

The logo for CienciaUAT, featuring the text "CienciaUAT" in a bold, orange, sans-serif font. The "U" is slightly larger and more prominent than the other letters.

CienciaUAT

ISSN: 2007-7521

ISSN: 2007-7858

Universidad Autónoma de Tamaulipas

Acosta-Gonzaga, Elizabeth; Ramirez-Arellano, Aldo  
Estudio comparativo de técnicas de analítica del aprendizaje para predecir  
el rendimiento académico de los estudiantes de educación superior  
CienciaUAT, vol. 15, núm. 1, 2020, Julio-Diciembre, pp. 63-74  
Universidad Autónoma de Tamaulipas

DOI: <https://doi.org/10.29059/cienciauat.v15i1.1392>

Disponible en: <https://www.redalyc.org/articulo.oa?id=441970372004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH The logo for redalyc.org, featuring the text "redalyc.org" in a red, sans-serif font. The "red" part is in a smaller font size than "alyc.org".

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto



"imagen: ijeab La portada ha sido diseñada usando imágenes de Freepik.com"

# Estudio comparativo de técnicas de analítica del aprendizaje para predecir el rendimiento académico de los estudiantes de educación superior

## Comparative study of learning analytics techniques to predict student academic performance in higher education

Elizabeth Acosta-Gonzaga\*, Aldo Ramirez-Arellano

### RESUMEN

La deserción escolar involucra diversos factores, entre ellos, el compromiso del estudiante, a través del cual se puede predecir su éxito en la escuela. Ese compromiso tiene varios componentes, tales como conductual, emocional y cognitivo. La motivación y el compromiso están fuertemente relacionadas, ya que la primera es un precursor del compromiso. El objetivo de este estudio fue comparar la eficacia de la regresión lineal contra dos técnicas de minería de datos para predecir el rendimiento académico de los estudiantes en la educación superior. Se hizo un estudio transversal explicativo en el que se encuestó a 222 estudiantes universitarios de una institución pública de la Ciudad de México. Se realizó un análisis de regresión lineal jerárquico (RL) y de técnicas de analítica del aprendizaje, como redes neuronales (RN) y máquinas de vector soporte (SVM). Para evaluar la exactitud de las técnicas de analítica del aprendizaje se realizó un análisis de varianza (ANOVA). Se compararon las técnicas de analítica del aprendizaje y de regresión lineal usando la validación cruzada. Los resultados mostraron que el compromiso conductual y la autoeficacia tuvieron efectos positivos en el desempeño del estudiante, mientras que la pasividad mostró un efecto negativo. Asimismo, las técnicas de RL y de SVM pronosticaron igualmente el desempeño académico de los estudiantes. La RL tuvo la ventaja de producir un modelo simple y de fácil interpretación. Por el contrario, la técnica de SVM generó un modelo más complejo, aunque, si el modelo tuviese como objetivo el pronóstico del desempeño, la técnica SVM sería la más adecuada, ya que no requiere la verificación de ningún supuesto estadístico.

**PALABRAS CLAVE:** compromiso del estudiante, motivación del estudiante, desempeño académico, analítica del aprendizaje.

### ABSTRACT

The issue of school dropout involves factors such as students' engagement that can predict his or her success in school. It has been shown that student engagement has three components: behavioral, emotional and cognitive. Motivation and engagement are strongly related since the former is a precursor of engagement. The aim of this study was to compare the efficiency of linear regression against two data mining techniques to predict the students' academic performance in higher education. A descriptive cross-sectional study was carried out with 222 students from a public higher education institution in Mexico city. An analysis of hierarchical linear regression (LR) and learning analytics techniques such as neural networks (NN) and support vector machine (SVM) was conducted. To assess the accuracy of the learning analytics techniques, an analysis of variance (ANOVA) was carried out. The techniques were compared using cross validation. The results showed that behavioral engagement and self-efficacy had positive effects on student achievements, while passivity showed a negative effect. Likewise, the LR and SVM techniques had the same performance on predicting students' achievements. The LR has the advantage of producing a simple and easy model. On the contrary, the SVM technique generates a more complex model. Although, if the model were aimed to forecast the performance, the SVM technique would be the most appropriate, since it does not require to verify any statistical assumption.

**KEYWORDS:** student engagement, student motivation, academic achievements, learning analytics.

\*Correspondencia: eacostag@ipn.mx/ Fecha de recepción: 7 de noviembre de 2019/ Fecha de aceptación: 12 de julio de 2020/ Fecha de publicación: 31 de julio de 2020.

Instituto Politécnico Nacional, Unidad Profesional Interdisciplinaria de Ingeniería Ciencias Sociales y Administrativas, SEPI. Av. Té, núm. 950 esquina, Resina, col. Granjas México, Alcaldía Iztacalco, Ciudad de México, México, C. P. 08400.

## INTRODUCCIÓN

En la educación superior el problema de la deserción escolar involucra múltiples factores, y el compromiso del estudiante es uno de ellos. Este juega un rol esencial, ya que impulsa el aprendizaje y puede predecir su éxito en la escuela (Reschly y Christenson, 2012). El compromiso de un estudiante únicamente ocurre si este tiene el deseo de comprometerse; en otras palabras, un estudiante puede querer algo y, sin embargo, no hacer nada para obtenerlo (González y col., 2016). Por lo tanto, la motivación influye en el compromiso y este guía al desempeño del estudiante (Martin, 2012). La teoría de la expectativa-valor de Pintrich y de-Groot (1990) muestra cómo interviene la motivación en la realización de metas. Esta teoría concibe a la motivación con dos componentes. El componente de expectativa (autoeficacia y control de la creencia del aprendizaje), que se refiere a la confianza de los alumnos en su capacidad para realizar una tarea y que ellos son los responsables de su propio rendimiento. El componente de valor se refiere a la motivación extrínseca e intrínseca hacia una meta, así como el valor que el alumno otorga a una tarea.

