

Guia de como elaborar um modelo preditivo para prevenir a evasão de alunos na EPT

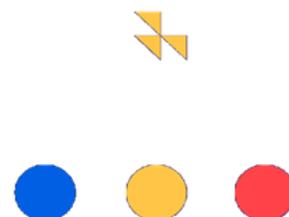


Produto Educacional: Guia de como elaborar um modelo preditivo para prevenir a evasão de alunos na EPT

Autores

Mestrando: Fabrício William da Cunha

Orientador: Prof. Dr. Adalberon Moreira de Lima Filho



Redação

Mestrando: Fabrício William da Cunha

Revisão e edição

Mestrando: Fabrício William da Cunha

Orientador: Prof. Dr. Adalberon Moreira de Lima Filho

Capa e diagramação

Mestrando: Fabrício William da Cunha

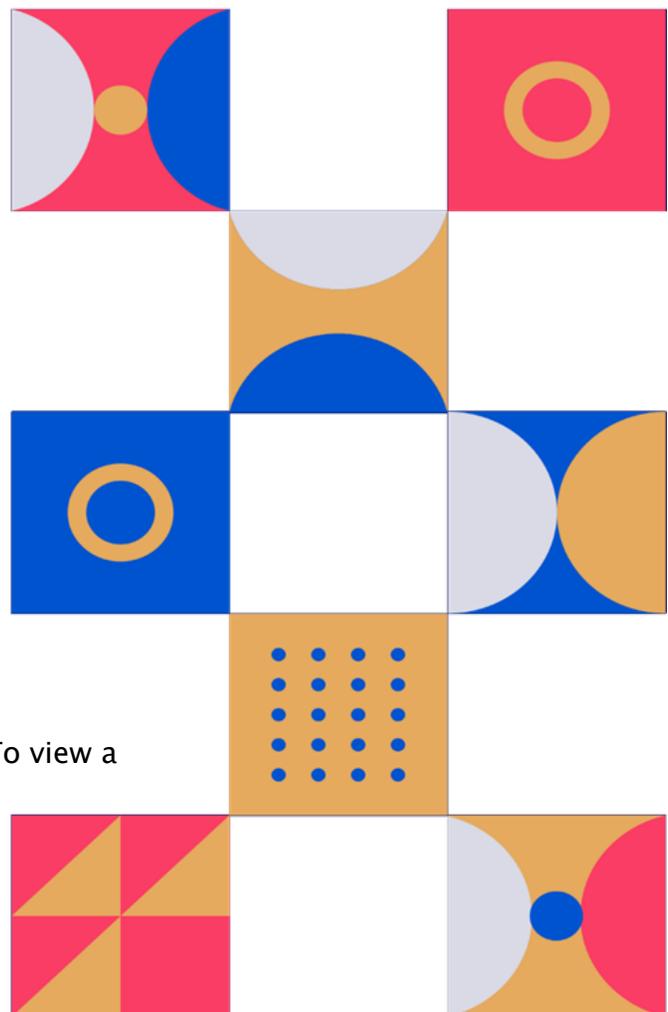
Produção editorial

Mestrando: Fabrício William da Cunha

Orientador: Prof. Dr. Adalberon Moreira de Lima Filho

This work is licensed under the Creative Commons Atribuição–NãoComercial 4.0 Internacional License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc/4.0/>.





**INSTITUTO
FEDERAL**
Alagoas

Dados Internacionais de Catalogação na Publicação
Instituto Federal de Alagoas
Campus Avançado Benedito Bentes
Biblioteca

C972g

Cunha, Fabrício William da.

Guia de como elaborar um modelo preditivo para prevenir a evasão de alunos na EPT / Fabrício William da Cunha, Adalberon Moreira de Lima Filho. – 2023.

26 f. : il.

Produto Educacional da Dissertação: Análise preditiva da evasão escolar no ensino médio integrado do Instituto Federal de Alagoas (Mestrado em Educação Profissional e Tecnológica) Instituto Federal de Alagoas, Campus Avançado Benedito Bentes, Maceió, 2023.

1. Educação. 2. Evasão Escolar. 3. Modelo Preditivo. I. Lima Filho, Adalberon Moreira. II. Título.

CDD: 370

Fernanda Isis Correia da Silva
Bibliotecária - CRB-4/1796

Sumário

1. [Introdução](#)
2. [O que é análise preditiva?](#)
3. [Modelo Preditivo](#)
4. [Como implementar modelo de análise preditiva?](#)
5. [Exemplificação da aplicação do modelo preditivo para prevenir a evasão em um curso técnico](#)

1

Introdução

1. Introdução

Você já ouviu falar em **análise preditiva**? Essa técnica, que já existe há alguns anos, agora vem sendo utilizada para antecipar tendências e fazer projeções no ambiente da gestão escolar.

Ela consiste no uso de dados históricos para **projetar e antecipar tendências**. Com o avanço computacional já é possível **analisar grandes volumes de dados, por meio da estatística, algoritmos e *machine learning*** para encontrar padrões e avaliar as futuras possibilidades a partir dos dados acadêmicos, socioeconômicos, pessoais e geográficos do aluno.

