ISSN 1870-4069

Aplicación del algoritmo K-means para el análisis de resultados de la prueba PLANEA 2017

Israel Gutiérrez González, Doricela Gutiérrez Cruz,
Jenny Elizabeth Juan Ramírez, Liliana Rodríguez Páez,
Ricardo Rico Molina, Mauricio Sánchez Medina

Universidad Autónoma del Estado de México,
Centro Universitario UAEM Nezahualcóyotl,
México

igutierrezg@uaemex.mx

Resumen. El uso de técnicas de minería de datos en el estudio de algunos de los factores involucrados en el rendimiento académico es actualmente un campo de trabajo que se encuentra permeando en los temas de investigación educativa a nivel internacional. Entre estas técnicas se encuentran el algoritmo K-Means el cual es un algoritmo de aprendizaje no supervisado y que se utiliza en este trabajo para analizar los resultados de la prueba PLANEA 2017 de escuelas de nivel medio superior en cada una de los Estados de la República Mexicana en el Área de Lenguaje y Comunicación. En la presente investigación, se analizaron los factores entidad, turno, financiamiento y nivel de marginación de las escuelas evaluadas y se encuentra que hay un importante impacto de los factores estudiados sobre el aprovechamiento por parte de los estudiantes.

Palabras clave: Rendimiento académico, K-Means, PLANEA 2017.

Application of the K-Means Algorithm for the Analysis of PLANEA 2017 Test Results

Abstract. The use of data mining techniques in the study of some of the involved factors in academic achievement is currently a work area that is permeating international educational research topics. Among these techniques is the K-Means algorithm, which is an unsupervised learning algorithm and is used in this work to analyze the results of the PLANEA 2017 test of the high school in each of the States of the Mexican Republic in the Language and Communication Area. In the present investigation, the factors entity, shift, financing and marginalization level of the evaluated schools, were analyzed and it was found that there is an important impact of the studied factors on the students' achievement.

Keywords: Academic achievement, K-Means, PLANEA 2017.

1. Introducción

La evaluación del rendimiento académico es uno de los principales indicadores que toman importancia fundamental dentro de los sistemas educativos a nivel mundial,

donde no solo se mide la efectividad de un sistema educativo, sino que apoyan a los docentes a mejorar la enseñanza y los procesos de evaluación encaminados a alcanzar una educación de calidad [1] e involucrando al estudiante en términos de desarrollo de capacidades o de adquisición de aptitudes generales, así como de competencias curriculares, entre otros [2]. Debido a esto, en el plano internacional, existen diversos organismos especializados en la evaluación y cuyos resultados tienen un impacto importante en las políticas estatales en cuanto a educación se refiere.

En este sentido, la Organización para la Cooperación y el Desarrollo Económico (OCDE) la cual es un organismo internacional que, entre otras funciones, brinda asistencia a sus miembros (México forma parte de esta organización) para la planificación y gestión de sus sistemas educativos, ha establecido la necesidad latente de una educación de calidad como elemento sustancial para mejorar el bienestar social y económico de las personas en todo el mundo [3]. Con miras a comprender el estado de la educación en los países, este organismo se ha encargado de realizar una evaluación trienal mediante el Programa para la Evaluación Internacional de Alumnos (PISA por sus siglas en inglés) la cual consiste en la aplicación de una prueba para evaluar los conocimientos y habilidades que han adquirido los alumnos de entre 15 y 16 años.

En México, también se cuenta con organismos encargados de medir el impacto educativo a través de la aplicación de pruebas a distintos niveles de educación como son el Instituto Nacional para la Evaluación de la Educación (INEE), y con el desarrollo de nuevas iniciativas de la Secretaria de Educación Pública (SEP), en particular con la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE) [4]. Los resultados que se obtienen de estas pruebas normalmente no son sometidos a un riguroso análisis estadístico y así mismo, estos datos pueden contener una enorme riqueza de información que, con la metodología adecuada, puede ser aplicada para la obtención de patrones interesantes que permitan explicar algunos comportamientos. En este sentido la minería de datos permite aplicar diversos algoritmos de agrupamiento como son COBWEB, EM, y K-means [5]. En el caso de K-means se han utilizado en la predictividad de proceso petitorio en una organización pública [6]. Así como para analizar la ejecución de procesos de negocios [7]. Además, para la identificación de los perfiles de estudiantes [8].

Por lo tanto, el objetivo de este trabajo fue aplicar el algoritmo K-Means para el análisis de resultados de la prueba PLANEA 2017.

2. Trabajos relacionados

Una cantidad importante de trabajos han sido realizados por investigadores del área, de los cuáles se han obtenido resultados que han impactado la forma de crear o transformar las políticas públicas encaminadas al mejoramiento de la educación. El enorme esfuerzo que se requiere para lograr esta transformación ha ido involucrando a su paso a especialistas tales como psicólogos, sociólogos, filósofos, pedagogos, matemáticos, y especialistas del área de la informática y del análisis de datos.

