



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**

**Impact Factor Oracle
Documentación Técnica**



Presentado por Gadea Lucas Pérez
en Universidad de Burgos — 12 de junio
de 2023

Tutores: Virginia Ahedo y Álvar Arnaiz

Índice general

Índice general	i
Índice de figuras	iii
Índice de tablas	v
Apéndice A Plan de Proyecto Software	1
A.1. Introducción	1
A.2. Metodología Scrum y Kanban	1
A.3. Herramienta	3
A.4. Planificación temporal	3
A.5. Estudio de viabilidad	20
Apéndice B Especificación de Requisitos	33
B.1. Introducción	33
B.2. Objetivos generales	33
B.3. Catálogo de Historias de Usuario	34
B.4. Requisitos de la aplicación	37
B.5. Especificación de requisitos	39
Apéndice C Especificación de diseño	49
C.1. Introducción	49
C.2. Diseño de los datos	49
C.3. Diseño procedimental	52
C.4. Diseño arquitectónico	56
C.5. Diseño de Interfaces	57

Apéndice D Documentación técnica de programación	65
D.1. Introducción	65
D.2. Estructura de directorios	65
D.3. Manual del programador	67
D.4. Compilación, instalación y ejecución del proyecto	77
D.5. Otros módulos del proyecto	79
D.6. Pruebas del sistema	81
 Apéndice E Documentación de usuario	 87
E.1. Introducción	87
E.2. Requisitos de usuarios	87
E.3. Instalación	88
E.4. Manual del usuario	88
 Apéndice F Glosario	 99
 Bibliografía	 107

Índice de figuras

A.1. Sprint 1: Burndown chart	5
A.2. Sprint 2: Burndown chart	6
A.3. Sprint 3: Burndown chart	6
A.4. Sprint 5: Burndown chart	9
A.5. Sprint 6: Burndown chart	10
A.6. Sprint 7: Burndown chart	11
A.7. Sprint 8: Burndown chart	11
A.8. Sprint 9: Burndown chart	13
A.9. Sprint 10: Burndown chart	14
A.10.Sprint 11: Burndown chart	16
A.11.Sprint 12: Burndown chart	17
A.12.Sprint 13: Burndown chart	18
A.13.Sprint 14: Burndown chart	19
A.14.Gráfico acumulativo de todo el proyecto	20
 B.1. Diagrama de Casos de Uso	 41
 C.1. Diagrama Relacional	 52
C.2. Diagrama de secuencia de inicio de sesión	53
C.3. Diagrama de secuencia de registro	53
C.4. Diagrama de secuencia de la consulta de revistas	54
C.5. Diagrama de secuencia de la consulta del histórico del JCR de una revista	55
C.6. Diagrama de secuencia de la predicción del JCR de una revista	56
C.7. Arquitectura Cliente-Servidor de dos capas	56
C.8. Arquitectura Modelo-Vista-Controlador de tres capas	57
C.9. <i>Mockup</i> de inicio de sesión	58
C.10. <i>Mockup</i> de registro	59

C.11. <i>Mockup</i> de página principal	59
C.12. <i>Mockup</i> de predicción	60
C.13. Interfaz definitiva de inicio de sesión	61
C.14. Interfaz definitiva de registro	61
C.15. Interfaz definitiva de página principal	62
C.16. Interfaz definitiva de predicción	62
D.1. Diagrama de paquetes del repositorio	66
D.2. Fases de la metodología TDD	82
D.3. Pruebas ejecutadas exitosamente	85
E.1. Interfaz con los campos de inicio de sesión	89
E.2. Interfaz con el registro	89
E.3. Interfaz con la contraseña «a la vista»	90
E.4. Interfaz con mensajes de error	90
E.5. Interfaz con selección de revistas	91
E.6. Interfaz con selección de modelos	92
E.7. Icono de listado de revistas	92
E.8. Tooltip que aconseja rellenar todos los campos	92
E.9. Interfaz con el histórico del JCR	93
E.10. Interfaz con la predicción del JCR	93
E.11. Interfaz con el listado de revistas	94
E.12. Interfaz con la información adicional de una revista	94
E.13. Interfaz con los datos del usuario	95
E.14. Interfaz con edición de los datos de usuario	95
E.15. Barra de navegación con menú desplegable de usuario	96
E.16. Barra de navegación con menú desplegable de idiomas	96
E.17. Barra inferior de la aplicación	97
E.18. Breadcrumbs	97
E.19. Responsividad frente al cambio de tamaño de pantalla	98

Índice de tablas

A.1. Tareas del sprint 1	4
A.2. Tareas del sprint 2	5
A.3. Tareas del sprint 3	7
A.4. Tareas del sprint 4	7
A.5. Tareas del sprint 5	8
A.6. Tareas del sprint 6	9
A.7. Tareas del sprint 7	10
A.8. Tareas del sprint 8	12
A.9. Tareas del sprint 9	13
A.10.Tareas del sprint 10	14
A.11.Tareas del sprint 11	15
A.12.Tareas del sprint 12	16
A.13.Tareas del sprint 13	17
A.14.Tareas del sprint 14	18
A.15.Conceptos de Seguridad Social	23
A.16.Otros conceptos	23
A.17.Resumen de costes del personal	25
A.18.Costes de software mensuales	27
A.19.Resumen de costes del proyecto	27
A.20.Información de los paquetes y sus licencias	31
B.1. HU Extracción de datos	35
B.2. HU Modelos de predicción	36
B.3. HU Aplicación web	36
B.4. CU-1 Inicio de sesión	42
B.5. CU-2 Registro.	43
B.6. CU-3 Consultar revistas	44
B.7. CU-4 Consultar histórico	45

B.8. CU-5 Consultar predicción.	46
B.9. CU-6 Cerrar sesión	47
D.1. Casos de prueba para el prototipo de Crossref	84

Apéndice A

Plan de Proyecto Software

A.1. Introducción

La planificación temporal es esencial para el éxito de cualquier proyecto, especialmente en el desarrollo de *software*. En esta sección se detallará cómo se ha llevado a cabo el cronograma siguiendo una metodología ágil tipo Scrum, mediante el uso de *sprints*. Se describirán los pasos necesarios para planificar y gestionar de manera eficiente el tiempo y los recursos disponibles, asegurando así el cumplimiento de los objetivos del proyecto en el plazo establecido.

A.2. Metodología Scrum y Kanban

La metodología Scrum es un marco de trabajo ágil que se utiliza para la gestión de proyectos. Se basa en la colaboración entre el equipo de desarrollo y el cliente, así como en la entrega continua de productos funcionales en ciclos cortos de tiempo conocidos como *sprints* [14]. Scrum se centra en la flexibilidad y la adaptabilidad, permitiendo más fácilmente reajustes y los cambios de manera rápida y efectiva ante situaciones inesperadas.

En el contexto del presente trabajo, el uso de la metodología Scrum será beneficioso puesto que nos permitirá tener un enfoque más dinámico y flexible, lo que nos posibilitará adaptarnos a las necesidades del proyecto a medida que avanzamos. Además, Scrum nos proporcionará una mayor transparencia y comunicación, lo que nos capacitará para tomar decisiones informadas y asegurar que el proyecto se entregue a tiempo y con éxito.

Por otro lado, se ha complementado la metodología Scrum con el modelo Kanban¹ propio de la metodología Kanban [13]. Esta combinación nos permitirá visualizar y organizar de manera efectiva las tareas, los hitos y los recursos disponibles, facilitando así la gestión y el seguimiento del proyecto.

Antes de continuar con la planificación del proyecto, es importante familiarizarnos con algunos conceptos clave que son fundamentales para comprender la metodología Scrum. A continuación, se proporcionará una breve descripción de los roles, artefactos y la medición de las tareas en *story points*.

Artefactos en Scrum

En cuanto a los artefactos de Scrum, hay tres principales: el *product backlog*, el *sprint backlog* y el incremento.

1. El *product backlog* es una lista priorizada de todas las funcionalidades, características o mejoras que se desean implementar en el producto final.
2. El *sprint backlog* es una selección de elementos del *backlog* del producto que se abordarán en un *sprint* específico.
3. Y finalmente, el incremento es el resultado tangible y potencialmente entregable al final de cada *sprint*, que incorpora nuevas funcionalidades o mejoras al producto.

Roles en Scrum

En Scrum, existen tres roles principales: el Scrum Master, el Product Owner y el Equipo de Desarrollo.

1. El Scrum Master es responsable de garantizar que el equipo siga las prácticas y principios de Scrum, eliminando cualquier obstáculo que pueda afectar a la productividad.
2. El Product Owner es el encargado de definir y priorizar los elementos del *product backlog*, asegurándose de que se cumplan las necesidades del cliente.

¹Este modelo se refiere comúnmente a un lienzo o plantilla visual utilizada para estructurar y desarrollar ideas de negocio.

3. Por último, el Equipo de Desarrollo es responsable de la implementación de las tareas y la entrega del producto.

Dado que en este caso el equipo está conformado solamente por los dos tutores y la alumna, es importante adaptar los roles y las prácticas de Scrum para que se ajusten a esta situación específica. Se pueden asignar los roles de Scrum Master y Product Owner a cada uno de los tutores y, el rol de Equipo de Desarrollo, correspondería a la alumna.

Medición de las tareas

En cuanto a la medición de las tareas en *story points*, Scrum utiliza esta técnica para estimar el esfuerzo relativo requerido para completar una tarea. Los *story points* son una medida abstracta que no se relaciona directamente con el tiempo, sino con la complejidad y el esfuerzo necesario para completar una tarea en comparación con otras. Esta medida ayuda al equipo a planificar su capacidad de trabajo en cada *sprint* y a realizar estimaciones más precisas en función de su experiencia previa.

A.3. Herramienta

Para llevar a cabo este proyecto, se han utilizado herramientas colaborativas como GitHub y ZenHub. GitHub proporciona un entorno seguro y eficiente para almacenar y gestionar el código fuente, permitiendo trabajar de manera transparente y controlada. Por otro lado, ZenHub ha sido fundamental en la organización de las tareas y el seguimiento de los avances, brindando una visión clara y estructurada del progreso del proyecto.

A.4. Planificación temporal

El presente proyecto comienza en septiembre y se extiende hasta junio. Durante este período de tiempo, es importante asegurarnos de que todas las tareas e hitos estén claramente definidos.

Para lograr esto, se han realizado reuniones periódicas para discutir el progreso del proyecto y asegurarnos de que estamos en el camino correcto. Además, hemos implementado *sprints* regulares (de entre una y dos semanas de duración) para asegurarnos de que estamos avanzando de manera constante y cumpliendo con nuestras metas a tiempo.

Con esta planificación temporal sólida, estamos seguros de que podremos completar el proyecto a tiempo y cumplir con los objetivos establecidos. Sin embargo, es importante ser flexibles y estar preparados para hacer ajustes según sea necesario a medida que avanzamos en el proyecto, puesto que se realiza al mismo tiempo que transcurre el curso académico.

En las siguientes tablas se recogen los distintos *sprints* junto con su duración, objetivos y tareas. Además, se adjuntará el gráfico *burndown*² de cada uno de los *sprints*. Por otro lado, una descripción detallada de las tareas y su duración puede verse en GitHub/ZenHub.

Sprint 1

Sprt.	Duración	Objetivos	Tareas
1	19/09/2022 03/10/2022	Comenzar el desarrollo del proyecto	<ul style="list-style-type: none">▪ Elegir un modelo de referencias bibliográficas.▪ Definición de conceptos.

Tabla A.1: Tareas del sprint 1

²Gráfico *burndown*: se trata de una herramienta comúnmente utilizada en la gestión de proyectos ágiles para realizar un seguimiento de un *sprint*. Este gráfico muestra la cantidad de trabajo que queda por hacer en relación con el tiempo disponible para hacerlo. A medida que se completa el trabajo, la línea del gráfico debería disminuir hasta alcanzar cero en el momento en que finaliza el *sprint*.



Figura A.1: Sprint 1: Burndown chart

Sprint 2

Sprt.	Duración	Objetivos	Tareas
2	03/10/2022 17/10/2022	Tareas de investigación y documentación	<ul style="list-style-type: none">■ Búsqueda de antecedentes.■ Familiarizarse con el procesador de textos L^AT_EX.

Tabla A.2: Tareas del sprint 2



Figura A.2: Sprint 2: Burndown chart

Sprint 3



Figura A.3: Sprint 3: Burndown chart

Sprt.	Duración	Objetivos	Tareas
3	17/10/2022 14/11/2022	Investigación y creación de un prototipo inicial	<ul style="list-style-type: none"> ■ Búsqueda de información sobre MIAR. ■ Documentación sobre el JCR y el índice de impacto. ■ Prototipo que permita la búsqueda de artículos en GS. ■ Extracción de datos: ¿qué datos se pueden extraer? ■ Reflexión sobre el diseño de la BBDD. ■ Prueba del prototipo.

Tabla A.3: Tareas del sprint 3

Sprint 4

Sprt.	Duración	Objetivos	Tareas
4	14/11/2022 28/11/2022	Base de datos y prototipo	<ul style="list-style-type: none"> ■ Creación de la base de datos en MariaDB. ■ Prueba de la BBDD. ■ Mejoras del prototipo.

Tabla A.4: Tareas del sprint 4

NOTA

Como se puede observar no se ha añadido un gráfico *burn-down* en este *sprint*. Esto se debe a que las tareas no fueron cerradas a tiempo. Sin embargo, es preciso aclarar que se trata de un error en la configuración de las tareas. Si bien todas las tareas fueron completadas a tiempo, fue preciso borrarlas en ZenHub y volverlas a crear más adelante cuando el *sprint* ya había terminado debido a que estaban inicialmente mal configuradas.

Sprint 5

Sprt.	Duración	Objetivos	Tareas
5	28/11/2022 12/12/2022	Base de datos y prototipo	<ul style="list-style-type: none"> ■ Se muda la BBDD a Postgres. ■ Pruebas e investigación para extraer el DOI.

Tabla A.5: Tareas del sprint 5



Figura A.4: Sprint 5: Burndown chart

Sprint 6

Sprt.	Duración	Objetivos	Tareas
6	09/01/2023 23/01/2023	Documentación y obtención del DOI	<div><div></div>Obtención del DOI a través de Crossref.</div> <div><div></div>Avance de la memoria y los anexos.</div> <div><div></div>Mejora de la BBDD.</div>

Tabla A.6: Tareas del sprint 6

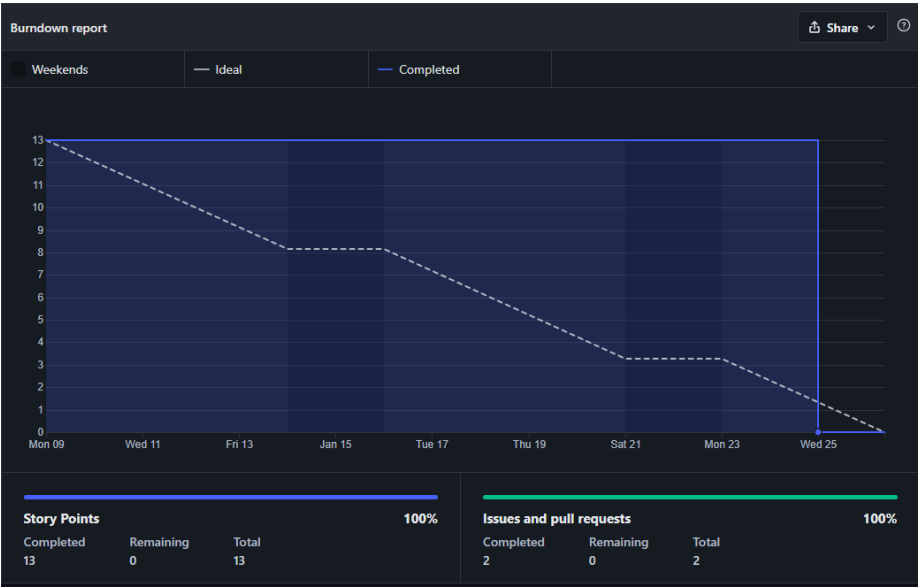


Figura A.5: Sprint 6: Burndown chart

Sprint 7

Sprt.	Duración	Objetivos	Tareas
7	06/02/2023 20/02/2023	Preparación para la conexión remota a los servidores de la universidad.	<ul style="list-style-type: none">■ Conectarse a la VPN de la universidad.■ Crear un servicio virtualizado en Miniconda.■ Mejorar la BBDD.

Tabla A.7: Tareas del sprint 7



Figura A.6: Sprint 7: Burndown chart

Sprint 8

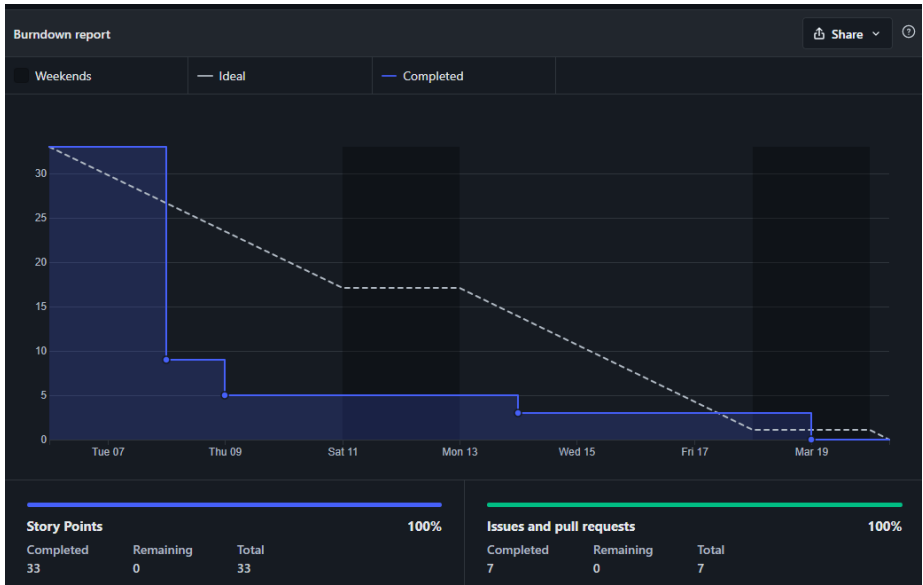


Figura A.7: Sprint 8: Burndown chart

Sprt.	Duración	Objetivos	Tareas
8	20/02/2023 06/03/2023	Conexión a los servidores y mejora de los prototipos. Avances en la API y documentación.	<ul style="list-style-type: none"> ■ Mejorar el prototipo. ■ Prototipo con varios hilos de ejecución (usar servidores de la universidad). ■ Nuevo prototipo con Selenium. ■ Sistema inteligente de requests. ■ Revisión de PoP. ■ API: Autenticación y back-end.

Tabla A.8: Tareas del sprint 8

Sprint 9

Sprt.	Duración	Objetivos	Tareas
9	06/03/2023 20/03/2023	Finalización del prototipo de extracción de Crossref.	<ul style="list-style-type: none">■ Plan de prueba para Crossref.■ Prototipo de Crossref.■ Comprobar límite de llamadas a Crossref.■ Crear entorno virtualizado.■ Fichero requirements.■ Licencia Crossref para Cited-by.■ Cálculo del IF.

Tabla A.9: Tareas del sprint 9

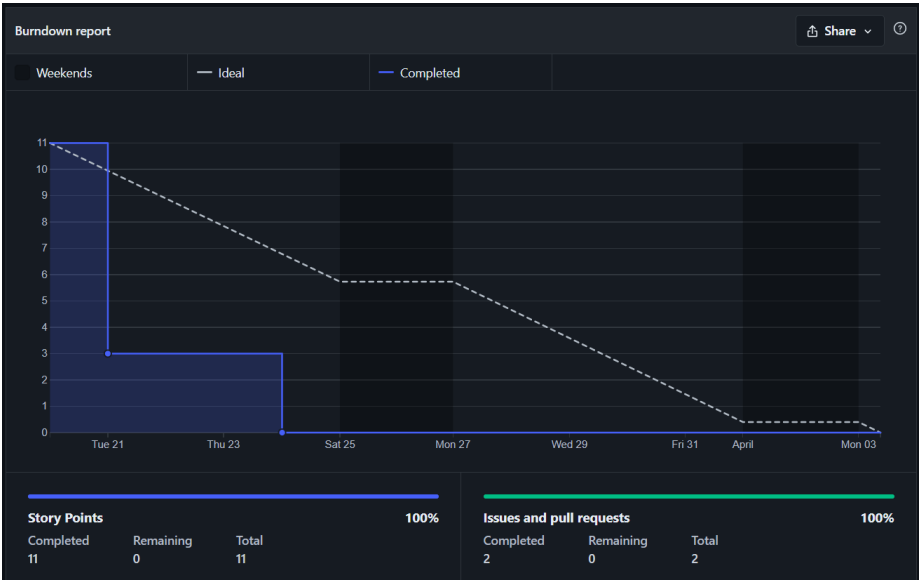


Figura A.8: Sprint 9: Burndown chart

Sprint 10

Sprt.	Duración	Objetivos	Tareas
10	03/04/2023 17/04/2023	Procesamiento de los datos, creación de los modelos de predicción y diseño de interfaces.	<ul style="list-style-type: none">■ Normalizar listados JCR.■ Comparar JCR obtenidos con los valores reales.■ Modelos de Crossvalidation.■ Diseño de las interfaces de la aplicación.■ Diagramas UML de la API.