Com um volume de dados disponíveis e uma ampla oferta de softwares acessíveis até mesmo para usuários não técnicos, **a análise preditiva deixou de ser uma tarefa exclusiva de cientistas de dados e estatísticos, sendo incorporada às áreas da gestão escolar para a tomada de decisão em situações como: a prevenção da evasão escolar.**

Neste guia, apresentamos **o conceito de análise preditiva, modelo preditivo, como implementá-la e aplicação prática da análise preditiva na prevenção da evasão escolar nos cursos técnicos integrados ao ensino médio.**

Boa leitura!



**O que é análise
preditiva?**

2. O que é análise preditiva?

A análise preditiva é o modelo mais conhecido de análise, **possibilita a tomada de decisão** com base na “antecipação” do provável resultado da decisão. Ajuda nos fatos que estão acontecendo no momento; **deduzir possíveis cenários que poderiam ocorrer e classificá-los conforme a probabilidade de acontecerem** BARI (2020).

Em geral, o objetivo das suas **aplicações é descobrir, interpretar e comunicar padrões significativos observados nos dados**. A análise preditiva conta, cada vez mais, com segurança e consolidação para descobrir padrões e **avaliar as probabilidades de um resultado ou um acontecimento futuro**.

Para isso essa técnica avançada utiliza tecnologias como a **mineração de dados, *machine learning*, inteligência artificial e estatística** para coletar, processar, interpretar e traduzir os dados, para dar um passo além da estatística ao invés de tirar conclusões sobre ontem, o objetivo é antecipar tendência e prever os comportamentos do amanhã tudo a partir de dados armazenados no registro acadêmico do aluno.

2.1 Atenção aos conceitos!

2.1.1 Mineração de dados

O conceito principal de mineração de dados é a extração de conhecimento útil a partir dos dados coletados para resolver problemas de interesses, com processos sistemáticos unidos às ferramentas computacionais, evoluindo o processo entre etapas previamente definidas. (FAYYAD, 1996).

2.1.2 Machine learning

O *machine learning*, também chamado em português de aprendizado de máquina, é um método de análise de dados e de informações que permite tomar decisões baseadas nas interpretações dos dados.

2.1.3 Inteligência Artificial

Um conjunto de tecnologias baseadas em redes neurais artificiais, algoritmos, sistemas de aprendizado, etc., que conseguem simular capacidades humanas e tornar as máquinas mais inteligentes para resolver problemas.

2.1.4 Estatística

A estatística é um conjunto de métodos e técnicas (ferramentas) que auxiliam a tomada de decisão quando na presença de incerteza. (Gráficos, tabelas, testar hipóteses, prever comportamentos futuro desses dados para então tomar decisão) “é a arte de torturar os dados até que eles confessem a verdade”.

2.2 Como a análise preditiva pode ajudar na evasão escolar?



Na realidade de uma Instituição de Ensino (IE), o gestor ou a equipe podem usar essa abordagem para compreender e melhorar o processo de tomada de decisão. Prever a evasão consiste em identificar os sinais que antecedem a saída do aluno antes da conclusão do curso, por meio de algoritmos que calculam a probabilidade de acordo com cada situação.

2.3 Benefícios da análise preditiva

- ❖ A análise preditiva pode ajudá-lo a tomar medidas proativas para permanecer no caminho certo;
- ❖ Identificar causas da evasão, perfil do alunado que evade no curso;
- ❖ Ações que aumentem os laços dos alunos com as IÉ;
- ❖ Identificar com antecedência alunos com probabilidade de evadir, evitando assim, futuros abandonos; e
- ❖ Tomar decisões mais acertadas.



2.4 Por que a análise preditiva é importante para a evasão escolar?

A análise preditiva realizada é o tipo de análise de dados, mais conhecida e a partir dela foi possível prever cenários futuros baseados também nos dados fornecidos pelo sistema acadêmico. E, por meio de algoritmos de aprendizado de máquina, chegar à previsão da possibilidade de evasão de alunos com determinada variável.

Diante disso, a análise preditiva se mostra bastante promissora ao proporcionar às instituições de ensino uma **visão estratégica sobre oportunidades e riscos que o futuro pode reservar**.

Um exemplo...



Para melhorar os índices da evasão escolar é preciso agir antecipadamente. Para agir antecipadamente precisamos de informações também de maneira antecipada. Para obter tais informações, é preciso projetar e/ou prever os alunos em risco de evasão.

Modelos preditivos atendem a essa necessidade.

3

Modelo preditivo

3. Modelo preditivo

Na prática, a ideia é implementar um sistema capaz de coletar dados. Além disso, registrá-los e transformá-los em informações úteis para a tomada de decisão.

No modelo preditivo trabalharemos com os dados selecionados. Um modelo preditivo é, de forma simplificada, uma função matemática que pode ser aplicada a uma abundante quantidade de dados. A ideia é evidenciar padrões capazes de apontar as próximas tendências. É como se fosse possível prever com eficiência o futuro, de forma matemática, com probabilidade e estatística.

