



Universidade Federal do Paraná  
Setor de Ciências Exatas  
Departamento de Estatística  
Programa de Especialização em Data Science e Big  
Data

Karoline de Araujo Rodrigues

# **Previsão de Incêndios via Modelos de Vetores de Suporte: uma aplicação para Porto Velho - RO**

**Curitiba  
2022**

Karoline de Araujo Rodrigues

**Previsão de Incêndios via Modelos de Vetores de Suporte:  
uma aplicação para Porto Velho - RO**

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Anderson Ara

**Curitiba  
2022**

# Previsão de Incêndios via Modelos de Vetores de Suporte: uma aplicação para Porto Velho - RO

Karoline de Araujo Rodrigues<sup>1</sup>, Anderson Ara<sup>2</sup>

<sup>1</sup>Departamento de Estatística, Universidade Federal do Paraná Rua Evaristo F. F. da Costa 418, Jardim das Americas, 82590-300, Curitiba, PR, Brasil \*

<sup>2</sup>Departamento de Estatística, Universidade Federal do Paraná Rua Evaristo F. F. da Costa 418, Jardim das Americas, 82590-300, Curitiba, PR, Brasil †

Desde 2020 o Brasil tem registrado recordes em incêndios florestais, a Região Norte possui 80% de seu território formado pela Floresta Amazônica e é uma das mais afetadas por esses incêndios. Os impactos que temos devido a esses incêndios, é visto desde o aumento de casos de doenças respiratórias causadas pela fumaça até a alteração de ciclos de chuvas, visto que a Amazônia ajuda na regulação das chuvas que impactam todo o planeta. Por isso, usar mecanismos como aprendizado de máquina que permitam a previsão de novos focos de incêndio são cada vez mais necessários para auxiliar os órgãos responsáveis a atuarem nesses locais de forma preventiva. Neste estudo, optamos por focar na previsão de incêndios para o município de Porto Velho, capital de Rondônia. O objetivo é aplicar o modelo de Vetores de Suporte para regressão em séries temporais usando o método SVR-Mediana, assim como a comparação com o uso de Vetores de Suporte para Regressão sem nenhuma modificação. Através do uso do SVR-Mediana foi possível obter medidas de performance preditiva mais satisfatórias, com o MASE de 0,11 e um RMSSE de 0,15.

**Palavras-chave:** previsão de incêndios, SVR, SVR mediana, series temporais

Since 2020 Brazil has recorded records of forest fires, the North Region, which has 80% of its territory formed by the Amazon Forest, is one of the most affected by these fires. The impacts we have due to these fires are seen from the increase in cases of respiratory diseases caused by smoke from fire to the change in rainfall cycles, as the Amazon helps in regulating the rains that impact everyone. Therefore, using mechanisms such as machine learning that allow the prediction of new fire outbreaks is increasingly necessary to support the responsible agencies to act in these places in a preventive way. In this study, we chose to focus on the prediction of fires for the municipality of Porto Velho, the capital of Rondônia, and the objective will be to apply the Support Vector model for regression in time series using the SVR-Median method, as well as the comparison with the use of Support Vectors for Regression without any modification. Using the SVR-Median, it was possible to obtain more satisfactory predictive performance measures, with a MASE of 0.11 and an RMSSE of 0.15.

**Keywords:** fire forecasting, Support Vector Regression, Median SVR, time series

## 1. Introdução

No Brasil, é uma prática comum o uso de fogo em atividades agropecuárias, para fazer pastagem, eliminação de pragas e preparo do solo, essa prática se controlada não acarreta danos (Coutinho, 2007)[7], mas se utilizada de forma incorreta e sem controle podem se tornar incêndios florestais, esses incêndios muitas vezes são difíceis de controlar e acabam invadindo áreas protegidas. Além disso, também acarretam problemas para a população aumentando problemas respirató-

