

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“SISTEMA DE RECOMENDACIÓN PARA RUBROS DE
CONSUMO EN UNA ENTIDAD FINANCIERA USANDO
KNN RECOMMENDER”**

**TRABAJO DE SUFICIENCIA PROFESIONAL
PARA OPTAR TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO**

EBSON DAVID ALLENDE QUINTANA

LIMA – PERÚ

2024

TSP_DAQ_vs25042024.docx

INFORME DE ORIGINALIDAD

21 %
INDICE DE SIMILITUD

20 %
FUENTES DE INTERNET

6 %
PUBLICACIONES

7 %
TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	notebook.community Fuente de Internet	3 %
2	repositorio.lamolina.edu.pe Fuente de Internet	3 %
3	Submitted to Universidad Nacional Agraria La Molina Trabajo del estudiante	1 %
4	hdl.handle.net Fuente de Internet	1 %
5	purl.org Fuente de Internet	1 %
6	eprints.ucm.es Fuente de Internet	1 %
7	sedici.unlp.edu.ar Fuente de Internet	1 %
8	ebin.pub Fuente de Internet	1 %
9	link.springer.com Fuente de Internet	

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN**

**“SISTEMA DE RECOMENDACIÓN PARA RUBROS DE
CONSUMO EN UNA ENTIDAD FINANCIERA USANDO
KNN RECOMMENDER”**

**PRESENTADO POR:
EBSON DAVID ALLENDE QUINTANA**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR
TÍTULO DE INGENIERO ESTADÍSTICO E INFORMÁTICO**

SUSTENTADO Y APROBADO ANTE EL SIGUIENTE JURADO:

.....
Dr. Rino Nicanor Sotomayor Ruiz

PRESIDENTE

.....
Mg. Jesús Walters Salinas Flores

ASESOR

.....
DR. Cesar Higinio Menacho Chiok

MIEMBRO

.....
MS. Jaime Carlos Porras Cerrón

MIEMBRO

Lima – Perú

2024

DEDICATORIA

La realización del presente trabajo de suficiencia profesional va dedicada a mis principales motores de vida: mis padres, que me apoyaron incansablemente en mi formación universitaria; mi compañera de vida, que me apoya día a día a continuar creciendo como persona y profesional; mis hijos que me motivan día a día a ser una mejor persona y un mejor profesional.

AGRADECIMIENTOS

Agradecimientos para el profesor Jesús Salinas, quién además de ser mi amigo en la actualidad, fue mi profesor durante mi formación universitaria y siempre me motiva a plasmar mis conocimientos sobre la aplicación de la estadística a los negocios; agradezco también a mis compañeros de actividades profesionales que nutrieron mi conocimiento de la realidad y contribuyeron a la implementación de las soluciones analíticas en las que he participado.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
1.1.	Problemática.....	1
1.2.	Objetivos	1
1.2.1.	Objetivo general	1
1.2.2.	Objetivos específicos.....	2
II.	REVISIÓN DE LITERATURA.....	3
III.	DESARROLLO DEL TRABAJO	13
3.1.	Funciones desempeñadas	13
3.2.	Puesta en práctica de lo aprendido en la carrera	13
3.2.1.	Descripción de las técnicas estadísticas y/o informáticas utilizadas en la solución de la situación problemáticas en el ejercicio de su actividad laboral.	13
3.2.2.	Propuesta de alternativa de solución a la situación problemática siguiendo las seis fases de la Metodología para el Desarrollo de Proyectos	18
3.3.	Contribución en la solución de situaciones problemáticas.....	23
3.4.	Análisis de la contribución en términos de competencias y habilidades	25
IV.	RESULTADOS Y DISCUSIÓN	26
4.1.	Nivel de beneficio obtenido por el centro laboral	26
V.	CONCLUSIONES	29
VI.	RECOMENDACIONES.....	30
VII.	REFERENCIAS BIBLIOGRÁFICAS.....	31
VIII.	ANEXOS	32

ÍNDICE DE TABLAS

Tabla 1 Ejemplo de matriz de preferencias para un sistema de recomendación de filtrado colaborativo basado en usuarios	15
Tabla 2 Puntuación promedio de cada usuario.....	16
Tabla 3 Diferencia de la puntuación por cada ítem respecto a la puntuación promedio de cada usuario	16
Tabla 4 Diferencia de la puntuación por cada ítem respecto a la puntuación promedio de cada usuario elevado al cuadrado	17
Tabla 5 Similaridad entre los usuarios	17
Tabla 6 Evolutivo del ratio de apertura de las campañas por correos electrónicos.....	23
Tabla 7 Evolutivo del ratio de conversión de las campañas por correos electrónicos	24
Tabla 8 Evolutivo del ratio de apertura de las campañas por banca móvil	26
Tabla 9 Extrapolación del evolutivo del ratio de apertura de las campañas por correos electrónicos.....	27
Tabla 10 Evolutivo del ratio de conversión de las campañas por banca móvil.....	27
Tabla 11 Extrapolación del evolutivo del ratio de conversión de las campañas por correos electrónicos	28

ÍNDICE DE FIGURAS

Figura 1 Sagemaker como servicio de desarrollo de modelos en AWS.....	20
Figura 2 Matriz de evaluación de recomendaciones hechas por un sistema	21
Figura 3 Ejemplo de diseño de experimento	21
Figura 4 Arquitectura propuesta para la personalización de los servicios de los clientes de un banco.....	22

ÍNDICE DE ANEXOS

Anexo 1 Código Python usado en el entrenamiento del algoritmo KNN Recommender ...	32
Anexo 2 Marco de categorías de roles laborales por el nivel de rigurosidad analítica por Royal Society. (2019).....	34

RESUMEN

El presente trabajo de suficiencia profesional de Estadística e Informática se desarrolló con la información de un banco del sistema financiero peruano, banco que se ubica en los primeros lugares por su alta participación en la colocación de créditos directos en el sistema financiero peruano. Los directores orientaron sus esfuerzos y estrategias a transformar digitalmente al banco, en el proceso de transformación digital, uno de los pilares es la personalización de la comunicación con los clientes, comunicación tanto de ofertas como de beneficios, con la transformación digital, el banco mejoró en la aceptación de sus campañas (mejores ratios de efectividad de campañas), en la satisfacción de sus clientes (mayor NPS - Net Promote Score) y en la bancarización de más peruanos (crecimiento del 10% respecto al año anterior). El banco se apoyó en el uso de una herramienta estadística que le permitió direccionar sus acciones comerciales y de comunicación de manera personalizada.

El rol que se desempeñó en el banco tuvo la misión de crear la herramienta estadística capaz de personalizar la comunicación de beneficios del banco a sus clientes, para lograrlo se desarrollaron las siguientes tareas: relevamiento de la problemática del negocio con los directores del banco, diagnóstico de la situación del negocio, mapeo de las fuentes de datos disponibles y necesarias para el desarrollo de la herramienta estadística, estructuración de las fuentes de datos, entrenamiento y validación del algoritmo de personalización KNN Recommender, medición del impacto al negocio, presentación al negocio y acompañamiento en la implementación de la solución en los sistemas digitales del banco.

Palabras clave: Transformación digital; Personalización; Efectividad de Campañas; Net Promote Score; Bancarización; KNN-Recommender.

ABSTRACT

This work on professional proficiency in Statistics and Computer Science was developed with information from a bank in the Peruvian financial system, a bank that is in the first places due to its high participation in the placement of direct credits in the Peruvian financial system. The directors directed their efforts and strategies to digitally transform the bank. In the digital transformation process, one of the pillars is the personalization of communication with customers, communication of both offers and benefits. With digital transformation, the bank improved in the acceptance of their campaigns (better campaign effectiveness ratios), in the satisfaction of their customers (higher NPS - Net Promote Score) and in the banking of more Peruvians (10% growth compared to the previous year). The bank relied on the use of a statistical tool that allowed it to direct its commercial and communication actions in a personalized way.

The role he played in the bank had the mission of creating the statistical tool capable of personalizing the communication of the bank's benefits to its clients. To achieve this, the following tasks were developed: survey of the business problems with the bank directors, diagnosis of the business situation, mapping of the data sources available and necessary for the development of the statistical tool, structuring of the data sources, training and validation of the KNN Recommender personalization algorithm, measurement of the impact on the business, presentation to the business and support in the implementation of the solution in the bank's digital systems.

