

Formación de Científicos de Datos mediante Aprendizaje Basado en Proyectos

Fernando Martínez-Plumed and José Hernández-Orallo

Abstract—Los conceptos de innovación, creatividad, resolución de problemas, comunicación efectiva, autonomía y pensamiento crítico son fundamentales para convertirse en un buen científico de datos. Adaptarse a los nuevos recursos y herramientas tecnológicas también es una habilidad importante, que se basa en la naturaleza curiosa e inquisitiva asociada con la ciencia de datos, y se alimenta de los ecosistemas de ciencia de datos en constante cambio en la industria. En este sentido, el aprendizaje basado en proyectos (ABP) tiene claros beneficios para involucrar a los estudiantes en cursos de ciencia de datos. Sin embargo, el carácter exploratorio de los proyectos de ciencia de datos, que no comienzan con una especificación clara de qué hacer, sino con algunos datos para analizar, plantea algunos desafíos para la aplicación del ABP. Nuestro objetivo es mejorar las experiencias y resultados de aprendizaje en ciencia de datos de los estudiantes a través del uso de ABP. En este artículo, compartimos nuestras experiencias con ABP y presentamos una rúbrica de evaluación que se enfoca en valor, innovación y narrativa, que puede usarse como una estructura de apoyo para los cursos de ciencia de datos. Nuestro análisis de un curso de ciencia de datos ABP a nivel de MSc, junto con los datos de las encuestas de estudiantes, muestra cómo la metodología y la rúbrica se alinean bien con el carácter exploratorio de la ciencia de datos y las habilidades proactivas, curiosas e inquisitivas requeridas por los científicos de datos.

Index Terms—Ciencia de Datos, Aprendizaje Basado en Proyectos, Herramientas de Evaluación.

I. INTRODUCCIÓN

LA búsqueda de nuevas metodologías de enseñanza en la educación superior y el aprendizaje a lo largo de la vida es un tema ampliamente debatido en universidades y otros entornos de formación. La facilidad de acceso a la información (así como las múltiples y variadas formas de obtenerla y contrastarla) ha influido en los perfiles de estudiantes jóvenes y más maduros, en cursos físicos o en línea. Además, a diferencia del conocimiento enciclopédico, hoy en día una persona casi nunca puede dominar todo el conocimiento en un campo muy específico. La generación acelerada de nuevos conocimientos nos insta a equipar los sistemas educativos y de formación con nuevas técnicas de aprendizaje que hagan que todos los actores, incluyendo instructores, estudiantes y empleadores, sean más aptos para este proceso de cambio continuo. Esto motiva un debate abierto en torno a la búsqueda de nuevas metodologías que hagan que los estudiantes aprendan de manera más efectiva, con el objetivo de formar profesionales adaptados a esta nueva sociedad. Los nuevos modelos buscan procesos de aprendizaje que sean más centrados en el estudiante que en el profesor. Para que esto ocurra, las metodologías de enseñanza tienen que cambiar.

Las denominadas metodologías activas [1] juegan un papel preponderante en la consecución de este objetivo.

Una gran metodología activa es el Aprendizaje Basado en Proyectos (ABP), una estrategia de aprendizaje cooperativo que comprende el aprendizaje como un proceso de comunicación y se centra en el aprendizaje tanto a nivel individual como miembro de un grupo. En el ABP, la resolución de un problema, el proyecto, impulsa todo el proceso [2] y la adquisición de habilidades. Los estudiantes son responsables de su propio progreso y los maestros desempeñan el papel de proveer materiales, retroalimentación (o *feedback*) y apoyo cuando se solicita, además de ser asesores para facilitar el trabajo de los estudiantes. Los estudiantes también pueden ver inmediatamente su proyecto como un campo de pruebas constante y eficiente para nuevas ideas. Además, los estudiantes tienen muchas más probabilidades de entender y aplicar conceptos si pueden usar su conocimiento para efectuar cambios en el mundo real.

En informática, el ABP es un elemento atractivo para los estudiantes [3]. Inicialmente introducido en cursos de ingeniería de software, se ha extendido a otros ámbitos [4]–[8]. La peculiaridad de los proyectos de ciencia de datos exige una combinación eficaz de ciencia de datos y ABP. En proyectos exploratorios de ciencia de datos, a diferencia de otros cursos basados en ABP, no hay una especificación inicial; muchos incluso carecen de un objetivo claro. Los *datos* son los actores principales. Debemos explorar qué posibles operaciones pueden liberar y utilizar su valor. La metodología para proyectos de ciencia de datos es menos prescriptiva y más inquisitiva. A diferencia de la *minería de datos*, orientada a objetivos y concentrada en el proceso, la ciencia de datos es dirigida por datos y exploratoria [9].

En minería de datos, un proyecto suele seguir una secuencia de etapas que inicia desde el ‘objetivo empresarial’, llevándolo a un ‘objetivo de minería de datos’ [10]. De aquí, surgen metodologías como CRISP-DM [11] para guiar los pasos habituales. No obstante, en ciencia de datos, el contexto adquiere mayor importancia. Por eso, surgen procesos nuevos que desafían a CRISP-DM, incluyendo adaptación de contexto y reutilización de modelos [12], o propuestas de trayectorias flexibles para los proyectos [9]. Esta visión de la ciencia de datos indica que los proyectos pueden seguir diversos caminos y el orden de las actividades depende del dominio y decisiones del científico de datos. La Figura 1 muestra un espacio de actividades exploratorias, orientadas a objetivos y gestión de datos, que pueden —o no— ser realizadas en proyectos de ciencia de datos con diferentes trayectorias sin un orden preestablecido. Varias veces estos proyectos no surgen de especificaciones claras, sino de la exploración y búsqueda

de novedades en los datos. Este viaje exploratorio culmina en una narrativa de datos que debe ser correctamente presentada y enfatizada. Esto sugiere que la ciencia de datos necesita mayor que pequeños ajustes a la metodología ABP; se requiere una revisión importante de la metodología y rúbricas para evaluar las habilidades del científico de datos [12].

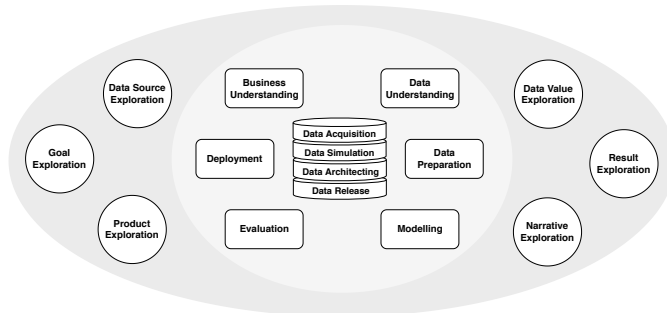


Fig. 1. El mapa de trayectorias de Ciencia de datos (de [9]), que contiene el círculo exterior de actividades exploratorias, el círculo interior de actividades de Minería de Datos (o dirigidas a objetivos), y en el núcleo las actividades de gestión de datos.

A pesar del rápido crecimiento en la necesidad de habilidades de ciencia de datos durante muchos años, lo que ha llevado a un aumento en el número de programas de ciencia de datos [13], se ha escrito poco sobre cómo educar de la mejor manera a los científicos de datos [14]–[17]. En cambio, podemos encontrar diferentes modelos pedagógicos genéricos como la enseñanza invertida [18], el uso de métodos y herramientas para apoyar el aprendizaje [19], o la consideración de esfuerzos colaborativos entre instructores de diferentes disciplinas académicas [20] (intentando aumentar la participación y el compromiso [21], [22] y la creatividad [17]). Menos comúnmente, también encontramos trabajos que analizan la importancia de incorporar un practicum sustancial en el currículo de los cursos de ciencia de datos (ver, por ejemplo, [23], [24]). La principal novedad de este artículo es, por lo tanto, presentar una metodología personalizada basada en ABP para los cursos de ciencia de datos cuyo objetivo principal es motivar a los estudiantes a mejorar habilidades como la innovación y la creatividad, la resolución de problemas, la comunicación efectiva, la autonomía y el pensamiento crítico, promoviendo también el trabajo colaborativo. Para lograr esto, hacemos las siguientes contribuciones:

- Describimos un enfoque de enseñanza replanteado hacia ABP en un curso de ciencia de datos con proyectos del mundo real diferenciados, abordando también varios aspectos de enseñanza para la mejora de el proceso de aprendizaje y su evaluación.
- Traemos recursos didácticos innovadores que los instructores pueden usar para organizar sus cursos, incluyendo una especial rúbrica para evaluar el equilibrio entre innovación-riesgo vs. resultado para el proyecto, la cual es usada por los estudiantes como andamiaje.
- Ilustramos el funcionamiento de la metodología ABP y el uso de las nuevas rúbricas presentadas aquí a través de los resultados de un estudio de caso en un curso de MSC en ciencia de datos.

