



*Asesorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada, Toluca, Estado de México. 7223898475*

RFC: ATII20618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.

<http://www.dilemascontemporaneoseducacionpoliticayvalores.com/>

Año: IX

Número: Edición Especial.

Artículo no.:19

Período: Octubre, 2021

TÍTULO: Caracterización del aprendizaje de algoritmos mediante minería de datos en el nivel superior.

AUTORES:

1. Ing. Jonathan Isael Duque Hernández.
2. Máster. Mario Humberto Rodríguez-Chávez.
3. Dr. Said Polanco-Martagón.

RESUMEN: La presente investigación se enfoca en el desarrollo de un modelo de caracterización de aprendizaje para estudiantes universitarios en el área de programación, basado en una taxonomía de nivel cognoscitivo de (Marzano, R. J., 2001). Se realizó una recopilación de respuestas de cuestionarios obtenidas de los estudiantes para un tratamiento y análisis de datos; de esta forma, obtener información necesaria y caracterizar el aprendizaje del estudiante identificando sus debilidades y fortalezas. Tras realizar pruebas se obtuvieron 5 agrupamientos que representan los niveles cognoscitivos, donde el 29% representa la mayor población en el primer nivel y el 13% se encuentra en el segundo nivel como la menor población.

PALABRAS CLAVES: caracterización de aprendizaje, algoritmos, objetos de aprendizaje, nivel cognoscitivo, modelo.

TITLE: Characterization of algorithm learning through data mining in the higher level.

AUTHORS:

1. Eng. Jonathan Isael Duque Hernández.
2. Master. Mario Humberto Rodríguez-Chávez.
3. PhD. Said Polanco-Martagón.

ABSTRACT: This research focuses on the development of a learning characterization model for university students in the programming area, based on a cognitive level taxonomy of (Marzano, R. J., 2001). A compilation of questionnaire responses obtained from the students was carried out for data treatment and analysis; in this way, obtain the necessary information and characterize the student's learning, identifying their strengths and weaknesses. After performing tests, 5 groupings were obtained that represent the cognitive levels, where 29% represent the largest population in the first level and 13% are in the second level as the smallest population.

KEY WORDS: characterization of learning, algorithms, learning objects, cognitive level, model.

INTRODUCCIÓN.

La deserción académica puede ser ocasionada por diversos factores; en el trabajo realizado por (Páramo, G., & Correa Maya, C., 1999) se indica que el ambiente familiar, educativo, situación económica y el bajo desempeño son algunas de las causas.

Se le conoce como deserción al abandono definitivo de clase o la no continuidad de formación académica (Páramo, G., & Correa Maya, C., 1999); también mencionan que existe una magnitud alarmante de deserción.

Algunos estudiantes con bajo desempeño mantienen características similares a los alumnos que han desertado (Páramo, G., & Correa Maya, C., 1999), entre las características que más destacan son: bajo aprovechamiento de oportunidades educativas, padres con bajo interés en la formación

académica, bajo deseo por el conocimiento, desmotivación hacia la carrera y/o universidad y resistirse a desarrollar actividades formativas.

La deserción es un tema preocupante en la Universidad Politécnica de Victoria (UPV), según registros del Sistema Integral de Información de la UPV (SIUPV); la Figura 1 muestra índices de deserción y retención desde el año 2014.

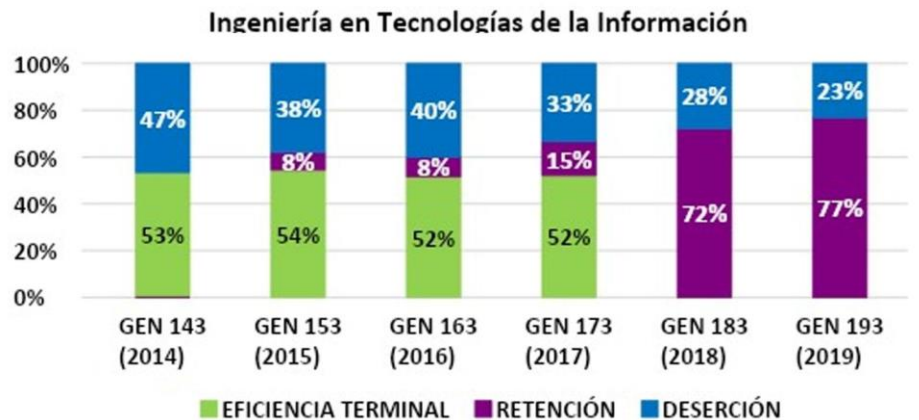


Figura 1. Índice de deserción académica según los registros del SIUPV.

