

## **SIDE – SISTEMA INTELIGENTE DE DETECÇÃO DE EVASÃO: APLICAÇÃO EM UMA INSTITUIÇÃO DE EDUCAÇÃO FEDERAL**

### **SIDE – INTELLIGENT ESCAPE DETECTION SYSTEM: APPLICATION IN A FEDERAL EDUCATION INSTITUTION**

### **SIDE – SISTEMA INTELIGENTE DE DETECCIÓN DE DESERCIÓN: APLICACIÓN EN UNA INSTITUCIÓN EDUCATIVA FEDERAL**

**Matheus Fogaça Bicudo de Almeida**

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – Campus Itapetininga

**Adriana Marques**

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – Campus São Paulo  
Agência Nacional de Água e Saneamento

**RESUMO.** O controle da evasão escolar requer ações integradas e um compromisso coletivo para garantir que todos os estudantes tenham acesso a uma educação de qualidade. Instituições de ensino superior devem tomar medidas para prevenir a evasão, o que pode incluir a oferta de programas de suporte acadêmico e financeiro, a criação de oportunidades de estágios e empregos para os estudantes, a sensibilização sobre a importância da educação superior, entre outras medidas. O aprendizado de máquina é uma ferramenta valiosa para prevenir a evasão escolar, pois permite a coleta e análise de dados para identificar os estudantes em risco, personalizar programas de apoio, monitorar o progresso e avaliar a eficácia de programas de prevenção. No Brasil, segundo a PNAD (Pesquisa Nacional por Amostra de Domicílios), no ano de 2019, cerca de 89,2% da população entre 15 e 17 anos estava incluída na taxa de escolarização, demonstrando o longo caminho a ser percorrido para alcançar a universalização proposta na Constituição de 1988. Dessa forma, a presente pesquisa teve como objetivo desenvolver um sistema de identificação de evasão utilizando Data Mining e Machine Learning, como um meio eficiente na identificação de evasão em instituições de ensino. Com base em informações dos próprios discentes do Instituto Federal de Educação, Ciência e Tecnologia – IFSP campus Itapetininga, obtiveram-se dados como local de residência, meio de transporte utilizado para chegar à instituição, motivação da matrícula, dentre outros fatores significativos na análise da probabilidade de evasão dos estudantes. O programa demonstrou quem possui maior probabilidade de evadir do corpo estudantil, ajudando os gestores a alocar os recursos e tomar as decisões necessárias. Os resultados obtidos nos dois modelos tiveram precisão de 86% em um modelo para análise das evasões e de até 100% em outro modelo, com pequenas oscilações devido à falta de coerência de alguns dos dados fornecidos.

**Palavras-chave:** Modelagem. Data mining. Gestão escolar. Instituições educativas. Permanência escolar.

Matheus Fogaça Bicudo de Almeida e Adriana Marques

**ABSTRACT.** Controlling school dropout requires integrated actions and a collective commitment to ensure that all students have access to quality education. Higher education institutions must take measures to prevent dropout, which may include offering academic and financial support programs, creating internship and job opportunities for students, raising awareness about the importance of higher education, among other measures. Machine learning is a valuable tool for preventing dropouts, as it allows data collection and analysis to identify at-risk students, customize support programs, monitor progress, and evaluate the effectiveness of prevention programs. In Brazil, according to the PNAD (Pesquisa Nacional por Amostra de Domicílios), in 2019, around 89.2% of the population, between 15 and 17 years old, was included in the schooling rate, demonstrating the long road to be traveled to reach the universalization proposed in the 1988 Constitution. Thus, this research aimed to develop a system of evasion using Data mining and Machine learning, as an efficient means of identifying evasion in educational institutions. Based on information from the students of the Federal Institute of Education, Science and Technology – IFSP campus Itapetininga, data were obtained such as place of residence, means of transport used to reach the institution, motivation for enrolling, among other factors that are significant in the analysis of the probability of a student dropping out. The program demonstrated who is most likely to drop out of the student body, helping managers to allocate resources and make the necessary decisions. The results obtained in the two models had an accuracy of 86% in a model for the analysis of dropouts and another model with an accuracy of up to 100% with small oscillations due to the lack of coherence of some of the data provided.

**Keywords:** Modeling. Data mining. School management. Educational institutions. School stay.

**RESUMEN.** Controlar la deserción escolar requiere acciones integradas y un compromiso colectivo para asegurar que todos los estudiantes tengan acceso a una educación de calidad. Las instituciones de educación superior deben tomar medidas para prevenir la deserción, que pueden incluir ofrecer programas de apoyo académico y financiero, crear pasantías y oportunidades laborales para los estudiantes, crear conciencia sobre la importancia de la educación superior, entre otras medidas. El aprendizaje automático es una herramienta valiosa para prevenir la deserción escolar, ya que permite la recopilación y el análisis de datos para identificar a los estudiantes en riesgo, personalizar los programas de apoyo, monitorear el progreso y evaluar la efectividad de los programas de prevención. En Brasil, según la PNAD (Pesquisa Nacional por Amostra de Domicílios), en 2019, alrededor del 89,2% de la población, entre 15 y 17 años, estaba incluida en la tasa de escolaridad, lo que demuestra el largo camino por recorrer para llegar a la universalización propuesta en la Constitución de 1988. Así, esta investigación tuvo como objetivo desarrollar un sistema de evasión utilizando la minería de datos y el aprendizaje automático, como un medio eficiente para identificar la evasión en las instituciones educativas. Con base en información de los estudiantes del Instituto Federal de Educación, Ciencia y Tecnología – IFSP campus Itapetininga, fueron obtenidos datos como lugar de residencia, medio de transporte utilizado para llegar a la institución, motivación para matricularse, entre otros factores que son significativos en el análisis de la probabilidad de que un estudiante abandone los estudios. El programa demostró quién tiene más probabilidades de abandonar el cuerpo estudiantil, ayudando a los administradores a asignar recursos y tomar las decisiones necesarias. Los resultados obtenidos en los dos modelos tuvieron una precisión del 86% en un modelo para el análisis de abandonos y otro modelo con una precisión de hasta el 100% con pequeñas oscilaciones debido a la falta de coherencia de algunos de los datos proporcionados.

**Palabras clave:** Modelado. Data mining. Gestión escolar. Instituciones educativas. Permanencia escolar.

## 1 INTRODUÇÃO

A evasão escolar refere-se ao abandono do ambiente escolar antes de concluir o ensino obrigatório, seja no ensino fundamental, técnico ou superior. Pode ocorrer por uma série de fatores, como desinteresse pelo aprendizado, problemas financeiros, dificuldades sociais ou familiares, falta de motivação, bullying, entre outros. A evasão pode ter graves consequências, tanto para o indivíduo como para a sociedade como um todo. As pessoas que abandonam a escola precocemente têm maiores dificuldades em encontrar emprego e ganhar um salário competitivo, o que pode levar a um ciclo interminável de pobreza e exclusão social. Além disso, pode afetar negativamente o desenvolvimento econômico de uma região ou país. Por essas razões, é importante que sejam implementadas políticas públicas eficazes para prevenir e combater a evasão escolar. Isso inclui a melhoria da qualidade da educação, a criação de programas de apoio para os estudantes, a sensibilização da sociedade sobre a importância da educação, entre outras medidas (Silva Filho; Araújo, 2017).