- Evaluamos el éxito de la metodología de enseñanza a través de un cuestionario con escalas de valoración (tipo Likert) y respuestas abiertas, incluyendo preguntas sobre motivación, objetivos y logro de competencias.

Las encuestas muestran que más del 85% de los estudiantes indican una alta motivación y que tanto los objetivos como las competencias cumplen con sus expectativas.

El resto del artículo está estructurado de la siguiente manera. La Sección II explica las nuevas habilidades que se requiere en los perfiles de ciencia de datos. Las Secciones III y IV presentan un estudio de caso en curso y la adaptación del ABP para un curso de ciencia de datos. La Sección V presenta las herramientas de evaluación basadas en proyectos y las rúbricas. La Sección VI analiza algunos ejemplos ilustrativos. Finalmente, la sección VII discute la evaluación de la metodología y los resultados de los estudiantes, cerrando el artículo en la Sección VIII.

II. APRENDIZAJE BASADO EN PROYECTOS Y CIENCIA DE DATOS

El tradicional rol del profesor, dictando clases magistrales y evaluando al final del curso, ha sido cuestionado, buscándose nuevos modelos desde finales de 1960 [25]. Estos modelos redujeron las conferencias y desarrollaron habilidades creativas de los estudiantes con preguntas y problemas abiertos. El método ABP se empezó a usar a nivel universitario en medicina [26], y después en ingeniería e informática [27], [28]. Enseñar informática resulta propicio para la implementación de actividades y asignaturas orientadas a proyectos [27]. En este esquema, los profesores proponen proyectos, normalmente basados en problemas reales, que los estudiantes deben resolver en grupos. Los estudiantes deciden cómo abordar los proyectos y qué actividades realizar. Este aprendizaje puede ser valioso para fomentar el desarrollo de habilidades genéricas [29]–[31]:

- **Trabajo en equipo:** Trabajar en equipos desarrolla habilidades de coordinación, comunicación, responsabilidad y planificación, etc.
- **Iniciativa:** Se fomenta la motivación de los estudiantes mediante la búsqueda y comprensión de nueva información, utilizando todos los recursos disponibles.
- **Proactividad:** El estudiante no es un receptor de conocimiento, sino un agente activo en su aprendizaje y resolución de problemas.
- **Innovación y creatividad:** No es garantía de éxito la conformidad con la norma. Más bien, se animan las ideas innovadoras que permiten destacar sobre el resto.
- **Pensamiento abstracto:** El ABP facilita la interdisciplinariedad y el pensamiento de orden superior.
- **Evaluación formativa y no punitiva:** El objetivo de la evaluación es que el estudiante aprenda de sus errores, proporcionando una experiencia de aprendizaje más enriquecedora.
- **Pensamiento crítico:** El ABP confronta a los estudiantes con situaciones del mundo real y tienen que comparar pros y contras para cada decisión tomada.

La Ciencia de Datos, con raíces en estadística e informática, es un campo adecuado para adoptar metodologías ABP en

la enseñanza y evaluación. Es un campo interdisciplinario que utiliza métodos científicos y sistemas para entender datos en diferentes formas, extrayendo y aplicando conocimientos en varios dominios. Es la evolución moderna de disciplinas enfocadas en análisis de datos, incluyendo estadística, minería de datos, tecnología de bases de datos, aprendizaje automático y analítica predictiva para entender y analizar fenómenos reales [32].

Durante las últimas décadas, la ciencia de datos se ha vuelto cada vez más popular, llegando a ser parte integral de cada modelo de negocio. Mientras que el área clásica de la minería de datos para obtener valor de los datos ha crecido exponencialmente en tamaño y complejidad, también se ha vuelto mucho más exploratorio bajo el paraguas de la ciencia de datos. En este último, las etapas impulsadas por datos y por conocimientos interactúan, en contraposición al proceso tradicional de minería de datos, que parte de objetivos de negocio precisos que se traducen en una tarea clara de minería de datos, que finalmente convierte "los datos en conocimientos". En otras palabras, no solo ha cambiado la naturaleza de los datos, sino también los procesos para extraer valor de ellos.

Se espera entonces que este nuevo perfil de aplicaciones y proyectos requerirá nuevas habilidades, así como la consolidación de la ciencia de datos como una nueva profesión. Se espera que los científicos de datos cubran una amplia gama de habilidades blandas, como ser proactivos, curiosos e inquisitivos, además de ser capaces de comunicar resultados, liderar un equipo, ser creativos, etc. [33]–[36]. La mayoría de los nuevos pasos exploratorios implican habilidades blandas. Además, la comprensión de nuevos dominios debe jugar un papel interactivo y exploratorio en la mayoría de los proyectos de ciencia de datos [9]. No es de extrañar que el carácter más flexible, menos sistemático, de las nuevas actividades de exploración y gestión de datos (ver Figura 1) destaque los desafíos que caracterizan a la ciencia de datos y allane el camino para seguir metodologías de ABP basadas en la innovación en la formación de científicos de datos.

III. ESTUDIO DE CASO PRÁCTICO

En años académicos anteriores, los autores han adaptado y personalizado la metodología ABP para su curso de ciencia de datos (CDA). Esta adaptación se realizó motivada por la experiencia obtenida de otros cursos similares donde los autores observaron que la aplicación de las metodologías de enseñanza clásicas (por ejemplo, conferencias, prácticas de laboratorio y trabajo independiente de los estudiantes) puede no ser la más adecuada para las competencias de un curso de ciencia de datos. Con el propósito de mejorar el aprendizaje y rendimiento de los estudiantes, aumentando la motivación y participación, los autores prepararon este curso prestando especial atención a las siguientes cuestiones: ¿Aprenderán los estudiantes algo útil? ¿Hasta qué punto aprenderán soluciones originales? Y, lo que es más importante, ¿qué sucederá una vez que se incorporen al mercado laboral y tengan que enfrentarse a problemas reales? Guiados por estas preguntas, diseñamos varios escenarios de evaluación alternativos para evaluar el

rendimiento de nuestros estudiantes en base al desarrollo de proyectos de objetivos abiertos que serían realizados de manera colaborativa (maximizando la utilidad, la pasión, la curiosidad y la competencia), inspirados por la naturaleza basada en proyectos del mundo real de la ciencia de datos.

CDA es una asignatura optativa (impartida en inglés) en el Máster de Informática en la Escuela de Informática y de Ingeniería (ETSInf) de la Universidad Politécnica de Valencia (UPV), España. Es un curso de cuatro meses y se le asignan 6 créditos ECTS¹ distribuidos en 1.5 créditos de teoría en aula, 3 créditos de seminario y 1.5 créditos de práctica de laboratorio.

Como asignatura optativa, desde su creación en el año académico 2016/17, CDA ha tenido alrededor de 30 estudiantes cada año. Los estudiantes provienen de diferentes programas de Informática de la UPV y de otras universidades europeas (el 61% de los estudiantes en el año 2021/22 eran de fuera de la UPV). Este curso se centra en preparar a los estudiantes para el papel de científico de datos en una organización, de modo que puedan identificar problemas y oportunidades relacionados con los datos, y desplegar y comunicar productos basados en datos utilizando herramientas efectivas. Los objetivos principales se pueden resumir de la siguiente manera:

- O1:** Reconocer el valor de los datos y las oportunidades de negocio para el desarrollo de productos basados en datos, en el contexto de Big Data.
- O2:** Determinar las tecnologías que se necesitan para manejar los datos de manera eficiente en diferentes entornos, tamaños y formatos, con el fin de facilitar la comprensión y análisis de los datos.
- O3:** Estimar la complejidad y los recursos que se necesitan para un proyecto de análisis de datos y establecer las medidas de coste y éxito.
- O4:** Transmitir los resultados, implicaciones y valor del análisis, creando representaciones visuales efectivas.

Además, las competencias y habilidades para alcanzar estos objetivos son las siguientes:

- C1:** Poseer y entender conocimientos que proporcionen una base u oportunidad para la originalidad en el desarrollo y/o aplicación de ideas en un contexto de investigación.
- C2:** Aplicar el conocimiento adquirido y la resolución de problemas en entornos nuevos o desconocidos dentro de contextos más amplios y multidisciplinarios, siendo capaz de integrar este conocimiento.
- C3:** Integrar conocimientos y afrontar la complejidad de formular juicios a partir de información incompleta o limitada, incluyendo reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- C4:** Comunicar sus conclusiones, y los conocimientos y razones última que las sostienen, a públicos especializados y no especializados de una manera clara e inequívoca.
- C5:** Poseer habilidades para el aprendizaje que permitan un estudio continuado de manera autodirigida o autónoma.

