

## GraduAI – Sistema com Aprendizagem de Máquina para Avaliação de Risco de Evasão

Leonardo de Almeida Teodoro<sup>1</sup>, Arthur Pedrosa Ferreira<sup>1</sup>, Marco André Abud Kappel (orientador)<sup>1</sup>

<sup>1</sup> Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET/RJ – Campus Nova Friburgo, Nova Friburgo, RJ, Brasil

leonardo.teodoro@aluno.cefet-rj.br, arthur.ferreira@aluno.cefet-rj.br,  
marco.kappel@cefet-rj.br

**Abstract.** *The dropout rate in Brazilian higher education institutions has become a growing concern. This study aims to identify dropout and academic success patterns using machine learning techniques on public education data. The results achieved an accuracy rate of approximately 80% in predictions. With this predictive model, an API was developed to provide information about students' chances of success and dropout, considering relevant characteristics such as age, extracurricular activities, and social support. This approach aims to assist students in planning their academic careers more efficiently, reducing the likelihood of dropout.*

**Resumo.** *A evasão nas instituições de ensino superior brasileiras tem se tornado uma preocupação crescente. Este trabalho possui como objetivo identificar padrões de evasão e sucesso acadêmico utilizando técnicas de aprendizado de máquina em dados públicos de educação. Os resultados alcançaram uma taxa de acerto nas previsões de aproximadamente 80%. Com posse desse modelo preditivo, foi desenvolvida uma API que fornece informações sobre as chances de sucesso e evasão dos estudantes, considerando características relevantes como idade, atividades extracurriculares e apoio social. Essa abordagem visa auxiliar os estudantes no planejamento de suas carreiras acadêmicas de forma mais eficiente, reduzindo a probabilidade de evasão.*

### 1. Introdução

A evasão é um problema que ocorre em vários níveis da educação e em diversos lugares do mundo. Atualmente, é um problema que atrai a atenção de pesquisadores devido às suas consequências para a sociedade (Freitas, et al., 2020). Além disso, a redução da taxa de evasão é um grande desafio para as instituições públicas de ensino superior brasileiras (PRESTES e FIALHO, 2018). Com o número de alunos matriculados aumentando a cada ano (COSTA e DIAS, 2016), a desproporção na razão entrada/saída nas universidades brasileiras pode configurar o desperdício de recursos públicos, gerando problemas econômicos, acadêmicos e sociais (FILHO *et al.*, 2007), (SALES, BALBY e CAJUEIRO, 2016).

Hoje, as técnicas de Aprendizado de Máquina são amplamente utilizadas para a determinação e classificação de padrões em massas de dados. Estas técnicas são caracterizadas como um tipo de inteligência artificial que é capaz de obter conhecimento sem programação explícita, através do processamento de um conjunto de dados (SARAVANAN e SUJATHA, 2018). Ademais, a emergente política de Dados Abertos ganhou força com a ampla utilização de algoritmos que dependem de dados. Esta política estabelece a publicação de dados com livre acesso à população, disponibilizando-os online para utilização em diferentes finalidades (SILVA *et al.*, 2020).

Diante disso, e ao considerar os dados de educação brasileira disponibilizados anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que contém não só informações acadêmicas dos alunos e dos cursos, como também informações estruturais sobre a instituição de ensino, foi possível idealizar um modelo classificatório que revelasse as probabilidades de sucesso e de evasão dos estudantes. A aplicação de aprendizado de máquina no contexto educacional é conhecida como Mineração de Dados Educacionais (MDE).

O presente trabalho tem como objetivo primário construir um sistema web que utiliza um modelo de aprendizado de máquina confiável para apresentar ao estudante a probabilidade de aprovação ou de evasão no curso pretendido. Para isso, o sistema coleta dados do estudante, do curso, da instituição de ensino e do campus no qual ele deseja se matricular. A partir destas informações, o resultado da análise é demonstrado pelo sistema por meio de gráficos, tabelas e estatísticas que informam quais foram os fatores que mais influenciaram no resultado obtido.

Como objetivo secundário, este trabalho visa apresentar gráficos, tabelas e dados estatísticos sobre a evasão, além de fornecer maneiras para os usuários identificarem os cursos nos quais terão mais chances de se formar e os principais atributos que podem ser trabalhados para aumentar essas chances. Por fim, este trabalho propõe uma *Application Programming Interface* (API) para a consulta da probabilidade de evasão dos alunos. Dessa forma, o estudo almeja auxiliar a preencher a lacuna de conhecimento sobre as probabilidades de conclusão dos cursos e os fatores que possuem maior influência neste cenário. Com isso, espera-se que não só a entrada, como toda a trajetória do estudante em uma instituição de ensino superior pública brasileira seja facilitada e bem-sucedida, o que irá contribuir para a redução da evasão e os problemas socioeconômicos causados por ela.