Ekubo E.A. (2019) comparte su trabajo sobre recolección de datos educativos, donde busca identificar causas o factores que lleven a una deserción académica, expone un proceso de recolección basado en la literatura de minería de datos educacional. Este primer proceso consta en preparar un conjunto de datos para aplicar técnicas de minería de datos posteriormente.

El método expuesto comienza con una obtención de permisos para adquirir el consentimiento de los estudiantes, agendan reuniones para dar información respecto a la solicitud de datos, busca información disponible en la literatura sobre el tema, realiza cuestionarios a estudiantes con bajo rendimiento académico y crea un formato de conveniencia para el conjunto de datos. Algunos datos pueden ser innecesarios o ilegibles hablando de un campo vacío; para esto es necesario realizar una

limpieza de datos, y finalmente, obtener un dataset listo para ser procesado y analizado. En la Figura 2 se puede observar el procedimiento de recolección de datos.

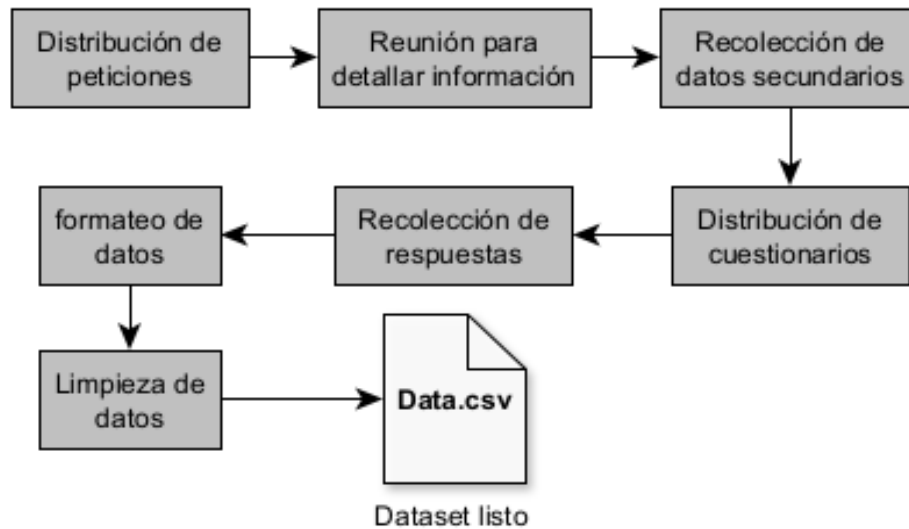


Figura 2. Proceso de recolección de datos según (Ekubo EA, 2019).

Ekubo, E. A. (2019) determinó que el algoritmo MDBC (Hacer un cluster basado en la densidad) ha realizado con éxito una agrupación con dos dataset de estudiantes, mediante las similitudes de características que indican un bajo rendimiento académico. Se determinó en ambos clúster, que los estudiantes presentan una media de ausencias en clase, mayor a una semana; además, estos estudiantes cuentan ausencia paterna en la educación de distintas maneras; es decir, que a sus padres no les interesa la educación. A esto se le conoce como aprendizaje no supervisado (ANS).

Masashi Sugiyama (2016) indica, que el ANS se basa en trabajar con un conjunto de datos obteniendo información que un humano no puede percibir a simple vista, en otras palabras, extrae datos sin etiquetar, mientras que el aprendizaje supervisado se utiliza frecuentemente en problemas como reconocimiento de escritura, de voz, imágenes, en sistemas de recomendación y pronósticos, obteniendo excelentes resultados ya que se tratan de datos etiquetados.

Ekubo, E. A. (2020) utiliza múltiples técnicas de clasificación tales como: J48, Logistic regression, MLP, NB (Naive Bayes) y SMO, los resultados son comparados y determina que el algoritmo MLP (Perceptrón Multicapa) ha obtenido los mejores resultados. El algoritmo MLP clasificó el bajo rendimiento de los alumnos y obtuvo alrededor de 1596 estudiantes clasificados correctamente y 47 estudiantes clasificados incorrectamente, mientras que el resto obtuvo 1580, 1578, 1566 y 1416 clasificaciones correctas.

La metodología utilizada por (Ekubo, E. A. 2020) consiste en recolectar datos como se explicó anteriormente en cuestión de minería de datos educacional, realiza un sistema predictivo para caracterizar el bajo rendimiento de estudiantes universitarios de último año, por medio de técnicas de aprendizaje automático. En la figura 3 se observa el proceso de minería de datos, seguido de una modelización para caracterización, la utilización de técnicas para la selección de características y técnicas para evaluación de modelos.