¹El Sistema de Transferencia y Acumulación de Créditos Europeo (ECTS) es un medio estándar para comparar los créditos académicos de la educación superior en toda la Unión Europea y otros países europeos colaboradores.

C6: Comprender y aplicar la responsabilidad ética, la legislación y la ética profesional de la actividad de la profesión de Informática.

C7: Integrar tecnologías, aplicaciones, servicios y sistemas específicos de Informática, con carácter generalista, y en contextos más amplios y multidisciplinarios.

Este curso resalta la importancia de los datos y el papel del "científico de datos" en contextos reales, utilizando diferentes datos y herramientas de ciencia de datos. Se busca que los estudiantes adquieran conocimientos teóricos para manejar diferentes actividades de un proyecto de ciencia de datos de forma autónoma (como exploración de datos, análisis, limpieza y transformación de datos, modelado, evaluación y producción) y sepan aplicarlos hacia un objetivo común. La idea es que los estudiantes experimenten la realización de un proyecto real: desde la concepción hasta la entrega de un producto, trabajando en equipo. Durante el curso, los objetivos se alcanzarán mediante el trabajo de los estudiantes, desarrollando competencias diversas.

Además, el curso fomenta y desarrolla habilidades genéricas o competencias transversales (TC) [37], útiles en diversas situaciones y beneficiosas para los estudiantes al ingresar al mercado laboral. Nuestro curso de ciencia de datos desarrolla las siguientes habilidades:

TC1: Comunicación efectiva: Comunicarse de manera efectiva implica haber desarrollado la habilidad para transmitir conocimientos y expresar ideas y argumentos de manera clara, rigurosa y convincente, tanto oralmente como por escrito, utilizando recursos adecuados y adaptándose a las circunstancias y al tipo de audiencia.

TC2: Innovación y creatividad: El desarrollo de esta competencia requiere tanto pensar de manera diferente para proporcionar perspectivas distintas (creatividad) como comprometer ciertos recursos por iniciativa propia para explorar una oportunidad, asumiendo el riesgo que esto conlleva (emprendimiento).

TC3: Análisis y resolución de problemas: Esta competencia se refiere a la necesidad de que los estudiantes puedan aplicar procedimientos estructurados para resolver problemas y tomar decisiones, promoviendo así su capacidad para aprender, entender y aplicar conocimientos de forma autónoma, así como para entender los mecanismos de expansión y difusión del conocimiento.

IV. DEFINICIÓN DE PROYECTO DE CIENCIA DE DATOS

Con el objetivo de adoptar una metodología ABP en CDA, debemos seleccionar cuidadosamente los proyectos a desarrollar y, específicamente, cuáles deben ser las principales características de estos proyectos orientados a los datos y exploratorios, también intentando liberar las habilidades prescriptivas e inquisitivas de los estudiantes. Primero, fue necesario establecer qué tipo de proyectos resolver. Para aumentar la motivación y mejorar las habilidades de innovación y curiosidad de los estudiantes, decidimos que deberían ser problemas del mundo real para que los estudiantes pudieran aportar contribuciones novedosas en un campo determinado. Sin embargo, el desafío principal (y el riesgo) es que la propuesta del proyecto se conciba como un proyecto de científico de datos autónomo y, por

lo tanto, los estudiantes son responsables de la definición y desarrollo del proyecto. No se proporcionarán especificaciones, pautas o recomendaciones claras a los estudiantes, a diferencia de otros cursos de informática o ingeniería de software basados en ABP. Tampoco se les proporciona ningún tipo de tema predefinido, etapas, plantilla o tecnología a utilizar. La razón de esto es que, si los estudiantes sugieren el proyecto y cómo resolverlo, su motivación es muy alta [38]. De esta manera, los propios estudiantes deben desarrollar la idea de un nuevo producto a partir de los datos, o la mejora de un procedimiento existente con conocimiento adquirido a través de los datos. Los estudiantes elegirían desarrollar productos basados en datos sobre un tema que les interesa genuinamente, también pensando en lo que un empleador o el público general querría ver. En cualquier caso, los instructores deben proporcionar alguna orientación y aplicar ciertas técnicas para mantener a los estudiantes motivados, como ayudar a los estudiantes a establecer sus proyectos, seleccionar los temas que pueden ser más atractivos, proporcionar repositorios de datos abiertos, etc., así como fomentar una competencia sana entre ellos si los proyectos a desarrollar están relacionados.

Los proyectos orientados a los datos también deben ser de principio a fin, requiriendo que los estudiantes formen los equipos, exploren las opciones de valor de los datos, realicen la investigación de mercado, especifiquen los objetivos, identifiquen las fuentes de datos, diseñen soluciones, realicen el análisis exploratorio de los datos, construyan, evalúen, implementen y mantengan modelos, y lidiar con los problemas de comunicación y presentación si es aplicable. Los estudiantes también administrarán su horario, aunque se proporcionan algunas directrices junto con los plazos para las diferentes fases a desarrollar, como la formación de equipos, la investigación de mercado, la recopilación de datos, la implementación, la implementación final, la presentación de informes, etc.). Durante el desarrollo del proyecto, el instructor desempeña el papel de un "asesor" (un orientador) destinado a proporcionar asistencia cuando sea necesario, principalmente en el desarrollo de las diferentes fases de los proyectos (p. ej., ayudar en la búsqueda de nuevas ideas, proporcionar fuentes útiles de datos, responder a preguntas técnicas, etc.). Desde un punto de vista pedagógico, hay una serie de "tareas" que el instructor debe realizar para que el proyecto sea exitoso:

- **Seminarios:** Los seminarios servirán para introducir herramientas, métodos, etc., con los que el estudiante no está familiarizado y que podrían ser algo difíciles de aprender por sí mismos desde el principio. Posteriormente, serán los estudiantes quienes profundicen en ellos por su cuenta.
- **Seguimiento grupal:** El instructor debe conocer la evolución mostrada por cada grupo en la resolución del proyecto y el grado de implicación de cada estudiante en él (no todos los estudiantes aprenden de la misma forma o al mismo ritmo). El seguimiento efectivo se puede realizar durante las clases, o de una manera más personalizada a través de reuniones grupales, y depende de reunir información y dar feedback sobre las interacciones de los grupos, así como anticiparse y prepararse para posibles

TABLE I
RÚBRICA DE EVALUACIÓN UTILIZADA EN CURSOS DE CIENCIA DE DATOS BASADOS EN ABP. LOS ÍTEMS TIENEN EL MISMO PESO.

