

RECOMMENDER SYSTEM OF UNIVERSITY PROGRAMS FOR THE PROFESSIONAL ORIENTATION OF HIGH SCHOOL STUDENTS

SISTEMA DE RECOMENDACIÓN DE PROGRAMAS UNIVERSITARIOS PARA LA ORIENTACIÓN PROFESIONAL DE ESTUDIANTES DE EDUCACIÓN MEDIA

MSc. Johanna Alexandra Orozco Cacique, PhD. Andrés Moreno Barbosa

Universidad de los Andes,
Facultad de Ingeniería
Cra 1 N° 18A – 12, Bogotá, Colombia.
Tel.: 57-1-3394999, 57-1-3394949
E-mail: {ja.orozco, dar-more}@uniandes.edu.co

Abstract: High school students facing the selection of a university program usually use web search engines, published program's information, psychological testing, and vocational reports. Although useful, these traditional resources fail to consider a data-driven decision-making approach taking into account the students' academic and socio-demographic profile. One data-driven decision-making tool used to help students select a university program are Academic Recommender Systems (ARS). Current systems fail to consider three main important factors in conjunction: consider the performance of university students, provide an explanations model justifying recommendations, and validate their work with ranking metrics. This paper presents a proposal to support the selection of a university program based on a hybrid RS which combines user-user collaborative and content-based filtering. Besides this, an explanations model is incorporated to justify the provided recommendations.

Keywords: Recommender Systems, decision-making, university program, university, explanations model.

Resumen: Los estudiantes de educación media que se enfrentan a la selección de programas universitarios usualmente usan buscadores web, información de programas y pruebas vocacionales. Si bien son útiles, estos recursos tradicionalmente no habilitan la toma de decisiones basada en datos objetivos que tengan en cuenta el perfil académico y sociodemográfico. Una herramienta de toma de decisiones basada en datos que ayuda a los estudiantes a seleccionar un programa universitario es Sistemas de Recomendación Académica (ARS). Los sistemas actuales no tienen en cuenta tres factores importantes en conjunto: considerar el rendimiento de los estudiantes universitarios, proporcionar un modelo de explicaciones que justifique las recomendaciones y validar su trabajo con métricas de *ranking*. Este documento presenta una propuesta para apoyar la selección de un programa universitario basado en un RS híbrido que combina Filtrado Colaborativo usuario-usuario y Filtrado Basado en Contenido. Además, incorpora un modelo de explicaciones para justificar las recomendaciones generadas.

Palabras clave: Sistemas de Recomendación, toma de decisiones, programas académicos, universidad, modelo de explicaciones.

1. INTRODUCCION

La educación es un asunto de interés general en todo el mundo, porque está evidenciado que el

crecimiento y desarrollo económico de un país están relacionados directamente con el hecho de que sus habitantes tengan acceso a la educación y sea de calidad (Vicentini, 2016). Por esto muchos países, ya sean o no desarrollados, enfocan recursos del estado para llegar a más personas en los diferentes niveles de su trayectoria educativa y para que dicha educación sea considerada adecuada, de acuerdo con estándares internacionales.

En el año 2012, el Ministerio de Educación Nacional, hizo un análisis de brechas en la educación (Ministerio de Educación Nacional, 2012), descubriendo que la deserción de estudiantes de educación superior está estrechamente ligada a los procesos de orientación vocacional y profesional. Por tal motivo, creó un portafolio de estrategias para reducir la deserción, orientado a promover la articulación de los estudiantes de la educación media con la oferta académica y las características de la educación superior. De esta forma, los estudiantes contarán con un mayor acervo de información para tomar mejores decisiones sobre los programas académicos de su interés. Este es un tema de gran relevancia para las universidades, por esto hacen seguimiento al problema de abandono de la educación superior, calculan el riesgo de deserción y evalúan cada una de las situaciones que influyen en la deserción, tales como la situación del estudiante, el programa académico y la institución. (Perez *et al.*, 2018).

En Colombia existen diferentes opciones para apoyar a los estudiantes de educación media en la decisión de seleccionar un programa de educación superior. Sin embargo, estas se basan solo en la centralización de información de las universidades y sus programas académicos y en la aplicación de pruebas o test vocacionales, que muestran al estudiante como podrían orientar su vida profesional solo atendiendo a un aspecto de su ser, el psicológico. A partir de lo anteriormente expuesto, este trabajo busca ofrecer una alternativa complementaria, mediante el uso de sistemas de recomendación, enfocada en el acceso a la educación universitaria, dando la posibilidad de apoyar la selección de su programa académico a través del análisis de elementos como su perfil académico y socioeconómico fusionándolos con información disponible en otras fuentes.

