

Content Based Recommender System for Enrollment of Elective Subjects in Engineering Careers at the Public University of Arequipa

Jerson Erick Herrera Rivera, Bachiller¹

¹Universidad Nacional de San Agustín de Arequipa, Perú, jherrerar@unsa.edu.pe

Abstract— One of the main problems faced by university students (specifically the case of the Professional School of Systems Engineering of the National University of San Agustín de Arequipa) is to make the right decision in relation to the elective courses to be enrolled based on available information (preferences, syllabus, timetables, content of the subject, possible performance, teacher, and others). Under these circumstances, this research work seeks to develop a Recommender System based on the content of all subjects (described in detail in their respective summits) that has studied and giving appropriate importance with the grades that the student has achieved. Thus, the system will recommend the student in which subject lines to register to obtain better results in the academic field. The recommendation given will be based on the similarity between the contents of each subject with respect to the elective subject lines. In order to reach the objective, data has been analyzed of the students of the aforementioned school, between 2011 and 2016. The results obtained indicate that the elective courses have courses with very similar contents among them, which guarantee effective lines of learning. In addition, it was ensured that the proposed system achieved its objective of providing objective recommendations to students during the decision making process on which elective subjects to follow.

Keywords— Recommender system, Content based model, enrollment, term frequency, inverse document frequency, curriculum.

Digital Object Identifier (DOI): http://dx.doi.org/10.18687/LACCEI2019.1.1.41 ISBN: 978-0-9993443-6-1 ISSN: 2414-6390
--

Sistema de Recomendación Basado en Contenido Para Matrícula de Asignaturas Electivas en Carrera de Ingeniería en Universidad Pública de Arequipa

Jerson Erick Herrera Rivera, Bachiller¹

¹Universidad Nacional de San Agustín de Arequipa, Perú, jherrera@unsa.edu.pe

Abstract– Uno de los principales problemas que enfrentan los estudiantes universitarios (en específico el caso de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de San Agustín de Arequipa) es tomar la decisión correcta en relación a las líneas de asignaturas electivas a matricularse basado en información disponible (preferencias, sílabus, horarios, contenido de la asignatura, posible rendimiento, docente, y demás). Bajo estas circunstancias, el presente trabajo de investigación busca desarrollar un Sistema de Recomendación basado en el contenido de todas las asignaturas (descrito a detalle en sus respectivas sumillas) que haya estudiado y dándole la importancia adecuada con las calificaciones que el estudiante haya alcanzado. Así, el sistema recomendará al estudiante en qué líneas de asignaturas matricularse para obtener mejores resultados en el ámbito académico. La recomendación dada estará basada en la similitud entre contenidos de cada asignatura con respecto a las líneas de asignatura electivas. Para alcanzar el objetivo se ha analizado data de los estudiantes de la escuela mencionada, entre los años 2011 y 2016. Los resultados obtenidos indican que las líneas de asignatura electivas cuentan con cursos con contenidos muy similares entre ellos, que garantizan líneas de aprendizaje eficaces. Además, se garantizó que el sistema propuesto alcanzara su objetivo de brindar recomendaciones objetivas a estudiantes durante la toma de decisión sobre qué línea de asignaturas electivas seguir.

Keywords– Sistemas de recomendación, modelo basado en contenido, matrícula, term frequency, inverse document frequency, plan de estudios.

I. INTRODUCCIÓN

La Escuela Profesional de Ingeniería de Sistemas (EPIS) de la Universidad Nacional de San Agustín de Arequipa (UNSA) cuenta con una malla curricular de una cantidad considerable de asignaturas, muchas de los cuales son requisitos unas de otras, considerando además que cuenta con líneas de asignaturas electivas desde el segundo semestre del cuarto año de universidad. Existen 3 líneas de cursos electivos:

- 1) Línea A: a) Negocios Electrónicos, b) Tópicos Avanzados en Bases de Datos, c) Gestión de Sistemas de Información y Tecnologías de Información.
- 2) Línea B: a) Introducción al Desarrollo de Software de Entretenimiento, b) Computación Gráfica, Visión Computacional y Multimedia, c) Desarrollo de Software para Juegos.
- 3) Línea C: a) Introducción al Desarrollo de Nuevas Plataformas, b) Desarrollo Avanzado en Nuevas

Plataformas, c) Plataformas Emergentes.