Com base nesta problemática, é possível combinar a tecnologia do *Data Mining* e de IA (Inteligência Artificial) em prol da resolução e redução dos níveis de evasão escolar, já que a evasão escolar é um problema presente em todas as regiões do mundo e que impacta diretamente a sociedade e a economia de forma negativa (Silva Filho; Araújo, 2017).

A evasão é uma questão que acontece ao redor do globo. A importância de antecipar o percentual de evasão, ou seja, reconhecer antecipadamente a evasão, ajuda a desenvolver planos prévios e soluções/procedimentos alternativos para lidar com esse tema. Vários tipos de estrutura de aprendizagem podem ser utilizados para limitar as chances de ser um desistente (Asha *et al.*, 2020).

O *Data Mining* ou mineração de dados é um processo no qual se utiliza a tecnologia para procurar e encontrar padrões ou anormalidades em um determinado grupo de dados, sendo possível encontrar, a partir destes, possíveis problemas ou fatores a serem analisados (Han; Kamber; Pei, 2012). *Machine Learning* ou aprendizado de máquina é um

meio em que o computador/máquina realiza a análise dos dados autonomamente, sem necessitar da intervenção humana direta (Goldschmidt; Passos, 2005).

O que faz dessas ferramentas perfeitas para serem utilizadas a favor da evasão escolar, pois este problema é algo de suma importância na sociedade atual devido ao grande impacto social e econômico que a taxa de evasão pode causar a longo prazo. Com o auxílio do *Data Mining* e do *Machine Learning*, a evasão escolar pode ser reduzida em valores altamente consideráveis, sendo benéfico para as instituições de ensino (Pimentel *et al.*, 2019).

A evasão escolar pode ocorrer devido a diversos fatores, variando de país para país e, dentro destes, de região para região, cultura para cultura, dentre outros. Cabe, assim, às instituições de ensino buscar meios de resolver tais problemas, como é o caso da UNESCO no mundo e do PNAD no Brasil.

Devido ao impacto da evasão e à influência da tecnologia, muitas pesquisas vêm sendo desenvolvidas em pequena escala pelo mundo, porém sem a devida atenção ao sistema social em que nos encontramos atualmente (Pimentel *et al.*, 2019).

Com o desenvolvimento de uma pesquisa aplicada, pode-se ver que a evasão está presente em todo local em que se tem uma instituição de ensino e o quão impactante a mesma é para o local (Han *et al.*, 2019).

Com base no exposto, a presente pesquisa visou desenvolver um modelo de predição de evasão mediante o software *Orange Data Mining*, usando as informações dos cursos de edificações e eletromecânica do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP campus Itapetininga, agrupando as principais características que levam à evasão nesses cursos técnicos. Buscando uma eficiência de ao menos 95% nas previsões das evasões para ter dados similares a pesquisas realizadas em outras regiões.

Embora a problemática da evasão não seja uma questão simples, podem ser encontrados métodos que visem soluções eficientes de predição e formulação de possíveis respostas a esse problema.

## **2 REVISÃO TEÓRICA**

A evasão escolar pode ter várias interpretações. Para alguns autores, evadir é o ato de desistir dos estudos sem dar continuidade aos mesmos. Já para outros, é o momento em que um aluno tranca a sua matrícula na escola, simbolizando assim uma descontinuidade nos estudos. A evasão escolar é um grande problema enfrentado pela área da educação. Socialmente, implica problemas para o indivíduo cuja escolaridade é baixa devido ao abandono dos estudos. Em países como o Brasil, onde a taxa de evasão é alta, pode acarretar prejuízos econômicos para a nação devido à falta de desenvolvimento local, causando estagnação social e econômica do indivíduo evadido e do país (Costa; Borges, 2019).

A evasão universitária é uma grande preocupação política em todo o mundo devido às suas consequências para o indivíduo, as instituições e a sociedade. As intervenções para reduzir a evasão devem tanto ajudar os alunos a ter acesso ao apoio familiar e desenvolver conexões pessoais na universidade para compensar a ausência de apoio familiar, quanto oferecer apoio acadêmico focado para alunos que não têm uma forte qualificação (Sosu; Pheunpha, 2019).

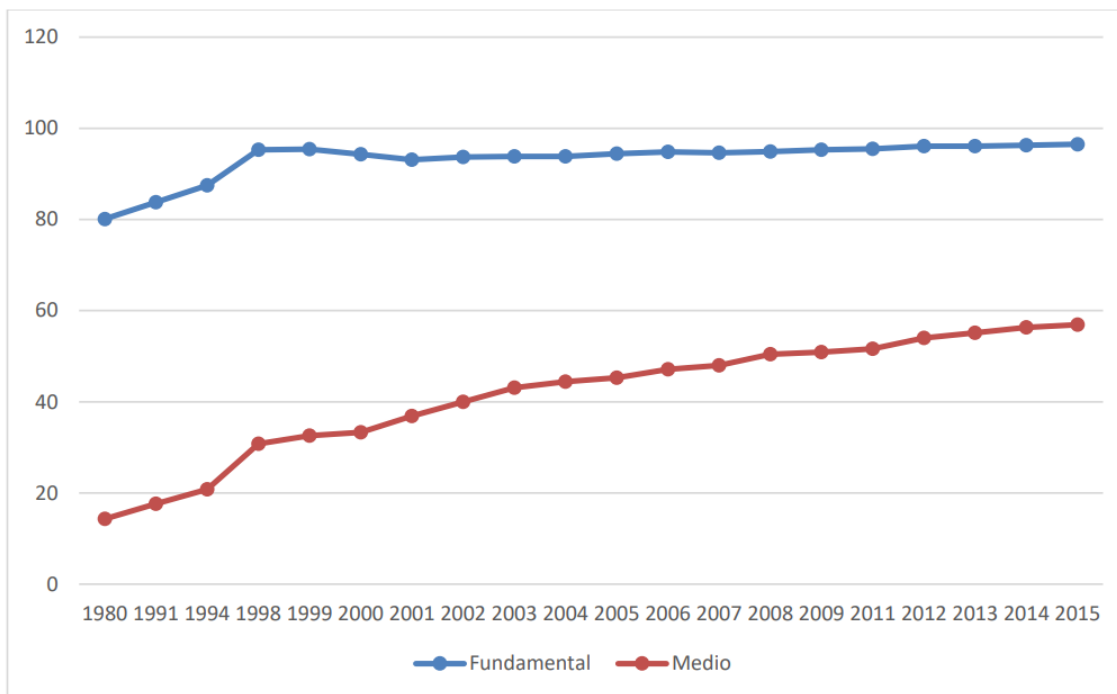
O problema é mundial. As taxas de evasão nos países europeus são uma das grandes questões a enfrentar num futuro próximo, conforme consta da estratégia Europa 2020. Em 2017, uma média de 10,6% dos jovens (18-24 anos) na UE-28 abandonaram precocemente o ensino e a formação, de acordo com as estatísticas do Eurostat (Agrusti; Bonavolontà, 2019).

