

Framework para predições e recomendações em dados acadêmicos

Bruno Bastos Stoll¹, Davidson Cury¹, Crediné Silva de Menezes²

¹ Departamento de Informática – Universidade Federal do Espírito Santo (UFES)

² Programa de Pós-Graduação em Educação – Universidade Federal do Rio Grande do Sul (UFRGS).

brunobstoll@hotmail.com, {dede, credine}@inf.ufes.br

Abstract. *In virtual environments, records of student productions, activities, and interactions are collected automatically, enabling analysis of development and learning. With this, it is possible to identify patterns of how students learn. Therefore, favoring the appropriate basis for the appropriate recommendations. The goal is to present a framework for generating messages of recommendations based on predictive systems in an innovative way. To validate the work was performed the analysis of a public data set.*

Resumo. *Nos ambientes virtuais, os registros das produções, atividades e interações dos alunos são coletadas automaticamente, possibilitando a análise do desenvolvimento e aprendizado. Com isso, oportuniza-se a identificação de padrões de como os alunos aprendem. Logo, favorecendo a base apropriada para as devidas recomendações. O objetivo é apresentar um framework para geração de mensagens de recomendações baseado em sistemas preditivos, de forma inovadora. Para validar o trabalho foi realizada a análise de um conjunto de dados público.*

1. Introdução

Na educação a distância (EaD), a percepção das individualidades e a intervenção no processo de aprendizagem de cada aluno, para professores torna-se uma tarefa desafiadora. As interações dos alunos devem ser periodicamente analisadas pelos professores, que por vezes, se torna uma tarefa impossível, pois o volume de dados poderá ser muito grande. O tempo imposto aos professores pelas instituições poderá ser outro problema. Por isso, entre a coleta de dados, a análise e a intervenção no aprendizado dos alunos, o professor poderá ter pouco tempo.

Esses dados contêm conhecimentos importantes sobre como os alunos aprendem e como eles interagem com o sistema de aprendizado. Contudo, analisar e monitorar esse volume de interações entre professor, alunos e conteúdo é uma importante e complexa tarefa do professor no processo de aprendizagem online [Medeiros 2013]. Por meio do uso de análise desses dados é possível apontar os diferentes fatores que influenciam o comportamento do aluno e identificar aspectos sutis, muitas vezes imperceptíveis por professores e gestores de cursos em EaD [Verbert 2004]. Em consequência disso, é fundamental personalizar o processo de aprendizado para cada aluno, conforme sua preferência e necessidade.

O objetivo geral deste artigo é apresentar um framework para geração de mensagens de recomendações baseado em sistemas preditivos. E como objetivo específico, avançar na pesquisa de intervenção no processo de aprendizado, usando as características coletadas em um ambiente de aprendizado virtual (AVA). A solução proposta é um framework focado em recomendação e predição, com uma abordagem dinâmica ao contexto dos dados. Desenvolvemos um sistema que use este framework e analisamos um conjunto de dados público. Na metodologia, usamos a abordagem qualitativa, de natureza aplicada, com o objetivo descritivo e com procedimentos experimentais.

O artigo foi organizado em 6 seções. Na seção 2 descreve fundamentação teórica com conceitos sobre Análise de Aprendizado, Aprendizado de Máquina e Sistemas de Recomendações. Na seção 3 são apresentados os trabalhos correlatos. Na seção 4 é descrita a proposta. Na seção 5 são apresentados os experimentos. E na seção 6 é apresentada a conclusão e são descritos os trabalhos futuros.

2. Fundamentação Teórica

Há diferentes definições sobre Análise de Aprendizado, do inglês *Learning Analytics* (LA). Segundo [Elias 2011] é descrito como um campo emergente, no qual as ferramentas de análise de dados são usadas para melhorar o processo de aprendizado e a educação. O [Long 2011] descreve como a mediação, coleta, análise e publicação de dados sobre os alunos e seus contextos, com o objetivo de compreender e otimizar o aprendizado e os ambientes em que ele ocorre. Já [Brown 2012] descreve como o uso de dados e modelos para prever o progresso e o desempenho do aluno, e a capacidade de agir com base nessas informações. E [Johnson et al. 2016] descreve LA como a aplicação de Web Analytics, tendo em vista, o perfil do aluno, um processo de coleta e análise de detalhes de interações individuais, em atividades de aprendizado online com o objetivo de melhorar a pedagogia, capacitar a aprendizagem ativa, atingir populações de estudantes em risco e avaliar fatores que afetam a conclusão e o sucesso na aprendizagem.

