

Dropoutless: plataforma colaborativa de predição de evasão

Laís Pisetta Van Vossen¹, Maria Teresa Silva Santos¹,
Luciana Bolan Frigo², Isabela Gasparini¹

¹Universidade do Estado de Santa Catarina (UDESC)
Joinville, Santa Catarina, Brasil

²Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Santa Catarina, Brasil

{lais.vossen,maria.santos2805}@edu.udesc.br

luciana.frigo@ufsc.br, isabela.gasparini@udesc.br

Abstract. *Student dropout is a recurring problem that affects various spheres of society, representing a waste of resources for the state and a decrease in qualified labor for companies. This article proposes a collaborative platform, named Dropoutless, whose aim is to reduce dropout rates in university disciplines by creating co-participative predictive models using AutoML techniques. As a result, the proposed tool allows for greater flexibility in the data used for dropout prediction and promotes the leadership of educators in changing this scenario.*

Resumo. *A evasão estudantil é um problema recorrente que afeta diversas esferas da sociedade, representando um desperdício de recursos para o Estado e uma queda da mão de obra qualificada para as empresas. Neste artigo é proposta uma plataforma colaborativa, nomeada Dropoutless, cujo objetivo é reduzir a evasão em disciplinas universitárias através da criação de modelos de predição de forma coparticipativa utilizando técnicas de AutoML. Como resultado, a ferramenta proposta permite uma flexibilização nos dados utilizados na predição de evasão e promove o protagonismo dos educadores na mudança desse cenário.*

1. Introdução

A redução das taxas de evasão em cursos do ensino superior é uma das metas mais desafiadoras para as universidades. Iniciativas como o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni) tem como um de seus objetivos o combate a evasão como forma de diminuir as desigualdades sociais no país¹.

Para o Estado, a evasão representa um desperdício de recursos, com a estrutura acadêmica sendo subutilizada com a saída de estudantes que ela poderia comportar. Além disso, o abandono dos estudos pode ser visto como um indicativo de falhas ou ineficiência no processo de ensino [Gilioli 2016]. Este fenômeno também afeta outras esferas da sociedade, podendo ser a causa de uma carência de mão de obra especializada, que leva a menor competitividade e eficiência das empresas nacionais [Mussliner et al. 2021].

A importância dos educadores na retenção de estudantes é notável, sendo um dos fatores mantenedores dos estudantes no curso visto que 80% dos estudantes diplomados

¹<http://reuni.mec.gov.br/o-que-e-o-reuni>

e apenas 46% dos estudantes evadidos afirmaram ter boas relações com os professores [Silva et al. 2012]. Isso evidencia a necessidade de investir em ações diretas do professor para com seus estudantes.

O presente artigo tem como objetivo propor uma ferramenta, nomeada Dropoutless, que busca prever a evasão de forma colaborativa, através da criação coparticipativa de modelos de *Machine Learning* (ML) para ajudar professores a prever a evasão de estudantes em suas disciplinas e possibilitar a tomada de ação para a mudança desse cenário. Através da criação de um sistema colaborativo, é possível contornar um dos maiores problemas na utilização de técnicas de ML que é a escassez de dados [Bansal et al. 2022].

A plataforma proposta também conta com um ranking de modelos de predição, que podem ser ordenados pelas métricas de avaliação desejadas pelo usuário, além de filtrados por *tags*, para apresentar apenas os modelos que levam em consideração determinadas características na predição. Desta forma, a ferramenta promove a colaboração através do uso e melhoria contínua dos modelos desenvolvidos, que se adaptam a inúmeros cenários acadêmicos e de classificação de evasão.

Para o desenvolvimento de Dropoutless, utilizou-se de ferramentas de *Machine Learning* automatizadas (AutoML), que são capazes de produzir modelos de predição de forma personalizada, além de tratar das colunas não completamente preenchidas com a entrada automática de números neutros ou zero. Seu uso é possibilitado até mesmo por pessoas que desconhecem a área de computação e inteligência artificial.

