

# **Análise de Técnicas de Aprendizagem de Máquina Supervisionado Aplicadas ao Problema de Classificação de Alunos do Ensino Superior Privado com Risco de Evasão**

**Natallyson Saraiva de Oliveira, Prof. Dr. Sergio Cleger Tamayo**

<sup>1</sup>Fundação Centro de Análise, Pesquisa e Inovação Tecnológica (FUCAPI)  
Manaus – AM – Brasil

{natallyson12, sergio.cleger}@gmail.com

**Abstract.** *This project addresses the problem of dropout of undergraduate students from private higher education institutions (HEI). They were used in the database as voluntary responses via the online form of 201 students from the Center for Analysis, Research and Technological Innovation Foundation (FUCAPI). To solve this problem, three methods of machine learning used without the supervised learning paradigm were presented. The Decision Tree method achieves the best result, reaching an F1-micro-score of 85.24%.*

**Resumo.** *Este projeto aborda o problema de evasão de alunos de graduação, provenientes de instituições de ensino superior (IES) de cunho privado. Foram utilizadas na base de dados as respostas voluntárias via formulário online de 201 alunos da Fundação Centro de Análise, Pesquisa e Inovação Tecnológica (FUCAPI). Para solução deste problema, foram elencados três métodos de aprendizado de máquina baseados no paradigma de aprendizado supervisionado. O método de Árvore de Decisão obteve o melhor resultado, alcançando F1-score-micro de 85,24%.*

## **Introdução**

De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), existem 2.152 Instituições de Ensino Superior (IES) de natureza privada em território nacional, representado 87,9% das IES do Brasil [INEP 2018]. Um dos fatores que têm influenciado no aumento do número de IES privadas, bem como as vagas disponíveis nas mesmas, são os programas governamentais, como o ProUni e o Fies, que proporcionam um leque de opções para o ingresso de novos alunos no IES [Norte 2018].

Dadas as informações fornecidas por instituições governamentais e o cenário da presença de alunos de baixa renda no ensino privado, constrói-se o perfil do aluno das IES privadas no Brasil, onde o número de alunos oriundos de escolas públicas e de baixa renda vem crescendo gradativamente nos últimos anos [Souza 2017].

A facilitação do ingresso dos alunos de baixa renda no ensino superior por parte das IES privadas através de bolsas próprias ou de programas governamentais tem como ponto positivo a oportunidade para que estes possam obter um título de graduação, porém, o ingresso não implica diretamente na conclusão do ensino superior, já que a quantidade alunos que concluem a graduação, seja em tempo hábil ou não, é expressivamente inferior aos que ingressam anualmente nas IES, chegando a 10,4% de abandono por parte dos

alunos beneficiados com o ProUni. Já a taxa de alunos que não conseguem concluir o curso no tempo previsto chega a 41,7% [da Costa and de Barros Ferreira 2016].

Entender os motivos pelos quais os alunos do ensino superior abandonam um curso de graduação é alta relevância, tanto para as IES, tendo em vista a estagnação no crescimento do ensino superior de natureza privada nos últimos anos, quanto para o Estado, já que é necessária a presença de profissionais capacitados para o desempenho das mais diversas tarefas no mercado de trabalho.

## 1. Objetivos

Levando em consideração o cenário apresentado, seguem os objetivos deste trabalho.

### 1.1. Objetivo Geral

Utilizar métodos de aprendizado de máquina supervisionado para predição de abandono temporário ou definitivo de alunos do ensino superior privado, aplicados à Fundação Centro de Análise, Pesquisa e Inovação Tecnológica, FUCAPI.

### 1.2. Objetivos Específicos

- Construção de uma base de dados proveniente dos alunos de cursos de graduação diversos, considerando aspectos socioeconômicos;
- Documentação dos modelos de aprendizado de máquina baseados em paradigma de aprendizado supervisionado para aplicação no problema em questão;
- Documentação dos resultados obtidos;

## 2. Trabalhos Relacionados

Sendo um problema presente em todas as IES, a predição de evasão acadêmica tem sido tratada por diversos pesquisadores, utilizando diferentes técnicas de predição.