Los Sistemas de Recomendación (RS) son un conjunto de herramientas de *software* y técnicas que proporcionan sugerencias para temas que son de utilidad para un usuario. Las sugerencias se refieren a diversos procesos de toma de decisiones, tales como qué elementos comprar o

qué música escuchar, entre otros. El desarrollo de un RS es un esfuerzo multidisciplinario que involucra a expertos de diversos campos, tales como inteligencia artificial, interacción con la computadora humana, minería de datos, estadísticas, sistemas de apoyo a las decisiones, marketing y comportamiento del consumidor (Ricci *et al.*, 2015). Debido a que la problemática de este trabajo está asociada a la toma de decisiones de estudiantes, es adecuado orientar la solución a este tipo de sistemas, equilibrando las necesidades y propósitos de los estudiantes con la implementación de métodos, algoritmos, ecuaciones y aspectos técnicos desarrollados para presentar una solución adecuada, tal y como se plantea en el *framework* propuesto por Dietmar Jannach y Gediminas Adomavicius (Jannach y Adomavicius, 2016).

Este artículo tiene el propósito de apoyar a los estudiantes de educación media en tomar la decisión de seleccionar un programa universitario para continuar sus estudios, a través de un prototipo funcional que implementa un modelo de recomendación que consta de la hibridación de dos recomendadores; el primero de Filtrado Colaborativo (CF) usuario-usuario y el segundo centrado en Filtrado Basado en Contenido (CB) a través del procesamiento de lenguaje natural. Adicionalmente, se incluye un modelo de explicaciones centrado en la necesidad de justificar a los estudiantes las recomendaciones generadas, de manera que sean comprendidas sin la intervención de alguna persona experta en orientación vocacional/profesional. También, se incluyó la validación off-line a través de métricas de ranking, no evidenciada en los trabajos revisados como parte del estado del arte.

Es de resaltar que los resultados obtenidos son de interés tanto para los estudiantes como para sus familias quienes están directamente involucrados. Asimismo, para los asesores, psicólogos o psicorrientadores de colegios quienes dan apoyo a los estudiantes y las Instituciones de Educación Superior quienes se ven afectados con la deserción de sus estudiantes debido a la toma de decisiones equivocada. También, para los empleadores y empresarios quienes buscan mano de obra calificada para vincular laboralmente. Por último, para el Ministerio de Educación Nacional y el Gobierno de Colombia.

El artículo continúa en la sección 2 con la revisión de estado del arte. La sección 3 incluye la estrategia de solución. La sección 4 muestra la evaluación del modelo y finalmente en la sección 5 se concluye este trabajo.

2. ESTADO DEL ARTE

El estado del arte permite descubrir los avances académicos relacionados al campo de estudio objeto de la investigación, para identificar problemas relevantes a resolver. En este sentido, la estrategia definida está orientada hacia la búsqueda de trabajos de educación o académicos que apoyen la toma de decisiones de estudiantes, principalmente, en la selección de materias, programas o universidades y que su enfoque sea hacia RS, minería de datos o sistemas o aprendizaje de máquina.

En este documento se incluyen sólo artículos relevantes encontrados a partir de dos métodos de búsqueda. El primer método es el uso de las siguientes palabras clave: *system recommender vocational orientation*. Con este método se obtuvieron 474 artículos. El segundo método incluye algunas búsquedas sistemáticas actualizadas como (Rivera *et al.*, 2018), (Iatrellis *et al.*, 2017), (Peña-Ayala, 2014), las cuales permitieron validar la importancia de los artículos a analizar.

Los trabajos filtrados debían cumplir con los siguientes criterios de selección:

- Corresponder a RS que ayudan a estudiantes que están buscando orientación vocacional y/o profesional a mejorar su selección de materias, programas y universidades.
- Estar incluidos en *Scopus*, base de datos mundial de bibliografía de investigación revisada por pares, en campos de la ciencia, la tecnología, la medicina, las ciencias sociales y las artes y humanidades.