O problema da evasão escolar no Brasil é tão relevante que, no ano de 2020, foi realizado um censo pelo Instituto Brasileiro de Geografia e Estatística (IBGE) sobre evasão

escolar e analfabetismo. O censo demonstrou que, de 50 milhões de jovens de 14 a 29 anos, cerca de 10,1 milhões não completaram a educação básica, seja por abandono ou por nunca terem frequentado a escola, o que demonstra a necessidade de garantir a permanência dos alunos nas escolas, principalmente no ensino médio. Tudo isso devido às consequências geradas para o desenvolvimento dos países e por ser um fato conhecido que um indivíduo sem escolaridade acaba sendo vítima da exclusão socioeducativa (Jornal Nacional, 2020).

Na Figura 1, apresenta-se a taxa de escolarização desde 1980 até 2015, indicando que o ensino fundamental teve quase 100% de eficácia e se manteve constante. No entanto, a escolarização no ensino médio apresentou uma taxa de menos de 60% no último ano, mostrando a taxa da população discente que não conclui o ensino básico.

**Figura 1 – Taxa líquida de escolarização por níveis de ensino**



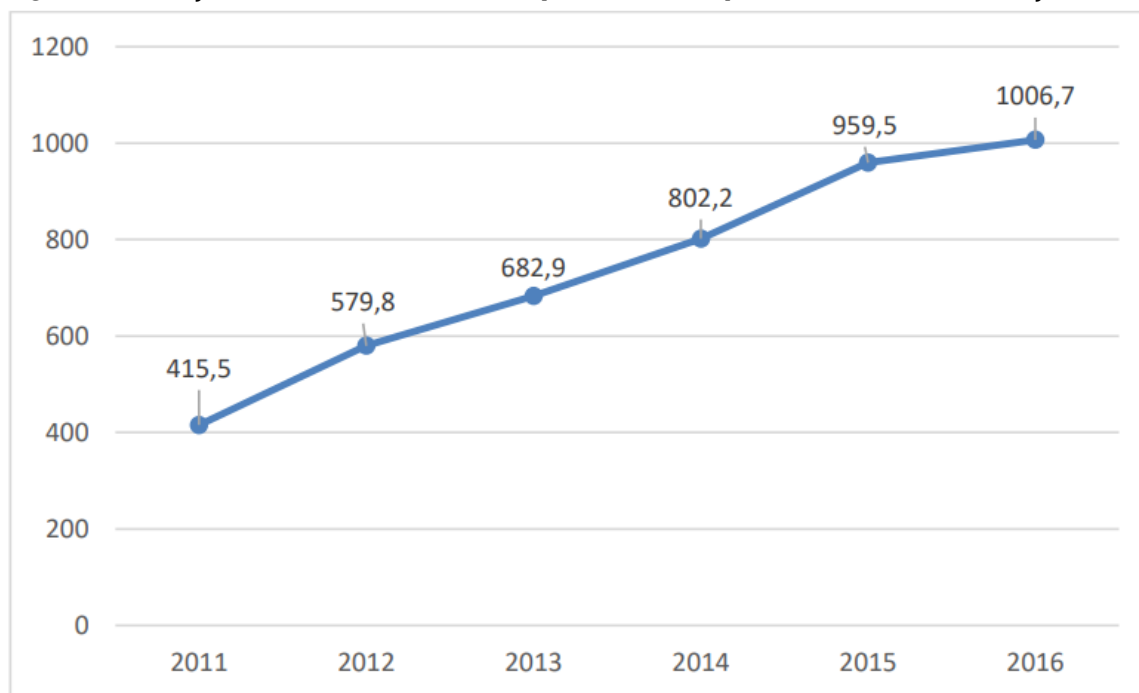
Fonte: Ministério da Transparência e Controladoria-Geral da União (2018).

## 2.1 Programas de Controle da Evasão

O controle da evasão escolar é uma tarefa importante para garantir que os estudantes concluam o ensino e tenham as habilidades e conhecimentos necessários para participar ativamente da economia do país. Medidas para controlar a evasão incluem ações integradas e compromisso por parte das instituições e da sociedade como um todo.

No Brasil, existem programas oferecidos para escolas, universidades e faculdades públicas, particulares e federais sendo o PNAES (Plano Nacional de Assistência Estudantil) o programa responsável por distribuir o dinheiro público para sistemas de ensino público, o qual fornece desde bolsas para serem desenvolvidas na instituição a até auxílios moradia e alimentação como apresentado na Figura 2 (ANPEC, 2018).

**Figura 2 – Orçamento em Milhões disponibilizado pelo PNAES a instituições federais**



Fonte: Ministério da Transparência e Controladoria-Geral da União (2018).

Para as instituições particulares, existe o programa FIES (Financiamento ao Estudante do Ensino Superior), que financia da metade ao total dos custos do curso, e o

PROUNI (Programa Universidade Para Todos), que distribui bolsas de estudo que cobrem o mesmo valor do FIES. Esses programas oferecem, respectivamente, auxílio para alunos com renda mensal per capita maior que um salário-mínimo por mês e para alunos com renda menor que um salário-mínimo e meio por mês (ANPEC, 2018).

Para o ensino público federal, foi criado em 2007 o REUNI (Reestruturação e Expansão das Universidades Federais), que visa instituir condições para a expansão do ingresso às universidades públicas e garantir a continuidade nos estudos, incluindo também a inclusão e assistência estudantil. Dentre os três programas, o REUNI foi o que apresentou melhor desempenho (ANPEC, 2018).

Devido a esses benefícios provenientes do governo, o índice de matrículas para graduação aumentou consideravelmente, principalmente nas instituições públicas. No Censo de Educação Superior de 2015, mencionava-se que, de 1980 até 2015, houve um aumento de 75,7% de matrículas em cursos de graduação no país em instituições particulares, enquanto nas públicas o aumento foi de 24,3%.

