UNIVERSIDAD PRIVADA DEL NORTE

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

"IMPLEMENTACIÓN DE AGENTE INTELIGENTE DE RECOMENDACIÓN BASADO EN FILTRADO DE CONTENIDO EN LA EXPERIENCIA DE LOS USUARIOS DE ALTERNATE EARTHS"

Tesis para optar el título profesional de:

Ingeniero de Sistemas Computacionales

Autor(es):
Br. Lezcano Menchola, Deyner Francisco
Br. Quispe Hernández, Aixa Josseline

Asesor:
Mg. Ing. Díaz Amaya, Lourdes Roxana

Trujillo – Perú
2018

APROBACIÓN DE LA TESIS

El (La) asesor(a) y los miembros del jurado evaluador asignados, **APRUEBAN** la tesis desarrollada por el (la) Bachiller **Francisco Lezcano Menchola y Aixa Quispe Hernández**, denominada:

"IMPLEMENTACIÓN DE AGENTE INTELIGENTE DE RECOMENDACIÓN BASADO EN FILTRADO DE CONTENIDO EN LA EXPERIENCIA DE LOS USUARIOS DE ALTERNATE EARTHS"

Mg. Ing. Lourdes Roxana Díaz Amaya
ASESOR

Mg. Ing. Víctor Enemesio Dávila Rodríguez
JURADO
PRESIDENTE

Ing. Luis Mauricio Gutiérrez Magán
JURADO

Mg. Ing. Rolando Javier Berrú Beltrán
JURADO

DEDICATORIA

A Dios, por darme todo lo necesario para cumplir mis metas y por enseñarme que despertar cada día es una oportunidad para hacer de este, un mundo mejor.

A mi madre, por hacer de mi la persona que soy, y por enseñarme con su ejemplo, que cualquier problema es fácil de resolver por más grande que parezca.

A mis hermanos, que con su apoyo incondicional me ayudan a seguir sin miedo hacia adelante.

Br. Francisco Lezcano Menchola

A Dios por todas sus bendiciones y su infinito amor. Todo el tiempo Dios es bueno.

A mi madre por estar conmigo siempre, a mi padre por el apoyo y su paciencia.

A mi hermano por su cariño.

A mi sobrina por ser mi alegría y el motivo de ser mejor cada día.

Br. Aixa Quispe Hernández

AGRADECIMIENTO

Dedicamos la presente investigación principalmente a Dios, por iluminarnos con su sabiduría y guiarnos por el camino correcto a través de nuestros años universitarios. A nuestras familias, con su paciencia y ejemplo, han moldeado nuestra vida en base a valores y virtudes los cuales han hecho de nosotros las personas de bien que somos.

A la Ingeniera Lourdes Díaz Amaya, por su dedicación y compromiso en el apoyo brindado para el desarrollo de la presente tesis.

Finalmente, a nuestra alma mater, la Universidad Privada del Norte, a la cual estamos inmensamente agradecidos puesto que, con su exigencia nos preparó para afrontar los obstáculos del mundo laboral.

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

Alternate Earths es una plataforma de colaboración de Comics, especializada en la creación de universos ficticios únicos poblados con personajes e historias originales, que ofrece a sus usuarios las herramientas necesarias para que exhiban su arte en un mercado especializado. Dado que el mundo del comic es muy extenso, y considerando que la plataforma no impone restricción alguna en la cantidad de información que sus usuarios puedan generar, poco después de haber sido lanzado al ambiente de producción, Alternate Earths se convirtió en un sitio que ofrecía grandes cantidades de información a sus usuarios, sin considerar si esta era relevante o no. Por ello, surge la necesidad de desarrollar una herramienta basada en tecnologías de machine learning, capaz de filtrar la información y exponerla de forma eficiente y rápida a los usuarios considerando su comportamiento dentro de la plataforma. El desarrollo de la mencionada herramienta se estimó como una investigación pre-experimental. Asimismo, para su modelado e implementación se empleó la metodología CRISP, propuesta por IBM para la construcción de proyectos de minería de datos.

Palabras clave: Machine Learning, Crisp, Alternate Earths, Comics

ABSTRACT

Alternate Earths is a comic's collaboration platform, specialized in the creation of unique fictional universes populated with characters and original stories, which offers its users the necessary tools to display their art in a specialized market. Given that the comic world is very extensive and considering that the platform does not impose any restriction on the amount of information that its users can generate, soon after being launched into the production environment, Alternate Earths became a site that offered large amounts of information to its users, regardless of whether this was relevant or not. Therefore, the need arises to develop a tool based on machine learning technologies, capable of filtering information and exposing it efficiently and quickly to users considering their behavior within the platform. The development of the tool was estimated as pre-experimental research. Likewise, for its modeling and implementation, the CRISP methodology was used, proposed by IBM for the construction of data mining projects.

Keywords: Machine Learning, Crisp, Alternate Earths, Comics

CAPÍTULO 1. INTRODUCCIÓN

1.1. Realidad problemática

Hace menos de 30 años el Internet ingresó en las sociedades modernas como medio de comunicación y soporte a procesos de información y transacciones; es una de las maravillas de la ingeniería que además brinda a quienes la utilizan, la posibilidad de llegar a una audiencia global para educarse, entretenerse y hacer negocios. Su crecimiento fomentó la creación de muchos sitios web, los cuales se enfocaron en brindar información de todo tipo. De la misma forma, la demanda por una herramienta para acceder a toda esa información marcó el inicio de los navegadores web, los cuales se presentan como un portal que abre camino a toda esa información que diariamente crece.

Los navegadores web crearon un hito en el consumo masivo de información por internet. Pero al igual que otras tecnologías tuvieron que adaptarse a la creciente demanda. Actualmente, los navegadores no son simples herramientas que filtran información a partir de una consulta dada por el usuario final, sino que también se han enfocado en estudiar el comportamiento de los usuarios, con la finalidad de ofrecer resultados en base a preferencias personales. Por otro lado, alrededor de los años 70, el interés de analizar grandes bloques de información incrementó; apareciendo así el concepto de minería datos.

La minería de datos surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos. Inicialmente, los datos son la materia prima, luego, al atribuirle el usuario algún significado especial, pasan a convertirse en información. La minería de datos se enfoca en la búsqueda de patrones, comportamientos, agrupaciones, secuencias, tendencias o asociaciones que puedan generar algún modelo que nos permita comprender mejor el dominio para ayudar en una posible toma de decisión.

Con el pasar de los años, generar conocimiento a partir de información se fue convirtiendo en una práctica muy demandada, siendo aplicada en técnicas como el de aprendizaje automático (machine Learning). El Machine Learning es una rama de la inteligencia artificial encargada de crear programas de software capaces de generalizar comportamientos a partir de los datos, basándose en técnicas y herramientas estadísticas. Así mismo, entre estos sistemas podemos destacar a los sistemas de recomendación, los cuales, a partir de un análisis de información, pueden descubrir el comportamiento de un usuario/cliente, con el fin de recomendar información con alto nivel de interés. Actualmente, los motores de recomendación se encuentran presentes en las plataformas digitales más conocidas de internet:

- YouTube: Mostrar recomendaciones personalizadas que ayudan a los usuarios a encontrar videos de alta calidad relevante para sus intereses.
- Google: Los resultados de búsqueda que esta plataforma ofrece, están ligados a temas considerados de alto interés para el usuario.

- Netflix: Con el fin de captar la atención constante de sus usuarios, la plataforma genera recomendaciones en base a los videos ya vistos previamente.

En el Perú, no ajenos a esta tendencia, diferentes sectores hacen uso de los sistemas de recomendación. Entre los ejemplos podemos destacar:

- Reducción de costos operativos en el sector minero, empresas mineras como Río Tinto usan automóviles que operan tanto de forma remota como en modo automático. Estos equipos se basan en tecnología de inteligencia artificial, la cual les permiten tomar decisiones acerca de las rutas que se emplean.
- Actualmente los bancos ofrecen productos a sus clientes según un análisis en sus comportamientos financieros. Un ejemplo de ellos lo podemos encontrar en bancos como Interbank o BCP, que ofrecen créditos pre aprobados a personas que cumplen ciertos requisitos.
- Los supermercados usan sistemas inteligentes que sugieren proximidad de ubicación entre dos o más productos a partir del comportamiento de compra de los clientes. Un ejemplo clásico de esto es ubicar los pañales cerca de las botellas de cerveza, puesto que existe una tendencia de asociar la compra de pañales con la compra de cervezas por parte de clientes masculinos.

La presente investigación se realizará sobre el sitio web Alternate Earths, la cual es una plataforma de colaboración especializada en la creación de universos ficticios únicos poblados con personajes e historias originales. En el mundo del comic, existe una gran variedad de historietas, por lo que AE trata de ofrecer a sus usuarios la mayor cantidad de ellas, con el fin de captar más audiencia. Lamentablemente, a pesar de los esfuerzos del sitio por tener una gran acogida, presenta una baja audiencia, por lo que se realizó un análisis en el cual se identificó los siguientes puntos críticos:

- El sitio Alternate Earths cuenta con una sección de búsqueda básica, la cual no considera las preferencias del usuario para mostrar.
- Se presentan múltiples secciones de información como recomendación a los usuarios basados en indicadores generales (más vistos / más populares), que muchas veces, no van acorde con los gustos de los usuarios.
- Según el análisis realizado por la plataforma Google Analytics, los usuarios presentan un reducido tiempo de permanencia en el sitio Alternate Earths

Al igual que otras páginas, Alternate Earths (AE) muestra a sus usuarios gran cantidad de información posible, por lo que un sistema capaz de predecir los gustos de los usuarios a partir de su comportamiento podrá mejorar significativamente la experiencia de navegación y aumentar la audiencia.

1.2. Formulación del problema

¿Cómo afecta un agente inteligente de recomendación basado en filtrado de contenido en la experiencia de los usuarios de Alternate Earths (AE)?

1.3. Justificación

- **Justificación Práctica**

  Debido a que el sitio Alternate Earths (AE) almacena una gran cantidad de información no filtrada, es decir redundante o no deseada para el usuario del sitio, se tiene la necesidad de implementar un agente de recomendación. Esta investigación plantea la implementación de un agente de recomendación basado en filtrado de contenido que permita mejorar la experiencia de los usuarios del sitio de Alternate Earths (AE). Esto podría aumentar el número de usuarios registrados y activos, al mismo tiempo mejoraría la experiencia del usuario en el sitio.

- **Justificación Teórica**

  Esta investigación se realiza con el propósito de aportar conocimiento sobre cómo implementar algoritmos referentes a los sistemas de recomendación, los cuales ayudan a clasificar la información de más a menos importantes y a optimizar las búsquedas en los sistemas de información en general.

- **Justificación Valorativa**

  Esta investigación servirá al sitio Alternate Earths para conocer las preferencias de sus usuarios y generar perfil de usuario, optimizar sus búsquedas, y recomendar según los gustos de sus usuarios.

- **Justificación Académica**

  Este trabajo sirve de base para estudiantes interesados en la implementación de sistemas de recomendación basado en filtrado de contenido; ya que cuenta con información importante sobre cómo implementar algoritmos y técnicas, ayudando a clasificar la información según la importancia del usuario y brindando nuevas sugerencias de interés al mismo.

1.4. Limitaciones

En el desarrollo de la investigación se presentaron las siguientes limitaciones:

- Limitada experiencia por parte de los tesistas en el campo de inteligencia artificial, por lo que será necesaria una investigación exhaustiva del tema.
- Inexistencia de técnicas de medición para obtener preferencias de usuarios instaladas en el sitio Alternate Earths; el equipo desarrollará las herramientas necesarias para tomar las mediciones correspondientes del proyecto.
- Limitado número de usuarios activos en Alternate Earths, dificultando una eficiente obtención de datos, por ello la muestra identificada será significativa con respecto

a la población con el fin de considerar a la mayor cantidad posible de usuarios en el estudio.

- El sitio Alternate Earths no guarda actualmente información del comportamiento de sus usuarios, por lo que en el presente proyecto se contempla la implementación de las funcionalidades necesarias para hacerlo.

1.5. Objetivos

1.5.1. Objetivo general

Determinar cómo afecta un agente inteligente de recomendación basado en filtrado de contenido en la experiencia de los usuarios de Alternate Earths.

1.5.2. Objetivos específicos

- Analizar la funcionalidad y eficiencia de los algoritmos implementados en el sistema de recomendación
- Determinar el porcentaje del error de precisión de las recomendaciones generadas del sistema de recomendación
- Determinar el nivel de aceptación de los usuarios hacia el sistema de Alternate Earths
- Determinar el nivel de interés de los usuarios hacia la información recomendada por el sistema de recomendación

CAPÍTULO 2. MARCO TEÓRICO

2.1. Antecedentes

(Castro Gallardo, 2018) Un Nuevo Modelo Ponderado para Sistemas de Recomendación Basados en Contenido. Jaén, España. En esta investigación se concluyó que un sistema de recomendación es una potente herramienta de personalización en dominios en los que la cantidad de productos disponibles desborda la capacidad de un usuario para evaluar cuál de ellos le puede resultar útil o interesante, estos sistemas ayudan al usuario, filtrar productos existentes en el sistema y presentarle una lista de los mismos, acorde con sus gustos o necesidad, influyendo positivamente en la experiencia del usuario con el sistema.

(De Campos, Fernández, Huete, & Rueda, 2018) Uso de Conocimiento Estructurado en un Sistema de Recomendación Basado en Contenido. Granada, España. En esta investigación se presentó un modelo de sistema de recomendación basado en contenido que usa una representación de la información de forma estructurada para mejorar las recomendaciones, y mediante la experimentación se demostró que el modelo estructurado propuesto se comporta mejor que los modelos Baseline y Naive Bayes; con el objetivo de sugerir al usuario nuevos productos en función a su similitud con el contenido (descripción) de otros productos que éste ha juzgado anteriormente.

(Nuñez, y otros, 2018) Sistema de Recomendación de Contenidos para Libros Inteligentes. Oviedo, España. En esta investigación se propuso un modelo para definir una plataforma de recomendación de contenidos, basados en el comportamiento de los usuarios, que permita implementar un sistema de recomendación que ayude a los usuarios a descubrir contenidos de su interés; basado principalmente en la retroalimentación implícita, permite recomendar contenidos sin la necesidad que lo usuarios tengan que valorar explícitamente los contenidos que le parecen interesantes. Con el desarrollo del proceso de conversión de información implícita, donde se evaluaron e implementaron distintas acciones que fueron definidas con el modelo de transformación matemática, con el fin de analizar el comportamiento de los usuarios en un entorno de libros electrónicos.

2.2. Bases teóricas

2.2.1. Inteligencia Artificial

2.2.1.1. Definición

El concepto de inteligencia artificial se refiere no solamente a la capacidad que tienen las máquinas para intentar comprender como los humanos piensan, sino que también se esfuerza en construir entidades inteligentes. La inteligencia artificial es una de las ciencias más recientes que abarca una variedad de sub-campos, que van desde áreas de propósito general, como el aprendizaje y la

percepción, a otras más específicas como demostración de teoremas matemáticos, procesamiento de lenguaje natural, diagnóstico de enfermedades, entre otras. La inteligencia artificial sintetiza y automatiza tareas intelectuales y es, por tanto, relevante para cualquier ámbito de la actividad intelectual humana. (Russel & Norving, 2010).

2.2.2. Agente Inteligente

2.2.2.1. Definición

Es una entidad de software que puede realizar tareas específicas para un usuario y posee un grado de inteligencia suficiente para ejecutar parte de sus tareas de forma autónoma y para interactuar con entorno de forma útil. (Pardo, 2018). Además, según (Russel & Norving, 2010) “Un agente inteligente es aquél que puede percibir su ambiente mediante sensores y actuar sobre ese mundo mediante efectores (o actuadores)”. Estos autores definen como meta de la Inteligencia Artificial (IA): diseñar un agente inteligente o racional que opere o actúe adecuadamente en sus ambientes. Esto significa que el agente debe hacer lo correcto de acuerdo a sus percepciones y que emprenderá la mejor acción posible en una situación dada. La racionalidad dependerá de: la secuencia de percepciones (todo lo que el agente ha percibido hasta ahora); la medida de éxito elegida; cuánto conoce el agente del ambiente en que opera y las acciones que el agente esté en condiciones de realizar.

2.2.2.2. Propiedad de Agentes Inteligente

Según (Wooldrige & Jennings, 1995) definen a un agente como: “un sistema de computación situado en algún entorno, que es capaz de una acción autónoma y flexible para alcanzar sus objetivos de diseño.” De esta definición se desprende que es un sistema de software con las siguientes propiedades fundamentales:

- Autonomía (actuar sin intervención, control): un sistema autónomo es capaz de actuar independiente, exhibiendo control sobre su estado interno.
- Habilidad Social (lenguaje de comunicación): es la capacidad de interacción con otros agentes (posibles humanos) a través de algún tipo de lenguaje de comunicación de agentes.
- Reactividad (percepción - acción): la mayoría de los entornos interesantes son dinámicos. Un sistema reactivo es aquel que mantiene una interacción continua con el entorno y responde a los cambios que se producen él, en tiempo de respuesta adecuado.

- Proactividad (dirigido a la meta, toma iniciativa): generalmente se espera que un agente haga cosas para nosotros. Un sistema proactivo es aquél que genera e intenta alcanza metas, no es dirigido solo por eventos, toma iniciativa y reconoce oportunidades.

2.2.3. Machine Learning

2.2.3.1. Definición

"Machine Learning es la programación de computadoras para optimizar un criterio de rendimiento utilizando datos de ejemplos o experiencias pasadas" (Alpydin, 2010). Esta técnica crea sistemas que aprenden automáticamente; es decir identifican patrones complejos en millones de datos y se mejoran de forma autónoma con el tiempo para generar decisiones y resultados fiables. (Gonzales, 2014)

2.2.4. Recuperación de Información

2.2.4.1. Definición

Es la actividad de encontrar material importante a partir de fuentes no estructuradas (normalmente texto) que satisface la necesidad de información de grandes colecciones (generalmente almacenadas en sistemas informáticos). La recuperación de información se está convirtiendo rápidamente en la forma dominante de acceso a la información, superando la búsqueda tradicional de bases de datos. (Manning, Raghavan, & Schütze, 2008)

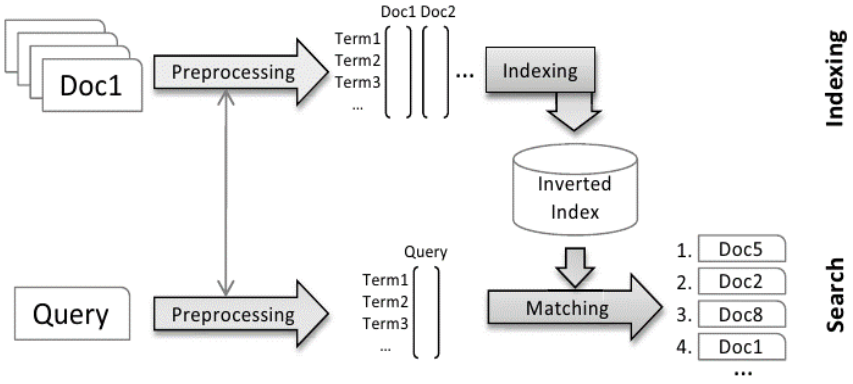


Figura 1. Ciclo de Recuperación de Información

2.2.4.2. Modelo de Espacio Vectorial

La representación de un conjunto de documentos como vectores en un espacio vectorial común se conoce como modelo de espacio vectorial y es fundamental para una serie de operaciones de recuperación de información que van desde la puntuación de documentos en una consulta, hasta su clasificación y agrupamiento. (Manning, Raghavan, & Schütze, 2008)

El modelo de espacio vectorial se basa en el grado de similaridad de una consulta dada por el usuario con respecto a los documentos de la colección cuyos términos fueron ponderados mediante técnicas como TF-IDF. El modelo vectorial fue presentado por (Salton, Wong, & Yang, 1975) y posteriormente asentado en (Salton & McGill, 1983) junto con Mc Gill y se basa en tres principios esenciales (Martínez, 2006)

- La equiparación parcial, esto es, la capacidad del sistema para ordenar los resultados de una búsqueda, basado en el grado de similaridad entre cada documento de la colección y la consulta.
- La ponderación de los términos en los documentos, no limitándose a señalar la presencia o ausencia de los mismos, sino adscribiendo a cada término en cada documento un número real que refleje su importancia en el documento.
- La ponderación de los términos en la consulta, de manera que el usuario puede asignar pesos a los términos de la consulta que reflejen la importancia de los mismos en relación a su necesidad informativa.

  Si bien en el modelo booleano un documento de la colección puede ser representado por la presencia o ausencia de los términos indexados en el fichero diccionario de la siguiente forma:

Tabla 1. Ejemplo de Vector Binario

	Código	Batman	Marvel	Gótica	Superman	Spiderman
Batman vs Superman	1	1	1	1	1	0

En el modelo de espacio vectorial se emplea el peso de los términos para cada documento, que refleja la relevancia de los términos del documento de cara a su representatividad en la colección, adquiriendo una forma como la que sigue (nótese que se reemplazó el uso de la representación binaria por la frecuencia del término, lo cual nos permite priorizar una palabra sobre otra).

Tabla 2. Ejemplo de Vector de Pesos TF-IDF

	Código	Batman		Marvel		Gótica		Superman	
Batman vs Superman	1	**TF**	**IDF**	**TF**	**IDF**	**TF**	**IDF**	**TF**	**IDF**
		1	0.098	6	0.428	5	0.812	9	0.6598

De esta forma, se puede aterrizar el vector descrito de la siguiente forma:

Ecuación 1. Representación del vector de un documento

$$Doc(d) = \left(P_{(1,d)}, P_{(2,d)}, P_{(3,d)}, P_{(n,d)}\right)$$

- $(\boldsymbol{d})$ Identificador del documento, documento 1, documento2
- $(\boldsymbol{n})$ Identificador del término por su posición en el fichero diccionario; término 1; término2, término 3

2.2.4.3. Proceso de Equiparación Mediante la Fórmula del Coseno

Tal como se ha explicado en la fórmula del producto escalar, el proceso de equiparación es posible cuando en el vector de la consulta y en el del documento existen términos coincidentes. Pero este enfoque no supone la representación del vector de la consulta y del documento. De hecho, una de las claves del modelo de espacio vectorial es precisamente la posibilidad de determinar el ángulo que forman los vectores del documento y de la consulta que se está comparando.

