

## **ANEXO V. MEMORIA FINAL DE PROYECTOS. MODALIDADES 1, 2, 3 Y 4**

**CURSO ACADÉMICO 2018/2019**

### **DATOS IDENTIFICATIVOS:**

#### **1. Título del Proyecto**

*Potenciando el perfil profesional Científico de Datos mediante dinámicas basadas en competiciones y un grupo de trabajo sobre Aprendizaje Automático.*

#### **2. Código del Proyecto**

2018-1-5015

#### **3. Resumen del Proyecto**

*El Científico de Datos es un perfil profesional que requiere un amplio conocimiento en estadística, métodos de optimización, lenguajes de programación y su aplicación práctica en problemas reales. Por tanto, el papel de los Ingenieros Informáticos como Científicos de Datos es fundamental en la sociedad, ya que sus aptitudes y competencias adquiridas durante su formación se ajustan fuertemente a lo que requiere este puesto laboral.*

*El Científico de Datos está siendo uno de los puestos laborales más demandados en la actualidad y está creciendo a un ritmo cada vez mayor: las ofertas de empleo en el sector del Big Data aumentaron un 92% en 2015. Incluso en plataformas de búsqueda de empleo como LinkedIn, los roles de los científicos de datos han crecido más del 650% desde 2012. Así, debido a la necesidad de formación de nuevos científicos en el campo de la ciencia de datos, surgen plataformas en las que éstos pueden adquirir experiencia. Una de estas plataformas es Kaggle, que ofrece la posibilidad de que investigadores y empresas de todo el mundo publiquen datos para resolver un determinado problema. Los investigadores pueden acceder a dicha plataforma para formarse y resolver los problemas propuestos, adquiriendo experiencia práctica en problemas reales. Kaggle está disponible como herramienta de cloud computing, y por tanto, tiene una perspectiva abierta y distribuida. Esto hace que los usuarios puedan cooperar al mismo tiempo que competir, haciendo más sencilla la adquisición y el desarrollo de las competencias necesarias en ciencia de datos.*

*Estas competencias son imprescindibles para cubrir las necesidades de empresas públicas, privadas o gubernamentales. Es por ello que ya existen asignaturas en el plan de estudios del Grado de Ingeniería Informática, con contenidos estrechamente relacionados, abordando el aprendizaje automático, metaheurísticas, big data, y los modelos computacionales, entre otros. No obstante, aunque el nivel teórico de los contenidos cubre estas competencias, creemos que es necesario una aplicación práctica para enfrentar al alumnado a problemas reales. El equipo de trabajo externo podría proporcionar bases de datos reales relacionadas con finanzas, marketing, etc.*

*Por tanto, este proyecto pretende hacer uso de Kaggle como herramienta TIC en una asignatura del Grado de Ingeniería Informática, concretamente “Introducción al aprendizaje automático”. De esta forma, el alumnado puede adquirir las competencias más importantes en el área de ciencia de datos (preprocesamiento, entrenamiento, validación, ajuste de parámetros y predicción) en un entorno muy motivador, junto al entusiasmo del profesorado.*

*Entre las actividades realizadas se incluye la creación de un grupo de trabajo, en el que el profesorado tutorice, guíe y preste su experiencia a la hora de abordar competiciones reales; impartición de talleres para aprender el uso del lenguaje de programación Python junto con la librería scikit-learn a través del Aula de Software Libre de la UCO, y creación de una práctica de la asignatura que implique una competición real con fines docentes para familiarizar al alumnado con la herramienta.*

#### 4. Coordinador/es del Proyecto

Nombre y Apellidos	Departamento	Código del Grupo Docente
Juan Carlos Fernández Caballero	Informática y A.N.	146

#### 5. Otros Participantes

Nombre y Apellidos	Departamento	Código del Grupo Docente	Tipo de Personal
Antonio Manuel Durán Rosal	Informática y A.N.	146	Becario FPU, miembro del grupo Ayrna. Doctor en Informática.
Javier Sánchez Monedero	Escuela de Periodismo, Medios y Cultura, Universidad de Cardiff, investigador asociado.		Personal externo, antiguo profesor de la UCO.
Mónica de la Paz Marín	Jefatura del Servicio de Gestión de la Investigación de la Universidad de Córdoba.	146	PAS. Miembro del grupo Ayrna.
David Guijo Rubio	Informática y A.N	146	Becario FPU, miembro del grupo Ayrna.
Julio Camacho Cañamón	Graduado en Ingeniería Informática, Máster universitario en Investigación en Inteligencia Artificial. Máster en Profesorado de Educación Secundaria y Bachillerato. Máster en Competencias Docentes Avanzadas.		Personal externo, antiguo colaborador honorario con el Informática y A.N, y contratado con cargo a proyecto.
Sergio Gómez Bachiller	Coordinador del Aula de Software Libre.		PAS
Carlos Castillo Botón	Graduado en Ingeniería Informática.		Antiguo contratado con cargo a proyecto en el grupo Ayrna.
Víctor Manuel Vargas Yun	Informática y A.N	146	Becario FPU, miembro del grupo Ayrna.
Francisco Martínez Estudillo	Departamento de Métodos Cuantitativos de Universidad Loyola Andalucía (sede de Córdoba).		Personal externo, miembro del grupo Ayrna.
Mercedes Torres Jiménez	Departamento de Métodos Cuantitativos de Universidad Loyola Andalucía (sede de Córdoba).		Personal externo, miembro del grupo Ayrna.

