



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 BASADO EN CONTENIDOS
PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”**

Autor:

Pacheco Pazmiño Henry Wladimir

Tutor:

Msc.Ing. Iza Carate Miryan Dorila

Latacunga - Ecuador

Agosto - 2018



Universidad
Técnica de
Cotopaxi



Ingeniería
Informática Y Sistemas
Computacionales

DECLARACIÓN DE AUTORÍA

Yo **PACHECO PAZMIÑO HENRY WLADIMIR** declaro ser autor del presente proyecto de investigación: **“ESTUDIO DE ALGORITMOS DE FILTRADO BASADO EN CONTENIDOS PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”**, siendo la Ing. MSc. Miryan Iza tutora 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.

Pacheco Pazmiño Henry Wladimir

C.I: 0503338394



Universidad
Técnica de
Cotopaxi



Ingeniería
Informática Y Sistemas
Computacionales

AVAL DEL TUTOR DE PROYECTO DE INVESTIGACIÓN

En calidad de tutor de investigación sobre el título: **“ESTUDIO DE ALGORITMOS DE FILTRADO BASADO EN CONTENIDOS PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”**. De **PACHECO PAZMIÑO HENRY WLADIMIR** 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 Aplicada** de la **Universidad Técnica de Cotopaxi** designe, para su correspondiente estudio y calificación.

Latacunga, 31 de Julio de 2018


Msc. Ing. Iza Carate Miryan Dorila

C.I: 0501957617



Universidad
Técnica de
Cotopaxi



Ingeniería
Informática Y Sistemas
Computacionales

APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores. Aprueba 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 postulante **Pacheco Pazmiño Henry Wladimir** con Cd: **0503338394** con el título del proyecto de Investigación: **“ESTUDIO DE ALGORITMOS DE FILTRADO BASADO EN CONTENIDO PARA SISTEMAS RECOMENDADORES DE INFORMACIÓN”**, ha considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido a acta de Sustentación de Proyecto.

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

Latacunga, 31 de julio del 2018

Por constancia firman:

Lector 1 (Presidente)
Mtr. Alex Cevallos
CC: 0502594427

Lector 2
Msc. Ing. Silvia Bravo
CC:0502437122

Lector 3
Phd. Gustavo Rodriguez
CC: 1757001357

AGRADECIMIENTO

Este presente documento de investigación en primer lugar agradezco a Dios por colmarme de bendiciones, salud y éxitos en el transcurso de la etapa de mis estudios.

A la Universidad Técnica de Cotopaxi por haberme brindado la oportunidad de estudiar y ejercer mi profesión competitiva.

A mi tutora de investigación Msc Miryan Iza, por su esfuerzo y dedicación quien con sus conocimientos, experiencia y paciencia ha logrado en mí motivar para culminar mis estudios con éxito.

A mis docentes por guiarme en la formación de mis actitudes y aptitudes, por haberme compartido sus conocimientos, sus experiencias.

Henry

DEDICATORIA

El presente trabajo de investigación está dedicado a mi madre Carmita Pacheco, a mi esposa Johana Pacheco y a mi hijo Josué Pacheco que siempre están apoyándome, motivándome a seguir adelante.

A toda mi familia por su apoyo, consejos, comprensión, amor, ayudada en los momentos difíciles necesarios para estudiar. Me han dado todo lo que soy como persona, mis valores, mis principios, mi coraje, mi carácter, mi empeño, mi perseverancia para conseguir mis objetivos.

Henry

ÍNDICE DE CONTENIDO

PROYECTO DE INVESTIGACIÓN.....	i
DECLARACIÓN DE AUDITORÍA.....	ii
AVAL DEL TUTOR DE PROYECTO DE INVESTIGACIÓN	iii
APROBACIÓN DEL TRIBUNAL DE TITULACIÓN.....	iv
AGRADECIMIENTO	v
DEDICATORIA.....	vi
ÍNDICE DE CONTENIDO	vii
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS	xi
RESUMEN.....	vii
AVAL DE TRADUCCIÓN.....	xv
1. INFORMACIÓN GENERAL:	1
2. RESUMEN DEL PROYECTO	3
3. JUSTIFICACIÓN DEL PROYECTO	3
4. BENEFICIARIOS DEL PROYECTO	4
5. EL PROBLEMA DE INVESTIGACIÓN	4
6. OBJETIVOS.....	5
6.1. Objetivo general	5
6.2. Objetivos específicos.....	5
7. ACTIVIDADES Y SISTEMA DE TAREAS EN RELACIÓN A LOS OBJETIVOS	5
8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA	7
8.1. Antecedentes.....	7
8.2. Sistemas de recomendación.....	8
8.2.2. Aplicación de los sistemas de recomendación	10
8.3. Tipos de los algoritmos empleados en la presente investigación	18

8.3.1 Basados en contenido (BC)	18
8.3.2. Basado en reglas	23
8.3.3.1. Dificultades que podemos encontrar cuando usamos el filtrado colaborativo	27
8.3.3.2. Filtrado colaborativo basado en el usuario	27
8.3.4. Basado en enfoques mixtos	28
8.4. Data sets que utilizan filtrado basado en contenido	28
8.4.1. Amazon.....	28
8.4.1. My BestBet.....	28
8.4.1. Jinni	29
9. HIPÓTESIS.....	29
10. METODOLOGÍA DE LA INVESTIGACIÓN	30
10.1. Investigación descriptiva.....	30
10.2. Diseño Descriptiva	30
11. ANÁLISIS DE RESULTADOS	30
11.1. Análisis descriptivo de sistemas basados en contenido	30
11.2. Análisis descriptivo de sistemas basados en reglas.....	31
11.5. Análisis de los cuatro algoritmos.....	34
11.5.1. Comparación de Data sets que utilizan filtrado basado en contenido	344
11.6. Data sets (Conjunto de datos).....	355
11.6.1. Jinni	35
11.6.2. My BestBet.....	36
11.6.3. Amazon.....	37
12. IMPACTOS TÉCNICOS Y SOCIALES	38
12.1. Impacto técnico	38
12.2. Impacto social	38
12.3. Impacto ambiental.....	39
12.4. Impacto económico	39

13. PRESUPUESTO PARA EL PROYECTO DE INVESTIGACIÓN	39
14. CONCLUSIONES Y RECOMENDACIONES	40
14.1. Conclusiones	40
14.2. Recomendaciones.....	40
15. BIBLIOGRAFÍA	41

ÍNDICE DE TABLAS

Tabla 1. Sistema de tareas en relación a los objetivos planteados.	6
Tabla 2. Diferencias entre los principales tipos de sistemas de recomendación	9
Tabla 3. Tabla de software que emplea filtrado basado en contenido.....	34
Tabla 4. Datos de Jinni	35
Tabla 5. Tabla de valoraciones de la herramienta MyBestBets	36

ÍNDICE DE FIGURAS

Figura 1. Arquitectura de un sistema de recomendación basado en contenido.....	21
Figura 2. Reglas empleadas en un cajero automático.....	24
Figura 3. Arquitecturas de sistemas basados en reglas.....	25

UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS.

TÍTULO: “Estudio de algoritmos de filtrado basado en contenidos para sistemas recomendadores de información”

Autor: Pacheco Pazmiño Henry Wladimir

RESUMEN

Los algoritmos de filtrado basados en contenidos tienen el fin de solventar la dificultad de la masiva información que se presenta del modo más frecuente en la personalización de contenido en cimiento a posibles registros de un usuario que recepta posteriormente de una búsqueda, ya que obtiene una inmensidad de posibilidades que no requiere o no está de acuerdo con lo que encuentra, perdiendo interés por seguir utilizando la aplicación. Los sistemas de recomendación basados en algoritmos de contenidos son aplicaciones inteligentes que utilizan las valoraciones de los usuarios sobre ciertos elementos del conjunto total para predecir valoraciones en el resto de los elementos y recomendar cuales son los de mayor valoración predicha, se basa en un algoritmo inteligente al ser capaz de receptar intereses de usuarios para su posterior implementación generando una ayuda para realizar tareas eficazmente. Mediante la vigente investigación de los algoritmos de filtrado de información basados en contenido se demuestra el gran provecho que presentan a las personas que hacen práctica de un software para disponer sus peticiones de manera adecuada y de pronta consulta. Este documento permite adquirir una representación de los sistemas de filtrado de información en el cual se cubre los múltiples aspectos que comprenden los diversos esquemas basados en modelo y en datos aplicados por filtros de información, el cual presenta la medio para obtener respuestas de información útil para el usuario. Para la presentación de resultados se comparó la lógica de cuatro algoritmos de filtrado de información incluyendo el basado en contenido, con los cuales se demuestra el funcionamiento que tienen mediante el desarrollo y la visualización de la aplicación en la web. Los impactos mencionados en el presente proyecto muestran el gran interés de continuar la investigación para la futura implantación en softwares de información que aportaría de parte y parte es decir, al usuario con el útil encuentro de lo que necesita y con las empresas que lo desarrollan aumentando sus ingresos económicos. Se comprueba que es posible establecer algoritmos que ayuden al fácil manejo de contenido específicos en aplicaciones informáticas siendo estas de mucha utilidad para la interacción del

usuario con el sistema, se recomienda la implementación de estos algoritmos en sistemas ya existentes dentro de la universidad con el cual se potenciará el aprendizaje, la experiencia y la satisfacción de los usuarios.

Palabras claves: Filtrado de información, Sistemas Inteligentes, Algoritmos, Aplicaciones Informáticas.

UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS.

TOPIC: “Study of content-based filtering algorithms for information-recommending systems”

Author: Pacheco Pazmiño Henry Wladimir

ABSTRACT

The filtering algorithms have a content that contains the most advanced information that is presented in the most frequent way in the personalization of content in a register of records of a user who received a subsequent search, since it obtains an immensity that does not require or does not he agree with found finds, losing interest in continuing to use the application. Classification systems are based on content algorithms that are used to detect the most important elements in the world, to be able to receive the interests of users for its subsequent implementation generating a help to perform tasks effectively. By means of the last investigation of the information filtering algorithms in content, it demonstrates the great benefit that people make software practices in order to obtain their suggestions in an appropriate way and with prompt consultation. This document allows to obtain a representation of the information filtering systems in which the multiple aspects are used in information filter, which presents the means to obtain answers of useful information the user. For the results, the logic of four information filtering algorithms is compared with the content, which demonstrates how they work through the development and visualization of the application on the web. The impacts are in the present project show the great interest to continue the investigation for the future implantation in information software that contribute part, to the user with the use of what he needs and with the companies his economic income. Is possible to establish algorithms that help the easy handling for specific content in computer applications, being very useful for the user's interaction with the system, it is recommended the implementation of these algorithms in systems within the university with the which will enhance learning, experience and user satisfaction.

Key words: Information Filtering, Intelligent Systems, Algorithms, Computer Applications



Universidad
Técnica de
Cotopaxi

CENTRO DE IDIOMAS

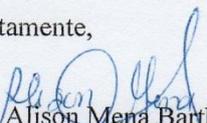
AVAL DE TRADUCCIÓN

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 señor Egresado de la Carrera de Ingeniería en Informática y Sistemas Computacionales: PACHECO PAZMIÑO HENRY WLADIMIR, cuyo título versa “**ESTUDIO DE ALGORITMOS DE FILTRADO BASADO EN CONTENIDO 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



CENTRO
DE IDIOMAS

1. INFORMACIÓN GENERAL:

Título del proyecto:

Estudio de algoritmos de filtrado basados en contenidos 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: Miryan Dorina

Apellidos: Iza Carate

Fecha de nacimiento: 09 de Septiembre de 1973

Teléfonos: 0998364480

E-mail: miryan.iza@utc.edu.ec

Estudios: - Universidad Técnica de Cotopaxi

- Escuela Superior Politécnica del Ejército

Títulos obtenidos: -Ingeniería en Sistemas Computacionales e Informática
- Msc. Ingeniería en Software

Datos Personales del Autor:

Nombres: Henry Wladimir

Apellidos: Pacheco Pazmiño

Fecha de nacimiento: 11 de Mayo de 1989

C.C: 0503338394

Teléfono: 0994154139

Correo electrónico: henrypacheco1989@gmail.com

Estudios: - Universidad Técnica de Cotopaxi
- Universidad de las Fuerzas Armadas

Área de conocimiento:

Red de Estudios Cienciométricos

Línea de investigación:

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

Sublínea de investigación de la Carrera:

Robótica e Inteligencia Artificial

2. RESUMEN DEL PROYECTO

El manejo de la gran cantidad de información de internet que existe en este mundo, es cada vez más complicado, sobre todo a la hora de buscar información. Es aquí donde han ido consolidándose herramientas que ayudan a disminuir la sobrecarga de información; entre estas herramientas existen los SR (Sistemas Recomendadores), los cuales filtran los ítems de la información a partir de técnicas que identifican los principales atributos que más se acomodan a las preferencias de los usuarios, recomendaciones que se generan a partir de gustos de otros usuarios o del contenido de la información (Cadena y Carantón, 2017)

El estudio realizado en este tema de investigación, contiene información recolectada y analizada de referencias bibliográficas acerca de sistemas recomendadores basado en contenido, el cual se evidencio, mediante la comparación de cuatro algoritmos de filtrado con los cuales se describen las características empleadas en cada uno, para determinar donde se puede aplicar un algoritmo recomendador basado en contenidos en ayuda a sistemas recomendadores a ser más útiles para los usuarios en la filtración de datos. Con lo cual se demuestra que un algoritmo recomendador basado en contenido se enfoca en la interacción y preferencia de datos o elementos similares aquellos por los que el usuario ha mostrado interés.

Palabras claves

- Filtrado de información, Sistemas Basados en Contenido, Algoritmos, Sistemas recomendadores.

3. JUSTIFICACIÓN DEL PROYECTO

Sin dudas, en la próxima generación de sistemas de información inteligentes, los agentes de información desempeñarán un papel fundamental en la búsqueda activa y la búsqueda de información relevante en nombre de sus usuarios en entornos complejos y abiertos, como Internet. Considerando que puede definirse pertinente únicamente para un usuario específico, y en el contexto de un dominio o tema concreto. Debido a esto el desarrollo de sistemas de clasificación de documentos inteligentes, personalizados, basados en contenidos, se está convirtiendo cada vez más atractivos hoy en día.

Por otra parte, los perfiles de aprendizaje que representan los intereses del usuario dentro de un dominio en particular, más tarde utilizados para el filtrado basado en contenido, se ha demostrado que es una tarea desafiante. Especialmente porque, dependiendo del usuario, los atributos relevantes para cada clase cambian en el tiempo. Esto hace que el problema no sea adecuado para los algoritmos de aprendizaje de máquinas tradicionales y de atributos fijos.

Un sistema de filtrado basado en el contenidos selecciona elementos basados en la correlación entre contenido de los artículos y las preferencias del usuario en comparación con un filtro colaborativo sistema que elige elementos basados en la correlación entre personas con preferencias. Por ello es una herramienta que potencia al software al ser capaz de interpretar la necesidad de un determinado usuario, recolectando sus interacciones para conocer lo que más le interesa, comparando sus gustos o características personales para recomendarle artículos, objetos o personas relacionado a su personalidad.

4. BENEFICIARIOS DEL PROYECTO

Beneficiarios directos:

Como beneficiario directo es la revista de Investigación de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

Beneficiarios indirectos:

Estudiantes, profesores.