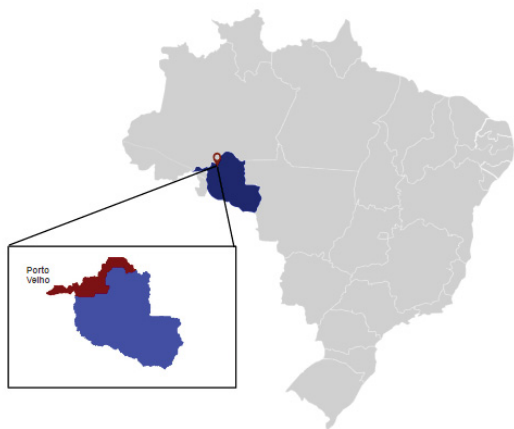
rios devido a quantidade de fumaça. Essas queimadas ocorrem com mais frequência no inverno, entre julho e outubro, durante esses meses a vegetação fica mais suscetível ao fogo por conta do clima mais seco, além de outros fatores meteorológicos como a baixa umidade, sendo a atividade humana a maior responsável pelas queimadas (Anderson et al., 2005) [1]

Durante a análise exploratória dos dados, Porto Velho se destacava como a segunda cidade com maior focos de incêndio em 2021, com 7,7% do total de incêndios do país. A capital de Rondônia possui uma população estimada para 2021 de 548.952 pessoas de acordo com site do IBGE[13] e em 2020 foi o 4º mu-

\*kr.araujo.12@gmail.com

†ara@ufpr.br

nicípio (IBGE, 2020) [14] com maior quantidade de cabeças de gado do País (1,3 milhão de cabeças). O desmatamento na região também tem crescido, em 2021 cerca de 32% da sua área total já havia sido desmatada, sendo a cidade com maior área desmatada do estado de Rondônia (PRODES, 2021) [15].



**Figura 1:** Mapa Brasil com Destaque na cidade de Porto Velho. Fonte: Elaborado pela autora Adaptado de IBGE

Esse é o resultado em grande parte se dá ao aumento das atividades agropecuárias e plantio de soja na região que tem crescido desde 2015. E caso não existam mudanças nos próximos anos relacionadas a diminuição do desmatamento, em 2050 a expansão agrícola terá eliminado 40% da floresta Amazônica (Soares-Filho et al., 2006) [20], por isso usar mecanismos que possam auxiliar na previsão desses incêndios são essenciais.

Um dos mecanismos que pode ser utilizado na previsão de novos focos de incêndio, é a aplicação de modelos de previsão para séries temporais em que a variável  $Y$  é dependente do tempo. Nesse caso, não temos uma variável contínua, mas sim uma contagem ao longo do tempo que será utilizada em conjunto com outras variáveis meteorológicas para previsão de novos focos.

Assim, para ter uma boa performance na previsão de séries temporais, é necessário uma série estacionária, nesse artigo iremos explorar o método sugerido por Hung, Hung e Lin (2014) [10] que utiliza a método tradicional de médias móveis (MA) em conjunto com o Vetores de Suporte para regressão aplicado para previsões de séries temporais, além disso também foi utilizado outros métodos tradicionais como a diferenciação para obter uma série estacionária. Sendo assim, o objetivo desse estudo é apresentar várias comparações entre métodos para séries temporais para prever o número de novos focos de incêndio na cidade de Porto Velho-RO.

## 2. Conjunto de dados

A base de dados com os focos de incêndio foi retirada do Bdequeimadas (banco de dados de queimadas do Instituto Nacional de Pesquisas Espaciais) [3]. Os período de extração dos dados é de 2007 até abril de 2022 e contém os dados de todos os municípios, onde cada linha representa um foco de incêndio naquela latitude e longitude junto com as variáveis meteorológicas daquela localidade, como os dias sem chuva, precipitação e o Risco de Fogo que indica o quanto a vegetação está propícia a pegar fogo. Como os dados eram para cada latitude e longitude e o objetivo é fazer a previsão por município, criamos uma nova coluna com a contagem, mantendo a data e agrupando os dados por dia e município.

Para os dados meteorológicos por município, utilizamos os dados do Bdmp (Banco de dados meteorológicos do Instituto Nacional de meteorologia) [2], esses dados por dia foram adicionados na tabela com focos de incêndio. A estrutura da base de dados e as variáveis disponíveis são apresentadas na Tabela 1.

