

## Técnicas estadísticas y logro de aprendizaje: revisión bibliográfica

### *Statistical techniques and learning achievement: literature review.*

Lilian Daniela Suárez-Riveros<sup>a</sup>, Wilmer Pineda-Ríos<sup>b</sup>, Iván Mauricio Mendivelso-Ramírez<sup>c</sup>

<sup>a</sup>Maestranda en Ciencia de Datos, Matemática - Universidad Escuela Colombiana Ingeniería Julio Garavito. lilian.suarez@mail.escuelaing.edu.co. ORCID: 0000-0002-8329-0765.

<sup>b</sup>PhD (c) en Estadística, MSc en Matemáticas, Matemático - Universidad Nacional de Colombia. Docente catedra - Universidad Escuela Colombiana Ingeniería Julio Garavito. wilmer.pineda@escuelaing.edu.co. ORCID: 0000-0001-7774-951X.

<sup>c</sup>MSc en Antropología Social – Universidad de los Andes, Estadístico - Universidad Nacional de Colombia. Docente catedra - Universidad Escuela Colombiana Ingeniería Julio Garavito. ivan.mendivelso@escuelaing.edu.co

**Forma de citar:** Suárez-Riveros, L.D, Pineda-Ríos, W, Mendivelso-Ramírez, I.M. (2021), Técnicas estadísticas y logro de aprendizaje: revisión bibliográfica. *Eco Matemático*, 12 (2), 112-124

Recibido: 19/04/2021

Aceptado: 16/06/2021

#### Palabras clave

Mapeamiento  
informacional  
bibliográfico, modelos  
estadísticos, métodos  
estadísticos, logro de  
aprendizaje.

**Resumen:** El objetivo de este escrito fue describir las diferentes técnicas estadísticas que han sido empleados para comprender o explicar el logro de aprendizaje, en estudiantes en diferentes niveles educativos. Desde el punto de vista teórico se consolidaron las categorías a priori, provenientes de las técnicas estadísticas (Modelos Multinivel, Modelos geoespaciales, Regresión, Clustering, Análisis Descriptivo, Redes Neuronales, Árboles de decisión, Bosques aleatorios, NaiveBayes y Support Vector Machine), así como la conceptualización de Logro de Aprendizaje. El enfoque metodológico para la revisión se hizo a partir del mapeamiento informacional bibliográfico. Entre los resultados se encontraron 50 documentos de diferentes bases de datos (Elsevier (1), Google Scholar (6), IEEE (4), Scielo (2), ScienceDirect (5), Scopus (31), y Springer (1)), que estudian diferentes regiones del mundo (Asia (17), América del sur (13), América del norte (8), Europa (6), África (5), Oceanía (4), Centro América (3), junto con la orientación a explicar (17), comprender (31) o comprender y explicar (2). Adicionalmente, se identificó un conjunto de variables emergentes en los diferentes reportes, entre las que se encuentra, con mayor relevancia, el nivel socioeconómico, género, afectividad, antecedentes y características y posibilidades de los padres.

\*Autor para correspondencia: [lilian.suarez@mail.escuelaing.edu.co](mailto:lilian.suarez@mail.escuelaing.edu.co)

DOI: 10.22463/17948231.3323

2462-8794© 2021 Universidad Francisco de Paula Santander. Este es un artículo bajo la licencia CC BY 4.0

## Keywords

Bibliographic informational mapping, statistical models, statistical methods, learning achievement.

**Abstract:** The aim of this document was to describe the different statistical techniques that have been used to understand or explain the achievement of learning in Students at different educational levels. From the theoretical point of view, the a priori categories from statistical techniques were consolidated (Multilevel Models, Geospatial Models, Regression, Clustering, Descriptive Analysis, Neural Networks, Decision Trees, Random Forests, Naive Bayes and Support Vector Machine) as well as the conceptualization of Learning Achievement. The methodological approach for the review was based on the bibliographic informative mapping. Among the results are 50 documents from different databases (Elsevier (1), Google Scholar (6), IEEE (4), Scielo (2), Science Direct (5), Scopus (31), y Springer (1)), who study different regions of the world (Asia (17), South America (13), North America (8), Europe (6), Africa (5), Oceania (4), Central America (3)), along with the orientation to explain (17), understand (31) or understand and explain (2). Additionally, they identified a set of emerging variables in the different reports, among which the socioeconomic level, gender, affectivity, background and characteristics and possibilities of the parents are most relevant..

## Introducción

La Calidad de la Educación (CE) se ha venido considerando como un elemento estratégico en las políticas de gobierno de diferentes países, al que se asocian otras variables relacionadas con el cierre de brechas sociales, mejoramiento de la calidad de vida de las personas, entre otras. Al parecer existe una relación educación – salud, en razón a que cuando se incrementan los índices educativos, también hay un mejoramiento en la salud, el incremento en la productividad, la movilidad social, la reducción de la pobreza y en la construcción de ciudadanía (Comisión Económica para América Latina y el Caribe, 2016). Siendo el sistema educativo un sistema con responsabilidades multinivel, la evaluación debiera corresponder a una articulación de esa estructura, llevando a cabo procesos de certificación y todas las actividades necesarias; sin embargo, los resultados de la actividad educativa, al parecer, tienen un sesgo hacia los logros de los estudiantes en pruebas estandarizadas (George-Reyes, 2020). Además, desde la óptica docente, ha pasado a un proceso de transformación, con ocasión de la pandemia debido al COVID-19 (Suárez et al., 2021), hecho que adicionalmente, ha implicado la transformación del perfil docente (Suárez et al., 2020).

