

Listas de contenidos disponibles en [ScienceDirect](#)

Informática y Educación: Inteligencia Artificial

página de inicio de la revista: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence



Aprendizaje de árboles de decisión a través de un modelo predictivo para estudiantes académicos Rendimiento en entornos inteligentes de aprendizaje móvil

Vasiliki Matzavela^{*}, Efthimios Alepis

Departamento de Informática, Universidad de Piraeus Piraeus, Grecia

INFORMACIÓN DEL ARTÍCULO

Palabras clave:

Minería de datos educativos
Aprendizaje del árbol de decisión
Educación adaptativa
Clasificación
modelo predictivo
Evaluación
Entornos de decisiones inteligentes

ABSTRACTO

En el área del aprendizaje automático y la ciencia de datos, el aprendizaje del árbol de decisiones se considera una de las técnicas de clasificación más populares. Por lo tanto, un algoritmo de árbol de decisión genera un modelo predictivo y de clasificación, que es simple de entender e interpretar, fácil de mostrar gráficamente y capaz de manejar datos numéricos y categóricos. Los sistemas inteligentes de m-learning, disfrutaron recientemente de un crecimiento explosivo de interés, para una educación más efectiva y un aprendizaje adaptativo a la medida de las habilidades de aprendizaje de cada estudiante. El objetivo de este trabajo es mejorar aún más la personalización en el rendimiento académico de los estudiantes, que incluye pruebas dinámicas con un modelo predictivo. Uno de los principales objetivos de esta investigación es crear pruebas dinámicas adaptativas para evaluar el rendimiento académico de los estudiantes, mientras se comparan constantemente los resultados de la evaluación que exhibe el perfil individual del estudiante, con los resultados del algoritmo del árbol de decisión que formula un modelo predictivo para el conocimiento de los estudiantes. nivel, de acuerdo con los pesos del árbol de decisión.

1. Introducción

Durante la última década y especialmente en los últimos años, durante la pandemia del Covid-19, los sistemas de aprendizaje adaptativo evolucionaron con un crecimiento explosivo en la educación semipresencial y en el aprendizaje digital. Este fenómeno indica la creciente demanda de sistemas inteligentes de aprendizaje m diseñados dinámicamente para el nivel cognitivo de los estudiantes, basados en características personalizadas. Existe la necesidad de métodos innovadores emergentes para extraer con precisión información educativa que se centre en las necesidades educativas individuales del estudiante (Virvou y Alepis, 2013).

Teniendo en cuenta los desafíos y problemas que enfrenta la comunidad educativa moderna, se emplearon sistemas inteligentes y algoritmos que mejoran la educación y los niveles de enseñanza en las instituciones educativas. El aprendizaje en línea y semipresencial se ha convertido en una estrategia educativa común en la educación terciaria y, por lo tanto, los instructores deben reconceptualizar cuestiones fundamentales de la enseñanza, el aprendizaje y la evaluación en espacios no tradicionales (Virvou, Troussas, Caro, Espinosa).

Estos temas incluyen conceptos tales como la validez y confiabilidad de la evaluación en entornos en línea en relación con el cumplimiento de los propósitos previstos, así como la comprensión de cómo funciona la evaluación adaptativa dentro del aprendizaje en línea y combinado (Kabassi & Alepis, 2020). El examen en línea es un componente cada vez más importante de los cursos en línea.

En la mayoría de los escenarios de examen en línea, la supervisión presencial está ausente. Los resultados señalan que un pequeño aumento en las tasas de aprobación podría afectar significativamente el éxito general, es decir, la disminución de las tasas de deserción (Lakshmi et al., 2013), (Matzavela & Alepis, 2021).

La Minería de Datos Educativos (EDM) es una aplicación de Técnicas de Minería de Conocimiento a partir de datos educativos, y su objeto es analizar datos, con el fin de resolver problemas de investigación en el campo de la Educación. El EDM examina cómo guiar a los alumnos en el aprendizaje. Sus datos provienen de diferentes fuentes, como bases de datos de sistemas educativos, sistemas de Internet, registro de aprendices, etc. El objetivo es mejorar el proceso de aprendizaje y actualizar los sistemas de apoyo al aprendizaje.

Aunque los investigadores (ver por ejemplo (Virvou & Alepis, 2005)) están estudiando el proceso de aprendizaje automático durante décadas, EDM difiere en que utiliza resultados experimentales no de situaciones de aprendizaje teórico sino de hechos reales. Las investigaciones en el área de EDM y Learning Analytics a partir de datos educativos, tienen como objetivo brindar soluciones a problemas relacionados con los procesos educativos.

Los investigadores de EDM responden a preguntas como: qué secuencia educativa es más efectiva para el estudiante, qué acciones del estudiante resultan en satisfacción y progreso en el aprendizaje, cuáles son las características que conducen al mejor resultado de un proceso de aprendizaje, (Calders & Pechenizkiy, 2012). Por otro lado, los investigadores de Learning Analytics responden a preguntas como: ¿Qué nota es probable que se obtenga?

^{*} Autor correspondiente.

Direcciones de correo electrónico: vasiliki.mat@yahoo.com (V. Matzavela), talepis@unipi.gr (E. Alepis).

<https://doi.org/10.1016/j.caeai.2021.100035> Recibido

el 24 de agosto de 2021; Aceptado el 3 de octubre de 2021

Disponible en línea el 5 de octubre de 2021

2666-920X/© 2021 Los autores.

Publicado por Elsevier Ltd.

Este es un artículo de acceso abierto bajo la licencia CC BY-NC-ND

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

por un estudiante sin supervisión, si los estudiantes no pueden seguir el flujo de un curso, o si los estudiantes corren el riesgo de no completar un curso.

El aprendizaje del árbol de decisiones es un método utilizado en la minería de datos (Topirceanu & Grosseck, 2017). El objetivo es crear un modelo que prediga el valor de una variable de destino, en función de varias variables de entrada. En minería de datos, los árboles de decisión también se pueden describir como la combinación de y técnicas computacionales para ayudar a la descripción, categorización y generalización de un conjunto de datos dado (Quinlan, 1986).

El estudio de (Song & Ying, 2015) aplicó la técnica de decisión inteligente, con el fin de elegir la mejor predicción y análisis. El algoritmo DT-Quest genera no solo las decisiones generales, sino también las decisiones específicas en casos excepcionales relevantes. Se estudia la eficacia de DT-Quest habiendo realizado un proceso experimental sobre 213 alumnos. Los experimentos muestran que el algoritmo propuesto es más efectivo que los enfoques tradicionales de aprendizaje automático para predecir el nivel cognitivo de los estudiantes.

La estructura de Decision Tree Learning, que se formula en el aprendizaje automático, se puede implementar en modelos predictivos educativos (Krouska et al., 2020). Los pesos de los árboles de decisión surgieron de parámetros que consistían en la divulgación de una plétora de resultados, con poda o sin ella. El Algoritmo DT-Quest fue desarrollado con atributos de clasificación, mientras que fue fundamentado con su integración con datos específicos. Entre los casos de desempeño de los estudiantes, se identificaron modalidades importantes para los niveles de conocimiento de los estudiantes y los perfiles educativos de los estudiantes.

Libre de ambigüedades, se ha concretado un modelo educativo predictivo del rendimiento académico de los estudiantes de educación terciaria en línea, al tiempo que se ha integrado una evaluación, (Matzavela et al., 2017), con variables que involucran y consideran el perfil de un estudiante. Los resultados de este estudio siguieron el camino de Decision Tree Learning con DT-Quest Algorithm, presentando comparaciones entre varios factores de identificación, como el rendimiento y grado del estudiante, género, educación de los padres, ingresos de los padres, si el estudiante es el primer hijo y si él/ella está trabajando.

En particular, se define un árbol de decisión con pesos, para crear pruebas/ejercicios adaptativos que identifiquen los niveles de conocimiento de los estudiantes y para distinguir dinámicamente las necesidades de cada estudiante.

El desafío de un modelo que prediga el rendimiento académico de los estudiantes es cada vez mayor debido al enorme aumento de datos en la base de datos, mientras que el análisis del rendimiento académico de los estudiantes constituye un factor importante para la eficacia del aprendizaje adaptativo. Los factores identificadores anteriores determinan las variables del perfil de los estudiantes, que a su vez inciden en su rendimiento académico. Después de evaluar las calificaciones obtenidas por los estudiantes en un año académico, las conclusiones ayudarán a los instructores a mejorar el aprendizaje adaptativo en la educación terciaria. Se evaluó la efectividad de nuestro sistema en términos de interacciones de e-learning adaptativo, utilizando los resultados obtenidos a través del estudio de las relaciones del rendimiento académico del estudiante con las variables de su perfil, como se mencionó anteriormente.

Este documento está organizado de la siguiente manera: la segunda sección contiene trabajos relacionados con el aprendizaje adaptativo a través de la evaluación móvil en el rendimiento académico de los estudiantes. Además, se citan encuestas en los campos de EDM y Learning Analytics. La sección 3 gira en torno a las cuatro etapas del método presentado: recopilación de datos, clasificación, creación de un modelo predictivo y evaluación. Las secciones 4 y 5 se refieren a la minería de datos y el aprendizaje del árbol de decisiones. La sexta sección trata de un proceso experimental, presenta la recolección de datos, describe el algoritmo DT-Quest, estudia algunos casos de uso y finalmente presenta los resultados relevantes. En las secciones 7 y 8 se presenta una discusión y las conclusiones, respectivamente. También está disponible un apéndice que presenta el árbol de decisión relacionado con el algoritmo utilizado en nuestro enfoque.

2. Trabajo relacionado

Se ha planteado la cuestión del uso excesivo de la tecnología durante la enseñanza. Ese problema fue abordado por Anshari et al., (Anshari et al., 2017). Realizaron una investigación, investigando el hecho de que el uso de dispositivos portátiles durante el aprendizaje puede causar una distracción a los estudiantes.

Dado que la implementación de la evaluación móvil extenderá el aprendizaje móvil, existe la posibilidad de que aumente el problema de la interferencia. Por lo tanto, se deben realizar los estudios correspondientes.

