



Facultad de Ingeniería
INGENIERÍA INDUSTRIAL

Proyecto de Grado – Segundo Semestre 2017

Trabajo de grado en modalidad de aplicación

**Diseño de un aplicativo web que recomiende asignaturas
electivas a estudiantes de ingeniería industrial de la Pontificia
Universidad Javeriana**

Valentina Carvajal Vera ^{a, c}, María Camila Villasmil Vergel ^{a, c},

Ricardo Otero Caicedo ^{b, c}

^aEstudiante de Ingeniería Industrial

^bProfesor, Director del Proyecto de Grado, Departamento de Ingeniería Industrial

^c Pontificia Universidad Javeriana, Bogotá, Colombia

Resumen

Las asignaturas electivas buscan el desarrollo integral de los estudiantes, la Pontificia Universidad Javeriana ofrece una gran cantidad de opciones, lo que, usualmente, dificulta el proceso de elección e inscripción de este tipo de asignaturas. Este problema lleva a los estudiantes a retirar las asignaturas o ver asignaturas electivas cuyo contenido no es de su agrado. Estas situaciones no promueven, necesariamente, el proceso de formación integral. Por esta razón, se hace necesaria una herramienta que recomiende asignaturas electivas a los estudiantes, según sus preferencias de aprendizaje. Este problema se va abordar a través de un algoritmo de recomendación mixto que base su predicción en el historial académico de los estudiantes. Se realizó una implementación web de este sistema de recomendación para facilitar el proceso de elección e inscripción de asignaturas electivas, promoviendo, la esencia de la educación integral.

Palabras claves: Asignaturas electivas, sistemas de recomendación.

Abstract

Elective subjects seek the student's integral development, the Pontificia Universidad Javeriana offers a great amount of options, which, usually, makes the election and inscription processes of this kind of subjects more difficult. This problem leads the students to withdraw the subjects or to enroll into subjects whose contents do not reflect their likings. These situations don't necessarily promote the process of integral formation. This reason makes necessary a tool that recommends elective subjects to students, according to their learning preferences. And the problem will be addressed through a hybrid recommender algorithm that bases its prediction on students' academic records. A web implementation of this recommendation system was made to facilitate the process of electing and registering elective subjects, promoting the essence of integral education.

Key Words: elective subject, recommendation systems.

Justificación y planteamiento del problema

El destino de la sociedad se proyecta desde la base de la educación. Los niveles de progreso y desarrollo de las comunidades dependen en gran medida, del grado y la calidad de educación que tengan las personas que la conforman. Así, el panorama se vuelve un poco más exigente cuando se hace referencia a las Instituciones de Educación Superior (IES), debido a que allí se forman los profesionales que pondrán sus conocimientos y experiencias al servicio de la comunidad, en la solución de problemas cada día más complejos y que requieren de la articulación de diferentes áreas del conocimiento para su solución. Por esto, una gran parte de los esfuerzos de las IES están direccionados al desarrollo de las habilidades de trabajo interdisciplinario, buscando flexibilizar el pensamiento de los estudiantes, potencializar su pensamiento crítico (Ackerman, 1998) y apoyar el desarrollo de habilidades analíticas, cognitivas, de síntesis e integración de conceptos (Carvajal Escobar, 2010). A raíz de esta necesidad, las IES han replanteado sus currículos con el fin de apoyar la formación integral de los estudiantes desde programas flexibles e interdisciplinarios.

Según Abraham Magendzo (1986) la flexibilidad es la posibilidad que tiene el currículo de ser modificado y adaptado a las necesidades de las IES y a las realidades del contexto social, político y económico de cada país, pero, sobre todo, como respuesta directa a los gustos y preferencias personales de los estudiantes. Por otra parte, la interdisciplinariedad de los programas académicos se presenta como una alternativa ante la dificultad de los estudiantes a la hora de integrar los conocimientos de las distintas asignaturas en la solución de problemas, la sectorización del pensamiento y la individualización del conocimiento. De esta forma el aspecto interdisciplinario complementa el aspecto flexible de la formación integral.

Fundamentalmente, el objetivo de la formación integral es direccionar procesos educativos que busquen acercar a los estudiantes a su realización plena, resultando en individuos éticos, felices, satisfechos y socialmente responsables. Esto contribuye al mejoramiento de la calidad de vida del entorno social, puesto que ningún ser humano se forma para sí mismo, sino que lo hace para contribuir en su propio contexto sociocultural (Guerra, Mórtigo Rubio, & Berdugo Silva, 2013).

Las IES son conscientes de la importancia de la formación integral de sus estudiantes. Por esto, suelen diseñar los currículos de sus programas desde dos componentes, uno de ellos básico o fundamental, compuesto por asignaturas que proveen a los estudiantes conocimientos técnicos, herramientas y habilidades propias de su área de formación y un componente flexible, que ofrece asignaturas complementarias y electivas, brindándole al plan de estudios el sentido de interdisciplinariedad y formación integral, respectivamente.

Ahora bien, la responsabilidad en la conformación del componente flexible de los currículos recae únicamente sobre el estudiante, ya que es éste quien selecciona libremente las asignaturas que desea inscribir y que conformarán esta parte de su plan de estudios. Durante el proceso de inscripción de las asignaturas electivas, los estudiantes se enfrentan a diferentes inconvenientes: i) es común que las universidades ofrezcan una gran cantidad de asignaturas, lo que dificulta la elección de los estudiantes, ii) usualmente, los estudiantes deben ingresar a los sistemas de información de la universidad para seleccionar todas las posibles asignaturas candidatas que cumplen con sus expectativas de aprendizaje, iii) el sistema de información universitaria usualmente provee una descripción de las asignaturas muy básica. Estos tres aspectos, ocasionan que, en algunos casos, los estudiantes no conozcan sobre la existencia de algunas asignaturas y, por lo tanto, sus elecciones finales no necesariamente satisfagan sus convicciones de formación.

Para conocer más a fondo el proceso de inscripción de asignaturas electivas de la PUJ, se realizó una encuesta a los estudiantes de ingeniería industrial, donde se les preguntó cómo logran encontrar información acerca de las electivas, qué plataformas usan, en qué se basan su elección, entre otras.

La encuesta fue enviada a 1.000 estudiantes, obteniendo 128 respuestas. Un 70% de los estudiantes toma decisiones acerca de la inscripción de asignaturas electivas a partir de consejos de sus compañeros, mientras que

un 60% busca en el listado del sistema de información de la universidad y finalmente un 50% busca información en redes sociales. Por otra parte, con respecto al criterio a partir del cual seleccionan una asignatura, el 75% de los estudiantes basa su elección en la afinidad existente entre el contenido de la asignatura y sus gustos e intereses, aun cuando esta cifra es alta, el 72% de los estudiantes ha inscrito asignaturas electivas cuyo contenido no fue de su agrado y el 48% ha retirado electivas. Esto significa que, aunque los estudiantes buscan asignaturas afines a sus gustos, una gran mayoría de ellos no se sienten satisfechos con las asignaturas que han inscrito. Adicionalmente y como posible explicación a la insatisfacción y deserción de los estudiantes en este tipo de asignaturas, un 24% de los estudiantes no sabe dónde buscarlas, mientras que un 20% asegura que existen muchas opciones lo que hace complejo realizar la elección correcta.

Para solucionar este tipo de inconvenientes, en algunos contextos educativos se han implementado sistemas de recomendación, los cuales generan una selección de ítems para recomendar a un usuario ya sea basado en sus datos históricos o en su similaridad con otros usuarios. Este tipo de sistemas hacen más eficientes los procesos de selección de cualquier tipo de elementos ya que, se limitan a presentar sólo aquellos ítems que tienen una mayor probabilidad de afinidad con el usuario.

Antecedentes

Los sistemas de recomendación son filtros de información que proporcionan ítems atendiendo las preferencias e intereses de los usuarios, buscando predecir qué ítem puede ser interesante para el usuario (Almudena Ruiz, 2011). Estos proporcionan como sugerencias un subconjunto de elementos pertenecientes a las alternativas que se consideren más apropiadas para el usuario, de esta manera, permiten que los usuarios tengan una visión personalizada de grandes agrupaciones de ítems o elementos, aliviando el problema de sobrecarga de opciones de diversos elementos. Existen tres tipos de sistemas de recomendación: los filtros colaborativos, filtros basados en contenido y filtros mixtos (Altuna Castillo, Guibert Estrada, & Estrada, 2014).

De manera formal, los sistemas de recomendación definen un conjunto de ítems, $I = \{I_1, I_2, I_3, \dots, I_n\}$ que el usuario puede elegir y una matriz de transacciones, $D = \{D_1, D_2, D_3, \dots, D_n\}$, en donde cada una de sus filas representa un conjunto de ítems I que son elegidos por un usuario. De esta manera, la matriz D conserva la información sobre las elecciones anteriores que los usuarios han realizado sobre el conjunto de ítems I . Una regla importante de asociación para los filtros colaborativos será la implicación $X \rightarrow Y$, donde X y Y son subconjunto de I , la cual expresa que, si el usuario ha elegido un conjunto de ítems X , es probable que también esté interesado en el conjunto de ítems Y . Para los filtros basados en contenido se asignan pesos a los atributos de los ítems, si en el perfil de un usuario se ha determinado los atributos del conjunto de ítems X que están entre sus elecciones anteriores, es probable que se recomiende un conjunto de ítems Y que compartan atributos con X (Pinho Lucas, Octubre de 2010).

