UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS

VITÓRIA

2023

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia. Orientador: Prof. Dr. Guilherme A. A. Pereira

VITÓRIA

2023

ALBERSON DA SILVA MIRANDA
MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS/ ALBERSON DA SILVA MIRANDA. – VITÓRIA, 2023-
36p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Guilherme A. A. Pereira

Dissertação (Mestrado) – UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA, 2023.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Universidade xxx. III. Faculdade de xxx. IV. Título

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia.

Aprovada em xx de xx de 20xx.

COMISSÃO EXAMINADORA

Prof. Dr. Guilherme A. A. Pereira
Universidade Federal do Espírito Santo
Orientador

Professor
Instituição

Professor
Instituição

RESUMO

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Palavras-chave: palavra-chave1. palavra-chave2. palavra-chave3.

ABSTRACT

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Keywords: keyword1. keyword2. keyword3.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

MinT	*Minimum Trace*
MCRL	Modelo Clássico de Regressão Linear
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados
ANN	*Artificial Neural Network*
SVR	*Support Vector Regression*

LISTA DE SÍMBOLOS

t	Tempo dentro da amostra
T	Último tempo dentro da amostra, quantidade de observações numa série
h	Horizonte de previsão, tempo fora da amostra
Ω	Conjunto de dados dentro da amostra
y	Série temporal dentro da amostra
$\hat{y}$	Série temporal estimada
$\tilde{y}$	Série temporal reconciliada
n	Número de séries na hierarquia
m	Número de séries no menor nível da hierarquia
k	Número de níveis na hierarquia
$\boldsymbol{S}$	Matriz de soma
$\boldsymbol{G}$	Matriz de reconciliação
$\{...\}$	Conjunto
$\lvert\{...\}\rvert$	Cardinalidade de um conjunto

SUMÁRIO

1 INTRODUÇÃO

[Escrever *outline* da dissertação]

1.1 Problema de pesquisa

1.1.1 Séries hierárquicas e séries agrupadas

Séries temporais hierárquicas são aquelas que podem ser agregadas ou desagregadas naturalmente em uma estrutura aninhada (HYNDMAN; ATHANASOPOULOS, 2021). Para ilustrar, tome a série do PIB brasileiro. Ela pode ser desagregada por estado que, por sua vez, pode ser desagregada por município.

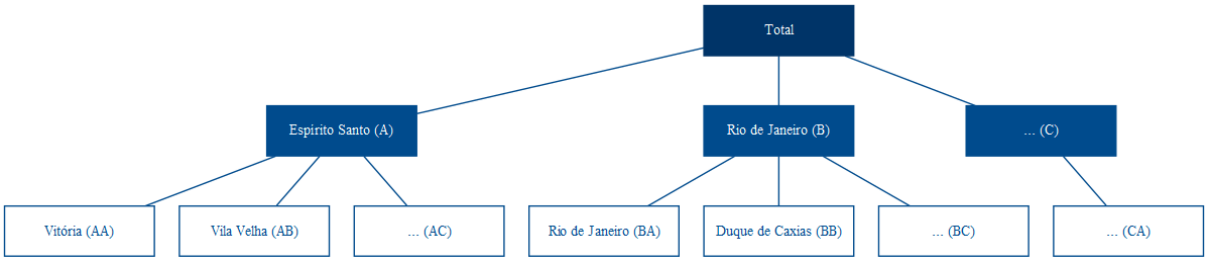

Figura 1 – Séries Hierárquicas

Essa estrutura pode ser representada por equações para qualquer nível de agregação.

$$y_t = y_{A,t} + y_{B,t} + y_{C,t} \tag{1}$$

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BC,t} + y_{CA,t} \tag{2}$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \tag{3}$$

Assim, o agregado nacional pode ser representado apenas pelos agregados dos estados, através de (1), ou como o agregado dos municípios (2). Já o agregado para o estado do Espírito Santo é representado por (3).

Alternativamente, podemos descrever a estrutura completa de forma matricial:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \\ y_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{BC,t} \\ y_{CA,t} \end{bmatrix} \tag{4}$$

Por outro lado, o PIB pode ser também desagregado de forma cruzada de acordo com a atividade econômica — lavoura, rebanho, indústria de transformação, extrativa, bens de capital, bens intermediários, comércio de vestuário, automotivos, serviços etc. Essa estrutura não pode ser desagregada naturalmente de uma única forma, como é a hierarquia de estados e municípios. Não pode ser aninhada por um atributo como a própria geografia. A esse tipo de estrutura dá-se o nome de séries agrupadas.

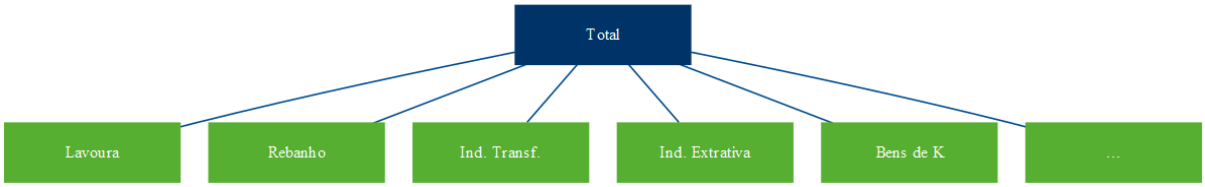

Figura 2 – Séries Agrupadas

Combinando as duas, temos a estrutura de séries hierárquicas agrupadas. Ao contrário da estrutura hierárquica, que só pode ser agregada de uma forma — como com os municípios abaixo dos estados —, a adição da estrutura agrupada pode ocorrer tanto acima (Figura 3) quanto abaixo (Figura 4) da hierárquica.

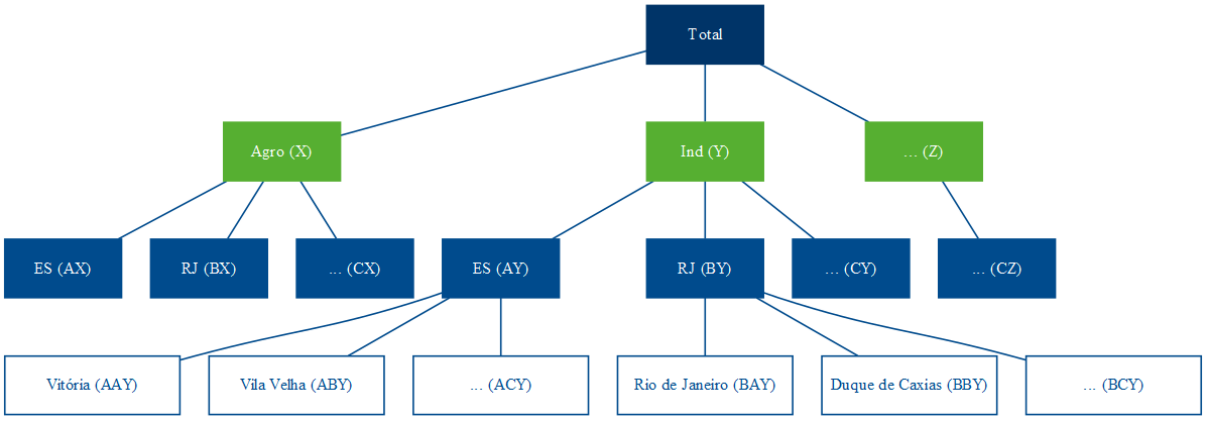

Figura 3 – Séries Hierárquicas Agrupadas (a)

