



Universidad  
Técnica de  
Cotopaxi

**UNIVERSIDAD TÉCNICA DE COTOPAXI**

**FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS  
CARRERA DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS  
COMPUTACIONALES**

**PROYECTO DE INVESTIGACIÓN**

**“ESTUDIO DE ALGORITMOS DE FILTRADO COLABORATIVO PARA  
SISTEMAS RECOMENDADORES DE INFORMACIÓN”**

**Autor:**

Chuquitarco Chasiluisa Jenny Alexandra

**Tutor:**

Ing. Edwin Edison Quinatoa Arequipa

Latacunga - Ecuador

Agosto - 2018



### DECLARACIÓN DE AUTORÍA

Yo **CHUQUITARCO CHASILUISA JENNY ALEXANDRA** declaro ser autor del presente proyecto de investigación: **“ESTUDIO DE ALGORITMOS DE FILTRADO COLABORATIVO PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”**, siendo el Ing. Edwin Quinatoa, el tutor del presente trabajo; y eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

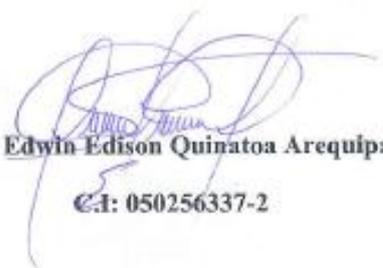
  
**Chuquitarco Chasiluisa Jenny Alexandra**  
C.I: 050300904-5



### **AVAL DEL TUTOR DE PROYECTO DE INVESTIGACIÓN**

En calidad de tutor de investigación sobre el título: **"ESTUDIO DE ALGORITMOS DE FILTRADO COLABORATIVO PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN"**. De **CHUQUITARCO CHASILUISA JENNY ALEXANDRA** de la Carrera de Ingeniería en Informática y Sistemas Computacionales, considero que dicho informe investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos a evaluación del Tribunal de Validación de Proyecto de Investigación que el **Consejo Directivo de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi** designe, para su correspondiente estudio y calificación.

Latacunga, 31 de Julio de 2018

  
**Ing. Edwin Edison Quinatoa Arequipa**

**C.I: 050256337-2**



### APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueben el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la Facultad de Ciencias de la Ingeniería y Aplicadas, por cuanto, el postulante: **Chuquitarco Chasiluisa Jenny Alexandra** con C.I. 050300904-5, con el título de Proyecto de Investigación: **"ESTUDIO DE ALGORITMOS DE FILTRADO COLABORATIVO PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN"**, ha considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, 31 de julio del 2018

Para constancia firman:

Lector 1  
Ing. Myrian Iza

C.I: 050195761-7

Lector 2  
Ing. Victor Medina

C.I: 050137395-5

Lector 3  
Ing. Jorge Rubio

CI: 050222229-2



## AGRADECIMIENTO

Me gustaría agradecer a todos mis maestros ya que ellos me enseñaron valorar los estudios y a superarme cada día, también agradezco a mis padres, esposo e hijo porque ellos estuvieron en los días más difíciles de mi vida durante esta etapa.

Y agradezco a Dios por darme mucha paciencia y sabiduría, también agradezco a mi querida empresa Fénix Corp por la confianza que han puesto en mí para lograr uno de mis objetivos.

Jenny.

## **DEDICATORIA**

Dedico este proyecto de investigación a mis padres quienes me apoyaron incondicionalmente todo este tiempo.

A mi esposo Diego y a mi hijo Alejandro quienes me dieron la fuerza para continuar, cuando parecía que me iba a rendir.

A mis docentes quienes nunca desistieron al enseñarme y continuar depositando su esperanza en mí

A todos los que me apoyaron para concluir este proyecto de investigación, para ellos es esta dedicatoria, pues es a ellos a quienes les debo por su apoyo incondicional.

Jenny.

## ÍNDICE DE CONTENIDO

PROYECTO DE INVESTIGACIÓN.....	i
DECLARACIÓN DE AUTORÍA .....	<b>¡Error! Marcador no definido.</b>
AVAL DEL TUTOR DE PROYECTO DE INVESTIGACIÓN .....	<b>¡Error! Marcador no definido.</b>
APROBACIÓN DEL TRIBUNAL DE TITULACIÓN.....	<b>¡Error! Marcador no definido.</b>
AGRADECIMIENTO .....	v
DEDICATORIA.....	vi
ÍNDICE DE CONTENIDO .....	vii
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS .....	xi
RESUMEN .....	xii
AVAL DE TRADUCCIÓN.....	xiv
1. INFORMACIÓN GENERAL: .....	1
2. RESUMEN DEL PROYECTO .....	2
3. JUSTIFICACIÓN DEL PROYECTO .....	3
4. BENEFICIARIOS DEL PROYECTO .....	4
5. EL PROBLEMA DE INVESTIGACIÓN .....	4
6. OBJETIVOS .....	4
6.1. Objetivo general .....	4
6.2. Objetivos específicos.....	5
7. ACTIVIDADES Y SISTEMA DE TAREAS EN RELACIÓN A LOS OBJETIVOS .....	5
8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA .....	6
8.1. Antecedentes .....	6
8.2. Sistemas de Recomendación basados en Filtrado Colaborativo .....	8
8.3. Técnicas de filtrado colaborativo .....	9
8.3.1. Filtrado Colaborativo basado en memoria .....	10

8.3.1.1.	Métricas de similaridad .....	12
8.3.1.1.1.	Índice Jaccard .....	12
8.3.1.1.2.	Correlación de Pearson .....	13
8.3.1.	Filtrado Colaborativo basado en modelos .....	14
8.3.1.1.	Algoritmos de agrupamiento .....	15
8.3.1.2.	K-Means .....	17
8.3.1.3.	Técnicas de factorización matricial (MF) .....	18
8.3.1.4.	Robust SVD (RobSVD) .....	18
8.3.	Librerías para sistemas recomendadores de información .....	18
8.4.1.	Crab .....	18
8.4.2.	LenksKit .....	19
8.4.3.	MyMediaLite .....	19
8.4.4.	RecommenderLab .....	19
8.5.	Implementaciones para Big Data .....	20
8.5.1.	LIBMF .....	20
8.5.2.	MLlib .....	21
8.5.3.	Comparación entre implementaciones .....	21
9.	HIPÓTESIS .....	21
10.	METODOLOGÍA DE LA INVESTIGACIÓN .....	21
10.2.	Diseño narrativo .....	22
11.	ANÁLISIS DE RESULTADOS .....	22
11.1.	Comparación entre implementaciones de herramientas .....	22
11.2.	Data sets .....	23
11.2.1.	MovieLens .....	24
11.2.2.	Jester .....	24
11.2.3.	Lastfm .....	25
11.2.4.	Bookcrossing .....	25

11.2.5. Netflix .....	26
11.2.6. Yahoo! Webscope.....	26
11.3. Análisis comparativo del sistema de plusvalía (alquiler y venta de bienes raíces).....	27
11.4. Análisis comparativo de los Data sets.....	30
12. IMPACTOS TÉCNICOS Y SOCIALES .....	31
12.1. Impacto técnico .....	31
12.2. Impacto social .....	32
12.3. Impacto ambiental.....	32
12.4. Impacto económico .....	32
13. PRESUPUESTO PARA EL PROYECTO DE INVESTIGACIÓN .....	32
14. CONCLUSIONES .....	34
15. RECOMENDACIONES .....	35
16. BIBLIOGRAFÍA.....	36
15. ANEXOS.....	38

## ÍNDICE DE TABLAS

Tabla 1. Sistema de tareas en relación a los objetivos planteados. ....	5
Tabla 2. Categorías de Filtrado Colaborativo.....	8
Tabla 3. Algoritmo de K-Means Clustering .....	17
Tabla 4. Librerías para sistemas de recomendación .....	22
Tabla 5. Estados del arte.....	23
Tabla 6. Datos de MovieLens.....	24
Tabla 7. Tabla de valoraciones de la herramienta Jester .....	24
Tabla 8. Cuadro comparativo que detalla el análisis de la aplicación web. ....	29
Tabla 9. Cuadro comparativo de los datasets .....	31

## ÍNDICE DE FIGURAS

Figura 1. Diagrama de técnicas de filtrado colaborativo.....	12
Figura 2. Fórmula que calcula la similitud entre dos usuarios o elementos.....	11
Figura 3. Aplicación web de Amazon .....	12
Figura 4. Aplicación web del supermercado Tía.....	13
Figura 5. Valoración de aplicación móvil .....	14
Figura 6. Formación del vecindario a partir de agrupamiento .....	16

# UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS.

**TÍTULO:** “Estudio de algoritmos de filtrado colaborativo para sistemas recomendadores de información”

**Autor:** Chuquitarco Chasiluisa Jenny Alexandra

## RESUMEN

En la actualidad la gran cantidad de información y de servicios/productos ofrecidos a los usuarios en la red, hace cada vez más difícil las tareas de filtrar entre tanta información y poder encontrar aquello que se ajuste a los gustos y/o necesidades de cada uno. En los últimos 20 años los sistemas recomendadores se han vuelto cada vez más complejos, tratándose de ajustarse lo máximo posible al perfil y necesidades de los distintos usuarios dando lugar a algoritmos basados en costosos cálculos computacionales. Es por ello que esta investigación proporciona información relevante acerca de la estructura y funcionamiento del algoritmo colaborativo y como este influye a potenciar sistemas de filtrado de información el cual presenta resultados en algunas empresas que registran de manera automática los intereses de los usuarios, recopilando las preferencias o gustos. Para la elaboración del presente estudio se utilizó el método de investigación cualitativa. Basándose en esta metodología de investigación se pretende explicar el concepto científico original de algoritmos de filtrado colaborativo, sin alterar su fundamento teórico y que de este modo el lector pueda entender e interpretar el documento de estudio. Finalmente se obtendrá mediante tablas comparativas los aspectos más relevantes, referente a los algoritmos de filtrado colaborativo.

**Palabras claves:** Algoritmos, Filtrado colaborativo, Predicción, Similitud, Sistemas Recomendadores.

# UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS.

**TOPIC:** “Study of collaborative filtering algorithms for information recommending systems”

**Author:** Chuquitarco Chasiluisa Jenny Alexandra

## ABSTRACT

Now, the large amount of information and services products offered to users in the network makes it difficult to filter the information and find what fits the tastes and / or needs for each one. In the last 20 years, the recommender systems have become increasingly complex, trying to adjust as much as possible to the profile and needs of different users, giving rise to algorithms based on costly computational calculations. That is why this research provides relevant information about the structure and functioning of the collaborative algorithm and how it influences to enhance information filtering systems which presents results in some companies that automatically register the interests of users, collecting preferences or likes. For the elaboration of the present study, the qualitative research method was used. Based on this researching methodology, the aim is to explain the original scientific concept of collaborative filtering algorithms, without altering its theoretical foundation and thus allowing the reader to understand and interpret the study document. Finally, the most relevant aspects will be obtained through comparative tables, referring to the collaborative filtering algorithms.

**Keywords:** Algorithms, Collaborative filtering, Prediction, Similarity, Recommender Systems.



## *AVAL DE TRADUCCIÓN*

En calidad de Docente del Idioma Inglés del Centro de Idiomas de la Universidad Técnica de Cotopaxi; en forma legal CERTIFICO que: La traducción del resumen del proyecto de investigación al Idioma Inglés presentado por el señorita Egresada de la Carrera de Ingeniería en Informática y Sistemas Computacionales: CHUQUITARCO CHASILUISA JENNY ALEXANDRA, cuyo título versa “ESTUDIO DE ALGORITMOS DE FILTRADO COLABORATIVO PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”, lo realizo bajo mi supervisión y cumple con una correcta estructura gramatical del Idioma.

Es todo cuanto puedo certificar en honor a la verdad y autorizo al peticionario hacer uso del presente certificado de la manera ética que estimaren conveniente.

Latacunga, julio del 2018

Atentamente,

  
Msc. Alison Mena Barthelotty  
DOCENTE CENTRO DE IDIOMAS  
C.C. 0501801252



## **1. INFORMACIÓN GENERAL:**

### **Título del proyecto:**

Estudio de Algoritmos de Filtrado Colaborativo para Sistemas Recomendadores de Información.

### **Fecha de Inicio:**

Marzo de 2018

### **Fecha de finalización:**

Agosto de 2018

### **Lugar de ejecución:**

Latacunga -Ecuador

### **Facultad que auspicia:**

Facultad de Ciencias de la Ingeniería y Aplicadas

### **Carrera que auspicia:**

Ingeniería en Informática y Sistemas Computacionales

### **Equipo de Trabajo:**

### **Datos personales del coordinador de proyecto de investigación:**

Nombres: Edwin

Apellidos: Quinatoa

Fecha de nacimiento: 13 de abril del 1981

Teléfonos: 0998542905

Correo electrónico: edwin.quinatoa@utc.edu.ec

### **Datos Personales del Autor:**

Nombres: Jenny Alexandra

Apellidos: Chuquitarco Chasiluisa

Fecha de nacimiento: 21 de diciembre de 1990

C.C: 0503009045

Teléfono: 0984019544

Correo electrónico: jenny.chuquitarco@gmail.com

**Área de conocimiento:**

Red de estudio cuantitativa

**Línea de investigación:**

Tecnologías de la Información y Comunicación (TICS) y Diseño Gráfico.

**Sub línea de investigación de la Carrera:**

Robótica e Inteligencia Artificial.