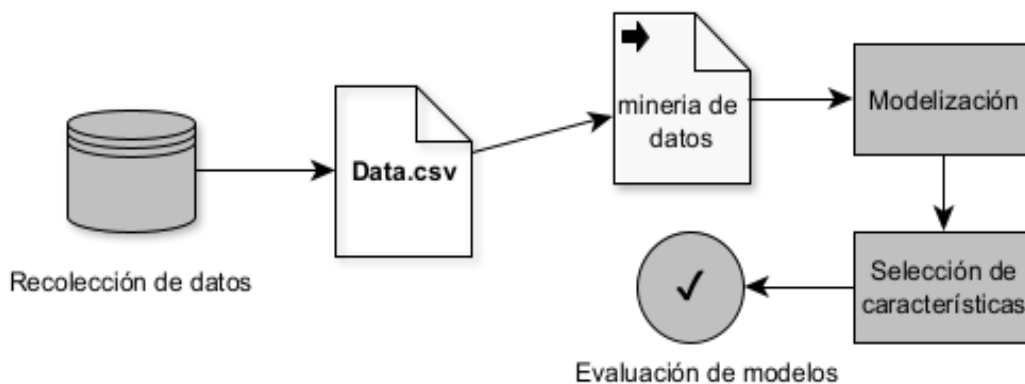


Figura 3. Metodología de caracterización de rendimiento de estudiantes según Ekubo, E. A. (2020).

Dang, Shilpa (2015) ha compartido resultados al someter cuatro de las técnicas de clustering más utilizadas, en cinco distintos datasets de rendimiento académico de estudiantes, los cuales algunos se encuentran disponibles de forma pública. Las técnicas que pone a prueba son: K-Means, DBSCAN (agrupamiento espacial basado en densidad de aplicaciones con ruido), HCA (Agrupación jerárquica)

y MDBA (Algoritmo de abeja distribuido modificado). Los resultados son importantes para realizar un análisis y llegar a un veredicto sobre la utilización de estas técnicas en el presente trabajo de investigación; sin embargo, los datasets no recopilan las características que se buscan sobre el rendimiento académico en alumnos de nivel superior para el área de algoritmos computacionales. Dang, Shilpa, (2015) concluyó que el algoritmo K-Means obtuvo los mejores resultados frente a datasets con instancias desde 100 a 2924 y de 8 a 1013 atributos, ya que realiza el agrupamiento en un tiempo mínimo en comparación a DBSCAN, HCA y MDBA. El algoritmo MDBA se acerca un poco al rendimiento de K-Means, HCA muestra demasiada variación con datos ruidosos, así mismo demuestra que DBSCAN no es una buena opción para datos que representen una alta varianza en la densidad, es decir que los datos se muestran dispersos, muy alejados de la media.

Las técnicas de caracterización (TC) ayudan a identificar de manera detallada a un sujeto, después de realizar un agrupamiento de características, según Dash, M., & Liu, H. (1997), las TC son aquellos métodos que se encargan de representar la información que compone a un elemento, se utilizan para clasificar un conjunto de datos en distintas clases o tipos de objetos. En la Figura 4 se puede apreciar el objetivo de una caracterización con 2 subconjuntos de características o atributos.

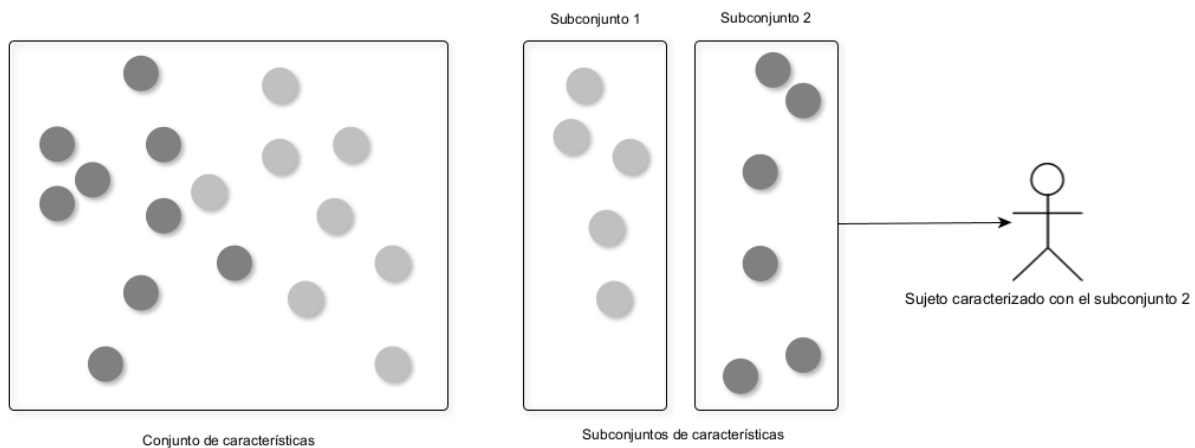


Figura 4. Caracterización de un sujeto u objeto.

