



Hackathon en docencia: Aprendizaje Automático aplicado a Ciencias de la Vida

Hackathon in teaching: Machine Learning applied to Life Sciences

David Guijo-Rubio¹, Víctor M. Vargas², Javier. Barbero-Gómez³, Jose V. Die⁴,
Pablo González-Moreno⁵

Fecha de recepción: 23/02/2022; Fecha de revisión: 20/07/2022; Fecha de aceptación: 01/09/2022

Cómo citar este artículo:

Guijo-Rubio D., Vargas, V. M., Barbero-Gómez, J., Die, J. V., & González-Moreno, P. (2022). *Hackathon en docencia: Aprendizaje Automático aplicado a Ciencias de la Vida*. *Revista de Innovación y Buenas Prácticas Docentes*, 11(2), 19-37.

Autor de Correspondencia: dguijo@uco.es

Resumen:

La programación ha sido tradicionalmente una competencia perteneciente a las ingenierías, que recientemente está adquiriendo una importancia significativa en áreas como Ciencias de la Vida, donde resulta fundamental para la resolución de problemas de análisis de datos. Este trabajo es un caso de estudio enmarcado en la necesidad de mejorar las habilidades, sobre análisis de datos en el alumnado de Ciencias de la Vida y de la base temática en los estudiantes de ingeniería. Mediante la herramienta del *hackathon* y el trabajo en equipo, se combinó al alumnado de ambas disciplinas y se le enfrentó a una serie de problemas de análisis de datos. Se establecieron equipos de trabajo que recibieron una formación previa al comienzo de la competición. De cada equipo, se valoró la metodología empleada para la obtención de los datos, su análisis, interpretación de resultados, y exposición de las diversas tareas. Se hizo un análisis descriptivo de los resultados del Proyecto mediante encuestas al alumnado, así como su percepción sobre las actividades realizadas. El Proyecto ha conseguido que el alumnado resuelva los problemas planteados, difícilmente abordables con equipos unidisciplinarios, generando un aprendizaje común y una experiencia multidisciplinaria altamente satisfactoria tanto para el alumnado como para el profesorado.

Palabras clave: experiencia profesional; interdisciplinariedad; inteligencia artificial; perfil profesional

Abstract:

Programming has traditionally been an engineering competence, but recently it is acquiring significant importance in several areas, such as Life Sciences, which is considered essential for problem-solving based on data analysis. This work is a case study framed within the need to improve not only the data analysis skills of life science students, but also the biological background concerning the given issue of engineering students. Using hackathon and teamwork-based tools, students from both disciplines have been made and challenged with a series of problems in the area of Life Sciences. To solve these problems, we established work teams trained before the competition's beginning. Their results were assessed concerning the approach to obtain the data, perform the analysis, and finally interpret and present the results to solve the challenges. The project outcomes were assessed using structured surveys for students and their overall perception. The project succeeded, meaning students solved the proposed problems and achieved the activity's goals. These goals would have been difficult to address with teams composed of students from the same field of study. The hackathon succeeded in generating a shared learning and a multidisciplinary experience for their professional training, being highly rewarding for both students and faculty members.

Key Words: professional experience; interdisciplinarity; artificial intelligence; professional profile

¹ Universidad de Córdoba (España), dguijo@uco.es, 0000-0002-8035-4057

² Universidad de Córdoba (España), vvargas@uco.es, 0000-0002-0700-275X

³ Universidad de Córdoba (España), jbarbero@uco.es, 0000-0002-9317-1428

⁴ Universidad de Córdoba (España), jose.die@uco.es, 0000-0002-7506-8590

⁵ Universidad de Córdoba (España), pablo.gonzalez@uco.es, 0000-0001-9764-8927

1. INTRODUCCIÓN

El aprendizaje y aplicación de lenguajes de programación se ha instalado en los últimos años en disciplinas más allá de las tradicionales relacionadas con el área de Ciencias de la Computación, permeando en todas las áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), en general y en el ámbito de Ciencias de la Vida en particular, principalmente por la facilidad que propone a la hora de resolver numerosos procedimientos actuales, sin conllevar una carga de trabajo o recursos excesivos (Markowitz, 2019; Perkel, 2021; Kuter, & Wędrychowicz, 2021). Este nuevo escenario exige a las Universidades preparar profesionales que ocupen el perfil de analista de datos, incorporando el aprendizaje de lenguajes de programación y el trabajo en equipos interdisciplinarios, como competencias transversales en otras titulaciones más allá de las tradicionales donde se venía realizando. El perfil del analista de datos, de enorme importancia en la actualidad y con gran proyección a futuro, tanto en el área de las Ciencias de la Computación (Bureau, 2020), como en aquellas relacionadas con las Ciencias de la Vida (Brunner, & Kim, 2016; Ramamurthy, 2016; Hicks, & Irizarry, 2018).

Se han desarrollado diversas experiencias docentes innovadoras relacionadas con la Ciencia de Datos, siendo bastante exitosas las relacionadas con el aprendizaje en base a proyectos. Por ejemplo, Pournaras (2017) ilustra la experiencia y las lecciones aprendidas en relación a un novedoso curso interdisciplinario de Ciencia de Datos. En este trabajo se propone un curso basado en proyectos abiertos a distintas disciplinas. A pesar del reto al que se enfrentan los estudiantes, el estudio valora la motivación y libertad que dan los proyectos para mejorar el aprendizaje. De manera similar, Chua et al. (2019) presentó un programa piloto para introducir la Ciencia de Datos en un instituto de Singapur. En este caso de estudio que duró cinco meses, el alumnado pertenecía a una misma disciplina, a la cual introdujeron la Ciencia de Datos. Como conclusión, los estudiantes confirmaron que el modelo de curso facilitó el aprendizaje de la Ciencia de Datos de una manera clara, interesante y motivadora.

Con el objetivo de presentar una primera actividad con este trasfondo, se obtuvo un Proyecto de Innovación Docente (PID) por la Universidad de Córdoba durante el curso 2020/2021. Este Proyecto ha hecho una aproximación a este nuevo escenario incluyendo a estudiantes y profesorado de diferentes áreas de conocimiento (ingeniería forestal, informática y agronómica), aplicando herramientas docentes innovadoras como el *hackathon* (Kross, & Guo, 2019; National Academies of Sciences, Engineering and Medicine, 2018). El *hackathon* consistió en una serie de sesiones de trabajo donde se plantearon problemas reales de las Ciencias de la Vida con equipos formados por perfiles de ambos campos (Ciencia de Datos y Ciencias de la Vida), que compartían conocimientos a lo largo de todo el proceso de trabajo. Esta herramienta del *hackathon* posibilitó además el desarrollo de otras competencias, como por ejemplo el trabajo en equipo entre diferentes áreas, fomentando la interdisciplinariedad y el conocimiento de primera mano de un perfil profesional. Dicha interdisciplinariedad es de vital importancia para los Científicos de Datos: desde la perspectiva práctica, en la que deben tener conocimientos del proceso de análisis en cuestión y desde la perspectiva técnica, en la que deben tener

conocimientos sobre las herramientas, algoritmos y técnicas a utilizar en cada situación (Asamoah et al., 2015).

