

APLICAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA COM SELEÇÃO DE VARIÁVEIS NA PREVISÃO DE RECEITAS PÚBLICAS DE 8 CAPITALS:

Relato dos resultados preliminares mais relevantes com dados de São Luís

Cláudia Patrícia S. Pimentel[†]
Universidade Estadual do Maranhão
São Luís, MA
cacaupimentel@gmail.com

Fábio M. F. Lobato
Universidade Federal do Oeste do Pará
Santarém, PA
fabio.lobato@ufopa.edu.br

Antonio F. L. Jacob Jr
Universidade Estadual do Maranhão
São Luís, MA
antoniojunior@professor.uema.br

ABSTRACT

Forecasting public revenues at the municipal level is quite relevant for public administration in Brazil since it defines the budget of public policies. However, it is a complex process, and its consequences that permeate the divergence between executed and budgeted values have harmful effects on the population. So, aiming to mitigate this problem, we analyzed to optimize public revenue forecasting using feature selection through model combination (ensemble) of wrapper filter approaches. We assessed the proposed solution in light of the Brazilian Fiscal Responsibility Law, using São Luís, Maranhão, as a case study. Our results suggested that our approach is suitable for reducing complexity and improving forecasting performance. The method can be used in other Brazilian state capitals, boosting data-driven public policy-making.

KEYWORDS

Public Revenue Forecasting, Constitutional and Legal Transfers, Machine Learning; Feature Selection; CRISP-DM.

1 INTRODUÇÃO

No âmbito dos Entes Subnacionais, a busca para atender os desafios da sociedade quando se trata de desigualdades sociais perpassa por uma constante demanda de desenvolvimento social e econômico os quais se refletem nas políticas públicas [1]. Atendendo o imperativo constitucional, tais políticas são financiadas pela arrecadação de tributos, os quais deverão planejar e especificar os gastos e investimentos priorizados para atender os serviços públicos, logo, a Constituição Federal (CF) de 1988 estabeleceu o Orçamento Público [2].

Com o advento da Lei de Responsabilidade Fiscal (LRF), o órgão responsável para orientar a estimativa das receitas e fixação das despesas nas peças orçamentárias do setor público é a Secretaria do Orçamento Federal (SOF) [2], [3]. Tomando como

estudo de caso a capital do Maranhão, São Luís, ao se comparar os valores inicialmente orçados com os executados mensalmente observam-se altas distorções nos valores das receitas transferidas do Município, no acumulado de janeiro a agosto de 2021, tendo períodos com subestimação entre 13% a 317% e 39% de uma superestimação [4].

Tendo a compreensão que a SOF indica as diretrizes metodológicas de previsão das receitas, concedendo liberdade de escolha por parte do gestor, desde que seja justificada e descrita [1], faz mister a análise dos dados e seus relacionamentos com as variáveis exógenas que podem influenciar o comportamento das receitas, em atendimento ao Art. 12 da LRF [3]. Percebendo-se o mérito, a magnitude e a relevância da previsão das receitas orçamentárias para a esfera municipal, este estudo traz à baila o seguinte problema: Quais as variáveis exógenas aplicada na previsão das Receitas, inicialmente, com dados do Município de São Luís poderiam explicar as prerrogativas dos efeitos preço e quantidade exigidos na LRF?

Neste contexto, o objetivo deste artigo é apresentar os resultados preliminares do estudo de caso para a escolha das variáveis exógenas que atendam as prerrogativas do Art. 12 da LRF, no que tange a possíveis descobertas dos efeitos preço e quantidade para previsão das receitas mensal, adotando métodos estatísticos convencionais e técnicas de Aprendizado de Máquina (AM) [3]. Em especial a comparação dos métodos de seleção de variáveis exógenas que sejam mais relevantes para a estimativa das peças orçamentárias, no horizonte trimestral, seguindo a Metodologia *Cross-industry standard process for data mining* (CRISP-DM) [5].

O presente artigo está dividido em seis seções, com esta introdução. A segunda faz um levantamento dos trabalhos acadêmicos que abordam o tema. Na terceira, fez-se uma breve descrição do referencial teórico. Para a quarta seção a ênfase foi para o relato dos experimentos do estudo de caso e a quinta seus resultados. Finaliza-se com as conclusões e os possíveis ajustes.

2 TRABALHOS CORRELATOS

A busca de trabalhos foi feita nas bases do Portal de Periódicos Capes e no site do Google Acadêmico. Foram utilizados os termos: “previsão receitas públicas” ou “previsão receitas”, textos com

idioma em inglês e português, no período de 2016 até agosto de 2021. Dentro de três categorias que refletem na aplicação do problema, encontraram-se 48 publicações, das quais 15 no Setor Público, 25 com Séries Temporais e 8 com seleção de variáveis. Elegeram-se como critério de seleção os artigos que estivessem na área de atuação do Setor Público. Destacando-se, nos 15 trabalhos escolhidos, que ficou evidente a preeminência da utilização de métodos estatísticos, 13 documentos, e apenas 2 usaram algum método de AM, e a receita mais explorada, em 6 pesquisas, foi o Imposto sobre Circulação de Mercadorias e prestação de Serviços de transporte interestadual e intermunicipal e comunicação (ICMS), no âmbito estadual.

2.1 Publicações com foco no setor público

Na literatura foram encontradas duas pesquisas que comparam o modelo proposto no Manual Técnico de Orçamento (MTO) da SOF com modelos paramétricos [6,7]. O primeiro [6] demonstrou a supremacia do modelo Autorregressivo Integrado de Médias Móveis (ARIMA) em valores globais anuais das receitas. Na conclusão, ele não garantiu que possa funcionar para municípios com alta dependência financeira das receitas transferidas ou mesmo a ocorrência de situações extremas, como a Pandemia, que quebram a estrutura das séries. Já o segundo [7] difere ao fazer a comparação com Médias Móveis Autorregressivas (ARMA).

Carmo, Boldt e Komati [8] trazem a primeira publicação com uma comparação de 8 métodos, entre **estatísticos** e **AM**. Aproveitando código aberto da competição M4, extraindo o componente de sazonalidade. Em seus resultados apontaram que o modelo combinado estatístico prevaleceu sobre os demais. O segundo artigo que utilizou AM foi o de Silva e outros autores [9] cujo intuito foi identificar a correlação entre variáveis ambientais, econômicas e sociais com os *royalties* arrecadados de petróleo e minério. Os autores aplicaram **Análise de Componentes Principais (PCA)** resultando em nenhuma correlação.

Tanto no âmbito estadual quanto municipal, pesquisadores elaboraram estudos comparativos da previsão da receita utilizando métodos paramétricos [10,11,12,13,14,15]. Em discussões de eficiência da previsão orçamentária identificou-se que as municipais são melhores em anos eleitorais [16].

