

Projeto 2 de Inteligência Artificial - SIN 323

Kayky Cristof Eduardo Domingos Silva - 8118

¹Instituto de Ciências Exatas e Tecnológicas (IEP)
Universidade Federal de Viçosa (UFV), Rio Paranaíba, MG, Brasil

kayky.silva@ufv.br

Abstract. *This report describes the development of the second project for the Artificial Intelligence course SIN-323, taught by Professor Larissa Ferreira Rodrigues Moreira. The objective of this project is to analyze mortality from different diseases in Brazil, considering temporal patterns and the risk factors associated with these diseases. We used statistical techniques and machine learning to understand mortality trends over the years.*

Resumo. *Este relatório descreve como foi feito o segundo projeto da disciplina de Inteligência Artificial SIN-323 ministrada pela professora Larissa Ferreira Rodrigues Moreira, onde o objetivo desse projeto é analisar a mortalidade por diferentes doenças no Brasil, tendo em vista padrões temporais e os fatores de riscos associados a estas doenças, utilizamos técnicas de estatística e de aprendizado de máquina para conseguir compreender as tendências de mortalidade ao longo dos anos.*

1. Introdução

As doenças representam uma das principais causas de morte no Brasil e no mundo, gerando impactos na qualidade de vida da população. Doenças cardiovasculares representam uma parte significativa dessas mortes. No entanto, diversas outras doenças contribuem para este cenário, a análise dos dados de mortalidade por diferentes **CIDs (Classificação internacional de doenças)** fornece uma análise para prevenção de doenças. O principal objetivo deste projeto é realizar uma análise geral sobre a mortalidade por diferentes doenças no Brasil, identificando padrões temporais e os fatores de risco associados a estas mortes. Utilizaremos a base de dados do **Sistema de Informações sobre mortalidade (SIM) do Datasus**, e aplicaremos **técnicas estatísticas** e de **aprendizado de máquina** para compreender as tendências de mortalidade ao longo dos anos, com o objetivo de fornecer estratégias para as políticas públicas de saúde para a prevenção. Utilizamos a linguagem de programação **PYTHON** para fazer essas análises em combinação com sua vasta gama de bibliotecas para trabalhar com análise de dados.

2. Análise Exploratória dos Dados

2.1. Pré-processamento

Neste projeto, utilizamos diversas bases de dados coletadas do Datasus com o objetivo de identificar e analisar as principais variáveis associadas às causas de determinadas patologias. A análise abrangeu o período de 2020 a 2023, levando em consideração as regiões do país relacionadas a cada patologia, a faixa etária da população e o sexo dos indivíduos.

Inicialmente, realizamos a padronização dos dados para garantir que pudessem ser manipulados corretamente utilizando a biblioteca **Pandas**. Durante esse processo, também identificamos inconsistências e dados irrelevantes que poderiam distorcer a análise. Esses dados foram tratados e removidos conforme necessário.

Uma das primeiras observações feitas foi a presença de dados irrelevantes na base, que não são patologias propriamente ditas. O primeiro grupo identificado foi **088-092: Gravidez, Parto e Puerpério**, que se refere a estados fisiológicos naturais do ciclo reprodutivo feminino, englobando condições e complicações associadas à gravidez, parto e período pós-parto. O segundo grupo identificado foi **101-103: Sintomas, Sinais e Achados Anormais Clínicos e Laboratoriais, Não Classificados em Outra Parte (NCOP)**, que abrange sintomas, sinais e resultados de exames que não podem ser atribuídos diretamente a uma doença ou condição específica. Por fim, o terceiro grupo identificado foi **104-113: Causas Externas de Morbidade e Mortalidade**, que se refere a eventos externos, como acidentes e violência, e que não estão relacionados a condições internas do corpo humano.

Dado que esses dados não são relevantes para o objetivo da análise, decidimos removê-los da base e excluí-los do cálculo das causas totais de mortes. Após esse processo de limpeza, os dados foram devidamente pré-processados e padronizados para garantir sua compatibilidade com a biblioteca Pandas, permitindo uma análise mais eficiente e precisa.

2.2. Análise de dados: Resumo geral sobre os datasets

Assim, fizemos a análise geral de cada dataset, os números de linhas e colunas, os tipos de dados, os valores ausentes e duplicados, são informações valiosas, conhecer bem o dataset, para aplicar as técnicas de estatística e de machine learning posteriormente.

```
Resumo do Dataset: Mortalidade 2020 (Região)
- Número de Linhas: 117
- Número de Colunas: 8

Tipos de Dados:
Causa - CID-BR-10      object
1 Região Norte         float64
2 Região Nordeste      float64
3 Região Sudeste       float64
4 Região Sul           float64
5 Região Centro-Oeste  float64
Total                  int64
Ano                    int64
dtype: object
```

Figura 1. Dados sobre mortalidade por região do ano de 2020

```
Valores Ausentes:
Causa - CID-BR-10      0
1 Região Norte         2
2 Região Nordeste      2
3 Região Sudeste       1
4 Região Sul           2
5 Região Centro-Oeste  5
Total                  0
Ano                    0
dtype: int64
```

Figura 2. Dados sobre mortalidade por região do ano de 2020

2.3. Análise de dados: Estatísticas descritivas

Por fim, realizamos as estatísticas descritivas de cada conjunto de dados, com o objetivo de organizar e apresentar de forma clara as características principais de cada um. Filtramos apenas os dados numéricos para esta análise e calculamos a média, o desvio padrão, os valores mínimo e máximo, além dos quartis (25%, 50%, 75%) e também valores extremos (mínimo e máximo). Esses valores fornecem informações importantes sobre a centralização e a dispersão dos dados.

```

Estatísticas Descritivas: Sexo Masc2023 (Sexo)

Resumo Estatístico:
  Óbitos p/Residênc
count      113.000000
mean       17984.495575
std        64736.315785
min         1.000000
25%        733.000000
50%       3315.000000
75%       11077.000000
max       640851.000000

Valores Extremos:
- Óbitos p/Residênc: Mínimo = 1, Máximo = 640851
-----

```

Figura 3. Estatísticas descritivas sobre o dataset de mortalidade do sexo masculino em 2023.

3. Análise Temporal

3.1. Combinação dos Dados

O primeiro passo para a realização da análise temporal foi a combinação de quatro conjuntos de dados. Esses conjuntos continham informações sobre o número de mortes relacionadas a diferentes CIDs, organizadas por região e ano. A combinação dos dados foi essencial para permitir uma análise abrangente da mortalidade ao longo do período estudado (2020-2023). Após a consolidação dos dados, utilizamos a biblioteca **matplotlib** para construir o gráfico apresentado a seguir, que ilustra as tendências temporais observadas.

