UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

LUCAS MOREIRA RIBEIRO
Orientador: Prof. Dr. Jadson Castro Gertrudes
Coorientador: Prof. Dr. Wendel Coura-Vital

PREDIÇÃO DAS TAXAS DE INCIDÊNCIA DE LEISHMANIOSE VISCERAL UTILIZANDO ALGORITMOS DE APRENDIZADO DE MÁQUINA

Ouro Preto, MG
2024

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

LUCAS MOREIRA RIBEIRO

PREDIÇÃO DAS TAXAS DE INCIDÊNCIA DE LEISHMANIOSE VISCERAL UTILIZANDO ALGORITMOS DE APRENDIZADO DE MÁQUINA

Monografia II apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Jadson Castro Gertrudes
Coorientador: Prof. Dr. Wendel Coura-Vital

Ouro Preto, MG
2024

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

R484p Ribeiro, Lucas Moreira.
Predição das taxas de incidência de leishmaniose visceral utilizando algoritmos de aprendizado de máquina. [manuscrito] / Lucas Moreira Ribeiro. - 2024.
51 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Jadson Castro Gertrudes.
Coorientador: Prof. Dr. Wendel Coura-Vital.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Graduação em Ciência da Computação .
ISBN: 978-65-00-95115-8.

1. Inteligência artificial. 2. Leishmaniose visceral. 3. Aprendizado do computador. 4. Redes neurais (Computação). 5. Epidemiologia. I. Gertrudes, Jadson Castro. II. Coura-Vital, Wendel. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 004.8

Bibliotecário(a) Responsável: Paulo Vitor Oliveira - CRB6/2551

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

FOLHA DE APROVAÇÃO

Lucas Moreira Ribeiro

Predição das taxas de incidência de leishmaniose visceral utilizando algoritmos de aprendizado de máquina

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal
de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 16 de Fevereiro de 2024.

Membros da banca:

Jadson Castro Gertrudes (Orientador) - Doutor - Universidade Federal de Ouro Preto
Wendel Coura-Vital (Coorientador) - Doutor - Universidade Federal de Ouro Preto
Hugo Eduardo Ziviani (Examinador) - Mestre - PPGCC/UFOP
Pedro Henrique Lopes Silva (Examinador) - Doutor - Universidade Federal de Ouro Preto

Jadson Castro Gertrudes, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 16/02/2024.

Documento assinado eletronicamente por **Anderson Almeida Ferreira**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 17/02/2024, às 10:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Jadson Castro Gertrudes**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 18/02/2024, às 18:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0669180** e o código CRC **0B502D12**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: 3135591692 - www.ufop.br

Dedico esse trabalho a vózinha mais amorosa do mundo (in memoriam), a senhora sempre terá um lugar no meu coração.

Agradecimentos

Agradeço primeiramente aos meus pais e aos meus irmãos, por serem um pilar fundamental em minha jornada, que sempre estiveram ao meu lado, apoiando-me e acreditando em meu potencial.

Aos integrantes do Laboratório de Epidemiologia das Doenças Infecciosas, em especial a Josefa Monteiro, que foi a coorientadora da minha primeira iniciação cientifica, ao Rafael Vieira e a Tamara Coelho, que me ajudaram na escrita dos relatórios e na apresentação do Encontro dos Saberes.

Ao Wendel Coura-Vital, o coorientador deste trabalho e orientador das minhas iniciações cientificas, o qual me ajudou muito na graduação e me deu base para este trabalho.

Ao corpo docente e técnico da UFOP que ajudaram na minha formação.

Ao meu orientador, Jadson, que me direcionou no desenvolvimento desse trabalho.

Resumo

A Leishmaniose Visceral representa uma doença tropical negligenciada, registrando anualmente entre 50.000 e 90.000 novos casos em todo o mundo. Originalmente concebida como uma enfermidade predominantemente rural, essa condição evoluiu devido à urbanização e à adaptação bem-sucedida do vetor ao ambiente doméstico, emergindo agora também em grandes centros urbanos. Atualmente, várias cidades de médio e grande porte são consideradas endêmicas para essa enfermidade. Diversos fatores estão impulsionando a disseminação da doença, incluindo elementos como temperatura, precipitação, umidade, desmatamento, urbanização e condições socioeconômicas, entre outros. Com esse cenário em mente, o presente estudo utilizou dois modelos de aprendizado de maquina, um modelo de rede neural profunda, o *Long Short Term Memory* (LSTM) e o *Extreme Gradient Boosting* (XGBoost) para prever a incidência de Leishmaniose Visceral a partir de series temporais de incidência e variáveis socioeconômicas dos municípios. Os resultados do modelo XGBoost foram insatisfatórios, apresentando um coeficiente de determinação (R^2) de $-0,27$. Em contrapartida, o modelo LSTM mostrou-se promissor, com um R^2 de $0,231$. Apesar de parecer baixo em primeira análise, esse valor é significativo, considerando as complexas características da doença e o emprego de um modelo com variáveis limitadas. O Erro Percentual Absoluto Médio (MAPE) do LSTM foi de 1,73%, evidenciando um baixo nível de erro nas predições. É importante notar que grande parte das observações de incidência era nula. Essa predominância de valores nulos pode ter influenciado o modelo a prever incidências mais baixas, sendo um ponto a ser considerado na interpretação dos resultados.

Palavras-chave: Redes neurais artificiais; Aprendizado profundo; Long Short Term Memory; Leishmaniose Visceral; Epidemiologia.

Abstract

Visceral leishmaniasis is a neglected tropical disease, with between 50,000 and 90,000 new cases worldwide every year. Originally conceived as a predominantly rural disease, this condition has evolved due to urbanization and the successful adaptation of the vector to the domestic environment, now also emerging in large urban centers. Currently, several medium-sized and large cities are considered endemic for this disease. Several factors are driving the spread of the disease, including elements such as temperature, rainfall, humidity, deforestation, urbanization and socioeconomic conditions, among others. With this scenario in mind, this study used two machine learning models, a deep neural network model, LSTM and XGBoost to predict the incidence of Visceral Leishmaniasis from time series of incidence and socioeconomic variables of the municipalities. The results of the XGBoost model were unsatisfactory, with an R^2 of -0.27. In contrast, the LSTM model proved promising, with an R^2 of 0.231. Despite appearing low at first glance, this value is significant, considering the complex characteristics of the disease and the use of a model with limited variables. The MAPE of the LSTM was 1.73%, showing a low level of error in the predictions. It is important to note that most of the incidence observations were null. This predominance of null values may have influenced the model to predict lower incidences, which is a point to consider when interpreting the results.

Keywords: Neural Networks; Deep Learning, Long Short Term Memory, Leishmaniose Visceral, Epidemiology.

Lista de Ilustrações

Figura 2.1 – Fêmea de Flebotomíneo adulto. 4
Figura 2.2 – Diagrama de Venn mostrando que o aprendizado de máquina é uma subárea da inteligência artificial e o aprendizado profundo é uma subárea do aprendizado de máquina 5
Figura 2.3 – Abordagem clássica vs Aprendizado de Máquina 6
Figura 2.4 – Algoritmo de Aumento de Gradiente 7
Figura 2.5 – Um neurônio biológico comparado com uma ANN: (a) neurônio biológico; (b) neurônio artificial; (c) sinapse biológica; e (d) equivalente a uma sinapse em uma ANN. 8
Figura 2.6 – Fluxograma do treinamento de uma Rede Neural para uma entrada 9
Figura 2.7 – Rede Neural Recorrente vs Rede Feed-Foward. 10
Figura 2.8 – Arquitetura LSTM 11
Figura 3.1 – Diagrama de seleção e limpeza dos dados 16
Figura 3.2 – Fluxo do tratamento dos dados 18
Figura 4.1 – Comparação do desempenho dos otimizadores e funções de perda para o LSTM 20
Figura 4.2 – Histograma de Resíduos 21
Figura 4.3 – Valores preditos vs Valores observados 22

Lista de Tabelas

Tabela 3.1 – Topologia da rede neural LSTM. 18
Tabela A.1 – Municípios utilizados para o treino e teste dos modelos 29

Siglas

R^2	coeficiente de determinação. vii, viii, 19, 20, 23
AI	Inteligência Artificial. 5, 14
ANN	Rede Neural Artificial. 7
CGIAE	Coordenação-Geral de Informações e Análise Epidemiológica. 15
CGZV	Coordenação-Geral de Zoonoses e Vetores. 15
DAENT	Departamento de Análise da Situação de Saúde. 15
DEDT	Departamento de Epidemiologia e Doenças Transmitidas por Vetores. 15
IEPS	Instituto de Estudos para Políticas de Saúde. 15
LSTM	*Long Short Term Memory*. vii, viii, 2, 3, 11, 12, 14, 15, 17, 20, 22–24
LV	Leishmaniose Visceral. 1–5, 12, 15, 23, 24
MAPE	Erro Percentual Absoluto Médio. vii, viii, 19, 21, 23
MS	Ministério da Saúde. 15
ReLU	*rectified linear unit*. 8
RF	*Random Forest*. 13
RNN	Rede Neural Recorrente. 10, 11
SVSA	Secretaria de Vigilância em Saúde. 15
XGBoost	*Extreme Gradient Boosting*. vii, 2, 3, 6, 7, 13, 15

Sumário

1 Introdução 1
1.1 Justificativa 2
1.2 Objetivos 3
1.3 Organização do Trabalho 3
2 Revisão Bibliográfica 4
2.1 A Leishmaniose Visceral & Epidemiologia 4
2.2 Aprendizado de máquina 5
2.3 XGBoost 6
2.4 Redes Neurais Artificiais 7
2.4.1 Treinamento de uma Rede Neural Artificial 9
2.4.2 Algoritmo de retropropagação 10
2.5 Redes Neurais Recorrentes 10
2.5.1 Long Short Term Memory 11
2.6 Trabalhos Relacionados 12
3 Metodologia 15
3.1 Desenho do estudo 15
3.2 Fonte de dados 15
3.3 Dados 15
3.4 Tratamento dos dados 17
3.5 Parametrização e treinamento do Modelo LSTM 17
3.6 Parametrização e treinamento do Modelo XGBoost 19
3.7 Métricas para avaliação do modelo 19
4 Resultados 20
4.1 Avaliação dos Modelos & Discussão 20
5 Considerações Finais 23
5.1 Conclusão 23
5.2 Trabalhos Futuros 24

Referências 25

Apêndices 28
APÊNDICE A Lista de municípios 29
APÊNDICE B Series temporais de cada município do conjunto de teste e sua incidência predita 31
APÊNDICE C Series temporais de cada município do conjunto de treinamento . . 37

Índice 51

1 Introdução

Consideradas como doenças negligenciadas, as leishmanioses representam um complexo de doenças com uma grande diversidade clínico-epidemiológica, constituindo um sério problema de saúde pública (Desjeux, 2004; World Health Organization, 2022). A Leishmaniose Visceral (LV) é uma doença tropical que contabiliza aproximadamente 50.000 a 90.000 novos casos em todo o mundo anualmente, com a maior parte dos casos ocorrendo em seis países: Índia, Sudão do Sul, Sudão, Brasil, Etiópia e Somália (World Health Organization, 2022; World Health Organization, 2023). Nas Américas, a LV está presente em 13 países, sendo 97% dos casos notificados no Brasil, com taxa de incidência de 2,0 / 100.000 habitantes e letalidade em torno de 7% (Organização Pan-Americana da Saúde, 2020).

Com base nos dados do Ministério da Saúde (2006), no Brasil a LV é causada pelo protozoário *Leishmania infantum* e transmitida pelo vetor *Lutzomia longipalpis*, e o principal reservatório doméstico é o cão. Inicialmente considerada uma doença predominantemente rural, a LV passou por um processo de urbanização a partir da década de 80, com a adaptação bem-sucedida do vetor ao ambiente doméstico. Consequentemente, a doença começou a ser identificada em grandes centros urbanos. Atualmente, diversas cidades de médio e grande porte são consideradas endêmicas da doença (Harhay *et al.*, 2011). Atualmente a LV está amplamente distribuída pelo Brasil, tendo ocorrido notificação em todas as 27 unidades da federação (Organização Pan-Americana da Saúde, 2020). O maior número de casos concentra-se na região nordeste, seguido da região sudeste e centro-oeste, sendo Minas Gerais o estado que apresentou o maior número de casos (Organização Pan-Americana da Saúde, 2020).

Diversos fatores podem contribuir para a expansão da LV e dentre eles podemos destacar: temperatura, precipitação, umidade, cobertura vegetal, desmatamento, urbanização, imunossupressão, condições socioeconômicas, processos migratórios, inundações e incursão de humanos para áreas naturais (Oryan; Akbari, 2016; Desjeux, 2001; Werneck, 2009; Werneck, 2010). As condições socioeconômicas abrangem aspectos como níveis de renda, acesso a serviços de saúde, saneamento básico, moradia adequada e educação, os quais podem influenciar significativamente na incidência e propagação da doença.

Com o aumento da disponibilidade de dados epidemiológicos, sociais e ambientais, surge a oportunidade de explorar o potencial do aprendizado de máquina, incluindo o aprendizado profundo (*deep learning*) para predizer a ocorrência da doença. O objetivo é identificar relações complexas entre grupos de variáveis e a capacidade de estimar a taxa de incidência da doença com base em fatores socioeconômicos, climáticos, ambientais, antrópicos, dentre outros, visando prever a incidência da Leishmaniose Visceral em diversas cidades e regiões. Essa abordagem pode fornecer *insights* valiosos para o planejamento e implementação mais eficaz de estratégias

de controle, auxiliando na alocação de recursos e na priorização de áreas de maior risco.

Diante dessa perspectiva, a criação de um modelo de predição utilizando séries temporais, poderia ser empregado para antecipar o número de casos em uma determinada região, considerando parâmetros conhecidos. Ao capturar variações sazonais, tendências e outros padrões oscilatórios nas variáveis relacionadas à LV, o modelo poderia fornecer informações cruciais para a quantificação precisa do risco de incidência em diferentes períodos temporais.

Portanto, neste trabalho, foi desenvolvido um modelo de *deep learning*, mais especificamente o *Long Short Term Memory* (LSTM) e o *Extreme Gradient Boosting* (XGBoost), que utilizou séries temporais de dados de incidência da LV e variáveis socioeconômicas por município do Estado de Minas Gerais, para prever incidências futuras da doença.

1.1 Justificativa

As medidas de controle adotadas pelo Ministério da Saúde brasileiro, por meio do Programa de Vigilância e Controle da Leishmaniose Visceral (PVC-LV), preconizam o diagnóstico e tratamento precoces dos indivíduos infectados; controle vetorial, com a aplicação de inseticidas de efeito residual; eutanásia dos cães soropositivos, além de ações de educação em saúde (Ministério da Saúde, 2006). No entanto, apesar de todas essas ações do programa de controle, observa-se uma dificuldade em reduzir a incidência da doença, especialmente em grandes centros urbanos (Rocha *et al.*, 2018). Portanto, a capacidade de prever a incidência da doença nos municípios é um fator chave para direcionar políticas públicas de controle da doença naquela área.

Outros estudos, como o de Duarte *et al.* (2022), analisam a influência de variáveis climáticas, utilizando métodos estatísticos menos complexos, como a regressão linear, apesar de trabalharem com séries temporais. Enquanto o atual trabalho, utiliza variáveis socioeconômicas e propõe a utilização de algoritmos de aprendizado de máquina para prever a incidência de casos com base na série histórica dos municípios da macrorregião central de Minas Gerais.

Mussumeci e Codeço Coelho (2020) têm uma proposta semelhante a este trabalho, de criar modelos de aprendizado de máquina, especificamente o LSTM e o *Random Forest*, que se baseia em series temporais para predizer novos casos, porém, analisando o comportamento de outra doença tropical, a Dengue ao invés da LV.

Silva *et al.* (2017) fazem uma análise espacial e temporal da incidência no Estado de Minas Gerais, além de utilizar a *Regressão de Poisson* para medir a variação do número médio de casos de um ano para o ano seguinte. No presente estudo, foi utilizado um modelo mais complexo, e utilizou mais dados para gerar um modelo robusto.

Com base no presente trabalho, torna-se evidente a relevância de empregar abordagens mais avançadas, como algoritmos de aprendizado de máquina, na previsão da incidência de doenças como a Leishmaniose Visceral. Justifica-se esta pesquisa porque a eficácia das medidas

de controle existentes tem sido limitada, e a implementação de modelos mais complexos pode proporcionar uma compreensão mais abrangente da propagação da doença. Além disso, esse estudo é fundamental para direcionar políticas públicas de saúde mais eficazes, considerando não apenas aspectos clínicos da doença, mas também o contexto social e econômico dos municípios afetados, que desempenham um papel crucial na disseminação e na resposta à doença.

1.2 Objetivos

O principal objetivo deste trabalho é treinar dois modelos de aprendizado de máquina, um utilizando o XGBoost e o outro utilizando o LSTM que utilizará as séries temporais prever as taxas de incidência da LV, baseado em dados socioeconômicos dos municípios e avaliá-los.

Para realizar o objetivo principal, é necessário cumprir os seguintes objetivos específicos:

- Realizar uma revisão de literatura sobre predição de casos de LV por meio de aprendizado de máquina;
- Modelagem da arquitetura de rede neural LSTM para predição da taxa de incidência dos casos de LV;
- Modelagem e otimização de hiperparâmetros do modelo XGBoost para predição da taxa de incidência dos casos de LV;
- Avaliar e comparar os modelos LSTM e o XGBoost que foram gerados.

1.3 Organização do Trabalho

O restante do trabalho está organizado da seguinte forma: O Capítulo 2 apresenta a revisão bibliográfica e o embasamento teórico necessário para entendimento do trabalho. O Capítulo 3 apresenta a metodologia utilizada para a realização de experimentos do trabalho. No Capítulo 4 são apresentados os resultados obtidos no trabalho. Por fim, o Capítulo 5 apresenta as conclusões e novas etapas a serem abordadas em trabalhos futuros.

2 Revisão Bibliográfica

Neste capítulo são apresentados os conceitos necessários para entendimento do trabalho, além dos trabalhos relacionados ao tema.

2.1 A Leishmaniose Visceral & Epidemiologia

A Leishmaniose Visceral (LV), é uma doença parasitária causada pelo protozoário do gênero Leishmania, com manifestações clínicas graves e potencialmente fatais se não tratada. Sua transmissão ocorre principalmente através da picada de fêmeas do vetor, que são flebotomíneos pertencentes ao gênero *Phlebotomus* em regiões do Velho Mundo e *Lutzomyia* nas Américas (Ready, 2014). Uma imagem do vetor transmissor é apresentada na Figura 2.1.

Figura 2.1 – Fêmea de Flebotomíneo adulto.

Fonte: Ministério da Saúde (2006).

O parasita se multiplica no intestino do vetor e é transmitido para os seres humanos durante a alimentação do flebotomíneo, iniciando o ciclo de infecção. Outro aspecto relevante da epidemiologia da leishmaniose visceral é a presença de reservatórios da doença, que incluem cães em áreas urbanas e mamíferos silvestres em regiões rurais. Eles desempenham um papel importante na manutenção e propagação do parasita (Ready, 2014).

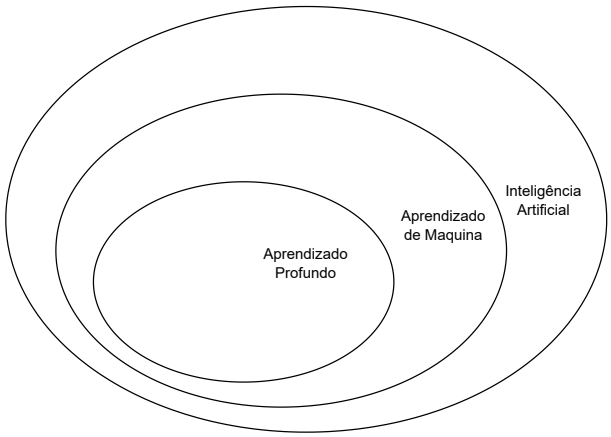
A distribuição geográfica da leishmaniose visceral é ampla e variável, com focos endêmicos em áreas específicas, muitas vezes relacionados às condições socioeconômicas, climáticas e ecológicas favoráveis ao vetor e ao parasita. É importante salientar que o vetor não precisa de água parada, apenas um substrato úmido rico em matéria orgânica, por isso, a remoção de água parada em recipientes não é um meio de prevenção. Populações mais vulneráveis, como aquelas que vivem em condições de pobreza, falta de saneamento básico e acesso limitado à assistência médica, estão em maior risco de contrair a doença (Ministério da Saúde, 2006).

A LV apresenta uma diversidade de manifestações clínicas que vão desde formas assintomáticas até quadros graves e potencialmente letais. Os sintomas incluem febre, perda de peso, hepatoesplenomegalia (aumento do fígado e do baço) e anemia normocítica normocrômica (Paltrinieri *et al.*, 2016). A falta de tratamento adequado pode levar ao óbito, especialmente em crianças (Scarpini *et al.*, 2022).