No âmbito internacional, investigam-se as causas dos erros das previsões de receitas e despesas, fazem sugestões de boas práticas [17,18]. Eles focam em 3 categorias de erros: política, econômica e técnica [17,19]. Inferiram que o componente econômico explica a receita e o político as despesas [17].

2.2 Apontamentos do estado da arte e da prática

Vários Autores afirmam que, na prática, os Municípios utilizam o modelo oferecido pela SOF [6,7,20]. Vale ressaltar que eles se concentram nas previsões anuais, não abordando o detalhamento da classificação das receitas, prerrogativa essencial para o Ente Subnacional, conforme fases de Entendimento do negócio e dos dados do CRISP-DM. Todavia, ao compararem várias cidades preocuparam-se com a heterogeneidade entre elas e agruparam-nas

por critérios, um cuidado que esta pesquisa também está considerando, haja vista a escolha apenas de Capitais [6,7].

Nota-se que os artigos estão voltados para utilização de métodos estatísticos, em maioria univariados, direcionados a substituir o modelo da SOF. Diferindo da literatura, não se pretende substituí-lo, uma vez que se entende que este sugere a utilização de regressão multivariada. Portanto, pretende-se testar vários modelos para verificar o que melhor se ajusta, uma forma de atender às prerrogativas imposta pela legislação vigente na escolha de variáveis [3], considerando as sugestões de trabalhos futuros que motivaram a elaboração deste estudo [8].

Características	Correlatos	Este Artigo
Propôs substituir o modelo da SOF	✓	✗
Utilizou algum método estatístico	✓	✓
Aplicou algum método de AM	[21,22]	✓
Fez seleção de variáveis	[8,9]	✓
Usou variáveis exógenas para previsão	[8]	✓
Fez previsão de Receitas Transferidas	✗	✓
Atendeu as prerrogativas do art. 12 da LRF	✗	✓
Atuou no âmbito municipal com previsões	[7,15,16]	✓

Quadro 1 - Comparativo dos trabalhos correlatos com o Artigo

Diante do exposto, entende-se que há poucas publicações voltadas para o estudo de métodos de AM em previsão de receitas públicas [21,22], e menos ainda com a proposta de seleção de variáveis para cada receita, bem como a relevância de atender os ditames do art. 12 da LRF [3] - ver resumo do Quadro 1. Portanto, considera-se a proposta de aplicação de técnicas de AM para previsão das receitas públicas, em séries temporais, uma boa contribuição a esta lacuna da literatura.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo teve o intuito de organizar o embasamento teórico do artigo, norteando-se pelos principais assuntos que envolvem os objetivos fixados e o resumo da fase de Entendimento do Negócio do CRISP-DM.

3.1 Receitas Públicas

De acordo com a SOF, as receitas públicas são recursos financeiros que ingressam nos cofres públicos e, quando representam a disponibilidade deles para o erário, e se dividem em receitas orçamentárias e extraorçamentárias [1]. A Lei nº 4.320, em seu art. 11, ao estabelecer o processo de classificação por categoria econômica, dividiu-as em corrente e de capital [23]. Analisando o MTO foi possível compreender que a previsão das receitas é individualizada [24,1].

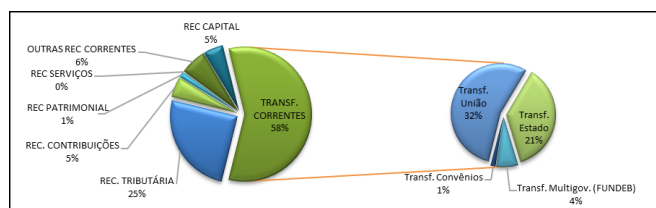
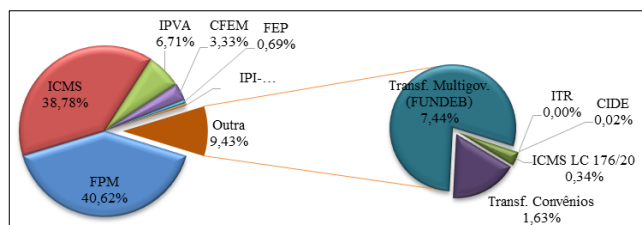


Gráfico 1: Composição das receitas de São Luís/MA

3.1.1 Transferências Correntes.

Em cumprimento à determinação constitucional, ordenamento legal ou voluntário instituem-se as transferências de recursos intergovernamentais, caracterizadas por transações ocorridas entre esferas distintas de governo (federal, estadual e municipal) [1,25]. Ainda sob a ótica da STN, o tema das transferências das receitas é considerado de difícil compreensão, tanto para os cidadãos quanto para os que estão envolvidos no contexto [1]. Cada uma das receitas possui particularidades distintas: legislação, cálculos de coeficiente para distribuição e periodicidade *etc.* Para ilustrar a importância e relevância do tema, como um exemplo, produziu-se o Gráfico 1, com base no Balanete da Receita [4]¹, em que a hegemonia das Transferências Correntes impera com 58%, dentre elas identificam-se que 53% contemplam a soma das Transferências da União (32%) e do Estado (21%). Ao se detalhar os 58%, ao nível do MTO da SOF [1], elaborou-se o Gráfico 2, no qual é perceptível a importância de algumas transferências em detrimento de outras.

Gráfico 2: Decomposição das Receitas de Transferências Correntes do município de São Luís/MA¹

Fica evidente a importância desta pesquisa, dada a dependência financeira dos municípios por estes repasses. Por esse motivo as Receitas transferidas escolhidas para este trabalho foram: Fundo de Participação dos Municípios (FPM), Compensação Financeira de Recursos Minerais (CFEM) e sua variante dos 15% da CFEM dos Municípios afetados por atividades de Mineração (CFEM AFETADOS), Fundo Especial de Petróleo (FEP), ICMS, Imposto sobre a Propriedade de Veículos Automotores (IPVA) e Imposto sobre Produtos Industrializados das Exportações (IPI).

3.1.2 Método de previsão das receitas públicas.

De acordo com orientações da SOF, a previsão das receitas constitui um estágio ou etapa alinhada aos fenômenos econômicos

[24,7] cujos horizontes temporais, nos termos do Art. 12 da LRF, devem alcançar 2 anos subsequentes [3]. Todavia, quando se combina com o Art. 13, o qual constitui a obrigatoriedade de seguir metas bimestrais com as receitas desdobradas em sua classificação, a granularidade de previsão passa a ser mensal, estratificada por receita em atendimento aos ditames dos Arts. 8º (cronograma de execução mensal de desembolso) e 9º (execução de metas bimestrais) [3,7].