Ítem	Excelente (≈ 10)	Bueno (≈ 7.5)	Satisfactorio (≈ 5)	Necesita mejora (≈ 2.5)
VALOR de los datos Peso: 20%	Los estudiantes han identificado el valor de los datos con los que han trabajado, su aplicabilidad en su mundo contemporáneo, las personas que pueden beneficiarse de este trabajo y posibles aplicaciones o incluso una idea de emprendimiento futuro como resultado. También han identificado las limitaciones y problemas de sostenibilidad, así como el impacto global del uso de los datos y la idea propuesta en la sociedad.	Los estudiantes han identificado el valor de los datos y su aplicabilidad, los beneficiarios y limitaciones finales. También identifican el impacto del producto de datos y los riesgos de su uso.	Los estudiantes han identificado brevemente el valor final de los datos y su aplicabilidad. El proyecto está a veces dominado por los detalles sin ver el panorama general.	Los estudiantes describen los datos, pero no logran transmitir cuál será el propósito de todo esto y por qué proporcionará valor y por qué será novedoso.
ALTERNATIVAS e innovación Peso: 20%	Los estudiantes han buscado propuestas alternativas (bibliografía, sitios web, aplicaciones) para el mismo dominio, los mismos datos o aplicación, y han comparado (cuantitativamente) sus resultados con ellos al menos a nivel abstracto, y vieron si lo que presentan es lo mismo o innovador, está por debajo del estado actual del arte, cubre necesidades reales, etc. Se acogen bien los estudios preliminares de mercado.	Los estudiantes han buscado algunas propuestas alternativas para el mismo dominio, datos o aplicación, también haciendo algunas comparaciones cualitativas o abstractas, señalando las innovaciones del proyecto en términos generales.	Los estudiantes han buscado pocas propuestas alternativas de manera general, con poca o ninguna relación con el proyecto presentado. Se hacen comparaciones muy genéricas y las innovaciones no se presentan de manera clara y específica.	Los estudiantes han pasado por alto las propuestas alternativas y no se han hecho comparaciones.
Integración de HER- RAMIENTAS TÉCNICAS Peso: 20%	Los estudiantes han dominado diferentes nuevas herramientas técnicas (p. ej., IDEs de desarrollo y notebooks, librerías de análisis y modelización de datos Python/R, software de visualización, scrapping de datos/web, gestión de APIs, etc.) y su integración para cumplir adecuadamente sus objetivos. El trabajo y la experiencia vistos durante el curso se reflejan en las soluciones técnicas, las cuales toman la mejor opción del estado de la técnica y la literatura. Las soluciones muestran iniciativa y originalidad.	Los estudiantes han integrado varias herramientas (p. ej., librerías de análisis de datos, modelización y visualización Python/R, scrapping de datos, etc.), y las han utilizado adecuadamente. Las soluciones reflejan una cantidad importante de esfuerzo y adecuación para sus necesidades.	Los estudiantes han utilizado algunas herramientas (p. ej., librerías básicas y software) apropiadamente para sus objetivos. El trabajo y la experiencia vistos durante el curso se reflejan en las soluciones técnicas.	Los estudiantes utilizan herramientas inapropiadas. No se ha puesto suficiente esfuerzo en encontrar las herramientas adecuadas o aprender nuevas.
ESFUERZO del proyecto Peso: 20%	Los estudiantes han trabajado con diferentes repositorios de datos, y han realizado un gran esfuerzo de curación, integración y recopilación (p. ej., a través de un apropiado proceso de integración de datos ETL). Además, han probado muchos modelos diferentes y variantes de sus características (p. ej., a través de mallas de ajuste y búsqueda de hiperparámetros), han seleccionado las métricas adecuadas y los protocolos de evaluación (p. ej., división, retención, k-fold, etc. recetas de validación). Los estudiantes muestran las lecciones aprendidas y cómo han resuelto los problemas o encontrado una solución. Muestran una clara evidencia de trabajo en equipo.	Los estudiantes han utilizado, cuidado e integrado diferentes repositorios de datos (con algún esfuerzo de integración a través de procesos ETL apropiados), han probado diferentes modelos, han seleccionado las métricas adecuadas y los protocolos de evaluación. Muestran algunas lecciones aprendidas.	Los estudiantes han utilizado pocos repositorios de datos (con poca integración), y han probado pocos modelos. Hay algunos problemas en la evaluación de los modelos, las métricas elegidas o las lecciones aprendidas proporcionadas.	Los estudiantes han utilizado muy pocos repositorios de datos (con poca o ninguna integración), han probado muy pocos modelos y/o los han evaluado incorrectamente, con una selección de modelo incorrecta o sobreajuste, y sacando conclusiones erróneas.
Calidad de la EXPOSICIÓN Peso: 20%	Los estudiantes han sido capaces de transmitir las ideas muy claramente, la motivación del trabajo y los insights. La calidad de las diapositivas y los gráficos es impecable, con el elemento correcto para ilustrar cada punto de la historia. Hacen gestos, usan recursos de expresión y realmente involucran a la audiencia. Están contando una historia. Responden las preguntas de manera correcta y precisa.	Los estudiantes han sido capaces de transmitir las ideas principales y los resultados. La presentación está bien organizada y apoyada por gráficos. Hacen algunos gestos para evitar ser monótonos. Están contando una historia. Responden las preguntas correctamente.	Los estudiantes transmiten de qué trata su proyecto. La presentación no parece tener una organización clara, y los gráficos se usan por disponibilidad más que por oportunidad. Hacen una presentación monótona. Responden a la mayoría de las preguntas correctamente.	Los estudiantes no son capaces de transmitir las ideas clave de su proyecto. La presentación es desordenada y no está bien respaldada por gráficos. Son aburridos, cometen muchos errores o no saben cómo seguir. Responden a muchas preguntas de manera incorrecta o vaga.

problemas (por ejemplo, progreso inadecuado, miembros que no contribuyen, etc.).

- **Feedback grupal:** Para ayudar a los estudiantes a avanzar con su proyecto, los instructores deben proporcionar constantemente una retroalimentación cualitativa y/o cuantitativa para reducir la brecha entre los resultados de aprendizaje esperados y los reales. El feedback se proporciona siguiendo dos mecanismos: 1) discusiones cara a cara donde se discuten aspectos específicos de los proyectos; o 2) comentarios y sugerencias escritas. Vemos esto en más detalle en la siguiente sección.

V. EVALUACIÓN DEL PROYECTO

La calificación de la materia se basa en tres actos de evaluación: 1) Pruebas cortas en clase (dos evaluaciones, cada una el 10% de la nota final); 2) evaluaciones prácticas (tres evaluaciones, cada una el 10% de la nota final); y 3) un trabajo (proyecto) que incluye una presentación oral (50% de la nota final), siendo este último el componente principal del ABP. Para evaluar el rendimiento de los estudiantes en estos proyectos, los estudiantes crean un portafolio de proyectos definiendo las actividades, que también se utiliza para evaluar el trabajo de los estudiantes. El proyecto debe llevarse a cabo en grupos de 2 a 4 personas. Dado que los estudiantes tienen que trabajar en equipo, deben emplear competencias transversales como la

flexibilidad, la organización, la resolución de problemas, las habilidades de negociación, el liderazgo, etc.

Todos los proyectos requieren una presentación final (realizada por todos los miembros del grupo) describiendo el desarrollo y los resultados principales. Vale la pena mencionar que un error muy común es calificar este tipo de proyectos desde un punto de vista excesivamente utilitarista. El producto del proyecto no está destinado a ser directamente utilizable por la sociedad. Aunque el proyecto debe entregar un producto, la evaluación debería centrarse en *qué tan ingeniosos son los estudiantes durante el proceso*. El hecho de que haya "problemas" que los estudiantes detecten y resuelvan, debe ser fomentado y no desalentado, y así debe reflejarse en la calificación.

Los estudiantes podrán solicitar comentarios durante el desarrollo del proyecto. Además, los estudiantes tienen una (primera) pre-evaluación (ensayo) de su presentación, recibiendo retroalimentación punto por punto. Dos semanas después de este primer intento, pueden hacer la (segunda) presentación/evaluación final. La nota de la pre-evaluación puede ser considerada final si los estudiantes deciden no reenviar su trabajo para la evaluación final. No hay más oportunidad de recuperación después de esta evaluación final. Este proceso de ensayo-mejora-reenvío del proyecto se lleva a cabo en dos semanas. Durante las presentaciones, los estudiantes de otros

grupos deberían hacer preguntas y expresar lo que el proyecto les transmite. Los instructores escriben un informe detallado siguiendo una rúbrica de evaluación (mostrada en la Tabla I). Esta primera evaluación viene con una nota provisional. Basándose en esta retroalimentación, que incluye las diferentes secciones de la rúbrica, especialmente la presentación, el grupo trabaja en mejorar el alcance del proyecto y puede hacer una presentación final en la última semana.

La rúbrica tiene como objetivo no solo una evaluación sistemática sino también proporcionar una estructura de trabajo y guía para los estudiantes acerca de dónde enfocarse. En el desarrollo de la rúbrica, se intentó identificar los estándares y habilidades clave para los científicos de datos [39], [40], así como los criterios de evaluación esenciales. La rúbrica permitiría evaluar el conjunto de conocimientos y competencias requeridas en proyectos basados en datos [15], [20]. También resalta habilidades significativas y ciertos aspectos relacionados con la naturaleza exploratoria de la ciencia de datos, desde el descubrimiento del valor de los datos hasta la construcción y evaluación final del modelo. La presentación de los resultados debería incluir una narrativa atractiva para guiar al público a través del producto final. A nivel técnico, la rúbrica evaluará la capacidad de los estudiantes para utilizar herramientas y métodos de código abierto, aplicar técnicas de machine learning y minería de datos y probar e interpretar modelos predictivos apropiadamente.

La rúbrica también evalúa competencias transversales del curso a través de distintos ítems. Así, el ítem *calidad de la exposición* evalúa la comunicación eficaz (CT1), y *alternativas e innovación* mide la originalidad y viabilidad de ideas (CT2). Finalmente, las habilidades de resolución de problemas (CT3) se evalúan con *integración de herramientas técnicas* (Tabla I).