Keywords: Digital transformation; Personalization; Campaign Effectiveness; Net Promote Score; Banking; KNN-Recommender.

I. INTRODUCCIÓN

1.1. Problemática

En los últimos años, los empresarios peruanos identificaron la necesidad de transformar sus organizaciones en digitales, en distintos foros empresariales como por ejemplo en el CADE (Conferencia Anual de Ejecutivos) digital se plasmó la necesidad de transformar las organizaciones, hacerlas digitales y contar con la capacidad de personalizar la experiencia de sus clientes. En esta coyuntura, las herramientas estadísticas tomaron un rol protagónico. La transformación digital se ha posicionado en la agenda inmediata de muchas organizaciones. El banco en el que se llevó a cabo este estudio fue una de esas instituciones. Los antecedentes de este banco muestran que durante varios años se gestionó las relaciones con sus clientes mediante campañas masivas que se caracterizaban por su falta de eficacia y eficiencia, lo que resultaba en una repercusión negativa para la marca del banco. Numerosos clientes se encontraban insatisfechos debido al envío masivo de correos electrónicos o mensajes de texto muy genéricos, la aceptación de campañas desplegadas por correo o mensaje de texto tenían resultados históricos por debajo de lo esperado.

El objetivo del presente trabajo de suficiencia profesional fue elaborar un sistema de recomendación usando el algoritmo KNN Recommender basándose en las transacciones efectuadas en los distintos rubros de consumo y permita personalizar el ofrecimiento de beneficios en el segmento de clientes más rentable del banco.

1.2. Objetivos

1.2.1. Objetivo general

Elaborar un sistema de recomendación de rubros de consumo para el segmento rentable de clientes del banco usando el algoritmo KNN Recommender.

1.2.2. Objetivos específicos

- Identificar entre los clientes del banco quienes pertenecen al segmento más rentable.
- Identificar los rubros de consumo atípicos a nivel de transacciones y a nivel monetario.
- Crear un ponderador a nivel de rubros de consumo que acompañe la recomendación del algoritmo.
- Medir el impacto de la recomendación personalizada en términos de aceptación y conversión.

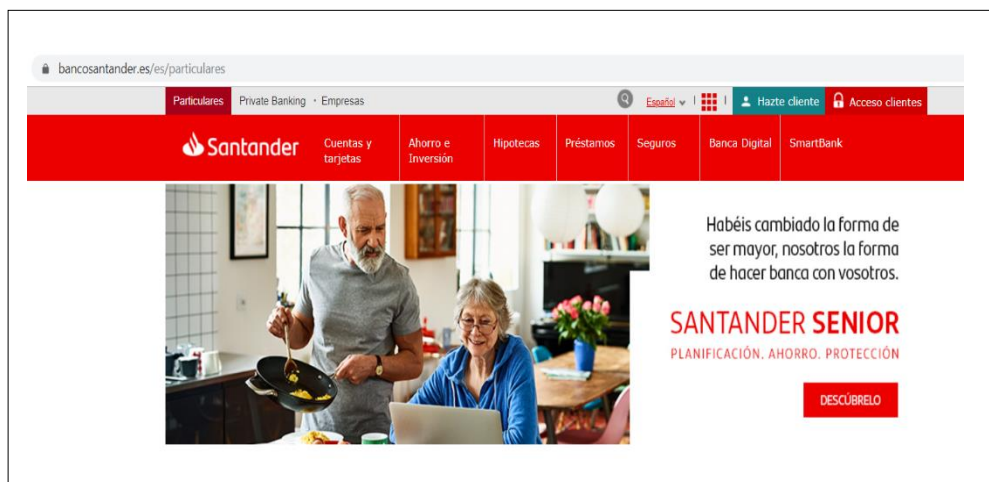
II. REVISIÓN DE LITERATURA

2.1. Antecedentes

En el presente trabajo de suficiencia profesional se tomó como referencia a los siguientes bancos, que han implementado algoritmos de sistemas de recomendación en sus canales digitales para el ofrecimiento de sus ofertas personalizadas: Santander (España) en la banca personas, Nidwaldner Kantonalbank (Alemania) en la banca personas y el Banco de Crédito (Perú) en el negocio de personas. El banco Santander tiene como oferta para sus clientes y no clientes un amplio repertorio de productos: 7 tipos de cuentas corrientes, 8 tipos de tarjetas de crédito, 8 tipos de tarjetas de débito, 10 tipos de fondos de inversión, 5 tipos de hipotecas, 7 tipos de préstamos personales y 8 tipos de seguros. En la Figura 1, se tiene el detalle de cada una de las categorías de los productos que el banco Santander tiene disponible para sus clientes y no clientes; la forma tradicional de mostrar a detalle cada uno de sus productos no considera la personalización de ofertas para cada uno de sus clientes y/o no clientes; por el contrario, es una manera estándar de mostrar la oferta disponible para todos los clientes y no clientes.

Figura 1

Página web versión genérica para la banca personas

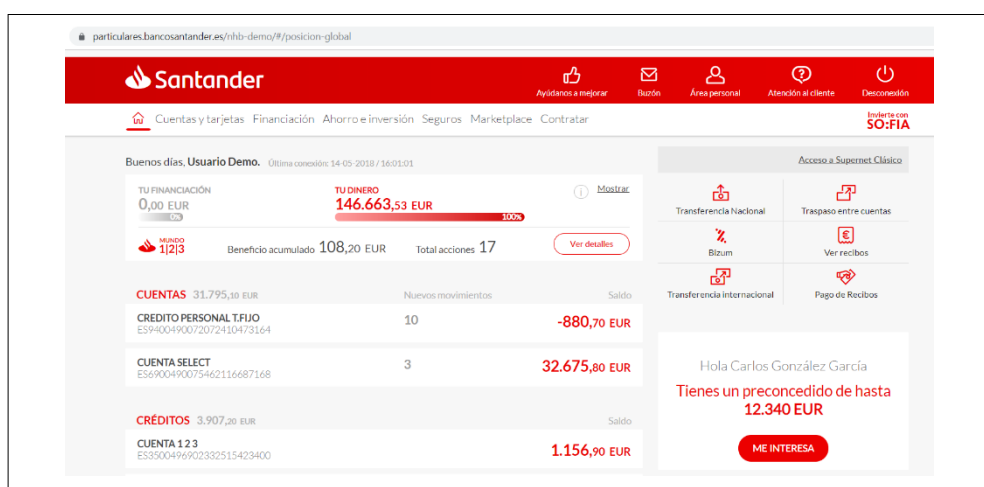


FUENTE: Banco Santander (España)

El banco Santander desarrolló e implementó algoritmos de sistemas de recomendación para que sus clientes puedan ingresar a la banca online y puedan visualizar el mejor producto que el banco tenga para ofrecerle basado en sus patrones de consumo e información sociodemográfica. En la Figura 2 se muestra la recomendación que un usuario genérico como Carlos Gonzales García puede recibir producto de la implementación de los algoritmos de sistemas de recomendación.

Figura 2

Banca Online versión modo prueba para la banca personas

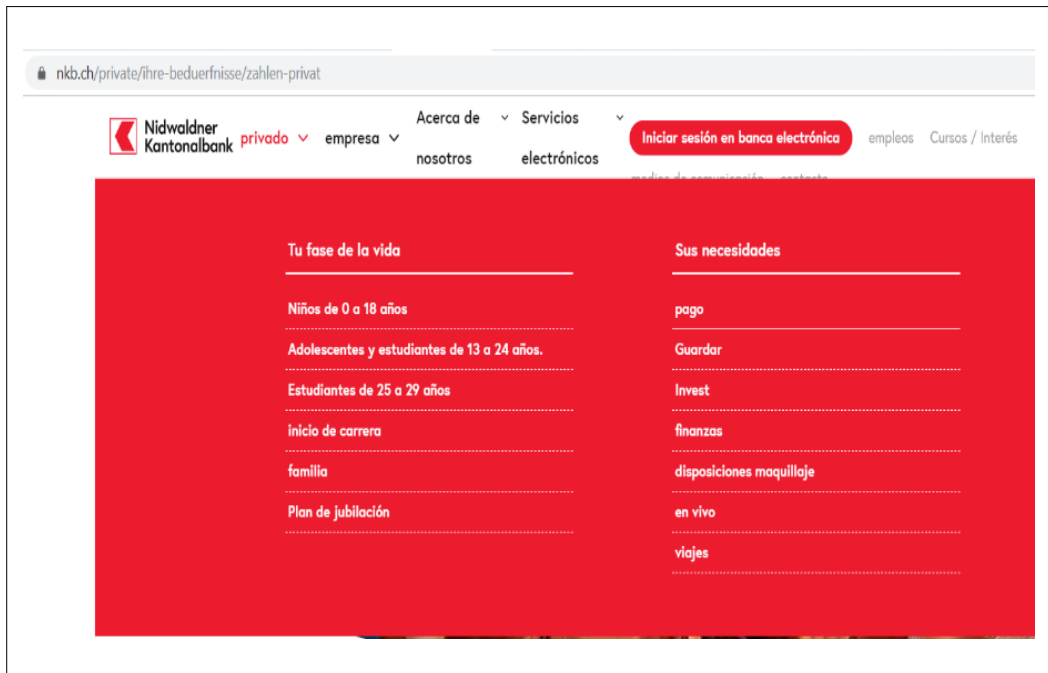


FUENTE: Banco Santander (España)