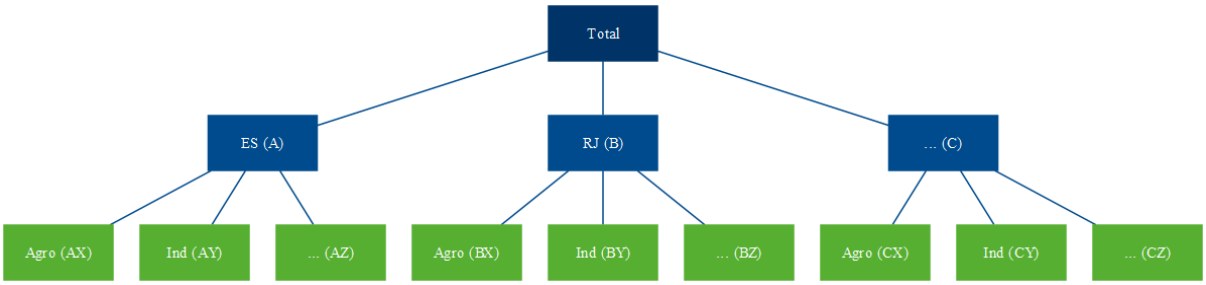

Figura 4 – Séries Hierárquicas Agrupadas (b)

Na notação matricial, a estrutura da Figura 4 é representada como abaixo. Formalmente, o primeiro membro da igualdade é composto pelo vetor $\boldsymbol{y}_t$ n-dimensional com todas as observações no tempo t para todos os níveis da hierarquia. O segundo membro é composto pela matriz de soma $\boldsymbol{S}$ de dimensão $n \times m$ que define as equações para todo nível de agregação, e pela matriz $\boldsymbol{b}_t$ composta pelas séries no nível mais desagregado.

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{Z,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{AZ,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{BZ,t} \\ y_{CX,t} \\ y_{CY,t} \\ y_{CZ,t} \end{bmatrix} \tag{5}$$

1.1.2 Abordagens top-down, bottom-up e middle-out

Talvez as formas mais intuitivas de se pensar em previsões para esses tipos de estrutura sejam as abordagens top-down e bottom-up. Tome a estrutura descrita na Figura 1, por exemplo. Podemos realizar a previsão para o horizonte de tempo h do agregado do PIB brasileiro,

representado no topo da hierarquia por *Total* (6), e então distribuir os valores previstos proporcionalmente entre os estados e municípios.

$$\hat{y}_{T+h|T} = E[y_{T+h}|\Omega_T] \tag{6}$$

Essa é a abordagem top-down. Nela, a previsão para os níveis mais desagregados da hierarquia são determinadas por uma proporção p_i do nível agregado. Por exemplo, as previsões para Vitória são dadas pela equação 7.

$$\tilde{y}_{AA,T+h|T} = p_1 \hat{y}_{T+h|T} \tag{7}$$

Para isso, temos de definir uma matriz com todos esses pesos, que, seguindo a formulação de Hyndman e Athanasopoulos (2021), vamos chamar de $\boldsymbol{G}$:

$$\boldsymbol{G} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

$\boldsymbol{G}$ é uma matriz $m \times n$ que multiplica a matriz $\hat{\boldsymbol{y}}_{T+h|T}$ que, por sua vez, é composta pelas previsões base — as previsões individuais para todos os níveis de agregação. A equação para a abordagem *top-down* será, então:

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T} \tag{9}$$

Na notação matricial para a estrutura da Figura 1, temos:

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \boldsymbol{S} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{y}_{T+h|T} \\ \hat{y}_{A,T+h|T} \\ \hat{y}_{B,T+h|T} \\ \hat{y}_{C,T+h|T} \\ \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{AC,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{BC,T+h|T} \\ \hat{y}_{CA,T+h|T} \end{bmatrix} \tag{10}$$

O que nos dá uma proporção do total para cada elemento no nível mais desagregado.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \boldsymbol{S} \begin{bmatrix} p_1\hat{y}_{T+h|T} \\ p_2\hat{y}_{T+h|T} \\ p_3\hat{y}_{T+h|T} \\ p_4\hat{y}_{T+h|T} \\ p_5\hat{y}_{T+h|T} \\ p_6\hat{y}_{T+h|T} \\ p_7\hat{y}_{T+h|T} \end{bmatrix} \tag{11}$$

Substituindo a matriz $\boldsymbol{S}$, temos as equações que definem cada previsão da estrutura em função de proporções da previsão do agregado.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1\hat{y}_{T+h|T} \\ p_2\hat{y}_{T+h|T} \\ p_3\hat{y}_{T+h|T} \\ p_4\hat{y}_{T+h|T} \\ p_5\hat{y}_{T+h|T} \\ p_6\hat{y}_{T+h|T} \\ p_7\hat{y}_{T+h|T} \end{bmatrix} \tag{12}$$

Já a abordagem bottom-up parte do raciocínio inverso e define as previsões de cada elemento da estrutura a partir das previsões dos elementos mais desagregados. Para tanto, basta modificar a matriz $\boldsymbol{G}$.

$$\boldsymbol{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

O que resulta nas equações desejadas. Portanto, $\boldsymbol{G}$ define a abordagem — se *top-down* ou *bottom-up* —, e $\boldsymbol{S}$ define a maneira da qual as previsões são somadas para formar as equações de previsão para cada elemento da estrutura. Portanto, chamo $\boldsymbol{G}$ de matriz de reconciliação.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{AC,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{BC,T+h|T} \\ \hat{y}_{CA,T+h|T} \end{bmatrix} \tag{14}$$

Quando m — a quantidade de elementos do nível mais desagregado — é muito grande, tornando muito custoso obter $\hat{\boldsymbol{y}}_{\boldsymbol{t}}$, e não se deseja uma abordagem estritamente *top-down*, pode-se combinar as duas formas. Ainda na estrutura hierárquica descrita na Figura 1, obter de forma criteriosa modelos Arima, por exemplo, para cada um dos municípios é muito custoso em tempo. Por outro lado, pode-se realizar a previsão para os estados e então obter de maneira *top-down* as previsões para os municípios, enquanto o nível mais agregado é obtido de maneira *bottom-up*.

$$\boldsymbol{G} = \begin{bmatrix} 0 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

Esse método é chamado de *middle-out*. Nele, o total é o somatório das proporções de um nível intermédiário escolhido, ao invés de proporções do total. Isso permite uma abordagem mais econômica, em termos de custo computacional e de tempo, ao mesmo tempo em que mantém em algum grau as características individuais das hierarquias.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{BC,t} \\ \tilde{y}_{CA,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1\hat{y}_{A,T+h|T} \\ p_2\hat{y}_{A,T+h|T} \\ p_3\hat{y}_{A,T+h|T} \\ p_4\hat{y}_{B,T+h|T} \\ p_5\hat{y}_{B,T+h|T} \\ p_6\hat{y}_{B,T+h|T} \\ p_7\hat{y}_{C,T+h|T} \end{bmatrix} \tag{16}$$

1.1.3 Coerência e reconciliação

Seja somando as previsões do nível mais desagregado para formar os níveis superiores da hierarquia (*bottom-up*) ou distribuindo proporcionalmente as previsões do nível mais agregado (*top-down*), o vetor $\tilde{\boldsymbol{y}}_t$ representa as previsões *coerentes*. Isso significa que as previsões são totalizadas corretamente — as previsões de cada elemento agregado corresponde ao somatório das previsões dos níveis inferiores da hierarquia. Isso é garantido pela multiplicação das matrizes $\boldsymbol{SG}$.