E com o aumento de matrículas, surge a necessidade de oferecer mais vagas à medida que a população se desenvolve e cresce, ocasionando assim mais investimentos governamentais em novas universidades e novas instituições públicas. O mesmo censo indicou que o número de universidades federais criadas de 2003 até 2010 aumentou em 24%, ou seja, 14 novas instituições.

## **2.2 Evasão Escolar nos Institutos federais**

Devido ao fato de que os institutos federais oferecem um ensino técnico profissionalizante, os números de evasão escolar tendem a ser superiores quando comparados com as escolas públicas e particulares. Isso se deve a uma série de fatores, desde a falta de preparo para o tipo de ensino até o perfil dos estudantes que frequentam a instituição, já que as federais atendem muitos alunos cujas condições financeiras são



menores e que não possuem na família um histórico de estudo, o que é um alto indicador da evasão escolar.

Com base nessa problemática, existem políticas públicas que visam diminuir a desigualdade social e a falta de informação da família sobre o estudo, buscando assim maior incentivo. Entre elas encontram-se a Constituição Brasileira de 1988, a Lei de Diretrizes e Bases da Educação e o sistema de cotas, além do Decreto Presidencial nº 7.234/2010 (Costa; Borges, 2019). O PNAES define que os critérios e a metodologia de seleção dos alunos de graduação a serem beneficiados são de competência das instituições, devendo atender prioritariamente estudantes oriundos da rede pública de educação básica ou com renda familiar per capita de até um salário-mínimo e meio (ANPEC, 2018).

Em alguns casos, a taxa de evasão chega a valores muito altos, apresentando um índice de aprovação muito baixo, no qual existe uma relação inversamente proporcional entre a conclusão do ensino e o grau de ensino: quanto maior for o grau de ensino oferecido pelo curso da instituição federal, menor será a taxa de aprovação (Henriques, 2018). Devido a essa relação, segundo aponta a Plataforma Nilo Peçanha, as instituições federais apresentam um índice de eficiência acadêmica de 43,76%. No Brasil, a taxa de evasão média dos institutos federais é maior que a própria conclusão do ensino, sendo a evasão superior à conclusão em 5,73%, o que corresponde a 17.742 estudantes.

### **2.3 Machine Learning Aplicada à Evasão escolar**

O Machine Learning (Aprendizado de Máquina) pode ser aplicado à prevenção da evasão escolar de várias maneiras. Algumas dessas aplicações incluem:

- Predição de risco de evasão, já que algoritmos podem ser treinados com dados de estudantes para identificar os que estão em risco de abandonar a escola. Esses algoritmos podem levar em consideração vários fatores, como

o desempenho acadêmico, a presença em aulas, o envolvimento em atividades extracurriculares, entre outros;

- Análise de dados para identificar fatores críticos que possam estar contribuindo para a evasão escolar;
- Personalização de programas de apoio para estudantes em risco de evasão. Por exemplo, um algoritmo pode recomendar a um estudante um tutor particular ou um programa de estudo para ajudá-lo a superar suas dificuldades;
- Monitoramento contínuo do progresso dos estudantes e identificação precoce de sinais de evasão escolar;
- Otimização de programas de prevenção da evasão escolar ao longo do tempo (Ernández-García *et al.*, 2021)

O Machine Learning (Aprendizado de Máquina) aplicado à evasão escolar é algo que vem sendo cada vez mais desenvolvido e recorrente nos dias atuais, devido à expansão da tecnologia e à crescente necessidade de estudar meios para diminuir os índices de evasão escolar no mundo todo. Isso tem levado a pesquisas que antes eram muito utilizadas por empresas em análise de dados sequenciais e outros fins a atuarem na área educacional (Carrano; Albergaria; Infante; Rocha, 2019).

Com isso, diferentes softwares e programas vêm sendo desenvolvidos, obtendo resultados que variam de acordo com a região e o tipo de ensino analisado. Segundo o estudo realizado por Pimentel *et al.* (2019), “a evasão em cursos de graduação presencial no Brasil atingiu 30,1% na rede privada e 18,5% na rede pública. Nos cursos de educação a distância (EaD), o índice chegou a 36,6% na rede privada e a 30,4% na pública.”

Devido a isso, toda tecnologia aplicada à evasão visa demonstrar e prever quais alunos possuem possibilidade de evadir do ensino. A partir de um sistema inteligente, também conhecido como Inteligência Artificial (IA), são inicialmente recolhidos dados quantitativos sobre o curso, como o número de alunos matriculados, o número de alunos que evadiram e quantos concluíram o ensino. Isso gera um resultado inicial que demonstra

a eficiência, para depois analisar os dados dos alunos que evadiram e estabelecer um padrão (Araruna *et al.*, 2015).

Com base nas informações levantadas, é possível remanejar as ações promovidas pela instituição para melhorar a eficiência do combate à evasão escolar. Com os dados sequenciados e padronizados pela IA, elimina-se o erro humano na análise e são demonstrados no sistema os fatores mais impactantes que podem levar um aluno a evadir. Isso torna muito mais fácil e rápido agir para garantir a continuidade no ensino dos discentes (Maria; Damiani; Pereira, 2016).

No entanto, são escassos os estudos que investigam o momento ideal para identificar os alunos em risco de evasão e o tipo de dados a serem utilizados. Um estudo realizado na Espanha criou um modelo de previsão baseado em árvores de classificação, usando dados sociodemográficos e acadêmicos de 935 alunos do primeiro ano. As informações foram coletadas em três momentos diferentes ao longo do primeiro semestre do primeiro ano universitário. Identificou-se a necessidade de uma intervenção precoce no primeiro ano para evitar a evasão e, igualmente, que os dados de desempenho acadêmico servem como um bom preditor (Ortiz-Lozano *et al.*, 2020).

O ensino técnico nos institutos federais de ensino é uma importante oferta de educação no Brasil, que tem como objetivo formar profissionais técnicos em áreas estratégicas para o desenvolvimento econômico e social do país. Os institutos federais oferecem cursos técnicos de nível médio em áreas como agropecuária, meio ambiente, construção civil, informática, eletrônica, entre outras (Minghelli, 2018).

Além disso, a maioria dos institutos federais de ensino também oferece cursos técnicos superiores, que são cursos de formação de nível técnico superior com duração de dois anos. Esses cursos técnicos superiores oferecem uma formação mais especializada e avançada, preparando os estudantes para o mercado de trabalho ou para prosseguir seus estudos em uma faculdade (Minghelli, 2018).

### **3 METODOLOGIA**

Para realizar o modelo proposto no trabalho, foi utilizada a plataforma Orange para criar a base do modelo, já que esse programa já possui uma IA própria em funcionamento, o que facilita o processo de criação do modelo. Para utilizá-lo de forma eficaz, foi realizado um curso básico online para aprender as principais ferramentas do programa e como relacionar dados com estas para iniciar o estabelecimento de padrões sequenciais de informações.