TABLE II
RÚBRICA DE CO-EVALUACIÓN DE GRUPOS UTILIZADA EN NUESTRO CURSO ABP.

Ítem	Descripción
Contribución	¿Cuál es el porcentaje de la contribución total que se puede atribuir a su compañero de equipo X? (considerando el resultado del proyecto, no las horas de trabajo, ya que algunas personas son más eficientes que otras)
Disposición	En una escala del 0 al 100, ¿cómo valoraría la actitud colaborativa de su compañero de equipo X? (disposición, ayuda, búsqueda de consenso en lugar de conflictos, etc.)

Además, decidimos que era necesario adoptar también técnicas de evaluación cruzada para evaluar mejor el rendimiento de cada estudiante, de manera que la calificación esté más en línea con la contribución de cada estudiante dentro del grupo. En este sentido, la evaluación del proyecto requiere que cada estudiante complete una evaluación de sus compañeros sobre el rendimiento de los miembros de su grupo, como se muestra en la Tabla II. Esta evaluación se lleva a cabo después de la presentación del proyecto. La suma de "contribución" para todos los participantes de un equipo en particular debe ser 100. Dado el par de marcas para todos los miembros del equipo en un grupo, los profesores utilizan un procedimiento (no divulgado para evitar optimizarlo en lugar de ser honesto), para derivar un coeficiente entre 0 y 1.2 para

cada estudiante, de acuerdo con los valores y la armonía de estas evaluaciones cruzadas. Este coeficiente debería estar cerca de 1 si el estudiante ha hecho una contribución justa a la parte del proyecto, ha demostrado una actitud colaborativa hacia sus compañeros de equipo, etc. Finalmente, este coeficiente multiplica la puntuación dada por el instructor para obtener la calificación individual final del proyecto.

VI. EJEMPLOS DE PROYECTOS

Con propósito ilustrativo, mostramos el feedback proporcionado (vía correo electrónico) para un par de proyectos con diferentes niveles de madurez. El primero propuso un enfoque de aprendizaje automático para analizar y predecir el empleo y los factores económicos relacionados basándose en datos del mercado de valores (IBEX35). Para la primera evaluación, el proyecto aún estaba en una etapa preliminar y adolecía de una falta de motivación y justificación clara. A su vez, la narrativa era deficiente, no realizaba una comparación con trabajos relacionados y la coherencia y cohesión en cuanto al análisis y los resultados era mejorable. Después del feedback (ver Tabla III), en la reevaluación, el grupo mejoró su puntuación final (promedio) de 2.8/10 a 5/10, abordando algunas de las cuestiones y comentarios planteados.

TABLE III
REEDBACK PROPORCIONADA PARA UN PROYECTO INMADURO SOBRE EL ANÁLISIS DE LA RELACIÓN DEL EMPLEO Y OTROS INDICADORES SOCIALES BASADO EN DATOS DEL MERCADO DE VALORES.

Ítem	Feedback
Valor de los DATOS	Bien hecho: El empleo es un problema importante y cuatro grandes empresas pueden tener un efecto en el empleo. Para mejorar: No está claro cuál es el valor que se encuentra aquí. Las empresas pueden aumentar los beneficios y reducir los empleados o al revés. Pero la relación es compleja y usando cuatro empresas va a ser inconcluso de todas formas. Necesitaríamos una pregunta concreta de importancia que se debe responder, como las que los estudiantes introducen en las conclusiones, y quién va a beneficiarse de los resultados de este estudio y cómo.
Nota: 3/10	
ALTERNATIVAS e innovación	Bien hecho: - Para mejorar: Hay una enorme literatura económica sobre el efecto del empleo y las tendencias de las empresas, pero normalmente los estudios usan modelos económicos y más datos. Sugiero a los estudiantes que consideren algunos datos que normalmente no se analizan en los estudios económicos, para hacer esto un poco más innovador y no competir con los economistas expertos. Debe ser algo diferente para destacar.
Nota: 0/10	
Integración de herramientas TÉCNICAS	Bien hecho: Hay cierto procesamiento al leer el CSV e integrar y preparar los datos. Para mejorar: El análisis es simplemente una colección de gráficas de barras y de líneas. Necesitamos usar algunas de las herramientas de modelado para encontrar conglomerados, tendencias, secuencias, etc. Esto puede requerir más datos que solo cuatro empresas, o información más detallada para cada empresa.
Nota: 3/10	
ESFUERZO del proyecto	Bien hecho: Hay cierto esfuerzo al recopilar e integrar los datos. Para mejorar: Los datos son insuficientes y el análisis es insuficiente. Debería realizarse más esfuerzo en muchos otros aspectos del proyecto, incluyendo valor, novedad, modelado y presentación.
Nota: 5/10	
Calidad de la EXPOSICIÓN	Bien hecho: En inglés. "La diapositiva del bien y del mal" es el tipo de diapositivas que ayudan a captar la atención. Los estudiantes dieron buenas respuestas a las preguntas. Para mejorar: Comienzan muy planos, sin siquiera aclarar cuál es el objetivo. Deberían pensar en algo que capte la atención, en lugar de un resumen. Por ejemplo, las conclusiones al final plantean preguntas, algunas de las cuales podrían enfatizarse al principio. Los presentadores deberían dar un traspaso más suave entre ellos, manteniendo el flujo. Al final, la presentación no parece totalmente integrada; le falta una narrativa. Sería bueno tener algunos aspectos destacables al final. Respecto al estilo de la presentación, los tres presentadores deberían transmitir más entusiasmo. Las gráficas son muy simples y toscas y las diapositivas en general podrían mejorarse significativamente (las primeras diapositivas son capturas de pantalla de R y el resto es muy esquemático).
Nota: 9/10	

El segundo proyecto estaba relacionado con una mejor comprensión de la exposición de los conductores a los riesgos de tráfico a lo largo de rutas especificadas, incluyendo

TABLE IV
FEEDBACK PROPORCIONADO PARA UN PROYECTO MADURO SOBRE EL ANÁLISIS DE RIESGOS DE TRÁFICO PARA LOS CONDUCTORES A LO LARGO DE RUTAS ESPECIFICADAS.

Ítem	Feedback
Valor de los DATOS	Bien hecho: Hay dos áreas principales de valor, la recomendación de rutas seguras y el análisis de los datos para la formulación de políticas. Se descubrieron o confirmaron algunos hallazgos interesantes, como la influencia de la edad. Para mejorar: Se podría obtener más valor de otras formas, especialmente si se hacen recomendaciones diferentes dependiendo de la edad o de la misma ruta pero con recomendaciones de velocidad, etc.
Nota: 9/10	
ALTERNATIVAS e innovación	Bien hecho: Han comparado con RouteWise (navegación segura basada en rutas) y documentos. Para mejorar: Los gobiernos y las aseguradoras analizan los datos de tráfico/accidentes. Es difícil ser novedoso aquí, pero el uso de información externa (p. ej., densidad, actividad industrial, edad por área, etc.) podría hacer que el análisis sea diferente.
Nota: 8/10	
Integración de herramientas TÉCNICAS	Bien hecho: Los estudiantes han integrado ideas del curso. Se utilizaron mapas en varios puntos. Para mejorar: Más modelado, como formas de predecir la probabilidad de accidente de una ruta teniendo en cuenta las condiciones, la edad del conductor, etc.
Nota: 8/10	
ESFUERZO del proyecto	Bien hecho: Trabajaron con grandes bases de datos, un esfuerzo importante en la preparación de datos y la transformación de atributos. Para mejorar: Se podría haber explorado más el modelo de recomendación de rutas.
Nota: 9/10	
Calidad de la EXPOSICIÓN	Bien hecho: Me gustó mucho la diapositiva "1.2 Millones", ya que atrae la atención del público a un problema importante. Se concentraron en hallazgos sorprendentes y su explicación (como los niños conduciendo). Buenos ejemplos (p. ej., rotondas). Buenas respuestas a las preguntas. Para mejorar: Se dedicó demasiado tiempo al atributo del clima. Los gráficos utilizados para representar leve vs grave no son la mejor opción, ya que todo es relativo (proporción) pero no vemos las magnitudes.
Nota: 9/10	

la propuesta de rutas alternativas seguras. Para la primera evaluación, el nivel de preparación del proyecto fue muy alto y los estudiantes hizo un buen trabajo en el establecimiento de los principales requisitos, objetivos y motivación. Realizaron un análisis exploratorio de datos correcto e ilustrativo, preprocesamiento, modelado y evaluación, así como una exposición clara de las lecciones aprendidas. Los estudiantes se mostraron satisfechos con su calificación promedio después de su primera presentación (8.6/10) y decidieron no mejorar ni reenviar su trabajo en la evaluación final. Esta fue la calificación final del proyecto (antes de usar los ajustes de co-evaluación).

