



UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior (Jaén)

Trabajo Fin de Máster

DESARROLLO DE MÉTODOS DE EXPLICABILIDAD PARA SISTEMAS DE RECOMENDACIÓN GRUPAL

Alumno/a: Sahraoui Ouardighi, Mohamed

Tutor/a: Prof. D. Luis Martínez López

Cotutor/a: Prof. D. Raciél Yera Toledo

Dpto.: Informática

Julio, 2023



Universidad de Jaén
Escuela Politécnica Superior de Jaén
Departamento de Informática

Don Luis Martínez López y Don Raciél Yera Toledo, tutores del Proyecto Fin de Carrera titulado: “**Desarrollo de métodos de explicabilidad para sistemas de recomendación grupal**”, que presenta Mohamed Sahraoui Ouardighi, autoriza su presentación para defensa y evaluación en la Escuela Politécnica Superior de Jaén.

Jaén, Julio de 2023

El alumno:

Mohamed Sahraoui Ouaridighi

Los tutores:

Prof. D. Luis Martínez López

Prof. D. Raciél Yera Toledo

Agradecimientos

En primer lugar, me gustaría expresar mi sincero agradecimiento a todas las personas que han contribuido y apoyado en la realización de este Trabajo de Fin de Máster.

Quiero comenzar agradeciendo a mi familia por su constante apoyo y por brindarme una sólida base educativa. A mis padres, les estoy enormemente agradecido por su aliento incondicional, su amor y su guía en cada etapa de mi vida. Su apoyo inquebrantable ha sido fundamental para alcanzar este logro.

También quiero extender mi gratitud a mis amigos, quienes han sido un pilar fundamental durante este proceso. Su ánimo, aliento y compañía me han impulsado a superar desafíos y a mantenerme motivado en todo momento. Su amistad ha sido un regalo invaluable y estoy agradecido por tenerlos a mi lado.

Agradezco de manera especial a los profesores con los que he tenido el privilegio de aprender a lo largo de mi carrera. Su dedicación, conocimiento y pasión por la enseñanza me han inspirado y me han brindado una base sólida en mi campo de estudio. Estoy agradecido por todas las oportunidades de aprendizaje que me han brindado.

Además, quiero expresar mi más sincero agradecimiento a Raciél, mi co-tutor en este Trabajo de Fin de Máster. Su orientación, experiencia y comentarios constructivos han sido fundamentales para el desarrollo y éxito de este proyecto. Su apoyo y disposición para responder mis preguntas y brindar asesoramiento han sido invaluable.

Por último, quiero dedicar una mención especial a Luis. Además de haber sido mi profesor en un par de asignaturas, ha sido una guía constante a lo largo de mi carrera y en el desarrollo de este proyecto. Su conocimiento profundo, paciencia y pasión por la materia han sido una fuente de inspiración. Ha sido un verdadero privilegio aprender de él y estoy enormemente agradecido por su apoyo.

Índice

Índice de Figuras	7
Índice de Tablas	9
Resumen	11
Abstract	11
CAPÍTULO 1: INTRODUCCIÓN	13
1. Introducción.....	15
2. Motivación.....	16
3. Propósito del proyecto.....	17
4. Objetivos.....	17
5. Metodología y planificación del proyecto.....	18
5.1. Metodología.....	18
5.2. Planificación.....	20
6. Estructura del TFM.....	21
7. Conclusión.....	22
CAPÍTULO 2: SISTEMAS DE RECOMENDACIÓN	23
1. Introducción.....	25
2. Técnicas de recomendación.....	27
2.1. Basado en Contenido.....	29
2.2. Filtrado Colaborativo.....	32
2.2.1. Basado en Memoria.....	33
2.2.2. Basado en Modelo.....	35
2.3. Híbrido.....	37
3. Sistema de recomendación grupal.....	40
3.1. Estrategias de agregación de los modelos individuales.....	42
3.2. Métodos de agregación.....	45
4. IA Transparente: Sistemas de Recomendación Explicables.....	47
5. Conclusión.....	51
CAPÍTULO 3: DISEÑO Y DESARROLLO DEL MODELO DE EXPLICABILIDAD DE SISTEMA DE RECOMENDACIÓN GRUPAL	53
1. Introducción.....	55
2. Descripción de los componentes del modelo.....	55
2.1. Componente de recomendación.....	56
2.2. Componente de explicación.....	59
2.2.1. Estrategia de filtrado de preferencias del usuario.....	60

2.2.2. Estrategia de imputación de las calificaciones faltantes.....	61
2.2.3. Generación de explicaciones.....	62
3. Algoritmos utilizados.....	63
3.1. Filtrado colaborativo.....	64
3.2. Filtrado basado en contenido.....	64
3.3. Agregación de preferencias.....	66
3.4. Agregación de recomendaciones.....	67
4. Conclusión.....	67
CAPÍTULO 4: EVALUACIÓN EXPERIMENTAL.....	69
1. Introducción.....	71
2. Configuración experimental.....	72
3. Método de evaluación.....	73
3.1. Ganancia Acumulada Descontada Normalizada (nDGC).....	74
3.2. F1-Score.....	75
3.3. Fidelidad.....	77
4. Experimentos y resultados.....	78
4.1. Evaluación del rendimiento del sistema.....	78
4.1.1. Influencia del método de agregación.....	79
4.1.2. Influencia del tamaño de grupo.....	85
4.1.3. Influencia de la longitud de la lista de recomendación.....	87
4.1.4. Influencia de la similitud intergrupala.....	88
4.2. Evaluación de explicabilidad del modelo propuesto.....	94
4.2.1. Influencia de métodos de generación de explicaciones.....	95
4.2.2. Influencia del algoritmo de imputación de las calificaciones faltantes....	100
4.2.3. Influencia del tamaño de grupo.....	102
4.2.4. Influencia de los umbrales mínimos de soporte y confianza.....	104
4.2.5. Influencia de la similitud intergrupala.....	107
5. Conclusión.....	108
CAPÍTULO 5: CONCLUSIONES Y TRABAJOS FUTUROS.....	111
1. Conclusiones.....	113
2. Trabajos futuros.....	115
BIBLIOGRAFÍA.....	119

Índice de Figuras

Figura 1.1: Ciclo de vida de metodología incremental.....	19
Figura 1.2: Estructura de Desglose de Trabajo.....	20
Figura 2.1: Número de documentos publicados sobre los sistemas de recomendación en la base de datos Scopus.....	25
Figura 2.2: Sistema de recomendación basado en Contenido.....	30
Figura 2.3: Arquitectura de alto nivel de un sistema de recomendación basado en contenido.....	32
Figura 2.4: Sistema de recomendación basado en Filtrado Colaborativo.....	33
Figura 2.5: Descomposición de valores singulares (SVD) en tres matrices.....	36
Figura 2.6: Diseño paralelo de un recomendador híbrido.....	38
Figura 2.7: Diseño secuencial de un recomendador híbrido.....	38
Figura 2.8: Clasificación de las técnicas de para combinar estrategias de recomendación en los tres tipos de diseños híbridos.....	40
Figura 2.9: Número de documentos publicados sobre los sistemas de recomendación grupales en la base de datos Scopus.....	41
Figura 2.10: Sistema de recomendación basado en agregación de recomendaciones....	42
Figura 2.11: Sistema de recomendación basado en agregación de preferencias.....	43
Figura 2.12: Número de documentos publicados sobre los sistemas de recomendación explicables en la base de datos Scopus.....	48
Figura 2.13: Explicaciones basadas en métodos más simples.....	49
Figura 2.14: Explicaciones basadas en LIRE.....	50
Figura 3.1: Componentes del modelo de explicabilidad de sistemas de recomendación grupal.....	56
Figura 3.2: Generación de reglas de asociación utilizando el algoritmo Apriori.....	59
Figura 4.1: Impacto del método de agregación en la métrica F1-Score.....	82
Figura 4.2: Impacto del método de agregación en la métrica nDGC.....	83
Figura 4.3: Impacto del tamaño de grupo en la métrica F1-Score.....	85

Figura 4.4: Impacto del tamaño de grupo en la métrica nDGC.....	85
Figura 4.5: Impacto de la longitud de lista de recomendación en la métrica F1-Score.....	87
Figura 4.6: Impacto de la longitud de lista de recomendación en la métrica nDGC.....	87
Figura 4.7: Distribución de los valores de similitud entre los usuarios en el conjunto de datos de MovieLens 100k.....	90
Figura 4.8: Impacto de la similitud intergrupala mínima en la métrica F1-Score.....	90
Figura 4.9: Impacto de la similitud intergrupala mínima en la métrica nDGC.....	91
Figura 4.10: Relación entre la similitud promedio y las métricas de evaluación en sistemas de recomendación grupal: Análisis a través de diagramas de dispersión.....	93
Figura 4.11: Número de las reglas generadas para cada estrategia de filtrado e imputación.....	96
Figura 4.12: Número de elementos cubiertos por las reglas de asociación generadas....	96
Figura 4.13: Distribución de los valores de soporte y confianza en las reglas de asociación generadas.....	97
Figura 4.14: Impacto del método de generación de reglas de asociación en la métrica de Fidelidad.....	98
Figura 4.15: Impacto del algoritmo de imputación en la métrica de Fidelidad.....	101
Figura 4.16: Impacto del tamaño de grupo en la métrica de Fidelidad.....	102
Figura 4.17: Impacto del umbral mínimo de soporte en la métrica de Fidelidad.....	105
Figura 4.18: Impacto del umbral mínimo de confianza en la métrica de Fidelidad.....	107
Figura 4.19: Impacto de la similitud intergrupala mínima en la métrica de Fidelidad.....	108

Índice de Tablas

Tabla 2.1: Ejemplo de calificaciones de cuatro usuarios para cinco películas.....	35
Tabla 3.1: Ejemplo de conjunto de entrenamiento.....	60
Tabla 3.2: Ejemplo de transacciones generadas con estrategia de extracción de los elementos relevantes.....	60
Tabla 3.3: Ejemplo de la matriz de calificaciones con predicciones de las calificaciones faltantes.....	61
Tabla 3.4: Ejemplo de transacciones generadas con estrategia de imputación de las calificaciones faltantes.....	61
Tabla 4.1: Matriz de confusión para sistemas de recomendación.....	76
Tabla 4.2: Matriz de confusión extendida.....	77

Resumen

Este trabajo de fin de máster se centra en el desarrollo de métodos de explicabilidad para sistemas de recomendación grupal. El objetivo principal es proporcionar explicaciones claras y comprensibles sobre las recomendaciones generadas en un contexto grupal, donde múltiples usuarios están involucrados y se consideran las preferencias y restricciones del grupo. El enfoque metodológico se basa en una revisión de la literatura existente sobre técnicas de recomendación y explicabilidad en sistemas de recomendación. A partir de esta revisión, se diseñarán y desarrollarán métodos de explicabilidad adaptados para sistemas de recomendación grupal. Además, se llevarán a cabo experimentos y análisis comparativos para evaluar la efectividad y calidad de los métodos de explicabilidad desarrollados. Los resultados obtenidos permitirán determinar la eficacia de los métodos propuestos y su capacidad para generar explicaciones comprensibles y relevantes para los usuarios del sistema.

Abstract

This master's thesis addresses the challenge of providing explainability in group recommendation systems. The main objective is to deliver clear and comprehensible explanations for recommendations generated within a group context, while considering the preferences and constraints of multiple users. To achieve this goal, an extensive literature review is conducted to identify relevant approaches and methodologies in recommendation techniques and explainability. Building upon this review, tailored explainability methods specifically adapted for group recommendation systems are designed and implemented. Furthermore, experiments and comparative analyses are performed to evaluate the effectiveness and quality of the developed explainability methods. The results contribute to enhancing the understandability and trustworthiness of recommendations in group settings, fostering user satisfaction and engagement.

Keywords: Recommendation Systems, Group Recommendation, Artificial Intelligence, Machine Learning, Explainable Artificial Intelligence, Explainability, Transparency, User Understanding, Post-Hoc.

INTRODUCCIÓN

CAPÍTULO 1

1. Introducción

Durante las últimas décadas, ha proliferado el crecimiento de servicios online tales como, Youtube (www.youtube.com), Amazon (www.amazon.com), Netflix (www.netflix.com), etc., donde los usuarios tienen acceso a una cantidad inmensa de información y productos. El fenómeno de la sobrecarga de información ha hecho que los sistemas de recomendación (RS) hayan surgido para ayudar a los usuarios de estos servicios a superar el problema de sobrecarga de información y puedan encontrar de forma simple, rápida y eficaz aquellos productos y/o servicios que buscan o necesitan (Milano et al., 2020). Así en el comercio electrónico (sugerir a los compradores artículos que podrían interesarles) o en la publicidad online (sugerir a los usuarios los contenidos correctos, de acuerdo con sus preferencias), son cada vez más habituales en nuestra navegación diaria online.

En efecto, la mayoría de RSs implementados en la actualidad utilizan algoritmos de "caja negra" (*Black-Box*), lo que significa que su funcionamiento interno no es completamente transparente y no se puede entender fácilmente cómo se generan las recomendaciones (Herlocker et al., 2000). Estos algoritmos suelen basarse en técnicas de aprendizaje automático, como filtrado colaborativo (Sarwar et al., 2001), filtrado basado en contenido (Adomavicius & Tuzhilin, 2005) o enfoques híbridos que combinan varias técnicas (Stern et al., 2009).

La falta de interpretabilidad en los algoritmos de los sistemas de recomendación plantea preocupaciones en términos de privacidad y confianza. Los usuarios pueden tener dificultades para entender por qué se les hacen ciertas recomendaciones y pueden tener dudas sobre cómo se utilizan sus preferencias. Además, la falta de transparencia en estos algoritmos puede resultar en la propagación de sesgos no deseados, ya que los usuarios pueden recibir recomendaciones limitadas o sesgadas basadas en su historial de navegación o preferencias pasadas. Es fundamental abordar estos desafíos y desarrollar sistemas de recomendación más transparentes y explicables que promuevan la confianza y la equidad en el proceso de recomendación (Herlocker et al., 2000).

Además, otro desafío importante en los sistemas de recomendación es la recomendación a grupos en lugar de usuarios individuales. Esto se presenta en situaciones como el consumo de contenido en familia o la planificación de viajes en grupo.

La recomendación a grupos es aún más compleja (Castro, Yera, & Martínez, 2017), ya que se debe considerar cómo combinar los modelos de usuarios individuales para generar recomendaciones grupales efectivas. Es esencial abordar los desafíos específicos asociados con la recomendación a grupos y desarrollar estrategias adecuadas para servicios de streaming, páginas web de eventos y agencias de viajes, entre otros (Castro, Yera, & Martínez, 2017). En este capítulo se expone la motivación para el desarrollo de un sistema de recomendación grupal (GRS) y se detallan los objetivos específicos para llevar a cabo su desarrollo.

2. Motivación

En la actualidad, los sistemas de recomendación (RS) desempeñan un papel fundamental en diversos contextos, como el comercio electrónico, la publicidad en línea, las redes sociales y el entretenimiento digital. Estos sistemas utilizan algoritmos y técnicas de aprendizaje automático para analizar los datos de los usuarios y los productos, con el objetivo de generar recomendaciones personalizadas. Sin embargo, a medida que el consumo de contenido se vuelve más social y colaborativo, surge la necesidad de desarrollar sistemas de recomendación grupal (GRS) que puedan brindar recomendaciones adecuadas para grupos de usuarios en lugar de individuos (Castro, Yera, & Martínez, 2017).

Los GRS son especialmente relevantes en situaciones en las que las decisiones de consumo de contenido se toman en grupo, como la planificación de actividades sociales, la elección de programas de televisión o películas para ver en familia, o la selección de restaurantes o destinos de viaje con amigos (Castro, Yera, & Martínez, 2017). En estos casos, es crucial considerar las preferencias y necesidades de todos los miembros del grupo para llegar a una decisión conjunta. Sin embargo, el desarrollo de GRS plantea desafíos únicos, ya que implica la combinación de modelos de usuarios individuales en un modelo grupal (De Pessemier et al., 2014), lo que puede ser complejo debido a las diferencias y posibles conflictos entre las preferencias de los miembros del grupo.

Aunque la importancia de los GRS es evidente, la investigación en este campo aún es limitada. Pocos estudios se han centrado específicamente en la explicabilidad en sistemas de recomendación grupal. Esto destaca la necesidad de abordar esta brecha y explorar en mayor medida la motivación y los beneficios de la explicabilidad en los GRS.

Por tanto, el objetivo principal de este proyecto de investigación es analizar y proponer un modelo de explicabilidad para sistemas de recomendación grupal. La falta de investigaciones en este campo resalta la necesidad de explorar y desarrollar métodos y técnicas que permitan mejorar la comprensión de las recomendaciones generadas en un contexto grupal. Al hacerlo, se busca abordar las preocupaciones de transparencia y confianza de los usuarios al entender cómo se generan las recomendaciones y proporcionar justificaciones claras y comprensibles.

3. Propósito del proyecto

El propósito de este proyecto es desarrollar un modelo de recomendación grupal transparente y explicativo. Se busca superar la opacidad de los sistemas de recomendación convencionales, donde los usuarios desconocen cómo se generan las recomendaciones. En este proyecto, se prioriza la explicabilidad como característica clave del sistema de recomendación grupal. Se investigarán y aplicarán técnicas que permitan a los usuarios comprender de forma clara y comprensible el proceso de generación de recomendaciones grupales.

4. Objetivos

Con el fin de lograr el propósito mencionado anteriormente, este Trabajo de Fin de Máster (TFM) busca alcanzar los siguientes objetivos específicos:

- ❖ Revisar la literatura existente sobre sistemas de recomendación, con el fin de comprender a fondo los diferentes aspectos y enfoques utilizados en este campo;
- ❖ Plantear y discutir estrategias de agregación de los modelos de recomendación de usuarios individuales para generar una recomendación grupal. Se investigarán y se analizarán los enfoques existentes para combinar las preferencias y opiniones de los miembros del grupo;
- ❖ Discutir y seleccionar métodos de explicabilidad que puedan ser implementados en el sistema de recomendación grupal desarrollado;
- ❖ Diseñar y desarrollar el modelo de recomendación grupal, incorporando las estrategias de agregación y el método de explicabilidad seleccionado.
- ❖ Utilizar métricas de evaluación de rendimiento y de explicabilidad, con el fin de medir la efectividad y la calidad del sistema de recomendación grupal

desarrollado. Se compararán y analizarán los resultados obtenidos en diferentes experimentos para evaluar el desempeño del sistema y su capacidad para generar recomendaciones explicables y relevantes.

- ❖ Planificar y realizar experimentos para analizar la influencia de los diferentes parámetros en el rendimiento del sistema de recomendación grupal. Se llevarán a cabo pruebas y análisis de sensibilidad para entender cómo variaciones en los parámetros del sistema pueden afectar las recomendaciones generadas y la explicabilidad del sistema.

5. Metodología y planificación del proyecto

En esta sección, se abordará la metodología utilizada y la planificación del proyecto en el desarrollo de este Trabajo de Fin de Máster.

5.1. Metodología

En el marco de este proyecto, se optó por utilizar una metodología incremental como estrategia de gestión del proyecto. La metodología incremental se basa en la idea de dividir el proyecto en etapas pequeñas y manejables, donde se van desarrollando funcionalidades de manera progresiva (Berggren, 2019).

La elección de esta metodología se fundamentó en su adecuación para proyectos que requieren una adaptación flexible a medida que se avanza en el desarrollo. Al desarrollar métodos de explicabilidad para sistemas de recomendación grupal, era necesario iterar y mejorar continuamente los resultados obtenidos. La metodología incremental se adaptó perfectamente a esta necesidad, permitiendo una evolución progresiva y constante de los métodos propuestos. En la figura 1.1 se puede ver el ciclo de vida de esta metodología.

Las características fundamentales de la metodología incremental que nos llevaron a seleccionarla son las siguientes (Berggren, 2019):

- ❖ **Iterativo y evolutivo:** La metodología incremental se basa en ciclos iterativos donde se desarrollan funcionalidades incrementales en cada iteración. Esto permite una mejora continua del sistema a medida que se avanza en el proyecto.

- ❖ **Flexibilidad y adaptabilidad:** Al dividir el proyecto en etapas más pequeñas, se facilita la incorporación de cambios y ajustes en cada iteración. Esto permite una mayor flexibilidad para responder a los requisitos cambiantes y las necesidades del sistema.
- ❖ **Retroalimentación constante:** La metodología incremental fomenta una comunicación continua con los miembros del equipo y las partes interesadas. Se busca obtener retroalimentación y comentarios en cada iteración, lo que permite una mayor alineación con los objetivos del proyecto.
- ❖ **Entregas parciales y valor temprano:** Mediante el enfoque incremental, se logra generar entregas parciales y funcionales en cada iteración. Esto permite obtener valor temprano y validar de manera progresiva la efectividad de los métodos de explicabilidad desarrollados.

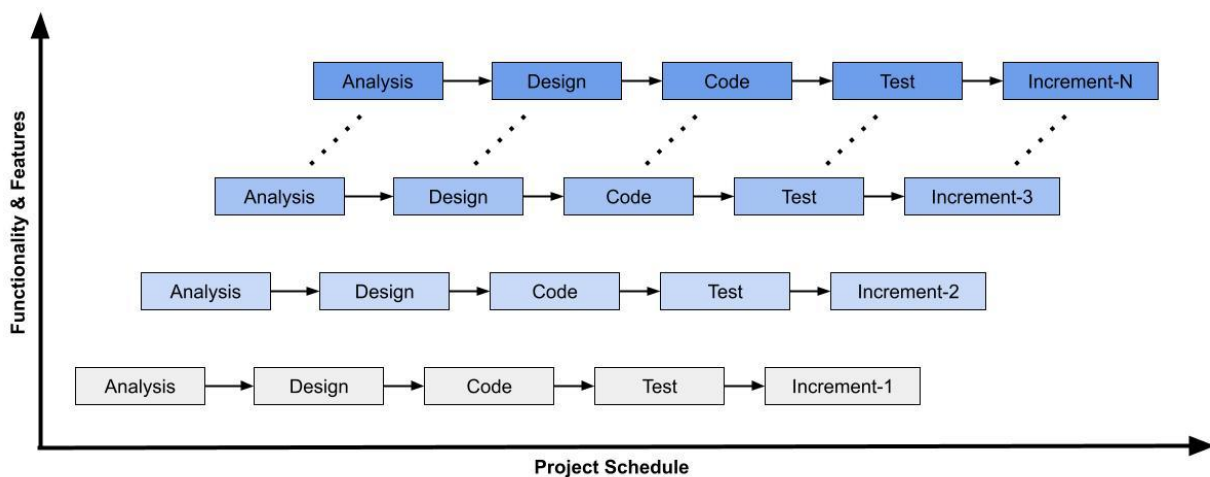


Figura 1.1: Ciclo de vida de metodología incremental

En el caso específico de este Trabajo de Fin de Máster, la metodología incremental se adaptó de manera óptima a las necesidades. El objetivo final de esta investigación consistía en desarrollar métodos de explicabilidad para sistemas de recomendación grupal, lo cual requería un enfoque flexible y adaptable a medida que se avanzaba en el desarrollo. La metodología incremental permitió una evolución continua de los métodos, con entregas parciales y mejoras iterativas en cada fase del proyecto.

En el próximo apartado, se proporcionará un análisis detallado de la estructura de desglose de trabajo, donde se presentarán todas las tareas definidas para el periodo de tiempo establecido, junto con su correspondiente planificación.

5.2. Planificación

En esta sección, se presentará la Estructura de Desglose de Trabajo (EDT) del proyecto, que representa visualmente la descomposición jerárquica de las tareas. Cada tarea será desglosada en niveles y subniveles, lo que permitirá una comprensión clara de las actividades requeridas para alcanzar los objetivos del proyecto (Devi & Reddy, 2012). Se brindará una descripción detallada de cada tarea.

En la figura 1.2 se muestra un esquema que representa la EDT establecida para este proyecto. Esta representación visualiza las principales tareas que determinan el trabajo a realizar, dividiéndolo en paquetes y actividades.

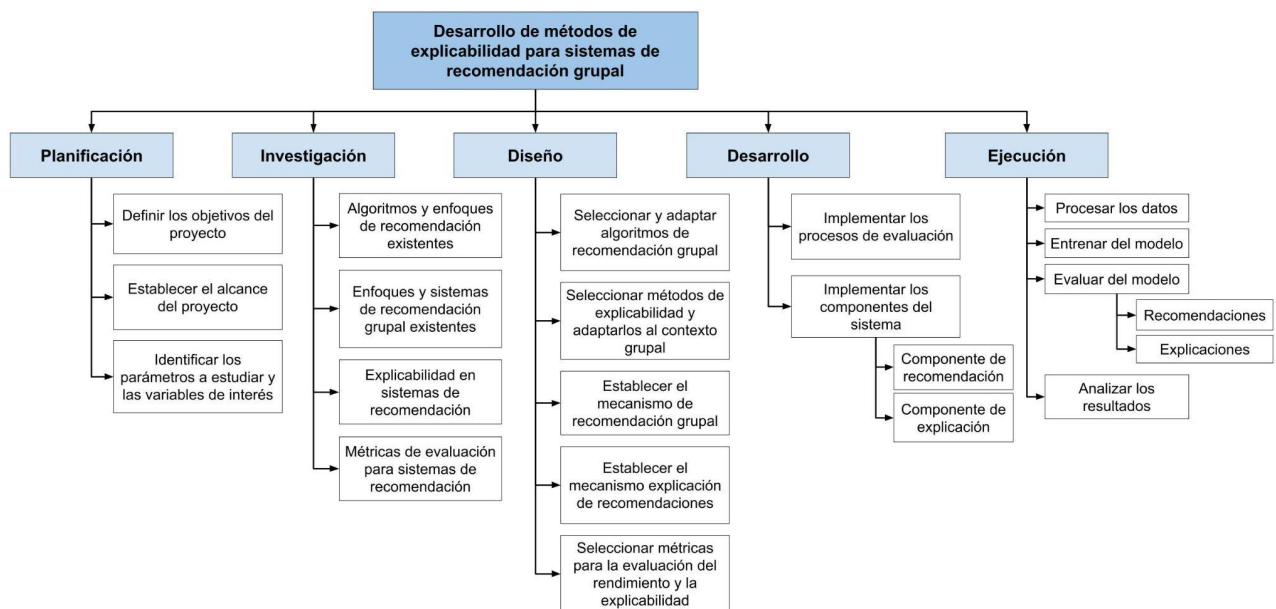


Figura 1.2: Estructura de Desglose de Trabajo

La realización de este proyecto implica una serie de etapas clave que garantizan un enfoque sistemático y efectivo. Comienza con la planificación, donde se definen los objetivos del proyecto y se establece el alcance del trabajo. Esto incluye identificar los parámetros a estudiar y las variables a evaluar, lo que proporciona una base sólida para el desarrollo.

La etapa de investigación es fundamental para comprender el panorama actual. Se investigan los algoritmos y enfoques de recomendación existentes, así como los sistemas de recomendación grupal ya implementados. Además, se estudian los métodos de

explicabilidad utilizados en estos sistemas y se analizan las métricas de evaluación para medir el rendimiento y la explicabilidad.

El diseño es una fase crucial donde se toman decisiones estratégicas. Se seleccionan y adaptan los algoritmos de recomendación grupal más adecuados para el proyecto, teniendo en cuenta las necesidades específicas. Asimismo, se eligen los métodos de explicabilidad que se ajusten al contexto grupal y se establecen los mecanismos para las recomendaciones grupales y las explicaciones asociadas. Además, se seleccionan las métricas adecuadas para evaluar la calidad y el rendimiento del sistema.

El desarrollo se centra en la implementación práctica de los componentes del sistema. Se lleva a cabo la implementación de los procesos de evaluación y se crean los componentes de recomendación y explicación. Estos componentes son fundamentales para generar recomendaciones precisas y comprensibles, y para proporcionar explicaciones claras y útiles que satisfagan las necesidades del grupo.

La ejecución implica el procesamiento de los datos, el entrenamiento del modelo de recomendación grupal y la evaluación del rendimiento y la explicabilidad del sistema. Se analizan los resultados obtenidos en términos de recomendaciones y explicaciones generadas. Esto permite obtener información valiosa sobre la efectividad del sistema y posibles áreas de mejora.

6. Estructura del TFM

El Trabajo de Fin de Máster se ha llevado a cabo siguiendo una metodología investigadora compuesta por varias fases. En primer lugar, en el Capítulo 1, se definió la introducción del trabajo y su motivación, se estableció el propósito del proyecto y se describieron los objetivos a alcanzar. Además, se explicó la metodología y planificación seguida.

En el Capítulo 2 se realizó un estudio detallado sobre los sistemas de recomendación, describiendo las diferentes técnicas existentes y sus características. Se presentaron técnicas como el filtrado colaborativo, basado en contenido e híbridos, y se introdujo la novedosa técnica de sistemas de recomendación grupal, además de hablar sobre la importancia de la explicabilidad en los sistemas de recomendación.

Posteriormente, en el Capítulo 3 se describió el modelo de explicabilidad propuesto, el cual está compuesto por dos componentes principales: el componente de recomendación y el componente de explicación. Se presentó la arquitectura del modelo y se explicó su funcionamiento.

En el Capítulo 4 se detallaron los aspectos de la evaluación experimental realizada, incluyendo la descripción del dataset utilizado, los métodos de generación de grupos y las métricas de evaluación. Además, se presentaron los diferentes experimentos realizados para evaluar el modelo de explicabilidad propuesto. Luego, se presentaron los resultados obtenidos de la evaluación experimental, analizando los resultados y presentando los hallazgos más relevantes.

Finalmente, en el Capítulo 5 se presentaron las conclusiones del trabajo realizado, se discutieron las limitaciones y se propusieron líneas de trabajo futuro para mejorar el modelo de explicabilidad desarrollado y para explorar nuevas áreas de investigación en el ámbito de los sistemas de recomendación grupal.

7. Conclusión

En este capítulo se aborda la importancia de los sistemas de recomendación (RS) en el contexto actual de sobrecarga de información en servicios en línea. Se destaca la falta de transparencia y explicabilidad en los algoritmos de los RS, lo que plantea preocupaciones en términos de privacidad y confianza. Además, se introduce la necesidad de desarrollar sistemas de recomendación grupal (GRS) para situaciones en las que las decisiones de consumo se toman en grupo. El objetivo principal de este proyecto de investigación es analizar y proponer un modelo de explicabilidad para los GRS, buscando mejorar la comprensión de las recomendaciones generadas en un contexto grupal. Se establecen los objetivos específicos y se describe la metodología incremental utilizada, junto con la planificación del proyecto. En futuros capítulos, se abordará la revisión de la literatura, las estrategias de agregación de modelos individuales, los métodos de explicabilidad y el diseño e implementación del sistema de recomendación grupal.

SISTEMAS DE RECOMENDACIÓN

CAPÍTULO 2

1. Introducción

Un sistema de recomendación (RS) es una subclase de los sistemas de filtrado de información que busca proporcionar sugerencias de elementos que pueden ser útiles para un usuario (Resnick et al., 1997; Burke et al., 2007; Mahmood & Ricci., 2009). Estas sugerencias se relacionan con varios procesos de toma de decisiones, pueden ser como qué producto comprar (Amazon: www.amazon.com), qué música escuchar (Spotify: open.spotify.com) o qué Película ver (Netflix: www.netflix.com), entre otros.