5. EL PROBLEMA DE INVESTIGACIÓN

A medida que la World Wide Web sigue creciendo a un ritmo exponencial, el tamaño y la complejidad de muchos sitios web crecen junto con él. Para los usuarios de estos sitios Web se vuelve cada vez más difícil y lento para encontrar la información que están buscando. (Meteren, 2017)

Es por ello que el problema principal de la mayoría de sistemas de recomendación es que presentan una serie de opciones que no ayudan al usuario a tener contenido relacionado con su búsqueda y cada vez tendría que volver a realizar la misma consulta, redundando la misma información ya encontrada.

Para nuestro caso de estudio se emplea el proyecto de red de estudio cuantitativos para la Universidad Técnica de Cotopaxi donde existen un sistema encargado de recopilar la información de los docentes investigadores relacionados a su producción científica, ósea publicación de artículos científicos, libros y ponencias, esto acarrea una importante y significativa cantidad de información en relación a los autores de estos trabajos, se necesita un algoritmo que permita sugerir área de conocimiento, temáticas, líneas y sublíneas de investigación según la relación o similaridad en cuanto a estos indicadores cuantitativos con respecto a autores y coautores.

¿Cómo establecer niveles de recomendación de información a partir de la producción científica de los investigadores de la Universidad Técnica de Cotopaxi, orientados a una mayor eficiencia en establecimiento de comunidades colaborativa de conocimientos?

6. OBJETIVOS

6.1. Objetivo general

- Determinar la validez de los algoritmos de filtrado de información basado en contenidos mediante la evaluación de su arquitectura para aportar con información relevante a las necesidades de los usuarios.

6.2. Objetivos específicos

- Realizar una revisión bibliográfica para determinar el estado del arte.
- Comparar algoritmos de filtrado de información basado en contenido en los Data set (Conjunto de datos) para conocer cuál de ellos aporta más a la búsqueda personalizada del usuario.
- Analizar la fortaleza del algoritmo de recomendación basado en contenido que presenta en sistemas recomendadores de información.

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.

Objetivo	Actividad (tareas)	Resultado de la actividad	Medios de Verificación
Realizar una revisión bibliográfica para determinar el estado del arte	<ul style="list-style-type: none"> • Recopilación de información bibliográfica de: • Libros, artículos científicos, tesis, páginas web. • Lectura crítica de la información recogida. • Documentación de la fundamentación teórica del proyecto de investigación 	<ul style="list-style-type: none"> • Determinar que algoritmo es el más apropiado para realizar búsquedas personalizadas. 	<ul style="list-style-type: none"> • Investigación Documental • Análisis comparativo
Comparar algoritmos de filtrado de información basado en contenidos con otros algoritmos recomendadores para conocer cuál de ellos aporta más a la búsqueda personalizada del usuario.	<ul style="list-style-type: none"> • Comparar características entre los algoritmos de filtrado seleccionado • Establecer un cuadro comparativo que presenta los detalles de funcionamiento de los algoritmos 	<ul style="list-style-type: none"> • Fundamentación teórica del proyecto de investigación. 	<ul style="list-style-type: none"> • Libros • Investigación Documental • Artículos
Analizar la fortaleza del algoritmo de recomendación basado en contenidos que presenta en sistemas recomendadores de información.	<ul style="list-style-type: none"> • Investigar aplicaciones populares que emplean algoritmos de filtrado de información. • Determinar la importancia de los algoritmos de recomendación basado en contenido 	<ul style="list-style-type: none"> • Obtener ventajas principales 	<ul style="list-style-type: none"> • Investigación Documental

Fuente: El investigador

8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA

8.1. Antecedentes

A mediados de los años 90 surgen los Sistemas de Recomendación (SR) con el objetivo de entregar a los usuarios resultados adaptados y cercanos a sus necesidades. Los SR están demostrando ser una herramienta útil para hacer frente a una parte del fenómeno de la sobrecarga de información a través de Internet. Su evolución ha acompañado a la evolución de la web. (Rodríguez, 2016)

Chesani define un SR como aquel capaz de realizar predicciones a partir del hecho que a un usuario le guste o no cierto ítem al que podría acceder, para esto es necesario almacenar la información de los usuarios por medio de perfiles de usuario. En la actualidad se han desarrollado muchas técnicas de SR, cada una puede ser usada para diferentes enfoques y en diferentes contextos, de acuerdo a las necesidades y a la información disponible, debido que estas técnicas necesitan gran cantidad de información sobre los usuarios y recursos a recomendar para entregar resultados de calidad, esto se logra con la caracterización de los estudiantes y la retroalimentación que se realice de las recomendaciones. (Chesani, 2014)

El crecimiento de Internet ha sido increíblemente acelerado, hay cantidades ingentes de datos casi sobre cualquier tópico, lo que hace de Internet la fuente más completa de información para casi todo. Sin embargo, hay más cantidad de artículos, libros, películas, programas de televisión, entre otras cosas, de los que las personas podemos manejar efectivamente, con lo cual hay una sobrecarga de información hacia los usuarios finales. El tiempo que se necesita para hacer una búsqueda sobre el conjunto total de tópicos es demasiado, y los procesos de búsqueda tradicionales pueden ser improductivos la mayoría de las veces o con resultados poco exitosos. (Velez, 2016)

Los sistemas de filtrado de información están dirigidos, en principio, a personas que carecen de la experiencia, la capacidad o el tiempo necesarios para evaluar la inmensa cantidad de tópicos que generalmente están a su disposición en un sitio Web y se basan en la premisa de que, habitualmente, la gente tiene en cuenta las recomendaciones provistas por otros para tomar decisiones rutinarias. (IMGBIBLIO, 2014)

En los últimos años ha aumentado el uso de elementos agradables en entornos no recreativos para aumentar la participación y la fidelidad del público. Esta estrategia, denominada

ludificación, y conocida también como gamificación, consiste en integrar técnicas, dinámicas y otros elementos propios de los juegos y del ocio en actividades no lúdicas de la vida cotidiana. El objetivo es potenciar la motivación, así como reforzar la conducta para solucionar problemas, mejorar la productividad, alcanzar un objetivo y activar el aprendizaje (Posada-Prieto, 2015)

La evolución de la Web ha propiciado el desarrollo de fuentes de información de variada índole, a menudo sustentadas por una amplia comunidad que las alimenta y las supervisa de forma altruista y colaborativa, como es el caso de Wikipedia. En el ámbito de la recomendación social en particular, existen numerosas plataformas que, por medio de un conjunto de estrategias conocidas como funware, incentivan la conducta de los usuarios para aumentar su implicación y explotar su conocimiento colectivo. Las dinámicas motivadoras a este efecto incluyen premios (generalmente intangibles, como puntos o insignias) que activan un círculo virtuoso de estímulo y recompensa (Werbach, 2015)

8.2. Sistemas de recomendación

Los sistemas de recomendación ayudan al usuario a seleccionar elementos de una gran cantidad de opciones. Estos sistemas han tenido gran aceptación debido a que los usuarios están acostumbrados a recibir recomendaciones de amigos y colegas. Se deben tener en cuenta diferentes aspectos para diseñar sistemas de recomendación. (Balavanovic y Shoham , 2016)

Una de las variables importantes es el volumen de la información, ya que de éste depende el detalle de las recomendaciones. Factores como el tiempo de vida, el tipo de elemento (películas, gente, artículos, etcétera) y la cantidad generada influyen de manera directa en el momento de la recomendación. (Balavanovic y Shoham , 2016)

Existen también implicaciones sociales. Establecer un perfil de intereses de las recomendaciones puede ocasionar problemas en casos de imparcialidad. Además, la privacidad de los participantes debe tomarse en cuenta ya que no todos los usuarios gustan de exhibir sus preferencias o de no ser reconocidos por su aportación. (Balavanovic y Shoham , 2016)

Debido a que mantener un sistema de recomendación es caro, se han considerado diferentes modelos para costear dichos sistemas:

- El consumidor paga por el servicio
- Los anuncios de publicidad mantienen el sistema
- El dueño del elemento a evaluar paga por la evaluación de su elemento.

En los sistemas de recomendación existen dos paradigmas para la selección de elementos, basados en contenido y filtrado colaborativo. En los sistemas basados en contenido el usuario recibirá información similar a la que ha mostrado interés en el pasado, mientras en el filtrado colaborativo las sugerencias serán de elementos que han gustado a gente con intereses similares a los suyos. (Balavanovic y Shoham , 2016)

A continuación se presenta la tabla en donde se realiza la comparación de varios algoritmos de filtrado de información:

Tabla 2. Diferencias entre los principales tipos de sistemas de recomendación

Tipo de recomendadores	Objetivos del usuario	Input
Filtrado colaborativo	Proporciona recomendaciones calculadas a partir de las valoraciones e interacciones de todos los usuarios (incluido el usuario objetivo) con los productos del sistema	Valoraciones del usuario + Valoraciones del resto de usuarios
Basado en contenido	Proporciona recomendaciones basándose en las descripciones o atributos de los productos del sistema, favoreciendo los gustos del usuario objetivo	Valoración del usuario + atributos de los productos
Basado en conocimiento	Proporciona recomendaciones en base a las restricciones del usuario sobre el conjunto de los problemas del sistema	Restricciones del usuario + atributos de los productos + conocimiento del dominio

Fuente: (Guillermo, 2017)

8.2.1 Función de los sistemas de recomendación

El objetivo de un sistema de recomendación es servir dinámicamente contenido personalizado (libros, películas, páginas, productos, anuncios) a los usuarios, en base a su perfil, preferencias o intereses. (Guillermo, 2017)

Los sistemas de recomendación son una de las formas más comunes de despersonalización. Aunque su origen se remonta a varias décadas atrás, es ahora, con la democratización de las técnicas de Machine Learning que aportan las soluciones Cloud y el auge del People-Based marketing, cuando muchas empresas están apostando decididamente por esta disciplina.

Es el caso de negocios nacidos online, como Amazon, Netflix, Spotify, LinkedIn o Facebook, y por supuesto, e-retailers que quieren potenciar el cross-selling y dar vida a su catálogo de productos. (Guillermo, 2017)

El objetivo final es ofrecer al usuario lo que está buscando, incluso antes de que él mismo lo sepa, ahorrándole tiempo y ofreciéndole una experiencia única.

8.2.2. Aplicación de los sistemas de recomendación

Existe una gran diversidad de tipos de productos a recomendar. El uso mayoritario que se hace de los sistemas de recomendación es en el campo del comercio electrónico, en la compra-venta de productos, libros, películas, viajes y otro tipo de servicios. Aunque esta es la utilidad típica de los recomendadores, lo cierto es que su uso se ha ampliado a la recomendación de otro tipo de productos y en otros contextos. (Guillermo, 2017)

Por ejemplo en la recomendación de personas (en una red social como Facebook o LinkedIn), a la cual se le llama recomendación recíproca, o en el área de la publicidad computacional, que proporciona recomendaciones de productos a usuarios al realizar búsquedas, como en las búsquedas de Google. También existen otras plataformas como YouTube, donde se recomiendan vídeos, IMDb, donde los ítems son películas o TripAdvisor, que recomienda hoteles, viajes, etc. (Guillermo, 2017)

8.2.2.1. El Marketing y los Sistemas de Recomendación

En general, durante todo el RecSys'09 (Recommender Systems) se mostró un alto interés en la aplicación de los Sistemas de Recomendación al marketing. Michel Wedel es un experto en marketing de la Escuela de Negocios Robert H. Smith, de la Universidad de Maryland, que

impartió una sesión especial con el título “Up-Close and Personalized: A Marketing View of the Future of Recommendation Algorithms”, en la que este interés quedó totalmente claro. (Monsalve, 2015)

Según Michel, en primer lugar, diversas investigaciones en el campo del Marketing sobre el comportamiento de los usuarios han demostrado que: (Monsalve, 2015)

- Los sistemas de recomendación (SR) ayudan a reducir los precios pagados, mejorar las decisiones e influir en las opiniones de los usuarios.
- Los SR generalmente mejoran la calidad y aceptación las recomendaciones.
- Las decisiones finales tomadas por los usuarios dependen del contexto y el formato de visualización.

Estos puntos coinciden además con lo comentado por Francisco Martín en su charla “Top 10 Lessons Learned Developing, Deploying, and Operating Real-World Recommender Systems“. (Monsalve, 2015)

Otro punto importante de la charla de Michel daba pistas sobre los criterios que debe tener una empresa que quiera utilizar Sistemas de Recomendación a la hora de implantar este tipo de servicios: (Monsalve, 2015)

- En primer lugar, recomendaba no crear SR propios y confiar en proveedores externos.
- Los SR son especialmente recomendables para empresas que tienen cantidades masivas de datos, y cualquier red social media actual los tiene.
- No hay que confiar en sistemas que utilicen únicamente el filtrado colaborativo, y buscar proveedores que ofrezcan sistemas híbridos que utilicen también etiquetas y contenidos.
- Pasar de los SR estáticos que actualizan las recomendaciones cada cierto tiempo a los sistemas dinámicos que actualizan los intereses de los usuarios con cada acción que éstos realizan.
- Aprovechar la información que ofrecen las redes sociales sobre las amistades y similitudes entre usuarios para mejorar los resultados de las recomendaciones. En este sentido los denominados sistemas basados en confianza tuvieron también mucha repercusión durante todo el Congreso.

- Utilizar no sólo las votaciones de los usuarios, sino también la información implícita que se proporciona a través de los clicks en enlaces y las compras de productos.
- Mostrar al usuario fotos de los productos recomendados, ya que algunos estudios demuestran que los resultados de las recomendaciones son mejores en estos casos.

El sistema generalizado de preferencias ya había integrado características en nuestros sistemas de recomendación, por lo que se ve que están haciendo las cosas según las tendencias del mercado. Aun así, tenemos claro que aún queda un largo trabajo en el campo de los SR, y vamos a seguir poniendo gran parte de nuestro esfuerzo en mejorar nuestros sistemas. (Monsalve, 2015)

8.2.2.2. Sistema de recomendación con conocimiento web

Una parte importante de la web que hoy manejamos cuenta con la presencia de sistemas de recomendación y estos se van apoderando cada vez más de las aplicaciones más comunes y de mayor consumo por los usuarios, entre estas aplicaciones se sitúan los buscadores, aplicaciones de compra y en cualquier sitio donde se maneje la toma de decisión de un usuario o se muestran opciones de acuerdo a las necesidades que presenten en ciertos campos los clientes, con menor o mayor refinamiento tendremos un sistema de recomendaciones funcionando muchas veces de manera inadvertida para mucha gente.