2.2 Aprendizado de máquina

O campo do aprendizado de máquina é uma subárea da Inteligência Artificial (AI) (Figura 2.2) e propõe uma abordagem diferente na resolução de problemas. Em contraste com a abordagem convencional de AI simbólica, onde os seres humanos inserem regras e dados para produzir respostas (conforme mostrado na Figura 2.3), o aprendizado de máquina reformula esse processo. (Chollet, 2017)

Figura 2.2 – Diagrama de Venn mostrando que o aprendizado de máquina é uma subárea da inteligência artificial e o aprendizado profundo é uma subárea do aprendizado de máquina

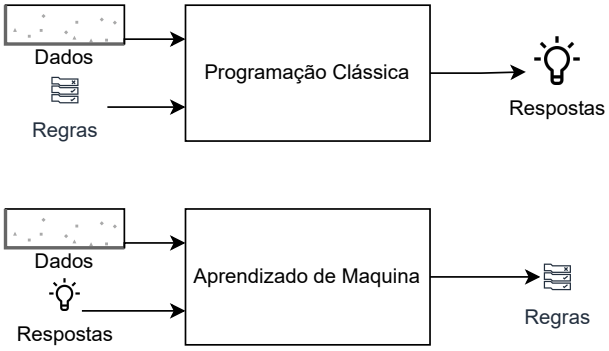


Fonte: Elaborado pelo autor.

No aprendizado de máquina, os humanos fornecem dados juntamente com as respostas desejadas associadas a esses dados. O modelo de aprendizado de máquina, então, assume a responsabilidade de derivar as regras a partir desses exemplos. Essas regras aprendidas podem ser posteriormente aplicadas a novos conjuntos de dados, gerando respostas. Esse processo, envolve apresentar uma variedade de exemplos relevantes para uma tarefa específica, permitindo que o sistema identifique padrões, que por sua vez, capacita-o a gerar respostas. (Chollet, 2017)

Há quatro tipos principais de aprendizado: O aprendizado supervisionado, que envolve o mapeamento de dados de entrada para alvos conhecidos com base exemplos rotulados; o aprendizado não supervisionado, que busca transformações nos dados sem recorrer a alvos específicos; O aprendizado auto supervisionado, que opera sem rótulos humanos, utilizando algoritmos heurísticos para gerar os próprios rótulos. Por fim, o aprendizado por reforço, em

Figura 2.3 – Abordagem clássica vs Aprendizado de Máquina



Fonte: Adaptado de Chollet (2017).

que um agente aprende a escolher ações para maximizar recompensas com base em informações ambientais. (Chollet, 2017) O presente trabalho utiliza o aprendizado supervisionado.

Neste trabalho, a técnica empregada para a predição das taxas de incidência é o aprendizado de máquina.

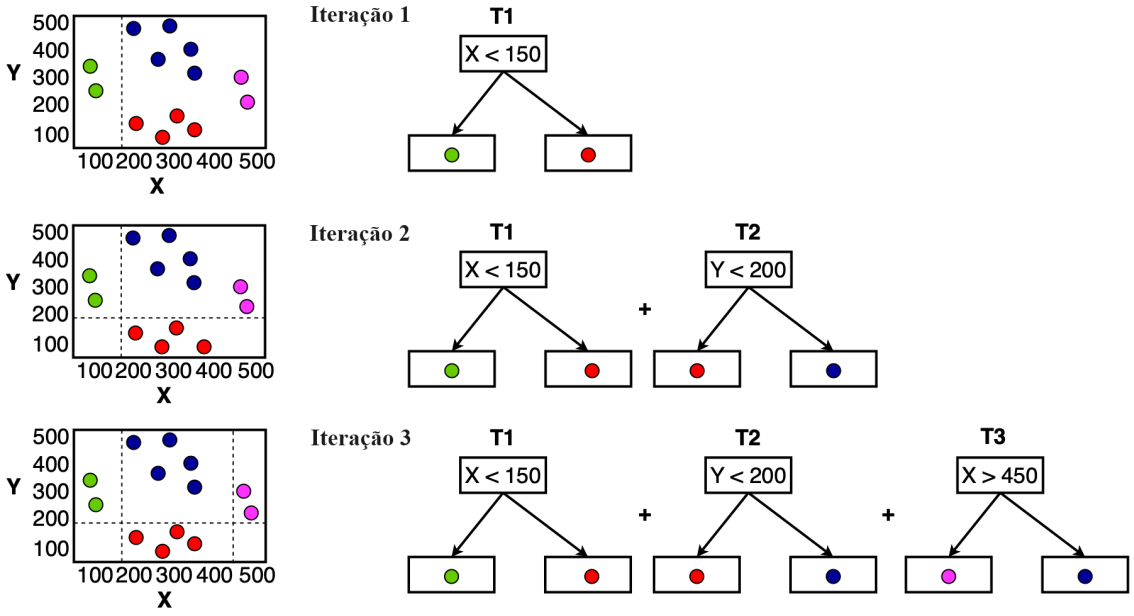
2.3 XGBoost

XGBoost, ou *Extreme Gradient Boosting* é um algoritmo de aprendizado de máquina baseado em árvores de decisão com a técnica de aumento de gradiente, proposto por Chen e Guestrin (2016). Ele utiliza o método de *ensemble* chamado *boosting* para combinar várias arvores de decisão sequencialmente. Esses componentes são explicados a seguir.

O método de *ensemble* é uma classe de métodos de aprendizado de máquina que treina vários *learners* para o mesmo problema e os combina para conseguir um melhor resultado. Os *learners* não necessariamente precisam ser o mesmo modelo e tão pouco utilizar os mesmos dados do problema (Wang *et al.*, 2019). Porém, no caso do XGBoost é utilizado apenas arvores de decisão, agrupadas sequencialmente (*boosting*). Existem três principais formas de agrupar modelos: o *Bagging*, em que modelos são treinados de forma independente em diferentes amostras do conjunto de dados (*bootstrap samples*), e suas previsões são combinadas, geralmente por meio da média ou votação, a fim de reduzir a variância; o *Boosting*, em que modelos são treinados sequencialmente, dando mais peso aos erros cometidos pelos modelos anteriores para melhorar o desempenho global, focando principalmente na redução do viés; e por fim, o *Stacking*, em que um meta-modelo é criado para combinar as previsões de modelos diversos, utilizando parte do conjunto de treino para modelos fracos e outra parte para um modelo mais robusto, como redes neurais, integrando sinergicamente as capacidades preditivas de diferentes modelos.

Árvore de decisão é um algoritmo de aprendizado de máquina que infere regras de decisão simples a partir dos dados utilizando uma estrutura semelhante a uma árvore, onde cada árvore contém um nó raiz, vários nós internos e várias folhas, o nó raiz contém todas as amostras, cada

Figura 2.4 – Algoritmo de Aumento de Gradiente



Adaptado de Chhetri *et al.* (2022)

ramificação divide os dados de acordo com o atributo selecionado, escolhido por um critério (como entropia de informação, proporção de granulação, índice de Gini). Esse processo se repete dividindo os nós internos até que a maioria dos dados seja classificada corretamente (Wang *et al.*, 2019).

Gradient boosting é um algoritmo semelhante ao *boosting*, porém, é utilizado para problemas de regressão (Bentéjac; Csörgõ; Martínez-Muñoz, 2021). Ele utiliza os resíduos (ou erros) do modelo anterior para produzir um novo modelo de forma incremental, como apresentado na Figura 2.4. Possuí uma natureza aditiva, adicionando arvores em cada iteração e minimizando suas perdas (Chhetri *et al.*, 2022).

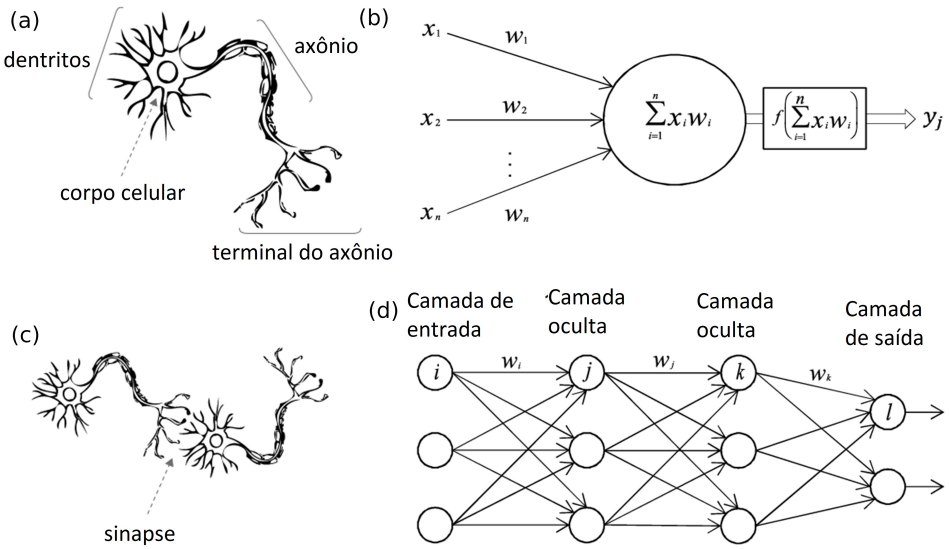
Por fim, o *Extreme Gradient Boosting* adiciona ao *Gradiente boosting* uma técnica de regularização L1 e L2 para evitar o sobre ajuste. Além disso, ele é capaz de paralelizar a construção da árvore e consegue lidar com valores ausentes no conjunto de dados (Belyadi; Haghighat, 2021).

Em relação ao presente trabalho, o XGBoost é empregado como um dos modelos.

2.4 Redes Neurais Artificiais

Rede Neural Artificial (ANN) é um tipo de modelo de aprendizado de máquina que foi inspirado no funcionamento do cérebro humano. Ela consiste em camadas interconectadas de nós, chamados de neurônios, que processam a informação. Cada neurônio recebe informação

Figura 2.5 – Um neurônio biológico comparado com uma ANN: (a) neurônio biológico; (b) neurônio artificial; (c) sinapse biológica; e (d) equivalente a uma sinapse em uma ANN.



Fonte: Adaptado de Meng, Hu e Ancey (2020).

de outros neurônios e aplica uma função de ativação para produzir uma saída (Rosa, 2013). A Figura 2.5 exemplifica como um neurônio biológico se assemelha com um artificial.

A função de ativação é um elemento essencial para o funcionamento de redes neurais, pois introduz a não-linearidade necessária para a aprendizagem de relações complexas de dados (Buduma; Locascio, 2017). Cada neurônio em uma rede neural processa sua entrada através de uma função de ativação antes de transmitir o resultado para camadas subsequentes. Como exemplificado na Figura 2.5 (b), em que, no neurônio, faz-se um produto escalar do vetor de entrada $\vec{x}$ com o vetor de pesos $\vec{w}$ podendo ser somado um *bias* após esta etapa - que não foi representado na figura - em seguida, aplica-se a função de ativação, representada por f. Esta, pode ser por exemplo uma função Sigmoid, Tangente hiperbólica ou uma *rectified linear unit* (ReLU). A recomendação padrão é utilizar a função *rectified linear unit* (ReLU) (Goodfellow; Bengio; Courville, 2016).

Existem muitos tipos de redes neurais, podendo ser classificadas por sua estrutura, fluxo de dados, densidade, função de ativação, entre outros. A *feed-foward Neural Network* é um desses tipos, em que os dados se propagam por meio de uma única direção, da camada de entrada para a camada de saída, sem formar ciclos ou retornar para camadas anteriores. Como exemplifica a Figura 2.5 (d), em que existe a camada de entrada (*input layer*), que recebe os dados; as camadas ocultas (*hidden layer*), responsáveis por aprender as representações e características dos dados, em que cada neurônio da camada oculta recebe as saídas dos neurônios da camada anterior e a camada de saída (*output layer*), que produzem o resultado final.

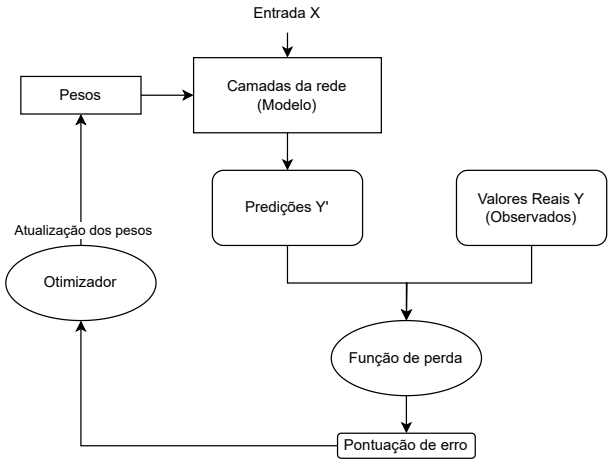
2.4.1 Treinamento de uma Rede Neural Artificial

Treinar uma rede neural artificial, significa minimizar os erros da rede. Para isso, a maioria dos algoritmos utilizam de um processo de otimização. A otimização consiste na tarefa de minimizar ou maximizar alguma função objetivo $f(\mathbf{x}; \mathbf{w})$ alterando o $\mathbf{w}$. Uma função que está sendo minimizada, pode ser chamada de *loss function*, *cost function* ou *error function*. (Goodfellow; Bengio; Courville, 2016).

A *loss function* ou função de perda é usada para calcular o erro. A partir dela, o algoritmo de otimização usa essa pontuação como um sinal de *feedback* para ajustar os valores dos pesos um pouco, em uma direção que vai diminuir a pontuação de perda para aquela mesma entrada. Este é um algoritmo central no *deep learning*, também chamado de *backpropagation* (Chollet, 2017).

A Figura 2.6 ilustra um fluxograma de uma iteração do processo de treinamento mencionado anteriormente. Inicialmente, são atribuídos valores arbitrários aos pesos w. Em seguida, a entrada $\mathbf{X}$ percorre a rede neural, gerando as predições Y'. Posteriormente, a *loss function* compara os valores preditos com os valores reais, resultando em uma pontuação de erro (*Loss score*). Por fim, essa pontuação é repassada para o otimizador, que realiza cálculos e atualiza os pesos da rede neural. Esse procedimento é repetido um número suficiente de vezes, iterando uma ou mais vezes para cada entrada em um conjunto de entradas. Dessa maneira, são produzidos pesos que minimizam a função de perda, tornando a rede neural progressivamente ajustada.

Figura 2.6 – Fluxograma do treinamento de uma Rede Neural para uma entrada



Fonte: Adaptado de Chollet (2017).

Para maior entendimento do processo descrito, ele pode ser dividido em duas fases:

- *Forward phase*: Os dados da entrada são propagados através da rede, camada por camada, com cada neurônio fazendo seu respectivo cálculo, até chegar na camada de saída. Nesta etapa os pesos já estão inicializados (Aggarwal, 2018).

- *Backward phase*: É a fase onde ocorre a etapa do *backpropagation*, que é responsável pela correção dos pesos usando uma técnica de otimização, como o Gradiente Descendente. Este, atualiza os dados gradualmente, diminuindo o erro e previsão de cada iteração (Aggarwal, 2018).

2.4.2 Algoritmo de retropropagação

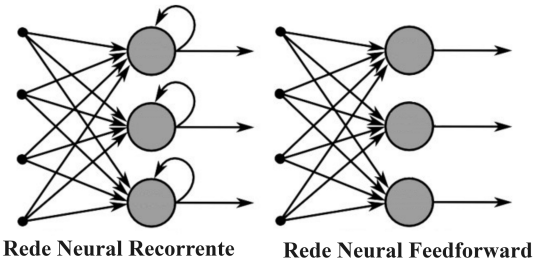
O *backpropagation*, ou retropropagação do erro, é uma técnica que calcula os gradientes das funções de perda em relação aos pesos da rede. Ele envolve o cálculo das derivadas parciais da *loss function* em relação a cada peso e viés (*bias*) da rede, a fim de determinar como esses parâmetros devem ser ajustados para minimizar o erro. O Gradiente Descendente é então usado para fazer os ajustes de fato, seguindo as direções dadas pelos gradientes calculados durante o processo de *backpropagation*.

Existem três variações do Gradiente Descendente, o Gradiente Descendente Estocástico, o *Mini-Batch Gradient Descent* e o *Batch gradient descent* (Ruder, 2017). Cada uma dessas, diferem na quantidade de dados que é usado para calcular o gradiente da função objetivo. Com base na quantidade de dados, deve-se ponderar entre a precisão da atualização do parâmetro e o tempo necessário para realizar a atualização.(Ruder, 2017)

2.5 Redes Neurais Recorrentes

Uma Rede Neural Recorrente (RNN) é um tipo de rede neural capaz de processar dados sequenciais, como por exemplo séries temporais ou linguagem natural. Ela é projetada para capturar as dependências temporais nos dados de entrada mantendo um estado oculto que é atualizado a cada intervalo de tempo. Esse estado oculto serve como uma memória das entradas passadas e é usado para fazer previsões sobre as entradas futuras (Goodfellow; Bengio; Courville, 2016). Elas podem ser construídas de diversas maneiras. Essencialmente, qualquer função envolvendo recorrência pode ser considerada uma RNN.

Figura 2.7 – Rede Neural Recorrente vs Rede Feed-Foward.



Fonte: Adaptado de DATA SCIENCE ACADEMY (2018).

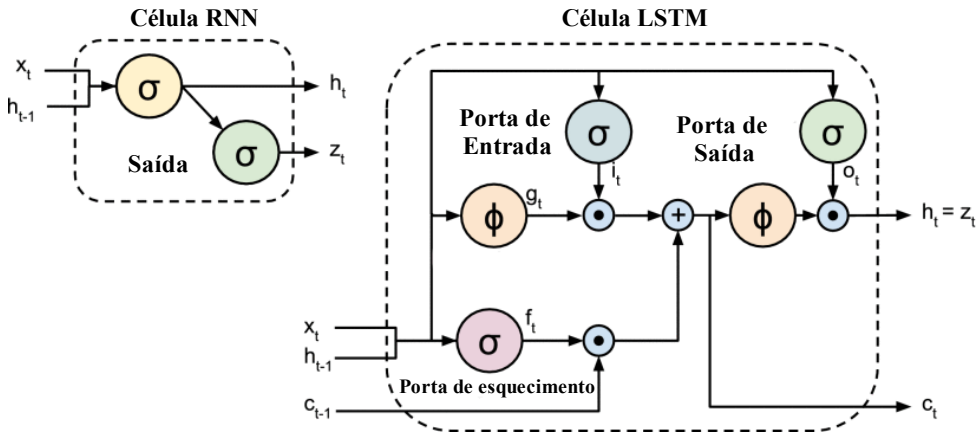
A Figura 2.7 mostra a diferença de uma RNN para uma rede *feed-foward*. O que caracteriza uma RNN é o *loop* de *feedback*, que faz com que existam duas fontes de entrada, o presente e o passado recente. Dessa forma, a rede é capaz de "lembrar" de dados das camadas anteriores (DATA SCIENCE ACADEMY, 2018).

As RNN apresentam um problema conhecido como *vanishing gradient*, em que o gradiente da função de perda em relação a rede se tornam muito pequenos à medida que são propagados através das camadas da rede. Isso pode fazer com que os pesos das camadas inferiores sejam atualizadas muito lentamente ou não sejam atualizados, o que pode levar ao baixo desempenho da rede (Hochreiter; Schmidhuber, 1997).

2.5.1 Long Short Term Memory

A *Long Short Term Memory* (LSTM) é uma arquitetura de rede neural recorrente que se propõe a resolver o problema de *vanishing gradient*, apresentada originalmente por Hochreiter e Schmidhuber (1997). Ela usa uma célula de memória que é uma unidade independente que pode armazenar informações de intervalos de tempo prolongados. Também usa três portas - *input gate*, *output gate* e *forget gate* - para controlar o fluxo de informações para dentro e fora da célula de memória (Hochreiter; Schmidhuber, 1997).

Figura 2.8 – Arquitetura LSTM



Fonte: Adaptado de Rassem, El-Beltagy e Saleh (2017).

A Figura 2.8 mostra a arquitetura de uma célula de memória de uma RNN comparada com uma LSTM. A LSTM possuí uma porta de entrada, que controla o fluxo de novas informações para a célula de memória, uma porta de saída, que controla o fluxo de informações para fora da célula de memória e a porta de esquecimento, que controla informações que não são mais necessárias (Rassem; El-Beltagy; Saleh, 2017).

Isso permite que a LSTM lembre ou esqueça informações seletivamente por longos períodos de tempo, o que é importante para resolver problemas com intervalos de tempo mínimos longos entre sinais relevantes. Em contraste, RNNs normais não possuem uma célula de memória

ou *gates* e sofrem com o problema do *vanishing gradient*, o que dificulta o aprendizado de dependências de longo prazo. Portanto, a LSTM é recomendado para tarefas não triviais em que a *random search* é inviável e leva a muito mais execuções bem-sucedidas do que seus concorrentes e aprende muito mais rápido (Hochreiter; Schmidhuber, 1996).

