

# Sistema Inteligente de Controle de Qualidade para Chapas de Aço Inoxidável

Patrik Lima Pereira



# Documento de Requisitos / *Product Backlog*

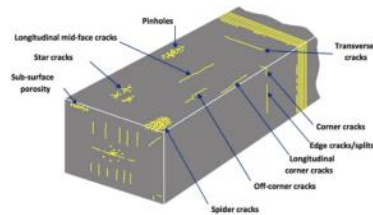
## Projeto Final do Bootcamp CDIA

Início: 12/05/2025

Apresentação para a banca: 15/05/2025 e 16/05/2025

Tempo de apresentação: 10 minutos

### Contextualização



Uma empresa do setor siderúrgico contratou você para a criação de um sistema inteligente de controle de qualidade para chapas de aço inoxidável. Essa empresa forneceu um conjunto de dados contendo informações extraídas a partir de imagens de superfície das chapas, com o objetivo de detectar e classificar defeitos automaticamente. Cada amostra no conjunto de dados é composta por 31 indicadores que descrevem aspectos geométricos e estatísticos do defeito identificado, como área, perímetro, índices de orientação, luminosidade e proporção de bordas. Além dessas características, cada amostra é rotulada com uma das sete possíveis classes de defeitos (seis categorias específicas e uma categoria genérica de "outros defeitos").

O sistema deverá ser capaz de, a partir do cadastro de uma nova imagem (ou conjunto de medições), **prever a classe do defeito e retornar a probabilidade associada**. Além disso, a empresa espera que você extraia *insights* da operação e dos defeitos e gere visualizações de dados.



## Reuniões (*Daily Scrum*)

### Datas:

- 09/05/2025 às 15h00
- 12/05/2025 às 9h00

### Participantes:

#### Scrum Master

- Guilherme Bernieri

#### Líder Técnico:

- Rodrigo Kobashikawa Rosa

**Observação:** Devido ao formato do Bootcamp, **não é possível emular completamente o *Scrum*** conforme o framework original

Requisitos funcionais,  
requisitos não funcionais e  
requisitos de sistemas.

# Metodologia

## Desenvolvimento da PoC (*Proof of Concept*)

### Abordagem

- Iterativa e incremental, com enfoque na prototipagem.
- **Ferramentas:** Google Colab, Python, Biblioteca Scikit-learn e Git para versionamento.

### Processo

- Uso de técnicas de Machine Learning com refatorações sucessivas em Python.
- **Etapas:** Pesquisa, Análises (Resumo Estatístico, Tratamentos de Dados e Pré-processamento, Visualização e Interpretação dos Dados), Configurações, Codificação em Python, Depuração e Testes.
- Validação das métricas com função de parâmetros desejáveis.

### Planejamento

- Fases planejadas e testadas em função do Documento de Requisitos e as “*Daily Scrum*”.
- Incorporação sucessiva de requisitos funcionais, não funcionais e requisitos de sistema.

## **Apresentação para o Cliente (*Stakeholder*)**

# Análise Preliminar de Viabilidade

É viável desenvolver um sistema inteligente de controle de qualidade para chapas de aço inoxidável?

**Sim!** O sistema pode a partir de imagens (ou conjunto de medições) **prever a classe do defeito e retornar a probabilidade associada** de superfície das chapas, com o objetivo de detectar e classificar defeitos automaticamente.

## Alta Confiabilidade

- A PoC, mesmo com os achados, retornou uma taxa de acertos de 68% em classificação.
- Detecção automática dos 7 tipos de defeitos com suas probabilidades associadas.

## Redução de Custos e Tempo

- Elimina inspeção manual, reduzindo erros humanos e paradas na produção.
- Processamento rápido de imagens (dados de 32 indicadores por amostra).

