

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E
SISTEMAS

**APLICAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA COM
SELEÇÃO DE VARIÁVEIS NA PREVISÃO DE RECEITAS PÚBLICAS DE
CAPITAIS BRASILEIRAS:**

Estudo de caso das receitas transferidas de São Luís

CLÁUDIA PATRÍCIA SILVA PIMENTEL

SÃO LUÍS - MA

2023

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E
SISTEMAS

CLÁUDIA PATRÍCIA SILVA PIMENTEL

**APLICAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA COM
SELEÇÃO DE VARIÁVEIS NA PREVISÃO DE RECEITAS PÚBLICAS DE
CAPITAIS BRASILEIRAS:**

Estudo de caso das receitas transferidas de São Luís

Dissertação apresentada ao curso de Mestrado Profissional em Engenharia da Computação e Sistemas na Universidade Estadual do Maranhão como pré-requisito para obtenção do título de Mestre sob orientação do Prof.º Msc. Antônio Fernando Lavareda Jacob Jr e coorientação do Prof.º Dr. Fábio Manoel França Lobato.

SÃO LUÍS - MA

2023

Pimentel, Cláudia Patrícia Silva.

Aplicação de técnicas de aprendizagem de máquina com seleção de variáveis na previsão de receitas públicas de capitais brasileiras: estudo de caso das receitas transferidas de São Luís / Cláudia Patrícia Silva Pimentel.

– São Luís, 2023.

146 f

Dissertação (Programa de Pós-Graduação em Engenharia de Computação e Sistemas) - Universidade Estadual do Maranhão, 2023.

Orientador: Prof. Me. Antônio Fernando Lavareda Jacob Júnior.

Coorientador: Prof. Dr. Fábio Manoel França Lobato.

1. Previsão de receitas públicas. 2.Transferências constitucionais e legais. 3.Aprendizagem de máquina. 4.Seleção de variáveis. 5.CRISP-DM. I.Título.

CDU: 681.5:[336.2:352](812.1)

CLÁUDIA PATRÍCIA SILVA PIMENTEL

**APLICAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA COM
SELEÇÃO DE VARIÁVEIS NA PREVISÃO DE RECEITAS PÚBLICAS DE
CAPITAIS BRASILEIRAS:**

Estudo de caso das receitas transferidas de São Luís

Dissertação apresentada ao Programa de Pós-graduação em Engenharia da Computação e Sistemas (curso de Mestrado Profissional) da Universidade Estadual do Maranhão (UEMA), como pré-requisito para obtenção do título de Mestre em Engenharia da Computação e Sistemas.

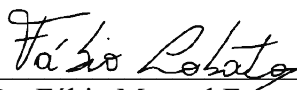
Aprovada, 10 de fevereiro de 2023.



Prof.º Dr. Ricardo Marcondes Marcacini
Examinador Externo
ICM/USP



Prof.º Dr. Omar Andres Carmona Cortes
Examinador Interno
PECS/IFMA



Prof.º Dr. Fábio Manoel França Lobato
Coorientador
PECS/UFOPA



Digitally signed by ANTONIO FERNANDO
LAVAREDA JACOB JUNIOR:78522552215
Reason: I attest to the accuracy and
integrity of this document
Location: São Luís-MA
Date: 2023-04-13 20:48-03:00

Prof.º Msc. Antônio Fernando Lavareda Jacob Junior
Orientador
UEMA

*Ao Deus da minha salvação que me concedeu um segundo sopro de vida, a meus pais, a
Natália, minha filha, por me presentear com Luís Gabriel.*

*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou
sobre aquilo que todo mundo vê.”*

Arthur Schopenhauer

AGRADECIMENTOS

O biênio deste Mestrado foi um retorno a uma antiga rota da trajetória de minha vida que exigiu muito esforço, dedicação e sacrifícios pessoais, no que tange a administração do tempo. Por isso quero deixar registro o meu profundo agradecimentos as pessoas que tiveram uma parcela de contribuição ou mesmo participaram junto comigo de todo este processo.

À Deus pelas multiformes surpresas que Ele tem me proporcionado, mesmo em tantos momentos de dor e sofrimento, os quais forjaram uma nova criatura em mim, mais resiliente e convicta de que “todas as coisas cooperam para o bem daqueles que O amam e seguem os Seus propósitos”, pois, dias melhores virão sob a lente do olhar do coração: muito obrigada!

Aos meus pais, Nocrato e Geny, por estarem comigo, incondicionalmente, principalmente pelos “toques de recolher” e pelos “lanchinhos”, uma vida em família: muito obrigada!

A minha eterna “princesa”, Natália, e meu genro, Eder Marques, que mesmo em meio aos seus protestos amorosos, foram compreensivos e trilharam junto comigo esta caminhada: muito obrigada!

Ao meu irmão, Carlos Pimentel, pelas palavras de incentivo, pelos momentos que deixou seu descanso para fazer cursos comigo, pelas sugestões, pelas correções de códigos, pelas “traduções simultâneas dos cursos ministrados por indianos”, por tudo, pois não tenho palavra para relatar, nem como mensurar o tamanho da sua contribuição: muito obrigada!

A minha amada amiga e assistente técnica, Elivania Ferreira, por sua dedicação, por ter me inscrito no processo e insistido que eu tentasse, pois quando nem eu mesma acreditava em mim ela o fez, sempre com seu tom bem-humorado: muito obrigada!

A Nair Goytacaz, grande profissional e amiga de trabalho, pelo incentivo e revisão do pré-projeto de inscrição quando eu não queria submetê-lo: muito obrigada!

A José Rodrigues, amigo, ex-Secretário Municipal de Fazenda, por acreditar na minha capacidade profissional, pelas orientações e ensinamentos e ter me integrado no contexto das Receitas Transferidas da Secretaria de Fazenda: muito obrigada!

A Délcio Rodrigues, ex-Secretário de Fazenda, por ter assumido a pasta e promovido a Chefe das Receitas transferidas, proporcionando mais oportunidade de aprendizado: muito obrigada!

A José Azzolini, atual Secretário Municipal de Fazenda, que autorizou a flexibilização do meu horário para assistir às aulas: muito obrigada!

Aos meu Orientador, Prof. Msc. Jacob, por ter me concedido a oportunidade de ser sua orientanda, pela sua compreensão com a fragilidade da minha saúde e suas sábias sugestões: muito obrigada!

Ao meu Coorientador, Prof. Dr. Fábio, por ter me acolhido, pelas excelente aulas e sugestões, bem como pelos chamados ao foco quando eu pensava em desviar: muito obrigada!

Ao Prof. Dr. Omar Carmona, pois quando pensei que não conseguiria desenvolver mais nada, ele ministrou o curso da Huawei, um divisor de águas para aguçar as minhas ideias: muito obrigada!

Aos demais Professores e colegas que sempre contribuem com uma parte do aprendizado: muito obrigada!

Aos meus Pastores, Fernando e Kátia Mendes, pelo incentivo, pela compreensão, pelo apoio moral e espiritual, durante todo os anos que passei a me congregar com eles: muito obrigada!

A minha turma do discipulado, dos adolescentes e dos jovens, nos momentos em que tive que me ausentar, ou ser substituída por outro professor na Escola Bíblica Dominical: muito obrigada!

A minha turma de trabalhos voluntários de resgate no aprendizado de Excel para o dia a dia, pela compreensão do cancelamento das aulas para cursar este mestrado: muito obrigada!

Ao Marcos Vinícius, pela ajuda com o código de leitura dos arquivos em “pdfs”, o qual adiantou muito o meu trabalho: muito obrigada!

A todos aqueles que contribuíram direta ou indiretamente com esta retomada aos estudo e a área de tecnologia, outrora abandonada: muito obrigada!

RESUMO

Considera-se a previsão de receitas de grande relevância para os tomadores de decisão, bem como para o planejamento. Quando se trata do campo de aplicação prática, voltado para o Setor Público, no que tange a esfera municipal, observam-se distorções entre os valores orçados e previsto, mesmo aplicando as regras previstas na legislação vigente. A questão tem sido investigada por pesquisadores com uma trajetória de avanços de métodos de regressão estatísticos e as aplicações de técnicas de aprendizagem de máquina, mas a problemática das divergências das previsões continua e a legislação exige justificativas. Neste contexto, faz-se necessário investigar se os efeitos preço e quantidade podem ser identificados por técnicas de aprendizagem de máquina e os erros de previsão das receitas poderiam ser mitigados se as variáveis fossem usadas pelo regime de competência do ingresso de recursos. Neste sentido, esta pesquisa tem o objetivo de realizar um estudo de caso, com os dados de São Luís, para escolher as variáveis que atendam as prerrogativas legais, adotando a metodologia CRISP-DM, por meio da comparação da lista de importância de algoritmos *ensembles*, *Random Forests*, *Gradient Boosting* e *XGBoost*, com um modelo combinado das abordagens de filtro com *wrapper*, submetendo-os aos mesmos algoritmos para escolher as variáveis com menores métricas de avaliação dentro de uma sequência de menores erros das receitas transferidas. No documento foram relatadas, como um comparativo de execução, as etapas e tarefas do CRISP-DM em sua primeira iteração, utilizando os dados dos Portais da Transparência, no período de 2010 a 2021. Nos resultados, comparou-se dois conjuntos de dados, um com todos os repasses, incluindo os valores extraordinários e outro com apenas as cotas oficiais. O Modelo Combinado obteve, na maioria dos resultados, as melhores métricas, especialmente, nos repasses extraordinários, corroborando com o estado da arte que já consagra esta abordagem, mas a aplicação do teste de Friedman não descartou a hipótese nula, pois as métricas dos dois conjuntos não apresentaram diferenças significativas. Na modelagem a RNN foi complexa obteve a melhor métrica, todavia, com exceção dos recursos da Mineração, a diferença dos valores foi melhor em outros algoritmos e o teste de Friedman também não teve diferenças significativas. Como resposta a questão de pesquisa foi possível identificar com clareza o efeito quantidade nos dois conjuntos de dados, mas o do preço não foi tão evidente nos resultados, aparecendo mais quando se testou apenas as cotas oficiais.

Palavras-chave: Previsão de Receitas pública, Transferências Constitucionais e Legais, Aprendizagem de Máquina; Seleção de variáveis; CRISP-DM.

ABSTRACT

Revenue forecasting is considered of great relevance for decision makers, as well as for planning. When it comes to the field of practical application, aimed at the Public Sector, regarding the municipal sphere, distortions are observed between the budgeted and predicted values, even applying the rules provided for in current legislation. The issue has been investigated by researchers with a history of advances in statistical regression methods and the application of machine learning techniques, but the problem of forecast divergences continues and legislation requires justification. In this context, it is necessary to investigate whether the price and quantity effects can be identified by machine learning techniques and whether revenue forecasting errors could be mitigated if the variables were used on an accrual basis for the inflow of resources. In this sense, this research aims to carry out a case study, with data from São Luís, to choose the variables that meet the legal prerogatives, adopting the CRISP-DM methodology, by comparing the list of importance of ensembles algorithms , Random Forests, Gradient Boosting and XGBoost, with a combined model of the wrapper filter approaches, submitting them to the same algorithms to choose the variables with the lowest evaluation metrics within a sequence of lowest errors of the transferred recipes. The document reports, as an execution comparison, the steps and tasks of the CRISP-DM in its first iteration, using data from the Transparency Portals, in the period from 2010 to 2021. In the results, two sets of data were compared, one with all transfers, including extraordinary amounts and another with only official quotas. The Combined Model obtained, in most of the results, the best metrics, especially in the extraordinary transfers, corroborating with the state of the art that already enshrines this approach, but the application of the Friedman test did not discard the null hypothesis, since the metrics of the two sets showed no significant differences. In the modeling, the RNN was complex and obtained the best metric, however, with the exception of the Mining resources, the difference in values was better in other algorithms and the Friedman test also did not have significant differences. As an answer to the research question, it was possible to clearly identify the quantity effect in both sets of data, but the price effect was not so evident in the results, appearing more when only the official quotas were tested.

Keywords: *Public Revenue, Constitutional and Legal Transfers, Machine Learning; Deep Learning; CRISP-DM.*

LISTA DE GRÁFICOS

Gráfico 1 - Composição das receitas no orçamento do município de São Luís/MA	29
Gráfico 2 – Decomposição das Receitas de Transferências Correntes do município de São Luís/MA.....	30
Gráfico 3 - Distribuição por ano e tipo de abordagem das publicações no período de 2015 até agosto de 2021	43
Gráfico 4 - Comparativo dos valores das Predições x Originais da Receita da CFEM nos testes de cotas oficiais	66
Gráfico 5 - Comparativo dos valores das Predições x Originais da Receita do ICMS nos testes de cotas oficiais	67
Gráfico 6 - Comparativo dos valores das Predições x Originais da Receita do IPI nos testes de cotas oficiais	70
Gráfico 7 - Comparativo dos valores das Predições x Originais da Receita do FEP nos testes de cotas oficiais	71
Gráfico 8 - Comparativo dos valores das Predições x Originais da Receita do FPM nos testes de cotas oficiais	73
Gráfico 9 - Comparativo dos valores das Predições x Originais da Receita do IPVA nos testes de cotas oficiais	74

LISTA DE FIGURAS

Figura 1 - Diagrama da estrutura do CRISP-DM com as suas 6 fases.....	40
Figura 2 - Tarefas genéricas (em negrito) e os artefatos (em itálico) das fases do CRISP-DM	41
Figura 3 - Esquema do plano de projeto inicial.....	52
Figura 4 - Esquema do algoritmo do modelo combinado	58
Figura 5 - Painel de monitoramento de ingresso de recursos.....	64
Figura 6 - Diagrama de entidade relacional usado pela API	84

LISTA DE QUADROS

Quadro 1 - Estrutura da codificação da receita por dígito e significado	28
Quadro 2 - Classificação econômica da receita orçamentária	28
Quadro 3 - Listagem das séries históricas por Capital, ano de início, protocolo das suas Ouvidorias e respostas	49
Quadro 4 - Avaliação das etapas concluídas do projeto inicial	50
Quadro 5 – Listagem dos Modelos que foram planejados e executados na primeira iteração .	60
Quadro 6 - Resumo comparativo da execução do plano inicial de implantação	63
Quadro 7 - Resultados das métricas da seleção das variáveis da Receita da CFEM de cotas oficiais	65
Quadro 8 - Resultados das métricas da seleção das variáveis da Receita da CFEM com repasses extraordinários	66
Quadro 9 - Resultados das métricas da seleção das variáveis da Receita do ICMS de cotas oficiais	67
Quadro 10 - Resultados das métricas da seleção das variáveis da Receita do ICMS com repasses extraordinários	68
Quadro 11 - Resultados das métricas da seleção das variáveis da Receita do IPI com repasses extraordinários	68
Quadro 12 - Resultados das métricas da seleção das variáveis da Receita do IPI de cotas oficiais	69
Quadro 13 - Resultados das métricas da seleção das variáveis da Receita do FEP de cotas oficiais	70
Quadro 14 - Resultados das métricas da seleção das variáveis da Receita do FEP com repasses extraordinários	71
Quadro 15 - Resultados das métricas da seleção das variáveis da Receita do FPM de cotas oficiais	72
Quadro 16 - Resultados das métricas da seleção das variáveis da Receita do FPM com repasses extraordinários	72
Quadro 17 - Resultados das métricas da seleção das variáveis da Receita do IPVA de cotas oficiais	73
Quadro 18 - Resultados das métricas da seleção das variáveis da Receita do IPVA com repasses extraordinários	74

Quadro 19 - Resultado do treinamento dos algoritmos de AM para a CFEM	76
Quadro 20 - Resultado do treinamento dos algoritmos de AM para o ICMS	77
Quadro 21 - Resultado do treinamento dos algoritmos de AM para o FEP	78
Quadro 22 - Resultado do treinamento dos algoritmos de AM para o FPM.....	79
Quadro 23 - Resultado do treinamento dos algoritmos de AM para o IPI.....	80
Quadro 24 - Resultado do treinamento dos algoritmos de AM para o IPVA.....	81
Quadro 25 - Lista de rotas da API implementada para	85

LISTA DE TABELAS

Tabela 1 - Comparativo da execução financeira e orçamentária inicial das Receitas Transferidas, em São Luís	21
Tabela 2 - Percentual de repasses dos recursos do FPM aos Municípios.....	32
Tabela 3 - Resumo dos trabalhos correlatos por Autor, método, técnicas e local.....	44
Tabela 4 - Distribuição quantitativa e percentual das 928 observações por receita	54
Tabela 5 - Mensuração do IPI em uma <i>pivot table</i> com disposição anual, mês a mês da publicação contábil	56
Tabela 6 - Distribuição das 906 observações por receita após ajustes da fase de pré-processamento	56
Tabela 7 - Quantidade de observações por subconjunto das receitas e divisão nos métodos ..	57
Tabela 8 - Diferenças entre Valores preditos e realizados dos treinos para a CFEM	77
Tabela 9 - Diferenças entre Valores preditos e realizados dos treinos para o ICMS	78
Tabela 10 - Diferenças entre Valores preditos e realizados dos treinos para o FEP	79
Tabela 11 - Diferenças entre Valores preditos e realizados dos treinos para o FPM	80
Tabela 12 - Diferenças entre Valores preditos e realizados dos treinos para o IPI	81
Tabela 13 - Diferenças entre Valores preditos e realizados dos treinos para o IPVA.....	82
Tabela 14 - Resumo das menores diferenças entre os valores previsto e realizados no primeiro bimestre de 2022	82

LISTA DE ABREVIATURAS E SIGLAS

AB	Arrecadação bruta
ABRASF	Associação Brasileira das Secretarias de Finanças das Capitais
AL	Arrecadação líquida
AM	Aprendizagem de Máquina
API	Interface de Programação de Aplicação, do Inglês <i>Application Programming Interface</i>
AR	<i>AutoRegressions</i> , traduzido como autorregressivo
ARIMA	Modelo Autoregressivo Integrado de Médias Móveis, do inglês <i>AutoRegressive Integrated Moving Average</i>
ARMA	<i>AutoRegressive Moving Average</i> , traduzido como Médias móveis autorregressivas
CBO	<i>Congressional Budget Office</i> , traduzido por Escritório de Orçamento do Congresso
CF	Constituição Federal
CFEM	Compensação Financeira de Recursos Minerais aos Municípios produtores
CFEM AFETADOS	15% da CFEM aos Municípios não produtores afetados por atividades de Mineração
CIDE	Contribuição de Intervenção no Domínio Econômico
CIFPM	Coeficiente Individual do FPM
CIRPe	Coeficiente relativo ao inverso da renda per capita do estado
CNN	Rede Neural Convolucional, tradução de <i>Convolutional Neural Network</i>
CNM	Confederação Nacional de Municípios
CPe	Coeficiente relativo da população da capital do estado e em relação ao total de população de todas as capitais
CRISP-DM	Processo Padrão da Indústria Cross para Mineração de Dados, tradução do inglês: <i>Cross Industry Standard Process for Data Mining</i>
CSV	"Valores separados por um delimitador", no inglês <i>"Character-separated values"</i>
CTN	Código Tributário Nacional
EUA	Estados Unidos da América
FEP	Fundo Especial do Petróleo
FPM	Fundo de Participação dos Municípios
FSS	<i>Feature Subset Selection</i>

FUNDEB	Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação
GRU	<i>Gated Recurrent Unit</i>
GT-8	Grupo de Trabalho responsável pela Contabilidade Aplicada ao Setor Público
HTTP	<i>Hypertext Transfer Protocol</i> , ou Protocolo de Transferência de Hipertexto, em português
ICMS	Imposto sobre Circulação de Mercadorias e prestação de Serviços de transporte interestadual e intermunicipal e comunicação
IPi	Imposto sobre Produtos Industrializados
IPI-EXP	Imposto sobre Produtos Industrializados das Exportações
IPM	Índice de Participação dos Municípios
IPVA	Imposto sobre a Propriedade de Veículos Automotores
IR	Imposto de Renda
JSON	<i>JavaScript Object Notation</i> , ou traduzindo, Notação de Objeto em <i>Javascript</i>
KNN	<i>K-Nearest Neighbor</i> , ou K vizinhos mais próximos
LOA	Lei Orçamentária Anual
LRF	Lei de Responsabilidade Fiscal, Lei Complementar 101/2000
LSTM	<i>Long Short-Term Memory</i> , traduzida como Memória longa de curto prazo
MA	<i>Moving Average</i> , traduzido como média móvel
MAE	Erro absoluto médio, do inglês <i>Mean Absolut Error</i>
MAPE	Média Percentual Absoluta do Erro
MLP	<i>Multilayer Perceptron</i>
MSE	Erro quadrático médio, no original <i>Mean Squared Error</i>
MTO	Manual Técnico de Orçamento
PCA	Análise de Componentes Principais, em língua materna <i>Principal Component Analysis</i>
R^2	<i>R-square</i> , traduzido por R-quadrado
ReLU	<i>Rectified Linear Unit</i> ou unidade linear retificada
REST	<i>Representational State Transfer</i> , em tradução livre, Transferência de estado representacional
RMSE	<i>Root Mean Squared Error</i> , em português raiz quadrática média
RNN	<i>Recurrent Neural Networks</i> , ou Redes Neurais Recorrentes
SARIMA	<i>Seasonal Auto Regressive Integrated Moving Average</i> , traduzido como Média móvel integrada autorregressiva sazonal

SARIMAX	<i>Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors</i> , em português Média móvel integrada autorregressiva sazonal com fatores exógenos
SOF	Secretaria do Orçamento Federal
STN	Secretaria do Tesouro Nacional
VA	Valor Adicionado
XGBoost	<i>Extreme Gradient Boosting</i>

SUMÁRIO

1	INTRODUÇÃO	20
1.1	PROBLEMA	21
1.2	JUSTIFICATIVA	22
1.3	QUESTÕES DE PESQUISA	24
1.4	OBJETIVOS	24
1.4.1	Objetivo Geral	24
1.4.2	Objetivos Específicos	24
1.5	ESTUDO DO MÉTODO	25
1.5.1	Metodologia de Pesquisa	25
1.5.2	Procedimentos Metodológicos	25
1.6	ESTRUTURA DO DOCUMENTO	26
2	FUNDAMENTAÇÃO TEÓRICA.....	27
2.1	RECEITAS PUBLICAS	27
2.1.1	Transferências Correntes	28
2.1.2	Método de previsão das receitas públicas.....	30
2.1.3	Transferência da união específica e de suas entidades	31
2.1.4	Transferências dos Estados.....	33
2.2	MÉTODOS DE PREVISÃO DE RECEITAS	35
2.2.1	Métodos paramétricos	36
2.2.2	Métodos não paramétricos.....	37
2.2.3	Seleção de Variáveis.....	38
2.3	CRISP-DM	39
3	TRABALHOS CORRELATOS.....	43
3.1	RESUMO QUANTITATIVO.....	43
3.2	ESTUDOS APLICADOS PREVISÃO DE RECEITAS NO SETOR PÚBLICO	44

3.3 DISCUSSÃO DA LITERATURA E A PRÁTICA.....	46
4 RELATO DAS ATIVIDADE DO ESTUDO DE CASO DE SÃO LUIS APLICANDO DO CRISP-DM.....	48
4.1 PRIMEIRA FASE DE ENTENDIMENTO DO NEGÓCIO	48
4.2 SEGUNDA FASE DE ENTENDIMENTO DOS DADOS	51
4.3 TERCEIRA FASE DE PREPARAÇÃO DOS DADOS.....	54
4.4 QUARTA FASE DE MODELAGEM	59
4.5 QUINTA FASE DE AVALIAÇÃO.....	61
4.6 SEXTA FASE DE IMPLANTAÇÃO	63
5 RESULTADOS INICIAIS OBTIDOS COM OS DADOS DE SÃO LUIS	65
5.1 PRIMEIRA ITERAÇÃO DA SELEÇÃO DE VARIÁVEIS.....	65
5.2 PRIMEIRA ITERAÇÃO DA MODELAGEM DOS MODELOS.....	75
5.3 PRIMEIRA ITERAÇÃO DA IMPLANTAÇÃO DA API.....	83
6 CONSIDERAÇÕES FINAIS.....	86
6.1 TRABALHOS FUTUROS.....	88
6.2 LIMITAÇÕES DA PESQUISA	89
Referências	90
APÊNDICE A: Questionário de Dissertação de Mestrado aplicado às Capitais da ABRASF	103
APÊNDICE B: Resumo geral das receitas.....	109
APÊNDICE C: Relatório inicial de coleta de dados da primeira iteração com a Capital São Luís.....	111
APÊNDICE D: Relatório de descrição dos dados da primeira iteração com a Capital São Luís	124
APÊNDICE E: Relatório de exploração de dados com o <i>pandas profiling</i> de São Luís	130
APÊNDICE F: Comprovante do depósito de IPI no Agente Financeiro	136

APÊNDICE G: Relatório de limpeza dos dados de São Luís.....	137
APÊNDICE H: Resumo do ranking de importância das variáveis do conjunto de dados de São Luís por receita.....	140
ANEXOS	142

1 INTRODUÇÃO

Desde os anos 1800, em que estatísticos investigavam a previsão da órbita ao redor do sol (IZBICKI e SANTOS, 2020, p. 3), até a divulgação das aplicações de Aprendizagem de Máquina (AM), como o reconhecimento de caracteres e o filtro de *spam*, a história foi marcada por uma trajetória de avanços na computação que impulsionaram os da Ciência de dados fomentando a conversão de grandes volumes de dados em informações com o intuito de apoiar a tomada de decisões com previsões (PETROPOULOS, APILETTI, *et al.*, 2022, p. 710; PARMEZAN, SOUZA e BATISTA, 2019, p. 303). Assim, a análise de séries temporais com diagnóstico do comportamento do passado para prever o futuro, embasada em muitas métricas de desempenho, ganhou visibilidade (GÉRON, 2021, p. 7). E, sob a ótica preditiva, métodos de regressão de AM foram disseminados com o objetivo de fornecer, em diversos contextos, os que apresentem bons estimadores para determinar a relação entre as variáveis aleatórias, visando estimar uma função de regressão com risco baixo (IZBICKI e SANTOS, 2020, p. 12).

De acordo com Petropoulos e outros autores (2022, p. 710), o campo da previsão registrou um grande crescimento, com destaque para a combinação de diversos métodos, entre os paramétricos e os que não são, desde que seus erros não sejam correlacionados. Os pesquisadores ainda enfatizam que é essencial verificar a natureza dos atributos em previsões multivariadas, principalmente em aplicação com cenários econômicos, requerendo variáveis explicativas com relevância. Todavia, ao discorrerem sobre aplicação no orçamento governamental, no âmbito internacional, os pesquisadores informam que há um direcionamento para desenvolver sistemas de monitoramento de finanças públicas de curtíssimo prazo, com dados anuais. Eles criticam o viés tendencioso das metas políticas e a retenção de informações privilegiada que poderiam melhorar a previsão; concluem que nesta área possuem um grande desafio: a ausência de um sistema capaz de produzir previsões consistentes entre os países (PETROPOULOS, APILETTI, *et al.*, 2022, p. 774).

No Brasil, a Carta Magna instituiu os instrumentos orçamentários com o intuito de padronizar e sistematizar o orçamento público (NAIBERT, 2021, p. 14). Destarte, com fulcro no Art. 165 da CF/88, estas ferramentas estimam as receitas e fixam as despesas (BRASIL, 1988). “[...]. Assim, as receitas são estimadas porque os tributos arrecadados (e outras fontes) podem sofrer variações ano a ano, enquanto as despesas são fixadas para garantir que o governo não gaste mais do que arrecada” (BRASIL, CGU, s.d.).

Com o advento da Lei de Responsabilidade Fiscal (LRF), o órgão responsável para

orientar a estimativa das receitas é a Secretaria do Orçamento Federal (SOF), sendo a metodologia especificada em seu artigo 12 (BRASIL, 2000). Assim, os procedimentos, no estágio de previsão, devem projetar as receitas do período seguinte refletindo o comportamento da arrecadação delas, auxiliada por modelos estatísticos e matemáticos (BRASIL; SOF, 2021a). Todavia, quando se comparam os valores inicialmente orçados com os executados mensalmente¹, observam-se altas distorções nos valores das receitas transferidas, pela Tabela 1, tendo períodos com subestimação entre 13% a 317%, e 39% de uma superestimação.

