



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

Proyecto de Innovación y Mejora de la Calidad Docente

Convocatoria 2017

Nº de proyecto 245

Big data en educación II: metodologías adaptativas en el proceso de enseñanza-aprendizaje  
desde el diagnóstico del estudiante

Adolfo Hernández Estrada

Facultad de Comercio y Turismo

Estadística e Investigación Operativa II (Métodos de Decisión)

## 1. Objetivos propuestos en la presentación del proyecto

Este proyecto se presentó como continuidad a los llevados a cabo en convocatorias anteriores (conv. 2006: nº 390, conv. 2007: nº 559, conv. 2011: nº 246, conv. 2014: nº 244, conv 2015: nº 379). En estos proyectos se habían desarrollado materiales para el aprendizaje autónomo del alumno, una herramienta de aprendizaje teórico, dos herramientas de evaluación y autoevaluación, un paquete de vídeos que facilitan el aprendizaje en modelos de aula de aprendizaje inverso, siempre en el entorno del aprendizaje de la Estadística, tanto en el desarrollo de asignaturas convencionales como en el desarrollo de Trabajo Fin de Grado.

En particular, este proyecto es continuación directa del proyecto 83 de la convocatoria 2016: "Big data en educación: tipologías de los estudiantes a partir del estudio de las interacciones dentro del triángulo pedagógico". Este proyecto del curso pasado tuvo como objetivo la construcción de un modelo de análisis de los datos que facilita la plataforma Moodle, desde las distintas interacciones entre el docente, el estudiante y la materia desarrollada, de forma que esta información pudiera transformarse en conocimiento y su comprensión pudiera ayudar a la mejora de la práctica docente de profesores de otros campos del saber. Este proyecto del curso pasado, aunque se limitó al estudio de los cursos donde eran profesores los integrantes del equipo, nos mostró que el análisis de los datos puede ser un elemento de innovación en la mejora de los escenarios educativos vigentes, dado que gran parte de la información generada en la actualidad no aporta conocimientos ni beneficio por la falta de un análisis detallado. Es ahí donde se ancló el trabajo del proyecto de este curso, ya teniendo acceso a una amplia base de datos: desde el desarrollo de nuevos modelos de análisis de los datos que facilita Moodle, desde las interacciones entre el docente, el estudiante y los contenidos expuestos. En concreto, se abordó la detección precoz del abandono y el estudio del rendimiento.

Los objetivos que se enumeran a continuación fueron los que se incluyeron en el proyecto.

### Objetivo principal:

Plantear metodologías adaptativas en la práctica docente que permitan ajustarse a las características personales y sociales del estudiante, como sujeto que aprende.

### Objetivos secundarios:

1. Explicar el rendimiento académico no únicamente con tareas académicas, sino también desde lo que sucede fuera del acto educativo.
2. Utilizar los foros como medio de aprendizaje social de contenidos reales, como piloto de análisis para el diseño de otro tipo de prácticas.
3. Diseñar e implementar prácticas educativas innovadoras que faciliten al estudiante la aplicación real de los contenidos, utilizando herramientas tecnológicas.
4. Desarrollar un modelo de evaluación de competencias basado en el comportamiento individual del alumno y no solo en resultados de tareas académicas, utilizando técnicas Big data que permitan obtener una perspectiva completa de los datos del alumno provenientes de distintas fuentes.
5. Desarrollar modelos estadísticos de clasificación, que permitan detectar de manera previa posibles situaciones de abandono académico antes de que se materialicen.
6. Definir tipologías de estudiantes de acuerdo a su estilo de aprendizaje en otras facultades de la Universidad Complutense, teniendo como referencia -no como límite- el estudio realizado en la Facultad de Comercio y Turismo y en la Facultad de Ciencias Económicas y Empresariales.

## 2. Objetivos alcanzados

1. Explicar el rendimiento académico no únicamente con tareas académicas, sino también desde lo que sucede fuera del acto educativo.

Las tareas académicas desarrolladas en el aula nos han aportado hasta ahora una información parcial de la actuación del estudiante como sujeto que aprende. Gracias al uso de la plataforma digital Moodle, podemos recoger el trabajo y la interacción social del estudiante en relación a sus resultados de rendimiento (Chaparro, Iglesias y Pascual, 2010). Estos datos recogidos en Moodle nos permiten recopilar perfiles de comportamiento relacionados con la actividad del aula “con el fin de analizar cómo estos atributos afectan el nivel de actividad de los estudiante” (Hidalgo, 2018, p. 126), de manera que podremos en un futuro crear grupos de características similares para favorecer la personalización del aprendizaje.

