

Projeto Final

Denise do Rocio Maciel

GitHub:<https://github.com/denise25maciel/bootcamp/>

Kaggle:<https://www.kaggle.com/code/denisemaciel/notebook>



ETAPA 1 - ENTENDIMENTO
DO PROBLEMA



ETAPA 2 - SOLUÇÃO
PROPOSTA



ETAPA 3 – SOLUÇÃO
“IDEAL”

Resultado final

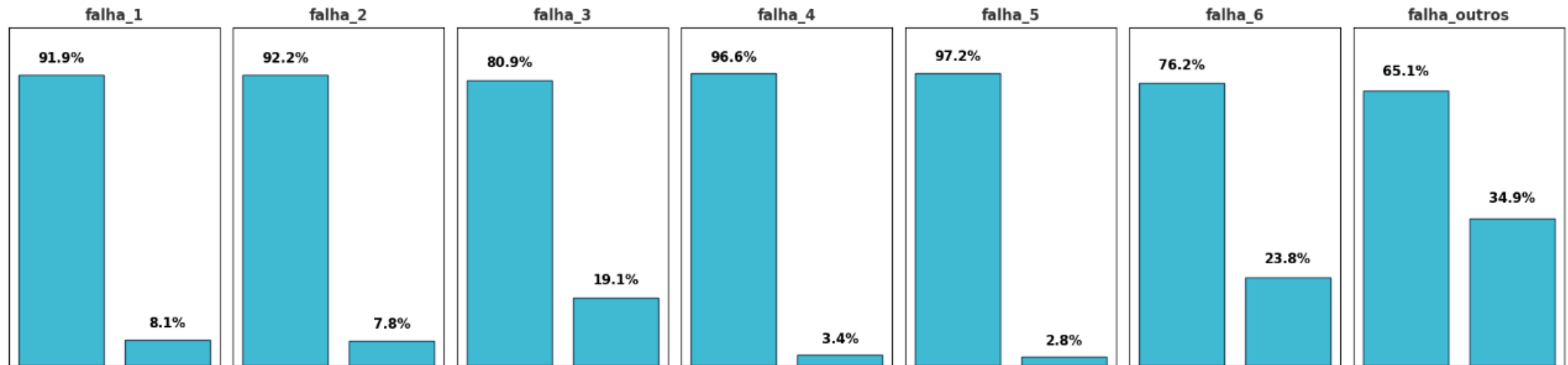
US\$ 2,5 trilhões (ano)

Uso do aço
inoxidável



Quanto custam essas falhas?

Total de erros por tipo





ETAPA 1 - ENTENDIMENTO
DO PROBLEMA



ETAPA 2 - SOLUÇÃO
PROPOSTA



ETAPA 3 – SOLUÇÃO
“IDEAL”

Resultado final

Solução proposta



Multiclasse

Um erro por item

Uma coluna alvo

Probabilidade clássica

Etapas desenvolvidas

- Insights sobre os dados
- Remoção de outliers e valores nulos
- Normalização
- Aplicação do modelo de Machine Learning
- Avaliação

Etapas desenvolvidas

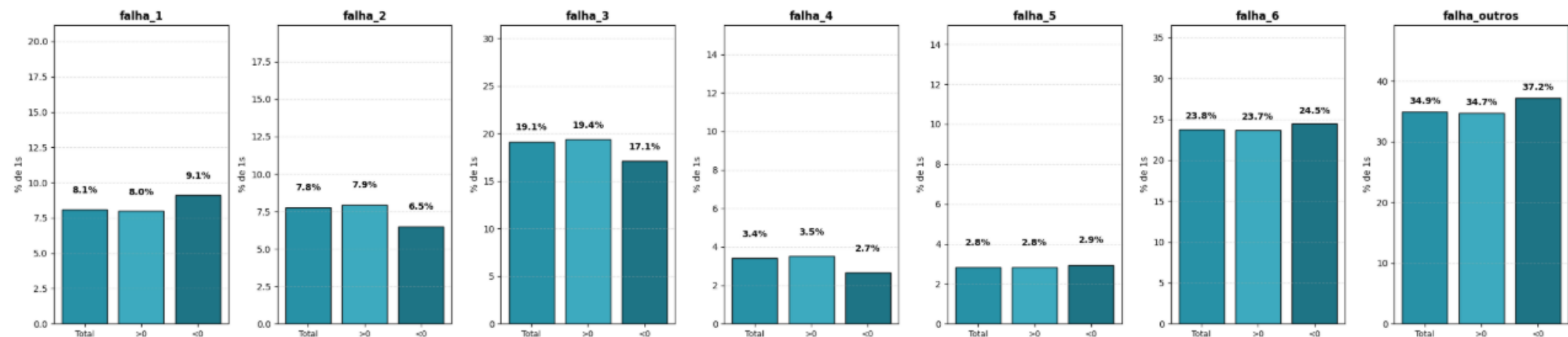
- **Insights sobre os dados**
 - >> **Considera-se grave o modelo deixar de identificar um erro**
(apesar de só haverem dados com erro no dataset)
 - >> Modelo **multiclasse**
 - >> Classes **desbalanceadas**
 - >> Cada item apresenta **um e apenas um erro**
 - >> Classes em **One Hot Encoder** adaptadas para **One Label Encoder**
- Remoção de outliers e valores nulos
- Normalização
- Aplicação do modelo de Machine Learning
- Avaliação

