UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMATICA

MODELO DE RECOMENDACION DE ACCIONES BASADO EN COMPETENCIAS PARA EL CURSO DE MATEMATICA II DE LA ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMATICA DE LA UNIVERSIDAD NACIONAL DE MOQUEGUA, AÑO 2021.

TESIS

PRESENTADO POR:

VERONICA QUISPE GARCÍA

PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO DE SISTEMAS E INFORMATICA

MOQUEGUA – PERÚ

2022

UNAM
UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Folio **Nº 006**

ACTA DE SUSTENTACIÓN

I. Modalidad:

Tesis : [X]
Trabajo de Suficiencia Profesional : [ ]
Otro: ..... : [ ]

II. Documento de aprobación para la sustentación:

Resolución de facultad Nº 071-2022-FACIA-UNAM

Fecha: 16 de mayo 2022

III. El jurado ha dictaminado la (el) Tesis, TITULADO:

"Modelo de recomendación de acciones basado en competencias para el curso de matemática II de la Escuela Profesional de Ingeniería de sistemas e Informática de la Universidad Nacional de Moquegua, año 2021".

Bachiller: Verónica Quispe García

Asesor Principal: Dra. Janeza Flores Gutiérrez

Co-Asesor: Dr. Salamón Rey Ramos Rivera

IV. Se dictamina:

Como: Aprobado por mayoría la sustentación de manera:

Presencial [X] Virtual [ ]

Con nota: Doce (12) por lo que; Sí [X] No [ ]

Opta el Título Profesional de INGENIERO DE SISTEMAS E INFORMÁTICA .

Siendo las 14:00 horas, del 15 de junio del año 2022 se procede con la firma en conformidad de lo dictaminado por el Jurado.

Dr. Anibal Flores G.
Presidente

Msc Alex Zuñiga I.
Primer Miembro

Mg. Juan Claros P.
Segundo Miembro

Bach. Verónica Quispe G.
Bachiller

DEDICATORIA

La presente Tesis se la dedico a Nuestro Señor Padre Celestial y a su madre, nuestra Virgen María, quienes con su bendición guían cada uno de mis pasos día a día.

A mi padre, quien partió al lado de nuestro Dios hace más de un año, pero que siempre me acompaña, a mi madre a quien le debo cada logro por su amor y apoyo incondicional; a mis hermanos por ser las personas que siempre están ahí a pesar de cualquier dificultad.



AGRADECIMIENTOS

Agradezco a nuestro Padre Celestial y a nuestra Virgen María por bendecirme, guiarme y brindarme la perseverancia necesaria para culminar cada uno de mis metas.

A mi familia por acompañarme y apoyarme en cada paso que doy y por tener confianza en mí.

A mis Asesores, Dra. Vaneza Flores Gutiérrez y Dr. Salomón Rey Ramos Rivera, por compartir conmigo sus conocimientos y experiencias como profesionales, por brindarme su tiempo y paciencia en el desarrollo de esta investigación.

Finalmente agradezco a la Universidad Nacional de Moquegua, por haberme permitido formarme en sus aulas y conocer personas que el día de hoy forman parte de mi vida y finalmente agradezco a todas las personas que con su apoyo podré lograr este gran anhelo.



RESUMEN

Los modelos de recomendación hoy en día son apropiados para la generación y entrega de recomendaciones muy útiles a los usuarios y en el ámbito de educación son muy aprovechables para el apoyo del rendimiento académicos de los estudiantes.

La presente investigación tiene como objetivo proponer un modelo de recomendación basado en competencias que determine acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua. El modelo de recomendación propuesto se obtendrá con la técnica de filtrado colaborativo basado en usuarios, y su desarrollo será mediante la metodología CRISP-DM. En la fase de construcción del modelo se trabaja con cuatro etapas, *pre-procesamiento de los datos de entrada*: matrices con las calificaciones de los estudiantes (antiguos y actuales) y matrices con las puntuaciones de las acciones a recomendar, *filtro de candidatos antiguos*, *cálculo de la similitud* en base al coeficiente de Jaccard y por último la *puntuación y entrega de las acciones a recomendar*.



ABSTRACT

The recommendation models today are appropriate for the generation and delivery of very useful recommendations to users and in the field of education are very useful for supporting the academic performance of students.

The present research aims to propose a competency-based recommendation model that determines actions for the Mathematics II course of the Professional School of Systems and Computer Engineering of the National University of Moquegua. The proposed recommendation model will be obtained with the user-based collaborative filtering technique, and its development will be through the CRISP-DM methodology. In the construction phase of the model we work with four stages, *preprocessing of the input data*: matrices with the qualifications of the students (former and current) and matrices with the scores of the actions to recommend, *filtering of former candidates*, *calculation of the similarity* based on the Jaccard coefficient and finally the *scoring and delivery of the actions to recommend*.



INDICE GENERAL

RESUMEN 4
ABSTRACT 5
INDICE GENERAL 6
INTRODUCCIÓN 10
I. PLANTEAMIENTO DEL PROBLEMA 11
1.1 Descripción de la realidad problemática 11
1.2 Formulación del problema de investigación 12
1.2.1 Problema general 12
1.2.2 Problemas específicos 12
1.3 Antecedentes 13
1.4 Objetivos 15
1.4.1 Objetivo general 15
1.4.2 Objetivos específicos 15
1.5 Justificación e importancia 16
1.5.1 Justificación 16
1.5.2 Importancia 16
1.6 Hipótesis 17
1.6.1 Hipótesis general 17
1.6.2 Hipótesis específicas 17
1.7 Variables de la investigación 17
1.7.1 Identificación de las variables 17
1.7.2 Operacionalización de las variables 17
II. MARCO TEÓRICO 19
2.1 Bases teóricas de la investigación 19
2.1.1 Educación basada en competencias 19
2.1.2 Matemáticas en educación universitaria 20
2.1.3 Minería de Datos 21
2.1.4 Modelo CRISP-DM 21
2.1.5 Modelos de recomendación 22
2.1.6 Sistemas de recomendación 22
2.1.7 Técnicas de implementación de un sistema de recomendación 24

2.1.8 Filtrado colaborativo 25
2.1.9 Métricas de similaridad 26
2.1.10 Coeficiente de similitud de Tanimoto (Jaccard) 26
2.1.11 Métricas de evaluación de un sistema de recomendación 27
2.2 Definición de términos 28
III. METODOLOGÍA 30
3.1 Tipo y Nivel de la investigación 30
3.2 Alcance de la investigación 31
3.3 Ámbito temporal y espacial 31
3.4 Población y muestra 31
3.5 Técnicas e instrumentos para la recolección de datos 32
3.6 Procedimientos 33
3.7 Análisis y procesamiento de los datos 34
3.7.1 Comprensión del negocio 34
3.7.2 Comprensión de los datos 38
3.7.3 Preparación de los datos 42
3.7.4 Modelado 45
3.7.5 Evaluación 58
3.7.6 Implementación 62
IV. RESULTADOS 63
V. DISCUSIÓN DE RESULTADOS 65
VI. CONCLUSIONES 67
VII. RECOMENDACIONES 69
VIII. REFERENCIAS 69
IX. ANEXOS 72



INDICE DE TABLAS

Tabla 1. Operacionalización de la variable independiente 18

Tabla 2. Operacionalización de la variable dependiente 18

Tabla 3. Registros de estudiantes matriculados en Matemática II 31

Tabla 4. Registros de estudiantes matriculados en Matemática II 32

Tabla 5. Porcentaje de estudiantes desaprobados en Matemática II 36

Tabla 6. Atributos de los registros académicos 39

Tabla 7. Acciones a recomendar 39

Tabla 8. Codificación de las Evaluaciones 44

Tabla 9. Codificación de las acciones a recomendar 45

Tabla 10. Calificaciones del estudiante actual CCE 51

Tabla 11. Calificaciones del estudiante actual CMP 54

Tabla 12. Puntuaciones generadas para el estudiante CCE 57

Tabla 13. Puntuaciones generadas para el estudiante CMP 58

Tabla 14. Error Medio Absoluto para la similitud de Jaccard 59

Tabla 15. Eror Medio Absoluto para la similitud de Pearson 59

Tabla 16. Matriz de confusión, estudiante CCE 60

Tabla 17. Matriz de confusión estudiante CMP 61

Tabla 18. Precisión y Recall obtenidos 61

Tabla 19. Acciones a recomendar generadas por el modelo 63



INDICE DE FIGURAS

Figura 1. Fases de la metodología CRISP-DM 22

Figura 2. Elementos de un Sistema de Recomendación 23

Figura 3. Funcionamiento del modelo de recomendación....25

Figura 4. Etapas del modelado de la creación del modelo de recomendación 34

Figura 5. Objetivos Institucionales de la UNAM 35

Figura 6. Fases de la metodología CRISP-DM 38

Figura 7. Registro académico del curso de Matemática 2020-1 41

Figura 8. Registro académico del curso de Matemática 2020-2....41

Figura 9. Puntuación de las acciones a recomendar 42

Figura 10. Limpieza de datos....43

Figura 11. Estructuración de los datos 43

Figura 12. Integración de los datos 44

Figura 13. Etapas de construcción del modelo de recomendación 45

Figura 14. Matriz de las calificaciones de los estudiantes antiguos....47

Figura 15. Matriz de las calificaciones de los estudiantes actuales 48

Figura 16. Filtrado de los estudiantes antiguos 49

Figura 17. Matriz de estudiantes que aprobaron la última AE 50

Figura 18. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo AAC .51

Figura 19. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo LVL...52

Figura 20. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo LVA ..52

Figura 21. Similitud según Jaccard con respecto al estudiante CCE 53

Figura 22. Similitud según Pearson con respecto al estudiante CCE 53

Figura 23. Similitud según Jaccard con respecto al estudiante CMP 55

Figura 24. Similitud según Pearson con respecto al estudiante CMP....56

Figura 25. Matriz de puntuaciones de los estudiantes antiguos 57

Figura 26. Matriz de estudiantes con un nivel mayor a 10 de AE6 64

Figura 27. Matriz con la similitud según Jaccard....65



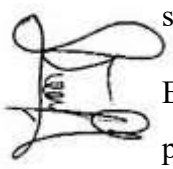
INTRODUCCIÓN

En los últimos años, en la educación se viene generando gran cantidad de información día tras día, sobre todo con respecto al ámbito académico. En muchas ocasiones, estos grandes volúmenes de datos esconden patrones que pueden ser usados para encontrar contenidos que sean de interés para apoyar en el rendimiento académico de los estudiantes, quienes son el pilar fundamental de toda universidad y buscan herramientas que les sean de apoyo para la culminación exitosa de sus estudios.

Con el uso de herramientas de predicción y recomendación, se busca apoyar en el proceso de aprendizaje de los estudiantes por medio de recomendaciones de acciones acordes a sus necesidades, los cuales deberán verse reflejados en el desarrollo de las competencias que presenten dificultades en un curso en específico.

Esta investigación pretende presentar un modelo de recomendación que apoye al estudiante universitario a encontrar recomendaciones de acciones para el curso de Matemática II de la Escuela de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua, en el entorno no presencial que actualmente se está dando, cuya contribución principal del estudio, es establecer matrices bases que contengan el perfil de los estudiantes y de las acciones a recomendar. Éstas últimas, serán establecidas por un especialista en la materia, a su vez, se tomará en cuenta el conocimiento social de los estudiantes antiguos. Por ser una investigación que inicia en cero, se busca establecer la relación de usuario y de los ítems propios de los modelos de recomendación.

La investigación está dada bajo un enfoque cuantitativo de tipo descriptivo y explicativo, presentando un modelo de recomendación con técnicas de filtrado colaborativo basado en usuarios. Para ello, se desarrollará la metodología CRISP-DM, cuyas fases están enfocadas a este tipo de investigación, y dentro de ellas, para la creación del modelo de recomendación se realiza un pre-procesamiento de dos tipos de datos de entrada, luego se realiza un filtro, cálculo de similitud y finalmente se dan las acciones a recomendar.



Este estudio está compuesto por las siguientes secciones: en la primera, tenemos el planteamiento del problema, el cual describe la problemática, los antecedentes, objetivos, justificación, importancia, hipótesis y variables de la investigación. En la segunda sección

se detalla el marco teórico con la definición de términos. La tercera sección explica la metodología de la investigación. Los resultados se presentan en la cuarta sección, seguido de la discusión de resultados, conclusiones, recomendaciones, referencias y por último los anexos que respaldan la investigación.

I. PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

En los últimos años se han venido realizando numerosos estudios sobre los inconvenientes que existen en los entornos de aprendizaje de los estudiantes. Hay que tener en cuenta que no todos los estudiantes asimilan los conocimientos de igual manera, mientras que para unos resulta muy sencillo aprender conceptos y procesos de un curso, para otros esto puede llegar a convertirse en algo muy pesado.

Por otro lado, con el paso del tiempo, se van acumulando grandes volúmenes de datos históricos de estudiantes, ello conlleva a la acumulación de una enorme cantidad de conocimiento social que hacen que resulte difícil, para los nuevos usuarios, encontrar de manera fácil y rápida, contenidos específicos que necesitan o se acomoden a sus necesidades. Es ahí donde aparecen los sistemas de recomendación, que, por medio de diversas técnicas y estrategias, hacen posible encontrar patrones de recuperación de información (filtrados) que dan solución al problema de sobrecarga de datos.

Estos sistemas son aplicados en todos los ámbitos, en el comercio, entretenimiento, en la educación, entre otros. Si nos centramos en educación, más aún en educación universitaria, su uso es mucho más apropiado como estrategias para la disminución de la deserción universitaria.

Estudios como el de (Cabrera et al., 2006), considera que la política universitaria debe preocuparse en analizar los perfiles de los estudiantes y las necesidades que ellos tengan, como preferencias profesionales, habilidades, capacidades entre otras. El principal factor de deserción es el institucional, por lo que se recomienda que los procesos de enseñanza-aprendizaje pueden ser mejorados por los mismos docentes, aplicando metodologías, talleres o herramientas aportadas por la misma universidad. (Mori Sánchez, 2012)

Por lo expuesto, es necesario contar con un modelo de recomendación que brinde adecuadas recomendaciones (acciones) que apoyen el proceso de aprendizaje de los estudiantes universitarios. Actualmente nuestra sociedad afronta una situación delicada,



debido a la pandemia provocada por el virus SARS-CoV-2, lo que ha conllevado a que la educación se de manera no presencial, obligando de alguna manera a innovar nuevas formas que permitan al estudiante logre alcanzar las competencias establecidas en un curso en específico.

Es por ello que esta investigación se lleva a cabo en la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua, analizando las calificaciones de los estudiantes que se matricularon en el curso de Matemática II en el periodo 2020-1 y 2020-2. Estudios recientes como el de (Vargas & Montero, 2016) afirman que las matemáticas en educación universitaria es uno de los cursos donde se refleja muchos problemas de aprendizaje, lo cual repercute en el rendimiento académico y conlleva a la reprobación o deserción. La situación descrita, no es ajena a la Universidad en la cual se lleva a cabo la investigación, por lo que se busca implementar un modelo de recomendación de acciones basadas en competencias que contribuya a la culminación con éxito del curso y disminuya la reprobación o deserción antes mencionada.

1.2 Formulación del problema de investigación

1.2.1 Problema general

Un modelo de recomendación basado en competencias, ¿puede determinar acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua?

1.2.2 Problemas específicos

- ¿Cómo realizar el pre-procesamiento de dos tipos de datos de entrada orientados a una competencia del curso de Matemática II?
- ¿Cómo realizar el filtro de estudiantes antiguos a través de su rendimiento académico para elegir el nivel de competencia esperado en el curso de Matemática II?
- ¿Cómo calcular la similitud de la curva de rendimiento entre el estudiante actual y los estudiantes antiguos?
- ¿Cómo obtener las recomendaciones a entregar al estudiante actual del curso de Matemática II?



1.3 Antecedentes

(Chavarriaga et al., 2014) publicaron el artículo titulado “Recommender System Based on Student Competencies Assessment Results”, el mismo que en español es “Sistema de recomendación basado en los resultados de la evaluación de las competencias de los estudiantes”. En este artículo describen un prototipo de sistema de recomendación de ítems que pueden ayudar a el proceso de aprendizaje de los estudiantes de un curso determinado basado en evaluación por competencias.

Usan técnicas de filtrado colaborativo y filtrado basado en conocimiento, aprovechando las experiencias y las calificaciones de los estudiantes antiguos. El sistema de recomendación recibe como datos de entrada las calificaciones de los estudiantes vinculados al curso en el pasado y las calificaciones de los estudiantes actuales, a quienes en un punto dado se les va a dar las recomendaciones.

Los autores concluyen que el sistema fomenta el conocimiento social ya que los mismos estudiantes pueden aportar con ítems que puedan ayudar en el proceso de aprendizaje de los estudiantes en un fututo, recolectando experiencias exitosas de estudiantes antiguos y sus niveles de competencias alcanzados.

Gibaja, et al. (2016) en su artículo “Aplicación móvil para la recomendación de asignaturas en los estudios de grado universitarios” tienen como objetivo el diseño e implementación de una aplicación móvil para la recomendación de asignaturas optativas basándonos en un sistema de recomendación colaborativo. Para obtener resultados preliminares, diseñan una encuesta donde los alumnos de tercer y cuarto curso del Grado de Ingeniería Informática valoran las asignaturas que hayan cursado y esa información es introducida en la base de datos, como información inicial que será empleada por el sistema de recomendación. Lo resaltante de esta investigación, es que hacen uso de cuatro algoritmos recomendadores, Correlación de Pearson, Distancia Euclídea, Coeficiente de Tanimoto (Jaccard) y Probabilidad LogLikelihood. Los resultados son sometidos a tres métricas de evaluación, Error Medio Absoluto, Precisión y Recall. Al finalizar, concluyen indicando que los mejores resultados se obtuvieron con la correlación de Pearson, obteniendo un 86.6% de precisión y 68.1% de Recall, dejando como segunda opción el coeficiente de Jaccard, quien obtuvo el 64% de precisión y un 72% de recall.



Vaca (2017) presentó la Tesis Doctoral denominada "Sistema de recomendación de tareas basado en competencias educativas registradas en un portafolio electrónico", la misma que dentro de sus objetivos está la construcción de un sistema inteligente de recomendación de tareas adecuadas al desarrollo de competencias de los alumnos de educación universitaria. Con este planteamiento se busca que el sistema indique que tareas debe realizar el alumno para adquirir las competencias plasmadas.

El autor plantea que estas tareas deben estar incluidas previamente por los docentes en un repositorio en el e-portafolio. Esta investigación se centra en desarrollar el sistema de recomendación en la asignatura "Habilidades Comunicativas" (curso de inglés), el autor elige una sola competencia transversal para hacer el desarrollo menos complejo, ésta es "CT3 – ser capaz de comunicarse de forma efectiva en inglés", identificando cinco indicadores: Reading, writing, listening, speaking y grammar.

Como conclusión de la investigación se plasma que el sistema de recomendación de tareas debe constar de cuatro partes: un repositorio de tareas, que son los ítems a recomendar; el modelo del usuario (en este caso del alumno) el cual almacena las calificaciones de alumnos desde años atrás; el recomendador que utiliza el alumno para consultar en el repositorio y generar la recomendación y un ranking de tareas generadas que se muestran al alumno.

Mendoza, et al. (2019) publicaron el artículo denominado Métricas de Similaridad y evaluación para sistemas de recomendación de filtrado colaborativo, en el cual presentan una comparación de las distintas métricas de similitud para la técnica de filtrado colaborativo basados en ítems, utilizan datasets de producción para realizar las pruebas necesarias. En su investigación, construyen módulos en Python para las métricas en cuestión con las cuales obtienen recomendaciones utilizando tres técnicas generadoras de las mismas. Finalmente concluyen indicando que las recomendaciones obtenidas dependen del enfoque colaborativo del sistema de recomendación, es decir, depende de la similitud utilizada para generar los vecinos cercanos, ya sea de ítems o de usuarios.

Sánchez, et al. (2019) en su artículo "Persuade Me - A User-based Recommendation System Approach" proponen un sistema de recomendación basado en perfiles que genera un conjunto de posibles estrategias persuasivas que pueden usarse con usuarios de características similares. Plantean mecanismos efectivos para aumentar el éxito de las recomendaciones de este tipo y a su vez proporcionan las estrategias que se deben seguir



para aliviar el problema del arranque en frío, es decir, cuando recién se iniciará un estudio de este tipo y que puede ser extrapolado a otros ámbitos. Los autores concluyen que, con su propuesta predicen la calificación que los usuarios darán al conjunto de estrategias persuasivas que les resulten apropiadas para darlas como recomendaciones.