El estudio de los RSs se considera relativamente nuevo en comparación con los estudios de otras técnicas clásicas de sistemas de información como por ejemplo las bases de datos, motores de búsqueda, etc. Los RSs surgieron como una área de investigación independiente a mediados de la década de 1990 (Goldberg et al. 1992; Mahmood & Ricci., 2009; Troiano et al., 2014).

En los últimos años, el interés en los RSs ha aumentado drásticamente. Ha habido varios números especiales en revistas académicas que cubren investigaciones y desarrollos en el campo de RS (Fig. 2.1).

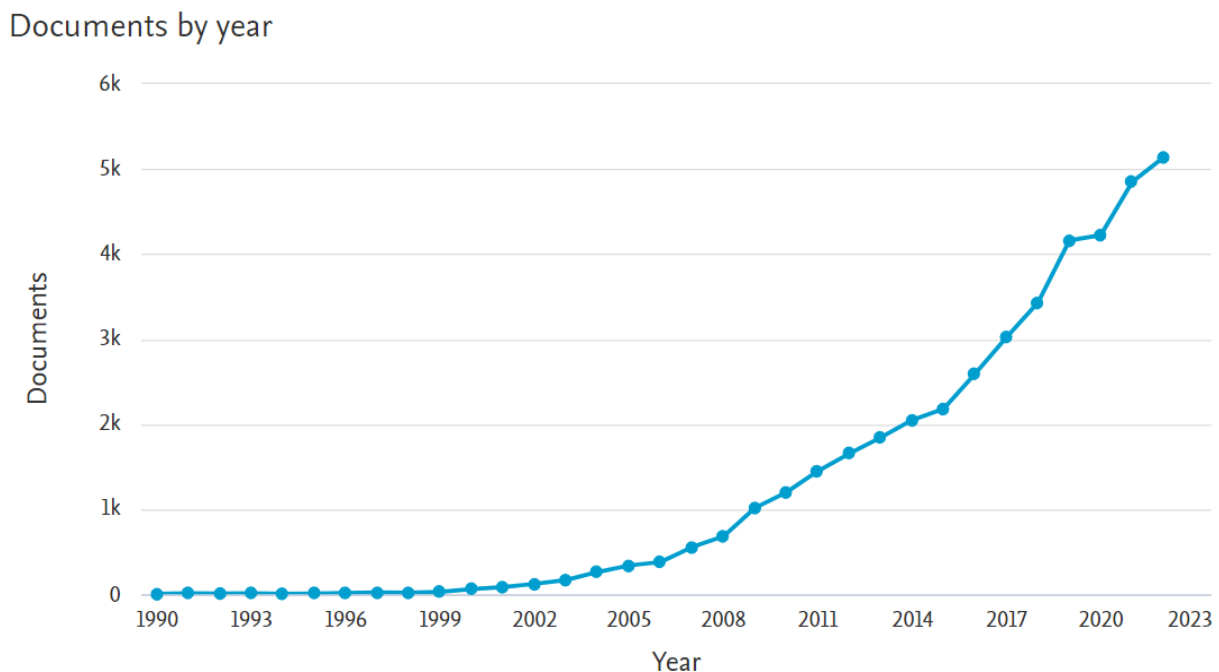


Figura 2.1: Número de documentos publicados sobre los sistemas de recomendación en la base de datos Scopus

Entre las revistas que tienen números dedicados a RS encontramos: *ACM Transactions on Intelligent Systems and Technologies* (2023), *User Modeling and User-Adapted Interaction* (2022), *Information Processing and Management* (2022), *Applied Soft Computing* (2020), *Future Generation Computer Systems* (2018), entre otras.

Hay varias razones por las cuales se ha despertado un gran interés en esta tecnología y por las cuales los proveedores de servicios pueden estar interesados en aprovecharla:

- ❖ **Aumentar el número de artículos vendidos:** Probablemente, es la función más importante para un RS comercial, que es, poder vender una cantidad adicional de productos en comparación con la cantidad que se vende sin ningún tipo de recomendación. Se puede lograr este objetivo mediante la recomendación de elementos que se adapten a las necesidades y deseos de los usuarios. Las aplicaciones no comerciales tienen también objetivos similares, aunque a veces no hay ningún costo para el usuario asociado con la selección de un elemento. Por ejemplo, un proveedor de una plataforma de noticias tiene como objetivo aumentar el número de noticias leídas. En general, podemos decir que el objetivo del proveedor por su adopción de un RS es aumentar la tasa de conversión, es decir, la cantidad de usuarios que aceptan la recomendación y consumen un artículo.
- ❖ **Vender artículos más diversos:** Otra función muy importante de un RS es ayudar al usuario a seleccionar elementos que pueden ser difíciles de encontrar sin una recomendación precisa. Por ejemplo, en una RS de películas como Netflix, el proveedor de servicios está interesado en alquilar todas las películas del catálogo, no sólo las más populares. Esto podría ser difícil, pero con un RS que sugiere o anuncia películas menos populares a los usuarios correctos, podría ser alcanzable.
- ❖ **Aumentar la satisfacción del usuario:** Un RS bien diseñado también puede mejorar la experiencia del usuario con la aplicación. El usuario encontrará recomendaciones interesantes y relevantes que le sugieran artículos que se ajusten a sus necesidades y deseos. La combinación de recomendaciones precisas, y una interfaz utilizable aumentará la evaluación subjetiva del sistema por parte del usuario. Esto, a su vez, aumentará el uso del sistema y la probabilidad de que se acepten las recomendaciones.

- ❖ **Aumentar la fidelidad de los usuarios:** Un usuario debe ser leal a una aplicación que, cuando se visita, reconoce al antiguo cliente y lo trata como un visitante valioso. Esta es una característica normal de un RS ya que muchos RS calculan recomendaciones, aprovechando la información adquirida del usuario en interacciones previas, por ejemplo, sus calificaciones de elementos. En consecuencia, cuanto más interactúa el usuario con el sitio, más refinado se vuelve su modelo de usuario (su representación por el sistema), y más se puede personalizar la recomendación para que coincida con sus preferencias.
- ❖ **Entender mejor lo que quiere el usuario:** Otra función importante de un RS, que se puede aprovechar para muchas otras aplicaciones, es la descripción de las preferencias del usuario. El proveedor de servicios puede entonces decidir reutilizar este conocimiento para una serie de otros objetivos, como mejorar la gestión de las existencias o la producción del artículo.

Estos puntos indican que un sistema de recomendación (RS) dentro de un sistema de información puede desempeñar diversas funciones y objetivos, lo que implica la utilización de una amplia gama de fuentes y técnicas de recomendación. En los siguientes apartados, se discutirán los datos que maneja un RS y las diferentes técnicas utilizadas para generar recomendaciones precisas y correctas.

2. Técnicas de recomendación

Para elaborar una recomendación, los RS recopilan activamente varios tipos de datos que son principalmente sobre los elementos que se sugieren y los usuarios que reciben estas recomendaciones (Milano et al., 2020). Dado que las fuentes de datos disponibles para los RSs pueden ser muy diversas, en última instancia, si se van a explotar o no, depende de la técnica de recomendación utilizada.

Existen dos tipos de técnicas para la recomendación de contenido. En primer lugar, hay técnicas que dependen en gran medida de la información (Apartado 2.1), ya que utilizan datos detallados tanto de los usuarios (por ejemplo, sus relaciones sociales y actividades, edad, etc.) como de los artículos (descripciones asociadas, categorías, etc.). Por otro lado, hay técnicas que requieren de menor información (Apartado 2.2), ya que utilizan datos más simples y básicos, como las calificaciones que los usuarios han otorgado a los artículos.

En cualquier caso, como clasificación general, los datos utilizados por los RSs se refieren a tres tipos de objetos: artículos (*Items*), usuarios (*Users*) y transacciones (*Transactions*), es decir, relaciones entre usuarios y artículos.

- ❖ **Item (Artículo):** Un ítem, es el objeto recomendado. Un ítem puede ser caracterizado por su utilidad, su valor y/o su complejidad. Notamos que cuando un usuario adquiere un ítem, siempre incurrirá en un costo, que incluye el costo cognitivo de buscar el ítem y el costo monetario real que pagará por el ítem. Por lo tanto, a la hora de desarrollar un RS hay que tener en cuenta tanto la complejidad del ítem (representación textual, estructura, etc) como el costo, (sea monetario o cognitivo), y sobre todo, cuando hablamos de seleccionar ítems que se han considerado muy complejos tales como pólizas de seguro, inversiones financieras, coches, viajes, trabajos, etc (Montaner et al. 2004).
- ❖ **User (Usuario):** Los usuarios de un RS, pueden tener varios objetivos y características. Para generar una recomendación precisa y correcta, el RS explota el conjunto de los datos que tiene sobre los usuarios. La selección de los datos que se van a utilizar se basa en con qué técnica de recomendación estemos utilizando. Por ejemplo, en el enfoque de filtrado colaborativo (Apartado 2.2) el usuario está modelado por sus evaluaciones para los ítems. En un RS demográfico, el usuario está modelado por su género, edad, profesión, educación, etc. Así que con los datos que tiene el RS sobre el usuario puede construir su perfil (Pazzani et al., 1997; Fisher et al., 2001), es decir, codificar sus preferencias, necesidades, etc. Con dicho perfil, el RS puede generarle una recomendación precisa y correcta que se ajusta con sus deseos.
- ❖ **Transaction (Transacción):** En general, una transacción, refiere a la interacción entre el usuario y un ítem determinado. Estas transacciones se almacenan durante la interacción del usuario con el RS, y se utilizan luego para generar la recomendación. Una transacción almacenada puede contener la referencia del ítem seleccionado y una descripción de la transacción, a veces se puede incluir una calificación del usuario para el ítem seleccionado.

En general, la calificación del usuario (*Rating*) es la forma más popular de datos de transacciones que recopila un RS. Estas calificaciones pueden recopilarse explícita o implícitamente. En la recopilación explícita de calificaciones, se le pide al usuario

que dé su opinión sobre un elemento en una escala de calificación. De acuerdo con (Schafe et al., 2007), las calificaciones pueden tomar una variedad de formas:

- **Calificaciones numéricas**, como las estrellas de 1 a 5 proporcionadas en el recomendador de libros asociado con Amazon.
- **Calificaciones ordinales**, como “muy de acuerdo, de acuerdo, neutral, en desacuerdo, muy en desacuerdo”, donde se le pide al usuario que seleccione el término que mejor indica su opinión sobre un ítem.
- **Calificaciones binarias**, que modelan opciones en las que simplemente se le pide al usuario que decida si un elemento determinado es bueno o malo.
- **Calificaciones unarias**, pueden indicar que un usuario ha observado o comprado un artículo, o que calificó el artículo de otra manera positivamente.

Brevemente, para llevar a cabo la personalización de las recomendaciones, los RS explotan la información que tienen sobre los usuarios y los ítems. Esta información se puede estructurar de varias formas y la selección de la información a modelar depende de la técnica de recomendación utilizada. A continuación, veremos en detalle algunas de las técnicas y algoritmos de recomendación más populares.

2.1. Basado en Contenido

Los RSs basados en contenido (CB-RS) se han convertido en un paradigma relevante en el desarrollo de RSs (Bobadilla et al., 2013; Yera & Martínez, 2017). El principio general de este enfoque es identificar las características comunes de los ítems que han recibido una calificación favorable de un usuario u , y luego recomendarle nuevos elementos que comparten las mismas características (Billsus et al., 1998; Billsus et al., 2004; Lang, 1995).

En los RSs basados en contenido, se supone que la información del ítem i están codificadas mediante un vector de características P_i . Para los ítems en forma de documentos de texto como los artículos de noticias (Billsus et al., 2004; Lang, 1995), este vector a menudo contiene los pesos de frecuencia inversa de término-frecuencia de documento (TFIDF) (Salton, 1988) de las palabras clave más informativas. También se han propuesto varios enfoques para la representación de estas características, incluida la creación de perfiles binarios para representar directamente la presencia de una

característica (Adomavicius & Tuzhilin, 2005), el uso de análisis de semántica latente para representar elementos textuales (Castro et al., 2019), el uso de perfiles basados en etiquetas (Gedikli & Jannach, 2013) y otras estrategias.

Además, para cada usuario u , normalmente se representa con un vector P_u en el que se codifican sus preferencias a partir del contenido de los ítems I_u con los cuales ha tenido transacciones (ítems vistos, comprados, leídos, etc). Una técnica para calcular estos perfiles, utilizada en varios RSs basados en contenido, es el algoritmo de Rocchio (Zitnick & Kanade, 2004; Buckley et al., 1995). Esta técnica actualiza el perfil P_u del usuario u cada vez que este usuario califica un ítem i sumando los pesos de P_i a P_u , en proporción a la apreciación de u por i (Ec. 1):

$$P_u = \sum_{i \in I_u} r_{ui} \times P_i \quad (1)$$

Donde

P_u	:	Perfil del usuario u
P_i	:	Perfil del ítem i
r_{ui}	:	Calificación del usuario u para el ítem i
I_u	:	Set de ítems calificados por el usuario u

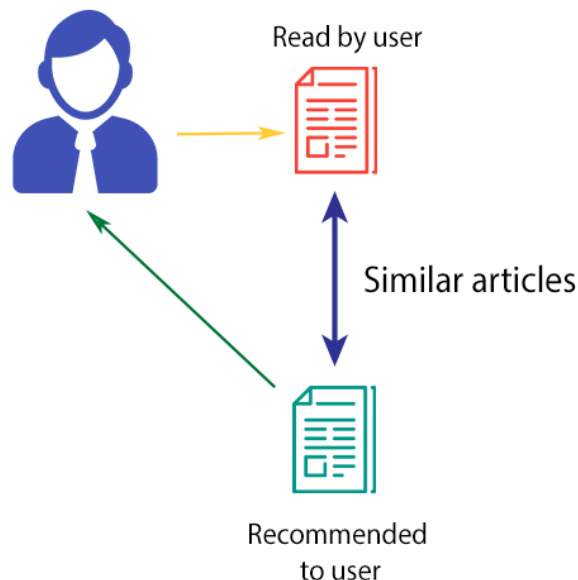


Figura 2.2: Sistema de recomendación basado en Contenido

Al tener listos los perfiles de los ítems y el perfil del usuario, se elige una métrica de similitud con la que se calcula la similitud entre el perfil del usuario y cada uno de los

perfiles de los ítems. Por fin, recomendar aquellos ítems más similares (Fig. 2.2). Considerando ambos perfiles P_u y P_i , la similitud $s_{u,i}$ entre el ítem i y el usuario (Ec. 2):

$$s_{u,i} = \text{Score}(P_u, P_i) \quad (2)$$

Donde

u	:	Usuario
i	:	Item
P_u	:	Perfil del usuario u
P_i	:	Perfil del ítem i
$s_{u,i}$:	Similitud entre el usuario u y el ítem i

En los RSs basados en contenido con una arquitectura de alto nivel se realiza el proceso de recomendación en tres pasos principales, cada uno de los cuales es manejado por un componente separado (Fig. 2.3) (Adomavicius & Tuzhilin, 2005):

- ❖ **CONTENT ANALYZER:** La responsabilidad principal del componente es representar el contenido de los ítems (por ejemplo, documentos, páginas web, etc.) provenientes de fuentes de información en una forma adecuada para los siguientes pasos de procesamiento. Los datos de los ítems se analizan mediante técnicas de extracción de características para cambiar su representación. Esta representación es la entrada del *PROFILE LEARNER* y al *FILTERING COMPONENT*;
- ❖ **PROFILE LEARNER:** Este componente recopila datos representativos de las preferencias del usuario e intenta construir su perfil a partir de estos datos. Por lo general, la estrategia de generalización se realiza a través de técnicas de aprendizaje automático (Mitchell, 1997), que pueden inferir un modelo de intereses del usuario a partir de elementos que les gustaron o no en el pasado;
- ❖ **FILTERING COMPONENT:** Este módulo explota el perfil de usuario para sugerir ítems relevantes comparando la representación del usuario con la de los ítems. El resultado es un juicio de relevancia binario o continuo, este último caso resulta en una lista clasificada de elementos potencialmente interesantes. Como ya hemos mencionado, la coincidencia se realiza calculando la similitud entre el vector prototipo y los vectores de elementos.

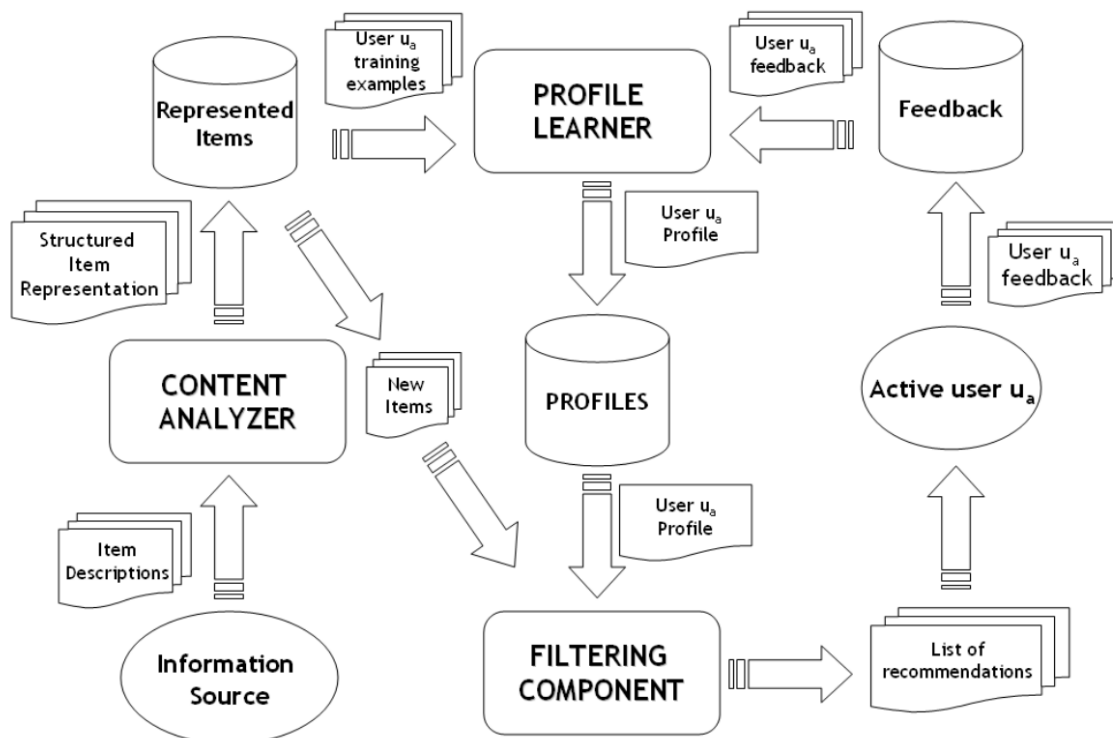


Figura 2.3: Arquitectura de alto nivel de un sistema de recomendación basado en contenido

Los RSs basados en el contenido se pueden adaptar en gran medida a los intereses del usuario, porque el método se basa en hacer coincidir las características o del perfil de los ítems con el perfil del usuario. El filtrado basado en contenido también es valioso para las empresas con bases de datos extensas que contienen un solo tipo de producto, como los teléfonos inteligentes, donde las recomendaciones deben basarse en muchas características discretas.

2.2. Filtrado Colaborativo

El filtrado colaborativo (CF-RS) se basa en la idea fundamental de hacer predicciones de calificaciones utilizando usuarios o ítems con características similares (Fig. 2.4) (Sarwar et al., 2001). La implementación más simple y original de este enfoque (Schafer et al., 2007) recomienda al usuario activo los ítems que han sido apreciados en el pasado por otros usuarios con gustos similares. La similitud en los gustos entre dos usuarios se calcula en función de sus historiales de calificaciones. Por esta razón, (Schafer et al., 2001) se refiere al filtrado colaborativo como "correlación de usuario a usuario". Esta técnica se considera la más popular y ampliamente implementada en los sistemas de recomendación (RS).

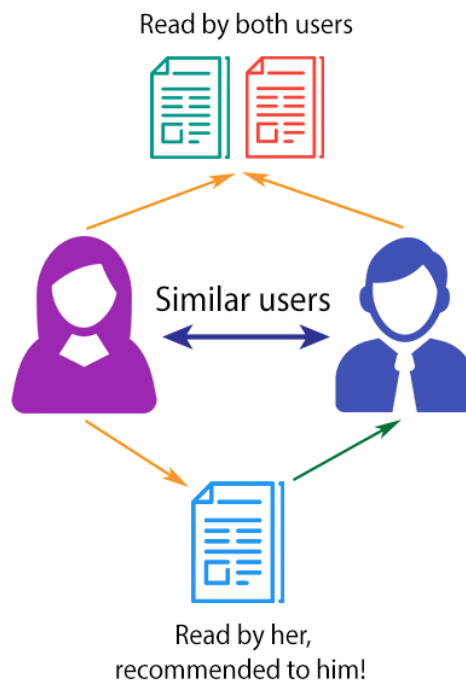


Figura 2.4: Sistema de recomendación basado en Filtrado Colaborativo

El enfoque de CF-RS supera algunas de las limitaciones de CB-RS. Especialmente en el caso de que el contenido de los artículos no esté disponible o sea difícil de obtener. Además, un CF-RS brinda una recomendación basada en la calidad de los elementos evaluados por pares, en lugar de depender del contenido que puede ser un mal indicador de calidad. Finalmente, a diferencia de CB-RS, los CF-RSs pueden recomendar elementos con contenido muy diferente, siempre que otros usuarios ya hayan mostrado interés por estos elementos diferentes.

Los métodos de CF-RS se pueden agrupar en las dos clases generales de métodos, basados en memoria y modelos.

2.2.1. Basado en Memoria

En el filtrado colaborativo basado en memoria o también conocido por basado en vecindario (John et al., 1998), las calificaciones de los usuarios son almacenadas directamente en el sistema y se utilizan para predecir las calificaciones de los elementos nuevos. Esto se puede hacer de dos maneras:

- ❖ **Basado en Usuario (UCF):** Los sistemas basados en usuarios evalúan el interés de un usuario u por un elemento i utilizando las calificaciones de este ítem por parte de otros usuarios, llamados vecinos, que tienen patrones de calificación similares. Los

vecinos del usuario u suelen ser aquellos usuarios v que tienen algunas calificaciones de ítems en común con el usuario u . Las predicciones se calculan como (Ec. 3) (John et al., 1998).

- ❖ **Basado en Ítem (ICF):** Los enfoques Basados en Ítems, por otro lado, predicen la calificación de un usuario u para un elemento i en función de las calificaciones de u para elementos similares a i . En tales enfoques, dos elementos son similares si varios usuarios del sistema han calificado estos ítems de manera similar. La predicción de la calificación que puede dar el usuario u se puede calcular usando un promedio ponderado simple como (Eq. 4) (John et al., 1998).

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in K_u} (r_{v,i} - \bar{r}_v) \times s_{u,v}}{\sum_{v \in K_u} |s_{u,v}|} \quad (3)$$

$$\hat{r}_{u,i} = \frac{\sum_{j \in K_i} r_{u,j} \times s_{i,j}}{\sum_{j \in K_i} |s_{i,j}|} \quad (4)$$

Donde	\bar{r}_u :	Promedio de las valoraciones del usuario u
	\bar{r}_v :	Promedio de las valoraciones del usuario v
	$\hat{r}_{u,i}$:	Predicción de valoración del usuario u al ítem i
	$r_{v,i}$:	Valoración del usuario v al ítem i
	$r_{u,j}$:	Valoración del usuario u al ítem j
	K_i :	Vecinos del ítem i
	K_u :	Vecinos del usuario u
	$s_{i,j}$:	Similitud entre el ítem i y el ítem j
	$s_{u,v}$:	Similitud entre el usuario u y el usuario v

Ejemplo 2.1. El usuario Eric tiene que decidir si alquilar o no la película "Titanic" que aún no ha visto (Tabla. 2.1). Él sabe que Lucy tiene gustos muy similares en cuanto a películas, ya que ambos odiaban "The Matrix" y amaban "Forrest Gump", por lo que le pide su opinión sobre esta película. Por otro lado, Eric descubre que él y Diane tienen gustos diferentes, a Diane le gustan las películas de acción mientras que a él no, y él descarta su opinión o considera lo contrario en su decisión.

	The Matrix	Titanic	Die Hard	Forrest Gump	Wall-E
John	5	1		2	2
Lucy	1	5	2	5	5
Eric	2	?	3	5	4
Diane	4	3	5	3	

Tabla 2.1: Ejemplo que muestra las calificaciones de cuatro usuarios para cinco películas

Predicción con enfoque Basado en Usuario. Supongamos que queremos utilizar el enfoque de recomendación basada en usuario para predecir la calificación de Eric para la película "Titanic" usando las calificaciones de Lucy y Diane para esta película. Además, suponga que los pesos de similitud entre estos vecinos y Eric son respectivamente 0,75 y 0,15. Según la ecuación 3, la calificación prevista sería:

$$\hat{r} = \frac{0.75 \times 5 + 0.15 \times 3}{0.75 + 0.15} \approx 4.67$$

Predicción con enfoque Basado en Ítem. Supongamos que nuestra predicción se hace nuevamente usando dos vecinos más cercanos, y que los elementos más similares a "Titanic" son "Forrest Gump" y "Wall-E", con pesos de similitud respectivos de 0,85 y 0,75. Dado que Eric otorgó calificaciones de 5 y 4 a estas dos películas. Según la ecuación 4, la calificación pronosticada se calcula como:

$$\hat{r} = \frac{0.85 \times 5 + 0.75 \times 4}{0.85 + 0.75} \approx 4.53$$

2.2.2. Basado en Modelo

Los enfoques basados en modelos son también una técnica ampliamente utilizada en las tareas de recomendación, que buscan predecir las preferencias de los usuarios sobre elementos en un sistema de información. Estos enfoques utilizan un modelo predictivo que se entrena con datos históricos de las interacciones usuario-elemento, y luego se utiliza para predecir las calificaciones de los usuarios para nuevos elementos (Grčar et al., 2006).

Uno de los algoritmos más populares en los enfoques basados en modelo es la Descomposición en Valores Singulares (SVD, por sus siglas en inglés), una técnica que

se utiliza para reducir la dimensionalidad de los datos al encontrar los factores latentes que subyacen a las interacciones usuario-item (Fig. 2.5) (Grčar et al., 2006).

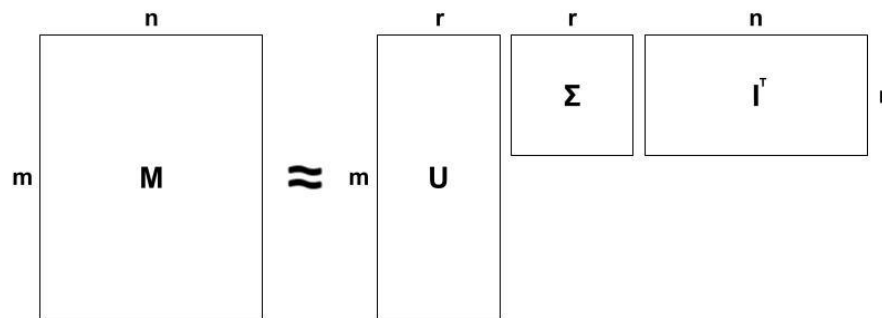


Figura 2.5: Descomposición de valores singulares (SVD) en tres matrices

El concepto en la parte posterior de SVD es que las interacciones usuario-ítem se pueden representar por medio de una matriz, en la que cada fila representa un usuario, cada columna representa un elemento y cada celda representa el puntaje que la persona le dio al objeto (Matriz M). SVD descompone esta matriz en tres matrices más pequeñas: una matriz de componentes del consumidor (matriz U), una matriz de componentes del artículo (matriz I) y una matriz diagonal de valor novedoso (matriz Σ) (Grčar et al., 2006).

La matriz U representa las características latentes de los usuarios que influyen en sus calificaciones, mientras que la matriz I representa las características latentes de los elementos que influyen en las calificaciones de los usuarios. La matriz Σ proporciona información sobre la importancia relativa de cada uno de estos factores. Una vez que se han encontrado estas tres matrices, se pueden utilizar para predecir las calificaciones de los usuarios para nuevos elementos (Grčar et al., 2006).

SVD es un enfoque muy popular en la recomendación porque es muy efectivo en la identificación de los factores latentes que subyacen a las interacciones usuario-elemento. Además, a diferencia de otros métodos de reducción de dimensionalidad, como el análisis de componentes principales (PCA) (Abdi & Williams, 2010), SVD tiene en cuenta la estructura de los datos de interacción usuario-elemento y puede identificar patrones de interacción complejos que no se pueden capturar con otros métodos (Grčar et al., 2006).

Además de SVD, existen varios enfoques basados en modelos que se han utilizado en las tareas de recomendación. Estos incluyen *Bayesian Clustering* (Breese et al., 1998), *Latent Semantic Analysis* (Hofmann, 2003), *Latent Dirichlet Allocation* (Blei et

al., 2003), *Maximum Entropy* (Zitnick et al., 2004), *Boltzmann Machines* (Salakhutdinov et al., 2007) y *Support Vector Machines (SVM)* (Grčar et al., 2006).

Bayesian Clustering es un enfoque que utiliza técnicas de agrupamiento para identificar grupos de usuarios con preferencias similares, mientras que *Latent Semantic Analysis* y *Latent Dirichlet Allocation* se basan en modelos probabilísticos para identificar los temas o conceptos latentes que subyacen a las interacciones usuario-elemento. *Maximum Entropy* se utiliza para modelar la probabilidad de que un usuario dé una calificación determinada en función de las características de los elementos, mientras que *Boltzmann Machines* utiliza redes neuronales para modelar las interacciones usuario-elemento. Por último, *Support Vector Machine* se utiliza para identificar la función de decisión que mejor separa las calificaciones positivas y negativas.

Aunque cada uno de estos enfoques tiene sus propias ventajas y limitaciones, SVD sigue siendo uno de los enfoques más populares y efectivos en las tareas de recomendación debido a su capacidad para identificar los factores latentes que subyacen a las interacciones usuario-elemento y su capacidad para manejar grandes conjuntos de datos. Sin embargo, es importante tener en cuenta que no existe un enfoque único que funcione para todas las situaciones y que la elección del enfoque debe basarse en las características específicas del problema de recomendación en cuestión.

2.3. Híbrido

Aunque en nuestra propuesta nos centramos con CF-RS y CB-RS, hemos de indicar que un estima categoría extendida de RS tanto en teoría como práctica son los sistemas de recomendación híbridos que combina diferentes técnicas de recomendación con el fin de obtener una mejor optimización del sistema para evitar algunas limitaciones y problemas de los RSs de las clases anteriores en solitario (Stern et al., 2009). La idea detrás de las técnicas híbridas es que una combinación de algoritmos proporcionará recomendaciones más precisas y efectivas que un solo algoritmo, ya que las desventajas de un algoritmo pueden ser superadas por otro algoritmo (Schafer et al., 2007).

Existen diversas maneras de diseñar sistemas de recomendación híbridos, que combinan diferentes enfoques y algoritmos para mejorar la precisión y eficacia de las recomendaciones. Entre estas opciones, se destacan tres formas principales de diseño: el

diseño en conjunto (*Joint design*), el diseño monolítico (*Monolithic design*) y los sistemas mezclados (*Mixed systems*) (Aggarwal, 2016).

- ❖ **Joint design:** Se aplican diferentes algoritmos de recomendación sobre un mismo conjunto de datos de entrada para generar un único resultado más robusto. Esto puede hacerse en paralelo, donde los resultados de cada algoritmo se combinan para obtener una lista de recomendaciones final (Fig. 2.6), o de manera secuencial, donde la salida de un algoritmo se utiliza como entrada para el siguiente algoritmo y así sucesivamente hasta obtener el resultado final (Fig. 2.7).

