

## Técnicas de clasificación para predecir el desempeño de los estudiantes en pruebas estandarizadas

## Classifying techniques for predicting how students perform on standardized tests

**Roberto Contreras Juárez**

*Benemérita Universidad Autónoma de Puebla.*

*Facultad de Ciencias de la Computación.*

*Puebla, Puebla, México.*

[roberto.contreras@correo.buap.mx](mailto:roberto.contreras@correo.buap.mx)

<https://orcid.org/0000-0002-3271-6754>

**Carmen Cerón Garnica**

*Benemérita Universidad Autónoma de Puebla.*

*Facultad de Ciencias de la Computación.*

*Puebla, Puebla, México.*

[carmen.ceron@correo.buap.mx](mailto:carmen.ceron@correo.buap.mx)

<https://orcid.org/0000-0001-6480-6810>

**Etelvina Archundia Sierra**

*Benemérita Universidad Autónoma de Puebla.*

*Facultad de Ciencias de la Computación.*

*Puebla, Puebla, México.*

[etelvina.archundia@correo.buap.mx](mailto:etelvina.archundia@correo.buap.mx)

<https://orcid.org/0000-0001-9686-5305>

**Karen Josefina Rivera Torres**

*Benemérita Universidad Autónoma de Puebla.*

*Facultad de Ciencias de la Computación.*

*Puebla, Puebla, México.*

[karen.riveratorres@alumno.buap.mx](mailto:karen.riveratorres@alumno.buap.mx)

<https://orcid.org/0009-0000-8439-8543>

## **Técnicas de clasificación para predecir el desempeño de los estudiantes en pruebas estandarizadas**

### **Classifying techniques for predicting how students perform on standardized tests**

#### **Resumen.**

Los exámenes de ingreso administrados por las instituciones educativas son uno de los principales desafíos que enfrentan los estudiantes que desean continuar su educación. Estos exámenes son pruebas estandarizadas que miden las habilidades y conocimientos de los estudiantes, lo que ayuda a seleccionar estudiantes que tienen más probabilidades de tener éxito en sus estudios futuros. Este trabajo de investigación tiene como propósito primordial realizar un análisis cuantitativo y cualitativo de las calificaciones finales de Español, Matemáticas, Ciencias Naturales y Sociales de estudiantes egresados de secundaria para determinar patrones que ayuden a predecir su desempeño en el examen de ingreso al bachillerato. El desempeño se mide con base al nivel de habilidad/conocimiento  $\theta$  obtenido por los estudiantes en la prueba. El análisis se realiza sobre una población de 7,118 estudiantes utilizando técnicas de agrupamiento y clasificación de WEKA y ha demostrado que matemáticas es la base principal para un mejor desempeño.

**Palabras clave:** Clasificación, Desempeño académico, Minería de datos, Predicción.

#### **Abstract**

Entrance exams administered by educational institutions are one of the major challenges faced by students who wish to continue their education. These exams are standardized tests that measure students' skills and knowledge, which helps to select students who are more likely to succeed in their future studies. The main purpose of this research is to conduct a quantitative and qualitative analysis of the final scores of Middle School graduates in Spanish, Mathematics, Natural Sciences and Social Sciences in order to identify patterns that help predict their performance on the High School entrance Exam. Performance is measured based on the level of skill/knowledge  $\theta$  obtained by students in the test. The analysis is performed on a population of 7,118 students using WEKA clustering and classification techniques and has shown that mathematics is the primary basis for better performance.

**Keywords:** Classification, Academic performance, Data mining, Prediction.

## Introducción

### *Situación problemática*

La minería de datos se refiere al descubrimiento de conocimiento en bases de datos y consiste básicamente en la extracción de información implícitamente oculta en los datos, Piotesky *et al.* (1991). Así pues, la minería de datos puede considerarse una metodología de análisis de datos utilizada para identificar patrones inicialmente desconocidos ocultos en conjuntos de datos.

La minería de datos se ha utilizado en muchos campos, especialmente en el ámbito comercial, pero en los últimos años ha cobrado gran interés en el ámbito educativo. La minería de datos aplicada a la educación se conoce como minería de datos educativos (EDM por sus siglas en inglés), que se encarga de extraer información útil de las bases de datos de las instituciones educativas mediante la aplicación de herramientas y técnicas de minería de datos para obtener una mejor comprensión o un mayor conocimiento, Al-Razgan *et al.* (2014).

Las técnicas de minería de datos aplicadas en entornos educativos han demostrado ser herramientas eficaces para predecir el rendimiento académico de los estudiantes, lo que ha permitido identificar los factores que más influyen en su aprendizaje, ayudando así a los docentes a mejorar el proceso de enseñanza, realizando acciones pedagógicas más eficientes y oportunas, Menacho (2017).