## 2. Trabalhos Relacionados

A aplicação de Aprendizado de Máquina no contexto educacional é conhecida como Mineração de Dados Educacionais. Alguns trabalhos elaborados nos últimos anos indicam ampla utilização de algoritmos como *Naive Bayes*, *K-Nearest Neighbors*, *Neural Networks*, *Random Forest* e *Decision Tree*, além de modelos de regressão, com o objetivo de extrair informações e detectar padrões em dados educacionais.

Diversos trabalhos visam entender e combater o problema da evasão através de algoritmos de Aprendizado de Máquina: (DA SILVA *et al.*, 2019), (DEL BONIFRO *et al.*, 2020) (NIETO *et al.*, 2019), (Santos, Menezes, de Carvalho, & Montesco, 2019). Entre os trabalhos mais recentes, no contexto do Brasil, (DA SILVA *et al.*, 2019) e (TEODORO e KAPPEL, 2020) buscam compreender a evasão considerando todos os cursos disponíveis na base do INEP. Já outros estudos preferem utilizar a amostra de apenas alguns cursos, como é o caso de (Santos, Menezes, de Carvalho, & Montesco,

2019). Porém, o que é comum nesses trabalhos é que não utilizam os modelos preditivos criados junto com um sistema web para revelar aos estudantes de universidades públicas brasileiras as informações relevantes para o sucesso de sua formação.

O artigo de da Silva *et al.* (2019) utiliza MDE nos dados do INEP, sobre o censo e o indicador de fluxo do ensino superior, disponibilizados em 2013. São registrados 22.972 estudantes, que abrangem tanto a rede pública quanto a rede privada de ensino superior brasileiro. Os alunos possuem características demográficas, acadêmicas e socioeconômicas. Por fim, os autores elaboram três modelos que, na verdade, são combinações de técnicas de aprendizado de máquina. Através da avaliação pelo erro absoluto médio e o erro quadrático médio, a regressão linear combinada com K-Nearest-Neighbors foi o melhor modelo elaborado. Também são reveladas as principais características para elaborar o modelo, dentre elas estão a quantidade de alunos remanescentes no curso e se o curso é lecionado no turno da noite.

Em Teodoro e Kappel (2020) a base de dados estudantis disponibilizada pelo INEP sobre o ano de 2017 foi analisada com o intuito de extrair o padrão dos estudantes mais propensos a evadir na rede pública de ensino superior. No total, há 376.746 estudantes entre os dados, e são elaborados cinco modelos preditivos: *K-Nearest Neighbors*, *Decision Tree*, *Random Forest*, Redes Neurais e *Naive Bayes*. No treinamento dos modelos, os dados apresentaram características socioeconômicas e demográficas sobre os alunos, os cursos e as instituições em que estão inscritos. Ao compará-los, foi evidente que a técnica de *Random Forest* obteve os melhores resultados, com acurácia média de aproximadamente 80%. O estudo termina revelando as 30 características consideradas mais importantes na previsão da evasão por esse modelo.