Orange é uma plataforma de ciência de dados e Machine Learning aberta, que permite que usuários sem conhecimento técnico possam explorar, analisar e visualizar dados, bem como construir e testar modelos. A plataforma oferece uma ampla gama de ferramentas para lidar com tarefas comuns de ciência de dados, como pré-processamento de dados, seleção de modelos e avaliação de desempenho.

A plataforma Orange é baseada em software livre e é amplamente utilizada em muitos campos, incluindo pesquisa acadêmica, empresas, governo e setor público. É especialmente útil para análise de dados explicativos, aprendizado supervisionado e aprendizado não supervisionado.

Os dados foram extraídos da base de dados do SUAP (Sistema Unificado de Administração Pública), sistema utilizado pelo Instituto Federal de São Paulo, no qual foram obtidos os dados de evasão escolar do Curso Técnico em Edificações concomitante ou subsequente ao Ensino Médio e do Curso Técnico em Eletromecânica Integrado ao Ensino Médio desde seu início até os dias atuais, sendo respectivamente os períodos de 2016-2020 e 2012-2020. Os dados, em Excel, foram padronizados para poder ser importados no software Orange.

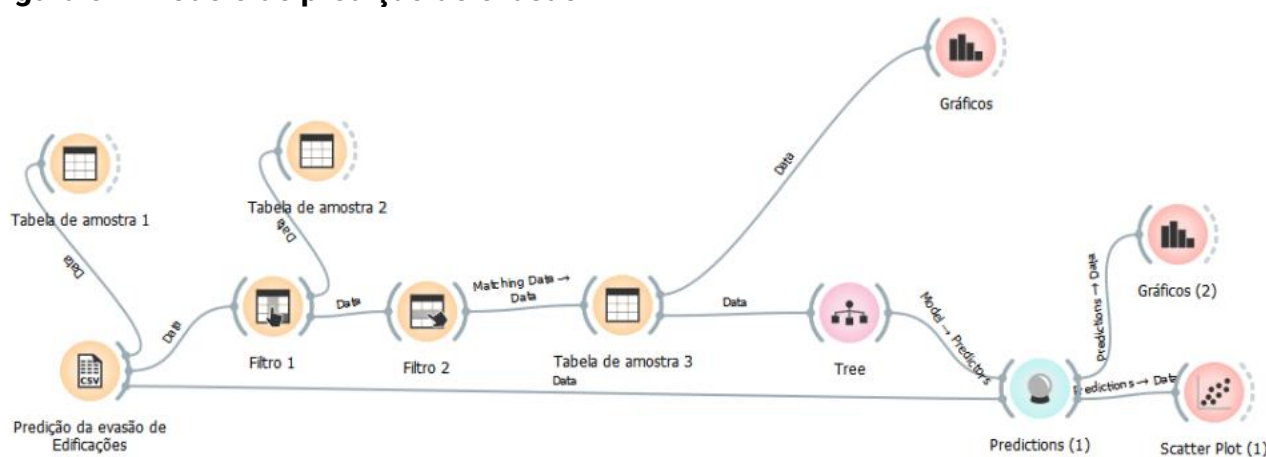
O Sistema Unificado de Administração Pública (SUAP) é uma plataforma eletrônica que permite a interação entre os cidadãos e a administração pública. Ele visa simplificar e modernizar o processo de solicitação de serviços públicos e garantir uma comunicação mais eficiente e transparente entre o cidadão e o Estado.

O software *Orange Data Mining* é um conjunto abrangente de software baseado em componentes para aprendizado de máquina e mineração de dados, desenvolvido no Laboratório de Bioinformática da Faculdade de Ciência da Computação e da Informação da Universidade de Ljubljana, Eslovênia, em conjunto com a comunidade de código aberto. É software livre e pode ser redistribuído e/ou modificado sob os termos da *GNU General Public License*, conforme publicado pela *Free Software Foundation* (Universidade de Ljubljana, 2022).

Para a criação do modelo na plataforma *Orange*, foi necessário encontrar informações para utilizar como variáveis relacionadas ou predecessoras à evasão, como a escola de origem, a renda familiar, o sexo e a etnia. Após a concessão de alguns arquivos, foram estudadas quais seriam as variáveis disponíveis para análise e, com base nisso, foi criado o seguinte modelo conforme mostra a Figura 3.

No início do modelo, tem-se o documento com os dados dos alunos que evadiram no curso de Edificações, para testar o programa e prever a evasão que aparece em porcentagem. Como se trata de dados de alunos evadidos, o resultado deve ser de 100%.

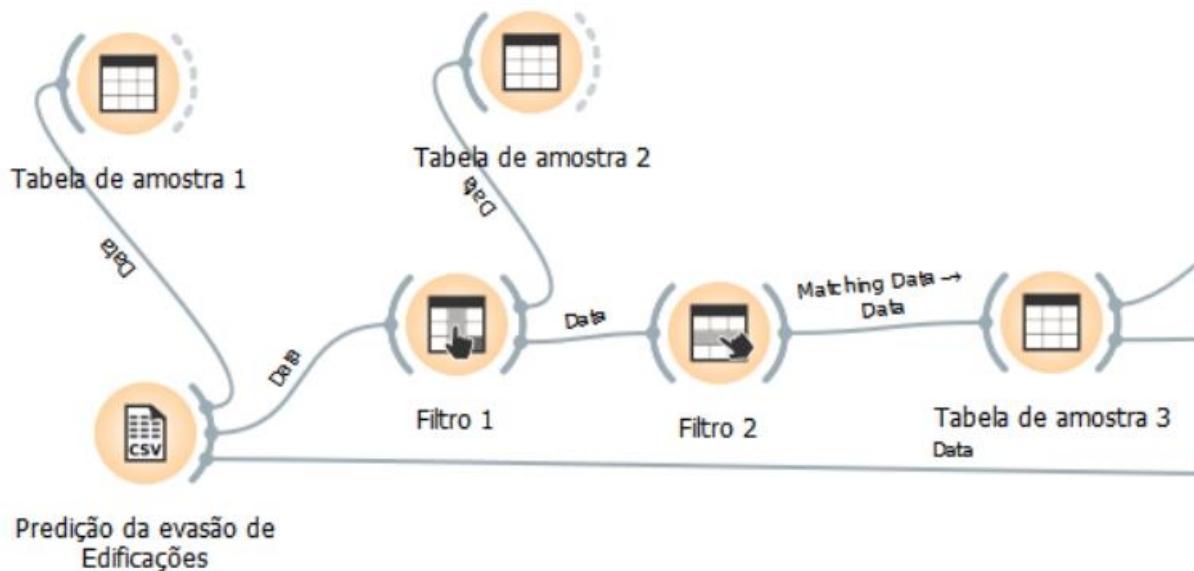
**Figura 3 – Modelo de previsão de evasão**



**Fonte:** Elaborado pelos autores, com base na pesquisa realizada.

E com isso, a programação dividiu-se em duas partes. Na primeira parte da programação, foram utilizados dois blocos de “filtro” para que informações de texto que não precisam ser analisadas ou que podem gerar problemas na hora da análise ficassem de fora no final. Além disso, foram utilizados três blocos denominados “tabela de amostra”, que servem para mostrar como os dados estão organizados. A partir daí, é possível ver se os dados estão sendo filtrados e devidamente analisados, conforme mostra a Figura 4.