#### 6. Asignaturas implicadas

Nombre de la asignatura	Titulación
Introducción al aprendizaje automático	Grado en Ingeniería Informática, curso 3º, especialidad en Computación.

## 7. Memoria final del proyecto

A continuación se expone la memoria final del presente proyecto de innovación docente del curso 2018-2019, bajo la Modalidad 1 del “Plan de Innovación y buenas Prácticas docentes 2018/2019” de la UCO.

### 7.1 Introducción

El Científico de Datos es un perfil profesional atribuido a personas que trabajan en la ciencia de datos y muy ligado a los Ingenieros Informáticos. Requiere fundamentos en estadística, métodos de optimización y una buena formación en lenguajes de programación de alto nivel.

Un aspecto fundamental de este perfil es que se requiere experiencia práctica en problemas reales, y está siendo uno de los puestos laborales más demandados en la actualidad, creciendo a un ritmo cada vez mayor: las ofertas de empleo en el sector del *Big Data* aumentaron un 92% en 2015 [1]. Incluso en plataformas de búsqueda de empleo como *LinkedIn*, los roles de los científicos de datos han crecido más del 650% desde 2012 y cientos de compañías están contratando esos roles [2]. Algunos analistas confirman que las empresas que adopten técnicas de analítica de Big Data tendrán una ventaja competitiva de 20% en todas las métricas financieras sobre sus competidores [3].

La Ciencia de Datos es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva y tiene una estrecha relación con la Inteligencia Artificial (IA). La ciencia de datos se nutre mucho del Aprendizaje Automático (rama de la IA), que es capaz de “aprender de los datos”, es decir, de conseguir que el ordenador aprenda a realizar una tarea nueva a partir de una base de datos de ejemplos, sin haber sido explícitamente programado para hacer dicha tarea.

La enorme demanda que existe hacia el perfil del Científico de Datos ha hecho que hayan proliferado distintas plataformas en las que los científicos pueden demostrar su habilidad y las empresas pueden reclutar nuevos empleados en función de su capacidad de resolver problemas de este estilo. De entre ellas, la más importante es *Kaggle* [4], fundada en 2010 con el objetivo de albergar competiciones de aprendizaje automático. *Kaggle* ofrece una plataforma en la que empresas, investigadores y científicos de datos de todo el mundo publican sus bases de datos para resolver un determinado problema. *Kaggle* organiza concursos en los que los Científicos de Datos compiten por la oportunidad de conseguir una entrevista en empresas líderes en Ciencia de Datos como *Facebook*, *Netflix*, *Winton Capital* y *Walmart*, por un premio en metálico, o simplemente por darse a conocer y aportar soluciones a la ciencia. Los participantes pueden competir, ver los resultados que obtienen e iterativamente mejorar sus algoritmos.

El conjunto de usuarios registrados en *Kaggle* recoge personas de todo tipo de disciplinas profesionales, lo que enriquece las soluciones y permite resolver problemas de todo tipo. *Kaggle* está disponible en la nube como herramienta de *cloud computing* y esto le confiere una perspectiva abierta y distribuida. Además, la competición hace que los usuarios terminen colaborando entre sí, copiando estrategias y complementándolas con sus propias formas de proceder. La participación en este tipo de competiciones enriquece muchísimo el desarrollo de competencias en Ciencia de Datos, debido a que, en los distintos foros de la plataforma, se pueden consultar estrategias de programación muy bien diseñadas.

Actualmente, las competencias en ciencia de datos son imprescindibles para cubrir las necesidades de empresas públicas, privadas o instituciones gubernamentales. Las universidades deberían poder satisfacer esta demanda, ofreciendo profesionales formados, al menos parcialmente, en estas tareas. Debido a esta necesidad, el plan de estudios de titulaciones como el Grado en Ingeniería Informática incluye asignaturas con contenidos relacionados. Aunque a nivel teórico y a cierto nivel práctico sí que se cubren competencias básicas necesarias para este perfil, consideramos que es necesario complementar esos conocimientos con una experiencia práctica que enfrente al alumnado a la complejidad que suponen los problemas reales.

