

A Holistic Recommendation System for Higher Education Academic Activities

Rafael Alves da Silva
Departamento de Informática
Universidade Federal de Viçosa
Viçosa, Minas Gerais, Brazil
rafaelalves@ufv.br

Lucas N. Ferreira
Departamento de Informática
Universidade Federal de Viçosa
Viçosa, Minas Gerais, Brazil
lucas.n.ferreira@ufv.br

ABSTRACT

Several recommender systems have been proposed to suggest courses to students based on their transcripts. In this paper, we evaluate whether these systems can be generalized to other academic activities, such as research projects and extracurricular activities. We follow previous grade-aware and content-based approaches, where course descriptions are represented by embedding vectors to predict students' grades. We extend this method by representing other academic activities as embedding vectors in the same space as the course descriptions using Jina, a recent BERT-based embedding model. This shared embedding space allows us to predict students' potential "grades" for other academic activities, different from courses, which we interpret as their level of interest. We first evaluate the Jina embeddings as features for predicting course grades compared to *TF-IDF* and Word2Vec. The results show that Jina embeddings outperform both for predicting course grades. Using the Jina embeddings as shared space for academic activities representation, we conducted a user study to evaluate our assumption that predicted grades are good proxies for student interests. We asked students to rate recommendations for courses, undergraduate research projects, and community activities with respect to novelty, diversity, personalization, accuracy, and satisfaction. The results indicate that students prefer our recommender system over a collaborative filtering baseline in all metrics.

Keywords

recommendation, content-based, embeddings, courses, extracurricular activities

1. INTRODUCTION

Course recommendation systems have been increasingly explored in educational contexts for different purposes, including improving proficiency [3], optimizing time to graduation [23], preparing students for specific target courses [13], and others. While formal courses typically fulfill the majority

of the credits in undergraduate courses, other important activities complement students' academic lives, such as undergraduate research and extracurricular community activities.

Participating in undergraduate research projects improves students' critical thinking and problem-solving skills, while also enhancing their advanced writing abilities and motivation to learn [1]. Nevertheless, many first-year students are unaware of available research opportunities and often only discover them later in their undergraduate careers [19]. Additionally, participation in extracurricular activities provides a sense of community, belonging, and well-being, while also being linked to improved academic performance and social connections. These benefits enhance students' overall university experience and positively impact student retention rates [5, 16].

This paper presents a holistic recommendation system that leverages previous grade-aware recommender systems [6, 9, 20] and neural word embeddings [29, 21, 26, 14] to recommend not only courses but also research projects and extracurricular activities. Different than previous works [20, 23], our goal is not to directly optimize GPA or reduce time to graduation, but to support students in discovering relevant and engaging academic opportunities throughout their studies, providing a more personalized academic experience. While we don't optimize GPA directly, we believe that students tend to perform better when engaging with activities aligned with their interests, which indirectly leads to academic success.

Our system is based on the assumption that students' interests can be inferred from their grades and that they are likely to perform better in courses whose topics align with their interests. Thus, the system starts by extracting embeddings from the textual content of courses with *jina-embeddings-v3*, a state-of-the-art BERT-based model. Since this model captures complex semantic and contextual information, we can also use it to generate embeddings from research projects and extracurricular activity descriptions, mapping them into the same space. Thus, we train a linear regression model to predict students' grades on courses not taken. We then apply this model to estimate the student's expected performance in other academic activities through linear regression. This estimated performance is then used as a signal for recommending courses, undergraduate research projects, and extracurricular activities.

Rafael Alves da Silva, and Lucas N. Ferreira. A Holistic Recommendation System for Higher Education Academic Activities. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 365–372. International Educational Data Mining Society (2025).

© 2025 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.
<https://doi.org/10.5281/zenodo.15870264>

We evaluate our holistic system with objective metrics and a user study. The objective metrics evaluated the error of our linear regression model when trained with different word embedding techniques: *jina-embeddings-v3*, TF-IDF, and Word2Vec. The user study measured the students' perceived quality of our recommendations compared to two baselines: a collaborative filtering approach and a random control method. Quality was assessed with 5 metrics: novelty, diversity, personalization, accuracy, and satisfaction.

Results showed that *jina-embeddings-v3* considerably outperformed TF-IDF and was slightly better than Word2Vec in the task of course grade prediction. While Word2Vec had a great trade-off between performance and model capacity, we use *jina-embeddings-v3* because it is language-independent and has better prediction performance. The user study showed that our holistic recommender is preferred by the students when compared to collaborative filtering and random baselines in almost all factors. Statistical tests showed a significant difference in the scores across the approaches in most metrics.

Our main contributions include (1) a grade-aware recommendation framework that integrates course content embeddings to improve grade prediction and (2) an expanded recommendation scope beyond courses to include research projects and extracurricular activities.