[Manhães et al. 2012] utilizaram algoritmos da ferramenta Weka, aplicados as notas dos primeiros semestres de alunos do curso de Engenharia Civil da Escola Politécnica da Universidade Federal do Rio de Janeiro, UFRJ, entre 1994 a 2005. Utilizando métodos como *Multi Layer Perceptron* e *Naive Bayes*, foi obtida acurácia média de 75% a 80%.

[Martinho et al. 2017] utilizaram Redes Neurais do tipo ARTMAP-Fuzzy, aplicadas a dados socioeconômicos obtidos no ano de matrícula, para predição do risco de evasão de alunos dos cursos de Automação Industrial, Controle de Obras e Sistemas para Internet, do Instituto Federal de Educação, Ciência e Tecnologia do Mato Grosso, IFMT, em Cuiabá. Os experimentos apresentaram acurácia média de 97%.

[Rovira et al. 2017] utilizaram métodos como regressão logística, máquinas de vetor de suporte e AdaBoost, para predizer, além de outras coisas, a tendência de abandono estudantil por parte de alunos dos cursos de Direito, Matemática e Ciências da Computação da Universidade de Barcelona. Com dados baseados nas notas dos alunos, o método de regressão logística obteve *F1-score* de 82% quando aplicado ao conjunto de dados dos alunos de Ciências da Computação, sendo este seu melhor resultado.

### 3. Materiais e Métodos

Dados os objetivos apresentados anteriormente, é necessário discorrer sobre os conceitos que envolvem as técnicas a serem empregadas para o solucionamento do problema em questão, bem como sobre a construção da base de dados a partir da qual os algoritmos selecionados irão adquirir o conhecimento necessário.

#### 3.1. Aprendizado de Máquina

Aprendizado de Máquina (AM) é uma sub-área da Inteligência Artificial (IA) cujo objetivo é a construção de sistemas computacionais capazes de adquirir conhecimento automaticamente, através dos dados a eles apresentados [Monard and Baranauskas 2003]. A literatura apresenta diversas formas de subdivisão para o AM, sendo a principal delas a divisão entre Aprendizado Não-Supervisionado e Aprendizado Supervisionado.

No aprendizado não-supervisionado, os dados do problema são apresentados ao sistema e este, de forma indutiva, tenta separar os dados de acordo com suas características, formando agrupamentos ou *clusters*. Já no aprendizado supervisionado os dados apresentados são rotulados de acordo com suas respostas desejadas, ou seja, o indutor possui a informação de qual grupo ou valor de referência os exemplos pertencem, cabendo ao indutor construir aprender a classificar (problemas de *classificação*) ou determinar valores contínuos (problemas de *regressão*) para novos exemplos que forem apresentados a ele [Monard and Baranauskas 2003].

Cada paradigma comporta diversos métodos de aprendizado, cada um com suas peculiaridades e semelhanças. Dentre os algoritmos de AM supervisionado serão descritos 3 dos mais recorrentes na literatura, são eles:

- Redes Neurais Artificiais;
- Classificador Naive Bayes; e
- Árvore de Decisão.

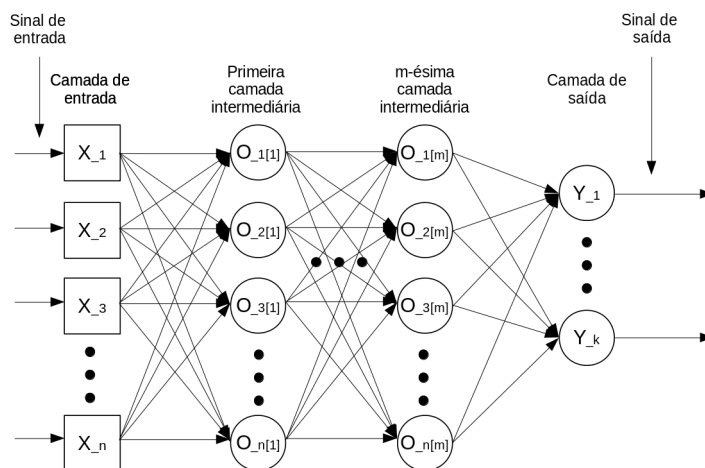
##### 3.1.1. Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são sistemas compostos por unidades simples de processamento, comumente chamadas de neurônios pela sua semelhança com as células do sistema nervoso humano. O processamento realizado pelas RNAs é dado de forma paralela e distribuída, tendo em vista a forma de agrupamento e conexionismo entre os neurônios (paralelismo) e a conexão as funções matemáticas presentes em cada um dos neurônios (processamento distribuído) [de Pádua Braga et al. 2000].