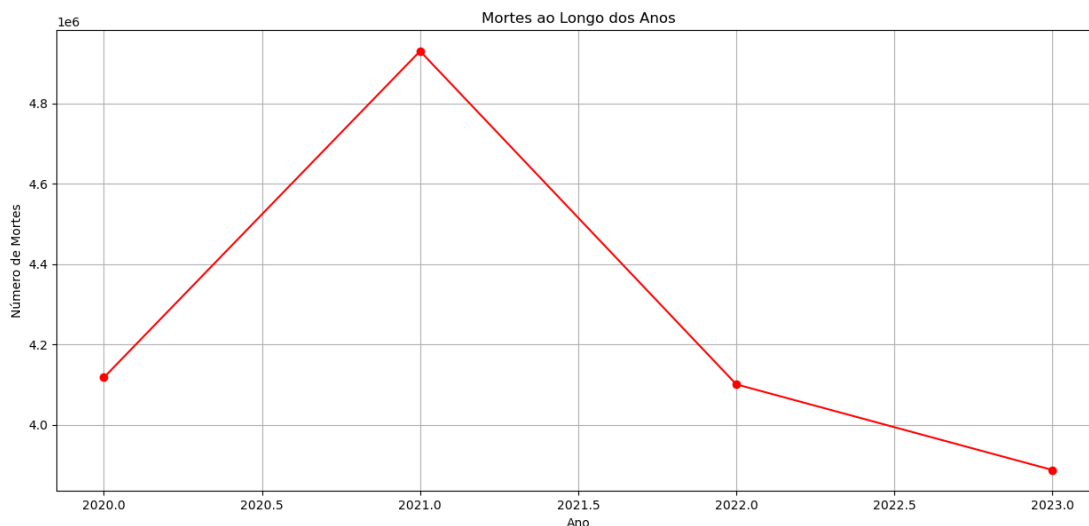


Figura 4. Gráfico que representa a taxa de mortalidade durante o período(2020-2023)

Contudo, já identificamos algumas informações importantes: houve um aumento significativo no número de mortes durante o ano de 2021, seguido por uma queda nos anos subsequentes. No entanto, ainda é necessário aprofundar a análise para compreender melhor os dados.

Quais foram as causas desse aumento abrupto na taxa de mortalidade de um ano para o outro? Quais CIDs foram responsáveis pelo maior número de mortes ao longo do

período analisado? Quais foram as principais causas de morte em cada região? Variáveis como sexo e faixa etária devem ser consideradas como fatores relevantes?

Essas questões são fundamentais para a análise temporal que será explorada em maior detalhe ao longo do estudo.

3.2. Mortes por diferentes CIDs ao longo do período por região

Com os dados agrupados, foi possível identificar as principais CIDs por região ao longo do período analisado. Realizamos uma análise detalhada de cada região do Brasil, destacando as quatro principais CIDs responsáveis pelas mortes. A seguir, apresentamos os gráficos de cada região:

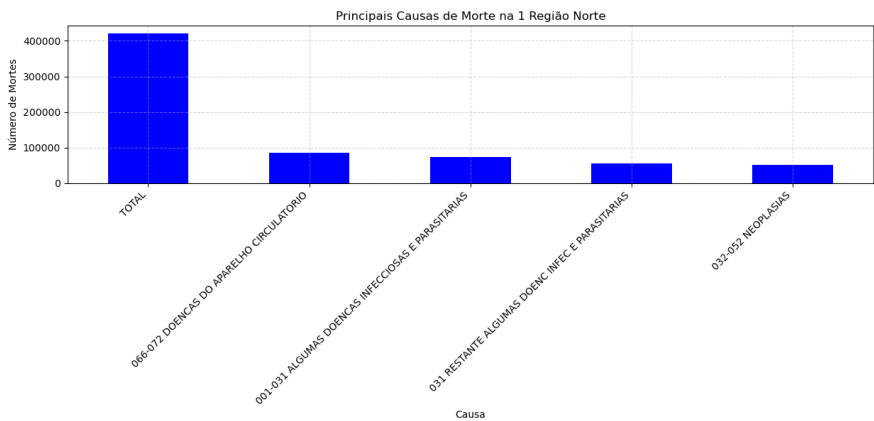


Figura 5. Taxa de mortalidade durante o período (2020-2023) na região Norte.

A primeira região analisada foi a região Norte. Conforme observado, a principal causa de mortes está relacionada a doenças do aparelho circulatório, seguida por doenças infecciosas e parasitárias, e, por último, neoplasias (câncer).

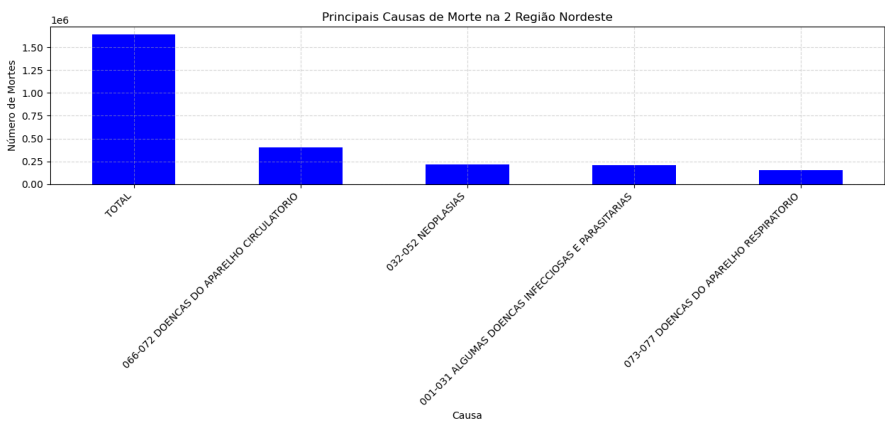


Figura 6. Taxa de mortalidade durante o período (2020-2023) na região Nordeste.

A segunda região analisada foi a região Nordeste. Assim como na região Norte, as doenças do aparelho circulatório foram a principal causa de mortes, seguidas por neoplasias (câncer), doenças infecciosas e parasitárias, e, por fim, doenças do aparelho respiratório.