El banco Nidwaldner Kantonalbank en la banca personas tiene para sus clientes y no clientes ofrecimientos masivos que decantan en aproximadamente 100 productos distribuidos en las siguientes categorías: cuentas corrientes, cuentas de ahorro, tarjetas de crédito, cuentas de inversión y préstamos personales, en la Figura 3 se muestra un catálogo genérico que dependiendo de la necesidad de la persona la página web del banco te direcciona a enlaces con mayor información de cada uno de los 100 productos mencionados.

Figura 3

Página web versión genérica para la banca personas

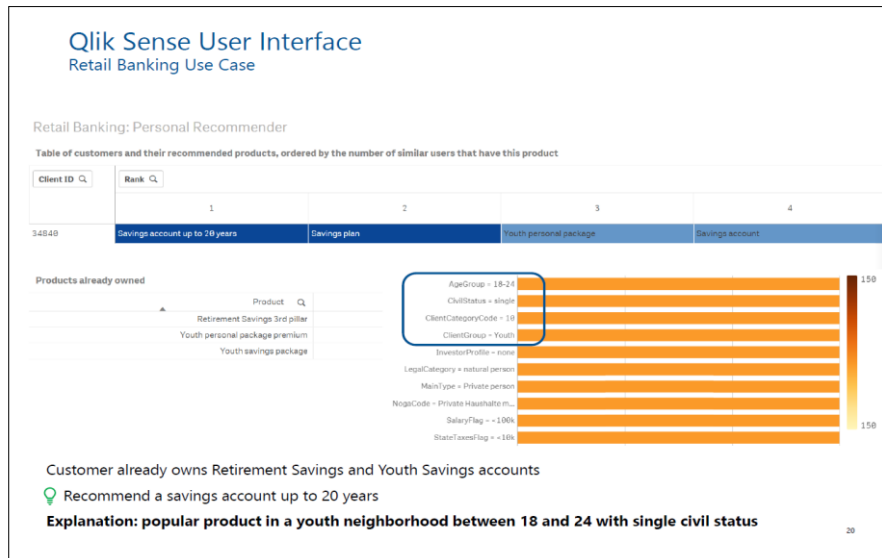


FUNTE: Banco Nidwaldner Kantonbank (Alemania)

Para lograr la personalización de sus ofertas entre sus clientes, el banco alemán implementó un motor de recomendación, en la Figura 4 se muestra la información que proporciona el motor de recomendación, considera la tenencia de productos financieros, enumera los cuatro mejores productos recomendados y una descripción de los motivos de la recomendación basándose principalmente en características sociodemográficas.

Figura 4

Prototipo de motor de recomendación de Nidwaldner Kantonalbank



FUENTE: InCube Group (Suiza)

El banco de Crédito del Perú a través de su aplicativo móvil ParaTi desarrolló un motor de recomendación basándose en los rubros de consumo más frecuentes que presenta la comunidad de usuarios. En la Figura 5 se muestra la recomendación que un cliente del banco de crédito tiene asignado como primera y segunda opción: descuentos en accesorios tecnológicos y restaurantes respectivamente.

Figura 5

Recomendación en el aplicativo móvil ParaTi



FUENTE: Banco de Crédito (Perú)

2.2. Sistema bancario en el Perú

En el Perú, el sistema bancario está compuesto por 17 entidades privadas y 2 gubernamentales. En los últimos años, APOYO & ASOCIADOS INTERNACIONALES S.A.C. (2022) mencionan que alrededor del 80% del total de los créditos directos han sido registrados por 4 bancos, entre ellos están el Banco de Crédito del Perú, BBVA, Interbank y Scotiabank.

2.3. Transformación digital y la personalización en la banca

Muchas organizaciones transformaron sus procesos de negocio en digitales tomando como referencia las estrategias tecnológicas y de transformación digital usadas por empresas de comercio electrónico como Amazon o empresas de entretenimiento digital como Netflix, que usan algoritmos de sistemas de recomendación para personalizar sus ofertas y la comunicación con sus clientes.

Chirkina y Rankov (2018) definen los retos de la banca para conseguir la personalización a los clientes de los bancos:

- No existen datos de retroalimentación directa, a diferencia de Amazon o Netflix, los clientes de la banca no otorgan calificaciones a los productos financieros que compran o su banco les propone.
- Las características de los productos financieros cambian a lo largo del tiempo.
- El impacto financiero de las recomendaciones.

A pesar de los desafíos mencionados, los clientes de la banca esperan ofrecimientos personalizados, ante ello los sistemas de recomendación son los más apropiados.

2.4. Sistemas de recomendación

Burke *et al.* (2011) mencionan que la definición de un sistema de recomendación ha evolucionado a través de los últimos 14 años. Resnick y Varian (1997) describen a los sistemas de recomendación como soporte de colaboración entre usuarios, donde los usuarios proporcionan recomendaciones como entradas, que luego el sistema agrega y dirige a los destinatarios apropiados. En algunos casos la transformación primaria está en el agregación; en otros, el valor del sistema reside en su capacidad de hacer buenas coincidencias entre los recomendadores y aquellos que buscan recomendaciones. Burke (2002) expande la definición a cualquier sistema que produce recomendaciones individualizadas o tiene el efecto de guiar al usuario de forma personalizada a ítems interesantes o útiles respecto a un gran espacio de posibles opciones. La definición de Burke fue formalizada y generalizada por Adomavicius y Tuzhilin en la que se resalta que para cada usuario se busca elegir el ítem que permita maximizar la utilidad, posteriormente en 2005. Los autores son cuidadosos al mencionar que el objetivo es recomendar al ítem con mayor utilidad, no necesariamente la utilidad es calculada de manera explícita.

Además, detallan que los principios fundamentales de los sistemas de recomendación son:

- El sistema de recomendación es personalizado, las recomendaciones están destinadas a optimizar la experiencia de un usuario y no para representar el consenso de un grupo.
- El sistema de recomendación está destinado a ayudar al usuario y que este seleccione entre opciones discretas. En general los artículos ya se conocen de antemano y no se generan con respecto a alguna tendencia.

Basado en cómo el sistema analiza y filtra los datos de acuerdo a los requerimientos, Gigimol y Sincy (2016) mencionan que hay diferentes tipos de sistemas de recomendación como los basados en contenido, basados en el filtrado colaborativo, híbridos, conscientes del contexto y basados en el conocimiento.

De los diferentes tipos de sistemas de recomendación, el presente trabajo de suficiencia profesional se basó en el sistema de recomendación de filtrado colaborativo basado en usuarios.

Jannach *et al.* (2010) mencionan que los sistemas de recomendación de filtrado colaborativo tienen como objetivo explotar información del comportamiento pasado o las opiniones de los usuarios de una comunidad para predecir en qué ítems el usuario actual del sistema pueda estar probablemente interesado. Estos tipos de sistemas hoy son muy utilizados a nivel industrial, en particular como herramientas en sitios online de consumo masivo para personalizar el contenido de las necesidades de un determinado cliente o promocionar ítems adicionales e incrementar las ventas.

Gigimol y Sincy (2016) mencionan que los sistemas de recomendación basados en el filtrado colaborativo dependen de juntar y analizar los datos de las preferencias de los usuarios y hacer una predicción de los ítems que le deberían gustar en base a la similitud que existe con otros usuarios, dicha similitud es calculada por comparación con puntajes de otros usuarios del mismo ítem. Además, mencionan que existen dos tipos de filtro colaborativo: los que se basan en ítems y los que se basan en usuarios.