El crecimiento es notorio y la complejidad de estos sistemas ha crecido, son mucho más potentes, pueden por ejemplo casi adivinar que podrías comer un miércoles al mediodía, o si la película de estreno que está en las salas de cine es de tu preferencia y hasta planificar tus vacaciones, es justamente en este último caso donde enfocamos nuestro proyecto, aprovechando el mercado de turismo en España que es ricamente productivo el cual ha generado mucho beneficio a lo largo del tiempo, en conocimiento que España ha sido, es y será un destino turístico por excelencia, tener al alcance herramientas que faciliten al usuario la toma de decisiones en cuanto a que lugar hospedarse siendo el mismo el que más se ajuste a sus gustos y necesidades es primordial, en la web hay varios sistemas que ya realizan este tipo de servicios como Tripadvisor o Trivago entre otros, nosotros hemos creado el nuestro, dedicado al sector hotelero para las principales y más visitadas ciudades de España. (Yáñez, 2015)

Hemos desarrollado bajo la perspectiva mencionada antes, un Sistema Recomendador de Hoteles con conocimiento Web, bajo dos conceptos que hemos llamado por sus características, a uno analizador estadístico y a otro analizador sentimental, la idea general es que un usuario pueda evaluar un hotel de su interés de varias formas en nuestra aplicación, de las que entraremos en detalle más adelante en esta memoria, saber la reputación de dicho hotel bajo el análisis estadístico y también con un análisis sentimental que obedece a manejo de la información web que se encuentre relacionada al hotel en cuestión, con el fin de dar una evaluación en cada caso, mostrando una puntuación, dicho resultado llevara al usuario a tomar la decisión más adecuada que se perfila a sus necesidades, siendo esta última parte el objetivo principal que tenemos con nuestro proyecto, que es satisfacer de forma eficiente las dudas y preferencias de los usuarios, sin que pierdan el control de decidir y tener siempre la última palabra. (Yáñez, 2015)

8.2.2.3. Sistemas de recomendación en el comercio electrónico y la e-educación

La cantidad de información disponible y su dispersión, hace que los individuos acudan a fuentes de información externas a la hora de tomar decisiones de compra o cuando están buscando información con fines académicos y/o investigativos, a fin de obtener recomendaciones que optimicen el uso de sus recursos. En el contexto de internet, se han desarrollado aplicaciones denominadas Sistemas de Recomendación (SR), utilizadas para cumplir con estos fines, beneficiando tanto a las organizaciones como a los consumidores/usuarios. (Martinez, 2013)

Las conclusiones obtenidas indican que la investigación desarrollada en torno al papel de fuentes de información impersonales en la toma de decisiones ha sido escasa. Sin embargo, su importancia en el proceso de comercio y de educación en internet se refleja en la posibilidad de personalizar la información a las preferencias o necesidades de sus diferentes usuarios, como ha sido demostrado en las aplicaciones de SR en el ámbito comercial. (Martinez, 2013)

8.2.2.4. Sistema de recomendación de música basado en aprendizaje semi-supervisado

Actualmente los sistemas de recomendación son cada vez más utilizados por usuario y empresas que buscan por más y mejores contenidos digitales en Internet. Idealmente, los sistemas de recomendación deben aprender los gustos y preferencias de sus usuarios con la intención de facilitarles el proceso de búsqueda. A medida que el Internet se ha vuelto una fuente importante de información, tanto para usuarios expertos como para usuarios inexpertos,

también se ha convertido en un canal importante para la distribución de contenidos digitales muy diversos, por ejemplo: música, video, imágenes, etc. (Alvarado, 2015)

En particular, la búsqueda de música representa una tarea tediosa y a menudo difícil para los usuarios. Una de las razones por las que esto sucede es debido a que los usuarios buscan por contenidos que satisfagan sus gustos personales, razón principal por la cual los sistemas de recomendación se vuelven aplicaciones indispensables. Las ventajas de un sistema de recomendación no solo aplican al usuario que las utiliza; las empresas dedicadas a la venta de música por Internet pueden obtener atractivas ventajas de este tipo de sistemas. (Alvarado, 2015)

Así por ejemplo, un sistema de recomendación puede ser utilizado para ofrecer mayor diversidad de productos (contenidos musicales) relacionados a los gustos de sus consumidores y en consecuencia incrementar el número de sus ventas. En este sentido, el entender las necesidades del usuario se vuelve un factor fundamental para lograr dichos objetivos. Actualmente, este esquema de negocio es seguido por empresas como Netflix, donde el proporcionar a sus clientes contenidos apropiados a sus preferencias les ha permitido volverse un ejemplo exitoso de la utilidad de los sistemas de recomendación que aprenden a partir de los gustos de los usuarios. (Alvarado, 2015)

Un sistema de recomendación de música (Music Recommendation System o MRS por sus siglas en Inglés) es un sistema con la capacidad de proveer a sus usuarios suscritos recomendaciones musicales tomando en cuenta los gustos y preferencias del usuario o grupo de usuarios. (Alvarado, 2015)

Entre los enfoques normalmente empleados por los sistemas de recomendación podemos mencionar seis principales, los cuales son: i) Sistemas Basados en Contenido, los cuales se apoyan en la valoración dada por el usuario; ii) Filtrado Colaborativo, basado en las valoraciones que ha realizado una población y realizando un consenso de las mismas; iii) Filtrado Demográfico, basado en las características del usuario como su sexo, país, lenguaje o edad; iv) Basados en Comunidad, utilizada en las redes sociales se apoya en la valoración de las amistades del usuario para realizar recomendaciones; v) Basados en Conocimiento, son sistemas que utilizan la información proporcionada por el usuario para otorgar una clasificación, y vi) Filtrado Híbrido, que combina dos o más de los tipos de sistemas de recomendación mencionados. (Alvarado, 2015)

8.2.3. Técnicas para desarrollar filtros de información

Hoy en día, existen numerosas técnicas para desarrollar filtros de información, algunas de ellas alcanzan tasas de error inferior al 10 % en varios experimentos. Entre estas técnicas se encuentran:

8.2.3.1. Árbol de decisión

Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios. Se pueden usar para dirigir un intercambio de ideas informal o trazar un algoritmo que anticipe matemáticamente la mejor opción. (lucidchart, 2015)

Uno de los algoritmos más conocidos es el ID3 el cual nos muestra su funcionamiento de la manera siguiente:

En el grupo de datos S contiene valores positivos o negativos sobre una variable dicotómica para calcular la entropía de S relativa a su clasificación booleana se debe definir:

P_p ; es la probabilidad de que las respuestas sean positivas según el conjunto S.

P_n ; es la probabilidad de que las respuestas sean negativas según el conjunto S.

$(P_n=1-P_p)$ X son los datos.

Se puede observar que de los 14 resultados, 10 tienen resultados positivos y 4 tienen resultados negativos. La probabilidad de cada resultado es: $P_p=10/14=0,71$ y que $P_n=4/14=0,28$ La entropía de X se define con base a las probabilidades anteriores, así:

$$H(S)=-P_p \log_2 P(P-) P_n \log_2 P_n \quad (1)$$

Según la ecuación (1), la entropía del conjunto de los 14 datos respecto a la variable “Administrar fármacos” se calcula de la siguiente manera: Si la entropía toma un valor de cero es cuando todos los miembros pertenecen a una misma clase ya sea negativa o positiva debido que $\log_2(1) = 0$. Por tal motivo la entropía se encuentra siempre en un intervalo de cero a uno, alcanzando a un máximo cuando esta proporción es de 0,5 es decir existe una máxima aleatoriedad.

B. Concepto de ganancia de información $H(S) = \sum_{i=1}^C -P_i \log_2(P_i) = -\frac{10}{14} \log_2 \frac{10}{14} - \frac{4}{14} \log_2 \frac{4}{14} = 0.86$ o como se dijo anteriormente la entropía es una medida de desorden e impureza en un conjunto de datos. Para la clasificación de los datos se utiliza una medida llamada ganancia de información, esta medida reduce la entropía $H(A, S) \equiv \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} H(S_v)$ (2) obtenida al realizar la división de los datos en los subconjuntos de entrenamiento. Donde S es un grupo de muestras clasificadas en C clases, A son los atributos y Sv es un subconjunto de S Valores. “A” es una lista de los posibles valores de cada atributo. La fórmula de ganancia de información se define como: $G(S, A) \equiv H(S) - H(S, A)$ (3) En la anterior expresión el primer término H(S) corresponde a la entropía de S, el segundo término corresponde al valor esperado de la entropía después de que S ha sido particionado de acuerdo al atributo A. Como podemos observar el segundo término de la fórmula de ganancia no es más que la sumatoria de entropías de cada subconjunto Sv, ponderado por la fracción $\frac{|S_v|}{|S|}$ (Lucidchart, 2015)

8.2.3.2. Las máquinas de vectores de soporte

Una máquina de vectores de soporte (SVM) es un algoritmo de aprendizaje supervisado que se puede emplear para clasificación binaria o regresión. Las máquinas de vectores de soporte son muy populares en aplicaciones como el procesamiento del lenguaje natural, el habla, el reconocimiento de imágenes y la visión artificial. (mathworks, 2016)

Una SVM primero mapea los puntos de entrada a un espacio de características de una dimensión mayor (los puntos de entrada están en R^2 entonces son mapeados por la SVM a R^3) y encuentra un hiperplano que los separe y maximice el margen m entre las clases en este espacio se aprecia mediante la formulas en el caso 1 $W^T x + b = -1$ y caso 2 $W^T x + b = 1$

Maximizar el margen m es un problema de programación cuadrática (QP) y puede ser resuelto por su problema dual introduciendo multiplicadores de Lagrange. Sin ningún conocimiento del mapeo, la SVM encuentra el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características que son llamadas kernels. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada que son llamados vectores de soporte. (Famaf, 2017)

8.2.3.3. Redes neuronales

Las redes neuronales son un modelo para encontrar esa combinación de parámetros y aplicarla al mismo tiempo. En el lenguaje propio, encontrar la combinación que mejor se ajusta es

"entrenar" la red neuronal. Una red ya entrenada se puede usar luego para hacer predicciones o clasificaciones, es decir, para "aplicar" la combinación. (Guillermo, 2016)

Matemáticamente, la función de red de una $f(x)$ se define como una composición de otras $g_i(x)$. Este se representa como una estructura de red, con flechas que representan las dependencias entre variables. Un tipo ampliamente utilizado de la composición es la suma ponderada no lineal, donde $f(x) = k(\sum_i w_i g_i(x))$ (denominado comúnmente como la función de activación) es una función predefinida, como la tangente hiperbólica o función sigmoide. La característica importante de la función de activación es que proporciona una transición suave como valores de entrada de cambio, es decir, un pequeño cambio en la entrada produce un pequeño cambio en la producción. Será conveniente para la siguiente para referirse a una colección de funciones $g_i(x)$ implemente como un vector $g = (g_1, g_2, \dots, g_n)$ (Guillermo, 2016)

8.2.3.4. Discriminantes lineales

El Análisis Discriminante equivale a un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Se trata de encontrar relaciones lineales entre las variables continuas que mejor discriminen en los grupos dados a los objetos. Además, se trata de definir una regla de decisión que asigne un objeto nuevo, que no sabemos clasificar previamente, a uno de los grupos prefijados. (mathworks, 2016)

Considere un conjunto de observaciones x (también llamadas características, atributos, variables o medidas) para cada muestra de un objeto o evento con clase conocida y . Este conjunto de muestras es conocido como conjunto de entrenamiento. El problema de clasificación consiste en encontrar a un buen estimador para la clase y de cualquier ejemplo con la misma distribución (no necesariamente del conjunto de entrenamiento) dada solamente una observación x .

La discriminante lineal enfrenta el problema asumiendo que las funciones de densidad de las probabilidades condicionales $p(x|y = 0)$ y $p(x|y = 1)$ siguen una distribución normal con parámetros de media y covarianza (u_0, E_0) (u_1, E_1) respectivamente. Bajo esta suposición, la solución óptima de Bayes es predecir puntos como en la segunda clase si el registro de las proporciones de probabilidad está debajo de algún umbral T , de tal modo que:

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x} - \vec{\mu}_0) + \ln |\Sigma_0| - (\vec{x} - \vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x} - \vec{\mu}_1) - \ln |\Sigma_1| > T$$

Más allá de cualquier suposición el clasificador resultante es referido como análisis discriminante cuadrático (ADC).

La discriminante lineal en lugar de eso hace la suposición simplificadora adicional de homocedasticidad (las covarianzas de las clase son idénticas, con $(E0 = E1 = E)$ y que las covarianzas tengan rango completo. En este caso, varios términos se cancelan:

$$\begin{aligned} \vec{x}^T \Sigma_0^{-1} \vec{x} &= \vec{x}^T \Sigma_1^{-1} \vec{x} \\ \vec{x}^T \Sigma_i^{-1} \vec{\mu}_i &= \vec{\mu}_i^T \Sigma_i^{-1} \vec{x} \text{ pues } \Sigma_i \text{ es Hermitiano} \end{aligned}$$

y el anteriormente citado criterio de decisión se convierte en un umbral en el producto

$$w \cdot x > c$$

para alguna constante de umbral c , donde

$$\begin{aligned} \vec{w} &= \Sigma^{-1} (\vec{\mu}_1 - \vec{\mu}_0) \\ c &= \frac{1}{2} (T - \vec{\mu}_0^T \Sigma_0^{-1} \vec{\mu}_0 + \vec{\mu}_1^T \Sigma_1^{-1} \vec{\mu}_1) \end{aligned}$$

Esto quiere decir que el criterio de la entrada x , siendo y una clase, es puramente una función de esta combinación lineal de las observaciones conocidas.

Esto es frecuentemente útil para ver esta conclusión en términos geométricos: el criterio de una entrada x en una clase y es puramente una función de proyección de puntos x en el espacio-multidimensional sobre el vector w (así, sólo se considera su dirección). En otras palabras, la observación forma parte de y , si la correspondiente w está ubicada en un cierto lado de un hiperplano perpendicular a w . La localización del plano está definida por el umbral c . (mathworks, 2016)

8.3. Tipos de los algoritmos empleados en la presente investigación

8.3.1 Basados en contenido (BC)

Este grupo de algoritmos se basa en la utilización de la descripción de cada ítem para recomendar, sin utilizar información de otros usuarios para generar la recomendación al usuario objetivo. Por ejemplo, para recomendar tweets en Twitter se pueden considerar las

palabras o hashtags como la descripción del contenido del tweet (ítem) objetivo; y el contenido de los tweets escritos y retweeteados por el usuario objetivo como base para inferir los gustos de éste, a comparar con el ítem objetivo. Así, estos algoritmos intentan recomendar ítems que son parecidos a los que le han gustado al usuario anteriormente. (Pepa, 2014)

Un recomendador basado en contenido trabaja con datos que el usuario proporciona, ya sea explícitamente (calificación) o implícitamente (haciendo clic en un enlace). En función de esos datos, se genera un perfil de usuario, que luego se utiliza para hacer sugerencias al usuario. A medida que el usuario proporciona más entradas o toma medidas sobre las recomendaciones, el motor se vuelve cada vez más preciso. (Shuvayan Das, 2015)

8.3.1.1. Ventajas de los algoritmos basados en contenidos

Entre las principales ventajas de estos algoritmos tenemos: (Parra, 2016)

- A diferencia del Filtrado Colaborativo, si los ítems tienen descripciones suficientes, nos evitamos el problema del nuevo ítem, es decir todo depende del gusto del usuario.
- Las representaciones del contenido son variadas y permiten utilizar diversas técnicas de procesamiento del texto, uso de información semántica, inferencias, etc.
- Es sencillo hacer un sistema más transparente: usamos el mismo contenido para explicar las recomendaciones.

8.3.1.2. Arquitectura de un sistema de recomendación basado en contenidos

Los componentes principales son: Analizador del Contenido, aprendizaje del perfil de usuario, filtrado de contenido.

