

# **Predição de Mortes por Tuberculose Utilizando Aprendizado de Máquina**

**Tales de Campos Hernandez<sup>1</sup>, Maria Amelia Eliseo<sup>1</sup>**

<sup>1</sup>Faculdade de Computação e Informática (FCI)  
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

[10408846@mackenzie.br](mailto:10408846@mackenzie.br); RA 10408846

**Resumo.** *Este projeto tem como objetivo aplicar técnicas de Inteligência Artificial para analisar e prever o número estimado de mortes por tuberculose (todas as formas) a cada 100.000 habitantes em diferentes países e períodos históricos. Utilizando dados da Organização Mundial da Saúde (OMS) disponibilizados via Gapminder, serão explorados padrões temporais, regionais e socioeconômicos relacionados à tuberculose, com o intuito de identificar fatores críticos que afetam a mortalidade. Pretende-se aplicar modelos de Machine Learning para criar um sistema preditivo que auxilie no monitoramento e combate da doença.*

**Palavras-chave:** *Inteligência Artificial; Aprendizado de Máquina; Tuberculose; Predição; Saúde Pública.*

**Abstract.** *This project aims to apply Artificial Intelligence techniques to analyze and predict the estimated number of tuberculosis deaths (all forms) per 100,000 inhabitants in different countries and historical periods. Using data from the World Health Organization (WHO) available via Gapminder, temporal, regional, and socioeconomic patterns related to tuberculosis will be explored to identify critical factors affecting mortality. The goal is to apply Machine Learning models to create a predictive system that aids in monitoring and combating the disease.*

**Keywords:** *Artificial Intelligence; Machine Learning; Tuberculosis; Prediction; Public Health.*

## **1. Introdução**

### **a. Contextualização**

A tuberculose é uma das doenças infecciosas mais letais do mundo, afetando milhões de pessoas todos os anos. Apesar dos avanços na medicina, ainda representa um grande desafio de saúde pública, especialmente em países de baixa e média renda.

### **b. Justificativa**

A análise de dados históricos pode auxiliar governos e organizações internacionais a direcionarem políticas públicas e recursos para o combate da

tuberculose. O uso de Inteligência Artificial permite identificar padrões e prever cenários futuros, o que pode contribuir para a redução da mortalidade.

### **c. Objetivo**

- Realizar uma análise exploratória dos dados históricos de mortalidade por tuberculose;
- Aplicar modelos de Machine Learning (classificação/regressão) para prever a evolução da mortalidade;
- Avaliar o desempenho dos modelos e interpretar os resultados obtidos.

### **d. Opção do projeto**

Framework: utilização do scikit-learn para análise e modelagem preditiva.

## **2. Descrição do Problema**

Apesar dos esforços da OMS e governos locais, a tuberculose continua a ser uma das principais causas de morte em diversos países. A previsão da mortalidade com base em dados históricos pode auxiliar na tomada de decisão e no direcionamento de recursos. O problema consiste em desenvolver um modelo de aprendizado de máquina capaz de prever a taxa de mortes por tuberculose em diferentes regiões.

## **3. Aspectos Éticos e Responsabilidade**

O uso da IA em saúde exige responsabilidade. Modelos preditivos não substituem diagnósticos médicos, mas podem apoiar políticas públicas. É importante garantir que os dados sejam tratados de forma ética, evitando vieses geográficos ou socioeconômicos que possam comprometer a interpretação. Além disso, os resultados devem ser apresentados de forma transparente, sem criar pânico ou interpretações equivocadas.

## **4. Dataset, Análise Exploratória e Preparação dos Dados**

O dataset utilizado foi obtido no portal Gapminder (<https://www.gapminder.org/data/>), com origem nos dados oficiais da Organização Mundial da Saúde (WHO) (<https://www.who.int/tb/en/>). Ele contém o número estimado de mortes por tuberculose (todas as formas) por 100.000 habitantes em diversos países e territórios, no período de 2000 a 2023.

O arquivo possui 214 linhas (países/regiões) e 24 colunas de anos (2000 a 2023), além da coluna de identificação do país.