Los filtros basados en contenido tienen al ítem como base de predicción y utilizan características del elemento para realizar la recomendación. Un ejemplo de este filtro se da en el servicio de música en *streaming*, donde los datos que se disponen de cada canción son: grupo, cantante, discográfica, género y calificaciones que el usuario ha hecho sobre el tema. De esta forma, se realizan las predicciones a partir de la extracción de la información del ítem, el perfil del usuario y sus calificaciones anteriores (Gomez Verdejo, 2012). Este tipo filtro ha sido utilizado en la educación en los recursos educativos abiertos (Terry González, Estrada Sentí, & Arteaga Gómez, 2016), redes personales de aprendizaje (Woerndl, Huebner, Bader, & Gallego-Vico, 19 de Septiembre 2013), material didáctico de aprendizaje (Ghauth & Abdullah, 2010) y en los recursos educativos para la elección de objetos de aprendizaje (Casali, Gerling, Deco, & Bender, 2010). Usualmente, los filtros basados en contenido requieren de la creación manual de los atributos de cada ítem, sin embargo, cuando la información relacionada a los ítems no está estructurada, sino que únicamente se presentan descripciones en lenguaje natural, es posible utilizar técnicas de minería de texto para extraer los atributos de los ítems. Algunas aplicaciones se han realizado en tiendas de ropa online (Ghani, Probst, Liu, Crema, & Fano, 2006), estudio de revisión de cámaras digitales (Aciar, Zhang, Simoff, & Debenham, 2006), servicios móviles (Li, y otros, 2010) y recomendación de películas (Domingues, Sundermann,

& Garcia Manzato, 2014).

Los filtros colaborativos basan su lógica en las características de los usuarios. Los datos obtenidos de los usuarios se convierten en el centro de este filtro. Analizan las preferencias, elecciones anteriores y calificaciones que ha dado el usuario a otros ítems y buscan otros usuarios similares y que han tomado decisiones parecidas a él (González, 2014). Un ejemplo de este tipo de filtro se da en la recomendación de películas, como en la plataforma *Movie Recommender System*, en la cual se genera un modelo de recomendación donde se obtienen las afinidades de los usuarios a través de las evaluaciones que estos realizan, los usuarios son quienes proporcionan de manera explícita los parámetros del producto que prefiere (Becerra Cortés, Jimenez, & Gonzalez, 2015). Los filtros colaborativos han sido utilizados en tiendas on-line (Galán Nieto, 2007), en la recomendación de noticias (Galán Nieto, 2007) y en la recomendación de películas *MovieLens* (Jiang & Yang, 2016). En el campo de la educación se han aplicado estos filtros en plataformas MOODLE de aprendizaje y enseñanza (Itmazi, 2005) y en los Sistemas Tutores Inteligentes (STI), aplicaciones informáticas que se encargan de enseñar de forma similar a como lo haría un tutor humano experimentado (Altuna Castillo, Guibert Estrada, & Estrada, 2014).

Igualmente, se han desarrollado sistemas de recomendación con algoritmos de filtrado mixto, los cuales combinan los enfoques basados en contenido y colaborativos, reduciendo el retraso en la entrega de recomendaciones a usuarios nuevos del sistema. Un ejemplo de estos algoritmos se da en el sistema de contenidos audiovisuales que realiza sus recomendaciones en función de los niveles de interés de los usuarios, la información de los contenidos y las métricas de similitud y relaciones entre los perfiles de los usuarios (Duran Dorado & Arciniega Herrera, 2013). Estos algoritmos de recomendación se han implementado en el turismo (Leiva, Guevara, Rossi, & Aguayo, 2014), en la recomendación de restaurantes (Burke, 2002) y en bibliotecas digitales (Gao, Xing, Du, & Wang, 2007). En el campo de la educación se han aplicado en *Learning Objects*, plataforma que busca reutilizar recursos educativos para apoyar los procesos de enseñanza y aprendizaje (Rodríguez Marín, Duque, & Ovalle, 2016).

Las plataformas de enseñanza online utilizan los sistemas de recomendación con el fin de enfocar al usuario en temáticas de su preferencia, ya sea en cursos ya vistos, en temáticas relacionadas, en foros en los que ya ha participado. Un ejemplo se da en el sistema MOOCs, curso abierto masivo en línea, donde se ofertan cursos online accesibles a nivel mundial de manera abierta. La plataforma de MOOCs edX ha diseñado un algoritmo de recomendación basado en filtrado colaborativo, el algoritmo se centra en recomendar problemas con los que han interactuado, realizado y aprobado usuarios similares al usuario al que se va a realizar dicha recomendación (Collado Sánchez, 2014). Asimismo los sistemas de recomendación son usados en repositorios de recursos educativos y sistemas de gestión de aprendizaje como CoFIND y CYCLADES, que permiten a los usuarios encontrar recursos educativos que mejor se adaptan a sus necesidades o intereses (Ruiz Iniesta, 2011).

En la siguiente tabla se presentan algunos de los artículos consultados en los cuales se explica la utilización de los sistemas de recomendación en el campo de aprendizaje, en plataformas o cursos educativos. En ellos se recomiendan profesores, cursos, materias y material de estudio como resultado de la aplicación de filtros basados en contenido, colaborativos o mixtos.

Artículo	Tipo de Sistema de Recomendación	Ambiente
(Collado Sánchez, 2014)	Filtro Colaborativo	Plataforma de aprendizaje MOOCs
(Terry González, Estrada Sentí, & Arteaga Gómez, 2016)	Filtro basado en contenido	Los recursos educativos abiertos

(Woerndl, Huebner, Bader, & Gallego-Vico, 19 de Septiembre 2013)	Filtro basado en contenido	El campo de redes personales de aprendizaje
(Rodríguez Marín, Duque, & Ovalle, 2016)	Filtro Mixto	La plataforma de enseñanza y aprendizaje Learning Objects
(Ruiz Iniesta, 2011)	Filtro Colaborativo	Repositorios de recursos educativos y sistemas de gestión de aprendizaje CoFIND v CYCLADES
(Altuna Castillo, Guibert Estrada, & Estrada, 2014)	Filtro Colaborativo	Los Sistemas Tutores Inteligentes (STI), aplicaciones informáticas
(Casali, Gerling, Deco, & Bender, 2010)	Filtro basado en contenido	Recursos educativos para la elección de objetos de aprendizaje.
(Itmazi, 2005)	Filtro Colaborativo	Plataforma de aprendizaje MOODLE

Los sistemas de recomendación son evaluados por medio de métricas para definir qué tan adecuadas son las recomendaciones. Las métricas permiten medir qué tan cerca está la selección de ítems realizada por el modelo contra la verdadera selección que haría el usuario según su preferencia (Rodríguez Buitrago, 2012). Según (Gunawardana & Shani, 2009) se encuentran diversas métricas para realizar la evaluación de los algoritmos de recomendación como lo son: elecciones finales del usuario, precisión, cobertura, confianza, novedad, serendipia, diversidad, utilidad, riesgo, robustez, privacidad, adaptabilidad, escalabilidad. Aquí podrían decir. En conjunto, estas métricas pretenden evaluar desde qué tan acertada es la recomendación, qué tan diversas son las recomendaciones, hasta qué tanta sorpresa genera las recomendaciones en el usuario.

Para evaluar el sistema de recomendación de las asignaturas electivas se proponen las métricas de precisión o exactitud, la cual mide la predicción de los sistemas de recomendación sobre las preferencias del usuario o su historial. Serendipia, mide lo sorprendente que son las recomendaciones. Finalmente, la diversidad, se refiere a qué tan diferentes son los objetos recomendados respecto a los ítems disponibles. Se seleccionaron estas métricas debido a que el aplicativo busca recomendar según las preferencias o historial del usuario, para esto se debe definir ítems nuevos relevantes para el usuario, seleccionar ítems de mayor calidad y buscar pluralidad y multiplicidad en los ítems recomendados.

Basándose en esto, para el proceso de inscripción de asignaturas electivas, se propone un algoritmo híbrido mixto, que recomiende a los estudiantes de ingeniería industrial asignaturas electivas basados en su historial de inscripción o preferencias, reduciendo el total de opciones de asignaturas electivas ofrecidas por la PUJ.

Objetivos

Diseñar un aplicativo web que recomiende a los estudiantes de Ingeniería Industrial de la pontificia universidad Javeriana las asignaturas electivas según su historial de inscripción.