El PID se basó principalmente en la resolución de dos retos: por un lado, la tarea de visualización, asociada a un problema del área de Biotecnología y Genética, consistente en mostrar los componentes principales y categorías de ontología génica en el proceso de la anotación funcional de genes. Por otro lado, enfrentarse al reto del análisis predictivo, en el área de la Ingeniería Agroforestal, consistente en el estudio de la biodiversidad forestal en Andalucía, así como de la interpretación de los resultados obtenidos.

La participación en el *hackathon* se realizó por equipos interdisciplinarios de las dos áreas, fomentando, por un lado, la cooperación y colaboración entre personas de un mismo equipo, permitiendo así compartir conocimientos y experiencias previas de ambos componentes: los participantes del área de Ciencias de Datos pudieron compartir conocimientos básicos, profundizar en la resolución de nuevos problemas y presentar ideas para su resolución. Los participantes del área de Ciencias de la Vida pudieron aportar su conocimiento teórico a los dos problemas a resolver, aportando conocimiento experto, además de profundizar en el conocimiento de nuevas herramientas para la resolución de problemas complejos difícilmente abordables sin ellas. Por otro lado, también se fomentó la competitividad sana entre equipos, con el objetivo principal de alcanzar la mejor solución posible.

Tras la realización de las diferentes sesiones de trabajo, los resultados obtenidos fueron altamente satisfactorios, consiguiendo muchos de los objetivos marcados por el PID y animando al alumnado a formarse de forma complementaria a su formación universitaria curricular.

2. DESARROLLO DE LA EXPERIENCIA DE INNOVACIÓN

El PID se enfocó en buscar la sinergia entre dos áreas diferentes, Ciencia de Datos y Ciencias de la Vida para desarrollar una experiencia transversal y multidisciplinar. Más concretamente, las asignaturas involucradas fueron:

- Grado en Ingeniería Informática:
 - Introducción a Modelos Computacionales (4º curso).
- Máster en Geomática. Teledetección y modelos espaciales aplicados a la gestión forestal:
 - Fundamentos Matemáticos de Programación (1º curso).
- Máster en Biotecnología:
 - Análisis e Interpretación de Genomas (1º curso).

2.1 Objetivos.

Los objetivos principales del PID fueron los siguientes:

1. Mejorar la adquisición de competencias relacionadas con el Aprendizaje Automático (en inglés, *Machine Learning*), basado en el análisis de datos para resolver problemas complejos.
2. Ampliar los conocimientos sobre los lenguajes de programación y los conceptos básicos asociados a las Ciencias de la Vida y la Naturaleza.

3. Utilizar dinámicas de competición con fines docentes para la resolución de los problemas.
4. Fomentar un ambiente de cooperación y competitividad al mismo tiempo entre el alumnado, con la finalidad de progresar y alcanzar los mejores resultados en la resolución de ambos problemas.
5. Enfrentar al alumnado a la complejidad real de los dos problemas desde un punto de vista multidisciplinar.
6. Desarrollar habilidades de exposición ante público especializado como no especializado por medio de una presentación del trabajo realizado.

2.2 Competencias.

Durante la ejecución del PID se desarrollaron las siguientes competencias:

- CTEC7. *Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.*
- CB7. *Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.*
- CG4. *Capacidad generalizada para analizar la información de datos experimentales de forma masiva.*
- CB4. *Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.*

2.3 Materiales y métodos.

Los objetivos y competencias anteriores se llevaron a cabo mediante las actividades que se describen a continuación.

2.3.1 Actividad 1. Instauración del sistema de comunicación y difusión del evento.

Para la difusión del evento y como núcleo central para el material y recursos del evento se creó un sitio web² público con toda la información relativa a las actividades del PID, así como un formulario de inscripción.

Dentro de este sitio web se habilitó una dirección³ con todo el material necesario para el desarrollo de las diferentes actividades, así como información sobre todo el evento. El enlace a esta página se comunicó a todo el alumnado perteneciente a las anteriores asignaturas. Del mismo modo, se dio difusión general entre las escuelas y departamentos de la Universidad, a través de redes sociales y boletines electrónicos. Un total de 20 estudiantes formalizaron su inscripción en el PID.

² <https://biodatauco.github.io/>

³ https://biodatauco.github.io/informacion_participantes

2.3.2. Actividad 2. Realización de guías formativas. Presentación de los retos.

Con el objetivo de generar unos conocimientos básicos en un área diferente al de la formación universitaria propia de cada participante, se elaboraron dos “Guías de Ciencia de Datos”, una por cada lenguaje de programación usado en esta disciplina, *Python*⁴ y *R*⁵. De acuerdo con los contenidos de sus titulaciones, estos son los lenguajes en los que el alumnado presenta mayor fluidez. Las guías a modo de tutorial se desarrollaron haciendo uso de la plataforma *Google Colaboratory*⁶, de forma que se facilitara el uso de las herramientas por parte del alumnado sin necesidad de instalar ninguna aplicación en sus ordenadores personales. En estas guías, se presentaron las nociones básicas de carga de datos, preprocesamiento y modelado que usarían durante el desarrollo de las siguientes actividades para resolver los problemas siguientes:

- **Visualización de información y extracción de características moleculares.** Este problema, asociado al área de Biotecnología y Genética, consiste en analizar, desde un punto gráfico, las características principales y las diferentes categorías de ontología génica en el proceso de la anotación funcional de genes.
- **Obtención de modelos de Aprendizaje Automático y Minería de Datos.** Este problema, relacionado con el área de la Ingeniería Agroforestal, consiste en realizar un análisis predictivo y estudiar la biodiversidad forestal de Andalucía.

Por cada uno de los dos retos anteriores, se le proporcionó al alumnado una segunda guía con información teórica más relacionada con la temática de cada reto, así como información de las bases de datos sobre las que se trabajaría. Adicionalmente, se proporcionaron vídeos explicativos con los contenidos teóricos de cada reto.

Tras las sesiones informativas, se propuso al alumnado una dinámica colaborativa y competitiva para resolver ambos retos. Previo al desarrollo de las sesiones se establecieron cinco grupos de trabajo, cada uno con cuatro estudiantes pertenecientes a estudios de las diversas titulaciones consideradas y con conocimientos complementarios en las diferentes áreas de trabajo. De los grupos establecidos inicialmente, cuatro llegaron a completar todas las actividades docentes.