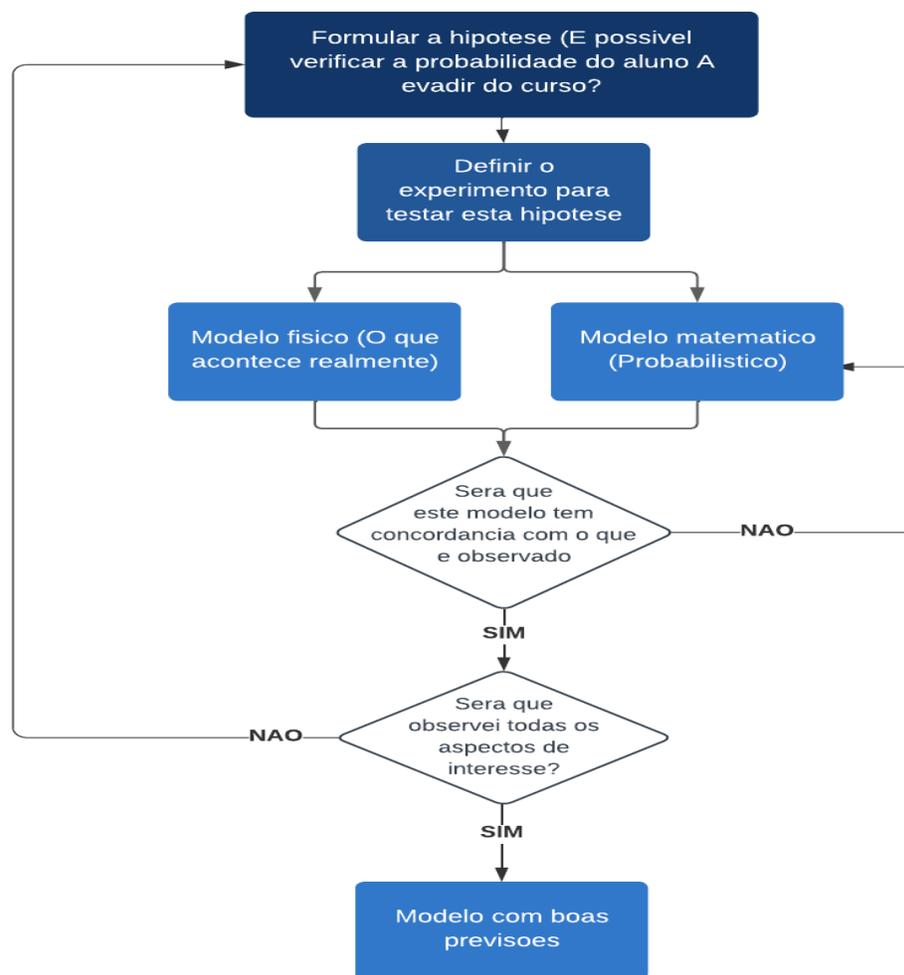
Este tipo de modelo de análise preditiva tem por finalidade minimizar os riscos da evasão escolar e auxiliar na tomada de decisão para encaminhamento de ações mais assertivas pela gestão institucional.

Prever a evasão consiste em identificar os sinais que antecedem a saída do aluno antes da conclusão do curso, por meio de algoritmos estatísticos e técnicas de *Machine Learning* para calcular probabilidades de resultados, a partir de dados armazenados de acordo com cada situação.

O modelo preditivo permite o cruzamento de diversas variáveis que podem determinar quais fatores impactam na evasão escolar, ou seja, no abandono.



3.1 Fluxograma do Modelo preditivo



O Modelo é capaz de auxiliar na tomada de decisão gerencial tempestiva quanto a previsão da evasão de estudantes em uma instituição, por meio da análise de variáveis acadêmicas, geográficas, pessoais e socioeconômicas, gerando a probabilidade de evasão de cada estudante no curso técnico integrado ao ensino médio.

Com um modelo preditivo bem estruturado e calibrado, riscos e oportunidades são identificados com antecedência suficiente para tomar decisões mais acertadas.