Não fosse essa pré multiplicação, nada garantiria a coerência das previsões. Tomando a estrutura da Figura 1 como exemplo, seria um acaso improvável que as previsões do agregado para o estado do Espírito Santo sejam exatamente a soma das previsões individuais de seus municípios. Isso porque não há qualquer razão para que cada série siga o mesmo processo (e.g., arima) com coeficientes idênticos.

Os métodos de gerar previsões coerentes a partir de previsões base são chamados de métodos de *reconciliação*. Os métodos de reconciliação tradicionais apresentados, *top-down* e *bottom-up*, utilizam informação limitada. No método *top-down*, utiliza-se apenas informações do nível mais agregado — por isso, apenas a primeira coluna em (8) é diferente de zero. Já na abordagem *bottom-up*, utiliza-se apenas as informações dos níveis mais desagregados, o que resulta na submatriz identidade $m \times m$ em (13), enquanto as colunas que representam os níveis mais agregados são nulas.

Alternativamente, podemos pensar numa matriz $\boldsymbol{G}$ qualquer que utilize toda a informação disponível e tenha algumas propriedades que garantam que as previsões coerentes tenham o menor erro o possível. Esse é o problema de pesquisa trabalhado na *reconciliação ótima*.

1.2 Motivação

A projeção da carteira de crédito é um dos elementos fundamentais para o planejamento dos bancos comerciais. Juntamente com as projeções de depósitos, provisões para créditos de liquidação duvidosa, eficiência operacional, entre outros indicadores-chave, essas projeções determinam a temperatura das expectativas da instituição em relação a elementos cruciais como rentabilidade, dividendos e posição no mercado (*market-share*), e isso é essencial para os acionistas e investidores. Essas projeções precisam ser tão precisas quanto possível, para que se possa calcular o risco de transacionar com a instituição financeira, seja como investidor ou cliente.

Embora não exista penalidades específicas para instituições financeiras que erram (por uma boa margem) em suas projeções, elas podem sofrer consequências negativas em outros aspectos, como na avaliação de seus desempenhos por parte dos investidores e clientes. Os investidores e clientes podem considerar as projeções equivocadas como um sinal de falta de competência ou confiança na instituição financeira, o que pode afetar negativamente a reputação e a imagem da instituição. Isso pode levar a uma redução no número de investimentos e depósitos, o que irá afetar diretamente sua saúde financeira.

Além disso, nos casos em que algum grupo se sentir lesado, as instituições financeiras podem enfrentar ações judiciais se suas projeções forem consideradas enganosas ou fraudulentas. Por exemplo, se uma instituição financeira fizer projeções excessivamente otimistas para incentivar os investidores a comprar seus títulos e, posteriormente, as projeções se mostrarem incorretas, os investidores podem entrar com uma ação judicial contra a instituição alegando fraude.

Por isso, é importante que as instituições financeiras sejam transparentes e precisas em suas projeções, fornecendo informações confiáveis e atualizadas para seus clientes e investidores. Entretanto, pouco foi produzido nesse sentido e a aplicação das técnicas envolvendo séries hieárquicas e agrupadas podem contribuir para a qualidade das projeções.

Atualmente, os métodos analíticos, especificamente o MinT, são os mais populares na literatura da reconciliação ótima. Entretanto, tais métodos são sujeitos a uma série de restrições, como as do MCLR, e têm sua capacidade preditiva reduzida quando suas hipóteses são violadas.

Em previsões de séries temporais, o objetivo na maioria dos casos é prever valores futuros com a maior acurácia possível. Em vista disso, métodos de *machine learning* são mais gerais, no sentido de permitir parâmetros não lineares e poderem aproximar virtualmente qualquer função. Além disso, são focados na capacidade preditiva, muitas vezes em detrimento da explicativa. Espera-se, portanto, que esses métodos alcancem melhor performance no problema da reconciliação ótima, devendo receber mais atenção.

1.3 Objetivos

O objetivo geral da dissertação é estudar o problema da reconciliação ótima de previsões pontuais a partir de métodos de *machine learning*.

Como objetivos específicos, tenho:

1. Estudar métodos para estimação da matriz de reconciliação aplicando algoritmos e fluxos de trabalho de *machine learning*, como *tuning* e *resampling*;
2. Identificar possíveis vantagens e limitações da abordagem por *machine learning* na reconciliação de previsões pontuais a partir de aplicação do método estudado na previsão de saldos de crédito do Banestes.

2 REVISÃO DE LITERATURA

2.1 Previsão de saldos de crédito de instituições financeiras

A literatura relacionada à economia bancária é abundante no tema risco de crédito. Podemos verificar através de pesquisa bibliométrica no *Google Scholar* que, salvo exceções, os trabalhos mais citados estão relacionados com a previsão de perdas e suas consequências (Tabela 1). A nível macroeconômico, a previsão de agregados de crédito é uma preocupação de bancos centrais e se manifestam em trabalhos como Bader (2014) e Çolak et al. (2019). No entanto, a pesquisa não revelou nenhum trabalho que se dedique à previsão de saldos de crédito por modalidade e agência.

Apesar da escassez de estudos sobre a previsão de saldos de crédito por modalidade e agência, outras tópicos da economia bancária foram objeto de estudo para previsão de séries

Tabela 1 – Trabalhos mais citados com os termos “banking forecasting”

Autor	Título	Citações	Ano
Jamshidi, N. Hussin	Forecasting patronage factors of Islamic credit card as a new e-commerce banking service: An integration of TAM with perceived religiosity and trust	95	2016
Bernoth, A. Pick	Forecasting the fragility of the banking and insurance sectors	75	2011
Antunes, D. Bonfim, N. Monteiro	Forecasting banking crises with dynamic panel probit models	41	2018
Karminsky, A. Kostrov	The back side of banking in Russia: Forecasting bank failures with negative capital	23	2017
Poorzaker Arabani	The improvement of forecasting ATMS cash demand of Iran banking network using convolutional neural network	18	2019
Yangibayevich, N.B. Absalomovich	Forecasting of processes the turnover and structure of the financial resources of the banking sector	15	2019
Al Wadi, A. Hamarsheh, H. Alwadi	Maximum overlapping discrete wavelet transform in forecasting banking sector	14	2013
Dastoori, S. Mansouri	Credit scoring model for iranian banking customers and forecasting creditworthiness of borrowers	14	2013

temporais, inclusive hierárquicas. Sezer, Gudelek e Ozbayoglu (2019) produziram revisão de literatura de trabalhos publicados entre 2005 e 2019 que realizaram previsão de séries temporais financeiras utilizando *deep learning* e os agruparam em preços de ações individuais, índices (e.g., IBovespa, Dow Jones), preços de commodities, tendência e volatilidade de ativos, preços de títulos, câmbio e preços de criptomoedas.

Outro tema abordado na literatura é a previsão da demanda por moeda em caixas eletrônicos. Gorodetskaya, Gobareva e Koroteev (2021) realizaram uma revisão de literatura recente sobre o assunto e apresentaram sua abordagem para o problema.

