

UMA PROPOSTA DE SISTEMA WEB PARA AVALIAR OS RISCOS DA EVASÃO ESCOLAR NO IFCE

A PROPOSAL OF WEB SYSTEM TO EVALUATE THE RISK OF DROPOUT OF STUDENTS AT IFCE

Igor Pereira de Lima*.
Daniel Victor Saraiva**.
Carina Teixeira de Oliveira***.

RESUMO

A evasão escolar é um dos diversos problemas enfrentados pela educação brasileira. Essa problemática é responsável por causar prejuízos econômicos e sociais ao país. Sendo assim, este artigo apresenta uma proposta de um sistema web para auxiliar gestores e professores na tomada de decisão com intuito de diminuir os números da evasão escolar no Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). Para a construção do sistema, são utilizados resultados de estudos baseados em técnicas de Aprendizado de Máquina (AM) e mineração de dados educacionais (*Educational Data Mining* – EDM). O sistema permite que o usuário final tenha acesso, através de uma interface amigável, às probabilidades que determinado estudante possui de obter *êxito* ou *evadir* do curso matriculado. Através dessa solução computacional, é possível que a gestão e os professores desenvolvam estratégias para a permanência dos estudantes.

Palavras-chave: Aprendizado de Máquina. Mineração de Dados Educacionais. Sistema Web. Evasão Escolar.

ABSTRACT

School dropout is one of the many problems faced by the Brazilian education. This problem is responsible for causing economic and social losses to the country. Thus, this paper presents a proposal of a web system to assist managers and teachers in decision making in order to reduce the numbers of school dropouts in the Federal Institute of Education, Science and Technology of Ceará (IFCE). To develop the system, there were used results of studies based on Machine Learning (ML) and Educational Data Mining (EDM) techniques. The system allows the end user to have access, through a friendly interface, to the probabilities that a given student has to

* Graduando em Bacharelado em Ciência da Computação no Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). E-mail: igor.pereira.lima02@aluno.ifce.edu.br

** Mestre em Ciência da Computação, Pesquisador do Laboratório de Redes de Computadores e Sistemas (LAR/IFCE). E-mail: victordvs@hotmail.com

*** Doutora em Computação, Docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Pesquisadora do Laboratório de Redes de Computadores e Sistemas (LAR/IFCE). E-mail: carina@lar.ifce.edu.br

succeed or drop out of the enrolled course. Through this computational solution, it is possible for management and teachers to develop strategies for the permanence of students.

Keywords: *Machine Learning. Educational Data Mining. Web System. School Dropout.*

1 INTRODUÇÃO

No Brasil, o direito à educação é garantido e amparado pela Constituição Brasileira de 1998 (BRASIL, 1988). Sendo assim, a educação no país se instituiu como um dever do Estado e, com o auxílio da família, tem o objetivo de garantir o desenvolvimento do indivíduo, no suporte ao exercício da cidadania. Porém, a evasão escolar surge no país como um dos desafios que historicamente se faz presente em discussões entre educadores, pesquisadores e todo sistema educacional em busca de minimizar essa problemática (ALMEIDA; OLIVEIRA, 2020).

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), 11,8% dos jovens na faixa etária entre 15 e 17 anos de idade se encontravam fora da escola (IBGE, 2019). Essa porcentagem representa 1,2 milhões de jovens. Já o levantamento do módulo de educação da Pesquisa Nacional por Amostra de Domicílios (PNAD), no ano de 2020, apresenta que o número de jovens entre 14 e 29 anos que deixaram de frequentar a escola chegou a 20,2% (PNAD, 2020). Esse número representa cerca de 10,1 milhões de jovens que desistiram do ambiente escolar. Os dados do PNAD ainda concluíram que a maioria desses jovens são do sexo masculino. Segundo o levantamento, a necessidade de trabalhar foi um dos motivos desse elevado índice de desistência (PNAD, 2020).

Muito se tem discutido sobre os motivos que levam um estudante a abandonar a sua carreira acadêmica. Dentro dessas discussões, é levado em consideração a importância da família e da instituição de ensino na vida acadêmica de um jovem (FERREIRA et al., 2021). Segundo Almeida e Oliveira (2020), alguns fatores podem contribuir para o aumento da evasão escolar no país. Entre esses fatores, além da necessidade de procurar emprego, estão as dificuldades de aprendizado, falta de estrutura familiar e até mesmo gravidez na adolescência.

Diante desse cenário, a evasão escolar ocupa uma grande relevância perante as políticas públicas de educação elaboradas pelos governantes, pois, as consequências causadas por esse altos índices de evasão impactam no desenvolvimento da sociedade (QUEIROZ, 2006). Além disso, cada jovem que desiste de frequentar a escola sofre as consequências de terem as suas capacidades intelectuais e cognitivas afetadas e ainda enfrenta diversos desafios para se inserir no mercado de trabalho (ALVES; LAVOR; PEREIRA, 2017).

Nas instituições de ensino, uma dessas consequências é o cálculo negativo do rendimento escolar. Esse cálculo é fundamental para gerar o Índice de Desenvolvimento da Educação Básica (IDEB), que possui o objetivo de medir a qualidade de ensino no Brasil (IDEB, 2021; INEP, 2021).