Prasad y Sarath (2011) presentan una descripción detallada del algoritmo de filtrado colaborativo basado en usuarios, en el que mencionan que el algoritmo genera una lista de recomendaciones por usuario a nivel de ítems de acuerdo con la vista de otros usuarios. Los supuestos por los que parte son si las calificaciones de algunos ítems calificados por algunos usuarios son similares, la calificación de otros ítems calificados por estos usuarios también será similar. El sistema de recomendación de filtrado colaborativo basado en usuarios utiliza técnicas estadísticas para buscar los vecinos más cercanos del usuario-ítem (usuario al que se busca generar la recomendación), basarse en la calificación del ítem por los vecinos más

cercanos para predecir la calificación del ítem al usuario recomendado, y luego producir la lista de recomendaciones correspondiente.

El componente de filtrado colaborativo que utiliza un algoritmo basado en los vecinos cercanos busca un subconjunto de usuarios que se eligen en función de su similitud con el usuario-ítem (usuario al que se busca generar la recomendación) y una combinación ponderada de las calificaciones de su comunidad cercana se utilizan para generar predicciones para el usuario-ítem.

2.5. Definiciones aplicadas

IBM (2021) presenta una descripción detallada de una metodología probada para abordar proyectos de minería de datos, metodología CRISP-DM cuyo nombre responde a las siglas Cross-Industry Standard Process for Data Mining, la metodología está compuesta por 6 fases que son la comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación, las fases tienen dependencia entre ellas, a pesar de ello la metodología es flexible y puede ser personalizada según la necesidad del negocio. El presente trabajo de suficiencia profesional se desarrolló bajo el paraguas de la metodología CRISP-DM.

2.5.1. Comprensión del negocio

IBM (2021) presenta a la primera fase de la metodología CRISP-DM, fase de comprensión del negocio, fase que es importante para identificar las expectativas de la organización a través de sus principales directivos, mapeados los objetivos comerciales de la organización es necesario traducirlos en criterios de aprobación del rendimiento del proceso de minería de datos. En esta fase es importante identificar los factores de riesgo del proyecto y los planes de contingencia para ellos.

2.5.2. Comprensión de los datos

IBM (2021) presenta a la segunda fase de la metodología CRISP-DM, fase de comprensión de los datos, fase que permite determinar la calidad de los datos que serán usados para el proyecto, en esta fase se describen los datos y se busca mitigar potenciales problemas a presentarse en la fase de preparación de los datos.

En esta fase es importante identificar a los datos disponibles y no disponibles para el proyecto, describir a los datos según la cantidad de datos y la clasificación de los datos según su tipo (cuantitativos o cualitativos).

2.5.3. Preparación de los datos

IBM (2021) presenta a la tercera fase de la metodología CRISP-DM, fase de preparación de los datos, fase que es de mucha importancia dado que en esta etapa se refleja la realidad del negocio en un conjunto de datos que será el insumo principal de la etapa de modelado, en la etapa de preparación de datos se fusionan archivos independientes, seleccionan casos por criterios técnicos y/o de negocio, agregan casos y se derivan variables en nuevas variables. Esta etapa suele ser la que mayor tiempo del proyecto demanda.

2.5.4. Modelado

IBM (2021) presenta a la cuarta fase de la metodología CRISP-DM, fase de modelado de los datos, en esta fase debemos partir por identificar el rol de las variables a ser analizadas, variables que fueron construidos en la etapa previa, etapa de preparación del dato.

Por ejemplo, buscamos predecir a través de una recomendación personalizada la siguiente elección del cliente de una organización, si contamos con información como las elecciones históricas de los clientes podemos afirmar que el rol de esta variable es el de la variable objetivo, además, si existen variables adicionales como edad, estado civil, entre otras, sobre las que tenemos hipótesis relevadas con el negocio podemos afirmar que estas variables son predictoras o característica.

Teniendo identificado el rol de las variables, definimos qué técnicas de modelado son las más adecuadas, aprendizaje supervisado que cuentan con una variable objetivo en función de predictoras y el aprendizaje no supervisado que no cuenta con una variable objetivo pero si cuenta con variables predictoras o característica.

Dependiendo de la elección de la técnica de modelado habrán requerimientos o asunciones a tomar en cuenta para que el modelo aprenda a identificar patrones recurrentes y pueda generar predicciones futuras.

2.5.5. Evaluación

IBM (2021) presenta a la quinta fase de la metodología CRISP-DM, fase de evaluación de los resultados a través de indicadores técnicos de los modelos y resultados comerciales del negocio, es importante contar con una estrategia de monitoreo y calibración de modelos según los hallazgos o alertas de la medición de los indicadores mencionados.

2.5.6. Despliegue

IBM (2021) presenta a la sexta fase de la metodología CRISP-DM, fase del despliegue, hallados los patrones en los datos estructurados en etapas previas, es necesario contar con una estrategia de integración de los resultados de los modelos a los procesos de negocio, de esta forma resulta más sencillo contar con la adopción comercial del uso de los modelos.

III. DESARROLLO DEL TRABAJO

3.1. Funciones desempeñadas

Las funciones desempeñadas en la entidad financiera del presente trabajo de suficiencia profesional fueron:

- Identificar la problemática del negocio es crucial para un diagnóstico preciso. Posteriormente, se promueve un enfoque de co-creación con el negocio para desarrollar una solución adaptada. Se propone entonces la creación de una herramienta estadística específica para abordar las necesidades identificadas.
- Se debe determinar qué algoritmos, como KNN, son candidatos para la recomendación y el tipo de aprendizaje apropiado para la solución.
- Definir las fuentes de datos disponibles y necesarias para el desarrollo de la herramienta estadística.
- Se procede a entrenar el algoritmo seleccionado como óptimo dentro de la solución propuesta.
- Validar que el algoritmo cumpla con los criterios mínimos establecidos.
- Presentar la herramienta estadística al negocio.
- Acompañar en la implementación de la herramienta estadística en los sistemas digitales del banco.

3.2. Puesta en práctica de lo aprendido en la carrera

3.2.1. Descripción de las técnicas estadísticas y/o informáticas utilizadas en la solución de la situación problemáticas en el ejercicio de su actividad laboral

La definición de Burke (2002) fue formalizada y generalizada por Adomavicius y Tuzhilin en la que se resalta que para cada usuario se busca elegir el ítem que permita maximizar la utilidad, posteriormente en 2005. Los autores son cuidadosos al mencionar que el objetivo es recomendar al ítem con mayor utilidad, no necesariamente la utilidad es calculada de manera explícita.

Además, detallan que los principios fundamentales de los sistemas de recomendación son:

- El sistema de recomendación es personalizado, las recomendaciones están destinadas a optimizar la experiencia de un usuario y no para representar el consenso de un grupo.
- El sistema de recomendación está destinado a ayudar al usuario y que este seleccione entre opciones discretas. En general los artículos ya se conocen de antemano y no se generan con respecto a alguna tendencia.

De los diferentes tipos de sistemas de recomendación, el presente trabajo de suficiencia profesional se basó en el sistema de recomendación de filtrado colaborativo basado en usuarios.

El sistema de recomendación de filtrado colaborativo basado en usuarios utiliza técnicas estadísticas para buscar los vecinos más cercanos del usuario-ítem (usuario al que se busca generar la recomendación), basarse en la calificación del ítem por los vecinos más cercanos para predecir la calificación del ítem al usuario recomendado, y luego producir la lista de recomendaciones correspondiente.

El componente de filtrado colaborativo que utiliza un algoritmo basado en los vecinos cercanos busca un subconjunto de usuarios que se eligen en función de su similitud con el usuario-ítem (usuario al que se busca generar la recomendación) y una combinación ponderada de las calificaciones de su comunidad cercana se utilizan para generar predicciones para el usuario-ítem.

El algoritmo se puede resumir en los siguientes pasos:

Paso 1. Todos los usuarios son ponderados con respecto a la similitud con usuario-ítem (usuario al que buscamos generar la recomendación).

La similitud entre usuarios se mide con el coeficiente de correlación de Pearson entre sus vectores de calificación.

Paso 2. Se selecciona a los n usuarios que tienen una alta similaridad con el usuario-ítem (usuario al que buscamos generar la recomendación) usando el algoritmo de los vecinos más

cercanos.