Los trabajos presentados por Ekubo, E. A. (2020) y el resto de la literatura que buscan realizar este tipo de modelo, se basan en la recopilación de datos en un rango de años definido; se buscó realizar este proceso de manera continua e indefinida, esto da a entender que los estudiantes de una institución pueden llegar a tener cambios repentinos en su rendimiento académico, si se mantiene el proceso de recopilación, se prestarán más oportunidades de análisis y se podría detectar a tiempo cualquier señal de aviso.

En cuestión de caracterización, existen métodos pedagógicos que pueden identificar el dominio de temas de un alumno; sin embargo, la taxonomía de Marzano, R. J. (2001) ofrece una segmentación de niveles cognoscitivos que determinan el dominio para un tema en cuestión; esta taxonomía es fácilmente adaptable a los temas de construcción de algoritmos, ya que contiene segmentos como: Conocimiento, Comprensión, Análisis, Aplicación y Metacognición, y estos segmentos se comprenden por atributos o características como: Identificar, Interpretar, Distinguir, Aplicar, Elaborar, entre otros; de esta manera, es más fácil identificar qué atributos o características definen el conocimiento de un estudiante, ya que es factible elaborar objetos de aprendizaje que comprendan estos segmentos y atributos.

Debido a que no todos los estudiantes aprenden de la misma manera, se requirió partir del conocimiento previo con el que cuenta cada alumno; Murphy, P., (2008) menciona que un estudiante no aprende igual a los demás, debido a que cada uno tiene un conocimiento previo distinto; por tanto, se utilizaron técnicas de caracterización en estudiantes de nivel superior, la caracterización de los estudiantes se basa en el análisis de resultados de un cuestionario contestado por los alumnos de la Universidad Politécnica de Victoria.

El objetivo de la presente investigación es el procesamiento de características que determinen el conocimiento de aprendizaje de algoritmos computacionales, identificando las habilidades o aptitudes con la que cuentan los alumnos a través de inteligencia artificial, haciendo uso de metodología de

aprendizaje automático. Los estudiantes candidatos a este proceso son aquellos que han presentado malos resultados a lo largo de la carrera, desde calificaciones hasta el interés y motivación, así mismo, cualquier estudiante puede ser caracterizado de tal forma que se busque mantener en un buen estado cognoscitivo y evitar que su rendimiento decaiga. Es importante establecer el tipo de aprendizaje que requiere cada alumno para fomentar aptitudes que conforman al estudiante de ITI en la UPV.

DESARROLLO.

Con base en la literatura, se detalla la metodología propuesta que se utilizó en el proceso de caracterización para el desarrollo de esta investigación; a continuación, se explica de manera breve, algunos procesos, métodos y herramientas, que ayudaron al desarrollo de este proyecto con eficiencia. Ekubo, E. A. (2020) realiza una metodología similar, lo que diferencia este proceso con el que se utilizó en el presente trabajo expuesto y de otros autores, es la manera de recopilación de datos, ya que además de utilizar encuestas, o exámenes, también se analizan registros de los alumnos de la carrera de ITI de la UPV; además se extrae información del aprendizaje y rendimiento de los estudiantes de un sistema de tutoría inteligente (ITS) sobre objetos de aprendizaje. Esto permite que el análisis sea constante y más fácil de apreciar la evolución de conocimiento del estudiante. En la Figura 5 se representa la metodología que se ha seguido en base a la literatura para alcanzar el objetivo, la cual se describe en las siguientes secciones.

