UTADEO	MACROPROCESO ENSEÑANZA Y FORMACIÓN ESTUDIANTIL	CÓDIGO:	FOR-EFE-GDB-007
	PROCESO GESTIÓN DE BIBLIOTECAS	VERSIÓN:	0
	AUTORIZACIÓN DE PUBLICACIÓN DE TESIS O TRABAJO DE GRADO	FECHA:	16/07/2019

Bogotá, D.C., Fecha: día 09 mes 02 año 2021

Señores
Sistema de Bibliotecas

Universidad de Bogotá Jorge Tadeo Lozano

Ciudad
Estimados Señores:

Yo, nosotros los abajo firmantes, en calidad de autor (es) de la tesis y/o trabajo de grado **titulado Evaluating Models for a Higher Education Course Recommender System** presentado y aprobado en el año 2020 como requisito para optar al título de **MAGISTER EN INGENIERIA Y ANALITICA DE DATOS**; autorizo (amos) al **Sistema de Bibliotecas Universidad de Bogotá Jorge Tadeo Lozano** para que con fines académicos, preserve, conserve, organice, edite, modifique tecnológicamente y divulgue el documento anteriormente mencionado a través del catálogo en línea de las Bibliotecas Institucionales, el Repositorio Institucional, las bases de datos y redes con las que establezca convenio la Universidad y el Sistema de Bibliotecas.

•Autorizo (amos) a usuarios internos y externos de la Institución a consultar y reproducir el contenido del trabajo de grado para fines académicos nunca para usos comerciales, cuando mediante la correspondiente cita bibliográfica se le de crédito a la obra y su(s) autor(s).

• Autorizo (amos) aplicar la licencia del estándar internacional Creative Commons (Attribution-NonCommercial-NoDerivatives 4.0 International) que indica que cualquier persona puede usar la obra dando crédito al autor, sin poder comerciar con la obra y sin generar obras derivadas.
• El (los) autor(es) certifica(n) que la tesis y/o trabajo de grado no infringe ni atenta contra derechos industriales, patrimoniales, intelectuales, morales o cualquier otro de terceros, así mismo declaran que la Universidad Jorge Tadeo Lozano se encuentra libre de toda responsabilidad civil, administrativa y/o penal que pueda derivarse de la publicación del trabajo de grado y/o tesis en calidad de acceso abierto por cualquier medio.
En cumplimiento con lo dispuesto en la Ley 1581 de 2012 y especialmente en virtud de lo dispuesto en el Artículo 10 del Decreto 1377 de 2013, autorizamos a la Universidad Jorge Tadeo Lozano a proceder con el tratamiento de los datos personales para fines académicos, históricos, estadísticos y administrativos de la Institución. De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, aclaramos que ***"Los derechos morales sobre el trabajo son propiedad de los autores",*** los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.
Cordialmente,

JENNY MAYERLY DIAZ DIAZ	_____	_____
Nombre Completo	Nombre Completo	Nombre Completo
_____	_____	_____
Firma	Firma	Firma
C.C 1033710037 De: Bogotá	C.C_____De:_____	C.C_____De:_____
Correo electrónico: jennym.diazd@utadeo.edu.co	Correo electrónico: _____	Correo electrónico: _____

Vo.Bo de publicación por parte del director, asesor o tutor del trabajo de grado o tesis.

Nombre: Ixent Galpin

Firma:
Documento de identidad: 303.572

Evaluating Models for a Higher Education Course Recommender System using State Exam Results*

Jenny Mayerly Díaz-Díaz[0000−0001−8779−3506] and Ixent Galpin[0000−0001−7020−6328]

Facultad de Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano
Bogotá, Colombia
{jennym.diazd,ixent}@utadeo.edu.co

Abstract. When young people approach the end of their schooling, they are faced with a plethora of often daunting decisions, including whether to go to University or other further education institution, and what further education course is most suitable for them. In this paper, we propose using result data obtained from the Colombian *Saber 11/T&T/Pro* state exams as input for a higher education course recommender system. We compare five different recommender models by analyzing precision, recall, ROC curves and prediction error. Our findings are that user-based collaborative filtering, and the model that recommends the most popular courses, are the ones that perform best. We note that while the context of this work is in Colombia, most other countries have similar or equivalent state exams. It can therefore be expected that our research findings can be more generally applied to other contexts. As further work, we hope to deploy this recommender system as a mobile telephone application for young people to use to help them choose higher education courses.

Keywords: recommender system · collaborative filtering · higher education courses

1 Introduction

When young people approach the end of their schooling, they are faced with a plethora of often daunting decisions, including whether to go to University or a further education institution, and what further education course is most suitable for them. This paper addresses the problems regarding the question of what to study at University. For many completing high school and making the leap to the adult world, it is perhaps one of the most difficult decisions that they face. For most, it is not clear what their professional field of action will be, or where they should study. In this paper, we propose a Recommender System (RS) to advise students finishing secondary school on further education courses suitable for them.

Recommender Systems [12] are typically useful in settings where there is a broad range of items to choose from. For example, in a movie recommendation setting such as [4], based on preferences elicited from users, a typical RS approach is to recommend

* This work is supported by Minciencias from the Call for Research Proposals 804-2018, Project Grant 115680463846.

movies to a user based on preferences of users with similar tastes, as is the case with User-based Collaborative Filtering (UBCF) [5]. The problem of selecting a higher education course is unlike that of selecting a movie to watch. While a person may watch a movie relatively frequently, most people only choose a further education course once in their lives, due to the time and financial costs entailed by this decision. Furthermore, the consequences of choosing an incorrect movie are typically, at worst, a couple of hours of wasted time. In contrast, making the wrong decision when leaving school can have significant ramifications for the rest of that person's life.