- Recolectar y analizar información sobre la base de datos académica de ingeniería industrial con el fin de obtener los patrones de inscripción de materias electivas.
- Crear un sistema de recomendación de materias electivas y evaluarlo a partir de las métricas: precisión, serendipia y diversidad.
- Diseñar un aplicativo web para recomendar asignaturas electivas.
- Medir la percepción de los estudiantes sobre las recomendaciones realizadas por el aplicativo web, antes de la inscripción de las asignaturas.

Contexto

La Pontificia Universidad Javeriana (en adelante, PUJ) define las asignaturas electivas como aquellas de libre elección que responden a intereses particulares de los estudiantes. Adicionalmente, establece que deben constituir al menos el 10% del plan de estudios en los programas de pregrado (Javeriana, 2003). Así las cosas, todas aquellas asignaturas creadas con carácter electivo o que hagan parte de programas diferentes al de ingeniería industrial, se consideran asignaturas electivas para los estudiantes de dicho programa, quienes deben elegir las asignaturas a inscribir entre aproximadamente 3000 opciones, dependiendo del número de materias ofrecidas semestre a semestre por la universidad.

A partir de un análisis descriptivo preliminar realizado a la base de datos académica, que contiene registros de las asignaturas inscritas por los estudiantes de ingeniería industrial desde el primer periodo del 2006 hasta el primer periodo del 2017, se pudo establecer que aproximadamente el 70% de las asignaturas electivas inscritas por los estudiantes son de 2 créditos.

Adicionalmente, se encontró que con un 6% de participación, la asignatura electiva más inscrita por los estudiantes es Creatividad Organizacional, seguida de Introducción a la Administración y Hoja de Cálculo nivel avanzado cada una con una participación del 2% (Gráfico 1).

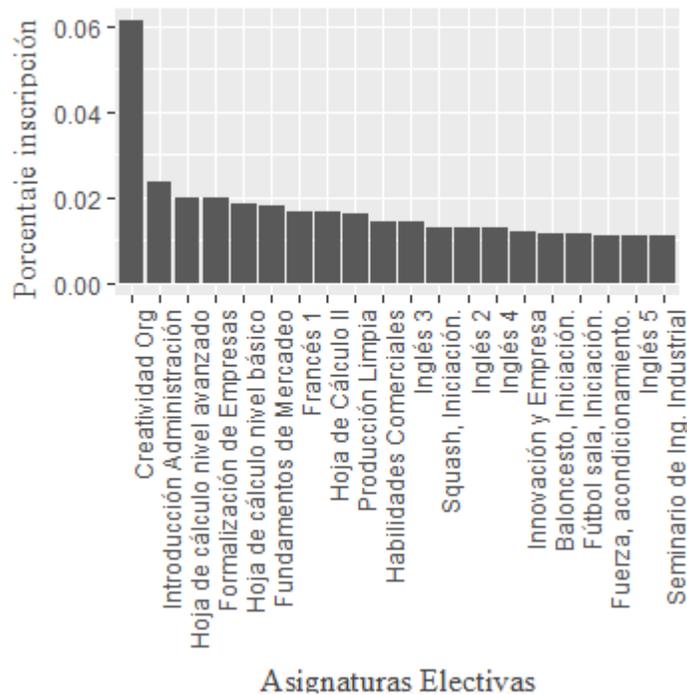


Gráfico 1. Porcentaje de estudiantes inscritos por asignatura.

Asimismo, se pudo establecer que los departamentos de lenguas y diseño industrial, son los más populares entre los estudiantes de ingeniería industrial, con un porcentaje de participación del 17.5% y 17% entre todos los departamentos ofrecidos (Gráfico 2).

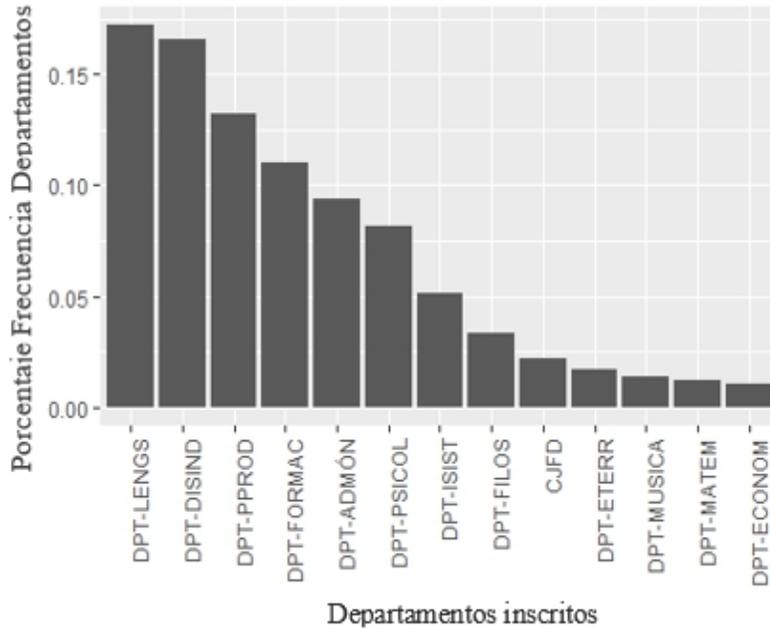


Gráfico 2. Departamentos más populares

De igual forma, se evidenció que las asignaturas Creatividad Organizacional y Hoja de Cálculo Nivel Básico tienen las mejores calificaciones promedio con 4.8 aproximadamente cada una (Gráfico 3).

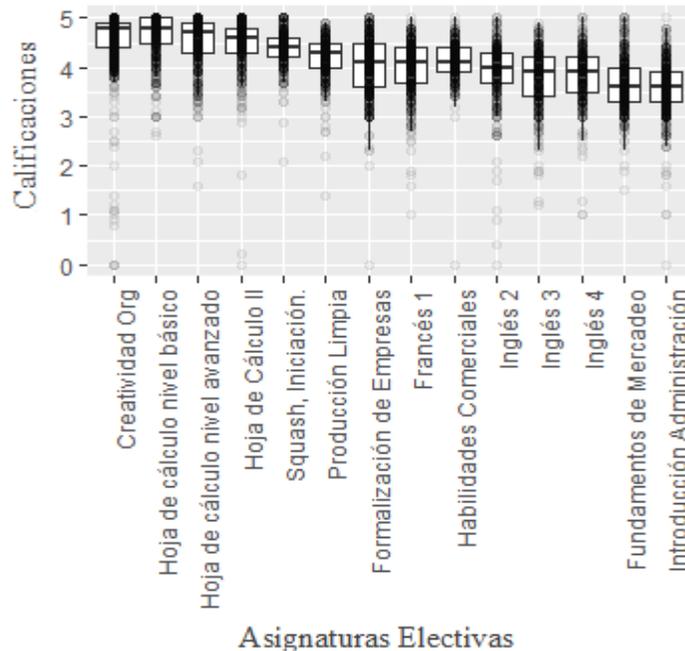


Gráfico 4. Calificaciones por asignatura.

A partir de esto, y con el fin de facilitar el proceso de inscripción de asignaturas electivas a los estudiantes de ingeniería industrial, se propuso en este proyecto el diseño de un aplicativo web que recomiende a los estudiantes asignaturas con base en sus preferencias; se pretende que el aplicativo permita a los estudiantes enfocarse en las asignaturas cuyo contenido sea de su agrado, reduciendo las opciones a la hora de tomar la decisión, disminuyendo

el tiempo de búsqueda y minimizando los niveles de insatisfacción por falta de compatibilidad entre los contenidos de las diferentes asignaturas y los intereses del estudiante.

Metodología

Cross Industry Standard Process for Data Mining o por sus siglas, CRISP-DM es un modelo de proceso estándar para el desarrollo de proyectos de minería de datos, que brinda una perspectiva más cercana a los objetivos reales del proyecto bajo un desarrollo sistemático. Esta metodología cuenta con 6 fases: entendimiento del problema, entendimiento de los datos, preparación de los datos, modelado, evaluación e implementación (Rodríguez, 2010).

Durante el desarrollo de las actividades del proyecto, se realizaron las fases de la metodología CRISP-DM. Para el desarrollo de la fase 1 y 2: entendimiento del problema y los datos, se tuvo que recolectar y analizar información sobre la base de datos académica de ingeniería industrial. Se realizó un análisis descriptivo sobre los patrones de comportamiento de inscripción de asignaturas electivas por medio del análisis descriptivo de la base de datos en R-Studio.

Seguidamente se implementó la fase 3: preparación de los datos. En esta fase se tomó la decisión de depurar la base de datos, para facilitar el manejo de la información y generar una mejor recomendación de las materias. De esta manera, se filtró la base de datos de los estudiantes de ingeniería industrial, seleccionando sólo aquellos que no retiraron la asignatura electiva inscrita y que la aprobaron.

Adicionalmente en esta fase se tomó la decisión de elegir los filtros de recomendación que se implementarían. Para el filtro de recomendación colaborativo, se creó una matriz binaria que incluía las asignaturas electivas y los estudiantes de ingeniería industrial, esta matriz fue usada en la técnica seleccionada para el diseño de este filtro. Para el filtro basado en contenido, por su parte, se preparó la descripción de las asignaturas a partir de la información de syllabus.

