

Predição de Evasão Utilizando Técnicas de Classificação: Um Estudo de Caso do Instituto Federal do Ceará

Fernando W B H Filho¹, Diogo S Siqueira¹, Bruno C Leal²

¹Instituto Federal do Ceará (IFCE)
Paracuru – CE – Brasil

²Instituto Federal do Piauí (IFPI)
Floriano – PI – Brasil

fernando.wagner@ifce.edu.br, diogo.silva.siqueira08@aluno.ifce.edu.br

brunoleal@ifpi.edu.br

Abstract. *The students' dropout has consolidated over time as one of the main problems present in national and international educational institutions. The literature lists numerous causes and factors that may justify student dropout, and several actions have been and are being proposed in an attempt to minimize dropout rates. This paper faces the students' dropout issue as a classification problem. Based on this premise, it aims to propose a methodology for predicting student dropping out, using the leading machine learning classifiers over previously collected data, taking into account the study case of Paracuru campus of Instituto Federal do Ceará (IFCE).*

Resumo. *A evasão escolar tem se consolidado ao longo do tempo como uma das principais problemáticas presentes nas instituições de ensino nacionais e internacionais. A literatura elenca inúmeras causas e fatores que podem vir a justificar uma evasão discente e várias ações já foram e vêm sendo propostas no intuito de tentar minimizar os índices de evadidos. O presente artigo encara a evasão escolar como um problema de classificação e, baseado nesta premissa, tem como objetivo propor uma abordagem para predição de evasão de discentes, utilizando os principais classificadores de aprendizagem de máquina, a partir de dados previamente coletados, levando em consideração o estudo de caso do campus Paracuru do Instituto Federal do Ceará (IFCE).*

1. Introdução

A evasão de alunos tem sido um dos problemas mais relevantes para as instituições de ensino (IEs), em especial, as que focam em oferta de ensino técnico e superior. A gestão de ensino das instituições frequentemente promovem discussões e debates com seus colaboradores acerca de como minimizar os índices e impactos deste fenômeno. Governos ao redor do mundo também esboçam preocupação considerável com o fenômeno da evasão, uma vez que esta impacta os índices educacionais de uma nação que, por sua vez, tem como consequência a queda no crescimento econômico e na tendência ao aumento das desigualdades sociais.

A literatura [Moreira et al. 2018, Vieira et al. 2017] expressa que a evasão pode ter como motivação as mais variadas causas, que podem consistir tanto de fatores inter-

nos, como externos à instituição. A Tabela 1 mostra os principais fatores, devidamente classificados, que podem se fazer presentes na vida do discente evadido.

Tabela 1. Exemplo de fatores relacionados à evasão.

Fatores internos	Fatores externos
Falta de recursos e de segurança nas escolas;	Falta de transporte local/municipal ao estudante;
O excesso de alunos nas salas de aula;	Vulnerabilidade socioeconômica do estudante;
Falta de qualificação dos professores;	Ausência de ambiente/condições de estudo em casa;
Matrizes curriculares e projetos de curso desatualizados e desalinhados com as necessidades atuais do mercado;	Falta de oportunidades de trabalho na área do curso do discente;

Mediante detecção dos fatores de evasão, ações vêm sendo propostas e implementadas em todos os âmbitos, sempre com o objetivo de tentar diminuir os índices de evasão como um todo. Como exemplo dessas ações, podemos citar a promoção de dinâmicas para aproximação de alunos e seus responsáveis da instituição, a frequente revisão e reestruturação da matriz curricular de cursos, investimento na infraestrutura das dependências físicas da instituição, em especial laboratórios e bibliotecas, a capacitação de servidores e colaboradores, a ampliação programas de bolsas de pesquisa, extensão e monitorias, bem como auxílios para os alunos mais carentes. Além disso, a criação e manutenção de ações inclusivas, que têm como objetivo propagar a tolerância e diversidade étnico-raciais, de gênero e necessidades especiais, e.g., problemas de visão, audição, desenvolvimento intelectual ou motor.

Muitas das ações de engajamento e manutenção dos alunos requerem uma detecção prévia daqueles que têm ou que poderiam vir a ter uma predisposição a evadir-se da instituição, como por exemplo ações que pregam um acompanhamento mais próximo por parte de profissionais do setor pedagógico (e até mesmo do próprio docente) junto ao discente. Nesse contexto levanta-se, então, a problemática: seria possível utilizar-se de técnicas de aprendizagem de máquina para tentar prever discentes propensos a uma possível evasão? Este trabalho tem como objetivo tentar responder esta pergunta a partir de um estudo de caso com dados do IFCE - campus Paracuru.

