

# Aplicação de uma metodologia de previsão de evasão: um estudo de caso no campus UFV-Florestal

Lucas M. Souza<sup>1</sup>, Pedro C. C. Mundim<sup>1</sup>, Daniel M. Barbosa<sup>1</sup>,  
Gláucia Braga e Silva<sup>1</sup>, Thais Regina de M. B. Silva<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa (UFV)  
Rodovia LMG 818, km 06 – 35690-000 – Florestal – MG – Brazil

{lucas.souza59, pedro.mundim, danielmendes, glaucia, thais.braga}@ufv.br

**Abstract.** *Dropout in higher education is a global challenge, making it essential to identify students who are likely to leave their courses and when this is likely to happen. This study, inspired by the methodology of a reference article, adopts two approaches applied to a dataset from a Brazilian university: machine learning models to predict who will drop out, and survival analysis to estimate when it will occur. The machine learning results were similar to those of the reference article. The survival analyses showed some differences, but maintained similar trends. The findings reinforce the validity of the reference article's methods in a new context, contributing to the generalization of the results.*

**Resumo.** *A evasão no ensino superior é um desafio global, tornando essencial a identificação de alunos propensos a abandonar o curso e o momento em que isso ocorre. Este estudo, inspirado na metodologia de um artigo de referência, adota duas abordagens aplicadas a uma base de dados de uma universidade brasileira: modelos de aprendizado de máquina para prever quem irá evadir e análise de sobrevivência para estimar quando isso ocorrerá. Os resultados do aprendizado de máquina foram semelhantes aos do artigo base. Já as análises de sobrevivência apresentaram algumas diferenças, porém mantendo tendências similares. Os achados reforçam a validade dos métodos do artigo base em um novo contexto, contribuindo para a generalização dos resultados.*

## 1. Introdução

A evasão no ensino superior impacta não apenas as instituições de ensino, mas também os estudantes e a sociedade como um todo [Da Cruz et al. 2023], tornando-se um desafio para o desenvolvimento socioeconômico dos países [Duran et al. 2023]. Estudos como [Silva et al. 2023] analisam a evasão de forma geral e [Duran et al. 2023] analisam a evasão especificamente nas áreas de tecnologia, ambos associando-a a diversos atributos, incluindo aspectos acadêmicos, socioeconômicos, étnico-raciais, de gênero e emocionais, tanto internos quanto externos às instituições.

Outros estudos têm buscado identificar quais atributos levam à evasão, bem como os métodos e modelos preditivos mais eficazes para a identificação de alunos em situação de possível evasão [Santos et al. 2019]. Em [de Oliveira and Barbosa 2023], é possível observar que tais quesitos tornaram-se temas de pesquisa muito relevantes e recorrentes, demonstrando a preocupação e urgência quanto à problemática.

Além dos fatores acadêmicos, socioeconômicos e emocionais já discutidos, é possível englobar ainda o aspecto temporal ao analisar a evasão no ensino superior e utilizar ambos os aspectos para obter uma boa visão sobre a situação, o que possibilita um escopo diverso de abordagens [Campos 2016]. Assim, este trabalho tem como objetivo adotar uma abordagem semelhante ao trabalho de [Pachas et al. 2021] (artigo tomado como base), porém adaptada quanto aos atributos disponíveis neste estudo de caso, para avaliar perspectivas acerca das taxas de evasão na Universidade Federal de Viçosa, campus UFV-Florestal, buscando responder às seguintes perguntas de pesquisa: (1) Quais os atributos relevantes para determinar a evasão dos alunos?; (2) Quais alunos possuem chance de evadir?; (3) Quanto tempo leva até os alunos evadirem?

Para responder às questões propostas, técnicas de Aprendizado de Máquina (ML) e de Análise de Sobrevivência foram utilizadas, com o objetivo de oferecer um panorama geral da evasão no campus UFV-Florestal. O estudo busca demonstrar e comparar o desempenho de algoritmos de ML, identificar os momentos nos quais ocorrem mais evasões e analisar o desenvolvimento desse quadro ao longo da graduação. Além disso, os resultados obtidos foram comparados com os de [Pachas et al. 2021], a fim de identificar possíveis tendências.

Este trabalho contribui com uma investigação dos atributos que mais influenciam o risco de evasão, com base nos dados do campus UFV-Florestal; uma análise comparativa dos modelos de classificação para identificar e prever alunos em risco de evasão; uma análise de sobrevivência para estimar quando isso ocorrerá; e um estudo comparativo ao artigo de [Pachas et al. 2021].