Owen, et al. (2012) en su libro "Mahout in Action" describen las funcionalidades de las métricas más usadas para filtrado colaborativo basado en usuarios, donde la implementación de la similitud de usuario es primordial. Dentro de ello, afirman que el coeficiente de Jaccard es una de las métricas a usarse cuando se ignora el valor de las preferencias, es decir, no es necesario considerar el valor de la calificación que los usuarios hacen hacia los ítems a recomendar para el cálculo de la similitud de los usuarios. Este mismo criterio es afirmado por Abalde y Armentano (2014), quienes indican que para poder realizar las predicciones de las utilidades se calcula un conjunto de vecinos usando la medida de similitud entre usuarios, seleccionando los k más cercanos y que el coeficiente de Jaccard es el indicado para este tipo de situaciones.

1.4 Objetivos

1.4.1 Objetivo general

Proponer un modelo de recomendación basado en competencias que determine acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.

1.4.2 Objetivos específicos

- Realizar el pre-procesamiento de datos a partir de dos tipos de datos de entrada (calificaciones pasadas y actuales de los estudiantes) orientado a una competencia del curso de Matemática II.
- Realizar el filtro de estudiantes antiguos a través de su rendimiento académico para elegir el nivel de competencia esperado del curso de Matemática II.
- Calcular la similitud de la curva de rendimiento entre el estudiante actual y los estudiantes antiguos usando el coeficiente de Jaccard.
- Puntuar los ítems de recomendación y almacenarlos en un repositorio para la posterior entrega de las recomendaciones al estudiante actual del curso de Matemática II.



1.5 Justificación e importancia

1.5.1 Justificación

Los grandes volúmenes de datos que se generan día a día en el ámbito académico universitario resulta ser una carga en lugar de ser beneficioso, detrás de esta cantidad de datos muchas veces se esconden patrones que pueden intervenir en el proceso de aprendizaje y verse reflejado en el rendimiento académico. Es ahí donde un modelo de recomendación es la mejor opción para poder procesar estos datos, realizar un filtrado y brindar recomendaciones personalizadas para los estudiantes.

El modelo de recomendación de acciones propuesto para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua está enfocado en la búsqueda y planteamiento de las mejores recomendaciones (acciones) que apoyen a los estudiantes en su proceso de aprendizaje en el entorno no presencial que actualmente se da, y también en la presencialidad, una vez que se reanude. A su vez, esto puede contribuir en la solución del problema que hoy en día las universidades enfrentan, como es la reprobación de cursos, repetición de ciclos académicos y la deserción estudiantil.

1.5.2 Importancia

La contribución principal de esta investigación está dada en la creación de matrices bases para almacenar los perfiles de los elementos básicos de un modelo de recomendación, como son las calificaciones de los estudiantes, acciones o ítems a recomendar y las puntuaciones que los estudiantes asignan a cada una de las acciones a recomendar. Cuando un estudio de este tipo arranca en frío, es importante definir estas matrices bases que serán el inicio a investigaciones de esta índole. Según Vaca (2017), es importante que el sistema de recomendación conste de cuatro partes: un repositorio de tareas que son los ítems a recomendar, el modelo del usuario, el recomendador y un ranking de tareas generadas que se muestran al alumno; las mismas que la presente investigación busca definir.



1.6 Hipótesis

1.6.1 Hipótesis general

El modelo de recomendación basado en competencias genera acciones para el curso Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.

1.6.2 Hipótesis específicas

- Con respecto al Objetivo específico 1, no se plantea hipótesis especifica porque corresponde a un objetivo descriptivo.
- Los estudiantes antiguos presentan un nivel de competencia esperado mayor a 10 que corresponden a una calificación aprobada.
- Las coincidencias de las calificaciones de los estudiantes antiguos y actuales se determinan a través del coeficiente de Jaccard.
- La puntuación de los ítems de recomendación se determina a través de la ecuación de puntuación para recomendar las mejores acciones al estudiante actual del curso de Matemática II.

1.7 Variables de la investigación

1.7.1 Identificación de las variables

- Variable Independiente (VI): Modelo de recomendación basado en competencias.
- Variable Dependiente (VD): Recomendaciones para el curso Matemática II.

1.7.2 Operacionalización de las variables

Las dimensiones, definición conceptual, definición operacional, indicadores y las escalas de medición para cada una de las variables se muestran en la tabla 1 y tabla 2.



Tabla 1. Operacionalización de la variable independiente

Variable Independiente (VI): Modelo de recomendación basado en competencias		
Definición conceptual: Herramienta que busca mejorar las experiencias de los usuarios, tales como encontrar ítems (recomendaciones) personalizados para los estudiantes asistiéndolos a lo largo de algún curso diseñado bajo el enfoque de evaluación por competencias. (Chavarriaga, et al., 2014)		
Definición operacional: El modelo de recomendación partirá a partir de unas matrices de datos de entrada con las calificaciones de los estudiantes y recomendaciones (acciones) trabajadas según las dimensiones (competencias) del curso de matemática II.		
Dimensiones	**Indicadores**	**Escala de medición**
D.1. Similitud de la curva de rendimiento de los estudiantes.	Coeficiente de Similitud de Tanimoto (Jaccard). **	0 - 1**
D.1. Definición Conceptual: La similitud son las coincidencias en la curva de aprendizaje de los estudiantes entre un estudiante actual y los antiguos. (Chavarriaga, et al., 2014)		
D.1. Definición Operacional: La similitud del rendimiento permitirá, a partir de las calificaciones de los estudiantes, obtener los vecinos k más cercanos para la evaluación de las recomendaciones.		

** Extraído de Mendoza, et al.(2019)

Fuente. Elaborado por la autora

Tabla 2. Operacionalización de la variable dependiente

Variable Dependiente (VD): Recomendaciones para el curso de Matemática II		
Definición conceptual: Cabero y López (2009) en su artículo describen un instrumento didáctico de modelo y estrategias de enseñanza de cursos universitarios en red, donde su dimensión de aspectos técnicos presenta el componente de "Significatividad/ comprensión", el cual alude a los procedimientos y recomendaciones didácticas orientadas hacia la comprensión y el aprendizaje significativo, tales como: mapas, esquemas conceptuales, metáforas, ejemplos, preguntas, glosario, simulaciones y todos aquellos recursos que contribuyen a relacionar los nuevos contenidos y experiencias con ideas y experiencias personales. Otros componentes que mencionan son la interacción social, integración y la funcionalidad.		
Definición operacional: Ítems a recomendar (acciones) debidamente clasificadas y establecidas por un especialista en la materia, así mismo, parte de ellas serán adquiridas mediante conocimiento social de parte de los estudiantes para ser almacenadas en un repositorio.		
Dimensiones	**Indicadores**	**Escala de medición**
	1.1. Analiza las propiedades de las teorías de ante derivadas e integrales. *	

C.1. Reconoce los diferentes métodos de integración	**1.2.** Identifica los conceptos básicos de integrales. * **1.3.** Compila los estudios sobre los temas de métodos de integración. * **1.4.** Formula y demuestra los problemas de integración. *	0-20 **
C.2. Aplica el cálculo integral en el desarrollo y resolución de problemas óptimos de forma gráfica, analítica y numéricamente.	**2.1.** Analiza las propiedades de las integrales definidas. * **2.2.** Identifica los conceptos básicos de integrales definidas. * **2.3.** Compila los estudios sobre los temas de integración. * **2.4.** Formula y demuestra los problemas de integración definida. *	0-20 **

C.1. Definición conceptual: Es la competencia N° 1 plasmada para la evaluación del curso de Matemática II, cálculo integral. (Alvarado y García, 2016)

C.1. Definición operacional: A partir de la competencia 1 se definen indicadores que son las tareas que deben cumplirse para alcanzar la competencia en mención. De la misma forma se plantea los ítems de recomendación, ya que están alineados a los indicadores que pertenecen a la competencia 1.

C.2. Definición conceptual: Es la competencia N° 2 plasmada para la evaluación del curso de Matemática II, cálculo integral. (Alvarado y García, 2016)

C.2. Definición operacional: A partir de la competencia 2 se definen indicadores que son las tareas que deben cumplirse para alcanzar la competencia en mención. De la misma forma se plantea los ítems de recomendación, ya que están alineados a los indicadores que pertenecen a la competencia 2.

* Indicadores propios para la evaluación de las competencias del curso de matemática II, cálculo integral extraído de Alvarado y García (2016).

**Escala de medición vigesimal establecida para la calificación de las asignaturas según el artículo 100 del Reglamento Académico de la Universidad Nacional de Moquegua aprobado con Resolución de Comisión Organizadora N° 213 – 2017 - UNAM.

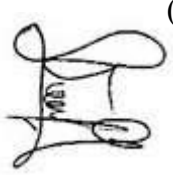
Fuente. Elaborado por la autora

II. MARCO TEÓRICO

2.1 Bases teóricas de la investigación

2.1.1 Educación basada en competencias

Coronado (2009) define la competencia como "conjunto integrado y dinámico de saberes, habilidades, capacidades y destrezas, actitudes y valores puestos en juego en la toma de decisiones, en la acción, en el desempeño del sujeto en un determinado espacio (profesional, laboral, etc.)".



Rieckmann (2017) menciona que las competencias se pueden desarrollar: “Se desarrolla en la acción - en la base de las experiencias y la reflexión”.

Según las definiciones dadas, podemos afirmar que las competencias son una combinación de destrezas, habilidades y conocimientos necesarios para desempeñar una tarea específica, son el resultado de la interacción de las destrezas, habilidades y el conocimiento.

En cuanto a una enseñanza basada en competencias, las dimensiones metas son: la capacidad académica, capacidad profesional, participación ciudadana y el desarrollo de la personalidad. (Rieckmann, 2017)

2.1.2 Matemáticas en educación universitaria

El rendimiento académico constituye una de las dimensiones más importantes del proceso de enseñanza-aprendizaje, es denominado como desempeño académico o aptitud escolar.

Jiménez (2000) define el rendimiento académico como “un nivel de conocimientos sobre un área o materia comparado con la edad y el nivel académico”.

Sánchez (2000) indica que el rendimiento académico es la suma de factores diferentes y complejos en la persona que aprende y se mide mediante calificaciones cuantitativas obtenidas, cuyos resultados muestran las materias aprobadas o desaprobadas, la deserción y el grado de éxito académico.

A nivel universitario, el rendimiento académico de los estudiantes se caracteriza por bajas calificaciones y alto porcentaje de ausencia en los exámenes, lo que ocasiona la reprobación, repetición de cursos o ciclos académicos hasta deserción sin completar sus estudios. (Lázaro, et al., 2017)

El rendimiento académico en matemáticas a nivel mundial, según Vargas y Montero (2016), se caracteriza por ser bajo y resulta ser una disciplina poco accesible por gran parte de estudiantes.

Castillo, et al. (2020) presentan un artículo de investigación en el cual concluyen que, a nivel general, tanto docentes como estudiantes coinciden que los conocimientos adquiridos en secundaria no son los suficientes como para enfrentar el curso de Matemáticas a nivel universitario. Las razones expuestas por el autor en mención no son muy ajenas a la realidad de la Escuela Profesional de Ingeniería de Sistemas de la



Universidad Nacional de Moquegua, por lo que la presente investigación tomara como referencia el curso de Matemática II para desarrollar el modelo de recomendación de acciones que sirva de apoyo al rendimiento académico.

2.1.3 Minería de Datos

Es una técnica que extrae data de grandes contenedores de información, realizando un proceso de recopilación, preparación y explotación de datos

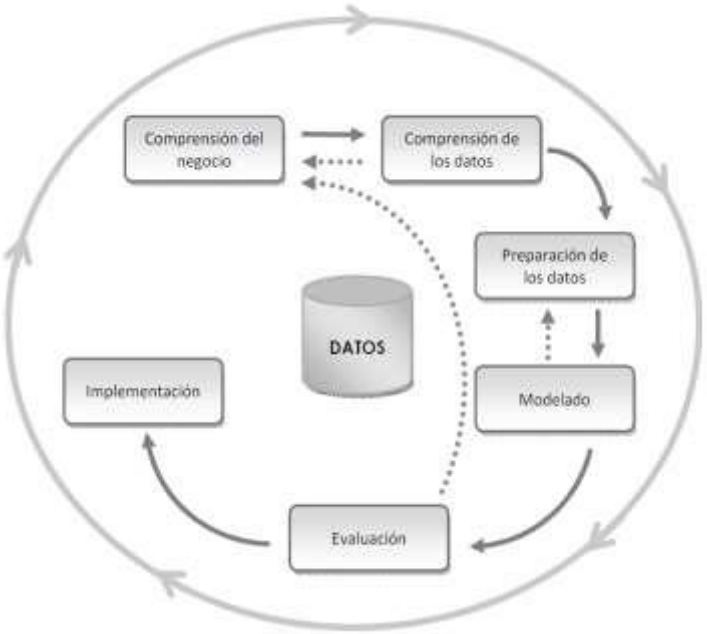
Ruiz, F. et al (2006) Los sistemas de recomendación y predicción forman partes de las tecnologías de estudio de la minería de datos, la cual aprovecha el conocimiento de las personas acerca de determinados objetos para ayudarse entre sí, dando a conocer nuevos objetos que probablemente sean de su interés.

2.1.4 Modelo CRISP-DM

Azevedo y Santos (2008) indican que CRISP-DM (siglas en inglés Cross Industria Standard Procesos for Data Mining), proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos. Este modelo, proporciona las fases de un proyecto de manera similar a la ingeniería de software, contempla el análisis de datos estableciendo un buen contexto que influye en la elaboración de modelos.

Ruiz, F. et al (2006) Indican que se debe centrar en un buen diseño o perspectiva, remitiéndose a las mejores prácticas para obtener buenos resultados, por lo que recomiendan al modelo CRISP-DM como un estándar para la planeación, organización, dirección y control de un proyecto de minería de datos. El ciclo de vida de un proyecto de minería de datos consiste en seis fases que se muestra en la figura 1.

Figura 1. Fases de la metodología CRISP-DM



Fuente. Extraído de Azevedo y Santos (2008)

2.1.5 Modelos de recomendación

Minsky, M. (1965) define "Un Modelo para un sistema y un experimento es cualquier objeto al que puede aplicarse E para responder preguntas acerca de S", con esta definición se afirma que un modelo puede ser un objeto material, entendimiento o comprensión acerca de cómo funciona un sistema o programa computacional. Existen modelos que son codificables, como los modelos matemáticos o modelos algorítmicos. (D. Cova, 2016)

Con lo afirmado por los autores en mención, podemos deducir que los modelos de recomendación todos aquellos algoritmos que nos proveen de recomendaciones de ciertos criterios o necesidades, los mismos que forman parte de un Sistema de recomendación que se define en adelante.

2.1.6 Sistemas de recomendación

Ricci, et al. (2010) definen los sistemas de recomendación como herramientas de software, plataforma o módulos que proporcionan sugerencias personalizadas de utilidad a los usuarios.

Los sistemas de recomendación son sistemas de filtrado de información que asisten a los usuarios en la búsqueda de información, productos o servicios que se ajusten a sus

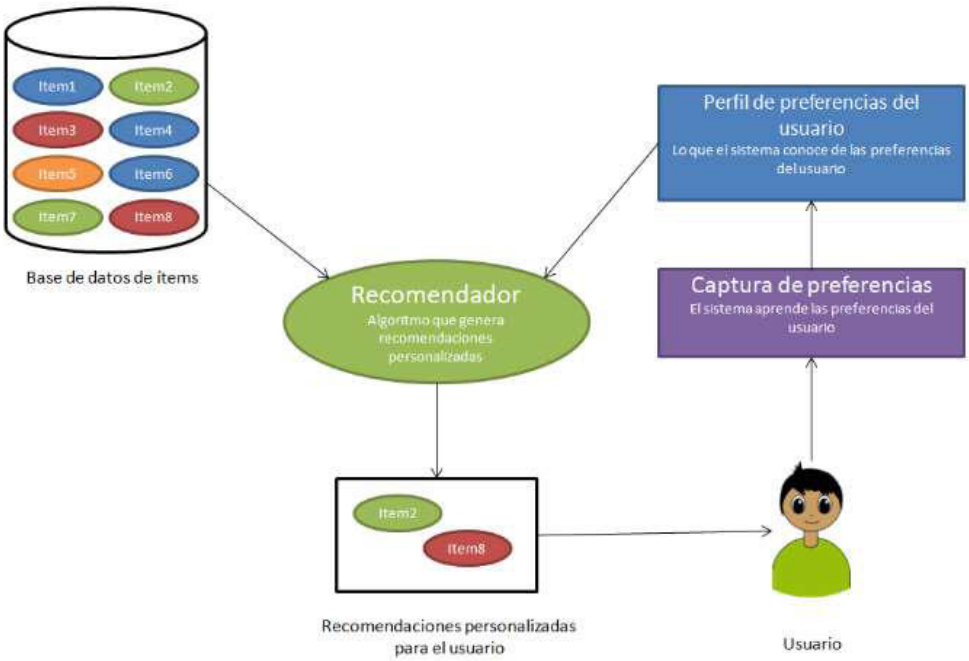
preferencias. Es decir, estos sistemas proveen de recomendaciones personalizadas a un determinado usuario. (Burke, 2002)

Los sistemas de recomendación ayudan a unir a los usuarios con ítems de manera que el sistema almacena los intereses y las preferencias de cada usuario y le propone recomendaciones que se adapten a esos intereses. (Dietmar Jannach, 2010)

Hoy en día, un sistema de recomendación tiene varios dominios de aplicación, sobre todo en el ámbito de comercio electrónico, en el periodismo, entretenimiento y otros, esto es posible gracias al procesamiento de datos previos, analizando similitudes y calificaciones para entregar las recomendaciones.

Según Vaca (2017), para obtener la información necesaria y llevar a cabo las recomendaciones, un sistema de recomendación debe contar con tres elementos principales: ítems, usuarios e interacciones (ver figura 2). Los ítems son los objetos que el sistema va a recomendar al usuario, éstos tienen sus propias características y se les asignara un valor (puntuación).

Figura 2. Elementos de un Sistema de Recomendación



Fuente. Extraído de Vaca Sánchez (2017)

En el ámbito educativo, los sistemas de recomendación tienen como objetivo encontrar buenos ítems personalizados para un estudiante, por ejemplo, encontrar recursos o realizar predicciones, de tal manera que apoyen al rendimiento académico. (Adolphs, 2009) (Wan, et al., 2001).

2.1.7 Técnicas de implementación de un sistema de recomendación

Dietmar, et al. (2011) afirman que para la implementación de un sistema de recomendación existen distintas técnicas: basadas en filtrado colaborativo, basadas en contenido, basadas en conocimiento y los híbridos. Mucho depende de la información con la que se cuenta o se necesite. El mismo autor define lo siguiente:

- **S.R. Basados en contenido:** las recomendaciones se realizan basadas en las similitudes de los ítems. Dichas recomendaciones se basan en la premisa que, si un usuario escogió un elemento, entonces le gustarán otros similares.
- **S.R. Basados en conocimiento:** las recomendaciones se realizan basadas en la información que se conoce por medio de la interacción del usuario. La premisa de estas recomendaciones se basa en satisfacer las necesidades en lugar de los intereses, para ello es necesario una base de conocimientos donde se definen los criterios que deben cumplir los ítems para ser recomendados.
- **S.R. Basados en filtrado colaborativo:** las recomendaciones se realizan basadas en las calificaciones que los usuarios hacen a los ítems a recomendar. Se trabaja con una matriz de puntuaciones de los ítems y trabajan con la premisa de que, si a un usuario le han gustado ciertos elementos porque los ha calificado, entonces le gustará elementos que hayan sido calificados de manera similar por otros usuarios. Esta técnica es la que se estará desarrollando en la presente investigación.
- **S.R. Híbridos:** Son aquellos que combinan las técnicas antes mencionadas, con la finalidad de subsanar las debilidades individuales que puedan presentar.