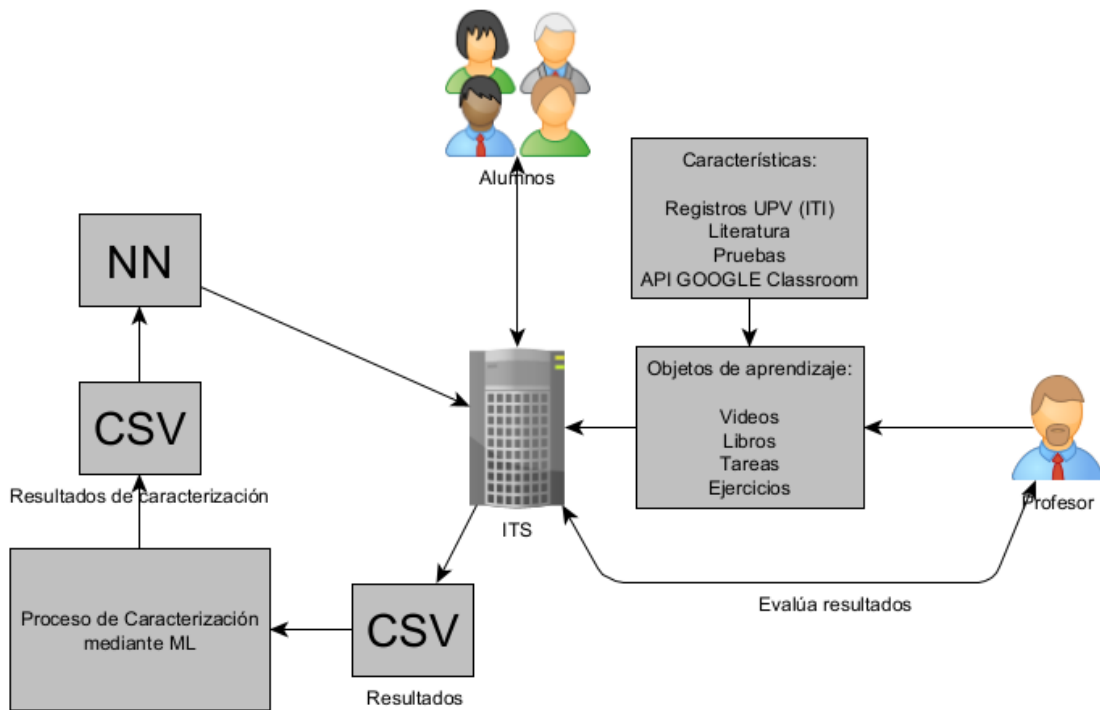


Figura 5. Proceso de adquisición de datos y caracterización.

Adquisición y Tratamiento de Datos.

El procedimiento de recolección de datos que se aborda para este trabajo está basado en el proceso de recolección y minería de datos educativos expuesto por Ekubo E.A., (2019). Las métricas pautadas en Marzano, R. J., (2001) se utilizan para en el proceso de estructuración de datos, el cual se divide en 3 fases: Adquisición, Reconocimiento de características, y Estructuración final.

Adquisición de datos.

La caracterización de aprendizaje de estudiantes requiere registros previos de cada alumno, es importante conocer un histórico del estudiante para poder determinar un resultado o una predicción; para este trabajo, se consideró una encuesta o examen de prueba en alumnos de nuevo ingreso y alumnos de último año de la UPV. La encuesta está formada por 5 reactivos para cada nivel cognoscitivo, donde cada reactivo considera características de acuerdo a lo que dicta Marzano, R. J., (2001). En este caso, la opción para la encuesta o examen inicial es determinante, ya que se puede

obtener información sobre el conocimiento del aspirante, de esta manera, se comparan los resultados con respuestas de alumnos que se encuentran a mediación o en la culminación de sus estudios superiores.

Estructuración de datos.

Al conseguir una cantidad de registros, que se considere óptima para inicializar un proceso de caracterización, serán guardados como archivo csv; los datos obtenidos podrán representar la adquisición o la portación de características en los estudiantes, ya sea, mediante.

- 1 = Si la tiene.
- 0 = No la tiene.

o alguna puntuación, tal que.

- pts < 80 = No la tiene.
- pts >= 80 = Si la tiene.

El medio de obtención de este puntaje o de elección entre 1 y 0 se encuentra entre cada objeto de aprendizaje; los objetos de aprendizaje que cada maestro proporcione deben ser relacionados con la taxonomía de Marzano, R. J., (2001); cada objeto abarca ciertos temas que representan un nivel cognitivo y los atributos que se puedan obtener de ese nivel, en las Tablas 1 y 2 se pueden observar ejemplos de estructuración de datos.

Tabla 1: Ejemplo de estructuración de datos en representación binaria.

Alumno	característica_1	característica_2	característica_3	característica_4
#_matrícula	1	0	1	1
#_matrícula	0	1	1	0

Tabla 2: Ejemplo de estructuración de datos en representación de puntaje.

Alumno	característica_1	característica_2	característica_3	característica_4
#_matrícula	76	55	80	92
#_matrícula	86	79	99	67

Tratamiento de datos.

Antes de iniciar el proceso de caracterización, es necesario realizar una limpieza de datos, como ha mencionado Ekubo E.A., (2019), en un dataset puede contener datos que causarán ruido para el proceso que se desea llevar a cabo, pueden ser campos en blanco (espacios en blanco), caracteres especiales o nulos (campos sin llenar).

El dataset obtenido inicialmente por la encuesta cuenta 117 columnas de 39 esperadas, de las cuales gran parte de los registros contienen campos en blanco, con caracteres especiales, nulos e incluso columnas repetidas debido a que se utilizó Google Forms como plantilla; por lo tanto, fue necesario realizar una limpieza y obtener un documento formateado a conveniencia.

