



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería  
Informática**

**Sistema de Recomendación  
basado en Aprendizaje  
Profundo  
Documentación Técnica**



Presentado por Raúl Negro Carpintero  
en Universidad de Burgos — 3 de julio  
de 2019

Tutor: Bruno Baruque Zanón



---

# Índice general

---

|   |            |
|---|------------|
| <b>Índice general</b>                                   | <b>I</b>   |
| <b>Índice de figuras</b>                                | <b>III</b> |
| <b>Índice de tablas</b>                                 | <b>v</b>   |
| <b>Apéndice A Plan de Proyecto Software</b>             | <b>1</b>   |
| A.1. Introducción . . . . .                             | 1          |
| A.2. Planificación temporal . . . . .                   | 2          |
| A.3. Estudio de viabilidad . . . . .                    | 4          |
| <b>Apéndice B Especificación de Requisitos</b>          | <b>9</b>   |
| B.1. Introducción . . . . .                             | 9          |
| B.2. Objetivos generales . . . . .                      | 9          |
| B.3. Catalogo de requisitos . . . . .                   | 9          |
| B.4. Especificación de requisitos . . . . .             | 11         |
| <b>Apéndice C Especificación de diseño</b>              | <b>27</b>  |
| C.1. Introducción . . . . .                             | 27         |
| C.2. Diseño de datos . . . . .                          | 27         |
| C.3. Diseño procedimental . . . . .                     | 29         |
| C.4. Diseño arquitectónico . . . . .                    | 31         |
| C.5. Diseño de la interfaz . . . . .                    | 36         |
| <b>Apéndice D Documentación técnica de programación</b> | <b>39</b>  |
| D.1. Introducción . . . . .                             | 39         |
| D.2. Estructura de directorios . . . . .                | 39         |

|  |           |
|--|-----------|
| D.3. Manual del programador . . . . .                            | 40        |
| D.4. Compilación, instalación y ejecución del proyecto . . . . . | 41        |
| D.5. Pruebas del sistema . . . . .                               | 44        |
| <b>Apéndice E Documentación de usuario</b>                       | <b>45</b> |
| E.1. Introducción . . . . .                                      | 45        |
| E.2. Requisitos de usuarios . . . . .                            | 45        |
| E.3. Instalación . . . . .                                       | 45        |
| E.4. Manual del usuario . . . . .                                | 48        |
| <b>Bibliografía</b>  | <b>61</b> |

---

# Índice de figuras

---

|  |    |
|--|----|
| B.1. Diagrama en caso de obtener un nuevo modelo . . . . .           | 12 |
| B.2. Diagrama en caso de obtener los resultados y las métricas . . . | 13 |
| B.3. Diagrama en caso de querer añadir valoraciones . . . . .        | 14 |
| B.4. Diagrama en caso de mostrar la ayuda . . . . .                  | 15 |
| C.1. Diagrama de secuencia para obtener un nuevo modelo . . . . .    | 30 |
| C.2. Diagrama de secuencia para obtener métricas y predicciones . .  | 30 |
| C.3. Diagrama de secuencia para añadir valoraciones . . . . .        | 31 |
| C.4. Diagrama UML del proyecto . . . . .                             | 32 |
| C.5. Diagrama UML del paquete <i>vista</i> . . . . .                 | 33 |
| C.6. Diagrama UML del paquete <i>controlador</i> . . . . .           | 34 |
| C.7. Diagrama UML del paquete <i>modelo</i> . . . . .                | 35 |
| C.8. Interfaz web . . . . .  | 37 |
| D.1. Descargar proyecto . . . . .                                    | 42 |
| D.2. Navegación hasta <i>src</i> . . . . .                           | 43 |
| D.3. Ejecución del proyecto . . . . .                                | 43 |
| D.4. Interfaz web de la aplicación . . . . .                         | 44 |
| E.1. Descargar proyecto . . . . .                                    | 46 |
| E.2. Navegación hasta <i>src</i> . . . . .                           | 46 |
| E.3. Ejecución del proyecto . . . . .                                | 47 |
| E.4. Interfaz web de la aplicación . . . . .                         | 47 |
| E.5. Página principal de la GUI . . . . .                            | 48 |
| E.6. Elegir entre modelo clásico o basado en Aprendizaje Profundo .  | 49 |
| E.7. Elegir 1 de los 3 modelos clásicos . . . . .                    | 50 |
| E.8. Parte de los parámetros disponibles para los modelos clásicos . | 51 |
| E.9. Página donde elegir si usar un dataset nuevo o no . . . . .     | 51 |

|  |    |
|--|----|
| E.10. Página donde elegir si usar un dataset nuevo o no . . . . .      | 52 |
| E.11. Página donde seleccionar los datasets ya guardados . . . . .     | 53 |
| E.12. Se pregunta si se quieren utilizar los timestamps o no . . . . . | 54 |
| E.13. Se pide seleccionar las matrices . . . . .                       | 55 |
| E.14. Métricas del modelo y datos de los datasets . . . . .            | 56 |
| E.15. Elegir usuario cuyas predicciones se quieren ver . . . . .       | 57 |
| E.16. Predicciones del usuario . . . . .                               | 58 |

---

# Índice de tablas

---

|  |    |
|--|----|
| A.1. Coste de personal . . . . .                           | 5  |
| A.2. Coste hardware . . . . .                              | 5  |
| A.3. Coste software . . . . .                              | 6  |
| A.4. Costes varios . . . . .                               | 6  |
| A.5. Coste total . . . . .                                 | 7  |
| A.6. Herramientas y sus licencias . . . . .                | 8  |
|  |    |
| B.1. CU-01 Gestión de los datos intermedios . . . . .      | 16 |
| B.2. CU-02 Guardar matrices . . . . .                      | 16 |
| B.3. CU-3 Cargar matrices . . . . .                        | 17 |
| B.4. CU-4 Gestión de los modelos . . . . .                 | 17 |
| B.5. CU-5 Mostrar modelos . . . . .                        | 18 |
| B.6. CU-6 Seleccionar modelos . . . . .                    | 18 |
| B.7. CU-7 Guardar modelos . . . . .                        | 19 |
| B.8. CU-8 Cargar modelos . . . . .                         | 19 |
| B.9. CU-9 Gestión de los resultados . . . . .              | 20 |
| B.10.CU-10 Guardar resultados . . . . .                    | 20 |
| B.11.CU-11 Mostrar resultados . . . . .                    | 21 |
| B.12.CU-12 Mostrar predicciones . . . . .                  | 21 |
| B.13.CU-13 Gestión de los usuarios . . . . .               | 22 |
| B.14.CU-14 Añadir valoraciones . . . . .                   | 23 |
| B.15.CU-15 Gestión de los conjuntos de datos . . . . .     | 23 |
| B.16.CU-16 Mostrar los conjuntos de datos . . . . .        | 24 |
| B.17.CU-17 Seleccionar el conjunto de datos . . . . .      | 24 |
| B.18.CU-18 Añadir conjunto de datos . . . . .              | 25 |
| B.19.CU-19 Ayuda de la aplicación . . . . .                | 25 |
| B.20.CU-20 Mostrar información sobre los modelos . . . . . | 26 |

|   |    |
|---|----|
| B.21.CU-21 Mostrar información sobre los conjuntos de datos . . . . | 26 |
| C.1. Formato ratings.csv . . . . .                                  | 28 |
| C.2. Formato users.csv . . . . .                                    | 28 |
| C.3. Formato items.csv . . . . .                                    | 28 |



## Apéndice A

---

# Plan de Proyecto Software

---

### A.1. Introducción

La planificación es un punto fundamental en cualquier proyecto software. En ella, se tiene que estimar qué cosas hay que hacer y cuánto tiempo y dinero va a llevar terminirlas. De esta manera, se podrá comprobar si se llega a tiempo a los distintos plazos y así evitar retrasos, lo que supondría pérdidas de dinero y tiempo.

La planificación se puede dividir en dos partes clave:

- **Planificación temporal:** consiste en dividir el proyecto en etapas (*sprints*) con una duración de x días. En estas etapas hay que dejar claro qué cosas se quieren conseguir y cuánto tiempo o esfuerzo se va a dedicar a cada una de ellas. Los *sprints* no tienen por qué durar siempre lo mismo, se pueden ajustar a las necesidades de cada momento.
- **Estudio de viabilidad:** con este estudio se podrá ver si el proyecto puede seguir adelante (es viable) o si no, con lo que habría que tomar medidas para cambiar la situación. Se divide en:
  - **Viabilidad económica:** según la cual se deberá calcular por cuánto tiene que venderse el servicio para compensar el dinero, el tiempo y el esfuerzo puestos en el proyecto. Se incluye salario del personal y la compra de software de terceros necesario para la realización del proyecto.
  - **Viabilidad legal:** se estudia el licenciamiento del software que se va a utilizar para proceder dentro de la legalidad.

## A.2. Planificación temporal

Para el desarrollo del proyecto se decidió seguir la metodología ágil *Scrum*. Al ser un proyecto académico con una sola persona trabajando en él, no se siguen a raja tabla todos los rasgos característicos de *Scrum*. Las características que se han seguido son:

- Desarrollo incremental en cada *sprint*.
- Cada *sprint* tiene una duración aproximada de dos semanas.
- Se realizan reuniones al final de cada *sprint* con el objetivo de repasar lo hecho y pensar en los objetivos del siguiente *sprint*.

A continuación se desarrolla el contenido de cada *sprint*.

### Sprint 1 (- 13/12/2018)

Este *sprint* se corresponde con el milestone **Búsqueda de información inicial**. En él se estudia el capítulo dedicado a los sistemas de recomendación del libro *Mining of Massive Datasets* [13].

También se realiza el curso de *fast.ai* [7]. Además, se escoge la librería de *LightFM* y se obtiene un modelo inicial.

### Sprint 2 (13/12/2018 - 18/01/2019)

Este *sprint* se corresponde con el milestone **Explorar modelo con LightFM**. En él se intenta comprender el funcionamiento de *LightFM* y se utiliza el conjunto de datos de *Movielens* con el modelo.

### Sprint 3 (18/01/2019 - 16/02/2019)

Este *sprint* se corresponde con el milestone **Recomendación híbrida y por contenido con LightFM**. En él se obtienen las primeras versiones del modelo híbrido y por contenido de *LightFM*.

También se obtiene un modelo inicial con *PyTorch* y se empieza a pensar en las métricas que se usarán para evaluar los modelos.

### Sprint 4 (16/02/2019 - 03/03/2019)

Este *sprint* se corresponde con el milestone **Trabajar en la documentación**. En él se empieza a ampliar los apartados de la documentación relativos a los conjuntos de datos utilizados.

**Sprint 5 (03/03/2019 - 13/03/2019)**

Este *sprint* se corresponde con el milestone **División de los datos en LightFM**. En él se dividen los conjuntos de datos en *train* y *test* y se obtienen de nuevo los modelos.

**Sprint 6 (13/03/2019 - 28/03/2019)**

Este *sprint* se corresponde con el milestone **Medidas de calidad LightFM**. En él se obtienen las métricas finales para los modelos de *LightFM*.

Además, se guardan las matrices y los modelos en archivos *pickle* como medida de persistencia.

**Sprint 7 (28/03/2019 - 05/04/2019)**

Este *sprint* se corresponde con el milestone **Últimos pasos con LightFM**. En él se intenta acabar todo lo relacionado con los modelos clásicos (tanto código como documentación).

**Sprint 8 (05/04/2019 - 24/04/2019)**

Este *sprint* se corresponde con el milestone **Ampliar documentación**. En él se amplía la documentación relativa a *LightFM* y se empiezan a obtener los primeros diagramas UML con la estructura del proyecto hasta el momento.

**Sprint 9 (24/04/2019 - 08/05/2019)**

Este *sprint* se corresponde con el milestone **Cambios en el diseño, Flask y DL**. En él se empieza a obtener la interfaz web con *Flask* y se escoge *Spotlight* como la librería de aprendizaje profundo.

También se modifica la documentación para ir acorde a los cambios del código.

**Sprint 10 (08/05/2019 - 22/05/2019)**

Este *sprint* se corresponde con el milestone **Primeros pasos con Spotlight**. En él se obtiene un primer modelo de *Spotlight* y se unifican los archivos de Entrada y Persistencia para que solo haya uno de cada.

Se continúa trabajando en la documentación.

### **Sprint 11 (22/05/2019 - 05/06/2019)**

Este *sprint* se corresponde con el milestone **Completar modelos de Spotlight**. En él se obtienen el resto de modelos de *Spotlight* y se generalizan los métodos de lectura de datos y obtención de matrices.

Se continúa trabajando en la documentación.

### **Sprint 12 (05/06/2019 - 12/06/2019)**

Este *sprint* se corresponde con el milestone **Diseño GUI**. En él se piensa qué forma va a tener la interfaz web obtenida a través de *Flask*.

### **Sprint 13 (12/06/2019 - 20/06/2019)**

Este *sprint* se corresponde con el milestone **Obtención GUI**. En él se termina de crear la interfaz web y se empiezan a emplear los métodos obtenidos anteriormente para que la aplicación sea funcional.

### **Sprint 14 (20/06/2019 - 03/07/2019)**

Este *sprint* se corresponde con el milestone **Último milestone**. En él se termina todo lo que queda por hacer en el proyecto. La interfaz es totalmente funcional y la documentación está terminada.

## **A.3. Estudio de viabilidad**

### **Viabilidad económica**

En este apartado se emularán los costes que hubiera tenido una empresa si hubiera dedicado tiempo y recursos a desarrollar este proyecto, así como el beneficio que podría sacar gracias a su venta.

Por lo tanto, los dos elementos fundamentales son: los costes y los beneficios.

### **Costes**

Los costes se pueden dividir de la siguiente manera:

**Costes humanos:**

En este apartado se incluye el sueldo de los trabajadores. Como el proyecto se ha podido llevar a cabo por una sola persona, se van a calcular los costes como si el equipo de trabajo estuviera compuesto por una persona. Además, la situación de esta persona sería la de recién graduado y sin experiencia laboral en este campo. La duración sería de unos 5 meses.

Según un informe de *Adecco* en 2015 [5], los informáticos junior cobraban alrededor de 17.248 € brutos al año. Sabiendo esto, el desglose quedaría:

| Concepto              | Coste             |
|-----------------------|-------------------|
| Salario mensual neto  | 1.202,76 €        |
| Seguridad Social      | 91,27 €           |
| Retención IRPF        | 143,31 €          |
| Salario mensual bruto | 1.437,33 €        |
| <b>Total 5 meses</b>  | <b>7.186,65 €</b> |

Tabla A.1: Coste de personal

Los datos se han calculado gracias a una herramienta proporcionada por la sección de *El País Economía* [1]. Para ello, se tiene en cuanto el salario bruto anteriormente proporcionado así como la categoría profesional del trabajador, su edad y la comunidad en la que reside.