Mendoza, et al. (2019) en su artículo indican que, sin importar la técnica a usar, el proceso de obtener las recomendaciones se basa en la similaridad, ya sea de los ítems o del usuario, para ello se hace uso de distintas métricas de similitud.



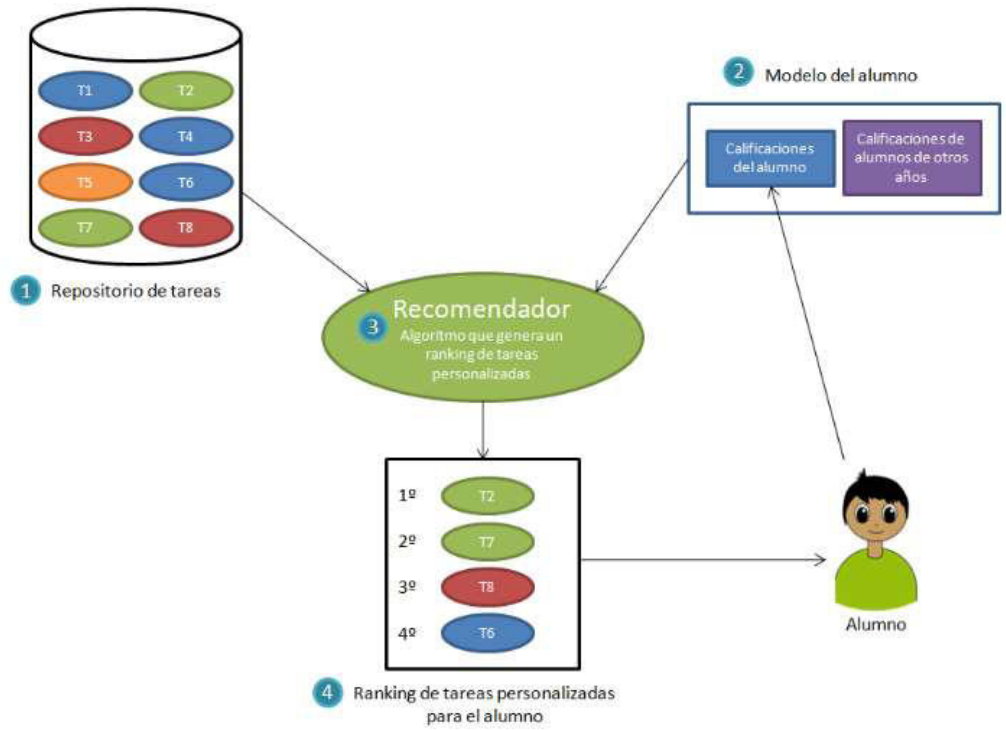
2.1.8 Filtrado colaborativo

Es una técnica que trabaja con algoritmos que se basan en las elecciones anteriores de un usuario y en las elecciones de otros usuarios con similar historial de valoraciones, sugiriendo así nuevos ítems de recomendación para el usuario. Dentro de este tipo de técnica se encuentras dos clases: los basados en usuario y los basado en ítems. (Abad, et al., 2010)

La diferencia radica al momento de realizar el cálculo de las similitudes, una se enfoca a buscar usuarios similares y la otra en búsqueda de los ítems con calificación similar. La presente investigación, se basa en realizar el filtrado colaborativo en base a los usuarios (estudiantes), ya que se analizará la similitud de su curva de rendimiento académico de los estudiantes antiguos con los estudiantes actuales a recomendar.

Para ello, en la figura 3 se muestra los componentes que tendrá el modelo recomendador propuesto:

Figura 3. Funcionamiento del modelo de recomendación



Fuente. Extraído de Vaca Sánchez (2017)

2.1.9 Métricas de similaridad

El paso fundamental de todo modelo de recomendación, es calcular que tan parecidos son los elementos para seleccionar los más parecidos. Varios autores presentan distintas métricas para este tipo de técnica. Mendoza, Laureano y Pérez (2019) hacen un resumen de ellas, las cuales son:

- Diferencia cuadrática media (MSD).
- Similitud basada en coseno.
- Similitud basada en correlación.
- Similitud de coseno ajustado.
- Coeficiente de similitud de Tanimoto (Jaccard).
- Similitud basada en Log-likelihood.
- Coeficiente de correlación de Pearson restringido (CPCC).
- El coeficiente de correlación de Pearson ponderado (WPCC).
- Proximity, Impact and Popularity (PIP).
- Proximity, Significance, Singularity (PSS).
- New Heuristic Similarity Model (NHSM).
- Descomposición en valores singulares (SVD).

Esta investigación se basa en el artículo presentado por Chavarriaga, et al. (2014), por lo que se tomará como medida de similitud el Coeficiente de Tanimoto (Jaccard).

2.1.10 Coeficiente de similitud de Tanimoto (Jaccard)

El coeficiente de Jaccard representa la cantidad de elementos por la que dos usuarios muestran cierta preferencia, dividida por la cantidad de elementos por los cuales el usuario en evaluación expresa cierta preferencia, es decir, es la relación del tamaño de la intersección entre el tamaño de la unión de los elementos preferidos de ambos y se denota con la siguiente ecuación:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

La Ecuación 1 es utilizada como una medida de similitud ya que cumple con las propiedades requeridas para medir las recomendaciones. (Owin y Robin, 2009).

2.1.11 Métricas de evaluación de un sistema de recomendación

Mendoza, et al. (2019) en su artículo presentado afirman que es necesario establecer métricas que permitan analizar la validez de las recomendaciones, es decir, es primordial conocer que tan alejados están los valores predichos por el algoritmo recomendador. Los más utilizados son: Error Medio Absoluto (MAE), coverage, precisión y recall. El primero corresponde a una métrica de tipo estadístico, mientras que los otros evalúan la efectividad de la decisión. (Galán, 2007)

Error de Medio Absoluto (MAE): Es el parámetro de evaluación más utilizado, mide la desviación de las recomendaciones predichas y los valores reales. A menor MAE, las evaluaciones del usuario son mejor predichas.

La Ecuación 2 se utiliza para calcular el MAE de un usuario, que es el valor medio de la diferencia entre las predicciones realizadas y las votaciones reales del usuario.

$$MAE_u = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (2)$$

Para el cálculo del MAE del sistema se utiliza la Ecuación 3, definida como el promedio del error medio absoluto cometido en cada usuario.

$$MAE_s = \frac{\sum_{i=1}^{N} MAE_u}{N} \quad (3)$$

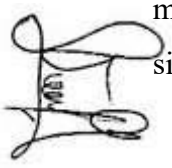
Coverage: La finalidad de esta medida es medir la capacidad de recomendación que tienen los k vecinos de cada usuario, es decir, indica el porcentaje de ítems que han podido ser predichos respecto del total de los que podían haberlo sido. Esta métrica se calcula con la Ecuación 4.

$$coverage_u = 100. \frac{\#\{i \in \frac{I}{r_{u,i}} = * \wedge p_{u,i} \neq *\}}{\#\{i \in I \,/ r_{u,i} = *\}} \quad (4)$$

Para evaluar el coverage del sistema se realiza mediante la Ecuación 5, calculando el valor promedio de la métrica de coverage para cada usuario.

$$coverage_s = \frac{1}{\#U} . \sum_{u \in U} coverage_u \quad (5)$$

Precisión y Recall: Estas dos métricas habitualmente aparecen unidas, cuya finalidad es medir la calidad de las recomendaciones dadas por un sistema. La precisión comprueba si los ítems recomendados a los usuarios activos eran relevantes para él. El recall es el



encargado de comprobar qué porcentaje de los ítems relevantes para un usuario le fueron recomendados.

Con la Ecuación 6 se calcula la precisión de un usuario u, el cual está definido como el porcentaje de ítems relevantes entre los recomendados.

$$precisión_u = \frac{\#Y_u}{\#Y_u + \#N_u} \quad (6)$$

Para calcular la precisión del sistema se hace uso de la Ecuación 7, la misma que plasma el valor promedio de la precisión para cada usuario.

$$precisión_s = \frac{1}{\#U} . \sum_{u \in U} precisión_u \quad (7)$$

El recall de un usuario u está definido por el porcentaje de ítems relevantes recomendados respecto del total de ellos, tal y como se plasma en la Ecuación 8.

$$recall_u = \frac{\#Y_u}{\#Y_u + \#O_u} \quad (8)$$

Por último, para el cálculo de la métrica recall del sistema se utiliza la Ecuación 9, el cual está definido como el valor promedio del recall de cada usuario.

$$recall_s = \frac{1}{\#U} . \sum_{u \in U} recall_u \quad (9)$$

2.2 Definición de términos

2.2.1 Sistemas de Recomendación (SR)

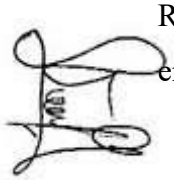
Los sistemas de recomendación son sistemas de filtrado de información que asisten a los usuarios en la búsqueda de información, productos o servicios que se ajusten a sus preferencias. Es decir, estos sistemas proveen de recomendaciones personalizadas a un determinado usuario. (Burke, 2002)

2.2.2 Filtrado Colaborativo basado en usuario (FCU)

Es un método de filtrado colaborativo utilizado para encontrar los *k* usuarios más cercanos dentro de una vecindad. (Del Pino, et al., 2011).

2.2.3 Similaridad

Representación numérica del grado de semejanza entre dos usuarios (estudiantes), basada en las preferencias sobre los ítems (acciones) en común. (Del Pino, et al., 2011)



2.2.4 Métrica de similaridad

Medida que determina el grado de similaridad entre dos objetos. (Del Pino, et al., 2011)

2.2.5 Modelo de datos

Representación de un repositorio de información de estudiantes y preferencias asociadas a las acciones. (Del Pino, et al., 2011)

2.2.6 Vecindad

Conjunto de usuarios (estudiantes) semejantes, construido en base a las similitudes de los usuarios (estudiantes). (Del Pino, et al., 2011)

2.2.7 Encuesta

Casas, et al. (2003) definen como la encuesta como una técnica de investigación. Ésta es ampliamente utilizada ya que nos permite obtener y elaborar datos de modo rápido y eficaz.

2.2.8 Cuestionario

Casas, et al. (2003) en su artículo sobre técnicas de investigación, afirma que, para toda encuesta su instrumento básico es el cuestionario, por lo que en esta investigación se aplicará cuestionarios a los alumnos de nuestra muestra para obtener las recomendaciones puntuadas.

2.2.9 Software estadístico y gráfico R (framework RecommenderLab)

Un marco que permite desarrollar y probar algoritmos de recomendación con un paquete del software R que proporciona una infraestructura para desarrollar y probar algoritmos recomendadores de distintos tipos. (Torres, 2015)

2.2.10 Umbral

Grado de similaridad del sistema de recomendación. El umbral se establece para que, de acuerdo a ello, toda recomendación con una puntuación mayor o igual a ese umbral es relevante, y menor es no relevante. (Gibaja, et al., 2016)

2.2.11 Error medio absoluto (MAE)

El error absoluto medio mide la desviación absoluta media entre una valoración prevista y la valoración real del usuario. (Herlocker, et al., 2004)



2.2.12 Precisión

Mide el porcentaje de relevancia de ítems recomendados a los usuarios activos. (Mendoza, et al., 2019)

2.2.13 Recall

Mide el porcentaje de ítems relevantes que fueron recomendados al usuario. (Mendoza, et al.,2019)

III. METODOLOGÍA

3.1 Tipo y Nivel de la investigación

3.1.1 Tipo

Según la definición de Hernández-Sampieri, et al. (2010), la presente investigación es aplicada ya que se busca dar solución a un problema.

3.1.2 Diseño

Para esta investigación se recurrirá a un diseño no experimental que se aplicará de manera transversal.

Hernández-Sampieri, et al. (2010), definen el diseño no experimental como: *"... la investigación que se realiza sin manipular deliberadamente variables. Es decir, se trata de investigación donde no hacemos variar en forma intencional las variables independientes."* En la investigación no experimental se desarrollará un modelo de recomendación en base a los datos recolectados y las recomendaciones formuladas por un especialista y el conocimiento social de los estudiantes.

"Los diseños de investigación transeccional o transversal recolectan datos en un solo momento, en un tiempo único. Su propósito es describir variables y analizar su incidencia e interrelación en un momento dado (o describir comunidades, eventos, fenómenos o contextos)" (Hernandez-Sampiere, et al., 2010). Este estudio es del tipo transversal porque se recolecta datos en un único período, los cuales son procesados para construir un modelo de datos de recomendación. Esta recolección de datos se da en un contexto de enseñanza no presencial correspondiente a los periodos académicos 2020-1 y 2020-2.



3.2 Alcance de la investigación

La investigación tiene un alcance descriptiva y explicativa, ya que se especificarán características propias de los estudiantes para plasmar perfiles del objeto en estudio llegando a responder el comportamiento de las variables. (Hernández-Sampieri, et al., 2010).

3.3 Ámbito temporal y espacial

3.3.1 Ámbito temporal

La investigación se ejecutó en un período de 09 meses:

- Fecha de inicio: 01 abril 2021
- Fecha de término: 30 diciembre 2021

3.3.2 Ámbito espacial

El problema de la investigación está enfocado en el siguiente espacio geográfico:

- Institución intervenida: Universidad Nacional de Moquegua
- Dirección: Urb. Ciudad Jardín S/N (Filial Ilo de la UNAM).
- Provincia: Ilo
- Región: Moquegua
- País: Perú

3.4 Población y muestra

3.4.1 Población

Registros de estudiantes matriculados en el curso de Matemática II en la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua, desde el periodo 2017 hasta 2021-1, tal y como se muestra en la tabla 3.

Tabla 3. Registros de estudiantes matriculados en Matemática II

Año	1	2	Total
2017	16	19	35
2018	15	22	37
2019	24	21	45
2020	24	33	57



2021	24	-	24
Total	**103**	**95**	**198**

Fuente. Datos proporcionados por la DASA de la UNAM

3.4.2 Muestra

La muestra establecida para esta investigación es de tipo no probabilística. Hernandez-Sampiere, et al. (2010) mencionan que para este tipo de muestra *"la elección de los elementos no depende de la probabilidad, sino de las causas relacionadas con las características del investigador o del que hace la muestra"*.

Esta investigación se dará en los únicos períodos que forman parte de la enseñanza no presencial, por lo cual, la muestra está dada por los registros de estudiantes matriculados en la Escuela Profesional de Ingeniería de Sistemas e Informática en el curso de Matemática II periodo 2020-1, 2020-2 y 2021-1 tal y como se muestra en la tabla 4.

Tabla 4. Registros de estudiantes matriculados en Matemática II

Año	1	2	Total
2020	24	33	57
2021	24	-	24
Total	48	33	81

Fuente. Datos proporcionados por la DASA de la UNAM

Como se muestra en la tabla 4, son pocos los registros históricos siendo una muestra pequeña para estudios de este tipo, sin embargo, es válido trabajar con ellos, estudios como el de Gibaja, et al. (2016) que en su artículo denominado "Aplicación móvil para la recomendación de asignaturas en los estudios universitarios de grado", en la etapa experimental participan 28 estudiantes, y en base a las similitudes con ellos se generan las recomendaciones a entregar obteniendo resultados apropiados.

3.5 Técnicas e instrumentos para la recolección de datos

3.5.1 Técnicas



- Encuesta
- Filtrado Colaborativo basado en usuario

- Medición de Similitud Basada en Coeficiente de Jaccard
- Salida de un Repositorio de recomendaciones (ítems). Chavarriaga, et al. (2014) indican en su artículo que cuando un estudiante califica algún ítem, esta acción tiene que registrarse en repositorio de recomendaciones.

3.5.2 Instrumentos

- Cuestionarios
- Software Estadístico y gráfico R (framework RecommenderLab)
- Python (framework Crab).

3.6 Procedimientos

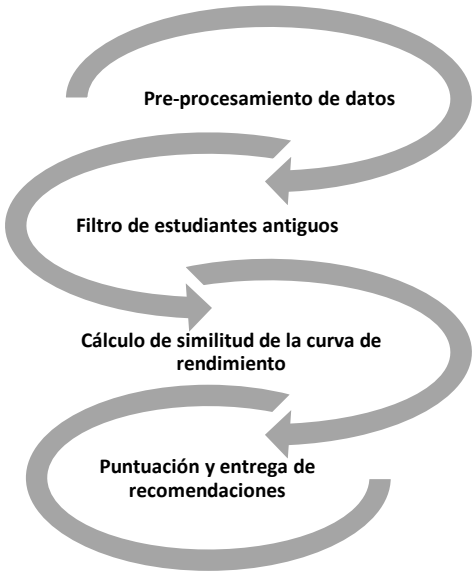
Los procedimientos a seguir para creación del modelo de recomendación están alineados a las fases de la metodología CRISP-DM, los cuales son:

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Implementación

Así mismo, dentro de la fase del modelado, se desarrollan las cuatro etapas para la obtención del modelo en sí, las cuales están plasmadas en la figura 4.



Figura 4. Etapas del modelado de la creación del modelo de recomendación



Fuente. Elaborado por la autora

3.7 Análisis y procesamiento de los datos

En este apartado se describirá las actividades que permitirán la obtención del modelo de recomendación siguiendo las fases de la metodología CRISP-DM.

3.7.1 Comprensión del negocio

Esta primera fase de la metodología CRISP-DM plantea los objetivos y requisitos del proyecto desde el enfoque del negocio, es decir, desde la perspectiva de la UNAM con el objeto de disminuir el porcentaje de estudiantes desaprobados en el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática por medio del modelo de recomendación de acciones propuesto.

3.7.1.1 Determinar los objetivos del negocio

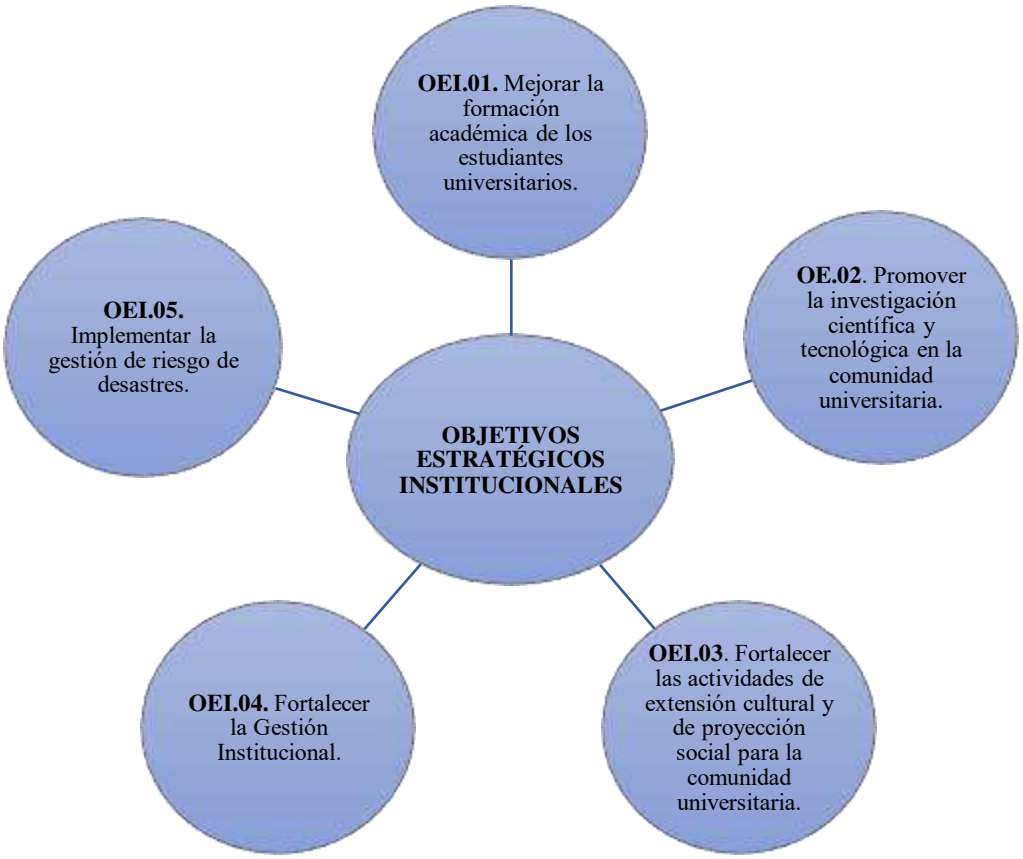
El proyecto tiene por objeto realizar un modelo de recomendación que determine acciones para el curso en cuestión, de tal manera que se pueda entregar esas acciones a los estudiantes actuales del curso de Matemática II de la EPISI con la finalidad de contribuir en uno de los OEI de la UNAM y disminuir el porcentaje de estudiantes desaprobados en el curso.