No contexto desta pesquisa o LSTM é utilizado como o modelo de aprendizado profundo para a predição das taxas de incidência da LV

2.6 Trabalhos Relacionados

Silva *et al.* (2017) realizam um estudo ecológico utilizando series temporais e análise da distribuição espacial da LV no Estado de Minas Gerais, Brasil, entre 2002 a 2013 usando duas metodologias para entender, caracterizar e quantificar a expansão da doença. A primeira trata-se de mapas temáticos de incidência e a segunda um modelo linear generalizado de Poisson (log-linear). A Regressão de Poisson mediu a variação do número médio de casos de um ano para o seguinte, para cada mesorregião. O estudo identifica as áreas com maior risco de transmissão e destaca a necessidade de medidas específicas de controle de doenças nessas áreas.

O trabalho de Rahmanian *et al.* (2020) é um estudo de série temporal que visa analisar os padrões temporais e prever a ocorrência de LV na província de Ardabil, Irã, entre 2000 e 2019. Foi utilizado dados anuais e mensais de 602 casos de LV humana na província e constatou que a ocorrência da doença não tem um padrão sazonal. O modelo ARIMA (5, 0, 1) foi selecionado como o melhor modelo para prever casos de LV humana, e os resultados podem ser usados como um sistema de alerta precoce para formuladores de políticas públicas e médicos de cuidados primários na prontidão de problemas de saúde pública antes do surto da doença. O estudo também destaca a importância da análise de séries temporais de dados de vigilância sobre doenças infecciosas para estimular novas hipóteses, prever eventos observados e estabelecer um sistema de controle de qualidade.

O trabalho de Mussumeci e Codeço Coelho (2020) propõem e comparam modelos de aprendizado de máquina para prever a incidência semanal de dengue em 790 cidades no Brasil. O documento destaca a importância da implantação oportuna de medidas de controle antes da estação de alta transmissão para gerenciar com eficácia doenças sazonais, como a dengue. A disponibilidade de previsões precisas de incidência pode ser decisiva no controle dessas doenças. Este usa séries temporais multivariadas como preditores e também utiliza séries temporais de cidades semelhantes para capturar o componente espacial da transmissão de doenças. O modelo de rede neural recorrente LSTM obteve o melhor desempenho na previsão de incidência futura de dengue em cidades de diferentes tamanhos. O artigo também inova ao propor usar uma abordagem multivariada. O artigo explora o conceito de similaridade epidemiológica para selecionar os preditores para cada cidade e, assim, compensar o curto intervalo de tempo de nossas séries com um amplo conjunto de séries semelhantes. Também é discutida a etapa de pré-processamento

de dados, onde a remoção de *outliers* e o preenchimento de dados ausentes são as principais técnicas para melhorar a qualidade dos dados. Além disso, a seleção de recursos ideais é realizada usando a *Neighbour Count based Dragonfly Electric Fish Optimization* baseada em contagem de vizinhos. O artigo conclui que os modelos de aprendizado de máquina propostos podem fornecer previsões precisas da incidência de dengue, que podem ser usadas para implantar medidas de controle em tempo hábil para gerenciar com eficácia doenças sazonais como a dengue.

O trabalho de Gioia, Barros e Silva (2022) tem como objetivo analisar a relação entre variáveis socioeconômicas e doenças tropicais negligenciadas para auxiliar gestores no desenho de políticas públicas para redução de casos. O estudo avalia quais variáveis socioeconômicas são mais importantes para a classificação de risco de três doenças negligenciadas: hanseníase, leishmaniose cutânea e dengue, com base em algoritmos de aprendizado de máquina. A área de estudo foi delimitada aos municípios do estado de Goiás e do Distrito Federal – Brasil. Três algoritmos baseados em árvores de decisão foram avaliados: *Random Forest* (RF), XGBoost e C5.0. O hiperparâmetro otimizado foi o número de ensaios, que neste estudo foi definido igual a 20. As variáveis socioeconômicas mais importantes para a predição das classes de risco para dengue e hanseníase foram semelhantes, destacando-se as condições de baixa renda, alfabetização e raça como preditores mais importantes variáveis. O estudo constatou que, para as classes de risco da dengue, tanto o algoritmo de RF quanto o XGBoost apresentaram resultados semelhantes. O estudo indicou como importantes as variáveis abastecimento de água, alfabetização, raça e moradia. Porém, para as classes de risco de leishmaniose tegumentar, os algoritmos apresentaram acurácia inferior, inviabilizando a avaliação de possíveis variáveis preditivas para o modelo. Os três algoritmos avaliados revelaram desempenho preditivo aproximado; no entanto, o RF foi ligeiramente superior.

O trabalho de Li *et al.* (2020) visa prever o número de casos de leishmaniose visceral na prefeitura de Kashgar de Xinjiang, China usando a média móvel integrada autorregressiva (ARIMA). e modelo híbrido ARIMA-EGARCH. O estudo utilizou dados mensais de casos de leishmaniose visceral de 2004 a 2016, com os dados amostrais entre 2004 e 2015 usados para estimativa e os dados amostrais em 2016 usados para previsão. O estudo constatou que o modelo ARIMA (2, 1, 2) (1, 1, 1)12 era consistente com os dados reais coletados de 2004 a 2015. No entanto, os casos previstos não cumpriram o número de casos observados. Portanto, os pesquisadores tentaram estabelecer um modelo híbrido ARIMA-EGARCH para ajustar a leishmaniose visceral. Finalmente, o modelo ARIMA (2, 1, 2) (1, 1, 1)12- EGARCH (1, 1) mostrou uma boa estimativa ao lidar com agrupamento de volatilidade na série de dados. O modelo combinado foi determinado como o melhor modelo de predição com o erro quadrático médio (RMSE) de 7,23 na fase de validação, o que significa que esse modelo tem alta validade e racionalidade e pode ser usado para predição de curto prazo de leishmaniose visceral e podem ser aplicadas na prevenção e controle da doença. O estudo fornece uma base científica para controlar a disseminação da leishmaniose visceral na província de Kashgar em Xinjiang, China.

Os trabalhos supracitados, em sua maioria utilizam outras abordagens de AI que não a LSTM ou a XGBoost, porém, trazem *insights* para o processo de predição, como métricas utilizadas. O trabalho de Mussumeci e Codeço Coelho (2020) por outro lado, utiliza LSTM, no entanto, é estudado casos de dengue, mas, apesar disso, ele se tornou relevante como base para a construção da topologia da rede LSTM deste trabalho. Por conta dos trabalhos analisarem outras regiões do mundo e em períodos diferentes, nenhum deles é diretamente comparável com este.

3 Metodologia

3.1 Desenho do estudo

Foi realizado um estudo epidemiológico na Macrorregião de Saúde “Centro” de Minas Gerais, Brasil. Foram coletados dados de números de casos de LV e dados sociodemográficos de 79 municípios, estes, foram divididos em conjunto de treinamento e teste, como detalhado e listado no Apêndice A. Estes dados foram usados para treinar uma rede neural LSTM e uma XGBoost para predizer a incidência de casos da doença baseado em características de um determinado município em anos anteriores.

3.2 Fonte de dados

O número de casos confirmados por município foi retirado do Sistema de Informação de Agravos de Notificação (Sinan) no período de 2007 à 2021. Eles foram validados pelo Grupo Técnico de Leishmanioses / Coordenação-Geral de Zoonoses e Vetores (CGZV) / Departamento de Epidemiologia e Doenças Transmitidas por Vetores (DEDT)/ Secretaria de Vigilância em Saúde (SVSA) / Ministério da Saúde (MS). Os dados da população residente por ano por município foram retirados de estimativas preliminares elaboradas pelo Ministério da Saúde / SVSA / Departamento de Análise da Situação de Saúde (DAENT) / Coordenação-Geral de Informações e Análise Epidemiológica (CGIAE) no período de 2000 à 2021. Os demais dados, como indicadores socioeconômicos foram retirados do IEPS Data[1], que é uma iniciativa do Instituto de Estudos para Políticas de Saúde (IEPS), uma organização sem fins lucrativos, independente e apartidária, cujo único objetivo é contribuir para o aprimoramento das políticas públicas do setor de saúde no Brasil.

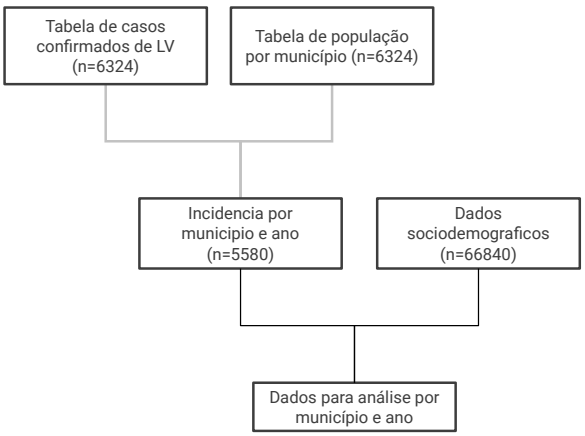
3.3 Dados

Os dados são conjuntos de séries temporais, onde cada serie representava o histórico de casos de leishmaniose por município, além de variáveis sociodemográficas, também para cada município e ano.

A Figura 3.1 ilustra as etapas para selecionar os dados que foram utilizados no projeto. O processo envolveu a combinação das tabelas de casos por município e ano com a tabela de habitantes por município e ano. Em seguida, foi realizado o cálculo da incidência por município e por ano, onde a quantidade de casos confirmados foi dividida pela população do município e

[1] https://iepsdata.org.br/

Figura 3.1 – Diagrama de seleção e limpeza dos dados



Fonte: Elaborado pelo autor.

multiplicada por 100.000 para cada ano. Após obter a incidência por município e ano, os dados sociodemográficos foram associados à incidência usando o código do município e o ano como índice. Por fim, filtrou-se apenas pela macrorregião central de Minas Gerais. Durante essas etapas, registros que não tinham correspondência entre as tabelas foram excluídos. A tabela final contém as seguintes colunas:

- *ds*: Ano do registro.
- *População*: Número de habitantes do município.
- *idhm*: O Índice de Desenvolvimento Humano Municipal (IDHM) é um indicador composto que mede o desenvolvimento humano de uma determinada região. Ele considera fatores como saúde, educação e renda, fornecendo uma visão geral do nível de qualidade de vida e desenvolvimento socioeconômico.
- *idhm_renda*: Essa é uma componente específica do IDHM, que avalia o nível de renda da população de uma região. É calculado considerando a renda média per capita dos habitantes.
- *idhm_educ*: Essa é outra componente do IDHM, que avalia o nível educacional da população. Leva em conta indicadores relacionados à alfabetização e à escolaridade das pessoas.
- *idhm_long*: O IDHM Longevidade é a terceira componente do IDHM, que mede a expectativa de vida ao nascer da população. Ele reflete a qualidade dos cuidados de saúde, saneamento básico e outras condições que afetam a longevidade das pessoas.
- *renda_dom_pc*: Essa coluna indica a renda média por domicílio, calculada dividindo a renda total dos domicílios pelo número de domicílios na região. É um indicador econômico que reflete a distribuição de renda dentro da área.

- *pct_san_adeq*: O percentual de saneamento adequado é a proporção da população que tem acesso a saneamento básico adequado, o que inclui serviços de água potável e esgoto sanitário. É um indicador de qualidade de vida e saúde pública.

- *pct_rural*: Essa coluna representa o percentual da população que vive em áreas rurais, em relação à população total da região. Pode refletir características demográficas e econômicas da área.

- *desp_tot_saude_pc_mun*: Os gastos totais com saúde por habitante no município. Isso inclui despesas do governo local com serviços de saúde, como hospitais, clínicas, programas de prevenção, entre outros.

3.4 Tratamento dos dados

Antes de utilizar os dados para alimentar na rede neural, os mesmos foram submetidos a um processo de normalização utilizando a classe *StandardScaler* do pacote *Scikit-learn*. Esse procedimento foi essencial para padronizar as diferentes escalas dos dados, normalizando cada *feature* e também a variável *target* para a mesma escala, garantindo assim, que a rede neural pudesse convergir de forma mais eficiente durante o treinamento. A predição do modelo foi posteriormente convertida para a escala original.

Após esta etapa, os dados que estavam em um *pandas DataFrame*, foram convertidos para um *array* tridimensional do *numpy*, de forma que cada serie temporal estivesse em uma dimensão do tensor, e suas observações nas outras duas dimensões restantes. A Figura 3.2 exemplifica o fluxo.

Para cada cidade foi utilizado o período de 2010 à 2020 para prever a incidência de 2021, por isso, o *array* de *target* continha apenas o ultimo *timepoint* da série.

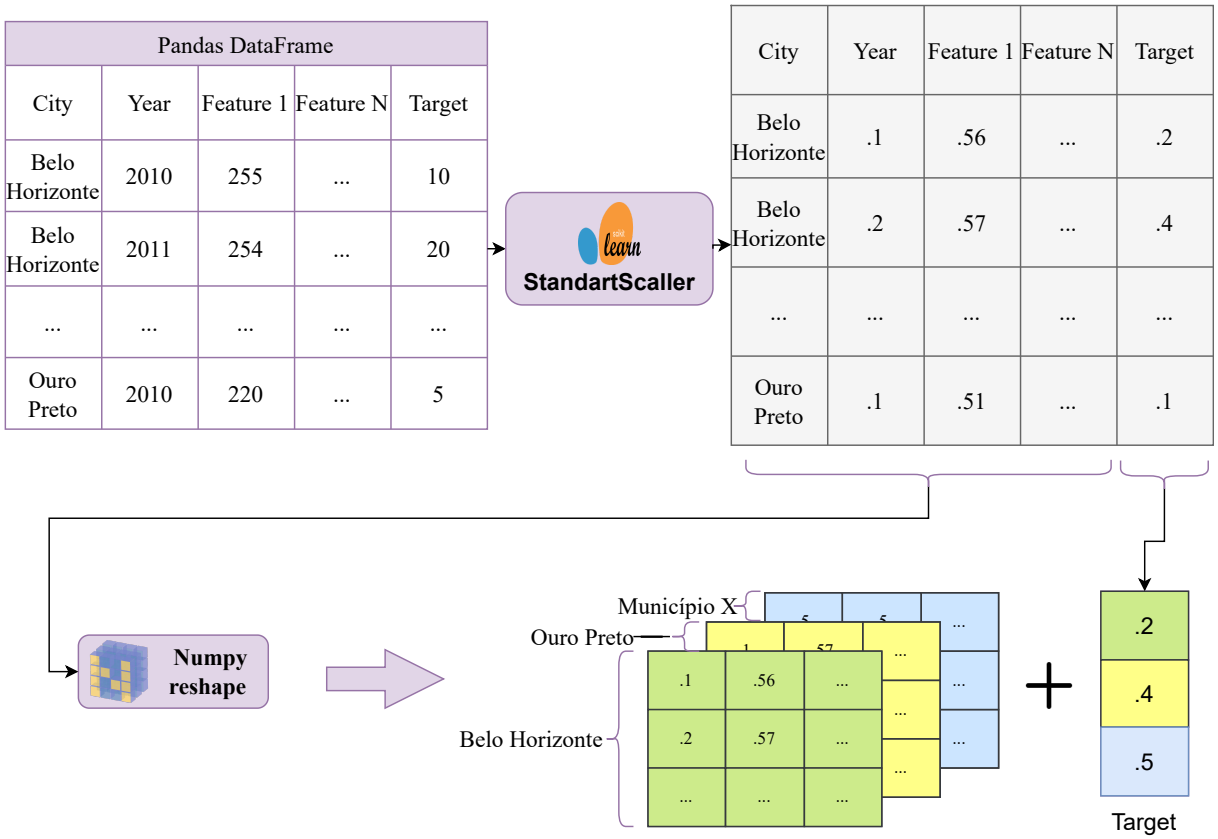
3.5 Parametrização e treinamento do Modelo LSTM

Foi ajustado um modelo LSTM, utilizando a biblioteca *keras* com o *tensorflow*, conforme a topologia descrita na Tabela 3.1, com base no modelo de Mussumeci e Codeço Coelho (2020). Foi testado com 4 *loss functions* diferentes e 6 otimizadores, como mostrado na Figura 4.1. Uma explicação mais detalhada sobre o significado desses elementos pode ser encontrada na subseção 2.4.1.

Para treinamento do modelo foi necessária, a separação do conjunto de teste e o conjunto de treino. O particionamento foi de: 75% das séries temporais dos municípios foram usados para treinamento do modelo e o restante (25%) para teste.

Após esta etapa, o modelo foi treinado utilizando 50 épocas, uma quantidade definida com base em avaliações empíricas de diferentes números de épocas. Observou-se que 50 épocas

Figura 3.2 – Fluxo do tratamento dos dados



Utiliza-se o *StandartScaller do scikitlearn* para colocar normalizar os dados, então, as *features* são convertidas em um *array* multidimensional *numpy*, onde cada município fica em uma dimensão z, enquanto cada observação da serie temporal fica na dimensão x,y. O *target* de cada serie temporal, que será usado para treinar e validar o modelo é colocado em um *array* a parte.

Fonte: Elaborado pelo autor.

Tabela 3.1 – Topologia da rede neural LSTM.

Camada (tipo)	Forma da Saída	Parâmetros
lstm_1 (LSTM)	(None, 11, 11)	968
lstm_2 (LSTM)	(None, 11, 11)	1012
lstm_3 (LSTM)	(None, 11)	1012
dense_1 (Dense)	(None, 1)	12

Fonte: Elaborado pelo autor.

resultavam em um erro menor. Os detalhes de como esta etapa funciona estão descritos na subseção 2.4.1.

3.6 Parametrização e treinamento do Modelo XGBoost

Foi utilizado a biblioteca *mlforecast* e a biblioteca *optuna* para fazer a optimização dos hiperparâmetros do modelo, são eles: a taxa de aprendizagem do modelo; a profundidade máxima das árvores; a soma mínima do peso da instância necessária em um filho; a fração das instâncias que são amostradas aleatoriamente durante o processo de treinamento de cada árvore e a proporção de subamostra de colunas ao construir cada árvore. Também foi feita a divisão de dados de treinamento e teste em uma proporção de 75% das séries temporais para treinamento e 25% delas para teste.

3.7 Métricas para avaliação do modelo

As métricas utilizadas para a análise do modelo foram o coeficiente de determinação (R^2) e o Erro Percentual Absoluto Médio (MAPE). O R^2 é uma métrica que representa a proporção da variabilidade nos dados de saída que é explicada pelo modelo. Em outras palavras, indica o quão bem o modelo se ajusta aos dados observados. Quanto mais proximo de 1, melhor o modelo se ajusta aos dados.

Por sua vez, o Erro Percentual Absoluto Médio (MAPE) é uma métrica que calcula a porcentagem média da diferença absoluta entre os valores previstos pelo modelo e os valores reais. É uma medida de precisão que expressa o erro médio como uma porcentagem do valor real. Quanto menor o valor do Erro Percentual Absoluto Médio (MAPE), melhor é a precisão do modelo em fazer previsões.

4 Resultados

Neste capítulo são apresentados, interpretados e analisados todos os resultados alcançados no trabalho.

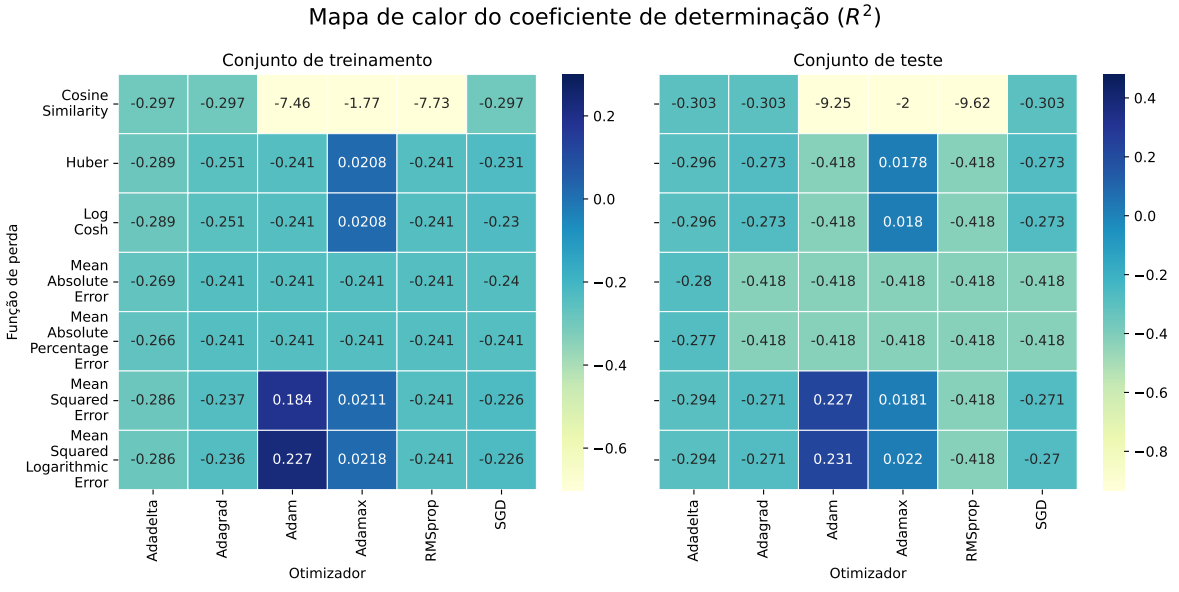
4.1 Avaliação dos Modelos & Discussão

Um dos modelos apresentou um resultado promissor, enquanto o outro apresentou um resultado não satisfatório. Na predição do modelo XGBoost, mesmo utilizando uma ferramenta de otimização de hiperparâmetros, o coeficiente de determinação (R^2) do modelo foi de -0.27 no melhor caso, ou seja, se fosse criado um modelo que sempre retorna o valor da média das observações, ele teria um erro menor do que o modelo do XGBoost.