El dataset se encuentra segmentado en: “Datos Generales del alumno,” “Conocimientos sobre desarrollo de software”, “Prácticas de programación”, “Estudios, conocimientos y habilidades”, y “Tutorías y asesorías”; la sección de “Prácticas de programación” es donde se encuentran los reactivos relacionados con los niveles cognoscitivos que dicta Marzano, R. J., (2001) y es la sección que tiene más peso en el dataset. En la figura 6 se pueden observar campos repetidos y con inconsistencias.

Marca temp	Puntuación	1. Edad:	1. Edad: [Pur	1. Edad: [Cor	2. Correo ins	2. Correo ins	2. Correo ins	3. Cuatrimes
2021/04/13	40.00 / 160	20	-- / 0		1930315@up	-- / 0		Cuarto
2021/04/13	100.00 / 160	20	-- / 0		1930318@up	-- / 0		Quinto
2021/04/13	100.00 / 160	20	-- / 0		1830331@up	-- / 0		SÃ@ptimo
2021/04/13	81.40 / 160	19	-- / 0		1930420@up	-- / 0		Quinto
2021/04/13	71.40 / 160	20	-- / 0		1830078@up	-- / 0		SÃ@ptimo
2021/04/13	70.00 / 160	19	-- / 0		1930136@up	-- / 0		Quinto
2021/04/13	100.00 / 160	42	-- / 0		1930438@up	-- / 0		Cuarto
2021/04/13	60.00 / 160	20	-- / 0		1930669@up	-- / 0		Quinto
2021/04/13	90.00 / 160	20	-- / 0		1830326@up	-- / 0		SÃ@ptimo
2021/04/13	70.00 / 160	21	-- / 0		1730355@up	-- / 0		Noveno
2021/04/13	110.00 / 160	21	-- / 0		1830510@UF	-- / 0		SÃ@ptimo
2021/04/13	80.00 / 160	21	-- / 0		1830178@up	-- / 0		SÃ@ptimo

Figura 6. Dataset inicial.

Después de realizar una limpieza al conjunto de datos, el dataset muestra un total de 22 columnas cómo: “Puntuacion_total”, “Edad”, “Matrícula”, “Cuatrimestre”, “Estado_civil”, “Sexo” y columnas cómo: 1, 2, 3...16, donde cada número representa un reactivo y 3 reactivos representan un nivel cognoscitivo de Marzano, R. J., (2001).

La columna “Puntuacion_total” comprende valores desde 0 hasta 160, dónde 160 representa el último nivel cognoscitivo, cada 3 reactivos hacen referencia a distintas características y las columnas restantes muestran un poco de información del estudiante.

En la Figura 7 se muestra que el estudiante con matrícula 1 solamente cuenta con 10 puntos en los primeros 3 reactivos; esto quiere decir, que cuenta con un 30 por ciento de las características que conforman el primer nivel; además, el estudiante cuenta con 40 puntos totales; esto quiere decir, que se podría identificar su aprendizaje al nivel de Recuperación de la taxonomía de Marzano, R. J., (2001).

Puntuacion_t	Edad	matricula	Cuatrimestre	Estado_civil	sexo	1	2	3
40	20	1	4	Soltero	M	0	0	10
100	20	2	5	Soltero	M	10	0	10
100	20	3	7	Soltero	M	10	0	10
80	19	4	5	Soltero	M	10	0	10
70	20	5	7	Soltero	M	10	0	10
70	19	6	5	Soltero	M	10	0	10
100	42	7	4	Soltero	M	10	0	10
60	20	8	5	Soltero	M	10	0	10
90	20	9	7	Soltero	F	10	10	10
70	21	10	9	Otro	M	10	0	10
110	21	11	7	Soltero	M	10	10	10
80	21	12	7	Soltero	M	10	0	10
130	20	13	7	Soltero	M	10	10	10
130	18	14	1	Soltero	M	10	10	10
100	18	15	1	Soltero	F	10	0	10
70	21	16	9	Otro	M	10	0	10

Figura 7. Dataset final.

Caracterización de aprendizaje en los estudiantes.

El proceso de caracterización se realizó en base a las características que representan los niveles cognitivos de la taxonomía de Marzano, R. J., (2001).

Agrupamiento.

Se realizó un agrupamiento al dataset final con el algoritmo K-Means, debido a que tanto Dang, Shilpa, (2015) como Ekubo, E. A. (2019) y gran parte en la literatura obtienen mejores resultados utilizándolo; en este caso, el algoritmo determinó 5 grupos que representan el nivel de los estudiantes. Las instancias totales del dataset son de 78; en la Figura 8 se observan las cantidades de instancias por cada grupo, donde se comprende que el grupo “cluster0” es igual a “Nivel de Recuperación” hasta llegar a “cluster4” que identifica al “Nivel de metacognición”.