Recommending the best higher education course to study using a technique such as UBCF may be especially challenging due to, as mentioned previously, the relatively small number of higher education courses a person typically takes in their lives. This means that there is likely to be very few ratings per user. In effect, everyone in the rating dataset may turn out to be a cold-start user [9]. To circumvent this potential issue, in this paper we propose using the scores obtained for the various state exams as a proxy for ratings in a Recommender System. In Colombia, there are two sets of state exams: *Saber 11* is typically taken during eleventh grade, when nearing the end of school. *Saber Pro* and *Saber T&T* are taken when the higher education course, for professional or technical courses respectively. These can be taken as an indication of how successful the student was in that course, given that the content of the exams is related to the higher education course taken.

The *Saber 11* exam is broken down into different thematic modules such as, for example, language comprehension or mathematical reasoning. Our proposal involves using the score for each thematic module to predict how well a student will do in the *Saber Pro* or *Saber T&T* exams, using classical recommender system techniques. Based on the predicted *Saber Pro* or *Saber T&T* scores, we can then recommend the most suitable higher education courses to the student. In this paper, we evaluate the performance of various RS technique approaches using recommenderlab [3], a framework in R that enables the comparison of different recommendation methods, to determine the most suitable approach for such a recommender system.

This paper is structured as follows: Section 2 presents a background on the Colombian state exams, and various recommender system algorithms which we evaluate. In Section 3, we describe related work. Section 4 detailed the data sources used, how they were integrated, and the application of normalization techniques. We subsequently, in Section 5, illustrate how the deployment of the algorithm would be, by illustrating the example recommendation output with different techniques. Section 6 describes the experimental evaluation carried to determine the performance of the aforementioned recommender system techniques. Finally, Section 7 concludes.

2 Background

2.1 Colombian State Exams

For a large number of young people it is clear that they want to continue studying, but they do not know what subject to study, or what higher education institution is the most

suitable for them. In Colombia, the *ICFES*[1] state entity offers educational evaluation services at all levels, and in particular, supports the National Ministry of Education in conducting state exams, such as the *Saber 11* exam, which aims to evaluate eleventh grade students to enable them to access higher education. Furthermore, it acts as an important barometer to monitor the quality of education offered by secondary education establishments. It can also be taken by those who have already obtained school-leaving baccalaureate diploma, or have passed a baccalaureate validation exam. On the other hand, the state *Saber Pro* exam is aimed at students who have passed 75 % of the credits of their respective professional university education programs.

Saber 11 Exam: The *Saber 11* exam is designed to assess the degree of development of skills of students who are finishing eleventh grade at School. The results are used by different actors in the educational system and beyond, *viz.*:

- it provides students with elements for their self-evaluation and the development of their life project.
- it allows higher education institutions to select suitable candidates for their education programs and to monitor their academic progress.
- it offers educational establishments references for their self-evaluation processes and orientation of their pedagogical practices.
- allows educational authorities to build quality indicators.

Annually, approximately 600,000 students take the *Saber 11*.

Exam Saber Pro and T&T: The *Saber Pro and T&T* exam has a compulsory first section for all those who take the exam, which is made up of 5 modules that assess generic skills (critical reading, quantitative reasoning, citizenship skills, written communication and English). In addition, there are modules associated with specific themes and contents that students have the possibility to take according to their area of study.

Approximately 400,000 students take the *Saber Pro and T&T* exams every year.

2.2 Recommender System Algorithms

This section briefly summarizes several algorithms used by Recommender Systems and which are supported by recommenderlab, which are used for the evaluation in Section 6.

- *User-based Collaborative filtering (UBCF)*: This algorithm produces recommendations that are made based on the terms of similarity between users. In other words, this algorithm recommends items that are of the same affinity as other users. The objects to be recommended are chosen from those that have received the highest score from other users with similar tastes or interests, thus reflecting a similar pattern of preferences [2].
- *Item-based Collaborative filtering (IBCF)*: This algorithm recommends items based on items that are deemed to be similar to those preferred by that user [6].

[1] ICFES is an abbreviation in Spanish for *Colombian Institute for the Evaluation of Education*. See https://www.icfes.gov.co/.

- *Singular value decomposition (SVD):* This algorithm is based on matrix factorization. Broadly speaking, it reduces the number of features in the user element ratings matrix by reducing its dimensions [7].
- *Random:* This algorithm selects items at random to the user.
- *Popular:* This algorithm recommends items based on their overall popularity, based on those which have the highest scores.
- *Rerecommend:* This algorithm recommends items which have been previously highly rated. It has two variants: `recommendHPR` recommends items according to the highest predicted ratings for a user. `recommendMF` recommends the most frequent item in a user item rating matrix.
- *Hybrid Recommender:* The approach combines the recommendation algorithms mentioned previously, according to the weights specified.

3 Related Work

The choice of a technical or professional career is an important act in the life of a person, for which a study was developed to make a recommendation of a possible professional or technical study option, with the support of an efficient tool and the objective of evaluating the knowledge of a grade eleven student, which is the *Saber 11* exam, by simplifying these classification elements, a more effective and subjective prediction is made for the user.

As a first measure, the different recommendation systems allow us to help make a better recommendation for the user and regret that in the field of education it is not well used [10,13]. For the majority of the objective users they lack the necessary knowledge to carry out decision making and that this in turn guides them to select the best option.

The recommendation systems with the academic profiles in the student stage and in the higher education of the base users, offers an alternative to users who want to continue with the next educational stage, that is, their higher education.

There are different techniques or forms for the implementation of a recommendation system, based on the similarity metrics and the information available to form the catalog, which will serve as the main source for the recommender when making the classification [11].

Most of the recommendation models are based on calculating different statistics that provide the number of possibilities of recommendation to the user on a given item, although some differ from the techniques and information on the academic characteristics of the students [1].

There are several documentations on the analysis carried out on education, where some focus on high school students applying various classification algorithms, which assesses the importance of education based on different metrics, providing great collaboration with teachers [8].