Variável	Descrição
<b>Quantitativas:</b>	
focos	Quantidade de focos naquela localidade.
PRECIPITACAO_TOTAL	Total de chuvas em mm nas últimas 24h.
PRESSAO_ATMOSFERICA_MEDIA	Média diária da força exercida por uma coluna de ar sobre a superfície.
TEMP_DO_PONTO_ORVALHO	Média diária da temperatura na qual o vapor de água começa a se condensar.
TEMP_MAXIMA	Medida máxima da temperatura do ar ocorrida no período de 24h.
TEMP_MEDIA_COMPENSADA	Média das três leituras mais a máxima e a mínima.
TEMP_MINIMA	Medida mínima da temperatura do ar no período de 24h.
UMIDADE_RELATIVA_MEDIA	Média da quantidade de vapor de água contida na atmosfera.
UMIDADE_RELATIVA_MINIMA	Medida mínima da umidade relativa do ar nas últimas 24h.
VENTO_RAJADA_MAXIMA	Maior velocidade diária do vento.
VENTO_VELOCIDADE_MEDIA	Quantificação média do dia do movimento do ar numa unidade de tempo.
<b>Qualitativas:</b>	
Data	Dados diários entre 2007-2022.

**Tabela 1:** Descrição das variáveis que compõem a base de dados do Bdequeimadas e Bdmp.

Das variáveis meteorológicas, Rajada máxima de vento e velocidade média do vento não serão consideradas nesse artigo, pois possuem observações somente até 2019.

A base total possui 5379 linhas, em que 67% dos focos são zeros e também uma sazonalidade entre os meses de julho e outubro fazendo com que os dados possuam uma grande variância. Os dados possuem o comportamento de uma série intermitente, em que temos muitos zeros e valores espalhados pela série de forma contínua (Silver, 1981) [17].

### 2.1. Recursos Computacionais

O *software* Python, versão 3.9.6 [16], foi utilizado para realizar o tratamento dos dados e criação das funções

utilizadas para criar o modelo, além de bibliotecas auxiliares como Sklearn[18] para o modelo de SVR e função de *grid* para os hiper-parâmetros. O ambiente computacional utilizado neste estudo foi um Windows 10 com 8GB de RAM.

### 3. Metodologia

#### 3.1. Máquinas de vetores de suporte

A metodologia do Maquinas de Vetores de Suporte (SVM) foi criada por Cortes e Vapnik em 1995 com o objetivo de resolver problemas de classificação. Como o principal problema em uma classificação binária é a separação de dois grupos, uma maneira de fazer essa separação é construindo uma linha, um plano ou hiperplano (dependendo da dimensionalidade dos dados) que faça essa separação em dois grupos da melhor forma possível. Essa linha, plano ou hiperplano é chamada de superfície de decisão linear. A ideia por trás dessa superfície de decisão é maximizar a distância entre os vetores de suporte tendo assim um hiperplano ótimo (Cortes, Vapnik 1995)[4].

#### 3.2. Vetores de suporte para regressão

Vetores de Suporte para regressão (SVR) é um modelo utilizado quando temos uma única variável de resposta  $y \in \mathbb{R}$  e os valores de  $x$  são as variáveis explicativas. Enquanto no SVM buscamos um hiperplano que melhor separa os dados, no SVR o objetivo é analisar os dados dentro de um hipertubo, encontrando uma função que tenha um desvio máximo  $\varepsilon$  do valor real obtido de  $y$  para os dados de treino (SMOLA, 2004)[19].

Inicialmente temos uma função linear onde  $b$  é uma constante e  $w$  são os pesos, assim:

$$f(x_i) = b + w \cdot x$$

Em que  $f(x_i)$  é encontrado a partir da minimização da norma:

$$\frac{\|w\|^2}{2} = \frac{w \cdot w}{2}$$

e sujeita às condições:

$$\begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon \\ (w \cdot x_i) + b - y_i \leq \varepsilon \end{cases}$$

Essas condições vêm do uso da função de perda  $\varepsilon - insensível$ . Em que a perda é igual a 0 se a discrepância entre os valores preditos e o observado é menor que  $\varepsilon$  que é a margem do hipertubo (VAPNIK, 2000)[21]. A função de perda é dada por  $L$ :

$$L(y, f(x)) = L(|y - f(x)|_\varepsilon),$$

Em que definimos:

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & \text{se } |y - f(x)| \leq \varepsilon, \\ |y - f(x)| - \varepsilon, & \text{caso contrário} \end{cases}$$

Com o uso da função de perda  $\varepsilon - insensível$  temos que o problema de otimização está restrito aos resíduos absolutos inferiores a  $\varepsilon$  (SMOLA, 2004)[19], mas como em muitos casos vamos precisar permitir alguns erros e para isso podemos inserir variáveis de folga  $\xi_i, \xi_i^*$  para lidar com as restrições de outra maneira que eram impossíveis no problema de otimização, permitindo uma maior flexibilização do modelo.