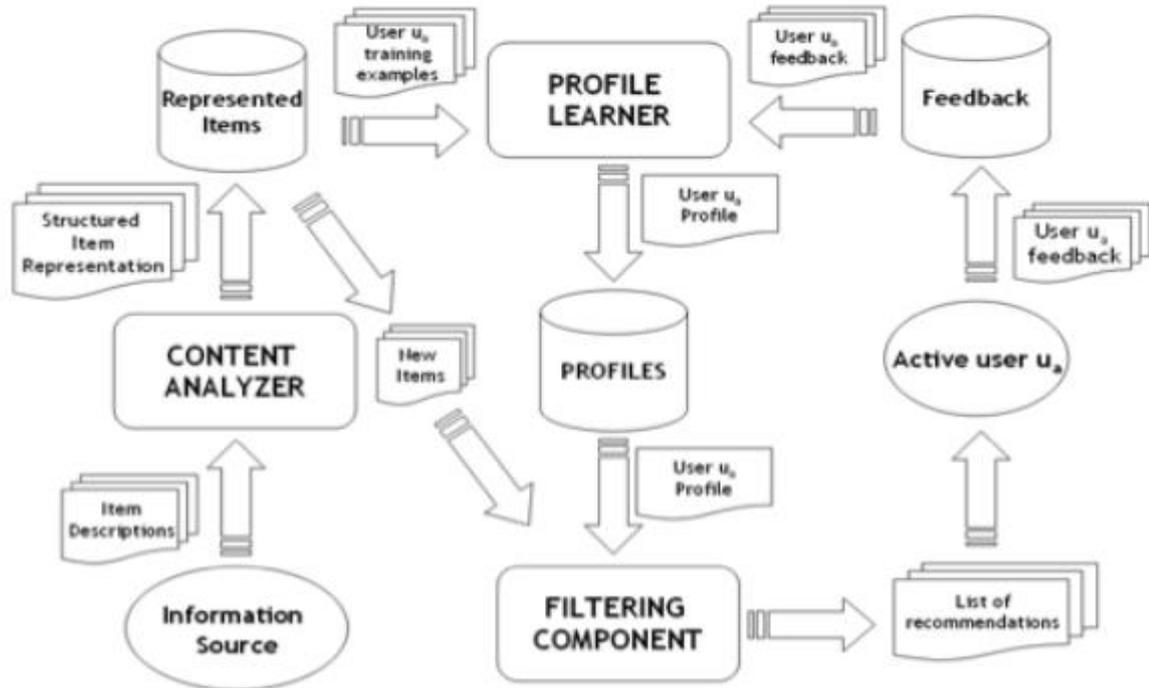
Los sistemas de filtrado de información (SI) basados en contenido necesitan técnicas adecuadas para representar los artículos y la producción del perfil del usuario, y algunas estrategias para comparar el perfil de usuario con la representación del artículo. La arquitectura de alto nivel de un contenido basado en el sistema de recomendación se muestra en la Figura 1. El proceso de recomendación se realiza en tres pasos, cada uno de los cuales es manejado por un componente separado;

- Analizador de contenido - Cuando la información no tiene estructura (por ejemplo, texto), algunos necesitan un tipo de paso denominado pre procesamiento para extraer información relevante estructurada. La principal responsabilidad del componente es

representar el contenido de los elementos (por ejemplo, documentos, páginas web, noticias, descripciones de productos, etc.) procedentes de información fuentes en una forma adecuada para los próximos pasos de procesamiento. Los elementos de datos son analizados por técnicas de extracción de características para cambiar la representación de elementos desde el espacio de información original hasta el objetivo (por ejemplo, páginas web representadas) como vectores de palabras clave). Esta representación es la entrada al perfil aprendizaje y componente filtrante. (Gemmis, 2012)

- Aprendizaje de perfil - Este módulo recopila datos representativos de las preferencias del usuario e intenta generalizar estos datos para construir el perfil del usuario. Generalmente, la estrategia de generalización se realiza a través de técnicas de aprendizaje automático, que son capaces de inferir un modelo de intereses del usuario a partir de elementos que les gusten o no me gustó en el pasado. Por ejemplo, el aprendizaje de perfil de un recomendador de página web puede implementar un método de retroalimentación de relevancia en el que el aprendizaje técnico combina vectores de ejemplos positivos y negativos en un prototipo vector que representa el perfil del usuario. (Gemmis, 2012)
- Componente filtrante - Este módulo explota el perfil del usuario para sugerir elementos relevantes al hacer coincidir la representación del perfil con la de los elementos que se van a recomendado. El resultado es un juicio de relevancia binario o continuo el último caso resulta en una lista clasificada de artículos potencialmente interesantes. En el ejemplo mencionado anteriormente, el emparejamiento se realiza mediante el cálculo de la similitud del coseno entre el vector prototipo los vectores de elementos (Gemmis, 2012)

Figura 1. Arquitectura de un sistema de recomendación basado en contenidos



Fuente: (Parra, 2016)

8.3.1.3. Modelo de espacio vectorial basado en palabras clave

La mayoría de los sistemas de recomendación basados en el contenido utilizan modelos de recuperación relativamente simples, como la concordancia de palabras clave o el Modelo de espacio vectorial (VSM) con TF-IDF básico ponderación. VSM es una representación espacial de documentos de texto. En ese modelo, cada documento está representado por un vector en un espacio n-dimensional, donde cada dimensión corresponde a un término del vocabulario general de una colección de documentos dada. Formalmente, cada documento se representa como un vector de pesos de término, donde cada el peso indica el grado de asociación entre el documento y el término. Dejar $D = \{d_1, d_2, \dots, d_N\}$ denotan un conjunto de documentos o corpus, y $T = \{t_1, t_2, \dots, t_n\}$ ser el diccionario, es decir el conjunto de palabras en el corpus. T se obtiene al aplicar algunas operaciones estándar de procesamiento de lenguaje natural, como tokenización, stopwords eliminación y derivación. Cada documento d_j se representa como un vector en una espacio vectorial n-dimensional, entonces $d_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}$, donde w_{kj} es el peso para término t_k en el documento d_j .

La representación de documentos en el VSM plantea dos problemas: ponderar los términos y midiendo la similitud del vector de características. El término de ponderación más utilizado

esquema, la ponderación TF-IDF (Frecuencia de documento de frecuencia de documento de término), es basado en observaciones empíricas con respecto al texto:

- Los términos raros no son menos relevantes que los términos frecuentes (suposición de IDF);
- Las múltiples apariciones de un término en un documento no son menos relevantes que el solo ocurrencias (suposición TF);
- Documentos largos no son preferibles a documentos cortos (suposición de normalización). En otras palabras, términos que ocurren frecuentemente en un documento (TF = term-frequency), pero rara vez en el resto del corpus (IDF = inversa-documento-frecuencia), son más probable que sea relevante para el tema del documento. Además, al normalizar el resultado del vector de peso evita que los documentos más largos tengan una mejor oportunidad de recuperación. Estas suposiciones están bien ejemplificadas por la función. (Gemmis, 2012)

$$TF - IDF(t_k, d_j) \cdot \log \frac{N}{nk}$$

Donde N denota el número de documentos en el corpus, y n k denota el número de documentos en la colección en la que el término tk aparece al menos una vez.

$$TF(t_k, d_j) = \frac{f_{k,j}}{\max_{f,j}}$$

Donde el máximo se calcula sobre las frecuencias fz, j de todos los términos tz eso ocurre en el documento dj . Para que los pesos caigan en el intervalo [0,1] y para los documentos a representar por vectores de igual longitud, pesos obtenidos por la ecuación generalmente se normaliza mediante la normalización del coseno:

$$(w_{k,j}) = \frac{TF - IDF(Tk, dj)}{\sqrt{\sum_{s=1}^T TF - IDF(ts, dj)^2}}$$

Determinar la cercanía entre dos documentos. Se han derivado muchas medidas de similitud para describir la proximidad de dos vectores; entre esas medidas, la similitud del coseno es la más ampliamente utilizado:

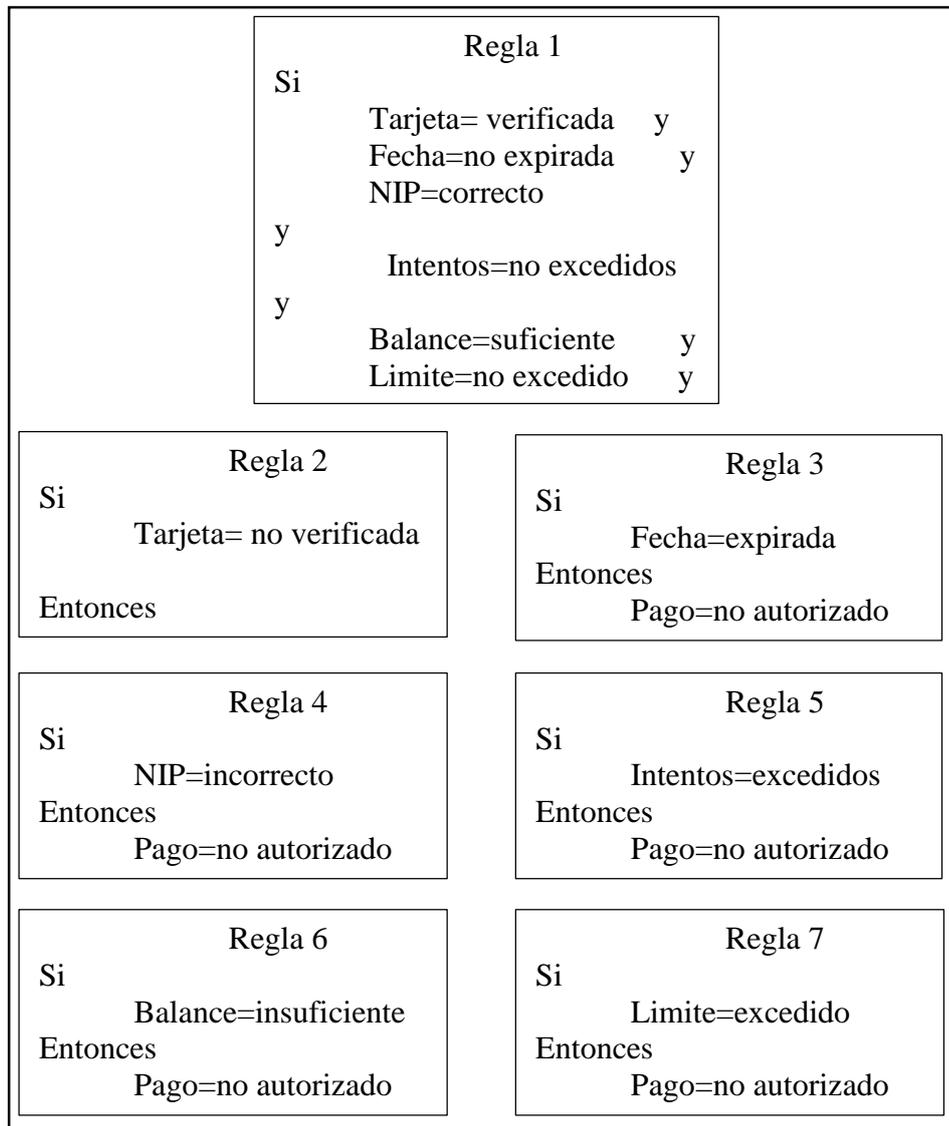
$$\text{sim}(d_i, d_j) = \frac{\sum_k w_{ki} \cdot w_{kj}}{\sqrt{\sum_k w_{ki}^2} \cdot \sqrt{\sum_k w_{kj}^2}}$$

En los sistemas de recomendación basados en contenido que dependen de VSM, tanto los perfiles de usuario como los artículos se representan como vectores de términos ponderados. Predicciones del interés de un usuario en un elemento particular se puede derivar mediante el cálculo de la similitud del coseno (Gemmis, 2012)

8.3.2. Basado en reglas

Los sistemas basados en reglas se han convertido en una de las herramientas más eficientes para tratar de manera eficiente una buena colección de problemas, ya que las reglas deterministas constituyen la más sencilla de las metodologías utilizadas en sistemas expertos. En estos sistemas, la base de conocimiento de la que se parte contiene las variables y el conjunto de reglas que definen el problema, y el motor de inferencia es capaz de extraer conclusiones aplicando métodos de la lógica clásica sobre esta base. Una regla en este contexto es una proposición lógica que relaciona dos o más objetos del dominio e incluye dos partes, la premisa y la conclusión, que se suele escribir normalmente como “Si premisa, entonces conclusión”. Cada una de estas partes es una expresión lógica con una o más afirmaciones objeto-valor conectadas mediante operadores lógicos (y, o, o no). (Sancho, 2017)

La siguiente figura es un ejemplo de las reglas que se podrían extraer de un sistema de cajero automático (Ver figura 2)

Figura 2. Reglas empleadas en un cajero automático

Fuente: (Sancho, 2017)

8.3.2.1. Ventajas

Entre las ventajas de los Sistemas Basados en Reglas, podemos destacar:

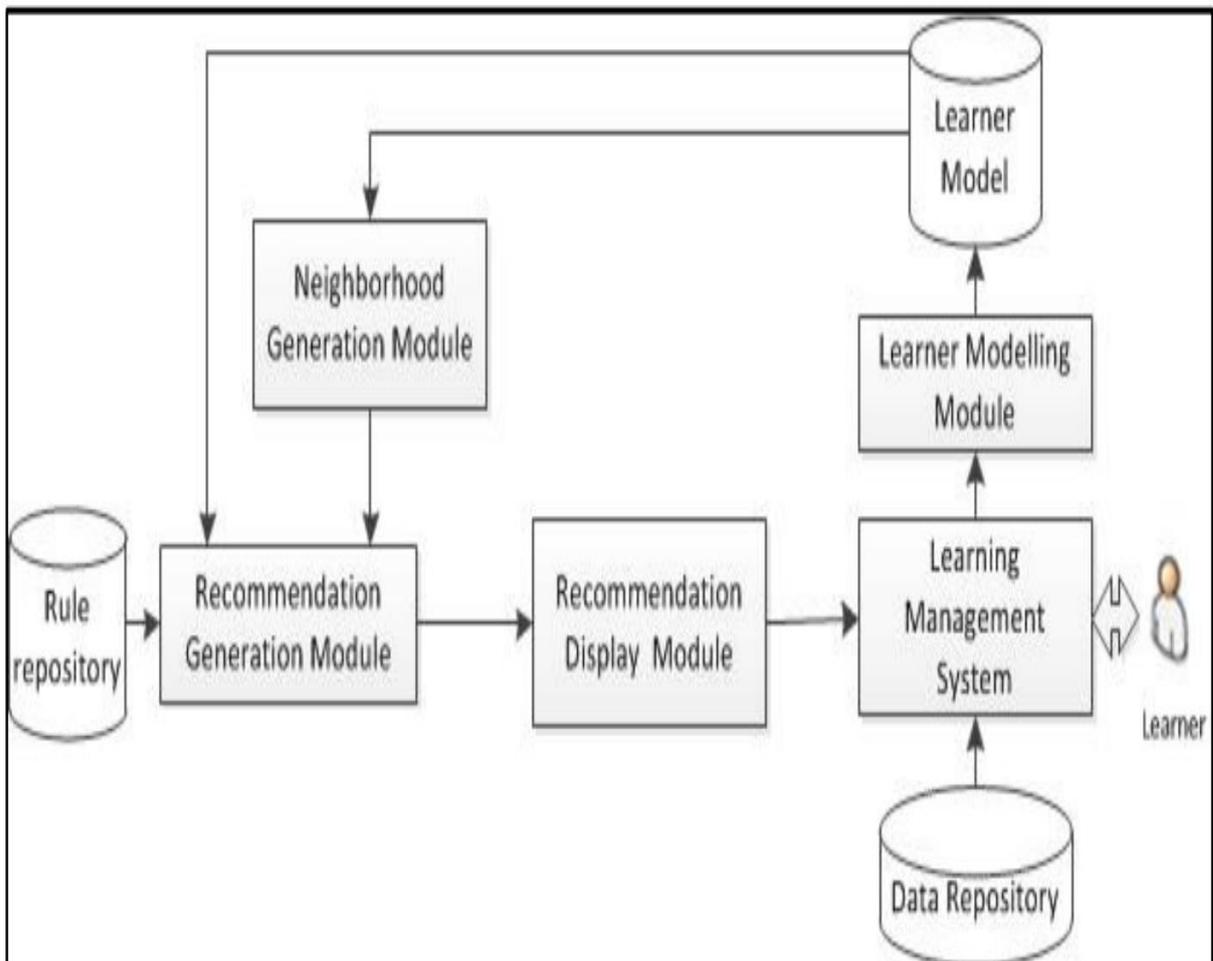
- Representan de forma natural el conocimiento explícito de los expertos: normalmente, los expertos humanos explican el procedimiento de resolución de problemas por medio de expresiones del tipo "Si estamos en esta situación, entonces yo haría esto", que se adapta fielmente al modelo seguido aquí.
- Estructura uniforme: Todas las reglas de producción tienen la misma estructura "Si entonces". Cada regla es una pieza de conocimiento independiente de las demás.
- Separación entre la base de conocimiento y su procesamiento.