Paso 3. Se calcula la predicción $P_{a,u}$ desde una combinación ponderada. La similaridad entre dos usuarios es calculada usando el coeficiente de correlación de Pearson.

$$P_{a,u} = \frac{\sum_{i=1}^{rn} (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{rn} (r_{a,i} - \bar{r}_a)^2 \times \sum_{i=1}^{rn} (r_{u,i} - \bar{r}_u)^2}}$$

Donde $r_{a,i}$ es la calificación dada al ítem i por el usuario a; y \bar{r}_a es la calificación promedio dada por el usuario a.

En el paso 3, las predicciones se calculan como el promedio ponderado de desviaciones de la media del vecino:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \times P_{a,u}}{\sum_{u=1}^n P_{a,u}}$$

Donde $P_{a,i}$ es la predicción para el usuario-ítem a (usuario al que recomendar) para el ítem i. $P_{a,u}$ es la similitud entre los usuarios a y u, n es el número de usuarios en el vecindario.

A continuación, se detalla un ejemplo que describe el funcionamiento de un sistema de recomendación de filtrado colaborativo basado en usuarios.

Tabla 1

Ejemplo de matriz de preferencias para un sistema de recomendación de filtrado colaborativo basado en usuarios

	Item1	Item2	Item3	Item4	Item5
Jennifer	5	3	4	4	?
Usuario1	3	1	2	3	3
Usuario2	4	3	4	3	5
Usuario3	3	3	1	5	4
Usuario4	1	5	5	2	1

En la Tabla 1 se muestran las puntuaciones de Jennifer (usuario al que se le va realizar la recomendación) y otros 4 usuarios (sobre los que se identificará la red de vecinos más cercanos a Jennifer respecto a las puntuaciones efectuadas). La escala de puntuación para todos los usuarios va del 1 al 5, donde el 5 denota que es extremadamente preferido y el 1 denota extremadamente no preferido.

A partir de las puntuaciones de la Tabla 1, se calcula la similaridad entre Jennifer y cada uno de los usuarios usando el coeficiente de correlación de Pearson.

Se deben realizar cálculos preliminares para llegar al cálculo de la similaridad entre usuarios, como se muestra a continuación:

En la Tabla 2, se calcula la puntuación promedio que cada usuario ha realizado.

Tabla 2

Puntuación promedio de cada usuario

Jennifer	4.00
Usuario1	2.40
Usuario2	3.80
Usuario3	3.20
Usuario4	2.80

En la Tabla 3, se efectúa el cálculo de la puntuación que cada usuario le ha dado al ítem correspondiente respecto a su puntuación promedio.

Tabla 3

Diferencia de la puntuación por cada ítem respecto a la puntuación promedio de cada usuario

	Item1	Item2	Item3	Item4
Jennifer	1.00	-1.00	0.00	0.00
Usuario1	0.60	-1.40	-0.40	0.60
Usuario2	0.20	-0.80	0.20	-0.80
Usuario3	-0.20	-0.20	-2.20	1.80
Usuario4	-1.80	2.20	2.20	-0.80

En la Tabla 4, se efectúa el cálculo de la puntuación que cada usuario le ha dado al ítem correspondiente respecto a su puntuación promedio elevado al cuadrado.

Tabla 4

Diferencia de la puntuación por cada ítem respecto a la puntuación promedio de cada usuario elevado al cuadrado

	Item1	Item2	Item3	Item4
Jennifer	1.00	1.00	0.00	0.00
Usuario1	0.36	1.96	0.16	0.36
Usuario2	0.04	0.64	0.04	0.64
Usuario3	0.04	0.04	4.84	3.24
Usuario4	3.24	4.84	4.84	0.64

En la Tabla 5, a partir de los cálculos previos efectuados en las tablas 3 y 4 se puede identificar la similitud que existe entre Jennifer y los otros usuarios. Tomando en consideración que el coeficiente de correlación de Pearson varía de -1 a +1, los usuarios que tengan mayor cercanía a +1 serán los usuarios con un comportamiento más similar al mostrado por Jennifer, por lo tanto, los usuarios 1 y 2 son aquellos que forman parte de su red de vecinos más cercanos.

Tabla 5

Similaridad entre los usuarios

Jennifer - Usuario1	1.00
Jennifer - Usuario2	1.00
Jennifer - Usuario3	0.00
Jennifer - Usuario4	-1.00

Para entender la puntuación o el grado de interés que tendría Jennifer por el ítem 5 realizamos la predicción de este ítem tomando en cuenta las puntuaciones efectuadas por su red de vecinos más cercanos (usuarios 1 y 2).

$$\text{Predicción (Item5)} = 4 + \left\{ \frac{1}{2} * [1 * (3 - 2.4) + 1 * (5 - 3.8)] \right\} = 4.9$$

Dados estos esquemas de cálculo, ahora se pueden calcular predicciones de calificación para Jennifer para todos los artículos que aún no ha visto a partir de los valores de predicción más altos se genera la lista de recomendaciones.

Finalmente, no está claro qué se puede recomendar a los nuevos usuarios o cómo se manejan nuevos artículos para los que no existen calificaciones.

3.2.2. Propuesta de alternativa de solución a la situación problemática siguiendo las seis fases de la Metodología para el Desarrollo de Proyectos

El presente trabajo de suficiencia profesional se desarrolló bajo el paraguas de la metodología CRISP-DM, (Cross Industry Standard Process for Data Mining), cuyas fases principales son la comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación. El desarrollo de la solución estadística teniendo en cuenta cada una de las etapas mencionadas permitió retroalimentar el desarrollo, corrigiendo de manera oportuna en base a lo descubierto en cada etapa.

a. Fase de comprensión del negocio o problema

En esta fase se hicieron y contestaron preguntas de negocio con la finalidad de entender y dejar registro de las particularidades del negocio para poder usarse a favor del correcto aprendizaje del algoritmo de recomendación.

En esta fase se consultó y entrevistó a los expertos de negocio y analistas para entender los tipos de análisis que ya existen; lo que funciona y no; lo que está resultando difícil de ejecutar. Se registró la necesidad de identificar a los clientes rentable del banco y tener un sistema de recomendación de beneficios expresado en rubros de consumo potenciales y pueda ser desplegado a través de los medios digitales del banco: app y página web del banco.

b. Fase de comprensión de los datos

Cubierta la fase de comprensión del negocio y ante la necesidad de explotar profundamente los datos se hizo necesaria la etapa de la comprensión de los datos, que fue vital para continuar acertadamente en el proceso de generación de valor.

En esta fase se consultó y entrevistó a los expertos de los datos para entender los tipos de datos que existían y se estaban usando. Se dejó registrado todo el activo representado por los datos explotados recurrentemente.

Esta fase confirmó el alcance y beneficio de la elaboración de un sistema de recomendación de rubros de consumo para los clientes rentables del banco.

Además, permitió entender que se tenían 500 rubros de consumo distintos que podían ser recomendados por el algoritmo KNN Recommender.

c. Fase de preparación de los datos

Teniendo como referencia las dependencias encontrados en la fase de entendimiento de los datos, se inició la preparación de los datos a ser consumidos por el algoritmo KNN Recommender. En esta etapa se tuvo en consideración preparar los datos de manera tal que los algoritmos puedan encontrar los patrones necesarios para lograr los objetivos que el proyecto amerita. Dentro de los retos que se tomaron en cuenta en esta fase fue el tratamiento de los valores perdidos y extremos a partir de las transacciones realizadas por los clientes del banco en los 500 rubros de consumo diferentes. Se construyeron variables derivadas desde los datos originales que respondieron al formato de datos necesario a ser usado por el algoritmo KNN Recommender. Se obtuvieron dos matrices de datos: una primera matriz de datos que permitía realizar los análisis necesarios para identificar a los clientes rentables del banco y una segunda matriz de datos que tenía a todos los clientes del banco a nivel de filas y a los 500 rubros de consumo a nivel de columnas, el contenido de esta matriz de datos estaba representado por el número de transacciones realizadas por cada cliente con respecto al rubro de consumo determinado (simulando el efecto de valorización o puntuación de cada usuario en un sistema de recomendación) como promedio semanal teniendo en cuenta una ventana de tiempo de los últimos siete días.