El estudio teórico sobre el compromiso del estudiante ha mostrado que es un constructo formado por varios componentes (González y col., 2016). Skinner y Belmont (1993) y Reschly y Christenson (2012) propusieron estudiar el compromiso del estudiante a través del componente de comportamiento y el componente emocional, mientras que investigadores como Fredricks y col. (2004); Appleton y col. (2006); Appleton y col. (2008), además de los anteriores, agregan un componente cognitivo. Por otro lado, está el término opuesto del compromiso del estudiante conocido como desafección o desapego (en inglés *disaffection*) (Skinner y Pitzer, 2012; Fernández y col., 2013; Salmela-Aro y Upadyaya, 2014). Los estudiantes que se encuentran en un estado de distanciamiento muestran emociones negativas tales como aburrimiento, depresión, ansiedad o incluso enojo. Normalmente, los estudian-

tes no comprometidos son pasivos y no se esfuerzan ante los retos escolares (Skinner y Belmont, 1993).

Reschly y Christenson (2012) definen al compromiso conductual como la participación de los estudiantes en actividades académicas, sociales o extracurriculares. El compromiso cognitivo se refiere a la inversión psicológica y esfuerzo del estudiante en el aprendizaje involucrando el uso de estrategias de aprendizaje (Newmann y col., 1992; González y col., 2016). Es decir, se basa en la autorregulación y el esfuerzo por el dominio del tema (Fredricks y col., 2004; Reschly y Christenson, 2012). El compromiso emocional se define como las respuestas afectivas de los estudiantes, ya sean emociones positivas o negativas a las tareas escolares (Fredricks y col., 2004), considerando también la interacción con maestros y compañeros de clase (Reschly y Christenson, 2012).

Desde el punto de vista psicológico se ha demostrado que la motivación y las emociones se influyen una a la otra de forma directa (Mayer y col., 1997). En un contexto escolar esto puede observarse cuando un estudiante que está muy comprometido con sus actividades escolares despliega emociones positivas tales como entusiasmo, optimismo, curiosidad e interés (Skinner y Belmont, 1993). Sin embargo, cuando se encuentra desinteresado puede mostrar aburrimiento (Skinner y Pitzer, 2012; Symonds y col., 2016). El desinterés del estudiante hacia su trabajo escolar y su escolaridad está relacionado con factores emocionales, personales y sociales (Symonds y col., 2016). De hecho, es importante evaluar el desinterés emocional de los estudiantes, ya que este se encuentra relacionado con un bajo rendimiento académico (Li y Lerner, 2011). Los efectos de la motivación en los estudiantes también han sido estudiados en el compromiso académico. Por ejemplo, Wang y Eccles (2013) demostraron que el ambiente escolar impactó la motivación y, a su vez, impactó a los tres tipos de compromiso, conductual, emocional y cognitivo.

La investigación de Chatzistamatiou y col. (2015) registró, que los estudiantes que valoraron las actividades de matemáticas y se sintieron eficaces en matemáticas (valor de la tarea y autoeficacia, componentes motivacionales) usaron estrategias cognitivas y metacognitivas para lograr su rendimiento académico. Asimismo, en el estudio de King y McInerney (2016) se observó que existe una relación entre la motivación, las estrategias de aprendizaje (compromiso en el aprendizaje) y el desempeño académico de los estudiantes.

Al proceso de captura, almacenamiento y análisis de datos para predecir patrones de aprendizaje que se usan como fuente de información significativa para mejorar los procesos educativos se le conoce como analítica del aprendizaje o análisis del aprendizaje (*learning analytics*). Investigadores como Siemens y Gasevic (2012) y Gasevic y col. (2015) definieron la analítica del aprendizaje como las técnicas para medir, recopilar, analizar e informar los datos producidos por el estudiante, que junto con el análisis de modelos se usan para descubrir información y relaciones sociales que predicen y optimizan el aprendizaje. Para extraer la información valiosa de grandes volúmenes de datos de los estudiantes, la analítica del aprendizaje hace uso de la minería de datos educativa (*educational data mining*). Las técnicas de minería de datos incluyen a la regresión lineal y la logística, así como herramientas de clasificación como son las redes neuronales (RN), Naïve-Bayes, máquina de vector soporte (VSM, por sus siglas en inglés: Support Vector Machine) y árboles de decisión (Urbina-Nájera y Mora, 2017; Tanuar y col., 2018). Las técnicas mencionadas poseen ventajas sobre las técnicas estadísticas tradicionales, ya que estas últimas deben cumplir con requisitos como normalidad, linealidad, multicolinealidad y homocedasticidad sin embargo, estos requisitos muchas veces se omiten, lo que provoca que un modelo tenga un ajuste deficiente (Ramirez-Arellano y col., 2019). Por lo tanto, y considerando el compromiso académico y la motivación, se podría preguntar: ¿Las técnicas de *learning analytics* podrían

predecir con la misma efectividad el desempeño académico de los estudiantes que las técnicas estadísticas tradicionales como la regresión lineal?