Como es sabido, la educación es uno de los factores más importantes para el desarrollo de un país, esto obliga a las instituciones educativas a considerar como su objetivo primordial el ofrecer programas educativos de calidad y una forma efectiva de lograrlo es a través de la evaluación del desempeño o rendimiento académico de los estudiantes. Por supuesto, este

proceso de evaluación es bastante complejo, ya que no puede limitarse a una simple calificación, pues existen muchos factores que pueden incidir en las buenas o malas calificaciones y conocerlos ayuda a los responsables de los procesos de enseñanza–aprendizaje a planear y personalizar sus programas educativos con base en la información recibida.

La EDM es uno de los enfoques de minería de datos que puede ayudar eficazmente a descubrir las complejas relaciones que se esconden tras las calificaciones. El diagrama de la Tabla 1 explica las diferentes etapas de la metodología para extraer información de los datos educativos para la toma de decisiones académicas.

**Tabla 1.**

*Metodología EDM*



Fuente: *Figura 1 en Durairaj (2014).*

En este trabajo se aplican algunas de las técnicas de clasificación en minería de datos para buscar predecir, a partir de las calificaciones finales en Español, Matemáticas, Ciencias Naturales y Ciencias Sociales, el desempeño de los estudiantes en una prueba estandarizada utilizada como examen de ingreso al nivel medio superior (bachillerato) de la Benemérita Universidad Autónoma de Puebla. La prueba estandarizada consta de 120 reactivos divididos en un componente de habilidades del pensamiento y cuatro

componentes de conocimiento–razonamiento: Español, Matemáticas, Ciencias Naturales y Ciencias Sociales.

En las pruebas estandarizadas es común utilizar dos formas para determinar el poder discriminativo de un ítem: el *índice de discriminación* y el *coeficiente de discriminación*. El índice de discriminación considera sólo el 54% de los encuestados, 27% con el rendimiento más alto y el 27% con el rendimiento más bajo. Por otro lado, “el coeficiente de discriminación se calcula mediante el coeficiente de correlación del punto biserial,  $r_{bis}$ , y considera a todos y cada uno de los encuestados”, Backhoff *et al.* (2020).

Tanto el índice de discriminación como el coeficiente de discriminación permiten determinar el poder discriminatorio de un ítem y comprobar si los alumnos "adecuados" son los que obtienen las respuestas correctas, cuánto poder predictivo tiene el ítem y cómo puede contribuir a las predicciones, Henrysson (1971).

### ***Antecedentes***

Existe un gran número de trabajos que utilizan la minería de datos educativos para predecir el rendimiento de los alumnos que toman en cuenta distintos factores. Entre las técnicas de minería de datos utilizadas se encuentran los algoritmos de clasificación y el descubrimiento de reglas de agrupamiento o clustering. En el caso de los algoritmos de clasificación, muchos investigadores han señalado que los algoritmos de árboles de decisión son bastante útiles para predecir el rendimiento académico de los estudiantes, Kaur *et al.* (2018).

Según Kabakchieva (2013), “las técnicas de minería de datos se utilizan cada vez más en las universidades para analizar datos educativos con el fin de extraer información y conocimientos que sirvan de apoyo a la toma de decisiones educativas”. Los diversos estudios que se han realizado sobre el

rendimiento académico han identificado varios factores o atributos que podrían afectar al rendimiento de los estudiantes.

Ketui *et al.* (2019), utiliza cinco modelos de clasificación para identificar la asignatura adecuada para los estudiantes de ciencias y, a continuación, comparan el rendimiento de cada uno de los modelos utilizando cuatro medidas: Precision, Recall, F-measure y Accuracy Scores.

Asimismo, Hamsa *et al.* (2016) examina las capacidades de las técnicas de minería de datos en el campo de la educación superior proporcionando un modelo que utiliza técnicas de clasificación para evaluar el rendimiento de los estudiantes basándose en las notas parciales, las notas finales y las puntuaciones de admisión.

Dole *et al.* (2015), aplica el algoritmo de Naive Bayes y el árbol de decisión para predecir la graduación de los estudiantes y la calificación final (aprobado y reprobado) sobre una colección de datos entre el primer año y los datos tomados durante la escuela secundaria, teniendo en cuenta 11 atributos.

Sivasakthi (2017), centra su atención en la predicción del rendimiento en programación introductoria de los estudiantes de primer año de licenciatura en el curso de Aplicación Informática mediante un modelo predictivo de minería de datos que utiliza algoritmos basados en la clasificación. Los datos recogidos contienen los datos demográficos de los estudiantes, la calificación en programación introductoria en la universidad, y la calificación en programación introductoria en la prueba que contiene 60 preguntas. Los datos recogidos se aplicaron en varios algoritmos de clasificación como Percepción Multicapa, Naive Bayes, SMO, J48 y REPTree utilizando WEKA.

Meghji *et al.* (2019), utiliza las características de comportamiento como la toma de notas, la atención, la entrega de tareas y la posición de asiento

extraídas de datos reales de estudiantes para predecir el rendimiento de los estudiantes. Varios clasificadores mostraron un buen rendimiento en términos de precisión y valores kappa.