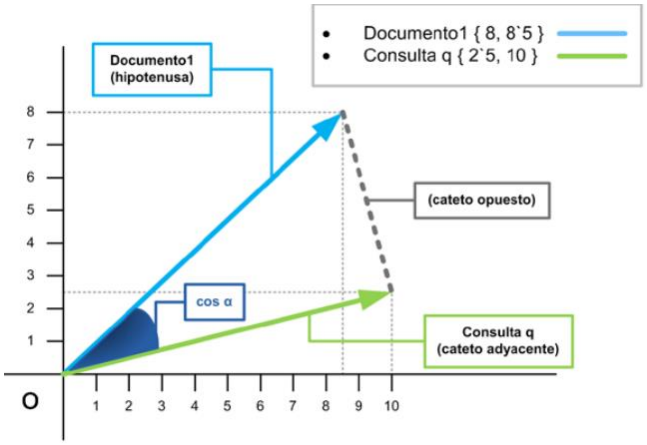


Figura 2. Similitud del coseno

Es posible medir cuál es la desviación de un documento con respecto a una consulta, por el número de grados del ángulo que forman. Esto es posible porque crean una estructura triangular a la que se aplica el cálculo del ángulo que forma la hipotenusa (en este caso el vector del documento1) y el adyacente (el vector q de la consulta dada por el usuario) que resulta ser el coseno del triángulo. En el caso de la Ilustración 2, se comprueba visualmente, cierta distancia del vector de la consulta con respecto al documento1; cuando ambos vectores se muestran tan próximos como para superponerse, implicará que el ángulo que forman será menor y que su nivel de coincidencia será superior. De hecho, un coseno de 0º implicaría una similaridad máxima.

$$SimCos(d_{(d)},q) = \frac{\sum_{n=1} \left(P_{(n,d)} \times P_{(n,q)} \right)}{\sqrt{\sum_{n=1} \left(P_{(n,d)} \right)^2 \times \sum_{n=1} \left(P_{(n,q)} \right)^2}}$$

Figura 3. Fórmula para el cálculo de la similaridad del coseno

2.2.5. Sistema de recomendación

2.2.5.1. Definición

Los sistemas de recomendación son sistemas que acceden a la información mediante la técnica de filtrado presentando un resultado personalizado al usuario. El proceso que siguen consiste en dos etapas: creación de perfil de usuario, donde se recopila información sobre él y la recomendación de ítems atendiendo a las preferencias del usuario. (Jannach, Zanker, Felfering, & Friedrich, 2010)

2.2.5.2. Basado en Filtrado Colaborativo

El filtrado colaborativo es una técnica empleada por los Sistemas de Recomendación que utiliza la información de preferencias y calificación de un grupo de usuarios respecto a los ítems de un repositorio con el fin de predecir o inferir la preferencia de un usuario en particular sobre un ítem y a partir de esto generar una recomendación acertada.

(Alfredo, Víctor, Manuel, & Cristobal, 2011) El filtrado colaborativo presenta características y problemas específicos tales como la escasez de datos referentes a calificaciones y preferencias lo que afecta el rendimiento del sistema siendo el problema más común el conocido como arranque en frío que se presenta con el ingreso de un nuevo usuario o ítem para el cual no existe suficiente información previa.

Las técnicas de filtrado colaborativas pueden ser agrupadas en tres categorías (Alfredo, Víctor, Manuel, & Cristobal, 2011): los algoritmos basados en memoria, que usan una base de datos usuario-ítem para generar una predicción y se basa en que cada usuario es parte de un grupo con intereses similares, al identificar los vecinos similares de un nuevo usuario se pueden producir predicciones de preferencias de nuevos ítems, la similaridad entre usuarios puede calcularse por ejemplo con los algoritmos de correlación de Pearson, Spearman Rank, Kendall o vector coseno. La segunda categoría son los algoritmos de filtrado basados en modelo, con los que se intenta que el sistema aprenda a reconocer patrones basados en datos de entrenamiento, entre los algoritmos para esta categoría están los modelos bayesianos, de

clustering y basados en regresiones. La tercera categoría es el filtrado híbrido, que combina el filtrado colaborativo con otras técnicas de recomendación.

2.2.5.3. Basado en Filtrado de Contenido

Los sistemas basados en contenido buscan utilizar la información disponible tanto del ítem como del usuario para calcular la recomendación más adecuada. Pazzani y Billsus, en su artículo “Content-Based Recommendation Systems” (Michael & Daniel, 2018), tratan los sistemas que presentan la recomendación de un ítem basada en una descripción del mismo y un perfil de los intereses del usuario, los mismos que deben tratar y definir 3 aspectos importantes que son el modelamiento del ítem, el modelamiento de las preferencias del usuario y el método o algoritmo para realizar la recomendación.

Los sistemas que implementan un enfoque de recomendación basado en contenido analizan un conjunto de documentos y/o descripciones de elementos previamente calificados por un usuario, y construyen un modelo o perfil de los intereses del usuario en función de las características de los objetos clasificados por ese usuario. (Mlandenic, 2018)

El perfil es una representación estructurada de los intereses del usuario, adoptado para recomendar nuevos artículos interesantes, el proceso de recomendación consiste básicamente en hacer coincidir los atributos del perfil de usuario contra los atributos de un objeto de contenido. El resultado es un juicio de relevancia que representa el nivel de interés del usuario en este objeto; si un perfil refleja con precisión las preferencias del usuario, es una gran ventaja para la efectividad de un proceso de acceso a la información. Por ejemplo, podría usarse para filtrar los resultados de búsqueda al decidir si un usuario está interesado en una página web específica o no y, en el caso negativo, evitando que se muestre.

2.2.5.4. Arquitectura de Alto Nivel de Sistemas Basados en Contenido

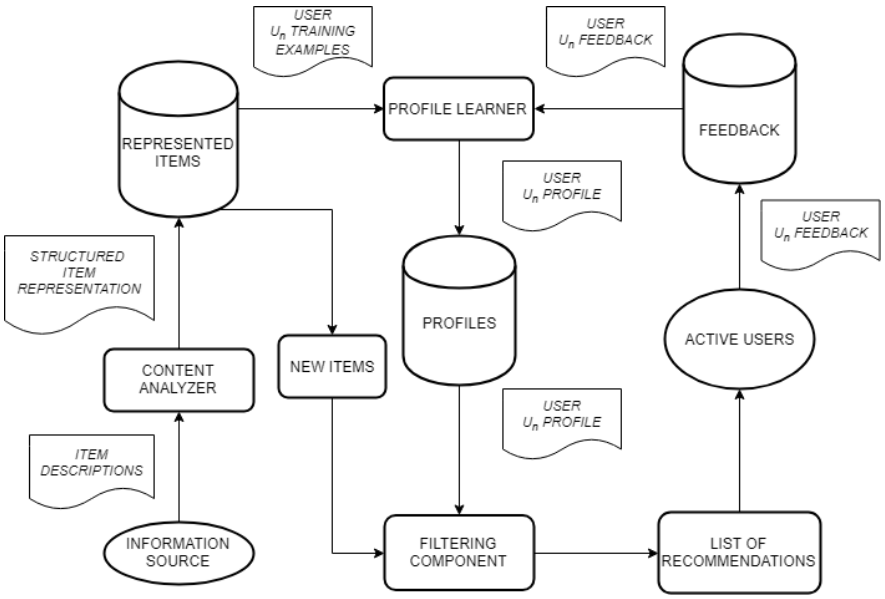


Figura 4. Arquitectura de Alto Nivel de Sistema de Recomendación Basado en Contenido

Los sistemas de filtrado de información 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. En la figura (agregar número) describe la arquitectura de alto nivel de un sistema de recomendación basado en contenido

CONTENT ANALYZER (ANALIZADOR DE CONTENIDO): cuando la información no tiene estructura (Ejemplo, texto) se necesita un tipo de paso de 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.) procedente de fuentes de información en una forma adecuada para los próximos pasos de procesamientos. Los ítems de la información son analizados por técnicas de extracción de características para cambiar la representación del ítem desde el espacio original hasta el objetivo (Por ejemplo, página web representada) como vectores de palabras claves, esta representación es la entrada al PROFILE LEARNER (Perfil de aprendizaje) y FILTERING COMPONENT (Componente de filtro)

PROFILE LEARNER (PERFIL DE APRENDIZAJE): este módulo recopila datos representativos de las preferencias del usuario e intenta generalizar estos datos para construir el perfil del usuario. Por lo general, la estrategia de

la generalización se realiza a través de las técnicas de aprendizaje automático, que son capaces de inferir un modelo de intereses del usuario a partir de elementos que le gusten o no le gustaron en el pasado. Por ejemplo, el perfil de aprendizaje de un recomendador de página web puede implementar un método de retroalimentación (Salton & Buckey) de relevancia en el cual la técnica de aprendizaje combina vectores positivos y negativos en un prototipo de vector que representa el perfil del usuario. Los ejemplos de entrenamiento son páginas web en las que un feedback positivo o negativo ha sido proporcionado por el usuario.

FILTERING COMPONENT (COMPONENTE DE FILTRADO): 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 recomendar. El resultado es un juicio de relevancia binario o continuo (cálculo usando medida de similitud), el resultado final es 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 y los elementos del vector.

2.2.6. Técnicas Avanzadas de Recuperación de Información

2.2.6.1. Definición

Son todos aquellos procesos destinado a la recuperación de información, desde la generación de las colecciones, su depuración, indexado, tratamiento textual, clasificación, almacenamiento, recuperación mediante modelos booleanos, vectoriales, probabilísticos, basados en el lenguaje, así como todos aquellos elementos que inciden en cualquier aspecto relacionado como por ejemplo el interfaz del consulta, el comportamiento del usuario, la retroalimentación de las consultas y la representación de la información. (Blázquez, 2018)

2.2.6.2. Mecanismo de Depuración para la Extracción y Procesamiento de Textos

Cuando se generan colecciones de documentos basados en páginas web y se efectúa un proceso de extracción del código fuente, mediante el empleo de técnicas cURL y DOM. Al hacerlo no sólo se adquiere el texto sujeto a recuperación, sino todo el conjunto de etiquetas en formato HTML y CSS que lo acompañan. Si tales etiquetas no son eliminadas, no se puede iniciar el procesamiento de la información y su correspondiente tratamiento. Por ello, se demuestra la importancia de aplicar mecanismos de depuración del código fuente, que facilite la extracción limpia de los textos, que serán la materia prima con la que se componen las colecciones sobre las que se recupera la

información. Para comprender los procesos que se llevan a cabo, se muestra la siguiente (**Tabla N° 03**) en la que puede ver un orden en la consecución de los mismos.

Tabla 3. Mecanismo de Depuración para la Extracción y Procesamiento de Textos

Preparación de la colección de Documentos
• Decodificación de cadena de caracteres • Extracción de texto del HTML
Normalización de Textos
• Tokenización • Conversión a minúsculas, eliminación de signos de puntuación y caracteres especiales • Eliminación de palabras vacías
Indexación
• Reducción morfológica • Almacenamiento en Mongo DB

TOKENIZACIÓN: es el proceso que descompone los textos de una colección en sus unidades mínimas, las palabras o términos propiamente dichos. A tales, elementos se les denomina “Tokens” que conforman una lista de ítems que se utiliza para su análisis PNL (Procesamiento del Lenguaje Natural). Para llevar a cabo tal proceso se utiliza los espacios entre las palabras del texto como divisores de los distintos “Tokens”.

Tabla 4. Ejemplo de Tokenización

Ejemplo de Tokenización			
Institución cuya finalidad consiste en la adquisición, conservación, estudio y exposición de libros y documentos. Local donde se tiene considerable número de libros ordenados para la lectura.			
Token Resultantes			
Institución	conservación,	Documentos.	de
cuya	estudio	Local	libros
finalidad	y	donde	ordenados
consiste	exposición	se	para
en	de	tiene	la
la	libros	considerable	Lectura.
adquisición	y	número	

CONVERSIÓN A MINUSCULAS, ELIMINACIÓN DE SIGNOS PUNTUACIÓN Y CARACTERES ESPECIALES: para permitir un proceso de indexación limpio, previamente es necesario convertir el texto a minúsculas y eliminar todos los signos de puntuación del texto. De hecho, este punto de la depuración

simplifica a posteriori el reconocimiento y proceso de cruzado de la consulta del usuario, permitiendo que independientemente de que escribiera su consulta correctamente, pueda ser recuperada bajo cualquier circunstancia.

ELIMINACIÓN DE PALABRAS VACÍAS: Las palabras vacías, irrelevantes o "stop words" son aquellas que por sí solas carecen de significación y que, por su altísima frecuencia de aparición en los textos, generan un ruido innecesario para la recuperación de información. La eliminación de estos términos (preposiciones, artículos determinados, artículos indeterminados, pronombres, conjunciones, contracciones y ciertos verbos y adverbios) mejora la afinación en los modelos de recuperación. Los estudios correspondientes a este fenómeno fueron iniciados por Hans Peter Luhn en 1958 con su investigación sobre el índice KWIC, una técnica de indexación que organizaba las palabras según su consideración como claves para la recuperación o no de la información, teniendo en cuenta el contexto del documento. Este proceso derivó en la acuñación del término "palabra vacía" para referirse a aquellas con un bajo poder discriminatorio y representativo del contenido del documento. Los análisis estadísticos efectuados por Luhn, demostraron que la indexación era un proceso más rápido, cuando se prescinde de tales términos y favoreciendo la economía de espacio requerido para el almacenamiento de la información. También se demostró, que entre un 30 y un 50% de las palabras de un texto corresponden a tal categoría. De hecho y pese a ser práctica habitual, hasta nuestros días, se siguen utilizando listas de palabras vacías para la depuración de los textos. No obstante, la técnica de eliminación de palabras vacías, se viene suavizando, debido a la introducción de técnicas de PNL (Procesamiento del Lenguaje Natural) que tienen en cuenta la significación de tales palabras cuando están acompañadas de sustantivos, en casos en los que no pueden ser separadas o eliminadas por conformar una denominación propia, así como por pérdidas en la significación semántica de un sintagma, frase o palabra.

Tabla 5. Ejemplo de Palabras Vacías

Palabras Vacías, Claves		
Frase	**Palabras Vacías**	**Descripción del caso**
Those were the days	Those, were, the	La frase corresponde al título de una canción de Boris Fomin
Es así o de esta otra manera	es, así, o, de, esta, otra	Título de un artículo sobre estilo gramatical

De aquí a la eternidad	de, aquí, a, la	Película de 1953 del director Fred Zinnemann
Cómo ha de ser el privado	cómo, ha, de, ser, el	Comedia teatral de Francisco de Quevedo

2.2.6.3. El proceso de Indexación

Indexar es la acción de construir un fichero inverso de forma automática o manual. Este proceso es necesario para localizar y recuperar rápidamente cada uno de los términos del texto de un documento. Esto significa que a cada palabra se le asigna un identificador del documento en el que aparece, un indicador de la posición que ocupa en el texto (párrafo, línea, número de carácter de inicio) y un número de identificación para ese término propiamente dicho (único e irrepetible). De esta forma se conoce la posición exacta de cada término en los documentos de la colección y posibilita el posterior análisis de frecuencias. Con el objetivo de reducir al máximo el tamaño de los archivo o tablas de la base de datos para conseguir la mejor relación entre tiempo de ejecución de las consultas y exhaustividad del fichero inverso. A esta misión se le denomina “Compresión de la Indexación” y en ella se circunscriben los procesos de depuración, tales como la supresión de palabras vacías, la normalización de palabras, la transliteración de caracteres especiales y la reducción morfológica o “Stemming”.

Tabla 6. Proceso de la Indexación

La compresión de la Indexación

- **Depuración de los textos de los documentos de la colección.**
- **Supresión de palabras (primer filtro)**
- **Reducción morfológica**
  - **Stemming**
  - **Lematización**
- **Supresión de palabras vacías (segundo filtro)**
- **La ley de Zipf y la frecuencia de aparición**
- **Técnicas de cortes Luhn: Cut – on y Cut – Off**
- **Cálculo del punto de transición**

REDUCCIÓN MORFOLÓGICA: es el proceso por el cual se depuran todos los términos de un texto, reduciendo su número de caracteres, simplificando su forma original, género, número, desinencia, prefijo, sufijo en su forma de

palabra más común o normalizada, debido a que la mayor parte de ellas tienen la misma significancia semántica. Este proceso reduce el tamaño de los términos, del diccionario y mejorar el “Recall” o exhaustividad de los resultados en la recuperación de información.

STEMMING: Efectúa una reducción de palabras a sus elementos mínimos con significados, las raíces de las palabras, de hecho "Stem" significa tallo o raíz. De esta forma, los procesos de Stemming, acotan las terminaciones de las palabras a su forma más genérica o común. **(Ver Tabla N° 07)**

Tabla 7. Ejemplo cásico de Stemming

Término	Stem
che	che
checa	chec
checar	chec
checo	chec
console	consol
consoles	consol
consolidated	consolid
consolidating	consolid
consoling	consol
consolingly	consol
conspirator	conspir
conspirators	conspir
conspire	conspir
conspired	conspir

Como se ven el análisis de la **Tabla N° 07**, tanto en inglés como en español y cualquier idioma, un término puede ser reducido a su común denominador, permitiendo la recuperación de todos los documentos cuyas palabras tengan la misma raíz común, por ejemplo (catálogo, catálogos, catalogación, catalogar, catalogado, catalogando, catalogándonos). Todos los términos derivan en tal caso de “catalog”, haciendo posible que la recuperación sea completa en más de 8 supuestos distintos. No obstante, no siempre esta técnica permite funcionar perfectamente todas las consultas que un usuario pueda plantear, es el caso de eliminar prefijos y sufijos cuya raíz puede ser compartida por múltiples palabras. **(Ver Tabla N° 8)**

Tabla 8. Ejemplo de conflictos de los procesos de stemming

Término con Prefijo	Raíz / Stem	Término con el que causaría confusión
Prevalencia	valenc	Valencia, valencia, valenciano, ambivalencia, polivalencia,
Precatalogar	catalog	Descatalogar, catalogo,

2.2.7. Lógica difusa

2.2.7.1. Definición

Es una lógica multivariada que permite representar matemáticamente la incertidumbre y la vaguedad, proporcionando herramientas formales para su tratamiento. (Fakhfakh, Ammar, & Amar, 2014).

2.2.7.2. Base Teórica

Conjuntos Borrosos

Según (Ying, 2000), los conjuntos borrosos son una extensión de los clásicos, donde se añade el concepto de conjunto o subconjunto borroso y se le asocia un determinado valor lingüístico, definido por una palabra o etiqueta lingüística, donde ésta es el nombre del conjunto o subconjunto. Por cada conjunto se define una función de pertenencia o membresía denominada $u_{A(x)}$, indica el grado en que la variable x está incluida en el concepto representado por la etiqueta A $(0 \leq u_{A(x)} \leq 1)$, si esta función toma el valor 0 significa que tal valor x no está incluido en A y si toma el valor 1 el correspondiente valor de x está absolutamente incluido en A. En la Figura 5 se puede apreciar un ejemplo donde el conjunto velocidad (con variable x) está subdividido en 3 subconjuntos $\{Baja, Media, Alta\}$, con sus respectivas funciones de membresía $\{u_{Baja(x)}\ u_{Media(x)}\ u_{Alta(x)}\}$

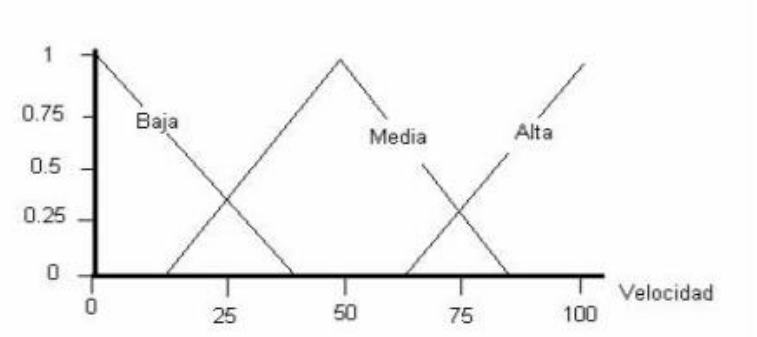


Figura 5. Ejemplo de subconjuntos borrosos

Inferencia de Mamdani

Según (Gonzáles, 2018), la inferencia difusa puede definirse como el proceso de obtener un valor de salida para un valor de entrada empleando la teoría de conjuntos difusos. La inferencia de Mamdani es el método más ampliamente utilizado, propuesto por Ebrahim Mamdani en 1975, el proceso se realiza en cuatro pasos:

- **Fuzificación de las variables de entrada**

  El primer consiste en tomar los valores crisp de las entradas y determinar el grado de pertenencia de estas entradas a los conjuntos difusos asociados. El valor crisp naturalmente estará limitado en el universo del discurso de la variable.

- **Evaluación de reglas**

  Se toma las entradas del paso anterior, y se aplican a los antecedentes de las reglas difusas. Si una regla tiene múltiples antecedentes, se utiliza el operador AND u OR para obtener un único número que represente el resultado de la evaluación. Este número (el valor de verdad) se aplica al consecuente.

- **Agregación de las saldas de las reglas**

  La agregación es el proceso de unificación de las salidas de todas las reglas; es decir, se combinan las funciones de pertenencia de todos los consecuentes previamente recortados o escalados, combinando para obtener un único conjunto difuso por cada variable de salida.

- **Defuzificación**

  El resultado final habitualmente es necesario expresarlo mediante un valor crisp. En esta etapa se toma como entrada el conjunto difuso anteriormente obtenido para dar un valor de salida. Existen varios métodos de defuzificación, pero probablemente el más ampliamente usado es el centroide.

2.2.8. Seguridad

2.2.8.1. Definición HMAC

Según (Bellare, Canetti, & Krawczyk, 1996) HMAC es un código de autenticación de mensaje basado en funciones hash criptográficas. La autenticación HMAC es un mecanismo para calcular un código de autenticación de mensaje utilizando una función HASH en combinación con una clave secreta compartida entre las dos partes involucradas en el envío y recepción de datos (FRONT END – BACK END – HTTP SERVICE). El uso principal del HMAC es verificar la integridad, autenticidad e identidad del remitente del mensaje.

2.2.8.2. Funcionamiento de HMAC

Las siguientes definiciones se usan a lo largo del funcionamiento de HMAC como un estándar aprobado por: FIPS (Federal Information Processing Standard) | NIST (National Institute of Standards and Technology)

- **Clave criptográfica (Key):** parámetro utilizado junto con un algoritmo criptográfico que determina la operación específica de ese algoritmo. En este estándar, el algoritmo HMAC utiliza la clave criptográfica para producir un MAC en los datos.
- **Función HASH:** es una función matemática aprobada que mapea una cadena de longitud arbitraria (hasta un tamaño máximo predeterminado) a una cadena de longitud fija; puede ser usada para producir una suma de comprobación, llamada valor de hash o resumen de mensaje, para una cadena potencialmente larga o mensaje.
- **Código de autenticación de mensaje basado en Key-Hash cifrado (HMAC):** es un código de autenticación de mensajes que usa clave de criptográfica junto con una función hash.
- **Código de autenticación de mensaje (MAC):** es una suma de comprobación criptográfica que resulta de pasar datos a través de un algoritmo de autenticación de mensaje llamado HMAC mientras que el resultado de aplicar HMAC se llama MAC.
- **Clave secreta:** es una clave criptográfica que está asociada de manera única con una o más entidades. El uso del término “secreto” en este contexto no implica un nivel de clasificación, más bien, el término implica la necesidad de proteger la clave de la divulgación o sustitución.