Algunos países, señalan que un elemento directamente asociado con la calidad de la educación (Urquizo, 2018) es la evaluación, específicamente los resultados a nivel de aula (Montagud-Mascarell & Gandía-Cabedo, 2014), del programa (Ruiz & Moya, 2020), niveles educativos, cuando se trata de procesos de certificación (Peláez-Valencia et al., 2020) de la calidad de la educación entre otros.

Las pruebas estandarizadas se constituyen en fuente de información de modelos estadísticos orientados a comprender-interpretar las variables que permiten explicar-predecir el logro de aprendizaje. Lo anterior, al parecer, puede ser utilizado para la concreción de estrategias, políticas y en general, rutas tendientes a mejorar la CE. La pregunta que se aborda en el presente escrito es ¿qué técnicas estadísticas han sido utilizadas para comprender o explicar el logro de aprendizaje de estudiantes en los diferentes niveles educativos? como una pregunta secundaria, y a la vez interesante, se propone ¿qué variables han emergido en los diferentes estudios que contribuyen en comprender o explicar los logros de aprendizaje de los estudiantes en diferentes niveles educativos?

## Elementos teóricos

En el ámbito educativo, la conceptualización de rendimiento académico, logro académico o Logro de Aprendizaje (LA) reviste algunas complejidades que no han permitido a la comunidad académica tener un consenso al respecto. Para Navarro (2003), el rendimiento académico no está solamente asociado con el conocimiento de un área específica, sino además, con todos los factores asociados que contribuyen a obtener los resultados. El LA se entiende como las puntuaciones obtenidas en las pruebas estandarizadas, cuyo resultado está asociado a los diferentes dimensiones contextuales y de desarrollo de la persona. En este artículo, se consideran principalmente tres técnicas estadísticas: modelos estadísticos, modelos computacionales y pruebas estadísticas.

## Metodología

El enfoque metodológico utilizado en el presente estudio se fundamenta en el Mapeamiento Informacional Bibliográfico (MIB), cuya finalidad se orienta a comprender e identificar las ideas del texto en forma sintética, es decir, se trata de reconfigurar en un nuevo y breve texto, el documento consultado, guardando la coherencia y sin perder de vista su objetivo, en pro de construir los antecedentes propios de un campo de investigación (Molina et al., 2012).

El proceso metodológico seguido en el desarrollo del estudio tuvo las siguientes etapas: Etapa 1 - Formulación de la pregunta de objetivos de investigación; Etapa 2 - Definición del periodo de tiempo para indagar documentos, Etapa 3 - Descargue, clasificación y verificación de pertinencia de los documentos; Etapa 4 - MBI; Etapa 6 - Reporte del estudio.

El instrumento utilizado como apoyo para la clasificación y aplicación del MIB, con las categorías a priori para el análisis, se elabora en hoja de cálculo y tiene la siguiente estructura:

- **No.** Número consecutivo para identificar cada documento.
- **Documento.** Información del documento con norma APA 7.0.
- **Modelos multinivel.** Esta técnica estadística asume que los datos están jerarquizados, las variables explicativas están en los diferentes niveles, en tanto que la variable dependiente, está en el nivel más bajo.
- **Geographically Weighted Regression (GWR)/ Inverse Distance Weighted (IDW) / Autocorrelación espacial.** Conjunto de técnicas estadísticas orientado a identificar patrones espaciales, en el conjunto de datos en estudio. GWR es una extensión de la regresión lineal, que tiene en cuenta la estructura espacial. IDW es un método de interpolación, donde se establece que el impacto de cada dato conocido, es inversamente proporcional a la distancia del dato desconocido.
- **Regresión.** Técnicas estadísticas utilizadas para predecir (continuas [regresión lineal], categóricas [regresión logística] o econometría [regresión cuantílica]), a partir de relaciones lineales, entre el destino y sus variables predictoras.
- **Clustering.** Se consideran dos tipos de algoritmos de Clustering, k-means es un algoritmo iterativo no supervisado, que agrupa los objetos en k-grupos y KNN es un algoritmo que busca las observaciones más cercanas para predecir y clasificar.
- **Análisis descriptivo e inferencial.** Conjunto de técnicas estadísticas en las que se encuentran pruebas estadísticas, medidas de tendencia central, análisis de correlación, entre otras.
- **Redes neuronales.** Es un modelo constituido por una estructura de neuronas interconectadas con

una señal de entrada y una señal de salida, inspirado en el sistema nervioso humano.

- **Árboles de decisión.** Son diagramas que dan probabilidades, mediante las cuales se determina una ruta de acción.

- **Bosques aleatorios.** Es un algoritmo supervisado, construido con árboles de decisiones, que resuelve predicciones y clasificaciones.