Compreendida a granularidade do horizonte temporal de previsão, faz-se mister comentar a equação (1), sugerida pela SOF para estimativa de receitas dos Órgãos da Administração Públicas [7,14,15,24]. Para melhor compreensão das legendas da equação (1), Albuquerque, Medeiros e Silva [7] lecionam que, excluindo o efeito legislação (fora do escopo deste estudo, tendo como exemplos recentes [7]) e a variação dos erros, os demais estão intimamente ligados aos indicadores econômicos, que influenciam na arrecadação de cada receita. Ainda sobre a equação (1), alguns Autores alegam que a SOF não detalha, muito menos elucida como devem ser aplicadas as variações [7,14].

$$Pm_t = Am_{t-1} * (1 + Ep_t) * (1 + Eq_t) * (1 + El_t) * (1 + E\varepsilon_t) \quad (1)$$

Legendas:

Pm_t = Previsão da receita mensal no tempo t ;

Am_{t-1} = Arrecadação mensal da receita para o período anterior

$(1 + Ep_t)$ = Variação do efeito preço no período t ;

$(1 + Eq_t)$ = Variação do efeito quantidade no período t ;

$(1 + El_t)$ = Variação do efeito legislação no período t ;

$(1 + E\varepsilon_t)$ = Variação do erro na previsão.

3.2 Seleção de Variáveis

Embora a literatura dissemine diversas definições para a tarefa de seleção de atributos a qual consiste na utilização de “critérios de seleção e algoritmos de busca distintos para avaliar e encontrar de forma heurística o subconjunto de atributos mais adequado.” [15]. Harrison [27] afirma que atributos irrelevantes e correlacionados prejudicam o modelo de AM, causando instabilidade nos coeficientes de uma regressão, ou mesmo na importância deles em modelos de árvores. Neste mesmo sentido, Baranauskas [28] preconiza que este processo de busca heurística, *Feature Subset Selection* (FSS), tem sua natureza determinada por quatro questões: ponto de partida, organização, estratégia e critério de parada. O Pesquisador divide o ponto de partida, relacionado ao critério do estado inicial dos atributos do *Dataset*, em três tipos: *forward* (inicialmente vazio), *backward* (completo, depois eliminando atributos) e *outward* (arranjo das anteriores, adotando um ponto e seguindo-se dele). A organização trata da forma como serão selecionados os atributos e as iterações realizadas. O critério de parada consiste em estabelecer um limite, um parâmetro para encerramento. Assim, as abordagens *Filtro*, *Wrapper* e *Embedded* (Embutida) são as fundamentais [29]. Em suma, Baranauskas [28] leciona que o Filtro utiliza o FSS para escolher os atributos

¹ Nota: Os valores foram totalizados de janeiro de 2010 a agosto de 2021, a conta dedutiva não foi descontada nas suas receitas originárias, com exceção do Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos

Profissionais da Educação (FUNDEB) sacado das transferências da União e dos Estados que foram deduzidos nas Transferências Multigovernamentais.

independente do algoritmo de indução, o *Embedded* o embute no algoritmo e o *Wrapper* aplica-o como uma caixa-preta.

Os *Ensembles* fazem parte da abordagem *Embedded* e os algoritmos desta abordagem possuem um método para listar a importância das variáveis do modelo. Como por exemplo, o *Random Forest*, *Gradient Boosting*, *Extreme Gradient Boosting* (XGBoost). O *Random Forest*, ou Floresta Aleatória, trabalha sobre a premissa da “sabedoria da multidão” combinando um grande número de árvores de decisão aleatórias cuja a variância é reduzida, provendo um resultado da previsão média de todas elas [27,30,31]. O *Gradient Boosting* é uma combinação de vários modelos de árvores de decisão pequenas com o intuito de melhorar a explicação da variação não efetivada pela árvore anterior, corrigindo os erros anteriores [30]. O XGBoost é um algoritmo vencedor de competições do *Kaggle*, cria uma árvore de decisão fraca, sendo seu foco prover melhorias com adição de outras árvores para ganhar em precisão e velocidade ao corrigir os erros dos resíduos usando o gradiente descendentes [27,30].

4 RELATO DO ESTUDO DE CASO

Nesta seção foram abordados os principais procedimentos, seguindo as etapas da metodologia CRISP-DM apoiando-se na revisão sistemática de literatura inglesa que o considera como um padrão em ciência de dados, nas pesquisas e na prática [32,29].

Em termos do estudo completo, a população abrangida compreenderá uma amostra de 8 Capitais que aceitaram participar desta pesquisa, a saber: Belém (PA), Belo Horizonte (MG), Campo Grande (MS), Fortaleza (CE), Maceió (AL), Palmas (TO), Salvador (BA) e São Luís (MA). Contudo aqui o relato será da inclusão dos dados apenas da última Capital, São Luís (MA), concentrando-se na fase de Preparação dos dados.

4.1 Preparação dos dados, ferramental e dataset

Todos os dados utilizados foram obtidos nos portais da transparência das três esferas governamentais, o que em alguns casos pode-se utilizar técnicas de *web scraping* e *crawling*, mas em outros precisa-se solicitá-los por e-mail, salvar em arquivos ou mesmo acessando uma Interface de Programação de Aplicação, do Inglês *Application Programming Interface* (API). Portanto, não envolvem informações sigilosas, uma vez que lidam com dados disponibilizados em meios eletrônicos de acesso público [33].

Foram priorizadas as ferramentas que não tivessem custos. Neste sentido, a ferramenta *Power BI Desktop* foi escolhida para unificar o *Dataset*, pelo fato de ser do tipo *self-service*² com interfaces amigáveis e curva de aprendizagem rápida, além de ter um dos melhores custo-benefício pelas pesquisas da *Garther* [34]. Para codificação, usando a linguagem *Python* na versão 3.9.10, foi escolhido o *Jupyter Notebook* por ter a uma versão online no *Google Colab*, facilitando o uso sem necessidade de instalação nos equipamentos das Secretarias de Finanças.