Autores como Yadav y col. (2012) han usado técnicas de minería de datos, como árboles de decisión, para predecir el desempeño académico de los estudiantes. Otros investigadores también han usado estas técnicas para predecir patrones en el compromiso académico de los estudiantes (Wong y Chong, 2018; Vytasek y col., 2020), así como para analizar el compromiso académico y la motivación del estudiante (Caruth, 2018). El trabajo de Moubayed y col. (2018) mostró que existe una correlación significativa entre el nivel de compromiso del estudiante y su rendimiento académico en un grupo de alumnos inscritos en un curso de aprendizaje mixto, usando un algoritmo de reglas de asociación *a priori*.

En la misma línea, el trabajo de Arnold (2010) demostró que el uso de técnicas de analítica de aprendizaje ayudó a hacer predicciones de cómo dar asesoría a los estudiantes de un curso abierto masivo en línea (MOOC, por sus siglas en inglés: Masive Online Open Course) de programación. También, la investigación de Halibas y col. (2018), examinó el compromiso del estudiante dentro y fuera del salón de clases mediante el uso de técnicas de analítica de datos y radio frecuencia. Asimismo, el estudio de Liu y col. (2013) analizó el compromiso conductual de los estudiantes en un curso sobre escritura mediante un sistema de analítica de datos para visualizar patrones del comportamiento de los estudiantes mientras escribían.

El objetivo del presente estudio fue comparar la eficiencia de la regresión lineal contra dos técnicas de minería de datos para predecir el rendimiento académico de los estudiantes en la educación superior.

## MATERIALES Y MÉTODOS

En este estudio participaron estudiantes universitarios matriculados en diversas unidades

de aprendizaje de las licenciaturas de Ingeniería Industrial y Administración Industrial de una institución pública de la Ciudad de México. Se obtuvo una muestra de 222 alumnos. Para obtener una calificación aprobatoria, los estudiantes acreditaron la unidad de aprendizaje realizando 3 exámenes y diversas tareas académicas, como exposiciones en clase y escritura de ensayos. Después de cada examen, se pidió a los estudiantes que respondieran dos instrumentos (se explican en la sección de instrumentos). Los datos se recopilaron de septiembre a diciembre de 2018. A todos los estudiantes se les hizo saber el objetivo de esta investigación, participaron voluntariamente y dieron su consentimiento para que sus respuestas fueran consideradas para este estudio.

## Método de investigación

### Instrumentos

Para recopilar los datos referentes a emociones y conductas de los estudiantes, se utilizó el instrumento compromiso y desapego del estudiante en la escuela (SED, por sus siglas en inglés: Student Engagement and Disaffection in school) de Skinner y col. (2008), el cual fue adaptado para el contexto de estudio y validado en una universidad de México (Ramírez-Arellano y col., 2018). Este instrumento se usa para medir el compromiso emocional de los alumnos y se obtuvo a través de 5 preguntas que evaluaron emociones positivas y que aportaron datos para los constructos disfrute, entusiasmo, interés, diversión y autorrealización. Este instrumento también mide el desapego emocional e incluye 12 preguntas que arrojan datos para los constructos aburrimiento, frustración, desinterés, ansiedad y tristeza. De igual forma, el compromiso conductual evaluó 5 preguntas referentes a conducta positivas como atención, esfuerzo e involucramiento. El desapego conductual también se midió con 5 preguntas que involucraron emociones negativas y midieron factores como pasividad, distracción y desapego mental. Por último, se incluyó al compromiso emocional, desapego emocional, compromiso conductual y desapego conductual. De esta forma, se integraron en el

análisis 20 constructos – que son el total de constructos del SED -. Las respuestas con valores bajos de desafección emocional significaron que los alumnos experimentaron emociones negativas. Por ejemplo, si el alumno otorgó 5 puntos a la pregunta “cuando trabajamos en algo en clase, me siento aburrido” significa que el estudiante experimentó aburrimiento. Todas las respuestas de los estudiantes se basaron en una escala tipo Likert de 5 puntos, donde 1 significó totalmente en desacuerdo y 5 significó totalmente de acuerdo.

El cuestionario de estrategias de motivación para el aprendizaje (MSLQ, por sus siglas en inglés: Motivated Strategies for Learning Questionnaire) (Pintrich y de-Groot, 1990; Pintrich, 1991) fue previamente adaptado para el contexto de este estudio y validado (Ramírez-Arellano y col., 2018). Este instrumento evalúa los constructos metacognitivos, cognitivos, estrategias de aprendizaje y motivacionales de los estudiantes. El constructo motivacional mide 2 componentes: valor y expectativas. El componente de valor incluye la motivación intrínseca y extrínseca hacia la meta, las creencias acerca de la importancia de una tarea o valor de la tarea y control de las creencias de aprendizaje. El componente de expectativa incluye a la autoeficacia y ansiedad a los exámenes. Los constructos metacognitivos, cognitivos y estrategias de aprendizaje, aunque son parte de este instrumento (MSLQ), no se integraron en el análisis debido a que no mostraron estar correlacionados con el rendimiento académico del estudiante (calificación global). Los estudiantes contestaron 81 preguntas que se respondieron usando una escala de Likert, que va desde 1 (no es del todo cierto) hasta 7 (muy cierto). Los alumnos respondieron a los instrumentos SED y MSLQ dos semanas antes de finalizar cada curso. La fiabilidad de los instrumentos SED y MSLQ se determinó a través del coeficiente de alfa de Cronbach.