O LSTM foi testado utilizando diferentes funções de perda e otimizadores, então foi calculado o R^2 de cada uma dessas combinações e apresentado na Figura 4.1. No conjunto de teste, a pior combinação foi o otimizador RMSprop com a função de perda Similaridade de Cosseno, com $-9,62$ de R^2. A melhor foi a combinação do otimizador Adam com a função de perda Erro Médio Quadrático Logarítmico, com $0,231$ de R^2.

Nesse contexto, mesmo que o coeficiente de determinação de $0,231$ possa parecer baixo à primeira vista, é considerado promissor dada a complexidade da doença. A Leishmaniose Visceral envolve diversos fatores, como o vetor transmissor, o hospedeiro vertebrado (cão),

Figura 4.1 – Comparação do desempenho dos otimizadores e funções de perda para o LSTM

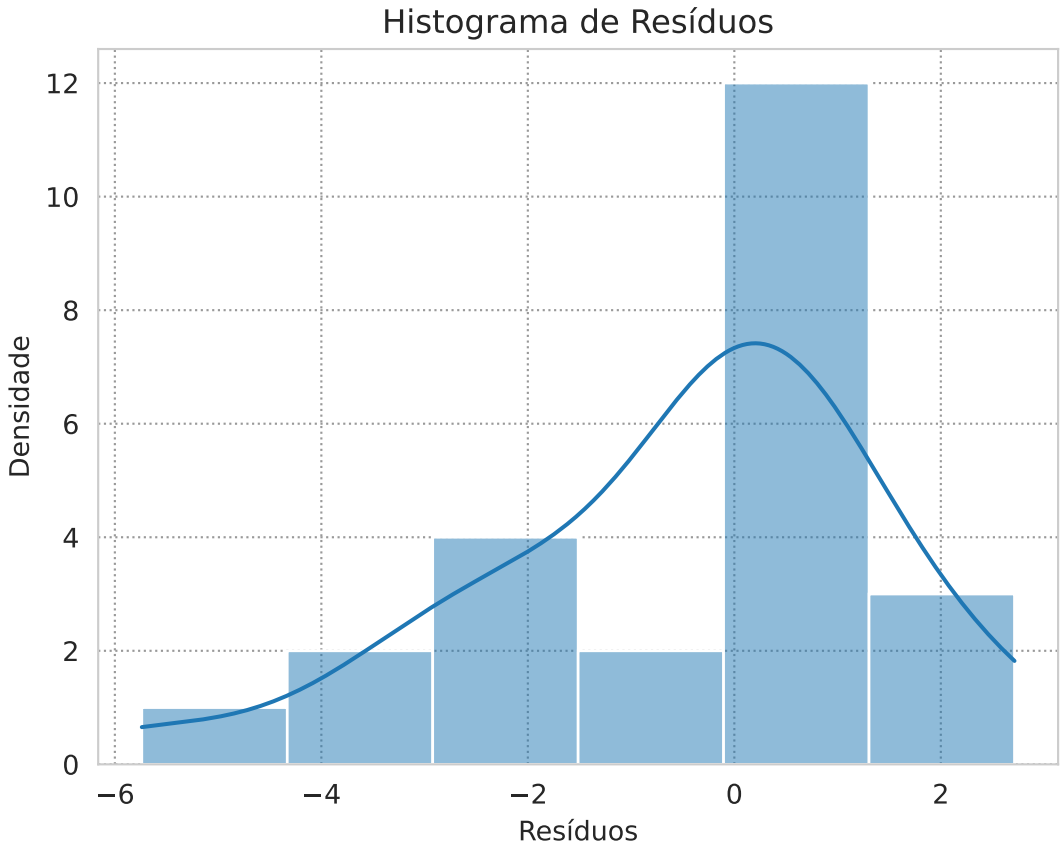


Fonte: Elaborado pelo autor

questões ambientais, climáticas e o desmatamento. Dado o desafio de modelar uma enfermidade tão complexa com poucas variáveis, é natural que o modelo tenha uma explicação limitada.

Analisando o modelo mais promissor (LSTM), com o melhor conjunto de otimizador e função de perda, observa-se que o Erro Percentual Absoluto Médio (MAPE) é de 1,78%, ou seja, um erro baixo. Ao examinar o histograma de resíduos na Figura 4.2, nota-se que a maior parte da densidade está próxima de zero, sugerindo que em muitas das previsões, os valores estavam próximos da incidência observada.

Figura 4.2 – Histograma de Resíduos



Histograma dos resíduos das predições do modelo de todos os municípios com Estimação de Densidade Kernel; Resíduos superiores a zero indicam que o modelo identificou uma incidência superior àquela observada, enquanto resíduos negativos denotam que o modelo identificou uma incidência inferior ao observado. Resíduos iguais a zero denotam que o modelo realizou uma previsão precisa, sem apresentar discrepâncias.

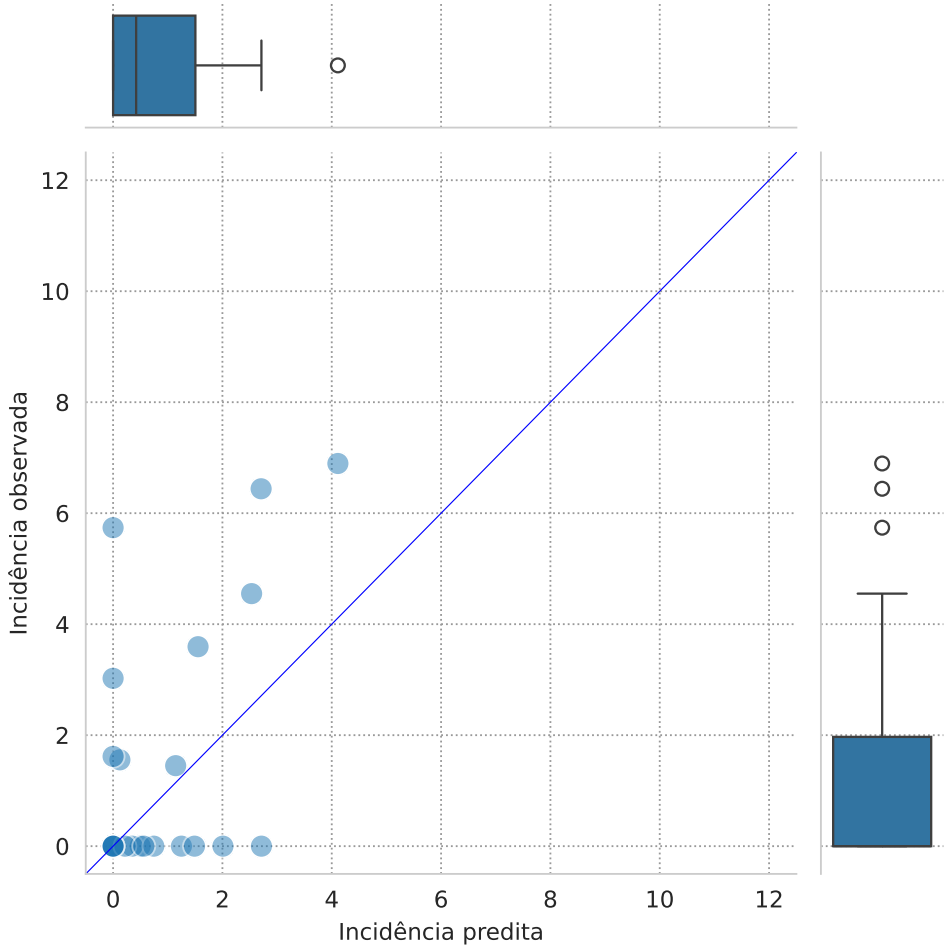
Fonte: Elaborado pelo autor

Na Figura 4.3, nota-se que mais de 50% da incidência observada no conjunto de teste é nula. Essa condição pode estar exercendo influência sobre o modelo, levando-o a realizar previsões predominantemente mais baixas, conforme evidenciado na própria figura. Nela, observa-se que 75% das predições situam-se entre 0 e aproximadamente $1\frac{1}{2}$.

No Apêndice C é possível ver as series temporais da incidência de cada município e a

previsão do modelo LSTM para o ano de 2021.

Figura 4.3 – Valores preditos vs Valores observados



Na figura é apresentado as distribuições observadas versus preditas para todas as predições do modelo, geradas para todos os municípios. O eixo x codifica a distribuição de pontos preditos, enquanto o eixo y os valores observados. Cada ponto (x, y) dá origem ao mapa de pontos, representando uma predição e uma observação, em que as áreas em azul escuro são onde se concentram a maioria dos valores. Se o valor previsto e o valor observado estiverem próximos da linha $x = y$, o valor previsto é consistente com o historicamente esperado para aquela cidade naquele ano.

Fonte: Elaborado pelo autor

5 Considerações Finais

5.1 Conclusão

O presente trabalho teve como objetivo principal desenvolver dois modelos de aprendizado de máquina, um utilizando a arquitetura LSTM, e o outro utilizando o XGBoost para prever os casos de LV com base em parâmetros sociais e econômicos específicos de cada município da Região Central do Estado de Minas Gerais. O objetivo era antecipar a incidência da doença nos municípios, contribuindo para a criação de políticas públicas e atuação mais eficiente por parte das autoridades competentes no controle dessa enfermidade.

A metodologia adotada envolveu a utilização de múltiplas series temporais, uma para cada município, o qual possuía dados epidemiológicos e socioeconômicos para treinar os modelos. Foram consideradas séries temporais do período de 2010 à 2020, visando prever a incidência para o ano de 2021.

Os resultados obtidos indicaram que o desempenho do modelo XGBoost não foi satisfatório para o problema proposto, mesmo após a optimização de seus hiperparâmetros, resultando em um coeficiente de determinação de $-0,27$. Por outro lado, o modelo *Long Short Term Memory* (LSTM) apresentou resultados promissores, com um Erro Percentual Absoluto Médio (MAPE) de $1,78\%$. Embora o R^2 de $0,231$ possa parecer baixo, é importante considerar a complexidade intrínseca da Leishmaniose Visceral. Dessa forma, é razoável esperar que um modelo que utiliza um número limitado de variáveis ofereça uma explicação igualmente limitada para uma doença tão complexa.

Além disso, o modelo demonstrou uma inclinação para prever uma incidência baixa, visto que 75% das predições se situaram abaixo de $1\frac{1}{2}$. No entanto, é necessário salientar que a incidência observada para o ano predito é igualmente baixa, com mais de 50% dos casos registrando uma incidência de 0. Nesse contexto, o modelo parece estar alinhado com os dados disponíveis. Contudo, é importante ressaltar que há a possibilidade de que, em municípios com incidência mais elevada, o modelo continue prevendo valores relativamente baixos.

A contribuição deste trabalho está na aplicação de técnicas de aprendizado de máquina para prever a incidência de uma doença em nível municipal. A proposta de utilizar redes neurais recorrentes e o algoritmo XGBoost com series temporais, abre caminho para novas pesquisas de modelos para prever a incidência de doenças.

Em resumo, este estudo representa um passo na aplicação de técnicas para abordar problemas de saúde pública, destacando a importância da integração de dados e inteligência artificial para prever e controlar doenças através da alocação eficiente de recursos do Estado.

5.2 Trabalhos Futuros

Considerando os desdobramentos e conclusões evidenciadas neste estudo, há várias perspectivas para refinamento e ampliação do trabalho. O estudo atual se concentrou em uma macrorregião específica de Minas Gerais, o que oferece espaço para pesquisas futuras abrangerem todos os municípios do estado ou mesmo do país. A inclusão de uma variedade mais ampla de localidades permitirá uma visão mais abrangente da dinâmica da Leishmaniose Visceral, considerando as especificidades de cada região.

Além disso, oportunidades adicionais residem na exploração de diferentes topologias do LSTM. A comparação e análise de desempenho entre diferentes topologias pode oferecer *insights* valiosos sobre qual modelo se adapta melhor aos dados e às características da incidência de LV.

Atualmente, o modelo realiza previsões para o próximo ano, fundamentando-se nos 11 anos anteriores. Como direção para pesquisas posteriores, sugere-se a redução do período necessário para previsões e a ampliação do horizonte de tempo predito pelo modelo. Isso permitirá uma visão mais prospectiva e de longo prazo, auxiliando ainda mais nas estratégias de planejamento e controle da LV.

A inclusão de mais variáveis sociais, ambientais e epidemiológicas pode aprimorar a capacidade preditiva do modelo. Explorar dados adicionais, como mobilidade populacional, padrões climáticos e características específicas dos vetores, pode enriquecer a complexidade do modelo, proporcionando uma compreensão mais abrangente dos fatores que influenciam a incidência da LV.

A presença recorrente de casos nulos pode ter impactado nas predições do modelo. Portanto, futuras pesquisas podem se dedicar à avaliação e implementação de técnicas específicas para lidar com dados desbalanceados, mitigando o efeito desses casos nulos na precisão das previsões. Ao abordar essas direções para pesquisas futuras, é possível aprimorar ainda mais a aplicação de modelos de aprendizado de máquina na previsão e controle da Leishmaniose Visceral, contribuindo para uma abordagem mais eficaz no enfrentamento dessa doença tropical negligenciada.

Referências

AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2018. ISBN 3319944622.

BELYADI, H.; HAGHIGHAT, A. Chapter 5 - supervised learning. In: BELYADI, H.; HAGHIGHAT, A. (Ed.). **Machine Learning Guide for Oil and Gas Using Python**. Gulf Professional Publishing, 2021. p. 169–295. ISBN 978-0-12-821929-4. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780128219294000044>.

BENTÉJAC, C.; CSÖRGÕ, A.; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. **Artificial Intelligence Review**, v. 54, p. 1937–1967, 3 2021. ISSN 0269-2821. Disponível em: <https://link.springer.com/10.1007/s10462-020-09896-5>.

BUDUMA, N.; LOCASCIO, N. **Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms**. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2017. ISBN 1491925612.

CHEN, T.; GUESTRIN, C. Xgboost. In: . ACM, 2016. v. 13-17-August-2016, p. 785–794. ISBN 9781450342322. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>.

CHHETRI, T. R. *et al.* A combined system metrics approach to cloud service reliability using artificial intelligence. **Big Data and Cognitive Computing**, v. 6, n. 1, 2022. ISSN 2504-2289. Disponível em: <https://www.mdpi.com/2504-2289/6/1/26>.

CHOLLET, F. **Deep Learning with Python**. 1st. ed. USA: Manning Publications Co., 2017. ISBN 1617294438.

DATA SCIENCE ACADEMY. **Deep Learning Book**. 2018. <https://www.deeplearningbook.com.br/>. Acesso em: 05 Agosto 2023.

DESJEUX, P. The increase in risk factors for leishmaniasis worldwide. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 95, n. 3, p. 239–243, 06 2001. ISSN 0035-9203. Disponível em: <https://doi.org/10.1016/S0035-9203(01)90223-8>.

DESJEUX, P. Leishmaniasis: current situation and new perspectives. **Comparative Immunology, Microbiology and Infectious Diseases**, v. 27, n. 5, p. 305–318, 2004. ISSN 0147-9571. Advances on some vector-borne diseases. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0147957104000232>.

DUARTE, R. V. *et al.* Influence of climatic variables on the number of cases of visceral leishmaniasis in an endemic urban area. **Journal of Global Health Economics and Policy**, JoGH Ltd, v. 2, jul. 2022. Disponível em: <https://doi.org/10.52872/001c.36750>.

GIOIA, T.; BARROS, J. R.; SILVA, R. R. da. Fatores socioeconómicos e algoritmos de machine learning aplicados à predição de risco de doenças negligenciadas.: Estudo de caso nos municípios do estado de goiás e distrito federal, brasil. **Finisterra**, v. 57, n. 121, p. 109–123, Dez. 2022. Disponível em: <https://revistas.rcaap.pt/finisterra/article/view/28635>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HARHAY, M. O. *et al.* Who is a typical patient with visceral leishmaniasis? characterizing the demographic and nutritional profile of patients in brazil, east africa, and south asia. **The American Society of Tropical Medicine and Hygiene**, The American Society of Tropical Medicine and Hygiene, Arlington VA, USA, v. 84, n. 4, p. 543 – 550, 2011. Disponível em: <https://www.ajtmh.org/view/journals/tpmd/84/4/article-p543.xml>.

HOCHREITER, S.; SCHMIDHUBER, J. Lstm can solve hard long time lag problems. In: **Proceedings of the 9th International Conference on Neural Information Processing Systems**. Cambridge, MA, USA: MIT Press, 1996. (NIPS'96), p. 473–479.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>.

LI, H.-l. *et al.* Predicting the number of visceral leishmaniasis cases in kashgar, xinjiang, china using the arima-egarch model. **Asian Pacific Journal of Tropical Medicine**, v. 13, n. 2, p. 67–73, 2020. ISSN 1995-7645. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1995764519307274>.

MENG, Z.; HU, Y.; ANCEY, C. Using a data driven approach to predict waves generated by gravity driven mass flows. **Water**, v. 12, 02 2020.

Ministério da Saúde. **Manual de Vigilância e Controle da Leishmaniose Visceral**. [S.l.]: Ministério da Saúde, 2006. <https://bvsms.saude.gov.br/bvs/publicacoes/manual_vigilancia_controle_leishmaniose_visceral.pdf>.

MUSSUMECI, E.; Codeço Coelho, F. Large-scale multivariate forecasting models for dengue - lstm versus random forest regression. **Spatial and Spatio-temporal Epidemiology**, v. 35, p. 100372, 2020. ISSN 1877-5845. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877584520300502>.

Organização Pan-Americana da Saúde. **Leishmanioses. Informe Epidemiológico das Américas**. 2020. Organização Pan-Americana da Saúde. Acesso em 21 Jan 2024. Disponível em: <https://iris.paho.org/handle/10665.2/53091>.

ORYAN, A.; AKBARI, M. Worldwide risk factors in leishmaniasis. **Asian Pacific Journal of Tropical Medicine**, v. 9, n. 10, p. 925–932, 2016. ISSN 1995-7645. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1995764516301572>.

PALTRINIERI, S. *et al.* Laboratory tests for diagnosing and monitoring canine leishmaniasis. **Veterinary Clinical Pathology**, v. 45, p. 552–578, 12 2016. ISSN 0275-6382. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/vcp.12413>.

RAHMANIAN, V. *et al.* Temporal analysis of visceral leishmaniasis between 2000 and 2019 in ardabil province, iran: A time-series study using arima model. **Journal of Family Medicine and Primary Care**, v. 9, n. 12, p. 6061–6067, December 2020.

RASSEM, A.; EL-BELTAGY, M.; SALEH, M. Cross-country skiing gears classification using deep learning. 06 2017.

READY, P. D. Epidemiology of visceral leishmaniasis. **Clinical Epidemiology**, Dove Medical Press, v. 6, p. 147–154, 2014. PMID: 24833919. Disponível em: <https://www.tandfonline.com/doi/abs/10.2147/CLEP.S44267>.

ROCHA, I. *et al.* Effectiveness of the brazilian visceral leishmaniasis surveillance and control programme in reducing the prevalence and incidence of leishmania infantum infection. **Parasites & Vectors**, v. 11, 11 2018.

ROSA, J. L. G. Biologically plausible artificial neural networks. In: SUZUKI, K. (Ed.). **Artificial Neural Networks**. Rijeka: IntechOpen, 2013. cap. 2. Disponível em: <https://doi.org/10.5772/54177>.

RUDER, S. **An overview of gradient descent optimization algorithms**. 2017.

SCARPINI, S. *et al.* Visceral leishmaniasis: Epidemiology, diagnosis, and treatment regimens in different geographical areas with a focus on pediatrics. **Microorganisms**, Switzerland, v. 10, n. 10, set. 2022.

SILVA, T. A. M. da *et al.* Spatial and temporal trends of visceral leishmaniasis by mesoregion in a southeastern state of brazil, 2002-2013. **PLOS Neglected Tropical Diseases**, Public Library of Science (PLoS), v. 11, n. 10, p. e0005950, out. 2017. Disponível em: <https://doi.org/10.1371/journal.pntd.0005950>.

WANG, Y. *et al.* A hybrid ensemble method for pulsar candidate classification. **Astrophysics and Space Science**, v. 364, p. 139, 8 2019. ISSN 0004-640X. Disponível em: <http://link.springer.com/10.1007/s10509-019-3602-4>.

WERNECK, G. Forum: Geographic spread and urbanization of visceral leishmaniasis in brazil. introduction. **Cadernos de saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública**, v. 24, p. 2937–40, 01 2009.

WERNECK, G. L. Expansão geográfica da leishmaniose visceral no brasil. **Cadernos de Saúde Pública**, Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, v. 26, n. 4, p. 644–645, Apr 2010. ISSN 0102-311X. Disponível em: <https://doi.org/10.1590/S0102-311X2010000400001>.

