

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

2024

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

2024

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE
SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS/ ALBERSON DA SILVA MI-
RANDA. – VITÓRIA, 2024-

70p. : 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, 2024.

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

Figura 1 – Séries Hierárquicas	29
Figura 2 – Séries Agrupadas	30
Figura 3 – Séries Hierárquicas Agrupadas (a)	30
Figura 4 – Séries Hierárquicas Agrupadas (b)	30
Figura 5 – Crescimento de árvores em algoritmos de <i>boosting</i>	39
Figura 6 – <i>Support vector classifiers</i> . Fonte: Hastie, Tibshirani e Friedman (2009).	41
Figura 7 – O Estado do Espírito Santo e suas meso e microrregiões.	44
Figura 8 – Série temporal do agregado de crédito do Banestes no ES	46
Figura 9 – Série temporal do agregado de crédito do Banestes por mesorregião do ES	46
Figura 10 – Série temporal do agregado de crédito do Banestes por microrregião do ES	47
Figura 11 – Verbetes no agregado do ES	47
Figura 12 – Verbetes por mesorregião do ES	48
Figura 13 – Previsões base	48
Figura 14 – Esquema de modelagem de previsões contínuas	49

LISTA DE QUADROS

Quadro 1 – Variáveis do dataset	43
---	----

LISTA DE TABELAS

Tabela 1 – Estrutura do dataset	43
Tabela 2 – Microrregiões por mesorregião	43
Tabela 3 – Municípios por microrregião	44
Tabela 4 – Contagem de únicos no dataset ESTBAN	45
Tabela 5 – Resultados do teste de Ljung-box para as previsões base ($\alpha = 0.05$)	48
Tabela 6 – Conjunto de dados para predição dos modelos de ML	50
Tabela 7 – Resultados Estban: Tempo de processamento dos métodos de ML (em horas)	51
Tabela 8 – Resultados Estban: Acurácia dos métodos analíticos de reconciliação	52
Tabela 9 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estraté- gia rolling forecast.	52
Tabela 10 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estraté- gia fitted base.	52
Tabela 11 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estraté- gia refit.	53
Tabela 12 – Resultados Tourism: Tempo de processamento dos métodos de ML (em horas)	53
Tabela 13 – Resultados Tourism: Acurácia dos métodos analíticos de reconciliação . . .	53
Tabela 14 – Resultados Tourism: Acurácia dos métodos de ML de reconciliação. Estraté- gia rolling forecast.	54
Tabela 15 – Resultados Tourism: Acurácia dos métodos de ML de reconciliação. Estraté- gia fitted base.	54
Tabela 16 – Intervalos de hiperparâmetros para {xgboost}	69
Tabela 17 – Intervalos de hiperparâmetros para {lightgbm}	69
Tabela 18 – Intervalos de hiperparâmetros para {ranger}	69
Tabela 19 – Intervalos de hiperparâmetros para {e1071} (svm)	70
Tabela 20 – Intervalos de hiperparâmetros para {glmnet}	70

LISTA DE ABREVIATURAS E SIGLAS

MinT	<i>Minimum Trace</i>
MCRL	Modelo Clássico de Regressão Linear
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados
MQGF	Mínimos Quadrados Generalizados Factíveis
ANN	<i>Artificial Neural Network</i>
SVR	<i>Support Vector Regression</i>
SFN	Sistema Financeiro Nacional
Favar	<i>Factor Augmented Vector Autoregression</i>
Lasso	<i>Least Absolute Shrinkage and Selection Operator</i>

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
\mathbf{S}	Matriz de soma
\mathbf{G}	Matriz de reconciliação
$\{\dots\}$	Conjunto
$ \{\dots\} $	Cardinalidade de um conjunto

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Previsão de saldos de crédito de instituições financeiras	21
1.2	Previsão de séries temporais hierárquicas	22
1.3	Objetivos	24
1.4	Outline da dissertação	25
2	REVISÃO DE LITERATURA	25
2.1	Previsão de saldos de crédito de instituições financeiras	25
2.2	Previsão de séries temporais hierárquicas e agrupadas	26
2.2.1	Abordagens de nível único	26
2.2.2	Reconciliação ótima de séries temporais hierárquicas	27
3	MÉTODOS PARA RECONCILIAÇÃO DE SÉRIES TEMPORAIS HI- ERÁRQUICAS E AGRUPADAS	29
3.1	Notação algébrica	29
3.2	Abordagens top-down, bottom-up e middle-out	31
3.3	Coerência e reconciliação	35
3.4	Métodos analíticos de reconciliação ótima	36
3.5	Métodos de reconciliação ótima baseados em aprendizado de máquina	36
3.5.1	Elastic Net	37
3.5.2	Gradient Boosting	38
3.5.3	Random Forest	38
3.5.4	Support Vector Machines	39
4	METODOLOGIA	41
4.1	Dados e variáveis	42
4.2	Análise exploratória dos dados	43
4.3	Previsões base	45
4.4	Design do experimento	48
4.5	Otimização de hiperparâmetros	51
5	RESULTADOS	51
5.1	ESTBAN	51
5.2	TOURISM	51
	REFERÊNCIAS	55

APÊNDICES	61
APÊNDICE A – DEMONSTRAÇÕES	63
ANEXOS	67
ANEXO A – CONJUNTO DE HIPERPARÂMETROS	69

1 INTRODUÇÃO

1.1 Previsão de saldos de crédito de instituições financeiras

Embora no séc. XX ainda houvesse espaço para uma gestão guiada apenas por instinto (WALLANDER, 1999), atualmente é impensável um banco não realizar previsões de seus resultados e comunicar suas expectativas ao mercado. Nesse documento, ou *guidance*, a projeção da carteira de crédito — o total de empréstimos e financiamentos, dentre outros itens — é frequentemente a primeira informação fornecida, uma vez que é um dos principais elementos 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.

Ainda que não existam 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, ela pode ser acusada de fraude¹ ou, ao menos, gestão temerária² — ambos caracterizados como crime contra o Sistema Financeiro Nacional.

Não são raros os casos em que bancos manipulam seus demonstrativos para se apresentarem mais saudáveis do que realmente o são para atrair clientes e investidores. Em 2001, o Banco Santos era classificado como o oitavo maior banco brasileiro. Em 2004 se encontrava sob intervenção do Bacen: ao mesmo tempo em que continuava a expandir seus negócios, o banco escondia um rombo de cerca de R\$ 2,2 bi em créditos que já haviam sido liquidado mas eram mantidos no balanço (MOURA, 2007). De forma semelhante, em 2009, o Banco PanAmericano

¹ Art. 3º: Divulgar informação falsa ou prejudicialmente incompleta sobre instituição financeira. Pena: Reclusão, de 2 (dois) a 6 (seis) anos, e multa. Art. 4º: Gerir fraudulentamente instituição financeira. Pena: Reclusão, de 3 (três) a 12 (doze) anos, e multa (BRASIL, 1986).

² Art. 4º, parágrafo único: Se a gestão é temerária: Pena: Reclusão, de 2 (dois) a 8 (oito) anos, e multa (BRASIL, 1986).

publicou em suas demonstrações contábeis carteira de crédito de R\$ 9,9 bi. Entretanto, cerca de 25% já havia sido liquidado e era mantido no balanço artificialmente como objetivo de evitar a redução da atratividade do banco no mercado (COELHO et al., 2015).

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. No entanto, há também motivações estratégicas para essa atividade. Beccalli et al. (2015) mostraram que, em uma amostra de 55 bancos europeus, a utilização de *guidance* está associada a um aumento de 15% na probabilidade do banco atingir ou superar as expectativas de mercado. Isso, por sua vez, está associado a um incremento de até 5% no retorno por ação em relação aos bancos que não alcançaram ou superaram as expectativas.

1.2 Previsão de séries temporais hierárquicas

No que concerne a elaboração dessas previsões, os bancos, assim como em diversas outras indústrias, se enquadram em uma categoria de negócio que requerem previsões de múltiplas séries temporais correlacionadas que são resultados de agregação. Por exemplo, o total de empréstimos de uma instituição financeira corresponde ao agregado dos empréstimos de cada uma de suas agências; o total de vendas de uma rede nacional de farmácias corresponde ao agregado de vendas de suas unidades em cada estado; o total da produção de uma petrolífrica multinacional corresponde ao total produzido em cada país por cada uma de suas plataformas. A essas estruturas naturais de agregação dá-se o nome de *séries temporais hierárquicas*.

Pode-se realizar previsões individualmente para todos os níveis da estrutura. No caso de uma instituição financeira, isso significa realizar previsões, por exemplo, para cada agência, para o agregado de cada região (ou superintendência regional) e para o agregado da instituição. Infelizmente, não há qualquer razão, exceto para métodos de previsão muito simples, para que essas previsões sejam *coerentes* (i.e. que a soma das previsões individuais seja igual à previsão do agregado). Além disso, realizar as previsões individualmente ignoraria os relacionamentos existentes entre as séries temporais na estrutura. Para fazer com que essas previsões se tornem coerentes entre si é que foram desenvolvidos os chamados métodos de *reconciliação*, sendo os mais simples o *top-down*, *bottom-up* e uma combinação das duas, a *middle-out*.

A prática usual em *budgeting*³, principalmente para empresas com muitas filiais, é a *top-down*, ou seja, realizar previsões para o total agregado e então distribuí-las para cada unidade seguindo alguma lógica proporcional. No caso dos bancos de varejo, com muitas agências espalhadas pelo território, especialmente em um país grande como o Brasil, esse método pode ser muito prático.

Esse é o caso do Banestes. Com 134 agências distribuídas pelos 78 municípios capixabas,

³ O orçamento é um documento no qual é definido o planejamento financeiro de um empresa, geralmente para o ano seguinte, estabelecendo metas e objetivos. Nele são projetadas as expectativas da empresa e é base de comparação para saber como os resultados estão se desviando da performance esperada.

realizar o *budgeting* para R\$ 5,5 bi de faturamento⁴ não é uma tarefa trivial. Além de uma estrutura hierárquica de alta dimensionalidade por conta da quantidade de agências, se tratando de um banco múltiplo⁵ que opera com diversas carteiras, as n modalidades de crédito⁶ expandem a estrutura para um total de $n \times 96$ séries temporais a serem estimadas.

Dada tal complexidade, a abordagem *top-down* se coloca como uma opção viável em termos de tempo de processamento e análise. No entanto, conforme descemos na hierarquia, menos precisa ela se torna e, além disso, as características individuais das séries temporais do menor nível hierárquicos são ignoradas. Isso significa que, se no agregado a previsão para uma carteira de crédito for de crescimento de 10%, todas as 96 agências devem seguir a mesma estimativa, divergindo apenas na proporção de participação de cada uma no total.

Tomando o caminho inverso, a abordagem *bottom-up* consiste em realizar previsões para cada série temporal individualmente e, então, agregá-las para obter a previsão para o total. Essa abordagem pode ser mais precisa, pois leva em consideração as características individuais de cada série temporal do nível mais desagregado. No entanto, ela é mais custosa em termos de tempo de processamento e análise. Nesse sentido, cabe ao analista avaliar o *trade-off* entre os ganhos de precisão percebidos com a geração de previsões individuais e a economia de tempo e processamento em realizar o contrário (GROSS; SOHL, 1990).

Além disso, ambas são abordagens de nível único, isto é, são realizadas as previsões para um único nível e então os demais níveis são obtidos agregando ou desagregando. O problema com esses tipos de abordagem é que elas utilizam informação incompleta (HYNDMAN; ATHANASOPOULOS, 2021). Por exemplo, suponha-se que se escolha estimar modelos para cada uma das 96 agências e agregá-las (*bottom-up*). Nesse caso, ignora-se a influência que os níveis mais agregados — aqui a carteira de crédito da região ou de todo o estado — pode ter na estimação do saldo de crédito de cada agência. Por outro lado, se escolher estimar modelos para os níveis mais agregados (*top-down*), ignora-se a informação individual de cada agência.

Uma terceira possibilidade é a *reconciliação ótima*. Ela é uma abordagem que busca resolver esse problema e consiste em realizar previsões para todos os níveis hierárquicos e, então, estimar um modelo para reescrever as previsões do nível mais desagregado como uma combinação linear de todos os elementos da hierarquia. Obtidas as novas previsões no menor nível, ela são então agregadas, gerando previsões coerentes nos níveis superiores. Dessa forma, a informação de todos os níveis é utilizada na estimação dos modelos e na geração das previsões, ao mesmo tempo em que a variância do erro de previsão é minimizado (HYNDMAN; AHMED et al., 2011).

⁴ Conforme demonstrativos publicados referentes ao exercício de 2022 (BANCO DO ESTADO DO ESPÍRITO SANTO, 2022).

⁵ Para ser classificado como banco múltiplo, a instituição financeira deve operar com, no mínimo, duas carteiras dentre: comercial; investimento ou desenvolvimento; crédito imobiliário; de crédito, financiamento e investimento, e; arrendamento mercantil (CONSELHO MONETÁRIO NACIONAL, 1994).

⁶ Crédito consignado, rural, imobiliário, pessoal, capital de giro, desconto de títulos etc.

Atualmente, os métodos analíticos baseados na minimização do traço da matriz da variância-covariância dos erros, desenvolvidos em [Wickramasuriya, Athanasopoulos e Hyndman \(2019\)](#), são os mais populares na literatura da reconciliação ótima. Esses métodos divergem apenas na forma da qual se dá o relacionamento entre os diferentes elementos da hierarquia: se os erros de previsão são descorrelacionados e equivariantes ao longo de toda estrutura (MQO), o que é impossível em séries temporais hierárquicas; se os erros são descorrelacionados e equivariantes apenas dentro do mesmo nível hierárquico (MQP estrutural); se os erros são descorrelacionados mas ponderados pela variância da série (MQP); ou se são correlacionados e variantes ao longo de toda a estrutura (estimadores *MinT Sample* e *MinT Shrink*).