RNAs com uma única camada intermediária são capazes de aproximar qualquer função contínua, enquanto RNAs com duas camadas intermediárias são capazes de aproximar qualquer função matemática, porém, nem todas os problemas reais estão contidos nesse conjunto de funções, sendo necessário o agrupamento de neurônios em mais de duas camadas intermediárias. Para problemas cuja necessidade seja de duas ou mais camadas intermediárias uma das arquiteturas mais utilizadas é a rede Perceptron de Múltiplas Camadas (MLP, do inglês *Multi Layer Perceptron*) [de Pádua Braga et al. 2000].

A Figura 1 mostra como são organizadas internamente as RNAs do tipo MLP. A camada de entrada, ou retina, é responsável pela recepção dos dados do ambiente externo.

As camadas intermediárias são completamente conectadas, o que significa que todos os neurônios da camada  $m$  estão ligados a todos os neurônios da camada  $m + 1$ . A camada de saída possui  $k$  neurônios, onde  $k$  é o número total de classes dos dados de entrada (em problemas de classificação). Quando o problema se caracteriza como regressão, a camada de saída possui apenas um neurônio responsável pela emissão do valor contínuo de resposta.



**Figure 1. Arquitetura de uma rede MLP.**

### 3.1.2. Classificador Naive Bayes

A forma de aprendizado utilizada pelo Classificador Naive Bayes (CNB) segue o paradigma estatístico, onde o classificador faz a inferência das classes através de cálculos probabilísticos. O fato da classificação ser feita através de cálculos estatísticos faz com o CNB tenha como principal ponto positivo a possibilidade de enviesamento dos cálculos, através da manipulação das equações, caso haja conhecimento e domínio pleno sobre os dados. Porém, devidos o mesmo motivo, há a possibilidade do modelo necessitar de um alto poder computacional para obtenção de resultados consistentes [Pardo and das Graças Volpe Nunes 2002].

Dada a probabilidade  $P(x|y_i)$ , onde  $x$  é um exemplo apresentado ao CNB, a mesma pode ser decomposta no produto  $P(x_1|y_i) \times \dots \times P(x_d|y_i)$ , onde  $x_j$  é o  $j$ -ésimo atributo do exemplo  $x$ . Sendo assim, a probabilidade de  $x$  pertencer a classe  $y_i$  poder ser expressa pela Equação 1 [Faceli et al. 2011].

$$P(y_i|x) \propto P(y_i) \prod_{j=1}^d P(x_j|y_i) \quad (1)$$

### 3.1.3. Árvores de Decisão

Em uma Árvore de Decisão (AD), o problema apresentado é dividido em subproblemas de forma recursiva, afim de simplificar a inferência de uma resposta para o mesmo. A

qualidade da resposta depende da capacidade que o algoritmo de indução utilizado possui para dividir as instâncias do problemas em subespaços, onde cada subespaço é ajustado de acordo com características particulares. Quando o problema apresentado é do tipo *classificação*, a árvore utilizada é chamada de *árvore de decisão*, enquanto as árvores utilizadas para resolver problemas de *regressão* são chamadas de *árvores de regressão* [Faceli et al. 2011].

A estrutura recursiva de uma AD poder ser definida a partir das seguintes características:

- *Nó folha*: usualmente corresponde a uma das classes do atributo alvo dos exemplos apresentados;
- *Nó de decisão ou divisão*: contém um teste condicional sobre algum atributo descritivo do exemplos apresentados. Para cada resultado do teste condicional, existe uma aresta ligando-o à uma subárvore ou *nó folha*, onde cada subárvore possui estrutura recursiva idêntica a árvore pai.

A Figura 2 exemplifica um processo de diagnóstico de um paciente como saudável ou doente através de uma árvore de decisão simples.

