

# SISTEMA RECOMENDADOR

$$\begin{aligned}
 & |b(T, a, b)| \leq 2 \\
 & \varphi(\sigma_1 t) \varphi(\sigma_2 t) = \varphi(\sqrt{\sigma_1^2 + \sigma_2^2} t) \\
 & \rho(\alpha) = \frac{\sum_{k=1}^{\infty} p_k^{\alpha} \log_2 \frac{1}{p_k}}{\sum_{k=1}^{\infty} p_k^{\alpha}} \quad (i_k \sigma_k^2 = \lambda_i \cdot c_i k) \quad \eta_1 = \sum_{k=1}^n a_k \frac{x_k}{\eta_k} \quad \log \varphi(u) = -\frac{\sigma^2 u^2}{2} \quad i^2 = -1; j^2 = -1; k^2 = -1 \\
 & \sum_{k=1}^n p_k^{\alpha} \log_2 \frac{1}{p_k} = \frac{1}{12\pi} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt \quad S(\alpha, \tau) = \frac{2}{\pi} \int_0^{\pi} \frac{\sin \alpha t}{t} dt \quad P(\eta_{\infty} < x) = F(x) \quad \lim_{n \rightarrow \infty} \frac{(2n)}{(n)} = e^{-2x} \\
 & S_n = A_n U \pi A_n \quad W_k = \binom{n}{k} p^k (1-p)^{n-k} \quad P(\eta < y | \xi = x) = \sup_{y' < y, y' \geq 0} P(\eta < y' | \xi = x) \\
 & |A_n| = \frac{n!}{2} \left| \int_{|x| > A} f(x) \log_2 \frac{1}{f(x)} dx \right| < \varepsilon \quad g^{-1} \cdot g = e \quad \gamma = \sqrt{\frac{2u}{\pi}} \left( \frac{\eta_{2n}}{\sqrt{2n}} + \frac{\eta_{2n} - \eta_{2n}}{\sqrt{2n}} \right) \quad f(t|y) = \frac{2e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \left( \frac{e^{-\frac{t^2}{2}}}{(1 - \frac{t^2}{u^2})^{\frac{3}{2}}} \right) \quad \Delta N = \sum_{n=1}^N \frac{1}{n} \\
 & \int dG_k(x) \geq \frac{1}{2} \sum_{k=0}^{+\infty} e^{-\frac{k^2 \pi^2}{2}} = H(k) \quad \prod_{k \leq b} \bigcup_{i=1}^{n-1} M_i; \bigcap_{n=0}^{\infty} X_n \quad f_n(t) = \frac{2^{-n} e^{-t} e^{-2t}}{(n-1)!} \quad H_r(x) = \frac{Gr(x)}{1+Gr(x)} \quad U_k^+ = (2n) - (2n-c) \\
 & f_{n-1}(t) = \int_0^1 f_n(u) f_1(t-u) du = \frac{2^{n-1} e^{-t} e^{-2t}}{n!} \quad \lim_{t \rightarrow 0} f_n(t) = 0 \quad C_{iv} = \sum_{j=1}^n a_{ij} b_{ij} \quad \lim_{n \rightarrow \infty} \frac{f_n(t)}{n} = p_k \quad R = \int_{-\infty}^{+\infty} \varphi(t) dt \quad U_k^+ = (2n) - (2n-c) \\
 & \log \varphi(t) = i \gamma t - c |t|^{\alpha} \left[ 1 + i \beta \frac{t}{|t|} \omega(t, \alpha) \right] \quad B(u) = \sum_{k=1}^r \varphi^*(b_k u) \quad \lim_{n \rightarrow \infty} P \left( \frac{\sum_{i=1}^n a_i - a_{in} - \log \frac{1}{q}}{\sqrt{\frac{1-q}{q}}} \right) \quad C_n(\alpha) \geq \frac{n!}{\prod_{k=1}^n n_k(\alpha)!} \quad \frac{1}{m} \varphi(t) = \varphi \left( c \left( \frac{n}{m} \right) t \right) \\
 & \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = F(x) \left( \frac{1}{\sqrt{2\pi}} \right)^{-1} \quad |\Psi_S(t)| = \left| \int_{-\infty}^{+\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{+\infty} e^{-v|x|} dF(x) = \varphi_S(iv) \quad g^{-1} N_g = \{g^{-1} n_g | n \in N\} \quad Q = F^{-1}(C\varphi) \quad q_n(\alpha) = \frac{p_k^{\alpha}}{\sum_{j=1}^n p_j^{\alpha}} \quad P(C\pi_2) = \\
 & \Gamma_m = \Gamma_r | \Gamma_{m-r} \quad |X \cup Y| = |X| + |Y| - |X \cap Y| \quad \lim_{n \rightarrow \infty} \frac{1}{n} k_n \left( \frac{x}{\sqrt{n}} \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad p_n(b_k) = \frac{C_n}{p_k^{\alpha}} \quad P \left( \lim_{n \rightarrow \infty} \sup \frac{|h_n|}{\sqrt{2n \log \log n}} \leq 1 \right) = 1 \quad (A) = 1 - \sqrt{1 - e^{2t}} \\
 & f: X \rightarrow X \cap W \quad Q(A) = \int_A \chi(\omega) dP \quad f'(x) = -\log_2 \left( \frac{\sum_{k=1}^r p_k^{\alpha} \log_2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^{\alpha}} - \left( \frac{\sum_{k=1}^r p_k^{\alpha} \log_2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^{\alpha}} \right)^2 \right) \quad f g(u_i) = f \left( \sum_{j=1}^{dim V_k} a_{ji} v_j \right) = \sum_{j=1}^{dim V_k} a_{ji} \left( \sum_{k=1}^{dim V_k} b_{kj} w_k \right) \left( \frac{2\ell}{2\ell} \right) \approx \frac{1}{\sqrt{2\ell}} \\
 & q \left( c^{-x} \sqrt{\frac{1-q}{nq}} - 1 \right) = x \sqrt{\frac{q(1-q)}{n}} + o \left( \frac{1}{\sqrt{n}} \right) \quad \prod_{k=1}^r \left[ g_k \left( \frac{t}{\sqrt{t_k}} \right) \right]^{N_k \alpha_k} = e^{-\frac{t^2}{2}} \quad P_{jk}^{(m)} = \sum_{r=0}^m p_{jr}^{(r)} p_{ek}^{(m-r)} \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} Re \left\{ \varphi(t) \frac{e^{it\alpha} - e^{it\beta}}{it} \right\} dt \quad P(\omega_1 | \gamma \varepsilon) \leq \frac{C_q}{\log N} \\
 & \liminf_{N \rightarrow \infty} \int_{-\infty}^{+\infty} f_N(x)^{\alpha} dx \geq \int_{-\infty}^{+\infty} f(x)^{\alpha} dx \quad M(l\sigma_j - 1|s) = \int_0^{\infty} (1-x-1)^s e^{-x} dx \quad \lim_{N \rightarrow \infty} \int_{-1}^{+1} f_N(x) \log_2 \frac{1}{f_N(x)} dx = \int_{-1}^{+1} f(x) \log_2 \frac{1}{f(x)} dx \quad N_{k_1} - \varepsilon_k = (2n) - \varepsilon_k = (2n) \\
 & D^2(J_n) \leq \frac{k}{n} + 2k \left( \frac{1}{2} \sum_{k=1}^n R_k(b_k) \right) \quad \det(M') = \det(M) + \det(M^*) = \det(M) \quad h(xy) = \frac{1}{2\pi} \left[ \sqrt{2} e^{-\frac{x^2}{2}} - e^{-x^2} \right] \quad |M(\varepsilon_n, \varepsilon_m)| \leq C_2 \sqrt{\frac{n}{m-n}}
 \end{aligned}$$