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, justificando a pesquisa e atenção ao tema. Nesse sentido, trabalhos como [Spiliotis et al. \(2021\)](#) desenvolvem métodos de reconciliação ótima de séries temporais utilizando algoritmos de *machine learning*, especificamente o *gradient boosting*, particularmente o *XGBoost*, e *Random Forest*, obtendo resultados superiores aos métodos analíticos tradicionais, especialmente quando as séries não possuem características semelhantes.

Tomando como ponto de partida as conclusões de [Spiliotis et al. \(2021\)](#), este trabalho busca continuar e estender a exploração de métodos de *machine learning* para a tarefa de reconciliação ótima, especificamente os métodos de regressão regularizada lasso, ridge e *elastic net*, e *support vector machine* (SVM), além de outra implementação de *gradient boosting*, o *LightGBM*, verificando se sua performance se mantém superior à dos métodos analíticos em um contexto de séries temporais financeiras de alta dimensionalidade.

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*.

Os objetivos específicos são:

1. Aplicar os métodos *elastic net*, SVM, *gradient boosting* e *random forest* para reconciliação ótima de previsões pontuais de séries temporais hierárquicas e agrupadas e atestar sua validade em dados financeiros de alta dimensionalidade;
2. Comparar a performance, em termos de funções de custo e tempo de processamento, do método *elastic net* nas configurações $\alpha = 0$ (*lasso*), $\alpha = 1$ (*ridge*) e uma combinação de

ambos obtida via reamostragem;

3. Comparar a performance, em termos de funções de custo e tempo de processamento, dos métodos de *machine learning* contra os métodos analíticos *bottom-up* e *MinT Shrink*;

1.4 Outline da dissertação

[Escrever *outline* da dissertação]

2 REVISÃO DE LITERATURA

2.1 Previsão de saldos de crédito de instituições financeiras

A nível macroeconômico, a previsão do agregado de crédito das instituições financeiras é uma preocupação de bancos centrais ao redor do mundo. No Brasil, [Bader, Koyama, Sérgio Mikio e Tsuchida, Marcos Hiroyuki \(2014\)](#) aprimoram o método FAVAR (*Factor Augmented Vector Autoregression*) com uma etapa de análise de correlação canônica para identificar as melhores, em termos de correlação com as variáveis de crédito do SFN, combinações lineares de componentes principais. Esse método, que chamaram de FAVAR canônico, alcançou resultado superior aos FAVAR em 1 e 2 estágios na previsão das variáveis de crédito utilizadas, que foram: a concessão de crédito total com recursos livres, o saldo da carteira de crédito total com recursos livres, o saldo da carteira de crédito total com recursos direcionados, a taxa de inadimplência da carteira de crédito total com recursos livres e a taxa média de juros das operações de crédito total com recursos Livre. O trabalho abordou apenas o nível mais agregado, no total do SFN.

[Çolak et al. \(2019\)](#) produzem uma série de indicadores para monitoramento dos períodos de expansão e desaceleração moderada ou excessiva de crédito no setor bancário turco. Os autores utilizam séries filtradas do agregados de crédito, crédito comercial, crédito direto ao consumidor, crédito imobiliário e financiamento de veículos, além de diversos setores da economia, como agricultura, manufatura, construção, comércio, dentre outros, para prever os ciclos de crédito no sistema bancário turco. Os autores concluem que os indicadores com maior poder de explicação para as variáveis macroeconômicas são a taxa de crescimento real do crédito e a taxa de resposta ao impulso do crédito.

Já para níveis abaixo do agregado de crédito (i.e. o total de crédito da economia), poucos trabalhos foram encontrados. Tangenciando o tema da previsão de saldos de crédito, outros tópicos da economia bancária foram objeto de estudo para previsão de séries temporais. [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. Apesar da extensa revisão de literatura realizada pelos autores, nenhum dos trabalhos

citados combinaram estruturas hierárquicas com *machine learning*.

Gorodetskaya, Gobareva e Koroteev (2021) propõem o que chamaram de “uma metodologia universal” para aplicação automática de *machine learning* na previsão de séries temporais do setor bancário, que poderia ser aplicada em qualquer tipo de problema. A metodologia consiste em obter preditores a partir da própria variável defasada, das estatísticas básicas da variável (máximo, mínimo, média, variância etc.), e de anomalias periódicas detectadas, e selecioná-las pela medida de importância. Realizaram uma revisão de literatura recente sobre o assunto e apresentaram sua abordagem para o problema da previsão da demanda por moeda em caixas eletrônicos.

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

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* 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; 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 $j = 1, \dots, m$, consiste em tomar a média das proporções da série desagregada $y_{j,t}$ em relação ao agregado $y_{1,t}$:

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_{1,t}} \quad (1)$$

Já a proporção das médias históricas consiste em tomar a proporção das médias das séries

desagregadas em relação à média do agregado⁷.

$$p_j = \frac{\sum_{t=1}^T \frac{y_{j,t}}{T}}{\sum_{t=1}^T \frac{y_{1,t}}{T}} \quad (2)$$

Athanasopoulos, Ahmed e Hyndman (2009) desenvolvem o método proporções de previsão, que consiste em um método *top-down* em que os pesos são calculados a partir das proporções das previsões fora da amostra ao invés do passado. A vantagem do método é que os pesos estarão os mais próximos das características mais recentes da série, ao invés de serem baseados em dados históricos. A desvantagem é que se deve realizar previsões para toda a hierarquia, perdendo o ganho de agilidade dos demais métodos *top-down*.

$$p_j = \prod_{i=0}^{K-1} \frac{\hat{Y}_{j,t+h}^i}{\sum \hat{Y}_{j,t+h}^{(i+1)}}$$

Li et al. (2016) compararam dois algoritmos de *machine learning* para previsão da produção de energia solar no estado da Flórida/EUA: ANN (*Artificial Neural Network*) e SVR (*Support Vector Regression*). 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 a partir 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, eles não utilizam esses algoritmos para a reconciliação ótima, caracterizando a abordagem do trabalho ainda como nível único.

2.2.2 Reconciliação ótima de séries temporais hierárquicas

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 método⁸. 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 de uma matriz de pesos 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 individuais, que tem

⁷ Isso é equivalente a tomar a proporção direta entre os somatórios das séries. Note que, pelas propriedades do operador de somatório, $\sum_{t=1}^T \frac{y_t}{T} = \frac{y_1}{T} + \dots + \frac{y_T}{T} = \frac{y_1 + \dots + y_T}{T} = \frac{\sum_{t=1}^T y_t}{T}$. Então, a equação Equação 2 pode ser simplificada para $p_j = \frac{\sum_{t=1}^T y_{j,t}}{\sum_{t=1}^T y_{1,t}}$.

⁸ Uma revisão dessa literatura pode ser encontrada em Athanasopoulos, Ahmed e Hyndman (2009).

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 de pesos por mínimos quadrados ordinários (MQO).

A estimação por esse método resulta numa reconciliação ótima que depende apenas da estrutura hierárquica e independe da variância e covariância das previsões individuais — o que não é uma conclusão satisfatória, uma vez que é impossível que uma série hierárquica seja descorrelatada, pois um nível superior é, necessariamente, o agregado do inferior.

Hyndman, Lee e Wang (2016) tentam aperfeiçoar o método usando as variâncias das previsões individuais 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 individuais 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{e}_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{y}_{t+h} , ao invés dos erros das previsões individuais \hat{y}_{t+h} . Dado que isso implica na minimização da variância de \tilde{e}_{t+h} , ou seja, na minimização do somatório da diagonal, o traço, da matriz de variância-covariância de \tilde{e}_{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 individuais.

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, os autores estendem essa interpretação geométrica para o contexto probabilístico, fornecendo 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.

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. Para o conjunto de dados utilizados, os autores 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*.

3 MÉTODOS PARA RECONCILIAÇÃO DE SÉRIES TEMPO-RAIS HIERÁRQUICAS E AGRUPADAS

3.1 Notação algébrica

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 de um país fictício com três estados, cada um com dois municípios. Essa série pode ser desagregada por estado que, por sua vez, pode ser desagregada por município (Figura 1).

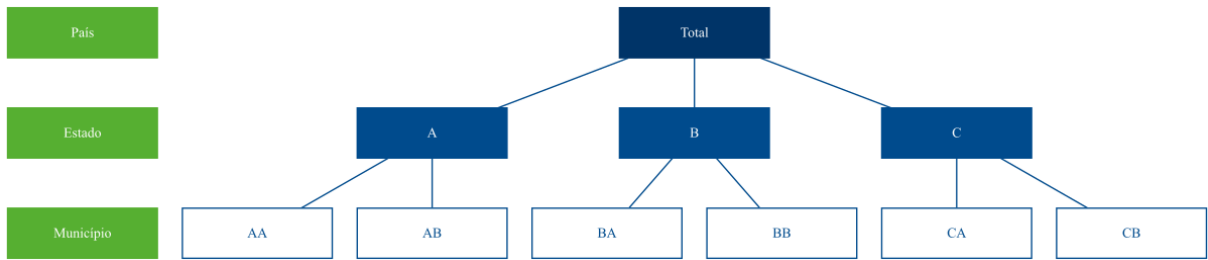


Figura 1 – Séries Hierárquicas

Essa estrutura pode ser representada através de equações para qualquer nível de agregação. Dessa forma, o agregado nacional pode ser descrito pelos agregados dos estados, Equação (3), ou como o agregado dos municípios, Equação (4). Já o agregado para o estado A é representado pela Equação (5).

$$y_t = y_{A,t} + y_{B,t} + y_{C,t} \quad (3)$$

$$y_t = y_{AA,t} + y_{AB,t} + y_{BA,t} + y_{BB,t} + y_{CA,t} + y_{CB,t} \quad (4)$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} \quad (5)$$

Alternativamente, podemos descrever a estrutura completa de forma matricial:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \end{bmatrix}_{m \times 1} \quad (6)$$

Uma outra forma de desagregarmos o PIB é por atividade econômica — agricultura, indústrias extrativas, indústria de transformação, eletricidade e gás, construção etc. Essa estrutura não pode ser desagregada naturalmente de uma única maneira, 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.

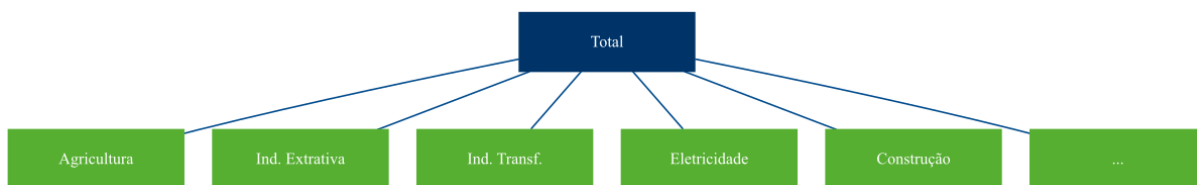


Figura 2 – Séries Agrupadas

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 estados⁹, a adição da estrutura agrupada pode ocorrer tanto acima (Figura 3) quanto abaixo (Figura 4) da hierárquica.

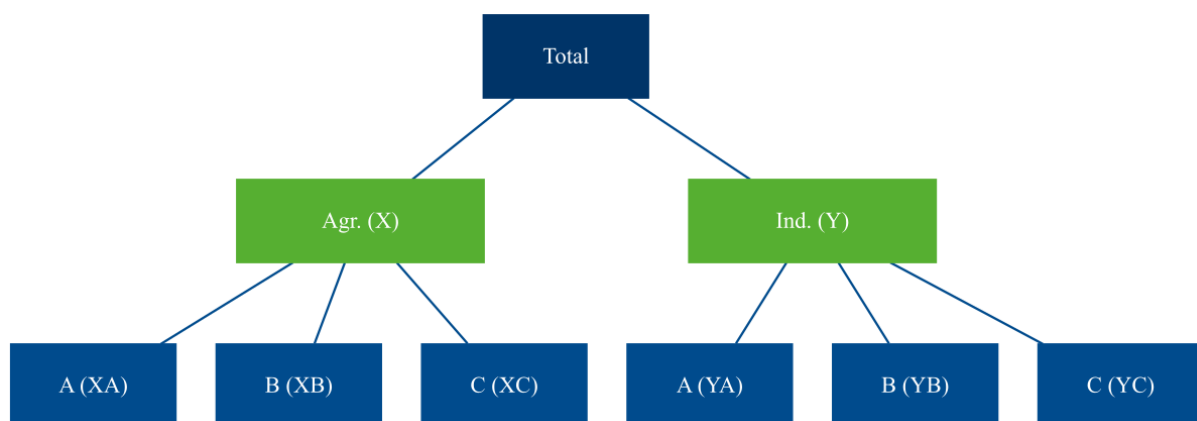


Figura 3 – Séries Hierárquicas Agrupadas (a)

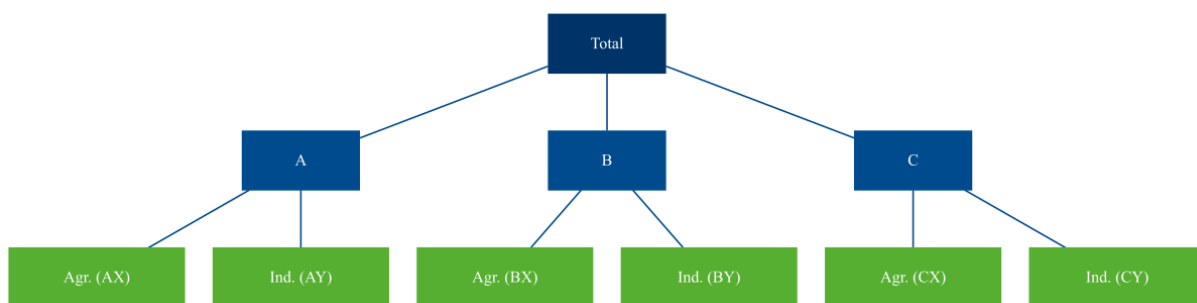


Figura 4 – Séries Hierárquicas Agrupadas (b)

Na notação matricial, a estrutura da Figura 4 é representada como abaixo. Formalmente, o primeiro membro da igualdade é composto pelo vetor \mathbf{y}_t n -dimensional com todas as observações no tempo t para todos os níveis da hierarquia. O segundo membro é composto pela matriz

⁹ Essa estrutura é única no sentido que o somatório dos municípios totaliza o estado, mas não se pode somar estados para totalizar um município. Outro exemplo de estrutura hierárquica é a série de vendas de uma empresa: pode-se agregar as vendas de cada filial para obter o total, mas não o contrário.

de soma \mathbf{S} de dimensão $n \times m$ que define as equações para todo nível de agregação, e pelo vetor \mathbf{b}_t composta pelas séries no nível mais desagregado.