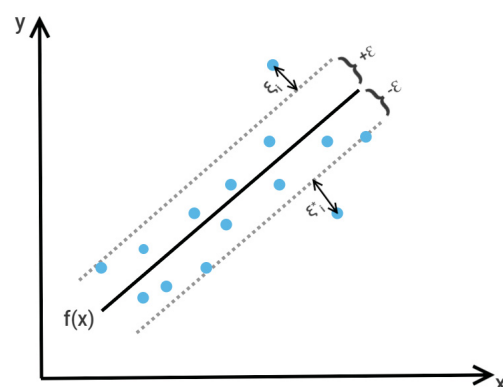
A formulação dada por Vapnik (1995)[22]:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i, \xi_i^*)$$

Em que  $\xi_i, \xi_i^* \geq 0$  são as variáveis de folga, sujeito a:

$$\begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Vemos a inclusão de uma constante  $C > 0$  que determina a penalidade aplicada às observações que estão fora da margem  $\varepsilon$ . A inclusão dessa variável de folga permite uma maior adaptabilidade dos dados e também evitar o *overfitting*. Na Figura 2 podemos observar o comportamento em um SVR de margens suaves.



**Figura 2:** SVR com margens suaves. Fonte: Elaborada pela Autora Adaptado de: Chanklan, R. Kaoungku, N. Suksut, K. Kerdprasop, K. Kerdprasop, Nittaya. (2018). Runoff prediction with a combined artificial neural network and support vector regression. International Journal of Machine Learning and Computing

### 3.3. SVR Mediana

SVR Mediana é um método proposto por Hung, Hung e Lin(2014)[10] em um estudo para previsão de saques em caixas eletrônicos na Inglaterra. A ideia desse método é que a mediana das médias móveis (MA - *moving average*) seja utilizada para previsão da variável  $Y_t$ , pela sua sequência de  $k$  médias aritméticas  $MA_k$ . O uso da média aritmética pode ser visto como uma forma de suavização para séries temporais evitando flutuações indesejadas, porém, essa suavização pode também levar a média móvel a reagir aos eventos com algum atraso. Para exemplificar temos:

$$Y_t \approx MA_k = \frac{1}{k}(Y_{t-1} + Y_{t-2} + \dots + Y_{t-k+1})$$

No método proposto por Hung, Hung e Lin (2014)[10] foi aplicada a mediana de 4 semanas para modelar e prever a série temporal, a variável alvo utilizada é a diferença entre o valor real e a mediana semanal da média móvel. Essa diferença vai ser chamada de  $D_t$  que vai ser a variável resposta a ser predita pelo SVR.

$$\begin{aligned} \text{week\_}MA_2 &= \frac{1}{2}(Y_{t-7} + Y_{t-14}) \\ \text{week\_}MA_3 &= \frac{1}{3}(Y_{t-7} + Y_{t-14} + Y_{t-21}) \\ \text{week\_}MA_4 &= \frac{1}{4}(Y_{t-7} + Y_{t-14} + Y_{t-21} + Y_{t-28}) \end{aligned}$$

$$Y_t \approx D_t + \text{mediana}(\text{week\_}MA_2, \text{week\_}MA_3, \text{week\_}MA_4)$$

Devido às características dos dados utilizados, foi considerado um método de tunagem via *grid* para automatizar o cálculo do SVR-Mediana. Dessa forma, generalizamos a suavização pela média e mediana para qualquer número de dias  $d$  e qualquer número de  $k$ . Assim, foram testados o valor de  $d$  de 1 até 30 dias para trás e o valor de  $k$  entre 4 e 30.

A generalização do método é indicada na Equação 1:

$$\begin{aligned} MA_2 &= \frac{1}{2}(Y_t + Y_{t-d}) \\ MA_3 &= \frac{1}{3}(Y_t + Y_{t-d} + Y_{t-2d}) \\ &\vdots \\ MA_k &= \frac{1}{k} \left( \sum_{j=0}^{k-1} Y_{t-jd} \right) \end{aligned} \quad (1)$$

$$Y_t \approx D_t + \text{mediana}(MA_2, \dots, MA_k)$$

Aplicamos essa mesma função para as outras variáveis do modelo, deixando todos os dados no mesmo formato da variável alvo. Também utilizamos dois tipos de *kernel*: Gaussiano e Laplaciano, para cada  $d$  e  $k$ , totalizando 3240 modelos diferentes até chegar no melhor resultado, que será apresentado com mais detalhes na Seção 5.

