

INFORME TÉCNICO

---

análisis del éxito de  
actividades de **formación**  
**del profesorado** basado  
en ciencia de **datos**

Manuel Ángel Jiménez Gómez

CENTRO DEL PROFESORADO DE CÓRDOBA



# Resumen

Este informe presenta el proceso y los resultados del trabajo titulado "*Análisis del éxito de las actividades de formación del profesorado basado en ciencia de datos*". En este trabajo se han considerado algunos datos, previamente disponibles, de las actividades formativas realizadas por el CEP de Córdoba durante los cursos 2017/2018 y 2018/2019, para tratar de extraer nueva información sobre cuestiones que preocupan a la organización, como el abandono, la satisfacción o la demanda de los participantes hacia las diferentes actividades formativas. Aunque el conjunto de datos no ha permitido generar modelos predictivos ni descriptivos suficientemente robustos para todas las preguntas planteadas, se han conseguido hacer visibles desconocidas e interesantes relaciones existentes entre estos datos: hallazgos que aportan a la red de formación nueva información a considerar para el diseño de actividades más exitosas.

El presente trabajo es fruto del Trabajo de Fin de Grado de Ingeniería Informática, mención en Computación, en la Escuela Politécnica Superior de la Universidad de Córdoba.

Realizado por: **Manuel Ángel Jiménez Gómez**  
*mjimenez@uco.es*

Dirigido por: **Sebastián Ventura Soto**  
*sventura@uco.es*

**José María Luna Ariza**  
*jmluna@uco.es*

Córdoba, junio de 2020

# Índice

Prefacio	1
1. Introducción	3
2. Datos	7
2.1. Extracción y anonimización de datos	7
2.2. Datos disponibles	7
3. Experimentación y resultados	14
3.1. El análisis predictivo	14
3.2. El análisis descriptivo	15
3.3. Predicción del abandono	16
3.4. Predicción de la satisfacción	17
3.5. Predicción de la demanda	19
3.6. Descripción de tipos de actividades	21
4. Conclusiones y recomendaciones	24
4.1. Limitaciones de los datos	24
4.1. Orientaciones extraídas	25
4.3. Ampliaciones recomendadas	26
Agradecimientos	27
Referencias	28

# *Prefacio*

El principal objetivo de la formación de un profesional es mejorar su rendimiento para mejorar los procesos en los que interviene, y así, mejorar el producto de esos procesos. Si hablamos de formación del profesorado, hablamos de mejorar la instrucción y producir, por ende, una mejora en los aprendizajes del alumnado. Son muchos los autores que tratan de encontrar una trazabilidad entre la formación docente y el éxito de su alumnado, pero esta no es una tarea sencilla.

Francesc Imbernón (Imbernón Muñoz, 2007) afirma, en esta línea, que mucha formación no se corresponde con mucho cambio en la escuela. El éxito se persigue buscando las claves de lo que funciona y, principalmente, este pasa por la colectividad o comunidad frente al individualismo o la colegialidad "artificial" o impuesta. En cualquier caso, estas afirmaciones tratan de posicionarse en esa trazabilidad, que, como se ha dicho, es difícil de observar.

*Cómo, cuándo y en qué condiciones aprende el profesorado*, es parte del título de un libro de Christopher Day (Day, 2005), que comparte axiomas con Imbernón analizando la importancia de los contextos personales, profesionales y el liderazgo, y hace especial hincapié en la reflexión sobre la práctica y la investigación-acción como parte esencial de la tarea de los docentes.

La llamada colegialidad real o colectividad también es la medida que más influye positivamente en el rendimiento escolar según el ranking de Hattie: un meta-análisis de más de 800 investigaciones en el que se analizan 252 acciones y su influencia en el aprendizaje (Hattie, 2013).

Por tanto, parece deseable que el profesorado se forme en destrezas hacia la reflexión colectiva y ese trabajo colegiado, para que la transferencia de su formación repercuta de forma positiva en el éxito del alumnado. Para lograr la adherencia del profesorado hacia la formación continua es esencial hacer cosas

que funcionen, pero igualmente importante es lograr su satisfacción, que, a su vez, repercutirá en bajar las tasas de abandono y aumentar la demanda.

La mencionada trazabilidad entre la formación y su impacto es llamada eficacia por autores como Carmen Aneas (Aneas Novo, Sánchez Rodríguez, & Sánchez Rivas, 2018), que han aportado herramientas para realizar mediciones estimativas de esta eficacia y han comparado la transferencia a la práctica de diferentes modalidades formativas en Andalucía. Por su parte, Pineda (Pineda Herrero, 2000) revisa tres modelos de evaluación del impacto de la formación en las organizaciones, propuestos por Kirkpatrick, Phillips y Wade. Aunque es el primero de ellos el que más se adapta a la escuela, todos toman en consideración datos recogidos expresamente para esta evaluación, al igual que la propuesta de Aneas.

La principal diferencia entre el trabajo que nos ocupa y los descritos en la literatura revisada reside en que este estudio se limita a medir lo que permiten los datos que ya se tienen almacenados en los sistemas de información de la organización, y no recaba nueva información específicamente para la evaluación. Tampoco se abordan problemas como la transferencia, *a priori* posible con la información disponible, pero cuyo coste superaría con creces las limitaciones temporales de este estudio.

No se ha descrito, hasta el momento, ningún trabajo que tenga como objeto el tratamiento de datos provenientes de la formación del profesorado en su dominio completo y con el objetivo de mejorar el propio sistema formativo, algo que quedaría dentro, según la clasificación de Ventura y Romero (Romero & Ventura, 2020) de las técnicas DDE (*Data-Driven Education*) o DDDM (*Data Driven Decision Making in Education*) aplicadas a la formación del profesorado; conceptos muy similares a la gobernanza digital educativa en palabras de Williamson (Williamson, 2015).