Posteriormente en la fase 4: modelado, se tomó la decisión de implementar un sistema de recomendación híbrido, que combinara el filtro de recomendación colaborativo y el filtro basado en contenido. Mediante la construcción de perfiles de usuario según sus intereses, el filtro colaborativo resulta en una recomendación basada en la afinidad de preferencias que se asume entre estudiantes que decidieron inscribir algunas asignaturas comunes en semestres anteriores; esta característica restringe al filtro colaborativo únicamente a estudiantes que tienen historial de inscripción y en consecuencia, no podría ser aplicada para aquellos que aún no han cursado ninguna asignatura electiva previamente.

El filtro de recomendación de contenido, se basa en la descripción del syllabus de las asignaturas electivas. Se compara la similitud entre los términos de las materias vistas por los estudiantes y las asignaturas electivas ofrecidas por la PUJ. Este filtro no tiene gran utilidad cuando se trata de asignaturas cuya descripción es muy reducida o se encuentra incompleta.

Para mitigar las características que se restringen en ambos filtros, se implementó un sistema de recomendación híbrido. A partir de esta combinación se recomiendan asignaturas electivas resultantes de la aplicación del filtro de recomendación colaborativo y basado en contenido.

La fase 5: evaluación, consistía en evaluar las recomendaciones usando los registros de la base de datos académica, para lo cual se aplicaron dos pruebas de evaluación de factibilidad del algoritmo en cada uno de los filtros. En cada prueba, se verificó si las recomendaciones generadas estaban relacionadas con el perfil del estudiante.

Finalmente en la fase 6: implementación, se definieron los parámetros de entrada y salida del aplicativo, se programó la lógica del sistema de recomendación en JavaScript y en Google Services, se llevaron a cabo pruebas de ejecución del aplicativo y se evaluó en el aplicativo web los requerimientos de la norma ISO 9126.

Adicionalmente en la implementación se realizó el lanzamiento del aplicativo, se diseñó y aplicó una encuesta donde se evaluó el aplicativo a través de las métricas Precisión, Serendipia y Diversidad, seleccionadas ya que el aplicativo busca recomendar según las preferencias o historial del usuario; para esto, se deben definir ítems nuevos relevantes para el estudiante, seleccionar ítems de mayor calidad y buscar pluralidad y multiplicidad en los ítems recomendados. Para terminar en esta fase se realizó el análisis de los resultados de la encuesta aplicada a los estudiantes de ingeniería industrial.

Descripción del filtro Colaborativo

La técnica de filtrado colaborativo seleccionada para el sistema de recomendación, se basa en reglas de asociación. Este método se usa principalmente para encontrar patrones frecuentes entre el comportamiento de los usuarios en grandes bases de datos. Fue propuesto por Agrawal en 1993 inicialmente para encontrar relaciones entre los diferentes artículos que compraban las personas en el supermercado, estudio conocido como el *análisis de la cesta de mercado* (DongHwan, Xi, & Jeong, Octubre 2017).

En el filtro colaborativo del sistema de recomendación, se aplicó el algoritmo a priori que genera reglas de asociación a partir de la identificación de conjuntos de ítems frecuentes en bases de datos transaccionales y extrae las asociaciones o co-ocurrencias a partir de esos datos. (Belyi, y otros, Mayo 2016).

Así, las reglas de asociación del sistema de recomendación de asignaturas electivas, buscan predecir aquellas asignaturas electivas que más probablemente le interesarán al estudiante, a partir de su historial de inscripción y los patrones de comportamiento de otros estudiantes con historiales de inscripción (y gustos) similares.

Inicialmente, para la generación de las reglas de asociación, se contó con la base de datos académica que contiene los datos académicos de un poco más de 3.400 estudiantes del programa de ingeniería industrial de la PUJ, desde el primer semestre del 2006 hasta el primer semestre del 2017. Dicha base de datos tiene dos columnas, la primera de ellas contiene el número de identificación único para cada estudiante de la universidad (ID estudiante) y la segunda, contiene el número de identificación único para cada asignatura electiva inscrita por los estudiantes (ID materia inscrita), como se muestra en el fragmento de la misma a continuación:

ID Estudiante	ID Materia Inscrita
10036045	15814
10036045	17615
10021263	7033
10036045	19589
10034775	15814
10021263	1377
10021263	15814
10036045	624
10036045	1377
10018586	17615
10018586	18776
10018586	18638
10018586	18775
10018586	18640

Fragmento de la base de datos académica

Con el fin de garantizar la correcta aplicación del algoritmo a priori, se creó una matriz binaria o matriz de transacciones con los datos de la base de datos académica. En esta matriz, las filas corresponden a cada uno de los IDs de los estudiantes sin repeticiones y las columnas corresponden a cada una de las asignaturas que fueron

inscritas por los estudiantes, sin repeticiones. En las celdas del contenido de la matriz aparece un número uno (1) en el caso que el estudiante de esa fila, haya cursado la asignatura de esa columna, y en caso contrario aparece un cero (0). Por lo tanto al finalizar, obtuvimos una matriz con cerca de 3.400 filas (correspondientes a los estudiantes de la base de datos académica) y cerca de 800 columnas (correspondientes a las asignaturas cursadas por los estudiantes del programa). Para el fragmento de la base de datos académica que se presentó anteriormente, la matriz de transacciones sería la siguiente:

		ID Asignatura electiva									
		15814	17615	7033	19589	1377	624	18776	18638	18775	18640
ID Estudiante	10036045	1	1	0	1	1	1	0	0	0	0
	10021263	1	0	1	0	1	0	0	0	0	0
	10034775	1	0	0	0	0	0	0	0	0	0
	10018586	0	1	0	0	0	0	1	1	1	1

Matriz de transacciones para el fragmento de la base de datos académica

Las reglas de asociación generadas, constan de un antecedente y un consecuente. El antecedente de la regla implica la condición con base en la cual se recomendará o no la asignatura del consecuente. Es decir, a aquellos estudiantes que cursaron todas las asignaturas del antecedente, se les recomendará la asignatura que compone el consecuente de la misma.

El soporte y la confianza de la regla, son dos medidas que definen en gran parte el desempeño y la calidad de las reglas generadas. Es por esto, que cuando se establece un soporte determinado, se está definiendo la proporción de transacciones (para nuestro caso, estudiantes) que deben cumplir con el antecedente de la regla. Por otro lado, al definir una confianza determinada, se está estableciendo la proporción de las transacciones en las que se cumple tanto el antecedente como el consecuente sobre las transacciones en las que se cumple el antecedente (soporte de la regla).

Finalmente, para la generación de las reglas a partir del algoritmo a priori, en primer lugar se construyen los conjuntos formados por sólo un *ítem* o asignatura que superan el soporte mínimo. Posteriormente, estos se utilizan para construir los conjuntos de dos asignaturas, y así sucesivamente hasta que se llegue a un tamaño en el que no existan conjuntos de asignaturas con el soporte requerido. Una vez que se han seleccionado los conjuntos de asignaturas que cumplen con el soporte mínimo, el siguiente paso consiste en generar, a partir de estos conjuntos, las reglas de asociación que tengan un nivel de confianza mínimo. Finalmente, del conjunto de reglas generadas, las que resultan más interesantes son aquellas que tienen su valor de soporte más alto (Wilford, Rosete, & Rodríguez, 2009).

Para el caso de la generación de recomendaciones del filtro colaborativo, y en vista de la diversidad de patrones de comportamiento a la hora de inscribir asignaturas, se estableció un soporte mínimo para las reglas del 1% y una confianza mínima del 25%, es decir, 34 estudiantes tuvieron que cursar las asignaturas del lado izquierdo de la regla y de estos, 9 debieron además, cursar la asignatura del lado derecho para que esa regla se generara y se recomendará a partir de la misma. Se tomaron estos porcentajes debido a la cantidad de reglas que se generaron y el número de materias que se recomendaban.

Basándonos en el ejemplo anterior podemos examinar la siguiente regla de asociación:

1. 15814 → 1377 (Creatividad Organizacional → Introducción a la administración). Esta regla implica que a aquellos estudiantes que cursaron la asignatura 15814 – Creatividad Organizacional (antecedente), se les recomendará la asignatura 1377 – Introducción a la administración (consecuente).

El soporte de esta regla está dado por la proporción de estudiantes que inscribieron la asignatura 15814 – Creatividad Organizacional, sobre el total de estudiantes de la base de datos académica (BD), así:

$$\text{Soporte}_{(15814 \rightarrow 1377)} = \frac{\text{Estudiantes que inscribieron 15814}}{\text{Estudiantes totales de la BD}} = \frac{3}{4} = 75\%$$

Por otro lado, el soporte de esta regla está dado por la proporción de estudiantes que inscribieron tanto la asignatura 15814 – Creatividad Organizacional como la asignatura 1377 – Introducción a la administración, entre la proporción de estudiantes que inscribieron la asignatura 15814.

$$\text{Confianza}_{(15814 \rightarrow 1377)} = \frac{\text{Estudiantes que inscribieron 15814 y 1377}}{\text{Estudiantes que inscribieron 15814}} = \frac{2}{3} = 67\%$$

Finalmente, a partir de esta regla se pueden generar recomendaciones ya que el soporte de 75% supera al soporte establecido del 1%, y la confianza del 67% supera la confianza exigida del 25%.