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t \quad (7)$$

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{CX,t} \\ y_{CY,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{CX,t} \\ y_{CY,t} \end{bmatrix}_{m \times 1} \quad (8)$$

3.2 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, representado no topo da hierarquia por *Total* (Equação 9), e então distribuir os valores previstos proporcionalmente entre os estados e municípios.

$$\hat{\mathbf{y}}_{T+h|T} = E[\mathbf{y}_{T+h} | \Omega_T] \quad (9)$$

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 o município AA são dadas pela Equação 10.

$$\tilde{\mathbf{y}}_{AA,T+h|T} = p_1 \hat{\mathbf{y}}_{T+h|T} \quad (10)$$

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

$$\mathbf{G} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \quad (11)$$

\mathbf{G} é uma matriz $m \times n$ que multiplica o vetor $\hat{\mathbf{y}}_{T+h|T}$ que chamamos de *previsões base*, isto é, as previsões individuais para todos os níveis de agregação. A equação para a abordagem *top-down* será, então:

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h|T} \quad (12)$$

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

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \begin{bmatrix} \hat{y}_{T+h|T} \\ \hat{y}_{A,T+h|T} \\ \hat{y}_{B,T+h|T} \\ \hat{y}_{C,T+h|T} \\ \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{CA,T+h|T} \\ \hat{y}_{CB,T+h|T} \end{bmatrix}_{n \times 1} \quad (13)$$

Multiplicando as matrizes \mathbf{G} e $\hat{\mathbf{y}}$ obtemos o vetor \mathbf{b}_t^{10} , que contém as previsões para os níveis mais desagregados, agora definidas como proporção do total.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \end{bmatrix}_{m \times 1} \quad (14)$$

¹⁰ Ver Equação 7.

Note que, por se tratar de um método *top-down*, é necessário apenas o primeiro elemento do vetor de previsões base, ou seja, a previsão do nível mais agregado (Equação 14). Sendo essa exatamente uma das vantagens do método *top-down*, na prática, podemos anular os demais elementos de $\hat{\mathbf{y}}$.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \begin{bmatrix} \hat{y}_{T+h|T} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{n \times 1} \quad (15)$$

Substituindo a matriz \mathbf{S} , temos as equações que definem as previsões reconciliadas.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \end{bmatrix}_{m \times 1} \quad (16)$$

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 \mathbf{G} .

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{m \times n} \quad (17)$$

Portanto, \mathbf{G} define a abordagem — se *top-down* ou *bottom-up* —, e \mathbf{S} define a maneira da qual as previsões são somadas para formar as equações de previsão para cada elemento da estrutura. Portanto, chamamos \mathbf{G} de matriz de reconciliação.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{CA,T+h|T} \\ \hat{y}_{CB,T+h|T} \end{bmatrix}_{m \times 1} \quad (18)$$

Quando m — a quantidade de elementos do nível mais desagregado — é muito grande, tornando muito custoso obter $\hat{\mathbf{y}}_t$, e não se deseja uma abordagem estritamente *top-down*, pode-se combinar as duas formas:

$$\mathbf{G} = \begin{bmatrix} 0 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_6 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \quad (19)$$

Esse método é chamado de *middle-out*. Nele, o vetor \mathbf{b}_t é reescrito como proporções de algum nível intermediário arbitrariamente 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.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} p_1 \hat{y}_{A,T+h|T} \\ p_2 \hat{y}_{A,T+h|T} \\ p_3 \hat{y}_{A,T+h|T} \\ p_4 \hat{y}_{B,T+h|T} \\ p_5 \hat{y}_{B,T+h|T} \\ p_6 \hat{y}_{B,T+h|T} \end{bmatrix}_{m \times 1} \quad (20)$$

3.3 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{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 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 estocástico.

Os métodos de gerar previsões coerentes (\tilde{y}_t) a partir de previsões base (\hat{y}_t) são chamados de métodos de *reconciliação*. Os métodos de reconciliação tradicionais apresentados, *top-down*, *bottom-up* e *middle-out*, utilizam informação limitada. No método *top-down*, utiliza-se apenas informações do nível mais agregado — por isso, apenas a primeira coluna em (Equação 11) é diferente de zero. Já na abordagem *bottom-up*, utiliza-se apenas as informações dos níveis mais desagregados, o que resulta na submatriz identidade $m \times m$ na Equação 17, enquanto as colunas que representam os níveis mais agregados são nulas. Por fim, a abordagem *middle-out* não utiliza a mesma, mas utiliza a mesma quantidade de informação que a *top-down* (Equação 19).

Alternativamente, podemos pensar numa matriz 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*.

3.4 Métodos analíticos de reconciliação ótima

Os métodos analíticos de reconciliação ótima são aqueles que estimam a matriz de reconciliação, \mathbf{G} , através de regressão linear. Isso resulta na redefinição das previsões do nível mais desagregado como uma combinação linear¹¹ das previsões de todos os elementos de todos os níveis da hierarquia, utilizando, assim, toda a informação disponível.

O estado-da-arte para esse tipo de método é o *MinT* (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2019). Nele, o objetivo é minimizar o erro das previsões reconciliadas:

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t} \quad (21)$$

Essa equação pode ser reescrita como $\tilde{\mathbf{e}}_t = \mathbf{S}\mathbf{G}\hat{\mathbf{e}}_t$, que tem variância dada por¹²

$$\text{Var}[\tilde{\mathbf{e}}] = \mathbf{S}\mathbf{G}\hat{\mathbf{W}}_{t+h|t}\mathbf{G}'\mathbf{S}' \quad (22)$$

em que $\hat{\mathbf{W}}_{t+h|t}$ é a matriz de variância-covariância dos erros de previsão base.

A abordagem consiste então em se obter um valor de \mathbf{G} que minimize o traço de $\text{Var}[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}]$. Isso resultaria no melhor (variância mínima) estimador linear não viesado¹³ (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2019).

$$\mathbf{G} = (\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger \quad (23)$$

em que $\hat{\mathbf{W}}_{t+h|t}^\dagger$ é a inversa generalizada de Moore-Penrose para $\hat{\mathbf{W}}_{t+h|t}$ ¹⁴. Essa formulação corresponde a um problema de regressão por mínimos quadrados generalizados, considerando \mathbf{S} como a matriz de preditores e \mathbf{G} os coeficientes a serem estimados. Consequentemente, as previsões ótimas reconciliadas são dadas por

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}(\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger\hat{\mathbf{y}}_{t+h|t} \quad (24)$$

3.5 Métodos de reconciliação ótima baseados em aprendizado de máquina

Embora os métodos de ML também sejam utilizados no contexto de previsão de séries temporais, é importante ressaltar que este não é caso. As previsões de séries temporais são re-

¹¹ Por essa razão, esses métodos são também chamados de métodos de combinação.

¹² Ver demonstração A.3.

¹³ A ausência de viés é garantida pela [condição de ausência de viés em \$\tilde{\mathbf{y}}\$](#) .

¹⁴ A necessidade da inversa generalizada aqui é trivial, uma vez que a inversa regular, do tipo \mathbf{A}^{-1} , requer matriz quadrada, o que não acontece no caso de séries temporais hierárquicas, uma vez que, necessariamente, tem-se $n > m$ ($n = m + \text{número de nós de agregação}$). Além disso, $\hat{\mathbf{W}}$ é posto incompleto (ver demonstração A.4).

alizadas anteriormente, na obtenção das previsões base \hat{y}_t . Nada impede que essas previsões sejam obtidas por um modelo de ML, mas isso não é o foco deste trabalho.

A aplicação dos métodos de ML aqui ocorrem na reconciliação, ou seja, na combinação contemporânea, *cross section*, das previsões base que, por sua vez, podem ter sido obtidas através de qualquer método.

3.5.1 Elastic Net

O *elastic net* (ZOU; HASTIE, 2005) é um método de regressão regularizada que combina as normas L_1 e L_2 , as penalidades do *lasso* e do *ridge*, respectivamente. A função objetivo a ser minimizada é dada por

$$L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1 \quad (25)$$

em que λ_1 e λ_2 são os parâmetros de regularização e β é o vetor de coeficientes a serem estimados. A solução para essa função objetivo é dada por¹⁵

$$\hat{\beta} = \arg \min_{\beta} |\mathbf{y} - \mathbf{X}\beta|^2, \text{ sujeito a } (1 - \alpha)|\beta|_1 + \alpha|\beta|^2 \leq t \quad (26)$$

com $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ e $t \in \mathbb{R}^+$.

A função $(1 - \alpha)|\beta|_1 + \alpha|\beta|^2$ é a penalidade *elastic net*, uma combinação das penalidades *lasso* e *ridge*. O parâmetro α controla a mistura das duas penalidades, incluindo os casos extremos. Note que $\alpha = 0 \implies \lambda_2 = 0$, resultando em uma penalidade exclusivamente *lasso*, enquanto $\alpha = 1 \implies \lambda_1 = 0$, e a penalidade é apenas do tipo *ridge*.

Portanto o *elastic net* é um método de *shrinkage*, uma vez que a penalidade *ridge* reduz o tamanho dos coeficientes, e de *seleção de variáveis*, uma vez que a penalidade *lasso* tende a anular os coeficientes de variáveis irrelevantes. Essas propriedades são desejáveis para a reconciliação de séries temporais, uma vez que a estrutura hierárquica pode conter séries insignificantes para a previsão de outras séries.

Diferentemente dos métodos analíticos estudados, o *elastic net* não possui uma solução fechada. Portanto, é necessário utilizar métodos iterativos para encontrar o valor ótimo de $\hat{\beta}$ e Zou e Hastie (2005) utilizam validação cruzada *k-fold* para encontrar quais os valores de λ_1 e λ_2 que minimizam o resíduo. Nesse sentido, dado a metodologia de processo iterativo envolvendo calibragem de hiperparâmetros e reamostragem, podemos classificar o *elastic net* como um método de *machine learning*.

¹⁵ Sob o valor otimizado ainda é aplicada correção de escala na forma $(1 + \lambda_2)\hat{\beta}$. Ver Zou e Hastie (2005).

3.5.2 Gradient Boosting

Como cada uma das diversas implementações de *gradient boosting* possui sua teoria adjacente, e não é de objetivo deste trabalho detalhar o funcionamento de cada uma delas, trabalharemos apenas sua intuição.

Assim como os métodos de floresta aleatória, os métodos de *gradient boosting* também são métodos de conjuntos de árvores. A diferença se dá na forma como os modelos são treinados. *Gradient boosting* são métodos que combinam as predições de vários modelos fracos para formar um conjunto — *ensemble*, na definição mais usual, ou comitê (*committee*), na definição de [Hastie, Tibshirani e Friedman \(2009\)](#) quando usado para classificação — mais complexo e preciso (forte).

Um estimador fraco é aquele que tem desempenho apenas ligeiramente melhor que o acaso. O propósito do *boosting* é produzir uma sequência de estimadores fracos, cada um deles focado nos erros dos estimadores anteriores. A cada iteração, as observações classificadas incorretamente (no caso de uma tarefa de classificação) ou de maior variância (no caso de uma tarefa de regressão) na iteração anterior têm seu peso aumentado, e vice-versa. Dessa forma, o modelo subsequente formado na próxima iteração é obrigado a se concentrar nas observações onde as iterações anteriores falharam. Isso que significa transformar um conjunto de estimadores fracos em um conjunto forte.

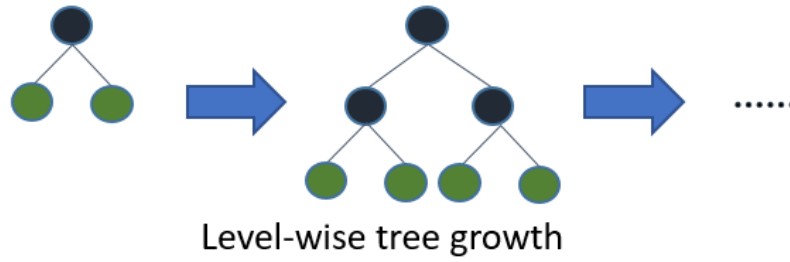
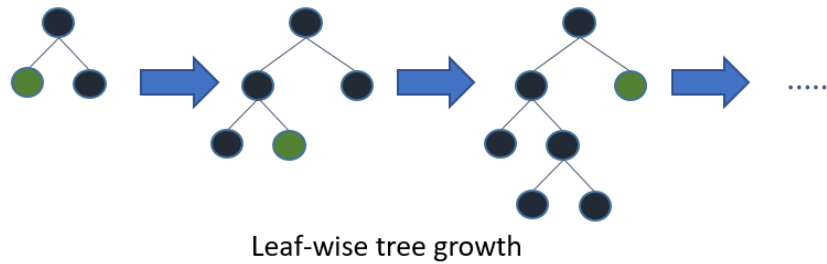
O *gradient boosting* é uma extensão do *boosting* que utiliza o gradiente da função de perda como critério de otimização, de forma que esta se dá na direção em que a função de perda decresce mais rapidamente a cada iteração. Os métodos utilizados neste trabalho são o *XGBoost* ([CHEN; GUESTRIN, 2016](#)) e o *LightGBM* ([KE et al., 2017](#)).