Objetivo conseguido.

2. Utilizar los foros como medio de aprendizaje social de contenidos reales, como piloto de análisis para el diseño de otro tipo de prácticas.

Se han diseñado y realizado diferentes actividades para trabajar en grupo, utilizando las herramientas de Moodle “Foro”, como soporte virtual del debate, y “Wiki” para la autogestión de la creación de los grupos de trabajo. En concreto, se ha propuesto que los alumnos participen en foros en modalidad asincrónica, de naturaleza distinta tanto desde su planteamiento con soporte material, como en la tipología de actividad a realizar por los estudiantes, pero con un objetivo común, crear opinión científica relacionada con algunos aspectos de la Estadística. Para ello, se ha pedido a los participantes que ahonden sobre ciertos contenidos del programa de las asignaturas (conceptos como muestra y población, así como sobre las técnicas de inferencia, su aplicación y utilidad), a partir del debate virtual en torno al contenido de materiales reales (vídeos y notas de prensa) publicados por el Instituto Nacional de Estadística (INE). Con la finalidad de que las actividades se puedan desarrollar en distintos grupos de alumnos de forma homogénea, se ha redactado una guía común para todos los grupos, en la que se detallan el contenido de las actividades, los objetivos y competencias que se pretenden alcanzar y su evaluación (ver Anexo 1.1.). La participación en estas actividades implica la aceptación de un contrato didáctico en el que se recogen los compromisos adquiridos por alumnos y profesores (ver Anexo 1.2.).

Objetivo conseguido.

3. Diseñar e implementar prácticas educativas innovadoras que faciliten al estudiante la aplicación real de los contenidos, utilizando herramientas tecnológicas.

Conocer al estudiante nos facilita el diseño de actividades que se ajusten de una manera real a su estilo de aprendizaje y sus posibilidades de trabajo diarias. Además conviene recordar que un aprendizaje basado en competencias facilita la aplicación de contenidos a lo que será la práctica futura en el entorno laboral, por lo que nuestro objetivo se ha encaminado de alguna manera a este trabajo bidireccional, la práctica real y la personalización de la propia tarea. Los contenidos de las asignaturas con las que hemos trabajado tienen unas características que las configuran como especiales, dado que no tiene sentido enseñar contenidos estadísticos sin situarlos en un contexto determinado que facilite no únicamente el trabajo de comprensión y análisis, sino de interpretación. Por lo tanto las actividades se basan en la metodología de aprendizaje basado en problemas (AbP), donde a los estudiantes se les facilita una situación de datos reales que se sitúan en un determinado contexto, y que requieren un análisis fundamentado. El hecho de que el trabajo se configure en grupos hace que además se pueda mejorar la comunicación entre los estudiantes, y con los profesores y además se incentive la “participación e implicación de los alumnos mediante el planteamiento de problemas reales y motivadores” (Martín-Peña, Díaz-Garrido y Sánchez-López, 2015).

Objetivo conseguido.

4. Desarrollar un modelo de evaluación de competencias basado en el comportamiento individual del alumno y no solo en resultados de tareas académicas, utilizando técnicas Big data que

permitan obtener una perspectiva completa de los datos del alumno provenientes de distintas fuentes.

Se ha construido una base de datos relacional con información combinada de tres fuentes: registros de actividad en la plataforma educativa Moodle-UCM de todos los alumnos y profesores en todas las asignaturas de los Grados impartidos en la Facultad de Comercio y Turismo (FCyT) de la Universidad Complutense de Madrid (UCM), datos de la aplicación de Gestión Académica UCM (GEA) relativos a la calificación final en las convocatorias ordinaria y extraordinaria de los alumnos correspondientes y datos de la aplicación de Gestión Académica UCM (GEA) de tipo socio-económico (sexo, nota de acceso a la Universidad, año de ingreso en el Grado y condición o no de becario de los alumnos correspondientes). Para evaluar los resultados de aprendizaje de las actividades desarrolladas, se han elaborado rúbricas *ad hoc* (Anexo 2) que incorporan la información proveniente de distintas fuentes que integran ítems propios de evaluación de contenido, resultados del análisis de los datos de los registros de Moodle de los foros (número de intervenciones del grupo y de cada miembro, tipo de acción realizada, fecha, IP UCM o externa), tipología de estudiante de los listados de Gestión Académica UCM (erasmus, repetidor, número de convocatorias agotadas,...) y nuevas variables integradas en un cuestionario final de autoevaluación del trabajo individual y en grupo por parte del alumno. Así, se *combinan variables visibles para el estudiante, como es su propio trabajo, con variables integradas en la base de datos*. Esta evaluación personalizada supone contar con una gran cantidad de datos que, analizados de forma conveniente, aportan información relevante para la evaluación de la competencia de trabajo en equipo.