As amostras utilizadas neste estudo advêm dos Relatórios de Balancete das Receitas da Prefeitura de São Luís no período de 2010 a 2020 [4], em formato “pdf”, transformados em um *Dataset*. As variáveis foram incluídas nas séries temporais com um mês anterior a sua divulgação. Os dados foram unificados criando uma chave primária por data, sob a topologia estrela. Alguns dos dados divergem na periodicidade, não foram utilizados métodos de decomposição mensal por produzirem uma margem de erro [35]. Após a unificação do *Dataset*, ele foi transformado em um arquivo do tipo “csv”. Na exploração dos dados, gerou-se um arquivo “html” do relatório do *pandas profiling* [7] o qual analisou as 928 observações do *Dataset* em 85 variáveis e identificou 22 com percentual de zeros e 3 constantes que foram retiradas, restando 65 colunas, elencadas como seguem:

- *POPULACAO*: População Municipal série histórica do IBGE;
- *Per_EntreCapitais*: Percentual da População de São Luís em relação a todas as Capitais;
- *PerPopulacaoUF*: Percentual da População de São Luís em relação aos Municípios da Unidade da Federação (UF);
- *PIB_MenBR*: Produto Interno Bruto do Brasil;
- *PIB_Mun*: Produto Interno Bruto municipal;
- *RndPrCap_Mun*: Cálculo da renda per capita municipal;
- *VA_BrtAgro*: Valor adicionado bruto (VAB) da agropecuária;
- *VA_BrtIndust*: VAB da Indústria municipal;
- *VA_BrtServExclAdm*: VAB dos serviços, exceto defesa, administração, educação e saúde e seguridade social municipal;
- *VA_BrtAdm*: VAB da administração, defesa, educação e saúde públicas e seguridade social municipal;
- *VA_BrtTotal*: VAB total municipal;
- *VA_ImpostosProd*: Impostos sobre produtos municipais;
- *SalarioMinimo*: Valor do limite mínimo para os salários pagos;
- *DolarComercial*: Média entre o maior e o menor valor cotado, no fechamento, do dólar comercial (EUA) no mês;
- *PetoleoBrentBR*: Preço médio, entre a maior e menor cotação do mês, do petróleo Brent;
- *Gasolina_BR*: Preço médio da Gasolina comum divulgado pela Agência Nacional de Petróleo (ANP);
- *ResvInternac*: Reservas Internacionais, usada nos financiamentos e/ou estabilizar o mercado de câmbio;
- *Importacao*: Importação de bens: corresponde ao valor total de ingresso de mercadorias originárias do exterior por UF;
- *EXPT_SLZ_ProdIndQuim*: Valores dos Produtos das indústrias químicas ou conexas exportados por São Luís;
- *EXPT_SLZ_TOTAL*: Valor de todas as categorias de produtos exportados por São Luís;
- *Exportacao*: Exportação de bens: corresponde ao valor total das mercadorias vendidas ao exterior por UF;
- *ICMS*: Recolhimento do ICMS na UF;
- *ICMS_DvAtTrb*: Arrecadação da dívida ativa tributária da UF;
- *ICMS_SimpNacio*: Arrecadação do Simples Nacional da UF
- *ST INTERNO*: Arrecadação de Substituição Tributária de movimento interno da UF;

² Conceitua-se para os softwares em que “[...] suas ferramentas estão disponíveis para diversos níveis de uma organização: não só para profissionais de TI, como também para analistas de negócios e executivos [...]” ([34]).

- *ST_EXTERNO*: Modalidade de Substituição Tributária de movimento externo da UF;
- *ICMS_TOTAL*: Soma do ICMS com ICMS SimpNacio, mais *ST_INTERNO* e *ST_EXTERNO* da UF;
- *IPVA*: Arrecadação do IPVA recolhido na UF;
- *CFEM_Mun*: Arrecadação por Município produtor;
- *CFEM_MA*: Arrecadação da CFEM da UF de Produtores;
- *CFEM_Total*: Arrecadação da CFEM do Brasil;
- *ICMS_DvAtTrb_Muns*: ICMS_DvAtTrb paga no Município;
- *IPVA_Muns*: Arrecadação do IPVA recolhido no Município;
- *ICMS_Muns*: Arrecadação do ICMS recolhido no Município;
- *TOTAL_ArrecMun*: Total da arrecadação no Município;
- *Veic_PgtolIPVA*: Veículos emplacados no Município;
- *TxDesocupados*: Pessoas sem trabalho, maiores de 14 anos;
- *TxIPCA*: Índice de Preço ao Consumidor Amplo;
- *Tx_SELIC_Perd*: Taxa que o governo define o “custo do dinheiro”, baliza as taxas das instituições financeiras;
- *IGPM*: Índice Geral de Preços - Mercado;
- *INPC*: Índice Nacional de Preços ao Consumidor;
- *IBC_Br*: Índice de Atividade Econômica do Banco Central, atividades da indústria, dos serviços e da agropecuária;
- *IPP*: Índice de Preços ao Produtor, indústria geral, extrativas e de transformação;
- *PIM_PF*: Pesquisa Industrial Mensal - Produção Física Industrial, por seções e atividades;
- *PercResSP_Nom*: Percentual do resultado nominal, soma do resultado primário com os juros a paga;
- *PercResSP_Prim*: Percentual do resultado primário, diferença do quanto o governo gasta e arrecada;
- *SINAP_Alto*: Custo de projeto por m², por tipo de projeto e padrão de acabamento Alto;
- *SINAP_Baixo*: Mesmo SINAP, mas acabamento Baixo;
- *SINAP_Minimo*: Mesmo SINAP, mas acabamento Mínimo;
- *SINAP_Normal*: Mesmo SINAP, mas acabamento Normal;
- *Coef_CIFPM_SLZ*: Coeficiente distribuição do FPM São Luís;
- *Coef_IPM*: Índice de Participação dos Municípios, coeficiente de distribuição estadual da receita do ICMS de São Luís;
- *FinMA_MnRestPrim*: Resultados primários, componente não financeiros do resultado fiscal da dívida municipal, subtração do resultado com os juros, ambos nominais;
- *FinMA_Mn_JrNom*: Fluxo dos juros, sobre a dívida interna e externa valores nominais municipais;
- *FinMA_MnRestNom*: Corresponde a variação nominal dos saldos da dívida líquida, é composta pelos juros reais, dívida atualizada e resultado primário municipal;
- *Ajustes_MA_Munic*: Outros fluxos (ajustes) do Estado do Maranhão e municípios, valores em R\$ (milhões);
- *DvLq_Gov_MA*: Dívida líquida do Estado do MA;
- *DvLq_MA_ExcCapital*: Dívida líquida do MA, exceto a capital;
- *DvLq_MA_Munic*: Dívida líquida de todos os municípios do Estado do Maranhão, incluindo a capital;
- *DvLq_SLZ_MA*: Dívida líquida de São Luís;
- *RECEITA*: Categoria da série temporal: CFEM, FEP, FPM, CFEM AFETADOS, ICMS, IPVA e IPI;
- *x_1*: Valor do atributo “Realizado” no mês anterior;
- *x_2*: Valor do atributo “Realizado” de 2 meses anteriores;

- *x_3*: Valor do atributo “Realizado” de 3 meses anteriores;
- *Realizado*: Variável “**target**”, valor bruto da Receita executada.