Um das principais diferenças entre os dois métodos é a forma como as árvores são construídas. O *XGBoost* cresce suas árvores de forma *level-wise*, ou seja, cresce todas as folhas do último nível de uma árvore de uma vez, adicionando mais um nível de profundidade completo a cada iteração (Figura 5a). Já o *LightGBM* cresce suas árvores de forma *leaf-wise*, ou seja, cresce uma folha por vez, aprofundando a árvore apenas no nó que resulta na maior variação negativa na função de perda (Figura 5b).

3.5.3 Random Forest

Floresta aleatória é um método de aprendizado de máquina que utiliza conjuntos de árvores de decisão descorrelatadas¹⁶ para classificação ou regressão. O método consiste em treinar várias árvores de decisão em subconjuntos aleatórios dos dados de treinamento e, então, combinar suas predições ([HASTIE; TIBSHIRANI; FRIEDMAN, 2009](#)). A aleatoriedade é introduzida de duas formas: na seleção das observações e na seleção das variáveis preditoras.

¹⁶ As árvores de decisão são descorrelatadas no sentido que, ao contrário os métodos de *boosting trees*, a próxima árvore não é construída com base na iteração anterior (processo de fortalecimento).

(a) *level-wise*(b) *leaf-wise***Figura 5** – Crescimento de árvores em algoritmos de *boosting*.

A intuição para seu algoritmo para regressão é simples e a ideia geral é, para cada árvore de decisão, particionar recursivamente nós de tamanho N em dois nós filhos de forma a aumentar a complexidade do modelo e minimizar a função de custo. Se a próxima partição de um nó resultar em um ou ambos nós filhos de tamanho menor que um mínimo estabelecido via hiperparâmetro, a partição é interrompida (Algoritmo 1).

3.5.4 Support Vector Machines

A intuição do métodos de SVMs é mais facilmente compreendida a partir de uma tarefa de classificação de duas classes. Nesse caso, o objetivo é encontrar um hiperplano (i.e. um sub-espaco de dimensão $n - 1$) que separe as classes de forma que a margem entre o hiperplano e os pontos de cada classe seja a maior possível. Então, dado um conjunto de N pares de observações e suas respectivas classes, $\{(x_i, y_i)\}_{i=1}^N$, em que $x_i \in \mathbb{R}^p$ e $y_i \in \{-1, 1\}$, queremos encontrar o hiperplano definido por

$$\{x : f(x) = x^T \beta + \beta_0 = 0\}$$

sendo $f(x)$ a distância ortogonal entre x e o hiperplano $f(x) = x^T \beta + \beta_0 = 0$, e M a margem entre as classes, definida como a distância ortogonal entre os pontos mais próximos de cada classe e o hiperplano (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Portanto, o problema de otimização é dado por

Algoritmo 1: Floresta aleatória para regressão

T_1, T_2, \dots, T_b é uma floresta aleatória de B árvores de decisão.

$i \in \mathbb{N}$ é a quantidade de observações em um nó.

$n_{min} \in \mathbb{N}$ é o número mínimo de observações em um nó.

for $b = 1 \rightarrow B$ **do**

1. Toma uma amostra de tamanho N dos dados de treino.

while $i < n_{min}$ **do**

 a. Seleciona j variáveis aleatoriamente.

 b. Escolhe a melhor variável para *split* dentre j .

 c. Divide o nó em dois nós filhos.

if Nenhum dos nós filhos tem $i < n_{min}$ **then**

 d. Repete os passos a-c para cada nó filho.

end

end

end

2. Produz um conjunto de árvores $\{T_b\}_1^B$.

3. Realiza a predição para cada árvore T_b em um ponto x e calcula a média das predições:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

$$\begin{aligned} \max_{\beta, \beta_0} \quad & M \\ \text{sujeito a} \quad & y_i(x_i^T \beta + \beta_0) \geq M, \quad i = 1, \dots, N \end{aligned} \tag{27}$$

Quando pode-se encontrar um hiperplano com $y_i f(x_i) > 0 \quad \forall i$, tem-se o caso da construção de uma solução única para um hiperplano entre duas classes perfeitamente separadas pela maior margem possível (Figura 6a).

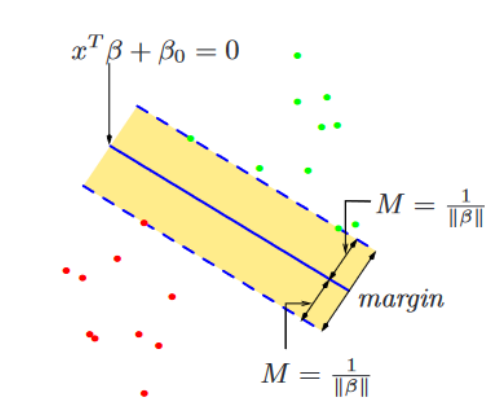
Entretanto, no mundo real, dificilmente um problema de classificação será linearmente separável. Para contornar esse obstáculo, é possível introduzir variáveis de folga $\xi_i \geq 0$ para cada observação (Figura 6b), de forma que a restrição se torne

$$y_i(x_i^T \beta + \beta_0) \geq M(1 - \xi_i), \quad i = 1, \dots, N \tag{28}$$

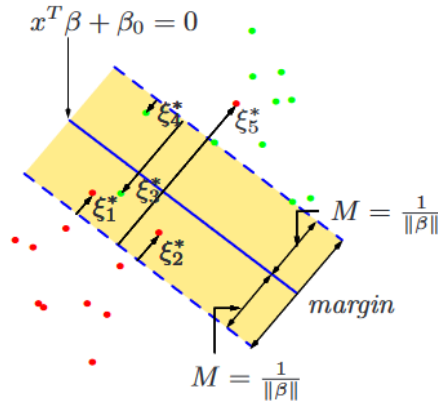
e a função objetivo se torne

$$\min \|\beta\| \quad \text{sujeito a} \quad \left\{ \begin{array}{l} y_i(x_i^T \beta + \beta_0) \geq M(1 - \xi_i) \quad \forall i, \\ \xi_i \geq 0, \sum \xi_i \leq \text{constante} \end{array} \right\} \tag{29}$$

Para estender essa ideia para problemas de regressão, é necessário introduzir uma função de perda ϵ -insensível, que penaliza apenas os erros maiores que ϵ , de forma que os erros menores



(a) Hiperplano entre duas classes perfeitamente separáveis.



(b) Hiperplano entre duas classes não separáveis.

Figura 6 – *Support vector classifiers*. Fonte: [Hastie, Tibshirani e Friedman \(2009\)](#).

são ignorados durante a otimização, assim como as observações localizadas no lado correto do hiperplano são ignoradas no problema de classificação. A função objetivo se torna

$$\min \|\beta\| \quad \text{sujeito a} \quad \left\{ \begin{array}{l} |y_i - x_i^T \beta - \beta_0| \leq \epsilon \quad \forall i, \\ \sum |\beta_j| \leq \text{constante} \end{array} \right\} \quad (30)$$

4 METODOLOGIA

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

O *software* usado foi o R ([R CORE TEAM, 2023](#)). As previsões base e os métodos de reconciliação analíticos foram realizados com o pacote `{fable}` ([O'HARA-WILD; HYNDMAN; WANG, 2023](#)) e suas extensões. Já metodologia de *machine learning* (reamostragem, otimização de hiperparâmetros, treino e predição) foi executada com o pacote `{mlr3}` ([LANG et al., 2023](#)) e suas extensões. As implementações do XGBoost e LightGBM foram realizadas com seus pacotes homônimos `{xgboost}` ([CHEN; HE et al., 2023](#)) e `{lightgbm}` ([SHI et al., 2023](#)),

do Elastic Net com o pacote `{glmnet}` (FRIEDMAN et al., 2023), a do Random Forest com o pacote `{ranger}` (WRIGHT; WAGER; PROBST, 2023) e a do SVM com o pacote `{e1071}` (MEYER et al., 2023).

4.1 Dados e variáveis

Os dados usados nesse trabalho são dados terciários obtidos do *datalake* público Base dos Dados (DAHIS et al., 2022). A fonte primária são os bancos comerciais e múltiplos com carteira comercial que disponibilizam mensalmente os saldos dos principais verbetes do balancete via documento 4500¹⁷ 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 Município¹⁸.

O que compõe os verbetes de crédito, ou seja, os valores das séries temporais a serem trabalhadas, são os saldos de crédito ativo (empréstimos e financiamentos), que correspondem ao principal mais os juros calculados até 59 dias de atraso¹⁹.

Além das estatísticas bancárias, foram obtidas 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 para UF e na raiz do CNPJ. Ademais, foram selecionados os principais verbetes, que são *empréstimos e títulos descontados* e *financiamentos*, e mantidas apenas as agências atualmente em atividade. Quanto ao período, há dados disponíveis desde 1988. Entretanto, utilizaremos os dados a partir de 2003 pois, se tratando de uma hierarquia larga, o custo computacional deve ser levado em conta. Por essa razão, apesar do Banestes contar com 134 agências, foram mantidas apenas as agências com série completa, ou seja, que já estavam em atividade em 2003.

Por fim, as variáveis mantidas no *dataset* foram as descritas no Quadro 1. Dessa forma, temos séries mensais de saldos de créditos de 79 agências bancárias, distribuídas por 62 municípios, a partir de janeiro de 2003, totalizando 37.920 observações.

Esses dados então são organizados de forma hierárquica por estado, mesorregião, microrregião, município e agência bancária; e, de forma agrupada, por verbete. Isso significa que, para cada nó de agregação, são adicionadas mais três observações ao *dataset*: duas para cada verbete e a terceira para a soma de ambos. Com a estrutura hierárquica e agrupada, o conjunto alcança 114 480 observações. Por fim, o *dataset* adquiriu a estrutura apresentada na Tabela 1.

¹⁷ 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.

¹⁸ <https://www4.bcb.gov.br/fis/cosif/estban.asp?frame=1>

¹⁹ Não são consideradas crédito ativo as operações de crédito liquidadas ou que tenham sido transferidas para prejuízo. São transferidas para prejuízo as operações de crédito em atraso há mais 6 meses após sua classificação

Quadro 1 – Variáveis do dataset

Variável	Descrição
ref	Data de referência do relatório ESTBAN
nome mesorregiao	Nome da mesorregião do ES
nome microrregiao	Nome da microrregião do ES
verbete	Descrição da rubrica do balancete
nome	Nome do município
cnpj agencia	CNPJ da agência bancária
saldo	Saldo do verbete

Tabela 1 – Estrutura do dataset

ref	mesorregiao	microrregiao	municipio	agencia	verbete	saldo
2022 dez	sul espirito santense	itapemirim	presidente kennedy	28127603005560	<aggregated>	15.609.729
2022 dez	sul espirito santense	itapemirim	presidente kennedy	<aggregated>	emprestimos e titulos descontados	15.339.010
2022 dez	sul espirito santense	itapemirim	presidente kennedy	<aggregated>	financiamentos	270.719
2022 dez	sul espirito santense	itapemirim	presidente kennedy	<aggregated>	<aggregated>	15.609.729
2022 dez	sul espirito santense	itapemirim	<aggregated>	<aggregated>	emprestimos e titulos descontados	107.975.016
2022 dez	sul espirito santense	itapemirim	<aggregated>	<aggregated>	financiamentos	2.105.064
2022 dez	sul espirito santense	itapemirim	<aggregated>	<aggregated>	<aggregated>	110.080.080
2022 dez	sul espirito santense	<aggregated>	<aggregated>	<aggregated>	emprestimos e titulos descontados	611.867.606
2022 dez	sul espirito santense	<aggregated>	<aggregated>	<aggregated>	financiamentos	11.718.310
2022 dez	sul espirito santense	<aggregated>	<aggregated>	<aggregated>	<aggregated>	623.585.916
2022 dez	<aggregated>	<aggregated>	<aggregated>	<aggregated>	emprestimos e titulos descontados	4.150.513.082
2022 dez	<aggregated>	<aggregated>	<aggregated>	<aggregated>	financiamentos	86.448.098
2022 dez	<aggregated>	<aggregated>	<aggregated>	<aggregated>	<aggregated>	4.236.961.180

4.2 Análise exploratória dos dados

O estado do Espírito Santo está localizado no sudeste brasileiro e é dividido em 78 municípios, que estão agrupados em 4 mesorregiões e 13 microrregiões.

As microrregiões que compõem cada mesorregião são apresentadas na Tabela 2 e na Figura 7b. Já os municípios que compõem cada microrregião são apresentados na Tabela 3. A Figura 7c mostra a quantidade de agências por município.

Tabela 2 – Microrregiões por mesorregião

Mesorregião	Microrregiões
central espirito santense	afonso claudio, guarapari, santa teresa, vitoria
litoral norte espirito santense	linhares, montanha, sao mateus
noroeste espirito santense	barra de sao francisco, colatina, nova venecia
sul espirito santense	alegre, cachoeiro de itapemirim, itapemirim

Na amostra selecionada para este trabalho, com exceção dos municípios de Colatina e Cachoeiro de Itapemirim, que contam com 2 agências cada, todos os demais municípios no interior contam com uma única agência. Já na microrregião da capital Vitória, se encontram municípios com múltiplas unidades. Isso pode afetar a decisão de incluir ou não o nível de município na hierarquia.

de risco em H, que é a mínima (CONSELHO MONETÁRIO NACIONAL, 1999).