Tabela 1 - Comparativo da execução financeira e orçamentária inicial das Receitas Transferidas, em São Luís

Valores Brutos Nominiais em R\$ 1.000,00

Receitas	CFEM	CIDE	FPM	FEP	ICMS	ICMS LC 176/20	IPi	IPVA	ITR	TOTAL
Executada	40.204,15	289,85	490.357,55	8.287,54	468.207,49	4.058,64	5.373,17	81.000,43	6,07	1.097.784,89
Orçada	34.680,41	477,98	356.533,33	5.543,77	411.371,16	0,00	3.813,38	53.410,39	1,46	865.831,89
% Exec E/O	15,93%	-39,36%	37,53%	49,49%	13,82%	100,00%	40,90%	51,66%	317,10%	26,79%

Fonte: Elaborada pela Autora com base no Decreto de Execução Orçamentária (SÃO LUÍS-MA, 2021a, p. 11-14) e no Balancete da Receita (SÃO LUÍS-MA, 2021b).

Nota: Receitas com siglas populares do Portal da Transparência, valores acumulados de janeiro a agosto/2021.

Tal ocorrência, corrobora para o entendimento, no exemplo em comento, que se torna imprescindível revisar as metodologias utilizadas. Haja vista que a norma jurídica direcionada a todos os Entes Federados² exige a elaboração de um relatório de avaliação bimestral contendo, dentre outras informações, a revisão da estimativa das receitas e dos seus parâmetros, uma vez que as divergências podem ocasionar o contingenciamento das despesas (BRASIL, 2021c, p. 46-47; BRASIL, 2000).

Neste contexto, o presente estudo visa automatizar a comparação da performance de previsão das receitas, mensalmente, adotando métodos paramétricos e não paramétricos, seguindo o Processo Padrão da Indústria CRoss para Mineração de Dados (CRISP-DM, do inglês: *Cross Industry Standard Process for Data Mining*) (WIRTH e HIPPE, 2000). Em especial, a comparação dos métodos de seleção de variáveis que sejam mais relevantes para a estimativa das peças orçamentárias, em atendimento às prerrogativas legais dos efeitos da variação do índice de preços e do crescimento econômico, elegendo o modelo mais adequado para cada receita, baseando-se nas menores métricas de erros, usando os dados de 2010 a 2021 dos portais da transparência de São Luís (BRASIL, 2000).

1.1 PROBLEMA

Considerando que a SOF determina as diretrizes metodológicas de previsão das receitas

¹ Usando como exemplo os dados já migrados da Capital do Maranhão, São Luís.

² no art. 9º e §4º do art. 55 da LRF (BRASIL, 2000), no §3 da CF (BRASIL, 1988st),

(BRASIL; SOF, 2021b), faz mister a análise dos dados e seus relacionamentos com as variáveis que podem influenciar o comportamento das receitas, em atendimento à legislação vigente (BRASIL, 2000). Destarte, percebendo-se o mérito, a magnitude e a relevância da previsão das receitas orçamentárias para a esfera municipal, este estudo trouxe à baila o seguinte problema: ***Qual o método e as variáveis aplicados a previsão das receitas, no âmbito municipal, atendendo aos ditames da legislação, gerariam menos distorções percentuais entre orçado e executado?***

1.2 JUSTIFICATIVA

O referido problema envolve um tema bastante complexo, com várias legislações tanto para a formalização das informações quanto para a arrecadação e repasse de cada receita, bem como a disponibilização dos dados que se encontram em diversas fontes e formatos (BRASIL, 1988; 2000; 2021a; 2021b; 2021c; TCU, 2021d; CGU, s.d.). Nesta perspectiva, como se trata de uma tarefa de extração, tratamento e predição cujo alvo é encontrar um modelo em função dos dados disponíveis para projetar o valor de cada receita, optou-se por usar as técnicas de AM, as quais lidam, amplamente, com tal problemática (MAKRIDAKIS, SPILIOTIS e ASSIMAKOPOULOS, 2022, p. 11; SEKAR, 2022, p. 43,61; GÉRON, 2021, p. 6-7; ACKERMANN e SELLITTO, 2022, p. 12; SAMPAIO, BERNARDINI, *et al.*, 2019, p. 82; PARMEZAN, SOUZA e BATISTA, 2019, p. 303;332;335).

Norteando-se pelo teorema “*no free-lunch*” (traduzindo, não existe almoço grátis) de David Wolpert (1996 *apud* SAMPAIO, BERNARDINI, *et al.*, 2019, p. 83,88-89; *apud* GÉRON, 2021, p. 27) que defende a inexistência de um único algoritmo que seja o melhor para todas as situações, tornando-se imperativa a realização de experimentos para verificar aquele que apresentar melhor resultado de predição por intermédio de um processo de avaliação e de conhecimento do domínio de aplicação. Aproveitando o ensejo, entende-se que a metodologia adequada seja “*(...) um processo que parte de um objetivo de negócios, relativamente claro, e dados que já foram coletados e estão disponíveis para processamento computacional posterior*”³ (MARTÍNEZ-PLUMED, CONTRERAS-OCHANDO, *et al.*, 2021, p. 3050), direcionando-se para a escolha do CRISP-DM, considerado, mesmo depois de 20 anos, confiável, simplificado, de uso popular na prática, na indústria e na pesquisa que independe do setor de aplicação (SCHRÖER, KRUSE e GÓMEZ, 2021, p. 527,529; MARTÍNEZ-PLUMED,

³ Texto original: “*(...) a process starting from a relatively clear business goal and data that have already been collected and are available for further computational processing.*” (MARTÍNEZ-PLUMED, CONTRERAS-OCHANDO, *et al.*, 2021, p. 3050).

CONTRERAS-OCHANDO, *et al.*, 2021, p. 3051; HARRISON, 2020, p. 22).

Assim, motivando-se pela sugestão dos trabalhos futuros de Carmo, Boldt e Komati, (2019, p. 128) e por publicações na literatura cuja conclusão afirma que há um “(...) *amplo ‘cardápio’ de desenhos alternativos de pesquisa (...) com potenciais contribuições práticas que beneficiem o processo de gestão e tomada de decisão nas organizações*” (AGUIAR, 2018, p. 11). Concordando com Petropoulo e outros autores (2022, p. 775), entende-se que há poucas publicações voltadas para o estudo de métodos de AM em previsão de receitas públicas, e menos ainda com a proposta de seleção de variáveis para as Receitas Transferidas municipais⁴.

Outrossim, por se tratar de um Mestrado Profissional, além de colaborar com o preenchimento desta lacuna, buscou-se fornecer um guia simplificado para técnicos e profissionais interessados no tema, na construção de *datasets* e em modelos preditivos para suas aplicações. Considerando as nuances específicas desta área, mediante disponibilização dos artefatos produzidos nesta pesquisa, tornando-se um diferencial, possibilitando a replicação, ampliação ou melhoria da previsão de receitas no setor público.

Ademais, buscou-se fazer uma aplicação das técnicas de AM a um problema do mundo real, uma alternativa para apoiar a implantação das regras da SOF para a LRF, com o intuito de contribuir com a disseminação das técnicas de AM no setor público, especialmente pelo fato de ser um tema relevante neste momento em que se vivencia uma volatilidade econômica (CARMO, BOLDT e KOMATI, 2019). Haja vista que alguns autores (NAIBERT, 2021; LIMA, LUI, *et al.*, 2020; PAMPLONA, FIIRST, *et al.*, 2019), em seus debates sobre orçamento, enfatizam que é notória a carência de mão de obra técnico-administrativa capacitada, na esfera municipal, para elaboração dos instrumentos orçamentários. Corroborando com os argumentos destes autores, das 8 capitais que aceitaram participar deste estudo, em resposta ao questionário⁵ disponibilizado ao Grupo de Trabalho responsável pela Contabilidade Aplicada ao Setor Público (GT-8) da Associação Brasileira das Secretarias de Finanças das Capitais (ABRASF), apurou-se que apenas duas ofertam cursos de capacitação relacionados à previsão das receitas públicas (Estatística) e não há registro de convênio com Universidades para tal finalidade. Ratificando a lacuna levantada pelos autores (capacitação), nestas Capitais, reforça-se a necessidade de propostas como a do presente estudo, pois, no

⁴Discutido no tópico 3 TRABALHOS CORRELATOS.

⁵Elaborado no *GoogleForms*, cópia no APÊNDICE A e disponível em: https://docs.google.com/forms/d/e/1FAIpQLSewulBw3Os8-58q0jDKy9VxSDBKKHRpk1_Z70N0ivWVy2-33w/viewanalytics.

referido questionário foi informado que apenas uma possui um sistema de projeção de receitas.

Por fim, enfatiza-se que a previsão de receitas públicas é uma questão de suma importância para o ente federado cujos reflexos atingem toda a população, “(...)no quesito do controle social e interesse público, [...] além de contribuir para corrente teórica na comparabilidade das modelagens de previsão” (PAMPLONA, FIIRST, *et al.*, 2019, p. 93), uma vez que as políticas públicas estão sujeitas ao montante estimado para atender as necessidades de financiamento dos governos (BRASIL; SOF, 2021a; PAMPLONA, FIIRST, *et al.*, 2019, p. 94).

1.3 QUESTÕES DE PESQUISA

Norteando-se pelo problema, duas questões foram levantadas relacionadas a seleção de variáveis e o agrupamento delas por competência:

- a) Q01: Os efeitos preço e quantidade da LRF podem ser identificados pelos métodos de AM, com possibilidade de respaldar as justificativas das peças orçamentárias, no que tange a utilização de metodologias?
- b) Q02: A predição das receitas deve considerar as variáveis pelo regime de competência trazendo-as para o períodos do fato gerador de cada receita transferida?

1.4 OBJETIVOS

Para responder as perguntas e averiguar se há uma solução ou mitigação do problema comentado na justificativa, o objetivo geral e específicos são descritos nos subitens a seguir.

1.4.1 Objetivo Geral

Realizar um estudo em 8 Capitais para a escolha das variáveis que atendam as prerrogativas dos efeitos preço e quantidade da LRF, no que tange a possíveis descobertas de qualquer outro fator relevante de impacto para estimativa das peças orçamentárias, nos horizontes dos ditames legais, adotando métodos estatísticos clássicos e técnicas de AM.

1.4.2 Objetivos Específicos

- a) Identificar as variáveis que impactam na execução de cada receita usando técnicas de seleção de variáveis baseada na importância e modelos combinados;
- b) Elaborar um comparativo dos modelos com previsão de cada receita por peça orçamentária utilizando modelos estatísticos clássicos e de AM;

- c) Disponibilizar os códigos para teste e uso em outras capitais com o intuito de disseminar IA e suas técnicas de AM ao setor público.

1.5 ESTUDO DO MÉTODO

Esta seção foi dividida em duas subseções com o intuito de caracterizar a metodologia que será utilizada e os procedimentos metodológicos para o planejamento de como alcançar os objetivos.

1.5.1 Metodologia de Pesquisa

No contexto científico, tanto Gil (2017) quanto Wazlawick (2021) concordam que a pesquisa pode ser classificada em função de diferentes critérios de finalidade, objetivos e abordagem do problema. Sob o ponto de vista da finalidade, esta pesquisa será do tipo aplicada, uma vez que estará voltada a averiguar qual o método e as variáveis, aplicadas à problemática das distorções percentuais entre os valores orçados e executados das receitas pública, visando a resolução ou mitigação de “[...] *problemas identificados no âmbito das sociedades em que os pesquisadores vivem.*” (GIL, 2017, p. 32).

Em se tratando do objetivo mais geral, adotar-se-á uma mescla na abordagem entre descritiva e exploratória. Como descritiva, o foco paira sobre o levantamento dos dados (WAZLAWICK, 2021, p. 43), assim como em descobrir a existência de associações entre variáveis, estando bem próxima de uma pesquisa explicativa, no que tange a identificação dos efeitos preço e quantidade da LRF associados a seleção de variáveis (GIL, 2017, p. 32). Quanto ao enfoque exploratório, relacionou-se à apreciação do conjunto das receitas, investigando possíveis anomalias conhecidas ou não (WAZLAWICK, 2021, p. 42).

Ao se tratar da abordagem do problema a pesquisa teve um cunho de delineamento incorporado, tanto quantitativo quanto qualitativo (GIL, 2017, p. 113). Mas, a predominância foi quantitativa ao se utilizar procedimentos estatísticos e técnicas de AM para responder à questão do estudo de caso. Contudo, ao se utilizarem os questionários, com o intuito de subsidiar um diagnóstico inicial de cada capital, incorporam-se elementos qualitativos.

1.5.2 Procedimentos Metodológicos

Quanto aos procedimentos técnicos será um estudo de caso, seguiram-se as etapas da metodologia CRISP-DM apoiando-se na revisão sistemática de literatura inglesa de Schröer,

Kruse e Gómez (2021) os quais afirmam que é o padrão em ciência de dados, sendo popularmente utilizado nas pesquisas e na prática (SEKAR, 2022).

Em termos do estudo, a população abrangida compreende uma amostra de São Luís, capital do Maranhão. Todos os dados utilizados serão obtidos dos portais da transparência das três esferas governamentais e publicações nos diários oficiais, o que em alguns casos pode-se utilizar técnicas de *web scraping* e *crawling*, mas em outros far-se-á mister solicitá-los por e-mail, salvos em planilhas eletrônicas, arquivos do tipo "Valores separados por um delimitador", no inglês "*Character-separated values*" (CSV) 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 pormenorizados sobre a execução orçamentária e financeiros, disponibilizados em meios eletrônicos de acesso público (BRASIL, 2009).

As receitas escolhidas foram as transferências correntes especificadas na subseção 2.1.1. O intervalo escolhido foi definido pela disponibilidade dos dados, abrangendo de 2010 a 2021, com a padronização de periodicidade mensal, entre variáveis, receitas efetivamente executadas e sua previsão orçamentária inicial do ano de 2021. Os resultados serão apresentados às 8 Capitais que farão sua avaliação por intermédio de um outro questionário.

1.6 ESTRUTURA DO DOCUMENTO

O presente documento está estruturado em 6 partes, incluindo esta introdução. Sendo que, no capítulo 2, foi apresentada a fundamentação teórica descrevendo os elementos básicos relacionados com as Receitas Públicas e os seus métodos de previsão, discorrendo sobre os paramétricos e os de AM, bem como a seleção de variáveis, com o intuito de embasar os principais assuntos envolvidos na solução do problema. No capítulo 3, relataram-se os trabalhos correlatos para fins de comparação e complementação da abordagem do tema. Já no capítulo 4, a finalidade foi registrar como as fases do CRISP-DM ocorreram, desde a compreensão do negócio, preparação e implementação dos dados. O capítulo 5 foram agrupados os resultados obtidos na primeira iteração, decorrente da avaliação dos testes na escolha da seleção de variáveis, da implementação da API e dos modelos de AM. E por fim, no capítulo 6 as considerações finais e os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo teve o intuito de organizar o embasamento teórico para a proposta da qualificação norteando-se pelos principais assuntos que envolvem os objetivos fixados. Assim, foram abordadas as receitas públicas, seu conceito e classificação, abrangendo as transferências correntes, destacando as mais relevantes, bem como seus métodos de previsão e horizonte temporal. Com o mesmo enquadramento, discorre-se sobre os métodos paramétricos e não paramétricos, as técnicas de seleção de variáveis, finalizando com a metodologia CRISP-DM.

2.1 RECEITAS PÚBLICAS

De acordo com a SOF, as receitas públicas são recursos financeiros que ingressam nos cofres públicos e, quando representam a disponibilidade de recursos financeiros para o erário, são divididas em receitas orçamentárias e extraorçamentárias (BRASIL, SOF, 2022). Em se tratando de ingresso das receitas públicas nos cofres do Ente, a receita orçamentária é uma fonte de recursos utilizada pelo Estado em seus planos e ações, cuja finalidade principal é atender às necessidades públicas e sociais, à vista disso, o art. 57, da Lei nº 4.320, determina a obrigatoriedade de classificar as receitas arrecadadas como orçamentárias, mesmo que não estejam previstas na Lei Orçamentária Anual (LOA) (BRASIL, 1964). Para a prática cotidiana, Albuquerque, Medeiros e Silva (2013, p. 177) lecionam que tais ingressos devem ser classificado sob a seguinte ótica.

- 1) Categoria econômica/natureza, identificando origens;
- 2) Fonte/destinação de recurso, especificado a vinculação; e
- 3) Indicador de resultado primário, verificando o impacto fiscal (ALBUQUERQUE, MEDEIROS e SILVA, 2013, p. 177).

No que tange à identificação da origem das receitas, na mesma lei, nos termos do art. 8º, estabelece-se a utilização de uma codificação decimal (BRASIL, 1964). Deste modo, a codificação por natureza torna-se a unidade básica de informação para alocação orçamentária, a qual busca refletir de forma transparente o fato gerador que originou o ingresso dos recursos no caixa, alinhando-se aos eventos econômicos (BRASIL, MINISTÉRIO DA FAZENDA, 2007). Por esse motivo, foi estabelecida uma codificação estruturada em níveis que é normatizada por meio de portarias interministeriais da Secretaria do Tesouro Nacional (STN) e da SOF (BRASIL, SOF, 2022). Esta codificação atualmente é composta por oito dígitos cujas posições estão especificadas no Quadro 1.

Quadro 1 - Estrutura da codificação da receita por dígito e significado

DÍGITO:	1º	2º	3º	4º a 7º	8º
SIGNIFICADO:	Categoria Econômica	Origem	Espécie	Desdobramentos para identificação de peculiaridades da receita (rubrica)	Tipo

Fonte: Adaptado de BRASIL, SOF (2022).

No mesmo dispositivo, em seu art. 11, ao estabelecer o processo de classificação por categoria econômica, dividiu-se em corrente e de capital e instituiu-se o detalhamento da origem, contemplado no Quadro 2 com a devida codificação (BRASIL, 1964).

Quadro 2 - Classificação econômica da receita orçamentária

Categoria Econômica (1º Dígito)	Origem (2º Dígito)
1. Receitas Correntes; 7. Receitas Correntes Intraorçamentárias.	1. Impostos, Taxas e Contribuições de Melhoria; 2. Contribuições; 3. Receita Patrimonial; 4. Receita Agropecuária; 5. Receita Industrial; 6. Receita de Serviços; 7. Transferências Correntes; 9. Outras Receitas Correntes.
2. Receitas de Capital; 8. Receitas de Capital Intraorçamentárias.	1. Operações de Crédito; 2. Alienação de Bens; 3. Amortização de Empréstimos; 4. Transferências de Capital; 9. Outras Receitas de Capital.

Fonte: Adaptado de Brasil, SOF (2022).

Quanto à classificação por fonte e destinação de recurso visam ao atendimento da especificação da vinculação do ingresso dos recursos por imposição do parágrafo único⁶ do art. 8º, combinado com o inciso I⁷ do art. 50 da LRF (BRASIL, 2000). Deste modo, pode-se interpretar que os balanços, ou mesmo os relatórios disponibilizados nos sites da Transparência dos Governos apresentam classificações analíticas e sintéticas, as quais precisam ser linearizadas para evitar duplicações. Outrossim, compreende-se que para a previsão das receitas o tratamento deve ser individualizado.

2.1.1 Transferências Correntes

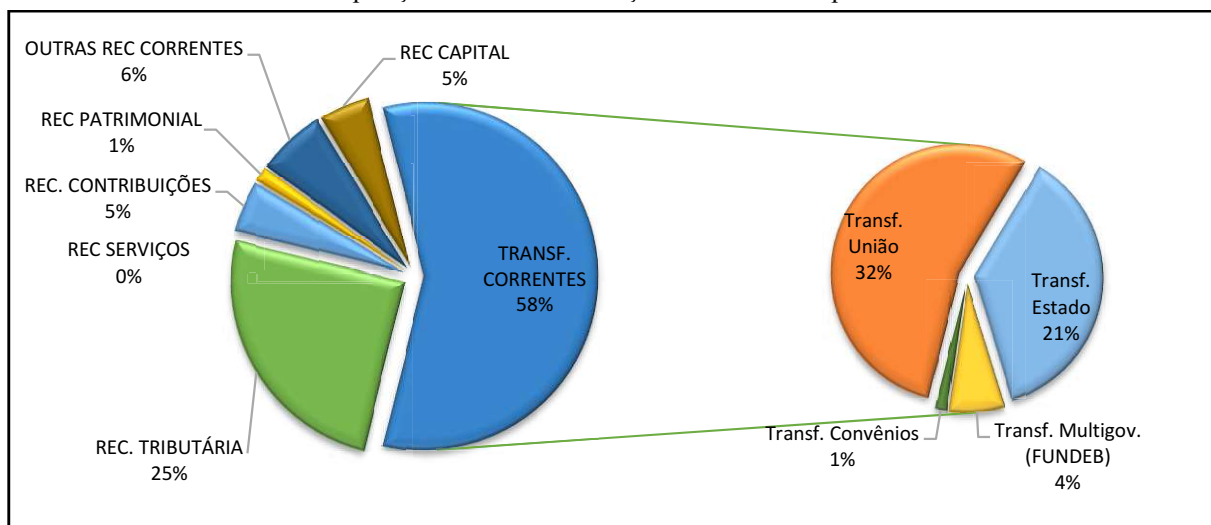
Em cumprimento à determinação constitucional, ordenamento legal ou voluntário instituem-se as transferências de recursos intergovernamentais, caracterizadas por transações ocorridas entre esferas distintas de governo (BRASIL, SOF, 2022; MINISTÉRIO DA

⁶ Parágrafo único. Os recursos legalmente vinculados a finalidade específica serão utilizados exclusivamente para atender ao objeto de sua vinculação, ainda que em exercício diverso daquele em que ocorrer o ingresso. (BRASIL, 2000).

⁷ I - a disponibilidade de caixa constará de registro próprio, de modo que os recursos vinculados a órgão, fundo ou despesa obrigatória fiquem identificados e escriturados de forma individualizada; (BRASIL, 2000).

FAZENDA, mar. 2016, p. 4).

Gráfico 1 - Composição das receitas no orçamento do município de São Luís/MA



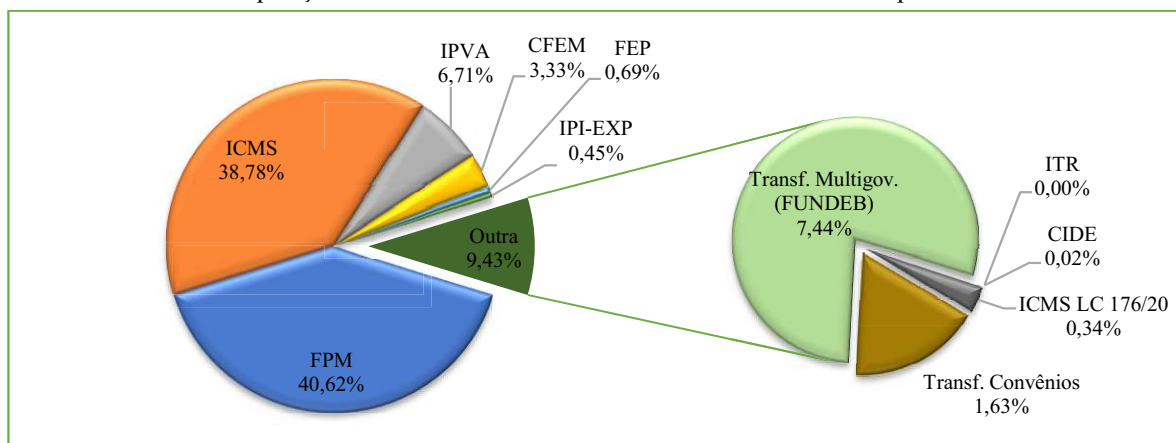
Fonte: Elaborado pela Autora (2023) com base no Balancete da Receita (SÃO LUÍS-MA, 2021b)⁸.

Ainda sob a ótica da STN, o tema das transferências fiscais é considerado de difícil compreensão tanto para os cidadãos quanto para os que estão envolvidos no contexto (BRASIL, SOF, 2022). Cada uma das receitas possui particularidades distintas: legislação, cálculos de coeficiente para distribuição e periodicidade, dentre outras nuances. Para ilustrar a importância e relevância das Receitas Transferidas, como um exemplo, produziu-se o Gráfico 1, em que a hegemonia das Transferências Correntes impera com 58%, dentre elas identificam-se que 53% contemplam as Transferências da União (32%) e do Estado (21%)⁸.

Quando se detalham os 58% das transferências da União e do Estado, ao nível do Manual Técnico de Orçamento (MTO) da SOF (BRASIL, SOF, 2022), chega-se a conceber o Gráfico 2, no qual é perceptível a importância de algumas transferências em detrimento de outras, motivo pelo qual a elas tendem a ter mais destaque do que as demais.

⁸ 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.

Gráfico 2 – Decomposição das Receitas de Transferências Correntes do município de São Luís/MA



Fonte: Elaborado pela Autora (2023) com base no Balancete da Receita ((SÃO LUÍS-MA, 2021b))⁸.

Destarte, fica evidente a importância desta pesquisa de previsão das Receitas Transferidas ao Município, dada a sua dependência financeira 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 aos Municípios produtores (CFEM) e sua variante que distribui 15% da CFEM aos Municípios não produtores afetados por atividades de Mineração (CFEM AFETADOS), Fundo Especial de Petróleo (FEP), Imposto sobre Circulação de Mercadorias e prestação de Serviços de transporte interestadual e intermunicipal e comunicação (ICMS), Imposto sobre a Propriedade de Veículos Automotores (IPVA) e Imposto sobre Produtos Industrializados das Exportações (IPI-EXP).

2.1.2 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 (BRASIL; SOF, 2021a; BRASIL, MINISTÉRIO DA FAZENDA, 2007; PAMPLONA, FIIRST, *et al.*, 2019, p. 94; ALBUQUERQUE, MEDEIROS e SILVA, 2013, p. 199) cujos horizontes temporais, nos termos do Art. 12 da LRF, devem alcançar 2 anos subsequentes. Todavia, quando se combina com o Art. 13, o qual constitui a obrigatoriedade de seguir metas bimestrais com as receitas desdobradas em sua classificação ((BRASIL, 2000; ALBUQUERQUE, MEDEIROS e SILVA, 2013, p. 201), 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) (BRASIL, 2000; ALBUQUERQUE, MEDEIROS e SILVA, 2013, p. 313).

Compreendida a granularidade do horizonte temporal de previsão, faz-se mister comentar a equação [1] projetada pela SOF para estimativas de receitas dos Órgãos da

Administração Públicas (BRASIL, MINISTÉRIO DA FAZENDA, 2007; SILVA, SHIAKU, *et al.*, 2018, p. 94; ALMEIDA e FERREIRA, 2018, p. 28; PAMPLONA, FIIRST, *et al.*, 2019, p. 95).

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

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.

Corroborando para o entendimento das legendas da equação [1], Albuquerque, Medeiros e Silva (2013, p. 201) instruem que, excluindo o efeito legislação (fora do escopo deste estudo)⁹ e a variação dos erros, os demais estão intimamente ligados aos indicadores econômicos que influenciam na arrecadação de cada tributo. Para uma visão mais holista do assunto uma consulta a Duarte e outros autores (2021, p. 173-199) ou Black (2015). Ainda sobre a equação [1], alguns auto/res (SILVA, SHIAKU, *et al.*, 2018, p. 442; PAMPLONA, FIIRST, *et al.*, 2019, p. 95) alegam que a SOF não detalha, muito menos elucida como devem ser aplicadas as variações.

Ainda sob a ótica da STN, o tema das transferências fiscais é considerado de difícil compreensão tanto para os cidadãos quanto para os que estão envolvidos no contexto. Em função disto, foi criado, em 2014, o Grupo de Trabalho Transferências da União “[...] com o objetivo de desenvolver um ambiente centralizado e integrado que consolidasse as informações sobre as transferências de recursos financeiros da União [...] a fim de promover a transparência das ações executadas por meio dessas transferências [...]”, culminando em uma melhor sistematização da classificação e organização da legislação sobre o assunto (BRASIL, MINISTÉRIO DA FAZENDA, mar. 2016, p. 7).

2.1.3 Transferência da união específica e de suas entidades

A partilha da arrecadação de receitas federais é prevista na Constituição nos arts. 157 ao 162, com o propósito de “[...] amenizar as desigualdades regionais, na busca incessante de promover o equilíbrio socioeconômico entre Estados e Municípios.” (BRASIL, s.d.). Para tal atribuição,

⁹ Tendo como exemplos recentes (PROQUE, BETARELLI JUNIOR, *et al.*, 2021).

Cabe ao Tesouro Nacional, em cumprimento aos dispositivos constitucionais, efetuar as transferências dos recursos aos entes federados, nos prazos legalmente estabelecidos. No caso do Fundeb, compete ao Fundo Nacional da Educação (FNDE), entidade do Ministério da Educação, realizar os repasses, na condição de agente daquele fundo. Em relação aos Fundos de Participação (FPE e FPM), compete ainda ao Tesouro Nacional divulgar aos estados e municípios as previsões de receita e os valores liberados com as respectivas bases de cálculo. (BRASIL, s.d.).