2.3.3 Actividad 3. Acceso, tratamiento y preprocesamiento de bases de datos con información biológica.

Esta actividad da comienzo al primero de los dos retos principales a los que se enfrentó el alumnado. Esta actividad se realizó en una sesión presencial remota a través de una de las plataformas de las que dispone la Universidad de Córdoba. Tuvo una duración estimada de 4 horas. El principal objetivo fue familiarizarse con una de las fases previas que mayor importancia tiene en el Análisis de Datos, el preprocesamiento. La primera tarea que realizar cuando nos enfrentamos a

⁴ <https://www.python.org>

⁵ <https://www.r-project.org/>

⁶ <https://research.google.com/colaboratory/>

un problema de Ciencia de Datos es acceder a fuentes de información fiables y analizar los posibles errores a los que nos podemos enfrentar.

Para este primer reto, se seleccionó un tema relacionado con el área de la Biotecnología y Genética. El problema se centró en la anotación funcional de la señalización molecular por estrés en *Arabidopsis* (Berardini et al., 2004). Las poliaminas son compuestos nitrogenados presentes en las plantas que se acumulan en respuestas al estrés. La acumulación de poliaminas específicas en plantas, produce la activación de las defensas, posiblemente a partir de la activación de la ruta metabólica del ácido salicílico. El punto de partida del problema fue el estudio de un trabajo reciente por parte de investigadores de las Universidades de Barcelona y Düsseldorf, donde se demuestra expresión diferencial en la transcripción de las defensas vegetales desencadenada a partir de la acción de las poliaminas (Liu et al., 2020). El objetivo de esta actividad fue doble: por el lado biológico, se pretendía determinar si existían diferencias cuantitativas/cualitativas en la activación transcripcional de plantas tratadas con putrescina y termoespermina; por el lado computacional, se requería generar herramientas automáticas de análisis de la anotación funcional basadas en el vocabulario estándar de ontología génica (Ashburner et al., 2000).

Esta actividad se evaluó principalmente con la documentación del código realizado para su resolución, así como un análisis de las ventajas e inconvenientes presentes en cada una de las técnicas computacionales y de los métodos desarrollados.

2.3.4 Actividad 4. Visualización de los datos y extracción de características estadísticas.

Una de las principales técnicas para el análisis inicial de los datos es la visualización de los mismos por medio de diversos gráficos. Son numerosas las técnicas existentes en esta área, las cuales han generado un enorme interés por las numerosas ventajas que proporcionan, sirviendo como fase inicial a la hora de analizar un problema concreto. Entre otras técnicas, se le mostró al alumnado cómo realizar un análisis visual de la distribución de los datos, conociendo si existe algún error en ellos o qué tipo de distribución está presente en los datos. Por otro lado, la visualización también es de enorme interés para el análisis de los resultados, de forma que nos permita analizar los resultados de los diversos modelos o cómo ha sido la predicción de cada modelo y su ajuste con la realidad. En este sentido, se les presentó una propuesta para abordar el reto, consistente en dos secciones, una descriptiva, donde se podría emplear un análisis de componentes principales (Jolliffe, 2005) o análisis *cluster* (Duran y Odell, 2013), y otra funcional, donde se realizaría la anotación funcional propiamente dicha.

Esta actividad se evaluó con la interpretación biológica específica asociada al problema presentado. Además, se valoró positivamente el análisis y las conclusiones obtenidas de todo el proceso.

2.3.5 Actividad 5. Generación y evaluación de modelos de Aprendizaje Automático y Minería de Datos.

Esta actividad dio comienzo al segundo reto presentado al alumnado. De igual forma que con el primer reto, también se realizó una sesión presencial remota

con duración estimada de 4 horas. La temática de este segundo problema estuvo más centrada en descubrir los patrones ocultos de la diversidad vegetal de Andalucía. Para ello, se le proporcionó al alumnado una base de datos con los inventarios de vegetación en zonas naturales de Andalucía, así como del número de especies de plantas leñosas que existen (CMA, 2016). Esta base de datos es única en Andalucía, conteniendo aproximadamente 50000 inventarios de vegetación para describir toda la variabilidad ambiental de Andalucía. Para cada uno de estos inventarios, se identificaron todas las especies de plantas presentes y se registraron numerosas variables relacionadas con clima, topografía y suelo, las cuales, generalmente, explican los patrones de diversidad en Andalucía.

El objetivo de esta actividad fue también dual, por el lado de Ingeniería Forestal, se pretendía realizar un análisis de la riqueza de plantas leñosas existentes en Andalucía. Para ello, debieron analizar las variables más relevantes. Por el lado de la Ingeniería Informática, se pedía obtener un modelo predictivo (Phyu, 2009) de diversidad vegetal con alta precisión.

Esta actividad se evaluó tanto con la documentación del código realizado para la generación y evaluación de los diferentes modelos de Aprendizaje Automático, así como con la motivación para elegir dichos modelos, profundizando en las ventajas y desventajas de los mismos.

2.3.6 Actividad 6. Interpretación y justificación de resultados.

En la segunda fase del reto relacionado con el análisis predictivo, se animó al alumnado a interpretar los resultados obtenidos en la fase de evaluación de los modelos. La idea principal fue, dado un modelo y una predicción obtenida, entender el funcionamiento detrás de la metodología, así como interpretar los resultados y generar conocimiento a partir de ellos. Para esta fase, se requirió, tanto realizar un análisis visual de los resultados, similar al que fue desarrollado por la Actividad 4, como el análisis de las variables introducidas en los modelos, ya que no todas las combinaciones proporcionan buenos resultados.

Esta actividad se evaluó según la justificación correcta de la selección de las variables a utilizar en el modelo (tenían que ser explicativas, relevantes y con sentido ecológico), por la interpretación del ajuste a la realidad ecológica de Andalucía realizado y finalmente, por el grado de ajuste a los datos de validación. En este último caso se trata de generar el mejor modelo predictivo, es decir, el que predice mejor unos datos de diversidad no conocidos.

2.3.7 Actividad 7. Presentación y exposición de las conclusiones obtenidas a partir de la realización del PID. Entrega de premios. Evaluación.

Esta actividad consistió en la presentación y exposición de los trabajos realizados para resolver los dos retos presentados a los grupos. Esta actividad se realizó también en una sesión presencial remota, donde los grupos tuvieron la oportunidad de mostrar al resto de los participantes y al profesorado el trabajo realizado mediante una breve presentación.

Se solicitó que la presentación estuviera dividida en dos fases, una por cada reto que hayan resuelto de forma que el alumnado pueda sintetizar la información, resumiendo el procedimiento escogido para resolver cada reto, así

como las principales decisiones tomadas por el grupo.

Para esta presentación se valoraron las aportaciones de visualización y modelado de datos, así como de análisis de los resultados. Tras cada presentación, tanto el profesorado como el alumnado realizaron preguntas dirigidas al grupo y relacionadas con su trabajo para valorar el nivel de comprensión sobre el mismo.

