



## TECHNICAL REPORT

Aluno: Ingrid Melo de Oliveira

### 1. Introdução

O dataset usado nesta análise possui informações importantes para a predição de diabetes em pacientes, com variáveis que representam dados clínicos e laboratoriais. O estudo tem como objetivo realizar a comparação do desempenho de dois modelos de classificação: K-Nearest Neighbors (KNN) e Regressão Logística, analisando métricas como precisão, recall, f1-score, matriz de confusão e AUC (Área sob a Curva ROC).

### 2. Observações

Foi observado durante a implementação um aviso de convergência apresentado pela Regressão Logística, que sugeriu a necessidade de aumentar o número de iterações ou escalar de dados. Mas, isso não foi prejudicial para se obter os resultados, que foram obtidos normalmente e analisados. Nenhum outro problema técnico foi encontrado.

### 3. Resultados e discussão

#### 3.1 K-Nearest Neighbors (KNN)

- **Matriz de Confusão**

[132 18]

[49 32]

**132** Verdadeiros Negativos (TN): Classificados corretamente como **não diabéticos (0)**.

**18** Falsos Positivos (FP): Classificados incorretamente como **diabéticos (1)**.

**49** Falsos Negativos (FN): Classificados incorretamente como **não diabéticos (0)**.

**32** Verdadeiros Positivos (TP): Classificados corretamente como **diabéticos (1)**.



- **Métricas do Knn**

<i>Métricas</i>	<i>Classe 0 - Negativos</i>	<i>Classe 1 - Positivos</i>
<i>Precisão</i>	73%	64%
<i>Recall</i>	88%	40%
<i>F1-score</i>	80%	49%
<i>Acurácia</i>	71%	-

*O modelo com melhor desempenho na classe 0 (não diabético), foi o KNN que apresentou uma acurácia de 71%, e obteve um recall de 88%. No entanto, para a classe 1 (diabético), o recall foi apenas 40%, isso indica que o modelo apresentou dificuldades na identificação correta de pacientes com diabetes.*

### **3.2 Regressão Logística**

- **AUC-ROC: 0.8360**

*Mostra que o modelo tem boa capacidade de separação entre as classes.*

- **Matriz de Confusão**

*[129 21]*

*[39 42]*

**129** Verdadeiros Negativos (TN).

**21** Falsos Positivos (FP).

**39 Falsos Negativos (FN).**

**42 Verdadeiros Positivos (TP).**



- **Métricas da Regressão Logística**

<b>Métricas</b>	<b>Classe 0 - Negativos</b>	<b>Classe 1 - Positivos</b>
<i>Precisão</i>	77%	67%
<i>Recall</i>	86%	52%
<i>F1-score</i>	81%	58%
<i>Acurácia</i>	74%	-

*O modelo de Regressão Logística apresentou uma acurácia de 74%, superando ligeiramente o KNN. O modelo apresentou um AUC de 0.836, isso indica uma melhor capacidade de separar as classes. A precisão e o recall para a classe 1 (diabético) foram maiores que as do KNN, com um recall de 52% e uma precisão de 67%, o que indica um melhor equilíbrio entre as métricas.*

#### **4. Conclusões**

*De acordo com os resultados a Regressão Logística teve um melhor desempenho geral do que o KNN, pois apresentou maior acurácia, melhor equilíbrio entre recall e precisão e um AUC mais alto. Portanto isso indica que, para este dataset, a Regressão Logística é a melhor escolha, por ser mais eficiente para a predição da presença de diabetes.*

#### **5. Próximos passos**

- *Testar outras abordagens de padronização ou normalização dos dados para que*

*a convergência da Regressão Logística seja melhor.*

- *Tentar usar outros modelos de aprendizado de máquina, como Random Forest e SVM, para avaliar o desempenho e verificar se apresentam melhores resultados.*
- *Ajustar hiperparâmetros dos modelos para otimizar suas performances.*