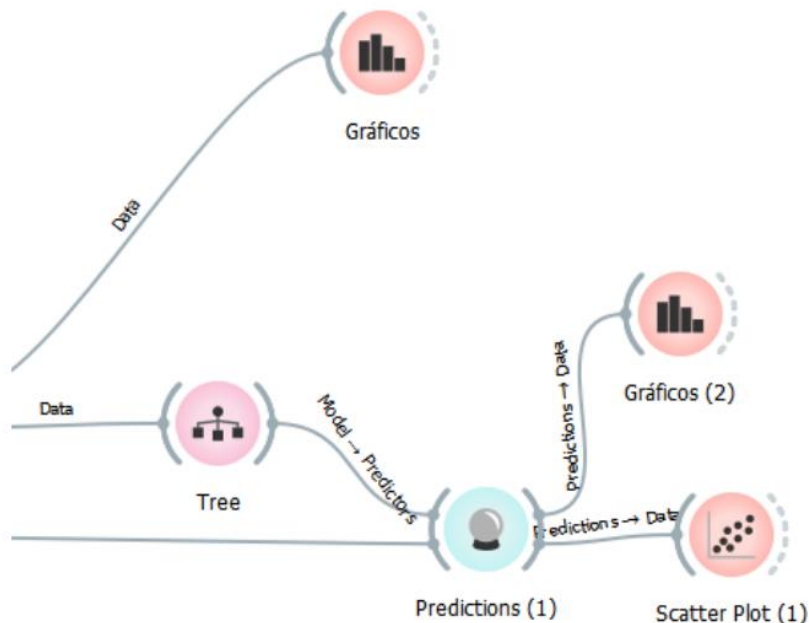
**Figura 4 – Parte inicial da programação de predição da evasão**



Fonte: Elaborado pelos autores, com base na pesquisa realizada.

E a segunda parte da programação, conforme mostra a Figura 5, é a parte que realiza a predição dos dados a partir do bloco “Tree” ligado ao bloco “Prediction”. Com a predição realizada por estes, pode-se gerar os gráficos com os resultados logo na sequência através dos blocos “gráficos” e “Scatter plot”

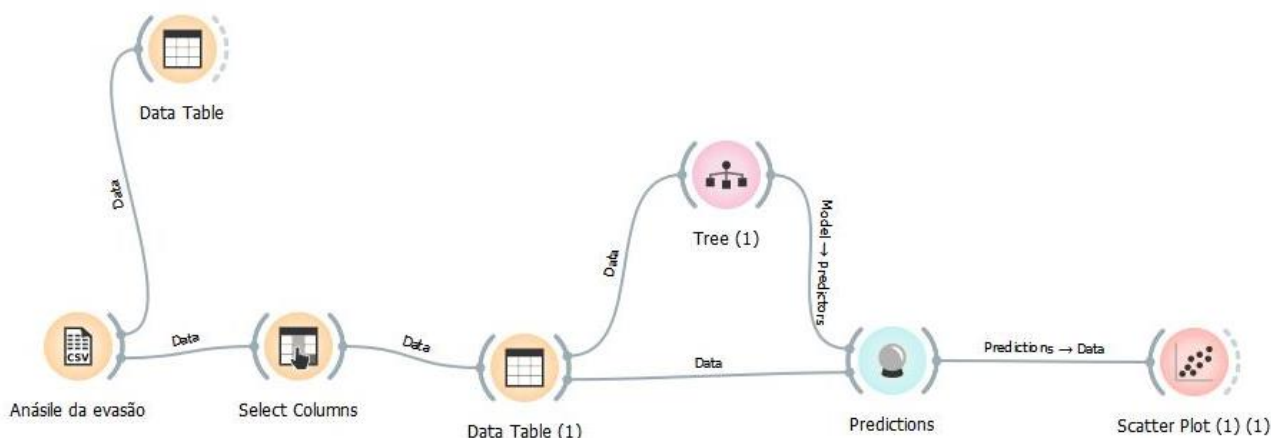
Figura 5 – Parte final da programação de predição da evasão



Fonte: Elaborado pelos autores, com base na pesquisa realizada.

Outro modelo realizado é o modelo de “Análise de evasão”, o qual é responsável por analisar os dados oferecidos para estudo, e a partir destes determinar quais são as variáveis mais impactantes para a evasão no curso, tanto para edificações como eletromecânica, que utiliza os blocos já mencionados conforme apresentado na Figura 6:

Figura 6 – Modelo de análise de evasão



Fonte: Elaborado pelos autores, com base na pesquisa realizada.

## **4 RESULTADOS E DISCUSSÕES**

Com base no que foi programado, notou-se que o modelo gerou uma resposta satisfatória e funcional. O programa acompanhou a crescente conclusão e entrada de estudantes, gerando a tendência de permanência no ensino. O programa gerou uma confiabilidade de 88% com base nas informações repassadas, sendo que desses 12% de erro, são provenientes de incoerências nos dados fornecidos para a elaboração do gráfico.

Outro resultado obtido foi o de predição da evasão. Como citado na metodologia, os dados analisados são apenas de alunos evadidos, logo a previsão deveria apresentar resultados próximos a dizer que o aluno irá evadir. As chances de os alunos evadirem variaram de 80% a 100%, com casos pontuais que passaram por 50%. No entanto, essas diferenças ocorrem devido às informações da evasão passadas e, em parte, pela falta de precisão do programa.

Na Tailândia, foram usadas técnicas de data mining para prever evasão e determinar os fatores causadores da mesma. Foram usadas árvores de decisão e modelos de indução de regras. Os resultados indicam uma acurácia promissora, de 75% a 81% (Meedech; lam-On; Boongoen, 2015). Um estudo realizado na Colômbia, para detecção de evasão de estudantes de graduação em Engenharia de Sistemas após sete anos de matrícula, identificou que algoritmos simples alcançam níveis confiáveis de precisão para identificar preditores de evasão (94%). Os resultados de Árvores de Decisão, Regressão Logística e Naive Bayes foram comparados para propor a melhor opção (Pérez; Castellanos; Correal, 2018).