VII. RESULTADOS Y EVALUACIÓN DE LA METODOLOGÍA ABP

A. Aplicación de la rúbrica

La Figura 2 muestra los resultados promedio de las evaluaciones del curso CDA en 2021. El curso contó con un total de 29 estudiantes, que formaron 10 grupos. La serie roja (*evaluación única*) en la Figura 2 representa el rendimiento promedio (con intervalos de confianza) de los estudiantes que estaban satisfechos con su puntuación en la pre-evaluación y decidieron no hacer la reevaluación (8 grupos y 25 estudiantes en total). La serie azul (*pre-evaluación*) representa el rendimiento promedio en la fase de pre-evaluación de aquellos estudiantes que decidieron hacer la reevaluación (2 grupos y 4 estudiantes en total). Finalmente, la serie verde (*reevaluación*) representa el rendimiento de los estudiantes anteriores en la reevaluación después de recibir comentarios.

La Figura 2 muestra que en elementos de la rúbrica como "Esfuerzo del proyecto", "Integración de herramientas

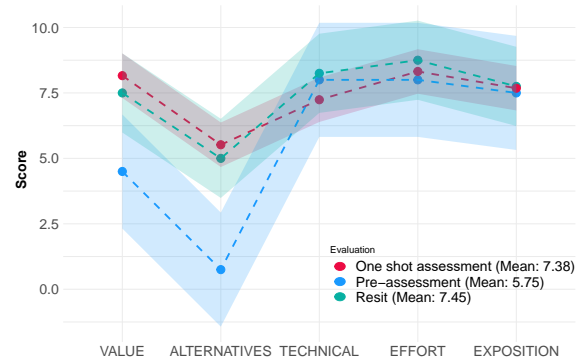


Fig. 2. Rendimiento de los estudiantes en los diferentes elementos de la rúbrica (para el curso CDA Otoño 2021). Puntuación promedio resumida por procedimiento/fase de evaluación: evaluación única (rojo), pre-evaluación (azul) y evaluación final (verde).

técnicas" y "Exposición", los estudiantes lo hicieron extremadamente bien. Curiosamente, los resultados muestran que los estudiantes por debajo de la media rinden muy mal en el elemento de la rúbrica "Alternativas e Innovación", donde los estudiantes deberían realizar una búsqueda de literatura científica e investigación de mercado para propuestas alternativas, también comprobando cómo se levanta su proyecto con respecto al estado del arte. Este elemento

mejoró significativamente después de los comentarios proporcionados. Parece sorprendente que la capacidad de buscar alternativas sea un predictor tan importante del rendimiento del equipo.

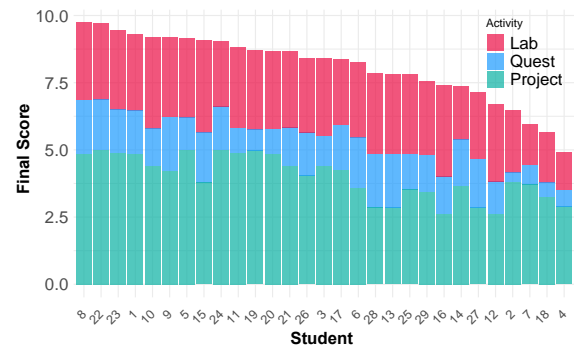


Fig. 3. Puntuaciones finales de los estudiantes, desglosadas por actividad.

A nivel individual, vemos en la Figura 3 que las calificaciones del proyecto no correlacionan mucho (correlación de orden de Spearman) con las obtenidas en la puntuación media de las tres evaluaciones prácticas (0.21) y en los dos cuestionarios (0.05). Sin embargo, la correlación entre las prácticas y los cuestionarios es mucho mayor (0.47). Esta baja correlación podría venir simplemente del hecho de que las prácticas y los cuestionarios son calificados individualmente y la calificación del grupo es colectiva, por lo que depende de los otros miembros del grupo. Esto también puede ser entendido como una indicación de que la ABP puede llevar a un menor rendimiento de los estudiantes en ciertos resultados del curso (más tradicionales) mientras que todavía se obtiene un mejor

rendimiento en las tareas más experimentales, pragmáticas y enfocadas al mundo real.

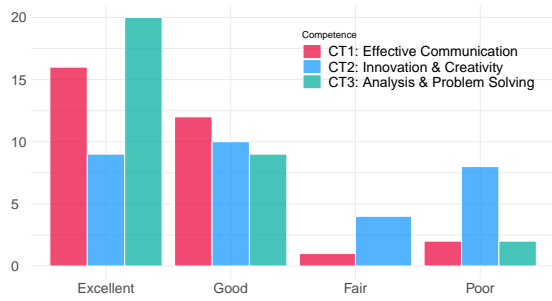


Fig. 4. Evaluación de las competencias a través de los elementos en la rúbrica de la Tabla I.

Por su parte, como se menciona en la sección V, la evaluación del proyecto final a través de la rúbrica también nos permite evaluar la adquisición de competencias transversales mediante los ítems incluidos. En este sentido, las calificaciones finales obtenidas en los correspondientes ítems serán utilizadas para obtener los valores requeridos por la universidad, que deben ser expresados mediante una clasificación en una escala de 4 valores: *Pobre* ($0 \leq x \leq 0.3$), *Justo* ($0.3 < x \leq 0.5$), *Bueno* ($0.5 < x \leq 0.8$) y *Excelente* ($0.8 < x \leq 1.0$). La Figura 4 muestra los resultados generales obtenidos. En general, la mayoría de los estudiantes han adquirido satisfactoriamente las diferentes competencias transversales. Esto coincide con la percepción personal de los profesores del curso en comparación con otros cursos impartidos por ellos donde se siguen metodologías de enseñanza más tradicionales. Encontramos los mayores problemas con la competencia de innovación y creatividad, donde alrededor del 25% de los estudiantes no adquirió la competencia de la manera esperada. El bajo rendimiento de estos estudiantes en el elemento de la rúbrica "Alternativas e Innovación" (ya mostrado en la Figura 2) muestra un aspecto claro del curso que podemos intentar mejorar en los próximos años (por ejemplo, fomentando sesiones de lluvia de ideas con los compañeros de equipo, actividades de investigación, incorporando más comentarios de los compañeros de equipo y profesores, trabajando en clase en temas relacionados como lidiar con la incertidumbre, la ambigüedad, la independencia, la tenacidad, etc.).

B. Encuesta de metodología de curso

Para evaluar si la metodología de enseñanza ha aumentado la motivación de los estudiantes para lograr las competencias establecidas en el curso de ciencia de datos, hemos desarrollado un cuestionario para ser completado por los estudiantes a través de Google Forms. El cuestionario contiene 10 preguntas, 8 de ellas basadas en escalas de puntuación tipo Likert de 5 puntos (desde "Muy en desacuerdo" hasta "Muy de acuerdo") y las 2 restantes son abiertas. Completar el formulario era opcional y recibimos un total de 20 respuestas. Los detalles de las preguntas se pueden encontrar en la Tabla V.

Las preguntas Q1 a Q3 se refieren a la motivación de los estudiantes con respecto a la metodología seguida en el curso. Las preguntas Q4 a Q8 se refieren a los objetivos

y competencias del curso. Finalmente, las preguntas Q9 y Q10 intentan sacar a relucir el pensamiento crítico sobre la metodología seguida en el curso.

TABLE V
CUESTIONARIO DE EVALUACIÓN DE METODOLOGÍA DE ENSEÑANZA.