#### Coste hardware:

En este apartado se tiene en cuenta el coste de todos los elementos hardware que se han necesitado durante la realización del proyecto. Normalmente, estos elementos se amortizan a los 5 años.

| Concepto                    | Coste            | Coste amortizado 5 meses |
|-----------------------------|------------------|--------------------------|
| Ordenador Personal (Bundle) | 1.200 €          | 100 €                    |
| Hardware alquilado (al mes) | 437,34 €         | -                        |
| <b>Total</b>                | <b>3.386,7 €</b> | <b>2.286,7 €</b>         |

Tabla A.2: Coste hardware

Ya que el proyecto involucra al aprendizaje profundo, sería interesante alquilar un buen hardware en el que ejecutar la aplicación.

A decir verdad, esta es la peor situación posible, ya que se está pagando de continuo por un hardware [?] que va a estar apagado la mayor parte del tiempo. Sería más rentable alquilarlo por semanas o minutos.

#### Coste software:

En este apartado se tiene en cuenta el software no gratuito. La amortización del software suele estar por los 2 años.

| Concepto       | Coste        | Coste amortizado 5 meses |
|----------------|--------------|--------------------------|
| Windows 10 Pro | 259 €        | 53,96 €                  |
| <b>Total</b>   | <b>259 €</b> | <b>53,96 €</b>           |

Tabla A.3: Coste software

Este caso también es poco favorable, ya que con el equipo comprado anteriormente podría venir Windows instalado. Además, se podría optar por un equipo con Linux y ahorrarse directamente el coste del S.O.

#### Costes varios:

Resto de costes que no tienen un grupo dedicado para ellos solos.

| Concepto                   | Coste           |
|----------------------------|-----------------|
| Alquiler oficina al mes    | 450 €           |
| Conexión a internet al mes | 80 €            |
| Material de oficina        | 150 €           |
| <b>Total</b>               | <b>11.800 €</b> |

Tabla A.4: Costes varios

#### Coste total:

Teniendo en cuenta todos los costes anteriores, desarrollar este proyecto costaría 21.327,31 €.

| Concepto     | Coste              |
|--------------|--------------------|
| Humano       | 7.186,65 €         |
| Hardware     | 2.286,7 €          |
| Software     | 53,96 €            |
| Varios       | 11.800 €           |
| <b>Total</b> | <b>21.327,31 €</b> |

Tabla A.5: Coste total

Esta cifra podría verse reducida mucho si se elige otro tipo de plan al alquilar el hardware.

## Beneficios

Una buena manera de poner precio a los sistemas de recomendación sería cobrar por predicción. Al poder utilizar casi cualquier conjunto de datos, cualquier cliente podría utilizar el sistema y obtener predicciones para los usuarios de su plataforma.

A estos clientes se les podría cobrar a  $x$  céntimos la predicción. O incluso se podría ofrecer un contrato para que el cliente no tuviera que pagar cada vez que quisiera predicciones, sino que pagara un tipo de suscripción.

## Viabilidad legal

Al desarrollar software, hay que tener muy presente qué herramientas se están utilizando y bajo qué licencias se han publicado.

En este proyecto, todas las librerías y herramientas que se han utilizado han sido gratuitas. A continuación se listan todas ellas junto con su tipo de licencia:

| Software | Descripción                  | Licencia   |
|----------|------------------------------|------------|
| LightFM  | Librería de modelos clásicos | Apache 2.0 |

continúa en la página siguiente

continúa desde la página anterior

| Software                        | Descripción  | Licencia          |
|---------------------------------|--|-------------------|
| Spotlight                       | Librería de modelos de deep learning                 | MIT               |
| NumPy                           | Paquete de Python utilizado en ciencia computacional | BSD               |
| Pandas                          | Herramienta usada para tratar los .csv               | BSD               |
| PyTorch                         | Librería de deep learning                            | BSD               |
| CUDA                            | Librería de programación paralela de nVidia          | Propietario, Free |
| Python                          | Lenguaje de programación usado en el proyecto        | PSFL              |
| Flask                           | Herramienta utilizada para la interfaz web           | BSD               |
| WTForms                         | Herramienta con la que poder obtener formularios web | BSD               |
| Werkzeug                        | Asegura el nombre de los ficheros a guardar          | BSD               |
| Gunicorn                        | Servidor Web para Python                             | MIT               |
| Sublime Text                    | Editor de Texto                                      | Propietario, Free |
| Texmaker                        | Editor para la documentación                         | GNU GPL           |
| L <sup>A</sup> T <sub>E</sub> X | Utilizado en la documentación                        | LPPL              |
| Anaconda                        | Utilizado con Jupyter y Spyder                       | BSD               |

Tabla A.6: Herramientas y sus licencias



## *Apéndice B*

---

# Especificación de Requisitos

---

### B.1. Introducción

Este anexo recoge los objetivos generales y la especificación de requisitos del proyecto.

### B.2. Objetivos generales

El proyecto persigue los siguientes objetivos generales:

- Comprender los sistemas de recomendación tanto clásicos como basados en aprendizaje profundo.
- Recoger y evaluar los resultados obtenidos por los dos modelos sobre diferentes conjuntos de datos.
- Comparar los resultados.

### B.3. Catalogo de requisitos

Los requisitos derivados de los objetivos del proyecto son los siguientes:

#### Requisitos funcionales

- **RF-1 Gestión de los datos intermedios:** el programa tiene que ser capaz de gestionar los datos intermedios:

- **RF-1.1 Guardar matrices:** el programa tiene que ser capaz de guardar las matrices generadas por los modelos para ahorrar tiempo en siguientes ejecuciones.
- **RF-1.2 Cargar matrices:** el programa tiene que ser capaz de cargar las matrices que el usuario quiera.
- **RF-2 Gestión de los modelos:** el programa tiene que ser capaz de gestionar los modelos:
  - **RF-2.1 Mostrar modelos:** el programa tiene que ser capaz de mostrar todos los tipos de modelos disponibles.
  - **RF-2.2 Seleccionar modelos:** el programa tiene que ser capaz de dejar al usuario seleccionar el tipo de modelo que quiera.
  - **RF-2.3 Guardar modelos:** el programa tiene que ser capaz de guardar los modelos obtenidos para ahorrar tiempo en futuras ejecuciones.
  - **RF-2.4 Cargar modelos:** el programa tiene que ser capaz de cargar los modelos previamente guardados.
- **RF-3 Gestión de los resultados:** el programa tiene que ser capaz de gestionar los resultados:
  - **RF-3.1 Guardar resultados:** el programa tiene que ser capaz de guardar los resultados generados por los modelos.
  - **RF-3.2 Mostrar resultados:** el programa tiene que ser capaz de mostrar los resultados obtenidos por los distintos modelos.
  - **RF-3.3 Mostrar predicciones:** el programa tiene que ser capaz de mostrar las predicciones obtenidas por los distintos modelos.
- **RF-4 Gestión de los usuarios:** el programa tiene que ser capaz de introducir las valoraciones de nuevos usuarios:
  - **RF-4.1 Añadir valoraciones:** el programa tiene que ser capaz de simular la entrada de nuevos usuarios permitiendo el añadido de nuevas valoraciones.
- **RF-5 Gestión de los conjuntos de datos:** el programa tiene que:
  - **RF-5.1 Mostrar los conjuntos de datos:** el programa tiene que ser capaz de mostrar los distintos conjuntos de datos de prueba.
  - **RF-5.2 Seleccionar el conjunto de datos:** el programa tiene que ser capaz de mostrar dejar que el usuario seleccione uno de los conjuntos de datos de prueba.
  - **RF-5.3 Añadir conjunto de datos:** el programa tiene que ser capaz de dejar que el usuario añada un conjunto de datos propio.

- **RF-6 Ayuda de la aplicación:** el programa tiene que ofrecer información al usuario:
  - **RF-6.1 Mostrar información sobre modelos:** el programa tiene que ser capaz de mostrar información sobre los distintos modelos que se pueden escoger.
  - **RF-6.2 Mostrar información sobre conjuntos de datos:** el programa tiene que ser capaz de mostrar información sobre los distintos conjuntos de datos de prueba.

### Requisitos no funcionales

- **RNF-1 Usabilidad:** la interfaz gráfica tiene que ser intuitiva y fácil de usar.
- **RNF-2 Soporte:** el programa tiene que dar soporte a versiones iguales o mayores a Python 3.
- **RNF-3 Localización:** el programa tiene que estar preparado para soportar varios idiomas.

## B.4. Especificación de requisitos

En esta sección se mostrará el diagrama de casos de uso y se desarrollará cada uno de ellos.