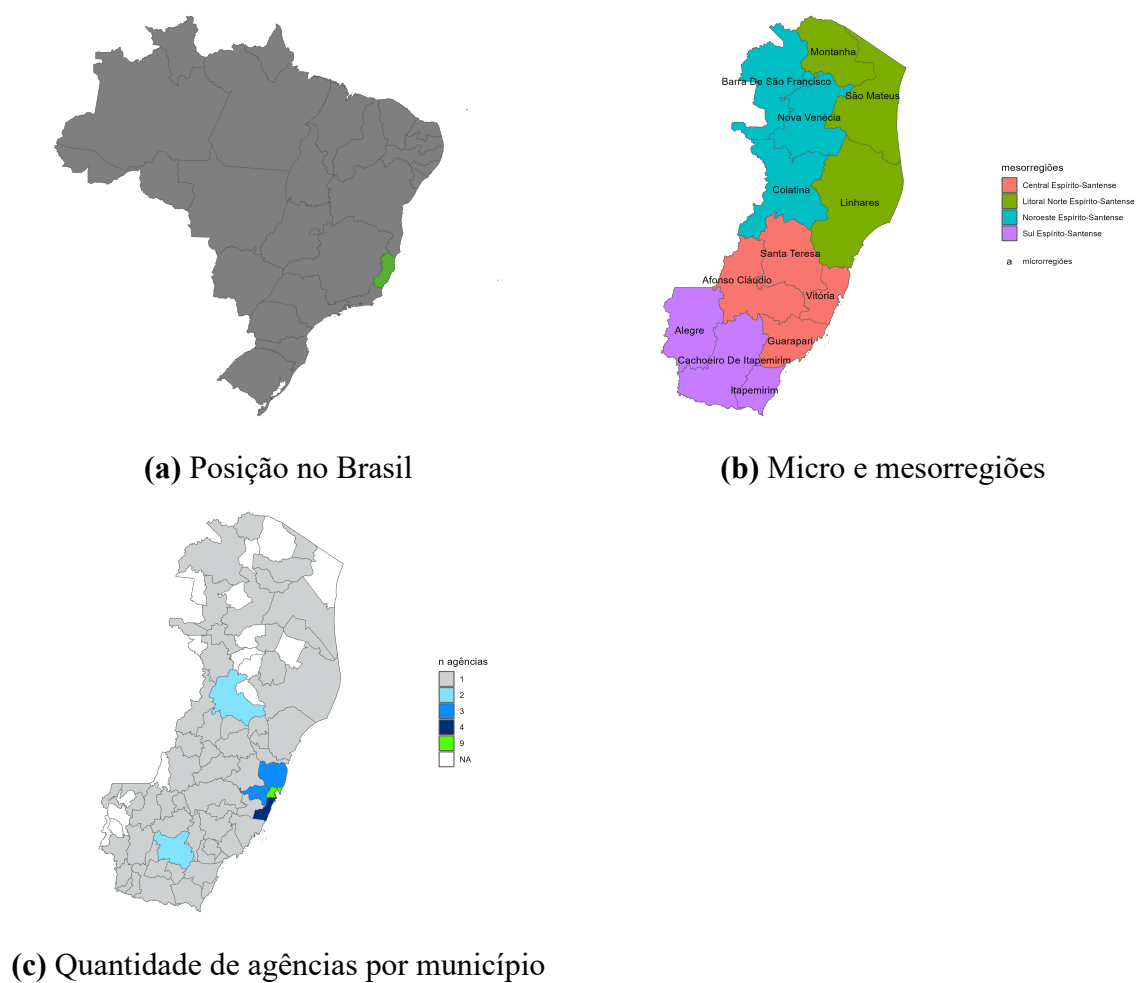


Figura 7 – O Estado do Espírito Santo e suas meso e microrregiões.

Tabela 3 – Municípios por microrregião

Microrregiões	Municípios
afonso claudio	afonso claudio, conceicao do castelo, domingos martins, laranja da terra, marechal floriano, venda nova do imigrante
alegre	alegre, guacui, ibatiba, iuna, muniz freire
barra de sao francisco	barra de sao francisco, ecoporanga, mantenopolis
cachoeiro de itapemirim	apiaca, atilio vivacqua, bom jesus do norte, cachoeiro de itapemirim, castelo, jeronimo monteiro, mimoso do sul, muqui, sao jose do calcado, vargem alta
colatina	baixo guandu, colatina, pancas
guarapari	alfredo chaves, anchieta, guarapari, iconha, piuma, rio novo do sul
itapemirim	itapemirim, marataizes, presidente kennedy
linhares	aracruz, fundao, ibiracu, joao neiva, linhares, rio bananal
montanha	mucurici, pinheiros
nova venecia	aguia branca, boa esperanca, nova venecia, vila valerio
santa teresa	itaguacu, itarana, santa leopoldina, santa maria de jetiba, santa teresa, sao roque do canaa
sao mateus	jaguare, pedro canario, sao mateus
vitoria	cariacica, serra, viana, vila velha, vitoria

O tamanho de uma estrutura hierárquica, em termos de observações, é determinada por seu nível mais desagregado. Assim, sendo 240 meses e 79 agências, a estrutura hierárquica deve contar com $240 \times 79 = 18.960$ observações. Sendo também uma estrutura agrupada por 2 verbetes, a quantidade de observações é multiplicada pela quantidade de níveis transversais, totalizando as 37.920 observações do *dataset* antes da adição dos nós de agregação. Portanto, podemos atestar a completude das séries temporais sem a necessidade de qualquer inspeção adicional.

Tabela 4 – Contagem de únicos no dataset ESTBAN

	Únicos
ref	240
nome mesorregiao	4
nome microrregiao	13
nome	62
cnpj agencia	79
verbeta	2

A série temporal do agregado de crédito no Banestes no Espírito Santo é apresentada na Figura 8. Como frequentemente ocorre em séries temporais de natureza social, a inspeção visual sugere a presença de quebras estruturais na média²⁰. Essas quebras podem afetar a qualidade das previsões base a serem reconciliadas.

Em relação à distribuição, a mesorregião Central Espírito-santense concentra mais crédito do que o somatório das demais regiões (Figura 9), sendo a microrregião de Vitória a responsável por essa concentração (Figura 10). Na ótica dos verbetes, o crédito para financiamentos é uma pequena fração do saldo de empréstimos, independentemente da mesorregião (Figura 11 e Figura 12).

4.3 Previsões base

Uma vez que o foco deste trabalho está no incremento de performance proporcionado pela reconciliação ótima sobre um conjunto de previsões base e não na qualidade destas, elas foram obtidas por meio de métodos básicos para previsão de séries temporais, especificamente o algoritmo de Hyndman-Koehler-Snyder-Groser para suavimento exponencial (ETS) (HYNDMAN; KOEHLER et al., 2002), sem tratamentos adicionais para *outliers*, quebras estruturais ou transformações. Esses métodos são amplamente utilizados na literatura de séries temporais e, portanto, servem como *benchmark* para a avaliação dos métodos de reconciliação ótima.

²⁰ Aqui, destacam-se os períodos de 2008 a 2010, que corresponde à crise financeira global; de 2014 a 2016, que corresponde à crise econômica brasileira, e; a partir de 2020 com a pandemia do Covid-19. Também é sugestivo que a forma como o governo se comporta durante esses eventos pode ter impacto direto no comportamento do crédito.

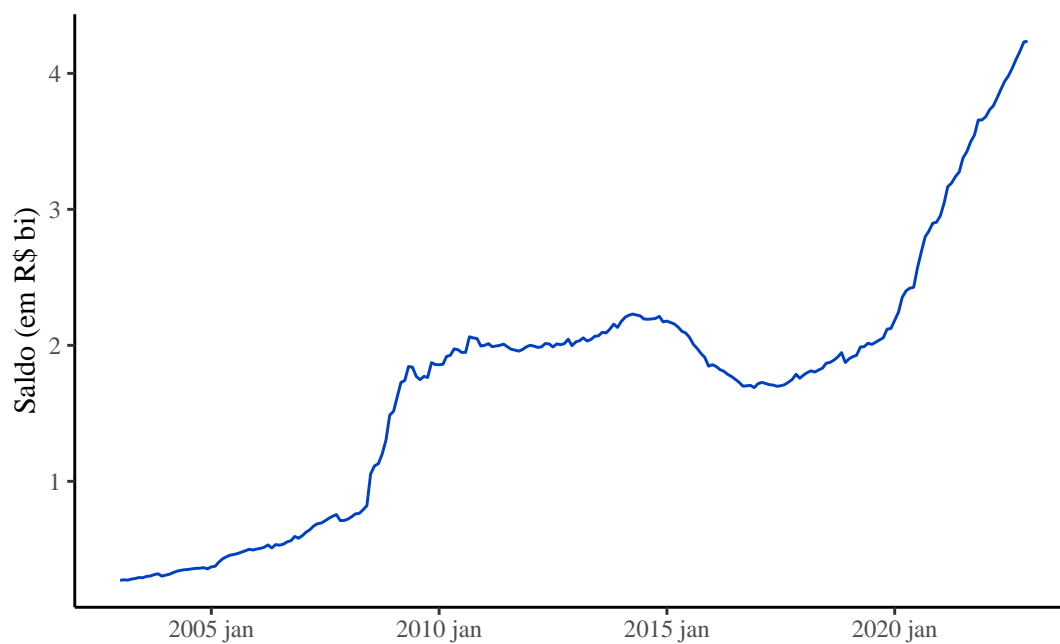


Figura 8 – Série temporal do agregado de crédito do Banestes no ES

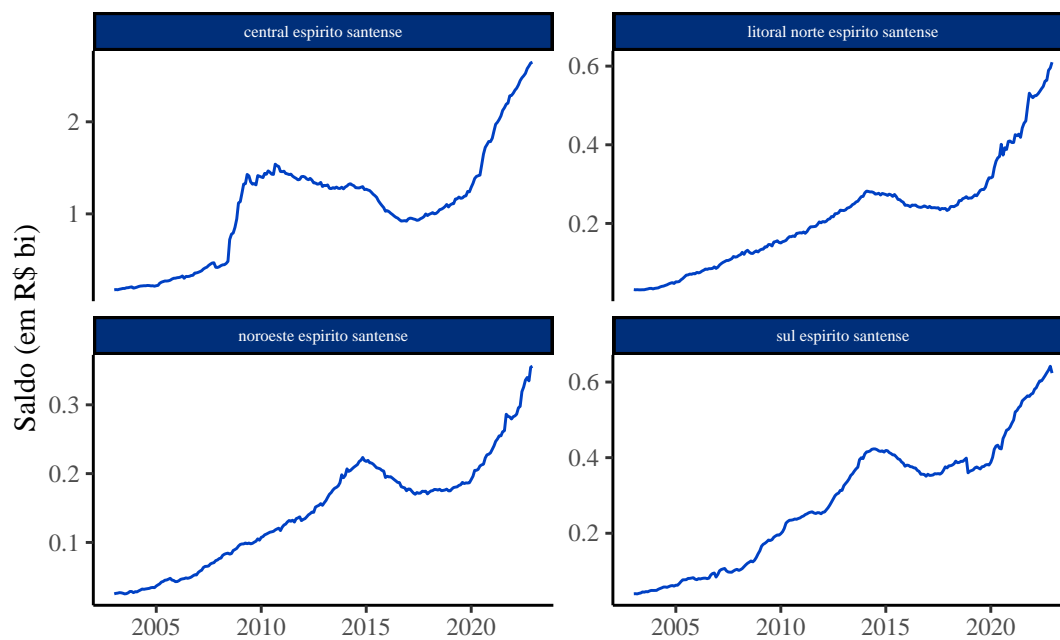


Figura 9 – Série temporal do agregado de crédito do Banestes por mesorregião do ES

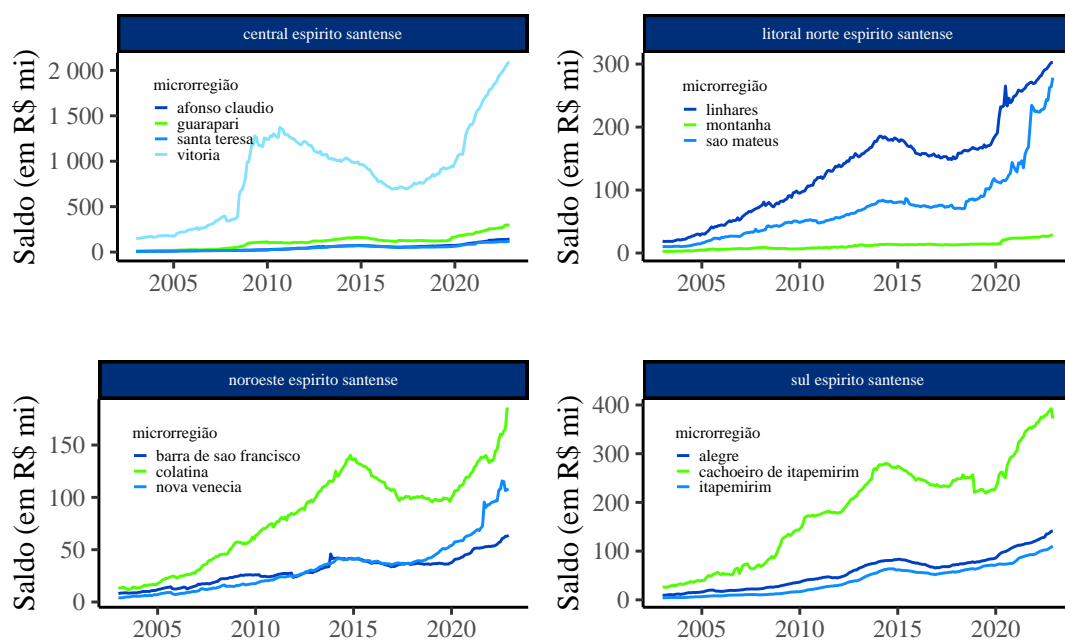


Figura 10 – Série temporal do agregado de crédito do Banestes por microrregião do ES