De esta manera, se incluyeron 16 artículos académicos que cumplieron con los criterios establecidos, además de ser relevantes para la problemática de investigación que se aborda en este documento. Los trabajos (Castellano *et al.*, 2011), (Unelsrød, 2011), (Maridueña-Arroyave *et al.*, 2016) fueron encontrados en las búsquedas sistemáticas mencionadas anteriormente y a través de *Scopus* están los restantes (Vialardi *et al.*, 2009), (Ognjanovic *et al.*, 2016), (Cho y Kang, 2010), (Pinto *et al.*, 2016), (Maridueña-Arroyave y Febles-Estrada, 2016), (Ragab *et al.*, 2012), (Iyengar *et al.*, 2017), (Bokde *et al.*, 2016), (Hu *et al.*, 2017), (Meryem *et al.*, 2016), (Xu *et al.*, 2016) y (Farzan y Brusilovsky, 2006).

2.1 Recomendación de materias

Los trabajos que tienen como objetivo recomendar materias a los estudiantes se pueden visualizar en la Tabla 1. Están divididos en estrategias de contenido, conocimiento y recomendación híbrida. En general, muestran soluciones diferentes para la selección de

materias, sin embargo, los trabajos híbridos en su mayoría presentan un listado Top-N de las recomendaciones.

Tabla 1: Búsqueda de materias

Tarea	Métrica computacional	Documento
Basado en Contenido (Content-Based)	Accuracy, Precision, recall.	(Vialardi <i>et al.</i> , 2009)
	Accuracy, RMSE.	(Ognjanovic <i>et al.</i> , 2016)
	No se evidencia.	(Farzan y Brusilovsky, 2006)
Basados en Conocimiento (Knowledge-Based)	Minimizar el tiempo de graduación maximizando el desempeño del GPA.	(Xu <i>et al.</i> , 2016)
Sistemas de recomendación híbridos (Hybrid)	Función de utilidad para ponderar lista de recomendaciones.	(Castellano <i>et al.</i> , 2011)
	Hit-ratios, Precision-ratios y Recall-ratios	(Cho y Kang, 2010)
	Precision, Recall	(Unelsrød, 2011)
	No se evidencia.	(Ragab <i>et al.</i> , 2012)

Aunque tienen un objetivo común, los algoritmos usados y las métricas evaluadas son diferentes para cada uno de ellos. Esto puede ser debido a la diversidad de información que se tiene como insumo para el entrenamiento y pruebas de los modelos, así como las estrategias definidas por los autores de cada trabajo para generar las predicciones. Además, se pueden revisar diversas alternativas para apoyar la toma de decisiones, unas basadas en información personal histórica de estudiantes y otras basadas en información propia de los programas, así como la validación de resultados presentados a través de listas Top-N.

2.2 Recomendación de universidades

Estos trabajos recomiendan universidades, aplicando estrategias de CF, CB y Sistemas Híbridos, pero en la mayoría no hay evidencia de cómo se midió la eficiencia de los modelos.

Tabla 2: Búsqueda de Universidades

Tarea	Métrica computacional	Documento
Basado en Filtrado Colaborativo (Collaborative Filtering-Based)	No se evidencia	(Iyengar <i>et al.</i> , 2017)
	Accuracy, Precision, Recall, F1 Metric.	(Bokde <i>et al.</i> , 2016)
Basado en Contenido (Content-Based)	No se evidencia	(Hu <i>et al.</i> , 2017)
Sistemas de recomendación híbridos (Hybrid)	No se evidencia	(Ragab <i>et al.</i> , 2012)

Para recomendar universidades, dos autores tienen en común que definen características que los estudiantes deben evaluar o han evaluado históricamente de las universidades. En los otros

dos trabajos, las recomendaciones están basadas en la evaluación de los perfiles de los estudiantes.

El trabajo de (Iyengar *et al.*, 2017) implementa el algoritmo de CF de manera similar a esta investigación, que busca generar un Top-N de predicciones a partir de la distancia entre un usuario y otros con información histórica a partir del análisis de su perfil.