- **Naive Bayes.** Algoritmo supervisado clasificador, basado en la aplicación del teorema de Bayes.

- **Support Vector Machine (SVM).** Técnica de aprendizaje supervisado para clasificación y regresión, que mapea las observaciones originales en un espacio de mayor dimensión.

- **Modelos de Riesgo proporcional de Cox.** Modelo de supervivencia o semiparamétrico.

- **Análisis de ruta.** Método que posibilita evaluar modelos teóricos de relaciones causales entre variables.

- **Curva de Kuznets.** Representación gráfica de U invertida entre el desarrollo económico en el tiempo y la desigualdad, propuesta por Simon Kuznets, ampliada a otras variables.

Los documentos consultados en el presente estudio debían estar publicados en línea. Las bases de datos con convenio institucional a las que se tuvo acceso son Elsevier, Google Scholar, IEEE, Scielo, ScienceDirect, Scopus, y Springer. La distribución geográficamente están en orden de mayor a menor representación América del sur, Europa, Asia, América del norte, Centro América, África y Oceanía.

## Resultados y discusión

Esta sección se ha organizado en dos partes, la primera relacionada con los documentos a partir de las variables a priori y la segunda, organizada desde las variables emergentes.

### Las variables a priori

En la producción académica, los modelos de regresión son los más utilizados en la predicción del LA. Apoyarse en el análisis descriptivo e inferencial, derivado del procesamiento estadístico de información, es el segundo elemento epistemológico utilizado de preferencia en la comprensión del LA y en algunos casos, como complemento al modelamiento estadístico predictivo del LA. Los árboles de decisión, las redes neuronales y los modelos multinivel, se constituyen en las terceras técnicas estadísticas de preferencia usadas para la comprensión o explicación del LA. Los árboles de decisión han sido utilizados con un enfoque principal de comprensión en el LA, así como articulado con otros modelamientos estadísticos para la explicación del LA (Cornell-Farrow & Garrard, 2020; Rebai et al., 2019; Wandera et al., 2019; Xu et al., 2019; Yang et al., 2020).

### Las variables emergentes en el Logro de Aprendizaje

En el conjunto de estudios consultados, las variables emergentes se organizan acorde a su recurrencia, en tres partes: alto nivel, medio y bajo nivel de coincidencia.

Tabla I. Conjunto de variables con alto nivel de recurrencia en los estudios.

Variable emergente	Descripción	Trabajos
Nivel socio-económico	Esta variable hace referencia a los ingresos familiares. En algunos países, se vincula con el estrato socioeconómico en el que está ubicada la vivienda.	(Benito et al., 2014; Chacón-Vargas & Roldán-Villalobos, 2021; Contreras et al., 2019; Lisboa- Bartholo & Da-Costa, 2016; Martínez-Mateus & TurriagoHoyos, 2015; Masci et al., 2018; Maulida & Kariyam, 2017; Rodríguez-Hernández et al., 2021; Salal & Abdullaev, 2020; Shah et al., 2019; Zhang & Campbell, 2014)
Género	Referida a la clasificación biológica de los participantes, en hombres o mujeres.	(Chacón-Vargas & Roldán-Villalobos, 2021; Gaeta-Rivas et al., 2021; Kumari et al., 2018; Lau et al., 2019; Lisboa- Bartholo & Da-Costa, 2016; Orjuela, 2014; Qazdar et al., 2019; Rodríguez-De-Souza-Pajuelo et al., 2021; Shah et al., 2019)
Afectividad	El afecto es una noción compleja, en los estudios se presenta en la motivación intrínseca (realizar acciones por la satisfacción, sin incentivos externos) o extrínseca (generada por estímulos externos), que pueden relacionarse con otras actividades académicas o desarrollo personal.	(Cvencek et al., 2017; Dagnew, 2017; Froiland & Oros, 2014; Giannakas et al., 2021; Jovanović et al., 2021; Mineshita et al., 2021; Pollak & Parnell, 2018; Rebai et al., 2019; Rodríguez-De-Souza-Pajuelo et al., 2021; Xu et al., 2019; Yang et al., 2020)
Antecedentes	Calificaciones que han obtenido los	(Chacón-Vargas & Roldán-Villalobos, 2021; Cornell-Farrow & Garrard, 2020;
	estudiantes en mediciones previas a las consideradas en el estudio.	Febro, 2019; Hasan et al., 2018; Kumari et al., 2018; Lau et al., 2019; Maisarah-Samsudin et al., 2021; Maulida & Kariyam, 2017; Qazdar et al., 2019; Rebai et al., 2019; Rodríguez-De-Souza-Pajuelo et al., 2021; Rodríguez-Hernández et al., 2021; Tapasco-Alzate et al., 2020; Wang et al., 2019)
Padres	En esta variable se identifican las características de los padres, en cuanto al nivel educativo y la ocupación.	(Chacón-Vargas & Roldán-Villalobos, 2021; Cornell-Farrow & Garrard, 2020; Kumari et al., 2018; Lau et al., 2019; Masci et al., 2018; Maulida & Kariyam, 2017; Qiu & Wu, 2019; Salal & Abdullaev, 2020; Shah et al., 2019)

En la Tabla I se encuentran las variables que han sido estudiadas con mayor recurrencia; en diversos estudios el nivel socio-económico se presenta como una variable que influye en el logro de aprendizaje, en áreas como la matemática y la lectura (Qiu & Wu, 2019), al parecer las condiciones de las familias se asocia con el hacinamiento (Contreras et al., 2019). Un hecho relevante señala, que el nivel socio-económico es una variable significativa en los modelos que se aproximan a comprender o explicar, excepto en Finlandia (Benito et al., 2014), país que ocupa puesto privilegiado en sectores económicos, con indicadores de pobreza por debajo de 12% y en general, un índice de calidad de vida favorable.