**2. RESUMEN DEL PROYECTO**

Una de las técnicas de recomendación de información es el filtrado colaborativo que generalmente sugieren elementos que han gustado a usuarios con preferencias similares. Por esta razón, se recolectan preferencias sobre los elementos y a partir de ellas se calculan las recomendaciones. En el diseño de estas aplicaciones se deben tener en cuenta algunos aspectos entre los cuales se destacan: cómo se recolectan las preferencias, qué se sugiere a los nuevos usuarios dado que se cuenta con poca información (cold start), la privacidad de la información y el volumen de datos (Betarte, Machano, Molina, 2016).

Mediante el estudio realizado en esta investigación, información que ha sido tomada de fuentes bibliográficas y a su vez analizada se ha determinado que los algoritmos de Filtrado Colaborativo son los más utilizados para sistemas recomendadores de información, el cual utiliza a usuarios con gustos similares para hacer recomendaciones, dicho estudio se demostrará, mediante la comparación de aplicaciones que cuentan con este tipo de filtrado colaborativo. En este documento, se estudian, los métodos más representativos de cada categoría. Luego, se presenta un estudio exhaustivo, a través de experimentos en tareas de

predicción de ratings y recomendación de ítems mediante una línea de tiempo en diferentes contextos que ofertan los Data set.

### **Palabras claves**

- Algoritmos, Filtrado colaborativo, Predicción, Similitud, Sistemas Recomendadores.

### **3. JUSTIFICACIÓN DEL PROYECTO**

A partir del auge de los sitios de distribución de contenido multimedia un considerable número de sistemas de recomendación han sido introducidos para mejorar el tratamiento de los usuarios. Entre ellos se encuentra el sistema de recomendación de FilmAffinity, que presenta contenidos multimedia asociados al área del cine. Este sistema cuenta con una base de datos con fichas técnicas detalladas de cuantiosa cantidad de películas, documentales, cortometrajes y series de televisión. Su funcionamiento se basa en el cálculo de la media de puntuaciones realizadas por los usuarios tanto a cada contenido como a sus críticas (Rojas, 2014).

Los impactos que genera la utilización de algoritmos colaborativos en el desarrollo de sistemas recomendadores prometen revolucionar la interacción entre las empresas y el usuario, siendo una gran herramienta para facilitar la navegación en búsqueda de lo que concretamente se necesita conocer, su aplicación al campo laboral potencia la entrega de servicios a gran velocidad a los clientes finales.

En la actualidad la cantidad excesiva de información ofrecidos a los usuarios en la red, hace cada vez más difícil las tareas de filtrar entre tanta información y poder encontrar aquello que se ajuste a los gustos y/o necesidades de cada uno, razón por la cual los algoritmos de filtrado colaborativo resultan ser los mejores para sistemas recomendadores.

Los resultados emitidos serán de gran utilidad para el grupo de investigación de la facultad de CIYA de la Universidad Técnica de Cotopaxi, se les presenta detalladamente como estos sistemas recomendadores ayudan a mejorar la experiencia del usuario al momento de utilizar aplicaciones informáticas. Su representación, evaluación y resultados se presenta mediante un cuadro comparativo en base al historial de empresas que han implementado los algoritmos de filtrado colaborativo hasta la fecha, para ayudar a la experiencia del usuario al momento de mostrar información de sus gustos o preferencias.

La presente investigación muestra como el algoritmo de filtrado colaborativo es una gran herramienta para futuras implementaciones en sistemas recomendadores, el cual servirá para mejorar la visibilidad y la forma de dar a conocer nuevos productos o servicios a más usuarios.

#### **4. BENEFICIARIOS DEL PROYECTO**

##### **Beneficiarios directos:**

Como beneficiario directo son los investigadores de la facultad de CIYA de la Universidad Técnica de Cotopaxi.

##### **Beneficiarios indirectos.**

Todos los docentes y estudiantes de la Universidad Técnica de Cotopaxi perteneciente a la facultad de ciencias de Ingeniería y Aplicadas de la carrera de Ingeniería en Informática y Sistemas Computacionales.

#### **5. EL PROBLEMA DE INVESTIGACIÓN**

Debido a la información masiva que poseen los Investigadores de la Universidad Técnica de Cotopaxi resulta problemático clasificar contenidos de acuerdo a una necesidad específica de búsqueda, en la mayoría de casos resulta exhaustivo encontrar algo importante ya que no se cuenta con sugerencias aproximadas a lo que se requiere. Así como la mayoría de aplicaciones informáticas empleadas no cuentan con sistemas recomendadores.

Es decir, no le proveen al usuario la facilidad de recibir recomendaciones de productos o servicios que le interesa, provocando una reacción negativa generalmente causando que el usuario abandone la aplicación.

#### **6. OBJETIVOS**

##### **6.1. Objetivo general**

- Analizar los algoritmos de filtrado colaborativo mediante la revisión en fuentes bibliográficas para determinar las características más relevantes de la investigación.

## 6.2. Objetivos específicos

- Revisar fuentes bibliográficas acerca de los Algoritmos de filtrado colaborativo.
- Analizar las técnicas que emplea el filtrado colaborativo que permitirá establecer sus características de funcionamiento.
- Elaborar cuadros comparativos que contemple los resultados obtenidos en la investigación.

## 7. ACTIVIDADES Y SISTEMA DE TAREAS EN RELACIÓN A LOS OBJETIVOS

La tabla 1 que se presenta a continuación describe la correlación directa entre cada uno de los objetivos, las actividades que los mismos involucran, los resultados de estas actividades y sus respectivos medios de verificación con el fin de obtener una secuencia lógica en el desarrollo del presente proyecto de investigación.

**Tabla 1.** Sistema de tareas en relación a los objetivos planteados.

<b>Objetivo</b>	<b>Actividad (tareas)</b>	<b>Resultado de la actividad</b>	<b>Medios de Verificación</b>
Revisar fuentes bibliográficas acerca de los algoritmos de filtrado colaborativo.	<ul style="list-style-type: none"> <li>• Recopilación de información bibliográfica de:</li> <li>• Libros, artículos científicos, tesis, páginas web.</li> <li>• Lectura crítica de la información recogida.</li> <li>• Documentación de la fundamentación teórica del proyecto de investigación</li> </ul>	<ul style="list-style-type: none"> <li>• Fundamentación teórica del proyecto de investigación.</li> </ul>	<ul style="list-style-type: none"> <li>• Marco teórico</li> <li>• Tabla de comparación</li> </ul>
Analizar los métodos que emplea el filtrado colaborativo que	<ul style="list-style-type: none"> <li>• Determinar que técnica se emplea para el buen funcionamiento de</li> </ul>	<ul style="list-style-type: none"> <li>• Obtención de las técnicas más adecuadas para el funcionamiento de</li> </ul>	<ul style="list-style-type: none"> <li>• Detalle de técnicas presentadas en el documento.</li> </ul>

permitirá establecer sus características de funcionamiento.	filtrado colaborativo de información.	los filtrados colaborativos.	
Elaborar cuadros comparativos que contemple los resultados obtenidos en la investigación.	<ul style="list-style-type: none"> <li>Resaltar los aspectos más importantes de la investigación realizada.</li> </ul>	<ul style="list-style-type: none"> <li>Presentación de resultados mediante cuadros comparativos del estudio.</li> </ul>	<ul style="list-style-type: none"> <li>Detalle de técnicas presentadas en el documento.</li> </ul>

**Fuente:** La investigadora

## 8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA

### 8.1. Antecedentes

Los Sistemas de Recomendación (Recommender Systems, RecSys) son técnicas de filtrado de información que nacen con el objetivo de ayudar a los usuarios a tomar decisiones frente a grandes cantidades de información. En las últimas décadas se han estudiado y desarrollado diferentes tipos de técnicas que varían en términos de la información utilizada y la forma en que realizan recomendaciones. (A. & B, 1999)

Los sistemas de recomendación basados en el filtrado colaborativo aparecen después de los sistemas de recomendación basados en el contenido. En un principio los sistemas de recomendación basados en el contenido fueron de gran utilidad para una primera aproximación de filtrado de información, pero presentaban un inconveniente: no podían medir la calidad de las recomendaciones que realizaban. Frente a este problema, a principios de los noventa aparecieron dos posibles alternativas:

- a. Esperar que técnicas de inteligencia artificial mejorasen la clasificación automática de los documentos o ítems.
- b. Introducir la opinión de las personas en el proceso de las recomendaciones.

Con la adopción del punto “a”, se mejoraban los algoritmos de recomendación basados en el contenido, mientras que el desarrollo del punto “b” produjo la creación de sistemas de

recomendación social o de filtrado colaborativo. El primer sistema que permitió añadir opiniones de los usuarios fue el llamado “Tapestry”. El sistema permite almacenar opiniones o anotaciones de los usuarios sobre los contenidos de mensajes como un tipo de meta información. Por su parte, el sistema brindaba a los usuarios la posibilidad de realizar búsquedas sobre el contenido de un documento, así como de la meta información producida por los usuarios. (Rodríguez, 2016).

De manera clásica se distinguen seis principales clases o tipos de sistemas de recomendación.

**Basados en el Contenido:** Elaboran recomendaciones centrándose en la información que pueden extraer de los objetos. Por ejemplo, el género o los actores en una película.

**Filtrado Colaborativo:** Utilizan a usuarios con gustos afines u objetos similares para estimar recomendaciones.

**Basados en el Contexto:** Recomiendan según las características generales de los usuarios como la edad, género, ciudad, etc.

**Basados en Conocimiento:** Consideran las necesidades e intereses del usuario para realizar recomendaciones.

**Basados en Comunidades:** Recomiendan objetos en función de las preferencias de los amigos de los usuarios. Siguen el refrán, “Dime con quién andas y te diré quién eres”.

**Introducción Híbridos:** Es la combinación de dos o más enfoques mencionados anteriormente.

Esta investigación se encuentra enfocada en sistemas de recomendación basados en Filtrado Colaborativo (Collaborative filtering, CF). Existen tres fases principales en el funcionamiento de los sistemas colaborativos:

1. El sistema crea un perfil de cada usuario con sus preferencias respectivas.
2. Se mide el grado de similitud entre los distintos usuarios dentro del sistema y se crean grupos de usuarios con gustos parecidos.
3. El sistema estima la preferencia futura de un usuario en base a la información recopilada anteriormente.

Dentro de los sistemas de recomendación basados en filtrado colaborativo existen dos clasificaciones principales, los basados en memoria y los basados en modelos.

Los métodos basados en memoria utilizan métricas de similitud para encontrar usuarios con gustos parecidos y realizar recomendaciones. Las métricas más comunes son el coeficiente de correlación de Pearson y la similitud coseno.

Los métodos basados en el modelo se dedican a buscar patrones en los datos para estimar o aprender un modelo que sea capaz de realizar predicciones. El modelo puede ser un algoritmo de data mining o de machine learning. Las técnicas de filtrado colaborativo basadas en el modelo incluyen redes bayesianas, modelos de factores latentes, o clustering, entre otros. (Torres, 2015)

Las técnicas más representativas de cada categoría, junto a sus principales ventajas y desventajas, se presentan en la Tabla 2 (Galán, 2014).

**Tabla 2.** Categorías de Filtrado Colaborativo

	<b>BASADOS EN MEMORIA</b>	<b>BASADOS EN MODELOS</b>
<b>Técnicas</b>	– <b>K vecinos más cercanos (kNN)</b>	– <b>Clustering</b> – <b>Factores Latentes</b> – <b>Métodos Espectrales</b>
<b>Ventajas</b>	– <b>Rápida implementación</b> – <b>Permiten agregar datos fácilmente y de manera incremental</b>	– <b>Mayor escalabilidad</b> – <b>Superan problemas de escasez de datos</b>
<b>Desventajas</b>	– <b>El desempeño decrece en matrices dispersas</b> – <b>No escalan bien en data sets a gran escala</b>	– <b>Construir el modelo es costoso</b> – <b>Puede existir pérdida de información útil</b>

**Fuente:** (Torres, 2015)

Los algoritmos de filtrado colaborativo surgieron como resultado de los problemas que tenía el filtrado basado en contenido, siendo el pionero en tomar calificaciones proporcionado por el usuario para en un futuro consecuente permita conocer los gustos que el usuario califico.

## **8.2. Sistemas de Recomendación basados en Filtrado Colaborativo**

Los sistemas de filtrado colaborativo son sistemas de recomendación que utilizan la información que los usuarios aportan sobre los ítems para realizar las recomendaciones. Así pues, tratan de predecir las valoraciones de un (varios) ítem(s) para un usuario en particular

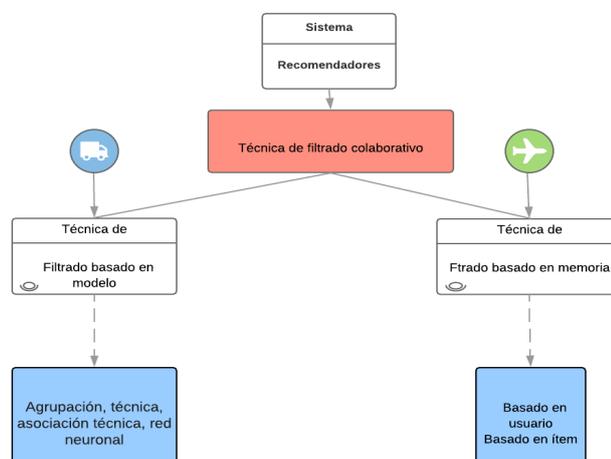
(llamado usuario activo) mediante el uso de ítems valorados por otros usuarios, es decir, teniendo en cuenta la opinión de los otros usuarios. Como consecuencia, estos sistemas se basan en las tendencias marcadas por grupos de personas con los mismos gustos o preferencias que el usuario activo para realizar las recomendaciones o sugerencias.