El proyecto se enmarca en un contexto de educación no presencial, correspondiente a los ciclos 2020-1, 2020-2 y 2021-1- El desarrollo del curso en evaluación fue dado por un

mismo docente, quien facilito la información necesaria para llevar a cabo el diseño del modelo de recomendación.

Los objetivos del negocio son cinco, los cuales son los objetivos estratégicos institucionales de la universidad, los mismos que están planteados en el Modelo Educativo periodo 2020-2022 y se detallan en la figura 5.

Figura 5. Objetivos Institucionales de la UNAM



Fuente. Obtenido del Plan Estratégico Institucional 2020-2022

De los objetivos referenciados se observa el objetivo "OEI.01. Mejorar la formación académica de los estudiantes universitarios", el cual se alinea a los objetivos del presente proyecto, considerando que uno de los indicadores de este objetivo es el porcentaje de egresados que culminan sus estudios según la duración del programa curricular establecido, lo cual no se logra con éxito cuando el estudiante desaprueba un curso.

El curso de Matemática es uno de los cursos cuyo rendimiento académico a nivel mundial se caracteriza por ser bajo y resulta ser una disciplina poco accesible por gran parte de los

estudiantes de ingeniería (Vargas y Montero, 2016), en el caso de la EPISI y más específicamente en el curso de Matemática II, no es ajeno a esta realidad, cuyas estadísticas de desaprobados se encuentran en la tabla 5:

Tabla 5. Porcentaje de estudiantes desaprobados en Matemática II

CURSO MATEMÁTICA II			
N°	**SEMESTRE ACADÉMICO**	**% APROB.**	**% DESAPR.**
1	2018 - 1	46,70%	53,30%
2	2018 - 2	50%	50%
3	2019 - 1	56,52%	43,48%
4	2019 - 2	85%	15%
5	2020 - 1	62,50%	37,50%
6	2020 - 2	72,72%	27,28%

Fuente. Obtenido de la Dirección de Actividades y Servicios Académicos de la UNAM.

Como se puede observar, generalmente más del 25% de los estudiantes desaprueban el curso, generando el incumplimiento del objetivo institucional. Con el modelo de recomendación se buscar mejorar esa situación, disminuyendo los porcentajes de desaprobados.

3.7.1.2 Valoración de la situación

Para el desarrollo del proyecto se contó con el apoyo del docente del curso de Matemática II, quien estuvo a cargo de los ciclos en evaluación, es por ese medio que se obtuvo los datos referentes a las calificaciones de los estudiantes. Así mismo, no se cuenta con un modelo o sistema que almacene datos referentes a estrategias o acciones que se puedan recomendar a los estudiantes, por lo que se trabajó en la obtención de ello.

En cuanto a los recursos utilizados, el principal es el Software Rstudio, versión libre para la preparación, análisis y modelado de los datos. De los datos proporcionados por el docente se restringió los nombres e identificaciones de los estudiantes como parte del resguardo de identidad.



3.7.1.3 Determinar los objetivos del proyecto

El presente estudio plantea una metodología de investigación estructurada que contiene, entre otros puntos, la determinación del problema, objetivos de la investigación y sus

respectivas hipótesis. Por lo mismo, el objetivo general y objetivos específicos son los mismo consignado en la sección 1.4 Objetivos.

OG: Objetivo General. Proponer un modelo de recomendación basado en competencias que determine acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.

OE_1: Objetivo Específico 1. Realizar el pre-procesamiento de datos a partir de dos tipos de datos de entrada (calificaciones pasadas y actuales de los estudiantes) orientado a una competencia del curso de Matemática II.

OE_2: Objetivo Específico 2. Realizar el filtro de estudiantes antiguos a través de su rendimiento académico para elegir el nivel de competencia esperado del curso de Matemática II.

OE_3: Objetivo Específico 3. Calcular la similitud de la curva de rendimiento entre el estudiante actual y los estudiantes antiguos usando el coeficiente de Jaccard.

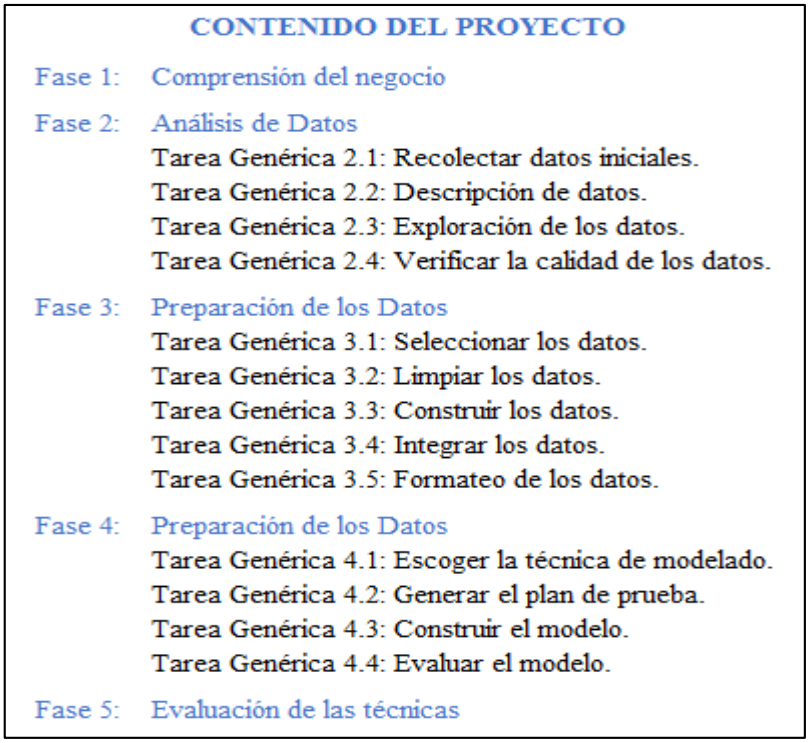
OE_4: Objetivo Específico 4. Puntuar los ítems de recomendación y almacenarlos en un repositorio para la posterior entrega de las recomendaciones al estudiante actual del curso de Matemática II.

3.7.1.4 Realizar el plan del proyecto

En esta fase se formuló el plan del proyecto abarcando cinco de las seis fases de la metodología CRISP-DM con sus respectivas tareas generales, tal y como se hace referencia en la figura 6. Cabe resaltar que no se incluye la última fase, la de implementación, ya que este proyecto tiene como objetivos la creación del modelo, como primera etapa, quedando como trabajo futuro la implementación.



Figura 6. Fases de la metodología CRISP-DM



Fuente. Elaborado por la autora

3.7.2 Análisis de los datos

En esta segunda fase de la metodología aplicada, se realiza un proceso analítico de recolección inicial de los datos, los cuales son todos los registros académicos del curso de Matemática II de la Escuela Profesional Ingeniería de Sistemas e Informática comprendido entre los periodos académicos 2020-1 y 2021-1.

3.7.2.1 Recolectar datos iniciales

La información requerida para la presente propuesta son los registros académicos del curso de Matemática II del periodo 2020-1, 2020-2, y 2021-1. Este último registro fue solicitado antes del término del periodo, precisamente para generar las recomendaciones antes de la última evaluación del curso.

Los datos con los que cuenta cada registro académico son sus nombres, apellidos y las calificaciones en cada una de las evaluaciones calificadas (competencias).

Así mismo, se recolecto mediante la técnica de encuestas la recolección de acciones a recomendar para el correcto desarrollo del curso en cuestión, las mismas que fueron

trabajadas con el docente del curso y se plasmaron en una matriz para una posterior calificación de los estudiantes antiguos.

3.7.2.2 Descripción de los datos

La primera información recolectada fueron los registros académicos, los cuales constan de los atributos que se muestran en la tabla 6.

Tabla 6. Atributos de los registros académicos

Atributos	Observación
1. *Apellidos y Nombres*	*Identificación del estudiante*
2. *Cuestionarios* 3. *Ejercicios de Límites* 4. *Parcial 1* 5. *Ejercicios de Integración* 6. *Ejercicios de Optimización* 7. *Parcial 2*	*Notas obtenidas por el estudiante en dicha evaluación*

Fuente. Elaborado por la autora

Así mismo, mediante encuestas aplicadas a los estudiantes antiguos y en coordinación con el docente del curso se obtuvieron acciones que ayudaron a los estudiantes antiguos del curso en poder aprobar la última evaluación del curso, las cuales se muestran en la tabla 7.

Tabla 7. Acciones a recomendar

ÍTEMS	RECOMENDACIONES	DESCRIPCIÓN
Ítem - Acción 1	Cap. 6 - Aplicaciones de la Integral – Cálculo Trascendentes Tempranas Cuarta Edición – Dennis G. Zill – Warren S. Wright	Volúmenes de sólidos: métodos de las rebanadas. Volúmenes de sólidos: método de discos. Volúmenes de sólidos: método de arandelas.
Ítem - Acción 2	Canal Julioprofenet	Explicación de coordenadas Polares. Volumen calculado con una Integral doble en Coordenadas Polares.
Ítem - Acción 3	Canal Academática	Volumen por el método de capas cilíndricas – Ejemplos. Momentos y centro de masa.



		Aplicaciones integrales dobles Momento y centro de masa. Integrales dobles sobre regiones polares generales. Integrales dobles sobre regiones polares generales – Ejemplo. Integrales dobles sobre regiones polares generales – Ejemplo.
Ítem - Acción 4	Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S.	Cálculo de áreas. Área de una región comprendida entre dos curvas. Volúmenes de sólidos de revolución. Longitud de una curva plana.
Ítem - Acción 5	Integrales Básicas – Canal Profe Alex	Integrales definidas. Área bajo la curva. Área entre curvas.
Ítem – Acción 6	Otras herramientas y bibliografías de apoyo	Eduardo Espinoza Ramos – Análisis Matemático I y II Máximo Mitacc - Fernando Hoyos-Félix Villanueva y otros Calculadoras online

Fuente. Elaborado por la autora

Por último, una vez plasmada las acciones que ayudaron a los estudiantes antiguos en el correcto desenvolvimiento de su último examen se procedió a enviar las acciones de la tabla 6 para que los estudiantes antiguos puedan realizar la puntuación a cada una de ellas, la puntuación va de 1 a 5, siendo esta ultima la mejor puntuación para la recomendación.

3.7.2.3 Exploración de los datos

La información respecto a los registros académicos se tiene en hojas de cálculo proporcionado por el docente, estos registros se encuentran estructuras en filas y columnas cuyas celdas contienen las notas de cada una de las actividades evaluadas en el periodo académico, tal y como lo muestra la figura 7 y 8.



Figura 7. Registro académico del curso de Matemática 2020-1

MATEMÁTICA II Salomón R. Ramos	NOTAS DE MATEMATICA II SEMESTRE 2020-I									
						7		8		9
	ASISTENCIA	EJERCICIOS DE LÍMITES	CUESTIONARIO 1	CUESTIONARIO 2	ASISTENCIA	Ejercicios de Derivación	CUESTIONARIO 3	Ejercicios de Optimización	EXAMEN PARCIAL	EXAMEN RESUELTO
[redacted]	-	18.00	16.00	4.00	20.00	14.00	4.00	14.00	-	14.00
[redacted]	-	13.00	20.00	16.00	20.00	12.00	16.00	10.00	-	10.00
[redacted]	-	18.00	16.00	4.00	20.00	5	12.00	5	-	-
[redacted]	-	5	16.00	8.00	-	5	8.00	5	-	10.00
[redacted]	-	5	12.00	4	-	5	12.00	5	-	-
[redacted]	-	5	20.00	8.00	.00	5	16.00	5	-	-
[redacted]	-	5	20.00	8.00	20.00	5	4.00	18.00	-	6.00
[redacted]	-	15.00	12.00	4.00	20.00	13.00	12.00	5	-	8.00

Fuente. Proporcionado por el docente de curso

Figura 8. Registro académico del curso de Matemática 2020-2

Carrera profesional:	INGENIERIA DE SISTEMAS E INFORMATICA
Fecha:	13-12-2020

Estudiantes	UNIDAD 1								UNIDAD 2							
	Semanas															
	4	5		6		9			10							11
	CUESTIONARIO N°	CUESTIONARIO 2	EJERCICIOS DE	Parcial 1A	EXAMEN RESUELT	ENTREGA DE	CUESTIONARIO N°	CUESTIONARIO N°	puntos en clase	puntod adicional	EXAMEN PARCIAL	EXAMEN RESUELT	RECUPERACION	SOLUCIÓN DE RECUPER	PRACTICAS Y TRABAJO	EXAMEN PACIAL
[redacted]	5	8	13	-	5	15	8,5	5	4	2	-	6	7	-	14,6	12
[redacted]	6	7	16	-	4	15	6,5	6,5	4	2	-	2	3	2	15,3	7
[redacted]	4	5	15	-	2	15	10	7	4	2	-	3	4	3	15	6
[redacted]	8	9	18	-	8	18	6	4,5	6	2	-	7	8	-	17,4	16
[redacted]	3	0	5	-	4	5	3,5	3	4	2	-	3	4	-	5,88	8
[redacted]	2	6	5	-	0	5	5	3	4	2	-	2	3	-	7,5	3
[redacted]	6	5	16	-	3	15	3,5	0	4	2	-	2	4	3	12,4	7
[redacted]	3	7	16	-	2	5	6	5	4	2	-	3	4	3	11,5	6
[redacted]	7	7	14	-	2	16	2,5	4,5	4	2	-	3	5	4	13,8	7

Fuente. Proporcionado por el docente de curso

Así mismo, la puntuación que se obtuvo de los estudiantes antiguos a las acciones a recomendar se exportar a una hoja de cálculo, la cual contiene los datos de los estudiantes y la puntuación que le dieron a cada una de las recomendaciones tal y como se muestra en la figura 9.



Figura 9. Puntuación de las acciones a recomendar

NOMBRE Y APELLIDOS	PERIODO EN EL QUE LLEVASTES MATEMATICA II	Cap. 6 - Aplicaciones de la Integral – Cálculo Trascendentes Tempranas Cuarta Edición - Dennis G. Zill - Warren S. Wright	Canal Julioprofenet	Canal Academatica	Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S.	Integrales Básicas – Canal Profe Alex	Libros recomendados y bibliografía de apoyo - Eduardo Espinoza Ramos - Análisis Matemático I y II - Máximo Mraco - Fernando Hoyos - Félix Villanueva y otros - Calculadoras online
Kevin Ronald Camilo Flores	2020-1	1	3	4	4	3	1
Flavio Mauricio Alarcon Torres	2020-1	4	1		4	2	5
Walter Enrique Callupe Aduviri	2020-2		5	3	4	3	1
Alex Salvador Castillon Villcaporna	2020-2	2	5	1	5	5	
Erika Cynthia Laqui Lope	2020-2	4	2	4	3	1	5
Selene Madel Maquera Ortega	2020-2	5	2	2	2	1	

Fuente. Elaborado por la autora

3.7.2.4 Verificar la calidad de los datos

Para esta actividad se procedió a verificar cada una de las celdas en las bases de datos que se obtuvieron. Esto con la finalidad de encontrar campo vacíos o nulos, de tal manera que no interfieran en el estudio de los mismos.

Así mismo, para validar los registros académicos con ayuda del docente a cargo, se trabajó y validó la correcta interpretación de cada una de las actividades evaluadas.

3.7.3 Preparación de los datos

En esta fase se procedió a trabajar con los registros académicos proporcionados del curso, para la selección de los datos relevantes, alineándolos a los requerimientos que necesitamos para dar inicio a la creación del modelo de recomendación. El objetivo es tener la data limpia para dar inicio a la siguiente fase.

3.7.3.1 Seleccionar los datos

A partir de los archivos BD_MAT_2020_1.xlsx, BD_MAT_2020_2.xlsx y BD_MAT_2021_1.xlsx proporcionados por el docente, se procede a transformarlos en formatos con extensión CSV, obteniendo los archivos con sus estancias respectivas para ser utilizadas en el software estadístico Rstudio y proceder con la limpieza de los datos.

3.7.3.2 Limpiar los datos

La limpieza que se realizó fue en los registros académicos de los estudiantes, ya que es en ellos donde la columna de nombres y apellidos no es relevante para esta investigación y por un tema de protección de los datos *(ver figura 10)*, sin embargo, antes de esa limpieza es necesario una reestructuración puesto que se necesita tener un identificador del estudiante para realizar el modelo de recomendación.



Figura 10. Limpieza de datos

NOMBRES	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
[redacted]	2020-1	20,00	15,00	8,00	17,00	15,00	12,00
[redacted]	2020-1	8,00	5,00	8,00	16,00	14,00	8,00
[redacted]	2020-1	16,00	15,00	8,00	15,00	14,00	7,00
[redacted]	2020-1	12,00	5,00	8,00	5,00	5,00	0,00
[redacted]	2020-1	12,00	13,00	8,00	16,00	15,00	9,00
[redacted]	2020-1	12,00	15,00	8,00	16,00	11,00	7,00
[redacted]	2020-1	16,00	5,00	8,00	14,00	5,00	5,00
[redacted]	2020-1	8,00	15,00	4,00	16,00	16,00	7,00
[redacted]	2020-1	20,00	18,00	8,00	14,00	5,00	5,00
[redacted]	2020-2	13,00	13,00	15,00	13,50	14,60	12,00

Fuente. Elaborado por la autora

3.7.3.3 Construir los datos

La estructuración que se realizó de los registros académicos de los estudiantes consta de la creación de una nueva columna como identificador de los mismos, esto con el propósito de tener la referencia con que estudiante se realiza la comparación de la curva de rendimiento académico. Esta columna es la de *ID*, la cual se creó a partir de las iniciales de los nombres y apellidos de los estudiantes, tal y como se puede observar en la figura 11.

Figura 11. Estructuración de los datos

NOMBRES	ID	CICLO	AE1	AE2	AE3	AE4	AE5
[redacted]	MMJ	2020-1	20,00	15,00	8,00	17,00	15,00
[redacted]	PMN	2020-1	8,00	5,00	8,00	16,00	14,00
[redacted]	PZA	2020-1	16,00	15,00	8,00	15,00	14,00
[redacted]	QHY	2020-1	12,00	5,00	8,00	5,00	5,00
[redacted]	TFG	2020-1	12,00	13,00	8,00	16,00	15,00
[redacted]	VDP	2020-1	12,00	15,00	8,00	16,00	11,00
[redacted]	VMW	2020-1	16,00	5,00	8,00	14,00	5,00
[redacted]	VCA	2020-1	8,00	15,00	4,00	16,00	16,00

Fuente. Elaborado por la autora

3.7.3.4 Integrar los datos

En esta tarea lo que se realizó es tener una sola base de datos en la hoja de cálculo que contiene tanto los registros académicos de los estudiantes del periodo 2020-1 y 2020-2, así como sus respectivas puntuaciones a las recomendaciones propuestas para su última evaluación del curso. Están integrados tal y como se puede visualizar en figura 12.



Figura 12. Integración de los datos

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	AC1	AC2	AC3
TPG	2020-1	12,00	13,00	8,00	16,00	15,00	9,00	5		
VDP	2020-1	12,00	15,00	8,00	16,00	11,00	7,00			3
VMW	2020-1	16,00	5,00	8,00	14,00	5,00	5,00	3		2
VCA	2020-1	8,00	15,00	4,00	16,00	16,00	7,00	1	2	4
ZAC	2020-1	20,00	18,00	8,00	14,00	5,00	5,00	3		3
ALY	2020-2	13,00	13,00	15,00	13,50	14,60	12,00	3	5	3
BVI	2020-2	13,00	16,00	15,00	13,00	15,30	7,00	5	5	4
BGB	2020-2	9,00	15,00	15,00	17,00	15,00	6,00		2	2
CRL	2020-2	17,00	18,00	18,00	10,50	17,40	16,00		5	4
CAW	2020-2	3,00	5,00	5,00	6,50	5,90	8,00		5	3
CPK	2020-2	8,00	5,00	5,00	8,00	7,50	3,00		1	1
CVA	2020-2	11,00	16,00	15,00	3,50	12,40	7,00	2	5	1

Fuente. Elaborado por la autora

3.7.3.5 Formateo de los datos

Los registros académicos de los estudiantes del curso de Matemática II fueron formateados con nuevas nomenclaturas para un mejor desarrollo del modelo de recomendación, *siendo el campo de identificación de los estudiantes la primera columna denominada ID*, la misma que cuenta con las iniciales del estudiante. Así mismo, las evaluaciones correspondientes fueron codificadas con textos alfanuméricos según el detalle de la tabla 8.