Otra de las medidas de desempeño de las reglas de asociación y que relaciona tanto al soporte como la confianza, es la “mejora de la confianza” o *lift*, medida que corresponde a la razón entre la confianza y el soporte de la regla, es decir que en la medida que el *lift* aumenta la correlación entre el antecedente y el consecuente de la regla de asociación es mayor (Anping, y otros, 2016)

En el caso del filtro colaborativo del sistema de recomendación de asignaturas electivas, con los parámetros descritos anteriormente se generaron aproximadamente 350 reglas, cada una con su *lift* asociado, a partir del cual se ponderarán las recomendaciones en la hibridación de los filtros

Es importante resaltar, que en algunos casos, el consecuente es el mismo para diferentes reglas, es decir, que en algunos casos más de una regla recomienda la misma asignatura. Por esta razón, cuando un estudiante cumple con más de una regla cuyo consecuente es el mismo, la asignatura se le recomendará una única vez y el *lift* asociado a esta recomendación será el mayor entre los lifts de todas las reglas que el estudiante cumplió que recomendaban la misma asignatura.

Descripción del filtro de contenido

Los filtros basados en contenido tienen al ítem como base de predicción y utilizan características del elemento para realizar la recomendación (Apaza, Vera Cervantes, Cruz Quispe, & Ochoa Luna, 2014). Este filtro analiza las descripciones de cada elemento para identificar cuáles de estos son de particular interés para el usuario (Pazzani & Billsus, 2007). Debido a que en la PUJ no existe información estructurada de las asignaturas electivas permitiendo comprender los atributos de las materias, se decidió extraer temas a partir de los syllabus de las asignaturas electivas.

Los syllabus contienen los resúmenes de las asignaturas que se ofrecen a los estudiantes con la finalidad de que a la hora de inscribir las asignaturas, los estudiantes cuenten con una idea general del contenido de las mismas. Finalmente, con base en los atributos extraídos de las descripciones de las asignaturas, se procede a la medición de la similitud entre asignaturas.

Para facilitar el análisis de la descripción de las asignaturas y generar la recomendación basada en sus atributos, se implementó el algoritmo Latent Dirichlet Allocation (en adelante, LDA). Modelo de bolsa de palabras que representa documentos como mezclas de temas que contienen palabras asociadas (Apaza, Vera Cervantes, Cruz Quispe, & Ochoa Luna, 2014), LDA se utiliza como descriptor de características de los cursos. Cada uno de los documentos, es una mezcla de temas que sigue una distribución multinomial y la aparición de cada palabra en un documento, responde al tema asignado a este, siguiendo una distribución Dirichlet (Griffiths, 2002). En este

modelo, los temas en los documentos están ocultos y las palabras asignadas a aquellos, son variables observadas, es decir latentes. En la figura 1, se puede observar el proceso generativo realizado por LDA para conseguir probabilidades a posteriori a partir de ciertos sucesos a priori (Chandía Sepúlveda, 2016).

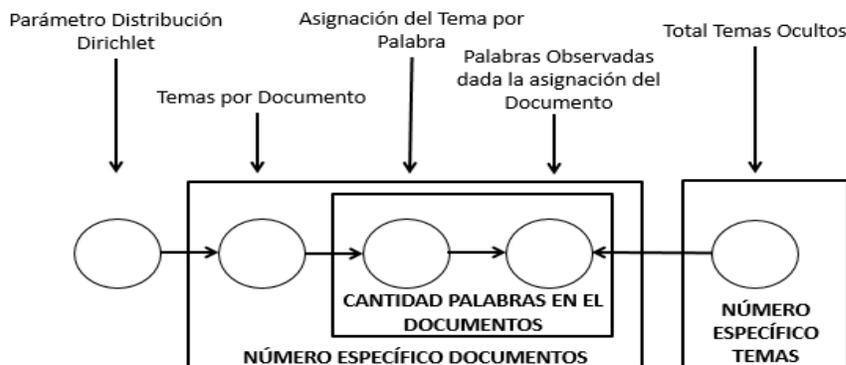


Figura 1. Proceso Modelo LDA. (Chandía Sepúlveda, 2016)

Aunque LDA es un modelo sencillo, no ocurre lo mismo con su inferencia exacta. Una solución para esto, es la aplicación de algoritmos de inferencia aproximada, como lo es el algoritmo Gibbs Sampling, mediante el cual, se puede obtener una secuencia de observaciones que son a su vez, aproximaciones de una distribución de probabilidad multivariable. Así dicha secuencia es usada para aproximar la distribución Dirichlet, empleada para la generación de las palabras en los temas de cada uno de los documentos (Chandía Sepúlveda, 2016).

En la creación de los documentos del modelo, fue importante decidir el número de palabras que tendría cada uno de los temas; en este filtro se eligieron diez palabras por tema. Este número de palabras se seleccionó debido a que para generar la recomendación se requería de una descripción completa de los temas por asignatura y muy pocas palabras, dejaban incompleta la descripción del tema; y un amplio número de estas, alcanzaban a incluir términos que no aportaban conocimiento del tema. El algoritmo LDA solo podría ser aplicado a aquellas asignaturas cuya descripción fuera amplia y completa.

Durante el desarrollo del modelo se presentaron algunos problemas, uno de ellos fue que el modelo LDA tomaba las palabras de los syllabus con mayor frecuencia y muchas de estas no reflejan la descripción de la asignatura, dejando por fuera palabras que reflejan en esencia el contenido de una materia. Para dar solución a estos problemas, se depuraron los textos de los syllabus, removiendo las stopwords, palabras vacías, sin significado, como lo eran artículos, preposiciones, pronombres, algunos verbos, entre otros. Adicionalmente, se removió la puntuación a las palabras, se convirtió el texto en minúscula y finalmente se obtuvo la raíz de las palabras para dejaren el texto únicamente la raíz de los términos.

Un ejemplo de este algoritmo se puede ver con el siguiente syllabus de la asignatura Fundamentos de Mercadeo, obteniendo un tema que describe la materia, junto con los términos que se extrajeron, este procedimiento se realizó para cada uno de las asignaturas electivas ofrecidas por la PUJ.

Syllabus: *“Fundamentos de Mercadeo está diseñado para entregarle a los estudiantes los conocimientos y herramientas básicas de esta disciplina que le permitan contar con una visión de los principales elementos que la constituyen, además de presentar como esta disciplina ha cambiado con el paso del tiempo y como es actualmente.*

Este curso proporcionará los principios fundamentales del Mercadeo a través del estudio y análisis de cómo las diferentes ciencias han afectado esta disciplina. Al finalizar, el estudiante deberá tener una perspectiva crítica que le permita interpretar situaciones propias del mercadeo a partir de sus bases teóricas”.

Debido a que los syllabus contenían palabras de la misma familia, se decidió trabajar con la raíz de estas así

como se ve en el siguiente ejemplo:

Palabra	Raíz de la Palabra
Estudio	Estud
Estudiar	
Estudiante	
Estudiantes	

Ejemplo de la raíz de las palabras de los textos

En vista de que los syllabus de las asignaturas solo se centran en una idea principal, se procedió a obtener un tema por asignatura. Luego para este tema se asignaron 10 términos siguiendo la distribución Dirichlet, donde se buscan las palabras con mayor probabilidad y que mejor describiera el syllabus.

Palabras clave: “disciplin, mercade, permit, basic, cambi, actual, afect, bas, conoc, constitu”.

Para conocer y relacionar las asignaturas electivas con el interés del estudiante, se generó el vector de intereses de cada uno de estos. A partir de las asignaturas electivas vistas de los estudiantes, se obtuvo un conjunto de palabras que las describen. Este vector de palabras de interés, se comparó finalmente, con cada uno de los términos extraídos de los syllabus de las asignaturas electivas.

Para la recomendación de las asignaturas electivas en este filtro, se calculó la similitud de coseno. Medida que se toma para determinar qué tan alineados están los gustos de cada usuario con los atributos de cada elemento y no tener en cuenta la magnitud, es decir no toma como referencia sí un usuario es más exigente que otro o si un documento tiene más palabras que otro (EckoThemes, 2017). Este método establece qué tan relacionado está el conjunto de palabras de cada asignatura electiva con respecto al vector de interés del estudiante.

El cálculo de la similitud de coseno se obtuvo mediante dos vectores. El primer de ellos, contenía la frecuencia de las palabras de las materias que no ha cursado el estudiante en el vector de intereses. Por su parte, con el segundo vector, contenía la frecuencia de las palabras de las materias que no ha cursado el estudiante en los términos asociados a la materia evaluada. Con estos dos vectores se aplicó la fórmula de similitud de coseno, donde X representaba el vector uno y Y representaba el vector dos. Al generar la recomendación se eligieron las 10 asignaturas cuya similitud de coseno fue superior.