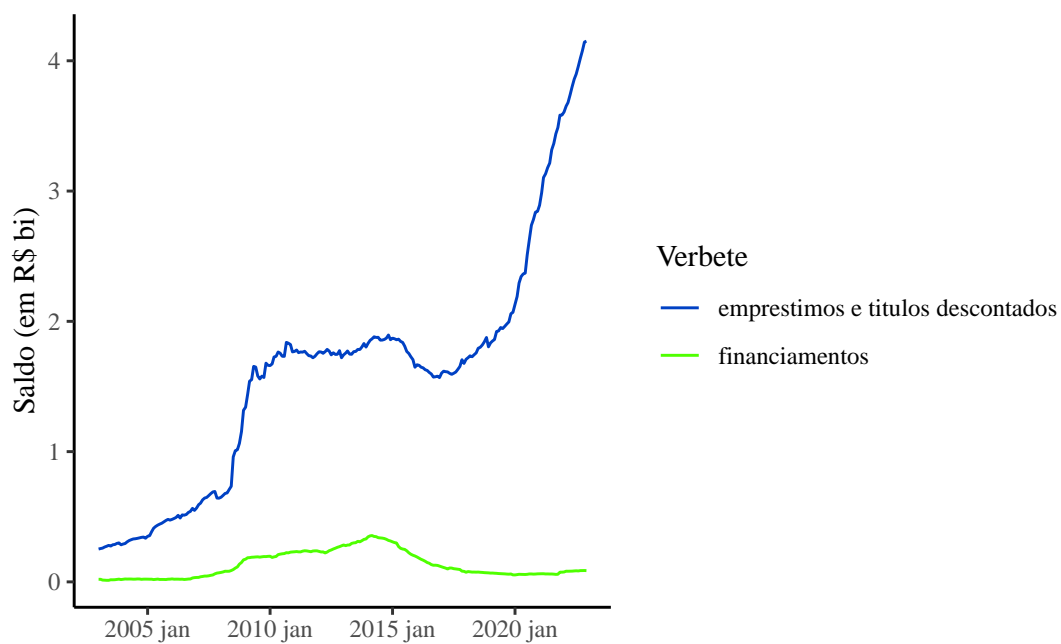


Figura 11 – Verbetes no agregado do ES

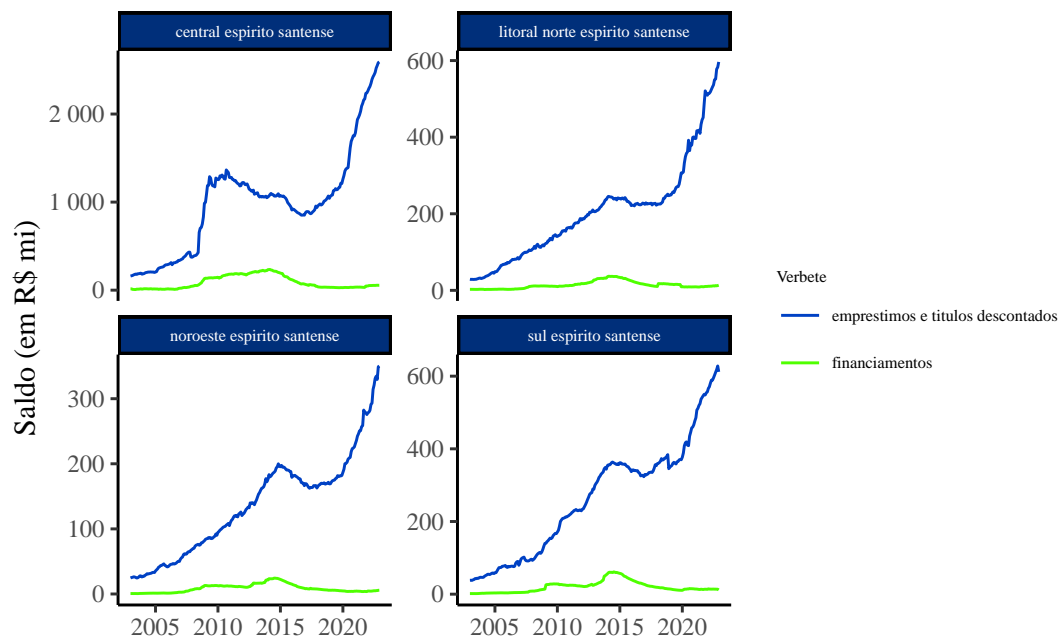


Figura 12 – Verbete por mesorregião do ES

Os modelos foram treinados com dados de 2003 a 2021 e as previsões foram realizadas para o ano de 2022 (Figura 13). Para verificar a qualidade básica do ajuste, foram realizados testes de Ljung-Box para os resíduos das previsões. Os resultados são apresentados na Tabela 5. Considerando nível de significância de $\alpha = 0.05$ e 12 defasagens, cerca de 80% dos modelos não rejeitam a hipótese nula de que os resíduos são ruído branco. Isso sugere que os modelos de previsão base são, em sua maior parte, adequados para o propósito deste trabalho.



Figura 13 – Previsões base

Tabela 5 – Resultados do teste de Ljung-box para as previsões base ($\alpha = 0.05$)

Resultado	Proporção
não rejeita	0.7945493
rejeita	0.2054507

4.4 Design do experimento

O objetivo do experimento consiste em comparar o ganho de desempenho dos métodos de reconciliação em relação às previsões base, ou seja, dada a performance das previsões

individuais, o quanto mais precisas elas se tornam ao aplicar um determinado método de reconciliação. Para isso, serão utilizadas a abordagem de nível único *bottom-up*²¹, o método analítico de reconciliação ótima *MinT-Shrink* e os métodos de reconciliação ótima baseados em *machine learning*: *gradient boosting* (*Xgboost* e *LightGBM*), *Random Forest*, *Elastic Net* (Lasso, Ridge e Elastic Net) e *Support Vector Machines*.

A metodologia para obtenção das previsões reconciliadas por métodos baseados em *machine-learning* será semelhante ao de Spiliotis et al. (2021). Ela consiste em:

1. Previsão contínua (*rolling forecast*): A amostra treino é separada em p subamostras para todas as séries em todos os níveis de agregação, com a primeira separação abrangendo y_1 até Y_Q , a segunda y_1 até Y_{Q+1} e assim sucessivamente até a última separação abrangendo y_1 até Y_{Q+p-1} . Para cada subamostra foi treinado um modelo (de mesmo algoritmo usado na seção 4.3) e obtida previsão um passo a frente. Para este trabalho, a amostra treino foi dividida em 50%, com $Q = 120$ e $p = 108$.

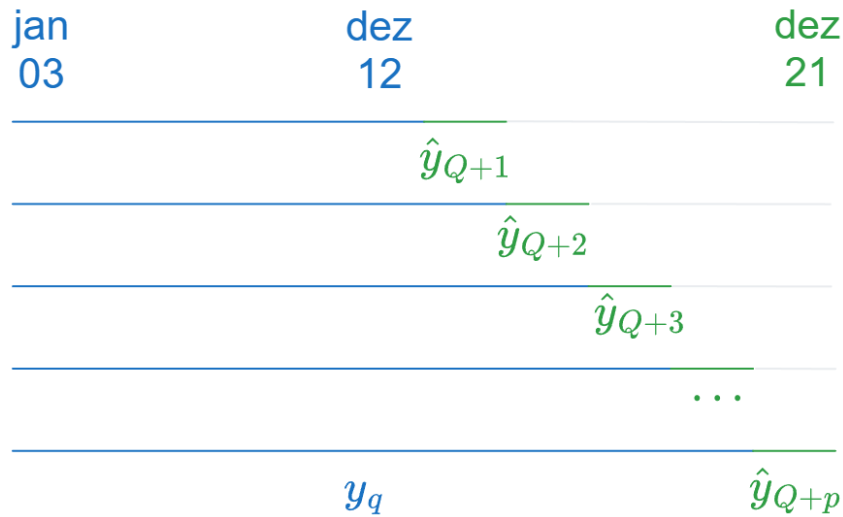


Figura 14 – Esquema de modelagem de previsões contínuas

2. Treino dos modelos de *machine learning*: Para cada série do nível mais desagregado, y_m , é treinado um modelo de ML com $n + 1$ variáveis, compostas pelas n séries — que incluem todos os níveis de agregação —, mais a própria y_m como *target* (Tabela 6). Cada uma das n séries contam com p previsões obtidas no passo 1. Isso resulta em um modelo de reconciliação ótima para cada elemento do menor nível da hierarquia, combinando informações disponíveis de todos os níveis hierárquicos.
3. Reconciliação ótima: Com os modelos treinados, passa-se as previsões base obtidas na seção 4.3 como regressores para se obter as previsões reconciliadas das séries do nível mais desagregado \tilde{y}_m .

²¹ Como o *dataset* Estban é agrupado, a desagregação não é única e as abordagens *top-down* e *middle-out* não podem ser aplicadas (ATHANASOPOULOS; HYNDMAN et al., 2023).

4. Agregação: Assim como nos métodos analíticos de combinação ótima, a obtenção das previsões reconciliadas para os demais níveis de hierárquicos \tilde{y}_n se dá através da agregação semelhante ao *bottom-up*, mas ao invés de se somar as previsões base \hat{y}_m , somam-se as previsões reconciliadas \tilde{y}_m .

Tabela 6 – Conjunto de dados para predição dos modelos de ML

Target	Variável 1	Variável 2	...	Variável n
$y_{1,Q+1}$	$\hat{y}_{1,Q+1}$	$\hat{y}_{2,Q+1}$...	$\hat{y}_{n,Q+1}$
$y_{2,Q+2}$	$\hat{y}_{1,Q+2}$	$\hat{y}_{2,Q+2}$...	$\hat{y}_{n,Q+2}$
...
$y_{m,Q+p}$	$\hat{y}_{1,Q+p}$	$\hat{y}_{2,Q+p}$...	$\hat{y}_{n,Q+p}$

Dessa forma, essa metodologia é semelhante à aplicada na reconciliação ótima analítica, se afastando principalmente em três pontos: (i) a utilização de algoritmos de ML ao invés de MQG, (ii) a não atribuição de peso de forma obrigatória para todos os nós da hierarquia e (iii) no ajuste de um modelo individual para cada série do nível mais desagregado, permitindo maior especialização e sendo capaz de se adaptar melhor aos diferentes padrões de cada série (Spiliotis et al. (2021)).

Um ponto negativo na metodologia proposta por Spiliotis et al. (2021) é o processo de *rolling origin* (passo 1). Esse processo requer a realização de previsões para dentro da amostra treino, o que pode ser um problema para séries temporais com poucas observações ou de série incompleta. No caso do *dataset* Estban, algumas das agências foram criadas após o período escolhido para o *split* em Q (dezembro/2012), sendo necessário sua exclusão do dataset e invalidando o uso da metodologia para essas unidades. Nesse sentido, o MinT se mostra uma opção mais viável para aplicações no mundo real.

Nesses casos, uma alternativa para permitir o uso dos métodos de ML é substituir o processo de *rolling origin* e usar os valores ajustados dos modelos das previsões base \hat{y} como input para os modelos de ML. Além de permitir a inclusão de séries incompletas — agências criadas durante o período, no caso do *dataset* Estban —, essa abordagem também aumenta o tamanho da amostra treino em Q observações, o que pode melhorar a performance dos modelos de ML. Chamaremos essa estratégia de *fitted base forecasts*.

Outra possibilidade de substituição do passo 1, caso as séries sejam de tamanho suficiente, é o processo de reajuste. Esse processo consiste no reajuste de um modelo para um novo conjunto de dados, conservando os hiperparâmetros originais porém reestimando os coeficientes (e.g., treina-se um modelo autoregressivo $AR(p)$ de coeficientes ϕ_p e então passa-se um novo conjunto de dados fora da amostra, mantendo o hiperparâmetro p e reestimando ϕ_p , obtendo novos valores ajustados). Utiliza-se então os valores reajustados para treinar os modelos de ML. Nessa estratégia, doravante denominada de *refit*, os modelos foram treinados até Q

e então reajustados para $Q + p$. Uma restrição dessa abordagem é que, fixados os hiperparâmetros anteriores, não necessariamente todos modelos alcançarão convergência no reajuste de seus coeficientes para o novo conjunto de dados.

4.5 Otimização de hiperparâmetros

A maior parte dos métodos de *machine learning* são altamente parametrizáveis, sendo sua performance de generalização (para fora da amostra) sensível à escolha de seus hiperparâmetros. Quando disponível, os hiperparâmetros a serem otimizados e seus espaços de busca seguiram a recomendação em [Bischl et al. \(2021\)](#).

Os conjuntos de hiperparâmetros e seus intervalos são apresentados no Anexo A. Para a otimização, foram utilizadas dois calibradores: (i) busca em grade (com resolução de 10 combinações), mais custoso em tempo de processamento, para os métodos com menor quantidade de hiperparâmetros a serem otimizados, e (ii) otimização bayesiana (na configuração padrão do pacote {mlr3MBO}), mais eficiente para os métodos com maior quantidade de hiperparâmetros. A estratégia de reamostragem utilizada foi a validação cruzada *k-fold* com $k = 10$.

A otimização bayesiana foi usada em todos os métodos, exceto no *elastic net*, uma vez que apenas um (no caso do *lasso* e *ridge*) ou dois hiperparâmetros foram otimizados. A medida de performance utilizada para a otimização foi a raiz do erro quadrático médio (*root mean squared error* — RMSE).

Por fim, cada modelo foi calibrado individualmente, ou seja, cada agência possui um conjunto de hiperparâmetros otimizados para cada um dos 7 métodos de ML empregados para reconciliação ótima.

