



Vigilada Mineducación

SISTEMA DE RECOMENDACIÓN CONTEXTUAL
Contextual Recommendation System

ANDRES FRANCO ZAPATA

Trabajo de Grado

Asesor, docente

Daniel Zuluaga Suarez

Marta Silvia Del Socorro Tabares Betancur

UNIVERSIDAD EAFIT
ESCUELA DE INGENIERÍAS
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA
MEDELLÍN
2022

Sistemas de Recomendación Contextual

Estudiante: Andrés Franco Zapata

Asesor: Daniel Zuluaga Suarez, Marta Silvia Del Socorro Tabares

Maestría en Ciencias de los Datos y Analítica

Universidad EAFIT

September 30, 2021

Abstract

El objetivo del presente trabajo es desarrollar una metodología para implementar un sistema de recomendaciones sobre un conjunto de datos masivos a partir de técnicas de filtro colaborativos, bajo las 6 etapas del marco de trabajo CRISP-DM como metodología para articular las necesidades del negocio con los resultados de un modelo matemático que permite cuantificar las oportunidades de negocio mediante la exposición de ofertas al cliente.

Keywords— sistemas de recomendación, filtros colaborativos, organización financiera, ofertas

1 Introducción

Las recomendaciones hacen parte de la historia de la humanidad, los primeros hombres no tenían el conocimiento de qué plantas eran aptas para el consumo, por lo tanto, no debería ser una sorpresa que las comunidades de nuestros antepasados tuvieran la necesidad de alimentarse de estas, desconociendo las consecuencias, lo mejor que podían hacer, era esperar a que otros las consumieran por mayor necesidad y conocer así los resultados, o simplemente acceder a ellas por supervivencia. El aprendizaje obtenido, ayudaba a que otros decidieran qué acciones tomar cuando se encontraban con cierta planta o fruto, de hecho, si evaluamos el comportamiento de algunos insectos como las hormigas cuando están buscando comida, son muy distantes unas a otras realizando tareas de búsqueda y recolección, pero cuando una de ellas encuentra algo de alimento, libera feromonas en su camino al hormiguero, lo cual permite que otras sigan el rastro y colaboren transportándolo.

En la era digital la información se ha convertido en uno de los activos más importantes que tienen las empresas para generar valor en nuevos productos y servicios que apalanquen al negocio. La banca, por ejemplo, genera millones de transacciones todos los días en la interacción con sus clientes; esto son datos que tienen un alto impacto en el negocio, donde las capacidades analíticas de las organizaciones deben continuamente desarrollar los procesos de análisis y procesamiento

de la información para transformarla en oportunidades de negocio y conocimiento de los clientes, logrando así mejorar la experiencia de estos, evaluar las tendencias del mercado, medir el impacto de sus servicios y generar innovación para la industria y sus clientes.

La llegada de nuevas tecnologías ha generado un crecimiento importante en la información, la aparición de los smartphones, tablets, IoT (internet de las cosas) y demás dispositivos móviles han permitido a los consumidores utilizar los recursos de internet con mayor facilidad para interactuar con empresas. Estas acciones han logrado que las industrias puedan perfilar a sus clientes, predecir comportamientos para conocer mejor a sus stakeholders. Los sistemas de recomendación obtienen mayor relevancia con toda la información disponible, puesto que estos son mecanismos de filtrado que permiten trabajar los problemas con altos volúmenes de datos, filtrando información relevante e identificando las preferencias, comportamientos o intereses del usuario (Lü et al., 2012).

Los sistemas de recomendación permiten estimar las capacidades de predicción sobre los intereses de un cliente, generando así un valor agregado en las estrategias comerciales gracias a la capacidad de presentar productos más afines al consumidor, por lo tanto, el uso de sistemas de recomendación se convierte en una solución que proporciona información relevante a los usuarios a los cuales se pretende ofrecer mejores productos y servicios y lograr una cotidianidad en el consumidor.

El reto entonces, en este nuevo servicio, se encuentra en conocer cómo se debe analizar la información disponible en los sistemas bancarios, sobre todas las interacciones generadas por los clientes utilizando canales como datáfonos (POS) y operaciones por internet de los productos como tarjeta de crédito y cuentas depósitos (ahorros y corriente), entendiendo al cliente como el centro del negocio y partiendo de la información transaccional y demográfica existente para apalancar las decisiones con la segmentación de las empresas y consumidores e identificando los criterios que los definen, para comprender sus necesidades y expectativas, todo esto con el propósito de plantear estrategias que impulsen la experiencia de usuario y logren fidelizar al consumidor a través de mejores recomendaciones que permitan la recordación para transformar el servicio en una cotidianidad.

En general, la industria del sector financiero cuenta con instituciones generadoras de información, por lo tanto, el análisis de información es una herramienta que ayuda a transformar los datos en conocimiento para entender los hábitos transaccionales, gustos y características generales de cada uno de los clientes y usuarios. Lo anterior permitirá desarrollar estrategias de innovación transformando los servicios financieros orientados hacia el cliente y así apalancar las ventas comerciales, brindando conocimiento de los clientes que interactúan con el comercio y logrando conocer detalles como: el día de más ventas, las horas más concurridas, la participación en el mercado según la competencia, el cliente objetivo, entre otras características que generen el crecimiento del negocio. Por otra parte, los consumidores se beneficiarán de todas estas estrategias que buscan aumentar la venta de productos y servicios mediante un análisis de mercado que utiliza la información demográfica del cliente y las características propias de cada comercio, construyendo así ofertas creadas a partir de sistemas de recomendación que filtran los intereses propios del consumidor y estimarán sus necesidades de consumo.

El problema entonces se encuentra en el análisis de información y la forma en la que se interpreta para tomar decisiones basadas en datos a la hora de plantear estrategias que apoyen el negocio. La evolución de la industria a pensar en servicios y poner al cliente como parte del negocio hacen que los sistemas de recomendación tengan un alto impacto y una gran oportunidad para identificar las posibles interacciones con nuevos comercios no conocidos por el usuario y premiar la fidelización, generando oportunidades de negocio y satisfacción en los clientes construyendo

relaciones basadas en la confianza ([Kindström, 2010](#)).

1.1 Contribución

La finalidad de este trabajo es ahondar en las metodologías necesarias de cara al análisis de datos para generar una propuesta de un sistema de recomendación basado en filtros colaborativos para usuarios de diferentes industrias, a través de la información transaccional de los clientes de un Banco. Lo anterior parte de diferentes investigaciones del entorno como industrias, instituciones financieras y expertos, para mostrar cómo desde el análisis de datos y algoritmos de filtros colaborativos, será posible identificar la relación cualitativa y algunos factores comunes y claves en la interacción entre comercios y consumidores para lograr un crecimiento en ventas y conocimiento a profundidad del comportamiento del usuario.

1.2 Alcance

El alcance de este trabajo es definir una metodología que permita desarrollar un sistema de recomendación.