2.3 Recomendación de programas académicos

Tabla 3: Búsqueda de programas académicos

Tarea	Métrica computacional	Documento
Basado en Contenido (Content-Based)	No se evidencia.	(Maridueña-Arroyave <i>et al.</i> , 2016)
	No se evidencia.	(Pinto <i>et al.</i> , 2016)
	WPM Potencia media LSP Puntuación de preferencia lógica.	(Maridueña-Arroyave y Febles-Estrada, 2016)
Sistemas de recomendación híbridos (Hybrid)	No se evidencia.	(Meryem <i>et al.</i> , 2016)

En los trabajos que recomiendan programas académicos, se tiene en común que todos aplican el paradigma CB; asimismo, en el sistema híbrido se les da mayor peso a los resultados generados desde el recomendador de CB. A partir de esto es posible establecer que el paradigma CB es de los más importantes para lograr predicciones de programas académicos. Además, todos los programas académicos están definidos a partir de vectores de calificaciones dadas por los usuarios a algunas características definidas para cada programa o de materias, de las que se mide la importancia para cada programa. También, hay un aporte valioso para el objetivo de este trabajo relacionado con dos artículos orientados a la definición de palabras clave de los programas académicos.

En general se evidencia que la mayoría de los artículos habla de soluciones que no logran concretarse en una validación en línea que pueda obtener la retroalimentación de los usuarios y por ende realizar mejoras enmarcadas en las sugerencias y observaciones de ellos. Adicionalmente, en los recomendadores incluidos sólo en (Unelsrød, 2011) se habla específicamente de la importancia de las explicaciones en los resultados, permitiendo realizar evaluaciones de los usuarios más detalladas y que buscan indagar en la importancia de las recomendaciones generadas.

3. ESTRATEGIA DE SOLUCIÓN

Los RS tienen sus raíces en diversas áreas de investigación, como la recuperación de información (*information retrieval*), el filtrado de información y la clasificación de textos, y aplican métodos de diferentes campos, como el aprendizaje automático, la extracción de datos y los sistemas basados en el conocimiento (Jannach *et al.*, 2011).

Existen diversos tipos de RS, por ejemplo, los que recomiendan *ítems* basados en nichos de idioma, país, edad, entre otras características de los usuarios, son conocidos como Demográficos. Los Basados en Conocimiento que tienen como objetivo principal encontrar la utilidad de la recomendación para el usuario, evaluando los requerimientos iniciales, reparando inconsistencias de los requisitos y explicando los resultados de las recomendaciones. Finalmente, los que solucionan la problemática planteada en este trabajo, son los Sistemas de Recomendación Híbridos que se basan en la combinación de técnicas, de manera que se puedan usar las ventajas de una para solucionar las desventajas de otra (Ricci *et al.*, 2015).

El prototipo funcional desarrollado como producto de este trabajo, implementa un Sistema de Recomendación Híbrido Mixto que produce dos listados de recomendaciones de programas. El primer listado se genera a través de una estrategia de CF que produce recomendaciones de programas académicos a partir de la información de los vecinos más cercanos al usuario activo, usando el algoritmo *Kneighbors Classifier*. El segundo listado de programas se genera a partir de una estrategia CB mediante el procesamiento de lenguaje natural vectorizando conceptos asociados a cada programa académico a partir de una estrategia *Bag of Words*.

Las métricas computacionales, miden que tan buenos son los listados de recomendaciones a partir del *Hit-Ratio* (Lee, 2010) y el MRR (*Mean Reciprocal Rank*) (Dias y Fonseca, 2013) que buscan las ocurrencias reales de un *ítem* en un listado de recomendaciones.

El modelo de explicaciones está centrado en un estilo Híbrido de 2 dimensiones que combina explicaciones de estilo Humano con otras explicaciones de estilo de Características (Papadimitriou, 2012). El estilo Humano proporciona explicaciones basadas en usuarios similares, es decir, considera que el usuario activo y los usuarios que se usaron para generar la explicación tienen intereses similares. Se incluye en la solución un componente que agrupa las explicaciones proporcionadas a los estudiantes en un histograma que muestra la relación entre cada programa académico

recomendado con los vecinos del Usuario Activo. El estilo de Características está basado en la definición de palabras clave que son usadas como características de los *ítems*. Se considera que este estilo ayuda a los usuarios a precisar las decisiones (Papadimitriou, 2012). Este estilo está relacionado, en el trabajo actual, con las recomendaciones que se generan a partir de la selección de conceptos asociados a otros programas recomendados al estudiante con anterioridad. Al incluir este estilo se busca refinar las recomendaciones iniciales.