## Capítulo 1

### Introducción

La ciencia de datos facilita un nuevo tipo de relación entre la información y las decisiones a tomar por los gestores públicos: la gobernanza digital. Como en todo avance, se presentan dos caras de la misma moneda: la utópica y la distópica. La misma tecnología que puede servir para una toma de decisiones más rápida y acertada sobre temáticas complejas, puede ser utilizada para ejercer un mayor control sociopolítico. Autores como Williamson (Williamson, 2015) o Kitchin (Kitchin & Dodge, 2011) alertan sobre cómo el software está modelando nuestra realidad cotidiana y la emergencia de que sea la gestión pública la que regule y tome el control de esta gobernanza digital en sectores sensibles y estratégicos como la Educación.

Administraciones nacionales y autonómicas en España, no se deciden a emplear esta inteligencia de la organización y ponerla al servicio del propio sistema para su mejora. Y es que escándalos en el ámbito internacional, relacionados con el uso de ciencia de datos para la manipulación y atentados a la privacidad, hacen contra-pedagogía y encasillan la gobernanza digital dentro de las actuaciones controvertidas pendientes que cualquier dirigente político, hasta el momento, ha mantenido en cuarentena.

En la última década contamos con numerosas revisiones e informes internacionales que sitúan a la formación del profesorado como variable clave en el éxito educativo. La famosa y discutida publicación "Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos" (Barber & Mourshed, 2007), mejor conocida como "informe McKinsey", establece como título de dos de sus secciones las siguientes afirmaciones: "la calidad de un sistema educativo tiene como techo la calidad de sus docentes", y "la única

manera de mejorar los resultados es mejorando la instrucción". Con estas premisas, y el objetivo de mejora continua del sistema educativo, los centros del profesorado tratan de promover la actualización científica y metodológica de docentes de enseñanzas básicas y medias en Andalucía.

La formación del profesorado no depende en exclusiva de la red de asesora de formación, ya que para acceder a la función docente se debe acreditar una formación inicial universitaria específica: grado de Magisterio para las etapas de Educación Infantil y Primaria y el MAES<sup>1</sup> para enseñanzas medias. Por tanto, la formación permanente debe paliar los déficits de la formación inicial y ocuparse de la actualización profesional. Esto, en Andalucía, se atiende con una vasta oferta de actividades (más de cinco mil cada año) en diferentes formatos: cursos presenciales, semipresenciales, en línea, jornadas, congresos, encuentros, grupos de trabajo, formaciones en centro... ofertada por 32 centros coordinados por el Servicio de Planes de Formación, dependiente de la Dirección General de Formación del Profesorado e Innovación Educativa.

A pesar del absoluto convencimiento por parte de la administración, inspección y un mayoritario sector del profesorado de la importancia que la formación continua desempeña en la mejora de la actividad docente, esta sigue teniendo un carácter voluntario en todo el territorio nacional. Esta voluntariedad hace que los centros del profesorado no sólo se preocupen por organizar actividades útiles y conectadas con las necesidades de docentes, alumnado y centros; también tienen que preocuparse de que el profesorado resulte satisfecho y no abandone.

En el CEP de Córdoba, un grupo de asesorías expertas lleva años intentando seguirle la pista al éxito de las actividades. Entendiendo éxito en un sentido amplio: demanda de participación, satisfacción de los participantes, número de certificados emitidos, impacto de la actividad en la práctica docente o transferencia de lo aprendido hacia el alumnado.

---

<sup>1</sup> Máster Universitario en Profesorado en Enseñanza Secundaria Obligatoria y Bachillerato Formación Profesional y Enseñanzas de Idiomas

Este trabajo ha tratado de extraer información mediante técnicas de minería y de explorar la utilidad que tienen para este fin, los datos que la Consejería de Educación de la Junta de Andalucía trata y almacena en referencia a la formación permanente de su profesorado, con el principal objetivo de conseguir mejorar esta oferta formativa y contribuir a aclarar algunas cuestiones referentes al abandono, satisfacción, o demanda de las actividades que el profesorado realiza.

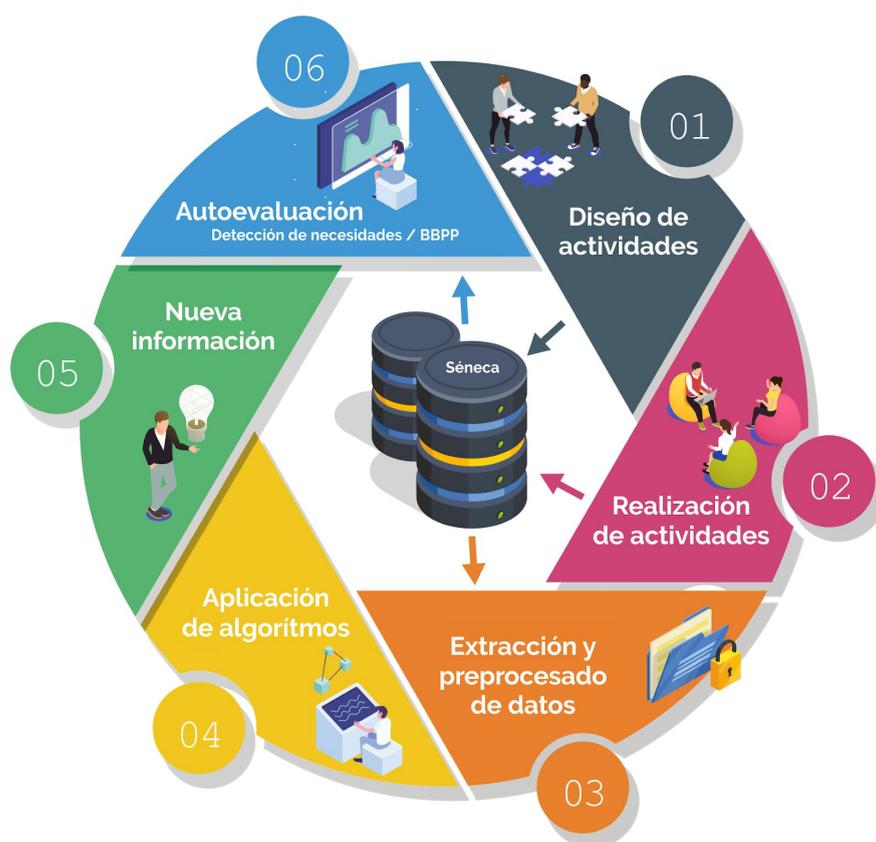


Figura 1. Flujo de información del estudio

Los problemas que se han tratado de resolver han sido variados, intentando responder a las siguientes preguntas:

- ¿Se puede predecir la demanda de una actividad?
- ¿Se puede predecir el abandono de una actividad?
- ¿Se puede predecir la satisfacción de una actividad?

- ¿Cuáles son las claves?
- ¿Qué variables se ven implicadas?
- ¿Qué tienen en común las diferentes actividades formativas?
- ¿Qué dependencias o correlaciones se pueden observar entre las diferentes variables o grupos de ellas?

Se ha contado con la cesión anonimizada de datos de casi mil actividades celebradas durante los cursos académicos 2017/2018 y 2019/2020 organizadas por el CEP de Córdoba. Sobre este conjunto se han aplicado diferentes algoritmos de aprendizaje automático, tanto no supervisado (enfoque descriptivo), como supervisado (enfoque predictivo).

El proceso de minería ha descubierto nueva información relevante y estratégica que se desconocía por la organización, aunque tan sólo se han podido validar modelos predictivos robustos para dar respuesta a la satisfacción de los participantes en las actividades. También se ha determinado la limitación del conjunto de datos para abordar algunos de los problemas propuestos.

Se considera esencial que se prolongue en el tiempo el ciclo ilustrado en la figura 1, es decir, que la nueva información descubierta sea utilizada por la organización para mejorar su práctica y vuelva a ser observado desde este mismo prisma en los cursos sucesivos, incorporando la toma de decisiones basadas en datos o la gobernanza digital dentro de la cultura de autoevaluación y mejora de la organización.

## Capítulo 2

### Datos

#### 2.1. Extracción y anonimización de datos

Todos los datos que se han utilizado en este estudio han sido previamente anonimizados por el equipo directivo del CEP de Córdoba y carecen de información sensible que pueda verse afectada por la normativa vigente sobre protección de datos de carácter personal (LOPD-GDD, 2018).

#### 2.2. Datos disponibles

En *Séneca*<sup>2</sup> se almacena, respecto a una actividad formativa, datos referentes a su diseño, solicitudes, participantes admitidos, participantes certificados, valoración de la actividad por los participantes y por las asesorías responsables, información económica... Pero no todo es accesible, desde las vistas disponibles, una vez finalizada la actividad. Tampoco se puede obtener gran parte de esta información si no es mediante un costoso procesamiento interactivo. Por tanto, la información de la que se dispone para el primer análisis que plantea este trabajo es la que se especifica a continuación. Se enumeran y describen las características que podrán considerarse sobre cada actividad formativa.

- **Curso:** curso académico en que se realiza la actividad. En la muestra toma los valores 2017-2018 o 2018-2019.
- **Inicio:** fecha de inicio de la actividad.

---

<sup>2</sup>Sistema de información Séneca. Junta de Andalucía.  
<https://www.juntadeandalucia.es/educacion/portalseneca>

- **Fin:** fecha de finalización de la actividad.
- **Modalidad:** modalidad de la actividad formativa. Existen las siguientes 12 categorías.
  - Formación en Centros
  - Formación Específica en Centros
  - Grupos de trabajo
  - Conferencia
  - Congreso
  - Curso
  - Curso a Distancia
  - Curso con seguimiento
  - Curso Semipresencial
  - Encuentro
  - Jornadas
  - Otros
- **Tipo:** existen 10 categorías según el tipo de actividad formativa.
  - Plan General
  - Actividad en Centro
  - A distancia
  - Formación Profesional
  - Provincial
  - Regional
  - Servicios Centrales
  - Formaciones en Centro
  - Grupos de trabajo
  - Programas Educativos
- **Línea:** pertenencia de la actividad a una de las 5 líneas estratégicas definidas en el III Plan Andaluz de Formación Permanente del Profesorado [13].

1ª. La formación del profesorado vinculada a la mejora de las prácticas educativas, el rendimiento y el éxito educativo de todo el alumnado.

2ª. La formación del profesorado como factor clave para el perfeccionamiento continuo y la capacitación profesional docente.

3ª. La formación del profesorado como impulsora del conocimiento compartido y producido en los centros educativos, la investigación y la innovación educativa y las buenas prácticas.

4ª. La formación del profesorado como apoyo en la progresiva transformación de los centros educativos en entornos colaborativos de aprendizaje y formación en los que participan todos los miembros de la comunidad educativa.

5ª. La formación del profesorado de formación profesional, enseñanzas artísticas y de idiomas y educación permanente como herramienta para conectar la educación con la realidad productiva y el empleo.