Este trabajo consta de varias fases para la implementación de un sistema de recomendación como estrategia para la creación de ofertas especializadas para los clientes de una institución bancaria. Se profundizará en varios factores como el estudio del entorno actual, la investigación de una metodología de análisis de datos para la creación de estrategias que generen recomendaciones acertadas por parte de negocios a usuarios y se pondrá en contexto como ejemplificación el caso de una institución financiera.

Es importante resaltar que en este trabajo no se hará profundidad a una retroalimentación por parte de los clientes como tampoco el despliegue del modelo en un ambiente productivo para presentar las ofertas a los clientes en los canales de distribución. Por temas legales tampoco se presentarán resultados que puedan comprometer información sensible de clientes, ni cifras exactas en relación a los resultados esperados, sin embargo, se generarán apreciaciones aproximadas y conclusiones cualitativas de cara a la metodología.

2 Revisión Literaria

2.1 Marco Teórico

Los sistemas de recomendación se han convertido en una rama importante de la investigación desde mediados de los años 90, para ayudar a las personas y empresas a tomar decisiones basadas en las recomendaciones de otros usuarios. En ([Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994](#)) se presenta el desarrollo de "GroupLens" un sistema distribuido para recolectar y utilizar las calificaciones de algunos usuarios con el fin de estimar el interés de otros consumidores en noticias. La arquitectura implementada por GroupLens almacena la información del consumo de noticias que se han leído para luego generar calificaciones, proporcionar nuevas predicciones para nuevos consumidores y retroalimentar las preferencias del usuario lector. La arquitectura presentada en ([Konstan et al., 1997](#)) permite comprender la importancia de almacenar la información sobre el consumo de artículos, en este caso noticias, con las que se ha generado una interacción para luego generar calificaciones, proporcionar nuevas características para identificar las necesidades

de los consumidores y finalmente mejorar las preferencias del usuario. El propósito de esta implementación es habilitar las capacidades de los sistemas, respondiendo a los obstáculos que la industria tiene y adaptándose en el tiempo para alcanzar los siguientes objetivos que marcaron un antes y un después en el desarrollo de sistemas con estas características:

- La apertura a usuarios nuevos y existentes
- Facilidad de uso e interpretación de las calificaciones
- Compatibilidad con los servicios
- Escalabilidad en el crecimiento de nuevos usuarios
- Privacidad en las preferencias y gustos del usuario

Con base en lo anterior, varias compañías han tomado como ejemplo este caso para crear proyectos que han permitido a las organizaciones implementar soluciones para mejorar la experiencia de los clientes a la hora de encontrar productos que posiblemente no descubrirían sin una recomendación especializada. A continuación, se presentan tres casos de éxitos modernos en la industria implementando sistemas de recomendación.

Amazon es una compañía de Estados Unidos que se dedica a la venta de productos de diferentes artículos como libros, software, ropa, música entre otros, actualmente se dedica también a prestar servicios en nube (Huckman, Pisano, & Kind, 2008) exponiendo sus capacidades y aplicaciones a las necesidades de otros. En Amazon los sistemas de recomendación han sido una metodología que les ha permitido aprovechar las capacidades de estos algoritmos para personalizar las tiendas según las necesidades del cliente, por ejemplo, presentar artículos de bebés para una nueva madre o recomendar libros de programación para un ingeniero de software. Los algoritmos de recomendación son una forma de marketing especializado mediante la estructuración de una experiencia basada en la interacción de compra personalizada para cada cliente, estas metodologías permiten la escalabilidad con grandes volúmenes de información y el tiempo es un factor determinante para la maduración de estos sistemas (Linden, Smith, & York, 2003; Smith & Linden, 2017).

Netflix es una empresa de servicios de streaming de contenidos audiovisuales, otro gran ejemplo en el uso de sistemas de recomendación. La compañía ha utilizado estos algoritmos para brindar a sus usuarios la capacidad de seleccionar el contenido con el que se desea poder establecer interacciones positivas con el material propuesto según los gustos de la persona. En octubre del 2006 la compañía publicó un desafío llamado “Netflix Prize” donde se retó a toda la comunidad de la minería de datos, machine learning e informáticos para desarrollar un sistema de recomendación que lograra superar la precisión de su sistema Cinematch (Bennett, Lanning, et al., 2007) para un conjunto de datos de más de 100 millones de películas calificadas anónimamente. El equipo ganador del gran premio usó como base para sus predictores modelos de filtros colaborativos que buscan identificar las interacciones entre los usuarios y elementos que generan calificaciones diferentes, además del tratamiento preciso en las características o sesgos de la información (Koren, 2009)

Youtube es una web dedicada a compartir videos, los usuarios día a día acuden a consumir los servicios de la plataforma con el fin de encontrar contenido de interés. En la figura x, adaptada de (Alvarado, Heuer, Vanden Abeele, Breiter, & Verbert, 2020) se presentan los cuatro actores que influyen en las predicciones generadas por el sistema de recomendación. Uno de los intereses principales que debe tener en cuenta la compañía, es ser capaz de entregar las recomendaciones más acertadas a cada uno de sus clientes, dándoles así la oportunidad de decidir cuales son los factores y características más importantes para acceder al contenido audiovisual. La plataforma sugiere videos utilizando la interacción del consumidor mediante acciones como: visto, marca de

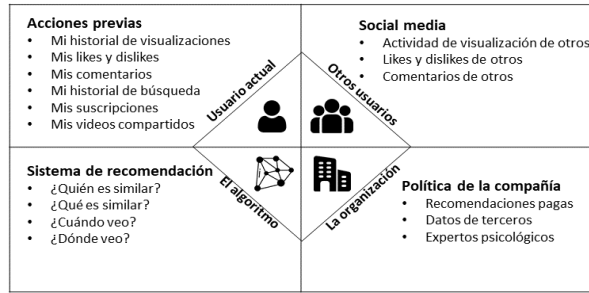


Figure 1: Creencias de los usuarios acerca de las recomendaciones de Youtube

favoritos, me gustas, entre otras acciones.

Profundizando en lo anterior, se plantea que los sistemas de recomendación pueden ser una metodología que genera valor en la transformación de la información como medio para construir nuevos servicios que permitan a las organizaciones sacar provecho de sus datos y de la interacción que generan con sus clientes. Los sistemas de recomendación se han definido como herramientas para el procesamiento de altos volúmenes de datos implementando algoritmos matemáticos para desarrollar y aumentar los procesos en el uso de recomendaciones (Adomavicius & Tuzhilin, 2005), estos sistemas buscan transformar los datos en información utilizando todos los eventos generados por los usuarios de las organizaciones usando filtros colaborativos (Schafer, Frankowski, Herlocker, & Sen, 2007), filtros basados en contenido (Lops, De Gemmis, & Semeraro, 2011) o filtros híbridos (Gomez-Uribe & Hunt, 2015).