2. RELATED WORK

This work is directly related to systems that recommend higher education courses, particularly *grade-aware recommendation systems* which use predicted grades as a ranking metric. It is also related to previous works that use word embeddings to represent the course content. This section briefly discusses common approaches to these problems.

2.1 Grade-aware Course Recommendation

Collaborative filtering is a widely used approach to grade-aware course recommendation. For example, Ceyhan et al. [6] recommend courses to a given student s by predicting their grades on courses not taken. To predict the grade of a course c not taken by s , the system computes the average grade in c among the top- k most similar students to s . Similarly, Elbadrawy et al. [9] also applied collaborative filtering but combined predicted grades with course popularity rankings to recommend those with the highest chances of success. Morsy et al. [20] proposed two different methods, with one incorporating predicted grades to enhance course ranking so that a student's overall GPA is maintained or improved. Our work is similar to these previous works because we also recommend courses based on predicted grades. However, we employ a content-based approach instead of collaborative filtering. Moreover, besides courses, we also recommend other academic activities such as research projects and community service.

Other important related works include methods that do not recommend courses but focus on grade prediction. For example, Hu et al, [12] applied a linear regression model to predict student grades based on their demographic and academic background, instructor information (e.g., rank and tenure status), and course features (e.g., difficulty and credit hours). Polyzou et al, [22] showed that a single linear model

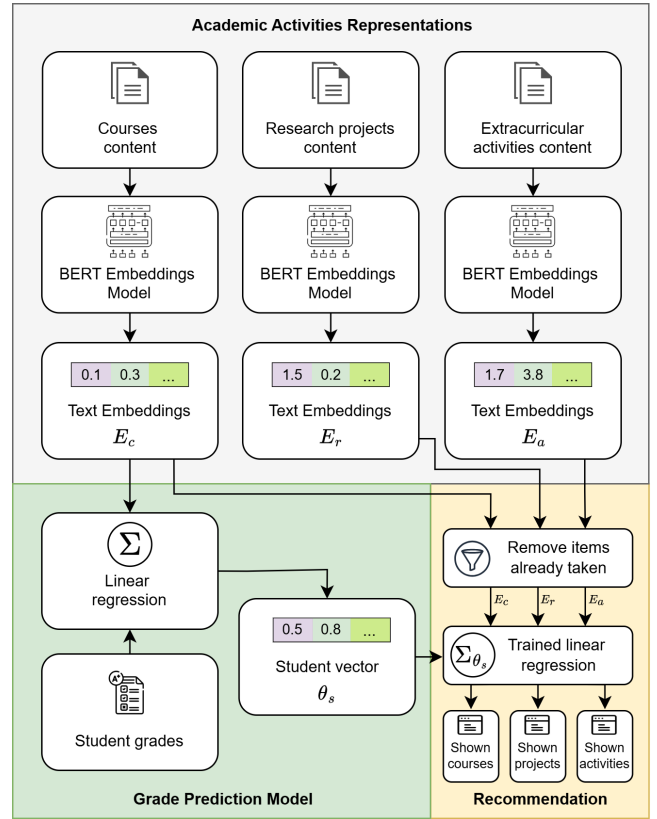


Figure 1: Overview of the holistic recommender system.

may fail to capture the various prior course combinations for programs with greater flexibility and numerous elective options. To address this limitation, they proposed student-specific regression to estimate course-specific linear models for each student. More recently, Li et al. [18] extracted semantic, syntactic, and frequency-based features to build a course-similarity graph that is fed to a graph neural network to predict the student grade. Our work differs from these previous approaches because our goal is not to predict grades, but to provide academic recommendations using grade prediction as proxies for students' interests.

2.2 Course Embeddings

Representing courses using word embeddings has recently been applied to different course recommendation systems. For example, Adilaksa et al. [2] represents course content using TF-IDF to recommend elective courses based on cosine similarity to core courses. Premalatha et al. [24] tackles the same problem, but employs a Word2Vec embedding instead. Elective courses are then recommended by predicting their knowledge domains and comparing them with the domain expertise of the student according to performance in previous core courses.

Word embeddings have also been used to represent and recommend Massive Open Online Courses (MOOC). For example, Zhou et al. [31] combined course and user embeddings to predict the most likely course for a target student with an Attention Neural Factorization Machine [29]. Mrhar et al. [21] also recommends MOOCs, but uses an LSTM net-