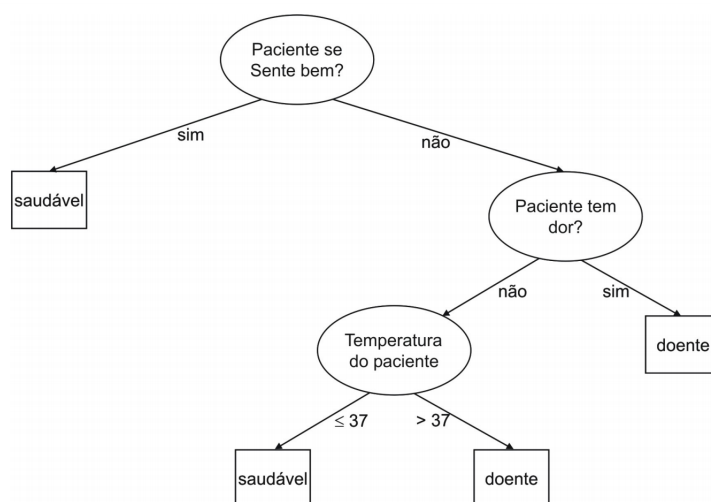


Figure 2. Exemplo de árvore decisão simples [Monard and Baranauskas 2003].

### 3.2. Coleta de Dados

A base de dados construída para o problema em questão foi obtida através das respostas de formulários digitais na plataforma Google Forms, provenientes de alunos de IES oriundos, em sua grande maioria, de natureza privada.

As perguntas presentes no formulário são voltadas para obtenção de informações socioeconômicas, afim de traçar o perfil dos estudantes com graduação em curso ou com matrícula trancada ou cancelada, sendo estes status o atributo alvo dos exemplos.

As resposta do formulário foram recebidas entre os dias 02 de Agosto de 2019 e 16 de Agosto de 2019, totalizando 15 dias de coleta de dados. No total, foram recebidas 246 formulários de resposta, sendo esta a base de dados inicial para o problema.

Conforme citado anteriormente, existem três agrupamentos principais de dados. O primeiro deles é o conjunto de dados pessoais, composto por: I – Data de nascimento; II – Endereço; e III – Estado civil [Solteiro(a), Casado(a) ou Viúvo(a)].

Já os dados de teor financeiro não restringe-se apenas a dados econômicos. Saber se o aluno possui curso técnico e/ou outra graduação já concluída implica, geralmente, em um status financeiro mais elevado, devido a qualidade e o maior número de oportunidades disponíveis.

Outra informação importante é o tipo de escola em que o aluno cursou o ensino médio, tendo em vista que, a variação entre escolas públicas estaduais e privadas geralmente reflete, em sua maioria, na classe social em que a família está inserida, enquanto origem de escolas públicas federais geralmente está relacionada a um ensino técnico. Os dados financeiros coletados foram: I – Tipo de escola em que cursou o ensino médio [Pública Estadual, Pública Federal ou Privada]; II – O aluno está trabalhando [Sim ou Não]; III – O aluno possui curso técnico [Sim ou Não]; e IV – O aluno já possui uma graduação [Sim ou Não].

Os dados acadêmicos coletados são relacionados ao curso de graduação vigente, sendo eles: I – Nome do curso de graduação; II – Turno em que estuda [Manhã, Tarde ou Noite]; III – Semestre de ingresso no curso; IV – Período atual do curso; V – Status atual da matrícula [Ativa, Trancada ou Cancelada]; VI – Já teve a matrícula Trancada ou Cancelada anteriormente: [Sim ou Não]; e VII – Caso já tenha tido a matrícula Trancada ou Cancelada anteriormente, qual o motivo?

Como já era previsto, o maior número de alunos que preencheram o formulário possui “Status atual da matrícula” Ativa, totalizando 212 resposta, o que equivale 85, 48% das respostas obtidas.

Com a disparidade entre as classes, decidiu-se por utilizar como atributo alvo o questionamento “Já teve a matrícula Trancada ou Cancelada anteriormente?”, onde as respostas negativas totalizaram 168, contra 80 respostas positivas. Mesmo a quantidade de respostas negativas sendo 2, 1 vezes maior que a quantidade de respostas positivas, a porcentagem de 67, 74% de respostas negativas, contra 32, 26% de respostas positivas deixa o atributo alvo mais equilibrado, amenizando o problema de enviesamento das predições.