### **Durante a análise exploratória:**

- Identificou-se que há valores ausentes em alguns anos iniciais (2000–2010), mas a partir de 2011 os dados estão completos.

- Observou-se que a média global de mortalidade por tuberculose vem diminuindo desde 2000, embora alguns países, sobretudo na África Subsaariana, apresentem taxas ainda muito elevadas.
- Em 2023, os países mais críticos incluem Lesoto, República Centro-Africana e Gabão, com valores superiores a 100 mortes por 100k habitantes.

#### **Para a preparação dos dados:**

- Os valores ausentes foram tratados por meio de imputação pela média.
- Foi feita a normalização (StandardScaler) para garantir escalas comparáveis entre as variáveis temporais.
- Definiu-se como variáveis explicativas (X) os anos de 2000 a 2022, e como variável alvo (y) o ano de 2023, permitindo avaliar modelos de regressão para prever a mortalidade recente a partir de dados históricos.
- Por fim, os dados foram divididos em treino (80%) e teste (20%), garantindo a avaliação adequada dos modelos de Machine Learning.

## **5. Metodologia e Resultados Esperados**

### **Metodologia**

O projeto utilizará técnicas de Ciência de Dados e Aprendizado de Máquina para prever a taxa estimada de mortes por tuberculose (TB) por 100.000 habitantes em diferentes países. O processo metodológico será estruturado em etapas:

- 1. Coleta e Organização dos Dados**
  - a. Utilização de dataset disponibilizado pela Organização Mundial da Saúde (WHO), contendo estimativas anuais de mortalidade por todas as formas de tuberculose entre 2000 e 2023.
- 2. Análise Exploratória dos Dados (EDA)**
  - a. Identificação de padrões, tendências históricas e diferenças regionais.
  - b. Verificação de valores ausentes e inconsistências nos registros.
  - c. Visualização temporal das séries de mortalidade.
- 3. Pré-processamento e Preparação dos Dados**
  - a. Tratamento de valores ausentes por imputação.
  - b. Normalização das variáveis para garantir melhor desempenho dos modelos.
  - c. Separação entre variáveis explicativas (anos 2000–2022) e variável alvo (ano 2023).

#### 4. Modelagem Preditiva

- a. Teste de diferentes algoritmos supervisionados de regressão, incluindo Regressão Linear, Ridge, Lasso e Random Forest.
- b. Validação cruzada e comparação entre modelos com base em métricas como  $R^2$  (coeficiente de determinação), MAE (Erro Absoluto Médio) e RMSE (Raiz do Erro Quadrático Médio).

#### 5. Avaliação e Interpretação dos Resultados

- a. Comparação entre valores reais e previstos para países do conjunto de teste.
- b. Análise da acurácia dos modelos e discussão sobre limitações (ex.: variáveis externas não incluídas, como condições socioeconômicas, campanhas de saúde pública e acesso a tratamento).

#### Resultados Esperados

- Construção de um modelo capaz de prever a mortalidade por tuberculose com boa precisão, utilizando apenas dados históricos de séries temporais.
- Geração de insights sobre a evolução da doença em diferentes países.
- Contribuição prática para a formulação de políticas públicas de saúde, permitindo a identificação antecipada de regiões com maior risco.

#### 6. Referências

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). \*Global Tuberculosis Report 2023. Disponível em: <<https://www.who.int/teams/global-tuberculosis-programme/data>>. Acesso em: 22 set. 2025.

GÉRON, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2. ed. Sebastopol: O'Reilly Media, 2019.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. An Introduction to Statistical Learning: with Applications in R. Nova York: Springer, 2013.

#### 7. Bibliografia

ALPAYDIN, Ethem. \*Introduction to Machine Learning\*. 4. ed. Cambridge: MIT Press, 2020.

KUHN, Max; JOHNSON, Kjell. \*Applied Predictive Modeling\*. Nova York: Springer, 2013.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. \*Data Mining: Concepts and Techniques\*. 3. ed. Burlington: Morgan Kaufmann, 2012.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). \*Tuberculosis Data Portal\*. Disponível em: <<https://www.who.int/teams/global-tuberculosis-programme/data>>. Acesso em: 22 set. 2025.