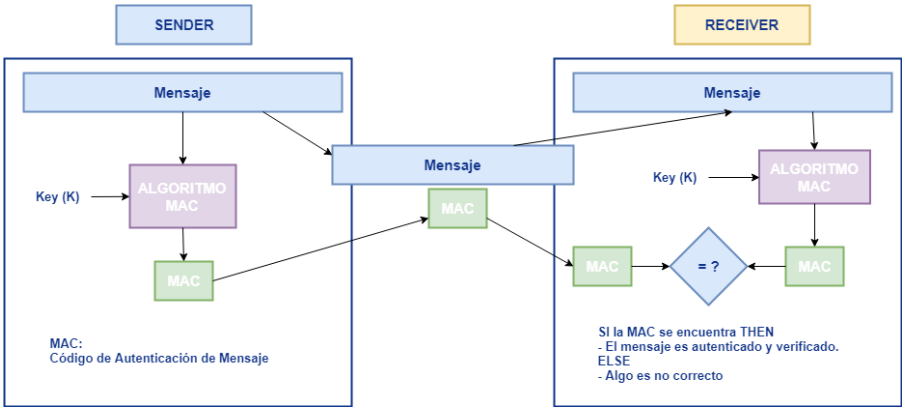


Figura 6. Funcionamiento de HMAC

2.2.9. Metodología CRISP

2.2.9.1. Definición

Según el manual CRISP de (IBM, 2012) Cross- Industry Standard Process for Data Mining es un método probado para orientar trabajos de minería de datos y machine learning. Es flexible y se puede personalizar fácilmente. CRISP incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene por qué ser ordenada desde la primera hasta la última.

2.2.9.2. Fases

Este modelo contiene 6 fases con fechas que indican las dependencias más importantes y frecuentes entre fases, además de mostrar dependencias bidireccionales.

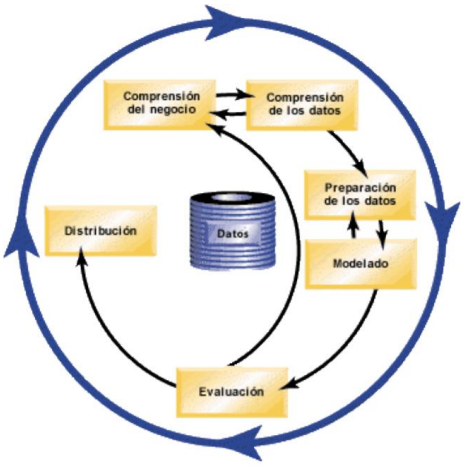


Figura 7. Secuencia del Proceso CRISP

Fuente: IBM Manual CRISP de IBM SPSS Modeler

a) Comprensión del negocio

En esta primera fase es probablemente la más importante y se realizan las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva de negocio, con el fin de convertirlos en objetivos técnicos.

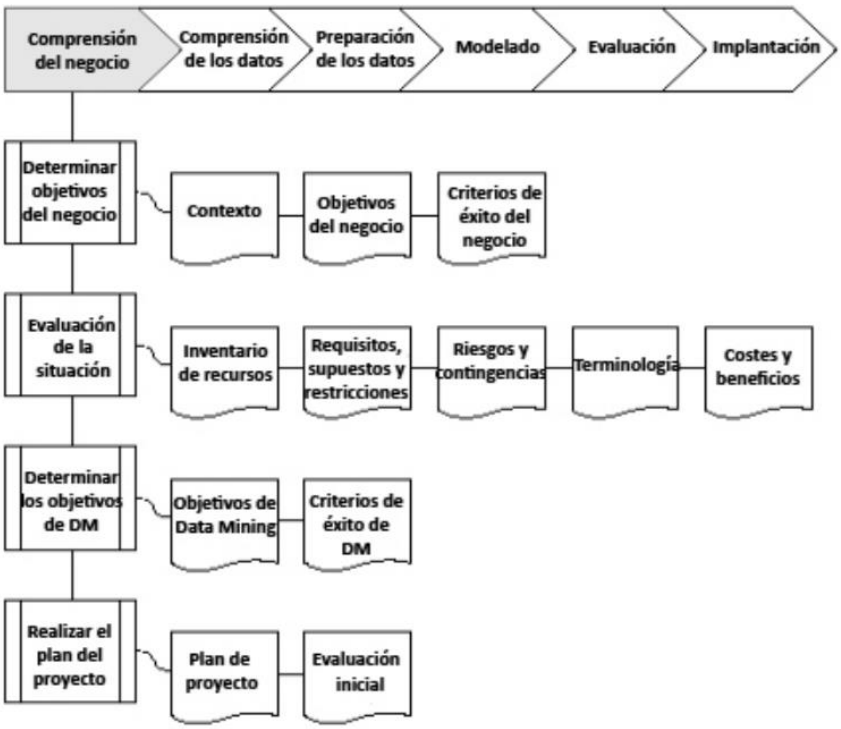


Figura 8. Fase de comprensión del negocio

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

A continuación, se describen las tareas que componen esta fase:

- Determinar objetivos del negocio: según (IBM, 2012) en esta tarea se tiene que determinar cuál es el problema que se desea resolver, con el fin de obtener la máxima información posible de los objetivos comerciales.
- Evaluación de la situación: según (IBM, 2012), en esta tarea se debe calificar los objetivos comerciales, por lo cual es importante realizar la evaluación de la situación actual.
- Determinar los objetivos: según (IBM, 2012), después que el objetivo comercial ha quedado claro, se deberá documentar los objetivos técnicos y proporcionar datos reales para resultados deseados.
- Realizar el plan de proyecto: después de tener los objetivos claros, esto permitirá elaborar un plan de proyecto que permitirá informar a todos los usuarios relacionados con los objetivos, recursos, riesgos del proyecto.

b) Comprensión de los datos

En esta segunda fase de la metodología CRISP se realiza la recolección inicial de los datos para poder establecer un primer contacto con el problema, familiarizarse con los datos y averiguar su calidad.

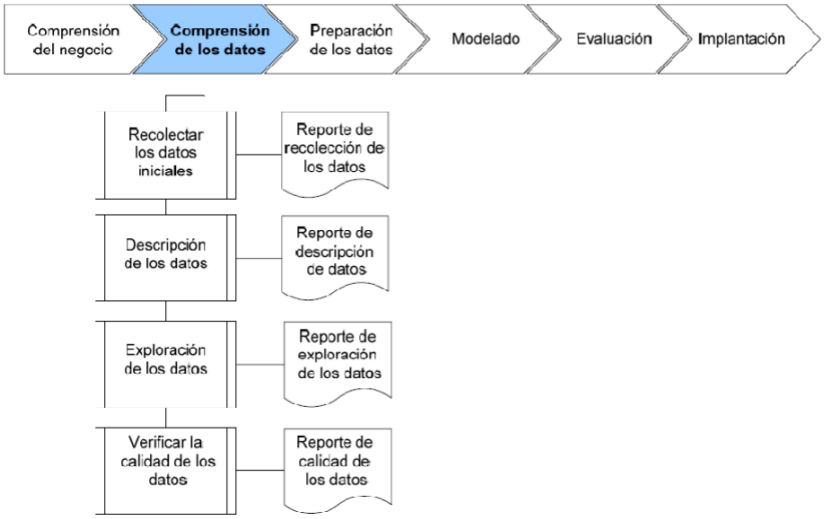


Figura 9. Fase de comprensión de los datos

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

Las principales tareas que engloba son las siguientes:

- **Capturar datos iniciales:**

  Hacerse con los datos (o, primeramente, con la posibilidad de acceder a los mismos) que se han identificado dentro de los recursos clave del proyecto.

- **Describir los datos:**

  Realizar una caracterización general de los datos obtenidos: su formato, cantidad (número de registros y campos) y cualquier otra característica descubierta en este primer vistazo general. Esta caracterización debe servir para evaluar si los datos obtenidos satisfacen los requerimientos relevantes.

- **Explorar los datos:**

  Realizar un análisis preliminar de los datos utilizando diferentes herramientas de consulta, visualización y elaboración de informes. En esta exploración nos deberíamos fijar en la distribución de los atributos clave, en las relaciones entre subconjuntos pequeños de los atributos o en las propiedades de determinadas “subpoblaciones” dentro del total de los datos.

- **Verificar la calidad de los datos:**

  En este examen de la calidad de los datos deberíamos fijarnos en cuestiones como las siguientes: si están completos los datos (cubren todos los casos que se requieren), si son correctos, cómo de frecuentes son los errores, si hay missing values (cómo se representan, donde y con qué frecuencia ocurren).

c) Preparación de los datos

En esta fase de la metodología se trata de preparar los datos para adecuarlos a las técnicas que se van a emplear sobre ellos. Esto implica seleccionar el subconjunto de datos que se van a utilizar, limpiarlos para mejorar su calidad, añadir nuevos datos a partir de los existentes y darles el formato requerido.

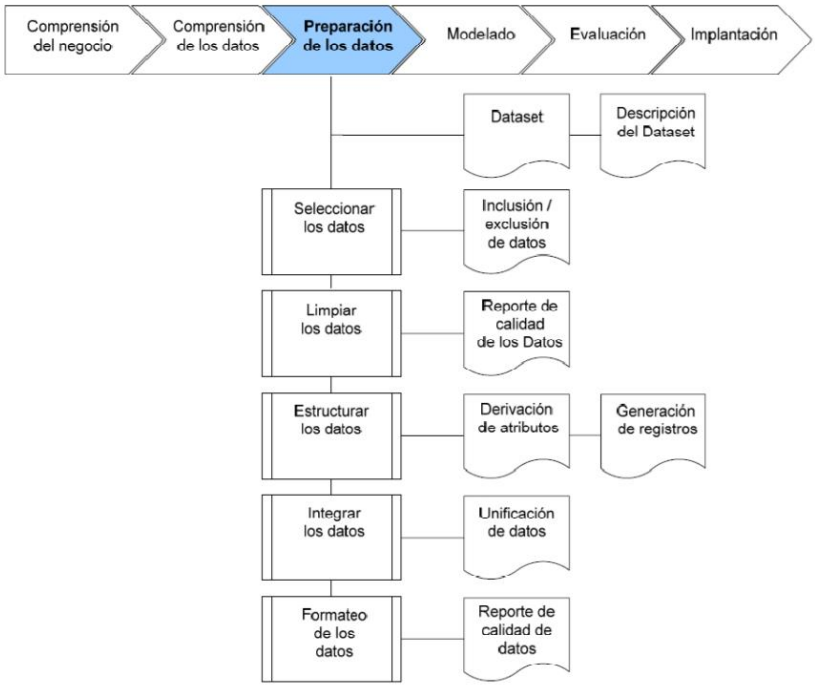


Figura 10. Fase de preparación de datos

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

Las principales tareas que engloba son las siguientes:

- **Selección de datos:**

  Decisión sobre los datos a emplear en el análisis, usando criterios relativos a la relevancia para los objetivos, la calidad de los datos o restricciones técnicas. La selección a realizar se refiere tanto a los

atributos o campos de los registros.

Limpieza de datos:

Se debe "elevar" el nivel de calidad de los datos al requerido por las técnicas de análisis. Esta tarea incluye la inserción de valores por defecto adecuados, o el uso de modelado para estimar los valores ausentes (missing values). Se deben documentar las decisiones y acciones para resolver los problemas de calidad de datos que ya fueron identificados en la fase anterior.

Construcción de datos:

A partir de los datos originalmente capturados, se generan atributos derivados, nuevos registros o valores transformados de atributos existentes, en función de los requerimientos para preparar la entrada a las herramientas de modelado.

Integración de datos

Esta tarea se enfoca a la combinación de múltiples tablas o registros para crear nuevos, uniendo por ejemplo datos sobre un mismo objeto pero que se encuentran dispersos en diferentes fuentes, o realizando agregaciones que resumen información contenida en varios registros.

Dar formato a datos

Estas transformaciones se refieren a modificaciones sintácticas que se hacen sobre los datos, sin alterar su significado pero que pueden ser requeridas por la herramienta de modelado a utilizar. Por ejemplo, puede que haya requisitos en el orden de los atributos, o que la herramienta de modelado requiera que los registros estén ordenados según el atributo resultado. En otros casos es necesario presentarlos en un orden más aleatorio del que vienen inicialmente.

d) Modelado

En esta fase se seleccionan y aplican diferentes técnicas (algoritmos) de modelado, calibrando sus parámetros para conseguir sus valores óptimos. Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada al problema.
- Disponer de datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica.

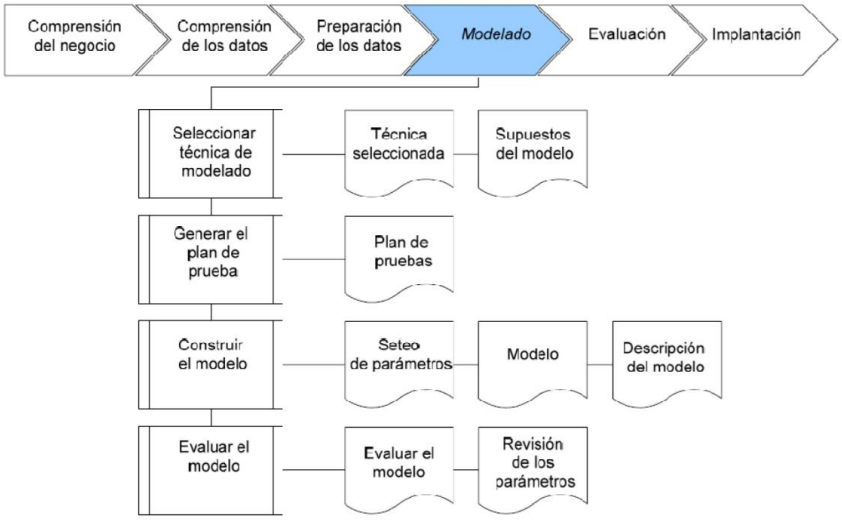


Figura 11. Fase de modelado

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

e) Evaluación

En esta etapa, formalizará su evaluación en función de si los resultados del proyecto cumplen los criterios del rendimiento comercial. El objetivo final de la fase es decidir la aprobación o no del uso de los resultados del análisis de datos.

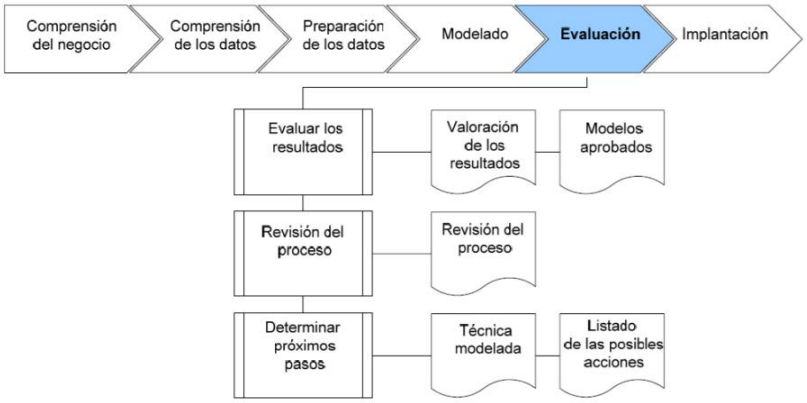


Figura 12. Fase de Evaluación

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

f) Implantación o Distribución

Según (IBM, 2012) indica que este proceso consiste en utilizar sus nuevos conocimientos para implementar las mejoras en su organización. Además, la distribución puede significar que utilice los conocimientos adquiridos para aplicar modificaciones en su organización.

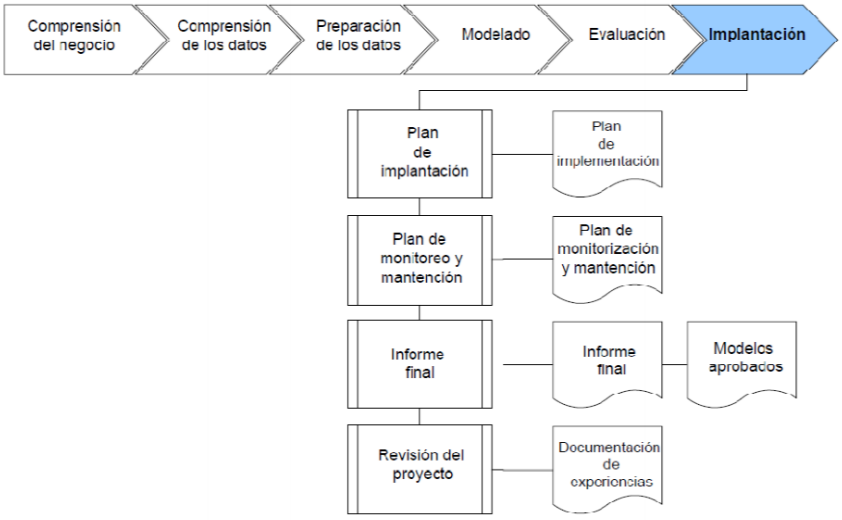


Figura 13. Fase de Implantación

Fuente: Aplicación de la metodología CRISP – Galán Cotrina

La fase de distribución de CRISP-DM incluye dos tipos de actividades:

- Planificación y control de la distribución de los resultados.
- Finalización de tareas de presentación como la producción de un informe final y la revisión de un proyecto

2.2.10. Contexto Tecnológico

2.2.10.1. Lenguaje de programación

C Sharp

C# es un lenguaje elegante, con seguridad de tipos y orientado a objetos, que permite a los desarrolladores crear una gran variedad de aplicaciones seguras y sólidas que se ejecutan en .NET Framework .NET. Puede usar C# para crear aplicaciones cliente de Windows, servicios web XML, componentes distribuidos, aplicaciones cliente-servidor, aplicaciones de base de datos y muchas, muchas más cosas. Visual C# proporciona un editor de código avanzado, prácticos diseñadores de interfaz de usuario, un depurador integrado y muchas otras herramientas que facilitan el desarrollo de aplicaciones basadas en el lenguaje C# y .NET Framework. (Microsoft, 2015)

2.2.10.2. Framework

ASP.NET WEB API

Es un nuevo framework de la familia .NET que contiene como objetivo el facilitarnos en gran medida la construcción de aplicaciones RESTful orientadas a ofrecer servicios. ASP.NET Web API es un marco que facilita la creación de servicios HTTP disponibles para una amplia variedad de clientes, entre los que se incluyen exploradores y dispositivos móviles. ASP.NET Web

API es la plataforma perfecta para crear aplicaciones RESTful en .NET framework.

2.2.10.3. Entorno de desarrollo

Visual Studio 2017

Es un entorno de desarrollo integrado (IDE) para sistemas operativos Windows. Soporta múltiples lenguajes de programación, tales como: C++, C#, Visual Basic .NET, f#, Java, Python, Ruby y PHP, al igual que entornos de desarrollo web, como ASP.NET MVC, Django, etc., a lo cual hay que sumarle las nuevas capacidades online bajo Windows Azure. (WIKIPEDIA)

2.2.10.4. Gestores de datos

SQL Server

SQL Server es un sistema de gestión de base de datos relacionales de Microsoft que está diseñado para el entorno empresarial. SQL Server se ejecuta en T-SQL – Transact SQL – que es un conjunto de extensiones de programación de Sybase y Microsoft que añaden varias características a SQL estándar, incluye control de transacciones, excepción y manejo de errores, procesamiento de filas, así también cómo variables declaradas. (Rouse, 2015)

Mongo DB

MongoDB es una base de datos de propósito general potente, flexible y escalable. Combina capacidad de escalar con características tales como índices secundarios, consultas de rango, clasificación, agregaciones e índices geoespaciales.

MongoDB es una base de datos orientada a documentos no relacional. Una base de datos orientada a documentos reemplaza el concepto de fila con un modelo más flexible llamado documento. Permite realizar documentos anidados y arreglos, el enfoque orientado a documentos hace posible representar relaciones jerárquicas complejas con un simple registro. MongoDB está diseñado para escalar de manera muy rápida. (Chodorow, 2013)

CAPÍTULO 3. HIPÓTESIS

3.1. Formulación de la Hipótesis

El desarrollo y aplicación de un agente inteligente de recomendación basado en filtrado de contenido mejora significativamente la experiencia de los usuarios de Alternate Earths, aumentando el nivel de aceptación del sistema y el nivel de interés de los usuarios a la información del sitio.

3.2. Operacionalización de variables

3.2.1. Variable Dependiente

Tabla 9. Operacionalización de variable dependiente.

VARIABLE	DEF. CONCEPTUAL	DEF. OPERACIONAL	DIMENSIONES	INDICADORES
EXPERIENCIA DE LOS USUARIOS DE ALTERNATE EARTHS	Según la ISO 9241-210, Es el resultado de las percepciones y respuestas de una persona por el uso de un producto, sistema o servicio.	Conjunto de preferencias de un usuario ante la utilización del sitio web, cuyo	Nivel de aceptación del sistema	Número de suscriptores
				Tiempo de permanencia en el sitio.
			Nivel de interés de información	Índice de interés en el sitio.

3.2.2. Variable Independiente

Tabla 10. Operacionalización de variable independiente

VARIABLE	DEF. CONCEPTUAL	DEF. OPERACIONAL	DIMENSIONES	INDICADORES
AGENTE INTELIGENTE DE RECOMENDACIÓN	Agente inteligente es una entidad software que, basándose en su propio conocimiento, realiza un conjunto de operaciones para satisfacer las necesidades de un usuario o de otro programa, bien por iniciativa propia o porque alguno de estos se lo requiere.	Agente inteligente de recomendación ofrece información recomendada a los usuarios, en base a semejanzas ente las características de los mismos, validando la funcionalidad y la eficiencia de los resultados.	Funcionalidad	Porcentaje de precisión de las recomendaciones
				Porcentaje de error de precisión en las recomendaciones
			Eficiencia	Tiempo de respuesta promedio de los algoritmos.

CAPÍTULO 4. DESARROLLO

En el presente capítulo se explicará la planificación del proyecto y se detallará la secuencia de pasos realizados en cada etapa de la implementación del agente inteligente de recomendación.

4.1. Comprensión del negocio

4.1.1. Evaluación Actual

La plataforma Alternate Earths no presenta ningún tipo de restricciones a sus usuarios con respecto a la cantidad de contenido que pueden crear. Esto conlleva a que los usuarios, al navegar por las diferentes páginas del sitio, sean bombardeados por información poco relevante. La gran variedad de temas expuestos por la plataforma genera un desinterés constante por parte de los usuarios, sobre todo porque dicha información es mostrada según fecha de creación.