- Capacidad para trabajar con conocimiento incompleto e incertidumbre (introduciendo variantes) (Werbach, 2015)

8.3.2.1. Arquitecturas de sistemas basados en reglas.

El sistema ha sido diseñado para integrarse en cualquier sistema de gestión de aprendizaje. La arquitectura del sistema de recomendación propuesto se presenta en la figura 3

Figura 3. Arquitecturas de sistemas basados en reglas.



Fuente: (Imran, 2014)

Como se muestra en la Figura. 3, el sistema de recomendación consta de cuatro módulos:

- Learner Modeling Module tiene como objetivo recopilar información sobre el aprendizaje de los usuarios basados en el modelo de estilo de aprendizaje, conocimiento previo, nivel de experiencia y rendimiento.

- Neighborhood Generation Module tiene como objetivo encontrar usuarios que comparten las mismas características como un usuario objetivo (un usuario para quien se calcula una recomendación).
- El módulo de generación de recomendaciones tiene como objetivo generar recomendaciones adecuadas para un usuario objetivo. Estas recomendaciones se generan en base a un conjunto de reglas que se utilizan para clasificar la idoneidad de los niveles de dificultad y luego seleccionar el tareas de aprendizaje apropiadas.
- El módulo de visualización de recomendaciones muestra recomendaciones para el usuario. (Imran, 2014)

8.3.3. Filtrado colaborativo

El filtrado colaborativo es una técnica utilizada por los sistemas de recomendación para solventar los problemas derivados de la sobreinformación que los consumidores sufren en Internet. Esta tendencia crece cada día más, debido a su enorme funcionalidad son más los usuarios que se valen de esta herramienta en sus búsquedas.

Antes del nacimiento de Internet el consumidor no tenía ninguna fuente de información salvo la propia publicidad del producto. El mercado ha pasado de esta escasez de información a la saturación de los mismos. En este contexto, surgen los filtrados colaborativos. Las empresas incorporan estas herramientas en su página y los propios usuarios construyen una inteligencia colectiva mediante un sistema de recomendaciones que son luego estudiados y traducidos mediante algoritmos estadísticos. Una de las empresas pioneras en incorporar esta herramienta dentro de su web fue la famosa tienda online Amazon.com, que informa a sus usuarios de los productos que podían interesarles partiendo de los que ya había clicado. (Del Campo, 2017)

Los filtrados colaborativos sirven para hacer predicciones automáticas sobre los intereses de un usuario mediante la recopilación de preferencias o gustos del mismo consumidor u otros consumidores con intereses comunes.

8.3.3.1. Dificultades que podemos encontrar cuando usamos el filtrado colaborativo

- Escasez de datos: los sistemas de filtrado colaborativo se basan en conjuntos de datos. Si esta muestra de datos es escasa puede ser muy costosa y poco eficaz. En ocasiones un problema común es empezar de cero, ya que no se pueden recopilar preferencias con precisión y fiabilidad **(Del Campo, 2017)**
- Sinónimos: la diversidad de etiquetas con nombres similares a veces no son reconocidos por los sistema de filtrados cuando en realidad el usuario está buscando el mismo elemento y se pierde información. Por ejemplo: un usuario que busca ordenadores o computadoras, son sinónimos pero el buscador no los relaciona. **(Del Campo, 2017)**
- Shilling attacks: en los sistemas de recomendación cualquiera puede hacer evaluaciones, pudiendo un usuario votar positivamente sólo a sus productos y servicios y dar negativo a sus competidores, falseando la eficacia de esta herramienta. **(Del Campo, 2017)**
- Diversidad: los filtros intentan buscar una diversidad para poder recomendar entre múltiples opciones. En ocasiones este filtro van reduciendo esta variedad dando sólo visibilidad a los productos con mayor popularidad. **(Del Campo, 2017)**

8.3.3.2. Filtrado colaborativo basado en el usuario

Hay dos opciones, la correlación de Pearson o la similitud del coseno. Deje $u_{\{i, k\}}$ denota la similitud entre el usuario i y el usuario k y $v_{\{i, j\}}$ denota la calificación que el usuario i le da al elemento j con $v_{\{i, j\}} = ?$ si el usuario no ha calificado ese artículo. Estos dos métodos se pueden expresar como los siguientes:

$$u_{ik} = \frac{\sum_j (v_{ij} - v_i)(v_{kj} - v_k)}{\sqrt{\sum_j (v_{ij} - v_i)^2 \sum_j (v_{kj} - v_k)^2}}$$

Formula de Person

$$\cos(u_i, u_j) = \frac{\sum_{k=1}^m v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^m v_{ik}^2 \sum_{k=1}^m v_{jk}^2}}$$

Formula de la similitud del coseno

Ambas medidas se usan comúnmente. La diferencia es que la Correlación de Pearson es invariante al agregar una constante a todos los elementos. (Huang, 2018)

8.3.4. Basado en enfoques mixtos

Un enfoque mixto, como el que tiene Spotify con sus playlist diarias, posiblemente sería el más acertado, detectando, por ejemplo, cuál es el tema más interesante para ese usuario y haciendo las recomendaciones sobre ese tema (filtrado basado en contenido + filtrado colaborativo). Posteriormente podrían añadirse reglas para ponderar estas recomendaciones basadas en aspectos demográficos (filtrado basado en reglas), pero, en principio, el propósito de este ejercicio es demostrar la implementación de una arquitectura funcional que cumpla el propósito de la recomendación de la manera más comprensible. El razonamiento detrás del enfoque mixto es poner primero los atributos relacionados a las últimas actividades del usuario (enfoque basado en el contenido), pero deja una gran cantidad de posiciones de puntaje subsecuentes para el siguiente acercamiento. (Glauber, 2015)

8.4. Data sets que utilizan filtrado basado en contenido

8.4.1. Amazon

Amazon, S.L. es una compañía estadounidense de comercio electrónico y servicios de computación en la nube a todos los niveles con sede en la ciudad estadounidense de Seattle, Estado de Washington. Su lema es From A to Z (traducido al español: «De la A a la Z»). Fue una de las primeras grandes compañías en vender bienes a través de Internet. Amazon también posee Alexa Internet, a9.com, Shopbop, Internet Movie Database (IMDb), Zappos.com, DPreview.com y Twitch.

Durante la década 2000-2010, Amazon ha desarrollado una base de clientes de alrededor de 30 millones de personas. Amazon.com es principalmente un sitio minorista con un modelo de ingresos por ventas. Amazon hace su dinero tomando un porcentaje del precio de venta de cada artículo que se vende a través de su sitio web. Amazon también permite a las compañías anunciar sus productos pagando para ser listadas como productos (Marquero, 2017)

8.4.1. My BestBet

MyBestBets es una aplicación web y móvil única que ayuda a los jóvenes adultos de bajos ingresos a tomar las decisiones correctas en lo que respecta a la educación postsecundaria y

las vías de capacitación: elecciones que llevarán a carreras de alta demanda y alto crecimiento.

Al usar MyBestBets, los jóvenes pueden:

- Identificar sus intereses, talentos y aspiraciones;
- Utilizar conjuntos de datos sofisticados y visualmente dinámicos para investigar posibles carreras;
- Organizar y rastrear el progreso hacia sus metas postsecundarias;
- Manténgase conectado con el personal y los entrenadores profesionales que se preocupan;
- Desarrollar habilidades y capital social a través de experiencias del mundo real; y
- Asumir la propiedad y dar forma de forma activa a su propia trayectoria educativa y profesional

MyBestBets es vanguardista, tanto en la combinación como en la presentación de datos personales y de mercado de trabajo de formas muy atractivas y en el apoyo a un enfoque de aprendizaje combinado que profundiza sus relaciones con sus asesores o asesores. (Mybestbets, 2010)

8.4.1. Jinni

La página web de recomendación de películas más popular es Jinni, perteneciente a la compañía Entertainment Genome. El principal objetivo es recomendar películas acordes a los gustos del usuario. Sin embargo, Jinni ofrece más funciones como la búsqueda semántica, enfocada en la interpretación de las consultas mediante la identificación de conceptos en el contenido, en lugar de keywords; el filtrado a través del estado de ánimo, el argumento que se desea, el tiempo disponible u otros parámetros. Además, incluye opciones adicionales para refinar la búsqueda. La tecnología de Jinni implica una taxonomía creada por profesionales del cine, con nuevos títulos indexados a través de procesamiento del lenguaje natural y los métodos de aprendizaje automático para analizar automáticamente los exámenes y los metadatos (Jannach, 2015)

9. HIPÓTESIS

¿Cuáles son los factores que influyen en los algoritmos de filtrado basados en contenidos para la búsqueda de información?

10. METODOLOGÍA DE LA INVESTIGACIÓN

10.1. Investigación descriptiva

Determinar las estructura y el comportamiento más importante de los algoritmos de filtrado basados en contenido con lo cual ayudará a la comprensión del funcionamiento del algoritmo mencionado. Y así el resultado del estudio permitirá ver la necesidad de implementar los algoritmos a las aplicaciones de desarrollo que faciliten las búsquedas de sus contenidos.

10.2. Diseño Descriptiva

Mediante la aplicación empleada para demostrar el algoritmo de filtrado basado en contenidos se determinó que los resultados presentados están en función a la preferencia de búsqueda del usuario.

11. ANÁLISIS DE RESULTADOS

11.1. Análisis descriptivo de sistemas basados en contenidos

Los sistemas de recomendación pueden basarse en las preferencias del usuario, en su perfil sociodemográfico o de cliente, en las características de los productos y objetos a recomendar, es una forma más común de que los sistemas de filtrado pueden personalizar su información, para ofrecer al usuario lo que está buscando incluso antes de que el mismo sepa ahorrándole tiempo y ofreciendo una experiencia única en base al sistema que se encuentra navegando. Las recomendaciones a un usuario se basan en sus propias preferencias, buscando elementos similares a aquellos por los que ha mostrado interés. Así el presente algoritmo se basa en la constancia de uso, es decir, mientras más interactúa el usuario con la aplicación esta ira almacenando lo más explorado y así tiene una base informativa de lo que puede necesitar en la siguiente interacción.

La aproximación de obtener un filtrado de contenido es entre el 81 al 83% de probabilidad

Prácticamente un recomendador basado en contenido trabaja con datos que el usuario proporciona, ya sea explícitamente (calificación) o implícitamente (haciendo clic en un enlace). En la presente práctica realizada se muestra el ingreso de libros que pertenecen a su respectivo autor, en donde un usuario puede acceder a cada libro y leerlo es ahí donde el

algoritmo basado en contenido actúa, al registrar el número de libros leídos de cada autor entonces el programa hace referencia a los gustos por autor, es decir, si leyó más de dos veces al autor x, la aplicación asume que le gusta ese autor y este le presenta más opciones o libros de este autor. (Ver Anexo, Figura 3)

En la siguiente figura se muestra un formulario para el ingreso de un nuevo libro, esto nos servirá para realizar los filtrados. (Ver Anexo, Figura 4)

Una vez ingresado los datos del libro se realizó una simulación en la cual un lector ingresa al mismo libro tres veces, el algoritmo realiza el filtrado en base al número de veces leída y muestra como favoritos a los libros que presenten dicha condición. (Ver Anexo, Figura 5)

11.2. Análisis descriptivo de sistemas basados en reglas

Es un sistema que remueve la información redundante o no deseada de un flujo de información utilizando métodos semiautomáticos o computarizados previo a la presentación de la información al usuario así como también las recomendaciones se basan en reglas previamente definidas como se muestra en la práctica realizada al diseñar un grafo de cinco vértices a las cuales les dimos valores así también como a sus aristas o distancias el cual mediante el manejo del algoritmo de filtrado se logró conseguir el camino más rápido desde un vértice inicial para llegar a un vértice final.

De manera similar, muchos bancos y aseguradoras online personalizan banners y otros acuerdo al segmento de cliente: jóvenes, pensionistas, autónomos, etc.

En esta parte se realiza el análisis de un sistema de filtrado en base a un ejemplo práctico, simulado diferentes camino unidos por puntos al cual un usuario no conoce la forma más rápida de trasladarse de un punto a otro, es decir se presenta el requerimiento que necesita para realizar la funcionalidad del ejemplo que se presentará. Basándose en un sistema que remueve la información redundante al emplear un algoritmo para la determinación del camino más corto dado un vértice origen al resto de vértices en un grafo con pesos en cada arista (Ver Anexo, Figura 6)

En la codificación se emplea el cálculo matemático para encontrar la vía más rápida entre punto y punto (Ver Anexo, Figura 7)

El resultado se refleja al introducir datos entre vértices que tiene el grafo y al ejecutar el botón calcular camino me da como resultado los vértices que forman la ruta más corta. (Ver Anexo, Figura 8)

11.3. Análisis descriptivo de algoritmos de filtrado colaborativo:

Este algoritmo presenta un enfoque en el cual se compara lo que gusta o no a otra gente para predecir las preferencias de un usuario dado, necesitan gran cantidad de información sobre las interacciones llevadas a cabo por los usuarios del sistema para generar consultas específicas.

Los resultados obtenidos de acuerdo al sistema desarrollado es la representación del algoritmo colaborativo ya que se registra información de usuarios relacionándose a la creación de un perfil en donde de acuerdo al perfil seleccionado se le muestra otros usuarios que comparten algún aspecto en común en este caso se observa de acuerdo a la ubicación ingresada en un principio.