## 1ª Entrega

## Proyectos de Programación (PROP)

21 de Noviembre de 2021

## Grupo 2.3

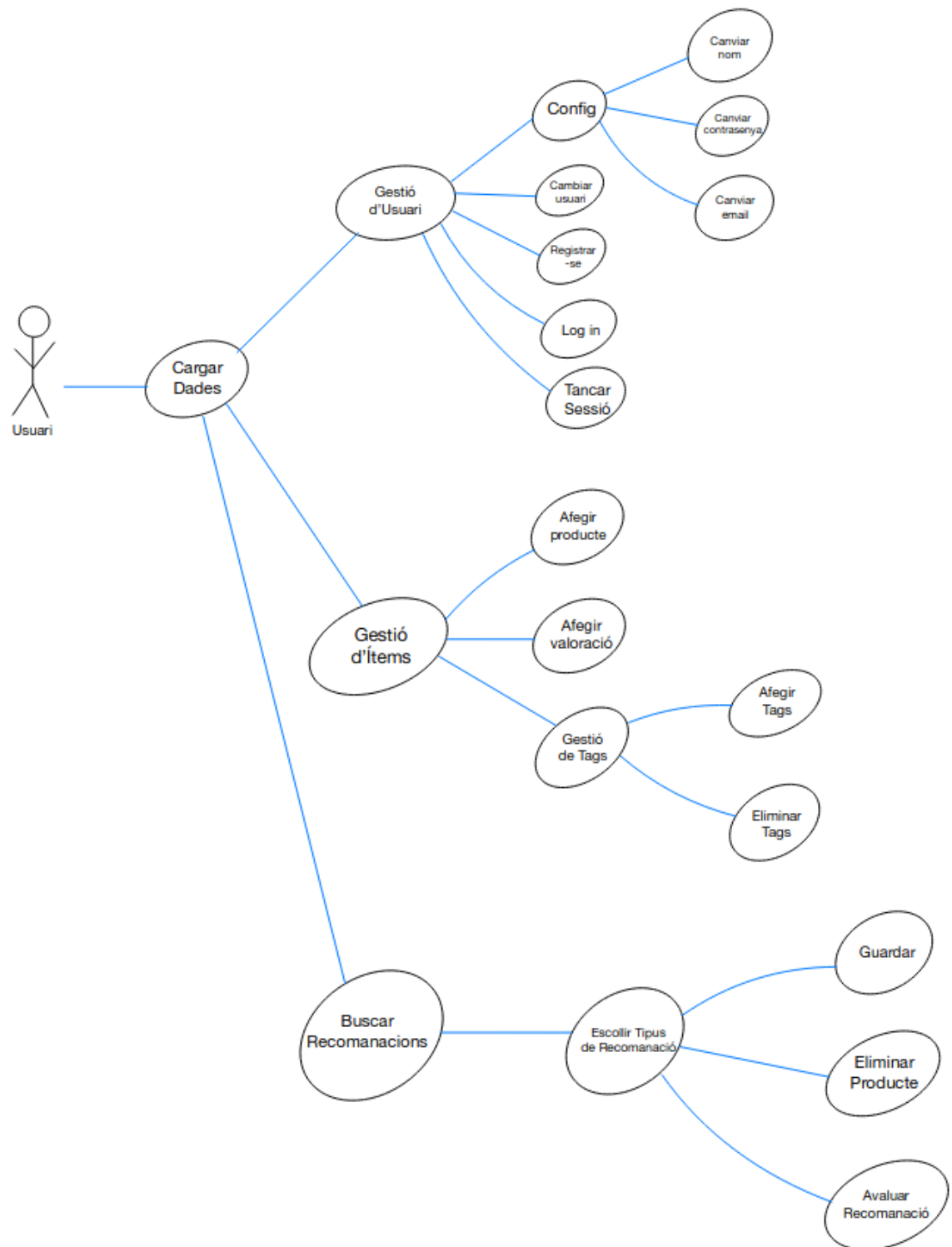
Marc Delgado -	marc.delgado.sanchez
Miguel Gutiérrez -	miguel.gutierrez.jariod
Oriol Martí -	oriol.marti.jariod
Manel Piera -	manel.piera

# ÍNDICE

<b>1. Casos de uso.....</b>	<b>2</b>
1.1 Diagrama	
1.2 Descripción	
1.2.1 Cargar Datos	
1.2.2 Gestión de Usuario	
1.2.11 Gestión de Ítems	
1.2.17 Buscar Recomendaciones	
<b>2. Clases del modelo conceptual de datos.....</b>	<b>16</b>
2.1 Diagrama	
2.2 Descripción de las clases	
2.2.1 CSVParserItem	
2.2.2 CSVParserRate	
2.2.3 Content	
2.2.4 K-NN	
2.2.5 K-Means	
2.2.6 SlopeOne	
2.2.7 CollaborativeFiltering	
2.2.8 Evaluation	
<b>3. Descripción de los algoritmos.....</b>	<b>24</b>
3.1 K-Means	
3.2 Slope One	
3.3 K-Nearest-Neighbours	
<b>4. Estructuras de datos y costes.....</b>	<b>30</b>
4.1 Decisiones sobre estructuras de datos	
4.2 Costes de las funciones	
<b>5. Relación clases y miembros.....</b>	<b>36</b>
<b>6. Manual de uso de la aplicación y drivers.....</b>	<b>37</b>

# 1. Casos de uso

## 1.1 Diagrama



## 1.2 Descripción

### 1.2.1 Cargar Datos

Actor Principal	Actor
Precondición	Ha de existir un conjunto de datos guardados.
Disparador	El actor quiere cargar los datos que hay guardados en el sistema de recomendaciones.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere cargar los datos previamente guardados.</li><li>2. El sistema muestra por pantalla los datos existentes del sistema de recomendaciones.</li></ol>
Extensiones	<ol style="list-style-type: none"><li>1. El actor cancela la acción de cargar los datos.</li></ol> <p>Acaba el caso de uso.</p>

### 1.2.2 Gestión de Usuario

Actor Principal	Actor
Precondición	Para gestionar un usuario previamente se ha de tener unos datos que identifiquen a este para poder realizar la gestión.
Disparador	El actor quiere gestionar su usuario previamente cargado.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere gestionar su usuario.</li><li>2. El sistema ofrece la opción de gestionar una serie de elementos con respecto al usuario</li></ol>
Extensiones	<ol style="list-style-type: none"><li>1. El actor cancela la acción de gestión del usuario.</li></ol> <p>Acaba el caso de uso.</p>

### 1.2.3 Configuración

Actor Principal	Actor
Precondición	Para realizar la configuración el actor debe estar logueado con una cuenta.
Disparador	El actor quiere configurar alguna opción relacionada con su cuenta.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere configurar su cuenta.</li><li>2. El sistema verifica que el actor está logueado como usuario.</li><li>3. El sistema ofrece las opciones de configuración</li></ol>
Extensiones	<p>2a. El actor no está logueado como usuario</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <ol style="list-style-type: none"><li>1. El actor cancela la acción de acceso a la configuración.</li></ol> <p>Acaba el caso de uso.</p>

### 1.2.4 Cambiar Nombre

Actor Principal	Actor
Precondición	El actor debe de haberse logueado y haber accedido a la configuración para poder realizar el cambio de su nombre de usuario.
Disparador	El actor quiere cambiar su nombre de usuario.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere cambiar el nombre de usuario de su cuenta.</li><li>2. El sistema ofrece el cambio al nuevo nombre.</li><li>3. El actor introduce el nuevo nombre.</li><li>4. El sistema guarda el nombre del usuario y lo almacena en el sistema.</li></ol>
Extensiones	<p>3a. El actor introduce un nombre incompatible.</p> <p>3a1. El sistema informa que el nuevo nombre no es válido.</p> <p>3a2. El actor vuelve a realizar el punto 3.</p> <ol style="list-style-type: none"><li>1. El actor cancela la acción de cambiar el nombre.</li></ol> <p>Acaba el caso de uso.</p>

### 1.2.5 Cambiar Contraseña

Actor Principal	Actor
Precondición	El actor debe de haberse logueado y haber accedido a la configuración para poder realizar el cambio de su contraseña de usuario.
Disparador	El actor quiere cambiar su contraseña de usuario.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere cambiar la contraseña de usuario de su cuenta.</li><li>2. El sistema ofrece el cambio a la nueva contraseña.</li><li>3. El actor introduce la nueva contraseña.</li><li>4. El sistema guarda la contraseña del usuario y lo almacena en el sistema.</li></ol>
Extensiones	<p>3a. El actor introduce una contraseña incompatible.</p> <p>3a1. El sistema informa que la nueva contraseña no es válida.</p> <p>3a2. El actor vuelve a realizar el punto 3.</p> <ol style="list-style-type: none"><li>1. El actor cancela la acción de cambiar la contraseña.</li></ol> <p>Acaba el caso de uso.</p>

### 1.2.6 Cambiar Email

Actor Principal	Actor
Precondición	El actor debe de haberse logueado y haber accedido a la configuración para poder realizar el cambio de su email de usuario.
Disparador	El actor quiere cambiar su email de usuario.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere cambiar el email de usuario de su cuenta.</li><li>2. El sistema ofrece el cambio al nuevo email.</li><li>3. El actor introduce el nuevo email.</li><li>4. El sistema guarda el email del usuario y lo almacena en el sistema.</li></ol>
Extensiones	<p>3a. El actor introduce un email incompatible.</p>

	<p>3a1. El sistema informa que el nuevo email no es válido.</p> <p>3a2. El actor vuelve a realizar el punto 3.</p> <p>1. El actor cancela la acción de cambiar el email.</p> <p>Acaba el caso de uso.</p>
--	---

### 1.2.7 Cambiar Usuario

Actor Principal	Actor
Precondición	El actor debe de haberse logueado previamente con una cuenta de usuario para poder realizar el cambio a otra cuenta. Esta última debe de estar registrada en el sistema.
Disparador	El actor quiere cambiar de cuenta de usuario.
Escenario Principal de éxito	<p>1. El actor indica que quiere cambiar de cuenta.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El sistema ofrece el cambio de cuenta.</p> <p>4. El actor introduce los datos requeridos para iniciar la nueva cuenta.</p> <p>5. El sistema realiza el cambio y se inicia con la nueva cuenta.</p>
Extensiones	<p>4a. El actor introduce una cuenta invalida.</p> <p>4a1. El sistema informa al actor que los datos introducidos no se corresponden.</p> <p>4a2. El actor vuelve al punto 4.</p> <p>1. El actor cancela la acción de cambiar de cuenta.</p> <p>Acaba caso de uso.</p>

### 1.2.8 Registro

Actor Principal	Actor
Precondición	El actor no debe de disponer de una cuenta logueada al realizar el registro. Los parámetros de registro del usuario serán únicos.
Disparador	El actor quiere registrarse con una cuenta de usuario al sistema.
Escenario	1. El actor indica que quiere registrarse al sistema.