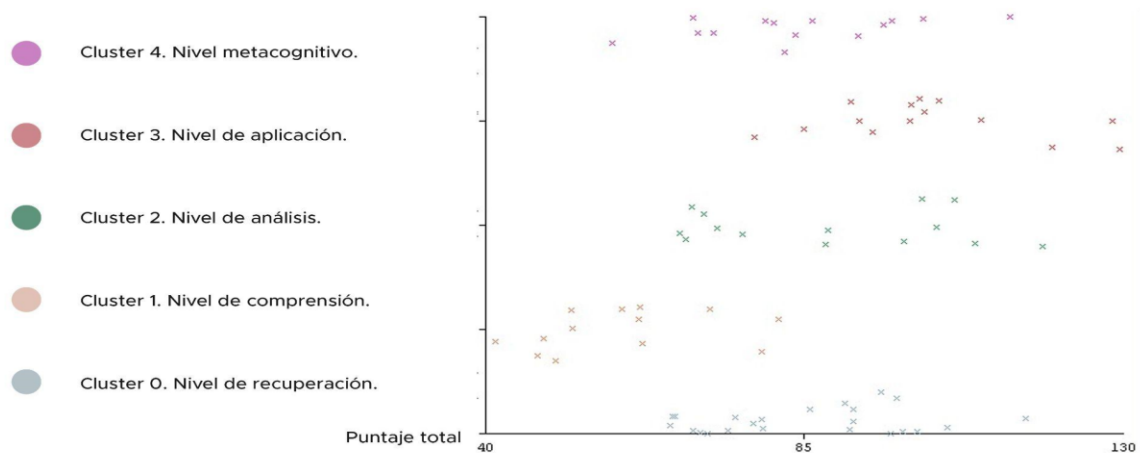
Final cluster centroids:

Attribute	Full Data (78.0)	Cluster#				
		0 (23.0)	1 (13.0)	2 (14.0)	3 (14.0)	4 (14.0)
Puntuacion_total	84.2308	85.6522	58.4615	88.5714	102.1429	83.5714
matricula	40.9103	38.6957	45.1538	32.4286	33.9286	56.0714
1	8.8462	10	5.3846	9.2857	9.2857	9.2857
2	0.8974	0	0	2.8571	2.1429	0
3	10	10	10	10	10	10
4	4.8718	0	3.8462	10	8.5714	5
5	9.359	9.5652	10	8.5714	10	8.5714
6	7.9487	9.5652	5.3846	7.8571	8.5714	7.1429
7	2.0513	1.3043	1.5385	0.7143	5	2.1429
8	7.1795	10	6.1538	5	3.5714	9.2857
9	5.8974	8.6957	0	9.2857	5.7143	3.5714
10	2.8205	1.7391	0	6.4286	4.2857	2.1429
11	3.0769	2.6087	2.3077	5.7143	3.5714	1.4286
12	9.4872	10	10	7.8571	10	9.2857
13	5.3846	6.5217	1.5385	2.1429	9.2857	6.4286
14	2.3077	0.4348	0.7692	0.7143	2.1429	8.5714
15	4.1026	5.2174	1.5385	2.1429	10	0.7143
16	0.2564	0.8696	0	0	0	0

Figura 8. Agrupamiento K-Means.

Tras el análisis de esta población, el “Nivel de Recuperación” cuenta con 23 instancias, siendo el grupo con mayor concentración, mientras que el “Nivel de Comprensión” cuenta con 19 instancias y es el grupo con menor concentración. La Figura 9 representa una gráfica del agrupamiento realizado, donde el eje X representa el puntaje total obtenido por los alumnos.

Figura 9. Gráfica de agrupamiento K-Means sobre niveles cognitivos de Marzano, R. J., (2001) con vista al puntaje total.



Nota. Es posible observar diferentes vistas de agrupamiento y no afecta en los resultados obtenidos.

Al seleccionar una instancia del “Cluster0” y una del “Cluster3” se observa que la instancia 39 es irregular y las ponderaciones se concentran en los primeros niveles, mientras que la instancia 25 tiene 5 ausencias de puntaje, por lo que K-Means ha situado en el “Cluster3”. Por lo tanto, se determinó que el alumno con instancia 39 requiere reforzar temas de programación que comprendan los siguientes 4 niveles, mientras que el alumno con instancia 25 solo necesita reforzar temas para el “Nivel de Análisis” y “Nivel de aplicación”. Es posible observar que existe una variación entre ambas instancias al encontrarse alejadas, y en la Figura 10 se observa una comparación de resultados entre ellas.