Finalmente, el rendimiento académico se calculó considerando la calificación global (CG) del curso, la cual se obtuvo teniendo en cuen-



ta las actividades de aprendizaje y 3 exámenes parciales, exposiciones y ejercicios en clase, así como la escritura de ensayos. La calificación global del estudiante se recopiló al final del semestre.

### Análisis de datos

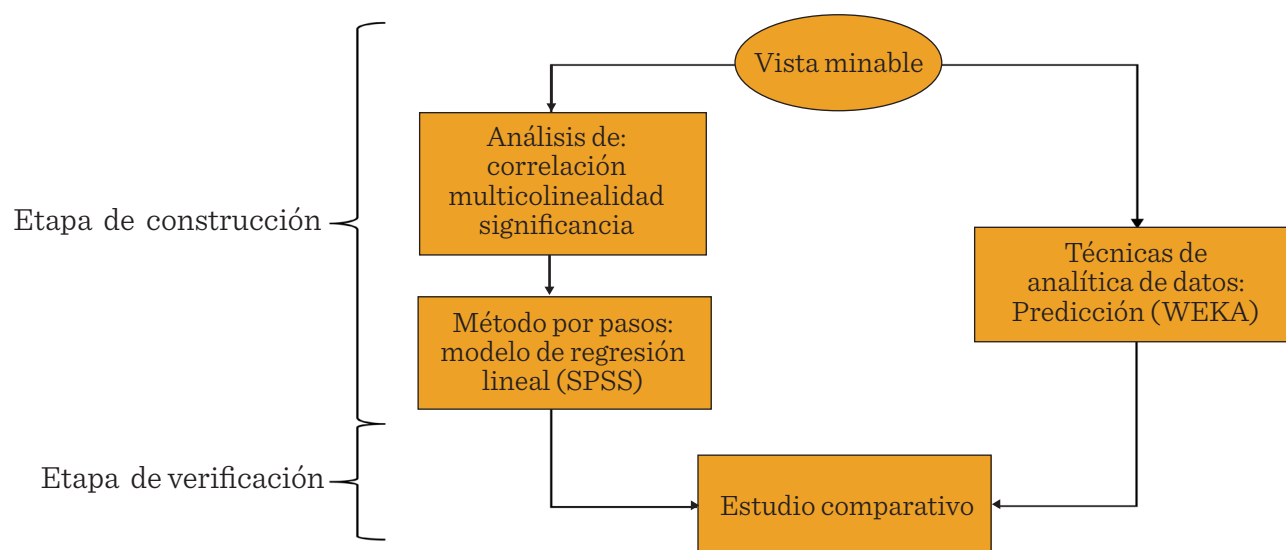
Los resultados de los instrumentos SED, MSLQ y de la calificación global de cada alumno se usaron para construir una vista minable con 27 constructos, incluyendo 20 constructos del instrumento SED - disfrute, entusiasmo, interés, diversión, autorrealización, aburrimiento, frustración, desinterés, ansiedad, tristeza, atención, esfuerzo, involucramiento, pasividad, distracción, desapego mental, compromiso emocional, desapego emocional, compromiso conductual y desapego conductual - y 6 constructos del instrumento MSQL - motivación intrínseca, motivación extrínseca, valor de la tarea, control de las creencias de aprendizaje, autoeficacia, ansiedad a los exámenes - y la calificación global. El proceso de análisis de datos se muestra en la Figura 1.

Después de seguir el proceso de análisis de datos propuesto en la Figura 1, en la primera etapa, de construcción, se consideraron los 27 constructos de la vista minable (datos prin-

cipales) para construir los modelos de la analítica del aprendizaje (RN, SVM) y para el modelo de regresión lineal (RL) se seleccionaron los constructos que tuvieron un nivel de significancia ( $P < 0.01$ ) con el desempeño académico o calificación global de cada estudiante.

Siguiendo en la etapa de construcción, se realizó un análisis de los constructos para determinar los valores de multicolinealidad del modelo resultante RL y la verificación de que estuvieran dentro de rango aceptable. Según Meyers y col. (2006), los valores aceptables para que un modelo no presente multicolinealidad son: índice de condición, con valores menores a 20; proporciones de varianza, con valores por debajo de 0.90; valores de tolerancia mayores a 0.1; y valores del factor de inflación de la varianza menores de 10.

Posteriormente, se realizó un análisis de regresión jerárquico (método por pasos), que tuvo por objetivo explicar las relaciones entre el compromiso emocional, conductual, emociones y el rendimiento académico. Para ello, se usó el paquete estadístico para ciencias sociales (SPSS, por sus siglas en inglés: Statistical Package for the Social Sciences) versión 22.



■ Figura 1. Proceso de análisis de datos.

Figure 1. Data Analysis Process.

Asimismo, durante la primera etapa, a través de las técnicas de analítica de datos (minería de datos) se determinaron los constructos que predicen el rendimiento académico del estudiante. Para ello, se construyeron 2 modelos con los 27 constructos de la vista minable, usando las técnicas de analítica del aprendizaje como redes neuronales (RN) y máquinas de vector soporte (SVM). Se usó el *software* entorno Waikato para el análisis del conocimiento (WEKA, por sus siglas en inglés: Waikato Environment for Knowledge Analysis) (Hall y col., 2009).