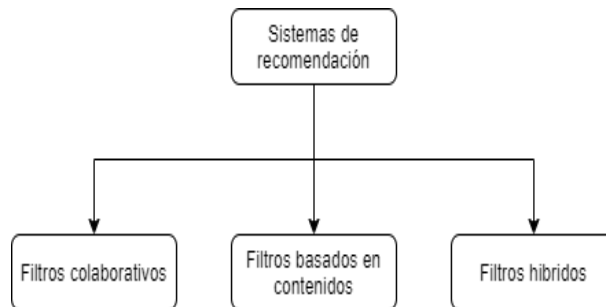


Figure 2: Técnicas en Sistemas de Recomendación

La metodología de filtros basados en contenido es un algoritmo que tiene como base la predicción, haciendo énfasis en el análisis de los atributos únicos de cada uno de los elementos, generando así una estimación (Aggarwal et al., 2016). Usualmente esta técnica es muy utilizada en los sistemas de recomendación basados en documentos (textos, libros, investigaciones, entre otras). Esta técnica no requiere información de otros usuarios para generar una recomendación, puesto que, parte del perfil de cada elemento y su interacción con el usuario (Zhang & Iyengar, 2002; Bobadilla, Ortega, Hernando, & Gutiérrez, 2013).

Los filtros colaborativos se basan generalmente en la relación entre usuarios e ítems. El algoritmo utiliza la información de las interacciones o calificaciones de los usuarios de los productos, asumiendo que los usuarios que estuvieron de acuerdo en el pasado van a estarlo en el futuro y

tendrán gustos por productos similares por los que mostraron algún tipo de interés en el pasado (Breese, Heckerman, & Kadie, 2013; Herlocker, Konstan, Terveen, & Riedl, 2004). En esta metodología existen dos clases: primero están los memory-based, que buscan a través de las similitudes entre usuarios o productos, identificar vecindades para dar recomendaciones (Zhao & Shang, 2010; Jannach, Zanker, Felfernig, & Friedrich, 2010). Segundo están los model-based, los cuales buscan a través de minería de datos y diferentes algoritmos de machine learning, predecir las calificaciones de productos que no han sido calificados por los usuarios (Su & Khoshgoftaar, 2009). Los model-based incluyen algoritmos de factores latentes, los cuales realizan reducción de dimensionalidad mediante la factorización de la matriz usuario-producto, a través de métodos de optimización que son resueltos por técnicas iterativas como gradiente descendente estocástico y alternative least squares (ALS) (Mehta & Rana, 2017).

La técnica de filtrado híbrido combina diferentes aspectos de los filtros colaborativos y filtros basados en contenidos implementando los algoritmos en diferentes combinaciones utilizando el resultado como retroalimentación del otro o desarrollando una recomendación unificada (Thorat, Goudar, & Barve, 2015; Burke, 2002).

2.2 Estado del Arte

Los algoritmos de inteligencia artificial y el Big Data han ido obteniendo mayor relevancia en los últimos años gracias al aumento exponencial de la información y el incremento en la capacidad de procesamiento y almacenamiento de esta (Yu, Song, et al., 2020). Los sistemas de recomendación han sido un área del conocimiento que ha crecido debido a los grandes volúmenes de datos que generan las industrias como el comercio por internet, servicios de streaming, entretenimiento entre otros.

La banca, por ejemplo, se enfrenta a dos grandes retos en la implementación de sistemas de recomendación utilizando filtros colaborativos. El primer desafío consiste en la falta de información explícita que tienen sobre las interacciones del cliente con el banco, debido a la relación transaccional que existe entre ambos; el segundo problema, parte de que se hace difícil generar recomendaciones a usuarios nuevos o que no cuenten con alguna interacción. Los autores, proponen una metodología para resolver los desafíos implementando un primer sistema de recomendación con calificaciones implícitas y el desarrollo de un segundo sistema híbrido utilizando información de preferencia y las características demográficas del cliente, logrando mejorar las ventas de productos bancarios (Oyebode & Orji, 2020). A continuación, se hace una revisión literaria en la implementación de sistemas de recomendación para el sector bancario:

En (Sharifhosseini, 2019), presentan la implementación de un sistema de recomendación como un mecanismo para identificar los comercios que un cliente no ha visitado, pero que es probable que visite, utilizando tarjetas débito. Los autores proponen un algoritmo de filtros colaborativos usando factorización de matrices para estimar las relaciones de un cliente con un comercio.

El mercado de valores donde se compran y venden títulos de renta fija o variable, es otro campo donde se pueden generar recomendaciones para apoyar a los usuarios en sus negociaciones. Los filtros colaborativos ayudan a identificar los comportamientos similares entre empresas, identificando las transacciones que tienen ganancias y pueden ser sugeridas entre clientes para la compra de dichos títulos (Sayyed, Argiddi, & Apte, 2013).

El auge de servicios y la digitalización de las industrias se han transformado en nuevos desafíos para el negocio, los clientes han pasado de ser compradores a convertirse en activos. El mundo y los cambios de generación han puesto sobre la marcha la necesidad de tener mayores canales

de distribución donde los usuarios puedan realizar las interacciones con el sistema financiero de maneras más ágiles y efectivas. La personalización de servicios y la necesidad puntual de cada usuario es una prioridad para el negocio, los clientes cada vez demandan más velocidad y facilidad en los procesos. Los filtros colaborativos son una alternativa para poder entender la interacción que tuvo un cliente con un canal o producto y personalizarlo de acuerdo a sus necesidades, la arquitectura y la capacidad de los canales para exponer los servicios y las personalizaciones necesarias para mejorar la experiencia del cliente deben ser una prioridad ([Abdollahpouri & Abdollahpouri, 2013](#)).

Para ([Hernández-Nieves, Hernández, Gil-González, Rodríguez-González, & Corchado, 2020](#)), la evolución de los servicios financieros y la transformación en la entrega de servicios personalizados al cliente ha traído la necesidad de implementar tecnologías de la información al habilitar las capacidades del negocio para procesar los datos y exponer los servicios imprescindibles para el negocio, buscando mejorar la experiencia. El artículo propone una arquitectura basada en redes de niebla y realiza una simulación de los datos en el contexto de la banca tradicional.

Basado en las implementaciones mencionadas anteriormente y en general en la literatura de sistemas de recomendación, el crecimiento de la información ha llevado a la banca a pensar en el cambio de los canales tradicionales a la digitalización de sus servicios para la adopción de las nuevas necesidades de las personas a un mundo más expuesto a internet. Por lo tanto, las metodologías de filtros colaborativos para resolver el problema de sobrecarga de información se tornan cada vez más relevantes, buscando apoyar al negocio en prestar nuevos servicios que apoyen la experiencia del cliente logrando la satisfacción de los consumidores y un crecimiento evidenciable en ventas ([Yahyapour, 2008](#)).

3 Metodología

Para el año 1999 se presenta la primera versión de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), como una guía detallada para orientar los proyectos relacionados con datos. Esta metodología pronto se convirtió en una de las propuestas metodológicas más trabajadas alrededor del mundo para resolver los problemas de minería de datos y analítica ([Martínez-Plumed et al., 2019](#)). En el desarrollo de este trabajo se implementó la metodología iterativa CRISP-DM, la cual adapta elementos similares al de un proyecto de ingeniería de software para enriquecer el ciclo de vida de un proyecto de análisis de datos. El modelo proporciona 6 fases: comprensión del negocio, comprensión de los datos, análisis de los datos, modelamiento de la información, evaluación y despliegue; todas estas fases giran alrededor de la información y determinan las actividades que se deben desarrollar en un determinado espacio de tiempo de manera reiterada, asignando tareas concretas y definiendo los objetivos de lo que se desea lograr en cada una de ellas.