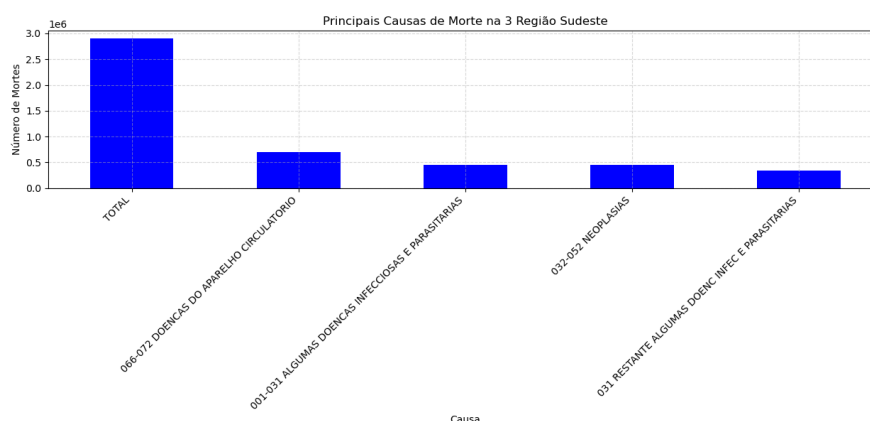


Figura 7. Taxa de mortalidade durante o período (2020-2023) na região Sudeste.

A terceira região analisada foi a região Sudeste. De maneira semelhante, as doenças do aparelho circulatório foram a principal causa de mortes, seguidas por doenças infecciosas e parasitárias, neoplasias (câncer) e outros grupos de doenças relacionadas a infecções.

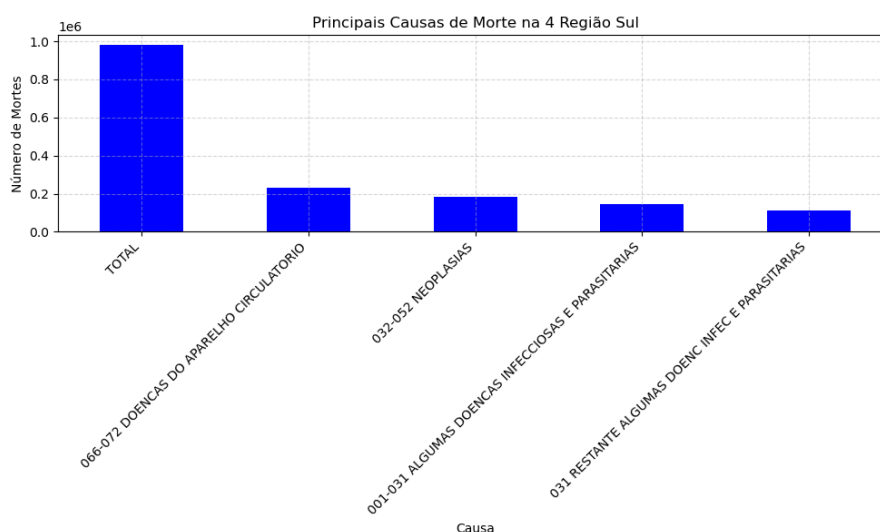


Figura 8. Taxa de mortalidade durante o período (2020-2023) na região Sul.

A quarta região analisada foi a região Sul. Nessa região, as doenças do aparelho circulatório também se destacaram como a principal causa de mortes, seguidas por neoplasias (câncer), doenças infecciosas e parasitárias, e outros grupos relacionados a infecções.

Por fim, a última região analisada foi a região Centro-Oeste. Assim como nas demais regiões, as doenças do aparelho circulatório foram a principal causa de mortes, seguidas por doenças infecciosas e parasitárias, outros grupos relacionados a infecções e, por último, neoplasias (câncer).

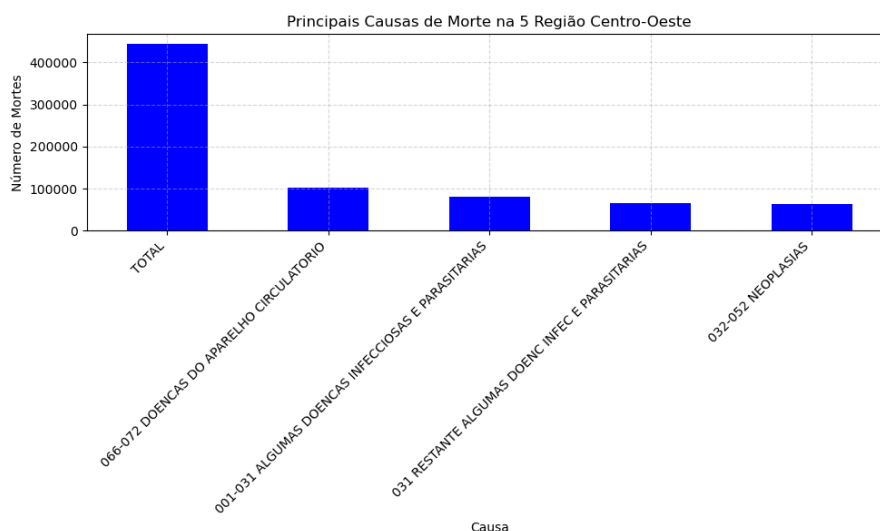


Figura 9. Taxa de mortalidade durante o período (2020-2023) na região Centro-Oeste.

Com base nas análises realizadas, concluímos que, durante o período estudado, o principal grupo de doenças que mais ocasionou mortes foi o das doenças do aparelho circulatório. Mas por que isso ocorre? Essas enfermidades estão entre as que mais causam mortes em todo o mundo devido a diversos fatores, como a alta prevalência de riscos, o diagnóstico tardio e o impacto direto nos órgãos vitais. Além disso, o envelhecimento da população está diretamente relacionado ao aumento dessas doenças, enquanto o estilo de vida moderno, caracterizado pelo sedentarismo, tabagismo e outros hábitos prejudiciais, também contribui significativamente para a mortalidade.

Outro ponto relevante é o número expressivo de mortes relacionadas a infecções e parasitas, especialmente em regiões mais vulneráveis. Esses óbitos são consequência de uma combinação de fatores socioeconômicos, ambientais e biológicos. Entre os principais motivos, podemos destacar as baixas condições de saneamento básico e higiene, a prevalência de doenças endêmicas em áreas mais carentes e a desigualdade social, que impacta diretamente nesses índices de mortalidade. O impacto de pandemias, como a COVID-19, também contribuiu para o aumento desses números durante o período analisado.