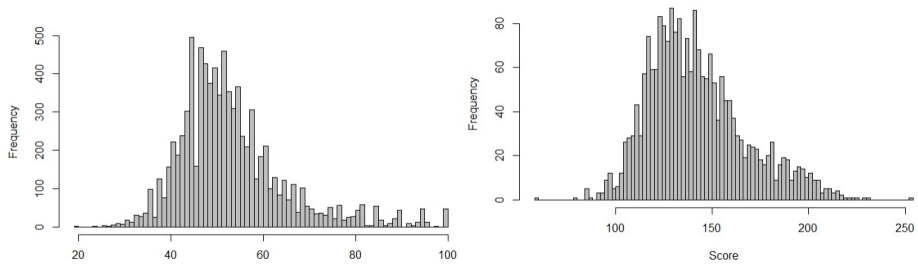
4 Data Preparation

4.1 Data Sources

In order to build the recommender models evaluated in this paper we used three datasets:

- The results of the *Saber 11* exam spanning the periods 2012-1 to 2015-1 comprising 1,751,870 records, corresponding to candidates registered with a high school educational establishment (i.e., we did not include records from candidates who took the exam independently).
- The results from *Saber Pro* and T&T between the years 2016 to 2018, comprising 657,443 records.
- A reference dataset to enable us to map minor identity card numbers to adult identity card numbers. This was necessary because in Colombia these identity documents use different numbering schemes. This dataset is created from the registration questionnaire used during the registration for the *Saber Pro* and Saber T&T exams.

The distribution of the results for the *Saber 11* exam is shown in Figure 1a, and the combined results for the *Saber Pro* and *T&T* exams is shown in Figure 1b. Note that the result ranges between the exams differ. The scores for the *Saber 11* exam are in the range of 0 to 100, whereas for the *Saber Pro* and *T&T* exams they are in the range of 0 to 300.



(a) *Saber 11* exam (b) *Saber Pro* and *T&T* exams

Fig. 1. Histograms showing results for Colombian state exams.

4.2 Integration of Data Sources

In order to create a user $\times$ item ratings matrix, we combined these sources into a single dataframe as follows:

- Join between *Saber Pro*/*Saber T&T* variable that indicates the identity document that the student used when taking *Saber 11* and the identity document variable registered in *Saber 11*.
- Join between variables of identity documents reported at the time of registering for *Saber Pro*/*Saber T&T* and *Saber 11*.
- Phonetic Join between *Saber* Pro students who did not find their *Saber 11* registration with the previous steps and students registered in *Saber 11*. This phonetic join takes into account names, surnames, gender and date of birth.

This dataset generated contains a sample of 2,142 users with their respective results from both exams. These results are found as variables in columns. An excerpt of the dataframe resulting from the integration of the data sources is shown in Table 1.

User	Social Sciences	English	Reading	Maths	Science	Architecture	Business
1	64	95	71	86	100	205	
2	61	85	64	65	62		
3	67	90	62	61	65		
4	51	85	59	67	66		
5	60	60	59	55	55		225

Table 1. Excerpt of the Final dataset showing dataframe resulting from integration of data sources.

4.3 Data Normalization

As can be seen from the histograms in Figure 1, the results for the *Saber 11* and *Saber Pro* and *Saber T&T* exams have different ranges. As such, it is necessary to normalize the results so that they have the same ranges for both types of exams. We do this by using the scalar method, which scales all the results in the range of 0 to 1.

Figures 2a and 2b show the normalization of the *Saber 11* exam data. Figures 2c and 2d show the normalization of the *Saber Pro/T&T* exam data. After having carried out the normalization separately from both exams, we have a final data set normalized and denormalized, as shown in the following Figure 2f and 2e respectively.

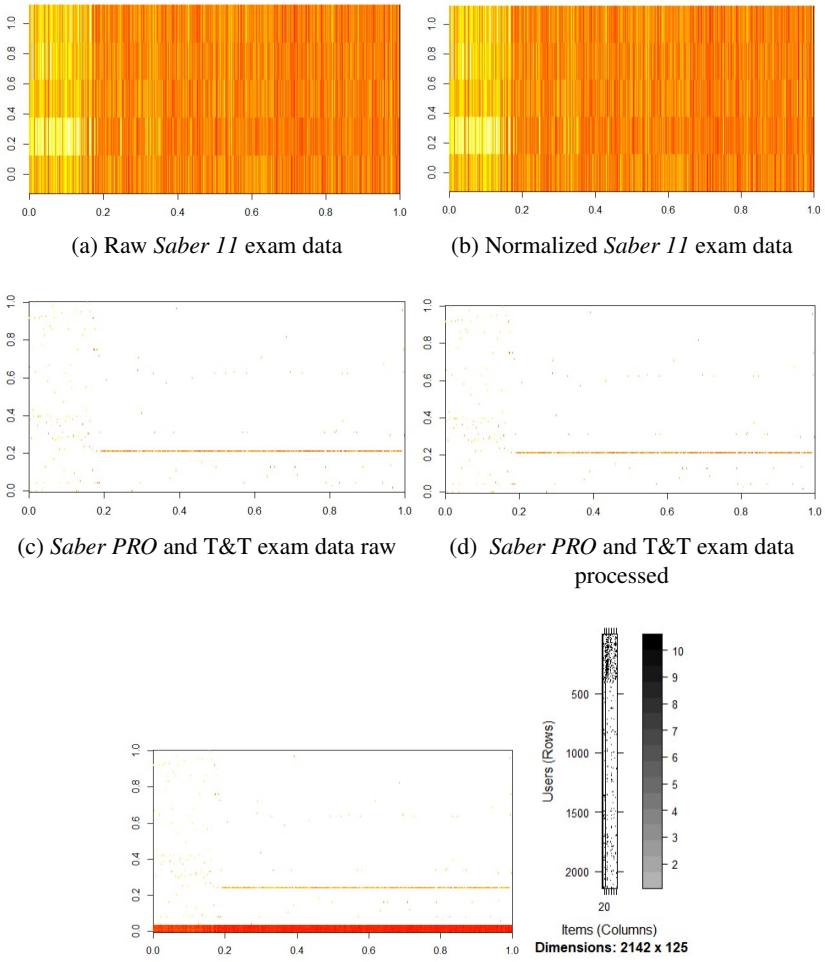
Comparison of normalized and desnormalized data histograms The histogram shows a distribution where they all occur at an almost identical frequency and the most frequent positive ratings with a steady decline toward the rating. Since this distribution may be the result of users with a biased rating, we look at the rating distribution after normalization below.