## Achados

- Análise das imagens da chapa revelou alguns problemas, juntamente com outros dados fornecidos e atrelados, que precisam ser investigados

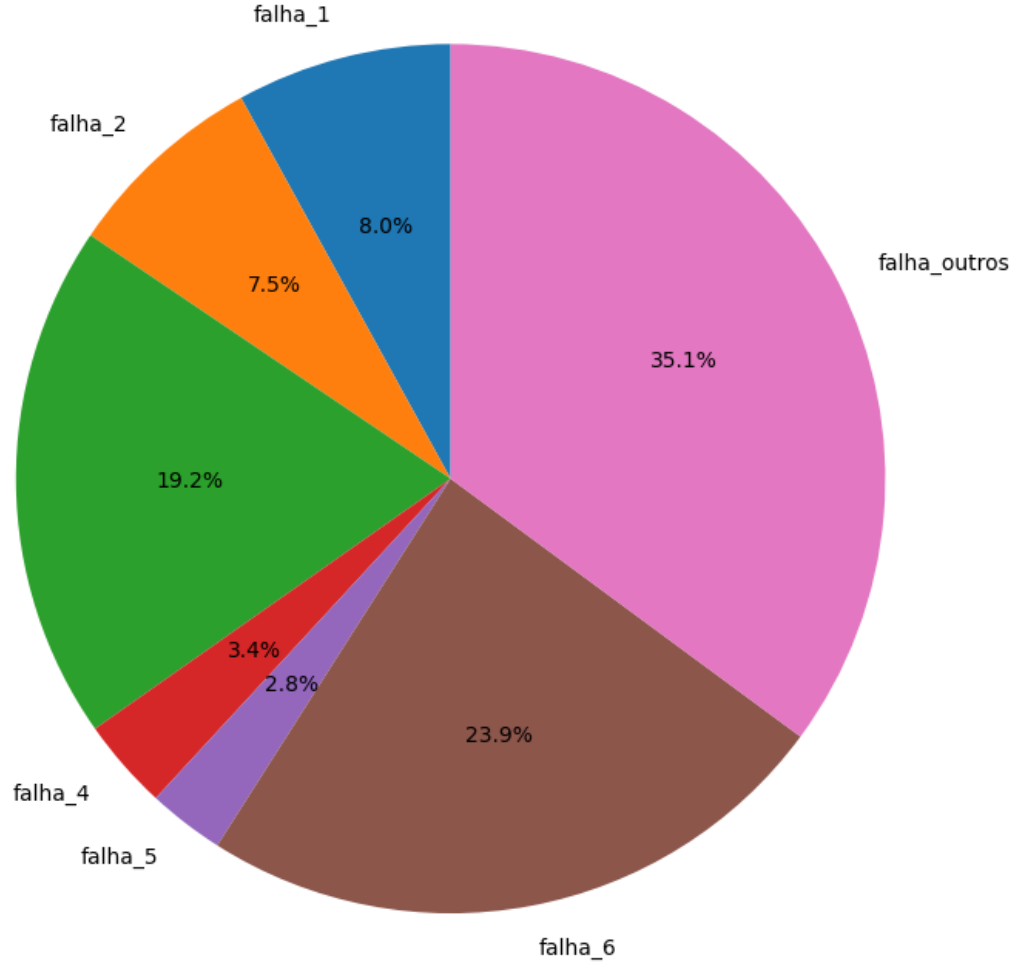
## Análise Preliminar de Viabilidade - Achados

**A base de dados fornecida para a prova de conceito encontra-se em um estado inconsistente, ou seja, não corresponde à realidade do mundo real, com dados e medidas impossíveis ou improváveis, tais como:**

