

DOI: <https://doi.org/10.23913/cemys.v11i21.349>

Artículos científicos

Logro académico en Lenguaje y Comunicación de estudiantes de Educación Media Superior: Caso Veracruz basado en un enfoque de inteligencia computacional

Academic achievement in Language and Communication of Upper Middle Education students: Veracruz Case based on a Computational Intelligence approach

Yaimara Céspedes González

Universidad Veracruzana

Xalapa, Veracruz, México

yclespedes@uv.mx

<https://orcid.org/0000-0001-7802-3219>

Alma Delia Otero Escobar

Universidad Veracruzana

Xalapa, Veracruz, México

aotero@uv.mx

<https://orcid.org/0000-0001-9266-6587>

Resumen

La educación juega un rol fundamental en la sociedad, por lo que es importante contar con un sistema educativo robusto. Para medir la calidad de un sistema educativo, el rendimiento académico se considera un indicador de interés. Los resultados de las pruebas estandarizadas, como es el caso de PLANEA en México, ofrecen la posibilidad de medir el rendimiento académico. Por lo que, esta investigación presenta el análisis del rendimiento académico en Lenguaje y Comunicación de estudiantes de Educación Media Superior de Veracruz, a través de un enfoque basado en la Inteligencia Computacional. Para llevar a cabo este trabajo de investigación, se utilizó un enfoque de aprendizaje no supervisado (K-means) que permite modelar un conjunto de variables de entrada y asignar elementos (vectores de datos) con características similares a determinados clústeres. El método de solución se conformó de cuatro pasos: adquisición de datos, análisis exploratorio de datos, selección de variables y aplicación de algoritmos. Como resultado, se observó una segmentación diferenciada para el período de evaluación. A partir de esta segmentación se identificaron algunas características que se destacan en cada uno de los seis

grupos obtenidos, como: el turno, grado de marginación, soporte, estudiantes programados y estudiantes evaluados.

Palabras claves: Educación Media Superior; Inteligencia Computacional; Lenguaje y Comunicación; Rendimiento Académico.

Abstract

Education plays a fundamental role in society; hence it is important to have a strong educational system. In this sense, the academic achievement of students is considered an indicator of interest to measure the quality of education offered by educational systems. The results of standardized test, such as PLANEA in Mexico, offer the possibility of measuring academic achievement. Therefore, this research presents the analysis of the academic achievement in language and communication of high school students in the state of Veracruz, through an approach based on computational intelligence. To carry out this research work, an unsupervised learning approach (K-means) was used, an algorithm that models a set of input variables and assigns elements (data vectors) with similar characteristics to certain clusters. The solution method consisted of four steps: i) data acquisition, ii) exploratory data analysis, iii) variable selection, and iv) algorithm application. As a result, a differentiated segmentation was observed for the evaluation period. From this segmentation, some characteristics that stand out in each of the six groups obtained were identified, such as: shift, degree of marginalization, support, scheduled students and evaluated students.

Keywords: Academic Achievement; Computational Intelligence; Language and Communication; Upper-Middle Education.

Fecha Recepción: Junio 2023

Fecha Aceptación: Diciembre 2023

Introducción

La educación se considera uno de los pilares fundamentales para el desarrollo social y económico de un país, ya que los estudiantes que actualmente cursan la educación obligatoria (primaria, secundaria y bachillerato) serán la fuerza laboral y económica del mañana (Molero-Castillo *et al.*, 2018). Con base en lo anterior, se vuelve importante disponer de un sistema educativo de calidad. En general, los sistemas educativos son estructuras económicas, políticas y sociales que ofrecen a la población la oportunidad de obtener una educación dentro de un determinado país (Wallace, 2009; Hatos, 2014).

Para mejorar la calidad educativa y los resultados en el aprendizaje, es necesario analizar la situación actual con base en evidencias y tener información sobre las fortalezas y debilidades de

un sistema educativo. Por lo tanto, un sistema de seguimiento y evaluación de indicadores relevantes puede proporcionar esta evidencia (Sauvageot, 1997), tal es el caso del rendimiento académico; el cual es un indicador que permite medir la calidad de la educación que ofrecen los sistemas educativos. Por su parte, el rendimiento académico es una medida del logro de los estudiantes en diversas materias académicas (Briones *et al.*, 2022). Por esta razón, existe interés en saber en qué medida los estudiantes adquieren competencias clave en dominios esenciales, como es el caso de Lenguaje y Comunicación.

Una forma de medir el rendimiento académico es a través de los resultados de pruebas estandarizadas (Ward *et al.*, 1996). Tal es así que, algunos países aplican evaluaciones a nivel nacional a estudiantes de diferentes niveles educativos para medir su rendimiento académico. En México, se aplican diferentes evaluaciones, las cuales aportan datos, cuyo análisis permite adoptar mejores políticas educativas (INEE, 2016). Uno de ellos es el examen PISA, que mide la adquisición de competencias clave para la vida adulta por parte de jóvenes de 15 años en matemáticas, lectura y ciencias en los países que conforman la OCDE -Organización para la Cooperación y el Desarrollo Económicos- (OECD, 2024). Además, se aplica la evaluación estandarizada ERCE (Estudio Regional Comparativo y Explicativo), que mide el rendimiento de los estudiantes en las áreas de lectura, escritura, matemáticas y ciencias (UNESCO, 2019).

En México a través del Instituto Nacional para la Evaluación de la Educación (INEE) se realiza la prueba estandarizada PLANEA (abreviatura de Plan Nacional para la Evaluación del Aprendizaje), la cual evalúa dos áreas de competencia, Matemáticas y Lenguaje y Comunicación (SEP, 2024), y se aplica en diferentes niveles educativos (primaria, secundaria y bachillerato). El análisis de esta información apoya la mejora educativa en el país, ya que este examen proporciona datos relevantes sobre el rendimiento académico de los estudiantes.

En el caso del nivel educativo medio superior, la aplicación de este examen permite conocer en qué medida los estudiantes lograron dominar un conjunto de aprendizajes esenciales en materias básicas como Lenguaje y Comunicación y Matemáticas. Este análisis es particularmente interesante, ya que, según Ibarrola (2012), algunos estudiantes al finalizar este nivel abandonan o retrasan sus estudios, pues no alcanzan los conocimientos necesarios para ingresar a las universidades del país.

En particular, el nivel educativo medio superior proporciona los conocimientos necesarios para continuar con los estudios a nivel profesional, ya que en él se construyen las bases educativas fundamentales para tener éxito en la universidad. Ante este escenario, hacer un diagnóstico de los conocimientos que adquieren los estudiantes en este nivel es importante, ya que, a partir de él, se

podrían plantear estrategias para mejorar el nivel académico y garantizar condiciones homogéneas y favorables para que los estudiantes continúen con una carrera universitaria (SEP, 2024).