As shown in Figure 4 the distribution of the scores is closer to a normal distribution after unifying the two data sets.

5 An Example Run

After the algorithms have been evaluated together, each one will be implemented to evaluate their respective predictions.

User-based collaborative filtering (UBCF) To predict the evaluation that a user will make of an item that has not yet been seen, users with similar profiles are searched and the evaluations of these other users on the item are used as an estimate of the user's evaluation. The prediction was made based on the following top-N recommendation (N = 3) for the first 11 selected users. Note that the figure is omitted for space reasons.



(a) Raw *Saber 11* exam data

(b) Normalized *Saber 11* exam data

(c) *Saber PRO* and T&T exam data raw

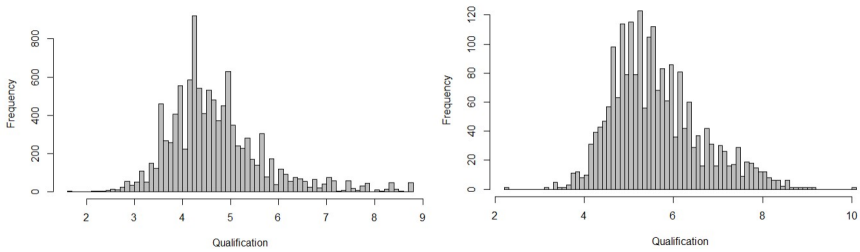
(d) *Saber PRO* and T&T exam data processed

(e) Data raw final

(f) Data processed final

Fig. 2. Normalization of the exam data

Item-based collborative filtering (IBCF) To predict the estimation or evaluation that a user will make of an item that he/she has not yet seen, other similar items that have had or received ratings and that the user has also rated are searched. Estimates are used the user has made similar items as a prediction of their assessment or score on the item. This recommendation system could be confused with the content-based one, the difference is that each item is not defined by its attributes but by the score it has received. The prediction was made based on the following top-N recommendation (N = 3) for the first 11 selected users, as evidenced by the results in Figure 5a.



(a) *Saber 11* Exam (b) *Saber PRO and T&T* Exam

Fig. 3. Histograms showing distributions of exam scores after normalization.

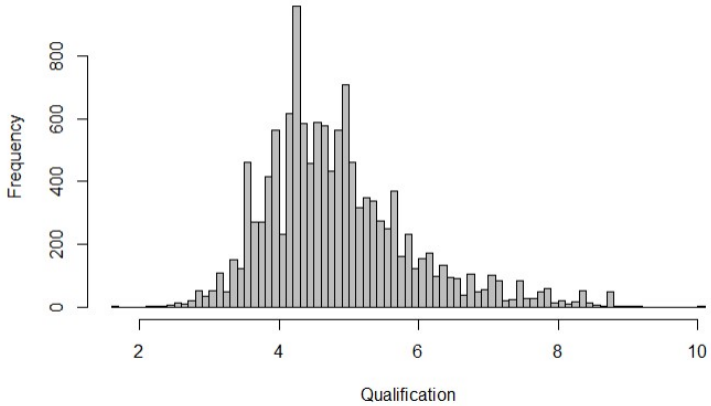


Fig. 4. Histogram showing distribution resulting from Normalization of the integrated data set.

RE-RECOMMEND In Figure 5b the prediction will be displayed, according to the following top-N recommendations (N = 3) for the first 11 selected users.

SVD In Figure 5c the prediction is displayed, according to the following top-N recommendations (N = 3) for the first 11 selected users

Popular Recommender In this case, the model has a list of top-N recommendations (N = 3) for the first 11 selected users, where it shows popularity of professional and technical careers. Note that the result is not shown here for space reasons.

Hybrid Recommender Create and combine recommendations using various recommendation algorithms and given the weight of weights, such as:

Case 1, in Figure 5d:

```
method = "UBCF" Peso = 0.6
method = "POPULAR" Peso = 0.3
method = "IBCF" Peso = 0.1
```

```
[[1]]
[1] "BANCA_FIINZAS_INTERNACIONALES" "INGENIERIA ELECTRONICA"        "DEPORTE"

[[2]]
[1] "MERCADEO_NEGOCIOS_INTERNACIONALES" "NEGOCIOS INTERNACIOLES_CICLO3"
[3] "INSTRUMENTACION_QUIRURGICA"

[[3]]
[1] "TERAPIA RESPIRATORIA" "NUTRICION_DIETETICA"  "DISEÑO_VESTUARIO"

[[4]]
[1] "TERAPIA RESPIRATORIA" "DISEÑO_VESTUARIO"     "LIC_HUM_LEN_CASTE"

[[5]]
[1] "NUTRICION_DIETETICA"        "TERAPIA RESPIRATORIA"       "INSTRUMENTACION_QUIRURGICA"

[[6]]
[1] "ESTUDIOS_LITERARIOS"            "INGENIERIA_SISTEMAS_COMPUTACION" "DEPORTE"

[[7]]
[1] "BANCA_FIINZAS_INTERNACIONALES" "INGENIERIA FIINCIERA"          "DEPORTE"

[[8]]
[1] "NUTRICION_DIETETICA"        "INSTRUMENTACION_QUIRURGICA" "TERAPIA RESPIRATORIA"

[[9]]
[1] "INGENIERIA_INDUSTRIAL"          "DEPORTE"                        "CIENCIAS_MILITARES_AEROIUTICAS"

[[10]]
[1] "DEPORTE"                        "INGENIERIA_INDUSTRIAL"          "CIENCIAS_MILITARES_AEROIUTICAS"

[[11]]
[1] "SEGURIDAD_SALUD_TRABAJO"                    "ADMINISTRACION_EMPRESAS_TURUISTICAS_HOTELERAS"
[3] "ODONTOLOGIA"
```