Por fim, outra das principais causas de morte foram as neoplasias (câncer). As principais razões para isso incluem o diagnóstico tardio, já que muitos casos só são identificados em estágios avançados, quando as opções de tratamento são mais limitadas e as chances de cura são reduzidas. Além disso, fatores de risco como tabagismo, dieta pobre em nutrientes, sedentarismo e poluição também desempenham um papel importante. A complexidade do tratamento das neoplasias é outro ponto crítico, pois alguns tumores têm a capacidade de se metastatizar (espalhar-se para outros órgãos), dificultando ainda mais o manejo da doença.

Com base nesses dados, já temos algumas diretrizes para começar a discutir a prevenção dessas doenças e reduzir suas taxas de mortalidade.

3.3. Mortes por diferentes CIDs por faixa etária(2023)

Após está análise de mortes por região, fizemos o estudo de mortes por diferentes CIDs baseado em cada faixa etária no qual também é uma variável importante. Quais as principais causas de morte de acordo com a idade dos indivíduos? Quais os fatores de risco para cada faixa etária? Com base no dataset tivemos essas informações no qual iremos mostrar nos gráficos abaixo, discutir e tirar conclusões. Começaremos com as idades de forma crescente.

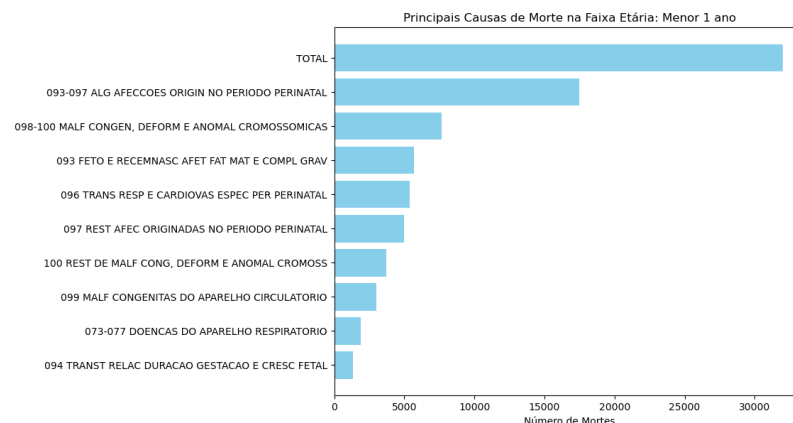


Figura 10. Taxa de mortalidade na faixa etária menor do que 1 ano.

Nessa faixa etária são basicamente recém-nascidos, as principais causas de morte são infecções durante o período pré-natal o (sistema auto-imune ainda não se desenvolveu), mal formações nos indivíduos que podem ser relacionados a problemas genéticos e doenças no sistema circulatório e respiratório, nessa idade cuidados pré-natais são de extrema importância.

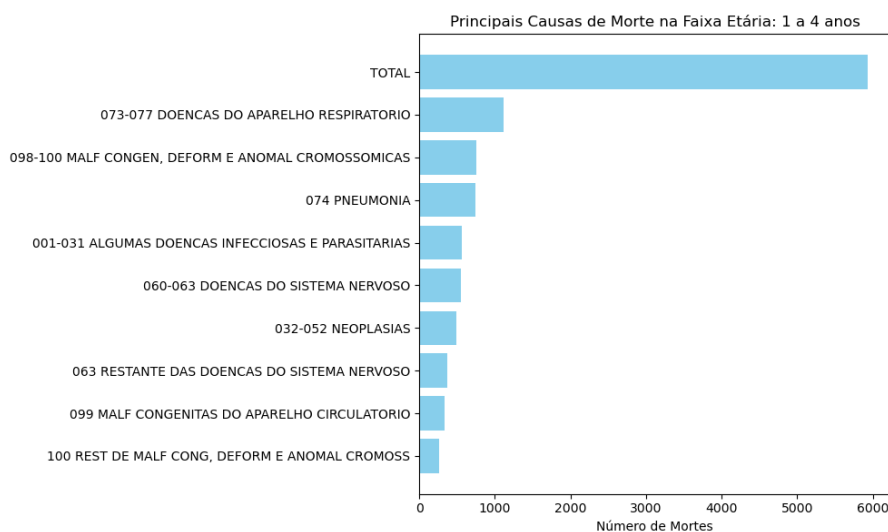


Figura 11. Taxa de mortalidade na faixa etária de 1 a 4 anos.

As causas de morte começam a se alterar entre crianças de 1 a 4 anos. A principal causa de óbito nessa faixa etária são as doenças do aparelho respiratório, uma vez que

as crianças têm maior exposição a ambientes externos, como creches e escolas, onde o contato com outros indivíduos aumenta o risco de infecções respiratórias. Em segundo lugar, estão as mal formações congênitas. Crianças que sobreviveram ao período pré-natal com mal formações podem enfrentar complicações à medida que crescem, o que pode levar ao óbito nessa faixa etária. Por fim, a pneumonia é uma das maiores ameaças à saúde infantil, especialmente em contextos de vacinação insuficiente.

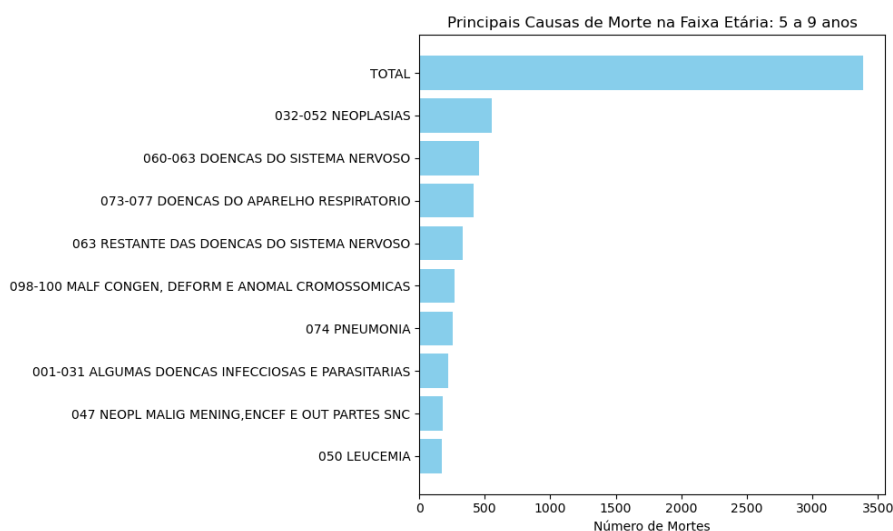


Figura 12. Taxa de mortalidade na faixa etária de 5 a 9 anos.