Perante essas circunstâncias, é de total interesse do Estado encontrar políticas públicas que contribuam para a permanência e o êxito do estudante no âmbito escolar. Por exemplo,

uma das áreas da pesquisa que vem sendo utilizada nos estudos em busca de encontrar as justificativas para a problemática da evasão escolar é a Mineração de Dados Educacionais (do inglês, *Educational Data Mining*, ou EDM) (RODRIGUES, 2016). Pode-se destacar que a EDM possui técnicas que auxiliam na análise de um conjunto de dados, podendo traçar, por exemplo, perfis de estudantes que possuem um risco maior de evasão (SANTOS; SARAIVA; OLIVEIRA, 2021). Com esses perfis é possível estimar diversos fatores que podem contribuir para descobrir os motivos que levam um estudante a deixar de frequentar a instituição de ensino. dentre eles, podem ser citados problemas demográficos, socioeconômicos, culturais, familiar, educacional, etc (RODRIGUES, 2016). Com o uso da EDM, é possível que fatores sejam apontados sobre o comportamento dos estudantes, fazendo que professores e gestores encontrem soluções pedagógicas que possam influenciar na permanência dos estudantes no âmbito escolar.

A atuação da EDM é fundamental para analisar e entender comportamentos que contribuam para a evasão escolar, além de possibilitar alguma intervenção apropriada (SILVA; CARVALHO; MACIEL, 2021). Com o uso de bibliotecas/algoritmos de aprendizado de máquina (em inglês, *Machine Learning*), aplicados a bases de dados com registros educacionais, é possível desenvolver mecanismos que auxiliem na tomada de decisão. E assim auxiliar professores e gestores a melhorar as suas metodologias de aprendizagem e incentivando aos estudantes a permanência na instituição.

O Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) é uma instituição que tem por sua especialidade a oferta de educação tecnológica e profissional, com diferentes modalidades de ensino. Entre elas, estão o ensino médio, técnico, graduação e pós-graduação, além de contar com projetos de Pesquisa e Extensão (IFCE, 2022).

Nos últimos anos, a instituição apresentou números bastante expressivos em relação aos índices de evasão. Por exemplo, no curso Técnico de Informática do IFCE - Campus Aracati, desde o período 2013/1 até o período 2022/2 foram realizadas 580 matrículas, entre as quais, 45,52% (264) correspondem a discentes evadidos e 33,10% (192) são egressos com êxito (IFCE, 2022). Para completar o levantamento, é importante destacar que 12,24% (71) são discentes que estão mantendo as matrículas ativas e 9,14% (53) correspondem a matrículas que estão suspensas temporariamente. Portanto, diante dos dados apresentados, destaca-se que o número de discentes evadidos no curso Técnico de Informática do IFCE-Campus Aracati é maior que o número de estudantes concludentes (IFCE, 2022).

Logo, este trabalho tem objetivo implementar um sistema web que permite avaliar riscos de evasão de estudantes, e assim auxiliar professores e gestores do IFCE na redução dos índices da evasão escolar da instituição. Para tanto, alguns objetivos específicos foram estabelecidos para chegar ao produto final. O primeiro desses objetivos é utilizar técnicas de *Machine Learning* que possuem um bom desempenho nas métricas de validação para fazer a previsão de estudantes com risco de desistência acadêmica. O segundo é construir uma ferramenta computacional capaz de utilizar o modelo preditivo criado para exibir os resultados alcançados para os gestores e professores do IFCE. E, por último, contribuir na redução dos índices de evasão em visão micro (IFCE - Campus Aracati) e macro (no *campi* do IFCE).

A metodologia para a realização deste trabalho foi dividida em seis etapas importantes. Ao final dessas etapas, a contribuição acadêmica deste trabalho é um sistema web para avaliar os riscos de evasão de estudantes do IFCE. Essa metodologia foi elaborada para que fosse construído primeiramente um *back-end* para manipular os dados que foram utilizados para compor o sistema, e ainda um *front-end* dinâmico e objetivo para que as informações sejam visualizadas por um usuário final.

2 TRABALHOS RELACIONADOS

Esta seção apresenta trabalhos que realizaram uma pesquisa com foco na previsão de risco de estudantes utilizando Mineração de Dados Educacionais e algoritmos de aprendizagem de máquina. Os trabalhos a seguir foram divididos em soluções para o cenário nacional e também no cenário internacional, de maneira cronológica. Ao final, são realizadas comparações entre os trabalhos pesquisados e a proposta apresentada nesse trabalho.

2.1 Cenário Nacional

Schreiber et al. (2017) apresentam a criação de um *software* chamado *SDBayes: a System to Predict Student Drop-out*, que tem o objetivo de analisar uma base de dados de uma Instituição de Ensino Superior (IES) e assim conseguir analisar e prever as possibilidades que o discente tem de evasão e de permanência no curso em que está realizando a graduação. O sistema construído usa redes *bayesianas*, que é uma das técnicas de aprendizagem de máquina para realizar a previsão dos dados que são utilizados durante a pesquisa. Para a construção do modelo preditivo é realizado primeiramente uma entrevista com os gestores e coordenadores de cada curso superior da instituição. Após esse levantamento de informações, chegou-se à conclusão que cada curso tem um perfil específico de discentes, portanto, um aluno de determinado curso tem uma probabilidade diferente de evadir. Como resultado, o *software* emite para a gestão da instituição a probabilidade de evasão e também de permanência dos estudantes pertencentes aos cursos de Ciência da Computação, Administração e Engenharia de Produção.