El supuesto sobre el que trabaja el sistema es que el usuario activo tendrá buen desempeño en los programas académicos recomendados si usuarios “estudiantes universitarios” con contexto y características (sociodemográficas, evaluación examen de estado, información del colegio donde provienen, entre otras) similares, tuvieron buen desempeño en el programa académico en el que están matriculados.

3.1 Filtrado Colaborativo

El CF tradicionalmente se encarga de la predicción de ratings, para el trabajo propuesto se usa el método de vecinos más cercanos, orientado hacia la clasificación multiclase para predecir para cada usuario la clase con la probabilidad más alta. Por lo tanto, el principio detrás de esta implementación del método, es encontrar un número predefinido de muestras de entrenamiento (registros históricos de estudiantes) cercanas a la distancia del nuevo punto y predecir la etiqueta a partir de estas. El número de muestras puede ser una constante definida por el usuario (aprendizaje del vecino más cercano a k) o variar según la densidad local de puntos (aprendizaje del vecino basado en el radio). La distancia puede ser, en general, cualquier medida métrica. Los métodos basados en vecinos se conocen como de aprendizaje automático no generalizados, ya que simplemente “recuerdan” todos sus datos de entrenamiento.

Definiendo, los estudiantes se encuentran en el conjunto U que contiene la información de los usuarios del sistema, y se representa por un vector de C posiciones así: $X_u \in \mathbb{R}^C$ donde C corresponde al número de características o variables independientes de cada usuario; después de aplicar las técnicas de normalización a los datos $C = 70$.

Entonces, se calcula la probabilidad de que el usuario u tenga asociada un ítem o clase (cat):

$$prob_{programa(u,cat)} = \sum_{k \in N(u)} \frac{\frac{1}{d(k,u)}}{total_{peso(u)}} * \frac{\mathbb{I}(cat(k) == cat)}{|N(u)|} \quad (1)$$

donde $N(u)$ es el vecindario del usuario activo u , k es un elemento dentro de este vecindario, $d(k,u)$ es la distancia que hay en el espacio entre el elemento del vecindario y el usuario activo. El total peso se calcula de la siguiente manera:

$$total_{peso(u)} = \sum_{k \in N(u)} \frac{1}{d(k,u)} \quad (2)$$

Para determinar el vecindario de un usuario se mide a través de la distancia *Manhattan* entre los puntos. La distancia *Manhattan*, distancia rectilínea o distancia L1 dice que la distancia entre dos puntos es la suma de las diferencias (absolutas) de sus coordenadas. Es decir, la distancia d , entre dos vectores u y v , en un espacio vectorial real n -dimensional con un sistema de coordenadas cartesianas fijo, es la suma de las longitudes de las proyecciones del segmento de línea entre los puntos sobre el sistema de ejes coordenados:

$$d(u,v) = \|u - v\|_1 = \sum_{i=1}^n |u_i - v_i| \quad (3)$$

donde $u = (u_1, u_2, \dots, u_n)$ y $v = (v_1, v_2, \dots, v_n)$ son vectores.

Para este trabajo se hacen predicciones del Top-N de programas académicos en términos de ordenar por programa más probable. Dicha predicción está dada por el vecindario de vectores de estudiantes más similares a las características del Usuario Actual para el cual se quiere generar la predicción, de tal manera que al extrapolar los programas académicos se recomiende una lista de N ítems. Para encontrar el mejor ítem se recomienda a un usuario activo u_a una lista $L(u_a)$ de clases predichas ordenadas de mayor a menor probabilidad y donde esas clases corresponden a N ítems que pueden interesarle. Sea $T(u)$ el subconjunto de ítems de prueba que un usuario u encontró relevante.

3.2 Filtrado Basado en Contenido

La recomendación CB se centra en la disponibilidad de descripciones de los ítems (creados manualmente o extraídos automáticamente) y en un perfil de usuario que asigna importancia a estas características. Una ventaja de esta estrategia es que no se requiere información de otros usuarios para hacer una recomendación. Otra ventaja es que los nuevos ítems se pueden recomendar inmediatamente una vez que los atributos del ítem estén disponibles (Jannach *et al.*, 2011).