Para un individuo o usuario cuyos gustos o tendencias son similares a los de un grupo de personas determinado, se establece que este usuario pasa a formar parte de este grupo. Por consiguiente, si una o varias personas del grupo valoran un nuevo ítem de forma positiva, el sistema de filtrado colaborativo establece que es muy probable que los demás miembros del grupo al que pertenece(n), también valoren positivamente el nuevo ítem. Los sistemas de filtrado colaborativo, por lo general, se adaptan al siguiente funcionamiento: (Seguido, 2014)

- Los usuarios elaboran valoraciones sobre ítems.
- Se analizan las valoraciones realizadas por los usuarios con tal de establecer grupos o vecinos cercanos de usuarios con preferencias similares.
- Una vez obtenidos los grupos de vecinos para un usuario, se realizan las recomendaciones al usuario activo teniendo en cuenta sus vecinos más cercanos y sus correspondientes valoraciones.

### 8.3. Técnicas de filtrado colaborativo

El filtrado colaborativo identifica patrones de interconexiones de usuario para hacer recomendaciones específicas Roturas colaborativas de filtrado hacia abajo en dos enfoques primarios basados en memoria y basados en modelos (Ver figura 3).



Fuente: (A. & B, 1999)

### 8.3.1. Filtrado Colaborativo basado en memoria

GroupLens Research fue uno de los primeros en estudiar los sistemas de recomendación automatizados, que mantienen la esencia de Tapestry, pero implementan importantes mejoras en las métricas utilizadas a la hora de predecir. Se establece que no todas las calificaciones de los usuarios son igualmente útiles, dándoles mayor importancia a usuarios con opiniones similares (Research, 2018)

El algoritmo de los K-vecinos más cercanos (kNN, k Nearest Neighbors) es el primer sistema de filtrado colaborativo automatizado. Fue presentado por GroupLens en su motor de recomendación Usenet, un sistema de discusión distribuido dentro de una organización donde los usuarios leían y publicaban mensajes, considerado el precursor de los foros en Internet que se utilizan extensamente en la actualidad. Posteriormente, Ringo y BellCore, emplearon variantes del mismo método en sus recomendaciones de música y vídeo (Torres, 2015)

Los algoritmos basados en memoria utilizan toda la base de datos de usuarios y objetos para generar predicciones. Cada usuario es parte de un grupo de personas con intereses similares. Identificando ese grupo de usuarios, llamado vecindario, se pueden combinar sus preferencias para realizar predicciones.

El algoritmo de los vecinos más cercanos (kNN) prevalece entre los métodos de filtrado colaborativo basados en memoria. Según sea el caso, puede ser utilizado para encontrar los k usuarios más cercanos (USER-BASED) o los k objetos más cercanos (ITEM-BASED). De aquí en más, se estudiará el algoritmo kNN pensando en los usuarios más cercanos. De manera análoga, puede ser aplicado sobre los ítems. Las etapas son las siguientes: a) calcular la similitud, que refleja la distancia, correlación, o peso, entre dos usuarios. b) encontrar los k usuarios más cercanos al usuario activo. c) elaborar una predicción ponderando todos los ratings entre usuarios afines. (Hodges, 2013)

El enfoque basado en la memoria toma los datos de calificación del usuario para calcular similitudes entre usuarios y artículos para hacer una recomendación.

El enfoque basado en la memoria más famoso es algoritmos basados en vecindad para calcular la relación entre cualquiera de los artículos o usuarios. Aquí se predice una recomendación para un usuario basada sobre clasificaciones de elementos similares (o vecinos) por el mismo usuario. El algoritmo basado en vecindad calcula la similitud entre dos

usuarios o elementos que producen una predicción para el usuario tomando el promedio ponderado de todas las calificaciones. Una famosa forma de calcular la similitud entre dos usuarios  $x$ ,  $y$  es usar el coseno similar después de identificar a los usuarios más similares de  $K$  superior a un usuario activo, las matrices de elementos de usuario correspondientes se agregan a elegir el conjunto de elementos que se recomendarán. Comúnmente se llama el algoritmo  $K$ -vecino más cercano (KNN) (Aberger, 2014).

Si bien es relativamente simple componer y explicar estos algoritmos el rendimiento se degrada a medida que los datos se vuelven cada vez más escaso y no suele funcionar bien en grandes conjuntos de datos, que ocurren con frecuencia en la práctica.

**Figura 1.** Fórmula que calcula la similitud entre dos usuarios o elementos

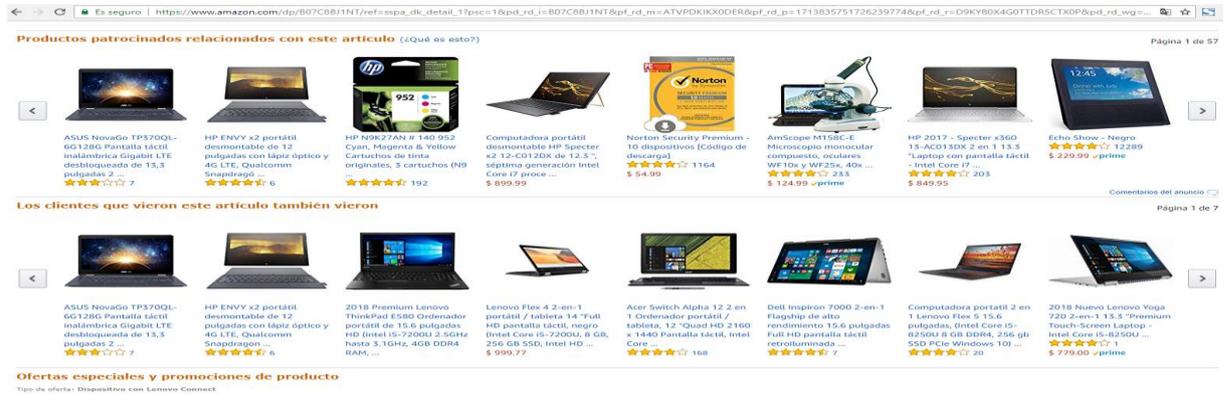
$$similarity(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}}$$

**Fuente:** (Aberger, 2014).

Existen muchos desafíos para las tareas de filtrado colaborativo basado en memoria ya que se requiere que los algoritmos tengan la capacidad de manejar datos muy dispersos, escalar con el creciente número de usuarios y elementos, hacer recomendaciones satisfactorias en un corto período de tiempo y tratar otros problemas como la sinonimia (la tendencia del mismo o elementos similares para tener diferentes nombres), ataques de chelín, ruido de datos y problemas de protección de privacidad.

Los sistemas de filtrado colaborativo basado en memoria de la primera generación, como GroupLens, usan los datos de clasificación del usuario para calcular la similitud o el peso entre usuarios o elementos y hacen predicciones o recomendaciones de acuerdo con esos valores de similitud calculados, estos algoritmos se implementan notablemente en sistemas comerciales como <http://www.amazon.com/> (vea un ejemplo en la Figura 3) y Barnes and Noble, porque son fáciles de usar para su ejecución y altamente efectivos. También promete una mayor lealtad de los clientes, mayores ventas, más ingresos publicitarios y el beneficio de las promociones específicas (Moya, 2013)

**Figura 2.** Aplicación web de Amazon



Fuente: (Amazon, 2018)

### 8.3.1.1. Métricas de similitud

En primer lugar, es fundamental encontrar a los usuarios con gustos afines al usuario activo. Para ello, se utilizan distintas funciones que miden el grado de similitud, correlación o distancia de los ratings entre usuarios. A continuación, se revisarán las métricas más comunes en sistemas de recomendación. (Jaccard, 2016)

#### 8.3.1.1.1. Índice Jaccard

Es uno de los primeros métodos propuestos para medir similitud entre usuarios. Utilizando el coeficiente de Jaccard es posible conocer la similitud (o disimilitud) que existe entre el conjunto de ítems calificados por un usuario con respecto a otro. (Jaccard, 2016)

Entonces, el índice de Jaccard entre el usuario activo  $a$  y un usuario  $u$  dentro del sistema, es el siguiente:

$$sim_{a,u} = Jaccard(a, u) = \frac{|I_a \cap I_u|}{|I_a \cup I_u|}$$

Donde  $I_a \cap I_u$  corresponde al conjunto de ítems evaluados tanto por el usuario activo  $a$  como por el usuario  $u$ . Por otro lado,  $I_a \cup I_u$  es el conjunto de ítems evaluados por uno u otro usuario (o ambos).

Un ejemplo de la utilización del índice de Jaccard es la aplicación web del supermercado Tía el cual le permite al usuario agregar artículos al carrito de compras sin importar la calificación

que tenga este por el contrario le presenta artículos que solo se parezcan, delimitando características que ayuden a un filtrado más específico, siendo el tiempo de respuesta 0.002 segundos según la búsqueda realizada.

**Figura 3.** Aplicación web del supermercado Tía



Fuente: (tia, 2018)

La limitante del coeficiente de Jaccard es que no considera el rating de cada ítem, si éste fue evaluado o no. Esto implica que, si un usuario en particular ha calificado una gran cantidad de ítems, según esta métrica, tendrá afinidad con la mayoría de los usuarios dentro del sistema.

### 8.3.1.1.2. Correlación de Pearson

Es la primera formulación estadística aparecida para el filtrado colaborativo (Castellano, 2017). La correlación que existe entre el usuario activo  $a$  y un usuario  $u$  dentro del sistema es

$$\text{sim}_{a,u} = \frac{\sum_{i \in \mathcal{I}_a \cap \mathcal{I}_u} (r_{a,i} - \bar{r}_a) \cdot (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in \mathcal{I}_a \cap \mathcal{I}_u} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in \mathcal{I}_a \cap \mathcal{I}_u} (r_{u,i} - \bar{r}_u)^2}}$$

Donde  $r_{a,i}$  y  $r_{u,i}$  son los ratings que los usuarios  $a$  y  $u$  emitieron sobre el ítem  $i$ . En tanto,  $\bar{r}_a$  y  $\bar{r}_u$  son los promedios de los ratings entregados por los usuarios  $a$  y  $u$  sobre todos los ítems que ambos calificaron. La sumatoria se realiza sobre  $i \in \mathcal{I}_a \cap \mathcal{I}_u$ , es decir, el conjunto de ítems calificados tanto por el usuario  $a$  como por el usuario  $u$  (Castellano, 2017)

La idea detrás de este cálculo es que el índice de correlación entre el usuario activo y el resto de los usuarios se utilice para asignarle más o menos peso a las calificaciones. El valor del índice de correlación varía en el intervalo  $[-1, 1]$ , es decir:

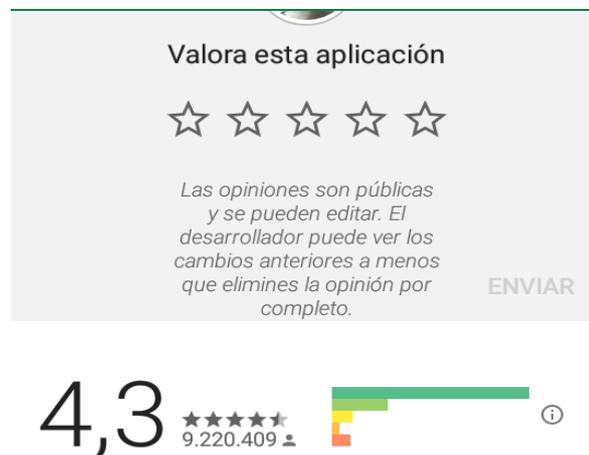
- Si es cercano a 1, significa que los gustos del usuario activo  $a$  y del usuario  $u$  están muy relacionados, por lo tanto, las opiniones de ese usuario en particular tendrán más peso en el cálculo de la predicción.

- Si el índice es cercano a  $-1$ , las opiniones de ese usuario influirán en sentido contrario cuando se generen predicciones, o sea, las votaciones positivas del usuario se considerarán negativas y viceversa.
- Si el índice es 0, las variables no tienen relación y por ende los gustos de ese usuario no se utilizarán en la estimación de ratings.

Otros coeficientes que miden el grado de correlación son: Spearman, similar a Pearson, excepto que los ratings son ordenados en un ranking y reemplazados por su posición respectiva; y el Coeficiente de Kendall, parecido a Spearman pero utiliza rangos relativos para calcular la correlación (Orseti, 2016)

La aplicación play store de Android hace uso del índice de la correlación de Pearson, al indicar al usuario que valore el funcionamiento de las aplicaciones que se encuentran alojadas ahí, esto permite que otros usuarios interesados en la temática tengan acceso a las aplicaciones con mayor ranking y dando como resultado un tiempo de respuesta de 0.001 segundo en presentar nuevas aplicaciones similares a la calificada según la búsqueda hecha.

**Figura 4.** Valoración de aplicación móvil



Fuente: (Google, 2018)

### 8.3.1. Filtrado Colaborativo basado en modelos

Esta técnica utiliza los datos para crear un modelo a través del cual establecen el conjunto de usuarios similares al usuario activo. El diseño y desarrollo del modelo permite al sistema aprender y reconocer patrones complejos presentes en los datos, y entonces, realizar predicciones a través del modelo aprendido. Las técnicas de filtrado colaborativo basadas en

modelos que se analizarán son: algoritmos de agrupamiento, métodos de descomposición matricial y métodos espectrales.