Principal de éxito	<p>2. El sistema verifica que no hay cuenta existente logueada.</p> <p>3. El sistema ofrece los parámetros necesarios para el registro.</p> <p>4. El actor introduce los parámetros para realizar el registro.</p> <p>5. El sistema verifica que los parámetros son correctos</p> <p>6. El sistema informa al autor que se ha realizado con éxito el registro.</p>
Extensiones	<p>2a El sistema rechaza el registro porque hay una cuenta logueada e informa del incidente al actor.</p> <p>2a1. El actor debe de salir de la cuenta existente para realizar el registro.</p> <p>4b El actor introduce parámetros incompatibles.</p> <p>4b1. El sistema informa al actor que sus datos no son válidos para el registro.</p> <p>4b2. El actor sigue el proceso desde el punto 4.</p> <p>1. El actor cancela la acción de registro.</p> <p>Acaba el caso de uso.</p>

### 1.2.9 Log In

Actor Principal	Actor
Precondición	Para acceder al log in es necesaria una cuenta de usuario existente registrada en el sistema.
Disparador	El actor quiere loguear con una cuenta de usuario al sistema.
Escenario Principal de éxito	<p>1. El actor indica que quiere loguearse en el sistema con una cuenta de usuario.</p> <p>2. El sistema muestra por pantalla los datos que el actor ha de introducir para cargar sus datos.</p> <p>3. El actor introduce los datos necesarios para cargar sus datos, como detalla el sistema.</p> <p>4. El sistema informa que se ha iniciado correctamente y se han cargado los datos debidamente.</p>
Extensiones	<p>3a. El actor indica un email que no contiene datos guardados.</p> <p>3a.1 El sistema informa que no es un email válido.</p>



	<p>3a.2 El actor sigue el proceso desde el punto 3.</p> <p>3b. El actor indica un email inválido.  3b.1 El sistema informa que no es un email válido.  3b.2 El actor sigue el proceso desde el punto 3.</p> <p>3c. El actor indica una contraseña errónea.  3c.1 El sistema informa que no es la contraseña correcta.  3c.2 El actor sigue el proceso desde el punto 3.</p> <p>3 El actor cancela la acción de Cargar Datos.</p> <p>Acaba el caso de uso.</p>
--	---

### 1.2.10 Cerrar Sesión

Actor Principal	Actor
Precondición	Para cerrar la sesión el actor debe de tener una cuenta logueada en curso.
Disparador	El actor quiere cerrar la sesión de la cuenta de usuario existente.
Escenario Principal de éxito	<p>1. El actor indica que quiere cerrar la sesión en curso.</p> <p>2. El sistema indica que se ha cerrado la sesión correctamente.</p>
Extensiones	<p>1a. El actor cancela el cierre de la sesión existente.</p> <p>Acaba caso de uso.</p>

### 1.2.11 Gestión Items

Actor Principal	Actor
Precondición	Para gestionar unos ítems, el usuario debe tener una cuenta logueada en curso.
Disparador	El actor quiere hacer una gestión relacionada con los ítems del sistema.
Escenario Principal de éxito	<p>1. El actor indica que quiere gestionar los ítems del sistema.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El sistema ofrece al actor la opción de hacer diversas gestiones referentes a los ítems.</p>

Extensiones	<p>2a. El actor no está logueado como usuario</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>1. El actor cancela la acción de gestionar los ítems del sistema.</p> <p>Acaba el caso de uso.</p>
-------------	--

### 1.2.12 Añadir Producto

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere añadir un producto nuevo al sistema.
Escenario Principal de éxito	<p>1. El actor indica que quiere añadir un producto al sistema.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El actor introduce el identificador único del ítem, así como todos sus tags.</p>
Extensiones	<p>2a. El actor no está logueado como usuario</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. Ya existe un ítem con ese identificador.</p> <p>3a1. El sistema informa que los identificadores de ítem deben ser únicos y el dado ya ha sido usado previamente.</p> <p>3a2. El actor vuelve al punto 3 e introduce un identificador distinto.</p> <p>1. El actor cancela la acción de añadir un producto al sistema.</p> <p>Acaba el caso de uso.</p>

### 1.2.13 Añadir Valoración

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso y el ítem que se quiere valorar debe existir en el sistema y haber sido adquirido por el actor.

Disparador	El actor quiere valorar un ítem que ha consumido.
Escenario Principal de éxito	<ol style="list-style-type: none"> <li>1. El actor indica que quiere valorar un ítem.</li> <li>2. El sistema verifica que el actor está logueado como usuario.</li> <li>3. El sistema verifica que el ítem a valorar existe.</li> <li>4. El sistema verifica que el ítem a valorar ha sido consumido previamente por el actor.</li> <li>5. El actor introduce la valoración correspondiente.</li> </ol>
Extensiones	<p>2a. El actor no está logueado como usuario</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. El ítem a valorar no existe.</p> <p>3a1. El actor debe introducir un identificador de ítem presente en el sistema.</p> <p>4. El ítem a valorar no ha sido consumido por el actor.</p> <p>4a1. El actor debe introducir un identificador de un ítem que haya consumido previamente.</p> <p>1. El actor cancela la acción de añadir una valoración.</p> <p>Acaba el caso de uso.</p>

#### 1.2.14 Gestión de Tags

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso y debe haber añadido al sistema el ítem que se quiere modificar.
Disparador	El actor quiere modificar los tags de un ítem del sistema.
Escenario Principal de éxito	<ol style="list-style-type: none"> <li>1. El actor indica que quiere modificar los tags de un ítem.</li> <li>2. El sistema verifica que el actor está logueado como usuario.</li> <li>3. El sistema verifica que el ítem que se quiere modificar ha sido añadido por el actor.</li> <li>4. El sistema muestra las opciones para modificar los tags de un ítem.</li> </ol>
Extensiones	<p>2a. El actor no está logueado como usuario</p> <p>2a1. El sistema informa que es necesario haber logueado con una</p>

	<p>cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. El ítem a modificar no fue añadido por el usuario.</p> <p>3a1. El actor debe introducir un ítem que haya añadido él al sistema.</p> <p>1. El actor cancela la acción de modificar los tags de un ítem</p> <p>Acaba el caso de uso.</p>
--	--

### 1.2.15 Añadir Tag

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso, debe haber añadido al sistema el ítem al que quiere añadir un Tag y el ítem no debe contener el Tag.
Disparador	El actor quiere añadir un tag a un ítem del sistema.
Escenario Principal de éxito	<p>1. El actor indica que quiere añadir un tag a un ítem.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El sistema verifica que el ítem al que se quiere añadir un Tag ha sido añadido por el actor.</p> <p>4. El actor introduce el Tag correspondiente.</p>
Extensiones	<p>2a. El actor no está logueado como usuario.</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. El ítem a añadir un Tag no fue añadido por el usuario.</p> <p>3a1. El actor debe introducir un ítem que haya añadido él al sistema.</p> <p>4a. El ítem ya contiene el Tag que quiere añadir el usuario.</p> <p>4a1. El actor debe introducir un Tag que no tenga el ítem.</p> <p>1. El actor cancela la acción de añadir tag a un ítem.</p> <p>Acaba el caso de uso.</p>

### 1.2.16 Eliminar Tag

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso, debe haber añadido al sistema el ítem al que quiere añadir un Tag y el ítem debe contener el Tag.
Disparador	El actor quiere eliminar un tag de un ítem del sistema.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere eliminar el tag de un ítem.</li><li>2. El sistema verifica que el actor está logueado como usuario.</li><li>3. El sistema verifica que el ítem que se quiere eliminar un Tag ha sido añadido por el actor.</li><li>4. El actor elimina el tag correspondiente.</li></ol>
Extensiones	<p>2a. El actor no está logueado como usuario.</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. El ítem a eliminar un Tag no fue añadido por el usuario.</p> <p>3a1. El actor debe introducir un ítem que haya añadido él al sistema.</p> <p>4a. El ítem ya no contiene el Tag que quiere añadir el usuario.</p> <p>4a1. El actor debe introducir un Tag que tenga el ítem.</p> <p>1. El actor cancela la acción de eliminar tag a un ítem.</p> <p>Acaba el caso de uso.</p>

### 1.2.17 Buscar Recomendaciones

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere conocer ítems recomendados.
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere ítems recomendados.</li><li>2. El sistema verifica que el actor está logueado como usuario.</li><li>3. El sistema muestra las opciones para hacer la recomendación de ítems.</li></ol>

Extensiones	<p>2a. El actor no está logueado como usuario.</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>1. El actor cancela la acción de buscar una recomendación.</p> <p>Acaba el caso de uso.</p>
-------------	--

### 1.2.18 Escoger Tipo de Recomendación

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere escoger qué algoritmo de recomendación va a usar.
Escenario Principal de éxito	<p>1. El actor elige el algoritmo deseado de recomendación. Collaborative Filtering o Content Based Filtering.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El sistema muestra los ítems recomendados al usuario.</p>
Extensiones	<p>2a. El actor no está logueado como usuario.</p> <p>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</p> <p>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>1. El actor cancela la acción de escoger un algoritmo de recomendación.</p> <p>Acaba el caso de uso.</p>

### 1.2.19 Guardar

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere guardar una recomendación.
Escenario Principal de éxito	<p>1. El actor indica que quiere guardar una recomendación.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p>

	3. El sistema guarda la recomendación.
Extensiones	<p>2a. El actor no está logueado como usuario.  2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.  2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>1. El actor cancela la acción de guardar la recomendación.</p> <p>Acaba el caso de uso.</p>

### 1.2.20 Eliminar Producto

Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere eliminar un producto de una recomendación
Escenario Principal de éxito	<p>1. El actor indica que quiere eliminar un producto de una recomendación.</p> <p>2. El sistema verifica que el actor está logueado como usuario.</p> <p>3. El sistema comprueba que el producto a eliminar está en una recomendación guardada.</p> <p>4. El sistema elimina el producto de la recomendación.</p>
Extensiones	<p>2a. El actor no está logueado como usuario.  2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.  2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</p> <p>3a. El producto a eliminar no se encuentra en una recomendación guardada.  3a1. El sistema informa que el producto a eliminar no está en una recomendación guarda.  3a2. El actor debe volver indicar un producto a eliminar.</p> <p>1. El actor cancela la acción de eliminar un producto de la recomendación.</p> <p>Acaba el caso de uso</p>