O artigo está organizado de forma que: a Seção 2 traz os trabalhos relacionados; a Seção 3 aborda os materiais e métodos utilizados; a Seção 4 traz os resultados; e a Seção 5 traz as considerações finais.

## 2. Trabalhos Relacionados

Em [Silva et al. 2023], foi realizada uma revisão sistemática da literatura (RSL) com o objetivo de identificar os principais atributos associados à predição de abandono no ensino superior, com base em estudos primários. Os autores utilizaram as bases de dados SCOPUS e IEEE, selecionando 52 trabalhos para análise. Os resultados indicaram que aproximadamente 29 atributos foram considerados, os quais foram organizados em demográficos, acadêmicos e de aprendizagem, destacando-se o desempenho acadêmico, gênero, etnia, local de residência e base de conhecimento prévia como os mais relevantes.

Já o estudo de [Opazo et al. 2021] tem como objetivo comparar o desempenho de modelos de Aprendizado de Máquina na previsão da evasão de alunos do primeiro ano de engenharia em duas universidades chilenas. Os resultados indicam que os modelos de Aprendizado de Máquina podem ser eficazes na identificação de atributos relacionados à evasão, sendo a aplicação de múltiplos modelos útil para detectar atributos relevantes e analisar tendências. A pesquisa ainda destaca que o rendimento médio e as preferências dos alunos são atributos importantes na previsão da evasão.

O trabalho de [Hoed et al. 2018] analisa as taxas de evasão em cursos de graduação em Computação no Brasil, utilizando dados do INEP e um estudo de caso na Universidade de Brasília (UnB), com o objetivo de verificar a influência de atributos como

abstração algorítmica, número de candidatos por vaga e gênero dos alunos. Para isso, questionários foram aplicados aos alunos desligados dos cursos para identificar motivos específicos de evasão. Os resultados destacam a dificuldade com abstração algorítmica e conhecimento matemático como atributos-chave para a evasão, além de evidências de que as taxas de abandono variam entre os gêneros.

Além do aprendizado de máquina, uma outra abordagem que pode ser utilizada para analisar a evasão é o aspecto temporal por meio de técnicas de Análise de Sobrevivência. Os autores de [Lima Junior et al. 2012] utilizam essas técnicas aplicadas em um estudo sobre o tempo de permanência de estudantes em um curso de graduação em Física da Universidade Federal do Rio Grande do Sul (UFRGS). Como resultado desta pesquisa, constatou-se que a análise de sobrevivência, a qual é originalmente atrelada às áreas da saúde, é adequada para descrever quando a evasão e a diplomação ocorrem, bem como quais são os atributos relacionados à permanência a longo prazo.

Por fim, ao combinar ambas as técnicas, [Pachas et al. 2021] tem como objetivo prever tanto os alunos que abandonam quanto o momento em que isso ocorre na Universidade Católica de San Pablo (UCSP). Por meio de técnicas de Aprendizado de Máquina e Análise de Sobrevivência, a pesquisa analisa os atributos que contribuem para a evasão e o tempo até o abandono. Os resultados indicam que o número de reprovações é um fator determinante no risco de evasão, enquanto, ao contrário de outros estudos, o desempenho acadêmico e o gênero não se mostraram significativos para a previsão do abandono.

Outros estudos como [Rodrigues et al. 2024], [Sousa et al. 2024] e [Correia et al. 2024] abordam o estudo da evasão em outras universidades. Os resultados mostram, de forma geral, nesses trabalhos, que a evasão acontece principalmente nos primeiros períodos do curso e apontam o rendimento acadêmico como um dos principais fatores para o abandono.

No presente trabalho, adota-se uma metodologia análoga à de [Pachas et al. 2021], com o objetivo de identificar e entender os atributos que levam à evasão e o quadro ao longo do tempo no cenário do campus UFV-Florestal. Diferente dos trabalhos de [Opazo et al. 2021] que aplicam Aprendizado de Máquina e [Lima Junior et al. 2012] que aplica Análises de Sobrevivência, este estudo trabalha com as 2 abordagens, permitindo uma análise mais ampla do problema. Além disso, é explorado o percentual da evasão em cursos variados, buscando um quadro mais geral, diferentemente do apresentado por [Hoed et al. 2018] que foca na Computação.

### **3. Materiais e Métodos**

Nesta seção, será explicada a forma de coleta dos dados dos discentes, incluindo os métodos, ferramentas e recursos computacionais utilizados.

