

PIXEL BIT

Nº 60 ENERO 2021
CUATRIMESTRAL

e-ISSN:2171-7966
ISSN:1133-8482

Revista de Medios y Educación



P
I
X
E
L

B
I
T



PIXEL-BIT

REVISTA DE MEDIOS Y EDUCACIÓN

Nº 60 - ENERO - 2021

<https://revistapixelbit.com>



EDITORIAL
UNIVERSIDAD DE SEVILLA

EQUIPO EDITORIAL (EDITORIAL BOARD)

EDITOR JEFE (EDITOR IN CHIEF)

Dr. Julio Cabero Almenara, Departamento de Didáctica y Organización Educativa, Facultad de CC de la Educación, Universidad de Sevilla (España).

EDITOR ADJUNTO (ASSISTANT EDITOR)

Dr. Juan Jesús Gutiérrez Castillo, Departamento de Didáctica y Organización Educativa. Facultad de CC de la Educación, Universidad de Sevilla (España).

Dr. Óscar M. Gallego Pérez, Secretariado de Recursos Audiovisuales y NN.TT., Universidad de Sevilla (España)

CONSEJO DE REDACCIÓN

EDITOR

Dr. Julio Cabero Almenara. Universidad de Sevilla (España)

EDITOR ASISTENTE

Dr. Juan Jesús Gutiérrez Catillo. Universidad de Sevilla. (España)

Dr. Óscar M. Gallego Pérez. Universidad de Sevilla (España)

EDITORES ASOCIADOS

Dra. Urtza Garay Ruiz, Universidad del País Vasco. (España)

Dra. Ivanovna Milqueya Cruz Pichardo, Pontificia Universidad Católica Madre y Maestra. (República Dominicana)

VOCALES

Dra. María Puig Gutiérrez, Universidad de Sevilla. (España)

Dra. Sandra Martínez Pérez, Universidad de Barcelona (España)

Dr. Selín Carrasco, Universidad de La Punta (Argentina)

Dr. Jackson Collares, Universidades Federal do Amazonas (Brasil)

Dra. Kitty Gaona, Universidad Autónoma de Asunción (Paraguay)

Dra. Elvira Esther Navas, Universidad Metropolitana de Venezuela (Venezuela)

Dr. Angel Puentes Puente, Pontificia Universidad Católica Madre y Maestra. Santo Domingo (República Dominicana)

Dr. Fabrizio Manuel Sirignano, Università degli Studi Suor Orsola Benincasa (Italia)

CONSEJO TÉCNICO

Edición, maquetación: Manuel Serrano Hidalgo, Universidad de Sevilla (España)

Dra. Raquel Barragán Sánchez, Universidad de Sevilla (España)

Antonio Palacios Rodríguez, Universidad de Sevilla (España)

Diseño de portada: Lucía Terrones García, S.A.V, Universidad de Sevilla (España)

Revisor/corrector de textos en inglés: Rubicelia Valencia Ortiz, MacMillan Education (México)

Revisores metodológicos: evaluadores asignados a cada artículo

Responsable de redes sociales: Manuel Serrano Hidalgo, Universidad de Sevilla (España)

Administración: Leticia Pinto Correa, S.A.V, Universidad de Sevilla (España)

CONSEJO CIENTÍFICO

Jordi Adell Segura, Universidad Jaume I Castellón (España)

Ignacio Aguaded Gómez, Universidad de Huelva (España)

María Victoria Aguiar Perera, Universidad de Las Palmas de Gran Canaria (España)

Olga María Alegre de la Rosa, Universidad de la Laguna Tenerife (España)

Manuel Área Moreira, Universidad de la Laguna Tenerife (España)

Patricia Ávila Muñoz, Instituto Latinoamericano de Comunicación Educativa (México)

Antonio Bartolomé Pina, Universidad de Barcelona (España)

Angel Manuel Bautista Valencia, Universidad Central de Panamá (Panamá)

Jos Beishuizen, Vrije Universiteit Amsterdam (Holanda)
Florentino Blázquez Entonado, Universidad de Extremadura (España)
Silvana Calaprice, Università degli studi di Bari (Italia)
Selín Carrasco, Universidad de La Punta (Argentina)
Raimundo Carrasco Soto, Universidad de Durango (México)
Rafael Castañeda Barrena, Universidad de Sevilla (España)
Zulma Cataldi, Universidad de Buenos Aires (Argentina)
Manuel Cebrián de la Serna, Universidad de Málaga (España)
Luciano Cecconi, Università degli Studi di Modena (Italia)
Jean-François Cerisier, Université de Poitiers, Francia
Jordi Lluís Coiduras Rodríguez, Universidad de Lleida (España)
Jackson Collares, Universidades Federal do Amazonas (Brasil)
Enricomaria Corbi, Università degli Studi Suor Orsola Benincasa (Italia)
Marialaura Cunzio, Università degli Studi Suor Orsola Benincasa (Italia)
Brigitte Denis, Université de Liège (Bélgica)
Floriana Falcinelli, Università degli Studi di Perugia (Italia)
María Cecilia Fonseca Sardi, Universidad Metropolitana de Venezuela (Venezuela)
Maribel Santos Miranda Pinto, Universidade do Minho (Portugal)
Kitty Gaona, Universidad Autónoma de Asunción (Paraguay)
María-Jesús Gallego-Arrufat, Universidad de Granada (España)
Lorenzo García Aretio, UNED (España)
Ana García-Valcarcel Muñoz-Repiso, Universidad de Salamanca (España)
Antonio Bautista García-Vera, Universidad Complutense de Madrid (España)
José Manuel Gómez y Méndez, Universidad de Sevilla (España)
Mercedes González Sanmamed, Universidad de La Coruña (España)
Manuel González-Sicilia Llamas, Universidad Católica San Antonio-Murcia (España)
Ángel Pio González Soto, Universidad Rovira i Virgili, Tarragona (España)
António José Meneses Osório, Universidade do Minho (Portugal)
Carol Halal Orfali, Universidad Tecnológica de Chile INACAP (Chile)
Mauricio Hernández Ramírez, Universidad Autónoma de Tamaulipas (México)
Ana Landeta Etxeberria, Universidad a Distancia de Madrid (UDIMA)
Linda Lavelle, Plymouth Institute of Education (Inglaterra)
Fernando Leal Ríos, Universidad Autónoma de Tamaulipas (México)
Paul Lefrere, Cca (UK)
Carlos Marcelo García, Universidad de Sevilla (España)
Francois Marchessou, Universidad de Poitiers, París (Francia)
Francesca Marone, Università degli Studi di Napoli Federico II (Italia)
Francisco Martínez Sánchez, Universidad de Murcia (España)
Ivory de Lourdes Mogollón de Lugo, Universidad Central de Venezuela (Venezuela)
Angela Muschitiello, Università degli studi di Bari (Italia)
Margherita Musello, Università degli Studi Suor Orsola Benincasa (Italia)
Elvira Esther Navas, Universidad Metropolitana de Venezuela (Venezuela)
Trinidad Núñez Domínguez, Universidad de Sevilla (España)
James O'Higgins, de la Universidad de Dublín (UK)
José Antonio Ortega Carrillo, Universidad de Granada (España)
Gabriela Padilla, Universidad Autónoma de Tamaulipas (México)
Ramón Pérez Pérez, Universidad de Oviedo (España)
Angel Puentes Puentes, Pontificia Universidad Católica Madre y Maestra. Santo Domingo (República Dominicana)
Julio Manuel Barroso Osuna, Universidad de Sevilla (España)
Rosalía Romero Tena, Universidad de Sevilla (España)
Hommy Rosario, Universidad de Carabobo (Venezuela)
Pier Giuseppe Rossi, Università di Macerata (Italia)
Jesús Salinas Ibáñez, Universidad Islas Baleares (España)
Yamile Sandoval Romero, Universidad de Santiago de Cali (Colombia)
Albert Sangrà Morer, Universidad Oberta de Catalunya (España)
Ángel Sanmartín Alonso, Universidad de Valencia (España)
Horacio Santángelo, Universidad Tecnológica Nacional (Argentina)
Francisco Solá Cabrera, Universidad de Sevilla (España)
Jan Frick, Stavanger University (Noruega)
Karl Steffens, Universidad de Colonia (Alemania)
Seppo Tella, Helsinki University (Finlandia)
Hanne Wacher Kjaergaard, Aarhus University (Dinamarca)



FACTOR DE IMPACTO (IMPACT FACTOR)

SCOPUS (CiteScore Tracker 2020: 1,8)- FECYT: Ciencias de la Educación. Cuartil 1. Posición 16. Puntuación: 39,80- DIALNET MÉTRICAS (Factor impacto 2019: 1,336. Q1 Educación. Posición 12 de 226) ERIH PLUS - Clasificación CIRC: B- Categoría ANEP: B - CARHUS (+2018): B - MIAR (ICDS 2019): 9,9 - Google Scholar (global): h5: 23; Mediana: 44 - Criterios ANECA: 20 de 21

Píxel-Bit, Revista de Medios y Educación está indexada entre otras bases en: SCOPUS, Fecyt, Iresie, ISOC (CSIC/ CINDOC), DICE, MIAR, IN-RECS, RESH, Ulrich's Periodicals, Catálogo Latindex, Biné-EDUSOL, Dialnet, Redinet, OEI, DOCE, Scribd, Redalyc, Red Iberoamericana de Revistas de Comunicación y Cultura, Gage Cengage Learning, Centro de Documentación del Observatorio de la Infancia en Andalucía. Además de estar presente en portales especializados, Buscadores Científicos y Catálogos de Bibliotecas de reconocido prestigio, y pendiente de evaluación en otras bases de datos.

EDITA (PUBLISHED BY)

Grupo de Investigación Didáctica (HUM-390). Universidad de Sevilla (España). Facultad de Ciencias de la Educación. Departamento de Didáctica y Organización Educativa. C/ Pirotecnica s/n, 41013 Sevilla.
 Dirección de correo electrónico: revistapixelbit@us.es . URL: <https://revistapixelbit.com/>
 ISSN: 1133-8482; e-ISSN: 2171-7966; Depósito Legal: SE-1725-02
 Formato de la revista: 16,5 x 23,0 cm

Los recursos incluidos en Píxel Bit están sujetos a una licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Unported (Reconocimiento-NoComercial-CompartirIgual)(CC BY-NC-SA 4.0), en consecuencia, las acciones, productos y utilidades derivadas de su utilización no podrán generar ningún tipo de lucro y la obra generada sólo podrá distribuirse bajo esta misma licencia. En las obras derivadas deberá, asimismo, hacerse referencia expresa a la fuente y al autor del recurso utilizado.


©2021 Píxel-Bit. No está permitida la reproducción total o parcial por ningún medio de la versión impresa de la Revista Píxel- Bit.

índice

- 1.- Una herramienta tangible para facilitar procesos de diseño y análisis didáctico** // A tangible tool to facilitate learning design and analysis discussions: Translation and cross-cultural adaptation of the ACAD Toolkit.
Peter Goodyear, Lucila Carvalho, Pippa Yeo-man, Linda Castañeda, Jordi Adell **7**
- 2.- The acquisition of ICT skills at the university level: the case of the Faculty of Business Studies and Tourism of the University of Huelva** // La adquisición de competencias TIC en el ámbito universitario: el caso de la Facultad de Ciencias Empresariales y Turismo de la Universidad de Huelva (**Bilingüe**)
Alfonso Infante-Moro, Juan C. Infante-Moro, Julia Gallardo-Pérez **29**
- 3.- Diseño y validación de un instrumento para la taxonomía de los robots de suelo en Educación Infantil** // Design and validation of an instrument for the taxonomy of floor robots in Early Childhood Education
Juan Francisco Álvarez Herrero **59**
- 4.-Relaciones entre redes sociales y recursos educativos digitales en la universidad: comparativa España – Colombia** // Relationship between social media and digital resources of instruction in the university: comparative Spain – Colombia
Luis Matosas López, Marianela Luzardo-Briceño, Alba-Soraya Aguilar-Jiménez, Ludym Jaimes-Carrillo **77**
- 5.- Revisión de la producción científica sobre MOOC entre 2016 y 2019 a través de SCOPUS** // A review of the scientific production on MOOCs from 2016 to 2019 using SCOPUS
Julio Ruiz-Palmero, Daniel López-Álvarez, Enrique Sánchez-Rivas **95**
- 6.- Active aging and internet use to improve the quality of life of the seniors** // Envejecimiento activo y uso de internet para mejorar la calidad de vida de las personas mayores (**Bilingüe**)
Pedro Román-Graván, Manuel Pérez-Hurtado, Pedro Tadeu **109**
- 7.- The content posting practices of young people on social networks** // Prácticas adolescentes de publicación de contenidos en redes sociales
José Luis Rodríguez-Illera, Francesc Martínez-Olmo, María José Rubio-Hurtado, Cristina Galván-Fernández **135**
- 8.- Las Competencias en el uso de las Tecnologías de la Información y la Comunicación en el alumnado universitario**// Competences in the use of Information Technologies and Communication in university students
Eva Ordóñez-Olmedo, Esteban Vázquez-Cano, Samuel Arias-Sánchez, Eloy López-Meneses **153**
- 9.- The Influence of Students' Gender on the Use of Virtual Campuses. A Case Study** // La influencia del sexo en el uso de los campus virtuales. Estudio de caso (**Bilingüe**)
Daniel David Martínez Romera, Manuel Cebrián de la Serna, Gloria Priego de Montiano **169**
- 10.- Uso del smartphone en jóvenes universitarios: una oportunidad para el aprendizaje** // Smartphone use in university students: An opportunity for learning
Alberto Dafonte Gómez, Marcelo Fabián Maina, Oswaldo García Crespo **211**

The Influence of Students' Gender on the Use of Virtual Campuses. A Case Study

La influencia del sexo en el uso de los campus virtuales. Estudio de caso

 **Dr. Daniel David Martínez Romera**
Profesor ayudante doctor. Universidad de Cádiz. España

 **Dr. Manuel Cebrián de la Serna**
Catedrático de Universidad. Universidad de Málaga. España

 **Dra. Gloria Priego de Montiano**
Profesora ayudante doctor. Universidad de Córdoba. España

Recibido: 2020/03/17 Revisado: 2020/05/05 Aceptado: 2020/10/10 Preprint: 2020/12/08 Publicado: 2021/01/01

ABSTRACT

A comparative analysis has been carried out in order to study the similarities and differences in the use and management of space and time by students, based on their gender. The study analyses the data records generated by students from three different degree courses in two universities, throughout four academic years. The methodology includes exploratory statistical analysis and learning analytics for the detection of spatial, temporal and behavioural patterns. Results show a consistent convergence in most cases, although they also show marginal behavioral trends, both for the days of the week, the hours of the day, and for the different contents of the virtual campuses. This is more evident in its spatial aspect, which highlights some clear differences in the processes of distribution and concentration of the events under study: at least in this case, women do not act in the same way as men. Ultimately, the study proposes new forms of synergy between educational work and the application of the Social Sciences' disciplinary contents thus strengthening the transfer of knowledge from specific didactics on both educational curriculum and teacher training.

RESUMEN

Se estudian los registros de datos generados por estudiantes de tres titulaciones distintas de dos universidades, en cuatro cursos académicos, a través de sus campus virtuales. La metodología ha considerado el análisis estadístico exploratorio y las analíticas de aprendizaje, especialmente la minería de datos, para la detección de patrones de espaciales, temporales y de comportamiento. Los resultados han mostrado una convergencia consistente en la mayoría de casos, si bien ha evidenciado tendencias marginales de comportamiento para los días de la semana, las horas del día, y distintos contenidos de los campus virtuales. Ha permitido discriminar algunas diferencias en los procesos de distribución y concentración espacial: las mujeres no se mueven de la misma forma que los hombres y presentan diferencias secundarias en la gestión del tiempo. En última instancia, el estudio plantea nuevas formas de sinergia entre el trabajo educativo y la aplicación de los contenidos disciplinares de las Ciencias Sociales; fortaleciendo así la transferencia del conocimiento desde las didácticas específicas tanto sobre el currículo educativo como en la formación del profesorado.

