

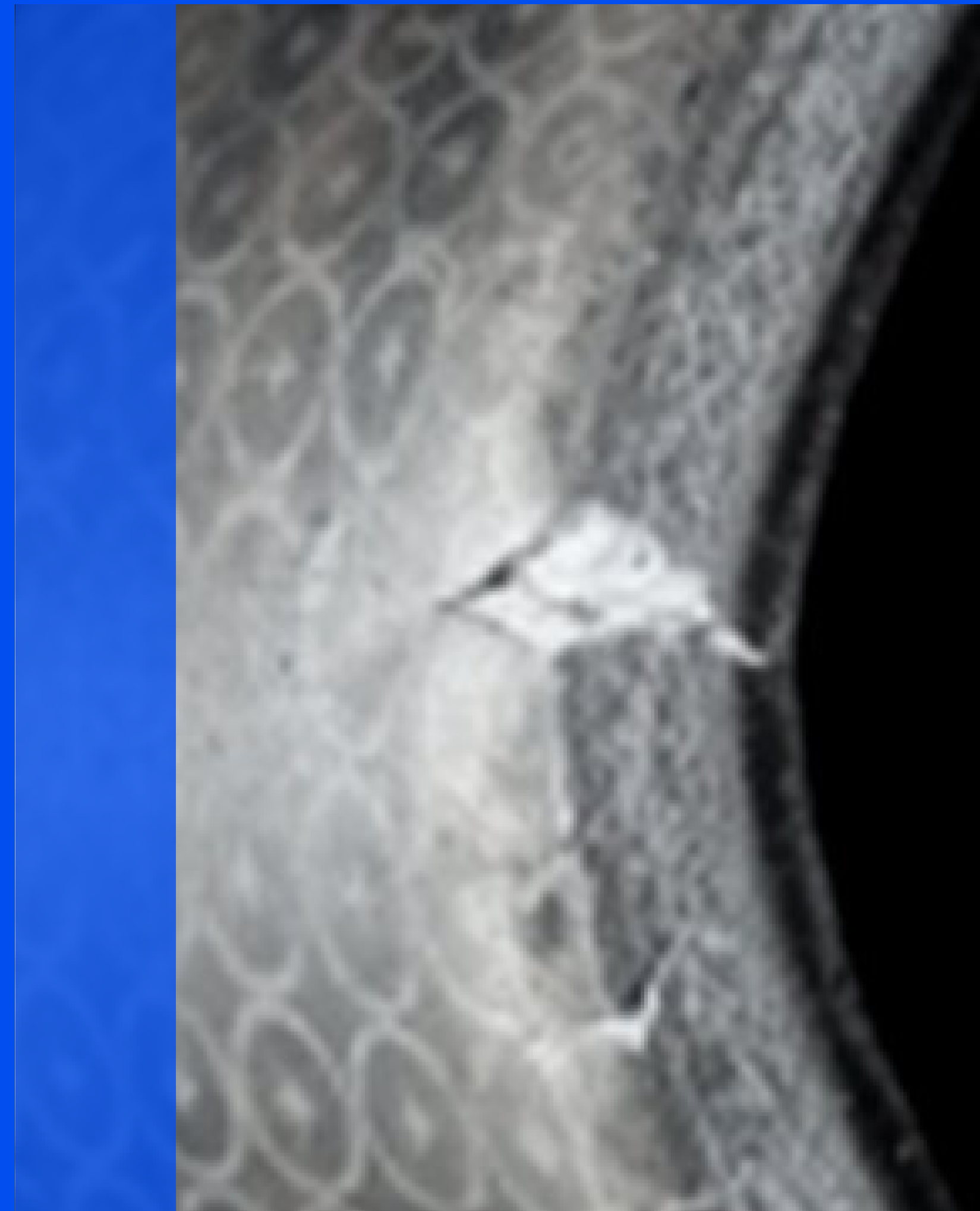
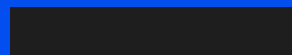