Durante la etapa de verificación se compararon las técnicas de minería de datos y el modelo de regresión lineal usando la validación cruzada. Es decir, de los datos recabados se seleccionaron de manera aleatoria, alrededor de 60 % para construir los modelos y el 40 % restante se usó para evaluar su exactitud. Este proceso de selección se repitió 10 veces para garantizar que los modelos no tuvieran un sobreajuste, ofreciendo una evaluación confiable de la exactitud de los modelos. Comparar múltiples clasificadores en varios conjuntos de datos es una tarea recurrente para evaluar el rendimiento de las técnicas de minería de datos. Por lo tanto, un análisis de varianza (ANOVA) es una prueba estadística apropiada para determinar cuál técnica es la mejor: la de minería de datos o la de regresión lineal.

## RESULTADOS

Los resultados de la fiabilidad de los instrumentos SED y MSLQ obtuvieron valores de alfa de Cronbach en el rango de 0.769 a 0.824. El valor de 0.7 es considerado como el valor mínimo aceptable de fiabilidad (Taber, 2018), por lo tanto, ambos instrumentos mostraron niveles aceptables de fiabilidad. Asimismo, se obtuvieron los resultados de la estadística descriptiva que se presentan en la Tabla 1.

De los 27 constructos de la vista minable (datos principales), se seleccionaron aquellos que tuvieron una correlación de Pearson significativa ( $P < 0.01$ ) - 14 constructos - con el de-

sempño académico (Tabla 2). Posteriormente fueron eliminados 11 constructos (compromiso emocional, entusiasmo, disfrute, diversión, autorrealización, interés, esfuerzo, involucramiento, atención, motivación intrínseca hacia la meta, valor de la tarea) al presentar alta colinealidad, por lo que, se excluyeron del conjunto de constructos para el modelo de regresión lineal. La Tabla 3 muestra los tres constructos resultantes del modelo final de regresión lineal de este trabajo, el cual no presentó problemas de multicolinealidad debido a que se obtuvieron valores aceptables del índice de condición, entre 1 y 6.752, proporciones de varianza entre 0.001 a 0.46, valores de tolerancia entre 0.463 y 0.576 valores del factor de inflación de la varianza entre 1.73 y 2.16, ya que en el índice de condición se reportaron valores menores a 20; en proporciones de varianza, valores por debajo de 0.90; en tolerancia, valores mayores a 0.1 y en el factor de inflación de la varianza, valores menores de 10.

En el análisis de la varianza ANOVA de una vía para comparar el coeficiente de determinación ajustado del modelo de regresión lineal (RL) y los obtenidos por las técnicas de minería de datos: RN y SVM (Tabla 4), el efecto principal del modelo en el coeficiente de determinación ajustado fue significativo:  $F(2, 297) = 21.69$ ,  $P < 0.001$ ,  $\eta^2 = 0.127$  (Tabla 4).

La prueba *post hoc* de Tukey (Tabla 4) indicó que no hay diferencia significativa entre los modelos RL y SVM. Sin embargo, el modelo RN tuvo un coeficiente de determinación ajustado significativamente menor que el modelo RL y SVM. Por lo tanto, los modelos RL y SVM pronostican con la misma certeza el desempeño de los estudiantes, lo cual significa que el uso de la regresión lineal es igualmente efectiva que la técnica SVM. La Tabla 5 muestra los resultados.

El resumen de los modelos generados por las técnicas RL, RN y SVM se muestran en la Tabla 6. También se presentan el coeficiente de determinación ajustado ( $R^2$  ajustado) y la

■ Tabla 1. Estadística descriptiva de los instrumentos SED, MSLQ y de la calificación global (27 constructos).

Table 1. Descriptive statistics of the SED, MSLQ instruments and the global grade (27 constructs).

| Constructos                             | M       | DE        | Min a Max   |
|-----------------------------------------|---------|-----------|-------------|
| Compromiso conductual*                  | 1.181 5 | 0.744 63  | 1.00 a 3.30 |
| Compromiso emocional*                   | 1.064 6 | 0.747 77  | 1.00 a 3.70 |
| Desapego conductual*                    | 3.920 3 | 0.960 26  | 1.25 a 4.75 |
| Desapego emocional*                     | 3.847 8 | 0.959 09  | 1.43 a 3.3  |
| Disfrute*                               | 1.512 0 | 0.728 27  | 2.00 a 4.0  |
| Entusiasmo*                             | 4.823 2 | 0.828 98  | 2.00 a 4.0  |
| Interés*                                | 4.896 4 | 0.770 99  | 2.75 a 3.25 |
| Diversión*                              | 4.688 1 | 0.955 96  | 1.31 a 5.0  |
| Autorrealización*                       | 2.514 8 | 0.695 31  | 2.25 a 4.25 |
| Aburrimiento*                           | 3.742 8 | 0.984 66  | 2.25 a 5.0  |
| Frustración*                            | 3.940 3 | 0.961 04  | 2.06 a 4.56 |
| Desinterés*                             | 3.679 1 | 1.032 16  | 2.32 a 5.0  |
| Ansiedad*                               | 3.987 2 | 0.986 08  | 2.13 a 4.50 |
| Tristeza*                               | 3.543 9 | 1.091 43  | 2.45 a 5.0  |
| Atención*                               | 4.825 4 | 0.765 07  | 1.75 a 3.50 |
| Esfuerzo*                               | 4.943 6 | 0.727 65  | 1.05 a 5.0  |
| Involucramiento*                        | 4.554 1 | 0.890 44  | 1.75 a 4.25 |
| Pasividad*                              | 3.887 4 | 1.021 95  | 2.11 a 5.0  |
| Distracción*                            | 3.991 5 | 0.992 619 | 2.0 a 4.75  |
| Desapego mental*                        | 3.864   | 0.997 86  | 2.13 a 5.0  |
| Motivación extrínseca hacia la meta     | 2.829 3 | 1.054 9   | 1.00 a 4.9  |
| Motivación intrínseca hacia la meta     | 2.859 5 | 1.042 3   | 1.00 a 4.9  |
| Valor de la tarea                       | 2.952 1 | 1.0437    | 1.00 a 5.0  |
| Autoeficacia                            | 2.878 3 | 1.017 7   | 1.00 a 4.8  |
| Ansiedad a los exámenes                 | 2.193 5 | 1.024 3   | 1.00 a 4.8  |
| Control de las creencias de aprendizaje | 2.827 7 | 1.032 6   | 1.00 a 4.8  |
| Calificación global                     | 7.442 0 | 1.049 0   | 1.00 a 10   |