PALABRAS CLAVES · KEYWORDS

Higher education; Student behaviour; Social Sciences; Comparative analysis; Data processing
Enseñanza superior; Comportamiento del alumno; Ciencias Sociales; Análisis comparativo; Procesamiento de datos

1. Introduction

When, in 2002, Martin Dougiamas, unhappy with the available software and motivated by the desire to incorporate pedagogical constructivism into it, released Moodle, i.e. the first version of his LCMS (Learning Content Management System), there were no indications that would become a de facto standard, thus putting virtual digital support at the forefront of education, especially at university level.

Nowadays, whether open source or not, it is unthinkable for a university subject, and also increasingly so in secondary education, not to have a virtual campus that is used as a preferential or exclusive way of work. This makes even more sense considering the new formats of educational analysis offered by this platform (Hamotoglu et al., 2020; Cantabella et al., 2019; Quintanas Mendes et al., 2019; Fakir & Touya, 2014; Lu & Law, 2012), the analytical techniques that can be used (Álvarez Vázquez et al, 2020; Charanya & Kesavan, 2019; Cohen, Rahimi & Zilka, 2019; Manne, Yelisetti, Kakarla, & Fátima, 2014) and the social and behavioural dimensions that can be addressed (Pakanen, Alavesa, Arhippainen & Ojala, 2020; Castro, Menacho & Pérez-Molina, 2018; Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017; Parise, 2016).

In our opinion, the two most relevant educational advantages are: the teaching spatial-temporal delocalisation, be it partial or total; and those originating from the interaction through technology, both among users and with the contents of the subject.

With regards to the former, it is clear that the man-machine symbiosis that was predicted at the beginning of computing (Licklider, 1960), is constantly increasing. Although the effects have not yet been well explored, they have been perceived for some time in Education (Orostica Verdugo, 2020; Desai, Chavan & Tendulkar, 2020; Cabanillas, Luengo & Carvalho, 2019; Morales, Infante-Moro & Gallardo-Pérez, 2019; Ballesteros Regaña et. al, 2010; Correa & Paredes, 2009), Geography (Capel, 2009) and other Social Sciences (Sarduy Domínguez et al. 2020; Arias & Defiore, 2019; Gómez Collado, Contreras Orozco & Gutiérrez Linares, 2016). The school timetable and calendar, i.e. the class time and weekly distribution of the teaching programme, is still organised with the same rationale that was used prior to the current technological reality. The traditional spatial-temporal concurrence of the three elements of the educational system (teacher, student and object of study) can be delocalised with LCMSs, making it very flexible. This means perceiving it as a new object of study, imbued with a chrono geographic (Tartia, 2020; Thulin, Vihelmsen & Scwanen, 2020; Vega Valverde, 2017) and behavioural complexity clearly different to the one devised during the institutionalisation of education of the 19th century and even to the current one which is regulated by subsequent regulatory frameworks.

As for the latter, teachers are still largely unaware of the new technical tools that can be used to teach and work together on a school subject. Be it the use of cloud-based collaborative tools, or the possibility of consulting learning material at any time and place, the truth is that the way students interact with content and tasks has nothing to do with what was possible, and available, just one or two generations ago.

Spain's entry into the European Higher Education Area (BOE, Spain's Official Bulletin, no. 224 of 18 September 2003) highlighted the importance of these issues. The introduction of the ECTS (European Credit Transfer System) credits was an attempt to calculate not only the traditional learning time in class, but also the hours of study, tutorials, seminars, work and any other dedicated time necessary or required to pass a subject.

Controlling the progress of students outside the classroom does not seem to be a reasonable or desirable solution. Also, focusing on objective results (exam notes, presentations, assignments, etc.) casts doubts on the importance of the process of continuous training and teacher-student feedback required by all modern educational system. Teachers are often questioned about these processes; hence they are now required to have additional skills (Sola-Martinez et al., 2020; Abbott, 2011). This is especially true from the perspective of learning analytics which is what this work is based on, given its versatility (Ruipérez-Valiente et al. 2020; Papoušek, Pelánek & Stanislav, 2016; Sclater, Peasgood & Mullan, 2016; Gašević, Dawson & Siemens, 2015; Siemens & Long, 2011)

This is why there is a need to analyse students' (digital) behaviour patterns on virtual campuses as an indirect way of finding out about their learning performance and spatial-temporal behaviour. This is not necessarily an invasive process if basic ethical considerations are set out in line with the humanist approaches already established in Geography (Tuan, 1977 & 1990).

Each identified user, while on a virtual campus, generates a multitude of data records that provide information on the place, time and actions (opening a link, a presentation, a document, writing a comment on a forum, handing in an assignment, etc.). Each piece of record is called an event (Caballé & Xhafa, 2013). This is very interesting information to which it is possible to add other variables such as gender taken from the user's ID data, thus enriching the content.

The data can be analysed starting from a modest statistical base, to then incorporate the latest techniques and methodologies, thanks to which the work and the result achievement schedules are clearly feasible when it comes to evaluating or even rectifying the didactic strategy (Badilla Quintana, Vera Sagredo & Lytras, 2017; Minguillón, Santanach & Appel, 2016).

In view of the above, the goal is carrying out a comparative analysis between the sexes in order to check the potential similarities and differences in the use of virtual campuses. This will be done from three perspectives: the time of use, the place of use and the content visited. This is in line with other research that points to the relevance of exploring this type of variables crossover to adequately understand student behaviour on virtual campuses (Álvarez Vázquez, 2019; Cabanillas García et al., 2019; De la Iglesia-Villasol, 2019).

2. Methodology

This study analyses the data collected from more than 1000 students over a 4-year period, between the 2016/17 and 2019/20 academic years. The students were divided into 14 groups belonging to 4 undergraduate and postgraduate subjects of three degrees in 2 Spanish universities. Three of the subjects belong to the field of Social Sciences Didactics and a fourth one to School Didactics and Organisation:

- Didactic and Technological Resources in Education in Pedagogy Degree (A1 - P-RDTE)
- Geography and History Curriculum in Master's Degree in Teaching (B1 - M-CCSS-C)
- Design and Development of Geography and History Programmes and Training Activities in the Teaching Master's Degree (B2 - M-CCSS-D)
- Didactics of Social Sciences in Infant Education (C1 - I-DCCSS)

Table 1

Events generated by subject and group

Teacher	Subject Area	(Group) Subject	ID	Year	Ab (Af: M/F)	Events
A	A1	P-RDTE (2)	A1	2017/18	111 (111: 17 / 94)	27.316
		P-RDTE (1)	A2	2018/19	58 (57: 5 / 52)	17.036
		P-RDTE (1)	A3	2018/19	59 (58: 11 / 47)	16.208
		P-RDTE th(1)	A4	2019/20	60 (59: 10 / 49)	23.236
		P-RDTE (1)	A5	2019/20	52 (51: 7 / 44)	10.928
B	B1	M-CCSS-C (1)	B1	2018/19	72 (72: 44 / 28)	5.775
		M-CCSS-C (1)	B2	2019/20	41 (39: 21 / 18)	3.451
	B2	M-CCSS-D(1)	B3	2018/19	72 (70: 23 / 47)	9.942
		M-CCSS-D (1)	B4	2019/20	41 (39: 21 / 18)	7.728
C	C1	I-DCCSS (1)	C1	2016/17	166 (166: 5 / 161)	24.003
		I-DCCSS (1)	C2	2017/18	164 (164: 13 / 151)	24.458
		I-DCCSS (1)	C3	2018/19	188 (188: 118 / 180)	22.899
				Σ	1084 (1074: 295 / 889)	192.980
				\bar{x}	87,36	15.060

Caption: A_b indicates the number of initial students; A_r after cleansing and normalisation. M/F: breakdown by gender (Male / Female).

Table 1 contains the fundamental descriptors of each subject (year, students, and total number of events), in order to establish the dimensions of the reference scenario. Nevertheless, given the internal heterogeneity of the variables (students and events), a proportional form of comparison (evaluations / students) must be used in order to define the relevance of each interaction with the platform when working on a subject. In this way, and

using teachers as a sorting criterion, the result would be: A (281.4), C (140.7) and B (119.0); with a joint average of 180.4 events per subject and person.

Using these data, an attempt is made to detect behavioural patterns among the different groups, as well as the individual characteristics, by taking into account the nature of the contents researched, the moment the consultation takes place and whether the gender has any influence on this behaviour (Martínez Romera, 2017). This is therefore a Big Data analytical context, part of Learning Analytics in education (Siddiqa et al. 2016; Tsai, Lai, Chao & Vasilakos, 2015; Sin and Muthu, 2015).

An important aspect, and one common to all subjects, is that the majority of the students are females, especially in the subject of Early Childhood Education, something that must be taken into account in the internal and external comparison processes. In this study, it implies the normalisation of the variables according to the sex, before comparing them.

Data mining has proved to be a useful tool, especially in educational contexts where it is produced in large quantities (Koedinger et al., 2015; Sin & Muthu, 2015). This is clearly the case in the situation under study, and in general in teaching-learning processes supported by virtual environments. Therefore, two applications will be used for data analysis:

Orange (v3.19), an open source software for the application of data mining and learning machines successfully used on large volumes of data, such as the one under study (Demsar et al., 2013; Demsar et al., 2004). The user relationship model is the visual programming language (VPL), which makes it especially interesting in social or educational research contexts where computer or statistical-mathematical skills are unique, especially in Geography (Dobesova, 2011).

Weka (v3.8), a software which focuses on the same analytical areas as the previous one, but with a higher degree of specialisation in terms of the number and diversity of implemented and implementable algorithms (Sebastian & Puthiyidam, 2015; Ahmadi & Ahmad, 2013). It is more procedural and less interactive, but its design is suitable for higher data volumes under the same technical (hardware) conditions as Orange.

Because of the differences in the configuration of Moodle (the LCMS that was used to build all the virtual campuses of the universities involved in this study) and the need to carry out an initial data control, a cleansing and normalisation process is carried out beforehand. This is done using a database in which data is manipulated using SQL, regular expressions, edit distance, and data mining. The overall process is defined by the knowledge discovery in databases (KDD) strategy, and involves a sequence of chained tasks: selection, pre-processing (cleansing), transformation, mining and interpretation (García et al., 2016; Ramírez-Gallego et al., 2016; Taleb, Dssouli & Serhani, 2015). The following step is the selection and re-classification of variables, as well as the evaluation of their possible use in future analysis.

Five variables are created from other pre-existing ones in order to provide functionality to the analyses: genders are assigned, time structures are simplified (hour and day of the week) and the modal values of the events generated are reorganised. In addition, each set of events for the same subject and academic year receives a unique identifier (Subject). The results of the process prior to the start of the analysis are shown in Table 2.

Table 2

Cleansing, standardisation of variables and final reference scenario

Variable	Presence (Teacher)	Standardised Values		Comment (Standardisation)
Day	ABC	1391	7	New variable: Day of the week (Monday to Sunday)
Time	ABC	20644	24	New variable: Time of day (0-23h).
Full name of participant	ABC	186642	186642	New variable: Sex (male of female)
Participant involved	AB	186642	186642	Reflective, does not point to third parties. Little interest.
Context of the event.	ABC	824	13	New variable: Context (1-13) with values added according to frequency. Elimination of redundancies and null records.
Component	AB	21	19	Minor standardisation required
Name of the event	ABC	145	24	New variable: Event (1-24) Elimination of redundancies and null records.
Description	AB	186642	186642	Name of the event + Moodle internal user ID. A non-downloadable ID invalidates its cleansing. Little interest.
Origin	AB	3	3	Semantics of the variable with little interest for this study. Virtually all records acquire the value 'web'.
IP address	ABC	12892	12892	Useful for geo-referencing and route analysis (chronogeography).
<i>Final students</i>	<i>1074</i>	<i>(99.1% of the total)</i>		
<i>Final events</i>	<i>182743</i>	<i>(99.1% of the total)</i>		

The technological support required and has been completed with the use of specialised hardware for massive calculations, thanks to the supercomputing services of the Information Systems Area of the parent university. The methods and techniques used for the development of the research have been diverse, and applied according to the phase of analysis:

1. Data collection: downloads from Moodle were compiled with spreadsheets, then entered into a database.
2. Cleansing and normalisation: advanced editing filters, SQL language, edit distances, regular expressions, and manual editing in borderline cases.
3. Data manipulation: SQL and Python languages.
4. IP decoding through renowned specialized services (IPInfo.io).

5. Data mining: descriptive statistics and selection of learning automata for data mining based on ideas developed by Kaur, Singh & Josan (2015).

3. Analysis

3.1. Weekly Structure

Group A subjects generally show a clear concentration on Tuesdays, which is a school day. The behaviour by gender at this time is quite similar. The days before and after offer modest values, but somewhat higher than in the rest of the week, during which the time of use is scarce, below 2-3%. Only in A4 and A5 is there a clear increase in the previous days (Saturday to Monday) and a more modest one for the rest of the days.

In the subjects of the first academic year the data show occasional deviations, but in the whole of the week they tend to converge in brackets of 2-3 percentage points; whereas, during the last year, females students clearly work more frequently every day of the week, except Saturdays and Sundays, when there is a significant reversal. Sunday is the most relevant day for this with 21.53% male to 9.14% female interactions.

Group B subjects were taught consecutively on the same days (Monday and Tuesday). B2 had an additional session on Wednesdays. The correspondence with class days is noticeable in almost all academic years, except for the third, where there is a peak in activities on Thursdays. In any case, a downward trend is observed as the week goes by, until the day before classes. The time distribution is more homogeneous than in the previous case (with peaks above 60 and 70%), with maximums always below 25% and minimums above 6%.