Un árbol de decisión representa una estructura jerárquica formada por un conjunto de condiciones. Las nuevas instancias de datos se clasifican en clases según el camino de las condiciones satisfechas hasta que se alcanza un nodo hoja, que representa una etiqueta de clase, Frank *et al.* (2016).

En el caso de los algoritmos de clasificación, muchos investigadores han implementado varios algoritmos como ID3, J48/C4.5, RandomTree, MultilayerPerceptron y RandomForest utilizando la herramienta WEKA, Kaur (2018).

Los métodos de clasificación difieren en su mecanismo interno para procesar y extraer características relevantes de los datos de entrenamiento para construir un modelo de clasificación. Un árbol de decisión, como J48 y REPTree de WEKA, representa una estructura jerárquica de nodos y hojas.

Cada nodo representa una prueba lógica, y dependiendo del resultado de la prueba, el nodo se ramifica a uno u otro nodo hijo. Las nuevas instancias de datos se clasifican en clases basándose en la ruta de las condiciones satisfechas hasta que se alcanza un nodo hoja; “el nodo hoja representa una etiqueta de clase”, Frank *et al.* (2016).

Por último, el clustering es un proceso que divide los datos en grupos de objetos similares. La agrupación desempeña un papel importante en aplicaciones de minería de datos como la recuperación de información y la minería de textos, entre muchas otras. Según Sembiring (2012), “el algoritmo *k*-means es probablemente el mejor de los algoritmos de clustering existentes”, su objetivo es elegir el mejor centro de clúster que

será el centroide, pero requiere que los atributos numéricos sean cambiados a nominales.

Por supuesto, estos estudios e investigaciones tienen un gran potencial para proporcionar información útil a la hora de aplicar políticas académicas adecuadas para prevenir el abandono y el fracaso escolar. Los educadores han utilizado clasificadores para predecir diversas facetas del aprendizaje de los alumnos.

### ***Objetivo(s)***

El objetivo general de este trabajo es predecir el desempeño en el examen de ingreso de los alumnos que desean ingresar al nivel medio superior, identificando áreas de oportunidad para fortalecer aquellas asignaturas que más apoyen al logro de un mejor desempeño al momento de enfrentar el examen.

Los objetivos específicos son:

1. Aplicar algunos de los métodos de clasificación que ofrece WEKA para predecir el rendimiento en la prueba de ingreso que realizan los estudiantes que desean acceder al nivel medio superior.
2. Aplicar el algoritmo  $k$ –means clustering para confirmar que el agrupamiento sugerido por el índice de discriminación es el adecuado.
3. Identificar las características académicas que conducen a un mejor desempeño en el examen de ingreso.

## **Materiales y Métodos**

### ***Clasificación de la investigación***

Para alcanzar el objetivo de este trabajo, el análisis cuantitativo y cualitativo de los datos se llevó a cabo aplicando minería de datos con el paquete de

software WEKA (Waikato Environment for Knowledge Analysis), que es un entorno de experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas de análisis de datos más relevantes, principalmente las procedentes del aprendizaje automático, sobre cualquier conjunto de datos de usuario, Frank *et al.* (2016) y Hall *et al.* (2009).

Teniendo en cuenta el análisis realizado para obtener el índice de discriminación de los ítems, los alumnos que realizaron el examen de ingreso se agrupan en tres categorías:

- **Alto:** corresponde al 27% de los alumnos que tuvieron un rendimiento alto.
- **Bajo:** corresponde al 27% de los alumnos que tuvieron un rendimiento bajo.
- **Medio:** aquellos alumnos que no pertenecen ni al grupo alto ni al bajo.

En este sentido, se aplican los algoritmos de clasificación incluidos en WEKA, J48, NaiveBayes, MultilayerPerceptron y RandomForest, con la única finalidad de asegurar la validez del estudio y así obtener una predicción significativa.

Además, como indica Meghji *et al.* (2019), es “fundamental evaluar el rendimiento y la utilidad de los clasificadores antes de utilizar los resultados para predecir y/o tomar decisiones sobre las estrategias a implementar en los procesos de enseñanza-aprendizaje de los alumnos”.

Por esta razón, los clasificadores se evaluaron utilizando las siguientes medidas: Precision, Recall, F-measure, Accuracy scores and Kappa statistic.

- **Precision.** Se entiende como la medida de los casos positivos correctamente identificados de todos los casos positivos predichos. Por lo tanto, es útil cuando el coste de los falsos positivos es elevado.

$$\text{Precision} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}}$$

- **Recall.** Es la medida de los casos positivos correctamente identificados de todos los casos positivos reales. Es importante cuando el coste de los falsos negativos es elevado.

$$\text{Recall} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}}$$

- **F-measure.** Es la media armónica de Precision y Recall y da una mejor medida de los casos clasificados incorrectamente que la métrica Accuracy.

$$\text{F-measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Accuracy.** Es la medida de todos los casos correctamente identificados. Es la más utilizada cuando todas las clases tienen la misma importancia.

$$\text{Accuracy} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Total datos}}$$

- **Kappa statistic.** Es el coeficiente estadístico propuesto por Cohen que permite medir la concordancia entre los resultados de dos o más variables cualitativas, Cohen (1960).