### 8.3.1.1. Algoritmos de agrupamiento

Los algoritmos de agrupamiento (en inglés, clustering) consisten en congregar a usuarios y/o ítems de acuerdo a sus ratings. Estas técnicas asumen que hay grupos de usuarios (o ítems) con similares características y, por tanto, una vez que el usuario ha sido asignado a un grupo, la predicción o recomendación se obtiene a partir de las puntuaciones otorgadas por otros miembros del mismo conjunto. Un clúster es una colección de datos, agrupados de acuerdo a algún criterio. Este criterio puede representar la distancia o similitud. Para medir la cercanía entre los datos se utiliza, por lo general, una función de la familia de distancias Minkowski. (Grané, 2015)

La distancia Minkowski de orden  $p$  entre dos puntos

$$x = (x_1, x_2, \dots, x_n) \text{ e } Y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$$

Está definida como:

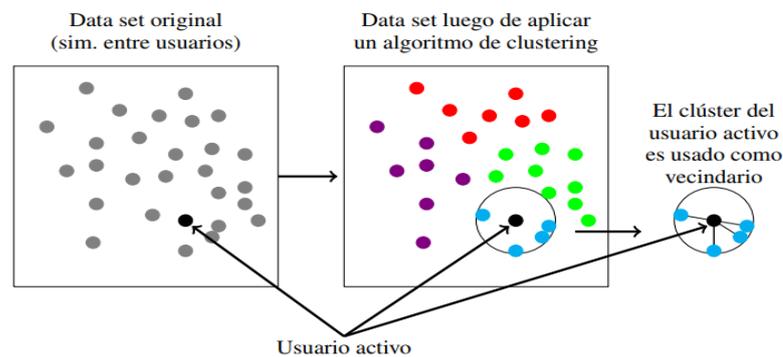
$$d(X, Y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

Cuando  $p = 1$ ,  $d$  es la distancia Manhattan; para  $p = 2$ ,  $d$  es la distancia Euclidiana.

Clustering ha sido aplicado en filtrado colaborativo para identificar grupos de usuarios con ratings similares o relacionados. Se han utilizado diferentes técnicas de agrupación para dividir los datos en clústers de usuarios con votaciones semejantes. Posteriormente, la predicción se genera empleando algún algoritmo clásico de filtrado colaborativo, solo que esta vez, se usa el clúster del usuario activo como vecindario. (Sarwar, 2016)

A continuación, se explica el proceso de agrupamiento y formación del vecindario (Ver figura1)

**Figura 5.** Formación del vecindario a partir de agrupamiento



**Fuente:** (Torres, 2015)

Formalmente este método se divide en tres etapas:

1. Aplicar el algoritmo de clustering para producir  $p$  particiones de usuarios. El conjunto de datos  $A$  es dividido en  $A_1, A_2, \dots, A_p$ , donde  $A_i \cap A_j = \emptyset$ , para  $1 \leq i, j \leq p$ ; y  $A_1 \cup A_2 \cup \dots \cup A_p = A$ .
2. Determinar el vecindario para un usuario a dado. Si  $a \in A_i$  entonces la partición entera  $A_i$  es usada como vecindario.
3. Una vez obtenido el vecindario, se utiliza un algoritmo clásico de filtrado colaborativo para predecir calificaciones. Los métodos de clustering pueden ser clasificados dentro de tres categorías: jerárquicos, basados en densidad y de particionamiento.

Clustering basado en densidad normalmente se encarga de buscar grupos con gran concentración de objetos separados por regiones dispersas que representan el ruido. (Torres, 2015)

La agrupación jerárquica crea una descomposición jerárquica del conjunto de datos utilizando algún criterio. Por último, la idea detrás de los métodos de particionamiento es dividir a los usuarios en grupos y usar estos grupos como vecindarios. Luego de haber utilizado el algoritmo de clustering se tienen calificaciones más densas de los usuarios hacia los ítems. Entonces, la matriz original de ratings, antes dispersa, se ha transformado en una matriz sin tantos espacios vacíos. (Torres, 2015)

Un ejemplo claro de clustering es CEAGA que es una agrupación industrial, creada en 1997, integrada por las empresas del Sector de Automoción de Galicia para posibilitar el desarrollo de proyectos de carácter cooperativo orientados a la mejora competitiva, cuenta con una

aplicación web en base a las características de su personal administrativo y sus clientes quienes mantienen una integración solida de necesidades.

### 8.3.1.2. K-Means

K-Means Clustering es un algoritmo iterativo en el cual los objetos se mueven a través de un conjunto de clústers hasta que se alcanza el ajuste deseado. K-means se utiliza para reducir el espacio de búsqueda, creando  $k$  grupos compuestos por ítems con características similares o usuarios con gustos afines.

Este método inicia seleccionando arbitrariamente  $k$  objetos como centroide de los  $k$  grupos. Luego, cada elemento es asignado al clúster con su centroide más cercano y se vuelve a calcular la media del grupo considerando a los nuevos elementos. Esa media es ahora considerada como el nuevo centroide del clúster. Posteriormente se calcula nuevamente la similitud de cada objeto y se asigna al centroide más cercano, recalculando la media y repitiendo el proceso de manera iterativa hasta que se alcance el mínimo deseado. Finalmente, se logra tener objetos dentro de cada grupo con características similares entre ellos, pero disímiles con elementos de otros grupos. La idea principal se resume en la tabla 3, el Algoritmo 1. (Ng, 2015)

**Tabla 3.** Algoritmo de K-Means Clustering

---

**Algoritmo 1 K-MEANS CLUSTERING**

---

- 1: Escoger al azar  $k$  puntos
  - 2: Establecer esos puntos como centroides iniciales de los  $k$  clústers
  - 3: **repeat**
  - 4:     **for each** punto  $p$  **do**
  - 5:         Asignar  $p$  al clúster con centroide más cercano
  - 6:         Recalcular el centroide del clúster al cual fue asignado  $p$
  - 7:     **end for**
  - 8: **until** Los centroides no varíen
- 

**Fuente:** (Ng, 2015)

Como ejemplo tenemos a mlpack que es una biblioteca de aprendizaje de máquina rápida y flexible, escrita en C ++, que tiene como objetivo proporcionar implementaciones rápidas y extensibles de algoritmos de aprendizaje automático de vanguardia. Mlpack proporciona estos algoritmos como simples programas de línea de comandos, enlaces de Python y clases de C ++ que luego pueden integrarse en soluciones de aprendizaje automático a gran escala.

### **8.3.1.3. Técnicas de factorización matricial (MF)**

Los modelos de factores latentes abordan el filtrado colaborativo con la intención de descubrir características ocultas en los ratings que expliquen el comportamiento de usuarios y objetos. Incluyen pLSA, redes neuronales, Latent Dirichlet Allocation y modelos que factorizan la matriz de ratings. Las técnicas de factorización o descomposición matricial han demostrado ser exitosas en sistemas de recomendación y han ganado popularidad debido a tres razones principales. Permiten lograr buena escalabilidad, proveen predicciones más precisas y entregan flexibilidad en el modelo (Nepomuceno, 2012)

### **8.3.1.4. Robust SVD (RobSVD)**

En general, los métodos de factorización matricial no garantizan robustez en el modelo. Son altamente susceptibles a valores atípicos que pueden pasar inadvertidos en términos numéricos, pero ocasionar importantes alteraciones Hawkins propone un método robusto a través del algoritmo AL1-SVD, donde la idea es estimar secuencialmente los vectores propios usando la norma L1.

Este método presenta dos diferencias principales con respecto al SVD. La primera, tiene que ver con la ortogonalidad. En el SVD convencional, todos los vectores propios son ortogonales. En este algoritmo todos son (en general) no ortogonales. La otra diferencia es que en el SVD clásico los vectores propios son calculados en orden decreciente según el valor propio. Lo cual no se cumple, necesariamente, en AL1. Se han utilizado diferentes funciones para asegurar robustez en el modelo. Zhang, se enfocaron en la función de Huber porque es fácil de implementar y rápida en cómputos. Además, desarrollaron el eficiente algoritmo iterative reweighted least squares (IRLS) para solucionar el problema de minimización. Este algoritmo, que puede ser visto como un proceso de regularización robusto, actualiza iterativamente las matrices. (Crab, 2018)

## **8.3. Librerías para sistemas recomendadores de información**

### **8.4.1. Crab**

Crab es un motor de recomendación rápido y flexible para Python que integra los algoritmos clásicos de información para la filtración de información en el mundo de los paquetes científicos de Python (numpy, scipy, matplotlib). El objetivo del motor es proporcionar un

amplio conjunto de componentes a partir de los cuales se puede construir un sistema de recomendación personalizado a partir de un conjunto de algoritmos (Crab, 2018)

#### **8.4.2. LensKit**

LensKit es un software gratuito y de código abierto, disponible bajo los términos de la licencia GNU Lesser General Public.

El desarrollo de LensKit comenzó en GroupLens Research en la Universidad de Minnesota y ahora está coordinado por People and Information Research Team (PIReT) en Boise State University .

- Múltiples proyectos de investigación de GroupLens , particularmente MovieLens y BookLens , usan LensKit en producción, proporcionando experiencia, informes de fallas y otros valiosos comentarios (y códigos)

Además, provee un entorno flexible para realizar evaluaciones de las recomendaciones. (lenskit, 2018)

#### **8.4.3. MyMediaLite**

MyMediaLite es una biblioteca de sistema recomendada para Common Language Runtime (CLR, a menudo llamado .NET).

Aborda los dos escenarios más comunes en el filtrado colaborativo:

- Predicción de calificación (por ejemplo, en una escala de 1 a 5 estrellas), y
- Predicción de elementos a partir de comentarios positivos (p. ej., clics, "me gusta" o acciones de compra).

MyMediaLite es software libre (software de fuente abierta), puede ser utilizado y distribuido bajo los términos de la Licencia Pública General de GNU (GPL) (mymedialite, 2018)

#### **8.4.4. RecommenderLab**

RecommenderLab Un marco para desarrollar y probar algoritmos de recomendación (Paquete R), este este paquete R proporciona una infraestructura para probar y desarrollar algoritmos de recomendación. El paquete admite conjuntos de datos de clasificación (por ejemplo, de 1 a 5 estrellas) y unario (0-1). Los algoritmos soportados son: (recommenderlab, 2018)

- Filtrado colaborativo basado en el usuario (UBCF)
- Filtrado colaborativo basado en elementos (IBCF)
- SVD con imputación media de columna (SVD)
- Funk SVD (SVDF)
- Recomendación de recomendación basada en reglas (AR)
- Artículos populares (POPULARES)
- Elementos elegidos aleatoriamente para comparación (ALEATORIO)
- Recomiéndonos los artículos que le gustan (RERECOMMEND)
- Recomendaciones híbridas (HybridRecommender)

Para la evaluación, el marco admite protocolos given-n y all-but-x con

- Tren / prueba dividida
- Validación cruzada
- Repetición de muestreo de arranque

Las medidas de evaluación son:

- Errores de calificación: MSE, RMSE, MAE
- Recomendaciones de Top-N: TPR / FPR (ROC), precisión y recuperación

## **8.5. Implementaciones para Big Data**

Las librerías y paquetes presentados anteriormente son útiles cuando se trabaja con conjuntos de datos en pequeña y mediana escala. Si bien esta investigación no está enfocada en grandes conjuntos de datos, es importante mencionar implementaciones que sean capaces de proveer soluciones al análisis de datos en “Big Data”.

### **8.5.1. LIBMF**

Es una herramienta de código abierto utilizada para aproximar matrices dispersas a través del producto entre dos matrices en un espacio latente. Esta factorización matricial es comúnmente utilizada en filtrado colaborativo para reducir la dimensionalidad de la matriz de entrada y

evitar trabajar con todo el conjunto de datos. Las principales características de LIBMF son: Además de las técnicas tradicionales de factorización matricial, se agregan factores de normalización para corregir el sesgo y mejorar el rendimiento de las recomendaciones. Permite el paralelismo multinúcleo. Esto hace de este paquete sumamente eficiente, acelerando las operaciones vectoriales. Como referencia, para sets de datos de 250M ratings, a LIBMF le toma menos de ocho minutos converger en un mínimo razonable.

Las principales características de LIBMF son:

- Además de las técnicas tradicionales de factorización matricial, se agregan factores de normalización para corregir el sesgo y mejorar el rendimiento de las recomendaciones.
- Permite el paralelismo multinúcleo. Esto hace de este paquete sumamente eficiente, acelerando las operaciones vectoriales. Como referencia, para sets de datos de 250M ratings, a LIBMF le toma menos de ocho minutos converger en un mínimo razonable

### **8.5.2. MLlib**

Es una implementación de funciones comunes de máquinas de aprendizaje. Actualmente, cubre cuatro tipos de problemas de aprendizaje automático: clasificación binaria, regresión, clustering y filtrado colaborativo. Dentro del filtrado colaborativo, MLlib soporta métodos basados en el modelo, en el cual los usuarios y productos son descritos por un conjunto de factores latentes que pueden ser usados para predecir datos faltantes. En particular, se encuentra implementado el algoritmo de Mínimos Cuadrados Alternantes (ALS) para solucionar el problema de optimización.

### **8.5.3. Comparación entre implementaciones**

Para la visualización de resultados de la comparación entre implementación de los sistemas recomendadores de filtrado colaborativo se presenta en el punto 11 (Análisis y resultados).

## **9. HIPÓTESIS**

¿El éxito de las búsquedas de información obedece al uso de algoritmos de filtrado colaborativo en sistemas recomendadores?