Objetivo conseguido.

5. Desarrollar modelos estadísticos de clasificación, que permitan detectar de manera previa posibles situaciones de abandono académico antes de que se materialicen.

Se han utilizado los algoritmos más referenciados en la literatura relacionada con este tipo de métodos: análisis discriminante lineal (LDA), regresión logística (LOGISTIC), modelo de los k vecinos más próximos (KNN MODEL), análisis discriminante cuadrático (QDA), redes neuronales (RED NEURONAL) y árboles aleatorios (RANDOM FOREST). Puede encontrarse el código de estos métodos en lenguaje Python en el Anexo 3.

Objetivo conseguido.

6. Definir tipologías de estudiantes de acuerdo a su estilo de aprendizaje en otras facultades de la Universidad Complutense, teniendo como referencia -no como límite- el estudio realizado en la Facultad de Comercio y Turismo.

Los resultados preliminares relativos a este objetivo nos hicieron reflexionar sobre la conveniencia de definir primero tipologías de profesores. Por tanto se han llevado a cabo estudios clúster de profesores y, sobre los dos grupos obtenidos, nuevos resultados clúster para definir tipologías de alumnos. Pueden encontrarse los resultados en el Anexo 4.

Objetivo conseguido.

REFERENCIAS:

Chaparro, J., Iglesias, S. & Pascual, F. (2010, October). Uso del registro de actividad de Moodle para un estudio del rendimiento académico de alumnos en entornos en línea y presencial. In *4th International Conference On Industrial Engineering and Industrial Management* (pp. 753-760).

Hidalgo, B. G. (2018). Minería de datos en los Sistemas de Gestión de Aprendizaje en la Educación Universitaria. *Campus Virtuales*, 7(2), 115-128.

Martín-Peña, M.L., Díaz-Garrido, E. y Sánchez-López, J.M. (2015). Coordinación interdisciplinar mediante aprendizaje basado en problemas. Una aplicación en las asignaturas dirección de producción y estadística empresarial. *Revista de Investigación Educativa*, 33(1), 163-178.

### **3. Metodología empleada en el proyecto**

El plan de trabajo siguió una cronología temporal:

1. Selección de titulaciones. y 2. Selección de materias.

El apoyo del vicerrector de Tecnologías de la Información de la UCM y de la decana de la Facultad de Comercio y Turismo (FCyT) nos ha posibilitado tener acceso institucional a todas las asignaturas de todos los estudios de Grado que se imparten en esta Facultad. Dado el carácter de estas titulaciones los datos analizados se corresponden con una gran variedad de áreas de conocimiento: asignaturas impartidas por un total de 34 departamentos diferentes de la UCM. Por tanto, los resultados obtenidos pueden ser extensibles a otros campos de conocimiento.

3. Planteamiento e implementación de prácticas innovadoras.

En función de la tipología de las materias seleccionadas, hemos diseñado prácticas ajustadas al perfil individual y grupal de los estudiantes, garantizando su implementación en grupos experimentales frente a grupos control. En todo este proceso de trabajo se ha prestado especial atención a la situación de abandono, aunque cabe señalar que este trabajo tendrá resultados más eficientes cuando podamos determinar que hemos desarrollado un estudio longitudinal en varios cursos académicos. Hemos partido de un primer escenario en la plataforma virtual, en torno a los foros de aprendizaje, considerando que es éste un medio de aprendizaje colaborativo que a la vez se ajusta a las características personales de cada estudiante por la autonomía que puede suponer. Así, se han utilizado una muestra de asignaturas como piloto de una actividad de aprendizaje cooperativo utilizando simulaciones y/o datos de entornos de trabajo reales. El trabajo que realicen los estudiantes en este medio se ha analizado en relación con la interacción global en la plataforma y sus características personales, lo que nos ha facilitado el diseño posterior de otras prácticas de tipología colaborativa en base a los resultados obtenidos, así como el desarrollo de un modelo de evaluación de competencias. Se ha detallado parte de este trabajo en la valoración de los objetivos.

4. Obtención de los datos.

Se ha contado con la información correspondiente a la Facultad de Comercio y Turismo de la UCM (aproximadamente 3000 alumnos y más de 150 profesores) para el curso 2016-2017. La información institucional a la que se ha tenido acceso recoge tres aspectos distintos, tal como se ha señalado en el objetivo 4. Debe tenerse en cuenta que cada vez que un alumno o profesor accede a Moodle se genera un registro, por lo que el volumen manejado de información se corresponde con millones de registros, a los que hay que añadir los datos provenientes de las otras fuentes de información. A partir de esta información se ha construido una base de datos relacional. Un esquema de su composición puede encontrarse en el Anexo 5. Los trabajos de construcción y depuración de esta base de datos se han hecho en lenguaje SQL. Puede encontrarse un ejemplo en el Anexo 6.