Además, en esta etapa se identificó a los clientes rentables del banco a través del cálculo de percentiles basados en sus ingresos y sus deudas en el sistema financiero, sobre este grupo de clientes se fusionaron las matrices de datos procesadas.

d. Fase del modelado

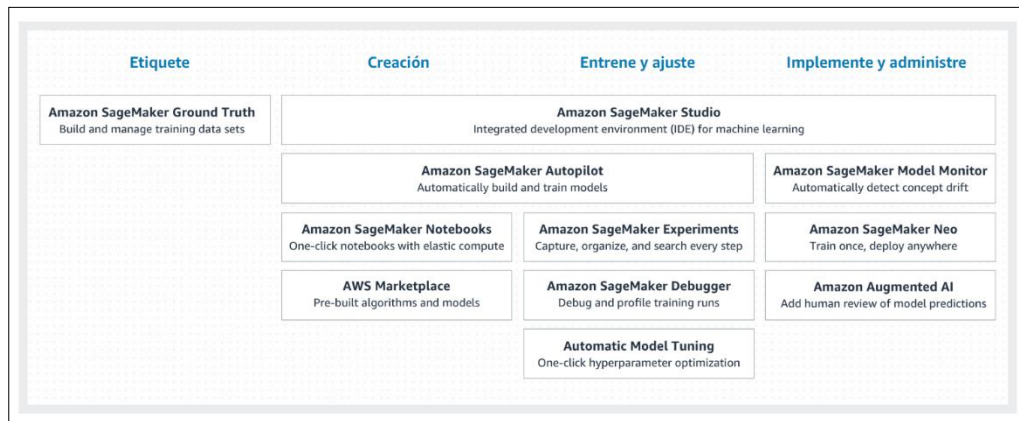
En la fase del modelado, el algoritmo KNN Recommender descubrió los patrones necesarios a partir de las transacciones realizadas a nivel de los diferentes rubros de consumo por los clientes rentables del banco, con la finalidad de predecir los rubros de consumo en los que

aún no existía interés, pero si existía potencial de consumo basado en la afinidad de comunidades.

Por la naturaleza iterativa del algoritmo, el entrenamiento del KNN Recommender tuvo un alto costo computacional y se usó un servicio de la nube de Amazon Web Services (Sage Maker) que permitió entrenar e implementar el algoritmo de aprendizaje automático, como se muestra en la Figura 6, Sagemaker tienes múltiples funcionalidades que permiten el desarrollo e implementación de modelos.

Figura 6

Sagemaker como servicio de desarrollo de modelos en AWS



Esta fase del proyecto se desarrolló con la ayuda del software Python, como se muestra en el anexo 1 se construyeron funciones que permitieron el entrenamiento del algoritmo.

e. Fase de evaluación

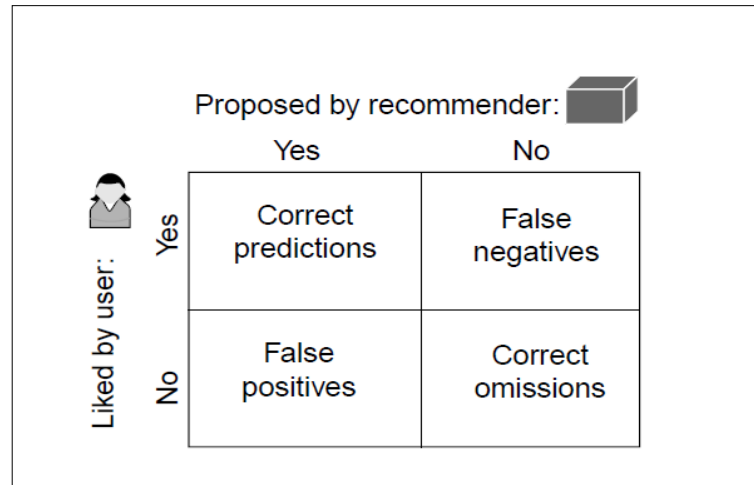
Jannach *et al.* (2010) establecen que los sistemas de recomendación tienen tradicionalmente ha ser evaluados utilizando experimentos fuera de línea que intentan estimar un error de predicción tomando como referencia las recomendaciones hechas utilizando un conjunto de datos existente de transacciones como se muestra en la Figura 7.

Sin embargo, muchos señalan las limitaciones de tales métodos, mientras que otros sostienen que la calidad de un sistema de recomendación nunca se puede medir directamente porque hay demasiadas funciones objetivos diferentes. Sin embargo, el uso generalizado de los

sistemas de recomendación hace que sea crucial desarrollar métodos realistas y evaluar con precisión su verdadero rendimiento y efecto sobre los usuarios.

Figura 7

Matriz de evaluación de recomendaciones hechas por un sistema

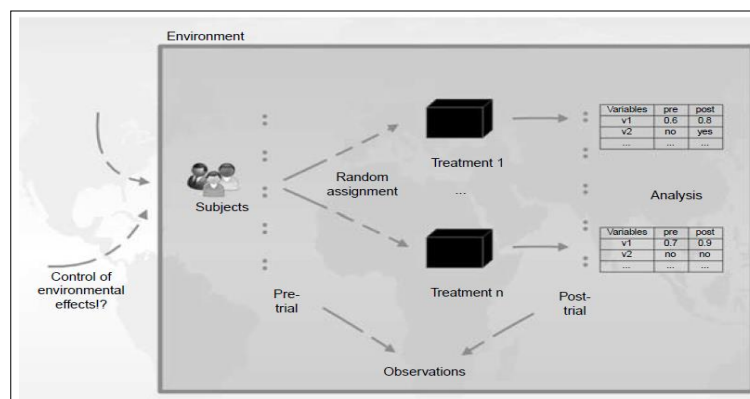


FUENTE: Jannach *et al.*(2010)

Por ello, en el presente trabajo de suficiencia profesional se utilizó un método de validación con pruebas A/B testing, en las que un grupo de clientes tenían una asignación genérica basada en reglas generalizadas de juicio de experto y otro grupo en base a una asignación dada por la recomendación del algoritmo KNN Recommender como se muestra en la Figura 8.

Figura 8

Ejemplo de diseño de experimento



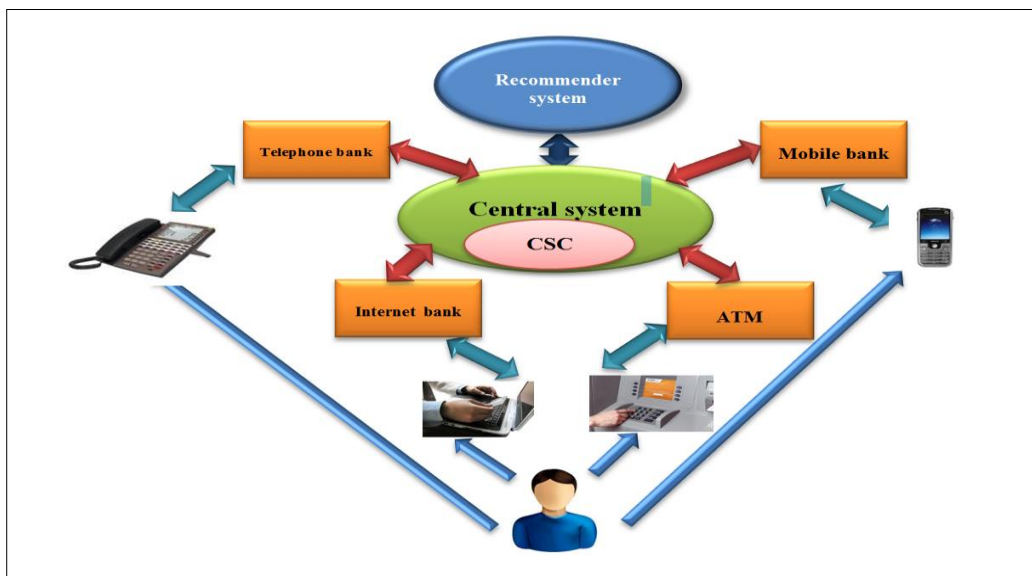
FUENTE: Jannach *et al.* (2010)

f. Fase de implementación

Basado en lo expuesto por Abdollahpouri y Abdollahpouri (2013), se planteó una arquitectura como la mostrada en la Figura 9, donde los sistemas informáticos del banco estén conectados con las demás dependencias necesarias y la solución estadística generada pueda generar el valor esperado.