### ***Datos***

Para este trabajo de investigación, se utilizaron los datos de 7,118 alumnos que presentaron el examen de ingreso al nivel medio superior que ofertó la Benemérita Universidad Autónoma de Puebla en el año 2022.

Se tomaron en cuenta las calificaciones finales de secundaria en Español, Matemáticas, Ciencias Naturales y Ciencias Sociales, así como el nivel de desempeño alcanzado en el examen de ingreso. Es importante señalar que sólo se tomaron en cuenta los datos académicos; no fueron considerados datos sensibles como nombre, ciudad de origen, escuela, nivel socioeconómico, entre otros.

Las calificaciones finales son números enteros, donde el mínimo es 6 y el máximo es 10. Por otro lado, el desempeño del estudiante en la prueba de ingreso se midió utilizando el parámetro  $\theta$ , que corresponde a un valor numérico en el intervalo cerrado  $[-3,+3]$ , e indica el nivel de habilidad/conocimiento que posee un alumno, Baker *et al* (2017).

Para un mejor manejo, el desempeño del alumno,  $\theta$ , se agrupó en tres categorías, a saber,

- **Alto.** Estudiantes con un desempeño  $\theta > 0.6966$ .
- **Medio.** Estudiantes con un desempeño  $-0.5447 \leq \theta \leq 0.6966$ .
- **Bajo.** Estudiantes con un rendimiento  $\theta < -0.5447$ .

La Figura 1 muestra un ejemplo de los datos considerados en esta investigación.

**Figura 1.**

*Muestra de la Base de Datos*

NMS_Calif.arff					
Relation: NMS_Calif					
No.	1: Span Numeric	2: Math Numeric	3: Nat Numeric	4: Social Numeric	5: Level Nominal
1	9.0	9.0	8.0	8.0	High
2	8.0	7.0	8.0	9.0	Medium
3	9.0	8.0	8.0	8.0	High
4	6.0	7.0	7.0	7.0	Low
5	9.0	10.0	8.0	8.0	High
6	8.0	9.0	9.0	8.0	High
7	8.0	8.0	9.0	9.0	High
8	8.0	8.0	8.0	8.0	High
9	6.0	7.0	7.0	7.0	Low
10	9.0	7.0	8.0	8.0	High
11	7.0	7.0	7.0	7.0	Low
12	7.0	8.0	7.0	8.0	Medium
13	8.0	7.0	8.0	7.0	Low
14	8.0	8.0	9.0	8.0	High
15	7.0	8.0	9.0	7.0	Medium
16	6.0	7.0	6.0	7.0	Low
17	8.0	7.0	8.0	9.0	Medium
18	7.0	7.0	7.0	8.0	Medium
19	7.0	8.0	7.0	7.0	Medium
20	7.0	7.0	7.0	8.0	Low

Fuente: *Elaboración propia usando WEKA*

La simbología utilizada se describe a continuación

- Span: calificación final en la asignatura Español.
- Math: nota final de la asignatura Matemáticas.
- Nat: nota final de la asignatura Ciencias Naturales.
- Social: nota final de la asignatura Ciencias Sociales.
- Level: grupo de rendimiento en la prueba de acceso.

### ***Agrupación y clasificación***

Para confirmar que la categorización del desempeño era adecuada, se procedió a la agrupación de los datos. La agrupación de datos es el proceso de dividir los datos en grupos con características similares. El algoritmo utilizado para agrupar los datos fue *k*-means clustering, que está incluido en WEKA.

El algoritmo de agrupación *k*-means es un procedimiento para agrupar un conjunto de elementos según un criterio, normalmente la distancia, mientras que el algoritmo NaiveBayes asume que, dada una variable de clase, el valor de una característica particular no está relacionado con la presencia o ausencia de otras características, Durairaj *et al.* (2014).

Por otro lado, la clasificación es la técnica más utilizada en la minería de datos. Implica la creación de un modelo con la clasificación de un conjunto de datos mediante la aplicación de algoritmos de clasificación incluidos en WEKA, Frank *et al.* (2016).

En los métodos de clasificación, es imposible ignorar los criterios de evaluación de los clasificadores. Estas evaluaciones ayudan a estimar la bondad de un clasificador y se conocen como proceso de validación, permitiendo medir eficazmente la capacidad predictiva del modelo generado a partir del clasificador.

Una de las alternativas para verificar la bondad de los clasificadores utilizados es la matriz de confusión. Esta matriz permite visualizar, a través de la tabla de contingencia, la distribución de los errores cometidos por el clasificador, tal y como se observa en la Tabla 2.

