**ESTADO DEL ARTE**

Actualmente con el crecimiento de las plataformas que brindan cursos en línea, crece la necesidad de aprovechar de mejor manera los recursos que estas tienen y mejorar los servicios que ofrecen, La minería de datos en la educación (EDM) como indica en [1] se indican de una manera más técnica que EDM nace como un paradigma orientado a al diseño de modelos, tareas, métodos o algoritmos para explorar datos del ámbito educativo y de [2] se sabe que EDM permite resolver incógnitas sobre como aprovecha el estudiante los elementos que le brinda la plataforma, que tanto influye estos elementos en el proceso de aprendizaje y que esta aprendiendo realmente el estudiante, para después de realizar una serie de análisis entregar información que sea de utilidad para basados en ellas se mejore la forma en la que la información es suministrada al estudiante, para que el docente tome decisiones sobre cual es la mejor manera de interaccionar con el estudiante o dar importancia a las variables que mas repercuten dentro del proceso de aprendizaje.

**Sistemas de aprendizaje e importancia de los datos.**

EDM cuenta con varias maneras de aportar a los dos sistemas de educación (tradicional y a distancia) debe saber a cuál de los dos se quiere realizar el análisis para con esto tener claridad de cuales métodos tienen una mayor efectividad en el análisis de los datos. encontramos en [3] unas definiciones y claras diferenciaciones de estos sistemas de educación.

En el ambiente tradicional que es el más ampliamente usado y necesita de un lugar físico para la reunión de los integrantes del grupo, se debe tener en cuenta que los datos que se almacenan para un posterior análisis, son menores y requiere de un mayor esfuerzo por parte de los tutores o coordinadores el almacenar estos. Lo que de cierta manera limita los alcances que pueda aportar un análisis con EDM.

Como para este caso se trabajará con ambientes a distancia este documento se enfocará más en estos.

Los ambientes a distancia son los que consisten unas herramientas o técnicas que permiten al estudiante acceder a la información suministrada por el tutor sin necesidad de que exista un contacto físico entre ellos, como estos están soportados en un servidor web la información que pueden ofrecer es mas amplia y con mayores facilidades para su captura, documentos como los

* Log file: contiene la información de las interacciones entre agentes del sistema de educación, estas interacciones pueden ser realizadas por el servidor, la cual contendrá información del tiempo, respuestas y entradas, por el cliente el cual será el estudiante y dará información de su interacción con la plataforma y por último por el proxi la cual tendrá información de la comunicación entre cliente y servidor.
* Contenido: información dentro de la plataforma que fue dispuesta al estudiante.
* Estructura: información que describa la organización que tiene el contenido.
* Uso: datos que contienen los patrones de uso que se le da a la información que existe en la plataforma.
* Perfil del usuario: información personal del estudiante.

Al igual que una buena fuente de información y comprensión de los datos que esta puede ofrecer, es importante saber el tratamiento que se le debe dar a la información, dado que por lo general la información vine dada de forma o en formatos que no son adecuados para aplicar las herramientas que brindan en EDM, para esto se recomienda realizar un procesamiento previo a la información para adecuarla en [4] indican una variedad de formatos posibles o herramientas para tratar información, este caso se centrará en la adecuación de los datos para el uso de lenguajes de programación como Python que permite realizar análisis estadístico de datos científicos o R que es una herramienta que permite el análisis estadístico de datos y aporta análisis de redes sociales dada la cantidad de paquetes que tiene incluidos para el análisis de social networks adapting (SNA).

Por la facilidad que ofrece para el ordenamiento de datos y comprensión a la lectura de estos por su interfaz Microsoft Excel una opción bastante tentativa ya que también permite el análisis de datos científicos. La manera en que esta herramienta representa los datos permite identificar problemas estructurales y semánticos en las variables ya representa estas de una manera clara además que es fácil de usar en Python o R.