MACHINE LEARNING

EINSTEIN GUSTAVO



CLASSIFICAÇÃO DE FALHAS EM CHAPAS DE AÇO



CARACTERÍSTICAS DO PROBLEMA:



DATA SET
MULTIVARIÁVEL



ATRIBUTOS INTEIROS
E REAIS

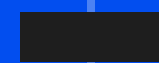


TAREFA PRINCIPAL:
CLASSIFICAÇÃO

TIPOS DE FALHAS A SEREM CLASSIFICADAS



PASTRY



Z SCRATCH



K SCATCH



STAINS



DIRTINESS



BUMPS



OTHER FAULTS

VARIÁVEIS

INDEPENDENTES

| | | | | |
|-----------------------|-----------------------|-----------------------|---|-------------------|
| X_MINIMUM | X_MAXIMUM | Y_MINIMUM |  | |
| X_MINIMUM | X_MAXIMUM | Y_MINIMUM | | |
| PIXELS_AREAS | Y_MAXIMUM | X_PERIMETER | | |
| Y_PERIMETER | SUM_OF_LUMINOSITY | MINIMUM_OF_LUMINOSITY | | LOGOFAREAS |
| MAXIMUM_OF_LUMINOSITY | LENGTH_OF_CONVEYER | TYPEOFSTEEL_A300 | | LOG_X_INDEX |
| TYPEOFSTEEL_A400 | STEEL_PLATE_THICKNESS | EDGES_INDEX | | LOG_Y_INDEX |
| EMPTY_INDEX | SQUARE_INDEX | OUTSIDE_X_INDEX | | ORIENTATION_INDEX |
| EDGES_X_INDEX | EDGES_Y_INDEX | OUTSIDE_GLOBAL_INDEX | | LUMINOSITY_INDEX |
| | | | | SIGMOIDOFAREAS |
| | | | | |

01. ESTUDO COMPARATIVO

Acurácia
desejada

02. EDA

Definir modelo
de classificador

METODOLOGIA

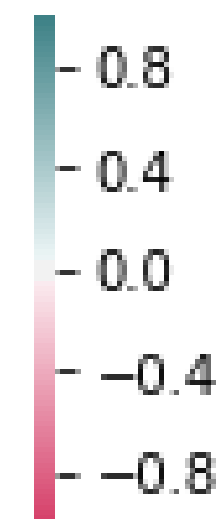
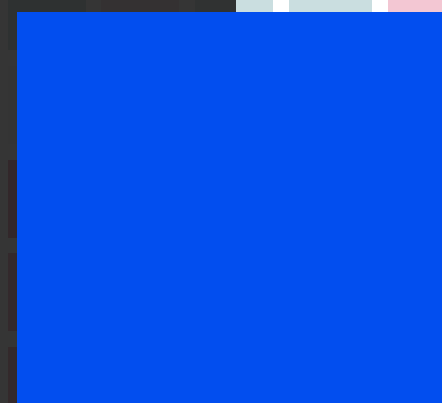
03. APLICAÇÃO

