

Modelización basada en datos a través del sistema de gestión del aprendizaje Moodle: Un estudio empírico basado en una asignatura de enseñanza de las matemáticas

Adrián Pérez-Suay, Steven Van Vaerenbergh, Pascual D. Nebot, Ana B. Pascual-Venteo, Francesc J. Ferri

Abstract—Este trabajo aborda el problema de inferir el rendimiento de los estudiantes a partir de la información adquirida en un Sistema de Gestión del Aprendizaje (SGA). En particular, se exploran las capacidades ofrecidas por Moodle, un SGA ampliamente utilizado. El estudio se realiza sobre datos adquiridos de cuatro clases de una misma asignatura, para las que inferimos el rendimiento de los estudiantes mediante sus notas académicas. La metodología desarrollada describe el grado en que la información adquirida permite predecir las notas de los estudiantes pertenecientes a la evaluación continua, mientras que la predicción de las notas finales de los estudiantes tiene una mayor complejidad intrínseca que requeriría un estudio más profundo de las variables relevantes. A continuación, seguimos un proceso totalmente basado en datos para descubrir similitudes entre clases. En particular, proponemos el uso de una medida de estimación de la dependencia, el criterio de independencia normalizado de Hilbert-Schmidt. Mostramos cómo esta medida de dependencia es útil para determinar relaciones entre clases, basadas en su metodología de enseñanza particular, utilizando únicamente datos adquiridos del SGA. Esto abre la puerta a explorar las capacidades de los SGA en la búsqueda de similitudes entre los cursos ofrecidos a lo largo de una plataforma educativa. Con el fin de ayudar a la comunidad y servir como una forma común de comparación, proporcionamos el código fuente de la metodología propuesta.

Index Terms—Cursos educativos, Instituciones educativas, Sistema de gestión del aprendizaje, Moodle SGA, Modelos analíticos, Analítica del aprendizaje, Modelización predictiva, Rendimiento de los estudiantes.

I. INTRODUCTION

En los últimos años, los sistemas de gestión del aprendizaje (SGA) se han implantado plenamente en la mayoría de las universidades. Debido al efecto de la pandemia de COVID-19, muchos cursos presenciales se trasladaron a cursos totalmente en línea o a entornos de *formación combinada* (en inglés, *blended learning*) [29], en función de decisiones educativas,

Adrián Pérez-Suay, E-mail: Adrian.Perez@uv.es y Pascual D. Nebot trabajan en el Departament de Didàctica de la Matemàtica, Universitat de València, Av. Tarongers 4, 46022 València, España. Steven Van Vaerenbergh trabaja en el Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Av. de los Castros 48, 39005 Santander, España. Ana B. Pascual-Venteo trabaja en el Laboratorio de Procesado de Imagen (IPL) de la Universidad de Valencia, Catedrático A. Escardino - 46980 Paterna, Valencia (España). Francesc J. Ferri trabaja en el Departament d'Informàtica, Universitat de València, València, España. Pascual-Venteo y Steven Van Vaerenbergh fueron apoyados por el proyecto con referencia RENOVA_PID, UV-SFPIE_PID-2079995.

sanitarias y/o políticas. Este cambio en las tendencias educativas afecta tanto a la forma habitual de enseñar, aprender e interactuar de profesores y alumnos.

Un SGA es una herramienta fundamental en torno a la cual gira la tecnología educativa. El SGA suele ser una plataforma online diseñada para planificar, dinamizar, tutorizar, evaluar, gestionar y administrar actividades de aprendizaje [27]. Se encarga principalmente de la gestión de usuarios (alumnos, profesores y administradores), recursos y actividades, y del seguimiento del proceso de aprendizaje mediante evaluaciones e informes. Además, proporciona herramientas de comunicación entre los miembros de la comunidad formativa (mensajería interna, chats, videoconferencia, foros, etc.). El uso de una plataforma virtual de este tipo permite tanto al profesorado como al alumnado disponer de una fuente de información común: por un lado, el profesor puede compartir información relevante con los alumnos, como apuntes, teoría, tareas, vídeos, cuestionarios, etc. Por otro, los alumnos disponen de un lugar centralizado donde se asigna la información y al que pueden subir sus tareas.

Los entornos de aprendizaje basados en SGA están diseñados para adquirir datos de los estudiantes, que incluyen datos de tareas y asistencia, todos ellos almacenados en archivos de registro como acciones relacionadas con la interacción de los estudiantes mientras utilizan la plataforma [20]. A modo de ejemplo, una asignatura de un solo semestre podría tener 60 estudiantes, cada uno de los cuales interactúa con el sistema, generando un archivo de registro que contiene varias decenas de miles de acciones de los estudiantes. Dada la cantidad de asignaturas que se imparten en una universidad media, así como la cantidad de titulaciones por facultad, la cantidad de datos generados en cada curso por los alumnos a través de estas plataformas se convierte rápidamente en un volumen muy grande.

Un uso interesante de los datos del alumnado recogidos por un SGA radica en analizar y visualizar la información sobre el progreso de una clase. Idealmente, se puede utilizar para mejorar el procedimiento de enseñanza-aprendizaje, que actualmente es un tema de investigación muy activo [12], [31]. Uno de los principales objetivos es descubrir patrones en la actividad de los estudiantes que puedan ser útiles para hacer inferencias, realizar una detección temprana de problemas y proporcionar medidas estimadas de la eficacia del curso. En particular, el problema de inferir el rendimiento

de los estudiantes está recibiendo atención en el campo de la investigación, ya que se trata de un problema abierto de crucial importancia. Disponer de buenas estimaciones de las calificaciones académicas podría ayudar a comprender mejor las necesidades, los fracasos, los éxitos y, de forma más general y ambiciosa, a entender cómo está funcionando y afectando a los alumnos la metodología prevista.

Recientemente, el análisis y la predicción del rendimiento de los estudiantes ha recibido especial atención en el campo de la investigación educativa [7], [27]. Tradicionalmente, las inferencias sobre las calificaciones de los estudiantes se realizaban principalmente utilizando diferentes fuentes de información fuera de línea, como atributos que incluyen las calificaciones de los estudiantes, datos demográficos, sociales y características relacionadas con la escuela. Estos datos se recopilaban principalmente mediante informes escolares. Un ejemplo de esta metodología puede encontrarse en [9]. El uso de SGA ha llevado a incluir datos del alumno adquiridos de una plataforma online a la hora de realizar inferencias, incluyendo acciones relativas al comportamiento del alumno a lo largo del curso tales como actividad, interacción con la plataforma, asistencia y evaluaciones.

En [7], los autores proponen realizar predicciones del rendimiento del alumnado en base a cuatro factores: acceso al curso en la plataforma, cuestionario, tarea y edad. Encontraron que la edad no es un factor relevante y que realmente no ayuda a inferir el rendimiento del estudiante. Los resultados comunicados muestran un coeficiente de correlación al cuadrado relativamente bajo, lo que pone de manifiesto la complejidad intrínseca del problema. En otro trabajo relacionado [17], se propone un procedimiento para obtener modelos predictivos del rendimiento de los estudiantes que sean compatibles entre diferentes cursos.