Na variável alvo, “Realizado”, devido a inconsistência sequencial, detectou-se a ausência de valores em alguns meses, mas ajustou-se com auxílio do site do agente financeiro, inclusive *outliers* que eram valores de até quatro meses depositados em um mês. Assim, obteve-se 778 observações e os subconjuntos das receitas tiveram sua divisão no conjunto de treino e teste como elencados na Tabela 1. A divisão utilizou a função *train_test_split* do pacote *sklearn.model_selection*, com argumento ‘*shuffle*’ como ‘*False*’ para não misturar os dados, pois trata-se de série temporal.

Conforme legislações vigentes, os valores dos indicadores vigoram até que um novo índice seja publicado. Logo, os valores ausentes foram tratados anteriormente na unificação das observações do *Dataset*, usando técnica da vizinhança mais próxima, repetindo o elemento imediatamente anterior. Ademais, foi aplicado o teste de *Shapiro-Wilk*, da biblioteca *stats.shapiro* do pacote *scipy*, em que a hipótese de distribuição normal foi rejeitada para todas as variáveis, exceto a série do “*TxIPCA*”. A padronização dos dados foi aplicada após a divisão das sub amostras de treino e teste com a função *RobustScaler* da biblioteca *sklearn.preprocessing*, uma vez que este faz um tratamento dos outliers ao remover a mediana e dimensionar os dados conforme o intervalo de *quantis*.

Receita	Observações	Divisões dos conjuntos de treino			
		<i>Random Forest</i>	<i>XGBoost</i>	<i>Gradient Boosting</i>	<i>Modelo Combinado</i>
CFEM	126	60%	50%	60%	70%
FPM	132				
FEP, ICMS, IPI e IPVA	130*4				
TOTAL	778				

Tabela 1: Quantidade de observações por subconjunto das receitas e divisão nos métodos

4.2 Métricas e métodos de seleção de variáveis

Os erros de previsão utilizados foram obtidos no subconjunto de teste pela diferença do valor original e o previsto pelos modelos. Foram calculados utilizando o pacote *sklearn.metrics*. O R-Quadrado (R²) ou coeficiente de determinação, calcula a porcentagem da variância que os dados do modelo podem prever ou explicar [31]. O Erro Quadrático Médio (MSE) é a diferença entre o valor original e o predito, fazendo uma média do erro das previsões ao quadrado, quanto mais próximo de 1, melhor a previsão do modelo. A Raiz do Erro Quadrático Médio (RMSE) é utilizada quando há diferenças de unidade, sendo a raiz quadrada do MSE. O Erro Absoluto Médio (MAE) é o cálculo da média das distâncias entre os valores originais e previstos pelo modelo, quanto menor melhor. Deste modo, escolheu-se como critério de seleção de modelos o menor RMSE, embora as demais métricas tenham sido calculadas e transcritas nas tabelas, a que definiu as escolhas foi o RMSE [31,36].

Foram testados, como abordagem embutida, individualmente, três *Ensembles*: *XGBRegressor*, *RandomForestRegressor* e *GradientBoostingRegressor*, os dois primeiros do pacote *sklearn.ensemble* e o terceiro do *XGBoost*. Cada Ensemble foi submetido a um laço para treinar apenas os dados de uma receita por vez, dividido e escalonado o subconjunto, sendo submetido ao treino e teste, e separando para visualização gráfica as *features* que tivessem importância superior a 0,005.

Já o modelo combinado, utilizado para comparação com os *ensembles* anteriores, a adaptação do modelo de seleção de *feature importance* de Pan He e outros autores [37] cujos atributos são submetidos a dois laços crescentes. O primeiro inicia com dois atributos, incrementando o conjunto sequencialmente em uma variável “k”, um ponto de partida do tipo *outward* [28]. Com uma abordagem híbrida, este laço utilizou uma estratégia de filtro com a técnica de ranqueamento de atributos, com a intenção de avaliar a quantidade fixa de atributos, no algoritmo denominado por “seletor” e o critério de parada foi exaustivo [29,31].

Dentro do segundo laço de “k” atributos, para atingir o objetivo, utilizou-se a combinação de cinco algoritmos seletores: *KBest*, *LinearRegression*, *Ridge*, *Lasso* e *RandomForestRegressor*. Tal combinação ocorreu com os algoritmos de *Ensembles* citados. Ao realizar o treinamento, naturalmente os atributos ganham coeficientes que já são elegíveis para classificar a importância de cada um dos atributos. Neste momento, extraem-se do modelo as melhores *features* com auxílio do pacote *Feature Selection* da biblioteca *sklearn*. Todos os dados resultantes são armazenados em um *DataFrame* em que se pode fazer comparações da quantidade de melhores *features* em cada modelo com suas métricas de teste.

A intenção desta combinação de modelos não é a menor métrica, mas a que conseguir ter uma estabilidade de menores erros em cada subconjunto dos rótulos da “RECEITA” [37]. Assim, foram estipulados os critérios pautados na verificação das seguintes etapas: 1) Separou-se as dez menores métricas pelo RMSE e R2 [31]; 2) Plotou-se um gráfico de todos os resultados excluindo-se os seletores que tiveram maiores variância ou outliers extremos; 3) Retirou-se os outliers com o Método de *Tukey*; 4) Selecionou-se a lista do seletor com menor RMSE e com maior; 5) Escolheu-se o menor RMSE, mediano das quatro melhores *features* escolhidas; 6) Extrauiu-se a lista dos atributos escolhidos.

5 RESULTADOS

Considerando o foco do estudo como a seleção de atributos para ulterior previsão de receitas, os resultados preliminares dos modelos, considerados mais relevantes para serem apresentados se concentram nas Receitas da CFEM, do ICMS, do IPVA e do IPI. O experimento inicial usou o *Dataset* completo, com todas as variáveis, no horizonte trimestral, com o intuito de averiguar se o modelo conseguiria distinguir as que atendem aos efeitos preço e quantidade [3]. Ademais, o PCA foi aplicado apenas para avaliar se a redução de dimensionalidade poderia ser uma alternativa, contudo, ele criou, em média, 20 componentes para explicar 95%

dos subconjuntos das receitas, mas selecionou variáveis similares para a maioria das receitas com métricas diferentes, variáveis estas que diferem das indicações de estudos econométricos, por este motivo não foi incluído nas comparações.