- **Temática:** tema concreto sobre el que versa la actividad formativa. Existen 65 temas definidos agrupados en una docena de descriptores más generales.
  - Integración de las competencias clave
    - Conocimiento y desarrollo de las competencias.
    - Evaluación.
    - Metodología y Planificación
  - Escuela Inclusiva-Atención Diversidad
    - Atención al alumnado altas capacidades
    - Atención al alumnado de carácter compensatorio.
    - Atención al alumnado dificultades aprendizaje
    - Atención alumnado NEE.
    - Convivencia
    - Igualdad
  - Sociedad del conocimiento
    - Actualización en Lenguas Extranjeras
    - AICLE
    - Currículum integrado - MCERL

- Dimensión Europea
- Fomento del Plurilingüismo
- Conocimiento Tecnológico (TIC)
- Uso didáctico (TAC)
- Emprendimiento
- Cultura autoevaluación
- Planes de mejora
  - Mejora continua.
  - Función asesora
  - Función coordinadora
  - Función directiva
  - Función inspectora
  - Función orientadora
  - Función tutorial
- Desarrollo profesional - competencias específicas
  - Actualización didáctica ámbito artístico
  - Actualización didáctica ámbito artístico deportivo
  - Actualización didáctica ámbito científico tecnológico.
  - Actualización didáctica ámbito cívico social
  - Actualización didáctica ámbito deportivo
  - Actualización didáctica ámbito lingüístico.
  - Actualización didáctica de las Matemáticas
- Desarrollo profesional - competencias docentes
  - Didáctica y Gestión del aula
  - Formación inicial
  - Formación novel.
  - Habilidades sociales y emocionales
  - Salud Laboral / Autoprotección
- Investigación e innovación - Programas educativos.
  - Ámbito cultural
  - Ámbito de economía y emprendimiento
  - Ámbito de hábitos de vida saludable

- Ámbito digital
  - Ámbito lingüístico
  - Ámbito medioambiental
- Investigación e innovación
  - Investigación en el aula
  - Trabajo colaborativo
- Participación comunidad - Escuela participativa
  - Actividades con la comunidad educativa
  - Comunidades de aprendizaje.
  - PASEN y otras plataformas
- Entornos específicos
  - Atención alumnado inmigrante
  - Atención domiciliaria
  - Aulas hospitalarias
  - Centros rurales
  - Residencias escolares
- FP, Enseñanzas Artísticas, EEOOII, EPER
  - Cualificaciones profesionales.
  - FP – Orientación al empleo
  - FP en alternancia
  - FP Actualización profesional y metodológica
  - Gestión de calidad
  - ARE – Fomento y difusión de la cultura
  - ARE – Orientación al empleo.
  - ARE Actualización profesional y metodológica.
  - EEOOII – Actualización niveles MCERL
  - EEOOII Actualización profesional y metodológica
  - EPER – Orientación profesional y académica
  - EPER Actualización profesional y metodológica
- **Ámbito:** hace referencia a la extensión del público al que la actividad va dirigida: centro docente, ámbito CEP (parte de la provincia), provincial o regional (toda la comunidad autónoma).

- **Asesorías:** personas (identificadores disociados) que intervienen en el diseño, organización y coordinación de la actividad formativa.
- **Horas presenciales:** duración, en horas, de la fase presencial de la actividad.
- **Horas no presenciales:** duración, en horas, de la fase no presencial de la actividad.
- **Gastos:** coste de la actividad expresado en euros.
- **Plazas:** aforo máximo.
- **Solicitudes:** número de demandantes de participación.
- **Participantes:** número de solicitudes admitidas. También disponemos de este dato diferenciado por género: participantes hombres y participantes mujeres.
- **Certificados:** número de participantes que terminan la actividad con una valoración positiva. También disponemos de este dato diferenciado por género: certificados hombres y certificados mujeres.
- **Encuesta de satisfacción:** se dispone de una serie de items que el profesorado participante valora al finalizar una actividad:
  - Grado de consecución de los objetivos.
  - Adecuación de los contenidos a las necesidades formativas.
  - Diseño de las sesiones / módulos (online).
  - Idoneidad de espacios utilizados / entornos virtuales (online).
  - Dinámica, recursos y metodología utilizada.
  - Nivel de interacción entre participantes / foros (online).
  - Utilidad en la práctica profesional.
  - Grado de aplicación en su contexto educativo.
  - Recursos, bibliografía y materiales propuestos
  - ¿En qué medida recomendaría a sus compañeros y compañeras participar en una actividad como esta? (NPS) [14].
  - Capacidad de motivar y comunicarse con los participantes.
  - Rigor y solvencia en los contenidos desarrollados.
  - Capacidad para fomentar un ambiente participativo/foros.

- Adaptación de las actividades/tareas a las necesidades del grupo.

No todos los datos están disponibles al mismo tiempo, existen variables que se determinan en la fase de diseño y otras que no serán conocidas hasta que se abra la inscripción o se termine la actividad formativa. La figura 2 muestra cómo se han incorporado los datos a este estudio simulando el momento temporal en el que estarían disponibles, y así poder generar modelos predictivos más o menos precoces.

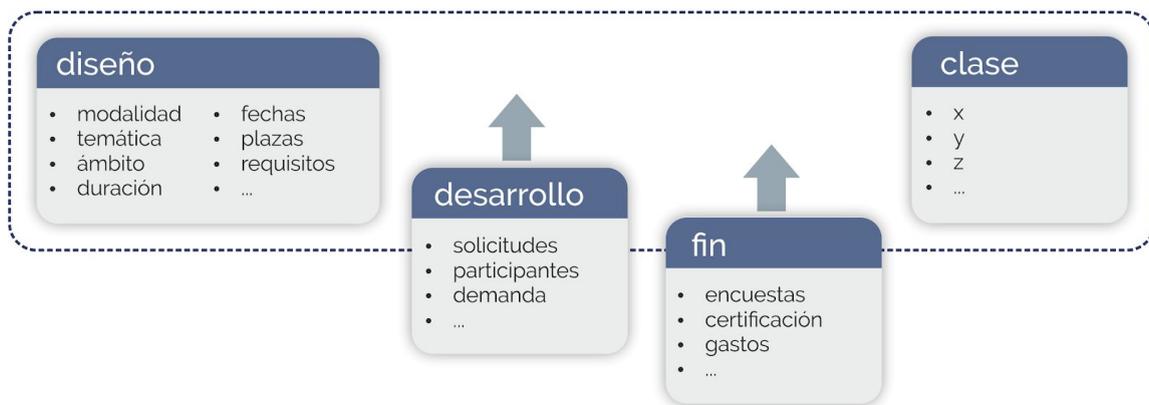


Figura 2. Datos disponibles según la fase de la actividad

El conjunto de datos ha sido dividido por modalidades, en subconjuntos más similares y comparables. Se han agrupado del siguiente modo:

- Formaciones en centros (FC) y formaciones específicas en centros (FEC).
- Grupos de trabajo (GT).
- Actividades diversas: cursos, jornadas, encuentros, congresos...