O restante deste trabalho está estruturado como descrito a seguir. A Seção 2 aborda os principais conceitos acerca do problema de classificação. Em seguida, na Seção 3 são apresentados e discutidos os principais trabalhos relacionados. A Seção 4 aborda a metodologia, bem como informações dos dados e tecnologias utilizadas. Na Seção 5 são apresentados os resultados obtidos em avaliação experimental e suas discussões. E, finalmente, a Seção 6 consiste das conclusões e trabalhos futuros.

2. Aprendizagem de Máquina – O Problema de Classificação

A aprendizagem de máquina (*machine learning*) é uma subárea da inteligência artificial que envolve conceitos de várias áreas de estudo, como matemática, estatística, mineração de dados, e que consiste do desenvolvimento de técnicas e algoritmos visando o aperfeiçoamento de determinada tarefa a partir da interpretação e entendimento dos dados já existentes relativos a eventos e acontecimentos passados.

Ao lidar com estes dados, os algoritmos de aprendizagem geram um modelo que, em tese, tem a capacidade de prever a probabilidade de um determinado evento acontecer. Desta forma, o uso de técnicas de *machine learning* (ML) têm crescido consideravelmente para criação de modelo de análise preditiva para um dado objetivo. Como exemplo de casos de uso, tem-se a prevenção de acidentes, a predição de tendências, a fidelização e descoberta de clientes em um determinado negócio, dentre outros [Mueller and Massaron 2016].

A classificação é uma das categorias de problemas dentro da área de ML que visa rotular um determinado registro com base em um modelo previamente construído através de um algoritmo classificador, em relação aos dados interpretados. Em essência, trata-se de um problema de decisão, no qual, a partir do modelo gerado [Bruce and Bruce 2019]: *i)* determina-se uma probabilidade de corte para a classe de interesse acima da qual consideramos um registro como pertencente àquela classe; *ii)* usar o modelo para estimar a probabilidade do registro pertencer a uma das classes; *iii)* se tal estimativa estiver acima da probabilidade de corte, atribuir o novo registro a classe apropriada.

Podemos utilizar modelos de classificação para, por exemplo, definir se um *email* trata-se de um *spam* ou não, se um determinado usuário é propenso a comprar certo produto e, até mesmo, se um certo tipo de acidente de trânsito poderá acontecer dadas determinadas circunstâncias.

Neste trabalho utilizamos e comparamos os principais classificadores existentes na literatura para definição de modelos para predição da evasão de um determinado discente, considerando o caso do IFCE - campus Paracuru. Entende-se que os resultados de uma boa classificação podem auxiliar a gestão acadêmica de uma instituição no tratamento do problema da evasão, pois permite agir através de ações preventivas e tomadas de decisão a nível institucional.

3. Trabalhos Relacionados

Dada a relevância do problema de evasão e o aumento na utilização de soluções em aprendizado de máquina, existe considerável literatura abordando cada um destes pontos. O trabalho desenvolvido por [Moreira et al. 2018] apresenta uma revisão da literatura acerca do problema da evasão em cursos de educação profissional e tecnológica, focando no levantamento das principais causas deste fenômeno. Como resultado, elencou-se a dificuldade na conciliação entre estudo e trabalho, deficiências no processo de aprendizagem, falta de incentivo e capacitação dos docentes, o excesso de conteúdo acadêmico, além de problemas de saúde e gravidez que venham a surgir. O autor conclui que há uma necessidade de maior atenção por parte do poder público para propiciar a permanência e êxito dos alunos nos cursos.

Em [Sousa et al. 2015], os autores focaram no uso de métodos estatísticos para determinação dos principais fatores que impactam no índice de reprovação e evasão no âmbito do IFCE - campus Fortaleza, com dados captados de alunos relativos ao período compreendido entre 2006 a 2013. Utilizaram-se da estatística descritiva e do método econométricos de efeitos fixo e aleatórios para geração de um modelo de regressão, estabelecendo “pesos” que determinam o quanto uma variável impacta na reprovação e na evasão de um discente.