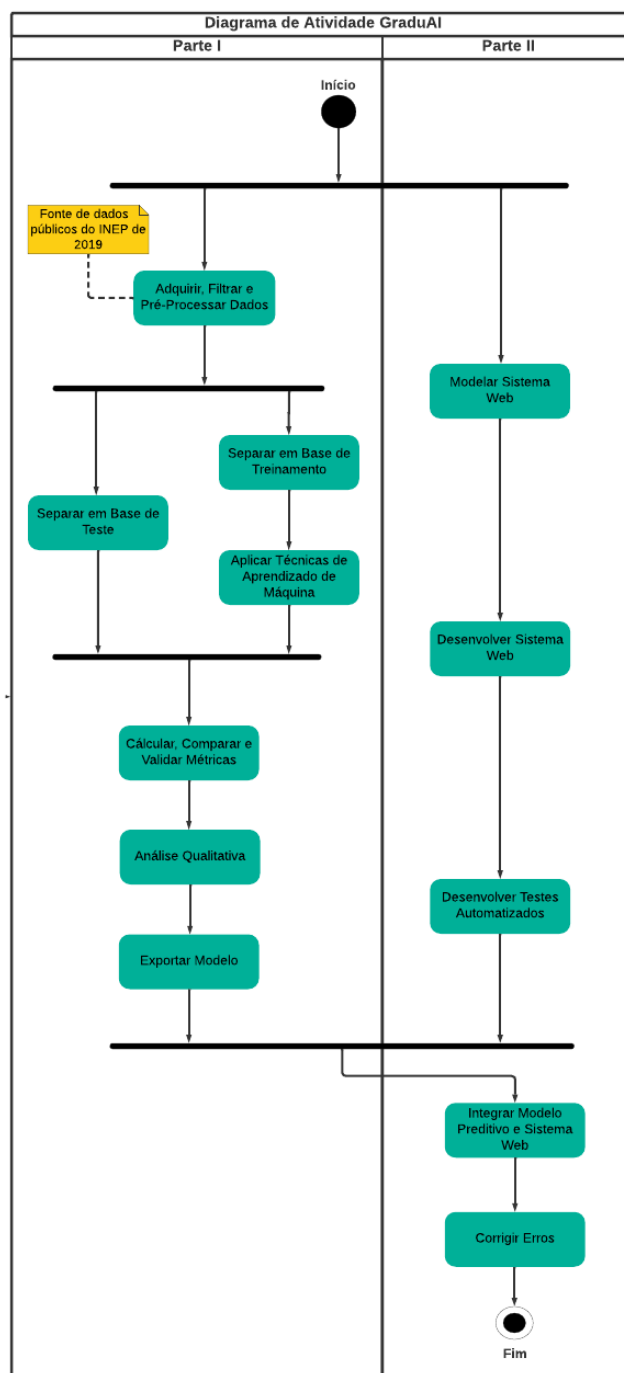
O presente trabalho tem a vantagem de identificar os alunos em risco de evasão em um momento anterior à inscrição no curso e na instituição de ensino desejada. Para isso, são utilizados apenas dados que estão disponíveis no censo do INEP. Certos atributos como notas e a quantidade de disciplinas em que o aluno está inscrito podem ser importantes para a previsão de evasão, como é exposto em (Santos, Menezes, de Carvalho, & Montesco, 2019), (DEL BONIFRO *et al.*, 2020) e (NIETO *et al.*, 2019), mas estas informações só podem ser obtidas uma vez que o aluno já está com a graduação em curso. Além disso, se considerarmos os estudos analisados que utilizam dados acadêmicos coletados ao longo do curso, como notas e progressão, eles atingem uma acurácia média de 80,37%, enquanto o estudo que não utiliza esses dados registra 79,54%, um valor relativamente próximo.

### 3. Metodologia

A Figura 1 mostra um diagrama de atividade que resume a metodologia adotada para o trabalho, que será aprofundada neste capítulo e separada em duas partes principais:

**Parte I.** Construção do modelo para previsão de evasão.

**Parte II.** Construção do sistema web para aplicação do modelo e visualização dos resultados.



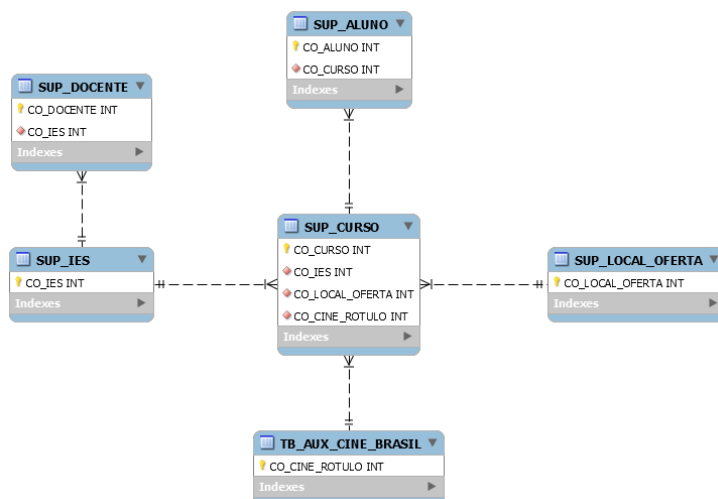
**Figura 1. Diagrama de atividade da metodologia.**

### 3.1. Parte I: Mineração de Dados Educacionais

A Parte I do trabalho envolve a obtenção, preparação e pré-processamento dos dados, assim como o treinamento dos algoritmos.

O desenvolvimento da Parte I do trabalho foi realizado com a linguagem Python e a biblioteca Scikit-learn (Pedregosa, et al., 2011). Estas ferramentas foram escolhidas por serem simples e eficientes para análise preditiva de dados, além de muito utilizadas na comunidade científica e reutilizáveis em vários contextos.

Os dados disponibilizados pelo INEP foram obtidos, filtrados e pré-processados. Para a montagem da base de dados utilizada pelos modelos preditivos, foram exploradas as relações entre as tabelas disponibilizadas pelo INEP. A relação entre estas tabelas, pode ser vista na Figura 2, que representa um diagrama de entidade-relacionamento com apenas os atributos que identificam as ligações entre as tabelas. Todos os dados, assim como a descrição detalhada de cada um, podem ser encontrados no portal do INEP<sup>1</sup>.



**Figura 2. Diagrama entidade-relacionamento dos microdados disponibilizados pelo INEP em 2019.**

A partir disso, foi feita a separação da base original em base de teste e de treinamento, com as proporções de 25% e 75%, respectivamente. No treinamento, foi aplicada a técnica *Random Forest*, por ter sido a que obteve os melhores resultados em trabalhos anteriores e por permitir a fácil identificação das características mais determinantes para a classificação dos dados. Depois, foi realizada a validação cruzada do modelo. Por fim, as importâncias das variáveis foram calculadas e os resultados foram comparados com os do ano anterior.

