

Asesor virtual para la recomendación de películas y series de entretenimiento con base en preferencias y opiniones del usuario.

Trabajo Terminal No. 2023-A026

Alumnos: Becerra Ramírez Luis Arturo, Lira Dávila Alan David, Ramírez Gutiérrez Enrique

Directores: Fabiola Ocampo Botello

e-mail: alirad1500@alumno.ipn.mx

Resumen – Hoy en día, el medio audiovisual cuenta con un catálogo de películas y series muy extenso, siendo uno de los entretenimientos más populares entre las personas, es por ello que existen distintas plataformas con servicio de streaming que ofrecen cientos o quizás miles de títulos en sus catálogos. Este trabajo tiene el propósito de desarrollar un asesor virtual para la recomendación de distintas películas y series mediante el perfil e historial de usuario al igual que sus preferencias y opiniones registradas dentro de la aplicación.

Palabras Clave – Bases de Datos, Ingeniería de Software, Minería de Datos, Minería de Textos

1. Introducción.

Con el paso del tiempo han aparecido diferentes servicios de streaming, por ejemplo, Netflix, Amazon Prime Video, Disney+, etc., donde el número de usuarios aumenta considerablemente año con año, y teniendo en cuenta el confinamiento global debido a la pandemia del COVID-19 el tráfico en internet aumentó y los usuarios necesitaban contenido audiovisual para distraerse [1], la opción más lógica era utilizar alguna de las plataformas mencionadas anteriormente. La mayoría de las plataformas de streaming ofrecen títulos exclusivos, es por eso que gran parte de los usuarios contratan más de un servicio, y es aquí donde se genera el problema, el catálogo a disposición del usuario no es compartido entre plataformas, y aumenta considerablemente el número de películas o series para ver y las personas pueden perderse fácilmente entre tanto contenido sin encontrar lo que más les gustaría consumir. Por lo tanto, se planea desarrollar una página web que implemente técnicas de minería de datos como parte del método de descubrir conocimiento en base de datos (Knowledge Discovery in Databases) para que los usuarios que utilicen la página obtengan recomendaciones sobre los títulos que les podrían gustar basados en los perfiles e historial de usuario.

El método KDD se refiere al proceso de identificar patrones válidos, novedosos, potencialmente útiles y principalmente entendibles. Buscando obtener patrones que puedan ser útiles usando como base, datos obtenidos de las bases de datos. Existen algunas etapas para el desarrollo de este proceso [2].

- Selección de datos.
- Almacenamiento de datos.
- Implementación del almacén de datos.
- Limpieza de datos.
- Procesamiento.
- Selección de técnica.
- Transformación de datos.
- Minería de datos.
- Evaluación.
- Difusión.

Además existen algoritmos que permiten generar recomendaciones tales como el filtrado basado en contenido (Content Based Filtering), cuyo funcionamiento es recomendar al usuario productos similares basado en los gustos del usuario en el pasado, distancia Euclidiana, este algoritmo en espacios n dimensionales encuentra

elementos similares cercanos entre sí, lo que permite recomendar elementos en función de la distancia, correlación de Pearson, nos dice cuanto se correlacionan dos elementos, cuanto más alta sea la correlación, más será la similitud, y filtrado colaborativo (Collaborative Filtering) este algoritmo utiliza comportamiento del usuario para recomendar elementos [3].

Un asesor virtual es un agente de software que ayuda a los usuarios a automatizar y realizar tareas con la mínima interacción hombre-maquina [4], ahora bien, el asesor virtual a desarrollar será capaz de ofrecer recomendaciones basándose en el perfil de usuario empleando alguno o algunos algoritmos de recomendación anteriormente mencionados sin excluir la utilización de algún otro algoritmo existente.

Los agentes de software ofrecen diversos beneficios a los usuarios finales al automatizar tareas repetitivas. Dichos agentes no requieren interacción del usuario y funcionan como agente para un usuario u otro programa, trabajando de forma autónoma y continua en un entorno en particular. pero de igual forma puede aprender de su experiencia en el funcionamiento en un entorno durante un largo periodo de tiempo [5].