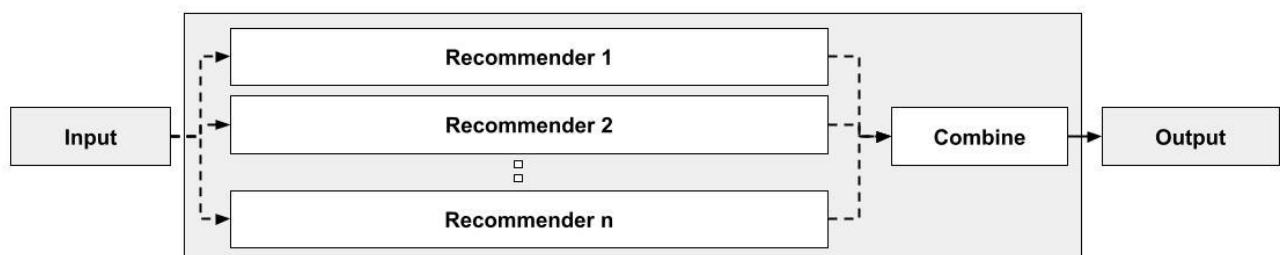


Figura 2.6: Diseño paralelo de un recomendador híbrido (Aggarwal, 2016)

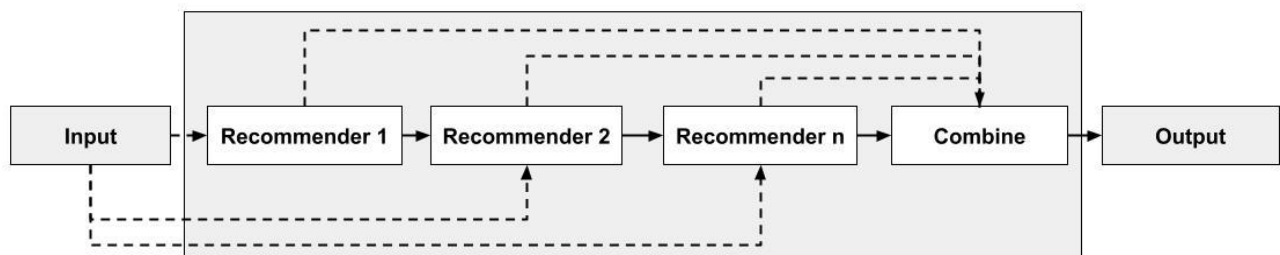


Figura 2.7: Diseño secuencial de un recomendador híbrido (Aggarwal, 2016)

- ❖ **Monolithic design:** Se utiliza un único algoritmo de recomendación para procesar varios conjuntos de datos de entrada y generar un único resultado de recomendación. En este enfoque, no se combinan diferentes algoritmos, sino que se confía en un solo algoritmo para proporcionar las recomendaciones.
- ❖ **Mixed systems:** Se utilizan diferentes algoritmos de recomendación sobre el mismo conjunto de datos de entrada, similar al diseño en conjunto. Sin embargo, en lugar de generar un único resultado, se obtienen varios resultados de diferentes algoritmos, y luego se combinan o presentan de alguna otra manera para que el usuario pueda elegir entre ellos o recibir múltiples recomendaciones.

Dentro de estos tres tipos de diseños, existen diversas estrategias de hibridación que se pueden utilizar (La Figura 2.8 muestra la clasificación de estas técnicas en los tres tipos de diseños vistos antes), incluyendo la ponderación de puntajes (*Weighted*), la conmutación (*Switching*), la presentación mixta de recomendaciones (*Mixed*), la combinación de características (*Feature Combination*), el enfoque en cascada (*Cascade*), el aumento de características (*Feature Augmentation*) y el enfoque de meta-nivel (*Meta-Level*). Cada una de estas estrategias tiene ventajas y limitaciones únicas, por lo que es fundamental elegir la estrategia adecuada según las necesidades específicas del sistema de recomendación y los datos disponibles (Burke, 2002). A continuación, explicamos cada una de estas estrategias:

- ❖ **Weighted:** se combinan los puntajes (o votos) de varias técnicas de recomendación para producir una sola recomendación. Cada técnica puede ser asignada con un peso diferente, lo que afecta la influencia que tiene en la recomendación final.
- ❖ **Switching:** el sistema alterna entre diferentes técnicas de recomendación. Por ejemplo, si un usuario tiene un historial de compras específico, se puede utilizar un algoritmo de recomendación basado en contenido, mientras que si el usuario tiene una gran cantidad de interacciones sociales en la plataforma, se puede utilizar un enfoque basado en filtrado colaborativo.
- ❖ **Mixed:** se presentan recomendaciones de varios recomendadores diferentes al mismo tiempo. Estas recomendaciones se pueden presentar como una lista combinada, o como diferentes opciones de recomendación para que el usuario pueda elegir entre ellas.
- ❖ **Feature Combination:** las características de diferentes fuentes de datos de recomendación se combinan en un solo algoritmo de recomendación. Por ejemplo, se pueden combinar datos de historial de compras y datos de interacciones sociales para generar recomendaciones más precisas.
- ❖ **Cascade:** un recomendador refina las recomendaciones dadas por otro. En este enfoque, se utilizan múltiples algoritmos de recomendación en cascada, donde el resultado de un recomendador se utiliza como entrada para el siguiente recomendador.
- ❖ **Feature Augmentation:** la salida de una técnica se utiliza como una característica de entrada para otra técnica. Por ejemplo, el resultado de un algoritmo de

recomendación basado en contenido puede utilizarse como una característica de entrada para un algoritmo de filtrado colaborativo.

- ❖ **Meta-Level:** el modelo entrenado por un recomendador se utiliza como entrada para otro. Por ejemplo, el modelo de regresión logística entrenado por un algoritmo de recomendación basado en contenido puede utilizarse como entrada para un algoritmo de aprendizaje profundo para generar recomendaciones más precisas.

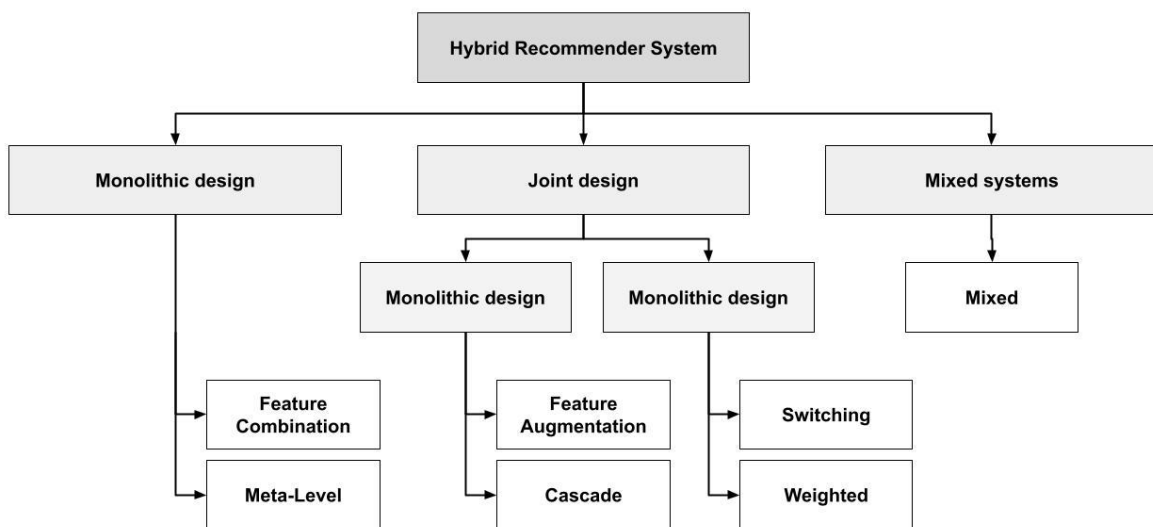


Figura 2.8: Clasificación de las técnicas de para combinar estrategias de recomendación en los tres tipos de diseños híbridos (Aggarwal, 2016)

El uso de sistemas de recomendación híbridos tiene ventajas significativas en comparación con los sistemas de recomendación basados en un solo enfoque. La combinación de múltiples técnicas de recomendación puede suprimir las debilidades de una técnica individual en un modelo combinado.

3. Sistema de recomendación grupal

En los últimos años, ha habido un creciente interés en los sistemas de recomendación grupales (GRS) (Fig. 2.9), especialmente en entornos sociales en línea donde los usuarios interactúan en grupos. A medida que el número de usuarios y la cantidad de información disponible en línea aumenta, se vuelve cada vez más difícil para los usuarios encontrar la información relevante y personalizada. Los GRSs tienen como objetivo abordar este problema proporcionando recomendaciones no solo basadas en las

preferencias individuales de un usuario, sino también en las preferencias del grupo al que pertenece.

Los investigadores en sistemas de recomendación como Castro, Yera y Martínez (2018) y De Pessemier et al. (2014) han comenzado a considerar la existencia de elementos sociales, tales como programas de televisión, lugares turísticos, música, películas, destinos de vacaciones, etc., que a menudo son consumidos por grupos en lugar de usuarios individuales. En estas situaciones, recomendar elementos que satisfagan las preferencias de todos los miembros del grupo es más difícil de lograr en comparación con la recomendación individual.

Documents by year

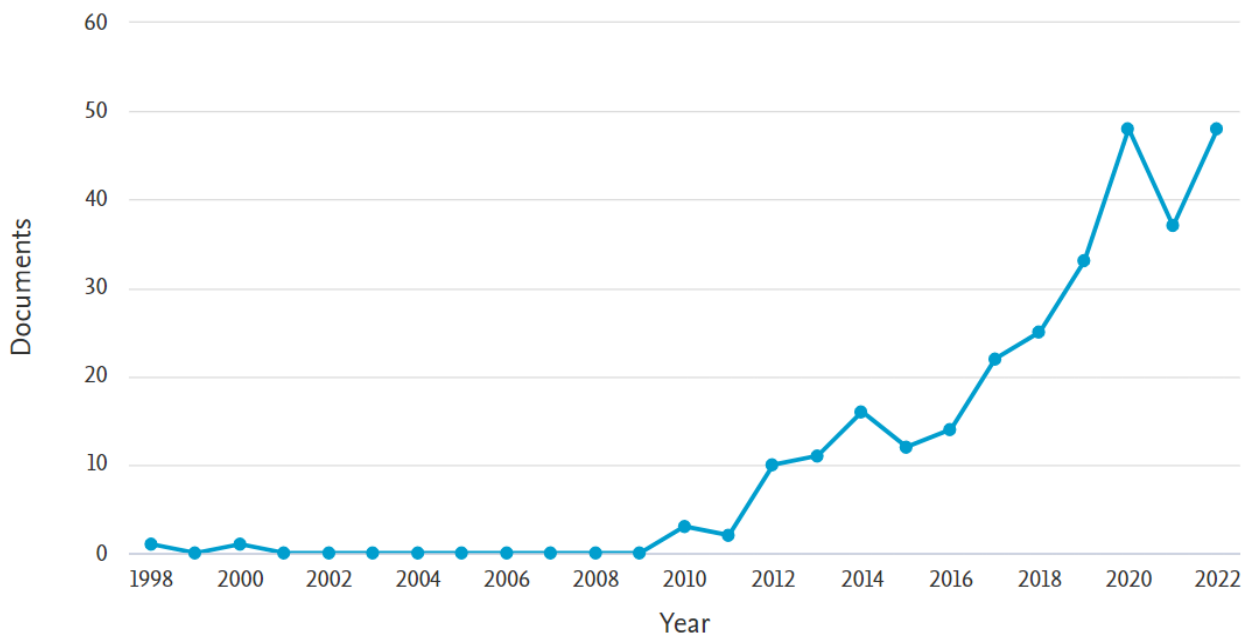


Figura 2.9: Número de documentos publicados sobre los sistemas de recomendación grupales en la base de datos Scopus

Por lo tanto, los sistemas de recomendación en grupo (GRSs) se enfocan en ampliar los sistemas de recomendación (RS) individuales para recomendar un artículo a un grupo de usuarios, mediante la agregación de información asociada a cada usuario individual (Castro, Barranco, Rodríguez y Martínez, 2018; Castro et al., 2017). A continuación, se describen las estrategias más utilizadas para crear un modelo grupal a partir de la agregación de modelos individuales.

3.1. Estrategias de agregación de los modelos individuales

Las primeras publicaciones científicas sobre GRS datan de finales de los noventa (McCarthy & Anagnost, 1998). Desde entonces, muchos investigadores ya han investigado cómo se pueden adaptar los actuales algoritmos de recomendación de última generación para generar recomendaciones grupales. En la literatura, las recomendaciones grupales se generaron principalmente agregando las recomendaciones individuales de los usuarios en recomendaciones para todo el grupo (agregación de recomendaciones) o agregando los modelos de preferencias individuales de los usuarios en un modelo de preferencia del grupo (agregación de preferencias) (De Pessemier et al., 2014).

- ❖ **Agregación de recomendaciones:** Esta estrategia genera recomendaciones para cada usuario individual utilizando un algoritmo de recomendación tradicional (recomendador interno). Posteriormente, las listas de recomendaciones de todos los miembros del grupo se agregan en una lista de recomendaciones del grupo que satisface a todos ellos (Fig. 2.10).

Esta estrategia tiene algunas ventajas. Por ejemplo, las recomendaciones resultantes se pueden vincular directamente con las recomendaciones individuales, lo que las hace fáciles de explicar basándose en las explicaciones del recomendador interno (Herlocker et al., 2000).

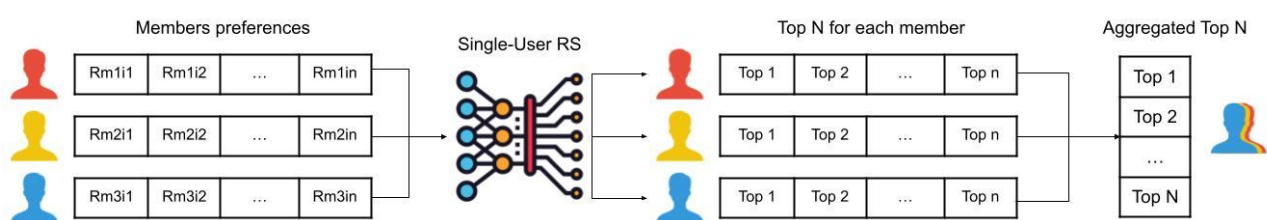


Figura 2.10: Sistema de recomendación basado en agregación de recomendaciones

- ❖ **Agregación de preferencias:** Esta estrategia combina las preferencias de los usuarios en preferencias grupales. De esta forma, las opiniones y preferencias de los miembros individuales del grupo constituyen un modelo de preferencia grupal que refleja los intereses de todos los miembros (Fig. 2.11). En la literatura, se han propuesto diferentes enfoques para agregar las preferencias de los miembros, pero aún no existe un consenso sobre la solución óptima (Baltrunas et al., 2010).

Después de agregar las preferencias de los miembros, el modelo de preferencia del grupo se trata como un pseudo-usuario para producir recomendaciones para el grupo utilizando el algoritmo interno. En comparación con la agregación de recomendaciones, la agregación de las preferencias de los usuarios aumenta la posibilidad de encontrar recomendaciones. Por otro lado, la agregación de las preferencias puede dar lugar a sugerencias grupales que se encuentran fuera del rango de cualquier lista de recomendaciones individuales, lo que puede hacer las recomendaciones difíciles de explicar (Herlocker et al., 2000).

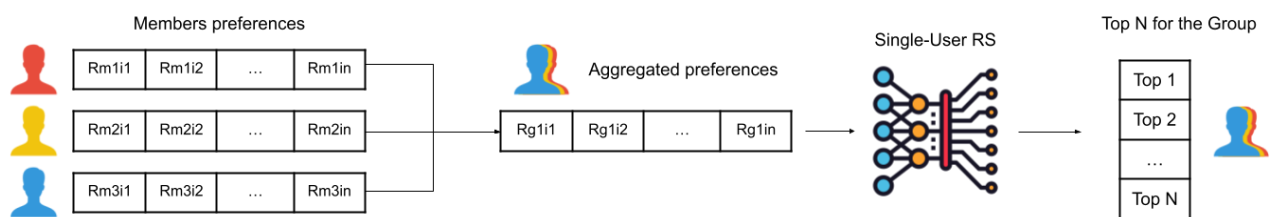


Figura 2.11: Sistema de recomendación basado en agregación de preferencias

Ha habido un número relativamente pequeño de estudios sobre los GRSs hasta el momento, entre ellos encontramos:

- ❖ **MusicFX** (McCarthy & Anagnost, 1998) selecciona estaciones de música que transmiten a un gimnasio lleno de gente. Los miembros del gimnasio califican todas las estaciones de antemano y MusicFX reproduce una de las estaciones con la calificación promedio más alta. El sistema intenta así maximizar la satisfacción del grupo. Un requisito previo crucial de MusicFX es que los usuarios deben calificar todas las estaciones por adelantado. Para la selección de estaciones de música entre un pequeño número de opciones, MusicFX sería un buen sistema, pero MusicFX o sus variaciones no serían buenos para recomendar elementos como libros entre un gran número de opciones, que deben ser calificados por los usuarios con antelación.
- ❖ **PolyLens** (O'Connor et al., 2001) recomienda películas a pequeños grupos de personas que ven películas juntas. Aplica el algoritmo CF estándar para encontrar recomendaciones para cada uno de los miembros del grupo y luego las combina en una recomendación grupal. A diferencia de MusicFX, PolyLens intenta satisfacer a todos los usuarios hasta cierto punto, sin necesariamente maximizar la satisfacción

promedio del grupo. PolyLens basa sus recomendaciones en la satisfacción esperada del miembro del grupo menos satisfecho. Por lo tanto, se puede recomendar una película que es marginalmente aceptable para la mayoría de los miembros del grupo en lugar de una que no le gusta a un miembro pero que todos los demás disfrutarían inmensamente.

- ❖ **Pocket RestaurantFinder** (McCarthy, 2002) recomienda restaurantes a grupos de comensales, en función de sus ubicaciones y preferencias culinarias. Los posibles comensales llenan perfiles de sus preferencias en dispositivos portátiles. El perfil incluye atributos de selección de restaurantes, como qué tan lejos están dispuestos a viajar, cuánto están dispuestos a gastar, etc. Pocket RestaurantFinder agrupa las preferencias de los miembros del grupo y presenta un conjunto de restaurantes potenciales clasificados en el orden de conveniencia esperado para el grupo.
- ❖ **Adaptive Radio** (Chao et al., 2004) es un sistema que transmite canciones a un grupo de oyentes que comparten un mismo entorno. La mayoría de los RSs de música determinan qué tipo de canciones prefieren los usuarios, generalmente mediante encuestas o perfiles en línea de sus hábitos de escucha. Adaptive Radio, por el contrario, realiza un seguimiento de las canciones que no les gustan a los usuarios y evita reproducirlas. Como efecto secundario del uso de preferencias negativas, Adaptive Radio a menudo reproduce canciones desconocidas o marginalmente aceptables para los usuarios.

En un GRS, debemos considerar hasta qué grado las recomendaciones para un grupo coinciden con los gustos de los individuos, ya que es difícil y, a menudo, imposible satisfacer a todos los miembros del grupo. La satisfacción del grupo se puede medir como la satisfacción promedio de los miembros, la satisfacción del miembro más satisfecho o la satisfacción del miembro menos satisfecho (Kim et al., 2010).

En los estudios existentes, por ejemplo, MusicFX y Pocket RestaurantFinder, el valor promedio se usa generalmente, pero puede generar recomendaciones sesgadas hacia los usuarios que brindan más calificaciones de contenidos o productos. Otro método, utilizado por PolyLens y Adaptive Radio, adopta la satisfacción del miembro menos satisfecho. Este método puede ignorar las preferencias de la mayoría mientras apoya a un pequeño número de miembros insatisfechos (Kim et al., 2010).

3.2. Métodos de agregación

En los sistemas de recomendación de grupo (GRS), la agregación de opiniones y evaluaciones de múltiples individuos es un aspecto fundamental. Existen diferentes estrategias y métodos de agregación que pueden ser utilizados para llegar a una decisión colectiva, y cada uno de ellos se enfoca en distintos aspectos y criterios, lo que puede hacerlos más apropiados para diferentes situaciones y contextos (Cantador y Castells, 2012; Beliakov, Calvo y James, 2015; Masthoff, 2004).

En la literatura, se ha empleado una clasificación ampliamente utilizada que se basa en tres categorías establecidas para los métodos de agregación (Senot et al., 2010):

- ❖ La primera categoría son las estrategias basadas en la mayoría, que fortalecen la opción más popular (Senot et al., 2010). Algunos ejemplos de estas estrategias son: el método de *Borda Count*, que crea un ranking de ítems en cada usuario asignándoles puntos en función de su posición en la lista de preferencias; el método de *Copeland Rule*, que cuenta la frecuencia con la que un elemento supera a otros elementos utilizando votos mayoritarios; y el método de *Plurality Voting*, que simplemente escoge el elemento con más votos.
- ❖ La segunda categoría son las estrategias basadas en el consenso o democráticas, que promedian todas las opciones disponibles (Senot et al., 2010). Por ejemplo, el método de *Average* realiza el promedio de los valores individuales de opinión o evaluación de los individuos. Otros métodos democráticos incluyen el *Additive Utilitarian*, que suma los valores individuales, y el *Multiplicative Utilitarian*, que multiplica los valores individuales. Estos métodos buscan encontrar un punto medio o consenso entre las opiniones de los individuos; y
- ❖ La tercera categoría son las estrategias *Borderline* o basadas en roles, que toman en cuenta un subconjunto de roles de usuario o cualquier otro criterio relevante (Senot et al., 2010). Por ejemplo, el método de *Least misery* toma el mínimo de los valores individuales, es decir, busca minimizar la insatisfacción o descontento de los individuos, mientras que el método de *Most pleasure* toma el máximo de los valores individuales, buscando maximizar la satisfacción o placer de los individuos.

En nuestra propuesta, vamos a evaluar cinco métodos de agregación diferentes: Average (Promedio), Median (Mediana), Most Pleasure (Máximo placer), Least Misery

(Mínima miseria) y Random (Aleatorio). A continuación, se proporcionan las fórmulas correspondientes a cada uno de estos métodos:

1. Average (Promedio): La fórmula del método de Average consiste en calcular el promedio de los valores individuales de opinión o evaluación de los individuos. Siendo n el número de individuos, la fórmula sería (Ec. 5):

$$v = \frac{1}{n} \sum_{i=1}^n v_i \quad (5)$$

2. Median (Mediana): El método de Median encuentra el valor central en una lista ordenada de valores individuales. Si hay un número par de valores, se toma el promedio de los dos valores del medio. Siendo n el número de individuos, los pasos para calcular la mediana son:

- ❖ Ordenar los valores individuales de menor a mayor.
- ❖ Si n es impar, la mediana es el valor en la posición $(n + 1) / 2$.
- ❖ Si n es par, la mediana es el promedio de los valores en las posiciones $(n / 2)$ y $(n / 2) + 1$.

3. Most Pleasure (Máximo placer): El método de Most Pleasure busca maximizar la satisfacción o placer de los individuos. La fórmula simplemente consiste en seleccionar el valor más alto de entre los valores individuales (Ec. 6).

$$v = \max(v_1, v_2, \dots, v_n) \quad (6)$$

4. Least Misery (Mínima miseria): El método de Least Misery busca minimizar la insatisfacción o descontento de los individuos. La fórmula consiste en seleccionar el valor más bajo de entre los valores individuales (Ec. 7).

$$v = \min(v_1, v_2, \dots, v_n) \quad (7)$$

5. Random (Aleatorio): El método de Random selecciona aleatoriamente uno de los valores individuales como resultado de la agregación. No sigue una fórmula específica, ya que la elección se basa en un proceso de selección aleatoria.

Estos métodos de agregación ofrecen diferentes enfoques para llegar a una decisión colectiva y pueden ser evaluados en términos de su desempeño y eficacia en relación con los objetivos y criterios específicos de evaluación de un sistema de recomendación de grupo.

4. IA Transparente: Sistemas de Recomendación Explicables

La inteligencia artificial (IA) ha avanzado rápidamente en los últimos años, permitiendo el desarrollo de aplicaciones y sistemas que son capaces de realizar tareas que antes solo podían ser realizadas por humanos. Sin embargo, a medida que la IA se ha vuelto más compleja y sofisticada, también ha surgido la necesidad de comprender y explicar cómo funciona, especialmente en contextos en los que se toman decisiones importantes que afectan a las personas, como en sistemas de recomendación (Arrieta et al., 2020).

La explicabilidad en la IA, también conocida como Inteligencia Artificial explicable (XAI, por sus siglas en inglés), se refiere a la capacidad de un modelo de IA para ser comprensible y explicado de manera clara y transparente, de manera que las personas puedan entender cómo se obtienen las predicciones o recomendaciones realizadas por el modelo. La explicabilidad en la AI se considera un aspecto fundamental para asegurar la confianza y aceptación de los sistemas de IA en la sociedad, ya que permite a las personas entender y cuestionar las decisiones tomadas por el modelo (Arrieta et al., 2020).

Como los sistemas de recomendación son un tipo de aplicación de IA, también enfrentan desafíos en términos de explicabilidad, ya que los algoritmos utilizados en estos sistemas suelen ser complejos y opacos, lo que dificulta la comprensión de cómo se generan las recomendaciones. Esto puede generar desconfianza en los usuarios, especialmente cuando se trata de decisiones importantes, como la compra de un producto o la elección de una película para ver.

Es por ello que la XAI se ha convertido en un tema de investigación y desarrollo activo en el campo de los sistemas de recomendación. En los últimos años, los investigadores y desarrolladores han propuesto diversas técnicas y enfoques para mejorar la explicabilidad de los sistemas de recomendación (Fig. 2.12), lo que ha llevado a la

emergencia de los sistemas de recomendación explicable (XRS), que son sistemas de recomendación que incorporan técnicas y enfoques de XAI para mejorar la explicabilidad de las recomendaciones generadas.

Documents by year

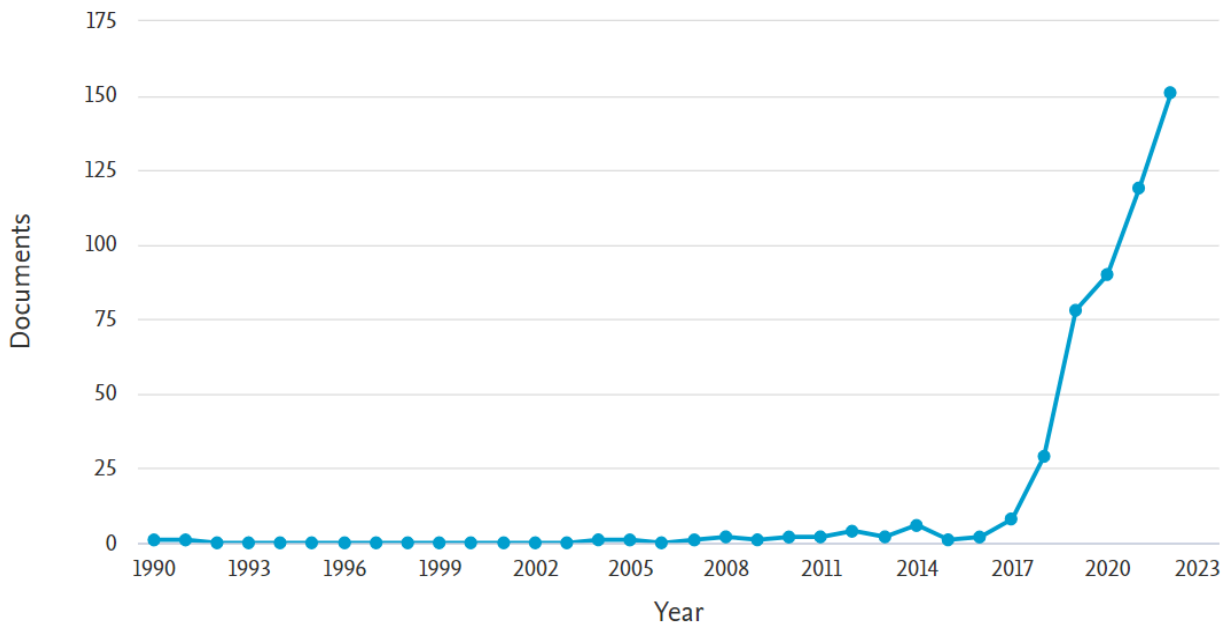


Figura 2.12: Número de documentos publicados sobre los sistemas de recomendación explicables en la base de datos Scopus

Dentro de las técnicas de XAI aplicadas a los sistemas de recomendación, existen dos tipos de algoritmos explicables: los intrínsecos y los post-hoc (Zhong & Negre, 2022). Los algoritmos intrínsecos se centran en explicar el funcionamiento interno del método principal de recomendación de caja negra, es decir, brindan explicaciones sobre cómo se generan las recomendaciones desde el algoritmo complejo utilizado. Por otro lado, los algoritmos post-hoc se enfocan en explicar la salida del método principal de recomendación de caja negra, sin tener en cuenta cómo se generan internamente las recomendaciones (Zhong & Negre, 2022).

En nuestra propuesta, nos enfocaremos en los algoritmos post-hoc, ya que son más genéricos y pueden adaptarse a cualquier modelo de recomendación de caja negra. Estos algoritmos analizan la salida del sistema de recomendación y buscan generar explicaciones comprensibles para los usuarios. Su objetivo es proporcionar información sobre los factores o características que influyeron en la generación de las recomendaciones, sin necesidad de modificar el algoritmo principal de recomendación.

Un enfoque desarrollado por Shmaryahu et al. (2020) que busca proporcionar explicaciones claras y comprensibles utilizando los algoritmos simples e interpretables. Este enfoque se basa en un marco general y se compone de varias etapas (Fig. 2.13). Primero, un modelo de recomendación de caja negra genera una lista de ítems recomendados según las preferencias del usuario. A continuación, se aplican algoritmos de similitud basados en el contenido de los ítems para generar explicaciones, destacando los ítems similares a los recomendados. Estas explicaciones se clasifican y se muestra la mejor explicación al usuario. Por ejemplo, un algoritmo basado en el contenido puede calcular la similitud entre el ítem recomendado y los ítems previamente evaluados por el usuario, resaltando las características comunes entre ellos. De esta manera, se brinda al usuario una explicación clara y significativa sobre las recomendaciones de ítems.