Para el análisis de este sistema de filtrado me basaré en un ejemplo práctico, teniendo en cuenta que este sistema utiliza recomendaciones como la comparación de usuarios similares, en este ejemplo un usuario se registra utilizando sus datos personales los cuales serán registrados en un a base de datos, un caro ejemplo lo tenemos con el perfil de usuario que se crean en Facebook, en donde se ingresa nombres y direcciones de usuarios para luego empleando el algoritmo de filtrado colaborativo se presente los usuarios que se relacionan por el sitio en donde se encuentran.

El código utilizado realiza el proceso de registro de datos del usuario (Ver Anexo, Figura 9)

La codificación del algoritmo empleado para el filtrado basado en contenido se refleja a continuación (Ver Anexo, Figura 10)

En la siguiente figura se simula una lista de varios usuarios en la cual se deberá seleccionar uno de ellos para obtener resultados de búsqueda filtrados por la dirección del usuario seleccionado (Ver Anexo, Figura 11)

En esta figura se muestra los resultados de la búsqueda que se realiza al seleccionar un usuario, presentando personas que quizá conozca basados en la dirección del usuario seleccionado (Ver Anexo, Figura 12)

11.4. Filtrado basado en enfoques mixtos

Este tipo de algoritmos combinan diferentes enfoques. Por ejemplo, pueden basarse en las preferencias de usuarios de gustos similares dentro de un determinado segmento de cliente, o en las preferencias de usuarios respecto a productos de una determinada categoría. Mediante este algoritmo se puede tener sugerencias de productos similares, esto ayuda a que el usuario no tenga que realizar más búsquedas, así tenemos en la aplicación realizada, que si escogió manzanas el sistema emite consejos en base a su categoría en este caso frutas y sugiere productos similares a este.

Para el filtrado de información basado en enfoques mixtos se realizara básicamente un aplicación que permite realizar una búsqueda de algo específico de acuerdo al ingreso de productos con su respectiva categoría y al seleccionar un producto simularé un carrito de compras el cual contiene productos asociados al ingresado, es decir aparte del producto seleccionado a comprar, se presentara una serie de sugerencias relacionados al producto escogido.

En el código utilizado se emplea el ingreso del nombre del producto, su categoría y la lógica empleada para demostrar el algoritmo (Ver Anexo, Figura 13)

En la siguiente interfaz permite al usuario ingresar productos en base a una categoría (Ver Anexo, Figura 14)

Una vez registrado los productos el sistema permite enlistar los productos sin importar la categoría que posee para que el usuario pueda seleccionar uno de ellos y realizar una compra (Ver Anexo, Figura 15)

Con la implementación del algoritmo basado en enfoques mixtos en esta aplicación se obtiene el resultado del producto seleccionado y la sugerencia de productos en base a su categoría (Ver Anexo, Figura 16)

11.5. Análisis de los cuatro algoritmos.

En la experiencia adquirida en el uso de los cuatro algoritmos mencionados en ejemplos anteriores, utilizando la descripción y el funcionamiento obtenido por las referencias bibliográficas puedo decir que, su eficiencia depende del sistema donde se pude aplicar y el número de registros a comparar para obtener los resultados de filtrado, con mi criterio el

algoritmo de filtrado de contenidos es el mejor ya que tenemos recomendaciones más acertadas al gusto y preferencias del usuario, la aplicación del algoritmo en el ejemplo ya mencionado se basa en las veces que el usuario ingresaba al mismo contenido para ello se realizó un contador el cual se iba incrementando hasta ponerlo de favorito, ya una vez como favorito obtenía los datos de referencia esto ayuda a que las búsquedas o sugerencias sean más acertadas.

11.5.1. Comparación de Data sets que utilizan filtrado basado en contenido

A continuación presentamos a las siguientes empresas Jinni, My BestBest y Amazon que están utilizando los algoritmos de filtrado basados en contenidos. 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 basado en contenido

Tabla 3. Tabla de software que emplea filtrado basado en contenido

Cantidad de usuarios	Aplicaciones	Lenguaje	Versión	Fecha de actualización	Tipo de algoritmo	Ejemplo práctico
270.000	Jinni	Python	1.1.1	05 Agosto 2015	Basado en reglas	Grafo esquemático de conexión de puntos
59,681	My BestBet	C++	2.1	01 Marzo 2016	Basado en contenido	Preferencia de usuarios
30 millones	Amazon	JavaScript PHP	10	15 abril 2018	Basado en enfoques mixtos	Sugerencia de productos

Fuente. El investigador

Se puede apreciar que Amazon está en constante actualización. Esto es una gran ventaja con respecto al resto, considerando la rapidez que progresan los algoritmos cada año. Y los usuarios de Amazon son de 30 millones, es evidente que es el mayor en manipular información de alta cantidad.

Los lenguajes de programación de Amazon, es JavaScript y PHP (acrónimo recursivo de PHP: Hypertext Preprocessor) son lenguajes de código abierto muy popular especialmente adecuado para el desarrollo web.

En base a este análisis comparativo se puede decir que el tipo de algoritmo basado en enfoques mixtos que utiliza Amazon al ser una de las empresas globales en el rating de ventas en línea, es el más utilizado hasta la fecha en relación de otros tipos de algoritmos. Hasta la fecha han utilizado el tipo de algoritmo mixto y utilizan lenguajes de programación más populares y lo más importante que cuentan con una mayor cantidad de información para ser procesados y presentar mayor exactitud en las preferencias del usuario al momento de realizar recomendaciones.

11.6. Data sets (Conjunto de datos)

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 empresa globales que utilizan algoritmos de filtrado basado en contenido.

11.6.1. Jinni

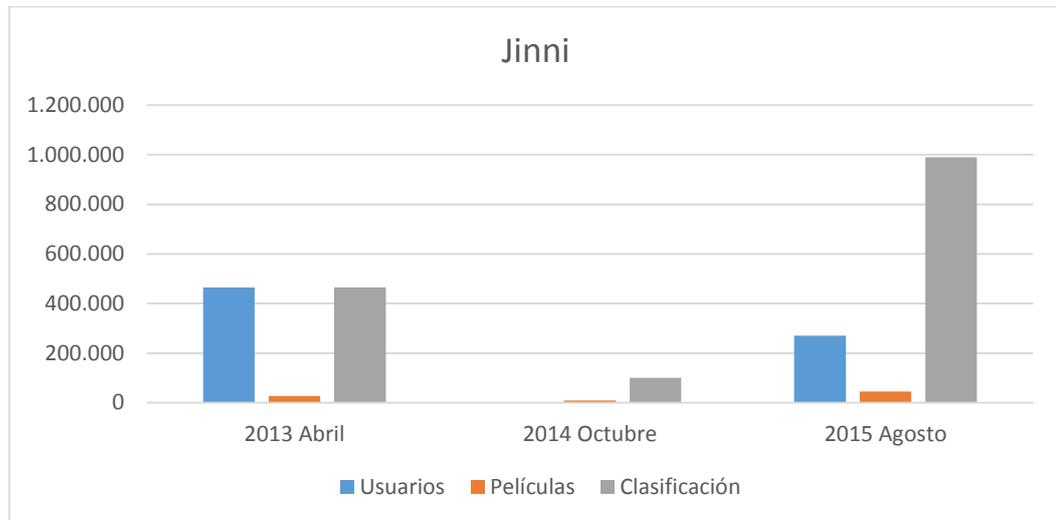
Jinni es página web de recomendación de películas más popular perteneciente a la compañía Entertainment Genome. El principal objetivo es recomendar películas acordes a los gustos del usuario. Sin embargo, Jinni ofrece más funciones como la búsqueda semántica, enfocada en la interpretación de las consultas mediante la identificación de conceptos en el contenido. Los datos han sido recolectados durante varios periodos de tiempo, dependiendo del tamaño del conjunto (Jannach, 2015)

Jinni ha publicado dos datasets estables hasta la fecha agosto del 2015, los cuales se detallan a continuación:

Tabla 4. Datos de Jinni

Aplicaciones	Clasificaciones	Películas	Usuarios	Fecha
-	465.000	27.000	465.000	04/2013
1.300	100.000	9.000	700	10/2014
750.000	990000	45.000	270.000	08/2015

Fuente. El investigador

Figura 4: Clasificaciones más altas

Fuente: El investigador

Se puede observar que las clasificaciones de las películas se elevó en Agosto del 2015 referente dos atrás.

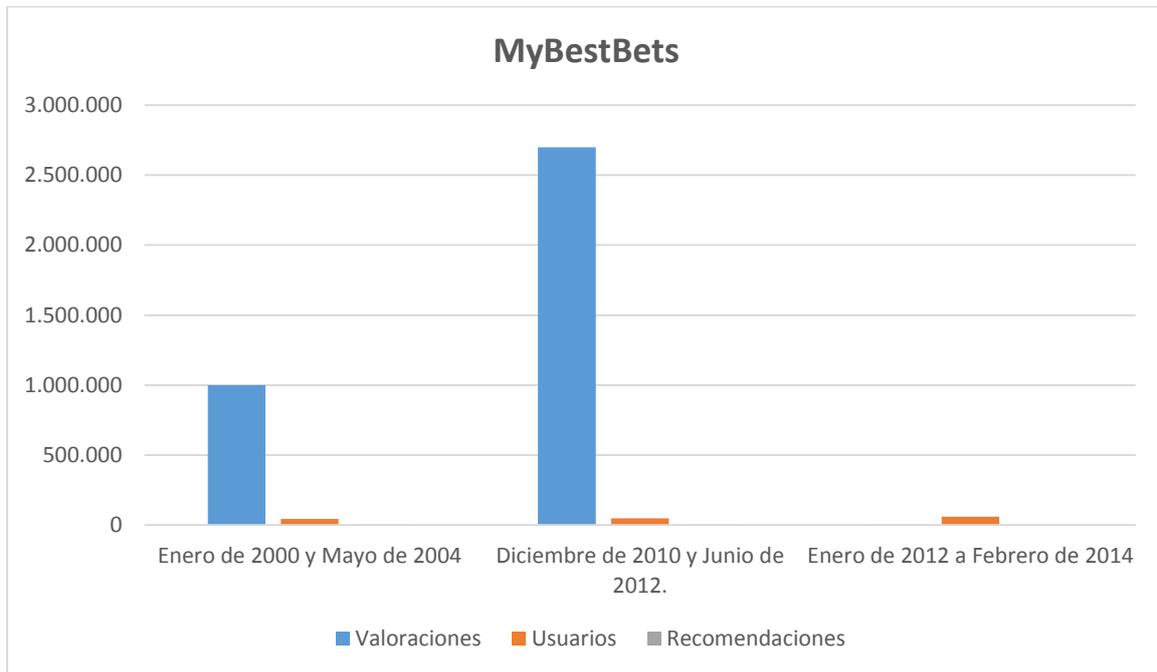
11.6.2. My BestBet

MyBestBets es una aplicación web y móvil única que ayuda a los jóvenes adultos de bajos ingresos a tomar las decisiones correctas en lo que respecta a la educación postsecundaria y las vías de capacitación: elecciones que llevarán a carreras de alta demanda y alto crecimiento.

Tabla 5. Tabla de valoraciones de la herramienta MyBestBets

Conjuntos	Valoraciones	Recomendaciones	Usuarios	Fecha
Agrupación de datos 1	más de 1.1 millones	(-1.00 a +50.00) de 50	43.421	Enero de 2000 y Mayo de 2004
Agrupación de datos 2	más de 2.7 millones	(-1.00 a +50.00) de 100	49.132	Diciembre de 2010 y Junio de 2012.
Agrupación de datos 2+	con más de 4,000		59,681	Enero de 2012 a Febrero de 2014

Fuente. El investigador

Ilustración 1: Valoraciones más altas

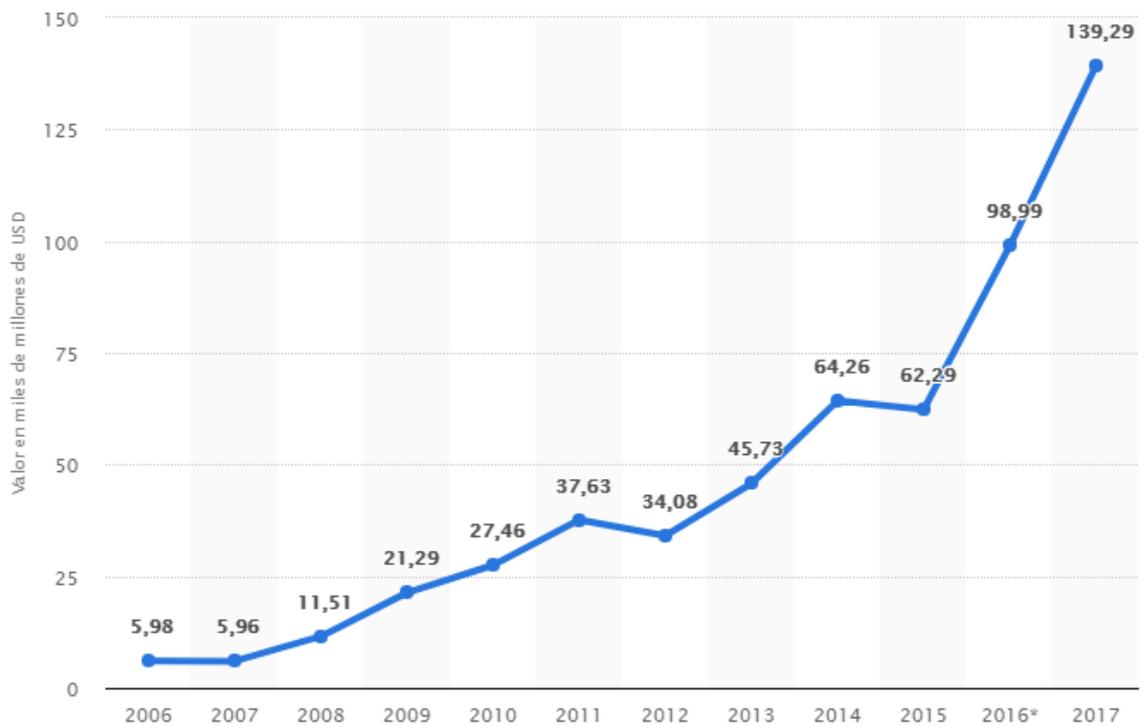
Fuente: El investigador

Se puede observar que las valoraciones de toma de decisiones para las capacitaciones se elevó en Diciembre de 2010 y Junio de 2012, mientras tanto de Enero del 2012 a Febrero del 2014 se ha bajado.

11.6.3. Amazon

Amazon, S.L. es una compañía estadounidense de comercio electrónico y servicios de computación en la nube a todos los niveles con sede en la ciudad estadounidense de Seattle, Estado de Washington. Su lema es From A to Z (traducido al español: «De la A a la Z»). Fue una de las primeras grandes compañías en vender bienes a través de Internet. Amazon también posee Alexa Internet, a9.com, Shopbop, Internet Movie Database (IMDb), Zappos.com, DPreview.com y Twitch.

Durante la década 2000-2010, Amazon ha desarrollado una base de clientes de alrededor de 30 millones de personas. Amazon.com es principalmente un sitio minorista con un modelo de ingresos por ventas. Amazon hace su dinero tomando un porcentaje del precio de venta de cada artículo que se vende a través de su sitio web. Amazon también permite a las compañías anunciar sus productos pagando para ser listadas como productos.