\*Constructos del instrumento SED fueron evaluados con una escala de Likert de 1 a 5. El resto de los constructos pertenecen el instrumento MSLQ y se evaluaron con una escala de Likert de 1 a 7. La calificación global se evaluó del 0 al 10.



■ **Tabla 2. Constructos con nivel de significancia ( $P < 0.01$ ).**  
Table 2. Constructs with level of significance ( $P < 0.01$ ).

| Constructos                                 | Desempeño académico (calificación global) |
|---------------------------------------------|-------------------------------------------|
| Compromiso emocional                        | - 0.382                                   |
| Compromiso conductual                       | - 0.445                                   |
| Componentes del compromiso emocional (SED)  |                                           |
| Entusiasmo                                  | 0.336                                     |
| Disfrute                                    | 0.338                                     |
| Diversión                                   | 0.336                                     |
| Autorrealización                            | 0.338                                     |
| Interés                                     | 0.408                                     |
| Componentes del compromiso conductual (SED) |                                           |
| Esfuerzo                                    | 0.441                                     |
| Involucramiento                             | 0.418                                     |
| Componentes del desapego conductual (SED)   |                                           |
| Pasividad                                   | - 0.128                                   |
| Distracción                                 | 0.420                                     |
| Componentes motivacionales (MSLQ)           |                                           |
| Motivación intrínseca hacia la meta         | 0.183                                     |
| Valor de la tarea                           | 0.210                                     |
| Autoeficacia                                | 0.223                                     |

■ **Tabla 3. Valores de multicolinealidad para el modelo RL propuesto.**  
Table 3. Multicollinearity values for the proposed RL model.

| Dimensión | Índice de condición | Compromiso conductual | Autoeficacia | Pasividad |
|-----------|---------------------|-----------------------|--------------|-----------|
| 1         | 1.00                | 0.01                  | 0.001        | 0.004     |
| 2         | 3.325               | 0.38                  | 0.03         | 0.002     |
| 3         | 6.752               | 0.40                  | 0.05         | 0.46      |

La intersección entre la dimensión y el constructo señala el nivel de varianza. Un nivel de varianza mayor a 0.90 y un índice de condición mayor a 20 indican un alta multicolinealidad entre constructos de la misma dimensión.

■ **Tabla 4. La prueba *post hoc* de Tukey.**  
Table 4. Tukey's *post hoc* test.

|     | R <sup>2</sup> ajustado    | RL    | RN    | SVM   |
|-----|----------------------------|-------|-------|-------|
| RL  | 0.768 (0.23) <sup>a</sup>  | ---   | 0.001 | 0.05  |
| RN  | 0.583 (0.248) <sup>b</sup> | 0.001 | ---   | 0.001 |
| SVM | 0.797 (0.267) <sup>c</sup> | 0.05  | 0.001 | ---   |

<sup>a,b,c</sup>Letras distintas indican diferencia significativa entre métodos. Se señala el nivel de significancia (0.05 o 0.001). Se indica el valor promedio y entre paréntesis la desviación estándar. El coeficiente de determinación ajustado fue significativo:  $F(2, 297) = 21.69$ ,  $P < 0.001$ ,  $\eta^2 = 0.127$ . RL = regresión lineal; RN = redes neuronales; SVM = máquinas de vector soporte.

■ Tabla 5. Resumen del análisis de regresión lineal jerárquico para los constructos de RL que predicen el desempeño académico de los estudiantes (n = 222).

Table 5. Summary of the hierarchical lineal regression analysis for the LR constructs that predict the students' academic performance (n = 222).

| Modelo | Constructo predictor  | B       | ES    | B         | R <sup>2</sup> (ajustado) |
|--------|-----------------------|---------|-------|-----------|---------------------------|
| 1      | compromiso conductual | 0.627   | 0.085 | 0.445***  | 0.198                     |
| 2      | compromiso conductual | 1.111   | 0.082 | 0.789***  | 0.442                     |
|        | autoeficacia          | 0.644   | 0.060 | 0.629***  |                           |
| 3      | compromiso conductual | 1.204   | 0.90  | 0.855***  | 0.77                      |
|        | autoeficacia          | 0.563   | 0.068 | 0.550***  |                           |
|        | pasividad             | - 0.179 | 0.073 | - 0.174** |                           |

B = Coeficiente de regresión no estandarizado; ES = Error Estándar; B = Coeficiente de regresión estandarizado; \*\*P < 0.05; \*\*\*P < 0.001.

■ Tabla 6. Comparación entre los modelos resultantes usando técnicas de estadística tradicional y de minería de datos.