5. Análisis de datos.

A partir de la base de datos construida en el paso anterior pueden, en primer lugar, desarrollarse métodos de visualización de datos. Puede encontrarse un ejemplo en el Anexo 7. A continuación se desarrollaron modelos estadísticos de clasificación, que permitieron detectar posibles situaciones de abandono académico antes de que se materialicen. Tal como se ha comentado en la sección de objetivos se utilizaron los algoritmos más referenciados en la literatura relacionada con este tipo de métodos. Por último, se definieron tipologías de profesores y, para cada una de ellas, de estudiantes de acuerdo a su estilo de aprendizaje.

6. Definición de recomendaciones.

En base al análisis efectuado se está elaborando un plan de ruta que recoge prácticas que faciliten el aprendizaje comprensivo en entornos similares a la realidad del futuro desempeño profesional y recomendaciones que posibiliten la reducción del número de abandonos. Estas recomendaciones se han complementado con una búsqueda de referencias que han facilitado el análisis de modelos similares en otros entornos. Algunos de los documentos analizados han sido:

Álvarez-Pérez, P. R. & López-Aguilar, D. (2017). Estudios sobre deserción académica y medidas orientadoras de prevención en la Universidad de la Laguna (España). *Paradigma*, 38(1), 48-71.

Cervero, A., Bernardo, A., Esteban, M., Tuero, E., Carbajal, R., & Núñez, J. C. (2017). Influencia en el abandono universitario de variables relacionales y sociales. *Revista de Estudios e Investigación en Psicología y Educación*, 12, 46-49.

Freixa, M., Llanes, J. & Venceslao, M. (2018). El abandono en el recorrido formativo del estudiante de ADE de la Universidad de Barcelona. *Revista de Investigación Educativa*, 36(1), 185-202.7.

Himmel, E. (2018). Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación*, 17, 91-108.

Orozco, G. H., Sosa, M. R. & Martínez, F. (2018). Modelos didácticos en la educación superior: Una realidad que se puede cambiar. *Profesorado, Revista de Currículum y Formación del Profesorado*, 22(2), 447-469.

Tuero, E., Cervero, A., Esteban, M., & Bernardo, A. (2018). ¿Por qué abandonan los alumnos universitarios? Variables de influencia en el planteamiento y consolidación del abandono. *Educación XX1*, 21(2).

## 7. Conclusiones.

Se establecieron conclusiones a partir de todos los puntos anteriores.

Tanto la metodología como los resultados se presentaron en diferentes congresos y una publicación (ver sección 5).

## 4. Recursos humanos

La mayoría del equipo lo forman profesores de las Facultades de Comercio y Turismo y de Ciencias Económicas y Empresariales. Además, se cuenta con la participación de 3 estudiantes al considerarse que la inclusión de su visión mejorará este estudio en prácticamente todas sus fases. La totalidad de los profesores tuvieron sus asignaturas dadas de alta en el Campus Virtual de la UCM.

Este problema presentaba todas las características asociadas a lo que en los últimos años ha dado en denominarse “Big Data” (Mayer-Schönberger y Cukier, 2013). Estas técnicas tienen su origen en métodos estadísticos aplicados en bases de datos con alta dimensión. En ese sentido, la pertenencia de la mayoría de los profesores al departamento de Estadística e Investigación Operativa II (Métodos de Decisión) resultó totalmente apropiada.

Además, varios de los profesores participantes en el proyecto tenían experiencia específica en el tratamiento de datos multivariantes, tanto a nivel docente (impartición de la asignatura “Técnicas Estadísticas Multivariantes en Gestión Comercial”, de tercer curso del grado en Comercio) como investigador.

Cabe destacar la presencia en el equipo de una profesora del departamento Sociología III (Estructura Social. Sociología de la Educación) que, además de aportar su experiencia en estos temas, es la actual coordinadora del Máster Oficial UCM en Consumo y Comercio, lo que ha aportado la perspectiva de este ámbito formativo. También formó parte del equipo una profesora, doctora en Ciencias de la Educación y actual coordinadora del Campus Virtual en la Facultad de Comercio y Turismo. Desde ese cargo pudimos coordinar toda la transferencia de información institucional. También formó parte del equipo un profesor del departamento de Organización de Empresas, miembro del Grupo de investigación de la Comunidad de Madrid: Social Bigdata-CM (<http://socialbigdata.transyt-projects.com/>) con experiencia en software e-learning tras su paso por la empresa Oracle.