O aprendizado de máquina é uma subárea da Inteligência Artificial, com o objetivo de desenvolver técnicas computacionais sobre o aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Sistema, que usa aprendizado de máquina, toma decisões baseada em experiências acumuladas através de soluções bem-sucedidas de um problema anterior [Richert 2013]. Existem vários algoritmos de aprendizado de máquina, destacamos os algoritmos que usamos em nossa pesquisa. A árvore de decisão, do inglês *Decision Tree*, é um modelo de árvore recursivo baseado em partição para prever qual é o rótulo de uma instância [Zaki et al. 2014]. O K Vizinhos Próximos, do inglês *K-Nearest Neighbors* (kNN) é um algoritmo de aprendizado de máquina que calcula a distância dos k (instâncias) vizinhos mais próximos [Zaki et al. 2014]. A Máquina de Suporte a Vetor, do inglês *Support Vector Machine* (SVM), é um algoritmo de aprendizado de máquinas com o objetivo de encontrar o hiperplano ideal que maximiza a margem entre as classes [Zaki et al. 2014]. Através de seu kernel, SVM pode ser usado para encontrar limites não-lineares [Zaki et al. 2014]. O classificador de Bayes, que é uma abordagem de classificação probabilística e usa o teorema de Bayes para predizer valores [Zaki et al. 2014]. O Perceptron em Multicamadas, em inglês *Multi-Layer Perceptron* (MLP), é uma rede

neural que apresenta uma ou mais camadas intermediárias de neurônios e uma camada de saída [Carvalho et al. 2011].

Com o aumento da quantidade de conteúdo disponível na internet é comum se deparar com uma diversidade grande de opções. Por isso, temos dificuldade em escolher uma entre várias alternativas de conteúdo. Sistemas de recomendação (SR) vêm a auxiliar na solução desse problema. Segundo [Costa et al. 2013], SR são técnicas que auxiliam usuários na busca e seleção de um conteúdo focado em seu perfil, funcionando literalmente como filtros de informações. O princípio técnico de um SR é o uso de um algoritmo para combinar dados de contexto e entrada para produzir sugestões aos usuários visando a geração de recomendações.

3. Trabalhos Relacionados

Os trabalhos descritos abaixo propõem técnicas de análise de dados educacionais, técnicas de predição ou descrição, recomendações e visualização, propondo diferentes soluções e abordagens.

Há pesquisas que usam técnicas de mineração de dados voltadas a identificar padrões que levam alunos a evadir, por meio de dados acadêmicos. O trabalho [Pascoal et al. 2016] propõe uma abordagem para auxiliar instituições na tomada de decisão, para combater a evasão através de dados acadêmicos e socioeconômicos. O trabalho [Oliveira Jr et al. 2015] propõe uma abordagem computacional para a identificação de padrões a serem utilizados na análise da evasão de estudantes em cursos presenciais de graduação, a fim de auxiliar os tomadores de decisão das instituições de ensino.

Os trabalhos voltados à previsão de desempenho de alunos usando algoritmos de aprendizado de máquina são o foco de muitas pesquisas. O trabalho [Gotardo et al. 2013] apresenta uma abordagem que usa algoritmos de aprendizado acoplados oferecendo recomendação sobre o desempenho do aluno. O trabalho [de Brito et al. 2014] propõe a previsão de desempenho dos alunos no primeiro período do curso de Ciência da Computação, da UFPB, através das suas notas de ingresso no vestibular. O trabalho [Gottardo et al. 2012] propõe o uso de inferências sobre o desempenho de estudantes, a partir de dados coletados em séries temporais, com o objetivo principal de investigar a viabilidade de obtenção dessas informações, em etapas iniciais de realização do curso de forma a apoiar a tomada de ações proativas. Já o trabalho [Vieira and Veiga 2014] tem o foco em métodos de árvores de decisão para a previsão de desempenho de alunos. O trabalho [Govindarajan et al. 2015] faz uso de dados acadêmicos na finalidade de reduzir taxas de insucesso de aprendizado.