En la figura 3, adaptada de ([Wirth & Hipp, 2000](#)) observamos el proceso iterativo y dependiente de cada una de las fases desarrolladas durante la ejecución del proyecto, cada una de estas etapas está diseñada para ayudar a buscar soluciones enfocadas a las necesidades del negocio haciendo que sean alcanzables de cara a los hallazgos identificados en las fases posteriores a la comprensión del negocio, esto permite un crecimiento constante que permite a todo el equipo identificar y redescubrir nuevas oportunidades logrando así poder convertir la información en una oportunidad que apalanque al negocio.

Gracias al papel itinerante de la metodología CRISP-DM, el proceso de refinamiento permite volver a alguna fase, gracias al análisis de las necesidades en cada una de estas, permitiendo al

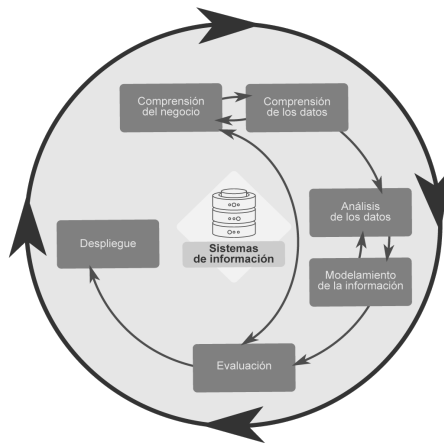


Figure 3: Metodología CRISP-DM

equipo refrescar el objetivo general a través de conclusiones y oportunidades de mejora sobre los procesos y actividades desarrolladas con el fin de implementar mejoras de cara a las necesidades del negocio.

4 Desarrollo

Durante el desarrollo de este trabajo se realizaron dos iteraciones de la metodología CRISP-DM, gracias a los hallazgos identificados en las fases de Comprensión de los datos y Modelamiento de la información. A continuación, se realizará una breve explicación de cada una de las fases de la metodología CRISP-DM y se detallarán las actividades realizadas para la implementación del modelo de recomendación.

4.1 Comprensión del Negocio

El objetivo principal de esta fase es definir y alinear los objetivos del proyecto de análisis de información con las necesidades del negocio, buscando que los datos sean una manera de producir un resultado que permita a la organización obtener un beneficio con el desarrollo del proyecto. Los entregables de esta fase son:

- Establecer los objetivos de negocio
- Evaluar la situación actual
- Obtener un plan de proyecto

La necesidad del negocio consiste en que los equipos comerciales y del negocio tengan insumos para tomar mejores decisiones en la creación de ofertas especializadas para un grupo de personas que tengan una relación o afinidad con un tipo de comercio específico, permitiendo apoyar a los negocios en sus ventas y mejorar la relación entre la organización y el cliente que tendrá acceso a las ofertas mediante los canales de distribución. Los clientes y comercios son el enfoque principal para generar una recomendación y se debe tener presente en el desarrollo las casuísticas

de negocio como:

- Recomendar ofertas a clientes activos
- Habilitar las capacidades para recomendar comercios a clientes nuevos
- Identificar los comercios que hacen parte a un conglomerado de empresas
- Identificar los comercios más afines al consumidor
- Regionalizar las ofertas según la información demográfica de clientes y comercios

Para desarrollar la temática central de este trabajo que son los sistemas de recomendación, se debe en primer lugar, evaluar la situación actual de estos, teniendo en cuenta cada uno de los insumos necesarios para poder ejecutarlo: los clientes como portadores de la información, la tecnología usada, las áreas y los equipos integrados, entre otros; uno de los factores más importantes para realizar un análisis de datos que permita generar recomendaciones es precisamente eso, los datos. Por tanto, es necesario ofrecer una contextualización reciente que brinde una visión para identificar cómo se encuentra el entorno actual que permitirá otorgar a las compañías el insumo más importante del proceso.

El panorama actual trae una unión inseparable entre los clientes como portadores de la información y la tecnología usada. Según (Gil, Castro, Bermúdez, et al., 2017) se evidencia un crecimiento en temas de tecnología de la información a nivel mundial en los últimos quince años con más de siete mil millones de usuarios que cuentan con un teléfono móvil y tres mil doscientos millones de personas que tienen internet, y es que es gracias a la digitalización que podemos obtener los datos de los usuarios para crear sistemas de recomendación, estas cifras a simple vista parecen grandes, pero es importante resaltar que esto solo representa aproximadamente un 53% de la población mundial.

Lo anterior pareciese mostrar un entorno desalentador para los sistemas de recomendación, teniendo en cuenta que las estadísticas pueden ser incluso más desfavorables para Colombia, que debido a su brecha digital persistente, complica la recolección de datos por parte de las industrias Colombianas. Sin embargo, en el 2020, la pandemia ha sido un punto a favor puesto que no solo ha servido de potencializador para mejorar la digitalización en Colombia si no también la inclusión financiera. (Ramos Pérez et al., 2021)

Para esta fase, también como insumo principal es de gran utilidad contar con actores como los dueños de producto y áreas de conocimiento del cliente como el equipo comercial, también los expertos en temas de sistemas como científicos de datos. Los integrantes de cada frente presentaron las necesidades para el desarrollo del sistema de recomendación y refinaron las necesidades según los hallazgos identificados en cada una de las 2 iteraciones.

En conclusión, a pesar de que las cifras demuestran que aún faltan grandes cambios para romper la brecha digital en Colombia, los eventos del último año han permitido un crecimiento de inclusión en donde cada vez más personas están teniendo acceso al internet, lo cual es una oportunidad para conectar con las industrias y comenzar a recolectar datos que puedan ser analizados por los negocios desde un punto de vista más amplio. Gracias a la apertura a nuevos públicos objetivos, será posible generar los insumos básicos para sistemas de recomendación más acertados, todo esto, de la mano de un equipo de trabajo que busque una forma de mejorar la experiencia del cliente de manera directa a través de las compras y las interacciones de los consumidores y mediante los productos y canales de distribución, facilitando así la relación de consumo entre el usuario y el negocio. Las ofertas son un medio para llegar a los clientes e impulsar su experiencia de manera indirecta utilizando estrategias de marketing para presentar la publicidad que el cliente desee consumir (Meyer, Schwager, et al., 2007).

El cliente es el principal foco de este trabajo puesto que él es el centro de acciones que a través de diferentes herramientas se convertirán en datos, que posteriormente pasarán por un análisis para ofrecer soluciones que pueden convertirse en ofertas, estas son un insumo que buscan mejorar la experiencia entre los clientes y el sector bancario, apoyando los intereses de cada parte.

4.2 Comprensión de los Datos

La necesidad del negocio parte de utilizar la información transaccional de los clientes de una institución bancaria como insumo principal de datos para construir un sistema de recomendación.