### **3.3. Limpeza de Dados**

A fase de limpeza de dados foi dividida em duas etapas. A primeira etapa consistiu na análise das respostas submetidas por alunos do ensino superior e a exclusão de exemplos não condizentes com o público alvo do projeto. Tendo em vista que os dados foram coletados através de formulários online, alunos e ex-alunos de diversas IES tiveram acesso ao mesmo, gerando uma heterogeneidade na origem e situação socioeconômica do conjunto de exemplos.

Primeiramente foram removidos os dados de alunos oriundos de outras IES. Tal ação de limpeza foi realizada removendo dados de alunos onde o número de matrícula informado no formulário não condiz com a estrutura do número de matrícula da FUCAPI, assim, removendo possíveis alunos de IES da rede pública e de outras IES privadas.

Também foram removidos exemplos com ex-alunos do ensino superior, tendo estes concluído sua graduação na FUCAPI e sem nenhuma graduação em curso na mesma

instituição. Concluídas as duas remoções de exemplos, a base de dados foi reduzida de 246 para 201 a quantidade de exemplos.

A segunda etapa da limpeza de dados consistiu em normalizar as respostas dos usuários. Todas as respostas em formato de texto foram separadas as categorias numéricas. O endereço fornecido pelos usuários foi resumido a utilização do CEP em formato numérico e o semestre de ingresso foi categorizado de forma crescente, iniciando o semestre mais atual para o mais antigo, com crescimento unitário a cada semestre.

### 3.4. Separação de dados

Durante a análise das respostas obtidas pelos formulários, selecionou-se como atributo alvo o motivo pelo qual os alunos afastaram-se, seja temporariamente ou definitivamente, da graduação, sendo estes motivos classificados em:

<b>Não houve trancamento:</b> 0	<b>Financeiro:</b> 2
<b>Pessoal:</b> 1	<b>Administrativo:</b> 3
	<b>Acadêmico:</b> 4

Dada a relação direta entre já ter trancado ou cancelado a matrícula e possuir um motivo para isso, e o fato de uma das duas respostas ser utilizado como atributo alvo, a outra resposta não foi considerada como atributo descritivo.

Em questão quantitativa, os dados foram separados em 70% para o conjunto de treinamento e 30% para o conjunto de validação, sendo a separação de forma aleatória.

## 4. Discussão dos Resultados

Para a implementação em Python dos métodos de aprendizado de máquina selecionados, foi utilizada a biblioteca Scikit-learn, mas especificamente as classes: *MLPClassifier* para redes neurais artificiais; *SVC (Support Vector Classifier)* para máquinas de vetor de suporte; *BernoulliNB*, *MultinomialNB* e *GaussianNB* para classificador naive bayes, utilizando os algoritmos de classificação de Bernoulli, multinomial e gaussiana, respectivamente; e *DecisionTreeClassifier* para árvores de decisão.

O procedimento experimental aplicado consistiu nas seguintes etapas:

1. Separação dos atributos descritivos e atributo alvo;
2. Separação dos conjuntos de treinamento e teste;
3. Construção dos modelos de aprendizado de máquina utilizando variação nos seus parâmetros de inicialização;
4. Treinamento e teste dos modelos;
5. Avaliação e documentação dos resultados.

Tendo em vista o desbalanceamento na quantidade de exemplos entre as classes presentes no atributo alvo, utilizar a acurácia como métrica de resultado poderia gerar resultados inconsistentes. Sendo assim, a métrica utilizada foi a *Medida F* ou *F1-score*, onde é levado em consideração um equilíbrio entre a precisão e a revocação dos resultados, sendo está uma métrica menos suscetível a desequilíbrio nos resultados dado o desbalanceamento no atributo alvo.