Tabla A.10: Tareas del sprint 10



Figura A.9: Sprint 10: Burndown chart

Sprint 11

Sprt.	Duración	Objetivos	Tareas
11	24/04/2023 15/05/2023	Creación de un prototipo de aplicación web y exportación de los modelos de predicción	<ul style="list-style-type: none">■ Prototipo de API■ API con manejo de errores■ API con conexión a BBDD■ API con interfaz de prueba■ API con internacionalización■ Generar ficheros binarios para los modelos de predicción.

Tabla A.11: Tareas del sprint 11

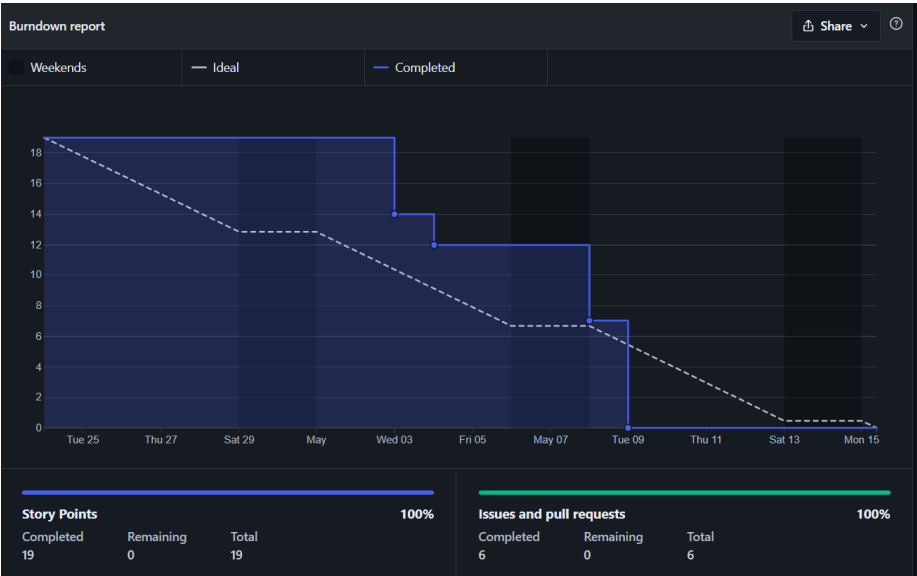


Figura A.10: Sprint 11: Burndown chart

Sprint 12

Sprt.	Duración	Objetivos	Tareas
12	15/05/2023 22/05/2023	Aplicación con todas las funcionalidades básicas en base al prototipo y corrección de la carga de información en la BBDD.	<ul style="list-style-type: none">■ Funcionalidades básicas de la API■ Cargar datos correctos

Tabla A.12: Tareas del sprint 12

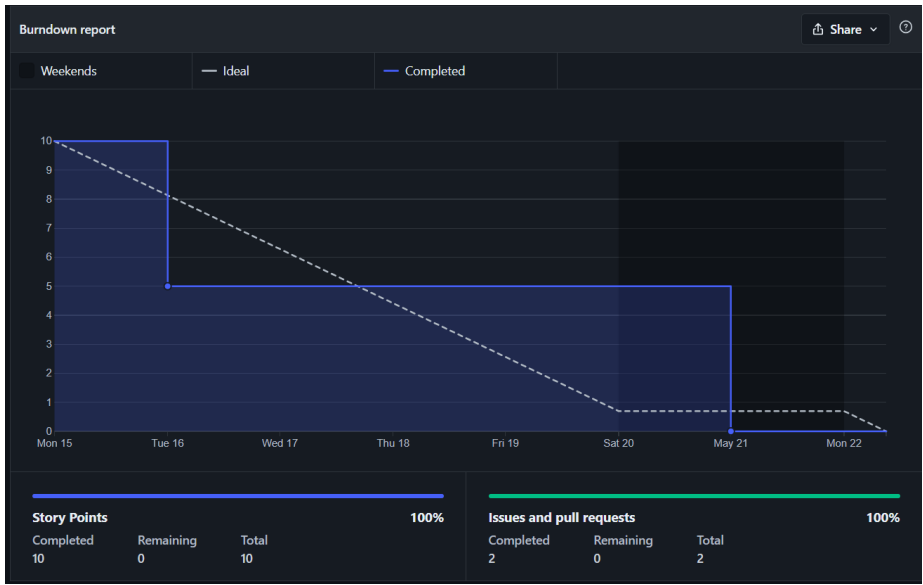


Figura A.11: Sprint 12: Burndown chart

Sprint 13

Sprt.	Duración	Objetivos	Tareas
13	15/05/2023 22/05/2023	Lanzar la aplicación en Heroku y mejora de funcionalidades.	<ul style="list-style-type: none">■ Serie temporal■ Añadir el cuartil a las revistas■ Configurar API en Heroku■ Configurar BBDD en Heroku

Tabla A.13: Tareas del sprint 13

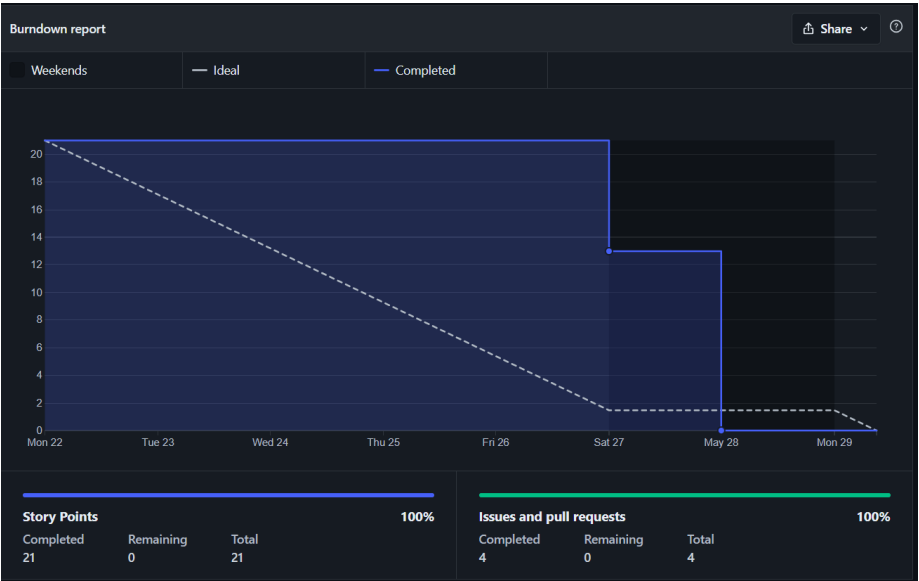


Figura A.12: Sprint 13: Burndown chart

Sprint 14

Sprt.	Duración	Objetivos	Tareas
14	29/05/2023 22/05/2023	Último sprint. Retoques finales y solución de bugs	<ul style="list-style-type: none">■ Solucionar login■ Obtener cuartiles■ Seguridad de contraseñas■ Documentación■ Bugs

Tabla A.14: Tareas del sprint 14

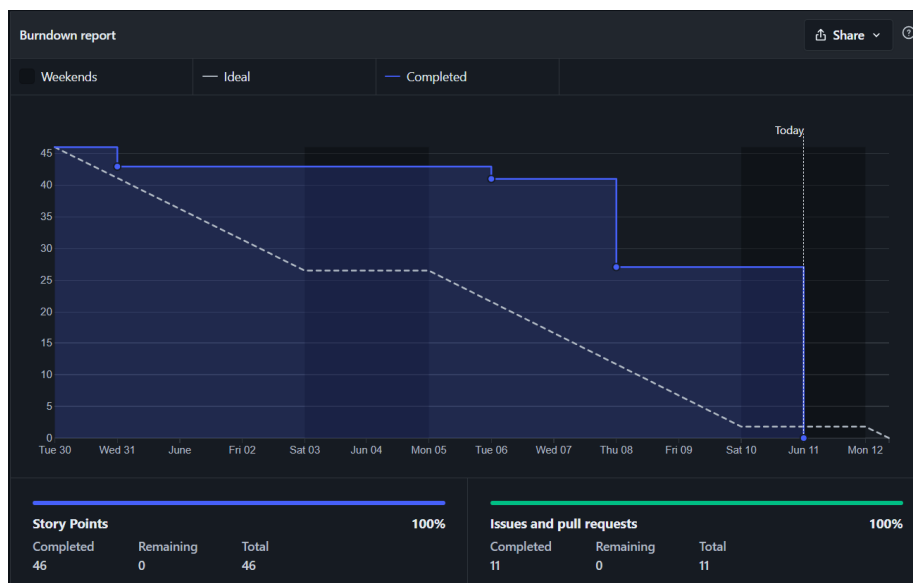


Figura A.13: Sprint 14: Burndown chart

Overview

Finalmente, se incluye un gráfico acumulativo³ (ver figura A.14) que muestra la cantidad total de trabajo realizado de forma más resumida.

El eje horizontal representa el período de tiempo, mientras que el eje vertical indica la cantidad acumulada de tareas completadas. A medida que avanza el tiempo, se puede observar cómo se han completado gradualmente las tareas y cómo ha progresado el proyecto. Además, las diferentes líneas de colores representan las diferentes etapas o estados de las tareas en un proyecto. Cada color corresponde a un estado específico, como «Nuevas tareas», «En progreso», «Completado», «Cerradas», etc. Estas líneas muestran cómo se distribuyen las tareas a lo largo del tiempo y cómo evolucionan desde su creación hasta su finalización. Al observar el gráfico, se puede identificar fácilmente la proporción de tareas en cada estado y cómo cambia esa proporción a medida que progresa el proyecto.

³Gráfico acumulativo: en este gráfico, el eje vertical representa la cantidad total de trabajo completado, mientras que el eje horizontal representa el tiempo transcurrido durante el proyecto. A medida que se completa el trabajo, la línea del gráfico se eleva, lo que indica el progreso general del proyecto

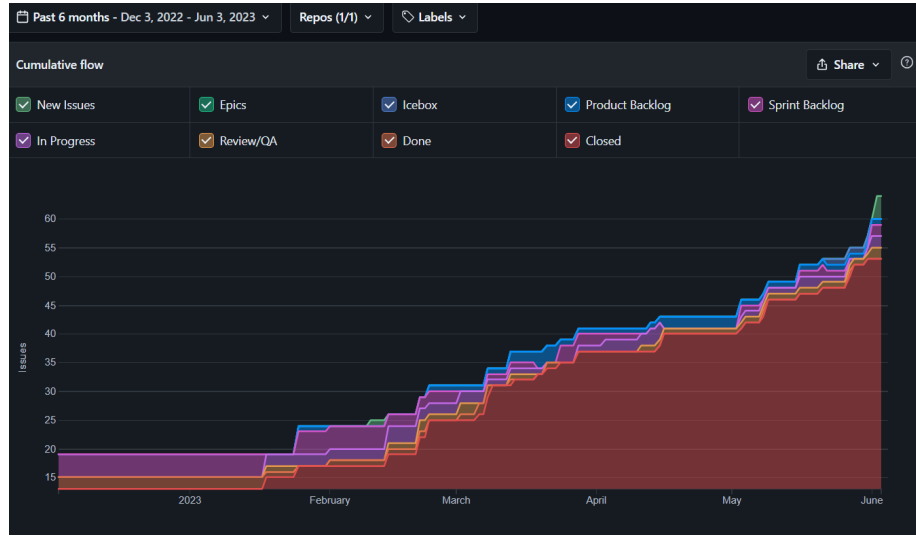


Figura A.14: Gráfico acumulativo de todo el proyecto

A.5. Estudio de viabilidad

El estudio de viabilidad es un componente esencial en el proceso de evaluación de cualquier proyecto o iniciativa. En el contexto de nuestra investigación, se ha desarrollado una aplicación *open source* destinada a beneficiar y servir a la comunidad científica.

Dado que la intención no es obtener beneficios económicos directos de este proyecto, se considerará un enfoque empresarial realista para asegurar la viabilidad y sostenibilidad a largo plazo de nuestro proyecto.

En las siguientes secciones, se analizarán en detalle los aspectos de viabilidad económica (sección A.5) y legal (sección A.5), con el objetivo de establecer un marco sólido que asegure la continuidad y éxito del proyecto.

Viabilidad económica

En primer lugar, la viabilidad económica se refiere a la evaluación de los recursos financieros necesarios para llevar a cabo el proyecto y asegurar su sostenibilidad a largo plazo. Aunque nuestra aplicación no tiene como objetivo obtener ganancias, es importante considerar los costos asociados al desarrollo, mantenimiento y mejora continua de la plataforma. Además, se deben analizar posibles fuentes de financiación, como subvenciones, donaciones o colaboraciones con instituciones interesadas en respaldar este tipo de iniciativas científicas abiertas.

Algunos aspectos relevantes para poder hacer estos cálculos son los siguientes:

- La empresa se situaría en España, lo que implica tener en cuenta las regulaciones y requisitos legales y fiscales del país.
- El proyecto tiene una duración estimada de 8 meses y una semana, desde octubre hasta la segunda semana de junio. Esto equivale aproximadamente a 34 semanas, considerando una duración promedio de 4 semanas por mes.
- El equipo encargado del proyecto está formado por una desarrolladora (la alumna), así como por el *Product Owner* y el *Scrum Master*, quienes actúan como tutores de la alumna en el desarrollo del proyecto.

Para todo esto, es fundamental considerar los costes y beneficios asociados al proyecto. Los **costes** se refieren a los desembolsos monetarios necesarios para llevar a cabo la iniciativa, incluyendo los gastos de inversión, los costes operativos y los costes de mantenimiento [11]. Por otro lado, los **beneficios** se refieren a las ganancias económicas que se esperan obtener como resultado de la implementación del proyecto. Estos beneficios pueden incluir ingresos generados por la venta de productos o servicios, ahorros en costes operativos, incremento en la productividad, entre otros [11].

Costes

Comenzaremos por calcular los costes totales de la supuesta empresa.

1. Costes de los empleados.

Se procederá al cálculo del salario bruto de cada integrante del equipo de acuerdo con la normativa laboral vigente en España para un proyecto de desarrollo de *software*. El salario bruto se determina considerando factores como la categoría profesional y la experiencia de cada empleado.

- a) Desarrollador: en este caso, la alumna es considerada una programadora *junior*. En España, un programador *junior* (con menos de 3 años de experiencia laboral) puede esperar un salario medio de alrededor de 19 700 € brutos por año [9]. Además, según el artículo 34 del Estatuto de los Trabajadores, la duración fijada en convenio colectivo de la jornada laboral es un máximo de 40

horas de trabajo efectivo [4]. Ahora, para calcular el salario bruto mensual, se puede utilizar la siguiente fórmula⁴:

$$\text{Salario Bruto Mensual} = \frac{\text{Salario Bruto Anual}}{14 \text{ pagas}}$$

Sustituyendo el valor del salario bruto anual, obtenemos:

$$\text{Salario Bruto Mensual} = \frac{19\,700 \text{ €}}{14 \text{ pagas}} \approx 1\,407,14 \text{ €}$$

- b) Product Owner: el salario bruto promedio anual para un Product Owner en España es de 41.866 € [5]. Utilizando la misma fórmula para calcular el salario bruto mensual, obtenemos:

$$\text{Salario Bruto Mensual} = \frac{41\,866 \text{ €}}{14 \text{ pagas}} \approx 2\,990,43 \text{ €}$$

- c) Scrum Master: el salario bruto promedio anual para un Scrum Master en España es de 40.626 € [6]. Utilizando la misma fórmula para calcular el salario bruto mensual, obtenemos:

$$\text{Salario Bruto Mensual} = \frac{40\,626 \text{ €}}{14 \text{ pagas}} \approx 2\,901,85 \text{ €}$$

Una vez obtenido el salario bruto, se deben aplicar los impuestos correspondientes de acuerdo con la legislación fiscal vigente. Los impuestos a tener en cuenta incluyen las contribuciones a la Seguridad Social [7], el impuesto sobre la renta⁵ (IRPF) [1] y otros impuestos relacionados con la contratación laboral. La información detallada sobre los impuestos y las obligaciones legales se encuentra en los documentos oficiales y normativas correspondientes proporcionados por las autoridades fiscales y laborales citados anteriormente.

De forma resumida, Para el año 2023, se han establecido los siguientes porcentajes⁶, los cuales se detallan en la Tabla A.15 y Tabla A.16.

Con base en los cálculos previos de los salarios brutos y los impuestos aplicados, se puede determinar el coste total de la empresa en cuanto a los empleados.

⁴Entendemos habitualmente 14 pagas (12 meses y 2 pagas extra).

⁵En cuanto al IRPF, el porcentaje aplicado puede variar según el nivel de ingresos y la situación fiscal personal. Según la Agencia Tributaria de España, se pueden encontrar las tablas y los porcentajes correspondientes en su documentación oficial [2].

⁶El gobierno pone a disposición la plataforma ipyme [8], que proporciona información valiosa para facilitar el cálculo de los tramites empresariales.

Concepto	Descripción	Porc. (%)
Cuota Contingencias Comunes	Gastos comunes en el ámbito laboral	23,60
Cuota Formación Profesional	Inversión en formación y capacitación	0,60
Desempleo	Prestaciones por desempleo	5,50
Accidentes de Trabajo	Prevención y compensación por accidentes laborales	1,50
FOGASA	Fondo de Garantía Salarial	0,20

Tabla A.15: Conceptos de Seguridad Social

Concepto	Descripción	Porcentaje
IRPF	Impuesto sobre la Renta de las Personas Físicas	Variable
IVA	Impuesto sobre el Valor Añadido (tipo general)	21,00 %

Tabla A.16: Otros conceptos

- a) Desarrollador: Dado que el salario bruto mensual de la desarrolladora es de aproximadamente 1 641,67 €, se aplicarán los porcentajes correspondientes de los conceptos de la Seguridad Social según lo establecido en la Tabla A.15. El total de los conceptos de Seguridad Social se calcula sumando los porcentajes aplicados al salario bruto mensual:

$$\text{Total SS} = 0,236 + 0,006 + 0,055 + 0,015 + 0,002 \approx 0,314$$

El IRPF es un impuesto variable y su porcentaje de retención en nómina depende del nivel de ingresos y la situación fiscal personal. Por lo tanto, no se tendrá en cuenta para el resultado de este cálculo.

Por último, el gasto de la empresa se calcula dividiendo el salario bruto mensual entre 1 menos el total de la Seguridad Social. Sustituyendo los valores correspondientes, obtenemos:

$$\text{Gasto mensual} = \frac{1\,407,14\text{ €}}{1 - 0,314} \approx 2\,051,22\text{ €}$$

- b) Product Owner: Como el salario bruto anual para el Product Owner es de 3 488,83 €, aplicando los porcentajes de la Tabla A.15, el gasto total de la empresa para el Product Owner sería:

$$\text{Salario mensual} = \frac{2\,990,43\text{ €}}{1 - 0,314} \approx 4\,359,22\text{ €}$$

De las 154 horas de la jornada mensual (22 días laborables por 7 horas al día), hemos calculado una dedicación del 7 % de su tiempo. Por tanto, se va aplicar este 7 % al gasto mensual del Product Owner y del Scrum Master. Se ha estimado una dedicación de 2 horas de reuniones semanales más 3 horas repartidas a lo largo del mes.

$$\text{Gasto mensual} = 4\,359,22\text{ €} \times 0,07 \approx 305,13\text{ €}$$

- c) Scrum Master: De manera similar, para el Scrum Master, partimos de un salario bruto mensual aproximado de 3 385,50 €. Aplicando los porcentajes de la Tabla A.15, el gasto total de la empresa para el Scrum Master sería:

$$\text{Salario mensual} = \frac{2\,901,85\text{ €}}{1 - 0,314} \approx 4\,230,10\text{ €}$$

Igualmente, al Scrum Master se le aplica un coeficiente reductor del 7 % de su salario, dado que ha invertido el mismo tiempo que el Product Owner.

$$\text{Gasto mensual} = 4\,230,10\text{ €} \times 0,07 \approx 296,11\text{ €}$$

A modo de resumen, se agrupa el resultado de todos los cálculos en la Tabla A.17.