La evidencia señala que educación básica o media los hombre tienen mejores resultados en matemáticas (Orjuela, 2014), en tanto que a nivel de pregrado son las mujeres (Lau et al., 2019) y a nivel de posgrado no hay diferencias significativas. En lo afectivo la motivación intrínseca está relacionada con el LA (Rodríguez-De-Souza-Pajuelo et al., 2021), en condiciones de la pandemia la enseñanza depende de los profesores y

estudiantes tengan estabilidad socioemocional (De Agüero Servín et al., 2021). Los antecedentes académicos de los estudiantes al parecer es un factor predictivo en la aprobación de los cursos (Lau et al., 2019), en general en el éxito académico en el pregrado (Chacón-Vargas & Roldán-Villalobos, 2021) y es contrario con la deserción (Febro, 2019). Finalmente, la estabilidad económica (Qiu & Wu, 2019) y la formación profesional de la familia (Chacón-Vargas & Roldán-Villalobos, 2021) se relacionan con el LA.

Tabla II. Conjunto de variables con nivel medio de recurrencia en los estudios.

Variable Emergente	Descripción	Trabajos
Ubicación	Esta característica hace referencia a la georreferenciación de la institución; en específico, está ubicada en el casco urbano de la (s) ciudad (es) o en la zona rural.	(Ariza et al., 2021; Froiland & Oros, 2014; Lau et al., 2019; Murillo & Carrillo, 2021; Rebai et al., 2019; Wandera et al., 2019)
Recursos materiales	Las encuestas o entrevistas de los estudios consideran en la caracterización de la población estudiada, la existencia de electrodomésticos, teléfono, computador, conexión a internet, medio de transporte (automóvil/moto), acceso a bibliotecas, número de habitaciones de la vivienda, entre otras.	(Ariza et al., 2021; Chacón-Vargas & Roldán-Villalobos, 2021; Contreras et al., 2019; Salai & Abdullaev, 2020; Wandera et al., 2019; Xu et al., 2019)
Relaciones con pares	Esta variable hace referencia a la forma en cómo interactúan los estudiantes con otros estudiantes, en cuanto a la aceptación o rechazo, apoyo u oposición a su estancia, derivados de, entre otros, la diversidad cultural.	(Jovanović et al., 2021; Maulida & Kariyam, 2017; Sbroglio-Rizzotto & Aniceto-França, 2020; Wang et al., 2019)
Características institucionales	Condiciones estructurales y de personal de las instituciones escolares, como calendario del colegio, tipo (privado/público), jornada, tamaño de la clase, tamaño de la escuela, infraestructura, entre otras.	(Ariza et al., 2021; Castrillón et al., 2020; Lisboa-Bartholo & Da-Costa, 2016; Masci et al., 2018; Rebai et al., 2019; Rodríguez-Hernández et al., 2021; Zhang & Campbell, 2014)
Experiencia docente	Las habilidades y destrezas adquiridas como consecuencia de sus actividades propias de la docencia.	(Ariza et al., 2021; Castrillón et al., 2020; Khan & Ghosh, 2018; Lisboa-Bartholo & Da-Costa, 2016; Qiu & Wu, 2019; Zhang & Campbell, 2014)

En la Tabla II se encuentran las variables que, en forma media, tienen coincidencias en los estudios consultados. En este nivel se encuentran la ubicación geográfica de la institución, la disponibilidad de recursos materiales, relaciones con pares, las características institucionales y la experiencia docente.

En regiones como Sudáfrica, se encuentra que la georreferenciación de las instituciones puede influir en el LA obtenido por los estudiantes (Wandera et al., 2019), hecho que no es generalizable, dado que hay estudios en los que no se logra identificar diferencias significativas (Lau et al., 2019; Rebai et al., 2019). En países donde existen disparidades regionales en el acceso a la educación, es posible que éstas se constituyan en un factor que afecta las puntuaciones de los estudiantes, por lo cual esta variable es un elemento a

profundizar en siguientes investigaciones, asociado con la gestión que las instituciones y las regiones realizan en sectores que se han identificado, o se prevén, están vinculados con la educación.

La presencia de recursos en el contexto cercano o de vivienda de los estudiantes, puede favorecer el desempeño de estos y en consecuencia, de las escuelas (Wandera et al., 2019). Los estudiantes que disponen de solo computador logran menores puntuaciones en diversas áreas, por el contrario, los que tienen mayor interacción computador-internet, incrementan sus puntuaciones (Ariza et al., 2021; Chacón-Vargas & Roldán-Villalobos, 2021). La línea de investigación de la incorporación de los recursos tecnológicos asociados al acceso de la internet, sigue teniendo elementos a estudiar.