#### **3.1. Coleta de Dados**

Os dados utilizados neste estudo são da base acadêmica da UFV-Florestal, a qual abrange informações demográficas e acadêmicas de 5.892 alunos, de seus 10 cursos superiores: Agronomia, Administração, Ciência da Computação, Engenharia de Alimentos, Gestão Ambiental e os cursos de Licenciatura em Ciências Biológicas, Educação Física, Física, Matemática e Química. O período considerado na base vai desde o primeiro semestre de 2008 até o primeiro semestre de 2024. A base contém 20 atributos, os quais estão

detalhados na Tabela 1. Ela foi explorada e limpa, retirando-se algumas células para evitar influenciar de maneira errônea as análises.

**Tabela 1. Atributos da base de dados.**

Nº	Atributo
1	Identificador
2	Ano de Nascimento
3	Gênero
4	Campus
5	Identificador do Curso
6	Curso
7	Estado de Nascimento
8	Semestre de admissão
9	Semestre de saída
10	Código da situação acadêmica do aluno
11	Situação acadêmica do aluno
12	Situação acadêmica agrupada
13	Modalidade inscrita no ENEM
14	Nota no ENEM
15	Número de reprovações
16	Raça
17	Coeficiente de rendimento acadêmico
18	Município de endereço do aluno ao final
19	Estado de endereço do aluno ao final
20	Município de nascimento

O trabalho de [Pachas et al. 2021] utilizou 7 atributos; neste trabalho, devido à grande gama de atributos contida na base da UFV-Florestal e visando utilizar ao máximo esses dados, foram explorados todos os que se mostraram relevantes para a análise, sendo importante ressaltar que nem todos os atributos utilizados por [Pachas et al. 2021] foram possíveis de ser utilizados, pois não há um semelhante na base da UFV-Florestal.

### 3.2. Análise de Dados

Para a realização das análises deste trabalho, foi utilizada a plataforma Jupyter em conjunto com a linguagem de programação Python, além de algumas de suas bibliotecas, sendo elas: matplotlib<sup>1</sup>, seaborn<sup>2</sup>, pandas<sup>3</sup>, NumPy<sup>4</sup>, scikit-learn<sup>5</sup> e lifelines<sup>6</sup>. Além disso, a metodologia de análise dos dados seguiu proposta no artigo de [Pachas et al. 2021], tendo como variável-alvo **situação aluno agrupada**, que resume as situações dos alunos em 4 categorias, sendo elas: evasão, conclusão, matriculado e retenção. Para este trabalho, essa variável-alvo foi categorizada em **evasão** e **não evasão**, sendo que a categoria **não evasão** abrange as outras três situações.

#### 3.2.1. Aprendizado de Máquina

Para responder às duas primeiras questões de pesquisa, foi realizada uma análise exploratória e uma engenharia de atributos, com foco no uso de algoritmos de aprendizado de

<sup>1</sup>matplotlib: <https://matplotlib.org>

<sup>2</sup>seaborn: <https://seaborn.pydata.org>

<sup>3</sup>pandas: <https://pandas.pydata.org>

<sup>4</sup>NumPy: <https://numpy.org/>

<sup>5</sup>scikit-learn: <https://scikit-learn.org/stable/>

<sup>6</sup>lifelines: <https://lifelines.readthedocs.io/en/latest/>

máquina. A estratégia envolveu a exploração dos atributos disponíveis na base de dados, buscando aprimorar a previsão dos algoritmos. Inicialmente, foram removidos os dados de alunos que possuíam algum atributo nulo, desde que a quantidade desses casos fosse consideravelmente pequena em comparação ao total de alunos da base. Atributos completamente nulos também foram descartados. Para os casos de valores ausentes nas notas do ENEM, a mediana foi utilizada para preencherê-los. Além disso, valores nominais foram mapeados em valores numéricos. Após todas essas transformações, foram considerados os atributos mostrados na Tabela 2.

**Tabela 2. Atributos para aprendizado de máquina.**

Nº	Atributo
1	gênero
2	Curso
3	Estado de nascimento
4	Semestre de admissão
5	Semestre de saída
6	Curso
7	Situação do aluno agrupada
8	Modalidade inscrita no ENEM
9	Nota do ENEM
10	Número de reprovações
11	Raça
12	Coeficiente de Rendimento
13	Região de nascimento
14	Ano Nascimento

Já os algoritmos de aprendizado de máquina utilizados neste trabalho foram: Regressão Logística (LR), Máquina de Vetor de Suporte (SVM), Naive Bayes (NB), Árvore de Decisão (DT) e Floresta Aleatória (RF). Para avaliar a capacidade preditiva dos algoritmos, foram utilizadas métricas usuais da literatura (acurácia, precisão, revocação, F1-score e AUC).