No estudo realizado por Silva (2019), o autor mostra o resultado de uma proposta de desenvolvimento de uma ferramenta que permite acompanhar os estudantes. Com foco principal nos atributos socioeconômicos e acadêmicos relevantes para detectar a possibilidade de evasão. A ferramenta foi desenvolvida no Instituto Federal do Rio Grande do Norte (IFRN), que durante o período de 2015 a 2017 apresentou taxas de evasão de 13,4% em cursos técnicos (Técnico Integrado em Edificações, Informática, Logística e Técnico Subsequente em Edificações) e 31,5% em cursos superiores (Tecnologia em Logística e Redes de Computadores). Diante disso, o trabalho tem como objetivo propor um modelo descritivo para auxiliar a gestão do IFRN na tomada de decisão com intuito de reduzir os números da evasão. A base de dados utilizada no trabalho foi dos alunos do curso de nível médio Técnico Integrado de Informática (TII) e do curso de nível superior de Tecnologia em Redes de Computadores (TRC). As técnicas de mineração de dados foram utilizadas para a construção do modelo descritivo. Os algoritmos

foram capazes descobrir padrões e fazer a previsão da evasão. O algoritmo de *Decision Tree* alcançou os índices de acurácia de 76% na base de dados do curso de TII e 70% de acurácia no curso de TRC. Por fim, o modelo foi integrado na construção da ferramenta (sistema web), proporcionando uma visualização analítica (gráficos) de cada aluno cadastrado.

Em Beltran et al. (2019), é descrita a criação de uma plataforma de aprendizado de máquina para detecção e monitoramento de alunos do ensino superior com risco de evasão. Para a construção dessa plataforma, os autores elaboraram, primeiramente, um modelo preditivo utilizando aprendizagem de máquina. Esse modelo foi planejado a partir do particionamento da base de dados utilizada no trabalho com o auxílio do algoritmo *K-means*. Com a base de dados particionada, os seguintes algoritmos de aprendizagem de máquina foram aplicados: *AdaBoost*, *Bagging*, *Decision Tree* e *Naive Bayes*. O sistema criado para utilizar o modelo de predição criado se caracteriza como um sistema educacional que possui uma função principal, que é listar para o usuário final a lista de estudantes classificados em ‘situação normal’ e ‘risco de evasão’.

2.2 Cenário Internacional

No trabalho de Hilmarsson (2019), o autor apresenta um estudo que possui o objetivo de melhorar a retenção de estudantes nas escolas secundárias da Irlanda. Diante disso, o trabalho propõe a criação de um sistema de gestão administrativa e de aprendizagem para ser utilizado por toda rede de escolas secundárias. Com dados fornecidos pelo Ministério da Educação, Ciência e Cultura da Irlanda, os algoritmos de aprendizagem de máquina foram utilizados para prever o risco de evasão dos alunos. Após a avaliação dos algoritmos, os classificadores com os melhores desempenhos foram *Gradient Boosting*, *Random Forest* e *AdaBoost*. O classificador *Gradient Boosting* apresentou uma precisão de 84%, se destacando dos demais algoritmos. Por fim, o modelo preditivo foi integrado a uma interface final (sistema web), na qual o usuário pode visualizar os dados e a previsão de evasão dos alunos através de um painel de controle.

Já no trabalho realizado por Guzmán-Castillo et al. (2022), os autores mostram a criação de um sistema que emprega aprendizagem de máquina, que permite calcular o risco de desistência de universitários de uma determinada Instituição de Ensino Superior. Com mais de 15.000 estudantes, os dados utilizados para a construção da amostra foram coletados, consolidados e complementados com informações disponibilizadas através do Ministério Colombiano de Educação Nacional e da Direção Colombiana de Desenvolvimento Social. Foram considerados os seguintes algoritmos de aprendizado de máquina: *Bayesian GLM*, *Decision Trees*, *Logitboost algorithms*, *Random Forest* e *Stochastic Gradient Boosting (SGB)*. Os algoritmos selecionados apresentaram uma grande eficiência para realizar a previsão da evasão na instituição.

2.3 Comparativo dos Trabalhos Relacionados

A Tabela 1 apresenta um comparativo dos trabalhos relacionados, assim como mostra as características do presente trabalho.

Tabela 1 – Comparativo dos Trabalhos Relacionados.

Trabalho	Algoritmo(s) de Classificação	Melhor Desempenho	Solução Computacional
SCHREIBER et al., 2017	<i>Redes Bayesianas</i>	<i>Redes Bayesianas</i> (82,01%)	<i>Software</i> chamado de 'SDBayes'
SILVA., 2019	<i>Decision Tree</i>	<i>Decision Tree</i> (76%)	Sistema <i>web</i> chamado de 'Lince'
BELTRAN et al., 2019	<i>AdaBoost, Bagging, Decision Tree, Naive Bayes, KNN, Neural Networks</i>	<i>Neural Networks</i> (97%)	Plataforma <i>web</i>
HILMARSSON., 2019	<i>Gradient Boosting, Random Forest e AdaBoost</i>	<i>Gradient Boosting</i> (86%)	Interface <i>web</i>
GUZMÁN - CASTILHO et al., 2022	<i>AdaBoost, Bayesian GLM, Decision Tree, Logiboot algorithms, Random Forest, SGB</i>	<i>Bayesian GLM e AdaBoost</i> (80%)	Sistema <i>web</i>