Ilustración 2: Ventas de Amazon

Fuente: El investigador

Se puede observar que la línea de ventas de Amazon sigue aumentando cada año.

12. IMPACTOS TÉCNICOS Y SOCIALES

12.1. Impacto técnico

En la investigación realizada se demuestra que las técnicas empleadas por este algoritmo son las referencias e interés del usuario con el contenido agilizando los procesos de búsqueda y generando recomendaciones a su necesidad.

12.2. Impacto social

La presente investigación tiene un aporte positivo en presentar a los estudiantes e investigadores los beneficios que tienen los algoritmos de filtrado basado en contenidos como es el brindar un acceso eficiente a la búsqueda de información.

12.3. Impacto ambiental

Al obtener más búsquedas más rápidas y concretas los usuarios suelen utilizar mucho la computadora o dispositivos que interactúan con información esto hace que de manera inconsciente utilicemos más recursos como electricidad, consecuentemente este recurso emplea agua y la modificación del entorno natural para crear centrales eléctricas para su funcionamiento.

Por otra parte el recurso tecnológico reduce la utilización de papel para conocer información de lo que desearía conocer los usuarios. Además contribuye a que las bibliotecas no mantengan documentos físicos.

12.4. Impacto económico

El presente estudio no presenta ningún costo en buscar la información que la comunidad universitaria requiere.

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

Recursos	Cantidad	Unidad	V. Unitario (\$)	Valor Total
Equipos	1	Laptop	0	0
Materiales y suministros	1	Resma de papel boom y material estudiantil.	50	50
Recursos Básicos	1	Energía eléctrica, internet, agua.	240	240
SUB TOTAL				290
IVA 12%				34.8
TOTAL 2				324.8

Fuente: El investigador

El proyecto de investigación tiene una inversión total de 324.8 USD.

14. CONCLUSIONES Y RECOMENDACIONES

14.1. Conclusiones

- Los algoritmos basados en contenidos realizan recomendaciones utilizando las características del usuario y los ítems que le han interesado al usuario, es un paso más para que el manejo de los datos se convierta en el futuro en el activo más importante para la humanidad.
- Se comparó cuatro algoritmos de filtrado de información para presentar el funcionamiento de cada uno y de esta manera demostrar la lógica que tienen en el filtrado de información, destacando el filtrado basado en contenido como un algoritmo que se basa en las preferencias del usuario.
- De acuerdo a los principales algoritmos de filtrado de información seleccionados se analizó su funcionamiento, de esta manera se desarrolló una aplicación web para cada una y así se demuestra el proceso que realiza el algoritmo en estudio en comparación a los otros algoritmos.
- Este tipo de investigación son excelentes para la implementación en aplicaciones informáticas, pero también se debe tomar en cuenta que son muy costosos.

14.2. Recomendaciones

- Se sugiere profundizar la investigación acerca de los algoritmos para filtrado de contenidos.
- Desacollarar sistemas recomendadores para una mejor interacción entre el usuario.
- Tomar como base las aplicaciones realizadas en el presente estudio para implementar en aplicaciones existentes.
- Analizar la funcionalidad del sistema a que ámbito va ser aplicado, ya que el algoritmo de contenido se basa en una estructura básica para su desempeño.

15. BIBLIOGRAFÍA

- Alvarado. (12 de 10 de 2015). *Sistema de recomendaciones de música basado en aprendizaje semi-supervisado*. Obtenido de <https://pdfs.semanticscholar.org/ee6b/83078bdd410fd6a37822a6dd1c119ea3a24c.pdf>
- Balavanovic y Shoham . (4 de 10 de 2016). *Sistemas de recomendación*. Obtenido de <http://ict.udlap.mx/people/lulu/documento/capitulo4.html>
- Cadena y Carantón. (15 de 09 de 2017). *repository.udistrial.edu.co*. Obtenido de <http://repository.udistrial.edu.co/bitstream/11349/7498/1/CadenaD%C3%ADazXeniaViviana2017.pdf>
- Chesani. (14 de 10 de 2014). *Recommmendation Systems*. Obtenido de <https://www.unipi.it/index.php/lauree/corso/10276>
- Del Campo. (04 de 08 de 2017). *E-Commerce*. Obtenido de https://www.iebschool.com/blog/filtrado-colaborativo-sirve-e-commerce/#filtrado_colaborativo
- es.estadistica. (2017). *es.estadistica*. Obtenido de <https://es.statista.com/estadisticas/656838/valor-de-marca-global-de-amazon/>
- Famaf. (2017). *famaf.unc.edu*. Obtenido de www2.famaf.unc.edu.ar/institucional/biblioteca/trabajo/601/17565.pdf
- Gemmis. (2012). <http://facweb.cs.depaul.edu>. Obtenido de <http://facweb.cs.depaul.edu/mobasher/classes/ect584/Papers/ContentBasedRS.pdf>
- Glauber. (2015). <http://ceur-ws.org>. Obtenido de <http://ceur-ws.org/Vol-1120/paper2.pdf>
- Guillermo. (26 de 01 de 2016). *xataka*. Obtenido de <https://www.xataka.com/robotica-e-ia/las-redes-neuronales-que-son-y-por-que-estan-volviendo>
- Guillermo. (01 de 06 de 2017). *SISTEMAS DE RECOMENDACIÓN BASADOS EN*. Obtenido de <http://eprints.ucm.es/43975/1/Sistemas%20de%20Recomendaci%C3%B3n%20basados%20en%20t%C3%A9cnicas%20de%20predicci%C3%B3n%20de%20enlaces%20para%20jueces%20en%20l%C3%ADnea%20-%20Marta%20Caro%20Mart%C3%ADnez.pdf>
- Huang, S. (28 de 06 de 2018). *hackernoon*. Obtenido de <https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75>
- IMGBIBLIO. (10 de 08 de 2014). *Evaluación de sistemas recomendadores de contenidos educativos*. Obtenido de <http://imgbiblio.vaneduc.edu.ar/fulltext/files/TC114681.pdf>
- Imran, H. (2014). *researchgate*. Obtenido de https://www.researchgate.net/publication/279752268_A_Rule-Based_Recommender_System_to_Suggest_Learning_Tasks?enrichId=rgreq-e9222921847c99bbc5c62c04369fca62-

XXX&enrichSource=Y292ZXJQYWdlOzI3OTc1MjI2ODtBUzoyNDgwMDgwNzU5MDI5NzZAMTQzNjE0MTE0MTY0Mg%3D%3D&el=1

Jannach. (05 de 10 de 2015). *Evaluacion de sistemas de recomendadores de contenidos educativos*. Obtenido de <http://imgbiblio.vaneduc.edu.ar/fulltext/files/TC114681.pdf>

lucidchart. (12 de 10 de 2015). *lucidchart*. Obtenido de <https://www.lucidchart.com/pages/es/qu%C3%A9-es-un-diagrama-de-%C3%A1rbol-de-decisi%C3%B3n>

Marquero, M. (16 de Agosto de 2017). *La nación*. Obtenido de <https://www.nacion.com/economia/negocios/amazon-acelera-contrataciones-en-costa-rica-con-la-apertura-de-2-500-plazas-mas/S3E5SCRXSBCFNLSHVKS53EGRVU/story/>

Martinez. (12 de 08 de 2013). *Sistemas de recomendación en el comercio electrónico y la Educación*. Obtenido de <http://www.unilibre.edu.co/CriterioLibre/images/revistas/12/CriterioLibre12art08.pdf>

mathworks. (04 de 10 de 2016). *mathworks*. Obtenido de <https://es.mathworks.com/discovery/svm-maquina-vectores-soporte.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>

Monsalve. (04 de 05 de 2015). *El marketing y los sistemas de recomendación*. Obtenido de <https://www.brainsins.com/es/blog/el-marketing-y-los-sistemas-de-recomendacion/99688>

Mybestbets. (31 de 07 de 2010). *mybestbets*. Obtenido de <http://www.mybestbets.org/>

Parra. (23 de 08 de 2016). *Filtro basado en contenido*. Obtenido de http://dparra.sitios.ing.uc.cl/classes/recsys-2016-2/clase6_contentBased_1.pdf

Pepa. (10 de 02 de 2014). *Suite de algoritmos de recomendación*. Obtenido de <http://imgbiblio.vaneduc.edu.ar/fulltext/files/TC114681.pdf>

Posada-Prieto. (04 de 04 de 2015). *El funware en los sitios de recomendación social*. Obtenido de <https://franganillo.es/funware.pdf>

Rodriguez. (2016). Sistema de recomendación de objetos. *Revista Científica Teknos*, 10.

Sancho. (15 de 10 de 2017). *Fernando Sancho Caparrini*. Obtenido de <http://www.cs.us.es/~fsancho/?e=103>

Shuvayan Das. (11 de 08 de 2015). *Beginners Guide to learn about Content Based Recommender Engines*. Obtenido de <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/>

Velez. (29 de 05 de 2016). Recuperado el 16 de 11 de 2017

Werbach. (04 de 04 de 2015). *El funware en los sitios de recomendación social*:. Obtenido de <https://franganillo.es/funware.pdf>

Yáñez. (12 de 10 de 2015). *Sistema de recomendación con conocimiento web*. Obtenido de <http://eprints.ucm.es/32695/>

ANEXOS

Anexo1.- Glosario de términos

Termino	Descripción
IMDb	Base de datos de películas en Internet
Jannach	Empresario que basa sus productos en sistemas de recomendación
People-Based marketing	Es una herramienta que saber cómo adaptarse a las necesidades y gustos de cada persona
E-retailers	Venta al detalle o Comercio minorista
Cross-selling	En marketing, se llama 'venta cruzada '
Ludificación	Es el uso de técnicas, elementos y dinámicas propias de los juegos y el ocio en actividades no recreativas con el fin de potenciar la motivación
Gamificación	Es una técnica de aprendizaje que traslada la mecánica de los juegos al ámbito educativo-profesional con el fin de conseguir mejores resultados
Funware	Generalmente emplea mecánicas de juego como puntos, tablas de clasificación, insignias, desafíos y niveles.
IMDb	Base de datos de películas en Internet
TripAdvisor	Es un sitio web estadounidense que proporciona reseñas de contenido relacionado con viajes.

Anexo2.- Fotografías.

Fotografía1. Aplicación web para demostrar algoritmo del filtrado basado en contenido



Fuente: El investigador

Figura 5. Codificación del algoritmo de filtrado basado en contenido

```
8         $this->usuario->insertarLibro($data);
9         redirect('contenido/index','refresh');
10    }
11    public function guardarleido()
12    {
13        $existe=$this->usuario->buscarUsuarioVentas($this->input->post('libro'));
14        if($existe){
15            $id=$existe->id_ve;
16            $nuca=$existe->contador+1;
17            $data = array(
18                'contador' => $nuca
19            );
20            $this->usuario->actualizar($id,$data);
21            redirect('contenido/index','refresh');
22        }else{
23            $idlibro=$this->usuario->buscarTodosLibroId($this->input->post('libro'));
24            $data = array(
25                'id_li' => $this->input->post('libro'),
26                'autor' => $idlibro->autor,
27                'id_usu' => 1,
28                'contador' => 1
29            );
30            $this->usuario->insertarLeidos($data);
31            redirect('contenido/index','refresh');
32        }
33    }
34    }
35    public function actualizarContenido($id)
36    {
37        $data = array(
38            'contador' => 0
39        );
40        $this->usuario->actualizar($id,$data);
41        redirect('contenido/index','refresh');
42    }
43    }
```

Fuente: El investigador

Figura 6. Ingreso de datos del libro



Formulario de ingreso de datos del libro. El formulario está contenido en un recuadro azul con un efecto de sombra. En la parte superior del recuadro hay una barra azul con los íconos de redes sociales: Facebook (f), Twitter (t), LinkedIn (in) y YouTube (v). El formulario mismo contiene tres campos de entrada blancos con bordes azules: el primero con el texto "Ingrese Nombre", el segundo con "Ingrese la descripción", y el tercero con "--seleccione--" y una flecha hacia abajo. Debajo de estos campos hay un botón rectangular negro con el texto "INGRESAR" en letras blancas mayúsculas.

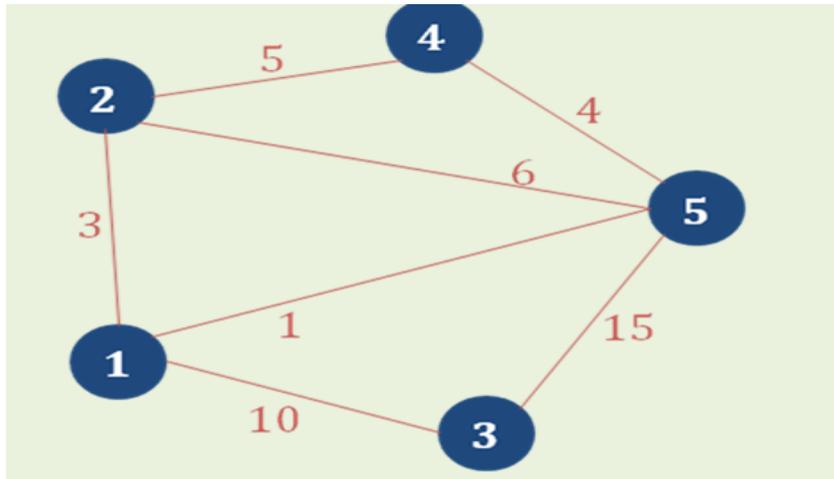
Fuente: El investigador

Figura 7. Filtrado de libros de acuerdo a la preferencia del usuario



Fuente: El investigador

Figura 8. Grafo esquemático de conexión de puntos



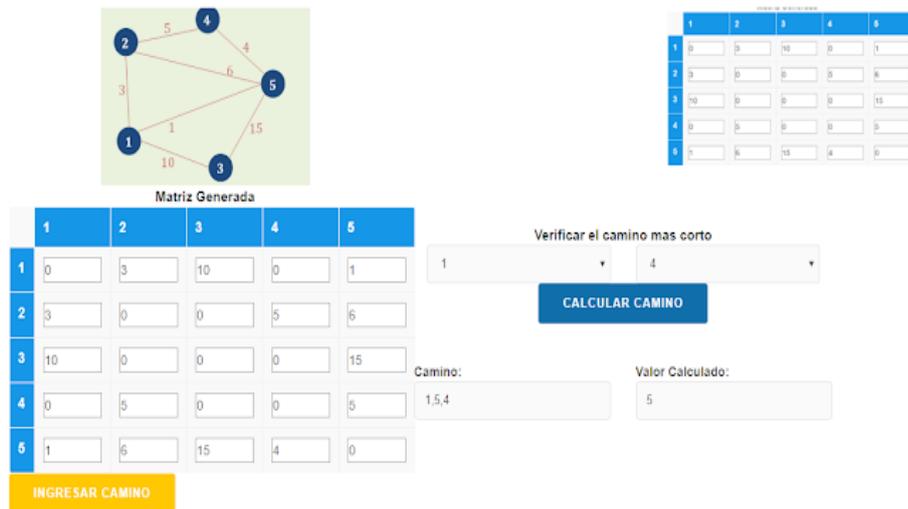
Fuente: El investigador

Figura 9. Algoritmo de filtrado basado en reglas.

```
199 mat=<</table>;
191 $("#contenedorPeso").html(mat);
192 }
193 var objetoAux1=new Object();
194 var matriz;
195 function guardarPesos(){
196     control=true;
197     for(i=0;i<5;i++){
198         for(j=0;j<5;j++){
199             if($("#*"+i+j).val()==""){
200                 control=false;
201                 i=999;
202             }
203         }
204     }
205 }
206 if(control){
207     $("#desde").empty();
208     $("#hasta").empty();
209     $("#contenedorRutas").fadeIn();
210     for(i=0;i<5;i++){
211         for(j=0;j<5;j++){
212             var auxiliar_valor_0=parseInt($("#*"+i+j).val());
213             if(auxiliar_valor_0!=0){
214                 matriz[i][j]=auxiliar_valor_0;
215             }else{
216                 matriz[i][j]=100000;
217             }
218         }
219     }
220 }
221 }
222 }
223 }
224 }
225 }
226 }
227 }
228 desdeAux=<option value="">Inicio</option>;
229 hastaAux=<option value="">Final</option>;
230 }
```

Fuente: El investigador

Figura 10. Representación del algoritmo de filtrado basado en reglas.