De las líneas de asignaturas electivas descritas previamente, todo estudiante está en la obligación de seguir solo dos de ellas. La primera asignatura de cada línea no tiene requisito para ser llevada, pero la segunda y tercera tienen como requisito la asignatura previa de cada línea de asignaturas electivas.

Los estudiantes deben escoger las líneas de asignaturas más convenientes para ellos, de acuerdo a diferentes criterios (asignaturas interesantes acorde a sus preferencias, asignaturas en las cuales su rendimiento es mayor, etc.). Sin embargo, la toma de la decisión implica tareas que necesitan tiempo para ser analizadas e incluyen actividades como: buscar los contenidos de cada curso de cada línea de asignaturas electivas, examinar cuidadosamente los sílabus, solicitar acceso al plan de estudios para analizar la sumilla de cada asignatura involucrada, revisar estadísticas de la asignatura, o pedir consejos a diferentes estudiantes que ya cuenten con la experiencia de la asignatura, aunque los comentarios podrían ser demasiado subjetivos dependiendo de dicha experiencia.

La decisión al escoger en cuáles líneas de asignaturas matricularse trae consigo algunas complicaciones. Por ejemplo, la línea que se escoja debe ser terminada, así, si el estudiante escoge la línea A y B, él deberá matricularse y aprobar todas las asignaturas de dichas líneas, caso contrario no podrá obtener el grado de egresado o bachiller. Otra complicación, es que el estudiante no puede volver a escoger las líneas electivas una vez se haya matriculado en alguna. Ello significa que no existe la posibilidad de cambiar de líneas de asignaturas electivas una vez hayan sido elegidas la primera vez.

Además de las restricciones descritas previamente, la tasa de desaprobación y deserción es alta en la EPIS en comparación con otras escuelas profesionales dentro de la universidad. Es necesario por tanto una herramienta que sugiera adecuadamente a los estudiantes en qué líneas asignaturas electivas deberían matricularse basados en sus preferencias y rendimiento en todas las asignaturas que hayan estudiado previamente obteniéndose una recomendación objetiva y exacta, todo ello haciendo uso de las herramientas y técnicas de un Sistema de Recomendación, específicamente con un modelos basado en contenido.

El presente trabajo está organizado como sigue: la sección 2 da un *overview* de trabajos relacionados aplicando modelos de sistemas de recomendación en el ámbito educativo, especialmente haciendo uso de modelos basados en contenido.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2019.1.1.41>

ISBN: 978-0-9993443-6-1 ISSN: 2414-6390

La sección 3 describe la propuesta de solución, los objetivos que se buscan alcanzar con el desarrollo de la investigación y las contribuciones a la escuela profesional, principalmente a los estudiantes. La sección 4 detalla el procedimiento para desarrollar el sistema de recomendación basado en contenido. La sección 5 explica los resultados del modelo basado en contenido del sistema de recomendación desarrollado. Finalmente, la sección 6 describe las conclusiones alcanzadas.

II. TRABAJOS RELACIONADOS

Michael O'Mahony y Barry Smith en [1], han desarrollado un Sistema de Recomendación para matrículas en la universidad de Dublín, donde los alumnos estudian 12 módulos por año, de los cuales 10 son propios del área de estudio; y 2 módulos son ajenos a la carrera. Así, los autores desarrollaron el sistema basados en filtrado colaborativo y basado en contenido. El primero sugería módulos electivos basados en elecciones pasadas de estudiantes con comportamiento similar. El segundo, hacía uso de campos de texto asociados que detallan la descripción del módulo y los resultados de aprendizaje; tras esto removía las *stopwords* y los sufijos, para luego calcular la similitud entre módulos y determinar cuáles serían recomendados al estudiante.