Dado que el número total de variables consideradas como inputs o covariables puede ser muy grande, es necesario centrarse en las fuentes de información más relevantes y ricas. Por ejemplo, [24] propone la integración de dispositivos móviles en la enseñanza y el aprendizaje para medir el rendimiento escolar del alumnado. Otra fuente de información es la sugerida en [21], que propone el uso de la información generada por los foros de discusión online. Estos foros recogen la interacción del alumnado y son un lugar donde los estudiantes pueden resolver y discutir las dudas y problemas de sus compañeros, lo que también sirve para mantener informado al instructor sobre estas dudas.

La predicción del rendimiento de los estudiantes en los programas de estudio también se propuso en [5], donde los autores ofrecen un estudio de las progresiones típicas, y las combinan con los resultados de la predicción. Una de las conclusiones de este trabajo fue que podría ser beneficioso identificar dos grupos importantes de estudiantes, los estudiantes de bajo y alto rendimiento, que adquirieron notas bajas y altas durante el período estudiado, respectivamente. Los resultados presentados indican que considerando sólo un pequeño número de cursos es posible obtener inferencias representativas para los cursos restantes. En [25], los autores ofrecen una visión general de las distintas fuentes de información. Concluyen que los modelos proporcionados basados en SGA y learning

analytics tienen un alto poder predictivo.

Un conjunto diferente de estudios aplica enfoques de aprendizaje estadístico o técnicas de inteligencia artificial para abordar la predicción del rendimiento de los estudiantes [28]. En [4], los autores proponen utilizar técnicas de agrupación junto con un clasificador para categorizar el perfil de los estudiantes. En [16], los autores utilizan la librería de modelado de ecuaciones estructurales para validar que es factible obtener una medida de la actividad en SGA mediante el análisis de datos de archivos de registro en Moodle. Motivado por el poder de las técnicas de Deep Learning (DL) para resolver problemas complejos, el trabajo en [19] reporta los resultados del uso de una Red Neuronal Artificial (RNA) sobre un conjunto de datos públicos. Cabe mencionar que los modelos de DL requieren una gran cantidad de datos que, afortunadamente, pueden estar disponibles en un SGA. La metodología propuesta en [30] permite realizar el modelado de estudiantes con atributos relacionados con el rendimiento y atributos no relacionados con el rendimiento utilizando el concepto de Student Attribute Matrix (SAM). Por último, algunos estudios, como el de [2], abordan el escenario en el que el número de estudiantes es relativamente bajo, por ejemplo, inferior a 30. En estos casos, el conjunto de datos disponible para la modelización de los estudiantes es muy limitado. En estos casos, el conjunto de datos disponible puede ser pequeño, lo que requiere técnicas como las máquinas de vectores soporte y el análisis discriminante lineal.

En las universidades españolas, la pandemia COVID-19 ha llevado a la suspensión parcial o total de las clases presenciales, promoviendo la docencia online o híbrida [1], generando un gran aumento de las evaluaciones participativas. La búsqueda de medios alternativos a la evaluación tradicional, típicamente exámenes presenciales, ha contribuido a asignar mayor peso a las evaluaciones continuas frente a los exámenes finales. En la evaluación continua es necesario disponer de pruebas suficientes para asegurar la adquisición de las competencias de cada asignatura. Se recomienda [1] que dicha evaluación constituya al menos el 60% de la nota final, o en el caso de que exista un examen final, que suponga más del 40% de la nota. Para evitar el fraude, un factor clave consiste en perseguir el máximo nivel de identificación del alumno en la modalidad online [15].

Para hacer frente a la creciente cantidad de tareas, trabajos y actividades que necesitan ser supervisadas por el profesorado, proponemos explotar la información que el SGA almacena sobre diferentes tipos de eventos en una asignatura. El SGA almacena todos estos eventos en un log-file que básicamente guarda un registro detallado de todas las diferentes acciones y eventos realizados por los alumnos y profesores dentro de esa asignatura. En particular, centramos el estudio en la plataforma Moodle, para la que proponemos un método específico de preprocesamiento de datos para convertir el archivo de registro de eventos del SGA en un formato estándar para el modelado basado en datos. Una vez convertido a este formato, realizamos dos experimentos diferentes con estos datos para obtener diferentes perspectivas del proceso de enseñanza-aprendizaje. En un primer experimento, procesamos los datos almacenados siguiendo prácticas similares a estudios

TABLE I
INFORMACIÓN ADQUIRIDA POR MOODLE. CADA FILA DEL ARCHIVO DE REGISTRO CONTIENE LAS SIGUIENTES VARIABLES, ENUMERADAS AQUÍ CON UNA BREVE DESCRIPCIÓN Y UN EJEMPLO DE FORMATO.

Variable	Descripción	Formato/Categorías
Time	Marca temporal del Evento	dd/mm/yy, hh:mm
UserFullName	Nombre del usuario realizando un Evento	primer apellido segundo apellido, nombre
AffectedUser	Usuario afectado por el Evento	primer apellido segundo apellido, nombre
EventContext	Nombre de categoría y Etiqueta del Evento	Assignment, Attendance, External tool, Course, File, Forum, Label, Other, URL
Component	Etiqueta	Assignment, Attendance, Excel spreadsheet, External tool, File, File submissions, Forum, Grader report, Logs, Overview report, Recycle bin, Single view, Submission comments, System, URL, User report
EventName	Information about EventContext	Ver Tabla II
Description	Que usuario realiza el evento y a quien afecta	user names
Origin	Etiqueta de procedencia del Evento	cli, restore, web
IPAddress	Dirección del protocolo de internet	Etiqueta numérica asignada al dispositivo conectado utilizando el protocolo de internet para la comunicación

previos [3], [7] para realizar tareas de inferencia como la regresión sobre el rendimiento del alumnado. En un segundo experimento, cuantificamos las similitudes entre las diferentes clases de una manera totalmente basada en datos. Por último, proporcionamos una configuración experimental exhaustiva y detallada y compartimos el código fuente de los métodos propuestos.

Los datos utilizados en este estudio se extraen de cuatro clases diferentes pertenecientes al mismo curso académico, durante las cuales los usuarios interactúan con la interfaz del SGA, que, a su vez, almacena las acciones de los usuarios en un archivo de registro. Estas actividades se denominan *eventos* realizadas por los usuarios y almacenan información valiosa como la marca de tiempo, el nombre del usuario, el contexto del evento o el nombre del evento entre otros, véase la Tabla I para una descripción más detallada. Algunas de estas variables han demostrado ser útiles en estudios previos [3], [7] para hacer inferencias sobre el rendimiento de los estudiantes en base a la actividad realizada a través de plataformas online. Siguiendo los estudios mencionados, nos centramos en las variables *nombre del evento* y *contexto del evento*. En la primera, la plataforma crea una fila en la tabla de datos para registrar cada uno de los nombres de evento detallados en la Tabla II. La información registrada en la columna nombre de evento hace referencia a la actividad del alumno, profesor y administrador en el curso. En esta última, la columna *contexto del evento* almacena la categoría del evento con respecto a *tareas, asistencia, curso, herramienta externa, archivo, etiqueta, otro, URL*; se refiere a los diferentes materiales fuente cargados por el profesor durante el curso. A lo largo de este trabajo nos referiremos a esta columna como *dimensión metodológica*, ya que recoge aspectos útiles para caracterizar la metodología docente seguida por el profesor.