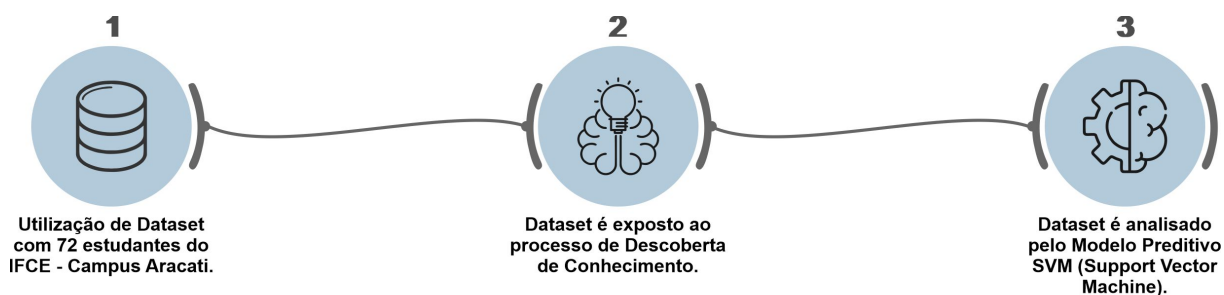
Fonte: Elaborada pelos autores (2022).

O presente trabalho possui um grande diferencial em relação os trabalhos analisados, que é descrever a criação de um *sistema web* que tem o objetivo de mostrar ao usuário final (nesse cenário a gestão do IFCE) as chances que os estudantes matriculados nos cursos oferecidos pelo IFCE possuem de obter 'sucesso' ou 'risco de evasão' através das probabilidades. Por exemplo, ao analisar o estudante é possível extrair a porcentagem que ele possui de evadir do curso, caso essa porcentagem esteja alta a gestão poderá intervir e ajudá-lo a permanecer na instituição.

3 PROPOSTA

A metodologia utilizada para construção do sistema *web* para avaliar riscos de evasão de estudantes do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) segue uma sequência lógica composta por seis etapas principais (Figura 1). As três primeiras etapas da metodologia foram baseadas no processo de descoberta de conhecimento em Bases de Dados utilizado no trabalho de Saraiva (2020). O trabalho citado consiste em apresentar soluções baseadas em Mineração de Dados para avaliar os riscos de evasão de estudantes. Já na Etapa 4 da metodologia iniciou o desenvolvimento da proposta deste trabalho, que teve continuidade nas Etapas 5 e 6. O objetivo geral do sistema é poder informar através de uma interface amigável se o estudante possui chances de "sucesso" ou "risco de evasão" no curso em que está matriculado, por meio de suas respectivas porcentagens. Em outras palavras, a porcentagem que determinado aluno tem de obter sucesso no curso e a porcentagem que o mesmo possui de risco de evasão.

Figura 1 – Etapas da Metodologia.



As etapas 1, 2 e 3 foram realizadas a partir da metodologia do trabalho Saraiva (2020).



As etapas 4, 5 e 6 foram realizadas para o desenvolvimento da proposta deste trabalho.

Fonte: Elaborada pelos autores (2022).

3.1 Etapa 01: Utilização de *Dataset* com estudantes do IFCE - Campus Aracati.

Na primeira etapa da metodologia, foram coletados dados de 72 estudantes do IFCE - Campus Aracati matriculados no curso Técnico em Informática durante o período de 2015/1 a 2018/1. Por ser o primeiro passo para a construção da solução computacional, essa etapa se destacou por ser o momento onde se requer um entendimento dos dados utilizados, além de fazer a relação entre a problemática abordada (a evasão escolar no IFCE) e os dados coletados, e assim obter uma melhor compreensão das etapas seguintes na construção da proposta.

3.2 Etapa 2: *Dataset* é exposto ao processo de Descoberta de Conhecimento.

A segunda etapa da metodologia consistiu em desenvolver processos de identificação de padrões no *dataset* citado na etapa anterior. Para isso, foi aplicado todo o entendimento de dados já adquirido no início do processo e, assim, estabelecida a finalidade de cada tipo de dado explorado. Os dados coletados possuíam finalidades socioeconômicas (renda per capita e escolaridade dos pais), dados acadêmicos (disciplinas, notas, faltas, etc), dados demográficos (sexo, etnia, endereço, etc) e, por fim, dados gerais (forma de ingresso, tipo de escola de origem, etc).

Em seguida, foi necessário preparar os dados, já que uma das adversidades mais frequentes encontradas em uma base de dados é a existência de dados inconsistentes, discrepantes, redundantes e ausentes. É necessário que se identifique essas adversidades antecipadamente,

além de realizar o tratamento de dados adequado. Nesse estudo, foram utilizadas técnicas de pré-processamento de dados para melhorar e manter a integridade e a qualidade dos dados (FACELI et al., 2011; CASTRO; FERRARI, 2016).

- **Eliminação de dados:** para o conjunto de dados utilizado nesta pesquisa, é preciso destacar que alguns atributos não são necessários para a construção do modelo preditivo, além de esta mesma base de dados possuir atributos com muitos valores ausentes. Portanto, é necessário também a eliminação de alguns dados.
- **Limpeza dos dados:** a limpeza de dados atua na remoção de dados que estão inconsistentes ou estão fora do padrão. Também se utiliza de técnicas para tratar dados que estão faltando. Para isso, é realizado o procedimento da troca entre o dado que está em falta por uma medida de tendência (média, moda e mediana).

Ao final dessa etapa, a base de dados possuía um número reduzido de registros por conta de todo processo de preparação de dados. A Tabela 2 apresenta os atributos selecionados para a construção do modelo preditivo. Esses atributos representam aspectos socioeconômicos e desempenho acadêmico dos estudantes que contribuem para a construção do modelo preditivo.