El equipo docente lo completaron tres profesores más: uno de la Universidad Politécnica de Madrid, otro de la Universidad Internacional de La Rioja (con amplia experiencia en el uso de plataformas virtuales, dado el carácter no presencial de las enseñanzas que imparten) y una profesora de la Universidad de Alcalá, especialista en las áreas de didáctica de las matemáticas y educación

adaptativa. En definitiva, se ha tratado de reunir especialistas tanto en educación como en el tratamiento cuantitativo y cualitativo de datos.

Dada la complejidad técnica del problema planteado, también se incorporó al equipo un miembro del PAS de la UCM. Se trata de un técnico especialista de los servicios informáticos, con experiencia previa de varios años como programador, desarrollador de software educativo y técnico de soporte interno para Microsoft Ibérica SRL; actualmente es T.E. II de aulas y servicios informáticos de la UCM.

Por último, aunque no fue formalmente un miembro del equipo, ha trabajado con nosotros durante el desarrollo del proyecto un colaborador en tareas informáticas y estadísticas. Su nombre es Carlos Enrique Huane Estrada, y él fue el principal responsable de la construcción de la base de datos relacional y de la implementación en Python de los diferentes métodos estadísticos utilizados.

## **5. Desarrollo de las actividades**

Las fases de trabajo siguieron los objetivos expuestos anteriormente, y que consideramos se describen de manera suficiente tanto en el punto 2 de objetivos conseguidos como en el punto 3 de metodología.

Respecto a la difusión, durante el desarrollo de este proyecto se han presentado los primeros avances en congresos:

Pérez, M., Martínez, E. y Arteaga, B. (2017). Description of an experience in university classroom teaching/learning: can the learning of statistics be improved through virtual forums? Conferencia: ICERI2017, 10a International Conference of Education, Research and Innovation. Sevilla, España. doi: 10.21125 / iceri.2017.1105

La publicación a la que ha dado lugar el trabajo de este Proyecto ha sido:

Martínez, E., Pérez, M. y Arteaga, B. (2018). El entorno virtual como espacio de aprendizaje: un recurso complementario para evaluar la competencia de trabajo en equipo. En A.I. Allueva Pinilla y J.L. Alejandro Marco. *Casos de éxito en aprendizaje ubicuo y social mediado con tecnologías*. Zaragoza: Universidad de Zaragoza.

## **6. Anexos**

### **Anexo 1. Detalle de documentos de gestión de foros.**

#### **1.1. Detalle del documento para estudiantes referido al Foro “análisis de datos”.**

Utilizando bases de datos reales, se pedirá a los grupos de trabajo que den respuesta a problemas aplicando diferentes técnicas estadísticas aprendidas en la asignatura.

El desarrollo del trabajo en este Foro consta de dos fases:

1. Fase de resolución de los problemas. Cada grupo puede contribuir al foro publicando información que consideren útil, comentando su experiencia (técnicas estadísticas utilizadas, hipótesis iniciales asumidas,...), planteando dudas y tratando de aclarar las dudas de otros grupos.

2. Fase de resultados. En esta fase cada grupo debe publicar en el foro los resultados obtenidos (incluyendo un resumen detallado del trabajo estadístico realizado). Una vez que todos los resultados estén publicados, cada grupo debe hacer una valoración crítica de los resultados presentados (defensa de las técnicas estadísticas utilizadas, posibles errores en la metodología, hipótesis no asumidas, ....)

El objetivo de este Foro es que el alumno identifique los aspectos de la realidad propuesta a los que puede dar solución con el conocimiento adquirido en la asignatura y que detecte aspectos para los cuales le falta aún conocimiento.

### **1.2. Detalle del contrato didáctico para trabajo en los foros**

La participación en esta actividad implica un compromiso por parte del alumno y de la profesora, que se materializa en:

El alumno se compromete a cumplir las siguientes normas de participación:

a- Cualquier intervención conlleva respeto al resto de los grupos y a sus intervenciones.

b- Trabajar en grupo:

b-1. cada grupo estará constituido por tres alumnos.

b-2. los grupos deben registrarse en el campus virtual, en la wiki (link) creada al efecto.

b-3. a la hora de formar los grupos se debe tener en cuenta que:

- se considera que las aportaciones al foro son el resultado del trabajo en equipo y no del miembro que publica.

- se califica el trabajo de grupo, es decir, todos los miembros del grupo tendrán la misma calificación.