Existem trabalhos que fazem uso de ferramentas prontas ou soluções de análise de dados em ambientes acadêmicos. O trabalho [Arnold and Pistilli 2012] propõe a oportunidade do uso de feedback em tempo real aos professores. O trabalho [Martin and Ndoye 2016] faz uso das ferramentas Tableau e Many Eyes para analisar dados acadêmicos. O trabalho [Dyckhoff et al. 2012] apresenta detalhes teóricos sobre a implementação de uma ferramenta, que permita aos professores explorarem e correlacionar o uso de objetos de aprendizado e o comportamento de alunos com resultados de avaliações com base em indicadores gráficos. O trabalho [Greller and Drachsler 2012] propõe um framework genérico, dando suporte a práticas educacionais e discute sobre as dimensões chave da análise do aprendizado. O trabalho [Padilha and

Souza 2017] apresenta o uso das ferramentas Gephi e RapidMiner para análise de interações entre alunos e professores.

No trabalho [Clow 2012] foi descrito sobre a importância do fechamento do ciclo da análise dos dados acadêmicos, com o uso de feedback através de intervenções apropriadas. Segundo [Moissa et al. 2014], em seu mapeamento sistemático da literatura, poucos trabalhos mencionam a intervenção, no entanto, há uma grande quantidade de trabalhos que não possibilitam intervenção alguma. Também foi apresentado que soluções de Predição e Intervenção não estão relacionadas aos tipos de intervenção de Recomendações e Feedback, sugerindo que esse pode ser um aspecto a ser abordado pela área de análise de aprendizado.

Em geral, as propostas na literatura são restritas e apresentam apenas experimentos ou um escopo fechado e reduzido. A maioria dos trabalhos não combinam as diversas técnicas abordadas pelo nosso trabalho. Os trabalhos que têm sistemas de previsão usam somente um algoritmo, enquanto o nosso trabalho se utiliza de vários algoritmos e seleciona dinamicamente o mais adequado para o conjunto de dados. Já outros trabalhos são aplicados em uma única instituição de ensino não dando possibilidade de ler outras bases de dados.

4. Descrição da Proposta

O framework implementado em nossa pesquisa possui dois módulos, o primeiro faz a previsão de desempenho dos alunos e o segundo gera as mensagens de recomendações, que serão sugeridas aos professores. Conforme apresentado na figura 1, o framework deverá receber dados acadêmicos previamente tratados, devolverá a previsão do desempenho dos alunos e sugerir as mensagens de recomendações.