Tabla 8. Codificación de las Evaluaciones

EVALUACIONES	CODIFICACIÓN	DESCRIPCIÓN
1. Cuestionario 1	AE1	Actividad Evaluada 1
2. Ejercicios de Límites	AE2	Actividad Evaluada 2
3. Parcial 1	AE3	Actividad Evaluada 3
4. Ejercicios de Integración	AE4	Actividad Evaluada 4
5. Ejercicios de Optimización	AE5	Actividad Evaluada 5
6. Parcial 2	AE6	Actividad Evaluada 6

Fuente. Elaborado por la autora

La información referente a las acciones a recomendar ha sido codificada con textos alfanuméricos para hacer una referencia corta hacia ellos, tal y como se visualiza en la tabla 9.



Tabla 9. Codificación de las acciones a recomendar

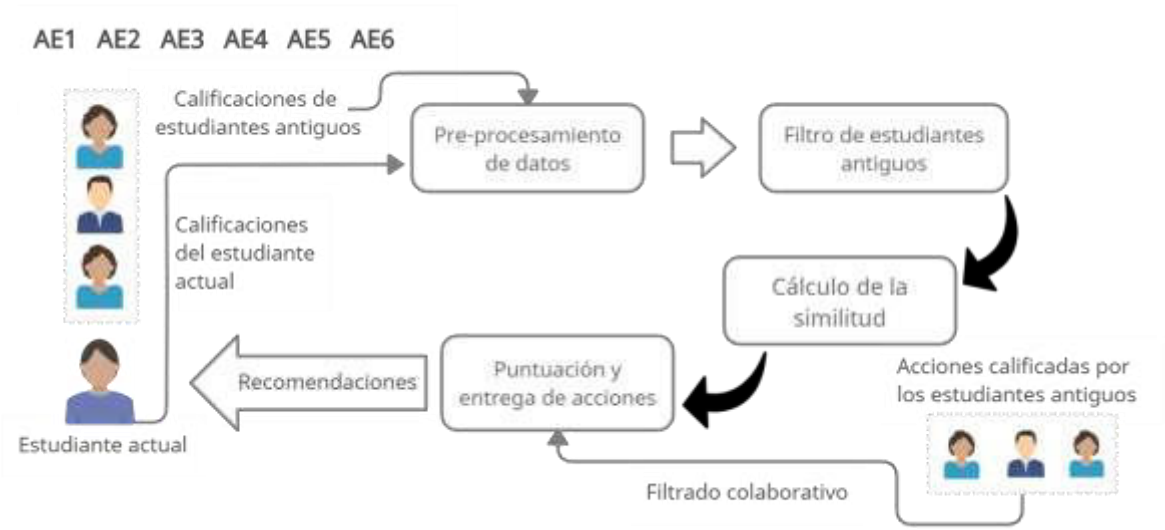
ÍTEMS	RECOMENDACIONES	CODIFICACIÓN	DESCRIPCION
Ítem - Acción 1	Cap. 6 - Aplicaciones de la Integral…	AC1	Acción a recomendar N° 1
Ítem - Acción 2	Canal Julioprofenet	AC2	Acción a recomendar N° 2
Ítem - Acción 3	Canal Academática	AC3	Acción a recomendar N° 3
Ítem - Acción 4	Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S.	AC4	Acción a recomendar N° 4
Ítem - Acción 5	Integrales Básicas – Canal Profe Alex	AC5	Acción a recomendar N° 5
Ítem – Acción 6	Otras herramientas y bibliografías de apoyo	AC6	Acción a recomendar N° 6

Fuente. Elaborado por la autora

3.7.4 Modelado

En esta etapa de la metodología se aplica la técnica apropiada para lograr los objetivos planteados y generar el modelo de recomendación. Para construir el modelo se hará uso del framework de Rstudio y se desarrolló en etapas que se plasman en la figura 13.

Figura 13. Etapas de construcción del modelo de recomendación



Fuente. Elaborado por la autora

3.7.4.1 Seleccionar técnica de modelado

Existen distintas técnicas de modelos de recomendación, dentro de las cuales están los modelos de recomendación basados en filtrado colaborativo, en conocimiento, basados en contenido o híbridos. Cada uno de ellos se aplican a un determinado contexto de estudio, tal y como se describió en el marco teórico. La técnica de filtrado colaborativo es ampliamente utilizada debido a su simplicidad y buenos resultados, más aún, cuando el arranque es en frio. (Gibaja et al., 2016). Filtrado colaborativo basado en usuarios es la técnica desarrollada en esta investigación ya que se hizo un análisis de similitud entre estudiantes y de acuerdo a la puntuación dada a las acciones se le recomienda las más apropiadas.

3.7.4.2 Generar plan de modelado

Se aplicó un procedimiento para probar la calidad y validez del modelo de recomendación, el cual consiste en aplicar el error medio absoluto, precisión y recall como métricas de evaluación. El desarrollo de estas métricas está detallado en el punto 3.7.5.

3.7.4.3 Construir el modelo

La construcción del modelo empieza con la carga de datos y se procederá a ejecutar diversas actividades agrupadas con el siguiente procedimiento:

a. Realizar un pre-procesamiento de datos de entrada

Los datos de entrada para el modelo de recomendación son de dos tipos: Las calificaciones de los estudiantes antiguos del curso y las calificaciones del o los estudiantes actuales a quienes se les va a recomendar las acciones. En ambos casos las actividades evaluadas tienen que ser las mismas, motivo por el cual la aplicación del modelo se desarrolla en el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática que fue dictada por el mismo docente.

La *figura 14 muestra parte de los datos de la matriz de las calificaciones de los estudiantes antiguos del curso.*



Figura 14. Matriz de las calificaciones de los estudiantes antiguos

	ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
15	LVA	2020-1	20	14	12	16.0	14.0	15
16	MMJ	2020-1	20	15	8	17.0	15.0	12
17	PMN	2020-1	8	5	8	16.0	14.0	8
18	PZA	2020-1	16	15	8	15.0	14.0	7
19	QHY	2020-1	12	5	8	5.0	5.0	0
20	TFG	2020-1	12	13	8	16.0	15.0	9
21	VDP	2020-1	12	15	8	16.0	11.0	7
22	VMW	2020-1	16	5	8	14.0	5.0	5
23	VCA	2020-1	8	15	4	16.0	16.0	7
24	ZAC	2020-1	20	18	8	14.0	5.0	5
25	ALY	2020-2	13	13	15	13.5	14.6	12
26	BVJ	2020-2	13	16	15	13.0	15.3	7
27	BGB	2020-2	9	15	15	17.0	15.0	6
28	CRL	2020-2	17	18	18	10.5	17.4	16
29	CAW	2020-2	3	5	5	6.5	5.9	8
30	CFK	2020-2	8	5	5	8.0	7.5	3
31	CVA	2020-2	11	16	15	3.5	12.4	7

Fuente. Elaborado por la autora

Esta matriz es del tipo *m* x *n,* donde *m* es la cantidad de estudiantes antiguos y *n* la cantidad de Actividades Evaluadas en el curso (AE). En esta investigación son 57 estudiantes antiguos y 6 actividades evaluadas.

La *figura 15 muestra parte de los datos de la matriz de las calificaciones de los estudiantes actuales a quienes se les brindara las acciones a recomendar.*



Figura 15. Matriz de las calificaciones de los estudiantes actuales

	ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
1	ACH	2021-1	10	18	13	18	12	NA
2	AHJ	2021-1	10	18	12	17	10	NA
3	ACG	2021-1	14	18	8	18	11	NA
4	BHA	2021-1	11	14	12	17	7	NA
5	CAW	2021-1	8	0	5	0	0	NA
6	CLG	2021-1	6	13	10	0	0	NA
7	CQA	2021-1	9	14	13	18	7	NA
8	CLH	2021-1	11	17	14	17	10	NA
9	CVA	2021-1	11	17	14	16	7	NA
10	CCJ	2021-1	12	15	10	17	8	NA
11	CMF	2021-1	14	18	10	17	13	NA
12	CLJ	2021-1	12	13	14	14	12	NA
13	CBV	2021-1	0	0	0	0	0	NA
14	COE	2021-1	11	18	12	18	14	NA

Fuente. Elaborado por la autora

Como se puede observar en la figura 15, la última Actividad Evaluada (AE6) aún no tiene una nota, esto debido a que es esa evaluación a la que se le asignara las acciones a recomendar.

b. Realizar un filtro de candidatos antiguos con respecto al actual.

Esta fase corresponde a realizar el filtrado de los estudiantes que no aprobaron la Actividad Evaluada para la cual el modelo de recomendación ofrecerá las acciones a recomendar. En esta investigación la AE (Actividad Evaluada) en la cual se dará las acciones a recomendar será la AE6 (Parcial 2), por lo cual se removerá de la matriz de entrada a todos aquellos estudiantes antiguos que no aprobaron dicha actividad (nota mayor a 10) tal y como se muestra en la figura 16.



Figura 16. Filtrado de los estudiantes antiguos

	ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
1	AAC	2020-1	16	18	4	14.0	14.0	14
2	ATF	2020-1	20	13	16	12.0	10.0	10
3	AAE	2020-1	16	18	12	5.0	5.0	0
4	ACR	2020-1	16	5	8	5.0	5.0	10
5	CAW	2020-1	12	5	12	5.0	5.0	0
6	CFK	2020-1	20	5	16	5.0	5.0	0
7	CCJ	2020-1	20	5	4	5.0	18.0	6
8	CMW	2020-1	12	15	12	13.0	5.0	8
9	CAH	2020-1	16	14	8	17.0	14.0	9
10	GAP	2020-1	12	14	8	18.0	17.0	8
11	HCR	2020-1	8	5	12	5.0	5.0	0
12	LLE	2020-1	4	18	8	17.0	5.0	7
13	LVL	2020-1	20	17	4	15.0	5.0	10
14	LFR	2020-1	12	5	12	16.0	14.0	7
15	LVA	2020-1	20	14	12	16.0	14.0	15
16	MMJ	2020-1	20	15	8	17.0	15.0	12

Fuente. Elaborado por la autora

Inicialmente se contaba con 57 estudiantes antiguos, al aplicar el filtro nos quedamos con una matriz de 14 estudiantes antiguos que aprobaron la última actividad evaluada, tal y como se aprecia en la figura 14.

La siguiente etapa del cálculo de la similitud, se trabajó con los estudiantes que están detallados en la figura 14, ya que son ellos los que obtuvieron una nota aprobatoria en la última actividad evaluada (AE6).



Figura 17. Matriz de estudiantes que aprobaron la última AE

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
AAC	2020-1	16	18	4	14.0	14.0	14
LVA	2020-1	20	14	12	16.0	14.0	15
MMJ	2020-1	20	15	8	17.0	15.0	12
ALY	2020-2	13	13	15	13.5	14.6	12
CRL	2020-2	17	18	18	10.5	17.4	16
CTE	2020-2	16	16	16	15.5	17.1	11
CLJ	2020-2	17	15	13	11.0	15.0	11
KLG	2020-2	14	16	16	12.5	15.6	12
LVL	2020-2	12	5	15	7.5	10.9	13
LLD	2020-2	14	16	16	13.5	15.9	13
MAC	2020-2	12	16	16	7.0	13.8	13
MGE	2020-2	15	17	16	12.5	16.1	11
MVR	2020-2	13	15	15	9.5	14.1	12
RSA	2020-2	15	17	16	15.0	17.0	14

Fuente. Elaborado por la autora

c. Calcular la similitud de la curva de rendimiento entre los estudiantes.

Esta fase corresponde al cálculo de la similitud entre los estudiantes antiguos y el estudiante actual, de acuerdo a las coincidencias en su curva de rendimiento académico se obtiene la similitud. Su cálculo está dado por distintas medidas de similitud, en este estudio se evalúan dos, el coeficiente de Jaccard según la ecuación 10 y la correlación de Pearson en la ecuación 11.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}} \tag{11}$$

En esta etapa se aplicó el cálculo de la similitud para dos estudiantes actuales, con calificaciones totalmente distintas entre sí, con la finalidad de observar como es el comportamiento del cálculo de las puntuaciones y entrega de las acciones a recomendarles para su última evaluación.

El primer estudiante actual seleccionado, a quien se le brindará las acciones a recomendar, tiene las calificaciones que se muestra en la tabla 10.



Tabla 10. Calificaciones del estudiante actual CCE

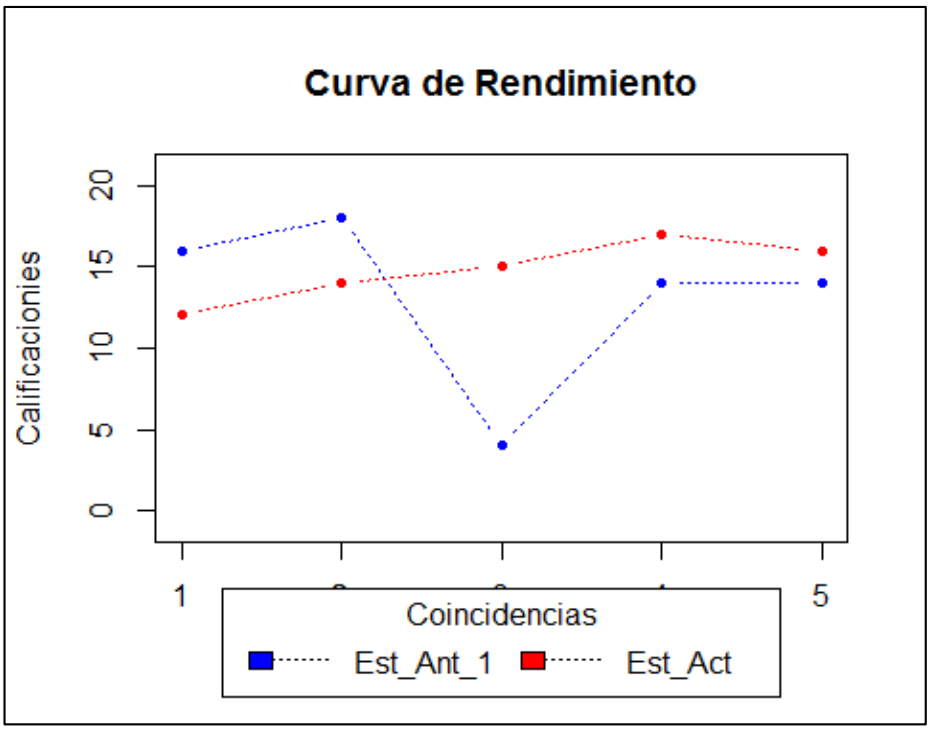
ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
CCE	2020-1	12,00	14,00	15,00	17,00	16,00	

Fuente. Elaborado por la autora

De acuerdo a la figura 17 que plasma a los estudiantes que aprobaron la última evaluación del curso y la tabla 10 con las calificaciones del estudiante actual en las cinco Actividades Evaluadas, se procede a evaluar la curva de rendimiento académico entre los estudiantes antiguos y el actual. Para ello se muestra en las figuras 18, 19 y 20 la comparación de sus curvas de rendimiento.

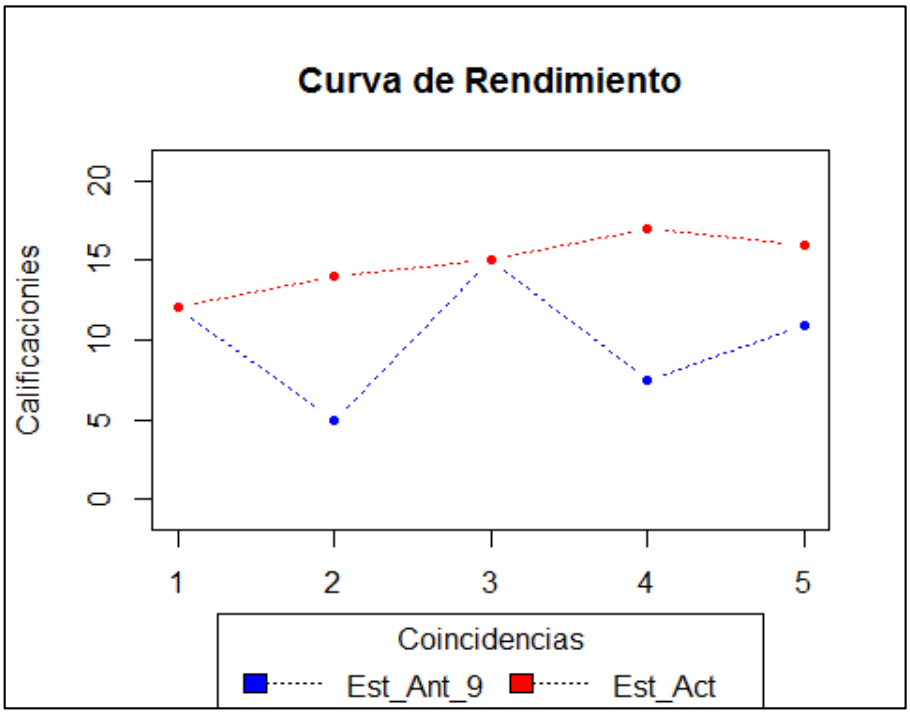
En estas figuras se puede apreciar las similitudes en cuanto a las notas obtenidas en cada AE, ya que a partir de ellas se hará el cálculo de la similitud.

Figura 18. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo AAC



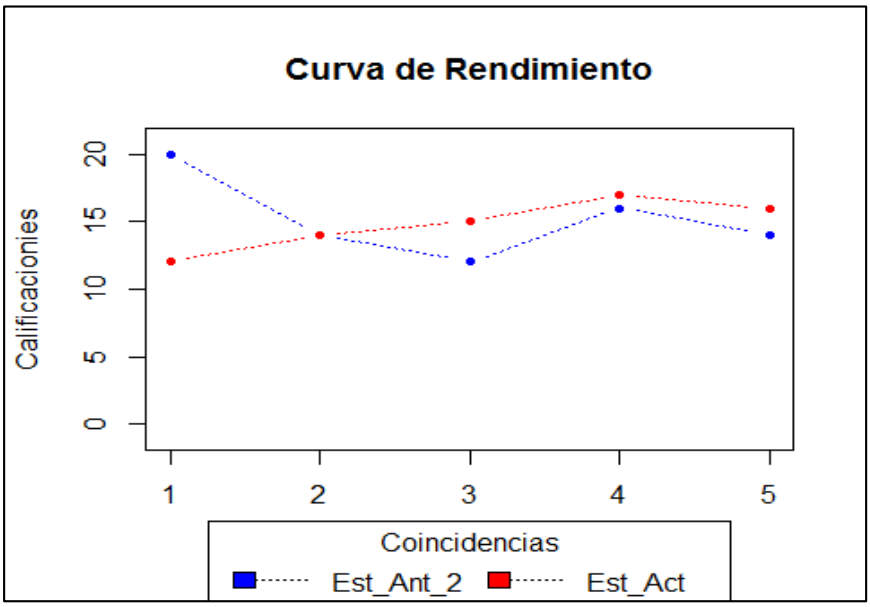
Fuente. Elaborado por la autora

Figura 19. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo LVL



Fuente. Elaborado por la autora

Figura 20. Curva de rendimiento del estudiante actual CCE y el estudiante antiguo LVA



Fuente. Elaborado por la autora

Se procede a realizar el cálculo de la similitud según Jaccard con la ecuación 10, la figura 21 muestra el coeficiente de Jaccard entre el estudiante actual y los estudiantes antiguos del curso.