Assim como os Governos Estaduais, a STN também veicula, em seu site, os comunicados descendias para os repasses das Cotas Parte do FPM e do IPI-EXP. A transparência da consumação do depósito é informada pelo Agente Financeiro, o qual, porquanto, disponibiliza, em seu, um sistema de consultas dos Documentos de Arrecadação de Receitas Federais, reunidos no ANEXO D.

2.1.3.1 FPM

O FPM é uma transferência constitucional estabelecida no Art. 159, inciso I, alíneas "b", "d" e "e", da CF (BRASIL, 1988st). Por ser um repasse da União para os Estados e o Distrito Federal, é composto pela arrecadação do Imposto de Renda (IR) e do Imposto sobre Produtos Industrializados (IPI), conforme distribuição da Tabela 2, cuja periodicidade do repasse do valor principal, é transferida em decêndios subsequentes ao seu período de arrecadação (CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016).

Tabela 2 - Percentual de repasses dos recursos do FPM aos Municípios

Descrição Repasse	Participação Percentual		Periodicidade do Repasse
	IR	IPI	
FPM do art. 159, I, "b"	22,50%	22,50%	Lei Complementar 62/89, disciplina os repasses decendiais de todos os meses.
FPM EC-55/07 do art. 159, I, "d"	1,00%	1,00%	Emenda Constitucional 55, de 2007, estabeleceu um repasse no 1º decêndio de dezembro de cada ano.
FPM EC-84/14 do art. 159, I, "e"	1,00%	1,00%	Emenda Constitucional 84, de 2014, estabeleceu um repasse no 1º decêndio de julho de cada ano.
FPM EC-112/2021 do art. 159, I, "f"	1,00%	1,00%	Emenda Constitucional 112, de 2021, estabeleceu um repasse no 1º decêndio de setembro de cada ano
TOTAL	24,50%	24,50%	

Fonte: Adaptado de CNN (2016, p. 27).

O cálculo do montante da transferência do FPM, de acordo com o percentual da Tabela 2, é feito somando a arrecadação bruta (AB) do período de IR com IPI, subtraindo-se as deduções (DE) de restituições, incentivos fiscais e chega-se a arrecadação líquida (AL), da qual, para as Capitais separa-se 10% para partição (CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016, p. 27; BRASIL, 1989a).

$$Ve = \frac{0,1 * AL * CPe * CIRPe}{\sum(CPe * CIRPe)} \quad [2]$$

Norteando-se pelas regras do Código Tributário Nacional (CTN) e do Decreto-Lei n.º 1.881/81, utiliza-se para o cálculo do Coeficiente Individual do FPM (CIFPM) o índice inverso da Renda per capita do Estado para as Capitais (CIRPe) e o relativo ao total da população das Capitais (CPe), de 3 anos anteriores e sua população com 2 anos, conforme a equação [2] (BRASIL, 1989a; 1981; CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016, p. 27).

2.1.3.2 CFEM

A CFEM foi instituída pela Lei Federal 7.990/1989 em repasses mensais, consiste nas compensações financeiras para os Estados e Municípios pelo resultado da exploração de bens finitos, como o petróleo ou gás natural, bem como da exploração de recursos hídricos para fins de geração de energia elétrica, de recursos (BRASIL, 1989b). Com o advento das recentes alterações na legislação, este recurso foi dividido em CFEM do Produtor, cujos Municípios recebem da AL recebem 60%, e CFEM AFETADOS, em que 15% é distribuída aos demais que são afetados pelas atividades de mineração que envolvam infraestruturas de transporte ferroviário e dutoviário, operações portuárias e onde se localizem pilas estéril, as barragens de rejeitos, bem como as instalações de beneficiamento de minerais (BRASIL, 1989b; 2017; 2018; CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016, p. 47). Para maiores detalhes das distribuições, deve-se consultar o site da Agência Nacional de Mineração que disponibiliza os critérios por ano.

2.1.3.3 FEP

Segundo ensina a CNM sobre a Lei 7.453/1985, este Fundo distribui parte dos *royalties* de mar a todos os Estados e Municípios. Sendo destinados mensalmente aos Municípios 80% e 20% aos Estados, tendo como critério de partilha o CIFPM (BRASIL, 2012; 1985; CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016, p. 54). Este repasse foi e continua sendo palco de constantes debates e contestações por parte dos Municípios Produtores de Petróleo, no que se refere a distribuição deste recurso, com propostas até de extinção.

2.1.4 Transferências dos Estados

As transferências de competência do Governo do Estado, compreendidas no Quadro 2, sob o grupo de códigos 17200000000, contemplam duas que são lançadas, arrecadas e

recolhidas por ele: o ICMS e o IPVA. Já as outras duas participações, a Contribuição de Intervenção no Domínio Econômico (CIDE) e o IPI-EXP, são constitucionais e obrigatórias, de natureza compensatória, advindas da União; mas, por determinação legal, são derivadas da partilha de 25% da cota recebida pelos Estados (BRASIL, 1989a). Neste enquadramento, por imposição legal, ICMS e IPI-EXP são distribuídos sob o cálculo do Índice de Participação dos Municípios (IPM), formalizado e publicado em Portarias pelo Governo do Estado (Id.). Além disso, a Lei Complementar nº 63, de 11 de janeiro de 1990, exige que:

Art. 8º Mensalmente, os Estados publicarão no seu órgão oficial a arrecadação total dos impostos a que se referem os arts. 2º e 3º desta Lei Complementar e o valor total dos recursos de que trata o art. 7º, arrecadados ou transferidos no mês anterior, discriminadas as parcelas entregues a cada Município. (BRASIL, 1990).

Em atendimento ao impositivo legal, a Autoridade Administrativa Pública Estadual, periodicamente, na página de transparência, um *link* para o seu sistema no qual divulga os repasses semanais do ICMS e do IPVA, fixado, em lei, seu depósito foi definido para as terça-feira de cada semana¹⁰, os quais estão exemplificados no ANEXO B (para o Estado do Maranhão). Ao encerramento de cada mês, disponibilizam, ainda, no mesmo site, as Portarias com o demonstrativo das transferências de receitas estaduais aos municípios e um consolidado dos aludidos valores comunicados, apensados no ANEXO C. Vale ressaltar que cada Estado tem sua legislação de cálculo do IPM e publicação.

2.1.4.1 ICMS

A CF, em seu Art. 158, inciso “IV”, determina que 25% da arrecadação do ICMS dos Estados pertence aos seus respectivos Municípios (BRASIL, 1988; CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM), 2016, p. 63).

Até o exercício financeiro de 2022, o cálculo do IPM era distribuído com 75% da proporção do Valor Adicionado (VA)¹¹, entre todos os entes municipais e 25% de acordo com Lei Estadual, no caso do Maranhão, 5% da Área do município, 5% da População e 15% de Divisão Igualitária (MARANHÃO, GOVERNO DO ESTADO DO, 1992). Já para os exercícios posteriores, por imperativo constitucional, os percentuais passaram para 65% do VA e 35% como dispuser a Lei Estadual, sendo obrigatório que 10% advenham de indicadores da

¹⁰ Art. 5º Até o segundo dia útil de cada semana, o estabelecimento oficial de crédito entregará, a cada Município, mediante crédito em conta individual ou pagamento em dinheiro, à conveniência do beneficiário, a parcela que a este pertencer, do valor dos depósitos ou remessas feitos, na semana imediatamente anterior, na conta a que se refere o artigo anterior. (BRASIL, 1990).

¹¹ Valor Adicionado consiste na apuração do valor da riqueza gerada e/ou sua contribuição por empresas de capital aberto, tais como empregados, financiadores, acionistas, governo e outros, bem como a parcela da riqueza não distribuída (BRASIL, 2007).

educação (BRASIL, 1988; 2020).

2.1.4.2 IPVA

Destinam-se 50% da AL do IPVA de domicílio da placa do veículo automotor, por determinação constitucional, conforme art. 158, inciso “III” (BRASIL, 1988st). Assim, este é o cálculo mais simplificado dentre as receitas transferidas, e no Estado do Maranhão, sua periodicidade de repasses ocorrem junto com o ICMS, ou seja, semanalmente, mas cada Ente pode ter suas datas de depósitos.

2.1.4.3 IPI-EXP

O IPI, de competência da União, foi constituído com esse nome pela primeira vez em nossa estrutura jurídica pelo art. 11 da EC 18/1965 (BRASIL, 1965). A Constituição de 1988 (art. 159 § II e §§ 2º e 3º) determinou a transferência da parte da arrecadação do IPI para os estados, o Distrito Federal e os Municípios, fixado em 10% da arrecadação “proporcionalmente ao valor das respectivas exportações de produtos industrializados” (BRASIL, 1988). Do montante recebido pelos Estados, 25% devem ser distribuídos aos seus respectivos municípios pelos critérios do IPM.

Com a aprovação da LC 61/1989, que regulamentou as transferências e estabeleceu normas para a participação dos Estados e do Distrito Federal na arrecadação do IPI, relativo às exportações, foi possível iniciar tais transferências intergovernamentais, que ocorreram a partir de fevereiro de 1990 (BRASIL, 1989).

2.2 MÉTODOS DE PREVISÃO DE RECEITAS

De acordo com Petropoulos e outros autores (2022, p. 710), “a teoria da previsão é baseada na premissa de que o conhecimento atual e passado pode ser usado para fazer previsões sobre o futuro”¹². Assim, os escritores consideram que tal teoria se desenvolve por meio de métodos e modelos, os quais são classificados em qualitativos e quantitativos, destes apenas os quantitativos estão no escopo desta proposta por produzirem resultados objetivos com procedimentos padronizados (ACKERMANN e SELLITTO, 2022, p. 86). Tais métodos quantitativos são agrupados em duas abordagens de distribuição de dados: paramétricos e não paramétricos descritos nas subseções que seguem (PARMEZAN, SOUZA e BATISTA, 2019, p. 308).

¹² Tradução livre, original: “The theory of forecasting is based on the premise that current and past knowledge can be used to make predictions about the future” (PETROPOULOS, APILETTI, et al., 2022, p. 710).

2.2.1 Métodos paramétricos

No entendimento de Parmezan, Souza e Batista (2019, p. 308), para utilização dos métodos paramétricos ou estatísticos, exige um conhecimento prévio de distribuição de dados, pois sua dependência de ajustes de parâmetros é crucial para otimizar os resultados de suas previsões. Já para Ackermann e Sellitto (2022, p. 89) denominam-os como “Séries Temporais” em sua modelagem univariada, sem variáveis explicativas externas. Vários autores (ACKERMANN e SELLITTO, 2022, p. 96; BROWNLEE, 2020, p. 12; PARMEZAN, SOUZA e BATISTA, 2019, p. 305-306) comentam a necessidade de pressupostos e identificação de padrões de comportamento ou efeitos associados à decomposição delas em componentes de tendência, sazonalidade, ciclicidade e aleatoriedade.

Dentre os modelos clássicos mais conhecidos na literatura com abordagem univariada, incia-se com o *Naives* que é considerado como um modelo de passeio aleatório, uma vez que valor futuro é a repetição do seu antecessor (MAKRIDAKIS, SPILIOTIS e ASSIMAKOPOULOS, 2020, p. 57).

O modelo *AutoRegressions*, traduzido como autorregressivo (AR) é considera como o bloco de construção mais básico, sendo adicionado pelas demais técnicas, pois faz a previsão por correlações entre os valores passados (KORSTANJE, 2021, p. 45). Outro bloco básico de construção é o modelo de *Moving Average*, traduzido como média móvel (MA) o qual explora a autocorrelação dos erros (PARMEZAN, SOUZA e BATISTA, 2019, p. 308; KORSTANJE, 2021, p. 86). A primeira combinação dos blocos básicos AR e MA originaram o *AutoRegressive Moving Average*, traduzido como Médias móveis autorregressivas (ARMA) com ajustes de dois parâmetros, um para cada bloco (ACKERMANN e SELLITTO, 2022, p. 93; KORSTANJE, 2021, p. 89).

Para modelar séries não estacionárias¹³, utiliza-se o Modelo Autoregressivo Integrado de Médias Móveis, do inglês *AutoRegressive Integrated Moving Average* (ARIMA), pois o parâmetro de integração mensura a diferença necessária para tornar a série estacionária (KORSTANJE, 2021, p. 113; PARMEZAN, SOUZA e BATISTA, 2019, p. 308). O *Seasonal Auto Regressive Integrated Moving Average*, traduzido como Média móvel integrada autorregressiva sazonal (SARIMA) é sugerido quando as séries apresentem componentes de tendência e sazonalidade (KORSTANJE, 2021, p. 115; PARMEZAN, SOUZA e BATISTA,

¹³ Segundo Parmezan, Souza e Batista (2019, p. 305)) “uma série estacionária desenvolve-se aleatoriamente em torno de uma média constante, refletindo algum equilíbrio estável” (tradução livre).

2019, p. 311).

Para os modelos multivariados, também denominados causais, Ackermann e Sellitto (2022, p. 87; KORSTANJE, 2021, p. 133) lecionam que há uma variável dependente que pode ser prevista em função de duas ou mais variáveis independentes, causais ou explicativas. Assim, passam a ter adições aos modelos de vetores com as variáveis, como por exemplo: *Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors*, em português Média móvel integrada autorregressiva sazonal com fatores exógenos (SARIMAX) (KORSTANJE, 2021, p. 125, 133, 147).

2.2.2 Métodos não paramétricos

O termo “métodos não paramétricos” é atribuído aos modelos que não exigem um conhecimento prévio da distribuição dos dados para descrever suas propriedades (PARMEZAN, SOUZA e BATISTA, 2019, p. 312), pois suas relações são aprendidas automaticamente e suportam modelos univariados e multivariados (PETROPOULOS, APILETTI, *et al.*, 2022, p. 745). Nestes métodos se enquadram os baseados em computação os quais identificam relacionamentos não lineares e interativos, a AM (ACKERMANN e SELLITTO, 2022, p. 94; GÉRON, 2021, p. 6). Géron (2021, p. 6) sugere o uso de AM para problemas mais complexos, com extensas regras, inúmeros ajustes, adaptável a novos dados e quando as soluções tradicionais não solucionam. Dentre as inúmeras aplicações de AM destacam-se as previsões financeiras com dados governamentais com algoritmos de regressão, sendo os principais passando a serem descritos (PETROPOULOS, APILETTI, *et al.*, 2022).

O modelo *K-Nearest Neighbor*, ou K vizinhos mais próximos (KNN), como um modelo de regressão, ele calcula a predição pela média dos alvos (KORSTANJE, 2021, p. 169; HARRISON, 2020, p. 186; BRUCE e BRUCE, 2019, p. 271). 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 (KORSTANJE, 2021, p. 194; NIELSEN, 2020, p. 267). O *Extreme Gradient Boosting* (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 descendente (KORSTANJE, 2021, p. 195; NIELSEN, 2020, p. 267; HARRISON, 2020, p. 198; WADE, 2020, p. 54).

O *Randon Forest*, ou Floresta Aleatória, é um conjunto de árvores de decisão cultivadas sem *pruning* (poda), cujos nós têm um número fixo de atributos aleatórios, originalmente

proposto por Breiman (2001 *apud* Marron et al., 2014, p. 4) “*como um método para usar a randomização na entrada e na construção interna das árvores de decisão.*” Trabalham 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 (KORSTANJE, 2021, p. 179; NIELSEN, 2020, p. 265; HARRISON, 2020, p. 123; BRUCE e BRUCE, 2019, p. 293).

As Redes Neurais tem uma arquitetura mais complexa, custo computacional maior, uma vez que tentam imitar o processamento do cérebro humano com algoritmo de aprendizado *Backpropagation*, instituindo uma unidade básica: o *perceptron* ((KORSTANJE, 2021, p. 225-226); (PARMEZAN, SOUZA e BATISTA, 2019, p. 313)).

As *Recurrent Neural Networks*, ou Redes Neurais Recorrentes (RNN) são semelhantes às Redes Neurais, diferindo na movimentação dos neurônios em várias direções, pois suas conexões formam ciclos, *backward*, já que geram saídas recursivas, cessando informações anteriores, por isso são indicadas para dados sequenciais (KORSTANJE, 2021, p. 228; GÉRON, 2021, p. 385; RAFFERTY, 2021, p. 23; PARMEZAN, SOUZA e BATISTA, 2019, p. 314). Mesmo reconhecendo o poder das RNNs, elas têm o gradiente instável e uma limitação na memória de curto prazo, assim adotam-se arquiteturas de *Long Short-Term Memory*, traduzida como Memória longa de curto prazo (LSTM) e sua versão simplificada *Gated Recurrent Unit* (GRU) as quais detectam as dependências dos dados a longo prazo, todavia necessitam de grande quantidade de dados e recurso computacional (GÉRON, 2021, p. 385, 398 e 401; KORSTANJE, 2021, p. 243; RAFFERTY, 2021, p. 23; PARMEZAN, SOUZA e BATISTA, 2019, p. 315).

2.2.3 Seleção de Variáveis

Embora a literatura dissemine diversas definições para a tarefa de seleção de atributos (feature selection) 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.” (ALMEIDA e FERREIRA, 2018, p. 38). Harrison (2020, p. 90) ensina 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.

Corroborando para este entendimento, Baranauskas (2021) 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 autor 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. E o âmago das questões, a estratégia, permeia a avaliação por intermédio de métricas e algoritmos básicos de indução. Assim, as abordagens Filtro, *Wrapper* e *Embedded* (Embutida) são as fundamentais (ALMEIDA, 2018, p. 38). Em suma, Baranauskas (2021) leciona que o Filtro utiliza o FSS para escolher os atributos 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*, floresta aleatória, um conjunto de árvores de decisão cultivadas sem *pruning* (poda), cujos nós têm um número fixo de atributos aleatórios, originalmente proposto por Breiman (2001 *apud* Marron et al., 2014, p. 4) “como um método para usar a randomização na entrada e na construção interna das árvores de decisão.”

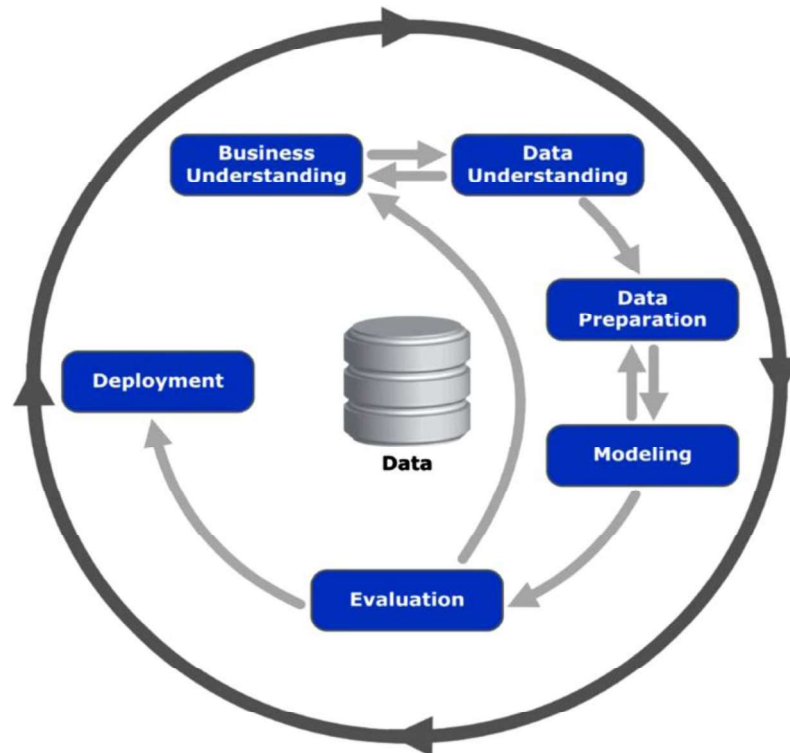
2.3 CRISP-DM

O CRISP-DM foi introduzido em 1996 com o intuito para fornecer uma padronização de diretrizes para projetos de mineração de dados em seis fases ilustradas na Figura 1 Compreensão de Negócios, Compreensão de Dados, Preparação de Dados, Modelagem, Avaliação e Implantação (SEKAR, 2022, p. 44; CHAPMAN, CLINTON, *et al.*, 2000). Como o CRISP-DM fornece apenas diretrizes, a literatura padroniza as etapas de cada fase, como um *roadmap*, elas são divididas em tarefas genéricas com os artefatos para serem produzidos conforme a Figura 2.

A fase de “Entendimento do Negócio” (*Business Understanding*) iniciou o *roadmap* para produção de uma base de compreensão do modelo de negócio. Nesse sentido, esta fase concentra-se na identificação dos objetivos, uma verdadeira etapa de elicitação de requisitos, cuja interdependência com as demais fases é extremamente vivenciada, principalmente, com a etapa de entendimento dos dados. Norteando-se pelo roteiro da Figura 2, os artefatos produzidos foram: “Determinar os objetivos do negócio” com descrição do contexto, objetivos e critérios de sucesso; “Avaliação da situação” contemplou o inventário de recursos, requisitos, suposições e restrições, riscos e contingências, terminologias e custos e benefícios; “Determinar objetivos

da mineração de dados” traçou objetivos e critérios de sucesso; e “Produzir um Plano de Projeto” com avaliação inicial de ferramentas e técnicas (SEKAR, 2022, p. 45; PLOTNIKOVA, DUMAS e MILANI, 2021, p. 5; CHAPMAN, CLINTON, *et al.*, 2000).

Figura 1 - Diagrama da estrutura do CRISP-DM com as suas 6 fases



Fonte: Sekar (2022, p. 44) e Chapman e outros autores (2000, p. 13).

O “Entendimento dos dados” (*Data understanding*) pretendeu verificar a integridade e precisão dos dados disponíveis proporcionando informações adicionais sobre o negócio. Assim, conforme a Figura 2, esta fase entregou ,em quatro relatórios, um em cada tarefa: em “Coletar dados iniciais”, o Relatório Inicial de Coleta de Dados; em “Descrever dados”, o Relatório de dicionário de dados; em “Explorar dados”, o Relatório de exploração de dados; e em “Verificar a qualidade dos dados”, o Relatório de qualidade dos dados (SEKAR, 2022, p. 45; CHAPMAN, CLINTON, *et al.*, 2000).

A “Preparação dos dados” (*Data preparation*) compreendeu a etapa mais longa do processo de mineração de dados em que os dados foram extraídos, transformados e carregados para uso em algoritmos de AM. Tal dispêndio de tempo, ocorreu em anuência a execução das seis atividades descritas na Figura 2, que foram: “Conjunto de dados” com a sua descrição; “Selecionar dados” com a fundamentação da inclusão e/ou exclusão deles; “Limpar dados” com

a produção de uma Relatório de limpeza de dados; “Construir dados” com a geração de atributos derivados e novos; “Integrar dados” com a verificação da viabilidade de mesclar dados; “Formatar dados” examina-se a necessidade de reformatar os dados (SEKAR, 2022, p. 45; PLOTNIKOVA, DUMAS e MILANI, 2021; CHAPMAN, CLINTON, *et al.*, 2000).

Figura 2 - Tarefas genéricas (em negrito) e os artefatos (em itálico) das fases do CRISP-DM



Fonte: Adaptado de (CHAPMAN, CLINTON, *et al.*, 2000, p. 15) (tradução nossa).

A “Modelagem” (*Modeling*) tem uma grande dependência da “Preparação dos dados” uma vez que esta fase necessita de dados com formato adequado para serem modelados e analisados com a finalidade de obter *insights* sobre estes, os quais foram concebidos na primeira fase de “Entendimento do negócio”. Seguindo as orientações da Figura 2, têm-se as seguintes tarefas: “Selecionar a técnica de modelagem”, tendo como pressupostos os dados disponíveis para mineração, as metas e os requisitos, e foi executada a divisão entre conjuntos de treino e teste, bem como a verificação da quantidade e dos tipos de dados. Passando-se para a atividade de “Suposição de modelagem”, em que se averiguou a necessidade de ajustes nos dados para aplicação de algoritmos. “Gerar projeto de teste” definiram-se quais os critérios foram testados, como taxa de erro, tempo de processamento, qual a mensuração do êxito do modelo; “Modelo de construção” verificou-se e documentou-se as configurações dos parâmetros, a execução dos modelos e seus resultados. E “Modelo de avaliação” em que se avaliam o critérios em função dos objetivos e a interpretação dos resultados para fazer uma revisão dos modelos (SEKAR, 2022, p. 45; PLOTNIKOVA, DUMAS e MILANI, 2021; CHAPMAN, CLINTON, *et al.*, 2000).

Plotnikova, Dumas e Milani (2021) enfatizam que a fase de “Avaliação” (*Evaluation*) almejou a confirmação do atendimento do(s) objetivo(s) do negócio por intermédio do modelo construído. Aproveitando o ensejo, os autores protestaram contra o CRISP-DM por não apresentar um roteiro de como deve ser esta validação de qualidade em cenários reais. Todavia, os pesquisadores, assim como Duboue (2020, p. 12), definem duas etapas de validação simultâneas: uma técnica e a outra, verificando a significância do modelo no ambiente de negócios, ambas ocorrendo na fase de Modelagem (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 8). De acordo com Sekar (2022, p. 46) “[...] após essa etapa, pode ser necessário atualizar o(s) objetivo(s) de entendimento do negócio e repetir os ciclos anteriores (como mostrado na Figura 4-1) até que a precisão seja melhorada.”

Na fase de Implantação (*Deployment*) fez mister tornar o resultado compreensível para os gestores e técnicos, assim, Plotnikova, Dumas e Milani (2021, p. 9-10) sugeriram a busca de indicadores de desempenho e visualizações mais simples, considerando os cenários de implantação com a identificação dos seus requisitos. Sekar (2022, p. 46) sugeriu que “[...] no caso de um painel, ele é compartilhado com usuários autorizados para que eles possam utilizar insights em seu processo de tomada de decisão.” Esta fase tem por trabalho planejar a implementação, o monitoramento e a manutenção, produzir um relatório final e realizar uma revisão do projeto final (CHAPMAN, CLINTON, *et al.*, 2000).

3 TRABALHOS CORRELATOS

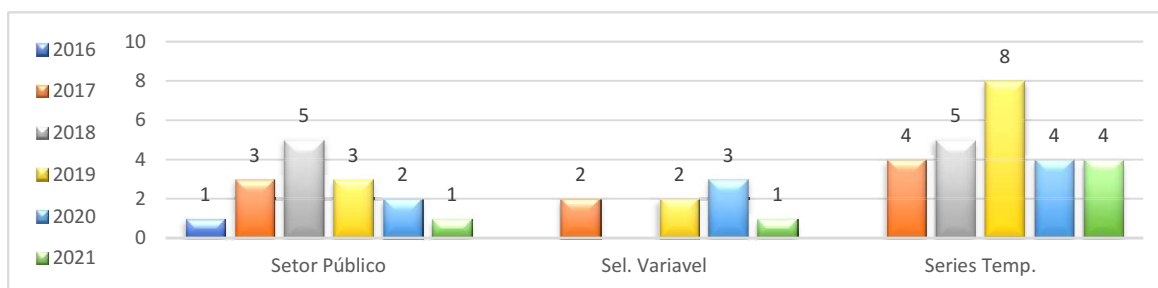
Como mencionado anteriormente, a aplicação do problema envolve a previsão das receitas do Setor Público cujas consequências permeiam a divergência entre valores executados e previstos para evitar contingenciamento de despesas. Neste sentido, a busca de trabalhos científicos foi feita nas bases do Portal de Periódicos Capes e no site do Google Acadêmico. Foram utilizados os termos: “previsão receitas públicas” OR “previsão receitas”, textos com idioma em inglês e português, no período de 2016 até agosto de 2021. Também foram procurados os termos: “Séries temporais” AND “Previsão” OR “Índices Econômicos” OR “Mercado financeiro” OR “Forecast”.

Dentre as publicações apresentadas nas pesquisas, elegeram-se como critérios de seleção os artigos que estivessem na área de atuação do Setor Público, ou mesmo relacionados a Séries Temporais, principalmente com atividades afins, como previsão de índices econômicos, correlação entre esses índices, demandas de previsões ou mesmo mercado financeiro, além da seleção de variáveis, pois a proposta de solução a envolve, ao buscar as melhores variáveis para previsão de cada receita. Destarte, dividiu-se as duas subseções, sendo a primeira um resumo quantitativo das publicações mais relevantes por ano e assunto, e a outra, trata-se dos estudos aplicados de previsão de receitas no setor público e a discussão da literatura.

3.1 RESUMO QUANTITATIVO

Dentro de três categorias que refletem na aplicação do problema identificado, encontraram-se 48 publicações, das quais 15 no Setor Público, 25 com Séries Temporais e 8 com seleção de variáveis, conforme a distribuição por ano de publicação no Gráfico 3.

Gráfico 3 - Distribuição por ano e tipo de abordagem das publicações no período de 2015 até agosto de 2021



Fonte: Elaborado pela Autora.