Implementação
e validação do
modelo

PESQUISA COMPARATIVA

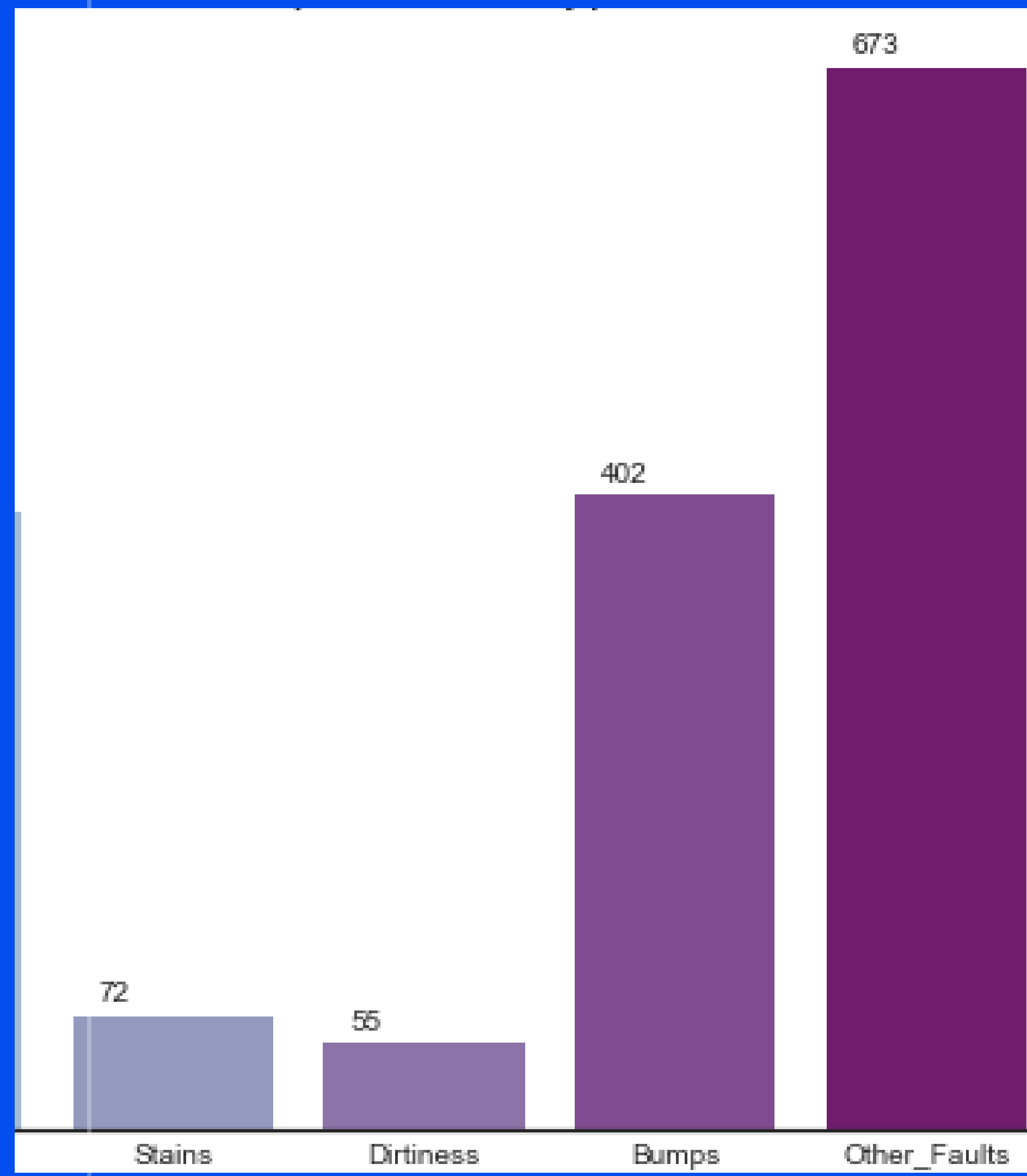
BASEADO NOS POUCOS
REPOSITÓRIOS DISPONÍVEIS A
ACURÁCIA DESEJADA ESTARIA
ENTRE 45% E 75% DE
ASSERTIVIDADE.

FOI ADOTADO ENTÃO UM VALOR DE
60% DE ACURÁCIA AO MENOS PARA
CONSIDERAR O MODELO UM SUCESSO.