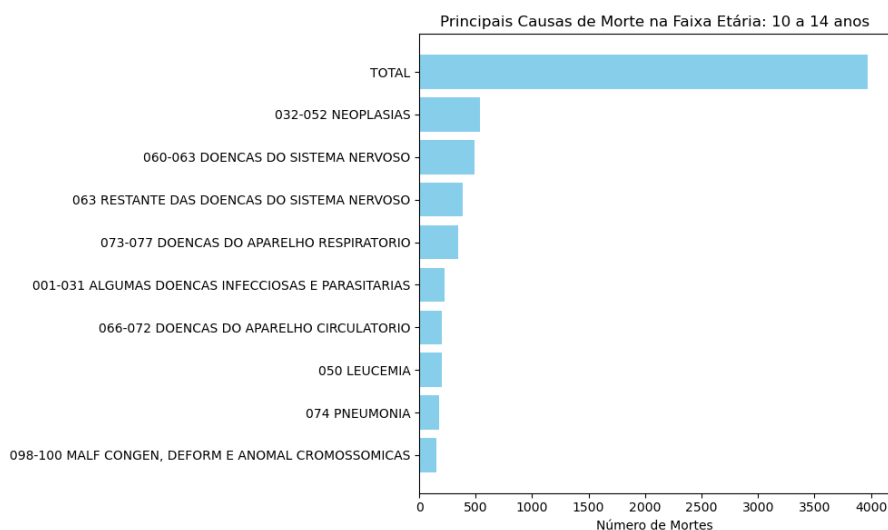


Figura 13. Taxa de mortalidade na faixa etária de 9 a 14 anos.

A partir destes gráficos, podemos identificar inferências importantes sobre determinadas CIDs e como elas impactam. A partir desta faixa etária, observa-se uma taxa crescente de mortes relacionadas a **neoplasias (câncer)**, que se torna a principal causa de morte até a fase de jovens adultos. O câncer surge devido a uma combinação de fatores **genéticos** e **ambientais**. Além disso, ele está intimamente relacionado ao crescimento celular, já que o câncer se origina a partir de mutações nas células. Essas mutações podem

ocorrer durante o desenvolvimento do indivíduo, o que explica a prevalência de cânceres em diferentes idades, incluindo a infância e a juventude. Para entender o motivo pelo qual uma doença causa um número significativo de mortes, é fundamental compreender seu funcionamento e os processos biológicos que a envolvem.

Além disso, as **doenças do sistema nervoso** representam uma parcela de mortes, sendo influenciadas por fatores **genéticos, infecções, lesões** e o **desenvolvimento contínuo** do sistema nervoso. O crescimento do sistema nervoso em desenvolvimento podem tornar as crianças e jovens mais suscetíveis a condições neurológicas graves. Algumas dessas doenças, como tumores cerebrais e distúrbios neurológicos, são desafiadores de diagnosticar, pois seus sintomas iniciais são sutis ou confundidos com outras condições. Além disso, o tratamento pode ser complexo e muitas vezes não eficaz se a condição for diagnosticada tardiamente.

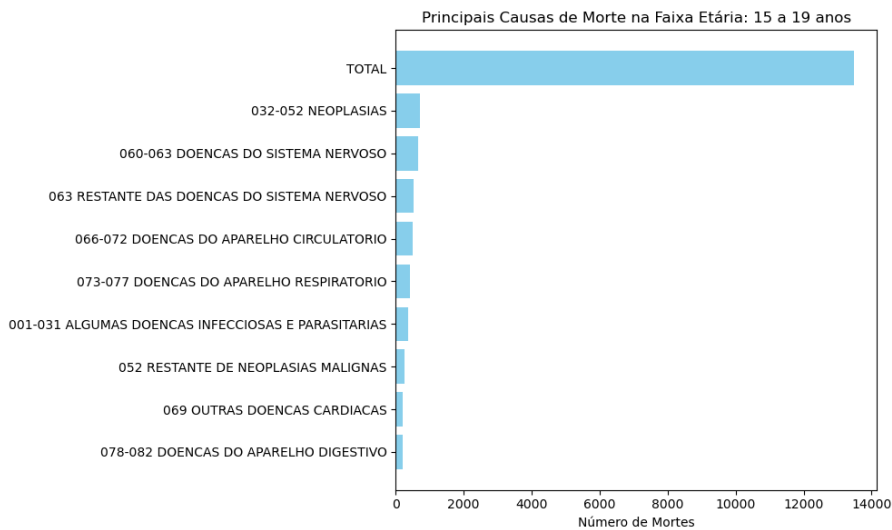


Figura 14. Taxa de mortalidade na faixa etária de 14 a 19 anos.

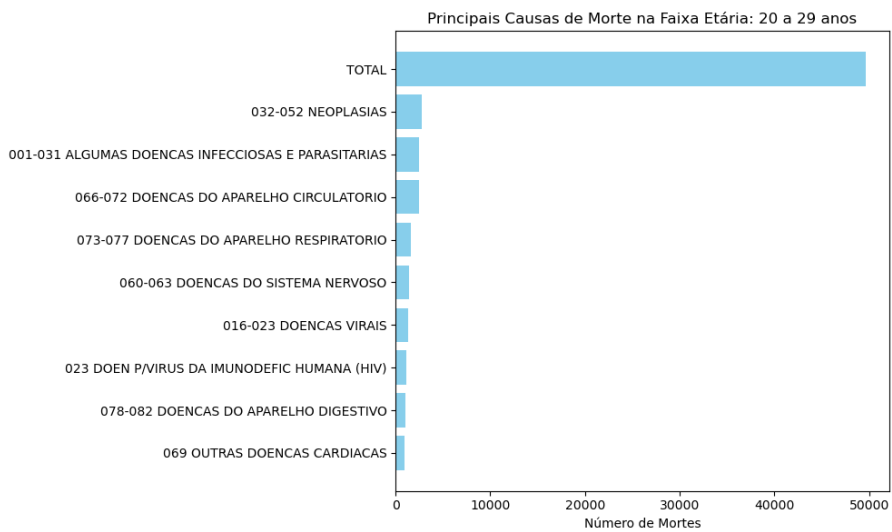


Figura 15. Taxa de mortalidade na faixa etária de 19 a 29 anos.

