

Descrição da base de Dados

Os dados disponibilizados contêm informações socioeconômicas de alunos que contrataram um financiamento estudantil por meio do FIES.

Table 1: Descrição das variáveis

Codigo	Descricao
SG_SEXO	Sigla Sexo
NU_IDADE	Idade em anos
DS_RACA_COR	Descrição Raça/Cor
SG_UF	Sigla Unidade da Federação
DS_ESTADO_CIVIL	Descrição Estado Civil
ST_DEFICIENCIA	Possui deficiência?
ST_ENSINO_MEDIO_ESCOLA_PUBLICA	Cursou EM em Escola Pública?
ST_BOLSISTA_PROUNI	É Bolsista do PROUNI?
VL_RENDA_FAMILIAR_BRUTA_MENSAL	Valor da Renda Familiar Bruta Mensal
VL_RENDA_PESSOA_BRUTA_MENSAL	Valor da Renda Bruta Mensal do Aluno
VL_RENDA_PERCAPITA	Valor da Renda Per Capita na Família
NU_SEMESTRE_FINANCIADO	Número de Semestres Financiados
VL_FINANCIAMENTO	Valor do Financiamento
ST_INADIMPLENCIA	Situação Inadimplência: S ou N (Sim ou Não)

Objetivo

Confeccionar um modelo preditivo, utilizando apenas as técnicas apresentadas na disciplina, para classificar os alunos em “bons” ou “mau-pagadores”. Considerando que, do ponto de vista do gestor do fundo, o custo de um **falso negativo** é bastante alto, ou seja, é pior se classificar um aluno como “bom pagador” quando, na verdade, o estudante seria um “mau-pagador” – você deve olhar principalmente para a sensibilidade (ou recall) do seu modelo.

Sumário Executivo:

- Descrição do melhor modelo/combinção de técnicas que apresentaram o melhor resultado em termos preditivos;
- Discorrer sobre a performance do seu melhor modelo (descritivo);
- Apresentar um modelo alternativo ao melhor modelo. Discorrer sobre os resultados e explicar o motivo pelo qual você escolheu esse modelo/combinção de técnicas como alternativa;
- Discorrer sobre quais foram os outros modelos e técnicas testadas;

Resultados (saídas do R):

- Sumário e acurácia do melhor modelo;
- Sumário e acurácia do modelo alternativo;

Anexos:

- Códigos utilizados para ler e tratar os dados; ajustar o melhor modelo e o modelo alternativo;

Observações Finais:

- É necessário tratar os dados antes de realizar a análise preditiva: eliminar colunas que possuem a mesma informação, missing values, transformar variáveis categóricas em fatores, etc. Além de interferir na capacidade preditiva do seu modelo, poderá invalidar completamente sua análise.