Figura 21. Similitud según Jaccard con respecto al estudiante CCE

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	Similitud
AAC	2020-1	16	18	4	14.0	14.0	14	0.0
LVA	2020-1	20	14	12	16.0	14.0	15	0.2
MMJ	2020-1	20	15	8	17.0	15.0	12	0.2
ALY	2020-2	13	13	15	13.0	14.6	12	0.2
CRL	2020-2	17	18	18	10.5	17.4	16	0.0
CTE	2020-2	16	16	16	15.5	17.1	11	0.0
CLJ	2020-2	17	15	13	11.0	15.0	11	0.0
KLG	2020-2	14	16	16	12.5	15.6	12	0.0
LVL	2020-2	12	5	15	7.5	10.9	13	0.4
LLD	2020-2	14	16	16	13.5	15.9	13	0.0
MAC	2020-2	12	16	16	7.0	13.8	13	0.2
MGE	2020-2	15	17	16	12.5	16.1	11	0.0
MVR	2020-2	13	14	15	12.0	14.1	12	0.4
RSA	2020-2	15	17	16	15.0	17.0	14	0.0

Fuente. Elaborado por la autora

Como se puede observar, los estudiantes LVL y MVR tienen la similitud más alta respecto al rendimiento académico del estudiante actual CCE, sin embargo, se aprecia que la similitud es menor a 0.5 debido a que la curva de rendimiento del estudiante actual seleccionado tiene poca similitud con respecto a todos los estudiantes antiguos del conjunto de datos.

Figura 22. Similitud según Pearson con respecto al estudiante CCE



ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	Similitud
AAC	2020-1	16	18	4	14.0	14.0	14	-0.259759522
LVA	2020-1	20	14	12	16.0	14.0	15	-0.548472936
MMJ	2020-1	20	15	8	17.0	15.0	12	-0.323753540
ALY	2020-2	13	13	15	13.0	14.6	12	0.302741038
CRL	2020-2	17	18	18	10.5	17.4	16	-0.585043828
CTE	2020-2	16	16	16	15.5	17.1	11	0.048539700
CLJ	2020-2	17	15	13	11.0	15.0	11	-0.843527392
KLG	2020-2	14	16	16	12.5	15.6	12	-0.218181994
LVL	2020-2	12	5	15	7.5	10.9	13	-0.167035311
LLD	2020-2	14	16	16	13.5	15.9	13	-0.002117447
MAC	2020-2	12	16	16	7.0	13.8	13	-0.391717629
MGE	2020-2	15	17	16	12.5	16.1	11	-0.419619570
MVR	2020-2	13	14	15	12.0	14.1	12	-0.144688292
RSA	2020-2	15	17	16	15.0	17.0	14	0.129968811

Fuente. Elaborado por la autora

En la figura 22 se puede apreciar las similitudes de acuerdo a la correlación de Pearson, donde se observa que el estudiante ALY es el único con mayor similitud al estudiante actual.

El segundo estudiante actual seleccionado, a quien se le brindará las acciones a recomendar, tiene las calificaciones que se muestra en la tabla 11.

Tabla 11. Calificaciones del estudiante actual CMP

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
CMP	2020-1	13,00	14,00	15,00	12,00	9,00	

Fuente. Elaborado por la autora

Al calcular la similitud de este estudiante actual con respecto a los estudiantes antiguos se obtiene los resultados que se muestran en la Figura 23.

Figura 23. Similitud según Jaccard con respecto al estudiante CMP

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	Similitud
AAC	2020-1	16	18	4	14.0	14.0	14	0.0
LVA	2020-1	20	14	12	16.0	14.0	15	0.2
MMJ	2020-1	20	15	8	17.0	15.0	12	0.0
ALY	2020-2	13	13	15	13.0	14.6	12	0.4
CRL	2020-2	17	18	18	10.5	17.4	16	0.0
CTE	2020-2	16	16	16	15.5	17.1	11	0.0
CLJ	2020-2	17	15	13	11.0	15.0	11	0.0
KLG	2020-2	14	16	16	12.5	15.6	12	0.0
LVL	2020-2	12	5	15	7.5	10.9	13	0.2
LLD	2020-2	14	16	16	13.5	15.9	13	0.0
MAC	2020-2	12	16	16	7.0	13.8	13	0.0
MGE	2020-2	15	17	16	12.5	16.1	11	0.0
MVR	2020-2	13	14	15	12.0	14.1	12	0.8
RSA	2020-2	15	17	16	15.0	17.0	14	0.0

Fuente. Elaborado por la autora

En este segundo escenario, la similitud más alta es con el estudiante MVR, obteniendo un 0.8 y con el estudiante ALY un 0.4 de similaridad, siendo estos dos estudiantes los que influyen más en la predicción de la puntuación de las acciones a recomendar.

Los resultados obtenidos con la similaridad de Pearson para el mismo estudiante se muestran en la figura 24.



Figura 24. Similitud según Pearson con respecto al estudiante CMP

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	Similitud
AAC	2020-1	16	18	4	14.0	14.0	14	-0.35369044
LVA	2020-1	20	14	12	16.0	14.0	15	-0.12888746
MMJ	2020-1	20	15	8	17.0	15.0	12	-0.39346377
ALY	2020-2	13	13	15	13.0	14.6	12	-0.10466885
CRL	2020-2	17	18	18	10.5	17.4	16	0.21220734
CTE	2020-2	16	16	16	15.5	17.1	11	-0.67471198
CLJ	2020-2	17	15	13	11.0	15.0	11	-0.07619393
KLG	2020-2	14	16	16	12.5	15.6	12	0.19360327
LVL	2020-2	12	5	15	7.5	10.9	13	0.11287390
LLD	2020-2	14	16	16	13.5	15.9	13	0.09376728
MAC	2020-2	12	16	16	7.0	13.8	13	0.34126940
MGE	2020-2	15	17	16	12.5	16.1	11	0.17216110
MVR	2020-2	13	14	15	12.0	14.1	12	0.26822836
RSA	2020-2	15	17	16	15.0	17.0	14	-0.21718612

Fuente. Elaborado por la autora

En este caso, las similitudes de acuerdo a la correlación de Pearson más resaltantes son con los estudiantes MAC y MVR.

d. Puntuación de los ítems (acciones) a recomendar, para su posterior almacenamiento y entrega a los estudiantes actuales.

El objetivo de esta fase es puntuar automáticamente las acciones a recomendar, para ello se hace uso de las acciones a recomendar que se obtuvieron en un inicio ya calificadas por los estudiantes antiguos y parte de ella se muestra en la figura 25.



Figura 25. Matriz de puntuaciones de los estudiantes antiguos con respecto a cada una de las acciones a recomendar.

ID	CICLO	AC1	AC2	AC3	AC4	AC5	AC6
AAC	2020-1	5	1	1	5	NA	1
ATF	2020-1	4	1	NA	4	2	5
AAE	2020-1	3	3	1	5	1	5
ACR	2020-1	4	1	5	2	1	3
CAW	2020-1	5	5	1	3	1	NA
CFK	2020-1	1	3	4	4	3	1
CCJ	2020-1	3	2	5	1	2	NA
CMW	2020-1	3	NA	2	1	3	5
CAH	2020-1	5	2	2	3	2	5
GAP	2020-1	3	4	NA	1	2	5
HCR	2020-1	NA	NA	4	NA	4	1
LLE	2020-1	3	2	NA	1	NA	NA
LVL	2020-1	1	2	4	4	NA	3

Fuente. Elaborado por la autora

Para predecir la puntuación de las acciones a recomendar que se entregará al estudiante actual, se aplicó la media ponderada, la cual trabaja con la similitud de Jaccard calculada y las puntuaciones almacenadas de los estudiantes antiguos. La ecuación de la media ponderada es la siguiente:

$$AC_n = \frac{\sum_j^k (Cj_j * R_j)}{\sum_j^k Cj_j} \quad (12)$$

La ecuación 12 da la puntuación a la Acción a recomendar AC_n, donde k es el número de estudiantes que puntuaron la AC_n, Cj_j es el coeficiente de Jaccard entre el estudiante j y el estudiante actual, y R_j es la puntuación que el estudiante j le dio a la AC_n.

La puntuación obtenida para cada AC que se recomendará al estudiante actual CCE se muestra en la tabla 12.

Tabla 12. Puntuaciones generadas para el estudiante CCE

Similaridad	AC1	AC2	AC3	AC4	AC5	AC6
Jaccard	2.5	1.833333	4.375	3.875	2.428571	1.6

Pearson	3.190107	1.984460	3.403923	3.310166	3.321990	2.921148

Fuente. Elaborado por la autora

En el segundo escenario, para el estudiante CMP, las puntuaciones obtenidas de acuerdo a las dos medidas de similitud se muestran en la tabla 13.

Tabla 13. Puntuaciones generadas para el estudiante CMP

Similaridad	AC1	AC2	AC3	AC4	AC5	AC6
Jaccard	2.5	2.285714	4.5	4.375	2.428571	2
Pearson	3.1015848	0.7395853	3.6989778	2.9825607	2.9019883	1.3569227

Fuente. Elaborado por la autora

3.7.5 Evaluación

Para comprobar la validez de las recomendaciones, se llevó a cabo un proceso de validación donde los dos primeros ciclos académicos referenciados en la muestra (2020-1 y 2020-2) de la investigación, son utilizados para llevar a cabo el entrenamiento del modelo de recomendación y en el ciclo restante, correspondiente al período 2021-1, es donde se aplicó, determinó y analizó los resultados de las recomendaciones generadas.

Las métricas de evaluación utilizadas serán Error Medio Absoluto (MAE), Precisión y Recall, éstos parámetros de evaluación también fueron considerados por los autores Gibaja, et al. (2016) para evaluar el error, precisión y la relevancia de los ítems recomendados.

3.7.5.1 Error Medio Absoluto (MAE)

El Error Medio Absoluto es el error que el modelo de recomendación comete al realizar las predicciones de la puntuación de las acciones a recomendar, por lo tanto, cuando menor sea el valor del MAE, mejor será su eficacia. Para ello, se calcula con la ecuación 13.

$$MAE = \frac{\sum_{i=1}^{N} |p_i - v_i|}{N} \quad (13)$$



Donde p_i es la predicción, v_i es el valor de la puntuacion real y N es el número de las predicciones que se calcularon. La tabla 14 muestra la comparación y el valor del MAE de las puntuaciones predecidas por el modelo de recomendación para el estudiante actual *CCE y CMP.*

Tabla 14. Error Medio Absoluto para la similitud de Jaccard

Estudiante	AC	Puntuación Estimada	Puntuación Real	MAE
Estudiante CCE	AC1	2.5	4	0.866
	AC2	1.833333	3	
	AC3	4.375	4	
	AC4	3.875	5	
	AC5	2.428571	2	
	AC6	1.6	1	
Estudiante CMP	AC1	2.5	3	0.872
	AC2	2.285714	2	
	AC3	4.5	4	
	AC4	4.375	4	
	AC5	2.428571	5	
	AC6	2	3	

Fuente. Elaborado por la autora

En la tabla 15 se muestra las diferencias de las puntuaciones con la similitud de la correlación de Pearson y el Error Medio Absoluto obtenido.

Tabla 15. Eror Medio Absoluto para la similitud de Pearson

Estudiante	AC	Puntuación Estimada	Puntuación Real	MAE
Estudiante CCE	AC1	3.190107	4	1.22
	AC2	1.984460	3	
	AC3	3.403923	4	
	AC4	3.310166	5	
	AC5	3.321990	2	



	AC6	2.921148	1	
Estudiante CMP	AC1	3.1015848	3	1,12
	AC2	0.7395853	2	
	AC3	3.6989778	4	
	AC4	2.9825607	4	
	AC5	2.9019883	5	
	AC6	1.3569227	3	

Fuente. Elaborado por la autora

En esta métrica cada elemento del conjunto de prueba es tratado de la misma manera, lo que significa que se le da el mismo peso al error cometido al calcular la predicción, con la similitud según Jaccard en ambos estudiantes se obtiene un valor similar aproximado de 0.87, que es el valor alejado de la verdadera puntuación.

Sin embargo, en la tabla 15 se observa que el MAE calculado para ambos estudiantes de acuerdo a la similitud de correlación de Pearson es mayor.

3.7.5.2 Precisión y Recall

Para realizar las pruebas de *precisión y recall* se definió un criterio de relevancia, el cual indica si es relevante o no la acción a recomendar. Como los estudiantes califican las acciones en un rango de 1 a 5, para controlar los aciertos y los fallos, se considera un umbral que establece como relevante, aquella acción que tenga una puntuación mayor que las valoraciones medias de dicho estudiante. (Gibaja, et al., 2016).

Para el estudiante *CCE* son relevantes las acciones cuya puntuación sea mayor a 3.16, las cuales están plasmadas en la tabla 16 y para el estudiante *CMP* puntuaciones mayores a 3.5 son relevantes, las mismas que se visualizan en la tabla 17.

Tabla 16. Matriz de confusión, estudiante CCE

		Realidad		
Predicción		**Relevantes**	**No Relevantes**	**Total**
	Relevantes	$Tp = 2$	$Fp = 0$	2
	No Relevantes	$Fn = 1$	$Tn = 3$	4

Total	3	3	$N = 6$

Fuente. Elaborado por la autora

Tabla 17. Matriz de confusión estudiante CMP

		Realidad		
Predicción		**Relevantes**	**No Relevantes**	**Total**
	Relevantes	$Tp = 2$	$Fp = 0$	2
	No Relevantes	$Fn = 1$	$Tn = 3$	4
	Total	3	3	$N = 6$

Fuente. Elaborado por la autora

La precisión es la proporción de ítems relevantes seleccionados con respecto al total de ítems seleccionados, se calcula con la ecuación 14. Esta métrica indica el porcentaje de valores que se han clasificado como relevantes son realmente relevantes.

$$Precisión = \frac{Tp}{Tp+Fp} \quad (14)$$

El recall o exhaustividad se define como la proporción de ítems relevantes seleccionados con respecto al total de ítems relevantes existentes. se obtiene mediante la ecuación 15. Esta métrica mide cuantos valores relevantes son correctamente predichos.

$$Recall = \frac{Tp}{Tp+Fn} \quad (15)$$

A continuación, la tabla 18 muestra los valores obtenidos para cada métrica en cada estudiante evaluado.

Tabla 18. Precisión y Recall obtenidos

Método	**Estudiante**	**Precisión**	**Recall**
Jaccard	Estudiante CCE	1	0.66
	Estudiante CMP	1	0.66
Pearson	Estudiante CCE	0.75	1
	Estudiante CMP	1	0.33



Fuente. Elaborado por la autora

Con el método de Jaccard se puede observar que para ambos estudiantes se obtiene una precisión de 1 lo que indica que de lo que se le recomendó al estudiante, el 100% realmente era de su interés, y un recall de 0.66, lo que indica que de todo lo que realmente le interesa al estudiante, se le recomendó el 66%, mientras que con el método de similaridad de Pearson se obtiene que la precisión y el recall son opuestos entre estudiantes, e incluso se llega a recomendar solo un 33% del interés del estudiante.

Los resultados obtenidos con Jaccard resultan prometedores, considerando que el número de estudiantes empleado en el estudio es muy reducido. Con estas premisas, se considera que los resultados son muy positivos, y que sería interesante contar con más información, para poder hacer un estudio más preciso, y conseguir mejorar los resultados.

3.7.6 Implementación

Esta fase que corresponde a la Metodología CRISP-DM no forma parte de esta investigación ya que dentro de los objetivos plasmados esta la creación del modelo de recomendación, sin embargo, queda como trabajo a futuro la implementación del modelo ya que al término de la investigación ya se contará con las matrices bases para este tipo de estudio.



IV. RESULTADOS

Los resultados se presentan con respecto a las hipótesis planteadas para la presente investigación.

Hipótesis general: *El modelo de recomendación basado en competencias genera acciones para el curso Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.*

Con respecto a esta hipótesis general se crea el modelo de recomendación que provee las recomendaciones para una competencia especifica *(AE6)* a los estudiantes actuales. Estas acciones se entregan en orden de relevancia según la puntuación y están plasmadas en la tabla 19.

Tabla 19. Acciones a recomendar generadas por el modelo

	Descripción	Puntuación
AC3	Canal Academática	4.5
AC4	Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S.	4.375
AC1	Cap. 6 - Aplicaciones de la Integral	2.5
AC5	Integrales Básicas – Canal Profe Alex	2.428571
AC2	Canal Julioprofenet	2.285714
AC6	Otras herramientas y bibliografías de apoyo	2

Fuente. Elaborado por la autora

Hipótesis específica 1: *Los estudiantes antiguos presentan un nivel de competencia esperado mayor a 10 que corresponden a una calificación aprobada.*

La AE6 es la competencia en la que se entregarán las recomendaciones, por lo que se realizó un filtro de los estudiantes antiguos que no hayan obtenido el nivel mayor a 10, quedando sólo aquellos que si obtuvieron el nivel esperado.

En la figura 26 se observa los 14 estudiantes que obtuvieron un nivel mayor a 10 en la competencia en evaluación. Estos estudiantes son los que presentan el nivel esperado para realizar las similitudes con el estudiante actual.



Figura 26. Matriz de estudiantes con un nivel mayor a 10 de AE6

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6
AAC	2020-1	16	18	4	14.0	14.0	14
LVA	2020-1	20	14	12	16.0	14.0	15
MMJ	2020-1	20	15	8	17.0	15.0	12
ALY	2020-2	13	13	15	13.5	14.6	12
CRL	2020-2	17	18	18	10.5	17.4	16
CTE	2020-2	16	16	16	15.5	17.1	11
CLJ	2020-2	17	15	13	11.0	15.0	11
KLG	2020-2	14	16	16	12.5	15.6	12
LVL	2020-2	12	5	15	7.5	10.9	13
LLD	2020-2	14	16	16	13.5	15.9	13
MAC	2020-2	12	16	16	7.0	13.8	13
MGE	2020-2	15	17	16	12.5	16.1	11
MVR	2020-2	13	15	15	9.5	14.1	12
RSA	2020-2	15	17	16	15.0	17.0	14

Fuente. Elaborado por la autora

Hipótesis específica 2: *Las coincidencias de las calificaciones de los estudiantes antiguos y actuales se determinan a través del coeficiente de Jaccard.*

Con respecto a esta hipótesis, el modelo presenta las coincidencias entre el estudiante actual y los estudiantes antiguos, buscando así cuantas coincidencias existen entre cada par de estudiantes y éstas dan la similitud, calculada con la ecuación del coeficiente de Jaccard. La figura 27 muestra la matriz de los estudiantes antiguos del curso a la cual se le añadió en la última columna el vector de similitud calculado, se filtró para que nos muestre de manera descendente pudiendo observar los estudiantes que más se asemejan al estudiante actual en cuanto a su rendimiento académico.



Figura 27. Matriz con la similitud según Jaccard

ID	CICLO	AE1	AE2	AE3	AE4	AE5	AE6	Similitud
MVR	2020-2	13	14	15	12.0	14.1	12	0.8
ALY	2020-2	13	13	15	13.0	14.6	12	0.4
LVA	2020-1	20	14	12	16.0	14.0	15	0.2
LVL	2020-2	12	5	15	7.5	10.9	13	0.2
AAC	2020-1	16	18	4	14.0	14.0	14	0.0
MMJ	2020-1	20	15	8	17.0	15.0	12	0.0
CRL	2020-2	17	18	18	10.5	17.4	16	0.0
CTE	2020-2	16	16	16	15.5	17.1	11	0.0
CLJ	2020-2	17	15	13	11.0	15.0	11	0.0
KLG	2020-2	14	16	16	12.5	15.6	12	0.0
LLD	2020-2	14	16	16	13.5	15.9	13	0.0
MAC	2020-2	12	16	16	7.0	13.8	13	0.0
MGE	2020-2	15	17	16	12.5	16.1	11	0.0
RSA	2020-2	15	17	16	15.0	17.0	14	0.0

Fuente. Elaborado por la autora

Hipótesis específica 3: *La puntuación de los ítems de recomendación se determina a través de la ecuación de puntuación para recomendar las mejores acciones al estudiante actual del curso de Matemática II.*

A través de la media ponderada y la similitud de Jaccard calculada, se obtuvo las predicciones de cada una de las actividades evaluadas, de manera que se almacenan en una matriz de puntuación y se entrega en orden de relevancia al estudiante actual, tal y como se muestra en la tabla 12 y 13.

V. DISCUSIÓN DE RESULTADOS

1. Esta investigación construye un modelo de recomendación basado en la técnica de filtrado colaborativo, que provee acciones a recomendar al estudiante actual, de manera similar al estudio presentado por Chavarriaga et al. (2014), donde a partir de estudiantes vinculados al curso en el pasado y las calificaciones de los estudiantes actuales calculan la similitud y obtienen la puntuación de las acciones a recomendar, con la finalidad de que los estudiantes puedan alcanzar el nivel de competencia



esperado. En ambos estudios, se establece la importancia de la curva de aprendizaje o rendimiento, para el diseño del modelo de recomendación.