No que diz respeito à previsão de séries temporais em largas hierarquias, Prayoga, Suhartono e Rahayu (2017) trabalharam na previsão do fluxo de caixa do Banco da Indonésia, utilizando uma hierarquia de 3 níveis — 40 agências no nível mais desagregado, as 6 grandes ilhas do país como nível intermediário e o total no nível mais agregado. Os autores realizaram um *benchmark* de 5 modelos para previsão da série no nível mais agregado e utilizando o método *top-down* para obter as previsões no nível mais desagregado, concluindo pela efetividade do método *top-down* por proporções históricas. Entretanto, os autores não incluíram reconciliação ótima, a estimativa *bottom-up* ou mesmo outros métodos *top-down para efeito de comparação, o que limita o alcance do trabalho.

2.2 Previsão de séries temporais hierárquicas e agrupadas

A segunda etapa da revisão de literatura consistiu na pesquisa bibliográfica relacionada à reconciliação ótima de previsões de séries temporais hierárquicas e agrupadas e sua interseção com o tema *machine learning*.

A pesquisa bibliométrica na base de dados do Google Acadêmico, pesquisando pelas palavras-chave "*hierarchical forecast reconciliation*" para qualquer lugar no corpo do texto, encontrando 27.600 resultados. Ordenando os resultados pelo número de citações[1], verifiquei que o trabalho mais citado é Hyndman e Athanasopoulos (2021).

Tabela 2 – Trabalhos mais citados com os termos "hierarquical forecast reconciliation"

Autor	Citações	Ano
Hyndman, G. Athanasopoulos	5222	2018
Dellarocas, X. Zhang	2082	2007
Hyndman, A. Lee, E. Wang, S. Wickramasuriya	1023	2013
Badre, M. D'esposito	974	2009
Hong, S. Fan	912	2016
Tashman	798	2000

Utilizando essa obra como texto base, obtive os textos referenciados no capítulo 11 "*Forecasting Hierarquical and Grouped Time-Series*", subcapítulo 3 "*Forecast Reconciliation*", além de Hyndman, Lee e Wang (2016), onde o método por MQP foi desenvolvido porém não está citado nas referências do capítulo.

Quadro 1 – Artigos de referência em Hyndman e Athanasopoulos (2021)

Autor	Ano
Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., Shang, H. L.	2011
Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., Hyndman, R. J.	2021
Wickramasuriya, S. L., Athanasopoulos, G., Hyndman, R. J.	2019
Rob J. Hyndman and Alan J. Lee and Earo Wang	2016

Adicionando o termo "*machine learning*" e refinando a pesquisa para encontrar as palavras chave no título dos trabalhos, 10 resultados foram encontrados.

2.2.1 Abordagens de nível único

Uma abordagem de nível único é uma abordagem em que as previsões são realizadas para um único nível da hierarquia. A partir dessas previsões, os demais níveis são obtidos, ou desagregando (no caso dos níveis inferiores), ou agregando (no caso dos níveis superiores) essas informações (HYNDMAN; ATHANASOPOULOS, 2021). Os métodos *top-down*, *bottom-up* e *middle-out*, demonstrados na introdução, são abordagens de nível único.

Enquanto há apenas uma única forma de se agregar níveis na hierarquia (*bottom-up*), a desagregação (*top-down*) pode ser realizada de, ao menos, duas dezenas de maneiras (GROSS;

[1] Para a funcionalidade, ver https://github.com/WittmannF/sort-google-scholar.

Quadro 2 – Trabalhos encontrados na busca estendida

Autor	Ano	Citacoes
Li, S.M.M. Rahman, R. Vega, B. Dong	2016	114
Mancuso, V. Piccialli, A.M. Sudoso	2021	17
Abolghasemi, R.J. Hyndman, G. Tarr	2019	13
Afolabi, Su Guan, K.L. Man, P.W.H. Wong, X. Zhao	2017	20
Saatloo, A. Moradzadeh, H. Moayyed	2021	6
Abolghasemi, G. Tarr, C. Bergmeir	2022	0
C. Neto, B.L. Fernando	2022	0
Moon	2012	0
Yan, C. Sheng	2018	1
Varone, C. Ieracitano, A. Özyüksel, T. Hussain	2022	0

SOHL, 1990). Dois dos métodos mais intuitivos são a média das proporções históricas e a proporção das médias históricas.

Na média das proporções históricas, cada proporção p_j, com $j = 1, ..., m$, consiste em tomar a média das proporções da série desagregada $y_{j,t}$ em relação ao agregado y_t:

$$p_j = \frac{1}{T} \sum_{t=1}^{T} \frac{y_{j,t}}{y_t} \tag{17}$$

Já a proporção das médias históricas consiste em tomar a proporção das médias das séries desagregadas em relação à média do agregado[2].

$$p_j = \frac{\sum_{t=1}^{T} \frac{y_{j,t}}{T}}{\sum_{t=1}^{T} \frac{y_t}{T}} \tag{18}$$

[escrever método de proporções de previsões]

Li et al. (2016) compararam dois algoritmos de *machine learning* para previsão da produção de energia solar no estado da Flórida/EUA: ANN e SVR. Argumentando que tradicionalmente as previsões nesse problema são realizadas com os dados de produção total da planta, eles propõem uma abordagem hierárquica *bottom-up*, com previsões base de cada inversor solar. Os autores concluem que a abordagem hierárquica *bottom-up* é mais precisa do que a previsão do agregado, ao menos na previsão um passo a frente. Embora os autores utilizem algoritmos de *machine learning* para as previsões base, eles não utilizam esses algoritmos para a reconciliação ótima, caracterizando a abordagem do trabalho ainda como nível único.

[2] Isso é equivalente a tomar a proporção direta entre os somatórios das séries. Note que, pelas propriedades do operador de somatório, $\sum_{t=1}^{T} \frac{y_t}{T} = \frac{y_1}{T} + \ldots + \frac{y_T}{T} = \frac{y_1 + \ldots + y_T}{T} = \frac{\sum_{t=1}^{T} y_t}{T}$. Então, a equação 18 pode ser simplificada para $p_j = \frac{\sum_{t=1}^{T} y_{j,t}}{\sum_{t=1}^{T} y_t}$.

2.2.2 Métodos analíticos para reconciliação ótima

Previsões pontuais de séries temporais hierárquicas não é um assunto novo. Ao menos desde a década de 70, pesquisas foram publicadas acerca de abordagens *bottom-up* e *top-down*, suas vantagens e desvantagens, e tentativas de se definir qual é o melhor método[3]. Entretanto, é apenas em Hyndman, Ahmed et al. (2011) que é formalizada uma abordagem prática que utiliza toda a informação disponível, (i.e. as previsões de todos elementos de todos os níveis da hierarquia) a partir da estimação da matriz $\boldsymbol{G}$ via regressão linear por mínimos quadrados generalizados (MQG).

Entretanto, para ser capaz de estimar o modelo por MQG, é necessária a matriz de variância-covariância dos erros. Hyndman, Ahmed et al. (2011) usam a matriz de erros de coerência, ou seja, a diferença entre as previsões reconciliadas e as previsões base, que tem posto incompleto e não identificada e, portanto, não pode ser estimada. Os autores contornam esse problema adotando no lugar da matriz de variância-covariância dos erros uma matriz diagonal constante, ou seja, assumem variância constante dos erros de reconciliação, e estimam a matriz $\boldsymbol{G}$ por mínimos quadrados ordinários (MQO).