Para poner de manifiesto la importancia del problema abordado, recordemos que se han instalado SGA en la mayoría de las universidades de todo el mundo. Moodle es una solución SGA de código abierto ampliamente utilizada para organizar la metodología y los recursos de los cursos. Además, existe un amplio catálogo de plugins que amplían la funcionalidad básica de Moodle, incluyendo plugins que pueden ser utiliza-

TABLE II
CATEGORÍAS DE NOMBRES DE EVENTOS DE MOODLE UTILIZADAS PARA MEDIR LA ACTIVIDAD DE LOS ESTUDIANTES EN LA PLATAFORMA.

Number	Event Name
1	A file has been uploaded
2	A submission has been submitted
3	Attendance taken by student
4	Comment created
5	Comment deleted
6	Course module instance list viewed
7	Course module viewed
8	Course user report viewed
9	Course viewed
10	Grade item created
11	Grade overview report viewed
12	Grade user report viewed
13	Remove submission confirmation viewed
14	Session report viewed
15	Submission created
16	Submission form viewed
17	Submission updated
18	The status of the submission has been updated
19	The status of the submission has been viewed
20	User graded
21	User list viewed
22	User profile viewed

dos para realizar tareas de análisis.

Aquí, proponemos una forma externa y totalmente automática de obtener información valiosa a partir de los archivos de registro de Moodle siguiendo un enfoque estadístico. Es robusto ya que se adapta a los cambios en el número de variables dentro del *nombre del evento*. El código fuente es compartido ¹ para permitir la reproducibilidad del enfoque propuesto.

Este trabajo está estructurado de la siguiente manera. Tras la introducción en la Sección I, en la que se define el problema y se enumera la literatura relacionada, detallamos la metodología propuesta en la Sección II. La sección III contiene los detalles de los datos utilizados en los experimentos, y la descripción completa de los experimentos que se llevaron a cabo se da en la sección IV. Por último, en la Sección V se ofrece una discusión de los resultados y las principales conclusiones de este trabajo.

¹https://www.uv.es/pesuaya/data/code/learn_moodle.zip

II. METODOLOGÍA

En esta sección, presentamos y discutimos los métodos que se utilizarán en los experimentos. En primer lugar, especificamos la transformación de datos propuesta para convertir el archivo de registro de eventos SGA sin procesar en un formato de datos común adecuado para aplicar métodos de aprendizaje automático. En segundo lugar, detallamos las técnicas estadísticas utilizadas para realizar tareas de regresión sobre estos datos. En tercer lugar, explicamos el método de estimación de la dependencia elegido, que nos permite encontrar similitudes entre clases en el segundo experimento.

A. Datos y preprocesado

El SGA almacena los datos de registro sin procesar de forma secuencial en un archivo de registro. Los observadores de eventos no pueden modificar los datos de eventos ni interrumpir el envío de eventos, ya que se trata de un canal de comunicación unidireccional. Las variables almacenadas por Moodle se detallan en la Tabla II, y son la principal fuente de información para realizar la inferencia considerada en este trabajo. En concreto, el archivo de registro de Moodle consiste en una secuencia de eventos realizados por los usuarios, y se almacena en formato de texto plano. Este archivo de registro es específico de las acciones de un estudiante dentro de un curso, y no incluye otra información de registro como errores. Por lo tanto, proporciona una gran cantidad de información útil para determinar y caracterizar la actividad del alumno durante el proceso de enseñanza-aprendizaje. Utilizamos un conjunto de variables útiles para hacer inferencias [3], [7], y a continuación detallaremos cómo extraer estas variables del archivo de registro.

La figura 1 ilustra el procedimiento de extracción de datos de un archivo de registro en bruto y su conversión en un tensor \mathbf{T} que representa el recuento de eventos. El componente metodológico, que se refiere a *Contexto del evento*, depende de cada asignatura y profesor. Para cada elemento metodológico, contamos qué variables *Nombre del evento* se han registrado. En concreto, construimos un tensor \mathbf{T} en el que cada elemento $\mathbf{T}_{c,a,s}$ contiene la cantidad de veces que el alumno s ha realizado el evento con nombre a y contexto c dentro del periodo de tiempo del curso. Denotamos el número total de *Contextos de Evento* como C , con $1 \leq c \leq C$, el número total de *Nombres de Evento* como A , con $1 \leq a \leq A$, y el número total de *estudiantes* como s , con $1 \leq s \leq S$, obteniendo el tensor $\mathbf{T} \in \mathbb{R}^{C \times A \times S}$. Una vez que los datos se transforman desde el archivo de registro en bruto a un tensor \mathbf{T} , esta información se puede resumir adecuadamente para aplicar métodos de aprendizaje automático para la inferencia. Por ejemplo, si \mathbf{T} se aplana con respecto a su primera dimensión (el *Contexto de Evento*) promediando los recuentos de ocurrencia de eventos, la información puede ser revelada sobre la actividad del usuario a lo largo de la plataforma SGA, que es equivalente al enfoque propuesto en [3], [7].

B. Extracción de características y reducción de la dimensionalidad

El Análisis de Componentes Principales (ACP) es una de las técnicas más utilizadas para realizar reducción de dimensionalidad, eliminación de ruido y análisis exploratorio de datos. PCA consta de tres pasos principales: 1) Centrar los datos (de modo que su media sea el vector cero), 2) calcular la matriz de covarianza, y 3) calcular los vectores propios, que son las direcciones principales de los datos, y los valores propios, que se refieren a la cantidad de varianza de los datos explicada por cada vector propio.

En este estudio, utilizamos el PCA primero para realizar una visualización exploratoria de los datos, y después también para obtener una representación de bajo rango de los datos con el fin de mejorar la robustez de las técnicas estadísticas posteriores.

C. Regresión

En los experimentos se emplean varios métodos de regresión, que pueden clasificarse en lineales y no lineales. Los métodos lineales obtienen su resultado como una combinación lineal de las variables de entrada, mientras que los no lineales se basan en relaciones más complejas y no lineales entre las variables de entrada, y dan lugar a modelos más flexibles que en muchas situaciones podrían ajustarse mejor a los datos.

En el entorno de regresión actual, se nos dan algunas variables observadas de valor real que se almacenan en una matriz de datos $\mathbf{X} \in \mathbb{R}^{N \times D}$. Aquí, N se refiere al número total de muestras y D es el número de variables o covariables del estudio. El objetivo es hacer inferencia de otra variable, $y \in \mathbb{R}^n$.