2. En la propuesta de Gibaja et al. (2016), un estudio similar a la presente propuesta, donde su inicio es un arranque en frio, con una participación de tan solo 28 alumnos, donde realizan la comparativa entre cuatro medidas de similitud, obtienen como mejor propuesta el uso de la correlación de Pearson, con una MAE de 0.9, una precisión del 86.6 % y recall de 68%, dejando como segunda propuesta el coeficiente de Jaccard, sin embargo, el presente estudio, obtiene resultados opuestos, al comparar estos dos métodos, el mas apropiado y que se ajusta a las necesidades en el coeficiente de Jaccard, con el que se obtiene un 100% de precisión y un 66% de recall, recomendando lo que realmente al estudiante le interesa.

3. De acuerdo a Vaca (2017), quien afirma que todo modelo de recomendación debe tener un *repositorio de tareas* que son los ítems a recomendar; *el modelo del estudiante*, el cual almacena las calificaciones de alumnos antiguos; *el recomendador* que utiliza el alumno para consultar en el repositorio (el cual genera las recomendaciones) y un *ranking de tareas generadas* que se muestran al alumno, la presente investigación se centra en la importancia de cuatro etapas de un modelo de recomendación, las cuales son similares a las del autor antes mencionado, *el pre-procesamiento de datos y el filtrado de los estudiantes*, ambas corresponden al modelo del estudiante; *el cálculo de la similitud*, el cual corresponde al recomendador; *por último el cálculo de las puntuaciones y entrega de las recomendaciones*, corresponden al ranking de las acciones a recomendar. Con lo explicado se puede afirmar que este estudio coincide con la afirmación el autor en mención.

4. Mendoza, et al. (2019), quien desarrolla un modelo de recomendación con filtrado colaborativo basado en ítems, afirma que la validez de las recomendaciones obtenidas por ese tipo de filtrado colaborativo depende si las características de los ítems a recomendar seran tomadas en cuenta o no, de no interesar esas características y solo se quiere recomendar ítems similares en calificación, ese enfoque es el más adecuado. Se confirma lo afirmado por el autor, ya que la presente investigación se centra en un filtrado colaborativo que recomienda acciones similares en cuanto a calificación, y no se busca evaluar las características de las acciones a recomendar,



puesto que se debería optar por un enfoque basado en contenido. Sin embargo, se puede complementar que, la presenta investigación se desarrolló en un enfoque basado en usuarios, llevándose a cabo una previa evaluación de similitud de usuarios.

5. Sánchez, et al. (2019) en su artículo "Persuade Me - A User-based Recommendation System Approach" proponen un sistema de recomendación basado en perfiles que genera un conjunto de posibles estrategias persuasivas que pueden usarse con usuarios de características similares. Plantean mecanismos efectivos para aumentar el éxito de las recomendaciones de este tipo y a su vez proporcionan las estrategias que se deben seguir para aliviar el problema del arranque en frío, es decir, cuando recién se iniciará un estudio de este tipo y que puede ser extrapolado a otros ámbitos. Es esa estrategia la que se usó para dar inicio a la creación del modelo de recomendación de esta investigación, al no tener información sobre las acciones a recomendar y sus puntuaciones, se hace un estudio con los estudiantes antiguos para recopilar acciones que les fue de apoyo en el desarrollo del curso de matemática y posteriormente, mediante un formulario tipo encuesta se les pidió que le asignen una puntuación a cada una de ellas, coincidiendo con el autor antes mencionado que es una estrategia apropiada para los arranques en frio.

6. El coeficiente de Jaccard es una métrica apropiada ya que ignora el valor de las preferencias sobre un ítem a recomendar, tal y como lo menciona Owen, et al. (2012), es decir, no es necesario considerar el valor de la calificación que los usuarios hacen hacia los ítems a recomendar para el cálculo de la similitud de los usuarios. Con este estudio se afirma lo mencionado, ya que con el coeficiente de Jaccard se evalúa la similitud entre estudiantes, en cuanto a su rendimiento académico. En esta investigación se obtuvo que el estudiante actual presenta una similitud de 0.4285714 con respecto a otros tres estudiantes, siendo la similitud más alta. Posteriormente ésta similitud se hace uso para predecir las puntuaciones a las recomendaciones.

VI. CONCLUSIONES

1. Se logra la construcción del modelo de recomendación que determina y entrega acciones a recomendar para el estudiante actual del curso de Matemática II. Los procesos desarrollados y explicados en esta investigación dan paso a la obtención del modelo por medio de la técnica de filtrado colaborativo basado en usuarios, cuya



evaluación con la métrica del MAE se obtiene que la diferencia de puntuación de las acciones a recomendar es de 0.866 y 0.872 con respecto a las verdaderas puntuaciones de los estudiantes evaluados, así como un 100% de precisión y un 66% de recall con el coeficiente de Jaccard como medida de similitud.

2. Se realizó el pre-procesamiento de datos de entrada, a partir de la creación de las matrices que contienen las calificaciones de los estudiantes antiguos y de los estudiantes actuales, las cuales están orientados a una competencia específica del curso de Matemática, en este caso de la última actividad evaluada (AE6).

3. A través de la curva del rendimiento académico de los estudiantes se logró evaluar y filtrar aquellos estudiantes antiguos que no alcanzan el nivel de competencia esperado en la última evaluación, por lo que de los 57 estudiantes antiguos solo se trabajaron con 14 estudiantes, que a partir de ellos se hizo las comparaciones de su similitud con los estudiantes actuales.

4. En el cálculo de la similitud entre el estudiante actual y los estudiantes antiguos, se obtiene la relación de la intersección y la unión de las coincidencias de las calificaciones obtenidas por ambos estudiantes, ello con el cálculo de la similitud según el coeficiente de Jaccard, obteniendo los tres estudiantes más parecidos al actual con una similitud de 0.8 y 0.4 para cada uno de los estudiantes evaluados.

5. La similitud obtenida permitió calcular los ítems (acciones) de recomendación mediante la Media Ponderada, una técnica generadora de recomendaciones. Al tener las predicciones de las puntuaciones para cada AC, se las almacenó en una matriz para su posterior entrega al estudiante actual, ordenada de manera descendente según el valor obtenido, siendo la AC3 – Canal Academática la que obtuvo una puntuación de 4.375, AC4 - Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S. con una puntuación de 3.875 y AC1 - Cap. 6 - Aplicaciones de la Integral con 2.5. Éstas tres acciones, de las seis en evaluación, son las apropiadas para los estudiantes actuales.



VII. RECOMENDACIONES

1. Las matrices de datos de entrada tienen que estar bien estructuras ya que a partir de ellas nace el modelo de recomendación, para un buen resultado se recomienda que para los estudios de este tipo las forma de calificar cada actividad evaluada sea de la misma manera o por el mismo docente, ello permitirá que la similitud obtenida sea de calidad.

2. Se recomienda trabajar con la mayor cantidad de datos posibles para obtener mejores resultados, esto debido a que en esta investigación al ser un arranque en frío se trabajó con una muestra pequeña y es por ello que las similitudes calculadas es su mayoría son inferiores a 0.5. Queda como trabajo a futuro el desarrollo del modelo con una muestra mayor para la obtención de similitudes más altas.

3. Para obtener las recomendaciones optimas y adecuadas a cada estudiante se recomienda seguir almacenando los datos de entrada, es decir, agregar las calificaciones de los estudiantes que van culminado el curso en evaluación.

4. Aplicar otras técnicas medidas de similitud para una correcta evaluación y comparación de las acciones a recomendar por el modelo, de tal manera que se pueda perfeccionar la calidad de las recomendaciones obtenidas.

VIII. REFERENCIAS

Abad, C., Kam, A., & Calva, L. (2010). Sistema de Recomendación de Películas. Escuela Superior Politecnica del Litoral.

Abalde, R., & Armentano, M. (2014). Inducción de preferencias a partir del contexto de elección del usuario en filtrado colaborativo. 15th Argentine Symposium on Articial Intelligence, ASAI 2014. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/41726/Documento_completo.pdf?sequence=1&isAllowed=y

Adolphs, R. (2009). The social brain: neural basic of social knowledge. Annu. Rev. Psychol, 60, 693–716.

Alvarado, M., & García, C. (2016). Cálculo diferncial en competencias. Ciudad de México: Grupo Editorial Patria S.A.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction volume 12, pages331–370.

Cabero, J., López, Eloy (2009). Descripción de un instrumento didáctico para el Análisis de modelos y estrategias de enseñanza de Cursos universitarios en red (A.D.E.C.U.R). Revista de Medios y Educación

Casas, J., Repullo, J., & Donado, J. (2003). La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). Aten Primaria. Obtenido de https://pdf.sciencedirectassets.com/277730/1-s2.0-S0212656703X75836/1-s2.0-S0212656703707288/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEN7%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQD74tt2dDC9It3v0VKlGqW2%2FqxbTGoPnd8TbdL2NSOZtwIgU4poWC69Mk

Castillo, M., Gamboa, R., & Hidalgo, R. (2020). Factores que influyen en la deserción y reprobación de estudiantes de un curso universitario de matemáticas. UNICIENCIA, 34, 219-245. doi:http://dx.doi.org/10.15359/ru.34-1.13

Chavarriaga, O., Florian-Gaviria, B., & Solarte, O. (2014). Recommender System Based on Student Competencies Assessment Results. IEEE, 6.

D. Cova, W. (2016). Sistemas, Modelos Simulación y un toque de Epistemolgía. Argentina: Editorial de la Universidad Tecnológica Nacional - edUTecNe. Obtenido de http://www.edutecne.utn.edu.ar/monografias/sistemas_modelos_simulaci%F3n.pdf

Del Pino, J., Salazar, G., & Cedeño, V. (2011). Adaptación de un recomendador de filtro colaborativo basado en el usuario para la creación de un recomendador de materias de pregrado basado en el historial académico de los estudiantes. Revista Tecnológica ESPOL - RTE, 24, 29-34.

Dietmar Jannach, M. Z. (2010). Recommender Systems: An Introduction.

Dietmar, J., Zanker, M., Felferning, A., & Friedrich, G. (2011). Recommender Systems. New York: Cambridge.



Galán, S. (2007). Filtrado Colaborativo y Sistemas de Recomendación. Obtenido de http://www.it.uc3m.es/~jvillena/irc/practicas/06-07/31.pdf

Gibaja, Z., Gámez, J., Soto, J., Arauzo, A., Moyano, J., Ramírez, A., & Martínez, M. (2016). Aplicación móvil para la recomendación de asignaturas en los estudios de grado universitarios. REVISTA DE BUENAS PRÁCTICAS DOCENTES EN LA UCO.

Herlocker, J., Konstan, J., Terveen, L., & Riedl, J. (2004). Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems, Pages 5–53. doi:https://doi.org/10.1145/963770.963772

Hernández Sampiere, R., Fernández, C., & Baptista, P. (2010). Metodología de la Investigación. (I. Editores, Ed.) Quinta.

Lázaro, N., Callejas, Z., Griol, D., & Durán, M. (2017). La deserción estudiantil en educación superior: S.O.S. en carreras de ingeniería informática. Conferencia Latinoamericana sobre Abandono en la Educación Superior CLABES, 1-9.

Mendoza, G., Laureano, Y., & Pérez, M. (2019). Métricas de Similaridad y evaluación para sistemas de recomendación de filtrado colaborativo. Riti Journal.

Minsky, M. (1965). Matter, Mind and Models. Obtenido de https://groups.csail.mit.edu/medg/people/doyle/gallery/minsky/mmm.html

Mori, M. (2012). Deserción universitaria en estudiantes de una Universidad Privada de Iquitos. Revista digital de investigación en docencia universitaria. doi:DOI: https://doi.org/10.19083/ridu.6.42

Owen, S., Anil, R., Dunnig, T., & Friedman, E. (2012). Mahout in Action . Manning Publications Co. All rights reserved.

Owin, S., & Robin, A. (2009). Mahout in Action. Greenwich: Manning Publications PublicationsCo.

Ricci, F., Rokach, L., & Shapira, B. (2010). Recommender Systems Handbook.

Rieckmann, M. (2017). La educación basada en competencias en la educación superior: propuestas teóricas-conceptuales y prácticas. Seminario Internacional "Educación basada en Competencias". Obtenido de https://www.researchgate.net/publication/320065929_La_educacion_basada_en_compet



encias_en_la_educacion_superior_propuestas_teoricas-conceptuales_y_experiencias_practicas

Ruiz, F., Diaz, A., Chang, H. & Echevarría, F. (2006). Sistema de predicción y recomendación personalizada basada en ranking de ítems homogéneos usando filtrado colaborativo. Obtenido de https://www.dspace.espol.edu.ec/handle/123456789/3205.

Sánchez, F. (2019). Fundamentos Epistémicos de la investigación Cualitativa y Cuantitativa: Consensos y Disensos. Scielo. Obtenido de http://www.scielo.org.pe/pdf/ridu/v13n1/a08v13n1.pdf

Sánchez, R., Casado, D., Borges, C., & López, D. (2019). "Persuade Me" A user-based Recommendation System Approach. IEEE. Obtenido de https://ieeexplore.ieee.org/document/9060163

Torres, N. (2015). Memoria de Titulación Sistemas de Recomendación basados en métodos de filtrado colaborativo. UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA, Santiago de Chile.

Vaca Sánchez, J. M. (2017). Sistema de Recomendación de Tareas basado en competencias educativas registradas en un portafolio electrónico.

Vargas, M., & Montero, E. (2016). Factores que determinan el rendimiento académico en Matemáticas en la Universidad Nacional de Ingeniería (UNI), Nicaragua: un modelo de ecuaciones estructurales. Universitas Psychologica, 1-11. Obtenido de https://doi.org/10.11144/Javeriana.upsy15--4.fdra

Wan, X., Jamaliding, Q., & Okamoto, T. (2001). Analizing Learners Relationship to Improve the Quality of Recommender System for Group Learning Support. J. Comput 6, 254-262.

Yago, H. (2019). Una propuesta de sistema de recomendación basado en competencias y modelado del estudiante ontológico. Alcalá.

IX. ANEXOS



ANEXO N° 1

LINEAS DE CODIGO DEL MODELO DE RECOMENDACIÓN

Importando la data de Calificaciones de Matemática II

```
notas1 <- read.csv2("BD_MAT_2020_1.csv")
notas2 <- read.csv2("BD_MAT_2020_2.csv")
notas3 <- read.csv2("BD_MAT_2021_1.csv")
```

Importando la data de las puntuaciones de las acciones a recomendar

```
punt <- read.csv2("BD_PUNT_ACCIONES.csv")
```

Cadena para realizar la limpieza de la data

```
limp <- c("ITEM", "NOMBRES")
```

Preparacion de los datos

```
notas1 <- notas1[ , !(names(notas1) %in% limp)]
notas2 <- notas2[ , !(names(notas2) %in% limp)]
notas3 <- notas3[ , !(names(notas3) %in% limp)]

punt <- punt[ , !(names(punt) %in% limp)]

str(notas1)

## 'data.frame':    24 obs. of  8 variables:
##  $ ID   : chr  "AAC" "ATF" "AAE" "ACR" ...
##  $ CICLO: chr  "2020-1" "2020-1" "2020-1" "2020-1" ...
##  $ AE1  : num  16 20 16 16 12 20 20 12 16 12 ...
##  $ AE2  : num  18 13 18 5 5 5 5 15 14 14 ...
##  $ AE3  : num  4 16 12 8 12 16 4 12 8 8 ...
##  $ AE4  : num  14 12 5 5 5 5 5 13 17 18 ...
##  $ AE5  : num  14 10 5 5 5 5 18 5 14 17 ...
##  $ AE6  : num  14 10 0 10 0 0 6 8 9 8 ...

str(notas2)

## 'data.frame':    33 obs. of  8 variables:
##  $ ID   : chr  "ALY" "BVJ" "BGB" "CRL" ...
##  $ CICLO: chr  "2020-2" "2020-2" "2020-2" "2020-2" ...
##  $ AE1  : num  13 13 9 17 3 8 11 10 14 16 ...
##  $ AE2  : num  13 16 15 18 5 5 16 16 14 16 ...
##  $ AE3  : num  15 15 15 18 5 5 15 5 16 16 ...
##  $ AE4  : num  13 13 17 10.5 6.5 8 3.5 11 7 15.5 ...
##  $ AE5  : num  14.6 15.3 15 17.4 5.9 7.5 12.4 11.5 13.8 17.1 ...
##  $ AE6  : num  12 7 6 16 8 3 7 6 7 11 ...
```



```
str(notas3)
```

```
## 'data.frame':    24 obs. of  8 variables:
##  $ ID   : chr  "ACH" "AHJ" "ACG" "BHA" ...
##  $ CICLO: chr  "2021-1" "2021-1" "2021-1" "2021-1" ...
##  $ AE1  : num  10 10 14 11 8 6 9 11 11 12 ...
##  $ AE2  : num  18 18 18 14 0 13 14 17 17 15 ...
##  $ AE3  : num  13 12 8 12 5 10 13 14 14 10 ...
##  $ AE4  : num  18 17 18 17 0 0 18 17 16 17 ...
##  $ AE5  : num  12 10 11 7 0 0 7 10 7 8 ...
##  $ AE6  : logi  NA NA NA NA NA NA ...
```

Integración de la data

```
notasEstAnt <- rbind(notas1,notas2)
notasEstAnt <- cbind(notasEstAnt, punt)
```

data_Aprobaron$Similitud<-vec_similitud

MODELADO FILTRADO DE ESTUDIANTES ANTIGUOS

Estudiantes que aprobaron la última evaluación

```
data_Aprobaron<-notasEstAnt[notasEstAnt$AE6> 10,]
punt <-data_Aprobaron[,9:16]
data_Aprobaron <- data_Aprobaron[,1:8]
```

Gráficos de la similitud de la curva de rendimiento

```
AE <- c(1:5)
```

Calificaciones del “1” estudiante antiguo

```
E1 <- c()
for (i in 3:7){
  E1 <- c(E1,data_Aprobaron[1,i])
}
```

Calificaciones del “2” estudiante antiguo

```
E2 <- c()
for (i in 3:7){
  E2 <- c(E2,data_Aprobaron[2,i])
}
```

Calificaciones del “9” estudiante antiguo

```
E9 <- c()
for (i in 3:7){
  E9 <- c(E9,data_Aprobaron[9,i])
}
```

Calificaciones del estudiante actual CMP

```
EA <- c()
for (i in 3:7){
```



```
  EA <- c(EA,notas3[15,i])
}
```

Gráficos de las curvas del rendimiento entre el estudiante antiguo_2 y el actual

```
plot( c(1,5), c(-1,21), type = "n", xlab = "AE",
      ylab="Calificacionies", main = "Curva de Rendimiento" )

lines( AE,  E2, type = "b", lwd = 1.3, lty = 3, col = "blue", pch = 20
)

lines( AE,  EA, type = "b", lwd = 1.3, lty = 3, col = "red", pch = 20
)


legend(x = "bottom", inset = c(0, -0.5),lty = 3, lwd = 1.3, legend = c
("Est_Ant_2", "Est_Act"), fill = c("blue", "red"),
       xpd=TRUE, horiz=TRUE, title = "Coincidencias")
```

Gráficos de las curvas del rendimiento entre el estudiante antiguo_9 y el actual

```
plot( c(1,5), c(-1,21), type = "n", xlab = "AE",
      ylab="Calificacionies", main = "Curva de Rendimiento" )

lines( AE,  E9, type = "b", lwd = 1.3, lty = 3, col = "blue", pch = 20
)

lines( AE,  EA, type = "b", lwd = 1.3, lty = 3, col = "red", pch = 20
)


legend(x = "bottom", inset = c(0, -0.5),lty = 3, lwd = 1.3, legend = c
("Est_Ant_9", "Est_Act"), fill = c("blue", "red"),
       xpd=TRUE, horiz=TRUE, title = "Coincidencias")
```

CALCULO DE LA SIMILITUD SEGUN JACCARD DEFINIENDO FUNCION DE JACCARD
J(A,B)=|A∩B|/(AUB)

```
jaccard <- function(a, b) {
  cont=0
  for (i in 1:5){
    if (a[i]==b[i])
      {cont=cont+1}
  }
  result <- cont/5
  return (result)
}
```