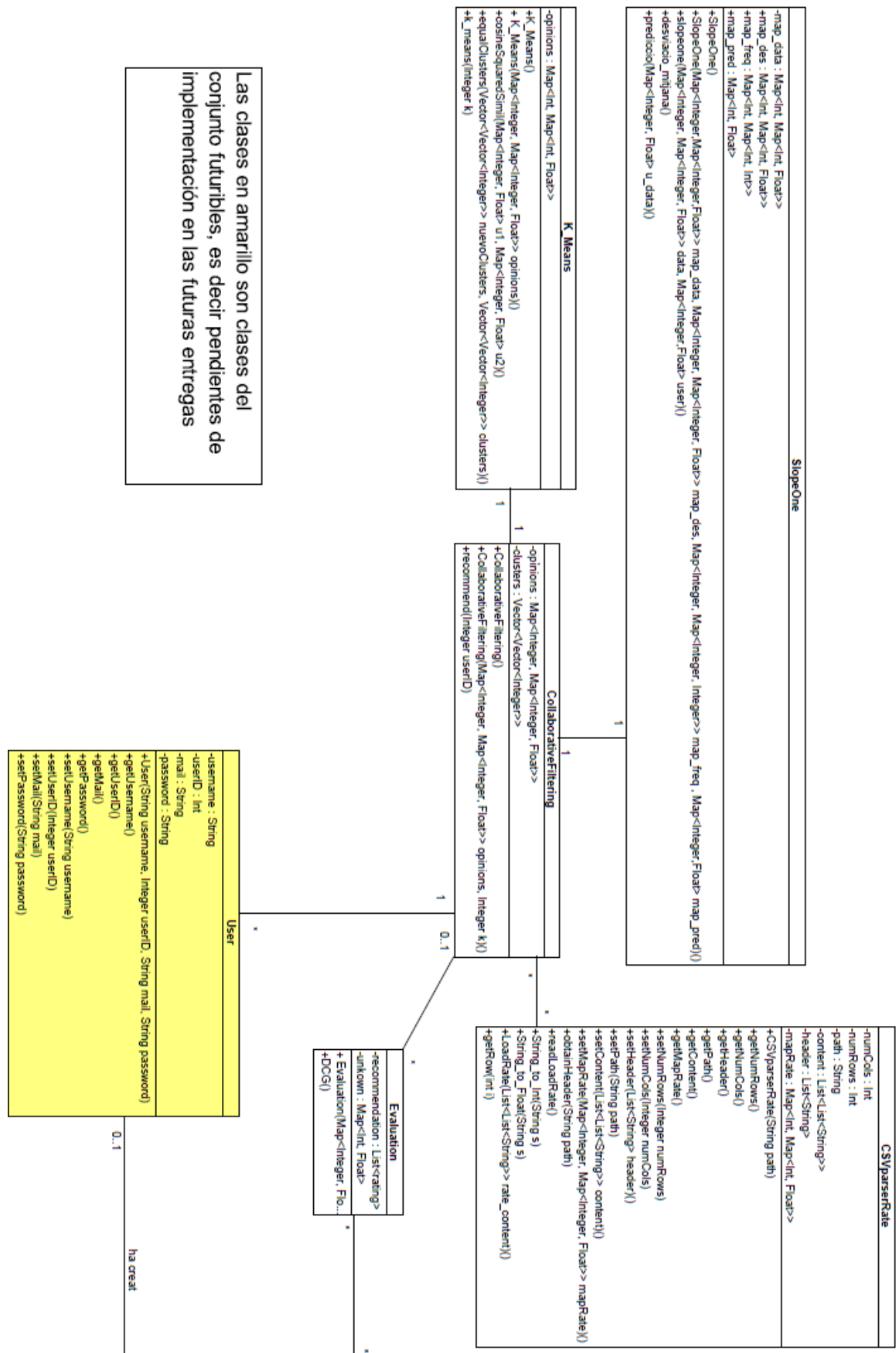
### 1.2.21 Evaluar Recomendación

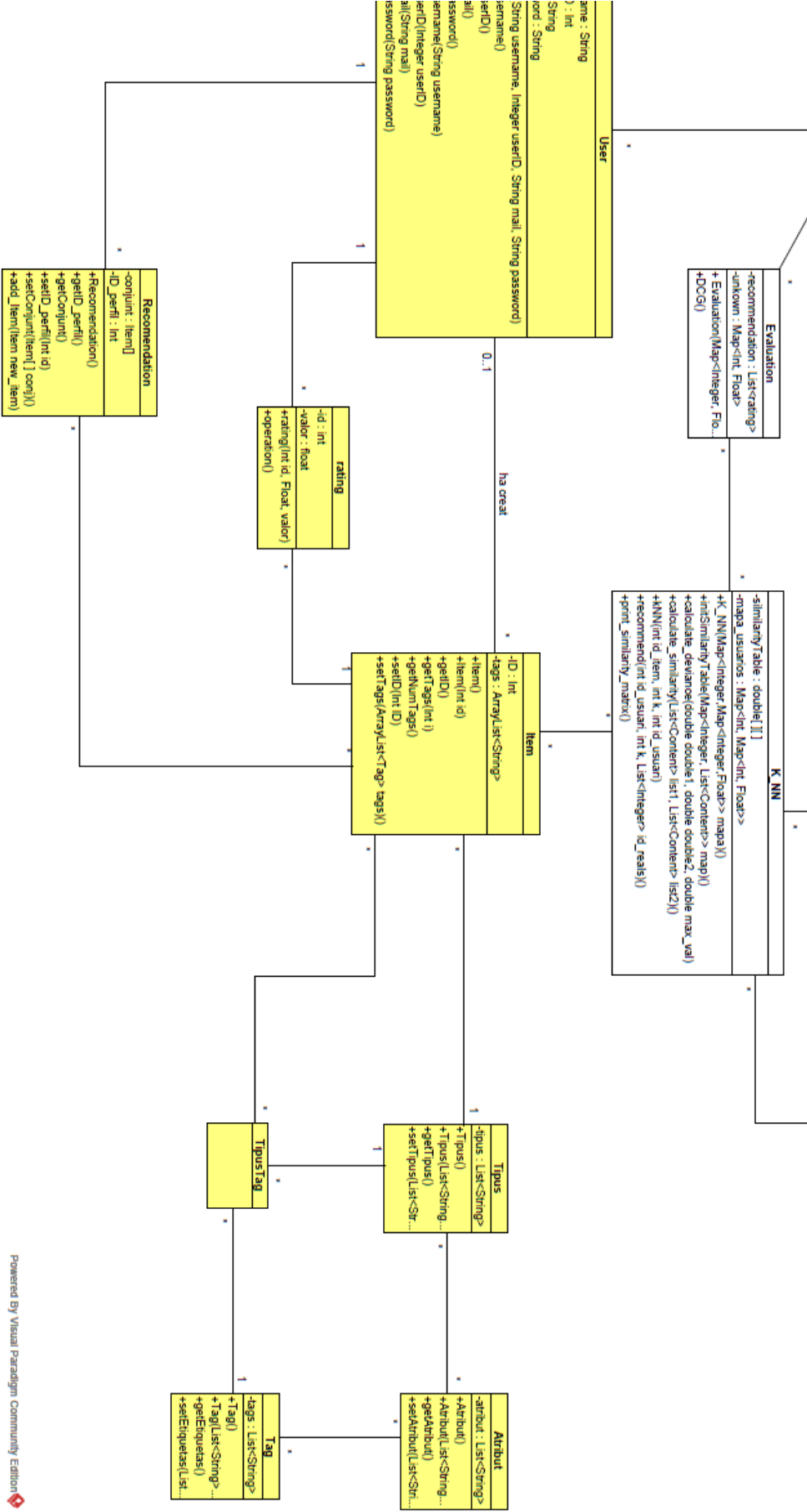
Actor Principal	Actor
Precondición	El actor debe tener una cuenta logueada en curso.
Disparador	El actor quiere evaluar una recomendación
Escenario Principal de éxito	<ol style="list-style-type: none"><li>1. El actor indica que quiere evaluar una recomendación.</li><li>2. El sistema verifica que el actor está logueado como usuario.</li><li>3. El actor evalúa la recomendación.</li><li>4. El sistema guarda la evaluación de la recomendación y lo almacena en el sistema</li></ol>
Extensiones	<ol style="list-style-type: none"><li>2a. El actor no está logueado como usuario.<ol style="list-style-type: none"><li>2a1. El sistema informa que es necesario haber logueado con una cuenta de usuario para poder acceder a la configuración.</li><li>2a2. El actor debe loguearse como usuario y volver a realizar el punto 1.</li></ol></li><li>3a. El actor introduce una evaluación no apta.<ol style="list-style-type: none"><li>3a1. El sistema informa que la evaluación introducida no es soportada.</li><li>3a2. El actor debe volver a introducir una evaluación válida.</li></ol></li><li>1. El actor cancela la acción de evaluar una recomendación.</li></ol> <p>Acaba el caso de uso</p>