1) Métodos lineales:

a) *Least Squares Regression*: El método de regresión por mínimos cuadrados (LS) es una técnica de inferencia estándar que tiene como objetivo estimar un conjunto de pesos $\mathbf{w} \in \mathbb{R}^D$ tal que la variable de salida estimada *haty* corresponde al modelo lineal $\hat{y} = \mathbf{X}\mathbf{w}$. En este caso, los pesos \mathbf{w} se obtienen minimizando la función de coste $\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2$.

b) *Ridge regression (RIDGE)*: Cuando los datos de entrada están contaminados con ruido, la regresión lineal por mínimos cuadrados estándar puede dar lugar a estimaciones deficientes. En este caso, se puede añadir un término de regularización al problema original de mínimos cuadrados, conocido como regularización de Tikhonov, obteniendo el problema de regresión ridge. El término de regularización puede ajustarse para adaptarse a la cantidad de ruido en la entrada.

2) *Métodos no lineales*: Los métodos anteriores se basan en la inferencia a través de un modelo lineal y son óptimos bajo estos supuestos. Sin embargo, cuando se trata de un escenario más complejo en el que los datos contienen relaciones no lineales, se requieren métodos de regresión no lineal. En este estudio incluimos las técnicas no lineales de los métodos Kernel [23], que son una clase de técnicas estándar para regresión no lineal, clasificación, estimación de dependencia y otras tareas. En particular, consideramos las técnicas basadas en kernel de Regresión de Vectores de Soporte [10] (SVR), y Kernel Ridge Regression [26] (KRR).

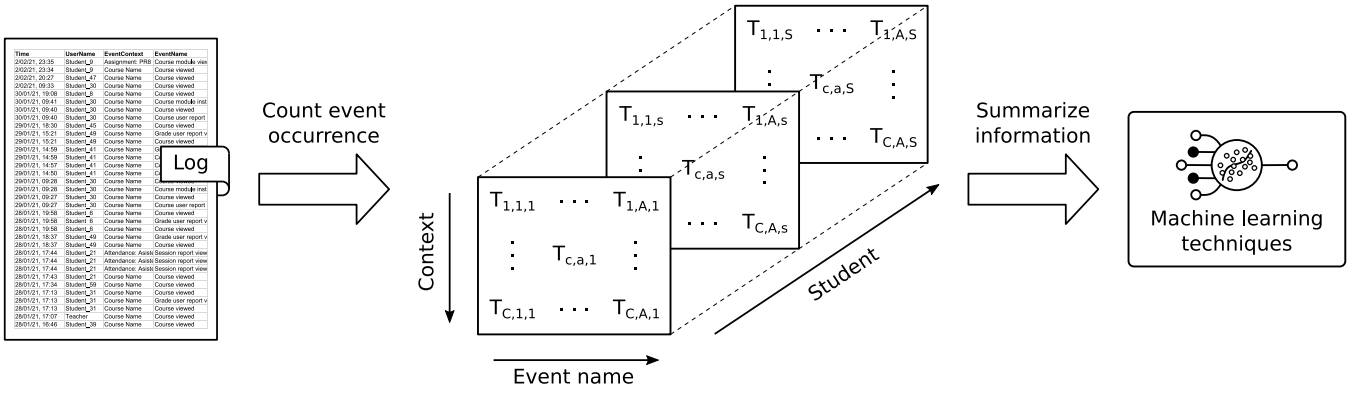


Fig. 1. Esquema de transformación de los datos sin procesar en formato de archivo de registro (log-file) se transforman en vectores de características del alumno. Cada elemento del tensor $T_{c,a,s} \in \mathbb{R}$ en la representación tensorial se refiere al número de ocurrencias del c -ésimo Contexto de Evento, el a -ésimo Nombre de Evento para el s -ésimo estudiante.

En los métodos kernel, los datos de entrada \mathbf{x}_i se mapean en un Espacio de Hilbert Kernel Reprodutor \mathcal{H} a través de una función de mapeo no lineal $\mathbf{x}_i \rightarrow \phi(\mathbf{x}_i) \in \mathcal{H}$. Por lo general, esta función de mapeo es implícita, pero se puede calcular mediante el uso de una función kernel definida positiva. Uno de los kernel más utilizados es el kernel de función de base radial (RBF), que es $k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / (2\sigma^2)}$. La solución del modelo de pesos se expresa a través del Teorema del Representante [22].

D. Estimación de la dependencia

Una medida típica que indica las relaciones entre variables es la correlación. Aunque el coeficiente de correlación es bien conocido y ampliamente utilizado en la comunidad científica, se trata de una medida lineal con limitaciones para captar la dependencia estadística entre variables [18]. Por esta razón, empleamos una medida de dependencia que tiene en cuenta las relaciones no lineales en los datos. En concreto, nos centramos en el Criterio de Independencia de Hilbert-Schmidt (HSIC) [14], que es una medida de independencia no paramétrica.

El estimador empírico del HSIC se calcula utilizando funciones kernel y para un par de muestras multivariantes de tamaño N , como $\mathbf{X} \in \mathbb{R}^{N \times D_x}$ y $\mathbf{Y} \in \mathbb{R}^{N \times D_y}$. En particular, definamos un mapeo (posiblemente no lineal) $\phi: \mathcal{X} \rightarrow \mathcal{F}$ de tal manera que el producto interno entre las características está dada por una definida positiva (p.d.) función kernel $K_x(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$. El espacio de características \mathcal{F} tiene la estructura de un espacio de Hilbert de núcleo reproductor (RKHS). Denotemos ahora otro mapa de características $\psi: \mathcal{Y} \rightarrow \mathcal{G}$ con función kernel p.d. asociada $K_y(\mathbf{y}, \mathbf{y}') = \langle \psi(\mathbf{y}), \psi(\mathbf{y}') \rangle$. Entonces, el operador de covarianza cruzada entre estos mapas de características es un operador lineal $\mathcal{C}_{\mathbf{xy}}: \mathcal{G} \rightarrow \mathcal{F}$ tal que $\mathcal{C}_{\mathbf{xy}} = \mathbb{E}_{\mathbf{xy}}[(\phi(\mathbf{x}) - \mu_x) \otimes (\psi(\mathbf{y}) - \mu_y)]$, donde \otimes es el producto tensorial, $\mu_x = \mathbb{E}_{\mathbf{x}}[\phi(\mathbf{x})]$, and $\mu_y = \mathbb{E}_{\mathbf{y}}[\psi(\mathbf{y})]$. Más detalles se pueden encontrar en [6], [11]. La norma al cuadrado del operador de covarianza cruzada, $\|\mathcal{C}_{\mathbf{xy}}\|_{\text{HS}}^2$, se denomina Criterio de Independencia de Hilbert-Schmidt (HSIC) y puede expresarse en términos de kernels [13]. Dados los conjuntos de datos

muestrales $\mathbf{X} \in \mathbb{R}^{N \times D_x}$, $\mathbf{Y} \in \mathbb{R}^{N \times D_y}$, y el conjunto de datos $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, con n pares $\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$ agrupados colectivamente en una matriz muestral $\mathbf{Z} \in \mathbb{R}^{n \times (d_x + d_y)}$ extraída de la distribución de probabilidad conjunta $\mathbb{P}_{\mathbf{xy}}$, un estimador empírico de HSIC es [13]:

$$\text{HSIC}(\mathcal{F}, \mathcal{G}, \mathbb{P}_{\mathbf{xy}}) = \frac{1}{N^2} \text{Tr}(\mathbf{K}_x \mathbf{H} \mathbf{K}_y \mathbf{H}) \quad (1)$$

donde $\text{Tr}(\cdot)$ es el operador de traza (la suma de las entradas diagonales), \mathbf{K}_x , \mathbf{K}_y son las matrices del núcleo para las variables aleatorias de entrada \mathbf{x} e \mathbf{y} , y $\mathbf{H} = \mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T$ centran los datos en los espacios de características \mathcal{F} y \mathcal{G} , respectivamente. Aquí δ representa el símbolo de Kronecker, donde $\delta_{i,j} = 1$ si $i = j$, y cero en caso contrario.