Centrándonos en el camino de la m-evaluación en el rendimiento académico de los estudiantes, se observó que existe una falta notable de estudios en ese campo. Muy pocos investigadores, hasta el día de hoy, se han ocupado del m-assessment, que es la evolución del m-learning. Por otro lado, la mayor parte de las encuestas que se han publicado en los últimos años hace referencia a la evaluación del proceso de m-learning al que se refiere. De hecho, la evaluación se lleva a cabo utilizando métodos tradicionales, como cuestionarios no dinámicos o pruebas educativas tradicionales (Troussas et al., 2020). También se han estudiado cuestionarios dinámicos relacionados con redes neuronales artificiales (Matzavela & Alepis, 2017, agosto).

Las encuestas en los campos de EDM y Learning Analytics responden a preguntas muy complejas sobre lo que un estudiante es consciente y el alcance de su participación. Los investigadores han experimentado con nuevas técnicas de modelado y con una variedad de tipos de datos de los nuevos sistemas educativos, que prometen en gran medida la predicción de los resultados de aprendizaje de un estudiante. Las áreas de aplicación más amplias de EDM y Learning Analytics se encuentran en Online Learning. Estas áreas se distinguen en el modelado de la siguiente manera: Comportamiento, conocimiento, experiencia de usuario, perfil de usuario, situaciones sectoriales y análisis de tendencias (Johnson & Adams, 2011, pp.

1–22). Uno de los métodos más extendidos de EDM y Data Mining en general para la producción de modelos predictivos, es el método de clasificación. El concepto básico de Clasificación en Educación es el examen de un tema de interés en el contexto de un curso o clase, a partir de las características específicas que lo rigen (Bienkowski et al., 2012). La categorización se considera una actividad clave de muchos Sistemas de Apoyo al Aprendizaje de los procedimientos educativos, ya sea directa o indirectamente. En la enseñanza en línea, la muestra de población suele ser grande y, en consecuencia, se aumentan los requisitos y la implementación de un modelo predictivo (clasificador) permite al docente formarse una mejor imagen de su clase en línea, (Siemens, d Baker). Los parámetros para extraer un modelo predictivo son los siguientes: La selección de un método de clasificación, un conjunto de datos cuyos valores se conocen y la correcta división del conjunto de datos en dos subconjuntos, el conjunto de entrenamiento y el conjunto de prueba. Se aplica un algoritmo de aprendizaje al conjunto de entrenamiento para dibujar un modelo predictivo. Luego, se prueba la capacidad predictiva del modelo resultante, aplicándolo al conjunto de prueba cuyos valores están ocultos (Desmarais, Lemieux).

Los investigadores, (Thomas & Galambos, 2004), se han ocupado de la evaluación de los conocimientos de los estudiantes a través de cuestionarios, mientras que un árbol de decisión en el fondo muestra que hay caminos-opciones disponibles para el mejor resultado. Un cuestionario está completamente respaldado por un árbol de decisión, con una excelente claridad y visualización de los datos, lo que nos ayuda a sacar conclusiones y hacer las correcciones necesarias, (Guan et al., 2019). La principal limitación de los árboles de decisión es la suposición de que todos los puntos se pueden categorizar; esto da como resultado que todas las posibles contradicciones se interpreten como errores. Un ejemplo de ello es el caso en el que todos los datos representan el desempeño de los estudiantes en una lección donde un porcentaje de instantáneas puede revelar contradicciones. A menudo se propone como solución el enfoque probabilístico de los valores característicos. Pero en este caso, el sistema se vuelve inestable, ya que cada nodo tiene su propia posibilidad de aparecer. Los árboles de decisión también son susceptibles de sobreajuste, especialmente cuando el conjunto de datos es relativamente pequeño (Korte & Vyten, 2008).

Papamitsiou y Economides, (Papamitsiou & Economides, 2014), utilizaron TAM en su estudio, en un esfuerzo por investigar si la actitud de los estudiantes influye en la adopción de la evaluación móvil. Prepararon un cuestionario de encuesta que fue respondido por los estudiantes, y los resultados llevaron a la conclusión de que la competencia, la autonomía y la relación

Hay 3 factores significativos que deben tenerse en cuenta al desarrollar el procedimiento. Nikou y Economides (Nikou & Economides, 2017) presentan un examen más completo de la evaluación de la evaluación móvil en otro artículo. Propusieron un modelo especializado basado en TAM, llamado Mobile-based Assessment Acceptance Model (MBAAM). Cuando se utiliza este modelo, se tienen en cuenta más factores, como la facilidad de uso, la utilidad y la intención de comportamiento, lo que lleva a una mayor comprensión. El resultado es una mejor experiencia para los alumnos, lo que favorece el aprendizaje. Aunque Cuku ħsi'c et al., (Cuku ħsi'c et al., 2014), no emplean dispositivos móviles, proponen una evaluación basada en ordenadores que se compara la evaluación moderna con los métodos tradicionales, lo que sugiere que existen efectos positivos en el desempeño de los estudiantes. Un análisis más detallado, con la inclusión de la evaluación móvil en la comparación entre la evaluación tradicional y la basada en computadora, muestra que tanto las computadoras como los dispositivos móviles tienen efectos positivos en la motivación de los alumnos y que podrían reemplazar las formas de evaluación anticuadas (Nikou & Economides, 2016).

Sobre la base de los estudios anteriores en m-assessment, hemos creado un algoritmo que examina el nivel cognitivo del alumno y, según el nivel, las preguntas clasificadas que se le pedirá que responda. Las preguntas, clasificadas en cuatro niveles de dificultad diferentes, se recuperan de una base de datos. EDM produce extracción de datos a partir de datos reales, los cuales son parametrizados y así se extraen los resultados correspondientes. A continuación se responderán las siguientes preguntas: cuál es la ruta más corta, cómo obtener el mejor resultado, qué examinar a través de la personalización. En el fondo, habrá un árbol de decisiones, que tiene la ventaja de la simplicidad y la versatilidad. También presenta datos fácilmente comprensibles, clasificación rápida de instantáneas y una gran dinámica de representación. Al definir los parámetros, comenzamos a construir el árbol de decisión con sus pesos. Nuestro objetivo es encontrar la ruta (ponderada) de longitud mínima para cada individuo. Queremos construir un Árbol de Determinación con la mínima profundidad, (asegurando al mismo tiempo que el nivel cognitivo del estudiante se comprueba satisfactoriamente), y luego construir las reglas de categorización.

3. Metodología

El método sugerido en este documento para mejorar la predicción del rendimiento académico de los estudiantes emplea una combinación de minería de datos y aprendizaje de árboles de decisión. Hay cuatro etapas principales en este método: recopilación de datos, clasificación, creación de un modelo predictivo y evaluación (ver, por ejemplo, (Bienkowski et al., 2012)).

En la recopilación de datos, las variables de evaluación se determinan a través de preguntas específicas para los estudiantes, los tipos de variables y la descripción de las variables. El análisis porcentual de los datos se logra centrándose en el perfil del alumno. Además, se concretaron los pesos del árbol de decisión para permitirnos definir dinámicamente el nivel de conocimiento del estudiante.

En la clasificación se integraron las respuestas individuales de 213 estudiantes de educación terciaria, mientras que las variables de la evaluación consisten en las características de los estudiantes (género, grado, educación de los padres, ingresos de los padres, primer hijo o no, el estudiante trabaja o no) .

En la creación del modelo predictivo se 'conformó' el Algoritmo DT-Quest con los parámetros del proceso experimental, que empleó los pesos del árbol de decisión para la elección dinámica de los ejercicios, con el fin de evaluar el rendimiento académico del estudiante. Los árboles de decisión tienen prioridad frente a los modelos predictivos alternativos, ya que son simples de entender e interpretar, fáciles de mostrar gráficamente y capaces de manejar datos tanto numéricos como categóricos.

En la evaluación, los resultados de la evaluación se compararon con las calificaciones y puntajes totales del desempeño del estudiante, y esto llevó a conclusiones significativas. La comparación entre la evaluación y el algoritmo del árbol de decisiones es una tarea desafiante pero complicada, con resultados alentadores. Las predicciones precisas del rendimiento podrían conducir a mejores resultados de aprendizaje y un mayor logro de objetivos en el aprendizaje adaptativo. Estas predicciones se centran en la mejora de los estudiantes

nivel cognitivo acorde al perfil del alumno, así como a una evaluación más personalizada.

El objetivo principal de la metodología propuesta es construir el modelo de clasificación y predicción del rendimiento académico de los estudiantes, que clasifique el perfil de un estudiante a través de la evaluación y prediga el nivel de conocimiento del estudiante utilizando el algoritmo del árbol de decisión.

4. Minería de datos

La minería de datos es el procedimiento de descubrir prototipos en conjuntos de datos grandes y complejos (Sarker). Hay dos aspectos en la minería de datos: la construcción de modelos y la detección de prototipos. La construcción de modelos en la minería de datos es muy similar al modelado estadístico, aunque surgen nuevos problemas debido al gran tamaño de los conjuntos de datos y al hecho de que la minería de datos es a menudo un análisis secundario de datos (Baker, 2010). Vivimos en un mundo donde se recopilan diariamente grandes cantidades de datos. Analizar tales datos es una necesidad importante, y dado que la necesidad es la madre de la invención (Platón), la minería de datos puede verse como el resultado de la evolución natural de la tecnología.

Un motor de búsqueda (por ejemplo, Google) recibe cientos de millones de consultas todos los días. Cada consulta puede verse como una transacción, donde el usuario describe la información que necesita (Kotsiantis, 2012). La industria de gestión de datos y bases de datos mejoró drásticamente el desarrollo de varias funcionalidades críticas. Hoy en día, numerosos sistemas de bases de datos ofrecen procesamiento de consultas y transacciones como práctica común. Desde la década de 1960, la tecnología de la información y las bases de datos ha evolucionado sistemáticamente desde sistemas primitivos de procesamiento de archivos hasta sistemas de bases de datos sofisticados y potentes. El mundo es rico en datos pero pobre en información. La minería de datos es la búsqueda de conocimiento en los datos y de prototipos interesantes (Sarker, 2018).

