

DiabeTech

Modelo SVM desenvolvido para identificar e gerar estatísticas sobre pacientes com risco de desenvolver diabetes, além de gerar números sobre casos diagnosticados e não diagnosticados.



INTRODUÇÃO AO PROJETO

A diabetes é uma condição crônica que afeta mais de 530 milhões de adultos no mundo, sendo 16 milhões no Brasil. Nesse cenário, a Inteligência Artificial, especialmente o Aprendizado de Máquina, mostra-se promissora para apoiar decisões clínicas por meio de modelos preditivos baseados em dados.

OBJETIVO

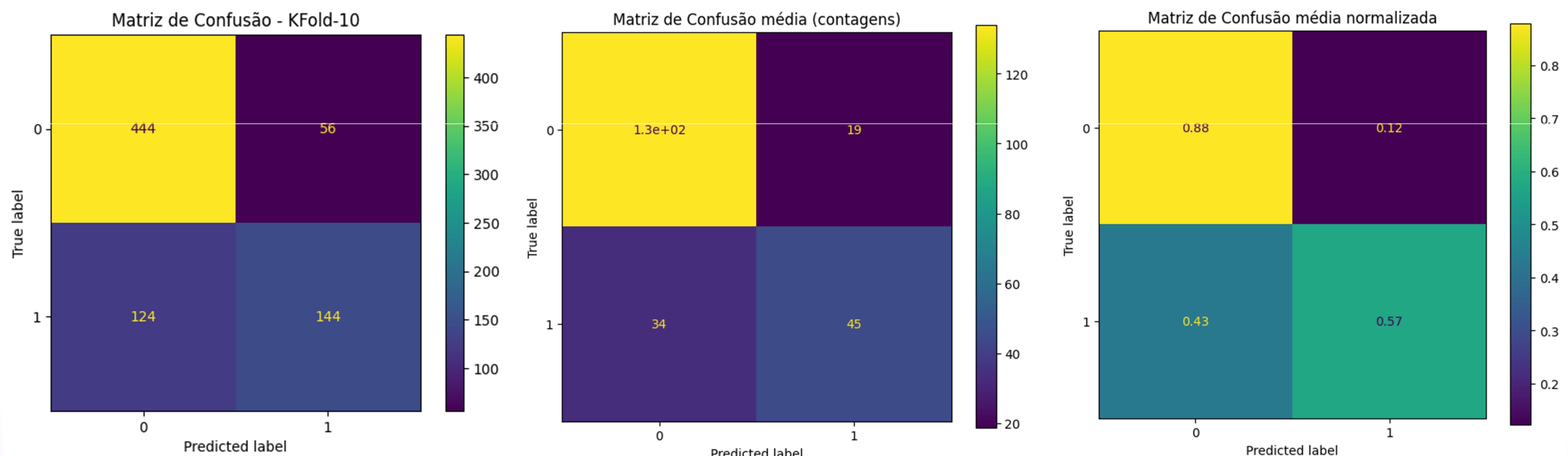
Comparar como o classificador SVM se sai usando duas formas diferentes de testar seu desempenho: KFold-10 e ShuffleSplit (com 30% dos dados para teste, repetido 10 vezes). Para isso, vamos analisar duas coisas: a acurácia (que mostra quantas vezes o modelo acerta) e a matriz de confusão (que mostra onde ele acerta e onde erra em cada classe).

METODOLOGIA

Foi obtido o dataset Pima Indians Diabetes Database (768 amostras, 8 atributos numéricos), as variáveis foram padronizadas usando o StandardScaler. O classificador usado foi o SVM, com os hiperparâmetros escolhidos pelo GridSearchCV.

Foram usados dois métodos para testar o modelo: (i) o primeiro dividiu os dados em 10 partes iguais e testou uma parte de cada vez; (ii) o segundo escolheu 30% dos dados de forma aleatória para testar, repetindo isso 10 vezes.

RESULTADOS OBTIDOS



Apesar da semelhança nos valores globais, a matriz de confusão mostrou diferenças entre os métodos. No KFold-10 houve tendência a mais falsos negativos, enquanto no ShuffleSplit a variação entre falsos positivos e negativos foi maior. Isso indica que a escolha do esquema de validação pode afetar a consistência na classificação de casos minoritários.

PESQUISADORES/DESENVOLVEDORES:
Graziela Pereira de Oliveira, Gustavo Di Risió,
Iago Lucas Fernandes de Faria, Rafael Souza
Santana e Rayane Da Luz Barbosa.
ORIENTADOR:
Fabiano Bezerra Menegidio.



Escaneie o QR code
para acessar o nosso
GitHub.