Tabela 2 – Seleção de Atributos.

Nome dos Atributos					
1. Período Letivo	2. Idade	3. Sexo	4. Renda Familiar	5. Etnia	6. Estado Civil
7. Grau de Instrução	8. Estado Civil dos Pais	9. Grau de Instrução dos Pais	10. Turno	11. Forma de Ingresso	12. Coeficiente de Rendimento
13. Total Aprovações	14. Total Aprovações Parcialmente	15. Total Aprovações com Dependência	16. Total Reprovações	17. Total Trancados	18. Situação Matrícula

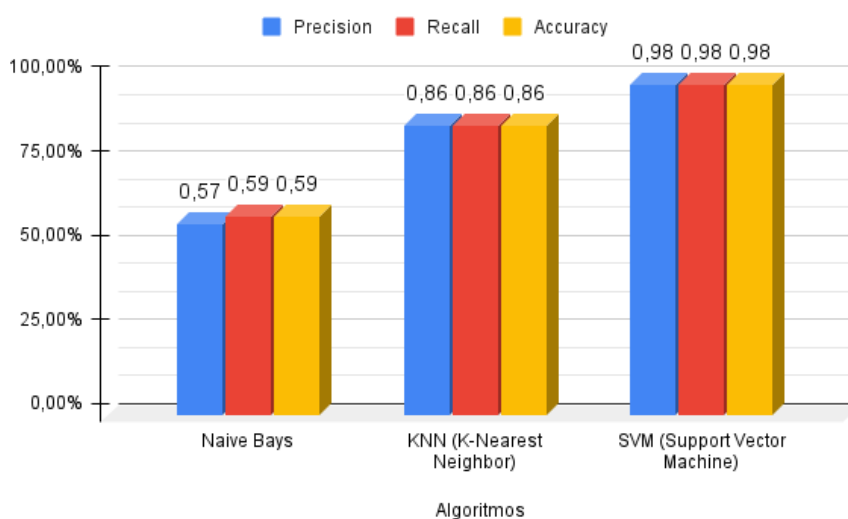
Fonte: Elaborada pelos autores (2022).

Com os dados preparados e os atributos selecionados, surgiu o momento de aplicar o processo de **Mineração de Dados**. Nesse momento, a base de dados dos estudantes do curso Técnico em Informática do IFCE foi submetida ao processo de mineração de dados educacionais para assim se extrair conhecimento. Para isso, se fez o uso dos algoritmos de Aprendizagem de Máquina (AM). A seleção desses algoritmos foi feita a partir de uma revisão no trabalho dos autores Santos, Saraiva e Oliveira (2021), onde é realizado um estudo que mostra os desempenhos dos principais modelos de predição (aprendizado supervisionado) que possuem o objetivo de classificar amostras de acordo com a classe determinada pelo supervisor. Depois de toda análise, os algoritmos escolhidos foram: probabilístico (*Naive Bayes*); baseado em distância (*N-Nearest Neighbor* - KNN) e baseados em otimização (*Support Vector Machine* - SVM).

Depois de aplicar os algoritmos selecionados, o objetivo foi fazer a comparação entre os resultados obtidos pelo modelos preditivos de aprendizado supervisionado que foram aplicados na base do Curso Técnico em Informática que foi preparada. Após, definiu-se qual classificador apresentou os melhores índices em classificar os estudantes em **egressos com êxito** (concluiu) e **egressos sem êxito** (evadiu).

Para validar o desempenho dos algoritmos aplicados, algumas métricas de validação foram usadas para analisar as atividades realizadas pelos algoritmos de aprendizagem de máquina propostos neste trabalho. A Figura 2 mostra através dos gráficos o desempenho de cada algoritmo utilizado na etapa de Mineração de Dados. Esses algoritmos foram avaliados seguindo suas métricas de validação, sendo elas *Precision*, *Recall*, *Accuracy*.

Figura 2 – Avaliação dos Algoritmos.



Fonte: Elaborada pelos autores (2022).

3.3 Etapa 3: Análise pelo algoritmo *Support Vector Machine* (SVM).

Após analisar os resultados dos algoritmos utilizados durante o processo de Mineração de Dados, concluiu-se que o algoritmo *Support Vector Machine* (SVM) possuía as melhores taxas de desempenho. Outro fator que reforçou a escolha deste algoritmo foi o armazenamento das porcentagem referentes às *chances de sucesso* de um discente e também ao seu *risco de evasão*. Para isso, foi necessário o uso das bibliotecas do *Python*¹ chamadas de *Pandas*² e *Scikit-learn*³. Essas porcentagens foram armazenadas em duas novas variáveis, que foram analisadas pelas classificações de Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN). Assim, determinado aluno pode ter uma porcentagem maior de *conseguir êxito* no curso em que está matriculado, conseqüentemente, ele tem uma porcentagem menor para *riscos de evasão*. Mas, pode acontecer

¹ <https://www.python.org/>

² <https://pandas.pydata.org/>

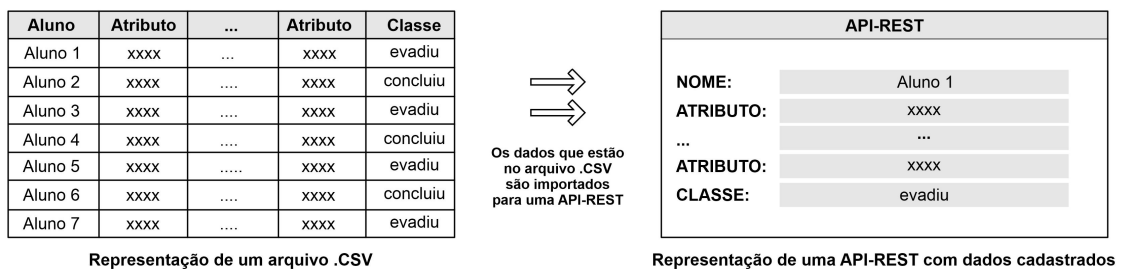
³ <https://scikit-learn.org/stable/index.html>

de o aluno possuir uma porcentagem maior para o *risco de evasão* e, conseqüentemente, suas chances de *conseguir êxito* serem menores.