5 RESULTADOS

5.1 ESTBAN

Tabela 7 – Resultados Estban: Tempo de processamento dos métodos de ML (em horas)

	xgb	ranger	glmnet	lasso	ridge	svm	lightgbm
one-step-ahead	19.2	5.5	1.4	0.9	1.0	1.3	3.3
fitted base	21.8	5.5	1.4	0.8	0.9	1.3	3.3
rolling forecast	36.4	5.4	1.3	0.8	0.9	1.3	3.4

5.2 TOURISM

Tabela 8 – Resultados Estban: Acurácia dos métodos analíticos de reconciliação

.model	agregado	mesorregiao	microrregiao	municipio	agencia	verbete	bottom	hierarquia
RMSSE								
base	0.20	0.69	0.85	0.67	0.72	0.18	0.66	0.66
bu	0.20	0.56	0.71	0.71	0.74	0.18	0.66	0.66
mint	0.17	0.62	0.72	0.69	0.71	0.14	0.63	0.64
MASE								
base	0.28	0.82	1.00	0.79	0.89	0.25	0.88	0.85
bu	0.24	0.57	0.77	0.82	0.89	0.22	0.88	0.84
mint	0.22	0.69	0.86	0.83	0.89	0.16	0.86	0.84

Tabela 9 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estratégia rolling forecast.

modelo	agregado	mesorregiao	microrregiao	municipio	agencia	verbete	bottom	hierarquia
RMSSE								
glmnet	0.28	0.76	1.18	1.21	1.25	0.77	1.06	1.09
lasso	0.20	0.73	1.05	1.00	1.04	0.50	0.84	0.88
lightgbm	1.41	1.63	1.53	1.29	1.32	0.88	0.97	1.09
ranger	1.23	1.40	1.30	1.12	1.17	0.73	0.86	0.95
ridge	0.42	0.78	1.13	1.51	1.53	0.92	1.36	1.37
svm	0.26	0.75	0.86	0.85	0.91	0.23	0.85	0.84
xgb	1.19	1.41	1.30	1.10	1.14	0.70	0.83	0.92
MASE								
glmnet	0.23	0.73	1.41	1.49	1.58	0.95	1.44	1.43
lasso	0.17	0.71	1.25	1.19	1.30	0.64	1.14	1.15
lightgbm	1.89	1.90	1.87	1.60	1.65	1.23	1.39	1.48
ranger	1.61	1.56	1.50	1.33	1.42	1.00	1.10	1.18
ridge	0.40	0.76	1.31	1.80	1.88	1.09	1.79	1.74
svm	0.31	0.68	0.86	0.99	1.11	0.29	1.25	1.14
xgb	1.54	1.56	1.48	1.30	1.37	0.95	1.02	1.12

Tabela 10 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estratégia fitted base.

modelo	agregado	mesorregiao	microrregiao	municipio	agencia	verbete	bottom	hierarquia
RMSSE								
glmnet	0.78	0.99	1.09	1.04	1.15	0.58	0.92	0.96
lasso	0.66	0.96	1.07	0.91	1.01	0.53	0.83	0.86
lightgbm	1.49	1.65	1.56	1.30	1.34	0.92	1.00	1.12
ranger	1.20	1.40	1.29	1.10	1.15	0.69	0.84	0.93
ridge	1.00	1.15	1.25	1.21	1.33	0.69	1.12	1.15
svm	0.39	0.93	0.93	0.93	0.96	0.32	0.90	0.90
xgb	1.20	1.37	1.28	1.08	1.13	0.70	0.82	0.92
MASE								
glmnet	1.05	1.14	1.32	1.27	1.45	0.80	1.29	1.28
lasso	0.89	1.09	1.32	1.11	1.28	0.73	1.17	1.16
lightgbm	2.03	1.93	1.91	1.61	1.68	1.30	1.43	1.51
ranger	1.58	1.56	1.49	1.31	1.40	0.95	1.04	1.13
ridge	1.34	1.35	1.55	1.50	1.70	0.94	1.58	1.55
svm	0.45	0.95	1.08	1.16	1.23	0.34	1.28	1.22
xgb	1.54	1.51	1.48	1.29	1.38	0.94	1.02	1.11

Tabela 11 – Resultados Estban: Acurácia dos métodos de ML de reconciliação. Estratégia refit.

modelo	agregado	mesorregiao	microrregiao	municipio	agencia	verbete	bottom	hierarquia
RMSSE								
glmnet	0.09	0.58	0.73	0.82	0.84	0.16	0.73	0.74
lasso	0.13	0.52	0.77	0.76	0.77	0.19	0.68	0.69
lightgbm	1.41	1.59	1.52	1.28	1.32	0.89	0.97	1.09
ranger	1.25	1.41	1.32	1.12	1.17	0.69	0.86	0.95
ridge	0.14	0.63	0.78	0.90	0.92	0.21	0.84	0.83
svm	0.19	0.74	0.77	0.79	0.83	0.30	0.81	0.79
xgb	1.22	1.35	1.25	1.08	1.14	0.71	0.84	0.93
MASE								
glmnet	0.09	0.58	0.83	0.97	1.01	0.21	0.94	0.92
lasso	0.14	0.52	0.88	0.91	0.93	0.22	0.89	0.88
lightgbm	1.88	1.83	1.83	1.58	1.65	1.24	1.39	1.47
ranger	1.64	1.58	1.53	1.33	1.42	0.95	1.07	1.16
ridge	0.16	0.63	0.89	1.05	1.09	0.23	1.07	1.02
svm	0.23	0.76	0.85	0.95	1.02	0.39	1.18	1.08
xgb	1.59	1.49	1.43	1.29	1.38	0.97	1.06	1.14

Tabela 12 – Resultados Tourism: Tempo de processamento dos métodos de ML (em horas)

	xgb	ranger	glmnet	lasso	ridge	svm	lightgbm
fitted base	15.8	3.7	1.5	1.3	1.4	2	3.5
rolling forecast	12.1	3.0	1.0	0.8	1.1	2	3.6

Tabela 13 – Resultados Tourism: Acurácia dos métodos analíticos de reconciliação

.model	agregado	state	region	purpose	bottom	hierarquia
RMSSE						
base	1.45	1.26	1.07	1.26	0.92	0.98
bu	2.58	1.63	1.11	2.00	0.92	1.01
mint	1.81	1.30	0.98	1.42	0.88	0.92
MASE						
base	1.53	1.40	1.13	1.33	0.98	1.04
bu	3.16	1.88	1.18	2.32	0.98	1.08
mint	2.09	1.45	1.02	1.51	0.94	0.98

Tabela 14 – Resultados Tourism: Acurácia dos métodos de ML de reconciliação. Estratégia rolling forecast.

modelo	agregado	State	Region	Purpose	bottom	hierarquia
RMSSE						
glmnet	1.99	1.39	1.09	1.54	0.99	1.04
lasso	1.93	1.37	1.10	1.52	1.03	1.07
lightgbm	4.33	2.76	1.65	3.46	1.14	1.35
ranger	2.13	1.36	1.03	1.71	0.91	0.97
ridge	1.26	1.18	1.01	1.20	0.92	0.96
svm	0.94	1.01	1.08	1.01	1.10	1.10
xgb	2.34	1.45	1.11	1.89	0.96	1.03
MASE						
glmnet	2.36	1.57	1.15	1.65	1.06	1.12
lasso	2.26	1.56	1.17	1.59	1.11	1.16
lightgbm	5.50	3.21	1.76	4.06	1.20	1.45
ranger	2.58	1.53	1.07	1.82	0.96	1.02
ridge	1.34	1.31	1.06	1.19	0.98	1.02
svm	1.07	1.10	1.14	1.03	1.18	1.17
xgb	2.89	1.65	1.16	2.12	1.01	1.09

Tabela 15 – Resultados Tourism: Acurácia dos métodos de ML de reconciliação. Estratégia fitted base.

modelo	agregado	State	Region	Purpose	bottom	hierarquia
RMSSE						
glmnet	2.17	1.40	1.10	1.77	0.97	1.03
lasso	1.90	1.45	1.09	1.61	0.97	1.03
lightgbm	4.33	2.76	1.65	3.46	1.14	1.35
ranger	2.12	1.36	1.03	1.72	0.91	0.96
ridge	1.57	1.16	0.97	1.29	0.90	0.93
svm	1.50	1.19	1.05	1.38	1.04	1.05
xgb	2.27	1.42	1.10	1.83	0.96	1.02
MASE						
glmnet	2.59	1.57	1.16	1.95	1.04	1.10
lasso	2.21	1.67	1.16	1.73	1.04	1.11
lightgbm	5.50	3.21	1.76	4.06	1.20	1.45
ranger	2.54	1.54	1.07	1.84	0.97	1.02
ridge	1.77	1.29	1.01	1.31	0.96	0.99
svm	1.75	1.30	1.09	1.36	1.10	1.11
xgb	2.79	1.62	1.15	2.04	1.01	1.08

REFERÊNCIAS

- ATHANASOPOULOS, G.; AHMED, R. A.; HYNDMAN, R. J. Hierarchical forecasts for Australian domestic tourism. en. **International Journal of Forecasting**, v. 25, n. 1, p. 146–166, jan. 2009. ISSN 0169-2070. DOI: [10.1016/j.ijforecast.2008.07.004](https://doi.org/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. 27.
- ATHANASOPOULOS, G.; HYNDMAN, R. J. et al. Forecast reconciliation: A review. en, 2023. Citado na p. 49.
- BADER, F. L. C.; KOYAMA, SÉRGIO MIKIO; TSUCHIDA, MARCOS HIROYUKI. Modelo favar canônico para previsão do mercado de crédito. pt. **Banco Central do Brasil**, v. 369, p. 38, nov. 2014. ISSN 1519-1028. Citado na p. 25.
- BANCO DO ESTADO DO ESPÍRITO SANTO. **Demonstrações Financeiras**. Vitória, 2022. Disponível em: https://www.banestes.com.br/ri/arquivos/informacoes/demoFinanceiras/2022_2S_demonFinanceiras.pdf>. Acesso em: 4 jun. 2023. Citado na p. 23.
- BECCALLI, E. et al. Earnings management, forecast guidance and the banking crisis. **The European Journal of Finance**, v. 21, n. 3, p. 242–268, fev. 2015. ISSN 1351-847X. DOI: [10.1080/1351847X.2013.809548](https://doi.org/10.1080/1351847X.2013.809548). Disponível em: <https://doi.org/10.1080/1351847X.2013.809548>>. Acesso em: 7 mai. 2023. Citado na p. 22.
- BISCHL, B. et al. **Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges**. [S.l.]: arXiv, nov. 2021. arXiv:2107.05847 [cs, stat]. DOI: [10.48550/arXiv.2107.05847](https://doi.org/10.48550/arXiv.2107.05847). Disponível em: <http://arxiv.org/abs/2107.05847>>. Acesso em: 4 set. 2023. Citado na p. 51.
- BRASIL. **Lei nº 7.492, de 16 de junho de 1986**. Brasília, DF: Presidência da República, jun. 1986. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/17492.htm>. Citado na p. 21.
- CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: PROCEEDINGS of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery, ago. 2016. (KDD '16), p. 785–794. ISBN 978-1-4503-4232-2. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>>. Acesso em: 23 jan. 2024. Citado na p. 38.
- CHEN, T.; HE, T. et al. **xgboost: Extreme Gradient Boosting**. [S.l.], 2023. R package version 1.7.5.1. Disponível em: <https://github.com/dmlc/xgboost>>. Citado na p. 41.

COELHO, A. N. B. et al. A responsabilidade da auditoria externa na fraude contábil do banco panamericano. pt. **RAGC**, v. 3, n. 7, set. 2015. Number: 7. ISSN 2317-0484. Disponível em: <<https://revistas.fucamp.edu.br/index.php/ragc/article/view/604>>. Acesso em: 31 mai. 2023. Citado na p. 22.

ÇOLAK, M. S. et al. **TCMB - Monitoring and Forecasting Cyclical Dynamics in Bank Credits: Evidence from Turkish Banking Sector**. en. Ankara, Turquia, 2019. Disponível em: <<https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Papers/2019/19-29>>. Acesso em: 6 mar. 2023. Citado na p. 25.

CONSELHO MONETÁRIO NACIONAL. **Resolução nº 2.099, de 17 de agosto de 1994**. Brasília, DF: Banco Central do Brasil, ago. 1994. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/1994/pdf/res_2099_v1_O.pdf>. Citado na p. 23.

_____. **Resolução nº 2.682, de 21 de dezembro de 1999**. Brasília, DF: Banco Central do Brasil, dez. 1999. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf>. Citado na p. 43.

DAHIS, R. et al. Data Basis (Base Dos Dados): Universalizing Access to High-Quality Data. en. **SSRN Electronic Journal**, 2022. ISSN 1556-5068. DOI: [10.2139/ssrn.4157813](https://doi.org/10.2139/ssrn.4157813). Disponível em: <<https://www.ssrn.com/abstract=4157813>>. Acesso em: 18 jun. 2023. Citado na p. 42.

FRIEDMAN, J. et al. **glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models**. [S.l.], 2023. R package version 4.1-8. Disponível em: <<https://glmnet.stanford.edu>>. Citado na p. 42.

GORODETSKAYA, O.; GOBAREVA, Y.; KOROTEEV, M. A Machine Learning Pipeline for Forecasting Time Series in the Banking Sector. en. **Economies**, v. 9, n. 4, p. 205, dez. 2021. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. ISSN 2227-7099. DOI: [10.3390/economies9040205](https://doi.org/10.3390/economies9040205). Disponível em: <<https://www.mdpi.com/2227-7099/9/4/205>>. Acesso em: 27 fev. 2023. Citado na p. 26.