Desde un punto de vista docente, está demostrado que utilizar nuevas tecnologías en el aula es una forma de facilitar el aprendizaje del alumnado, estimulando la motivación y creando un ambiente de colaboración. Uno de los aspectos más criticados por los egresados en las distintas titulaciones universitarias

es que la formación recibida es muy teórica y fuertemente generalista [5], con lo que se hace esencial llevar los conocimientos teóricos a la práctica.

Esto nos lleva a concluir que, debido a la demanda de competencias en ciencia de datos y a la necesidad de mejorar las estrategias docentes en el aula, la participación del alumnado en la plataforma *Kaggle* es muy ventajosa. Los alumnos se enfrentarían a problemas reales, se nutrirían de soluciones proporcionadas por otros alumnos u otros investigadores y la competitividad impulsaría su formación (al igual que sucede con las técnicas de *gamificación*).

Por ello, este proyecto ha utilizado *Kaggle* como herramienta TIC en una asignatura del Grado de Ingeniería Informática que oferta la Universidad de Córdoba, de manera que permita al alumnado aplicar y comprobar en la práctica los conocimientos teóricos adquiridos dentro de los programas de la asignatura. Las competencias más importantes en el área de ciencia de datos corresponden a preprocesamiento de bases de datos, entrenamiento, validación, ajuste de parámetros y evaluación de modelos predictivos. La plataforma *Kaggle* puede ser un entorno muy motivador para adquirir estas competencias, junto con el entusiasmo del profesorado. En concreto, la asignatura implicada en el proyecto ha sido “Introducción al Aprendizaje Automático”, 3º curso de Grado en Ingeniería Informática, especialidad en Computación.

Parece interesante que de cara a la asignatura implicada se utilicen competiciones más simples, con fines docentes. *Kaggle* ofrece esta posibilidad, por lo que el presente proyecto ha creado una competición docente y privada. Para crear una experiencia completa para el alumnado en *Kaggle*, se solicitó como presupuesto un premio simbólico en material de reprografía para los grupos de alumnos que consiguieran las puntuaciones más altas.

Como objetivo adicional, para seguir nutriendo la formación recibida en el aula con experiencias reales, se han formado grupos de trabajo en el que el profesorado ha tutorizado, guiado y prestado su experiencia a la hora de abordar competiciones reales. Como parte de ese grupo de trabajo, se han impartido dos talleres en ciencia de datos utilizando el lenguaje Python y la librería Scikit-learn, con el apoyo del Aula de Software Libre de la Universidad de Córdoba. De esta forma, el proyecto utiliza exclusivamente *software* libre, potenciando así la línea de actuación a favor del *software* libre marcada tanto por la Conferencia de Rectores de las Universidades Españolas (CRUE), con la creación del Grupo de Trabajo Software Libre dentro de CRUE-TIC, como por la Universidad de Córdoba con la creación del Aula de Software Libre. Las razones pedagógicas y sociales de por qué debe utilizarse *software* libre en los centros de enseñanza se sintetizan en el artículo [6].

Debemos indicar que el proyecto se ajusta a una de las líneas de acción prioritaria establecidas en la modalidad I del Plan de Innovación y buenas Prácticas docentes (2018/2019): transferencia del conocimiento teórico a la práctica.

## 7.2 Objetivos

A continuación se indican los objetivos planteados y conseguidos con este proyecto:

1. Mejorar la adquisición de aspectos prácticos de las competencias específicas del perfil profesional científico de datos, en la asignatura “Introducción al aprendizaje automático”, 3º curso de Grado en Ingeniería Informática, especialidad en Computación.
2. Instruir al alumnado en el uso de la plataforma *Kaggle* como un medio de aprendizaje y motivación, para que, de una manera amena y divertida, se compruebe la aplicabilidad que tiene los conceptos de aprendizaje automático que se estudian en la asignatura afectada.
3. Incluir una o varias sesiones prácticas en la asignatura involucrada. Concretamente, se han utilizado competiciones con fines docentes, organizadas por el profesorado de la asignatura, estableciendo unas pautas básicas para utilizar las herramientas estudiadas en la asignatura en dicha competición. Es decir, la experiencia se ha realizado en un entorno controlado y limitando el grado de dificultad.

4. Fomentar un ambiente tanto de cooperación como de competitividad entre el alumnado, con el objetivo de alcanzar los mejores resultados en los problemas.
5. Enfrentar a los alumnos a la complejidad real de los problemas de Aprendizaje Automático, mediante el grupo de trabajo.
6. Cubrir parte de las competencias CTEC4, CTEC5 y CTEC7 de la asignatura “Introducción al aprendizaje automático”, 3º curso de Grado en Ingeniería Informática, especialidad en Computación.