## Capítulo 3

# Experimentación y resultados

En ciencia de datos existen dos enfoques principales, los métodos predictivos y los métodos descriptivos. Los predictivos aprenden a partir de un conjunto de datos y obtienen modelos que permiten clasificar patrones en una clase determinada, por ejemplo, permitirían saber si una actividad va a tener suficiente demanda, si su abandono va a ser menor del 20% o si los resultados de las encuestas van a ser buenos. Mientras que el enfoque descriptivo no predice que ocurrirá, sino que nos habla de lo que ha pasado. La mayoría de los datos educativos que se almacenan en los sistemas de información son difícilmente consumibles por el ser humano directamente. Las técnicas descriptivas tratan de hacernos un resumen, sin más pretensión. Se han utilizado técnicas de agrupamiento y de reglas de asociación.

### 3.1. El análisis predictivo

Mediante métodos de clasificación se ha tratado de responder a tres preguntas principalmente. Con los datos disponibles, pormenorizados en la sección anterior...

- ¿Se puede predecir la **demanda** de una actividad?
- ¿Se puede estimar *a priori* la tasa de **abandono** que sufrirá?
- ¿Es posible conocer el grado de **satisfacción** que reflejarán las encuestas?

La utilización de clasificadores en entornos educativos requiere tanto que los modelos sean sensibles y robustos en una situación real, como comprensibles para los usuarios involucrados. Por esto, como en la mayoría de los estudios

revisados, se han seleccionado algoritmos cuyos modelos están basados en reglas o árboles de decisión (cuyos resultados se pueden considerar como modelos de "caja blanca"), frente a otros más difícilmente interpretables como máquinas vector-soporte, redes bayesianas o neuronales (con modelos de "caja negra"). Concretamente se han utilizado los algoritmos **RIPPER** (Cohen, 1995), **PART** (Frank & Witten, 1998) y **C4.5** (Quinlan, 1993).

## 3.2. El análisis descriptivo

Mediante agrupamiento y reglas de asociación se ha tratado de responder a dos de las preguntas planteadas en la definición del problema.

- ¿Qué tienen en común las diferentes actividades?
- ¿Qué relaciones existen entre las diferentes características?

Las técnicas de *clustering* o agrupamiento no predicen una pertenencia a una clase, sino que aportan información sobre la naturaleza del conjunto de datos, y más específicamente, sobre características comunes o parecidas que tienen en las instancias que forman parte de *clusters* o grupos, y que pueden pasar desapercibidas.

Se ha experimentado con dos algoritmos, considerados de referencia (Jung, Kang, & Heo, 2014), que plantean dos enfoques diferentes para la resolución del mismo problema, **K-medias** y **Expectation-Maximization**, el primero utiliza como parámetro de entrada el número de grupos en el que queremos separar el conjunto, mientras que el segundo calcula un óptimo en función de la máxima verosimilitud estadística.

Las reglas de asociación, en minería de datos, se utilizan para descubrir hechos que ocurren dentro del conjunto de datos, dependencias repetidas entre variables o conjuntos de variables y sus valores respecto a otros valores o conjuntos de valores de otras variables. Se trata de una exploración exhaustiva

del conjunto de datos en búsqueda de estas relaciones definidas en forma de reglas, y cuyo fin no es clasificar los patrones en ninguna clase ni categoría, simplemente describir hechos. Se ha utilizado el algoritmo **Apriori** (Agrawal & Srikant, 1994) que utiliza una búsqueda en anchura con un enfoque *bottom-up*, reduciendo iterativamente el soporte hasta que encuentra reglas con un mínimo de confianza predefinido.

### 3.3. Predicción del abandono

Se han utilizado los algoritmos citados en el apartado 3.1 sobre los diferentes conjuntos de datos (Completo, FC, FEC, GT y la "no autoformación") y no se ha logrado obtener un modelo que pueda predecir de forma robusta si una actividad sufrirá más o menos abandono, no obstante, merece la pena resaltar los siguientes hallazgos de estos experimentos.

- Existe una relación fuerte entre la puntuación NPS y el abandono que sufre una actividad. Y merece la pena aclarar, que las personas que abandonan una actividad no suelen realizar la encuesta. La relación se establece según las siguientes reglas.
  - $NPS \geq 80 \rightarrow \text{Abandono} < 20\%$
  - $80 > NPS \geq 50 \rightarrow 40\% > \text{Abandono} \geq 20\%$
  - $NPS < 50 \rightarrow \text{Abandono} \geq 40\%$
- Las actividades dirigidas a un pequeño público suelen ser mejor valoradas y tener menos abandono. Esto también ocurre cuando la actividad se celebra en el marco de un centro educativo. Así, los grupos de trabajo casi no tienen abandono, y menos aún los que tienen financiación.
- Así, las formaciones en centros de claustros pequeños no tienen problemas de abandono (inferior al 20% si hay menos de 20 participantes). Mientras que en claustros grandes (más de 45 participantes) el abandono es superior al 40% (y su NPS inferior a 50).
- Las formaciones vehiculadas por PLC o PRODIG no tienen abandono.