### Diagrama de casos de uso

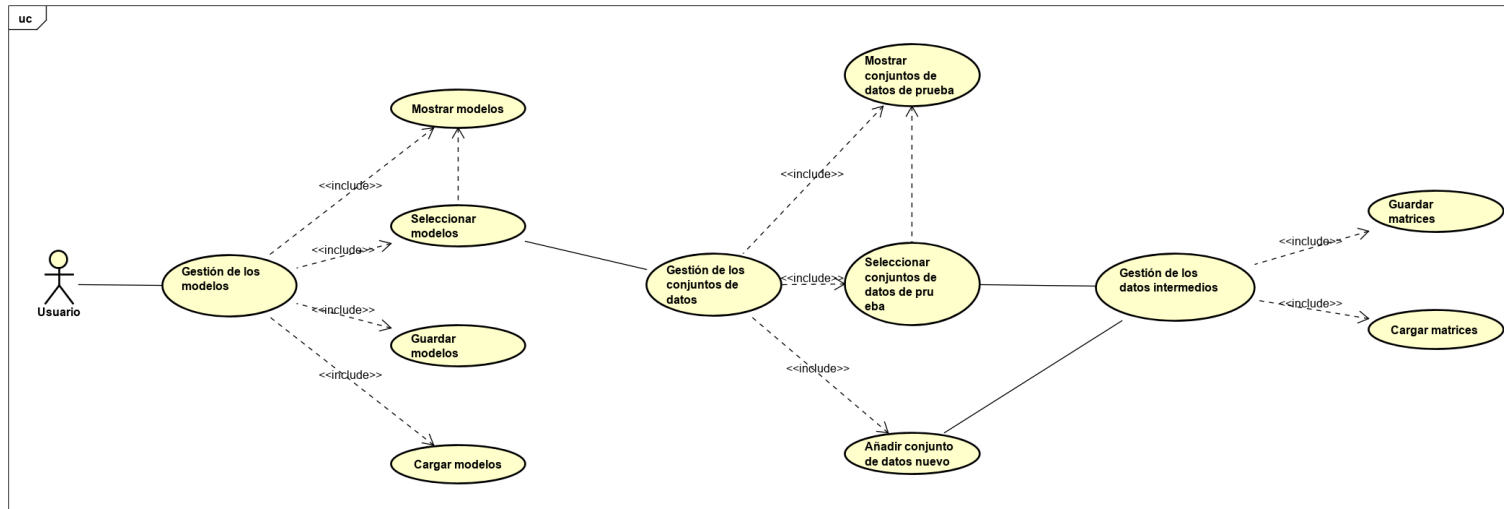


Figura B.1: Diagrama en caso de obtener un nuevo modelo

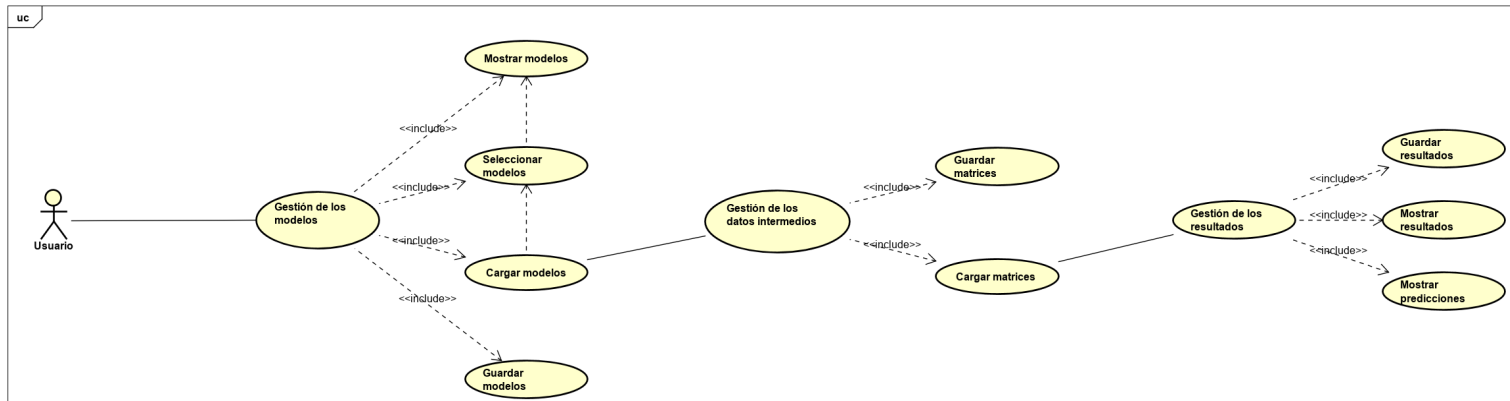


Figura B.2: Diagrama en caso de obtener los resultados y las métricas

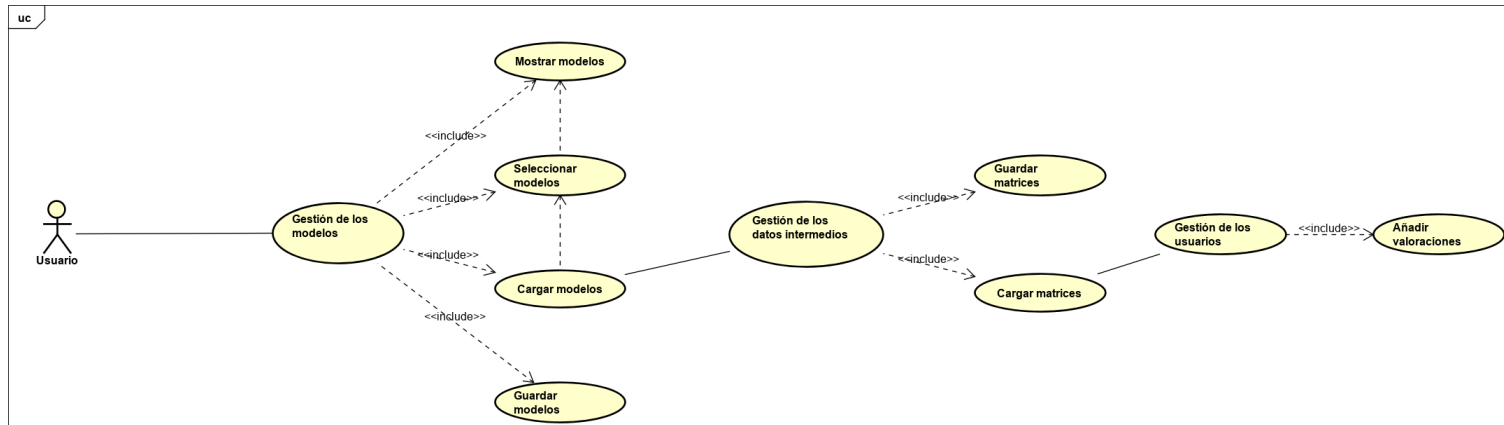


Figura B.3: Diagrama en caso de querer añadir valoraciones

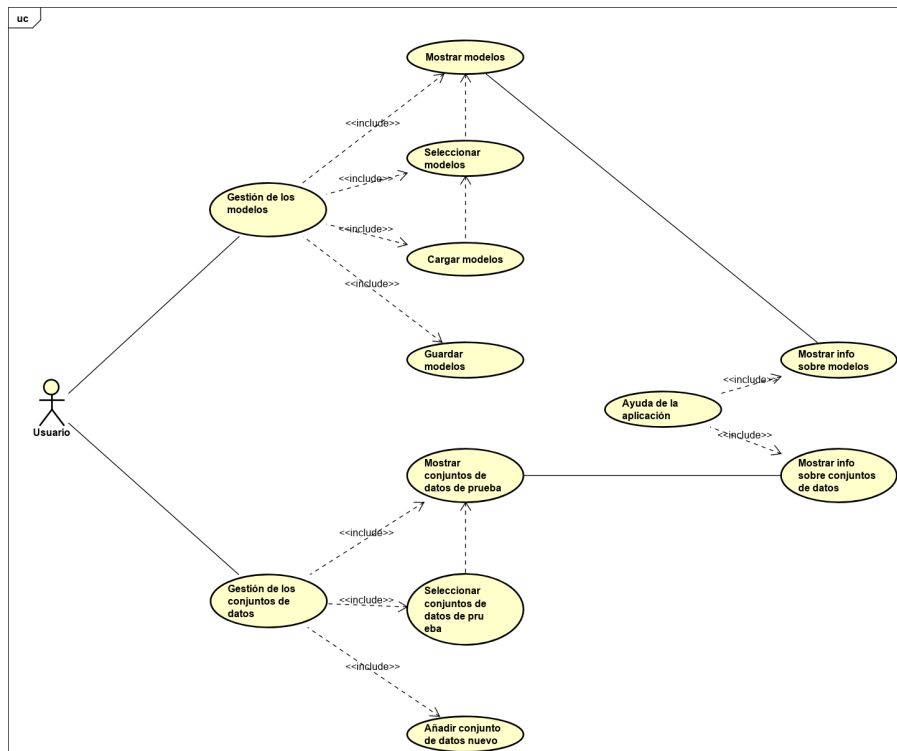


Figura B.4: Diagrama en caso de mostrar la ayuda

## Actores

Con la aplicación solo interactuará un actor, el usuario que esté probando la aplicación en un momento determinado.

## Casos de uso

A continuación, se desarrollará cada caso de uso:

---

CU-01

Gestión de los datos intermedios

continúa en la página siguiente

---

continúa desde la página anterior

| CU-01                       | Gestión de los datos intermedios                         |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero                                    |
| <b>Requisitos asociados</b> | RF-1 & RF-1.1 & RF-1.2                                   |
| <b>Descripción</b>          | Gestión de los datos intermedios.                        |
| <b>Precondiciones</b>       | Los datos intermedios tienen que existir.                |
| <b>Acciones</b>             | El usuario escoge los datos que quiere cargar o guardar. |
| <b>Postcondiciones</b>      | Los datos intermedios se cargan o se guardan.            |
| <b>Excepciones</b>          | Los datos intermedios no existen.                        |

Tabla B.1: CU-01 Gestión de los datos intermedios

| CU-02                       | Guardar matrices   |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero                                    |
| <b>Requisitos asociados</b> | RF-1 & RF-1.1  |
| <b>Descripción</b>          | Guardar las matrices de datos.                           |
| <b>Precondiciones</b>       | Tienen que existir las matrices.                         |
| <b>Acciones</b>             | El usuario guarda las matrices generadas por el sistema. |
| <b>Postcondiciones</b>      | Las matrices se guardan en archivos <i>.pickle</i>       |
| <b>Excepciones</b>          | Los matrices no existen.                                 |

Tabla B.2: CU-02 Guardar matrices

| CU-03                       | Cargar matrices       |
|-----------------------------|-----------------------|
| <b>Versión</b>              | 1.0                   |
| <b>Autor</b>                | Raúl Negro Carpintero |
| <b>Requisitos asociados</b> | RF-1 & RF-1.2         |

continúa en la página siguiente



continúa desde la página anterior

| CU-03                  | Cargar matrices  |
|------------------------|--|
| <b>Descripción</b>     | Cargar las matrices de datos.                                    |
| <b>Precondiciones</b>  | Tienen que existir los archivos <i>.pickle</i> con las matrices. |
| <b>Acciones</b>        | El usuario carga las matrices.                                   |
| <b>Postcondiciones</b> | Las matrices se cargan y se vinculan al sistema.                 |
| <b>Excepciones</b>     | No hay archivos <i>.pickle</i> con las matrices.                 |

Tabla B.3: CU-3 Cargar matrices

| CU-04                       | Gestión de los modelos  |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-2 & RF-2.1 & RF-2.2 & RF-2.3 & RF-2.4                                  |
| <b>Descripción</b>          | Gestión de los modelos.   |
| <b>Precondiciones</b>       | Se tienen que haber generado los modelos.                                 |
| <b>Acciones</b>             | El usuario escoge el tipo de modelo que quiere obtener, guardar o cargar. |
| <b>Postcondiciones</b>      | El modelo escogido se crea, guarda o carga.                               |
| <b>Excepciones</b>          | No hay modelos generados.   |

Tabla B.4: CU-4 Gestión de los modelos

| CU-05                       | Mostrar modelos   |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero                                       |
| <b>Requisitos asociados</b> | RF-2 & RF-2.1   |
| <b>Descripción</b>          | Se listan todos los tipos de modelos que se pueden obtener. |

continúa en la página siguiente

continúa desde la página anterior

| CU-05                  | Mostrar modelos                                |
|------------------------|--|
| <b>Precondiciones</b>  | -  |
| <b>Acciones</b>        | El usuario piensa que tipo de modelo utilizar. |
| <b>Postcondiciones</b> | -  |
| <b>Excepciones</b>     | No hay tipos de modelos que mostrar.           |

Tabla B.5: CU-5 Mostrar modelos

| CU-06                       | Seleccionar modelos  |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | RF-2 & RF-2.2  |
| <b>Descripción</b>          | Se selecciona un modelo.   |
| <b>Precondiciones</b>       | Se han mostrado todos los tipos de modelos.  |
| <b>Acciones</b>             | El usuario escoge el tipo de modelo que quiere crear, guardar o cargar.              |
| <b>Postcondiciones</b>      | El tipo de modelo que se quiere utilizar se queda seleccionado para trabajar con él. |
| <b>Excepciones</b>          | No existe el modelo seleccionado.  |

Tabla B.6: CU-6 Seleccionar modelos

| CU-07                       | Guardar modelos                               |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero                         |
| <b>Requisitos asociados</b> | RF-2 & RF-2.3                                 |
| <b>Descripción</b>          | Se guarda un modelo.                          |
| <b>Precondiciones</b>       | El modelo que se quiere guardar debe existir. |

continúa en la página siguiente

continúa desde la página anterior

| CU-07                  | Guardar modelos                                     |
|------------------------|---|
| <b>Acciones</b>        | Se guarda el modelo en un archivo <i>.pickle</i> .  |
| <b>Postcondiciones</b> | Se obtiene un archivo <i>.pickle</i> con el modelo. |
| <b>Excepciones</b>     | No hay ningún modelo que guardar.                   |

Tabla B.7: CU-7 Guardar modelos

| CU-08                       | Cargar modelos   |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | RF-2 & RF-2.4  |
| <b>Descripción</b>          | Se carga un modelo.  |
| <b>Precondiciones</b>       | El archivo <i>.pickle</i> con el modelo que se quiere cargar debe existir. |
| <b>Acciones</b>             | Se carga el modelo desde un archivo <i>.pickle</i> .                       |
| <b>Postcondiciones</b>      | Se obtiene el modelo.  |
| <b>Excepciones</b>          | No hay ningún archivo <i>.pickle</i> con el modelo deseado.                |

Tabla B.8: CU-8 Cargar modelos

| CU-09                       | Gestión de los resultados   |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-3 & RF-3.1 & RF-3.2 & RF-3.3   |
| <b>Descripción</b>          | Gestión de los resultados.  |
| <b>Precondiciones</b>       | El modelo tiene que estar creado.<br>Los resultados tienen que haber sido calculados. |

continúa en la página siguiente

continúa desde la página anterior

| CU-09                  | Gestión de los resultados   |
|------------------------|---|
| <b>Acciones</b>        | Tienen que existir los resultados.<br>Las predicciones tiene que haber sido calculadas.<br>Guardar los resultados.<br>Mostrar los resultados. |
| <b>Postcondiciones</b> | Mostrar las predicciones.<br>Se cargan los resultados.<br>Se muestran los resultados.   |
| <b>Excepciones</b>     | Se muestran las predicciones.<br>Los resultados no se han calculado.<br>Las predicciones no se han calculado.<br>No hay ningún modelo.        |

Tabla B.9: CU-9 Gestión de los resultados

| CU-10                       | Guardar resultados  |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-3 & RF-3.1   |
| <b>Descripción</b>          | Guardar resultados.   |
| <b>Precondiciones</b>       | Los resultados tienen que haber sido calculados.<br>El modelo tiene que estar creado. |
| <b>Acciones</b>             | Guardar los resultados.   |
| <b>Postcondiciones</b>      | Se guardan los resultados.  |
| <b>Excepciones</b>          | Los resultados no se han calculado.<br>No hay ningún modelo.                          |

Tabla B.10: CU-10 Guardar resultados

| CU-11 | Mostrar resultados |
|-------|--------------------|
|-------|--------------------|

continúa en la página siguiente

continúa desde la página anterior

| CU-11                       | Mostrar resultados  |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-3 & RF-3.2   |
| <b>Descripción</b>          | Mostrar los resultados.   |
| <b>Precondiciones</b>       | Los resultados tienen que haber sido calculados.<br>El modelo tiene que estar creado. |
| <b>Acciones</b>             | Mostrar los resultados.   |
| <b>Postcondiciones</b>      | Se muestran los resultados.   |
| <b>Excepciones</b>          | Los resultados no se han calculado.<br>No hay ningún modelo.                          |

Tabla B.11: CU-11 Mostrar resultados

| CU-12                       | Mostrar predicciones   |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | RF-3 & RF-3.3  |
| <b>Descripción</b>          | Mostrar las predicciones.  |
| <b>Precondiciones</b>       | El modelo tiene que estar creado.<br>Las predicciones tiene que haber sido calculadas. |
| <b>Acciones</b>             | Mostrar las predicciones.  |
| <b>Postcondiciones</b>      | Se muestran las predicciones.  |
| <b>Excepciones</b>          | Las predicciones no se han calculado.<br>No hay ningún modelo.                         |

Tabla B.12: CU-12 Mostrar predicciones

| CU-13 | Gestión de los usuarios |
|-------|-------------------------|
|-------|-------------------------|

continúa en la página siguiente

continúa desde la página anterior

| CU-13                       |         | Gestión de los usuarios  |
|-----------------------------|---------|--|
| <b>Versión</b>              |         | 1.0  |
| <b>Autor</b>                |         | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | asocia- | RF-4 & RF-4.1  |
| <b>Descripción</b>          |         | Gestión de los usuarios.   |
| <b>Precondiciones</b>       |         | Deben existir las matrices de datos.<br>Debe existir el usuario con el que se quieren añadir valoraciones. |
| <b>Acciones</b>             |         | Añadir valoraciones.   |
| <b>Postcondiciones</b>      |         | Se añaden valoraciones y/o usuarios a un conjunto de datos.  |
| <b>Excepciones</b>          |         | No existen las matrices de datos.<br>No existe el usuario al que se le quiere añadir valoraciones.         |

Tabla B.13: CU-13 Gestión de los usuarios

| CU-14                       |         | Añadir valoraciones  |
|-----------------------------|---------|--|
| <b>Versión</b>              |         | 1.0  |
| <b>Autor</b>                |         | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | asocia- | RF-4 & RF-4.1  |
| <b>Descripción</b>          |         | Añadir valoraciones de un usuario existente o nuevo.   |
| <b>Precondiciones</b>       |         | Deben existir las matrices de datos.<br>Debe existir el usuario con el que se quieren añadir valoraciones. |
| <b>Acciones</b>             |         | Añadir valoraciones sobre un usuario nuevo o uno ya existente.   |
| <b>Postcondiciones</b>      |         | Se añaden valoraciones y/o usuarios a un conjunto de datos.  |

continúa en la página siguiente

continúa desde la página anterior

| CU-14              | Añadir valoraciones  |
|--------------------|--|
| <b>Excepciones</b> | No existen las matrices de datos.<br>No existe el usuario al que se le quiere añadir valoraciones. |

Tabla B.14: CU-14 Añadir valoraciones

| CU-15                       | Gestión de los conjuntos de datos  |
|-----------------------------|--|
| <b>Versión</b>              | 1.0  |
| <b>Autor</b>                | Raúl Negro Carpintero  |
| <b>Requisitos asociados</b> | RF-5 & RF-5.1 & RF-5.2 & RF-5.3  |
| <b>Descripción</b>          | Gestión de los conjuntos de datos.   |
| <b>Precondiciones</b>       | Tienen que existir los conjuntos de datos de prueba.<br>Tiene que existir un conjunto de datos para usar.                  |
| <b>Acciones</b>             | Listar los conjuntos de datos de prueba.<br>Seleccionar un conjunto de datos de prueba.<br>Añadir conjunto de datos nuevo. |
| <b>Postcondiciones</b>      | Se carga el conjunto de datos de prueba seleccionado.<br>Se añade el conjunto de datos nuevo.                              |
| <b>Excepciones</b>          | No existen conjuntos de datos de prueba.<br>No existe ningún conjunto de datos que añadir.                                 |