Para profundizar un poco más en el tema y entender el funcionamiento de la distancia de coseno, se presenta un ejemplo. Tomamos el ID de un estudiante aleatoriamente, con las asignaturas vistas por el mismo. De esta forma obtenemos el vector de intereses y los términos de las materias no cursadas por el estudiante.

ID: 20010995

Materias vistas por ID: 1805, 7035, 17639, 15814, 105, 93, 106, 134, 142, 221, 318, 17646, 29838.

Se obtiene el Vector de Intereses, por medio de las materias vistas. A estas materias se les implementa el método LDA, obteniendo como resultado los términos de las asignaturas.

mundial	nacional	ambiental	defin	Ecologi	desarroll	ecolog
context	escritur	escuch	ingles	bicultural	cumpl	habl
actitud	alumn	ambit	creativ	profesional	capac	constru

acerqu	aplicacion	obtencion	apropiacion	Barr	bas	bien
emerg	etap	abord	atribut	Ciert	client	gestion
podr	innovacion	cienci	conoc	Proyect	horizont	insatisfech
innov	merc	negoci	basic	elemental	experimental	mediant
tecnic	vitral	articulacion	identific	Segment	brind	lanzamient
argument	clas	cual	comprender	Conten	cultural	disen
audit	autoaprendizaj	line	aprobacion	comprension	credit	academ
comunicacion	solucion	material	dpm	invencion	element	principi
activ	pais	ceram	composit	gener	prove	estrategi
acerc	requisit	product	calid	cas	nich	clasic
activ	pais	ceram	composit	gener	prove	estrategi
virtual	mercade					

Seguidamente se toma una asignatura electiva, para implementarle el algoritmo LDA, posteriormente se toman los términos de la materia, para finalmente aplicar el método de similitud de coseno.

Id Asignatura 1899: comunicacion, desarroll, cas, enmarc, strategi, cambi, planeación, basic, concret, context

Se aplica el método de similitud de coseno, donde X representaba el vector uno y Y el vector dos. Para este caso específico, la similitud de coseno se calcularía así:

$$\text{Similitud Coseno} = \frac{\sum XY}{\sqrt{(\sum X^2 * \sum Y^2)}} = \frac{11}{38.729} = 0.28$$

$$\text{Similitud Coseno} = 0.284$$

Finalmente de los resultados obtenidos por cada una de las asignaturas electivas, se seleccionaron las diez materias con similitud de coseno superior para generar la recomendación.

Es importante tener en cuenta a aquellos estudiantes que inscribirán asignaturas electivas por primera vez. Para estos estudiantes se hizo un procedimiento distinto, se implementó LDA para todas las asignaturas electivas y no para cada una de las asignaturas vistas, ya que en estos casos los estudiantes no tienen historial de inscripción. Se asignaron cuarenta temas y veinte términos para la creación del documento. Obteniendo esto se seleccionaron solo veinte temas y diez términos por tema, los que mejor describieran las asignaturas. De esta manera los estudiantes pueden elegir algunos temas de interés, de modo que el conjunto de palabras asociadas a los temas seleccionados se convertiría en el vector de intereses de los estudiantes. Con los términos del vector de intereses y los términos de las asignaturas electivas se dispone a realizar similitud de coseno para generar la recomendación.

Filtro Híbrido

Un sistema de recomendación híbrido es aquel que combina diferentes sistemas de recomendación para producir una sola salida con el objetivo de que se complementen con sus mejores características y generen mejores recomendaciones (Burke, 2002). Para el caso del Recomendador de Asignaturas Electivas (RAE, en adelante),

donde se establecieron 10 asignaturas para la recomendación final a entregar a los estudiantes, se seleccionaron dos métodos a partir de los cuales se hibridarían los resultados del filtro colaborativo y de contenido, como se ilustra en la **Figura 2**.

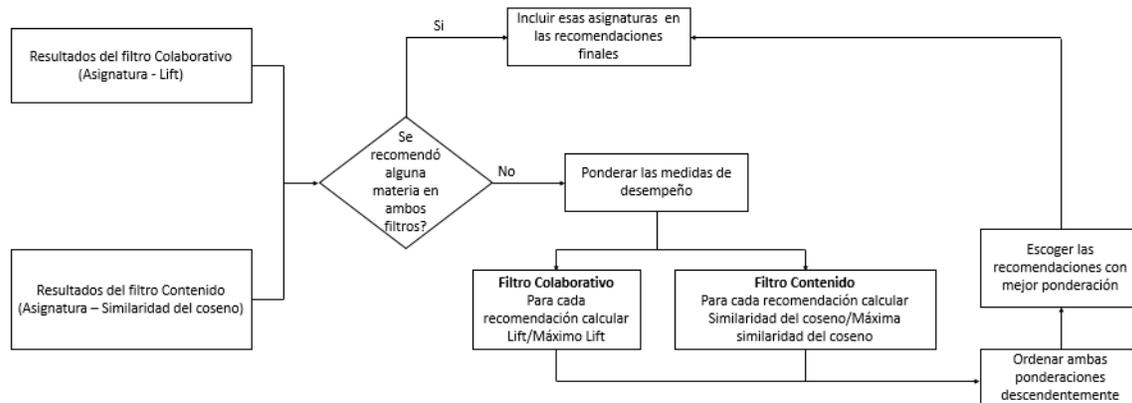


Figura 2. Hibridación de filtros

Inicialmente, se estableció una hibridación bajo el método intersección de resultados (Rodríguez Marín, Duque, & Ovalle, 2016), en el que las primeras recomendaciones que se harían al estudiante, serían aquellas que fueran resultado de la recomendación de ambos filtros, es decir, las asignaturas recomendadas en las primeras posiciones son las asignaturas que se recomendaron tanto en el filtro colaborativo como en el de contenido, en caso de que esta situación ocurra.

Las recomendaciones restantes se establecieron bajo el método ponderado (Rodríguez Marín, Duque, & Ovalle, 2016). Esta ponderación, en el caso del filtro colaborativo se realizó en función del lift de cada una de las reglas y en el filtro basado en contenido, se realizó en función de la similitud del coseno.

Para cada recomendación del filtro colaborativo, se le asignó una puntuación de 0 a 100%, resultante del cociente entre su lift y el mayor de los lifts de todas las recomendaciones del filtro para el estudiante.

Por su parte, para cada recomendación del filtro de contenido, se le asignó una puntuación de 0 a 100%, resultante del cociente entre su similitud del coseno y la mayor de las similitudes es del coseno de todas las recomendaciones del filtro para el estudiante.

Finalmente, se combinaron las recomendaciones de ambos filtros de forma descendente con base en las puntuaciones ponderadas anteriormente.

Experimentos computacionales

Para garantizar el cumplimiento de los requerimientos de desempeño, se realizó una serie de pruebas que permitieron verificar la factibilidad del aplicativo. El objetivo fue ver qué tan bueno era el sistema de recomendación antes de enviarlo a los estudiantes.

La primera prueba se realizó en el filtro colaborativo, basado en contenido y en el híbrido. Consistió en escoger 300 estudiantes aleatoriamente y obtenerlas materias vistas por estos y; a cada uno de ellos se le eliminó dos asignaturas. Posteriormente, se obtuvo el vector de intereses para las materias vistas, sin tener en cuenta las eliminadas. Finalmente se generó la recomendación para estas asignaturas, donde se calculó cuantas de las materias eliminadas fueron recomendadas, si fueron 0, 1 o 2 asignaturas de las eliminadas anteriormente.

En esta prueba se encontró que de los 300 estudiantes aleatoriamente seleccionados, el filtro colaborativo les recomendó a 10 estudiantes, las dos materias eliminadas; a 115 estudiantes, les recomendó una de las materias

eliminadas y a 175 estudiantes, no les recomendó ninguna de estas.

En el filtro basado en contenido, por su parte, se obtuvo que a 15 estudiantes, les recomendó las dos materias eliminadas; a 132 estudiantes, les recomendó una de las materias eliminadas y a 153 estudiantes, no les recomendó ninguna de las materias eliminadas.

Finalmente, en el sistema de recomendación híbrido se obtuvo que a 19 estudiantes, les recomendó las dos materias eliminadas; a 144 estudiantes, les recomendó una de las materias eliminadas y a 137 estudiantes, no les recomendó ninguna de las materias eliminadas.

La segunda prueba que se realizó consistió en escoger 300 estudiantes aleatoriamente y obtener las materias vistas por cada uno de estos. Seguidamente se calculó el vector de intereses para las materias vistas y se generó la recomendación para estas materias. Esta prueba se realizó en el filtro colaborativo, basado en contenido y en el híbrido, registrando la cantidad de materias diferentes que recomendó en cada filtro y el porcentaje de materias diferentes recomendadas sobre el total de materias recomendadas, con el fin de probar la métrica de diversidad del sistema.

