

FIAP X CCR

Abstração dos Dados para Machine Learning

Integrantes:

João Vinícius Alves (559369)

Juan Pablo Coelho (559445)

Matheus Mariotto (560276)

Sumário

Descrição 3

Variável Alvo 3

Descrição

Os dados utilizados no projeto são trazidos da base de dados aberta da Agência Nacional do Transporte Terrestre (ANTT) e contém informações sobre falhas ferroviárias entre 2020 e 2024. O link para acessar os dados brutos está [aqui](#). Abaixo está a descrição de cada coluna após processar os dados:

Nome da Variável	Descrição
Data	Data e horário da ocorrência.
Perímetro Urbano	Indica se a falha ocorreu em perímetro urbano.
Causa Direta	Motivo principal da falha.
Causa Contributiva	Possível fator adicional que contribuiu para a falha.
Natureza	Tipo de incidente ocorrido (ex: abalroamento, descarrilamento).
Interrupção (h)	Tempo de interrupção da operação ferroviária, em horas.
Número de Feridos	Número de pessoas feridas.
Número de Óbitos	Número de óbitos registrados.
Prejuízo (R\$)	Valor financeiro do prejuízo causado pelo incidente.

Variável Alvo

A variável alvo escolhida para este projeto é **Interrupção (h)**, pois representa o impacto operacional das falhas ferroviárias. O objetivo do modelo é prever o tempo de interrupção com base nas demais variáveis, permitindo antecipação de impactos e tomada de decisões mais rápidas.

Algumas variáveis possuem correlações importantes com a variável alvo:

- **Causa direta e natureza:** Algumas causas diretas tendem a gerar interrupções mais longas, como descarrilamentos.
- **Número de feridos e óbitos:** Ocorrências com vítimas podem demandar maior tempo de atendimento, aumentando a interrupção.
- **Prejuízo:** Eventos mais graves e dispendiosos podem estar associados a interrupções prolongadas.