Existen seis clasificaciones de agentes:

1. Agentes de interfaz de usuario. Dan asistencia al usuario para que aprenda a utilizar alguna aplicación, este tipo de agentes funcionan de manera gráfica, guiando al usuario paso a paso para completar una tarea.
2. Agentes colaborativos o cooperativos: Trabajan junto con otros agentes para la realización de sus tareas, se usan en sistemas donde se presentan funcionalidades que solo pueden ser obtenidas gracias a su trabajo en conjunto.
3. Agentes móviles: Son capaces de viajar por la WAN y la WWW, recabando información en beneficio de su propietario y regresando después de haber ejecutado las tareas asignadas por el usuario.
4. Agentes de información: Recopilan información de distintas fuentes para dar respuestas a las cuestiones planteadas por los usuarios.
5. Agentes reactivos: Responden a estímulos del entorno en el que se encuentran.
6. Agentes híbridos: Son la combinación de dos o más clasificaciones de agentes (interfaz, móvil, colaborativo, etc.). De este modo se maximizan las habilidades del agente [6].

En la tabla 1 se describen los productos con características similares a las cuales se propone desarrollar en este proyecto.

SOFTWARE	CARACTERÍSTICAS	PRECIO EN EL MERCADO
Buaala (https://www.buaala.com/)	El usuario recibe sugerencias personalizadas de los contenidos que más se adecuan a el y a sus amigos. Cuanto más participe el usuario dentro de Buaala, más lo conocerá y mejores resultados ofrecerá.	Acceso gratuito.
JustWatch (https://www.justwatch.com/)	Es la guía de televisión definitiva para plataformas de transmisión. Permite al usuario navegar por los catálogos de casi todos los servicios que existen, incluidos Netflix y Amazon Prime Vídeo.	Acceso gratuito.
TV Time (https://www.tvtime.com/)	Es una aplicación en donde el usuario puede seguir fácilmente los programas que ve regularmente y recibe notificaciones cuando se emiten nuevos episodios.	Acceso gratuito.

Watcha (https://watcha.com/)	Pide al usuario que califique algunas películas y programas de TV antes de comenzar. Una vez hecho esto, crea una calificación pronosticada, que es una estimación de cómo calificaría el usuario algo que aún no ha visto.	Acceso gratuito.
Solución Propuesta	Con base a las preferencias e historial de usuario y haciendo uso de técnicas de minería de datos se realizarán recomendaciones de interés a los usuarios	Acceso gratuito.

Tabla 1. Resumen de productos similares.

2. Objetivo.

General.

Desarrollar una página web que funcione como un asesor virtual, recomendando series y películas considerando un catálogo de productos de esta naturaleza, tomando como base las preferencias y opiniones de los usuarios.

Específicos.

- Desarrollar un asesor virtual capaz de recomendar series y películas con base en las preferencias e información recopilada de las opiniones de los usuarios.
- Generar una relación de las preferencias y opiniones de los usuarios para recomendarle otras películas y series que puedan ser de su agrado.
- Implementar un método de encriptación por MD5 para agilizar la carga de trabajo que realiza el servidor al momento de cifrar o descifrar los datos de los usuarios.
- Identificar los aspectos funcionales del sistema a desarrollar.
- Diseñar el modelo de datos que contenga los datos necesarios para generar las recomendaciones de series y películas.
- Diseñar la arquitectura del sistema.
- Analizar algunas técnicas de minería de datos para algoritmos de sistemas de recomendación.