Un sistema de base de datos, también llamado sistema de administración de base de datos (DBMS), consta de una colección de datos interrelacionados, conocida como base de datos, y un conjunto de programas de software para administrar y acceder a los datos. Una base de datos relacional es una colección de tablas a cada una de las cuales se le asigna un nombre único. Cada tabla consta de un conjunto de atributos (columnas o campos) y normalmente almacena un gran conjunto de tuplas (filas o registros). Se puede acceder a los datos mediante consultas a la base de datos escritas en un lenguaje de consulta relacional (SQL) o con la ayuda de interfaces gráficas de usuario (Parsazadeh et al., 2018). Hay una serie de funcionalidades de minería de datos, que se utilizan para especificar los tipos de prototipos que se encontrarán en las tareas de minería de datos. En general, tales tareas se pueden clasificar en dos categorías, descriptivas y predictivas. Las tareas de minería descriptiva realizan inducción sobre los datos actuales para hacer predicciones. La clasificación es el proceso de encontrar un modelo o función que describa y distinga clases de datos o conceptos. El modelo derivado se basa en el análisis de un conjunto de datos de entrenamiento. El modelo se utiliza para predecir la etiqueta de clase de los objetos para los que aún se desconoce la clase. El modelo derivado se puede representar de diversas formas, como reglas de clasificación, árboles de decisión, fórmulas matemáticas o redes neuronales (Rizvi et al., 2019). Un árbol de decisión es una estructura de árbol similar a un diagrama de flujo, donde el peso de cada nodo denota una prueba en un valor de atributo, cada rama representa un resultado de la prueba y las hojas del árbol representan clases o distribuciones de clase (Skrbinjek & Dermol, 2019) . A diferencia de la clasificación y la regresión, que analizan conjuntos de datos de entrenamiento y etiquetados por clase, la agrupación en clústeres analiza objetos de datos sin consultar las etiquetas de clase. La agrupación en clústeres se puede utilizar para generar etiquetas de clase para un grupo de datos. Cada grupo así formado puede verse como una clase de objetos, de los cuales se pueden derivar reglas. Un conjunto de datos puede contener objetos que no cumplen con el comportamiento general o modelo de los datos (Fayyad et al., 1996). Estos objetos de datos son valores atípicos. Muchos métodos de minería de datos descartan los valores atípicos como ruido o excepciones. Los valores atípicos se pueden detectar mediante pruebas estadísticas que asumen un modelo de distribución o probabilidad para los datos, o mediante el uso de medidas de distancia, donde los objetos que están alejados del "corazón

La minería de datos adopta técnicas de muchos dominios. El aprendizaje automático investiga cómo las computadoras pueden aprender en función de los datos. Un problema típico de aprendizaje automático es programar una computadora para que pueda

reconocer automáticamente los códigos postales escritos a mano en el correo, después de aprender de un conjunto de ejemplos. Los problemas clásicos del aprendizaje automático que están relacionados con la minería de datos son los siguientes: a) Aprendizaje supervisado, (que es sinónimo de clasificación). b) Aprendizaje no supervisado, (que es sinónimo de agrupamiento). c) Aprendizaje semisupervisado, que es una clase de técnicas de aprendizaje automático que utilizan ejemplos etiquetados y no etiquetados al desarrollar un modelo. d) Aprendizaje activo, que es un enfoque que permite a los usuarios desempeñar un papel activo en el proceso de aprendizaje (Han et al., 2011).

5. Aprendizaje del árbol de decisión

Decision Tree Learning es una herramienta general de modelado predictivo que tiene aplicaciones que abarcan varias áreas diferentes (Qin & Lawry, 2005). En general, los árboles de decisión se construyen a través de un enfoque algorítmico que identifica formas de dividir un conjunto de datos, en función de varias condiciones (Hamsa et al., 2016). Es uno de los métodos más utilizados y prácticos para el aprendizaje supervisado. Decision Tree Learning es un método de aprendizaje supervisado no paramétrico, utilizado tanto para tareas de clasificación como de regresión. Las reglas de decisión son generalmente de la forma declaraciones 'if-then-else'. Cuanto más profundo es el árbol, más complejas son las reglas y más ajustado es el modelo (Baldwin, Xie).

Un árbol de decisión es un gráfico similar a un árbol con nodos que representan el lugar donde elegimos un atributo y hacemos una pregunta, los bordes representan las respuestas a la pregunta y las hojas representan la etiqueta de clase real. Se utilizan en la toma de decisiones no lineales con forma de decisión lineal simple.

Los árboles de decisión clasifican los ejemplos clasificándolos en el árbol desde la raíz hasta cada hoja, y las hojas proporcionan la clasificación a los ejemplos correspondientes. Cada nodo del árbol actúa como un caso de prueba para algún atributo, y cada arista que desciende de ese nodo corresponde a una de las posibles respuestas al caso de prueba. Los árboles de decisión se pueden convertir fácilmente en reglas de clasificación.

Los árboles de decisión utilizados en la minería de datos (Ogunde & Ajibade, 2014) son de dos tipos principales: a) Árboles de clasificación, que se utilizan cuando el resultado previsto es la clase discreta a la que pertenece. b) Árboles de regresión, que se utilizan cuando el resultado previsto es un número real. El término análisis de árbol de clasificación y regresión (CART) es un término general que se usa para referirse a los dos procedimientos anteriores, introducido por primera vez por Breiman et al., en 1984. Los algoritmos de árboles de decisión notables incluyen: a) ID3, un algoritmo inventado por Ross Quintan, que se utiliza para generar un árbol de decisión a partir de un conjunto de datos. b) C4.5, que puede utilizarse para la clasificación. c) CARRO. d) CHAID, que es una técnica de árbol de decisión, basada en pruebas de significancia ajustada, y fue publicada en 1980 por Gordon V.

Kass. e) MARS, que es una forma de análisis de regresión introducida por Jerome H. Friedman en 1991.

Comparado con otros métodos de minería de datos, el método del árbol de decisión tiene varias ventajas: a) Es simple de entender e interpretar. b) Es fácil de mostrar gráficamente. c) Es capaz de manejar datos tanto numéricos como categóricos. d) Requiere poca preparación de datos. e) Funciona bien con grandes conjuntos de datos.

Los árboles pueden ser muy escasos. Además, un pequeño cambio en los datos de entrenamiento puede resultar en un gran cambio en el árbol y, en consecuencia, en las predicciones finales (Sarker et al., 2020).

El árbol de decisión generado por C4.5 se puede utilizar para la clasificación. En cada nodo del árbol, C4.5 elige el atributo de los datos que divide de manera más efectiva su conjunto de muestras en subconjuntos, incrustados en una de las clases discretas. El criterio de división es la ganancia de información normalizada. El atributo con la ganancia de información normalizada más alta se elige para tomar la decisión. El algoritmo C4.5 luego se repite en las sublistas más pequeñas. El árbol de decisiones es eficiente y, por lo tanto, es adecuado para conjuntos de datos grandes o pequeños. Es el método exploratorio más exitoso para descubrir estructuras de datos desviadas. Los árboles dividen recursivamente el espacio de datos de entrada para identificar segmentos donde los registros son homogéneos (Wu et al., 2008).

Los métodos para generar árboles de decisión a partir de datos, como C4.5, permiten

para una representación en forma de árbol de los resultados del aprendizaje (Lin & Fan, 2019). Se proponen enfoques de minería de datos para predecir el rendimiento de los estudiantes.

6. Proceso experimental

6.1. Recopilación de datos/ descripción del conjunto de datos

El propósito principal de usar técnicas de selección de datos es minimizar la redundancia y maximizar el subconjunto de datos relevantes, mientras se mantiene una alta precisión sin perder información importante. Utilizamos datos de un curso en una universidad en el que la asistencia a clases es obligatoria.

El perfil del estudiante se identificó luego de ajustar por variables de control que incluían género, grado, educación de los padres, ingresos de los padres, ser el primer hijo o no, trabajar o no (Guarín et al., 2015). El conjunto de datos incluye 213 estudiantes que tomaron el curso en la primavera de 2020. Todos estos estudiantes se habían registrado en el curso y habían declarado su participación real. La técnica de clasificación se seleccionó en función de su reputación en la literatura de minería de datos publicada y su superioridad en problemas de tipo predicción (Hand & Adams, 2014, pp. 1–7). El sistema toma en cuenta varias características de los estudiantes que son importantes para el logro de los objetivos de aprendizaje de los estudiantes. Estas características han sido reportadas en la literatura como aspectos significativos que influyen en el proceso educativo, y son las siguientes:

1. Género: El género de los estudiantes es una característica personal que puede afectar el proceso educativo. Los estudiantes varones tienen en cuenta principalmente los puntos de aprendizaje de cada curso, mientras que sus preferencias de aprendizaje pueden no ser claras, lo que podría ser diametralmente opuesto a las estudiantes mujeres.
2. Grado: Esta variable del nivel cognitivo de un individuo también es muy importante para el logro de la meta de aprendizaje. El grado de rendimiento académico podría verse afectado por cuestiones externas específicas (como la ansiedad, el cansancio, la ira). Usando un modelo predictivo a través de un algoritmo de árbol de decisión, un sistema de e-learning puede tener una ilustración más clara de su nivel de conocimiento.
3. Educación de los padres: La educación de los padres es un parámetro significativo que afecta las metas de aprendizaje de la mayoría de los estudiantes. El nivel educativo de los padres interactúa con la retroalimentación de conocimientos de los estudiantes e influye en el ritmo de aprendizaje de cada estudiante.
4. Ingresos de los padres: La adquisición de ingresos sólidos de los padres se considera un factor importante para la individualización de las necesidades de conocimiento de los estudiantes. Los ingresos de los padres podrían causar una divergencia sorprendentemente grande entre los estudiantes. Los sistemas de aprendizaje electrónico incorporaron métodos que atraen a todos los estudiantes independientemente de los ingresos de los padres. Algunos estudiantes pueden abandonar el aprendizaje debido a los bajos ingresos de los padres. Las oportunidades de aprendizaje que se brindan a los estudiantes en los sistemas de educación combinada y de aprendizaje electrónico son independientes de los ingresos de los padres.
5. Primer hijo: Una variable incorporada en el cuestionario que se entregó a los estudiantes, se refiere a si el estudiante es el primer hijo de su familia. Esta variable arrojó resultados interesantes, ya que influyó positivamente en el rendimiento de los estudiantes y negativamente en su tendencia a la deserción. Por ello, se ha incluido como una variable muy importante que puede contribuir positivamente a la mejora de la calidad del proceso educativo.
6. Estudiante que trabaja: Después de la investigación y el análisis porcentual, se encontró que los estudiantes que trabajan tienen dificultades para continuar sus estudios y completarlos. Además, su desempeño se ve afectado negativamente por puntajes más bajos, lo que genera dificultades para los instructores. Esta variable completa el conjunto de variables que constituyen la evaluación de los alumnos. Que el alumno trabaje o no puede afectar su rendimiento académico, por lo que se observa atentamente la variable y se anota su evolución.