- no se permite cambiar la composición del grupo.

- se deben respetar los plazos de participación en los foros.

c- Cumplir los plazos:

c-1. cada grupo debe participar con el número de veces que como mínimo se establezca en las instrucciones específicas de cada foro y siempre dentro de los plazos establecidos.

c-2. las publicaciones deben estar espaciadas en el tiempo, evitando la participación puntual, esto es, se debe participar durante todo el proceso y no sólo a inicio o final.

d. Participar activamente. Las publicaciones deben:

d.1. estar relacionadas con el tema y tener fundamentación estadística.

d.2. mencionar las intervenciones de otros grupos (interacción con el resto de grupos participantes).

d.3. estar argumentadas (por ejemplo, no se considera participación activa opiniones del tipo "estoy o no de acuerdo con lo que dice el grupo x")

d.4. Al menos una de las publicaciones debe proporcionar algún tipo de material adicional (videos relacionados, artículos sobre el tema, citas de autor, webs, otros recursos).

e. Responder al cuestionario inicial Técnicas de aprendizaje enlace(link)

f. Evaluar su trabajo personal y el trabajo en equipo realizado contestando de forma individual y a final de curso a un cuestionario.



## Anexo 2. Rúbrica de evaluación

Aspectos formales										SI	NO	OBSERVACIONES			
Cumple los plazos establecidos															
Interviene la última semana															
Se ha detectado plagio															
Incluye recurso															
Calidad ortográfica															
Incluye citas de autores, libros, artículos, páginas web, ...															
Valoración de los aspectos formales															
Aspectos específicos										SI	NO	OBSERVACIONES			
Responden a las intervenciones de otros grupos															
Su aportació ha sido citada															
Plantean nuevas ideas de debate															
												Muy bien	Bien	Apto	No apto
Calidad de la argumentación ESTADÍSTICA															
Calidad del material aportado															
Relación con el tema															
Originalidad de la intervención															
Valoración de los aspectos específicos															
Registros Moodle															
Algún contenido ha sido publicado	Mensaje actualizado	Mensaje creado	Módulo de curso visto	Tema visto	Mensaje creado	Total	Intervenciones	Nuevas Variables							
								Intervienen todos componentes	Interviene un solo componente	Intervención continua	Intervención por periodos	Homogeneidad publicaciones			
<b>Autoevaluación</b>															
El foro ha facilitado el trabajo en grupo															
Considero que todos los miembros del grupo han estado igualmente involucrados															
Valoración global del foro															

## Anexo 3. Código en lenguaje Python de los métodos de clasificación utilizados

# LDA

```
lda = LinearDiscriminantAnalysis(n_components=2)
```

```
scores_lda = cross_val_score(lda, X, y, cv=5)
```

```
lda.fit(X_train, y_train)
```

```
method.loc['score', 'lda'] = scores_lda.mean(); method.loc['std', 'lda'] = scores_lda.std() * 2
```

```
Z['lda'] = lda.predict(X_test)
```

```
np.unique(df.variableDependiente, return_counts=True)
```

# LOGISTIC

```
log = LogisticRegression(penalty='l2', C=0.1)
```

```
scores_log = cross_val_score(log, X, y, cv=5)
```

```
method.loc['score', 'log'] = scores_log.mean(); method.loc['std', 'log'] = scores_log.std() * 2
```

```
log.fit(X_train, y_train)
```

```
Z['log'] = log.predict(X_test)
```