**Tabla 2.**

*Matriz de confusión*

	Clase 1	Clase 2	Clase 3
Clase 1	Correctos	Error 1 a 2	Error 1 a 3
Clase 2	Error 2 a 1	Correctos	Error 2 a 3
Clase 3	Error 3 a 1	Error 3 a 2	Correctos

Fuente: *Elaboración propia*

Otra alternativa utilizada es el coeficiente  $\kappa$  (Kappa), el coeficiente estadístico propuesto por Cohen que permite medir la concordancia entre los resultados de dos o más variables cualitativas, Cohen (1960).

El índice  $\kappa$ , aplicado a la matriz de confusión, permite evaluar si la clasificación observada es consistente con la clasificación predicha por el clasificador. La Tabla 3 muestra los índices  $\kappa$  para datos categóricos, Landis *et al.* (1977).

**Tabla 3.**

*Valores del índice  $\kappa$ .*

Valor	Intensidad
$\kappa \leq 0.0$	Pobre
$0.0 < \kappa \leq 0.2$	Leve
$0.2 < \kappa \leq 0.4$	Justo
$0.4 < \kappa \leq 0.6$	Moderado
$0.6 < \kappa \leq 0.8$	Sustancial
$0.8 < \kappa \leq 1.0$	Casi perfecto

Fuente: *Página 165, Landis et al. (1977)*

Por último, los clasificadores se evalúan comparando las medidas: Precision, Recall, F-measure, Accuracy Scores and Kappa Statistic.

## Resultados y Discusiones

### *Resultados.*

Aunque WEKA es una potente herramienta tanto para desarrollar algoritmos de clasificación y filtrado como para preprocesar los datos de forma que tengan una estructura adecuada. En este trabajo se utilizan datos ya formateados, por lo que no es necesario preprocesarlos.

El primer algoritmo aplicado fue  $k$ -means clustering, la finalidad fue corroborar la agrupación propuesta del desempeño mostrado por los estudiantes en el examen de ingreso. Los clústeres generados por el algoritmo de agrupación con  $k = 3$ , así como sus centroides, se muestran en la Tabla 4.

**Tabla 4.**

### *Clústeres y centroides*

Atributo	Datos	Clúster 1	Clúster 2	Clúster 3
Span	7.63	8.21	7.65	7.02
Math	7.63	8.44	7.49	7.05
Nat	7.61	8.16	7.58	7.10
Social	7.59	8.30	7.56	7.20
Level	Medium	High	Medium	Low
Total	7,118	1,923	3,273	1,922

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

Ahora, es posible comparar las instancias agrupadas con la agrupación propuesta utilizando los índices de discriminación, véase la Tabla 5.

**Tabla 5.**

### *Instancias $k$ -means clustering*

Clúster	<i>k</i> -means		Índices	
	Instancias	%	Instancias	%
High	1,923	27	1,922	27
Medium	3,273	46	3,274	46
Low	1922	27	1922	27

Fuente: Elaboración *propia a partir de la información obtenida en WEKA*

Como puede observarse, el algoritmo de agrupación *k*-means (Tablas 4 y 5) confirma la agrupación propuesta mediante el uso del índice de discriminación, quedando establecido el objetivo específico 2.

Por otro lado, al aplicar los algoritmos de clasificación en WEKA, se utilizó una validación cruzada estratificada del número dado de particiones (folds). La validación cruzada consiste en proporcionar un número natural  $n$ , por defecto  $n = 10$ , para dividir los datos en  $n$  partes y, para cada parte, construir el clasificador con las restantes  $n - 1$  partes y probarlo en ellas; el proceso se repite para cada una de las particiones, Frank *et al.* (2016) y Hall *et al.* (2009).

En la Tabla 6 se muestra un resumen de los distintos métodos de clasificación utilizados para construir los modelos de predicción.

**Tabla 6.**

*Resultados WEKA*

Algoritmo	Correctos	Incorrectos	$\kappa$
J48	81.1042%	18.8958%	0.7046
NaivesBayes	79.2638%	20.7362%	0.6690
MultilayerPerceptron	80.5142%	19.4858%	0.6911
RandomForest	81.4414%	18.5586	0.7106

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

De acuerdo con la Tabla 6, los algoritmos RandomForest y J48 tienen el mayor porcentaje de clasificaciones correctas.

Para determinar la eficiencia de los algoritmos se realiza la evaluación, el primer indicador es la matriz de confusión de cada uno de ellos.