### 3.2. Parte II: Desenvolvimento do Sistema Web e API

A Parte II do trabalho envolve o desenvolvimento do sistema web e da API GraduAI, descrevendo os requisitos, a arquitetura, as tecnologias utilizadas, a importação dos dados e o controle de qualidade do *software*.

Para o desenvolvimento da aplicação web, foram selecionadas tecnologias que estão em evidência. Assim, foi possível criar um sistema web em NodeJS<sup>2</sup>, que no seu lado cliente (*frontend*) utilize o Javascript<sup>3</sup>, e no seu lado servidor (*backend*) utilize o TypeScript<sup>4</sup> e Python. Dessa forma, é possível utilizar os algoritmos de aprendizado de máquina e demais recursos fornecidos pela biblioteca Python no aplicativo NodeJS, como é feito no GraduAI.

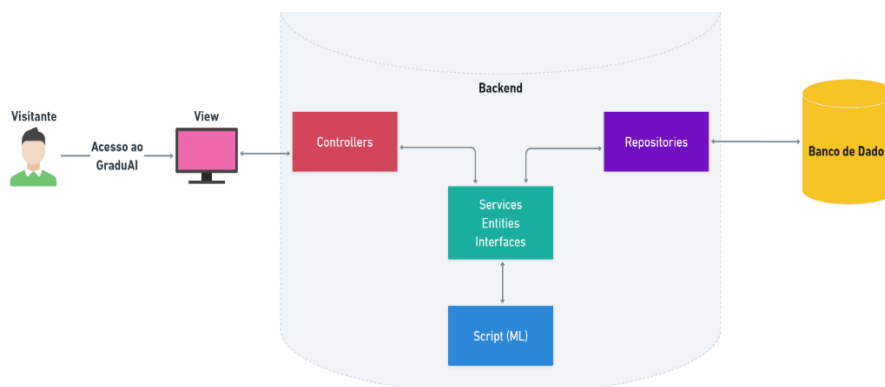
<sup>1</sup> <https://www.gov.br/inep/pt-br>

<sup>2</sup> <https://nodejs.org/en/>;

<sup>3</sup> <https://www.javascript.com/>;

<sup>4</sup> <https://www.typescriptlang.org/>;

O GraduAI utiliza a arquitetura em camadas, a fim de obter uma maior manutenibilidade e escalabilidade. As camadas utilizadas na aplicação foram: *controllers*, *services*, *repositories*, *entities*, *views* e *interfaces*. As *controllers* são responsáveis pelas requisições e controles dos fluxos. Enquanto isso, as *services* implementam as lógicas da regra de negócio. As *repositories*, fazem a manipulação do banco de dados e as *views* são responsáveis pela renderização da aplicação. Por fim, temos as *entities* e as interfaces, camadas que cuidam das entidades e interfaces do sistema. A Figura 3 demonstra como funciona a comunicação entre as camadas do sistema web e da API.



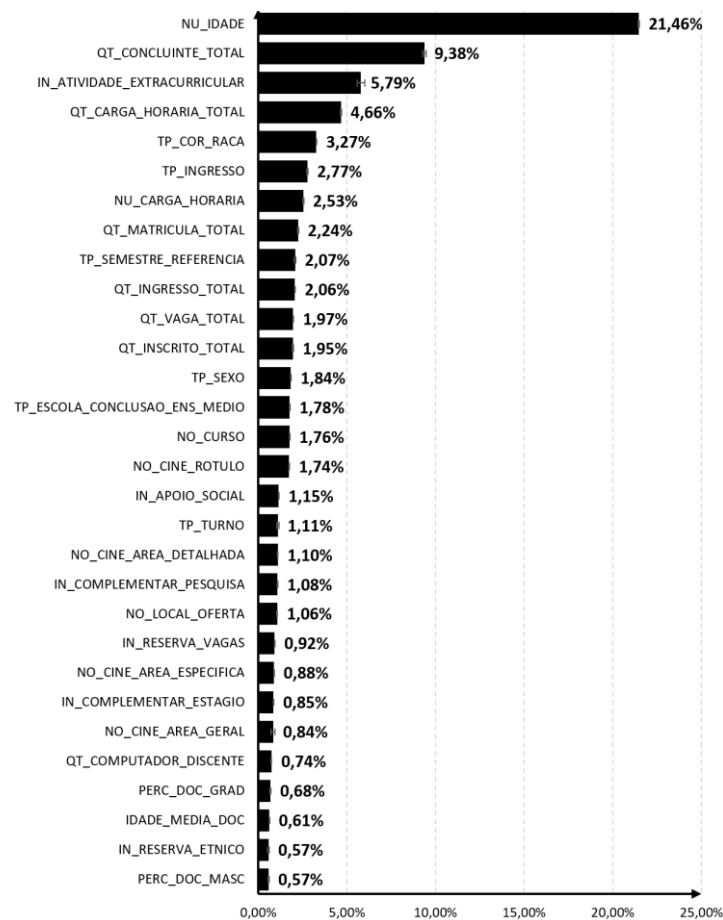
**Figura 3. A comunicação entre as camadas.**