- Muchos participantes se descuelgan en la fase online de actividades semipresenciales.
- Autoformaciones con menos de 15 participantes casi no sufren abandono.
- Las actividades dirigidas a perfiles profesionales (secretaría, dirección, jefatura de estudios, orientación...) o colectivos específicos (redes) tienen un abandono inferior al 20%.
- Las mujeres tienen un compromiso mayor con la autoformación.
- La mayoría de las actividades relacionadas con el uso educativo de la tecnología, casi siempre en formato semipresencial, tienen un abandono de entre un 20 y un 40%.
- La mayoría de las actividades de actualización didáctica del ámbito lingüístico llevadas a cabo durante el segundo trimestre tienen un abandono superior al 30%.
- Actividades sobre metodología y planificación tienen un abandono menor al 30%.
- Las actividades con baja demanda (20-70%) a penas tienen abandono.

Se han encontrado muchas más reglas interesantes pero con un soporte muy bajo. Esto podría indicar que si se aumentase el tamaño de la muestra se obtendrían mejores modelos.

### 3.4. Predicción de la satisfacción

Aunque existen hallazgos sobre las valoraciones que los participantes hacen de las actividades por medio de la encuesta, se pone de manifiesto que esta no tiene poder informativo o discriminatorio dentro de la muestra. Casi el 90% de las valoraciones se encuentra en la horquilla de 8 y 9 puntos sobre 10 (lo mismo ocurre en las 9 preguntas de la encuesta), lo cual manifiesta la buena percepción que tienen los participantes de los ítems observados, pero no ayuda a detectar dependencias entre otros factores y estas valoraciones. Lo mismo ocurre con los ítems de la encuesta referidos a ponentes y tutores: las valoraciones siempre son

muy positivas. El único indicador que ha demostrado tener valor discriminatorio es el NPS. No obstante, se enumeran hallazgos de interés para la organización.

- Las actividades, en ocasiones vinculadas con *planes y programas*, promovidas por servicios centrales y los PFEC suelen tener una valoración inferior a la media. La inmensa mayoría de los PFEC no contaron tampoco con financiación.
- FC y FEC con más de 15 participantes y sin ponente suelen tener una valoración inferior a la media. Lo mismo ocurre si hay más de 35 docentes, haya o no ponente.
- Existe la tendencia de que las actividades sin ponente externo (coste cero o muy bajo) sean peor valoradas.
- Los GT y las actividades para FP suelen puntuar por encima de la media.
- Las actividades con mucha demanda suelen tener buenos resultados en la encuesta.

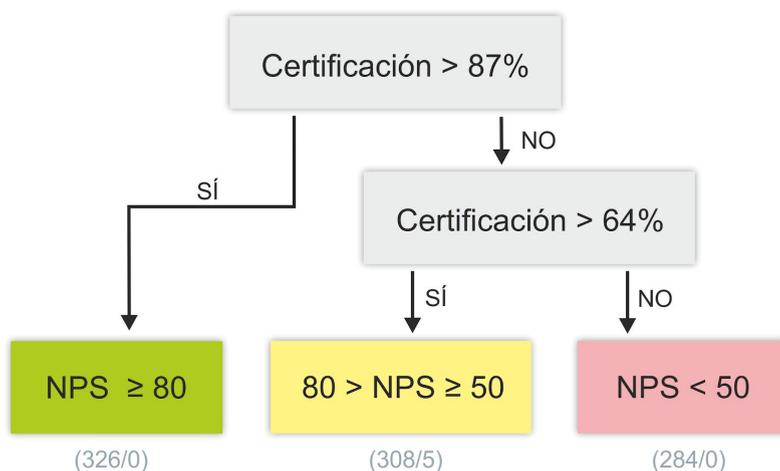


Figura 3. Árbol para predecir la recomendación (Precisión 99,45%)

- Como se muestra en la figura 3, las actividades con una certificación mayor al 87% reciben una puntuación NPS mayor o igual a 80, mientras que las que tienen una tasa de certificación entre ese 87 y un 64% obtendrán un NPS entre 50 y 80. Las actividades en las que la certificación baja del 64%, el NPS lo hace por debajo de 50.

- GT y FC con menos de 30 inscritos resultan más recomendables que FEC o FC con más de 29 participantes (figura 4). Concretamente, las FEC suelen tener menos de 50 puntos NPS.

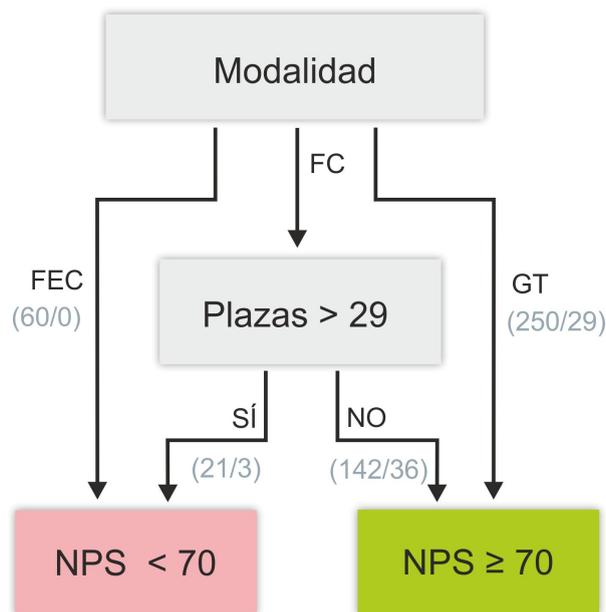


Figura 4. Árbol de decisión sobre autoformaciones y NPS (Precisión: 84%)

- Las actividades con aforo reducido son las más recomendables. Indicio que apunta la bondad de una atención más personalizada y a la posibilidad de establecer diferentes dinámicas que fomenten el intercambio entre los participantes.