La relación entre estudiantes a nivel de pares afecta el LA. Aquellos estudiantes que migran hacia las instituciones (Wang et al., 2019) o sufren hostigamiento (Sbroglio-Rizzotto & Aniceto-França, 2020), de alguna de las formas posibles, se ven afectados negativamente en el LA. Las

condiciones en las cuales se dan las relaciones personales entre pares, es elemento a considerar desde los diferentes puntos de vista, toda vez que tienen un efecto en el LA.

La educación se entiende como una actividad compleja, en la que las características institucionales en su contexto, como la formación de los profesores, los horarios y las relaciones interpersonales entre los diferentes actores y el complemento de actividades extracurriculares, la eficiencia de la gestión institucional, tienden a influir en el LA; (Ariza et al., 2021; Castrillón et al., 2020). De acuerdo con lo anterior, ver las instituciones como un conjunto de personas que requieren de un entorno adecuado para el desarrollo de la actividad educativa, parece influir positivamente en el LA de los estudiantes.

El docente, no solamente en su formación, sino también en el tiempo dedicado a liderar procesos educativos, puede influir positivamente en la obtención del LA (Ariza et al., 2021; Castrillón et al., 2020; Lisboa- Bartholo & Da-Costa, 2016; Qiu & Wu, 2019; Zhang & Campbell, 2014).

**Tabla III.** Conjunto de variables con bajo nivel de recurrencia en los estudios.

Variable Emergente	Descripción	Trabajos
Tipo de institución	Hace referencia a la característica de la Institución, en cuanto a si es pública, privada o alianza público-privada.	(Contreras et al., 2019; Murillo & Carrillo, 2021; Orjuela, 2014)
Habilidades cognitivas	Procesos relacionados con la mente, específicamente en los referidos y necesarios para el logro de aprendizaje.	(Froiland & Oros, 2014; Schuth et al., 2017)
Refuerzo de alimentación	Alimentación de los estudiantes en las jornadas escolares.	(Gaete-Rivas et al., 2021)
Composición familiar	Personas con las que regularmente vive el estudiante en casa.	(Chacón-Vargas & Roldán-Villalobos, 2021; Masci et al., 2018; Qiu & Wu, 2019; Salal & Abdullaev, 2020)
Edad	Corresponde al tiempo registrado de vida de los participantes, desde su nacimiento.	(Cornell-Farrow & Garrard, 2020; Shah et al., 2019)

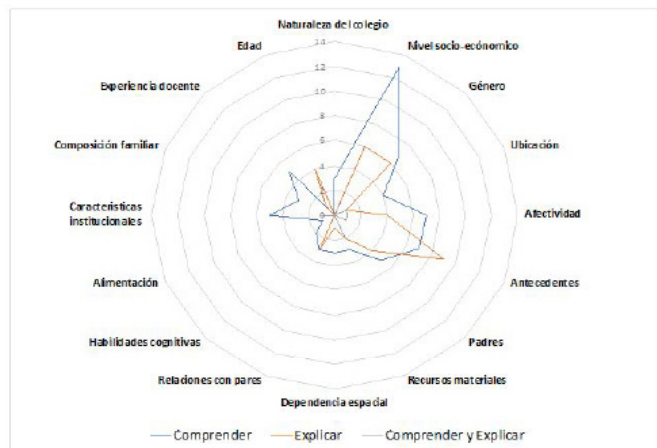


Figura 1. Variables emergentes en estudios de logro de aprendizaje.

La Figura 1 muestra las diferentes variables emergentes, organizadas de acuerdo a la intención con la que se incorporaron en las diferentes técnicas estadísticas. Algunas técnicas estadísticas se orientaron a comprender e interpretar el LA, a partir de las variables que emergieron en el estudio, otros modelos estadísticos, tuvieron como propósito la predicción del LA, a partir de algunas variables a priori, o que emergieron en el estudio realizado. Se puede identificar cómo el nivel socioeconómico es la variable más recurrente en la comprensión del LA. Los antecedentes previos de puntuaciones de los estudiantes, al parecer, contribuyen a predecir y explicar, en forma reiterada, el LA.

## Conclusiones

Los modelos estadísticos han sido utilizados para comprender – interpretar y explicar – y predecir el logro de aprendizaje. La revisión adelantada permite ver la riqueza y variedad de los modelos estadísticos utilizados en investigaciones, orientadas a develar las variables relacionadas con el logro de aprendizaje. El modelamiento estadístico es una fortaleza epistemológica, que contribuye a comprender la complejidad de la educación.

La recurrencia es distinta de la relevancia, en cuanto a las variables emergentes que se orientan a comprender o explicar el logro de aprendizaje. En los diferentes estudios consultados se encuentra un

nivel de recurrencia para las diferentes variables; de manera unánime no se tiene certeza de su incidencia simétrica, en las diferentes comunidades en las que se ha hecho el estudio. Luego del estudio del logro de aprendizaje, se identifican las variables emergentes, pero dado que éstos no son comparables (diferente muestra, interés de investigación, metodología, alcance entre otras), no es posible aproximarse a una generalización.