(a) IBCF algorithm

```
Recommendations as 'topNList' with n = 3 for 11 users.
[[1]]
[1] "PUNT_SOCIALES_CIUDADA" "NORMAL_SUPERIOR"       "PUNT_MATEMATICAS"

[[2]]
[1] "PUNT_CIENCIANATURALES" "PUNT_MATEMATICAS"      "PUNT_SOCIALES_CIUDADA"

[[3]]
[1] "PUNT_LECTURA_CRITICA"  "PUNT_INGLES"           "PUNT_SOCIALES_CIUDADA"

[[4]]
[1] "NORMAL_SUPERIOR"       "PUNT_LECTURA_CRITICA"  "PUNT_CIENCIANATURALES"

[[5]]
[1] "NORMAL_SUPERIOR"       "PUNT_SOCIALES_CIUDADA" "PUNT_LECTURA_CRITICA"

[[6]]
[1] "PUNT_LECTURA_CRITICA"  "NORMAL_SUPERIOR"       "PUNT_SOCIALES_CIUDADA"

[[7]]
[1] "PUNT_SOCIALES_CIUDADA" "PUNT_INGLES"           "PUNT_LECTURA_CRITICA"

[[8]]
[1] "PUNT_SOCIALES_CIUDADA" "NORMAL_SUPERIOR"       "PUNT_CIENCIANATURALES"

[[9]]
[1] "PUNT_SOCIALES_CIUDADA" "NORMAL_SUPERIOR"       "PUNT_CIENCIANATURALES"

[[10]]
[1] "PUNT_CIENCIANATURALES" "NORMAL_SUPERIOR"       "PUNT_SOCIALES_CIUDADA"

[[11]]
[1] "PUNT_CIENCIANATURALES" "PUNT_SOCIALES_CIUDADA" "PUNT_INGLES"
```

(b) Rerecommend algorithm

```
Recommendations as 'topNList' with n = 3 for 11 users.
[[1]]
[1] "NEGOCIOS_INTERNACIONALES" "LIC_HUM_LEN_CASTE"        "INGENIERIA_AMBIENTAL"

[[2]]
[1] "INGENIERIA_CIVIL"              "COMUNICACIÓN_SOCIAL_PERIODISMO" "LIC_HUM_LEN_CASTE"

[[3]]
[1] "PSICOLOGIA"                  "MATEMATICAS"                 "RELACIONES_INTERNACIONALES"

[[4]]
[1] "RELACIONES_INTERNACIONALES" "ARQUITECTURA"               "MATEMATICAS"

[[5]]
[1] "PSICOLOGIA"               "NEGOCIOS_INTERNACIONALES" "ADMINISTRACION_POLICIAL"

[[6]]
[1] "PUBLICIDAD"             "MATEMATICAS"            "INGENIERÍA MECATRÓNICA"

[[7]]
[1] "INGENIERIA_AMBIENTAL" "LIC_HUM_LEN_CASTE"    "DISEÑO_VESTUARIO"

[[8]]
[1] "COMUNICACIÓN_SOCIAL_PERIODISMO" "ADMINISTRACION_POLICIAL"        "NEGOCIOS_INTERNACIONALES"

[[9]]
[1] "DISEÑO_VESTUARIO"  "FILOSOFIA"         "LIC_HUM_LEN_CASTE"

[[10]]
[1] "FILOSOFIA"        "DISEÑO_VESTUARIO" "PUBLICIDAD"

[[11]]
[1] "ADMINISTRACION_SALUD_OCUPACIONAL" "DISEÑO_GRAFICO"
[3] "INGENIERÍA MECATRÓNICA"
```

(c) SVD algorithm

```
$weights
[1] 0.6 0.3 0.1

[[1]]
[1] "ADMINISTRACION_TECNOLOGICA"         "INGENIERIA_ELECTROMECANICA"
[3] "FISICA"                             "GEOLOGIA"
[5] "ADMINISTRACION_PUBLICA_TERRITORIAL"

[[2]]
[1] "ADMINISTRACION_TECNOLOGICA"         "INGENIERIA_ELECTROMECANICA"
[3] "FISICA"                             "GEOLOGIA"
[5] "ADMINISTRACION_PUBLICA_TERRITORIAL"

[[3]]
[1] "ADMINISTRACION_TECNOLOGICA"         "INGENIERIA_ELECTROMECANICA"
[3] "FISICA"                             "GEOLOGIA"
[5] "ADMINISTRACION_PUBLICA_TERRITORIAL"

[[4]]
[1] "ADMINISTRACION_TECNOLOGICA"         "INGENIERIA_ELECTROMECANICA"
[3] "FISICA"                             "GEOLOGIA"
[5] "ADMINISTRACION_PUBLICA_TERRITORIAL"

[[5]]
[1] "ADMINISTRACION_TECNOLOGICA"         "INGENIERIA_ELECTROMECANICA"
[3] "FISICA"                             "GEOLOGIA"
[5] "ADMINISTRACION_PUBLICA_TERRITORIAL"
```

(d) Hybrid algorithm Case 1

Fig. 5. Prediction result for different algorithms

Case 2, not shown for space reasons:

```
method = "UBCF" Peso = 0.3
method = "POPULAR" Peso = 0.5
method = "IBCF" Peso = 0.2
```