# K-NN MODEL

```
knn = KNeighborsClassifier(n_neighbors=10)
```

```
scores_knn = cross_val_score(knn, X, y, cv=5)
```

```

method.loc['score', 'knn'] = scores_knn.mean(); method.loc['std', 'knn'] = scores_knn.std() * 2
knn.fit(X_train, y_train)
Z['knn'] = knn.predict(X_test)

# QDA
qda = QuadraticDiscriminantAnalysis()
scores_qda = cross_val_score(qda, X, y, cv=5)
method.loc['score', 'qda'] = scores_qda.mean(); method.loc['std', 'qda'] = scores_qda.std() * 2
qda.fit(X_train, y_train)
Z['qda'] = qda.predict(X_test)

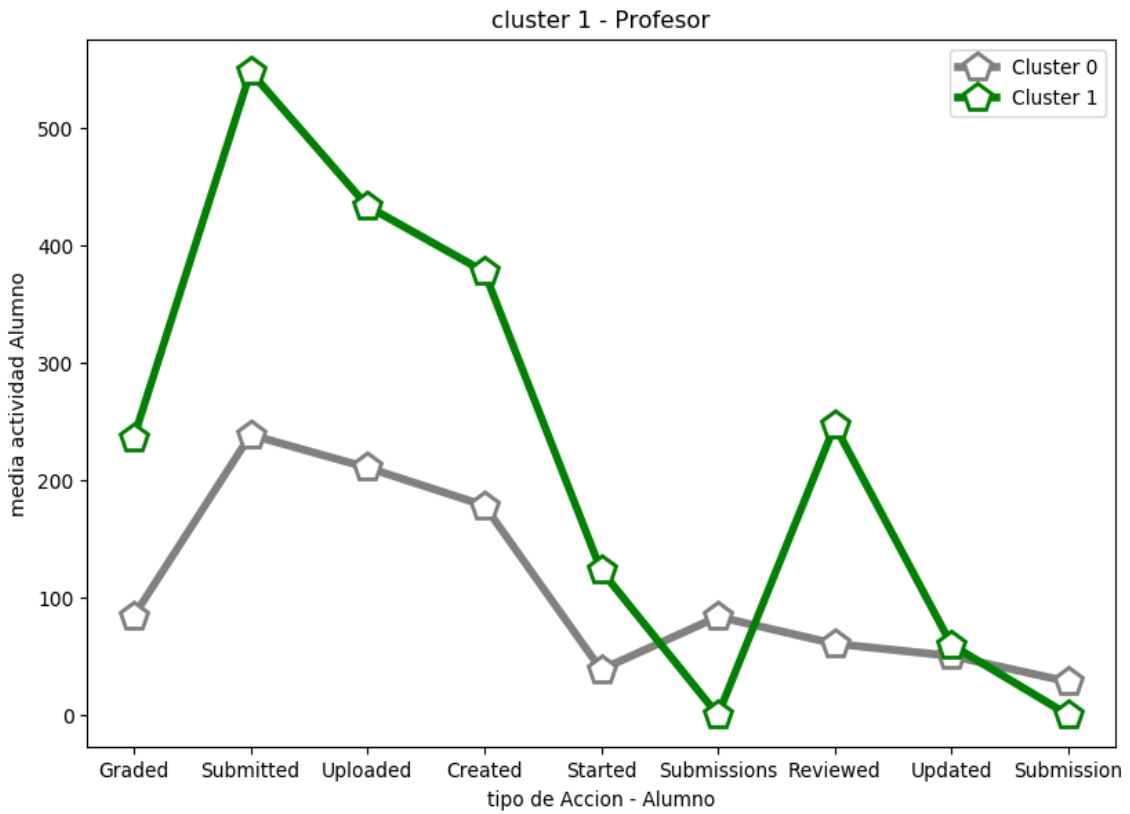
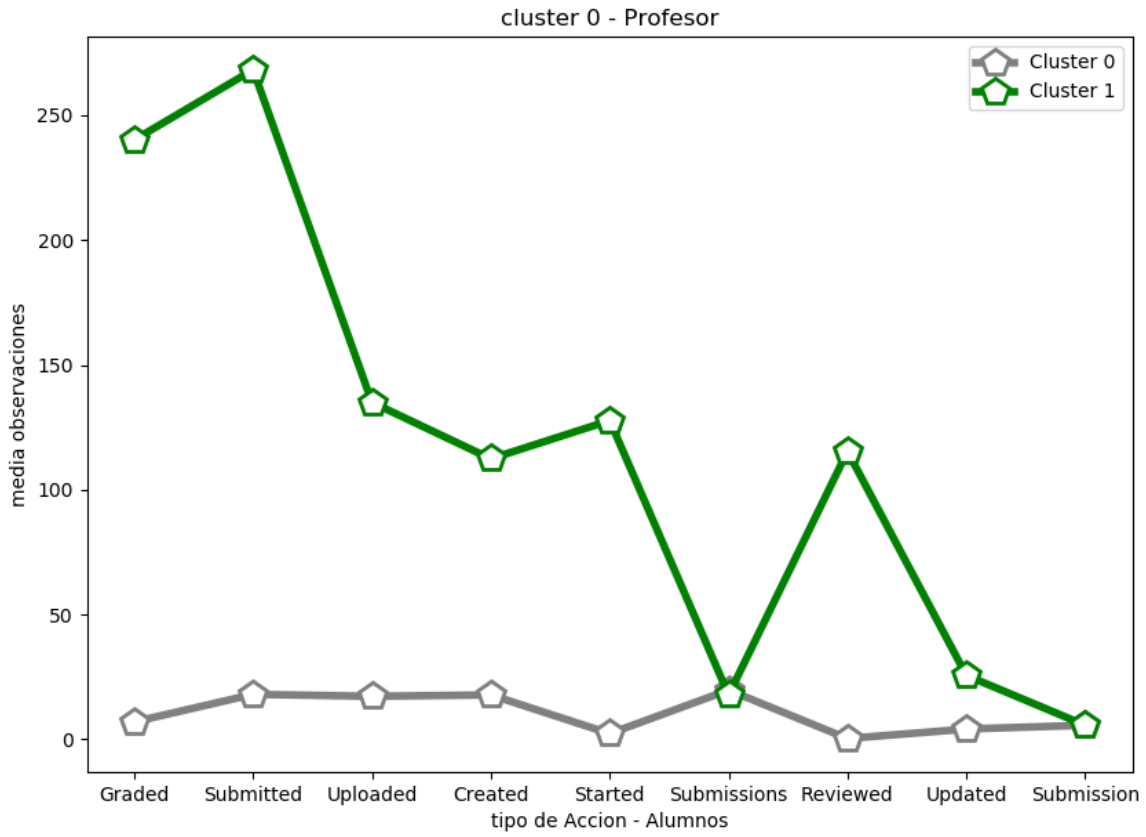
# Neural networks classifier
nnc = MLPClassifier(hidden_layer_sizes=(130), activation='relu', solver='lbfgs', learning_rate='adaptive',
random_state=2)
scores_nnc = cross_val_score(nnc, X, y, cv=5)
method.loc['score', 'nnc'] = scores_nnc.mean(); method.loc['std', 'nnc'] = scores_nnc.std() * 2
nnc.fit(X_train, y_train)
Z['nnc'] = nnc.predict(X_test)

# Neural networks regressor
nnr = MLPRegressor(max_iter=500, alpha=1e-7)
scores_nnr = cross_val_score(nnr, X, y, cv=5)
method.loc['score', 'nnr'] = scores_nnr.mean(); method.loc['std', 'nnr'] = scores_nnr.std() * 2
nnr.fit(X_train, y_train)
Z['nnr'] = nnr.predict(X_test)

# Random Forest Classifier
rfc = RandomForestClassifier(n_estimators=10)
scores_rfc = cross_val_score(rfc, X, y, cv=5)
method.loc['score', 'rfc'] = scores_rfc.mean(); method.loc['std', 'rfc'] = scores_rfc.std() * 2
rfc.fit(X_train, y_train)
Z['rfc'] = rfc.predict(X_test)