En la siguiente tabla se enumeran las variables que influyeron en los resultados

(ver [Tabla 1](#)). Estas variables fueron empleadas en la mayoría de las investigaciones (ver por ejemplo ([Cuku ýsíc et al., 2014](#)),([Kotsiantis, 2012](#))) en educación adaptativa.

Los resultados de la evaluación anterior identificaron el perfil de los estudiantes y denominaron la calidad del curso ([Virvou, Kabassi, Alepis, Kameas, Pierrakeas, Theodosiou](#)). Las variables se relacionaron con los requerimientos del proceso experimental. Después de completar el semestre, los estudiantes completaron la evaluación con las variables: género, grado, educación de los padres, ingresos de los padres, ser el primer hijo, trabajar, lo que llevó a la formación de la [Tabla 1](#).

El género indica una pequeña precedencia (57% frente a 43%) para las mujeres. Por otro lado, las alumnas obtuvieron titulación superior en un 87% frente al 72% de los alumnos varones. La mayoría de los estudiantes tenían padres bien educados (83%). El ingreso de los padres se registró como bajo en 12%, como medio en 69% y como alto en 19%. Los estudiantes eran primeros hijos en un 52%, mientras que un 48% no lo eran. Finalmente, el 33% de los estudiantes trabajaba y el 67% restante no (ver [Tabla 2](#)).

6.2. Algoritmo DT-Quest

En este sistema educativo adaptativo se creó un algoritmo de árbol de decisión, con el objetivo de definir dinámicamente el nivel de conocimiento de los estudiantes, siendo simple de entender e interpretar, así como fácil de visualizar gráficamente. Los puntos del examen se usaron como un indicador del rendimiento del curso de cada estudiante. La calificación está en una escala de 0 a 100, y se requieren 50 puntos para aprobar. El objetivo es mejorar aún más la personalización dinámica en el rendimiento académico de los estudiantes. El algoritmo que se ha creado para el propósito de este estudio se ha denominado Algoritmo DT-Quest y se representa en el siguiente diagrama de flujo. En un principio, se establecen estos parámetros ([Tabla 3](#)):

La justificación del algoritmo es la siguiente:

6.2.1. Inicializando las variables

La variable "i" se utiliza como contador, para mostrar las primeras 4 preguntas, una de cada nivel de dificultad.
"Calificación" representa la suma de los puntos obtenidos por cada respuesta correcta hasta el momento.
"MaxGrade" es la calificación máxima que el alumno podría haber obtenido hasta el momento, y se utiliza para decidir el nivel de dificultad de la siguiente pregunta, así como para controlar cuándo terminará el algoritmo.
Por supuesto, el nivel de dificultad depende de otros factores, que se indicarán más adelante.

6.2.2. El primer ciclo

ýt muestra las primeras 4 preguntas, una de cada nivel de dificultad, en orden creciente.

Si la respuesta es correcta, la variable "Nota" se incrementa en el peso de la pregunta correspondiente.
Por cada pregunta de nivel i que se plantea, se incrementa en i*5 la "NotaMáxima".

La "i" se incrementa en 1, y si es menor o igual a 4, se muestra la siguiente pregunta.

Tabla 1
Variables de valoración.

Variable	Tipo de variable	Descripción
Género	Nominal	Hombre, Mujer
Calificación	Númérico	0–100
Educación para padres	Nominal	Bien educado, No bien educado
Ingresos de los padresb	Nominal	Bajo medio alto
Ser el primer hijo	Nominal	Verdadero Falso
Laboral	Nominal	Verdadero Falso

^a Bien educados: título de educación secundaria o terciaria, No bien educados: título de educación primaria.
^B Renta Alta: > 2000€ al mes, Renta Media: Entre 1000€ y 2000€ al mes, Renta Baja: ý 1000€ al mes.

Tabla 2
Análisis porcentual de las variables.

Género	Masculino:	Mujer:	
	43%	57%	
Calificación	Masculino 0-45: 50-100: 28% 72%	Mujer 0-45: 50-100: 13% 87%	
educación de los padres	Bien educados: 83%	No bien educados: 17%	
Ingresos de los padres	Bajo: 12%	Medio: 69%	Alto: 19%
Ser el primer hijo	Cierto: 52%	Falso: 48%	
Laboral	Cierto: 33%	Falso: 67%	

Tabla 3
Parámetros iniciales.

Niveles de dificultad	yo: {1,2,3,4}
Grado máximo	100
peso de una pregunta	i*5

6.2.3. Segundo ciclo La

subrutina "QuestionLevel" se llama después de las primeras 4 preguntas y selecciona el nivel de la siguiente pregunta. Procesa parámetros tales

como:

El número de respuestas correctas; En caso de respuestas incorrectas, el nivel más alto entre estas preguntas; A lo largo de la prueba se pueden plantear como máximo 3 preguntas de cada nivel de dificultad; El peso de la siguiente pregunta debe ser tal que "MaxGrade" no supere 100; Si las preguntas de todos los niveles de dificultad ya se han respondido con éxito, se pronostican las siguientes preguntas permitidas, asegurando al mismo tiempo que se planteará la menor cantidad posible de preguntas en los pasos restantes de la prueba.

La variable "k" almacena el nivel de la siguiente pregunta que devolvió "Ques tionLevel".

Se muestra una pregunta de nivel k.
Si la respuesta es correcta, la variable "Grado" se incrementa en k*5 y en cualquier caso, "MaxGrade" se incrementa en k*5.
Si "MaxGrade" es inferior a 100, el algoritmo ejecuta el segundo ciclo una vez más.

6.2.4. Terminar

Cuando "MaxGrade" es exactamente igual a 100, se muestra el "Grade" final y el algoritmo termina.
En el Apéndice se presenta una parte del árbol T con raíz binaria creado por el algoritmo anterior, que cubre los casos estudiados en la sección [6.3](#) a continuación.

6.3. casos de uso

Analizaremos algunos casos que son útiles para sacar conclusiones.
A través de diferentes caminos y parámetros que hemos identificado, reconocemos los diferentes resultados. En el rendimiento académico de los alumnos, podemos identificar mediante los ejercicios adecuados al alumno bien preparado o al alumno que necesita estudiar más. Se seleccionaron casos con conductas extremas y alumnos con distintas desviaciones en su desempeño, con el fin de obtener resultados interesantes con el correspondiente interés de estudio. Al investigar estos casos específicos, hemos analizado los puntos más importantes del árbol de decisión. Los alumnos no conocen y no pueden ver todos los ejercicios de rendimiento académico, sino sólo el próximo ejercicio que planteará el algoritmo. Dependiendo del nivel de conocimiento individual, el modelo predictivo del Algoritmo DT-Quest muestra un ejercicio adecuado, mientras que, dependiendo de la respuesta dada, el siguiente

Se proporcionará ejercicio, no el mismo para todos los estudiantes. Las preguntas se alternan dinámicamente con el propósito de obtener un resultado óptimo del desempeño de cada alumno. El algoritmo del árbol de decisiones decide sobre el número y el nivel de dificultad de los ejercicios que se incluyen en el cuestionario académico personalizado creado. Los siguientes 4 casos destacan 4 modelos de estudiantes diferentes. En el primer caso, se hace referencia al alumno excelente que recibirá preguntas de acuerdo a su nivel. En el cuarto caso, observamos a un estudiante que ignora una amplia gama del plan de estudios, por lo que la personalización se adapta adecuadamente. Los casos 2 y 3 se refieren a estudiantes que han realizado una preparación moderada y han obtenido resultados correspondientemente moderados.

6.3.1. Caso 1

Al principio, el candidato debe responder una pregunta de cada nivel (es decir, 4 preguntas con una suma de 50 puntos). Si todas las preguntas se responden correctamente, significa que es muy probable que el candidato esté bien preparado; la siguiente pregunta debe ser de máxima dificultad (es decir, del 4º nivel). Si esa pregunta también se responde correctamente, la próxima pregunta será nuevamente del 4.º nivel, lo que significa que el peso total de las preguntas ahora es de 90 puntos (50 + 20 + 20). El último tema debe ser extraído inevitablemente del 2º nivel (es decir, con un peso de 10 puntos), para que "MaxGrade" alcance exactamente los 100 puntos con la menor cantidad de preguntas posible. Si el alumno también responde correctamente a esta pregunta, obtendrá una calificación final de 100/100. La calificación final de este candidato se calcula después de solo 7 preguntas, ya que es obvio que está muy bien preparado. Este caso se muestra por el camino rojo del árbol T (ver Figs. 1 y 2 del Apéndice).

6.3.2. Caso 2

En el caso 2, examinamos a un estudiante que no responde las preguntas de 1º y 3º nivel, mientras que responde correctamente las preguntas de 2º y 4º nivel. Después de las primeras 4 materias de prueba, ha obtenido 30 puntos de 50. Las preguntas del 1er y 3er nivel ahora tienen prioridad sobre los otros niveles y dado que el candidato debe responder la menor cantidad de preguntas posible, la siguiente pregunta debe ser desde el más alto de esos dos, ese es el 3er nivel. Si el candidato responde a la pregunta del 3er nivel, el siguiente tema se extraerá del 1er nivel, ya que aún no ha respondido a ese nivel. (Si el candidato no vuelve a contestar una pregunta de 3er nivel, se le dará una 3ra y última oportunidad para hacerlo). Si el candidato también responde correctamente la pregunta de primer nivel, habrá obtenido 50 puntos de 70. Esto significa que todavía se necesitan 30 puntos para que "MaxGrade" alcance 100 y termine el examen. Teniendo en cuenta que se deben plantear el menor número de preguntas posible, estos 30 puntos restantes se separarán en 20 y 10. Por tanto, el candidato deberá responder a una pregunta de 4º y 2º nivel respectivamente. Si ahora, por ejemplo, el alumno no responde correctamente la pregunta de 4º nivel mientras que responde correctamente la pregunta de 2º nivel, obtendrá una nota final de 60/100. Este caso se muestra por el camino verde del árbol T (ver Figs. 1 y 4 del Apéndice).