Além disso, o sistema foi construído utilizando conceitos e boas práticas de desenvolvimento de sistemas de software que trazem benefícios à todo o seu ecossistema. Foi adotado o conceito de *Mobile First* (DUONG *et al.*, 2014), portanto, seu funcionamento foi pensado para pequenas e grandes resoluções de tela, trazendo uma melhor experiência aos usuários. Ademais, foram adotadas boas práticas de codificação para obter maior legibilidade e manutenibilidade, com o *Clean Code* (MARTIN, 2009) e a Modularização (ALMEIDA, 2021).

O fluxo desenvolvido funciona através da coleta de dados por meio de formulários que estão disponíveis no GraduAI. A fim de diminuir o preenchimento das 151 características necessárias para alimentar o modelo preditivo, o formulário funciona preenchendo alguns campos de forma automática. Por exemplo, se o aluno seleciona o curso e a instituição, as informações destas entidades já serão enviadas junto com a submissão do formulário. Com isso, o aluno precisa apenas preencher dados referentes às suas pretensões acadêmicas e alguns dados pessoais.

#### **4. Resultados e Discussões**

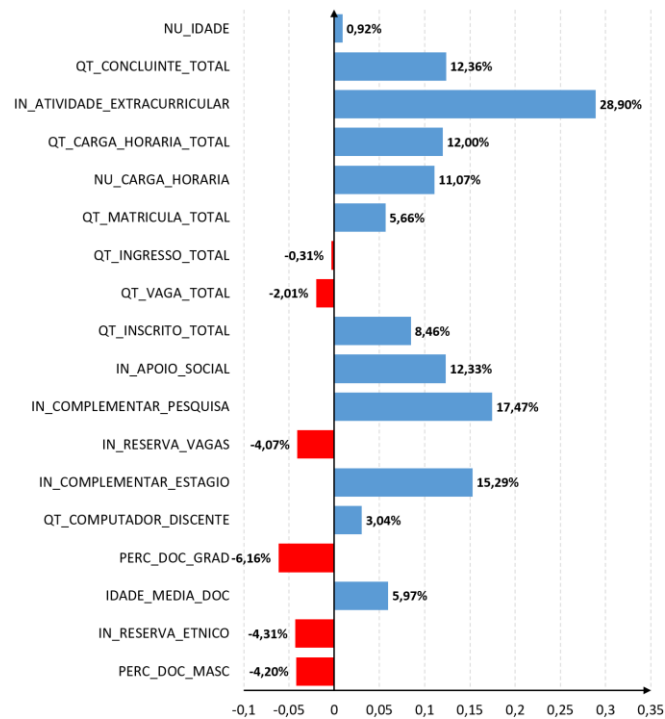
A análise das métricas obtidas após a fase de treinamento e validação cruzada de cada uma das técnicas mostra que o método *Random Forest* apresentou os melhores e mais equilibrados resultados, com acurácia média de 80%. Uma vez identificado e validado o modelo que obteve a melhor capacidade preditiva dentre todos os avaliados, foi possível extrair a informação de importância relativa que cada característica presente na base possui na realização das previsões. É importante salientar que esta informação representa apenas uma relação estatística entre as características e a evasão, o que não corresponde, necessariamente, a uma relação de causalidade. A Figura 4 mostra os 30 atributos mais importantes para a previsão de evasão para um estudante em uma instituição pública de ensino superior Brasil de acordo com o modelo construído.