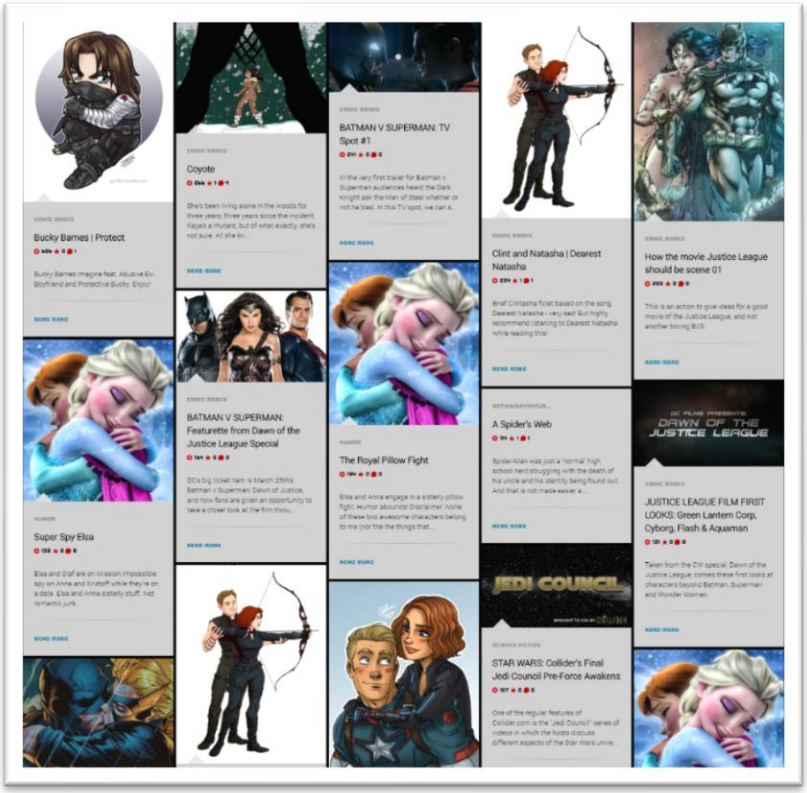


Figura 14. Página Principal de Alternate Earths

Asimismo, el sitio no incentiva al usuario a seguir navegando. Un claro ejemplo de este fenómeno lo podemos apreciar en la siguiente pantalla, en donde se visualiza que el sistema no recomienda enlaces a otra información.

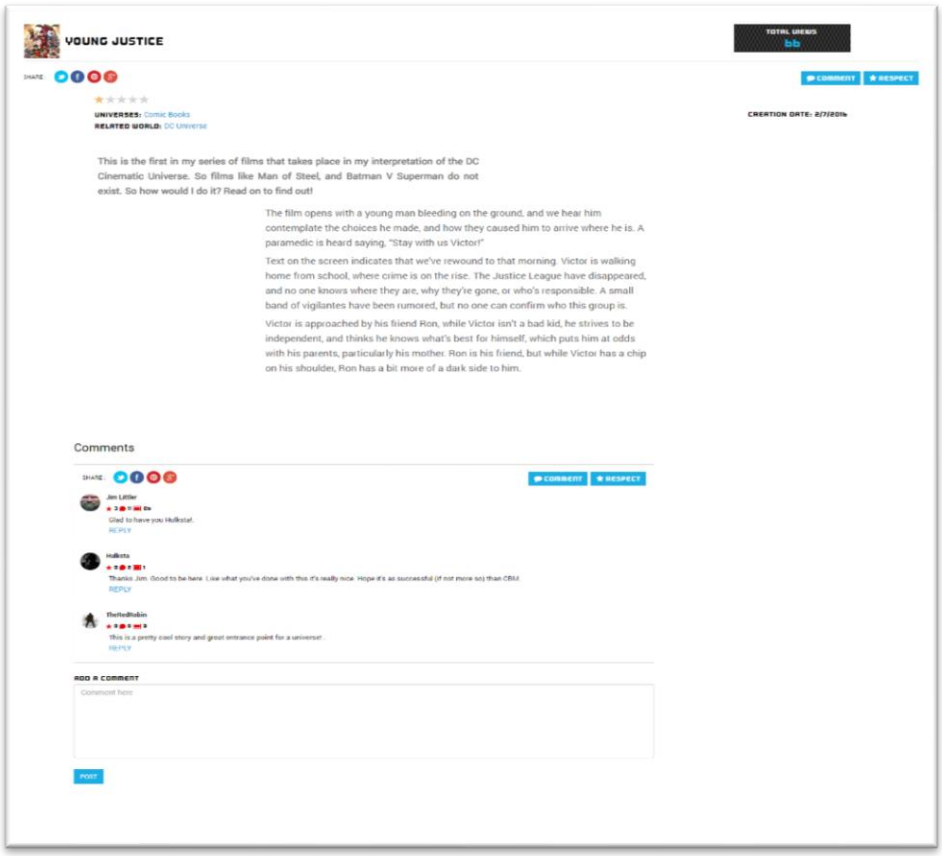


Figura 15. Página submission

4.1.2. Determinar Objetivos

4.1.2.1. Objetivo General

Convertirse en sitio de colaboración diseñado específicamente para desarrollar universos ficticios únicos poblados con personajes e historias originales.

4.1.2.2. Objetivos Específicos

- Convertirse en un repositorio de información de comics ya conocidos (Marvel, DC).
- Fomentar la discusión de temas relacionados a los comics
- Dar a conocer las noticias más relevantes del mundo del Comic

4.1.3. Plan de Proyecto

4.1.3.1. Recursos Humanos

En esta subsección se describe los nombres de los integrantes del equipo y el rol a desempeñar de cada uno.

Tabla 11. Integrantes del Equipo

NRO.	MIEMBRO	ROL
01	Lezcano Menchola, Francisco	Analista – Programador – Tester
02	Quispe Hernández, Aixa	Analista – Programador – Tester
03	Ing. Díaz Amaya, Lourdes	Asesor

4.1.3.2. Costes

En esta subsección se detallará los costos existentes en el proyecto de tipo: hardware, software, servicios y recursos humanos.

Tabla 12. Costo de Hardware

RECURSOS DE HARDWARE			
EQUIPO	**CANTIDAD**	**PRECIO UNITARIO**	**PRECIO FINAL (Depreciación 2 años)**
Laptop	2	3500.00	2500.00
		TOTAL	*S/. 5000.00*

Tabla 13.Costos de Software

RECURSOS DE SOFTWARE			
DESCRIPTION	**CANTIDAD**	**PRECIO**	**SUB TOTAL**
Windows 10	1	416.50	416.50
Office 365	1	350.00	365.50
Visual Studio 2017 Community	2	0.00	0.00
Mongo DB	2	0.00	0.00
SQL Azure	1	140	140.00
Servidor Azure	1	45.50	45.50
		TOTAL	*S/. 967.50*

Tabla 14. Costos de servicios

SERVICIOS			
EQUIPO	**PRECIO (MES)**	**MESES**	**PRECIO FINAL (Depreciación 2 años)**
Internet Movistar 20 MB	120.00	4	480.00
		TOTAL	*S/. 480.00*

Tabla 15. Costos de Recursos Humanos

RECURSOS HUMANOS			
PERSONAL	**MESES**	**PRECIO**	**SUB TOTAL**
(01) Analista – Programador – Tester	4	1500.00	6000.00
		TOTAL	*S/. 6000.00*

4.1.3.3. Fases de Desarrollo

En esta subsección se describe y especifica las tareas realizadas por cada fase de desarrollo. El proyecto ha sido divido en 6 fases las cuales han permitido obtener el agente de recomendación.

Tabla 16. Descripción de Fase de desarrollo Nro. 01

FASE 01
Nombre de la fase: Compresión del giro de negocio
Encargados: - Lezcano Menchola, Francisco - Quispe Hernández, Aixa
Descripción: En esta fase se obtuvo información de la situación actual del negocio y del sitio Alternate Earths (AE). Las tareas de esta fase son: - Determinar y evaluar situación actual del sitio. - Determinar objetivos del negocio. - Realizar la planificación del proyecto.

Tabla 17. Descripción de Fase de desarrollo Nro. 02

FASE 02
Nombre de la fase: Compresión de los datos
Encargados: - Lezcano Menchola, Francisco - Quispe Hernández, Aixa
Descripción: En esta fase se accedió a la base de datos del sitio, se determinó las tablas a usar para la obtención de datos. Las tareas de esta fase son: - Acceso a base de datos. - Identificación de tablas para la obtención de datos. - Mapeo de los campos de las tablas seleccionadas.

Tabla 18. Descripción de Fase de desarrollo Nro. 03

FASE 03
Nombre de la fase: Preparación de datos
Encargados: - Lezcano Menchola, Francisco

- Quispe Hernández, Aixa
Descripción:
Después de realizar la identificación de las tablas claves para la obtención de datos, se realiza algoritmos que permiten la preparación de datos. Las tareas de esta fase son: - Aplicar técnicas de preparación y/o limpieza de la información

Tabla 19. Descripción de Fase de desarrollo Nro. 04

FASE 04
Nombre de la fase: Modelado
Encargados: - Lezcano Menchola, Francisco - Quispe Hernández, Aixa
Descripción:
En esta fase se realiza investigación sobre tecnologías y plataformas de desarrollo que permitan cumplir los objetivos del agente de recomendación. Además de la implementación de las técnicas. Las tareas de esta fase son: - Seleccionar tipo de arquitectura para el modelo. - Seleccionar las técnicas y algoritmos a aplicar. - Modelar y generar documento en base de datos no relacional. - Implementación de la arquitectura elegida. - Implementación las técnicas y algoritmos a aplicar.

Tabla 20. Descripción de Fase de desarrollo Nro. 05

FASE 05
Nombre de la fase: Evaluación del modelo
Encargados: - Lezcano Menchola, Francisco - Quispe Hernández, Aixa
Descripción:
En esta fase se realiza pruebas del modelo para ver su funcionamiento. Las tareas de esta fase son: - Generar pruebas a los algoritmos para conocer su tiempo de respuesta.

Tabla 21. Descripción de Fase de desarrollo Nro. 06

FASE 06
Nombre de la fase: Implantación del agente
Encargados: - Lezcano Menchola, Francisco - Quispe Hernández, Aixa
Descripción:
En esta fase se la implantación del agente de recomendación en entorno de producción.

Las tareas de esta fase son:

- Planificación y ejecución de la implantación
- Elaborar informe final

4.1.3.4. Planificación Inicial

En esta subsección se define la prioridad (Bajo, Media o Alta) según la importancia que tenga cada fase, además del esfuerzo (Bajo, Medio o Alto) según el tiempo y trabaja que demandará en desarrollar la fase.

Tabla 22. Planificación inicial por fase

ITERACIÓN	FASE	PRIORIDAD	ESFUERZO
01	Comprensión de giro de negocio	Alta	Medio
02	Obtención de datos	Alta	Alto
03	Preparación de datos	Alta	Medio
04	Modelado	Alta	Medio
05	Evaluación del modelo	Alta	Alto
06	Implantación del agente	Alta	Bajo

4.1.3.5. Estimación de Tiempo

De acuerdo a la definición de prioridades y esfuerzo por fase de desarrollo se ha estimado el tiempo de desarrollo de cada fase, estimando un total de 89 días para la entrega del agente de recomendación.

Tabla 23. Tiempo estimado por fase de desarrollo

NRO.	FASE	TIEMPO ESTIMADO
01	Comprensión del giro de negocio	4 días
02	Obtención de datos	20 días
03	Preparación de datos	20 días
04	Modelado	30 días
05	Evaluación del modelo	10 días
06	Implantación del agente	5 días
	TOTAL	89 días

4.2. Comprensión de los datos

4.2.1. Recopilación de Datos Iniciales

La recopilación de datos se realizó a través de la siguiente fuente:

- **Base de datos – xComic**: ésta base de datos contiene todas las tablas que conforman el sitio Alternate Earths, la cual contiene información de los usuarios registrados activos e inactivos; además contiene los documentos, interacción del usuario con el sitio, la cual servirá como fuente inicial de obtención de datos.

4.2.2. Descripción de Datos

Para el desarrollo del proyecto, se utilizan todos los datos existentes desde su lanzamiento a producción del sitio Alternate Earths. Se elaboró un diccionario de datos de las tablas implicadas en la obtención de datos preliminares al análisis

Tabla 24. Descripción de Tabla Submission

Nombre de Tabla: TBL_SUBMISSION			
Campo	**Descripción**	**Tipo de Dato**	**Requerido**
ID	Almacena el Identificador del submission	Bigint	Si
TITLE	Título del submission	Nvarchar (600)	Si
DISTRIBUTORID	Almacena el identificador del usuario creador del submission	Bigint	Si
STATUS	Almacena del estado del submission	Smallint	Si
DESCRIPTION	Almacena una breve descripción del submission	Text	Si
SUMMARY	Almacena el contenido general del submission	Nvarchar (Max)	Si
TOTALFOLLOWERS	Almacena el número total de seguidores del submission	Int	No
TOTALCOMMENTS	Almacena el número total de comentarios del submission	Int	No
TOTALRESPECTS	Almacena el número total de respects del submission	Int	No
ALLKEYWORDS	Almacena las palabras claves del submission	Nvarchar (Max)	Si
TOTALVIEWS	Almacena el número de vistas del submission	Int	Si

Tabla 25. Descripción de la tabla Distributor

Nombre de Tabla: TBL_DISTRIBUTOR			
Campo	**Descripción**	**Tipo de Dato**	**Requerido**
ID	Almacena el identificador del usuario	bigint	Si
FIRSTNAME	Almacena el primero nombre del usuario	Varchar (200)	Si
LASTNAME	Almacena los apellidos del usuario	Varchar (100)	Si
GENDER	Almacena el género del usuario.	Smallint	Si
USERNAME	Almacena el nombre de usuario	Varchar (50)	Si
PASSWORD	Almacena el password del usuario	Varchar (50)	Si
EMAIL	Almacena el email del usuario	Nvarchar (Max)	Si
ABOUT	Almacena breve descripción del usuario	Int	No
LOCATION	Almacena ubicación del usuario	Int	No
TOTALRESPECTS	Almacena el número total de respects del usuario	Int	Si
TOTALFOLLOWERS	Almacena el número total de seguidores del usuario.	Int	Si
TOTALCOMMENTS	Almacena el número total de comentarios del usuario.	Int	Si
TOTALSUBMISSIONS	Almacena el número total submission creados por el usuario	Int	Si
TOTALARTWORKS	Almacena el número total de obras de arte creadas por el usuario.	Int	Si
TOTALPHOTOGRAPHS	Almacena el número total de fotografías creadas por el usuario.	Int	Si
TOTALWRITINGS	Almacena el número total de escritos creados por el usuario.	Int	Si
TOTALVIDEOS	Almacena el número total de videos creados por el usuario.	Int	Si
ALLKEYWORDS	Almacena palabras claves de los gustos y preferencias del usuario.	Int	Si
TOTALVIEWS	Almacena el número de vistas del usuario en el sitio	Int	Si

Tabla 26. Descripción de la Tabla News

Nombre de Tabla: TBL_NEWS			
Campo	**Descripción**	**Tipo de Dato**	**Requerido**
ID	Almacena el identificador de news.	Int	Si
NEWSSUBJECT	Almacena el nombre del tema del new.	Varchar (150)	Si
NEWSTEXT	Almacena el contenido del new.	Text	Si
LANGUAGEID	Almacena el idioma del new.	Int	Si
STATUS	Almacena el estado del new	Smallint	Si
SUMMARY	Almacena el resumen del new.	Varchar (500)	Si
THUMBNAILNAME	Almacena el nombre de la imagen del new.	Varchar (200)	Si
KEYWORDS	Almacena las palabras claves del new.	Varchar (100)	Si
TOPNEW	Almacena un identificador para saber si el new es o no un TOP.	bit	Si
TOTALFOLLOWERS	Almacena el número total de seguidores del new.	Int	Si
TOTALRESPECTS	Almacena el número total de respect del new.	Int	Si
TOTALCOMMENTS	Almacena el número total de comentarios del new.	Int	Si
TOTALVIEWS	Almacena el número total de vistas al new.	Int	Si
POSTDATE	Almacena la fecha de publicación del new.	Datetime	Si

4.3. Preparación de los datos

4.3.1. Selección de Datos

Tabla Submission: De acuerdo con el diccionario de datos (Tabla Nro. 24), se realiza un procedimiento almacenado en lenguaje Transact – SQL para seleccionar las columnas de la tabla "TBL_SUBMISSION" que serán utilizadas el algoritmo para el agente de recomendación.

```
-- ============================================
-- AUTHORS: Aixa Quispe - Francisco Lezcano
-- CREATION: [2017-09-09]
-- PROCESS: Recommendation system
-- NOTES: Get all active submission
-- MODIFICATION:
--   Francisco Lezcano [12/10/2017] Alias
--   Aixa Quispe [01/07/2018] Add keys as alias in columns
-- ============================================
CREATE PROCEDURE SP_XCOMICS_RECOMMENDER_SUBMISSIONS_LIST
AS
BEGIN
	SET TRANSACTION ISOLATION LEVEL READ UNCOMMITTED;
	SET NOCOUNT ON;

	SELECT		S.ID [SubmissionId],
			ISNULL(S.TITLE,'') [Submission Title],
			ISNULL(S.Description,'') [Description],
			ISNULL(S.Summary,'') [Submission Summary],
			ISNULL(S.TOTALFOLLOWERS,0) [Total Followers],
			ISNULL(S.TOTALCOMMENTS,0) [Total Comments],
			ISNULL(S.TOTALRESPECTS,0) [Total Respects],
			ISNULL(S.ALLKEYWORDS,0) [Key Words],
			ISNULL(S.TOTALVIEWS,0) [Total Views]
	FROM TBL_SUBMISSION AS S
	WHERE S.[STATUS] = 1 and S.PUBLICVIEWSTATUS = 1
END
```

Figura 16. Selección de submission activos

Results | Messages

	Submissio	Submission Title	Description	Submission Summ	Total Follow	Total Comme	Total Respe	Key Words	Total Views
34	2609	NERD DEBATE: Best Weapon for the...	<p><img...	Something that do...	0	0	0	NERD DEBATE: Best Weapon for the Zombie Apocalypse?...	18
35	2644	Creation Story: Birth of the Rogue Go	<p><img...	In the beginning, ...	0	0	0	Creation Story: Birth of the Rogue Gods,creation,story,birth...	24
36	2652	Story Idea: Werewolf vs. Vampire	<p><img...	No joke. I wrote u...	0	1	1	Story Idea: Werewolf vs. Vampire,story,idea,werewolf,vs,va...	51
37	2653	Training Day	<p><img...	A brief taste of my...	0	0	0	Training Day,training,day,	18
38	2666	"Ape Canyon" Sasquatch Story from ...	<p><sp...	What is Bigfoot? F...	0	0	1	"Ape Canyon" Sasquatch Story from One of the Survivors,"...	24
39	2667	Script Idea: Why hasn't Hollywood ma...	<p><img...	A monster in the w...	0	1	1	Script Idea: Why hasn't Hollywood made a decent Bigfoot H...	59
40	2748	Princess Peach Cosplayer	<p>Conte...	Princess Peach C...	0	0	0	Princess Peach Cosplayer,princess,peach,cosplayer,	17
41	2749	Wonder Woman Cosplayer		Wonder Woman h...	0	0	0	Wonder Woman Cosplayer,wonder,woman,cosplayer,	16
42	2750	Margie Cox as Star Sapphire		Booth babe and p...	0	1	1	Margie Cox as Star Sapphire,margie,cox,as,star,sapphire,	64
43	2751	Joker Cosplayer		Joker Babe	0	0	1	Joker Cosplayer,joker,cosplayer,	12
44	2755	Captain America	<p><str...	During the dark d...	0	0	0	Captain America,captain,america,	11
45	2756	Captain America and the Falcon		This is concept art...	0	0	0	Captain America and the Falcon,captain,america,and,the,fa...	10
46	2792	Page 2 of Battle between Juggernaut,...		Page 2 of Battle b...	0	0	0	Page 2 of Battle between Juggernaut, USAgent and Iron M...	8
47	2795	Page 3 of Battle between Juggernaut,...		Page 3 of Battle b...	0	0	0	Page 3 of Battle between Juggernaut, USAgent and Iron M...	11
48	2799	"Storm" of the X-Men		"Storm" of the X-M...	0	0	0	"Storm" of the X-Men ,storm,of,the,xmen,,	6
49	2800	Sith Master - Darth Sidious	<p><str...	Sith Master - Dart...	0	0	1	Sith Master - Darth Sidious,sith,master,darth,sidious,	25
50	2801	Janga Fett	<p><str...	Janga Fett, daddy...	0	0	0	Janga Fett,janga,fett,	13
51	2802	Sithster to Darth Maul		Could this be the ...	0	0	0	Sithster to Darth Maul,sithster,to,darth,maul,	6
52	2803	The Beholder	<p>The gl...	The Beholder (als...	0	0	0	The Beholder,the,beholder,	14
53	2878	Colt Python	<p>The C...	Best known as Ric...	0	0	0	Colt Python,colt,python,	14
54	2879	An ornate Pole Arm spear tip and it's ...		An ornate Pole Ar...	0	0	0	An ornate Pole Arm spear tip and it's sheath ,an,ornate,pol...	24
55	2880	Mermaid Chick with a Trident.		Mermaid Chick wit...	0	0	0	Mermaid Chick with a Trident.,mermaid,chick,with,trident,	8
56	2881	Hot Bow Woman		Hottie with a bow ...	0	0	0	Hot Bow Woman,hot,bow,woman,	20
57	2882	Fish Babe with Trident		Fish Babe with Tri...	0	0	0	Fish Babe with Trident,fish,babe,with,trident,	9
58	2883	The History of Weapons of War		The History of We...	0	0	0	The History of Weapons of War,the,history,of,weapons,of,w	24

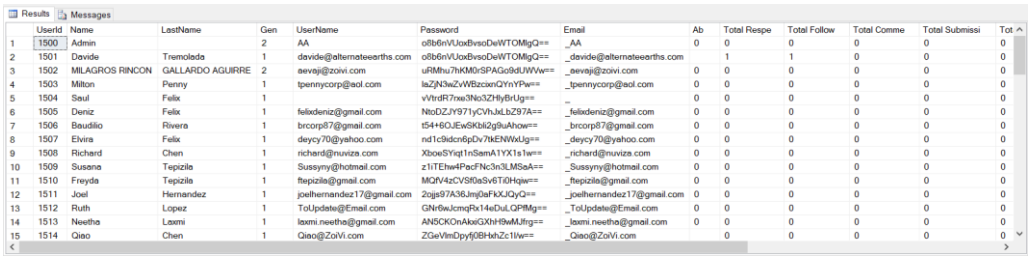
Figura 17. Información de la base de datos xComic sobre submission