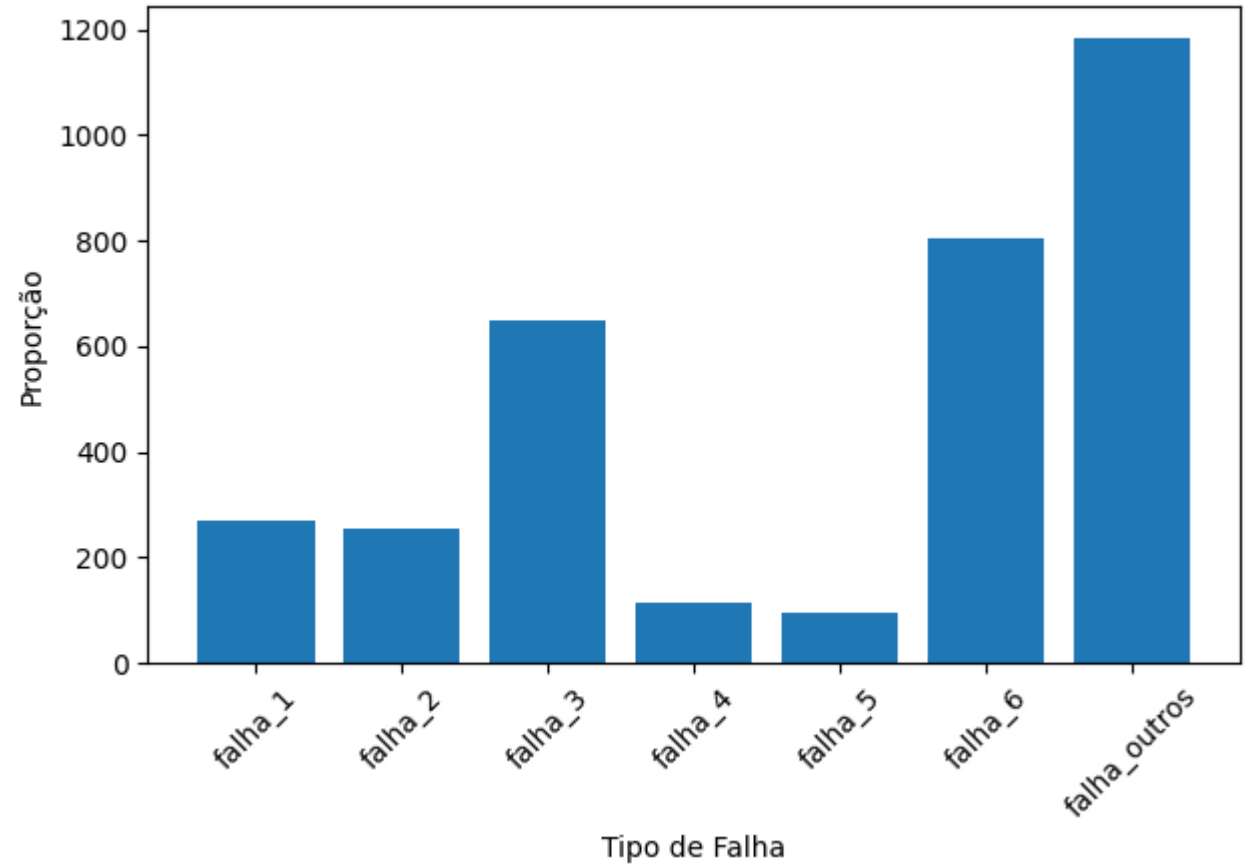
- Chapas com dimensões distintas possuindo o mesmo peso;
- Chapa de aço que não é do tipo A400 nem A300;
- Valores negativos impossíveis, com repetição constante, apontando padrões de erros;
- Proporção de defeito abrangendo, principalmente, outros tipos de defeitos e não mapeando corretamente o tipo de defeitos;
- Falta de padronização e muitos registros nulos;
- Pesos iguais para todos os tipos de chapas.