Instance: 25	Instance: 39
Instance_number : 24.0	Instance_number : 38.0
Puntuacion_total : 110.0	Puntuacion_total : 70.0
1 : 10.0	1 : 10.0
2 : 0.0	2 : 0.0
3 : 10.0	3 : 10.0
4 : 10.0	4 : 0.0
5 : 10.0	5 : 0.0
6 : 10.0	6 : 10.0
7 : 0.0	7 : 0.0
8 : 0.0	8 : 10.0
9 : 10.0	9 : 0.0
10 : 0.0	10 : 0.0
11 : 10.0	11 : 0.0
12 : 10.0	12 : 10.0
13 : 10.0	13 : 10.0
14 : 10.0	14 : 0.0
15 : 10.0	15 : 10.0
16 : 0.0	16 : 0.0
Cluster : cluster3	Cluster : cluster0

Figura 10. Comparativa entre dos instancias.

Tras un análisis entre ambas instancias, se pudo determinar lo siguiente:

- El alumno con instancia 39 se ha ubicado en el “Nivel de recuperación”.
- El alumno con instancia 25 se ha ubicado en el “Nivel de Aplicación”.

CONCLUSIONES.

Tras los resultados obtenidos, se determinó que el algoritmo K-Means ha realizado 5 agrupaciones, de las cuales el “cluster0” o “Nivel de Recuperación” representa la mayor concentración de población con un total de 25 instancias de un total de 78 registros; esto refleja que la mayor parte de una población determinada de estudiantes representaría un bajo rendimiento académico. La agrupación realizada en este trabajo de investigación tendrá relevancia para realizar una clasificación del aprendizaje en los estudiantes de ITI en la UPV.

Se ha considerado obtener más información mediante un sistema de tutoría inteligente que se encuentra en desarrollo; esta información extra representará el interés, el compromiso y la motivación de los estudiantes. La información recolectada se basará en factores importantes como el tiempo, el cual puede definir qué tan responsable es el estudiante, trabajos realizados, los cuales pueden determinar el interés por sus estudios, ausencias de entregas, las cuales pueden ser interpretadas como desmotivación o posible deserción.

REFERENCIAS BIBLIOGRÁFICAS.

1. Dang, Shilpa. (2015). Performance Evaluation of Clustering Algorithm Using Different Datasets. IJARCSMS. 3. 167-173.
2. Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.
3. Ekubo, E. A. (2019). Attributes of Low Performing Students In E-Learning System Using Clustering.

4. Ekubo EA (2019) Data Collection Experience on Educational Data Mining in Nigeria. Am J Compt Sci Inform Technol Vol.7 No.2: 37.
5. Ekubo, E. A., & Esiefarienrhe, M. B. (2019, December). Attributes of low performing students in e-learning system using clustering technique. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1324-1328). IEEE.
6. Ekubo, E. A. (2020). Predictive system for characterizing low performance of Undergraduate students using machine learning techniques (Doctoral dissertation, North-West University (South Africa)).
7. Marzano, R. J. (2001). Designing a new taxonomy of educational objectives. Experts in Assessment Series, Guskey, T. R., \& Marzano, R. J. (Eds.). Thousand Oaks, CA: Corwin.
8. Masashi Sugiyama. Chapter 1 - statistical machine learning. In Masashi Sugiyama, editor, Introduction to Statistical Machine Learning, pages 3 – 8. Morgan Kaufmann, Boston, 2016.
9. Murphy, P. (2008). Defining pedagogy. Pedagogy and practice: Culture and identities, 28-39.
10. Páramo, G., & Correa Maya, C. (1999). Deserción estudiantil universitaria. Conceptualización. Revista Universidad EAFIT, 35(114), 65-78.

DATOS DE LOS AUTORES.

1. **Jonathan Isael Duque Hernández.** Universidad Politécnica de Victoria. Estudiante de Maestría en Ingeniería del tercer cuatrimestre. México. E-mail: 1630206@upv.edu.mx
2. **Mario Humberto Rodríguez-Chávez.** Maestro en Sistemas de Información. Estudiante del Doctorado en Gestión e Innovación Educativa en la Unidad Académica Multidisciplinaria de Ciencias, Educación y Humanidades de la Universidad Autónoma de Tamaulipas. Universidad Politécnica de Victoria. Profesor investigador de tiempo completo. México. E-mail: mrodriguez@upv.edu.mx

3. **Said Polanco-Martagón.** Doctor en Ciencias Computacionales. Universidad Politécnica de Victoria. Profesor investigador de tiempo completo. México. E-mail: spolancom@upv.edu.mx

RECIBIDO: 15 de agosto del 2021.

APROBADO: 18 de septiembre del 2021.