VECTOR DE SIMILITUDES

```
vec_similitud<-c()
```

BUCLE PARA EL CALCULO DE LA SIMILITUD

```
for(i in 1:nrow(data_Aprobaron)){
  a<-c(data_Aprobaron[i,3:7])
  b<-EA
  similitud<-jaccard(a,b)
  vec_similitud<-c(vec_similitud,similitud)
}

#Agregando la columna de Similaridad

vec_similitud

##  [1] 0.0 0.2 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.8 0.0

data_Aprobaron$Similitud<-vec_similitud
```

SIMILUTUDES DE LOS ESTUDIANTES

```
data_Aprobaron

##     ID  CICLO AE1 AE2 AE3  AE4  AE5 AE6 Similitud
## 1  AAC 2020-1  16  18   4 14.0 14.0  14       0.0
## 15 LVA 2020-1  20  14  12 16.0 14.0  15       0.2
## 16 MMJ 2020-1  20  15   8 17.0 15.0  12       0.0
## 25 ALY 2020-2  13  13  15 13.0 14.6  12       0.4
## 28 CRL 2020-2  17  18  18 10.5 17.4  16       0.0
## 34 CTE 2020-2  16  16  16 15.5 17.1  11       0.0
## 36 CLJ 2020-2  17  15  13 11.0 15.0  11       0.0
## 40 KLG 2020-2  14  16  16 12.5 15.6  12       0.0
## 42 LVL 2020-2  12   5  15  7.5 10.9  13       0.2
## 43 LLD 2020-2  14  16  16 13.5 15.9  13       0.0
## 44 MAC 2020-2  12  16  16  7.0 13.8  13       0.0
## 45 MGE 2020-2  15  17  16 12.5 16.1  11       0.0
## 47 MVR 2020-2  13  14  15 12.0 14.1  12       0.8
## 52 RSA 2020-2  15  17  16 15.0 17.0  14       0.0
```

PUNTUACION Y ENTREGA DE ACCIONES A RECOMENDAR Media ponderada I_n=(∑j^k (CJj*Rj))/(∑j^k CJj)

```
vec_puntuacion<-c()

for (i in 3: ncol(punt)){
  sum<-0
  sum_similitud<-0
  for(j in 1: nrow(punt)){
    if(!is.na(punt[j,i]))
    {
      sum<-sum+(data_Aprobaron[j,9]*punt[j,i])
      sum_similitud<-sum_similitud+data_Aprobaron[j,9]
    }
```



```
  }
  puntuacion<-sum/sum_similitud
  vec_puntuacion<-c(vec_puntuacion,puntuacion)
}

vec_puntuacion
```

```
## [1] 2.500000 2.285714 4.500000 4.375000 2.428571 2.000000
```

EVALUACIÓN DEL MODELO DE RECOMENDACIÓN

Importamos la BD de la Puntuación de los estudiantes actual

```
punt_Actual <- read.csv2("BD_PUNT_ACTUAL.csv")
```

Cálculo del Valor del Error Medio Absoluto

```
sumatoria <-0
for(i in 1:6){
  dif<-abs(vec_puntuacion[i]-punt_Actual[15,i+2])
  sumatoria<-sumatoria+dif
}
MAE<-sumatoria/6
MAE
```

```
## [1] 0.8720238
```

APLICANDO MÉTRICAS DE SIMILARIDAD PARA EL MODELO DE JACCARD

Para la clasificación como Relevante las puntuaciones se evalúa de acuerdo al promedio del estudiante, Estudiante CMP valor promedio 3.5

```
punt_real <- c()
for(i in 1:6){
  if(punt_Actual[15,i+2]>3.5){
    punt_real<-c(punt_real,'Relevante')
  }
  else{
    punt_real<-c(punt_real,'No Relevante')
  }
}
punt_real
```

```
## [1] "No Relevante" "No Relevante" "Relevante"    "Relevante"    "Re
levante"
## [6] "No Relevante"
```

```
punt_prediccion <-c()
for(i in 1:6){
  if(vec_puntuacion[i]>3.5){
    punt_prediccion<-c(punt_prediccion,'Relevante')
  }
  else{
    punt_prediccion<-c(punt_prediccion,'No Relevante')
  }
```



```
}
punt_prediccion
```

```
## [1] "No Relevante" "No Relevante" "Relevante"    "Relevante"    "No
Relevante"
## [6] "No Relevante"
```

MATRIZ DE CONFUSION

```
table<- table(punt_prediccion, punt_real, dnn = c("prediccion", "real"
))

table
```

```
##                      Real
## Predicción     No Relevante Relevante
##   No Relevante            3         1
##   Relevante               0         2
```

Precision Tp/(Tp+Fp) Recall Tp/(Tp+Fn)

```
precision<-table[2,2]/((table[2,2]+table[2,1]))
precision
```

```
## [1] 1
```

```
recall<-table[2,2]/(table[2,2]+table[1,2])
recall
```

```
## [1] 0.6666667
```

Para la clasificación como Relevante las puntuaciones se evalúa de acuerdo al promedio del estudiante, Estudiante CCE valor promedio 3.16

```
punt_real <- c()
for(i in 1:6){
  if(punt_Actual[18,i+2]>3.16){
    punt_real<-c(punt_real,'Relevante')
  }
  else{
    punt_real<-c(punt_real,'No Relevante')
  }
}
punt_real
```

```
## [1] "No Relevante" "No Relevante" "Relevante"    "Relevante"    "Re
levante"
## [6] "No Relevante"
```

```
punt_prediccion <-c()
for(i in 1:6){
  if(vec_puntuacion[i]>3.16){
```



```
    punt_prediccion<-c(punt_prediccion,'Relevante')
  }
  else{
    punt_prediccion<-c(punt_prediccion,'No Relevante')
  }
}
punt_prediccion
```

```
## [1] "No Relevante" "No Relevante" "Relevante"    "Relevante"    "No
Relevante"
## [6] "No Relevante"
```

MATRIZ DE CONFUSION

```
table<- table(punt_prediccion, punt_real, dnn = c("prediccion", "real"
))

table
```

```
##                     Real
## Predicción     No Relevante Relevante
##   No Relevante            3         1
##   Relevante               0         2
```

Precision Tp/(Tp+Fp) Recall Tp/(Tp+Fn)

```
precision<-table[2,2]/((table[2,2]+table[2,1]))
precision
```

```
## [1] 1
```

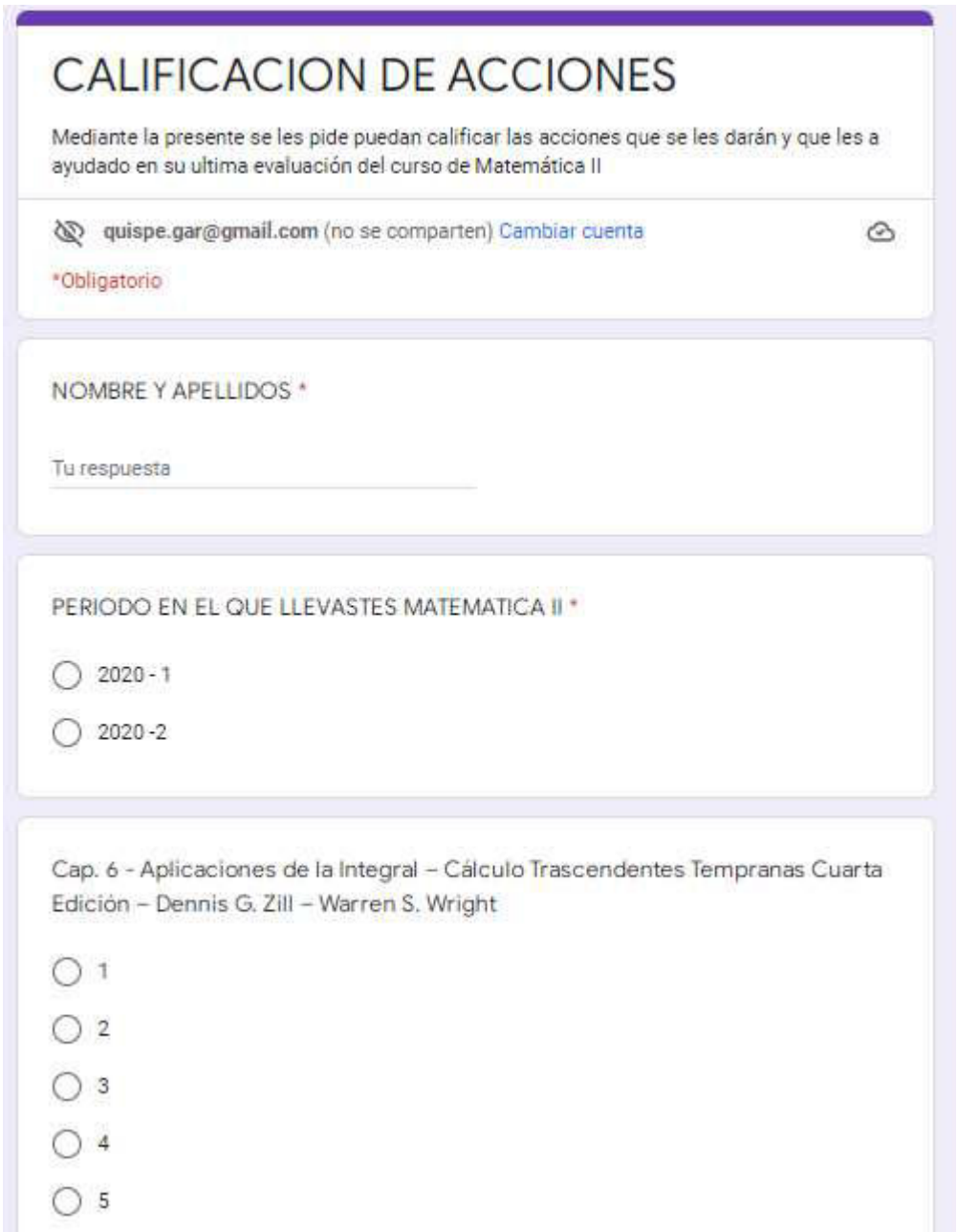
```
recall<-table[2,2]/(table[2,2]+table[1,2])
recall
```

```
## [1] 0.6666667
```



ANEXO N° 2

FORMULARIO/CUESTIONARIO PARA ASIGNACION DE PUNTUACION DE ACCIONES A RECOMENDAR



CALIFICACION DE ACCIONES

Mediante la presente se les pide puedan calificar las acciones que se les darán y que les a ayudado en su ultima evaluación del curso de Matemática II

quispe.gar@gmail.com (no se comparten) Cambiar cuenta

*Obligatorio

NOMBRE Y APELLIDOS *

Tu respuesta

PERIODO EN EL QUE LLEVASTES MATEMATICA II *

- 2020 - 1
- 2020 -2

Cap. 6 - Aplicaciones de la Integral – Cálculo Trascendentes Tempranas Cuarta Edición – Dennis G. Zill – Warren S. Wright

- 1
- 2
- 3
- 4
- 5

Canal Julioprofenet *

- 1
- 2
- 3
- 4
- 5

Canal Academatica

- 1
- 2
- 3
- 4
- 5

Cálculo Diferencial e Integral con aplicaciones – Prof. Elsie Hernández S.

- 1
- 2
- 3
- 4
- 5

ANEXO N° 3

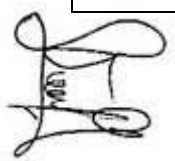
FORMULARIO/CUESTIONARIO PARA ASIGNACION DE PUNTUACION DE ACCIONES A RECOMENDAR



MATRIZ DE CONSISTENCIA

MODELO DE RECOMENDACIÓN DE ACCIONES BASADO EN COMPETENCIAS PARA EL CURSO DE MATEMÁTICA II DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA DE LA UNIVERSIDAD NACIONAL DE MOQUEGUA, AÑO 2021

PROBLEMAS	OBJETIVOS	HIPOTESIS	VARIABLES
Problema General Un modelo de recomendación basado en competencias, ¿puede determinar acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua?	**Objetivo General** Proponer un modelo de recomendación basado en competencias que determine acciones para el curso de Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.	**Hipótesis General** El modelo de recomendación basado en competencias genera acciones para el curso Matemática II de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua.	**Variable Independiente: Modelo de recomendación basado en competencias** **Dimensión:** Similitud de la curva de rendimiento de los estudiantes **Indicador:** Coeficiente de Similitud de Tanimoto (Jaccard) **Variable Dependiente: Recomendaciones para el curso de Matemática II** **Dimensión 1:** Reconoce los diferentes métodos de integración **Indicador 1.1:** Analiza las propiedades de las teorías de ante derivadas e integrales. **Indicador 1.2:** Identifica los conceptos básicos de integrales. **Indicador 1.3:** Compila los estudios sobre los temas de métodos de integración. **Indicador 1.4:** Formula y demuestra los problemas de integración. **Dimensión 2:** Aplica el cálculo integral en el desarrollo y resolución de problemas óptimos de forma gráfica, analítica y numéricamente**.** **Indicador 2.1: A**naliza las propiedades de las integrales definidas. **Indicador 2.2:** Identifica los conceptos básicos de integrales definidas.
Problemas Específicos 1. ¿Cómo realizar el pre-procesamiento de dos tipos de datos de entrada orientados a una competencia del curso de Matemática II? 2. ¿Cómo realizar el filtro de estudiantes antiguos a través de su rendimiento académico para elegir el nivel de competencia esperado en el curso de Matemática II? 3. ¿Cómo calcular la similitud de la curva de rendimiento entre el estudiante actual y los estudiantes antiguos? 4. ¿Cómo obtener las recomendaciones a entregar al estudiante actual del curso de Matemática II?	**Objetivos Específicos** 1. Realizar el pre-procesamiento de datos a partir de dos tipos de datos de entrada (calificaciones pasadas y actuales de los estudiantes) orientado a una competencia del curso de Matemática II. 2. Realizar el filtro de estudiantes antiguos a través de su rendimiento académico para elegir el nivel de competencia esperado del curso de Matemática II. 3. Calcular la similitud de la curva de rendimiento entre el estudiante actual y los estudiantes antiguos usando el coeficiente de Jaccard. 4. Puntuar los ítems de recomendación y almacenarlos en un repositorio para la posterior entrega de las recomendaciones al estudiante actual del curso de Matemática II.	**Hipótesis Específicos** Con respecto al Objetivo específico 1, no se plantea hipótesis especifica porque corresponde a un objetivo descriptivo. 1. Los estudiantes antiguos presentan un nivel de competencia esperado mayor a 10 que corresponden a una calificación aprobada. 2. Las coincidencias de las calificaciones de los estudiantes antiguos y actuales se determinan a través del coeficiente de Jaccard. 3. La puntuación de los ítems de recomendación se determina a través de la ecuación de puntuación para recomendar las mejores acciones al estudiante actual del curso de Matemática II.	



			Indicador 2.3: Compila los estudios sobre los temas de integración. **Indicador 2.4:** Formula y demuestra los problemas de integración definida.
TIPO, DISEÑO Y MÉTODO	**POBLACIÓN Y MUESTRA**	**TÉCNICAS E INSTRUMENTOS**	
Tipo de Investigación: Investigación Aplicada **Diseño de Investigación:** No experimental de tipo transversal **Alcance de Investigación:** Descriptiva – Explicativa **Método de Investigación:** Hipotética Deductiva	**Población:** Registros de estudiantes matriculados en la Universidad Nacional de Moquegua en el periodo 2020-1, 2020-2 y 2021-1. **Muestra:** Registros de estudiantes matriculados en la Escuela Profesional de Ingeniería de Sistemas e Informática del curso de Matemática II periodo 2020-1, 2020-2 y 2021-1.	**Técnicas:** - Aplicación de cuestionarios. - Filtrado Colaborativo basado en usuarios. - Medición Basada en Coeficiente de Jaccard - Salida de un repositorio de recomendaciones (ítems). **Instrumentos:** - Encuestas. - Software Estadístico y gráfico R (framework RecommenderLab) - Python (framework Crab).	



FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL UNAM

Formulario de autorización para la publicación electrónica en la página web del Repositorio Institucional de la Universidad Nacional de Moquegua, de conformidad con la Ley N° 30035 Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto y Resolución de Comisión Organizadora N° 041-2022-UNAM que aprueba el Reglamento que regula el Repositorio Institucional Digital de la UNAM.

1. DATOS DEL AUTOR

Nombres y Apellidos	: VERONICA QUISPE GARCIA
DNI/Carnet de Extranjería/Pasaporte N°	: DNI 70323204
Correo electrónico	: quispe.gar@gmail.com
Teléfono/Celular	: 953910894

2. GRADO, TITULO O ESPECIALIZACIÓN (Marque con una X)

Bachiller	Título Profesional	Magister	Doctor	Segunda Especialidad
	X			

3. TIPO DE PUBLICACIÓN (Marque con una X)

a) Articulo ( )
b) Tesis de pregrado (x)
c) Tesis de Maestría ( )
d) Tesis de Doctorado ( )
e) Monografía ( )
f) Reporte de caso científico o hallazgos ( )
g) Libro ( )
h) Parte de un libro ( )
i) Revisión ( )
j) Conferencia ( )
k) Presentación ( )
l) Artículo preliminar ( )
m) Articulo antes de ser publicado ( )
n) Artículo de prensa ( )
o) Documentos internos ( )
p) Imagen ( )
q) Documentos técnicos ( )
r) Ponencias presentadas en eventos científicos (Jornada, Congreso, Simposio Conferencias, etc.) ( )
s) Informes de suficiencia profesional ( )
t) Patentes producidas por académicos/as ( )

4. TIPO DE TRABAJO DE INVESTIGACIÓN (Marque con una X)

a) Tesis (x)
b) Trabajo de Investigación ( )
c) Trabajo de suficiencia profesional ( )
d) Trabajo académico ( )

5. DATOS DE LA PUBLICACION

a) Título : **MODELO DE RECOMENDACION DE ACCIONES BASADO EN COMPETENCIAS PARA EL CURSO DE MATEMATICA II DE LA ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMATICA DE LA UNIVERSIDAD NACIONAL DE MOQUEGUA, AÑO 2021**.

b) Campo del conocimiento OCDE (Indicar el código de clasificación OCDE de la obra : https://catalogos.concytec.gob.pe/vocabulario/ocde_ford.html) : **2.02.04**

c) Fecha de publicación: **2022**

6. DATOS DEL ASESOR:

a) Nombres y Apellidos : VANEZA FLORES GUTIERREZ
b) DNI/Pasaporte/Cédula : 40060249
c) ORCID : 0000-0001-7534-5810

FIRMA : .....



DOCUMENTO DE IDENTIDAD : 70323204
NOMBRES Y APELLIDOS : VERONICA QUISPE GARCIA

LICENCIAS

Licencia Estándar

Bajo los siguientes términos, autorizo el depósito de mi trabajo de investigación en el Repositorio Institucional Digital de la UNAM.

Con la autorización de depósito da mi trabajo de investigación, otorgo a la Universidad Nacional de Moquegua, una Licencia No Exclusiva para reproducir. distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi trabajo de investigación (incluido el resumen), en formato físico o digital, en cualquier medio, conocido o por conocerse, a través de los diversos servicios provistos por la Universidad, creados o por crearse, tales como el Repositorio Institucional Digital de la UNAM, Colección de Tesis, entre otros, en el Perú y en el extranjero, por el tiempo y veces que considere necesarias, y libre de remuneraciones.

En virtud de dicha Licencia, la UNIVERSIDAD NACIONAL DE MOQUEGUA, podrá reproducir mi trabajo de investigación en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con los propósitos de seguridad, respaldo y preservación.

Declaro que el trabajo de investigación es una creación de mi autoría y exclusiva titularidad, o coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicho trabajo de investigación no infringe derechos de autor de terceras personas.

La UNIVERSIDAD NACIONAL DE MOQUEGUA consignara el nombre del/los autor/es del trabajo de investigación, y no le hará ninguna modificación más de la permitida en la presente licencia.

Autorizo el depósito (marque con una X).

(X) **SI autorizo que se deposite inmediatamente.**

FIRMA : .....



DOCUMENTO DE IDENTIDAD : 70323204
NOMBRES Y APELLIDOS : VERONICA QUISPE GARCIA
FECHA : 09-08-2022.