Por fim, este trabalho se divide em sete seções, a Seção 2 apresenta conceitos de evasão e a definição adotada para este artigo, a Seção 3 descreve a ferramenta em mais detalhes e suas funcionalidades. Na Seção 4 é apresentado um caso de uso do Dropoutless e na Seção 5 é discutida a arquitetura adotada para a plataforma. Na Seção 6 é feita uma discussão das implicações éticas envolvidas com o desenvolvimento do Dropoutless. Por fim, a Seção 7 discute as conclusões do artigo e apresenta sugestões de trabalhos futuros.

2. Evasão

Determinar a causa da evasão é uma tarefa complexa pois envolve muitas variáveis. Alguns trabalhos dedicaram-se a prever a evasão estudantil em universidades, analisando fatores de contextos diversos tanto de cursos online, como os Massive Open Online Course (MOOC) [Liang et al. 2016], [Periwal and Rana 2017], quanto de estudantes da Universidade [Costa et al. 2020]. Isso demonstra a amplitude do tópico, atingindo numerosos cenários acadêmicos de formas diferentes.

Para a construção da ferramenta proposta, é necessário conceituar o que é considerado como evasão, visto que existem várias classificações para esse termo na literatura. Silva Santos et al. (2022) discorre sobre como a evasão é vista sob diferentes óticas, elencando três principais tipos de evasão, que são elas:

- *Dropout*: abandono definitivo do curso;
- *Optout*: mudança de curso ou instituição feita de forma intencional;
- *Stopout*: pausa por um período indeterminado de tempo.

Dadas as diferentes classificações de evasão apresentadas, este artigo trata sobre a evasão definitiva, ou *Dropout*, do estudante em relação a disciplina. Nota-se que o conceito adotado neste artigo difere-se do conceito tradicional de evasão pois trata do aban-

dono nas disciplinas apenas e não do curso ou instituição como um todo. A escolha desse conceito se deu por conta do público alvo da ferramenta, os professores responsáveis pelas disciplinas. Estes possuem os meios de determinar se um estudante evadiu ou não de suas disciplinas, porém estender isso ao curso ou instituição muitas vezes foge do escopo do educador e torna difícil determinar se ocorreu a evasão.

Este trabalho pode considerar evasão estudantes que não tiveram a frequência necessária, não entregaram atividades, não obtiveram a nota mínima para passar na disciplina, entre outras opções a depender das variáveis de entrada. Com isso, a definição será dada pelo usuário no momento em que fornecer os dados necessários para a ferramenta, visto que é ele quem vai classificar os estudantes como evadidos ou não, e graças ao uso do AutoML essa flexibilização de conceito e a adaptação mais ampla aos variados cenários é possível.

3. Dropoutless – Plataforma Colaborativa para predição de evasão de disciplinas

A ferramenta aqui apresentada, irá possibilitar a previsão antecipada da evasão, permitindo assim ações para evitar a evasão do estudante. Seu objetivo é utilizar as vantagens que um sistema colaborativo proporciona para criar modelos de predição de evasão eficazes e fáceis de usar.

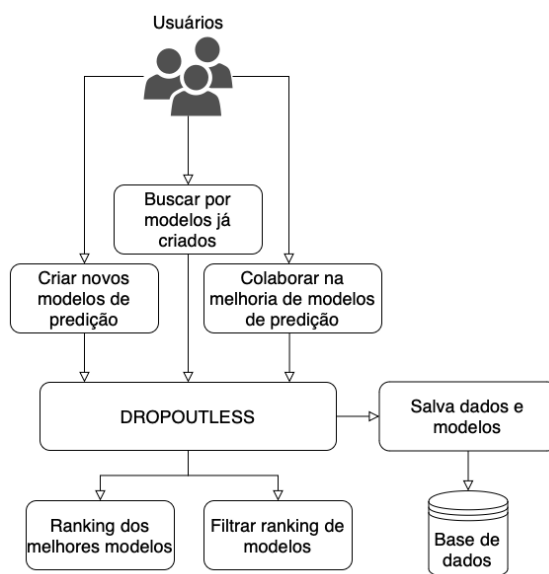


Figura 1. Diagrama de contexto do sistema Dropoutless.