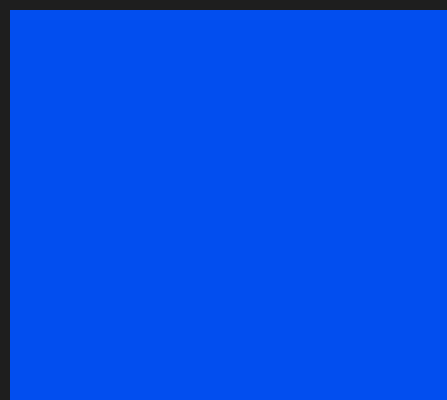


ANÁLISE EXPLORATÓRIA DE DADOS

—



PROBLEMAS IDENTIFICADOS NO DATASET



QUANTIDADE
DESBALANCEADA DE
AMOSTRAS DE CADA
TIPO DE FALHA



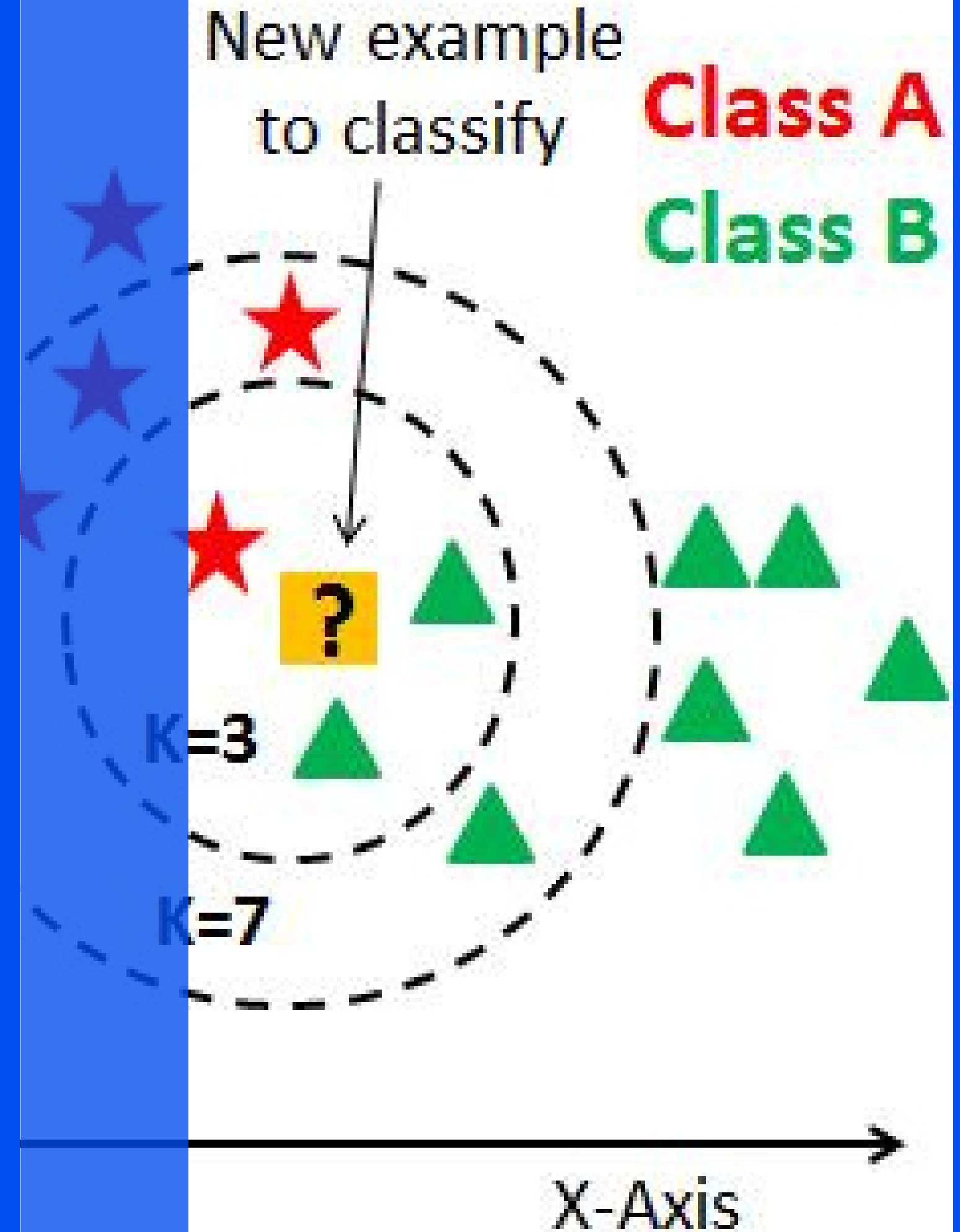
QUANTIDADE MUITO
GRANDE DE
VARIÁVEIS COM
POUCA
INTERDEPENDENCIA




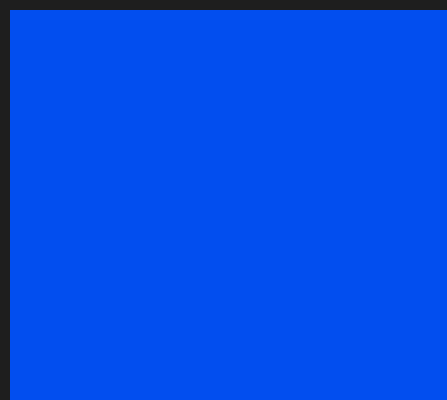
DIMENSÃO DOS
DADOS MUITO
DISTOANTES

K-NEAREST NEIGHBORS


DEVIDO O DATASET SER DE VARIÁVEIS COM