## **10. METODOLOGÍA DE LA INVESTIGACIÓN**

### **10.1. Investigación cualitativa**

Para la elaboración del presente estudio se utilizó el método de investigación cualitativa. Basándose en esta metodología de investigación se pretende explicar el concepto científico original de algoritmos de filtrado colaborativo, sin alterar su fundamento teórico y que de este modo el lector pueda entender e interpretar el documento de estudio.

## 10.2. Diseño narrativo

Mediante esta investigación se detalla la funcionalidad que presentan el algoritmo colaborativo teniendo como fundamentación teórica, artículos y referencias bibliográficas, que permite establecer cómo se puede emplear este algoritmo en un sistema de recomendación que facilite al usuario con las búsquedas que realice.

## 11. ANÁLISIS DE RESULTADOS

### 11.1. Comparación entre implementaciones de herramientas

En este punto se procede a dar a conocer el análisis de los resultados obtenidos en la presente investigación, utilizando la comparación de herramientas para sistemas de recomendación de filtrado colaborativo.

Tabla 4: Entrega un resumen de las librerías o paquetes de software detallados anteriormente: Crab, LensKit, MyMediaLite, RecommenderLab.

**Tabla 4.** Librerías para sistemas de recomendación

Librerías	Lenguaje	Versión	Fecha de actualización
Crab	Python	0.1	01 marzo 2012
MyMediaLite	C#	3.11	14 febrero 2017
LensKit	Java	2.2.1	08 junio 2018
RecommenderLab	R	0.2.3	1 julio 2018

**Fuente.** La Investigadora

Se puede apreciar LensKit, RecommenderLab están en constante actualizaciones. Esto es una gran ventaja con respecto al resto, considerando la rapidez que progresan los algoritmos cada año.

Tabla 5: Indica que algoritmos de filtrado colaborativo, estudiados en las herramientas se encuentran implementados por las diferentes librerías.

**Tabla 5.** Estados del arte

<b>Librerías</b>	<b>UBCF</b>	<b>IBCF</b>
Crab	Si	Si
MyMediaLite	Si	Si
LensKit	Si	Si
RecommenderLab	Si	Si

**Fuente.** La Investigadora

En general, todas las librerías cuentan con los métodos de filtrado colaborativo tradicionales basados en usuarios más cercanos (USER-BASED) y objetos más cercanos (ITEM-BASED).

Se ha optado por RecommenderLab, del entorno estadístico y gráfico R, debido a las siguientes principales razones:

- El ambiente de la librería facilita tanto el desarrollo de algoritmos, como la evaluación y comparación entre métodos.
- Cuenta con las principales técnicas de filtrado colaborativo y las métricas comúnmente empleadas para evaluar.
- R dispone de una gran cantidad de funciones que pueden ser adaptadas y utilizadas en la implementación de nuevos algoritmos.
- Posee una extensa documentación, tanto la librería en particular, como el lenguaje en que se programa.
- Libera actualizaciones frecuentes, con variedad de guías y manuales disponibles.

## 11.2. Data sets

Los sistemas de recomendación emplean Benchmark data para comparar el desempeño de sus algoritmos. Para realizar los experimentos se utilizarán conjuntos de datos ampliamente estudiados en la literatura. A continuación, se revisarán los Data Sets de mediana escala más comunes dentro de los sistemas de recomendación de algunas empresas globales que utilizan algoritmos de filtrado colaborativo.

### 11.2.1. MovieLens

MovieLens es un sitio web en donde la gente califica películas y el sistema recomienda de acuerdo a los gustos de las personas, es decir, es un recomendador basado en filtrado colaborativo. Forma parte de uno de los proyectos de investigación desarrollados por GroupLens y cuenta con cientos de miles de usuarios registrados que han entregado ratings sobre películas. Los datos han sido recolectados durante varios periodos de tiempo, dependiendo del tamaño del conjunto.

MovieLens ha publicado dos Data Sets estables hasta la fecha agosto del 2017, los cuales se detallan a continuación:

**Tabla 6.** Datos de MovieLens

Aplicaciones	Clasificaciones	Películas	Usuarios	Fecha
-	465.000	27.000	465.000	04/2015
1.300	100.000	9.000	700	10/2016
750.000	26.000.000	45.000	270.000	08/2017

**Fuente.** La Investigadora

Los Data sets de MovieLens, han sido largamente estudiados bajo diversos propósitos incluyendo sistemas de recomendación. Gran parte de las primeras publicaciones de filtrado colaborativo utilizaron estos datos para evaluar sus algoritmos, por lo que gozan de una gran popularidad y permiten comparar los resultados obtenidos con la literatura. (Latest, 2018)

### 11.2.2. Jester

Jester es un sistema de recomendación de bromas desarrollado en la Universidad de Berkeley (UC Berkeley) para estudiar el filtrado de información social.

**Tabla 7.** Tabla de valoraciones de la herramienta Jester

Conjuntos	Valoraciones	Bromas	Usuarios	Fecha
Conjunto de datos 1	más de 4.1 millones	(-10.00 a +10.00) de 100	73.421	Abril de 1999 y mayo de 2003
Conjunto de datos 2	más de 1.7 millones	(-10.00 a +10.00) de 150	59.132	Noviembre de 2006 y mayo de 2009.

Conjunto de datos 2+	de con más de 500,000 de		79,681	Noviembre de 2006 a noviembre de 2012
----------------------	--------------------------	--	--------	---------------------------------------

**Fuente.** La Investigadora

Los conjuntos de datos 2 y 2+ referente a un rango de 3 a 6 años presenta una valoración de 20 +. (Dataset, 2018)

### 11.2.3. Lastfm

Lastfm es un servicio para descubrir nueva música que ofrece recomendaciones personalizadas basándose en la música que escucha el usuario.

La curiosidad sobre la actuación relativa de los artistas más sonados con el tiempo. Entonces, graficamos los mejores seis artistas desde 2009 (Vea Anexo 4, Figura 8)

David Bowie es el gran ganador aquí, pasando del sexto lugar en 2009 al primer lugar hoy como mi artista más clasificado (desde la suscripción de Last.fm). Tomar en cuenta que estos datos desglosados difieren ligeramente de los recuentos totales de reproducciones por artista anterior, porque Last.fm aquí descarta los recuentos totales de reproducción de iTunes que primero scrobbed al registrarse. Por lo tanto, los scrobbed desagregados con fechas solo representan juegos de canciones posteriores a la inscripción además no todos los usuarios usan las aplicaciones de Last.fm

### 11.2.4. Bookcrossing

Bookcrossing (BC) es una iniciativa que comenzó en marzo de 2001 en Estados Unidos. Básicamente consiste en un Club de libros a nivel global que atraviesa el tiempo y el espacio, que no conoce límites geográficos y cuya meta, simplemente, es convertir el mundo en una Biblioteca (Vea Anexo 4, Figura 9)

En este punto, se puede observar un cuadro simple de recomendación basado en la popularidad basada en el recuento de calificaciones de los usuarios para diferentes libros. Es evidente que los libros escritos por JK Rowling son bastante populares.

### 11.2.5. Netflix

Netflix en los últimos 5-10 años ha capturado una gran población de espectadores. Con más espectadores, lo más probable es que aumente la variedad de espectáculos. Sin embargo, ¿entiende la gente la distribución de las calificaciones en los shows de Netflix?

Los espectadores de Netflix pueden darse cuenta de que muchos programas son de 2015 y más recientes. A continuación, hay evidencia de respaldo, basada en la muestra, que la mayoría de los espectáculos son de 2015 y más recientes (Ver Anexo 4, Figura 10)

Netflix muestra por año por clasificación en la figura representa una agrupación de clasificaciones (cada color de barra) por año. Es interesante observar que el número de videos de TV-14 se ha incrementado en 5 veces en 2016 desde 2012. No es sorprendente ver que el 82% de los programas de TV-MA que se han lanzado, fueron lanzados en los últimos 2 años (Ver Anexo 4, Figura 11)

La cantidad liberada que domina el mercado es TV-14, sin embargo, TV-MA tiene la mayor tasa de aumento en los últimos 5 años.

### 11.2.6. Yahoo! Webscope

Yahoo! Research provee una gran variedad de data sets recolectados desde todos los servicios que ofrece la firma, esto se enmarca en el programa Webscope. El Programa Yahoo! Webscope8 es una biblioteca de referencia de interesantes y científicamente útiles conjuntos de datos para uso no-comercial por parte de investigadores. La enorme cantidad de data sets ofrecidos por Yahoo! se dividen en distintos tipos y se detallan a continuación:

- Language Data: Este tipo de datos pueden ser utilizados para recuperación de información y algoritmos de procesamiento de lenguaje natural.
- Graph and Social Data: Se utilizan para trabajo con matrices, grafos, clustering, y algoritmos de aprendizaje automático.
- Ratings and Classification Data: Conjuntos de datos muy útiles en filtrado colaborativo, sistemas de recomendación y algoritmos de aprendizaje automático.
- Competition Data: Estos datos se utilizaron en competencias con académicos e investigadores. Dentro de los datos se encuentran conjuntos de mediana escala que han sido utilizados para evaluar sistemas de recomendación.

### **11.3. Análisis comparativo del sistema de plusvalía (alquiler y venta de bienes raíces)**

Describimos a continuación algunos detalles sobre nuestro estudio, y luego seguimos con una descripción de nuestros principales hallazgos para lo cual se aplicó las técnicas de filtrado en memoria y la técnica de filtrado en modelo para la comparación de caso práctico de la página oficial de la plusvalía Ecuador. Con fines de respaldar el presente estudio se creó una versión ligera del servicio de viviendas cuyo objetivo principal de este ejemplo fue familiarizar con los principios del filtrado colaborativo y su comportamiento en una gran cantidad de datos.

Para respaldar el ejemplo se ha tomado como base al sistema de Plusvalía que es el mayor portal de bienes raíces en Ecuador. La plusvalía es una normativa que pretende regular el valor del suelo al vender bienes inmuebles y de este modo evitar una ganancia ilegítima (Ver Anexo, Figura 12)

En base a la cantidad de información que oferta el sistema de plusvalía se ha encontrado un déficit al momento de recomendar ciertos viene que puede estar relacionado a los gustos de los usuarios quienes navegan por encontrar cierta información de interés anteriormente interactuados con otros usuarios de gustos similares y a la ves filtrar la información en base a los costos bajos y la cantidad de votaciones.

La aplicación se trata de compra y alquiler de viviendas donde los usuarios pueden interactuar con la información proporcionada que son almacenadas por todos los usuarios registrados. Una vez que hay suficiente información disponible, la aplicación tiene la capacidad de calcular la similitud de las viviendas con mayor votación, mayor me gustas, y precios accesibles. A cada usuario se le asigna las viviendas más queridas en la base de datos. En función de esto, se recomiendan viviendas en la página de inicio.

Para demostrar la técnica de filtrado en memoria, los usuarios deberán calificar y dar me gusta a la vivienda. Para demostrar la técnica de filtrado en modelo, los usuarios deberán buscar viviendas recomendadas de acorde a sus precios establecidos, ya que necesitar de mayor datos disponibles y similares a otros usuarios que le gusten y la votación sea mayor

Este ejemplo esta realizado en Framework Laravel ya que intenta aliviar el dolor del desarrollo al facilitar las tareas comunes utilizadas en la mayoría de los proyectos web, como la autenticación, el enrutamiento, las sesiones, las colas y el almacenamiento en caché. Es

accesible, pero potente, y proporciona herramientas necesarias para aplicaciones grandes y robustas.

**Escenarios:**

1. Usuario ingresa al sistema
2. Usuario se registra en el sistema
3. Usuario inicia sesión
4. Sistema le da la bienvenida
5. Sistema presenta información de viviendas existentes.
  - a. Si no existe viviendas, el usuario puede crear viviendas
  - b. El usuario crear nueva vivienda, requiere completar toda la información.
  - c. Sistema procesa información
  - d. Sistema presenta información
6. Usuario califica la vivienda de 1 al 5
7. El usuario da me gusta a la vivienda
8. El usuario puede dar no me gusta a la vivienda
9. Sistema procesa información para presentar información.
10. El usuario puede ver información completa de vivienda
  - a. Sistema presenta información de la vivienda
  - b. Sistema presenta cantidad y detalle de usuarios que le guste esa vivienda
  - c. Sistema presenta cantidad y detalle de usuarios que no le guste esa vivienda
11. El usuario elimina información de vivienda
12. El usuario realiza recomendaciones de vivienda
  - a. El sistema procesa información
  - b. El sistema recomienda viviendas más requeridas.
  - c. Sistema presenta información
    - i. De viviendas con mayor votación
    - ii. De viviendas con precios accesible
    - iii. Cantidad y detalle de usuarios que le guste o no esa vivienda
13. Usuario cierra sesión

En el siguiente gráfico se presenta la página principal de ejemplo práctico para recomendar viviendas (Ver Anexo 4, Figura 13)

La siguiente figura presenta la página para que puedan registrar los usuarios (Ver Anexo 5, Figura 14)

Inicio de sesión de aplicación web (Ver Anexo 4, Figura 15)

El siguiente gráfico presenta el listado de viviendas, se recomienda viviendas con mayor votación y mayor me gustas ofertada por usuarios que han interactuado con dicha información (Ver Anexo 4, Figura 16)

La siguiente figura muestra la página para ingresar nueva vivienda el cual será para ofertar e interactuar con los usuarios de acuerdo a su preferencia de gustos (Ver Anexo, Figura 17)

A continuación, se indica página para calificar, dar me gusta, eliminar y ver información de la vivienda con todos los usuarios que les gusten (Ver Anexo 4, Figura 18)

El siguiente gráfico se presenta la página que visualiza todos los usuarios que les guste o hayan dejado de gustar cierta información (Ver Anexo 4, Figura 19)

Por siguiente se obtiene la página para ver toda la información de la vivienda y la opción eliminar información de vivienda (Ver Anexo 4; Figura 20, 21)

La siguiente figura presenta la página donde se recomienda viviendas a los usuarios de acuerdo a sus referencias de me gustas, votaciones y precios accesibles (Ver Anexo 4, Figura 22)

En último lugar se presenta el código fuente donde se realiza el ejemplo práctico para el sistema recomendada de viviendas (Ver Anexo 4, Figura 23)

El siguiente cuadro comparativo presenta las 10 viviendas interactuadas por distintos usuarios, donde se puede observar la cantidad de me gustas, no me gustas y las calificaciones obtenidas. Para obtener los datos, se procedió a interactuar el sistema con 10 usuarios registrados en el sistema, los cuales dieron su respectiva ponderación a las viviendas.