En [2], Vialardi et al. proponen un Sistema de Recomendación de matrículas basado en el registro de rendimiento académico del estudiante. Para ello, trabaja con dos atributos que mejoran la importancia de la recomendación: a) dificultad inherente de un curso dado y b) medida de la competencia de un estudiante para un curso dado basado en notas obtenidas en cursos similares. Se evaluaron diferentes métodos de *data mining*: C4.5, KNN, Naïve Bayes, *Bagging* y *Boosting*, para alcanzar el mejor resultado para este dominio de aplicación. Concluyeron que *Bagging* es el método que garantiza la precisión en la predicción. Así, se desarrolló "*Student Performance Recommender System*" (SPRS) incluyendo un motor de aprendizaje que mostró un buen rendimiento bajo condiciones reales de ejecución.

En [3] se presenta a AACORN, un sistema basado en casos que recomienda cursos a estudiantes en la escuela de *Computer Science*, Telecomunicaciones, y Sistemas de Información (CTI) en la Universidad DePaul. La información de cada estudiante es organizada en base a cuatro características: el programa académico del estudiante, el plan curricular, el promedio general de calificaciones del estudiante, y la historia de cursos del estudiante. El sistema reutiliza la experiencia pasada de estudiantes para inferir los cursos apropiados que un alumno pueda matricularse en el siguiente periodo de estudios. La suposición básica de AACORN es que estudiantes similares tendrán historial de cursos similar. Dos estudiantes en el mismo programa y con intereses similares son propensos a tomar muchas veces los mismos cursos. De esta forma, un estudiante buscando una recomendación puede usar la experiencia de estudiantes que han completado el programa como si se tratara de una plantilla. Cada curso encontrado en la plantilla que el

estudiante no haya tomado es probablemente un buen curso para matricularse.

En [4] y [5] se considera la aplicación de algoritmos de *data mining*, como *clustering* (k-means) y reglas de asociación (algoritmo a priori) para recomendar cursos a estudiantes en sistemas *E-learning* (en este caso Moodle). Por ejemplo, si un estudiante le gustó el curso "Sistemas Operativos", es probable que el curso "Sistemas Distribuidos" sea de su agrado. Se desarrolló la solución utilizando el lenguaje de programación Java y la herramienta de minería Weka. Se evaluaron tres modelos: *clustering*, reglas de asociación y un algoritmo que combinaba ambas. Como resultado se concluye que el modelo combinado generaba más y mejores reglas, lo que permite recomendar al estudiante diferentes combinaciones de cursos a diferencia del modelo de reglas de asociación que sólo generaba una regla de asociación para la recomendación.

En [6] entienden la dificultad que tienen que enfrentar los estudiantes al escoger en cuáles cursos electivos matricularse. Una recomendación de cursos ayuda a los estudiantes en esta toma de decisión tan crucial. Así, se propone un Sistema de Recomendación de filtrado colaborativo, para brindar al estudiante una predicción exacta de la calificación que ellos pueden alcanzar si ellos escogen un curso en particular. Lo cual es de gran ayuda cuando ellos escojan un curso electivo, puesto que la nota es un parámetro muy importante para el estudiante en esta decisión. Se evaluó el modelo de filtrado colaborativo con un *dataset* real alcanzándose gran efectividad en términos de precisión.

III. PROPUESTA DE SOLUCIÓN

Los Sistemas de Recomendación (*Recommender Systems* - RS) [7] son técnicas y herramientas de software que brindan sugerencias de ítem para ser usados por un usuario. Las sugerencias dadas por un RS están dirigidas a apoyar a los usuarios en procesos de toma de decisiones, tales como qué productos comprar, qué música escuchar, o qué noticias leer [8]. Los RS son medios valiosos para que los usuarios online puedan hacer frente a la sobrecarga de información y ayudarlos a tomar mejores decisiones. Ahora son una de las más poderosas y populares herramientas de descubrimiento de información en la web.

El principal objetivo de los RS es brindar apoyo a los usuarios en la toma de sus decisiones (online). Particularmente, el fin es brindar recomendaciones accesibles de alta calidad para una comunidad grande de usuarios con características en común [9].