### 3.2.2. Análise de Sobrevidência

Para responder à terceira questão de pesquisa, ou seja, quando um aluno irá evadir, foram aplicadas análises de sobrevidência, uma vez que esse método lida diretamente com a dimensão temporal, representada pela variável  $T$ , que indica o tempo até o evento. A variável  $Y$ , por sua vez, sinaliza se o aluno evadiu ou não. Dessa forma, a análise de sobrevidência gera como resultado uma variável que representa o “tempo até a ocorrência de um evento”, com a função de probabilidade de sobrevidência sendo definida por:

$$S(t) = \text{Prob}(T > t) \quad (1)$$

que representa a probabilidade de um estudante não ter evadido até o tempo  $t$ , conforme [Pachas et al. 2021]. Além disso, a função de risco  $h(t)$  que computa a probabilidade da evasão de um estudante ocorrer no tempo  $t$  é dada por:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{\text{Prob}(t \leq T \leq t + \delta t \mid t \leq T)}{\delta t}, \quad (2)$$

Para otimizar o uso dos métodos de Análise de Sobrevidência, foi realizada uma segunda etapa de engenharia de atributos na base original. Foram selecionados os atributos mais relevantes para esse tipo de análise, mantendo os numéricos e que permitissem

explorar áreas específicas. Para representar os cursos, aplicou-se *One-Hot Encoding*, criando um atributo binário para cada um dos 10 cursos, indicando a matrícula do aluno como (1) ou sua ausência como (0). As situações dos alunos foram agrupadas em evasão e não evasão. Por fim, foi criado um atributo temporal, representando o número de semestres, calculado pela subtração do semestre de saída pelo semestre de entrada, após ambos serem convertidos para números inteiros. Os atributos finais são mostrados na Tabela 3.

**Tabela 3. Atributos para análise de sobrevivência.**

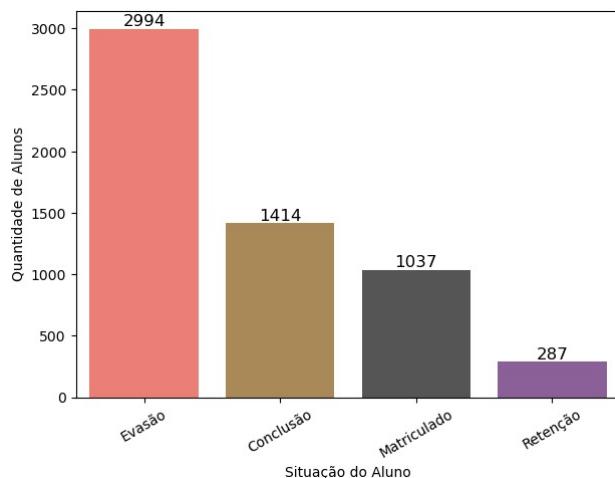
indicador	Atributo
X1	Nota do ENEM
X2	Número de reprovações
X3	Coeficiente de rendimento
X4	Gênero
X5	Cursos
Y	Evasão
T	Número de semestres

Neste trabalho, foram empregados os estimadores não paramétricos Kaplan-Meier (KM-estimator) e Nelson-Aalen (Na-estimator) conforme [Pachas et al. 2021] para estimar, respectivamente, a função de sobrevivência  $S(t)$  e a função de risco acumulado  $H(t)$ . Além disso, foi utilizado o modelo semiparamétrico de Cox para avaliar a significância dos atributos.

## 4. Resultados

Nesta seção, serão apresentados os resultados obtidos das duas abordagens apresentadas, bem como uma visão exploratória geral da base de dados do campus UFV-Florestal.

Para uma visão panorâmica inicial, ao explorar a base de dados, foi observado que mais da metade dos alunos que ingressaram na instituição no período analisado evadiram, como mostrado na Figura 1.



**Figura 1. Número de estudantes por situação.**

### 4.1. Abordagem 1 - Aprendizado de Máquina

Para prever quais estudantes podem evadir, utilizou-se os cinco algoritmos de aprendizado de máquina citados na Seção 3. Para a avaliação desses algoritmos, foram utilizadas

as métricas mencionadas também na Seção 3. Cada algoritmo foi executado cinco vezes, utilizando 70% da amostra para treinamento e 30% para teste. Os resultados dessas execuções estão apresentados na Tabela 4, que apresenta as médias das métricas para LR, SVM, NB, DT e RF, respectivamente.