### 3.4. Previsão de séries intermitentes

Uma série é considerada como intermitente quando possui valores aleatórios não contínuos e com muitos zeros (Silver, 1981)[17]. Esse é um problema normalmente encontrado em previsões de demandas em estoque, que passam por períodos de baixa e alta demanda e precisam de modelos que consigam prever esses dois cenários. Modelos tradicionais de previsão esperam uma variável contínua no tempo e não conseguem lidar bem com os zeros. Croston (1972)[5] foi pioneiro ao sugerir um novo método para esse tipo de série aplicado a previsão de componentes que tem sido utilizada até hoje como referência para previsão de séries intermitentes.

Para esse artigo, o entendimento da base como série intermitente pela quantidade de zeros e a sazonalidade permitiu que fossemos mais assertivos nos usos das métricas de avaliação do modelo, visto que as métricas normalmente utilizadas como *MAPE* (*Mean absolute percentage of error*) e *sMAPE* (*Symmetric mean absolute percentage error*) não conseguem lidar com zeros. Hyndman e Koehler (2006)[11] sugeriram novas métricas como *MASE* (*Mean absolute scaled error*) e *RMSSE* (*Root mean squared scaled error*) que usam a o erro médio absoluto escalado para avaliar previsões de séries intermitentes.

## 4. Pré processamento

Quando trabalhamos com modelos para séries temporais, um dos principais trabalhos está na limpeza e preparação dos dados. De acordo com BOX e JENKINS(2015)[8], uma série estacionária possui um equilíbrio estatístico com propriedades estatísticas que não mudam ao longo do tempo, tendo a média, a variância e a auto correlação constantes ao longo do tempo.

Como a agregação dos dados do INPE[3] fazendo a contagem da quantidade de focos por dia, não temos um  $Y$  contínuo e também temos vários períodos onde não houve focos de incêndio, 65% da base é de zeros. Além disso, a sazonalidade é bem definida nos meses com tempo mais seco entre julho e outubro, nos quais

estão concentrados a maior quantidade de incêndios. Por este fato, encontramos uma alta variabilidade nos dados.

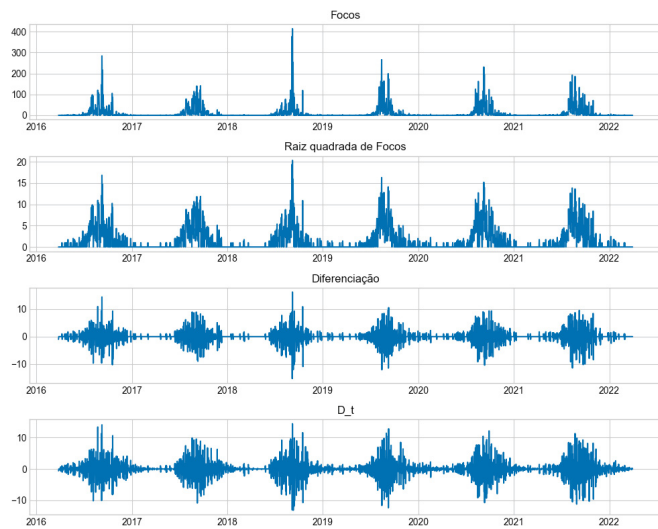
Assim, utilizamos algumas abordagens na variável resposta, antes da aplicação do SVR mediana, a primeira delas foi transformar a variável através da raiz quadrada:

$$\sqrt{Y_t}$$

Através da aplicação da raiz quadrada da variável  $Y$ , obtivemos uma diminuição na variância, para tentar diminuir ainda mais a variância fizemos a diferenciação dos dados em primeiro grau onde  $\Delta y$  vai ser o resultado da diferenciação de  $Y_t$ :

$$\Delta y = Y_t - Y_{t-1}$$

Com isso, aplicamos a função do SVR mediana para transformação dos dados. Na Figura 3 podemos observar como ficou a base de dados a cada transformação.