En los RS se pueden distinguir básicamente dos clases de enfoques de recomendación: basado en contenido y filtrado colaborativo. En los RS basado en contenido, los atributos descriptivos de los ítems son usados para hacer las recomendaciones. El término "contenido" hace referencia a esas descripciones. Así, los ratings y el comportamiento de los usuarios, son combinados con la información del contenido disponible en los ítems [10].

Haciendo el paralelo con el propósito de la investigación, los ítems equivalen a las diferentes asignaturas que forman parte del conjunto de la malla curricular. El contenido equivale al texto completo de las sumillas de todas las asignaturas de la malla curricular descritas en un documento denominado “Plan de Estudios”. Los atributos son el resultado del análisis de textos que se van a aplicar a cada sumilla. Los usuarios hacen referencia a los estudiantes de la EPIS que se encuentran ante la decisión de escoger en cuáles líneas de asignaturas electivas matricularse. Finalmente, los ratings equivalen al rendimiento (notas finales) y preferencias de cada estudiante sobre las asignaturas que haya cursado, y por tanto, el comportamiento del usuario hace referencia a la predisposición que tiene cada estudiante hacia un grupo específico de asignaturas, donde mostraría un mayor rendimiento con respecto al resto y definirían su perfil.

A. Objetivo General

Diseñar e implementar una arquitectura de Sistema de Recomendación basado en contenido que sugiera adecuadamente a los estudiantes en qué líneas de asignaturas electivas matricularse basados en el perfil del estudiante y en el perfil de todas las asignaturas de la carrera que haya estudiado, obteniéndose una recomendación objetiva y exacta.

Para ello es necesario analizar los preferencias y el rendimiento académico de cada estudiante (basado principalmente en las notas que haya alcanzado en cada matrícula) y analizar el contenido de todas las asignatura que forman parte de la malla curricular (para ello se va a analizar el contenido de cada sumilla haciendo uso de herramientas de análisis de textos).

B. Objetivos Específicos

- 1) Recolectar y estructurar la data de notas de estudiantes de las matrículas que se hayan dado entre los años 2011 y 2016.
- 2) Recolectar y estructurar la data de las sumillas que forman parte de la malla curricular.
- 3) Analizar el rendimiento de cada estudiante a través de las notas que haya obtenido en cada matrícula.
- 4) Determinar el perfil del estudiante definido por las asignaturas en las que haya mostrado mejor rendimiento.
- 5) Analizar las sumillas para definir un perfil de cada uno basado en el contenido de los mismos.
- 6) Generar una recomendación a cada estudiante de la línea de asignaturas electiva a seguir.
- 7) Verificar que las líneas de asignaturas electivas cuentan con cursos cuyos contenidos buscan alcanzar los mismos objetivos y competencias en los estudiantes.

C. Resume General

La presente investigación busca resolver uno de los diferentes problemas que se presentan a lo largo de los estudios superiores mediante el análisis del rendimiento de los estudiantes y el análisis de los contenidos de enseñanza, herramientas que forman parte de *Educational Data Mining*

(EDM) [11][12][13][14] y que suscita interés en la comunidad científica que estudia este campo tan crítico que involucra a la educación y a la minería de datos.

El sistema propuesto ha involucrado el análisis, diseño e implementación de un RS que permitirá a los estudiantes saber en qué líneas de asignaturas electivas deberían matricularse acorde a su rendimiento y a sus preferencias y reducir de ese modo, la posibilidad de desaprobación o deserción. Es importante el desarrollo del RS porque también será un ejemplo adicional de que los RS pueden ser empleados en diferentes entornos o fines y pueden ser útiles para resolver problemas en diversos campos (en este caso en el campo de la educación), ya que los RS han sido empleados principalmente para recomendar productos o servicios (por ejemplo: [15], [16], [17], [18], [19], [20], [21]) en el campo del marketing, películas, turismo, noticias y ventas. Algunas contribuciones que se buscan son: a) RS de líneas de asignaturas electivas a matricularse para estudiantes de la EPIS, b) identificación de asignaturas más afines a cada estudiante, c) *clustering* de asignaturas según el contenido de su sumilla, d) RS web en base a modelos de recomendación basado en contenido utilizando analizadores de texto.