World Health Organization. **Leishmaniasis: situation and trends**. 2022. World Health Organization. Acesso em 21 Jan 2024. Disponível em: <http://www.who.int/gho/neglected_diseases/leishmaniasis/en/>.

World Health Organization. **Leishmaniasis**. 2023. World Health Organization. Acesso em 21 Jan 2024. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>.

Apêndices

APÊNDICE A – Lista de municípios

Tabela A.1 – Municípios utilizados para o treino e teste dos modelos

(continua)

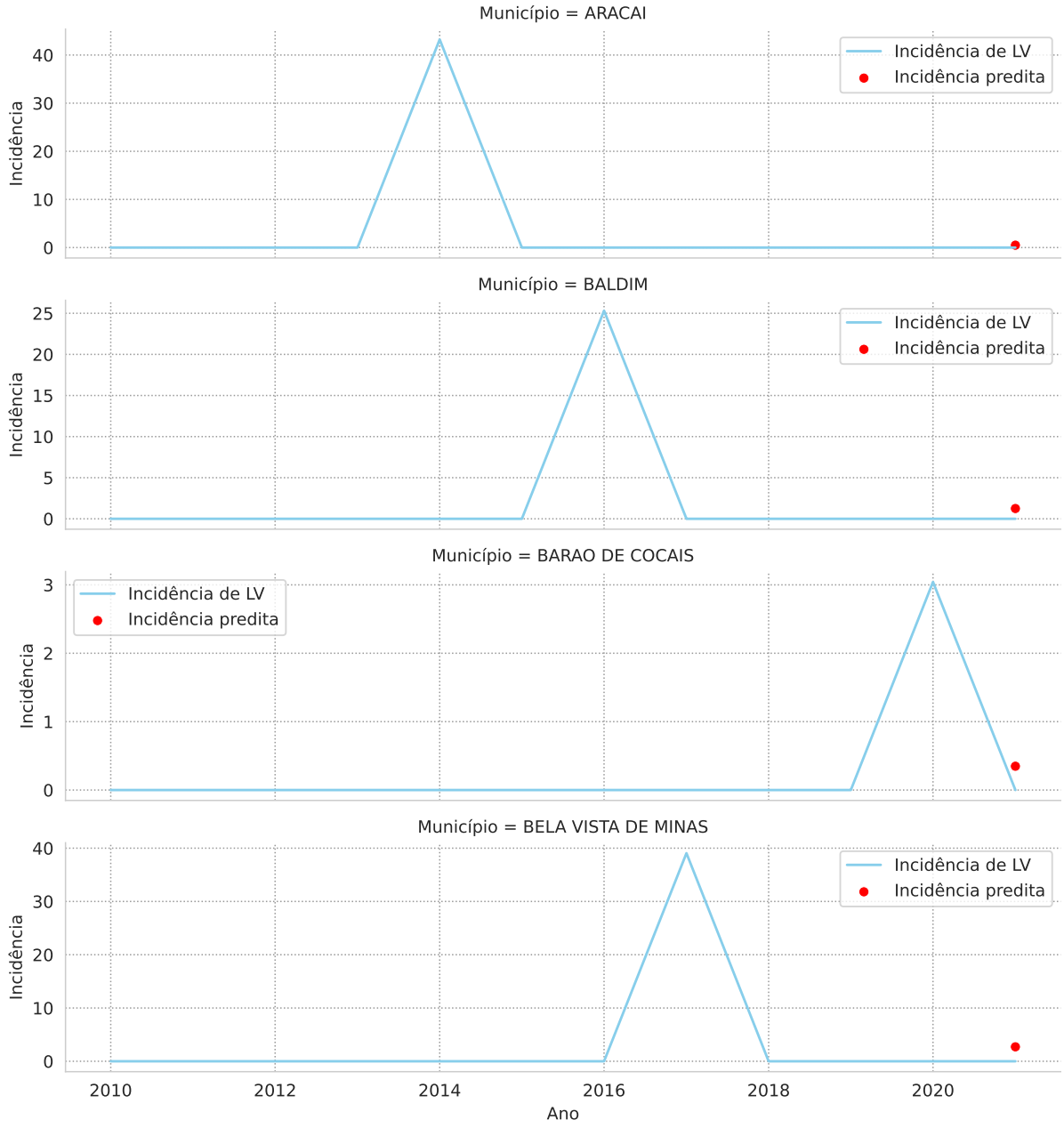
Municípios	
ABAETE	ARACAI*
AUGUSTO DE LIMA	BALDIM*
BARAO DE COCAIS*	BELA VISTA DE MINAS*
BELO HORIZONTE	BELO VALE
BETIM*	BOM JESUS DO AMPARO
BONFIM	BRUMADINHO
BUENOPOLIS	CACHOEIRA DA PRATA*
CAETANOPOLIS	CAETE
CAPIM BRANCO	CARMESIA
CATAS ALTAS	CONFINS
CONTAGEM	CORDISBURGO
CORINTO	CURVELO
DOM JOAQUIM	ESMERALDAS*
FELIXLANDIA*	FERROS
GUANHAES	IBIRITE
IGARAPE	INHAUMA
INIMUTABA	ITABIRA
ITABIRITO*	JABOTICATUBAS
JEQUITIBA	JOAO MONLEVADE
JUATUBA*	LAGOA SANTA
MARAVILHAS	MARIANA*

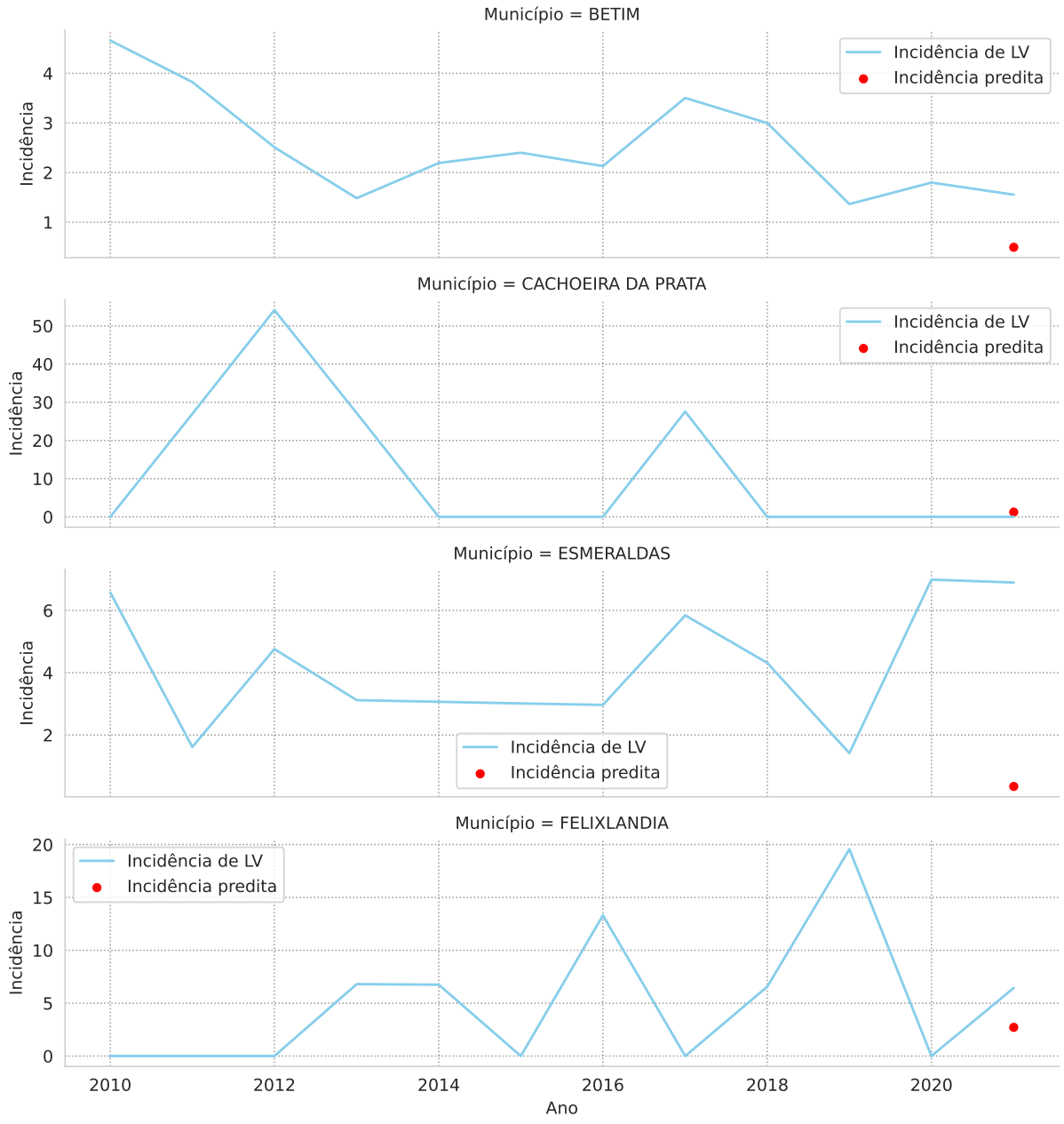
Tabela A.1 – *Municípios utilizados para o treino e teste dos modelos* (conclusão)

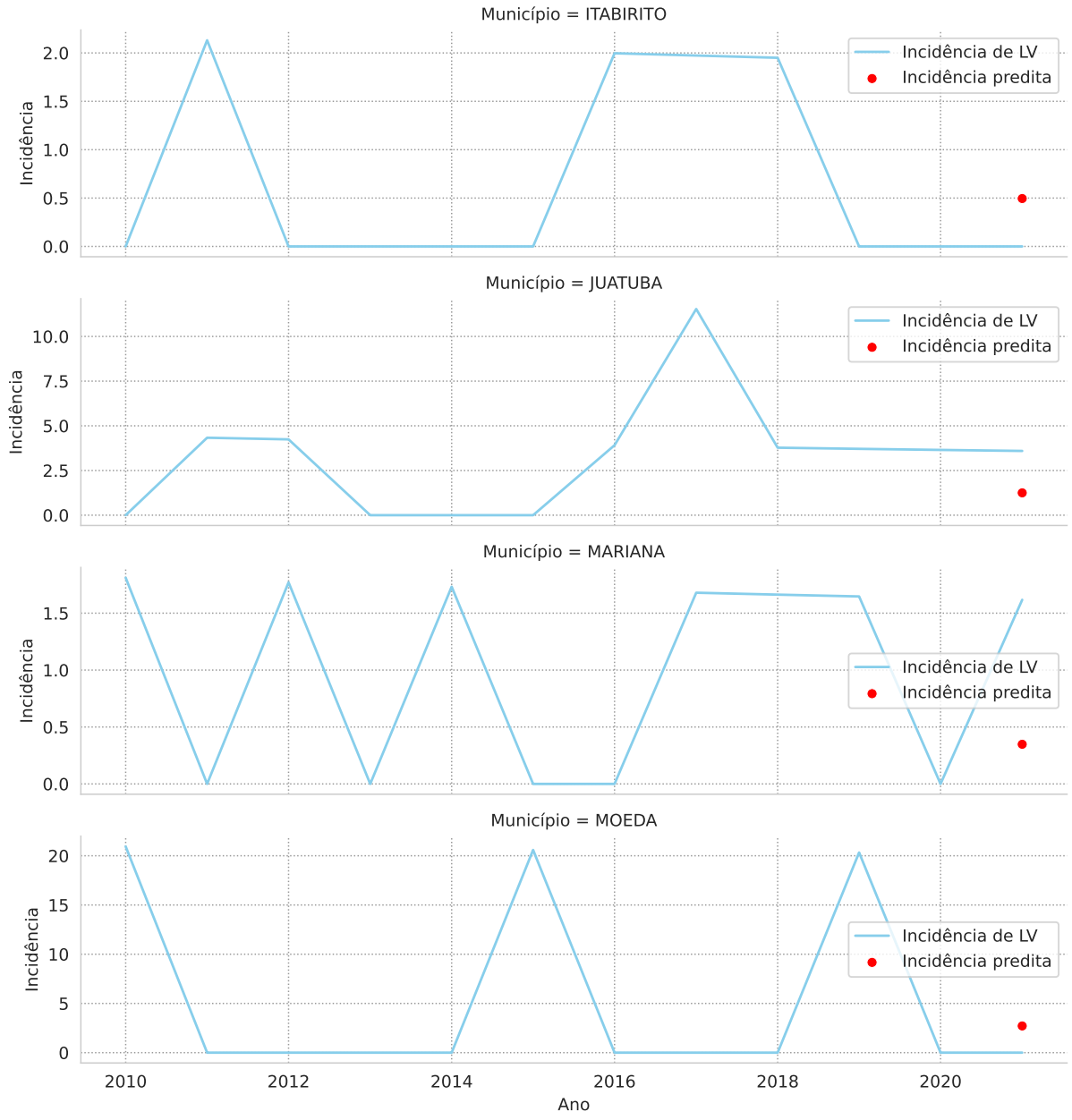
Municípios	
MARIO CAMPOS	MATEUS LEME
MATOZINHOS	MOEDA*
MONJOLOS*	MORRO DO PILAR
NOVA ERA	NOVA LIMA
NOVA UNIAO*	OURO PRETO
PAPAGAIOS	PARAOPEBA
PEDRO LEOPOLDO*	PEQUI*
POMPEU	PRESIDENTE JUSCELINO
PRUDENTE DE MORAIS	RAPOSOS
RIBEIRAO DAS NEVES	RIO ACIMA
RIO VERMELHO*	SABARA*
SANTA BARBARA	SANTA LUZIA
SANTA MARIA DE ITABIRA*	SANTANA DE PIRAPAMA
SANTANA DO RIACHO	SANTO HIPOLITO
SAO DOMINGOS DO PRATA*	SAO JOAQUIM DE BICAS
SAO JOSE DA LAPA	SARZEDO
SETE LAGOAS*	TAQUARACU DE MINAS
TRES MARIAS*	VESPASIANO*
VIRGINOPOLIS*	

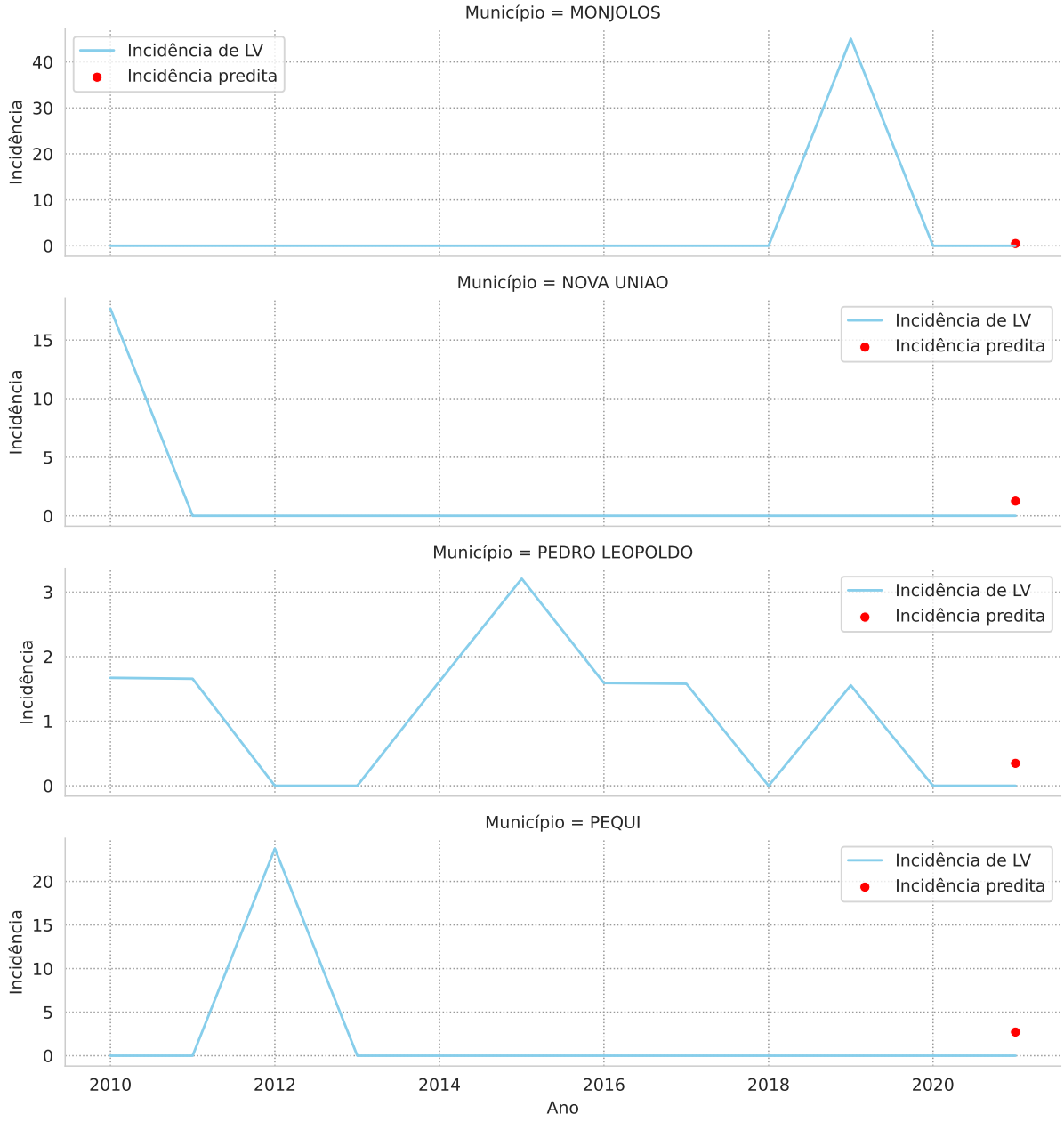
Municípios com asterisco (*) foram usados para teste dos modelos.

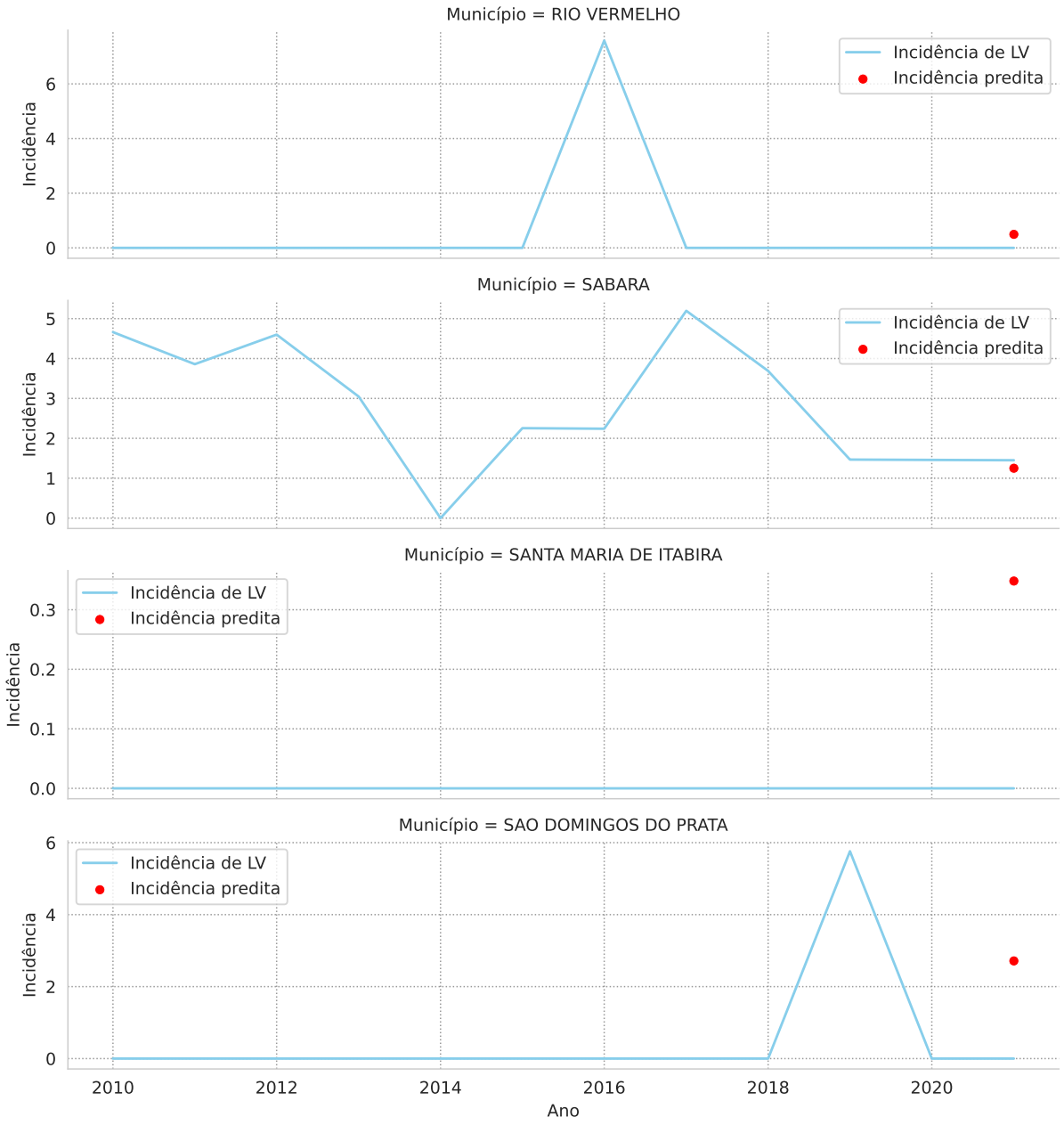
APÊNDICE B – Series temporais de cada município do conjunto de teste e sua incidência predita