### 7.3 Material y métodos

Los objetivos anteriormente mencionados se han llevado mediante la realización de las siguientes actividades:

1. **Actividad 1.** Búsqueda y adaptación, por parte de los integrantes del proyecto, de un problema a ser susceptible de resolverse mediante técnicas de aprendizaje automático. El profesorado tendrá que adaptar la base de datos a *Kaggle*, y configurar y entender cómo se gestiona la competición.

Para ello se ha recurrido a un problema real abordado por el grupo de investigación del que forman parte los participantes de este proyecto (Grupo AYRNA - <http://www.uco.es/ayrna/>), concretamente un problema de predicción de altura de ola en una zona del Golfo de Alaska [7]. El problema se ha adaptado a datos más actuales y a boyas concretas, y se ha incluido mucha más información en cuanto a las características a partir de las cuales se podría obtener un modelo de predicción. Ello obliga a que para obtener buenos resultados el alumnado tenga que emplear técnicas de selección de características y de pre-procesado de datos. Toda la información proporcionada al alumnado se ha adaptado al formato que exige la plataforma *Kaggle* y proviene de una boya real situada en el océano.

2. **Actividad 2.** Crear una práctica para la asignatura mencionada en las secciones anteriores, que introduzca la plataforma, explique su funcionamiento, ejemplifique el envío de resultados y proporcione indicaciones de cómo se establecen los distintos *rankings*.

Para ello, el equipo de trabajo se ha encargado de redactar una práctica de 3 sesiones de duración (6 horas, 2 horas por sesión de clase). También ha sido necesario crear la competición de *Kaggle* en la nube, disponer la base de datos con la información sobre la predicción de altura de ola, crear las normas de competición, configurar el entorno de trabajo en la nube y todo lo necesario para que la plataforma esté lista para subir resultados por parte de los alumnos.

3. **Actividad 3.** Llevar a cabo un conjunto de sesiones prácticas en las que se expliquen los contenidos elaborados en la Actividad 2. Evaluar al alumnado en función tanto de su posición en el *ranking* privado como de la forma en que se explique la solución aplicada.

Las clases prácticas las ha impartido, en presencia del coordinador de este proyecto y responsable de la asignatura, el Graduado en Ingeniería Informática y colaborador honorario del Dpto. De Informática y Análisis Numérico, Antonio Manuel Gómez Orellana, miembro actualmente del grupo de investigación AYRNA y con experiencia en Ciencia de Datos.

El 40% de la nota de la práctica se ha puntuado en función del *ranking* privado y el 60% restante en función del preprocesado y metodología seguida por el alumnado para la obtención de modelos de aprendizaje.

4. **Actividad 4.** Creación de un equipo de trabajo con alumnos. Este equipo de trabajo estará mentorizado por varios de los profesores que participan en el proyecto. Se planificarán reuniones periódicas en las que se establecerán distintas estrategias para competir en las competiciones reales, de mayor complejidad.

Se formaron equipos de trabajo, donde cada equipo cooperó internamente y compitió con el resto de equipos, siempre mentorizados por el profesorado y por Antonio Manuel Gómez Orellana, que acompañó durante toda la competición a los participantes de los equipos. Se mantuvieron con frecuencia reuniones con los equipos en horarios de tutorías y también fuera de esos horarios oficiales (a petición de los alumnos), tanto en forma presencial como virtual.

5. **Actividad 5.** Impartición de una serie de talleres en ciencia de datos, utilizando *software* libre, bajo el auspicio del Aula de Software Libre de la Universidad de Córdoba. La idea es que los propios miembros del grupo de trabajo participen en dichos talleres, ya sea como ponentes o como alumnos.

En la asignatura IAA se enseña al alumnado a crear modelos de predicción utilizando la herramienta software Weka [8], muy usada por la comunidad científica como entorno de trabajo para preprocesar datos y con multitud de algoritmos que crean modelos predictivos, pero también se ha querido añadir la posibilidad de usar Python y Scikit-Learn por ser otra opción que actualmente está teniendo una gran repercusión y uso a nivel mundial. Se contactó entonces con el Aula de Software Libre de la UCO, que se brindó a colaborar, preparar los ordenadores de las aulas y difundir los seminarios entre el alumnado. Concretamente se realizaron las siguientes 2 sesiones, impartidas por el profesor Pedro Antonio Gutiérrez Peña, Profesor del Dpto. De Informática de la Universidad de Córdoba y miembro del grupo de investigación AYRNA, y por el Graduado en Informática Antonio Manuel Gomez Orellana:

- **Visualización, aprendizaje supervisado y métodos de evaluación (jueves 21 febrero):** Se realizó en el Aula P1 del Ramón y Cajal del Campus de Rabanales, en horario de 16:00-18:00.
  - **Promoción de los seminarios:**
    - <https://consejo-eps.uco.es/aulasoftwarelibre/actividades/idea/seminarios-de-introduccion-a-la-ciencia-de-datos-y-aprendizaje-automatico-con-python>
    - <https://www.uco.es/aulasoftwarelibre/2303-seminarios-de-introduccion-a-la-ciencia-de-datos-y-aprendizaje-automatico/>
  - **Material para los seminarios:**
    - <https://github.com/ayrna/taller-sklearn-asl-2019>
- **Aprendizaje no supervisado (jueves 14 marzo):** Se realizó en el Aula P1 del Ramón y Cajal del Campus de Rabanales, en horario de 16:00-18:00.