O diagrama da Figura 1 é uma representação do contexto geral da ferramenta, fornecendo uma visão das responsabilidades do sistema e como ele interage com os usuários e a base de dados. É possível perceber que o sistema permite ao usuário interagir com a criação de modelos de predição, desde a criação do seu próprio modelo até a busca por modelos desenvolvidos por outros usuários e aperfeiçoados pela comunidade de forma colaborativa. Então, o sistema Dropoutless conta com sua principal funcionalidade:

- **Geração de modelos de predição com o uso da tecnologia AutoML:** descrito na Seção 5.1, essa funcionalidade possibilita ao usuário criar e treinar novos modelos

de predição, se adaptando ao conjunto de dados que o mesmo possui, ou ainda, utilizar um modelo pronto, previamente treinado em conjuntos de dados de outros usuários.

A forma como a colaboração ocorre no sistema Dropoutless se baseia no Modelo 3C de Colaboração, que analisa a colaboração em três dimensões: a comunicação, a coordenação e a cooperação [Fuks et al. 2011]. A dimensão da comunicação é a ação de troca de mensagens para que haja o entendimento comum entre os envolvidos, podendo ser feita de forma síncrona ou assíncrona. A coordenação é o elemento que organiza e coordena as ações e demandas geradas pela comunicação, é a dimensão que mantém a ordem no espaço compartilhado. Por fim, a dimensão da cooperação é onde todas as partes envolvidas trabalham juntas em uma atividade no espaço compartilhado.

Sendo assim, essas dimensões podem ser vistas no Dropoutless, no qual a comunicação entre usuários (prevista como trabalho futuro) ocorrerá através do compartilhamento de materiais dos comentários existentes nas publicações, em que os professores poderiam escrever suas opiniões sobre o conteúdo postado. A coordenação, que é definida como o gerenciamento de pessoas e recursos, pode ser vista na plataforma através dos usuários moderadores, que possuem os privilégios de agir de forma a moderar o conteúdo, tendo o poder de apagar modelos e excluir contas de usuários por mau uso da ferramenta. Por fim, a cooperação ocorre na criação de modelos de predição através do compartilhamento de dados com a plataforma, o que permite que o modelo treine com cada vez mais dados de diferentes contextos, podendo gerar uma melhoria nas predições geradas.

4. Cenário de uso

Esta seção tem como objetivo apresentar um cenário de uso da plataforma Dropoutless, com objetivo de analisar a viabilidade da solução e demonstrar como ela pode ser usada para prever a evasão de forma colaborativa.

A Figuras 2 e 3 apresentam as principais telas do sistema Dropoutless. Além delas, há ainda a tela de Login, a página inicial onde é apresentada uma descrição da ferramenta e como o usuário pode utilizá-la, a tela de ranking de modelos e a tela de gerenciamento dos modelos criados e salvos.

Para apresentar um caso de uso do Dropoutless, o usuário “João” é utilizado como exemplo. “João” é um professor do curso de Ciência da Computação, que leciona a disciplina de Algoritmos. Todos os semestre, “João” percebe estudantes que abandonaram a disciplina por diferentes motivos, desde faltas excessivas até notas baixas em provas. Visando reduzir a evasão das turmas de sua disciplina, “João” decide utilizar a ferramenta Dropoutless para identificar estudantes que possam evadir com antecedência e poder tomar ações para evitar isso.