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	4	0,460	0,284	0,211	0,744
<i>RandomForest</i>	1	0,447	0,239	0,200	0,758
<i>XGBoost</i>	16	4,612	2,500	21,267	-24,782
Mod. Combinado					
<i>Ridge-GradientBoosting</i>	3	0,199	0,145	0,040	0,946

Quadro 2: Resultados das métricas da Receita da CFEM

Para a Receita da CFEM, organizou-se o *ranking* das métricas, obtidas no Quadro 2, tendo, em 1º lugar o Modelo Combinado com o seletor *Ridge* aplicando as variáveis ao *GradientBoosting*, o qual escolheu 3 com RMSE de 0,199: “CFEM_Mun”, “ICMS” e “Coef_CIFPM_SLZ”. Dentre as posições do *ranking*, sente-se a falta de atributos relacionados aos efeitos preço [3]. Mas o efeito quantidade foi identificado pela 1ª variável mais importante ter sido a “CFEM_Mun”, uma vez que 60% do montante é repassado ao Município Produtor, conforme prescreve a legislação vigente³ [38]. A 2ª posição, faz um certo sentido, pois esta receita movimenta produtos e serviços refletindo no ICMS, mas a 3ª seria interessante investigar melhor, pois ela foi escolhida como redutora.

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	10	0,489	0,385	0,239	0,488
<i>RandomForest</i>	8	0,525	0,429	0,275	0,410
<i>XGBoost</i>	25	0,396	0,293	0,157	0,664
Mod. Combinado					
<i>KBest-XGBoost</i>	47	0,211	0,184	0,045	0,897

Quadro 3: Resultados das métricas da Receita do ICMS

Na amostra do ICMS, novamente o Modelo Combinado apresentou o melhor desempenho, ver Quadro 3, com 47 variáveis escolhidas pelo seletor *KBest* aplicados ao *XGBoost*, as quais foram: “CFEM_MA”, “IPVA_Muns”, “ICMS_DvAtTrb”, “CFEM_Total”, “VA_BrtAgro”, “EXPT_SLZ_TOTAL”, “IPP”, “x_2”, “Veic_PgtoIPVA”, “PIM_PF”, “x_1”, “EXPT_SLZ_ProdIndQuim”, “ICMS_Muns”, “ICMS”, “TOTAL_ArrecMun”, “Coef_CIFPM_SLZ”, “DvLq_Gov_MA”, “DvLq_SaoLuis_MA”, “DvLq_MA_Munic”, “TxDesocupados”, “IBC_Br”, “PerPopulacaoUF”, “PercResSP_Nom”, “DolarComercial”, “Coef_IPM”, “Gasolina_BR”, “ResvInternac”, “Importacao”, “PercResSP_Prim”, “VA_BrtIndust”, “VA_ImpostosProd”, “Exportacao”, “SalarioMinimo”, “POPULACAO”, “PIB_Mun”, “ICMS_SimpNacio”, “RndPrCap_Mun”, “SINAP_Alto”, “VA_BrtTotal”, “VA_BrtServExclAdm”, “SINAP_Normal”, “SINAP_Minimo”, “VA_BrtAdm”, “SINAP_Baixo”, “PIB_MenBR”, “ST_EXTERNO” e “ICMS_TOTAL”.

³ Inciso “VI”, do Art. 2º, da Lei 13.540/2017 [38].

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	9	0,656	0,326	0,430	0,489
<i>RandomForest</i>	6	0,540	0,320	0,292	0,653
<i>XGBoost</i>	22	0,609	0,343	0,371	0,560
Mod. Combinado					
<i>RandomForest</i>	36	0,499	0,343	0,249	0,729

Quadro 4: Resultados das métricas da Receita do IPVA

Embora o RMSE tenha sido de 0,211 (Quadro 3) e tenham-se *features* de efeito preço e quantidade, elas não atenderam às expectativas. Haja vista que as 8 primeiras colocações tiveram os maiores percentuais de importância, todavia, esperava-se que o algoritmo selecionasse os mais relevantes⁴ como “*ICMS_TOTAL*”, em último lugar, e o “*Coef_IPM*”, em 24^a, pois determinam a forma de cálculo dessa receita.

De acordo com o Quadro 4, o menor RMSE, de 0,499, foi do Modelo Combinado para a Receita de IPVA, selecionado pelo *RandomForest*, aplicado ao *GradientBoosting*, elegendo-se 36 atributos: “*IPVA*”, “*IPVA_Muns*”, “*EXPT_SLZ_TOTAL*”, “*x_3*”, “*FinMA_MnRestPrim*”, “*CFEM_Mun*”, “*IGPM*”, “*x_1*”, “*CFEM_MA*”, “*Tx_SELIC_Perd*”, “*x_2*”, “*FinMA_MnRestNom*”, “*TxIPCA*”, “*EXPT_SLZ_ProdIndQuim*”, “*Importacao*”, “*INPC*”, “*PetoleoBrentBR*”, “*PIM_PF*”, “*CFEM_Total*”, “*ICMS_DvAtTrb*”, “*ICMS_DvAtTrb_Muns*”, “*DolarComercial*”, “*ResvInternac*”, “*ST_EXTERNO*”, “*TOTAL_ArrecMun*”, “*TxDesocupados*”, “*ST_INTERNO*”, “*Veic_PgtoIPVA*”, “*VA_BrtAgro*”, “*ICMS_Muns*”, “*FinMA_Mn_JrNom*”, “*Coef_CIFFM_SLZ*”, “*ICMS_SimpNacio*”, “*IBC_Br*”, “*SINAP_Baixo*” e “*SINAP_Minimo*”. Destaca-se, neste ranking que a 1^a posição ficou com o “*IPVA*” (80,63%) e a 2^a para o “*IPVA_Muns*” (7,35%). Contudo, este último é o que define o repasse, pois 50% da arrecadação⁵ dos veículos emplacados no Município retornam para ele [2].

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	30	1,270	0,981	1,613	-1,106
<i>RandomForest</i>	8	1,122	0,874	1,260	-0,645
<i>XGBoost</i>	42	1,099	0,793	1,208	-0,577
Mod. Combinado					
<i>Lasso</i>	7	0,872	0,674	0,760	0,286

Quadro 5: Resultados das métricas da Receita da IPI

Considerando o ranking do Modelo Combinado para a receita de IPI, no Quadro 5, obteve-se o RMSE de 0,872 referente ao Modelo Combinado com seletor *Lasso* e *GradientBoosting*, escolhendo 7 variáveis: “*x_3*”, “*x_2*”, “*x_1*”, “*DolarComercial*”, “*ICMS*”, “*ICMS_DvAtTrb*” e “*ICMS_TOTAL*”. Ao analisá-las, até a 4^a, atenderia ao efeito quantidade [3] e a recursividade dos repasses, mas as outras não se encaixam bem, também não apresentou *features* do efeito preço.