**Tabla 8.** Cuadro comparativo que detalla el análisis de la aplicación web.

#	Tipo de casa	Vivienda	Me gusta por usuarios	No me gusta por usuario	Calificaciones	Técnica empleada
1	Casa de dos pisos de color	007	8	0	5	Basado en

	verde					modelo/ memoria
2	Suite privado	006	7	2	5	Basado en modelo/memoria
3	Suite por estrenar	002	6	2	4	Basado en modelo/memoria
4	departe maneto por estrenar	005	5	1	3	Basado en modelo/memoria
5	Casa de 1 piso	008	4	0	3	Basado en modelo/memoria
6	Casa de 3 pisos color blanco	009	3	2	2	Basado en memoria/modelo
7	Departamento completo	003	2	3	2	Basado en memoria/modelo
8	Casa de 8 pisos color verde	004	2	0	1	Basado en memoria/modelo
9	Departamento en 2 piso	010	1	0	0	Basado en memoria/modelo
10	Cas de 1 piso color rojo	001	0	0	0	Basado en memoria

**Fuente:** La investigadora

#### **11.4. Análisis comparativo de los Data sets**

Con la información recolectada de referencias bibliográficas por parte de la investigadora, a continuación, se realizará el siguiente cuadro comparativo con las empresas MovieLens, Jester, Lastfm, Bookcrossing y Netflix. En el cual se compara en base al desempeño de la empresa, para la distribución de la tabla se realizará en base a los siguientes parámetros: los datos obtenidos para la comparación se realizó mediante los dataset (Conjunto de Datos) de cada empresa datos a conocer en sus cuentas oficiales, en la columna referencia está el nombre de la empresa, en la columna tipo de calificador se referencia a que la calificación obtenida se puede realizar con ayuda del usuario o del propio sistema, la columna N° se determina el número de calificador, en calificación se describe la cantidad de calificación obtenidas por el calificador, en fecha se da a conocer la última actualización de los dataset de

la empresa y el tipo de modelo determinar la referencia que posee el algoritmo recomendador colaborativo los cuales son de memoria y modelo.

Con lo cual se pretende recomendar un tipo de modelo para futuros sistemas recomendadores colaborativos, pero se debe tomar en cuenta que los basado en memoria utiliza datos recogidos para calcular la similitud entre los usuarios o elementos comunes mientras que el basado en modelo elabora un modelo de las búsquedas del usuario, pero este proceso necesita un aprendizaje largo e intensivo.

**Tabla 9.** Cuadro comparativo de los datasets

<b>Referencias</b>	<b>Calificador</b>	<b>Nº</b>	<b>Calificaciones</b>	<b>Fecha</b>	<b>Tipo</b>
MovieLens	Usuarios	270.000	26.000.000	2017	Memoria
Jester	Usuarios	79.681	+ 500.000	2012	Memoria
Lastfm	Reproducciones	2.500	6( artistas)	2016	Modelo
Bookcrossing	Usuarios	5.506	409	2003	Memoria
Netflix	Show	100	60.000.000	2017	Modelo

**Fuente:** La investigadora

Las empresas MovieLens, Bookcrossing utilizan el tipo de memoria a diferencia de las empresas Lastfm, y Netflix utilizan el tipo de modelo, ya que cuentan con enormes cantidades de usuarios y servicios para recomendar a los distintos tipos de usuarios dependiendo a sus gustos.

En base a este análisis, recomendamos a futuros lectores que revisen este documento, aplicar la técnica de memoria en sistemas pequeños y en base que se vaya alimentando el sistema con datos extensos, obligatoriamente aplicar la técnica en modelo.

## **12. IMPACTOS TÉCNICOS Y SOCIALES**

### **12.1. Impacto técnico**

El presente estudio permitirá determinar cómo los algoritmos colaborativos mejoran las búsquedas en aplicaciones informáticas, proporcionando al usuario el interés de softwares inteligentes que faciliten las tareas que realice, dejando un lado a las que no poseen este tipo de algoritmos volviéndolas menos visitadas.

### 12.2. Impacto social

La presente investigación tiene un aporte positivo presentar los algoritmos más importantes empleados en el mercado para de esta manera sugerir que se emplee la tecnología de filtrado de información en la mejora de la experiencia del usuario al momento de realizar búsquedas en internet.

### 12.3. Impacto ambiental

Las personas que utilizan y llegan a depender de aplicaciones informáticas, inconscientemente emplean recursos como la electricidad esto a llevando a que cada vez se construyan más fuentes hidroeléctricas afectando su entorno.

Por otro lado, el uso de la tecnología reduce la publicidad impresa en papel para conocer información de lo que desearía conocer los usuarios.

### 12.4. Impacto económico

Estos algoritmos permitirán aumentar ventas, usuarios y reducir tiempo al buscar contenido, proporcionando a las empresas dar a conocer más sus productos así de esta manera elevar sus ingresos económicos.

## 13. PRESUPUESTO PARA EL PROYECTO DE INVESTIGACIÓN

La tabla 3, detalla la mano de obra, equipos y recursos empleados.

**Tabla 3.** Equipos y costo de mano de obra

<b>Recursos</b>	<b>Cantidad</b>	<b>Unidad</b>	<b>V. Unitario (\$)</b>	<b>Valor Total</b>
<b>Equipos</b>	1	Laptop	0	0
<b>Materiales y suministros</b>	1	Resma de papel boom y material estudiantil.1	90	90
<b>Recursos Básicos</b>	1	Energía eléctrica, internet, agua.	275	275
<b>Alimentación</b>	10	-	3	30
<b>Impresiones</b>	250	-	0.02	5

SUB TOTAL	400
IVA 12%	48
<b>TOTAL</b>	<b>448</b>

**Fuente:** La investigadora

El proyecto de investigación tiene una inversión total de 448.00 USD.

## 14. CONCLUSIONES

- Los sistemas de recomendación colaborativos son los más utilizados actualmente, teniendo en cuenta las calificaciones de todos los usuarios de un sistema se pueden ofrecer recomendaciones más variadas, pero manteniendo la apreciación del usuario.
- Los resultados obtenidos por la comparación revelan que un usuario al obtener recomendaciones de su interés hace que la aplicación sea más útil para realizar la búsqueda.
- La metodología cualitativa proporciona datos reales que detallan el objetivo del presente estudio y mediante el diseño narrativo se presenta los resultados del funcionamiento del algoritmo colaborativo.
- Mediante el caso práctico se pudo conocer el concepto y la estructura que el algoritmo colaborativo cumple al potenciar la aplicación en una búsqueda rápida y precisa de lo que el usuario necesita.

## **15. RECOMENDACIONES**

- Profundizar el estudio acerca de información que permita programar los algoritmos de filtrado colaborativo.
- Tomar como base la aplicación realizada en el presente estudio para implementar en softwares existentes.
- Se debe analizar la funcionalidad del sistema a que ámbito va ser aplicado, ya que el algoritmo colaborativo se basa en una estructura básica de funcionamiento (lógica de negocio) para su desempeño.
- Poner en práctica la presente investigación para desarrollar sistemas que ayuden a una mejor interacción entre el usuario y el sistema.

## 16. BIBLIOGRAFÍA

- A., B.-Y. R., & B, R.-N. (1999). *Modern Information Retrieval*,. Boston: Retrieval.
- Aberger, C. R. (10 de 06 de 2014). *stanford.edu*. Obtenido de <http://cs229.stanford.edu/proj2014/Christopher%20Aberger,%20Recommender.pdf>
- Amazon. (2018). Obtenido de [www.amazon.com](http://www.amazon.com)
- Betarte,Machano,Molina. (12 de 10 de 2016). *fing*. Obtenido de <https://www.fing.edu.uy/inco/grupos/pln/prygrado/InformePGMusica.pdf>
- Boeing, G. (2010). *Geoff Boeing*.
- Castellano. (10 de 2017). *sinbad2.ujaen.e*. Obtenido de [https://sinbad2.ujaen.es/sites/default/files/publications/DEA\\_EmilioCastellano.pdf](https://sinbad2.ujaen.es/sites/default/files/publications/DEA_EmilioCastellano.pdf)
- Crab. (2018). *muricoca.github*. Obtenido de <http://muricoca.github.io/crab/>
- Dataset. (2018). *eigentaste*. Obtenido de <http://eigentaste.berkeley.edu/dataset/>
- Galán. (14 de 11 de 2014). *researchGate*. Obtenido de [https://www.researchgate.net/publication/266878341\\_Filtrado\\_Colaborativo\\_y\\_Sistemas\\_de\\_Recomendacion](https://www.researchgate.net/publication/266878341_Filtrado_Colaborativo_y_Sistemas_de_Recomendacion)
- Google. (2018). *playstore*. Obtenido de [playstoreapp](http://playstoreapp)
- Grané. (2015). *Distancias estadísticas y*. Obtenido de [http://halweb.uc3m.es/esp/Personal/personas/agrane/ficheros\\_docencia/MULTIVARIANT/slides\\_Coorp\\_reducido.pdf](http://halweb.uc3m.es/esp/Personal/personas/agrane/ficheros_docencia/MULTIVARIANT/slides_Coorp_reducido.pdf)
- Hodges. (2013). *n Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges*. Obtenido de [www.international Statistical Review / Revue Internationale de Statistique](http://www.international Statistical Review / Revue Internationale de Statistique)
- Jaccard. (2016). *Bulletin de la Société Vaudoise des Sciences Naturelles*.
- Latest. (2018). *grouplens*. Obtenido de <https://grouplens.org/datasets/movielens/latest/>
- lenskit. (2018). *lenskit*. Obtenido de <http://lenskit.org/>
- Maes, S. S. (10 de 2015). *UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA*. Obtenido de [www.utfsa/Nicolas\\_Torres.pdf](http://www.utfsa/Nicolas_Torres.pdf)
- Martínez. (05 de 12 de 2012). <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com>. Obtenido de <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com/2012/12/modelo-vectorial.html>
- Meteren, R. v. (24 de 05 de 2017). *Using Content-Based Filtering for Recommendation*. Obtenido de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.5743&rep=rep1&type=pdf>

- Moya. (21 de 07 de 2013). *jarroba*. Obtenido de <https://jarroba.com/que-son-los-sistemas-de-recomendacion/>
- mymedialite. (2018). *mymedialite*. Obtenido de <http://www.mymedialite.net/>
- Nepomuceno. (2012). *idus.us.es*. Obtenido de <https://idus.us.es/xmlui/bitstream/handle/11441/27156/1-Memoria%20de%20Tesis%20doctoral.pdf?sequence=1>
- Ng, A. (2015). *es.coursera.org*. Obtenido de <https://es.coursera.org/lecture/machine-learning/k-means-algorithm-93VPG>
- Orseti. (22 de 01 de 2016). *Correlación de pearson y sperman*. Obtenido de <https://es.slideshare.net/miriorsetti/correlacin-de-pearson-y-sperman>
- recommenderlab. (2018). *CRAN.R-project.org*. Obtenido de <https://CRAN.R-project.org/package=recommenderlab>
- Research, G. (2018). *grouplens*. Obtenido de <https://grouplens.org/>
- Rodriguez. (2016). Sistema de recomendación de objetos. *Revista Científica Teknos*, 10.
- Rojas. (03 de 07 de 2014). *Sistema de recomendación por filtrado colaborativo*. Obtenido de <http://www.ijias.issr-journals.org/>
- Salton, G. (1991). *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. New York: Addison Wesley.
- Saluja, C. (2010). *towardsdatascience*. Recuperado el 20 de 07 de 2010, de <https://towardsdatascience.com/my-journey-to-building-book-recommendation-system-5ec959c41847>
- Sarwar, B. M. (10 de 2016). *In Proceedings of CSCW*. Obtenido de Using Filtering Agents to Improve Prediction Quality in the GroupLens Research
- Seguido. (23 de 06 de 2014). *upcommons*. Obtenido de <https://upcommons.upc.edu/bitstream/handle/2099.1/7193/Master%20Thesis%20Seguido.pdf?sequence=1>
- tia. (2018). *tia*. Obtenido de <https://www.tia.com.ec/celular-bmobile-smart-ax675/p?lid=1a4c8798-34d2-4ede-9b44-7101fcb5a17d>
- Torres. (10 de 2015). *UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA*. Obtenido de [www.usm.edu.ec/Nicolas\\_Torres.pdf](http://www.usm.edu.ec/Nicolas_Torres.pdf)
- Triola, M. F. (2012). *Pearson Educación*.
- Willden, C. (2018). *Kaggle*. Obtenido de <https://www.kaggle.com/chasewillden/netflix-shows>

# ANEXOS

**Anexo1.-** Glosario de términos

<b>Termino</b>	<b>Descripción</b>
Bipartito	Que está constituido por dos partes o que se establece entre dos partes.
Tapestry	Es un framework para desarrollar aplicaciones web de código abierto desarrollado sobre el lenguaje Java basado en componentes

## Anexo2.- Siglas

<b>Sigla</b>	<b>Descripción</b>
<b>KNN</b>	Vecinos más cercanos
<b>UBCF</b>	Used Based Collaborative Filtering
<b>IBCF</b>	Item Based Collaborative Filtering
<b>SVD</b>	Singular Value Decomposition, Descomposición en Valores Singulares
<b>CF</b>	Collaborative Filtering, Filtrado Colaborativo.
<b>IR</b>	Information Retrieval, Recuperación de Información.
<b>ITCC</b>	Information Theoretic Co-clustering.
<b>LSI</b>	Latent Semantic Indexing, Indexación Semántica Latente

### Anexo3.- Fotografías.

#### Fotografía 1. Aplicación web para demostrar el algoritmo del filtrado colaborativo

Vivienda Luis Jose ▾

Atras

Recomendaciones Usuarios que les gusta Casa de dos pisos color verde Usuarios que ha han dechado de gustar Casa de dos pisos color verde

### Recomendaciones para Casa de dos pisos color verde

Donde precio venta sea menor a: 35000.00 y donde precio alquiler sea menor a: 5000.00. Se ordena por clasificación más alta.