A análise dos gráficos revela uma queda significativa no número de mortes nesta faixa etária, o que pode ser atribuído ao fato de os indivíduos serem jovens adultos. Nessa fase da vida, as pessoas apresentam maior resistência física e imunológica em comparação com outras faixas etárias, como crianças ou idosos. Entre as principais causas de morte nesta faixa, destacam-se as neoplasias (câncer), seguidas pelas doenças do sistema nervoso.

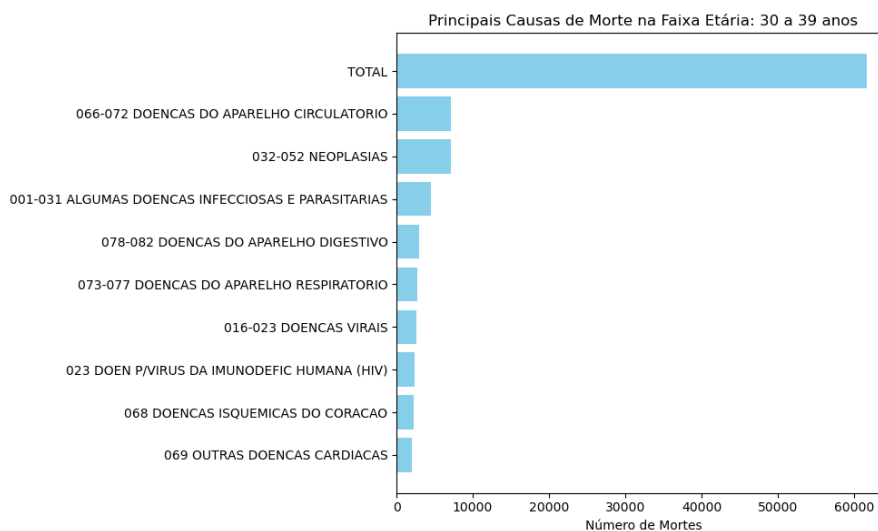


Figura 16. Taxa de mortalidade na faixa etária de 29 a 39 anos.

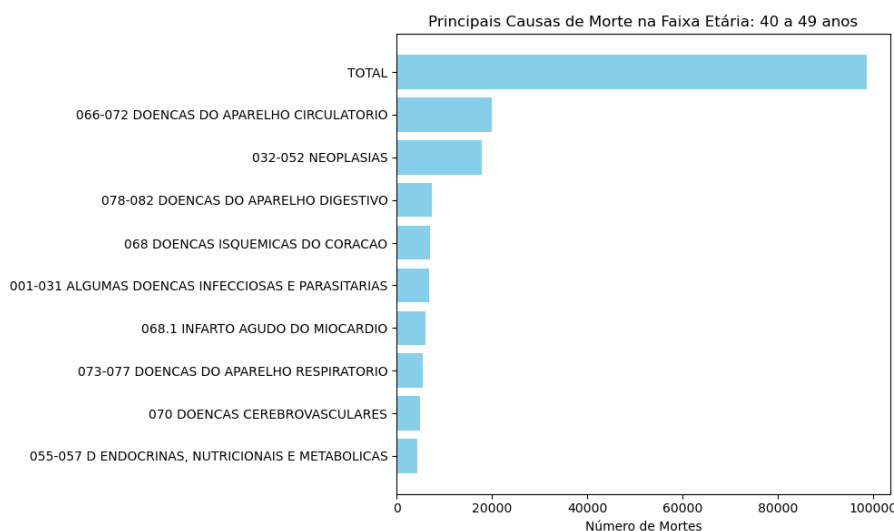


Figura 17. Taxa de mortalidade na faixa etária de 39 a 49 anos.

A partir desses dois gráficos, obtemos informações importantes: à medida que a população envelhece, a mortalidade aumenta, e mais indivíduos começam a falecer por CIDs específicas, como doenças do aparelho circulatório e neoplasias (câncer). Esse fenômeno pode ser explicado por diversos fatores. O envelhecimento do organismo está associado ao acúmulo de danos celulares e moleculares, à diminuição da capacidade de regeneração dos tecidos e ao declínio do sistema imunológico, conhecido como

imunossenescência. Além disso, fatores de risco acumulados ao longo da vida, como tabagismo, sedentarismo e má alimentação, contribuem significativamente para o aumento dessas condições.

No caso das doenças do aparelho circulatório, o envelhecimento provoca alterações como o endurecimento das artérias e o aumento da pressão arterial, elevando o risco de eventos cardiovasculares graves. Já as neoplasias (câncer) estão relacionadas ao aumento de mutações genéticas ao longo do tempo, devido à exposição prolongada a agentes mutagênicos e à redução da eficácia dos mecanismos de reparo do DNA.

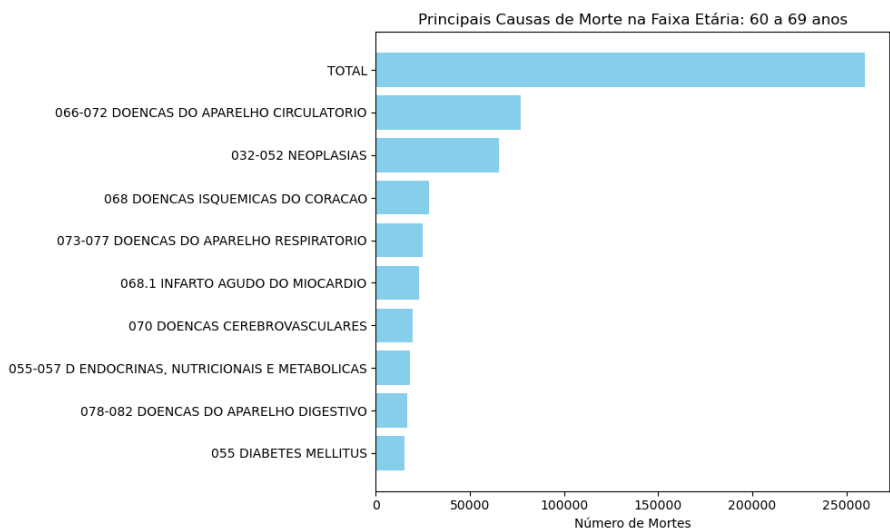


Figura 18. Taxa de mortalidade na faixa etária de 60 a 69 anos.