**Técnicas**

para el estudio de las técnicas que brinda la EDMse considera correcto ir ilustrando con casos de investigación realizadas con anterioridad.

En el estudio de las massive online open courses MOOC se pueden realizar una gran cantidad de estudios en el caso del [5] donde se busca concluir sobre el efecto que tiene la confusión de los estudiantes en el proceso de aprendizaje, en el desarrollo del problema y su método de análisis son identificables técnicas como el reconocimiento de patrones y la clasificación de información, esta última técnica es también utilizada en [6] donde se busca caracterizar los perfiles de aprendizaje de los estudiantes y para ellos utilizan una metodología con una serie de reglas que permiten emular el cuestionario VARK.

En [7] donde se estudia las interacciones sociales en un aula de case por medios de análisis cualitativo y SNA, donde esta ultima describe patrones de relaciones interpersonales, lo que en si es una técnica de reconocimiento de patrones, pero también se utilizan técnicas de agrupamiento. La técnica de agrupamiento esta presente en una gran cantidad de soluciones implementadas para solucionar problemas, en [8] donde se pretende la detección temprana de estudiantes en riesgo, en este estudio también se utiliza una técnica de predicción. Otro estudio donde se utiliza la técnica de agrupamiento es [9] que intenta proponer agrupamientos y dar un perfil a estudiantes de acuerdo a las interacciones que tengan con un tutor inteligente. Como ultimo ejemplo para esta técnica se tiene [10] donde se pretende utilizar métodos supervisados y no supervisados para la creación de modelos de usuario para la exploración de ambientes de aprendizaje, en este se usan no solo agrupamiento si no también clasificación como técnicas para la solución del problema.

Como se puede notar las técnicas son variadas dentro de EDM, observando que se puede clasificar, agrupar, predecir, reconocer patrones, asociación y otras que no han sido utilizadas en los anteriores estudios, pero que [11] da a conocer brindando ejemplo de métodos o técnicas que se pueden agrupar en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Modelos | Objetivos | Técnicas | Métodos y Tareas |
| Predicción | Pretende inferir características dentro de una base de datos a partir del diseño de un modelo que se lo permita. | Regresión, clasificación,  Agrupación, categorización | Social network análisis (SNA), logistic regression, multiple  Regression, decision tree, Pearson correlation, mean squared error, modelo oculto de markov |
| Descriptivo | Dar a conocer los aspectos más significativos dentro de un grupo de datos | Regresión,  Agrupación,  Patrones secuenciales, reglas de asociación, análisis de correlación, análisis numérico | decision tree, interactions network, cluster análisis  Association rules, performance sequence  Analysis, Bayes theorem, |

[1] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.

[2] Vargas Bonilla, J. F., Contreras Olivares, L., López Hincapié, J. D., & Montoya Lince, A. (2015). Ecucational data mining (EDM) para la determinación de comportamientos en estudiantes de ingeniería en la modalidad virtual de UDE.

[3] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert saystems with applications*, *33*(1), 135-146.

[4] Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, *42*(1), 85-106.

[5] Yang, D., Kraut, R., & Rose, C. P. (2016). Exploring the effect of student confusion in massive open online courses. *Journal of Educational Data Mining*, *8*(1), 52-83.

[6] Jhonny Alberto Lopera Hoyos, & Julian Esteban Giraldo Hurtado. caracterización de perfiles de aprendizaje de estudiantes mediante técnicas de minería de datos en la plataforma UDE@.

[7] Martınez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & De La Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, *41*(4), 353-368.

[8] Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk-predicting student dropouts using administrative student data from German universities and machine learning methods. *JEDM| Journal of Educational Data Mining*, *11*(3), 1-41.

[9] Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *JEDM| Journal of Educational Data Mining*, *5*(1), 104-146.

[10] Amershi, S., & Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory. *JEDM| Journal of Educational Data Mining*, *1*(1), 18-71.

[11] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.