6 CONCLUSÕES

A previsão de receitas públicas no nível municipal é bastante relevante para a administração pública no Brasil, pois define o orçamento das políticas públicas. Equívocos nesta predição possuem impactos deletérios para sociedade. Visando mitigar eventuais erros, o presente trabalho apresentou um estudo para seleção de atributos a fim de melhorar a previsão de receitas públicas usando como estudo de caso a capital maranhense, São Luís. Os resultados obtidos demonstram a viabilidade e potencialidade das análises conduzidas. Convém destacar que as ocorrências aqui relatadas, de divergências e ausências identificadas na série temporal do IPI foram comunicadas ao Gestor Municipal o qual notificou tanto o agente financeiro quanto o Governo do Estado do Maranhão. O fato foi de tal relevância que demandou a criação de um painel no Power BI Desktop para monitorar o evento. Assim, este ano, foi identificado que em janeiro ocorreu novamente tais ausências de depósito, mas em 21 de junho os valores foram repassados ao erário público⁶.

Nos resultados foi possível também analisar o reflexo destes eventos, pois mesmo o Modelo Combinado obtendo a supremacia na métrica do RMSE, corroborando com o estado da arte que já consagra esta abordagem por melhorar a precisão das previsões [22], as variáveis escolhidas não possuem uma representatividade aparente em relação às explicações de sua origem com as de todas as receitas, com exceção parcial da CFEM Produtores. Foi possível identificar o efeito quantidade, mas o do preço não [3]. Neste sentido, será necessário voltar à etapa de Pré-Processamento para reavaliar as séries temporais do *Dataset*, testando a inclusão dos dados das outras capitais. Ademais, sugere-se examinar a inclusão no modelo combinado de uma abordagem *stacking* para o efeito preço e aplicar o teste de Friedman, com nível de significância de 5%, seguido do teste *post-hoc* de *Nemenyi* com o intuito de criar uma lista para comparar os resultados mais significativos [40].

Por fim, para etapas posteriores do processo tem-se a intenção de submeter as variáveis selecionadas à predição dos modelos estatísticos clássicos e de AM, incluindo as sugestões de trabalho futuros de Carmo, Boldt e Komati [8], para ulteriores comparativos.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - DT - 308334/2020; pela Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) - PRONEM-FAPESPA/CNPq n° 045/2021. Agradecemos também aos revisores(as) pelas sugestões que muito auxiliaram na melhora do trabalho.

REFERÊNCIAS

- [1] BRASIL, SOF. Classificação da Receita. In Manual Técnico do Orçamento - MTO Online. Brasília, 2022.
- [2] BRASIL. Constituição Federal. 1988.
- [3] BRASIL. Lei Complementar n° 101. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências (Mato

⁴ Inciso “IV”, do art. 158, da Constituição Federal [2].

⁵ Inciso “III”, do art. 158, da Constituição Federal [2].

⁶ Consulta no Portal da Transparência, disponível em: <https://www42.bb.com.br/portalbb/da/beneficiarioList,802,4647,4652,0,1,1.bb?cid=87121>.