4

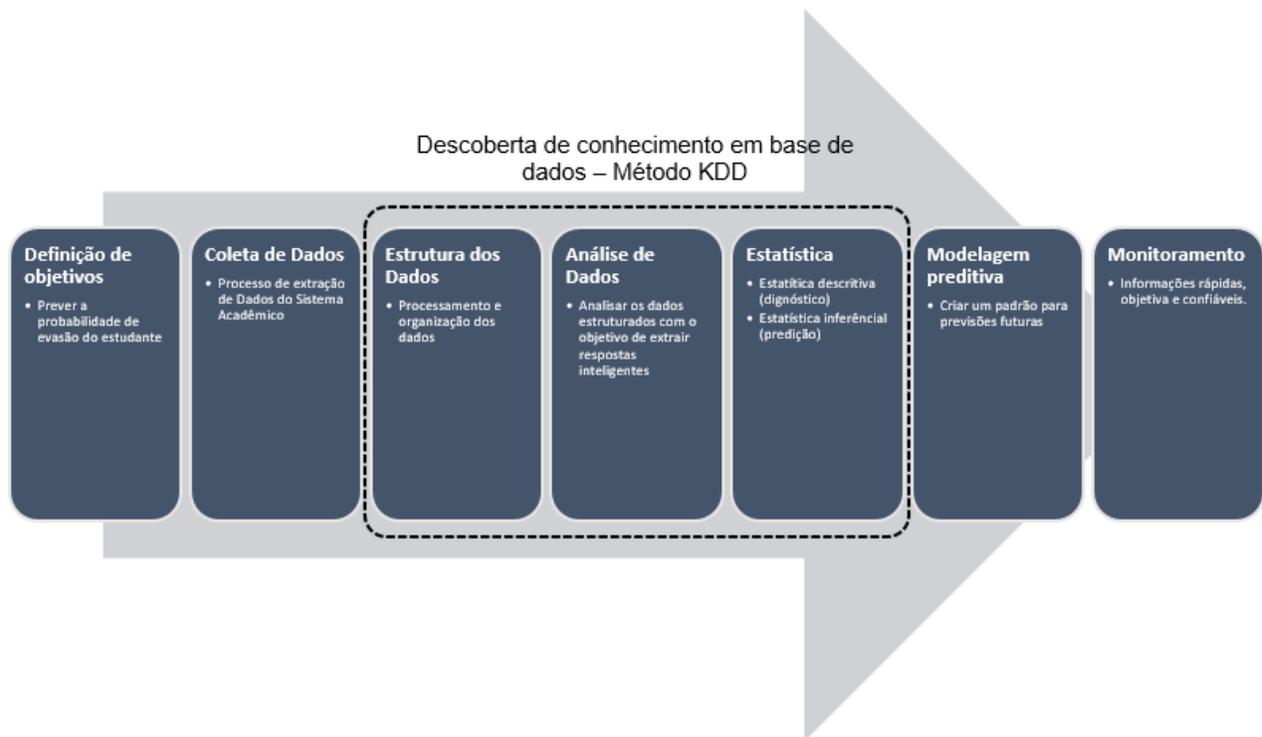
**Como implementar
modelo de análise
preditiva**

4. Como implementar modelo de análise preditiva?

Os modelos preditivos se utilizam de informações de eventos passados para prever comportamentos de alunos por meio da detecção de padrões, mesmo os mais sutis, dentro dos dados analisados.

Para criar um modelo preditivo, algumas etapas precisam ser cumpridas. A primeira delas é definir os objetivos, em seguida coletar dados. São eles que fornecem a base e sustentação para todo e qualquer modelo que se planeje criar.

Veja abaixo os passos que você não pode deixar de cumprir para ter sucesso!



É um processo conhecido como de Extração de Conhecimento em Bases de Dados (do inglês, *KDD – Knowledge Discovery in Database*), formalizado por Fayyad, Piatetsky-Shapiro e Smyth (1996). Em geral, esse processo é utilizado para mineração de dados e extração de conhecimento a partir dos dados pré-processados.

4.1 Defina objetivos concretos

O primeiro passo para implementar modelos preditivos eficazes é propor **objetivos**, é importante a partir de agora criar o modelo preditivo que será utilizado para alcançá-lo.

Definir objetivos é muito importante porque, com eles, serão apontadas as metas a serem analisadas a fim de apontar o alcance do que foi proposto.

4.2 Faça a coleta dos dados

Neste momento, será utilizado os dados contidos no Sistema acadêmico dos alunos evadidos e com matrícula ativa no curso para treinar o modelo.

Contudo, é importante determinar o que é realmente necessário para atingir os objetivos e lembrar que a **qualidade do que é coletado é essencial para garantir a confiabilidade das análises.**

4.3 Prepare os dados para análise (Estrutura dos dados)

Nesta fase, após recolher as informações necessárias, é preciso estruturar estas informações para facilitar o momento de análise. Também é conhecida como a fase da estratégia, pois é onde vai transferir os dados para um local, no qual consiga facilitar a visão do tanto de informação coletada.

São os dados estruturados, geralmente quantificáveis e organizados de maneira fixa; e os não estruturados, que não atendem a uma predefinição específica.

4.4 Análise os dados

Na análise de dados iremos trabalhar com os dados importantes definidos e devidamente estruturados, para a partir deste analisá-los para extrair respostas inteligentes.

Com os dados mais claros para serem visualizados, é neste momento de analisar as informações que se tem.

Esse é o momento de entender, por exemplo, o período letivo onde tem a maior taxa de evasão, para assim pensar em um plano de ação mais efetivo.

4.5 Estatística (Descritiva e Inferencial)

Quando falamos de **estatística na análise preditiva**, levamos em consideração duas técnicas bem conhecidas da área: **estatística descritiva e estatística inferencial.**

Na estatística descritiva temos como objetivo sumarizar e descrever um grande conjunto de dados, para a partir dele ser possível criar medidas de tendência central e medidas de variabilidade ou dispersão.

Já na estatística inferencial é o estudo de um grupo de amostra para tirar conclusões de um grupo maior, ou seja, envolve parte de interpretação dos dados. Levantamentos populacionais são um grande exemplo dessa técnica.

A estatística é fundamental para se ter uma análise preditiva rodando de forma correta.