| |
|---|
| <p>Title: Deep Learning</p> <p>Syllabus: Introduction; Supervised Learning; Shallow Neural Networks; Deep Neural Networks; Recurrent Neural Networks; Improving Performance of Neural Networks; Convolutional Neural Networks; Attention Mechanisms; Generative Models</p> <p>Units: (1) <i>Introduction:</i> History of Artificial Intelligence; Machine Learning vs. Deep Learning; Success of Deep Learning; Deep Learning Problems; Neural Network Architectures; Technologies; (2) <i>Supervised Learning:</i> Mathematical Formulation; Classification and Regression Problems; Datasets; Hypothesis Space; Loss Functions; Training, Validation e Testing; Model Evaluation; Generalization; Overfitting and Underfitting. (3) <i>Shallow Neural Networks:</i> Linearly Separable Problems; Perceptron; Logistic Regression; Gradient Descent; Implementing Neural Networks with Numpy. (4) <i>Deep Neural Networks:</i> Linearly Non-Separable Problems; Multilayer Perceptron; Activation Functions; Backpropagation; Implementing Neural Networks with Pytorch; Recurrent Neural Networks; Natural Language Processing (NLP); Recurrent Neural Networks (RNNs); Exploding/Vanishing Gradients; Gated Recurrent Unit (GRU); Long Short Term Memory (LSTM); Bidirectional RNNs; Word Embeddings; Implementing RNNs with Pytorch. (5) <i>Convolutional Neural Networks:</i> Computer vision; Filters and Convolutions; Padding and Stride; Convolution Over Volumes; Pooling Layers; Successful Architectures; Implementing Convolutional Neural Networks with Pytorch. (6) <i>Attention Mechanisms:</i> Encoder-decoder models; Self attention; Multi-head attention Transformers. (7) <i>Generative Models Generative:</i> Adversarial Networks; Autoregressive Models; Variational Autoencoders; Diffusion Models. (8) <i>Improving Performance of Neural Networks:</i> Hyperparameter Tuning; Input and Batch Normalization; Regularization; Advanced Optimization Algorithms; Data Augmentation.</p> |
|---|

(a) A deep learning course from the Exact Sciences academic center.

| |
|---|
| <p>Title: Satellite Image Analysis with Applications in Precision Agriculture</p> <p>Research Line: Digital Image Processing</p> <p>Objective: Remote sensing can be defined as the acquisition of information about a given object without physical contact. In precision agriculture, remote sensing images have several uses: crop species classification, disease and weed monitoring, water stress detection, and soil property mapping. The objective of this project is to develop and test image processing algorithms and machine learning techniques to process multi-and/or hyperspectral satellite images.</p> <p>The problems addressed in this project are:</p> <ul style="list-style-type: none"> - Pre-processing of images to detect clouds, shadows, or other irrelevant information; - Segmentation of agricultural fields in images using Machine Learning and Deep Learning techniques; - Identification and classification of crop types and other important information for farmers within the images. |
|---|

(b) A research project on deep learning for agricultural applications.

| |
|--|
| <p>Title: II Workshop on Computational Intelligence and Statistical Learning Applied to Agriculture</p> <p>Objective: The workshop aims to bring together professionals and students (undergraduate and graduate) from Statistics, Computer Science, Agronomy, Animal Science, Genetics and Breeding, Crop Science, and related fields to promote the presentation and discussion of new methodological developments in Computational Intelligence and Statistical Learning and their applications in Agriculture.</p> |
|--|

(c) An extracurricular activity on machine learning in agriculture.

Figure 2: Examples of course, research project, and extracurricular activity from our dataset. The research and activity texts were translated into English to facilitate reading, since they are only available in Portuguese.

work to predict semantic similarity between formal courses taken by the student and available online MOOCs. Jing et al. [14] and Yin et al. [30] used the Latent Dirichlet Allocation algorithm [4] to extract topics from the course description, combining it with historical access behavior to create a user interest profile. Users and courses were then mapped into the same latent space for similarity-based recommendations. Similar representations have been applied to help course search [26] and recommend learning materials for courses, such as readings [15] and videos [25].

Most prior works leveraging embedding representations recommend courses based on similarity metrics over the embeddings themselves. Our work is different because we employ the embeddings as input to a linear regression model to predict students' grades. Moreover, most works have been applied specifically for elective courses or MOOCs. Our work is different because we focus on formal higher education courses, supporting both elective and core courses, as well as extracurricular activities.

3. A HOLISTIC RECOMMENDER SYSTEM

Our main goal with our holistic recommender system is to help personalize students' academic trajectories during their undergraduate degrees. We envision our system as an advisor that helps students find interesting courses but also other extracurricular activities such as research projects and community activities. We address this problem by bridging previous works that predict grades to recommend courses [6, 9, 20, 12, 22, 18] and that represent courses with word embeddings from textual descriptions [31, 26, 15, 25].

We propose a content-based recommender system that rep-

resents courses, research projects, and other extracurricular activities using the same BERT model, as shown in Figure 1. Our system uses an independent regression model for each student, trained to predict their grades on courses not taken. At recommendation time, since all activities are in the same space, our system employs the learned regression model θ_s to predict the student s performance on all available activities. Thus, we can recommend the activities that have a higher predicted performance to the student s .