Estos estudiosos del tema han descrito de forma cualitativa y cuantitativa la forma en que, directa o indirectamente, factores de índole social, económico y político impactan sobre el rendimiento académico.

Al respecto, puede mencionarse el trabajo coordinado por Raúl Abreu y David Calderón [9], en donde en su informe del Índice Compuesto de Eficacia de los sistemas Escolares 2007 en México, analizaron algunas variables que resultan relevantes en el aprovechamiento educativo a nivel básico encontrando que la infraestructura de las escuelas, el ausentismo de los profesores y la incorporación de los maestros a la carrera magisterial son factores escolares que presentan una correlación significativa con el logro académico de los estudiantes.

Por otra parte, De Hoyos et al. [10] establecen la importancia que, los antecedentes y recursos escolares, factores institucionales y antecedentes familiares tienen sobre la dinámica del aprendizaje en estudiantes de México. Otros factores estudiados son los institucionales, características personales, ambiente del hogar antecedentes familiares del estudiante, los cuales se cree que son factores asociados al logro cognitivo [11-13]. Todd & Wolpin [11] establecen que la acumulación de conocimiento y logros académicos son el resultado de un proceso acumulativo que depende de la historia y herencia familiares, así como de los insumos escolares.

Así mismo, Glewwe y Kremer [12] definen diferentes actores asociados con el desempeño académico dentro de las instituciones de educación tales como los incentivos a la práctica y el ausentismo docentes. Particularmente en matemáticas y ciencias, en un análisis realizado sobre los resultados de la prueba PISA, Fuchs y Woessman [13] aseguran que existe una importante correlación entre factores tales como el nivel de estudios y la ocupación de los padres, así como el número de libros que se leen en la casa con el logro académico en esas áreas. Hanushek [14] por su parte, establece que los resultados en cuanto a la eficiencia cognitiva se refieren, están en función de las políticas públicas que el Estado implementa y en particular en función de la inversión que se hace en materia de Educación.

Una forma de abordar el análisis de los factores involucrados en la efectividad de los procesos de enseñanza-aprendizaje (como aquellos factores mencionados con anterioridad), es mediante el uso de técnicas de manejo de la información propias de la minería de datos que, en su modalidad de aplicación a los temas educativos, se conoce como minería de datos educativa (EDM) y que, desde ya hace algunos años, se encuentra realizando importantes análisis así como aportando soluciones de los más relevantes problemas educativos [15, 16]. En particular, se ha hecho un uso reiterado de las técnicas propias del aprendizaje automático o aprendizaje de máquina, las cuales permiten la búsqueda de información relevante o de patrones que proporcionan una mejor comprensión del comportamiento de los resultados y eventualmente pueden conducir a la predicción de dichos comportamientos.

En la literatura se pueden encontrar ejemplos de la aplicación de técnicas de minería de datos en investigación educativa [17-19], tales como el uso de algoritmos de agrupamiento (los cuales entran dentro de las técnicas del aprendizaje no supervisado) para el análisis de datos en investigación educativa tales como el estudio realizado por Ivancevic en donde, mediante el algoritmo K-medias, evalúa las implicaciones que tiene la selección de un asiento por un estudiante en el salón de clases con las evaluaciones [20].

Este mismo algoritmo fue utilizado por Chang et al. [21] para saber qué tipo de aprendizaje fue el más apropiado para los distintos tipos de estudiantes, lo que permitió que los profesores adaptaran sus materiales didácticos y sus programas de enseñanza

Tabla 1. Clasificación de niveles de aprovechamiento según PLANEA.

Nivel I	Nivel II	Nivel III	Nivel IV
Logro Insuficiente	Logro apenas indispensable	Logro Satisfactorio	Logro Sobresaliente

de acuerdo con las habilidades y necesidades específicas de cada estudiante y, a su vez, una mayor eficiencia académica.

Los algoritmos de aprendizaje automático también han sido utilizados para evaluar el aprendizaje cuando se hace uso de las plataformas electrónicas mediante lo que se conoce como e-learning [22, 23]. Al respecto, Franki y Moudgalya [24] mencionan que, la forma o formas en que los alumnos hacen uso del software es determinante para saber cómo es el aprovechamiento académico al realizar aprendizaje con recursos e-learning.

En cuanto a la evaluación del aprovechamiento académico y variables o factores involucrados, en México se han llevado a cabo diversas investigaciones de análisis de datos sobre los resultados que se obtienen de la aplicación de las diferentes pruebas a diferentes niveles educativos. Raymundo et al. [25] en 2010 por ejemplo, utilizaron un modelo que compara el grado obtenido por la prueba ENLACE con las calificaciones obtenidas por cada bimestre del ciclo escolar, este análisis lo aplicaron a los alumnos de la Ciudad de México durante los años 2006 al 2009, en la educación primaria. Otros autores que se ha involucrado en el área del análisis del aprovechamiento académico mediante las pruebas estandarizadas son Heredia et al. [26] quienes desarrollaron el sistema informático ANCONE y cual determinó las variables de mayor impacto en pruebas académicas derivado del análisis de los resultados de la prueba PLANEA 2015 y reportando que existen algunos atributos tales como las aspiraciones académicas de los alumnos, el nivel de estudios de los padres o los recursos familiares que toman una importante relevancia en el impacto académico.