Então, “João” realiza o cadastro na plataforma, clicando no link para criar uma conta na tela de login. De posse de uma conta, “João” seleciona a opção de criar um modelo, onde determina um nome para o seu modelo, faz o carregamento dos dados que coletou sobre sua turma e preenche o nome e descrição de todas as colunas presentes no arquivo do tipo *comma separated values* (CSV) selecionado. “João” optou por usar como dados as notas na primeira prova, a frequência dos estudantes e a entrega do primeiro trabalho, embora a ferramenta aceite quaisquer outros dados sobre provas, trabalhos,

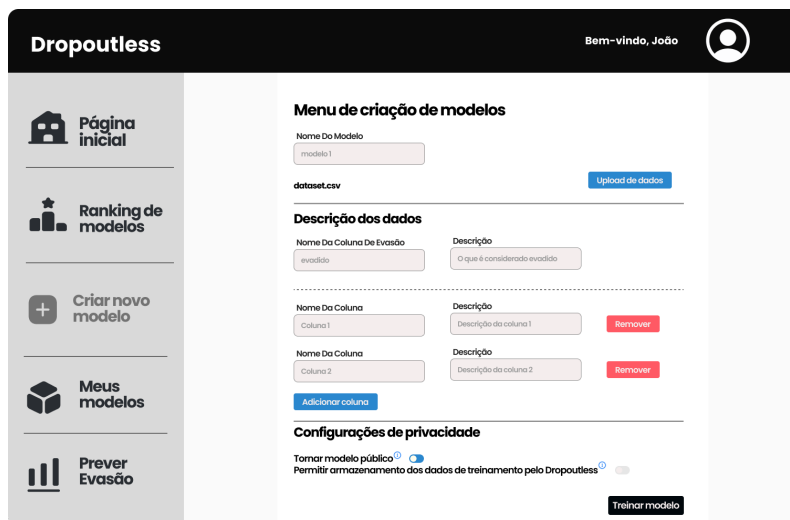


Figura 2. Tela de criação de modelos do sistema Dropoutless.

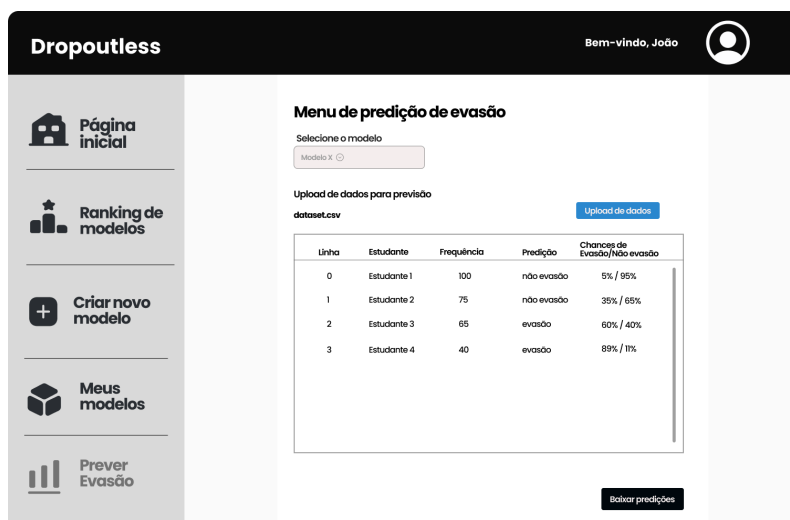


Figura 3. Tela de predição de evasão do sistema Dropoutless.

frequência, participação, entre outros. Nesta tela, “João” pode escolher se deseja compartilhar seu modelo para que outros usuários possam utilizá-lo para realizar predições sobre os dados que possuem. Por fim, clicando no botão “Treinar modelo” as configurações e arquivos são enviados para que o AutoML realize a criação e treinamento do modelo.

Quando o modelo estiver pronto, “João” poderá visualizá-lo na tela da Figura 3. Nesta tela, “João” deve fornecer os dados atuais dos estudantes dos quais deseja prever as chances de evasão, podendo observá-las na própria plataforma ou ainda baixar um arquivo CSV com os resultados.

Porém, devido a baixa quantidade de dados disponibilizados para o treinamento, visto que “João” apenas apresentou os dados da turma anterior, o modelo desenvolvido obteve um baixo desempenho. Então, para mitigar esse problema, “João” buscou no ranking de modelos por um que possuísse métricas de desempenho melhores e que utilizasse na predição a “Nota de Prova” e a “Frequência”, conforme evidenciado nas *tags*. Para

identificar quais métricas ele gostaria de maximizar na busca por um modelo, “João”pôde ler a explicação sobre o que cada uma delas representa, já que a explicação aparece ao passar o cursor em cima do ícone de informações. Com isso, “João”selecionou o melhor modelo e salvou ele clicando no coração para curtir.