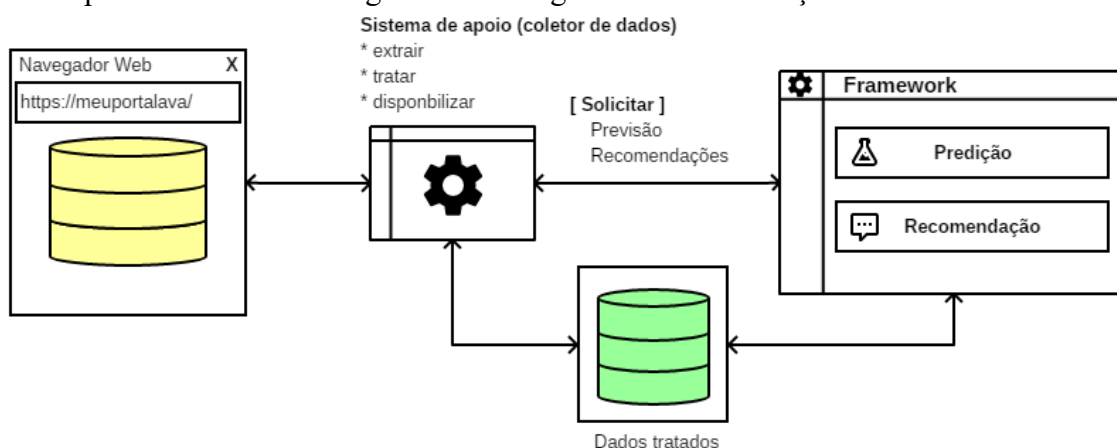


Figura 1. Arquitetura proposta

No módulo de predição, de forma antecipada, é realizada previsão dos resultados acadêmicos dos alunos, através de algoritmos de aprendizado de máquina. Podendo classificar os alunos em “sucesso” ou “em risco” acadêmico. A classificação “sucesso” acadêmico ocorre a previsão é de que o aluno atenda aos propósitos educacionais de aprender e tirar notas boas. Já o resultado “em risco” ocorre quando o aluno dá sinais que não está aprendendo, com probabilidade de não tirar notas boas, ou até mesmo desistir (evadir) de uma determinada disciplina ou do curso.

O [Pakalra and Olprod 2018] descreve que para um problema específico, vários algoritmos podem ser apropriados, no entanto, um algoritmo pode ser mais adequado do

que outros, mas nem sempre é possível saber com antecedência qual é a melhor opção. Como o framework está preparado para ler diferentes estruturas de dados acadêmicos, usamos a seleção automática do algoritmo que tiver melhor acurácia, dentre as opções: Decision Tree, KNN, SVM e MLP. Esses algoritmos foram selecionados, pois cada um tem uma abordagem diferente.

Serão usados os dados históricos de turmas anteriores para fazer o treinamento e teste dos algoritmos preditivos. Para avaliar a acurácia dos algoritmos usamos a técnica de validação cruzada, onde é feita a divisão dos dados já rotulados para treinamento e teste para verificar a quantidade de acertos [Carvalho et al. 2011]. O algoritmo selecionado, que tiver a melhor acurácia, classificará instâncias ainda não rotuladas.

Segundo [Brown 2012] existem as categorias de intervenções automáticas e a semiautomáticas. Na intervenção automática o sistema realiza ações pré-determinada automaticamente com pouca ou nenhuma participação. Já as intervenções semiautomáticas são alertas ou indicadores relevantes dos padrões de aprendizagem enviados para os envolvidos no processo de ensino. Optamos pela intervenção semiautomática, através do sistema de apoio, para que o professor tenha a oportunidade de visualizar e alterar as mensagens antes de enviá-las ou simplesmente as ignorar, caso não sejam apropriadas.

Algoritmo 1. Gerar mensagem - Características	
1.	Entrada: instanciaAluno
2.	Variáveis: conta, ResultadoAluno, msgRecomendação,
3.	listaCaracteristica, característica, itemMsg;
4.	Início
5.	ResultadoAluno = PreverResultado(instanciaAluno);
6.	Se ResultadoAluno igual 'sucesso' Então:
7.	sair();
8.	msgRecomendação = [];
9.	listaCaracteristica = ListarCaracPorImportancia();
10.	Para característica em listaCaracteristica:
11.	MudarCaracteristica(instanciaAluno, característica);
12.	itemMsg = GerarItemMsg(instanciaAluno, característica);
13.	AdicionarMsg(msgRecomendação, itemMsg);
14.	conta = conta + 1;
15.	Se conta maior 5 Então:
16.	ResultadoAluno = PreverResultado(instanciaAluno);
17.	Se ResultadoAluno igual 'sucesso' Então:
18.	Parar
19.	
20.	Retornar msgRecomendação
21.	Fim

No módulo de recomendações, para cada aluno classificado “em risco”, o framework gera uma mensagem ao professor. Essas mensagens são geradas através da métrica derivada do processamento do módulo de Predição. Essa métrica é a importância de cada característica para classificação das instâncias.

Usamos cinco características mais relevantes inicialmente, podendo aumentar a quantidade, caso necessário. Cada mensagem terá recomendações de mudanças no comportamento (características), que comprometem o desempenho acadêmico. Essas características correspondem aos dados coletados do AVA, que podem ser desde informações sobre o uso do ambiente, como também dados demográficos. Em nosso experimento, usamos apenas dados sobre o uso do AVA.