```

## Anexo 4. Tipologías de profesores y alumnos



## Anexo 5. Esquema de la base de datos relacional construida



## Anexo 6. Ejemplo código SQL

```

SUM(CASE WHEN eventName = '\\mod_forum\\event\\discussion_created' THEN 1 ELSE 0 END) as V23,
SUM(CASE WHEN eventName = '\\mod_folder\\event\\course_module_viewed' THEN 1 ELSE 0 END) as V24,
SUM(CASE WHEN eventName = '\\core\\event\\group_member_added' THEN 1 ELSE 0 END) as V25,
SUM(CASE WHEN eventName = '\\core\\event\\calendar_event_updated' THEN 1 ELSE 0 END) as V26,
SUM(CASE WHEN eventName = '\\gradereport_singleview\\event\\grade_report_viewed' THEN 1 ELSE 0 END) as V27,
SUM(CASE WHEN eventName = '\\mod_quiz\\event\\report_viewed' THEN 1 ELSE 0 END) as V28,
SUM(CASE WHEN eventName = '\\mod_ur1\\event\\course_module_viewed' THEN 1 ELSE 0 END) as V29,
SUM(CASE WHEN eventName = '\\mod_turnitintooltwo\\event\\list_submissions' THEN 1 ELSE 0 END) as V30,
count(*) as conexionTotal,
count(distinct userId) as numeroProfesor
from registros_log where userId in (select distinct userId from registros_log where rol in ('Profesor', 'Profesor sin permiso de edición')
and ((userId != 246037 and courseId != 75309) or
(userId != 200078 and courseId != 80403) or
(userId != 157863 and courseId != 83903) or
(userId != 69709 and courseId != 76264)) and cursoAcademico = '16/17'
and date between '2016-09-26' and DATE_ADD('2016-09-26', INTERVAL 17 WEEK))
group by courseId) t1
inner join
(select * from asignaturas_cv where cursoAcademico = '16/17') t2
on t1.courseId = t2.url
inner join
(select * from asignaturas_ucm where semestre = 1 and codigoPlan in (810,831,'DT12')) t3
on t3.codigoAsignatura = t2.codigoAsignatura
  
```

## Anexo 7. Ejemplo de visualización de datos