- **Tabla Distributor:** De acuerdo con el diccionario de datos (Tabla Nro. 25), se realiza un procedimiento almacenado en lenguaje Transact – SQL para seleccionar las columnas de la tabla "TBL_DISTRIBUTOR" que serán utilizadas el algoritmo para el agente de recomendación.

```
-- ============================================
-- AUTHORS: Aixa Quispe - Francisco Lezcano
-- CREATION: [2017-09-09]
-- PROCESS: Recommendation system
-- NOTES: Get all active user
-- ============================================
CREATE PROCEDURE SP_XCOMICS_RECOMMENDER_DISTRIBUTOR_LIST
AS
BEGIN
	SET TRANSACTION ISOLATION LEVEL READ UNCOMMITTED;
	SET NOCOUNT ON;

	SELECT		D.ID [UserId],
			ISNULL(D.FIRSTNAME,'') [Name],
			ISNULL(D.LASTNAME,'') [LastName],
			ISNULL(D.GENDER,'') [Gender],
			ISNULL(D.USERNAME,'') [Username],
			ISNULL(D.[PASSWORD],0) [Password],
			ISNULL(D.EMAIL,'') [Email],
			ISNULL(D.ABOUT,0) [About],
			ISNULL(D.TOTALRESPECTS,0) [Total Respects],
			ISNULL(D.TOTALFOLLOWERS,0) [Total Followers],
			ISNULL(D.TOTALCOMMENTS,0) [Total Comments],
			ISNULL(D.TOTALSUBMISSIONS,0) [Total Submissions],
			ISNULL(D.TOTALARTWORKS,0) [Total Artworks],
			ISNULL(D.TOTALPHOTOGRAPHS,0) [Total Photographs],
			ISNULL(D.TOTALWRITINGS,0) [Total Writings],
			ISNULL(D.TOTALVIDEOS,0) [Total Videos],
			ISNULL(D.ALLKEYWORDS,'') [Key Words],
			ISNULL(D.TOTALVIEWS,'') [Total Views]
	FROM TBL_DISTRIBUTOR AS D
	WHERE D.USERSTATUS = 5
END
```

Figura 18. Selección de usuarios activos



Results Messages

	UserId	Name	LastName	Gen	UserName	Password	Email	Ab	Total Respe	Total Follow	Total Comme	Total Submissi	Tot
1	1500	Admin		2	AA	o8b6nVUoxBvsoDeWTOMlgQ==	_AA	0	0	0	0	0	0
2	1501	Davide	Tremolada	1	davide@alternateearths.com	o8b6nVUoxBvsoDeWTOMlgQ==	_davide@alternateearths.com		1	1	0	0	0
3	1502	MILAGROS RINCON	GALLARDO AGUIRRE	2	aevaji@zoivi.com	uRMhu7hKM0rSPAGo9dUWVw==	_aevaji@zoivi.com	0	0	0	0	0	0
4	1503	Milton	Penny	1	tpennycorp@aol.com	laZjN3wZvWBzcknQYnYPw==	_tpennycorp@aol.com	0	0	0	0	0	0
5	1504	Saul	Felix	1		vVtrdR7nxe3Nio3ZtHlyBrUg==	_	0	0	0	0	0	0
6	1505	Deniz	Felix	1	felixdeniz@gmail.com	NtoDZJY971yCVhJxLbZ97A==	_felixdeniz@gmail.com	0	0	0	0	0	0
7	1506	Baudilio	Rivera	1	brcorp87@gmail.com	t54+6OJEwSKbli2g9uAhow==	_brcorp87@gmail.com	0	0	0	0	0	0
8	1507	Elvira	Felix	1	deycy70@yahoo.com	nd1c9idcn6pDv7tkENWxUg==	_deycy70@yahoo.com	0	0	0	0	0	0
9	1508	Richard	Chen	1	richard@nuviza.com	XboeSYiqt1nSamA1YX1s1w==	_richard@nuviza.com	0	0	0	0	0	0
10	1509	Susana	Tepizila	1	Sussyny@hotmail.com	z1iTEhw4PacFNc3n3LMSaA==	_Sussyny@hotmail.com	0	0	0	0	0	0
11	1510	Freyda	Tepizila	1	ftepizila@gmail.com	MQfV4zCVSf0aSvET/0Hqw==	_ftepizila@gmail.com	0	0	0	0	0	0
12	1511	Joel	Hernandez	1	joelhernandez17@gmail.com	2cjje97A36Jmj0aFkXJQyQ==	_joelhernandez17@gmail.com	0	0	0	0	0	0
13	1512	Ruth	Lopez	1	ToUpdate@Email.com	GNr6wJcmqRx14xDuLQPlMg==	_ToUpdate@Email.com	0	0	0	0	0	0
14	1513	Neetha	Laxmi	1	laxmi.neetha@gmail.com	AN5CKOnAkaGXhH9wMJftfg==	_laxmi.neetha@gmail.com	0	0	0	0	0	0
15	1514	Qiao	Chen	1	Qiao@ZoiVi.com	ZGeVlmDpyfj0BHbhZc1Iw==	_Qiao@ZoiVi.com		0	0	0	0	0

Figura 19. Información de la base de datos xComic sobre usuarios

- **Tabla News:** De acuerdo con el diccionario de datos (Tabla Nro. 26), se realiza un procedimiento almacenado en lenguaje Transact – SQL para seleccionar las columnas de la tabla “TBL_NEWS” que serán utilizadas el algoritmo para el agente de recomendación.

```
-- ============================================
-- AUTHORS: Aixa Quispe - Francisco Lezcano
-- CREATION: [2017-09-09]
-- PROCESS: Recommendation system
-- NOTES: Get all active news
-- ============================================
CREATE PROCEDURE SP_XCOMICS_RECOMMENDER_NEWS_LIST
AS
BEGIN
	SET TRANSACTION ISOLATION LEVEL READ UNCOMMITTED;
	SET NOCOUNT ON;
	SELECT		N.ID [News Id],
			ISNULL(N.NEWSSUBJECT,'') [Subject],
			ISNULL(N.NEWSTEXT,'') [New Text],
			ISNULL(N.LANGUAGEID,'') [Language Id],
			ISNULL(N.[STATUS],'') [Status],
			ISNULL(N.SUMMARY,0) [Summary],
			ISNULL(N.KEYWORDS,'') [Key Words],
			ISNULL(N.TOPNEW,0) [IsTopNew],
			ISNULL(N.TOTALRESPECTS,0) [Total Respects],
			ISNULL(N.TOTALFOLLOWERS,0) [Total Followers],
			ISNULL(N.TOTALCOMMENTS,0) [Total Comments],
			ISNULL(N.TOTALVIEWS,'') [Total Views]
	FROM TBL_NEWS AS N
	WHERE N.[STATUS] = 2
END
```

Figura 20. Selección de news activos

Results | Messages

	News Id	Subject	NewText	Languag...	Stat	Summary	KeyWords
19	1140	asdad	adadasd	0	2	asdasd	
20	1141	test-new		0	2	test-new	
21	1142	Develop hedline...	<p>Dwayne "The Rock&qu...	0	2	Dwayne "The Rock" Johnson was among the many st...	
22	1143	test-new	supercampeones	0	2	test-new	
23	1144	TITLE DEVELO...	<p>Last June we had the opportunity to vi...	0	2	Last June we had the opportunity to visit the set of Av...	
24	1145	new developme...		0	2	aaa bb ccc	
25	1146	prueba-prueba-...		0	2	summary	
26	1147	development ne...	hello	0	2	aaa bbb cc sss	
27	1148	develpment 123		0	2	summary here	
28	1149	one test here 1		0	2	11456	
29	1150	develop		0	2	aaaa	
30	1151	test-test|12312	naruto	0	2	summary	
31	1152	test-new-123123	marutop	0	2	summary	
32	1153	new title here fo...		0	2	aa	
33	1154	testing news	content here	0	2	summary h	
34	1160	Test	shamanking	0	2	Test	
35	1161	Prueba-new	prueba, prueba, prueba <...	0	2	prueba-new	
36	1162	prueba-new-new	new please	0	2	new-new	

Figura 21. Información de la base de datos xComic sobre news

4.3.2. Limpiar los datos

La limpieza de datos se realizó siguiendo los siguientes puntos.

- **Conversión de Data a Formato JSON**

La limpieza de datos se realizará del lado del agente de recomendación y se usará una base de datos no relacional (Mongo DB), para realizar el análisis es importante pasar toda la información de la base de datos xComic (SQL Server) a Mongo DB; por ello se implementó un método para convertir todo los submission en formato JSON y almacenarlos en un archivo texto antes de realizar la limpieza.

```
public Boolean Recommender_InitRecommender(ref BaseEntity Base)
{
  Base = new BaseEntity();
  Boolean success = false;
  List<object> lst = new List<object>();
  DataTable dt = Submissions_ListAllSubmissionForRecommender(ref Base);
    foreach (DataRow item in dt.Rows){
       lst.Add(new{
               ItemId = item["ID"],
                Text=clsUtilities.StripHTML(
                     clsWebUtility.HtmlDecode(item["DESCRIPTION"].ToString())
                     )
                });
     }
  using
(
StreamWriter file = File.CreateText(@"C:\xcomics\trunk\fileSubmissions.txt")
)
            {
                string jsonStr = clsWebUtility.JSONSerialize(lst);
                file.Write(jsonStr);
            }
            return success;
        }
```

Figura 22. Método inicializador del agente de recomendación (SUBMISSION)

```
public Boolean Recommender_InitRecommender(ref BaseEntity Base)
{
  Base = new BaseEntity();
  Boolean success = false;
  List<object> lst = new List<object>();
  DataTable dt = Submissions_ListAllNewsForRecommender (ref Base);
    foreach (DataRow item in dt.Rows){
       lst.Add(new{
               ItemId = item["ID"],
                Text=clsUtilities.StripHTML(
                     clsWebUtility.HtmlDecode(item["DESCRIPTION"].ToString())
                     )
                });
     }
  using
(
StreamWriter file = File.CreateText(@"C:\xcomics\trunk\fileNews.txt")
)
            {
                string jsonStr = clsWebUtility.JSONSerialize(lst);
                file.Write(jsonStr);
            }
            return success;
        }
```

Figura 23. Método inicializador del agente de recomendación (NEWS)

- **Procesamiento de Archivo y Limpieza**

  En esta etapa se tiene un archivo de texto con toda la información, para lo cual se implementó un método que lea el archivo y realice la limpieza, la cual trata de remover números, urls, correos electrónicos, código HTML, símbolos de monedas; además de realizar una división por cada signo de puntuación que encuentre en el texto.

```
/// <summary>
/// Tokenizes a string, returning its list of words.
/// </summary>
/// <param name="text">string</param>
/// <returns>string[]</returns>
public static string[] Tokenize (string text)
{
  // Strip all HTML.
  text = Regex.Replace(text, "<[^<>]+>", "");

  // Strip numbers.
  text = Regex.Replace(text, "[0-9]+", "number");

  // Strip urls.
  text = Regex.Replace(text, @"(http|https)://[^\s]*", "httpaddr");

  // Strip email addresses.
  text = Regex.Replace(text, @"[^\s]+@[^\s]+", "emailaddr");

  // Strip dollar sign.
  text = Regex.Replace(text, "[$]+", "dollar");

  // Strip usernames.
  text = Regex.Replace(text, @"@[^\s]+", "username");

  // Tokenize and also get rid of any punctuation
  return text.Split(" @$/#.-:&*+=[]?!(){},''\">_<;%\\".ToCharArray());
}
```

Figura 24. Método para limpiar la información

Fuente. Expresiones regulares en.NET

4.4. Modelado

4.4.1. Arquitectura de la Aplicación

a) Componente Web

Este componente contiene la plataforma web Alternate Earths, en la cual se administra: submission, news, worlds los cuales están relacionados al mundo del comic. Esta plataforma está desarrollada con ASP.NET con el framework 4.5.

b) Componente API

Este componente contiene los métodos de la limpieza de datos, y los métodos relacionados al agente de recomendación en el cual se aplican los algoritmos para crear el perfil de aprendizaje por usuario. Se ha usado un proyecto tipo web API, el cual permitirá proporcionará métodos que serán consumidos desde el componente web.

c) Base de datos Relacional

Este componente es la base de datos existente llamada xComic la cual almacena todo lo administrado desde el componente web. Esta base de datos está en SQL Azure.

d) Base de datos No Relacional

Este componente es la base de datos no relacional, donde se almacenará la información generada en el componente Web Api, en esta base de se guardará los perfiles de usuario. Este componente estará desplegado con Mongo DB.

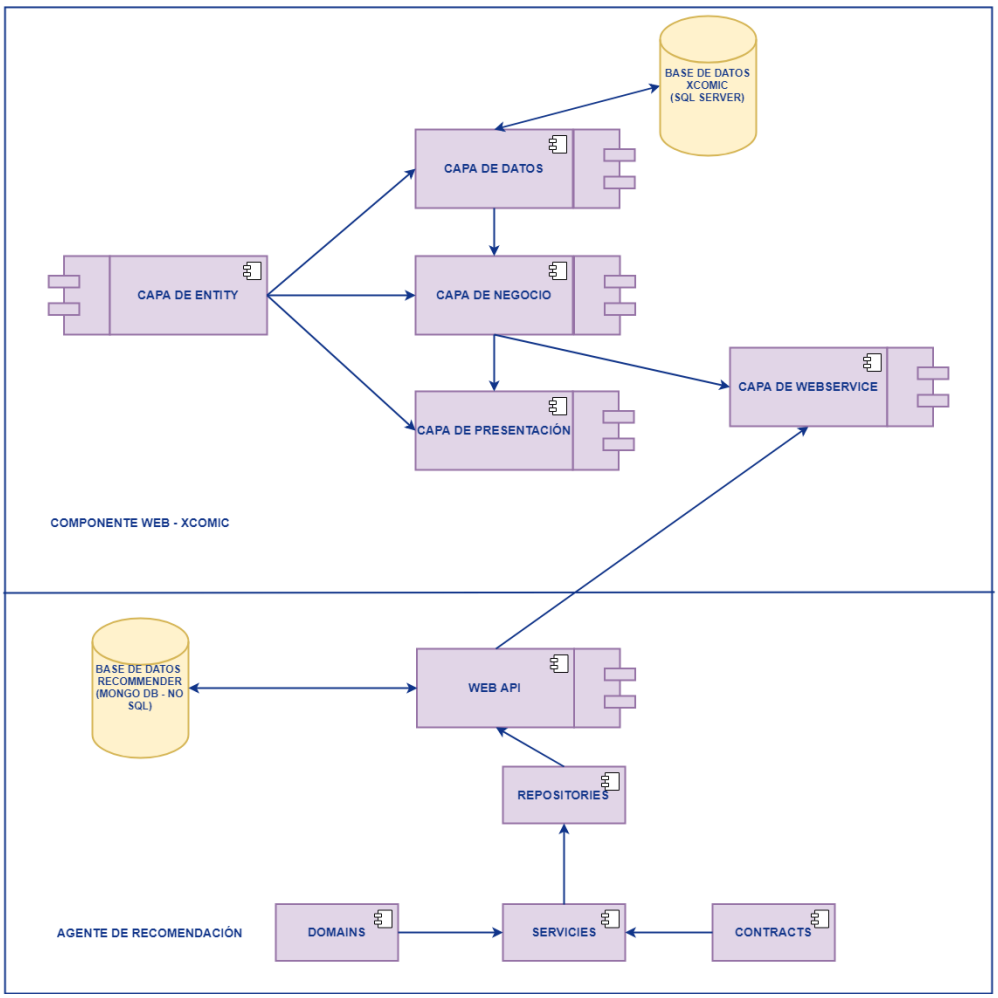


Figura 25. Componentes de agente de recomendación

4.4.2. Técnicas

El agente de recomendación desarrollado está basado en un modelo de espacio vectorial, el cual es modelo estadístico – matemático que se base en el grado de similaridad de una consulta dada por el usuario con respecto a los documentos de una colección que fueron ponderados mediante la técnica del TF – IDF.

4.4.2.1. Factor TF

El factor TF es la suma de todas ocurrencias o el número de veces que aparece un término en un documento, con este factor se puede conocer la importancia de los

términos para representar un documento de los submission y news del sitio Alternate Earths.

Tabla 27. Técnica de Factor TF

FÓRMULA MATEMÁTICA	PSEUDOCODIGO
$tf(n) = \sum D1(n)$	***Entrada:*** *documents strings[ ]* ***Salida:*** *valueTF double [ ]* *INICIO* *foreach (term in documents)* *valueTF[term] = documents.contains(term). count();* *return valueTF* *FIN*
Donde: **$tf(n)$:** suma de las ocurrencias de dicho termino **n:** frecuencia de aparición de un término **$D1$:** representación del documento	

4.4.2.2. Factor IDF

El factor IDF de un término es inversamente proporcional al número de documentos en los que aparece dicho término, cuyo valor determina la capacidad discriminatoria del término de un documento con respecto a la colección.

Tabla 28. Técnica de factor IDF

FÓRMULA MATEMÁTICA	PSEUDOCODIGO
$idf(n) = log_{10} \frac{N}{DF_n}$	***Entrada:*** *vocabularyTF strings[ ]* ***Salida:*** *valueIDF double [ ]* *INICIO* *foreach (term in vocabularyTF)* *valueIDF [term] = Log(**vocabularyTF. contains(term). count()** / vocabularyTF(term).ValueTF);* *return valueIDF* *FIN*
Donde: **$idf(n)$:** coeficiente que determina la capacidad discriminatoria de la palabra en el documento. **log_{10}:** utilizado para obtener un coeficiente bajo fácil de manejar **N:** número total de documentos de la colección **DF_n:** número de documento en los que aparece el término n	

4.4.2.3. Peso TF – IDF

Este peso corresponde al producto de ambos TF e IDF, este resultado es una representación de la importancia del término en cada documento.

Tabla 29. Técnica de peso TF – IDF

FÓRMULA MATEMÁTICA	PSEUDOCODIGO
$TF - IDF_{(n,d)} = TF_{(n,d)} \times IDF_{(n)}$	***Entrada:*** *vocabularyTFIDF strings[ ]* ***Salida:***
Donde:	

$TF - IDF_{(n,d)}$: peso de un término (n) en un documento (d) $TF_{(n,d)}$: frecuencia de aparición de un término (n) en un documento (d) $IDF_{(n)}$: factor IDF de un término (n)	*pesoTFIDF double [ ]* *INICIO* *foreach (term in vocabularyTFIDF)* *pesoTFIDF [term] =* ***vocabularyTFIDF(term).TF** * **vocabularyTFIDF(term).IDF*** *return pesoTFIDF* *FIN*

4.4.2.4. Proceso de Vectorización

Después de tener los pesos TF – IDF calculados, se debe realizar el proceso de vectorización, la cual tiene como objetivo ponderar la importancia de los términos de la consulta para poder generar el vector de la consulta del usuario, este paso es importante antes de realizar el proceso de equiparación de la consulta con los documentos de la colección y poder determinar cuáles de ellos son más relevantes.

Tabla 30. Proceso de vectorización

FÓRMULA MATEMÁTICA	PSEUDOCODIGO
$\overrightarrow{v_{(norm)}} = \frac{v}{\sqrt{v_1^2 + v_2^2 + v_3^2 + \cdots + v_n^2}}$ Donde: v: $tf - idf_{raw}$ v_n^2: vector que representa a un término n al cuadrado.	***Entrada:*** *vectors <string, double>* ***Salida:*** *NonralizedVector double [ ] [ ]* *INICIO* *foreach (v in vectors)* *var normalized = Normalize (v)* *NonralizedVector.add (normalized)* *Return NonralizedVector.ToArray()* *FIN*

	Nombre de Función: Normalize **Entrada:** Vector double [ ] **Salida:** Result double [ ] INICIO double sumSquared=0 double sqrtSumSquared=0 List<double> result Foreach (value in vector) sumSquared += value * value sqrtSumSquared= Math.Sqrt(sumSquared) Foreach (value in vector) result.add (value / sqrtSumSquared) RETURN result.ToArray() FIN

4.4.2.5. Proceso de Similaridad mediante Fórmula de Coseno.

Después de haber realizado los procesos anteriores, es posible medir cual es la desviación de un documento respecto a una consulta, por el número de grados del ángulo que forman. Este proceso es posible porque ambos vectores crean una estructura triangular a la que se le aplica el cálculo del ángulo que forma la hipotenusa (en este caso el vector del documento n) y el adyacente (el vector q de la consulta dada por el usuario) el cual resulta ser el coseno del triángulo. La interpretación de este ángulo es que si existe cierta distancia del vector de la consulta con respecto al documento m; implica que el ángulo que forma será menor y que su nivel de coincidencia será superior, si un coseno de 0° implicaría una similaridad máxima.

Tabla 31. Proceso de Similaridad mediante Fórmula de Coseno

FÓRMULA MATEMÁTICA
$$SimCos(d_{(d)}, q) = \frac{\sum_{n=1}(P_{(n,d)} \times P_{(n,q)})}{\sqrt{\sum_{n=1}(P_{(n,d)})^2 \times \sum_{n=1}(P_{(n,q)})^2}}$$
Donde: $P_{(n,d)} \times P_{(n,q)}$: Ponderación del vector valorado por el usuario y vector aun no valorado. $\sum_{n=1}(P_{(n,d)})^2$: Sumatoria de los valores TF – IDF al cuadrado del vector del documento 1 $\sum_{n=1}(P_{(n,q)})^2$: Sumatoria de los valores TF – IDF al cuadrado del vector del documento 2

4.4.3. Migración de SQL – No SQL

Es importante pasar toda la información de la base de datos xComic (SQL Server) a Mongo DB; por ello se implementó un método para convertir todo los submission en formato JSON y almacenarlos en un archivo texto que serán leídos y migrado a la base de datos no relacional.

```
public Boolean Recommender_InitRecommender(ref BaseEntity Base)
{
  Base = new BaseEntity();
  Boolean success = false;
  List<object> lst = new List<object>();
  DataTable dt = Submissions_ListAllSubmissionForRecommender(ref
Base);
    foreach (DataRow item in dt.Rows){
       lst.Add(new{
              ItemId = item["ID"],
               Text=clsUtilities.StripHTML(
                    clsWebUtility.HtmlDecode(item["DESCRIPTION"].To
             String())
                   )
              });
     }
  using
(
StreamWriter file =
File.CreateText(@"C:\xcomics\trunk\fileSubmissions.txt")
)
            {
                string jsonStr = clsWebUtility.JSONSerialize(lst);
                file.Write(jsonStr);
            }
            return success;
        }
```

Figura 26. Método inicializador del agente de recomendación (SUBMISSION)

Es importante resaltar que MongoDB es una base de datos NoSQL cuya información es guardada en documentos usando formato JSON. A diferencia de las bases de datos relaciones, MongoDB permite almacenar información sin usar una estructura predeterminada, haciendo de los sistemas que lo usan, plataformas fácilmente escalables. El agente inteligente de recomendación que la presente investigación plantea será usado en las diferentes secciones de Alternate Earths, por lo que está sujeto a ser enriquecido constantemente según los diferentes requerimientos que el sitio demande. Habiendo dicho esto, la principal razón por lo que se seleccionó a MongoDB como motor de base de datos para la presente investigación, es la flexibilidad que brinda para guardar la información.