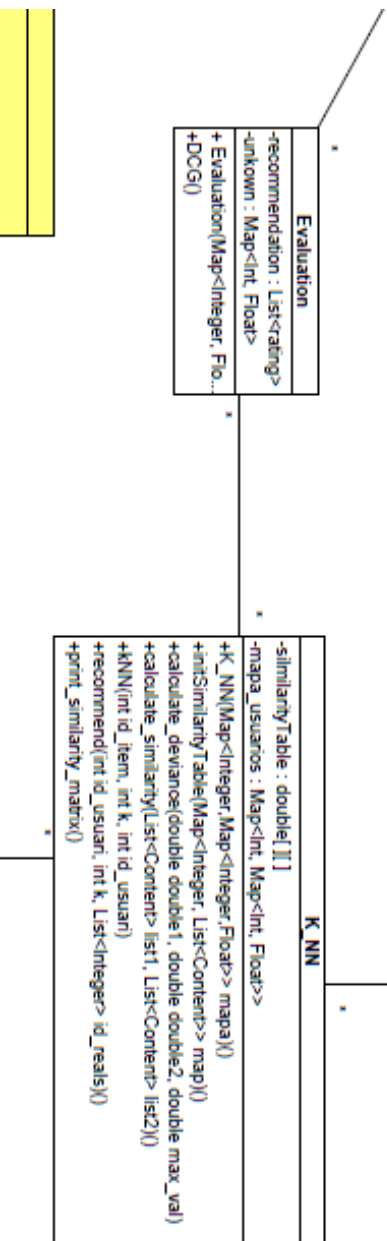
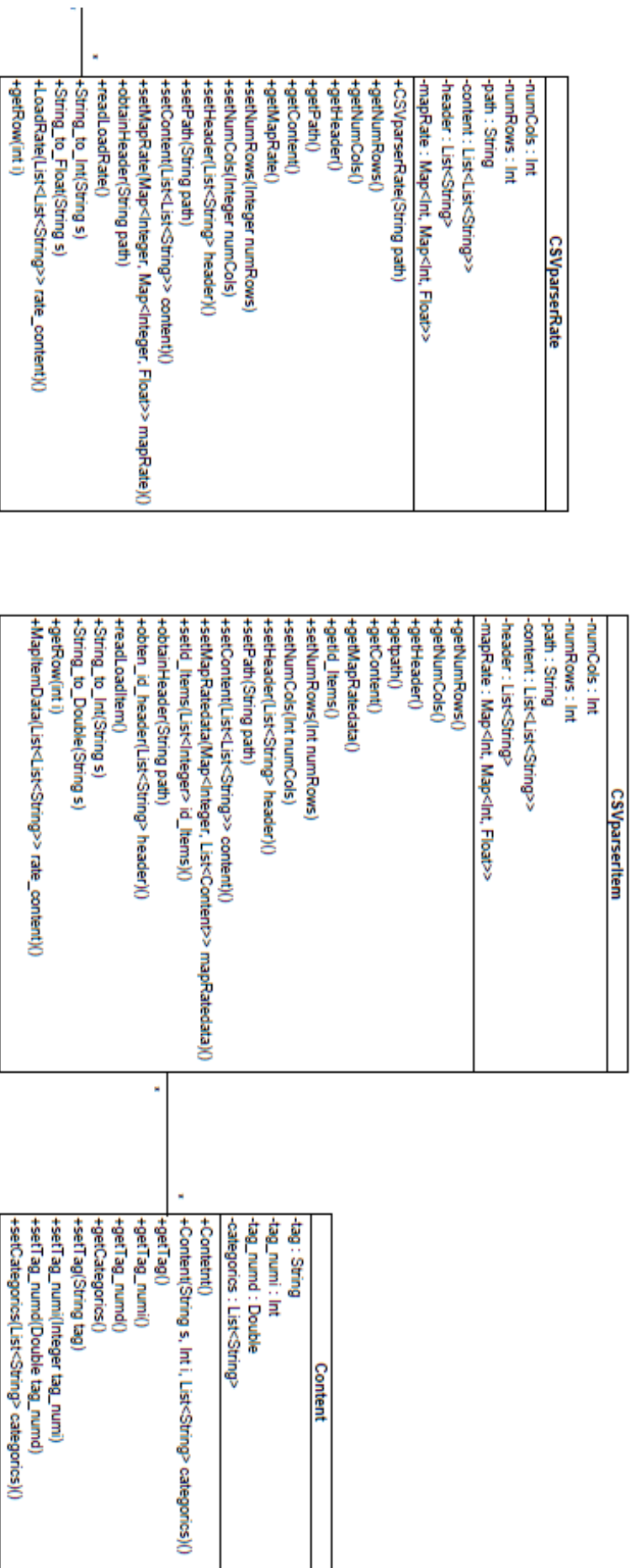


## 2. Clases del modelo conceptual de datos

### 2.1 Diagrama (para ver mejor abrir el pdf uml\_datos\_prop\_entrega1.pdf)








## 2.2 Descripción de clases

**Anotación:** En el doxygen tenemos una descripción detallada de los métodos de cada clase.

### 2.2.1 CSVParserItem

La clase CSVParserItem se encarga de realizar la lectura de los ficheros csv que tienen un formato como los ejemplos items.csv y procesa los datos con la finalidad de facilitar las operaciones de los datos en los algoritmos.

Básicamente se utiliza la librería Scanner para poder leer el documento y mediante una expresión regex nos separa el documento de forma que nos proporciona una lectura cómoda de los datos para así procesarlos convenientemente para el uso de los algoritmos. Esta lectura la realizamos mediante una *List<List<String>>* llamada *content*. También contamos con una *List<String>* *header* que contiene la cabecera del documento, es necesaria para identificar los identificadores y para evitar que un elemento tenga dos filas.

Con respecto al preproceso de los datos hemos planteado el siguiente escenario: con tal de facilitar las operaciones de los algoritmos, hemos tenido en cuenta los tipos booleanos, enteros, doubles y los elementos categóricos para hacer más eficiente el trato de datos, además de los strings, aunque sean bastante ineficientes. Hemos creado un *Map<Integer, List<Content>>* *mapRateData* donde los datos están indicados en base la categoría del elemento comentado anteriormente. A continuación usamos la clase Content para poder identificarlos y almacenarlos correctamente en nuestra estructura.

Finalmente contamos con una *List<String>* *id\_Items* que almacena todos los ids de los ítems teniendo en cuenta su posición en la lectura y procesado del documento csv.

### 2.2.2 CSVParserRate:

La clase CSVParserRate se encarga de realizar la lectura de los ficheros csv que tienen un formato como los ejemplos ratings.csv y procesa los datos con la finalidad de facilitar las operaciones de los datos en los algoritmos

En esta usamos la librería Scanner para poder leer el documento y mediante una expresión regex, menos complicada que la de la anterior clase, nos permite separar el documento en un formato que nos otorga una mayor facilidad para la

manipulación de los datos. Esta lectura la realizamos mediante una *List<List<String>>* llamada *content*. También contamos con una *List<String>* *header* que contiene la cabecera del documento, es necesaria para identificar el orden en los que entran los elementos para introducirlos correctamente en la estructura diseñada.

Para el preproceso de los datos hemos usado un *Map<Integer, Map<Integer,Float>>* *mapRate* para almacenar los de forma que por cada id de un usuario se tengan los id de los ítems que lo relacionan con su valoración correspondiente.

### **2.2.3 Content**

En la clase *content* se dotan una serie de elementos para hacer posible el preproceso de datos de la clase *CSVParserItem*. En esta podemos encontrar un string, que sirve de identificador o de elemento para procesar el contenido. Un entero el cual es usado para determinar si el objeto anterior se puede convertir en este tipo y si es así atribuirlo. Un double donde se ejerce la misma funcionalidad que con el entero. Por último cuenta con una *List<String>* *categorys* donde se almacenan el conjunto de elementos que forman parte de este grupo.

### **2.2.4 K-NN**

La clase K-NN implementa el algoritmo de k-nearest-neighbours para calcular recomendaciones usando la filosofía del content based filtering. El principal atributo de esta clase es la llamada *similarityTable*, una matriz de doubles que almacena las similitudes entre todos los ítems del sistema. Usando esta matriz, la clase hace uso del método *k-nn* para encontrar los k ítems más parecidos a un ítem dado (no valorados por el usuario).

También almacenamos el mapa que nos informa de los ítems que ha valorado cada usuario, así como de la valoración que le ha dado a cada ítem, y una lista de enteros, que almacena los id's reales de los ítems. Esta lista es necesaria porque, para hacer el cómputo del algoritmo, el método usa la posición de los ítems en la tabla de entrada del programa como id's, por una cuestión de comodidad y sinergia con la matriz de similitudes.

Para hacer una recomendación para un usuario dado, usamos el método *k-nn* sobre todos los ítems que el usuario ha valorado, y obtenemos los k ítems más parecidos a todos ellos. Una explicación más detallada se encuentra en el desarrollo del algoritmo de K-Nearest-Neighbours (sección 3.3).

### 2.2.5 K\_Means

La clase *K\_means* implementa el algoritmo *K means* (K medias) para organizar los usuarios en K grupos diferentes, estos grupos se constituyen por usuarios con opiniones parecidas. El único atributo de esta clase es *opinions*, el cual representa las valoraciones de los usuarios sobre los ítems. Esta representación se hace mediante un *Map<Integer, Map<Integer, Float>>*. En la sección 3.1 hay la explicación del algoritmo *K Means* para encontrar los susodichos K grupos.

### 2.2.6 Slope One

La clase *Slope One* implementa el algoritmo *Slope One*, el cual estima una valoración para un ítem no valorado de un usuario. Consta de dos métodos para calcular el algoritmo y un tercero para la llamada de estas y limpiar los datos. La clase consta de cuatro atributos, los cuales tres son matrices de maps y un map. Hay el map de datos *map\_data* que contiene todos los datos necesarios para calcular la predicción. Está representado por *Map<Integer, Map<Integer,Float>>*, que representan *UserID<ItemID,Valoración>*. Hay el map *map\_des* que contiene las diferencias entre dos ítems tal que *Map<Integer,Map<Integer,Float>>* que se traduce a *ItemID\_1<ItemID\_2,Diferencia>*. Hay el map de frecuencias *map\_freq* que se guardan las veces que hemos computado un par de ítems, se ve representado *Map<Integer,Map<Integer,Float>>* que toman el significado de *ItemID\_1<ItemID\_2,Veces\_computados>*. Y finalmente el map de predicción *map\_pred* el cual guarda la predicción hecha para el usuario que queremos. Está representado por *Map<Integer,Float>* que tienen el valor de *<ItemID,Valoracion\_estimada>*. En el apartado 3.2 hay una explicación más extensa y que profundiza más en el funcionamiento y cálculo del algoritmo.