As 15 publicações com aplicação no Setor Públicas foram resumidas na Tabela 3, uma vez que elas tratam do cerne da problemática abordada. Na referida totalização ficou evidente que a maioria das publicações utilizaram métodos estatísticos, 13 documentos, e apenas 2

usaram algum método de AM.

Tabela 3 - Resumo dos trabalhos correlatos por Autor, método, técnicas e local

Autor/Ano	Método		Previsão Receita	Técnicas usadas															Local	Sazonal	Natureza da metodologia de aplicação da publicação
	Estatístico	AM		ARIMA	SERIMA	Autoregressiva	Correlação	Theta	MLP	Holt	ARMA	Snape drift	ruv2	Koyck	FMOLS	PCA	AutoARIMA	desagregação temporal			
Naibert (2021)	1		ICMS	1															RS		Comparativa ARIMA <i>Box-Jenkins</i> (BJ) com MTO
Rodrigues, Aquino, <i>et al.</i> , (9-11 nov. 2020)	1		ICMS		1														RN	1	Comparativa Rec.Exc. c/HWA e HWM
An e Jalles (2020)	1		Geral			1	1												EUA	1	Comparativa dos erros CBO c/Setor Privado
Sousa, Correia, <i>et al.</i> (2019)	1		ICMS	0	1														CE	1	Previsão de nov/18-out/19 SERIMA c/ <i>drift</i> e BJ
Carmo, Komati e Boldt, (2019)	1	1	ICMS	1				1	1	1		1	1						ES	1	Comparativa de 8 métodos estatísticos com os de ML
Pamplona <i>et. al.</i> , (2019)	1		Geral								1								PR		Comparativa ARMA com MTO de 120 municípios
Silva, Shiaku, <i>et al.</i> , (2018)	1		Multas Trânsito								1								MG		Comparativa ARMA com MTO
Almeida e Ferreira, (2018)	1		ISSQN		1														PE	1	Comparativa entre previsão Linear e Autorregressiva, modelos de Tendência Linear
Fiirst, <i>et al.</i> , (2018)	1		Geral											1					PR		Desempenho da previsão das receitas dos municípios
Noch, Pattiasina e Purwanto, (2018)	1		PAD												1				Ind.		Medir a correlação das receitas próprias e encargos
Mikesell, (2018)			Geral																EUA		Apresenta 9 lições sobre projeções de receitas
Silva, Jacovine, <i>et al.</i> , (2017)		1	CFEM royalties petróleo													1			BR		Correlação entre variáveis ambientais, econômicas e sociais c/ <i>royalties</i> e CFEM
Azevedo, Silva e Gatsios, (2017)	1		ICMS	1															BR		Comparativa ARIMA com LOA e Arrecadação
Almeida, Divino, <i>et al.</i> , (2017)	1		não														1		BR		Calcular alíquota média sobre consumo, trabalho e capital
Silva, Albuquerque, <i>et al.</i> , (2016)	1		ICMS	1													1		PA	1	Explicar o comportamento da arrecadação do ICMS
TOTAIS	13	2		4	2	2	1	1	1	1	2	1	1	1	1	1	1	1		6	

Fonte: Elaborado pela Autora (2023).

As referidas publicações serão descritas no subitem seguinte por cada tipo de abordagem do problema. As outras publicações relacionadas a Séries Temporais e Seleção de Variáveis não foram descritas individualmente por estarem sendo abordadas de forma tangencial ao problema.

3.2 ESTUDOS APLICADOS PREVISÃO DE RECEITAS NO SETOR PÚBLICO

Na literatura foram encontradas duas pesquisas que comparam o modelo proposto no MTO da SOF com modelos paramétricos. O mais recente, a dissertação de Naibert (2021) demonstrou a supremacia do modelo ARIMA em 83% das previsões de 12 municípios do Rio Grande do Sul em valores globais anuais das receitas. Na conclusão de Naibert (2021), ele não

garantiu que o método 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. O segundo, de Pamplona e outros autores (2019), difere ao fazer a comparação com o ARMA em 120 cidades do Paraná com erro médio de 7,05% em 74 casos.

No âmbito Estadual e com uma visão da classificação das receitas, pesquisadores elaboraram estudos comparativos da previsão da receita de ICMS utilizando métodos paramétricos. Assim, Rodrigues e outros autores (2020) compararam os modelos SARIMA com alisamento exponencial *Holt-Winters* aditivo e multiplicativo, no Rio Grande do Norte, ao final o método aditivo obteve o melhor resultado. No Ceará, Sousa e outros autores (2019) utilizaram *Box-Jenkins* em modelo **estatístico** SERIMA com *drift*, obtendo uma Média Percentual Absoluta do Erro (MAPE) de 4,72%. Azevedo e outros pesquisadores (2017) obtiveram bons resultados com aplicação de ARIMA em 6 Estados (SP, MG, RJ, RS, PR e BA). Já no Pará, Silva e outros autores (2016) obtiveram melhores resultados com o modelo SARIMA, observando-se alguns *outliers*.

Carmo, Boldt e Komati (2019) trazem a primeira publicação com uma comparação de 8 métodos, entre **estatísticos** e **ML**, propõem uma previsão de ICMS do Estado do Espírito Santo. Seguindo métricas e equações, com código aberto, similares as competições M3 e M4 de Makridakis e outros autores (Makridakis et al. 2018a apud (CARMO, BOLDT e KOMATI, 2019)). Os métodos escolhidos pelos autores foram os da competição M4, incluindo o *Multilayer Perceptron* (MLP) e a combinação dos métodos univariados, extraíndo o componente de sazonalidade em todos eles. Como resultados apontaram a eficácia das técnicas em cascata para seleção de variáveis explicativas de importação e exportação, sendo o modelo combinado estatístico o que prevaleceu sobre os demais (CARMO, BOLDT e KOMATI, 2019).

Uma previsão para receitas de Multas de Trânsito de Uberlândia, em Minas Gerais, foi publicada por Silva e outros autores (2018) utilizando o método **estatístico** ARMA o qual teve êxito em 9 dos 16 períodos projetados. No artigo de Almeida e Ferreira (2018) foi proposta a previsão do Imposto Sobre Serviços de Qualquer Natureza do Recife, Capital de Pernambuco, utilizando modelos de Tendência Linear e Autorregressivo com êxito. Discutindo a eficiência da previsão orçamentária do Setor Público, Fiirst e outros autores (2018) fazem uma comparação das previsões das receitas dos Municípios do Paraná utilizando o modelo **estatístico** *Koyck*, cujos resultados obtiveram menores erros. Todavia em anos eleitorais as previsões municipais são melhores.

No âmbito internacional, An e Jalles (2020) investigam as causas dos erros das previsões fiscais das receitas e despesas do governo para a economia dos Estados Unidos da América (EUA) advindas do *Congressional Budget Office*, traduzido por Escritório de Orçamento do Congresso (CBO). Eles focam em 3 categorias de erro: política, econômica e técnica. Utilizaram autorregressões e médias móveis. Como resultado constatam que as médias do CBO para receitas são baixas e despesas altas, fazendo correlações com o PIB. Entendem que o componente econômico explica a receita e o político as despesas. Ainda fazem um comparativo das previsões oficiais do CBO com as dos Economistas do Setor Privado que apresentam “melhor prática”.

No estudo **estatístico** do Governo Provincial da Indonésia, Noch e outros autores (2018) buscam medir a elasticidade dos impostos e encargos locais da previsão de Receitas próprias, utilizando o método Mínimo Quadrado Totalmente Modificado. Inferindo que tanto os impostos locais quanto os encargos mostram bons resultados do que a previsão real mostrada. *Mikesell* (2018) traz uma abordagem mais teórica sobre o entendimento da estrutura tributária, e o quão importante as previsões são. O Autor elenca 9 (nove) lições sobre o comparativo dos erros de previsão das receitas do Estado de Indiana nos EUA.

O segundo artigo mais próximo do problema é o de Silva e outros autores (2017) cujo intuito foi identificar a correlação entre variáveis ambientais, econômicas e sociais com os *royalties* arrecadados nos principais municípios produtores de petróleo e minério. Os autores utilizaram o pacote *Vegan* do R para aplicar **Análise de Componentes Principais (PCA)**, em 183 observações as quais resultaram na evidenciação de nenhuma proporcionalidade apontando como justificativa a ausência de um dispositivo legal na aplicação destes recursos.

Almeida e outros autores (2017) propõem o cálculo de alíquotas tributárias relacionadas ao consumo, à renda do trabalho e a do capital, aplicando métodos de desagregação temporal. Concluindo que a tributação no Brasil onera a renda do trabalho e do capital, quando comparada com a do consumo.

3.3 DISCUSSÃO DA LITERATURA E A PRÁTICA

Vários autores (NAIBERT, 2021, p. 15; PAMPLONA, FIIRST, *et al.*, 2019; LIMA, LUI, *et al.*, 2020), afirmam que, na prática, os Municípios utilizam o modelo oferecido pela SOF. Vale ressaltar que os pesquisadores em suas tabelas demonstraram previsões anuais, não abordando o detalhamento da classificação das receitas, prerrogativas essenciais para o Ente Subnacional, conforme fases de Entendimento do negócio e dos dados do CRISP-DM. Todavia,

ao compararem várias cidades tiveram o cuidado com a heterogeneidade entre elas e agruparam-nas por população e estratificação por valores de receitas, um cuidado que esta pesquisa também está considerando, haja vista a escolha apenas de Capitais (NAIBERT, 2021, p. 16-17; PAMPLONA, FIIRST, *et al.*, 2019, p. 93).

Pode-se observar que os artigos que abordam o domínio da aplicação, previsão de receitas do Setor Público, estão voltados para utilização de métodos estatísticos, direcionados a substituir o modelo da SOF. Encontrou-se um artigo que faz uma comparação entre 8 métodos, introduzindo algoritmos de combinação de modelos e AM (CARMO, BOLDT e KOMATI, 2019), e neste mesmo artigo faz a sugestão de aplicar RNN e LSTM para previsões, uma das motivações para este desenvolver este estudo, conforme comentado na justificativa. O outro artigo relacionado (SILVA, JACOVINE, *et al.*, 2017), não tem como cerne a previsão de receitas, mas a identificação da correlação de variáveis, usando PCA. Assim, diferindo da literatura não se pretende substituir o modelo da SOF, uma vez que se entende que induz a utilização de autorregressão, mas multivariado, por isso pretende-se testar vários modelos para verificar o que melhor se ajusta, uma forma de atender às prerrogativas imposta pela legislação vigente.

Neste sentido, entende-se que há poucas publicações voltadas para o estudo de métodos de AM em previsão de receitas públicas, e menos ainda com a proposta de seleção de variáveis para cada receita, bem como a relevância de manter a sazonalidade da série temporal, no que tange as publicações dos instrumentos orçamentários, os relatórios bimestrais e a própria execução orçamentária que ocorre mensalmente, bem como identificar as variáveis do efeito preço e quantidade da LRF. Portanto, a proposta de aplicação de técnicas de AM para previsão das receitas públicas, em séries temporais com utilização de RNN e LSTM é uma indicação dos pesquisadores Carmo, Komati e Boldt, (2019).

4 RELATO DAS ATIVIDADE DO ESTUDO DE CASO DE SÃO LUÍS APLICANDO DO CRISP-DM

Mediante os objetivos outrora traçados, a proposta deste trabalho de regressão, que visa a comparação entre modelos estatísticos e de técnicas de AM como uma alternativa de apoiar a implantação das regras da SOF para a LRF na previsão das receitas públicas, utilizou o CRISP-DM para buscar as melhores previsões e variáveis explicativas com base nas séries históricas das capitais. Portanto, o capítulo teve com escopo registrar como as atividades ocorreram, desde a compreensão do negócio e dos dados, detalhando a escolha das variáveis e construção dos *datasets*, a fase de preparação dos dados, com os procedimentos adotados para ajustar as bases para a modelagem avaliação e parte da implementação foram resumidas nas seis subseções.

É importante registrar que toda a execução foi elaborada baseando-se na primeira iteração, utilizando os dados da Capital do Maranhão, São Luís, pretendia-se aplicar o mesmo processo, com foco na análise das questões propostas nos objetivos para a previsão das receitas públicas das 8 capitais que aceitaram participar deste experimento. Todavia devido a problemas físicos em hardwares, muitos registros, dados e documentos foram perdidos, atrasando o cronograma de execução.

4.1 PRIMEIRA FASE DE ENTENDIMENTO DO NEGÓCIO

Como relatado no subitem 2.3, esta fase concentra-se na identificação dos objetivos do negócio (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 5). Todavia é conveniente relembrar que até a SOF assume que é uma área de difícil compreensão o que fez desta fase um dos principais motivos para escolha desta metodologia (BRASIL, MINISTÉRIO DA FAZENDA, mar. 2016, p. 7).

Seguindo o *roadmap* desta etapa, Figura 2, a primeira tarefa foi determinar os objetivos do negócio, cujo primeiro artefato, de Histórico, foi descrito na introdução deste documento, uma prática comum (SCHRÖER, KRUSE e GÓMEZ, 2021, p. 529), contendo uma contextualização das previsões das receitas públicas e contemplando a definição dos objetivos. Quanto aos Critérios de sucesso do negócio, consistem em obter o menor erro, ou melhor, a menor divergência quantitativa entre os valores orçados e executados na previsão das receitas, já na seleção das variáveis precisam atender os critérios impostos pela SOF descritos no item 2.1.2 (BRAND, 2014, p. 77; CHAPMAN, CLINTON, *et al.*, 2000, p. 16-17). Quanto a mensuração do critério, o percentual de São Luís, cujo valor foi um erro igual ou menor que 5% para evitar um decreto de contingenciamento, informação obtida em diálogo informal com

o Secretário Adjunto de Planejamento.

Para avaliação da situação, no Inventário de recursos, São Luís, possui o Sistema de Gerenciamento de Banco de Dados é o *Oracle Database*, o software utilizado como Sistema Único e Integrado de Execução Orçamentária, Administração Financeira e Controle¹⁴ do Município é o do Grupo GIAP. Os experimentos foram conduzidos em um computador equipado com processador *Intel(R) Core(TM) i5-1030H CPU @ 2.5GHz*, com placa de vídeo *GeForce GTX 1650* de 4 GB, memória RAM de 8GB e disco SSD M2 de 512GB, no sistema operacional *Microsoft Windows 11*, utilizando o *Jupyter Notebook* com a linguagem *Python 3* para execução dos algoritmos de AM e para a unificação do *Dataset*, o software *Power BI Desktop*, versão: 2.96.901.0 64-bit, ferramenta gratuita da *Microsoft*.

Quadro 3 - Listagem das séries históricas por Capital, ano de início, protocolo das suas Ouvidorias e respostas

Município	Ano Inicial	Protocolo	Resposta Protocolo
Belém - PA	2010	---***---	Disponível no portal
Belo Horizonte - MG	2021	31.00519268/2022 e 31.00519425/2022	Enviaram um arquivo sem identificação das receitas com valores totais
Campo Grande - MS	2010	---***---	Disponível por API
Fortaleza - CE	2010	---***---	Disponível no portal
Maceió - AL	2011	2700.120866.2022 e 1000.120868.2022	Informaram que há outra versão do portal da transparência, mas limita-se até 2011.
Palmas - TO	2018	2022111436925 e 2022111436927	Solicitação negada, tem que fazer pedido na Sede da Secretaria de Finanças de Palmas-TO
Salvador - BA	2015	2022060557295 e 2022060557293	Portal de difícil acesso, ainda sem resposta.
São Luís - MA	2010	11111000064202111 e 11111000066202282	Mesmo não tendo resposta, já possuía os dados por varredura no Diário Oficial do Município

Fonte: Elaborado pela Autora (2023).

As questões de Requisitos, premissas foram encapsuladas no *pipeline* do software *Power BI Desktop* como metadados, tendo o código fonte disponibilizado ao Laboratório de Inteligência Computacional e Programação da Universidade Estadual do Maranhão. As restrições consistem em uma vasta legislação, bases de dados distintas e uma transparência municipal de séries históricas esparsas. No que concerne o artefato de Terminologia, foi iniciado um glossário, todavia observou-se que seria desnecessário, neste momento, uma vez que o Senado Federal¹⁵, a SOF¹⁶ e a STN¹⁷ possuem em seu site. Quanto aos custos não se aplicaram a este estudo e benefícios foram descritos no capítulo 1, na justificativa (BRAND, 2014, p. 78; CHAPMAN, CLINTON, *et al.*, 2000, p. 17-18).

¹⁴ Adoção de SIAFIC é uma determinação legal e está regulamentado pelo Decreto Nº 10.540, de 5 de novembro de 2020, o qual dispõe sobre o padrão mínimo de qualidade do Sistema Único e Integrado de Execução Orçamentária, Administração Financeira e Controle.

¹⁵ Disponível em: https://www12.senado.leg.br/orcamento/glossario?b_start:int=60. Acessado em 17. Julho 2022.

¹⁶ O Glossário da SOF está disponível em: <https://www.economia.gov.br/ acesso-%C3%A0-informa%C3%A7%C3%A3o/250-planejamento/or%C3%A7amento/5410-gloss%C3%A1rio-t%C3%A9cnico.html#v>. Acessado em 17. Julho. 2022.

¹⁷ Disponível em: <https://www.tesourotransparente.gov.br/sobre/glossario-do-tesouro-nacional>. Acessado em 17. Julho 2022.

A atividade de determinar o objetivo da Mineração dos Dados, tem como entregáveis os relatórios com esta finalidade e o de critérios de sucesso os quais podem ser resumidos como a unificação das várias bases de dados, criando um *Dataset* para cada experimento que for realizado, sendo os dados disponibilizados em um dashboard que poderá ser manipulado por outros técnicos da Capital, mas não houve tempo hábil para construir o referido painel, pela parda dos dados iniciais. Já os critérios de sucesso de mineração de dados estão pautados, na parte quantitativa, na precisão do modelo ou melhoria preditiva em comparação com um método existente, discutido no item 2.1.2, e na qualitativa, relacionada a escolha das variáveis sob a apreciação do corpo técnico da Capitais em ulterior reunião do GT-08 da ABRASF, o que não foi possível ocorrer, uma vez que o início de exercício demanda fechamento de balanço, relatórios de prestação de contas e abertura do orçamento, sendo agendada preliminarmente para abril de 2023 (BRAND, 2014, p. 79; CHAPMAN, CLINTON, *et al.*, 2000, p. 18).

Quadro 4 - Avaliação das etapas concluídas do projeto inicial

Processo	Tarefa	Dados São Luís	Demais Capitais	Observação
Fonte de dados	Dados obtidos com a ferramenta <i>Power BI Desktop</i> entre planilhas eletrônicas, <i>web scraping</i> , <i>web crawling</i> , <i>download</i> em <i>sites</i> , <i>APIs</i> e <i>scripts</i> em <i>Python</i>	✓	✓	Dificuldades com os dados dos balancetes, migrando para obter pelo portal do Banco do Brasil
Integração dos dados	Construir o <i>Dataset</i> , no <i>Power BI Desktop</i> , utilizando linguagem “M”, “DAX” e <i>Python</i> , gerando um arquivo do tipo <i>CSV</i>	✓	✗	Ausência de padronização nos dados das demais Capitais, demandam mais tempo para integrar, optando
Distribuição do Pré-Processamento	Fazer um pré-processamento do <i>Dataset</i> utilizando o <i>Power BI Desktop</i> e o <i>Jupyter Notebook</i>	✓ ✗	✗	Nem todas as variáveis de São Luís podem retirar os <i>outliers</i>
	Selecionar as variáveis pela importância delas nos algoritmos <i>ensembles</i>	✓	✗	Foram aplicados apenas nos dados de São Luís no horizonte temporal de 3 meses
Consumo de dados	Aplicar os algoritmos de predição com as <i>features</i> escolhidas	✓ ✗	✗	Não houve tempo hábil para aplicar todos os algoritmos
	Avaliar os menores erros com as métricas	✓	✗	
	Repetir aplicação todas as etapas 4 vezes para a previsão nos horizontes bimestral, trimestral, anual (LOA) e quadrienal (PPA)	✗	✗	
Construir API de Previsão	Criar uma API no padrão REST, usando linguagem <i>Python</i> e biblioteca <i>Flask</i> e comunicação em formato <i>Json</i>	✓ ✗	✓ ✗	Apenas uma versão inicial foi codificada para esta primeira iteração, não foi implementada num <i>dashboard</i>
	Apresentar resultados em reunião aos técnicos do GT-08 da ABRASF para discussão e avaliação por aplicação de questionário	✗	✗	O GT-08 não teve agenda disponível senão após abril de 2023

Fonte: Elaborado pela Autora (2023)

No que tange o plano de projeto inicial, resumido na Figura 3, não houve tempo hábil

para incluir as receitas das outras 7 Capitais no estudo, pelo fato de não terem sido enviados os dados por autorização formal, pois apenas Belém, Fortaleza e Campo Grande disponibilizaram em seus Portais da Transparência os dados desde 2010 até 2022. Tal fato demandou a solicitação formal as suas Ouvidorias, conforme relacionado no Quadro 3. Para melhor verificação da produção ocorrida na primeira iteração, elaborou-se um resumo das tarefas cumpridas e as não do projeto inicial elencadas no Quadro 4.

Quanto à última tarefa desta fase, avaliação inicial de ferramentas e técnicas, há várias ferramentas no mercado para as tarefas (SCHRÖER, KRUSE e GÓMEZ, 2021, p. 529), inclusive para documentação do processo CRISP-(DM (MARTÍNEZ-PLUMED, CONTRERAS-OCHANDO, *et al.*, 2021, p. 3050), mas foram priorizadas as que não tivessem custos, não dependessem de autorizações dos Setores de Tecnologia, questão de celeridade, principalmente, licitações para aquisição de softwares. Neste sentido, o *Power BI Desktop* é gratuito, outrossim, a Autora possui cursos de qualificação para a ferramenta, agilizando a utilização e disseminação do conhecimento para outros técnicos. O *Jupyter Notebook* tem a sua versão *online* no *Google Colab*, facilitando o uso sem necessidade de instalação nos equipamentos das Secretarias de Finanças.

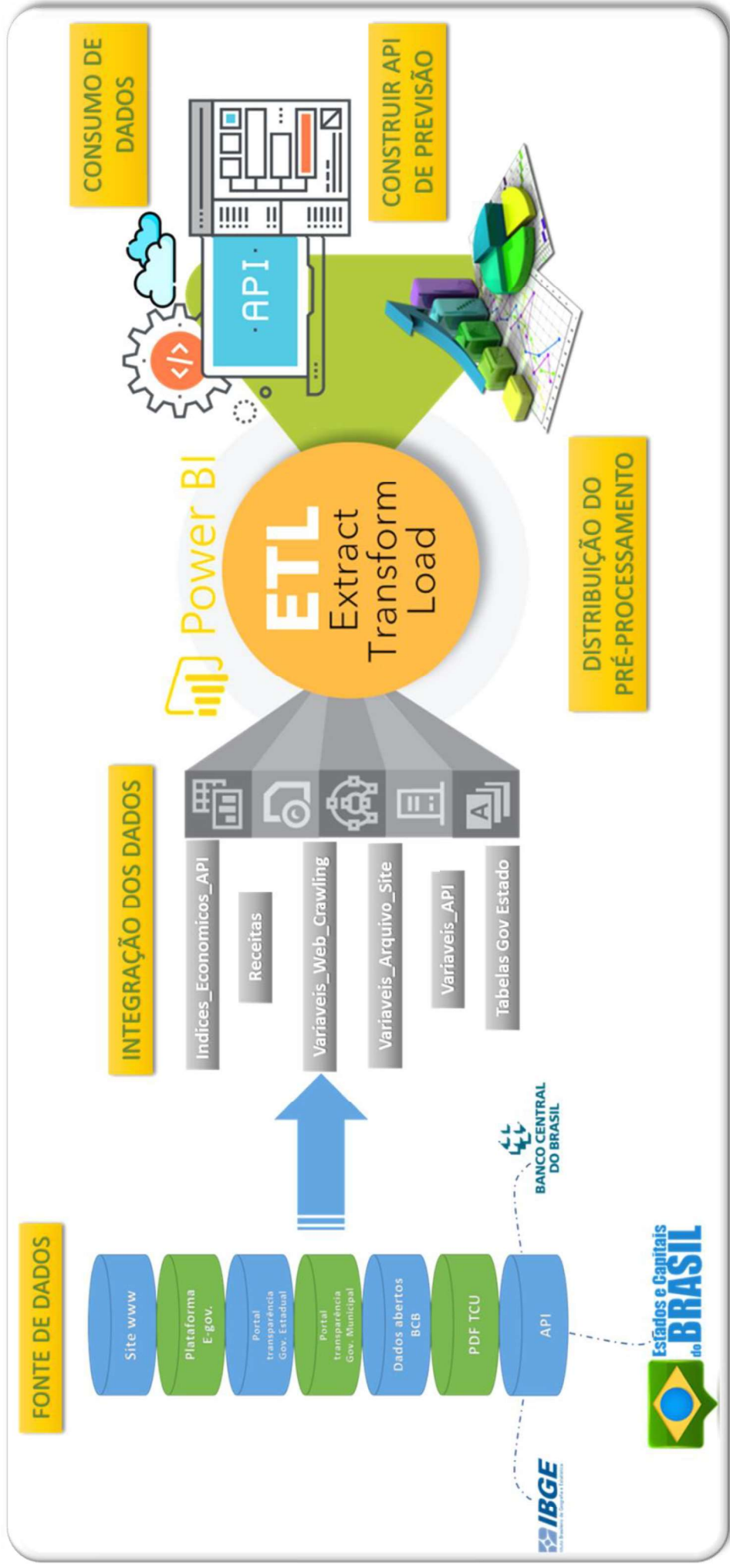
4.2 SEGUNDA FASE DE ENTENDIMENTO DOS DADOS

Na primeira iteração realizada, com os dados do Município de São Luís, foi obtida uma aplicação prática das quatro tarefas desta fase, tal fato levou-se a concordar com Plotnikova, Dumas e Milani quando confirmaram a interdependências desta etapa com a anterior e sua forte iteração, os autores ainda comentam que chegam a substituir dados com o intuito de aprimorá-los ou incluir novos. (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 6). Assim, os exemplos dos relatórios e tarefas desta fase estão relacionados aos dados de São Luís, as demais capitais não foram todas incluídas, ficará para a segunda iteração.

Assim, a “Coleta dos dados” de São Luís iniciou-se com um pedido formal ao site da Ouvidoria, mas como não foi respondida¹⁸ a solicitação dos dados, todos foram adquiridos nos Portais da Transparência, evitando quaisquer questões de sigilo fiscal, o que poderia ser um risco se fossem tratadas as receitas próprias (BRASIL, 2011). Sendo importante enfatizar que

¹⁸ Foi cadastrada no dia 22 de novembro de 2021, como o e-mail institucional de aluno da UEMA, sob protocolo nº 11111000064202111, o acesso às séries históricas dos dados das receitas municipais no site e-OUV, no Sistema de Ouvidorias do Poder Executivo Municipal de São Luís, disponível no endereço: <http://eouv.saoluis.ma.gov.br:82/publico/Manifestacao/RegistrarManifestacaoBs.aspx>. E, no mesmo site, em 24 de novembro de 2022, nova solicitação, sob protocolo 11111000066202282.

Figura 3 - Esquema do plano de projeto inicial



Fonte: Elaborado pela Autora (2023)

os dados obtidos não ferem nenhum dos ditames da Lei Nº 12.527, de 18 de novembro de 2011, a qual versa sobre o acesso à informação no setor público brasileiro, já que elas estão disponíveis nos sites de Transparência de Entidades do setor público (BRASIL, 2011).

Os requisitos para coleta dos dados foram observados na fase de entendimento dos dados, tendo suas descrições e especificações, bem como as relações de cada Receita Pública com as suas possíveis variáveis preditoras (explicativas ou dependente) elencados no APÊNDICE D, baseando-se na Fundamentação Teórica. Vale comentar que as Transferências da União possuem um padrão de divulgação, em detalhamento de repasses no Portal da Transparência do Governo Federal, já as Transferências dos Estados, provavelmente poderão apresentar mais buscas ou alterações desta proposta ao longo da execução por falta de padronização entre os Governos Estaduais.

A maior parte da tarefa de coleta dos dados foi elaborado na ferramenta *Power BI Desktop*, foi necessário distribuir o pré-processamento dos dados em vários grupos de relatórios¹⁹ para fazer extração, transformação e carregamento dos dados, que por *script* em *Python*, geraram arquivos em CSV já ajustados, os quais foram unificados em um relatório geral, conforme processos de integração dos dados e distribuição do pré-processamento da Figura 3. E, na prática, observou-se que quando os relatórios são divididos e unificados em outros, seu tempo de processamento reduz; lembrando que a versão da ferramenta é gratuita, por isso algumas limitações. Quanto ao artefato desta atividade, ele foi registrado na ferramenta, ficando os títulos no painel para visualização, assim o relatório inicial de coleta de dados, apenas da Capital São Luís, está resumido no APÊNDICE C.