**Figura 4. Atributos determinantes para a previsão da evasão.**

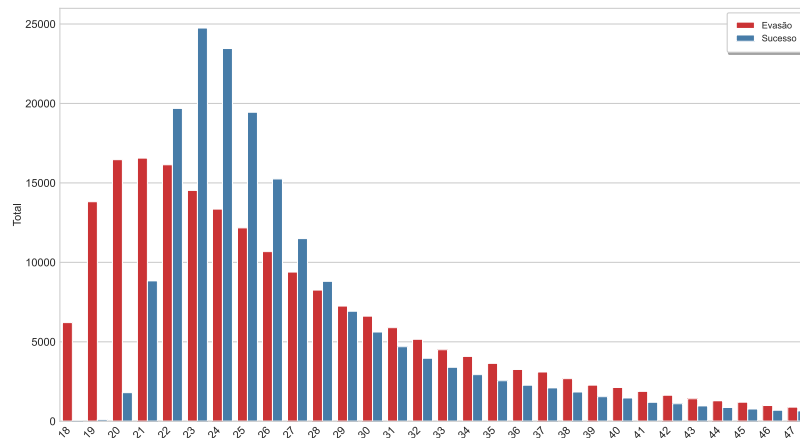
Para analisar como as variáveis numéricas se relacionam com a taxa de evasão, a correlação de Pearson entre elas e a evasão dos estudantes foi calculada, gerando os resultados mostrados na Figura 5. Como a variável de evasão pode assumir os valores ordinais 0 e 1, representando “evasão” e “sucesso”, respectivamente, uma correlação positiva indica que a característica está relacionada com o sucesso do aluno, enquanto uma correlação negativa indica que a característica está relacionada com a evasão. Isso não significa que existe dependência ou causalidade direta entre as características e a evasão, mas ajuda a identificar os padrões característicos de alunos que possuem maior chance de evasão, e que devem ser alvos de políticas públicas de alguma forma de assistência.

Por exemplo, um dos atributos determinantes, de acordo com a Figura 4, é a participação do aluno em atividades extracurriculares. A Figura 5 mostra uma forte correlação positiva desta característica com a variável de evasão, ou seja, pode-se dizer que se o aluno realiza atividades extracurriculares, sua probabilidade de obter sucesso na formação é maior. Este fato é corroborado pelo trabalho de Lamers, Santos e Toassi (2017), que mostra que o fato do aluno se dedicar a atividades que vão além da sala de aula possui forte relação com sua probabilidade de conclusão do curso. Esta mesma análise pode ser feita com relação às outras características identificadas.



**Figura 5. Correlação de Pearson dos atributos determinantes para a evasão.**

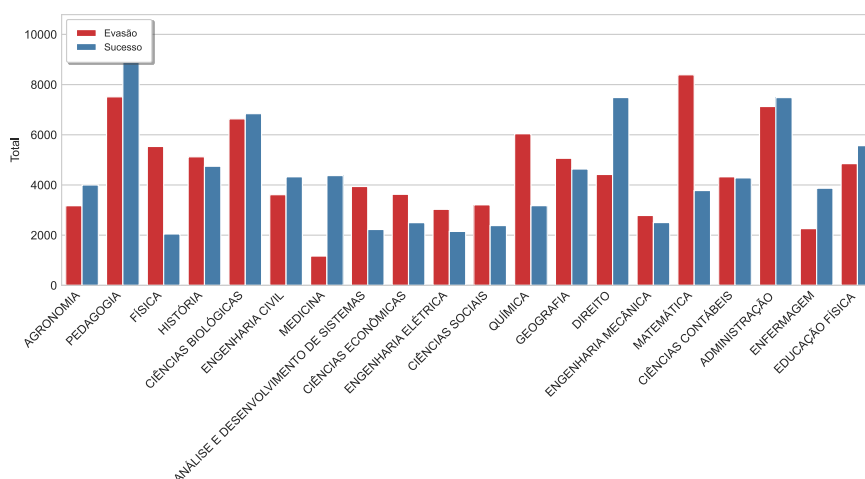
É importante ressaltar que a correlação de Pearson revela apenas a correlação linear entre duas variáveis. Algumas vezes, as variáveis podem possuir uma correlação não-linear, que não é facilmente observada. Percebe-se, ao analisar a Figura 5, que não existe uma forte correlação linear entre a idade dos alunos e a evasão. Investigando um pouco mais, podemos visualizar no histograma mostrado na Figura 6 que a idade dos alunos com evasão ou sucesso seguem distribuições de probabilidade diferentes. Este fato mostra a capacidade do classificador de identificar padrões normalmente difíceis de serem observados. Analisando a Figura 6, é possível notar que há uma desproporção entre sucesso e evasão para jovens entre 18 e 21 anos, onde a evasão está mais evidente. Além disso, também é observável que, no período de 22 até 28 anos, o sucesso é mais provável. Por fim, podemos observar que a probabilidade de evasão, a partir dos 29 anos, se torna maior que a de sucesso.



**Figura 6. Histograma da variável NU\_IDADE, separada em casos de evasão e sucesso.**



Como as variáveis categóricas nominais não podem ser relacionadas com a evasão pela correlação de Pearson, torna-se necessária uma análise individual para entender a relação delas com a evasão. Por exemplo, diferentes cursos apresentam maiores ou menores taxas de evasão. A Figura 7 ilustra esse fato, mostrando a proporção de casos de evasão e sucesso nos 20 cursos com maior número de alunos. A análise de Filho et al. (2007) indica que a área de conhecimento associada ao curso escolhido pelo aluno realmente tem relação direta com a taxa de evasão. Assim, pode-se afirmar que os resultados aqui obtidos e confirmados por trabalhos conceituais relacionados revelam as principais características, em um aspecto geral, dos estudantes com maior risco de evasão.