Fuente: El investigador

Figura 11. Ingreso de datos para el ejemplo de algoritmo colaborativo

```

public function index()
{
    $data['usu'] = $this->usuario->buscarTodosUsuarios();
    $this->load->view('cabeza');
    $this->load->view('colaborativo', $data);
    $this->load->view('pie');
}

public function guardar()
{
    $data = array(
        'nombre' => $this->input->post('nombre'),
        'direccion' => $this->input->post('direccion')
    );
    $this->usuario->insertarUsuarios($data);
    redirect('colaborativo/index', 'refresh');
}

public function listado($nombre, $id)
{
    $datos['usuario'] = $this->usuario->buscar($nombre, $id);
    $this->load->view('listado', $datos);
}
}
    
```

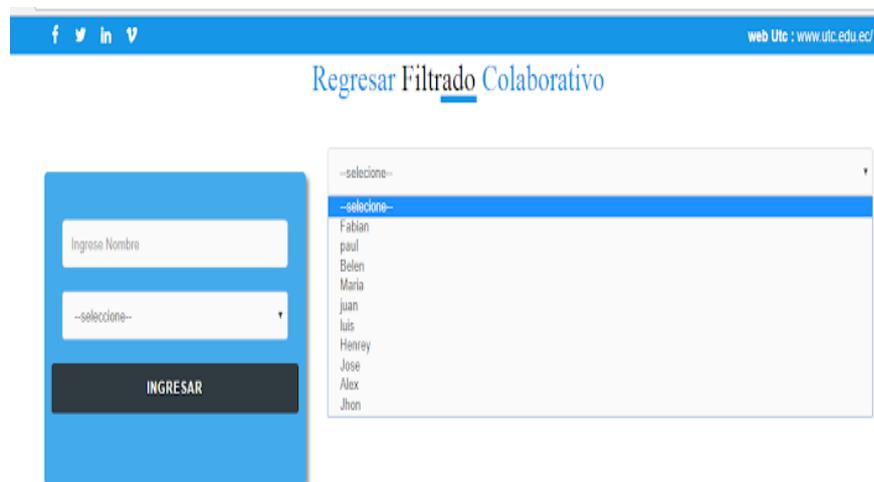
Fuente: El investigador

Figura 12. Representación del algoritmo colaborativo

```
parent::__construct();
}
public function insertarUsuarios($data){
    $this->db->insert('persona', $data);
    return $this->db->affected_rows();
}
public function buscar($data,$id)
{
    $this->db->like('direccion', $data);
    $this->db->where('id !=',$id);
    $query=$this->db->get('persona');
    if ($query->num_rows()>0) {
        return $query;
    }else{
        return false;
    }
}
public function buscarTodosUsuarios(){
    $query = $this->db->get('persona');
    if($query -> num_rows > 0){
        return $query;
    }else{
        return false;
    }
}
public function buscarTodosArticulos(){
    $this->db->order_by("nombre", "asc");
    $query = $this->db->get('articulos');
    if($query -> num_rows > 0){
        return $query;
    }else{
        return false;
    }
}
```

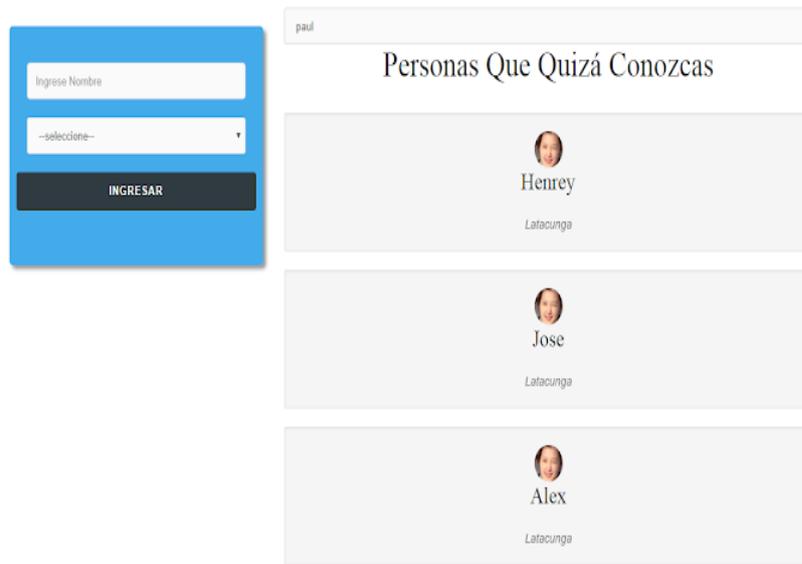
Fuente: El investigador

Figura 13. Interfaz para ingresar datos del usuario



Fuente: El investigado

Figura 14. Resultado del filtrado de búsqueda.



Fuente: El investigador

Figura 15. Representación del algoritmo de filtrado basado en enfoques mixto

```
}
public function buscarTodosArticulos(){
    $this->db->order_by("nombre", "asc");
    $query = $this->db->get('articulos');
    if($query->num_rows > 0){
        return $query;
    }else{
        return false;
    }
}

public function insertarArticulo($data){
    $this->db->insert("articulos", $data);
    return $this->db->affected_rows();
}

public function buscarTodosArticulosId($id){
    $query = $this->db->get_where('articulos', array('id' => $id));
    if($query->num_rows > 0){
        return $query->row();
    }else{
        return false;
    }
}

public function buscarSimilares($data, $id){
    $this->db->order_by("nombre", "asc");
    $query = $this->db->get_where('articulos', array('categoria' => $data, 'id !=' => $id));
    if($query->num_rows > 0){
        return $query;
    }else{
        return false;
    }
}
}
```

Fuente: El investigador

Figura 16. Interfaz para registrar productos por categoría.



Regresar Filtrado Mixto

Ingrese Nombre

--seleccione--

INGRESAR

Fuente: El investigador

Figura 17. Visualización de productos ingresados.



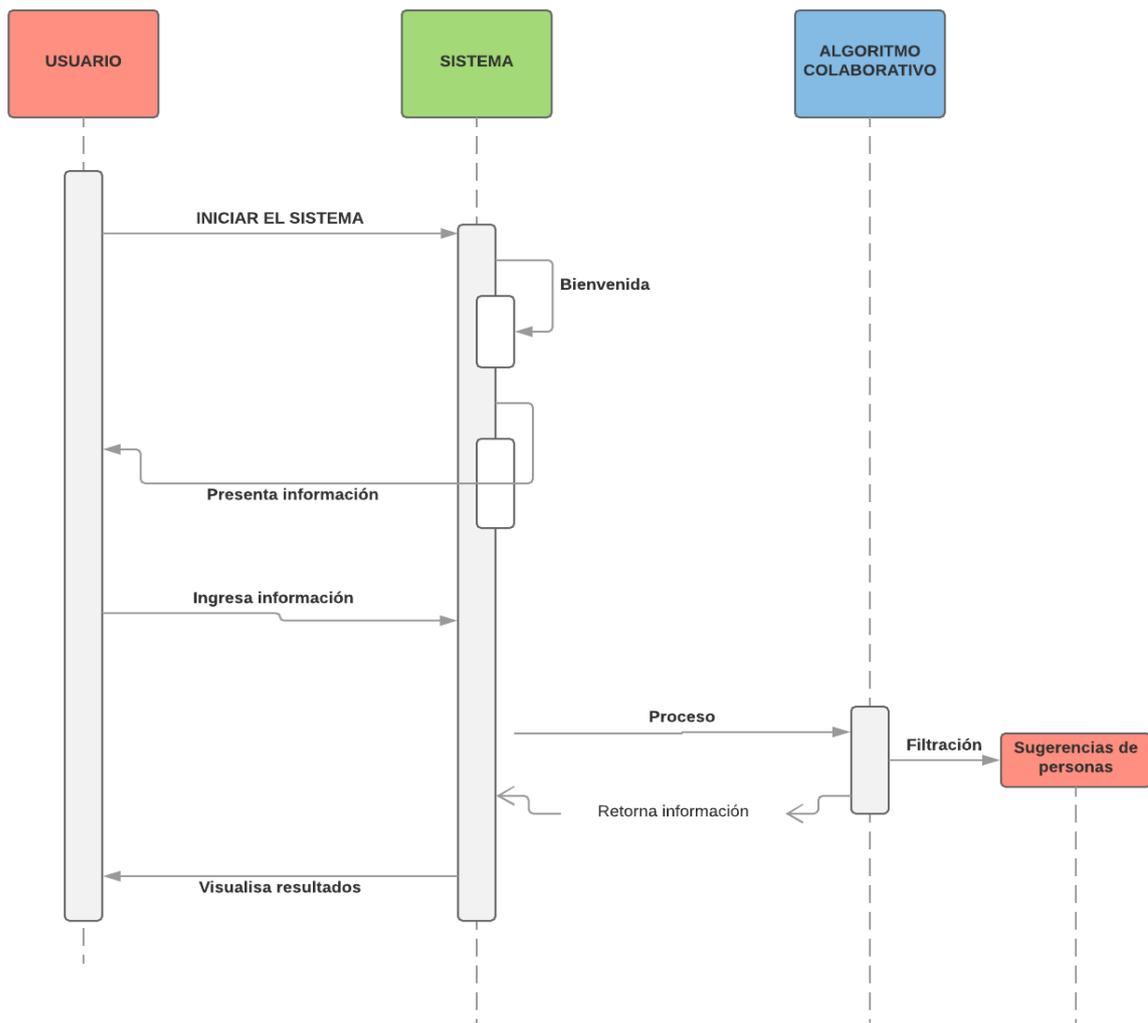
Fuente: El investigador

Figura 18. Simulación de carrito de compras con sugerencia de productos



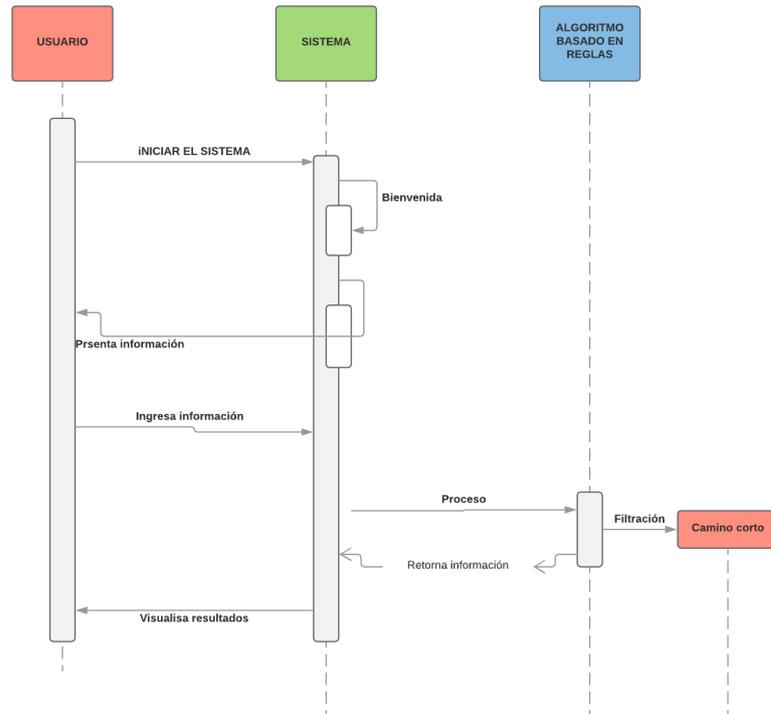
Fuente: El investigador

Figura 19: Diagrama de secuencia básico de Sugerencia de personas aplicando algoritmo colaborativo.



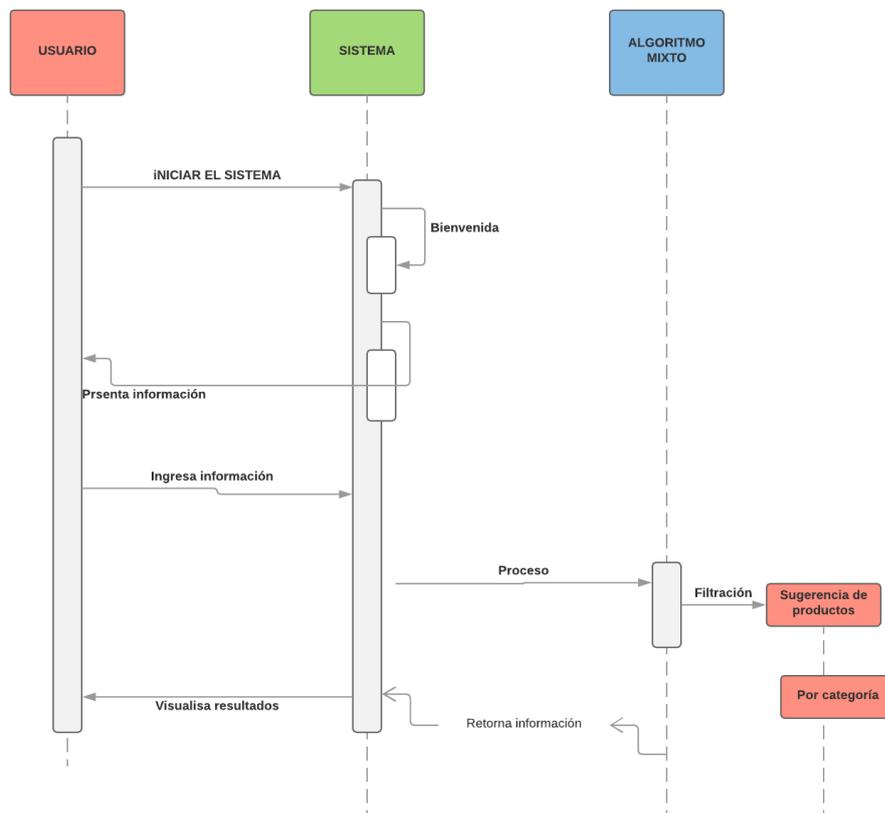
Fuente: El investigador

Figura 20: Diagrama de secuencia básico de camino corto aplicando algoritmo basado en reglas.



Fuente: el investigador

Figura 21: Diagrama de secuencia básico de sugerencia de productos por categoría aplicando algoritmo mixto.



Fuente: El investigador