Dado que los datos correspondientes a los resultados de PLANEA, para el Estado de Veracruz, se encuentran disponibles para el nivel de educación media superior, en su versión más actualizada, 2017. Surge el interés en analizar el rendimiento académico en Lenguaje y Comunicación, de estudiantes de educación media superior, particularmente en el estado de Veracruz, a través de un enfoque basado en la Inteligencia Computacional, el cual se compone de métodos y enfoques computacionales inspirados en abordar problemas complejos. En concreto, se utilizó un enfoque de aprendizaje no supervisado basado en K-means, el cual es un algoritmo de cuantificación vectorial que agrupa elementos similares en forma de grupos (clústeres).

El artículo está organizado en secciones. La Sección 2 presentan los antecedentes sobre la educación media superior en México, el rendimiento académico, la segmentación basada en aprendizaje automático y se presentan trabajos relacionados. La Sección 3 describe el método definido como propuesta de solución. La Sección 4 presenta los resultados obtenidos, basados en el análisis de los datos de PLANEA, y la Sección 5 resume las principales conclusiones y trabajos futuros.

Antecedentes

En México, la educación media superior ha adquirido relevancia. Desde 2012, se estableció su carácter obligatorio, al mismo tiempo que se han realizado esfuerzos para ampliar su cobertura y mejorar la calidad de la enseñanza. En este contexto, PLANEA logra mayor notoriedad, ya que brinda información a la sociedad sobre el estado de la educación en México, en términos de logros de aprendizaje de los estudiantes. Por lo tanto, este plan bien utilizado puede convertirse en una herramienta de diagnóstico para mejorar la calidad de la educación en el país.

Educación media superior

La educación media superior en México tiene una duración de tres años y se ubica entre el nivel básico y universitario. Por lo tanto, en este nivel se complementa la educación básica, y al mismo tiempo se prepara a los estudiantes para el ingreso a la universidad, instituciones de educación superior o para realizar alguna actividad productiva (Villa, 2007; Lorenzo y Zaragoza, 2014). Así, por su relevancia e impacto en el desarrollo del país, desde 2012 se busca implementar su obligatoriedad, extendiéndola a todas las clases sociales, aumentando la eficiencia terminal y mejorando el aprendizaje de los estudiantes (Villa, 2013; INEE, 2024a).

En este sentido, es importante comprender el papel que juegan las evaluaciones estandarizadas a gran escala, como PLANEA, para la educación media superior. El cual está orientado a la evaluación de estudiantes de todo el país que se encuentran en el último grado de educación media superior, ya sea en colegios federales, estatales, privados o autónomos. Este examen evalúa el desempeño del estudiante en dos dominios formativos: Lenguaje y Comunicación y Matemáticas. La prueba consta de 50 ítems para cada dominio.

En general, la evaluación en Lenguaje y Comunicación se centra en la comprensión lectora, es decir, la extracción de información, interpretación y reflexión acerca de la naturaleza del lenguaje y su uso como herramienta del pensamiento lógico (SEP, 2016; SEP, 2024). Además, a partir de 2018 incorporó la expresión escrita (INEE, 2018).

La primera evaluación de PLANEA en educación media superior se realizó en 2015. Los siguientes exámenes se administraron en 2016 y 2017. El examen no se aplicó durante la pandemia por COVID-19, y se retomó en 2022, cuando se realizó su cuarta y última aplicación, de la cual no se han publicado los resultados para el estado de Veracruz (PLANEABD, 2024). Con base en lo anterior, esta investigación se centra en el análisis de los datos relacionados con los resultados obtenidos en 2017 para la educación media superior, en el ámbito de Lenguaje y Comunicación.

El propósito de PLANEA es conocer en qué medida los estudiantes logran aprendizajes clave. Además, proporciona información contextualizada para mejorar los procesos de enseñanza, informando a la sociedad sobre los logros de aprendizaje de los estudiantes, brindando a las autoridades educativas y docentes, información relevante para monitorear, planificar y operar el sistema educativo. Para esto, PLANEA agrupa el rendimiento académico en cuatro niveles progresivos, como se muestra en la Tabla 1, donde cada nivel proporciona información sobre el grado en que los estudiantes dominan ciertos aprendizajes, que se consideran claves en la educación media superior (INEE, 2016; INEE, 2024b).

Tabla 1. Niveles del logro académico.

Niveles	Descripción
I	Los estudiantes tienen un conocimiento insuficiente de los aprendizajes esenciales. Esto sugiere posibles dificultades para continuar con estudios posteriores.
II	Los estudiantes tienen conocimientos elementales de los aprendizajes esenciales.
III	Los estudiantes tienen un conocimiento satisfactorio del aprendizaje esencial.
IV	Los estudiantes tienen un conocimiento sobresaliente de los aprendizajes esenciales. Lo que sugiere mayores posibilidades de éxito para continuar con estudios posteriores.

Fuente: Elaboración propia.

Segmentación basada en Inteligencia Computacional

Una característica notable de los métodos de inteligencia computacional es su capacidad similar a la humana para respaldar el proceso de toma de decisiones basado en información con imprecisión e incertidumbre. Es así como, el aprendizaje automático cobra relevancia, ya que consiste en un conjunto de algoritmos de análisis basado en datos, que permiten construir modelos, basados en observaciones (experiencias). El propósito es entrenar y aprender de estas observaciones (Méndez *et al.*, 2008; Mathur, 2019). Dentro del aprendizaje automático, los métodos no supervisados son algoritmos que basan su proceso de entrenamiento en un conjunto de datos sin etiquetas definidas. Es decir, no se conoce ningún valor objetivo o de clase, ni categórico, ni numérico. Por lo tanto, estos métodos no requieren intervención humana (Rouhiainen, 2018).

Las principales aplicaciones del aprendizaje no supervisado están relacionadas con la segmentación de datos, donde el objetivo es encontrar grupos con elementos similares. De tal forma que, la información interna de los elementos que forman un clúster sea similar y diferente a la de los elementos de otros clústeres (García y Gómez, 2006). Actualmente, existen dos tipos principales de algoritmos de segmentación (Pascual *et al.*, 2007):

- *Métodos jerárquicos*, producen una organización jerárquica de los elementos, en forma de árbol, en función del grado de similitud (medidas de distancia) de los elementos que componen el conjunto de datos, permitiendo diferentes niveles de agrupación.
- *Métodos de particionamiento*, también conocidas como particiones, que generan agrupaciones de elementos en forma de segmentos que delimitan los grupos, en base a mediciones de distancias con centros geométricos, conocidos como centroides.

Un algoritmo de agrupamiento jerárquico, del tipo aglomerado, inicia el agrupamiento a partir de cada elemento individual, considerándolo como un grupo unitario; y une iterativamente los dos grupos más cercanos hasta obtener un único grupo general. Los elementos similares, con menor distancia, son los primeros en unirse, y se van uniendo progresivamente según sus similitudes. Por el contrario, los métodos jerárquicos divisivos funcionan en la dirección opuesta. Un único grupo inicial de elementos se divide en dos subgrupos, de modo que los elementos de un subgrupo son diferentes de los elementos del otro. Estos subgrupos se dividen sucesivamente hasta tener tantos subgrupos como elementos tiene (Pascual *et al.*, 2007).