La estrategia CB está asociada a la construcción de una Bolsa de palabras o *Bag of words* (BoW) que se utiliza para representar documentos ignorando el orden de las palabras y representar el vocabulario que utilizaremos en nuestro modelo. Es decir, que cada elemento está formado por un vector en un espacio n -dimensional, donde cada dimensión corresponde a un término relevante tomado del texto asociado a cada programa académica (Ricci *et al.*, 2015). La bolsa está formada por “palabras clave en contexto”, las cuales se refieren a secuencias lineales de unidades lingüísticas (palabras) que son conocidas como n -gramas.

Para BoW, el vector de palabras se expresa como $BoW = (w_0, w_1, \dots, w_{|Con|}) \in \mathbb{R}^{|Con|} [0, 1]$ donde Con es el conjunto de palabras que definen el total de conceptos extraídos en el conjunto I . Para cada coordenada w_c en el vector, su peso representa una indicación numérica del grado de afinidad entre el ítem y el concepto representado por la coordenada (Moreno, 2014).

Para poder calcular los pesos de cada coordenada en el vector, se utiliza el esquema de ponderación de términos más comúnmente referenciado TF-IDF o Frecuencia de término - Frecuencia de documentos inversa. Esta estrategia consiste en asignar una ponderación a un concepto proporcional al número de veces que aparece en el documento (TF) e inversamente proporcional al número de documentos que aparece (IDF) (Moreno, 2014).

$$TF_{IDF}(t, i) = TF(t, i) \cdot \log \frac{N}{n_t} \quad (4)$$

donde N denota el número de documentos en el corpus y n_t denota el número de documentos en la colección en la cual el término t ocurre al menos una vez, finalmente $N = |I|$ y $t \in Con$.

Cuando el usuario selecciona los conceptos, estos pasan a ser una nueva entrada del modelo CB, construyendo una cadena de texto a la que se aplica los pasos de procesamiento de lenguaje natural, incluyendo la vectorización y la ponderación TF-IDF, para formar un vector similar al de los documentos definido para cada programa. Luego se compara con los textos definidos para cada programa buscando los más similares. Esta estrategia permite que se haga una nueva búsqueda para encontrar programas que se pueden recomendar al usuario.

3.3 Sistemas de Recomendación Híbridos

La construcción de sistemas híbridos que combinan las fortalezas de diferentes algoritmos

y modelos para superar algunas de las deficiencias y problemas de estos, se ha convertido en el objetivo de investigaciones recientes (Jannach *et al.*, 2011).

El diseño de hibridación implementado es el paralelizado mixto. Este tipo de RS funciona cuando se presenta al usuario una lista de ítems relevantes, en lugar de mostrarle una lista originada por un RS. La salida de dos o más sistemas de recomendación se combina en una sola lista (Burke, 2002). El resultado de la recomendación para el usuario u y el ítem i de una estrategia híbrida mezclada es el conjunto de n -tuplas $\{score, k\}$ para cada una de sus n que constituyen recomendadores rec_k :

$$rec_{mixed}(u, i) = \bigcup_{k=1}^n (rec_k(u, i), k) \quad (5)$$

4. EVALUACIÓN

La evaluación de los resultados del modelo se hizo a través de las métricas *Accuracy*, *Hit Ratio* y *MMR* para los datos de prueba y validación. Teniendo en cuenta que al estudiante se le presenta un Top-5 de programas académicos.

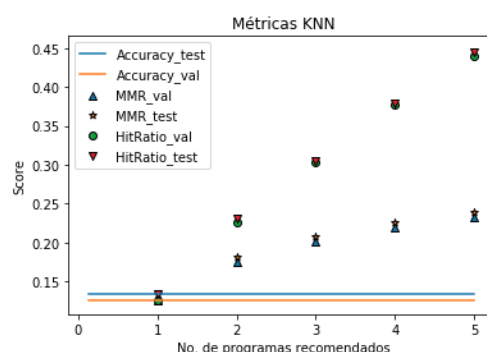


Fig. 1. Evaluación del modelo

En la figura 1, en el eje X se muestra el número de programas recomendados en el listado Top-N y en el eje Y el acierto de las métricas.

Los métodos basados en KNN tienen la ventaja de ser relativamente fácil de implementar, adaptarse rápidamente a los cambios recientes y tienen la ventaja de que, en comparación con otros enfoques de aprendizaje, un número relativamente pequeño de calificaciones es suficiente para hacer una predicción de calidad razonable. Sin embargo, la precisión de la predicción de los métodos de kNN puros puede ser inferior a la de otras técnicas más sofisticadas (Jannach *et al.*, 2011). En parte, esta puede ser una justificación válida para los resultados obtenidos en la primera prueba, pero también es importante analizar que el conjunto de datos se

encuentra desbalanceado por lo cual algunas clases no están representadas adecuadamente.