Soluções de predição por classificação também vêm ganhando destaque em áreas

e eventos de cunho gerencial. No trabalho desenvolvido em [Kantorski et al. 2016], os autores utilizam a ferramenta Weka para desenvolver e estabelecer o melhor modelo gerado dentre os classificadores disponíveis. O mesmo trabalho utilizou métricas como precisão, revocação (recall), para o cálculo da chamada medida- F (F -measure), que se propõe a ser uma medida que leva em consideração a cobertura e a corretude da solução.

Em âmbito internacional, [Şara et al. 2015], utiliza soluções de machine learning para predição da evasão (school dropout) no ensino médio (high school) utilizando dados de pouco mais de setenta mil alunos, separando 50% para a etapa de definição e treino do modelo, e os outros 50% para os testes de validação e corretude, obtendo bons resultados.

Por fim, o trabalho desenvolvido em [Maria et al. 2016] propõe o SPEED, um sistema computacional que utiliza redes bayesianas para prever os percentuais de chance de evasão dos alunos e, com isso, auxiliar os gestores educacionais na prevenção dessas situações. Os autores utilizaram dados oriundos de discentes dos cursos do SENAI da cidade de Tubarões/SC, além de uma metodologia como *cross-validation* (validação cruzada) para minimizar a possibilidade de vícios no modelo construído.

4. Dados Utilizados e Avaliação Experimental

Esta seção apresenta uma contextualização dos dados utilizados para predição de evasão de alunos para, em seguida, descrever a avaliação experimental e sua metodologia. Por fim apresenta os resultados e, mais importante, uma discussão para o contexto específico de predição da evasão de alunos com técnicas de *machine learning*.

4.1. Obtenção dos dados

Assim como inúmeras instituições ao redor do mundo, o IFCE também vem sofrendo com o problema da evasão, sobretudo após o período de expansão ocorrido ao longo última década. Como forma de combate ao problema, uma importante iniciativa foi criada: a plataforma IFCE em Números [Pró-reitoria do Instituto Federal do Ceará 2017]. A Figura 1 mostra uma consulta na plataforma a dados de cursos no período de 2019-2.

Nível de Ensino	Cód. do C.	Curso+Modalidade	Total geral	Em Curso	Retidos	Evadidos	Formados	
Total geral			43.522	24.399	10.017	5.589	3.517	CAMPI 32
Básico	131040	ESPAÑHOL BÁSICO I	2	2				CURSOS 519
	101009	ESTAMPARIA ARTESANAL EM TECIDO	3	3				INGRESSANTES 13.047
	141044	VIOLÃO- INTERMEDIÁRIO III	3			1	2	MATRICULADOS (total em curso + total de egressos) 43.522
	271027	CAD 3D - BÁSICO	3	1		2		Situação Atual das Matrículas
	061029	BÁSICO DE VIOLÃO	5	5				Em Curso 24.399
	111028	INGLÊS INTERMEDIÁRIO	5	3		2		Evadidos 5.589
	252013	DESENVOLVIMENTO DE JOGOS DIGITAIS	5	5				Formados 3.517
	252026	XADREZ BÁSICO	5	5				Retidos 10.017
	011057	INICIAÇÃO AO ACORDEÃO	6			4	2	Total ge.. 43.522
	091032	INICIAÇÃO A MATEMÁTICA	6			2	4	
	141045	HARMONIA E IMPROVISACÃO II	6			2	4	
	171028	INICIAÇÃO AO CAVAQUINHO	6			3	3	
	262032	MÚSICA & PROGRAMAÇÃO	6			2	4	
	321055	NEGÓCIOS DIGITAIS	6			2	4	
	111023	LÍNGUA BRASILEIRA DE SINAIS - BÁSICO	7	6	1			
	171024	CÁLCULO APLICADO	7			1	6	
	211019	AVANÇADO EM LÍNGUA ESPANHOLA	7		4	3		
	262035	ECOLOGIA E MEIO AMBIENTE	7			1	6	

Figura 1. Uma consulta a dados de cursos relativos ao período 2019-2

A plataforma IFCE em números surgiu da necessidade de se fornecer uma visão sistêmica da instituição e seu contexto, de modo a auxiliar no planejamento e na definição

de ações de combate a evasão e a reprovação discente. Essa plataforma surgiu da consolidação de dados dos sistemas acadêmicos de forma centralizada, e teve como premissa de desenvolvimento o fácil acesso e o alto nível de transparência, não somente para a gestão da instituição, mas para com a sua comunidade – docentes, técnico-administrativos, discentes, pesquisadores, dentre outros. A partir da plataforma, foi possível obter os dados de alunos agregados por campus, por curso e por período, sendo também possível compilar e exportar todos os dados desejados em formato de planilhas eletrônicas, e.g. formato “.csv”. A partir dessas planilhas, se fez possível a utilização de bibliotecas para extração, manipulação, processamento e posterior análise desses dados.