É feita a validação de cada uma dessas características de cada aluno classificado “em risco”. Caso o valor preenchido, para cada uma das características, não corresponder ao valor que tiver a maior porcentagem de estudantes com sucesso acadêmico, o sistema adicionará a característica como parte da mensagem de recomendação. O algoritmo 1 demonstra que o sistema submete novamente cada instância com essas características alteradas, validando se o algoritmo preditivo o classifica como “sucesso” acadêmico. Se o algoritmo classificar a instância em “sucesso” acadêmico, a mensagem será gerada e sugerida ao professor, que deverá julgar sua pertinência antes de enviar para o aluno. Caso não seja, será aumentada gradativamente a quantidade de características alteradas, até que o algoritmo classifique a instância em sucesso acadêmico.

Usamos a linguagem de programação Python para o desenvolvimento e várias bibliotecas e outros frameworks. Na figura 3 é apresentado o diagrama de componentes do Framework desenvolvido em nossa pesquisa, que descreve as comunicações, componentes, interfaces e dependências. Nele contém as entidades mapeadas usando SQLAlchemy, uma biblioteca de persistência de banco de dados. Contém o módulo de software chamado ML - Predição, que utiliza as bibliotecas: Pandas para tratar dos dataframes; Numpy para manipulação de dados genéricos de forma eficiente; e o framework Sklearn para uso dos algoritmos preditivos (Decision Tree, KNN, SVM, MLP e Naive Bayes). Usamos o banco de dados SQLite para armazenar os dados. E hospedamos o sistema que usa o Framework proposto no PythonAnywhere ¹, devido a facilidade de instalação de bibliotecas Python.

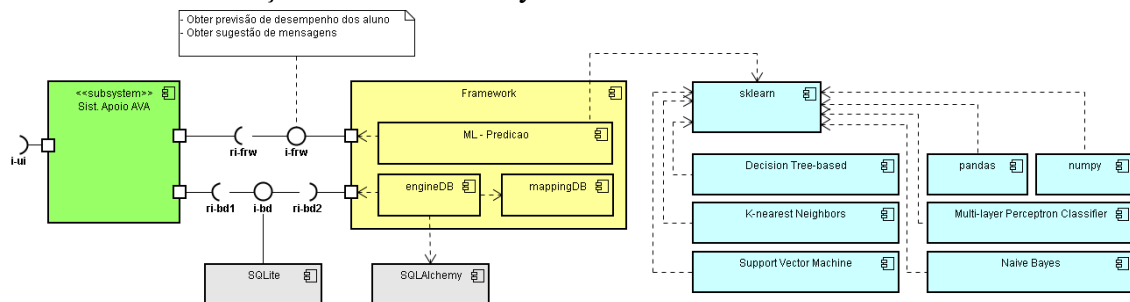


Figura 2. Diagrama de Componentes

5. Experimentos

Para validar as funcionalidades do framework criamos um sistema de apoio para apresentar os resultados dos módulos preditivos e de recomendações. Para isso, usamos um conjunto de dados público chamado OULA, *Open University Learning Analytics*, disponível no site *UCI Machine Learning Repository* ². Segundo, [Kuzilek et al. 2017], esses dados foram disponibilizados para apoiar nas pesquisas relacionadas à análise de dados educacionais, contendo informações de cursos apresentados na Universidade Aberta (*Open University - UO*). O que torna o conjunto de dados único é o fato de conter dados demográficos e dados de fluxo de cliques agregados das interações dos alunos no AVA. Isso possibilita a análise do comportamento do aluno, representado por suas ações. O conjunto de dados contém as informações sobre 22 cursos, 32.593 alunos

¹ <https://www.pythonanywhere.com/>

² <http://archive.ics.uci.edu/ml/>

com seus resultados de avaliação e registros de suas interações com o AVA representados por resumos diários de cliques de alunos (10.655.280 entradas).