**Figura 7. Total de casos de evasão e sucesso para os 20 cursos com maior número de alunos na base de dados.**

Com o sistema GraduAI<sup>5</sup>, todas estas informações estão disponíveis para qualquer estudante e instituição de educação. Além disso, o estudante pode inserir seus dados e fornecer informações do curso pretendido no sistema para descobrir uma estimativa da probabilidade de sucesso ou evasão, baseada no modelo preditivo desenvolvido no presente trabalho com técnicas de aprendizado de máquina. As Figuras 8 e 9 mostram um exemplo de telas do sistema.

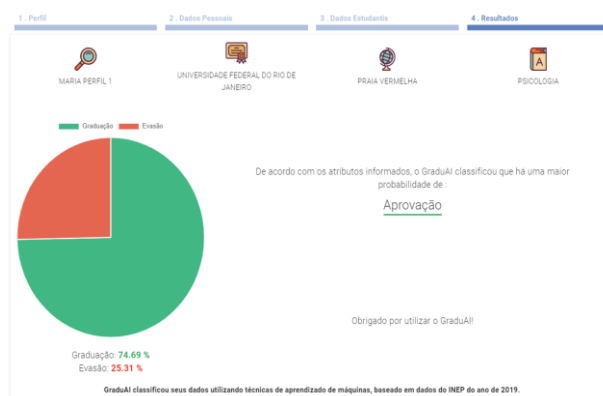
The form includes the following fields:

- 1. Perfil:** University (UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ)), Campus (PRAIA VERMELHA).
- 2. Dados Pessoais:** Course (PSICOLOGIA), Turn (Matutino), Ingresso (ENEM).
- 3. Dados Estudantis:** Reservas (Informa se o aluno pretende participar de programa de reserva de vagas), Semestre (Primeiro semestre), Escola Conclusão Ensino Médio (Privada), Atividades Extracurriculares, Bolsas, Apoio Social, Financiamento Estudantil (Não), PARFOR (Não).
- 4. Resultados:** (Empty field)

Buttons: "VOCÊ ESTÁ CRIANDO UM NOVO PERFIL" and "PRÓXIMO".

**Figura 8. Tela de preenchimento de dados do sistema GraduAI.**

<sup>5</sup> <https://www.graduai.com/>



**Figura 9. Tela de resultados da análise do sistema GraduAI.**

## 7. Conclusões

O estudo realizado teve como objetivo primário fornecer aos estudantes informações sobre a probabilidade de sucesso ou evasão em cursos de graduação, preenchendo uma lacuna de conhecimento que pode os auxiliar na tomada de decisão sobre os rumos a serem seguidos em suas carreiras. Para isso, foram criados um modelo preditivo, um sistema web e uma API que fornece a previsão das chances de conclusão e evasão.

Para elaboração do modelo preditivo, foram utilizadas 5 técnicas de aprendizado de máquina, que foram treinadas com a base de dados disponibilizada pelo INEP, sobre o ano de 2019. Dentre essas 5 técnicas, o algoritmo Florestas Randômicas se destacou, uma vez que registrou os melhores resultados, com aproximadamente 80% de acurácia, e revelou quais os atributos mais importantes para determinar o padrão dos alunos que irão evadir, como a idade, a prática de atividades extracurriculares e o recebimento de algum apoio social. Ao final, esse modelo foi exportado para a API construída, de forma que ficasse acessível através de uma rota.

O sistema web funcionou como um mediador entre os alunos e o modelo preditivo, que ficou disponível na API. Dessa forma, qualquer aluno com acesso à Internet pode descobrir suas chances de sucesso e evasão em determinado curso. Além disso, através da API, é possível que qualquer corpo administrativo forneça as informações dos alunos no momento de inscrição e obtenha as chances do aluno se formar ou evadir. Com essa informação e com as características que mais importantes para obtê-la, os corpos administrativos podem iniciar um plano de mitigação de evasão, a fim de manter os alunos inscritos dentro dos seus cursos e instituições. Diante disso, o objetivo foi alcançado.

Ainda há, entretanto, espaço para novas funcionalidades dentro do GraduAI. Seria interessante que, no resultado da análise, o sistema apresentasse automaticamente quais atributos que o aluno poderia mudar para aumentar suas chances. Além disso, para trabalhos futuros, seria interessante estudar como a pandemia do COVID-19 afetou a evasão nas universidades brasileiras. Repetir o processo de Mineração de Dados Educacionais realizado no presente trabalho, mas utilizando dados dos anos de 2020 e 2021, irá fornecer valiosas informações tanto do impacto da pandemia nas taxas de evasão, quanto das alterações nas características mais importantes para sua previsão. O conhecimento desta variação pode auxiliar em uma avaliação mais ampla das consequências da pandemia na dinâmica educacional em instituições públicas de ensino superior no Brasil.