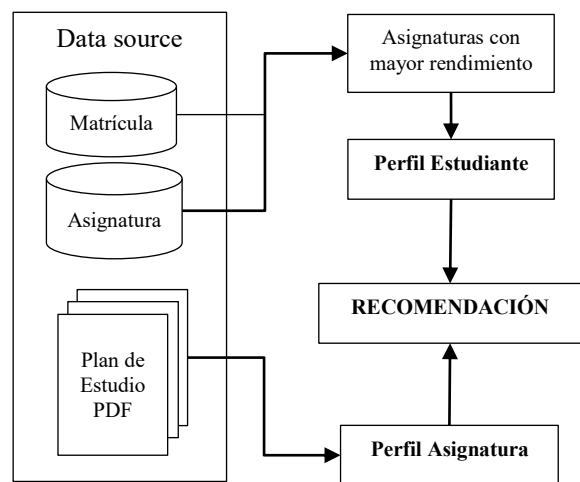


Fig. 1 Secuencia de actividades para el desarrollo del Sistema de Recomendación propuesto.

En la Fig. 1 se muestra la secuencia de actividades a realizar para el desarrollo de la propuesta y poder alcanzar los objetivos planteados anteriormente. Ello involucra la obtención de una base de datos con información de las asignaturas y de las matrículas. Tras esto, se ha realizado dos procesos: 1) construir el perfil del estudiante haciendo uso de las notas alcanzadas en todas sus matrículas y un análisis de texto de las sumillas de los cursos con mayor rendimiento y 2) construir el perfil de todas las asignaturas aplicando análisis de texto de sus sumillas (modelo basado en contenido de los RS). Finalmente se analiza el perfil del estudiante con el perfil de todas las asignaturas y se identifica objetivamente en cuáles dos líneas de asignaturas electivas debiera matricularse.

IV. DESARROLLO DE LA SOLUCIÓN

Para la implementación de la propuesta se utiliza el lenguaje de programación Python, en conjunto con el ambiente de desarrollo Jupyter [22]. Adicionalmente, fue necesaria la incorporación de las librerías de Python: pandas, numpy, tika, pickle, sklearn, scipy.

La data de matrículas, asignaturas y notas estaba almacenada en una base de datos Access. Asimismo, el contenido de las asignaturas fue obtenido del Plan de Estudios de la EPIS, específicamente en el apartado que reúne todas las sumillas.

A. Pre procesamiento de data

Los atributos de las matrículas son los siguientes: identificador de la matrícula, identificador del estudiante, nombre de la asignatura, grupo, número de matrícula, nota obtenida, condición (aprobado, desaprobado, evaluación por suficiencia, evaluación por jurado, abandono, retirado). Por otro lado, la información de cada sumilla de cada asignatura consigna información como: nombre de la asignatura, departamento académico al que pertenece, tipo de área curricular, tipo (teórico, práctico), competencias, reglamento de estudios generales, unidades didácticas, y contenidos.

La data de matrículas, asignaturas y notas tuvo que ser exportada como archivo csv (*comma-separated values*) para que pueda ser analizada en Jupyter haciendo uso de la librería pandas. Asimismo, las sumillas de las 71 asignaturas fueron extraídas del Plan de Estudios (en formato PDF) utilizando la librería tika y almacenado en archivos de texto para su posterior análisis de textos.

B. Definición del perfil del estudiante

La información de las notas ha sido organizada por clases. Una clase hace referencia a un conjunto de estudiantes que comparten la misma asignatura, grupo, ciclo, y año calendario. Tras esto las notas son normalizadas de modo, que representen adecuadamente el rendimiento de un estudiante dentro de las clases a las que pertenece.

Luego se determinó las asignaturas en las cuáles un estudiante mostró mejor rendimiento y las cuales determinarán su perfil de preferencia.