Figura 9

Arquitectura propuesta para la personalización de los servicios de los clientes de un banco



FUENTE: Abdollahpouri & Abdollahpouri (2013)

En esta fase se contempló las siguientes actividades:

- Actualizar las matrices de datos (identificar a los clientes rentables del banco y para actualizar el algoritmo KNN Recommender) de manera semanal.
- Habilitar un servidor en el cual se puedan alojar las matrices de datos y se pueda ejecutar del algoritmo KNN Recommender de manera semanal.
- Comunicar la matriz de datos resultante con los sistemas encargados de refrescar el contenido en los canales digitales del banco como el app y la página web.

3.3. Contribución en la solución de situaciones problemáticas

Los principales problemas del banco eran la efectividad en la comunicación y la conversión de los beneficios disponibles para sus clientes rentables.

El primer problema se originaba con el envío de comunicación masiva a través de correos electrónicos; el ratio histórico de apertura de correos electrónicos en el periodo de estudio (2018) no superaba el 1% como se muestra en la tabla 6, lo cual implicaba tener alcance limitado de la población objetivo.

Tabla 6

Evolutivo del ratio de apertura de las campañas por correos electrónicos

Campaña	# Clientes rentables	# Clientes c/ interés	Ratio de Apertura
Ene. 18	110,656	686	0.62%
Feb. 18	114,199	948	0.83%
Mar. 18	115,625	844	0.73%
Abr. 18	121,296	934	0.77%
May. 18	118,101	1,146	0.97%
Jun. 18	126,258	972	0.77%
Jul. 18	124,124	1,117	0.90%
Ago. 18	123,415	1,123	0.91%
Set. 18	116,317	1,082	0.93%
Oct. 18	122,707	982	0.80%
Nov. 18	124,842	986	0.79%
Dic. 18	124,482	1,083	0.87%

Con los datos históricos de apertura de correos electrónicos de los clientes rentables del banco se evidenció la necesidad de cambiar el canal de comunicación para tener un mayor alcance sobre la población objetivo. Ante dicha necesidad, a través del aplicativo móvil del banco se disponibilizó un botón que direccionaba al cliente hacia la imagen del beneficio ofrecido por el banco, para visualizar dicho beneficio era necesario que el cliente del banco dé un clic sobre el botón, esta funcionalidad solo estaba disponible para aquellos clientes rentables del banco que el sistema de recomendación identificaba con suficiente transaccionalidad en los últimos siete días con su tarjeta de crédito o débito.

El segundo problema recaía en el ofrecimiento de beneficios para el segmento de clientes más rentable del banco, los criterios que se usaban estaban basados en la edad del cliente, el género del cliente y la temporada del año, la conversión de los beneficios ofrecidos no

superaba el 1% como se muestra la Tabla 7, es decir 2.6 clientes convertían de 1K clientes que mostraron interés en el beneficio del banco (0.26% en promedio como ratio de conversión).

Tabla 7

Evolutivo del ratio de conversión de las campañas por correos electrónicos

Campaña	# Clientes c/interés	# Clientes c/ conversión	Ratio de conversión
Ene. 18	686	3	0.44%
Feb. 18	948	2	0.21%
Mar. 18	844	1	0.12%
Abr. 18	934	2	0.21%
May. 18	1,146	3	0.26%
Jun. 18	972	4	0.41%
Jul. 18	1,117	3	0.27%
Ago. 18	1,123	3	0.27%
Set. 18	1,082	2	0.18%
Oct. 18	982	2	0.20%
Nov. 18	986	3	0.30%
Dic. 18	1,083	3	0.28%

Con los datos históricos de conversión del beneficio ofrecido a los clientes rentables del banco se evidenció la necesidad de cambiar la estrategia de ofrecimiento para tener un mayor impacto sobre la población objetivo.

Es por ello que una vez que el cliente mostraba interés en la oferta, dándole clic al botón mostrado en el aplicativo móvil del banco, la recomendación del beneficio ofrecido por el banco pasó a ser generada por el algoritmo KNN Recommender, este algoritmo se basaba en los datos registrados en las transacciones efectuadas en los últimos siete días.

Muchos de los clientes que mostraban interés por el beneficio reportaban problemas en redimirlo, dado que no todos los establecimientos encargados de hacer efectivo el beneficio contaban con la información en línea. Ante este problema, el sistema del banco que registraba el interés y la conversión de los beneficios ofrecidos se tuvieron que tener en línea con los establecimientos afiliados con la finalidad de mejorar la experiencia de los clientes rentables del banco.

La necesidad personalizar estos servicios para los clientes se siente más que antes ya que las personas están más ocupadas que nunca y tienen un tiempo limitado para hacer sus tareas diarias. Ellos presentan una arquitectura para la implementación de este servicio de personalización en un entorno multicanal con la ayuda de un enfoque soportado con sistemas de recomendación basados en el filtro colaborativo en memoria como se muestra en el anexo 1. Lo expuesto por Abdollahpouri y Abdollahpouri (2013) sirvió como referencia para que los sistemas informáticos del banco se adapten a la solución estadística generada.

3.4. Análisis de la contribución en términos de competencias y habilidades

En la solución del problema del banco del presente trabajo de suficiencia profesional participaron diferentes profesionales con distinta formación, las tareas de los profesionales involucrados estuvieron compartidas entre analíticas y tecnológicas.

Blake en el reporte de Royal Society (2019) muestra un marco de categorías en el anexo 4 y dependiendo de la rigurosidad analítica necesaria relaciona la descripción de los roles y los profesionales necesarios.

Al profesional de Estadística e Informática a través de sus competencias y habilidades se le solicita la creación de modelos analíticos sofisticados, construcción de matrices de datos y extracción de conocimiento de los datos, siendo ubicado en el nivel más alto de exigencia analítica.

El rol que se desempeñó en el banco tuvo la misión de crear la herramienta estadística capaz de personalizar la comunicación del banco a sus clientes a nivel de beneficios a través del algoritmo KNN Recommender.

IV. RESULTADOS Y DISCUSIÓN

4.1. Nivel de beneficio obtenido por el centro laboral

El banco donde se aplicó el presente trabajo obtuvo los siguientes beneficios: la mejora en la efectividad de la comunicación y la mejora de la conversión de los beneficios disponibles para sus clientes rentables. En la Tabla 8 se muestra que en promedio el 85% de los clientes rentables del banco presentaban suficiente transaccionalidad con sus tarjetas de crédito o débito (105K de 125K clientes) y estaban en condiciones de visualizar la oferta a través de su aplicativo móvil, de ellos el 5.6% (5.9K clientes rentables) mostraron interés por el beneficio.

Tabla 8

Evolutivo del ratio de apertura de las campañas por banca móvil

Campaña	# Clientes rentables	# Clientes c/ consumo	# Clientes c/interés	Ratio de Apertura
Ene. 19	124,561	100,894	5,645	5.59%
Feb. 19	124,350	103,211	5,702	5.52%
Mar. 19	124,569	102,147	5,738	5.62%
Abr. 19	124,621	103,435	5,790	5.60%
May. 19	124,546	104,619	5,810	5.55%
Jun. 19	124,613	105,921	5,895	5.57%
Jul. 19	124,123	107,987	6,002	5.56%
Ago. 19	124,567	109,619	6,112	5.58%
Set. 19	124,703	109,739	6,105	5.56%

Para lograr cuantificar el impacto del uso del sistema de recomendación se tomaron muestras de control desde Ene.19 a Set. 19, las muestras de control fueron aleatorias y representativas de la población objetivo (clientes rentables del banco) sobre las muestras se aplicaron los tratamientos convencionales (campañas dirigidas a través de correos electrónicos), los resultados se extrapolaron a la población y se obtuvieron los resultados que se muestran en la Tabla 9, por lo tanto usar un sistema de recomendación basado en el algoritmo KNN Recommender nos llevó a casi sextuplicar el interés de los clientes rentables del banco por

los beneficios que les ofrecía el banco, el promedio de clientes con interés a través de las campañas de correo electrónico hubiera sido de 1K clientes con respecto a los 5.8K clientes interesados obtenidos con las campañas desplegadas a través de la banca móvil.

Tabla 9

Extrapolación del evolutivo del ratio de apertura de las campañas por correos electrónicos

Campaña	# Clientes rentables	# Clientes c/interés	Ratio de Apertura
Ene. 19	124,561	1,096	0.88%
Feb. 19	124,350	982	0.79%
Mar. 19	124,569	997	0.80%
Abr. 19	124,621	1,034	0.83%
May. 19	124,546	1,059	0.85%
Jun. 19	124,613	1,047	0.84%
Jul. 19	124,123	1,080	0.87%
Ago. 19	124,567	1,109	0.89%
Set. 19	124,703	1,122	0.90%

En la Tabla 10 se muestra que en promedio el 8% de los clientes rentables del banco que mostraron interés en el beneficio terminaron convirtiendo el beneficio (466 de 5.9K clientes).