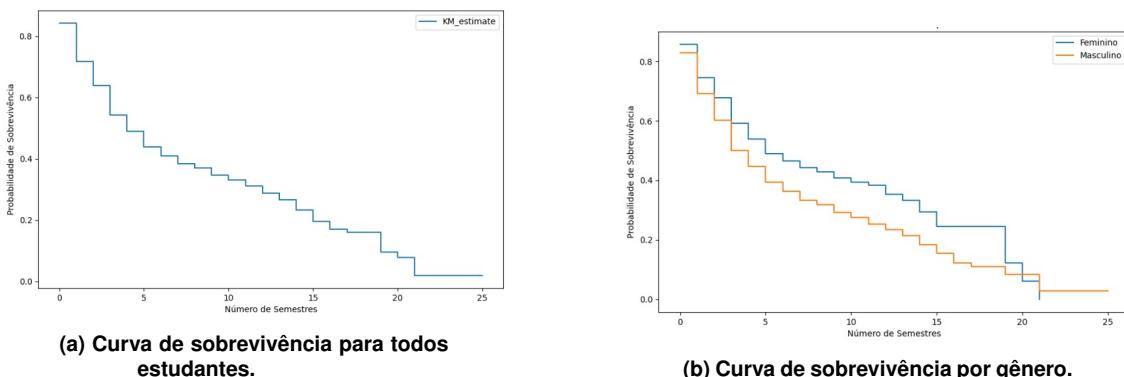
**Tabela 4. Média dos resultados**

Métodos	Acurácia	Precisão	Revocação	F1	AUC
LR	0.876	0.877	0.876	0.875	0.960
SVM	0.882	0.885	0.882	0.882	0.962
NB	0.856	0.874	0.856	0.859	0.957
DT	0.916	0.916	0.916	0.916	0.926
<b>RF</b>	<b>0.940</b>	<b>0.942</b>	<b>0.940</b>	<b>0.941</b>	<b>0.987</b>

Como é possível observar, RF apresentou os melhores valores para todas as métricas avaliadas, similarmente ao apresentado em [Pachas et al. 2021]. Dessa forma, pode-se afirmar que RF foi o modelo que classificou corretamente a maior proporção das instâncias devido à sua acurácia. Além disso, a precisão indica que previu corretamente a maioria das instâncias que associou à classe positiva. A revocação, por sua vez, mostra que o modelo identificou melhor as instâncias da classe positiva. O F1-score indica um bom equilíbrio entre precisão e revocação, reforçando um desempenho consistente. O valor de AUC revela que o modelo apresentou uma capacidade excelente de discriminar as classes, indicando uma boa capacidade de classificar corretamente uma instância aleatória.

#### 4.2. Abordagem 2 - Análise de Sobrevida

Para as análises de sobrevida, foi usado o estimador Kaplan-Meier (KM-estimator) para estimar as curvas de sobrevida para todos os estudantes e agrupadas por gênero, apresentadas na Figura 2a e na Figura 2b, respectivamente.

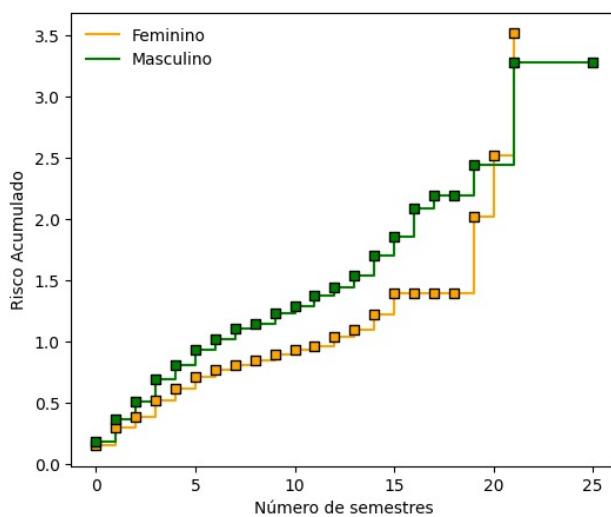


**Figura 2. Comparação das curvas de sobrevida geral e por gênero**

A Figura 2a mostra que a taxa de evasão é mais acentuada nos primeiros semestres, tornando-se significativamente menor a partir do 5º semestre, fato evidenciado devido às quedas das curvas nos gráficos, as quais representam a escala da ocorrência do evento observado (evasão). De forma semelhante, a Figura 2b mostra que, independentemente do gênero, a taxa de evasão segue o mesmo padrão da curva geral. No entanto, observa-se que a curva referente ao gênero masculino permanece consistentemente abaixo da feminina ao

longo da maior parte do período analisado, sugerindo uma possível maior taxa de evasão para o gênero masculino.