HSIC es una medida de independencia que toma el valor cero si y sólo si las muestras de entrada son independientes [13]. Sin embargo, este criterio de independencia no tiene límite superior y podría llevar a conclusiones engañosas en nuestro entorno. Para evitar esto, normalizamos la medida HSIC por el producto de las normas del núcleo como se menciona en [8], nos referimos a esto como NHSIC. El estimador empírico de NHSIC es:

$$\text{NHSIC}(\mathcal{F}, \mathcal{G}, \mathbb{P}_{\mathbf{xy}}) = \frac{\text{Tr}(\mathbf{K}_x \mathbf{H} \mathbf{K}_y \mathbf{H})}{\|\mathbf{K}_x \mathbf{H}\|_{\text{Fro}} \|\mathbf{K}_y \mathbf{H}\|_{\text{Fro}}}, \quad (2)$$

donde $\|\cdot\|_{\text{Fro}}$ es la norma de Frobenius, es decir $\|A\|_{\text{Fro}} = \sqrt{\text{Tr}(AA^T)}$.

III. DESCRIPCIÓN DE LOS DATOS

En esta sección explicamos los datos utilizados en los experimentos. Consideramos 4 clases diferentes dentro de la misma asignatura, impartidas por 3 profesores diferentes. Las clases se denotan como 4P, 4R, 4S y 4U, mientras que los profesores se denotan como profesores A, B y C. Se proporciona un resumen en la Tabla III. El profesor C aplica la misma metodología en las clases 4S y 4U, que está estrechamente relacionada con la metodología utilizada por el profesor A en la clase 4P. La primera mitad del curso en la clase 4R (impartida por los profesores A y B) la imparte el profesor A siguiendo exactamente la misma metodología que

en la clase 4P. La metodología utilizada por el profesor B no está relacionada con los profesores A y C, pero el profesor A y B comparten la mitad del curso en la clase 4R.

En la Tabla III, hay un total de 193 estudiantes en el estudio que han interactuado con Moodle. En conjunto, sus acciones constituyen más de 162000 líneas en el archivo de registro. La columna *Contexto del evento* se refiere a la dimensión metodológica, que contiene información sobre las diferentes tareas, ayudas, URLs y en general el material utilizado en la metodología por el Profesor. La columna *Nombre del Evento* indica el número de diferentes *Nombres del Evento* considerados, que es fijo ya que es un parámetro interno de la plataforma referido al número total de variables adquiridas, ver Tabla II.

Se han utilizado datos del curso académico 2020/21. El objeto del caso de estudio fue “La enseñanza de las matemáticas en la escuela infantil”, que se incluye en el 4º (y último) curso del Grado en Educación Primaria de la *Universitat de València*. Debido a los efectos de la pandemia del COVID-19, cada curso de docencia se dividió en dos grupos diferentes a lo largo de la facultad. No hemos eliminado este hecho en los datos, y lo utilizamos como factor de robustez de la propuesta. Por ejemplo, las clases 4P y 4R han dividido las tareas utilizando una tarea por grupo, lo que ha dado lugar a un mayor número de tareas. No procesamos estos datos a mano, siguiendo el espíritu de nuestro enfoque, que está totalmente automatizado y no requiere intervención humana.

TABLE III
INFORMATION OF CONSIDERED CLASSES IN THE STUDY REGARDING CLASS NAME, RESPONSIBLE TEACHER, NUMBER OF LINES IN THE LOG-FILE, AMOUNT OF DIMENSIONS IN THE EVENT CONTEXT (METHODOLOGICAL DIMENSION) AND EVENT NAME NUMBER OF COMPONENTS.

Class	Teacher	Students	Log lines	Event Context	Event Name
4P	A	46	35300	89	22
4R	A,B	45	45333	106	22
4S	C	59	44368	50	22
4U	C	43	37257	49	22

IV. EXPERIMENTOS

En esta sección abordamos dos experimentos diferentes. El primero se refiere a la inferencia de las calificaciones finales y continuas del rendimiento de los estudiantes utilizando los datos generados por el archivo de registro. El segundo tiene como objetivo encontrar dependencias entre clases en un procedimiento totalmente basado en datos.

Previamente a los experimentos y con fines de visualización, hemos procesado los datos brutos de los archivos y calculado los componentes principales. Estos componentes representan la mayor varianza entre los datos originales. En la Figura 2, podemos ver un diagrama de dispersión de los componentes principales primero y segundo de cada una de las cuatro clases. Los datos de la Fig. 2 a) representan la clase 4P, impartida por el profesor A, y aparecen más concentrados en valores similares; los datos de la Fig. 2 b) representan la clase 4R, impartida por los profesores A y B, mostrando una mayor

dispersión; los datos de las Figs. 2 c) y d) representan las clases 4S y 4U, impartidas por el mismo profesor C, y reflejan una distribución de datos similarmente cercana.

El análisis anterior proporciona una primera idea cualitativa de la similitud entre los datos obtenidos para las distintas clases. En el último experimento, descrito en la sección HSIC, analizaremos el uso de una medida de dependencia para cuantificarla. Allí se tratará de validar el uso de NHSIC para establecer similitudes a partir de la búsqueda de dependencias entre clases.

A. Rendimiento del alumnado

En primer lugar, validamos nuestro enfoque infiriendo las notas de los estudiantes utilizando los datos disponibles. Tras procesar los datos, convertimos los archivos de registro sin procesar de la plataforma Moodle en matrices de muestras (filas) y covariables (columnas). El objetivo es obtener estimaciones sobre el rendimiento de los estudiantes utilizando únicamente los datos adquiridos del SGA.

En concreto, planteamos el problema como uno de predicción tanto de la evaluación continua como de las notas finales de los estudiantes. Por un lado, el número de trabajos y tareas requiere un aumento de tiempo y esfuerzo por parte de los profesores para correcciones y evaluaciones. La información adquirida por la plataforma parece estar más relacionada y ser más útil para explicar las notas continuas. Por otro lado, predecir la evaluación final es una tarea más difícil, ya que incluye algunas variables que no se pueden medir bien dentro del SGA, como el tiempo de estudio y los conocimientos previos.

En ambos experimentos, al predecir las notas continuas y finales de los estudiantes, hemos dividido los datos en un 70% para el entrenamiento y la validación de los modelos y un 30% para probar los modelos finales. Se realizó una media de 25 ensayos para obtener los resultados presentados.