Q	Pregunta	Respuesta
1	¿Te motivó la forma en que se desarrolló el curso?	Likert (1-5)
2	¿Encontraste interesante el proyecto de ciencia de datos auto-gestionado propuesto en este curso?	Likert (1-5)
3	¿Crees que tu participación e interés en el curso ha aumentado al trabajar con tus compañeros en equipo?	Likert (1-5)
4	El objetivo del proyecto final es que comprendas el papel del científico de datos en las organizaciones, identificar problemas y oportunidades y desplegar soluciones utilizando herramientas de uso común. ¿Crees que estos objetivos se han cumplido después del curso?	Likert (1-5)
5	Después de haber completado y presentado el proyecto final, ¿te sientes más seguro para aplicar lo que has aprendido en tu futuro trabajo?	Likert (1-5)
6	¿Crees que la realización del proyecto final te ha permitido trabajar en habilidades como la innovación, la creatividad y el emprendimiento?	Likert (1-5)
7	¿Crees que la realización del proyecto final te ha permitido trabajar en habilidades para resolver problemas de forma autónoma?	Likert (1-5)
8	¿Crees que el proyecto final te ha permitido trabajar en habilidades de comunicación efectiva y hablar en público?	Likert (1-5)
9	Menciona aspectos positivos de este curso	Abierta
10	Menciona aspectos negativos de este curso	Abierta

La Figura 5 muestra los resultados de las preguntas basadas en escalas de puntuación (Q1 a Q8). En las preguntas referentes al bloque de motivación (Q1 a Q3), vemos que el 5% de los estudiantes marcaron las preguntas Q2 y Q3 como "En desacuerdo". Por su parte, el 85% de los estudiantes consideran que la metodología de enseñanza ha servido para motivarlos en el aprendizaje de la ciencia de datos e interesante el proyecto de ciencia de datos propuesto en el curso. Finalmente, el 75% de los estudiantes consideran que el haber trabajado en colaboración en un equipo ha mejorado no sólo su participación sino también su interés en la materia.

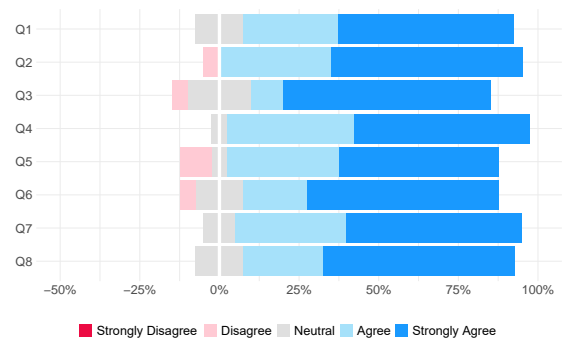


Fig. 5. Resumen de resultados para la encuesta en la Tabla V. 20 respuestas procesadas.

El segundo grupo de cuestiones (Q4-Q8) indica que los alumnos concuerdan mayoritariamente en que se han cumplido los objetivos y habilidades propuestos. Destacan Q5 y Q6, donde hay discrepancias. En Q5, aproximadamente un 10% siente inseguridad al aplicar lo aprendido en el curso, posiblemente por la distancia entre su profesión y la ciencia de datos o por sentir que necesitan formación más avanzada. Q6 revela que un 5% afirma no haber potenciado habilidades

como innovación, creatividad o emprendimiento, quizás por falta de curiosidad o dificultades para concebir y desarrollar proyectos. Nadie opinó negativamente en Q7-Q8, indicando que las aptitudes de resolución de problemas y comunicación se han trabajado bien en el curso. En general, las respuestas muestran alto grado de aceptación, con todas las cuestiones superando el 75% de respuestas positivas, y algunas (como Q4 y Q5) rozando el 100%. Estos resultados resaltan que la mayoría de los alumnos siente que se han cumplido los objetivos y competencias propuestas con la metodología de enseñanza usada.

Además de las preguntas tipo Likert, se utilizaron Q9 y Q10 para provocar los aspectos positivos y negativos de la metodología de enseñanza. A continuación, resumimos las ideas más relevantes que extrajimos de ambas preguntas. En cuanto a los aspectos positivos:

- Se valora el trabajo en equipo y la coordinación.
- Se potencia la creatividad y la solución de problemas.
- Los estudiantes elogian la estructura del curso, las prácticas desafiantes y emocionantes, así como la utilidad real del proyecto final.
- La sólida base en fundamentos de ciencia de datos es útil.

En cuanto a los aspectos negativos, destacamos lo siguiente:

- Fue difícil para algunos estudiantes encontrar ideas y datos interesantes para el proyecto final.
- Algunos estudiantes carecían de tiempo para desarrollar completamente algunas partes de su proyecto.
- Algunos estudiantes consideraban que los plazos para las prácticas y la entrega del proyecto eran demasiado ajustados.

En general, la mayoría de los estudiantes han reconocido que la metodología seguida y la participación y desarrollo de proyectos les ha obligado a aprender a través de la colaboración, la investigación y una mejor comunicación, mejorando también sus habilidades interpersonales.

Solemos recibir también comentarios positivos de los estudiantes a través de las encuestas de la universidad. Además de algunas puntuaciones, estas encuestas también proporcionan información útil sobre el rendimiento de los instructores. La encuesta contenía un total de 10 preguntas que buscaban analizar diferentes aspectos de la asignatura, así como tratar de verificar la consistencia de las respuestas, para evitar completar al azar. Concretamente, con una satisfacción media general de 8.5/10, la mayoría de los estudiantes reconocieron que los instructores muestran un verdadero compromiso e interés por sus estudiantes, saben ser flexibles y atender las necesidades individuales.

VIII. CONCLUSIONES

Este artículo ha presentado la experiencia, las innovaciones y las lecciones aprendidas de la aplicación de una estrategia de ABP a una asignatura de ciencias de datos. De hecho, este trabajo proporcionó un conjunto de recursos didácticos innovadores, herramientas de evaluación y conocimiento que podrían utilizarse para organizar cursos de ciencias de datos en general. Hemos visto cómo la ausencia de esquemas rígidos y el carácter exploratorio de las ciencias de datos, así como

el rango de habilidades proactivas, curiosas e inquisitivas que los científicos de datos deberían tener, hacen que la adopción de ABP sea más ventajosa, pero también más desafiante, precisamente porque las especificaciones y las etapas no son tan claras como en otros proyectos de ingeniería. Aún así, los resultados obtenidos siguen la misma línea en términos de la efectividad y relevancia de las metodologías de ABP en otras disciplinas de computación e ingeniería (ver, por ejemplo, [4], [5], [8]).

El objetivo para los estudiantes es aprender a desarrollar proyectos por sí mismos, considerando también el papel relevante del valor de los datos y la innovación. Al aplicar nuestras evaluaciones basadas en rúbricas, también ayudamos a garantizar que las discusiones y el feedback con los estudiantes son más efectivos, además de hacer el seguimiento y la evaluación más sistemáticos. La metodología de enseñanza ha sido evaluada a través de un cuestionario completado por los estudiantes y el análisis de las calificaciones obtenidas en la actividad. Ambos elementos muestran que se han logrado los objetivos y las competencias necesarias para trabajar como científicos de datos. Además, según las respuestas al cuestionario, parece que la metodología seguida ha aumentado la motivación de los estudiantes, captando su interés y atención. Esta percepción también es compartida por los profesores, en base a su experiencia en la enseñanza de otros cursos de temas similares que no siguen una metodología ABP.

Aunque la aplicación de esta metodología de enseñanza generó resultados positivos para los cursos de ciencias de datos basados en proyectos, es necesario hacer más trabajo para comparar aún más los resultados de esta metodología de ABP con otros enfoques pedagógicos potenciales. Por ejemplo, uno podría comparar estudios de casos o competiciones estilo Kaggle² con este curso enfocado en proyectos. Además, esta metodología de enseñanza aún puede ser refinada y mejorada. En función de los comentarios de los estudiantes y de la universidad, los ajustes específicos pueden incluir la reducción, pero no la eliminación, de la cantidad de autoaprendizaje que necesita tener lugar para que los estudiantes tengan éxito en la clase, así como el número de prácticas de laboratorio a realizar (considerando el percibido esfuerzo además del proyecto final). Además, se tendrán que incorporar en el curso actividades adicionales destinadas a mejorar las habilidades de innovación de los estudiantes, ya que es la parte donde los estudiantes suelen tener un peor rendimiento según las rúbricas de evaluación. Por otro lado, el compromiso del estudiante para mejorar y reenviar el proyecto final puede aumentar aún más a través de la posibilidad de permitirles participar en un evento de cierre, taller o feria de proyectos donde los estudiantes pueden presentar su trabajo a otros estudiantes, socios industriales, y terceros interesados.

Como trabajo futuro, queremos también considerar y fomentar el papel de la automatización en estos proyectos. Como en otras áreas que utilizan IA [41], es cada vez más común que los estudiantes utilicen herramientas sofisticadas que aceleren su proceso, desde el manejo de datos hasta Auto-ML [42]. Cómo equilibrar la entrada de 'esfuerzo del proyecto' en la

²<https://www.kaggle.com/>

rúbrica con un uso más inteligente de las herramientas que reducen el esfuerzo de los científicos de datos requiere una declaración muy clara sobre cómo declarar el uso de estas herramientas y cómo se califican positivamente.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their comments. This work was funded by valgrAI, the Norwegian Research Council grant 329745 Machine Teaching for Explainable AI, the Future of Life Institute, FLI, under grant RFP2-152, the EU (FEDER) and Spanish grant RTI2018-094403-B-C32 funded by MCIN/AEI/10.13039/501100011033 and by CIPROM/2022/6 funded by Generalitat Valenciana, EU's Horizon 2020 research and innovation programme under grant agreement No. 952215 (TAILOR), US DARPA HR00112120007 (RECoG-AI) and Spanish grant PID2021-122830OB-C42 (SFERA) funded by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe".