POUCA INTERDEPENDÊNCIA E DEVIDO À
SIMPLICIDADE DE IMPLEMENTAÇÃO, ESTE
FOI O MODELO DE CLASSIFICADOR
ESCOLHIDO.




RESULTADOS



DEVIDO À GRANDEZA
MUITO DIFERENTE
ENTRE OS
PARÂMETROS, UMA
NORMALIZAÇÃO SE
MOSTROU ESSENCIAL



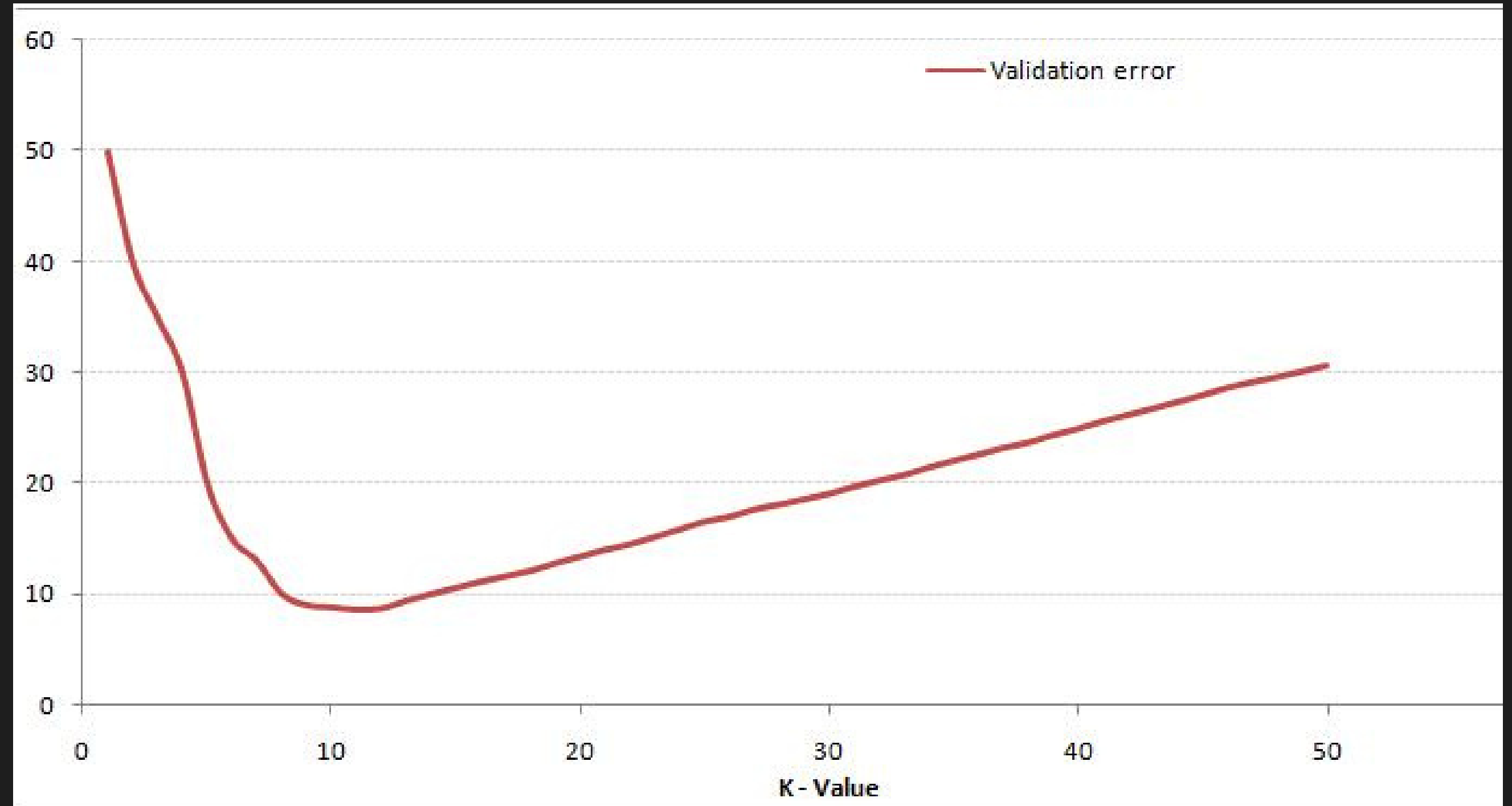
DURANTE A ANÁLISE
PCA, A MUDANÇA
PARA DIMENSÕES
MENORES NÃO
RESULTOU EM MAIOR
PRECISÃO, ENTÃO FOI
CONSIDERADO
APENAS A ROTAÇÃO,
E MANTEVE-SE OS 27
PARÂMETROS