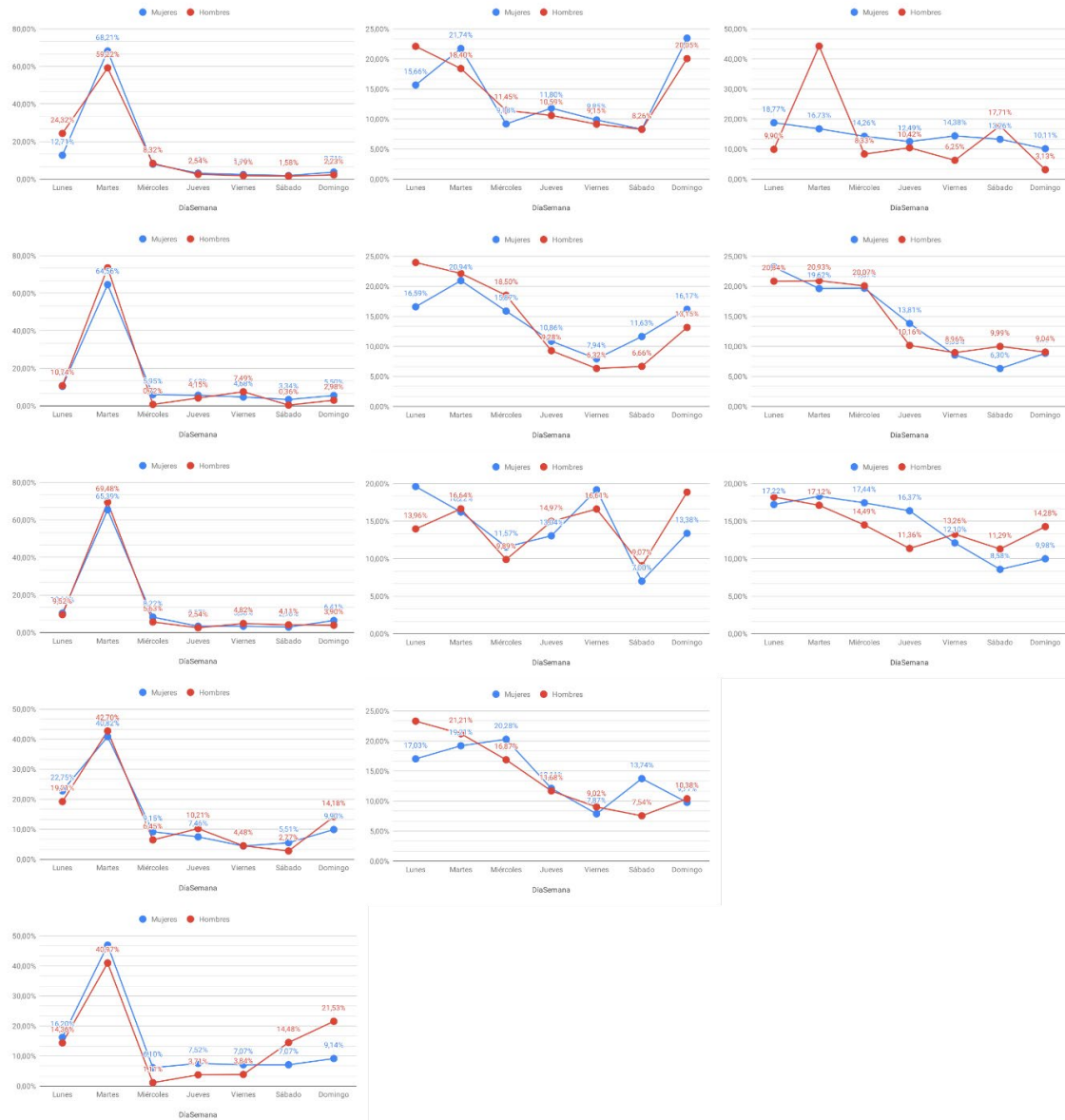
The behavioural patterns by gender reveal occasional differences, however, it is not possible to detect clear trends: it is more likely that the relative maximum of interactions is performed by males during school time and females outside school time. However, there are sufficient exceptions to explain the lack of inverse correlation.

Group C had lessons from Monday to Wednesday. The time distribution is more similar to group B than to group A, as it shows more moderate absolute maximums, except for the first year, with a male peak on Tuesdays close to 45%. The lack of female presence is especially noteworthy. For the rest, as with group B, there is a moderate downward trend with a small reduction or reversal only on Sundays.

As with group B, there is greater consistency in female trends (changes are less abrupt), which is particularly evident in the first year. The probability of a relative maximum for men compared to women is again higher at the weekend, especially in the last two years, and the opposite is true on the other days. Overall, Figure 1 data indicate that the subject most influenced by the time structure is A, while B and C develop more complex behaviours, especially when gender is included as a variable. However, this initial time analysis leads to a second, more detailed one of the daily pattern of students.

Figure 1

Standardised weekly structure by gender and ordered by ID (columns).



Translation: Women – Men / Monday Tuesday Wednesday Thursday Friday Saturday Sunday / Day of the week

3.2. Daily Structure

The subject of reference (A) presents a very homogeneous behaviour between men and women in almost all academic years, with the clear exception of the final one. Besides the deviations from the most critical hours, associated mainly with the specific lecture time, it is possible to establish a clear daily structure: night and early morning, from midnight to 8 A.M., with no or very little use, and when there, it is probably by a man; the morning hours, from 9 A.M. to 2 P.M., and the afternoon, from 3 P.M. to 8 P.M., are clearly the times of high interaction; after lunch, at 2 P.M., and dinner time, from 9 to 11 P.M., when the first

is a clear relative minimum, similar to mid-morning breakfast (around 11 A.M.), while the second is a time of low, mainly female interaction in the first three years and both genders in the last two years.

There is also a gradual depolarisation of the time structure as we move into the last academic years. Thus, at 12 noon the absolute maximums in the first year are close to 30% for men and 20% for women, while in the final year they have fallen to 12% for men at 12 noon and almost 15% for women at 6 P.M.

The subjects in group B only partially replicate the time coherence of both genders observed in A. Both the night and the early morning have a small but clear relevance, with some academic years close to a 3% interaction for both genders. Therefore, despite the modest values, it has been observed that the hours of the night when there is no one working on the virtual campus are rare. There is no clear gender-linked behaviour in this time frame. The structure of the rest of the day is much more chaotic, the moments of impasse are not always detectable (breakfast, approximately between 11 A.M. and 12 A.M.), although this is more the case during more formal meals such as lunch (between 2 P.M. and 3 P.M.) and dinner (9 P.M. to 10 P.M.).

The absolute peaks do not coincide in three of the four academic years and the relative peaks tend to be displaced, except in the first case:

During the first year, women recorded an absolute maximum at 11 A.M. and two relative maximums at 6 P.M. and 11 P.M., while men have an absolute maximum at 4 P.M. and two relative maximums at 11 A.M. and 11 P.M. The convergence of absolute and relative minimums is very significant: at 6 A.M. in the first case, and at 3 P.M. and 10 P.M. for the second. Women predominate in the central morning hours (9 A.M. to 1 P.M.) and men in the afternoon (5 P.M. to 10 P.M.).

The second year shows the most divergent behaviour of the four cases, both because of the lack of presence between maximums and minimums by gender, and because of the general time structure. Women start working at 6 A.M. and constantly increase the use of the platform until 8 P.M., with an absolute maximum at 5 P.M. and a relative minimum at 11 A.M. On the other hand, men begin working on the virtual campus at 8 A.M. and reach a relative maximum at 10 A.M., before steadily decreasing until 2 P.M., when a second increase begins, culminating at 7 P.M. with an absolute maximum. Between 6 A.M. and 9 P.M., there are four major shifts in the frequency of use, with differences of around 8 percentage points and usually around 2-3 percentage points.

The third year shows again a great convergence between men and women in all the identified periods, with the exception of the relative minimum of men at 10 A.M., much more intense than that of women, with a difference of more than 3 percentage points (3.19% compared to 6.48%). To a lesser degree, this happens again at the relative minimum at

3 P.M., or at the relative maximum of 6 P.M. to 7 P.M. The relative maximum at 9 P.M. is clearly male dominated (8% compared to 5.55%)

The fourth year is a mixed scenario: the first section, from night to dawn, is relatively similar. Women start work earlier (6 A.M.) than men (8 A.M.). The former do so incrementally until 12 noon, with an absolute maximum, while the latter show a moment of great activity concentrated in 2 hours (from 8 A.M. to 10 A.M.), and then remain below or similar to the female behaviour during the rest of the day. With specific exceptions, two maximums, a relative one at 2 P.M. and an absolute at 6 P.M., which coincide with the relative minimums for women. The male absolute maximum takes place at 7 P.M., and only clearly exceeds the female maximum at 10 P.M.

The subjects of group C offer two useful groups for the study, since the time structure obtained after the cleansing of the third-year data is clearly inconsistent, as it is an almost perfect linear distribution.

In the first year, there is a time structure for women that has already been detected in other cases: they increase a little before the start of the interactions and continue to do so for the best part of the day, in this case until 4 P.M., the absolute maximum, at which time they slow down until the early hours of the following morning. On the other hand, men begin their work one hour later, at 7 A.M., but stop at 11 A.M. after a modest relative maximum of 2.60%, they increase significantly until 3 P.M. (22.40%) and slow down almost with the same curve, except for the relative maximum at 5 P.M. The absolute maximum, at 10 P.M., shows the same pattern of work concentration. Thus, in two hours of peak activity men concentrate almost half of all their daily work, while women do so more homogeneously throughout the day.

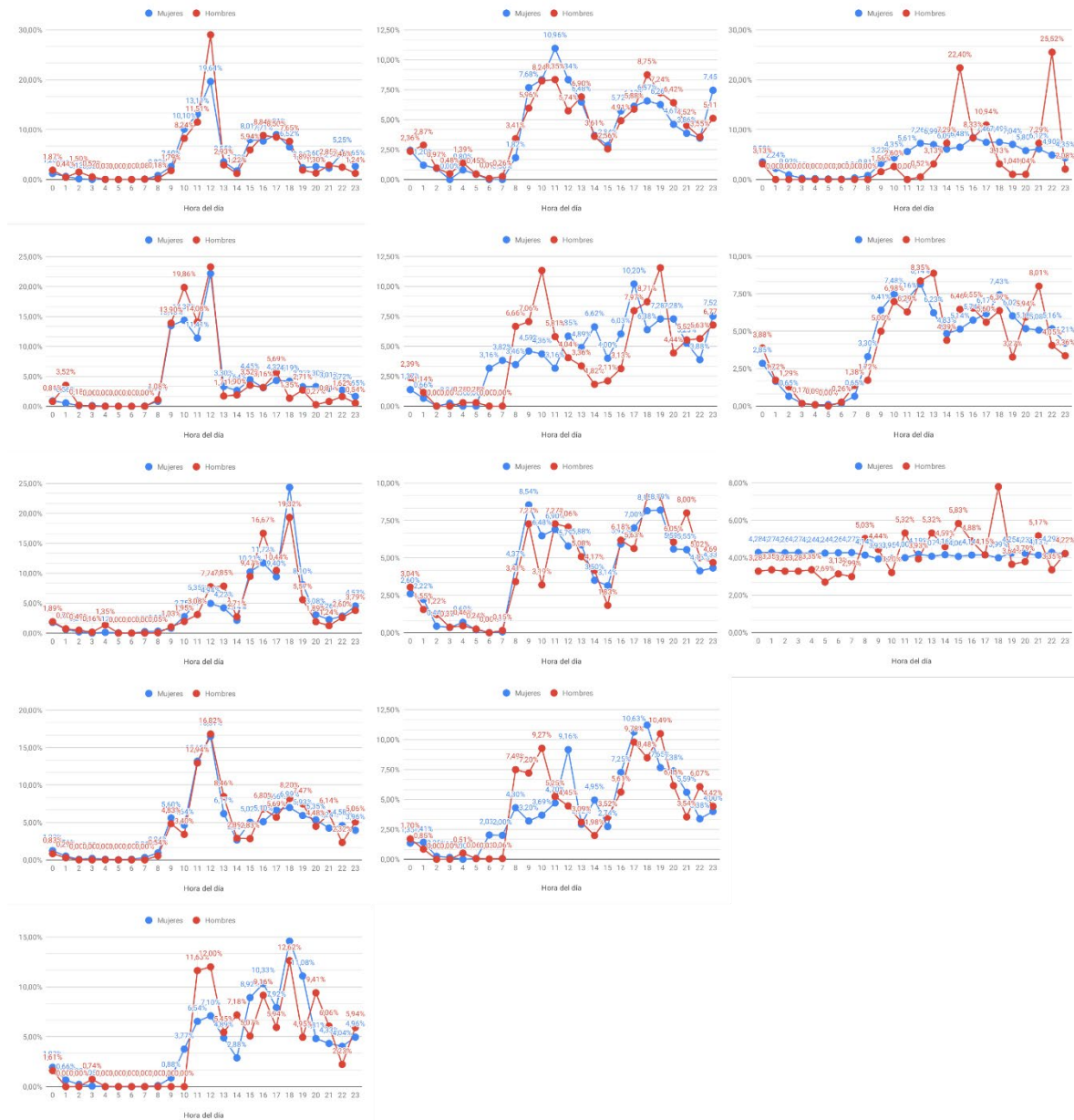
The second case reflects a high degree of concomitance, especially at night. Absolute and relative maximum levels are simultaneous and similar or have one hour gaps. Overall, only three moments of clear discrepancy are identified: two male peaks, an absolute one at 1 P.M. and a relative one at 9 P.M., which do not take place with women, and a relative low, at 7 P.M., which also deviates from the female trend by almost 2 percentage points. All of the above is shown in the graphical analysis of the data tabulation (Figure 2).

3.3. Spatial Structure

The IP (Internet Protocol) is an analysis variable that has proven to be relatively weak, especially after cleansing. The number of women under study is reduced to 861 (96,9%) and, above all, the number of men falls to 159 (53,9%). The remaining values in this field are either null, not adjusted to the format of the protocol (x.x.x.x, with x=[0..255]), or not geographically equivalent (not included in the spatial databases).

Figure 2

Daily event structure according to gender and ordered by ID (columns).



Translation: Women – Men / Time of day

IPInfo.io, a specialised service of recognised prestige in the sector, has been used for this decoding. The 12902 valid IPs outline the following structure by ID and gender (Table 3):

Table 3*Description of IP behaviour by ID and gender*

ID	Women			Men			Total		
	Students	IP	Ratio	Students	IP	Ratio	Students	IP	Ratio
A1	94	1046	11.1	17	192	11.3	111	1191	10.7
A2	52	694	13.3	5	82	16.4	57	741	13.0
A3	47	648	13.8	11	112	10.2	58	725	12.5
A4	49	646	13.2	10	124	12.4	59	738	12.5
A5	44	263	6.0	7	31	4.4	51	283	5.5
B1	28	239	8.5	44	373	8.5	72	602	8.4
B2	18	118	6.6	21	113	5.4	39	226	5.8
B3	47	486	10.3	23	266	11.6	70	733	10.5
B4	18	148	8.2	21	151	7.2	39	289	7.4
C1	161	3718	23.1	5	18	3.6	166	3734	22.5
C2	151	2391	15.8	13	189	14.5	164	2568	15.7
C3	180	2416	13.4	8	67	8.4	188	2472	13.1
<i>Total</i>	<i>861</i>	<i>12066</i>	<i>14.0</i>	<i>159</i>	<i>1536</i>	<i>9.7</i>	<i>953</i>	<i>12902</i>	<i>13.5</i>

The range of the overall ratio is set between 5.5 (A5) and 22.5 (C1), with an interval width of 17 and more than four times the minimum value. Therefore, it is an unconcentrated distribution, strongly influenced by ID and marginally by gender. With one notable exception, C1 (23.1 vs 3.6). Absolute values are not useful, given the internal gender disproportion, especially in the IDs of Group A and C. When such samples are less polarized (group B) the averages tend to converge.

The geolocation process has further reduced the number of usable IPs (9176), the rest could not be properly decoded. From smallest to biggest, the final description of spatial variables is as follows: countries (13); regions (41); towns (629); post codes (ZIP): 668. Geographical coordinates and time zone are added to this.

The geographical distribution by country is not very significant, 99.29% of the IPs used by women and 97.88% of those used by men are found within Spain. The United States have marginal relevance for men (1.73%), and Great Britain is added for women (0.31% and 0.24% respectively). The rest are purely testimonial.