Así mismo, se espera que el sistema de recomendación cuente con una gran cantidad de información, haciendo necesario el uso de soluciones de Big Data. MongoDB presenta una fácil integración con tecnologías como Hadoop, la cual se piensa usar en un futuro no muy lejano.

4.4.4. Diseño de Base de Datos No – SQL.

En la base de datos mongo DB, se crearon “tablas” las cuales permitirán almacenar los documentos extraídos del sitio Alternate Earths, y almacenar los valores de los cálculos realizados por los algoritmos, además de la generación del perfil de aprendizaje de los usuarios del sitio.

a) Estructura de Ítem

En esta tabla se almacena los documentos analizados, haciendo referencia al id de la base de datos relacional xComic, y almacena las palabras normalizadas del texto con su respectivo TF – IDF.

Figura 27. Estructura de "Tabla - Ítem" en mongo DB

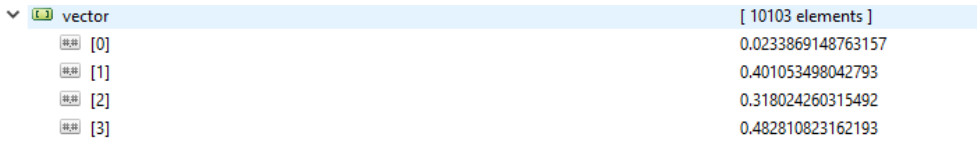


Figura 28. Estructura de "Tabla - Ítem Vector" en mongo DB

b) Estructura de Word

Contiene el análisis de las palabras de los documentos, cuantas veces se repite, su peso y los cálculos TF e IDF respectivamente.

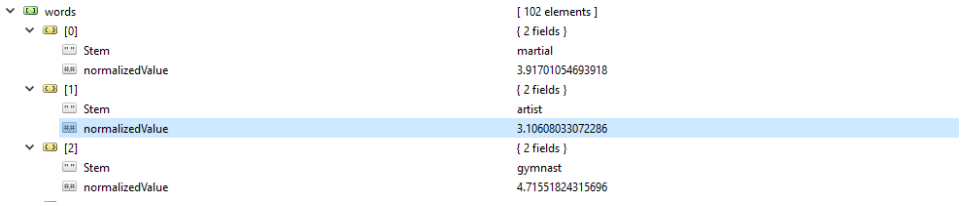


Figura 29. Ilustración 27. Estructura de "Tabla - Word" en mongo DB

4.5. Pruebas

En esta fase, se realizaron pruebas para verificar el tiempo de respuesta de los métodos que implementan los algoritmos de recomendación; para ello se utilizó herramientas propias del Visual Studio Enterprise 2017 – Web Performance y Load Test.

En las primeras pruebas el tiempo de respuesta de los algoritmos al generar los perfiles de usuario la respuesta del test web era de TIMEOUT, para lo cual se tuvo que realizar seguimiento y mejora en la implementación de estos algoritmos.

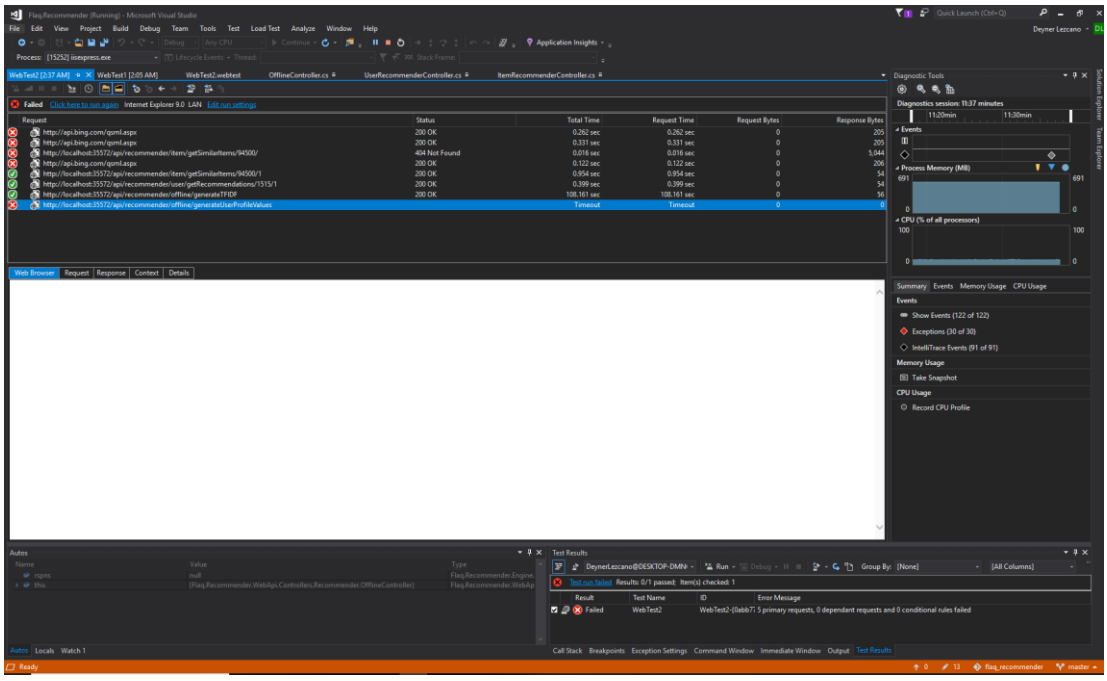


Figura 30. Test web del agente de recomendación - Primera Parte

Después de realizar las modificaciones en la implementación de los algoritmos de recomendación y de generación de perfil de usuario, se realizaron diferentes test web, para lo cual el problema de TIMEOUT había sido resuelto.

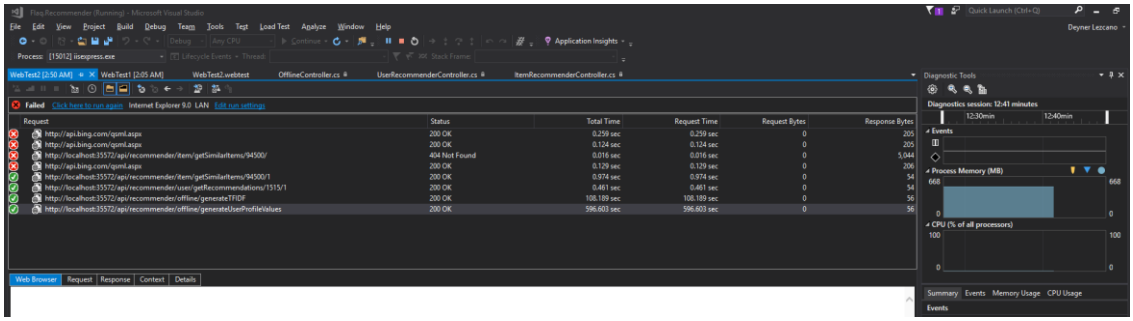


Figura 31. Test web del agente de recomendación - Segunda Parte

4.6. Implantación

En la fase de implantación se explicará cual es el plan de despliegue del agente de recomendación.

4.6.1. Despliegue de la Aplicación

Para el despliegue de la aplicación se hace uso de un servidor IIS – Windows 2008, un motor de base de datos Mongo DB, el agente de recomendación será desplegado en esta plataforma, es importante indicar que este servidor mantendrá comunicación constante con el servidor IIS del sitio Alternate Earth el cual hace uso de una base de datos de SQL Server.

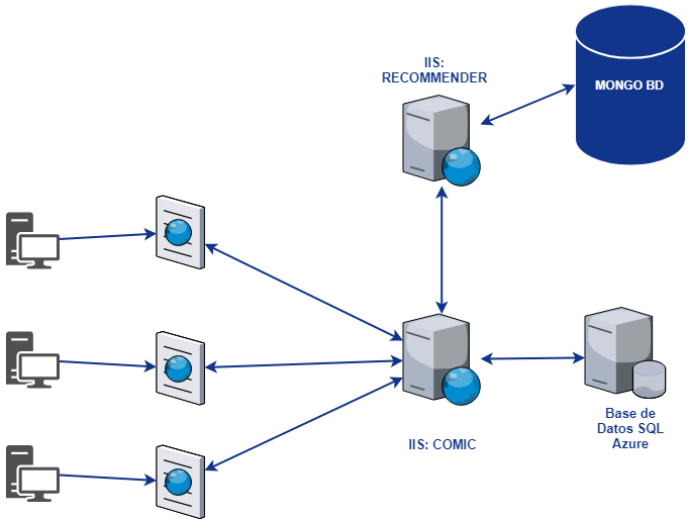


Figura 32. Diagrama de despliegue

CAPÍTULO 5. METODOLOGÍA

5.1. Diseño de Investigación

Investigación pre experimental, porque se pretende investigar la forma en que el desarrollo de un agente de recomendación basado en filtrado de contenido mejora la experiencia de los usuarios del sitio Alternate Earths (AE).

Se cuenta con un grupo G_1, la variable a evaluar "X" y dos observaciones O_1 y O_2.

G_1	X	O_1
		O_2

Dónde:

G_1: es grupo nro. 01

X: es el tratamiento

O_1: Observación antes de

O_2: Observación después de

5.2. Unidad de Estudio

Usuario del sitio Alternate Earths.

5.3. Población

La población considerada para la presente investigación comprende a todos los usuarios registrados en el sitio Alternate Earths, los cuales a la fecha 13 de enero del 2018 suman 163 usuarios activos en la plataforma.

5.4. Muestra

Para determinar la cantidad de usuarios que formarán parte de la muestra de la presente investigación se utilizó la siguiente formula:

$$n = \frac{N \times Z_a^2 \times p \times q}{d^2 \times (N-1) + Z_a^2 \times p \times q}$$

Aplicando los valores correspondientes en la formula se obtiene un total de **127 usuarios** como muestra.

VARIABLE	DESCRIPCIÓN DE VARIABLE	VALOR
n	Número total de la muestra	**127**
N	Tamaño de la población	163
p	Probabilidad de éxito | Proporción de éxito	0.50
q	Probabilidad de fracaso	0.50
Z	Nivel de Confianza	0.95
d	Error de máximo admisible en términos de proporción	0.002

5.5. Técnicas, instrumentos y procedimientos de recolección de datos

En esta sección se detallará los métodos e instrumentos que se utilizará para recolección de datos.

5.5.1. Para recolectar datos

Observación: está técnica se usa en la evaluación de la variable dependiente e independiente.

Ficha de observación digital: este instrumento se usa para registrar los datos obtenidos del análisis de los indicadores.

VARIABLE	DIMENSIÓN	INDICADOR	TÉCNICA	INSTRUMENTO	PROCEDIMIENTO
EXPERIENCIA DE LOS USUARIOS DE ALTERNATE EARTHS	Nivel de aceptación del sistema	Número de suscriptores Tiempo de permanencia en el sitio.	Observación	Ficha de observación digital	Almacenar los resultados de la interacción del usuario con la plataforma Alternate Earths en la ficha de observación digital.
	Nivel de interés de información	Índice de interés en el sitio.	Observación	Ficha de observación digital	Almacenar los resultados de la interacción del usuario con la plataforma Alternate Earths en la ficha de observación digital.

Tabla 32.Técnicas e instrumentos de la variable dependiente (Elaboración propia)

VARIABLE	DIMENSIÓN	INDICADOR	TÉCNICA	INSTRUMENTO	PROCEDIMIENTO
AGENTE INTELIGENTE DE RECOMENDACIÓN	Funcionalidad	Porcentaje de precisión de las recomendaciones	Observación	Ficha de observación digital	Almacenar los resultados del agente de recomendación en la ficha de observación digital.
		Porcentaje de error de precisión en las recomendaciones			
	Eficiencia	Tiempo de respuesta promedio de los algoritmos.	Observación	Ficha de observación digital	Almacenar los resultados del agente de recomendación en la ficha de observación digital.

Tabla 33. Técnicas e instrumentos de la variable independiente (Elaboración propia)

5.6. Métodos, instrumentos y procedimientos de análisis de datos

VARIABLE	DIMENSIÓN	INDICADOR	MÉTODOS	INSTRUMENTO	PROCEDIMIENTO
EXPERIENCIA DE LOS USUARIOS DE ALTERNATE EARTHS	Nivel de aceptación del sistema	Número de suscriptores	$\boldsymbol{S = NSF - NSI}$ Donde, S: numero de suscripciones NSF: número de suscriptores final NSI: número de suscriptores inicial	Hojas de cálculo	Aplicar la fórmula y analizar los resultados en la hoja de cálculo
		Tiempo de permanencia en el sitio.	$\boldsymbol{TPS = FF - FI}$ Donde: TPS: tiempo de permanencia en el sitio FF: Fecha y hora final de salida del sitio. FI: Fecha y hora inicial de ingreso al sitio. N: número de usuarios		

	Nivel de interés de información	Índice de interés en el sitio.	$$IIR = VA$$ Donde: IIR: índice de interés de las recomendaciones VA: valoración de recomendación (lógica difusa)	Implementación de algoritmo	Aplicar la fórmula y analizar los resultados en la hoja de cálculo

Tabla 34. Método y procedimientos para la variable dependiente (Elaboración propia)

VARIABLE	DIMENSIÓN	INDICADOR	MÉTODOS	INSTRUMENTO	PROCEDIMIENTO
AGENTE INTELIGENTE DE RECOMENDACIÓN	Funcionalidad	Porcentaje de precisión de las recomendaciones	$P = \left(\frac{VP}{VP + FN}\right) * 100$ Dónde, $\boldsymbol{P}$**:** precisión $\boldsymbol{VP}$: verdaderos positivos $\boldsymbol{FN}$**:** falso positivo	Hojas de cálculo	Aplicar la fórmula y analizar los resultados en la hoja de cálculo
		Porcentaje de error de precisión en las recomendaciones	$E = \left(\frac{\sum_{i=1}^{N} \lvert p_i - v_i \rvert}{N}\right) * 100$ Dónde, $\boldsymbol{p_i}$**:** predicción $\boldsymbol{v_i}$: valor de preferencial real $\boldsymbol{N}$: número de predicciones	Hojas de cálculo	Aplicar la fórmula y analizar los resultados en la hoja de cálculo
	Eficiencia	Tiempo de respuesta promedio de los algoritmos.	$t = \frac{N}{TE}$ Dónde: $\boldsymbol{N}$**:** número de predicciones $\boldsymbol{TE}$: tiempo de ejecución	Hojas de cálculo	Aplicar la fórmula y analizar los resultados en la hoja de cálculo

Tabla 35. Método y procedimientos para la variable independiente (Elaboración propia)

CAPÍTULO 6. RESULTADOS

A continuación, se presentarán los resultados obtenidos al desplegar el agente de recomendación en el servidor cuyas características son: Windows 2008 R2 Standard Edition – 64 bit, Memoria RAM 2 GB y procesador Core i7 – 6ta generación.

6.1. Análisis de Indicadores

Para verificar el funcionamiento del agente de recomendación se realizaron pruebas con 127 usuarios como muestra del sitio Alternate Earths, esta muestra estuvo expuesta a los ítems recomendados del agente de recomendación y se almacenó la interacción del usuario con la recomendación y con el sitio dónde se mostraron las recomendaciones, evaluándose los comentarios y las visitas. Para poder realizar el análisis de los indicadores se ha seguido la siguiente matriz de confusión con los posibles casos que se pueden presentar cuando se realiza una recomendación, dependiendo de si el ítem recomendado es o no relevante para el usuario.

	RECOMENDADO	**NO RECOMENDADO**
RELEVANTE	Verdadero positivo (VP)	Falso negativo (FN)
NO RELEVANTE	Falso positivo (FP)	Verdadero negativo (VN)

Tabla 36. Matriz de confusión

6.2. Resultados para los indicadores de las variables independientes

6.2.1. Indicador 01: Porcentaje de precisión de las recomendaciones

La muestra de usuarios para este proyecto de investigación son 127 usuarios, para los cuales el agente de recomendación a generados 5634 recomendaciones y según la matriz de confusión y la toma de información de la interacción de los usuarios con los sitios donde se muestra la recomendación se obtiene la siguiente información (Tabla Nro. 36)

	RECOMENDADO	**NO RECOMENDADO**
RELEVANTE	4056	451
NO RELEVANTE	1014	113

Tabla 37. Matriz de confusión con valores de las recomendaciones de la muestra

Considerando los datos de la Tabla Nro. 37, representamos en la fórmula de datos correspondiente.

$$P = \left(\frac{VP}{VP + FN}\right) * 100$$

$$P = \left(\frac{4056}{4056 + 451}\right) * 100$$

$$P = 89.9\ \%$$

Después de aplicar la fórmula, se puede decir que el porcentaje de precisión de las recomendaciones es 89.9%

6.2.2. Indicador 02: Porcentaje de error de precisión en las recomendaciones

Para calcular el porcentaje de error en la precisión de la recomendación, se tomará como muestras la interacción del usuario con las recomendaciones relevantes, según la muestra de 127 usuarios, existen 4056 interacciones con las recomendaciones, en las cuales se enfrentarán el valor de similitud del coseno con el valor difuso según la interacción del usuario / recomendación. (Tabla Nro. 38 - Referencial)

Usuario ID	Nro. Recomendación	Valor Coseno (p_i)	Valor Lógica Difusa v_i
1	1	0.703	0.553
1	2	0.719	0.521
1	3	0.715	0.539
1	4	0.735	0.715
1	5	0.692	0.641
1	6	0.754	0.592
1	7	0.712	0.591
1	8	0.729	0.625
1	9	0.782	0.704
1	10	0.699	0.597
1	11	0.719	0.569
1	12	0.761	0.685
1	13	0.750	0.692

Tabla 38. Descripción del segundo indicador de la variable independiente

Considerando los datos de la Tabla Nro. 38 (Ver Anexo – Información completa), representamos en la fórmula de datos correspondiente.

$$E = \left(\frac{\sum_{i=1}^{N} |p_i - v_i|}{N}\right) * 100$$

$$E = \left(\frac{\sum_{i=1}^{N} 405.97|}{4056}\right) * 100$$

$$E = 10.01\%$$

Después de aplicar la formula, se obtiene que el porcentaje de error de las recomendaciones es 10.01%.

6.2.3. Indicador 03: Tiempo de respuesta promedio de los algoritmos

Para medir la eficiencia en tiempo de un sistema de recomendación, este se calcula dividiendo el número de recomendaciones que se calculan por el tiempo que se consume al calcularlas, para lo cual se ha capturado el valor en microsegundos por cada ítem recomendado, para luego aplicar la formula correspondiente al indicador 3 y conocer el tiempo respuesta promedio del algoritmo que recomienda los ítems según la similitud del coseno.

Nro. Recomendación	Tiempo Ejecución
1	00:00:21
2	00:00:20
3	00:00:09
4	00:00:09
5	00:00:22
6	00:00:02
7	00:00:12
8	00:00:23
9	00:00:11

Tabla 39. Descripción del tercer indicador de la variable independiente

Considerando los datos de la Tabla Nro. 39 - Referencial, representamos en la fórmula de datos correspondiente.

$$t = \frac{N}{TE}$$

$$t = \frac{5634}{4\ horas\ 43\ minutos\ 21\ segundos}$$

$$t = 20\ minutos\ 12\ segundos$$

Según la fórmula aplicada el tiempo total para generar todas las recomendaciones ha sido de 20 minutos con 43 segundos, además por promedio simple podemos decir que el tiempo promedio para procesar una recomendación es de 3 segundos.

6.3. Resultados para los indicadores de las variables dependientes

6.3.1. Indicador 01: Número de suscriptores

6.3.1.1. Definición de variables

$\boldsymbol{PTS_a}$: Número de suscriptores antes del agente de recomendación.

$\boldsymbol{PTS_d}$: Número de suscriptores después del agente de recomendación.

6.3.1.2. Hipótesis Estadística

$\boldsymbol{Hipótesis\ H_0}$: El número de suscriptores antes del agente de recomendación es mayor e igual que el número de suscriptores después del agente de recomendación.

$$H_0 = PTS_a - PTS_d \leq 0$$

$\boldsymbol{Hipótesis\ H_1}$**:** El número de suscriptores antes del agente de recomendación es menor que el número de suscriptores después del agente de recomendación

$$H_1 = PTS_a - PTS_d > 0$$

6.3.1.3. Nivel de Significancia

Se define el margen de error, con una confiabilidad del 95%

Usando un nivel de significancia de $\propto = \mathbf{0.05}$ del 5%. Por lo tanto, el de confianza es de $\mathbf{1} - \propto = \mathbf{0.95}$ será del 95%

6.3.1.4. Estadígrafo de contraste

Puesto que la muestra es $n = 31$, usaremos el Z críticos debido a que nuestra muestra es mayor que 30. Aplicaremos la siguiente fórmula.

$$\overline{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i}}{\boldsymbol{n}}$$

$$\boldsymbol{\sigma^2} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i} - \overline{\boldsymbol{X}}}{\boldsymbol{n}}$$

$$\boldsymbol{Z_c} = \frac{\overline{\boldsymbol{X_A}} - \overline{\boldsymbol{X_D}} + \boldsymbol{X_A} - \boldsymbol{X_D}}{\sqrt{\left(\frac{{\boldsymbol{\sigma_A}}^2}{\boldsymbol{n_A}} + \frac{{\boldsymbol{\sigma_D}}^2}{\boldsymbol{n_D}}\right)}}$$

Según la tabla donde se han realizados los cálculos recolectados del índice de interés en el sitio tanto del antes y después, se procede a calcular Z:

$$\boldsymbol{Z_c} = \frac{\overline{\boldsymbol{TR_a}} - \overline{\boldsymbol{TR_s}}}{\sqrt{\left(\frac{{\boldsymbol{\sigma_a}}^2}{\boldsymbol{n_a}} + \frac{{\boldsymbol{\sigma_s}}^2}{\boldsymbol{n_s}}\right)}}$$

$$\boldsymbol{Z_c} = \frac{\mathbf{5.258 - 6.903}}{\sqrt{\mathbf{0.233 + 0.279}}}$$

$$Z_c = -3.209$$

Tabla 40. Descripción del primer indicador de variable dependiente

TPS_a	TPS_d	$TPS_a - \overline{TPS_a}$	$TPS_d - \overline{TPS}$	$(TPS_a - \overline{TPS_a})^2$	$(TPS_d - \overline{TPS_d})^2$
11	12	5.741935484	5.096774194	32.9698231	25.97710718
2	9	-3.258064516	2.096774194	10.61498439	4.396462019
4	7	-1.258064516	0.096774194	1.582726327	0.009365245
6	4	0.741935484	-2.903225806	0.550468262	8.428720083
4	9	-1.258064516	2.096774194	1.582726327	4.396462019
5	6	-0.258064516	-0.903225806	0.066597294	0.815816857
4	8	-1.258064516	1.096774194	1.582726327	1.202913632
7	5	1.741935484	-1.903225806	3.03433923	3.62226847
10	3	4.741935484	-3.903225806	22.48595213	15.2351717
6	3	0.741935484	-3.903225806	0.550468262	15.2351717
1	9	-4.258064516	2.096774194	18.13111342	4.396462019
5	7	-0.258064516	0.096774194	0.066597294	0.009365245
7	9	1.741935484	2.096774194	3.03433923	4.396462019
2	9	-3.258064516	2.096774194	10.61498439	4.396462019
6	12	0.741935484	5.096774194	0.550468262	25.97710718
4	4	-1.258064516	-2.903225806	1.582726327	8.428720083
3	8	-2.258064516	1.096774194	5.098855359	1.202913632
6	7	0.741935484	0.096774194	0.550468262	0.009365245
8	3	2.741935484	-3.903225806	7.518210198	15.2351717
4	12	-1.258064516	5.096774194	1.582726327	25.97710718
2	8	-3.258064516	1.096774194	10.61498439	1.202913632
1	5	-4.258064516	-1.903225806	18.13111342	3.62226847
1	6	-4.258064516	-0.903225806	18.13111342	0.815816857
7	5	1.741935484	-1.903225806	3.03433923	3.62226847
9	11	3.741935484	4.096774194	14.00208117	16.78355879
8	3	2.741935484	-3.903225806	7.518210198	15.2351717
5	6	-0.258064516	-0.903225806	0.066597294	0.815816857
6	11	0.741935484	4.096774194	0.550468262	16.78355879
6	8	0.741935484	1.096774194	0.550468262	1.202913632
3	3	-2.258064516	-3.903225806	5.098855359	15.2351717
10	2	4.741935484	-4.903225806	22.48595213	24.04162331
163	214		**Suma**	**223.9354839**	**268.7096774**
5.258064516	6.90322581				

6.3.1.5. Región Crítica

Para $\alpha = 0.05$ encontramos que $\boldsymbol{Z_\alpha = 0.05}$. Entonces la región crítica de la prueba es

$\boldsymbol{Z_{tab} = -1.64}$.