El modelo de vecinos más cercanos con los parámetros algoritmo = Auto, métrica = Mahattan, $p = 1$, $n_neighbors = 177$ y pesos = Distancia, es el implementado para el sistema de recomendación propuesto y estará disponible para los estudiantes de educación media. Es de carácter opcional y donde se puede dar crédito a instituciones y personas por su aporte.

5. CONCLUSIONES

Se debe hacer énfasis en que, al no seleccionar una estrategia de datos balanceados, los resultados van a estar sesgados hacia aquellas clases que están mejor representadas, por lo cual la implementación del enfoque CB puede alivianar la carga al permitir a los estudiantes revisar otros programas relacionados a los conceptos con los que sienten afinidad. Tomado de la tesis de maestría (Orozco, 2019).

La hibridación de dos modelos que buscan dar recomendaciones nuevas e inesperadas a partir de la estrategia de vecinos más cercanos y mejorarlas al presentar programas académicos adicionales producto de la afinidad que tiene el estudiante con conceptos generados a partir de procesamiento de lenguaje natural, hace de esta una nueva solución a la problemática de los estudiantes de educación media que está dirigida por la información y la interacción sólo de los estudiantes.

Como ya se ha dicho este sistema es una herramienta adicional a las existentes, que a diferencia de los trabajos revisados incluye un modelo de explicaciones que permite a los estudiantes comprender por sí mismos las recomendaciones de programas académicos que se le presentan. Es un estilo de explicaciones implementado en otros dominios, pero que puede ser ampliamente aceptado para el propósito de este sistema de recomendación.

También, se incluyó la validación off-line a través de métricas de ranking, no evidenciada en los trabajos revisados en el estado del arte.

Aún queda pendiente realizar la validación con usuarios reales que permita obtener la retroalimentación necesaria para detectar oportunidades de mejora visibles que permitan medir la satisfacción del consumidor, la confianza real del usuario y la eficacia de los resultados.

Como trabajo futuro se debería, ampliar la base de datos histórica de estudiantes de pregrado, con

la información de estudiantes de otras universidades y regiones del país, de manera que se pueda tener un conjunto de datos más amplio y que permitan eliminar el sesgo hacia la clase que está mayormente representada. También, obtener información de los programas académicos se puede implementar un modelo de enriquecimiento semántico y análisis ontológico con Dbpedia, para beneficiarse de las características técnicas de este tipo de modelos.

Además, incluir una evaluación del comportamiento de los estudiantes mediante su historial académico que permita encontrar características determinantes en la toma de decisiones y que no se hayan tenido en cuenta aún.

REFERENCIAS

- Vicentini V., E. (2016). “Marco sectorial de educación y desarrollo infantil temprano”.
- Ministerio de Educación Nacional. (2012). “Estrategias para reducir la deserción,” **Vol. 2**, No. 2, pp. 88–88.
- Perez B., Castellanos C. and Correal D. (2018). Predicting Student Drop-Out Rates Using Data Mining Techniques: A Case Study. *Applications of Computational Intelligence*. Springer International Publishing, pp. 111–125.
- Ricci F., Rokach L., y B. Shapira. (2015). *Recommender Systems Handbook*, 2a ed., vol. 247, no. 6403.
- Jannach D. y Adomavicius G. (2016). “Recommendations with a Purpose,” in *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, pp. 7–10.
- Rivera A.C., Tapia-Leon M. y Lujan-Mora S. (2018). Recommendation Systems in Education: A Systematic Mapping Study. *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*, Vol. 721.
- Iatrellis O., Kameas A. y Fitsilis P. (2017). “Academic Advising Systems: A Systematic Literature Review of Empirical Evidence,” *Education Sciences*, **Vol. 7**, no. 4, p. 90.
- Peña-Ayala A. (2014). “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Systems with Applications*, **Vol. 41**, no. 4 PART 1, pp. 1432–1462.
- Castellano E. J., Barranco M. J. y Martínez L. (2011). “Academic Orientation Supported by Hybrid Intelligent Decision Support System,” in *Efficient Decision Support*