Junto con las notas continuas y finales de los estudiantes, proponemos estudiar la idoneidad de los métodos a lo largo de la división de datos intragrupo e intergrupo. Proporcionamos resultados utilizando cinco métodos diferentes basados en el aprendizaje estadístico y damos mediciones de tres métricas, el coeficiente de correlación (R), el error cuadrático medio (RMSE) y el error absoluto medio (MAE).

a) *Predicción intragrupo*: En primer lugar, ofrecemos resultados sobre la estimación intragrupo de las notas del alumnado. El objetivo es validar el rendimiento de los métodos de inferencia para cada clase. La tabla IV contiene los resultados para las notas continuas y finales de las cuatro clases. Hemos marcado en negrita los mejores resultados para cada métrica y clase. curso. Las notas finales parecen difíciles de inferir, como se puede ver en los resultados al compararlos con los resultados continuos. En cuanto al bloque de marcas continuas, los mejores resultados los obtiene siempre uno de los métodos no lineales: SVM y KRR. La mayoría de los mejores resultados a lo largo de las diferentes métricas se encuentran en el método KRR y conducen a puntuaciones relativamente altas en correlación que rondan $R0,9$ en la evaluación continua. Este buen resultado no fue el mismo en

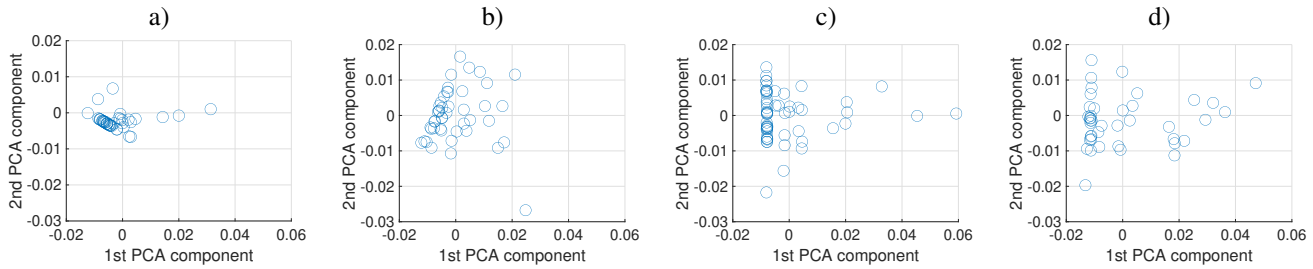


Fig. 2. Representación de los datos en la primera y segunda dimensión de PCA del conjunto de datos propuesto, en a) 4P, b) 4R, c) 4S, y d) 4U.

TABLE IV

COMPARACIÓN INTRAGRUPO. RENDIMIENTO SEGÚN LA REGLA DEL 70%-30% (PARA LOS CONJUNTOS DE ENTRENAMIENTO Y TEST) LOS VALORES SON UN PROMEDIO DE 25 REPETICIONES INDEPENDIENTES; COEFICIENTE DE CORRELACIÓN (R), RAÍZ CUADRADA DEL ERROR CUADRÁTICO MEDIO (RMSE) Y ERROR ABSOLUTO MEDIO (MAE) PARA LAS NOTAS CONTINUAS Y FINALES DEL ALUMNADO. EL MEJOR VALOR PARA LAS NOTAS CONTINUAS Y FINALES APARECE EN NEGRITA EN CADA FILA.

		Continua				Final			
		LS	RIDGE	SVR	KRR	LS	RIDGE	SVR	KRR
R	4P	0.04	0.20	-0.17	0.06	-0.04	0.02	0.01	0.03
	4R	0.27	0.58	0.82	0.85	0.02	-0.08	0.27	0.14
	4S	0.40	0.77	0.63	0.81	0.05	0.08	-0.06	0.02
	4U	0.64	0.87	0.81	0.89	0.18	0.61	0.60	0.54
RMSE	4P	2.81	0.86	0.86	0.86	3.63	1.24	1.23	1.28
	4R	15.01	1.30	1.18	0.67	5.81	1.22	1.18	1.19
	4S	4.35	0.56	0.64	0.52	12.90	1.07	1.14	1.11
	4U	0.77	0.32	0.48	0.32	2.48	0.84	0.87	0.89
MAE	4P	1.90	0.42	0.36	0.39	2.71	0.73	0.73	0.77
	4R	11.07	0.88	0.53	0.41	4.25	0.75	0.69	0.72
	4S	3.39	0.36	0.37	0.32	9.44	0.84	0.92	0.89
	4U	0.58	0.22	0.27	0.20	1.95	0.69	0.73	0.73

la clase 4P, donde el modelo de regresión Ridge (RIDGE) obtuvo el mejor resultado. Tanto los resultados obtenidos en RMSE como en MAE indican a KRR como la mejor opción en la evaluación continua y a los métodos RIDGE y/o SVR en la inferencia de las notas de la evaluación final.

b) *Predicción intergrupo*: La tabla V proporciona los resultados para predecir las notas del alumnado sin distinción de grupo de clase. En el caso continuo, el coeficiente de correlación es de 0.8, que es similar a los resultados obtenidos en el experimento intragrupo. En el caso de la nota final, el resultado obtenido es peor en términos de correlación que en el continuo. Por lo demás, este resultado es mejor tanto en términos de RMSE como de MAE. Este comportamiento puede deberse al tamaño relativamente pequeño de la muestra utilizada en el caso continuo. El método SVR parece ser un enfoque más robusto y plausible para construir el modelo de rendimiento de los estudiantes en esa configuración intergrupos. Se puede observar que este método ha obtenido los mejores resultados en las tres métricas. Vale la pena mencionar que a pesar de que KRR no ha obtenido los mejores resultados en la experimentación intergrupos, conduce a resultados similares con SVR. Como conclusión global de ambos experimentos intragrupo e intergrupo, se podría derivar que KRR es uno de los mejores algoritmos ya que ha obtenido buenos resultados en las tres métricas a la hora de hacer inferencias sobre el rendimiento del alumnado.

B. Descubrimiento de dependencias basado en datos

En el segundo experimento, utilizamos el Criterio de Independencia de Hilbert-Schmidt normalizado (NHSIC) como

TABLE V

COMPARACIÓN ENTRE GRUPOS. RENDIMIENTO SEGÚN LA REGLA DEL 70%-30% (PARA LOS CONJUNTOS DE ENTRENAMIENTO Y TEST) LOS VALORES SON UN PROMEDIO DE 25 REPETICIONES INDEPENDIENTES; COEFICIENTE DE CORRELACIÓN (R), RAÍZ CUADRADA DEL ERROR CUADRÁTICO MEDIO (RMSE) Y ERROR ABSOLUTO MEDIO (MAE) PARA LAS NOTAS CONTINUAS Y FINALES DE LOS ALUMNOS. EL MEJOR VALOR PARA LAS NOTAS CONTINUAS Y FINALES APARECE EN NEGRITA EN CADA FILA.