A linguagem *Python* foi utilizada para a realização dos experimentos práticos. Destacam-se também as bibliotecas *Pandas*¹, que permite a leitura e a manipulação de dados tabulares, oriundos de planilhas, e a *scikit-learn*², uma biblioteca contendo suporte e algoritmos para soluções em análise em aprendizagem de máquina.

4.2. Avaliação Experimental

Em um primeiro momento, houve a coleta dos dados do campus Paracuru, oriundos da plataforma IFCE em Números. Foram extraídos 1942 registros, dos quais 70% foram separados para treino e construção do modelo e 30% para testes de validação do mesmo. Essa etapa consiste de remover eventuais inconsistências e incorretudes nos dados, além de extraí-los das planilhas e adequá-los para a correta interpretação dos classificadores. Foi realizado, também, um processo para normalização dos dados.

Uma das principais transformações nesta fase foi o mapeamento de um valor categórico para um numérico. Por exemplo, a informação de que um discente estava retido ou não (isto é, se ele atrasou o curso) foi mapeada para 1 (em caso de retenção) e 0 (em caso de não haver retenção). Esta lógica foi aplicada a todos as demais variáveis, uma vez que os classificadores, via de regra, trabalham com valores numéricos.

Para a geração do modelo, foram levadas em consideração as seguintes variáveis acerca de cada discente: *i*) se é ingressante no período ou não; *ii*) se encontra-se retido no curso; *iii*) se é cotista ou não; *iv*) a etnia do aluno; *v*) sexo; *vi*) nível de ensino (técnico, graduação ou básico); *vii*) faixa etária; e *viii*) se é natural do mesmo município da instituição ou a outro município.

Após normalizados, todos os dados foram submetidos em validação cruzada com cinco iterações, aos seguintes classificadores para a geração dos modelos: *i*) Árvore de decisão (*Decision Tree*); *ii*) *Naive Bayes*; *iii*) *KNeighbors* (KNN); *iv*) Máquina de Vetor de Suporte (*Support Vector Machine*); *v*) *GradientBoosting*.

Uma vez obtidas as médias dos resultados de cada classificador com todos os registros do campus Paracuru, percebeu-se que havia um certo desbalanceamento na carga de dados, visto que o número de alunos não-evadidos é muito superior ao de alunos evadidos. A partir desta percepção, optou-se por balancear os dados, aproximando o quantitativo da amostra de não-evadidos ao de evadidos. Após o balanceamento, novos modelos foram gerados e os resultados captados.

¹pandas.pydata.org/

²scikit-learn.org/

5. Resultados e discussões

O primeiro passo para a geração dos resultados, após a execução dos classificadores, foi, a partir dos acertos e erros gerados, calcular a matriz de confusão. A matriz de confusão é uma matriz cujos valores da diagonal principal constituem os acertos, e a da diagonal inversa constituem os erros gerados pelo classificador. A Tabela 2 exhibe a matriz de confusão gerada a partir dos resultados do classificador árvore de decisão.

Tabela 2. Matriz de confusão gerada a partir do algoritmo de árvores de decisão

	não-evasão	evasão
não-evasão	289	92
evasão	145	57

A partir da matriz de confusão, é possível calcular a taxa de acurácia que o classificador gerou. A acurácia representa o nível de acerto do classificador e pode ser calculada dividindo a soma do total de acertos pelo total de registros avaliados. Desta forma, para um total de 583 registros de teste, a acurácia obtida a partir da matriz de confusão da Tabela 2 é $(289 + 57)/583 = 0,59$, ou seja, 59%. No entanto, apesar de bastante utilizada, a acurácia por si só pode mascarar problemas nos modelos gerados, em especial, oriundos do desbalanceamento entre os dados. Tome como exemplo que em uma determinada escola, 99% dos alunos não evadiram, então, um modelo que simplesmente determina que todo aluno não evadirá, terá em média 99% de acurácia. No entanto, é nula sua capacidade de identificar possíveis evasões.