2. Costes de hardware.

Para calcular los costes de *hardware*, consideraremos únicamente el portátil de la alumna. Se trata de un MSI, modelo Prestige 15 A12UD-053XES con un procesador 12th Gen Intel(R) Core(TM) i7-1280P 2.00

Empleado	Gasto Mensual (€)
Desarrollador	2 051,22
Product Owner	305,13
Scrum Master	296,11
Total Proyecto	2 652,46

Tabla A.17: Resumen de costes del personal

GHz y una RAM de 32 GB. Este portátil se encuentra en el mercado a un precio de 1 211,80 €.

El coste de amortización del portátil de la alumna se basará en su vida útil estimada. Supongamos que el portátil tiene una vida útil de 4 años. Para calcular el coste anual de amortización, dividiremos el coste del portátil entre el número de años de vida útil:

$$\text{Coste anual de amortización} = \frac{1\,211,80\text{ €}}{4\text{ años}} \approx 302,95\text{ €/año}$$

Por tanto, como el proyecto ha durado 8 meses, los costes de *hardware* suponen 302,95 por 66,66 % (proporción de los 8 meses sobre un año).

$$\text{Costes de hardware} = 302,95\text{ €} \times 66,66\% \approx 201,95\text{ €}$$

3. Costes de *software*.

Comenzaremos por contemplar las licencias correspondientes a los sistemas operativos usados.

- Windows 10 Pro: el valor aproximado de la licencia de Windows 10 Pro versión 22H2 es de 200 €. Al igual que en el caso del *hardware*, consideraremos una vida útil de 4 años y aplicaremos la amortización correspondiente.

$$\text{Coste anual de amortización} = \frac{200\text{ €}}{4\text{ años}} \approx 50\text{ €/año}$$

Aplicando de nuevo el 66,66 % del porcentaje anual que ha supuesto nuestro proyecto (8 meses), tenemos lo siguiente:

$$\text{Costes de licencia Windows} = 50 \text{ €} \times 66,66 \% \approx 33,33 \text{ €}$$

- Linux Debian: el portatil tiene una segunda partición con un sistema Debian que, debido a su filosofía de «*free software*» [3], es gratuito y no supone costes adicionales.

El resto del *software* utilizado está disponible de forma gratuita o puede ser usado gracias a las licencias de estudiantes que ofrece la universidad, por lo que no ha generado costes adicionales. Suponiendo que finaliza el plazo de estas licencias gratuitas, debemos considerar los siguientes costes:

- Heroku: se trata de un servidor de aplicaciones en la nube. Actualmente, se está utilizando una licencia gratuita de Heroku, gracias al *Student Pack* de GitHub proporcionado por la universidad. Sin embargo, una vez que la licencia gratuita expire, se deberá considerar el coste mínimo de los planes de Heroku. A saber, el coste mínimo para el Dyno básico (web Gunicorn API.paperrank.app:app) es de 7,00 \$ (que equivale aproximadamente a 6,53 € mensuales). Además, será necesario tener en cuenta el coste adicional para disponer (al menos) de la versión mini de Heroku Postgres. Este coste es de 5,00 \$ (4,66 € mensuales).
- GitHub y extensión de ZenHub: La licencia de GitHub (versión empresarial) tiene un coste de 17,95 € mensuales. Además, la extensión de ZenHub tiene un coste adicional de 12 €.

Se adjunta una tabla resumen con los costes de *software* estimados de forma mensual, considerando los elementos mencionados anteriormente (ver Tabla A.18).

Para obtener los gastos totales del proyecto, simplemente se multiplica el costo mensual por el número de meses:

$$\text{Gastos totales} = 45,31 \text{ €} \times 8 \text{ meses} \approx 362,48 \text{ €}$$

Finalmente, aunque durante el desarrollo del proyecto se ha decidido no utilizar las licencias de las APIs de Google Scholar, Web of Science y Scopus, es importante tener en cuenta que en futuras investigaciones o aplicaciones más complejas, podría considerarse la necesidad de

Concepto	Descripción	Coste Mensual (€)
Windows 10 Pro	Licencia software	4,17
Heroku Dyno web	<i>Cloud server</i>	6,53
Heroku Postgres	<i>Cloud server</i>	4,66
GitHub Enterprise	Licencia software	17,95
ZenHub	Extensión GitHub	12,00
Total (mensual)		45,31

Tabla A.18: Costes de software mensuales

acceder a estas fuentes de información científica y académica. Estas APIs proporcionan un acceso especializado a una amplia gama de datos y recursos, lo cual podría requerir una inversión adicional en forma de licencias y sus correspondientes costos. Es recomendable evaluar detenidamente los requisitos del proyecto y el presupuesto disponible antes de tomar la decisión de adquirir estas licencias, en caso de que se considere necesario para la viabilidad y calidad de la aplicación.

Para visualizar los costes de forma clara, se ha elaborado una última tabla resumen con todos los costes finales (ver Tabla A.19).

Concepto	Descripción	Coste Total (€)
Personal	Costes totales de empleados	12 416,43
Hardware	Costes por el portátil	302,70
Software	Costes totales por licencias	402,75
Coste Total del Proyecto		13 121,88

Tabla A.19: Resumen de costes del proyecto

Beneficios

Aunque la aplicación no tiene fines lucrativos y estará disponible de forma gratuita para la comunidad científica, es importante considerar una

forma hipotética de generar ingresos para sostener el proyecto. Una opción sería incluir anuncios en la página web de la aplicación.

La inclusión de anuncios publicitarios podría proporcionar una fuente potencial de ingresos. Sin embargo, es fundamental analizar las implicaciones y opciones disponibles antes de tomar una decisión. En primer lugar, se debe tener en cuenta que la incorporación de anuncios puede afectar la experiencia del usuario y la usabilidad de la aplicación. Es importante encontrar un equilibrio entre la generación de ingresos y la satisfacción de los usuarios.

Una opción es utilizar redes de publicidad en línea, como Google AdSense, que permiten mostrar anuncios relevantes y personalizados en la página web. Estas redes se encargan de gestionar los anuncios y los ingresos generados por ellos, ofreciendo a los propietarios de sitios web una parte de los ingresos publicitarios.

Otra posibilidad es buscar acuerdos directos con empresas o instituciones que deseen promocionar sus productos o servicios en la aplicación. Esto podría implicar la negociación de contratos publicitarios personalizados y la integración de anuncios específicos en la página web.

Es importante tener en cuenta que la inclusión de anuncios publicitarios requerirá un trabajo adicional en términos de diseño, implementación y gestión de la publicidad. Se deberá evaluar la viabilidad económica de los ingresos generados por los anuncios en comparación con los costes asociados. Es fundamental también considerar las implicaciones éticas y de privacidad asociadas con la publicidad en línea. Se deben seguir las políticas y regulaciones vigentes para garantizar la transparencia, el respeto de la privacidad de los usuarios y la integridad del proyecto.

Viabilidad legal

Por otro lado, la viabilidad legal implica evaluar las cuestiones jurídicas y normativas que puedan afectar la operación y distribución de la aplicación *open source*. A pesar de que nuestro proyecto está destinado a ser utilizado por la comunidad científica en general, es necesario asegurarnos de cumplir con las regulaciones y requisitos legales relacionados con la privacidad, protección de datos y derechos de propiedad intelectual.

Web Scraping

La viabilidad legal de la aplicación es un aspecto crucial a considerar, especialmente cuando se utiliza una técnica como *web scraping* para obtener

información de fuentes externas. Es importante comprender los aspectos legales y los posibles problemas que pueden surgir en relación con los permisos y derechos de propiedad intelectual (especialmente en los sitios web que requieren registro de usuario).

En el caso específico de la aplicación, se realiza *web scraping* utilizando la API de Crossref (<https://api.crossref.org>), por lo que el primer paso es revisar y comprender los términos de servicio y las políticas de esta API. Estos documentos pueden establecer las condiciones específicas para el uso de la API, incluyendo cualquier restricción o requisito legal. Es recomendable revisar las políticas de Crossref regularmente para asegurarse de cumplir con los términos actualizados.

Según la página oficial, Crossref ofrece su API de REST públicamente [15], lo que permite a los usuarios acceder y utilizar los metadatos depositados por sus miembros. Crossref proporciona dichos metadatos sin restricciones de *copyright* y los usuarios pueden utilizarlos para cualquier propósito. En este sentido, nuestro proyecto no supone ningún problema legal (lo cual no sucedía con los sitios de Google Scholar, Scopus y Web of Science).

Sin embargo, es importante tener en cuenta que algunos *abstracts* contenidos en los metadatos pueden estar sujetos a derechos de autor por parte de los editores o autores originales. Por lo tanto, es esencial que al utilizar la API de Crossref se respeten los derechos de autor y se cumpla con los términos y condiciones establecidos tanto por Crossref como por sus editores y autores originales. Además, Crossref ha implementado un sistema de «piscinas» separadas para usuarios «educados» que incluyen información de contacto y respetan ciertas prácticas de acceso.

Licencias de software

Dado que en el presente proyecto se han utilizado una variedad de herramientas y paquetes con diferentes licencias (ver Tabla A.20), es fundamental tenerlas en cuenta al elegir una licencia para nuestro *software*. A continuación, se enumeran y explican las opciones de licencia más relevantes:

- **Licencia MIT:** esta licencia es ampliamente utilizada y permite el uso, copia, modificación y distribución del *software*, tanto en proyectos comerciales como no comerciales, con la condición de mantener el aviso de derechos de autor y la exención de responsabilidad. Es una opción flexible y abierta que fomenta la colaboración y la adopción del *software*.
- **Licencia BSD:** Las licencias BSD son flexibles y permiten la redistribución y el uso en proyectos comerciales y no comerciales. Estas licencias suelen requerir el mantenimiento del aviso de derechos de autor y la exención de responsabilidad. También pueden incluir cláusulas adicionales que limitan la utilización del nombre del titular de los derechos de autor en la promoción del *software* derivado.
- **Licencia Apache:** La Licencia Apache es una licencia de código abierto que permite el uso, modificación y distribución del *software* bajo ciertas condiciones. Estas condiciones incluyen el mantenimiento del aviso de derechos de autor, la exención de responsabilidad y la necesidad de proporcionar una copia de la licencia en cualquier distribución del *software*.
- **Licencia GNU:** La Licencia Pública General de GNU (GPL) es una licencia de *software* libre que garantiza a los usuarios la libertad de utilizar, estudiar, modificar y distribuir el *software*. La GPL también exige que cualquier *software* derivado se distribuya bajo los términos de la GPL, lo que garantiza que las mejoras y modificaciones sigan siendo de código abierto.

Después de considerar cuidadosamente las licencias de las herramientas utilizadas en nuestro proyecto y dada la naturaleza del mismo, se ha considerado que la más adecuada es la Licencia GNU. Esta licencia garantiza la libertad del *software* y el mantenimiento de su código abierto. Esta opción es ampliamente utilizada y respaldada por la comunidad del *software* libre y de código abierto.

Paquete	Versión	Licencia
Babel	2.12.1	MIT License
beautifulsoup4	4.12.2	MIT License
crossrefapi	1.5.0	CC Attribution 4.0 International License
cryptography	40.0.2	OSI Approved: Apache SW License, BSD License
Flask	2.3.2	BSD License
flask-babel	3.1.0	BSD 3 License
Flask-Cors	3.0.10	MIT License
Flask-Login	0.6.2	MIT License
Flask-Migrate	4.0.4	MIT License
frozenset	1.3.3	Apache Software License (Apache 2)
gunicorn	20.1.0	MIT License
habanero	1.2.3	Copyright (C) 2021 Scott Chamberlain
httpcore	0.17.2	BSD License (BSD)
httpx	0.24.1	BSD License
idna	3.4	BSD License
imagesize	1.4.1	MIT License
ipykernel	6.23.1	BSD License (BSD 3-Clause License)
ipython	8.13.2	BSD License (BSD-3-Clause)
jedi	0.18.2	MIT License
Jinja2	3.1.2	BSD-3-Clause License
joblib	1.2.0	BSD License (BSD)
jupyter_client	8.2.0	BSD License (BSD 3-Clause License)
jupyter_core	5.3.0	BSD License (BSD 3-Clause License)
matplotlib	3.7.1	PSF
matplotlib-inline	0.1.6	BSD 3-Clause License
numpy	1.24.3	BSD License (BSD-3-Clause)
pandas	2.0.1	BSD License (BSD 3-Clause License)
pickleshare	0.7.5	MIT License
psycpg2	2.9.6	GNU Lesser General Public (LGPLv2)
requests	2.31.0	Apache Software License (Apache 2.0)
scholarly	1.7.11	Unlicensed
scikit-learn	1.2.2	BSD License (new BSD)
scipy	1.10.1	BSD License (Copyright (c) 2001-2002 Enthought,...)
selenium	4.9.1	Apache Software License (Apache 2.0)
SQLAlchemy	2.0.15	MIT License
xgboost	1.7.5	Apache Software License (Apache-2.0)
Bootstrap	5.2.3	MIT License
D3js	7.8.4	ISC
jQuery	3.6.4	MIT License
DataTables	1.13.4	MIT License
pytest	7.2.0	MIT License
Chartjs	4.0	MIT License

Tabla A.20: Información de los paquetes y sus licencias

Apéndice B

Especificación de Requisitos

B.1. Introducción

Para llevar a cabo el proyecto adecuadamente, se realizará un análisis exhaustivo de los requisitos funcionales y no funcionales de la aplicación. En la sección de Objetivos generales (B.2) se establecerán los resultados esperados y los criterios de éxito del proyecto. Seguidamente, se presentará el catálogo de requisitos (sección B.3), detallando las funcionalidades y características que la aplicación debe cumplir. Finalmente, se procederá a la especificación de requisitos (sección B.5), describiendo cada uno de los identificados en el catálogo y estableciendo su prioridad, complejidad y dependencias.

A través de este informe, se pretende sentar las bases y establecer las directrices necesarias para el desarrollo exitoso de la aplicación web que proporcionará a los usuarios la información clave sobre el impacto de las revistas científicas, con el objetivo de facilitar la toma de decisiones informadas en el ámbito de la publicación académica.

B.2. Objetivos generales

Los objetivos generales de este proyecto son los siguientes:

1. Realizar el proceso de extracción de datos relevantes para el cálculo del Índice de Impacto de las revistas científicas indexadas en el Journal Citation Reports (JCR) de Clarivate. Esta tarea constituye la parte

más desafiante del proyecto, ya que implica llevar a cabo una investigación exhaustiva y realizar pruebas utilizando diversas técnicas de *web scraping* y simulación con Selenium. Se requerirá desarrollar métodos eficientes para recopilar los datos necesarios que no estén fácilmente accesibles y estructurados.

2. Implementar modelos de predicción utilizando técnicas de *machine learning*, en particular utilizando la biblioteca **Scikit-learn**. Estos modelos permitirán predecir el Índice de Impacto de las revistas científicas en base en las características y métricas disponibles. Se explorarán diferentes algoritmos y técnicas para elegir el que mejor se ajuste a los datos y así obtener predicciones precisas y confiables
3. Desarrollar una aplicación web utilizando el *framework* Flask. Esta aplicación será la interfaz principal para que los usuarios puedan acceder y visualizar los resultados del proyecto. La aplicación web proporcionará una experiencia de usuario intuitiva y amigable, permitiendo la búsqueda y filtrado de revistas, así como la visualización de los índices de impacto calculados y las predicciones generadas por los modelos.

El objetivo central del proyecto es proporcionar a los usuarios una herramienta integral que les permita evaluar y comparar el impacto de las revistas científicas de manera eficiente y precisa. En particular, se pretende proporcionar valores del Índice de Impacto de años anteriores, y la predicción para el año en curso. Para lograr esto, se abordarán tres aspectos clave: la extracción de datos, los modelos de predicción y el desarrollo de la aplicación web. Cada uno de estos objetivos se complementa para lograr un sistema funcional y útil para la comunidad académica y científica.

Es importante reiterar que el proceso de extracción de datos es considerado como la parte más desafiante del proyecto debido a la necesidad de investigar y probar diferentes técnicas de *web scraping*. El éxito en esta etapa es fundamental para garantizar la disponibilidad de los datos necesarios para el cálculo del Índice de Impacto y para el entrenamiento de los modelos de predicción.

B.3. Catálogo de Historias de Usuario

En este caso, dado que se ha elegido seguir una metodología ágil, se considera un catálogo de historias de usuario (usadas en ambos marcos,

Scrum y Kanban). Para poder redactar una historia de usuario, hay tres elementos clave [10]:

- Perfil: El rol del usuario final.
- Necesidad: El objetivo que tiene la función de *software* para el usuario final.
- Propósito: El objetivo de la experiencia del usuario final con la función de *software*.

Habiendo identificado estos elementos, la historia de usuario se redactará siguiendo este formato: «**Como [perfil], quiero [necesidad], para lograr [propósito]**».

Se ha elaborado el catálogo de historias de usuario (HU) con el objetivo de cubrir todos los objetivos mencionados en el apartado anterior sin entrar en detalles (los requisitos más específicos con respecto a la aplicación web están definidos en la sección B.4). Para visualizar mejor el catálogo, se han clasificado las historias de usuario en tres categorías: extracción de datos (Tabla B.1), modelos de predicción (Tabla B.2) y aplicación web (Tabla B.3).

Identificador	Descripción
HU1	Como desarrollador, quiero obtener información relevante para el cálculo del Índice de Impacto (JCR) de las revistas científicas a partir de fuentes externas.
HU2	Como desarrollador, quiero aplicar técnicas de <i>web scraping</i> y simulación con Selenium para extraer los datos de manera eficiente y precisa.
HU3	Como desarrollador, quiero procesar y almacenar los datos extraídos en una base de datos para su posterior uso en el cálculo del Índice de Impacto y en los modelos de predicción.

Tabla B.1: HU Extracción de datos

Identificador	Descripción
HU4	Como desarrollador, quiero implementar modelos de predicción utilizando técnicas de <i>machine learning</i> (scikit-learn) para predecir el Índice de Impacto de las revistas científicas.
HU5	Como desarrollador, quiero evaluar y comparar diferentes algoritmos de <i>machine learning</i> para seleccionar los modelos más adecuados.
HU6	Como desarrollador, quiero entrenar los modelos utilizando los datos almacenados y ajustar sus parámetros para obtener predicciones precisas.

Tabla B.2: HU Modelos de predicción

Identificador	Descripción
HU7	Como usuario, quiero acceder a una aplicación web donde pueda buscar y filtrar revistas científicas para evaluar su Índice de Impacto.
HU8	Como usuario, quiero visualizar de forma clara y comprensible los índices de impacto calculados y las predicciones generadas por los modelos de predicción.
HU9	Como usuario, quiero interactuar con la aplicación web, seleccionar revistas específicas y obtener información detallada sobre cada una de ellas.

Tabla B.3: HU Aplicación web

Estas historias de usuario representan las necesidades y objetivos de los diferentes perfiles de usuarios en relación con la extracción de datos, los modelos de predicción y la aplicación web. Siguiendo una metodología ágil, estas historias de usuario servirán como guía para el desarrollo incremental y la entrega de funcionalidades de valor en cada iteración del proyecto.

B.4. Requisitos de la aplicación

En esta sección se enumerarán, en mayor detalle, los requisitos que se esperan de la aplicación web en concreto.

Requisitos funcionales

- **RF1:** Autenticación y Registro
 - **RF1.1:** La aplicación debe permitir a los usuarios autenticarse mediante un nombre de usuario y contraseña válidos.
 - **RF1.2:** La aplicación debe permitir a los usuarios registrarse proporcionando un nombre de usuario, una contraseña, una dirección de correo electrónico válida y confirmación de la contraseña.
 - **RF1.3:** Las contraseñas deben tener un formato válido (i.e.: una longitud mínima de 8 caracteres, caracteres especiales requeridos, etc.).
 - **RF1.4:** Los cuadros de texto de las contraseñas deben tener la opción de mostrar y ocultar el contenido por motivos de seguridad.
- **RF2:** Búsqueda y selección de revistas
 - **RF2.1:** En la interfaz principal debe mostrar un botón que redirija a otra interfaz con una lista de revistas y con un listado de revistas y la información más relevante sobre las mismas
 - **RF2.2:** La interfaz con la lista de revistas debe incluir una barra de búsqueda por texto y paginación para filtrar las revistas.
 - **RF2.3:** La interfaz principal debe permitir a los usuarios seleccionar la categoría y el nombre de la revista utilizando listas *dropdowns* con búsqueda por texto (activándose el botón para consultar el histórico del JCR de la revista).
 - **RF2.4:** La interfaz principal debe permitir a los usuarios seleccionar uno o más modelos de predicción (activándose el botón de predicción del JCR).
 - **RF2.5:** En la interfaz principal se deben validar los campos obligatorios y mostrar un mensaje de error en caso de que no se completen.