3.1 Dataset

We trained our grade prediction model with a collection of undergraduate courses from our university, the Universidade Federal de Viçosa (UFV), a public research university in Minas Gerais state, Brazil. UFV offers 45 majors organized into four broad academic centers: Agricultural Sciences (CCA), Biological and Health Sciences (CCB), Exact Sciences (CCE), and Humanities (CCH). Our dataset contains 14 years (2010–2024) of academic records, including 1.6 million student transcripts from 4,164 courses of 45 programs. The grades are measured on a scale from 0 to 100. Although student grades are treated as confidential information, details regarding course content, research projects, and extracurricular activities are publicly available on the university website.

Course documents have an average of 195 words, including the course name (4 words), a brief syllabus summarizing the main topics (40 words), and the detailed content of each unit (151 words including all units). All course descriptions are available in both Brazilian Portuguese and English. Figure 2a shows an example of a deep learning course in our dataset. The formatting of the text has been removed to illustrate

how the embedding model will process the course inputs.

Besides courses, our dataset also comprises 1,859 research projects and 11,206 extracurricular activities, which are not used for training the grade prediction model but to evaluate our recommender system in our user study. Research project documents have an average of 110 words, comprising the project title (14 words), its research line (5 words), and a summary of objectives (91 words). Extracurricular activity documents have a similar structure, with an average of 39 words, including an activity title (9 words) and a summary of its objectives (averaging 30 words). Research projects and extracurricular activities descriptions are available only in Brazilian Portuguese.

3.2 Academic Activities Representations

Considering that most formal academic activities are typically registered in universities as textual documents, we represent them in a shared embedding space using a pre-trained BERT model [7]. Particularly, we use *jina-embeddings-v3* [27] to extract embeddings, because it is a state-of-the-art model that supports 100 languages and a context size of 8192 tokens. This context size is large enough to represent academic activity textual descriptions in general, including courses. Additionally, this model supports Matryoshka Representation Learning [17] that allows reducing the embedding dimensionality without compromising performance. It is important to highlight that by representing academic activities with *jina-embeddings-v3*, our recommender system is practically language-agnostic, supporting the same 100 languages. From now on, for simplicity, we will refer to this model as Jina.

While Jina embeddings can be produced for any textual document, to make meaningful recommendations, our method needs to represent academic activities closely to their related courses in the embedding space. Thus, we assume that academic activities are described by a coherent and informative text that shares contextual or terminological overlap with course topics. For example, a research project on neural networks (Figure 2b) or an activity focused on artificial intelligence (Figure 2c) will have related terms such as “Learning”, “Image”, and “Classification”, and hence will be represented closely to deep learning courses in the embedding space (Figure 2a).

To evaluate whether the documents in our dataset have a meaningful embedding space, we sampled 3,000 academic activities in total (1,000 courses, 1,000 research projects, and 1,000 extracurricular activities) uniformly distributed among the 4 academic centers in our university. Figure 3 shows a plot of the embeddings in a two-dimensional space reduced using the t-SNE algorithm [28]. One can see that the produced shared embedding space places courses, research projects, and extracurricular activities from similar areas close together. For example, all Exact Science academic activities (in red) form a cluster at the center-top of the reduced embedding space.

3.3 Grade Prediction and Recommendations

We train a linear regression model $y = h_{\theta}(x)$ for each student to predict their grades y on courses represented as word embeddings x . We use linear regression instead of a more

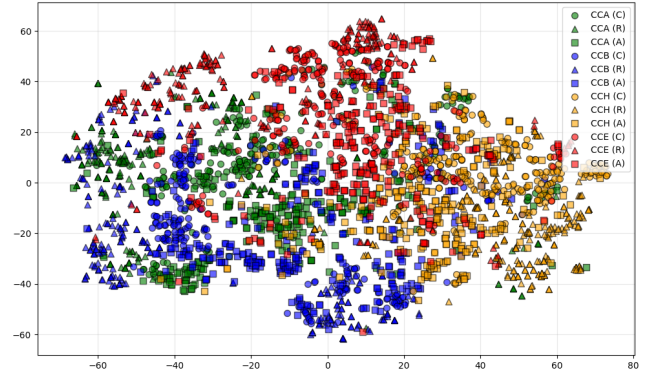


Figure 3: Embeddings from courses (C), research projects (R) and extracurricular activities (A).

complex model (e.g., a neural network) because we train a single model per student, and thus we don’t have much data for each model. Having a single model per student has the benefit that we can interpret the weights of the linear regression θ as a vector that represents the student’s interests/academic strengths.

Since Jina embeddings group research projects and extracurricular activities close to courses in space based on knowledge areas, as shown in Figure 3, we can use the model $y = h_{\theta}(x)$ trained on courses to predict grades on other activities x . Thus, we can generalize the predicted “grade” y as the student expected performance on that activity.