6. **Actividad 6.** Para valorar la eficacia del proyecto, se realizó una encuesta al alumnado antes y después de realizar la actividad, tratando de valorar los conocimientos prácticos que disponen sobre tareas aplicadas de ciencia de datos (preprocesamiento, modelado, evaluación de rendimiento...).

Se han realizado encuestas voluntarias de tipo test al alumnado, una antes de empezar la competición y los talleres-seminarios, y otra al finalizar el proyecto, tratando de valorar los conocimientos prácticos adquiridos. Para este fin el equipo de trabajo utilizó una herramienta que ofrece Google, <https://docs.google.com/forms/>, concretamente se usaron “test de autoevaluación”, accesibles en:

- **Cuestionario previo:**
  - <https://docs.google.com/forms/d/e/1FAIpQLSeBbN5epx8HxPj8ZduWnOevirtLvDkIpc94i5ApBDwP27nCOA/viewform>
- **Cuestionario posterior:**
  - [https://docs.google.com/forms/d/e/1FAIpQLScqn6KxUzkN\\_89gJ-CbknEwIx8FHIoVoGGZWYnwLjDaPXa1JA/viewform](https://docs.google.com/forms/d/e/1FAIpQLScqn6KxUzkN_89gJ-CbknEwIx8FHIoVoGGZWYnwLjDaPXa1JA/viewform)

Se incluyeron un total de 16 preguntas de tipo test para obtener un sondeo del nivel inicial de los alumnos. Por otro lado, se incluyen 4 preguntas para auto-evaluar los conocimientos sobre la plataforma Kaggle y las bibliotecas de Python para aprendizaje automático, de forma que el equipo de trabajo de este proyecto pueda ratificar si son tecnologías nuevas y conocidas por el alumnado. Las 4 preguntas siguen una escala de Likert, desde “muy reducido” hasta “lo domino en profundidad”: 1) Valora tus conocimientos sobre Python, 2) Valora tus conocimientos sobre scikit-learn, 3) Valora tus conocimientos sobre aprendizaje automático, 4) Valora tu capacidad para abordar una competición de Kaggle.

Los resultados que se han obtenido, descritos de manera breve para no alargar en demasía el contenido de la presente memoria, son los siguientes:

■ **Resultados antes de realizar el proyecto:**

- **Para las 16 preguntas tipo test:** Cada respuesta acertada vale 1 punto, con lo que se pueden obtener un máximo de 16. Dicho esto se puede observar en la Figura 1 que el promedio es de 9 sobre 16, más o menos un 6 sobre 10. La normal o mediana se encuentra en 8,42, con lo que refuerza la idea de que más o menos la mitad de los encuestados obtendrían un 5 sobre 10, y por último el intervalo de puntuaciones está entre 4 y 12, con lo cual indica que hay alumnos que tienen conocimientos muy bajos.

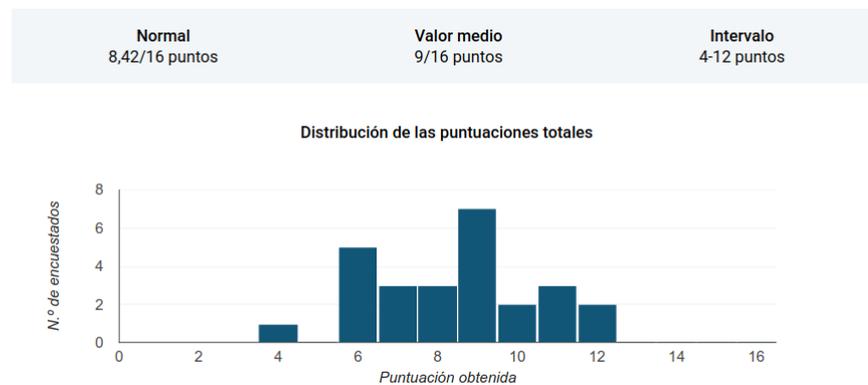


Figura 1: Distribución de puntuaciones previas a proyecto

- **Para las 4 preguntas sobre auto-evaluación de la nueva tecnología utilizada:** En este caso el mayor porcentaje de contestación se encuentra en las posiciones 1 y 2 de la escala de Likert, es decir, que el alumnado tiene conocimientos muy reducidos o bajos sobre este tipo de tecnologías a nivel práctico.