3.4 Etapa 4: Resultados alcançados pelo algoritmo *Support Vector Machine* (SVM).

Nesta etapa iniciou-se a construção da **proposta de criação de um sistema web para avaliar os riscos de evasão no IFCE**. Para isso, os resultados obtidos pelo algoritmo *Support Vector Machine* (SVM) foram exportados para um arquivo *Comma Separated Values* (.CSV). Esse tipo de arquivo se destaca por conter número, letras e compor dados em forma de tabela. Esse processo de exportação também foi realizado com auxílio das bibliotecas do *Python*¹, o *Pandas*². Juntamente com os resultados obtidos, o arquivo .CSV também foi composto pelos atributos selecionados registrados na Tabela 2 desta metodologia. Assim, o arquivo final pode ser consumido por uma *Application Programming Interface* (API) e ser disponibilizado para um usuário final através de uma aplicação *front-end*. Na Figura 3, é possível visualizar como o arquivo *Comma Separated Values* (.CSV) é composto, sendo assim, é possível compreender como os dados registrados neste arquivo podem ser consumidos pela *Application Programming Interface* (API).

Figura 3 – Representação do uso do arquivo '.csv'.



Fonte: Elaborada pelos autores (2022).

3.5 Etapa 5: Construção de uma API-REST em Django.

Dando continuidade a construção da proposta deste trabalho, a partir dessa etapa, iniciou-se a construção da API-REST que irá compor o sistema web a ser desenvolvido.

Um sistema web possui algumas etapas de construção, uma delas é denominada *back-end*. Para essa aplicação, o *back-end* é responsável por consumir, manipular os dados registrados no arquivo .CSV adquirido na etapa anterior, além de garantir consistência e segurança durante a manipulação desses dados mencionados. Para construir o *back-end* foi utilizado o *framework* baseado em *Python* chamado *Django*. Ao final dessa aplicação, obteve-se um modelo de interface capaz de utilizar mecanismos do protocolo *Hypertext Transfer Protocol* (HTTP): *Get*, *Post*, *Put*, *Delete*. Além disso, permitir as realizações das operações *Create*, *Read*, *Update* e *Delete* (CRUD).

3.6 Etapa 6: Construção do *front-end* para exibir os resultados.

Com a API-REST construída, foi desenvolvido o *front-end* da aplicação e assim concluir o sistema web. Com o trabalho proposto, o gestor ou professor pode monitorar/acompanhar a situação de um aluno utilizando uma interface objetiva e dinâmica, e assim avaliar as porcentagens que esse discente possui de ter sucesso ou evadir do curso matriculado, além de visualizar os seus perfis socioeconômicos através de gráficos (*dashboards*). Desse modo, o sistema é capaz de auxiliar diretamente na melhora de rendimento de um possível aluno com risco alto de desistência.

A Figura 4 apresenta a tela principal do sistema, na qual são destacadas as principais funcionalidades. Cada número marcado na figura é explicado a seguir:

Figura 4 – Tela principal do *front-end* construído.

NOME:	IDADE:	PROB. ÊXITO	STATUS
ALUNO 1	18 anos	78.0	●
ALUNO 2	23 anos	98.0	●
ALUNO 3	20 anos	50.1	●
ALUNO 4	19 anos	40.0	●

Fonte: Elaborada pelos autores (2022).