Tabla B.15: CU-15 Gestión de los conjuntos de datos

| CU-16                       | Mostrar los conjuntos de datos |
|-----------------------------|--------------------------------|
| <b>Versión</b>              | 1.0                            |
| <b>Autor</b>                | Raúl Negro Carpintero          |
| <b>Requisitos asociados</b> | RF-5 & RF-5.1                  |

continúa en la página siguiente

continúa desde la página anterior

| CU-16                  | Mostrar los conjuntos de datos                       |
|------------------------|--|
| <b>Descripción</b>     | Listar los conjuntos de datos de prueba.             |
| <b>Precondiciones</b>  | Tienen que existir los conjuntos de datos de prueba. |
| <b>Acciones</b>        | Listar los conjuntos de datos de prueba.             |
| <b>Postcondiciones</b> | Se selecciona un conjunto de datos de prueba.        |
| <b>Excepciones</b>     | No existen conjuntos de datos de prueba.             |

Tabla B.16: CU-16 Mostrar los conjuntos de datos

| CU-17                       | Seleccionar el conjunto de datos  |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-5 & RF-5.2   |
| <b>Descripción</b>          | Seleccionar un conjunto de datos de prueba.   |
| <b>Precondiciones</b>       | Tienen que existir los conjuntos de datos de prueba.  |
| <b>Acciones</b>             | Seleccionar un conjunto de datos de prueba.   |
| <b>Postcondiciones</b>      | Se carga el conjunto de datos de prueba seleccionado.   |
| <b>Excepciones</b>          | No existen conjuntos de datos de prueba.<br>No hay ningún conjunto de datos de prueba seleccionado. |

Tabla B.17: CU-17 Seleccionar el conjunto de datos

| CU-18                       | Añadir conjunto de datos        |
|-----------------------------|---------------------------------|
| <b>Versión</b>              | 1.0                             |
| <b>Autor</b>                | Raúl Negro Carpintero           |
| <b>Requisitos asociados</b> | RF-5 & RF-5.3                   |
| <b>Descripción</b>          | Añadir nuevo conjunto de datos. |
| <b>Precondiciones</b>       | -                               |

continúa en la página siguiente



continúa desde la página anterior

| CU-18                  | Añadir conjunto de datos                       |
|------------------------|--|
| <b>Acciones</b>        | Añadir conjunto de datos nuevo.                |
| <b>Postcondiciones</b> | Se añade el nuevo conjunto de datos.           |
| <b>Excepciones</b>     | No existe ningún conjunto de datos que añadir. |

Tabla B.18: CU-18 Añadir conjunto de datos

| CU-19                       | Ayuda de la aplicación  |
|-----------------------------|---|
| <b>Versión</b>              | 1.0   |
| <b>Autor</b>                | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> | RF-6 & RF-6.1 & RF-6.2  |
| <b>Descripción</b>          | Ayuda de la aplicación.   |
| <b>Precondiciones</b>       | Tiene que haber información sobre los modelos.<br>Tiene que haber información sobre los conjuntos de datos de prueba. |
| <b>Acciones</b>             | Mostrar información sobre los modelos.<br>Mostrar información sobre los conjuntos de datos de prueba.                 |
| <b>Postcondiciones</b>      | Se muestra la información sobre los modelos.<br>Se muestra la información sobre los conjuntos de datos de prueba.     |
| <b>Excepciones</b>          | No existe información sobre los modelos.<br>No existe información sobre los conjuntos de datos de prueba.             |

Tabla B.19: CU-19 Ayuda de la aplicación

| CU-20          | Mostrar información sobre los modelos |
|----------------|---------------------------------------|
| <b>Versión</b> | 1.0                                   |

continúa en la página siguiente

continúa desde la página anterior

| <b>CU-20</b>                |  | <b>Mostrar información sobre los modelos</b>   |
|-----------------------------|--|--|
| <b>Autor</b>                |  | Raúl Negro Carpintero                          |
| <b>Requisitos asociados</b> |  | RF-6 & RF-6.1                                  |
| <b>Descripción</b>          |  | Mostrar información sobre los modelos.         |
| <b>Precondiciones</b>       |  | Tiene que haber información sobre los modelos. |
| <b>Acciones</b>             |  | Mostrar información sobre los modelos.         |
| <b>Postcondiciones</b>      |  | Se muestra la información sobre los modelos.   |
| <b>Excepciones</b>          |  | No existe información sobre los modelos.       |

Tabla B.20: CU-20 Mostrar información sobre los modelos

| <b>CU-21</b>                |  | <b>Mostrar información sobre los conjuntos de datos</b>             |
|-----------------------------|--|---|
| <b>Versión</b>              |  | 1.0   |
| <b>Autor</b>                |  | Raúl Negro Carpintero   |
| <b>Requisitos asociados</b> |  | RF-6 & RF-6.2   |
| <b>Descripción</b>          |  | Mostrar información sobre los conjuntos de datos de prueba.         |
| <b>Precondiciones</b>       |  | Tiene que haber información sobre los conjuntos de datos de prueba. |
| <b>Acciones</b>             |  | Mostrar información sobre los conjuntos de datos de prueba.         |
| <b>Postcondiciones</b>      |  | Se muestra la información sobre los conjuntos de datos de prueba.   |
| <b>Excepciones</b>          |  | No existe información sobre los conjuntos de datos de prueba.       |

Tabla B.21: CU-21 Mostrar información sobre los conjuntos de datos

## Apéndice C

---

# Especificación de diseño

---

### C.1. Introducción

En este apéndice se explica cómo están conformados los datos que utilizan las librerías usadas en el proyecto, así como la forma en la que está estructurado el mismo.

### C.2. Diseño de datos

Todos los datos que se han utilizado a lo largo del proyecto están en formato .csv. Lo más normal es que para cada conjunto de datos se tengan los siguientes archivos:

- *ratings.csv*
- *users.csv*
- *items.csv*

La estructura de estos archivos suele ser: *idUser*, *idItem*, *rating*, *timestamp* para *ratings.csv*, *idUser*, *name*, *feature1*, ..., *featureN* para *users.csv* y *idItem*, *name*, *feature1*, ..., *featureN* para *items.csv*.

| Id Usuario | Id Ítem | Valoración | Timestamp  |
|------------|---------|------------|------------|
| 1          | 1       | 5          | 1562057352 |
| 1          | 2       | 1          | 1562057354 |
| 2          | 120     | 4          | 1562057370 |
| .          | .       | .          | .          |
| .          | .       | .          | .          |
| .          | .       | .          | .          |
| Usuario N  | Ítem N  | Valoración | timestamp  |

Tabla C.1: Formato ratings.csv

| Id Usuario | Nombre           | Feature 1 | ... | Feature N |
|------------|------------------|-----------|-----|-----------|
| 1          | Pepe             | Feature 1 | ... | Feature N |
| 2          | Ana              | Feature 1 | ... | Feature N |
| 3          | Luis             | Feature 1 | ... | Feature N |
| .          | .                | .         | .   | .         |
| .          | .                | .         | .   | .         |
| .          | .                | .         | .   | .         |
| Usuario N  | Nombre Usuario N | Feature 1 | ... | Feature N |

Tabla C.2: Formato users.csv

| Id Ítem | Nombre                | Feature 1 | ... | Feature N |
|---------|-----------------------|-----------|-----|-----------|
| 1       | Star Wars Episodio IV | Feature 1 | ... | Feature N |
| 2       | Spotlight             | Feature 1 | ... | Feature N |
| 3       | Seven                 | Feature 1 | ... | Feature N |
| .       | .                     | .         | .   | .         |
| .       | .                     | .         | .   | .         |
| .       | .                     | .         | .   | .         |
| Ítem N  | Nombre Ítem N         | Feature 1 | ... | Feature N |

Tabla C.3: Formato items.csv

Para poder trabajar con los datos primero se pasan a *DataFrames* de *pandas* [2].

## Datos con LightFM

Una vez obtenidos los *DataFrames* para cada *.csv*, es necesario convertirlos a la clase *Dataset* de *LightFM* [8] para poder trabajar con ellos.

Esta clase se encarga de convertir los datos almacenados en los *DataFrames* en *matrices COO* y *matrices CSR*.

## Datos con Spotlight

A diferencia de *LightFM*, en *Spotlight* no se trabaja con matrices dispersas, si no con arrays de *NumPy* (aunque se ofrece la posibilidad de transformar los arrays en *matrices COO* y *matrices CSR*).

*Spotlight* utiliza su propia clase, *Interactions* [10], para convertir los *DataFrames* y así poder utilizar los datos.

Otro aspecto a tener en cuenta en *Spotlight* son las secuencias, utilizadas en el modelo de secuencia implícito [12]. Las recomendaciones pueden verse como secuencias; dados los ítems con los que ha interactuado un usuario, ¿cuáles serán los próximos ítems con los que interactuará? De esta manera, una vez obtenidas las interacciones "normales", hay que pasarlas a interacciones de secuencia con el método `to_sequence()` [11].

## Persistencia

Para la realización del proyecto se necesita guardar de alguna manera tanto los datos intermedios (matrices de interacción) como los propios modelos y los resultados obtenidos al evaluarlos.

Esto se consigue gracias a *pickle* [6].

## C.3. Diseño procedimental

En este apartado se utilizan diagramas de secuencia para explicar el funcionamiento de la aplicación.

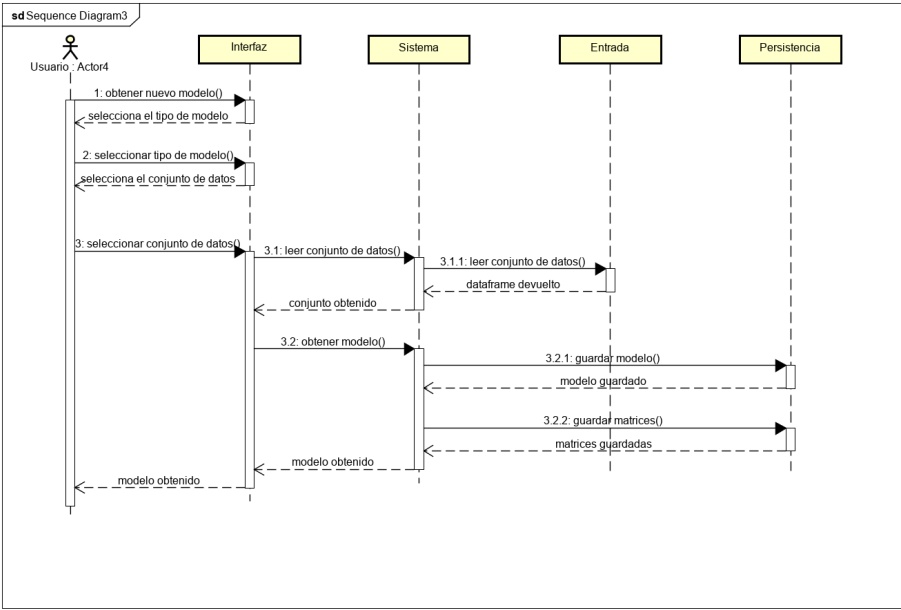


Figura C.1: Diagrama de secuencia para obtener un nuevo modelo

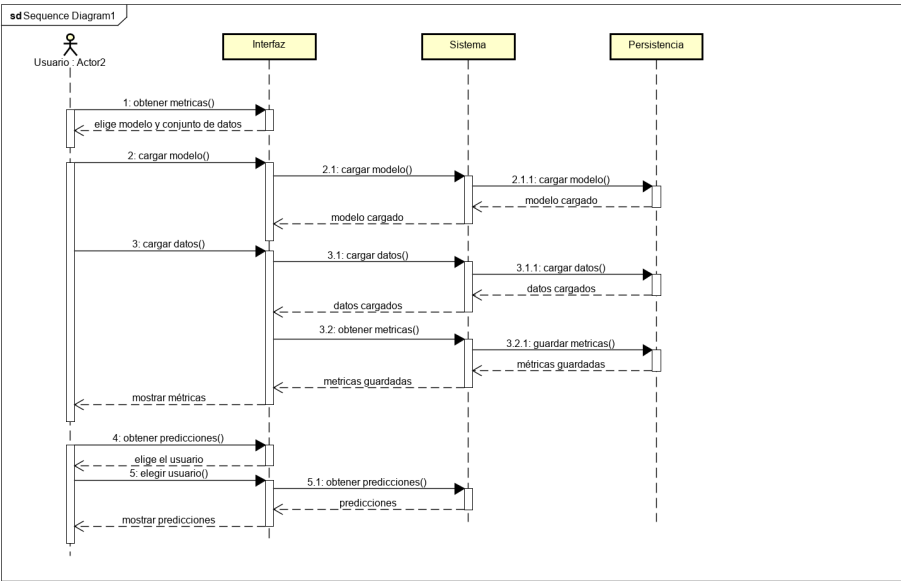


Figura C.2: Diagrama de secuencia para obtener métricas y predicciones

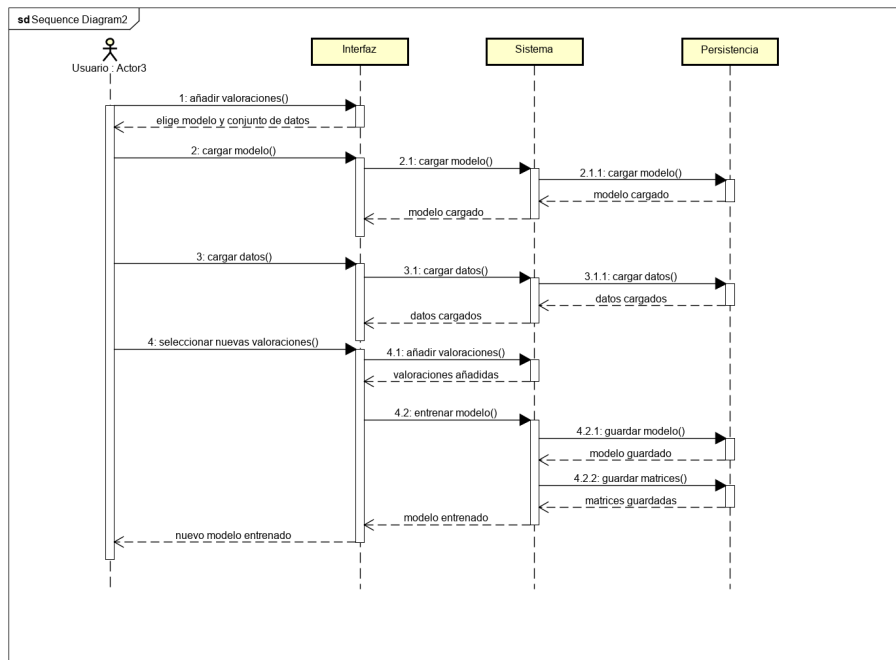


Figura C.3: Diagrama de secuencia para añadir valoraciones

## C.4. Diseño arquitectónico

Para la realización de este proyecto se ha seguido el patrón arquitectónico MVC (*Modelo Vista Controlador*). El objetivo de este patrón es dividir el código en función de su propósito. Sus partes son:

- *Modelo*: el acceso a los datos. Se corresponde con las clases de Entrada, Salida y Persistencia; que leen los datos para dárselo al sistema de recomendación y guardan los resultados.
- *Vista*: la visualización de los datos. Se corresponde con las clases de Interfaz, Flask y Forms, que muestran la información solicitada.
- *Controlador*: la manipulación de los datos. Se corresponde con los clases de Sistema, que crean los sistemas de recomendación.