### 2.2.7 CollaborativeFiltering

La clase *CollaborativeFiltering* implementa la combinación de los algoritmos *K Means* y *Slope One* para crear una recomendación para un usuario al que llamaremos A. En esta se usa el primer algoritmo para agrupar los usuarios parecidos en los mismos grupos. Posteriormente localizamos el conjunto de usuarios al que pertenece A y usamos este grupo de usuarios como input para el algoritmo *Slope One*. *Slope One* nos calcula entonces una predicción de las valoraciones que haría A de los ítems que no ha valorado, de las cuáles elegimos las 10 más altas para nuestra recomendación.

### 2.2.8 Evaluation

La clase Evaluation implementa el cálculo de la evaluación de las recomendaciones. A esta clase se le introduce la recomendación y un conjunto de valoraciones del usuario al que estamos haciendo la recomendación que inicialmente eran desconocidas para hacer la recomendación(unknown). La evaluación es el cálculo del Discounted Cumulative Gain, cuya fórmula es:

$$DCG = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

En la que se tiene una recomendación de  $p$  ítems ordenada de forma decreciente. La  $i$  representa la posición del ítem en la recomendación. El valor  $rel_i$  es la valoración que el usuario ha dado al ítem en la posición  $i$  de la recomendación. Cuanto mayor sea el DCG mejor es la recomendación.

### Clases futuras:

El siguiente conjunto de clases está formado por clases que, o bien, no están implementadas, o bien, están implementadas en una versión inicial y no se usan durante la ejecución del programa. Es posible que estas clases no existan en la versión final del proyecto.

### 2.2.9 Ítem

La clase representa cada ítem que se encuentra en el sistema de relaciones dado. Cada uno tiene un ID que lo identifica, siendo así único, y además consta de sus etiquetas.

### 2.2.10 User

La clase representa a cada usuario que tiene interacción con el sistema de relaciones. Cada uno requiere de un perfil, el cual está formado por un identificador único, un nombre al cual hacerse referencia, un correo electrónico el cual se vincula a la cuenta y finalmente una contraseña para poder acceder y obtener cierta seguridad. Podremos ver que este podrá tomar acciones con respecto a los ítems y sobre el filtraje de estos.

### **2.2.11 Recomanació**

La clase representa el sistema empleado para el filtraje de los ítems, según los algoritmos establecidos en el proyecto, en base a lo que el usuario detalla. Estos algoritmos son: k-means, kk-neighbourhood y slope one. El primero hace una agrupación en base a ... , el segundo es respecto ... y el último ....

### **2.2.12 Tipus**

La clase representa una clasificación sobre las características con respecto a un determinado ítem. Esta tiene un atributo, *List<String> tipus*, el cual contiene un conjunto de atributos que definen de qué tipo es un determinado ítem.

### **2.2.13 Atribut**

La clase representa la serie de estos que integra cada ítem. Cada uno de estos atributos forman parte de un cierto Tipus, otorgando así una serie de características. Por cada serie de ítems y por cada cierto tipo, el sistema de relaciones tiene uno o más atributos que describen cada uno de estos elementos.

### **2.2.14 Tag**

La clase representa el conjunto de etiquetas atribuidas a cada atributo. Cada una de ellas tiene un conjunto de etiquetas *List<String> tags* en el que están contenidas.

### **2.2.15 TipusTag**

La clase representa la asociación entre las clases Tipus y Tag, dado que es necesaria para hacer una clasificación más estricta y correcta de los diferentes ítems. Así de cara a las búsquedas nos encontramos una relación más homogénea de lo que se quiere encontrar.



### 3. Descripciones de los algoritmos

#### 3.1 K-Means

La aplicación del algoritmo de K means se hace en la clase *K-Means*. En primer lugar se usa la creadora de la clase pasando como parámetro un `Map<Integer, Map<Integer, Float>>` que representa las valoraciones de los usuarios de los ítems y se lo atribuimos al atributo *opinions* de la clase. Después de esta inicialización ejecutamos el método *k\_means* para encontrar los los k clústers de usuarios, para ello usamos el algoritmo más común que usa una técnica iterativa.

Consideramos que los usuarios forman un vector en el cual cada ítem representa una componente. Las componentes no nulas para un usuario son las valoraciones de los ítems que ha valorado. Las medias representan “usuarios medios” de cada conjunto de usuarios y por lo tanto también se entienden como vectores.

En primera instancia, inicializamos los k conjuntos cuyos índices van de 0 a k-1. Para ello recorremos el Map *opinions* atribuyendo a cada usuario un conjunto de forma que el usuario que aparece en la i-ésima posición, empezando desde i = 0, pertenece al (i%k)-ésimo conjunto. Una vez ya tenemos los k conjuntos inicializados, debemos calcular sus medias de forma que

$$M_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j$$

dónde  $S_i$  es el conjunto i-ésimo y  $M_i$  su media. No hay que olvidar que esto no son valores escalares sino que son vectores y realmente estamos sumando sus componentes.

Ahora entramos en la parte iterativa del algoritmo. Esta parte tiene dos fases:

1. Fase de asignación.
2. Fase de actualización.

**Fase de asignación:** Calculamos la distancia entre los usuarios y las medias y colocamos cada usuario en el conjunto que tiene la media más cercana (menor distancia). En nuestro caso no hemos usado una función de distancia sino que hemos usado una función de similitud, el coseno cuadrado, que calcula el cuadrado del coseno del ángulo que forman los vectores del usuario y la media. Consecuentemente cada usuario es asignado al conjunto con el que tiene mayor similitud con la media.

**Fase de actualización:** Volvemos a calcular la media de los conjuntos con la fórmula que hemos usado anteriormente.

Estos dos procesos se repiten indefinidamente hasta que en la fase de asignación los conjuntos creados son los mismos que en la iteración anterior.

### 3.2 Slope One

La aplicación del algoritmo de Slope One se hace en la clase *SlopeOne*. El algoritmo consiste en estimar/predecir una valoración para un ítem concreto que le daría un usuario, a partir de las valoraciones de ese ítem de los otros usuarios.

El cálculo del algoritmo empieza con la función *slopeone* que le llega por parámetros una matriz *Map<Integer, Map<Integer, Float>>* que representa las valoraciones de los usuarios de los ítems y se lo atribuimos al atributo *map\_data*, que será nuestra base de datos para el cálculo. Y un *Map<Integer, Float>* denominado *user* que representa las valoraciones de los ítems del usuario al que queremos estimar las valoraciones que aún no ha realizado en ciertos ítems.

Este método básicamente centraliza la ejecución del algoritmo, haciendo llamadas a los dos métodos principales del cálculo.

Primero llama a la función *desviacio\_mitjana()*. Aquí nos calculará con nuestra matriz *map\_data* la desviación de las valoraciones de cada ítem con los otros ítems, por usuario. Lo que nos construirá una matriz *map\_des* simétrica con la diagonal con valor 0.

Seguidamente llama a la función *prediccio(user)* que se le pasa por parámetro el map del usuario que estimaremos sus valoraciones. Y nos llenará el map *map\_pred* con las estimaciones calculadas por cada ítem, los ítems que sí que había evaluado el usuario y los que no, solamente nos interesan estos últimos.

Así que el último procedimiento del método *slopeone* es un bucle que itera a través del map *user* y elimina los items que sean compartidos entre este map y *map\_pred* dejando solamente los ítems estimados. Finalmente retorna *map\_pred*.

#### **Desviacio\_mitjana:**

En este método nos computa la matriz de desviaciones de cada ítem respecto los otros ítems. El resultado será el de una matriz simétrica con la diagonal a 0. La fórmula para el cálculo es la siguiente.

$$dev_{j,i} = \sum_{u \in S_{j,i}(X)} \frac{u_j - u_i}{card(S_{j,i}(X))}$$

Siendo  $u_j$  la valoración del usuario  $u$  al ítem  $j$ , menos  $u_i$  siendo la valoración del usuario  $u$  al ítem  $i$ , donde  $u$  es el vector de las valoraciones del usuario  $u$ .

De aquí sacamos la diferencia, y lo dividimos por  $card(S_{j,i}(X))$  el conjunto de usuarios que han valorado  $j$  e  $i$ . Obtenemos la desviación media.

Aplicada esta fórmula al código, tenemos que iterar primeramente por todos los usuarios de la matriz de datos, y a través de ella por todos los datos de cada usuario. Añadimos el ítem que tratamos, si no estaban, a las matrices de desviación y frecuencia, esta última nos servirá para calcular las cardinalidades de cada conjunto de usuarios. Y tenemos que volver a iterar a través de los datos de un segundo usuario. Entonces calculamos la diferencia entre las valoraciones de los dos usuarios, y lo guardamos en la matriz de desviaciones, pero no sin antes haber comprobado y sumado si es necesario con las diferencias para ese ítem acumuladas de otros usuarios anteriores.

Vamos sumando las veces que hemos tratado aquel ítem para un par de usuarios y se guarda en la matriz de frecuencia, teniendo en cuenta la acumulación de este valor para casos anteriores. Para finalizar después de haber recorrido todos los usuarios. Recorremos una última vez la matriz de desviaciones, y con la diferencia la dividimos con la cardinalidad del mapa de frecuencias y obtenemos la desviación media de cada ítem.