Según lo expuesto anteriormente, de importancia radical es el estudio de los distintos factores y variables que son actores vitales en el proceso de enseñanza-aprendizaje y que, eventualmente, pueden conducir al éxito o fracaso del alumnado.

3. Materiales y métodos

En este estudio, se procesan los resultados obtenidos de la prueba PLANEA 2017 para estudiantes del último grado del nivel medio superior de la base de datos del INEE y PLANEA [27, 28] y se analiza el impacto que cada una de las variables como ubicación de una institución, tipo de sostenimiento, nivel de marginación, turno y tipo de sistema educativo tiene sobre el rendimiento académico.

Los datos abiertos contienen un total de 16380 registros que corresponden al número total de escuelas evaluadas y de las cuales, para este análisis, se tomó la información de las columnas “entidad”, “nombre de la escuela”, “turno”, “subsistema educativo”, “grado de marginación”, “sostenimiento”, y “porcentaje de alumnos de la escuela en cada nivel de logro” como atributos y donde se tomaron en cuenta los resultados para el área de Lenguaje y Comunicación para cada uno de los 32 estados que conforman la República Mexicana. En su mayoría los datos utilizados son datos de tipo numérico, para el caso de los atributos no numéricos se agregó una clave numérica, así mismo para depurar los datos innecesarios se realizó la limpieza y depuración.

Particularmente, en este trabajo se estudia el impacto que sobre el aprovechamiento académico tienen los atributos "Subsistema Educativo", "Turno Escolar", "Nivel de marginación" y "Financiamiento de la escuela (sostenimiento)".

En la prueba PLANEA se evalúan 21 distintos sistemas educativos de nivel medio superior que se muestran en la Figura 2. Además, se tienen cuatro valores distintos para los turnos como son: discontinuo matutino, vespertino y nocturno.

Por otra parte, PLANEA clasifica el fenómeno de la marginación en cuanto a las carencias que padece la población en 5 niveles distintos; Muy Alto, Alto, Medio, Bajo y Muy Bajo. Estas carencias están en función de parámetros tales como falta de acceso a la educación, carencia de oportunidades sociales, residencia en viviendas inadecuadas, percepción de ingresos, ausencia de capacidades entre otros factores.

En cuanto al financiamiento o sostenimiento se refiere, PLANEA toma en cuenta cuatro distintas formas de procedencia de los recursos como son: Autónomo, Federal, Estatal y Particular.

Los niveles de logro en la prueba PLANEA son un parámetro de suma importancia en la interpretación de nuestros resultados y estos niveles plantean una forma de categorizar a las escuelas y a los estudiantes con base en los conocimientos y habilidades que demuestren. En el Nivel I corresponde al nivel de menor rendimiento académico y el nivel cuatro corresponde al de mayor aprovechamiento como se muestra en la Tabla 1.

3.1. Desarrollo

El procesamiento de los datos se llevó a cabo mediante la implementación del algoritmo K-medias, es una técnica muy conocida como método de agrupación de datos [29], el nombre viene dado porque representa cada uno de los grupos por la media (o media ponderada) de sus puntos, es decir, por sus "centroides". La representación mediante "centroides" brinda un sentido estadístico inmediato, donde se puede conocer una aproximación de las características por grupos. El algoritmo o criterio resulta más eficiente aplicándolo a atributos numéricos. Cuando esto ocurre, una medida de distancia muy común es la distancia Euclídea [7], para este análisis los datos se procesaron con un software desarrollado en lenguaje de programación JAVA permitiendo el seccionamiento de los datos en diferentes grupos o particiones. La forma de calcular los centroides, fue mediante el promedio de los valores numéricos (porcentajes de alumnos por nivel de logro según PLANEA) de los cuatro niveles de logro (Tabla 2) entre los diferentes objetos o registros de cada grupo; los K grupos que se consideraron para comparar las dispersiones de los datos respecto de los centroides fueron construidos mediante todas las posibles combinaciones entre el total de registros y se tomaron los grupos cuyas combinaciones dieran como resultado un valor menor del error cuadrático.