3. Justificación.

Durante los dos últimos años, los servicios de streaming en México se han vuelto mucho más populares, según el Instituto Federal de Telecomunicaciones (IFT) los mexicanos aumentaron en un 77% el gasto mensual en estas plataformas [7], además se reveló en un estudio realizado en el año 2020 por Sherlock Communications que en México el 22% de los entrevistados está suscrito a por lo menos 4 plataformas diferentes [8]. Es aquí donde se ha encontrado un problema para los consumidores, muchos de ellos están suscritos a más de una plataforma y cada una de ellas ofrece cientos de series y películas en sus catálogos por lo cual es complicado para los usuarios encontrar un título que se adapte a sus preferencias y gustos.

Según un comunicado realizado por Ana Fernanda de las Fuentes, directora de Estrategias de Educación Financiera en Cumplo México, los millennials suelen ser el sector de la población que más gasta en las suscripciones de servicios de streaming [9], es por ello que los millennials podrían ser nuestros usuarios potenciales, ya que ellos buscan siempre estar al tanto de las últimas novedades en cuanto a contenidos multimedia y por supuesto cuentan con muchos de estos servicios.

Nuestra propuesta para resolver este problema es desarrollar un asesor virtual que sea capaz de recomendar alguna serie o película que pueda ser del agrado del usuario con base en sus preferencias y opiniones de otros usuarios. En esta página web, los usuarios podrán establecer cuáles son las plataformas de streaming que tienen contratadas para que el asesor virtual pueda recomendar cualquier título que esté disponible en alguna de las plataformas que el usuario consume y que este dentro del conjunto de datos del sistema.

Ahora bien, el asesor virtual que se pretende desarrollar en este proyecto pertenece a la clasificación de agentes de software colaborativos, trabajando en conjunto con el agente de información ya que este recopila la información de las series y películas que se encuentran en el conjunto de datos, dando una recomendación final al usuario.

Es así que para este proyecto se planea contar con dos conjuntos de datos, el primero almacenará información relevante del catálogo de las series y películas existentes en algunos servicios de streaming, y el segundo se utilizará para almacenar la información del perfil del usuario.

Actualmente, cada una de las plataformas ofrece recomendaciones basadas únicamente en su catálogo, ignorando completamente si sus usuarios poseen otras plataformas de streaming, también muchas de ellas no consideran la posibilidad de que un perfil sea utilizado por diferentes personas, lo cual afecta su algoritmo de recomendación. Nuestra propuesta no tendría este problema ya que cada usuario registra los títulos que personalmente ha visto por lo que se genera un perfil más detallado de cada usuario.

4. Productos o Resultados esperados.

Se espera presentar para la evaluación de TT1 el sprint 1 que consiste en el diseño de la interfaz de usuario junto con el sprint 2 en donde se planea integrar la base de datos para almacenar la información del usuario, así como también un avance en la documentación del sistema.

Para la evaluación de TT2 se tiene planeado entregar el sistema completo, junto con su documentación y manuales técnico y de usuario.

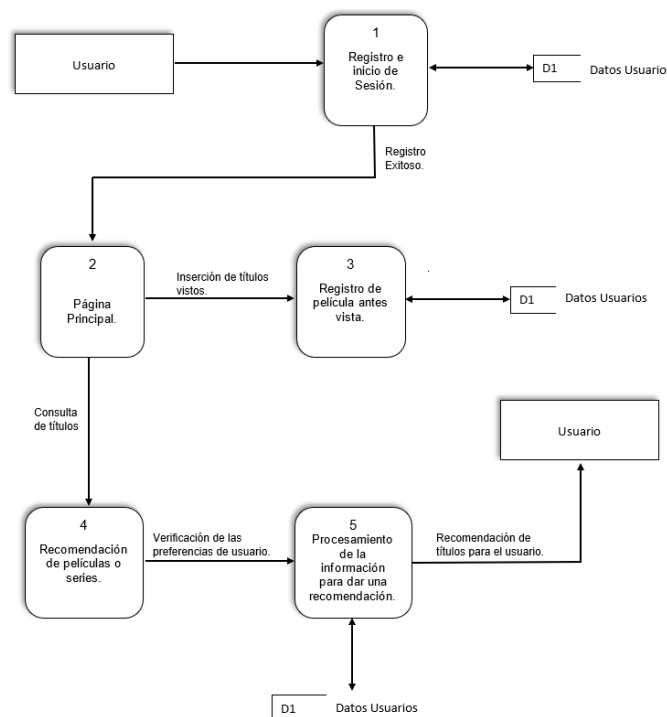


Figura 1. Arquitectura del sistema.