Metric	Continua				Final			
	LS	RIDGE	SVR	KRR	LS	RIDGE	SVR	KRR
R	0.69	0.77	0.80	0.79	0.19	0.19	0.37	0.24
RMSE	2.51	2.02	1.91	1.95	1.28	1.25	1.21	1.26
MAE	1.76	1.60	1.49	1.54	0.92	0.91	0.84	0.90

medida para descubrir dependencias entre clases basadas en su dimensión metodológica. Recordamos que la dimensión metodológica se refiere a la actividad planificada por el profesor, consistente en última instancia en las tareas cargadas, enlaces compartidos como vídeos, etc. La tabla III resume en su columna Event Context la cantidad de material virtual creado por el profesor en la plataforma.

El NHSIC con kernel RBF requiere ajustar adecuadamente el parámetro σ que se refiere a la escala de varianza. En particular, los valores más bajos de σ implican una estimación basada más localmente que podría interpretarse como una estimación más no lineal. Valores más altos de este parámetro conducen a una estimación más amplia que puede interpretarse como más lineal (menos compleja). Seguimos las recomendaciones de la bibliografía para ajustar este parámetro σ . En particular, seguimos la regla empírica de fijar este parámetro en la raíz cuadrada de la mediana de las distancias

al cuadrado [13], que representa una estimación razonable de la dependencia entre las variables comparadas.

La figura 3 muestra la dependencia entre clases medida a través del NHSIC. En particular, utilizamos dos kernels diferentes para la comparación: en a) la función kernel de base radial (RBF) y en b) el kernel lineal. Mostramos los resultados de un procedimiento bootstrap sobre el número de estudiantes fijado en 25 y un total de 100 ensayos. Los resultados obtenidos se comparan con la configuración explicada en la Sección III, que informa de que las clases 4S y 4U siguen la misma metodología y deberían ser similares; la clase 4P debería ser similar a 4R, 4S y 4U. De la figura puede deducirse que NHSIC con el kernel RBF encuentra más dependencia entre las clases 4S y 4U que con el kernel lineal. Las clases 4P y 4R son descubiertas como dependientes por ambos kernels, más notablemente por el lineal. Recordemos que la mitad de las clases 4P y 4R comparten la misma metodología. La relación entre 4P y 4S también es notable, ya que comparten la metodología pero no el Profesor. La estimación lineal es capaz de detectar la estructura subyacente entre las clases, pero parece producir colas más grandes que su homóloga no lineal, ya que los intervalos de confianza a veces se solapan.

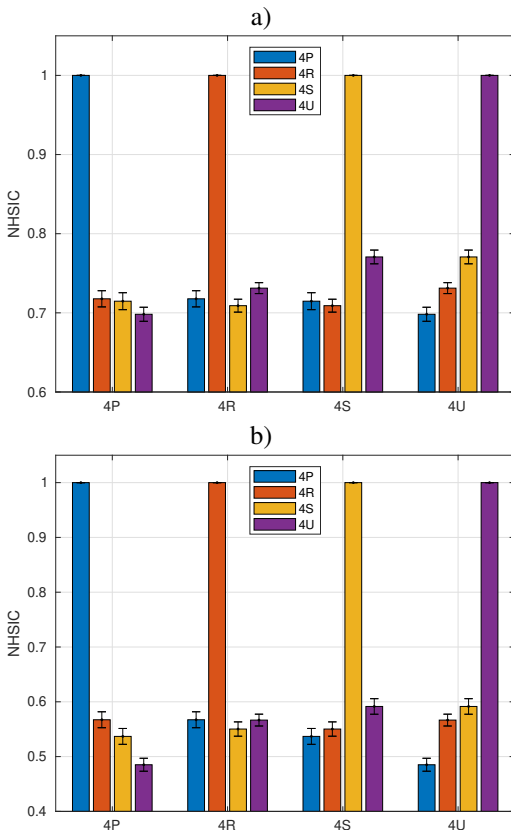


Fig. 3. Dependence among classes computed using: a) kernel RBF NHSIC, and b) linear NHSIC.

V. CONCLUSIONES & TRABAJO FUTURO

En este trabajo presentamos un estudio de las notas de predicción del alumnado en cuatro clases diferentes de la

misma asignatura pero con metodologías y profesores distintos. Hemos comprobado la validez de los métodos kernel como metodología para desarrollar tareas de regresión, hacer inferencias y realizar estimaciones de dependencia entre las metodologías. Toda la experimentación se ha diseñado siguiendo un procedimiento data-driven y en un proceso totalmente automático. Además, compartimos el código fuente que permite a la comunidad investigadora implementar y comparar los experimentos realizados en otros contextos.

En lo que respecta a la tarea de regresión, se procesaron los archivos de registro sin procesar del SGA y se utilizaron para realizar inferencias tanto de la evaluación continua como de las notas finales de los estudiantes. Llegamos a la conclusión de que la información de la plataforma es mucho más eficaz para predecir las notas de la evaluación continua que para predecir las notas finales. La razón es que el SGA almacena información sobre los alumnos que realizan su actividad a lo largo de la plataforma. Esta información adquirida es útil para medir en términos de tareas y asistencia lo que está más relacionado con la parte continua de la nota del alumnado. Viendo los resultados obtenidos en este estudio pero también en la literatura relacionada, la nota final es un escenario más complejo que implica el uso de variables que no se miden dentro del SGA.

Con respecto a la tarea de estimación de la dependencia, concluimos que el kernel NHSIC es capaz de encontrar las dependencias entre clases comparando sus dimensiones metodológicas. Es capaz de encontrar relaciones no lineales entre variables y exhibe las dependencias de forma razonable en comparación con la naturaleza de los datos experimentales propuestos. Siguiendo esta discusión, proponemos utilizar el NHSIC no lineal para realizar estimaciones de dependencia entre diferentes metodologías y descubrir clases y procedimientos de enseñanza similares. NHSIC fue capaz de mostrar las relaciones entre las clases siguiendo un enfoque basado en datos, que podría aplicarse para encontrar clases similares dentro de un SGA, y podría ser útil para recomendar la coordinación entre clases del mismo tema, por ejemplo.

Aunque las tareas mencionadas podrían considerarse bajo la misma teoría de aprendizaje estadístico, se basan en paradigmas diferentes. La primera, el rendimiento de los estudiantes, se basa en el aprendizaje supervisado cuando se dispone de etiquetas (en este caso, las etiquetas se refieren a las notas académicas de los estudiantes), ya que encuentra el modelo óptimo que corresponde a las características de los estudiantes con sus notas académicas. El segundo, la búsqueda de similitudes, se basa en un aprendizaje no supervisado en el que no se utilizan etiquetas. En su lugar, el estimador NHSIC encuentra las relaciones intrínsecas entre las metodologías utilizadas por los profesores en las distintas clases. El trabajo futuro incluye la creación de modelos de rendimiento del alumnado y la aplicación a las clases de metodologías similares descubiertas por NHSIC.

Como conclusión más general, encontramos que los métodos kernel son adecuados para realizar tareas de regresión sobre datos educativos, ya que son capaces de obtener modelos flexibles y no lineales que se ajustan bien a estos datos. Los métodos kernel también se emplearon en el problema

de estimación de la dependencia al tratar datos educativos. Por último, mientras que estos métodos son adecuados para la predicción del rendimiento de las notas de evaluación continua a través del SGA, la predicción de las notas finales constituye un problema más difícil que debe tener en cuenta más variables en comparación con la configuración del problema descrito.