Posteriormente, se llevó a cabo un análisis de las características de cada grupo con el objetivo de buscar algún comportamiento o patrón que nos pudiera ayudar con la interpretación de los resultados. A pesar de que el cálculo se realizó con K=2, K=3 y K=4, en este trabajo únicamente se reportan los resultados para K=3 grupos; para K=2 y K=4 los resultados y comportamientos de los grupos fueron análogos que en el caso K=3. Para llevar a cabo la clasificación, con cada uno de los tres grupos finales obtenidos se calculó el promedio del total de los valores numéricos (porcentajes de

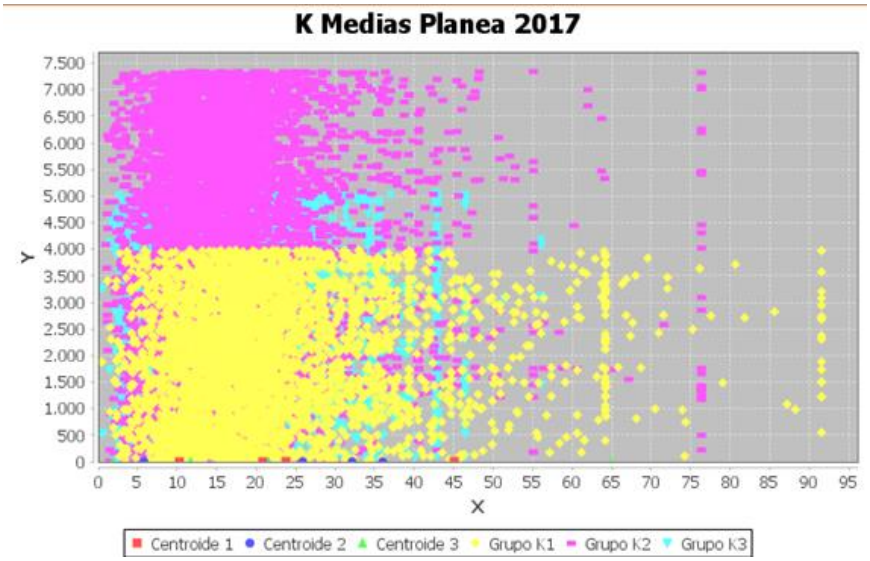

Fig. 1. Visualización de datos de Lenguaje y Comunicación con algoritmo K-medias, con representación de tres centroides y los grupos K1, K2 y K3.

Tabla 2. Distribución por número y porcentaje de escuelas en los tres grupos K1, K2 y K3 para el Área de Lenguaje y Comunicación.

Grupo	Número de escuelas	Promedio de porcentaje por nivel de logro			
		Nivel I	Nivel II	Nivel III	Nivel IV
K1	3988	10.295	20.840	45.084	23.783
K2	7344	32.141	36.012	25.888	5.799
K3	5048	64.900	21.551	11.703	1.845

alumnos) en cada nivel de logro y posteriormente se hizo la clasificación de acuerdo con las cantidades mayoritarias de porcentaje relacionando estas con los niveles de logro de PLANEA. Esto es, de las cuatro variables numéricas que corresponden a los porcentajes de alumnos en cada uno de los cuatro niveles de logro. En tanto que para las variables nominales tales como entidad, turno, subsistema, sostenimiento y marginación fue recuperada su correspondencia una vez que se aplicó el algoritmo a las variables numéricas.

4. Resultados

La visualización de los datos para el Área de Lenguaje y Comunicación en los tres grupos analizados K1, K2 y K3 se puede apreciar en la Figura 1. Para el grupo K=3 se realizaron 22 iteraciones para separar la base de datos PLANEA en donde los individuos comparten características similares.

Para el grupo K1 (3988 escuelas), grupo K2 (7344 escuelas) y grupo K3 (5048 escuelas).

En la Tabla 3 se muestra la cantidad de escuelas o instituciones en las que quedaron las particiones, así como los promedios de cada uno de los niveles de logro (promedios de las columnas) para cada grupo.

Con base en las cantidades mayoritarias de los promedios de escuelas mostrados en la Tabla 2, se hace una correspondencia de cada grupo con un nivel de logro. De esta forma, para el grupo K1 se puede ver que el porcentaje mayor (45.08%) corresponde al nivel III de logro, por lo que se puede decir que el grupo K1 concentra una cantidad importante de escuelas en nivel III y que en conjunto con el nivel IV (23.7%) podemos clasificar este grupo como de logro satisfactorio según los rangos propuestos por PLANEA.

Así mismo, para el grupo K2, la cantidad mayoritaria del promedio se encuentra en el nivel II (36.01%) por lo que se clasifica este grupo como de nivel de logro apenas indispensable. Finalmente, es claro que el grupo K3 tiene que clasificarse como un grupo de nivel de logro insuficiente al encontrarse la mayoría del porcentaje de escuelas (64.9%) dentro del nivel I de logro académico (logro insuficiente según PLANEA). Los grupos ya clasificados, según el criterio tomado para el análisis de datos son: K1 para satisfactorio, K2 logro apenas indispensable y K3 logro insuficiente.

En congruencia con este criterio de clasificación propuesto en este trabajo, adicionalmente en la Tabla 2 se puede observar el siguiente comportamiento: en el nivel IV de PLANEA de logro sobresaliente, se tiene los menores porcentajes de escuelas para cada uno de los grupos K, es decir, son pocas las escuelas con nivel académico destacado. Otro resultado interesante y que resalta esta misma congruencia es que el grupo K3 que se clasifico como de logro insuficiente, tiene un porcentaje mínimo de escuelas en el nivel IV (1.845) de PLANEA mientras, que en el grupo K1 de logro satisfactorio se tiene un mayor porcentaje de escuelas dentro del nivel IV (23.783), ósea, entre mayor es el nivel de desempeño académico (grupo K1), mayor es el porcentaje de escuelas dentro del nivel sobresaliente PLANEA (IV) en comparación con el porcentaje de escuelas para el mismo nivel PLANEA en el grupo K3 de menor nivel de desempeño académico.