### Predicción:

En este segundo método calculamos la segunda parte del algoritmo, la parte de la predicción de evaluación de ítem. El algoritmo nos propone la siguiente fórmula para hacer la predicción.

$$P^{S1}(u)_j = \frac{1}{card(R_j)} \sum_{i \in R_j} (u_i + dev_{j,i})$$

$$R_j = \{i \mid i \in S(u), i \neq j, card(S_{j,i}(X)) > 0\}$$

Pero realmente se puede simplificar el cálculo con valoraciones densas, las que entendemos con  $card(S_{j,i}(X)) > 0$  para casi todos los  $j$  e  $i$ . Así que al final queda que la predicción aproximada es:

$$P^{S1\approx}(u)_j = \bar{u} + \frac{1}{card(R_j)} \sum_{i \in R_j} dev_{j,i}$$

Vemos que se ha simplificado la valoración del usuario al ítem  $i$   $u_i$ , dentro del sumatorio, por la media de las valoraciones de  $u$ . Facilitando mucho el código donde simplemente tenemos que calcular un sumatorio de todas las desviaciones del ítem respecto al que estimaremos. Entre la cardinalidad de este conjunto más la suma de la media de valoraciones de los ítems del usuario al que estamos prediciendo.

Esto aplicado al código, se traduce con inicializar primeramente el mapa de predicción con todos los ítems que aparecen en el mapa de desviaciones. Después calcular la media de las valoraciones del usuario en concreto que estamos tratando. Y calcular la suma de todas las desviaciones de ese ítem con los otros entre la cardinalidad de este conjunto. Se suman la desviación definitiva con la media del usuario y tenemos la predicción para ese ítem. Y devuelve el mapa de predicción. Este código calcula la predicción para todos los ítems del conjunto, es decir los valorados y los que no, por el usuario, a nosotros solamente nos interesa los que no ha valorado así que en la función principal ya se limpia el mapa de predicción.

### 3.3 K-Nearest-Neighbours

La aplicación del algoritmo de K-Nearest-Neighbours se hace en la clase *K-NN*, usando la matriz de similitudes entre todos los ítems del sistema.

La clase contiene un método de inicialización de la matriz de similitudes, que se ejecuta cuando leemos el conjunto de ítems de entrada. La inicialización consiste en calcular las similitudes entre todos los ítems dados y almacenarlas en la matriz.

El cálculo de las similitudes entre ítems se hace comparando sus tags, que se reciben como instancias de la clase *Content*. La similitud entre dos ítems será la suma de las similitudes entre todos sus tags. Para normalizar este valor y mantenerlo en un rango entre 0 y 1, la similitud entre dos ítems distintos se divide entre la similitud entre un ítem y él mismo (que siempre será máxima en una misma fila de la matriz).

En función del tipo del tag (booleano, entero, double, categórico o string), el valor de la similitud es calculado de manera distinta y tiene una importancia diferente.

En nuestro caso, se valoran como más importantes las coincidencias en los tipos booleanos, seguido por los tipos categóricos, estos seguidos por los enteros y los doubles, y finalmente los strings.

Las coincidencias entre dos valores booleanos y dos valores categóricos aportan una similitud que se calcula de forma absoluta; la similitud prende valor máximo si los valores coinciden y, en cambio, prende valor mínimo (0) si no coinciden.

Las coincidencias entre dos valores enteros y dos valores double, por otra parte, aportan una similitud que se calcula de forma relativa a la variancia entre los valores dados, es decir, la similitud puede tomar cualquier valor entre el valor máximo (que se tomará si los números son iguales) y el valor mínimo (0). La similitud será igual a:

$$(1 - x) * X$$

siendo  $x$  la variancia entre los dos números dados,  $X$  el valor máximo que puede tomar la similitud entre enteros/doubles.

Una vez calculadas las similitudes entre todos los ítems del sistema, cada vez que queramos encontrar los  $k$  elementos más similares a un ítem dado, únicamente tendremos que recorrer la fila de la matriz correspondiente al ítem dado, y quedarnos con los  $k$  elementos con mayor similitud.

Este preciso procedimiento se aplica en el método  $kNN$ , el cual dado un  $id$  de un ítem  $x$  y un  $k$ , retorna los  $k$  elementos más parecidos a  $x$ . Esto lo conseguimos haciendo uso de una cola de prioridad que almacena parejas de  $id$ 's y similitudes. Para implementar esta cola de prioridad de parejas, se ha implementado una clase ordenable `Pair`, que almacena un entero (el  $id$  del ítem con el que se compara) y un double (la similitud entre los dos ítems). La cola de prioridad se ordena en función de la similitud, dejando los `Pairs` con menor similitud como los primeros que saldrán en caso de llegar un `Pair` con mayor similitud. Al terminar la iteración sobre la fila del ítem  $x$ , el método  $kNN$  retorna una lista con los  $k$  ítems más cercanos a  $x$ .

Además, este método podemos aplicarlo para calcular una recomendación para un usuario dado, llamémosle  $T$ . El método *recommend*, que recibe como parámetros el  $id$  del usuario que quiere una recomendación ( $T$  en este caso), y cuantos ítems quiere en la recomendación ( $k$ ) se encarga de calcular una recomendación de  $k$  elementos para el usuario  $T$ . Tomando todos los ítems que ha valorado  $T$  (a los que llamaremos  $x_i$ ), calculamos los  $k$  elementos más cercanos a cada uno de ellos (a los que llamaremos  $x_{ij}$ ). Para cada elemento presente en las listas de los  $k$ -NN, computamos la similitud global entre ese ítem y el usuario  $T$ . Esta similitud global se computa multiplicando la similitud entre un  $x_{ij}$  y  $x_i$  por la valoración que el usuario  $T$  le ha dado al ítem  $x_i$ , para todos los  $x_i$ .

Es decir, la similitud global entre  $x_{ij}$  (ítem no valorado por T) y T será:

$$Sim(x_{ij}, T) = \sum_{i=0}^n sim(x_{ij}, x_i) * val(x_i)$$

Calculado este valor para todos los  $x_{ij}$  presentes en todos los vectores k-NN (donde puede haber  $x_{ij}$  repetidos, no causa ningún problema), lo único que nos falta es tomar los k ítems con mayor similitud global entre ellos y el usuario T.

El cálculo del algoritmo se realiza usando la posición de los ítems en la tabla de entrada como identificador de los ítems, por una cuestión de comodidad y sinergia con la matriz de similitudes. Es por ello que, para retornar el id real de un ítem, necesitamos una lista que nos traduzca la posición de la tabla que ocupa un ítem a su identificador real. Esta lista es el otro parámetro global de la clase, que debe pasarse como parámetro al instanciar un objeto de la clase.

## 4. Estructuras de datos y costes

### 4.1 Decisión estructuras de datos

#### **ArrayList:**

ArrayList es una clase propia de Java muy similar a los arrays de los lenguajes de programación, pero esta nos permite añadir y quitar elementos sin tener en cuenta el tamaño del mismo. La hemos tenido en cuenta de cara a la agregación o eliminación de elementos de las listas que hemos desarrollado.

#### **List:**

Para la lectura de los documentos csv hemos optado por el desarrollo de las listas. Más concretamente hemos creado una `List<List<Strings>>` de forma que las listas de Strings contienen los elementos del csv, separados correspondientemente, y situados de la misma forma tal y como dispone el documento. Finalmente la lista que engloba el conjunto de las anteriores nos permite hacer referencia a las filas, como dispone el formato del documento csv. A priori hemos pensado que esta era una de las mejores formas de poder realizar la lectura. Una vez finalizado dicho proceso hemos elegido una nueva estructura con tal de tener la información de una forma más eficiente, ya que las consultas de las listas pueden resultar costosas.

#### **TreeMap:**

La estructura que hemos decidido utilizar es el TreeMap. Principalmente porque hemos visto que la entrada de elementos es un tanto significativa y al no saber si este número se podía expandir, optar por esta opción que nos ofrece coste en tiempo  $\log(n)$  nos parecía conveniente. Es cierto que por el momento la ordenación podría no tenerse mucho en cuenta, pero al tratar con una cantidad de datos significativa también nos ha hecho remarcar el uso de esta opción.

Esta clase está implementada a partir de un árbol binario de búsqueda equilibrada, esto significa que el equilibrio depende de las etiquetas, al agregar o eliminar los elementos. Por lo tanto el equilibrio es importante para mantener un buen rendimiento, que corresponderá directamente con la altura del árbol. Hemos de evitar que se desequilibre el árbol con tal de dar una altura mayor y que empiece así a afectar al rendimiento. Nos garantiza entonces que el tiempo de los métodos `contains`, `get` o `put` serán  $O(\log(n))$ .

El principal uso ha sido para la organización de la información de los documentos csv. Para el csv de los ítems hemos desarrollado un `Map<Integer, List<Content>>` el cual la clave hace referencia a la fila con la que se ha obtenido dichos elementos y la lista contiene el conjunto de valores leídos previamente pero modificados con tal de proporcionar una mejor resolución a los algoritmos. Para el caso de los otros documentos, al tener definidos el conjunto de valores, hemos implementado un `Map<Integer, Map<Integer, Float>>` atribuyendo la clave del primer map el id del usuario, en la segunda clave el id del ítem y el valor del último la valoración correspondiente.

### **Hashmap:**

Esta clase no la hemos tenido en cuenta de cara a la implementación pero creemos que es una buena idea resaltar sus características para hacer una mejor comparación y ver porque la opción que hemos escogido resulta más adecuada. El `HashMap` ofrece ventajas en términos de rendimiento, que es constante para operaciones como `get` y `put`, pero ante un crecimiento de la estructura a través del tiempo se debería de tener en cuenta ya que podría cambiar su coste.