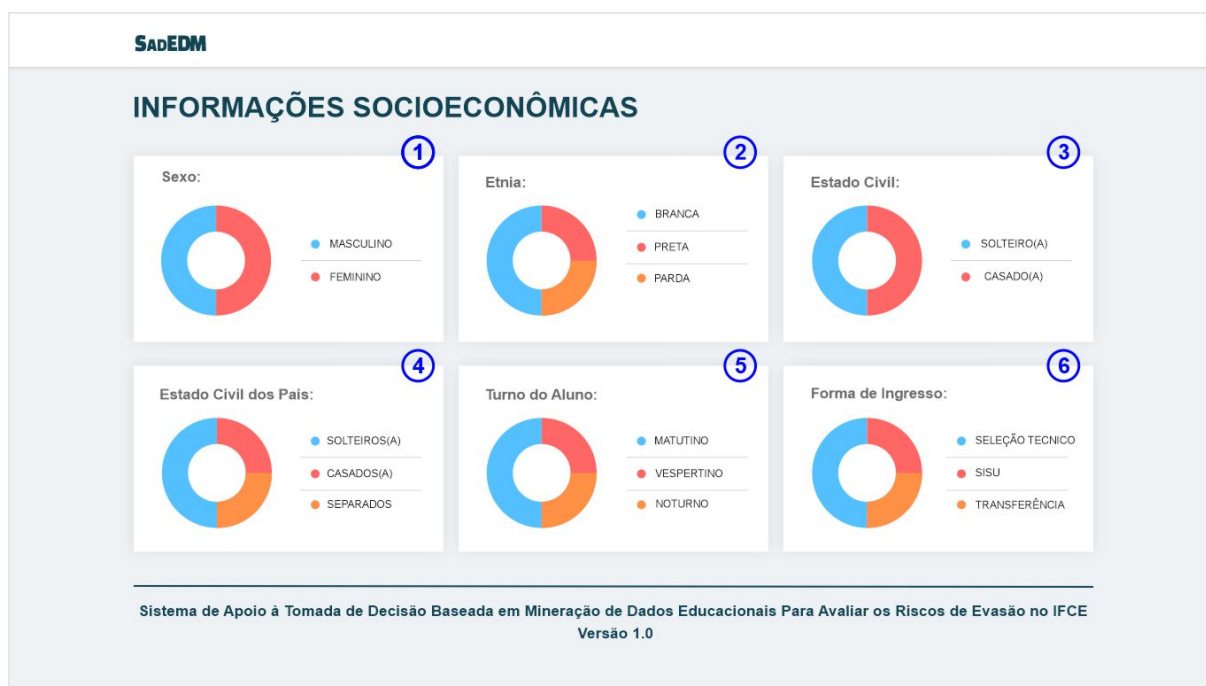
- **01, 02 e 03.** Os locais correspondentes aos itens 1, 2 e 3 são filtros que selecionam os dados dos alunos que pertencem aos seus respectivos campus, curso e semestre.
- **04.** Corresponde ao segundo filtro que está relacionado a porcentagem que os alunos possuem de sucesso no curso. Esses botões determinam pela cor, o *status* correspondente a suas porcentagens. Na cor vermelha, as chances de obter sucesso estão baixas. Na cor laranja, o sinal é de atenção, ou seja, esse estudante corre riscos de desistência. Na cor verde, as taxas de sucesso estão altas e as chances de evasão são baixas.
- **05.** A lista de alunos é exibida no item 5 após os filtros selecionados. Na lista de aluno constam nome, idade, porcentagem de sucesso e o *status* (cor verde, laranja ou vermelha). As porcentagens usadas no sistema para informar a situação dos estudantes analisado são

adquiridas através do cálculo realizado pelo algoritmo SVM (Support Vector Machine), assim como explorado na Etapa 3. Esse cálculo é realizado pela busca de um hiperplano que possui a capacidade de separar duas classes de amostras positivas e negativas em um determinado conjunto de dados.

- **06.** É o botão relacionado ao *dashboard* (gráficos), que exibe os dados socioeconômicos dos alunos que foram listados a partir da seleção feita através dos filtros (Figura 5).

Os dados exibidos na tela de *dashboard* (Figura 5) tem o objetivo de determinar um perfil de um possível aluno desistente e também de um aluno com chances altas de sucesso. Para isso, os dados observados são: 1 – Sexo; 2 – Etnia; 3 – Estado Civil; 4 – Estado Civil dos Pais; 5 – Turno de Aluno; e 6 – Forma de Ingresso.

Figura 5 – Tela do *dashboard* para a exibição dos dados socioeconômicos.



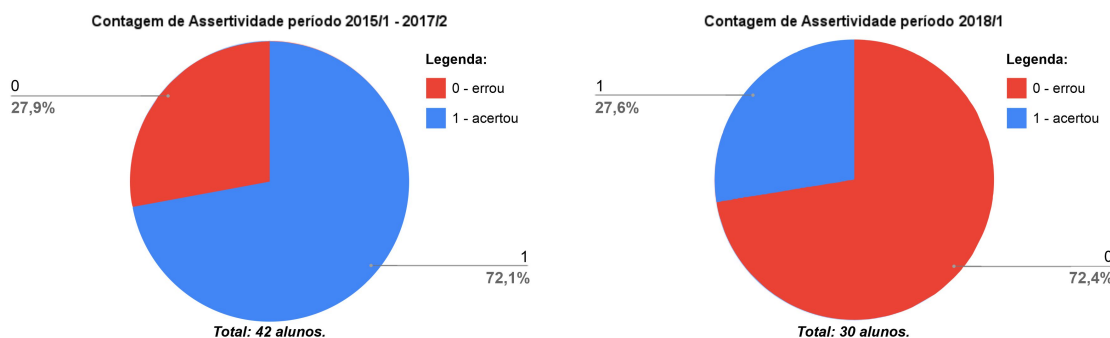
Fonte: Elaborada pelos autores (2022).

Por fim, vale a pena ressaltar que a assertividade do sistema em exibir as chances de ‘desistência’ e ‘sucesso’ passa pela consistência nos dados explorados pelo modelo preditivo (Etapa 3). Por exemplo, os alunos do primeiro semestre do ano de 2018 possuem alguns dados relacionados ao Rendimento Acadêmico que ainda não estão totalmente completos por conta do período de coleta. Com isso, a assertividade não é totalmente precisa. Por outro lado, a assertividade dos estudantes dos outros semestres analisados (2015/1 a 2017/2) possuem a assertividade bem próxima à exatidão. Isso porque os dados do Rendimento Acadêmico estão completos e, conseqüentemente, a classificação se torna mais precisa.

A Figura 6 ilustra a taxa de assertividades. Na Figura é possível observar que o gráfico com os estudantes do período de 2015/1 - 2017/2 possui uma taxa de assertividade de 72,1% e errando apenas 27,9% de um total de 42 alunos. Já o gráfico com os estudantes do período

de 2018/1, possui uma assertividade de apenas 27,6% e errando 72,4% de um total de 30 alunos. Logo, pode-se concluir que quanto mais os dados estiverem completos, maior vai ser a assertividade do sistema web construído, tornando-o mais confiável para o usuário final.