6.3.3. Caso 3

En otro caso, después de las primeras 4 preguntas, el estudiante no ha respondido solo la pregunta de 2do nivel. Entonces, la siguiente pregunta debe ser de ese nivel y si no responde, se seleccionará una tercera pregunta de ese nivel. Si tampoco responde a esta pregunta, la nota del alumno será 40 sobre 70 y 'MaxGrade' necesita 30 más para llegar a 100, como en el ejemplo anterior. Sin embargo, esta vez no se puede llegar a 30 con una pregunta de 20 y 10, ya que el estudiante ya ha respondido tres preguntas de 2º nivel, que tienen un peso de 10. Por lo tanto, 30 se obtendrá con 20, 5 y 5, que es de 4º nivel y se plantearán dos preguntas de 1º nivel. Si ahora, por ejemplo, el alumno no responde la pregunta de 4º nivel mientras que responde correctamente las dos preguntas de primer nivel obtendrá una nota final de 50/100. Este caso se muestra por el camino azul del árbol T (ver Figs. 1 y 3 del Apéndice).

6.3.4. Caso 4

Finalmente, trataremos el caso de un candidato que no resolvió ninguna de las primeras cuatro preguntas. Esto significa que el candidato no estaba lo suficientemente bien preparado para el examen; con base en el requisito de que debe enfrentar la menor cantidad de preguntas posible, se le presentará una pregunta del más alto nivel. Si el candidato no responde correctamente, se le planteará otra pregunta de ese nivel. Si ahora, por ejemplo, el alumno no responde también a esta pregunta, la pregunta final vendrá del 2º nivel, ya que el peso de las 6 preguntas anteriores es 90 y le faltan 10 más para llegar a 100. Entonces, se planteará una pregunta de segundo nivel. Si el alumno tampoco responde a esta pregunta, obtendrá una nota final de 0/100. Este caso se muestra por el camino naranja del árbol T (ver Figs. 1 y 5 del Apéndice).

7. Resultados

Considerando las notas de rendimiento académico de los estudiantes, las siguientes 4 tablas (Tablas 4, 5, 6 y 7) muestran los resultados teniendo en cuenta las variables de evaluación.

La Tabla 4 presenta los hallazgos del caso 1, es decir para estudiantes que obtuvieron calificaciones excelentes (entre 80 y 100 puntos). El 51% de ellos eran mujeres y el 49% eran hombres. Casi todos los padres de los estudiantes tenían buena educación, llegando al 94%, mientras que solo el 6% no lo tenía. En cuanto al ingreso de sus padres, para el 56% de ellos el ingreso fue alto, para el 40% fue medio, mientras que el ingreso fue bajo solo para el 4%. Finalmente, el 60% de los estudiantes eran los primeros hijos de sus familias, y solo el 27% de los estudiantes trabajaban.

En el caso 2, las notas de los alumnos oscilaron entre 51 y 79 puntos. Las mujeres eran el 58%, mientras que el 42% eran hombres. Una gran proporción de los estudiantes tenían padres con buena educación (88%) y solo el 12% de los padres de los estudiantes no tenían buena educación. Se observa que los menores ingresos de los padres se relacionan con la disminución de las calificaciones, siendo el ingreso para el 37% de ellos alto, el 46% medio y el 17% bajo. Exactamente la mitad de los alumnos eran primogénitos, y el 38% de los alumnos trabajaba (ver Tabla 5).

En el caso 3, los alumnos obtuvieron una calificación entre 31 y 50 puntos. El 52% de ellos eran hombres, mientras que el 48% eran mujeres. El porcentaje de estudiantes que tenían padres con buena educación fue del 28%, mientras que el 72% de los padres de los estudiantes no tenían buena educación. El ingreso de los padres fue alto para el 13% de los estudiantes, medio para el 39% y bajo para el 48%. Finalmente, el 31% de los alumnos son primeros hijos, y el porcentaje de alumnos que trabajaban alcanzaba el 51% (ver Tabla 6).

El caso 4 está formado por alumnos cuyas notas oscilaron entre 0 y 30 puntos. Las mujeres eran el 33%, mientras que el 67% eran hombres. La mayoría de ellos (89%) fueron criados por padres sin buena educación, mientras que solo el 11% tenían padres con buena educación. Solo el 7% de los padres de los estudiantes tenían ingresos altos, el 18% medianos y el 75% bajos. La mitad de los estudiantes eran primeros hijos de sus familias, mientras que el 78% de los estudiantes trabajaban (ver Tabla 7).

Tabla 4
Caso 1 (Grados 80–100).

	Género Educación de los padres	Padre ingreso	Primer hijo	Laboral
Masculino	49%			
Mujer	51%			
Bien educado		94%		
no bien educado		6%		
Alto		56%		
Medio		40%		
Bajo		4%		
Cierto			60%	
Falso			40%	
sí				27%
No				73%

Tabla 5

Caso 2 (grados 51 a 79).

	Género Educación de los padres	Padre ingreso	Primer hijo	Laboral
Masculino	42%			
Mujer	58%			
Bien educado	88%			
no bien educado	12%			
Alto		37%		
Medio		46%		
Bajo		17%		
Cierto			50%	
Falso			50%	
si				38%
No				62%

Tabla 6

Caso 3 (Grados 31–50).

	Género Padre educación	Padre ingreso	Primer hijo	Laboral
Masculino	52%			
Mujer	48%			
Bien educado	28%			
no bien educado	72%			
Alto		13%		
Medio		39%		
Bajo		48%		
Cierto			31%	
Falso			69%	
si				51%
No				49%

Tabla 7

Caso 4 (Grados 0-30).

	Género Padre educación	Padre ingreso	Primer niño	Laboral
Masculino	67%			
Mujer	33%			
Bien educado	11%			
Mal educado	89%			
Alto		7%		
Medio		18%		
Bajo		75%		
Cierto			50%	
Falso			50%	
si				78%
No				22%

8. Discusión

Las conclusiones de este estudio son muy prometedoras y aportan otro punto de vista para la evaluación del rendimiento de los alumnos. Utilizando datos de evaluación de los estudiantes para el curso, es útil para predecir los factores que afectan su rendimiento (Chrysafiadi et al., 2020). Además, esto ilustra un punto de vista original (en comparación con otros trabajos que tratan sobre el aprendizaje adaptativo, (Nikou & Economides, 2017), (Topirceanu & Grosseck, 2017)), mejorando la calidad educativa, que es vital para atraer estudiantes.

Al estudiar los resultados de nuestro trabajo, nos dimos cuenta de lo siguiente: En el caso 1, donde los estudiantes tenían calificaciones excelentes (variando de 80 a 100 puntos), el porcentaje de estudiantes mujeres era ligeramente superior al porcentaje de estudiantes varones. La gran mayoría de ellos habían sido criados por padres con altos ingresos (más de 2000 euros al mes). El 60% de ellos eran primeros hijos de sus familias, mientras que la mayoría de los estudiantes en este caso no trabajaba (73%).

El caso 2 se refiere a estudiantes con calificaciones que oscilan entre 51 y 79 puntos, siendo el 58% de ellos estudiantes mujeres. Nuevamente, la mayoría de los estudiantes (88%) tenían padres bien educados. Los padres del 46% de los alumnos percibían ingresos medios (entre 1000 y 2000 euros al mes). La mitad de los alumnos eran primeros hijos y el 38% de los alumnos en este caso estaban trabajando. El rendimiento académico disminuyó y esto se relacionó con los ingresos de los padres, mientras que no hubo correlación entre las calificaciones y el hecho de que el estudiante fuera el primer hijo o estuviera trabajando.

En el caso 3 (de 31 a 50 puntos), los alumnos varones tienen una pequeña precedencia sobre las alumnas (52% y 48% respectivamente). El porcentaje de padres con buena educación es del 28% y la mayoría relativa de los padres de los estudiantes (48%) tenían bajos ingresos. Ser el primer hijo aplicó al 31%, mientras que el 51% de los estudiantes tenía un trabajo. Las calificaciones estaban fuertemente influenciadas por la educación de los padres, los ingresos de los padres y por el hecho de que el estudiante estaba trabajando. Por el contrario, no hubo correlación entre las calificaciones y el hecho de que el estudiante fuera el primer hijo o no.

En el 4º caso, donde los estudiantes tenían calificaciones menores o iguales a 30 puntos, el 67% de ellos eran hombres. Para el 89% de los padres de los alumnos la educación era básica, mientras que el 75% de los padres de los alumnos tenían ingresos bajos (menos de 1000 euros al mes). La mitad de los alumnos eran primeros hijos y el 78% trabajaba. Las notas, en este caso, son las más bajas, y esto estaba relacionado con los ingresos y la educación de los padres, así como con el hecho de que el estudiante estaba trabajando. No se observó correlación entre las calificaciones y el hecho de que el estudiante fuera el primer hijo o no.

9. Conclusiones

Dado que el aprendizaje en línea y semipresencial se ha convertido en una estrategia educativa común en la educación terciaria, los profesores deben reconceptualizar cuestiones fundamentales de la enseñanza, el aprendizaje y la evaluación en entornos no tradicionales (Chrysafiadi y Virvou, 2014).

Este documento presentó un sistema que se adhiere a los principios generales de diseño e involucra elementos relacionados con el e-learning adaptativo. Por último, cabe destacar que las técnicas de personalización de nuestro sistema (Matzavela & Alepis, 2021) impresionaron a los alumnos. Esto era bastante esperado, ya que nuestro sistema emplea clasificación y un modelo predictivo para el aprendizaje individualizado y la mejora del rendimiento académico de los estudiantes. Los resultados del experimento fueron muy prometedores, provocando un alto nivel de aceptación de nuestro sistema por parte de los estudiantes.