Além disso, por meio do estimador Nelson-Aalen, computou-se as curvas da função de risco acumulado para o gênero, conforme a Figura 3, a qual, juntamente com o teste Log-Rank [Saccaro et al. 2019] indica uma diferença estatisticamente significativa entre os grupos, com a hipótese nula  $H_0: h_{\text{feminino}}(t) = h_{\text{masculino}}(t)$  sendo rejeitada, dado que o p-valor obtido é inferior a qualquer nível de significância usual. Isso sugere que, ao contrário dos resultados de [Pachas et al. 2021], o gênero tem, de fato, influência sobre o risco de evasão, sendo esse risco maior para os alunos do gênero masculino.



**Figura 3. Curva de função de risco acumulado.**

Adicionalmente, foi analisada a função de sobrevivência para cada curso ao longo de 8 semestres. Os resultados, apresentados na Tabela 5, indicam uma taxa de evasão mais elevada nos primeiros semestres para todos os cursos. Além disso, nos valores referentes aos períodos iniciais, observa-se uma maior taxa de evasão nos cursos de Licenciatura em Química, Matemática, Física, Gestão Ambiental e Engenharia de Alimentos. É relevante destacar que o curso de Ciência da Computação apresentou uma taxa de evasão inferior à dos cursos de Licenciatura, podendo ser evidenciada ao se observar que a função de sobrevivência do curso de Computação no 8º semestre é maior que a do curso de Física já no 3º semestre. Isso é especialmente significativo, considerando que a área de Computação, de maneira geral, registra índices elevados de evasão. Conforme evidenciado por [Mundim et al. 2024], estratégias de acolhimento/incentivo, como as utilizadas neste curso de Ciência da Computação, podem desempenhar um papel fundamental na mitigação desse problema, especialmente nos primeiros períodos, contribuindo diretamente para a redução dessas taxas.

Com base nos gráficos 2a, 2b e na Tabela 5, esse padrão de evasão precoce é consistente com os resultados observados na UCSP analisada por [Pachas et al. 2021].

Para concluir as análises, utilizou-se o Modelo de Cox. Inicialmente, os parâmetros do modelo foram resumidos, considerando um  $L_1$ -ratio igual a 1 e um penalizador de 0.0005. O índice de concordância obtido foi de 0.85, indicando uma boa capacidade preditiva do modelo. Os resultados estão apresentados na Tabela 6.

**Tabela 5. Função de sobrevivência por curso**

curso	1	2	3	4	5	6	7	8
Administração	0.740	0.651	0.576	0.535	0.502	0.465	0.447	0.440
Agronomia	0.810	0.748	0.656	0.608	0.530	0.519	0.491	0.480
Ciência da Computação	0.747	0.687	0.555	0.495	0.439	0.407	0.378	0.356
Engenharia de Alimentos	0.687	0.620	0.496	0.441	0.376	0.347	0.335	0.314
L.Ciências Biológicas	0.714	0.649	0.590	0.532	0.494	0.473	0.445	0.438
L.Educação Física	0.749	0.669	0.581	0.525	0.497	0.466	0.423	0.400
L.Física	0.598	0.482	0.353	0.304	0.235	0.216	0.187	0.174
L.Matemática	0.592	0.508	0.393	0.351	0.325	0.312	0.302	0.302
L.Química	0.676	0.592	0.488	0.397	0.323	0.295	0.259	0.244
Gestão Ambiental	0.673	0.584	0.520	0.482	0.435	0.367	0.340	0.301

Cursos com “L.” são cursos de licenciatura

atributo	coef	exp(coef)	coef(lower)	95%	coef(upper)	95%	p-value
Gênero	0.06	1.06	-0.02		0.13		0.13
ENEM	0.00	1.00	0.00		0.00		0.02
Num_Reprovacoes	-0.04	0.96	-0.04		-0.03		< 0.005
CRA	-0.05	0.95	-0.05		-0.05		< 0.005