Los algoritmos de partición asumen un conocimiento a priori del número de clústeres en el que se debe dividir el conjunto de datos, es decir, llegan a una división que optimiza un criterio

predefinido (Soto y Vázquez, 2006). Entre los algoritmos que utilizan este tipo de agrupación destaca K-means, cuya idea principal es definir K centroides (uno para cada clúster) para luego tomar cada elemento (un vector de datos) de la base de datos y colocarlo en el clúster cuyo centroide es el más cercano.

El centroide es un punto que ocupa la posición media en un grupo. El siguiente paso es recalcular el centroide de cada grupo y redistribuir todos los elementos según el centroide más cercano. El proceso se repite hasta que no haya cambios en los grupos formados (Pascual *et al.*, 2007). Además, en K-medias se puede utilizar el método del codo, con diferentes configuraciones de K, para obtener una aproximación del número apropiado de segmentos.

Trabajos relacionados

En los últimos tiempos, el análisis del rendimiento académico de los estudiantes de educación obligatoria ha sido objeto de estudio debido a su importancia, no solo en el nivel de educación primaria y secundaria, sino también en la educación media superior, por su impacto en la educación de la población mexicana. En este sentido, el análisis del rendimiento académico en el nivel de educación media superior es fundamental, ya que proporciona información sobre el logro de aprendizajes clave de los estudiantes de educación obligatoria y las brechas existentes entre los diferentes grupos poblacionales. Este tipo de análisis se puede realizar basándose en la observación de similitudes, tendencias y comportamientos, para lo cual, los algoritmos de aprendizaje automático son útiles. Algunos trabajos relacionados con algoritmos de segmentación y partición de datos en educación se presentan en la Tabla 2.

Tabla 2. Trabajos relacionados

Descripción	Datos	Método
En Maldonado <i>et al.</i> (2016), se centraron en el análisis de los logros en Lenguaje y Comunicación y Matemáticas en la educación media superior. De los resultados, se destacaron los niveles I y II (insuficiente y elemental) en ambos dominios evaluados.	PLANEA 2015	K-means
En Molero-Castillo <i>et al.</i> (2019) también se analizaron los resultados de Lenguaje y Comunicación alcanzados en el nivel de educación media superior. Se desarrolló una aplicación web para este análisis del rendimiento académico.	PLANEA 2017	Clúster particional
En Gutiérrez <i>et al.</i> (2020) se analizan los resultados en Lenguaje y Comunicación para el nivel medio superior. Se identificaron variables que impactan el rendimiento académico.	PLANEA 2017	K-means
En Céspedes-González <i>et al.</i> (2024) se persenta el análisis del rendimiento académico en Matemática para estudiantes de educación media superior en Veracruz a través de un enfoque basado en inteligencia computacional.	PLANEA 2017	K-means

Fuente: Elaboración propia.

Motivación

De acuerdo con el INEE (2016), tanto Matemáticas como Lenguaje y Comunicación son dos campos de formación que proporcionan herramientas indispensables para la vida cotidiana y para el desarrollo de aprendizajes en otras áreas. Ambas disciplinas representan una proporción considerable de la carga curricular en educación media superior. Por lo que, los aprendizajes en ambos campos se consideran un indicador de los resultados educativos en general. Con base en lo anterior, surge el interés de analizar el rendimiento académico en Lenguaje y Comunicación.

Para el caso de Matemáticas, como investigación preliminar, en Céspedes-González et al. (2024), se analizó el rendimiento académico de estudiantes de educación media superior de Veracruz. Para esto, se utilizó un enfoque de aprendizaje no supervisado basado en K-means. Como resultado, se obtuvo una segmentación diferenciada, a partir, de la cual se identificaron características que destacan en cada grupo. Lo anterior permitió comprender el comportamiento del rendimiento académico de esta materia en la educación media superior en Veracruz.

Con base en lo anterior, este trabajo de investigación utiliza también un enfoque basado en la Inteligencia Computacional, para analizar el rendimiento académico en Lenguaje y Comunicación. Para esto, se utilizó la base de datos PLANEA-2017. Con este análisis se pretende encontrar similitudes entre los registros de las escuelas de educación media superior del estado de Veracruz que participaron en la evaluación estandarizada.

Método

El método de solución definido para el análisis del rendimiento académico en Lenguaje y Comunicación de estudiantes de secundaria de Veracruz se dividió en cuatro etapas: i) adquisición de datos, ii) análisis exploratorio de datos, iii) selección de variables y iv) aplicación de algoritmos.

Adquisición de datos

Como fuente de datos se utilizaron registros de la base de datos del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA), específicamente datos de escuelas de educación media superior estatales, federales y privadas; reconocido por la Secretaría de Educación Pública de México.

El acceso a la versión actual de la fuente de datos se realizó a través de la página institucional PLANEA, cuyo principal objetivo es conocer en qué medida los estudiantes logran dominar un conjunto de aprendizajes esenciales en diferentes momentos de su educación obligatoria (INEE, 2016; INEE, 2024b), en este caso al finalizar la educación media superior. Para esto, se utilizó PLANEA como instrumento de evaluación en un examen compuesto por 100 ítems de opción múltiple, divididos en dos competencias formativas: a) 50 para Lenguaje y Comunicación, y b) 50 para Matemáticas.

En Lenguaje y Comunicación se evalúa la capacidad de los estudiantes para comprender, analizar, interpretar, reflexionar, evaluar y utilizar textos escritos, mediante la identificación de su estructura, funciones y elementos. Para esto, se utilizan diferentes tipos de textos (apelativos, argumentativos, expositivos y narrativos) que le permiten al estudiante evaluado mostrar sus habilidades lectoras. Este examen se enfoca en las siguientes competencias básicas (INEE, 2016):

- Identificar, ordenar e interpretar ideas, datos y conceptos explícitos e implícitos en un texto.
- Evaluar un texto mediante la comparación de su contenido.
- Identificar el uso normativo de la lengua, considerando la intención y su situación.
- Analizar un argumento preciso y coherente.
- Relacionar ideas y conceptos en composiciones coherentes y creativas.
- Evaluar la secuencia y relación lógica en el proceso de comunicación.

- Interpretar la idea principal y posible desarrollo de un mensaje escrito.

Para 2017 se evaluaron 1640 instituciones, entre estatales, federales y privadas, con validez oficial y reconocidas por la Secretaría de Educación Pública. PLANEA en educación media superior evalúa las competencias educativas adquiridas en cuatro niveles de rendimiento académico: a) I -insuficiente-, b) II -elemental-, c) III -satisfactorio-, y d) IV -sobresaliente-. Estos niveles describen el desempeño que un estudiante en el último año escolar puede obtener como calificación en la evaluación PLANEA. Los niveles representan las tareas y procesos cognitivos que deben alcanzar los estudiantes al egresar del nivel medio superior en Lenguaje y Comunicación. Estos niveles de competencia no sólo sirven para identificar el rendimiento académico de los estudiantes sino también para tener una visión global del desempeño de los colegios evaluados.