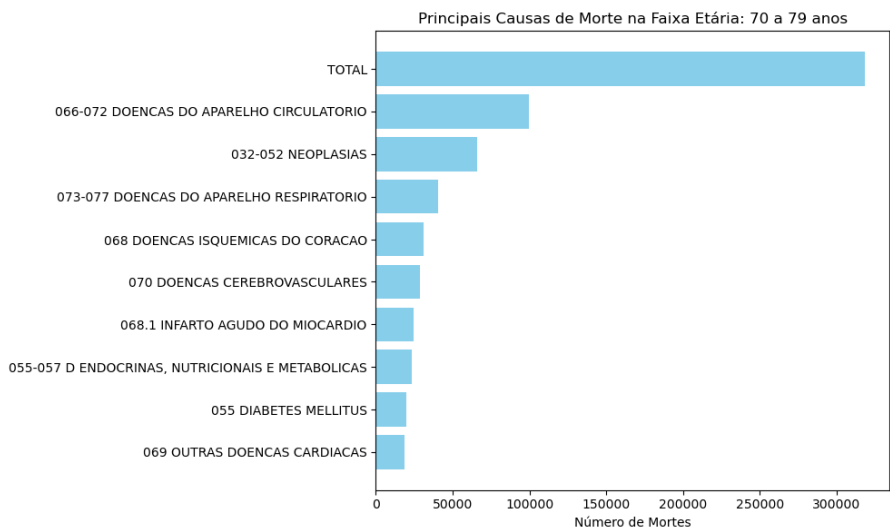


Figura 19. Taxa de mortalidade na faixa etária de 70 a 79 anos.

Nestes dois últimos gráficos, podemos inferir conclusões importantes. Conforme mencionado anteriormente, é nesta faixa etária que se encontra a maior taxa de mortalidade, algo natural com o envelhecimento da população. Contudo, é possível observar

informações relevantes: à medida que a população envelhece, há um aumento significativo nas mortes causadas por doenças do coração e neoplasias (câncer). Além disso, em idades mais avançadas, as doenças do aparelho respiratório também representam uma parcela considerável dessas mortes.

Esse aumento está diretamente relacionado ao processo de envelhecimento, que resulta em uma diminuição progressiva da imunidade, tornando os indivíduos mais suscetíveis a doenças crônicas e infecciosas. O declínio da capacidade do organismo de combater infecções e regenerar tecidos contribui para a maior prevalência de condições graves em faixas etárias mais avançadas.

A análise das mortes por diferentes CIDs ao longo das faixas etárias em 2023 revela padrões claros e importantes. Em idades mais jovens, as principais causas de morte estão relacionadas a fatores genéticos, imunidade em desenvolvimento e doenças infecciosas, como infecções respiratórias e malformações congênitas. À medida que os indivíduos avançam na idade, observa-se uma transição para doenças crônicas, como neoplasias (câncer) e doenças do aparelho circulatório, que se tornam predominantes.

Nas idades mais avançadas, a maior taxa de mortalidade é esperada devido ao acúmulo de fatores de risco ao longo da vida, ao declínio da imunidade e à maior vulnerabilidade do organismo a doenças graves. Esses dados reforçam a importância de estratégias de prevenção e intervenção específicas para cada faixa etária, desde cuidados pré-natais e vacinação na infância até o monitoramento de fatores de risco e promoção de hábitos saudáveis na vida adulta e na velhice.

3.4. Análise de Mortalidade por CIDs e Sexo (2023)

Utilizamos nossa base de dados para investigar se o sexo dos indivíduos influencia nas taxas de mortalidade, avaliando também se essa variável pode ser relevante na identificação de fatores de risco. O objetivo foi compreender como essas diferenças podem contribuir para estratégias de prevenção. A análise foi realizada para o ano de 2023, e os resultados das taxas de mortalidade estão representados abaixo.

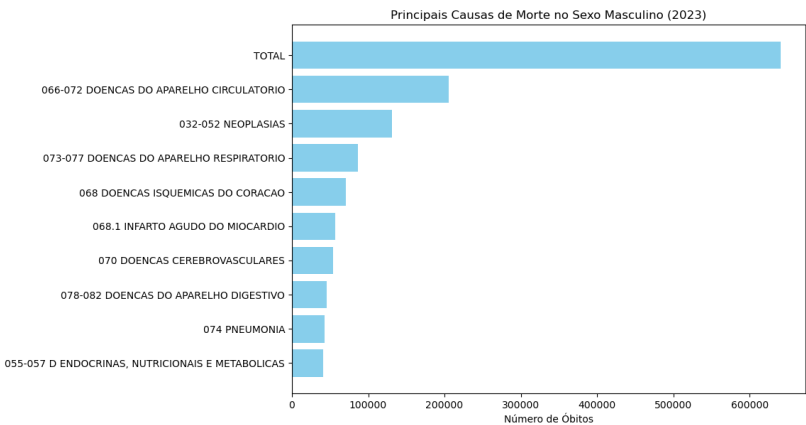


Figura 20. Taxa de mortalidade por diferentes CIDs para o sexo masculino no ano de 2023.

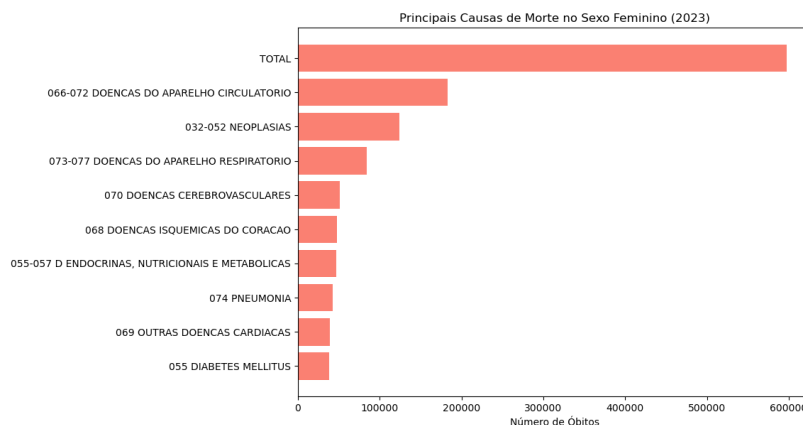


Figura 21. Taxa de mortalidade por diferentes CIDs para o sexo feminino no ano de 2023.

Os gráficos acima mostram que, em geral, não há grandes diferenças entre os sexos em relação aos tipos de CIDs mais prevalentes. Assim como em análises anteriores, as principais causas de mortalidade continuam sendo doenças do aparelho circulatório e neoplasias (câncer).

No entanto, algumas particularidades foram observadas. Entre os homens, as doenças relacionadas com aparelho digestivo aparecem com maior destaque, assim como o infarto. Já entre as mulheres, as diabetes emergem como uma causa relevante, representando uma parcela significativa das mortes associadas a diferentes CIDs. Apesar dessas observações, o estudo foi conduzido de maneira exploratória, sem maior aprofundamento em análises complementares.