Para as tarefas de “Descrever dados” e “Explorar dados” (CHAPMAN, CLINTON, *et al.*, 2000, p. 19), foram incluídos os “APÊNDICE D” e “E”. O primeiro, trata do Relatório de descrição de dados, a tabela elencou todas as 82 variáveis testadas na primeira iteração com os dados de São Luís. Foram identificadas, por tabela de origem da tarefa anterior, sua relação com cada receita, uma breve descrição, a periodicidade de publicação, frequência das observações e a identificação da fonte de origem. Sendo a variável “Data” separada como índice e a “Realizado” da tabela “*f_Programaca_Financeira*” a variável alvo. O segundo, versa sobre o Relatório de Exploração de Dados, este foi elaborado no *Jupyter Notebook*, utilizando o pacote *pandas profiling*²⁰ na linguagem *Python*.

¹⁹ O termo “relatório do *Power BI* é composto por uma ou mais páginas de visualizações, grafos e texto. Todas as visualizações em um relatório vêm de um único conjunto de dados[...]” (HART, RATHORE, *et al.*, 2022)).

²⁰ Maiores detalhamento disponíveis na documentação do pacote em <https://pandas-profiling.ydata.ai/docs/master/index.html>.

A atividade de “Verificar a qualidade dos dados” que remete a preparação de um Relatório de qualidade de dados ficou mais ágil com a utilização do pacote do *pandas profiling*, uma vez que foi possível identificar a divergência nos dados das RECEITAS, pois esta variável deveria ter a mesma quantidade de observações para todas, no caso 132 conforme distribuição da Tabela 4, com exceção da CFEM AFETADOS que iniciou a distribuição no exercício de 2019 e as receitas das Emendas Constitucionais (EC) do FPM que são repasses anuais.

Tabela 4 - Distribuição quantitativa e percentual das 928 observações por receita

Receita	Qnt. Observações	% do Total
FPM	132	14,20%
FEP	130	14,00%
ICMS	130	14,00%
IPVA	130	14,00%
CFEM	126	13,60%
IPI	124	13,40%
ITR	123	13,30%
CFEM AFETADOS	18	1,90%
FPM EC 55/07	12	1,30%
FPM EC 84/14	3	0,30%
Total	928	100,00%

Fonte: Elaborado pela Autora (2023), adaptado do APÊNDICE E.

A questão da identificação, principalmente, no repasse da cota-parte do IPI-Exportação, levou a notificação formal do Agente Financeiro e do Governo do Estado do Maranhão a prestarem esclarecimento, os quais culminaram no depósito dos valores descobertos na 1ª iteração apontados como ausentes dos meses de janeiro e maio, em junho de 2022 (ver APÊNDICE F). Ainda foram encontrados outros valores que não estão bem compreensíveis, quando são comparados os montantes que o Governo Federal divulga com o que o Governo Municipal de São Luís registra em seu Balancete e as publicações do Agente Financeiro. Os repasses dos comunicados extraordinários do FPM e IPI foram estornados dos montantes do *dataset*, haja vista que sua periodicidade é inconstante e constituem-se em *outliers*. Ademais, os valores atípicos, como exemplo da CFEM do Produtor que foi explicado por uma multa aplicada a uma empresa e foi estornado do montante como um *outlier*. Destarte, pretende-se repassar a cada uma das Capitais o mesmo relatório de descoberta da qualidade dos seus dados, caso tenha ocorrência semelhante.

4.3 TERCEIRA FASE DE PREPARAÇÃO DOS DADOS

O conjunto de dados para a fase de “Preparação de dados” teve sua descrição definida no APÊNDICE D. A proposta foi seguir a mesma ordem das atividades do CRISP-DM para esta etapa em que foram produzidos os dados para mineração (CHAPMAN, CLINTON, *et al.*,

2000, p. 25 e 65). Toda esta fase foi desenvolvida no *Jupyter Notebook*, e como embasamento para a tarefa de “Selecionar dados” utilizou-se as informações do APÊNDICE E: Relatório de exploração de dados com o *pandas profiling* de São Luís, tomando as seguintes decisões:

- 1) Os valores com “ZEROS”, relacionados no relatório, em comento, foram todos excluídos, pois não vão ter influências nas receitas com os percentuais acima de 50%;
- 2) Os valores constantes das variáveis “FtPopulacao”, “PrcInvRndPrCpMn” e “FtInvRndPrCap_Mun” são derivados dos cálculos do coeficiente CIFPM e para uma única Capital não há relevância, sendo retirado, mas para as 8 cidades juntas elas serão incluídas, com o intuito de testar a importância delas nos algoritmos;
- 3) Na série das receitas, especificamente sobre a variável, “RECEITA”, os dois repasses das Emendas Constitucionais do FPM, por serem anais e as demais mensais foram excluídas do estudo, a CIDE por seu repasse ser trimestral e o ITR por sua pequena relevância financeira.
- 4) Os valores trimestrais foram divididos por 3 para refletir uma média para esta primeira iteração, mas pretende-se testar a desagregação temporal proposta pelo Banco Central do Brasil (BCB, BANCO CENTRAL DO BRASIL, 2021);
- 5) Os valores são todos nominais, mesmo tendo construído uma função para o deflacionamento, que foi aplicada em algumas tabelas, mas não em todas, por isso, o *dataset* deflacionado será testado em uma iteração posterior (LOSSO, 2020).

Seguindo as orientações da atividade de “Limpar dados” (CHAPMAN, CLINTON, *et al.*, 2000, p. 49), voltou-se a analisar o APÊNDICE E, no qual observa-se que não há indicação de dados ausentes, todavia a falta de padronização na quantidade das observações da Tabela 4 é o maior indicio da inconsistência nos intervalos temporais, o que foi comprovado no APÊNDICE F. Assim, o Relatório de limpeza de dados e os demais das etapas de “Construir dados” e “Integrar dados” foram compilados no APÊNDICE G, com destaque para descoberta da ausência dos depósitos da Receita do IPI Exportação que originaram o APÊNDICE F. A Tabela 5 demonstras a interrupção temporal dos dados, sendo efetuada uma pesquisa dos depósitos junto ao site de transparência do Governo Federal. Tal pesquisa demonstrou que haviam divergências tanto na ausência destes dados quanto nos valores de outras observações, uma vez que foram identificados ingressos similares em meses subsequentes cujo total anual não igualou com a referida publicação (maiores detalhes no APÊNDICE G).

Tabela 5 - Mensuração do IPI em uma *pivot table* com disposição anual, mês a mês da publicação contábil

Ano / Mês	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	324.774,33	464.345,80	179.210,76	NaN	NaN	NaN	458.609,09	388.318,24	457.341,78	526.239,00	482.262,65	687.264,52
2	287.851,66	372.337,58	380.934,87	692.316,35	623.983,05	NaN	355.391,64	424.136,95	468.640,16	569.535,89	521.199,64	558.950,13
3	255.448,07	312.539,63	336.757,58	263.070,23	278.861,68	NaN	346.974,25	335.542,95	335.294,54	NaN	480.991,45	663.711,26
4	270.447,90	389.353,56	352.702,86	308.950,36	304.074,38	NaN	356.024,87	402.287,66	472.876,63	518.565,73	161.614,16	974.688,23
5	308.042,22	330.190,24	404.440,25	343.625,39	330.730,93	796.392,39	404.294,90	380.161,85	425.299,56	546.038,09	67.235,40	515.333,88
6	295.901,57	365.029,65	361.310,65	357.895,08	315.003,70	252.060,52	240.573,82	382.228,69	389.897,85	572.639,19	331.041,36	718.284,99
7	326.064,00	348.867,32	336.936,02	352.535,58	323.606,89	224.504,65	326.872,95	423.729,09	470.170,32	579.695,77	464.941,94	681.692,47
8	365.904,76	359.982,33	372.420,14	377.215,40	333.567,14	238.883,42	330.302,02	3.035.996,80	425.064,95	518.832,80	537.712,98	573.245,53
9	295.425,82	333.713,02	376.478,63	289.678,74	313.903,97	91.632,94	361.700,51	429.175,71	425.004,51	632.470,13	631.608,48	682.070,39
10	333.801,52	465.084,63	364.017,53	420.251,84	336.164,79	287.326,12	377.501,27	-2.092.561,35	431.376,09	559.728,86	644.985,83	716.718,83
11	371.729,34	442.117,79	368.524,21	406.881,35	386.017,34	265.540,18	368.748,69	438.002,00	414.146,66	205.501,62	870.233,46	672.568,80
12	414.046,87	500.156,26	368.524,21	434.868,38	380.886,83	262.204,70	455.847,10	604.315,53	499.031,95	690.900,60	648.137,50	529.333,59
Total	3.849.438,06	4.683.717,81	4.202.257,71	4.247.288,70	3.926.800,70	2.418.544,92	4.382.841,11	5.151.334,12	5.214.145,00	5.920.147,68	5.841.964,85	7.973.862,62

Fonte: Elaborado pela Autora (2023) com os dados do Balancete da Receita (SÃO LUÍS-MA, 2021b).

Para dirimir a questão de ausência dos dados não foram utilizadas técnicas complexas (como imputação de dados), mas apenas pesquisas nas bases de dados originais. Isto, pelo fato deste estudo possuir uma legislação que prescreve suas regras, como exemplo dos índices econômicos, os quais, no período entre publicações, o valor vigente é o índice anterior. Logo, na ausência em um período, basta repetir o anterior. E no caso das receitas, tendo como arquétipo a Tabela 5, todas as receitas dos períodos com ausência de dados e *outlier* foram reavaliadas, um reajuste do *Dataset*, já que se adotaram os valores dos repasses que deveriam ter ocorrido no mês de sua publicação. Neste sentido, a etapa de “Formatar dados” foi necessária a reindexação do conjunto e os dados incluídos foram formatados à proporção que estavam sendo alterados e/ou incluídos. Após os ajustes o *dataset* ficou com 906 observações distribuídas conforme listado na Tabela 6.

Tabela 6 - Distribuição das 906 observações por receita após ajustes da fase de pré-processamento

Receita	Exercícios financeiros	Qnt. Observações	% do Total
FPM	2010 a 2021	144	15,89%
FEP	2010 a 2021	144	15,89%
ICMS	2010 a 2021	144	15,89%
IPVA	2010 a 2021	144	15,89%
CFEM	2010 a 2021	144	15,89%
IPI	2010 a 2021	144	15,89%
CFEM AFETADOS	2018 a 2021	42	4,64%
Total		906	100,00%

Fonte: Elaborado pela Autora (2023), adaptado do APÊNDICE E.

Considerando o plano de projeto inicial (Figura 3), esta atividade visou responder a questão de pesquisa “Q01”, bem como o atendimento de 3 objetivos específicos, outrora delineados, na primeira iteração. Como o intuito perpassa pela identificação de um modelo causal²¹ (ACKERMANN e SELLITTO, 2022), foram escolhidos e testados modelos de AM

²¹ utilizado na Administração para previsão de demandas, por ser um método matemático que permite controlar o erro, contam com variáveis dependentes cuja relação “causa-efeito” vinculam-se às variáveis antecessoras, as quais recebem as alcunhas de “variáveis independentes ou causais” (ACKERMANN e SELLITTO, 2022, p. 87)).

para seleção das variáveis com abordagem embutida. Assim, os três *Ensembles*: *RandomForestRegressor*, *GradientBoostingRegressor* e *XGBRegressor*, os dois primeiros do módulo *sklearn.ensemble* (PEDREGOSA, VAROQUAUX, *et al.*, 2011) e o terceiro do *XGBoost* (XGBOOST DOCUMENTATION, 2021; WADE, 2020), foram aplicados individualmente para ulterior comparação dos seus resultados com o modelo combinados. O *Dataset* foi separado em subconjuntos dos rótulos da variável “RECEITA”, treinando seus dados em um laço, devidamente escalonados e normalizados, dividindo-os em treino e teste, conforme a Tabela 7, sendo armazenadas as *features* que alcançassem importância superior a 0,005. Utilizou-se a função *train_test_split* do módulo *sklearn.model_selection*, com argumento '*shuffle*' como '*False*' para não misturar os dados, pois trata-se de série temporal cuja dependência da ordem das observações podem interferir nas tendências e/ou ciclos dela (GÉRON, 2021, p. 71; CHOLLET, 2021, p. 146; KORSTANJE, 2021, p. 31; RAFFERTY, 2021, p. 17).

Tabela 7 - Quantidade de observações por subconjunto das receitas e divisão nos métodos

Receita	Observações	Divisões dos conjuntos de treino			
		<i>Random Forest</i>	<i>XGBoost</i>	<i>Gradient Boosting</i>	<i>Modelo Combinado</i>
CFEM AFETADOS	42	70%	70%	70%	70%
CFEM, FPM, FEP, ICMS, IPI e IPVA	144*6				
TOTAL	906				

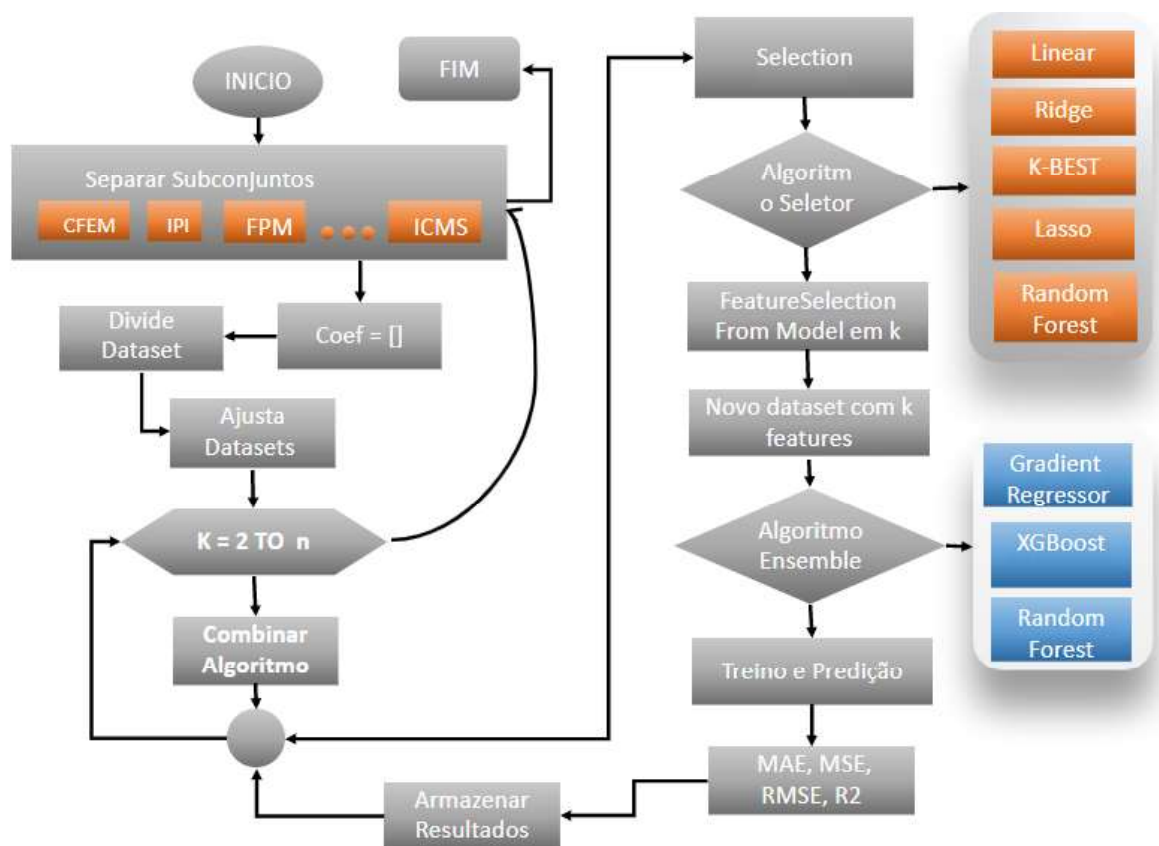
Fonte: Elaborado pela Autora (2023).

Já o modelo combinado (Figura 4), utilizado para comparação com os anteriores, teve como proposta a adaptação do modelo de seleção de *feature importance* de Pan He e outros autores (2014) para as previsões de anuncios do Facebook, cujos atributos são submetidos a um laço crescente, iniciando a partir de dois atributos e incrementando o conjunto sequencialmente, conforme esquema da Figura 4 (HE, PAN, *et al.*, 2014). Deste modo, o ponto de partida foi do tipo *outward* (BARANAUSKAS e MONARD, 1999). Com uma abordagem híbrida, o primeiro laço utilizou uma estratégia de filtro com a técnica de ranqueamento de atributos, com a intenção de avaliar a quantidade fixa de *features*, no algoritmo denominado por ‘*k*’ e o critério de parada foi exaustivo (ALMEIDA, 2018, p. 38; BRUCE e BRUCE, 2019, p. 190).

Dentro do segundo laço de “*k*” atributos (Figura 4), para atingir o objetivo, utilizou-se a combinação de cinco algoritmos, denominados de seletores: *KBest*, *LinearRegression*, *Ridge*, *Lasso* e *RandomForestRegressor* (PEDREGOSA, VAROQUAUX, *et al.*, 2007). Tal combinação fez uma permutação com os algoritmos alcunhados de *ensembles*: *RandomForestRegressor*, *XGBRegressor* e *GradientBoostingRegressor*. Ao realizar o

treinamento, naturalmente as *features* ganham coeficientes que já são elegíveis para classificar a importância de cada uma delas. Aproveitando o ensejo, extrai-se do modelo as melhores *features* com auxílio do pacote *Feature Selection* da biblioteca *sklearn* (PEDREGOSA, VAROQUAUX, *et al.*, 2007). No afã de evidenciar a eficiência na escolha do número de atributos, adicionou-se ao código múltiplos modelos em combinações cruzadas com quantidades diferentes de *features* (Figura 4). 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.

Figura 4 - Esquema do algoritmo do modelo combinado



Fonte: Elaborado pela Autora (2023).

A intenção da combinação de modelos da Figura 4 não foi obter o menor valor da métrica, mas o que conseguir ter uma estabilidade de menores erros em cada subconjunto dos rótulos da “RECEITA” (HE, PAN, *et al.*, 2014, p. 7; DUBOUE, 2020, p. 13). Diferente do método referenciado²², foram estipulados outros critérios pautados na verificação das seguintes etapas, com o mesmo procedimento de subconjunto por rótulo:

²² Pan He e outros autores (2014) tinham como proposta algoritmos de classificação e este estudo trata de regressão por isso adaptações foram feitas, mescladas com os critérios da listagem.

- 1) Separam-se as vinte menores métricas pelo *Root Mean Squared Error*, em português raiz quadrática média (RMSE) e o *R-square*, traduzido por R-quadrado (R^2) (BRUCE e BRUCE, 2019, p. 190);
- 2) Plota-se um gráfico de todos os resultados excluindo-se os setores que tiverem maiores variância ou *outliers* extremos;
- 3) Retira-se os *outliers* com o Método de *Tukey*, no intuito de obter uma região que possua uma certa estabilidade, ou seja, que não esteja vizinha de *outliers*;
- 4) Escolhe o menor RMSE, mediano das quatro melhores *features* escolhidas;
- 5) Extrai-se a lista dos atributos escolhidos para aplicar nos modelos de treinamento.

4.4 QUARTA FASE DE MODELAGEM

Nesta primeira iteração, registra-se que a técnica de modelagem selecionada foi de regressão em séries temporais multivariadas e univariadas. Por se tratar de séries temporais multivariadas, os algoritmos de LSTM e GRU foram testado com o método de janela deslizante (*sliding window*, tradução livre), utilizado por vários autores ((LAKSHMANAN, ROBINSON e MUNN, 2021; PAVLYUK, 2020; NIELSEN, 2020; BROWNLEE, 2020; 2019; 2017).

Segundo Jason Brownlee (2020, p. 16), janela deslizante “(...) em estatísticas e análise de séries temporais, isso é chamado de método *lag* ou *lag*. O número de passos de tempo anteriores é chamado de largura da janela ou tamanho do atraso....”(tradução nossa)²³. Como o foco é trabalhar com séries multivariadas, Tsay (2014, p. 1) ensina que

A análise de séries temporais multivariadas considera simultaneamente múltiplas séries temporais. É um ramo da análise estatística multivariada, mas lida especificamente com dados dependentes. É, em geral, muito mais complicada do que a análise de séries temporais univariadas, especialmente quando o número de séries consideradas é grande ... (TSAY, 2014, p. 1) (tradução nossa)²⁴.

Quanto a etapa de Geração do Projeto de Teste dividiu-se os conjuntos de dados na proporção de 11 anos (132 meses) para treino e 1 ano (12 meses) para teste, isto porque cada subconjunto já possui poucas observações (ver Tabela 6), mesmo esperado um aumento de 7 vezes, continua pequeno (GÉRON, 2021, p. 26; KORSTANJE, 2021, p. 30), utilizando a função *train_test_split* do módulo *sklearn.model_selection*, com argumento '*shuffle*' como '*False*' para não misturar os dados, pois trata-se de série temporal cuja dependência da ordem das

²³ Original: “(...) In statistics and time series analysis, this is called a lag or lag method. The number of previous time steps is called the window width or size of the lag....”.

²⁴ Original: “Multivariate time series analysis considers simultaneously multiple time series. It is a branch of multivariate statistical analysis but deals specifically with dependent data. It is, in general, much more complicated than the univariate time series analysis, especially when the number of series considered is large”

observações podem interferir nas tendências e/ou ciclos dela (GÉRON, 2021, p. 71; CHOLLET, 2021, p. 146; KORSTANJE, 2021, p. 31; RAFFERTY, 2021, p. 17).

Na etapa de “Modelo de construção”, a intensão era seguir as orientações de Plotnikova, Dumas e Milani, (2021, p. 8), quanto a pretensão de executar uma multimodelagem, para apresentar um conjunto de resultados de vários modelos e horizontes temporais, mas não foi possível para esta primeira iteração. Assim, empregou-se os modelos mais simples para obter algum resultado inicial, sem ajustes de parâmetros neste momento.

Neste sentido, elencam-se no Quadro 5 os modelos que inicialmente foram escolhidos, dentre os métodos estatísticos e os de AM que deveriam ter sido utilizados e os sugeridos por Carmo, Boldt e Komati (2019), pode-se cogitar a utilização de combinações híbridas em outras iterações. Como nem todos foram implementados, a produção de uma *baseline*²⁵ para subsidiar as análises entre as Capitais ficou comprometida, principalmente que pelo fato dos algoritmos de aprendizagem profunda exigirem mais observações (GÉRON, 2021). Importante registrar que o modelo da SOF, equação [1] (p. 31), faria parte deste baseline, utilizando média móveis.

Quadro 5 – Listagem dos Modelos que foram planejados e executados na primeira iteração

Não paramétricos (AM)	Planejado	Executado	Paramétricos (Estatísticos)	Planejado	Executado
XGBoost	✓	✗	Naives	✓	✓
Randon Forest	✓	✓	AR	✓	✗
KNN	✓	✗	MA	✓	✗
Gradient Boosting	✓	✗	ARMA	✓	✗
Redes Neurais	✓	✗	ARIMA	✓	✗
RNN	✓	✗	SARIMA	✓	✗
GRU	✓	✓	SARIMAX	✓	✗
LSTM	✓	✓	VAR	✓	✗
CNN	✗	✓	VARMA	✓	✗

Fonte: Elaborado pela Autora (2023).

A última etapa desta fase é a “Avaliação do modelo” cujo objetivo foi a obtenção dos melhores resultados dentre os experimentos provenientes dos modelos do Quadro 5 que foram executados. Buscou identifica a satisfação do critério de melhor de desempenho no teste com a projeção fornecida pelos modelos, no caso de São Luís, seria um modelo com máximo de 5% de divergência entre a projeção e o executado, mensurados pelas métricas de avaliação (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 8).

²⁵ “um modelo usado como ponto de referência para comparar o desempenho de outro modelo (normalmente mais complexo). Por exemplo, um modelo de regressão logística pode servir como uma boa linha de base para um modelo de Deep Learning. Para um problema específico, a linha de base ajuda os Cientistas de Dados a quantificar o desempenho mínimo esperado que um novo modelo deve alcançar para que o novo modelo seja útil.” (DATA SCIENCE ACADEMY, 2022)).

4.5 QUINTA FASE DE AVALIAÇÃO

Plotnikova, Dumas e Milani (2021) enfatizam que a fase de “Avaliar resultados” almeja pela confirmação do atendimento do(s) objetivo(s) do negócio por intermédio do modelo construído. Aproveitando o ensejo, os autores protestaram contra o CRISP-DM por não apresenta um roteiro de como deve ser esta validação de qualidade em cenários reais. Todavia os pesquisadores, assim como Duboue (2020, p. 12), definem duas etapas de validação simultâneas: uma técnica, ocorrendo na fase de Modelagem, e a outra, para esta fase, verificando a significância do modelo no ambiente de negócios (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 8).

Neste sentido, para a primeira avaliação, das comparações entre os modelos do Quadro 5 e as combinações de conjuntos de dados serão utilizadas as métricas R^2 , erro absoluto médio, do inglês *Mean Absolut Error* (MAE), erro quadrático médio, no original *Mean Squared Error* (MSE) e RMSE do módulo do *sklear.metrics* (PEDREGOSA, VAROQUAUX, *et al.*, 2011b).

Sendo conveniente enfatizar que o R^2 calcula a porcentagem da variância que os dados do modelo podem prever ou explicar a variável alvo, geralmente encontram-se no intervalo de 0 e 1 (HARRISON, 2020, p. 209; BRUCE e BRUCE, 2019, p. 186). É importante registrar que, em modelos de regressão com múltiplas variáveis

O R^2 aumenta quando uma nova variável é adicionada ao modelo, entretanto o R^2 ajustado nem sempre aumentará, bem como **poderá diminuir ou até ficar negativo**. Para este último caso, Stock e Watson (2004) explicam que o R^2 ajustado **pode ficar negativo** quando as variáveis explicativas, tomadas em conjunto, reduzirem a soma dos quadrados dos resíduos em um montante tão pequeno que esta redução não consiga compensar o fator $(n-1)/(n-k)$ (FÁVERO e BELFIORE, 2017, p. 540) (grifo nosso).

O MAE é o cálculo da média das distâncias entre os valores originais e previstos pelo modelo, “[...] como está em unidades do alvo, ele é mais interpretável...” (HARRISON, 2020, p. 210), quanto menor melhor (HYNDMAN e ATHANASOPOULOS, 2018). Semelhantemente, o 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 avaliação do modelo, sendo que erros maiores são penalizados por calcular a raiz quadrada ao final (HARRISON, 2020, p. 210). Quando há diferenças de unidade do sinal original (DUBOUE, 2020, p. 12; HYNDMAN e ATHANASOPOULOS, 2018) indica-se o RMSE, principalmente para comparar modelos de regressão (BRUCE e BRUCE, 2019, p. 185).

Pesquisadores indicam MAE e RMSE para comparar modelos que concedam uma avaliação de previsão concorrente (PETROPOULOS, APILETTI, *et al.*, 2022, p. 726).

Portanto, escolhem-se mais de uma métrica por reconhecer que os impactos dos erros não são iguais (HARRISON, 2020, p. 210), ademais, alguns autores sugerem que não se deve “[...] confiar em apenas uma medida, mas ver os resultados de diferentes ângulos.”²⁶ (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 8)). Caso haja ocorrência de todas as métricas se igualemente, Bruce e Bruce (2019, p.188) sugerem utilizar o princípio da “navalha de Occam”²⁷.

Para a segunda validação, da significância no ambiente de negócios, a qual foi alcunhada por Duboue (2020, p. 12), em problemas de regressão, como métrica de utilidade, aplicou-se o teste estatístico não paramétrico de *Friedman*, uma vez que ele não depende de distribuição normal, muito menos de homogeneidade de variância (DEMˆSAR, 2006, p. 27) fazendo vários *ranks* para cada métrica com nível de significância de 0,05, Caso a hipótese nula (não haja diferença significativa nos *ranks*) seja rejeitada, seguir-se-á a indicação de Parmezan, Souza e Batista (2019, p. 323) utilizando teste *post hoc* de *Nemenyi* para comparar os resultados (DEMˆSAR, 2006, p. 11). A execução utilizará o pacote *scikit-posthocs* para *Python* (TERPILOWSKI, 2021), apresentando os resultados de forma tabular e com diagramas de diferença crítica (DEMˆSAR, 2006, p. 15; PARMEZAN, SOUZA e BATISTA, 2019, p. 323).