Análisis exploratorio de datos

Sobre el conjunto de datos se realizó inicialmente un análisis exploratorio, el cual resultó útil para conocer los datos y comprender sus principales características. Para ejecutar este análisis se utilizó Python como herramienta. Así, con base en la exploración de datos, se observó que la estructura de datos original está conformada por 52 variables y 1640 registros (colegios), sin valores nulos, lo que demuestra que es un conjunto de datos robusto. También se observó que no existen valores atípicos o fuera de rango.

Selección de variables

De las 52 variables se seleccionaron 12 estrictamente relacionadas con el dominio Lenguaje y Comunicación, y un total de 1640 registros. Este conjunto de variables se obtuvo filtrando algunas variables categóricas, por ejemplo, clave de entidad, nombre de entidad, extensión, nombre del municipio (dejando solo el nombre de la localidad), subsistema educativo y colegios similares, entre otras. Así como la eliminación de todas las variables asociadas al dominio de Matemáticas, ya que no es objeto de estudio en este trabajo de investigación; y otras variables repetidas sobre los porcentajes de estudiantes de la escuela en cada nivel de rendimiento académico alcanzado.

En este sentido, las variables seleccionadas ofrecen información específica sobre los logros académicos alcanzados en Lenguaje y Comunicación por los estudiantes de educación media superior en Veracruz. La Tabla 3 muestra una descripción de las variables seleccionadas.

Tabla 3. Variables seleccionadas

N.	Variables	Descripción
1	Escuela	Nombre de la escuela que fue evaluada.
2	Horario	Turno en el que la escuela desarrolla sus actividades.
3	Localidad	Nombre de la localidad a la que pertenece la escuela evaluada.
4	Marginación	Grado de marginación de la localidad en la que se encuentra la escuela.
5	Soporte	Tipo de apoyo al que pertenece la escuela evaluada (estatal, federal o privado).
6	Total_Estudiantes	Total de estudiantes matriculados en el último ciclo de educación media superior.
7	Programados	Total de estudiantes programados para la evaluación.
8	Evaluados	Total de estudiantes evaluados en Lenguaje y Comunicación.
9	Logro_I	Número de estudiantes con un nivel insuficiente (logro I).
10	Logro_II	Número de estudiantes con un nivel elemental (logro II).
11	Logro_III	Número de estudiantes con un nivel satisfactorio (logro III).
12	Logro_IV	Número de estudiantes con un nivel sobresaliente (logro IV).

Fuente: elaboración propia.

Aplicación del algoritmo

El análisis de datos se basó en la aplicación del algoritmo K-means por sus características de funcionalidad y eficiencia, donde a través del Método Elbow se logró definir el número adecuado de clústeres, en los cuales se asignaron los vectores de datos (elementos), que incluyen el objeto de estudio. Además, mediante este algoritmo se optimiza la solución de problemas en los que los elementos están distribuidos en k clústeres, de modo que la suma de las varianzas internas de todos estos sean las más bajas posibles.

Además, debido a que los algoritmos particionales suponen un conocimiento a priori del número de clústeres en los que se debe dividir el conjunto de datos y, a diferencia de los algoritmos de clustering jerárquico, donde es posible definir cualquier número de clústeres en función de las necesidades del usuario, para este análisis el conocimiento previo del número adecuado de clústeres fue eficiente. Además, el uso de los centroides permitió representar, de forma general, cada uno de los clústeres obtenidos para describir sus comportamientos.

En este sentido, se utilizó como algoritmo K-means, el cual se implementó en Python para encontrar similitudes entre los vectores de datos de las escuelas de educación media superior de Veracruz, evaluadas como parte de PLANEA; de tal manera que se generaron clústeres con base en el siguiente proceso de asignación de elementos y actualización de los centroides:

1. Inicio: se establecieron centroides aleatorios para la formación de clústeres.
2. Asignación: cada elemento (vector de datos) fue asignado a su centroide más cercano.

3. Actualización: se calculó el promedio de todos los puntos asignados en el grupo para establecer el nuevo centroide.
4. Repita: los pasos 2 y 3 se repitieron iterativamente hasta que los centroides ya no cambiaron.

Para la implementación del algoritmo K-means se utilizó el módulo *sklearn.cluster*, que recopila algoritmos de segmentación de datos no supervisados. Se buscó asignar los elementos a través de las distancias mínimas, medidas por la distancia euclidiana, entre cada elemento y los centroides, logrando así una alta similitud intracluster y baja similitud intercluster. La ecuación de distancia euclidiana es la siguiente:

$$\text{dist}(p, c) = \sqrt{\sum_{i=1}^n (p_i - c_i)^2} \quad (1)$$

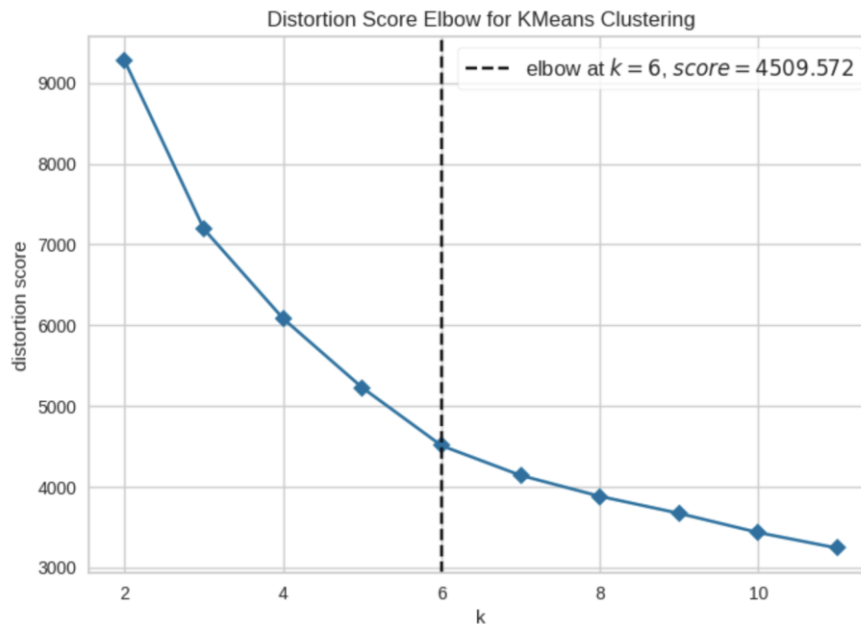
Donde: $P=\{p_1, \dots, p_n\}$ son los elementos del conjunto de datos y $C=\{c_1, \dots, c_k\}$ corresponde a los centroides. Debido a la necesidad de conocer a priori el número adecuado de clústeres, se estableció un rango de k configuraciones para la implementación del algoritmo. Este rango permitió ejecutar el algoritmo de forma iterativa para obtener los clústeres. Posteriormente, con base en la categorización resultante, se calculó la suma del error cuadrático (SSE) entre cada elemento del clúster formado, esto para cada configuración de k . En este sentido, SSE es la suma de las distancias euclidianas al cuadrado de cada elemento a su centroide más cercano, con base en la siguiente ecuación:

$$\text{SSE} = \sum_{k=1}^k \text{dist}(p_i, c_i) = \sum_{k=1}^k \sum_{p_i \in C_k} (p_i - c_i)^2 \quad (2)$$