It is however possible to detect spatial patterns of concentration by region on the basis the data attached (Table 4):

Table 4*Description of IP behaviour by ID and gender*

Region	F		M		Total		Region	F		M		Total	
	Abs	Rel	Abs	Rel	Abs	Rel		Abs	Rel	Abs	Rel	Abs	Rel
Madrid	2804	33.3	194	25.8	2998	32.7	New York	3	0.0	0.0	3	0.0	
Andalusia	2594	30.8	286	38.0	2880	31.4	Capital Region	1	0.0	1	0.1	2	0.0
Catalonia	808	9.6	70	9.3	878	9.6	Ceuta	2	0.0	0.0	2	0.0	
Valencia	426	5.1	28	3.7	454	4.9	Ankara	1	0.0	0.0	1	0.0	
Castille and León	285	3.4	17	2.3	302	3.3	Connaught	1	0.0	0.0	1	0.0	
Galicia	258	3.1	29	3.9	287	3.1	Coquimbo Region	1	0.0	0.	1	0.0	
Canary Islands	255	3.0	17	2.3	272	3.0	Faro	1	0.0	0.0	1	0.0	
Castille-La Mancha	157	1.9	7	0.9	164	1.8	Flanders		0.0	1	0.1	1	0.0
Balearic Islands	119	1.4	26	3.5	145	1.6	Hauts-de-France	1	0.0	0.0	1	0.0	
Basque Country	120	1.4	23	3.1	143	1.6	Île-de-France	1	0.0	0.0	1	0.0	
Extremadura	125	1.5	14	1.9	139	1.5	Istanbul	1	0.0	0.0	1	0.0	
Asturias	118	1.4	3	0.4	121	1.3	Karpoš	1	0.0	0.0	1	0.0	
Aragon	116	1.4	3	0.4	119	1.3	Lisbon	1	0.0	0.0	1	0.0	
Murcia	69	0.8	5	0.7	74	0.8	North Rhine-Westphalia	1	0.0	0.0	1	0.0	
La Rioja	46	0.5	2	0.3	48	0.5	Porto	1	0.0	0.0	1	0.0	
Navarre	33	0.4	9	1.2	42	0.5	Santiago Metropolitan	1	0.0	0.0	1	0.0	
Virginia	13	0.2	13	1.7	26	0.3	Thuringia	1	0.0	0.0	1	0.0	
Cantabria	16	0.2	3	0.4	19	0.2	Vaud		0.0	1	0.1	1	0.0
England	19	0.2		0.0	19	0.2	Wales	1	0.0	0.0	1	0.0	
Melilla	12	0.1	1	0.1	13	0.1	Washington	1	0.0	0.0	1	0.0	
California	9	0.1		0.0	9	0.1	<i>Total sum</i>	<i>8423</i>	<i>1.0</i>	<i>753</i>	<i>1.0</i>	<i>9176</i>	<i>1.0</i>

Madrid and Andalusia are the two most relevant regions, followed at a distance by Catalonia and Valencia. England is listed as the first foreign region, ahead of Melilla and Ceuta, the least relevant in quantitative terms in the whole country.

By gender, small discrepancies in the distribution can be observed: Madrid is more relevant for women than for men, with a difference of more than 8 percentage points. The opposite takes place in Andalusia, with a similar difference. Most IPs are activated by women in almost all other regions except Galicia, the Balearic Islands, the Basque Country, Extremadura, Navarra, Virginia and Cantabria. The rest are not statistically representative for our sample.

The towns with the highest activity correspond in part to the regions with highest activity, although secondary regions are also present, indicating a high degree of concentration in the internal distribution of the latter. Madrid and Seville clearly stand out (27.1 and 11.7% respectively), followed by Malaga (4.1%), Barcelona (3.9%), Cordoba (1.5%), Valencia and Granada (both with 1.3%). The rest show values of less than 1%.

The IPs used by women are again in the majority in Madrid (27.7% vs 20.3%), along with Seville and Barcelona (12% vs 8%, and 3.9% vs 3.7%)., The trend is reversed in the following towns: Malaga (3.7% vs 8.6%), Cordoba (1.3% vs 1.7%), Valencia (1.3% vs 1.7%) and Granada (1.2% vs 1.9%), although the differences, except in the first case, are not significant. A heat mapping technique has been used to demonstrate the usage intensity according to spatial distribution (Figure 3):

The colour scale has been adjusted in order to identify the local patterns of lesser intensity, as the concentration of the first two cities is very high. A marginal pattern is thus detected, linked to the Mediterranean coast all the way to the Bay of Cadiz. On this stretch important local concentrations can be observed (Costa del Sol and Levante), with the area around Barcelona (Mataró, Sabadell, Sitges) clearly standing out.

It is also possible to ascertain that in inland areas one-off situations are the norm, even in the surroundings of important cities such as Valladolid, Salamanca, Zamora or León. Areas of minor concentration are again detected around Galician cities, especially Coruña, Santiago de Compostela and Vigo.

3.4. Data Mining

Considering the heterogeneity of the data, four learning automata were used to validate such data before searching for patterns in Orange: logistic regression, random forest for classification, support vector machine (SVM) and Naive Bayes. Additionally, REPTree was used in Weka. The user manuals of each application were used to analyse and implement the results of the algorithms.

The objective variable was Gender. When a variable produced another derived variable, the latter was chosen, and the former was excluded from the matrix. Considering this scenario, the overall results of the methods were those shown in Table 5.

Figure 3

Heat map of the IPs

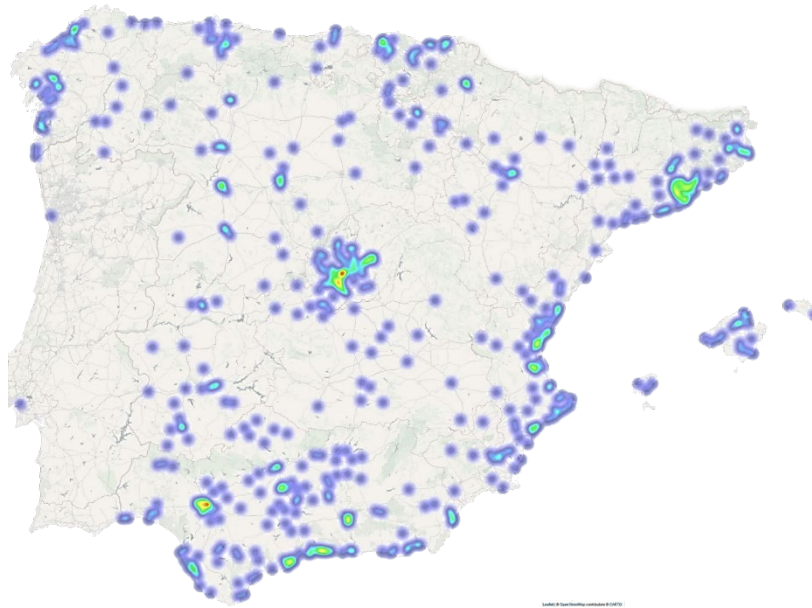


Table 5

Statistics of the methods tested on the gender category

Method	AUC	CA	F1	Precision	Recall
SVM Learner	0.510	0.420	0.492	0.759	0.420
Random Forest Learner (RFL)	0.839	0.884	0.868	0.869	0.884
Logistic Regression	0.784	0.869	0.852	0.848	0.869
Naive Bayes	0.769	0.837	0.840	0.843	0.837
REPTree	0.634	0.862	0.805	0.849	0.862

The best classifier for all indicators is RFL. According to Tape (n.d.), it is a good classifier, being in the .80-.90 bracket. Therefore, gender is a useful variable to determine the behaviour of students on the Virtual Campus. However, values above .90 would be desirable if solid patterns are to be established, as the false positive rate is high for men, as described in their confusion matrix (Table 6).

The classification of women is clearly more correct than that of men (success ratio of 0.97 to 0.35). Therefore, the female behaviour is more statistically consistent than the male one, although the weakness of the second sample may imply errors of interpretation due to inconsistency. It should be taken into account that men are a small percentage compared to women.

Table 6*Confusion matrix for RFL*

		Prediction			Derived statistics			
		F	M	Σ				
Real	F	152820	4339	157159	Positive F ratio	0.97	Negative F ratio	0.03
	M	16771	8813	25584	Positive M ratio	0.35	Negative M ratio	0.66
	Σ	169591	13152	182743	Accuracy	0.96	Precision	0.67

Caption: the values indicate the number of events. The meaning of the statistics is available on <https://bit.ly/2MfBCJN>.

The data mining has not yielded clear results due to the interference of the weekly and daily structure of the different subjects. Nevertheless, it is possible to point out some common trends: women are more likely to be working on the Virtual Campus both at the beginning of the week and at the beginning of the day. On the other hand, those who access or work on the Virtual Campus at night are more likely to be men.

An in-depth study of each academic year does lead to clear sets of rules by using a priori algorithms, both for the detection of general patterns, which affect a large part of the class, and marginal ones, useful for the detection of homogeneous subgroups. Clear examples of both situations are given for ID = B3. Ten rules of behaviour by gender were found to have an effect on the "Workshop" work component, nine for women and one for men.

In the first case they are recurrent, each involving 15 to 25 students for specific days and times (Monday, Tuesday, and/or Wednesday at 9 A.M., 7 P.M. and/or 8 P.M.) or for a specific time without day restriction (8 P.M.). There is only a partial overlap, so it establishes a set of behavioural patterns for most women, and accuracy values ranging from 0.98 to 0.99.

As for men, the rule affects seven students, i.e. less than a third of them. It is therefore a small group of well-defined behaviour (accuracy of 0.96) within a heterogeneous subset.

4. Discussion

The analysis of the relationship between the weekly calendar and gender has highlighted some interesting aspects. Although in general there is a high degree of similarity between the two patterns, especially at times of greater intensity, the situation shows nuances when the attention is focused on marginal situations. Moreover, this has been quite consistent over time.

It is interesting to note that in the subjects of the Pedagogy degree (group A), a change of pattern can be observed in the last year: women work more on the virtual campus during the week, while men clearly do so during the weekend, always in relative terms. The reasons for this break in the similarity pattern are not clear: both the teaching staff, the subject and

its time structure are the same. Unfortunately, longitudinal cohort studies are not possible here, so the influence of the student context is both conditioning and undetermined.

The results of the subjects in group B correspond to a postgraduate degree for students in Geography and History. These are therefore older students who, in principle, study fewer subjects compared to an undergraduate degree. We believe that this aspect can explain the greater homogeneity of their patterns. Here, absolute and relative maximums never exceed 25%, whereas in the previous case they almost reach 80% in some academic years.

Their weekly work is less concentrated and, although not exempt from general trends, it must be carefully differentiated by gender: in half of the cases there is female pre-eminence in weekend work, in another case it is the opposite, and the last one shows female pre-eminence only on one of the two days, Saturday. There are no significant differences during the week either, except for Monday, when, in three of the four courses, males show more participation.

Group C subjects belong to a Teaching degree and, against expectations, they reveal a pattern of behaviour more similar to B than to A. This is remarkable since A and B share a curriculum clearly linked to Education, while B's material is specifically designed for students of the faculties of Humanities, hence, without a previous regulated didactic background. However, some clearly exceptional situations, such as that observed in first-year students on Tuesdays, could not be explained on the basis of the structure of the campus or teachers' understanding of the ongoing situation.

Although the analysis of comparative behaviour by gender and time of day were expected to be clearer, concomitance has been in fact the main feature: the hourly concentration has again been more relevant in the subjects of group A and clearly lower in those of B and C, with a difference of 10 percentage points or more in their absolute maximums.

It is only when dealing with marginal, sometimes even main, behaviours that differences between genders can be observed: absolute peaks, i.e. moments of high intensity of work, are usually dominated by men, while women generally show higher average values in time brackets close to these peaks, which means that they work for longer stretches of time.

Overall, and in the sample studied, the analysis reveals the existence of both daily time patterns specific to each subject and gender patterns linked to specific academic years, but not in a structural way: there are academic years in which similarities in behaviour by gender explain much more than the differences do: However, in other years the differences are clear, and in such cases they are usually characterised by the tendency for men's work to be concentrated and women's work to be more evenly distributed over time.

The IP has proven to be a variable of great interest for geographical analysis and comparison by gender. Some clear differences have been found with regard to both dispersion and spatial concentration, which indicate a greater mobility of women outside their original locations, although with a greater tendency towards concentration in the destinations. On the other hand, the men under study show less mobility, but do so to a greater number of destinations; in both cases proportionally.

5. Conclusions

The study of learning behaviour on virtual campuses is a field of study still in its infancy in Higher Education, especially in the Didactics of Geography and Social Sciences. However, despite its complexity, the knowledge and insights it brings to the table make it useful, relevant and necessary. The time students spend in front of computer screens as part of their learning process cannot be ignored, especially in line with the European ECTS credit format.

The methodology used needs to be developed, especially with regard to data mining. This approach has proved interesting in an aspect that is particularly difficult to perceive: the homogeneous marginal behaviour, i.e. small group trends that are hard to detect because of their very nature.

It has become clear that the robustness of this type of analysis is based on the availability of significant volumes of data. Therefore, a correct data recording and unification is a key element that may question whether certain data should be included or not, as has happened with one of the subjects in group C.

The location analysis of the IPs is, in our opinion, an opportunity to take advantage of interesting geographic information, both from a basic and applied perspective. It requires, and opens the door to, future research that exploits its potential from totally heterogeneous perspectives such as humanist geography (Tuan, op. cit.) or quantitative geography (Hägerstrand, 1985).

The focus on the main variable, i.e. gender, has also been a determining factor in a context, i.e. education, where women are in the majority. This allows for the research of one part of the population but questions the analytical soundness of the other. This is why prediction errors in the confusion matrix are so disproportionate between men and women.

Nonetheless, the introduction of this analysis is presented in a natural way and retrieves a set of data not much used in the geographical or educational context. Beyond the similarities and discrepancies found from this comparative perspective, to which it is necessary to add further analysis before obtaining solid explanatory arguments, the study does appear to offer sufficient richness to be researched systematically, at least from the didactic perspective of geography.

In conclusion, we believe it necessary to continue with this type of research, and to add case studies that will allow future developments of more robust studies on the consistency of patterns divided by general and unique profiles. Research should also complement this type of results with explanatory analyses, for example, on the reasons why men and women have different behaviours, on the learning difficulties that can be detected in this subject field and what relationship all this has with the academic performance on the curriculum.

La influencia del sexo en el uso de los campus virtuales.

Estudio de caso

1. Introducción

Cuando Martin Dougiamas liberó la primera versión de su LCMS (Learning Content Management System) en 2002, Moodle, descontento por la presencia del software disponible y motivado por el deseo de incorporar el constructivismo pedagógico, nada hacía pensar que se convertiría en un estándar de facto, y con ello pondría en primera línea de interés didáctico el soporte digital virtual, especialmente el universitario.

Hoy día, se opte por formas propietarias o libres, es impensable que una asignatura universitaria, y cada vez más en la enseñanza media, no tenga asociada un campus virtual, si no es que se lo utiliza como vía preferente o exclusiva de trabajo. Algo que cobra especial sentido gracias a las nuevas formas de análisis educativo que ofrece esta plataforma (Hamotoglu et al., 2020; Cantabella et al., 2019; Quintanas Mendes et al., 2019; Fakir & Touya, 2014; Lu & Law, 2012), las técnicas analíticas que permite introducir (Álvarez Vázquez et al., 2020; Charanya & Kesavan, 2019; Cohen, Rahimi & Zilka, 2019; Manne, Yelisetti, Kakarla, & Fátima, 2014) y las dimensiones sociales y conductuales que hace posible abordar (Pakanen, Alavesa, Arhipainen & Ojala, 2020; Castro, Menacho & Pérez-Molina, 2018; Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017; Parise, 2016).

Bajo nuestro criterio, dos conjuntos de ventajas docentes se presentan como más relevantes: las que derivan de la deslocalización espaciotemporal del acto educativo, sea de forma parcial o total; y las que se originan en la forma tecnológica de interactuar, tanto entre los usuarios como con los contenidos de la asignatura.

Respecto a las primeras, es evidente que la simbiosis hombre-máquina que ya se intuyera en los inicios de la computación (Licklider, 1960), ha pasado a ser un hecho en expansión y de efectos todavía no bien explorados, aunque sí percibido desde hace tiempo en Educación (Orostica Verdugo, 2020; Desai, Chavan & Tendulkar, 2020; Cabanillas, Luengo & Carvalho, 2019; Morales, Infante-Moro & Gallardo-Pérez, 2019; Ballesteros Regaña et. al, 2010; Correa & Paredes, 2009), Geografía (Capel, 2009) y otras ciencias sociales (Sarduy Domínguez et al. 2020; Arias & Defiore, 2019; Gómez Collado, Contreras Orozco & Gutiérrez Linares, 2016). El horario y calendario escolar, entendidos como horario de clase y distribución semanal de la carga lectiva, se sigue planteando bajo una lógica previa a la vigente realidad tecnológica. La tradicional concurrencia espacio-temporal de los tres elementos del acto educativo (docente, discente y objeto de estudio) encuentra una alternativa deslocalizadora en los LCMS, que posibilita una concepción ubiquesta de aquél. Ello implica concebirlo como un nuevo objeto de estudio, impregnado de una complejidad cronogeográfica (Tartia, 2020; Thulin, Vihelmsen & Scwanen, 2020; Vega Valverde, 2017) y conductual claramente distinta a la que fue concebida en el siglo XIX, durante la institucionalización educativa, y aún a la regulada en la actualidad por los sucesivos marcos normativos.