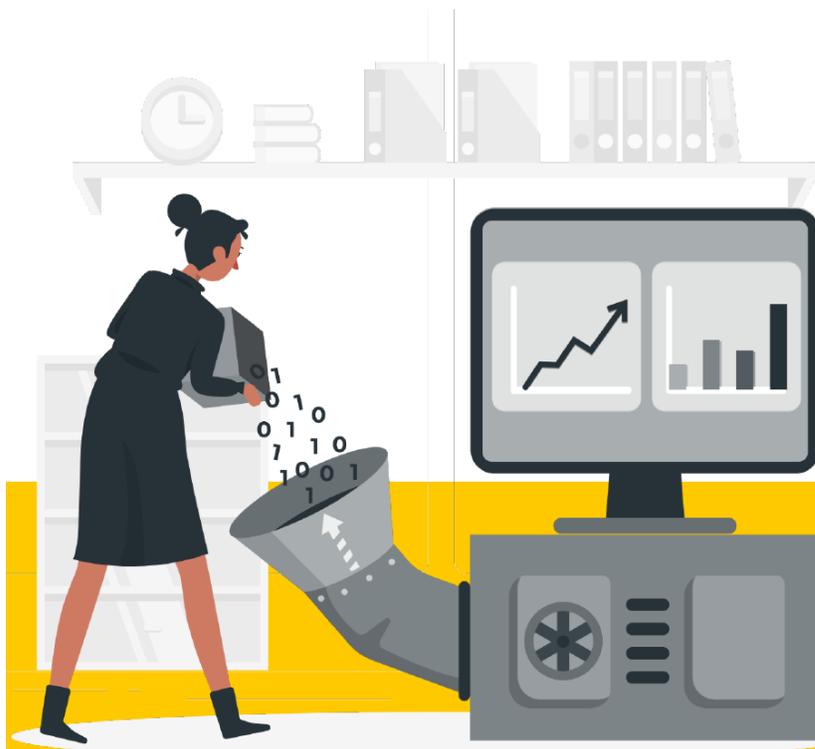
4.6 Criando uma modelagem preditiva

Na modelagem preditiva, com análise conduzida e os resultados reunidos, o próximo passo é **criar um padrão para previsões futuras**. Ele será construído, basicamente, a partir de variáveis consideradas na fase de análise e processar os novos dados coletados com mais agilidade. Desse modo, fica mais fácil visualizar respostas e gerar percepções.

4.7 Monitoramento do modelo de análise dos dados

Nesse momento, após utilizar todos os estágios anteriores, **é necessário manter um monitoramento para que os dados processados e principalmente as informações obtidas a partir da modelagem continue sendo confiáveis**.

Além de ter de uma forma mais rápida e objetiva as respostas que precisa otimizar e melhorar ainda mais o processo e o produto.



5

**Exemplificação da
aplicação do modelo
preditivo para
prevenir a evasão
em um curso técnico**

5. Exemplificação da aplicação do modelo preditivo para prevenir a evasão em um curso técnico

O propósito desse tipo de aplicação é estimar a probabilidade de um novo exemplo pertencer à classe de interesse.

Para problemas como esse, aplicamos um modelo linear aos dados, mas escolhendo outra função objetivo, mais adequada à estimativa de probabilidade de classe. **A regressão logística é um procedimento que podemos utilizar nesse caso.** Sendo assim, é uma técnica estatística comumente usada para prever resultados binários, como se um aluno desistirá ou não.

Chegar a probabilidade de um aluno desistir de um curso pode ser uma tarefa desafiadora, pois depende de várias variáveis, como dados demográficos do aluno, desempenho acadêmico, frequência, motivação e outros fatores pessoais e externos.

Aqui estão alguns passos que podemos seguir para estimar a probabilidade de um aluno desistir de um curso:

5.1 Reúna dados relevantes

Você deve coletar dados sobre o curso, incluindo o tamanho do curso, duração, nível de dificuldade e qualquer outra informação relevante. Além disso, colete dados sobre uma amostra de alunos, incluindo informações sobre fatores de risco que se acredita estarem associados ao abandono escolar, como desempenho acadêmico, frequência, dados demográficos do aluno, idade, sexo e situação socioeconômico, etc. Colete também dados a respeito da pretensão de cada novo em relação a abandonar ou não. Será necessário também extrair esses dados do sistema acadêmico dos alunos evadidos e dos alunos em curso; dos alunos em processo de matrícula deverá ser aplicados questionários e para validar a compatibilidade das variáveis com os fatores geradores da evasão devem ser realizadas entrevistas com servidores que trabalham diretamente com o curso da pesquisa.

Tabela 1 – Dados dos alunos

Categorias	Variáveis
Dados Pessoais:	Matrícula (Código do Aluno)/Data de Nascimento/Sexo/Estado Civil/Município residencial/Origem escolar (Pública ou privada)
Dados Acadêmicos:	Nota do exame de seleção de ingresso/Coeficiente de Rendimento /Forma de Ingresso (Via exame de seleção, transferência, outros).
Financeiros	Renda bruta familiar
Familiar	Quantidade de membros familiar
Cota	Tipo de cota (PPI e PcD)

Fonte: Elaborado pelo autor (2023).