Tabla 10

Evolutivo del ratio de conversión de las campañas por banca móvil

Campaña	# Clientes c/interés	# Clientes c/ conversión	Ratio de conversión
Ene. 19	5,645	432	7.65%
Feb. 19	5,702	439	7.70%
Mar. 19	5,738	459	8.00%
Abr. 19	5,790	466	8.05%
May. 19	5,810	462	7.95%
Jun. 19	5,895	475	8.06%
Jul. 19	6,002	486	8.10%
Ago. 19	6,112	488	7.98%
Set. 19	6,105	490	8.03%

El sistema de recomendación basado en el algoritmo KNN Recommender llevó al banco a tener una conversión en promedio mensual de 3 a 466 clientes rentables como se puede comparar con las tablas 10 y 11 respectivamente.

Tabla 11

Extrapolación del evolutivo del ratio de conversión de las campañas por correos electrónicos

Campaña	# Clientes c/interés	# Clientes c/ conversión	Ratio de conversión
Ene. 19	1,096	3	0.29%
Feb. 19	982	3	0.30%
Mar. 19	997	3	0.31%
Abr. 19	1,034	3	0.27%
May. 19	1,059	4	0.35%
Jun. 19	1,047	3	0.28%
Jul. 19	1,080	3	0.29%
Ago. 19	1,109	4	0.32%
Set. 19	1,122	4	0.33%

V. CONCLUSIONES

- El uso del sistema de recomendación como herramienta estadística cumplió un rol importante en el proceso de transformación digital del banco, personalizar la experiencia de sus clientes respecto a las campañas masivas generó beneficios como los obtenidos en la aceptación de la campaña (se creció de 1K a 5.8K clientes rentables) y en la conversión de la campaña (se creció de 3 a 466 clientes rentables con redención).
- Los sistemas de recomendación como estrategia de personalización de ofertas para los clientes de la banca fueron útiles a partir del despliegue en medios digitales como la banca móvil y teniendo conexión directa y en línea con los establecimientos de redención implicados.
- Los sistemas de recomendación capturaron el comportamiento de comunidades similares, para aislar efectos de dispersión en el consumo de los distintos perfiles de clientes del banco, se identificó a los clientes rentables del banco (124.5K clientes) y sobre ellos se aprendió de todos aquellos que tenían suficiente transacciones (85% de ellos).
- Se excluyó los consumos realizados en rubros atípicos como los establecimientos de limpieza, talleres de reparación y galerías de arte para facilitar el aprendizaje del algoritmo.

VI. RECOMENDACIONES

- Probar diferentes tipos de sistemas de recomendación como los basados en contenido, híbridos, conscientes del contexto y basados en el conocimiento.
- Reducir la temporalidad en la que se cuantifica la valorización de los clientes expresada en términos de transacciones.
- Probar otros criterios de valorización distintos al número de transacciones efectuadas en una ventana de tiempo determinada.
- Analizar el impacto al cambiar la medida de similaridad de la comunidad de vecinos más cercanos.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Abdollahpouri, H. & Abdollahpouri, A. (2013). An approach for personalization of banking services in multi-channel environment using memory-based collaborative filtering [Archivo PDF]. DOI:10.1109/IKT.2013.6620066
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370.
- Burke, R., Felfering, A. & Goeker, M. (2011). Recommender Systems: An Overview. *AI Magazine*, 32(3), 13-18.
- Chirkina, A. & Rankov, B. (2018). *Recommender System for private banking* [Blog Post]. Retrieved from <https://www.incubegroup.com/blog/recommender-system-for-private-banking/>
- Gigimol, S. & Sincy, J. (2016). A Survey on Different Types of Recommendation Systems. *International Research Journal of Advanced Engineering and Science*, 1, 111-113.
- IBM. (2021). *Guía de CRISP-DM de IBM SPSS Modeler*. https://www.ibm.com/docs/es/spss-modeler/saas?topic=SS3RA7_sub/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.html
- Jannach, D., Zanker, M., Felfernig, A. & Friedrich, G. (2010). *Recommender Systems in Computer Science and Information Systems – A Landscape of Research*. In: Huemer, C., Lops, P. (eds) E-Commerce and Web Technologies. EC-Web 2012. Lecture Notes in Business Information Processing, vol 123. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-32273-0_7
- Prasad, B. & Sarath, K. (2011). An implementation of user based collaborative filtering algorithm. *International Journal of Computer Science and Information Technology*, 2, 1283-1286.
- Resnick, P. y Varian, H.R. (1997). Sistemas de recomendación. *Comunicaciones de la ACM*, 40, 56-58. <https://doi.org/10.1145/245108.245121>
- Royal Society. (2019). *Dynamics of Data Science Skills*. <https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf>.

VIII. ANEXOS

Anexo 1

Código Python usado en el entrenamiento del algoritmo KNN Recommender

```
def CF_userbased(u_vec,K,data,indx=False):
    def FindKNeighbours(r,data,K):
        neighs = []
        cnt=0
        for u in xrange(len(data)):
            if data[u,r]>0 and cnt<K:
                neighs.append(data[u])
                cnt +=1
            elif cnt==K:
                break
        return np.array(neighs)

    def CalcRating(u_vec,r,neighs):
        rating = 0.
        den = 0.
        for j in xrange(len(neighs)):
            rating += neighs[j][-1]*float(neighs[j][r]-neighs[j][neighs[j]>0][-1].mean())
            den += abs(neighs[j][-1])
        if den>0:
            rating = np.round(u_vec[u_vec>0].mean()+(rating/den),0)
        else:
            rating = np.round(u_vec[u_vec>0].mean(),0)
        if rating>5:
            return 5.
        elif rating<1:
```

```

        return 1.
    return rating
#add similarity col
data = data.astype(float)
nrows = len(data)
ncols = len(data[0])
data_sim = np.zeros((nrows,ncols+1))
data_sim[:, :-1] = data
#calc similarities:
for u in xrange(nrows):
    if np.array_equal(data_sim[u, :-1], u_vec) == False: #list(data_sim[u, :-1]) !=
list(u_vec):
        data_sim[u, ncols] = sim(data_sim[u, :-1], u_vec, 'pearson')
    else:
        data_sim[u, ncols] = 0.
#order by similarity:
data_sim = data_sim[data_sim[:, ncols].argsort()][: :-1]
#find the K users for each item not rated:
u_rec = np.zeros(len(u_vec))
for r in xrange(ncols):
    if u_vec[r] == 0:
        neighs = FindKNeighbours(r, data_sim, K)
        #calc the predicted rating
        u_rec[r] = CalcRating(u_vec, r, neighs)
if indxs:
    #take out the rated movies
    seenindxs = [indx for indx in xrange(len(u_vec)) if u_vec[indx] > 0]
    u_rec[seenindxs] = -1
    recsvec = np.argsort(u_rec)[: :-1][np.argsort(u_rec) > 0]

    return recsvec
return u_rec

```

Anexo 2

Marco de categorías de roles laborales por el nivel de rigurosidad analítica por Royal Society. (2019)

TABLE 1

Framework categories showing increasing levels of analytical rigour across all Data Science and Analytics (DSA) jobs.

	FRAMEWORK	FUNCTIONAL ROLE	SAMPLE OCCUPATIONS
ANALYTICAL RIGOUR	Data Scientists and Advanced Analysts	Create sophisticated analytical models used to build new datasets and derive new insights from data	Data Scientist Economist Data Engineer Biostatistician Statistician Financial Quantitative Analyst
	Data Analysts	Leverage data analysis and modelling techniques to solve problems and glean insight across functional domains	Data Analyst Business Intelligence Analyst
	Data Systems Developers	Design, build and maintain an organisation's data and analytical infrastructure	Systems Analyst Database Administrator
	Analytics Managers	Oversee analytical operations and communicate insights to executives	Chief Analytics Officer Marketing Analytics Manager
	Functional Analysts	Utilise data and analytical models to inform specific functions and business decisions	Business Analyst Financial Analyst
	Data-Driven Decision Makers	Leverage data to inform strategic and operational decisions	IT Project Manager Marketing Manager