**Tabla 7.**

*Matriz de confusión J48*

	High	Medium	Low
High	1,628	294	0
Medium	269	2,655	350
Low	0	432	1,490

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

**Tabla 8.**

*Matriz de confusión NaivesBayes*

	High	Medium	Low
High	1,298	624	0
Medium	91	2,838	345
Low	0	416	1,506

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

**Tabla 9.**

*Matriz de confusión MultilayerPerceptron*

	High	Medium	Low
High	1,550	372	0
Medium	234	2,800	240
Low	0	541	1,381

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

**Tabla 10.**

*Matriz de confusión RandomForest*

	High	Medium	Low
High	1,636	286	0
Medium	285	2,637	352
Low	0	398	1,524

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

En general, los resultados mostrados en las Tablas 7, 8, 9 y 10 son buenos, ninguno de los algoritmos clasifica un dato de *High* a *Low* o viceversa. No obstante, RandomForrest y J48 muestran clasificaciones más consistentes (Tablas 7 y 10), los otros dos algoritmos concentran su clasificación en *Medium* (Tablas 8 y 9).

Para completar la evaluación de los algoritmos se analizan los parámetros *Precision*, *Recall*, *F-measure* y *Accuracy*. Los resultados se muestran en la Tabla 11.

**Tabla 11.**

*Clústeres y centroides*

Algoritmo	Precision	Recall	F-measure	Accuracy
J48	0.812	0.811	0.811	0.811
NaiveBayes	0.809	0.793	0.792	0.793
MultilayerPerceptron	0.812	0.805	0.805	0.805
RandomForest	0.815	0.814	0.814	0.814

Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

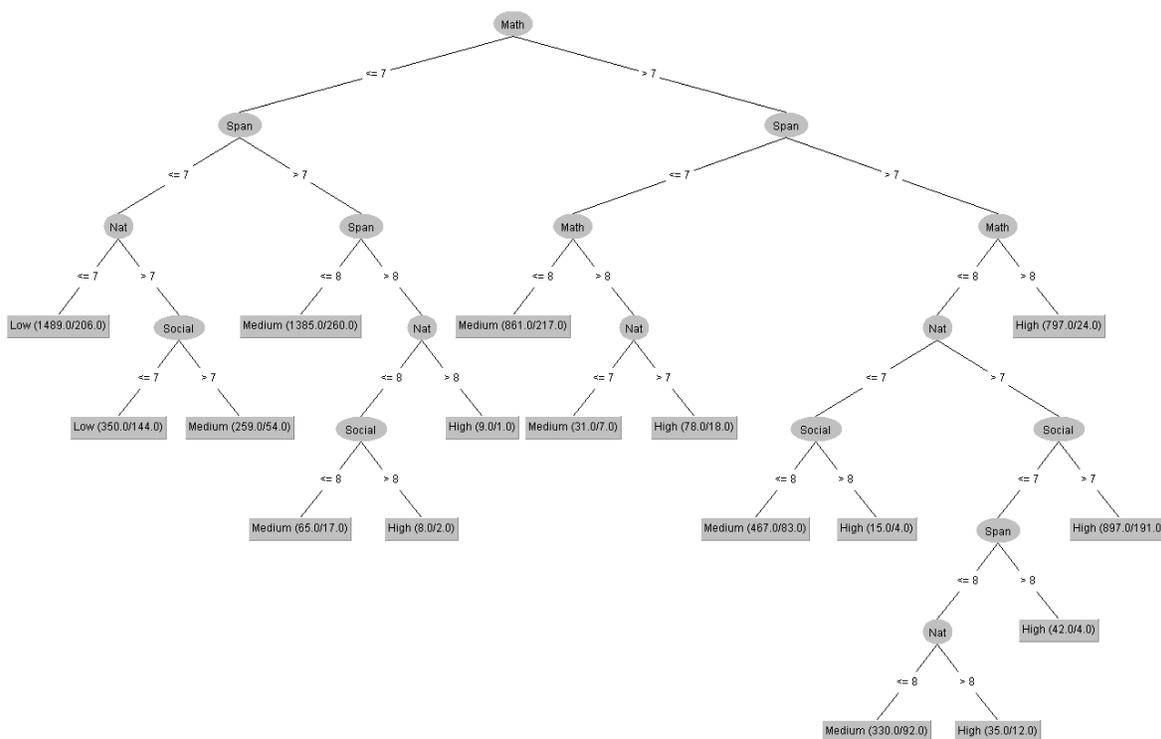
Los resultados de la Tabla 11 indican que los algoritmos tienen, en general, un buen desempeño, todos ellos tienen una *precisión* (superior a 0,8 y una *exactitud* (Accuracy) alta.

Una vez más, los algoritmos de clasificación RandomForest y J48 tienen mejores parámetros, pues *Recall* es prácticamente igual a *Precision*, mientras que la *F-measure* es buena en ambos algoritmos.

A pesar del desempeño del algoritmo RandomForest, sólo el algoritmo J48 permite la visualización del árbol de decisión, por lo que este algoritmo es utilizado como base para la predicción deseada.

La Figura 2 muestra el árbol de decisión generado por el algoritmo J48, donde cada nodo de hoja está clasificado en tres clases de rendimiento: **Alto** (High), **Medio** (Medium) y **Bajo** (Low).