### 3.5. Predicción de la demanda

Sólo tiene sentido tratar de predecir la demanda sobre el subconjunto de datos relativo a las actividades que se ofertan al público y cuya inscripción se realiza de manera individual. Por tanto, contamos para este problema con un conjunto de datos preparado para una clasificación binaria dividido en actividades cuya demanda ha superado el número de plazas ofertadas y las que han quedado con plazas vacantes. Tampoco tiene sentido buscar un modelo predictivo que



### 3.6. Descripción de tipos de actividades

En la muestra parece que ciertas variables toman valores, más o menos, parecidos y dependiendo unas de otras configurando tendencias que podrían verse como clases o tipos de actividades que se comportan de manera similar. Por ejemplo, entre las formaciones en centros destacan tres tipos, que se han representado en la tabla 1.

	FC-A	FC-B	FC-C
Presencia (%)	18	15	40
Horas de ponencias	alto	medio	bajo
Número de participantes	bajo	alto	medio
Abandono	bajo	alto	bajo
Valoración	alta	media	alta

Tabla 1. Tipos o agrupamientos de FC

Entre los grupos de trabajo encontramos dos comportamientos, divididos, principalmente por su asignación económica, aunque existen otros detalles a considerar, como se muestra la tabla 2.

	GT-A	GT-B
Presencia	10%	90%
Financiación	sí	no
Participantes hombres	medio	bajo
Participantes mujeres	medio	alto
Valoración	máxima	alta

Tabla 2. Tipos o agrupamientos de GT

Tan complejo como interesante resulta representar los tipos de actividades (cursos, jornadas, congresos, encuentros...) que se han descubierto. Se trata de 10 grupos diferentes que tendrían comportamientos comunes. En la tabla 3, enumerados desde A-01 hasta A-10, se dan todos los detalles.

	%	Núm. de participantes	Tasa de abandono	Gasto por participante	Sesiones presenciales	Horas online	Demanda	Valoración
A-01	13	bajo	muy baja	alto	5-8	3-7	normal	muy alta
A-02	16	medio	media	medio	3-4	3-7	alta	muy alta
A-03	1	medio	sin cert.	bajo	1-2	0-5	normal	normal
A-04	14	medio	baja	bajo	2-3	0-5	baja	muy baja
A-05	9	alto	muy alta	bajo	1-2	0-5	baja	normal
A-06	7	muy alto	sin cert.	bajo	2-3	0-20	normal	baja
A-07	21	bajo	alta	medio	3-4	0-5	normal	alta
A-08	4	muy alto	media	medio	0	>30	muy alta	alta
A-09	10	medio	alta	alto	5-8	3-20	normal	normal
A-10	6	muy alto	baja	medio	3-4	3-7	normal	normal

Tabla 3. Tipos o agrupamientos de actividades

A modo de leyenda, en la tabla 4 se especifican los valores medios para cada atributo y categoría.

	muy bajo/a	bajo/a	normal/medio	alto/a	muy alto/a
Núm. de participantes	10	20	30	40	100
Tasa de abandono	15%	20%	30%	40%	50%
Gasto por participantes	5	15	30	60	100
Demanda	60%	90%	100%	150%	400%
Valoración (NPS)	20	40	50	70	80

Tabla 4. Valores medios de referencia para categorías y atributos

De estos datos, a modo de análisis, se pueden resaltar las siguientes observaciones y posibles motivos:

- Entre las actividades no tienen certificación (8%), son peor valoradas aquellas que tienen una implicación mayor (presencial y *online*) para los participantes (1%).
- Las actividades que tienen alrededor de 20 participantes (32%), tienen menor abandono cuantas más sesiones presenciales (mejor más de 5), esto podría deberse a la posibilidad de poder certificar con alguna/s ausencia/s, o a la adherencia propia de una actividad para perfiles profesionales o tipo "red". Constituyen un 13% del total.
- Entre las actividades con una participación media (en torno a 30 personas, un 40% del total), podemos diferenciar aquellas que tienen una tasa de abandono bajo (14%), medio (16%) o alto (10%), y relacionar este resultado de manera directamente proporcional con el número de sesiones presenciales. Las que se mantienen en el término medio son las más demandadas y mejor valoradas.
- Existe otro grupo de actividades que se caracteriza por tener más público (sobre 40 participantes) y poca implicación (presencial y online), se quedan con plazas vacantes, el grado de satisfacción es medio y sufren mucho abandono. Constituyen un 9% del total.
- Si doblamos el número de participantes a la vez que la implicación encontramos actividades con menos abandono. Suman el 6%.
- Finalmente, la teleformación (4%) goza de mucha demanda, un alto grado de satisfacción y un abandono que se sitúa en torno al 30%. El gasto por participante es algo superior a la media.

## Capítulo 4

# Conclusiones y recomendaciones

A pesar de que no se han conseguido modelos robustos de predicción de las variables que pueden determinar el abandono o la demanda de una actividad formativa, la información extraída por este trabajo se considera muy valiosa y de ella surgen un conjunto de orientaciones o “manual de de buena praxis” que facilita tener en cuenta, en las fases de diseño y desarrollo de una actividad, las cuestiones que hacen que crezca la demanda y se reduzca el abandono o la insatisfacción; tratando, así, de mejorar procesos y toma de decisiones para un mejor aprovechamiento de los recursos.

### 4.1. Limitaciones de los datos

La exploración exhaustiva del conjunto de datos ha puesto de manifiesto algunas limitaciones del mismo. Se enumeran a continuación:

- La principal limitación del conjunto de datos es que se trata de una visión parcial, carente de gran parte de la información, del dominio del problema. Es decir, sólo representa una parte de los datos referentes a la formación, sin considerar a los participantes ni a los ponentes.
- Las puntuaciones medias recogidas en los ítems valorados en las encuestas no dan información discriminante sobre las actividades.
- La clasificación de las actividades en líneas, descriptores, temáticas, ámbito, etc. no siempre se realiza siguiendo los mismos criterios, lo cual hace que el uso de estas características no produzca soluciones suficientemente robustas.