5. Metodología.

Debido a la naturaleza del proyecto, es importante hacer uso de una metodología ágil que nos brinde flexibilidad, a su vez, sea colaborativa y nos permita el desarrollo iterativo. Las características de la metodología Scrum son idóneas para la realización de este proyecto [10].

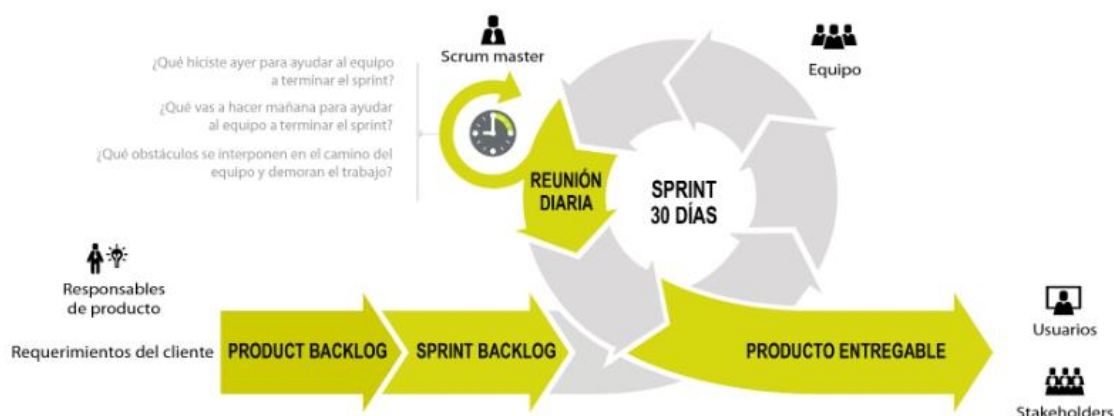


Figura 2. Diagrama de la metodología scrum.

Imagen tomada de <https://www.leadersummaries.com/es/libros/resumen/detalle/metodologia-scrum>

- Definir el Product Backlog, con los lineamientos aportados por el cliente.
- Conformar el equipo que llevará a cabo el Sprint.
- Decidir el tiempo de duración de cada Sprint (iteración).
- Designar un Scrum Master que oficie de líder para ese equipo particular, durante el trabajo.
- Nombrar un Product Owner que represente al cliente dentro del proyecto, comprendiendo sus necesidades específicas, y esté en comunicación constante con él.
- Ejecutar un Sprint Planning, para definir prioridades y tiempos específicos de cada tarea.
- Planificar e iniciar el Sprint.
- Realizar un Sprint Review y volver a ejecutar nuevos Sprints.
- Spring Retrospective

6. Cronograma.

Cronograma: Becerra Ramírez Luis Arturo

TT No: 2023-A026

Título del TT: “Asesor virtual para la recomendación de películas y series de entretenimiento con base en preferencias y opiniones del usuario.”

[illegible]

Título del TT: “Asesor virtual para la recomendación de películas y series de entretenimiento con base en preferencias y opiniones del usuario.”

[illegible]

Título del TT: “Asesor virtual para la recomendación de películas y series de entretenimiento con base en preferencias y opiniones del usuario.”