Dado que se trata de una medida de error, el objetivo de K-means es intentar minimizar este valor. Esta medida de error se utiliza para realizar el método del codo, en el que se traza una curva con los valores obtenidos del SSE para encontrar un punto de inflexión o codo, a través del cual se establece el número óptimo de clústeres a analizar. Bajo la idea del objetivo de K-means, que es la minimización de la varianza intracluster y la maximización de la varianza intercluster, la gráfica resultante permite observar que mientras el número de clústeres es mayor, la varianza

intracluster tiende a disminuir; fenómeno a través del cual es posible identificar el codo que permite establecer el número de clústeres. La Figura 1 muestra el trazo de dicha curva, en la cual se observó que el efecto de la rodilla (cambio de orientación) es en k igual a 6. Además, con base en lo anterior, se observó la cantidad de elementos que conforman los clústeres obtenidos (Figura 2).

Figura 1. Método del codo para identificar el número apropiado de conglomerados.



Fuente: Elaboración propia.

Figura 2. Método del codo para identificar el número apropiado de conglomerados.

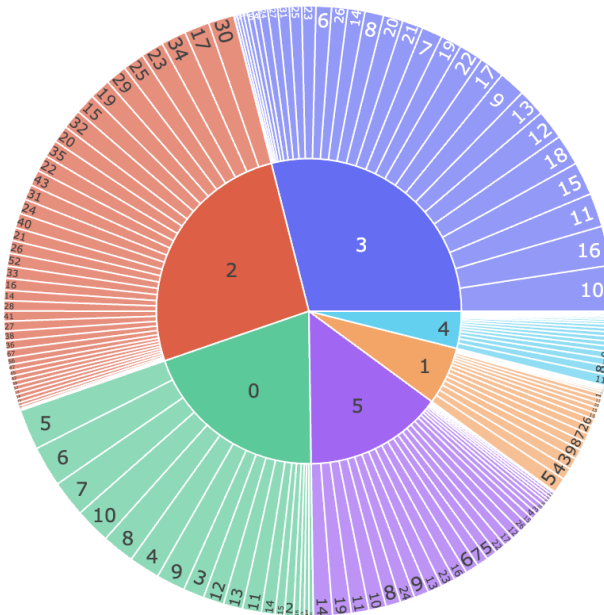
```
▶ PLANEALyC.groupby(['clusterP'])['clusterP'].count()
↳ clusterP
0      531
1      251
2      162
3      370
4       89
5      237
Name: clusterP, dtype: int64
```

Fuente: Elaboración propia.

Discusión de resultados

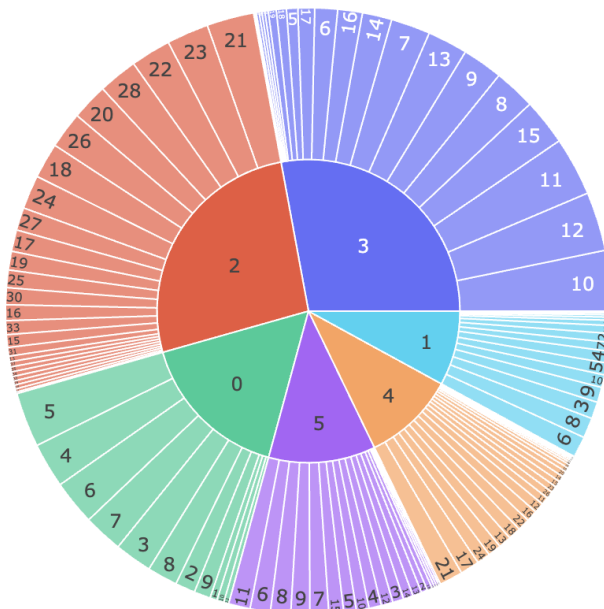
De los resultados obtenidos, derivados de la aplicación del algoritmo K-means, se observó una segmentación diferenciada para el periodo de evaluación. Así, a partir de las similitudes internas de los conglomerados y los valores de sus centroides, las Figuras 3 a 6 muestran un compendio del número de estudiantes que alcanzaron los diferentes logros académicos: i) deficiente -nivel I-, ii) elemental -nivel II, iii) satisfactorio -nivel III-, y iv) sobresaliente -nivel IV; a partir del cual se determinaron aspectos que se destacan en los seis clústeres, como el turno, grado de marginación, tipo de apoyo, estudiantes programados y estudiantes evaluados.

Figura 3. Segmentación en función del rendimiento académico I: Deficiente.



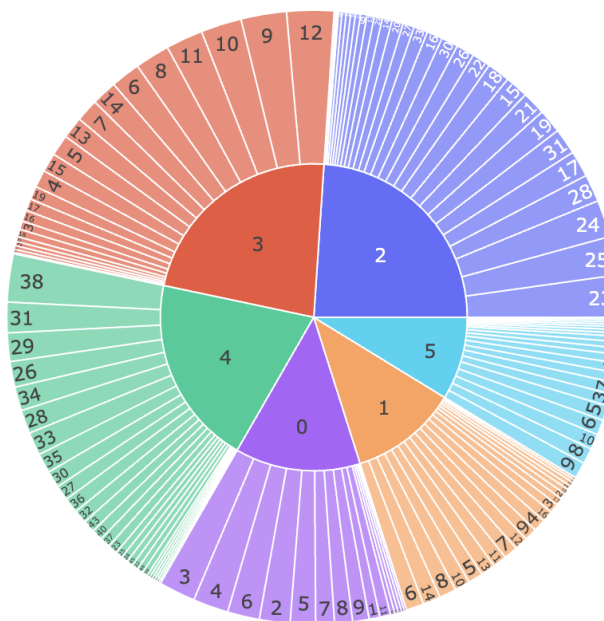
Fuente: elaboración propia

Figura 4. Segmentación en función del rendimiento académico II: Elemental.