**Figura 2.**  
*Árbol de decisión J48.*



Fuente: *Elaboración propia a partir de la información obtenida en WEKA*

En resumen, los resultados mostrados en las tablas 6, 7, 8, 9, 10 y 11, así como en la Figura 2, comprueban los objetivos específico 1 y 3, las clasificaciones realizadas por los algoritmos, en especial el J48 y el RandomForest, conducen a un buena predicción sobre el desempeño de los estudiantes que enfrentan el examen de ingreso.

En particular, el árbol de decisión mostrado en la Figura 2 establece, de manera fehaciente, el cumplimiento del objetivo general planteado en este trabajo de investigación.

### ***Discusiones.***

La mayoría de los autores consultados que utilizan técnicas de clasificación para predecir desempeño académico centran su atención en materias específicas o la graduación de los estudiantes y utilizan las calificaciones parciales o finales de materias específicas del nivel superior. Incluso Hamsa, *et al.* (2016) que si utiliza las puntuaciones obtenidas por estudiantes en el examen de admisión, solo evalúa el desempeño de los estudiantes del nivel superior. No obstante, las ideas y técnicas aplicadas para el desarrollo de sus investigaciones han sido cruciales para la buena conducción de este trabajo.

A diferencia de los autores citados, este trabajo se ha centrado en el análisis cuantitativo y cualitativo de las calificaciones finales de estudiantes graduados de secundaria para la predicción del desempeño en el examen de ingreso al bachillerato (nivel medio superior) y como ellos, se han aplicado técnicas de clasificación implementadas por el software WEKA.

El análisis predictivo se ha realizado a partir de las calificaciones finales obtenidas por los aspirantes al finalizar la secundaria en Español, Matemáticas, Ciencias Naturales y Ciencias Sociales. La predicción del desempeño se realizó sobre el nivel de habilidad/conocimiento mostrado en una prueba estandarizada, medido por el parámetro  $\theta$ , que puede tomar valores entre  $-3$  y  $+3$  y, categorizado en **Alto**, **Medio** y **Bajo**.

En primer lugar, el algoritmo de agrupación  $k$ -means muestra que la agrupación de los alumnos según su rendimiento es correcta y coincide con la propuesta por el índice de discriminación (Tablas 5).

Por otro lado, todos los algoritmos de clasificación proporcionan una *precisión* superior al 80% y la mayoría de ellos tienen *exactitud* mayor al 80%.

Además, todos los algoritmos de clasificación tienen un índice kappa de entre 0.69 y 0.72, lo que significa que tienen una concordancia sustancial.

Por último, el análisis mostró que los algoritmos de clasificación RandomForest y J48 tienen los mejores resultados paramétricos. Más aún, el árbol de decisión J48 permite crear reglas predictivas para determinar el nivel de desempeño de un estudiante en la prueba estandarizada.

Las reglas predictivas se crean a partir de la Figura 2, solo basta recorrer las ramas desde el nodo raíz hasta el nodo hoja deseado. Por ejemplo,

- La raíz Math indica que la base fundamental de la predicción está basada en la calificación final que obtiene un alumno en matemáticas; *menor o igual a 7 o mayor a 7*.
- Si  $\text{Math} < 7$ , la única manera que un alumno tenga un desempeño *Alto* es que la calificación final en Español sea superior a 8 y en Ciencias Naturales también sea superior a 8. Si la calificación final en Ciencias Naturales es menor o igual que 8, es necesario que la calificación final en Ciencias Sociales sea mayor que 8.
- Si  $\text{Math} > 7$ , ningún alumno tendrá un desempeño *Bajo*.

Para finalizar es importante confirmar que el objetivo planteado en este trabajo de investigación se ha confirmado con todos los análisis cuantitativos y cualitativos realizados a través del software WEKA.

## Conclusión

A diferencia de algunos de los autores consultados, el trabajo desarrollado considera únicamente aspectos académicos: calificaciones finales en

Español, Matemáticas, Ciencias Naturales y Ciencias Sociales. Como se desprende de los análisis, estos aspectos pueden predecir casi a la perfección el desempeño de los estudiantes en una prueba estandarizada utilizada como examen de ingreso al bachillerato.

Sin embargo, cabe señalar que la base fundamental para tener un desempeño medio o alto se basa en las habilidades y conocimientos que se tienen en Matemáticas, requiriendo una calificación final mínima de 8. Pero, si es 7 o menos, se requiere una calificación final alta en Español, Ciencias Naturales y Ciencias Sociales.

Aunque la construcción y aplicación de pruebas estandarizadas garantiza la equidad al tratar de minimizar factores no académicos como la escuela de procedencia, el nivel educativo de los padres, el nivel socioeconómico, entre otros, es bien sabido que el nivel académico de los estudiantes de zonas rurales puede ser inferior al de los estudiantes de zonas urbanas, Chao et al. (2015) y Ramos et al. (2012).