En esta segunda prueba se encontraron los siguientes resultados: el filtro colaborativo de los trecientos IDs aleatorios, recomendó 21 materias diferentes, representando un 24% del total de materias obtenidas de la base de datos de estudiantes de ingeniería industrial. Con el filtro basado en contenido, se obtuvo un resultado de 896 materias recomendadas, representando un 32% del total de materias ofrecidas por la PUJ. Finalmente, el filtro de recomendación híbrido, recomendó 905 materias diferentes, representando un 28% del total de materias electivas. Estos resultados permiten ver que para recomendar gran diversidad de materias se requiere del filtro colaborativo ya que abarca todas las asignaturas electivas que ofrece la universidad.

Para obtener el cálculo del porcentaje de materias recomendadas diferentes en esta prueba, se tuvo en cuenta que el total de registros de asignaturas electivas es distinto por filtro. En el filtro colaborativo se tomó como referencia las materias registradas en la base de datos de los estudiantes de ingeniería industrial, en el filtro basado en contenido todas aquellas asignaturas electivas que ofrece la PUJ y en el filtro mixto los registros de ambos filtros.

Aplicativo Web



La interfaz con la que tuvieron contacto los estudiantes, se diseñó en un formulario de Google con la menor cantidad de requerimientos posibles de modo que tomara unos pocos segundos para los estudiantes ingresar sus datos, y de esta forma, recibir las recomendaciones en la menor brevedad posible (máximo 3 minutos).

El aplicativo requiere únicamente que el estudiante ingrese su número único de identificación (ID), su correo electrónico institucional (donde recibirá el archivo con las recomendaciones finales) y seleccione una de las dos opciones para la generación de recomendaciones:

- Generar recomendaciones basadas en su historial académico, lo que significa que se correrá el sistema de recomendación híbrido bajo la lógica explicada en los apartados anteriores.
- Generar recomendaciones basadas en temas de su interés, lo que significa que el estudiante escoge entre 20 alternativas de temas que comprenden biología, música, filosofía, política, antropología, sistemas, salud, arte, derecho, economía, tecnología, teología, literatura, ecología, ética, lenguas, educación, electrónica, construcción y diseño con base en las que se generarán sus recomendaciones.

Estos temas se seleccionaron después de aplicar LDA a todas las asignaturas y agrupar los términos por tema.

El siguiente es un pantallazo de la primera parte de la interfaz web. Esta parte es la única para aquellos estudiantes que escojan la opción basada en el historial de asignaturas electivas.

Aplicativo RAE

Recomendador de Asignaturas Electivas

***Required**

ID *

Your answer

Correo electrónico institucional *

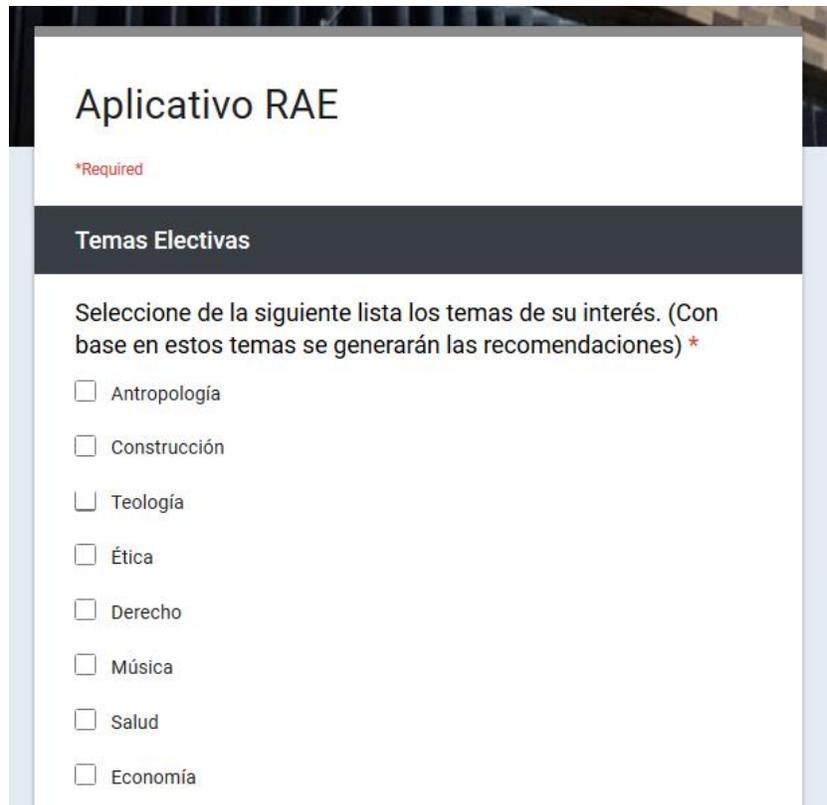
Your answer

¿Cómo desea que generemos sus recomendaciones? *

Con base en temas de su gusto, sin tener en cuenta su historial de materias electivas inscritas. (Opción Obligatoria NEOS)

Teniendo en cuenta su historial de materias electivas inscritas

Por otro lado, para el caso de los estudiantes que desean seleccionar temas de su interés, la interfaz donde se presentan las alternativas de los temas, es la siguiente:



The screenshot shows a web application interface titled 'Aplicativo RAE'. Below the title, there is a red asterisk and the word 'Required'. A dark grey header bar contains the text 'Temas Electivas'. Below this, a white box contains the instruction: 'Seleccione de la siguiente lista los temas de su interés. (Con base en estos temas se generarán las recomendaciones) *'. A list of eight topics follows, each with an unchecked checkbox: Antropología, Construcción, Teología, Ética, Derecho, Música, Salud, and Economía.

El link para ingresar al aplicativo web es el siguiente: <https://goo.gl/ktS89y>. Para llevar a cabo pruebas sobre el mismo, se puede usar el ID de la estudiante María Camila Villasmil 20079123.

Resultados

El aplicativo RAE fue utilizado por 304 estudiantes del programa de Ingeniería Industrial de la PUJ, de los que al 62% se le generaron recomendaciones basadas en temas de su gusto, sin tener en cuenta su historial de materias electivas inscritas (opción obligatoria para los estudiantes de primer semestre), mientras que al 38% restante se le generaron recomendaciones basadas en su historial académico de inscripción.

De las recomendaciones generadas a partir del filtro híbrido, aproximadamente en el 20% de los casos, se recomendó al menos una asignatura resultante tanto del filtro de contenido como del filtro colaborativo. Asimismo, del total de las asignaturas recomendadas el 85% resultó únicamente del filtro de contenido y el 15% restante resultó únicamente del filtro colaborativo.

Por otra parte, los temas más populares entre los estudiantes a quienes se les recomendó según los temas seleccionados fueron lenguas con el 12% de participación, economía y diseño con el 10% de participación cada una y artes y tecnología con el 9% de participación cada una.

Finalmente, entre los resultados del aplicativo, se evidenció que con un 15% de participación la asignatura que más se recomendó fue Comunicación y Ciudad Digital, seguida de Creatividad Organizacional con el 13% de participación.

Junto con el enlace del aplicativo, se envió a los estudiantes una encuesta que buscaba medir el aplicativo a la luz de las métricas establecidas. A partir de esta encuesta, se evidenció que para el 75% de los estudiantes las recomendaciones reflejaban sus gustos y preferencias, lo que demuestra la precisión de las recomendaciones

generadas por el aplicativo. Adicionalmente el 67% de los estudiantes manifestó no tener conocimiento de las asignaturas electivas recomendadas, lo que refleja que el aplicativo da a conocer a los estudiantes opciones completamente desconocidas y que pueden ser un gran agregado en su proceso de formación integral. Seguidamente se pudo observar que un 75% de los estudiantes que respondieron la encuesta consideraron como inesperadas algunas de las asignaturas recomendadas, evidenciando el nivel de serendipia o sorpresa generado por las recomendaciones.

Finalmente, el 70% de los estudiantes manifestó que inscribiría las asignaturas electivas recomendadas por el aplicativo, si estas llegaran a concordar con su horario de clase, lo que comprueba la eficacia del aplicativo RAE a la hora de satisfacer la necesidad que dio origen a este proyecto: estudiantes cada vez más críticos, analíticos, con una visión holística e integradores de conocimiento, a través de los que se proyectará el destino de nuestra sociedad.