REFERENCES

- [1] V. Abella García, M. Grande de Prado, B. Rodríguez Bravo, F.J. García Peñalvo, and A. Corell Almuzara. Guía de recomendaciones para la evaluación online de las universidades públicas de castilla y león. Technical report, University of Salamanca, 2020.
- [2] Lubna Mahmoud Abu Zohair. Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(1):27, Aug 2019.
- [3] Rebeca Cerezo Menéndez Alejandro Bogarín Vega, Cristóbal Romero Morales. Aplicando minería de datos para descubrir rutas de aprendizaje frecuentes en moodle. *EDMETIC*, 5(1):73–92, 2016.
- [4] Ammar Almasri, Rami S. Alkhalwaldeh, and Erbuğ Çelebi. Clustering-based emt model for predicting student performance. *Arabian Journal for Science and Engineering*, 45(12):10067–10078, Dec 2020.
- [5] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113:177–194, 2017.
- [6] C. Baker. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289, 1973.
- [7] Javier Bravo-Agapito, Sonia J. Romero, and Sonia Pamplona. Early prediction of undergraduate student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, 115:106595, 2021.
- [8] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research*, 13(28):795–828, 2012.
- [9] P. Cortez and Silva A. Using data mining to predict secondary school student performance. In A. Brito and Teixeira J., editors, *Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008)*, pages 5–12, 2008.
- [10] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In M. C. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, pages 155–161. MIT Press, 1997.
- [11] Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*, 5:73–99, 2004.
- [12] Diego Garcia-Saiz and Marta Zorrilla. Comparing classification methods for predicting distance students' performance. In *Proceedings of the Second Workshop on Applications of Pattern Analysis*, pages 26–32. JMLR Workshop and Conference Proceedings, 2011.
- [13] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In Sanjay Jain, HansUlrich Simon, and Etsuji Tomita, editors, *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2005.
- [14] Arthur Gretton, Ralf Herbrich, Alexander Smola, Olivier Bousquet, and Bernhard Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6(70):2075–2129, 2005.
- [15] F. D. Guillén-Gámez and I. García-Magariño. A technique for designing glossary activities with facial authentication. *Journal of E-Learning and Knowledge Society*, 11(1):125–138, 2015.
- [16] Thomas Lerche and Ewald Kiel. Predicting student achievement in learning management systems by log data analysis. *Computers in Human Behavior*, 89:367–372, 2018.
- [17] Javier López-Zambrano, Juan A Lara, and Cristóbal Romero. Towards portability of models for predicting students' final performance in university courses starting from moodle logs. *Applied Sciences*, 10(1):354, 2020.
- [18] Adrián Pérez-Suay and Gustau Camps-Valls. Sensitivity maps of the hilbert-schmidt independence criterion. *Applied Soft Computing*, 70:1054–1063, 2018.
- [19] Alberto Rivas, Alfonso González-Briones, Guillermo Hernández, Javier Prieto, and Pablo Chamoso. Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, 423:713–720, 2021.
- [20] Cristobal Romero and Sebastian Ventura. Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1355, 2020.
- [21] Cristóbal Romero, Manuel-Ignacio López, Jose-María Luna, and Sebastián Ventura. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68:458–472, 2013.
- [22] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. In David Helmbold and Bob Williamson, editors, *Computational Learning Theory*, pages 416–426, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- [23] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.
- [24] Yao-Ting Sung, Kuo-En Chang, and Tzu-Chien Liu. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94:252–275, 2016.
- [25] D. T. Tempelaar, B. Rienties, and Q. Nguyen. Towards actionable learning analytics using dispositions. *IEEE Transactions on Learning Technologies*, 10(1):6–16, Jan 2017.
- [26] Sergios Theodoridis. Learning in parametric modeling: Basic concepts and directions. In Sergios Theodoridis, editor, *Machine Learning (Second Edition)*. Academic Press, second edition edition, 2020.
- [27] Darren Turnbull, Ritesh Chugh, and Jo Luck. Learning management systems: An overview. In Arthur Tatnall, editor, *Encyclopedia of Education and Information Technologies*, pages 1–7. Springer International Publishing, Cham, 2019.
- [28] Steven Van Vaerenbergh and Adrián Pérez-Suay. A classification of artificial intelligence systems for mathematics education. In Philippe R. Richard, M. Pilar Vélez, and Steven Van Vaerenbergh, editors, *Mathematics Education in the Age of Artificial Intelligence: How Artificial Intelligence can Serve Mathematical Human Learning*, pages 89–106. Springer International Publishing, Cham, 2022.
- [29] Hien M. Vo, Chang Zhu, and Nguyet A. Diep. The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, 53:17–28, 2017.
- [30] Fan Yang and Frederick W.B. Li. Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education*, 123:97–108, 2018.
- [31] Marta Zorrilla and Diego García-Saiz. A service oriented architecture to provide data mining services for non-expert data miners. *Decision Support Systems*, 55(1):399–411, 2013.



Adrián Pérez-Suay es licenciado en Matemáticas (2007), máster en Computación Avanzada y Sistemas Inteligentes (2010) y doctor en Matemática Computacional e Informática (2015) por la Universitat de València. Es profesor ayudante en el Departamento de Educación Matemática de la Universitat de València.



Steven Van Vaerenbergh (M¹¹-SM¹⁵) se licenció en Ingeniería Eléctrica por la Universidad de Gante, Bélgica, en 2003, y se doctoró por la Universidad de Cantabria, España, en 2010, donde actualmente es profesor ayudante en el Departamento de Matemáticas, Estadística y Computación. Su investigación se centra en el uso de la inteligencia artificial en la enseñanza de las matemáticas, en particular en los entornos tecnológicos de aprendizaje, la geometría dinámica y el aprendizaje individualizado.



Pascual D. Diago es profesor asociado del Departamento de Didáctica de las Matemáticas de la Universitat de València. Actualmente imparte docencia en los grados de Educación Infantil, Educación Primaria y en el máster de investigación en Didácticas Específicas. Su investigación se centra en la resolución de problemas y el uso de las nuevas tecnologías en los procesos de enseñanza-aprendizaje de las matemáticas escolares.



Ana B. Pascual-Venteo licenciada en Matemáticas, es candidata al doctorado en el Laboratorio de Procesamiento de Imágenes, donde trabaja como investigadora. Además, es profesora del Departamento de Matemáticas, todo ello de la Universitat de València.



Francesc J. Ferri Licenciado en Física (Electricidad, Electrónica e Informática) en 1987 y Doctor en Reconocimiento de Patrones en 1993, ambos por la Universitat de València. Trabaja en el Departamento de Informática de la Universitat de València desde 1986. El Dr. Ferri es autor o coautor de más de 150 artículos técnicos en conferencias internacionales y revistas de reconocido prestigio en sus campos de interés. Sus intereses de investigación actuales incluyen la selección de características, los métodos de clasificación no paramétricos, el aprendizaje automático, la visión por ordenador, la recuperación de imágenes y diferentes aspectos de la escucha automática.