Los datos permiten identificar la interacción entre el comercio y los clientes. En este momento es donde se conocen los hábitos de consumo del cliente para poder generar ofertas. Además, es posible identificar otras características importantes tales como información demográfica, conductual, psicológica, entre otros atributos de clientes que acceden a productos y servicios del comercio.

Esta fase consiste en comenzar con la extracción y exploración de los datos iniciales que serán el insumo para el desarrollo de todo el proyecto. El principal objetivo de esta segunda fase es tener una descripción de los datos que serán utilizados y la verificación de calidad de estos.

Para la construcción y desarrollo del modelo de recomendación de ofertas y poder generar todo el contexto del cliente y los comercios, se trabajó con las fuentes de información histórica transaccional de todos los usuarios que han interactuado con un comercio adquirente o han realizado una transacción no presente (internet). Se utilizaron datos en una ventana de tiempo de 3 meses, tercer trimestre del año 2020, específicamente de los datos relacionados con tarjetas debito y credito. Las fuentes de información son:

- Maestro de Productos (Cuentas depósitos y Tarjetas)
- Maestro de Clientes (Información Demográfica)
- Maestro de Cuentas
- Históricos transaccionales de clientes
- Histórico transaccional de comercios (Adquirentes con la organización)

Durante el análisis de datos se declaró al negocio que la relación transaccional entre un comercio y un cliente solo puede depender del valor y la frecuencia de compra, esto se debe a que la industria brinda una información insuficiente sobre el tipo de producto o servicio adquirido por el usuario, evitando así hacer explícitos los datos más allá de la transacción, esto puede complicar la exactitud de la recomendación.

El mayor problema identificado con la información transaccional es que existen comercios que venden múltiples tipos de productos y servicios, por lo tanto, puede pasar que la relación con un comercio sea estricta hacia un producto específico y no esté relacionada con una sola categoría de todo su catálogo de ventas. Un ejemplo de lo anterior se presenta en la interacción de un cliente con un comercio retail donde adquirió un producto de tecnología, sin embargo, una posible recomendación podría ser un producto de la canasta familiar siempre y cuando esté amparado bajo la misma marca.

La información disponible solo tiene la relación entre cliente y comercio, pero no tiene la relación de los productos adquiridos en un comercio.

Una posible solución al problema presentado anteriormente sobre el detalle de la compra o adquisición, sería construir estrategias de alianzas comerciales entre las empresas adquirentes y

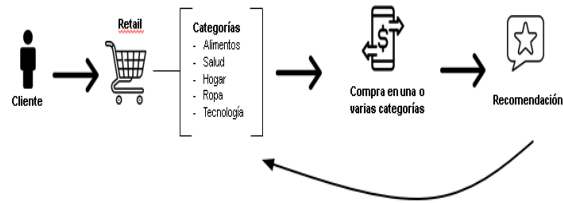


Figure 4: Recolección de datos en una interacción entre el cliente y el comercio

la institución bancaria, para determinar el detalle en las compras realizadas por sus clientes y obtener beneficios en el análisis de datos hacia ambas partes.

4.3 Análisis de los datos y selección de características

Durante esta fase se debe concluir cuál es el conjunto de datos que se utilizará para el desarrollo del modelo. Las actividades de extracción de datos de los sistemas transaccionales definen el punto de partida para la transformación final de la información y construcción de los datos que se utilizarán en la siguiente fase de modelado.

En el apartado anterior, se expusieron las fuentes transaccionales que se utilizaron como insumo para la construcción del conjunto de datos. Durante la preparación de estos se utilizó el proceso de extracción, transformación y carga, que permitió recopilar los datos de las diferentes fuentes de información en los sistemas y subdominios de la institución financiera para la transformación y limpieza. La carga de los datos a un sistema independiente a los ambientes productivos es parte fundamental de esta fase, debido a las limitaciones o afectaciones que puedan tener durante el proceso de transformación y uso de la información en ambientes productivos o transaccionales que puedan exigir un sobreesfuerzo a la infraestructura tecnológica.

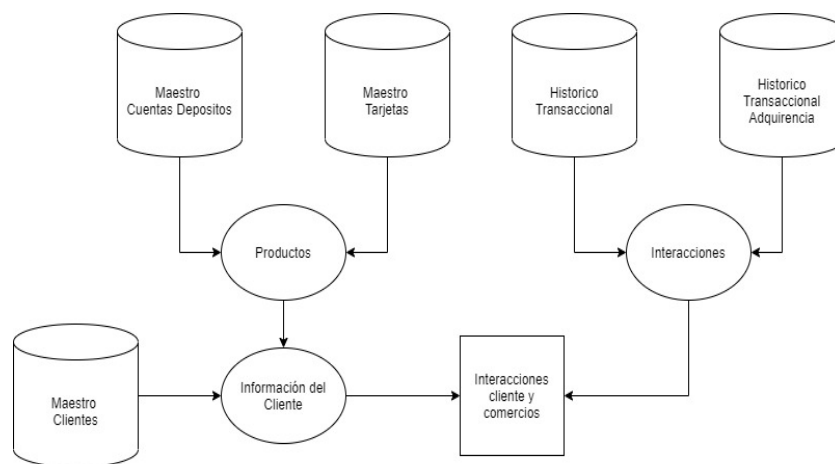


Figure 5: Modelo de extracción a alto nivel de la información

En la figura 5, podemos observar a alto nivel el proceso de extracción y relación de la información que se utilizó para la construcción del conjunto de datos, a partir de la interacción de un cliente con un comercio, utilizando la información de sus productos y su historial transaccional relacionado a compras o interacciones realizadas. La relación entre un comercio puede verse desde la perspectiva del cliente como comprador o como una venta del comercio adquirente hacia un cliente.

La base de datos construida cuenta con 30'029.517 observaciones, cada registro es la relación de un cliente con un comercio, cada usuario está acotado a un máximo de veinte interacciones. Los datos recopilados constan de tres variables que se presentan a continuación:

- `user_id`: Es el identificador único de un tarjetahabiente con los datos de las interacciones de dicho cliente con uno o varios comercios.
- `item_id`: Identificador único del comercio adquirente que ha generado ventas realizadas por los clientes con los pagos realizados con tarjeta crédito o débito.
- `n_trx`: Número de transacciones que ha generado un cliente utilizando sus productos de tarjetas en un comercio adquirente de la organización.

Este conjunto de datos construido, es el insumo principal que permitirá al equipo mediante el uso de técnicas de análisis de información implementar un sistema de recomendación basado en técnicas de filtros colaborativos que busca estimar la calificación implícita sobre la interacción de un cliente con un comercio.

4.4 Modelamiento de la Información

En la fase de modelado se busca identificar los algoritmos matemáticos que serán utilizados para implementar una solución usando filtros colaborativos como medio para poder ofrecer a los clientes bancarizados recomendaciones especializadas de acuerdo con sus interacciones con los comercios, también se diseña un plan de evaluación que permita identificar si se cumplen las necesidades del negocio.