Por fim, “João”realiza novas predições com o modelo salvo e verifica um estudante com altas chances de evasão. Determinado a ajudar esse estudante a concluir a disciplina de algoritmos, “João”desenvolve atividades extras para engajar o estudante e consegue impedir uma possível evasão.

O caso de uso de “João”apresentou como o usuário pode interagir com a plataforma e participar ativamente dela criando novos modelos de predição. Com isso, é possível perceber os benefícios que a criação de um sistema colaborativo pode trazer para o problema da evasão.

5. Arquitetura da plataforma

A Figura 4 apresenta a arquitetura do sistema Dropoutless, cuja base é de uma arquitetura de microsserviços. Nela é possível ver a divisão de cada parte da estrutura, onde o usuário interage com o *Frontend*, a parte gráfica da aplicação que funcionará na *Web*, e esta se comunica com as demais interfaces de programação de aplicações (APIs) para realizar as funções necessárias de forma desacoplada. As APIs se comunicam com um banco de dados único, responsável por armazenar as informações de postagem dos usuários, modelos criados e dados de criação de novos usuários.

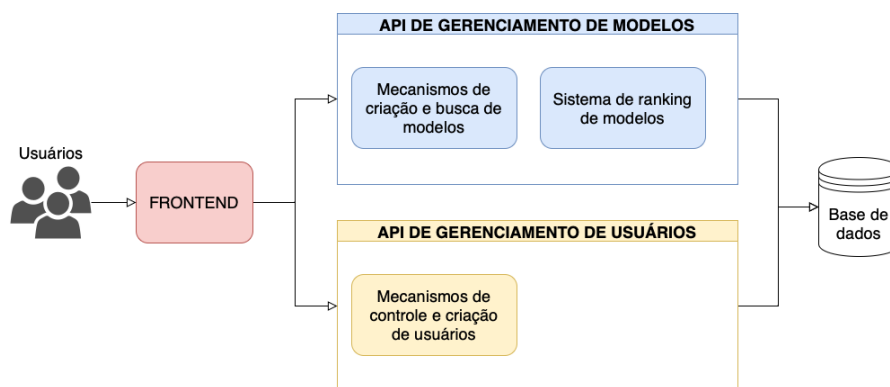


Figura 4. Representação gráfica da arquitetura do Dropoutless.

A API de gerenciamento de modelos é responsável por duas funções principais, a primeira é a de fornecer os meios de criação, busca e gerenciamento de modelos criados utilizando a biblioteca de AutoML, o Autogluon. Com isso, esse módulo fornecerá um canal de comunicação que permite criar novos modelos utilizando arquivos no formato CSV. Já o módulo de *ranking* de modelos será responsável pela organização dos mesmos de acordo com os filtros selecionados pelo usuário. O *ranking* determina quais são os melhores preditores baseando-se em métricas fornecidas pela biblioteca, como por exemplo a acurácia do modelo. Esta API também realiza o fluxo presente no diagrama da Figura 5, onde ocorre a melhoria contínua dos modelos de predição através de seu uso por novos usuários. Além do armazenamento do modelo treinado com os novos dados, no caso de melhora de desempenho em relação ao anterior, também é de responsabilidade desta API

realizar os procedimentos para garantir a anonimização dos dados para o armazenamento na base de dados.

O modelo mostra no ranking apenas os modelos que satisfazem os filtros fornecidos pelo usuário, como por exemplo, levar em consideração a quantidade de faltas do estudante para a predição, então nesse caso o ranking contaria apenas com modelos que usam dessa informação para a tomada de decisão.