La estructura del proyecto siguiendo este patrón quedaría de la siguiente forma:

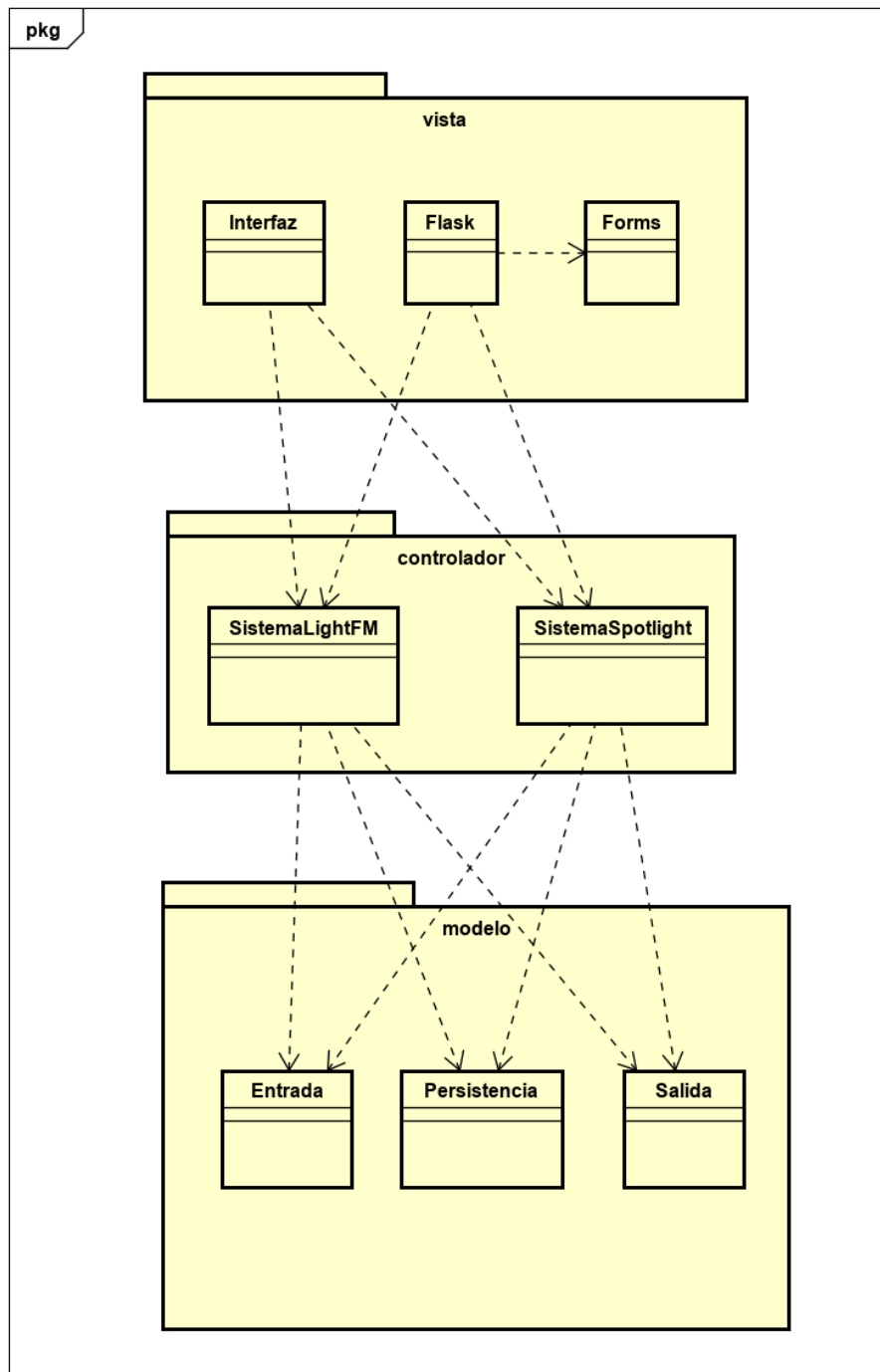


Figura C.4: Diagrama UML del proyecto



Por separado, los paquetes contienen:

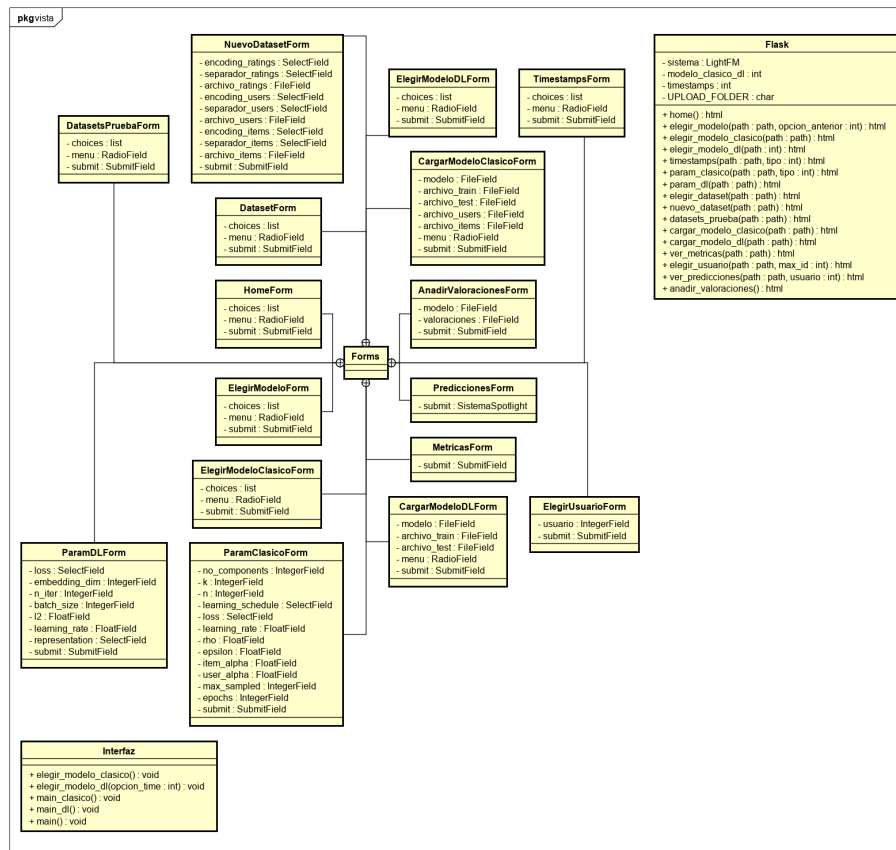
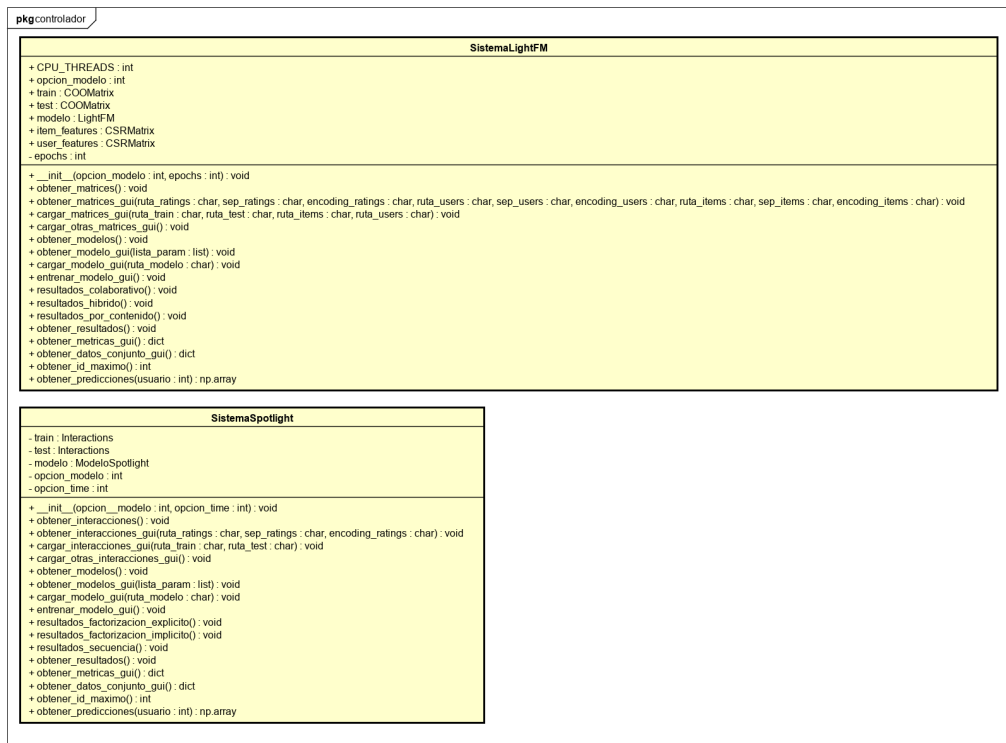


Figura C.5: Diagrama UML del paquete *vista*

Figura C.6: Diagrama UML del paquete *controlador*

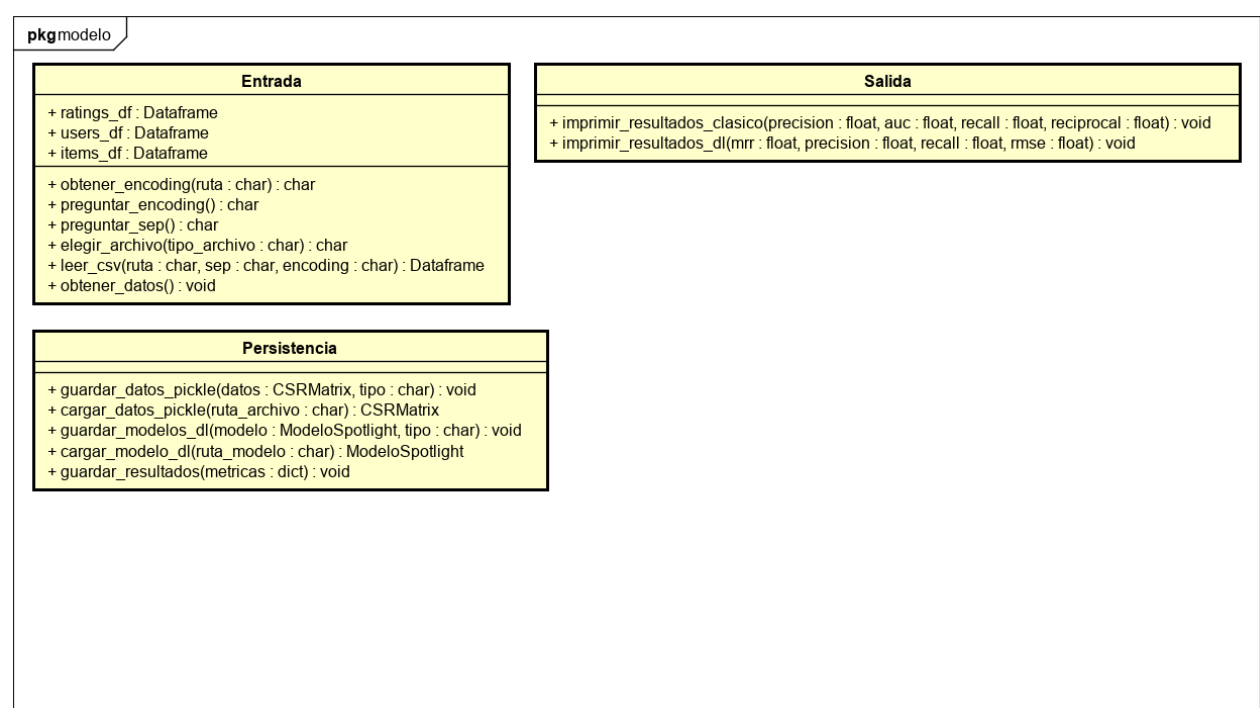


Figura C.7: Diagrama UML del paquete *modelo*

## **C.5. Diseño de la interfaz**

A continuación se muestra el diseño que tendrá la interfaz web (es posible que la versión final cambie).

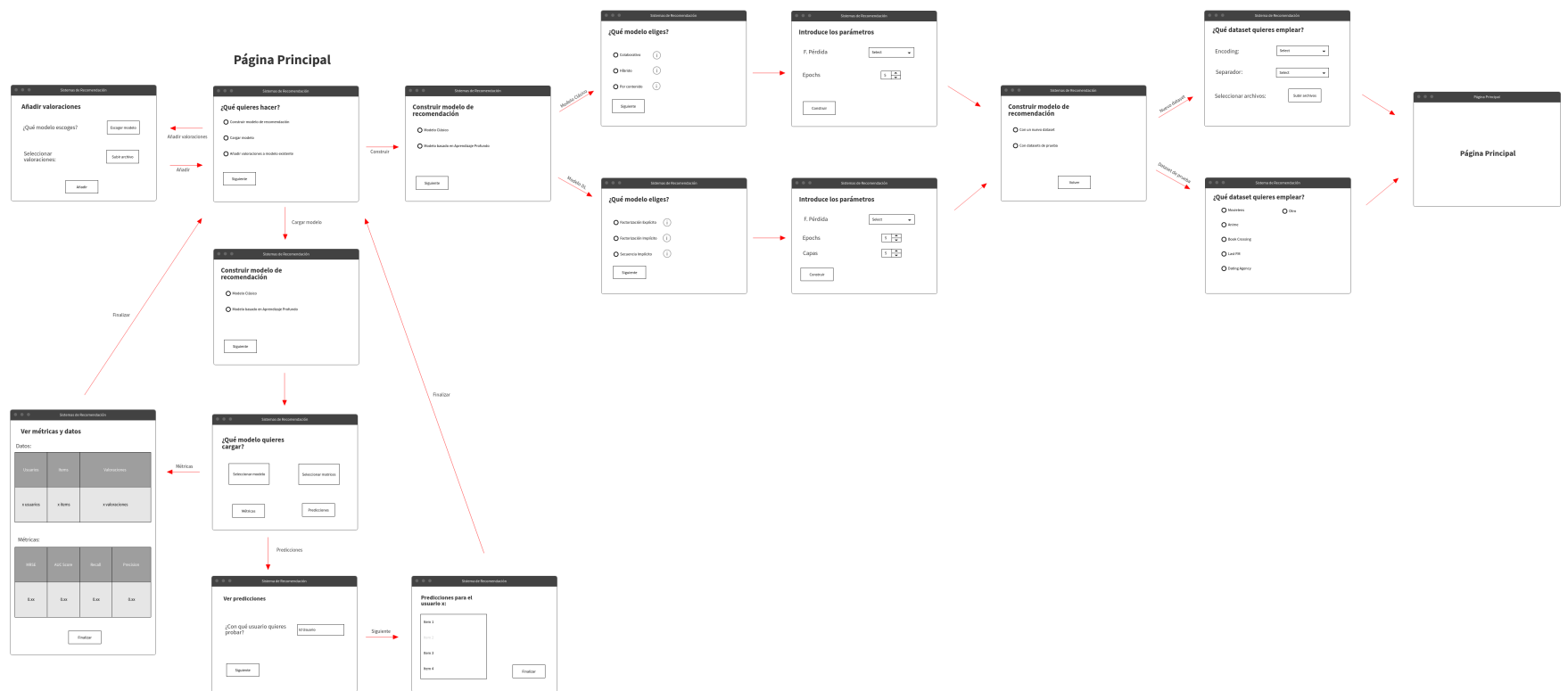


Figura C.8: Interfaz web



## Apéndice *D*

---

# Documentación técnica de programación

---

### D.1. Introducción

En este apéndice se explica cómo está estructurado el proyecto en cuanto a los directorios, así como los pasos para compilar, instalar y ejecutar el programa.