Por fim, as métricas discutidas antes tratam de médias e tentam resumir o comportamento do modelo em seu caso mais representativo. No entanto, isso não aborda a variância dos resultados. Este tópico é bem estudado no ML, onde falamos sobre o viés (aprender a coisa errada; erros devido às limitações do modelo) versus variância (aprender pontos dispersos; erros devido a amostragem finita de dados, pois amostras diferentes produzem modelos diferentes) (DOMINGOS, 2012 *apud* DUBOUE, 2020, p. 12).²⁸

A proposta da etapa de “Processo de revisão”, em que se pretendia aplicar outro questionário às 8 Capitais com os membros do GT-8 da ABRASF com a finalidade de avaliar os resultados obtidos, principalmente o percentual que cada Município adota para evitar a publicação de um Decreto de Contingenciamento de despesas. Todavia, como não houve agenda compatível com os técnicos do GT-08 da ABRASF para apresentação dos resultados e verificação das métricas específicas para cada Capital, bem como solicitar a revisão das solicitações dos protocolos negados, aguarda-se a divulgação e disponibilização oficial pelos Municípios, conforme protocolos do Quadro 3.

²⁶Tradução livre, original: “[...] rely on only one measure but can see the results from different angles.” (PLOTNIKOVA, DUMAS e MILANI, 2021, p. 8)).

²⁷ Princípio que dar preferência aos modelos mais simples para mais detalhe consultar Wazlawick (2021, p. 39).

²⁸ Tradução livre, original: “Finally, the metrics discussed before all deal with averages and try to summarize the behaviour of the model on its most representative case. Nevertheless, that does not address the **variance** of the results. This topic is well studied in ML, where we talk about the **bias** (learning the wrong thing; errors due to limitations of the model) versus **variance** (learning scattered points; errors due to finite data sampling and different samples produce different models) (DOMINGOS, 2012 *apud* (DUBOUE, 2020, p. 12)).

No que tange a etapa de “Determinar os próximos passos” ficou vinculada ao princípio denominado de “Exatidão ou Realismo Orçamentário” o qual versa que as estimativas deverão ser consistentes, tão exatas quanto possível, respaldando-se nos arts. 7º, 16, 17 e 18 do Decreto-Lei nº 200/67 (BRASIL, 1967; GONTIJO, GREGGIANIN e MENDES, s.d.). Corroborando para este entendimento combina-se com o princípio do “Equilíbrio” o qual deve ser compreendido em termos fiscais e contábeis, veiculando que não haja despesa sem receita para suportá-la, respaldado nos arts. 165 e 167 da CF (BRASIL, 1988). Em suma, o que importa é que a divergência entre orçado e executado seja mínima e ainda se faz imperativa a execução de novas iterações para avaliação dos demais modelos propostos.

4.6 SEXTA FASE DE IMPLANTAÇÃO

Para a fase de implantação, Plotnikova, Dumas e Milani (2021, p. 9-10) sugeriram a busca de indicadores de desempenho e visualizações mais simples com intuito de tornar o resultado compreensível para os gestores e técnicos. Assim, o Quadro 6 resume a etapa de “Planejar a implantação”, elabora-se um plano de preparação de implantação, sendo resumido como segue:

Quadro 6 - Resumo comparativo da execução do plano inicial de implantação

Plano inicial de implantação	Planejado	Executado	Observações
Utilizar um dashboard do Power BI Desktop para implementar os resultados;	✓	✗	Ainda pretende-se criar o painel, pois é uso fácil para os técnicos;
Modelos disponibilizados por meio de uma API;	✓	✗✓	Os modelos não foram implementados, somente a Seleção de Features;
Resultados dos erros das previsões tabulados por horizonte temporal, por método e modelo, com filtros;	✓	✗✓	A ser implementado em outra iteração, a API apenas tem os dados de São Luís;
Ranquear, por percentual de importância, as variáveis escolhidas em um gráfico de barras;	✓	✓	As receitas dos dados de São Luís foram concluídas;
Habilitação de exportação dos resultados em cada um dos objetos do painel (tabelas e gráficos) proporcionando a liberdade de manipulação para os usuários;	✓	✗✓	O painel não ficou pronto, mas a API já responde em arquivos "Json" o que esta implantado da seleção de features;
Incluir cartões com valores de previsão anual e diferenças entre previsto e executado.	✓	✗	Para outra iteração.

Fonte: Elaborado pela Autora (2023).

Quanto às tarefas de “Plano de Monitoramento e Manutenção”, “Produzir relatório final” e “Revisar Projeto, nenhum dos artefatos foi produzido, pois foi planejado com a demonstração do *Dashboard* em funcionamento, em uma das reuniões do GT-08 da ABRASF, em que se aproveitaria o ensejo para aplicar um questionário de avaliação dos resultados obtidos

para cada Capital.

Figura 5 - Painel de monitoramento de ingresso de recursos.



Fonte: Elaborado pela Autora (2023).

Contudo, ainda no período da qualificação, foi desenvolvido o painel da Figura 5 devido às ausências dos valores das receitas e inconsistência entre os totais publicados e os que ingressaram (ver APÊNDICE F e APÊNDICE G). O intuito foi de agilizar o monitoramento das publicações de repasses de transferências constitucionais (RTCs na Figura 5) com seus ingressos no Agente Financeiro, assim o painel da Figura 5 possui:

- 1) Na faixa lateral esquerda, filtros de receita, de ano e mês financeiros;
- 2) No retângulo superior, um título para identificar quais filtros foram selecionados;
- 3) Abaixo do anterior, três cartões de resumo, dos valores depositados e os publicados, as diferenças totais de um ano com o seu antecessor, a do último mês e as quantidades de depósitos e sua previsão, seguido do percentual faltante;
- 4) Ao lado dos filtros foi incluído um gráfico com os rótulos dos valores comparando os RTCs com o ingresso do recurso, mês a mês;
- 5) E as outras tabelas listam os recursos por data e por semana.

5 RESULTADOS INICIAIS OBTIDOS COM OS DADOS DE SÃO LUIS

Neste capítulo foram reunidos os resultados obtidos na primeira iteração, somente com os dados de São Luís, advindos da avaliação dos testes na escolha da seleção de variáveis, tabulando os resultados obtidos e as *features* escolhidas pela menor métrica, os resultados da modelagem com algoritmos de aprendizagem profunda e parte da implementação da API.

5.1 PRIMEIRA ITEAÇÃO DA SELEÇÃO DE VARIÁVEIS

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 atenderiam aos efeitos preço e quantidade (BRASIL, 2000). Esta amostra utilizou os dados de janeiro de 2010 a dezembro de 2021, com frequência mensal; valores nominais, testou-se os montantes em seus depósitos por competência das cotas oficiais e retirando-se todos os repasses extraordinários²⁹, logo sem os principais causadores de *outliers*.

Quadro 7 - Resultados das métricas da seleção das variáveis da Receita da CFEM de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	2	0,141	0,093	0,020	0,957
<i>RandomForest</i>	3	0,211	0,154	0,045	0,904
<i>XGBoost</i>	5	0,223	0,161	0,050	0,894
Mod. Combinado					
<i>KBest-GradientBoosting</i>	38	0,131	0,086	0,017	0,963

Featuares do Modelo Combinado KBest-GradientBoosting:							
Ajustes_MA_Munic	4,99	CFEM_Mun	0,59	SINAP_Baixo	0,13	VA_BrtTotal	0,03
IPVA	4,97	ST_EXTERNO	0,47	SINAP_Minimo	0,12	PIB_Mun	0,02
EXPT_SLZ_ProdIndQuim	3,71	ICMS_TOTAL	0,38	VA_BrtServExclAdm	0,12	RndPrCap_Mun	0,02
ICMS_SIMP_NACIO	3,4	SalarioMinimo	0,23	SINAP_Normal	0,11	Coef_IPM	0,01
Importacao	3,36	ResvInternac	0,17	Prod_Minerais	0,11	CFEM_MA	0,01
TxIPCA	2,44	PIB_MenBR	0,15	SINAP_Alto	0,1	Exportacao	0
IPP	1,84	TOTAL_ArrecMun	0,15	MetaisCcomuns	0,06	VA_ImpostosProd_	0
Gasolina_Brasil	1,47	ICMS_Muns	0,15	x_3	0,06	VA_BrtIndust	0
EXPT_SLZ_TOTAL	1,08	VA_BrtAdm	0,15	PercResSP_Nom	0,06		
Tx_SELIC_Perd	0,8	DolarComercial	0,13	DvLq Gov MA	0,04		

Fonte: Elaborado pela Autora (2023).

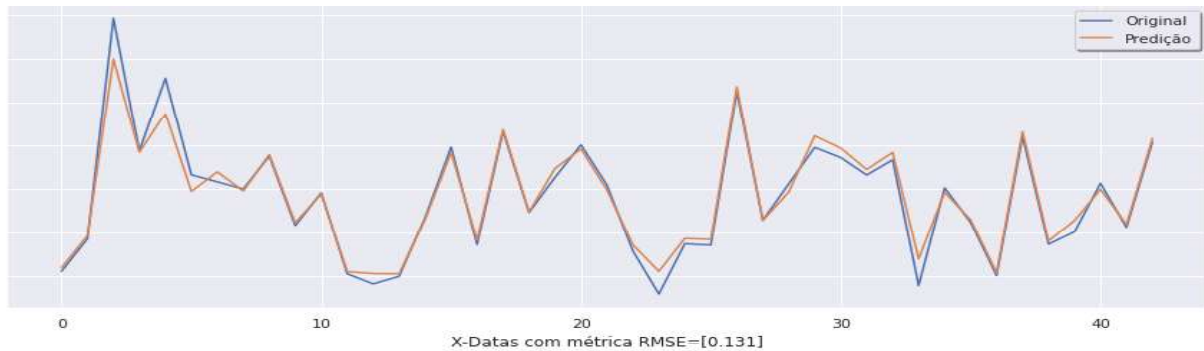
No Quadro 7 foi organizado o *ranking* das métricas da Receita da CFEM, tendo, em 1º lugar o Modelo Combinado com o seletor *KBest* aplicando as variáveis ao *GradientBoosting*, com configuração padrão, diferindo apenas na função de perda, '*huber*³⁰', e o *random_state* que foi zero, o qual escolheu 38 (ver Quadro 7), com RMSE de 0,131, proporcionando um resultado

²⁹ Situação já explicada no item 4.2.

³⁰ Permite uma combinação do 'squared_error', 'absolute_error' (PEDREGOSA, VAROQUAUX, *et al.*, 2011)).

satisfatório no conjunto de teste, conforme comparativo do Gráfico 4.

Gráfico 4 - Comparativo dos valores das Predições x Originais da Receita da CFEM nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

Diferente da amostra com repasses extraordinários, testada inicialmente, especificando 3 variáveis (Quadro 8), obtendo um RMSE de 0,199, também no Modelo Combinado, mas com seletor *Ridge*, identificou-se apenas variáveis de efeito quantidade (BRASIL, 2000), conforme Quadro 8. Ao aplicar o teste de *Friedman*³¹ nos resultados dos Quadro 7 e Quadro 8, a hipótese nula não foi rejeitada, logo, as métricas não possuem diferenças significativas entre os algoritmos (DEM'SAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 8 - Resultados das métricas da seleção das variáveis da Receita da CFEM com repasses extraordinários

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
<i>Ensemble</i>					
<i>GradientBoosting</i>	4	0,414	0,244	0,171	0,792
<i>RandomForest</i>	1	0,447	0,239	0,200	0,758
<i>XGBoost</i>	16	4,612	2,500	21,267	0,000
Mod. Combinado					
<i>Ridge-GradientBoosting</i>	3	0,199	0,145	0,040	0,946

Featuress do Modelo Combinado Ridge-GradientBoosting:			
CFEM_Mun	90,39	ICMS	21,15
Coef_CIFPM_SLZ	-14,95		

Fonte: Elaborado pela Autora (2023).

Dentre as posições do *ranking* do Quadro 7, registrou-se tanto atributos relacionados aos efeitos quantidade quanto preço (BRASIL, 2000). Sendo importante comentar que o efeito quantidade foi identificado pela 1ª variável dos dados com repasses extraordinários (Quadro 8), a “CFEM_Mun”, e nos dados das cotas oficiais ela ficou em 10ª posição (Quadro 7). Lembrando que a variável em questão corresponde a 60% do montante repassado ao Município Produtor, conforme prescreve a legislação vigente³² (BRASIL, 2017). As demais variáveis do conjunto seguem a cadeia produtiva, pois esta receita movimentada produtos e serviços e é comercializada em dólar, além de ser importada e exportada com produtos químicos e com forte influência na

³¹ Usando o as bibliotecas *scipy* e *scikit_posthocs*, com resultado da estatística = 6.000 e *p-value* = 0,112.

³² Inciso “VI”, do Art. 2º, da Lei 13.540/2017 (BRASIL, 2017).

indústria.

Quadro 9 - Resultados das métricas da seleção das variáveis da Receita do ICMS de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	11	0,302	0,241	0,091	0,829
<i>RandomForest</i>	10	0,345	0,279	0,119	0,777
<i>XGBoost</i>	17	0,303	0,227	0,092	0,828
Mod. Combinado					
<i>KBest-GradientBoosting</i>	35	0,255	0,198	0,065	0,879

Featuers do Modelo Combinado KBest-GradientBoosting:			
CFEM_Total	Coef_CIFPM_SLZ	VA_BrtAdm	SalarioMinimo
ICMS_DvAtTrb	Prod_Minerais	DolarComercial	ICMS_SIMP_NACIO
ICMS_Muns	VA_BrtIndust	Coef_IPM	DvLq_Gov_MA
TOTAL_ArrecMun	TxDesocupados	Gasolina_Brasil	PercResSP_Nom
IGPM	RndPrCap_Mun	SINAP_Minimo	ICMS
DvLq_Sao_Luis_MA	VA_ImpostosProd_	SINAP_Normal	PIB_MenBR
x_1	PIB_Mun	SINAP_Baixo	ST_EXTERNO
MetaisComuns	VA_BrtTotal	SINAP_Alto	ICMS_TOTAL
EXPT_SLZ_ProdIndQuim	VA_BrtServExclAdm	ResvInternac	

Fonte: Elaborado pela Autora (2023).

Na amostra do ICMS, novamente o Modelo Combinado apresentou o melhor desempenho, tanto nos dados oficiais (Quadro 9) quanto nos extraordinários (Quadro 10), escolhendo 35 e 47 variáveis, respectivamente, ambos com o seletor *KBest* aplicado ao *GradientBoosting*³³ e ao *XGBoost*³⁴, respectivamente. As métricas do RMSE para o experimento das cotas oficiais foi de 0,255 (Quadro 9) e da inclusão dos repasses extraordinários foi de 0,211 (Quadro 10), o que nominalmente o último seria menor, mas na prática a indicação é utilizar somente valores oficiais para projeção.

Gráfico 5 - Comparativo dos valores das Predições x Originais da Receita do ICMS nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

³³ Treinado com as mesmas configurações da receita da CFEM, pois foi padronizado para todas as receitas.

³⁴ Treinado com as configurações padronizadas, diferindo no objetivo configurado para "*reg:squarederror*", indicando uma regressão linear, a taxa de aprendizado ficou em 0,1, o "*max_depth*" foi 10, que indica a profundidade dos nós da árvore, conjugado com "*min_child_weight*" de 5, que controla a criação de nós filhos com peso mínimo, "*subsample*" de 0,6, controla a quantidade de observações nas etapas e "*seed*" foi 42 (WADE, 2020)).

Do resultado dos dados de cotas oficiais (Quadro 9), o comparativo dos valores preditos com os executados da receita do ICMS foi promissor, mas ao final aumentou o erro, conforme Gráfico 5. Aplicando-se o teste de *Friedman*³⁵, a hipótese nula também não foi rejeitada, logo, as métricas não possuem diferenças significativas entre os algoritmos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 10 - Resultados das métricas da seleção das variáveis da Receita do ICMS com repasses extraordinários

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	10	0,488	0,386	0,238	0,489
<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

Featuare do Modelo Combinado KBest -XGBoost:			
CFEM_MA	ICMS_Muns	Gasolina_BR	SINAP_Alto
IPVA_Muns	TOTAL_ArrecMun	ResvInternac	VA_BrtTotal
ICMS_DvAtTrb	Coef_CIFPM_SLZ	Importacao	ICMS
CFEM_Total	DvLq_Gov_MA	PercResSP_Prim	VA_BrtServExclAdm
VA_BrtAgro	DvLq_Sao_Luis_MA	VA_BrtIndust	SINAP_Normal
EXPT_SLZ_TOTAL	DvLq_MA_Munic	VA_ImpostosProd	SINAP_Minimo
IPP	TxDesocupados	Exportacao	VA_BrtAdm
x_2	IBC_Br	SalarioMinimo	SINAP_Baixo
Veic_PgtoIPVA	PerPopulacaoUF	POPULACAO	PIB_MenBR
PIM_PF	PercResSP_Nom	PIB_Mun	ST_EXTERNO
x_1	DolarComercial	ICMS_SimpNacio	ICMS_TOTAL
EXPT_SLZ_ProdIndQuim	Coef_IPM	RndPrCap_Mun	

Fonte: Elaborado pela Autora (2023).

No que tange às *variáveis* de efeito preço e quantidade (BRASIL, 2000) foram reconhecidas nos dois *datasets*, as de preço foram: “*IGPM*”, “*IPP*” e “*IBC_Br*”. Ambos, os modelos, selecionaram o “*ICMS_TOTAL*”, que é a mais relevantes³⁶ e o “*Coef_IPM*”, pois determinam a forma de cálculo desta transferência.

Quadro 11 - Resultados das métricas da seleção das variáveis da Receita do IPI com repasses extraordinários

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	30	1,272	0,982	1,619	0,000
<i>RandomForest</i>	8	1,122	0,874	1,260	0,000
<i>XGBoost</i>	42	1,099	0,793	1,208	0,000
Mod. Combinado					
<i>Lasso-GradientBoosting</i>	7	0,872	0,674	0,760	0,286

Featuare do Modelo Combinado Lasso-GradientBoosting:			
x_3	x_1	ICMS	ICMS_TOTAL
x_2	DolarComercial	ICMS_DvAtTrb	

Fonte: Elaborado pela Autora (2023).

³⁵ Apresentando como resultado da estatística = 5.400 e *p-value* = 0,145.

³⁶ Inciso "IV", do art. 158, da Constituição Federal (BRASIL, 1988st).

Considerando o ranking do Modelo Combinado, com seletor *Lasso* e *GradientBoosting*, para a receita de IPI, com repasses extraordinários, no Quadro 11, obteve-se o RMSE de 0,872 referente à escolha de 7 variáveis. Já com os dados das cotas oficiais, elencados no Quadro 12, o modelo combinado *Ridge* e *XGBoost* obteve o menor RMSE, de 0,462, e MSE, de 0,214, mas com 55 variáveis. Neste último modelo, as *features* identificadas são tanto de quantidade quanto de preços (“INPC”, “TxIPCA”, “IPP”, “IGPM”, “IBC_Br”, e “Tx_SELIC_Perd”) o que as dos dados com repasse extraordinário não identificaram, pois somente apresentaram efeito quantidade (BRASIL, 2000).

Quadro 12 - Resultados das métricas da seleção das variáveis da Receita do IPI de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	30	0,613	0,504	0,376	0,000
<i>RandomForest</i>	22	1,048	0,807	1,098	0,000
<i>XGBoost</i>	36	0,547	0,474	0,300	0,177
Mod. Combinado					
<i>Ridge-XGBoost</i>	55	0,462	0,397	0,214	0,413

Features do Mod. Combinado Ridge-XGBoost:							
VA_BrtAgro	50,62	INPC	0,32	CFEM_Mun	3,72	SINAP_Alto	-8,63
x_2	49,62	ICMS_SIMP_NACIO	8,89	IGPM	4,02	CFEM_Total	-9,51
Coef_CIFPM_SLZ	31,88	ICMS_TOTAL	8,82	TxDesocupados	4,08	PIB_Mun	-9,6
ST_EXTERNO	25,55	ICMS_DvAtTrb	8,34	Veic_PgtoIPVA	4,22	DolarComercial	-10,46
Exportacao	23,63	FinMA_MnRestPrim	7,24	MetaisCcomuns	4,4	RndPrCap_Mun	-10,83
CFEM_MA	18,47	FinMA_MnRestNom	7	IBC_Br	4,55	Ajustes_MA_Munic	-10,24
DvLq_Sao_Luis_MA	18,37	TxIPCA	6,7	Petroleo_Brent_Brasil	4,6	VA_BrtAdm	-10,64
Prod_Minerais	18,1	DvLq_Gov_MA	6,76	ResvInternac	5,51	ST_INTERNO	-10,2
DvLq_MA_ExcCap	17,54	TransacoesEspeciais	6,04	VA_BrtTotal	5,01	EXPT_SLZ_ProdIndQ	-10,22
Coef_IPM	15,76	EXPT_SLZ_TOTAL	5,72	ICMS_DvAtTrb_Muns	5,58	IPVA	-10,44
ICMS	14,3	VA_BrtServExclAdm	5,64	SINAP_Minimo	5,16	VA_ImpostosProd	-10,31
ICMS_Muns	13,9	FinMA_Mn_JrNom	5,6	IPVA_Muns	5,36	VA_BrtIndust	-10,62
TOTAL_ArrecMun	13,69	x_3	5,35	SINAP_Baixo	5,41	Tx_SELIC_Perd	-10,91
PIB_MenBR	11,68	IPP	5,59	SINAP_Normal	5,62		

Fonte: Elaborado pela Autora (2023).

Ao analisar o comparativo dos valores preditos com os realizados do Gráfico 6, não se obteve um bom resultado, mesmo com as métricas razoáveis, há um certo acompanhamento neste teste, mas com muitos atrasos e ao final apontou uma tendência de subida quando deveria ser de queda. O teste de *Friedman*³⁷ novamente, não rejeitou a hipótese nula também, assim, as métricas não possuem diferenças significativas entre os algoritmos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

³⁷ Obtendo o resultado da estatística = 5.400 e *p-value* = 0,145.

Gráfico 6 - Comparativo dos valores das Predições x Originais da Receita do IPI nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

O Modelo Combinado Lasso com *XGBoost* mostrou melhor RMSE no Quadro 13, de 0,889, com os dados das cotas oficiais da receita do FEP, e com os dos repasses extraordinários, no Quadro 14, obteve 0,621 com *LinearRegression* e *XGBoost*. Vale ressaltar que no conjunto dos repasses extraordinários houve um empate e o critério para escolher o Modelo Combinado, foi pelo seu maior R^2 . Contudo, o teste de *Friedman*³⁸ confirma que não se rejeitou a hipótese nula, portanto, as métricas não possuem diferenças significativas entre os algoritmos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 13 - Resultados das métricas da seleção das variáveis da Receita do FEP de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	21	1,348	0,957	1,816	0,000
<i>RandomForest</i>	17	1,438	1,117	2,067	0,000
<i>XGBoost</i>	22	1,045	0,759	1,093	0,181
Mod. Combinado					
<i>Lasso-XGBoost</i>	31	0,889	0,661	0,791	0,407

Features do Mod. Combinado Lasso-XGBoost:							
x_3	0	TxDesocupados	0	x_2	-1E-04	ST_EXTERNO	0
ResvInternac	0	SINAP_Normal	0	Importacao	0	ICMS_TOTAL	0
EXPT_SLZ_ProdIndQuim	0	SINAP_Minimo	0	Exportacao	0	ICMS_DvAtTrb	0
EXPT_SLZ_TOTAL	-1E-04	SINAP_Baixo	0	IBC_Br	0	ICMS	0
PIB_MenBR	0	SINAP_Alto	0	IGPM	0	DolarComercial	0
IPP	0	INPC	-1E-04	IPVA	0	x_1	0
Petroleo_Brent_Brasil	0	Tx_SELIC_Perd	-1E-04	ICMS_SIMP_NACIO	0	PreInvRndPrCpMn	-1E-04
TxIPCA	-1E-04	SalarioMinimo	0	ST_INTERNO	-1E-04		

Fonte: Elaborado pela Autora (2023).

Em se tratando das variáveis selecionadas pelos dois *datasets*, estão compatíveis (Quadro 13 e Quadro 14), sendo que os repasses extraordinários elegeram como principal característica o “*Coef_CIFPM_SLZ*”, o qual faz a distribuição dos repasses aos Municípios, atendendo ao requisito do efeito quantidade, todavia, faltou para as cotas oficiais. Em ambas,

³⁸ Com resultado da estatística = 4.800 e *p-value* = 0,187.

foram selecionados o “*Petroleo_Brent_Brasil*” e “*DolarComercial*”, além das variáveis de efeito preço, “*TxIPCA*”, “*INPC*”, “*IBC_Br*”, “*IGPM*” e “*Tx_SELIC_Pred*” (BRASIL, 2000).

Gráfico 7 - Comparativo dos valores das Predições x Originais da Receita do FEP nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

Mesmo com um bom contexto de variáveis selecionadas, o comparativo do Gráfico 7, dos valores previstos e originais não acompanharam muito bem os movimentos dos repasses.

Quadro 14 - Resultados das métricas da seleção das variáveis da Receita do FEP com repasses extraordinários

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	21	0,795	0,643	0,631	0,000
<i>RandomForest</i>	9	0,771	0,624	0,595	0,000
<i>XGBoost</i>	22	0,615	0,503	0,379	0,000
Mod. Combinado					
<i>LinearRegression-XGBoost</i>	39	0,621	0,493	0,385	0,314

Features do Mod. Combinado LinearRegression-XGBoost:					
Coef_CIFPM_SLZ	0,3206	ICMS_TOTAL	0,0301	ST_INTERNO	0,0119
DvLq_MA_ExcCapital	0,2098	PetoleoBrentBR	0,03	DolarComercial	0,0107
DvLq_Sao_Luis_MA	0,0698	Tx_SELIC_Perd	0,0263	SINAP_Baixo	0,0103
DvLq_Gov_MA	0,045	ICMS_Muns	0,0205	SINAP_Mínimo	0,0076
TOTAL_ArrecMun	0,0435	CFEM_MA	0,0149	TxDesocupados	0,0069
PIB_Mun	0,0336	FinMA_Mn_JrNom	0,0131	Gasolina_BR	0,0068
				SINAP_Alto	0,0068
				ResvInternac	0,0064
				Importacao	0,0053
				PIM_PF	0,0051

Fonte: Elaborado pela Autora (2023).

No FPM, as métricas do RMSE ficaram bem próximas, os repasses extraordinários, conforme Quadro 16, foram o menor, 0,534, e os das cotas oficiais alcançaram 0,578 registrado no Quadro 15. O algoritmo, nos dois conjuntos de dados, foi o mesmo: a combinação com do *Ridge* com *GradientBoosting*. Diante de tal mensuração, o teste de *Friedman*³⁹ também não rejeitou a hipótese nula, indicando que as métricas não possuem diferenças significativas entre os algoritmos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

³⁹ Apresentou resultado da estatística = 6.000 e *p-value* = 0,112.

Quadro 15 - Resultados das métricas da seleção das variáveis da Receita do FPM de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	28	0,703	0,566	0,494	0,157
<i>RandomForest</i>	17	0,930	0,686	0,866	0,000
<i>XGBoost</i>	19	0,676	0,521	0,456	0,221
Mod. Combinado					
<i>Ridge-GradientBoosting</i>	20	0,578	0,474	0,334	0,431

Featuers do Ensemble GradientBoosting:					
IPVA	20,27	Gasolina_Brasil	8,92	DvLq_Sao_Luis_MA	-8,4
x_1	20,26	ICMS_Muns	8,49	Importacao	-9,5
DvLq_MA_ExcCapit	15,59	TOTAL_ArrecMun	8,44	TransacoesEspeciais	-10,93
CFEM_Total	10,8	TxIPCA	8,4	CFEM_MA	-12,07
PIB_MenBR	10,52	Petroleo_Brent_Bras	8,37	IGPM	-13,07
				IBC_Br	-42,38
				x_3	-13,44
				FinMA_MnRestPrim	-15,95
				FinMA_MnRestNom	-17,29
				TxDesocupados	-82,46

Fonte: Elaborado pela Autora (2023).

A quantidade de variáveis baixou em seis unidades, identificada na ausência do “Coef_CIFPM_SLZ”, nas cotas oficiais (Quadro 15), mas a sua presença nos repasses extraordinários como redutora, requer uma reavaliação. Em ambos, observa-se a presença dos dois efeitos: quantidade e preço (BRASIL, 2000).