# Análise Preliminar de Viabilidade - Achados

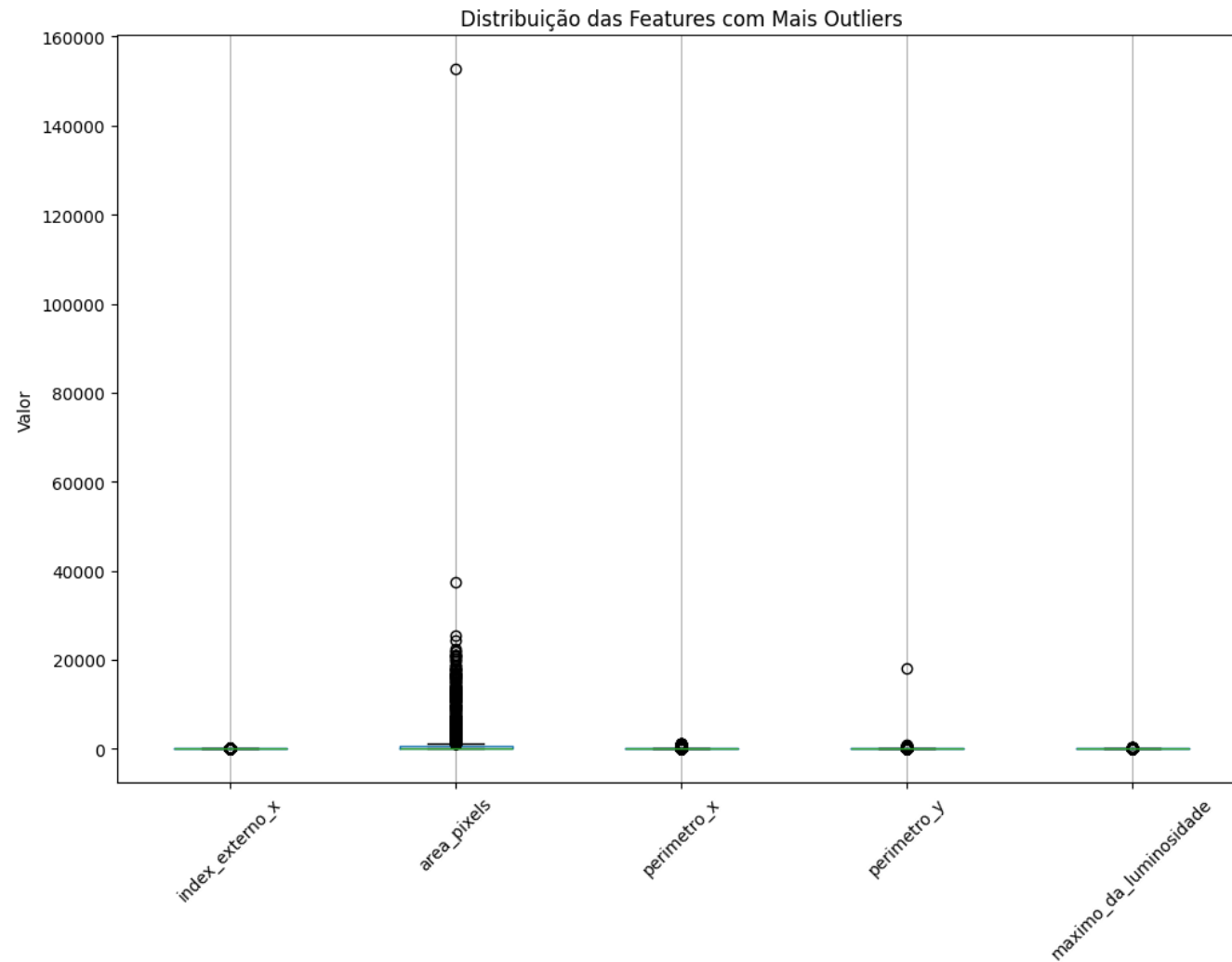
Proporção de Defeitos por Tipo de Falha



Proporção de Defeitos por Tipo de Falha



# Análise Preliminar de Viabilidade - Achados





# Análise Preliminar de Viabilidade - Achados

**Não foi possível identificar a origem dos problemas, mas formulam-se algumas hipóteses:**

- Anotações manuais ou registros de sistemas ocorrendo de maneira incorreta;
- Classificações incorretas;
- Sensores defeituosos;
- Produção de três tipos de aço;
- Cálculo de pesos ou pesagem incorretos;
- Software defeituoso analisando as imagens de modo impreciso;
- Problemas no processo;
- Sabotagem nos dados;
- Falta de padronização de alguns processos; e
- Corrupção da base de dados.