- Systems - Practice and Challenges From Current to Future*, InTech, pp. 241–262.
- Unelsrød H. F. (2011). “Design and Evaluation of a Recommender System for Course Selection,” Norges teknisk-naturvitenskapelige universitet.
- Maridueña-Arroyave M. R., Estrada A. F. y González R. C. (2016). “Modelo de recomendación para la orientación vocacional basado en la computación con palabras [Recommendation models for vocational orientation based on computing with words],” **Vol. 15**, no. 1, pp. 80–92.
- Vialardi C., Bravo J., Shafti L. y Ortigosa A. (2009). “Recommendation in Higher Education Using Data Mining Techniques,” *Proceedings of the 2nd International Conference on Educational Data Mining*, pp. 191–199.
- Ognjanovic I., Gasevic D., y Dawson S. (2016). “Using institutional data to predict student course selections in higher education,” *Internet and Higher Education*, **Vol. 29**, pp. 49–62.
- Cho J. y Kang E. Y. (2010). “Personalized curriculum recommender system based on hybrid filtering,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6483, pp. 62–71.
- Pinto F. M., Estefania M., Cerón N. y Andrade R. (2016). “iRecomendYou: A design proposal for the development of a pervasive recommendation system based on student’s profile for Ecuador’s students’ candidature to a scholarship,” *New Advances in Information Systems and Technologies*, vol. 445, pp. 537–546.
- Maridueña-Arroyave M. R. y Febles-Estrada A. (2016). “A college degree recommendation model [Modelo de recomendación de carreras universitarias],” *DYNA (Colombia)*, vol. 83, no. 199, pp. 29–34.
- Ragab A. H. M., Mashat A. F. S. y Khedra A. M. (2012). “HRSPCA: Hybrid recommender system for predicting college admission,” *International Conference on Intelligent Systems Design and Applications, ISDA*, pp. 107–113.
- Iyengar M., Sarkar A. y Singh S. (2017). “A Collaborative Filtering based model for recommending graduate schools,” *2017 7th International Conference on Modeling, Simulation, and Applied Optimization, ICMSAO 2017*, pp. 0–4.
- Bokde D. K., Girase S. y Mukhopadhyay D. (2016). “An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques,” *Proceedings - 2015 IEEE International Symposium on Nanoelectronic and Information Systems, iNIS 2015*, pp. 231–236.
- Hu Q., Kevin F. Y. y Craig P. (2017). “Towards a recommendation approach for university program selection using Primitive Cognitive Network Process,” *14th International Conference on Services Systems and Services Management, ICSSSM 2017 - Proceedings*, pp. 3–6.
- Meryem G., Douzi K. y Chantit S. (2016). “Toward an E-orientation Platform”.
- Xu J., Xing T. y Van Der Schaar M. (2016). “Personalized Course Sequence Recommendations,” *IEEE Transactions on Signal Processing*, **vol. 64**, no. 20, pp. 5340–5352.
- Farzan R. y Brusilovsky P. (2006). “Social navigation support in a course recommender system,” *Proceedings of the 4th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, pp. 91–100.
- Jannach D., Zanker M., Felfernig A. y Friedrich G. (2011). *Recommender systems: an introduction*, **Vol. 40**.
- Lee D., Park S. E., Kahng M., Lee S. y Lee S. G. (2010). “Exploiting contextual information from event logs for personalized recommendation,” *Studies in Computational Intelligence*, vol. 317, pp. 121–139.
- Dias R. y Fonseca M. J. (2013). “Improving music recommendation in session-based collaborative filtering by using temporal context,” *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, pp. 783–788.
- Papadimitriou A., Symeonidis P. y Manolopoulos Y. (2012). “A generalized taxonomy of explanations styles for traditional and social recommender systems,” *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 555–583.
- Moreno Barbosa A. D. (2014). “Privacy-enabled scalable recommender systems,” Université Nice Sophia Antipolis.
- Burke R. (2002). “Hybrid Recommender Systems: Survey and Experiments,” *User Modeling and User-Adapted Interaction*, **Vol. 12**, no. 4, pp. 331–370.
- Orozco Cacique J. (2019). “Sistema de Recomendación de programas universitarios para la Orientación Profesional de Estudiantes de Educación Media,” Universidad de los Andes.