo bailar y la facilidad de palabra.

Cómo trabajo futuro se incrementará el número de instancias en la base de datos así como la dimensionalidad de la misma, buscando abarcar más factores o actividades de los estudiantes. Se estudiará la aplicación y combinación de algoritmos evaluadores en varias fases, buscando obtener mejoras en el filtrado de atributos. Por otro lado, se buscarán reglas de asociación en la base de datos de alto soporte y confianza. Se compararán los resultados futuros y los obtenidos en este trabajo. Por último se recomienda para investigaciones relacionadas, no incluir atributos con información que obvie el rendimiento académico, tal como número de materias aprobadas, número de créditos asignados, calidad de la clase recibida, por citar algunos, ya que estos proveen directamente información del rendimiento del estudiante.

## REFERENCIAS

- [1] ENLACE, «Evaluación Nacional de Logro Académico en Centros Escolares,» 2014. [En línea]. Available: [http://enlace.sep.gob.mx/content/ms/docs/2014/estadisticas\\_de\\_resultados/1.-NVL\\_DOM\\_NAC\\_2008-2014.xls](http://enlace.sep.gob.mx/content/ms/docs/2014/estadisticas_de_resultados/1.-NVL_DOM_NAC_2008-2014.xls). [Último acceso: 12 03 2015].
- [2] Instituto Nacional de Estadística, Geografía e Informática., «<http://cuentame.inegi.org.mx/poblacion/asistencia.aspx?tema=P>,» INEGI, 2010. [En línea]. Available: <http://cuentame.inegi.org.mx/poblacion/asistencia.aspx?tema=P>. [Último acceso: 2015 12 03].
- [3] R. Baker, Data Mining for Education, E. B. a. B. M. Penelope Peterson, Ed., Pittsburgh, Pennsylvania: Elsevier Ltd., 2011, pp. 112-118.
- [4] M. A. Nahid H., «Analysis of factors that affect students' academic performance - Data Mining Approach,» *International Journal of advanced studies in Computer Science and Engineering*, pp. 1-4, 2014.
- [5] P. Cheewaprakobkit, «Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program,» *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. I, pp. 332-336, 2013.
- [6] R. C. V. S. Vera C., «Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos,» *IEEE RITA*, pp. 109-117, 2012.
- [7] L. Medina y F. Martell, «“Evaluación Integral del rendimiento escolar en educación superior. Instrumentos de evaluación,» 1° Reunión Internacional de Veracruz Evaluación en Educación Media Superior y Superior, 2008.
- [8] I. e. a. Witten, Data Mining, Practical Learning Tools and Techniques., Morgan Kauffman Publishers, 2011.
- [9] G. e. a. Glew, «Bullying, psychosocial adjustment and academic performance in elementary school,» *Archives of pediatrics & adolescent medicine*, vol. 159, nº 11, pp. 1026-1031, 2005.
- [10] C. e. a. Konishi, «Do school bullying and student-teacher relationships matter for academic achievement? A multilevel analysis,» *Canadian Journal of School Psychology*, vol. 25, nº 1, pp. 19-39, 2010.
- [11] A. Hills, «Scholastic and Intellectual Development and Sport,» *Sports and Children*, pp. 76-90, 1998.



- [12] R. Bailey, «Physical Education and Sport in Schools: A review of the Benefits and Outcomes,» *Journal of School Health*, vol. 76, nº 8, pp. 397-401, 2006.
- [13] R. e. a. Wagner, «Region 13, education service center,» Austin, 1999.
- [14] J. Spaeth, «Characteristics of the Works Settings and the Job as Determinants of Income,» *In Schooling and Achievement in American Society*, 1976.
- [15] D. J. Kaufmann K., *Dance Integrations: 36 Dance Lessons Plans for Science and Mathematics*, Montana: PA Hutchison Co., 2014.
- [16] e. a. Raychaudhuri, «Factors affecting Student's Academic Performance: A case study in Agartala municipal concial area,» *Bangladesh e-Journal of Sociology*, vol. 7, nº 2, 2010.