Quadro 16 - Resultados das métricas da seleção das variáveis da Receita do FPM com repasses extraordinários

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	35	0,778	0,598	0,605	0,000
<i>RandomForest</i>	10	1,056	0,857	1,116	0,000
<i>XGBoost</i>	35	0,767	0,576	0,588	0,000
Mod. Combinado					
<i>Ridge-GradientBoosting</i>	26	0,534	0,425	0,285	0,405

Featuers do Modelo Combinado Ridge-GradientBoosting:					
ICMS	47,16	IPP	8,59	ICMS_DvAtTrb_Muns	5,57
ICMS_TOTAL	20,09	SINAP_Baixo	8,17	DvLq_Sao_Luis_MA	-5,95
PetroleoBrentBR	18,46	PerPopulacaoUF	8,09	Importacao	-6,31
EXPT_SLZ_ProdIndQuim	16,31	IPVA	8,03	ResvInternac	-7,38
Gasolina_BR	12,68	CFEM_Total	7,14	ST_EXTERNO	-8,8
Ajustes_MA_Munic	9,38	Tx_SELIC_Perd	5,89	ICMS_SimpNacio	-9,38
INPC	9,17	SINAP_Mínimo	5,83	DvLq_MA_ExcCapital	-10,64
				ST_INTERNO	-12,6
				ICMS_DvAtTrb	-14,85
				Coef_CIFPM_SLZ	-21,78
				Coef_IPM	-23,94
				TxDesocupados	-27,91

Fonte: Elaborado pela Autora (2023).

Analisando o comparativo do Gráfico 8, percebe-se que a performance não foi tão boa, mas as predições acompanharam alguns picos e em outros períodos ficou atrasada, e no final ficou bem distante.

Gráfico 8 - Comparativo dos valores das Predições x Originais da Receita do FPM nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

O Modelo Combinado foi o que obteve menor RMSE nos dois *datasets* para a receita de IPVA, de acordo com o Quadro 17, ele alcançou 0,463 para as cotas oficiais e 0,499 para os repasses extraordinários, conforme Quadro 18. O primeiro caso foi selecionado por *Lasso* aplicando *XGBoost*, e no segundo, pelo *RandomForest* com o *GradientBoosting*. Mesmo com esta melhora na métrica, o teste de *Friedman*⁴⁰ não rejeitou a hipótese nula, indicando que as métricas não possuem diferenças significativas entre os algoritmos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 17 - Resultados das métricas da seleção das variáveis da Receita do IPVA de cotas oficiais

Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	4	0,550	0,330	0,303	0,643
<i>RandomForest</i>	5	0,501	0,345	0,251	0,704
<i>XGBoost</i>	25	0,693	0,472	0,480	0,435
Mod. Combinado					
<i>Lasso-XGBoost</i>	20	0,463	0,319	0,214	0,748

Features do Mod. Combinado Lasso-XGBoost:							
PrcInvRndPrCpM	0,01	SalarioMinimo	0,00	x_3	0,00	ICMS_DvAtTrb	-0,00
x_2	-0,00	Importacao	-0,00	ICMS_SIMP_NACIO	0,00	ICMS	0,00
INPC	0,00	Exportacao	-0,00	ST_INTERNO	-0,00	DolarComercial	0,00
Tx_SELIC_Perd	-0,00	IBC_Br	-0,00	ST_EXTERNO	0,00	x_1	-0,00
ResvInternac	0,00	IGPM	-0,00	ICMS_TOTAL	0,00	IPVA	0,00

Fonte: Elaborado pela Autora (2023).

Destacam-se, no *ranking* dos repasses extraordinários que a 1ª posição ficou com o “IPVA” (80,63%) e a 2ª para o “IPVA_Muns” (7,35%). Contudo, este último é o que define o repasse, pois 50% da arrecadação⁴¹ dos veículos emplacados no Município retornam para ele (BRASIL, 1988). Já no das cotas oficiais a 1ª variável ficou bem abaixo das demais, mas com menos variáveis alcançou uma melhora, e ambas registraram efeitos preço e quantidade

⁴⁰ Apresentou resultado da estatística = 5,400 e *p-value* = 0,145.

⁴¹ Inciso “III”, do art. 158, da Constituição Federal.

(BRASIL, 2000).

Quadro 18 - Resultados das métricas da seleção das variáveis da Receita do IPVA com repasses extraordinários

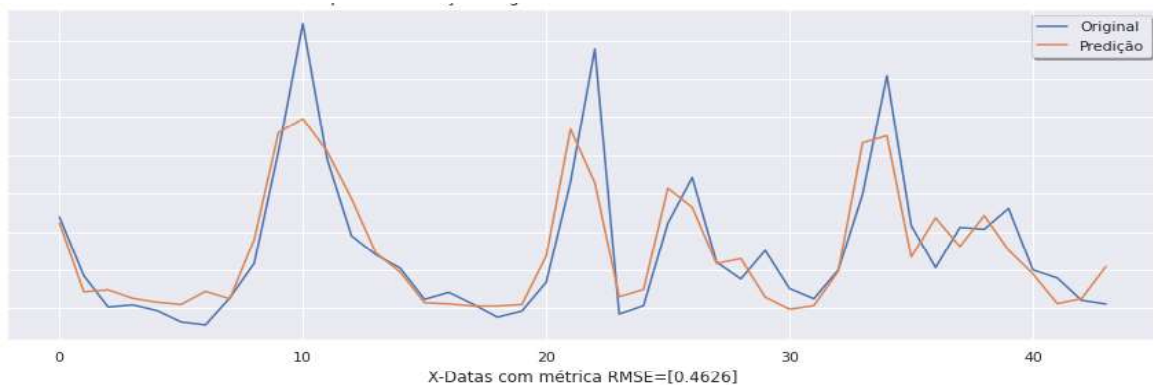
Tipos de Seletores	k Features	RMSE	MAE	MSE	R2
Ensemble					
<i>GradientBoosting</i>	9	0,655	0,326	0,429	0,490
<i>RandomForest</i>	6	0,540	0,320	0,292	0,653
<i>XGBoost</i>	22	0,609	0,343	0,371	0,560
Mod. Combinado					
<i>RandomForest-GradientBoosting</i>	36	0,499	0,343	0,249	0,729

Featuas do Modelo Combinado RandomForest-GradientBoosting:					
IPVA	80,63	Tx_SELIC_Perd	0,37	CFEM_Total	0,21
IPVA_Muns	7,35	x_2	0,35	ICMS_DvAtTrb	0,19
EXPT_SLZ_TOTAL	1,55	FinMA_MnRestNom	0,34	ICMS_DvAtTrb_Muns	0,17
x_3	1,19	TxIPCA	0,28	DolarComercial	0,17
FinMA_MnRestPrim	0,85	EXPT_SLZ_ProdIndQuim	0,28	ResvInternac	0,14
CFEM_Mun	0,77	Importacao	0,26	ST_EXTERNO	0,14
IGPM	0,71	INPC	0,25	TOTAL_ArrecMun	0,14
x_1	0,66	PetoleoBrentBR	0,23	TxDesocupados	0,13
CFEM_MA	0,44	PIM_PF	0,22	ST_INTERNO	0,13
				SINAP_Mínimo	0,07

Fonte: Elaborado pela Autora (2023).

No que se refere ao comparativo dos valores previstos com o teste nos originais, o Gráfico 9 demonstra que o modelo acompanhou a movimentação dos depósitos.

Gráfico 9 - Comparativo dos valores das Predições x Originais da Receita do IPVA nos testes de cotas oficiais



Fonte: Elaborado pela Autora (2023).

Vale ressaltar que foram obtidos valores negativos para o R^2 , os quais foram mantidos no código apenas para efeitos de escolha do melhor modelo, uma vez que a literatura indica que tal ocorrência é possível quando as variáveis explicativas têm pouco efeito sobre a variável independente, ou melhor, o modelo foi pior que um valor médio (FÁVERO e BELFIORE, 2017, p. 540). Assim, optou-se por seguir o mesmo procedimento que o Softwares *Minitab* orienta: substituir o valor negativo por ZERO, pois entende-se que seria um indicativo de um modelo sobreajustado (MINITAB, 07 maio 2019).

Ademais, a Análise de Componentes Principais, em língua materna *Principal Component Analysis* (PCA) foi aplicado apenas para avaliar se a redução de dimensionalidade poderia ser uma alternativa, contudo, ela criou, em média, 20 componentes para explicar 95% dos subconjuntos das receitas, mas selecionou variáveis similares para a maioria das receitas com métricas diferentes. Tais variáveis diferiram das indicações de estudos econométricos, por este motivo não foram incluídas nas comparações.

5.2 PRIMEIRA ITEAÇÃO DA MODELAGEM DOS MODELOS

Tratando-se da modelagem dos algoritmos de aprendizagem profunda, do Quadro 5 foram treinados, com a biblioteca *Keras*⁴², os modelos de Rede Neural Convolucional, tradução de *Convolutional Neural Network* (CNN), GRU, LSTM simples e complexa, e RNN simples e complexa. A arquitetura de todas as redes foram as mais básicas para esta primeira iteração. Todos os modelos codificados utilizaram parâmetros de treinamento iguais, a técnica de *Slide Windows* com *time step* de 2 e *lookback* de 12, número máximo de épocas foi de 100, mas aplicou-se *Checkpoint* para salvar a cada melhor desempenho no treinamento (função *CalBack* do pacote *Keras*). Sendo importante registrar que antes de fazer a predição o melhor modelo é resgatado e carregado para aplicar no conjunto de teste, padronizado para o ano de 2022. A função de otimização foi a “*Adam*” com taxa de aprendizagem de 0,0001, exceto para RNN que foi de 0,01. A função de perda foi o MSE e foi incluída a métrica RMSE.

A arquitetura da CNN foi treinada com uma camada *Conv1D*, seguida por outra camada *Flatten* e duas *Dense*. A camada convolucional, *Conv1D*, teve 64 mapas de filtros e kernel de tamanho 2. Na sequência, incluiu-se uma camada *Flatten*, seguida de uma *Dense* com saída de 8 vias e função de ativação *Rectified Linear Unit* ou unidade linear retificada (ReLU) para reduzir os mapas de recursos em um único vetor unidimensional. Por fim, utilizou outra camada *Dense* com quantidade configurável de vias de saída para o horizonte temporal que estiver sendo treinado (GÉRON, 2021, p. 402-403; BROWNLEE, 2019, p. 90).

Tanto a arquitetura de rede neural recorrente LSTM simples quanto a complexa foram bem parecidas, diferindo apenas na quantidade de camadas. A LSTM simples utilizou duas camadas LSTM, ambas com a função de ativação ReLU, sendo a primeira com 64 unidades LSTM na camada oculta e a segunda com 32, seguida de uma de *Dropout* com taxa de 0,2, finalizando com uma *Dense* com quantidade configurável de vias de saída para o horizonte

⁴² <https://keras.io/>.

temporal que estiver sendo treinado (BROWNLEE, 2019, p. 128). A complexa só duplicou as camadas de LSTM e antes da de *Dropout*, foi incluída uma *Dense* de 8 vias com função de ativação ReLU.

A GRU usou uma camada de entrada possuindo 64 unidades GRU na camada oculta, seguida de uma camada *Dense* de 8 vias com função de ativação ReLU, finalizando com outra *Dense* com quantidade configurável de vias de saída para o horizonte temporal que estiver sendo treinado (GÉRON, 2021, p. 402).

Por fim, a RNN simples utilizou uma camada de 64 unidades RNN na camada oculta, com função de ativação *softmax*. Na sequência, incluiu-se uma camada *Flatten*, finalizando com outra *Dense* com quantidade configurável de vias de saída para o horizonte temporal que estiver sendo treinado. E a RNN complexa inclui-se mais uma camada de 64 unidades RNN na camada oculta e outra camada *Dense* de 8 vias com função de ativação ReLU (GÉRON, 2021, p. 402).

Antes de relatar os resultados da modelagem, é necessário lembrar que a meta de diferença de São Luís seria de 5%, mas para todos os efeitos da legislação (BRASIL, 2000), neste estudo estão sendo testados apenas os efeitos preço e quantidade.

Quadro 19 - Resultado do treinamento dos algoritmos de AM para a CFEM

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	59	0,7	0,8734	0,7628	31.658,74	19.567,52	51.226,26
GRU	70	2,33	0,1385	0,0192	8.567,94	12.497,94	21.065,88
LSTM	86	3,97	0,1279	0,0164	15.718,46	15.274,30	30.992,76
LSTM_COMPLEX	76	4,45	0,1573	0,0247	16.807,47	16.861,99	33.669,46
RNN	65	3,39	0,1744	0,0304	16.805,94	16.587,29	33.393,23
RNN_COMPLEX	29	2,4	0,1127	0,0127	14.062,86	15.579,68	29.642,54

Fonte: Elaborado pela Autora (2023).

Passando-se a discorrer sobre os resultados, registra-se que o subconjunto da receita da CFEM AFETADOS não foi submetida à modelagem por ter uma quantidade inferior a 100 observações. Portanto, inicia-se o relato com a receita da CFEM, para a qual foram filtradas as 38 variáveis elencadas no Quadro 7, sendo este subconjunto aplicado aos algoritmos anteriormente citados cujos resultados foram tabulados no Quadro 19. A RNN complexa foi a que apresentou menor RMSE e MSE, em 29 épocas (a menor quantidade), num treino de 2,4 minutos, produzindo uma previsão bimestral de R\$ 29.642,54, com R\$ 14.062,86 para janeiro e R\$ 15.579,68 para fevereiro. Em termos de tempo de execução a melhor foi a CNN com 0,7

minutos, seguida da GRU com 2,33 minutos, mas conforme Tabela 8 suas previsões não foram boas.

Tabela 8 - Diferenças entre Valores preditos e realizados dos treinos para a CFEM

Modelo	RMSE	Jan-22 = R\$ 14.208,74			Fev-22 = R\$ 15.555,34			Bimestre = R\$ 29.764,08		
		Predito	Diferença	% Dif.	Predito	Diferença	% Dif.	Predito	Diferença	% Dif.
CNN	0,8734	31.658,74	-17.450,00	122,81%	19.567,52	-4.012,18	28,24%	51.226,26	-21.462,18	151,05%
GRU	0,1385	8.567,94	5.640,80	39,70%	12.497,94	3.057,40	21,52%	21.065,88	8.698,20	61,22%
LSTM	0,1279	15.718,46	-1.509,72	10,63%	15.274,30	281,04	1,98%	30.992,76	-1.228,68	8,65%
LSTM_COMPLEX	0,1573	16.807,47	-2.598,73	18,29%	16.861,99	-1.306,65	9,20%	33.669,46	-3.905,38	27,49%
RNN	0,1744	16.805,94	-2.597,20	18,28%	16.587,29	-1.031,95	7,26%	33.393,23	-3.629,15	25,54%
RNN_COMPLEX	0,1127	14.062,86	145,88	1,03%	15.579,68	-24,34	0,17%	29.642,54	121,54	0,86%
MÉDIA	0,2640	17.270,24	-3.061,50	21,55%	16.061,45	-506,11	3,56%	33.331,69	-3.567,61	25,11%

Fonte: Elaborado pela Autora (2023).

Comparando-se os valores realizados no primeiro bimestre de 2022, cujos depósitos totalizaram R\$ 29.764,08, sendo que o realizado em janeiro foi R\$ 14.208,74 e em fevereiro R\$ 15.555,34, obteve-se uma diferença de R\$ 121,54 a menor que o realizado com a RNN complexa. Mesmo fazendo a subtração entre o valor predito e o realizado para todos os algoritmos, a menor diferença no bimestre, em valores absolutos, continua sendo da RNN complexa, cujo valor ficou menor que a média entre todas as diferenças e atingiu um percentual de 0,86% superando a meta de 5% de diferença, conforme registrado na Tabela 8. Ao aplicar o teste de Friedman na receita da CFEM a hipótese nula não foi rejeitada, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 20 - Resultado do treinamento dos algoritmos de AM para o ICMS

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	88	1,38	0,3936	0,1549	67.869.504,00	64.909.236,00	132.778.740,00
GRU	88	5,39	0,7114	0,5061	57.261.072,00	54.582.236,00	111.843.308,00
LSTM	98	4,4	0,4926	0,2427	64.036.864,00	62.552.484,00	126.589.348,00
LSTM_COMPLEX	87	6,45	0,4888	0,2389	63.529.060,00	60.846.648,00	124.375.708,00
RNN	100	4,38	1,1521	1,3274	49.143.176,00	48.306.336,00	97.449.512,00
RNN_COMPLEX	9	1,49	0,4981	0,2481	63.302.520,00	63.237.356,00	126.539.876,00

Fonte: Elaborado pela Autora (2023).

Com a receita de ICMS, separando-se as 35 variáveis do Quadro 9, com as cotas oficiais, destacou o melhor desempenho do RMSE (0,3936) para a CNN em 88 épocas usando 1,38 minutos para treino (o menor tempo), prevendo R\$ 132.778.740,00 para o primeiro bimestre de 2022, com R\$ 67.869.236,00 para janeiro e R\$ 64.909.236,00 para fevereiro. No Quadro 20, ainda foi possível verificar que a menor quantidade de épocas foi da RNN complexa com 9, seguida da LSTM complexa com 87, a qual obteve a melhor predição, conforme Tabela 9.

Tabela 9 - Diferenças entre Valores preditos e realizados dos treinos para o ICMS

Modelo	RMSE	Jan-22 = R\$ 70.700.563,89			Fev-22 = R\$ 53.251.349,76			Bimestre = R\$ 123.951.913,65		
		Previsto	Diferença	% Dif.	Previsto	Diferença	% Dif.	Previsto	Diferença	% Dif.
CNN	0,3936	67.869.504,00	2.831.059,89	4,00%	64.909.236,00	-11.657.886,24	21,89%	132.778.740,00	-8.826.826,35	7,12%
GRU	0,7114	57.261.072,00	13.439.491,89	19,01%	54.582.236,00	-1.330.886,24	2,50%	111.843.308,00	12.108.605,65	9,77%
LSTM	0,4926	64.036.864,00	6.663.699,89	9,43%	62.552.484,00	-9.301.134,24	17,47%	126.589.348,00	-2.637.434,35	2,13%
LSTM_COMPLEX	0,4888	63.529.060,00	7.171.503,89	10,14%	60.846.648,00	-7.595.298,24	14,26%	124.375.708,00	-423.794,35	0,34%
RNN	1,1521	49.143.176,00	21.557.387,89	30,49%	48.306.336,00	4.945.013,76	9,29%	97.449.512,00	26.502.401,65	21,38%
RNN_COMPLEX	0,4981	63.302.520,00	7.398.043,89	10,46%	63.237.356,00	-9.986.006,24	18,75%	126.539.876,00	-2.587.962,35	2,09%
MÉDIA	0,6228	60.857.032,67	9.843.531,22	13,92%	59.072.382,67	-5.821.032,91	10,93%	119.929.415,33	4.022.498,32	3,25%

Fonte: Elaborado pela Autora (2023).

Quando se faz a comparação dos valores executados da receita de ICMS cujo primeiro bimestre de 2022 somou R\$ 123.951.913,65, com janeiro alcançando R\$ 70.700.563,89 e fevereiro R\$ 53.251.349,76, pela predição do modelo da CNN obteve-se uma diferença no bimestre, em valores absolutos, de R\$ 8.826.826,35 a maior, bem acima da média das diferenças das previsões, um percentual de 7,12% acima da meta. E quando se mensuram as diferenças nas predições dos demais modelos, elencados na Tabela 9, a LSTM complexa fica com o melhor desempenho no bimestre, em valores absolutos, totalizando R\$ 423.794,35 a maior que o realizado, muito abaixo que a média geral dos modelos e um percentual de 0,34% de diferença superando a meta. Mas a GRU fica com a melhor predição em fevereiro de 2022 e em janeiro a melhor predição é da LSTM simples, ver Tabela 9. O teste de Friedman na receita da ICMS não rejeitou a hipótese nula, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 21 - Resultado do treinamento dos algoritmos de AM para o FEP

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	100	1,38	2,5934	6,7256	741.352,50	643.499,69	1.384.852,19
GRU	99	5,39	1,7111	2,9277	845.861,94	749.098,69	1.594.960,63
LSTM	97	6,4	1,7073	2,915	720.756,12	678.136,31	1.398.892,43
LSTM_COMPLEX	98	8,05	2,216	4,9106	695.743,81	646.795,38	1.342.539,19
RNN	100	3,69	2,3791	5,6599	639.085,00	639.631,69	1.278.716,69
RNN_COMPLEX	50	1,79	1,4584	2,1269	901.232,06	881.691,69	1.782.923,75

Fonte: Elaborado pela Autora (2023).

Para a receita do FEP, aplicando as 31 variáveis do Quadro 13, o modelo de melhor RMSE (1,4584) foi o da RNN complexa em 50 épocas treinada em 1,79 minutos, prevendo para o primeiro bimestre de 2022 R\$ 1.782.923,75, resultado de R\$ 901.232,06 do mês de janeiro mais R\$ 881.691,69 de fevereiro, conforme o Quadro 21. Este último, ainda elenca o rank do menor tempo que foi da CNN e sua predição, ao comparar com o realizado foi melhor que a RNN complexa, conforme Tabela 10.

Tabela 10 - Diferenças entre Valores preditos e realizados dos treinos para o FEP

Modelo	RMSE	Jan-22 = R\$ 719.937,71			Fev-22 = R\$ 786.058,29			Bimestre = R\$ 1.505.996,00		
		Predição	Diferença	% Dif.	Predição	Diferença	% Dif.	Predição	Diferença	% Dif.
CNN	2,5934	741.352,50	-21.414,79	2,97%	643.499,69	142.558,60	18,14%	1.384.852,19	121.143,81	8,04%
GRU	1,7111	845.861,94	-125.924,23	17,49%	749.098,69	36.959,60	4,70%	1.594.960,63	-88.964,63	5,91%
LSTM	1,7073	720.756,12	-818,41	0,11%	678.136,31	107.921,98	13,73%	1.398.892,43	107.103,57	7,11%
LSTM_COMPLEX	2,216	695.743,81	24.193,90	3,36%	646.795,38	139.262,91	17,72%	1.342.539,19	163.456,81	10,85%
RNN	2,3791	639.085,00	80.852,71	11,23%	639.631,69	146.426,60	18,63%	1.278.716,69	227.279,31	15,09%
RNN_COMPLEX	1,4584	901.232,06	-181.294,35	25,18%	881.691,69	-95.633,40	12,17%	1.782.923,75	-276.927,75	18,39%
MÉDIA	2,0109	757.338,57	-37.400,86	5,20%	706.475,58	79.582,72	10,12%	1.463.814,15	42.181,85	2,80%

Fonte: Elaborado pela Autora (2023).

Ainda sobre os repasses do FEP, a comparação das previsões com o realizado no primeiro bimestre de 2022 que foi de R\$ 1.505.996,00, decorrência da soma do mês de janeiro R\$ 719.937,71 e fevereiro R\$ 786.058,29, produziram uma diferença, no bimestre, em valores absolutos, de R\$ 276.927,75 a maior que o realizado com a RNN complexa, um percentual de diferença de 18,39%, muito acima da meta, conforme Tabela 10. Porém ao analisar por mês, a LSTM simples foi melhor em janeiro, pois teve uma diferença de R\$ 818,41 a maior, em valores absolutos, um percentual de 0,11%, ínfimo em relação a meta. Em fevereiro, a menor diferença foi da GRU, R\$ 36.959,60 a menor (4,70%), que no final do bimestre também obteve o melhor desempenho, em valores absolutos, de R\$ 88.964,63, um percentual próximo da meta 5,91%, contudo foi o dobro da média, ver Tabela 10. Para a receita do FEP, o teste de Friedman também não rejeitou a hipótese nula, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 22 - Resultado do treinamento dos algoritmos de AM para o FPM

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	95	1,38	1,463	2,1402	50.814.036,00	56.530.560,00	107.344.596,00
GRU	99	4,14	1,3029	1,6976	47.706.856,00	50.673.852,00	98.380.708,00
LSTM	88	4,41	1,1361	1,2908	55.049.192,00	61.415.008,00	116.464.200,00
LSTM_COMPLEX	95	9,47	1,3313	1,7724	44.624.752,00	61.337.596,00	105.962.348,00
RNN	100	4,39	1,6601	2,7561	40.730.136,00	42.287.548,00	83.017.684,00
RNN_COMPLEX	15	1,88	0,832	0,6923	56.670.380,00	57.413.080,00	114.083.460,00

Fonte: Elaborado pela Autora (2023).

A receita do FPM, selecionando as 20 variáveis das cotas oficiais do Quadro 15, apresentou como menor RMSE (0,832) com a RNN complexa, treinando em 15 épocas (a menor quantidade), durante 1,88 minutos, prevendo para o primeiro bimestre de 2022 o montante de R\$ 114.083.460,00, resultado da soma de R\$ 56.670.380,00 do mês de janeiro com R\$ 57.413.080,00 de fevereiro, de acordo com o Quadro 22.

Tabela 11 - Diferenças entre Valores preditos e realizados dos treinos para o FPM

Modelo	RMSE	Jan-22 = R\$ 56.972.497,09			Fev-22 = R\$ 74.870.189,54			Bimestre = R\$ 131.842.686,63		
		Previsão	Diferença	% Dif.	Previsão	Diferença	% Dif.	Previsão	Diferença	% Dif.
CNN	1,463	50.814.036,00	6.158.461,09	10,81%	56.530.560,00	18.339.629,54	24,50%	107.344.596,00	24.498.090,63	18,58%
GRU	1,3029	47.706.856,00	9.265.641,09	16,26%	50.673.852,00	24.196.337,54	32,32%	98.380.708,00	33.461.978,63	25,38%
LSTM	1,1361	55.049.192,00	1.923.305,09	3,38%	61.415.008,00	13.455.181,54	17,97%	116.464.200,00	15.378.486,63	11,66%
LSTM_COMPLEX	1,3313	44.624.752,00	12.347.745,09	21,67%	61.337.596,00	13.532.593,54	18,07%	105.962.348,00	25.880.338,63	19,63%
RNN	1,6601	40.730.136,00	16.242.361,09	28,51%	42.287.548,00	32.582.641,54	43,52%	83.017.684,00	48.825.002,63	37,03%
RNN_COMPLEX	0,832	56.670.380,00	302.117,09	0,53%	57.413.080,00	17.457.109,54	23,32%	114.083.460,00	17.759.226,63	13,47%
MÉDIA	1,28757	49.265.892,00	7.706.605,09	13,53%	54.942.940,67	19.927.248,87	26,62%	104.208.832,67	27.633.853,96	20,96%

Fonte: Elaborado pela Autora (2023).

Calculando as diferenças entre os valores executados do FPM no primeiro bimestre de 2022 que foi R\$ 131.842.686,63, resultante da soma de R\$ 56.972.497,00 do executado em janeiro com R\$ 74.870.189,54 de fevereiro. Tal montante resultou em uma diferença de R\$ 17.759.226,63, no bimestre, melhor que a média, mas um percentual de 13,47% maior que a meta, ver Tabela 11. Contudo ao analisar as demais diferenças dos outros modelos na Tabela 11, verificou-se que a LSTM simples fez a melhor previsão para o bimestre com R\$ 15.378.486,63, 11,66% de diferença para o realizado e melhor que a média, contudo ainda fica acima da meta. Para a receita do FPM, o teste de Friedman também não rejeitou a hipótese nula, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 23 - Resultado do treinamento dos algoritmos de AM para o IPI

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	73	1,37	1,1137	1,2404	528.188,62	533.549,25	1.061.737,87
GRU	71	4,39	0,5713	0,3264	656.699,19	620.698,19	1.277.397,38
LSTM	87	3,03	0,7203	0,5188	651.104,06	590.318,00	1.241.422,06
LSTM_COMPLEX	98	6,43	1,0731	1,1515	570.966,06	534.375,69	1.105.341,75
RNN	100	3,38	1,3072	1,7088	494.063,12	483.199,00	977.262,12
RNN_COMPLEX	29	1,41	0,3797	0,1442	684.828,44	658.143,38	1.342.971,82

Fonte: Elaborado pela Autora (2023).

O IPI com um subconjunto de 55 variáveis das cotas oficiais do Quadro 12, elegeu o RMSE (0,3797) da RNN complexa treinada em 29 épocas durante 1,41 minutos, produzindo uma previsão para o primeiro bimestre de 2022 de R\$ 1.342.971,82, resultantes da soma de R\$ 684.828,44 do mês de janeiro com R\$ 658.143,38 de fevereiro.