Para maior capacidade de análise dos resultados, além da acurácia, as medidas de precisão (Equação 1), revocação (Equação 2) e especificidade (Equação 3), são frequentemente utilizadas em aplicações de classificação, uma vez que tendem a avaliar a capacidade dos modelos de identificar alunos que tendem a evadir, e que tendem a não evadir. No qual P é a precisão do modelo, R é a revocação do modelo, E é a especificidade do modelo, T_p é o número de evadidos que o modelo conseguiu prever corretamente, T_n é o número de não-evadidos previsto corretamente, F_n é o número de não-evadidos que o modelo acusou como evadidos (falso positivo), e F_p é o número de evadidos que o modelo acusou como não evadidos (falso positivo).

$$P = \frac{T_p}{T_p + F_p} \quad (1) \quad R = \frac{T_p}{T_p + F_n} \quad (2) \quad E = \frac{T_n}{T_n + F_n} \quad (3) \quad F = \frac{2 \times (P \times R)}{(P + R)} \quad (4)$$

Mais especificamente, a revocação mede a capacidade do modelo de identificar corretamente alunos evadidos em relação aos demais da amostra; a precisão é a capacidade de prever corretamente os alunos identificados como evadidos; e a especificidade mede a capacidade do modelo de prever um aluno não-evadido. Já a medida- F (Equação 4) representa uma média harmônica de precisão e revocação, uma vez que podem existir situações em que o modelo produz alta precisão, mas baixa revocação, ou vice-versa, portanto torna-se interessante levar em consideração o equilíbrio entre essas medidas.

Os experimentos consistiram de cinco execuções com validações cruzadas. Como resultado foi tomada a média dos valores obtidos em cada execução. O cálculo do desvio

padrão foi feito da mesma forma a partir dos valores absolutos. A Tabela 3 exibe os resultados da acurácia, precisão, revocação, especificidade e medida- F para cada classificador com balanceamento e sem balanceamento de carga, destacando em negrito os melhores resultados. Para balanceamento entre as classes foi utilizada a técnica de *Under-sampling*, eliminando aleatoriamente aproximadamente 45% da classe majoritária (não-evadidos). Os valores da esquerda (s/ bal.) representam os valores obtidos de cada um dos classificadores antes do processo de balanceamento, e os valores da direita (c/ bal.) são referentes aos obtidos após o processo de balanceamento. Logo abaixo de cada resultado apresenta-se o valor médio de desvio padrão.

Tabela 3. Resultados dos medidores para cada classificador para o campus Paracuru.

Classificador		Acurácia		Precisão		Revocação		Especificidade		Medida-F	
		s/ bal.	c/ bal.	s/ bal.	c/ bal.	s/ bal.	c/ bal.	s/ bal.	c/ bal.	s/ bal.	c/ bal.
Decision Tree	μ	0.63	0.58	0.46	0.60	0.32	0.50	0.69	0.57	0.38	0.54
	σ	0.016	0.005	0.029	0.015	0.020	0.016	0.014	0.013	0.018	0.007
NaiveBayes	μ	0.63	0.57	0.49	0.57	0.24	0.57	0.66	0.57	0.32	0.57
	σ	0.009	0.027	0.022	0.030	0.012	0.048	0.005	0.041	0.014	0.023
KNeighbors (k=5)	μ	0.62	0.58	0.46	0.59	0.36	0.54	0.69	0.56	0.40	0.57
	σ	0.005	0.018	0.030	0.037	0.039	0.026	0.018	0.023	0.027	0.031
Support Vector Machine	μ	0.67	0.60	0.54	0.58	0.04	0.61	0.68	0.62	0.08	0.59
	σ	0.009	0.023	0.141	0.026	0.025	0.055	0.011	0.024	0.042	0.035
Gradient Boosting	μ	0.66	0.59	0.55	0.60	0.23	0.59	0.68	0.59	0.32	0.59
	σ	0.016	0.014	0.048	0.044	0.018	0.076	0.014	0.020	0.026	0.023

A partir dos resultados obtidos, observa-se que, em especial, após o processo de balanceamento, os modelos obtiveram níveis consideráveis de precisão, revocação, especificação e medida- F , chegando a patamares de quase 60%. Percebe-se que os níveis de acurácia foram, em média, levemente maiores, na maioria dos modelos gerados, antes do processo de balanceamento. No entanto, isso não significa necessariamente que o modelo gerado foi melhor, uma vez que, após o balanceamento, os níveis de precisão e revocação aumentaram, e conseqüentemente, a medida- F também. Com relação à especificidade, os níveis diminuiram após o balanceamento, muito provavelmente porque o número de registros de não-evadidos diminuiram na carga de treino dos modelos.