6.3.1.6. Conclusión

Puesto que $\boldsymbol{Z_c = -3.209}$ calculado, es menor que $\boldsymbol{Z_{tab} = -1.64}$ y estando este valor dentro de la región de rechazo de la hipótesis nula; entonces se rechaza la H_0 y por consiguiente se acepta la H_1. Se concluye que el número de suscriptores es mayor con el agente de recomendación con un nivel de error del 5% y un nivel de confianza del 95%.

6.3.1.7. Discusión de Resultados

Comparación del indicador del número de suscriptores antes del agente de recomendación $\boldsymbol{NS_a}$ y número de suscriptores después del agente de recomendación $\boldsymbol{NS_d}$.

Tabla 41. Comparación del Indicador NS_a y NS_d

NSA	NSD	Incremento%
11.04%	**23.83**	**12.79%**

Como se puede observar que el indicador número de suscriptores es de 0.2388, mientras que antes era 0.1104 lo que representa un aumento del 12.79%.

6.3.2. Indicador 02: Tiempo de Permanencia en el sitio

Para determinar el tiempo de permanencia del sitio se aplicará la diferencia entre la fecha de salida y la fecha de ingreso al sitio por usuarios de la muestra, obteniendo el tiempo de permanencia del sitio. Así mismo, por existir una variación de resultados debido a la interacción del usuario antes y después del agente de recomendación, se realizará una comparación del tiempo de permanencia entre ANTES Y DESPUES del despliegue del agente de recomendación.

6.3.2.1. Definición de variables

$\boldsymbol{TPS_a}$: Tiempo de permanencia en el sitio antes del agente de recomendación.

$\boldsymbol{TPS_d}$: Tiempo de permanencia en el sitio después del agente de recomendación.

6.3.2.2. Hipótesis Estadística

$\boldsymbol{Hipótesis\ H_0}$: El tiempo de permanencia en el sitio antes del agente de recomendación es mayor e igual que el tiempo de permanencia en el sitio después del agente de recomendación.

$$H_0 = PTS_a - PTS_d \leq 0$$

$\boldsymbol{Hipótesis\ H_1}$**:** El tiempo de permanencia en el sitio antes del agente de recomendación es menor que tiempo de permanencia en el sitio después del agente de recomendación

$$H_1 = PTS_a - PTS_d > 0$$

6.3.2.3. Nivel de Significancia

Se define el margen de error, con una confiabilidad del 95%

Usando un nivel de significancia de $\propto = \mathbf{0.05}$ del 5%. Por lo tanto, el de confianza es de $\mathbf{1} - \propto = \mathbf{0.95}$ será del 95%

6.3.2.4. Estadígrafo de contraste

Puesto que la muestra es $n = 127$, usaremos el Z crítico debido a que nuestra muestra es mayor que 30. Aplicaremos la siguiente fórmula.

$$\overline{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i}}{\boldsymbol{n}}$$

$$\boldsymbol{\sigma^2} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i} - \overline{\boldsymbol{X}}}{\boldsymbol{n}}$$

$$Z_c = \frac{\overline{X_A} - \overline{X_D} + X_A - X_D}{\sqrt{\left(\frac{{\sigma_A}^2}{n_A} + \frac{{\sigma_D}^2}{n_D}\right)}}$$

Según la tabla N° 40, donde se han realizados los cálculos recolectados del tiempo de permanencia en el sitio tanto del antes y después, se procede a calcular Z:

$$Z_c = \frac{\overline{TR_a} - \overline{TR_s}}{\sqrt{\left(\frac{{\sigma_a}^2}{n_a} + \frac{{\sigma_s}^2}{n_s}\right)}}$$

$$Z_c = \frac{981639 - 2619906}{\sqrt{\left(\frac{0.0022}{127} + \frac{0.0233}{127}\right)}}$$

$$Z_c = \frac{0.0895 - 0.2388}{\sqrt{0.0000175 + 0.0002}}$$

$$Z_c = -10.54$$

Tabla 42. Descripción del segundo indicador de la variable dependiente.

N°	TPS_a	TPS_d	$TPS_a - \overline{TPS_a}$	$TPS_d - \overline{TPS}$	$\left(TPS_a - \overline{TPS_a}\right)^2$	$\left(TPS_d - \overline{TPS_d}\right)^2$
1	0.0714	0.0725	-0.0181	-0.1662	0.0003	0.0276
2	0.0451	0.0043	-0.0443	-0.2345	0.0020	0.0550
3	0.0368	0.1684	-0.0527	-0.0704	0.0028	0.0050
4	0.0645	0.3164	-0.0250	0.0776	0.0006	0.0060
5	0.0702	0.1829	-0.0192	-0.0559	0.0004	0.0031
6	0.1681	0.2207	0.0786	-0.0181	0.0062	0.0003
7	0.1481	0.3568	0.0586	0.1180	0.0034	0.0139
8	0.1210	0.0281	0.0315	-0.2107	0.0010	0.0444
9	0.1267	0.0102	0.0373	-0.2285	0.0014	0.0522
10	0.1387	0.0367	0.0492	-0.2020	0.0024	0.0408
11	0.1038	0.1604	0.0143	-0.0784	0.0002	0.0061
12	0.1598	0.2309	0.0703	-0.0079	0.0049	0.0001
13	0.1349	0.1324	0.0454	-0.1064	0.0021	0.0113
14	0.0604	0.2817	-0.0291	0.0430	0.0008	0.0018
15	0.1335	0.4864	0.0440	0.2477	0.0019	0.0613
16	0.0856	0.3418	-0.0039	0.1031	0.0000	0.0106
17	0.0844	0.0041	-0.0051	-0.2347	0.0000	0.0551
18	0.0121	0.0300	-0.0774	-0.2088	0.0060	0.0436
19	0.0561	0.0923	-0.0333	-0.1465	0.0011	0.0215
20	0.0070	0.0110	-0.0825	-0.2278	0.0068	0.0519
21	0.1502	0.4248	0.0607	0.1861	0.0037	0.0346
22	0.0609	0.2846	-0.0286	0.0459	0.0008	0.0021
23	0.0579	0.4338	-0.0316	0.1950	0.0010	0.0380
24	0.0494	0.0692	-0.0401	-0.1696	0.0016	0.0288
25	0.0817	0.3079	-0.0077	0.0691	0.0001	0.0048
26	0.0052	0.0813	-0.0842	-0.1575	0.0071	0.0248
27	0.1356	0.4571	0.0461	0.2183	0.0021	0.0477
28	0.1070	0.3450	0.0176	0.1062	0.0003	0.0113
29	0.0875	0.4176	-0.0020	0.1788	0.0000	0.0320
30	0.1321	0.4050	0.0426	0.1662	0.0018	0.0276
31	0.0123	0.3982	-0.0772	0.1595	0.0060	0.0254
32	0.0413	0.1591	-0.0482	-0.0797	0.0023	0.0063
33	0.1445	0.4900	0.0550	0.2512	0.0030	0.0631
34	0.0876	0.3384	-0.0018	0.0997	0.0000	0.0099
35	0.1442	0.0217	0.0547	-0.2170	0.0030	0.0471
36	0.0473	0.3229	-0.0421	0.0841	0.0018	0.0071
37	0.0636	0.3233	-0.0258	0.0845	0.0007	0.0071
38	0.1011	0.4175	0.0117	0.1787	0.0001	0.0319
39	0.0012	0.1083	-0.0883	-0.1305	0.0078	0.0170

40	0.0255	0.2195	-0.0639	-0.0193	0.0041	0.0004
41	0.0981	0.4649	0.0087	0.2262	0.0001	0.0512
42	0.0087	0.2572	-0.0807	0.0185	0.0065	0.0003
43	0.0125	0.3899	-0.0770	0.1511	0.0059	0.0228
44	0.0589	0.0661	-0.0306	-0.1727	0.0009	0.0298
45	0.0914	0.4595	0.0019	0.2208	0.0000	0.0487
46	0.0791	0.4548	-0.0104	0.2160	0.0001	0.0467
47	0.0701	0.0106	-0.0194	-0.2281	0.0004	0.0520
48	0.0556	0.2072	-0.0338	-0.0316	0.0011	0.0010
49	0.0816	0.2326	-0.0079	-0.0061	0.0001	0.0000
50	0.1372	0.0170	0.0477	-0.2217	0.0023	0.0492
51	0.0822	0.1239	-0.0072	-0.1149	0.0001	0.0132
52	0.1487	0.0405	0.0592	-0.1983	0.0035	0.0393
53	0.0772	0.4296	-0.0122	0.1909	0.0002	0.0364
54	0.0691	0.0082	-0.0203	-0.2306	0.0004	0.0532
55	0.0178	0.4194	-0.0716	0.1807	0.0051	0.0326
56	0.0455	0.1048	-0.0439	-0.1340	0.0019	0.0180
57	0.0056	0.2712	-0.0838	0.0324	0.0070	0.0011
58	0.0570	0.4164	-0.0324	0.1776	0.0011	0.0315
59	0.0994	0.4606	0.0100	0.2218	0.0001	0.0492
60	0.0767	0.4874	-0.0128	0.2487	0.0002	0.0618
61	0.0839	0.1713	-0.0055	-0.0674	0.0000	0.0045
62	0.1440	0.3025	0.0545	0.0637	0.0030	0.0041
63	0.1176	0.4700	0.0281	0.2313	0.0008	0.0535
64	0.1463	0.3866	0.0568	0.1478	0.0032	0.0219
65	0.0707	0.3387	-0.0187	0.0999	0.0004	0.0100
66	0.0953	0.3368	0.0059	0.0981	0.0000	0.0096
67	0.1219	0.4099	0.0324	0.1711	0.0011	0.0293
68	0.1026	0.4600	0.0132	0.2212	0.0002	0.0489
69	0.1281	0.3457	0.0386	0.1069	0.0015	0.0114
70	0.1280	0.0158	0.0386	-0.2230	0.0015	0.0497
71	0.1114	0.0705	0.0220	-0.1683	0.0005	0.0283
72	0.0208	0.4428	-0.0687	0.2041	0.0047	0.0416
73	0.0982	0.2428	0.0088	0.0040	0.0001	0.0000
74	0.0460	0.2595	-0.0435	0.0208	0.0019	0.0004
75	0.1252	0.4016	0.0357	0.1628	0.0013	0.0265
76	0.0140	0.0659	-0.0755	-0.1729	0.0057	0.0299
77	0.0921	0.4426	0.0026	0.2039	0.0000	0.0416
78	0.1023	0.0929	0.0129	-0.1459	0.0002	0.0213
79	0.1639	0.3846	0.0744	0.1459	0.0055	0.0213
80	0.0587	0.0528	-0.0308	-0.1860	0.0009	0.0346

81	0.1528	0.3542	0.0633	0.1155	0.0040	0.0133
82	0.1720	0.1496	0.0826	-0.0892	0.0068	0.0080
83	0.0894	0.3280	-0.0001	0.0893	0.0000	0.0080
84	0.0432	0.1632	-0.0462	-0.0756	0.0021	0.0057
85	0.0691	0.1676	-0.0204	-0.0711	0.0004	0.0051
86	0.1554	0.2949	0.0659	0.0562	0.0043	0.0032
87	0.0853	0.3682	-0.0042	0.1294	0.0000	0.0167
88	0.0652	0.1042	-0.0243	-0.1346	0.0006	0.0181
89	0.0741	0.0522	-0.0154	-0.1865	0.0002	0.0348
90	0.0824	0.1429	-0.0071	-0.0959	0.0001	0.0092
91	0.0509	0.1774	-0.0386	-0.0614	0.0015	0.0038
92	0.0228	0.0607	-0.0667	-0.1781	0.0044	0.0317
93	0.1669	0.3495	0.0775	0.1107	0.0060	0.0123
94	0.1503	0.1717	0.0608	-0.0671	0.0037	0.0045
95	0.0138	0.1233	-0.0757	-0.1154	0.0057	0.0133
96	0.1010	0.0695	0.0115	-0.1693	0.0001	0.0287
97	0.0560	0.0155	-0.0334	-0.2233	0.0011	0.0499
98	0.1240	0.4107	0.0345	0.1720	0.0012	0.0296
99	0.1487	0.3477	0.0593	0.1090	0.0035	0.0119
100	0.1191	0.1861	0.0297	-0.0527	0.0009	0.0028
101	0.0743	0.0606	-0.0152	-0.1781	0.0002	0.0317
102	0.1093	0.2928	0.0198	0.0540	0.0004	0.0029
103	0.1510	0.3532	0.0615	0.1144	0.0038	0.0131
104	0.1531	0.3703	0.0636	0.1315	0.0040	0.0173
105	0.1509	0.4033	0.0614	0.1646	0.0038	0.0271
106	0.0663	0.1446	-0.0232	-0.0942	0.0005	0.0089
107	0.0763	0.1242	-0.0132	-0.1145	0.0002	0.0131
108	0.0546	0.3594	-0.0349	0.1206	0.0012	0.0145
109	0.0378	0.2524	-0.0517	0.0137	0.0027	0.0002
110	0.1660	0.1130	0.0766	-0.1258	0.0059	0.0158
111	0.1532	0.0264	0.0637	-0.2124	0.0041	0.0451
112	0.1201	0.1320	0.0307	-0.1067	0.0009	0.0114
113	0.0574	0.0127	-0.0321	-0.2261	0.0010	0.0511
114	0.0815	0.2100	-0.0079	-0.0288	0.0001	0.0008
115	0.0500	0.0581	-0.0395	-0.1806	0.0016	0.0326
116	0.1553	0.2214	0.0658	-0.0174	0.0043	0.0003
117	0.1475	0.2898	0.0580	0.0510	0.0034	0.0026
118	0.1570	0.0399	0.0675	-0.1988	0.0046	0.0395
119	0.0193	0.4329	-0.0702	0.1942	0.0049	0.0377
120	0.1286	0.1563	0.0391	-0.0825	0.0015	0.0068
121	0.0048	0.3440	-0.0847	0.1052	0.0072	0.0111

122	0.1428	0.0543	0.0534	-0.1844	0.0028	0.0340
123	0.1533	0.4889	0.0638	0.2501	0.0041	0.0626
124	0.0523	0.2999	-0.0372	0.0611	0.0014	0.0037
125	0.1504	0.4521	0.0610	0.2134	0.0037	0.0455
126	0.0915	0.1360	0.0020	-0.1028	0.0000	0.0106
127	0.0071	0.2493	-0.0824	0.0105	0.0068	0.0001
Promedio	**0.0895**	**0.2388**			**0.0022**	**0.0233**
Sumatoria	**11.3616**	**30.3230**			**0.2829**	**2.9560**

6.3.2.5. Región Crítica

Para $\alpha = 0.05$ encontramos que $\boldsymbol{Z_{\alpha} = 0.05}$. Entonces la región crítica de la prueba es

$\boldsymbol{Z_{tab} = -1.64}$.

6.3.2.6. Conclusión

Puesto que $\boldsymbol{Z_c = -10.54}$ calculado, es menor que $\boldsymbol{Z_{tab} = -1.64}$ y estando este valor dentro de la región de rechazo de la hipótesis nula; entonces se rechaza la H_0 y por consiguiente se acepta la H_1. Se concluye que el tiempo de permanencia en el sitio es mayor con el agente de recomendación con un nivel de error del 5% y un nivel de confianza del 95%.

6.3.2.7. Discusión de Resultados

Comparación del indicador tiempo de permanencia en el sitio antes del agente de recomendación $\boldsymbol{PTS_a}$ y tiempo de permanencia en el sitio después del agente de recomendación $\boldsymbol{PTS_d}$.

Tabla 43. Comparación del Indicador TPS_a y TPS_d

TPSa		TPSd		Incremento	
	%		%		%
0.0895	**100%**	**0.2388**	**266.81**	**0.149**	**166.81**

Como se puede observar que el indicador tiempo promedio de permanencia en el sitio con el agente de recomendación es de 0.2388, mientras que antes era 0.0895 lo que representa un aumento del 166.81%.

6.3.3. Indicador 03: Índice de interés del sitio

Para obtener el índice de interés de los usuarios en el sitio, se tomará en cuenta las valoraciones de ellos tanto ANTES y DESPUÉS del agente de recomendación en el sitio, a través de la lógica difusa se obtiene un valor de aceptación de una escala del 0 al 1.

6.3.3.1. Definición de variables

$\boldsymbol{IIS_a}$: Índice de interés del sitio antes del agente de recomendación.

$\boldsymbol{IIS_d}$: Índice de interés del sitio después del agente de recomendación.

6.3.3.2. Hipótesis Estadística

$\boldsymbol{Hipótesis\ H_0}$: El índice de interés en el sitio antes del agente de recomendación es mayor e igual que el índice de interés en el sitio después del agente de recomendación.

$$H_0 = IIS_a - IIS_d \leq 0$$

$\boldsymbol{Hipótesis\ H_1}$**:** El índice de interés en el sitio antes del agente de recomendación es menor que el índice de interés en el sitio después del agente de recomendación

$$H_1 = IIS_a - IIS_d > 0$$

6.3.3.3. Nivel de Significancia

Se define el margen de error, con una confiabilidad del 95%

Usando un nivel de significancia de $\propto = \mathbf{0.05}$ del 5%. Por lo tanto, el de confianza es de $\mathbf{1} - \propto = \mathbf{0.95}$ será del 95%

6.3.3.4. Estadígrafo de contraste

Puesto que la muestra es $n = 127$, usaremos el Z crítico debido a que nuestra muestra es mayor que 30. Aplicaremos la siguiente fórmula.

$$\overline{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i}}{\boldsymbol{n}}$$

$$\boldsymbol{\sigma^2} = \frac{\sum_{i=1}^{n} \boldsymbol{X_i} - \overline{\boldsymbol{X}}}{\boldsymbol{n}}$$

$$\boldsymbol{Z_c} = \frac{\overline{\boldsymbol{X_A}} - \overline{\boldsymbol{X_D}} + \boldsymbol{X_A} - \boldsymbol{X_D}}{\sqrt{\left(\frac{\boldsymbol{\sigma_A}^2}{\boldsymbol{n_A}} + \frac{\boldsymbol{\sigma_D}^2}{\boldsymbol{n_D}}\right)}}$$

Según la tabla N° 43, donde se han realizados los cálculos recolectados del índice de interés en el sitio tanto del antes y después, se procede a calcular Z:

$$\boldsymbol{Z_c} = \frac{\overline{\boldsymbol{TR_a}} - \overline{\boldsymbol{TR_s}}}{\sqrt{\left(\frac{\boldsymbol{\sigma_a}^2}{\boldsymbol{n_a}} + \frac{\boldsymbol{\sigma_s}^2}{\boldsymbol{n_s}}\right)}}$$

$$Z_c = \frac{0.483 - 0.636}{\sqrt{\left(\frac{0.082}{127} + \frac{0.005}{127}\right)}}$$

$$Z_c = \frac{-0.153}{\sqrt{0.001 + 0.0000404}}$$

$$Z_c = -5.86$$

Tabla 44. Descripción del tercer indicador de la variable dependiente.