Importamos os dados das tabelas *studentVle* e *studentInfo* do conjunto de dados público. Juntamos as tabelas duas tabelas com uma base de nomes fictícios. Para melhorar o cenário do experimento apagamos o resultado de cinco alunos aleatoriamente de cada um dos sete módulos, gerando um total de 35 alunos. Transformamos os dados armazenados em linhas para serem mostrados em colunas dos valores das interações com o AVA (por exemplo: *click_homepage*, *click_quiz*, *click_glossary*, *click_forumng*, etc). E por fim, categorizamos os dados, transformando valores contínuos da quantidade de cliques em valores discretos, com as descrições de “pouco uso”, “uso moderado” e “muito uso” conforme os quartis. Onde o primeiro quartil corresponde a “pouco uso”, o segundo e terceiro quartil corresponde a “uso moderado” e o quarto quartil corresponde a “muito uso”.

No módulo de Predição geramos os modelos de algoritmo de aprendizado de máquina usando os dados transformados e tratados nas etapas anteriores. Conforme apresentado na figura 3, o algoritmo selecionado foi a Árvore de Decisão (*Decision Tree*), que teve a melhor acurácia, que deverá classificar os alunos em “em risco” ou “sucesso” acadêmico.

#	Algoritmo	Acurácia
*	Decision Tree-based	70.2027
	K-nearest Neighbors	66.2912
	Support Vector Machine	69.0149
	Multi-layer Perceptron Classifier	69.1583
	Naive Bayes	62.8302

* Algoritmo selecionado para prever os registros com a classe vazia

Figura 3. Acurácia dos algoritmos preditivos

A base da mensagem de recomendação é o valor da importância de cada característica processada pelo módulo de Predição. Conforme apresentado na figura 4, o algoritmo preditivo apresentou a característica “Acessos ao portal AVA”, com uma importância de 0,51518 (valor de referência do algoritmo). A maior porcentagem de alunos com sucesso acadêmico teve “muito acesso” nessa característica. Com um total de 93,53% dos alunos com sucesso acadêmico e consecutivamente 6,47%, com insucesso acadêmico. Demonstrando a relevância na mudança de comportamento dos estudantes. Todas as características são apresentadas com diferentes importâncias e porcentagens.

Na figura 5, bloco A, é apresentada a mensagem candidata sugerida pelo framework ao professor, que deverá julgar sua pertinência antes do envio ao aluno usando o sistema de apoio. A mensagem é composta por itens de recomendações de mudança no comportamento do aluno, que comprometem o sucesso acadêmico. O professor poderá editar cada item da mensagem ou até mesmo ignorar qualquer um que não considere apropriado. Na figura 5, bloco B, no qual a tela foi capturada com o auxílio do software de conexão remota chamado TeamViewer, é apresentada a mensagem recebida pelo aluno com as recomendações do professor.