#### 4.1. Redes Neurais Artificiais

Foram geradas 25.920 configurações de redes neurais, variando a quantidade de camadas internas em [1, 2], com quantidade de neurônios entre [2, 11]. Dentre as demais variações, a configuração que obteve melhor resultado foi a composta pelos seguintes parâmetros:

activation: identity	hidden_layer_sizes: (2,)	power_t: 0.5
alpha: 0.0001	learning_rate: adaptive	random_state: None
batch_size: auto	learning_rate_init: 0.003	shuffle: True
beta_1: 0.9	max_iter: 200	solver: lbfgs
beta_2: 0.999	momentum: 0.9	tol: 0.0001
early_stopping: True	n_iter_no_change: 10	validation_fraction: 0.1
epsilon: 1e-08;	nesterovs_momentum: True	verbose: True
		warm_start: False

Tal configuração alcançou *F1-score-micro* de 73,77%.

#### 4.2. Classificador Naive Bayes

Tendo em vista uma menor quantidade de parâmetros configuráveis para efetuar variações, foram gerados apenas 76 configurações diferentes de classificadores naive bayes. A configuração com melhor resultado utilizou o algoritmo de Bernoulli, alcançando *F1-score-micro* de 75,41%, tendo sido inicializada com os seguintes parâmetros:

alpha: 0.0 | binarize: 1.0 | class\_prior: None | fit\_prior: True

#### 4.3. Árvore de Decisão

Foram gerados um total de 2.400 configurações diferentes de árvores de decisão a partir da variação dos parâmetros de:

- *criterion*: Função para medir a qualidade de uma divisão;
- *splitter*: Estratégia a ser usada para divisão em cada nó;
- *min\_samples\_split*: Número mínimo de amostras necessárias para dividir um nó interno;
- *min\_samples\_leaf*: Número mínimo de amostras necessárias para estar em um nó folha;
- *min\_weight\_fraction\_leaf*: A fração ponderada mínima da soma total dos pesos exigida para o nó ser dividido em folhas; e
- *max\_features*: Número de recursos a serem considerados ao procurar a melhor divisão;

Os demais parâmetros configuráveis foram inicializados com os valores *default* da biblioteca.

Dentro os métodos de aprendizado de máquina utilizados, as árvores de decisão obtiveram os melhores resultados, alcançando *F1-score-micro* de 85,24% com a seguinte configuração:

A Figura 3 apresenta o gráfico resultante da configuração com o melhor resultado dentre as árvores de decisão geradas.



class_weight: None	max_leaf_nodes: None	min_samples_split: 9
criterion: gini	min_impurity_decrease: 0.0	min_weight_fraction_leaf: 0
max_depth: None	min_impurity_split: None	presort: False
max_features: sqrt	min_samples_leaf: 5	random_state: None
		splitter: best

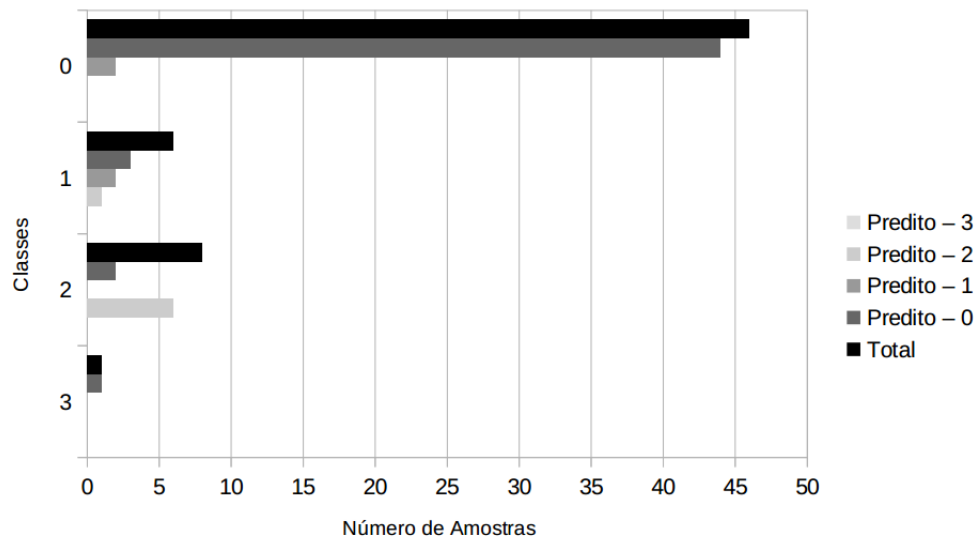


Figure 3. Resultado da melhor árvore de decisão.