Our holistic recommender system uses the trained linear regression to suggest interesting courses, research projects, and extracurricular activities to a given student s . These three different items are recommended independently of each other, but using the same procedure, as shown in the area shaded in yellow in Figure 1. The first step is to remove the activities already taken by s , to make sure we only recommend new ones. Then, we apply the trained linear regression to predict the grade y for each course $y_c = \theta_s \cdot E_c$, research project $y_r = \theta_s \cdot E_r$, and extracurricular activity $y_a = \theta_s \cdot E_a$ in our dataset. Finally, we show to student s three lists of items, one for courses, one for research projects, and one for extracurricular activities, where each list has the top- k items in our dataset with respect to predicted grade.

While the system relies on historical grade data, the potential cold-start problem is naturally mitigated by the structure of undergraduate curricula, which typically include a fixed set of courses in the first term (quarter, semester, etc.). As a result, our recommender system begins suggesting activities only from the second term onward. By then, students have usually completed enough courses to allow for meaningful recommendations, and the quality of these recommendations is expected to improve as more data becomes available throughout their academic journey.

4. EMPIRICAL EVALUATION

We evaluate our holistic recommendation systems with two experiments. First, we compare the error of the logistic regression when predicting course grades with Jina embeddings against two simpler approaches: TF-IDF and Word2Vec.

Second, we conducted a user study to evaluate the quality of the recommended academic activities with respect to novelty, diversity, personalization, accuracy, and satisfaction.

4.1 Experimental Setup

All variations of linear regression were trained using gradient descent with the same hyperparameter configuration: 100 epochs, learning rate $\eta = 0.01$, and L2 regularization $\lambda = 0.1$. Since we want our model to predict grades of courses not taken by a student, we split the students' transcripts over time, using 80% of their first course grades for training and the remaining 20% for testing. If a student took a course more than once, we considered only the most recent grade, as it best represents the student's current proficiency in that course. We also removed records of courses that do not use numerical grades for evaluation.

We normalized grades using the z-score within each class¹ since we observed differences in average grades across academic centers. For example, humanities courses have higher average grades than those in the exact sciences. Since our methodology infers a student's representation based on their grades and course content, these variations could introduce inconsistencies, especially because many programs from different academic centers share courses.

All course embeddings were represented with 768-dimensional vectors. Jina includes five LoRA (Low-Rank Adaptation) adapters [11] for specific tasks: retrieval-query, retrieval-passage, separation, classification, and text-matching. Considering our recommendation application, we evaluated Jina with the separation, classification, and text-matching adapters. We also reported the Jina performance without any adapters to measure the impact of each adapter.

Since TF-IDF embeddings have a dimensionality equal to the number of words across all documents (i.e., our vocabulary), they result in high-dimensional vectors. Therefore, we reduced its dimensionality by selecting only the most frequent terms to match the same 768 dimensionality as Jina. To measure the impact of this reduction, we also evaluated TF-IDF at full dimensionality.

Before passing these texts through TF-IDF and Word2Vec, we pre-processed them removing stop words and digits, and applying lemmatization. For Word2Vec, we evaluated both the Continuous Bag of Words (CBOW) and Skip-Gram models. Since the research projects and extracurricular activities in our dataset are all in Brazilian Portuguese, we used embeddings pre-trained in a large Brazilian Portuguese corpus, including Wikipedia, Google News, and scientific texts [10].

4.2 Objective Evaluation

We evaluated the logistic regression models on the testing set with Mean Absolute Error (MAE), Root Mean-Square Error (RMSE), and R^2 Score between the predicted and ground truth grades. Since grades are measured on a scale from 0 to 100, MAE and RMSE are on the same interval, where lower values indicate more accurate predictions. Table 1 shows the MAE, RMSE, and R^2 score for each logistic

¹A class is the set of grades for a course offered in a specific year and semester by a professor.

Table 1: Error for each course embedding. A value in bold is the best score for a particular quality metric (column).

| Embedding | MAE | RMSE | R^2 |
|-----------------------|--------------|--------------|--------------|
| TF-IDF (full) | 55.03 | 57.41 | -7.09 |
| TF-IDF (reduced) | 48.34 | 51.68 | -7.54 |
| Word2Vec (CBOW) | 14.90 | 19.48 | -0.21 |
| Word2Vec (Skip-Gram) | 13.78 | 18.56 | -0.10 |
| Jina (no task) | 17.14 | 21.08 | -0.42 |
| Jina (text-matching) | 19.21 | 23.39 | -0.75 |
| Jina (classification) | 13.62 | 17.83 | -0.01 |
| Jina (separation) | 13.57 | 17.81 | -0.01 |

regression model we evaluated. Jina (separation) and Jina (classification) embeddings outperformed all the other methods, which is evidenced by their lowest and practically equal MAE, RMSE, and R^2 scores. These results suggest that embeddings that better capture academic context lead to more accurate grade predictions.