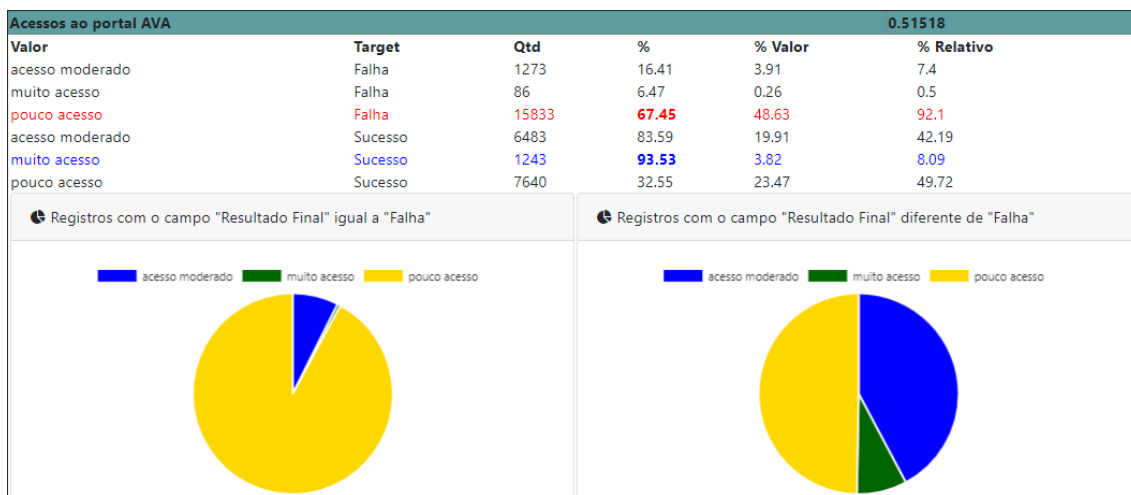


Figura 4. Pesos das características

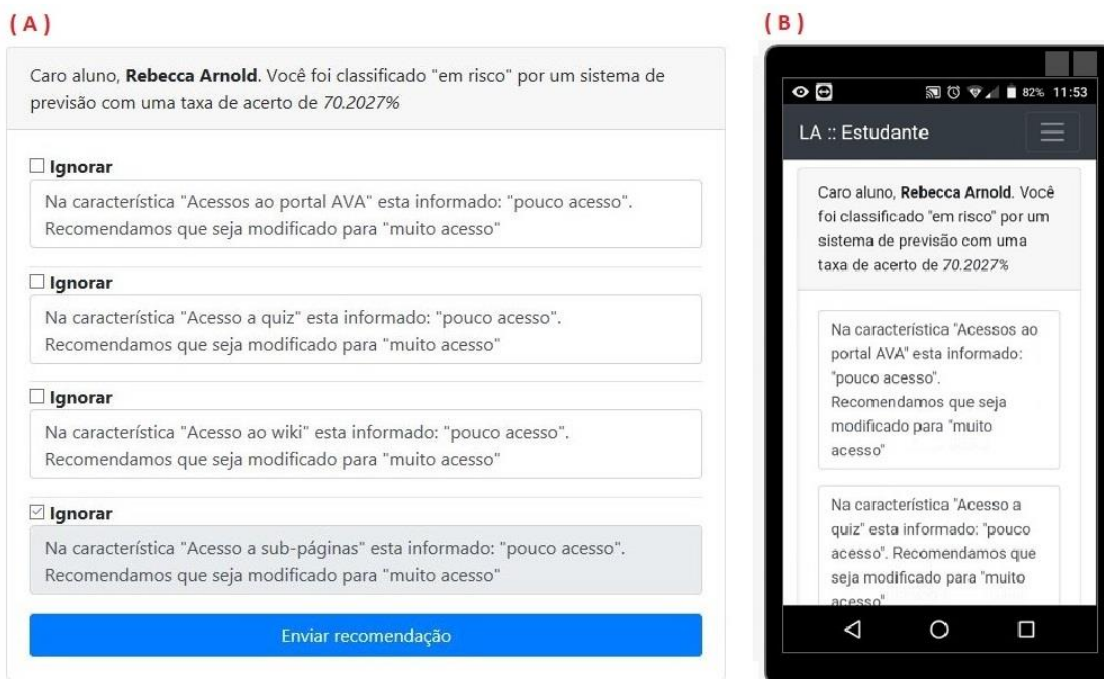


Figura 5. Mensagem: (A) sugerida ao professor; (B) recebida pelo aluno

6. Conclusão

Durante os últimos anos houve um aumento nas pesquisas voltadas à análise de dados acadêmicos. Principalmente devido ao aumento na procura de aprendizado online, sendo que ambientes virtuais são apropriados para essas pesquisas, devido a seus dados serem coletados automaticamente. Este trabalho busca auxiliar na evolução das pesquisas de intervenção no processo de aprendizado de alunos, através de sistemas de predição de desempenho e de recomendações. Os primeiros resultados do framework usando um conjunto de dados públicos indicam que as mensagens geradas são apropriadas aos alunos para evitar insucesso acadêmico. Portanto, almejamos que os estudos realizados nesse trabalho se consolidem em uma ferramenta aplicável em

qualquer ambiente de aprendizado online. Oferecendo aos professores a possibilidade real de compreenderem como os alunos aprendem e intervir em seu aprendizado de forma individualizada.