Os Institutos Federais de Educação, Ciência e Tecnologia são instituições que oferecem educação técnica e superior. Portanto, uma análise de evasão escolar pode revelar muita informação sobre o que está acontecendo. Igualmente, o uso de ferramentas tecnológicas para predição da evasão permite identificar variáveis a serem tratadas mediante programas que evitem essa situação, contribuindo assim para a formação de jovens, tanto de baixa como de alta renda.



Na Índia, um estudo realizado em um curso de engenharia determinou que as variáveis que melhor explicam a evasão de um aluno são os fatores socioeconômicos e a pontuação de renda fornecida pelo *University Admission Test* (UAT). De acordo com a técnica da árvore de decisão utilizada, concluiu-se que a retenção é de 78,3%. A qualidade dos classificadores permite garantir que suas previsões estejam corretas, com níveis estatísticos da curva ROC de 76%, 75% e 83% de sucesso para classificadores de rede Bayesiana e árvore de decisão (Viloria *et al.*, 2019).

Para muitos autores, os fatores pessoais são as principais causas de desistência dos alunos da universidade. A idade e o sexo são os fatores de previsão mais utilizados, pois são considerados fatores internos de variabilidade que são simples de definir e medir.

Fatores acadêmicos: Referem-se ao desenvolvimento dos alunos em seu processo formativo. Identificamos 40 fatores acadêmicos, que correspondem a 36% do total de fatores identificados. A análise desses fatores mostra que o teste de admissão é o fator mais utilizado na literatura. No entanto, vale ressaltar que o processo de aprendizagem na universidade tem uma estreita relação com os níveis de estudo anteriores, impactando mais as realizações educacionais. Da mesma forma, a pontuação que um aluno obtém no vestibular é considerada um indicador para explicar o sucesso ou o fracasso na trajetória na universidade. Nesse sentido, muitos estudos têm analisado a validade preditiva desse fator, considerando-o um preditor de características cognitivas e atitudinais que são de extrema importância para que os alunos tenham sucesso na universidade.

Fatores econômicos: Estão relacionados à capacidade dos alunos para satisfazer as exigências econômicas que se apresentam durante o programa acadêmico. Nesta dimensão, foram identificados 15 fatores que afetam a evasão, e eles correspondem a aproximadamente 13% do total de fatores analisados. Esses fatores referem-se ao conforto material e à capacidade dos pais de alocar mais e melhores recursos para o desempenho acadêmico de seus filhos, o que tem um impacto significativo no desempenho acadêmico.

Fatores sociais: São aspectos que afetam os alunos como um todo e que são determinados pelo seu lugar e espaço. Por outro lado, a dimensão social concentra-se na

importância da interação entre alunos e seu ambiente social; interação em relação à instituição, normas acadêmicas e hábitos de estudo.

Fatores institucionais: Os fatores que correspondem a esta categoria dizem respeito às características estruturais e funcionais de uma instituição; estes representam aproximadamente 3,53% do total de fatores analisados.

## **5 CONSIDERAÇÕES FINAIS**

A evasão é um problema presente em todas as escolas do mundo, e que em algumas regiões e países gera grandes problemas, tendo consequências no setor econômico e social, e com base nessa problemática foi desenvolvido um modelo programado em plataforma virtual para sanar as evasões ocorridas, visando a eficiência da análise de dados.

Por tanto o modelo desenvolvido alcançou uma boa taxa de acertos de evasões, apresentando uma pequena imprecisão devido á imprecisão dos dados fornecidos, os quais são identificados pelo programa, sendo possível saber a origem do erro. O modelo pode ser constantemente aprimorado, com isso futuramente pode haver mudanças que visem a eficácia do sistema, seja por utilização de novas variáveis, utilização de novos blocos, podendo também ser utilizados de outros sistemas com a mesma funcionalidade e a melhora de registros educacionais da instituição, sendo possível aumentar a eficácia do programa para mais de 95%.

No caso específico dos institutos federais (IFs) o ensino técnico é uma importante oferta de educação no Brasil, que tem como objetivo formar profissionais técnicos qualificados e preparados para o mercado de trabalho, contribuindo para o desenvolvimento econômico e social do país.

## REFERÊNCIAS

- AGRUSTI, F.; BONA VOLONTÀ, G. University dropout prediction through educational data mining techniques: a systematic review. **Journal of e-Learning and Knowledge Society**, v. 15, n. 3, p. 45-60, 2019. Disponível em: <https://doi.org/10.20368/1971-8829/1135017>. Acesso em: 1 nov. 2024.
- ANPEC - Associação Nacional dos Centros de Pós-Graduação em Economia. **Retenção e evasão no ensino superior brasileiro**: uma análise dos efeitos da bolsa permanência do PNAES. In: ENCONTRO DA ANPEC, 14., 2018, São Paulo. Anais [...]. São Paulo: Anpec, 2018. Disponível em: <https://link.ufms.br/osNnk>. Acesso em: 20 ago. 2022.
- ARARUNA, L. T.; SANTOS, T. A.; FIAMENGUI, A. H. R.; DECANINI, J. G. M. S. Sistema inteligente para identificação de grupos de risco relacionados à evasão escolar. **Sinergia** São Paulo, v. 16, n. 4, p. 263-268, 2015. Disponível em: <https://link.ufms.br/ZBNrk>. Acesso em: 1 nov. 2024.
- ASHA, P. et al. Predicting university dropout through data analysis. In: INTERNATIONAL CONFERENCE ON TRENDS IN ELECTRONICS AND INFORMATICS, 3., 2020, Tamil Nadu. **Anais** [...]. Tamil Nadu: ICOEI, 2020. Disponível em: <https://link.ufms.br/hbX70>. Acesso em: 1 nov. 2024.
- CARRANO, D.; ALBERGARIA, E. T.; INFANTE, C.; ROCHA, L. Combinando Técnicas de Mineração de Dados para Melhorar a Detecção de Indicadores de Evasão Universitária. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 1., 2019, Rio de Janeiro. **Anais** [...]. Rio de Janeiro: SBIE, 2019. Disponível em: <https://link.ufms.br/gk75P>. Acesso em: 1 maio 2022.
- COSTA, J. B.; BORGES, A. P. F. Evasão escolar no curso de Eletrotécnica do Ensino Médio Integrado do IF Sertão Pernambucano. **Semiário de Visu**, Petrolina, v. 3, n. 7, p. 366-370, 2019. Disponível em: <https://link.ufms.br/jjbkt>. Acesso em: 15 set. 2022.
- ERNÁNDEZ-GARCÍA, A. J. et al. A real-life machine learning experience for predicting university dropout at different stages using academic data. **IEEE Access**, v. 9, p. 133076-133090, 2021. Disponível em: <https://doi.org/10.1109/ACCESS.2021.3115851>. Acesso em: 1 nov. 2024.
- HENRIQUES, F. G. Evolução histórica, expansão e financiamento da educação superior no Brasil: das primeiras faculdades à expansão do início do século XXI. **Comunicologia**, Brasília, v. 11, n. 2, p. 124-145, 2018. Disponível em: <https://doi.org/10.31501/comunicologia.v11i2.10058>. Acesso em: 1 mar. 2022.
- JORNAL NACIONAL. IBGE mede o problema nacional da evasão escolar. **Jornal Nacional**, São Paulo, 2020. Disponível em: <https://link.ufms.br/Gq5W6>. Acesso em: 30 jul. 2022.