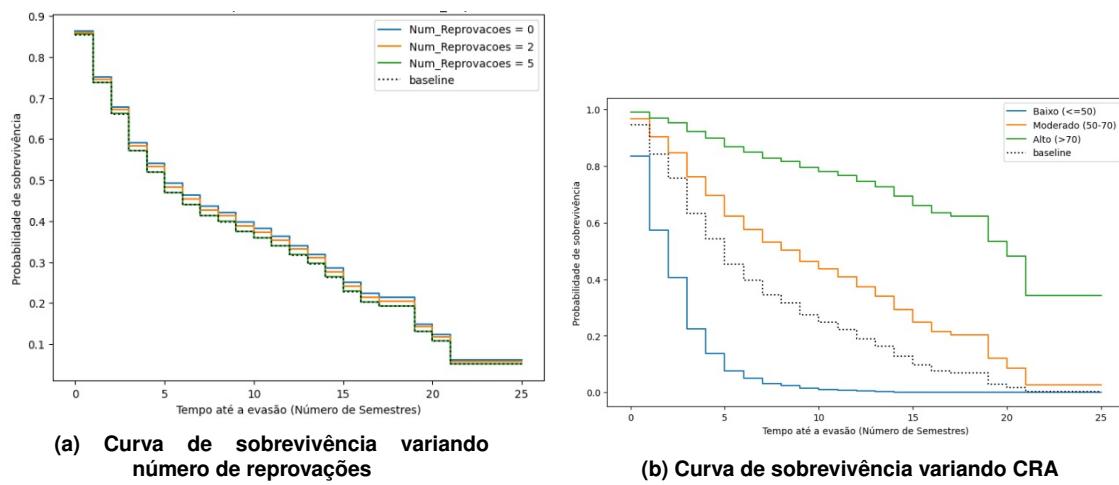
**Tabela 6. Sumarização dos parâmetros do modelo Cox**

De acordo com a Tabela 6, observa-se que as variáveis estatisticamente significativas para a previsão da evasão são o número de reprovações e o CRA (índice de rendimento), uma vez que seus p-valores são inferiores a 0.005. Ao analisar os demais atributos e seus respectivos p-valores, conclui-se que não há evidências estatísticas fortes que indiquem influência significativa na previsão da evasão.

A partir da análise da Tabela 6 e da Figura 4a, observa-se que o número de reprovações totais ao longo do curso não parece exercer um impacto significativo na probabilidade de evasão devido à proximidade das curvas, apesar do p-valor apresentado indicar uma significância estatística. No estudo de [Pachas et al. 2021], essa variável não foi considerada. Em seu lugar, foi analisado o número de reprovações por semestre, uma métrica semelhante; porém, em [Pachas et al. 2021], apresentou maior relevância para a previsão de evasão. Em contrapartida, os autores de [Pachas et al. 2021] destacam que o desempenho do estudante não foi um fator relevante para a instituição por eles estudada, enquanto que, neste trabalho, para o campus UFV-Florestal, o desempenho medido pelo CRA se mostrou um atributo importante, como evidenciado pela Figura 4b.

## 5. Considerações Finais

Este trabalho utilizou uma metodologia para analisar o quadro de evasão do campus UFV-Florestal com o intuito de determinar os atributos relevantes para a previsão de evasão, prever qual aluno irá evadir e quando isso acontecerá. Após analisar os resultados, nota-se que, diferentemente de [Pachas et al. 2021] e semelhante a outros trabalhos, rendimento e gênero se mostraram bem relevantes. Além disso, as reprovações também se mostraram um fator determinante neste trabalho e em [Pachas et al. 2021], ressaltando esses atributos como pontos centrais da questão da evasão. Quanto a prever os alunos que irão evadir, os modelos analisados apresentaram boas métricas, destacando-se o algoritmo de Floresta Aleatória para prever quem irá evadir. Já para a determinação de quando irá evadir, foi



**Figura 4. Comparação das curvas de sobrevivência para diferentes fatores**

observada uma tendência maior de evasão nos primeiros semestres para todos os casos analisados. Além disso, o modelo Cox revelou o número de reprovações e o CRA como as variáveis mais significativas para prever evasão. O teste Log-Rank confirmou o gênero como um fator decisivo, ressaltando uma maior tendência de evasão do gênero masculino.

Adicionalmente, ao analisar os resultados dos cursos, nota-se uma taxa de evasão melhor do que o esperado para o curso de computação, possibilitando a investigação dos motivos por trás destas taxas, sendo uma das possíveis causas as estratégias de acolhimento do curso. Com estes resultados, tem-se um panorama geral do quadro de evasão no campus UFV-Florestal, podendo ser utilizado por ela e por outras instituições para a identificação da necessidade de se realizarem medidas preventivas. Finalmente, os achados deste trabalho também reforçam a validade dos métodos de [Pachas et al. 2021] em um novo contexto, contribuindo para a generalização dos resultados do mesmo.