- **RF3:** Interfaz de Historial JCR
 - **RF3.1:** La interfaz de Historial JCR debe mostrar un gráfico con el JCR de la revista seleccionada en los últimos 5 años.
 - **RF3.2:** Asimismo, debe mostrar un gráfico con el cuartil de la revista seleccionada en los últimos 5 años.
- **RF4:** Interfaz de Predicción JCR
 - **RF4.1:** La interfaz de Predicción JCR debe permitir a los usuarios seleccionar uno o más modelos de predicción.
 - **RF4.2:** Debe mostrar la predicción del JCR de cada modelo para el año actual.
 - **RF4.3:** Debe mostrar un gráfico con la evolución temporal del JCR de la revista y la predicción de los dos últimos años en diferentes colores para cada modelo.
 - **RF4.4:** Los usuarios deben poder ocultar cualquier línea del gráfico haciendo clic sobre su leyenda.
- **RF5:** Barra inferior y superior de las interfaces
 - **RF5.1:** La barra inferior debe incluir enlaces a la política de privacidad y a los términos de uso de la aplicación.
 - **RF5.2:** La barra inferior debe incluir el logo de la Universidad de Burgos que redirige a su página web oficial.
 - **RF5.3:** La barra superior debe mostrar el logo de la aplicación que redirige siempre a la página principal.
 - **RF5.4:** La aplicación debe tomar el idioma configurado en el navegador o el idioma «más cercano» disponible por defecto. Además, la barra superior debe incluir un menú *dropdown* con opciones de internacionalización en español, inglés, italiano y francés.
 - **RF5.5:** La barra superior debe mostrar un enlace a la ayuda en línea.
 - **RF5.6:** La barra superior debe mostrar el nombre e imagen del usuario que ha iniciado sesión.
 - **RF5.7:** El menú *dropdown* del nombre del usuario debe permitir cerrar sesión (eliminando las variables de sesión y redirigiendo a la interfaz de inicio de sesión) o editar el perfil (redirigiendo

a una página donde se pueda consultar y modificar la información del usuario).

- **RF6:** Navegación
 - **RF6.1:** Todas las interfaces deben tener navegación en forma de *breadcrumbs* para indicar la ubicación actual del usuario en la aplicación.

Requisitos No Funcionales

- **RNF1:** Seguridad
 - **RNF1.1:** Las contraseñas deben ser almacenadas de forma segura utilizando técnicas de cifrado adecuadas.
 - **RNF1.2:** La aplicación debe protegerse contra ataques de fuerza bruta en los formularios de autenticación y registro.
- **RNF2:** Rendimiento
 - **RNF2.1:** La aplicación debe ser capaz de manejar de manera eficiente la carga de usuarios concurrentes.
 - **RNF2.2:** Las consultas a la base de datos deben optimizarse para proporcionar respuestas rápidas a los usuarios.
- **RNF3:** Usabilidad
 - **RNF3.1:** La interfaz de usuario debe ser intuitiva y fácil de usar para los usuarios.
 - **RNF3.2:** Los mensajes de error deben ser claros y descriptivos, brindando orientación sobre cómo solucionar los problemas.
- **RNF4:** Mantenibilidad
 - **RNF4.1:** El código fuente de la aplicación debe estar bien estructurado, modularizado y documentado.
 - **RNF4.2:** El código debe seguir las mejores prácticas de desarrollo de *software* y ser fácilmente mantenible y escalable.

B.5. Especificación de requisitos

En esta sección se enumerarán los distintos casos de uso para la aplicación web desarrollada.

Actores

Antes de comenzar con los casos de uso, se identifican los distintos actores que pueden interactuar con la aplicación y se definen a continuación:

- **Usuario registrado:** Referido a una persona que ha creado una cuenta o perfil en la aplicación. Tendrá la posibilidad de modificar sus credenciales.
- **Administrador:** Usuario registrado con privilegios de administrador. Actualmente no tiene ninguna funcionalidad adicional, pero se ha dejado con vistas a futuras ampliaciones del *software*.

Casos de Uso

En cuanto a los casos de uso en sí mismos, se han identificado los siguientes:

- **CU-1** Inicio de sesión (ver Tabla B.4).
- **CU-2** Registro (ver Tabla B.5).
- **CU-3** Consultar revistas (ver Tabla B.6).
- **CU-4** Consultar histórico (ver Tabla B.7).
- **CU-5** Predecir JCR (ver Tabla B.8).
- **CU-6** Cerrar sesión (ver Tabla B.9).

Diagrama de casos de usos

Finalmente, para poder visualizar todos estos casos, se incluye un diagrama de casos de uso (ver Figura B.1).



Figura B.1: Diagrama de Casos de Uso

CU-1	Inicio de sesión
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF1, RF1.1, RF1.4
Descripción	Este caso de uso describe el proceso de autenticación de los usuarios en la aplicación.
Precondición	El usuario debe estar registrado en la aplicación.
Acciones	<ol style="list-style-type: none"> 1. El usuario introduce su nombre de usuario y contraseña (pudiendo mostrar y ocultar su contenido). 2. El sistema verifica la validez de las credenciales ingresadas. 3. El sistema otorga acceso al usuario si las credenciales son válidas.
Postcondición	El usuario ha iniciado sesión en la aplicación.
Excepciones	<ul style="list-style-type: none"> • Si las credenciales son inválidas, se muestra un mensaje de error y se solicita al usuario que vuelva a ingresar las credenciales.
Importancia	Alta

Tabla B.4: CU-1 Inicio de sesión

CU-2	Registro
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF1, RF1.3, RF1.4
Descripción	Este caso de uso describe el proceso de registro de nuevos usuarios en la aplicación.
Precondición	-
Acciones	<ol style="list-style-type: none"> 1. El usuario selecciona la opción de registro en la interfaz de inicio de sesión. 2. El usuario proporciona un nombre de usuario, una contraseña (pudiendo mostrar y ocultar su contenido), una dirección de correo electrónico válida y confirma la contraseña. 3. El sistema verifica que la contraseña tenga un formato válido y que la dirección de correo electrónico sea válida. 4. El sistema verifica que la contraseña y la confirmación de la contraseña coincidan. 5. El sistema crea un nuevo usuario en la base de datos con la información proporcionada. 6. El usuario es redireccionado a la interfaz de inicio tras recibir un mensaje de éxito.
Postcondición	El usuario se ha registrado exitosamente.
Excepciones	<ul style="list-style-type: none"> • Si la contraseña no cumple con el formato válido (i.e.: no tiene una longitud mínima de 8 caracteres, no contiene caracteres especiales requeridos, etc.), se muestra un mensaje de error y se solicita al usuario que ingrese una contraseña válida. • Si la dirección de correo electrónico no es válida, se muestra un mensaje de error y se solicita al usuario que ingrese una dirección válida. • Si la contraseña y la confirmación de la contraseña no coinciden, se muestra un mensaje de error y se solicita al usuario que vuelva a ingresar la contraseña y su confirmación.
Importancia	Alta

Tabla B.5: CU-2 Registro.

CU-3	Consultar revistas
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF2, RF2.2
Descripción	Este caso de uso describe la interfaz que muestra una lista de revistas y su información relevante, con funcionalidades de búsqueda y paginación.
Precondición	El usuario ha iniciado sesión y ha accedido a la interfaz principal.
Acciones	<ol style="list-style-type: none"> 1. El usuario accede a la interfaz de revistas pulsando sobre el icono correspondiente. 2. Se obtiene la lista completa de revistas de la base de datos. 3. La interfaz muestra una lista de revistas con su información relevante. 4. La interfaz incluye una barra de búsqueda por texto para filtrar las revistas en base a palabras clave. Muestra también una paginación para navegar entre las diferentes páginas de resultados.
Postcondición	-
Excepciones	<ul style="list-style-type: none"> • Si la lista de revistas no se puede conseguir exitosamente de la base de datos, se redirige al usuario a una página de error.
Importancia	Alta

Tabla B.6: CU-3 Consultar revistas

CU-4	Consultar histórico
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF2.3, RF3.1, RF3.2
Descripción	Este caso de uso describe el proceso para consultar el histórico del JCR de una revista en particular.
Precondición	El usuario ha iniciado sesión en la aplicación y está en la interfaz principal.
Acciones	<ol style="list-style-type: none"> 1. La interfaz principal muestra listas desplegables (<i>dropdowns</i>) con opciones de categorías y nombres de revistas. El usuario selecciona las opciones que desee. 2. El sistema valida los campos obligatorios y activa el botón. 3. El usuario pulsa el botón y es redirigido a la interfaz de consulta del JCR. 4. Se carga la interfaz con la información solicitada (muestra dos gráficos, uno con los cuartiles y otro con los valores del JCR).
Postcondición	-
Excepciones	<ul style="list-style-type: none"> • Si el usuario no selecciona ambos menús desplegables, el botón no se activa. Un <i>tooltip</i> aconsejará completar la información.
Importancia	Alta

Tabla B.7: CU-4 Consultar histórico

CU-5	Consultar predicción
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF2.4, RF4.1, RF4.2, RF4.3, RF4.4
Descripción	Este caso de uso permite a los usuarios consultar la predicción del JCR de una revista utilizando uno o más modelos de predicción.
Precondición	El usuario ha iniciado sesión en la aplicación y está en la interfaz principal. El usuario ya ha seleccionado la categoría y el nombre de la revista.
Acciones	<ol style="list-style-type: none"> 1. El usuario selecciona uno o más modelos de predicción en la interfaz principal. 2. El sistema valida que al menos un modelo está seleccionado y activa el botón. 3. El usuario hace clic sobre el botón. 4. Se muestra la interfaz de Predicción JCR con la predicción del JCR de cada modelo para el año actual. 5. El usuario puede ocultar cualquier línea del gráfico haciendo clic sobre su leyenda.
Postcondición	-
Excepciones	<ul style="list-style-type: none"> • Si el usuario no selecciona alguno de los modelos, el botón no se activa. Un <i>tooltip</i> aconsejará completar la información.
Importancia	Alta

Tabla B.8: CU-5 Consultar predicción.

CU-6	Cerrar sesión
Versión	1.0
Autor	Gadea Lucas Pérez
Requisitos asociados	RF1, RF2
Descripción	Este caso de abarca el proceso de cierre de sesión.
Precondición	El usuario ha iniciado sesión en la aplicación.
Acciones	<ol style="list-style-type: none">1. El usuario selecciona (en el <i>dropdown</i> de la esquina superior derecha) la opción «cerrar sesión».2. Se eliminan las variables de sesión para el usuario actual.3. El usuario es redireccionado a la interfaz de inicio.
Postcondición	El usuario ya no está <i>logueado</i>
Excepciones	-
Importancia	Alta

Tabla B.9: CU-6 Cerrar sesión

Apéndice C

Especificación de diseño

C.1. Introducción

El objetivo principal de esta especificación de diseño es brindar una guía clara y completa para el desarrollo de la aplicación web, proporcionando una base sólida sobre la cual los desarrolladores puedan implementar el sistema de manera efectiva. Asimismo, este informe servirá como referencia y documentación para futuras etapas del proyecto y posibles mejoras o expansiones.

Así pues, en esta sección se describirán y analizarán en detalle los diferentes aspectos del diseño de la aplicación web, abarcando desde la estructura y organización de los datos hasta los procedimientos y algoritmos que permiten su funcionamiento adecuado. Además, se profundizará en el diseño arquitectónico, que define la distribución de los componentes y la interacción entre ellos.

C.2. Diseño de los datos

Comenzaremos detallando el diseño de los datos de la aplicación. En este caso, para el almacenamiento de la información, se ha utilizado una base de datos PostgreSQL, la cual es accedida a través de la aplicación haciendo uso de la librería de Python *psycopg2*.

Diccionario de datos

A continuación, se describirán las tablas que componen la base de datos y se explicará la estructura y los atributos de cada una de ellas.

Tabla «Users»

Esta tabla almacena la información de los usuarios registrados en la aplicación. Los atributos de la tabla son los siguientes:

- **username:** campo de tipo `VARCHAR(255)` que guarda el nombre de usuario.
- **password:** campo de tipo `VARCHAR(255)` que almacena la contraseña del usuario.
- **email:** campo de tipo `VARCHAR(255)` que sirve como clave primaria y guarda la dirección de correo electrónico del usuario.
- **admin:** campo de tipo `BOOLEAN` que indica si el usuario tiene privilegios de administrador.

Tabla «Modelos»

En esta tabla se registran los modelos utilizados en la aplicación. Los atributos de la tabla son los siguientes:

- **id:** campo de tipo `SERIAL` que actúa como clave primaria y asigna un identificador único a cada modelo.
- **nombre:** campo de tipo `TEXT` que almacena el nombre del modelo.
- **rmse:** campo de tipo `FLOAT` que guarda el valor del error cuadrático medio asociado al modelo.

Tabla «Revista»

Esta tabla contiene la información de las revistas presentes en la aplicación. Los atributos de la tabla son los siguientes:

- **nombre:** campo de tipo `VARCHAR(255)` que sirve como clave primaria y almacena el nombre de la revista.
- **ISSN:** campo de tipo `VARCHAR(9)` que garantiza la unicidad y no nulidad del International Standard Serial Number de la revista.
- **categoria:** campo de tipo `VARCHAR(255)` que guarda la categoría a la que pertenece la revista.

Tabla «Revista_jcr»

Esta tabla registra las métricas JCR (Journal Citation Reports) asociadas a las revistas. Los atributos de la tabla son los siguientes:

- **nombre:** campo de tipo **VARCHAR(255)** que guarda el nombre de la revista.
- **fecha:** campo de tipo **NUMERIC** que indica la fecha de la métrica JCR.
- **jcr:** campo de tipo **FLOAT** que almacena el valor de la métrica JCR para la revista en la fecha correspondiente.
- **citas:** campo de tipo **NUMERIC** que indica la cantidad de citas recibidas por la revista en la fecha dada.
- **diff:** campo de tipo **FLOAT** que representa la diferencia en la métrica JCR real en comparación con el JCR calculado por el alumno (en la fecha dada).
- **cuartil:** campo de tipo **VARCHAR(3)** que indica el cuartil al que ha pertenecido la revista en la fecha dada.

Estas tablas y sus respectivos atributos han sido diseñados de manera coherente para almacenar y organizar la información necesaria en la aplicación web de una forma sencilla.

Optimización de la base de datos

Para mejorar el rendimiento y la eficiencia de las consultas en la base de datos, se han aplicado diferentes técnicas de optimización. Una de estas técnicas es la creación de índices, que permiten acceder rápidamente a los registros de una tabla en función de los valores de una columna o conjunto de columnas. En el caso de la tabla «Revista», se ha creado un índice llamado **nombre_index** en la columna **nombre** mediante la instrucción **CREATE INDEX nombre_index ON revista (nombre)**. Esto facilita la búsqueda y recuperación de registros en la tabla «Revista» basándose en el valor de la columna **nombre**.

Además de crear índices, se ha utilizado el comando **ANALYZE** para actualizar las estadísticas de las tablas. Esto permite al sistema de gestión de bases de datos (SGBD) recopilar información actualizada sobre la distribución de los datos y la estructura de las tablas. En el caso de la tabla «Revista», se ha aplicado el comando **ANALYZE revista** para asegurar que las estadísticas se encuentren actualizadas

y que el SGBD pueda tomar decisiones más acertadas al ejecutar consultas sobre esta tabla. Del mismo modo, se ha realizado el análisis del resto de tablas de la base de datos.

Estas técnicas permiten acelerar las consultas y mejorar la eficiencia en la recuperación de datos, lo que resulta en una aplicación web más rápida y con mejor respuesta para los usuarios.

Diagramas de datos

A continuación se mostrará un sencillo diagrama relacional (ver Figura C.1).

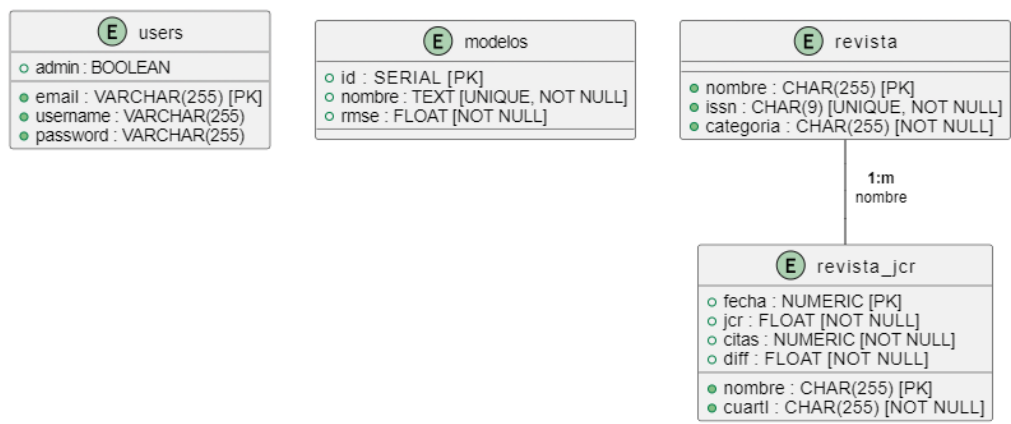


Figura C.1: Diagrama Relacional

C.3. Diseño procedimental

Todo el diseño procedimental se detalla en los siguientes diagramas de secuencia: C.2, C.3, C.4, C.5 y C.6.