Para el desarrollo de este proyecto se implementó una técnica de filtros colaborativos con información implícita para sistemas de recomendación, la cual busca a través de la interacción entre un comercio y el consumidor, identificar patrones similares de consumo que permitan estimar o predecir ofertas para un determinado usuario.

Los datos construidos a partir de la información transaccional de una institución financiera no tienen una calificación explícita por los usuarios, si no la frecuencia de consumo entre un cliente y un comercio, por lo tanto, la metodología que se utilizará está descrita en (Hu, Koren, & Volinsky, 2008), donde se presenta una alternativa para trabajar con información implícita utilizando una técnica para la factorización de matrices para minimizar la siguiente función de costo:

$$\min_{x^*, y^*} \sum_{(u,i)} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2) \quad (1)$$

El objetivo entonces es encontrar el vector $x_u \in \mathbb{R}$ para cada usuario u y el vector $y_i \in \mathbb{R}$ para cada comercio i . Para nuestro modelo r_{ui} es el número de interacciones que ha tenido un usuario con un comercio. El rating está dado por $p_{ui} = x_u^T y_i$, este vector es conocido como los factores de usuarios y comercios, los vectores se esfuerzan por mapear usuarios y comercios en un espacio de factores latentes donde puedan ser comparados, entonces para medir la confianza en cada observación tenemos a $c_{ui} = 1 + \alpha r_{ui}$, en este modelo se usó $\alpha = 40$ según los experimentos

de los autores de (Hu et al., 2008). La ecuación (1) es similar a las técnicas para información explícita, con dos distinciones (1) se deben tener presentes todas las variaciones en los niveles de confianza, (2) la optimización debe tener en cuenta todos los pares u, i , en lugar de solo los correspondientes a los datos. La función de regularización está dada por $\lambda(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$, donde λ en el caso de este trabajo, fue encontrado utilizando técnicas de validación cruzada para evadir el sobre entrenamiento del modelo.

Por las características de la información el resultado del modelo no puede ser evaluado con las métricas que usualmente se utilizan como el RMSE para los sistemas de recomendación con información explícita, es decir, la información disponible no tiene una calificación tangible del gusto del usuario si no una interacción que tuvo el usuario con un comercio, además en la información financiera no existe una retroalimentación del usuario de cuáles son los comercios que más le gustan o son de su interés, por lo tanto, la métrica que se utilizó para la evaluación del modelo es la misma que se utilizó en (Hu et al., 2008), algunas fuentes la llaman rank ordering error metric ROEM $ROEM = \frac{\sum_{u,i} r_{ui}^t rank_{ui}}{\sum_{u,i} r_{ui}^t}$, donde $rank_{ui}$ es la clasificación porcentual del comercio de una lista ordenada de todos los comercios con los que el usuario tuvo una interacción. De esta forma si $rank_{ui} = 0\%$ entonces el comercio es el más deseable por el usuario u , si $rank_{ui} = 100\%$ se interpreta como el comercio que menos le gusta al usuario.

El modelo Alternating Least Squares como algoritmo para la construcción de un sistema de recomendaciones fue desarrollado usando el motor de procesamiento distribuido Apache Spark (Meng et al., 2016), como herramienta de desarrollo para el entrenamiento y procesamiento de las matrices generadas para la estimación de nuevas ofertas para los clientes.

4.5 Evaluación

Luego de tener un modelo que cumpla con las necesidades del negocio, se procede a cuestionar las estrategias construidas durante toda la metodología para garantizar que la mayoría de las soluciones por parte de la industria se cumplan acorde al plan de trabajo, para esto se genera una evaluación de todos los procesos realizados hasta la última etapa.

La extracción de información para los procesos de modelamiento de datos, como también los insumos necesarios para generar y segmentar los clientes se ven reflejados en la figura 5, donde se extrae toda la información de usuarios activos y la información transaccional con la que se entrena el modelo. En el proceso de entrenamiento el algoritmo utiliza factores latentes para mapear en un espacio a los clientes y comercios, estos vectores son almacenados para poder predecir nuevos ratings.

Como se mencionó en la fase anterior, el modelo de recomendación cuenta con un rating implícito donde no existe una calificación explícita sobre los gustos de un usuario a un comercio específico, sino que se tiene la cantidad de interacciones que tiene un cliente con un comercio. En la siguiente sección se discutirá sobre los resultados del modelo y los pasos a seguir para la implementación de un sistema de recomendación que apalanque la estrategia de ofertas especializadas para todos los usuarios bancarizados.

Los modelos de recomendación con calificación implícita son un reto para los procesos de evaluación porque no se tiene una calificación tangible de si el consumidor está a gusto con el producto o servicio adquirido, ni la imagen o percepción que tiene con un negocio específico. La evaluación del modelo se hizo utilizando la métrica ROEM, la literatura sugiere tener la métrica por debajo del 50% para asegurar que la recomendación sea mejor que una decisión aleatoria.

$\sum_{u,i} r_{u,i}^t$	ROEM	Regla de acotación
9352667	38,78%	Conjunto de datos completo
9354485	38,82%	Mínimo 10 clientes por comercio
9333686	38,74%	Mínimo 20 clientes por comercio
9260998	38,74%	Mínimo 30 clientes por comercio
9185094	38,57%	Mínimo 40 clientes por comercio
9093712	38,48%	Mínimo 500 clientes por comercio
8709054	38,19%	Mínimo 100 clientes por comercio
8111101	37,64%	Mínimo 200 clientes por comercio
7669731	37,16%	Mínimo 300 clientes por comercio
7317638	36,90%	Mínimo 400 clientes por comercio
7029662	36,59%	Mínimo 500 clientes por comercio
6091140	35,46%	Mínimo 1000 clientes por comercio

Figure 6: Descripción de las reglas con las que se acotaron los comercios

La selección del mejor modelo se hizo utilizando validación cruzada donde obtuvimos un ROEM de 38.78%, con los parámetros de $\alpha = 40$, $\lambda = 0.1$ y $rank = 16$. Durante el proceso de evaluación identificamos que el ROEM del modelo se ve altamente afectado por clientes que tienen pocas interacciones con comercios, por lo tanto, se probaron modelos aumentando el número mínimo de interacciones de un cliente con diferentes comercios como también la interacción mínima de un comercio, es decir que si un comercio no tenía al menos 100 o más interacciones no se tenía en cuenta para el entrenamiento. El ROEM mejora mucho a medida que se hacen este tipo de restricciones, pero se pierde mucha información debido a las reglas duras previas al entrenamiento, por lo tanto, se decide entrenar el modelo con todos los datos disponibles.