TABLA I
ASIGNATURAS CON MAYOR RENDIMIENTO DE UN ESTUDIANTE

Asignatura	Nota	Nota normalizada
Redes y Comunicación de Datos	15	1.000000
Aspectos Formales de Especificación y Validación	17	1.000000
Sistemas Distribuidos	12	1.000000
Técnicas de Estudio	15	0.937500
Análisis y Diseño de Sistemas	13	0.928571
Innovación y Creatividad	13	0.928571
Tecnologías de la Información	13	0.928571
Tecnologías de Objetos	15	0.882353
Interacción Humano Computador	14	0.875000
Programación Web I	13	0.866667

En la Tabla I se muestran las asignaturas con mayor rendimiento de un estudiante con código 487 con respecto a su clase. Nótese que hay algunas notas que son mayores y no

ocupan los primeros lugares. Esto es debido a que en esos cursos hubo otros estudiantes que alcanzaron mejores calificaciones, y por tanto la nota normalizada no es tan alta como se esperaba.

C. Definición del perfil de las asignaturas

Con las sumillas de cada asignatura almacenadas en documentos de texto, y haciendo uso del objeto TfidfVectorizer [23] de la librería sklearn, se genera un vocabulario de características comunes a todos los documentos y lo más importante, se genera una matriz de características TF-IDF.

Los documentos son codificados por TF-IDF como vectores en un espacio euclidiano [9]. Las dimensiones del espacio corresponden a las características que aparecen en el vocabulario. TF (*term frequency*) describe qué tan frecuente un cierto término aparece en un documento (asumiendo que las palabras importantes aparecen más a menudo). IDF (*inverse document frequency*) es la medida que se combina con el TF; su objetivo es reducir el peso de términos que aparecen muy a menudo en todos los documentos. La idea es que esos términos muy frecuentes no son útiles para discriminar entre documentos, por lo que se debe dar más peso a las palabras que aparecen en unos pocos documentos.

Es necesario definir algunos parámetros durante el proceso de vectorización de características. Uno de ellos es el conjunto de *stopwords* [10]; estas palabras son muy frecuentes en los documentos y en el idioma, pero no representan mayor significado al definir el modelo. Son palabras como conjunciones, preposiciones, verbos modales, incluso números.

Otro parámetro son la posibilidad de extraer frases [10]. La idea es detectar palabras que aparecen juntas en los documentos y que tienen un significado diferente al estar juntas. Por ejemplo, la frase 'red neuronal' tiene un significado diferente al que tienen las palabras que lo constituyen.

D. Generación de Recomendación

A continuación se debe contrastar el perfil del estudiante contra la matriz TF-IDF generada previamente para identificar la similitud con el vocabulario generado con todas las sumillas. Con la nueva matriz y haciendo uso de sklearn [24] se calcula una matriz que indique la similitud entre las diferentes sumillas y el perfil del estudiante utilizando la similitud coseno.

Luego, se define una función que recibe como parámetro el código de asignatura o código de estudiante y devuelve una lista con los ítems más similares a la entrada de la función, acompañado del índice de similitud.

V. RESULTADOS

Se evaluó el sistema con la similitud entre las asignaturas electivas. Así, la Tabla II muestra las asignaturas más similares al primer curso de la línea A de asignaturas electivas, los cuales son precisamente las dos asignaturas adicionales de dicha línea. La Tabla III muestra que sólo una asignatura de la

línea B es muy similar a la primera; la tercera asignatura de esa línea electiva ocupa el noveno lugar de similitud.

TABLA II
ASIGNATURAS CON MAYOR SIMILITUD A LA ASIGNATURA “NEGOCIOS ELECTRÓNICOS”

Código Asignatura	Asignatura	Similitud
1704256	Negocios Electrónicos (E)	1.000000
1705273	Gestión de Sistemas y Tecnologías de la Información (E)	0.445930
1705166	Tópicos Avanzados en Bases de Datos (E)	0.353251
1704147	Inteligencia Artificial	0.309864
1704258	Introducción al Desarrollo en Nuevas Plataformas (E)	0.250146

TABLA III
ASIGNATURAS CON MAYOR SIMILITUD A LA ASIGNATURA “INTRODUCCIÓN AL DESARROLLO DE SOFTWARE DE ENTRETENIMIENTO”