Table 6. Comparison between the resulting models using traditional statistics and data mining techniques.

| Modelo | Técnica | R <sup>2</sup> (ajustado) | Número de variables en el modelo |
|--------|---------|---------------------------|----------------------------------|
| 1      | RL      | 0.768                     | 3                                |
| 2      | RN      | 0.583                     | 27                               |
| 3      | SVM     | 0.797                     | 27                               |

RL = regresión lineal; RN = redes neuronales; SVM = máquinas de vector soporte.

complejidad del modelo de medida con el número de variables que lo componen. Tomando en cuenta el coeficiente de determinación ajustado y la complejidad, el modelo RL podría preferirse sobre el SVM.

## DISCUSIÓN

La comparación entre técnicas de minería de datos y técnicas de estadística tradicional, como la regresión lineal, para predecir el rendimiento académico de los estudiantes universitarios, es un método útil para determinar qué técnica es más adecuada. Dependiendo del objetivo buscado, ambas son útiles para descubrir información y relaciones sociales que predicen y optimizan el aprendizaje (Siemens y Gasevic, 2012; Gasevic y col., 2015; Reyes y col., 2019).

En este estudio, el compromiso conductual mostró una relación positiva con el desempe-

ño académico de los estudiantes (Tabla 5). Este hallazgo está en línea con los trabajos de Fredricks y col. (2004); Wang (2017); Olivier y col. (2018). Asimismo, la pasividad -componente del desapego conductual- tuvo efectos negativos en el desempeño académico de los estudiantes (Skinner y Belmont, 1993; Skinner y col., 2008; Skinner y Pitzer, 2012), por lo que se puede suponer que cuando los estudiantes se involucran en realizar sus actividades académicas con un gran compromiso, presentan una repercusión positiva y significativa en su desempeño académico. De igual forma, se puede observar que los alumnos que muestran mayor pasividad impactan negativamente su desempeño académico.

Los resultados también revelaron que la autoeficacia, perteneciente al componente de expectativa de la motivación, mostró un efecto positivo con el desempeño académico de los

estudiantes (Sun y col., 2018; Ramirez-Arellano y col., 2019), lo que puede significar que la confianza del estudiante en la realización de sus actividades académicas favorece que obtenga mejores calificaciones.

Por otro lado, se puede observar que la técnica de minería de datos usada (SVM) demostró ser igualmente efectiva que las técnicas de estadísticas como la RL. Sin embargo, cabe destacar que las técnicas de la analítica de datos no necesitan comprobar los supuestos de la multicolinealidad y de las correlaciones significativas, lo cual es necesario en la RL para evitar el sobreajuste. Por otro lado, la técnica de minería de datos SVM generó un modelo más complejo, de 27 constructos (Tabla 6), ya que esta técnica asocia la importancia de las variables con el peso que el algoritmo asigna a la misma. Esto significa que una variable con un mayor predictivo tendrá asociado un mayor peso.

## CONCLUSIONES

Los resultados mostraron que el modelo de regresión lineal jerárquico (RL) y el modelo de máquinas de vector soporte (SVM) pronos-

tican igualmente el desempeño académico de los estudiantes. Sin embargo, debe considerarse el objetivo que persigue la construcción del modelo para optar por una de las dos técnicas. Si se busca determinar los factores psicológicos (motivación, compromiso conductual y emocional) que tienen una influencia significativa en el desempeño académico, la RL es la opción adecuada, ya que discrimina las variables no relevantes, produciendo un modelo simple y de fácil interpretación. Por el contrario, el SVM genera un modelo complejo que difícilmente puede ser usado para determinar los factores psicológicos más significativos que inciden en el desempeño académico. Si el modelo tuviese como objetivo el pronóstico del desempeño, la técnica SVM sería la más adecuada, ya que la técnica no requiere la verificación de ningún supuesto estadístico, tal como se pudo observar en la técnica de RL.

## AGRADECIMIENTOS

Los autores agradecen al Instituto Politécnico Nacional por el apoyo otorgado a la presente investigación a través de los proyectos SIP20195324 y SIP20194925.

## REFERENCIAS

- Appleton, J. J., Christenson, S. L., and Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5): 369-386.
- Appleton, J. J., Christenson, S. L., Kim, D., and Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44(5): 427-445.
- Arnold, K. E. (2010). Signals: Applying Academic Analytics. *EDUCAUSE Quarterly*, 33(1): 1.
- Caruth, G. D. (2018). Student Engagement, Retention, and Motivation: Assessing academic success in today's college students (Vol. 5). *Online Submission*, 5(1): 17-30.
- Chatzistamatiou, M., Dermizaki, I., Efklides, A., and Leondari, A. (2015). Motivational and affective determinants of self-regulatory strategy use in elementary school mathematics. *Educational Psychology*, 35(7): 835-850.
- Fernández, A. G., Paoloni, P. V., Rinaudo, M. C., and Donolo, D. (2013). Situational interest in Spanish language class on secondary education: Structural relations with engagement, disaffection, and performance. *Universitas Psychologica*, 12(3): 753-765.
- Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1): 59-109.
- Gasevic, D., Dawson, S., and Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1): 64-71.
- González, A., Rodríguez, Y., Faílde, J. M., and Carreira, M. V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences*, 45: 214-221.
- Halibas, A. S., Pillai, I. G., and Matthew, A. C. (2018). Utilization of RFID analytics in assessing student engagement. [En línea]. Disponible en: <https://doi.org/10.1109/IT-DREPS.2017.8277809>. Fecha de consulta: 8 de mayo de 2019.