6 Evaluation

For our evaluation, a cross-validation scheme was created (K = 10). We specified the `given = 3` parameter, meaning that for test users, all but three randomly selected items are withheld for evaluation. The parameterization of the evaluation scheme was as follows:

```
    scheme <- evaluationScheme(dtev, method="cross", k=10 given=3,goodRating=5)
scheme
Evaluation scheme with 3 items given
Method: 'cross-validation' with 10 run(s).
Good ratings: >=5.000000
```

```
Data set: 2142 x 125 rating matrix of class 'realRatingMatrix' with 12850 ratings.
```

We use the evaluation scheme created to evaluate the popular recommendation method. We evaluated the recommendation lists of top-1, top-3, top-5, top-10, top-15 and top-20:

```
resultsPopular <- evaluate(scheme, method="POPULAR", type
= "topNList", n=c(1,3,5,10,15,20))
```

The result is the classification report for the four runs, as shown on Figure 6.

	TP	FP	FN	TN	precision	recall	TPR	FPR
1	0.3194444	0.6805556	0.7129630	120.2870	0.31944444	0.4148418	0.4148418	0.005618585
3	0.3333333	2.6666667	0.6990741	118.3009	0.11111111	0.4330900	0.4330900	0.022037987
5	0.3379630	4.6620370	0.6944444	116.3056	0.06759259	0.4403893	0.4403893	0.038534230
10	0.3472222	9.6527778	0.6851852	111.3148	0.03472222	0.4501217	0.4501217	0.079793326
15	0.3657407	14.6342593	0.6666667	106.3333	0.02438272	0.4647202	0.4647202	0.120974613
20	0.3750000	19.6250000	0.6574074	101.3426	0.01875000	0.4708029	0.4708029	0.162233390

Fig. 6. Confusion matrix for the POPULAR recommendation algorithm

For the first run we have six confusion matrices represented by rows, one for each of the six different top-N lists that we used for the evaluation. n is the number of recommendations per list. TP, FP, FN, and TN are the inputs for true positives, false positives, false negatives, and true negatives in the confusion matrix. The remaining columns contain the performance calculated. Figure 7 shows the average of all the executions of the evaluation results.

	TP	FP	FN	TN	precision	recall	TPR	FPR
1	0.2259259	0.7740741	0.8180556	120.1819	0.22592593	0.3010066	0.3010066	0.006392765
3	0.2449074	2.7550926	0.7990741	118.2009	0.08163580	0.3249361	0.3249361	0.022771260
5	0.2643519	4.7356481	0.7796296	116.2204	0.05287037	0.3500705	0.3500705	0.039145927
10	0.2819444	9.7180556	0.7620370	111.2380	0.02819444	0.3656503	0.3656503	0.080339572
15	0.2962963	14.7037037	0.7476852	106.2523	0.01975309	0.3776552	0.3776552	0.121560225
20	0.3027778	19.6972222	0.7412037	101.2588	0.01513889	0.3844237	0.3844237	0.162846984

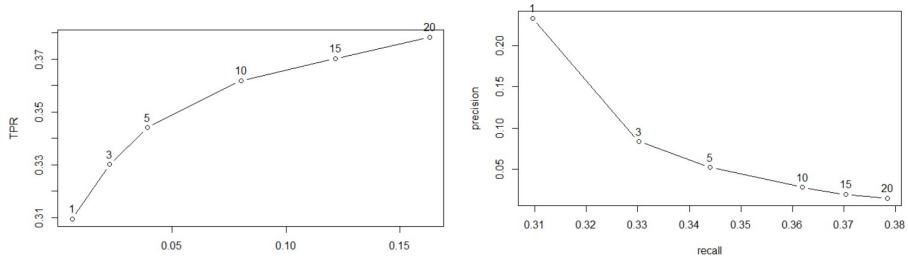
Fig. 7. Average number of executions

The results of the evaluation plotted on the ROC curve shown in Figure 9a is the curve of the true positive rate (TPR) against the false positive rate (FPR), which indicates that in iterations 3,5 and 10 it represents a moderately perfect diagnostic value.

For Figure 8b where we write down the curve with the size of the list N = 1 it is higher with respect to recall of the other iterations.

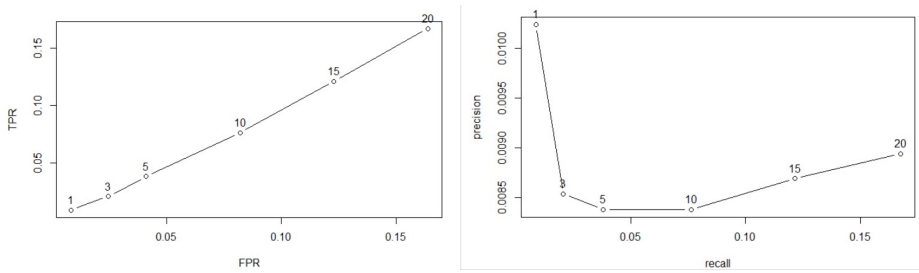
Algorithm evaluation We will evaluate the results of the RANDOM, UBCF, IBCF and SVD algorithms