Etapas desenvolvidas

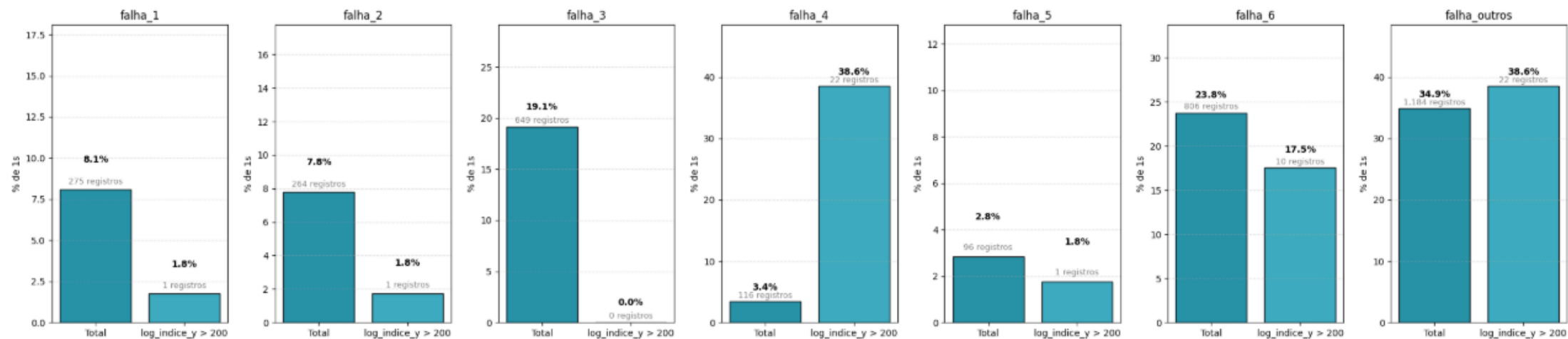
- Insights sobre os dados
- **Remoção de outliers e valores nulos**
 - >> Evitou-se remoção devido ao desbalanceamento
 - >> Média com outlier
 - >> Média sem outlier
 - >> Remoção de colunas sem desvio padrão
 - >> Remoção de colunas pouco significativas
- Normalização
- Aplicação do modelo de Machine Learning
- Avaliação

Outliers

Erros por grupo de outliers - coluna: perimetro_x



Percentual de registros com valor 1 — variável: log_indice_y



Etapas desenvolvidas

- Insights sobre os dados
- Remoção de outliers e valores nulos
- **Normalização**
 - >> Min-max scale, **distribuição não normal**
- **Aplicação do modelo de Machine Learning**
 - >> Radom Forest Classifier
- **Avaliação**
 - >> Recall e F1-Score
 - >> Probabilidade clássica



ETAPA 1 - ENTENDIMENTO
DO PROBLEMA



ETAPA 2 - SOLUÇÃO
PROPOSTA



ETAPA 3 – SOLUÇÃO
“IDEAL”

Resultado final

Solução “ideal”



Classificação binária + Classificação Multirótulo

Zero ou vários erros por item

N colunas alvo calculadas isoladamente

Probabilidade por Teorema de Bayes

Etapas desenvolvidas

- Conversar com **Stakeholders** sobre os prejuízos gerados por cada tipo de erro
- **Classificação binária** para detectar peças com e sem problemas. Utilizar métrica Recall.
- **Classificação Multirrótulo**, formato One Hot Encoder, para que cada item possa apresentar mais de um erro
- Análise por **One-vs-Rest (OvR)**, analisar cada rótulo como um classificador binário
- Aplicação do modelo de **Machine Learning LightGBM ou XGBoost**
- Avaliação por **Recall, F1 Score** e probabilidade por **teorema de Bayes**.