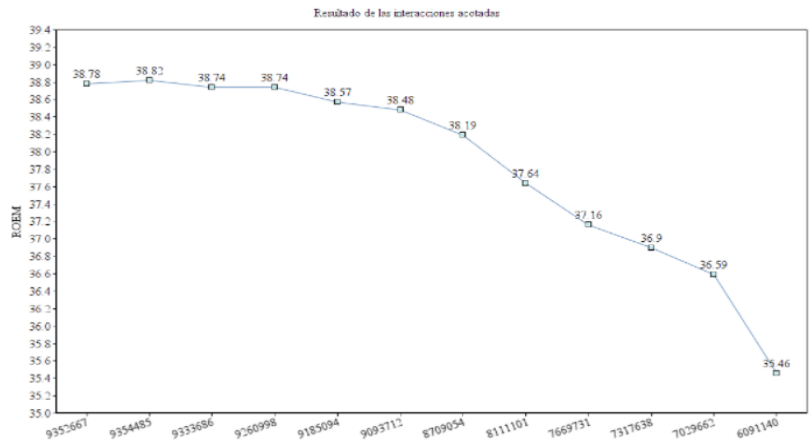


Figure 7: Evolución ROEM con las limitaciones por interacción

En las figuras 6 y 7, se presenta cómo se comporta el ROEM cuando se aplican las reglas para buscar maximizar las relaciones e interacciones entre comercios y clientes. Cuando se acota el conjunto de datos para que un comercio entre en el modelo de recomendación se llega a perder mucha información debido a que existen múltiples métodos de pago como tarjetas, compras

no presentes, transferencias digitales o efectivo. El mejor ROEM representa una pérdida de aproximadamente del 65% de la información de las interacciones, esto representa una de las problemáticas de los sistemas de recomendación, tener que entregar una estimación a un cliente o comercio con el que no fue entrenado el modelo. Por lo tanto, se decide dejar el modelo inicial para garantizar que se tenga presente todos los clientes y comercios en la ventana previamente establecida.

Durante la evaluación de clientes que no entraron en el modelo ya sea porque no hicieron transacciones monetarias durante la ventana de tres meses seleccionada para el entrenamiento del modelo, se generó la recomendación basado en dos puntos la segmentación comercial interna y la ubicación geográfica, es decir que se utilizaron los mejores comercios de personas que estaban en la misma región y tenían una relación comercial similar con otros clientes que sí entraron en el proceso de entrenamiento del modelo. La evaluación de esta necesidad se hizo con las recomendaciones entregadas a los clientes nuevos y sus interacciones por fuera de la ventana de tiempo antes y después para identificar si estos clientes en algún momento tuvieron relación con este comercio específico o realizaban alguna actividad con un comercio similar.

4.6 Despliegue

La última fase, determina el uso y el alcance que tiene el negocio sobre los hallazgos entregados por el modelo construido para la exposición de ofertas especializadas a los usuarios. En las necesidades del negocio está como factor principal tener un modelo actualizado, en el caso del alcance planteado para este trabajo, esta actualización se llevará a cabo cada tres meses con el fin de tener presentes los hábitos de consumo de los clientes y sus interacciones con nuevos comercios.

Es importante aclarar que, aunque el despliegue productivo no es la finalidad de este trabajo, es de gran valor entregar las recomendaciones a la institución financiera, puesto que estas servirán como insumo para alinear las estrategias de cara a las ofertas y a los clientes que utilizan los canales digitales en su día a día, para generar así mejores experiencias al usuario y la oportunidad a los comercios de conocer mejor a sus clientes.

Las recomendaciones son un insumo muy valioso porque son la entrada para conocer las preferencias del usuario y ofrecer resultados que pueden ser más acertados en la venta de un producto o servicio, todo esto se hace utilizando las herramientas construidas por otros equipos integrados que segmentarán internamente a los clientes, construirán perfiles transaccionales e identificarán los hábitos de consumo que aportarán un conocimiento especializado de las necesidades del cliente, permitiendo así la implementación de estrategias de venta de calidad al usuario. Todo esto es de gran valor para los comercios puesto que entrega información de las características de los usuarios consumidores y no consumidores ya que logra transformar su estrategia comercial en una posible atracción de nuevos clientes y compromiso con sus compradores más habituales.

5 Resultados

Como se presentó en la sección anterior se implementó el algoritmo Alternating Least Squares desarrollando la metodología propuesta por (Hu et al., 2008), para implementar un sistema de recomendación utilizando información implícita, basada en la interacción de un cliente con un comercio. Al no tener una retroalimentación explícita del usuario no entramos a detallar mediante encuestas el rendimiento en las estimaciones, por lo tanto, solo utilizamos la métrica

de evaluación ROEM como una estrategia para garantizar que la generación de más de setenta mil millones de recomendaciones, estuvieran respaldadas por un modelo con un error inferior al 50% para que fueran mejor que una decisión aleatoria como sugiere la literatura. El modelo construido para el desarrollo de este trabajo tuvo un ROEM de 38.78%, por lo tanto, este modelo es mejor que entregar una oferta a un cliente de manera aleatoria.

La calificación implícita de un sistema de recomendación dificulta la tarea de comprender la posible estimación del cliente a una interacción porque no existe una retroalimentación específica y entrar a evaluar las recomendaciones generadas se vuelve una tarea difícil, por lo tanto, la manera en la que se evaluaron las recomendaciones fue identificando los comercios más afines al cliente utilizando la información de entrenamiento, el perfil transaccional y los comercios más relevantes para el cliente y evaluar la calificación entregada por el modelo versus las interacciones ya conocidas por el cliente. las sugerencias mejor puntuadas del modelo estaban basadas en comercios similares.

Los resultados encontrados permitieron identificar que la similitud de las recomendaciones con las costumbres de los clientes están altamente relacionadas, por lo tanto, se identificó que si un usuario acostumbraba a visitar almacenes de cadena las calificaciones proporcionadas por la metodología para este tipo de comercios son coherentes con las necesidades y costumbres del cliente, de hecho, en varias ocasiones se identificó recomendaciones calificadas positivamente habían sido preferencias del cliente en otras ventanas de tiempo no consideradas en los datos de entrenamiento del modelo.

Para este trabajo las recomendaciones son ofertas y la relación entre el negocio y la fuerza comercial de la institución financiera debe ser transparente con el fin de poder garantizar una distribución que no genere fricción en el servicio. Los clientes son el centro y las ofertas son el insumo para generar interacciones, que apalanquen las ventas de los comercios y las necesidades de los consumidores, por lo tanto, es importante que se definan límites como: número de ofertas presentadas, periodicidad de las ofertas, tipo de ofertas y horarios para presentar, para no generar una experiencia negativa en la comunicación del servicio.

La alta posibilidad de generar recomendaciones para todos los usuarios llevó a tomar decisiones de negocio para la presentación de ofertas a los clientes, en un proceso posterior a la ejecución del modelo filtrando las recomendaciones generadas para cada cliente. Las ofertas solo deben ser entregadas a los clientes cuando el equipo comercial establezca una estrategia que habilite una campaña comercial de cara a la relación entre clientes y comercios.

La prioridad para la presentación está dada en dos momentos, el primero consiste en utilizar a los clientes que han tenido una interacción con el comercio y utilizar la frecuencia de esta para priorizar los clientes que entran en una campaña. Los usuarios que no han tenido alguna interacción con dicho comercio serán priorizados por la estimación del modelo de recomendación de mayor a menor calificación, los que no han sido participes en el entrenamiento del modelo serán asociados a las alertas mediante la segmentación que se tiene en la organización junto al alcance geográfico del negocio.