A estimação por esse método resulta numa reconciliação ótima que depende apenas da matriz $\boldsymbol{S}$, ou seja, da estrutura hierárquica, e independe da variância e covariância das previsões base $\hat{\boldsymbol{y}}_{\boldsymbol{T+h}}$ — o que não é uma conclusão satisfatória.

Hyndman, Lee e Wang (2016) tentam aperfeiçoar o método usando as variâncias das previsões base estimadas (dentro da amostra) como estimativa para a matriz de variância-covariância dos erros de reconciliação, de forma a as utilizar como pesos e realizar a reconciliação ótima por mínimos quadrados ponderados (MQP). Assim, previsões base mais acuradas têm peso maior do que as mais ruidosas. Entretanto, não fornecem justificativa teórica para usar a diagonal da matriz de variância-covariância de $\hat{\boldsymbol{e}}_{\boldsymbol{t}}$.

Wickramasuriya, Athanasopoulos e Hyndman (2019) argumentam que o que de fato interessa é que as previsões reconciliadas tenham o menor erro. Então, corrigem a abordagem de reconciliação ótima para o objetivo de minimização dos erros das previsões reconciliadas $\tilde{\boldsymbol{y}}_{\boldsymbol{t+h}}$, ao invés dos erros das previsões base $\hat{\boldsymbol{y}}_{\boldsymbol{t+h}}$. Dado que isso implica na minimização da variância de $\tilde{\boldsymbol{e}}_{\boldsymbol{t+h}}$, ou seja, na minimização do somatório da diagonal, o traço, da matriz de variância-covariância de $\tilde{\boldsymbol{e}}_{\boldsymbol{t+h}}$, eles chamaram esse método de Traço Mínimo (MinT, na sigla em inglês). Paralelamente, usam desigualdade triangular para demonstrar que as previsões reconciliadas obtidas por esse método são ao menos tão boas quanto as previsões base.

Panagiotelis et al. (2021) reinterpreta a literatura de coerência e reconciliação de previsões pontuais a partir de uma abordagem geométrica, trazendo provas alternativas para conclusões anteriores ao mesmo tempo em que fornece novos teoremas. Além disso, Panagiotelis et al. (2021) estende essa interpretação geométrica para o contexto probabilístico, fornecendo

[3] Uma revisão dessa literatura pode ser encontrada em Athanasopoulos, Ahmed e Hyndman (2009).

métodos paramétricos e não paramétricos (via *bootstrapping*) para reconciliação de previsões probabilísticas, ou seja, para reconciliar previsões $\hat{y}_t$ obtidas a partir de toda a distribuição, e não apenas a média.

2.2.3 Métodos de machine learning para reconciliação ótima

Spiliotis et al. (2021) propõem a utilização de *machine learning* para a reconciliação ótima de séries temporais, especificamente os algoritmos de árvore de decisão *Random Forest* e *XGBoost*. Os autores descrevem como vantagens desse método em relação aos anteriores a descrição de relacionamentos não lineares, performance preditiva e a desnecessidade da utilização de todos os elementos da hierarquia na combinação ótima. A abordagem utilizada foi:

1. dividir a amostra em treino e teste;
2. treinar um modelo de previsão na amostra treino e obter previsões um passo a frente para a amostra teste;
3. treinar um modelo de *machine learning* para cada série do menor nível da hierarquia, em que os parâmetros são as previsões obtidas no passo 2 e a variável explicada são os valores observados. Isso resulta em um modelo de reconciliação ótima para cada elemento do menor nível da hierarquia, combinando informações disponíveis de todos os níveis hierárquicos;
4. obter as previsões base $\hat{y}_t$;
5. passar as previsões base ao modelo treinado no passo 3 para se obter as previsões reconciliadas para o menor nível da hierarquia;
6. agregar as previsões reconciliadas para se obter as previsões nos demais níveis hierárquicos.

Para o conjunto de dados utilizados, Spiliotis et al. (2021) afirmam que os métodos de *machine learning*, especialmente o XGBoost, alcançaram, em média, melhor performance que as abordagens de nível único e o MinT. Além disso, concluíram que quanto maior é a diferença entre as séries, em todos os níveis hierárquicos, maior são os benefícios da abordagem por *machine learning*.

2.2.3.1 O processo de ajuste e sobreajuste

Considere uma função de ajuste f, um conjunto de pontos $D = d_1, ..., d_n$ com $d_i = (x_i y_i)´$, variáveis de decisão ou parâmetros $x_i \in \mathbb{R}^m$ e imagem $y_i = f(x_i) \in \mathbb{R}$. Diferentemente da abordagem clássica, em que, no caso do modelo clássico de regressão linear, há um modelo teórico de coeficientes estimados por mínimos quadrados ordinários (MQO) que é garantido pelo teorema de Gauss-Markov ser o melhor estimador linear não viesado (BLUE),

em *machine learning* o objetivo é encontrar, de forma iterativa, um meta-modelo que melhor aproxima a função f usando a informação contida em D, ou seja, queremos ajustar uma função de regressão $\hat{f}_D$ aos nossos dados D de forma que $\hat{y} = \hat{f}_D(x, \varepsilon)$ tenha o menor erro de aproximação ε.

Para verificar o quão bem o modelo $\hat{f}_D$ se aproxima da função real f, é necessário uma função de perda $L(y, \hat{f}(x))$ que, no caso de regressão, será a perda quadrática $(y - \hat{f}(y))^2$ ou a perda absoluta $|y - \hat{f}(y)|$. Esses valores são agregados pela média para formar as funções de custo erro médio quadrático (MSE) e erro médio absoluto (MAE).

Dada a função de perda, pode-se definir o risco associado ao modelo de função de ajuste

$$R(f, p) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} L(y, f(x))p(x, y)dxdy$$

em que $p(x, y)$ é a função densidade de probabilidade conjunta. Como não temos a função real mas procuramos uma função estimada que se aproxime dela, temos

$$GE(\hat{f}_D, p) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} L(y, \hat{f}_D(x))p(x, y)dxdy \tag{1}$$

que é o erro de generalização ou risco condicional associado ao preditor.

Então, podemos estimar o erro de generalização do modelo. Como não conhecemos a distribuição P, a substituímos pela amostra de teste D^* e ficamos com

$$\widehat{GE}(\hat{f}_D, D^*) = \sum_{(xy)' \in D^*} \frac{L(y, \hat{f}_D(x))}{|D^*|}$$

Se substituirmos a amostra teste pela amostra treino D usada para ajustar o modelo, teremos o chamado erro de resubstituição

$$\widehat{GE}_{\text{resub}} = \widehat{GE}(\hat{f}_D, D)$$

Naturalmente, nesse caso estaríamos usando os dados de treino tanto para treinar o preditor quanto para estimar o erro de generalização, o que nos levaria a uma estimativa enviesada do erro de generalização. Caso usássemos essa estimativa para seleção de modelos, esse viés favoreceria modelos mais adaptados à amostra.

O problema é que nesses modelos, dadas suficientes iterações, o erro de resubstituição tende a zero. Isso acontece porque conforme o preditor se adapta cada vez mais aos dados de treinamento ele irá memorizar a relação entre o conjunto de pontos D e a imagem $f(x_i)$, ou seja, irá se ajustar perfeitamente ao formato da função a ser modelada. E não necessariamente

um modelo perfeitamente ajustado se traduz na capacidade de predição de dados futuros (fora da amostra).