Figura 6 – Taxas de assertividade.



Fonte: Elaborada pelos autores (2022).

4 CONCLUSÕES

Este trabalho apresentou um sistema web capaz de auxiliar professores e gestores do IFCE a reconhecer estudantes em situação de risco de evasão a partir da utilização de Aprendizagem de Máquina e Mineração de Dados Educacionais. Assim, utilizando o sistema construído é possível apresentar o perfil dos estudantes com risco de evadir da instituição e indicar um conjunto de indicadores que podem estar associados ao baixo desempenho dos estudantes. Com isso, juntamente com a realização de um acompanhamento do rendimento acadêmico, algumas ações poderão ser promovidas para auxiliar esses estudantes a permanecer matriculado e frequentando a instituição.

Como trabalhos futuros, pretende-se apresentar os resultados obtidos para os gestores e professores do IFCE avaliarem o potencial do sistema e os benefícios que ele pode trazer para a instituição.

REFERÊNCIAS

ALMEIDA, T. C. de O.; OLIVEIRA, R. d. C. da S. As causas da evasão escolar de crianças e adolescentes da educação básica e sua relação com a violação de direitos humanos. **Educação Contemporânea-Volume 04 Educação Inclusiva, Reflexões**, p. 52, 2020.

ALVES, C. C.; LAVOR, L. A. M. de; PEREIRA, H. P. Evasão escolar: um desafio para a educação na atualidade. **Revista de Pesquisa Interdisciplinar**, v. 2, n. 1, 2017.

BELTRAN, C. A. R. et al. Plataforma de aprendizado de maquina para detecção e monitoramento de alunos com risco de evasão. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2019. v. 30, n. 1, p. 1591.

BRASIL. Constituição da república federativa do brasil. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 1988. Senado Federal. Disponível em: <https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>.

CASTRO, L. N. de; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. 1. ed. São Paulo: Saraiva, 2016.

FACELI, K. et al. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro: LTC, 2011.

FERREIRA, J. A. d. O. A. et al. Criação de um painel de controle para prevenção da evasão escolar no instituto federal de educação, ciência e tecnologia do Amazonas. Brasil, 2021.

GUZMÁN-CASTILLO, S. et al. Implementation of a predictive information system for university dropout prevention. **Procedia Computer Science**, v. 198, p. 566–571, 2022. ISSN 1877-0509. 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050921025266>>.

HILMARSSON, H. Æ. **Using Machine Learning for Predicting the Likelihood of Upper Secondary School Student Dropout**. Tese (Doutorado) — University of Iceland, 2019.

IBGE. **Instituto Brasileiro de Geografia e Estatística**, 2019. [Último Acesso: 03-12-2022]. Disponível em: <<https://censos.ibge.gov.br/>>.

IDEB. **Índice de Desenvolvimento da Educação Básica**, 2021. [Último Acesso: 03-12-2022]. Disponível em: <<http://ideb.inep.gov.br/>>.

IFCE. **IFCE em Números**, 2022. [Último Acesso: 03-12-2022]. Disponível em: <<https://ifceemnumeros.ifce.edu.br/>>.

INEP. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**, 2021. [Último Acesso: 03-12-2022]. Disponível em: <<https://www.gov.br/inep/>>.

PNAD. **Pesquisa Nacional por Amostra de Domicílios**, 2020. [Último Acesso: 03-12-2022]. Disponível em: <<https://censos.ibge.gov.br/>>.

QUEIROZ, L. D. Um estudo sobre a evasão escolar: para se pensar na inclusão escolar. **Rev Bras Estudos Pedag**, v. 64, n. 147, p. 38–69, 2006.

RODRIGUES, R. L. Uma abordagem de mineração de dados educacionais para previsão de desempenho a partir de padrões comportamentais de autorregulação da aprendizagem. **Repositorio Digital da UFPE**, 2016. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/21132>>.

SANTOS, V.; SARAIVA, D.; OLIVEIRA, C. Uma análise de trabalhos de mineração de dados educacionais no contexto da evasão escolar. In: **Anais do XXXII Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2021. p. 1196–1210. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/18142>>.

SARAIVA, D. V. **SoMDAR: Soluções baseadas em Mineração de Dados para Avaliar Riscos de evasão de estudantes**. Dissertação (Mestrado em Ciências de Computação) — Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Fortaleza, 2020.

SCHREIBER, J. N. C. et al. Software sdbayes: Um auxílio para a predição de evasão discente. In: . [s.n.], 2017. Disponível em: <<https://repositorio.ufsc.br/xmlui/handle/123456789/181197>>.

SILVA, E. M. C. Um modelo descritivo para auxiliar o acompanhamento da evasão escolar nos cursos técnicos e superiores no Instituto Federal do Rio Grande do Norte-Campus do São Gonçalo do Amarante. Dissertação (Mestrado) — Brasil, 2019.

SILVA, G. L. B.; CARVALHO, J. A. de; MACIEL, A. M. A. Desenvolvimento de um learning analytics dashboard a partir de modelos de mineração de dados educacionais. **Revista de Engenharia e Pesquisa Aplicada**, v. 6, n. 3, p. 59–69, 2021. Disponível em: <<https://doi.org/10.25286/repa.v6i3.1688>>.