REFERENCES

- [1] D. Nunan, C. Candlin, and H. Widdowson, *Syllabus design*. Oxford University Press Oxford, 1988, vol. 55.
- [2] J. D. Bigelow, "Using problem-based learning to develop skills in solving unstructured problems," *Journal of Management Education*, vol. 28, no. 5, pp. 591–609, 2004.
- [3] M. A. Almulla, "The effectiveness of the project-based learning (pbl) approach as a way to engage students in learning," *Sage Open*, vol. 10, no. 3, 2020.
- [4] R. Pucher and M. Lehner, "Project based learning in computer science—a review of more than 500 projects," *Procedia-Social and Behavioral Sciences*, vol. 29, pp. 1561–1566, 2011.
- [5] M. L. Fioravanti, B. Sena, L. N. Paschoal, L. R. Silva, A. P. Allian, E. Y. Nakagawa, S. R. Souza, S. Isotani, and E. F. Barbosa, "Integrating project based learning and project management for software engineering teaching: An experience report," in *SIGCSE*, 2018, pp. 806–811.
- [6] J. A. Macías, "Enhancing project-based learning in software engineering lab teaching through an e-portfolio approach," *IEEE Transactions on Education*, vol. 55, no. 4, pp. 502–507, 2012.
- [7] M. Souza, R. Moreira, and E. Figueiredo, "Students perception on the use of project-based learning in software engineering education," in *33th Brazilian Symp. on Software Engineering*, 2019, pp. 537–546.
- [8] J. W. McManus and P. J. Costello, "Project based learning in computer science: a student and research advisor's perspective," *Journal of Computing Sciences in Colleges*, vol. 34, no. 3, pp. 38–46, 2019.
- [9] F. Martínez-Plumed, L. Contreras-Ochando, C. Ferri, J. H. Orallo, M. Kull, N. Lachiche, M. J. R. Quintana, and P. A. Flach, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Trans. on Knowledge and Data Engineering*, 2019.
- [10] J. Hernández-Orallo, C. Ferri, and M. Ramírez-quintana, *Introduction to Data Mining*. Pearson, 2004.
- [11] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRISP-DM 1.0 step-by-step data mining guide," 2000.
- [12] F. Martínez-Plumed, L. Contreras-Ochando, C. Ferri, P. Flach, J. Hernández-Orallo, M. Kull, N. Lachiche, and M. J. Ramírez-Quintana, "CASP-DM: context aware standard process for data mining," *arXiv preprint arXiv:1709.09003*, 2017.
- [13] M. O'Neil, "As data proliferate, so do data-related graduate programs," *The Chronicle of Higher Education*, vol. 60, p. 1, 2014.
- [14] R. E. Anderson, M. D. Ernst, R. Ordóñez, P. Pham, and B. Tribelhorn, "A data programming cs1 course," in *46th ACM Technical Symp. on Computer Science Education*, 2015, pp. 150–155.
- [15] Y. Demchenko, A. Belloum, C. de Laat, C. Loomis, T. Wiktorski, and E. Spekschoor, "Customisable data science educational environment: From competences management and curriculum design to virtual labs on-demand," in *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*. IEEE, 2017, pp. 363–368.
- [16] J. W. Johnson, "Scaling up: Introducing undergraduates to data science early in their college careers," *Journal of Computing Sciences in Colleges*, vol. 33, no. 6, pp. 76–85, 2018.
- [17] Y. M. Kim, "The effects of pbl-based data science education program using app inventor on elementary students' computational thinking and creativity improvement," *Ilkogretim Online*, vol. 20, no. 1, 2021.
- [18] C. Dichev, D. Dicheva, L. Cassel, D. Goelman, and M. Posner, "Preparing all students for the data-driven world," in *Symposium on Computing at Minority Institutions, ADMI*, vol. 346, 2016.
- [19] C. Vera, J. Féliz, J. Antonio Cobos, M. J. Sánchez-Naranjo, and G. Pinto, "Experiences in education innovation: developing tools in support of active learning," *European Journal of Engineering Education*, vol. 31, no. 2, pp. 227–236, 2006.
- [20] D. A. Asamoah, D. Doran, and S. Schiller, "Interdisciplinarity in data science pedagogy: a foundational design," *Journal of Computer Information Systems*, vol. 60, no. 4, pp. 370–377, 2020.
- [21] B. Cassel and H. Topi, "Strengthening data science education through collaboration," in *WS on data science education*, vol. 7, 2015, p. 27.
- [22] Q. Cheng, F. Lopez, and A. Hadjixenofontos, "Integrating introductory data science into computer and information literacy through collaborative project-based learning," in *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2019, pp. 1–5.
- [23] D. White, "A project-based approach to statistics and data science," *Primus*, vol. 29, no. 9, pp. 997–1038, 2019.
- [24] L. Philip and W. K. Li, "Project-based learning via competition for data science students," *Harvard Data Science Review*, 2021.
- [25] N. Postman and C. Weingartner, "Teaching as a subversive activity," *delacorte press*, New York, 1969.
- [26] H. S. Barrows, R. M. Tamblyn et al., *Problem-based learning: An approach to medical education*. Springer, 1980, vol. 1.
- [27] J. Kay, M. Barg, A. Fekete, T. Greening, O. Hollands, J. H. Kingston, and K. Crawford, "Problem-based learning for foundation computer science courses," *CSE*, vol. 10, no. 2, pp. 109–128, 2000.
- [28] P. Dart, L. Johnston, and C. Schmidt, "Enhancing project-based learning: Variations on mentoring," in *Proceedings of 1996 Australian Software Engineering Conference*. IEEE, 1996, pp. 112–117.
- [29] W. J. Pluta, B. F. Richards, and A. Mutnick, "Pbl and beyond: Trends in collaborative learning," *Teaching and learning in medicine*, vol. 25, no. sup1, pp. S9–S16, 2013.
- [30] M. H. Baturay and O. F. Bay, "The effects of PBL on the classroom community perceptions and achievement of web-based education students," *Computers & Education*, vol. 55(1), pp. 43–52, 2010.
- [31] Y. Woo and T. C. Reeves, "Meaningful interaction in web-based learning: A social constructivist interpretation," *The Internet and higher education*, vol. 10, no. 1, pp. 15–25, 2007.
- [32] C. Hayashi, "What is data science? fundamental concepts and a heuristic example," in *Data science, classification, and related methods*. Springer, 1998, pp. 40–51.
- [33] D. Holtz, "8 Skills You Need to Be a Data Scientist," <https://blog.udacity.com/2014/11/data-science-job-skills.html>, 2014.
- [34] V. Dhar, "Data science and prediction," *Communications of the ACM*, vol. 56, no. 12, pp. 64–73, 2013.
- [35] M. Loukides, *What Is Data Science?* O'Reilly Media, Apr. 2011.
- [36] E. Commission, "European e-Competence Framework," 2016. [Online]. Available: <http://www.ecompetences.eu/>
- [37] M. J. Sá and S. Serpa, "Transversal competences: Their importance and learning processes by higher education students," *Education Sciences*, vol. 8, no. 3, p. 126, 2018.
- [38] R. Pucher, A. Mense, and H. Wahl, "Intrinsic motivation of students in project based learning in South Africa," *SAIEE*, vol. 94, pp. 7–14, 2003.
- [39] F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc., 2013.
- [40] H. Harris, S. Murphy, and M. Vaisman, *Analyzing the analyzers: An introspective survey of data scientists and their work*. O'Reilly Media, Inc., 2013.
- [41] F. Martínez-Plumed, S. Tolan, A. Pesole, J. Hernández-Orallo, E. Fernández-Macías, and E. Gómez, "Does AI qualify for the job? a bidirectional model mapping labour and ai intensities," in *AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 94–100.
- [42] T. D. Bie, L. D. Raedt, J. Hernández-Orallo, H. H. Hoos, P. Smyth, and C. K. I. Williams, "Automating data science: Prospects and challenges," *Communications of the ACM*, vol. 65, no. 2, pp. 76–87, 2022.