- Random Algorithm



(a) ROC curve (b) Precision vs. Recall

Fig. 8. Results of the POPULAR recommendation algorithm

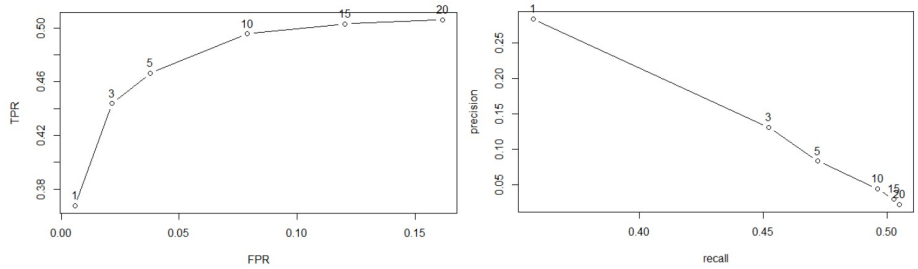


(a) ROC curve (b) Precision vs. Recall

Fig. 9. Results for the RANDOM recommendation algorithm

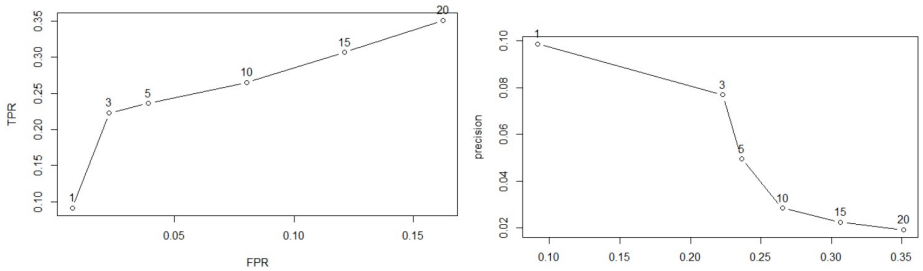
In Figure 9a it shows a random classification, since list N is located along the diagonal line or as the non-discrimination line is known.
In Figure 9b, it is evident that the precision is much more effective when the iteration N = 1, however its recall is better with respect to the other iterations

- UBCF Algorithm
  The results of the evaluation on the ROC curve shown in Figure 10a is the curve of (TPR) with respect to (FPR). The curve indicates that in iterations 3, 5 and 10 it represents a fairly good diagnostic value and even almost the same as the Popular algorithm.
  In Figure 10b, it can be seen that in the N = 1 iteration it has a greater precision with respect to the other iterations, however the iterations 10, 15 and 20 are predominant in recall.
- IBCF Algorithm
  For this algorithm, the evaluation on the ROC curve shown in Figure 11a. The curve indicates that in iterations 5, 10 and 15 they remain slightly constant, therefore their result is quite good.
  In Figure 11b, a drastic change is evident in the N = 5 iteration, since its precision is lower with respect to the recall of the other iterations.
- SVD Algorithm



(a) ROC curve (b) Precision vs. Recall

Fig. 10. Results for the UBCF recommendation algorithm

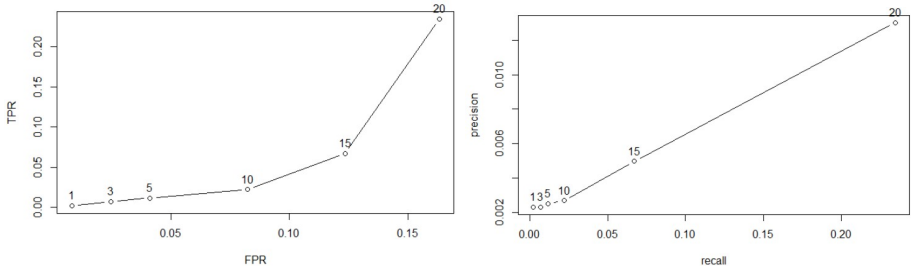


(a) ROC curve (b) Precision vs. Recall

Fig. 11. Results for the IBCF recommendation algorithm

The SVD algorithm, as regards its result of the ROC curve, does not have a favorable diagnostic, since even the N = 20 iteration is good, as can be seen in Figure 12a.

In Figure 12b, iterations 1, 3, 5 and 10 are the result of the best recall, however in terms of precision it was not, but it was for iterations 15 and 20.



(a) ROC curve (b) Precision vs. Recall

Fig. 12. Results for the SVD recommendation algorithm

6.1 Comparison of recommendations

Comparison of various recommendation algorithms is one of the main functions of Recommenderlab [3]. That is why we jointly evaluate these algorithms using the `CROSS` method. We then use the evaluation scheme to compare the five recommendation models:

```
algorithms <- list(
"random items" = list(name="RANDOM", param=NULL),
"popular items" = list(name="POPULAR", param=NULL),
"user-based CF" = list(name="UBCF", param=list(nn=50)),
"item-based CF" = list(name="IBCF", param=list(k=50)),
"SVD approximation" = list(name="SVD", param=list(k = 50)))
resultsalgorcross <- evaluate(scheme, algorithms, type = "topNList", n=c(1, 3, 5, 10,
15, 20))
```

Figures 13a and 13b shows the prediction time and the model time for each algorithm included in the execution list, using four-fold cross validation.

```
RANDOM run fold/sample [model time/prediction time]
         1  [0sec/0.08sec]
         2  [0.01sec/0.1sec]
         3  [0sec/0.12sec]
         4  [0sec/0.1sec]
         5  [0sec/0.06sec]
         6  [0sec/0.09sec]
         7  [0sec/0.09sec]
         8  [0sec/0.09sec]
         9  [0.01sec/0.08sec]
         10  [0.02sec/0.09sec]
POPULAR run fold/sample [model time/prediction time]
         1  [0sec/0.56sec]
         2  [0sec/0.62sec]
         3  [0sec/0.67sec]
         4  [0.02sec/0.62sec]
         5  [0.01sec/0.8sec]
         6  [0.02sec/0.5sec]
         7  [0sec/0.55sec]
         8  [0.02sec/0.54sec]
         9  [0sec/0.56sec]
         10  [0sec/0.57sec]
UBCF run fold/sample [model time/prediction time]
         1  [0sec/0.6sec]
         2  [0sec/0.54sec]
         3  [0sec/0.67sec]
         4  [0.02sec/0.67sec]
         5  [0sec/0.54sec]
         6  [0sec/0.6sec]
         7  [0sec/0.61sec]
         8  [0sec/0.74sec]
         9  [0sec/0.56sec]
         10  [0.01sec/0.5sec]
```