6 Conclusiones

Como se ha planteado a lo largo del documento, obtener una buena recomendación sobre las posibles ofertas presentadas a un usuario es fundamental para la relación entre la entidad financiera, los comercios y el cliente. En general, las recomendaciones ayudan al negocio a tomar decisiones apropiadas gracias a las ofertas acertadas de cara al cliente y por ende al consumidor,

este tipo de acciones no solo ayudarán a reducir costos a los ofertantes para que el cliente sienta que le están llegando las mejores opciones logrando que la consideración de compra sea mayor y aumente la relación de confianza entre el cliente y la organización.

Las recomendaciones pueden verse como una acción intrusiva hacia el consumidor, por lo tanto, es muy importante que el sistema de recomendación esté acompañado por un equipo comercial y de mercadeo que entienda las necesidades del cliente para poner límites a las campañas y plantear alcances orientados al cliente buscando siempre mantener los niveles de experiencia con la satisfacción y recomendación del cliente hacia la organización.

En todo este proceso, y cómo se ha visto a lo largo de esta investigación, el cliente persona, es el centro de todo y los datos de su comportamiento en temas de consumo cumplen el papel más relevante como insumo principal al momento de generar recomendaciones. Es por esto, que uno de los factores más importantes a tener en cuenta, es buscar la manera adecuada para integrar a este usuario final en el proceso haciéndole entender de una manera sencilla su importancia dentro del modelo y generar disposición y consentimiento por su parte a la hora de brindar o permitir que terceros puedan acceder a esta información.

Para obtener los datos de clientes con su consentimiento, es fundamental actuar de manera transparente, explicando de una forma muy accesible el panorama completo de cómo al permitirle al negocio conocer su información puede resultar beneficioso para él y sus futuras decisiones de compra, en este proceso es necesario crear una integración de varias áreas como la de análisis de datos, jurídico, mercadeo y comercial para entender los límites y crear diferentes estrategias que lograrán convencer al usuario y convertirlo en un aliado humano más allá de un insumo prioritario.

El trabajo futuro que le espera a los sistemas de recomendación con calificación implícita es poder identificar una metodología que permita calificar la relación entre el cliente y el producto para poder generar mejores recomendaciones basadas en dicha calificación, conociendo los factores que determinan una relación entre los consumidores. Los bancos son compañías que recopilan mucha información sobre los hábitos de consumo de sus clientes, por lo tanto, si se puede capturar una retroalimentación en el momento de acceder a una recomendación permitiría tener insumos en otras variables y que pueden determinar la capacidad en la relación del cliente.

References

- Abdollahpouri, H., & Abdollahpouri, A. (2013). An approach for personalization of banking services in multi-channel environment using memory-based collaborative filtering. In *The 5th conference on information and knowledge technology* (pp. 208–213).
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734–749.
- Aggarwal, C. C., et al. (2016). *Recommender systems* (Vol. 1). Springer.
- Alvarado, O., Heuer, H., Vanden Abeele, V., Breiter, A., & Verbert, K. (2020). Middle-aged video consumers' beliefs about algorithmic recommendations on youtube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1–24.

- Bennett, J., Lanning, S., et al. (2007). The netflix prize. In *Proceedings of kdd cup and workshop* (Vol. 2007, p. 35).
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, *46*, 109–132.
- Breese, J. S., Heckerman, D., & Kadie, C. (2013). Empirical analysis of predictive algorithms for collaborative filtering. *arXiv preprint arXiv:1301.7363*.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331–370.
- Gil, H. A. P., Castro, K. A. C., Bermúdez, G. M. T., et al. (2017). La brecha digital en colombia: Un análisis de las políticas gubernamentales para su disminución. *Redes De Ingeniería*, 59–71.
- Gomez-Uribe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 1–19.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5–53.
- Hernández-Nieves, E., Hernández, G., Gil-González, A.-B., Rodríguez-González, S., & Corchado, J. M. (2020). Fog computing architecture for personalized recommendation of banking products. *Expert Systems with Applications*, *140*, 112900. doi: <https://doi.org/10.1016/j.eswa.2019.112900>
- Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *2008 eighth ieee international conference on data mining* (pp. 263–272).
- Huckman, R. S., Pisano, G. P., & Kind, L. (2008). Amazon web services.
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.
- Kindström, D. (2010). Towards a service-based business model—key aspects for future competitive advantage. *European management journal*, *28*(6), 479–490.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, *40*(3), 77–87.
- Koren, Y. (2009). The bellkor solution to the netflix grand prize. *Netflix prize documentation*, *81*(2009), 1–10.
- Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, *7*(1), 76–80.
- Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73–105). Springer.
- Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics reports*, *519*(1), 1–49.
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández Orallo, J., Kull, M., Lachiche, N., ... Flach, P. A. (2019). Crisp-dm twenty years later: From data

- mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 1-1. doi: 10.1109/TKDE.2019.2962680
- Mehta, R., & Rana, K. (2017). A review on matrix factorization techniques in recommender systems. In *2017 2nd international conference on communication systems, computing and it applications (cscita)* (pp. 269–274).
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... others (2016). Mlib: Machine learning in apache spark. *The Journal of Machine Learning Research*, 17(1), 1235–1241.
- Meyer, C., Schwager, A., et al. (2007). Understanding customer experience. *Harvard business review*, 85(2), 116.
- Oyebode, O., & Orji, R. (2020). A hybrid recommender system for product sales in a banking environment. *Journal of Banking and Financial Technology*, 1–11.
- Ramos Pérez, J. F., et al. (2021). Digitalización e inclusión financiera en colombia durante la pandemia 2020.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 acm conference on computer supported cooperative work* (pp. 175–186).
- Sayyed, F., Argiddi, R., & Apte, S. (2013). Generating recommendations for stock market using collaborative filtering. *Int. J. Comput. Eng. Sci*, 3, 46–49.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291–324). Springer.
- Sharifhosseini, A. (2019, 10). A case study for presenting bank recommender systems based on bon card transaction data. In (p. 72-77). doi: 10.1109/ICCKE48569.2019.8964698
- Smith, B., & Linden, G. (2017). Two decades of recommender systems at amazon. com. *Ieee internet computing*, 21(3), 12–18.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.
- Thorat, P. B., Goudar, R., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4), 31–36.
- Wirth, R., & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29–39).
- Yahyapour, N. (2008). *Determining factors affecting intention to adopt banking recommender system: case of iran*.
- Yu, T. R., Song, X., et al. (2020). Big data and artificial intelligence in the banking industry. *World Scientific Book Chapters*, 4025–4041.
- Zhang, T., & Iyengar, V. S. (2002). Recommender systems using linear classifiers. *Journal of machine learning research*, 2(Feb), 313–334.
- Zhao, Z.-D., & Shang, M.-S. (2010). User-based collaborative-filtering recommendation algorithms on hadoop. In *2010 third international conference on knowledge discovery and data mining* (pp. 478–481).