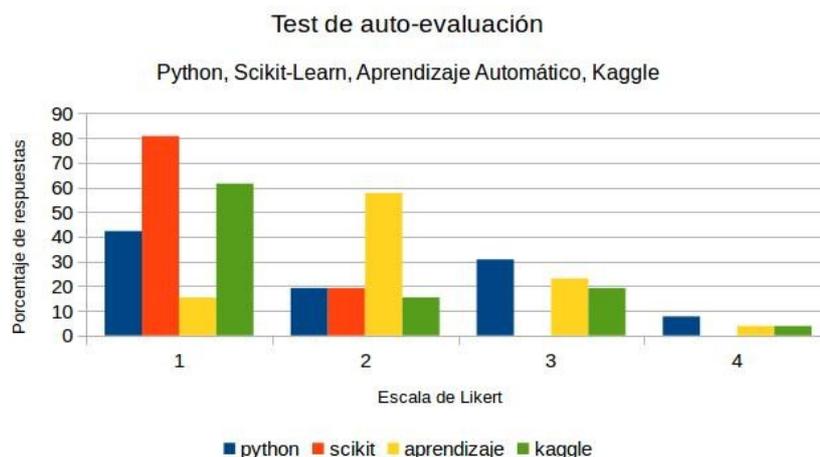


Figura 2: Autoevaluación previa con malos resultados

■ **Resultados después de realizar el proyecto:**

- **Para las 16 preguntas tipo test:** En la Figura 3 se muestra que el promedio de los que han contestado es de 10 sobre 16, subiendo un punto. La mediana sube también a 10.63, y el valor mínimo se encuentra en 7 puntos sobre 16. Esto quiere decir que el alumnado tiene mayor formación sobre los contenidos iniciales al terminar la experiencia docente, aunque se esperaba que las diferencias fueran mayores.

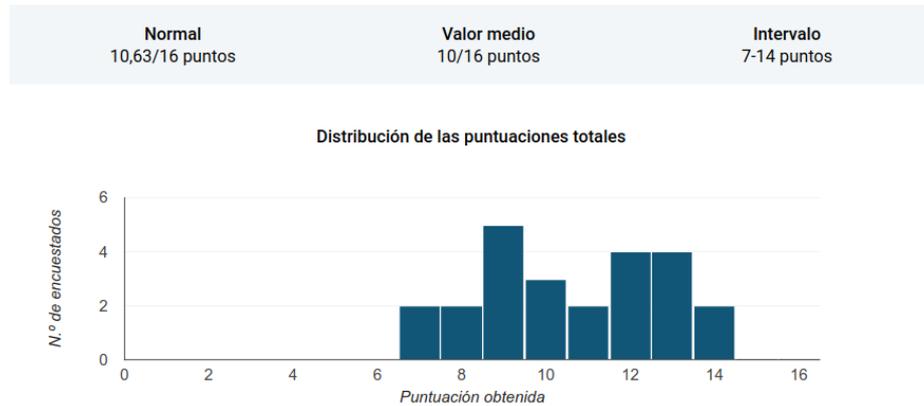


Figura 3: Distribución de puntuaciones posterior a proyecto

- **Para las 4 preguntas sobre auto-evaluación de la nueva tecnología utilizada:** En este caso los resultados afirman lo positivo de la formación después del proyecto con respecto a las tecnologías utilizadas, siendo las posiciones 3 y 4 de la escala de Likert las más usadas, es decir, que el alumnado tiene conocimientos de medios a altos sobre este tipo de tecnologías a nivel práctico.

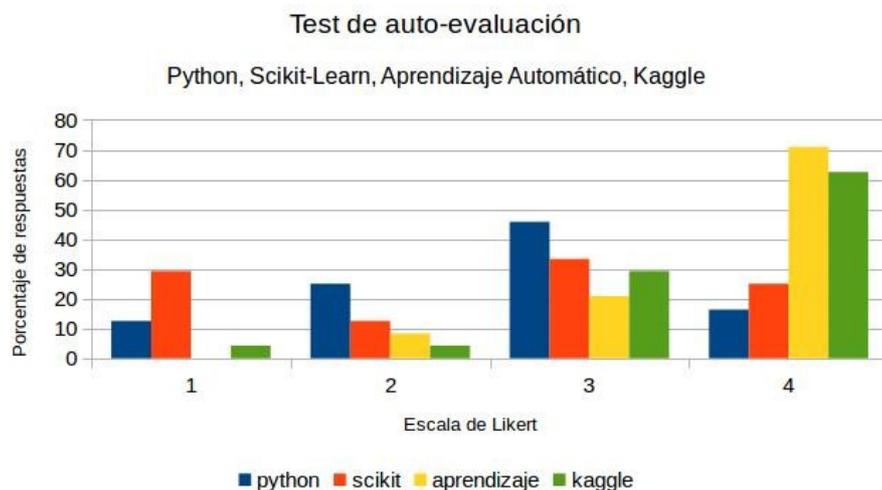


Figura 4: Autoevaluación posterior invirtiendo los resultados

**7.3.1 Presupuesto para el proyecto**

El presupuesto que se solicitó para este proyecto fue de 200 euros, a usar en concepto de premios para los alumnos con mejores puestos en el ranking de la competición, concretamente en tickets de material fungible en Copisterias Don Folio de Córdoba.