Respecto a las segundas, los nuevos medios técnicos para aprender y colaborar en una asignatura son un espacio todavía ignoto en buena medida para el docente. Sea por el uso de herramientas colaborativas con base en la nube (Internet), sea por la posibilidad de consultar las fuentes docentes en cualquier momento y lugar, lo cierto es que la forma en que el alumnado interactúa con los contenidos y las tareas no tiene nada que ver con las formas posibles, y disponibles, apenas una o dos generaciones atrás.

La entrada de España en el Espacio Europeo de Educación Superior (BOE nº 224 de 18 de septiembre de 2003) puso estas cuestiones en primera línea de interés reflexivo docente. Ya que la introducción de los créditos ECTS (European Credit Transfer System) supone un intento de cómputo no sólo de la carga lectiva tradicional (en clase), sino de las horas de estudio, tutorías, seminarios, trabajos y cualesquiera otra dedicación necesaria o requerida para la superación de una asignatura.

Controlar el devenir del alumnado fuera de clase no parece una solución razonable ni deseable, y centrarse en los resultados objetivos (notas de examen, de exposiciones, trabajos...) plantea dudas sobre el carácter enfático que todos los modelos educativos modernos demandan sobre los procesos de formación continua y de retroalimentación docente-discente; mediante los que frecuentemente es cuestionado un profesorado que ahora requiere de capacidades adicionales (Sola-Martínez et al., 2020; Abbott, 2011). Especialmente bajo el prisma de las analíticas de aprendizaje (learning analytics) en el que nos apoyamos en la presente, dada su versatilidad (Ruipérez-Valiente et al. 2020; Papoušek, Pelánek & Stanislav, 2016; Sclater, Peasgood & Mullan, 2016; Gašević, Dawson & Siemens, 2015; Siemens & Long, 2011).

Es por ello que consideramos necesario explorar el análisis de patrones de comportamiento (digital) del alumnado en los campus virtuales como una forma indirecta de conocimiento de desempeño discente y comportamiento espaciotemporal, no necesariamente invasivo si se establecen consideraciones éticas elementales, en línea con enfoques humanistas ya consolidados en Geografía (Tuan, 1977 & 1990).

Cada usuario identificado en un campus virtual, durante el uso que hace del mismo, genera multitud de registros de datos que informan del lugar, la hora y sus acciones (abrir un enlace, una presentación, un documento, realizar un comentario en un foro, entregar una tarea...); a cada uno de estos registros se le denomina evento (Caballé & Xhafa, 2013). Información muy interesante, a la que se pueden añadir incluso más variables, como el sexo, a partir de los propios datos de identificación del usuario, para dotar de mayor riqueza a la fuente.

Su análisis puede hacerse partiendo de un bagaje estadístico modesto, hasta incorporar las técnicas y metodologías de vanguardia. Gracias a estas últimas los tiempos de trabajo, y de obtención de resultados, pueden ser claramente factibles de cara al momento de evaluación o, incluso, de rectificación de la estrategia didáctica (Badilla Quintana, Vera Sagredo & Lytras, 2017; Minguillón, Santanach & Appel, 2016).

A tenor de todo lo anterior, nuestro objetivo es realizar un análisis comparativo entre sexos con la intención de comprobar las posibles semejanzas y diferencias en el uso de los campus virtuales desde tres perspectivas: el momento de uso, el lugar de uso y el contenido visitado. En línea con otras investigaciones que señalan la pertinencia de explorar este tipo de cruce de variables para conocer de forma adecuada la conducta del alumnado en los campus virtuales (Álvarez Vázquez, 2019; Cabanillas García et al., 2019; De la Iglesia-Villasol, 2019).

2. Metodología

El presente estudio analiza los datos recogidos sobre más de 1000 alumnos en un periodo de 4 años, entre los cursos 2016/17 y 2019/20; organizados en 14 grupos de clase pertenecientes a 4 asignaturas de grado y posgrado de tres titulaciones, pertenecientes a 2 universidades españolas. Tres de las asignaturas pertenecen al Área de conocimiento de la Didáctica de las Ciencias Sociales y una cuarta a Didáctica y Organización Escolar:

- Recursos Didácticos y Tecnológicos en Educación en Grado de Pedagogía (A1 - P-RDTE).
- Currículo de Geografía e Historia en Máster de Profesorado (B1 - M-CCSS-C).
- Diseño y Desarrollo de Programaciones y Actividades Formativas de Geografía e Historia en Máster de Profesorado (B2 - M-CCSS-D).
- Didáctica de las Ciencias Sociales en Grado de Educación Infantil (C1 - I-DCCSS).

La Tabla 1 recoge los descriptores fundamentales de cada asignatura (curso, alumnado y número total de eventos), con objeto de establecer las dimensiones del universo de referencia. Pero dada la heterogeneidad interna de las variables (alumnado y eventos), es necesario recurrir a una forma proporcional de comparación (Evaluaciones / Alumnado), para definir cómo de relevante es en cada caso la interacción con la plataforma en el desarrollo de la asignatura. En este sentido, y usando a los docentes como criterio de ordenación, el resultado sería: A (281,4), C (140,7) y B (119,0); con una media conjunta de 180,4 eventos por asignatura y persona.

A partir de estos datos se intenta detectar patrones de comportamiento común entre los diferentes grupos, así como singularidades propias, atendiendo a la naturaleza de los contenidos consultados, el momento en que se produce la consulta y la influencia, o no, del sexo en dicho comportamiento (Martínez Romera, 2017). Estamos, por tanto, en un contexto analítico propio del Big Data, como parte del Learning Analytics, en educación (Siddiqa et al. 2016; Tsai, Lai, Chao & Vasilakos, 2015; Sin y Muthu, 2015).

Tabla 1*Eventos generados por asignatura y grupo*

Docente	Materia	Asignatura (grupos)	ID	Curso	A _b (A _r : M/F)	Eventos
A	A1	P-RDTE (2)	A1	2017/18	111 (111: 17 / 94)	27.316
		P-RDTE (1)	A2	2018/19	58 (57: 5 / 52)	17.036
		P-RDTE (1)	A3	2018/19	59 (58: 11 / 47)	16.208
		P-RDTE (1)	A4	2019/20	60 (59: 10 / 49)	23.236
		P-RDTE (1)	A5	2019/20	52 (51: 7 / 44)	10.928
B	B1	M-CCSS-C (1)	B1	2018/19	72 (72: 44 / 28)	5.775
		M-CCSS-C (1)	B2	2019/20	41 (39: 21 / 18)	3.451
	B2	M-CCSS-D (1)	B3	2018/19	72 (70: 23 / 47)	9.942
		M-CCSS-D (1)	B4	2019/20	41 (39: 21 / 18)	7.728
C	C1	I-DCCSS (1)	C1	2016/17	166 (166: 5 / 161)	24.003
		I-DCCSS (1)	C2	2017/18	164 (164: 13 / 151)	24.458
		I-DCCSS (1)	C3	2018/19	188 (188: 118 / 180)	22.899
Σ					1084 (1074: 295 / 889)	192.980
\bar{x}					87,36	15.060

Leyenda: A_b indica número de alumnos iniciales; A_r tras depuración y normalización. M/F: desglose por sexos (Masculino / Femenino).

Un aspecto importante, y común a todas las asignaturas, es la clara feminización de la población, muy especialmente en la asignatura de Educación Infantil. Hecho que ha de ser tenido en cuenta en los procesos de comparación interna y externa, y que implica, en el presente estudio, la normalización por sexo de variables previa a su cotejo.

La minería de datos (Data Mining) ha demostrado ser un instrumento útil, especialmente en contextos educativos donde hay producción masiva (Koedinger et al., 2015; Sin & Muthu, 2015), algo claramente aplicable a la situación planteada, y en general a los procesos de enseñanza-aprendizaje apoyados en entornos virtuales. En consecuencia, para el análisis de datos se recurrirá a dos aplicaciones:

Orange (v3.19), software de código abierto para la aplicación de minería de datos y máquinas de aprendizaje utilizado con éxito en volúmenes significativos de datos, como el que nos ocupa (Demsar et al., 2013; Demsar et al., 2004). El modelo de relación con el usuario es el de programación visual (visual programming language, VPL), algo que lo hace especialmente interesante en contextos de investigación social o educativa donde las destrezas informáticas o estadístico-matemáticas son singulares, especialmente en Geografía (Dobesova, 2011).

Weka (v3.8), software centrado en los mismos ámbitos analíticos que el anterior, pero con un mayor grado de especialización en cuanto al número y diversidad de algoritmos implementados e implementables (Sebastian & Puthiyidam, 2015; Ahmadi & Ahmad, 2013). Presenta un modelo más procedural y menos interactivo, pero su diseño es apto para volúmenes de datos superiores a igualdad de condiciones técnicas (hardware), respecto al primero.

Debido a las diferencias de configuración de Moodle, el LCMS sobre el que se han construido todos los campus virtuales de este estudio, entre las universidades implicadas, así como por la necesidad de realizar un control de datos inicial, se plantea un procedimiento de depuración y normalización previo a la explotación. Se recurre para ello a un contexto de base de datos en el que se manipulan los datos mediante SQL, expresiones regulares, distancias de edición y minería de datos. El proceso global queda definido por la estrategia knowledge discovery in databases (KDD), e implica una secuencia de tareas encadenadas: selección, preproceso (depuración), transformación, minería e interpretación (García et al., 2016; Ramírez-Gallego et al., 2016; Taleb, Dssouli & Serhani, 2015). Se realiza un proceso de selección y reclasificación de variables, así como de calificación respecto a su posible utilidad en los análisis posteriores.

Se construyen 5 variables a partir de otras preexistentes con objeto de dotar de funcionalidad a los análisis: se asignan sexos, se simplifican las estructuras temporales (hora y día de la semana) y se reorganizan los valores modales de los eventos generados. Además, cada conjunto de eventos para una misma asignatura y curso recibe un identificador único (Asignatura). Con todo ello, los resultados del proceso previo al inicio del análisis se recogen en la Tabla 2.

El soporte tecnológico se ha completado con, y ha requerido, el uso de hardware especializado para cálculo masivo, gracias a los servicios de supercomputación del Área de Sistemas de Información de la universidad matriz. Los métodos y técnicas utilizados para el desarrollo de la investigación han sido diversos, y aplicados en función de la fase de análisis:

- Obtención de los datos: descargas desde Moodle, se hicieron a partir de hojas de cálculo, posteriormente integradas en una base de datos.
- Depuración y normalización: filtros avanzados de edición, lenguaje SQL, distancias de edición, expresiones regulares y edición manual en casos marginales.
- Manipulación de datos: lenguajes SQL y Python.
- Decodificación de IP mediante servicios especializados de reconocido prestigio (IPInfo.io).
- Explotación de datos: estadística descriptiva y selección de autómatas de aprendizaje para minería de datos a partir de las ideas desarrolladas por Kaur, Singh & Josan (2015).

Tabla 2*Depuración, estandarización de variables y universo final de referencia*

Variable	Presencia (Docente)	Valores	Normalizado	Comentario (Normalización)
Día	ABC	1391	7	Deriva nueva variable: Día de la Semana (Lunes a Domingo).
Hora	ABC	20644	24	Deriva nueva variable: Hora del día (0-23h).
Nombre Completo del Participante	ABC	186642	186642	Deriva nueva variable: Sexo (Masculino/Femenino).
Participante Afectado	AB	186642	186642	Reflexivo, no apunta a terceros. Escaso interés.
Contexto del Evento	ABC	824	13	Deriva nueva variable: Contexto (1-13), con valores agregados por frecuencias. Eliminación de redundancias y registros nulos.
Componente	AB	21	19	Normalización menor necesaria
Nombre del evento	ABC	145	24	Deriva nueva variable: Evento (1-24). Eliminación de redundancias y registros nulos.
Descripción	AB	186642	186642	Nombre del evento + ID interno Moodle de usuario. ID no descargable invalida su depuración. Escaso interés.
Origen	AB	3	3	Semántica de la variable de escaso interés para el estudio. La práctica totalidad de los registros adquieren el valor 'web'.
Dirección IP	ABC	12892	12892	De interés para georeferenciación y el análisis de itinerarios (cronogeografía).
<i>Estudiantes finales</i>	<i>1074</i>	<i>(99,1% del total)</i>		
<i>Eventos finales</i>	<i>182743</i>	<i>(94,7% del total)</i>		

3. Análisis

3.1. Estructura semanal

Las asignaturas del grupo A presentan en términos generales una clara concentración los martes, día de clase. El comportamiento por sexos en ese momento es bastante semejante. Los días previo y posterior ofrecen valores modestos, pero algo más elevados que en el resto de la semana, tiempo en el que es testimonial su uso con valores por debajo del 2-3%. Sólo en A4 y A5 hay un claro repunte de los días previos (sábado a lunes) y uno más modesto para el resto de días.

En las asignaturas del primer curso los datos presentan desviaciones puntuales, pero que en el conjunto de la semana tienden a converger en horquillas de 2-3 puntos porcentuales; mientras que en el último las féminas interactúan claramente con mayor frecuencia todos los días de la semana, excepto sábados y domingos, momento en el que se produce una fuerte inversión. Es el domingo el día más relevante para ello: 21,53% de interacciones masculinas por 9,14% femeninas.

Las asignaturas del grupo B se impartían de forma consecutiva los mismos días (lunes y martes). B2 contaba con una sesión adicional los miércoles. La correspondencia con los

picos de clase es notable en casi todos los cursos, con la salvedad del tercero, en el que el jueves es claramente un día de mayor uso relativo. En todo caso se observa la preeminencia de una tendencia descendente según transcurre la semana, hasta el día previo al inicio de las clases. La distribución temporal es más homogénea que en el caso anterior (con picos superiores al 60 y 70%), con máximos inferiores siempre al 25% y mínimos superiores al 6%.

Los patrones de comportamiento por sexo expresan diferencias puntuales, pero no permiten discernir modelos claros: es más probable que en periodo lectivo el máximo relativo de interacciones sea masculino y que en periodo no lectivo sea femenino; sin embargo, existen excepciones suficientes que explican la inexistencia de correlación inversa.

La asignatura del grupo C se impartía de lunes a miércoles. La distribución temporal de eventos es más semejante al grupo B que al A, en tanto que presenta máximos absolutos más moderados, excepción hecha del primer curso, con un pico masculino los martes, próximo al 45%. Especialmente llamativo por la inexistencia de reflejo femenino. Para el resto, al igual que sucede con B, se observa una tendencia descendente moderada que sólo encuentra reducción o inversión, pequeña en el mejor de los casos, el domingo.