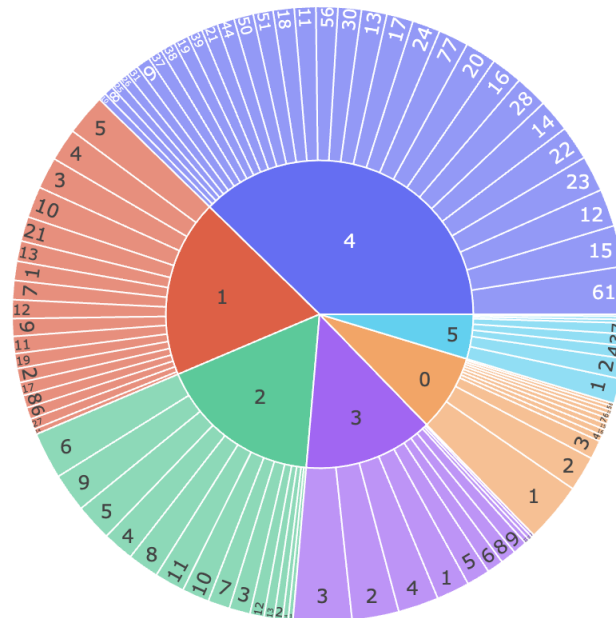
Fuente: elaboración propia

Figura 5. Segmentación en función del rendimiento académico III: Satisfactorio.



Fuente: elaboración propia

Figura 6. Segmentación en función del rendimiento académico IV: Sobresaliente.



Fuente: elaboración propia

Clúster etiquetado 0

Este conglomerado está conformado por 531 escuelas evaluadas, todas del turno matutino; con un alto grado de marginación, y apoyo principalmente estatal (530 escuelas) y una escuela federal (Centro de Bachillerato Tecnológico Agropecuario, CBTA). El número total de estudiantes programados para esta evaluación fue de 15 estudiantes en promedio, de los cuales 14 presentaron el examen. Con respecto a los logros académicos alcanzados, se observó:

- Más del 42% en promedio (6 estudiantes de 14) lograron un rendimiento deficiente (nivel I), lo que representa, según PLANEA, que los estudiantes tienen dificultades para identificar y extraer información específica y datos que se mencionan en ensayos, artículos de ciencia, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 28% (4 estudiantes de 14) alcanzaron un rendimiento elemental (nivel II), lo que representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes de un texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.

- El nivel III (satisfactorio) fue alcanzado por solo 3 estudiantes. Lo que significa que solo el 21% del total de estudiantes evaluados tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar sus propias conclusiones, descubrir características de los personajes de un cuento y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.
- En el nivel IV sobresale un solo estudiante en promedio (7%), quien obtuvo un logro destacado, lo que representa que comprende la función de los diferentes recursos utilizados en los textos, como: opiniones, explicaciones que apoyan argumentos, descripciones, figuras y otros. Además, distinguen los elementos estructurales y de contenido, siendo capaces de identificar el sentido global de un texto, la intención comunicativa de su autor y la secuencia lógica del proceso comunicativo.

Clúster etiquetado 1

Clúster conformado por 251 escuelas, la mayoría del turno matutino (222 escuelas) y otras del turno vespertino (29 escuelas); con un bajo nivel de marginalización; con apoyo principalmente particular (220 escuelas) y solo una escuela federal. El número total de alumnos programados para la evaluación fue de 21, de los cuales 18 presentaron el examen. Con respecto a los logros académicos alcanzados destacan:

- El 22% en promedio (4 estudiantes de 18) lograron un rendimiento deficiente (nivel I), lo que representa que los estudiantes tienen dificultades para identificar y extraer información específica y datos que se presentan en los ensayos, artículos, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas, que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 22% (4 estudiantes de 18) alcanzaron el rendimiento elemental (nivel II), lo que representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes de un texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.
- En el nivel III (satisfactorio) lograron destacar 6 estudiantes de los 18 evaluados, lo que representa el 34% en promedio. Esto significa que los estudiantes que lograron este nivel tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar

sus propias conclusiones, descubrir características de los personajes de un cuento y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.

- En el nivel IV (sobresaliente) fueron 4 los estudiantes que lograron un destacado desempeño (22%). Esto representa que estos estudiantes logran comprender la función de los diferentes recursos utilizados en los textos (opiniones, explicaciones que apoyan argumentos, descripciones, figuras y otros). Además, distinguen los elementos estructurales y de contenido, siendo capaces de identificar el sentido global de un texto, la intención comunicativa del autor y la secuencia lógica del proceso comunicativo.

Clúster etiquetado 2

Clúster conformado por 162 escuelas, la mayoría del turno matutino (101 escuelas) y algunas con actividades vespertinas (61 escuelas); con un bajo nivel de marginación, y con apoyo principalmente estatal (101 escuelas), también federal (56 escuelas) y en menor medida solo 5 escuelas particulares. El número total de alumnos programados fue de 79, de los cuales 72 presentaron el examen. Con respecto a los logros académicos alcanzados, se establece:

- Alrededor del 35% en promedio (25 estudiantes de 72) lograron un rendimiento deficiente (nivel I), lo que representa que los estudiantes tienen dificultades para identificar y extraer información específica y datos que se presentan en los ensayos, artículos, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas, que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 30% (22 estudiantes de 72) alcanzaron el rendimiento elemental (nivel II). Esto representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes de un texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.
- En el nivel III (satisfactorio) lograron destacar 20 estudiantes de los 72 evaluados, lo que representa el 28% en promedio. Esto significa que los estudiantes que lograron este nivel tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar sus propias conclusiones, descubrir características de los personajes y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.

- En el nivel IV (sobresaliente) destacaron solo 5 estudiantes (7%), lo que representa que estos logran comprender la función de los diferentes recursos utilizados en los textos (opiniones, explicaciones que apoyan argumentos, descripciones, figuras y otros). Además, distinguen los elementos estructurales y de contenido, siendo capaces de identificar el sentido global de un texto, la intención comunicativa del autor y la secuencia lógica del proceso comunicativo.

Clúster etiquetado 3

Clúster conformado por 370 escuelas, todas del turno vespertino; con un grado de marginación alto y medio, y principalmente de soporte estatal (358 escuelas), seguido de 11 escuelas federales y una escuela particular (Olmeca de Texistepec). El número total de alumnos programados fue de 34, de los cuales 32 presentaron el examen. Con respecto a los logros académicos alcanzados, destacan:

- Más del 37% en promedio (12 estudiantes de 32) lograron un rendimiento deficiente (nivel I), lo que representa que los estudiantes tienen dificultades para identificar y extraer información específica que se menciona en ensayos, artículos, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 31% (10 estudiantes de 32) alcanzaron un rendimiento elemental (nivel II), lo que representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes del texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.
- En el nivel III (satisfactorio) lograron destacar 8 estudiantes de los 32 evaluados (solo el 25% en promedio). Esto significa que los estudiantes que lograron este nivel tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar sus propias conclusiones, descubrir características de los personajes y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.
- En el nivel IV (sobresaliente) destacaron solo 2 estudiantes (6% en promedio), lo que representa que estos logran comprender la función de los diferentes recursos utilizados en los textos (opiniones, explicaciones que apoyan argumentos, descripciones, figuras y

otros). Además, distinguen los elementos estructurales y de contenido, siendo capaces de identificar el sentido global de un texto, la intención comunicativa del autor y la secuencia lógica del proceso comunicativo.