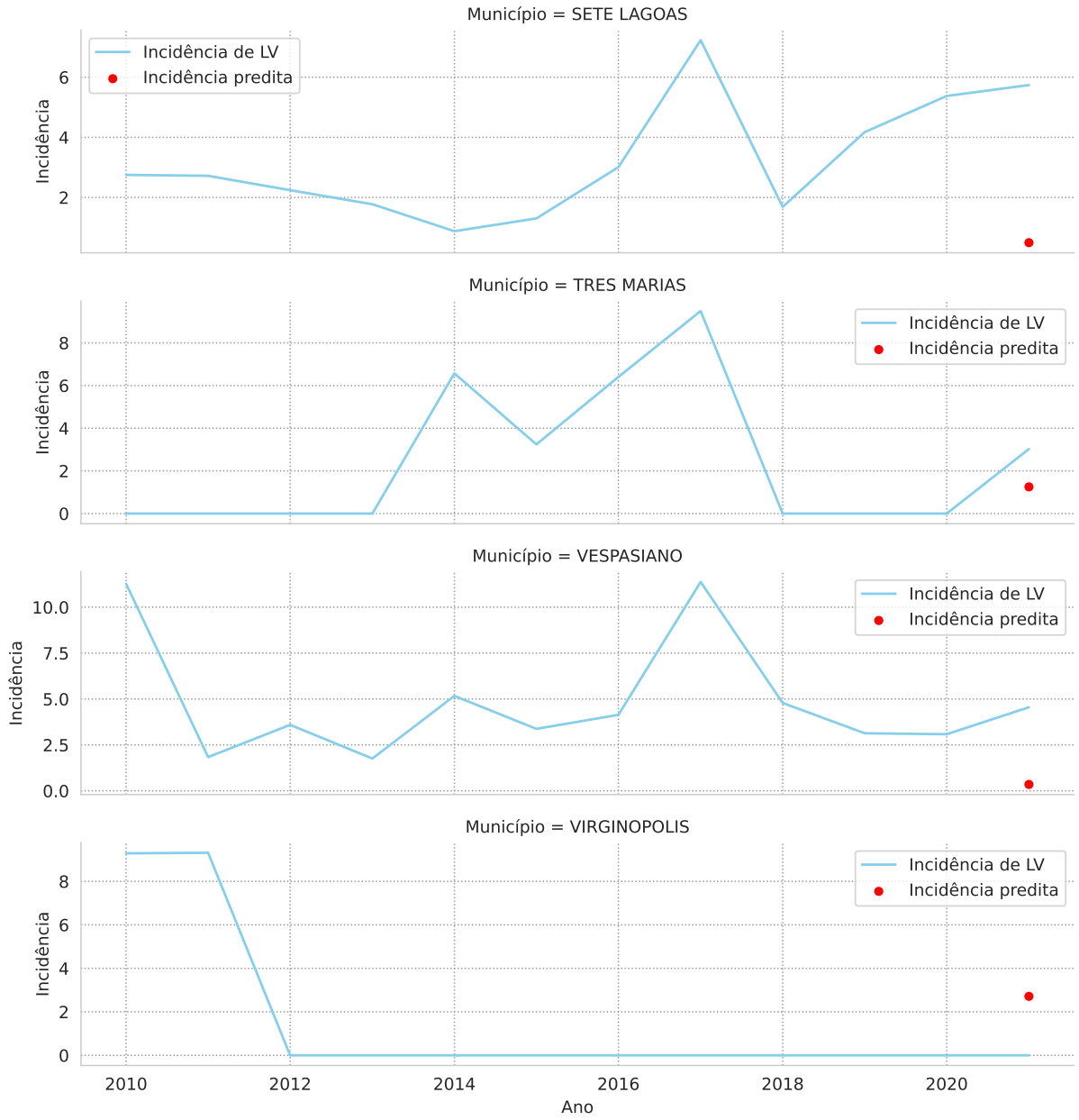



Município = BETIM
Incidência de LV
Incidência predita
Incidência
Município = CACHOEIRA DA PRATA
Município = ESMERALDAS
Município = FELIXLANDIA
2010
2012
2014
2016
2018
2020
Ano

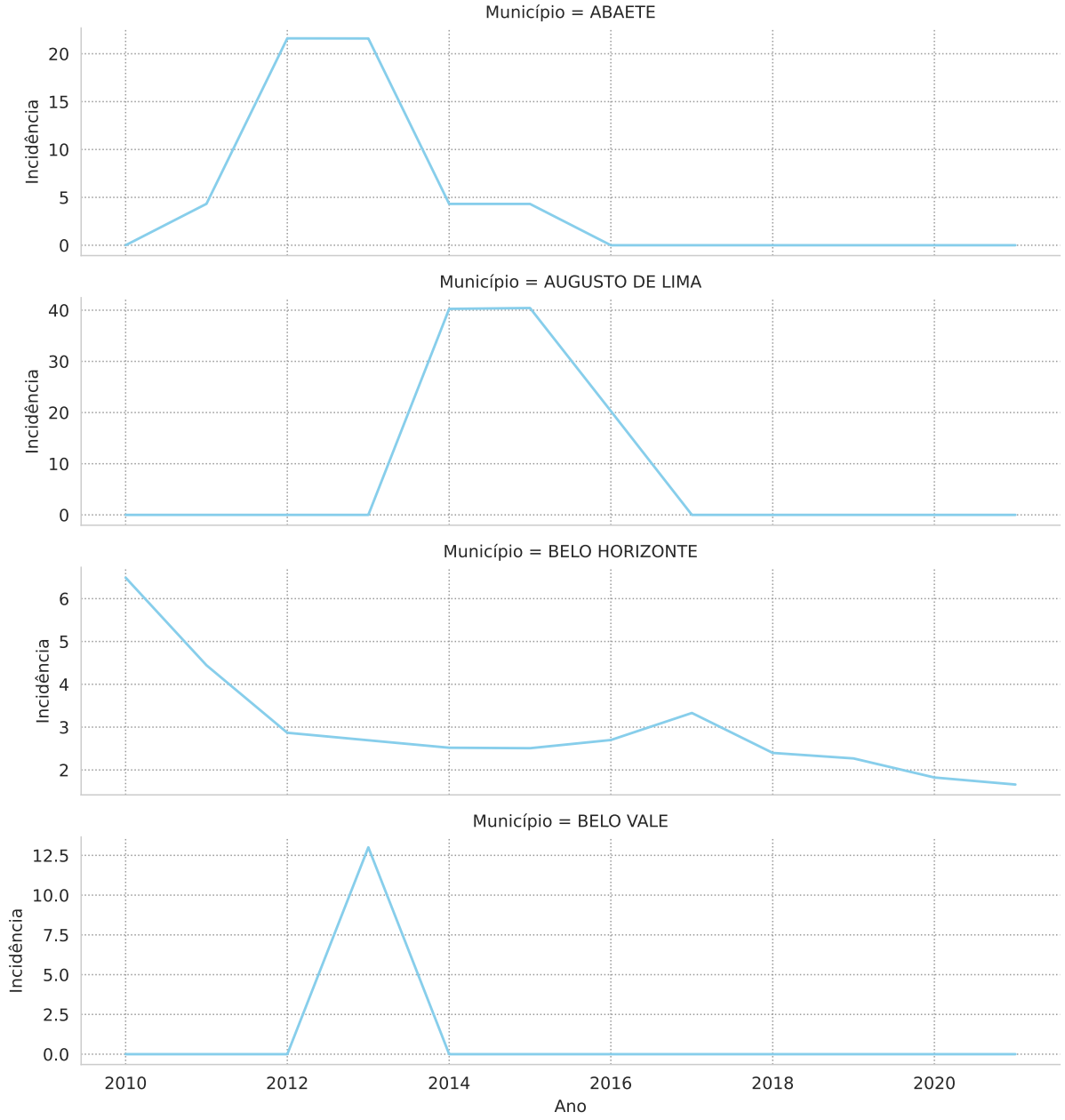

Município = ITABIRITO
Município = JUATUBA
Município = MARIANA
Município = MOEDA
Incidência de LV
Incidência predita
Incidência
Ano
2010
2012
2014
2016
2018
2020

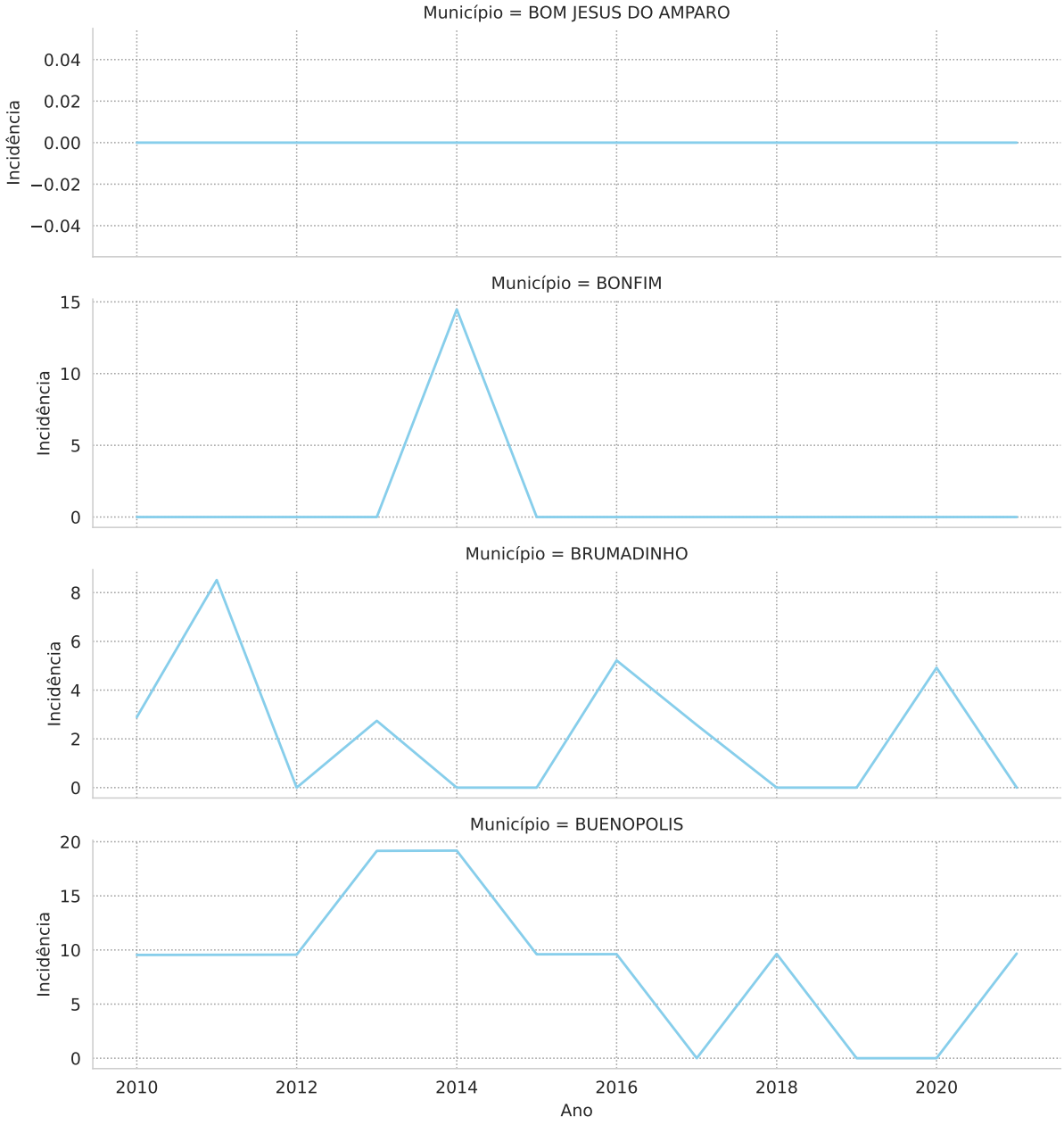

Município = MONJOLOS
Incidência de LV
Incidência predita
Incidência
0
10
20
30
40
Município = NOVA UNIAO
Incidência de LV
Incidência predita
Incidência
0
5
10
15
Município = PEDRO LEOPOLDO
Incidência de LV
Incidência predita
Incidência
0
1
2
3
Município = PEQUI
Incidência de LV
Incidência predita
Incidência
0
5
10
15
20
2010
2012
2014
2016
2018
2020
Ano

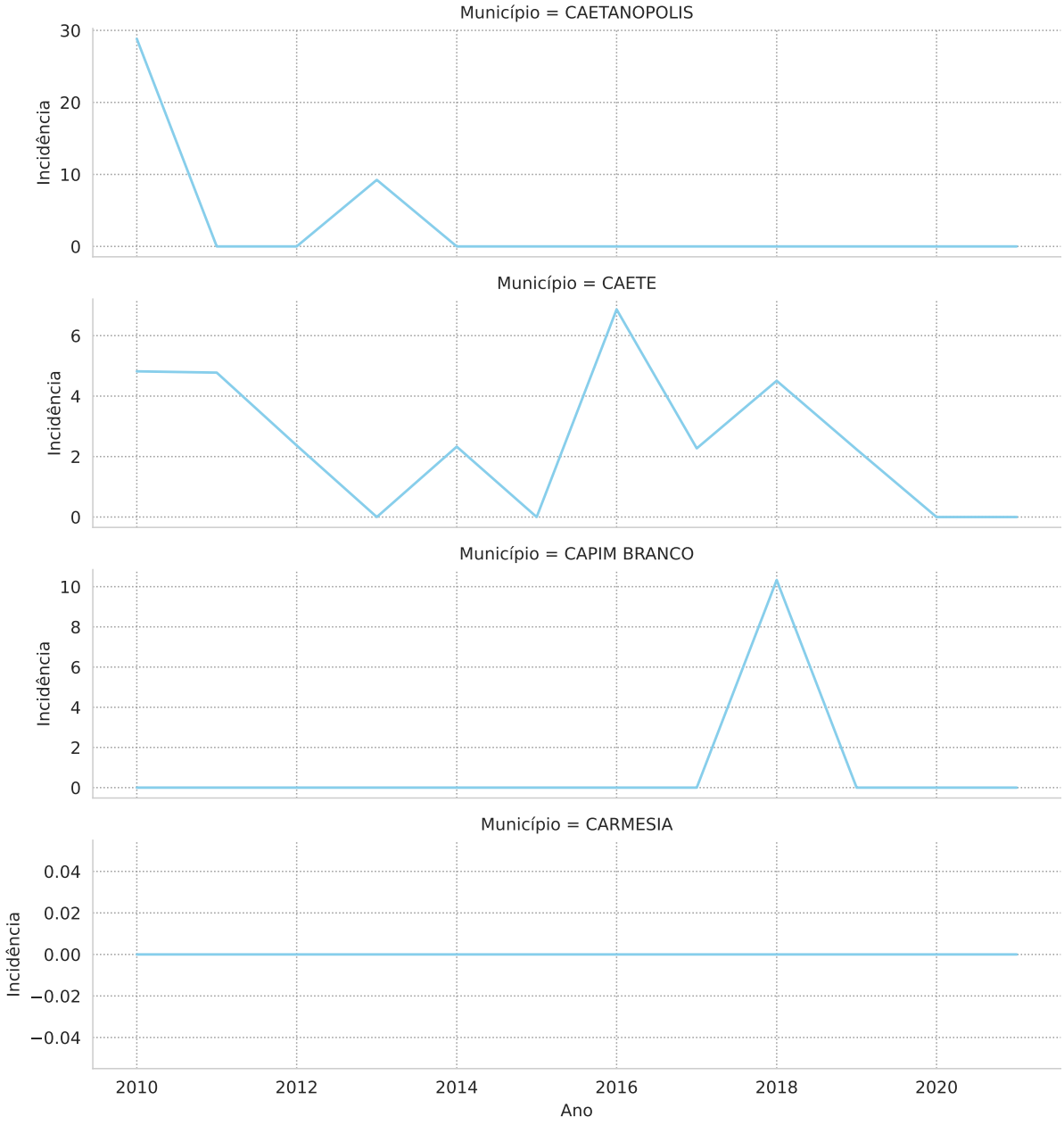

Município = RIO VERMELHO
Município = SABARA
Município = SANTA MARIA DE ITABIRA
Município = SAO DOMINGOS DO PRATA
Incidência de LV
Incidência predita
Incidência
Ano
2010
2012
2014
2016
2018
2020

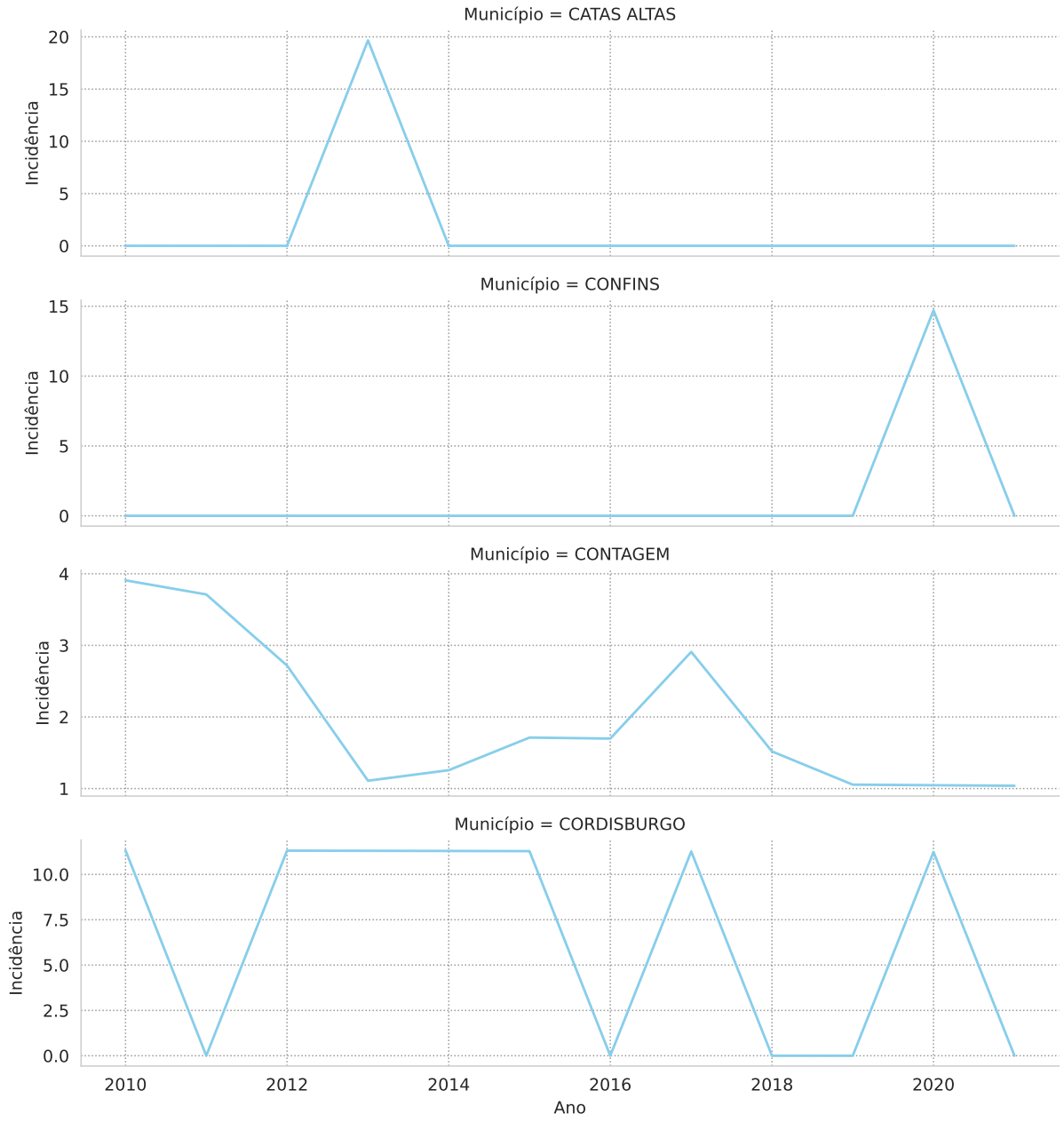

Município = SETE LAGOAS
Município = TRES MARIAS
Município = VESPASIANO
Município = VIRGINOPOLIS
Incidência de LV
Incidência predita
Incidência
Ano
2010
2012
2014
2016
2018
2020