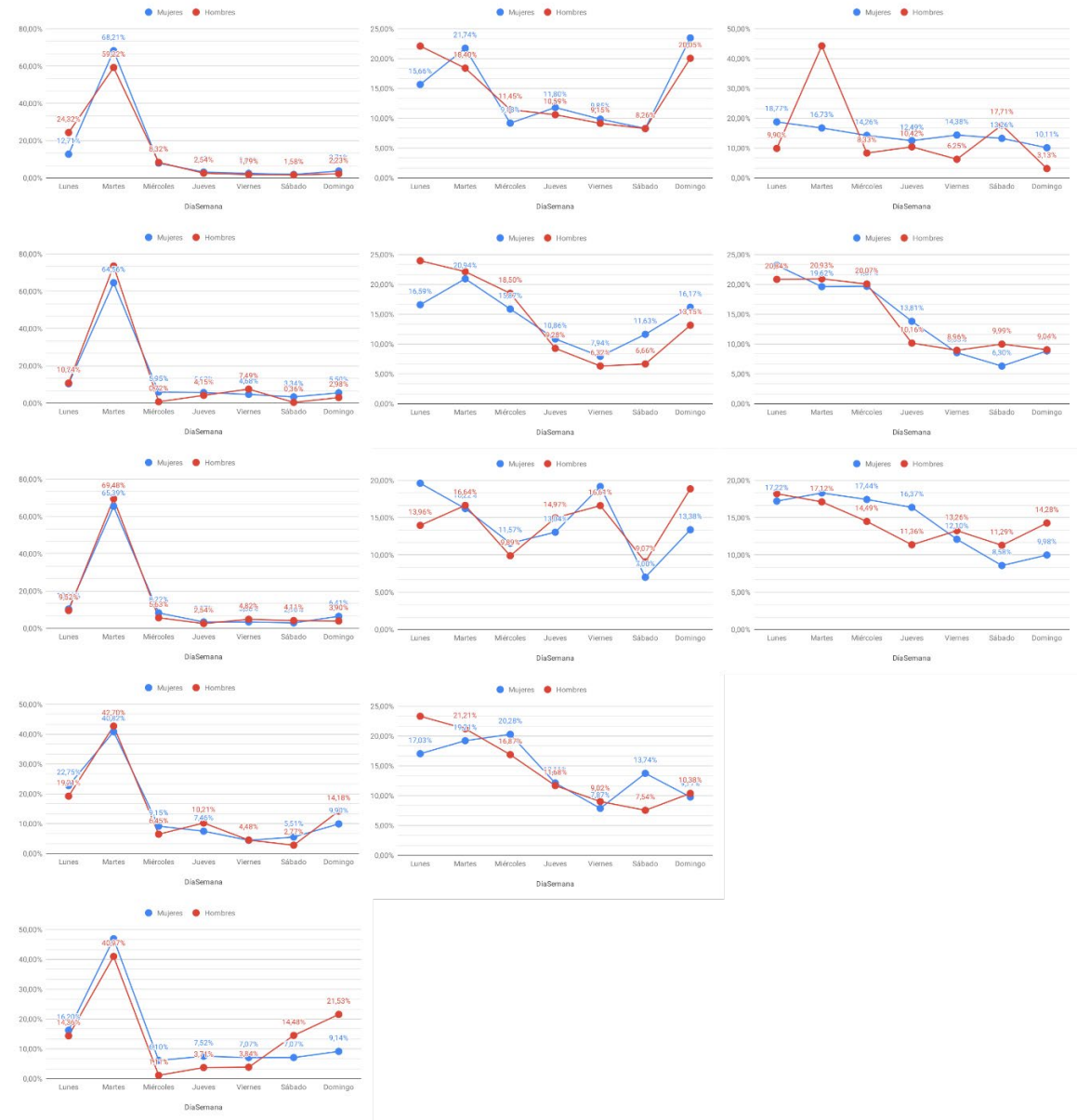
Al igual que sucede con el grupo B, se observa una mayor constancia en las tendencias femeninas (cambios bruscos menos acusados), algo especialmente claro en el primer curso. Así como vuelve a constatarse una mayor probabilidad de máximo relativo masculino frente al femenino durante el fin de semana, especialmente en los dos últimos cursos, y lo contrario durante el resto de días. En conjunto, Figura 1, los datos reflejan que la asignatura más influenciada por la estructura temporal es la A, mientras que B y C desarrollan comportamientos más complejos, especialmente cuando se incluye el sexo como variable. Con todo, este primer análisis temporal deriva en un segundo, de mayor detalle, relativo al patrón diario del alumnado.

3.2. Estructura diaria

La asignatura de referencia (A) presenta un comportamiento muy homogéneo entre hombres y mujeres en casi todos los cursos, con la clara excepción del último. Más allá de las desviaciones en el centro de gravedad sobre las horas más relevantes, asociadas fundamentalmente con la hora específica de clase, es posible establecer una estructura diaria clara: la noche y la madrugada, de 0 a 8h, de nula o muy escasa relevancia en el uso, cuando éste se produce se trata probablemente de un hombre; la jornada laboral de mañana, de 9 a 14h y la jornada laboral de tarde, de 15 a 20h, son momentos en los que se concentran claramente las interacciones; y la sobremesa, 14h, y la cena, de 21 a 23h, en el que el primero es un mínimo relativo claro, con cierto reflejo a media mañana con el desayuno (alrededor de las 11h), mientras que el segundo es un momento de baja interacción de preeminencia femenina en los tres primeros cursos y mixta en los dos últimos.

Figura 1

Estructura semanal normalizada por sexo y ordenada por ID (columnas)



También se observa una paulatina despolarización de la estructura temporal según avanzamos a los cursos más recientes. Así, los máximos absolutos del primer curso están próximos al 30% en el caso de los hombres, a las 12h, y al 20% para las mujeres, en el mismo tiempo, mientras que en el último curso ha descendido hasta fijarse en un 12% para los hombres a las 12h, y de casi 15% para las mujeres a las 18h.

Las asignaturas del grupo B emulan solo parcialmente la coherencia temporal de ambos sexos observadas en A. Tanto la noche como el inicio de la madrugada tienen una relevancia pequeña pero clara, con interacciones que rozan el 3% en algunos cursos para ambos sexos. Así, pese a los valores modestos, se constata que son raras las horas de la noche en las que no hay alguien trabajando con el campus virtual. No se observa un

comportamiento ligado a sexo claro en esta franja. La estructura del resto del día es mucho más caótica, los momentos de impasse no son siempre detectables (desayuno, de forma relativa entre 11 y 12h), pero sí algo más en los de naturaleza más formal almuerzo (entre 14 y 15h) y cena (21 a 22h).

Los máximos absolutos no coinciden en tres de los cuatro cursos y los relativos tienden a estar desplazados, salvo en el primer caso:

Durante el primer curso las mujeres registran un máximo absoluto a las 11h y dos relativos, a las 18h y 23h. Mientras que los varones presentan su máximo absoluto a las 16h y dos máximos relativos a las 11h y 23h. La convergencia de mínimos absolutos y relativos es muy significativa: a las 6h en el primer caso, y a las 15h y 22h para los segundos. Se observa una preeminencia de las mujeres en las horas centrales de la mañana (9 a 13h) y de los hombres por la tarde (17 a 22h).

Durante el segundo curso se detecta el comportamiento más divergente de los cuatro casos, tanto por la falta de reflejo entre máximos y mínimos por sexo, como por la estructura temporal general. Las mujeres inician las interacciones a las 6h e incrementan de forma constante el uso de la plataforma hasta las 20h, con un máximo absoluto a las 17h y un mínimo relativo a las 11h. Por su parte, los hombres inician el trabajo con el campus virtual a las 8h y alcanzan un máximo relativo a las 10h, para descender de forma sostenida hasta las 14h, momento en el que se inicia un segundo incremento que culmina a las 19h con máximo absoluto. Entre las 6h y las 21h, se producen cuatro inversiones de preeminencia en la frecuencia de uso, con diferencias que llegan a rondar los 8 puntos porcentuales y que suelen estar alrededor de los 2-3 puntos porcentuales.

El tercer curso vuelve a mostrar una gran convergencia entre hombres y mujeres en todos los tramos definidos, con la salvedad del mínimo relativo de los hombres a las 10h, mucho más acusado que el de las mujeres, entre los que median más de 3 puntos porcentuales (3,19% frente a 6,48%). De forma menos relevante esto vuelve a suceder en el mínimo relativo de las 15h, o en el máximo relativo de 18 a 19h. El máximo relativo de 21h sí está claramente masculinizado (8% frente a 5,55%).

El cuarto curso presenta una situación mixta: el primer tramo, noche a madrugada, es relativamente similar. Las mujeres inician el trabajo antes (6h) que los hombres (8h). Las primeras lo hacen de forma creciente hasta las 12h, con un máximo absoluto, mientras que los varones presentan un momento de gran actividad concentrado en 3h (de 8 a 10h), para quedar por debajo o similar al comportamiento femenino el resto del día. Con excepciones puntuales, dos máximos, relativo a 14h y absoluto a las 18h, que coinciden con mínimos relativos femeninos. El máximo absoluto masculino se localiza a las 19h, y sólo volverá a superar claramente al femenino en su máximo relativo de las 22h.

Las asignaturas del grupo C presentan dos grupos útiles para el estudio, ya que la estructura temporal obtenida tras la depuración del tercer curso es claramente inconsistente, al tratarse de una distribución lineal casi perfecta.

En el primer curso se observa una estructura temporal femenina ya detectada en otros casos: incrementan un poco antes el inicio de las interacciones y éstas experimentan un incremento sostenido durante buena parte del día, en este caso hasta las 16h, máximo absoluto, momento en el que inicia un lento declive hasta las primeras horas de la madrugada. Por su parte los varones inician una hora después sus interacciones, a las 7h, pero las interrumpen a las 11h tras un modesto máximo relativo de 2,60%, para incrementarse exponencialmente hasta las 15h (22,40%) y descender casi con la misma pendiente, salvedad hecha del máximo relativo de las 17h. El máximo absoluto, a las 22h, refleja el mismo esquema de concentración en las interacciones. Así, en dos horas de máxima actividad los hombres concentran casi la mitad de todas sus interacciones del día, mientras que las mujeres interactúan de forma más homogénea a lo largo del día.

El segundo caso refleja un alto grado de concomitancia, especialmente en las horas nocturnas. Los máximos absolutos y relativos son simultáneos y semejantes, o presentan decalajes de una hora. En conjunto sólo se identifican tres momentos de discrepancia clara: dos máximos masculinos, absoluto a las 13 y relativo a las 21h que no encuentran un reflejo semejante en las féminas, y un mínimo relativo, a las 19h, que también se aleja de la tendencia femenina en casi 2 puntos porcentuales. Todo lo anterior es reflejo del análisis gráfico derivado de la tabulación de datos (Figura 2).

3.3. Estructura espacial

La IP (Internet Protocol) es una variable de análisis que se ha mostrado relativamente débil, especialmente tras la depuración. Reduce a 861 (96,9%) las mujeres controladas, pero sobre todo a los hombres, que descienden a 159 (53,9%). El resto presentan en dicho campo valores nulos, no ajustados al formato de dicho protocolo (x.x.x.x, con x=[0..255]), o sin equivalencia geográfica (no figuran en las bases de datos espaciales).

Para dicha decodificación se ha recurrido a IPInfo.io, servicio especializado de reconocido prestigio en el sector. Las 12902 IP distintas válidas describen la siguiente estructura por ID y sexo (Tabla 3):

Figura 2

Estructura diaria de eventos normalizada por sexo y ordenada por ID (columnas).

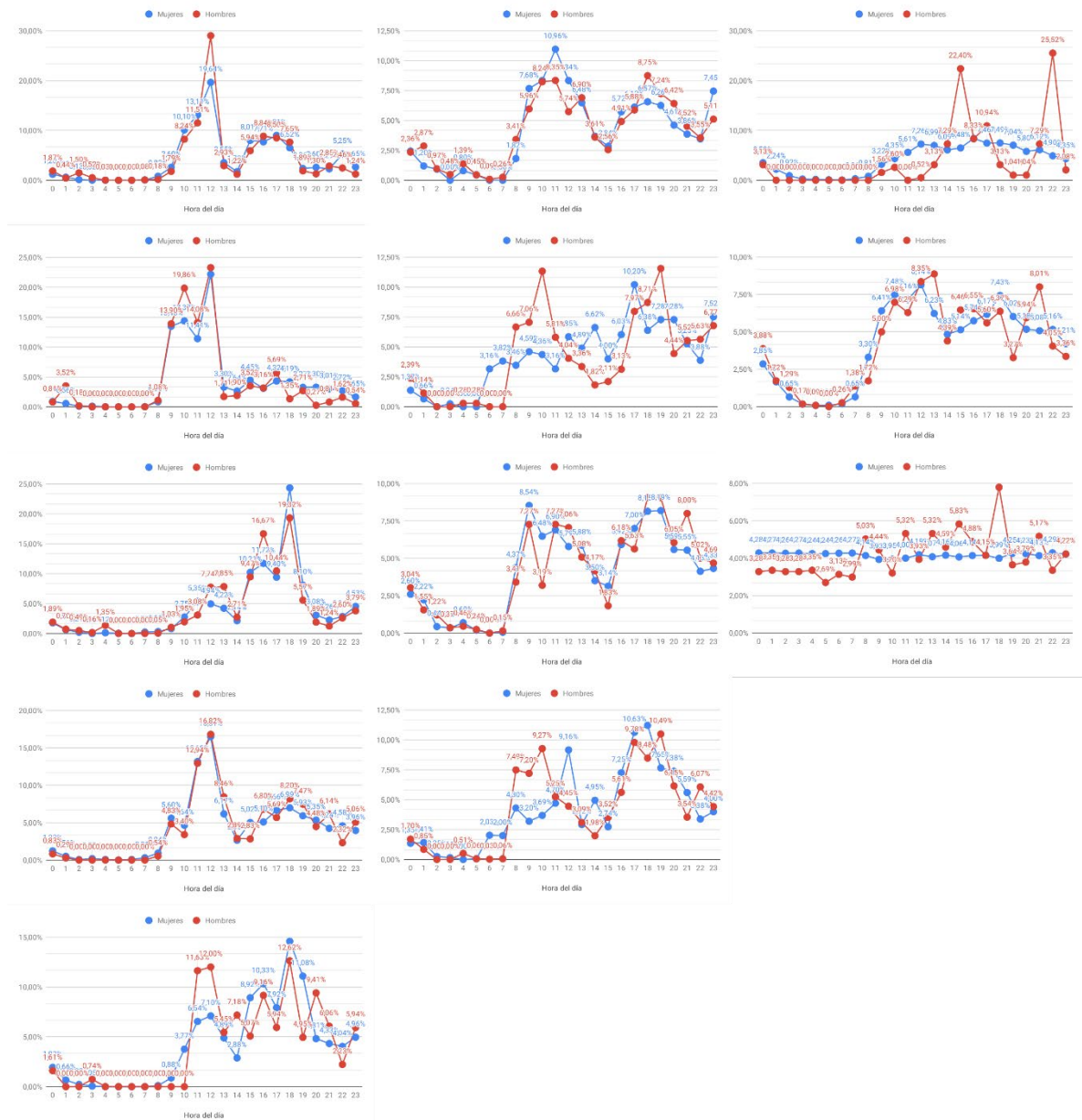


Tabla 3

Descripción del comportamiento de IP por ID y sexo.

ID	Mujeres			Hombres			Total		
	Alumnado	IP	Ratio	Alumnado	IP	Ratio	Alumnado	IP	Ratio
A1	94	1046	11,1	17	192	11,3	111	1191	10,7
A2	52	694	13,3	5	82	16,4	57	741	13,0
A3	47	648	13,8	11	112	10,2	58	725	12,5
A4	49	646	13,2	10	124	12,4	59	738	12,5
A5	44	263	6,0	7	31	4,4	51	283	5,5
B1	28	239	8,5	44	373	8,5	72	602	8,4

B2	18	118	6,6	21	113	5,4	39	226	5,8
B3	47	486	10,3	23	266	11,6	70	733	10,5
B4	18	148	8,2	21	151	7,2	39	289	7,4
C1	161	3718	23,1	5	18	3,6	166	3734	22,5
C2	151	2391	15,8	13	189	14,5	164	2568	15,7
C3	180	2416	13,4	8	67	8,4	188	2472	13,1
<i>Total</i>	<i>861</i>	<i>12066</i>	<i>14,0</i>	<i>159</i>	<i>1536</i>	<i>9,7</i>	<i>953</i>	<i>12902</i>	<i>13,5</i>

El rango de la ratio general se fija entre 5,5 (A5) y 22,5 (C1), con una amplitud de intervalo de 17 y superior al cuádruple del valor mínimo. Por tanto, se trata de una distribución no concentrada, fuertemente influenciada por ID y de forma marginal por el sexo. Con una salvedad notable, C1 (23,1 vs 3,6). Los valores absolutos no son útiles, dada la desproporción interna por sexos, especialmente en las ID del grupo A y C. Cuando dichas muestras están menos polarizadas, grupo B, las medias tienden a la convergencia.