- La información obtenida se debe, principalmente, a las reglas que representan hechos muy repetidos en el conjunto (con un soporte adecuado). Por tanto, no hay que perder de vista aquello que no surge en estas reglas porque su representación en la muestra no sea significativa. Este problema podría solucionarse mediante la observación de un conjunto de datos mayor.
- Las métricas usadas para analizar el éxito no consideran transferencia ni impacto de la actividad, ya que no hay información referente a estos objetivos en los datos disponibles.

#### 4.1. Orientaciones extraídas

A modo de decálogo, acompañan a este informe unas orientaciones a tener en cuenta para el mejor desempeño de la tarea de diseño de las actividades, con especial atención a dejar de hacer lo que no funciona.

1. Las actividades de autoformación que superan los 20 participantes suelen funcionar peor (más abandono y menor satisfacción), por lo que parece razonable, en estos casos, utilizar estrategias que dividan al grupo en grupos menores.
2. Las actividades contextualizadas a las necesidades de un centro tienen mejores resultados de abandono y satisfacción.
3. Por los resultados que obtienen, se deben fomentar los grupos de trabajo.
4. Se tiene que intentar, en la medida de lo posible, contar con ponentes externos en todas las autoformaciones (FC y FEC).
5. Si se diseña una actividad para pocos participantes, mejor cuantas más sesiones presenciales.
6. Las actividades con muchas sesiones presenciales no tienen más abandono por este hecho.
7. Si se diseña una actividad para muchos participantes, mejor cuanto más *online*.

8. Es conveniente establecer diferentes dinámicas que fomenten el intercambio y la comunicación entre los participantes.
9. Existe la necesidad de formular nuevas preguntas en la encuesta que revelen información relevante.
10. Existe la necesidad de un mayor consenso entre el equipo asesor a la hora de etiquetar actividades en líneas, descriptores, temáticas, ámbitos, etc.

### 4.3. Ampliaciones recomendadas

Este trabajo es susceptible de muchísimas mejoras, pero casi ninguna está al alcance del CEP de Córdoba. Se relacionan a continuación algunas ideas para ampliar este trabajo.

- Repetir el estudio con información de otros cursos académicos y/o de otros centros del profesorado.
- Trabajar con técnicas de procesamiento del lenguaje natural sobre los campos de texto libre de las encuestas, los títulos, objetivos, contenido, observaciones, etc... buscando palabras clave, o aplicando análisis de sentimiento o minería de opinión.
- Ampliar el conjunto de datos contando con datos de ponentes y participantes.
- Buscar relaciones entre la formación del profesorado y el rendimiento de su alumnado.
- Contar con otras fuentes de información: Aula Virtual, Colabor@ o Moodle Centros.

# *Agradecimientos*

Este trabajo no habría sido posible sin la colaboración del Centro del Profesorado de Córdoba, en cuanto a la sesión de datos y la ayuda en la interpretación de los resultados. Especialmente, ha colaborado en la extracción y anonimización de datos, Paco España, como vicedirector del centro. El análisis de los resultados se ha apoyado en una comisión, ya existente, que trata de de buscar los factores que inciden en el abandono de las actividades.

# Referencias

- Agrawal, R., & Srikant, R. (1994). **Fast algorithms for mining association rules**. *Proc. 20th Int. Conf. very Large Data Bases, VLDB*, , 1215 487-499.
- Aneas Novo, M. C., Sánchez Rodríguez, J., & Sánchez Rivas, E. (2018). **Diseño y validación de dos cuestionarios dirigidos a evaluar la eficacia de la formación del profesorado**. *Docencia e investigación*, Retrieved from <https://revista.uclm.es/index.php/rdi/article/view/1715>
- Barber, M., & Mourshed, M. (2007). **How the world's best-performing schools systems come out on top**. McKinsey & Company.
- Cohen, W. W. (1995). **Fast effective rule induction**. *Machine learning proceedings 1995* (pp. 115-123) Elsevier.
- Day, C. (2005). **Formar docentes: Cómo, cuándo y en qué condiciones aprende el profesorado**. Narcea Ediciones.
- Frank, E., & Witten, I. H. (1998). **Generating accurate rule sets without global optimization**. *Proc. 15th International Conference on Machine Learning, July 1998. Pages 144-151*
- Hattie, J. (2013). **Visible learning: A synthesis of over 800 meta-analyses relating to achievement**. London, United Kingdom: Taylor & Francis.
- Imbernón Muñoz, F. (2007). **10 ideas clave. la formación permanente del profesorado: Nuevas ideas para formar en la innovación y el cambio** EDITORIAL GRAO. Retrieved from <https://books.google.es/books?id=D71SQZdH-5MC>
- Jung, Y. G., Kang, M. S., & Heo, J. (2014). **Clustering performance comparison using K-means and expectation maximization algorithms**. *Biotechnology & Biotechnological Equipment: Advanced and Applied Convergence*, 28(sup1), S44-S48. doi:10.1080/13102818.2014.949045
- Kitchin, R., & Dodge, M. (2011). **Code/space: Software and everyday life**. Mit Press.
- Ley orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales**. BOE núm. 294 de 6 de diciembre de 2018. sec. I. pág. 119788. (2018).
- Pineda Herrero, P. (2000). **Evaluación del impacto de la formación de las organizaciones**. *Educar*, (27), 119-133.
- Quinlan, R. (1993). **C4.5: Programs for machine learning**. San Mateo, CA.: Morgan Kaufmann Publishers.
- Romero, C., & Ventura, S. (2020). **Educational data mining and learning analytics: An updated survey**. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), n/a. doi:10.1002/widm.1355
- Williamson, B. (2015). **Governing software: Networks, databases and algorithmic power in the digital governance of public education**. *Learning*, 40 doi:10.1080/17439884.2014.924527