Las Figura 5 muestra una copia escaneada de un ticket original, tanto en su parte delantera como trasera.

### 7.3.2 Ranking en la competición

En cuanto a diplomas y rankings, los resultados fueron los que se exponen a continuación, medidos en función del valor de la métrica F-Score, muy usada en Aprendizaje Automático para obtener la bondad de un modelo en cuanto a su capacidad de predicción. En <https://www.kaggle.com/c/competition-iaa-2018-2019> se puede ver la competición y sus datos públicos (los premios van en función del ranking privado). Se formaron un total de 18 equipos de trabajo en la competición, y el reparto en tickets para los ganadores se ha realizado entre los 4 primeros de la siguiente manera:

- 1er premio: 80 € - José Manuel Cuevas Muñoz y Adrián López Ortiz.
- 2o premio: 60 € - Tomás Espejo Gómez y Victoriano Pedrajas Fernández.
- 3er premio: 40 € - José Luis Gordillo Relaña y Ismael Abu-jadur García.
- 4o premio: 20 € - Alejandro Del Moral Rodríguez.

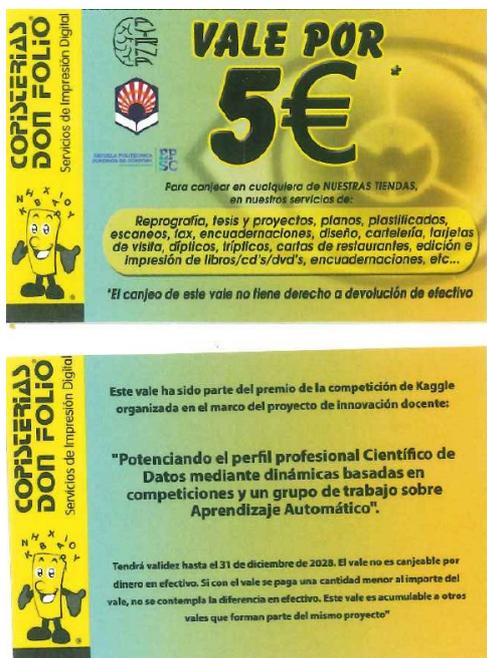


Figura 5: Ticket de 5 euros para material fungible en Don Folio Copisterias

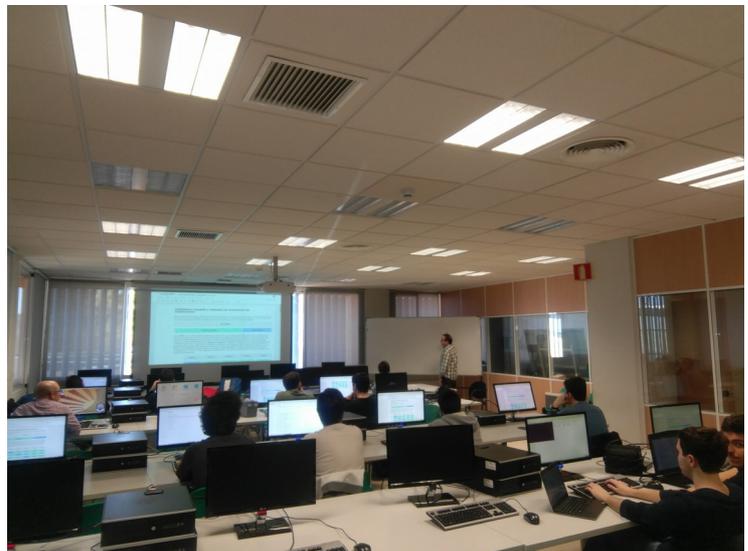


Figura 6: Uno de los seminarios impartidos con el apoyo del Aula de Software Libre

### 7.4 Resultados obtenidos y discusión

Los resultados que se han obtenido en esta experiencia docente, al entender de los participantes de este proyecto de innovación, son los siguientes:

- Se ha dado a conocer al alumnado una serie de herramientas profesionales y de actualidad para el modelado de problemas reales mediante técnicas de Aprendizaje Automático.
- Se ha conseguido que el alumnado se cerciore de que el perfil profesional Científico de Datos es cada vez más demandado, y que puede ser una exitosa salida profesional. Muestra de ello es que en la plataforma Kaggle participan las principales empresas de Ingeniería Informática del mundo.
- Se ha dado a conocer al alumnado las diversas tareas que realiza el científico de datos, entre ellas el preprocesado de datos, con la finalidad de poder profundizar en un futuro.
- Se han mejorado a nivel práctico las competencias a adquirir por el alumnado en la asignatura de IAA.