Prepare os dados, limpe-os e prepare-os para análise, incluindo a identificação de quaisquer dados ausentes ou discrepantes.

5.2 Identifique os fatores de risco

Identifique os fatores que podem aumentar o risco de um aluno desistir. Por exemplo, alunos com notas baixas ou registros de frequência, aqueles que trabalham em período integral ou aqueles que viajam longas distâncias podem estar em maior risco de abandono.



Dica:

Utilize o Microsoft Excel, no menu Análise de Dados, para realizar os cálculos estatísticos.

A estatística permite conhecer medidas de tendência central, como a média, também medidas frequentes, como a moda.

Selecione os preditores: Identifique os fatores de risco que deseja incluir no modelo de regressão logística. Estas serão as variáveis preditoras.

5.3 Analisar os dados

Divida os dados em dois conjuntos: um conjunto de treinamento com os dados dos alunos evadidos e um conjunto de teste com os dados dos alunos que seja saber a probabilidade de risco de evadir. O conjunto de treinamento será usado para construir o modelo de regressão logística e o conjunto de teste será usado para avaliar o desempenho do modelo.

Use a análise estatística para identificar padrões nos dados e determinar quais fatores estão

mais fortemente associados ao abandono.

Para tal é preciso utilizar da Estatística para encontrar tendência central, moda, desvio padrão, média. É como estes que montaremos o perfil do aluno evadidos:

Quadro 1 – Dados de um dos alunos e suas variáveis

Código do Aluno	Renda Familiar Total	Quantidade de membro familiar	Autodeclaração de Cor (Etnia)	Forma de Ingresso	Tipo de cota	Gênero	Idade da Evasão	Ano da Evasão	Cidade do Campus
54DF	Até 1 SM	4	Parada	Exame de Seleção	PPI	M	16	1°	0 (Não)

Fonte: Elaborado pelo autor (2023).

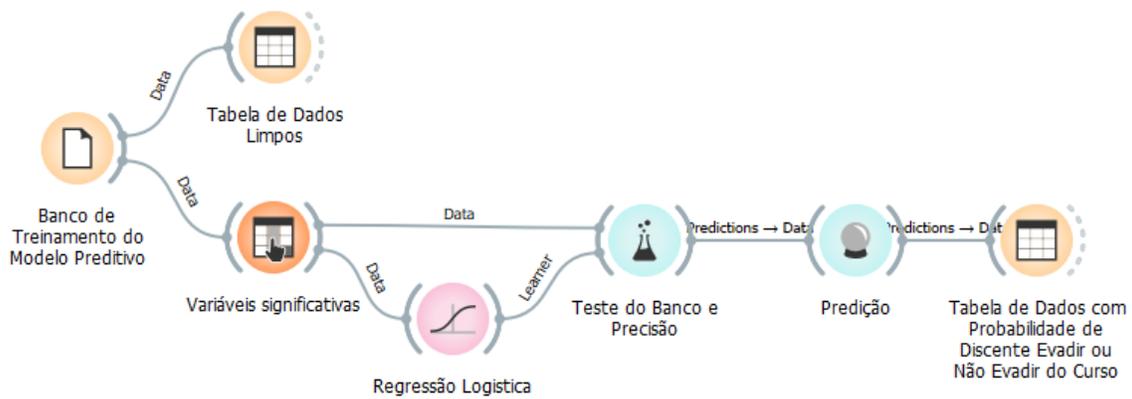
Você pode usar ferramentas como a regressão logística para prever a probabilidade de um aluno desistir com base nos fatores de risco.

Construa o modelo: Use o conjunto de treinamento para construir o modelo de regressão logística. O objetivo é encontrar os valores dos coeficientes do modelo que maximizam a semelhança dos dados observados entre um indivíduo evadido e um indivíduo com matrícula ativa no curso. Estas várias dos alunos compõe o Banco de Dados a ser utilizado no modelo.

Os modelos preditivos permitem o cruzamento de diversas variáveis que podem determinar **quais fatores impactam no abandono, ou seja, na evasão.**

Para realizar a regressão logística e execução do modelo preditivo, utiliza-se um software simulador capaz de gerar as probabilidades. Que neste caso, utilizamos o Software Orange. Ele é uma das poucas ferramentas de mercado, totalmente construída em *python*, que o analista de dados pode trabalhar de forma totalmente visual, com um amplo aspecto de atendimento a diversas demandas na área de mineração de dados por meio do uso de algoritmos de *Machine Learning*.

Figura 1 – Modelo preditivo no Orange



Fonte: Elaborado pelo autor (2023).

Figura 2 – Tabela de dados com as variáveis e para predição.