Usuario	IIS_a	IIS_d	$IIS_a - \overline{IIS_a}$	$IIS_d - \overline{IIS_d}$	$(IIS_a - \overline{IIS_a})^2$	$(IIS_d - \overline{IIS_d})^2$
1	0.308	0.553	-0.174	-0.083	0.030	0.007
2	0.900	0.521	0.417	-0.115	0.174	0.013
3	0.597	0.539	0.114	-0.097	0.013	0.009
4	0.647	0.715	0.164	0.078	0.027	0.006
5	0.787	0.641	0.304	0.005	0.092	0.000
6	0.734	0.592	0.252	-0.045	0.063	0.002
7	0.800	0.591	0.317	-0.045	0.101	0.002
8	0.499	0.625	0.016	-0.011	0.000	0.000
9	0.245	0.704	-0.238	0.068	0.056	0.005
10	0.829	0.597	0.346	-0.039	0.120	0.002
11	0.754	0.569	0.271	-0.067	0.073	0.005
12	0.474	0.685	-0.009	0.049	0.000	0.002
13	0.886	0.692	0.403	0.056	0.163	0.003
14	0.844	0.584	0.362	-0.053	0.131	0.003
15	0.471	0.656	-0.011	0.020	0.000	0.000
16	0.096	0.558	-0.387	-0.078	0.150	0.006
17	0.827	0.614	0.344	-0.022	0.118	0.000
18	0.022	0.642	-0.461	0.006	0.213	0.000
19	0.074	0.722	-0.409	0.086	0.167	0.007
20	0.521	0.747	0.038	0.111	0.001	0.012
21	0.458	0.543	-0.025	-0.093	0.001	0.009
22	0.855	0.719	0.372	0.083	0.138	0.007
23	0.887	0.568	0.405	-0.068	0.164	0.005
24	0.006	0.715	-0.477	0.079	0.227	0.006
25	0.476	0.776	-0.007	0.140	0.000	0.020
26	0.906	0.750	0.423	0.114	0.179	0.013
27	0.015	0.795	-0.468	0.158	0.219	0.025
28	0.218	0.661	-0.265	0.025	0.070	0.001
29	0.809	0.657	0.326	0.020	0.106	0.000
30	0.806	0.785	0.323	0.149	0.104	0.022

31	0.920	0.748	0.437	0.111	0.191	0.012
32	0.484	0.604	0.002	-0.032	0.000	0.001
33	0.071	0.629	-0.411	-0.007	0.169	0.000
34	0.181	0.527	-0.302	-0.109	0.091	0.012
35	0.276	0.674	-0.207	0.037	0.043	0.001
36	0.750	0.528	0.268	-0.108	0.072	0.012
37	0.581	0.549	0.098	-0.087	0.010	0.008
38	0.053	0.586	-0.430	-0.050	0.185	0.002
39	0.681	0.776	0.198	0.140	0.039	0.020
40	0.238	0.628	-0.245	-0.008	0.060	0.000
41	0.489	0.563	0.006	-0.073	0.000	0.005
42	0.114	0.737	-0.369	0.100	0.136	0.010
43	0.848	0.683	0.365	0.047	0.133	0.002
44	0.775	0.657	0.292	0.020	0.085	0.000
45	0.524	0.667	0.041	0.031	0.002	0.001
46	0.758	0.688	0.276	0.052	0.076	0.003
47	0.928	0.511	0.445	-0.126	0.198	0.016
48	0.790	0.604	0.307	-0.032	0.094	0.001
49	0.321	0.490	-0.162	-0.146	0.026	0.021
50	0.958	0.621	0.476	-0.016	0.226	0.000
51	0.178	0.644	-0.305	0.008	0.093	0.000
52	0.607	0.658	0.124	0.022	0.015	0.000
53	0.545	0.697	0.062	0.060	0.004	0.004
54	0.039	0.558	-0.444	-0.078	0.197	0.006
55	0.732	0.558	0.249	-0.078	0.062	0.006
56	0.418	0.622	-0.065	-0.014	0.004	0.000
57	0.280	0.689	-0.203	0.053	0.041	0.003
58	0.356	0.678	-0.127	0.042	0.016	0.002
59	0.855	0.677	0.372	0.041	0.138	0.002
60	0.465	0.561	-0.018	-0.076	0.000	0.006
61	0.930	0.665	0.447	0.028	0.200	0.001
62	0.677	0.671	0.194	0.035	0.038	0.001
63	0.271	0.485	-0.212	-0.151	0.045	0.023
64	0.267	0.598	-0.216	-0.038	0.047	0.001
65	0.072	0.561	-0.410	-0.075	0.168	0.006
66	0.561	0.648	0.078	0.012	0.006	0.000
67	0.558	0.587	0.075	-0.049	0.006	0.002
68	0.661	0.593	0.178	-0.043	0.032	0.002
69	0.849	0.588	0.367	-0.048	0.134	0.002
70	0.317	0.616	-0.166	-0.020	0.028	0.000
71	0.141	0.612	-0.342	-0.024	0.117	0.001

72	0.502	0.690	0.019	0.053	0.000	0.003
73	0.376	0.654	-0.107	0.018	0.011	0.000
74	0.169	0.682	-0.314	0.046	0.099	0.002
75	0.540	0.748	0.057	0.112	0.003	0.012
76	0.284	0.467	-0.199	-0.169	0.040	0.029
77	0.146	0.605	-0.337	-0.032	0.114	0.001
78	0.716	0.630	0.233	-0.006	0.054	0.000
79	0.304	0.709	-0.178	0.073	0.032	0.005
80	0.560	0.765	0.077	0.129	0.006	0.017
81	0.157	0.657	-0.326	0.020	0.106	0.000
82	0.066	0.543	-0.417	-0.093	0.174	0.009
83	0.492	0.642	0.010	0.006	0.000	0.000
84	0.493	0.532	0.011	-0.105	0.000	0.011
85	0.047	0.660	-0.436	0.024	0.190	0.001
86	0.492	0.704	0.010	0.068	0.000	0.005
87	0.823	0.727	0.340	0.091	0.116	0.008
88	0.563	0.669	0.081	0.032	0.006	0.001
89	0.691	0.653	0.208	0.017	0.043	0.000
90	0.037	0.607	-0.445	-0.029	0.198	0.001
91	0.568	0.545	0.086	-0.092	0.007	0.008
92	0.183	0.584	-0.300	-0.052	0.090	0.003
93	0.745	0.583	0.262	-0.053	0.069	0.003
94	0.097	0.528	-0.386	-0.108	0.149	0.012
95	0.244	0.541	-0.239	-0.095	0.057	0.009
96	0.070	0.674	-0.413	0.038	0.171	0.001
97	0.346	0.620	-0.137	-0.016	0.019	0.000
98	0.857	0.586	0.374	-0.050	0.140	0.003
99	0.101	0.668	-0.382	0.032	0.146	0.001
100	0.893	0.621	0.410	-0.015	0.168	0.000
101	0.147	0.493	-0.336	-0.143	0.113	0.020
102	0.824	0.716	0.341	0.080	0.116	0.006
103	0.753	0.658	0.271	0.022	0.073	0.000
104	0.893	0.670	0.411	0.034	0.169	0.001
105	0.140	0.715	-0.343	0.079	0.118	0.006
106	0.623	0.668	0.141	0.031	0.020	0.001
107	0.356	0.704	-0.127	0.068	0.016	0.005
108	0.234	0.611	-0.249	-0.026	0.062	0.001
109	0.444	0.622	-0.039	-0.014	0.002	0.000
110	0.272	0.765	-0.211	0.129	0.044	0.017
111	0.073	0.687	-0.410	0.051	0.168	0.003
112	0.676	0.646	0.193	0.010	0.037	0.000

113	0.307	0.677	-0.176	0.041	0.031	0.002
114	0.067	0.621	-0.416	-0.016	0.173	0.000
115	0.261	0.623	-0.221	-0.013	0.049	0.000
116	0.741	0.694	0.258	0.058	0.067	0.003
117	0.602	0.629	0.120	-0.007	0.014	0.000
118	0.800	0.702	0.317	0.066	0.101	0.004
119	0.041	0.732	-0.442	0.096	0.196	0.009
120	0.540	0.605	0.057	-0.031	0.003	0.001
121	0.639	0.551	0.156	-0.085	0.024	0.007
122	0.664	0.547	0.181	-0.089	0.033	0.008
123	0.690	0.604	0.207	-0.032	0.043	0.001
124	0.038	0.550	-0.444	-0.086	0.197	0.007
125	0.825	0.749	0.342	0.113	0.117	0.013
126	0.598	0.638	0.115	0.002	0.013	0.000
127	0.237	0.668	-0.246	0.032	0.061	0.001
Sumatoria	**61.316**	**80.794**			**10.398**	**0.652**
Promedio	**0.483**	**0.636**				

6.3.3.5. Región Crítica

Para $\alpha = 0.05$ encontramos que $\boldsymbol{Z_{\alpha} = 0.05}$. Entonces la región crítica de la prueba es

$\boldsymbol{Z_{tab} = -1.64}$.

6.3.3.6. Conclusión

Puesto que $\boldsymbol{Z_c = -5.86}$ calculado, es menor que $\boldsymbol{Z_{tab} = -1.64}$ y estando este valor dentro de la región de rechazo de la hipótesis nula; entonces se rechaza la H_0 y por consiguiente se acepta la H_1. Se concluye que el índice de interés del sitio es mayor con el agente de recomendación con un nivel de error del 5% y un nivel de confianza del 95%.

6.3.3.7. Discusión de Resultados

Comparación del indicador del índice de interés del sitio antes del agente de recomendación $\boldsymbol{IIS_a}$ y el índice de interés en el sitio después del agente de recomendación $\boldsymbol{IIS_d}$.

Importante:

Los perfiles de usuario son generados mediante el uso de lógica difusa. El modelo implementado (basado en el método Mamdani) acepta como datos de entrada a la cantidad de vistas y comentarios que los usuarios hacen en un artículo. Estos datos, forman parte de los datos de entrada tipo "Implícito".

Las vistas en un artículo no necesariamente indican que el articulo sea de gran interés del usuario. Por otro lado, los comentarios hechos por los usuarios sobre un artículo tienen relación sobre el impacto este genera sobre él. Por estas razones, para fines de la presente investigación, se consideró un peso de 0.2 a las vistas y 0.8 a los comentarios, obteniendo como valor resultante, el nivel de interés implícito del usuario a la información. (Youtube, 2018)

Tabla 45. Comparación del Indicador IIS_a y IIS_d

IISa		IISd		Incremento	
	%		%		%
61.316	**100%**	**80.794**	**137.77**	**19.478**	**37.77**

Como se puede observar que el indicador tiempo promedio de permanencia en el sitio con el agente de recomendación es de 0.2388, mientras que antes era 0.0895 lo que representa un aumento del 166.81%.

CAPÍTULO 7. DISCUSIÓN

Esta investigación tiene como propósito determinar cómo afecta la experiencia de los usuarios de Alternate Earths mediante la implementación e implantación de un agente inteligente de recomendación basado en filtrado de contenido. Después de haber procesado los indicadores de la variable independiente, se obtuvo lo siguiente:

Indicador 1: El porcentaje de precisión de las recomendaciones es del 89.9 %, lo que significa que el agente de recomendación brinda recomendaciones de ítems de manera correcta, de un total de 5634 recomendaciones: 5065 son correctas. Por eso, el porcentaje de precisión de las recomendaciones es aceptable para experiencia de los usuarios con el uso del agente de recomendación.

Indicador 2: El porcentaje de error de precisión de las recomendaciones es del 10.1 %, lo que el agente de recomendación brinda recomendaciones no relacionadas a los gustos del usuario, de un total de 5634 recomendaciones: 569 son incorrectas. Aunque el porcentaje de error de precisión de recomendaciones no es muy bajo, es aceptable puesto que muchas de estas recomendaciones no son mostradas al usuario.

Indicador 3: El tiempo de respuesta promedio de los algoritmos es de 20 minutos con 12 segundos para el procesamiento de 5634 recomendaciones entre relevante y no relevantes; a pesar que el tiempo de procesamiento es alto, el tiempo de procesamiento promedio por recomendación es de 3 milisegundos.

Luego de haber procesado los indicadores de la variable dependiente, se obtuvo lo siguiente

Indicador 1: El número de suscriptores según el análisis estadístico rechaza la hipótesis nula y acepta la hipótesis H_1, la cual indica que el número de suscriptores antes del agente de recomendación es menor que el número de suscriptores después del agente de recomendación, incrementando en un 12.79%.

Indicador 2: El tiempo de permanencia en el sitio, según el análisis estadístico rechaza la hipótesis nula y acepta la hipótesis H_1, la cual indica que el tiempo de permanencia en el sitio antes del agente de recomendación es menor que el tiempo de permanencia en el sitio después del agente de recomendación, incrementando en un 166.81%.

Indicador 3: El índice de interés en el sitio, según el análisis estadístico rechaza la hipótesis nula y acepta la hipótesis H_1, la cual indica que el índice de interés en el sitio antes del agente de recomendación es menor que el índice de interés en el sitio después del agente de recomendación, incrementando en un 37.77%

CONCLUSIONES

Al finalizar la investigación de este proyecto, se obtuvieron las siguientes conclusiones:

Se logró mejorar la experiencia de los usuarios de Alternate Earths mediante la implementación e implantación de un agente inteligente de recomendación basado en filtrado de contenido, cumpliendo los siguientes objetivos:

- Se logró determinar el nivel de aceptación de los usuarios hacia el sistema de Alternate Earth, puesto que el porcentaje de suscripciones aumentó en un 12.79% después del agente de recomendación. Además, el tiempo de permanencia en el sitio incrementó en un 166.81%.
- Se logró analizar el tiempo promedio de respuesta de los algoritmos, el cual para generar todas recomendaciones toma 20 minutos, y en promedio por recomendación 3 milisegundos.
- Se logró calcular el índice de interés en el sitio tanto del antes y después, y mediante la estadística se logró determina que incrementó en un 37.7%
- Se logró determinar el porcentaje de precisión de las recomendaciones siendo del 89.9 % y el porcentaje de error de precisión siendo del 10.1%.

RECOMENDACIONES

Con la finalización de esta investigación, se recomienda considerar los siguientes puntos:

- Para disminuir el proceso de inicio en frío del agente de recomendación, se recomienda agregar una sección en la creación de usuarios, donde se pueda recolectar los gustos y /o preferencias previas a la interacción con el sitio el usuario.
- Para poder obtener más información que alimente al agente de recomendación, se puede mejorar la forma en la que se recolecta la interacción implícita del usuario, mediante el análisis de información como: porcentaje de reproducción de videos o porcentaje de lectura de documentos.

REFERENCIAS

Alfredo, Z., Víctor, M., Manuel, P., & Cristobal, R. (2011). *A Hybrid Recommender Method for Learning Objects .* España: International Journal of Computer Applications.

Alpydin, E. (2010). *Introduction to Machine Learning.* Cambridge, Estados Unidos: Massachusetts Institute of Technology.

Bautista, Q. (2012). *Programación Extrema XP.*

Bellare, M., Canetti, R., & Krawczyk, H. (1996). *Keying Hash Functions for Message Authentication.* Springer: Gollmann.

Blázquez, M. (18 de Junio de 2018). *Monografías Electrónicas.* Obtenido de http://mblazquez.es/wp-content/uploads/ebook-mbo-tecnicas-avanzadas-recuperacion-informacion1.pdf

Castro Gallardo, J. (02 de Junio de 2018). *Sistemas Inteligentes basados en Análisis de Decisión Difuso.* Obtenido de SINBAD2 2016.: http://sinbad2.ujaen.es/sites/default/files/publications/TTII_JorgeCastro.pdf

Chodorow, K. (2013). *MongoDB The definitive guide.* O'Reilly Media.

De Campos, L., Fernández, J., Huete, J., & Rueda, M. (02 de Junio de 2018). *Research Gate.* Obtenido de ResearchGate 2018. All rights reserved.: https://www.researchgate.net/publication/237680987_USO_DE_CONOCIMIENTO_ESTRUCTURADO_EN_UN_SISTEMA_DE_RECOMENDACION_BASADO_EN_CONTENIDO1

Dussán, M. (2012). *Sistema de Recomendación web basado en la actividad de los usuarios de la universidad Nacional de Colombia.* Bogotá: Departamento de Ingeniería de Sistemas e Industrial.

Ekstrand, M., Riedl, J., & Konstan, J. (2011). *Collaborative Filtering Recommender Systems.*

Fakhfakh, R., Ammar, A., & Amar, C. (2014). Fuzzy User Profile Modeling for Information Retrieval. Tunisia.

Gonzales, A. (2014). *¿Qué es Machine Learning?* Obtenido de http://cleverdata.io/que-es-machine-learning-big-data/

Gonzáles, C. (2018). *Escuela Superior de Informática.* Obtenido de Escuela Superior de Informática: http://www.esi.uclm.es/www/cglez/downloads/docencia/2011_Softcomputing/LogicaDifusa.pdf

IBM. (2012). *FTP Software IBM.* Obtenido de FTP Software IBM: ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf

Jannach, Zanker, Felfering, & Friedrich. (2010). *An Introduction to Recommender.* Cambridge University Press.

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Martínez, C. (2006). *Los modelos clásicos de recuperación de.* México.

Michael, P., & Daniel, B. (18 de Junio de 2018). Obtenido de https://pdfs.semanticscholar.org/3444/6adc7d701a2c3a89c2fc5f6d3479eef407b0.pdf

Microsoft. (2015). Obtenido de https://docs.microsoft.com/es-es/dotnet/csharp/getting-started/introduction-to-the-csharp-language-and-the-net-framework

Mlandenic, D. (18 de Junio de 2018). *Text-learning and related intelligent agents: a survey.* Obtenido de IEEE Intelligent Systems and their Applications : https://ieeexplore.ieee.org/document/784084/

Nuñez, E., García, V., Pascual, J., Montenegro, C., Cueva, J., & Martínez, O. (02 de Junio de 2018). *Universidad de Oviedo .* Obtenido de Universidad de Oviedo : http://di002.edv.uniovi.es/~cueva/investigacion/tesis/Tesis-Edward.pdf

Pardo, C. (04 de Junio de 2018). AGENTES INTELIGENTES. España.

Rouse, M. (2015). *SQL Server.* Obtenido de SearchDataCenter: http://searchdatacenter.techtarget.com/es/definicion/SQL-Server

Russel, S., & Norving, P. (2010). *Inteligencia Artificial un enfoque moderno (3ra Edición).* España: Pearson Educación.

Salton, & McGill. (1983). *Introduction to Modern Information.* New York: Mc Graw Hill.

Salton, G., & Buckey, C. (s.f.). Relevance Feedback Information Retrieval. En *The SMART retrieval system - experiments in automated document processing* (págs. 313-323). Englewood Cliffs.

Salton, Wong, & Yang. (1975). Obtenido de http://www.cs.uiuc.edu/class/fa05/cs511/Spring05/other_papers/p613-salton.pdf

WIKIPEDIA. (s.f.). Obtenido de https://es.wikipedia.org/wiki/Microsoft_Visual_Studio

Wooldrige, & Jennings. (1995). *Intelligent Agents: theory and practice.* The Knowledge Engineering Review.

Ying, H. (2000). *Fuzzy Control and modeling.* IEEE Press Series on Biological Engineering.

Youtube. (11 de Julio de 2018). *Creator Academy*. Obtenido de Creator Academy: https://creatoracademy.youtube.com/page/lesson/impact-metrics#strategies-zippy-link-2

Anexos

Tabla 46. Tabla resumen de usuarios y recomendaciones

Usuario	T. Recomendados	T. Mostrados	T. No Mostrados	T. Relevantes M	T. Irrelevantes M	T. Relevantes No M	T. Irrelevantes No M
1	18	16	2	13	3	1	0
2	61	55	6	44	11	5	1
3	32	29	3	23	6	3	1
4	40	36	4	29	7	3	1
5	25	23	3	18	5	2	1
6	16	14	2	12	3	1	0
7	57	51	6	41	10	5	1
8	59	53	6	42	11	5	1
9	48	43	5	35	9	4	1
10	15	14	2	11	3	1	0
11	58	52	6	42	10	5	1
12	64	58	6	46	12	5	1
13	78	70	8	56	14	6	2
14	57	51	6	41	10	5	1
15	67	60	7	48	12	5	1
16	28	25	3	20	5	2	1
17	52	47	5	37	9	4	1
18	78	70	8	56	14	6	2
19	40	36	4	29	7	3	1
20	42	38	4	30	8	3	1
21	58	52	6	42	10	5	1
22	16	14	2	12	3	1	0

23	24	22	2	17	4	2	0
24	13	12	1	9	2	1	0
25	21	19	2	15	4	2	0
26	54	49	5	39	10	4	1
27	61	55	6	44	11	5	1
28	17	15	2	12	3	1	0
29	46	41	5	33	8	4	1
30	35	32	4	25	6	3	1
31	45	41	5	32	8	4	1
32	68	61	7	49	12	5	1
33	67	60	7	48	12	5	1
34	10	9	1	7	2	1	0
35	76	68	8	55	14	6	2
36	31	28	3	22	6	2	1
37	14	13	1	10	3	1	0
38	21	19	2	15	4	2	0
39	20	18	2	14	4	2	0
40	76	68	8	55	14	6	2
41	69	62	7	50	12	6	1
42	53	48	5	38	10	4	1
43	57	51	6	41	10	5	1
44	30	27	3	22	5	2	1
45	46	41	5	33	8	4	1
46	10	9	1	7	2	1	0
47	76	68	8	55	14	6	2
48	30	27	3	22	5	2	1

49	74	67	7	53	13	6	1
50	53	48	5	38	10	4	1
51	72	65	7	52	13	6	1
52	35	32	4	25	6	3	1
53	66	59	7	48	12	5	1
54	15	14	2	11	3	1	0
55	40	36	4	29	7	3	1
56	29	26	3	21	5	2	1
57	65	59	7	47	12	5	1
58	31	28	3	22	6	2	1
59	71	64	7	51	13	6	1
60	20	18	2	14	4	2	0
61	69	62	7	50	12	6	1
62	37	33	4	27	7	3	1
63	77	69	8	55	14	6	2
64	50	45	5	36	9	4	1
65	20	18	2	14	4	2	0
66	39	35	4	28	7	3	1
67	52	47	5	37	9	4	1
68	54	49	5	39	10	4	1
69	42	38	4	30	8	3	1
70	57	51	6	41	10	5	1
71	16	14	2	12	3	1	0
72	33	30	3	24	6	3	1
73	41	37	4	30	7	3	1
74	57	51	6	41	10	5	1

75	61	55	6	44	11	5	1
76	80	72	8	58	14	6	2
77	19	17	2	14	3	2	0
78	34	31	3	24	6	3	1
79	17	15	2	12	3	1	0
80	41	37	4	30	7	3	1
81	39	35	4	28	7	3	1
82	36	32	4	26	6	3	1
83	10	9	1	7	2	1	0
84	41	37	4	30	7	3	1
85	65	59	7	47	12	5	1
86	71	64	7	51	13	6	1
87	38	34	4	27	7	3	1
88	35	32	4	25	6	3	1
89	15	14	2	11	3	1	0
90	80	72	8	58	14	6	2
91	50	45	5	36	9	4	1
92	54	49	5	39	10	4	1
93	42	38	4	30	8	3	1
94	58	52	6	42	10	5	1
95	40	36	4	29	7	3	1
96	22	20	2	16	4	2	0
97	31	28	3	22	6	2	1
98	31	28	3	22	6	2	1
99	19	17	2	14	3	2	0
100	74	67	7	53	13	6	1

101	76	68	8	55	14	6	2
102	33	30	3	24	6	3	1
103	40	36	4	29	7	3	1
104	12	11	1	9	2	1	0
105	67	60	7	48	12	5	1
106	28	25	3	20	5	2	1
107	41	37	4	30	7	3	1
108	34	31	3	24	6	3	1
109	46	41	5	33	8	4	1
110	60	54	6	43	11	5	1
111	44	40	4	32	8	4	1
112	40	36	4	29	7	3	1
113	35	32	4	25	6	3	1
114	14	13	1	10	3	1	0
115	62	56	6	45	11	5	1
116	13	12	1	9	2	1	0
117	68	61	7	49	12	5	1
118	80	72	8	58	14	6	2
119	41	37	4	30	7	3	1
120	21	19	2	15	4	2	0
121	76	68	8	55	14	6	2
122	50	45	5	36	9	4	1
123	72	65	7	52	13	6	1
124	17	15	2	12	3	1	0
125	25	23	3	18	5	2	1
126	74	67	7	53	13	6	1

127	67	60	7	48	12	5	1