Observa-se o processo de balanceamento como algo positivo e que tende a gerar melhores valores nas métricas de avaliação. No entanto, via de regra, o quantitativo de alunos evadidos tende a ser bem menor que o número de alunos não-evadidos (matriculados, retidos, formados), o que faz com que a carga de dados total utilizada na fase de treino do classificador seja muitas vezes consideravelmente mais baixa do que poderia ser. Isso pode representar um fator limitante para a qualidade dos modelos gerados, pois sabe-se que uma carga de dados maior e mais diversa é propensa a gerar modelos com mais qualidade. O modelo gerado para o estudo de caso apresentado tende a melhorar com o tempo, uma vez que a cada semestre, novos dados serão gerados e poderão ser agregados na construção dos modelos, permitindo assim, ter uma maior amostragem para formar uma carga de dados balanceada entre evadidos e não-evadidos.

6. Conclusão

O presente trabalho apresentou uma abordagem para predição da evasão de alunos com técnicas de aprendizagem de máquina, definindo-o como um problema de classificação.

Nesta abordagem, utilizou-se como fonte os dados oriundos do portal IFCE em números e como geradores de modelos os principais classificadores utilizados na literatura. Os resultados atingiram acurácia em torno de 60% preservando uma medida-F em torno de 60%, gerando, principalmente, discussões e conclusões para criação de modelos no cenário abordado considerando várias métricas de avaliação.

Como trabalhos futuros, há a possibilidade de considerar novas variáveis, como os dados de rendimento acadêmico dos discentes. Além disso, pode-se investigar maneiras de melhorar os resultados dos classificadores, através de possíveis parametrizações, bem como tentar verificar quais as variáveis que mais impactam nos cálculos feitos pelos modelos construídos. Com isso, será possível tentar explicar quais as variáveis que podem estar mais fortemente associadas com um aluno que tende a evadir. Por fim, é almejado a construção de um sistema completo para que a pró-reitoria de ensino possa utilizar no auxílio da tomada de decisões visando diminuir os níveis e impactos da evasão na instituição.

Referências

- Bruce, A. and Bruce, P. (2019). *Estatística Prática para Cientistas de Dados*. Alta Books.
- Kantorski, G. Z., Schmitt, J. A., Flores, E. G., and Hoffmann, I. L. (2016). Predição da evasão dos alunos em cursos de graduação presenciais. *XXXVI Encontro Nacional de Engenharia de Produção*.
- Maria, W., Damiani, J. L., and Pereira, M. (2016). Rede bayesiana para previsão de evasão escolar. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 5, page 920.
- Moreira, L. K. R., Lambert, A. S., and Castro, R. C. A. M. (2018). Educação profissional e tecnológica: permanência e evasão em foco. *Revista Brasileira de Educação e Saúde*, 8(4):48–53.
- Mueller, J. P. and Massaron, L. (2016). *Machine learning for dummies*. John Wiley & Sons.
- Pró-reitoria do Instituto Federal do Ceará (2017). IFCE em números. <https://ifceemnumeros.ifce.edu.br/>. Acesso em 25 jun. 2020.
- Sousa, J. N. M., Tabosa, F. J., Simonassi, A., and Castelar, P. U. C. (2015). Principais fatores que impactam na reprovação e evasão dos alunos dos cursos tecnológicos no instituto federal de ciência e tecnologia do ceará. *XI Encontro de Economia do Ceará-IPECE*.
- Vieira, A. C. F., de Lima Gallindo, E., and Cruz, H. A. (2017). Plano estratégico para permanência e êxitos dos estudantes do IFCE. Disponível em <https://gestaoproen.ifce.edu.br/attachments/download/3052/>. Acesso em 23 jun. 2020.
- Şara, N.-B., Halland, R., Igel, C., and Alstrup, S. (2015). High-school dropout prediction using machine learning: a danish large-scale study. In *23rd European Symposium on Artificial Neural Networks, ESANN2015*, pages 319–324.