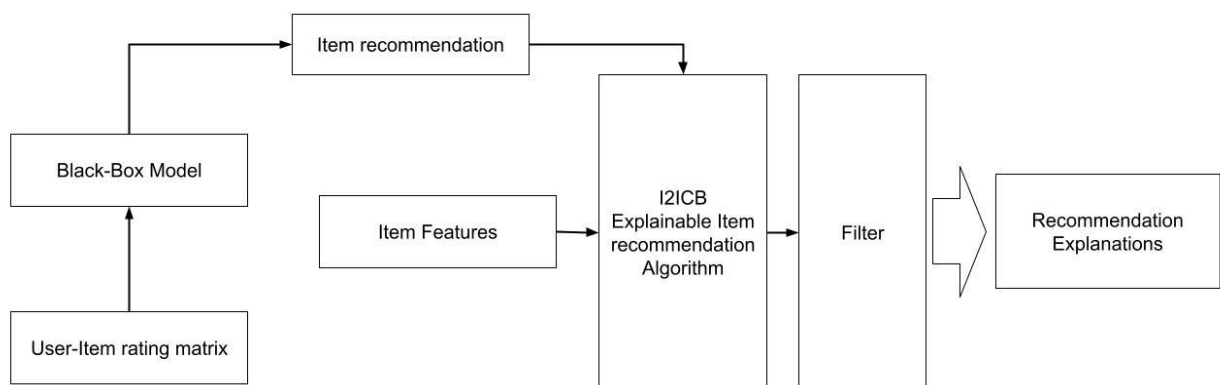


Figura 2.13: Explicaciones basadas en métodos más simples

Se ha propuesto otro enfoque para incorporar la explicabilidad utilizando la técnica LIRE (Chanson et al., 2021), una versión mejorada de *Local Interpretable Model-Agnostic Explanations* (LIME) (Ribeiro et al., 2016). Este enfoque consta de varias etapas (Fig. 2.14). Primero, un modelo de recomendación de caja negra genera una lista de recomendaciones basadas en las preferencias del usuario. A continuación, se crean instancias perturbadas basadas en el ítem que se desea explicar, donde se introducen cambios aleatorios en características interpretables. Estas instancias perturbadas se utilizan para entrenar un modelo interpretable (regresión lineal) (Ribeiro et al., 2016) que identifica las características más relevantes. Finalmente, se generan las explicaciones: se obtiene un conjunto de pesos asociados a cada característica interpretable, enfocados en mostrar la importancia local de la característica correspondiente en la instancia actual. Por lo tanto, tales características, en este caso representadas como otros nombres de ítems

(Fig. 2.14), pueden usarse como una forma de explicar la generación actual de recomendaciones.

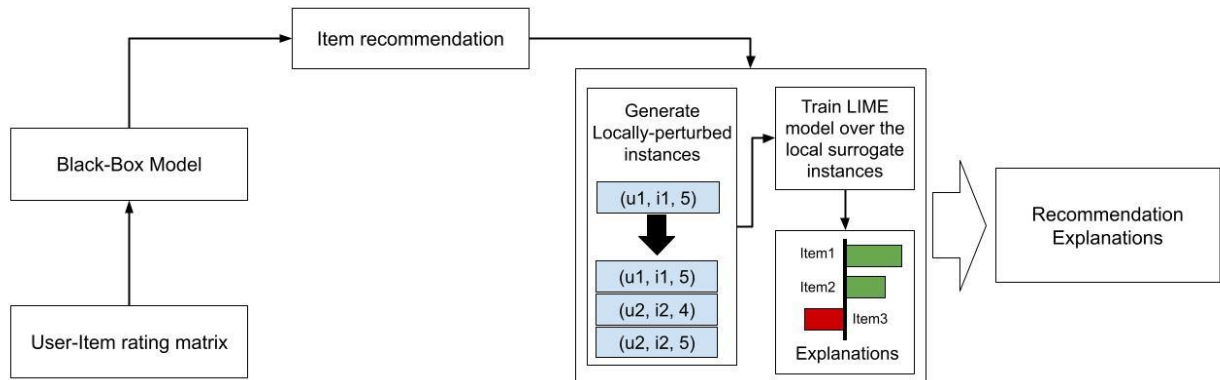


Figura 2.14: Explicaciones basadas en LIRE

Es importante destacar que la explicabilidad en los sistemas de recomendación explicable no implica necesariamente revelar toda la información utilizada por el sistema de recomendación, como datos sensibles o algoritmos propietarios. La explicabilidad se refiere más bien a proporcionar una descripción clara y comprensible del proceso de recomendación y las razones detrás de las decisiones tomadas por el modelo, sin comprometer la privacidad y seguridad de los datos.

La XAI y los sistemas de recomendación explicable tienen implicaciones no solo en el ámbito técnico, sino también en el ámbito ético y legal. La transparencia y explicabilidad en los sistemas de recomendación son importantes para garantizar la equidad y la no discriminación en las recomendaciones realizadas, así como para cumplir con regulaciones y normativas, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, que establece el derecho de los individuos a una explicación clara y comprensible de las decisiones automatizadas que les afectan.

La explicabilidad en los RSs no sólo es importante para los usuarios finales, sino también para los proveedores de servicios. Los proveedores de servicios necesitan comprender los factores que influyen en las recomendaciones de sus sistemas para poder optimizar su rendimiento y hacer mejoras.

5. Conclusión

En este capítulo hemos explorado diversas técnicas de recomendación, incluyendo el enfoque basado en contenido, el filtrado colaborativo basado en memoria y en modelo, así como los sistemas de recomendación híbridos. También hemos examinado los sistemas de recomendación grupal, centrándonos en las estrategias y métodos de agregación de los modelos individuales.

Una de las áreas emergentes en el campo de los sistemas de recomendación es la transparencia y explicabilidad. Con la creciente preocupación por la ética y la responsabilidad en la inteligencia artificial, es fundamental que los sistemas de recomendación sean explicables, es decir, que puedan proporcionar razones claras y comprensibles para las recomendaciones que realizan. En este capítulo, hemos abordado la importancia de la transparencia en los sistemas de recomendación, especialmente en el contexto de los sistemas de recomendación grupal.

La transparencia en los sistemas de recomendación no solo tiene implicaciones éticas, sino que también puede mejorar la confianza del usuario y fomentar una mayor adopción de estos sistemas. Los usuarios necesitan entender cómo se generan las recomendaciones y qué factores influyen en ellas, especialmente cuando se trata de decisiones importantes que pueden tener un impacto significativo en su vida o en la toma de decisiones de un grupo. Por lo tanto, los sistemas de recomendación grupal explicables pueden ser una solución prometedora para abordar la falta de transparencia en los sistemas de recomendación tradicionales.

Sin embargo, la implementación de sistemas de recomendación grupal explicables no está exenta de desafíos. La agregación de las preferencias de diferentes miembros del grupo, la combinación de diferentes técnicas de recomendación y la presentación de explicaciones claras y comprensibles son algunos de los desafíos clave a enfrentar. Es importante abordar estos desafíos para desarrollar sistemas de recomendación grupal explicables y efectivos que puedan ser utilizados en diversos contextos, como el comercio electrónico, la toma de decisiones en grupo y la personalización de contenido.

DISEÑO Y DESARROLLO DEL MODELO DE EXPLICABILIDAD DE SISTEMA DE RECOMENDACIÓN GRUPAL

CAPÍTULO 3

1. Introducción

En este capítulo se abordará la creación y diseño de un modelo de sistema de recomendación grupal que sea fácilmente interpretable y comprensible por los usuarios. El objetivo no solo es ofrecer recomendaciones personalizadas y relevantes, sino también brindar información clara y explicaciones sobre cómo y por qué se generan dichas recomendaciones.

Con el objetivo de lograr esta meta, se presentarán los elementos fundamentales del modelo de recomendación para grupos, que comprenden la representación de los grupos en el modelo y sus respectivas preferencias. Además, se explorarán diversas estrategias de agregación con el propósito de obtener las recomendaciones para grupos, las cuales extienden los algoritmos que normalmente se usan para individuos, tales como el filtrado colaborativo y el enfoque basado en contenido.

2. Descripción de los componentes del modelo

El modelo de recomendación grupal propuesto consta de dos componentes principales (Fig. 3.1), el primero es el motor de recomendación grupal, que se encarga de analizar las preferencias de un grupo de usuarios y proporcionar recomendaciones de items que sean de interés para todos los miembros del grupo. El segundo componente es el generador de explicaciones de recomendación, que tiene como objetivo explicar el razonamiento detrás de cada recomendación para que los usuarios puedan entender por qué se les está sugiriendo un determinado ítem.

En las siguientes secciones, se describe en detalle cada uno de estos componentes y su funcionamiento. Primero, se explicará cómo funciona el motor de recomendación grupal, incluyendo los algoritmos y técnicas utilizados para analizar las preferencias del grupo y generar recomendaciones.

Luego, se abordará el generador de explicaciones de recomendación y cómo se utiliza para proporcionar a los usuarios una explicación clara y concisa de cada recomendación. Con esta información detallada, se espera que los usuarios comprendan mejor las recomendaciones y puedan tomar decisiones informadas sobre qué ítems seleccionar.

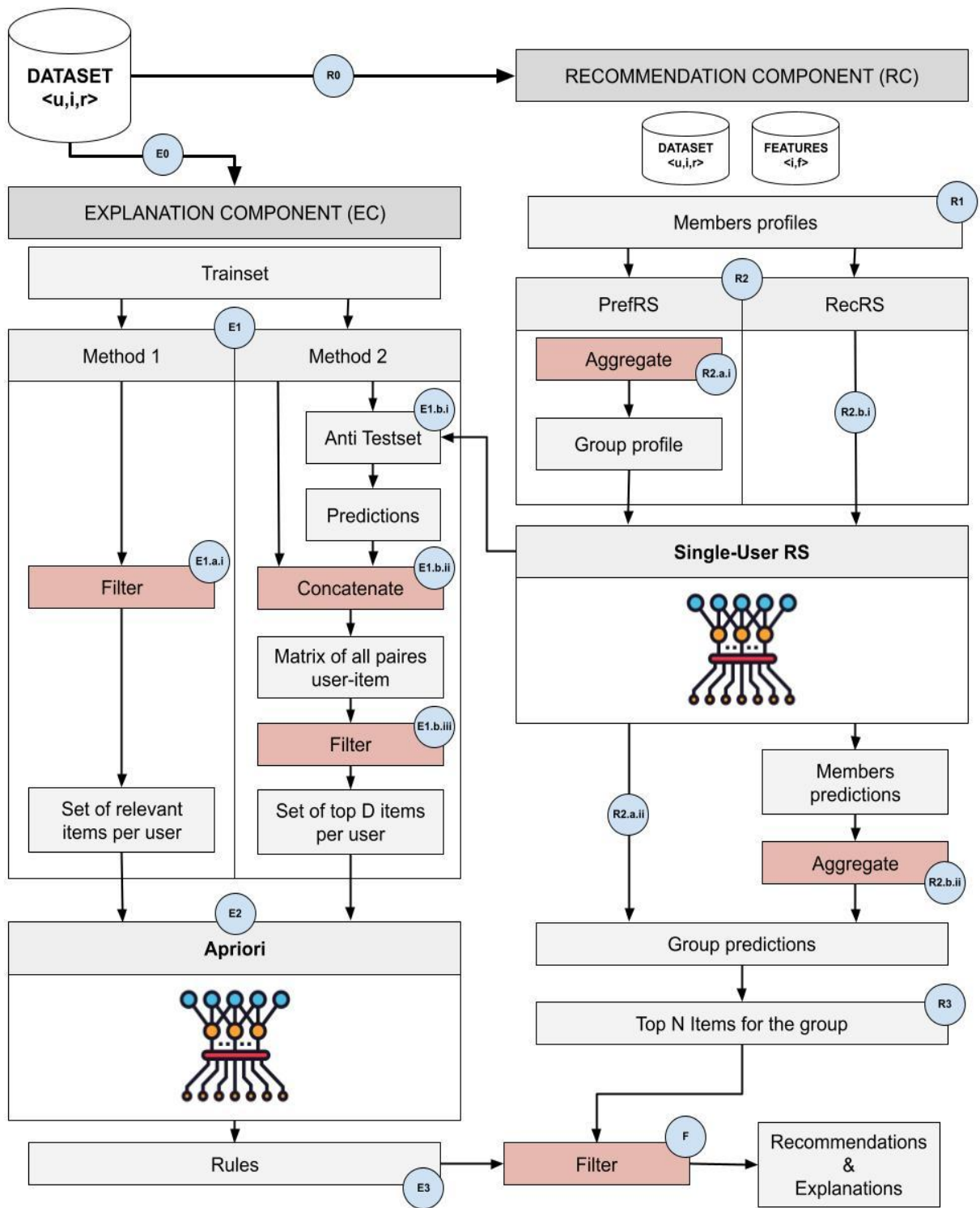


Figura 3.1: Componentes del modelo de explicabilidad de sistemas de recomendación grupal

2.1. Componente de recomendación

En el proceso de generación de recomendaciones grupales, el componente de recomendación (RC) desempeña un papel fundamental. El RC recibe un conjunto de

datos que contiene las calificaciones otorgadas por los usuarios a los elementos del sistema (Etapa R0 en Fig. 3.1). Utilizando esta información, el RC emplea diversas estrategias de recomendación grupal, como la agregación de preferencias (PrefRS) o la agregación de recomendaciones (RecRS), con el fin de predecir aquellos elementos que satisfacen los requisitos de todos los miembros del grupo.

Inicialmente, en ambas estrategias, se crean perfiles individuales de los usuarios mediante el uso de las calificaciones asignadas a los elementos del sistema (Etapa R1 en Fig. 3.1). Estos perfiles individuales representan las preferencias de cada miembro del grupo.

En la estrategia de agregación de preferencias individuales (PrefRS) (Algoritmo 3.1), en primer lugar, se combinan los perfiles individuales para construir un perfil de grupo que refleje las preferencias colectivas (Etapa R2.a.i en Fig. 3.1) (Algoritmo 3.1, línea 3). Esto se logra mediante diferentes métodos de agregación, como el cálculo de la media aritmética o la mediana de los perfiles individuales. A continuación, el perfil de grupo resultante se emplea para predecir las calificaciones (Etapa R2.a.ii en Fig. 3.1) y generar recomendaciones en el siguiente paso, utilizando algoritmos de recomendación tradicionales (Algoritmo 3.1, línea 6).

Algorithm 3.1: *Preferences Aggregation*

Input: Dataset D

Input: Internal model

Input: Aggregation method

Input: n value

- 1: **for** $g \in \text{Groups}$ **do**
 - 2: Retrieve the set of users U belonging to group g
 - 3: Aggregate the preferences of users U per item using the specified **Aggregation method**
 - 4: Append the aggregated profile to the dataset D , resulting in D'
 - 5: Train the internal model with the dataset D'
 - 6: Calculate scores for the aggregated profile using the trained **internal model**
 - 7: Sort the unseen items based on the score value in descending order
 - 8: Recommend the top n items from the sorted list to group g
 - 9: **end for**
-

Algoritmo 3.1: Estrategia de agregación de preferencias

Por otro lado, en la estrategia de agregación de recomendaciones individuales (RecRS) (Algoritmo 3.2), en primer lugar, se utilizan algoritmos de recomendación individuales para calcular las predicciones para cada usuario individual (Etapa R2.b.i en Fig. 3.1) (Algoritmo 3.2, línea 2). Estas predicciones se basan en el perfil individual de cada miembro y en su historial de calificaciones. En segundo lugar, se agregan las predicciones individuales de cada usuario para obtener una predicción para el grupo (Etapa R2.b.ii en Fig. 3.1) (Algoritmo 3.2, línea 7). Esta agregación se realiza mediante diferentes métodos, tal como se mencionó anteriormente.

Algorithm 3.2: Recommendation Aggregation

Input: Dataset D

Input: Internal model

Input: Aggregation method

Input: n value

- 1: Train the **internal model** with the dataset D
 - 2: Calculate individual scores using the trained **internal model**
 - 3: Use the trained model to get the individual's predictions
 - 5: **for** $g \in \text{Groups}$ **do**
 - 6: Retrieve the set of users U belonging to group g
 - 7: Aggregate the individual scores of users U per item using the specified **Aggregation method**
 - 8: Sort the unseen items based on the aggregated score value in descending order
 - 9: Recommend the top n items from the sorted list to group g
 - 10: **end for**
-

Algoritmo 3.2: Estrategia de agregación de recomendaciones

Finalmente, en ambos enfoques se utilizan los puntajes obtenidos para seleccionar los elementos que serán recomendados al grupo (Etapa R3 en Fig. 3.1). Estos puntajes son empleados por el algoritmo de recomendación para ordenar los elementos y elegir aquellos que serán recomendados. Por ejemplo, en el enfoque de filtrado colaborativo, se pueden basar en los puntajes predichos, mientras que en el enfoque basado en contenido se utilizan los valores de similitud calculados entre el perfil del grupo y los perfiles de los elementos. Es importante destacar que todos los miembros del grupo recibirán la misma lista de recomendaciones.

2.2. Componente de explicación

En este trabajo, nuestro enfoque para abordar la explicabilidad en el sistema de recomendación grupal se basa en el uso de reglas de asociación. Esta elección se justifica debido a su capacidad para proporcionar una representación intuitiva y comprensible de las relaciones entre los ítems del sistema y las preferencias de los usuarios.

Estas reglas establecen conexiones lógicas entre los elementos consumidos o evaluados por los usuarios, lo que permite comprender el proceso de recomendación. Al identificar las reglas más frecuentes o con alta confianza, se obtiene información valiosa sobre las relaciones entre los ítems del sistema, lo cual puede utilizarse para explicar las recomendaciones proporcionadas.

En el modelo propuesto, el componente de explicación de recomendaciones (EC) utiliza técnicas de minería de datos para generar reglas de asociación a partir del conjunto de calificaciones de los usuarios recibido en la etapa E0 (Fig. 3.1). La generación de estas reglas se realiza mediante el algoritmo Apriori (Agrawal et al., 1996), el cual requiere como entrada un conjunto de transacciones. Estas transacciones representan conjuntos de ítems que reflejan las preferencias de los usuarios con respecto a los diferentes ítems del sistema (Fig. 3.2). En otras palabras, una transacción es una lista de ítems que un usuario ha consumido, comprado, clasificado o interactuado de alguna manera en el sistema de recomendación.

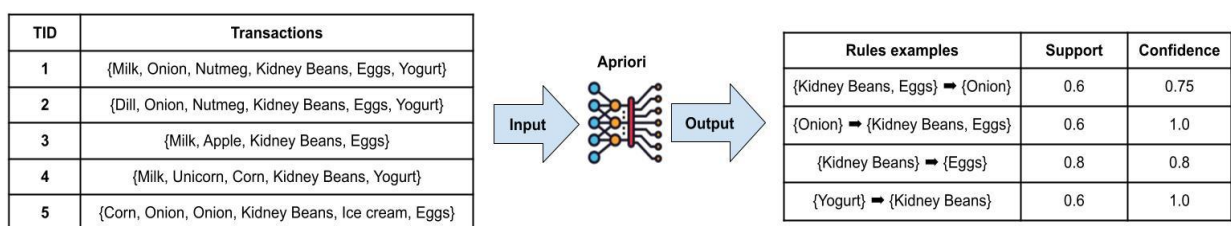


Figura 3.2. Generación de reglas de asociación utilizando el algoritmo Apriori

En el modelo propuesto, se han destacado dos estrategias principales para generar este conjunto de reglas de asociación a partir del conjunto de calificaciones de los usuarios.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
John	-	-	4	5	3	4
Lucy	2	4	5	-	-	5
Eric	-	3	4	-	5	4
Diane	1	4	-	3	4	-

Tabla 3.1: Ejemplo de conjunto de entrenamiento

En la Tabla 3.1, se presenta un ejemplo de conjunto de calificaciones entre 1 y 5 de 4 usuarios para 6 ítems. Utilizaremos esta tabla como ejemplo para explicar las dos estrategias de generación de reglas de asociación.

2.2.1. Estrategia de filtrado de preferencias del usuario

La estrategia de filtrado de preferencias del usuario (Fig. 3.1, “Method 1”) se enfoca en identificar los ítems relevantes para cada usuario al filtrar los ítems vistos por ellos. La idea subyacente consiste en considerar exclusivamente los ítems que han recibido valoraciones positivas o altas calificaciones por parte de los usuarios.

Tomando como ejemplo el conjunto de calificaciones presentado en la Tabla 3.1, se consideraron los elementos con calificaciones superiores a 3 como relevantes. Al aplicar esta estrategia a este conjunto de datos, se generan transacciones individuales para cada usuario, tal como se muestra en la Tabla 3.2. Cada transacción representa los ítems relevantes para un usuario en particular.

TID	Transacciones
1	{Item 3, Item 4, Item 5, Item 6}
2	{Item 2, Item 3, Item 6}
3	{Item 2, Item 3, Item 5, Item 6}
4	{Item 2, Item 4, Item 5}

Tabla 3.2: Ejemplo de transacciones generadas con estrategia de extracción de los elementos relevantes

Esta estrategia permite focalizarse en los ítems que han tenido un impacto positivo en la experiencia de los usuarios, lo cual contribuye a la generación de reglas de

asociación más sólidas y relevantes. Además, al utilizar exclusivamente los ítems relevantes en las transacciones, se evita la inclusión de elementos menos significativos que podrían diluir los patrones de preferencia de los usuarios.

2.2.2. Estrategia de imputación de las calificaciones faltantes

La estrategia de imputación de calificaciones faltantes (Fig. 3.1, “Method 2”) se basa en el uso del mismo algoritmo interno utilizado para la generación de recomendaciones, con el fin de completar la matriz usuario-ítem.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
John	4.6	4.3	4	5	3	4
Lucy	2	4	5	2.1	4.5	5
Eric	4.7	3	4	1.6	5	4
Diane	1	4	3.5	3	4	4.7

Tabla 3.3: Ejemplo de la matriz de calificaciones con predicciones de las calificaciones faltantes

Una vez generada esta matriz completa de calificaciones, se seleccionan los mejores D elementos para cada usuario, incluyendo aquellos que ya han sido vistos anteriormente. En el ejemplo dado, se consideró seleccionar 4 elementos de cada conjunto de elementos ($D = 4$). Al aplicar esta estrategia al conjunto de datos mostrado en la Tabla 3.1, obtenemos las predicciones que se muestran en la Tabla 3.3, se generan las transacciones correspondientes en la Tabla 3.4.

TID	Transacciones
1	{Item 1, Item 2, Item 3, Item 4}
2	{Item 2, Item 3, Item 5, Item 6}
3	{Item 1, Item 2, Item 5, Item 6}
4	{Item 2, Item 3, Item 5, Item 6}

Tabla 3.4: Ejemplo de transacciones generadas con estrategia de imputación de las calificaciones faltantes

Esta estrategia tiene como objetivo generar una lista de ítems para cada usuario que incluya tanto los ítems valorados positivamente como los ítems recomendados en base a las predicciones del algoritmo de recomendación grupal.

Al contar con reglas que relacionan los ítems valorados positivamente (preferencias) con los ítems que han recibido altas predicciones (recomendaciones), se puede proporcionar una justificación transparente y comprensible a los usuarios sobre las razones detrás de ciertas recomendaciones. Las reglas de asociación actúan como un enlace explicativo entre las preferencias conocidas del usuario y las recomendaciones sugeridas, lo cual aumenta la confianza del usuario en el sistema de recomendación y mejora la comprensión de las decisiones subyacentes.

2.2.3. Generación de explicaciones

Una vez que se han generado las transacciones, se pueden utilizar como entrada para el algoritmo Apriori para generar las reglas de asociación (Etapa E3 en Fig. 3.1). Estas reglas se caracterizan por su soporte y confianza.

El soporte se refiere a la frecuencia con la que una regla de asociación aparece en las transacciones del conjunto de datos. Por ejemplo, si se ha generado una regla de asociación que indica que los usuarios que compran productos A y B también tienden a comprar el producto C, el soporte de esta regla se refiere a la proporción de usuarios en el conjunto de datos que compran los productos A, B y C juntos.

Por otro lado, la confianza se refiere a la probabilidad de que un usuario compre el producto C dado que ya ha comprado los productos A y B. En otras palabras, si la confianza es alta, significa que es probable que la regla de asociación sea verdadera para una gran proporción de usuarios en el conjunto de datos.

Con el fin de brindar una explicación (Algoritmo 3.3) a cada miembro del grupo sobre la recomendación del ítem I, se aplica un filtro al conjunto de reglas $X \Rightarrow Y$, donde X representa el antecedente y Y el consecuente. De esta manera, para cada miembro del grupo, se seleccionan las reglas en las que el consecuente es igual a I y todos los ítems que componen el antecedente (X) han recibido una valoración positiva por parte del respectivo miembro (Algoritmo 3.3, línea 6). Cada miembro recibirá una explicación personalizada según sus preferencias individuales.

En base a esta selección, se proporciona la siguiente explicación personalizada (Algoritmo 3.3, línea 9): "Te recomendamos el ítem I debido a que has expresado una valoración positiva hacia los ítems X ".

Algorithm 3.3: *Explanation Generation*

Input: Association rules

Input: Recommendations

Input: Minimum thresholds support and confidence (s_th , c_th);

```

1:  for  $g \in \text{Groups}$  do
2:      get all rules  $X \rightarrow Y$ , where support  $\geq s\_th$  AND confidence  $\geq c\_th$ 
3:      Retrieve the set  $R$  of recommended items for group  $g$ 
4:      for  $u \in g$  do
5:          for  $i \in R$  do
6:              Get all rules, where and  $Y = i$  AND all items in  $X$  are positively evaluated by member  $u$ 
7:              Order the filtered set  $Z$  of association rules by support and confidence
8:              Select the best association rule with high support and confidence
9:              Provide an explanation to member  $u$  : "We recommend you item  $i$  because you liked  $X$ ."
10:         end for
11:     end for
12: end for

```

Algoritmo 3.3. Generación de explicaciones

Este enfoque nos permite generar explicaciones personalizadas y relevantes para cada miembro del grupo. Al considerar los valores de soporte y confianza, aseguramos que las reglas seleccionadas sean representativas y confiables en el contexto de las preferencias del usuario y los datos disponibles.

3. Algoritmos utilizados

En el componente de recomendación, se pueden emplear diferentes algoritmos para generar recomendaciones grupales. Estos algoritmos se dividen en dos categorías principales: filtrado colaborativo y filtrado basado en contenido. Entre los algoritmos utilizados, se encuentran IKNN, UKNN, SVD y CB, este último utilizando la representación de los perfiles mediante TF-IDF (Term Frequency-Inverse Document Frequency) (Salton, 1988).

3.1. Filtrado colaborativo

En el contexto del filtrado colaborativo, se busca identificar patrones de similitud entre los usuarios y los elementos del sistema para generar recomendaciones personalizadas. Los algoritmos utilizados en este enfoque son IKNN, UKNN y SVD.

- ❖ **IKNN (Vecinos más cercanos de elementos):** Este algoritmo utiliza la similitud de los patrones de calificación entre los elementos para generar recomendaciones. Se basa en encontrar los elementos más cercanos a un elemento dado según las calificaciones de los usuarios. La similitud puede calcularse utilizando medidas como la distancia euclidiana o la similitud del coseno (Adomavicius & Tuzhilin, 2005).
- ❖ **UKNN (Vecinos más cercanos de usuarios):** A diferencia del IKNN, este algoritmo busca encontrar usuarios similares basados en sus patrones de calificación. Se seleccionan los usuarios más cercanos a un usuario dado y se utilizan sus calificaciones para generar recomendaciones. La similitud entre usuarios también puede calcularse mediante medidas como la distancia euclidiana o la similitud del coseno (Adomavicius & Tuzhilin, 2005).
- ❖ **SVD (Descomposición en valores singulares):** Este algoritmo busca reducir la dimensión de la matriz de calificaciones mediante una descomposición matricial. La idea es encontrar patrones ocultos de similitud entre usuarios y elementos al descomponer la matriz de calificaciones en valores singulares. Estos patrones se utilizan para generar recomendaciones personalizadas (Koren et al., 2009).

3.2. Filtrado basado en contenido

En el filtrado basado en contenido, se utilizan las características de los elementos del sistema para identificar elementos similares y generar recomendaciones en función de las preferencias del usuario. El algoritmo utilizado en este enfoque es CB (Filtrado basado en contenido).

- ❖ **CB (Filtrado basado en contenido):** Este algoritmo utiliza las características de los elementos del sistema, como etiquetas, palabras clave o atributos, para identificar elementos similares. Se calcula la similitud entre los perfiles de los elementos y el perfil del grupo utilizando medidas como la similitud del coseno. A partir de la similitud calculada, se generan recomendaciones para el grupo.

Para construir los perfiles de los elementos utilizando TF-IDF, se han seguido los siguientes pasos:

1. Preprocesamiento de texto: Antes de calcular TF-IDF, es necesario realizar un preprocesamiento de texto en los elementos. Esto puede incluir eliminar palabras irrelevantes (stop words), aplicar lematización o stemming, y eliminar caracteres especiales o puntuaciones.
2. Construcción de la matriz de términos: Construir una matriz donde las filas representen los elementos y las columnas representan los términos o características del sistema. Cada celda de la matriz almacenará el conteo de ocurrencias (frecuencia de término) de un término en un elemento.
3. Cálculo de TF (Term Frequency): Calcular la frecuencia de término (TF) para cada término en cada elemento. La frecuencia de término se calcula dividiendo el número de veces que un término aparece en un elemento por el número total de términos en ese elemento (Ec. 8).

$$TF = (\text{Number of term occurrences}) / (\text{Total number of terms}) \quad (8)$$

4. Cálculo de IDF (Inverse Document Frequency): Calcular la frecuencia inversa de documento (IDF) para cada término en función de todos los elementos. La IDF se calcula tomando el logaritmo del inverso de la proporción de elementos que contienen el término en toda la colección (Ec. 9).

$$IDF = \log(\text{Total number of elements} / \text{Number of elements containing the term}) \quad (9)$$

5. Cálculo de TF-IDF: Calcular el TF-IDF para cada término en cada elemento multiplicando el valor de TF por el valor de IDF correspondiente al término (Ec. 10).

$$TF_IDF = TF \times IDF \quad (10)$$

6. Perfil de elemento: El perfil de un elemento se representa mediante un vector de características donde cada entrada es el valor de TF-IDF para un término.

Para construir los perfiles de los usuarios utilizando la ponderación de los elementos evaluados por la calificación dada por dicho usuario, se siguen los siguientes pasos:

7. Crear una matriz de calificaciones: Construir una matriz donde las filas representan los usuarios y las columnas representan los elementos. Cada celda de la matriz almacenará la calificación dada por un usuario a un elemento específico.
8. Ponderación de las calificaciones: Aplicar una ponderación a las calificaciones para reflejar el nivel de preferencia del usuario. Esto se logra multiplicando las calificaciones por un factor de ponderación adecuado que refleje la escala de preferencia utilizada. Por ejemplo, si se utiliza una escala de calificación de 1 a 5, donde 5 indica una preferencia alta y 1 indica una preferencia baja, se puede multiplicar cada calificación por el factor de ponderación correspondiente.
9. Perfil de usuario: El perfil de un usuario se construye como la suma ponderada de los perfiles de elementos calificados por el usuario, utilizando las calificaciones dadas por el mismo. Para cada característica del perfil del usuario, se calcula como la suma ponderada de los perfiles de elementos.

A continuación se presenta la fórmula para calcular el perfil del usuario (Ec. 11):

$$P_u = \sum_{i \in I_u} r_{ui} \times P_i \quad (11)$$

Donde

P_u	:	Perfil del usuario u
P_i	:	Perfil del ítem i
r_{ui}	:	Calificación del usuario u para el ítem i
I_u	:	Set de ítems calificados por el usuario u

3.3. Agregación de preferencias

En el caso de los algoritmos de filtrado colaborativo utilizados en la estrategia de agregación de preferencias, se agregan las calificaciones de los usuarios. Supongamos que tenemos un grupo de usuarios que han calificado películas en una escala del 1 al 5. Para recomendar una película a todo el grupo, se pueden sumar las calificaciones individuales de cada usuario para obtener una puntuación de grupo. Por ejemplo, si los usuarios A, B y C calificaron una película con 4, 3 y 5 respectivamente, la puntuación de grupo para esa película sería 12 (4 + 3 + 5) (De Pessemier et al., 2014).