**Figura 3:** Transformações realizadas na variável foco. Fonte: Elaborado pela autora

O uso somente do SVR mediana, sem as outras transformações, não deixou a série estacionária por conta da variância dos dados. Então a aplicação das duas, foi a melhor forma de deixar os dados com média e variância constante. Também utilizamos o teste de Dickey Fuller para verificar se a série era estacionária que com o  $p$ -value em 0.00, indica que a série é estacionária (Dickey and Fuller 1979)[6].

Nas covariáveis, o primeiro passo foi o tratamento dos dados nulos, para isso foi utilizado a média por dia

Variável	Qt Nulos	% do total	Média	Desvio padrão	Min	Max
precip_total_mm	1460	27,14	6,57	13,29	0,00	117,20
pressao_atmosferica_avg	1396	25,95	996,01	9,18	950,44	1010,07
temp_ponto_orvalho	1408	26,17	21,97	1,79	5,48	2458
temp_max	1134	21,08	32,26	2,41	21,00	40,70
temp_media_compensada	1642	30,52	26,04	1,29	17,52	30,67
temp_min	1125	20,91	22,33	1,70	11,40	26,00
umidade_relativa_media	1136	21,11	79,18	8,43	38,54	94,87
umidade_relativa_min	899	16,71	52,25	13,27	14,00	91,00

**Tabela 2:** Valores das covariáveis antes da transformação

dos valores de toda a série. Nas Tabelas 2 e 3, conseguimos observar o antes e depois da inclusão da média nos dias faltantes, a média permaneceu próxima ao valor anterior e o maior impacto está no desvio padrão.

Variável	Média	Desvio padrão	Min	Max
precip_total_mm	6,57	11,36	0,00	117,20
pressao_atmosferica_avg	996,02	0,29	995,64	997,05
temp_ponto_orvalho	21,97	0,11	21,69	22,17
temp_max	32,26	2,14	21,00	40,70
temp_media_compensada	26,04	1,07	17,52	30,67
temp_min	22,33	1,51	11,40	26,00
umidade_relativa_media	79,18	7,49	38,54	94,87
umidade_relativa_min	52,25	12,12	14,00	91,00

**Tabela 3:** Valores das covariáveis após da transformação

Outros tratamentos realizados incluem a inclusão de um *lag* nas covariáveis para que o valor das covariáveis em  $Y_t$  sejam sempre referente ao dia anterior. Após isso, foram aplicadas as mesmas transformações que foram realizadas na variável resposta, ou seja, a raiz quadrada, diferenciação e SVR Mediana. Também criamos uma nova variável com base nos números dos meses que se mostrou importante para melhorar a previsão do modelo. A base de dados foi dividida em treino e teste, sendo treino entre os anos de 2016-2021 e teste entre os meses de 01-2022 até 03-2022.

Para esse trabalho, iremos comparar a capacidade preditiva para cada tipo de modelo: SVR, SVR com diferenciação, SVR Mediana e SVR Mediana com diferenciação com o uso das covariáveis.

## 5. Resultados e discussão

Para cada valor de  $d$  e  $k$  do SVR Mediana foi aplicado a função de *gridSearch*[18] para busca dos hiper-parâmetros de  $\gamma$ , os valores de  $C$  e foram considerados os valores padrões de 1 e 0,01 respectivamente. Além disso, também foram testados dois *Kernels*: Laplaciano e Gaussiano em cada modelo.

A escolha do melhor parâmetro de  $d$  e  $k$  no SVR mediana foi realizado com o uso de um *grid*, onde o valor

de  $d$  ficou entre 1 e 30, ou seja os valores variaram de 1 dia para trás até 30 dias para trás, com esses valores é calculado os valores da média aritmética( $k$ ) que ficaram entre 4 até 30. Por exemplo, se o valor de  $d$  é igual a 1 e o valor de  $k$  igual a 10, teremos 10 períodos de 1 dia com intervalos de 1 dia cada.

Com a automatização da função do SVR Mediana (1) foi possível testar 810 modelos para cada tipo de SVR, totalizando 3240 modelos testados. A Tabela 4 apresenta o resultado dos 4 melhores modelos encontrados e seus respectivos  $d$  e  $k$ . Para fins de comparação, também testamos os dados apenas com o uso do SVR, com e sem diferenciação na variável  $\sqrt{Y_t}$ .