De forma geral, espera-se que o preditor reduza seu viés durante o treino apenas o suficiente para que seja capaz de generalizar sua predição para fora da amostra em um nível ótimo de acurácia. A partir desse ponto, a redução no viés é penalizada com o aumento da variância, ou seja, com a redução de sua capacidade de prever dados futuros (BISCHL et al., 2012). A esse processo se dá o nome de *overfitting* ou **sobreajuste**. Isso quer dizer que não podemos considerar a performance do preditor em D se desejamos estimar honestamente a performance real do modelo.

2.2.3.2 Reamostragem

Uma forma de se corrigir esse problema é dividindo a amostra em um conjunto para treino D_{treino} e outro conjunto para teste D_{teste} de forma que $D_{\text{treino}} \cup D_{\text{teste}} = D$ e $D_{\text{treino}} \cap D_{\text{teste}} =$. Assim, pode-se treinar o modelo em D_{treino} para se obter $\hat{f}_{D_{\text{treino}}}$ e calcular seu erro de generalização usando os dados de D_{teste}. Essa abordagem é chamada de *hold-out* e ela é de simples implementação e utilização, uma vez que as observações do conjunto teste são completamente independentes das observações com as quais o modelo foi treinado. A estimativa do erro de generalização então se torna

$$\widehat{GE}_{\text{hold-out}} = \widehat{GE}(\hat{f}_{D_{\text{treino}}}, D_{\text{teste}})$$

Dois problemas permanecem:

1. É necessária uma amostra grande, uma vez que deve-se ter dados suficientes tanto na amostra treino para ajustar um modelo adequado, quanto na amostra teste para realizar uma avaliação de performance estatisticamente válida.

2. Esse método não é suficiente para detectar variância e instabilidades na amostra treino. Modelos mais complexos, especialmente não lineares, podem produzir resultados muito diferentes com mudanças pequenas nos dados de treino.

É exatamente para lidar com essas situações que foram desenvolvidas as técnicas de reamostragem. Todas essas tecnicas geram repetidamente i subconjuntos de treino $D^{(i)}_{\text{treino}}$ e teste $D^{(i)}_{\text{teste}}$ com o dataset disponível, ajustam um modelo com cada conjunto de treino e atestam sua qualidade no conjunto de teste correspondente. A estimativa do erro de generalização então se torna

$$\widehat{GE}_{\text{samp}} = \frac{1}{k}\sum_{i=1}^{k} \widehat{GE}(\hat{f}_{D^{(i)}_{\text{treino}}}, D^{(i)}_{\text{teste}}) \tag{2}$$

O erro de generalização dado na equação (1) depende tanto do tamanho da amostra usada para treinar quanto para testar o modelo ajustado. Portanto, devemos garantir que o tamanho da amostra usado para verificar o erro de generalização de um modelo estimado a partir de n *data points* seja próximo de *n*. Se, por exemplo, o conjunto de treino for muito menor que a amostra total o erro será superestimado, uma vez que muito menos informação foi usada para calcular o estimador.

Da mesma forma, a qualidade do estimador do erro de generalização obtido em (2) a partir de uma estratégia de reamostragem também depende muito do tamanho dos conjuntos $D^{(i)}$ em relação à amostra original, da quantidade k de subconjuntos utilizados e da estrutura de dependência entre os subconjuntos $D^{(i)}$ — novamente, modelos mais complexos são mais sensíveis a alterações no dataset e a variância entre os subconjuntos tende a ser maior. O erro do estimador é geralmente medido pelo erro médio quadrático (MSE):

$$\text{MSE}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[(\widehat{GE}_{\text{samp}} - GE(\hat{f}_D, P))^2]$$

Esse estimador também pode ser representado como a soma do quadrado do viés e a variância:

$$\text{MSE}(\widehat{GE}_{\text{samp}}) = \text{Bias}(\widehat{GE}_{\text{samp}})^2 + \text{Variância}(\widehat{GE}_{\text{samp}})$$

Sendo que o viés expressa a diferença média entre um estimador e o valor real, enquanto a variância mede a dispersão média do estimador. Essas quantidades são definidas da seguinte forma:

$$\text{Bias}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[\widehat{GE}_{\text{samp}}] - \mathbb{E}[GE(\hat{f}_D, p)]$$

e

$$\text{Variância}(\widehat{GE}_{\text{samp}}) = \mathbb{E}[(\widehat{GE}_{\text{samp}} - \mathbb{E}[\widehat{GE}_{\text{samp}}])^2]$$

3 METODOLOGIA

Neste capítulo estão contidas explicações sobre os dados e variáveis, sobre o *design* da modelagem e sobre a avaliação dos modelos.

3.1 Dados e variáveis

Os dados usados nesse trabalho são dados terciários obtidos do *datalake* público Base dos Dados (CAVALCANTE; HERSZENHUT; DORNELLES, 2023). A fonte primária são os bancos comerciais e múltiplos com carteira comercial que disponibilizam mensalmente os saldos dos principais verbetes do balancete via documento 4500[4] ao Banco Central do Brasil, que os compila e publica, agrupados por agência bancária e por município, no relatório ESTBAN — Estatística Bancária Mensal e por Município[5].

Além das estatísticas bancárias, foram obtidos informações de regiões, mesorregiões e microrregiões dos estados, também a partir *datalake* Base dos Dados, com o objetivo de enriquecer a estrutura hierárquica dos dados do ESTBAN, limitada aos municípios.

Uma vez que o escopo deste trabalho se encerra ao Espírito Santo e ao Banestes, foram aplicados os filtros na para UF e na raiz do CNPJ. Além disso, foram mantidos apenas os verbetes relacionados a crédito. Quanto ao período, há dados disponíveis desde 1988. Entretanto, escolhi manter apenas os dados a partir de 2010 para evitar problemas relacionados a séries muito longas. Isso porque é seria muito otimista assumir que o processo de autocorrelação e o padrão sazonal se manteria ao longo de várias décadas[6]. Além disso, se tratando de uma hierarquia larga, o custo computacional deve ser levado em conta na escolha do período.

Por fim, as variáveis mantidas no *dataset* foram:

1. ref: data de referência do relatório ESTBAN
2. nome_mesorregiao: nome da mesorregião do ES:
    - Central Espírito-Santense
    - Litoral Norte Espírito-Santense
    - Noroeste Espírito-Santense
    - Sul Espírito-Santense
3. verbete:
    - empréstimos e títulos descontados
    - financiamentos
    - financiamentos imobiliários
    - financiamentos rurais
4. id_município: código do município
5. cnpj_agencia

[4] Esses documentos são relatórios eletrônicos obrigatórios demandados pelo Bacen às instituições financeiras que permitem ao regulador o conhecimento minucioso dos bancos e de seus clientes.

[5] https://www4.bcb.gov.br/fis/cosif/estban.asp?frame=1

[6] "It is, perhaps, unrealistic to assume that the seasonal pattern remains the same over nearly three decades. So we could simply fit a model to the most recent years instead" (HYNDMAN; ATHANASOPOULOS, 2021)

6. valor: saldo do verbete no município

Os dados foram organizados de forma hierárquica, do mais agregado para o mais desagregado, por estado, mesorregião, município e agência bancária; e de forma agrupada, por verbete (Figura 5).