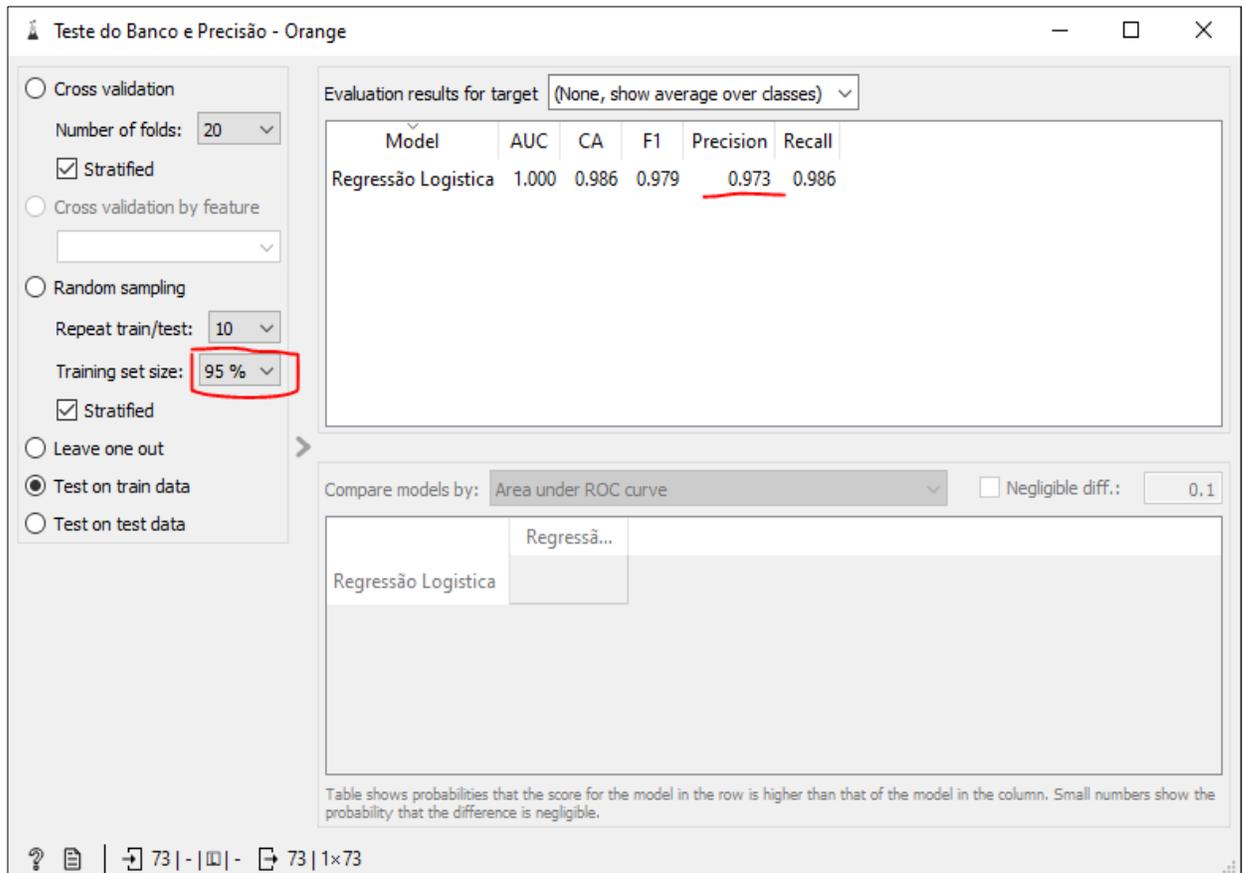
Nome	Ordem	Saida	cota	sexo	Sexo_Masc
1 AL...	1	2015	A0 (NÃO COTIS...	M	1
2 HY...	2	2015	A0 (NÃO COTIS...	F	0
3 JH...	3	2015	L2	M	1
4 AD...	4	2016	A0 (NÃO COTIS...	M	1
5 CL...	5	2016	L2	F	0
6 EM...	6	2016	L2	M	1
7 JO...	7	2016	A0 (NÃO COTIS...	M	1
8 M...	8	2016	L2	F	0
9 VIC...	9	2016	L2	M	1
10 CL...	10	2017	L2	M	1
11 LU...	11	2017	A0 (NÃO COTIS...	M	1
12 LUI...	12	2017	A0 (NÃO COTIS...	F	0
13 PA...	13	2017	L2	M	1

Fonte: Elaborado pelo autor (2023).

5.4 Valide suas descobertas

Teste seu modelo em relação a dados históricos para ver se ele prevê com precisão as taxas de abandono. Ajuste o modelo conforme necessário para melhorar sua precisão.

Figura 3 – Modelo preditivo, em teste, no Orange



Fonte: Elaborado pelo autor (2023).

Pois, para pesquisa foi definido um intervalo de confiança de 0,95 ou 95%. Em outras palavras, se p menor 0,05, os resultados são estatisticamente significantes.

5.5 Interprete, monitore e ajuste o modelo

Interprete o modelo: Após construir e avaliar o modelo, você pode interpretar os coeficientes para ver quais fatores de risco estão mais fortemente associados ao abandono. Você também pode usar o modelo para fazer previsões sobre a probabilidade de abandono de novos alunos com base em seus perfis de fatores de risco. Neste modelo abaixo, duas variáveis impactam diretamente na probabilidade de evasão, o “Ano_evasão” e a “Idade”.