2.3.8 Actividad 8. Evaluación del PID.

Los resultados del PID se evaluaron de forma descriptiva. En primer lugar, con el objetivo de reflejar la realidad del alumnado, se hicieron cuestionarios de autoevaluación de conocimientos antes y después de las actividades. Para ello, se utilizó la herramienta de Google Forms⁷. El primero de ellos se realizó de forma previa al comienzo del PID, con la finalidad de poder conocer los conocimientos de los que disponía el alumnado en las tres áreas de conocimiento. Mientras que el segundo se hizo de forma posterior a la realización del PID, con el objetivo de comprobar los conocimientos desarrollados por el alumnado durante el mismo. Se realizaron un total de 15 preguntas (5 por área) en las que el alumnado fue preguntado por conocimientos básicos.

Adicionalmente, se realizó otro análisis descriptivo, también por medio de Google Forms, de los resultados, poniendo especial hincapié en los objetivos cumplidos y mejoras futuras. En este cuestionario, se pidió al alumnado una autoevaluación de sus conocimientos sobre las tres disciplinas (4 preguntas sobre Ciencia de Datos, 4 sobre Genética y 3 sobre Agroforestal), antes y después de la actividad docente. La finalidad de realizar esta autoevaluación fue ver con qué grado el alumnado es consciente de los conocimientos de los que disponen desde su percepción. En estos cuestionarios no había respuestas correctas o erróneas, simplemente evalúan sus conocimientos desde su perspectiva.

Tanto los cuestionarios como las preguntas de autoevaluación contribuyeron a verificar que los objetivos 1, 2 y 4, se desarrollaron con el grado de satisfacción esperado. De forma similar, la competición que se organizó entre los diferentes grupos fue clave para llevar a cabo el objetivo 3.

El profesorado también evaluó de forma cualitativa la realización del PID, de una forma más transversal y grupal de todas las actividades realizadas.

Al final de la actividad, se realizó una encuesta más informal mediante la herramienta *Mentimeter*⁸, con la finalidad de reflejar las sensaciones que tuvo el alumnado durante su participación en el *hackathon*. Las sensaciones del alumnado contribuyeron a evaluar el cumplimiento del objetivo 5.

Finalmente, las exposiciones descritas en la Actividad 7, también fueron de gran utilidad para verificar que el objetivo 6 de este PID fuera llevado a cabo de manera satisfactoria.

⁷ <https://www.google.es/intl/es/forms/about/>

⁸ <https://www.mentimeter.com/>

3. RESULTADOS

3.1 Cuestionarios de evaluación de conocimientos.

Con el primer cuestionario hemos podido conocer el punto de partida y cuál era el área donde más hincapié había que realizar. Obviamente, el alumnado de una disciplina podría no tener conocimientos previos sobre las otras disciplinas, por lo que resultó interesante aplicar el cuestionario para ver el nivel de conocimientos en ellas. Los resultados de la encuesta de nivel que se realizó para un total de 14 estudiantes antes de la celebración de la actividad muestran que la puntuación media obtenida fue de 10,57 puntos sobre 15 y las puntuaciones individuales oscilaron entre los 8 y los 14 puntos (Figura 1).

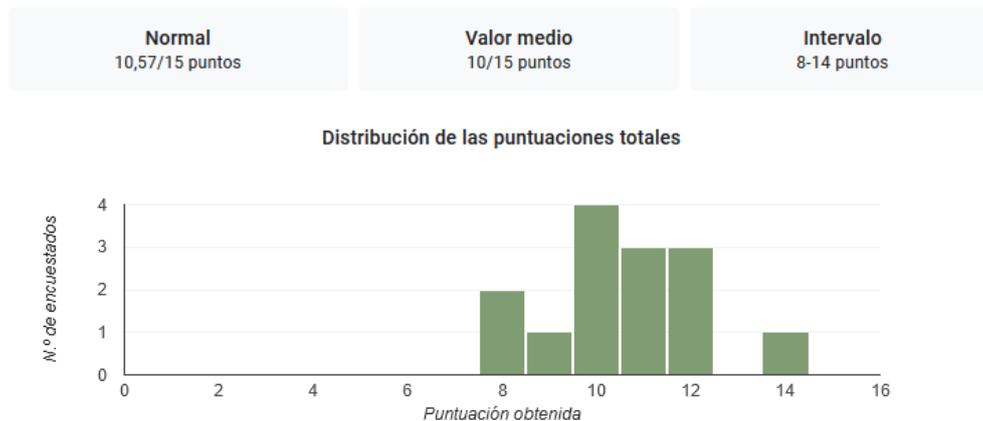


Figura 1. Resultados del cuestionario de conocimiento previo.

De igual forma, el mismo cuestionario se pasó al finalizar todas las actividades del PID, sin haber desvelado las respuestas a ninguna de las preguntas. La principal finalidad de repetir las mismas preguntas fue ver si el alumnado había afianzado los conceptos de los que ya disponían y mejorado aquellos relativos a las otras disciplinas. En este caso, el profesorado sí que esperábamos que los resultados del cuestionario mejorasen. En concreto, la puntuación media ascendió hasta 11,93 puntos sobre 15 mientras que el intervalo de puntuaciones individuales también se ha ampliado a valores entre los 8 y los 15 puntos (Figura 2). Además, se puede observar en la gráfica que la mayor parte de las puntuaciones se encuentran concentradas en torno a los 12 y los 15 puntos. Por tanto, se puede percibir una mejora notable con respecto a los resultados que se obtuvieron en el cuestionario inicial.

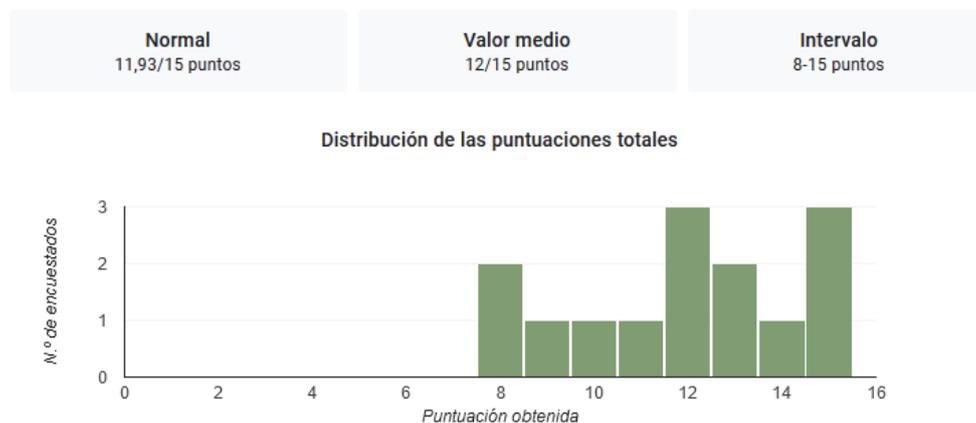


Figura 2. Resultados del cuestionario final.

3.2 Autoevaluación.

Los cuestionarios de autoevaluación también se pasaron antes y después de la realización del PID. La Figura 3 muestra los resultados de los formularios de autoevaluación antes de la realización.