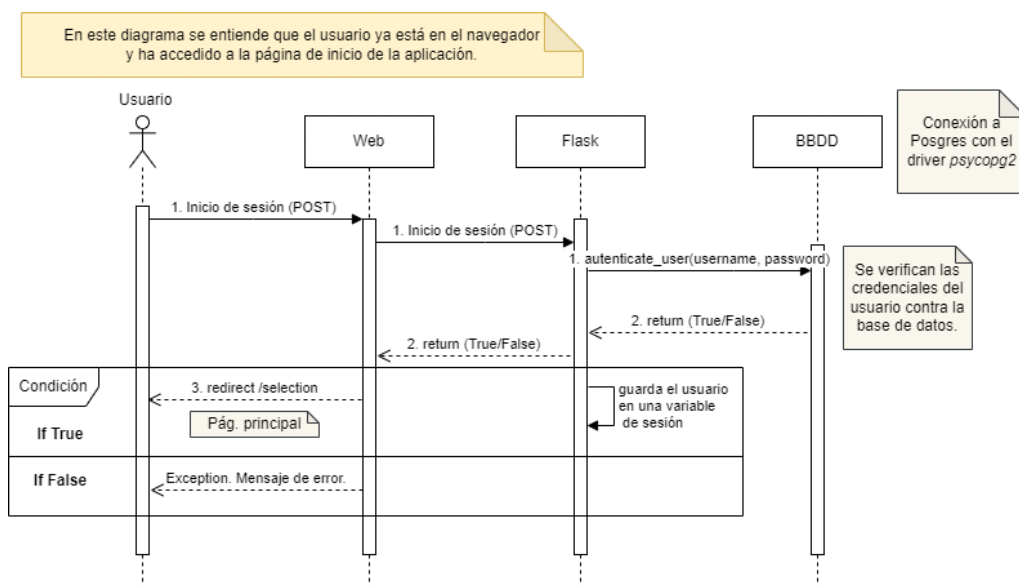


Figura C.2: Diagrama de secuencia de inicio de sesión

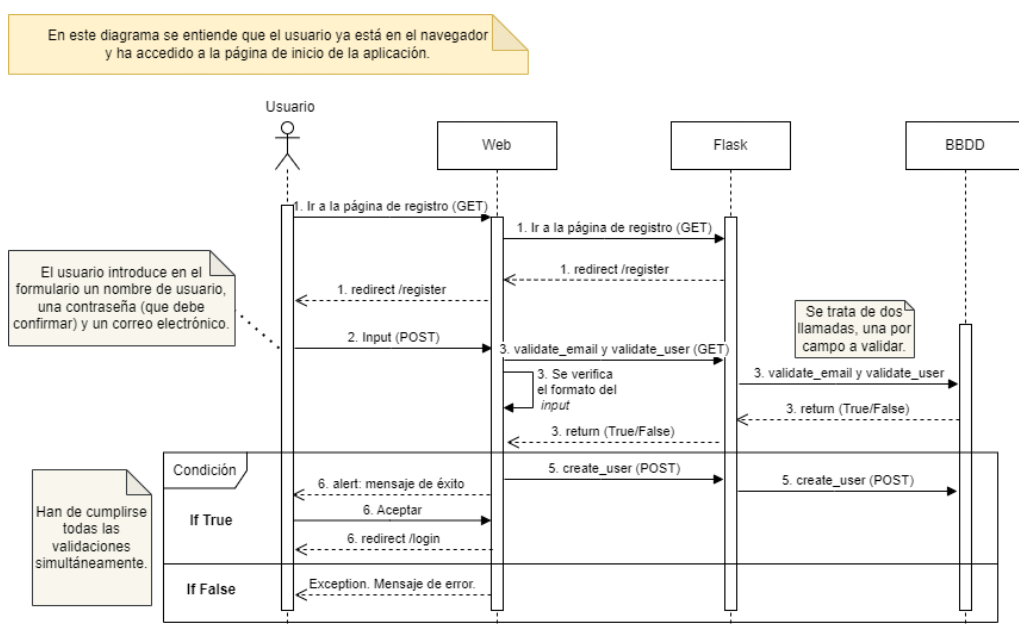


Figura C.3: Diagrama de secuencia de registro

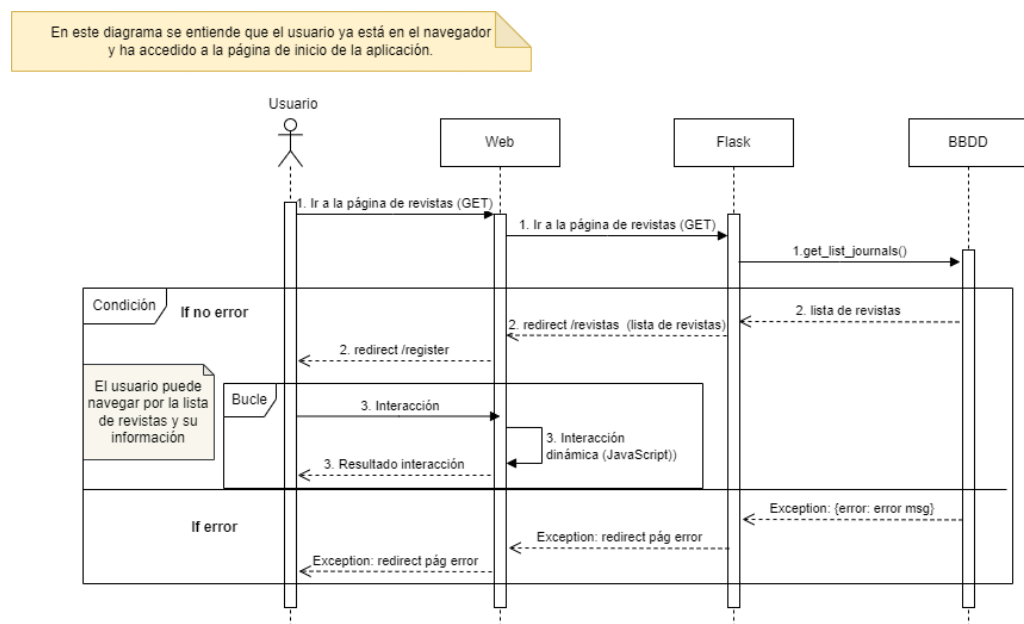


Figura C.4: Diagrama de secuencia de la consulta de revistas

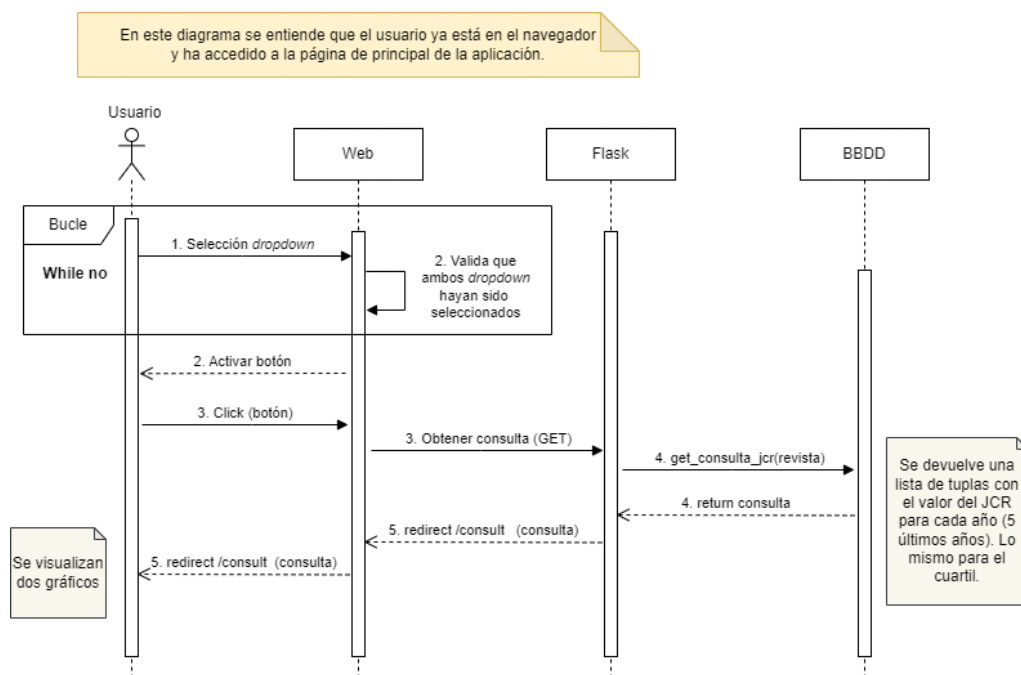


Figura C.5: Diagrama de secuencia de la consulta del histórico del JCR de una revista

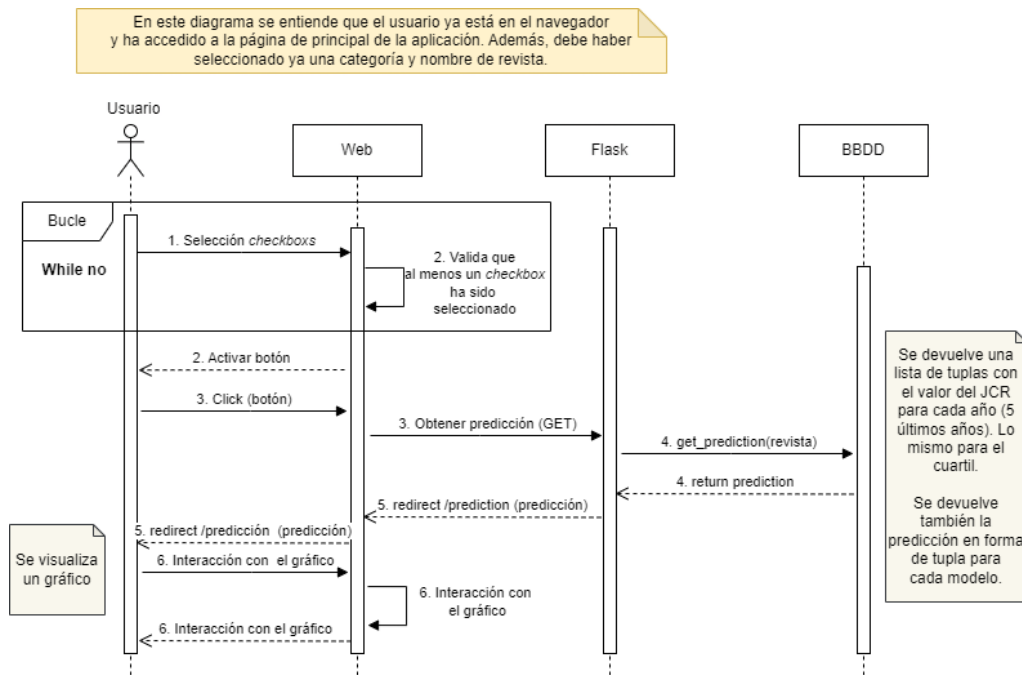


Figura C.6: Diagrama de secuencia de la predicción del JCR de una revista

C.4. Diseño arquitectónico

Se ha desarrollado una aplicación web sencilla que usa una arquitectura cliente-servidor de dos capas (ver Figura C.7).

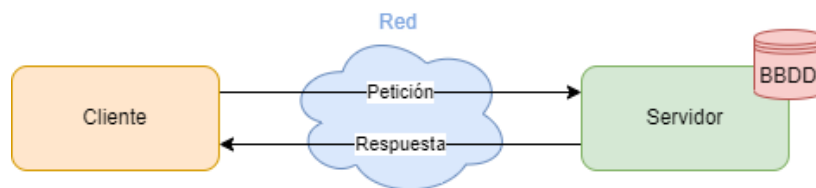


Figura C.7: Arquitectura Cliente-Servidor de dos capas

Una arquitectura Cliente-Servidor de dos capas para una aplicación web presenta varias ventajas significativas. En primer lugar, permite una clara separación de responsabilidades entre el cliente y el servidor, lo que facilita el mantenimiento y la escalabilidad del sistema. Además,

esta arquitectura permite un rendimiento óptimo, ya que las tareas se distribuyen eficientemente entre el cliente y el servidor.

Por otro lado, se ha programado el *backend* en tres capas, usando una arquitectura Modelo-Vista-Controlador (MVC). Esta arquitectura promueve una clara separación de responsabilidades, lo que facilita la organización y mantenimiento del código. El modelo se encarga del acceso a datos, la vista se ocupa de la presentación de la interfaz de usuario, y el controlador de la lógica de negocio (además de coordinar las interacciones entre el modelo y la vista).

Además, la arquitectura MVC favorece la reutilización de código, ya que los componentes son independientes y pueden ser modificados o reemplazados sin afectar a los demás. Esto aumenta la flexibilidad y permite un desarrollo más eficiente.

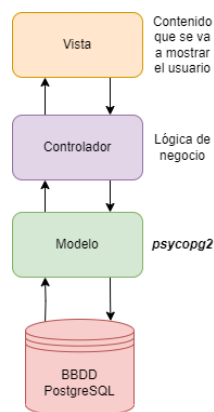


Figura C.8: Arquitectura Modelo-Vista-Controlador de tres capas

C.5. Diseño de Interfaces

La aplicación estará conformada por seis interfaces principales, que se detallan a continuación:

- Interfaces iniciales con funciones de inicio de sesión y registro, verificando la validez de los datos.
- Interfaz principal para seleccionar una revista y los modelos de predicción.
- Interfaces con el histórico y la predicción del JCR, además de gráficos representativos.

- Interfaz con una lista de las revistas disponibles con barra de búsqueda.
- Área de usuario donde modificar las credenciales.

En la etapa inicial del proyecto, se procedió al diseño de *mockups* de las interfaces de la aplicación. Un *mockups* se refiere a una representación visual estática de una interfaz de usuario, que muestra de manera aproximada la disposición de elementos, la estructura y la apariencia general de la aplicación. Los *mockups* permiten visualizar y evaluar el diseño antes de implementarlo, lo que ayuda a identificar posibles mejoras y realizar ajustes necesarios de manera temprana.


A continuación, adjunto los *mockups* diseñados al inicio del proyecto. Es preciso recordar que son una representación preliminar de las interfaces y sirven como punto de partida para el desarrollo de la aplicación. Estos diseños han evolucionado y se han ido refinando a medida que ha avanzado el proyecto.



Figura C.9: *Mockup* de inicio de sesión



The mockup shows a registration form titled "Registro" in a dark blue header. Below the header, there are three input fields: "Nombre de usuario", "Contraseña", and "Correo electrónico". The "Contraseña" field has a toggle icon (an eye) to its right. Below the input fields is a dark blue button labeled "Aceptar". At the bottom of the form, there is a link: "¿Ya eres usuario? [Inicia sesión aquí](#)".

Figura C.10: *Mockup* de registro

The mockup shows the main page titled "Paperrank" in a dark blue header. Below the header, there are three dropdown menus: "Seleccione una revista", "Seleccione un año", and "Seleccione un modelo". The "Seleccione un modelo" dropdown has two options: "Modelo 1" (unchecked) and "Modelo 2" (checked). Below the dropdowns is a dark blue button labeled "Calcular". At the bottom of the page, there are two links: "Política de privacidad" and "Términos de uso".

Figura C.11: *Mockup* de página principal

Figura C.12: *Mockup* de predicción

Para contrastar con los *mockups* iniciales, también se adjuntan las interfaces definitivas del proyecto. Estas interfaces representan la versión final del diseño de la aplicación y reflejan los cambios y mejoras implementados para lograr una experiencia de usuario óptima y satisfactoria.

El cambio más relevante a destacar se encuentra en la interfaz principal. Se puede apreciar que el menú de selección del año ha desaparecido. Esta decisión es lógica puesto que los datos de los que disponemos en la base de datos son limitados.



Figura C.13: Interfaz definitiva de inicio de sesión



Figura C.14: Interfaz definitiva de registro

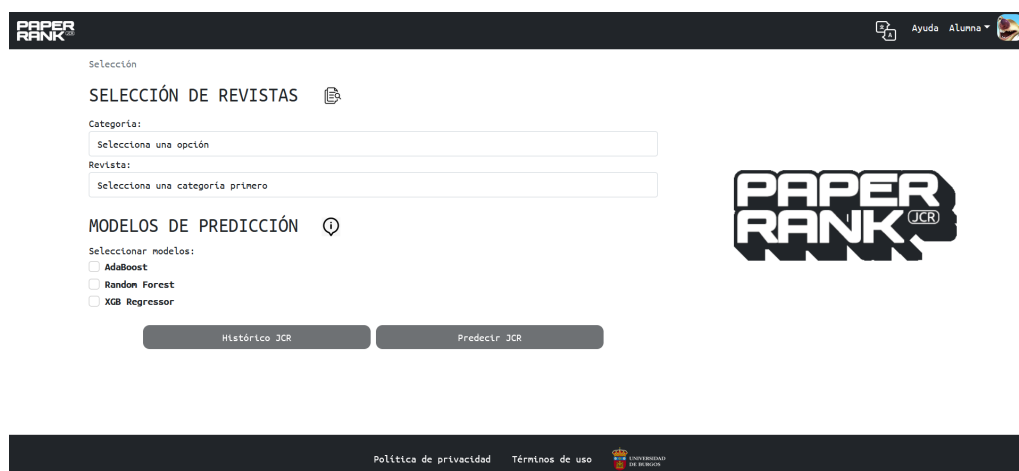


Figura C.15: Interfaz definitiva de página principal

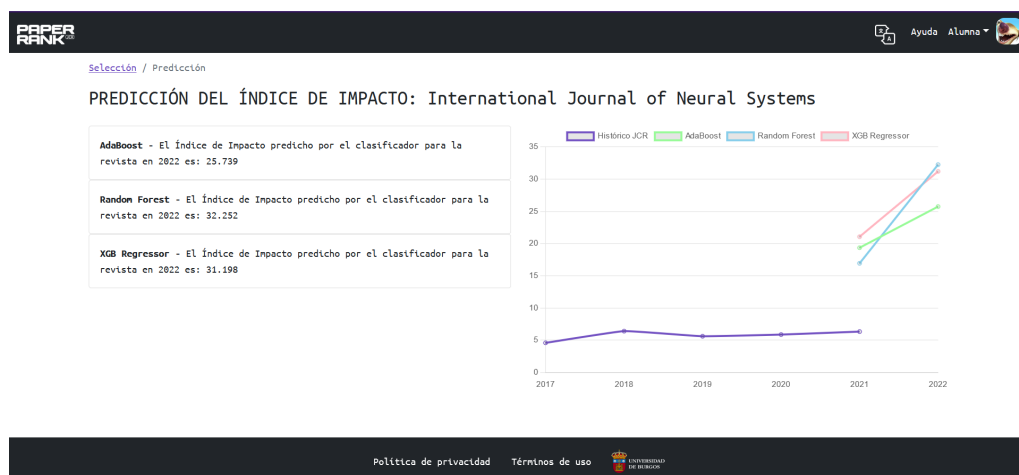


Figura C.16: Interfaz definitiva de predicción

Finalmente, es importante destacar que el diseño de las interfaces con CSS ha sido creado desde cero, sin utilizar plantillas preexistentes. Esta decisión ha permitido una mayor flexibilidad y adaptación a las necesidades específicas del proyecto. Para asegurar la responsividad de la aplicación, se han empleado elementos de Bootstrap, garantizando así una experiencia de usuario consistente en diferentes dispositivos y tamaños de pantalla. Además, tanto el logo de la aplicación como la animación inicial han sido creados por la alumna utilizando Photoshop,

lo que añade un toque personal y distintivo al aspecto visual de la aplicación.

Apéndice *D*

Documentación técnica de programación

D.1. Introducción

En este anexo se aportará la documentación necesaria sobre el diseño, implementación y funcionamiento de la aplicación y del resto de módulos de código. Esta documentación proporcionará a otros desarrolladores y programadores una guía completa y coherente para comprender, utilizar y mantener eficientemente este *software*.

D.2. Estructura de directorios

El repositorio (disponible en https://github.com/glp1002/JCR_Impact_Factor) se organiza en un total de 5 directorios principales (ver Figura D.1) que, en orden alfabético, serían los siguientes:

- **web_application**: contiene el código y los recursos de la aplicación web.
 - **backend**: en esta carpeta se organiza la estructura de clases para gestionar los datos y la lógica de negocio de la aplicación.
 - **diagram_models**: aquí se encuentran los datos binarios (*pickle*) de los modelos de predicción que va a usar la aplicación.
 - **static**: contiene todos los ficheros CSS, JavaScript y multimedia (imágenes, iconos y animaciones).

- **templates**: contiene todos los ficheros HTML y plantillas Jinja de la aplicación.
 - **translations**: contiene todos ficheros «.po» y «.mo» para la internacionalización al inglés, francés, italiano y español.
- **calculate_jcr**: contiene los módulos necesarios para el cálculo del JCR.
 - **listas_jcr**: contiene los ficheros con los datos extraídos y que nos permitirán calcular el JCR.
 - **resultados**: contiene los resultados del cálculo en diferentes ficheros CSV.
- **doc**: contiene la documentación del proyecto (memoria y anexos).
- **data_extraction**: contiene los modelos desarrollados para la fase de extracción de datos.
 - **listas_jcr**: contiene los ficheros con los datos extraídos y que nos permitirán calcular el JCR.
 - **resultados**: contiene los resultados del cálculo en diferentes ficheros CSV.
- **prediction_models**: contiene los modelos de aprendizaje automático de predicción del JCR.



Figura D.1: Diagrama de paquetes del repositorio

D.3. Manual del programador

Esta sección está destinada a proporcionar información detallada sobre cómo utilizar el programa desarrollado en el proyecto.

Desarrollo local con Flask

Durante la etapa de desarrollo, generalmente, se lanza la aplicación de forma local haciendo uso del *framework* de Flask. A continuación se detallan algunos aspectos relevantes a conocer por el programador.

Configuración de la base de datos local

Para poder ejecutar la aplicación en local, será necesario crear una base de datos PostgreSQL y conectarla al código de la aplicación.

1. Instalación de PostgreSQL

- Descargar e instalar PostgreSQL desde el sitio web oficial: <https://www.postgresql.org/download/>.
- Seguir las instrucciones de instalación para su sistema operativo específico.
- Durante la instalación, será necesario asegurarse de recordar la contraseña del usuario «postgres» (usuario predeterminado), ya que será necesaria más adelante. Un detalle a tener en cuenta es que en Linux, el usuario «postgres», no requiere una contraseña inicial. Sin embargo, en nuestro caso es preciso configurar una contraseña para poder establecer conexiones a través de Flask (para ello se puede usar, desde la consola psql, el comando

```
ALTER USER postgres with password 'contraseña'
```

2. Creación de la BBDD a través de un terminal

- Abrir un terminal y comprobar que el servidor PostgreSQL esté en funcionamiento.
- Ejecutar el siguiente comando para acceder a la interfaz de línea de comandos de PostgreSQL:

```
psql -U postgres
```

- Se solicitará la contraseña del usuario «postgres».
- Una vez en la interfaz de PostgreSQL, ejecutar el siguiente comando para crear una nueva base de datos:

```
CREATE DATABASE nombre_basedatos;
```

(Reemplazar `nombre_basedatos` con el nombre que desee para su base de datos.)

- Se puede verificar que la base de datos se haya creado correctamente ejecutando el comando `\l`, que mostrará una lista de todas las bases de datos disponibles.

3. Creación de la BBDD utilizando PgAdmin

- Abrir PgAdmin, la interfaz gráfica de administración de PostgreSQL.
- En la ventana de conexión, introducir los detalles de conexión necesarios, como el nombre del servidor, el puerto, el nombre de usuario y la contraseña.
- Hacer clic con el botón derecho en «Servers» y seleccionar «Create» y «Server».
- Completar la información requerida, como el nombre del servidor y la dirección IP.
- En la pestaña «Connection», introducir el nombre de usuario y la contraseña del usuario «postgres».
- Hacer clic en «Save» para guardar la configuración del servidor.
- En el panel izquierdo de PgAdmin, expandir el servidor recién creado y hacer clic con el botón derecho en «Databases».
- Seleccionar «Create» y «Database».
- Ingresar el nombre deseado para la base de datos y hacer clic en «Save» para crearla.