4. Modelo de Machine Learning

Nesta seção, apresentamos a implementação de um simples modelo de Machine Learning para prever o total de mortes relacionadas aos CIDs usando dados históricos do nosso dataset. A seguir, detalhamos a abordagem escolhida e como os dados se comportaram.

4.0.1. Regressão Linear

Optamos por um modelo de regressão linear, uma técnica de aprendizado supervisionado amplamente usada para prever e analisar tendências em dados numéricos. Este modelo ajusta uma linha reta para relacionar uma variável dependente (neste caso, o total de mortes) com uma variável independente (o ano). A regressão linear é simples, mas eficiente para identificar padrões lineares nos dados.

4.0.2. Construção do Modelo de Regressão Linear

A construção do modelo seguiu uma série de etapas. Primeiro, os dados foram carregados e passaram por um processo de limpeza para remover inconsistências e valores ausentes. Além disso, aplicamos uma transformação logarítmica ao total de mortes para estabilizar a variabilidade dos dados, facilitando o ajuste do modelo.

Em seguida, o dataset foi dividido em duas partes: 80% para treinamento e 20% para teste. Isso permitiu avaliar a capacidade do modelo de generalizar para novos dados. O modelo de regressão linear foi treinado usando o ano como variável independente e o total de mortes como variável dependente, ajustando uma linha reta para representar a melhor relação entre essas variáveis com base nos dados de treinamento.

Para medir a precisão do modelo, utilizamos métricas como o **Erro Quadrático Médio (MSE)** e o **Coeficiente de Determinação (R^2)**. Essas métricas indicaram a qualidade do ajuste e a eficácia do modelo em explicar as variações nos dados.

Por fim, empregamos o modelo para realizar previsões de mortes para os anos de 2024 a 2027. As estimativas geradas seguiram tendências coerentes com os dados históricos analisados, demonstrando o potencial do modelo para projeções futuras.

4.0.3. Visualização dos Resultados

O gráfico resultante mostra a linha de regressão ajustada aos dados históricos e estendida para previsões futuras. A linha contínua conecta os pontos reais e previstos, destacando a tendência crescente ou decrescente ao longo do tempo.

Este modelo de regressão linear, embora simples, proporciona uma base sólida para previsões iniciais e pode ser refinado para cenários mais complexos usando técnicas avançadas de Machine Learning.

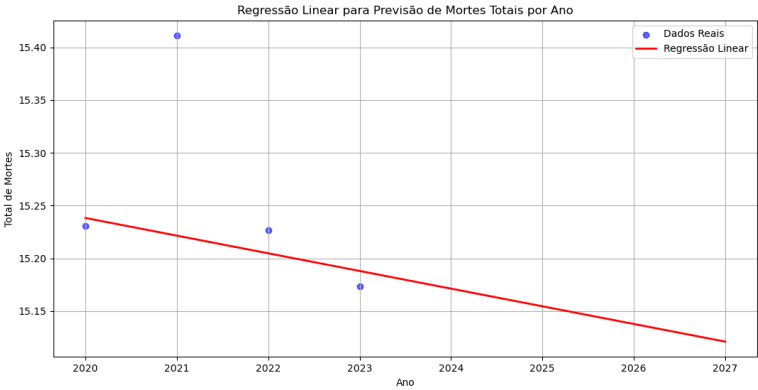


Figura 22. Gráfico de previsão de mortes ao longo do período de (2024-2027) utilizando regressão linear

Com base no período analisado, observamos um aumento significativo nas taxas de mortalidade no ano de 2021. Esse crescimento abrupto pode ser amplamente atribuído à pandemia de COVID-19, que impactou de maneira devastadora a saúde pública global. Após esse pico, os anos seguintes apresentaram uma queda consistente nas taxas de mortalidade, refletindo os esforços de vacinação em massa e melhorias nas estratégias de tratamento.

Esse comportamento se reflete diretamente no modelo de regressão linear utilizado para prever tendências futuras. O modelo sugere uma tendência de diminuição gradual

no número de mortes ao longo dos anos subsequentes, alinhando-se com o padrão decrescente observado após o impacto inicial da pandemia.

5. Conclusão

O presente projeto teve como objetivo principal analisar as tendências de mortalidade no Brasil ao longo do período de 2020 a 2023, destacando as principais causas de morte por diferentes CIDs, regiões, faixas etárias e gêneros. Por meio de técnicas de análise de dados e aprendizado de máquina, foram identificados padrões relevantes que reforçam a necessidade de ações específicas para reduzir a mortalidade por doenças do aparelho circulatório, neoplasias e infecções parasitárias.

Os resultados demonstraram que as doenças do aparelho circulatório continuam sendo a principal causa de morte em todas as regiões do Brasil, seguidas por neoplasias e doenças infecciosas. Além disso, fatores como envelhecimento populacional, desigualdade socioeconômica e hábitos de vida modernos desempenham papéis cruciais no aumento da mortalidade. No entanto, variações significativas foram observadas entre faixas etárias, com doenças respiratórias e malformações congênitas afetando principalmente crianças, enquanto doenças crônicas predominam em faixas etárias mais avançadas.

O uso de regressão linear para previsão de mortalidade destacou o impacto devastador da pandemia de COVID-19 em 2021, seguido por uma tendência de recuperação nos anos subsequentes. As previsões para o período de 2024 a 2027 indicam uma possível estabilização nas taxas de mortalidade, resultado de avanços nas políticas públicas e estratégias de saúde.

Este estudo reforça a importância de análises abrangentes de dados para subsidiar políticas públicas eficazes, especialmente no campo da saúde preventiva. As conclusões aqui apresentadas podem servir como base para o desenvolvimento de intervenções mais direcionadas, voltadas para a redução das desigualdades regionais e para a promoção de estilos de vida saudáveis em todas as faixas etárias.

6. Referências

References

- [1] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2ª Edição, 2009.
- [2] Osborne, J. W. *Best Practices in Data Transformation: The Log Transformation*. Practical Assessment, Research, and Evaluation, vol. 8, n.º 6, 2002.
- [3] Silva, P. H.; Santos, R. A. *Impacto da Pandemia de COVID-19 nas Taxas de Mortalidade no Brasil*. Revista Brasileira de Epidemiologia, vol. 24, n.º 3, 2022.
- [4] Pedregosa, F.; Varoquaux, G.; Gramfort, A., et al. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.