

# Machine Learning como ferramenta de auxílio no tratamento da evasão escolar

Fagner Ferreira de Santana<sup>1</sup>

<sup>1</sup> Universidade Federal da Bahia, Salvador, Bahia, Brasil

**Resumo:** A evasão escolar é um problema grave que atinge as Instituições de Ensino e ela pode ocorrer por diversos fatores. Neste trabalho, foi desenvolvida uma ferramenta de previsão que ajudará a detectar alunos em risco de evasão. As análises foram realizadas utilizando modelos de *Machine Learning* e dados de matrículas de alunos, concluintes ou evadidos, oriundos de cursos técnicos do SENAI-BA. O modelo com maior precisão foi o *Random Forest*, utilizando o método de seleção de variáveis Featurewiz.



# Problemas, Dados

## Problema

Configurar um modelo que permita estimar, a ocorrência ou não, de evasão em cursos técnicos.

Este problema pode ser dividido em duas etapas:

1. Identificação das melhores características;
2. Seleção do modelo de *Machine Learning* com melhor resultado.

## Dataset

A base é formada por informações referente às matrículas dos estudantes, concluintes ou evadidos, em cursos técnicos.

1. Oriunda do SENAI Bahia;
2. Dados entre 2012 e 2022;
3. 57.552 registros;
4. Contém 15 características;



# Métodos

## Etapa 1: Seleção de Variáveis

1. Stepwise Adaptado
2. Boruta
3. Featurewiz



## Etapa 2: Modelos Machine Learning

1. Regressão Logística
2. Naïve Bayes
3. Random Forest
4. Rede Neural



## Etapa 3: Combinação de Modelos

1. Voting Soft
2. Voting Hard
3. Stacking



# Resultados e Conclusões

Utilizando-se da técnica Featurewiz para a seleção de variáveis e o modelo *Random Forest*, foi possível identificar um subconjunto de 4 características, gerando o resultado para predição de 72% de acurácia.

	<i>Stepwise</i>	<i>Boruta</i>	<i>Featurewiz</i>
<i>Regressão Logística</i>	0,63	0,63	0,63
<i>Naive Bayes</i>	0,61	0,59	0,57
<i>Random Forest</i>	0,72	0,71	0,72
<i>Rede Neural</i>	0,68	0,70	0,66

O modelo não apresenta um ajuste com alta acurácia, porém pode servir de ferramenta de auxílio para direcionar decisões para tentar minimizar a evasão escolar. Este resultado pode ser melhorado com a introdução de novas características (*features*) que descrevam outros aspectos dos alunos não contempladas na base de dados atual.



4th Conference on  
**Statistics and  
Data Science**  
Salvador, Brazil (online)  
December 1-3, 2022