Después, bastará con descomentar en el fichero de configuración de la aplicación (`app.py`) la opción para conectar con la base de datos local e incluir ahí las credenciales de la base de datos nueva (y comentar la opción para conectarse a la base de datos de Heroku). Más adelante, en la subsección «Interacción con la base de datos», se dan más detalles al respecto.

NOTA

Para aquellos usuarios de sistemas operativos Windows, es importante tener en cuenta un detalle en el momento de obtener información de la base de datos en PostgreSQL. Cuando se recolectan los datos en formato CSV, es necesario cambiar la ruta de generación de estos archivos a la carpeta «Public»^a. Esto se debe a que el usuario por defecto que genera PostgreSQL (usuario «postgres»), no tiene los permisos suficientes para leer e importar datos de los CSV en cuestión ya que, para ello, se utiliza el comando *copy* (que solo puede ser ejecutado por «Superusers»)[12]. Sin embargo, la carpeta «Public» cuenta con los permisos necesarios para que el usuario que crea PostgreSQL pueda acceder y leer los datos. Por lo tanto, es importante seguir este paso para garantizar que se pueda obtener correctamente la información recogida previamente.

^aDirectorio que permite compartir archivos entre distintos usuarios en un mismo sistema Windows. Microsoft ha mantenido este directorio desde la versión de WindowsXP.

Instalación de dependencias

1. Configuración del entorno virtual

NOTA

Este paso es opcional pero muy recomendado, especialmente si se están manejando otras aplicaciones en el mismo dispositivo.

Antes de comenzar, es una buena práctica configurar un entorno virtual, ya que permite aislar las dependencias del proyecto y evitar conflictos con otras aplicaciones. Para configurar un entorno virtual, sigue los siguientes pasos:

- Abrir un terminal y navegar hasta la carpeta raíz del proyecto.
- Instalar `python3-venv`.
- Ejecutar el siguiente comando para crear un nuevo entorno virtual:

```
python3 -m venv nombre_entorno_virtual
```

(Reemplazar `nombre_entorno_virtual` con el nombre que se desee para el entorno virtual).

- Activar el entorno virtual ejecutando el siguiente comando:
 - En Windows:

```
nombre_entorno_virtual\Scripts\activate
```

- En Linux:

```
source nombre_entorno_virtual/bin/activate
```

2. Instalación de Flask y dependencias

Una vez configurado el entorno virtual (si se ha decidido utilizar uno), el siguiente paso es instalar Flask y las dependencias necesarias para la ejecución de la aplicación. Será preciso navegar hasta la carpeta raíz del proyecto. Después, se deberá seguir los siguientes pasos:

- Ejecutar el siguiente comando para instalar las dependencias especificadas en el archivo `requirements.txt` (se recomienda usar la versión de Python 3.11 para asegurar la compatibilidad de versiones con los requerimientos):

```
pip install -r requirements.txt
```

Este comando buscará las dependencias requeridas en el archivo `requirements.txt` y las instalará automáticamente. Si diese algún problema al instalar el paquete `psycopg2` (driver Python para Postgres), se puede instalar directamente el binario con la instrucción:

```
pip install psycopg2-binary
```

- Si fuera necesario agregar dependencias adicionales a su proyecto, puedes hacerlo editando el archivo `requirements.txt`. Es preciso asegurarse de seguir el formato adecuado, donde cada dependencia está en una línea separada y puede incluir una versión específica si es necesario. Por ejemplo:

```
flask==2.0.1  
sqlalchemy==1.4.20
```

3. Verificación de la instalación

Una vez completada la instalación de Flask y las dependencias, conviene verificar que todo se haya configurado correctamente ejecutando un comando de prueba. Por lo general, esto implica ejecutar un servidor de desarrollo local y abrir su aplicación web en un navegador. Aquí hay un ejemplo de cómo hacerlo:

- Ejecutar el siguiente comando para iniciar el servidor de desarrollo de Flask:

- En Windows:

```
win_start.cmd
```

- En Linux:

```
./lin_start.sh
```

Si no tuviera permisos de ejecución, bastaría con ejecutar previamente el siguiente comando:

```
chmod +x lin_start.sh
```

- Abrir un navegador web y dirigirse a la siguiente dirección URL: <http://localhost:5000>

Si todo ha sido configurado correctamente, se debería poder ver la aplicación Paperrank en funcionamiento en el navegador.

Lanzamiento de la aplicación

Para lanzar la aplicación, bastará con replicar el punto 3 del apartado anterior «Verificación de la instalación». Es preciso repetir este proceso cada vez que se realiza un cambio en el código.

Por otro lado, se puede lanzar Flask en modo *debug*. Para ello se debe establecer la variable de entorno `FLASK_ENV` antes de ejecutar el comando para iniciar el servidor Flask. La forma exacta de hacerlo puede variar ligeramente según el sistema operativo.

- En Windows:

Abrir un terminal (por ejemplo, PowerShell o CMD) y navegar hasta el directorio donde se encuentra la aplicación Flask.

Establecer la variable de entorno `FLASK_ENV` a «*development*» utilizando el siguiente comando:

```
set FLASK_ENV=development
```

- En Linux:

Abrir un terminal y navegar hasta el directorio donde se encuentra la aplicación Flask. Establecer la variable de entorno `FLASK_ENV` a «*development*» utilizando el siguiente comando:

```
export FLASK_ENV=development
```

Interacción con la base de datos en Flask

Para poder establecer una conexión con la base de datos que se ha creado, para posteriormente ejecutar sentencias SQL.

1. Configuración de la conexión a la BBDD

- Abrir el archivo de configuración de la aplicación Flask «app.py».
- Se puede observar el módulo `psycopg2` importado:

```
import psycopg2
```

- Configurar los parámetros de conexión (con las credenciales de la base de datos que se ha creado en el apartado anterior):

```
app.config['DATABASE'] = {  
    'host': 'localhost',  
    'port': '5432',  
    'database': 'nombre_basedatos',  
    'user': 'nombre_usuario',  
    'password': 'contraseña_usuario'  
}
```

Realizar la conexión a la base de datos:

```
def get_db():  
    if 'db' not in g:  
        g.db = psycopg2.connect(**app.config['DATABASE'])  
    return g.db
```

2. Ejecución de consultas en la base de datos

- Crear una función en el paquete «modelo» que ejecute consultas en la base de datos. A continuación se presenta un ejemplo práctico de cómo ejecutar una consulta **SELECT**:

```
def obtener_datos(self):  
    cursor = self.conn.cursor()  
    cursor.execute("SELECT * FROM tabla")  
    datos = cursor.fetchall()  
    cursor.close()  
    conn.close()  
    return datos
```

En este ejemplo, se establece una conexión a la base de datos, se ejecuta una consulta **SELECT** en una tabla y se obtienen los resultados. Luego, se cierran tanto el cursor como la conexión a la base de datos (sin necesidad de incluirlo directamente en el método).

Uso de plantillas en Flask

Las plantillas en Flask te permiten separar la lógica de presentación de la aplicación web del código Python. Esto mejora la legibilidad y mantenibilidad del código, así como la reutilización de componentes de interfaz de usuario. Flask utiliza el motor de plantillas Jinja2, que proporciona una sintaxis sencilla y poderosa. A continuación, se detallan los pasos para utilizar plantillas en Flask:

1. **Creación de una carpeta para las plantillas** Las plantillas de la página web se encuentran en la carpeta «templates» (llamada así por convención). Ahí deberán ubicarse todos los ficheros HTML.

```
/paperrank  
|-- app.py  
|-- /templates  
    |-- plantilla.html
```

2. **Creación de una plantilla**

- Crear un archivo HTML dentro de la carpeta «templates». Por ejemplo, «plantilla.html».

- Abrir el archivo «plantilla.html» y escribir el código HTML deseado, incluyendo las partes que se desee que sean dinámicas utilizando las etiquetas y sintaxis de Jinja2. Por ejemplo:

```
<!DOCTYPE html>
<html>
<head>
    <title>{{ titulo }}</title>
</head>
<body>
    <h1>Bienvenido a mi aplicación Flask</h1>
    <p>{{ mensaje }}</p>
</body>
</html>
```

3. Ejecución de la aplicación

- Guardar los cambios y ejecutar la aplicación Flask.
- Abrir un navegador web y accede a la URL correspondiente a la ruta definida en la aplicación Flask.
- Se debería ver la plantilla renderizada con los valores dinámicos insertados en los lugares correspondientes.

Internacionalización con Babel en Flask

La internacionalización es un aspecto muy importante al desarrollar una aplicación web, y Flask ofrece soporte para la internacionalización utilizando la extensión Flask-Babel. A continuación, se detallan los pasos necesarios para internacionalizar la aplicación Flask y cómo implementar la traducción en las plantillas HTML.

1. Configuración de Babel en la aplicación Flask

En el archivo principal de la aplicación «app.py» deberán realizarse los siguientes pasos:

- Se puede observar que la clase Babel está importada al inicio:

```
from flask_babel import Babel
```

- También se puede observar la presencia de una función para determinar el idioma a utilizar:

```
# Variables globales de internacionalización con Babel
app.config['BABEL_DEFAULT_LOCALE'] = 'en'
```

```

app.config['LANGUAGES'] = {
    'en': gettext('Inglés'),
    'es': gettext('Español'),
    'fr': gettext('Francés'),
    'it': gettext('Italiano')
}

# Obtiene el idioma preferido del navegador, si no se
# toma el idioma por defecto de la aplicación
def get_locale():
    lang = session.get('LANGUAGES', None)
    if lang == None:
        browser_locale = request.accept_languages.
            best_match(app.config['LANGUAGES'].keys())
        if browser_locale is not None:
            return browser_locale
        return app.config['BABEL_DEFAULT_LOCALE']
    else:
        app.config['BABEL_DEFAULT_LOCALE'] = lang
        return app.config['BABEL_DEFAULT_LOCALE']

```

- Inicializar Babel, pasando la función `get_locale` como parámetro `locale_selector`:

```
Babel(app, locale_selector=get_locale)
```

Además de estas configuraciones básicas (ya implamentadas), se puede añadir lógica adicional según las nuevas necesidades que aparezcan.

2. Internacionalización de las plantillas HTML

En las plantillas HTML, utilizar llaves dobles `{{ }}` para los textos que se deseen internacionalizar. Por ejemplo:

```
<p>{{ gettext("Ayuda") }}</p>
```

El método `gettext` permite traducir el texto proporcionado en el idioma seleccionado.

3. Generación de archivos de traducción

- Abrir un terminal en el directorio raíz de la aplicación.

- Ejecutar el siguiente comando para generar un archivo de configuración automáticamente:

```
pybabel extract -F babel.cfg -o messages.pot .
```

Esto creará un archivo «messages.pot» que contiene los textos extraídos de las plantillas HTML.

- Después, ejecutar el siguiente comando para inicializar los archivos de traducción para cada idioma que se desee:

```
pybabel init -i messages.pot -d translations -l es
```

Será necesario reemplazar «es» con el código del idioma correspondiente. Este comando creará un archivo «.po» en la carpeta de traducciones para cada idioma, donde será necesario ingresar las traducciones para cada texto.

4. Compilación de las traducciones

Una vez que se hayan ingresado las traducciones en los archivos «.po», se deberá ejecutar el siguiente comando para compilar las traducciones y generar los archivos binarios «.mo»:

```
pybabel compile -d translations
```

Esto creará los archivos binarios «.mo» necesarios para cargar las traducciones en la aplicación Flask.

Consideraciones adicionales

Es necesario tener en cuenta que el nombre de los directorios está elegido a propósito para ser entendido por Flask. Por ejemplo, las carpetas «templates» y «static» son reconocidas por Flask como las carpetas donde se encuentran los ficheros HTML, CSS y JavaScript por defecto. Si se quisiera renombrar estas carpetas, se deberá especificar el nuevo nombre a la hora de crear la aplicación en el fichero «app.py».

D.4. Compilación, instalación y ejecución del proyecto

Despliegue en Heroku

En esta sección, se describe el proceso de lanzamiento de la aplicación en Heroku (aprovechando el tiempo de despliegue gratuito que proporciona la cuentas GitHub Student).

NOTA

En el código de la aplicación web, al incluir `import` de clases propias, es preciso añadir al inicio de la ruta de paquetes el punto (por ejemplo, `import .miCarpeta.miClase`). Sin el punto inicial, Heroku no identificará los módulos por no estar en la raíz del repositorio.

Ficheros necesarios

Se deben ubicar los siguientes ficheros en la raíz del repositorio:

- `requirements.txt` → con todas las dependencias necesarias (`$ pip freeze > requirements.txt`)
- `Procfile` → indicando el módulo donde se encuentra `app.py` (fichero principal y su nombre).
- `runtime.txt` → indicar el lenguaje de programación y su versión (i.e.:`python-3.11.1`).

Creación de la aplicación

Para la creación de la aplicación en Heroku, se deben seguir los siguientes pasos previos:

1. Crear cuenta de Heroku (<https://www.heroku.com/what>), donde se solicitará elegir una forma de doble autenticación.
2. Solicitar el *Student Pack* (<https://www.heroku.com/github-students>) que ofrece GitHub Student para poder utilizar de forma gratuita (con limitaciones) los servicios de Heroku. Será necesario esperar algunos días antes de recibir una confirmación.

3. Descargar el *software* Heroku CLI (<https://devcenter.heroku.com/articles/heroku-cli>) adecuado para su sistema operativo.
4. Crear aplicación. En este caso, se recomienda hacerlo desde un terminal en la raíz del repositorio de GitHub.

Los comandos para crear la aplicación son los siguientes:

1. Instalar el Dyno básico: `$ pip install gunicorn`.
2. Ejecutar `$ heroku login`, que abrirá una pesataña en un navegador y solicitará las credenciales de la cuenta de Heroku creada.
3. Para crear la aplicación, bastará con ejecutar `$ heroku create nombre_app -region=eu`
Como respuesta, deberá obtener la dirección de la aplicación y del repositorio remoto de Heroku con el siguiente formato:
`https://nombre_app.herokuapp.com/ |`
`https://git.heroku.com/nombre_app.git`
4. Finalmente, ejecutar `$ git push heroku main` → para subir el contenido de la rama `main` del repositorio GitHub al repositorio remoto de Heroku. Este comando se deberá ejecutar cada vez que hay cambios nuevos que se quieran desplegar.
5. Para desplegar la aplicación bastará con ejecutar `$ heroku open`.

Finalmente, se puede comprobar que la aplicación se ha creado correctamente en el *dashboard* de Heroku (<https://dashboard.heroku.com/apps/>).

Añadir una base de datos

En este caso, como nuestra aplicación hace uso de una base de datos Postgres, se hará uso del complemento Heroku Postgres (<https://devcenter.heroku.com/articles/getting-started-monitoring-postgres-database>). Para ello será preciso seguir los siguientes pasos:

- Añadir (comando `$ addon` o directamente desde la web) un nuevo complemento: Heroku Postgres. Sin el *Student Pack* mencionado anteriormente, será necesario realizar un pago.
- Además será necesario configurar en el código las credenciales de la base de datos (proporcionadas por Heroku) para poder establecer las conexiones. Es importante ejecutar de nuevo el comando `$ git push heroku main` tras realizar estos cambios.

- Volver a ejecutar `$ heroku open` y la aplicación ya estará en funcionamiento.

Mantenimiento y gestión de la aplicación

Algunos comandos útiles (<https://devcenter.heroku.com/articles/heroku-cli-commands>) para el mantenimiento y gestión de la aplicación son los siguientes:

- Destruir la aplicación: `$ heroku apps:destroy -a nombre_app` (pide confirmación).
- Reiniciar todos los *Dynos*: `$ heroku restart`.
- Registro: `$ heroku logs` (Añadir la opción `-tail` si se quiere que el log se mantenga abierto).

Por otro lado, Heroku y Heroku Postgres presentan ciertas limitaciones que es importante tener en cuenta. En primer lugar, una de las restricciones notables es la incapacidad de manejar peticiones largas a las bases de datos. Esto implica que las consultas o transacciones que requieran un tiempo prolongado de ejecución pueden verse interrumpidas o finalizadas prematuramente. Esto puede ser problemático en casos donde se necesite procesar grandes volúmenes de datos o realizar operaciones complejas.

Además, otra limitación significativa está relacionada con las restricciones de versiones en la plataforma Python de Heroku. En particular, la plataforma exige que todas las aplicaciones sean compatibles con Python 3.11, que es una versión muy reciente. Esta rigidez en las versiones puede resultar restrictiva, ya que puede requerir ajustes y actualizaciones significativas en el código existente para asegurar la compatibilidad. Esta falta de flexibilidad puede ser un desafío para aquellos proyectos que dependen de bibliotecas o *frameworks* que aún no son totalmente compatibles con Python 3.11.

D.5. Otros módulos del proyecto

Además de la aplicación, es importante mencionar que para poder ejecutar fácilmente el módulo de extracción de datos de Crossref que se encuentra en el repositorio, se ha creado un entorno virtualizado.

Se trata de un entorno aislado que te permite instalar y gestionar de forma independiente las bibliotecas y paquetes de Python que se

necesitan para el proyecto, sin afectar a otros proyectos o al sistema operativo. Esto nos permite evitar problemas de dependencias y compatibilidad entre diferentes versiones de bibliotecas, además de facilitar la gestión y el mantenimiento del proyectos. Miniconda, por su parte, es una distribución de Python que incluye un gestor de paquetes llamado Conda, que te permite instalar y gestionar de forma sencilla bibliotecas y paquetes de Python. Así pues, se procede a crear un entorno virtualizado de Python utilizando Miniconda. El entorno en cuestión se encuentra empaquetado en la carpeta comprimida `environment.zip`. Para poder utilizar este entorno, bastará con realizar los siguientes pasos:

1. Instalar MiniConda en el dispositivo en el que se quiera ejecutar el prototipo.
2. Crear el entorno usando el siguiente comando en la consola de MiniConda: `conda env create -f environment.yml`. Es preciso estar en el mismo directorio que el fichero `environment.yml` o especificar la ruta completa.
3. A continuación, se debe activar el entorno con el siguiente comando: `conda activate crossref_entorno`.
4. Ejecutar el prototipo de Crossref.
5. Una vez finalizado, para desactivar el entorno, bastará con ejecutar el siguiente comando: `conda deactivate`.

D.6. Pruebas del sistema

En el desarrollo de *software*, es esencial contar con una metodología adecuada para garantizar la calidad del producto final. En este sentido, el Test Driven Development (TDD) se ha convertido en una de las metodologías más populares en la actualidad. TDD consiste en escribir las pruebas antes de comenzar a desarrollar el código, lo que permite detectar errores de manera temprana y asegurar que el *software* cumpla con los requisitos establecidos. En este apartado, presentaremos los proyectos de pruebas que hemos llevado a cabo utilizando TDD como metodología principal.

Esta metodología se divide en tres fases principales: la fase roja, la fase verde y la fase morada.

- En la fase roja, se escriben las pruebas unitarias para el código que se desea implementar, las cuales deben fallar inicialmente (puesto que aún no existe el código a probar).
- En la fase verde, se escribe el código mínimo necesario para que las pruebas pasen satisfactoriamente.
- Finalmente, en la fase morada, se realiza la refactorización del código escrito anteriormente para mejorar su calidad y mantenerlo limpio y legible.

Este enfoque iterativo permite a los desarrolladores garantizar la calidad y funcionalidad del software en cada paso del proceso de desarrollo y ayuda a reducir los errores y *bugs* que puedan surgir en el proceso.

Ciclo de vida TDD

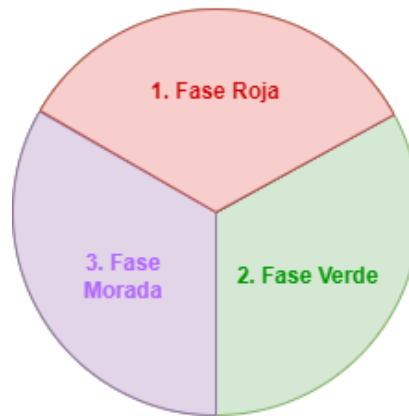


Figura D.2: Fases de la metodología TDD

Prototipo de Crossref

Este es el prototipo final para la extracción de los datos bibliográficos. A continuación, se documentará la información más relevante del proyecto de pruebas llevado a cabo.

Diseño y arquitectura

Puesto que el *software* a probar está compuesto de un solo *script* de Python, se ha generado un único módulo de pruebas llamado `crossref_test.py`.