Como trabalhos futuros, pretendemos submeter a ferramenta para avaliação de outros pesquisadores e professores para uma avaliação quantitativa sobre a aderência dela em um contexto acadêmico real. Pretendemos melhorar a capacidade preditiva do framework, usando algoritmos genéticos para que os ajustes dos parâmetros sejam dinâmicos. Por fim, tornar as mensagens mais amigáveis e menos técnicas.

Referências

- Arnold, K. E. and Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success. *In Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 267–270. ACM.
- Brown, M. (2012). Learning analytics: moving from concept to practice. *EDUCAUSE Learning Initiative*, pages 1–5.
- Carvalho, A., FACELI, K., LORENA, A., & GAMA, J. (2011). Inteligência Artificial—uma abordagem de aprendizado de máquina. *Rio de Janeiro: LTC*.
- Clow, D. (2012). The learning analytics cycle: closing the loop effectively. *In Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 134–138. ACM.
- de Brito, D. M., de Almeida Junior, I. A., Queiroga, E. V., and do R ego, T. G. (2014). Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. *In Brazilian Symposium on Computers in Education (Simposio Brasileiro de Informática na Educação-SBIE)*, volume 25, page 882.
- Dyckhoff, A. L., Zielke, D., Bultmann, M., Chatti, M. A., and Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15(3):58.
- Elias, T. (2011). Learning analytics. Learning.
- Gotardo, R., Cereda, P. R. M., and Junior, E. R. H. (2013). Predição do desempenho do aluno usando sistemas de recomendação e acoplamento de classificadores. *In Brazilian Symposium on Computers in Education (Simposio Brasileiro de Informática na Educação-SBIE)*, volume 24, page 657.
- Gottardo, E., Kaestner, C., and Noronha, R. V. (2012). Previsão de desempenho de estudantes em cursos EaD utilizando Mineração de dados: uma estratégia baseada em séries temporais. *In Brazilian Symposium on Computers in Education (Simposio Brasileiro de Informatica na Educação-SBIE)*, volume 23.
- Govindarajan, K., Kumar, V. S., Boulanger, D., et al. (2015). Learning analytics solution for reducing learners' course failure rate. *In Technology for Education (T4E), 2015 IEEE Seventh International Conference on*, pages 83–90. IEEE.
- Greller, W. and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3):42.

- Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. *Scientific data*, 4, 170171.
- Long, P. (2011). *LAK'11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, February 27-March 1, 2011, Banff, Alberta, Canada*. ACM.
- Martin, F. and Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, 13(3):7.
- Medeiros, F., Gomes, A. S., Amorim, R., & Medeiros, G. (2013, July). Architecture for Social Interactions Monitoring in Collaborative Learning Environments as a Support for the Teacher's Awareness. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on* (pp. 123-127). IEEE.
- Moissa, B., Gasparini, I., and Kemczinski, A. (2014). Learning Analytics: um mapeamento sistemático. *Nuevas Ideas en Informática Educativa TISE*, 2014:283–290.
- Oliveira Júnior, José Gonçalves de. *Identificação de padrões para a análise da evasão em cursos de graduação usando mineração de dados educacionais*. MS thesis. Universidade Tecnológica Federal do Paraná, 2015.
- Pakalra, and Olprod (2018) “Página de dicas úteis do algoritmo para Machine Learning para o Microsoft Azure Machine Learning Studio”, <https://docs.microsoft.com/pt-br/azure/machine-learning/studio/algorithm-cheat-sheet>, agosto.
- Padilha, T. and Souza, I. (2017). Uma experiência do uso de ferramentas de learning analytics para análise de interações. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 644.
- Pascoal, T., de Brito, D. M., Andrade, L., and do Rego, T. G. (2016). Evasão de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos. In *Brazilian Symposium on Computers in Education (Simposio Brasileiro de Informática na Educação-SBIE)*, volume 27, page 926.
- Richert, W. (2013). *Building machine learning systems with Python*. P. Publishing Ltd.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., and Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10):1500–1509.
- Vieira, M. A. and Veiga (2014), E. F. Árvores de decisão aplicadas na previsão de desempenho de alunos: Estado da arte.
- Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.