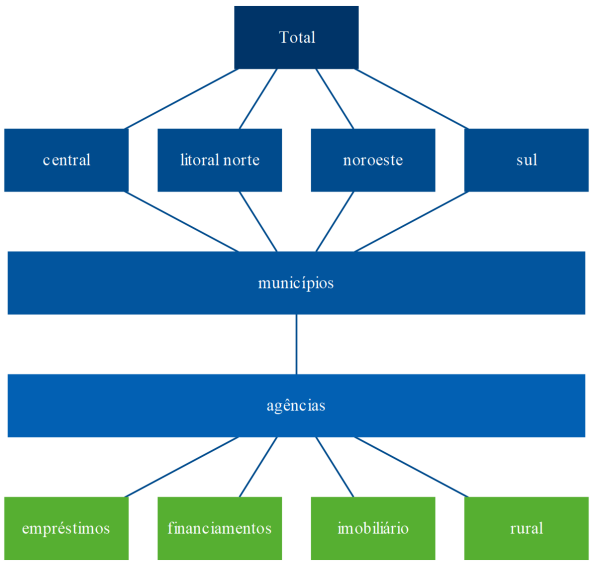

Figura 5 – Hierarquia dos dados

O detalhamento da construção do *dataset* se encontra no Anexo A.

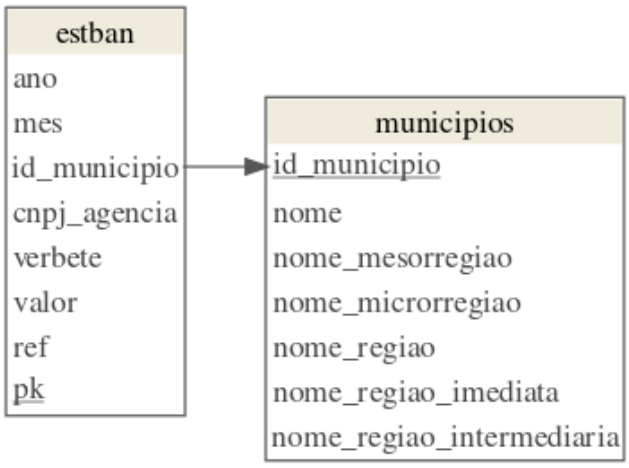

Figura 6 – Modelo de dados

3.2 Análise exploratória dos dados

Analisando os verbetes no agregado do estado,

4 RESULTADOS

[Escrever *outline* do capítulo]

Tabela 3 – Estrutura do dataset ESTBAN

	x
ref	154
nome_mesorregiao	4
verbete	4
id_municipio	72
cnpj_agencia	97

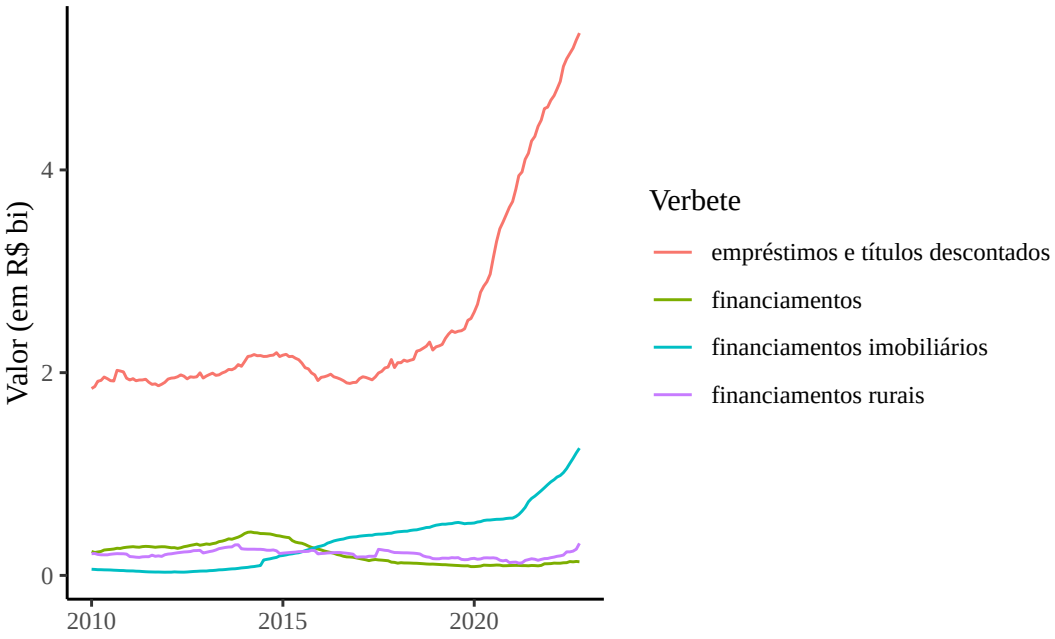

Figura 7 – Verbete no agregado do ES

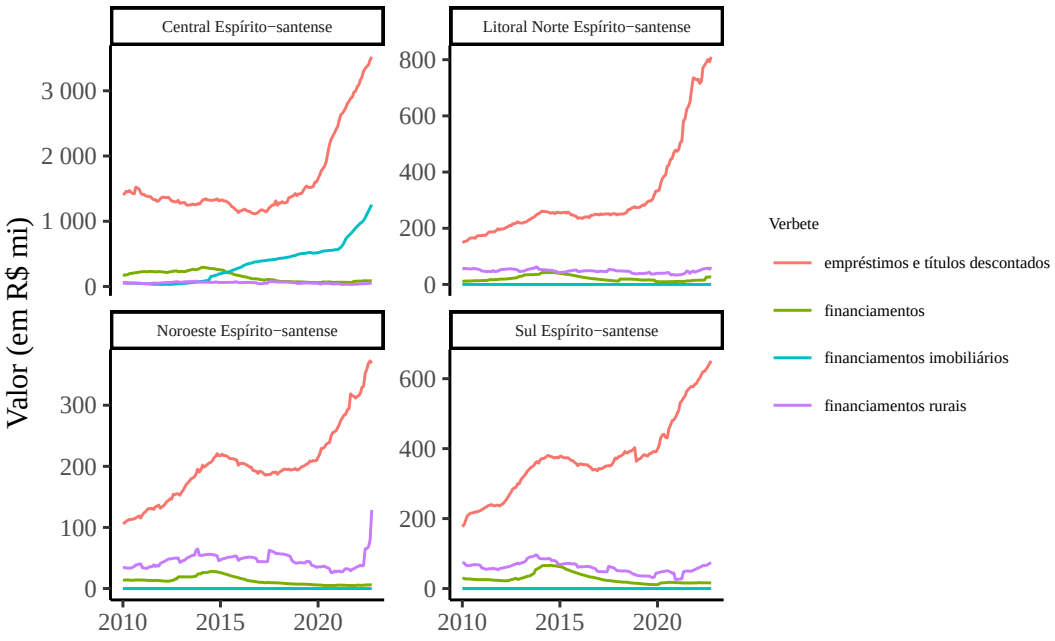

Figura 8 – Verbete por mesorregião do ES

REFERÊNCIAS

ATHANASOPOULOS, George; AHMED, Roman A.; HYNDMAN, Rob J. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, v. 25, n. 1, p. 146–166, 1 jan. 2009. ISSN 0169-2070. DOI: 10.1016/j.ijforecast.2008.07.004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207008000691>. Acesso em: 11 jan. 2023. Citado na p. 24.