El alumnado valoró cada uno de los aspectos comentados anteriormente con un valor del 1 al 4, siendo 1 el nivel más bajo de conocimiento y 4, el más alto. Los resultados iniciales reflejan que la mayor parte del alumnado indicó que no tenía mucho dominio sobre los temas del *hackathon* (Figura 3). Para la mayoría de las preguntas, los resultados se concentraron en la opción 1. Por otro lado, el número de respuestas en la opción 4 es mínimo en todas las preguntas.

Al finalizar todas las actividades del PID, se volvió a realizar el cuestionario de autoevaluación, con el objetivo de conocer si el alumnado era consciente de haber ampliado sus conocimientos. De igual forma, las preguntas fueron las mismas que en la autoevaluación previa.

Los resultados de la autoevaluación final fueron en promedio más altos (Figura 4). En el caso anterior, la mayoría de valores se encontraban concentrados en el 1, mientras que en este caso se encontraron más repartidos, apareciendo bastantes valores en el nivel 4, lo que indica que parte del alumnado tenía la sensación de haber obtenido nuevo conocimiento durante la realización de la actividad. Por lo general, la mayoría de los valores se encontraron concentrados en los niveles 2 y 3, indicando que la mayor parte del alumnado no se considera experto en todas las materias después de la realización de la actividad (lo cual es normal debido a la duración de la misma) pero sí reconocen tener un cierto grado de conocimiento en todos los temas trabajados. De este modo, se considera que los resultados son satisfactorios.

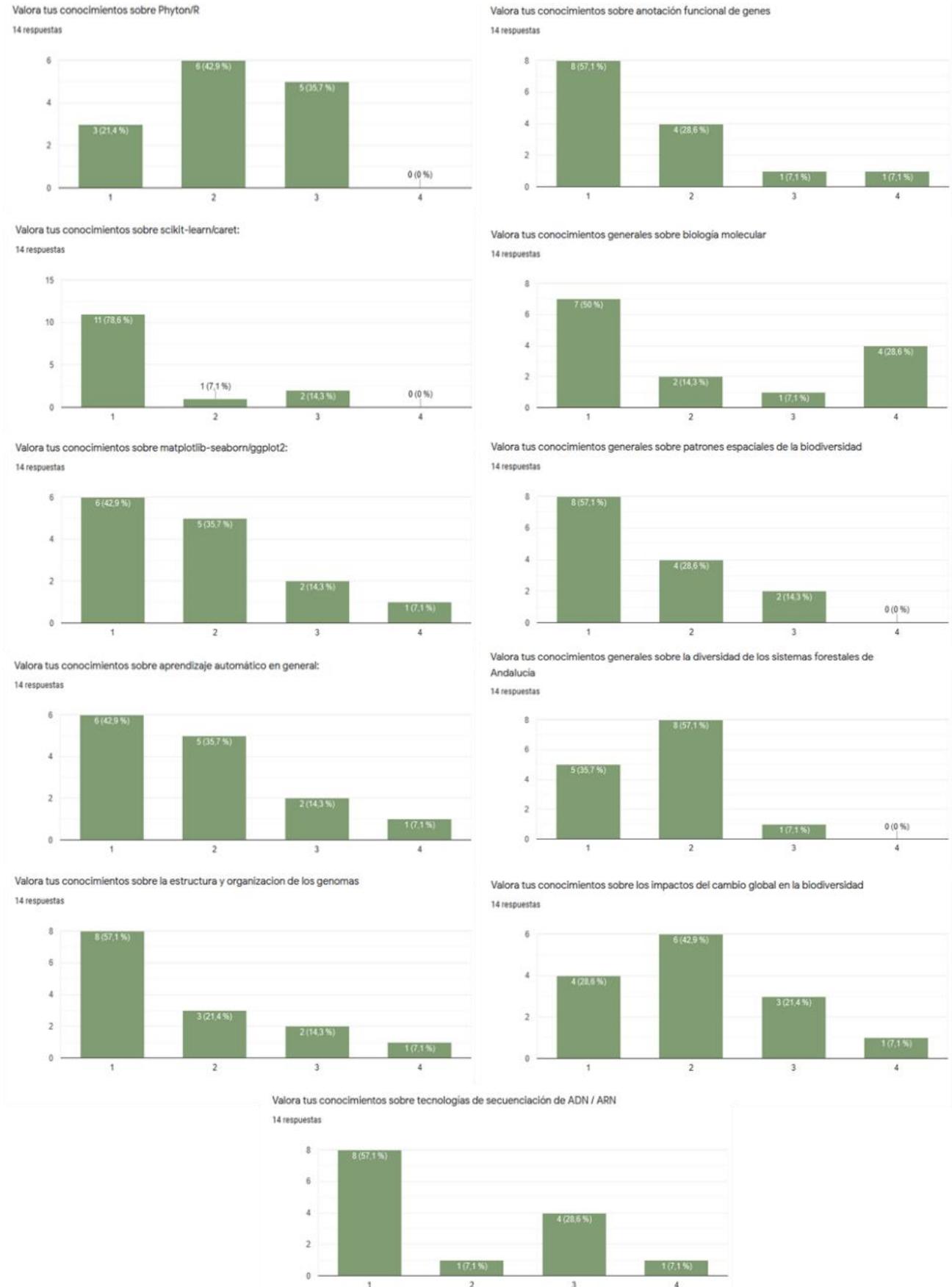
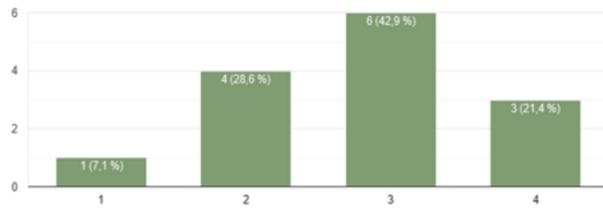


Figura 3. Resultados de la autoevaluación inicial.