[illegible]

7. Referencias.

- [1] Instituto Federal de Telecomunicaciones. (2021). Contratación y patrones de Consumo de los Usuarios de Servicio de Telecomunicaciones Fijas Antes y Durante la Pandemia Ocasionada por la COVID-19 [Diapositivas]. Instituto Federal de Telecomunicaciones. <http://www.ift.org.mx/sites/default/files/contenidogeneral/usuarios-y-audiencias/patronesdeconsumodurantelapandemia.pdf>
- [2] "KDD: Knowledge Discovery in Databases". Minerva Data. <https://mnrv.io/kdd-platform.html> (accedido el 21 de abril de 2022).
- [3] JAYWRKR, "Guía para construir un sistema de recomendación (Parte 1)," Medium, 2019. <https://medium.com/jaywrkr-tech/gu%C3%ADa-para-construir-un-sistema-de-recomendaci%C3%B3n-parte-1-2b1a65d6eac3> (accedido el 27 de mayo del 2022).
- [4] "Asistente virtual - EcuRed," EcuRed, 2019. https://www.ecured.cu/Asistente_virtual#:~:text=es%20un%20agente%20de%20software,desde%20su%20hogar%20u%20oficina (accedido el 27 de mayo del 2022).
- [5] "TIPOS DE AGENTES DE SOFTWARE," Personales.upv.es, 2022. <http://personales.upv.es/ccarrasc/doc/2003-2004/UMDL.htm#:~:text=El%20prototipo%20de%20la%20UMDL%20consistir%C3%A1%20en%20un,de%20interfaz%2C%20agentes%20mediadores%20y%20agentes%20de%20colecci%C3%B3n> (accedido el 28 de mayo del 2022).
- [6] "DEFINICIÓN DE AGENTES INTELIGENTES Y CLASIFICACIÓN," Personales.upv.es, 2022. http://personales.upv.es/ccarrasc/doc/2003-2004/AI_Web/definicion.html#:~:text=Agentes%20de%20interfaz%3A%20apoyan%20y%20dan%20asistencia%2C%20principalmente,cabo%2C%20solo%20los%20resultados%20que%20este%20le%20proporciona (accedido el 28 de mayo del 2022).
- [7] "¿Cuáles son los servicios de streaming disponibles en México?" Expansión. <https://expansion.mx/tecnologia/2021/12/29/servicios-streaming-mexico-2021> (accedido el 21 de abril de 2022).
- [8] "Informe de Consumo de Streaming | Sherlock Communications". Sherlock Communications. <https://www.sherlockcomms.com/es/informe-streaming-latam/> (accedido el 21 de abril de 2022).
- [9] "Pago de plataformas digitales: gasto hormiga notable para Millennials - América Retail". América Retail. <https://www.america-retail.com/mexico/pago-de-plataformas-digitales-gasto-hormiga-notable-para-millennials/> (accedido el 21 de abril de 2022).
- [10] "Ejemplo Scrum: Proyecto Web". Blog - ComparaSoftware. <https://blog.comparasoftware.com/ejemplo-scrum-proyecto-web/#:~:text=Cualquier%20proyecto%20web%20es%20un%20buen%20ejemplo%20Scrum.,será%20quien%20cumple%20el%20rol%20de%20Scrum%20Master> (accedido el 21 de abril de 2022).

8. Alumnos y directores.

Becerra Ramírez Luis Arturo. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta:2019630052, Tel. 5551458251, luisramirezlabr@hotmail.com

Firma: 

Lira Dávila Alan David. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2019630152, Tel. 5583593297, alirad1500@alumno.ipn.mx

Firma: 

Ramírez Gutiérrez Enrique. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta:2019630435, Tel. 5614423845, eramirezg1504@alumno.ipn.mx

Firma: 

Ocampo Botello Fabiola. - Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Ciudad Madero, Maestría en Ciencias de la Computación por el CINVESTAV-IPN y Doctorado en Educación Internacional por la Universidad Autónoma de Tamaulipas (UAT). Profesora de la Escuela Superior de Cómputo (ESCOM) desde 1995, Tel.: 57296000 Ext.: 52083, correo institucional: focampob@ipn.mx

Firma: 

CARÁCTER: Confidencial FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública. PARTES CONFIDENCIALES: Número de boleta y teléfono.