A PRESENÇA DE UMA
CLASSIFICAÇÃO
GENÉRICA CUJO
NÚMERO DE
AMOSTRAS É MUITO
SIGNIFICATIVO
PREJUDICOU O
MODELO

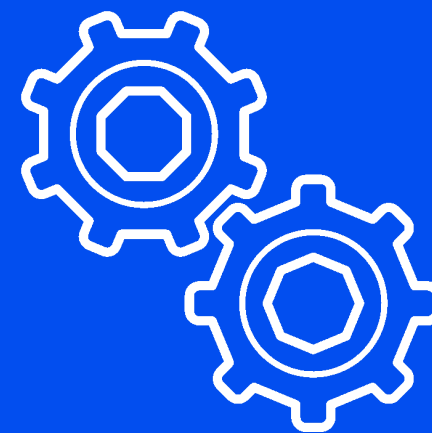
RESULTADOS

O NÚMERO DE
VIZINHOS MOSTROU
NÃO MELHORAR
MUITO O RESULTADO
APÓS 5 VIZINHOS,
ESTANDO OTIMIZADO
DENTRO DO
INTERVALO DE 5 A 10
VIZINHOS

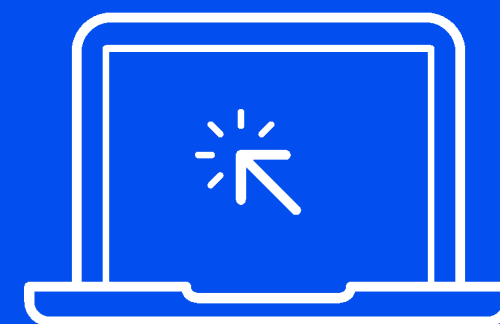


CONCLUSÃO

UMA VEZ QUE O OBJETIVO FOI ALCANÇADO (ACURÁCIA DE PELO MENOS 60%), A ESCOLHA DO KNN SE MOSTROU CORRETA E RESOLVEU O PROBLEMA



ACURÁCIA DO KNN INICIALMENTE: $(50 \pm 2)\%$



ACURÁCIA DO KNN APÓS NORMALIZAÇÃO DOS DADOS: $(65 \pm 2)\%$



ACURÁCIA DO KNN APÓS NORMALIZAÇÃO DOS DADOS E ANÁLISE DE COMPONENTES PRINCIPAIS: $(70 \pm 2)\%$