Al parecer, la elección de los métodos estadísticos en las diversas investigaciones reportadas y consultadas, están relacionadas con las preferencias de los investigadores y asociadas a los datos y tipos de datos disponibles para los modelos.

La revisión adelantada, permite evidenciar áreas de interés común en la comunidad académica de diferentes regiones, que, de alguna manera, se constituyen en campos que invitan a continuar las investigaciones y ampliar la cartografía de las variables que se deben considerar, en este caso, para la comprensión y predicción del logro de aprendizaje.

El caso de la variable nivel socioeconómico, que se identifica con un peso negativo en diversos estudios, excepto en Finlandia, podría ser evidencia que indique que, para tener mejores resultados en educación, entendidos desde los logros de aprendizaje, es una condición necesaria que la sociedad posea un nivel socioeconómico parecido al de éste país en referencia. Lo anterior daría entonces argumentos para que los países que hacen la apuesta a la calidad de la educación, tengan como meta la transformación del nivel socioeconómico de su sociedad.

## Agradecimientos

El equipo de investigación agradece a la Universidad Escuela Colombiana de Ingeniería Julio Garavito, a través de la Maestría en Ciencia de Datos, por su apoyo al proyecto de investigación, del cual se deriva el presente documento.



## Referencias

- Ariza, J., Saldarriaga, J., Reinoso, K., & Tafur, C. (2021). Tecnologías de la información y la comunicación y desempeño académico en la educación media en Colombia. *Lecturas de Economía*, 94, 47–86. <https://doi.org/10.17533/udea.le.n94a338690>
- Benito, R., Alegre, M., & González-Balletbò, I. (2014). School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems. *Comparative Education Review*, 58(1), 104–134. <https://doi.org/10.1086/672011>
- Castrillón, O. D., Sarache, W., & Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación universitaria*, 13(1), 93–102. <https://doi.org/10.4067/s0718-50062020000100093>
- Chacón-Vargas, É., & Roldán-Villalobos, G. (2021). Factores que inciden sobre el rendimiento académico de los estudiantes de primer ingreso del curso Matemática General del Instituto Tecnológico de Costa Rica. *Uniciencia*, 35(1), 265–283. <https://doi.org/10.15359/ru.35-1.16>
- Comisión Económica para América Latina y el Caribe. (2016). Desarrollo Social Inclusivo, una nueva generación de políticas para superar la pobreza y reducir la desigualdad en América Latina y Caribe. *Cepal*, 304. <https://www.cepal.org/es/publicaciones/39100-desarrollo-social-inclusivo-nueva-generacion-politicas-superar-la-pobreza>
- Contreras, D., Delgadillo, J., & Riveros, G. (2019). Is home overcrowding a significant factor in children's academic performance? Evidence from Latin America. *International Journal of Educational Development*, 67, 1–17. <https://doi.org/10.1016/j.ijedudev.2019.01.006>
- Cornell-Farrow, S., & Garrard, R. (2020). Machine learning classifiers do not improve the prediction of academic risk: Evidence from Australia. *Communications in Statistics Case Studies Data Analysis and Applications*, 6(2), 228–246. <https://doi.org/10.1080/23737484.2020.1752849>
- Cvencek, D., Fryberg, S., Covarrubias, R., & Meltzoff, A. (2017). Self-Concepts, Self-Esteem, and Academic Achievement of Minority and Majority North American Elementary School Children. *Child Development*, 89(4), 1099–1109. <https://doi.org/10.1111/cdev.12802>
- Dagnew, A. (2017). The relationship between students' attitudes towards school, values of education, achievement motivation and academic achievement in gondar secondary schools, Ethiopia. *Research in Pedagogy*, 7(1), 30–42. <https://doi.org/10.17810/2015.46>
- De Agüero Servín, M., Benavides Lara, M. A., Rendón Cazales, J., Pompa Mansilla, M., Hernández-Romo, A. K., Hernández-Martínez, A. M. del P., & Sánchez-Mendiola, M. (2021). Los retos educativos durante la pandemia de COVID-19: segunda encuesta a profesoras y profesores de la UNAM. *Revista de Investigación Educativa*, 22(5). <https://doi.org/10.22201/cuaieed.16076079e.2021.22.5.13>
- Febro, J. (2019). Utilizing feature selection in identifying predicting factors of student retention. *International Journal of Advanced Computer Science and Applications*, 10(9), 269–274. <https://doi.org/10.14569/ijacsa.2019.0100934>
- Froiland, J., & Oros, E. (2014). Intrinsic motivation, perceived competence and classroom engagement as longitudinal predictors of adolescent reading achievement. *Educational Psychology*, 34(2), 119–132. <https://doi.org/10.1080/01443410.2013.822964>