# Análise Preliminar de Viabilidade - Impactos

**O modelo pode apresentar baixa confiabilidade e previsões imprecisas quando colocado em ambiente de produção, resultando em:**

- Alto risco de decisões equivocadas baseadas na solução inteligente;
- Desperdício de tempo e recursos com retrabalho para ajustar o modelo, sem resolver a causa raiz;
- Impossibilidade de validar e homologar a solução para implantação em produção.

Sugere-se para próxima etapa a revisão completa do processo siderúrgico relacionado aos dados de chapas e suas falhas, com o objetivo de identificar as fontes das inconsistências no conjunto de dados, para iniciar os trabalhos com uma base de dados em um estado consistente

# **Apresentação para Equipe Técnica**

# Resumo Estatísticos

```
1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3390 entries, 0 to 3389
Data columns (total 39 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   id                    3390 non-null   int64
 1   x_minimo              3390 non-null   int64
 2   x_maximo              3334 non-null   float64
 3   y_minimo              3390 non-null   int64
 4   y_maximo              3390 non-null   int64
 5   peso_da_placa         3390 non-null   int64
 6   area_pixels           3390 non-null   int64
 7   perimetro_x           3390 non-null   int64
 8   perimetro_y           3390 non-null   int64
 9   soma_da_luminosidade  3290 non-null   float64
10  maximo_da_luminosidade 3292 non-null   float64
11  comprimento_do_transportador 3390 non-null   int64
12  tipo_do_aço_A300       3390 non-null   object
13  tipo_do_aço_A400       3314 non-null   object
14  espessura_da_chapa_de_aço 3349 non-null   float64
15  temperatura            3390 non-null   float64
16  index_de_bordas        3390 non-null   float64
```

```
17  index_vazio            3390 non-null   float64
18  index_quadrado          3354 non-null   float64
19  index_externo_x         3390 non-null   float64
20  indice_de_bordas_x      3390 non-null   float64
21  indice_de_bordas_y      3390 non-null   float64
22  indice_de_variacao_x    3390 non-null   float64
23  indice_de_variacao_y    3390 non-null   float64
24  indice_global_externo   3331 non-null   float64
25  log_das_areas           3390 non-null   float64
26  log_indice_x            3390 non-null   float64
27  log_indice_y            3390 non-null   float64
28  indice_de_orientacao    3390 non-null   float64
29  indice_de_luminosidade  3340 non-null   float64
30  sigmoide_das_areas      3390 non-null   float64
31  minimo_da_luminosidade  3390 non-null   int64
32  falha_1                 3390 non-null   object
33  falha_2                 3390 non-null   object
34  falha_3                 3390 non-null   bool
35  falha_4                 3390 non-null   object
36  falha_5                 3390 non-null   object
37  falha_6                 3390 non-null   bool
38  falha_outros            3390 non-null   object
dtypes: bool(2), float64(20), int64(10), object(7)
memory usage: 986.7+ KB
```

df.info() sem tratamento

# Tratamentos de Dados e Pré-processamento

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 3389 entries, 0 to 3388
```