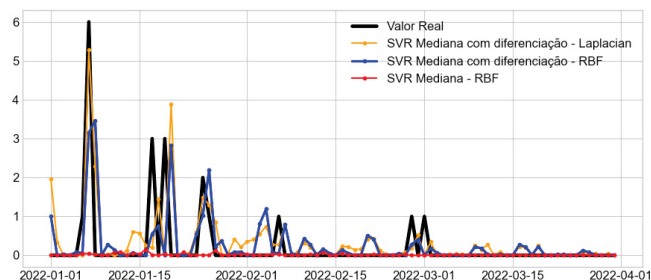
Para escolha das métricas foi necessário levar em conta as características dos dados, por ter muitos zeros, métricas comumente utilizadas como *MAPE* e *SMAPE* retornam valores indefinidos ou infinitos, por isso foram utilizados medidas baseadas em erros relativos, mas escalando o erro baseado no *MAE* (*Mean absolute error*) da previsão ingênua dos dados, essa métrica chamada *MASE* foi proposta por Hyndman e Koehler(2006)[11]. Em que a previsão ingênua supõem que a previsão no período seguinte vai ser igual ao valor do período mais recente. Como o resultado é independente da escala dos dados, quando o erro é  $< 1$ , a previsão foi melhor que a previsão ingênua realizada na amostra ingênua, caso o valor seja  $> 1$ , a previsão do modelo é pior que a previsão da amostra ingênua(HYNDMAN, 2006).[12]. Na Tabela 4 temos os 8 modelos que apresentaram os menores valores de *MASE* na base de treino.

Tipo do modelo	Kernel	d	k	MASE	RMSSE	RMSE	MAE
SVR Mediana com diferenciação	Laplaciano	17	5	0,114	0,142	0,762	0,367
SVR Mediana com diferenciação	RBF	17	14	0,115	0,154	0,780	0,347
SVR Mediana	RBF	5	19	0,137	0,287	0,830	0,830
SVR - sem modificações	RBF	-	-	0,155	0,328	0,836	0,211
SVR - sem modificações	Laplaciano	-	-	0,156	0,328	0,836	0,211
SVR Mediana	Laplaciano	1	4	0,170	0,281	0,845	0,280
SVR com diferenciação	RBF	-	-	0,278	0,434	1,108	0,378
SVR com diferenciação	Laplaciano	-	-	0,279	0,435	1,109	0,380

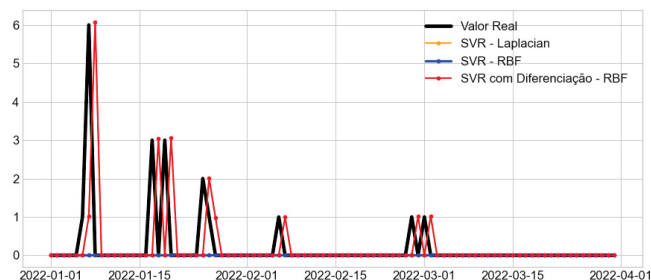
**Tabela 4:** Resultados obtidos nos modelos ordenado pelo menor resultado em MASE

Conseguimos observar nos Gráficos 4 e 5 que os dois modelos com a maior capacidade preditiva são com uso do SVR mediana e diferenciação, foram os modelos que tiveram o menor *MASE* e *RMSSE*. Já no SVR sem modificações, o modelo não consegue performar bem, melhorando apenas com uso de diferenciação.

Com esses resultados é possível observar a importância de aplicação de técnicas para deixar a série estacionária antes da aplicação do modelo, mesmo SVR sendo



**Figura 4:** SVR Mediana. Fonte: Elaborado pela autora



**Figura 5:** SVR. Fonte: Elaborado pela autora

um modelo flexível, em casos em que se tem períodos com muitos zeros, o uso da diferenciação em conjunto com SVR Mediana permitiu encontrar um modelo melhor do que com uso apenas do SVR, tradicionalmente utilizado.

## 6. Comentários finais

Esse trabalho teve como objetivo fazer a previsão de novos focos de incêndio através do Método SVR Mediana, vimos que a previsão de novos focos de incêndio é uma atividade complexa que envolve muitos detalhes principalmente pelas características dos dados utilizados.