#### 4.4. Comparação com os trabalhos relacionados

Os trabalhos de [Manhães et al. 2012] e [Martinho et al. 2017] utilizando como métrica a acurácia, que considera as taxas de acerto direto, desprezando a diferença entre as classes de origem das predições e o desbalanceamento dos dados. Sendo assim, os trabalhos foram utilizados para direcionamento teórico, não havendo condições técnicas de comparação entre a acurácia utilizando por estes e o *F1-score* da presente pesquisa.

Quando comparados os resultados aqui obtidos com a pesquisa realizada por [Rovira et al. 2017], é possível observar que, utilizando dados relacionados a causa dos problemas acadêmicos dos alunos, ao invés das notas obtidas por estes nas matérias da graduação, a presente pesquisa atingiu um *F1-score* de 85,24%, contra 82% obtido por [Rovira et al. 2017].

#### 5. Considerações Finais

Neste projeto foram aplicados métodos de aprendizagem de máquina baseados no paradigma de aprendizado supervisionado para predição de motivos que levam alunos do ensino superior a se afastarem temporariamente ou definitivamente do curso de graduação em IES privadas, aplicando tal conhecimento no escopo da FUCAPI.

Foram obtidos, após a fase de limpeza de dados, 201 exemplos úteis para os experimentos. Dentre os três métodos de aprendizagem de máquina elencados para a solução do problema, as árvores de decisão apresentaram o melhor resultado, alcançando *F1-score-micro* de 85,24%, contra 75,41% do classificador naive bayes e 73,77% das redes neurais artificiais.

É importante levar em consideração que os tipos de dados utilizados foram obtidos através de respostas livres de estudantes de IES que colaboraram com este projeto, portanto, uma base de dados oriunda de sistemas internos das IES, considerando notas, reprovações e porcentagem de presença nas matérias cursadas, podem fornecer conhecimento de diferente nível de relevância para o problema, podendo apresentar resultados diferentes dos aqui apresentados.

## References

- da Costa, D. D. and de Barros Ferreira, N. I. (2016). O prouni na educação superior brasileira: indicadores do acesso e permanência, no brasil (2005-2010). *Avaliação: Revista da Avaliação da Educação Superior*, 22(1):141–163.
- de Pádua Braga, A., de Leon Ferreira de Carvalho, A. P., and Ludermir, T. B. (2000). *Redes neurais artificiais: teoria e aplicações*. Livros Técnicos e Científicos.
- Faceli, F., Lorena, A. C., Gama, J., and de Leon Ferreira de Carvalho, A. C. P. (2011). *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro, RJ. ISBN 978-85-216-1880-5.
- INEP (2018). *Censo da educação superior: Notas estatísticas 2017*.
- Manhães, L. M. B., Da Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2012). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)*, volume 1.
- Martinho, V. R., Nunes, C., and Minussi, C. R. (2017). Predição do grupo de risco de evasão discente em cursos superiores presenciais utilizando uma rede neural artmap-fuzzy.
- Monard, M. C. and Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32.
- Norte, D. B. (2018). Apesar das falhas, fies e prouni melhoraram o acesso ao ensino superior. Disponível em <https://revistaensinosuperior.com.br/fies-e-prouni-expandiram-matriculas/>.
- Pardo, T. A. S. and das Graças Volpe Nunes, M. (2002). Aprendizado bayesiano aplicado ao processamento de línguas naturais. *Série de Relatórios Técnicos do Instituto de Ciências Matemáticas e de Computação – ICMC, Universidade de São Paulo*, (180).
- Rovira, S., Puertas, E., and Igual, L. (2017). Data-driven system to predict academic grades and dropout. *PLoS one*, 12(2):e0171207.
- Souza, D. (2017). Número de alunos pobres em faculdades privadas cresce 20%. Disponível em <https://educacao.estadao.com.br/noticias/geral,numero-de-alunos-pobres-em-faculdades-privadas-cresce-20,70001741988>.