```
Data columns (total 37 columns):
```

#	Column	Non-Null Count	Dtype
0	x_minimo	3389 non-null	int64
1	x_maximo	3389 non-null	int64
2	y_minimo	3389 non-null	int64
3	y_maximo	3389 non-null	int64
4	area_pixels	3389 non-null	int64
5	perimetro_x	3389 non-null	int64
6	perimetro_y	3389 non-null	int64
7	soma_da_luminosidade	3389 non-null	int64
8	maximo_da_luminosidade	3389 non-null	int64
9	comprimento_do_transportador	3389 non-null	int64
10	tipo_do_aço_A300	3389 non-null	int64
11	tipo_do_aço_A400	3389 non-null	int64
12	espessura_da_chapa_de_aço	3389 non-null	int64
13	temperatura	3389 non-null	float64
14	index_de_bordas	3389 non-null	float64
15	index_vazio	3389 non-null	float64
16	index_quadrado	3389 non-null	float64
17	index_externo_x	3389 non-null	float64
18	indice_de_bordas_x	3389 non-null	float64

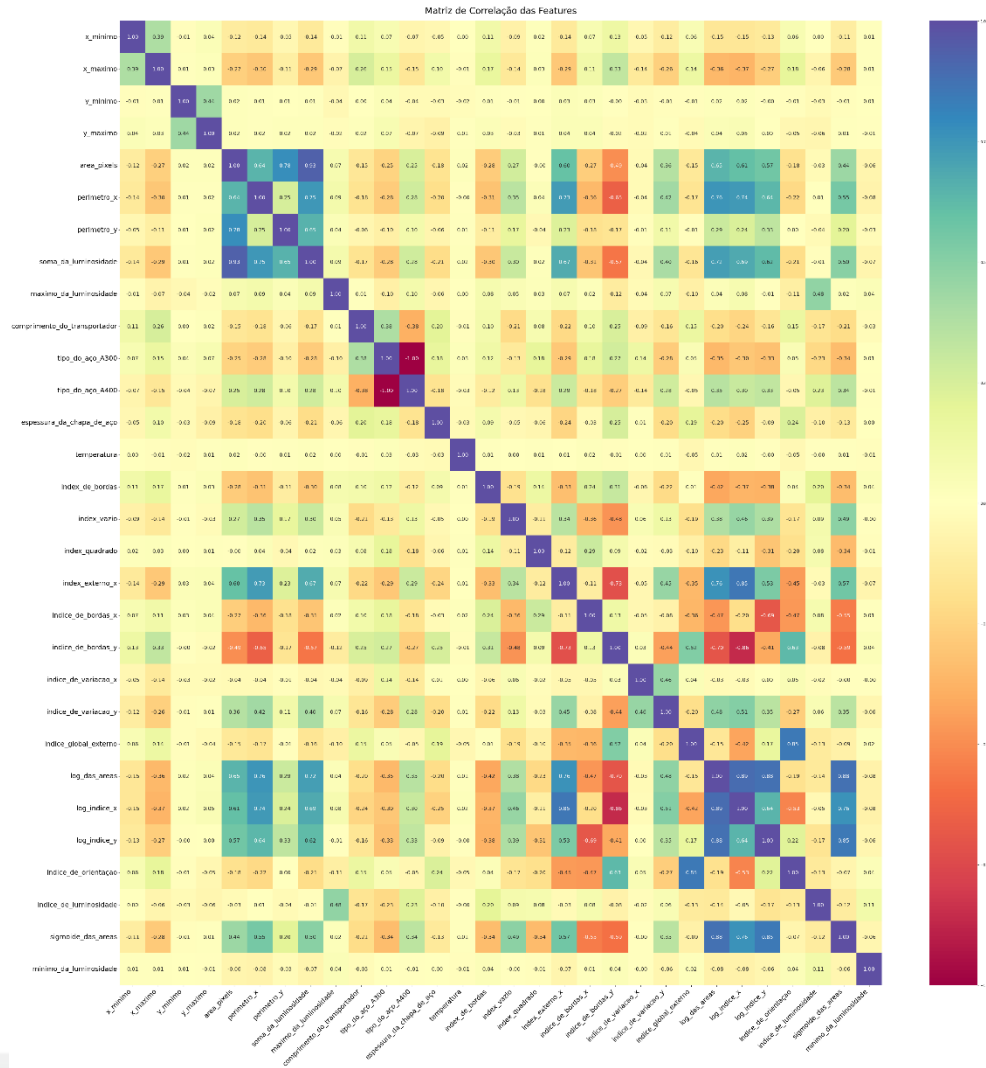
20	indice_de_variacao_x	3389 non-null	float64
21	indice_de_variacao_y	3389 non-null	float64
22	indice_global_externo	3389 non-null	float64
23	log_das_areas	3389 non-null	float64
24	log_indice_x	3389 non-null	float64
25	log_indice_y	3389 non-null	float64
26	indice_de_orientacao	3389 non-null	float64
27	indice_de_luminosidade	3389 non-null	float64
28	sigmoide_das_areas	3389 non-null	float64
29	minimo_da_luminosidade	3389 non-null	int64
30	falha_1	3389 non-null	int64
31	falha_2	3389 non-null	int64
32	falha_3	3389 non-null	int64
33	falha_4	3389 non-null	int64
34	falha_5	3389 non-null	int64
35	falha_6	3389 non-null	int64
36	falha_outros	3389 non-null	int64