El proceso de geolocalización ha reducido aún más el número de IP utilizables (9176), el resto no han podido ser debidamente decodificadas. La descripción de variables espaciales final es la siguiente, de menor a mayor escala: países (13); regiones (41); núcleos urbanos (629); códigos postales (ZIP): 668. A los que se suman las coordenadas geográficas y la zona horaria.

La distribución geográfica por países es poco significativa, el 99.29% de las IP utilizadas por mujeres y el 97.88% de las utilizadas por varones son se encuentran dentro de España. Estados unidos tiene relevancia marginal para éstos últimos (1.73%), y se añade Gran Bretaña en el caso de las mujeres (0.31% y 0.24% respectivamente). El resto tienen carácter testimonial.

Por regiones sí es posible detectar patrones espaciales de concentración a partir de los datos agregados (Tabla 4).

Madrid y Andalucía son las dos comunidades más relevantes, seguidas a distancia por Cataluña y Valencia. Inglaterra, considerada aquí por el servicio debido a su integración estatal en Gran Bretaña, figura como primera región extranjera. Por delante de Melilla y Ceuta, las menos relevantes cuantitativamente de todo el país.

Por sexos, es posible comprobar pequeñas discrepancias en la distribución: Madrid es más relevante para las mujeres que para los hombres, con más de 8 puntos porcentuales de diferencias. Situación que se invierte con Andalucía, con una diferencia similar. Las IP activadas por mujeres son mayoría en casi todas las demás regiones, salvo Galicia, Islas Baleares, País Vasco, Extremadura, Navarra, Virginia y Cantabria. El resto carecen de representatividad estadística para la muestra.

Tabla 4

Descripción del comportamiento de IP por ID y sexo

Región	F		M		Total		Región	F		M		Total	
	Abs	Rel	Abs	Rel	Abs	Rel		Abs	Rel	Abs	Rel	Abs	Rel
Madrid	2804	33,3	194	25,8	2998	32,7	New York	3	0,0	0,0	3	0,0	
Andalucía	2594	30,8	286	38,0	2880	31,4	Capital Region	1	0,0	1	0,1	2	0,0
Cataluña	808	9,6	70	9,3	878	9,6	Ceuta	2	0,0	0,0	2	0,0	
Valencia	426	5,1	28	3,7	454	4,9	Ankara	1	0,0	0,0	1	0,0	
Castille and León	285	3,4	17	2,3	302	3,3	Connaught	1	0,0	0,0	1	0,0	
Galicia	258	3,1	29	3,9	287	3,1	Coquimbo Region	1	0,0	0,0	1	0,0	
Canary Islands	255	3,0	17	2,3	272	3,0	Faro	1	0,0	0,0	1	0,0	
Castille-La Mancha	157	1,9	7	0,9	164	1,8	Flanders		0,0	1	0,1	1	0,0
Balearic Islands	119	1,4	26	3,5	145	1,6	Hauts-de-France	1	0,0	0,0	1	0,0	
Basque Country	120	1,4	23	3,1	143	1,6	Île-de-France	1	0,0	0,0	1	0,0	
Extremadura	125	1,5	14	1,9	139	1,5	Istanbul	1	0,0	0,0	1	0,0	
Asturias	118	1,4	3	0,4	121	1,3	Karpoš	1	0,0	0,0	1	0,0	
Aragon	116	1,4	3	0,4	119	1,3	Lisbon	1	0,0	0,0	1	0,0	
Murcia	69	0,8	5	0,7	74	0,8	North Rhine-Westphalia	1	0,0	0,0	1	0,0	
La Rioja	46	0,5	2	0,3	48	0,5	Porto	1	0,0	0,0	1	0,0	
Navarre	33	0,4	9	1,2	42	0,5	Santiago Metropolitan	1	0,0	0,0	1	0,0	
Virginia	13	0,2	13	1,7	26	0,3	Thuringia	1	0,0	0,0	1	0,0	
Cantabria	16	0,2	3	0,4	19	0,2	Vaud		0,0	1	0,1	1	0,0
England	19	0,2		0,0	19	0,2	Wales	1	0,0	0,0	1	0,0	
Melilla	12	0,1	1	0,1	13	0,1	Washington	1	0,0	0,0	1	0,0	
California	9	0,1		0,0	9	0,1	<i>Suma total</i>	<i>8423</i>	<i>1,0</i>	<i>753</i>	<i>1,0</i>	<i>9176</i>	<i>1,0</i>

Los núcleos urbanos más visitados corresponden en parte con las regiones más visitadas, pero también aparecen otros de regiones secundarias, lo que indica un elevado grado de concentración en la distribución interna de estas últimas. Así, encontramos Madrid y Sevilla claramente destacadas (27,1 y 11,7% respectivamente), seguidas por Málaga (4,1%), Barcelona (3,9%), Córdoba (1,5%), Valencia y Granada (ambas con 1,3%). El resto presenta valores inferiores al 1%.

Las IP usadas por mujeres vuelven a ser mayoría en Madrid (27,7% vs 20,3%), así como en Sevilla y Barcelona (12% vs 8%, y 3,9% vs 3,7%). En el resto de las ciudades se invierte la tendencia: Málaga (3,7% vs 8,6%), Córdoba (1,3% vs 1,7%), Valencia (1,3% vs 1,7%) y Granada (1,2% vs 1,9%), si bien las diferencias, salvo en el primer caso, son poco

significativas. Para comprender la intensidad local de este fenómeno en su distribución espacial, se ha recurrido a la técnica de mapa de calor (Figura 3):

La rampa de color ha sido ajustada para permitir discernir los patrones locales de menor intensidad, ya que la concentración de las dos primeras ciudades es muy alta. Se detecta así un patrón marginal ligado a la costa mediterránea, que se extiende hasta la bahía de Cádiz. En dicho recorrido es posible señalar concentraciones locales de cierta entidad (costa del sol y levante), y entre las que destaca, claramente, el entorno de Barcelona (Mataró, Sabadell, Sitges).

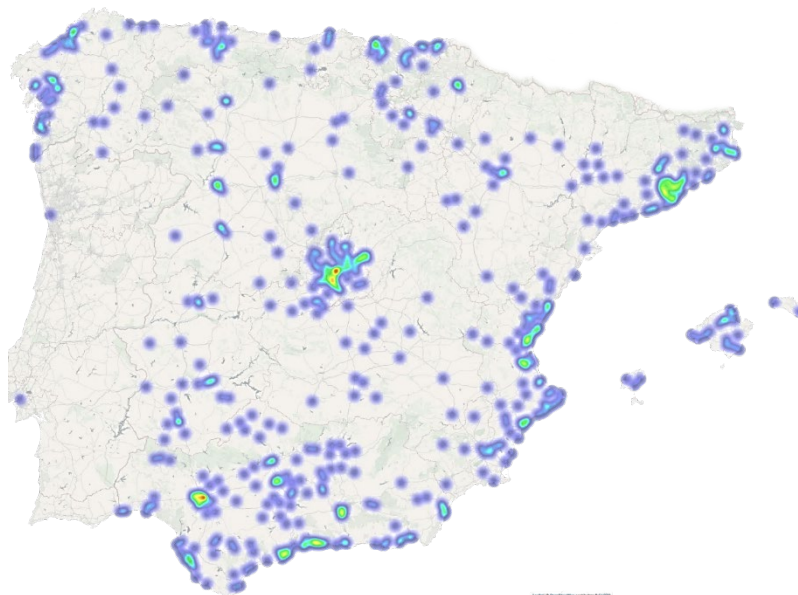
También permite comprobar que en las zonas de interior las situaciones puntuales son la norma, incluso en el entorno de ciudades relevantes como Valladolid, Salamanca, Zamora o León). Áreas de concentración menor se vuelven a detectar en torno a las ciudades gallegas, especialmente Coruña, Santiago de Compostela y Vigo.

3.4. Minería de datos

Ante la heterogeneidad de datos, cuatro autómatas de aprendizaje se lanzaron contra los mismos para realizar su validación, previa a la búsqueda de patrones en Orange: regresión logística, bosque aleatorio de clasificación, support vector machine (SVM) y Naive Bayes. Y uno adicional en Weka, REPTree. Para la aplicación y análisis de resultados de los distintos algoritmos se recurrió a los manuales de uso de cada aplicación.

Figura 3

Mapa de calor de las IP



La variable objetivo fue Sexo. Y en los casos en que una variable dio origen a otra derivada, se optó por la segunda y se excluyó de la matriz a la primera. Bajo dicho escenario, los resultados generales de los métodos fueron los reflejados en la Tabla 5.

Tabla 5

Estadísticos de los métodos probados sobre la clase sexo

Método	AUC	CA	F1	Precision	Recall
SVM Learner	0,510	0,420	0,492	0,759	0,420
Random Forest Learner (RFL)	0,839	0,884	0,868	0,869	0,884
Logistic Regression	0,784	0,869	0,852	0,848	0,869
Naive Bayes	0,769	0,837	0,840	0,843	0,837
REPTree	0,634	0,862	0,805	0,849	0,862

El mejor clasificador, para todos los indicadores, es RFL. Siguiendo a Tape (s.f.), es un clasificador bueno, al encontrarse en la horquilla .80-.90. En consecuencia, el sexo es una variable útil para determinar el comportamiento en Campus Virtual del alumnado. No obstante, valores por encima de .90 serían deseables si se quiere establecer patrones sólidos, ya que la tasa de falsos positivos es elevada en el caso de los hombres, como describe su matriz de confusión (Tabla 6).

Tabla 6

Matriz de confusión para RFL.

		Predicción			Estadísticos derivados			
		F	M	Σ				
Real	F	152820	4339	157159	Ratio F positivo	0,97	Ratio F negativo	0,03
	M	16771	8813	25584	Ratio M positivo	0,35	Ratio M negativo	0,66
	Σ	169591	13152	182743	Accuracy	0,96	Precision	0,67

Leyenda: los valores expresan número de eventos. El significado de los estadísticos se encuentra disponible en <https://bit.ly/2MfBCJN>.

La clasificación de mujeres es claramente más correcta que la de hombres (ratio de aciertos de 0,97 frente a 0,35). Por tanto, es más coherente estadísticamente el comportamiento femenino que el masculino, si bien la debilidad de la segunda muestra puede inferir errores de interpretación por inconsistencia. Los hombres son una fracción pequeña en comparación con las mujeres, y debe ser tenido en cuenta.

La minería realizada no arroja resultados claros, debido a la interferencia de la estructura semanal y diaria de las diferentes asignaturas. Con todo, es posible señalar algunas tendencias comunes: tanto al principio de la semana, como al principio del día, es más probable encontrar a una mujer trabajando con el Campus Virtual. Es más probable que sea un hombre quien acceda o trabaje en Campus Virtual durante la noche.

El estudio pormenorizado por cursos sí permite establecer conjuntos de reglas claras mediante algoritmos de carácter apriorístico. Tanto para la detección de patrones generales, que afectan a buen aparte de la clase, como marginales, útiles para la detección de subgrupos homogéneos. Ejemplos claros de ambas situaciones aparecen para ID = B3. Se encontraron 10 reglas de comportamiento por sexo con efecto sobre la componente de trabajo 'Taller', 9 para mujeres y 1 para hombres.

En el primer caso tienen carácter recurrente, cada una implica de 15 a 25 alumnas para días y horas específicos (lunes, martes o miércoles a las 9, 19 o 20h) o para una hora específica sin restricción de día (08h). Con solape solo parcial, de modo que establece un conjunto de patrones de comportamiento para la mayor parte de las mujeres, y valores de exactitud que oscilan entre 0,98 y 0,99.

Para los segundos la regla afecta a 7 alumnos, menos de un tercio de los mismos. Por tanto, se trata de un pequeño grupo de comportamiento bien definido (exactitud de 0.96) dentro de un subconjunto heterogéneo.

4. Discusión

El análisis de la relación entre calendario semanal y sexo ha señalado algunos aspectos interesantes. A pesar de que en líneas generales podemos hablar de un alto grado de semejanza entre ambos patrones, especialmente en los momentos de mayor intensidad, la situación ofrece matices cuando se centra la atención en las situaciones marginales. Algo que además ha sido bastante consistente en el tiempo.

En este sentido es interesante señalar que en las asignaturas de grado de pedagogía, grupo A, se observa un cambio de patrón en el último año: durante la semana las mujeres trabajan más en el campus virtual, mientras que los hombres lo hacen claramente más durante el fin de semana, siempre en términos relativos. Las razones de la ruptura del patrón de semejanza detectado no están claras: el profesorado es el mismo, así como la asignatura y su estructura temporal. Lamentablemente, los estudios longitudinales basados en cohortes no son posibles aquí, por lo que la influencia del contexto estudiantil es tan condicionante como indeterminada.

Los resultados exhibidos por las asignaturas del grupo B corresponden a un posgrado para alumnado Geografía e Historia, se trata por tanto de alumnado de una edad mayor y en un contexto de dedicación, en principio, a un número reducido de asignaturas si se compara con un curso académico de grado. Entendemos que este aspecto puede explicar la mayor homogeneidad de sus patrones. Ya que máximos absolutos y relativos nunca rebasan el 25%, cuando en el caso anterior casi llegan a suponer el 80% en algunos cursos.

Realizan un trabajo semanal menos concentrado, no exento pese a todo de tendencias generales, pero cuya matización por sexos debe realizarse con prudencia: en la mitad de los casos hay una preeminencia femenina en el trabajo de fin de semana, en otro es al contrario y la última expresa preeminencia femenina sólo en uno de los dos días, el sábado. Y durante la semana tampoco se constatan diferencias significativas, salvo para el lunes, que en tres de los cuatro cursos es momento de mayor interacción masculina.

Las asignaturas del grupo C pertenecen a un grado de magisterio, y en contra de lo que habría podido suponerse, expresan un patrón de comportamiento más semejante a B que a A. Algo que es destacable ya que A y B comparten una naturaleza curricular claramente vinculada a la Educación, mientras que B es una concreción explícitamente diseñada para estudiantes de las facultades de Filosofía y Letras, por tanto, sin un bagaje didáctico previo reglado. Con todo, alguna situación excepcional clara, como la que observable los martes del primer curso, no han podido ser explicadas a partir de la estructura del campus ni del conocimiento docente sobre la realidad del momento.

Aunque se esperaba que el análisis del comportamiento comparado por sexos y hora del día fuese más claro, lo cierto es que nuevamente la concomitancia ha sido la característica: la concentración horaria ha sido nuevamente más relevante en las asignaturas del grupo A y claramente inferior en las del B y C, a 10 puntos porcentuales o más de diferencia en sus máximos absolutos.

Sólo al abordar los comportamientos marginales, en ocasiones incluso principales, se observan diferencias entre los sexos: los máximos absolutos, momentos de gran intensidad de trabajo, suelen estar copados por los varones, mientras que, en términos generales, las mujeres presentan valores medios más elevados en horquillas temporales próximas a dichos máximos, expresando así una interacción más prolongada en el tiempo.