MARIA, W.; DAMIANI, J. L.; PEREIRA, M. Rede Bayesiana para previsão de evasão escolar. *In: V CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*, 5., 2016, Uberlândia. **Anais [...]**. Uberlândia: CBIE, 2016. Disponível em: <https://link.ufms.br/xBJR8>. Acesso em: 2 jul. 2022.

MEEDECH, P.; IAM-ON, N.; BOONGOEN, T. Prediction of student dropout using personal profile and data mining approach. *In: LAVANGNANANDA, K., PHON-AMNUAISUK, S., ENGCHUAN, W., CHAN, J. (ed.). Intelligent and Evolutionary Systems. Proceedings in Adaptation, Learning and Optimization*. Springer, Cham: 2015. v. 5. Disponível em: [https://doi.org/10.1007/978-3-319-27000-5\\_12](https://doi.org/10.1007/978-3-319-27000-5_12). Acesso em: 12 set. 2022.

MINGHELLI, M. Institutos Federais de Educação, Ciência e Tecnologia: um futuro incerto. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 23, n. 51, p. 157-165, 2018. Disponível em: <https://doi.org/10.5007/1518-2924.2018v23n51p157>. Acesso em: 1 nov. 2024.

MINISTÉRIO DA TRANSPARÊNCIA E CONTROLADORIA-GERAL DA UNIÃO. **Relatório de gestão exercício 2017**. 2018. Disponível em: <https://link.ufms.br/zil3h>. Acesso em: 15 set. 2022.

ORTIZ-LOZANO, J. M. *et al.* University student retention: best time and data to identify undergraduate students at risk of dropout. **Innovations in Education and Teaching International**, v. 57, p. 74-85, 2020. Disponível em: <https://doi.org/10.1080/14703297.2018.1502090>. Acesso em: 1 nov. 2024.

PÉREZ, B; CASTELLANOS, C; CORREAL, D. Applying data mining techniques to predict student dropout: a case study. *In: 1st COLOMBIAN CONFERENCE ON APPLICATIONS IN COMPUTATIONAL INTELLIGENCE*, 15., 2018, São Paulo. **Proceedings [...]**. Bogotá: IEEE, 2018. Disponível em: <https://doi.org/10.1109/ColCACI.2018.8484847>. Acesso em: 5 jul. 2022.

SILVA FILHO, R. B.; ARAÚJO, R. M. L. Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. **Educação Por Escrito**, [S. l.], v. 8, n. 1, p. 35, 2017. Disponível em: <https://link.ufms.br/WpSem>. Acesso em: 15 set. 2022.

SOSU, E. M.; PHEUNPHA, P. Trajectory of university dropout: investigating the cumulative effect of academic vulnerability and proximity to family support. **Front. Educ.**, v. 12, p. 89-104, 2019. Disponível em: <https://link.ufms.br/YbrX9>. Acesso em: 1 nov. 2024.

UNIVERSIDADE DE LJUBLJANA. **Orange**. 2022. Disponível em: <https://link.ufms.br/YklrB>. Acesso em: 1 nov. 2024.

VILORIA, A. *et al.* Integration of data technology for analyzing university dropout. **Procedia Computer Science**, v. 155, p. 569-574, 2019. Disponível em: <https://doi.org/10.1016/j.procs.2019.08.079>. Acesso em: 1 nov. 2024.

## Sobre os autores

### Matheus Fogaça Bicudo de Almeida

Possui experiência na área de Engenharia Civil através da formação técnica em edificações participando em congressos de iniciação científica. Possui experiência na Engenharia Mecânica e elétrica pela formação técnica em eletromecânica. cursando Bacharelado no ensino superior em Engenharia Mecânica no Instituto Federal de Ciências e Tecnologia de São Paulo - Câmpus Itapetininga. Possui participação como bolsista no projeto de extensão "Cursinho Popular" do IFSP-ITP na área de química.

E-mail: [matheusfogaça001@gmail.com](mailto:matheusfogaça001@gmail.com)

### Adriana Marques

Pós-doutorado no Instituto de Pesquisas Energéticas e Nucleares-IPEN (em andamento) sob supervisão da Professora Doutora Isolda Costa no tema de microplásticos. Pós-doutorado na Universidade Federal de São Paulo-Unifesp (em andamento) sob supervisão do Professor Doutor Rubens Lacerda de Sá no título "Operação acolhida, interiorização e hos(ti)pitalidade migrante". Doutora em Engenharia Mecânica pela Universidade de Brasília (UnB). Mestre em Engenharia Civil pela Universidade Estadual de Campinas (UNICAMP), graduação em Ciências Econômicas e Ciências Contábeis. Atualmente é Coordenadora de Estudos Econômicos da Superintendência de Estudos Hídricos e Econômicos - SHE da ANA. Foi Coordenadora do gabinete da Presidente da ANA, Secretária Executiva adjunta no Ministério da Igualdade Racial. Coordenadora Geral CGDEA – Enap. É professora do Instituto Federal de São Paulo (IFSP) campus SPO, membro do núcleo de estudos afro-brasileiros e indígenas - NEABI. Desde 2003 é fellow do Programa Lead Internacional, organização ligada ao desenvolvimento de lideranças na temática da sustentabilidade. Possui experiência como avaliadora de cursos de ensino superior e técnicos pelo INEP/MEC. Líder do Grupo de Pesquisa de Energia, Água e Saneamento (ENASA/CNPq). Editora-chefe da Revista Sinergia do IFSP. Atua na área ambiental, econômica e social em temas como comunidades tradicionais, recursos hídricos, planejamento energético e algoritmos racistas. Participou de projetos e treinamento pelas Nações Unidas (UN), Fundação Rockefeller e Fundação Kellog no Japão, México, Inglaterra e Itália. Organizadora e autora do livro "Arcabouço Pedagógico Para o Estudo Étnico-Racial e Ambiental" e do livro "Isotermas de Adsorção: Teoria e Prática".

E-mail: [adrimarks@ifsp.edu.br](mailto:adrimarks@ifsp.edu.br)

Submetido em 13 de fevereiro de 2023.

Aceito para publicação em 10 de outubro de 2024.

## Licença de acesso livre



A **Revista Edutec** utiliza a [Licença Creative Commons - Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/), pois acredita na importância do movimento do acesso aberto nos periódicos científicos.