```
dtypes: float64(16), int64(21)
```

```
memory usage: 979.8 KB
```

df.info() com tratamento

# Visualização e interpretação dos dados



Valor de	(+ ou -)	Interpretação
0.00 a 0.19		Uma correlação bem fraca
0.20 a 0.39		Uma correlação fraca
0.40 a 0.69		Uma correlação moderada
0.70 a 0.89		Uma correlação forte
0.90 a 1.00		Uma correlação muito forte

## Coeficiente de Pearson

```
1 #Resolvi preservar os tipos de aço A300 e A400,
2 #Apesar de poder eliminar um dos dois.
3 df = df.drop(['log_indice_x', 'sigmoide_das_areas'], axis=1)
4 df = df.drop(['indice_de_orientacao', 'log_indice_y'], axis=1)
5 df = df.drop(['soma_da_luminosidade'], axis=1)
6
```

# Visualização e interpretação dos dados



Dados Discrepantes			
Variável	Outliers	Total	%
soma da luminosidade	679	3389	20.04
index externo x	650	3389	19.18
area pixels	641	3389	18.91
perimetro x	631	3389	18.62
perimetro y	545	3389	16.08
log indice x	396	3389	11.68
maximo da luminosidade	386	3389	11.39
minimo da luminosidade	364	3389	10.74
espessura da chapa de aço	336	3389	9.91
y minimo	335	3389	9.88
indice de luminosidade	221	3389	6.52
x minimo	185	3389	5.46
y maximo	184	3389	5.43
indice de variacao x	150	3389	4.43
indice de variacao y	75	3389	2.21

log das areas	60	3389	1.77
index vazio	29	3389	0.86
temperatura	19	3389	0.56
log indice y	4	3389	0.12
indice de bordas y	2	3389	0.06
comprimento do transportador	0	3389	0.00
x maximo	0	3389	0.00
index de bordas	0	3389	0.00
index quadrado	0	3389	0.00
indice de bordas x	0	3389	0.00
indice global externo	0	3389	0.00
indice de orientacao	0	3389	0.00
sigmoide das areas	0	3389	0.00

Os dados discrepantes não foram removidos.

# Treinamento dos classificadores

```
29
30 #Dividir os dados em treino (75%) e teste (25%).|
31 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 42)
32
```



	Classificador	Acurácia	Precisão	Recall	F1-Score
0	KNN	0.584906	0.587348	0.584906	0.584557
1	SVM	0.660377	0.682522	0.660377	0.632082
2	Árvore de Decisão	0.586085	0.594325	0.586085	0.588794
3	Floresta Aleatória	0.674528	0.687507	0.674528	0.667343
4	Regressão Logística	0.672170	0.681907	0.672170	0.665354
5	Naive Bayes	0.521226	0.611662	0.521226	0.506888
6	AdaBoost	0.555425	0.552818	0.555425	0.488895





# Evolução do Modelo

- Obter um DataFrame consistente com, no mínimo, 100 mil registros.
- Realizar validação cruzada (k-fold).
- Criar estrutura para a seleção de hiperparâmetros.
- Testar a remoção de colunas com menor correlação de Pearson e avaliar as métricas.
- Refinar o modelo sucessivamente até tornar as métricas ótimas e as ferramentas de aferição de qualidade igualmente ótimas em função dos parâmetros desejados.
- Integrar a equipe técnica do patrocinador do modelo ao desenvolvimento do classificador para torná-lo ergonômico e homologável na indústria.