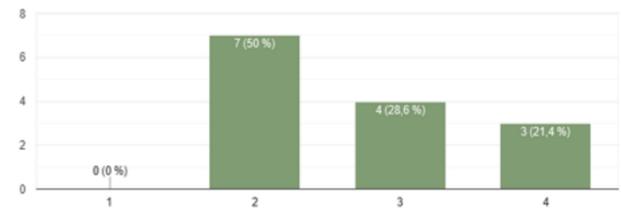
Valora tus conocimientos sobre Phytion/R

14 respuestas



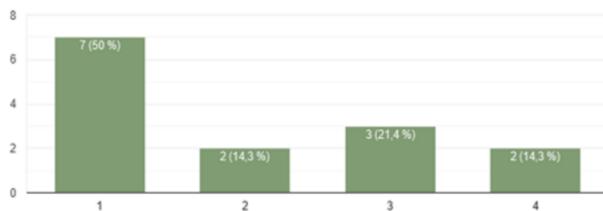
Valora tus conocimientos sobre anotación funcional de genes

14 respuestas



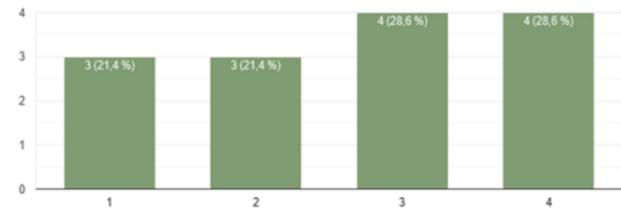
Valora tus conocimientos sobre scikit-learn/caret:

14 respuestas



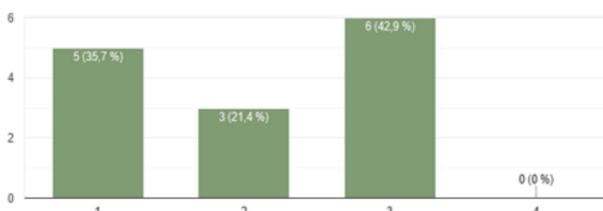
Valora tus conocimientos generales sobre biología molecular

14 respuestas



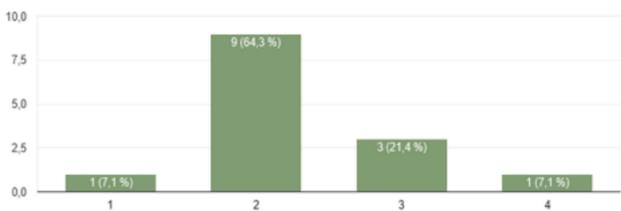
Valora tus conocimientos sobre matplotlib-seaborn/ggplot2:

14 respuestas



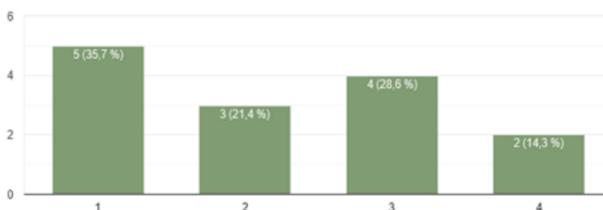
Valora tus conocimientos generales sobre patrones espaciales de la biodiversidad

14 respuestas



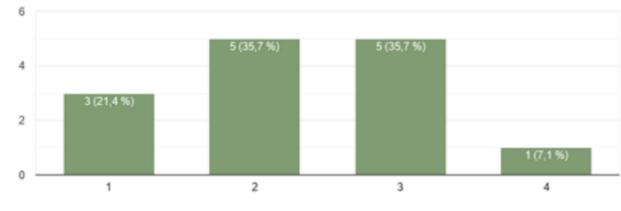
Valora tus conocimientos sobre aprendizaje automático en general:

14 respuestas



Valora tus conocimientos generales sobre la diversidad de los sistemas forestales de Andalucía

14 respuestas



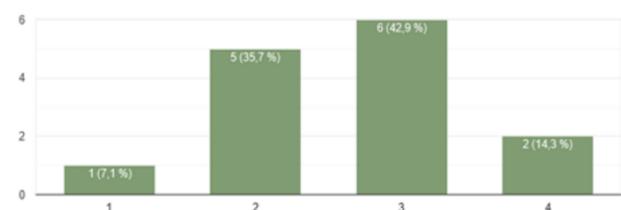
Valora tus conocimientos sobre la estructura y organización de los genomas

14 respuestas



Valora tus conocimientos sobre los impactos del cambio global en la biodiversidad

14 respuestas



Valora tus conocimientos sobre tecnologías de secuenciación de ADN / ARN

14 respuestas

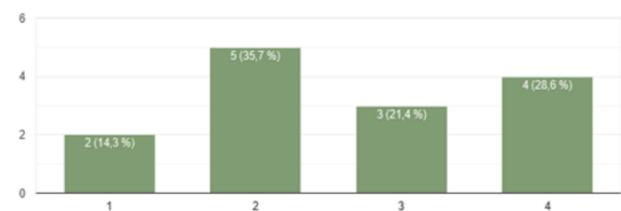


Figura 4. Resultados de la autoevaluación final.

3.3 Evaluación final.

3.3.1 Presentación de los resultados obtenidos por los grupos de trabajo.

Desde una perspectiva cualitativa, el profesorado también ha realizado una evaluación del PID a través de exposiciones de los grupos. En estas exposiciones, el alumnado, dividido por grupos, realizó una presentación de las soluciones propuestas a cada una de las dos tareas que se llevaron a cabo. Por un lado, una presentación más gráfica de las visualizaciones realizadas en el problema de Biotecnología, mientras que, por otro lado, realizaron una presentación más técnica sobre el código empleado para resolver el problema de predicción de la variedad forestal. Estas presentaciones nos permitieron ver con detalle el trabajo que había realizado cada grupo, la forma en que habían analizado el problema y el método que habían seguido para resolverlo. Además, fueron de gran interés debido a que permitió al resto del alumnado ver diferentes soluciones para el mismo problema y detectar elementos que habían pasado por alto durante su trabajo. Por último, también nos sirvieron al profesorado para elegir a los ganadores de la competición docente, no solo en base a los resultados finales que habían obtenido, sino al trabajo global realizado, desde el análisis del problema hasta la aportación de una solución final completa.

3.3.2 Análisis cualitativo sobre sensaciones del alumnado.

Las sensaciones del alumnado se recogieron con tres preguntas diferentes: una primera, para ver qué aspectos eran los que más le habían gustado del *hackathon* (Figura 5); una segunda para que pudieran valorar ciertos aspectos personales como su nivel de participación o la mejora que han visto en sus conocimientos gracias a la realización de la actividad (Figura 6); y, por último, en una tercera pregunta se les preguntó qué aspectos mejorarían del *hackathon*, de cara a tenerlos en consideración para futuras ediciones (Figura 7). Un total de 13 estudiantes participaron en este cuestionario final.

¿Qué es lo que más te ha gustado del hackathon?



Figura 5. Mentimeter: ¿Qué es lo que más te ha gustado del hackathon? Palabras con mayor tamaño fueron más repetidas por el alumnado en el cuestionario.

Los aspectos positivos en los que mayor parte del alumnado ha coincidido son la colaboración y el aprendizaje (Figura 5). Esto es bastante satisfactorio ya que esta respuesta refleja dos de los objetivos que más queríamos reforzar en el PID. Por otro lado, se observa que al alumnado le han gustado los retos propuestos y el hecho de haber realizado un trabajo multidisciplinar con grupos formados por estudiantes de diversas áreas.

Scales



Figura 6. Mentimeter: Autoevaluación de ciertos aspectos personales como su nivel de participación o la mejora que han visto en sus conocimientos gracias a la realización de la actividad.