- Se ha conseguido que el alumnado aporte y reciba nuevos conocimientos prácticos a partir de la gran comunidad existente en la plataforma Kaggle, formada por investigadores y Científicos de Datos de diversos ámbitos, ámpliamente consolidados.
- Se ha fomentado el uso de TICs en el aula y se ha conseguido una buena participación en seminarios (Figura 6).
- Se ha fomentado el trabajo en grupo, ya sea entre participantes de un mismo grupo de competición como entre grupos, haciendo que el feedback y la relaciones entre ellos culmine en una mejora de conocimiento para todos.
- Se ha aumentado la confianza del alumnado en sí mismo para involucrarse en problemas reales donde pueda aplicar sus conocimientos teórico-prácticos para dar una solución tangible.

### **7.5 Conclusiones**

Como conclusión final, ligada a los resultados obtenidos y la discusión realizada en la sección anterior, a nuestro parecer este tipo de proyectos de innovación docente mejoran de manera notable la capacidad del alumnado para enfrentarse a un problema real cuando terminen su titulación. En este sentido, el aplicar técnicas actuales que el equipo de participantes del proyecto usa en su investigación, supone darles la motivación y capacidad para el aprendizaje y reciclaje continuo de conocimientos a los que tienen que someterse los Ingenieros en Informática.

Es la segunda experiencia de Kaggle que se realiza y la primera ya ha dado lugar a que los alumnos se interesen cada vez más en realizar Trabajos Fin de Grado (TFGs) relacionados con la investigación y el aprendizaje automático.

El grado de implicación y de interés de muchos de los alumnos ha sido muy alto desde el inicio.

Con respecto al esfuerzo del equipo de trabajo hay que reseñar que ha sido elevado, ya que ha habido que adaptar al nivel del alumnado todo el material de los seminarios y la práctica docente en IAA, pero la satisfacción conseguida hace que haya merecido la pena. Se espera poder realizar nuevas ediciones en cursos académicos futuros.

### **7.6 Agradecimientos**

Agradecemos al Aula de Software Libre de la Universidad de Córdoba y a la Copistería Don Folio su colaboración en la realización de este proyecto de innovación docente. Por otro lado, agradecemos al Graduado Antonio Manuel Gómez Orellana su implicación en el proyecto como colaborador honorario del Dpto. De Informática y Análisis Numérico, y al Profesor Pedro Antonio Gutiérrez Peña por haber impartido parte de los seminarios y proporcionar material didáctico sobre Python y Scikit-Learn.

### **Bibliografía**

- [1] Andrea Núñez-Torrón Stock. “Las ofertas de empleo en el sector del Big Data aumentaron un 92% en 2015”. [Online]. Disponible en <http://www.ticbeat.com/tecnologias/ofertas-empleo-big-data/>
- [2] LinkedIn's Fastest-Growing Jobs Today Are In Data Science And Machine Learning. [Online]. Disponible en <https://www.forbes.com/sites/louiscolumbus/2017/12/11/linkedins-fastest-growing-jobs-today-are-in-data-science-machine-learning/#3d8fbec651bd>
- [3] Gustavo Tamaki. “La era del Big Data”. [Online]. Disponible en <http://www.techroi.pe/thechroi/13/82/la-hora-del-big-data>
- [4] Kaggle, Your Home for Data Science (2017, Junio). [Online]. Disponible en: <https://www.kaggle.com/>
- [5] Agencia Nacional de la Evaluación de la Calidad y Acreditación (ANECA), Los procesos de inserción laboral de los titulados universitarios en España. Factores de facilitación y de obstaculización, 2009. (2017, Junio). [Online]. Disponible en: [http://www.aneca.es/media/308144/publi\\_procesosil.pdf](http://www.aneca.es/media/308144/publi_procesosil.pdf)

- [6] Richard Stallman. Por qué las escuelas deberían usar exclusivamente software libre. [Online]. Disponible en: <http://www.gnu.org/philosophy/schools.es.html>
- [7] FERNANDEZ, J. C., SALCEDO-SANZ, S., GUTIÉRREZ, P. A., ALEXANDRE, E., HERVÁS-MARTÍNEZ, C., Significant wave height and energy flux range forecast with machine learning classifiers, Engineering Applications of Artificial Intelligence, vol. 43, pages 44–53, 2015.
- [8] FRANK, E., HALL, M. A., WITTEN, I. H., The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.