BADER, Fani Lea Cymrot. Modelo FAVAR Canônico para Previsão do Mercado de Crédito. *Banco Central do Brasil*, v. 369, 2014. ISSN 1519-1028. Citado na p. 20.

BISCHL, Bernd et al. Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. *Evolutionary computation*, v. 20, p. 249–75, 16 fev. 2012. DOI: 10.1162/EVCO_a_00069. Citado na p. 27.

CAVALCANTE, Pedro; HERSZENHUT, Daniel; DORNELLES, Rodrigo. *basedosdados: Base Dos Dados R Client*. [S.l.], 2023. R package version 0.2.2. Disponível em: <https://CRAN.R-project.org/package=basedosdados>. Citado na p. 29.

ÇOLAK, Mehmet Selman et al. *TCMB - Monitoring and Forecasting Cyclical Dynamics in Bank Credits: Evidence from Turkish Banking Sector*. 2019. Disponível em: <https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Paperss/2019/19-29>. Acesso em: 6 mar. 2023. Citado na p. 20.

GORODETSKAYA, Olga; GOBAREVA, Yana; KOROTEEV, Mikhail. A Machine Learning Pipeline for Forecasting Time Series in the Banking Sector. *Economies*, v. 9, n. 4, p. 205, dez. 2021. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. ISSN 2227-7099. DOI: 10.3390/economies9040205. Disponível em: <https://www.mdpi.com/2227-7099/9/4/205>. Acesso em: 27 fev. 2023. Citado na p. 21.

GROSS, Charles W.; SOHL, Jeffrey E. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, v. 9, n. 3, p. 233–254, 1990. _eprint: https://onlinelibrary.wiley.com/doi/pdf/1 ISSN 1099-131X. DOI: 10.1002/for.3980090304. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980090304>. Acesso em: 26 jan. 2023. Citado na p. 22.

HYNDMAN, R.J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. 3. ed. Melbourne, Austrália: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Citado nas pp. 12, 15, 22, 29.

HYNDMAN, Rob J.; AHMED, Roman A. et al. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, v. 55, n. 9, p. 2579–2589, 1 set. 2011. ISSN 0167-9473. DOI: 10.1016/j.csda.2011.03.006. Disponível em: <https://www.scienc

edirect.com/science/article/pii/S0167947311000971>. Acesso em: 11 jan. 2023. Citado na p. 24.

HYNDMAN, Rob J.; LEE, Alan J.; WANG, Earo. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, v. 97, p. 16–32, 1 mai. 2016. ISSN 0167-9473. DOI: 10.1016/j.csda.2015.11.007. Disponível em: <https://www.sciencedirect.com/science/article/pii/S016794731500290X>. Acesso em: 11 jan. 2023. Citado nas pp. 22, 24.

LI, Zhaoxuan et al. A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting. en. *Energies*, v. 9, n. 1, p. 55, jan. 2016. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. ISSN 1996-1073. DOI: 10.3390/en9010055. Disponível em: <https://www.mdpi.com/1996-1073/9/1/55>. Acesso em: 8 abr. 2023. Citado na p. 23.

PANAGIOTELIS, Anastasios et al. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, v. 37, n. 1, p. 343–359, 1 jan. 2021. ISSN 0169-2070. DOI: 10.1016/j.ijforecast.2020.06.004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207020300911>. Acesso em: 15 jan. 2023. Citado na p. 24.

PRAYOGA, I.G.S.A.; SUHARTONO, Suhartono; RAHAYU, S.P. Top-down forecasting for high dimensional currency circulation data of Bank Indonesia. *International Journal of Advances in Soft Computing and its Applications*, v. 9, p. 62–74, 1 jan. 2017. Citado na p. 21.

SEZER, Omer Berat; GUDELEK, Mehmet Ugur; OZBAYOGLU, Ahmet Murat. *Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019*. [S.l.]: arXiv, 29 nov. 2019. arXiv: 1911.13288[cs,q-fin,stat]. Disponível em: <http://arxiv.org/abs/1911.13288>. Acesso em: 7 mar. 2023. Citado na p. 21.

SPILIOTIS, Evangelos et al. Hierarchical forecast reconciliation with machine learning. *Applied Soft Computing*, v. 112, p. 107756, 1 nov. 2021. ISSN 1568-4946. DOI: 10.1016/j.asoc.2021.107756. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494621006773>. Acesso em: 11 jan. 2023. Citado na p. 25.

WICKRAMASURIYA, Shanika L.; ATHANASOPOULOS, George; HYNDMAN, Rob J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, v. 114, n. 526, p. 804–819, 3 abr. 2019. Publisher: Taylor & Francis. ISSN 0162-1459. DOI: 10.1080/01621459.2018.1448825. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825>. Acesso em: 11 jan. 2023. Citado na p. 24.

Anexos

ANEXO A – CÓDIGO PARA CONSTRUÇÃO DA BASE DE DADOS

```
# pacotes
library(magrittr, include.only = "%>%")

# municípios x regiões imediatas
municipios = basedosdados::read_sql("
SELECT
  id_municipio
  , nome
  , nome_mesorregiao
  , nome_microrregiao
  , nome_regiao
  , nome_regiao_imediata
  , nome_regiao_intermediaria
FROM `basedosdados.br_bd_diretorios_brasil.municipio`
WHERE sigla_uf = 'ES'
")

# importando dados
estban = basedosdados::read_sql("
SELECT
  CAST(ano AS STRING) AS ano
  , CAST(mes AS STRING) AS mes
  , id_municipio
  , cnpj_agencia
  , CASE
        WHEN id_verbete = '160' THEN 'operações de crédito'
        WHEN id_verbete = '161' THEN 'empréstimos e títulos descontados'
        WHEN id_verbete = '162' THEN 'financiamentos'
        WHEN id_verbete = '163' THEN 'financiamentos rurais'
        WHEN id_verbete = '169' THEN 'financiamentos imobiliários'
        WHEN id_verbete = '172' THEN 'outros créditos'
        WHEN id_verbete = '174' THEN 'provisão para operações de crédito'
        ELSE 'outros'
    END AS verbete
  , valor
```

```
FROM `basedosdados.br_bcb_estban.agencia`

WHERE
  cnpj_basico = '28127603'
  AND id_verbete IN ('161', '162', '163', '169')
")

# formatando datas
estban = within(estban, {
  mes = formatC(as.numeric(mes), format = "d", width = 2, flag = "0")
  ref = as.Date(paste(ano, mes, "01", sep = "-"))
})

# identificando agências em atividade
agencias_fim = subset(estban, ref == max(ref), select = cnpj_agencia) |>
  (\(x) unique(x$cnpj_agencia))()

# filtrando apenas agências em atividade
estban = subset(estban, cnpj_agencia %in% agencias_fim)

# mesclando com tabela municípios
estban_df = merge(estban, municipios, by = "id_municipio")

# adicionando estrutura hierárquica e agrupada
estban = tsibble::tsibble(
  estban_df,
  index = ref,
  key = c("id_municipio", "cnpj_agencia", "verbete", "nome_mesorregiao")
)

estban = estban |>
  fabletools::aggregate_key(
    (nome_mesorregiao / id_municipio / cnpj_agencia) * verbete,
    valor = sum(valor)
)
```