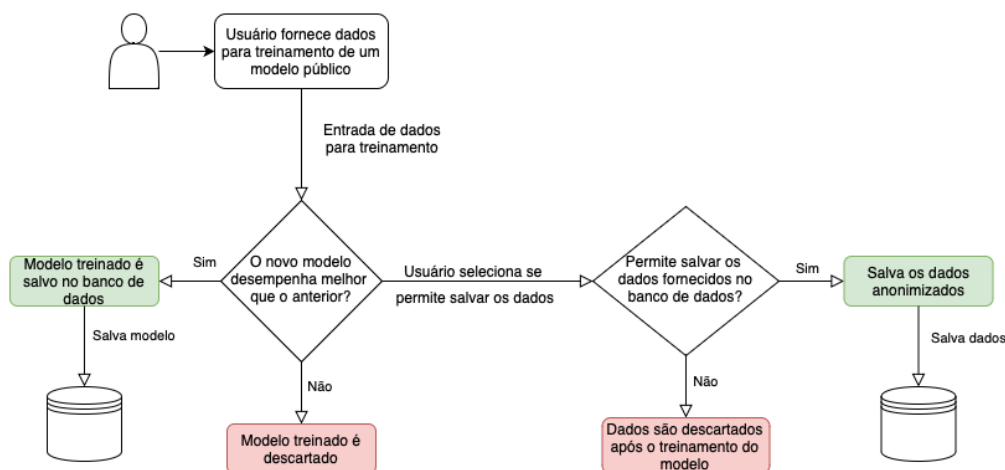


Figura 5. Fluxo da melhoria contínua de modelos.

Então, para o funcionamento de todas as partes, a API de gerenciamento de usuários é necessária, visto que ela é a responsável pelo cadastro e login daqueles que utilizam a ferramenta. Além disso, essa API gerencia os dados como nome, email, foto de perfil e senha que o usuário preenche na plataforma.

5.1. Modelo de predição de evasão

Diante da natureza do problema, os usuários que não são familiarizados com aprendizado de máquina utilizariam o sistema, portanto, é fundamental que a ferramenta possa realizar o máximo possível de processos de forma automatizada, de modo que o usuário leigo não precise entender o funcionamento do mesmo para gerar seu modelo. Por conta disso, a ferramenta Dropoutless emprega o uso de técnicas de AutoML para tal.

Segundo a AutoML², o *Machine Learning* automatizado pode ser pensado como um algoritmo de busca especializado em achar soluções ótimas para cada componente do pipeline de ML. Com isso, ele permite a automatização de processos chaves como a engenharia de *features*, o pré-processamento de dados e a otimização de hiperparâmetros na busca pelo melhor modelo.

Como a predição de evasão dinâmica e simples é o foco principal da ferramenta Dropoutless, então, a biblioteca de AutoML escolhida para o sistema é a Auto-gluon³, por se tratar de uma ferramenta *open-source* e gratuita, além de declarar, em sua documentação, que realiza os processo de limpeza e preparação dos dados, o que diminui a necessidade do usuário de entender certos conceitos para usar a ferramenta. No mais,

²<https://www.automl.org/automl/>

³<https://auto.gluon.ai/stable/index.html>

a escolha de uso da tecnologia AutoML agrega na forma de trabalho de um sistema colaborativo, visto que os usuários poderão colaborar dentro da plataforma na criação de modelos de ML cada vez melhores através do compartilhamento dos dados, contribuindo também para a redução de escassez dos mesmos, já que um único modelo poderia contar com informações de estudantes de disciplinas diversas de diferentes professores.

6. Implicações Éticas

Quando se armazena e manipula dados dos usuários, como o Dropoutless faz, deve-se ter precauções para garantir a privacidade de todas as partes. Sendo assim, é imprescindível à ferramenta a determinação de processos para lidar com a anonimização dos dados dos estudantes, visto que o objetivo dela é que o usuário possa usar os dados reais de suas turmas para gerar as previsões. Então, para lidar com essa situação, a ferramenta apresenta indicações de como tornar os dados anônimos antes de realizar o carregamento dos mesmos, através de ícone explicativos na tela de criação de modelo. Além disso, quando o usuário dá a permissão de armazenamento desses dados para treinamentos futuros de modelos, a API de gerenciamento dos modelos realiza a limpeza desses dados, transformando variáveis de texto em categorias numéricas através de técnicas como *One Hot Encoder* e *Label Encoder*, para impedir o armazenamento do nome dos estudantes.