- Gaete-Rivas, D., Olea, M., Meléndez-Illanes, L., Granfeldt, G., Sáez, K., Zapata-lamana, R., & Cigarroa, I. (2021). Hábitos alimentarios y rendimiento académico en escolares chilenos de quinto a octavo año básico. *Revista chilena de nutrición*, 48(1), 41–50. <https://doi.org/10.4067/S0717-75182021000100041>
- George-Reyes, C. (2020). Pruebas Estandarizadas Y Calidad De La Educacion En México. *Universidad y Sociedad Revista científica de la Universidad de Cienfuegos*, 12(4), 418–425. <http://scielo.sld.cu/pdf/rus/v12n4/2218-3620-rus-12-04-418.pdf>
- Giannakas, F., Troussas, C., Voyiatzis, I., & Sgouropoulou, C. (2021). A deep learning classification framework for early prediction of team-based academic performance. *Applied Soft Computing*, 106. <https://doi.org/10.1016/j.asoc.2021.107355>
- Hasan, R., Palaniappan, S., Rafiez-Abdul, A., Mahmood, S., & Uddin-Sarker, K. (2018). Student Academic Performance Prediction by using Decision Tree Algorithm. *4th International Conference on Computer and Information Sciences (ICCOINS)*, 1–5. <https://doi.org/10.1109/ICCOINS.2018.8510600>
- Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers and Education*, 172(April), 1–13. <https://doi.org/10.1016/j.compedu.2021.104251>
- Khan, A., & Ghosh, S. (2018). Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, 23(4), 1677–1697. <https://doi.org/10.1007/s10639-017-9685-z>
- Kumari, P., Jain, P., & Pamula, R. (2018). An Efficient use of Ensemble Methods to Predict Students Academic Performance. *4th Int'l Conf. on Recent Advances in Information Technology*. <https://doi.org/10.1109/RAIT.2018.8389056>
- Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1(982). <https://doi.org/10.1007/s42452-019-0884-7>
- Lisboa- Bartholo, T., & Da-Costa, M. (2016). Evidence of a school composition effect in Rio de Janeiro public schools. *Ensaio*, 24(92), 498–521. <https://doi.org/10.1590/S0104-40362016000300001>
- Maisarah-Samsudin, N., Milleana-Shaharudin, S., Filza-Sulaiman, N., Mohd-Fuad, M., Fareezuan-Zulfikri, M., & Hila-Zainuddin, N. (2021). Modeling student's academic performance during Covid-19 based on classification in support vector machine. *Turkish Journal of Computer and Mathematics Education*, 12(5), 1798–1804. <https://doi.org/10.17762/turcomat.v12i5.2190>
- Martínez-Mateus, W., & TurriagoHoyos, Á. (2015). Análisis de distribución geográfica y espacial de los resultados de las Pruebas Saber 11 del Instituto Colombiano para el Fomento de la Educación Superior -ICFES-. 2005-2012. Colombia. *Cuadernos Latinoamericanos de Administración*, 11(21), 39–50. <https://doi.org/10.18270/cuaderlam.v11i21.1618>
- Masci, C., Johnes, G., & Agasisti, T. (2018). Student and school performance across countries: A machine learning approach. *European Journal of Operational Research*, 269(3), 1072–1085. <https://doi.org/10.1016/j.ejor.2018.02.031>
- Maulida, J., & Kariyam. (2017). Students academic performance based on behavior. *AIP Conference*

- Proceedings*, 1911(December 2017). <https://doi.org/10.1063/1.5016003>
- Mineshita, Y., Kim, H., Chijiki, H., Nanba, T., Shinto, T., Furuhashi, S., Oneda, S., Kuwahara, M., Suwama, A., & Shibata, S. (2021). Screen time duration and timing: effects on obesity, physical activity, dry eyes, and learning ability in elementary school children. *BMC Public Health*, 21(422). <https://doi.org/10.1186/s12889-021-10484-7>
- Molina, A., Pérez, M., Castaño, N., Bustos, E., Suárez, O., & Sánchez, M. (2012). Mapeamiento informacional bibliográfico en el campo de la enseñanza de las ciencias, contexto y diversidad cultural: el caso del Journal Cultural Studies in Science Education (CSSE). *Revista EDUCyT, Extraordin*, 1997–222.
- Montagud-Mascarell, M. D., & Gandía-Cabedo, J. L. (2014). Virtual learning environment and academic outcomes: Empirical evidence for the teaching of Management Accounting. *Revista de Contabilidad-Spanish Accounting Review*, 17(2), 108–115. <https://doi.org/10.1016/j.rcsar.2013.08.003>
- Murillo, J., & Carrillo, S. (2021). Incidencia de la Segregación Escolar por Nivel Socioeconómico en el Rendimiento Académico. *Un Estudio desde Perú. Archivos analíticos de políticas educativas*, 29(49), 3–11. <https://doi.org/10.14507/epaa.29.5129>
- Navarro, R. E. (2003). EL rendimiento académico: concepto, investigación y desarrollo. REICE - *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 1–15. <https://doi.org/2152>
- Orjuela, J. (2014). Análisis del Desempeño Estudiantil en las Pruebas de Estado para Educación Media en Colombia mediante Modelos Jerárquicos Lineales. *Ingeniería*, 18(2). <https://doi.org/10.14483/udistrital.jour.reving.2013.2.a04>
- Peláez-Valencia, L. E., Trefftz, H., & Delgado-González, I. A. (2020). Acreditación Internacional de Carreras de Ingeniería. *Educación en Ingeniería*, 15(29), 28–33. <https://doi.org/dx.doi.org/10.26507/rei.v15n29.1044>
- Pollak, M., & Parnell, D. (2018). An Interdisciplinary Analysis of Course Meeting Frequency, Attendance and Performance. *Journal of the Scholarship of Teaching and Learning*, 18(3), 132–152. <https://doi.org/10.14434/josotl.v18i3.23752>
- Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammass, D. (2019). A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, 24(6), 3577–3589. <https://doi.org/10.1007/s10639-019-09946-8>
- Qiu, X., & Wu, S. sheng. (2019). Contextual variables of student math proficiency and their geographic variations in Missouri. *Applied Geography*, 109, 102040. <https://doi.org/10.1016/j.apgeog.2019.102040>
- Rebai, S., Ben Yahia, F., & Essid, H. (2019). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-Economic Planning Sciences*, 70(August 2018), 100724. <https://doi.org/10.1016/j.seps.2019.06.009>
- Rodríguez-De-Souza-Pajuelo, A. A., Tarazona-Luján, A. F., & Reyes-Bossio, M. (2021). Physical activity enjoyment and self-efficacy in school performance of 11-17-year-old students at educational institutions in Lima. *Journal of Physical Education and Sport*, 21(3), 2183–