Por fim, monitore o curso e os alunos e ajuste o modelo conforme necessário. À medida que novos dados ficam disponíveis, você pode atualizar o modelo para fazer previsões mais precisas.

Refinar o modelo: Se o modelo não tiver um bom desempenho, você pode tentar refiná-lo adicionando ou removendo variáveis preditoras, transformando os dados ou tentando diferentes

técnicas de modelagem.

A análise preditiva pode apontar comportamentos específicos dos alunos em sua jornada acadêmica.

Figura 4 - Modelo preditivo, variáveis e probabilidade de evasão de cada conforme as variáveis predictoras, no Orange

Predição - Orange

Show probabilities for: Classes in data Show classification errors Restore Original Order

Classificacao	Nome	são Lc	Regressão Logística (Evadido)	Regressão Logística (NaoEvadido)	Fol
Evadido	KAY...	Eva...	0.973205	0.0267948	1
Evadido	KEV...	Eva...	0.999921	7.88416e-05	1
Evadido	LUC...	Eva...	0.999726	0.000273621	1
Evadido	LUI...	Eva...	0.999519	0.000481163	1
Evadido	MA...	Eva...	0.998199	0.00180132	1
Evadido	MA...	Eva...	0.975977	0.0240229	1
Evadido	MIL...	Eva...	0.998243	0.00175713	1
Evadido	RAP...	Eva...	0.999696	0.000304458	1
Evadido	RO...	Eva...	0.999696	0.000304458	1
Evadido	SA...	Eva...	0.997256	0.00274355	1
Evadido	VIT...	Eva...	0.988936	0.0110638	1
Evadido	WE...	Eva...	0.996278	0.00372167	1
Evadido	WIS...	Eva...	0.999546	0.000454474	1
NaoEvadido	ALU...	Eva...	0.645264	0.354736	1

Show performance scores Target class: (Average over classes)

Fonte: Elaborado pelo autor (2023).

Apredizados

- A análise preditiva é, resumidamente, uma técnica que consiste no uso de históricos de dados para projetar previsões e antecipar tendências. Com ela, as instituições podem suprimir suposições e achismos que, atualmente, não são suficientes para responder às novas demandas do mercado.
- A técnica se mostra bastante promissora ao proporcionar às organizações uma visão estratégica sobre oportunidades e riscos que o futuro pode reservar.
- É importante observar que prever a probabilidade de evasão de um aluno não é uma ciência exata, e sempre existem outros fatores que podem influenciar a decisão de um aluno de permanecer ou sair de um curso. No entanto, analisando os dados disponíveis, você pode fazer estimativas informadas que podem ajudá-lo a identificar os alunos que podem precisar de suporte ou intervenções adicionais para ajudá-los a ter sucesso no curso.
- O modelo preditivo à evasão apresentado é válido para a realidade do curso objeto do estudo. Porém, o método trabalhado pode ser utilizado em outras Instituições de Ensino.
- No geral, a regressão logística pode ser uma ferramenta útil para prever a probabilidade de um aluno desistir com base em fatores de risco. No entanto, é importante ter em mente que a precisão das previsões dependerá da qualidade dos dados e da escolha das variáveis preditoras, e que a regressão logística não pode dar conta de todos os fatores que podem contribuir para a evasão.

- > Portanto, o Modelo de Análise Preditiva, propõe-se a aplicação da **regressão logística**, de maneira que se possa calcular a probabilidade de evasão dos alunos ao final do semestre, e permitir que os responsáveis pelos fatores dos estudantes possam realizar abordagens de intervenção baseadas nas variáveis que mais influenciam no cálculo da evasão, este modelo se mostra mais preciso na classificação dos alunos e, portanto, permite uma redução na quantidade de abordagens que seriam realizadas quando comparado aos demais.

- > Em resumo, as instituições de ensino e gestores escolares estão adotando a análise preditiva para:
 1. Prever as próximas movimentações do segmento discente;
 2. Identificar oportunidades de ações de intervenção;
 3. Prevenir evasões e reduzir riscos;
 4. Otimizar estratégias de combate ao abandono;
 5. Mapear o comportamento e variáveis de discentes;
 6. Melhorar as operações e aumentar a eficiência das ações de acesso, permanência e êxito.

Acesse este conteúdo pelo QR code, abaixo:



REFERÊNCIAS

FAYYAD, U.; SHAPIRO, G. P. – From Data Mining to knowledge Discovery in databases. AI Magazine, 17, Fall 1996.

BARI, Anasse; CHAOUCHI, Mohamed; JUNG, Tommy. Análise Preditiva para leigos. Alta Books, 2020.



Programa de Pós-graduação em Educação Profissional e Tecnológica - ProfEPT.

AO LEITOR

Este guia tem por objetivo propor auxiliar a insituição de ensino por meio de seus gestores, mais uma ferramenta que visa minimizar a evasão escolar em nossa comunidade acadêmica. Com interface amigável, para atender desde as necessidades mais simples às mais sofisticadas e possibilita análises que combinam dados externos e internos, monitorados automaticamente, com dados internos do próprio registro escolar dos discentes.

Agradecemos!