En el caso del algoritmo de filtrado basado en contenido utilizado en la estrategia de agregación de preferencias, se agregan las características de los vectores de usuarios generados mediante TF-IDF. Supongamos que cada usuario tiene un vector TF-IDF que

representa sus preferencias musicales. Para recomendar una canción al grupo, se pueden sumar los vectores de usuarios utilizando la técnica de suma de vectores. Esto proporcionará un vector de grupo que representa las preferencias combinadas de los usuarios.

3.4. Agregación de recomendaciones

En el caso de los algoritmos de filtrado colaborativo utilizados en la estrategia de agregación de recomendaciones, se agregan las predicciones generadas para cada usuario individual. Por ejemplo, supongamos que se utilizan los algoritmos IKNN, UKNN y SVD para generar predicciones de calificación para diferentes películas. Para recomendar una película al grupo, se pueden sumar las predicciones individuales para cada película y calcular la puntuación de grupo. La película con la puntuación más alta se recomendará al grupo.

En el caso del algoritmo basado en contenido (CB) utilizado en la estrategia de agregación de recomendaciones, se calculan las similitudes entre cada perfil de miembro del grupo y los perfiles de los elementos, se pueden aplicar diferentes técnicas de agregación. Una opción común es sumar las similitudes obtenidas para cada elemento y seleccionar el elemento con la similitud más alta como recomendación para el grupo. Esto significa que se consideran todas las similitudes calculadas entre los perfiles de los miembros del grupo y los perfiles de los elementos, y se selecciona aquel elemento que tenga la similitud más alta con los perfiles del grupo.

4. Conclusión

En conclusión, este capítulo ha presentado el diseño y desarrollo de un modelo de explicabilidad para un sistema de recomendación grupal. El objetivo fundamental del modelo es proporcionar recomendaciones relevantes y personalizadas a los grupos de usuarios, al tiempo que ofrece explicaciones claras y comprensibles sobre el proceso detrás de cada recomendación.

El modelo consta de dos componentes principales: el motor de recomendación grupal y el generador de explicaciones de recomendación. El motor de recomendación utiliza estrategias de agregación para analizar las preferencias del grupo y generar recomendaciones que satisfagan los intereses de todos los miembros. Se emplean

algoritmos de recomendación tradicionales, como el filtrado colaborativo y el enfoque basado en contenido, para generar estas recomendaciones grupales.

Por otro lado, el generador de explicaciones utiliza reglas de asociación generadas a partir de las calificaciones de los usuarios. Estas reglas permiten comprender las relaciones entre los ítems del sistema y las preferencias de los usuarios, proporcionando justificaciones claras sobre por qué se realizan ciertas recomendaciones. Se utiliza el algoritmo Apriori para generar estas reglas de asociación.

En conjunto, este modelo combina la generación de recomendaciones grupales con las explicaciones de las mismas, lo que permite a los usuarios comprender mejor y confiar en las recomendaciones recibidas. Al comprender el razonamiento detrás de las recomendaciones, los usuarios pueden tomar decisiones más informadas y personalizadas sobre qué ítems seleccionar.

EVALUACIÓN EXPERIMENTAL

CAPÍTULO 4

1. Introducción

En el presente capítulo, se abordan los aspectos clave de la evaluación experimental realizada para analizar y validar el rendimiento del modelo de explicabilidad propuesto en el contexto de sistemas de recomendación grupales. Se presentan en detalle los experimentos llevados a cabo, junto con la descripción del dataset utilizado y las métricas de evaluación utilizadas.

En primer lugar, se describe el dataset utilizado en el estudio. La elección de un dataset apropiado es esencial para reflejar las características y peculiaridades del dominio de aplicación. La selección de un conjunto de datos relevante y representativo permitirá obtener conclusiones más sólidas y transferibles a situaciones reales. A continuación, se presentan las métricas de evaluación utilizadas para medir el desempeño del modelo de explicabilidad propuesto. Estas métricas son cruciales para evaluar y comparar los resultados obtenidos en los experimentos. Se seleccionaron métricas adecuadas que reflejan la calidad de las recomendaciones grupales y permiten una evaluación objetiva y cuantitativa del modelo.

Finalmente, se detallan los experimentos diseñados para evaluar el impacto de los parámetros en el rendimiento del sistema de recomendación grupal y en la calidad de las explicaciones proporcionadas. Además, se mostrarán los resultados de dichos experimentos y se realizará un análisis correspondiente. Durante estos experimentos, se explorará cómo varían las métricas de evaluación al modificar estos parámetros, lo que permitirá comprender mejor la influencia de cada uno en el desempeño global del modelo.

Los experimentos se llevaron a cabo de manera sistemática, ajustando uno o varios parámetros a la vez y registrando los resultados de evaluación correspondientes. Se analizarán diferentes configuraciones y combinaciones de parámetros, con el objetivo de identificar aquellos que tienen un impacto significativo en la calidad de las recomendaciones y en la claridad de las explicaciones. Este enfoque experimental nos permitirá obtener información valiosa sobre la sensibilidad del modelo a los parámetros y proporcionará orientación para la optimización y ajuste del sistema de recomendación grupal. Además, nos ayudará a comprender cómo el modelo de explicabilidad puede adaptarse y mejorar en función de las necesidades y preferencias de los usuarios.

2. Configuración experimental

En este estudio, se utilizó el conjunto de datos ampliamente reconocido y utilizado en el campo de los sistemas de recomendación denominado "MovieLens 100K" (Harper et al., 2015) para llevar a cabo la evaluación de los sistemas de recomendación. Este conjunto de datos proporciona información detallada sobre las preferencias de los usuarios en relación con películas, lo que lo convierte en un recurso relevante y adecuado para evaluar el rendimiento del modelo de explicabilidad propuesto.

El conjunto de datos MovieLens 100K consta de una amplia colección de reseñas y valoraciones de películas realizadas por diversos usuarios. Con alrededor de 100,000 valoraciones de 943 usuarios distintos y más de 1,682 películas diferentes, este conjunto de datos proporciona una muestra significativa y diversa. Cada valoración en el conjunto de datos incluye un identificador de usuario, un identificador de película, una puntuación numérica y una marca de tiempo que indica cuándo se realizó la valoración.

Además de las valoraciones, el conjunto de datos también proporciona información adicional sobre las películas, como género, título y año de lanzamiento. Estos detalles complementarios resultan valiosos para comprender las preferencias de los usuarios y mejorar la calidad de las recomendaciones generadas por el modelo de explicabilidad.

La elección de utilizar el conjunto de datos MovieLens 100K se basa en su amplia disponibilidad, su tamaño lo suficientemente grande para realizar evaluaciones significativas y su representatividad en términos de las preferencias de los usuarios y la diversidad de películas. Al utilizar un conjunto de datos comúnmente utilizado en la comunidad de investigación, se facilita la comparación y replicación de los resultados obtenidos en este estudio, lo que contribuye a la validez y confiabilidad de las conclusiones alcanzadas.

Antes de calcular las recomendaciones, se realiza una transformación en el conjunto de datos para estimar de manera óptima las preferencias de los usuarios (Ec. 12). Esta transformación implica la normalización de las calificaciones, donde para cada calificación individual (r) de un usuario, se resta la calificación media del usuario (μ) y se divide la diferencia resultante por la desviación estándar de las calificaciones del usuario (σ).

$$r_{norm} = \frac{r - \mu}{\sigma} \quad (12)$$

Donde

- r : Calificación del usuario
- μ : Calificación media del usuario
- σ : desviación estándar de las calificaciones del usuario
- r_{norm} : Calificación normalizada

Esta normalización es necesaria para compensar las variaciones en las calificaciones de los usuarios y tener una perspectiva equilibrada de las preferencias. Teniendo en cuenta que algunos usuarios pueden ser más entusiastas y dar calificaciones principalmente positivas, mientras que otros pueden ser más críticos y dar calificaciones negativas con mayor frecuencia.

Es importante destacar que algunas métricas de similitud, como la correlación de Pearson, tienen en cuenta las diferencias en la interpretación de la escala de calificación por parte de los usuarios. En estos casos, el proceso de normalización no es necesario para calcular las similitudes entre usuarios. Sin embargo, cuando se agregan las calificaciones de los miembros del grupo para calcular una calificación grupal, la normalización sigue siendo significativa (Masthoff, 2004).

3. Método de evaluación

Cuando se trata de evaluar sistemas de recomendación grupal, utilizamos un enfoque ligeramente diferente al método tradicional para sistemas de recomendación individual. En primer lugar, creamos grupos artificiales seleccionando usuarios al azar y asignándoles a grupos de tamaño predefinido. Esto nos permite simular situaciones en las que un sistema de recomendación debe generar recomendaciones para un conjunto de usuarios en lugar de uno solo.

A continuación, generamos recomendaciones grupales para cada uno de estos grupos utilizando la información de calificaciones proporcionada por los miembros del grupo en el conjunto de entrenamiento.

El objetivo principal de estas recomendaciones grupales es adaptarse a las preferencias colectivas de todos los miembros del grupo. Es importante destacar que todos los miembros del grupo reciben la misma lista de recomendaciones, fomentando la colaboración y la toma de decisiones conjuntas.

Sin embargo, surge un desafío al evaluar el rendimiento de los sistemas de recomendación grupales. Dado que los conjuntos de prueba individuales de los miembros del grupo pueden contener elementos diferentes o incluso no contener elementos de la lista de recomendación grupal. Para abordar este problema, proponemos una solución que implica proyectar la lista de recomendación grupal al conjunto de prueba de los miembros, tal y como lo hace Baltrunas et al. (2010). En otras palabras, seleccionamos solo los elementos relevantes de la lista de recomendación grupal para cada miembro, teniendo en cuenta su historial de preferencias y características personales. Luego, calculamos métricas de evaluación, como precisión, recall, etc., utilizando estas listas, lo que permite una evaluación individualizada y justa para cada miembro del grupo.

Por ejemplo, supongamos que tenemos un grupo de tres personas (A, B, C) y el sistema de recomendación grupal genera la siguiente lista de recomendaciones: [Item1, Item2, Item3, Item4, Item5]. Sin embargo, cada miembro del grupo tiene diferentes preferencias y solo tiene acceso a ciertos elementos en su conjunto de prueba, que pueden ser diferentes o incluso no contener elementos de la lista grupal. En este caso, realizamos la proyección de la lista grupal a las listas individuales. Si el miembro A tiene en su conjunto de prueba los elementos [Item1, Item2, Item3], entonces la lista de recomendaciones resultante para el miembro A sería [Item1, Item2, Item3]. Realizamos la proyección de manera similar para los miembros B y C, teniendo en cuenta sus respectivos conjuntos de prueba.

Una vez que se han formado las listas de recomendaciones individuales, podemos calcular las métricas de evaluación utilizando estas listas. Esto proporciona una evaluación más precisa y justa del sistema de recomendación grupal, ya que se considera la disponibilidad de elementos en los conjuntos de prueba de cada miembro. En el resto de esta sección, discutiremos las métricas de calidad que se utilizaron para evaluar las recomendaciones grupales.

3.1. Ganancia Acumulada Descontada Normalizada (nDGC)

La Ganancia Acumulada Descontada Normalizada (nDCG, por sus siglas en inglés) es una métrica utilizada para evaluar la calidad del ranking de los elementos recomendados en sistemas de recomendación. Se basa en la relevancia y posición de los elementos en la lista recomendada (Järvelin & Kekäläinen, 2002).

El propósito del nDCG es medir qué tan bien el algoritmo de recomendación coloca los elementos más relevantes en las posiciones superiores de la lista recomendada. Cuanto más alto sea el nDCG, mejor será el ranking y, por lo tanto, más relevantes serán los elementos recomendados.

La fórmula para calcular el nDCG de una lista de recomendación, denotada como $rec = [c1, c2, c3, c4, \dots, cn]$, se define de la siguiente manera:

$$DCG_n^u = r_{uc1} + \sum_{i=2}^n \frac{r_{uci}}{\log_2(i)} \quad (13)$$

$$nDCG_n^u = DCG / IDC G \quad (14)$$

Donde	$nDCG$:	Ganancia Acumulada Descontada Normalizada
	DCG :	Ganancia acumulada descontada
	$IDCG$:	Ganancia acumulada descontada ideal
	n :	Longitud de la lista de recomendación
	r_{uci} :	Calificación del usuario u al ítem ci en la posición i
	u :	Usuario u

El nDCG se calcula a partir del DCG y el IDC G. El DCG se obtiene sumando las relevancias de los elementos en la lista recomendada, teniendo en cuenta la posición de cada elemento (Ec. 13). Se descuentan los valores según su posición para reflejar que los elementos más relevantes deben tener un mayor impacto en la evaluación.

Por otro lado, el IDC G representa la ganancia acumulada descontada ideal y se calcula considerando un ranking ideal en el que los elementos más relevantes ocupan las posiciones superiores. Proporciona un punto de referencia máximo para el DCG. Finalmente, el nDCG se obtiene dividiendo el DCG por el IDC G (Ec. 14), generando un valor entre 0 y 1. Un nDCG igual a 1 indica un ranking perfecto donde los elementos más relevantes se encuentran en las posiciones superiores.

3.2. F1-Score

El F1-Score es una métrica comúnmente utilizada en sistemas de recomendación para evaluar la precisión y el equilibrio entre la precisión y el recuerdo de un modelo de recomendación. Se basa en la combinación de dos métricas: precisión (precision) y recuerdo (recall) (Goutte & Gaussier, 2005).

La precisión mide la proporción de elementos relevantes (recomendados correctamente) sobre el total de elementos recomendados. Por otro lado, la recuperación mide la proporción de elementos relevantes recomendados correctamente sobre el total de elementos relevantes en el conjunto de datos. A continuación, en la Tabla 4.1 se presenta una matriz de confusión que ayuda a visualizar estos conceptos:

	Recomendado	No recomendado
Relevantes	TP	FN
Irrelevantes	FP	TN

Tabla 4.1. Matriz de confusión para sistemas de recomendación

Donde TP : Número de elementos relevantes recomendados
 FP : Número de elementos irrelevantes recomendados
 TN : Número de elementos irrelevantes no recomendados
 FN : Número de elementos relevantes no recomendados

La fórmula para calcular la precisión (Precisión) (Ec. 15) y el recuperación (Recall) (Ec. 16) es la siguiente:

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

A partir de la precisión y la recuperación, se puede calcular el F1-Score, que es una medida de la media armónica entre ambos valores. El F1-Score se calcula de la siguiente manera (Ec. 17):

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (17)$$

El F1-Score proporciona una medida equilibrada entre la precisión y la recuperación, ya que tiene en cuenta tanto los elementos relevantes recomendados correctamente como los elementos relevantes no recomendados. Un valor de F1-Score cercano a 1 indica un equilibrio entre precisión y recuperación, lo cual es deseable en un sistema de recomendación. En nuestro estudio, hemos considerado un elemento relevante para el usuario si su calificación es superior o igual a 4.

3.3. Fidelidad

El objetivo principal de este proyecto consiste en desarrollar un modelo de recomendación grupal altamente fidedigno y explicativo. Con el fin de evaluar la calidad de nuestro modelo de explicación, hemos establecido el concepto de "Fidelidad del Modelo", que nos permite medir qué tan fielmente reproduce los resultados del complejo modelo de recomendación. Para ilustrar estas definiciones, hemos ampliado la tradicional matriz de confusión incorporando una variable explicativa, tal como se muestra en la Tabla 4.2.

	Explicable		Inexplicable	
	Recomendado	No recomendado	Recomendado	No recomendado
Relevantes	A	B	C	D
Irrelevantes	E	F	G	H

Tabla 4.2. Matriz de confusión extendida

La Fidelidad del Modelo se convierte en nuestra métrica central de éxito en este proyecto, ya que consideramos que el éxito de un modelo de explicación debe ser independiente del éxito predictivo del modelo que está siendo explicado. En este sentido, hemos adaptado las métricas de recuperación de información propuestas por Abdollahi y Nasraoui (2016) a nuestro escenario de interpretabilidad post hoc.

En este contexto, consideramos el modelo de reglas de asociación como nuestro sistema de recuperación y, por lo tanto, los elementos encontrados por las reglas de asociación son considerados elementos explicables. Por otro lado, las recomendaciones generadas por el modelo complejo son consideradas como los elementos recomendados. El conjunto de resultados está definido como la intersección de los elementos explicables y los elementos recomendados, es decir, aquellas recomendaciones que pueden ser explicadas por el modelo de reglas de asociación.

$$Fidelity = \frac{Recommended\ items \cap Explainable\ items}{Recommended\ items} = \frac{A + C}{A + C + E + G} \quad (18)$$

De esta manera, la Fidelidad se define como la proporción de recomendaciones que pueden ser explicadas por el modelo de reglas de asociación (Ec. 18). Esta medida

nos permite evaluar qué tan bien nuestro modelo de explicación se ajusta a los resultados del modelo de recomendación complejo (Peake & Wang, 2018).

4. Experimentos y resultados

En esta sección, llevaremos a cabo una serie de experimentos para evaluar el rendimiento del sistema de recomendación grupal y analizar el impacto de varios parámetros en dicho rendimiento. Nuestra meta es comprender cómo factores como el tamaño del grupo, el tamaño de la lista de recomendaciones y la similitud entre las preferencias de los miembros del grupo afectan la efectividad del sistema.

Además de evaluar el rendimiento del sistema de recomendación grupal, también nos enfocaremos en evaluar la explicabilidad y confiabilidad de las recomendaciones. Analizaremos cómo los parámetros mencionados anteriormente influyen en la capacidad del sistema para proporcionar explicaciones claras y comprensibles para las recomendaciones realizadas. Para llevar a cabo los experimentos, hemos dividido el conjunto de datos en un 70% para entrenamiento y un 30% para pruebas. En cada experimento, hemos repetido el proceso 30 veces, realizando una reordenación aleatoria de los miembros de los grupos para probar diferentes combinaciones.

4.1. Evaluación del rendimiento del sistema

En esta sección, se amplían los resultados de la evaluación de las recomendaciones grupales generadas por diferentes algoritmos. El objetivo principal de esta evaluación es analizar la calidad de las recomendaciones desde múltiples perspectivas, utilizando diversas métricas de calidad como F1-Score y nDGC.

En primer lugar, se examina detalladamente la influencia del método de agregación en estas métricas de calidad. Se investiga cómo diferentes métodos de agregación afectan el rendimiento de las recomendaciones y cómo se refleja en las métricas evaluadas.

A continuación, se lleva a cabo una evaluación de las recomendaciones para grupos de distintos tamaños y con diferentes longitudes de lista de recomendaciones. Esto implica analizar el desempeño de los algoritmos en la generación de recomendaciones tanto para grupos pequeños como para grupos grandes. Además, se

consideran diferentes configuraciones de longitud de lista de recomendaciones para observar cómo afecta la calidad de las recomendaciones.

En este análisis, se incluyen grupos formados al azar, donde los miembros del grupo tienen preferencias diversas y no necesariamente similares. También se evalúan grupos con preferencias similares, lo que significa que los miembros del grupo tienen gustos y preferencias en común. Esto permite examinar cómo los algoritmos se desempeñan en la generación de recomendaciones para diferentes tipos de grupos y cómo las métricas de calidad se ven afectadas en cada caso.

En el análisis realizado, se asume que al evaluar el impacto de un parámetro en las métricas de calidad de las recomendaciones, los demás parámetros se mantienen fijos. Esto significa que se realiza una evaluación individual de cada parámetro, manteniendo constantes los valores de los demás parámetros para aislar el efecto de cada uno de ellos en el rendimiento de los algoritmos de recomendación.

4.1.1. Influencia del método de agregación

Como se mencionó previamente en el apartado 3.2 del capítulo 2, el método de agregación de datos es una función matemática que determina cómo se combinan las listas de recomendaciones individuales de los miembros del grupo para formar recomendaciones grupales en el caso de la estrategia de agregación de recomendaciones, o cómo se combinan las preferencias individuales de los miembros del grupo para formar una preferencia grupal en el caso de la estrategia de agregación de preferencias.

En el caso de la estrategia de agregación de recomendación, se utiliza un algoritmo de recomendación tradicional (como se discutió en el apartado 2.1 del capítulo 3) para calcular una predicción de calificación de usuario para cada ítem en el sistema y para cada usuario del grupo. Estas predicciones se utilizan para generar recomendaciones individuales ordenando los elementos de contenido en función de las calificaciones predichas. Para obtener recomendaciones grupales, las recomendaciones individuales de los miembros del grupo se combinan mediante el método de agregación de datos, que implica combinar los valores de predicción de cada lista de recomendaciones. Posteriormente, los elementos recomendados se ordenan en función del valor de

predicción agregado y se seleccionan los mejores N elementos para formar la lista de recomendaciones grupales.

En el caso de la estrategia de agregación de preferencias, las preferencias individuales de los miembros del grupo se combinan para formar una preferencia grupal mediante el método de agregación de datos. Esto implica combinar las calificaciones de los miembros para cada elemento y utilizar el resultado agregado como calificación grupal. A partir de estas calificaciones grupales, se generan las recomendaciones de grupo utilizando un algoritmo de recomendación tradicional. Una vez más, solo se seleccionan las mejores N recomendaciones para el grupo.

Un factor determinante en el proceso de selección del método de agregación puede ser la calidad resultante de las recomendaciones grupales. Por lo tanto, se investiga la influencia del método de agregación en la precisión de las recomendaciones grupales mediante la comparación de los siguientes cinco métodos de agregación propuestos en la literatura (Masthoff, 2004).

- ❖ **Average (AVG):** En este método, se calcula el promedio de las calificaciones o valores de predicción de las recomendaciones individuales. En el caso de la estrategia de agregación de recomendaciones, se promedian las calificaciones predichas para cada elemento recomendado. En la estrategia de agregación de preferencias, se promedian las calificaciones individuales de los miembros del grupo para cada elemento. Luego, se ordenan los elementos en función del promedio y se seleccionan los mejores N elementos para formar la lista de recomendaciones grupales.
- ❖ **Median (MED):** La mediana (*median*) es un método de agregación que difiere del promedio (*average*) en su propósito y en la forma en que se calcula. Mientras que el promedio calcula la suma de los valores y luego los divide por la cantidad de elementos, la mediana se utiliza para encontrar el valor central en un conjunto de datos ordenados. El propósito de utilizar la mediana como método de agregación en recomendaciones grupales es mitigar el impacto de los valores atípicos o extremos en las calificaciones o valores de predicción de las recomendaciones individuales.
- ❖ **Least Misery (LM):** Este método busca minimizar el sufrimiento o la insatisfacción en el grupo al seleccionar las recomendaciones grupales. Se selecciona el elemento

con la calificación o valor de predicción más bajo entre las recomendaciones individuales. En otras palabras, se elige la opción que cause menos "misericordia" a los miembros del grupo. Se ordenan los elementos en función de las calificaciones o valores de predicción y se selecciona el elemento con la calificación más baja para formar la lista de recomendaciones grupales.

- ❖ **Most Pleasure (MP):** A diferencia del método de *Least Misery*, el método de *Most Pleasure* busca maximizar el placer o la satisfacción en el grupo. Se selecciona el elemento con la calificación o valor de predicción más alto entre las recomendaciones individuales. En resumen, se elige la opción que brinde la mayor "satisfacción" a los miembros del grupo. Los elementos se ordenan en función de las calificaciones o valores de predicción, y se selecciona el elemento con la calificación más alta para formar la lista de recomendaciones grupales.
- ❖ **Random (ONE):** En este método, se selecciona aleatoriamente una de las recomendaciones individuales o preferencias para cada elemento. Esto implica que la selección de las recomendaciones grupales se realiza al azar, sin tener en cuenta las calificaciones o valores de predicción. Se pueden seleccionar aleatoriamente los elementos recomendados para formar la lista de recomendaciones grupales.

Para evaluar el impacto del método de agregación en el rendimiento del sistema de recomendación, se llevan a cabo una serie de experimentos comparando los resultados obtenidos con cada uno de estos métodos. En estos experimentos, se forman grupos seleccionando usuarios al azar, lo que implica que no se aplican restricciones adicionales ni a los grupos ni a sus miembros. En cada experimento, se estableció un tamaño de grupo fijo de 4 personas y se definió una longitud de lista de recomendaciones de 5 elementos.

Se evaluaron dos estrategias distintas para generar recomendaciones grupales: la agregación de recomendaciones y la agregación de preferencias. Esta elección permitió examinar detalladamente cómo el método de agregación afecta la calidad de las recomendaciones grupales, teniendo en cuenta diferentes configuraciones de grupo y enfoques de generación de recomendaciones. Cada experimento se repitió 30 veces y se calculó el promedio de los resultados obtenidos. De esta manera, se buscó obtener una visión más sólida y representativa de los efectos de cada método de agregación en la calidad de las recomendaciones grupales.

Al final de estos experimentos, se seleccionará un método de agregación basado en los resultados obtenidos. Esta selección será crucial para continuar con el resto de los experimentos, ya que permitirá centrar los esfuerzos en analizar y comparar de manera más exhaustiva el desempeño de los algoritmos considerando el método de agregación elegido. De esta manera, se podrá determinar con mayor precisión el impacto del método de agregación en la calidad de las recomendaciones grupales en diferentes contextos.

En las figuras 4.1 y 4.2 se presentan los resultados obtenidos en términos de F1 y nDGC respectivamente. Estos resultados se analizan en función de los algoritmos utilizados, la estrategia de agrupamiento empleada y los métodos de agregación considerados. En la serie de barras con el sufijo "Rec" se evalúan los algoritmos de recomendación en combinación con la estrategia de agregación de recomendaciones, mientras que el sufijo "Pref" se refiere a la estrategia de agregación de preferencias. Por ejemplo, la serie de barras "SVD-Pref" representa las recomendaciones grupales generadas al combinar las preferencias individuales de los miembros utilizando la estrategia de preferencias agregadas y calculando las recomendaciones para este perfil agregado utilizando el algoritmo de descomposición de valores singulares.

En las Figuras 4.1 y 4.2 se muestran el promedio de nDCG y F1, respectivamente, junto con el intervalo de confianza del 95% del promedio de nDCG. Estas gráficas representan la relación entre el algoritmo de recomendación utilizado, la estrategia de agrupamiento (Pref o Rec) y el método de agregación.

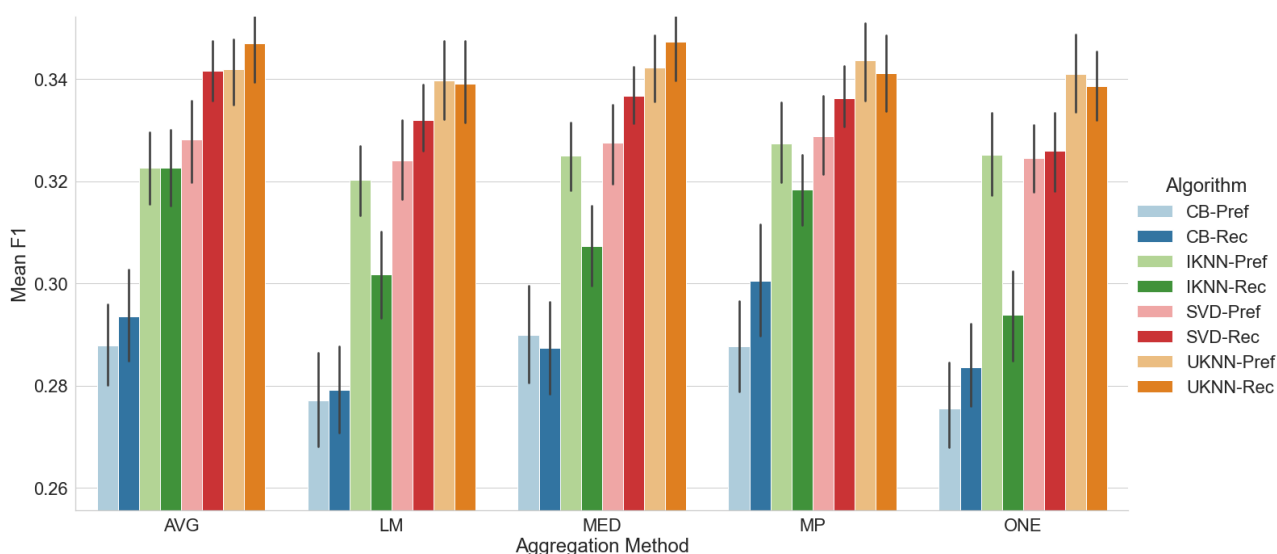


Figura 4.1: Impacto del método de agregación en la métrica F1-Score

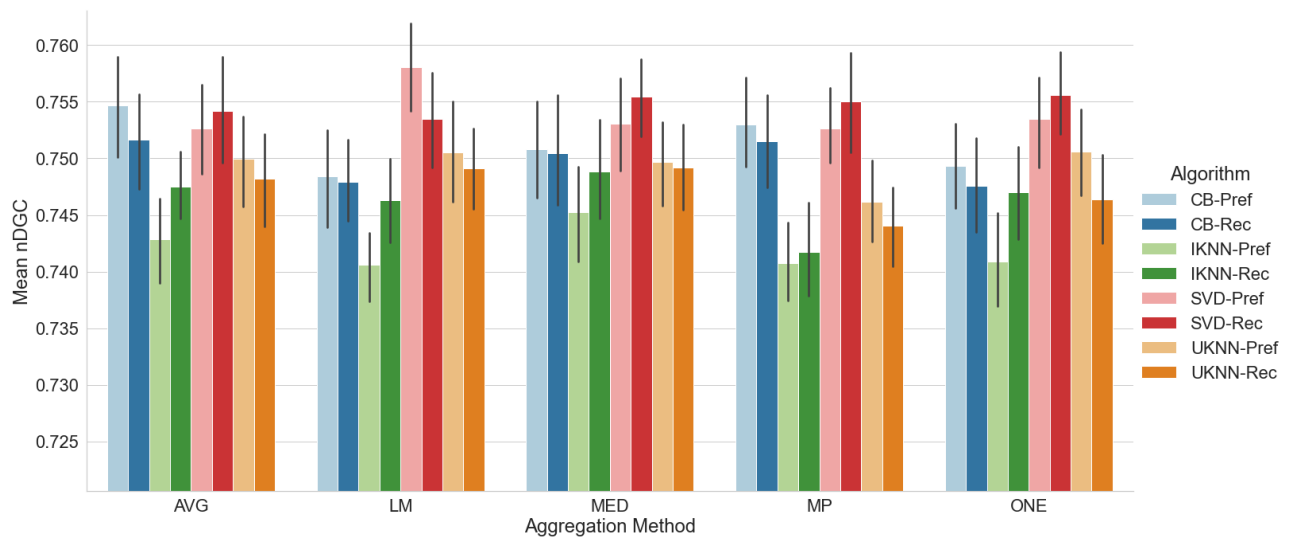


Figura 4.2: Impacto del método de agregación en la métrica nDGC

En el análisis de los resultados representados en las figuras 4.1 y 4.2, se evidencia que los valores de nDGC y F1 están fuertemente influenciados por la estrategia de agregación y el algoritmo utilizado. Dentro de este contexto, es interesante destacar el comportamiento de los algoritmos SVD-Rec y SVD-Pref, el cual muestra una menor variabilidad en ambas métricas, indicando una mayor estabilidad en su rendimiento. Esto implica que, sin importar el método de agregación empleado, los algoritmos de recomendación grupal que utilizan SVD como recomendador interno tienden a generar resultados consistentes y con una variabilidad reducida. Por otro lado, se observa que los algoritmos IKNN-Rec y IKNN-Pref presentan una mayor variabilidad en ambas métricas, evidenciando una mayor sensibilidad a los métodos de agregación utilizados, lo cual se traduce en resultados más variables según la estrategia seleccionada.

