

Atividade 2

No arquivo dados_filtrados.csv temos várias informações de candidatos do ENEN 2016, incluindo dados geográficos, escolares e desempenho nas provas. O objetivo dessa atividade é de usar as técnicas já aprendidas até o momento, principalmente as de regressão, para tentar criar um modelo que seja capaz de prever as notas da prova de matemática a partir das demais informações disponíveis.

Inicialmente, faça um estudo exploratório para tentar identificar quais atributos seriam mais adequados na predição das notas de matemática. Na base de dados há muitos atributos categóricos codificados em números. Fazendo uma boa análise exploratória, você poderá verificar quais atributos categóricos de fato possuem influência sobre a variável alvo. Descarte os atributos categóricos que não trazem informação relevante.

Faça a preparação dos dados, tratando dados faltantes, escalonamento (se necessário) e partição em treino e teste.

Criar diferentes modelos, utilizando os métodos de regressão multilinear e regressor Random Forest (`sklearn.ensemble.RandomForestRegressor`). Avaliar a performance dos modelos com base em viés e tendenciosidade.

O dicionário da tabela, explicando o significado de cada atributo, está na planilha `dicionario.xlsx`.

Você deve entregar um relatório discutindo sua abordagem, modelos e técnicas utilizadas e também uma avaliação das métricas de desempenho para dar suporte ao seu resultado.