É importante destacar ainda que este trabalho utilizou apenas dados acadêmicos e demográficos, além de considerar um período pandêmico com todas as suas particularidades, obtendo-se assim bons resultados nas métricas dos modelos e identificando possíveis causas para a evasão no campus UFV-Florestal. No entanto, limitou-se a uma gama específica de variáveis, não abrangendo todos os fatores que influenciam a evasão. Trabalhos futuros podem explorar dados econômicos e psicológicos, como renda familiar, custos indiretos, distância da cidade natal, saúde mental e problemas de adaptação, ampliando a robustez dos modelos e melhorando a precisão das previsões. Além disso, novas metodologias e técnicas podem ser aplicadas para aprimorar a análise e identificar atributos adicionais relacionados à evasão, contribuindo para abordagens mais eficazes no enfrentamento do problema.

## Referências

- Campos, J. D. d. S. (2016). Fatores explicativos para a evasão no ensino superior através da análise de sobrevivência: o caso da ufpe. Dissertação de mestrado, Universidade Federal de Pernambuco.
- Correia, R., Mendonça, H., Silva, C., and Toledo, D. (2024). Análise dos principais fatores que influenciam a evasão no ensino superior utilizando técnicas de mineração

- de dados educacionais. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 830–841, Porto Alegre, RS, Brasil. SBC.
- Da Cruz, R. C., Juliano, R. C., Monteiro Souza, F. C., and Correa Souza, A. C. (2023). A score approach to identify the risk of students dropout: an experiment with information systems course. In *Proceedings of the XIX Brazilian Symposium on Information Systems*, pages 120–127.
- de Oliveira, M. M. and Barbosa, E. F. (2023). Multilevel modeling for the analysis and prediction of school dropout: a systematic review. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 103–112. IEEE.
- Duran, R., Bim, S. A., Gimenes, I., Ribeiro, L., and Correia, R. C. M. (2023). Potential factors for retention and intent to drop-out in brazilian computing programs. *ACM Transactions on Computing Education*, 23(3):1–33.
- Hoed, R. M., Ladeira, M., and Leite, L. L. (2018). Influence of algorithmic abstraction and mathematical knowledge on rates of dropout from computing degree courses. *Journal of the Brazilian Computer Society*, 24(1):1–16.
- Lima Junior, P., Silveira, F. L. d., and Ostermann, F. (2012). Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira. *Revista brasileira de ensino de física*, 34:1403.
- Mundim, P., Barbosa, D., Silva, G., and Silva, T. (2024). Impacto da gamificação e de grupos de apoio no ensino-aprendizagem de programação em um curso de ciência da computação. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 419–430, Porto Alegre, RS, Brasil. SBC.
- Opazo, D., Moreno, S., Alvarez-Miranda, E., and Pereira, J. (2021). Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics*, 9(20):2599.
- Pachas, D. A. G., Garcia-Zanabria, G., Cuadros-Vargas, A. J., Camara-Chavez, G., Poco, J., and Gomez-Nieto, E. (2021). A comparative study of who and when prediction approaches for early identification of university students at dropout risk. In *2021 XLVII Latin American Computing Conference (CLEI)*, pages 1–10. IEEE.
- Rodrigues, H., Moraes, L., Santiago, E., Campos, J., Júnior, E. G., Wanderley, G., Garcia, A., Mello, C., Alvares, R., and Santos, R. (2024). Predicting student dropout on the information systems undergraduate program of unirio using decision trees. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 588–598, Porto Alegre, RS, Brasil. SBC.
- Saccaro, A., França, M. T. A., and Jacinto, P. d. A. (2019). Fatores associados à evasão no ensino superior brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de ciência, matemática e computação e de engenharia, produção e construção em instituições públicas e privadas. *Estudos Econômicos (São Paulo)*, 49(2):337–373.
- Santos, K. J. d. O., Menezes, A. G., de Carvalho, A. B., and Montesco, C. A. (2019). Supervised learning in the context of educational data mining to avoid university students dropout. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 207–208. IEEE.

Silva, E. C., Freitas, S., Soares Ramos, C., Muniz De Menezes, A. E., and Rodrigues De Araujo, L. K. S. (2023). A systematic review of the factors that impact the prediction of retention and dropout in higher education.

Sousa, R., Fachini-Gomes, J., Holanda, M., and Leão, M. (2024). Um estudo da evasão no curso de licenciatura em computação da universidade de brasília. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 715–725, Porto Alegre, RS, Brasil. SBC.