Suite
Precio de venta <b>\$ 470.00</b>
Precio de alquiler <b>\$ 34.00</b>
Habitaciones <b>4</b>
<div style="display: flex; border: 1px solid #ccc; border-radius: 10px; padding: 2px;"><span style="padding: 0 5px;">0</span> <span style="padding: 0 5px;">1</span> <span style="padding: 0 5px;">2</span> <span style="padding: 0 5px;">3</span> <span style="padding: 0 5px;">4</span> <span style="padding: 0 5px;">5</span></div>
<div style="display: flex; align-items: center;"><span style="color: red; font-size: 24px; margin-right: 10px;">♥</span> <span>3</span></div>
<a href="#">ver recomendaciones</a> <span style="margin-left: 10px;">🗑️</span> <span style="margin-left: 10px;">ℹ️</span>



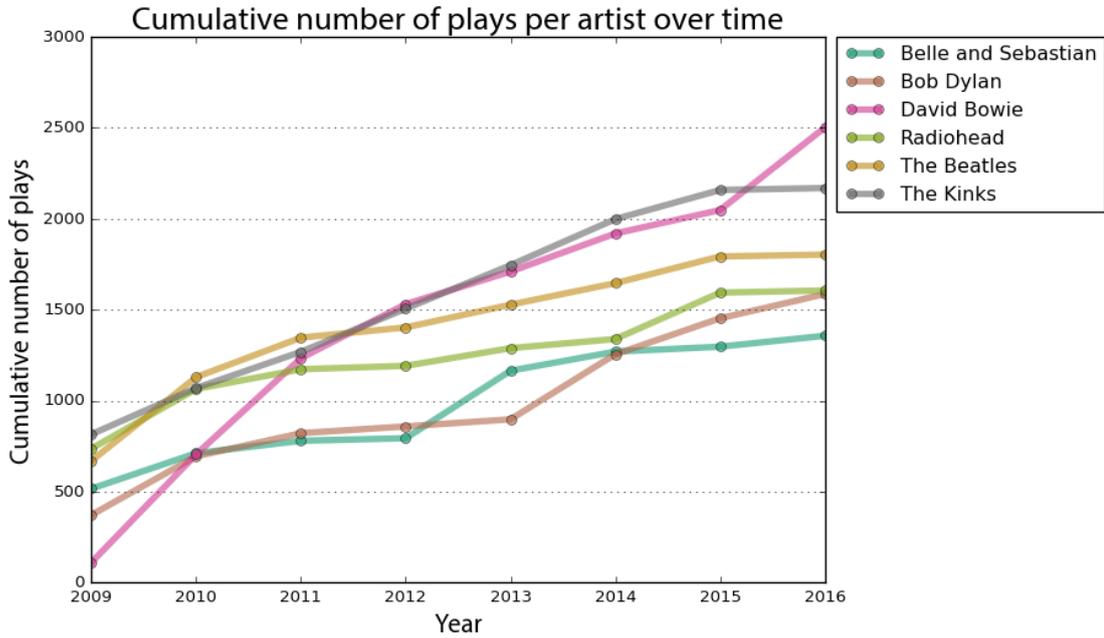
SUITE POR ESTRENAR
Precio de venta <b>\$ 7005.00</b>
Precio de alquiler <b>\$ 500.00</b>
Habitaciones <b>2</b>
<div style="display: flex; border: 1px solid #ccc; border-radius: 10px; padding: 2px;"><span style="padding: 0 5px;">0</span> <span style="padding: 0 5px;">1</span> <span style="padding: 0 5px;">2</span> <span style="padding: 0 5px;">3</span> <span style="padding: 0 5px;">4</span> <span style="padding: 0 5px;">5</span></div>
<div style="display: flex; align-items: center;"><span style="color: red; font-size: 24px; margin-right: 10px;">♥</span> <span>1</span></div>
<a href="#">ver recomendaciones</a> <span style="margin-left: 10px;">🗑️</span> <span style="margin-left: 10px;">ℹ️</span>



Departamento
Precio de venta <b>\$ 7899.00</b>
Precio de alquiler <b>\$ 300.00</b>
Habitaciones <b>2</b>
<div style="display: flex; border: 1px solid #ccc; border-radius: 10px; padding: 2px;"><span style="padding: 0 5px;">0</span> <span style="padding: 0 5px;">1</span> <span style="padding: 0 5px;">2</span> <span style="padding: 0 5px;">3</span> <span style="padding: 0 5px;">4</span> <span style="padding: 0 5px;">5</span></div>
<div style="display: flex; align-items: center;"><span style="color: red; font-size: 24px; margin-right: 10px;">♥</span> <span>3</span></div>
<a href="#">ver recomendaciones</a> <span style="margin-left: 10px;">🗑️</span> <span style="margin-left: 10px;">ℹ️</span>

**Anexo 4.-** Figuras de resultados de la investigación

**Figura 6.** Seis artistas más acumulados de 2009 al 2016



Fuente: (Boeing, 2010)

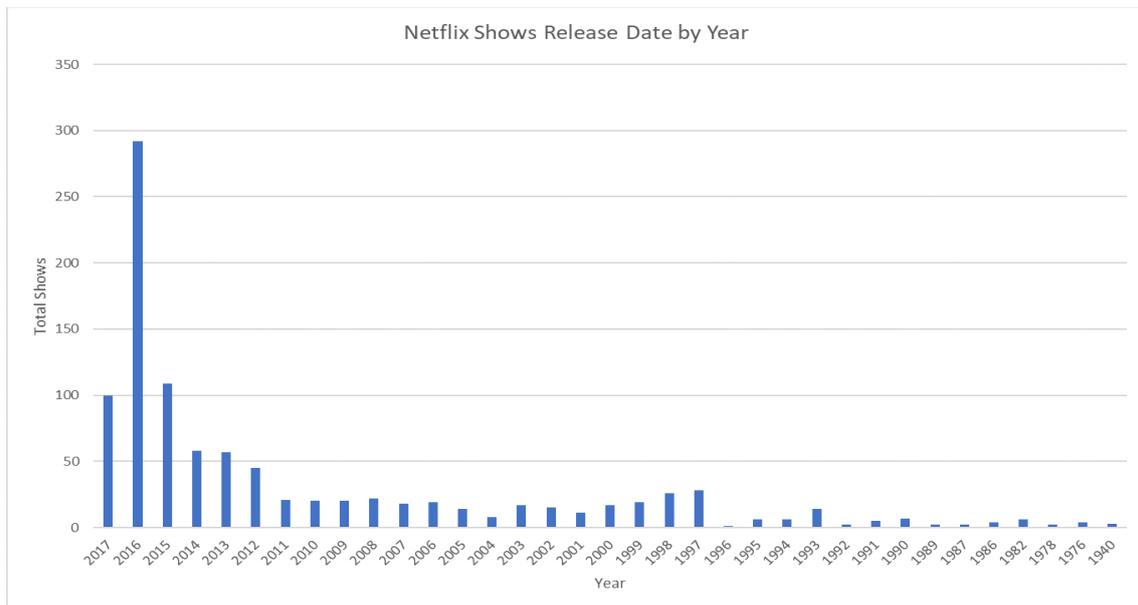
**Figura 7.** Se recomiendan los siguientes libros

Calificación	Isbn	Título del libro	Autor del libro	Año de publicación	Editor
408	0316666343	The Lovely Bones A Novel	Nice behold	2002	Little, Brown
2143	059035342X	Barry Potter and the Sorcerers Stone (Harry Potter (Paperback))	J K Bowling	1999	Arthur A Leone Books
5506	043935806X	Barry Potter and the Order of the Phoenix (Book 5)	1 K Bowling	2003	Scholastic
748	0385504209	The Da Vinci Code	Dan Brown	2003	Doubleday
522	0312195516	The Red Tent (Bestselling Backlmt)	Anita Damant	1998	Picador USA

37	0446310786	To Kill a Mockingbird	Harper Lee	1988	Little Brown & Company
5431	0439139597	Harty Potter and the Goblet of Fire (Book 4)	1 K Rowling	2000	Scholastic
1195	0345370775	Jurassic Park	Michael Cnchton	1999	Ballantne Books
5432	0439064864	Harry Potter and the Chamber of Secrets (Book 2)	1 K Bowling	1999	Scholastic
3839	0439136350	Harry Potter and the Pnsoner of Azkaban (Book 3)	1 K Bowling	1999	Scholastic

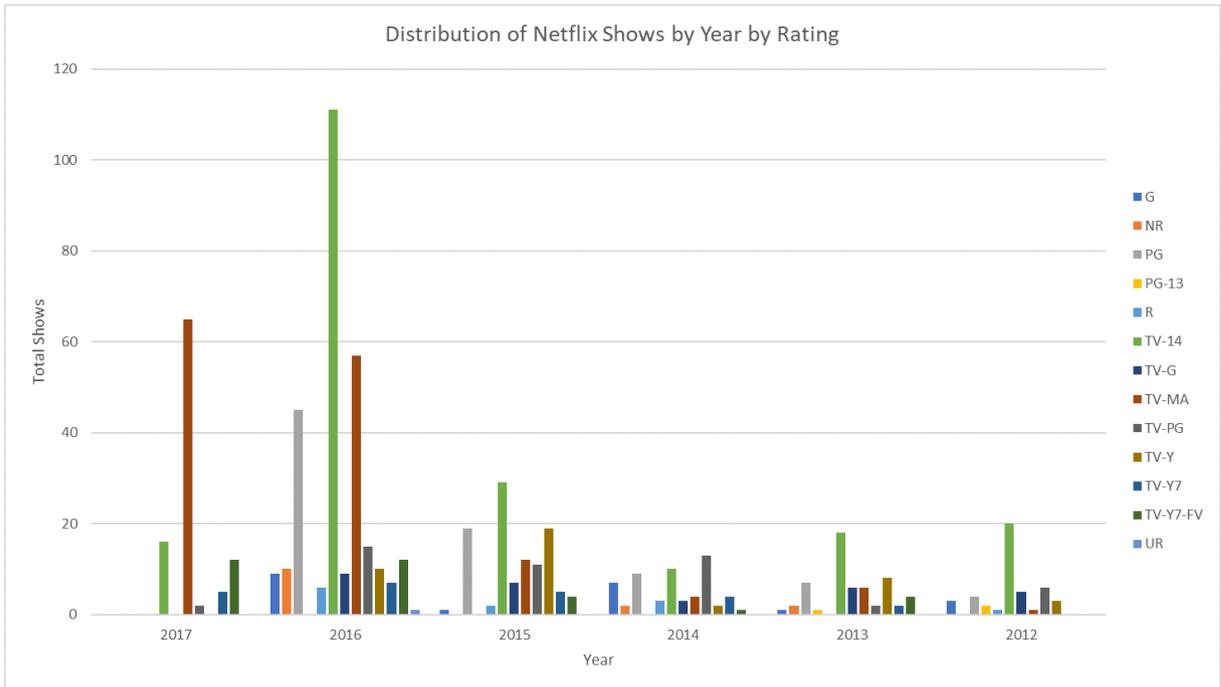
Fuente: (Saluja, 2010)

**Figura 8.** Muestra las fechas de lanzamiento por año



Fuente. (Willden, 2018)

**Figura 11.** Distribución de Netflix muestra por año por calificación



Fuente. (Willden, 2018)

**Figura 12.** Página oficial de plusvalía

The screenshot shows the Plusvalía website interface. At the top, it displays '11.392 Departamentos en venta en Ecuador'. Below this, there are navigation tabs for 'COMPRAR', 'ALQUILAR', 'TEMPORAL/VACACIONAL', and 'PROYECTOS'. A search bar and filter options are visible on the left. The main content area features two real estate listings:

- Edificio lucero'S - Carcelen:** Located at Mariano Pozo N77-295 y Juan de Selis - Ponceano. It features 7 units, 1 to 3 bedrooms, and 73 to 111 m² of covered area. The price starts at \$96,060. It is available for sale immediately.
- Ed Republica Residencial y Comercial - Se...:** Located at Av. República y Atahualpa - Centro Norte, Quito. It features 26 units, 1 to 4 bedrooms, and 77 to 425 m² of covered area. The price starts at \$138,310. It is currently under construction and available for sale immediately.

