**Metodologia**:  
Foram avaliados múltiplos modelos de regressão utilizando:

* 8 tipos de modelos (Linear, Ridge, Lasso, Árvore de Decisão, Random Forest, Gradient Boosting, SVR, KNN) que foram reduzidos para 3 depois de testes para ver qual o melhor
* Validação cruzada com otimização de hiperparâmetros
* Métricas: MAE (anos), R² (variância explicada)

**Resultados Principais**:

1. Comparação de Desempenho:

| **Modelo** | **COVID+ (R²)** | **COVID- (R²)** | **Melhores Parâmetros (COVID+)** | **Melhores Parâmetros (COVID-)** |
| --- | --- | --- | --- | --- |
| Regressão Linear | 0.09 | 0.13 | Nenhum | Nenhum |
| Random Forest | 0.10 | 0.16-0.17 | max\_depth=5, n\_estimators=150 | max\_depth=5, n\_estimators=100 |
| Gradient Boosting | 0.10 | 0.16 | learning\_rate=0.05, max\_depth=3 | learning\_rate=0.1, max\_depth=3 |
| Árvore de Decisão | 0.06 | 0.12 | max\_depth=5 | max\_depth=3 |

1. Margens de Erro:

* COVID+: MAE ≈10.7 anos (melhor modelo)
* COVID-: MAE ≈13.2 anos (melhor modelo)

**Principais Conclusões**:

1. **Desempenho dos Modelos**:

* Random Forest e Gradient Boosting mostraram desempenho marginalmente superior (R² 0.10-0.17)
* Todos os modelos apresentaram erros elevados (MAE >10 anos)
* As diferenças entre modelos foram inconsistentes (no pior caso Gradient Boosting superou Decision Tree por R² 0.10, mas tal não se repetiu consistentemenete)

1. **Impacto do Estado COVID**:

* Modelos performaram ligeiramente melhor para casos negativos
* Margens de erro maiores para previsões COVID-negativas (MAE +2.5 anos vs COVID+)

1. **Consistência de Parâmetros**:

* Parâmetros ótimos do Random Forest foram notavelmente consistentes:
  + max\_depth=5 para ambos os grupos
  + min\_samples\_leaf=1 na maioria das execuções
  + n\_estimators entre 50-200

**Limitações Críticas**:

1. **Restrições de Dados**:

* Amostras extremamente pequenas (COVID+: n=4, COVID-: n=3 nos dados de teste)
* Alta variabilidade nas métricas devido ao tamanho amostral
* Validação cruzada limitada a 2-3 divisões (5 para o dataset completo)

1. **Confiabilidade dos Modelos**:

* Valores de R² <0.2 indicam explicação de <20% da variância
* MAE >10 anos é clinicamente inaceitável
* Instabilidade na otimização de parâmetros

**Conclusões Finais**:

1. Com os dados de teste atuais:

* Não é possível prever com confiança a idade ao óbito para nenhum dos grupos
* Padrões observados podem ser artefactos da pequena amostra
* O desempenho ligeiramente superior de métodos ensemble sugere possíveis relações não-lineares (a validar com mais dados)

1. Interpretação Clínica:

* MAE >10 anos torna as previsões clinicamente inúteis
* Não foi estabelecida diferença significativa entre grupos COVID
* Análise de importância de variáveis não é confiável com esta amostra

**Recomendações**:

1. Essencial repetir a análise com o dataset completo (custom\_covid19.csv)
2. Para aplicação clínica:
   * Necessário incluir variáveis preditivas adicionais
   * Alvo: MAE <5 anos para utilidade prática

**Avaliação Global**:  
Os dados de teste disponíveis são insuficientes para construir modelos fiáveis de previsão de idade em pacientes falecidos com COVID-19. Embora a análise demonstre viabilidade metodológica, todos os modelos apresentaram desempenho marginal, com erros elevados.