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. (2009). The WEKA data mining software: An update. *SIGKDD Explor. Newsl.* 11(1): 10-18.
- King, R. B. and McInerney, D. M. (2016). Do goals lead to outcomes or can it be the other way around? Causal ordering of mastery goals, metacognitive strategies, and achievement. *British Journal of Educational Psychology.* 86(2): 296-312.
- Li, Y. and Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology.* 47(1): 233-247.
- Liu, M., Calvo, R. A., and Pardo, A. (2013). Tracer: A tool to measure and visualize student engagement in writing activities. *IEEE 13th International Conference on Advanced Learning Technologies.* 421-425.
- Martin, A. J. (2012). Part II Commentary: Motivation and engagement: Conceptual, operational, and empirical clarity. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 303-311). Boston, MA: Springer.
- Mayer, J. D., Chabot, H. F., and Carlsmith, K. M. (1997). Cognition, affect, and cognition in personality. In G. Matthews (Ed.), *Advances in Psychology* (pp. 31-63). North-Holland, The Netherlands: Elsevier.
- Meyers, L. S., Gamst, G., and Guarino A. J. (2006). *Applied multivariate research: Design and interpretation.* Thousand Oaks, CA: SAGE Publications. 1016 Pp.
- Moubayed, A., Injadat, M., Shami, A., and Lutfiyya, H. (2018). Relationship between student engagement and performance in e-learning environment using association rules, In *2018 IEEE World Engineering Education Conference (EDUNINE)*. [En línea]. Disponible en: <https://doi.org/10.1109/EDUNINE.2018.8451005>. Fecha de consulta: 29 de junio de 2019.
- Newmann, F. M., Wehlage, G. G., and Lamborn, S. (1992). *The Significance and sources of student engagement* (In Student engagement and achievement in American secondary schools, ed. F.M. Newmann.). New York: Teachers College Press. 11-39 Pp.
- Olivier, E., Archambault, I., and Dupéré, V. (2018). Boys' and girls' latent profiles of behavior and social adjustment in school: Longitudinal links with later student behavioral engagement and academic achievement? *Journal of School Psychology.* 69: 28-44.
- Pintrich, P. R. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. [En línea]. Disponible en: <https://eric.ed.gov/?id=ED338122>. Fecha de consulta: 7 de febrero de 2018.
- Pintrich, P. R. and de-Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology.* 82(1): 33-40.
- Ramirez-Arellano, A., Acosta-Gonzaga, E., Bory-Reyes, J., and Hernández-Simón, L. M. (2018). Factors affecting student learning performance: A causal model in higher blended education. *Journal of Computer Assisted Learning.* 34(6): 807-815.
- Ramirez-Arellano, A., Bory-Reyes, J., and Hernández-Simón, L. M. (2019). Emotions, motivation, cognitive-metacognitive strategies, and behavior as predictors of learning performance in blended learning. *Journal of Educational Computing Research.* 57(2): 491-512.
- Reschly, A. L. and Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 3-19). Boston, MA: Springer.
- Reyes, N. S., Morales, J., Moya, J. G., Teran, C. E., Rodríguez, D. N. y Altamirano, G. C. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologías de Informação.* (E17): 258-266.
- Salmela-Aro, K. and Upadyaya, K. (2014). School burnout and engagement in the context of demands-resources model. *British Journal of Educational Psychology.* 84(1): 137-151.
- Siemens, G. and Gasevic, D. (2012). Guest editorial-learning and knowledge analytics. *Journal of Educational Technology y Society.* 15(3): 1-2.
- Skinner, E. A. and Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology.* 85(4): 571-581.
- Skinner, E., Furrer, C., Marchand, G., and Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology.* 100(4): 765-781.
- Skinner, E. A. and Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 21-44). Boston, MA: Springer.
- Sun, Z., Xie, K., and Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *The Internet and Higher Education.* 36: 41-53.

Symonds, J., Schoon, I., and Salmela-Aro, K. (2016). Developmental trajectories of emotional disengagement from schoolwork and their longitudinal associations in England. *British Educational Research Journal*. 42(6): 993-1022.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*. 48(6): 1273-1296.

Tanuar, E., Heryadi, Y., Lukas, Abbas, B., S., and Gaol, F. L. (2018). Using machine learning techniques to earlier predict student's performance. *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*. 85-89.

Urbina-Nájera, A. B. y Mora, J. de la C. (2017). Breve revisión de aplicaciones educativas utilizando Minería de Datos y Aprendizaje Automático. *Revista Electrónica de Investigación Educativa*. 19(4): 84-96.

Vytasek, J. M., Patzak, A., and Winne, P. H. (2020). Analytics for student engagement. In M. Virvou, E. Alepis, G. A. Tsihrintzis, and L. C. Jain (Eds.), *Machine Learning Paradigms: Advances in Learning Analytics* (pp. 23-48). Switzerland: Springer Nature.

Wang, F. H. (2017). An exploration of online behaviour engagement and achievement in flipped classroom supported by learning management system. *Computers y Education*. 114: 79-91.

Wang, M. T. and Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*. 28: 12-23.

Wong, A. and Chong, S. (2018). Modelling adult learners' online engagement behaviour: Proxy measures and its application. *Journal of Computers in Education*. 5(4): 463-479.

Yadav, S. K., Bharadwaj, B., and Pal, S. (2012). Data mining applications: A comparative study for predicting student's performance, in *ArXiv:1202.4815 [Cs]*. [En línea]. Disponible en: <http://arxiv.org/abs/1202.4815>. Fecha de consulta: 7 de abril de 2019.