Word2Vec (Skip-Gram) and Word2Vec (CBOW) were the second and third-best approaches, respectively, with an MAE very close to Jina (separation) and Jina (classification), but RMSE slightly worse. This is a surprising result considering that Word2Vec is relatively simpler than Jina. We hypothesize that Word2Vec's good performance comes from its specific training with a Brazilian Portuguese corpus. While this helps improve performance, it limits its embeddings to a single language.

Jina (text-matching) and Jina (no task) performed worse than Jina (separation) and Jina (classification) and than the two Word2Vec models, showing that LoRA adapters considerably impact the model grade prediction performance. The TF-IDF approaches had the worst performance among all embedding approaches. This is an expected result since this is a very simple method based on word frequencies. The reduction of TF-IDF dimensionality resulted in a lower error but still considerably higher than Word2Vec and Jina.

4.3 User Study

We conducted our user study as a within-subject study where a participant is asked to evaluate our holistic recommendation system (with Jina) and two other baselines: a simple collaborative-filtering approach and a random recommender, which we use as a control method. We chose collaborative filtering as our baseline because it is one of the most common approaches for grade-aware course recommendation. The user study was approved by the research ethics committee of our university.

Our collaborative filtering baseline followed the work of Ceyhan et al. [6], where students are represented as a vector of their grades in a student \times course grade matrix. Differently from their work, we used the cosine similarity. Since only courses have grades associated, for research projects and extracurricular activities, we used a binary value, associating 1 if the student has participated and 0 otherwise.

We adapted the methodology of Ekstrand et al. [8] to evalu-

ate the quality of our recommendations. Thus, each participant was asked to evaluate each method according to five metrics using a five-point Likert Scale (1-5), with higher values indicating a better perception of quality:

1. Novelty (Nov.): the ability to recommend interesting items with which the student is unfamiliar.
2. Diversity (Div.): the ability to recommend items covering a broad range of topics.
3. Personalization (Per.): the student’s perception that the recommender understands their interests and preferences.
4. Accuracy (Acc.): the suitability of the recommended items for the student’s current academic progress.
5. Satisfaction (Sat.): the student’s overall satisfaction with the recommender and their perception of its usefulness.

Each participant evaluated three recommendation lists per method, one at a time: first a list of courses, then a list of research projects, and finally a list of extracurricular activities. Thus, each participant evaluated 9 recommendations in total. The methods were presented in a random order to avoid ordering bias, and the recommended lists were shown in Brazilian Portuguese. The entire study was designed to be completed in 10 to 15 minutes.

We invited undergraduate students currently enrolled at our university to participate via email. We received responses from 71 students, but 9 did not complete the questionnaire rating the three systems. Thus, we had answers from 62 students enrolled in 12 undergraduate courses within the four academic center: CCH (37%), CCE (27%), CCB (19%), and CCA (16%). On average, they are 25 ± 7 years old and are in the 6 ± 3 semester of their undergraduate course. All participants are Brazilians and all recommended activities have descriptions in Portuguese.

Table 2 shows the average scores for our holistic recommender system and the two baselines on the course recommendation task. For each condition, we ran a Friedman test across methods and if significant, a pairwise Nemenyi. All methods showed significant differences across recommenders ($p\text{-value} \leq 0.05$) based on the Friedman test. Our system is significantly better at recommending courses than the two baselines in almost all factors according to the Nemenyi post-hoc test, except for diversity, in which the control random method performed better. It is expected that the random control baseline had a higher perceived diversity since it has a high chance of recommending courses very different from the student’s major. However, high diversity alone is not typically helpful in educational contexts, since a good recommendation involves diverse but accurate and personalized courses.

The results for the research projects and extracurricular activities recommendation tasks followed a similar pattern. The Friedman test indicated significant differences across recommenders for all factors except diversity. However, in

Table 2: Average scores for courses (in 1-5). For each metric, the higher the score, the better. Values with shared letter superscripts ^{abc} in a given row were not significantly ($p\text{-value} > 0.05$) different in pairwise post-hoc comparisons. A value in bold is the best score for a particular quality metric (row).

| | Random | Collab. Filtering | Holistic |
|-------------|-------------------------------------|-------------------|-------------------------------------|
| <i>Nov.</i> | $2.73^a \pm 1.29$ | $2.75^a \pm 1.23$ | $3.30^b \pm 1.17$ |
| <i>Div.</i> | $4.02^a \pm 1.18$ | $3.40^a \pm 1.44$ | $3.48^a \pm 1.30$ |
| <i>Per.</i> | $1.52^a \pm 0.76$ | $1.98^a \pm 1.09$ | $3.60^b \pm 1.40$ |
| <i>Acc.</i> | $2.00^a \pm 1.02$ | $2.48^a \pm 1.28$ | $3.66^b \pm 1.38$ |
| <i>Sat.</i> | $1.77^a \pm 0.91$ | $2.02^a \pm 1.09$ | $3.05^b \pm 1.38$ |