En conjunto, y para la muestra estudiada, el análisis revela la existencia tanto de patrones temporales diarios propios de cada asignatura como de patrones de sexo ligados a cursos específicos, pero no de forma estructural: hay cursos en los que la semejanza en el comportamiento por sexo explica mucho más que las diferencias; pero hay otros en los que las diferencias son claras, y en dichos casos se caracterizan habitualmente por la tendencia a la concentración de las interacciones de los hombres y la mayor distribución temporal de las interacciones por parte de las mujeres.

La IP se ha mostrado como una variable de gran interés para el análisis geográfico y la comparación por sexos. Se han constatado algunas diferencias claras tanto respecto a la dispersión como concentración espacial, que indican una mayor movilidad de las mujeres fuera de las localizaciones originales, si bien con una mayor tendencia a la concentración en los destinos. Por su parte, los hombres del presente estudio presentan una menor movilidad hacia el exterior, pero lo hacen sobre un mayor número de destinos; en ambos casos de forma proporcional.

5. Conclusiones

El estudio del comportamiento discente en los campus virtuales es un ámbito de estudio todavía en ciernes en el ámbito de la Educación Superior, especialmente en la Didáctica de la Geografía y las Ciencias Sociales. Sin embargo, a pesar de la complejidad que introduce, los elementos de conocimiento y reflexión que pone sobre la mesa lo hacen tanto útil como pertinente y necesario. No se puede ignorar el tiempo dedicado por los discentes ante terminales de todo tipo como parte de su proceso de aprendizaje, especialmente en consonancia con el modelo de créditos europeo ECTS.

La metodología utilizada necesita ser ampliada, especialmente en lo relativo a la minería de datos. Aproximación que ha demostrado ser interesante en un aspecto especialmente complejo de percibir: el de los comportamientos marginales homogéneos; es decir, el de las tendencias de pequeño grupo, difíciles de detectar por su propia naturaleza.

Se ha puesto de manifiesto que la robustez de este tipo de análisis se basa en la disponibilidad de volúmenes significativos de datos. Por ello, la correcta grabación y unificación de datos es un elemento clave que puede llegar a cuestionar la inclusión de los mismos, como ha ocurrido en una de las asignaturas del grupo C.

El análisis locacional a partir de la IP es, bajo nuestro criterio, una oportunidad para aprovechar información geográfica de gran interés, tanto desde una perspectiva básica como aplicada. Abre las puertas a, y requiere, futuras investigaciones que aprovechen su potencial desde perspectivas tan heterogéneas como la geografía humanista (Tuan, op. cit.) o la cuantitativa (Hägerstrand, 1985).

También ha sido determinante el sesgo de la variable principal, sexo, en un contexto, el educativo, donde la feminización es muy acusada. Esto facilita la exploración de una parte de la población, pero cuestiona la solidez analítica de la otra. Por ello, los errores de predicción en la matriz de confusión son tan desproporcionados entre hombres y mujeres.

En todo caso, la introducción de este análisis se presenta de forma natural y recupera un conjunto de datos no demasiado utilizado en el contexto geográfico o el educativo. Más allá de las semejanzas y discrepancias encontradas desde esta perspectiva comparada, a la que es necesario añadir ulteriores análisis antes de obtener argumentos explicativos sólidos, parece evidente que sí ofrece una riqueza suficiente como para ser investigada de forma sistemática, al menos, desde la didáctica de la geografía.

En conclusión, entendemos necesario seguir profundizando en este tipo de investigaciones, e ir sumando aportaciones de casos que permitan en un futuro el desarrollo de estudios más robustos sobre la consistencia de patrones por perfiles generales y singulares. O que complementen resultados de este tipo con análisis explicativos sobre las razones que, por ejemplo, llevan a hombres y mujeres a tener comportamientos divergentes, qué dificultades de aprendizaje en la especialidad pueden detectarse a partir de ahí y qué relación guarda todo ello con el rendimiento académico sobre el currículo.

Referencias.

- Abbott, M. L. (2011). *Understanding educational statistics using microsoft excel and SPSS (1st ed.)*. John Wiley & Sons Inc. URL: <https://bit.ly/31FpWbQ>
- Ahmadi, F., & Ahmad, S. (2013). Data mining in teacher evaluation system using WEKA. *International Journal of Computer Applications*, 63(10), 14-18. <https://doi.org/10.5120/10501-5268>
- Álvarez Vázquez, M. P., Álvarez-Méndez, A., Bravo-Llatas, C., Barrios, J. & Angulo Carrere, T. (2020). Tipologías de estudiantes de fisioterapia según el uso que hacen del campus virtual. *RIDU* 12, 74-81. <https://doi.org/10.1344/RIDU2020.12.8>
- Álvarez Vázquez, M. P. (2019). Estudiantes y Campus Virtual. Utilidad del Learning Analytics para identificar luces y sombras y caminar hacia metodologías adaptativas. *Proyecto de Innovación y Mejora de la Calidad Docente*, convocatoria 2019/20. Universidad Complutense, Madrid. <https://bit.ly/2NMEBK7>
- Arias, F., & Defiore, C. (2019). Avances y actualizaciones en torno al Campus Virtual de la Facultad de Ciencias Sociales de la UBA. *XXI Congreso de la Red de Carreras de Comunicación Social y Periodismo*. <https://bit.ly/3ina0kg>
- Badilla Quintana, M. G., Vera Sagredo, A., & Lytras, M. D. (2017). Pre-service teachers' skills and perceptions about the use of virtual learning environments to improve teaching and learning. *Behaviour & Information Technology*, 36(6), 575-588. <https://doi.org/10.1080/0144929X.2016.1266388>
- Ballesteros Regaña, C., Cabero Almenara, J., Llorente Cejudo, M., & Morales Lozano, J. A.. (2010). Usos del e-learning en las universidades andaluzas: estado de la situación y análisis de buenas prácticas. *Pixel-Bit. Revista de Medios y Educación*, 37, 7-18.
- Caballé, S., & Xhafa, F. (2013). Distributed-based massive processing of activity logs for efficient user modeling in a Virtual Campus. *Cluster computing*, 16(4), 829-844. <https://doi.org/10.1007/s10586-013-0256-9>
- Cabanillas, J. L., Luengo, R., & Carvalho, J. L. (2019). Análisis de los objetos de aprendizaje y de la percepción docente del campus virtual de la Universidad de Extremadura. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 6(2), 41-61.
- Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A. & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a Big Data framework. *Future Generation Computer Systems*, 90, 262-272. <https://doi.org/10.1016/j.future.2018.08.003>
- Capel, H. (2009). La enseñanza digital, los campus virtuales y la Geografía. *Ar@cne*, nº 125. <https://bit.ly/2NO9EIn>.
- Castro, M., Menacho, A., & Perez-Molina, C. (2018, April). Mining LMS students' data on online task-based master degree studies. *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 661-668). IEEE. <https://doi.org/10.1109/EDUCON.2018.8363294>.

- Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J. C. (2017). Procrastinating behavior in computer-based learning environments to predict performance: A case study in moodle. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01403>.
- Charanya, R. & Kesavan, M. (2019). Analysis of Factors Influencing the Virtual Learning Environment in a Sri Lankan Higher Studies Institution. *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 240-244. Colombo: Sri Lanka. <https://doi.org/10.23919/SCSE.2019.8842719>
- Cohen, R., Rahimi, I. D., & Zilka, G. C. (2019). Self-efficacy, challenge, threat and motivation in virtual and blended courses on multicultural campuses. *Issues in Informing Science and Information Technology*, 16, 71-95. <https://doi.org/10.28945/4295>
- Correa, J. M., & Paredes, J. (2009). Cambio tecnológico, usos de plataformas de e-learning y transformación de la enseñanza en las universidades españolas: la perspectiva de los profesores. *Revista de Psicodidáctica*, 14(2), 261-277.
- De la Iglesia-Villasol, M.C. (2019). Huellas de los estudiantes en las plataformas virtuales. Aplicación para evaluar una metodología de aprendizaje activo. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 22(3), 173-191. <http://dx.doi.org/10.6018/reifop.22.3.371341>
- Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, & Zupan B (2013) Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14(Aug), 2349–2353.
- Demšar, J., Zupan, B., Leban, G., & Curk, T. (2004, September). Orange: From experimental machine learning to interactive data mining. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 537-539). Springer.
- Desai, R., Chavan, A. & Tendulkar, H. (2020). Virtual Campus. *Studies in Indian Place Names*, 40(53), 268-270.
- Dobesova, Z. (2011). Visual programming language in geographic information systems. *Proceedings of the 2nd international conference on Applied informatics and computing theory*, pp. 276-280. World Scientific and Engineering Academy and Society (WSEAS). URL: <https://bit.ly/2MoDWxY>.
- Fakir, M., & Touya, K. (2014). Mining students' learning behavior in moodle system. *Journal of Information Technology Research (JITR)*, 7(4), 12-26. <https://doi.org/10.1109/ICCCT2.2014.7066695>
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics*, 1(1), 9. <https://doi.org/10.1186/s41044-016-0014-0>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71. <https://doi.org/10.1007/s11528-014-0822-x>.
- Gómez Collado, M. E., Contreras Orozco, L., & Gutiérrez Linares, D. (2016). El impacto de las tecnologías de la información y la comunicación en estudiantes de ciencias sociales: un estudio comparativo de dos universidades públicas. *Innovación educativa*, 16(71), 61-80.

- Hägerstrand, Torsten (1985). Time geography: focus on the corporeality of man, society and environment. Aida, Shūhei, ed. *The science and praxis of complexity: contributions to the symposium held at Montpellier, France, 9–11 May, 1984*. United Nations University Press. pp. 193-216.
- Hamutoglu, N., Gemikonakli, O., Duman, I., Kirsekiz, A. & Kiyici, M. (2020). Evaluating students experiences using a virtual learning environment: satisfaction and preferences. *Educational Technology Research and Development*, 68, 437–462. <https://doi.org/10.1007/s11423-019-09705-z>
- Kaur, P., Singh, M. & Josan, G. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 57, 500-508. <https://doi.org/10.1016/j.procs.2015.07.372>.
- Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353. <https://doi.org/10.1002/wcs.1350>
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1, 4-11. <https://goo.gl/1AjMzS>
- Lu, J., & Law, N. W. Y. (2012). Understanding collaborative learning behavior from moodle log data. *Interactive Learning Environments*, 20(5), 451-466. <https://doi.org/10.1080/10494820.2010.529817>
- Manne, S., Yeliseti, S., Kakarla, M., & Fatima, S. (2014). Mining VRSEC student learning behaviour in moodle system using datamining techniques. *International Conference on Computing and Communication Technologies*. India, Hyderabad, pp. 1-7. <https://doi.org/10.1109/ICCCT2.2014.7066695>
- Martínez Romera, D. D. (2017). Profesorado en Formación y Ambientes Educativos Virtuales. *Campus Virtuales*, 6(2), 69-78.
- Minguillón, J., Santanach, F., & Appel, M. C. (2016). Using learning analytics to support applied research and innovation in higher education. *EUA 2016 Annual Conference*, Galway, April 7th. URI: <http://hdl.handle.net/10609/47501>
- Morales Salas, R. E., Infante-Moro, J. C., & Gallardo-Pérez, J. (2019). La mediación e interacción en un AVA para la gestión eficaz en el aprendizaje virtual. *Campus Virtuales*, 8(1), 49-61.
- Orostica Verdugo, K. V. (2020). Entornos Virtuales de Aprendizaje: Campus UVM ONLINE. *UTE* 1, 6-21. <https://doi.org/10.17345/ute.2019.1>
- Pakanen, M., Alavesä, P., Arhippainen, L., & Ojala, T. (2020). Stepping out of the classroom: Anticipated user experiences of web-based mirror world like virtual campus. *International Journal of Virtual and Personal Learning Environments (IJVPLE)*, 10(1), 1-23. <https://doi.org/10.4018/IJVPLE.2020010101>
- Papoušek, J., Pelánek, R., & Stanislav, V. (2016). Adaptive geography practice data set. *Journal of Learning Analytics*, 3(2), 317-321. <https://doi.org/10.18608/jla.2016.32.17>

- Parise, P. (2016). A preliminary look at online learner behavior: What can the moodle logs tell us? *Bulletin of Kanagawa Prefectural Institute of Language and Culture Studies* (6), 15-31.
- Quintas Mendes, A., Bastos, G., Amante, L., Lebres, L., & Cardoso, T. (2019). *Virtual pedagogical model: development scenarios*. Universidade Aberta.
- Ramírez-Gallego, S., García, S., Mouriño-Talín, H., Martínez-Rego, D., Bolón-Canedo, V., Alonso-Betanzos, A., & Herrera, F. (2016). Data discretization: taxonomy and big data challenge. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(1), 5-21. <https://doi.org/10.1002/widm.1173>
- Ruipérez-Valiente, J. A., Halawa, S., Slama, R. & Reich, J. (2020). Using multi-platform learning analytics to compare regional and global MOOC learning in the Arab world. *Computer & Education*, 146,. <https://doi.org/10.1016/j.compedu.2019.103776>
- Sarduy Domínguez, Y., Vialart Vidal, N., Vidal Leo, M. & Paredes Esponda, E. (2020). Estrategias para el desarrollo de repositorios de recursos educativos abiertos del campus virtual de salud pública. *XVIII Convención y Feria Internacional de Informática en Salud*. Habana: Cuba.
- Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education*. Jisc. <https://bit.ly/2tnQyig>
- Sebastian, S., & Puthiydam, J. J. (2015). Evaluating students' performance by artificial neural network using WEKA. *International Journal of Computer Applications*, 119(23), 36-39. <https://doi.org/10.5120/21380-4370>
- Siddiq, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, 71, 151-166. <https://doi.org/10.1016/j.jnca.2016.04.008>
- Siemens, G. & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5).
- Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytics - A literature review. *ICTACT Journal on Soft Computing*, 5(4), 1035-1049. <https://doi.org/10.21917/ijsc.2015.0145>
- Sola-Martínez, T., Cáceres-Reche, M. P., Romero-Rodríguez, J. M., & Navas-Parejo, M. R. (2020). Estudio Bibliométrico de los documentos indexados en Scopus sobre la Formación del Profesorado en TIC que se relacionan con la Calidad Educativa. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 23(2). <https://doi.org/10.6018/reifop.418611>
- Taleb, I., Dssouli, R., & Serhani, M. A. (2015). Big data pre-processing: A quality framework. In *2015 IEEE International Congress on Big Data* (pp. 191-198). IEEE. <https://doi.org/10.1109/BigDataCongress.2015.35>
- Tape, T. G. (s.f.). *Interpreting diagnostic tests*. University of Nebraska Medical Center. URL: <https://bit.ly/2NESlcs>.
- Tartia, J. (2020). *The Temporality and Rhythmicity of Lived Street Space*. Tampere University.

- Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big data*, 2(1), 21. <https://doi.org/10.1186/s40537-015-0030-3>
- Tuan, Y. F. (1990). *Topophilia: A study of environmental perceptions, attitudes, and values*. Columbia University Press.
- Tuan, Y. F. (1977). *Space and place: The perspective of experience*. University of Minnesota Press.
- Thulin, E., Vilhelmson, B. & Schwanen, T. (2020). Absent Friends? Smartphones, Mediated Presence, and the Recoupling of Online Social Contact in Everyday Life. *Annals of the American Association of Geographers*, 110(1), 166-183. <https://doi.org/10.1080/24694452.2019.1629868>
- Vega Valverde, M. (2017). Aplicaciones didácticas de la Cronogeografía. A propósito de nuestra relación espacio - tiempo. *Repositorio Abierto de la Universidad de Cantabria*. <http://hdl.handle.net/10902/>

Como citar:

- Martínez-Romera, D., Cebrián de la Serna, M., & Priego de Montiano, G. (2021). The Influence of Students' Gender on the Use of Virtual Campuses. A Case Study. [La influencia del sexo en el uso de los campus virtuales. Estudio de caso]. *Pixel-Bit. Revista de Medios y Educación*, 60, 169-210. | <https://doi.org/10.12795/pixelbit.78445>