En relación a las métricas específicas, se destaca que el algoritmo IKNN-Rec muestra la mayor variabilidad en la métrica F1, lo que sugiere una mayor variabilidad en su rendimiento en comparación con los demás algoritmos analizados. Por otra parte, los algoritmos que utilizan IKNN tanto CB como recomendadores internos exhiben las mayores diferencias en la métrica nDGC, lo cual indica que estos algoritmos pueden ser más susceptibles a los métodos de agregación empleados en comparación con los demás.

En cuanto a la mejor estrategia para lograr recomendaciones grupales precisas, es importante considerar el algoritmo utilizado. Al analizar diferentes algoritmos, se observa

que AVG es la opción destacada para CB-Pref y SVD-Rec, mientras que CB-Rec muestra buenos resultados con la estrategia MP. Por otro lado, el algoritmo UKNN-Pref alcanza un rendimiento óptimo con la estrategia LM, y para IKNN-Pref se encuentra que MED es la estrategia más adecuada. Estos hallazgos subrayan la importancia de tener en cuenta el algoritmo específico al seleccionar la estrategia de agregación más adecuada para maximizar el rendimiento del sistema de recomendación grupal.

En general, se destaca que el método de agregación AVG (promedio) muestra mejores resultados en comparación con los demás métodos de agregación analizados. Esta afirmación se basa en las métricas de rendimiento observadas, donde AVG ha demostrado ser más efectivo en la generación de recomendaciones grupales precisas y relevantes.

Una posible interpretación de estos resultados es que el método AVG tiene en cuenta las preferencias individuales de los usuarios de manera equilibrada. Al promediar las calificaciones o valores de relevancia asignados por los miembros del grupo, se logra una combinación de opiniones que tiende a reflejar mejor los gustos y preferencias del grupo en su conjunto. Esto permite generar recomendaciones que se ajustan de manera más precisa a los intereses y necesidades del grupo, ya que se basan en una representación agregada de sus opiniones individuales.

Además, el método AVG también puede ayudar a minimizar el efecto de valores extremos o sesgados. Si un miembro del grupo otorga calificaciones extremas o poco representativas, el promedio suavizará ese impacto, evitando que influya excesivamente en las recomendaciones grupales. Esto contribuye a obtener recomendaciones más equilibradas y generalmente más precisas en términos de relevancia para el grupo en su conjunto.

Dado que AVG ha mostrado buenos resultados en términos de precisión y relevancia en los experimentos realizados, es razonable seleccionarlo como el método de agregación preferido para continuar con los análisis y pruebas posteriores. Utilizar AVG como estrategia de agregación en los experimentos subsiguientes permitirá evaluar y comparar otros aspectos del sistema de recomendación grupal de manera más precisa y consistente, al basarse en un método que ha demostrado ser efectivo en generar recomendaciones precisas y relevantes.

4.1.2. Influencia del tamaño de grupo

En este experimento, se analiza cómo el tamaño del grupo impacta en la calidad de las recomendaciones grupales. Se exploran diferentes tamaños de grupo, desde 1 persona por grupo (recomendaciones individuales) hasta 10 personas por grupo. Sin embargo, es importante tener en cuenta que la longitud de la lista de recomendación se ha fijado a 5 en este experimento. A diferencia del primer experimento, se utiliza el método de agregación AVG en todas las combinaciones de estrategias de agrupamiento y algoritmos de recomendación.

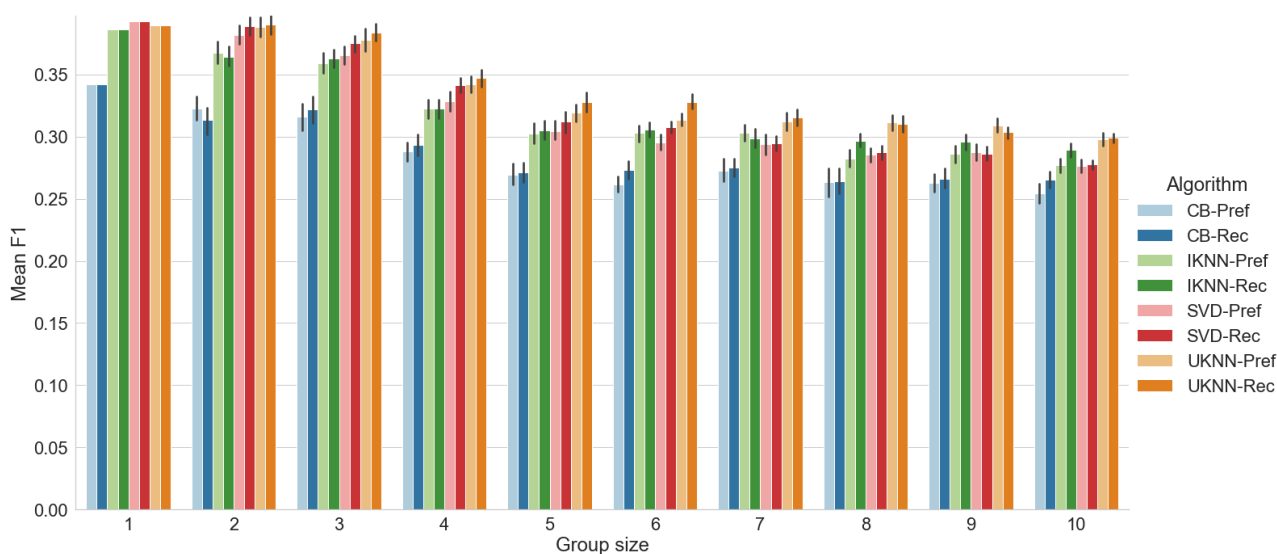


Figura 4.3: Impacto del tamaño de grupo en la métrica F1-Score

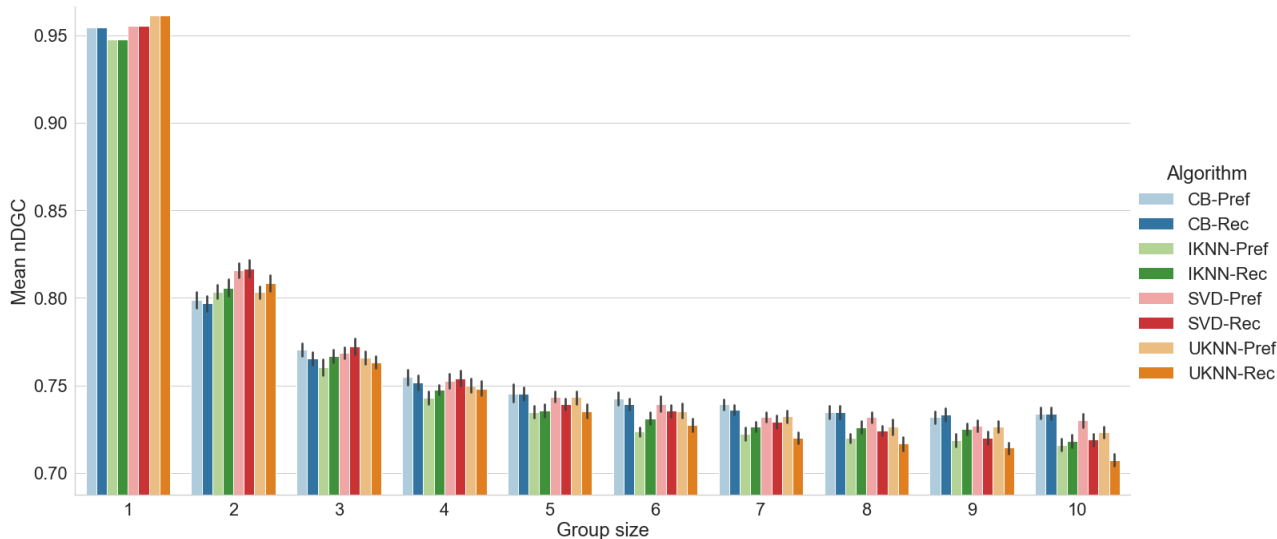


Figura 4.4: Impacto del tamaño de grupo en la métrica nDGC

El proceso de composición de grupos se realiza seleccionando usuarios aleatorios del conjunto de datos, repitiendo este proceso 30 veces. Luego, se calcula cada métrica de calidad 30 veces y se obtiene el promedio de estas mediciones como una estimación de la calidad de las recomendaciones grupales.

Los resultados se presentan en las Figuras 4.3 y 4.4, donde se muestran los promedios en el eje vertical, junto con los intervalos de confianza del 95% de los valores promedio. En el eje horizontal se indica el tamaño del grupo.

Tal como se anticipaba, en las Figuras 4.3 y 4.4 se observa que tanto el valor de F1 como el de nDGC disminuyen a medida que aumenta el tamaño del grupo. Este hallazgo es coherente con lo que se podría anticipar en un sistema de recomendación grupal. Cuando el tamaño del grupo aumenta, se vuelve más desafiante llegar a un consenso o acuerdo entre los miembros en términos de gustos, preferencias y necesidades. Esto puede generar mayor diversidad en las opiniones y dificultar la generación de recomendaciones que sean igualmente satisfactorias para todos los miembros del grupo.

Esta falta de consenso o acuerdo puede tener un impacto negativo en las métricas de calidad, como F1 y nDGC. El valor de F1, que combina precisión y exhaustividad, puede disminuir debido a la dificultad para encontrar recomendaciones que sean relevantes para todos los miembros del grupo. Por otro lado, el nDGC, que mide la utilidad acumulativa de las recomendaciones, también puede disminuir porque es más complicado alcanzar altos niveles de utilidad cuando las preferencias y necesidades de los miembros del grupo son diversas y no se pueden satisfacer completamente.

Sin embargo, el rendimiento de los algoritmos en función del aumento del tamaño del grupo, se observa un comportamiento consistente en relación a la métrica F1. A medida que el tamaño del grupo aumenta, se registra una disminución generalizada en el rendimiento de todos los algoritmos, sin embargo, se mantiene la misma tendencia relativa entre ellos. Es decir, si un algoritmo muestra un desempeño superior a otro en un grupo de tamaño reducido, es probable que esa misma comparación se mantenga al incrementar el tamaño del grupo hasta alcanzar su máximo.

Por otro lado, en lo que respecta a la métrica nDGC, se aprecia un cambio notable en el patrón de rendimiento. Inicialmente, los algoritmos basados en el paradigma de

basado en contenido exhiben un desempeño inferior en grupos pequeños. No obstante, a medida que el tamaño del grupo aumenta, los algoritmos basados en contenido superan al resto en cuanto al valor de nDGC. Estos resultados destacan la sensibilidad del valor de nDGC al tamaño del grupo, la cual depende del algoritmo utilizado.

4.1.3. Influencia de la longitud de la lista de recomendación

En este experimento, se examina cómo la longitud de la lista de recomendación influye en la calidad de las recomendaciones grupales, mientras se mantiene un tamaño de grupo constante de 4 personas. Se exploran diferentes longitudes de lista, variando desde 3 hasta 10 elementos. Como los otros experimentos, este procedimiento se repetirá 30 veces y se calculará el promedio de los valores obtenidos.

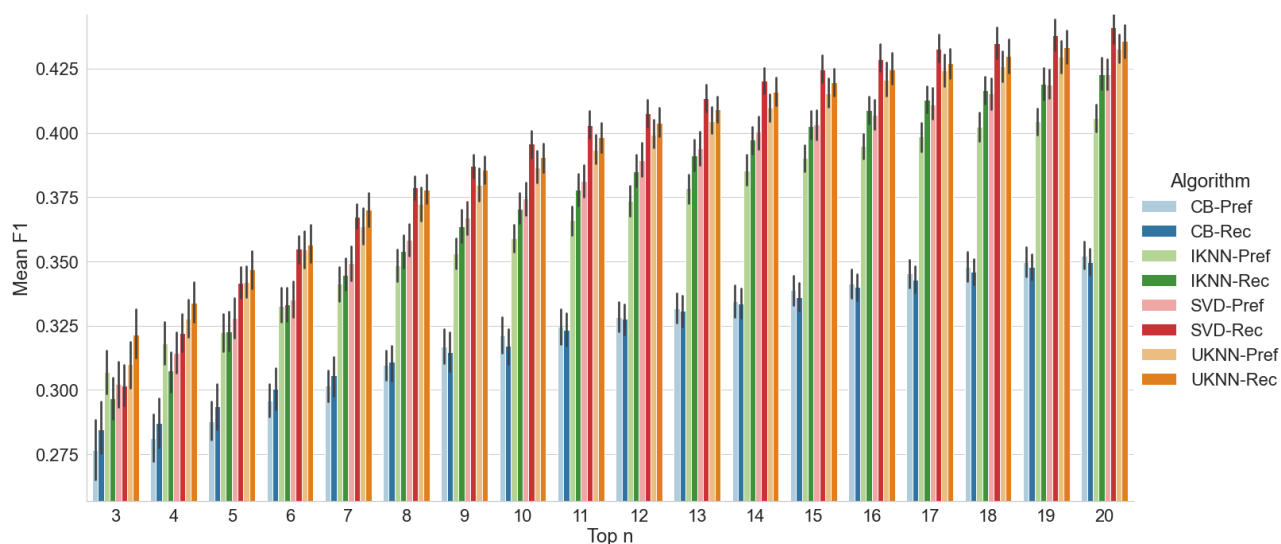


Figura 4.5: Impacto de la longitud de lista de recomendación en la métrica F1-Score

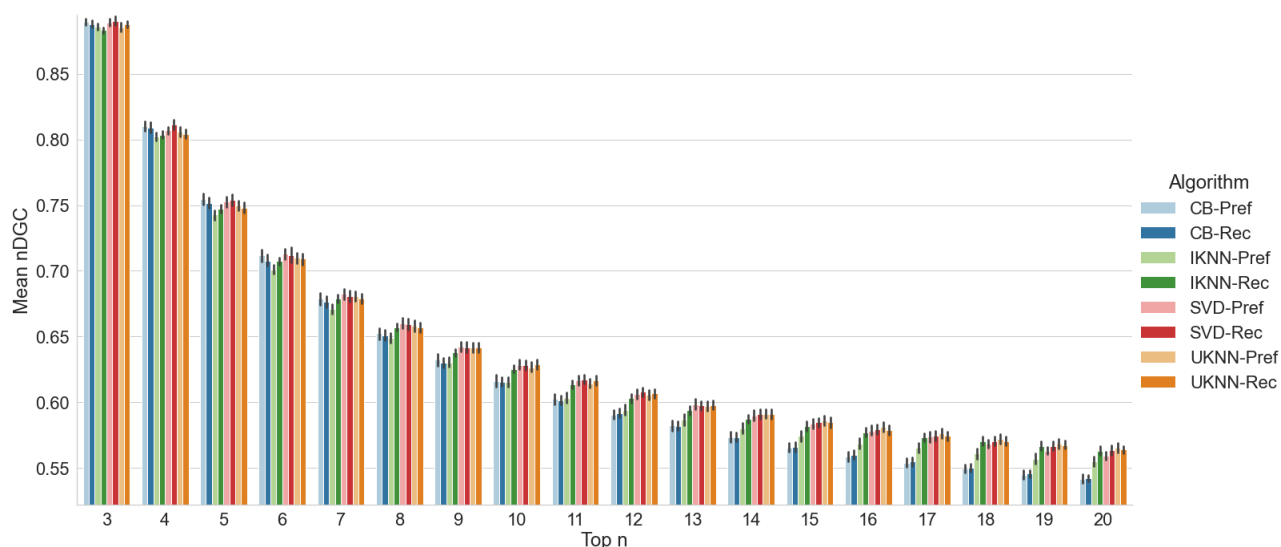


Figura 4.6: Impacto de la longitud de lista de recomendación en la métrica nDGC

Los resultados se presentan en las Figuras 4.5 y 4.6, donde se muestran los promedios en el eje vertical, junto con los intervalos de confianza del 95% de los valores promedio. En el eje horizontal se indica la longitud de la lista de recomendación.

En la Figura 4.5, se evidencia un incremento en el valor de F1 al aumentar la longitud de la lista de recomendaciones. Esto implica que al incluir más elementos en la lista, se logra mejorar tanto la precisión como la exhaustividad de las recomendaciones grupales. Este resultado puede interpretarse como una consecuencia directa de brindar más opciones a los usuarios. Al ampliar la lista de recomendaciones, se aumenta la probabilidad de incluir ítems relevantes para diferentes perfiles de usuarios. Esto resulta especialmente beneficioso cuando existe una alta diversidad en las preferencias de los miembros del grupo. Al ofrecer una gama más amplia de alternativas, se mejora la probabilidad de encontrar elementos de interés, lo que se traduce en un incremento en el valor de F1.

Por otro lado, en la Figura 4.6 se observa una disminución en el valor de nDGC a medida que la longitud de la lista de recomendación se incrementa. Esta tendencia sugiere que al agregar más elementos a la lista, se diluye la importancia de los ítems más relevantes en términos de su utilidad acumulativa. Una posible explicación de este resultado es que, al aumentar la longitud de la lista, también se incluyen elementos de menor calidad o relevancia. Esto afecta la utilidad acumulativa, ya que los ítems más relevantes se vuelven menos prominentes entre aquellos de menor relevancia. Como consecuencia, el valor de nDGC disminuye.

4.1.4. Influencia de la similitud intergrupala

Este tercer experimento tiene como objetivo investigar cómo la similitud de las preferencias entre los miembros del grupo influye en la calidad de las recomendaciones grupales. Para llevar a cabo este estudio, se formaron grupos compuestos por 4 personas con diferentes niveles de similitud entre ellas.

El procedimiento de medición consiste en establecer un umbral mínimo de similitud dentro del grupo. Este umbral representa el nivel mínimo de similitud requerido entre cada par de miembros del grupo. Es decir, cada par de usuarios dentro del mismo grupo debe tener una similitud igual o superior a este umbral mínimo. Para calcular estas similitudes

entre usuarios, se utiliza la correlación de Pearson, que se aplica a las calificaciones otorgadas por los usuarios en el conjunto de datos. La fórmula de correlación de Pearson es la siguiente entre dos usuarios (Ec. 19):

$$pearson_sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u) \times (r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2} \times \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \bar{r}_v)^2}} \quad (19)$$

Donde

- u : Primer usuario
- v : Segundo usuario
- r_{ui} : Calificación del usuario u para el ítem i
- \bar{r}_u : Promedio de las calificaciones del usuario u
- I_{uv} : Conjunto de los elementos calificados por los dos usuarios

Para formar los grupos, se sigue un proceso específico. El primer miembro del grupo se selecciona aleatoriamente, sin requerir ninguna similitud mínima. Luego, el segundo miembro se elige aleatoriamente del subconjunto de usuarios que son suficientemente similares al primer miembro, cumpliendo con la similitud mínima intragrupo establecida. Sin embargo, para garantizar que exista un nivel adecuado de coincidencia en las preferencias, además de la similitud, también se requiere que estos dos miembros tengan más de 5 elementos calificados en común.

Esta consideración adicional se basa en la idea de que la similitud por sí sola no garantiza una suficiente convergencia en las preferencias de los miembros del grupo. Al requerir que los usuarios en pares tengan más de 5 elementos comunes calificados, se busca asegurar que exista una base sólida de preferencias compartidas. Esto permite que el proceso de generación de recomendaciones tome en cuenta un conjunto significativo de elementos en común, lo que aumenta las probabilidades de generar recomendaciones satisfactorias para todos los miembros del grupo.

En la figura 4.7 se muestra la distribución de similitudes entre los usuarios en el conjunto de datos MovieLens 100K. Esta figura es importante para comprender la variabilidad en las similitudes entre los usuarios y para establecer umbrales mínimos de similitud que sean adecuados para formar grupos en los experimentos.

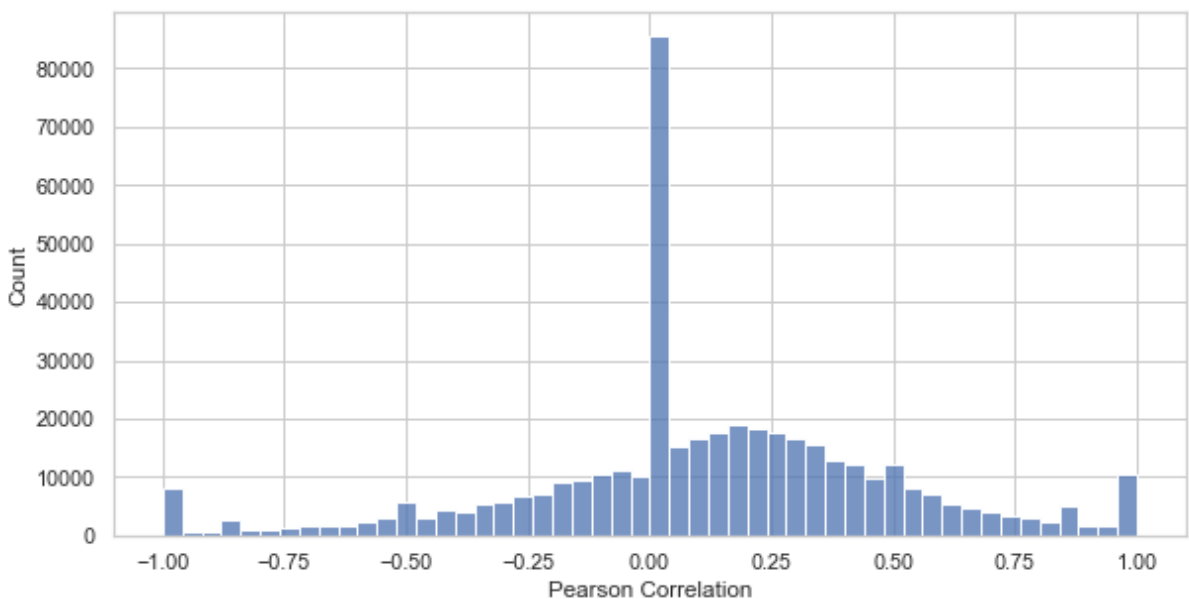


Figura 4.7: Distribución de las similitudes entre los usuarios en el conjunto de datos de MovieLens 100k

En este experimento, hemos utilizado umbrales mínimos de similitud que varían desde -0.6 hasta 0.6 en pasos de 0.1. La elección de detenernos en un umbral máximo de 0.6 nos permite asegurar que tenemos una cantidad adecuada de participantes que cumplen con los requisitos de similitud establecidos. Esto es importante para garantizar la viabilidad y representatividad de los grupos en el experimento. Se ha considerado la posibilidad de tener similitudes negativas (de -0.6 a 0) en los umbrales para poder simular situaciones en las que los miembros del grupo pueden tener preferencias en conflicto.

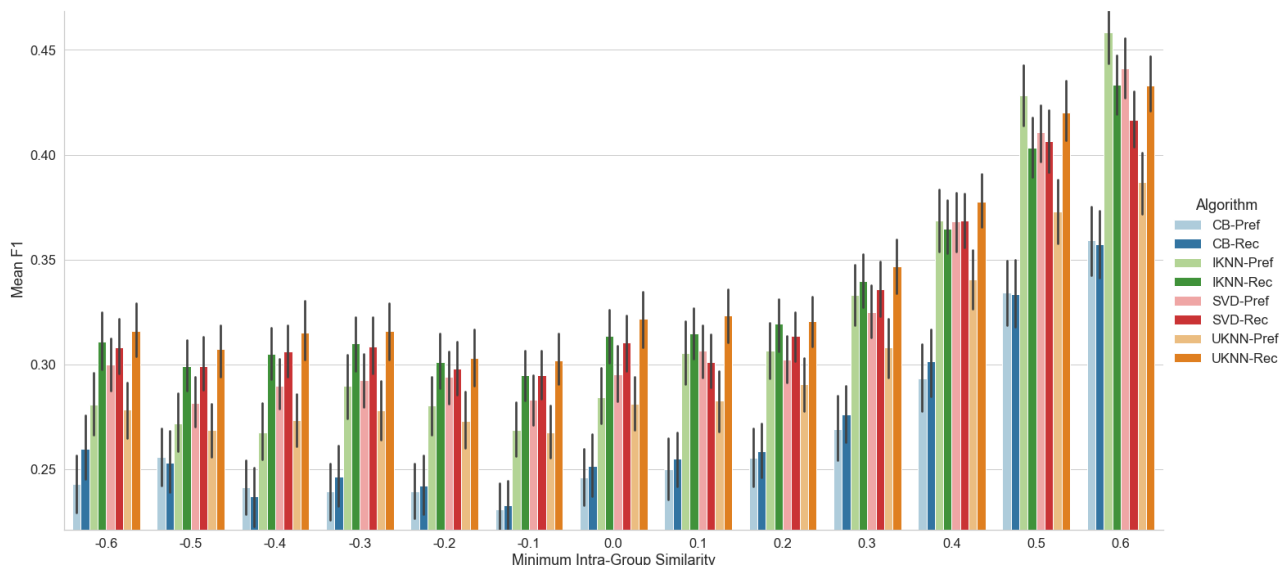


Figura 4.8: Impacto de la similitud intergrupual mínima en la métrica F1-Score

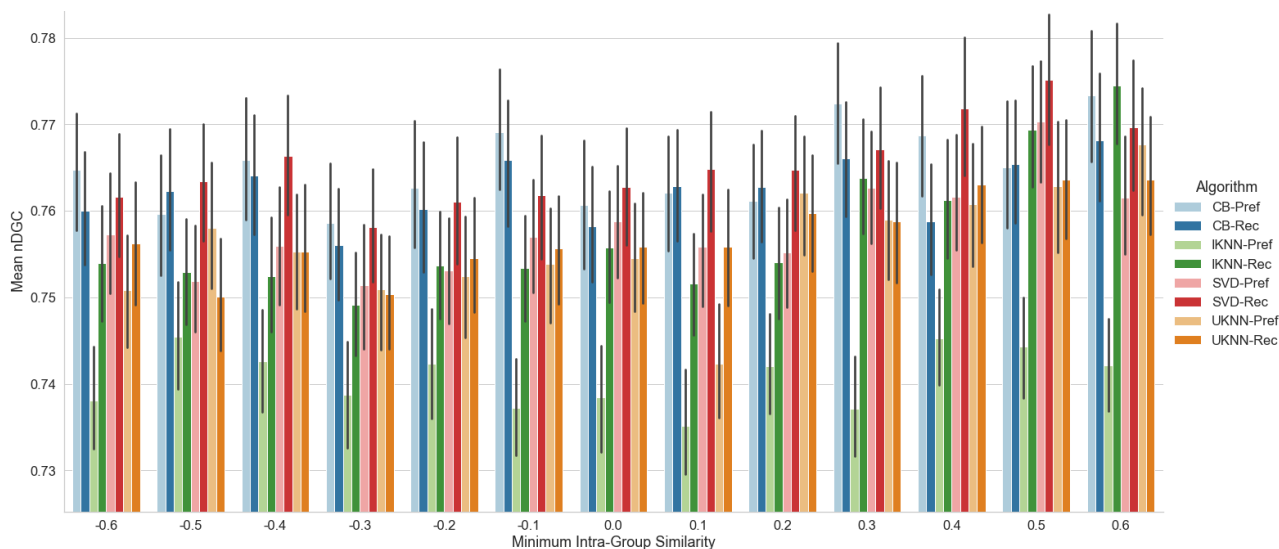


Figura 4.9: Impacto de la similitud intergrupal mínima en la métrica nDGC

Las figuras 4.8 y 4.9 presentan como varían las métricas F1 y nDGC en un sistema de recomendación grupal al variar los umbrales mínimos de similitud entre los miembros del grupo. En el eje horizontal se representan los umbrales de similitud, mientras que en el eje vertical se muestran los valores de F1 y nDGC. Cada barra de la figura representa los valores correspondientes a cada métrica obtenidos para diferentes algoritmos de recomendación.

Un análisis detallado revela una tendencia común para ambas métricas: a medida que se incrementa el umbral mínimo de similitud entre los usuarios del grupo, tanto el valor de F1 como el de nDGC experimentan un aumento significativo. Esto indica que establecer umbrales más altos de similitud entre los miembros del grupo contribuye a mejorar tanto la precisión (F1) como la calidad de las recomendaciones (nDGC) generadas por el sistema de recomendación grupal.

Es importante resaltar que esta relación entre el umbral de similitud y los valores de F1 y nDGC no es uniforme para todos los algoritmos evaluados. Se observa una variación en los valores de ambas métricas entre los diferentes algoritmos para cada umbral de similitud. Estos hallazgos sugieren que algunos algoritmos son más sensibles a los cambios en el umbral de similitud que otros, lo cual puede influir en la elección del algoritmo más adecuado para implementar un sistema de recomendación grupal.

La métrica F1 muestra un incremento significativo al aumentar el umbral mínimo de similitud en comparación con nDGC. Esto revela la sensibilidad de la métrica F1 a los

cambios en la similitud entre los miembros del grupo y su capacidad para proporcionar una medida precisa de la precisión del sistema de recomendación grupal. Una posible explicación de esta diferencia se basa en el enfoque distintivo de cada métrica. La métrica F1 se enfoca en la precisión del sistema de recomendación, considerando tanto los resultados verdaderos positivos como los falsos positivos.

A medida que el umbral de similitud se incrementa, es probable que se reduzcan los falsos positivos, lo que se traduciría en un aumento más marcado en la métrica F1. Por otro lado, la métrica nDGC evalúa la calidad general de las recomendaciones, teniendo en cuenta la posición de los elementos recomendados en el ranking. Es posible que algunos algoritmos logren mejorar la precisión sin afectar significativamente la posición de los elementos recomendados, lo que explicaría la relación más fuerte entre la métrica F1 y el umbral de similitud.

Sin embargo, se observa que el aumento significativo en las métricas de evaluación nDGC y F1 ocurre principalmente después de superar umbrales mínimos de similitud alrededor de -0.2 o -0.1 . Este patrón puede explicarse por el hecho de que cuando se forman grupos con similitudes por encima de los umbrales más pequeños, como los encontrados en el intervalo $[-0.6, -0.3]$, los usuarios agregados al grupo tienen valores de similitud mucho más altos que el umbral específico. Cabe destacar que este intervalo de similitud se muestra en la figura 4.7.

Para una mejor visualización de estos resultados, se presenta la figura 4.10, que utiliza una representación diferente. En esta figura, se muestran conjuntos de diagramas de dispersión en forma de matriz. Cada columna de la matriz corresponde a un algoritmo utilizado, mientras que cada fila representa una métrica de evaluación (nDGC o F1). En la intersección de cada columna y fila se encuentra un diagrama de dispersión que muestra la relación entre la similitud promedio entre los miembros de cada grupo y el valor promedio de la métrica correspondiente. Cada punto en el diagrama de dispersión representa un grupo individual y se posiciona según su similitud promedio en el eje horizontal y su valor de métrica en el eje vertical.

