Resultados das regressões utilizadas nos dados e na amostra dada:

Regressao Linear All\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.4251 | 14.6500 | 0.2541 |

Regressao Linear test\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 10.5341 | 12.2838 | 0.5206 |

Uma imagem com texto, captura de ecrã, Gráfico, file

Os conteúdos gerados por IA poderão estar incorretos.

Resultados da DecisionTreeRegressor All\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.2820 | 14.4702 | 0.2723 |

Melhores parâmetros:

|  |  |  |  |
| --- | --- | --- | --- |
| 'max\_depth' | 'max\_features' | 'min\_samples\_leaf' | 'min\_samples\_split' |
| 10 | None | 9 | 25 |

Resultados da DecisionTreeRegression test\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 10.9899 | 12.8290 | 0.4771 |

Melhores parâmetros:

|  |  |  |  |
| --- | --- | --- | --- |
| 'max\_depth' | 'max\_features' | 'min\_samples\_leaf' | 'min\_samples\_split' |
| 10 | None | 7 | 7 |

Uma imagem com texto, captura de ecrã, Gráfico, diagrama

Os conteúdos gerados por IA poderão estar incorretos.RandomForestRegression all\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.2436 | 14.4092 | 0.2784 |

Melhores parâmetros:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| max\_depth | max\_features | min\_samples\_leaf | min\_samples\_split | n\_estimators |
| None | sqrt | 4 | 10 | 100 |

Resultados da Random Forest Regression test\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.4008 | 13.1006 | 0.4547 |

Melhores parâmetros:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| max\_depth | max\_features | min\_samples\_leaf | min\_samples\_split | n\_estimators |
| 5 | sqrt | 1 | 2 | 100 |

Uma imagem com texto, diagrama, Gráfico, captura de ecrã

Os conteúdos gerados por IA poderão estar incorretos.

Resultados da Regressão K Neighbors all\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.4945 | 14.8440 | 0.2342 |

Melhores parâmetros:

|  |  |  |
| --- | --- | --- |
| Knn\_\_n\_neighbors | knn\_\_p | knn\_\_weights |
| 10 | 1 | ‘uniform’ |

Resultados da Regressão K Neighbors test\_data:

|  |  |  |
| --- | --- | --- |
| MAE | RMSE | R² |
| 11.7929 | 14.4576 | 0.3359 |

Melhores parâmetros:

|  |  |  |
| --- | --- | --- |
| Knn\_\_n\_neighbors | knn\_\_p | knn\_\_weights |
| 7 | 1 | ‘uniform’ |

Conclusões:

Com o Objetivo (O2) teve como principal foco de prever a idade a partir de atributos clínicos disponíveis no ficheiro csv disponibilizado. Foram utilizados e comparados quatro modelos de regressão: **Regressão Linear**, **Árvore de Decisão (Decision Tree)**, **Random Forest** e **k-Nearest Neighbors (k-NN).**

**Regressão Linear**: Modelo simples, desempenho modesto. Indica que há relação linear fraca entre os atributos e a idade.

**Decision Tree Regressor**: Um leve ganho sobre a regressão linear. Capta não-linearidades, mas sofre com overfitting.

**Random Forest Regressor** (melhor desempenho): Modelo mais robusto. Melhor equilíbrio entre viés e variância. Porém, a variância explicada ainda é baixa.

**k-Nearest Neighbors Regressor**: Desempenho inferior. Provavelmente afetado pelo tamanho do data\_set e pela alta dimensionalidade.

Em todos os modelos, os **erros médios absolutos (MAE)** estão entre **11 e 12 anos**, o que representa uma margem significativa de imprecisão ao estimar a idade. Além disso, os valores de **R² abaixo de 0.30** indicam que os modelos explicam menos de 30% da variação da idade com base nos atributos disponíveis.

Com os dados disponíveis, os resultados dos modelos revelam-nos que os atributos disponíveis não são suficientemente informativos para prever a idade de forma confiável.

O desempenho semelhante entre a regressão linear e modelos mais complexos (como árvores e florestas) reforça que **não há relações complexas relevantes** entre os dados e a idade — ou que tais relações são mascaradas por ruído ou co-linearidade.

O conjunto test\_data, com apenas 100 exemplos, apresentou métricas ligeiramente melhores em alguns modelos. No entanto, isso provavelmente se deve à menor variabilidade ou maior homogeneidade dentro desse conjunto, e **não pode ser interpretado como um ganho real de generalização**. Na prática, o desempenho em all\_data é mais representativo e confiável.

Para concluir, não é possível prever a idade com precisão usando apenas os atributos fornecidos. Embora existam correlações fracas, os modelos apresentam erro elevado e baixa explicabilidade.