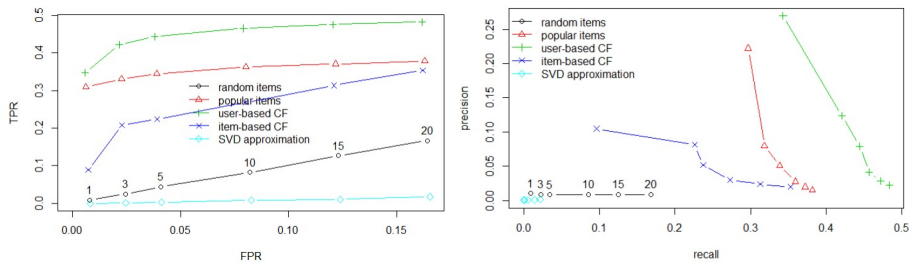
(a) RANDOM, POPULAR and UBCF

```
IBCF run fold/sample [model time/prediction time]
         1  [0.1sec/0.07sec]
         2  [0.1sec/0.09sec]
         3  [0.11sec/0.08sec]
         4  [0.11sec/0.09sec]
         5  [0.11sec/0.06sec]
         6  [0.14sec/0.08sec]
         7  [0.11sec/0.08sec]
         8  [0.12sec/0.08sec]
         9  [0.09sec/0.09sec]
         10  [0.11sec/0.08sec]
SVD run fold/sample [model time/prediction time]
         1  [0.07sec/0.2sec]
         2  [0.06sec/0.15sec]
         3  [0.08sec/0.19sec]
         4  [0.07sec/0.19sec]
         5  [0.05sec/0.17sec]
         6  [0.06sec/0.16sec]
         7  [0.08sec/0.2sec]
         8  [0.09sec/0.21sec]
         9  [0.07sec/0.18sec]
         10  [0.08sec/0.26sec]
```

(b) IBCF and SVD

Fig. 13. Evaluation of the five algorithms

Figure 14a is the ROC curve of the 5 algorithms. Together we observe that the Popular, IBCF, UBCF algorithms have a better diagnosis regarding the recommendations of a professional or technical career, with respect to the Random algorithm and SVD not they were so efficient in diagnosing.



(a) ROC curves (b) Precison vs. recall

Fig. 14. Comparison of performance for the five RS models

Regarding Figure 14b, it is observed that the UBCF, POPULAR and IBCF algorithms have a better recall with respect to the other remaining algorithms, additionally the RANDOM algorithm has a lower precision and a constant in its recall and the SVD algorithm has a lower precision and recall with respect to the other algorithms evaluated.

6.2 Comparison of Predicted Ratings

We evaluate not the top recommendations, but how well the algorithm ratings can predict. Plotting the results shows a bar graph with the mean root error and the mean absolute error as shown in Figure 15, the one with the largest error is the Random algorithm and the one with the least error is the Popular algorithm

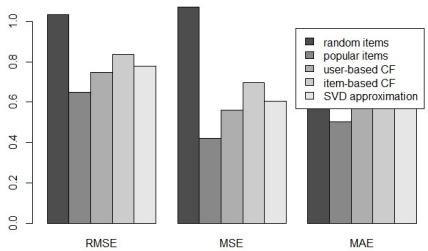


Fig. 15. Comparison of RMSE, MSE and MAE for recommendation methods

7 Conclusions

In this paper, we have shown that it is possible to cast the Colombian state exam results as a user $\times$ item recommendation matrix to construct a Recommender System for recommending higher education courses to young people. We evaluate five different

recommender system models. The worst performing model is the SVD approximation, which has worse results than the random baseline. The best results are obtained using the user-based collaborative filtering and popular models. We note that while the context of this work is in Colombia, most other countries have similar or equivalent state exams. It can therefore be expected that our research findings can be more generally applied to other contexts. As further work, we hope to deploy this recommender system as a mobile application for young people to use to help them choose higher education courses.

References

1. Caprara, G.V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G.M., Barbaranelli, C., Bandura, A.: Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. J. of Edu. Psychology **100**(3), 525 (2008)
2. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of the ACM **35**(12), 61–70 (1992)
3. Hahsler, M.: recommenderlab: A framework for developing and testing recommendation algorithms. Tech. rep. (2015)
4. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) **5**(4), 1–19 (2015)
5. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. on Inf. Systems (TOIS) **22**(1), 5–53 (2004)
6. Kitts, B., Freed, D., Vrieze, M.: Cross-sell: a fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities. In: Proceedings of the sixth ACM SIGKDD. pp. 437–446 (2000)
7. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
8. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., Addison, K.L.: A machine learning framework to identify students at risk of adverse academic outcomes. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1909–1918 (2015)
9. Liu, N.N., Meng, X., Liu, C., Yang, Q.: Wisdom of the better few: cold start recommendation via representative based rating elicitation. In: Proceedings of the fifth ACM conference on Recommender systems. pp. 37–44 (2011)
10. López, M.B., Montes, A.J.H., Ramírez, R.V., Hernández, G.A., Cabada, R.Z., Estrada, M.L.B.: Emoremsys: Sistema de recomendación de recursos educativos basado en detección de emociones. Revista Ibérica de Sistemas e Tecnologias de Informação (17), 80–95 (2016)
11. Olguín, G.M., de Jesús, Y.L., de Celis Herrero, M.P.: Métricas de similaridad y evaluación para sistemas de recomendación de filtrado colaborativo. Revista de Investigación en Tecnologías de la Información **7**(14), 224–240 (2019)
12. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Recommender systems handbook, pp. 1–35. Springer (2011)
13. Torres, E.: Evaluación del uso de algoritmos colaborativos para orientar académicamente al alumnado en bachillerato. Departamento de Informática-Universidad de Jaén **143** (2007)