De forma geral, os modelos via SVR Mediana obtiveram uma performance superior que o SVR tradicional. Porém, mesmo que o uso da diferenciação tenha melhorado a capacidade preditiva dos focos de incêndio, o método ainda não consegue identificar totalmente os picos e os períodos com ausência de focos de incêndio. Além desta oportunidade de melhoria, este trabalho pode ser continuado através da adaptação do SVR Mediana para séries temporais que considerem estruturas georreferenciadas, bem como métodos específicos para previsão de contagens, além de poder ser estendido para a previsão de vários municípios conjuntamente.



## 7. Agradecimentos

Gostaria de agradecer ao Professor Orientador Anderson Ara por toda ajuda e contribuição nesse artigo. Não posso deixar de agradecer também os meus colegas do grupo de estudos de SVM, pelas inúmeras dicas e por me ajudarem durante todo o processo de desenvolvimento.

## Referências

- [1] Anderson, L.O., Aragão, L.E.O.C, Lima, A., Shimabukuro, C., 2005. Detecção de cicatrizes de áreas queimadas baseada no modelo linear de mistura espectral e imagens índice de vegetação utilizando dados multitemporais dos sensor MODIS/TERRA no estado do Mato Grosso, Amazônia brasileira. Revista Acta Amazonica 35, 445-456.
- [2] Bdmep Disponível em: <<https://bdmep.inmet.gov.br/>>, acesso em 15.04.2022.
- [3] BdQueimadas Disponível em: <<https://queimadas.dgi.inpe.br/queimadas/bd-queimadas>>, acesso em 15.04.2022.
- [4] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273-297 (1995).
- [5] CROSTON, J. D. Forecasting and stock control for intermittent demands. Operational Research Quarterly, vol. 23, nº 3, p. 289-303, 1972.
- [6] Dickey, D. Fuller, Wayne. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. JASA. Journal of the American Statistical Association. 74. 10.2307/2286348.
- [7] Coutinho, A. C. Queimadas no Estado de Mato Grosso. EMBRAPA Monitoramento por satélite, 2007 <<http://www.qmdmt.cnpem.embrapa.br>>
- [8] George E.P Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung Time Series Analysis Forecasting and Control. 5ª edição John Wiley & Sons, 2016.
- [9] HAMEL, L. H. Knowledge discovery with support vector machines.[S.l.]:John Wiley & Sons, 2011.
- [10] HUNG, C.; HUNG, C.-N.; LIN, S.-Y. Predicting time series using integration of moving average and support vector regression. International Journal of Machine Learning and Computing, IACSIT Press, v. 4, n. 6, p. 491, 2014.
- [11] Hyndman, R.J., Koehler A.B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 22, 679-688.
- [12] Hyndman, R.J. 2006. Another look at forecast accuracy metrics for intermittent demand. Foresight: The International Journal of Applied Forecasting, 4(4), 43-46.
- [13] IBGE, Diretoria de Pesquisas, Coordenação agropecuária, Pesquisa da Pecuária Municipal 2020 , Rio de Janeiro V.48, p.1-12, 2020. Disponível em: [https://biblioteca.ibge.gov.br/visualizacao/periodicos/84/ppm\\_2020\\_v48\\_br\\_informativo.pdf](https://biblioteca.ibge.gov.br/visualizacao/periodicos/84/ppm_2020_v48_br_informativo.pdf)
- [14] IBGE. Panorama do Município. Disponível em: <https://cidades.ibge.gov.br/brasil/ro/porto-velho/panorama>
- [15] PRODES Disponível em: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>>, acesso em 17.06.22.
- [16] Python 3.9.6 - Software disponível em: <https://www.python.org/downloads/release/python-396/>
- [17] SILVER, E. Operations research in inventory management: a review and critique. Operations Research, vol. 29, nº 4, p. 628-645, 1981.
- [18] Sklearn - Biblioteca disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>
- [19] Smola, A. J., AND Schölkopf, B. A tutorial on support vector regression. Statistics and computing. Statistics and Computing 14. Kluwer Academic Publishers, n.1, p.200-201, 2004.
- [20] Soares-Filho, B., et al. (2006), Modelling conservation in the Amazon basin, Nature, 440, 520 – 523.
- [21] VAPNIK, V. N. The nature of statistical learning theory, ser. statistics for engineering and information science. New York: Springer, v. 21, n. 1003-1008, p. 182-183, 2000.
- [22] Vapnik V. 1995. The Nature of Statistical Learning Theory. Springer, New York.