### D.2. Estructura de directorios

El proyecto está dividido en los siguientes directorios:

- **/:** carpeta raíz. Contiene la carpeta *docs*, la carpeta *src* y los ficheros *.gitignore* y *README.md*.
- **/docs:** contiene todos los archivos relacionados con la documentación tanto en  $\text{\LaTeX}$  como en *pdf*. También contiene los archivos de bibliografía y las imágenes que se utilizan a lo largo de la documentación.
- **/docs/img:** carpeta que contiene las imágenes necesarias para apoyar la documentación.
- **/docs/tex:** carpeta que contiene los distintos apartados de la memoria y los anexos en  $\text{\LaTeX}$ .
- **/src:** carpeta que contiene todo el código de la aplicación.
- **/src/controlador:** contiene los ficheros *.py* con los sistemas de *LightFM* y *Spotlight*.
- **/src/modelo:** contiene los ficheros *.py* correspondientes a la *Entrada*, *Salida* y *Persistencia* de los datos.

## APÉNDICE D. DOCUMENTACIÓN TÉCNICA DE PROGRAMACIÓN

- **/src/vista:** contiene los ficheros *.py* correspondientes a la *Interfaz* (de texto), *Flask* (web) y *Forms* (formularios de los que se compone la interfaz web). Además, contiene los *.html* de la interfaz.
- **/src/vista/templates:** contiene las páginas *.html* que componen la interfaz web de la aplicación.
- **/src/uploads:** carpeta donde se guardan los modelos y las matrices obtenidas.
- **/src/notebooks:** esta carpeta contiene archivos que en su día se utilizaron de prueba. Se podrían borrar, pero se prefiere dejarlos para que haya constancia de que se trabajó en ellos.

### D.3. Manual del programador

Las herramientas necesarias para la realización del proyecto han sido:

- Python 3.6
- Spyder o Sublime Text
- LightFM
- Spotlight
- Flask
- WTForms

A continuación se detalla como instalar estas herramientas:

#### Python

Lenguaje de programación escogido para llevar a cabo el proyecto. Se puede descargar la última versión desde el siguiente enlace: [Descargar Python](#).

#### IDE

En este caso, se ha utilizado tanto *Spyder* como *Sublime Text*, además de los notebooks de *Jupyter*.

*Spyder* y *Jupyter* vienen dentro del paquete de *Anaconda* (que también incluye Python). [Guía para instalar Anaconda en Windows](#). En este otro enlace se descarga *Anaconda* directamente para cualquier sistema operativo: [Descargar Anaconda](#).

*Sublime Text* se puede obtener desde el siguiente enlace: [Descargar Sublime Text](#).



#### D.4. COMPILACIÓN, INSTALACIÓN Y EJECUCIÓN DEL PROYECTO

### LightFM

Librería utilizada para la obtención de los modelos de recomendación clásicos. Para instalar la librería es recomendable seguir los pasos descritos en el siguiente enlace: [Instalación de LightFM](#).

### Spotlight

Librería utilizada para la obtención de los modelos de recomendación basados en aprendizaje profundo. Para instalar *Spotlight* basta con ejecutar el siguiente comando: **conda install -c maciejkula -c pytorch spotlight=0.1.5** [9]. Como se puede observar, con este comando también instalamos *PyTorch*.

### Flask

Con *Flask* se obtiene la interfaz web. Tan solo hay que ejecutar el comando **pip install Flask** para instalarlo [4].

### WTForms

Esta librería se utiliza para poder generar todos los formularios que componen la interfaz web. Para instalarla vale con ejecutar **pip install WTForms** [3].

## D.4. Compilación, instalación y ejecución del proyecto

### Obtención del código

El código se encuentra disponible en el siguiente repositorio de GitHub: [Proyecto](#). Una vez dentro del repositorio, basta con descargar el proyecto desde el botón de descarga o clonación.

## APÉNDICE D. DOCUMENTACIÓN TÉCNICA DE PROGRAMACIÓN

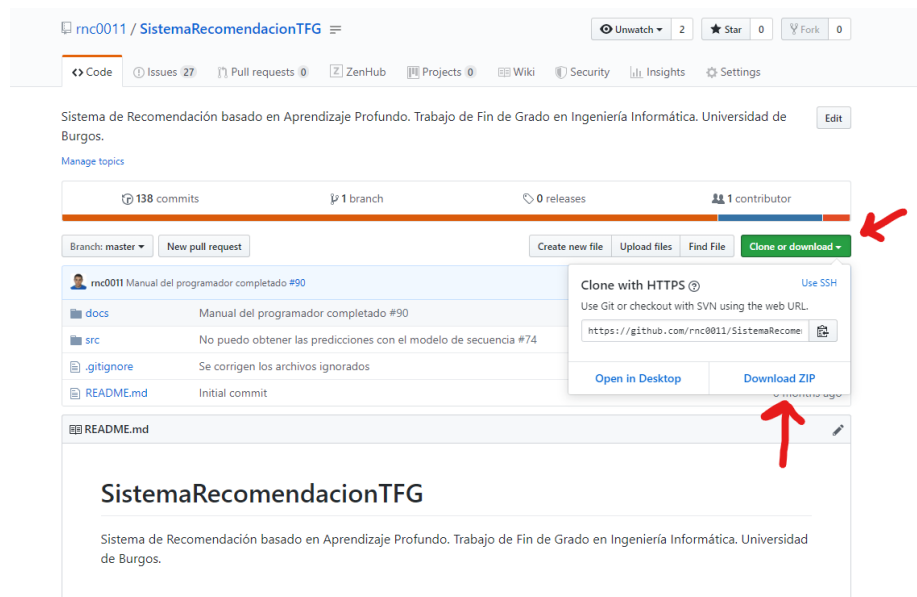


Figura D.1: Descargar proyecto

### Instalación

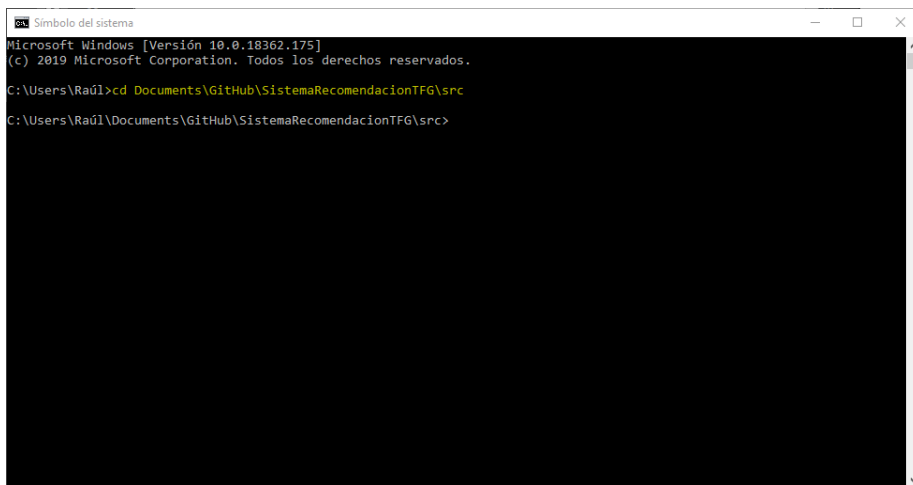
Para poder compilar y ejecutar el proyecto es necesario tener instaladas todas las herramientas listadas anteriormente.

No hace falta instalar una por una las librerías y herramientas anteriores. Basta con navegar hasta el directorio raíz del proyecto y ejecutar el comando **pip install -r requirements.txt**. Esto hará que se instalen de manera automática para el usuario todas las herramientas listadas en el archivo *requirements.txt*.

### Compilación y ejecución

Una vez descargado el repositorio e instaladas todas las herramientas necesarias, se puede proceder a ejecutar la aplicación. Para ello, navegamos desde la consola hasta la carpeta *src* del proyecto:

#### D.4. COMPILACIÓN, INSTALACIÓN Y EJECUCIÓN DEL PROYECTO

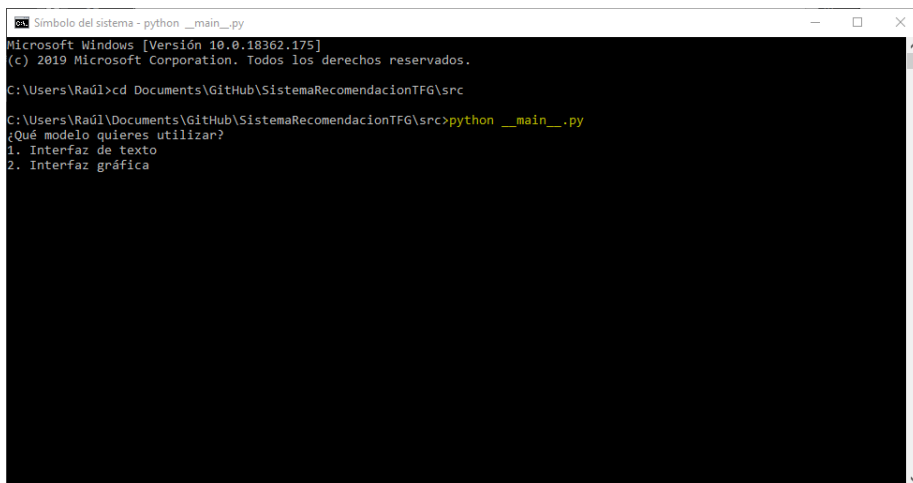


```
Simbolo del sistema
Microsoft Windows [Versión 10.0.18362.175]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Raúl>cd Documents\GitHub\SistemaRecomendacionTFG\src
C:\Users\Raúl\Documents\GitHub\SistemaRecomendacionTFG\src>
```

Figura D.2: Navegación hasta *src*

Para ejecutar el proyecto basta con ejecutar el archivo `__main__.py` que se encuentra dentro de *src*:



```
Simbolo del sistema - python __main__.py
Microsoft Windows [Versión 10.0.18362.175]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Raúl>cd Documents\GitHub\SistemaRecomendacionTFG\src
C:\Users\Raúl\Documents\GitHub\SistemaRecomendacionTFG\src>python __main__.py
¿Qué modelo quieres utilizar?
1. Interfaz de texto
2. Interfaz gráfica
```

Figura D.3: Ejecución del proyecto

Metiendo la opción *2* se ejecuta la interfaz web obtenida gracias a *Flask*. Se tiene que abrir el navegador e ir a la dirección: **`http://localhost:5000/home`**.

## ~~AN~~ÉNDICE D. DOCUMENTACIÓN TÉCNICA DE PROGRAMACIÓN

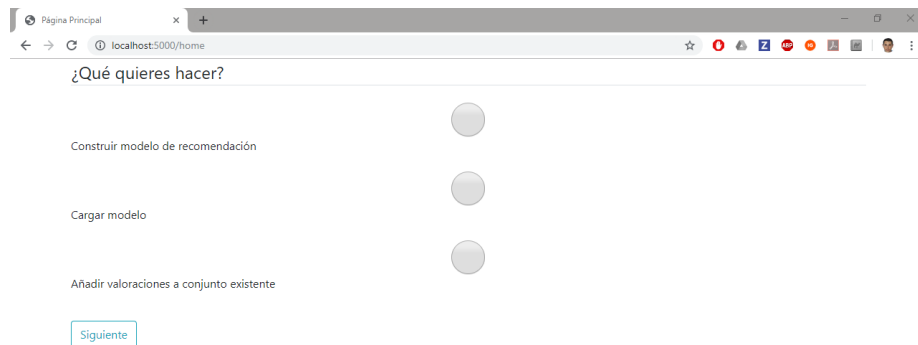


Figura D.4: Interfaz web de la aplicación

### D.5. Pruebas del sistema

Debido a la falta de tiempo durante la realización del proyecto, no ha sido posible realizar tests sobre el código. Como ya se ha comentado en la memoria, este aspecto se deja como línea de trabajo futura.

## Apéndice *E*

---

# Documentación de usuario

---

### E.1. Introducción

En este apéndice se explica cómo el usuario debe instalar y utilizar la aplicación. También se indican los requisitos que necesita el usuario para llevarlo a cabo.

### E.2. Requisitos de usuarios

- Navegador web.
- Conjuntos de datos, en caso de que se quieran utilizar otros además de los de prueba.

### E.3. Instalación

Como no se ha podido desplegar el proyecto con *Heroku* y *Gunicorn*, la aplicación no puede salir a internet. Si un usuario quiere comprobar el funcionamiento de los sistemas, tendrá que descargar e instalar *Flask* para poder acceder a la aplicación desde *localhost*.

Para instalar *Flask* basta con ejecutar **pip install -r requirements.txt**. Esto instalará todas las dependencias que se necesiten para ejecutar el proyecto, no solo *Flask*.

Antes es necesario descargar el proyecto desde el **repositorio** en el que se encuentra. Una vez descargado habría que navegar hasta la carpeta raíz para poder ejecutar el comando anterior.