En la Tabla 3 se muestra la cantidad y porcentajes de escuelas ordenados alfabéticamente según la entidad federativa y clasificadas en los tres grupos K1, K2 y K3. Se puede ver que el estado que concentra una mayor cantidad de escuelas en el grupo K1 (de logro satisfactorio) es Jalisco con un 40.09% de sus escuelas mientras que Chiapas es el estado que concentra un mayor porcentaje de sus escuelas evaluadas en el nivel K3 de logro insuficiente con un 63.17%.

Así mismo se puede ver que Jalisco, Ciudad de México y Querétaro son los estados que tienen un mayor porcentaje de escuelas dentro del grupo K1 de logro satisfactorio; a pesar de que se encuentran en los primeros lugares, aún tienen menos del 50% de sus escuelas con el nivel de logro satisfactorio. Por otra parte, se pueden identificar los estados de Chiapas, Guerrero y Tabasco como los estados con mayor rezago educativo al ubicarse como los primeros lugares en porcentaje de escuelas dentro del grupo K3 de logro insuficiente.

Así mismo, se realiza un análisis del comportamiento de la distribución de escuelas en cada uno de los tres grupos obtenidos en el caso del atributo “turno”, información que puede ser vista en la Tabla 4.

Tabla 3. Clasificación de los grupos obtenidos de acuerdo con los niveles PLANEA por entidad federativa para el Área de Lenguaje y Comunicación.

Entidad Federativa	Porcentaje de escuelas por grupo		
	K1	K2	K3
Aguascalientes	28.037	48.131	23.832
B. California	26.722	51.791	21.488
B. California S	33.028	43.119	23.853
Campeche	25.373	47.761	26.866
Chiapas	9.404	27.417	63.179
Chihuahua	18.333	41.042	40.625
CDMX	39.430	42.755	17.815
Coahuila	25.051	35.934	39.014
Colima	35.345	37.931	26.724
Durango	23.077	44.056	32.867
Guanajuato	21.842	52.463	25.696
Guerrero	11.873	27.425	60.702
Hidalgo	25.714	55.102	19.184
Jalisco	40.098	48.533	11.369
México	22.253	53.139	24.608
Michoacán	16.254	36.572	47.173
Morelos	28.829	46.847	24.324
Nayarit	13.580	43.210	43.210
Nuevo León	22.372	35.310	42.318
Oaxaca	14.580	51.347	34.073
Puebla	35.272	44.374	20.354
Querétaro	38.610	50.965	10.425
Quintana Roo	30.244	46.341	23.415
San Luis Potosí	14.004	38.293	47.702
Sinaloa	25.822	40.845	33.333
Sonora	35.802	42.284	21.914
Tabasco	13.095	30.952	55.952
Tamaulipas	22.613	41.709	35.678
Tlaxcala	19.681	56.915	23.404
Veracruz	23.110	48.780	28.110
Yucatán	29.178	49.008	21.813
Zacatecas	23.985	52.768	23.247

Tabla 4. Clasificación de los grupos obtenidos de acuerdo con los niveles PLANEA por turno para el Área de Lenguaje y Comunicación.

TURNO	Porcentaje de escuelas por grupo		
	K1	K2	K3
Discontinuo	60	34	6
Matutino	27.997	44.499	27.504
Vespertino	14.727	46.166	39.107
Nocturno	8.955	34.328	56.716

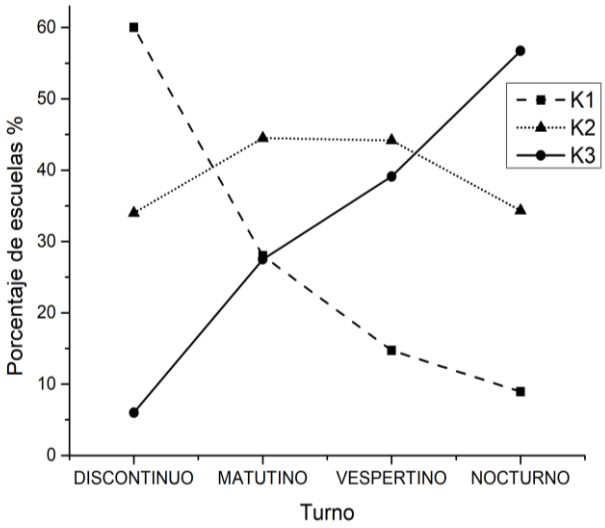

Fig. 2. Porcentaje de escuelas evaluadas por PLANEA por turno para el Área de Lenguaje y Comunicación: para el grupo K1 - logro satisfactorio, k2 - logro apenas indispensable, K3 – logro insuficiente.