Este problema motiva investigar que tan significativa resulta esta diferencia, y si lo es, intentar establecer una equivalencia entre las calificaciones asignadas a los alumnos rurales y urbanos. Este análisis podrá establecer, desde un punto de vista académico, que un 9 puede no representar el mismo nivel de conocimientos para ambos grupos. La pregunta que surge es si el nivel de desempeño mostrado en las pruebas estandarizadas permitiría establecer algún tipo de equivalencia.

Otra área de oportunidad que se vislumbra es explorar si las técnicas de clasificación y agrupamiento aplicadas a los resultados de una prueba estandarizada, como la utilizada en el examen de ingreso, son, en realidad, capaces de predecir el éxito académico, la deserción y la eficiencia terminal de los estudiantes de mejor manera que el coeficiente de discriminación.

## Referencias

- Al-Razgan, M., Al-Khalifa, A. S. & Al-Khalifa, M. S. (2014). Educational Data Mining: A Systematic Review of the Published Literature 2006–2013. *Proceedings of the First International Conference on Advanced Data and Information Engineering*. Springer, Vol. 285, pp. 711–719. doi: 10.1007/978-981-4585-18-7\\_80
- Backhoff E, Larrazolo N, Rosas M. (2020). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1), pp 11–285. <https://www.redalyc.org/articulo.oa?id=15502102>
- Baker, F. B. and Kim, S. H. (2017). *The Basics of Item Response Theory Using R*. Springer, Cham, New York, NY.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), pp. 37–46. doi: 10.1177/001316446002000104
- Chao-Fernández, R., Ferreiro-Seoane, F., & Mato-Vázquez, M. (2015). Analysis of Performance Differences among Students in Urban and Non-urban Centers Based on Compulsory Secondary Education Special Awards. *Revista Electrónica Educare*, 19(3), pp. 1–19. doi: 10.15359/ree.19-3.23
- Dole, L., & Rajurkar, J. (2015). A Decision Support System for Predicting Student Performance. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(15), pp. 7232–7237. doi: 10.15680/IJIRCCE.2014.0212015
- Durairaj, M. & Vijitha, C. (2014). Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), pp. 5987–5991. doi = 10.1.1.567.8824

- Frank, E., Hall, M. A. & Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, CA. doi: 10.5555/3086818
- Hall, M., Frank, E, Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. (2009). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *SIGKDD Explorations*, 11(1), pp. 10–18. doi: 10.1145/1656274.1656278
- Hamsa, H., Indiradevi, S. & Kizhakkethottam, J. J. (2016). The WEKA Data Mining Software: An Update. *Procedia Technology, Elsevier Ltd*, 25, pp. 326–332. doi: 10.1016/j.protcy.2016.08.114
- Hand, J. D., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*. The MIT Press, Cambridge, MA.
- Henrysson, S. (1971). *Gathering, Analysing, and Using Data on Test Items*. Educational Measurement, Washington, DC.
- Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies*, 13(1), pp. 61–72. doi: 10.2478/cait-2013-0006
- Kaur, Hardeep. (2018). A Literature Review from 2011 TO 2014 on Student's Academic Performance Prediction and Analysis using Decision Tree Algorithm. *Journal of Global Research in Computer Science* 9.5, pp. 10-15. <https://api.semanticscholar.org/CorpusID:54780296>
- Ketui, N., Wisomka W. & Homjun, K. (2019). Using Classification Data Mining Techniques for Students Performance Prediction. *The 4th International Conference on Digital Arts, Media and Technology and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering*, pp. 359–363. doi: 10.1109/ECTI-NCON.2019.8692227

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), pp. 159–174. doi: 10.2307/2529310
- Meghji, A. F., Mahoto, N. A., Unar, M. A. & Shaikh, M. A. (2019). Predicting Student Academic Performance using Data Generated in Higher Educational Institutes. *3C Tecnología. Glosas de innovación aplicadas a la pyme*, Edición Especial pp. 366–383. doi: 10.17993/3ctecno.2019.specialissue2.366–383
- Menacho Chioc, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), pp. 26–33). doi: 10.21704/ac.v78i1.811
- Piotesky–Shapiro, G. & Frawley, W. J. (1991). *Knowledge Discovery in Database*. The MIT Press, Cambridge, MA.
- Ramos, R., Duque, J. C. & Nieto, S. (2012). An analysis of rural and urban differences in the educational performance of Colombian students based on PISA microdata (In Spanish). *RiSE-group*. <https://2012.economicsofeducation.com/user/pdfsesiones/177.pdf>
- Sembiring, S. (2012). *An Application of Predicting Student Performance Using Kernel K–Means and Smooth Support Vector Machine*. Masters thesis, Universiti Malaysia Pahang.
- Sivasakthi, M. (2017). Classification and prediction based data mining algorithms to predict students' introductory programming performance. *International conference on inventive computing and informatics*, Coimbatore, India, pp. 346–350, doi: 10.1109/ICICI.2017.8365371