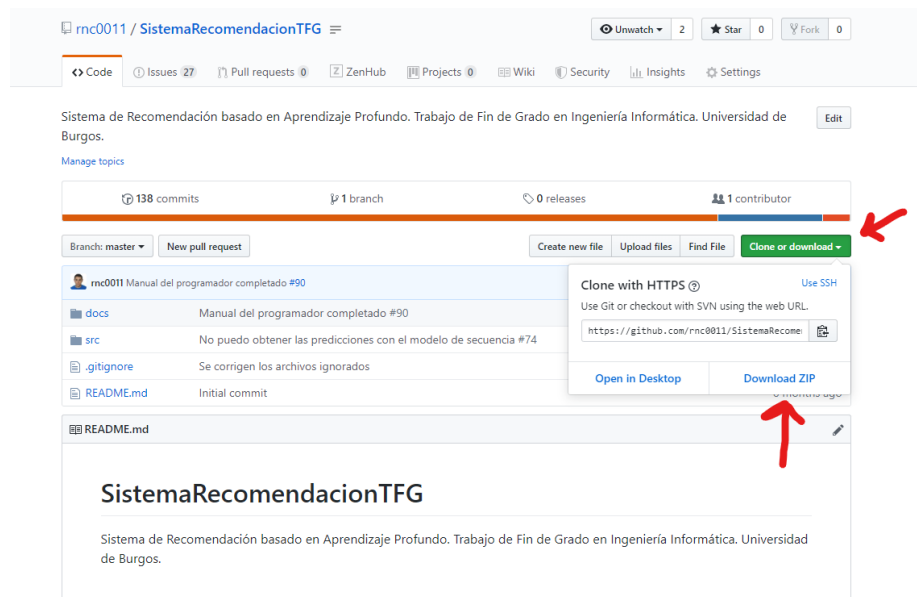
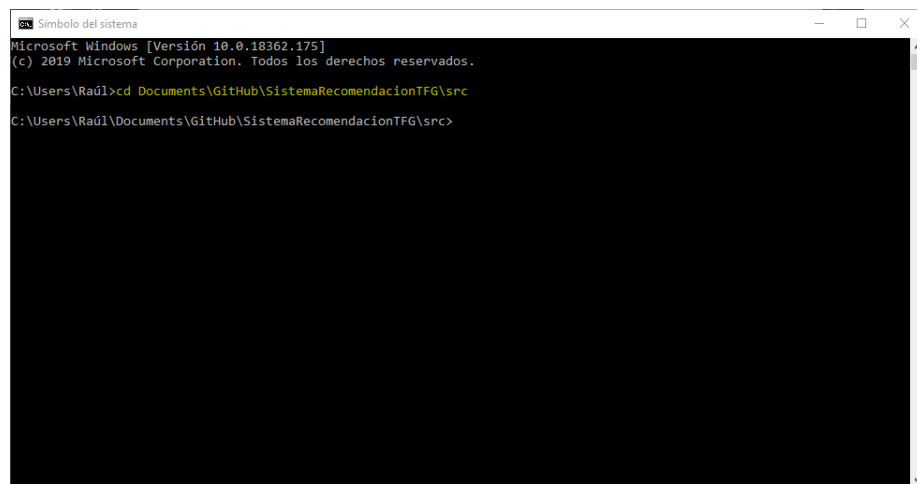
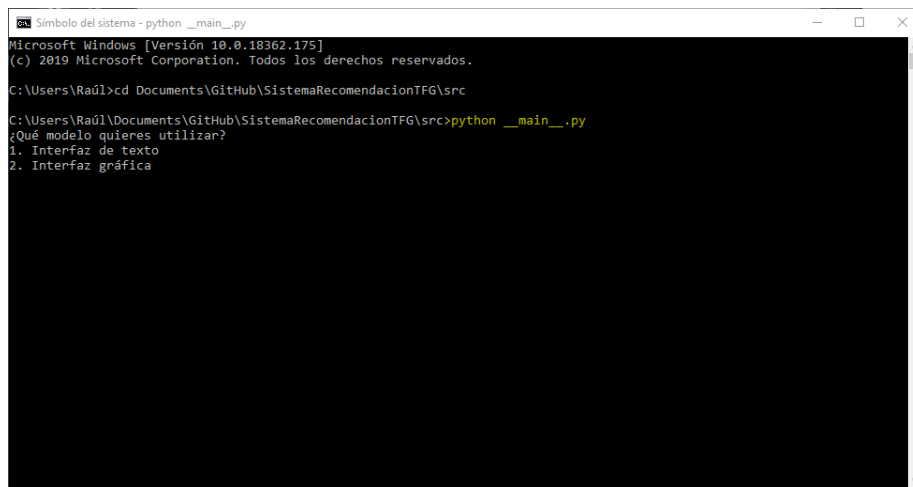


Figura E.1: Descargar proyecto

Figura E.2: Navegación hasta *src*



```
Simbolo del sistema - python __main__.py
Microsoft Windows [Versión 10.0.18362.175]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Raúl>cd Documents\GitHub\SistemaRecomendacionTFG\src

C:\Users\Raúl\Documents\GitHub\SistemaRecomendacionTFG\src>python __main__.py
¿Qué modelo quieres utilizar?
1. Interfaz de texto
2. Interfaz gráfica
```

Figura E.3: Ejecución del proyecto

Metiendo la opción *2* se ejecuta la interfaz web obtenida gracias a *Flask*. Se tiene que abrir el navegador e ir a la dirección: **<http://localhost:5000/home>**.

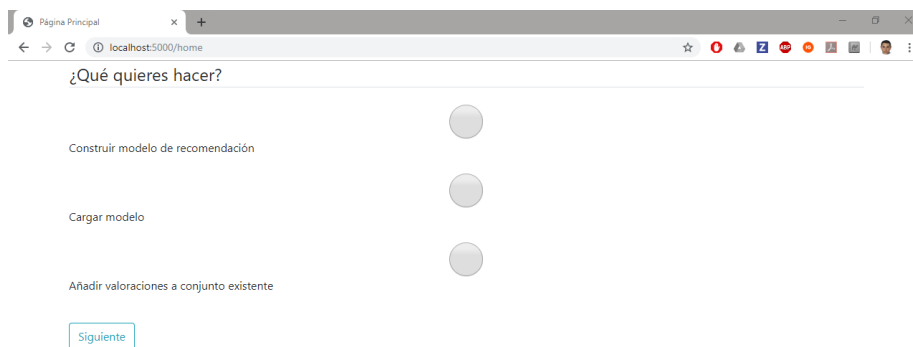


Figura E.4: Interfaz web de la aplicación

## E.4. Manual del usuario

Como se ha comentado en el apartado anterior, hay que navegar hasta el directorio *src* y ejecutar **python \_\_main\_\_.py**. La interfaz de texto se creó en su momento para comprobar el funcionamiento de los sistemas cuando aún no había ninguna interfaz gráfica. En cuanto se generó la interfaz web, la de texto se dejó de lado y ahora no cumple todas las funcionalidades que debería (y que sí cumple la interfaz de *Flask*). Es por eso que se recomienda encarecidamente el uso de la interfaz gráfica.

Una vez seleccionada la *GUI*, se debe ir a la aplicación web con la dirección **localhost:5000/home**.

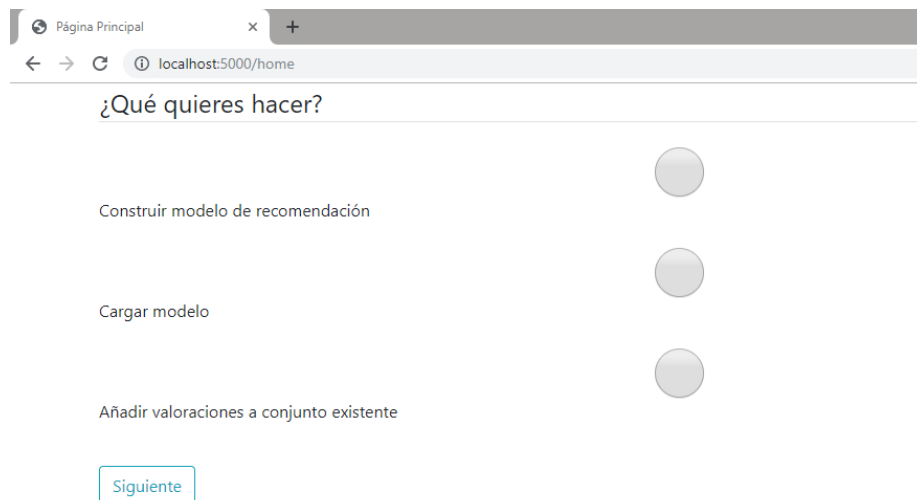


Figura E.5: Página principal de la GUI

En esta primera página se puede construir un modelo de recomendación de 0, ya sea con un conjunto de datos nuevo como con uno ya guardado de antemano; y cargar modelos para evaluar sus métricas y obtener predicciones.

Aunque la opción de añadir valoraciones a conjunto existente parezca estar disponible, no se ha logrado obtener esa funcionalidad. Así pues, el usuario solo tiene dos opciones desde la página principal.



## Construir modelo

Si la opción escogida por el usuario es la primera, se va a encontrar con lo siguiente:

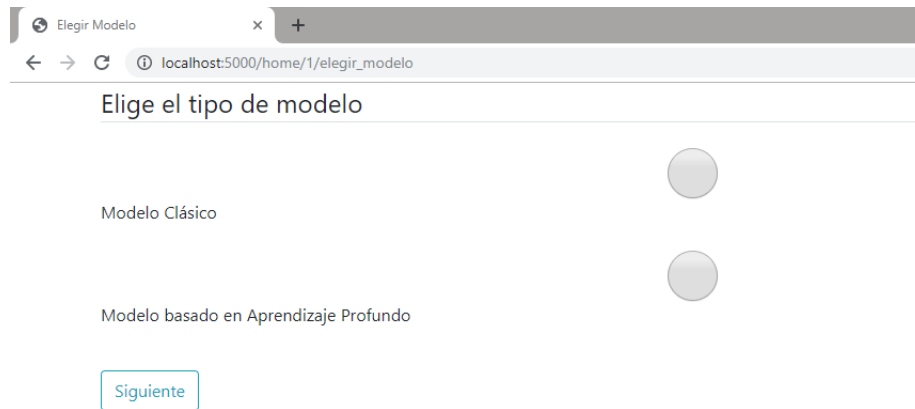


Figura E.6: Elegir entre modelo clásico o basado en Aprendizaje Profundo

Ahora tendrá que decidir qué tipo de modelo quiere obtener. Si escoge el modelo clásico, se le preguntará por el tipo de modelo clásico:

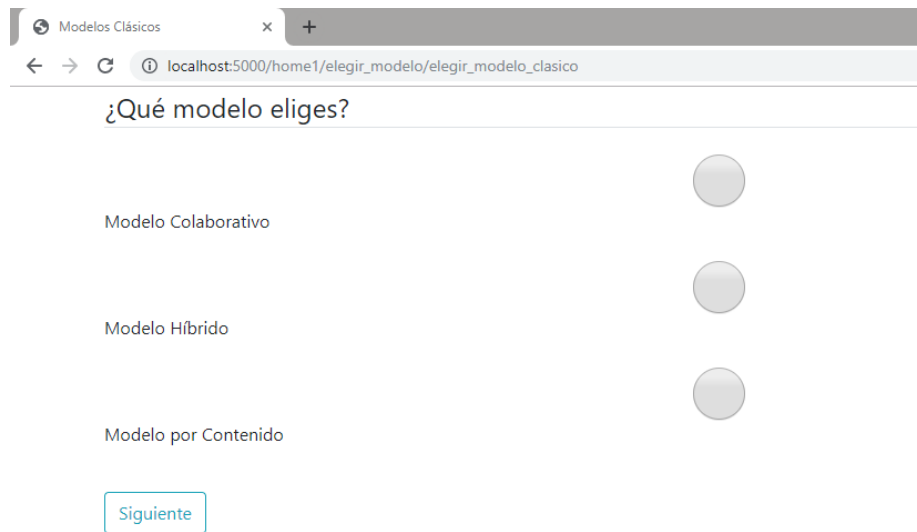


Figura E.7: Elegir 1 de los 3 modelos clásicos

Escoja el que escoja, el usuario tendrá a continuación que elegir los parámetros con los que se va a obtener el modelo. Si no selecciona ninguno, se usarán los que salen por defecto:

The screenshot shows a web browser window with the title 'Parámetros Modelo Clásico'. The address bar shows 'localhost:5000/home1/elegir\_modelo/elegir\_modelo\_clasico/1/param\_clasico'. The page title is 'Introduce los parámetros'. The form contains the following fields:

- No Components: 10
- K: 5
- N: 10
- Learning Schedule: adagrad (dropdown)
- Loss: logistic (dropdown)
- Learning Rate: 0.05
- Rho: 0.95
- Epsilon: 1e-06

Figura E.8: Parte de los parámetros disponibles para los modelos clásicos

En la siguiente página, se pedirá al usuario que escoja entre crear el modelo con un dataset nuevo o con uno ya guardado de antes:

The screenshot shows a web browser window with the title 'Elegir Dataset'. The address bar shows 'localhost:5000/home1/elegir\_modelo/elegir\_modelo\_clasico1/param\_clasico/elegir\_dataset'. The page title is 'Elegir Dataset'. The form contains two radio buttons and a button:

- Con un nuevo dataset (radio button)
- Con datasets de prueba (radio button)
- Siguiente (button)

Figura E.9: Página donde elegir si usar un dataset nuevo o no

Si el usuario decide utilizar un dataset nuevo, tendrá que seleccionar cada archivo (el de valoraciones, el de ítems y el de usuarios), además de indicar qué encoding y separador utilizan:

Nuevo Dataset

¿Que dataset quieres analizar?

Encoding Ratings  
utf-8

Separador Ratings  
|

Archivo .csv de valoraciones  
Seleccionar archivo Ningún archivo seleccionado

Encoding Users  
utf-8

Separador Users  
|

Archivo .csv de usuarios  
Seleccionar archivo Ningún archivo seleccionado

Encoding Items  
utf-8

Separador Items  
|

Archivo .csv de items  
Seleccionar archivo Ningún archivo seleccionado

Siguiente

Figura E.10: Página donde elegir si usar un dataset nuevo o no

Una vez seleccionados, la aplicación se quedará pensando un rato mientras leer los archivos de datos, obtiene las matrices, obtiene el modelo y lo entrena. Según va haciendo todo esto irá preguntando al usuario cómo quiere guardar los datos intermedios. Cuando el proceso acabe, se volverá a la página principal automáticamente.