Con base en los resultados reportados en la Tabla 4 y considerando los grupos K1 y K3 como representativos (como se hizo en el caso del análisis por entidad), en la Figura 2 se puede ver que el turno matutino es preponderante en el grupo K1 de logro satisfactorio sobre el turno vespertino, mientras que el porcentaje de escuelas evaluadas del turno vespertino es mayoritario en el grupo K3 de logro insuficiente.

Esto indica que existe una importante correlación entre el turno y el aprovechamiento académico siendo el turno matutino asociado a una mayor eficiencia. A pesar de que los turnos discontinuo y nocturno no tienen una cantidad significante de escuelas, por lo que no se pueden comparar con los turnos matutino y vespertino, se puede ver una tendencia que indica que el turno discontinuo tiene una preponderancia general en el grupo K1 y el turno nocturno en el grupo K3 para el caso de la evaluación en Lenguaje y Comunicación.

Análogamente a lo que se hizo en el caso anterior, en la Figura 3 se muestra el comportamiento de los porcentajes de escuelas distribuidas en los grupos K1, K2 y K3 para el análisis por subsistema.

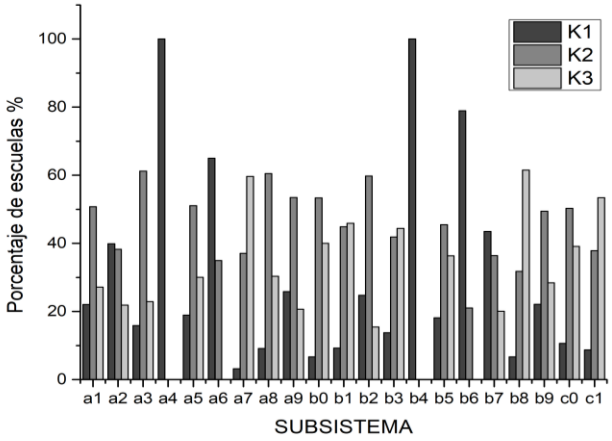

Fig. 3. Porcentaje de escuelas evaluadas por PLANEA por subsistema para el Área de Lenguaje y Comunicación: el grupo K1 - logro satisfactorio, K2 - logro apenas indispensable, K3 - logro insuficiente (a1-BACH ESTATAL DGE-CGE, a2-BACHI AUTONOMO, a3-CECYTE, a4-CETI, a5-COBACH, a6-COLBACH MEX, a7-CONALEP DF-OAX, a8-CONALEP EDOS, a9-DGB, b0-DGECYTM, b1-DGETA, b2-DGETI, b3-EMSAD, b4-IPN, b5-OTRAS ESTATALES, b6-OTRAS FEDERALES, b7-PARTICULARES, b8-PREECOS, b9-PREFECOS, c0-TELEBACHILLERATOS, c1-TELEBACHILLERATOS COMUNITARIOS).

Cabe destacar que, a pesar de que en el subsistema del Instituto Politécnico Nacional solo se evaluaron 35 escuelas en comparación con otras escuelas donde la cantidad fue mucho mayor, la totalidad de escuelas del IPN se encuentran dentro del grupo K1 mientras que sistemas como subsistemas tales como el CONALEP DF-OAX o los sistemas PREECOS (Preparatorias Estatales por Cooperación), tienen una cantidad mayoritaria de escuelas dentro del grupo K3 de nivel insuficiente.

Con respecto al nivel de marginación, en la Figura 4 se puede ver que existe una importante correlación entre el factor pobreza y el nivel de aprovechamiento académico al observar que, si el nivel de marginación es muy bajo, entonces se tiene una cantidad mayoritaria de escuelas en el grupo K1 (de logro satisfactorio) mientras que, si el nivel de marginación es muy alto, las escuelas se distribuyen mayoritariamente en el grupo K3 de nivel insuficiente.

4. Conclusiones y trabajo a futuro

En la presente investigación se ha aplicado al algoritmo K-medias para el análisis de los datos obtenidos a partir de los resultados de la prueba PLANEA llevada a cabo en el año 2017 en estudiantes del último grado del nivel medio superior en el Área de Lenguaje y Comunicación. Mediante el algoritmo se obtuvieron 3 grupos (K=3) a los que, aplicando nuestro criterio de acuerdo, a sus características, podemos clasificar como de logro satisfactorio (K1), logro apenas indispensable (K2) y logro insuficiente (K3).