Fase Roja

A lo largo de esta fase se definen las pruebas necesarias para crear un código que extraiga correctamente los datos de Crossref. Los casos de prueba diseñados son los siguientes:

ID	Nombre	Descripción	Salida esperada
C01	Revista existente	Método <code>test_journal_exist</code> . Se trata de extraer una revista que previamente se conoce que existe.	True
C02	Revista inexistente	Método <code>test_journal_exist</code> . Se trata de extraer una revista que previamente se conoce que no existe.	False
C03	Cargar datos de entrada	Método <code>test_load_data</code> . Se comprueba que se cargan bien los datos del CSV de entrada.	Carga correcta
C04	Formateado I	Método <code>test_formatData</code> . Se comprueba que se eliminan bien las etiquetas HTML y los espacios.	Cadena correcta
C05	Formateado II	Método <code>test_formatData</code> . Se comprueba que se eliminan bien las etiquetas «href».	Cadena correcta
C06	Formateado III	Método <code>test_formatData</code> . Se comprueba que una cadena correcta no cambia.	Cadena sin modificaciones
C07	Formateado IV	Método <code>test_formatData</code> . Se comprueba que no hay campos vacíos (rellenar con «n.m.»).	Cadena sin valores vacíos
C08	Formato de salida I	Método <code>test_getArticles_returns_list_of_dicts</code> . Se comprueba que se obtiene una lista de artículos.	Lista de diccionarios
C09	Formato de salida II	Método <code>test_getArticles_returns_valid_keys</code> . Se comprueba que las claves del diccionario son correctas.	Claves correctas
C10	Número de resultados	Método <code>test_getArticles_correct_article_count</code> . Se comprueba que se obtiene el número correcto de artículos que aparecen en Crossref.	Número correcto

ID	Nombre	Descripción	Salida esperada
C11	Excepción de ISSN	Método <code>test_getArticles_invalid_issn_raises_error</code> . Se intenta buscar un ISSN incorrecto.	False
C12	Excepción de año I	Método <code>test_getArticles_invalid_year</code> . Se introducen los años de búsqueda en el orden incorrecto.	False
C13	Excepción de año II	Método <code>test_getArticles_invalid_year</code> . Se introducen años inválidos.	False
C14	Ejecución completa I	Método <code>test_main</code> . Pre-requisito: CSV de entrada de datos. Se comprueba que se genera una gráfica al finalizar el proceso. Post-requisito: Imagen con la gráfica generado.	True
C15	Ejecución completa II	Método <code>test_main</code> . Precondición: CSV de entrada de datos. Se comprueba que se generan los ficheros CSVs de resultados al finalizar el proceso. Post-requisito: Ficheros CSV generados.	True
C16	Ejecución completa III	Método <code>test_main</code> . Precondición: CSV de entrada de datos. Se comprueba que se genera el fichero de log al finalizar el proceso. Postcondición: Ficheros de log generados.	True
C17	Ejecución completa IV	Método <code>test_main</code> . Precondición: CSV de entrada de datos y fichero de log generado correctamente. Se comprueba que la información de log es correcta.	True

Tabla D.1: Casos de prueba para el prototipo de Crossref

Fase Verde

Tras la creación del código, finalmente se consigue una versión que pasa todos los test programados en la fase anterior, como se puede observar en la siguiente ilustración:

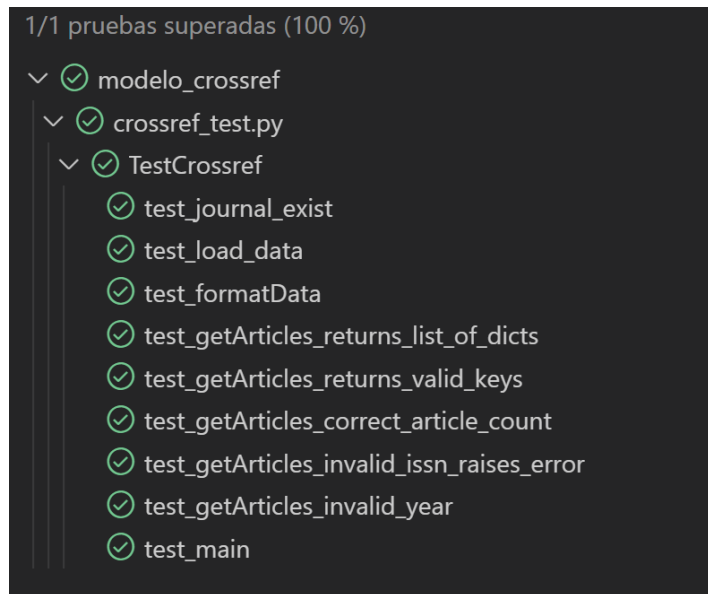


Figura D.3: Pruebas ejecutadas exitosamente

Fase Morada

En esta fase se han realizado acciones de refactorización. A saber:

- Documentación de código: Se han incluido las cabeceras y comentarios más importantes en el código para facilitar su comprensión y claridad. Asimismo, se han eliminado comentarios redundantes o innecesarios.
- Ordenación de *imports*: Se han ordenado los *imports* del código de forma que se identifique fácilmente la naturaleza de los paquetes y módulos empleados. De esta forma, se agrupan al inicio los módulos integrados de Python, posteriormente los paquetes de terceros y, finalmente, módulos o paquetes locales.
- Mensajes de salida I: En lugar de imprimir mensajes de salida en la consola, (ant. `print("[ERROR] Se ha producido un error")`), ahora se utilizan registros de *logging* para registrar los eventos de

la función (por ejemplo, `logging.error('Se ha producido un error')`). Los registros son una forma más eficiente y flexible de registrar información que imprimir en la consola.

- Mensajes de salida II: Utiliza f-strings en lugar de concatenar cadenas para mejorar el rendimiento y la legibilidad del código. Ejemplo: `logging.error('Se ha producido un error'+ codigo_error)` por `logging.error(f'Se ha producido un error {codigo_error}')`.
- Acceso a diccionarios I: En lugar de especificar los parámetros de búsqueda utilizando una cadena de consulta, se puede usar un diccionario y pasarlos como argumentos de palabras clave al método `requests.get()`. Esto puede hacer que el código sea más legible.
- Acceso a diccionarios II: En lugar de crear un diccionario vacío `author` y luego verificar si existe una lista de autores, se puede usar el método `get()` directamente en la lista de autores para obtener el primer elemento de la lista.
- Lista de comprensión: En lugar de crear un diccionario temporal `publicacion` para cada artículo, se puede usar una lista de comprensión para crear una lista de diccionarios de artículos. Esto puede hacer que el código sea más conciso.

Documentación de usuario

E.1. Introducción

Esta sección se enfoca en proporcionar información detallada y clara sobre el funcionamiento de la aplicación web desarrollada con Flask. Tiene como objetivo principal ayudar a los usuarios a comprender cómo utilizar la aplicación de manera efectiva y aprovechar al máximo sus características y funcionalidades.

E.2. Requisitos de usuarios

En este apartado se enumerarán los requisitos que el usuario debe cumplir para poder interactuar con la aplicación satisfactoriamente.

- Sistema Operativo: Indiferente, ya que se trata de una aplicación web (multiplataforma).
- Conocimientos básicos de navegación web: Los usuarios deben estar familiarizados con la navegación en sitios web y tener conocimientos básicos sobre cómo interactuar con formularios, botones y enlaces.
- Conexión a Internet: Para utilizar la aplicación web, los usuarios deben tener acceso a una conexión estable a Internet. Esto permitirá la comunicación con el servidor donde se encuentra alojada la aplicación y garantizará un funcionamiento adecuado de todas las funcionalidades en línea.
- Navegador web: Se recomienda a los usuarios utilizar un navegador web actualizado, como Google Chrome o Mozilla Firefox. Esto

asegurará una mejor compatibilidad con las tecnologías utilizadas en la aplicación y ofrecerá una experiencia de usuario óptima.

- **Dispositivo compatible:** Los usuarios deben contar con un dispositivo compatible, como un ordenador de escritorio, un portátil, una tableta o un teléfono móvil, que les permita acceder y utilizar la aplicación web.
- **Cuenta de usuario:** En función del rol que quiera desempeñar el usuario, podrá crear una cuenta e iniciar sesión cada vez (no es obligatorio).

E.3. Instalación

Dado que se trata de una aplicación web, el usuario no necesita instalar ningún *software* en su dispositivo. Bastará con acceder (asegurándose siempre de tener conexión a Internet) a la dirección donde se aloja la aplicación: <https://paperrank.herokuapp.com/>.

E.4. Manual del usuario

Interfaz de inicio

La interfaz inicial de la aplicación web presenta las siguientes funciones:

- **Log In:** Permite a los usuarios iniciar sesión en la aplicación. Se solicita al usuario un nombre de usuario y una contraseña para acceder (ver Figura E.1).
- **Sign In:** Permite a los nuevos usuarios registrarse en la aplicación. Durante el registro, se solicitará un nombre de usuario, una contraseña y una dirección de correo electrónico. Se verificará que el correo electrónico tenga un formato adecuado y que la contraseña cumpla con los requisitos básicos de seguridad. En caso contrario, se mostrará un mensaje de error con los requisitos (ver Figura E.2).
- **Mostrar/Ocultar Contraseña:** Se incluirá una opción para mostrar u ocultar la contraseña mientras se introduce, brindando al usuario mayor control y seguridad (ver Figura E.3).

En caso de que se introduzcan datos incorrectos en el inicio de sesión o registro, se mostrará un mensaje de error correspondiente (ver Figura E.4).

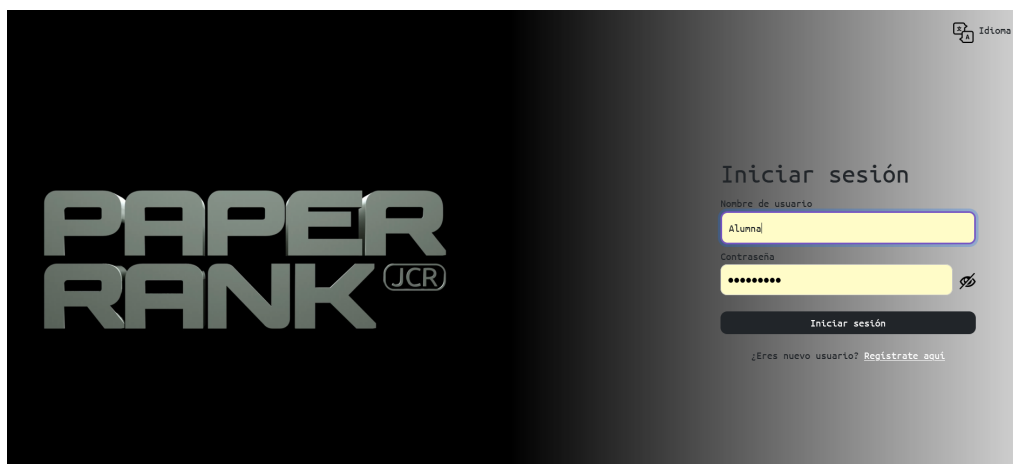


Figura E.1: Interfaz con los campos de inicio de sesión

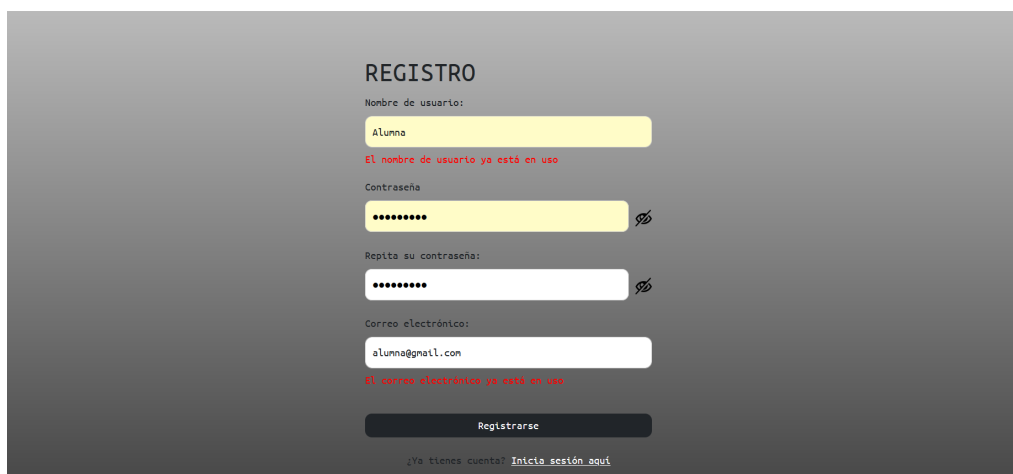


Figura E.2: Interfaz con el registro

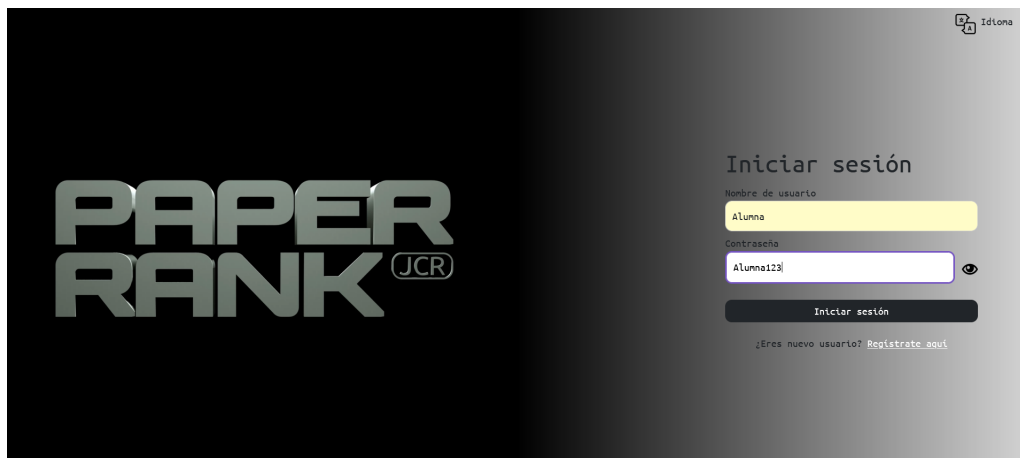


Figura E.3: Interfaz con la contraseña «a la vista»



Figura E.4: Interfaz con mensajes de error


Interfaz principal

La interfaz principal de la aplicación web permite al usuario realizar las siguientes acciones:

- **Selección de categoría y revista:** El usuario puede elegir una categoría y revista de una lista desplegable (o utilizar la barra de búsqueda para encontrar una revista específica) para la cual se desea predecir el JCR (ver Figura E.5).
- **Modelo de Predicción:** El usuario puede seleccionar uno o varios modelos de predicción para calcular el JCR (ver Figura E.6). Se proporcionará una lista de modelos disponibles para elegir.
- **Botones:**
 1. Para visualizar el listado de revistas, el usuario deberá pulsar sobre el icono junto al título de selección de revistas (ver Figura E.7).
 2. Para visualizar una gráfica con los valores del JCR de los últimos años de una revista (que deberá seleccionar previamente), pulsar sobre el botón «Histórico JCR».
 3. Para visualizar las predicciones del JCR, el usuario deberá pulsar sobre el botón «Predecir JCR».

Si alguno de los campos requeridos no está cumplimentado, los botones no se activarán y aparecerá un *tooltip* (ver Figura E.8).

Selección

SELECCIÓN DE REVISTAS 

Categoría:

COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE

Revista:

Artificial Intelligence

Artificial Intelligence
Pattern Recognition
Malaysian Journal of Computer Science
International Journal of Neural Systems
IEEE Transactions on Pattern Analysis and Machine Intelligence
Pattern Recognition Letters
Decision Support Systems

Figura E.5: Interfaz con selección de revistas

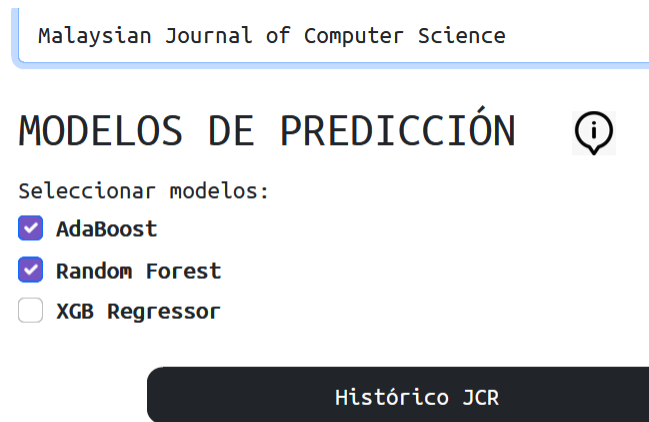


Figura E.6: Interfaz con selección de modelos



Figura E.7: Icono de listado de revistas

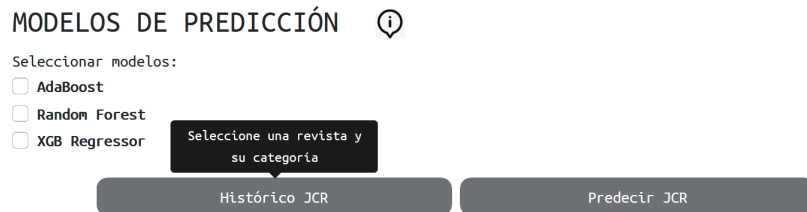


Figura E.8: Tooltip que aconseja rellenar todos los campos

Interfaz del histórico

En esta interfaz, a la que solo se puede acceder desde la interfaz principal, se mostrarán los resultados del cálculo del JCR de la revista seleccionada en los últimos 5 años en un gráfico. El usuario puede interactuar con el gráfico que mostrará un *tooltip* con los valores donde se sitúe el cursor.

[Selección](#) / [Consulta](#)

HISTÓRICO DEL ÍNDICE DE IMPACTO: Journal of Automated Reasoning

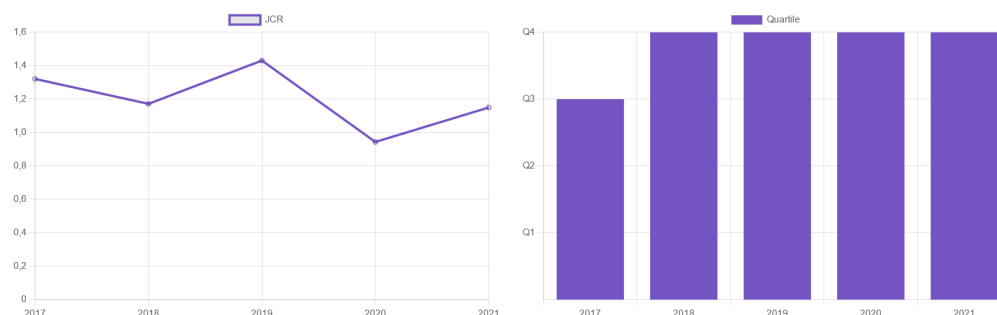


Figura E.9: Interfaz con el histórico del JCR

Interfaz de predicción

En esta interfaz (de nuevo solo se puede acceder desde la interfaz principal) se mostrará otro gráfico los resultados del cálculo del JCR de la revista seleccionada en los últimos 5 años, así como las predicciones de cada modelo. El usuario puede interactuar con el gráfico que mostrará un *tooltip* con los valores donde se sitúe el cursor.

[Selección](#) / [Predicción](#)

PREDICCIÓN DEL ÍNDICE DE IMPACTO: Journal of Automated Reasoning



Figura E.10: Interfaz con la predicción del JCR

Interfaz de listado de revistas

Esta interfaz (de nuevo solamente accesible desde la interfaz principal) muestra una lista de revistas disponibles en la aplicación. Se incluye una barra de búsqueda para facilitar la búsqueda de revistas específicas. La paginación está por defecto en 10 revistas por página. Además, el usuario podrá ver información adicional de cada revista al pulsar sobre el nombre de las mismas.

[Selección](#) / [Revistas](#)

LISTA DE REVISTAS

Nombre	ISSN	Categoría	JCR 2021	Cuartil 2021
CAAI Transactions on Intelligence Technology	2468-2322	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	-	-
IEEE Transactions on Games	2475-1502	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	1.237	Q4
Nature Machine Intelligence	2522-5839	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	25.898	Q1
Machine Learning: Science and Technology	2632-2153	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	6.013	-
Advanced Intelligent Systems	2640-4567	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	7.298	Q1

<< < > >>

Figura E.11: Interfaz con el listado de revistas

Nombre	ISSN	Categoría	JCR 2021	Cuartil 2021
Artificial Intelligence	0004-3702	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	14.05	Q1
Pattern Recognition	0031-3203	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	8.518	Q1
Malaysian Journal of Computer Science	0127-9084	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	0.436	Q4

INFORMACIÓN ADICIONAL DE LA REVISTA

JCR y cuartil de los últimos 5 años:

- Año - JCR - Cuartil
- 2017 - 0.65 - Q4
- 2018 - 0.625 - Q4
- 2019 - 0.723 - Q4
- 2020 - 0.622 - Q4
- 2021 - 0.436 - Q4

Figura E.12: Interfaz con la información adicional de una revista

Interfaz de perfil de usuario

El usuario podrá consultar y modificar (icono de lapicero) sus credenciales en cualquier momento. De igual forma se puede modificar la

imagen (tanto arrastrando la imagen como buscando en el navegador de archivos).