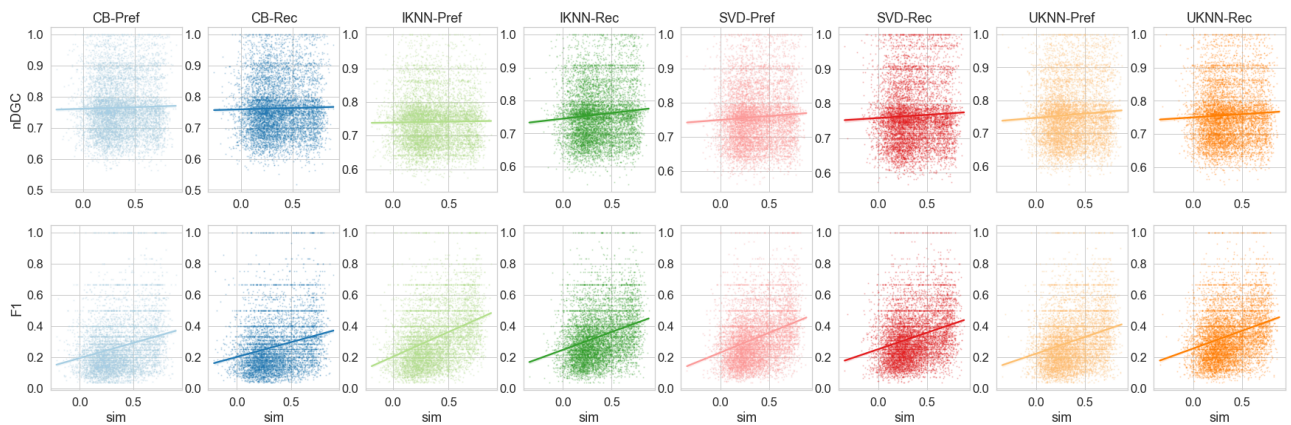


Figura 4.10: Relación entre la similitud promedio y las métricas de evaluación en sistemas de recomendación grupal: Análisis a través de diagramas de dispersión

Según los resultados presentados en los diagramas de dispersión, ahora se observa claramente que la similitud promedio de la mayoría de los grupos se encuentra por encima de aproximadamente -0.2 . Esto indica que, en general, los grupos formados tienen una similitud considerablemente alta entre sus miembros. Este hallazgo es consistente con la tendencia previamente mencionada, donde se observó un aumento significativo en las métricas de evaluación nDGC y F1 después de superar umbrales mínimos de similitud alrededor de -0.2 o -0.1 . Al formar grupos con una similitud promedio más alta, es más probable que se generen recomendaciones más precisas y de mayor calidad.

Al examinar los resultados de correlación de las métricas nDGC y F1 en diferentes algoritmos de recomendación grupal, se pueden hacer observaciones importantes. En cuanto al nDGC, ciertos algoritmos como CB-Pref y CB-Rec muestran una correlación baja con la similitud promedio, lo que sugiere una limitada relación entre la similitud y el nDGC. Esto implica que la mejora en la calidad de las recomendaciones grupales no está fuertemente influenciada por la similitud promedio entre los miembros del grupo. Por otro lado, al analizar la métrica F1, se nota que los algoritmos que utilizan SVD, UKNN o IKNN como recomendadores internos presentan una correlación moderada con la similitud promedio. Esto indica que un mayor grado de similitud entre los miembros del grupo se relaciona con una mayor precisión del sistema de recomendación grupal, considerando tanto los verdaderos positivos como los falsos positivos.

La diferencia en la influencia de la similitud entre los miembros del grupo en el enfoque basado en contenido y el filtrado colaborativo se debe a las características inherentes a cada enfoque. En el enfoque basado en contenido, la similitud se calcula en función de las características o atributos de los elementos que se recomiendan. En este enfoque se analizan las características de los elementos y buscan recomendaciones que sean similares a los elementos que han sido bien valorados o preferidos por un usuario. Por lo tanto, la similitud entre los miembros del grupo puede no tener un impacto significativo en la calidad de las recomendaciones generadas por estos algoritmos. En otras palabras, estos algoritmos son menos sensibles a la similitud entre los miembros del grupo, ya que se centran principalmente en las características de los elementos recomendados en lugar de las relaciones entre los usuarios.

En contraste, el enfoque de filtrado colaborativo se basa en la similitud entre usuarios para generar recomendaciones. Estos algoritmos encuentran similitudes entre los perfiles de preferencias de los usuarios y utilizan esta información para predecir las preferencias de un usuario en función de las preferencias de usuarios similares. Por lo tanto, en el enfoque de filtrado colaborativo, la similitud entre los miembros del grupo es un factor importante para determinar la calidad de las recomendaciones grupales. Si los miembros del grupo tienen una similitud promedio más alta, es más probable que sus preferencias sean más consistentes y que las recomendaciones generadas sean más precisas y relevantes para todos los miembros del grupo.

4.2. Evaluación de explicabilidad del modelo propuesto

En esta sección, se llevarán a cabo una serie de experimentos para evaluar la explicabilidad del modelo propuesto. El objetivo principal es comprender cómo diferentes parámetros y características del modelo afectan la métrica de fidelidad y el equilibrio de fidelidad en las explicaciones generadas.

Para iniciar, se llevará a cabo un experimento con el objetivo de analizar la fidelidad de las explicaciones generadas por diferentes métodos. Cabe destacar que se repetirá el experimento en 30 ocasiones para asegurar la robustez y consistencia de los resultados. Asimismo, se ha utilizado el método de agregación promedio (AVG) en todos los experimentos.

Posteriormente, se realizará otro experimento centrado en los umbrales mínimos de soporte y confianza de las reglas de asociación utilizadas para explicar las recomendaciones. Se llevarán a cabo pruebas variando estos umbrales y se analizará cómo afectan la fidelidad y el equilibrio de fidelidad en las explicaciones. También se explorará el impacto del tamaño del grupo en la fidelidad de las explicaciones, considerando la complejidad del grupo. Esto permitirá comparar la explicabilidad de un sistema de recomendación individual y la explicabilidad de un sistema de recomendación grupal. Además, se investigará cómo la similitud entre los usuarios del grupo influye en la fidelidad de las explicaciones. Se realizarán experimentos utilizando grupos con diferentes niveles de similitud en las preferencias de sus miembros, y se evaluará cómo esto se refleja en las explicaciones generadas.

En la etapa de evaluación de la explicabilidad del modelo propuesto, se procederá a generar grupos artificiales compuestos por miembros seleccionados de manera aleatoria utilizando el conjunto de datos completo para el entrenamiento. En esta evaluación, no será necesario contar con los valores reales de las calificaciones. Para llevar a cabo esta evaluación, se utilizará un conjunto de prueba denominado "Antitestset", el cual estará compuesto por pares de usuario-elemento que no se encuentren presentes en el conjunto de entrenamiento. Dado que el conjunto de prueba "Antitestset" tiene un tamaño considerable, se aplicará un muestreo aleatorio de 50% para reducir su tamaño.

4.2.1. Influencia de métodos de generación de explicaciones

En este experimento, realizaremos una evaluación de los métodos de generación de explicaciones en el modelo propuesto, centrándonos en dos enfoques clave: la estrategia de filtrado de preferencias del usuario (Apartado 2.2.1 del capítulo 3) y la estrategia de imputación de las calificaciones faltantes (Apartado 2.2.2 del capítulo 3). Nuestro objetivo principal es determinar cuál de estos métodos ofrece una mejor explicabilidad en el sistema de recomendación grupal.

A continuación, se muestra en la Figura 4.11 un diagrama de barras que ilustra el número de reglas generadas por cada estrategia, que incluye la estrategia de filtrado de elementos relevantes del conjunto de entrenamiento (trainset) y la imputación de las calificaciones faltantes utilizando varios algoritmos como UKNN, IKNN, SVD y CB (Con

D=65). Por otro lado, la Figura 4.12 presenta el número de elementos cubiertos por estas reglas de asociación.

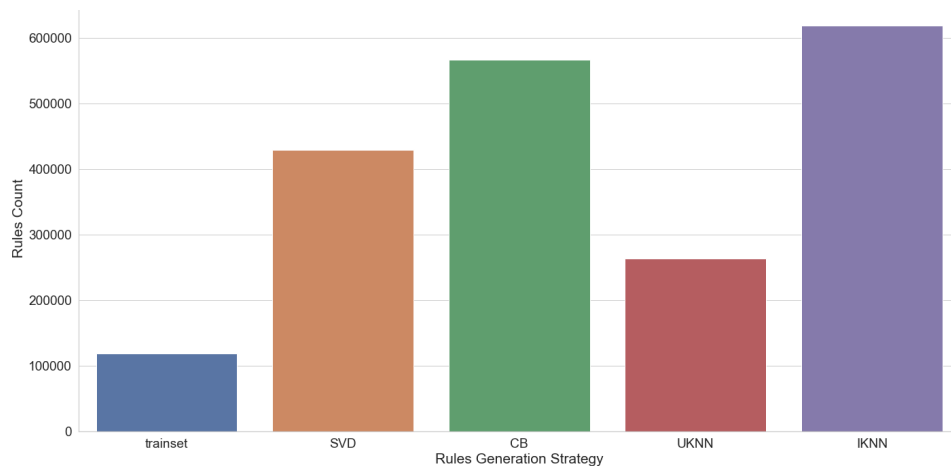


Figura 4.11: Número de las reglas generadas para cada estrategia de filtrado e imputación

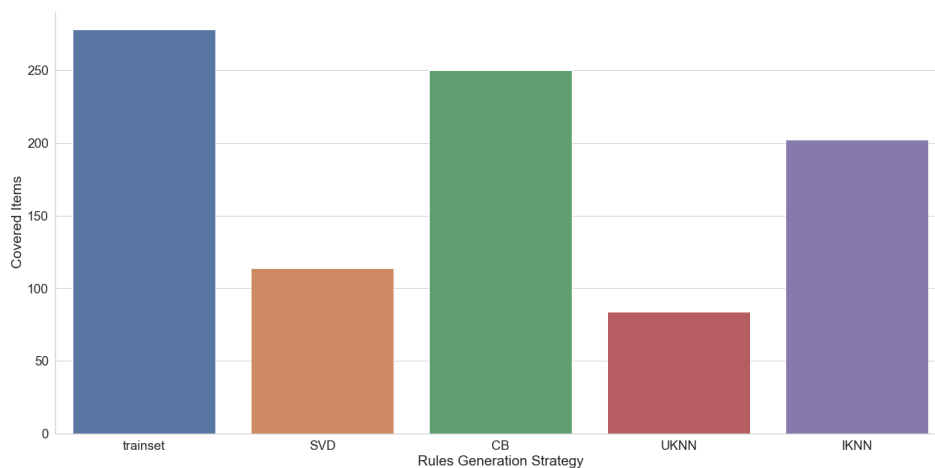


Figura 4.12: Número de los elementos cubiertos por las reglas de asociación generadas

En los resultados obtenidos, se observa que la estrategia de filtrado de elementos relevantes del conjunto de entrenamiento generó un menor número de reglas de asociación en comparación con las estrategias de imputación de calificaciones faltantes. Sin embargo, esta estrategia logró cubrir un mayor número de elementos en el sistema de recomendación.

Por otro lado, las estrategias de imputación de las calificaciones faltantes (SVD, CB, UKNN y IKNN) generaron un mayor número de reglas de asociación en comparación con el filtrado de elementos relevantes. Sin embargo, en cuanto a la cobertura de elementos, se observa que hay variaciones entre los diferentes algoritmos utilizados. Algunos algoritmos, como IKNN y CB, lograron cubrir un mayor número de elementos con

las reglas generadas, mientras que algoritmos como UKNN y SVD mostraron una cobertura más limitada, abarcando menos elementos en el sistema de recomendación.

A continuación, en la Figura 4.13, se presentan histogramas que muestran la distribución de los valores de soporte y confianza de las reglas de asociación generadas por cada estrategia. Los histogramas proporcionan una visualización clara de cómo varían estos valores en cada estrategia y permiten comparar su distribución.

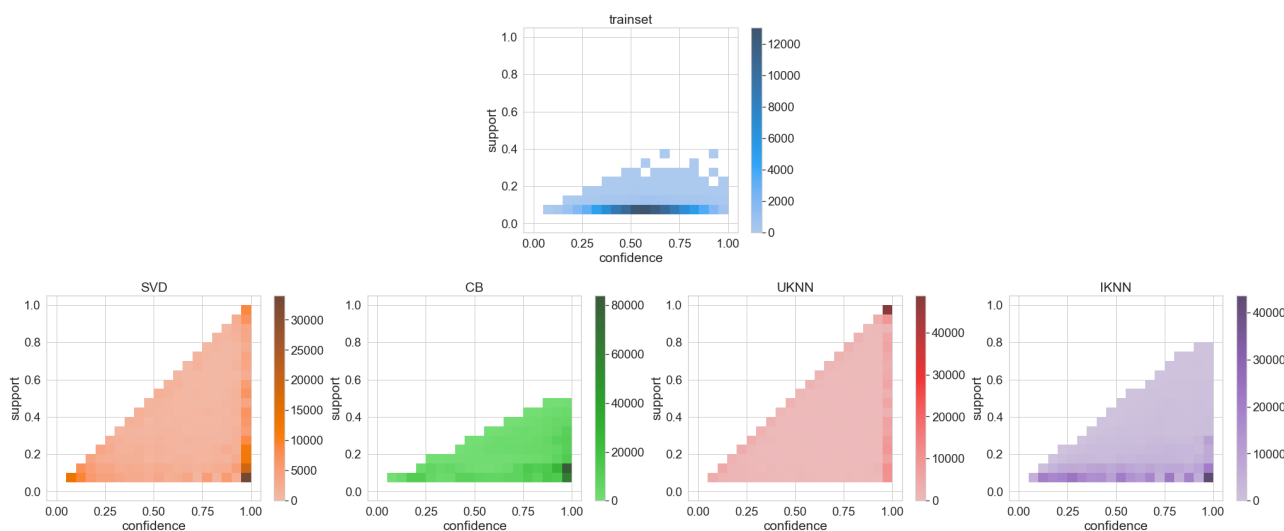


Figura 4.13: Distribución de los valores de soporte y confianza en las reglas de asociación generadas

Tras analizar los resultados obtenidos, representados por los valores de soporte y confianza de cada estrategia en la figura 4.13, se observaron patrones interesantes. La estrategia de filtrado de elementos relevantes del conjunto de entrenamiento (trainset) mostró principalmente un soporte bajo, con una amplia variabilidad en los valores. En cuanto a la confianza, se observaron niveles tanto bajos como altos.

En contraste, la estrategia de imputación de las calificaciones faltantes con el algoritmos SVD generó reglas con valores altos tanto de soporte como de confianza, lo que indica una alta confianza en las reglas generadas. El algoritmo CB mostró distribuciones equilibradas tanto en soporte como en confianza, con una presencia moderada de reglas y niveles altos de confianza. El algoritmo UKNN produjo reglas altamente confiables, con valores elevados de soporte y confianza. Sin embargo, IKNN presentó una variabilidad en los valores de soporte y confianza, sin mostrar un sesgo claro hacia valores altos o bajos.

En términos de confianza, todas las estrategias mostraron valores máximos de 1, lo que indica que se generaron reglas con un alto nivel de confianza en todas ellas. Esto implica que las reglas generadas son muy confiables y respaldadas por una fuerte relación entre los elementos asociados. No obstante, es importante tener en cuenta que la distribución de los valores de confianza puede variar entre las estrategias.

Por otro lado, las estrategias de filtrado de elementos relevantes, representadas por "trainset", y la imputación de las calificaciones faltantes con CB, presentaron valores máximos de soporte inferiores a 0.5, lo que indica una posible menor consistencia en el cumplimiento de las reglas. Sin embargo, todas las estrategias mostraron un alto nivel de confianza en las reglas generadas, lo que indica que las reglas son confiables.

Para comparar la fidelidad de las dos estrategias, se presentan los resultados obtenidos en la Figura 4.14. En este diagrama de barras se muestran los valores promedio de fidelidad para cada estrategia de generación de explicaciones y para diferentes algoritmos de recomendación grupal. En el eje vertical se muestra el valor promedio de fidelidad. En el eje horizontal se encuentran las dos estrategias: la estrategia de filtrado de preferencias relevantes del usuario (Relevant Item Filtering) y la estrategia de imputación de calificaciones faltantes (Missing Value Imputation). Para cada algoritmo de recomendación, se utilizó el mismo algoritmo interno para imputar la matriz de calificaciones y así generar las reglas de asociación utilizadas en la explicación de las recomendaciones.

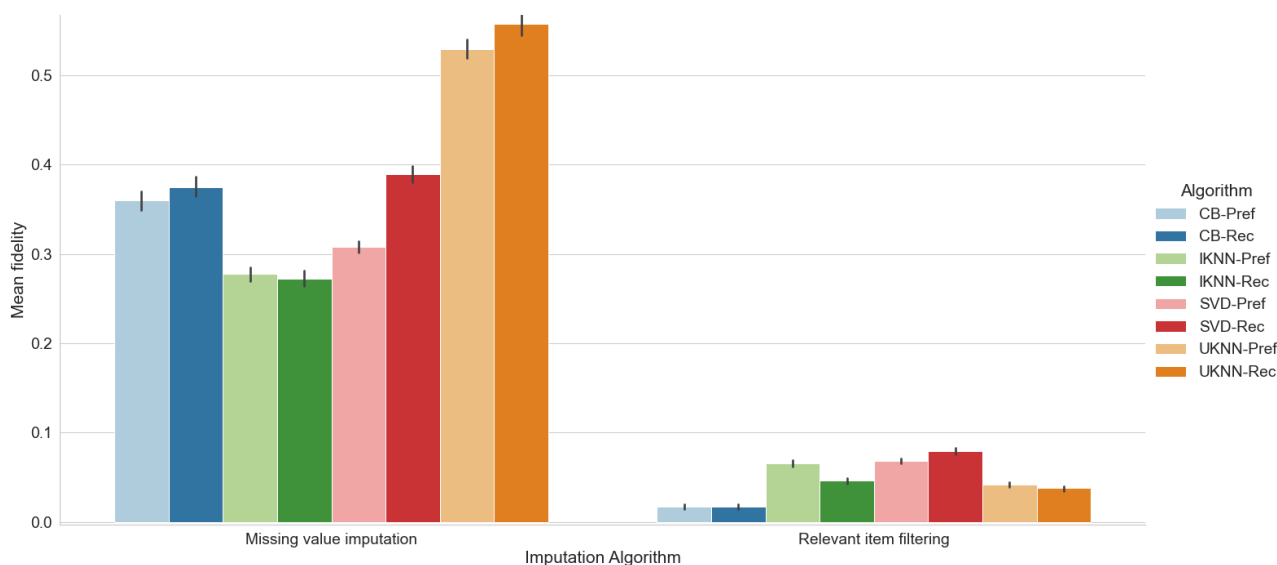


Figura 4.14: Impacto del método de generación de reglas de asociación en la métrica de Fidelidad

Para llevar a cabo este experimento, hemos establecido los parámetros de evaluación necesarios. En primer lugar, hemos fijado el umbral mínimo de soporte en 0.05 y la confianza en 0.5. Estos valores nos permitirán identificar reglas de asociación significativas y confiables. Además, hemos formado grupos con 4 miembros aleatorios del conjunto de datos disponible, lo que nos permite obtener resultados representativos. Cada grupo ha recibido una lista de recomendaciones de 5 elementos.

Basándonos en los resultados presentados en la Figura 4.14, podemos extraer las siguientes conclusiones e interpretaciones:

En primer lugar, se observa que la estrategia de imputación de calificaciones faltantes se destaca como la opción más efectiva en términos de fidelidad para todos los algoritmos de recomendación grupal evaluados. Esto implica que al emplear esta estrategia, se logra generar explicaciones más precisas y consistentes en comparación con la estrategia de filtrado de preferencias de los usuarios. Los algoritmos que utilizan la estrategia de imputación de calificaciones faltantes lograron superar un umbral del 25% de fidelidad, mientras que al utilizar la estrategia de filtrado de elementos relevantes todos los algoritmos no han podido superar el umbral del 10% de fidelidad.

La diferencia en los niveles de efectividad entre las dos estrategias se debe al enfoque utilizado en la estrategia de imputación de calificaciones faltantes. Esta estrategia genera transacciones que incluyen una combinación de ítems valorados positivamente por los usuarios y los ítems recomendados con altas predicciones. Al combinar estos dos tipos de ítems en las reglas de asociación, se pueden proporcionar explicaciones de alta calidad y mayor precisión. Las reglas resultantes actúan como un puente explicativo entre las preferencias conocidas del usuario y las recomendaciones sugeridas, lo que aumenta la relevancia y utilidad para los usuarios del sistema.

Es interesante notar también que al aplicar la estrategia de imputación de calificaciones faltantes, en la mayoría de los casos, los algoritmos de recomendación que utilizan la estrategia de agregación de recomendaciones son más explicables y obtienen resultados más precisos en comparación con la estrategia de agregación de preferencias. Esta tendencia se mantiene independientemente del algoritmo interno utilizado.

Una posible explicación de esto podría ser que la estrategia de agregación de predicciones considera las preferencias individuales de cada miembro del grupo al generar las recomendaciones. Al utilizar las predicciones individuales, se tiene en cuenta la información específica de cada usuario, lo que permite una mejor coincidencia entre las reglas de asociación y las preferencias individuales. En contraste, la estrategia de agregación de preferencias combina las preferencias de todos los miembros del grupo, lo que puede ocultar las preferencias individuales y diluir la relación entre las reglas de asociación y las preferencias de cada usuario. Esto puede resultar en una menor capacidad para explicar las recomendaciones utilizando las reglas de asociación generadas previamente.

Es interesante resaltar un patrón inesperado en los resultados de los experimentos de la estrategia de imputación de calificaciones faltantes. Aunque el algoritmo IKNN generó un mayor número de reglas de asociación (Figura 4.11) y cubrió una mayor cantidad de elementos en el conjunto de datos (Figura 4.12), su valor de fidelidad resultó ser el peor entre los algoritmos evaluados. Por otro lado, UKNN obtuvo mejores resultados a pesar de generar menos reglas de asociación y cubrir un menor número de elementos. Este hallazgo nos lleva a una conclusión crucial: la calidad de las reglas generadas es un factor fundamental, incluso más importante que la cantidad. En otras palabras, es primordial producir reglas de asociación que sean relevantes, precisas y de alta calidad para lograr una explicación efectiva de las recomendaciones.

4.2.2. Influencia del algoritmo de imputación de las calificaciones faltantes

En el experimento previo, se demostró que la metodología de imputación de calificaciones faltantes resultó considerablemente más eficiente. Sin embargo, surge la interrogante de si es necesario utilizar el mismo algoritmo interno de recomendación para la imputación de estos valores.

Para responder a esta pregunta, se realizaron experimentos de imputación cruzada utilizando una variedad de algoritmos tanto para la imputación como para la generación de recomendaciones. Los algoritmos de imputación utilizados incluyeron la Descomposición de Valores Singulares (SVD), Vecinos más Cercanos del Usuario (UKNN), Vecinos más Cercanos del Ítem (IKNN) y el Basado en Contenido (CB).

De manera similar, se emplearon los mismos algoritmos (SVD, IKNN, UKNN y CB) utilizando las estrategias de agregación de preferencias (Pref) y de recomendaciones (Rec). El objetivo principal de estos experimentos fue evaluar la calidad y coherencia de las explicaciones generadas para cada combinación de algoritmos. Mediante el análisis de los resultados obtenidos mediante la imputación cruzada, se buscó comprender cómo la elección del algoritmo de imputación afecta la explicabilidad del sistema de recomendación en relación a los grupos de usuarios.

En la Figura 4.15 se presentan los resultados obtenidos, mostrando un gráfico que representa los algoritmos utilizados para la imputación de las calificaciones faltantes en el eje horizontal y los valores promedio de fidelidad en el eje vertical. Cada barra en el gráfico corresponde a un algoritmo de recomendación grupal específico.

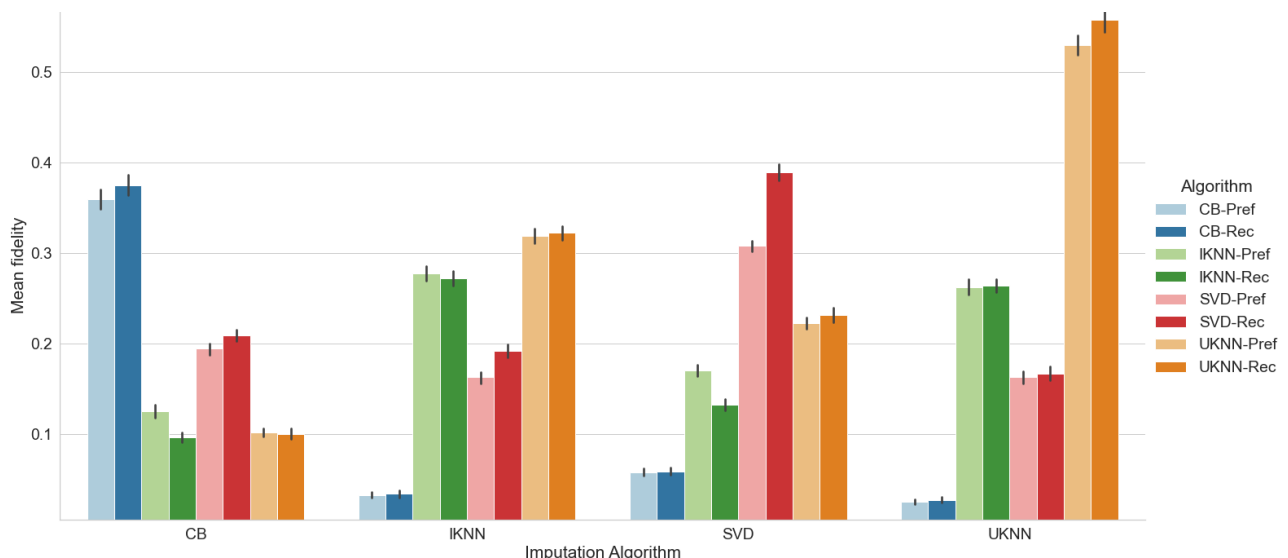


Figura 4.15: Impacto del algoritmo de imputación en la métrica de Fidelidad

Según los resultados presentados en la Figura 4.15, se puede observar que todos los algoritmos de recomendación grupal mostraron los mejores valores de fidelidad cuando se utilizó el mismo algoritmo interno de recomendaciones como algoritmo de imputación de las calificaciones faltantes. Esta tendencia se refleja en todas las barras del gráfico, donde se muestra un aumento significativo en los valores promedio de fidelidad cuando se emplea el mismo algoritmo tanto para la imputación como para la generación de recomendaciones. Esto indica que existe una mayor coherencia y calidad en las explicaciones generadas cuando se utiliza el algoritmo interno para imputar los valores faltantes.

Cuando el algoritmo de imputación es diferente al algoritmo principal utilizado para generar las recomendaciones, puede haber discrepancias en la forma en que se completan los valores faltantes. Esto puede llevar a inconsistencias en las explicaciones generadas y afectar su calidad y su confiabilidad. Sin embargo, al utilizar el mismo algoritmo interno como estrategia de imputación, se establece una coherencia en el proceso de completar los valores faltantes. Esto puede resultar en explicaciones más precisas y consistentes, lo que a su vez aumenta la fidelidad de las explicaciones y mejora la experiencia del usuario. Estos hallazgos sugieren que al elegir el algoritmo de imputación, es importante considerar su compatibilidad y coherencia con el algoritmo principal de recomendación utilizado. Utilizar el mismo algoritmo interno para la imputación puede mejorar la consistencia y calidad de las explicaciones, lo que contribuye a una mejor experiencia del usuario.

4.2.3. Influencia del tamaño de grupo

En este apartado, analizaremos la influencia del tamaño del grupo en la fidelidad de los sistemas de recomendación. Para ello, se realizó un experimento donde se varió el tamaño del grupo de recomendación, desde una recomendación individual (tamaño del grupo = 1) hasta una recomendación grupal (tamaños del grupo de 2 a 10). En el eje horizontal de la figura 4.16 se encuentran los diferentes tamaños del grupo, mientras que en el eje vertical se representa el valor promedio de fidelidad. Cada barra de la figura representa un algoritmo de recomendación.

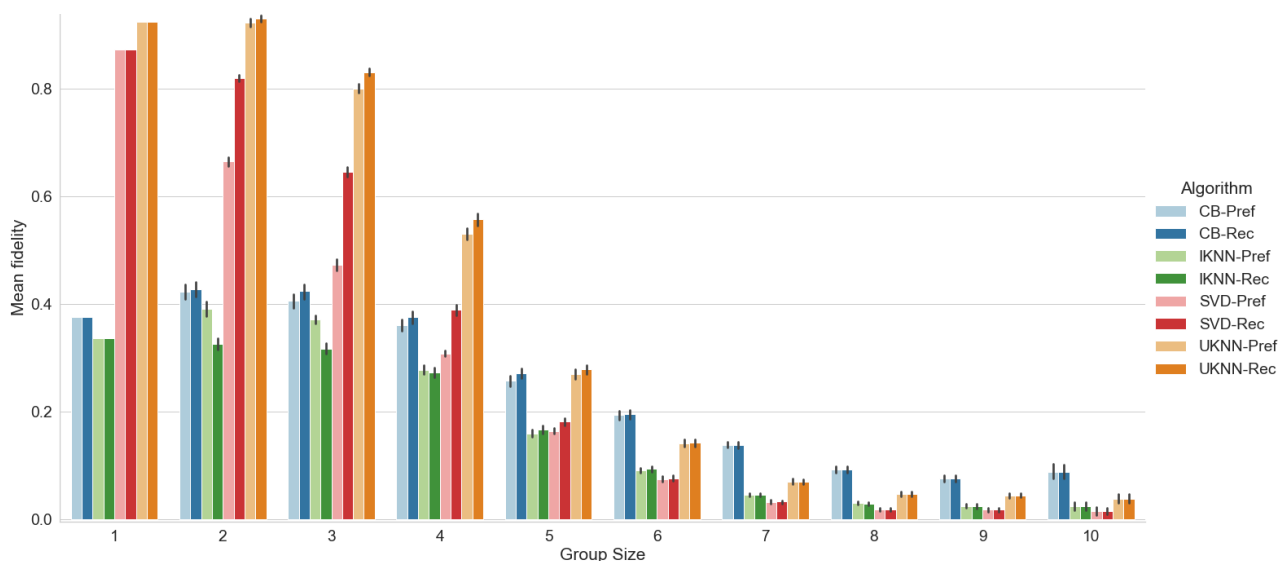


Figura 4.16: Impacto del tamaño de grupo en la métrica de Fidelidad

En la figura 4.16, podemos observar la relación entre el tamaño del grupo y el valor promedio de fidelidad. Conforme aumenta el tamaño del grupo, se observa una disminución en el valor de fidelidad. Esto significa que a medida que se incrementa la cantidad de miembros en el grupo de recomendación, la proporción de elementos recomendados que pueden ser explicados disminuye.

Es importante destacar que, sin embargo, los algoritmos basados en contenido exhiben valores de fidelidad superiores en comparación con los demás algoritmos. Este hallazgo puede interpretarse como una indicación de que los algoritmos basados en contenido ofrecen una mejor explicabilidad en el sistema, siendo tolerantes a las preferencias distintas y los conflictos en las preferencias de los miembros del grupo. El mismo patrón se observó en el experimento que evaluó el impacto del tamaño del grupo en la métrica de nDGC (Fig. 4.4). Los algoritmos basados en contenido obtuvieron valores más altos en comparación con los demás algoritmos, exactamente en grupos de tamaño grande.