Con base en nuestros resultados y tomando a los grupos K1 y K3 como representativos, se puede ver que nuestro criterio de clasificación es adecuado ya que se pudo observar una mayor cantidad de escuelas del grupo K1 en el nivel IV de

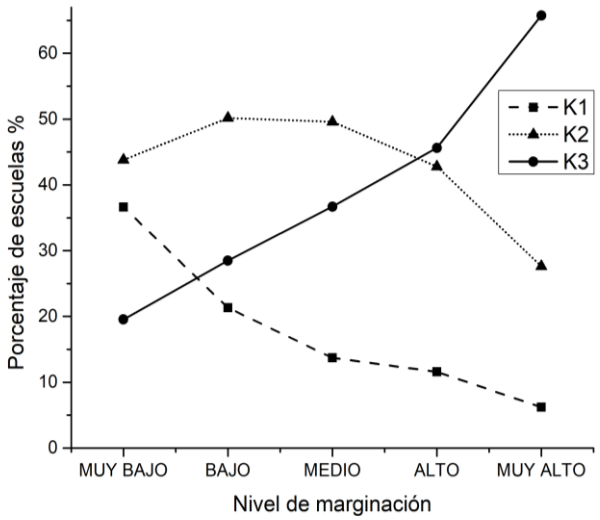

Fig. 4. Porcentaje de escuelas evaluadas por PLANEA por nivel de marginación para el Área de Lenguaje y Comunicación: para el grupo K1 - logro satisfactorio, k2 - logro apenas indispensable, K3 - logro insuficiente.

PLANEA (sobresaliente) en comparación con el número de escuelas del grupo K3 en ese mismo nivel; en contraste a este resultado, se obtiene que el grupo K3 concentra la mayoría de sus escuelas en el nivel I de logro insuficiente.

En cuando al atributo Entidad Federativa, reportamos que los estados de Jalisco, Ciudad de México, Querétaro, Sonora y Colima son aquellos que ostentan un mayor aprovechamiento académico al tener entre el 35 y el 40 por ciento de sus escuelas en el grupo K1, mientras que los estados de Chiapas, Guerrero, Tabasco, San Luis Potosí y Michoacán son los estados que tienen una mayor concentración de sus escuelas en el grupo K3 (entre 47 y 63 por ciento). Estos resultados son congruentes con aquellos que se reportan en la Encuesta Nacional de los Hogares 2016 realizada por el Instituto Nacional de Estadística y Geografía (INEGI) [30] en donde, en particular, los Estados de Chiapas, Michoacán y Guerrero se encuentran entre los estados de mayor rezago educativo, mientras que, en esta misma encuesta, la Ciudad de México y Sonora se encuentran entre los estados de menor rezago tal y como es reportado en este trabajo.

En cuanto al análisis de la relación que tiene el aprovechamiento académico con el turno, en este trabajo se obtiene que el turno matutino domina en el grupo K1 mientras que en el turno vespertino es dominante el grupo K3, esto significa que el turno matutino se caracteriza por tener un mayor aprovechamiento. Por otra parte, respecto al atributo Subsistema, con base en nuestros resultados es de destacar la eficiencia de algunas instituciones educativas como el Instituto Politécnico Nacional o los Centros de Enseñanza Técnica Industrial (CETI) cuya totalidad de escuelas se encontraron en el grupo K1. Respecto del nivel de marginación, otra de las variables estudiadas en este trabajo, se pudo observar que existe una alta correlación entre el nivel marginación y el aprovechamiento académico al ver que un mayor porcentaje de escuelas que pertenecen al grupo K1, de logro satisfactorio, se encuentran en zonas que tienen niveles de

marginación bajo o muy bajo, mientras que aquellas escuelas que se ubican en zonas de alta marginación se clasificaron dentro el grupo K3 de logro insuficiente. Una extensión del presente trabajo sería realizar el mismo estudio, pero ahora tomando en cuenta los resultados que se obtuvieron en el Área de Matemáticas (en el mismo año 2017) y ver si hay un patrón de comportamiento entre ambas áreas y así mismo, realizar el estudio tomando en cuenta los resultados de la prueba de años posteriores para ver la evolución de dicho comportamiento.