## Agradecimentos

Os autores agradecem ao CEFET-RJ pelo incentivo à esta pesquisa.

## Referências

- ALMEIDA, M. (2021). *INTEGRAÇÃO DE FERRAMENTAS PARA UM FIREWALL MODULAR E ESCALÁVEL COM USO DE CONTAINERS E COMPUTADORES DE PLACA ÚNICA*. Curitiba. Fonte: [http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/28033/1/CT\\_COTEL\\_2021\\_1\\_01.pdf](http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/28033/1/CT_COTEL_2021_1_01.pdf)
- Costa, S. L., & Dias, S. M. (2016). A permanência no ensino superior e as estratégias institucionais de enfrentamento da evasão. *JORNAL DE POLÍTICAS EDUCACIONAIS*, 9(18), pp. 51-60. doi:10.5380/jpe.v9i17/18.38650
- da Silva, P. M., Lima, M. N., Soares, W. L., Silva, I. R., Fagundes, R. A., & de Souza, F. F. (2019). Ensemble Regression Models Applied to Dropout in Higher Education. *8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 120-125). IEEE. doi:10.1109/BRACIS.2019.00030
- Del Bonifro, F., Zingaro, S., Lisanti, G., & Maurizio, G. (2020). Student Dropout Prediction. *International Conference on Artificial Intelligence in Education: AIED 2020*. 12163. Springer, Cham. doi:[https://doi.org/10.1007/978-3-030-52237-7\\_11](https://doi.org/10.1007/978-3-030-52237-7_11)
- Duong, T., Koehler, B., Liew, T., Mower, C., & Venkatraman, S. (2014). *IBM MobileFirst Strategy Software Approach*. IBM Redbooks.
- Filho, R. L., Motejunas, P. R., Hipólito, O., & Lobo, M. B. (set./dez. de 2007). A Evasão no Ensino Superior Brasileiro. *Cadernos de Pesquisa*, 37(132), 641-659. doi:10.1590/S0100-15742007000300007
- Freitas, F., Dewan, M., Hassan, M. M., de Albuquerque, V. C., Vasconcelos, F., & Peixoto, S. (1 de 9 de 2020). IoT System for School Dropout Prediction Using Machine Learning Techniques Based on Socioeconomic Data. *Electronics*, p. 1613. doi:<https://doi.org/10.3390/electronics9101613>
- Lamers, J., Santos, B., & Toassi, R. (03 de Abril de 2017). RETENÇÃO E EVASÃO NO ENSINO SUPERIOR PÚBLICO: ESTUDO DE CASO EM UM CURSO NOTURNO DE ODONTOLOGIA. *Educação em Revista*, 33, 1-26. doi:10.1590/0102-4698154730
- MARTIN, R. C. (2009). *Código Limpo: Habilidades Práticas do Agile Software* (1ª ed.). Alta Books.
- Nieto, Y., Gacia-Díaz, V., Montenegro, C., González, C. C., & González, R. (2019). Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions. *IEEE Access*, 7, 75007-75017. doi:10.1109/ACCESS.2019.2919343
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. Fonte: <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>

- Prestes, E. M., & Fialho, M. G. (2018). Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba. *Ensaio: Avaliação e Políticas Públicas em Educação*, 26(100), 869-889. doi:10.1590/s0104-40362018002601104
- Sales, A., Balby, L., & Cajueiro, A. (Agosto de 2016). Exploiting Academic Records for Predicting Student Drop Out: a case study in Brazilian higher education. *Journal of Information and Data Management*, 7(2), 166-180. Fonte: <https://periodicos.ufmg.br/index.php/jidm/article/view/343>
- Santos, K. J., Menezes, A. G., de Carvalho, A. B., & Montesco, C. A. (2019). Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout. *IEEE 19th International Conference on Advanced Learning Technologies (ICALT)* (pp. 207-208). IEEE. doi:10.1109/ICALT.2019.00068
- Saravanan, R., & Sujatha, P. (2018). A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. *Proceedings of the Second International Conference on Intelligent Computing and Control Systems*, pp. 945-949. doi:10.1109/ICCONS.2018.8663155
- Silva, S., Santos, É., Rabelo, L., & Luciano, M. (29 de Maio de 2020). Avaliando a política de Dados abertos no Legislativo brasileiro. *Revista Compolítica*, 10(1), pp. 137-160. doi:10.21878/compolitica.2020.10.1.351
- Teodoro, L. A., & Kappel, M. A. (2020). Machine Learning Applied to Academic Drop Out Prediction in Brazilian Public Universities (Aplicação de Técnicas de Aprendizado de Máquina para Predição de Risco de Evasão Escolar em Instituições Públicas de Ensino Superior no Brasil). *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, 28, 838-863. doi:10.5753/RBIE.2020.28.0.838