Outra situação que se deve observar ao lidar com ML, em especial aquelas que fazem classificações de pessoas, nesse caso como evasão ou não evasão, é que as classificações podem muitas vezes estar erradas. Então, o usuário que agir com base na classificação dada pelo modelo, deve ter em mente que o mesmo pode estar incorreto. Para amenizar possíveis tomadas de decisões mais drásticas influenciadas pelos resultados do modelo, o mesmo apresenta as porcentagens de chance do estudante evadir ou não evadir, demonstrando que não há certeza nas respostas, apenas há uma chance maior de um fenômeno acontecer em detrimento do outro.

7. Conclusões e Trabalhos Futuros

Este artigo teve como principal objetivo o desenvolvimento de uma ferramenta colaborativa de predição de evasão em disciplinas. Observou-se que os trabalhos envolvendo predição de evasão tendem a focar na evasão em cursos e com modelos de ML pré-desenvolvidos, o que torna a ferramenta Dropoutless um diferencial por seu foco na evasão em disciplinas e uso de AutoML em um sistema colaborativo para a melhora contínua das predições.

Como trabalhos futuros, podem ser citadas algumas melhorias na plataforma, como o estudo da otimização dos hiperparâmetros fornecidos para os algoritmos do AutoML, com o objetivo de obter desempenhos melhores nos modelos. Além disso, pode ser acrescentado um módulo de compartilhamento de materiais com o objetivo de permitir ao usuário publicar textos e arquivos que descrevem métodos de redução da evasão testados por ele. Além das melhorias na plataforma, ressalta-se a necessidade de testar a plataforma com usuários e dados reais, para verificar a eficiência da solução em um ambiente colaborativo.

Referências

- Bansal, M. A., Sharma, D. R., and Kathuria, D. M. (2022). A systematic review on data scarcity problem in deep learning: solution and applications. *ACM Computing Surveys (CSUR)*, 54(10s):1–29.
- Costa, A. G., Queiroga, E., Primo, T. T., Mattos, J. C. B., and Cechinel, C. (2020). Prediction analysis of student dropout in a computer science course using educational data mining. In *2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO)*, pages 1–6.
- Fuks, H., Raposo, A. B., Gerosa, M. A., Pimentel, M., Filippo, D., and Lucena, C. d. (2011). Teorias e modelos de colaboração. *Sistemas colaborativos*, pages 16–33.
- Gilioli, R. d. S. P. (2016). Evasão em instituições federais de ensino superior no brasil: expansão da rede, sisu e desafios. *Brasília: Câmara dos Deputados*, 49:1–55.
- Liang, J., Yang, J., Wu, Y., Li, C., and Zheng, L. (2016). Big data application in education: Dropout prediction in edx moocs. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 440–443.
- Mussliner, B. O., de Sousa, M., Mussliner, S., Meza, E. B. M., and Rodríguez, G. L. (2021). O problema da evasão universitária: um desafio à democratização do ensino superior público. *Brazilian Journal of Development*, 7(4):42738–42758.
- Periwal, N. and Rana, K. (2017). An empirical comparison of models for dropout prophecy in moocs. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pages 906–911.
- Silva, F. I. C. d., Rodrigues, J. D. P., Brito, A. K. A., and França, N. M. d. (2012). Evasão escolar no curso de educação física da universidade federal do piauí. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*.
- Silva Santos, M. T., Pisetta Van Vossen, L., Vasconcellos, D., Tomaselli Borchardt, G., Vaichulonis, G., Bolan Frigo, L., and Gasparini, I. (2022). Análise da evasão feminina nos cursos de ciência da computação das universidades públicas e presenciais de santa catarina. *RENOTE*, 20(1):233–242.