Referencias

1. Rivas, T., González, M.J., Delgado, M.: Descripción y propiedades psicométricas del test de evaluación del rendimiento académico (TERA). Interamerican Journal of Psychology, 44(2), pp. 279–290 (2010)
2. Martínez, F.: La evaluación formativa del aprendizaje en el aula en la bibliografía en inglés y francés. Revista Mexicana de Investigación Educativa, 17(54), pp. 849–875 (2012)
3. Organización para la Cooperación y Desarrollo Económico: El trabajo de la OCDE sobre educación y competencias, https://www.oecd.org/education/El-trabajo-de-la-ocde-sobre-educacion-y-competencias.pdf (2019)
4. Martínez, F.: Las evaluaciones educativas en América Latina. México: Instituto Nacional para la Evaluación de la Educación (2008)
5. Garre, M., Cuadrado, J.J., Sicilia, M.A., Rodríguez, D., Rejas, R.: Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software. REICIS, Revista Española de Innovación, Calidad e Ingeniería del Software, 3(1), pp. 6–22. https://www.redalyc.org/articulo.oa?id=922/92230103 (2007)
6. Martínez-Abad, F., Hernández-Ramos, J.P.: Técnicas de minería de datos con software libre para la detección de factores asociados al rendimiento. REXE, Revista de Estudios y Experiencias en Educación, 2(2), pp. 135–145. https://www.redalyc.org/articulo.oa?id=2431/243156768012 (2018)
7. Pérez, S., Jaime, J., Espín, R.: Departamento ingeniería industrial, Instituto Superior Politécnico José Antonio Echeverría. Revista investigación operacional, 33(3), pp. 210–221 (2012)
8. Chaparro, A., González, C., Caso, J.: Familia y rendimiento académico: configuración de perfiles estudiantiles en secundaria. Revista electrónica de investigación educativa, 18(1), pp. 53–68. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1607-40412016000100004&lng=es&tlng=es (2016)
9. Abreu, R., Martín del Campo, D.C.: Índice compuesto de eficacia de los sistemas escolares. Mexicanos Primero Visión 2030 A.C, Fundación IDEA A.C. (2007)
10. De Hoyos, R., Espino, J.M., García, V.: Determinantes del logro escolar en México. Primeros resultados utilizando la prueba ENLACE media superior. El Trimestre Económico, LXXIX 4(316), Fondo de Cultura Económica, pp. 783–811 (2012)
11. Todd, P.E., Wolpin, K.I.: On the specification and estimation of the production function for cognitive achievement. Economic Journal, Royal Economic Society, 113(485), pp. 3–33, https://ideas.repec.org/a/ecj/econjl/v113y2003i485pf3-f33.html (2003)
12. Glewwe, P., Kremer, M.: Chapter 16 Schools, Teachers, and Education Outcomes in Developing Countries. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education, pp. 945–1017 (2006)
13. Thomas, F., Ludger, W.: What accounts for international differences in student performance? a re-examination using PISA data. Empirical Economics, 32(2-3), pp. 433–464 (2007)

14. Hanushek, E.: The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature, 24(3), pp. 1141–1177 (1986)
15. Romero, C., Ventura, S.: Educational data mining: A survey from 1995 to 2005. Expert Syst. Appl., 33, pp. 135–146 (2007)
16. Romero, C., Ventura, S.: educational data mining: A review of the state of the Art. IEEE Transactions on Systems Man and Cybernetics Part C, (Applications and Reviews) 40(6), pp. 601–6182 (2007)
17. Beck, J. E., Woolf, B.: High-level student modeling with machine learning. In: Proc. 5th Int. Conf. Intell. Tutoring Syst, pp. 584–593 (2000)
18. Baker, R., Corbett, A., Wagner, A.: Off-task behavior in the cognitive tutor classroom: When students game the system. In: Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 383–390 (2004)
19. Wang, Y., Liao, H.: Data mining for adaptive learning in a TESL-based e-learning system. Expert Syst. Appl., 38(6), pp. 6480–6485 (2011)
20. Ivancevic, V., Celikovic, M., Lukovic, I.: The individual stability of student spatial deployment and its implications. Int. Symp. Comput. Edu. (SIIE), pp. 1–4 (2012)
21. Chang, W., Chen, S., Li, M., Chiu, J.: Integrating IRT to clustering student's ability with K-means. In: Proc. 4th Int. Conf. Innov. Comput., Inf. Control (ICICIC), pp. 1045–1048 (2009)
22. Aher, S., Lobo, L.: Applicability of data mining algorithms for recommendation system in e-learning. In: Conf. Adv. Comput., Commun. Inform. (ICACCI), pp. 1034–1040 (2012)
23. Zheng, Q., Ding, J., Du, J., Tian, F.: Assessing method for E-learner clustering. In: 11th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD), pp. 979–983 (2007)
24. Eranki, K., Moudgalya, K.: Evaluation of Web based behavioral interventions using spoken tutorials. In: IEEE 4th Int. Conf. Technol. Edu., pp. 38–45 (2012)
25. Campos, R., Romero, F.: Desempeño educativo en México: la prueba ENLACE. In: Serie documentos de trabajo del Centro de Estudios Económicos, El Colegio de México, Centro de Estudios Económicos, https://ideas.repec.org/p/emx/ceedoc/2010-19.html (2010)
26. Heredia, A., Chi, A., Guzmán, A., Martínez, G.: ANCONE: An interactive system for mining and visualization of students information in the context of PLANEA 2015. Computación y Sistemas, 24(1), pp. 151–176 (2020)
27. Instituto Nacional para la Evaluación de la Educación: Evaluaciones de Logro referidas al Sistema Educativo Nacional. Último grado de Educación Media Superior (2016-2017), INEE. https://www.inee.edu.mx/evaluaciones/planea/media-superior-ciclo-2016-2017 (2020)
28. Plan Nacional para la Evaluación de los Aprendizajes: Bases de datos de escuelas, PLANEA. http://planea.sep.gob.mx/ms/base_de_datos_2017/ (2019)
29. Tar, J., Bitó, J.F., Rudas, I., Várkonyi, T.A.: Decentralized Adaptive Control with Fractional Order Elimination of Obsolete Information. In: 4th International Conference on Emerging Trends in Engineering and Technology, ICETET 2011 (2011)
30. Instituto Nacional de Estadística y Geografía: Encuesta Nacional de los Hogares. https://www.inegi.org.mx/programas/enh/2016/ (2016)