Figura E.13: Interfaz con los datos del usuario



Figura E.14: Interfaz con edición de los datos de usuario

Características generales de las interfaces

Las interfaces de la aplicación web contarán con las siguientes características generales:

- **Barra Superior:** En la zona superior se muestra una barra que contiene el icono del perfil del usuario «logueado», así como un enlace de ayuda que redirigirá a la documentación en línea. Al hacer clic en el icono del perfil del usuario, se abrirá una lista desplegable donde podrá, o bien cerrar la sesión o bien acceder

a la interfaz de perfil de usuario (ver Figura E.15). También se incluirá un símbolo de internacionalización (ver Figura E.16), donde el usuario podrá seleccionar el idioma que desee entre los disponibles (español, inglés, italiano y francés).

- **Política de Privacidad y Términos de Uso:** En el pie de página de la aplicación web, se proporcionará un enlace a la política de privacidad y los términos de uso para que los usuarios puedan acceder a esta información importante (ver Figura E.17). También aparecerá el logo de la Universidad de Burgos, a cuya página oficial pueden redirigirse los usuarios al pulsar sobre él.
- **Navegación:** El usuario podrá retroceder fácilmente a interfaces anteriores a través de las *breadcrumbs* (ver Figura E.18).
- **Compatibilidad y Responsividad:** La aplicación web es compatible con diferentes navegadores y sistemas operativos para garantizar que los usuarios puedan acceder a ella desde cualquier dispositivo y ubicación (ver Figura E.19). Además, se ha diseñado con un enfoque de «responsividad» para que se adapte adecuadamente a diferentes tamaños de pantalla y dispositivos.

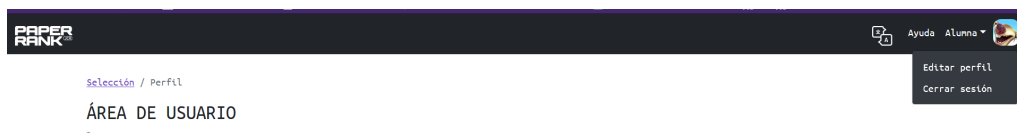


Figura E.15: Barra de navegación con menú desplegable de usuario

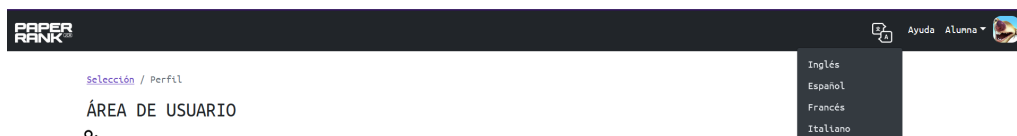


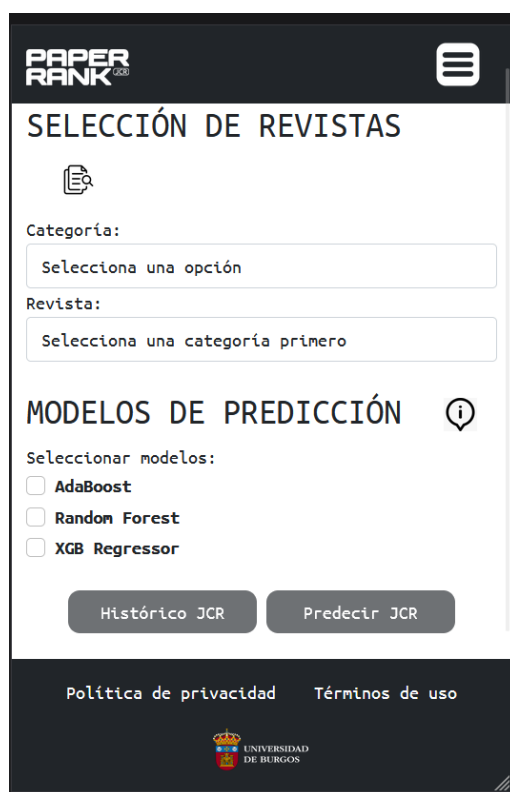
Figura E.16: Barra de navegación con menú desplegable de idiomas



Figura E.17: Barra inferior de la aplicación



Figura E.18: Breadcrumbs



The screenshot displays the PAPER RANK application interface on a mobile device. The top header is dark blue with the 'PAPER RANK' logo and a hamburger menu icon. The main content area is white and divided into two sections. The first section, 'SELECCIÓN DE REVISTAS', features a magnifying glass icon and two dropdown menus: 'Categoría:' with the placeholder 'Selecciona una opción' and 'Revista:' with the placeholder 'Selecciona una categoría primero'. The second section, 'MODELOS DE PREDICCIÓN', includes an information icon and a heading 'Seleccionar modelos:'. Below this are three checkboxes for 'AdaBoost', 'Random Forest', and 'XGB Regressor'. At the bottom of this section are two buttons: 'Histórico JCR' and 'Predecir JCR'. The footer is dark blue, containing links for 'Política de privacidad' and 'Términos de uso', and the logo of the 'UNIVERSIDAD DE BURGOS'.

Figura E.19: Responsividad frente al cambio de tamaño de pantalla

Apéndice *F*

Glosario

AngularJS: Entorno de trabajo de JavaScript para creación una página web simple, emulando una aplicación de escritorio.

API: Acrónimo de *Application Programming Interface* (Interfaz de Programación de Aplicaciones). Es un conjunto de reglas y protocolos que permite la interacción entre diferentes aplicaciones de *software*.

API REST: Una API basada en los principios del protocolo HTTP y diseñada para facilitar la comunicación y transferencia de datos entre sistemas. REST significa *Representational State Transfer* (Transferencia de Estado Representacional).

Babel: Una herramienta de traducción y localización de software que permite adaptar una aplicación para soportar diferentes idiomas y regiones.

Backend: La parte de un sistema o aplicación que se encarga del procesamiento y almacenamiento de datos, así como de la lógica de negocio. Normalmente, se refiere a la parte del sistema que no es visible para los usuarios finales.

BeautifulSoup4: Una biblioteca de Python que facilita el análisis y extracción de datos de documentos HTML y XML.

Bootstrap: Un *framework* de desarrollo web que proporciona estilos CSS y componentes JavaScript predefinidos, lo que facilita la creación de interfaces web responsivas y atractivas.

Breadcrumbs: En español migas de pan. Son un elemento de navegación utilizado en sitios web y aplicaciones para indicar la ubicación actual del usuario dentro de la estructura del sitio o la aplicación. Toman

su nombre de la historia del cuento de Hansel y Gretel, donde los personajes dejaron un rastro de migas de pan para encontrar el camino de regreso a casa.

Chartjs: Una biblioteca JavaScript para la creación de gráficos interactivos y visualmente atractivos en páginas web.

Crossref: Una organización sin fines de lucro que proporciona servicios y datos relacionados con la identificación y el enlace de contenido académico.

Crossvalidation: Un método utilizado en aprendizaje automático (*machine learning*) para evaluar y validar el rendimiento de un modelo utilizando diferentes subconjuntos de datos.

Cryptography: Una biblioteca de Python que proporciona herramientas para la implementación de algoritmos criptográficos, como el cifrado y la generación de firmas digitales.

CSS: Acrónimo de *Cascading Style Sheets* (Hojas de Estilo en Cascada). Es un lenguaje utilizado para describir la presentación y el estilo de un documento HTML.

Cuartil: En estadística, un cuartil es un valor que divide un conjunto de datos ordenados en cuatro partes iguales. Los cuartiles se utilizan para analizar y describir la distribución de un conjunto de datos, especialmente en el contexto de medidas de tendencia central y dispersión.

D3js: Una biblioteca JavaScript para la creación de visualizaciones de datos interactivas y dinámicas en páginas web.

DataTables: Un complemento de JavaScript que proporciona funcionalidad avanzada de tablas interactivas y filtrado en páginas web.

DOI: Acrónimo de *Digital Object Identifier* (Identificador de Objeto Digital). Es un identificador único utilizado para identificar de manera persistente un objeto digital, como un artículo científico.

Dropdown: Un *dropdown*, también conocido como menú desplegable o lista desplegable, es un elemento de interfaz gráfica que permite al usuario seleccionar una opción de una lista desplegable de opciones.

Enrutamiento: El enrutamiento (en el contexto de los *frameworks web*) se refiere a la forma en que una aplicación web maneja y responde a las solicitudes de los clientes para diferentes URL o rutas. En otras palabras, el enrutamiento determina cómo se mapean las

URL entrantes a las funciones o controladores correspondientes en la aplicación.

Flask: Un *framework* de desarrollo web minimalista y flexible para Python. Permite construir aplicaciones web rápidas y escalables.

flask-babel: Una extensión de Flask que simplifica la internacionalización y localización de aplicaciones web.

Flask-Cors: Una extensión de Flask que permite manejar la política de intercambio de recursos entre dominios (CORS) en aplicaciones web.

Flask-Login: Una extensión de Flask que proporciona funcionalidad de autenticación de usuarios y gestión de sesiones.

Flask-Migrate: Una extensión de Flask que facilita la gestión de migraciones de base de datos utilizando SQLAlchemy.

Frontend: La parte de un sistema o aplicación que interactúa directamente con los usuarios finales. Se refiere a la interfaz de usuario y la presentación visual.

Frossrefapi: Un paquete de Python que proporciona una interfaz para interactuar con la API de Crossref y acceder a metadatos de publicaciones académicas.

Frozenlist: Un paquete de Python que proporciona una lista inmutable (no modificable) que garantiza la integridad de los datos.

Git: El sistema de control de versiones más popular. Registra los cambios en los archivos de los programas, permitiendo la gestión de distintas versiones de código y la colaboración entre desarrolladores.

Google Scholar: Un motor de búsqueda especializado en literatura académica y científica. Proporciona acceso a artículos, tesis, resúmenes y otros recursos académicos.

Gráfico *burndown*: Un gráfico utilizado en la metodología ágil para representar el progreso de un proyecto a lo largo del tiempo, mostrando el trabajo pendiente y la tendencia de finalización.

Gunicorn: Un servidor web HTTP para aplicaciones Python. Es comúnmente utilizado para implementar aplicaciones Flask en producción.

Heroku: Una plataforma en la nube que permite implementar, gestionar y escalar aplicaciones web.

HTML: Acrónimo de *Hypertext Markup Language* (Lenguaje de Marcado de Hipertexto). Es el lenguaje estándar utilizado para crear páginas web.

Httpcore: Una biblioteca Python que proporciona una interfaz para realizar solicitudes HTTP de bajo nivel.

Httpx: Una biblioteca Python que proporciona una interfaz de alto nivel para realizar solicitudes HTTP.

Idna: Un módulo de Python que proporciona herramientas para la manipulación de nombres de dominio internacionalizados (IDN).

IF: Acrónimo de *Impact Factor* (Factor de Impacto). Es una métrica utilizada para evaluar la importancia relativa de una revista científica dentro de su campo.

Índice de Impacto: Una medida que indica la influencia relativa de una revista científica. Se basa en la frecuencia con la que los artículos de la revista son citados por otros investigadores.

Ipykernel: Un kernel de IPython que permite ejecutar código Python en el entorno de Jupyter Notebook.

Ipython: Un entorno interactivo de programación y exploración de datos que proporciona características adicionales sobre el intérprete de Python estándar.

JavaScript: Un lenguaje de programación interpretado utilizado principalmente para agregar interactividad y dinamismo a las páginas web.

JCR: Acrónimo de *Journal Citation Reports*. Es una base de datos que recopila información sobre las citas y el impacto de las revistas científicas.

Jedi: Una biblioteca de Python que proporciona funcionalidades avanzadas de autocompletado y análisis estático para el código Python.

Jinja2: Un motor de plantillas de Python utilizado en el framework Flask para generar contenido dinámico en las aplicaciones web.

Joblib: Una biblioteca de Python utilizada para la serialización y paralelización de tareas, especialmente en el contexto de machine learning.

jQuery: Una biblioteca de JavaScript rápida, pequeña y rica en características que simplifica la manipulación y el manejo de eventos en el código HTML.

JSON: JavaScript Object Notation, lenguaje de marcado que permite el intercambio de datos con texto sencillo.

Jupyter_client: Un cliente de Jupyter Notebook que permite la comunicación con los kernels y el manejo de sesiones interactivas.

Jupyter_core: Un conjunto de funcionalidades esenciales para el funcionamiento de los entornos Jupyter, incluyendo la gestión de notebooks y configuraciones.

Kanban: Una metodología ágil de gestión de proyectos que se centra en la visualización del flujo de trabajo y la optimización de la productividad.

L^AT_EX: Un sistema de composición de documentos utilizado principalmente para la creación de documentos científicos y técnicos con alta calidad tipográfica.

Logo: Abreviatura de logotipo. Es un símbolo gráfico o una representación visual que se utiliza para identificar una marca, empresa, organización o producto.

Machine Learning: Un campo de estudio de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y tomar decisiones basadas en datos.

MariaDB: Un sistema de gestión de bases de datos relacionales (RDBMS) derivado de MySQL. Ofrece compatibilidad con la mayoría de las características de MySQL y mejoras adicionales.

Matplotlib: Una biblioteca de Python ampliamente utilizada para la creación de gráficos estáticos, gráficos 2D y 3D, y visualizaciones de datos.

Matplotlib-inline: Una extensión de Jupyter Notebook que permite la visualización de gráficos Matplotlib de forma integrada en el notebook.

Metadatos: Datos que describen características o propiedades de un objeto. En el contexto de los documentos académicos, los metadatos pueden incluir información sobre los autores, título, resumen, fecha de publicación, etc.

MIAR: Acrónimo de *Matriz de Información para el Análisis de Revistas*. Es una base de datos en línea que proporciona información sobre las revistas académicas y científicas.

Miniconda: Una versión minimalista de Anaconda, un sistema de gestión de entornos y paquetes de Python, que proporciona solo los componentes esenciales.

Modularizado: Práctica que consiste en dividir un sistema o programa en módulos o componentes más pequeños y autónomos. Cada módulo se ocupa de una funcionalidad específica y se puede desarrollar, mantener y probar de forma independiente.

MySQL: Gestor de base de datos relacional ampliamente utilizado para aplicaciones web.

NumPy: Una biblioteca de Python utilizada para realizar operaciones matemáticas y numéricas eficientes en matrices y arreglos multidimensionales.

Open Source: Un término que se refiere a programas o software cuyo código fuente es de acceso público y puede ser utilizado, modificado y distribuido por cualquier persona.

Pandas: Una biblioteca de Python utilizada para el análisis y manipulación de datos estructurados. Proporciona estructuras de datos y herramientas para el manejo eficiente de tablas y series temporales.

Paquete: Un conjunto de módulos y archivos relacionados que se agrupan y distribuyen juntos para facilitar su uso y reutilización en aplicaciones de software.

PHP: Lenguaje de programación interpretado utilizado para desarrollo web del lado del servidor.

Pickleshare: Un módulo de Python que proporciona una forma sencilla de almacenar y compartir objetos de Python mediante la serialización y deserialización utilizando el formato "pickle".

Postgres: Un sistema de gestión de bases de datos relacional de código abierto (RDBMS) conocido también como PostgreSQL.

Product Owner: Un rol en la metodología ágil Scrum que representa a los interesados y es responsable de gestionar el backlog del producto y priorizar las funcionalidades.

Psycopg2: Un adaptador de base de datos de PostgreSQL para Python que permite interactuar con bases de datos PostgreSQL utilizando Python.

PyJWT: Una biblioteca de Python que proporciona herramientas para la codificación y decodificación de JSON Web Tokens (JWT).

Pytest: Un marco de pruebas de Python que facilita la escritura y ejecución de pruebas unitarias, de integración y funcionales.

Requests: Una biblioteca de Python utilizada para realizar solicitudes HTTP de manera sencilla y eficiente.

Scholarly: Un paquete de Python que proporciona una interfaz para interactuar con la API de Google Scholar y obtener información sobre publicaciones académicas.

Scikit-learn: Una biblioteca de Python ampliamente utilizada para el aprendizaje automático (machine learning). Proporciona una amplia gama de algoritmos y herramientas para el análisis de datos y la construcción de modelos predictivos.

Scipy: Una biblioteca de Python utilizada para el cálculo científico y el análisis de datos. Proporciona funcionalidades para el álgebra lineal, estadísticas, optimización y más.

Scopus: Una base de datos bibliográfica y de citas que contiene información sobre publicaciones académicas, patentes y conferencias.

Scrum Master: Un rol en la metodología ágil Scrum responsable de facilitar el proceso y asegurar que el equipo siga las prácticas y principios de Scrum.

Scrum: Un marco de trabajo ágil para la gestión y desarrollo de proyectos. Se centra en la colaboración, la adaptabilidad y la entrega incremental.

Selenium: Una suite de herramientas utilizada para la automatización de pruebas en aplicaciones web. Permite controlar navegadores web y simular interacciones de los usuarios.

Sprints: Iteraciones cortas y enfocadas en las que se realiza el trabajo en un proyecto ágil. Cada sprint tiene una duración fija y al final de cada uno se entrega un incremento potencialmente entregable del producto.

SQLAlchemy: Una biblioteca de Python que proporciona una capa de abstracción sobre los motores de bases de datos relacionales. Permite interactuar con bases de datos utilizando código Python en lugar de SQL directamente.

Story Points: Una medida utilizada en la metodología ágil para estimar el esfuerzo o la complejidad de una tarea. Ayuda a determinar la capacidad de trabajo de un equipo en un *sprint*.

Web of Science: Una base de datos bibliográfica y de citas ampliamente utilizada que proporciona información sobre publicaciones científicas, conferencias y patentes.

Web Scraping: La extracción de datos de páginas web de forma automatizada mediante el uso de programas o *scripts*.

XGBoost: Una biblioteca de aprendizaje automático (*machine learning*) que se enfoca en el algoritmo de Gradient Boosting.

ZenHub: Una herramienta de gestión de proyectos basada en GitHub que proporciona características adicionales como tableros Kanban, seguimiento de problemas y más.

Bibliografía

- [1] Agencia tributaria: Irpf. <https://sede.agenciatributaria.gob.es/Sede/irpf.html>.
- [2] Agencia tributaria: Retenciones. <https://sede.agenciatributaria.gob.es/Sede/Retenciones.shtml>.
- [3] Debian. our philosophy. <https://www.debian.org/intro/philosophy>.
- [4] Jornada laboral, permisos y vacaciones - condiciones de trabajo, incluyendo trabajadores desplazados - trabajo y jubilación - ciudadanos - tus derechos y obligaciones en la ue - tu espacio europeo - punto de acceso general. https://administracion.gob.es/pag_Home/Tu-espacio-europeo/derechos-obligaciones/ciudadanos/trabajo-jubilacion/condiciones-trabajo/jornada-permisos.html#-488e2841e18d.
- [5] Payscale. product owner salary. https://www.payscale.com/research/ES/Job=Product_Owner/Salary.
- [6] Payscale. scrum manager salary. <https://www.payscale.com/research/ES/Job=ScrumMaster/Salary>.
- [7] Seguridad social: Cotización. <https://www.seg-social.es/wps/portal/wss/internet/Trabajadores/CotizacionRecaudacionTrabajadores/10721/10957/583>.
- [8] Trámites para la puesta en marcha. <http://www.ipyme.org/es-ES/creaciondelaempresa/Paginas/Tramites.aspx>.

- [9] Jobted. slario programador, 2023. <https://www.jobted.es/salario/programador>.
- [10] Team Asana. Historias de usuario: 3 ejemplos para generar valor para el usuario [2022] • asana, Jan 2022. <https://asana.com/es/resources/user-stories>.
- [11] Anthony E. Boardman, David H. Greenberg, Aidan R. Vining, and David L. Weimar. *Cost-benefit analysis: Concepts and practice*. Prentice Hall, 2006.
- [12] Jorge Domínguez Chávez. *CLIENTE PSQL DE POSTGRESQL*. 04 2020.
- [13] Julia Martins. ¿qué es la metodología kanban y cómo funciona? [2022] • asana, Oct 2022. <https://asana.com/es/resources/what-is-kanban>.
- [14] Marta Palacio. *Scrum Master. Temario troncal 1*. 06 2021.
- [15] Rosa-Clark. Rest api - crossref, Apr 2020. <https://www.crossref.org/documentation/retrieve-metadata/rest-api/>.