La Figura 6 muestra los valores medios de las autoevaluaciones que ha hecho cada participante sobre su nivel de participación en el grupo y la mejora de sus conocimientos gracias a la realización de esta actividad. Como se observa, el nivel de participación de los participantes dentro de su grupo es muy alto, por lo que podemos confirmar que la forma en que se crearon los grupos y el tamaño de los mismos fue apropiado y correcto. Por otro lado, el alumnado considera que ha mejorado su conocimiento en análisis de datos, genética y estudios forestales.

¿Qué aspectos mejorarías?



Figura 7. Mentimeter: ¿Qué aspectos mejorarías? Palabras con mayor tamaño fueron más repetidas por el alumnado en el cuestionario.

Por último, la Figura 7 muestra los aspectos que se podrían mejorar según los participantes en el *hackathon*. El principal aspecto que se señala es la falta de tiempo para resolver los retos. En un futuro, se intentará ocupar menos tiempo en la introducción de los retos de manera que los grupos tengan más tiempo para trabajar. Otro aspecto que ha surgido es la preparación, por lo que se explorarán maneras alternativas de preparar a los participantes para los retos, quizá con tiempo adicional. Por último, también se hace referencia al hecho de que se celebrara de forma *online* en lugar de hacerlo presencial, lo cual se debió exclusivamente a la situación sanitaria generada por la COVID-19.

4. DISCUSIÓN

Los cursos de Ciencia de Datos tradicionalmente se han incorporado en los programas docentes de carreras de ciencias de la computación (e.g. Anderson et al 2014). Sin embargo, a medida que otras disciplinas y perfiles profesionales han ido incorporando la programación y modelización como competencias necesarias, es necesario ofrecer programas docentes para un perfil más variado. En este caso de estudio, mostramos los resultados y lecciones aprendidas de un curso de ciencia de datos orientado a perfiles interdisciplinarios de estudiantes incluyendo ciencias de la computación y ciencias de la vida. Como aspecto novedoso incorporamos como herramienta docente el *hackathon*, actividad basada en proyectos que se ha aplicado en cursos similares pero en otros contextos temáticos (Kross, & Guo, 2019; Pournaras 2017). En general, los resultados del proyecto de innovación docente han sido satisfactorios siguiendo la tendencia de experiencias similares (e.g. Chua et al. 2019; Kuter, & Wędrychowicz 2021).

4.1 Evaluación del PID

En primer lugar, se presentan los objetivos del PID que se consideran cumplidos en vista de los resultados obtenidos y descritos en la Sección 3:

- Se ha mejorado la capacidad del alumnado para acometer problemas reales mediante técnicas de Aprendizaje Automático, incluyendo un análisis inicial de los datos de los que disponían, la visualización de los mismos, el preprocesamiento necesario para poder trabajar con ellos y la generación de modelos que permitieron realizar predicciones en base a los datos de los que se dispone.
- Se han adquirido conceptos teóricos y prácticos relacionados con la genética y la diversidad forestal mediante la resolución de problemas relacionados con estas temáticas. Para poder acometerlos, ha sido necesario conocimiento experto sobre la temática del problema.
- Se han incrementado los conocimientos acerca de análisis de datos y lenguajes de programación mediante las guías y vídeos aportados por el profesorado de la actividad.
- Se ha conseguido fomentar la cooperación entre personas de diferentes grupos y disciplinas, a la vez que se ha fomentado la competitividad sana entre grupos para poder conseguir los mejores resultados. Otros casos de estudio identificaron como problemático mantener la motivación y nivel docente a estudiantes con diferente nivel de programación y perfil temático (Pournaras, 2017). En contraposición, hemos comprobado que

las competiciones con fines docentes consiguen motivar al alumnado, en este caso, para intentar buscar la mejor solución a los problemas propuestos.

- Se ha conseguido enfrentar al alumnado a problemas reales con la dificultad que conlleva y, además, se ha hecho en un tiempo limitado, de modo que la cooperación multidisciplinar dentro de cada grupo ha sido imprescindible para poder resolver los problemas planteados. Otros trabajos, también destacan la relevancia de resolver problemas reales para mantener la motivación de los estudiantes (Chua et al., 2019).
- El alumnado ha desarrollado sus capacidades para explicar en público los resultados de su trabajo, de manera que los estudiantes de las diferentes áreas puedan entender los resultados a los que han llegado y la forma en que lo han conseguido. Además, también han podido explicar las dificultades que se han encontrado a la hora de llegar a una solución y la forma en que han podido resolverlas.

Por otro lado, se describen los objetivos que consideramos que no se han logrado durante la realización de este PID:

- Los grupos se diseñaron buscando la presencia de, al menos, un perfil de cada disciplina. Sin embargo, en algún caso, no todos los participantes pudieron asistir a las sesiones.
- En algunos casos los retos planteados no se resolvieron en su totalidad, principalmente por la falta de tiempo. En cierto modo, esto era algo que se esperaba debido a que eran retos de cierta complejidad y el tiempo que se daba para resolverlos era bastante limitado.

4.2 Fortalezas y debilidades del PID.

Tras un análisis de los resultados, identificamos las fortalezas y debilidades observadas del PID que pueden considerarse lecciones aprendidas para futuros cursos en Ciencia de Datos interdisciplinarios. Como fortalezas destacamos los siguientes aspectos:

- La actividad realizada ha sido de gran interés para el alumnado participante por dos motivos: a) el alumnado que estaba algo familiarizado con la Ciencia de Datos ha tenido la oportunidad de trabajar sobre problemas reales de cierto interés y complejidad y b) los participantes pertenecientes a las disciplinas relacionadas con los problemas planteados han aprendido técnicas de Ciencia de Datos para trabajar con esos datos, analizarlos y obtener modelos a partir de ellos.
- El equipo de trabajo del PID está formado por personas de diferentes disciplinas, de forma que ha resultado más fácil trabajar en una actividad multidisciplinar.
- El alumnado de la actividad ha tenido una motivación extra gracias a la competición con fines docentes que se ha llevado a cabo y que los animaba a estudiar el problema con mayor profundidad para poder encontrar la mejor solución posible.

Por otro lado, se consideran las siguientes debilidades:

- Dificultad de coordinación y trabajo en grupo debido a la realización por videoconferencia.
- Dificultad para fijar el nivel de conocimiento inicial del que se debe partir debido a la pertenencia de los participantes a diferentes disciplinas.

4. CONCLUSIONES

Una vez analizados los resultados obtenidos tanto por medio de los formularios como de la percepción del profesorado, creemos que experiencias como la promovida por el Proyecto de Innovación Docente (PID) fomentan el interés del alumnado en la búsqueda de nuevos conocimientos. En relación con la participación, consideramos que ha sido altamente satisfactoria, no solo por el volumen del alumnado que decidió participar en una actividad totalmente voluntaria y sin beneficio directo en las asignaturas que estaban cursando, sino por el carácter complementario a su formación curricular. En este caso, la experiencia ha sido interdisciplinar, al contar con alumnado perteneciente a áreas diferentes, donde cada persona aportaba conocimientos y experiencia en tareas distintas pero complementarias.