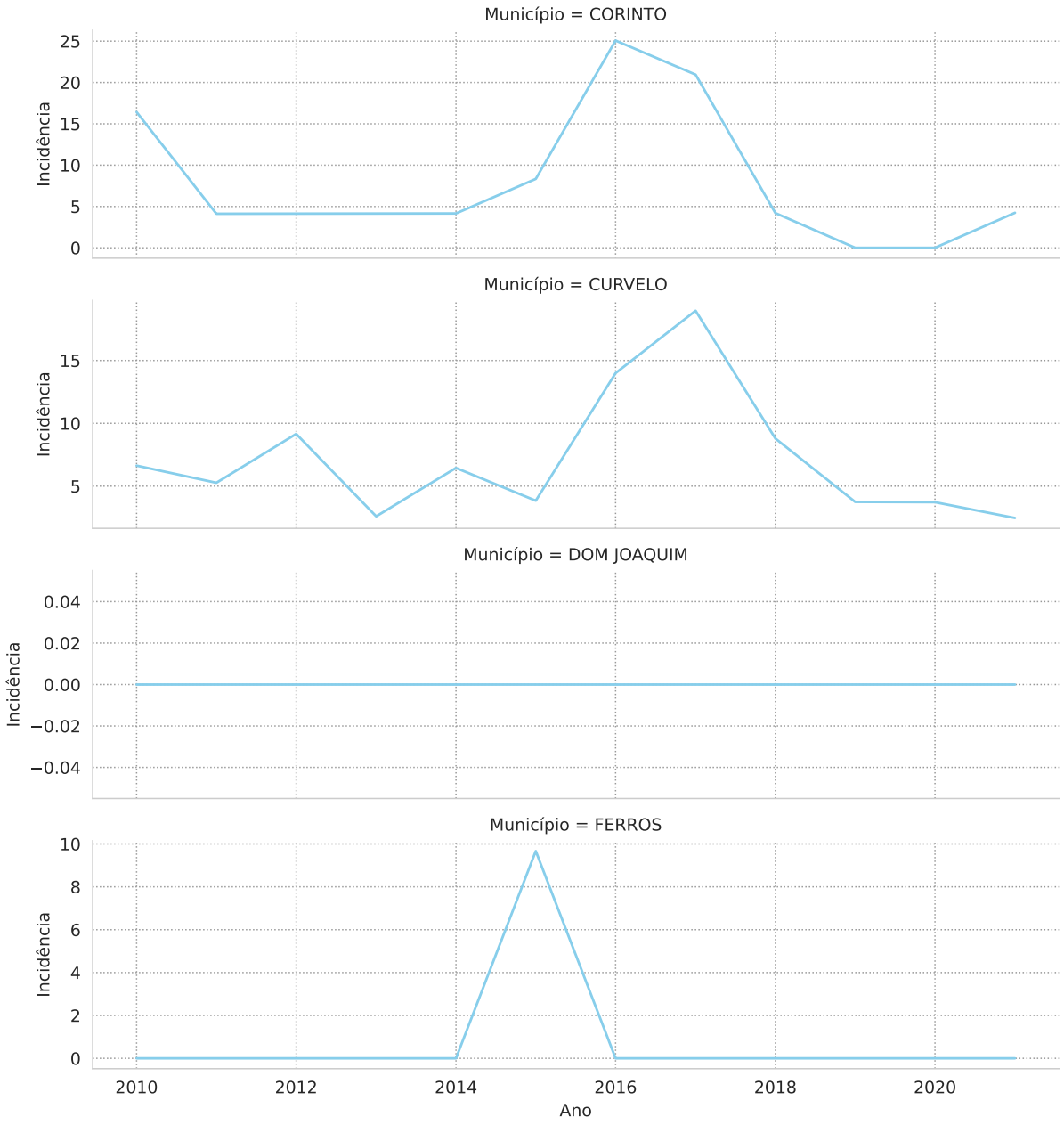
APÊNDICE C – Series temporais de cada município do conjunto de treinamento

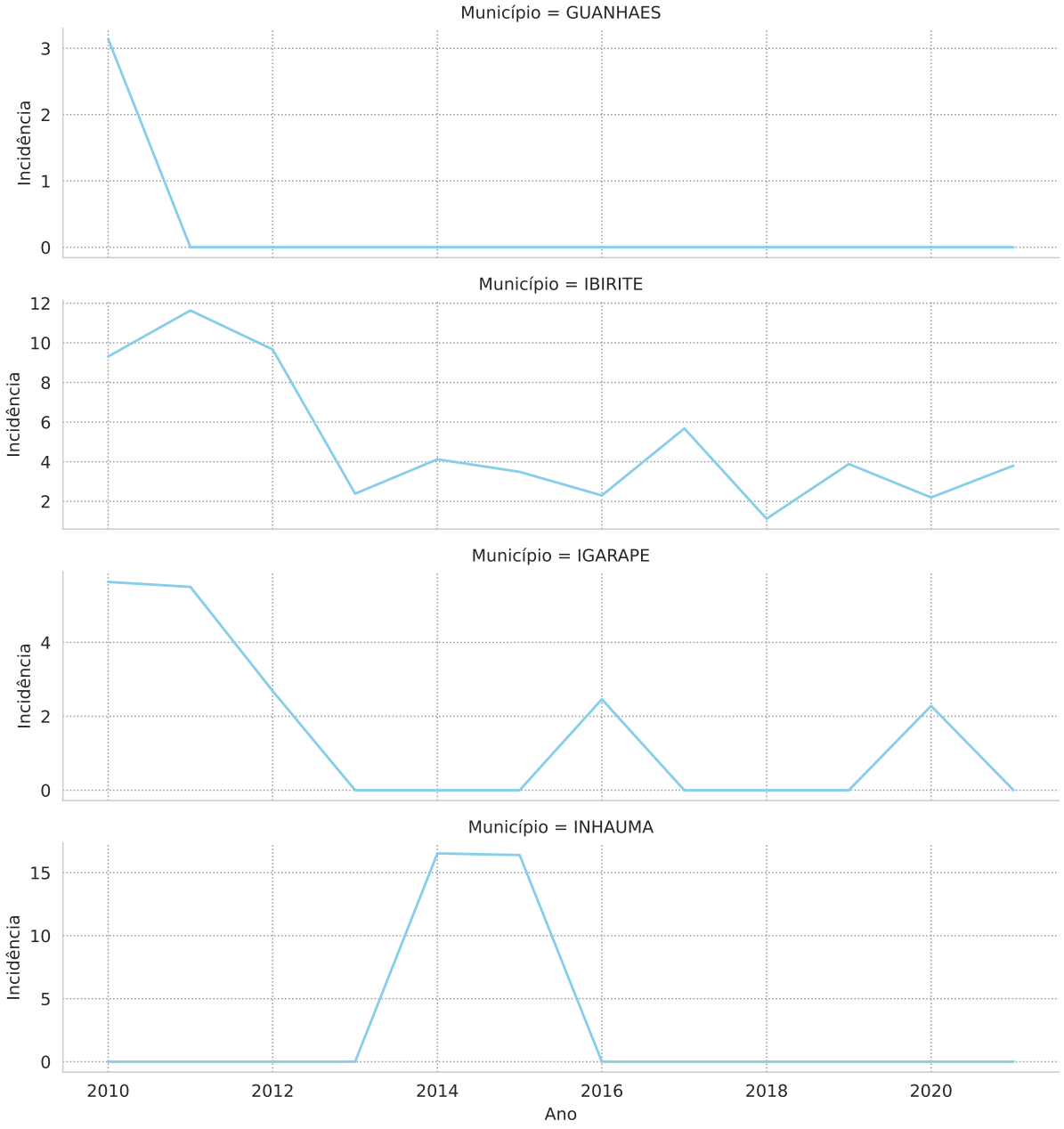


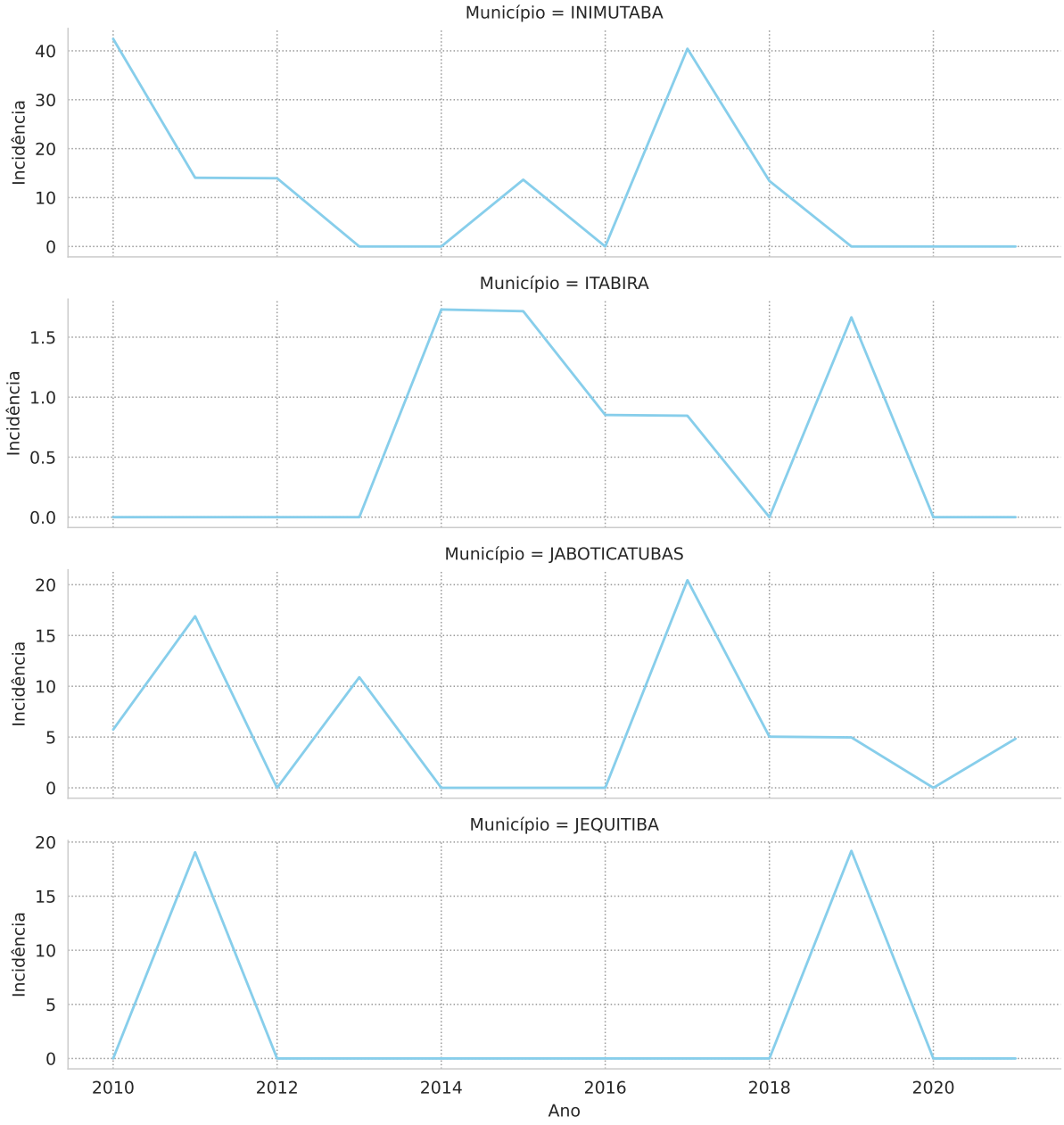

Município = BOM JESUS DO AMPARO
Município = BONFIM
Município = BRUMADINHO
Município = BUENOPOLIS
Incidência
Ano
2010
2012
2014
2016
2018
2020

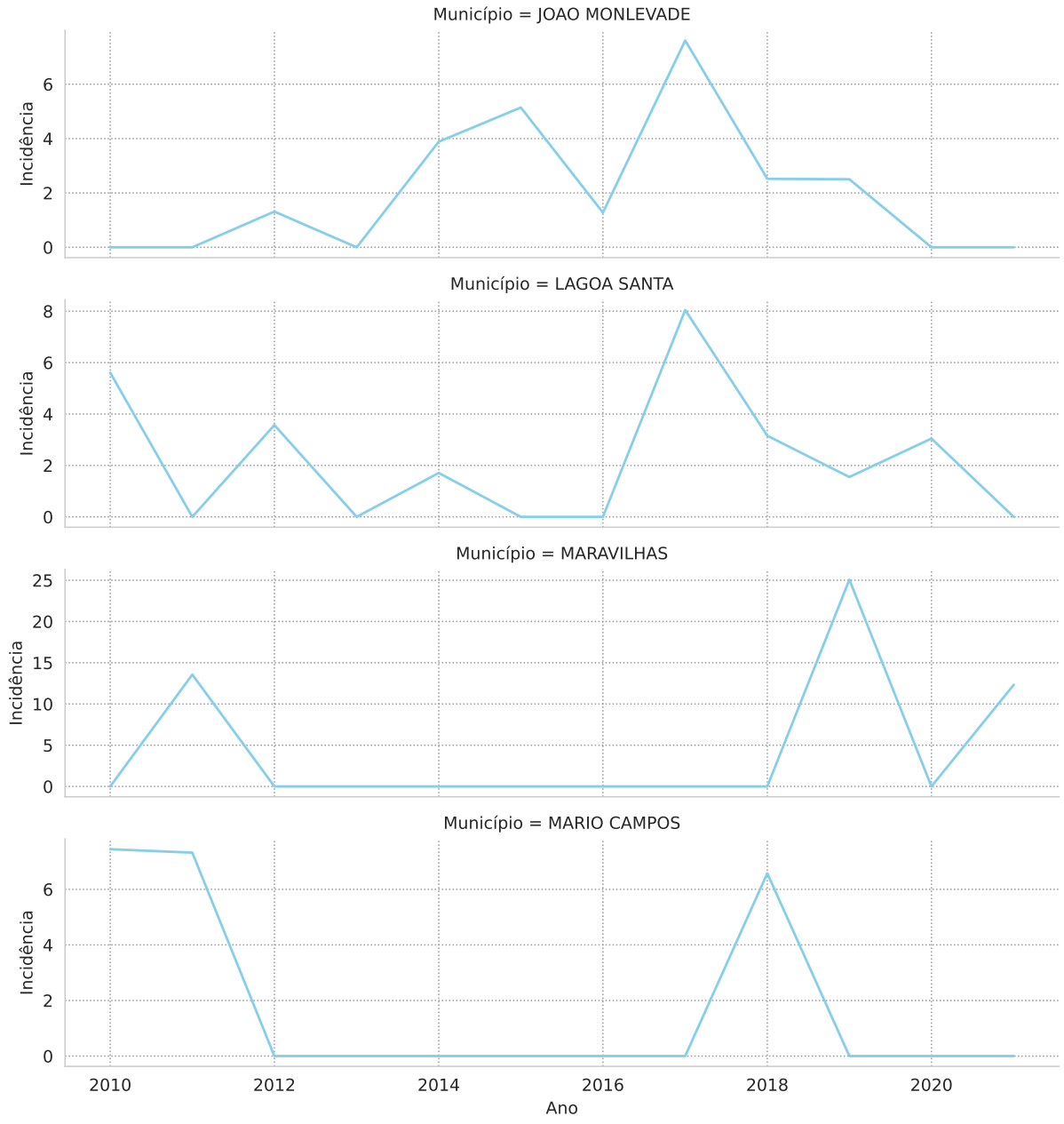

Município = CAETANOPOLIS
Incidência
30
20
10
0
Município = CAETE
Incidência
6
4
2
0
Município = CAPIM BRANCO
Incidência
10
8
6
4
2
0
Município = CARMESIA
Incidência
0.04
0.02
0.00
−0.02
−0.04
2010
2012
2014
2016
2018
2020
Ano

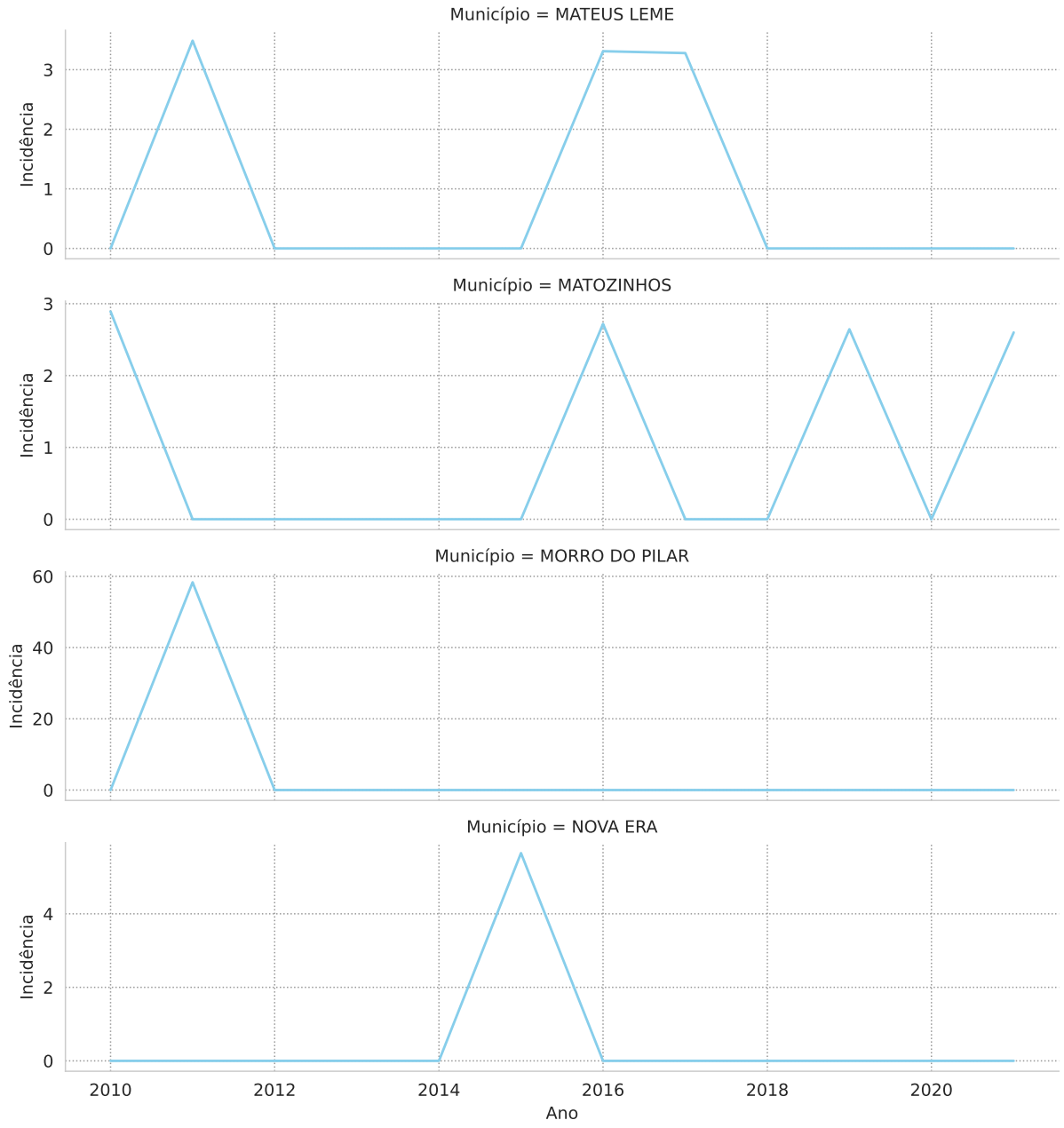

Município = CATAS ALTAS
Incidência
20
15
10
5
0
Município = CONFINS
Incidência
15
10
5
0
Município = CONTAGEM
Incidência
4
3
2
1
Município = CORDISBURGO
Incidência
10.0
7.5
5.0
2.5
0.0
2010
2012
2014
2016
2018
2020
Ano