1. Anexos

No. Anexo	Nombre	Desarrollo	Tipo de Archivo	Enlace Corto (https://goo.gl/)	Relevancia para el documento (1-5)
1	Código Aplicativo Web Java	Propio	Bloc Notas	https://drive.google.com/open?id=14SktHbUdveli8TrkKleRl3rMYbfv6kjU	5
2	Reglas de Asociación Filtro Colaborativo	Propio	Excel	https://drive.google.com/open?id=1euvJzrKpFfpguG-_nerCb_S7bxaPB56v	4
3	Base de Datos Académica PUJ	Universidad Javeriana	Excel	https://drive.google.com/open?id=1QqLe1UshxeiJXl9bHEuZ3zmd79sZffd1	3
4	Base de Datos Otros Departamentos PUJ	Universidad Javeriana	Excel	https://drive.google.com/open?id=1IQ7NECzojAEDIqeTOvbVIQ1gog7_evli	2
5	Resultado Filtro Basado en Contenido	Propio	Excel	https://drive.google.com/open?id=1iExwo88FVw56YczPFMuwZfwdJyhBV8Ib	4
6	LDA Neos - Materias Agrupadas	Propio	Excel	https://drive.google.com/open?id=1ZXtWx6mRgf84iy5OtK42Xdkrg7lLsLa	4
7	Asignaturas Programadas en el Periodo 1810	Universidad Javeriana	Excel	https://drive.google.com/open?id=1VTfhYhlpng8cQz6BPfqp36By43_Z0zW	2
8	Materias de Núcleo Común Ingeniería Industrial	Universidad Javeriana	Excel	https://drive.google.com/open?id=1SQ5B9zqHQlr7LK2y_XJgGOFK_yraBVok	3

Bibliografía

Acıar, S., Zhang, D., Simoff, S., & Debenham, J. (2006). Mining the real-time web: A novel approach to product recommendation. *International conference on Web Intelligence and Intelligent Agent Technology*, 37-40.

- Ackerman, P. L. (1998). Determinants of Individual Differences During Skill Acquisition: Cognitive Abilities and Information Processing. *Journal of Experimental Psychology*, 288-318.
- Altuna Castillo, E. J., Guibert Estrada, L., & Estrada, V. (2014). Sistema de recomendación basado en k-nn para condiciones de incertidumbre en un Sistema Tutor Inteligente. *Ciencias de la Información Vol. 45, No. 3*, 25-30.
- Anping, S., Xuehai, D., Jianjiao, C., Mingbo, L., Wei, C., & Ke, P. (2016). Multi-objective association rule mining with binary bat algorithm. *Intelligent Data Analysis Vol 20*, 105-128.
- Apaza, R. G., Vera Cervantes, E., Cruz Quispe, L., & Ochoa Luna, J. (2014). Online Courses Recommendation based on LDA. *National University of St. Agustin*, 42-48.
- Becerra Cortés, C., Jimenez, S., & Gonzalez, F. (2015). Products recommendation based on interpretable user profiles. *Tecnura 19(45)*, 89-100.
- Belyi, E., giabanelli, p., patel, i., Baladabhadrapathurni, N., Abdallah, a., hameed, w., & mago, v. (Mayo 2016). Combining association rule mining and network analysis for pharmacosurveillance. *Journal of supercomputing* , 2014-2034.
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *Volume 12, Issue 4*, 331–370.
- Calvo, J. (1 de Diciembre de 2016). *Topic modeling: ¿qué, cómo, cuándo?* Obtenido de morethanbooks : <http://www.morethanbooks.eu/topic-modeling-introduccion/>
- Carvajal Escobar, Y. (2010). Interdisciplinarity: A Challenge For Higher Education and Research. . *Luna Azul*.
- Casali, A., Gerling, V., Deco, C., & Bender, C. (2010). Sistema inteligente para la recomendación de objetos de aprendizaje.
- Chandía Sepúlveda, B. (2016). Aplicación y evaluación LDA para asignación de tópicos en datos de twitter . *PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO FACULTAD DE INGENIERÍA ESCUELA DE INGENIERÍA INFORMÁTICA* , 20-25.
- Collado Sánchez, A. (2014). Sistema de recomendación de recursos basado en filtrado colaborativo para la plataforma edX. *Universidad Carlos III de Madrid*.
- Domingues, M. A., Sundermann, C. V., & Garcia Manzato, M. (2014). Exploiting Text Mining Techniques for Contextual Recommendations. *Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*.
- DongHwan, C., Xi, W., & Jeong, W. K. (Octubre 2017). Assessing lightning and wildfire hazard by land properties and cloud to ground lightning data with association rule mining in alberta canada . *Sensors*, 1-32.
- Duran Dorado, D. F., & Arciniega Herrera, J. L. (2013). Hybrid-Mixed Filtering Algorithm for Audiovisual Content Recommendation to Virtual Communities. *Scientia et Technica Año XVIII, Vol. 18, No. 1*.
- EckoThemes. (2017). *Sistemas de recomendación basados en contenido (Content-Based)*. Obtenido de <http://www.p.valienteverde.com/sistemas-de-recomendacion-basados-en-el-contenido-content-based/>
- Galán Nieto, S. M. (2007). Filtrado Colaborativo y Sistemas de Recomendación. *Inteligencia en Redes de Comunicaciones*.
- Gao, F., Xing, C., Du, X., & Wang, S. (2007). Personalized Service System Based on Hybrid Filtering. *TSINGHUA SCIENCE AND TECHNOLOGY Volume 12, Number 1*, 1-8.

- Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text Mining for Product Attribute Extraction. *SIGKDD Explorations Volume 8, Issue 1*, 41- 48.
- Ghauth, K. I., & Abdullah, N. A. (2010). Learning materials recommendation using good learners' ratings and content-based filtering. *Educational Technology Research and Development Volume 58, Issue 6*, 711-727.
- Gomez Verdejo, V. (7 de Noviembre de 2012). *Sistema de Recomendación: Filtro Colaborativo*. Obtenido de <http://www.tsc.uc3m.es/~miguel/MLG/adjuntos/VanessaRecommender.pdf>
- gonzález, A. (19 de Septiembre de 2014). *Sistemas de recomendación de contenido con Machine Learning*. Obtenido de Clever Data: <http://cleverdata.io/sistemas-recomendacion-machine-%20learning/>
- Griffiths, T. (2002). *Gibbs sampling in the generative model of Latent Dirichlet Allocation*. Obtenido de citeseerx: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2E5FA3EFCF6CC36D7B10DBA8EB2B352B?doi=10.1.1.138.3760&rep=rep1&type=pdf>
- Guerra, Y., Mórtigo Rubio, A. M., & Berdugo Silva, N. C. (2013). Formación Integral: La Importancia de Formar Pensando En Todas Las Dimensiones Del Ser. . *Educación y Desarrollo Social*.
- Gunawardana, A., & Shani, G. (2009). A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research 10*, 2935-2962.
- Itmazi, J. (2005). Sistema flexible de gestión del e-learning para soportar el aprendizaje en universidades tradicionales y abiertas. *Dialnet*.
- Javeriana, P. U. (2003). *Reglamento de Unidades Académicas*. . Bogotá D.C.
- Jiang, W., & Yang, L. (2016). Research of improved recommendation algorithm based on collaborative filtering and content prediction. *11th International Conference on Computer Science & Education (ICCSE)*, 598-602.
- Kuang, W., Nianlong, L., & Sun, Z. (2011). Resource Recommendation Based on Topic Model. *The work was supported by the National High Technology Research and Development Program of China*, 370-374.
- Leiva, J. L., Guevara, A., Rossi, C., & Aguayo, A. (2014). Realidad Aumentada y Sistemas de Recomendación Grupales. *Estudios y Perspectivas en Turismo, Vol. 23*, 40-59.
- Li, Y., Nie, J., Zhang, Y., Wang, B., Yan, B., & Weng, F. (2010). Contextual recommendation based on text mining. *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 692-700.
- Pazzani, M., & Billsus, D. (2007). Content-Based Recommendation Systems. *Rutgers University, ASBIII, 3 Rutgers Plaza*.
- Pinho Lucas, J. (Octubre de 2010). Métodos de clasificación basados en asociación aplicados a sistemas de recomendación. *Universidad de Salamanca Departamento de Informática y Automática*, 1-237.
- Rodríguez Buitrago, O. (2012). Evaluación del Sistema de Recomendación de Patrones Pedagógicos (SRPP) en cursos de Geometría Euclidiana. *Tesis de Maestría Universidad del Cauca*.
- Rodríguez Marín, P. A., Duque, N., & Ovalle, D. (2016). Multi-agent system for Knowledge-based recommendation of Learning Objects. *Revista Virtual Universidad Católica del Norte No. 40*.
- Rodriguez, O. (2010). Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM.

Ruiz Iniesta, A. (2011). Sistemas de recomendación: Presente y Futuro en la web. *Universidad Complutense de Madrid. Madrid, España.*

Shani, G., & Gunawardana, A. (s.f.). Evaluating Recommendation Systems.

Terry González, Y., Estrada Sentí, V., & Arteaga Gómez, Y. (2016). System for recommending of open educational resources design patterns. *Ciencias de la Información Vol. 47, No. 1*, 3-8.

What is a good explanation of Latent Dirichlet Allocation? (9 de Junio de 2017). Obtenido de <https://www.quora.com/WhatisagoodexplanationofLatentDirichletAllocation>

Woerndl, W., Huebner, J., Bader, R., & Gallego-Vico, D. (19 de Septiembre 2013). Proactivity and context-awareness: Future of recommender systems design. *Universidad Politécnica de Madrid.*

Yali. (Octubre de 2017). *Market Basket Analysis: identifying products and content that go well together.* Obtenido de For Data Modelers & Consumers: <https://discourse.snowplowanalytics.com/c/data-modelers>