Es remarcable que la utilización de técnicas y herramientas de actualidad genera más interés en la participación. En este caso, se utilizó la herramienta del *hackathon*, muy popular en Ciencia de Datos, lo que nos permitió motivar el aprendizaje y reciclaje continuo de conocimientos a los que debe someterse hoy día el alumnado. El *hackathon* consistía en trabajar en dos retos diferentes, donde una de las principales herramientas han sido las competiciones grupales con fines docentes, con las que se ha podido trabajar tanto la cooperación entre estudiantes de un mismo grupo con titulaciones diferentes (la colaboración ha sido vital para alcanzar un buen resultado en ambos retos), como la competitividad sana entre los grupos (ha resultado de gran importancia para generar esa competencia entre grupos por alcanzar la mejor solución).

Al tratarse, de una actividad con connotación investigadora, varios participantes se han interesado por esta novedosa área, ya sea en la realización de Trabajos Fin de Grado, como en la solicitud de Proyectos de Investigación derivados de las titulaciones en Ciencias de la Vida, donde el interés por la Ciencia de Datos ha aumentado significativamente en los últimos años.

El profesorado asociado al PID ha realizado un esfuerzo para elaborar guías y vídeos explicativos, página web para difusión de la actividad y comunicación con los participantes, organizar herramientas para el desarrollo de la competición, y emplear distintos instrumentos para la evaluación final del PID. Dado que el balance del PID realizado es completamente positivo, podemos destacar que el PID abre diferentes líneas de trabajo, motivadas por los resultados cualitativos de cómo las dinámicas de competición y el trabajo en grupo interdisciplinar ayudan a mejorar los conocimientos del alumnado sobre diferentes áreas de conocimiento. Dichas líneas consisten en realizar actividades docentes de carácter interdisciplinar, utilizar dinámicas de competición como actividades docentes y emplear técnicas y herramientas novedosas para motivar al alumnado, como ha sido en este caso el *hackathon*.

AGRADECIMIENTOS

El Proyecto de Innovación Docente ha sido financiado por la Universidad de Córdoba con referencia 2020-2-5003 y título “*Hackathon en docencia: Aprendizaje Automático aplicado a Ciencias de la Vida y la Naturaleza*”. La docencia de David Guijo-Rubio está financiada por la Universidad de Córdoba con cargo a subvenciones a Universidades Públicas para la recualificación del sistema universitario español del Ministerio de Universidades financiado por la Unión Europea - NextGenerationEU. (Ref. UCOR01MS). La docencia de Víctor M. Vargas-Yun está financiada por el Programa Predoctoral de Formación al Profesorado Universitario (FPU) del Ministerio de Ciencia, Innovación y Universidades (Ref. FPU18/00358). La docencia de Javier Barbero-Gómez está financiada por el Programa Predoctoral de Formación de Personal Investigador (FPI) del Ministerio de Ciencia, Innovación y Universidades (Ref. PRE2018-085659). La docencia de José V. Die está financiada por el programa H2020-MSCA-IF-2018 (Grant agreement ID: 844431). La docencia de Pablo González-Moreno está financiada por un contrato Juan de la Cierva Incorporación (IJCI-2017-31733).

REFERENCIAS

- Anderson P., McGuffee J., & Uminsky D. (2014). Data science as an undergraduate degree. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*. (pp. 705–706). ACM. <https://bit.ly/3CcFt5c>
- Asamoah, D., Doran, D., & Schiller, S. (2015). *Teaching the foundations of data science: An interdisciplinary approach*. <https://bit.ly/3zti10m>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25, 25-29. <https://bit.ly/3sdol2Y>
- Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., ... & Rhee, S. Y. (2004). Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiology*, 135, 745-755. <https://bit.ly/3ohgnue>
- Brunner, R. J., & Kim, E. J. (2016). Teaching data science. *Procedia Computer Science*, 80, 1947-1956. <https://bit.ly/3AIDSGY>
- Bureau, E. T. (2021). *11.5 mn job openings by 2026, sky-high salaries: Why data science is booming*. [Online]. <https://bit.ly/3tsDVwZ>
- Chua, H. X., Chua, E. L., & Soo, K. (2019). Budding Data Scientists Hackathon. *ACM SIGKDD Explorations Newsletter*, 21(1), 38-40. <https://bit.ly/3SdJPP3>
- CMA (2016). *Cartografía de vegetación de la masa forestal de Andalucía (VEGE10_Memorias)*. Junta de Andalucía. <https://bit.ly/3L0IUhg>
- Duran, B. S., & Odell, P. L. (2013). Cluster analysis: a survey. *Springer Science & Business Media*, 100. <https://bit.ly/3fGIS0K>
- Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American Statistician*, 72(4), 382-391. <https://bit.ly/3nMYpiG>

- Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*. <https://bit.ly/3tLvrn7>
- Kross, S., & Guo, P. J. (2019). Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges. In *Proceedings of the 2019 Conference on Human Factors in Computing Systems* (pp. 1-14). <https://bit.ly/3rEhpRz>
- Kuter, K., & Wędrychowicz, C. (2021). Hosting a data science hackathon with limited resources. *Stat*, 10(1), e338. <https://bit.ly/3SkETHV>
- Liu, C., Atanasov, K. E., Arafaty, N., Murillo, E., Tiburcio, A. F., Zeier, J., & Alcázar, R. (2020). Putrescine elicits ROS-dependent activation of the salicylic acid pathway in *Arabidopsis thaliana*. *Plant, Cell & Environment*, 43 (11), 2755-2768. <https://bit.ly/3fJXbjX>
- Markowitz, F. (2017). All biology is computational biology. *PLoS Biology*, 15(3), 1-4. <https://bit.ly/3fK6h01>
- National Academies of Sciences, Engineering, and Medicine. (2018). Data science for undergraduates: Opportunities and options. *National Academies Press*. <https://bit.ly/3qJcYWt>
- Organización de Naciones Unidas. (2022). *Objetivos de Desarrollo Sostenible*. <https://bit.ly/2qk9f28>
- Perkel, J. M. (2021). Five reasons why researchers should learn to love the command line. *Nature*, 590(7844), 173-174. <https://bit.ly/3FI2j2J>
- Phyu, T. N. (2009). Survey of classification techniques in data mining. In *Proceedings of the international multiconference of engineers and computer scientists*, 1 (5). <https://bit.ly/3lqdr68>
- Pournaras, E. (2017). Cross-disciplinary higher education of data science—beyond the computer science student. *Data Science*, 1(1-2), 101-117. <https://bit.ly/3PTQpSP>
- Ramamurthy, B. (2016). A practical and sustainable model for learning and teaching data science. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 169-174). <https://bit.ly/3qKXscO>