De hecho, nuestro sistema proporciona una forma individualizada de aprendizaje adaptativo, mientras que el modelo predictivo propuesto es una forma novedosa de crear pruebas dinámicas y efectivas del nivel de conocimiento de los estudiantes. Aparte de eso, la individualización durante el proceso de aprendizaje impresionó a los estudiantes. Este artículo presenta una clasificación de las características de los estudiantes y un modelo predictivo utilizando el Algoritmo DT-Quest para la mejora del rendimiento académico de los estudiantes en entornos inteligentes de m-learning. Las características de los estudiantes empleadas por esta evaluación son género, grado, educación de los padres, ingresos de los padres, si el estudiante es el primer hijo, si el estudiante está trabajando. La evaluación es más efectiva cuando se adapta a las habilidades de aprendizaje de los estudiantes. El sistema crea un entorno de m-learning inteligente al ofrecer individualización; fue completamente evaluado por los estudiantes y los resultados mostraron una alta tasa de aceptación, manteniendo un alto nivel de rendimiento pedagógico. La comparación de los resultados obtenidos de las variables de evaluación y luego del modelo de predicción que surgió del algoritmo que creamos, mostró que existe una correlación entre el rendimiento de los estudiantes y sus características específicas. El estudio de esta correlación contribuye a la mejora de la evaluación del desempeño de los estudiantes en los sistemas educativos inteligentes en línea.

Los pasos futuros de la investigación pueden incluir el enriquecimiento del conocimiento del dominio con otros conceptos, a través de modelos predictivos originales y nuevas características dinámicas que ofrecen entornos efectivos de aprendizaje móvil en la educación terciaria. La aplicación que utiliza la evaluación móvil efectiva respaldada por el árbol de decisión, ya está en desarrollo y será

presentado en un trabajo futuro. Este enfoque podría abrir nuevos horizontes en los métodos para examinar el rendimiento académico en entornos inteligentes de e-learning y m-learning. La correlación entre la minería de datos y el aprendizaje del árbol de decisión, para la clasificación y predicción de resultados respectivamente, influye en el aumento de la efectividad de la educación adaptativa, de acuerdo con las necesidades individualizadas de cada estudiante y

nivel de conocimiento

Declaración de intereses en competencia

No hay conflicto de intereses con ninguno de los revisores sugeridos.

APÉNDICE

El árbol con raíz binaria T

Una parte del árbol T con raíz binaria creado por el algoritmo anterior, que cubre los casos de la prueba que se han estudiado anteriormente en los Casos 1, 2, 3 y 4, se presenta en las siguientes Figs. 1–5; las etiquetas de este árbol se determinan de la siguiente manera: La raíz está etiquetada



Cada vértice interno de T, en el nivel 1 l ÿ 2, tiene una etiqueta de la forma



, donde.

- mÿ{1,2,3,4} muestra el nivel de dificultad de la pregunta correspondiente, • nÿ{a,b,c} indica si esta pregunta es la primera, segunda o tercera respectivamente del nivel de dificultad m planteada al estudiante, y donde
- p es la "Calificación" del alumno después de su (1-1)^a respuesta, y • q es la "Calificación Máxima" correspondiente.

Finalmente, la etiqueta de cada hoja de T muestra el puntaje total del estudiante cuyo desempeño ha dictado dinámicamente al algoritmo para seguir este camino particular (es decir, esta secuencia particular de niveles de dificultad) desde la raíz hasta esta hoja.

En la Fig. 1 se muestra el subárbol T' del árbol T, con todos los puntos de nivel hasta el 4. Debido al tamaño muy grande del árbol T, sus subárboles T_i, i = 1,2,3,4, que corresponden a los cuatro casos desarrollados anteriormente (que son los subárboles con raíz en la hoja 1, 3, 6 y 8 de Tÿ respectivamente), se muestran en las Fig. 2 a 5.

La estrella que se presenta en cada camino desde la raíz hasta una hoja de T, indica el punto donde el algoritmo decide el nivel de dificultad y el número de las preguntas restantes.

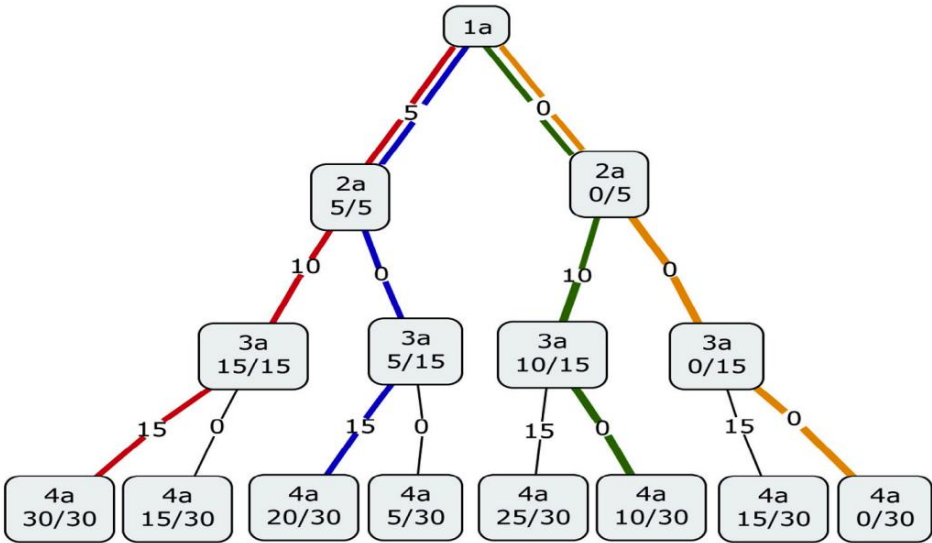


Fig. 1. El subárbol Tÿ del árbol T.

¹ El nivel l(v) de un vértice v de un árbol enraizado con raíz r, se define recursivamente como sigue: l(r) = 1; si u es el (único) vecino de v que está más cerca de r que de v, entonces l(v) = l(u)+1.

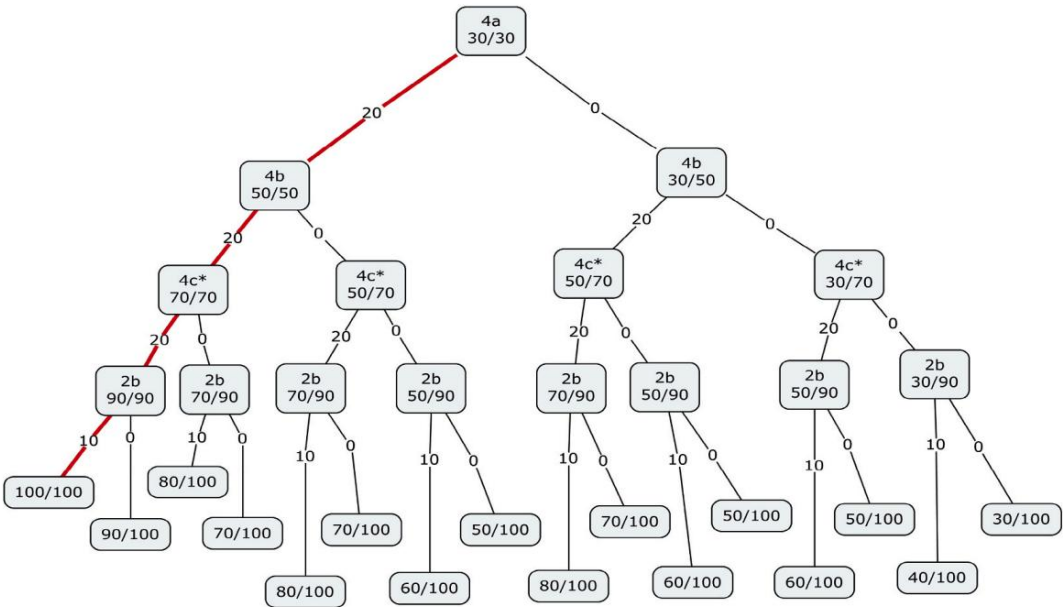


Fig. 2. El subárbol T1 del árbol T.