2189. <https://doi.org/10.7752/jpes.2021.s3278>
- Rodríguez-Hernández, C., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2, 100018. <https://doi.org/10.1016/j.caeai.2021.100018>
- Ruiz, J., & Moya, S. (2020). Evaluación de las competencias y de los resultados de aprendizaje en destrezas y habilidades en los estudiantes de Grado de Podología de la Universidad de Barcelona. *Educacion Medica*, 21(2), 127–136. <https://doi.org/10.1016/j.edumed.2018.08.007>
- Salal, Y., & Abdullaev, S. (2020). Deep Learning based Ensemble Approach to Predict Student Academic Performance: Case Study. *En 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. <https://doi.org/10.1109/ICISS49785.2020.9316044>
- Sbroglio-Rizzotto, J., & Aniceto-França, M. (2020). Does Bullying Affect the School Performance of Brazilian Students? An Analysis Using Pisa 2015. *Child Indicators Research*, 14, 1027–1053. <https://doi.org/10.1007/s12187-020-09790-0>
- Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, 49, 157–165. <https://doi.org/10.1016/j.learninstruc.2017.01.005>
- Shah, M., Kaistha, M., & Gupta, Y. (2019). Student Performance Assessment and Prediction System using Machine Learning. 4th International Conference on Information Systems and Computer Networks, *ISCON 2019*, 386–390. <https://doi.org/10.1109/ISCON47742.2019.9036250>
- Suárez, O., Hernández-Barbosa, R., Lizarazo-Osorio, J., & Orjuela-Osorio, C. (2021). La evaluación en tiempos de covid -19 : una mirada desde los docentes. *Revista Academia y Virtualidad*, 14(2), 31–43. <https://doi.org/https://doi.org/10.18359/ravi.5365>
- Suárez, O., Molina-Vasquez, R., Orjuela, P., & Lizarazo, J. (2020). Elementos para la construcción de un modelo de tutor virtual. Editorial Fundación Universidad Autónoma de Colombia.
- Tapasco-Alzate, O., Ruiz-Ortega, F., Osorio-García, D., & Ramírez-Ramírez, D. (2020). El historial académico de secundaria como factor predictor del rendimiento universitario. *Caso de estudio. Revista Colombiana de Educación*, 1(81), 147–169. <https://doi.org/10.17227/rce.num81-7530>
- Urquizo, H. G. (2018). Propuesta de medición y evaluación de Resultados de Aprendizaje según criterios de ABET y ASIIN. Proceedings of the LACCEI international Multi-conference for Engineering, *Education and Technology*, 2018-July(May). <https://doi.org/10.18687/LACCEI2018.1.1.435>
- Wandera, H., Marivate, V., & Sengeh, M. (2019). Predicting national school performance for policy making in South Africa. *6th International Conference on Soft Computing and Machine Intelligence, ISCOMI 2019*, 23–28. <https://doi.org/10.1109/ISCOMI47871.2019.9004323>
- Wang, Y., Pei, F., Zhai, F., & Gao, Q. (2019). Academic performance and peer relations among rural-to-urban migrant children in Beijing: Do social identity and self-efficacy matter? *Asian Social Work and Policy Review*, 13(3), 263–273. <https://doi.org/10.1111/aswp.12179>
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine

learning algorithms. *Computers in Human Behavior*, 98(April), 166–173. <https://doi.org/10.1016/j.chb.2019.04.015>

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y.-M., & Lim, H. (2020). Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, 24(24), 18777–18793. <https://doi.org/10.1007/s00500-020-05110-4>

Zhang, D., & Campbell, T. (2014). An examination of the impact of teacher quality and “Opportunity Gap” on student Science Achievement in China. *International Journal of Science and Mathematics Education*, 13(3), 489–513. <https://doi.org/10.1007/s10763-013-9491-z>