Si en el paso de usar un dataset nuevo o no, se escoge la segunda opción, el usuario se encontraría con lo siguiente:

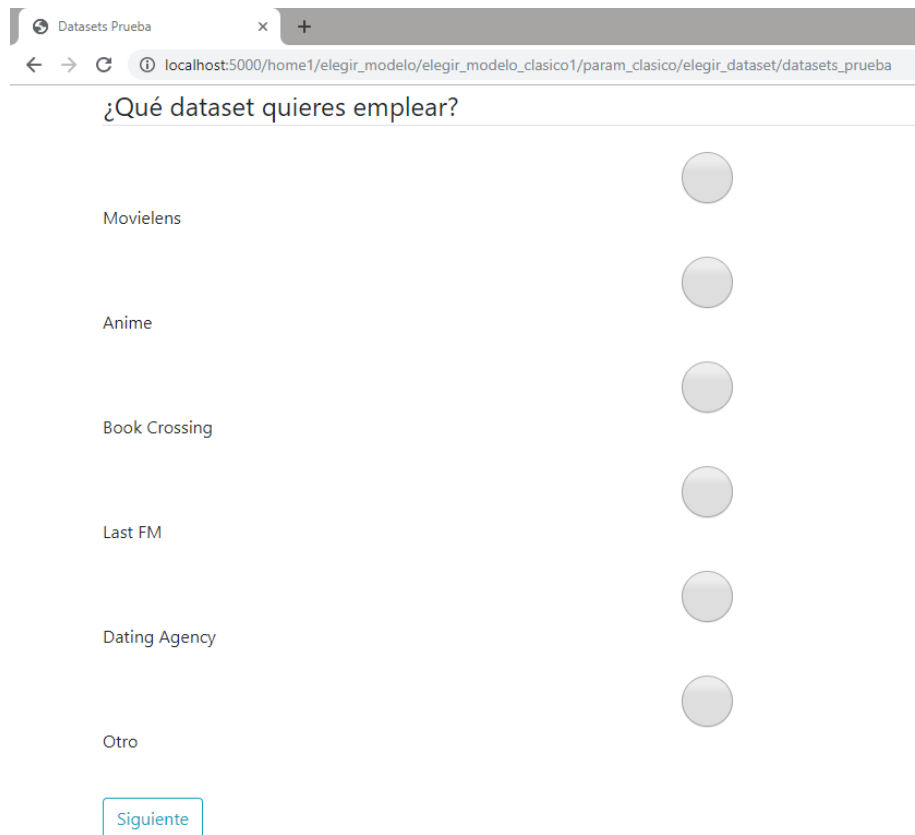
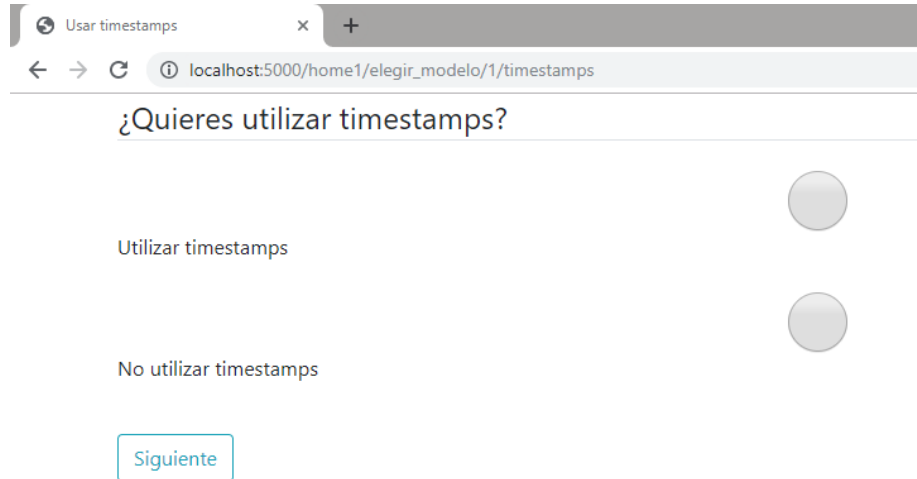


Figura E.11: Página donde seleccionar los datasets ya guardados

En caso de escoger uno entre *Movielens* y *Dating Agency*, la aplicación busca directamente las matrices, las carga y solo pregunta al usuario dónde guardar y con qué nombre el modelo resultante. Si, en cambio, se escoge otro dataset, la aplicación pide al usuario la ruta de las matrices que se debieron guardar cuando se metió ese dataset por primera vez.

Todo esto es completamente igual si al principio se escoge un modelo basado en aprendizaje profundo y no uno clásico, con la excepción, que una vez escogido el modelo de deep learning se pregunta al usuario si quiere utilizar los timestamps de los dataset o no. Si el modelo que se escoge es el

de secuencia, los timestamps tiene que estar sí o sí.



The screenshot shows a web browser window with a single tab titled "Usar timestamps". The address bar displays "localhost:5000/home1/elegir\_modelo/1/timestamps". The main content area features a heading "¿Quieres utilizar timestamps?". Below this heading are two radio button options: "Utilizar timestamps" and "No utilizar timestamps". At the bottom of the form is a blue button labeled "Siguiente".

Figura E.12: Se pregunta si se quieren utilizar los timestamps o no

## Cargar modelo

Si se escoge esta opción, la aplicación volverá a preguntar el tipo de modelo, si clásico o no. Una vez escogido, se tendrá que ir seleccionando las matrices que utiliza el modelo:

Cargar Modelo Clásico

localhost:5000/home2/elegir\_modelo/cargar\_modelo\_clasico

### ¿Qué modelo clásico quieres cargar?

Archivo .pickle con el modelo

Seleccionar archivo Ningún archivo seleccionado

Archivo .pickle de entrenamiento

Seleccionar archivo Ningún archivo seleccionado

Archivo .pickle de test

Seleccionar archivo Ningún archivo seleccionado

Archivo .pickle de usuarios

Seleccionar archivo Ningún archivo seleccionado

Archivo .pickle de items

Seleccionar archivo Ningún archivo seleccionado

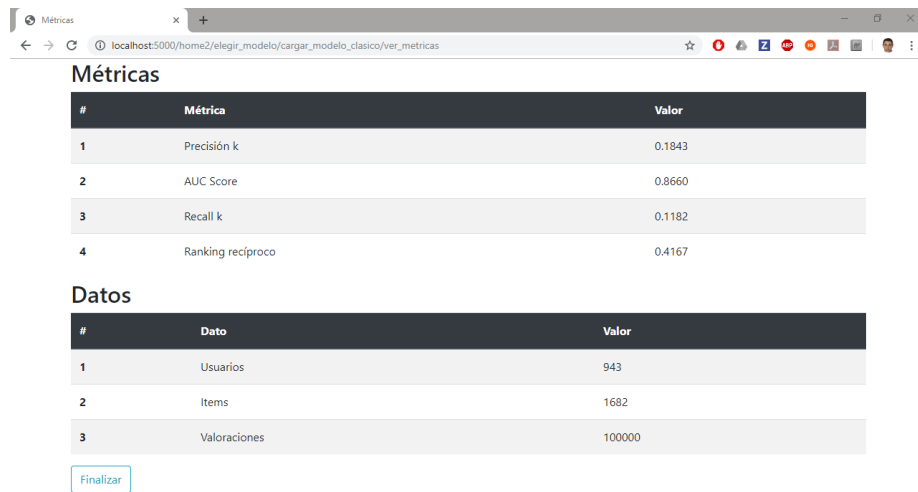
Métricas

Predicciones

Siguiente

Figura E.13: Se pide seleccionar las matrices

Una vez las matrices se han seleccionado, el usuario puede ver las métricas y los datos del modelo:



Métricas

| # | Métrica           | Valor  |
|---|-------------------|--------|
| 1 | Precisión k       | 0.1843 |
| 2 | AUC Score         | 0.8660 |
| 3 | Recall k          | 0.1182 |
| 4 | Ranking recíproco | 0.4167 |

Datos

| # | Dato         | Valor  |
|---|--------------|--------|
| 1 | Usuarios     | 943    |
| 2 | Items        | 1682   |
| 3 | Valoraciones | 100000 |

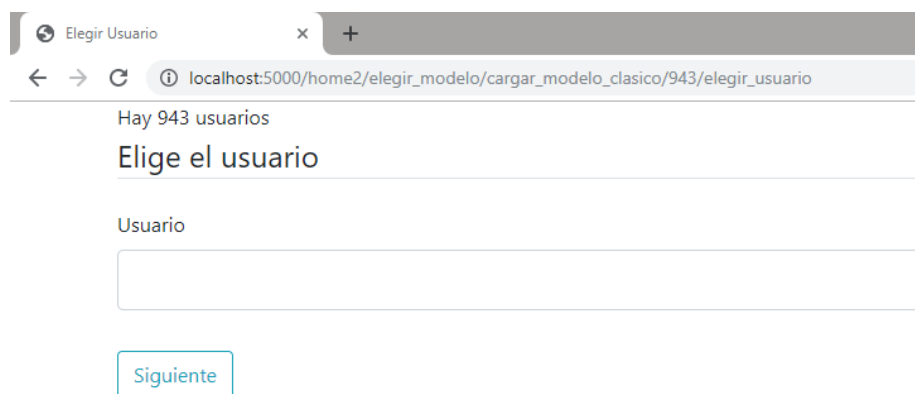
Finalizar

Figura E.14: Métricas del modelo y datos de los datasets

Al ver las métricas también se pide al usuario un nombre y una localización para guardar el archivo *.csv* que las guardará.

Si se escoge la opción de ver predicciones, saldrá la siguiente página:





The screenshot shows a web browser window with a single tab titled 'Elegir Usuario'. The address bar displays the URL 'localhost:5000/home2/elegir\_modelo/cargar\_modelo\_clasico/943/elegir\_usuario'. The page content includes the text 'Hay 943 usuarios' followed by the heading 'Elige el usuario'. Below the heading is a text input field with the label 'Usuario'. At the bottom of the visible form area is a button labeled 'Siguiente'.

Figura E.15: Elegir usuario cuyas predicciones se quieren ver

Una vez introducido el usuario, se verán las 20 mejores predicciones:

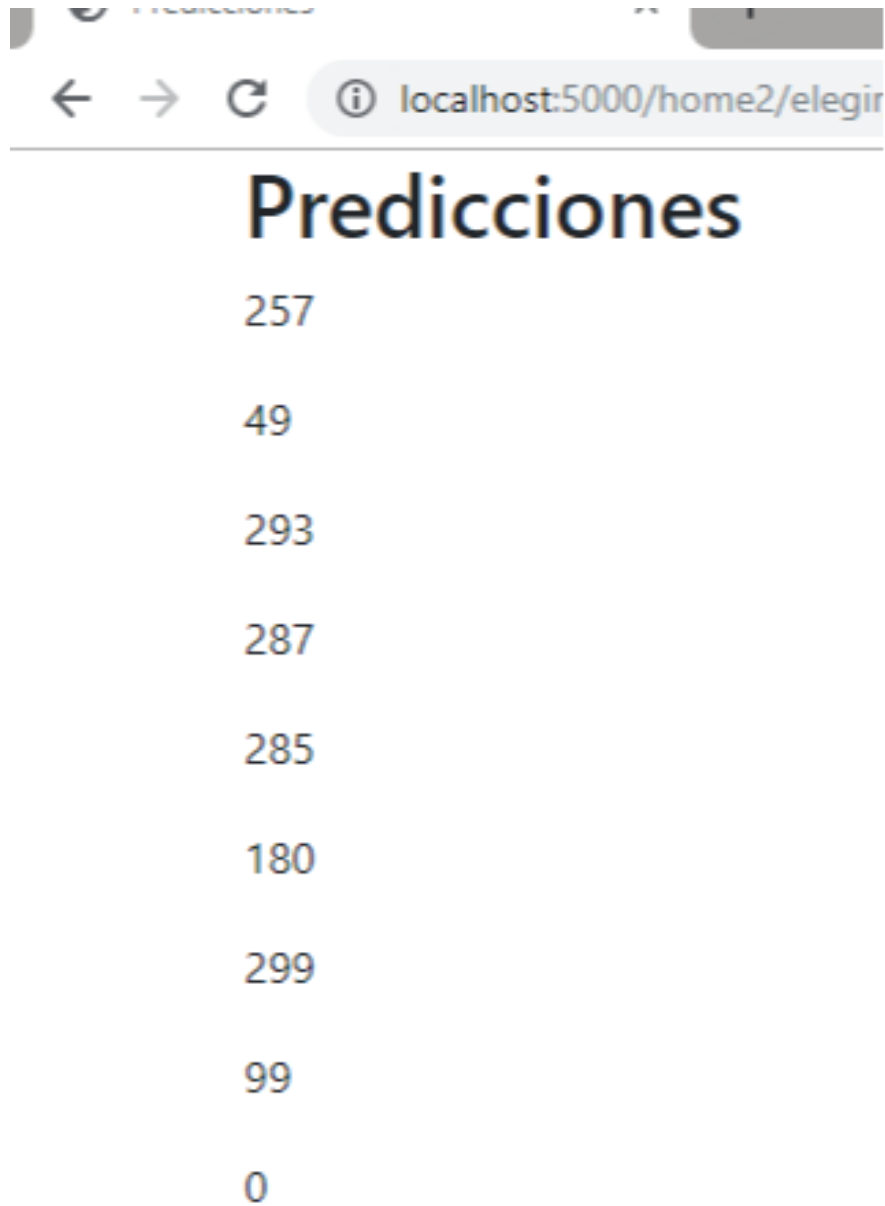


Figura E.16: Predicciones del usuario

Hay que tener en cuenta que las predicciones para el modelo de secuencia de *Spotlight* no se pueden obtener.

También hay que tener en cuenta que *LightFM* y *Spotlight* van de los índices del 0 al máximo - 1.

Es recomendable dar nombres muy significativos a los modelos, matrices,

etc.,



---

## Bibliografía

---

- [1] Calculador irpf. <https://cincodias.elpais.com/herramientas/calculadora-irpf/>.
- [2] Dataframe. <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>.
- [3] Download / installation. [https://wtforms.readthedocs.io/en/stable/crash\\_course.html#download-installation](https://wtforms.readthedocs.io/en/stable/crash_course.html#download-installation).
- [4] Install flask. <http://flask.pocoo.org/docs/1.0/installation/#install-flask>.
- [5] La realidad del perfil de informático junior en españa según los informes. <https://www.xataka.com/tecnologiazen/la-realidad-del-perfil-de-informatico-junior-en-espana-segun-los-informes>.
- [6] pickle - python object serialization. <https://docs.python.org/3/library/pickle.html>.
- [7] Practical deep learning for coders, v3. <https://course.fast.ai/>.
- [8] Maciej Kula. Dataset construction. <http://lyst.github.io/lightfm/docs/lightfm.data.html#lightfm.data.Dataset>.
- [9] Maciej Kula. Installation. <https://maciejkula.github.io/spotlight/#installation>.
- [10] Maciej Kula. Interactions. <https://maciejkula.github.io/spotlight/interactions.html>.

- [11] Maciej Kula. Interactions - to\_sequence(). [https://maciejkula.github.io/spotlight/interactions.html#spotlight.interactions.Interactions.to\\_sequence](https://maciejkula.github.io/spotlight/interactions.html#spotlight.interactions.Interactions.to_sequence).
- [12] Maciej Kula. Sequential models. <https://maciejkula.github.io/spotlight/index.html#sequential-models>.
- [13] Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman. *Mining of Massive Datasets*, chapter 9, page 513. Universidad de Standford, 2010.