Al analizar los resultados, es posible identificar algunas explicaciones sobre por qué se reduce la fidelidad a medida que aumenta el tamaño del grupo:

- ❖ **Diversidad de preferencias:** A medida que el tamaño del grupo crece, también lo hace la diversidad de preferencias y necesidades de sus miembros. Esto implica que encontrar recomendaciones que sean explicables y relevantes para todos los miembros se vuelve más desafiante. Los algoritmos de recomendación pueden tener dificultades para generar explicaciones claras y coherentes que satisfagan a todos los miembros del grupo.
- ❖ **Consenso y compromisos:** En un grupo de recomendación, es necesario llegar a un consenso entre las preferencias individuales para generar recomendaciones grupales. Esto puede llevar a compromisos en la elección de elementos recomendados, lo que a su vez puede disminuir la explicabilidad. Las explicaciones pueden volverse más genéricas o menos específicas para adaptarse a las preferencias diversas del grupo.
- ❖ **Personalización limitada:** A medida que el tamaño del grupo aumenta, la personalización individual se ve comprometida en favor de la consideración del

grupo en su conjunto. Esto significa que las recomendaciones pueden basarse más en preferencias generales del grupo en lugar de adaptarse a las preferencias individuales de cada miembro. Como resultado, las explicaciones pueden ser menos personalizadas y comprensibles para cada usuario.

En términos de fidelidad en sistemas de recomendación individual y sistemas de recomendación grupal, se puede concluir que los sistemas de recomendación individual tienden a ofrecer una mayor fidelidad en comparación con los sistemas de recomendación grupal. Esto se debe a que en los sistemas individuales, las recomendaciones se adaptan directamente a las preferencias y necesidades de un único usuario, lo que facilita la generación de explicaciones claras y personalizadas. Por otro lado, en los sistemas de recomendación grupal, la necesidad de encontrar un consenso entre las preferencias de los diferentes miembros puede reducir la fidelidad y la personalización de las explicaciones.

4.2.4. Influencia de los umbrales mínimos de soporte y confianza

En esta sección, examinamos la influencia de los umbrales mínimos de soporte y confianza en la generación de explicaciones en el modelo propuesto. Ajustaremos estos umbrales y evaluaremos cómo afectan la cantidad y la calidad de las reglas de asociación generadas. Analizaremos diferentes combinaciones de umbrales y mediremos su impacto en la capacidad del modelo para explicar las recomendaciones de manera precisa y significativa.

En la figura 4.17, el eje vertical representa el valor promedio de la fidelidad de los algoritmos de recomendación grupal, y el eje horizontal muestra los umbrales mínimos utilizados para el umbral de soporte. Los umbrales de soporte varían desde 0.05 hasta 0.95 con un incremento de 0.05. El valor de confianza se mantiene constante en 0.5.

En la figura 4.18, el eje vertical también representa el valor promedio de la fidelidad de los algoritmos de recomendación grupal, y el eje horizontal muestra los umbrales mínimos utilizados para el umbral de confianza. Los umbrales de confianza varían desde 0.1 hasta 0.9 con un incremento de 0.1. El valor de soporte se fija en 0.05.

En el análisis de los resultados obtenidos al variar el umbral mínimo de soporte (Fig. 4.17), se observó una tendencia común en todos los algoritmos evaluados. A medida

que se incrementa el valor del umbral mínimo de soporte, la fidelidad de los algoritmos de recomendación grupal tiende a disminuir. Esta disminución era esperada, ya que al utilizar umbrales mínimos más altos se reduce la cantidad de reglas de asociación utilizadas para explicar las recomendaciones.

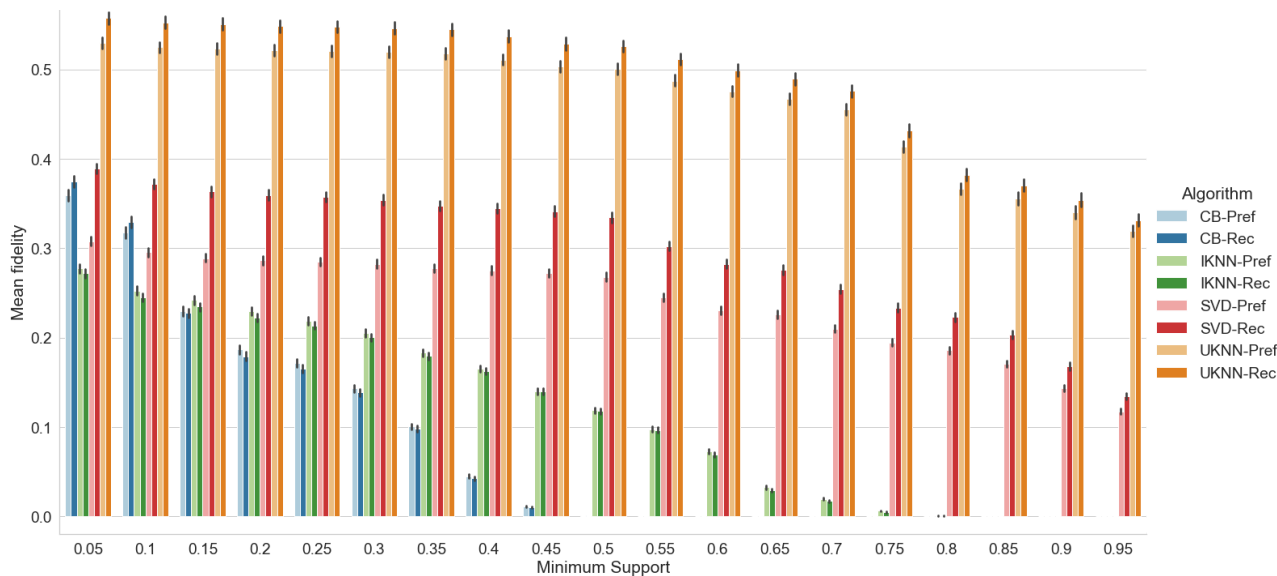


Figura 4.17: Impacto del umbral mínimo de soporte en la métrica de Fidelidad

Cada algoritmo mostró un umbral específico a partir del cual su fidelidad comenzó a disminuir y alcanzó un valor de fidelidad de 0 en un umbral particular. Los algoritmos SVD, UKNN (con ambas estrategias de agregación) siguieron la misma tendencia, comenzando a disminuir su fidelidad a partir de un umbral mínimo de soporte de 0.5. Sin embargo, ninguno de estos algoritmos alcanzó un valor de fidelidad de 0 incluso con umbrales de soporte muy altos, como 0.9 y 0.95.

Por otro lado, el algoritmo IKNN (con ambas estrategias de agregación) mostró una disminución en su fidelidad desde el principio, alcanzando un valor de fidelidad de 0 con un umbral mínimo de soporte de 0.85. Se observaron valores muy bajos de fidelidad en el rango de soporte de 0.65 a 0.8 para este algoritmo.

En el caso del algoritmo CB (con ambas estrategias de agregación), se observó una disminución dramática en su fidelidad desde el inicio, alcanzando un valor de fidelidad de 0 antes que los demás algoritmos, a partir de un umbral mínimo de soporte de 0.5. Es importante destacar que, aunque los algoritmos IKNN y CB mostraron una

disminución desde el principio, no experimentaron la misma magnitud de reducción en su fidelidad. El algoritmo CB disminuyó rápidamente en comparación con el algoritmo IKNN.

Los resultados obtenidos pueden ser explicados mediante el análisis de los histogramas en la figura 4.13, los cuales representan la distribución de los valores de soporte y confianza de las reglas de asociación generadas por cada estrategia.

En el caso de la imputación de calificaciones faltantes con SVD y UKNN, se puede observar que las reglas generadas tienen una distribución de soporte que abarca valores desde 0 hasta 1. Esto indica que existen reglas de asociación para una amplia gama de situaciones. Esta distribución uniforme contribuye a que los algoritmos SVD y UKNN, con ambas estrategias de agregación, mantengan valores altos de fidelidad incluso al aumentar el umbral mínimo de soporte. En otras palabras, estos algoritmos siguen siendo capaces de generar recomendaciones confiables y explicables a pesar de reducir la cantidad de reglas utilizadas.

En contraste, el algoritmo IKNN muestra una distribución de soporte más limitada, donde no se generan reglas de asociación con un valor de soporte superior a 0.8. Debido a esta limitación, al aumentar el umbral mínimo de soporte, la fidelidad del algoritmo IKNN con las dos estrategias de agregación disminuye rápidamente y alcanza un valor de 0 con un umbral de 0.85.

En el caso del algoritmo CB, se observa que no se generan reglas de asociación a partir de un umbral mínimo de soporte de 0.55. Al superar este umbral, la fidelidad del algoritmo CB alcanza un valor de 0 antes que los demás algoritmos.

Por otro lado, al observar la figura 4.18, se nota que todos los algoritmos presentan una disminución en la fidelidad a medida que aumenta el umbral mínimo de confianza. Esto ocurre debido a que, al incrementar el umbral de confianza, se reduce el número de reglas de asociación generadas, lo que a su vez disminuye la cantidad de reglas de asociación utilizadas para la explicación de recomendaciones.

Es interesante notar que todos los algoritmos siguen una tendencia similar de disminución en la fidelidad a medida que se incrementa el umbral de confianza. Esto se debe al hecho de que el número de reglas de asociación generadas disminuye de manera similar para todos los algoritmos.

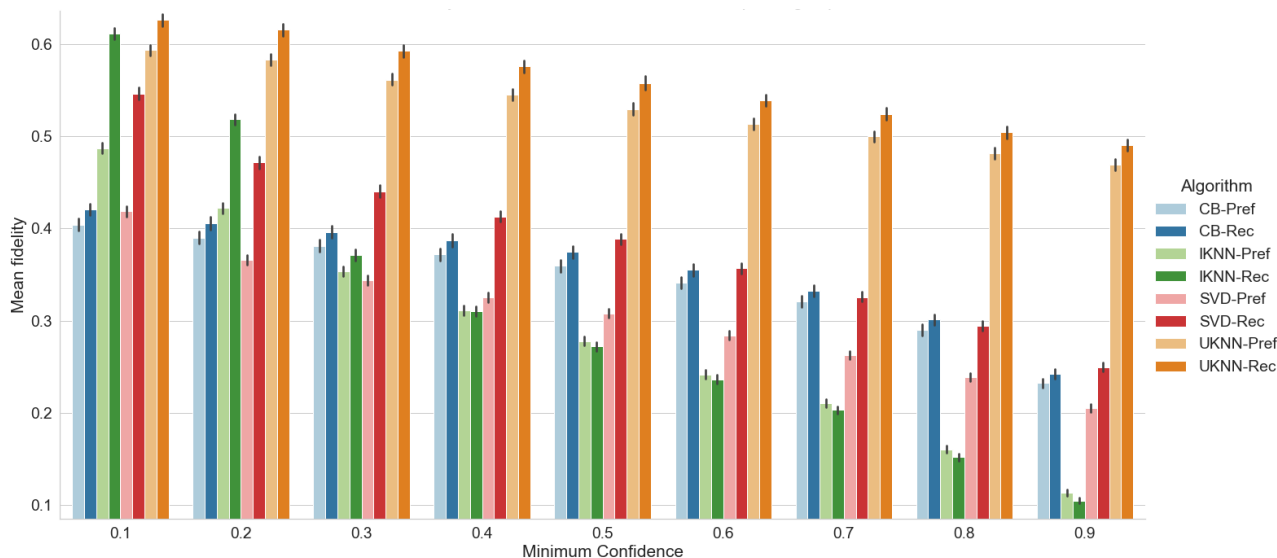


Figura 4.18: Impacto del umbral mínimo de confianza en la métrica de Fidelidad

Sin embargo, es importante destacar que en ningún caso la fidelidad alcanza el valor 0, incluso cuando se utiliza el umbral máximo de confianza. Esto se debe a que, al analizar la figura 4.13, se puede observar que todos los algoritmos generan reglas de asociación en todo el rango de valores de confianza, es decir, desde 0 hasta 1. A diferencia de lo que ocurre con el umbral de soporte, donde se observan diferencias significativas en la generación de reglas de asociación entre los algoritmos.

4.2.5. Influencia de la similitud intergrupar

En esta sección, exploramos la influencia de la similitud intergrupar en la generación de explicaciones en el modelo propuesto. Analizaremos cómo la similitud entre diferentes grupos de usuarios afecta la calidad y la coherencia de las explicaciones proporcionadas por el modelo.

La figura 4.19 presenta como varía el promedio de fidelidad de nuestro modelo de sistema de recomendación grupal al variar los umbrales mínimos de similitud entre los miembros del grupo. En el eje horizontal se representan los umbrales de similitud, mientras que en el eje vertical se muestra el valor promedio de fidelidad. Cada barra de la figura representa los valores obtenidos para diferentes algoritmos de recomendación grupal.

Al examinar los resultados, se observa que no existe una correlación clara entre la similitud entre los miembros del grupo y la métrica de fidelidad. En otras palabras, el valor

de similitud entre los miembros del grupo no parece tener un impacto significativo en la fidelidad del modelo.

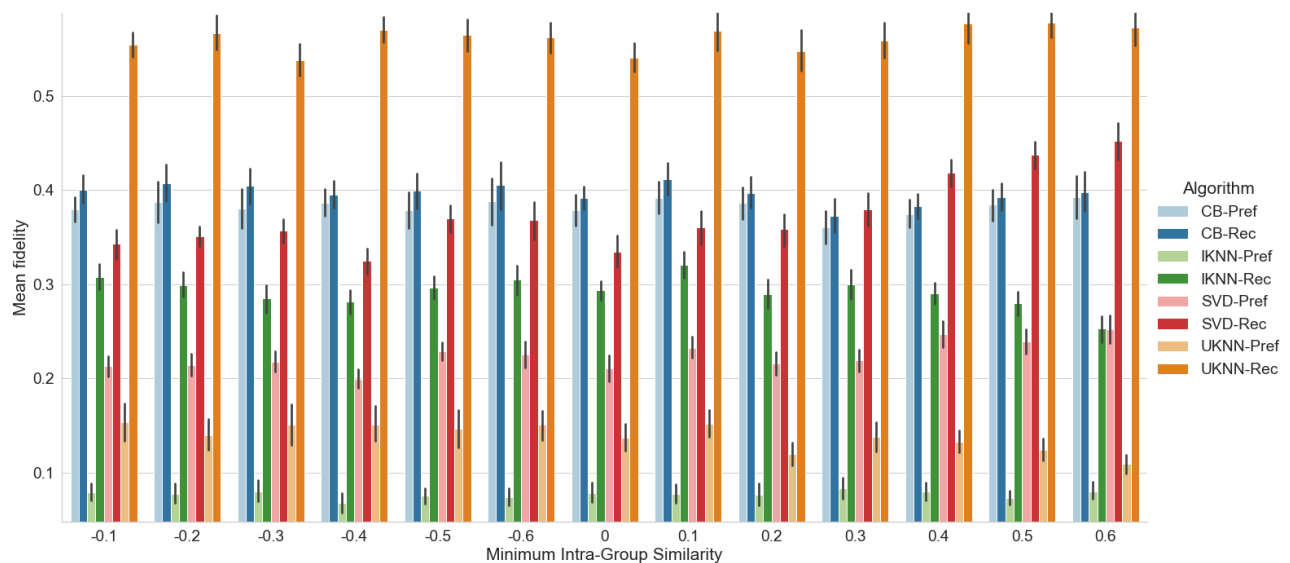


Figura 4.19: Impacto de la similitud intergrupual mínima en la métrica de Fidelidad

Esto puede indicar que nuestro modelo de sistema de recomendación grupal es robusto frente a las diferencias de similitud entre los miembros de un grupo. Es capaz de proporcionar explicaciones coherentes sin depender fuertemente de la similitud entre los usuarios del grupo. Estas conclusiones pueden tener implicaciones importantes, ya que sugieren que nuestro modelo puede ser aplicado de manera efectiva en diversos escenarios y grupos de usuarios, independientemente de su nivel de similitud.

5. Conclusión

En este capítulo se llevaron a cabo experimentos para evaluar el rendimiento y la explicabilidad de un sistema de recomendación grupal. Se compararon diferentes métodos de agregación y se evaluó su impacto en las métricas de rendimiento. También se analizó cómo los parámetros y características del modelo afectan la fidelidad y el equilibrio de las explicaciones generadas.

El objetivo principal fue identificar el método de agregación más efectivo y evaluar la influencia de las estrategias de generación de explicaciones en la explicabilidad del sistema. Se compararon métodos de agregación propuestos en la literatura y se analizaron enfoques de filtrado de preferencias del usuario y de imputación de

calificaciones faltantes. Estos experimentos brindaron información importante sobre la influencia de estos parámetros en el rendimiento y la explicabilidad del sistema.

Estos experimentos son un avance significativo en el desarrollo de sistemas de recomendación grupales más efectivos y explicables. Proporcionan una base sólida para la toma de decisiones informadas en el diseño e implementación de estos sistemas en entornos colaborativos. Además, abren oportunidades para investigaciones futuras, como la combinación de métodos de agregación, la influencia de diferentes algoritmos de recomendación y la consideración de restricciones adicionales en la formación de grupos.

CONCLUSIONES Y TRABAJOS FUTUROS

CAPÍTULO 5

1. Conclusiones

En resumen, se realizó una revisión de la literatura existente sobre sistemas de recomendación, investigando diferentes enfoques para combinar las preferencias y recomendaciones de los usuarios. También se analizaron las estrategias de explicabilidad disponibles y las métricas de evaluación para medir el rendimiento, la precisión y la explicabilidad del modelo propuesto.

Para evaluar el rendimiento y la precisión del modelo, se utilizaron métricas como el nDGC (Normalized Discounted Cumulative Gain) y el F1-Score. Estas métricas permitieron medir la efectividad y precisión de las recomendaciones generadas. En cuanto a la evaluación de la explicabilidad, se empleó la métrica de Fidelidad, que evaluó la proporción de elementos explicables en la lista de recomendaciones.

Se compararon diversas estrategias de explicabilidad, como la imputación de calificaciones faltantes y el filtrado de preferencias de los usuarios, con el objetivo de determinar cuál generaba explicaciones más precisas y coherentes. Además, se llevaron a cabo experimentos para evaluar el impacto de los parámetros del sistema, como el tamaño del grupo, el número de elementos recomendados, la similitud entre los miembros del grupo, la estrategia de agregación, etc. A continuación, se presentan los resultados obtenidos de estos experimentos:

En primer lugar, se determinó que el método de agregación AVG (promedio) es altamente efectivo para generar recomendaciones grupales precisas y relevantes. Al equilibrar las preferencias individuales de los miembros del grupo, este método logra ajustarse con precisión a los intereses y necesidades del grupo en su conjunto.

Además, se observó que el tamaño del grupo tiene un impacto en la precisión y calidad de las recomendaciones. A medida que el grupo se vuelve más grande, se vuelve más desafiante llegar a un consenso entre los miembros en términos de gustos y preferencias, lo que afecta la generación de recomendaciones satisfactorias para todos. Además, se encontró que los algoritmos basados en contenido tienen un mejor desempeño en grupos más grandes.

En relación con la longitud de la lista de recomendación, se determinó que aumentarla mejora tanto la precisión como la exhaustividad de las recomendaciones

grupales. Al proporcionar más opciones a los usuarios, se incrementa la probabilidad de incluir elementos relevantes para diferentes perfiles de usuarios. Sin embargo, es importante tener en cuenta que al agregar más elementos, se diluye la importancia de los ítems más relevantes, lo que afecta la calidad general de las recomendaciones.

Finalmente, se encontró que la alta similitud entre las preferencias de los usuarios del grupo contribuye a mejorar la precisión y calidad de las recomendaciones generadas. Sin embargo, la relación entre el umbral de similitud y las métricas de rendimiento no es uniforme para todos los algoritmos evaluados, resaltando la importancia de considerar las características y sensibilidades de cada algoritmo al implementar un sistema de recomendación grupal.

En los experimentos de explicabilidad, se observaron diversos factores que influyen en la fidelidad del modelo. En primer lugar, se determinó que la estrategia de imputación de calificaciones faltantes es la opción más efectiva en términos de fidelidad, generando explicaciones precisas y consistentes en comparación con la estrategia de filtrado de preferencias de los usuarios. Los resultados mostraron que emplear el mismo algoritmo interno de recomendaciones como algoritmo de imputación de las calificaciones faltantes proporcione una mayor coherencia y calidad en las explicaciones generadas.

Además, se notó que al utilizar la estrategia de imputación de calificaciones faltantes, los algoritmos de recomendación que emplean la estrategia de agregación de recomendaciones son más explicables y obtienen resultados más precisos en comparación con la estrategia de agregación de preferencias.

En cuanto al tamaño del grupo, se observó que a medida que este aumenta, la fidelidad disminuye. Sin embargo, los algoritmos basados en contenido mostraron una mayor tolerancia a grupos de mayor tamaño, tanto en términos de fidelidad en la explicación de las recomendaciones como en la precisión de las recomendaciones generadas. Su enfoque basado en contenido le permitió adaptarse mejor a las preferencias de los miembros del grupo y ofrecer recomendaciones más acertadas en grupos de mayor tamaño.

Además, se investigó el impacto de los umbrales mínimos de soporte y confianza de las reglas de asociación generadas en la fidelidad. Se observó que a medida que se

incrementa el umbral de confianza y de soporte, la fidelidad disminuye debido a la reducción en el número de reglas de asociación utilizadas para la explicación de las recomendaciones.

En cuanto al impacto de la similitud entre los miembros del grupo en la fidelidad, el experimento ha encontrado que no existe una correlación clara. Esto indica que el modelo es robusto y puede generar explicaciones coherentes, independientemente de las diferencias de similitud entre los usuarios del grupo. Estos hallazgos respaldan la aplicabilidad del modelo en diferentes contextos y grupos de usuarios.

2. Trabajos futuros

El presente estudio sobre el rendimiento y la explicabilidad del modelo de recomendación grupal ha proporcionado importantes conclusiones, pero también ha abierto la puerta a posibles trabajos futuros que podrían ampliar y mejorar la investigación en este campo. A continuación, se presentan algunas sugerencias para futuras investigaciones:

1. Desarrollo de métricas de explicabilidad específicas para recomendaciones grupales: Actualmente, la mayoría de las métricas de explicabilidad se centran en explicar las recomendaciones a nivel individual. Sería beneficioso desarrollar métricas que evalúen la calidad y la comprensibilidad de las explicaciones generadas para grupos de usuarios. Estas métricas podrían tener en cuenta factores como la coherencia entre las explicaciones de los diferentes miembros del grupo y la representatividad de las preferencias grupales.
2. Exploración de enfoques de explicación más avanzados: Aunque se han investigado diferentes enfoques de explicación, como el filtrado de preferencias de usuarios y la imputación de calificaciones faltantes, aún queda espacio para explorar enfoques más avanzados. Por ejemplo, el uso de técnicas de aprendizaje automático interpretable, como reglas de decisión o árboles de decisión, podría brindar explicaciones más intuitivas y comprensibles para los usuarios del grupo.
3. Evaluación de la satisfacción del usuario con las explicaciones grupales: Además de evaluar la calidad técnica de las explicaciones generadas, es importante

comprender cómo los usuarios perciben y utilizan estas explicaciones en la toma de decisiones. Realizar estudios empíricos para evaluar la satisfacción y comprensión de los usuarios con las explicaciones grupales proporciona información valiosa sobre su utilidad y eficacia en la práctica.

4. Incorporación de preferencias contextuales en las explicaciones: En muchos casos, las preferencias y decisiones de un grupo pueden estar influenciadas por el contexto en el que se encuentran. Sería interesante explorar cómo se puede capturar y utilizar la información contextual en la generación de explicaciones grupales. Esto podría implicar considerar factores como la ubicación geográfica, el momento del día o las condiciones ambientales al proporcionar explicaciones personalizadas y relevantes para el grupo.

Estos trabajos futuros podrían contribuir al avance de la explicabilidad de los sistemas de recomendación grupal, promoviendo una mayor confianza y comprensión por parte de los usuarios y facilitando una toma de decisiones más informada.

SIGLAS

AI: *Artificial Intelligence*

AVG: *Average*

CB-RS: *Content-Based Recommender System*

CB: *Content-Based*

CF-RS: *Collaborative Filtering Recommender System*

CF: *Collaborative Filtering*

EDT: *Estructura de Desglose de Trabajo*

GDPR: *Reglamento General de Protección de Datos*

GRS: *Group Recommender System*

ICF: *Item-Based Collaborative Filtering*

IKNN: *Item-Based K Nearest Neighbor*

LIME: *Local Interpretable Model-Agnostic Explanations*

LM: *Least Misery*

MED: *Median*

MP: *Most Pleasure*

nDGC: *Normalized Discounted Cumulative Gain*

PCA: *Principal Component Analysis*

PrefRS: *Recommender System based on Preferences Aggregation*

RS: *Recommender System*

RecRS: *Recommender System based on Recommendation Aggregation*

SVD: *Singular Value Decomposition*

TFIDF: *Term Frequency-Inverse Document Frequency*

TFM: *Trabajo de Fin de Máster*

UCF: *User-Based Collaborative Filtering*

UKNN: *User-Based K Nearest Neighbor*

XAI: *Explainable Artificial Intelligence*

XRS: *Explainable Recommender System*

BIBLIOGRAFÍA

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.
- Abdollahi, B., & Nasraoui, O. (2016). Explainable matrix factorization for collaborative filtering. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 5-6).
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734-749.
- Aggarwal, C. C. (2016). *Recommender systems (Vol. 1)*. Cham: Springer International Publishing.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in knowledge discovery and data mining*, 12(1), 307-328.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115.
- Baltrunas, L., Makcinskas, T., & Ricci, F. (2010, September). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 119-126).
- Beliakov, G., Calvo, T., & James, S. (2015). Aggregation functions for recommender systems. *Recommender systems handbook*, 777-808.
- Berggren, C. (2019). The cumulative power of incremental innovation and the role of project sequence management. *International journal of project management*, 37(3), 461-472.
- Billsus, D. & Pazzani, M. J. (1998), *Learning Collaborative Information Filters*, 46--54.
- Billsus, D. & Pazzani, M.J. (2004). User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction*, 10, 147-180.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132.

-
- Breese, J.S., Heckerman, D. and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, May 1998, 43-52.
- Buckley, C. & Salton, G. (1995). Optimization of Relevance Feedback Weights.. In E. A. Fox, P. Ingwersen & R. Fidel (eds.), SIGIR (p./pp. 351-357), : ACM Press. ISBN: 0-89791-714-6.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12, 331-370.
- Burke, Robin & Robin (2007). Hybrid Web Recommender Systems. LNCS. 4321. 10.1007/978-3-540-72079-9_12.
- Castro, J., Barranco, M. J., Rodríguez, R. M., & Martínez, L. (2018). Group recommendations based on hesitant fuzzy sets. *International Journal of Intelligent Systems*, 33, 2058–2077.
- Castro, J., Yera, R., & Martínez, L. (2017). An empirical study of natural noise management in group recommendation systems. *Decision Support Systems*, 94, 1–11.
- Castro, J., Yera, R., Alzahrani, A. A., Sanchez, P., Barranco, M., & Martínez, L. (2019). A big data semantic driven context aware recommendation method for questionanswer items. *IEEE Access*, 7, 182664–182678.
- Chanson, A., Labroche, N., & Verdeaux, W. (2021). Towards local post-hoc recommender systems explanations. In *Proceedings of the 23rd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP)*.
- Chao, D. L., Balthrop, J., & Forrest, S. (2005, November). Adaptive radio: achieving consensus using negative preferences. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work* (pp. 120-123).
- De Pessemier, T., Doms, S., & Martens, L. (2014). Comparison of group recommendation algorithms. *Multimedia Tools and Applications*, 72, 2497–2541.
- Devi, T. R., & Reddy, V. S. (2012). Work breakdown structure of the project. *Int J Eng Res*
- Fisher, G,. (2001). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction* 11, 65–86.
- García, A. & Martínez, J. (2022). Avances en explicabilidad de sistemas de inteligencia artificial. *Revista de Inteligencia Artificial*, 10(2), 45-68.

-
- Gedikli, F., & Jannach, D. (2013). Improving recommendation accuracy based on item-specific tag preferences. *ACM Transactions on Intelligent Systems and Technology*, 4, 11.
- Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 61-70.
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *European conference on information retrieval*, 345-359.
- Grčar, M., Fortuna, B., Mladenič, D., & Grobelnik, M. (2006). kNN versus SVM in the collaborative filtering framework. In *Data Science and Classification* (pp. 251-260).
- Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4), 1-19.
- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241-250).
- Hofmann, Thomas. (2003). Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis. *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*. 259-266. 10.1145/860435.860483.
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based indicators of IR performance.
- Kim, J. K., Kim, H. K., Oh, H. Y., & Ryu, Y. U. (2010). A group recommendation system for online communities. *International journal of information management*, 30(3), 212-219.
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37.
- Lang, K. (1995). Newsweeder: Learning to filter netnews. In *Machine Learning Proceedings 1995* (pp. 331-339). Morgan Kaufmann.
- Mahmood, T., & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia* (pp. 73-82).
- Masthoff, J. (2004). Group modeling: Selecting a sequence of television items to suit a group of viewers. *Personalized Digital Television: Targeting Programs to individual Viewers*, 93-141.
-

-
- McCarthy, J. F. (2002). Pocket restaurantfinder: A situated recommender system for groups. In *Workshop on mobile ad-hoc communication at the 2002 ACM conference on human factors in computer systems* (Vol. 8).
- McCarthy, J. F., & Anagnost, T. D. (1998). MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (pp. 363-372).
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *Ai & Society*, 35, 957-967.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill, New York.
- Montaner, M., López, B., & Esteva, J.L. (2004). A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*, 19, 285-330.
- O'Connor, M., Cosley, D., Konstan, J. A., & Riedl, J. (2001). PolyLens: A recommender system for groups of users. In *ECSCW 2001* (pp. 199-218).
- Pazzani, M., Billsus, D., (1997). Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine Learning* 27, 313–331.
- Peake, G., & Wang, J. (2018). Explanation mining: Post hoc interpretability of latent factor models for recommendation systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2060-2069).
- Resnick, P. & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40, 56-58.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
- Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning* (pp. 791-798).
- Salton, G. (ed.) (1988). *Automatic text processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Sarwar, B., Karypis, G., Konstan, J. A., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web, WWW 2001* (pp. 285-295).
- Schafer, Ben & J, Ben & Frankowski, Dan & Dan, & Herlocker, & Jon, & Shilad, & Sen, Shilad. (2007). *Collaborative Filtering Recommender Systems*.
-

-
- Schafer, J.B., Konstan, J.A. and Riedl, J. (2001). E-Commerce Recommendation Applications. *Data Mining and Knowledge Discovery*, 5, 115-153.
- Senot, C., Kostadinov, D., Bouzid, M., Picault, J., Aghasaryan, A., & Bernier, C. (2010). Analysis of strategies for building group profiles. In *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings 18* (pp. 40-51). Springer Berlin Heidelberg.
- Shmaryahu, D., Shani, G., & Shapira, B. (2020, January). Post-hoc Explanations for Complex Model Recommendations using Simple Methods. In *IntRS@ RecSys* (pp. 26-36).
- Stern, D. H., Herbrich, R., & Graepel, T. (2009, April). Matchbox: large scale online bayesian recommendations. In *Proceedings of the 18th international conference on World wide web* (pp. 111-120).
- Troiano, L., & Díaz, I. (2014). A model for preserving privacy in recommendation systems. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 56-65).
- Yera, R., & Martínez, L. (2017). Fuzzy tools in recommender systems: A survey. *International Journal of Computational Intelligence Systems*, 10, 776–803.
- Zhong, J., & Negre, E. (2022). Towards improving user-recommender systems interactions. In *2022 IEEE/SICE International Symposium on System Integration (SII)* (pp. 816-820). IEEE.
- Zitnick, C.L., & Kanade, T. (2004). Maximum Entropy for Collaborative Filtering. *ArXiv*, abs/1207.4152.