Table 3: Average scores for research projects (in 1-5). Values with shared letter superscripts ^{abc} in a given row were not significantly ($p\text{-value} > 0.05$) different in pairwise post-hoc comparisons. A value in bold is the best score for a particular quality metric (row).

| | Random | Collab. Filtering | Holistic |
|-------------|-------------------|-------------------|-------------------------------------|
| <i>Nov.</i> | $3.23^a \pm 1.46$ | $3.87^b \pm 1.22$ | $4.13^b \pm 1.19$ |
| <i>Div.</i> | $3.77^a \pm 1.21$ | $4.02^a \pm 1.02$ | $4.05^a \pm 1.08$ |
| <i>Per.</i> | $1.74^a \pm 1.02$ | $3.16^b \pm 1.36$ | $3.50^b \pm 1.34$ |
| <i>Acc.</i> | $2.16^a \pm 1.30$ | $3.44^b \pm 1.26$ | $3.81^b \pm 1.16$ |
| <i>Sat.</i> | $1.87^a \pm 1.09$ | $3.19^b \pm 1.37$ | $3.65^b \pm 1.38$ |

these two domains, our holistic system also outperformed the others in diversity, as shown in Tables 3 and 4. We believe this is related to the nature of these activities: although research projects and extracurricular activities have specific objectives, they often span multiple knowledge areas, which may lead to a greater perception of diversity compared to courses, which typically focus on specific topics.

For research projects, the pairwise Nemenyi post-hoc test showed no significant differences between the holistic and collaborative filtering recommenders. For extracurricular activities, there is also no significant difference between the holistic and the collaborative filtering recommender, except for novelty.

Overall, these findings suggest that our content-based holistic recommender system can be better than traditional collaborative filtering at recommending formal undergraduate courses. Considering its higher performance on research projects and extracurricular activity recommendations, we also confirm our hypothesis that predicted grades are good proxies for student interests.

5. CONCLUSIONS AND FUTURE WORK

This paper presented a content-based holistic recommendation system to suggest courses, research projects, and extracurricular activities to undergraduate students. Building upon previous work, our approach combined grade-aware recommendations and course content embeddings, representing research projects and academic activities in the same space as courses. A user study showed that our system outperformed a collaborative filtering approach in recommending the three types of academic activities.

Table 4: Average scores for extracurricular activities (in 1-5). Values with shared letter superscripts ^{abc} in a given row were not significantly (p-value > 0.05) different in pairwise post-hoc comparisons. A value in bold is the best score for a particular quality metric (row).

| | Random | Collab. Filtering | Holistic |
|------|--------------------------|--------------------------|--------------------------------|
| Nov. | 3.18 ^a ± 1.41 | 3.37 ^a ± 1.44 | 4.19^b ± 1.07 |
| Div. | 3.77 ^a ± 1.29 | 3.63 ^a ± 1.28 | 4.15^a ± 1.01 |
| Per. | 1.81 ^a ± 1.07 | 3.50 ^b ± 1.36 | 3.87^b ± 1.23 |
| Sui. | 2.31 ^a ± 1.28 | 3.66 ^b ± 1.28 | 3.98^b ± 1.03 |
| Sat. | 1.94 ^a ± 1.25 | 3.35 ^b ± 1.43 | 4.13^b ± 1.06 |

In theory, our system can recommend any type of textual document. Thus, future work can expand the scope of recommendations to other academic activities such as events and mini-courses, as long as they have meaningful textual content. Other avenues for future exploration are to expand our user study to students/activities in different languages, investigate our methodology in a graduate school context, and evaluate how our holistic recommender system affects students' academic experiences over time.

6. ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Finance Code 001

7. REFERENCES

- [1] Y. A. Adebisi. Undergraduate students' involvement in research: Values, benefits, barriers and recommendations. *Annals of medicine and surgery*, 81:104384, 2022.
- [2] Y. Adilaksa and A. Musdholifah. Recommendation system for elective courses using content-based filtering and weighted cosine similarity. In *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 51–55. IEEE, 2021.
- [3] B. Bakhshinategh, G. Spanakis, O. Zaiane, and S. ElAtia. A course recommender system based on graduating attributes. In *International Conference on Computer Supported Education*, volume 2, pages 347–354. SCITEPRESS, 2017.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [5] P. Buckley and P. Lee. The impact of extra-curricular activity on the student experience. *Active Learning in Higher Education*, 22(1):37–48, 2021.
- [6] M. Ceyhan, S. Okyay, Y. Kartal, and N. Adar. The prediction of student grades using collaborative filtering in a course recommender system. In *2021 5th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, pages 177–181. IEEE, 2021.
- [7] J. Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [8] M. D. Ekstrand, F. M. Harper, M. C. Willemsen, and J. A. Konstan. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 161–168, 2014.
- [9] A. Elbadrawy and G. Karypis. Domain-aware grade prediction and top-n course recommendation. In *Proceedings of the 10th ACM conference on recommender systems*, pages 183–190, 2016.
- [10] N. Hartmann, E. Fonseca, C. Shulby, M. Treviso, J. Rodrigues, and S. Aluisio. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. *arXiv preprint arXiv:1708.06025*, 2017.
- [11] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.
- [12] Q. Hu, A. Polyzou, G. Karypis, and H. Rangwala. Enriching course-specific regression models with content features for grade prediction. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 504–513. IEEE, 2017.
- [13] W. Jiang, Z. A. Pardos, and Q. Wei. Goal-based course recommendation. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 36–45, 2019.
- [14] X. Jing and J. Tang. Guess you like: course recommendation in moocs. In *Proceedings of the international conference on web intelligence*, pages 783–789, 2017.
- [15] M. R. Julianti, Y. Heryadi, B. Yulianto, and W. Budiharto. Recommendation system model for personalized learning in higher education using content-based filtering method. In *2022 International Conference on Information Management and Technology (ICIMTech)*, pages 1–6. IEEE, 2022.
- [16] A. E. King, F. A. McQuarrie, and S. M. Brigham. Exploring the relationship between student success and participation in extracurricular activities. *SCHOLE: A Journal of Leisure Studies and Recreation Education*, 36(1-2):42–58, 2021.
- [17] A. Kusupati, G. Bhatt, A. Rege, M. Wallingford, A. Sinha, V. Ramanujan, W. Howard-Snyder, K. Chen, S. Kakade, P. Jain, et al. Matryoshka representation learning. *Advances in Neural Information Processing Systems*, 35:30233–30249, 2022.
- [18] J. Li, S. Supraja, W. Qiu, and A. W. Khong. Grade prediction via prior grades and text mining on course descriptions: Course outlines and intended learning outcomes. *International Educational Data Mining Society*, 2022.
- [19] C. R. Madan, B. D. Teitge, et al. The benefits of undergraduate research: The student's perspective. *The mentor: An academic advising journal*, 15:1–3, 2013.
- [20] S. Morsy and G. Karypis. Will this course increase or decrease your gpa? towards grade-aware course recommendation. *arXiv preprint arXiv:1904.11798*, 2019.
- [21] K. Mrhar and M. Abik. Toward a deep recommender

- system for moocs platforms. In *Proceedings of the 3rd International Conference on Advances in Artificial Intelligence*, pages 173–177, 2019.
- [22] A. Polyzou and G. Karypis. Grade prediction with course and student specific models. In *Advances in Knowledge Discovery and Data Mining: 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19-22, 2016, Proceedings, Part I 20*, pages 89–101. Springer, 2016.
 - [23] M. Premalatha and V. Viswanathan. Course sequence recommendation with course difficulty index using subset sum approximation algorithms. *Cybernetics and Information Technologies*, 19(3):25–44, 2019.
 - [24] M. Premalatha, V. Viswanathan, and L. Čepová. Application of semantic analysis and lstm-gru in developing a personalized course recommendation system. *Applied Sciences*, 12(21):10792, 2022.
 - [25] F. Ramadhan and A. Musdholifah. Online learning video recommendation system based on course and syllabus using content-based filtering. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3):265–274, 2021.
 - [26] K. K. San, K. E. E. Chaw, and T. T. Nwe. Attention-based neural networks with tf-idf features and bert embeddings for enhanced course recommendation system. In *2024 5th International Conference on Advanced Information Technologies (ICAIT)*, pages 1–6. IEEE, 2024.
 - [27] S. Sturua, I. Mohr, M. K. Akram, M. Günther, B. Wang, M. Krimmel, F. Wang, G. Mastrapas, A. Koukounas, N. Wang, et al. jina-embeddings-v3: Multilingual embeddings with task lora. *arXiv preprint arXiv:2409.10173*, 2024.
 - [28] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
 - [29] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, and T.-S. Chua. Attentional factorization machines: Learning the weight of feature interactions via attention networks. *arXiv preprint arXiv:1708.04617*, 2017.
 - [30] S. Yin, K. Yang, and H. Wang. A mooc courses recommendation system based on learning behaviours. In *Proceedings of the ACM Turing Celebration Conference-China*, pages 133–137, 2020.
 - [31] W. Zhou, X. Zhou, J. He, Y. Nie, Q. Chen, and G. Kang. Online course recommendation by exploring user interaction and course description. In *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 2973–2978. IEEE, 2024.