GROSS, C. W.; SOHL, J. E. Disaggregation methods to expedite product line forecasting. en. **Journal of Forecasting**, v. 9, n. 3, p. 233–254, 1990. ISSN 1099-131X. DOI: [10.1002/for.3980090304](https://doi.org/10.1002/for.3980090304). Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980090304>>. Acesso em: 18 jun. 2023. Citado nas pp. 23, 26.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition**. 2nd 2009, Corr. 9th Printing 2017 ed. edição. New York, NY: Springer, fev. 2009. ISBN 978-0-387-84857-0. Citado nas pp. 38, 39, 41.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3. ed. Melbourne, Austrália: OTexts, 2021. Disponível em: <<https://otexts.com/fpp3/>>. Citado nas pp. 23, 26, 29, 31.

HYNDMAN, R. J.; KOEHLER, A. B. et al. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439–454, jul. 2002. ISSN 0169-2070. DOI: [10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207001001108>. Acesso em: 26 nov. 2023. Citado na p. 45.

HYNDMAN, R. J.; AHMED, R. A. et al. Optimal combination forecasts for hierarchical time series. en. **Computational Statistics & Data Analysis**, v. 55, n. 9, p. 2579–2589, set. 2011. ISSN 0167-9473. DOI: [10.1016/j.csda.2011.03.006](https://doi.org/10.1016/j.csda.2011.03.006). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167947311000971>. Acesso em: 11 jan. 2023. Citado nas pp. 23, 27.

HYNDMAN, R. J.; LEE, A. J.; WANG, E. Fast computation of reconciled forecasts for hierarchical and grouped time series. en. **Computational Statistics & Data Analysis**, v. 97, p. 16–32, mai. 2016. ISSN 0167-9473. DOI: [10.1016/j.csda.2015.11.007](https://doi.org/10.1016/j.csda.2015.11.007). Disponível em: <https://www.sciencedirect.com/science/article/pii/S016794731500290X>. Acesso em: 11 jan. 2023. Citado na p. 28.

KE, G. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: **ADVANCES in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>. Acesso em: 23 jan. 2024. Citado na p. 38.

LANG, M. et al. **mlr3: Machine Learning in R - Next Generation**. [S.l.], 2023. R package version 0.16.1. Disponível em: <https://CRAN.R-project.org/package=mlr3>. Citado na p. 41.

LI, Z. 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](https://doi.org/10.3390/en9010055). Disponível em: <https://www.mdpi.com/1996-1073/9/1/55>. Acesso em: 8 abr. 2023. Citado na p. 27.

MEYER, D. et al. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien**. [S.l.], 2023. R package version 1.7-13. Disponível em: <https://CRAN.R-project.org/package=e1071>. Citado na p. 42.

MOURA, D. d. **Análise dos fatores de convencimento do juízo brasileiro quanto à ocorrência de fraude contábil: um estudo de caso Múltiplo da Gallus, da Encol e do Banco Santos**. 2007. Dissertação de mestrado – Fundação Getúlio Vargas, Rio de Janeiro. Accepted: 2009-11-18T19:01:34Z. Disponível em: <http://bibliotecadigital.fgv.br:80/dspace/handle/10438/4038>. Acesso em: 31 mai. 2023. Citado na p. 21.

O'HARA-WILD, M.; HYNDMAN, R.; WANG, E. **fable: Forecasting Models for Tidy Time Series**. [S.l.], 2023. R package version 0.3.3. Disponível em: <<https://CRAN.R-project.org/package=fable>>. Citado na p. 41.

PANAGIOTELIS, A. et al. Forecast reconciliation: A geometric view with new insights on bias correction. en. **International Journal of Forecasting**, v. 37, n. 1, p. 343–359, jan. 2021. ISSN 0169-2070. DOI: [10.1016/j.ijforecast.2020.06.004](https://doi.org/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. 28.

PRAYOGA, I.; SUHARTONO, S.; RAHAYU, S. 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, jan. 2017. Citado na p. 26.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org/>>. Citado na p. 41.

SEZER, O. B.; GUDELEK, M. U.; OZBAYOGLU, A. M. **Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019**. Turquia: arXiv, 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. 25.

SHI, Y. et al. **lightgbm: Light Gradient Boosting Machine**. [S.l.], 2023. R package version 4.2.0. Disponível em: <<https://github.com/Microsoft/LightGBM>>. Citado na p. 41.

SPILOTIS, E. et al. Hierarchical forecast reconciliation with machine learning. en. **Applied Soft Computing**, v. 112, p. 107756, nov. 2021. ISSN 1568-4946. DOI: [10.1016/j.asoc.2021.107756](https://doi.org/10.1016/j.asoc.2021.107756). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494621006773>>. Acesso em: 11 jan. 2023. Citado nas pp. 24, 28, 49, 50.

WALLANDER, J. Budgeting — an unnecessary evil. en. **Scandinavian Journal of Management**, v. 15, n. 4, p. 405–421, dez. 1999. ISSN 0956-5221. DOI: [10.1016/S0956-5221\(98\)00032-3](https://doi.org/10.1016/S0956-5221(98)00032-3). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0956522198000323>>. Acesso em: 8 mai. 2023. Citado na p. 21.

WICKRAMASURIYA, S. L.; ATHANASOPOULOS, G.; HYNDMAN, R. J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. en. **Journal of the American Statistical Association**, v. 114, n. 526, p. 804–819, abr. 2019. ISSN 0162-1459, 1537-274X. DOI: [10.1080/01621459.2018.1448825](https://doi.org/10.1080/01621459.2018.1448825). Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825>>. Acesso em: 18 jun. 2023. Citado nas pp. 24, 28, 36.

WRIGHT, M. N.; WAGER, S.; PROBST, P. **ranger: A Fast Implementation of Random Forests**. [S.l.], 2023. R package version 0.15.1. Disponível em: <<https://github.com/imbs-hl/ranger>>. Citado na p. 42.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. en. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301–320, 2005. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2005.00503.x>. ISSN 1467-9868. DOI: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x). Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x>>. Acesso em: 5 jun. 2023. Citado na p. 37.

Apêndices

APÊNDICE A – DEMONSTRAÇÕES

Proposição A.1 (condição de ausência de viés em $\tilde{\mathbf{y}}$). *Se as previsões reconciliadas são não viesadas, então $\mathbf{SGS} = \mathbf{S}$, ou seja, \mathbf{G} é inversa generalizada de \mathbf{S} .*

Demonstração.

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.1})$$

Se $\hat{\mathbf{y}}_{t+h|t}$ é não viesado, então

$$\mathbb{E}[\hat{\mathbf{y}}_{t+h|t}] = \mathbb{E}[\mathbf{y}_{t+h|t}] = \mathbf{S}\mathbf{b}_t \quad (\text{A.2})$$

Da mesma forma, se espera-se que as previsões reconciliadas não sejam viesadas,

$$\mathbb{E}[\tilde{\mathbf{y}}_{t+h|t}] = \mathbb{E}[\mathbf{y}_{t+h|t}] = \mathbf{S}\mathbf{b}_t \quad (\text{A.3})$$

Substituindo (A.2) em (A.1), temos

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SGS}\mathbf{b}_t \quad (\text{A.4})$$

Logo, para manter a igualdade entre (A.1) e (A.4), $\mathbf{SGS} = \mathbf{S}$

□

Proposição A.2. $\tilde{\mathbf{e}}_t = \mathbf{SG}\hat{\mathbf{e}}_t$.

Demonstração.

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t} \quad (\text{A.5})$$

Substituindo (A.1) em (A.5),

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.6})$$

Lembrando que, por definição, $\mathbf{y}_{t+h} = \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t}$, então

$$\tilde{\mathbf{e}}_{t+h|t} = \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t} - \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.7})$$

$$= \hat{\mathbf{e}}_{t+h|t} + \hat{\mathbf{y}}_{t+h|t}(\mathbf{I} - \mathbf{SG}) \quad (\text{A.8})$$

Usando a definição novamente, temos que

$$\tilde{e}_{t+h|t} = \hat{e}_{t+h|t} + (\mathbf{y}_{t+h} - \hat{e}_{t+h|t})(\mathbf{I} - \mathbf{SG}) \quad (\text{A.9})$$

$$= \mathbf{y}_{t+h} - \mathbf{SG}\mathbf{y}_{t+h|t} + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.10})$$

$$= \mathbf{y}_{t+h}(\mathbf{I} - \mathbf{SG}) + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.11})$$

Substituindo (7) em (A.11), temos

$$\tilde{e}_{t+h|t} = \mathbf{S}\mathbf{b}_{t+h}(\mathbf{I} - \mathbf{SG}) + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.12})$$

$$= \mathbf{S}\mathbf{b}_{t+h} - \mathbf{S}\mathbf{b}_{t+h}\mathbf{SG} + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.13})$$

$$= \mathbf{S}\mathbf{b}_{t+h} - (\mathbf{G}'\mathbf{S}')(\mathbf{b}'_{t+h}\mathbf{S}') + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.14})$$

$$= \mathbf{S}\mathbf{b}_{t+h} - \mathbf{SG}\mathbf{S}\mathbf{b}_{t+h} + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.15})$$

Finalmente, pela condição de ausência de viés em $\tilde{\mathbf{y}}$, temos que

$$\tilde{e}_{t+h|t} = \mathbf{S}\mathbf{b}_{t+h} - \mathbf{S}\mathbf{b}_{t+h} + \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.16})$$

$$= \mathbf{SG}\hat{e}_{t+h|t} \quad (\text{A.17})$$

□

Proposição A.3. $\text{Var}[\tilde{e}_t] = \mathbf{SG}\hat{\mathbf{W}}\mathbf{G}'\mathbf{S}'$.

Demonstração. Por A.2, temos que

$$\text{Var}[\tilde{e}] = \mathbb{E}[(\mathbf{SG}\hat{e})(\mathbf{SG}\hat{e})'] \quad (\text{A.18})$$

$$= \mathbb{E}[\mathbf{SG}\hat{e}\hat{e}'\mathbf{G}'\mathbf{S}'] \quad (\text{A.19})$$

$$= \mathbf{SG}\hat{\mathbf{W}}\mathbf{G}'\mathbf{S}' \quad (\text{A.20})$$

Em que $\hat{\mathbf{W}}$ é a matriz de variância-covariância dos erros de previsão base. □

Proposição A.4. $\hat{\mathbf{W}}$ é posto incompleto.

Demonstração. Pela propriedade do vínculo do posto do produto de matrizes, ou seja, $\text{pos}(\mathbf{AB}) \leq \min(\text{pos}(\mathbf{A}), \text{pos}(\mathbf{B}))$, temos que

$$\text{pos}(\mathbf{SG}\hat{e}_{t+h|t}) \leq \min(\text{pos}(\mathbf{S}), \text{pos}(\mathbf{G}), \text{pos}(\hat{e}_{t+h|t})) \quad (\text{A.21})$$

Como S é a representação matricial de uma estrutura hierárquica, em que os nós pais totalizam os nós filhos, S apresenta, por hipótese, dependência linear e, conseqüentemente, posto incompleto.

Pela equação (A.21), segue que \tilde{e} é posto incompleto. Da mesma forma, $pos(\tilde{e}\tilde{e}') \leq \min(pos(\tilde{e}), pos(\tilde{e}'))$. Portanto, \hat{W} é posto incompleto. \square

Anexos

ANEXO A – CONJUNTO DE HIPERPARÂMETROS

Tabela 16 – Intervalos de hiperparâmetros para {xgboost}

Hiperparâmetro	Descrição	Intervalo	Trafo
nrounds	Número de iterações	[1, 5000]	NULL
eta	Taxa de aprendizado	$[-4, 0]$	10^x
max_depth	Profundidade máxima	[1, 20]	NULL
subsample	Subamostra	[0.1, 1]	NULL
colsample_bytree	Subamostra de colunas para uma árvore	[0.1, 1]	NULL
colsample_bylevel	Subamostra de colunas por nível de profundidade	[0.1, 1]	NULL
lambda	Regularização L2	$[-10, 10]$	2^x
alpha	Regularização L1	$[-10, 10]$	2^x

Tabela 17 – Intervalos de hiperparâmetros para {lightgbm}

Hiperparâmetro	Descrição	Intervalo	Trafo
num_iterations	Número de iterações	[1, 1000]	NULL
boosting	Algoritmo de boosting	{gbdt, dart, goss}	NULL
learning_rate	Taxa de aprendizado	$[-4, 0]$	10^x
num_leaves	Número de folhas	[2, 20]	NULL
lambda_l1	Regularização L1	$[-12, 12]$	2^x
lambda_l2	Regularização L2	$[-12, 12]$	2^x
feature_fraction	Subamostra de colunas	[0.1, 1]	NULL
bagging_fraction	Subamostra de linhas	[0.1, 1]	NULL
bagging_freq	Frequência de amostragem	[1, 10]	NULL

Tabela 18 – Intervalos de hiperparâmetros para {ranger}

Hiperparâmetro	Descrição	Intervalo	Trafo
min.node.size	Número mínimo de observações em um nó terminal	[1, 7]	2^x
mtry	Número de variáveis candidatas para split	[1,)	NULL
replace	Amostragem com reposição	{TRUE, FALSE}	NULL
sample.fraction	Fração de observações a serem amostradas	[0.1, 1]	NULL
num.trees	Número de árvores	[1, 2000]	NULL

Tabela 19 – Intervalos de hiperparâmetros para {e1071} (svm)

Hiperparâmetro	Descrição	Intervalo	Trafo
cost	Custo de ξ	$[0, 1]$	2^x
kernel	Kernel	{linear, polynomial, radial, sigmoid}	NULL
degree	Grau do polinômio	$[1, 5]$	NULL
gamma	Influência amostral	$[-12, 12]$	2^x
type	Tipo de SVM	{eps-regression}	NULL

Tabela 20 – Intervalos de hiperparâmetros para {glmnet}

Hiperparâmetro	Descrição	Intervalo	Trafo
alpha	Mix entre lasso e ridge	$[0, 1]$	NULL
lambda	Regularização	$[-12, 12]$	2^x