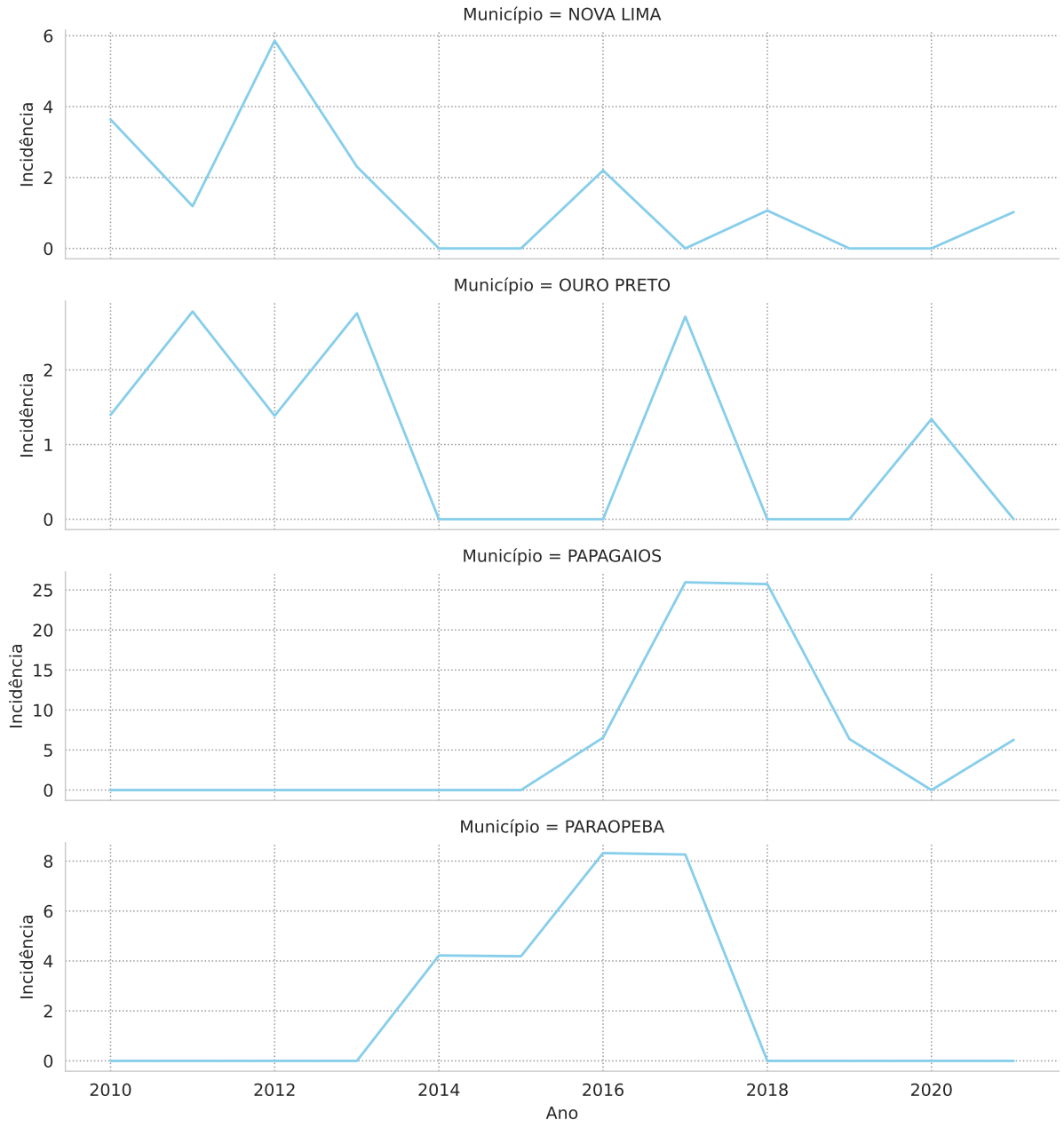

Município = CORINTO
Incidência
25
20
15
10
5
0
Município = CURVELO
Incidência
15
10
5
Município = DOM JOAQUIM
Incidência
0.04
0.02
0.00
−0.02
−0.04
Município = FERROS
Incidência
10
8
6
4
2
0
2010
2012
2014
2016
2018
2020
Ano

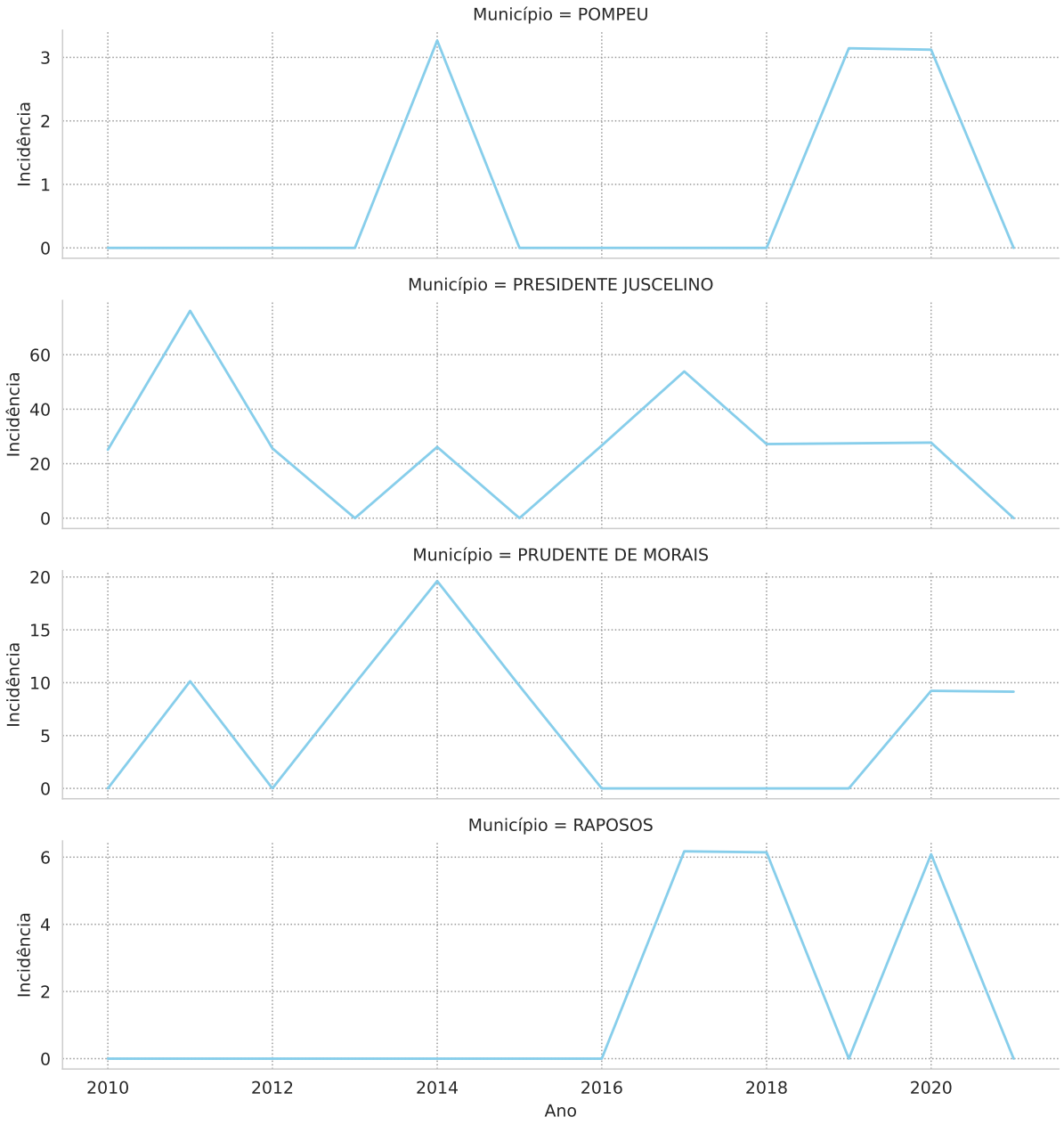

Município = GUANHAES
Incidência
0
1
2
3
Município = IBIRITE
Incidência
2
4
6
8
10
12
Município = IGARAPE
Incidência
0
2
4
Município = INHAUMA
Incidência
0
5
10
15
2010
2012
2014
2016
2018
2020
Ano

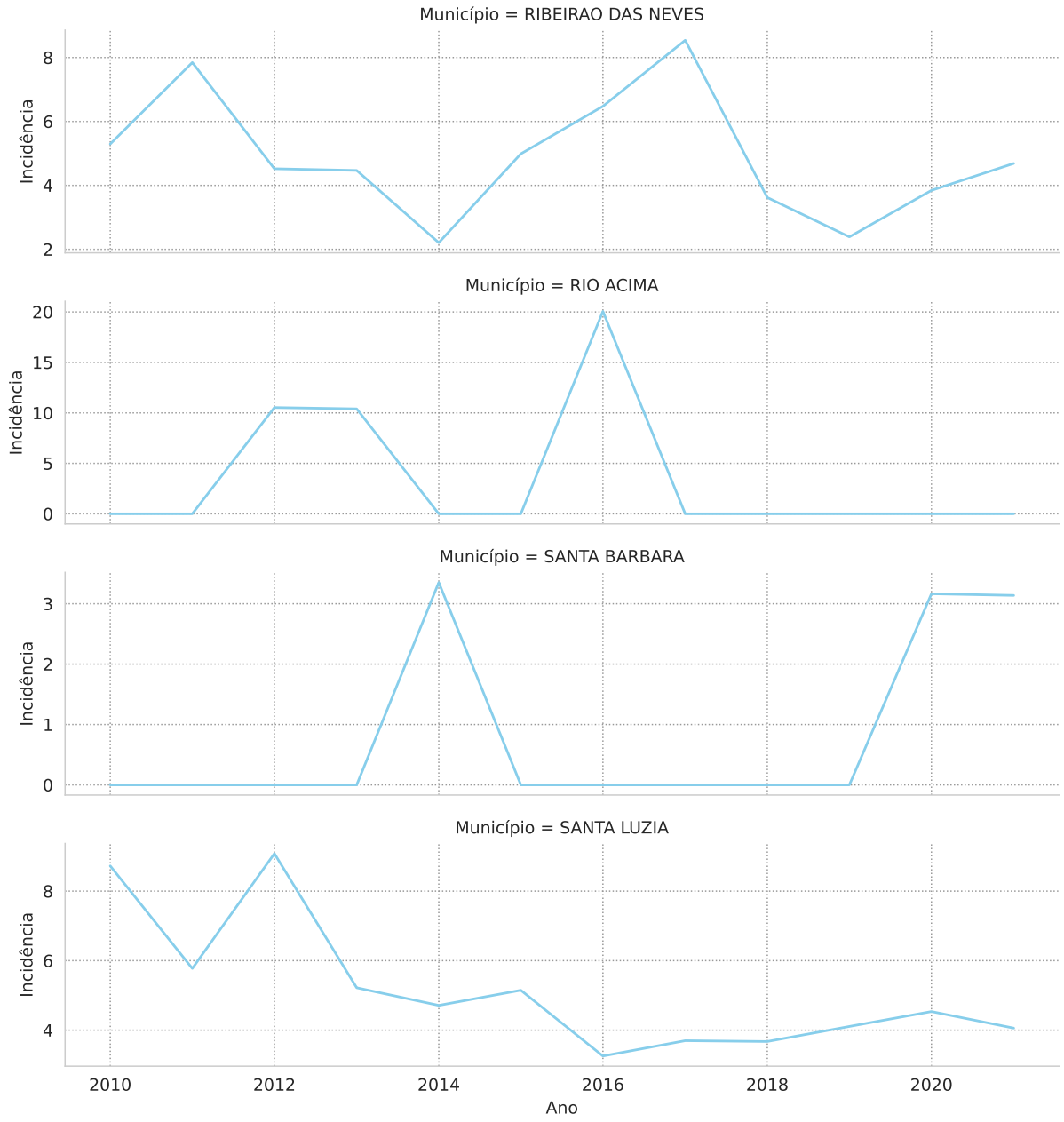

Município = INIMUTABA
Incidência
0
10
20
30
40
Município = ITABIRA
Incidência
0.0
0.5
1.0
1.5
Município = JABOTICATUBAS
Incidência
0
5
10
15
20
Município = JEQUITIBA
Incidência
0
5
10
15
20
2010
2012
2014
2016
2018
2020
Ano

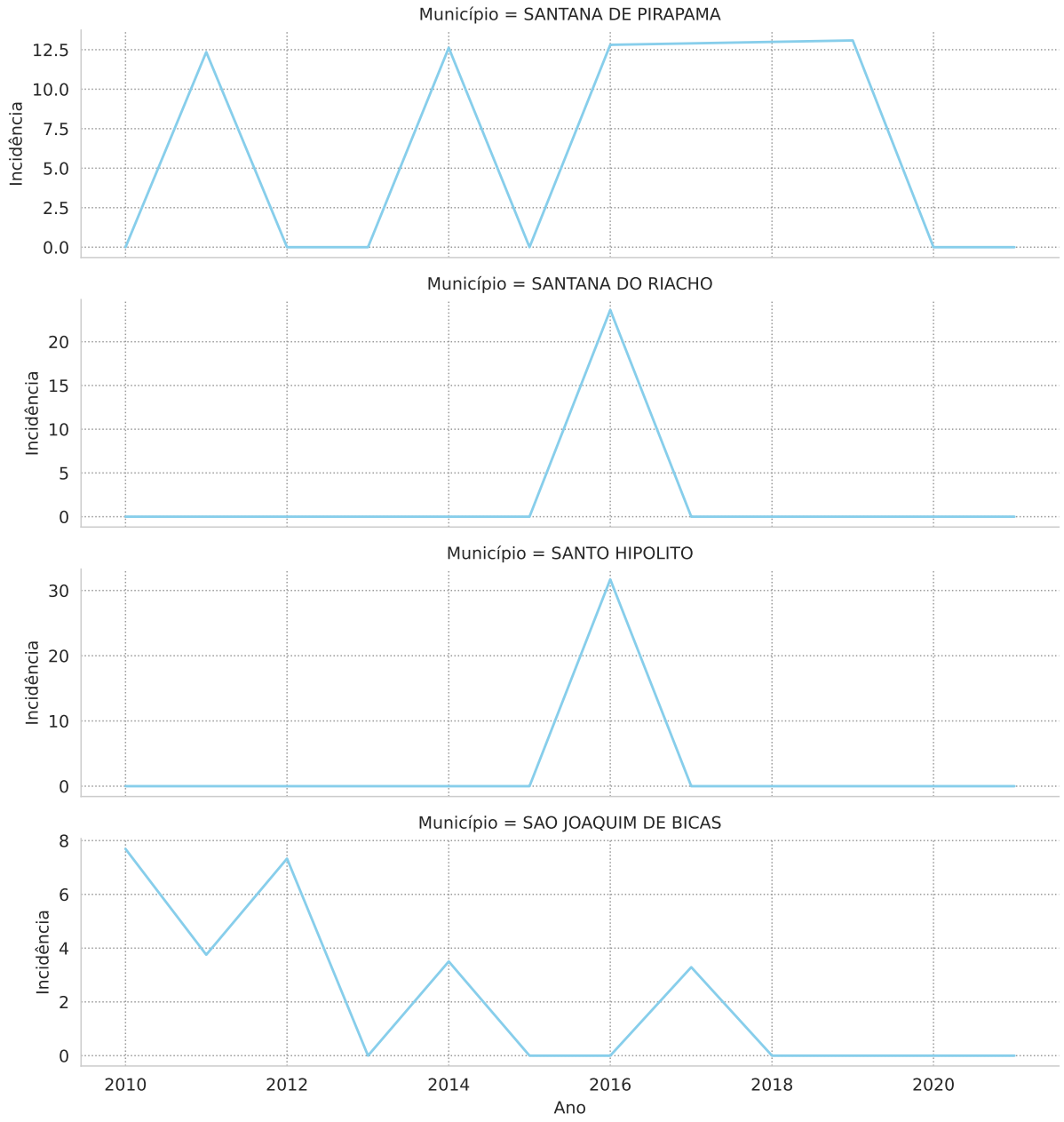

Município = JOAO MONLEVADE
Incidência
0
2
4
6
Município = LAGOA SANTA
Incidência
0
2
4
6
8
Município = MARAVILHAS
Incidência
0
5
10
15
20
25
Município = MARIO CAMPOS
Incidência
0
2
4
6
2010
2012
2014
2016
2018
2020
Ano

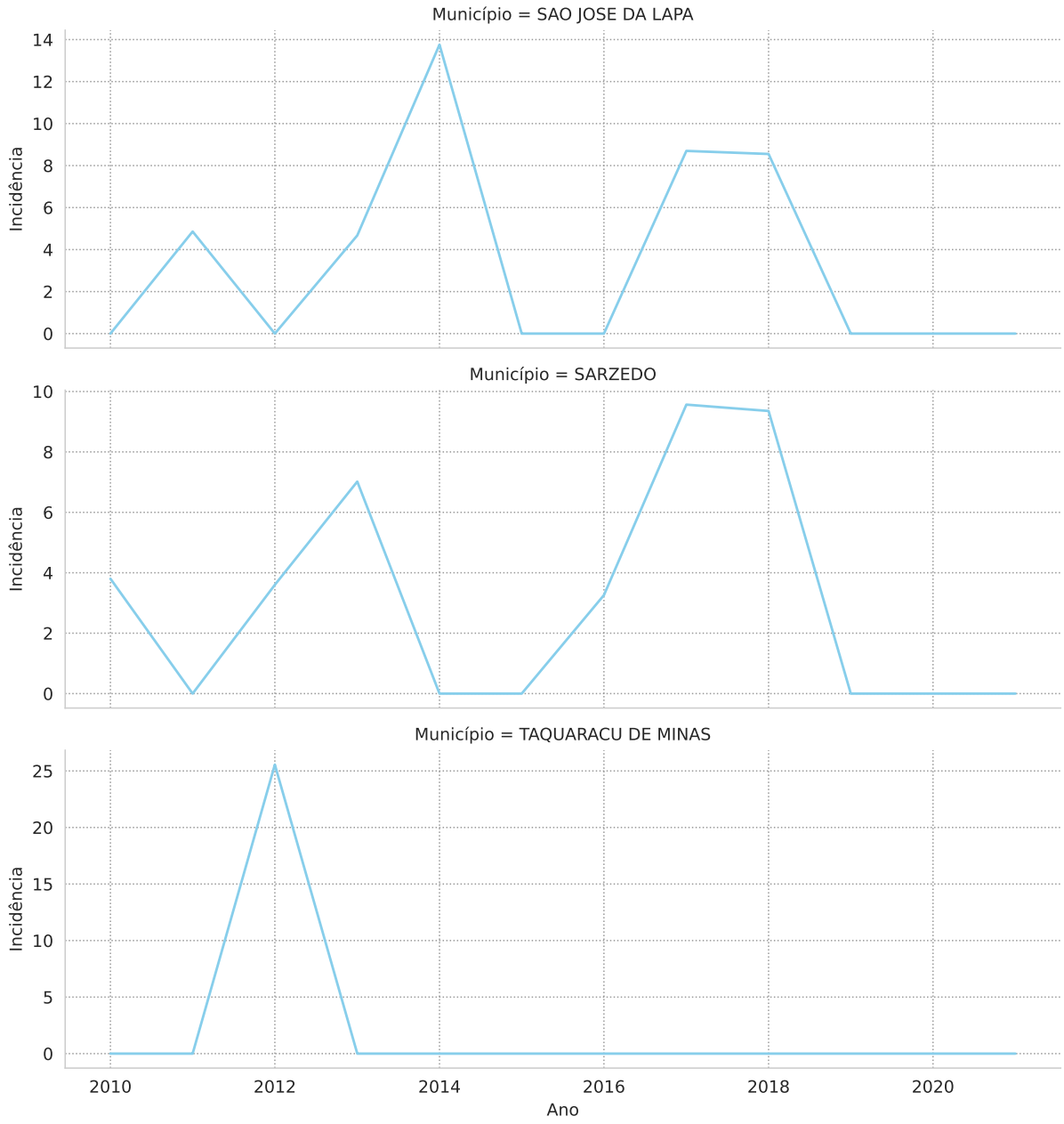

Município = MATEUS LEME
Município = MATOZINHOS
Município = MORRO DO PILAR
Município = NOVA ERA
Incidência
Ano
2010
2012
2014
2016
2018
2020


Município = NOVA LIMA
Incidência
0
2
4
6
Município = OURO PRETO
Incidência
0
1
2
Município = PAPAGAIOS
Incidência
0
5
10
15
20
25
Município = PARAOPEBA
Incidência
0
2
4
6
8
2010
2012
2014
2016
2018
2020
Ano


Município = POMPEU
Incidência
0
1
2
3
Município = PRESIDENTE JUSCELINO
Incidência
0
20
40
60
Município = PRUDENTE DE MORAIS
Incidência
0
5
10
15
20
Município = RAPOSOS
Incidência
0
2
4
6
2010
2012
2014
2016
2018
2020
Ano


Município = RIBEIRAO DAS NEVES
Incidência
8
6
4
2
Município = RIO ACIMA
Incidência
20
15
10
5
0
Município = SANTA BARBARA
Incidência
3
2
1
0
Município = SANTA LUZIA
Incidência
8
6
4
2010
2012
2014
2016
2018
2020
Ano


Município = SANTANA DE PIRAPAMA
Incidência
12.5
10.0
7.5
5.0
2.5
0.0
Município = SANTANA DO RIACHO
Incidência
20
15
10
5
0
Município = SANTO HIPOLITO
Incidência
30
20
10
0
Município = SAO JOAQUIM DE BICAS
Incidência
8
6
4
2
0
2010
2012
2014
2016
2018
2020
Ano


Município = SAO JOSE DA LAPA
Incidência
0
2
4
6
8
10
12
14
Município = SARZEDO
Incidência
0
2
4
6
8
10
Município = TAQUARACU DE MINAS
Incidência
0
5
10
15
20
25
2010
2012
2014
2016
2018
2020
Ano

Índice

Algoritmo de retropropagação, 10
Aprendizado de máquina, 5

Fonte de dados, 15

Leishmaniose Visceral, 4
Long Short Term Memory, 11

Métricas para avaliação, 19

Parametrização do Modelo LSTM, 17

Redes Neurais Artificiais, 7
Redes Neurais Recorrentes, 10

Treinamento de uma RNN, 9

XGBoost, 6, 19