Fuente: www.plusvalia.com

**Figura 13.** Ingresa al sistema, página principal de ejemplo práctico para recomendar viviendas



**Fuente:** La investigadora

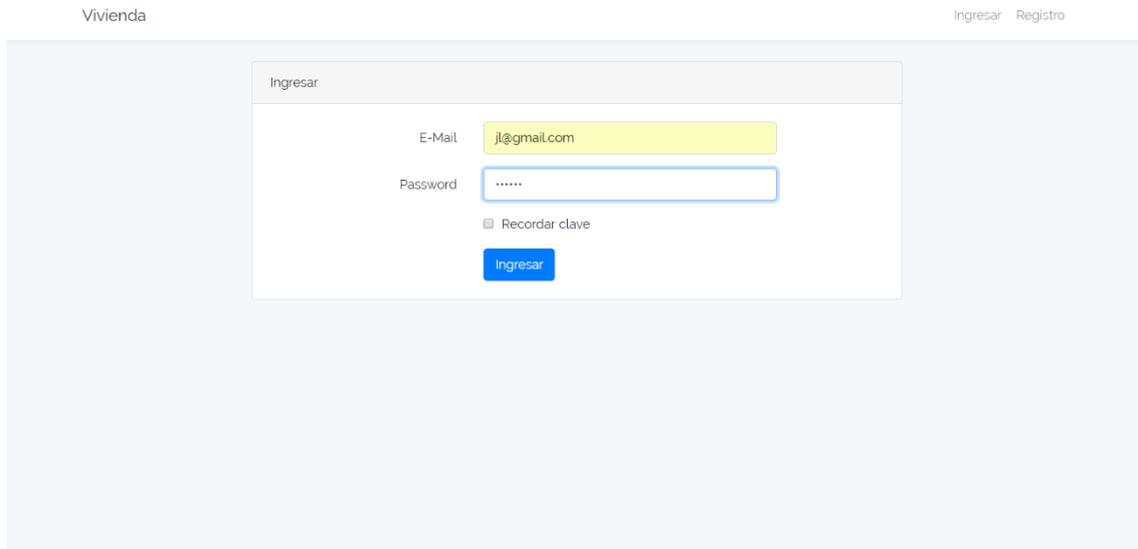
**Figura 14.** Registro de nuevo usuario

The screenshot shows a registration form titled "Registro" in a light gray box. The form is set against a background with "Vivienda" on the left and "Ingresar Registro" on the right. The form contains the following fields and a button:

- Usuario:** Input field containing "Luis Jose".
- E-Mail:** Input field containing "lj@gmail.com".
- Password:** Input field with masked characters ".....".
- Confirme Password:** Input field with masked characters ".....".
- Registrar:** A blue button with white text.

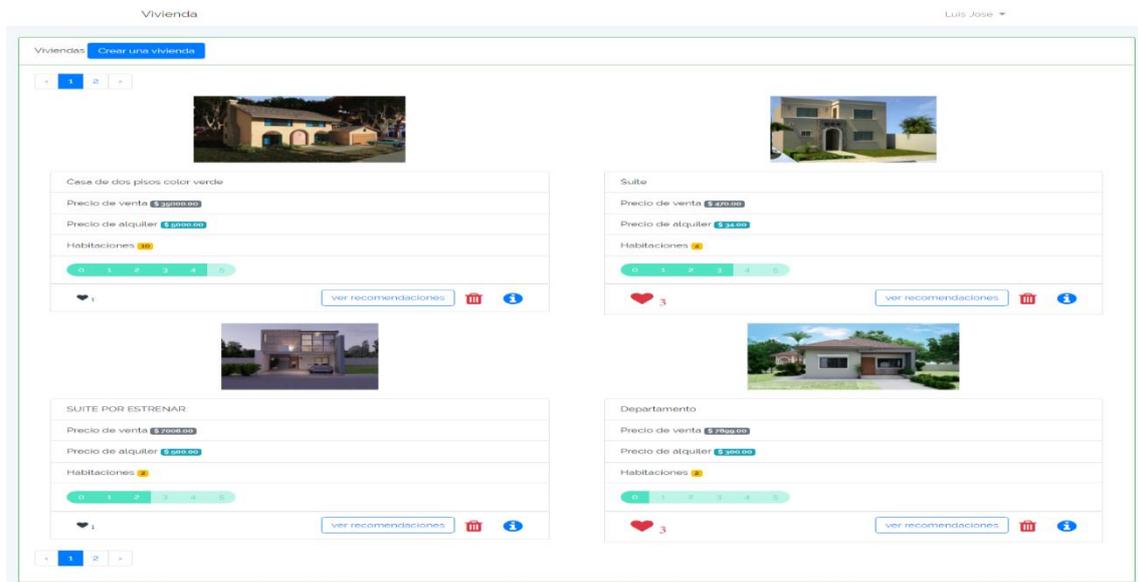
**Fuente:** La investigadora

**Figura 15.** Iniciar sesión



**Fuente:** La investigadora

**Figura 16.** Listado de viviendas, se recomienda viviendas con mayor votación y mayor me gustas ofertada por usuarios que han interactuado con dicha información.



**Fuente:** La investigadora

**Figura 17.** Crear nueva vivienda, página para ingresar nueva vivienda el cual será para ofertar e interactuar con los usuarios de acuerdo a su preferencia de gustos.

Vivienda Luis Jose ▾

Crear nueva Vivienda

Título	<input type="text"/>
	<small>El campo título es obligatorio.</small>
Precio de venta	<input type="text"/>
	<small>El campo precio venta es obligatorio.</small>
Precio de alquiler	<input type="text"/>
	<small>El campo precio alquiler es obligatorio.</small>
Condominio	<input type="text"/>
	<small>El campo condominio es obligatorio.</small>
Superficie total	<input type="text"/>
	<small>El campo superficie total es obligatorio.</small>
Superficie cubierta	<input type="text"/>
	<small>El campo superficie cubierta es obligatorio.</small>
Habitaciones	<input type="text"/>
	<small>El campo habitaciones es obligatorio.</small>
Número de baños	<input type="text"/>
	<small>El campo baños es obligatorio.</small>
Número de medios baños	<input type="text"/>
	<small>El campo medio baños es obligatorio.</small>
Número de garage	<input type="text"/>
	<small>El campo garage es obligatorio.</small>
Número de antigüedad	<input type="text"/>
	<small>El campo antigüedad es obligatorio.</small>
Descripción	<input type="text"/>
	<small>El campo descripción es obligatorio.</small>
Ubicación	<input type="text"/>
	<small>El campo ubicación es obligatorio.</small>
Foto	<input type="button" value="Seleccionar archivo"/> Ningún archivo seleccionado
	<small>El campo foto es obligatorio.</small>

**Fuente:** La investigadora

**Figura 18.** Calificar vivienda, pagina para calificar, dar me gusta, eliminar y ver información de la vivienda con todos los usuarios que les gusten.



**Fuente:** La investigadora

**Figura 19.** Me gusta y no me gusta vivienda, pagina donde se visualiza todos los usuarios que les guste o hayan dejado de gustar cierta información.



**Fuente:** La investigadora

**Figura 20.** Información de vivienda, página para ver toda la información de la vivienda

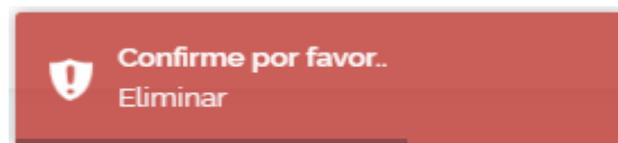
The screenshot shows a modal window titled "INFORMACIÓN" with a close button (x) in the top right corner. Below the title, there are three tabs: "Información" (selected), "Usuarios que le gusta", and "Usuarios que han dejado de gustar". The main content area displays a list of property details, each with a label and a corresponding value in a light gray box:

Título	Casa de dos pisos color verde
Precio de venta	35000.00
Precio de alquiler	5000.00
Condominio	56.00
Superficie total	500 metros cuadrados
Superficie cubierta	560 metros cuadrados
Habitaciones	10
Número de baños	5
Número de medios baños	2
Número de garage	2
Número de antigüedad	1
Descripción	Una linda casa de 2 pisos color rojo
Ubicación	Latacunga El salto, frente la farmacia sana sana

At the bottom right of the modal, there is a "Cerrar" button.

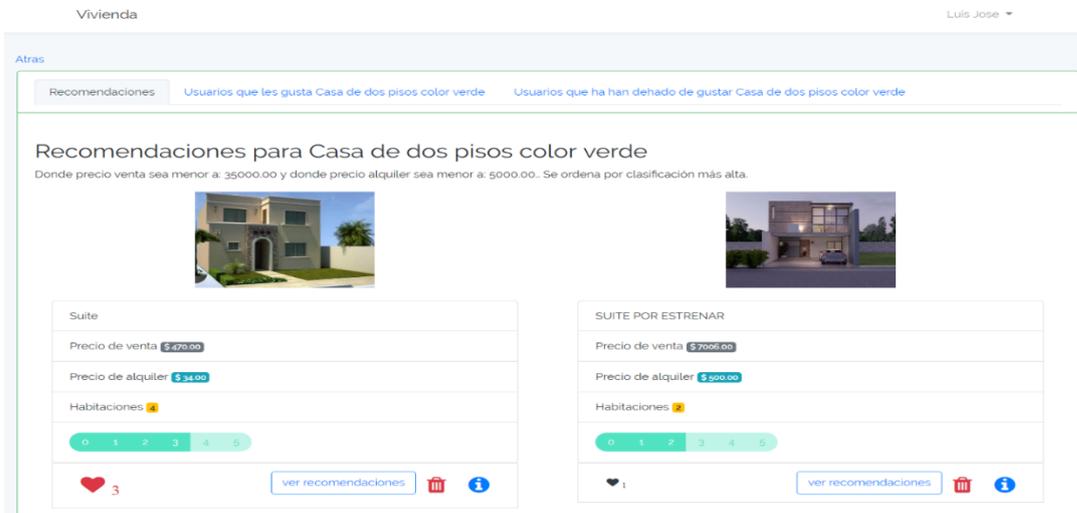
**Fuente:** La investigadora

**Figura 21.** Eliminar información de vivienda



**Fuente:** La investigadora

**Figura 22.** Recomendación de viviendas, pagina donde se recomienda viviendas a los usuarios de acuerdo a su referencia de me gustas, votaciones y precios accesibles.



**Fuente:** La investigadora

**Figura 23.** Función para recomendar viviendas, código fuente donde se realiza el ejemplo práctico para el sistema recomendada de viviendas.

```

File Edit Selection Find View Goto Tools Project Preferences Help
FOLDERS
  ejemplo
  app
    Console
    Exceptions
    Http
    Controllers
      Auth
        Controller.php
        HomeController.php
        Viviendas.php
    Middleware
      EncryptCookies.php
      RedirectIfAuthenticated.php
      TrimStrings.php
      TrustProxies.php
      VerifyCsrfToken.php
    Requests
    Kernel.php
    Providers
      Like.php
      User.php
      Vivienda.php
    bootstrap
    config
    database
    public
    resources
    routes
    storage
    tests
  env
  HomeController.php
class HomeController extends Controller
{
    public function __construct()
    {
        $this->middleware('auth');
    }
    public function index()
    {
        $vivienda=ModelVivienda::orderBy('clasificacion','desc')->paginate(4);
        $data = array('viviendas' => $vivienda );
        return view('home',$data);
    }
    public function nuevo()
    {
        return view('viviendas.nuevo');
    }
    public function crear(Request $request)
    {
        public function actualizarRango(Request $request)
        {
            $svi=ModelVivienda::find($request->input('id'));
            $svi->clasificacion=$request->input('valor');
            $svi->save();
            return json_encode($request->input('valor'));
        }
        public function like(Request $request)
        {
            $like=Like::where('vivienda_id',$request->input('id'))->where('user_id',Auth::id())->first();
            if ($like) {
                if ($like->estado) {
                    $like->estado=false;
                }else{
                    $like->estado=true;
                }
                $like->save();
            }else{
                $like=new Like;
                $like->vivienda_id=$request->input('id');
                $like->user_id=Auth::id();
                $like->estado=true;
                $like->save();
            }
            $like=Like::where('vivienda_id',$request->input('id'))->where('estado',true)->count();
            return response()->json(['estado'=>$like->estado,'i'=>$i]);
        }
        public function eliminar($id)
        {
            $svi=ModelVivienda::find($id);
            try {
                $svi->delete();
                Session::flash('info','Vivienda eliminado exitoso');
            } catch (\Exception $e) {
                Session::flash('warning','Vivienda no eliminada ya que contiene datos de información.'];
            }
            return redirect()->route('home');
        }
        public function recomendaciones($id)
        {
            $svi=ModelVivienda::find($id);
            $sviremin=ModelVivienda::where('precio_venta','<=', $svi->precio_venta)->where('precio_alquiler','<=', $svi->precio_alquiler)->orderBy('clasificacion','desc')->where('id','!=',$svi->id)->paginate(4);
            $data = array('remin' => $sviremin,'info'=>$svi );
            return view('viviendas.algoritmo',$data);
        }
    }
}
Line 91, Column 61
Spaces: 4
PHP

```

**Fuente:** La investigadora