Código Asignatura	Asignatura	Similitud
1704257	Introducción al Desarrollo de Software de Entretenimiento (E)	1.000000
1705274	Desarrollo de Software para Juegos (E)	0.852662
1704148	Diseño y Arquitectura de Software	0.364381
1704149	Sistemas Distribuidos	0.307367
1703137	Organización y Métodos	0.294254

TABLA IV
ASIGNATURAS CON MAYOR SIMILITUD A LA ASIGNATURA “INTRODUCCIÓN AL DESARROLLO DE NUEVAS PLATAFORMAS”

Código Asignatura	Asignatura	Similitud
1704258	Introducción al Desarrollo de Nuevas Plataformas (E)	1.000000
1705168	Desarrollo Avanzado en Nuevas Plataformas (E)	0.720941
1705275	Plataformas Emergentes (E)	0.625516
1704252	Gestión de Proyectos de Software	0.439836
1705160	Proyecto de Ingeniería de Software I	0.418877

Finalmente, la Tabla IV muestra el mismo comportamiento con respecto a la línea de asignaturas electivas A, donde las asignaturas más similares son precisamente las que pertenecen a esa línea de asignaturas electivas.

TABLA V
LÍNEAS DE ASIGNATURAS ELECTIVAS RECOMENDADAS

Línea	Asignaturas	Similitud
Línea A	“Negocios Electrónicos”, “Tópicos Avanzados en Bases de Datos”, “Gestión de Sistemas de Información y Tecnologías de Información”	0.470290
Línea B	“Introducción al Desarrollo de Software de Entretenimiento”, “Computación Gráfica, Visión Computacional y Multimedia”, “Desarrollo de Software para Juegos”	0.446106
Línea C	“Introducción al Desarrollo de Nuevas Plataformas”, “Desarrollo Avanzado en Nuevas Plataformas”, “Plataformas Emergentes”	0.311300

Con respecto a la recomendación de las líneas de asignaturas electivas, se alcanzaron recomendaciones

objetivas basadas en el rendimiento e interés de estudiantes. Por ejemplo, en la Tabla V se muestra las líneas de asignaturas recomendadas (línea A y línea B) de un estudiante que obtuvo el mejor rendimiento en las asignaturas: “Gestión de Proyectos de Software”, “Tópicos Avanzados en Ingeniería de Software”, “Mantenimiento, Configuración y Evolución de Software”, “Estadística Matemática, probabilidades y métodos empíricos” e “Inteligencia Artificial”, Así, el sistema recomienda a este estudiante seguir la línea A (Negocios Electrónicos, etc.) y la línea B (Introducción al Desarrollo de Software de Entretenimiento, etc.).

VI. CONCLUSIONES

Al analizar la similitud entre cada línea de asignaturas electivas se demostró que las líneas A (“Negocios Electrónicos”, “Tópicos Avanzados en Bases de Datos”, “Gestión de Sistemas de Información y Tecnologías de Información”) y C (“Introducción al Desarrollo de Nuevas Plataformas”, “Desarrollo Avanzado en Nuevas Plataformas”, “Plataformas Emergentes”) cuentan con asignaturas que tienen contenidos muy similares definidos en las sumillas que forman parte del Plan de Estudios de la Escuela Profesional de Ingeniería de Sistemas. Por otro lado, la línea B (“Introducción al Desarrollo de Software de Entretenimiento”, “Computación Gráfica, Visión Computacional y Multimedia”, “Desarrollo de Software para Juegos”), muestra un comportamiento diferente: la primer y tercer asignatura sí presentan similitud en su contenido, sin embargo la segunda asignatura de esta línea (Computación Gráfica, Visión Computacional y Multimedia), no presenta mayor similitud ya que esta ocupa el noveno lugar de similitud en comparación con su propia línea. Ello indica que debería revisarse mejor el contenido de esta asignatura, o quizás evaluar cambiarla por otra que sí sea similar y logren en conjunto alcanzar los objetivos que tienen planteados como una línea de asignaturas electivas sólida y consistente.