Clúster etiquetado 4

Clúster conformado por 89 escuelas, la mayor parte del turno matutino (75 escuelas) y 14 con actividades vespertinas; con un bajo nivel de marginación, y con soporte mayoritariamente estatal (49 escuelas), 21 escuelas federales y 19 particulares. El número total de alumnos programados por escuela, en promedio, fue de 79, de los cuales, en promedio, 73 presentaron el examen. Con respecto a los logros académicos alcanzados, destacan:

- El 9% en promedio (7 estudiantes de 73) lograron un rendimiento deficiente (nivel I), lo que representa que los estudiantes tienen dificultades para identificar y extraer información específica y datos que se mencionan en ensayos, artículos, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 20% (15 estudiantes de 73) alcanzaron un rendimiento elemental (nivel II), lo que representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes del texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.
- El nivel III (satisfactorio) fue alcanzado por 30 estudiantes. Lo que significa que el 41% del total de estudiantes evaluados tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar sus propias conclusiones, descubrir características de los personajes de un cuento y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.
- En el nivel IV sobresalen 21 estudiantes evaluados (29%), quienes obtuvieron un logro destacado, lo que representa que comprenden la función de los diferentes recursos utilizados en los textos (opiniones, explicaciones que apoyan argumentos, descripciones, figuras y otros). Además, distinguen los elementos estructurales y de contenido, siendo capaces de identificar el sentido global de un texto, la intención comunicativa del autor y la secuencia lógica del proceso comunicativo.

Clúster etiquetado 5

Clúster conformado por 237 escuelas, la mayor parte del turno vespertino (231 escuelas) y 6 con actividades nocturnas; con un nivel de marginación medio, y con soporte mayoritariamente estatal (221 escuelas), 5 escuelas federales y 11 particulares. El número total de alumnos programados fue de 24, de los cuales 22 presentaron el examen. Con respecto a los logros académicos alcanzados, destacan:

- El 45% en promedio (10 estudiantes de 22) lograron un rendimiento deficiente (nivel I), lo que representa que los estudiantes tienen dificultades para identificar y extraer información específica y datos que se mencionan en ensayos, cuentos y otros textos relacionados. Además, los estudiantes se limitan a buscar información local, probablemente para resolver preguntas directas que no requieren de un análisis en profundidad sobre el contenido del texto.
- Alrededor del 27% (6 de 22 estudiantes) alcanzaron un rendimiento elemental (nivel II), lo que representa que los estudiantes relacionan y ordenan información que aparece en diferentes partes del texto. Esto les permite identificar el tema central de uno o varios párrafos, reconocer el significado de las palabras que se utilizan y distinguir los diferentes hechos y opiniones.
- El nivel III (satisfactorio) fue alcanzado por 5 estudiantes. Lo que significa que el 23% del total de estudiantes evaluados tienen habilidades para relacionar la información de un texto, siendo capaces de elaborar sus propias conclusiones, descubrir características de los personajes y suponer sus posibles acciones. Además de resumir apartados de un ensayo o de un artículo científico.
- En el nivel IV sobresale un solo estudiante (5%), quien obtuvo un logro destacado, lo que representa que comprende la función de los diferentes recursos utilizados en los textos (opiniones, explicaciones que apoyan argumentos, descripciones, figuras y otros). Además, distingue los elementos estructurales y de contenido, siendo capaz de identificar el sentido global de un texto, la intención comunicativa del autor y la secuencia lógica del proceso comunicativo.

Conclusiones

El rendimiento académico en la educación media superior ha experimentado serias dificultades en los últimos años debido a la pandemia por COVID-19, que impidió su fortalecimiento. Siendo este tipo de educación obligatoria una parte importante del sistema educativo del país. Por lo que, es fundamental entender su comportamiento y, de la mano de algoritmos especializados de inteligencia computacional, es posible hacerlo.

Precisamente, el uso de métodos de inteligencia computacional, específicamente algoritmos de aprendizaje automático, permite identificar patrones de datos para promover una mejor comprensión del comportamiento del rendimiento académico en Lenguaje y Comunicación de los estudiantes de educación media superior en Veracruz, y que, junto con el uso de tecnologías de información, se busca que estas tengan crecimiento para enfrentar los nuevos desafíos en materia de tecnología.

Por otro lado, con el objetivo de identificar los logros académicos en Lenguaje y Comunicación de los estudiantes de educación media superior de Veracruz, a través de un enfoque basado en la inteligencia computacional, cobra relevancia, gracias a que estos métodos brindan una vía a través de la cual se puede abordar la resolución de problemas. Por esta razón, la aplicación del algoritmo de agrupamiento K-means permitió ampliar el entendimiento del comportamiento del rendimiento académico en la educación media superior.

De los seis conglomerados identificados, el clúster 5 se caracteriza por tener uno de los porcentajes más altos de estudiantes con rendimiento académico deficiente (45%) y elemental (27%). Caso similar ocurre con el clúster 0, donde el logro académico alcanzado por los estudiantes fue, principalmente, deficiente (42%) y elemental (28%). El patrón anterior también se repite en los clústeres 2 y 3, donde predomina los logros: deficiente y elemental. En el caso específico del clúster 3, el logro académico fue mayoritariamente deficiente (37%) y elemental (31%); mientras que en el caso del clúster 2, el 35% de los estudiantes obtuvieron un logro deficiente y el 30% elemental. Esto contrasta con la deteriorada situación educativa del nivel analizado, esto es, la educación media superior en el estado de Veracruz.

No obstante, en el único segmento con mayor número de estudiantes evaluados y que lograron rendimiento sobresaliente fue el clúster etiquetado 4, conformado por 89 escuelas, la mayoría de las cuales cuentan con soporte estatal. En este clúster el 41% obtuvo rendimiento satisfactorio y el 29% sobresaliente, siendo este el grupo con mejores rendimientos académicos.

Los resultados de rendimiento académico reflejan los diversos factores sociales, culturales y económicos en los que viven los estudiantes. Además, debido a la diversidad de instituciones educativas del sistema educativo, el rendimiento académico requiere de una atención diferenciada

en cada entidad, tipo de servicio y tipo de control administrativo. Para enfrentar esta situación, es necesario reducir las brechas de conocimientos, oportunidades y condiciones de enseñanza y aprendizaje, desde el inicio de la educación obligatoria.

Los resultados confirman los bajos niveles educativos a nivel estatal de los estudiantes de educación media superior en Lenguaje y Comunicación. Por lo tanto, existen retos en el sistema educativo actual por tener y ofrecer una mejor calidad educativa, ya que se evidencia la inequidad entre los estudiantes que asisten a los diferentes centros educativos. Si se realizaran seguimientos periódicos se podría saber si estas distancias se acortan.