Los dos factores a tener en cuenta son la carga y la capacidad. La primera se refiere a la amplitud de los “buckets” y la segunda al número de “buckets” que se han creado. A medida que crece el número de estos se ha de rehashear para crear más “buckets”, lo que puede ser una operación costosa dependiendo del número de entradas. Si el factor de carga se mantiene por debajo del 75% la implementación en java funciona mejor, lo que quiere decir es que para mantener el rendimiento siempre se asigna más memoria de la que se necesita para el almacenaje de las entradas, por eso debemos usarlo exclusivamente si se cumplen los siguientes aspectos: queremos un coste en tiempo constante, tenemos una idea de la dimensión de la colección y no añadiremos o eliminaremos los elementos con regularidad.



## 4.2 Costes de las diferentes funciones

### CSVparserItem:

Estas son las funciones con un coste más relevante:

- **obtain\_id\_header(List<String> header)** : Esta función que obtiene la posición del elemento donde se encuentra el id, tiene un coste lineal  $O(n)$ , siendo  $n$  el tamaño de header, ya que debe recorrer la lista.
- **readLoadItem()**: Esta función tiene un coste cuadrático ya que al encargarse de leer el documento csv necesariamente necesita recorrer todos los elementos.
- **MapItemData(List<List<String>> rate\_content)**: Esta función también tiene un coste cuadrático ya que tiene que recorrer cada elemento del rate\_content para procesarlo como dato para los algoritmos.

### CSVparserRate:

Estas son las funciones con un coste más relevante:

- **obtain\_id\_header(List<String> header)** : Esta función que obtiene la posición del elemento donde se encuentra el id, tiene un coste lineal  $O(n)$ , siendo  $n$  el tamaño de header, ya que debe recorrer la lista.
- **readLoadRate()**: Esta función tiene un coste cuadrático ya que al encargarse de leer el documento csv necesariamente necesita recorrer todos los elementos.
- **LoadRate(List<List<String>> rate\_content)**: Esta función también tiene un coste cuadrático ya que tiene que recorrer cada elemento del rate\_content para procesarlo como dato para los algoritmos.

### Content:

No tiene funciones con costes relevantes, se usa como soporte para el desarrollo de la clase CSVparserItem.

## K\_NN:

Estas son las funciones con un coste más relevante:

- **initSimilarityTable(Map<Integer, List<Content>> map):** La inicialización de la tabla de similitudes entre ítems tiene coste pseudo-cuadrático, ya que se aprovecha el hecho de que la matriz es simétrica (antes de aplicar la normalización) para computar la  $(n^2)/2$  operaciones de cálculo de similitud. Por lo que podemos decir que este método tiene coste  $O(n^2)$ , lo cual es asumible ya que solo se ejecuta una vez, cuando inicializamos el programa.
  - **calculate\_similarity(List<Content> list1, List<Content> list2):** Dadas dos listas de tamaño  $n$ , el cálculo de la similitud es lineal con  $n$ , ya que la comparación de tags se hace uno a uno y las dos listas se recorren simultáneamente. En caso de recibir un tag multivaluado, la comparación entre estos también se hace de manera lineal con el número de evaluaciones del tag, ya que se emplea un algoritmo de comparación basado en el cálculo de la intersección de las dos listas en tiempo lineal. Por lo que podemos concluir que este método tiene coste  $O(n)$ .
  - **kNN(int id\_item, int k, int id\_usuario):** El método para el cálculo de los  $k$  vecinos más cercanos a un ítem dado es lineal con el número de ítems, ya que recorremos la fila de la matriz de similitudes del ítem dado. Considerando las comparaciones, las asignaciones y la gestión de la cola de prioridad como  $O(1)$ , podemos concluir que este método tiene coste  $O(n)$ .
  - **Map<Integer,Float> recommend(int id\_usuario, int k):** El método para hacer una recomendación basada en el content based filtering a un usuario dado tiene coste pseudo-cuadrático (en la versión actual), ya que se calculan los  $k$  vecinos más cercanos de todos los ítems que el usuario ha valorado. En el caso peor, en el que el usuario ha valorado todos los ítems del sistema (menos un conjunto pequeño, para poder recibir una recomendación), el método kNN (de coste  $O(n)$ ) se ejecutaría  $n$  veces, dando lugar a un coste de  $O(n^2)$ . Ahora bien, este coste se puede reducir fácilmente a un coste lineal si, en vez de calcular los  $k$  ítems más parecidos de todos los que ha valorado el usuario, solo lo hiciéramos de los  $t$  ítems con mejor valoración.
- Aunque es sabido que esta implementación reduciría el coste de ejecución drásticamente en casos desfavorables, se ha valorado que, para esta primera entrega, las entradas no tendrán casos muy desfavorables y la precisión en la recomendación se valora más positivamente que el tiempo de ejecución. En caso de cambiar de opinión en el futuro, la implementación de esta modificación es sencilla y se aplicará sin problema.

## Pair:

No tiene funciones con costes relevantes, se usa como soporte para el desarrollo de la clase *K\_NN*, concretamente para la implementación de una cola de prioridad que almacene dos valores.

## K\_Means:

Estas son las funciones de coste más relevante:

- **cosineSquaredSimil(Map<Integer, Float> u1, Map<Integer, Float> u2):** Si consideramos que los dos usuarios tienen  $m$  valoraciones esta función tiene un coste  $O(m \cdot \log m)$ .
- **equalClusters(Vector<Vector<Integer>> nuevoClusters, Vector<Vector<Integer>> Clusters):** En el peor caso hay que recorrer todos los usuarios y por lo tanto el coste es lineal,  $O(n)$  siendo  $n$  el número de usuarios.
- **k\_means(Integer k):** En esta función tenemos el bucle *while* que se ejecuta un número indefinido de veces. Como el coste depende de muchos factores consideraremos que tenemos  $n$  usuarios con cuyo número medio de valoraciones es  $m$ , para el cálculo consideraremos que todos los usuarios tienen el mismo número de valoraciones. Dentro de este bucle, el proceso más costoso es calcular a qué cluster pertenecen los usuarios. Todos los usuarios deben encontrar la media más cercana ejecutando la función *cosineSquaredSimil* así que tenemos un coste  $O(m \cdot \log m) \cdot n = O(n \cdot m \cdot \log m)$ . Suponemos que el bucle *while* no se ejecutará más de  $n$  veces ya que como mínimo un usuario llegará a su cluster correcto en cada iteración. Tenemos entonces un coste de  $O(n^2 \cdot m \cdot \log m)$ .

## CollaborativeFiltering:

Sólo ejecuta funciones de las clases *K\_Means* y *Slope One*. La creadora, con parámetros, tiene el mismo coste que *k\_means*,  $O(n^2 \cdot m \cdot \log m)$ . La función *recommend* tiene el mismo coste que *Slope One*.

## Evaluation:

Funciones de coste más relevante:

- **Evaluation(Map<Integer, Float> unknown, Map<Integer, Float> recommendation):** El coste de esta función es ordenar la recomendación, así que tenemos un coste  $O(n \cdot \log n)$ , siendo  $n$  el número de ítems recomendados.
- **DCG():** Esta función sólo itera a través de toda la recomendación, tiene un coste lineal,  $O(n)$ .

## SlopeOne:

Estas son las funciones de coste relevante en la clase:

- **slopeone(Map<Int,Map<Int,Float>> data, Map<Int,Float> user):**  
En esta función solamente tenemos un bucle de tamaño  $n$  por lo tanto el coste es lineal  $O(n)$ .
- **desviacio\_mitjana():** En esta función nos encontramos con un triple bucle, cada bucle recorre el map entero, uno de usuarios y los otros dos de sus datos. En el caso peor sería que el número de usuarios y de ítems valorados por los usuarios fuese el mismo, es decir  $n$ . Al recorrer los maps al completo, tenemos que el coste de esta función es  $O(n^3)$ .
- **prediccio(Map<Int,Float> u\_data):** En esta función tenemos un doble bucle al calcular la mediana de desviaciones del ítem respecto al que estimaremos. En el caso peor es que el tamaño de los conjuntos de ítems a calcular sea el mismo, es decir  $n$ . Al recorrer los dos bucles al completo nos sale que el coste de esta función es  $O(n^2)$ .

En resumen, el coste de la clase sería el de  $O(n^3)$ .

## 5. Relaciones entre clases y miembros del equipo

### **Miguel:**

CSVparserItem

CSVparserRate

Content

### **Marc:**

K-NN

Pair

### **Manel:**

K-Means

CollaborativeFiltering

Evaluation

### **Oriol:**

SlopeOne

### **Clases futuras:**

Tag

Tipus

TipusTag

Atribut

User

Ítem

Recomanacio

## **6. Manual de uso de la aplicación y drivers:**

En este apartado detallaremos brevemente cómo iniciar la aplicación mediante el makefile creado. Cabe destacar que hemos creado drivers interactivos, por lo que se han de seguir las instrucciones. En los juegos de pruebas resaltamos el uso de estos y ejemplificamos con algún pequeño detalle.

Una vez adquirido el proyecto, el usuario, debe situarse en la carpeta src para correr el Makefile. Con el comando make se realizan los ejecutables, creando una carpeta llamada "EXE" en la primera capa y, una vez finalizado el proceso, se inicia el testeo principal. Para la ejecución de los drivers sugerimos que introduzca el comando make help para ver las opciones. Con estas podrá ejecutar los comandos pertinentes para proceder con cada driver. Finalmente con el comando make clean eliminamos la carpeta de los ejecutables creada anteriormente.

En la carpeta DOCS del proyecto existen descripciones detalladas del funcionamiento de los drivers, así como de los juegos de pruebas adjuntos.