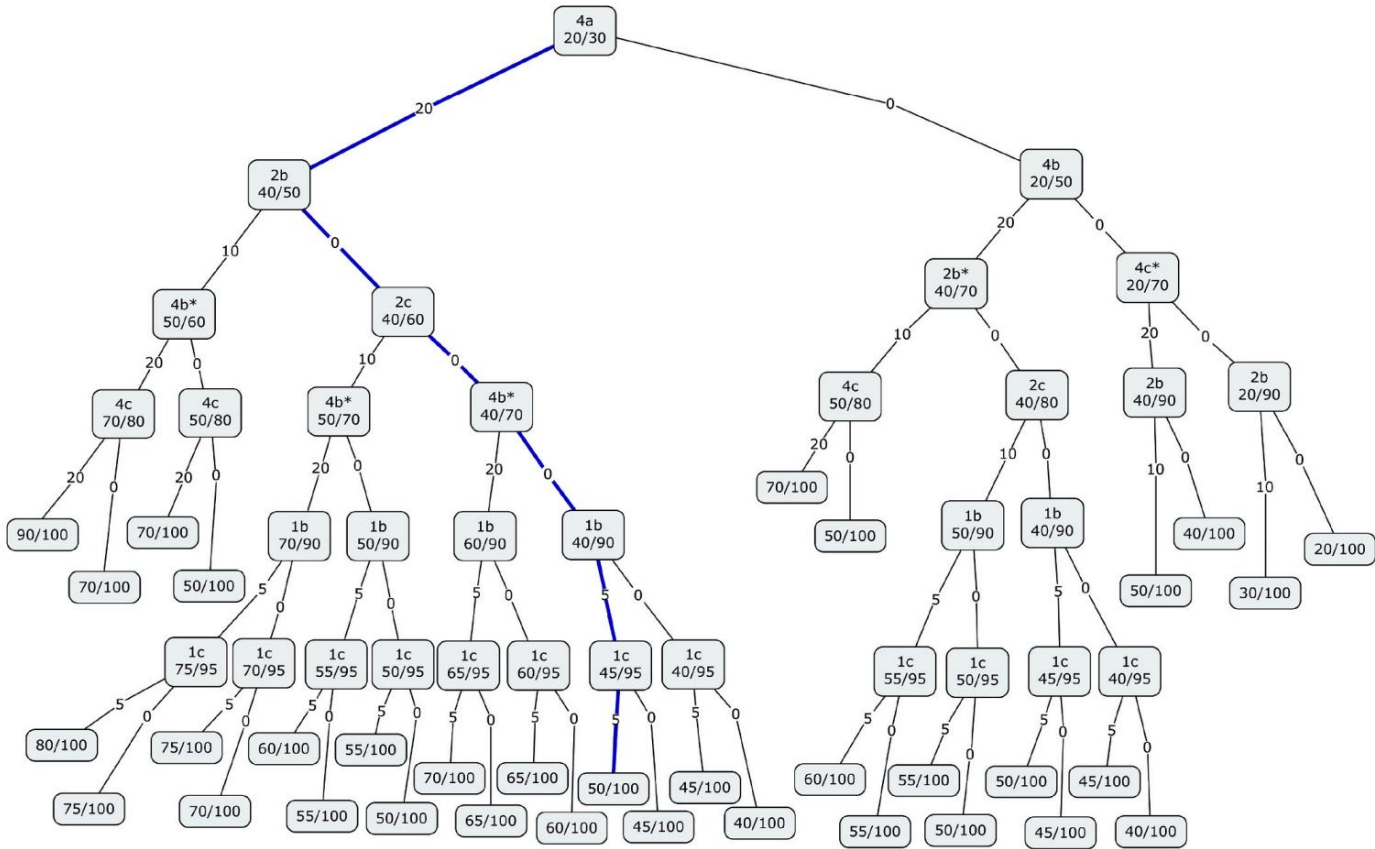


Fig. 3. El subárbol T2 del árbol T.

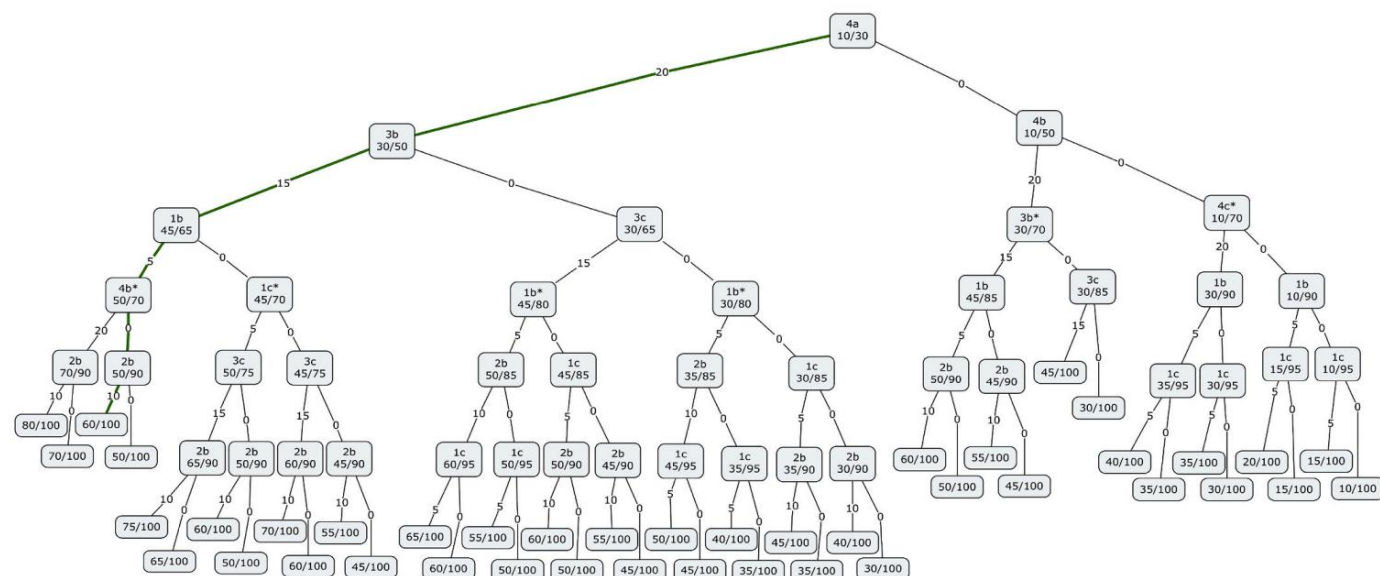


Fig. 4. El subárbol T3 del árbol T.

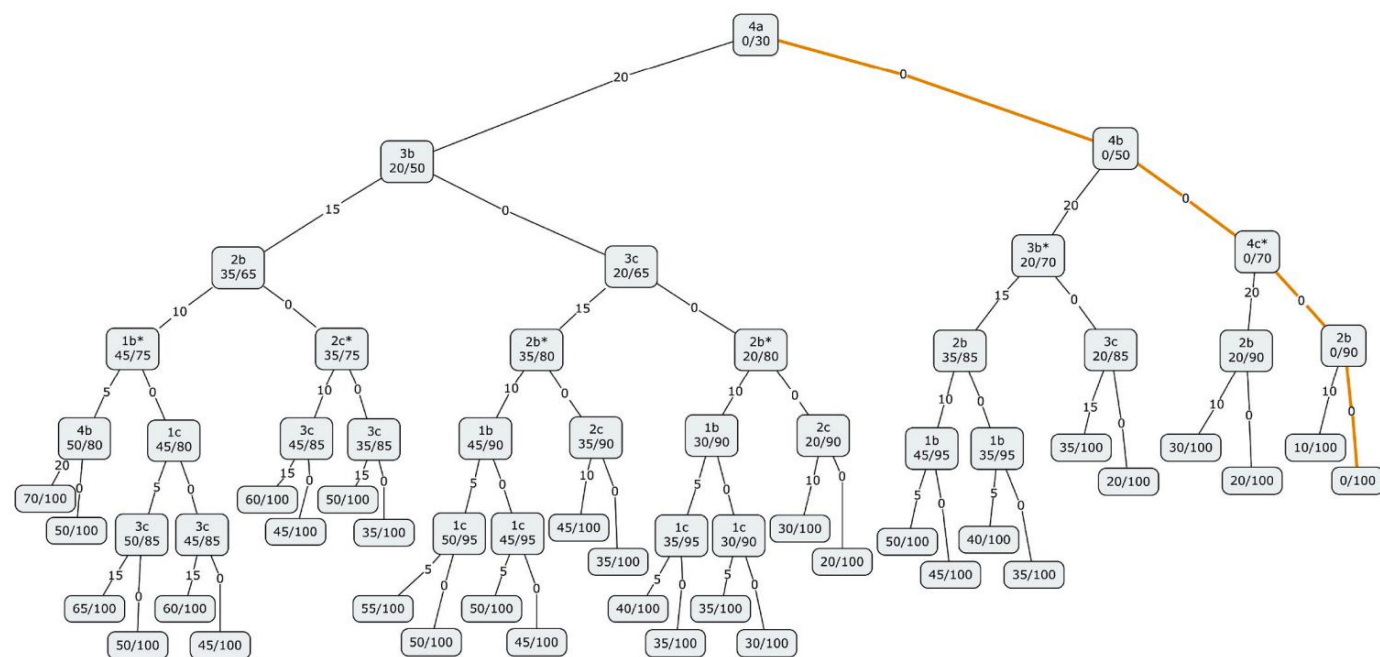


Fig. 5. El subárbol T4 del árbol T.

Fondos

Este estudio no fue financiado. El Dr. Alepis es el fundador de la empresa griega "Software Engineering Innovation Group – SEIG".

Aprobación ética

Este papel cumple totalmente con las normas éticas.

Consentimiento para participar

No aplica.

Consentimiento para publicación

No aplica.

Disponibilidad de datos y material

Todos los datos se indicaron en el documento.

Disponibilidad de código

El algoritmo está disponible, pero no el código personalizado.

Contribuciones de los autores

Este estudio fue realizado por Matzavela Vasiliki y Alepis Efthimios.

Referencias

Anshari, M., Almunawar, MN, Shahrill, M., Wicaksono, DK y Huda, M. (2017). Uso de teléfonos inteligentes en las aulas: ¿Ayuda para el aprendizaje o interferencia? *Educación y tecnologías de la información*, 22(6), 3063–3079.

Baker, RSJD (2010). Minería de datos para la educación. *Enciclopedia internacional de educación*, 7(3), 112–118.

Baldwin, JF y Xie, DW (2004). Reglas de lógica difusa simples basadas en árboles de decisión difusos para problemas de clasificación y predicción. En *Conferencia internacional sobre procesamiento inteligente de la información* (págs. 175–184). Boston, MA: Springer.

Blenkowski, M., Feng, M., Medios, B. y Minería, TED (2012). Octubre). Mejorar la enseñanza y el aprendizaje a través de la minería de datos educativos y el análisis del aprendizaje: resumen de un problema. En *Actas de la conferencia sobre tecnología avanzada para la educación* (págs. 1–57).

Calders, T. y Pechenizkiy, M. (2012). Introducción a la sección especial sobre minería de datos educativos. *Boletín de exploraciones de Acm Sigkdd*, 13(2), 3–6.

Chrysafiadi, K. y Virvou, M. (2014). Lógica difusa para la instrucción adaptativa en un entorno de aprendizaje electrónico para la programación de computadoras. *IEEE Transactions on Fuzzy Systems*, 23(1), 164–177.

Chrysafiadi, K., Virvou, M. y Sakkopoulos, E. (2020). Optimización de la programación aprendizaje de idiomas a través del modelado de estudiantes en un entorno educativo adaptativo basado en la web. En *Paradigmas de aprendizaje automático* (págs. 205–223). Cham: Springer.

Cukū jsi'c, M., Garajica, Z., & Jadri'c, M. (2014). Autoevaluación en línea y el éxito de los estudiantes en las instituciones de educación superior. *Informática y Educación*, 72, 100–109.

Desmarais, M. y Lemieux, F. (2013). Agrupación y visualización de secuencias de estado de estudio. En *Actas de la sexta conferencia internacional sobre minería de datos educativos* (págs. 224–227). Memphis: Sociedad Internacional de Minería de Datos Educativos, 2013.

Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). Desde la minería de datos hasta el descubrimiento de conocimiento en bases de datos. *Revista AI*, 17(3), 37–37.

Guan, N., Song, D. y Liao, L. (2019). Gráfico de conocimiento incrustado con conceptos. *Sistemas basados en el conocimiento*, 164, 38–44.

Guarín, C. E. L., Guzman, E. L., & Gonz' alez, F. A. (2015). A model to predict low academic performance at a specific enrollment using data mining. *IEEE Revista Iberoamericana de tecnologías del Aprendizaje*, 10(3), 119–125.

Hamsa, H., Indiradevi, S. y Kizhakkethottam, JJ (2016). estudiante academico modelo de predicción de rendimiento utilizando árbol de decisión y algoritmo genético difuso. *Procedia Technology*, 25, 326–332.

Hand, DJ y Adams, Nuevo México (2014). *Procesamiento de datos*. Wiley StatsRef: Referencia de estadísticas En línea.

Han, J., Pei, J. y Kamber, M. (2011). *Minería de datos: conceptos y técnicas*. Elsevier.

Johnson, L. y Adams, S. (2011). *Perspectivas tecnológicas para la educación terciaria del Reino Unido 2011-2016: un análisis regional del informe de horizonte de NMC*. El Consorcio de Nuevos Medios.

Kabassi, K. y Alepis, E. (2020). Analítica de aprendizaje en educación a distancia y móvil para el diseño de software personalizado. En *Paradigmas de aprendizaje automático* (págs. 185–203). Cham: Springer.

Korte, B. y Vyxen, J. (2008). *Teoría y algoritmos de optimización combinatoria* (4ª ed.). Saltador.

Kotsiantis, SB (2012). Uso de técnicas de aprendizaje automático para propuestas educativas: un sistema de apoyo a la decisión para pronosticar las calificaciones de los estudiantes. *revisión de inteligencia artificial*, 37(4), 331–344.

Krouska, A., Troussas, C. y Virvou, M. (2020). Aprendizaje profundo para el sentimiento de Twitter análisis: el efecto de la incrustación de palabras previamente entrenada. En *Paradigmas de aprendizaje automático* (págs. 111 a 124). Cham: Springer.

Lakshmi, TM, Martin, A., Begum, RM y Venkatesan, VP (2013). Un análisis sobre el rendimiento de los algoritmos de árboles de decisión utilizando datos cualitativos de los estudiantes. *Revista Internacional de Educación Moderna y Ciencias de la Computación*, 5(5), 18.

Lin, CL y Fan, CL (2019). Evaluación de algoritmos CART, CHAID y QUEST: un estudio de caso de defectos de construcción en Taiwán. *Revista de arquitectura asiática e ingeniería de construcción*, 18 (6), 539–553.

Matzavela, V., & Alepis, E. (2017, agosto). Una encuesta para la evolución de la adaptación aprendizaje en dispositivos móviles y electrónicos. En *2017, octava conferencia internacional sobre información, inteligencia, sistemas y aplicaciones (IISA)* (págs. 1 a 5). IEEE.

Matzavela, V. y Alepis, E. (2021). M-learning en la era COVID-19: clase física vs digital. *Educación y tecnologías de la información*, 1–21.

Matzavela, V., Chrysafiadi, K. y Alepis, E. (2017). Cuestionarios y redes neuronales artificiales : una revisión de la literatura sobre técnicas modernas en educación. En *la conferencia mundial de educación en ingeniería del IEEE de 2017 (EDUCON)* (págs. 1700–1704). IEEE.

Nikou, SA y Economides, AA (2016). El impacto de la autoevaluación basada en papel, computadora y móvil en la motivación y el rendimiento científico de los estudiantes. *Computadoras en el comportamiento humano*, 55, 1241–1248.

Nikou, SA y Economides, AA (2017). Evaluación basada en dispositivos móviles: integración aceptación y factores motivacionales en un modelo combinado de la teoría de la autodeterminación y la aceptación de la tecnología. *Computadoras en el comportamiento humano*, 68, 83–95.

Ogunde, AO y Ajibade, DA (2014). Un sistema de minería de datos para predecir las calificaciones de graduación de los estudiantes universitarios utilizando el algoritmo de árbol de decisión ID3. *Revista de informática y tecnología de la información*, 2(1), 21–46.

Papamitsiou, Z. y Economides, AA (2014). Análisis de aprendizaje temporal para adaptativo evaluación. *Revista de análisis de aprendizaje*, 1(3), 165–168.

Parsazadeh, N., Ali, R. y Rezaei, M. (2018). Un marco para el aprendizaje móvil cooperativo e interactivo para mejorar las habilidades de evaluación de información en línea. *Informática y Educación*, 120, 75–89.

Qin, Z. y Lawry, J. (2005). Aprendizaje de árboles de decisión con etiquetas difusas. *ciencias de la información*, 172(1–2), 91–129.

Quinlan, JR (1986). Inducción de árboles de decisión. *Aprendizaje automático*, 1(1), 81–106.

Rizvi, S., Rienties, B. y Khoja, SA (2019). El papel de la demografía en el aprendizaje en línea; Un enfoque basado en árboles de decisión. *Informática y Educación*, 137, 32–47.

Sarker, IH (marzo de 2018). Behavminer: extracción de comportamientos de usuarios a partir de datos de teléfonos móviles para servicios personalizados. En *la conferencia internacional IEEE de 2018 sobre talleres de computación y comunicaciones omnipresentes (talleres PerCom)* (págs. 452–453). IEEE.

Sarker, IH (2018). *Ciencia de datos móviles: hacia la comprensión de las aplicaciones móviles inteligentes basadas en datos*. preimpresión de arXiv arXiv:1811.02491.

Sarker, IH, Colman, A., Han, J., Khan, AI, Abushark, YB y Salah, K. (2020). Behavdt: un árbol de decisiones de comportamiento que aprende a construir un modelo predictivo centrado en el usuario y consciente del contexto. *Aplicaciones y redes móviles*, 25(3), 1151–1161.

Siemens, G. y Baker, RS (2012). Analítica de aprendizaje y minería de datos educativos: hacia la comunicación y la colaboración. En *Actas de la segunda conferencia internacional sobre análisis de aprendizaje y conocimiento* (págs. 252–254). ACM.

Skrbinjek, V. y Dermol, V. (2019). Predecir la satisfacción de los estudiantes usando un árbol de decisión. *Educación terciaria y gestión*, 25(2), 101–113.

Song, YY y Ying, LU (2015). Métodos de árboles de decisión: aplicaciones para la clasificación y predicción. *Archivos de psiquiatría de Shanghai*, 27(2), 130.

Thomas, EH y Galambos, N. (2004). ¿Qué satisface a los estudiantes? Extracción de datos de opinión de los estudiantes con regresión y análisis de árboles de decisión. *Investigación en Educación Superior*, 45(3), 251–269.

Topirceanu, A. y Grossec, G. (2017). Aprendizaje de árbol de decisión utilizado para la clasificación de arquetipos de estudiantes en cursos en línea. *Procedia Computer Science*, 112, 51–60.

Troussas, C., Krouska, A. y Virvou, M. (2020). Uso de un modelo de varios módulos para análisis de aprendizaje para predecir los estados cognitivos de los alumnos y proporcionar vías de aprendizaje y evaluación personalizadas. En *Paradigmas de aprendizaje automático* (págs. 9–22). Cham: Springer.

Virvou, M. y Alepis, E. (2005). Funciones educativas móviles en herramientas de autor para tutorías personalizadas. *Informática y Educación*, 44(1), 53–68.

Virvou, M. y Alepis, E. (2013). Modelado de usuarios en entornos de aprendizaje móvil para estudiantes con necesidades especiales. En *Servicios multimedia en entornos inteligentes* (pp. 7-17). Heidelberg: Springer.

Virvou, M., Kabassi, K., Alepis, E., Kameas, A., Pierrakeas, C. y Theodosiou, A. (2015). Estudio empírico hacia la creación de perfiles de usuarios educativos para los estudiantes de una universidad abierta. En *2015, sexta conferencia internacional sobre información, inteligencia, sistemas y aplicaciones (IISA)* (págs. 1 a 5). IEEE.

Virvou, M., Troussas, C., Caro, J. y Espinosa, KJ (2012). Modelado de usuarios para el aprendizaje de idiomas en Facebook. En *Conferencia internacional sobre texto, habla y diálogo* (pp. 345-352). Berlín, Heidelberg: Springer.

Wu, X., Kumar, V., Quinlan, JR, Ghosh, J., Yang, Q., Motoda, H., ... Zhou, ZH (2008). Los 10 mejores algoritmos en minería de datos. *Sistemas de conocimiento e información*, 14(1), 1–37.

Vasiliki Matzavela recibió un M.Sc. en Informática en 2009 y desde 2016 hasta ahora es Ph.D. Candidato en la sección Informática para la Educación. Ha participado en varias conferencias internacionales en Grecia y otros países. Ella está involucrada en la educación desde hace 18 años y enseña en la educación secundaria. Sus intereses se encuentran en las áreas de Software Educativo, desarrollo de Software Móvil, e-learning y m-learning.

El Dr. Efthimios Alepis recibió un B.Sc. en Informática en 2002 y un Ph.D. en 2009, ambos del Departamento de Informática de la Universidad del Pireo (Grecia). Es Profesor Asistente en el Departamento de Informática de la Universidad de Piraeus desde diciembre de 2013. Es autor/coautor de más de 120 artículos científicos que se han publicado en

V. Matzavela y E. Alepis

revistas internacionales, capítulos de libros y congresos internacionales. El Dr. Alepis es el fundador de la empresa griega "Software Engineering Innovation Group – SEIG", cuyas actividades incluyen, entre otras, la producción de software innovador, servicios de TI y

Informática y Educación: Inteligencia Artificial 2 (2021) 100035

organización de congresos internacionales. Sus intereses de investigación actuales se encuentran en las áreas de programación orientada a objetos, ingeniería de software móvil, interacción humano-computadora, computación afectiva, modelado de usuarios y software educativo.