En cuanto a la recomendación de líneas de asignaturas electivas a escoger para la matrícula, los resultados han sido positivos y reflejan adecuadamente el interés y rendimiento de los estudiantes que han probado el sistema. Por tanto, se puede afirmar que las recomendaciones alcanzadas apoyan objetivamente en la decisión de los estudiantes durante la matrícula.

ACKNOWLEDGMENT

Me gustaría agradecer a Víctor Cornejo Aparicio - Director de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de San Agustín - por el pronto apoyo con material necesario para la investigación.

REFERENCIAS

- [1] M. P. O’Mahony and B. Smyth, “A recommender system for on-line course enrolment: an initial study,” *Proc. 2007 ACM Conf. Recomm. Syst. - RecSys ’07*, p. 133, 2007.
- [2] C. Vialardi *et al.*, “A data mining approach to guide students through the enrollment process based on academic performance,” *User Model. User-adapt. Interact.*, vol. 21, no. 1–2, pp. 217–248,

- 2011.
- [3] R. Burke, "Aacorn: A CBR recommender for academic advising," *Tech. Rep. TR05-015*, 2005.
 - [4] S. B. Aher and L. Lobo, "Applicability of data mining algorithms for recommendation system in e-learning," *Proc. Int. Conf. Adv. Comput. Commun. Informatics - ICACCI '12*, p. 1034, 2012.
 - [5] S. B. Aher and L. M. R. J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data," *Knowledge-Based Syst.*, vol. 51, pp. 1–14, 2013.
 - [6] S. Ray and A. Sharma, "A collaborative filtering based approach for recommending elective courses," *Commun. Comput. Inf. Sci.*, vol. 141 CCIS, pp. 330–339, 2011.
 - [7] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, Second Edi., vol. 54, 2015.
 - [8] A. S. Lampropoulos and G. A. Tsihrintzis, *Machine Learning Paradigms Applications in Recommender Systems*. 2015.
 - [9] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*, vol. 40. Cambridge, 2011.
 - [10] C. C. Aggarwal, *Recommender Systems The TextBook*, vol. 40, no. 3. Springer, 2016.
 - [11] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 3, no. 1, pp. 12–27, 2013.
 - [12] R. Jindal and M. D. Borah, "A Survey on Educational Data Mining and Research Trends," *Int. J. Database Manag. Syst.*, vol. 5, no. 3, pp. 53–73, 2013.
 - [13] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1432–1462, 2014.
 - [14] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Comput. Educ.*, vol. 51, no. 1, pp. 368–384, 2008.
 - [15] M. D. Dussán Sarria, "Sistema de recomendación web basado en la actividad de los usuarios de la Universidad Nacional de Colombia," Universidad Nacional de Colombia, 2012.
 - [16] E. R. Núñez Valdéz, "Sistemas de Recomendación de Contenidos para Libros Inteligentes," UNIVERSIDAD DE OVIEDO, 2012.
 - [17] S. Landa Tinco and F. J. Ticona Pacheco, "Un Modelo de Recomendación de Productos utilizando Redes Sociales implementado en Tecnologías de la Web Semántica," UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS, 2015.
 - [18] H. Chen and A. L. P. Chen, "A music recommendation system based on music data grouping and user interests," *CIKM '01 Proc. tenth Int. Conf. Inf. Knowl. Manag.*, pp. 231–238, 2001.
 - [19] F. M. Hsu, Y. T. Lin, and T. K. Ho, "Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3257–3264, 2012.
 - [20] S. J and A. N, "CinemaScreen Recommender Agent: Combining collaborative and content-based filtering," *IEEE Intell. Syst.*, vol. 21, no. 1, pp. 35–41, 2006.
 - [21] S. T. R. Gopidi, "Automatic User Profile Construction for a Personalized News Recommender System Using Twitter," no. July, 2015.
 - [22] M. Ragan-Kelley *et al.*, "The Jupyter/IPython architecture: a unified view of computational research, from interactive exploration to communication and publication.," in *AGU Fall Meeting Abstracts*, 2014.
 - [23] D. Cournapeau and M. Brucher, "TfidfVectorizer," 2007. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html. [Accessed: 11-Jan-2019].
 - [24] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in {P}ython," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.