Futuras líneas de investigación

Si bien el objetivo de la investigación se cumplió, el cual fue analizar el rendimiento académico en Lenguaje y Comunicación de estudiantes de Educación Media Superior de Veracruz, a través de un enfoque basado en la Inteligencia Computacional, los resultados obtenidos del análisis dejaron abierta futuras líneas de investigación, entre las que destacan:

- (1) Realizar un análisis comparativo utilizando datos de PLANEA antes y después de la pandemia por COVID-19. Esto con el propósito de conocer en qué medida la pandemia afectó la educación media superior en Veracruz y otros estados del país.
- (2) Ampliar el análisis a otros niveles académicos de la educación obligatoria, como primaria y secundaria. Aunque es conocida la importancia de la educación media superior por ser la etapa que antecede a los estudios profesionales, resulta interesante también conocer los logros académicos en primaria y secundaria, como etapas previas a la educación media superior.
- (3) Realizar el análisis utilizando otros algoritmos de inteligencia computacional con el fin de ampliar la visión de la situación actual de los logros académicos alcanzados en la educación media superior.

Referencias

- Briones S., Dagamac R., David J. and Landerio C.(2022). Factors Affecting the Students' Scholastic Performance: A Survey Study. *Indonesian Journal of Educational Research and Technology*, 2(2), 97–102. Recuperado de <https://doi.org/10.17509/ijert.v2i2.41394>.
- Céspedes-González Y. Otero A., Ricardez J. y Molero-Castillo G. (2024). Academic achievement in mathematics of Higher-Middle Education students in Veracruz: An approach based on computational intelligence. *11th International Conference in Software Engineering Research and Innovation (CONISOFT)*.
- García C. y Gomez I. (2006). Algoritmos de aprendizaje: knn & kmeans. *Inteligencia en Redes de Comunicación, Universidad Carlos III de Madrid*, 2006.
- Gutiérrez I., Gutiérrez D., Juan J., Rodríguez L., Rico R. y Sánchez M. (2020). Aplicación del algoritmo Kmeans para el análisis de resultados de la prueba PLANEA 2017. *Research in Computing Science*, 149(8), 407–419. Recupera de: <https://api.semanticscholar.org/CorpusID:232314306>
- Hatos, A. (2014). Educational System. *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. Recuperado de https://doi.org/10.1007/978-94-007-0753-5_844
- Ibarrola M. (2012). Los grandes problemas del sistema educativo mexicano. *Perfiles educativos*, xxxiv(), 16–28.
- INEE (2016). Planea: Una nueva generación de pruebas. Mexico: INEE.
- INEE (2018). Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). Recuperado de: <http://planea.sep.gob.mx/content/general/docs/2018/PlaneaDocumentoRector18.pdf> [Consultado: 25 En. 2024].
- INEE (2024a). La educación Media Superior en México. Informe 2010-2011. Recuperado de: <https://www.inee.edu.mx/wp-content/uploads/2018/12/P1D235.pdf> [Consultado: 22 Mar. 2024].
- INEE (2024b). El ABC de PLANEA. Recuperado de: <https://www.inee.edu.mx/evaluaciones/planea/abc-planea/> [Consultado: 15 Feb. 2024].
- Lorenzo O. y Zaragoza J. E. (2014). Educación Media y Superior en México: análisis teórico de la realidad actual. *DEDiCA Revista De Educação E Humanidades*, 6, 59–72. Recuperado de <https://doi.org/10.30827/dreh.v0i6.6961>.
- Maldonado G., Molero-Castillo G., Rojano-Cáceres J. y Velázquez-Mena A. (2016). Análisis del logro académico de estudiantes en el nivel medio superior a través de minería de datos

- centrada en el usuario. *Research in Computing Science*, 125(1), 121–133. Recuperado de: <https://doi.org/10.13053/rcs-125-1-11>.
- Mathur P. (2019). Overview of machine learning in retail. *Machine Learning Applications Using Python*, Apress, Berkeley, CA, 147–157. Recuperado de https://doi.org/10.1007/978-1-4842-3787-8_7.
- Méndez P., Tomás J. y Marín R. (2008). Inteligencia artificial: Métodos, técnicas y aplicaciones. *Madrid, MacGraw-Hill*, pp. 1022, 2008.
- Molero-Castillo G., Maldonado-Hernández G., Mezura-Godoy C. y Benítez-Guerrero E. (2018). Interactive system for the analysis of academic achievement at the upper-middle education in Mexico. *Computación y Sistemas*, 22, 223–233. Recuperado de <https://doi.org/10.13053/CyS-22-1-2773>.
- Molero-Castillo G., Bárcenas E., Velázquez-Mena A. y Céspedes-González Y (2019). Analysis of Academic Achievement in Higher-Middle Education in Mexico through Data Clustering Methods. *Education Systems Around the World. IntechOpen*, p. 93. Recuperado de <https://doi.org/10.5772/intechopen.84744>
- OECD (2024). ¿Qué es Pisa?. Programme for International Students Assessment. Recuperado de www.oecd.org/pisa/pisa-es/ [Consultado: 01 Feb. 2024].
- Pascual D., Pla F. y Sánchez S. (2007). Algoritmos de agrupamiento. *Método Informáticos Avanzados*, pp. 164–174.
- PLANEABD (2024). Plan Nacional para la Evaluación de los Aprendizajes: Bases de datos PLANEA 2017. Recuperado de: http://planea.sep.gob.mx/ms/base_de_datos_2017/, [Consultado: 10 En. 2024].
- Rouhiainen L. (2018). Inteligencia artificial. *Madrid: Alienta Editorial*, pp. 352.
- Sauvageot C. (1997). Indicators for educational planning: a practical guide. Paris: International Institute for Educational Planning
- SEP (2016). Manual para Usuarios Plan Nacional para las Evaluaciones de los Aprendizajes en el nivel medio superior. *Centro Nacional de Evaluación para la Educación Superior*.
- SEP (2024). Plan Nacional para la Evaluación de los Aprendizajes en la Educación Media Superior. Recuperado de: <http://planea.sep.gob.mx/ms>. [Consultado: 20 Feb. 2024].
- Soto A., Ponzoni I y Vázquez G. (2006). Análisis numérico de diferentes criterios de similitud en algoritmos de clustering. *Mecánica Computacional*, XXV, 993–1011.

- UNESCO (2019). Estudio Regional Comparativo y Explicativo (ERCE 2019). Recuperado de: www.unesco.org/es/articles/estudio-regional-comparativo-y-explicativo-erce-2019 [Consultado: 15 Mar. 2024].
- Villa L. (2013). Modernización de la educación superior, alternancia política y desigualdad en México. *Revista de la Educación Superior*, 42(168), 81–103.
- Villa L.(2007). La educación media superior ¿igualdad de oportunidades?. *Revista de la educación superior*, 36(141), 93–110.
- Ward A., Stoker H. W. y Murray-Ward M. (1996). Achievement and Ability Tests - Definition of the Domain. *Educational Measurement*, 2, 2–5.
- Wallace D. (2009). Parts of the Whole: Approaching Education as a System. *Numeracy*, 2(2), 9. Recuperado de [https:// dx.doi.org/10.5038/1936-4660.2.2.9](https://dx.doi.org/10.5038/1936-4660.2.2.9)