- 4, 2000).
- [4] SÃO LUÍS-MA. *Receitas Total e detalhada* (2021).
- [5] Wirth, Rüdiger and Hipp, Jochen. *Crisp-dm: towards a standard process modell for data mining* (2000), 11.
- [6] Naibert, Nelson Terres. *Orçamento público: metodologia alternativa para previsão de receitas municipais. Dissertação (Mestrado em Administração Pública) - FGV, Rio de Janeiro, 2021.*
- [7] Pamplona, Edgar, Fiirst, Clóvis, Hein, Nelson, and Zonatto, Vinicius Costa da Silva. *Desempenho do Modelo ARMA na Previsão das Receitas Orçamentárias dos Municípios do Estado do Paraná (Jan-mar 2019).*
- [8] Carmo, Marcelo Magalhães do, Boldt, Francisco de Assis, and Komati, Karin Satie. *Previsão de receitas de ICMS do estado do Espírito Santo através de seleção de características em cascata e técnicas de Aprendizado de Máquina. In Anais do XVI Encontro Nacional de Inteligência Artificial de Aprendizagem de Máquina (Salvador/BA 2019), 118-129.*
- [9] Silva, Liniker Fernandes da, Jacovine, Laércio Antônio Gonçalves, Silva, Marcio Lopes da, Isbaex, Crismeire, and Rego, Lyvia Julienne Sousa. *Correlação das Variáveis Socioeconômicas e Ambientais com royalties Petrolíferos e CFEM Municipais (Setembro 28, 2017).*
- [10] Rodrigues, Rafael Ramon Fonseca, Aquino, Italo Robean Praxedes de, Diniz, Saulo Medeiros, Cirne, Gianinni Martins Pereira, Cabral, Luiz Marcelo Martins do Amaral Carneiro, and Carvalho, José Ribamar Marques de. *Séries Temporais de Holt-Winters na análise preditiva da arrecadação do ICMS no estado do Rio Grande do Norte. In Anais do XXVII Congresso Brasileiro de Custos - ABC (9-11 Novembro 2020), 14.*
- [11] Sousa, Katsuk Rodrigues, Correia, José Jonas Alves, Santos, Ramon Rodrigues dos, and Santos, Josete Florêncio dos. *Previsão de Receitas tributárias: uma análise do icms no ceará por meio de séries temporais (Junho 2019).*
- [12] Silva, Alexandre César Batista da, Albuquerque, Francivaldo dos Santos, Bezerra, Maria do Socorro Coelho, and Silva, Wilton Bernardino da. *Utilização de Séries Temporais na Previsão de Arrecadação de ICMS no Estado do Pará (2016).*
- [13] Azevedo, Ricardo Rocha de, Silva, José Marcos da, and Gatsios, Rafael. *Análise crítica dos modelos de previsão de série temporal com base no ICMS estadual. (2017).*
- [14] Silva, Alexandra Maria da, Shiaku, Aline Cristine, Lopes, José Eduardo Ferreira, and Pereira, Vinicius Silva. *Orçamento Público, [Como Prever as Receitas?] (Outubro 19 a 31, 2018).*
- [15] Almeida, Maria Eduarda da Silva and Ferreira, Gleidson Ramos. *Métodos de acompanhamento e previsão da receita pública: um estudo de município do Recife. Revista Capital Científico - Eletrônica (RCCe), 16, 2 (Abril/Junho 2018).*
- [16] Fiirst, Clóvis, Pamplona, Edgar, Zonatto, Vinicius Costa da Silva, Rodrigues Junior, Moacir Manoel, and Hein, Nelson. *Evidências sobre a previsão orçamentária da receita pública nos municípios do estado do Paraná: um estudo com a aplicação do Modelo de Koyck (Jan-Jun 2018).*
- [17] An, Zidong and Jalles, John Tovar. *"On the performance of US fiscal forecasts: government vs. private information" (2020).*
- [18] Mikesell, John L. *Often Wrong, Never Uncertain: Lessons from 40 Years of State Revenue Forecasting (Maio 28, 2018).*
- [19] Noch, Muhammad Yamin, Pattiasina, Victor, and Purwanto, Hafidz Arif. *Elasticity Measurement of Local Taxes and Charges In Forecast Of Own-Source Revenue (PAD) Of Provincial Government In Indonesia (Janeiro 11, 2018).*
- [20] Lima, Luciana Leite, Lui, Lizandro, Dias, Gianna Vargas Reis Salgado, Papi, Luciana Pazini, and Demarco, Diogo Joel. *Planejamento governamental nos municípios brasileiros: em direção a uma agenda de pesquisa. Cadernos EBAPE.BR, 18 (Abril-Junho 2020), 323-335.*
- [21] Aguiar, Andson Braga de. *O pequeno mundo da pesquisa em contabilidade gerencial no Brasil: discussão sobre desenhos alternativos de pesquisa. Revista De Contabilidade E Organizações, 12, 151933 (Novembro 2018), 16.*
- [22] Petropoulos, Fotios, Apiletti, Daniele, Assimakopoulos, Vassilios et al. *Forecasting: theory and practice. International Journal of Forecasting, 38 (2022), 705-871. ISSN: 0169-2070.*
- [23] BRASIL. *Lei Nº 4.320, de 17 de março de 1964. Estatui Normas Gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União, dos Estados, dos Municípios e do Distrito Federal (Março 17, 1964).*
- [24] BRASIL, MINISTÉRIO DA FAZENDA. *Receitas Públicas: manual de procedimentos aplicados à União, aos Estados, ao Distrito Federal e aos Municípios. STN/Coordenação Geral de Contabilidade, Brasília, 2007.*
- [25] BRASIL, MINISTÉRIO DA FAZENDA. *Princípios Básicos: o que você precisa saber sobre as transferências fiscais da União. Presidência da República, Brasília, DF, mar. 2016.*
- [26] Black, Clarissa. *Preços das commodities, termos de troca e crescimento econômico brasileiro nos anos 2000. Indicadores Econômicos FEE, 42 (2015), 27-44.*
- [27] Harrison, Matt. *Machine Learning guia de referência rápida: trabalhando com dados estruturados em Python. Novatec, São Paulo/SP, 2020.*
- [28] Baranauskas, José Augusto and Monard, Maria Carolina. *The MLC++ Wrapper for Feature Subset Selection Using Decision Tree, Production Rule, Instance Based and Statistical Inducers: Some Experimental Results. 87, ICMC-USP, São Carlos-SP, 1999.*
- [29] Almeida, Thissiany Beatriz. *Seleção de atributos usando abordagem Wrapper para classificação hierárquica multirrótulo. Dissertação (Mestrado em Ciência da Computação) - UTFPR, Ponta Grossa-PR, 2018. Orientadora: Profa. Dra. Helyane Bronoski Borges.*
- [30] Korstanje, Joos. *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR. Apress, Maisons Alfort, France, 2021.*
- [31] Bruce, Peter and Bruce, Andrew. *Estatística Prática para Cientistas de Dados: 50 conceitos essenciais. Alta Books, Rio de Janeiro-RJ, 2019.*
- [32] Schröer, Christoph, Kruse, Felix, and Gómez, Jorge Marx. *A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science, 181 (2021), 526-534.*
- [33] BRASIL. *Lei Complementar nº 131, de 27 de maio de 2009. Acrescenta dispositivos à Lei Complementar no 101, de 4 de maio de 2000 que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, a fim de determinar a disponibilização, em tempo real, de inform (2009).*
- [34] Lago, Karine and Alves, Laennder. *Fundamentos do Business Intelligence, Introdução ao Power BI. In Dominando o Power BI. DATAB Inteligência e Estratégia, São Paulo, 2019.*
- [35] Santos, Thiago Trafane Oliveira. *Contribuição dos Componentes para o Crescimento Real do PIB Trimestral no Contexto de Índices de Laspeyres de Base Móvel: uma decomposição analítica. Nota Técnica do Banco Central do Brasil nº 46. (Agosto 2018), 11.*
- [36] Hyndman, Rob J. and Athanasopoulos, George. *Forecasting: principles and practice. OTexts, Melbourne, Australia, 2018.*
- [37] He, Xinran, Pan, Junfeng, Jin, Ou et al. *Practical Lessons from Predicting Clicks on Ads at Facebook. In Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (ADKDD'14) (New York, NY, USA 2014), Association for Computing Machinery, 1-9.*
- [38] BRASIL. *Lei nº 13.540, de 18 de dezembro de 2017. Altera as Leis nº 7.990, de 28 de dezembro de 1989, e 8.001, de 13 de março de 1990, para dispor sobre a Compensação Financeira pela Exploração de Recursos Minerais (CFEM) (Dezembro 2017).*
- [39] Géron, Aurélien. *Mãos à Obra Aprendizado de Máquina com Scikit-Learn & TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes. Alta Books, Rio de Janeiro, 2019.*
- [40] Parmezan, Antonio Rafael Sabino, Souza, Vinicius M. A., and Batista, Gustavo E.A.P.A. *Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information Sciences, 484 (Maio 2019), 302-337.*
- [41] Proque, Andressa Lemes, Betarelli Junior, Admir Antonio, Silva, Libania Araujo, and Moraes, Ana Maria. *Efeitos econômicos da nova política tributária sobre a demanda de combustíveis para a economia brasileira. XXVI Encontro Regional de Economia (Outubro 27 e 28, 2021).*
- [42] Duarte, Guilherme Jardim, Souza, Pedro H. G. Ferreira de, Chiavegatto Filho, Alexandre Dias Porto et al. *Guia brasileiro de análise de dados: armadilhas & soluções. Enap, Brasília/DF, 2021.*
- [43] Sekar, Maris. *Machine Learning for Auditors: Automating Fraud Investigations Through Artificial Intelligence. Apress, Calgary, AB, Canada, 2022.*
- [44] Albuquerque, Claudiano Manoel de, Medeiros, Márcio Bastos, and Silva, Paulo Henrique Feijó da. *Gestão de Finanças Públicas: fundamentos e práticas de planejamento, orçamento e administração financeira com responsabilidade fiscal. Gestão Pública Editora e Treinamentos, Brasília, DF, 2013.*