Tabela 12 - Diferenças entre Valores preditos e realizados dos treinos para o IPI

Modelo	RMSE	Loss	Jan-22 = R\$ 702.313,98			Fev-22 = R\$ 558.950,13			Bimestre = R\$ 1.261.264,11		
			Predição	Diferença	% Dif.	Predição	Diferença	% Dif.	Predição	Diferença	% Dif.
CNN	1,1137	1,2404	528.188,62	174.125,36	24,79%	533.549,25	25.400,88	4,54%	1.061.737,87	199.526,24	15,82%
GRU	0,5713	0,3264	656.699,19	45.614,79	6,49%	620.698,19	-61.748,06	11,05%	1.277.397,38	-16.133,27	1,28%
LSTM	0,7203	0,5188	651.104,06	51.209,92	7,29%	590.318,00	-31.367,87	5,61%	1.241.422,06	19.842,05	1,57%
LSTM_COMPLEX	1,0731	1,1515	570.966,06	131.347,92	18,70%	534.375,69	24.574,44	4,40%	1.105.341,75	155.922,36	12,36%
RNN	1,3072	1,7088	494.063,12	208.250,86	29,65%	483.199,00	75.751,13	13,55%	977.262,12	284.001,99	22,52%
RNN_COMPLEX	0,3797	0,1442	684.828,44	17.485,54	2,49%	658.143,38	-99.193,25	17,75%	1.342.971,82	-81.707,71	6,48%
MÉDIA	0,8609	0,8484	597.641,58	104.672,40	14,90%	570.047,25	-11.097,12	1,99%	1.167.688,83	93.575,28	7,42%

Fonte: Elaborado pela Autora (2023).

A Tabela 12 listou os resultados das diferenças dos valores executados no primeiro bimestre de 2022 para a receita do IPI que foi de R\$ 1.261.264,11, resultado da adição de R\$ 702.313,98 do mês de janeiro com R\$ 558.950,13 de fevereiro. O algoritmo da RNN complexa, em valores absolutos, obteve uma diferença abaixo da média de R\$ 81.707,71, um percentual de 6,48%, um pouco maior que a meta. Quando os demais modelos são analisados, a menor diferença que se encontrou foi na GRU com diferença bimestral de R\$ 16.133,27 (valor absoluto) a maior que o realizado, ou seja, um percentual de 1,28% que supera a meta e a média dos modelos treinados (ver Tabela 12). Aplicando o teste de Friedman para a receita do IPI, este também não rejeitou a hipótese nula, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Quadro 24 - Resultado do treinamento dos algoritmos de AM para o IPVA

Modelo	Épocas	Tempo (min)	RMSE	Loss	Predição		
					jan-22	fev-22	Bimestre
CNN	97	1,22	0,4659	0,2171	7.605.182,00	3.642.430,75	11.247.612,75
GRU	27	3,40	0,6819	0,465	4.401.547,50	2.045.559,38	6.447.106,88
LSTM	96	4,42	0,417	0,1739	3.018.726,25	3.226.109,25	6.244.835,50
LSTM_COMPLEX	90	8,46	0,4047	0,1638	4.468.508,00	3.942.075,50	8.410.583,50
RNN	100	4,39	0,7008	0,4911	5.378.548,50	5.672.377,50	11.050.926,00
RNN_COMPLEX	96	3,40	0,2873	0,0825	2.600.249,25	3.975.441,50	6.575.690,75

Fonte: Elaborado pela Autora (2023).

O IPVA elegeu 20 variáveis listadas no Quadro 17 pelas cotas oficiais, as quais foram submetidas ao treinamento com os modelos do Quadro 24 e indicaram a RNN complexa como o melhor RMSE (0,2873), obtido em 96 épocas, em 3,40 minutos, produzindo uma previsão pra o primeiro bimestre de 2022 de R\$ 6.575.690,75, resultantes da totalização de R\$ 2.600.249,25 de janeiro e R\$ 3.975.441,50 de fevereiro.

Tabela 13 - Diferenças entre Valores preditos e realizados dos treinos para o IPVA

Modelo	RMSE	Jan-22 = R\$ 5.833.637,06			Fev-22 = R\$ 11.446.102,25			Bimestre = R\$ 17.279.739,31		
		Previsto	Diferença	% Dif.	Previsto	Diferença	% Dif.	Previsto	Diferença	% Dif.
CNN	0,4659	7.605.182,00	-1.771.544,94	30,37%	3.642.430,75	7.803.671,50	68,18%	11.247.612,75	6.032.126,56	34,91%
GRU	0,6819	4.401.547,50	1.432.089,56	24,55%	2.045.559,38	9.400.542,87	82,13%	6.447.106,88	10.832.632,43	62,69%
LSTM	0,417	3.018.726,25	2.814.910,81	48,25%	3.226.109,25	8.219.993,00	71,81%	6.244.835,50	11.034.903,81	63,86%
LSTM_COMPLEX	0,4047	4.468.508,00	1.365.129,06	23,40%	3.942.075,50	7.504.026,75	65,56%	8.410.583,50	8.869.155,81	51,33%
RNN	0,7008	5.378.548,50	455.088,56	7,80%	5.672.377,50	5.773.724,75	50,44%	11.050.926,00	6.228.813,31	36,05%
RNN_COMPLEX	0,2873	2.600.249,25	3.233.387,81	55,43%	3.975.441,50	7.470.660,75	65,27%	6.575.690,75	10.704.048,56	61,95%
MÉDIA	0,4929	4.578.793,58	1.254.843,48	21,51%	3.750.665,65	7.695.436,60	67,23%	8.329.459,23	8.950.280,08	51,80%

Fonte: Elaborado pela Autora (2023).

Finalizando o relato, a Tabela 13 elencou as diferenças dos montantes realizados no primeiro bimestre de 2022 para o IPVA que foi de R\$ 17.279.739,31, decorrente da soma de R\$ 5.833.637,06 do mês de janeiro com R\$ 11.446.102,25 de fevereiro. O algoritmo da RNN complexa, com menor RMSE, alcançou uma diferença de R\$ 10.704.048,56, um percentual de 61,95% que foi maior que a média e a meta. O referido modelo teve suas predições de janeiro e fevereiro inferiores a RNN simples e a CNN cujos percentuais de diferenças ficam bem mais baixas que a média dos modelos. Assim, em valores absolutos, a CNN apresentou a menor diferença (R\$ 6.032.126,56) com um percentual de 34,91%, conforme Tabela 13. Todavia, quando se aplicou o teste de Friedman para a receita do IPVA, este também não rejeitou a hipótese nula, não havendo diferenças significativas entre os modelos (DEMˆSAR, 2006; CALVO e SANTAFÉ, 2016, p. 250).

Tabela 14 - Resumo das menores diferenças entre os valores previsto e realizados no primeiro bimestre de 2022

Receitas	Modelo	Épocas	Tempo (min)	RMSE	Bimestral			Meta = 5%	
					Predição	Realizado	Diferença	% Dif.	ok
CFEM	RNN_COMPLEX	63	2,4	0,1127	29.642,54	29.764,08	121,54	0,86%	✓
FEP	GRU	100	5,39	1,7111	1.594.960,63	1.505.996,00	-88.964,63	5,91%	✗
	RNN_COMPLEX		1,79	1,4584	1.782.923,75		-276.927,75	18,39%	✗
FPM	LSTM		4,41	1,1361	116.464.200,00	131.842.686,63	15.378.486,63	11,66%	✗
	RNN_COMPLEX		1,88	0,832	114.083.460,00		17.759.226,63	13,47%	✗
ICMS	LSTM_COMPLEX	52	6,45	0,4888	124.375.708,00	123.951.913,65	-423.794,35	0,34%	✓
	CNN	50	1,38	0,3936	132.778.740,00		-8.826.826,35	7,12%	✗
IPI	GRU	97	4,39	0,5713	1.277.397,38	1.261.264,11	-16.133,27	1,28%	✓
	RNN_COMPLEX	96	1,41	0,3797	1.342.971,82		-81.707,71	6,48%	✗
IPVA	CNN	97	1,22	0,4659	11.247.612,75	17.279.739,31	6.032.126,56	34,91%	✗
	RNN_COMPLEX	96	3,4	0,2873	6.575.690,75		10.704.048,56	61,95%	✗

Fonte: Elaborado pela Autora (2023).

Diante do exposto, foi possível verificar que a RNN complexa, pelo critério da métrica do RMSE, foi o modelo que melhor performou entre as receitas, com exceção do ICMS em que a CNN foi a melhor. Todavia a menor diferença nos valores preditos para o primeiro bimestre

de 2022, em relação aos montantes realizados, foi diferente (LSTM complexa), resumido na Tabela 13. Sendo importante comentar sobre o desempenho do algoritmo na receita da CFEM, ao se analisar os percentuais de diferença para alcançar a meta (5%) da Tabela 13, foi um dos menores (0,86%) dentro do limite da meta, espelhando o resultado das *features* selecionados no Quadro 7 cujo RMSE refletiu nas previsões de teste com uma excelente aproximação. Contudo, há de se registrar que todos os parâmetros foram básicos, nenhum foi ajustado para melhorar as previsões ainda.

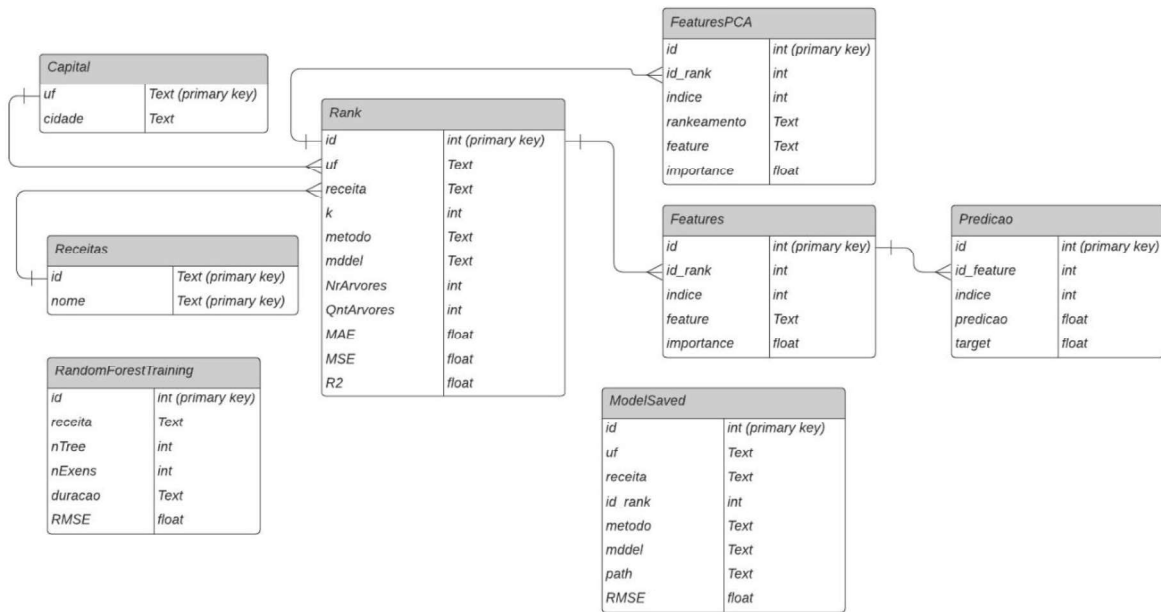
5.3 PRIMEIRA ITEAÇÃO DA IMPLANTAÇÃO DA API

No que tange a construção da API, ela é baseada em *Representational State Transfer*, em tradução livre, Transferência de Estado Representacional (REST), que de acordo com Higginbotham, (2022, p. 102; GÉRON, 2021, p. 65), consiste em um estilo de arquitetura para sistemas distribuídos evolutivos e escaláveis, baseada em do *Hypertext Transfer Protocol*, ou Protocolo de Transferência de Hipertexto, em português (HTTP). Sendo este termo cunhado por Roy Thomas Fielding⁴³ (2000 apud HIGGINBOTHAM, 2022, p. 103), o qual discutiu e apresentou, em sua tese, como os recursos do HTTP, com suas regras padronizadas, contribui para o estabelecimento de restrições que originaram um acordo de flexibilidade e capacidade de evolução para tais sistemas, principalmente baseados e rede e *web*.

A API foi desenvolvida em linguagem *Python*, utilizando a biblioteca *Flask*, tendo como base de dados independente a biblioteca *SQLite*. Tendo seu diagrama de entidade relacional representado na Figura 6. O intuito do banco é armazenar os resultados das avaliações dos algoritmos tanto da seleção de variáveis quanto das métricas dos modelos. Por isso, incluiu-se uma tabela *Capital*, para os dados do Município e seu Estado; uma para as Receitas; duas tabelas específicas para os treinos do *Random Forest* e PCA; uma de “*Rank*”, que registra as métricas dos resultados do treino dos algoritmos de seleção de variáveis que as armazena na tabela “*Features*” com suas respectivas previsões; finalizando com a tabela “*ModelSaved*” contendo os resultados dos modelos do Quadro 5.

⁴³ Para maiores detalhes consultar Roy Thomas Fielding, “Architectural Styles and the Design of Network-based Software Architectures” (PhD diss., University of California, 2000), dispone: <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.

Figura 6 - Diagrama de entidade relacional usado pela API



Fonte: Elaborado pela Autora (2023).

O sistema operacional vai ser Linux, inicialmente usando o Ubuntu, com uma versão mínima, bem como a versão do Miniconda⁴⁴, pois a intenção foi fazer a implementação em ambientes de *container*. O código foi distribuído, inicialmente, em quatro arquivos:

- 1) “*feature_selection.py*” que armazena o código de seleção das *features* por Município e Receitas e gravação dos dados no *SQLite*;
- 2) “*treinamento.py*” contém o código de treinamento das melhores variáveis por Município e Receitas;
- 3) “*functions_ML.py*” modelos de ML e salva na estrutura de diretório;
- 4) “*app.py*” que seria a API, com os comandos em *Flask* para integração com o banco, as mensagerias HTTP, integra os treinamentos e possui as rotas com retorno em formato *JavaScript Object Notation*, ou traduzindo, Notação de Objeto em *Javascript* (JSON).

Nesta primeira iteração, apenas para teste, o *dataset*, padronizado em formato CSV, foi carregado para uma pasta, denominada de “Dataset”, e quando o pedido HTTP do cliente é feito ocorre o reencaminhamento ao módulo “*app.py*”, que identifica a requisição pelas rotas criadas no Flask, sendo o seu o retorno por JSON. As rotas implementadas estão elencadas no Quadro

⁴⁴ Para mais informações consultar a documentação disponível em: <https://docs.conda.io/en/latest/miniconda.html>.

25.

Quadro 25 - Lista de rotas da API implementada para

API Routs	Pedido HTTP	Parâmetro	Descrição
# Rank receitas PCA json	http:127.0.0.1/ranks/pca/<string:estado>', methods=['GET'])	<string:estado> , sigla do Estado, exemplo= MA	Treina o PCA e retorna o json com os componetes criados, especificando as métricas alcançadas
Rank receitas PCA com detalhes em features por ESTADO	http:127.0.0.1/ranks/pca/features/<string:estado>', methods=['GET'])		Retorna o json com os componetes criados pelo PCA, bem como a listagem das melhores e piores variáveis
Rank receitas PCA com detalhes em features por ESTADO e RECEITA	http:127.0.0.1/ranks/pca/features/<string:estado>/<string:receita>', methods=['GET'])	<string:receita>', é a sigla da Receita com nome popular em maiúsculo, exemplo = FPM	O mesmo que a anterior, mas especificando apenas uma receita
# Treinamento do grid seach Random Forest	http:127.0.0.1/standalone/randomforest/training', methods=['GET'])		Ativa a função de treinamento do Random Forest, buscando os melhores paâmetros de número de árvos e amostras
# Treinamento do Xgboost por Estado	http:127.0.0.1/standalone/xboost/features/html/<string:estado>', methods=['GET'])	<string:estado> , sigla do Estado, exemplo= MA	Redireciona para a função de treinamento do Xgboost

Fonte: Elaborado pela Autora (2023).

Os modelos dos algoritmos seguiram uma estrutura de salvamento em arquivos, hierarquizados por pastas iniciando em “Model/”, criando uma subpasta “MODELO ML/”, especificando o nome do Estado, na pasta “Estado/”, seguindo para especificação das receitas, com a pasta “RECEITA/”, e nesta foi incluído o arquivo com o modelo para implementação.

A documentação completa ainda não foi produzida, ficando para próximas iterações, para este momento elaborou-se apenas a versão mais simplória de uma API para testar os dados, muitas outras rotas precisam ser incluídas, bem como protocolos de segurança e acesso.

6 CONSIDERAÇÕES FINAIS

Embora a proposta inicial visasse aplicar as mesmas técnicas expostas ao longo da dissertação nas 8 Capitais que responderam ao questionário do APÊNDICE A, com o intuito de testa a escolha das variáveis que atendessem as prerrogativas dos efeitos preço e quantidade da LRF, foi prejudicada, mas executou-se com os dados de São Luís. Atribui-se tal prejuízo, ao fato das séries históricas não estarem disponíveis nos portais da transparência da maioria das Capitais e as que tinham estavam muito despadronizadas, logo, requereria mais tempo para unificar o *dataset*. Por consequência, a implementação, não ocorreu em sua plenitude, pois era relevante que os cenários de identificação dos requisitos e dos dados das 8 Capitais fossem disponibilizados para responder a questão de pesquisa com maior abrangência.

Todavia o estudo de caso com os dados de São Luís foi executado, e a proposta passou a ser a apresentação dos resultados preliminares da primeira iteração, usando a Metodologia CRISP-DM, com um breve relato da criação do *dataset* e os ajuste básicos, focando na análise da seleção de *features*. Deste modo cumprindo parcialmente com o seu objetivo principal, ao selecionar as variáveis mais relevantes para os efeitos quantidade e preço das receitas transferidas escolhidas.

Nesse sentido, respondendo à questão de pesquisa “Q01”, sobre a possibilidade de identificação dos efeitos preço e quantidade da LRF, utilizando métodos de AM, com possibilidade de respaldar as justificativas das peças orçamentárias, no que tange a utilização de metodologias. No tópico 5 foi demonstrado que é viável a identificação destas variáveis, e como esta é apenas uma primeira iteração, o importante foi deixar os algoritmos descobrirem a importância dos atributos para cada receita, submetendo-os ao conjunto completo o que levou a resultados promissores.

Quanto a questão de pesquisa “Q02”, não foi possível responder com precisão que a predição das receitas deve considerar as variáveis pelo regime de competência trazendo-as para o períodos do fato gerador de cada receita transferida, uma vez que seria necessária a disponibilização dos dados por regime de competência e a tal situação depende de uma cooperação das Capitais.

No que tange ao objetivo específico de identificação das variáveis que impactam na execução de cada receita foi alcançado. A elaboração de um comparativo dos modelos foi executada parcialmente, haja vista que foram modelados apenas os algoritmos de aprendizagem

profunda nesta primeira iteração, mas parte da API que viabilizará esta meta foi codificada. Quanto à disponibilização dos códigos para teste e uso em outras capitais com o intuito de disseminar IA e suas técnicas de AM ao setor público, pretende-se apresentar os resultados e viabilizar os códigos fonte que já estão no drive do Laboratório de Inteligência Computacional e Programação da UEMA.

É importante registrar que a maior dificuldade se centrou na etapa de unificação do *dataset* pelo fato dos dados serem de diversas fontes com as mais variadas estruturas, além dos problemas físicos com hardwares, os quais atrasaram o cronograma. E a outra foi no pré-processamento, na identificação da descontinuidade da sequência temporal, pois mesmo com os ajustes os valores encontrados não coincidiram com o total de registro dos repasses na publicação do Balancete das Receitas de São Luís.

Vale registrar que, em relação aos dados de São Luís, a receita do IPVA, cujo repasse ocorre por 50% da arrecadação do imposto em cada Município, o qual foi incluído no *dataset* pelo atributo “*IPVA_Muns*” não alcançou relevância nos algoritmos de seleção de *features*. Voltando-se a exploração dos dados, verificou-se que os dados publicados não correspondiam ao montante particionado, induzindo a uma nova busca por outra fonte de dados que serão os dados abertos do Conselho Nacional de Política Fazendária.

Durante o processo, observou-se que devido à quantidade de ajustes que foram feitos para que os dados mantivessem uma integridade temporal foi muito grande. Tais ajustes só foram possíveis devido a transparência do Governo Federal, especialmente no IPI, mas pouco eficiente para pesquisar a CFEM AFETADOS, sendo este o principal motivo de não ser inclusa neste relato.

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, no exercício financeiro de 2022, foi identificado que em janeiro ocorreu novamente tais ausências de depósito, mas em 21 de junho os valores foram repassados ao erário público⁴⁵.

⁴⁵ Consulta no Portal da Transparência, disponível em: <https://www42.bb.com.br/portalbb/daf/beneficiarioList.802.4647.4652.0.1.1.bbx?cid=87121>.

No ICMS também ocorreu, mas há uma nota no site da transparência do Estado do Maranhão, informando uma mudança do “*Coef_IPM*”, em que os valores foram recalculados. Todavia estes montantes ainda continuam muito confusos, pois não são compatíveis com o registro da Contabilidade da Capital. Já o IPVA, ao reavaliar a série temporal da *feature* “*IPVA_Muns*” constatou-se que ela não tem representatividade no total repassado à Capital, tal fato também foi notificado ao Gestor Municipal. Diante do exposto, possivelmente, os dados não sejam tão representativos, sendo necessária a revisão das séries temporais do IPI, IPVA e ICMS, conforme orientações de Géron (2019).

Nos resultados foi possível analisar o reflexo destes eventos, pois mesmo o Modelo Combinado obtendo a supremacia na métrica do RMSE, principalmente, no *dataset* dos repasses extraordinários, corroborando com o estado da arte que já consagra esta abordagem por melhorar a precisão das previsões (PETROPOULOS, APILETTI, *et al.*, 2022, p. 737). Vale ressaltar que no comparativo dos dados com inclusão de repasses extraordinários foi possível identificar com clareza o efeito quantidade, mas o do preço não foi tão evidente nos resultados, aparecendo mais quando se testou apenas as cotas oficiais (BRASIL, 2000). 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, para averiguar se ocorrem as mesmas situações.

Diante do exposto, entende-se que boa parte dos objetivos foram atendidos, uma vez que a questão de pesquisa “Q01” foi respondida positivamente e a “Q02”, sobre o uso dos períodos de competências, necessita de mais teste para encontrar evidências concretas, a API foi codificada para a seleção de variáveis e os códigos fontes estão disponíveis para disseminação. Ademais, por se tratar do relato da primeira iteração, o trabalho não está acabado, pois ainda serão produzidos outros artefatos para disseminação das técnicas de IA para o Setor Público.

6.1 TRABALHOS FUTUROS

Ao final desta primeira iteração, com a avaliação dos resultados, percebe-se que ainda há muita atividade para ser executada, bem como artefatos para serem desenvolvidos e disponibilizados. A principal, será a integração dos dados das demais Capitais para averiguar se as variáveis de efeito preço e quantidade serão selecionadas. Registrando que é esperado que nas receitas de transferência estadual as variáveis poderão ser diferentes, considerando os diferentes setores de produção de riqueza das Unidades da Federação.

Outra, crucial, será a finalização da API com a possibilidade de testar outras variáveis de forma dinâmica, bem como gerar os modelos de AM para gerar previsões, incrementando com a geração de múltiplos cenários interligados ao Relatório FOCUS do Banco Central de projeções de índices econômicos. Seria muito interessante testar a transformação das variáveis em variações do mês atual com o anterior ou mesmo com o mesmo mês no ano anterior.

Com a finalização da API, a comparação dos modelos treinados com as variáveis para efeito preço e quantidade será de suma importância a comparação dos resultados dos modelos, tendo como *baseline* a fórmula da SOF, a equação [1] (p. 31), utilizando médias móveis, criando modelos em diferentes horizontes temporais.

Como alternativa, para a seleção de variáveis, seria conveniente testar outras abordagens, como *staking* e *voting*, separando as variáveis por tipo de efeito (preço e quantidade) com a finalidade de ter as mais importantes para aplicar nos modelos de previsão de receita. Uma verificação que se pretende testar é o desempenho das bases de dados com valores deflacionados, aplicando o teste de *Friedman* para identificar se há diferenças significativas.

Outro trabalho futuro interessante seria a aplicação de técnicas de *Explainable AI*, que encontra-se em voga na área de tecnologia, especialmente pela exigência da explicação e interpretação das predições e como foram obtidas, mas ainda não há uma exigência legal direcionada ao orçamento público.

Quanto ao efeito legislação que ficou fora do escopo deste estudo, requer um investimento maior, pois seria necessário o desenvolvimento de multiagentes, com processamento de linguagem natural, com uma ontologia para leitura dos textos da legislação tributária, aplicando técnicas para avaliar o impacto na previsão das receitas. Tal aplicação envolveria a construção de uma base de dados não relacionais, as *NoSQL*.

6.2 LIMITAÇÕES DA PESQUISA

O estudo de caso apresentou limitações importantes, no que se refere a quantidade de registro na base de dados, pois 12 anos renderam 144 meses, e uma boa parte das Capitais disponibiliza suas séries históricas a partir de 2015, tornando o quantitativo de observações ainda menor. A quantidade reduzida de observações dificulta a performance de modelos de aprendizagem profunda.

REFERÊNCIAS

- ACKERMANN, A. E. F.; SELLITTO, M. A. Métodos de previsão de demanda: uma revisão de literatura. **Innovar**, 32, Julho-Setembro 2022. 83-99.
- AGUIAR, A. B. D. O pequeno mundo da pesquisa em contabilidade gerencial no Brasil: discussão sobre desenhos alternativos de pesquisa. **Revista De Contabilidade E Organizações**, [S.l.], v. 12, n. 151933, p. 16, Novembro 2018. Acesso em: 6 Dezembro 2021.
- ALMEIDA, M. E. D. S.; FERREIRA, G. R. Métodos de acompanhamento e previsão da receita pública: um estudo de município do Recife. **Revista Capital Científico - Eletrônica (RCCe)**, Guarapuava-Paraná, 16, n. 2, Abril/Junho 2018. Disponível em: <https://revistas.unicentro.br/index.php/capitalcientifico/article/view/5013>. Acesso em: 7 Outubro 2021.
- ALMEIDA, T. B. **Seleção de atributos usando abordagem Wrapper para classificação hierárquica multirrótulo**. Dissertação (Mestrado em Ciência da Computação) - UTFPR. Ponta Grossa-PR, p. 97. 2018. Orientadora: Profa. Dra. Helyane Bronoski Borges.
- ALMEIDA, V. et al. Alíquotas Tributárias Efetivas Médias para a Economia Brasileira: uma abordagem macroeconômica, mar/jun 2017. Disponível em: <https://www.scielo.br/j/rbe/a/M6ZJ43KB8TMFBYXntNTGBVx/?format=pdf&lang=pt>. Acesso em: 6 Outubro 2021.
- AN, Z.; JALLES, J. T. "On the performance of US fiscal forecasts: government vs. private information", 2020. Disponível em: <https://doi-org.ez80.periodicos.capes.gov.br/10.1108/JES-08-2019-0388>. Acesso em: 04 Outubro 2021.
- AZEVEDO, R. R. D.; SILVA, J. M. D.; GATSIOS, R. Análise crítica dos modelos de previsão de série temporal com base no ICMS estadual., 2017. Disponível em: https://www.researchgate.net/publication/312542601_Analise_Critica_dos_Modelos_de_P_revisao_de_Serie_Temporal_com_Base_no_ICMS_Estadual. Acesso em: 30 Setembro 2021.
- BARANAUSKAS, J. A.; MONARD, M. C. **The MLC++ Wrapper for Feature Subset Selection Using Decision Tree, Production Rule, Instance Based and Statistical Inducers: Some Experimental Results**. ICMC-USP. São Carlos-SP, p. 40. 1999. (87).
- BCB, BANCO CENTRAL DO BRASIL. Nota Técnica do Banco Central do Brasil nº 52. **Procedimento de mensuração do Produto Interno Bruto nominal**, 2021. ISSN 52.

Disponível em: https://www.bcb.gov.br/content/publicacoes/notastecnicas/NT_52_202107.pdf. Acesso em: 20 Setembro 2021.

BLACK, C. Preços das commodities, termos de troca e crescimento econômico brasileiro nos anos 2000. **Indicadores Econômicos FEE**, Porto Alegre/RS, 42, 2015. 27-44. Disponível em: <https://revistas.planejamento.rs.gov.br/index.php/indicadores/article/download/3466/3533>.

BRAND, A. W. **Data Mining For Dummies**. Hoboken, New Jersey: John Wiley & Sons, 2014. ISBN 978-1-118-89316-6.

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. Brasília, DF, 17 Março 1964. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l4320.htm. Acesso em: 2022 Maio 26.

_____. **Emenda Constitucional Nº 18, de 1º de dezembro de 1965**. Reforma do Sistema Tributário., Brasília, 1965.

_____. **Decreto-Lei nº 200**. Dispõe sobre a organização da Administração Federal, estabelece diretrizes para a Reforma Administrativa e dá outras providências. Brasília-DF, 25 fevereiro 1967. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/Del0200.htm. Acesso em: 15 Agosto 2022.

_____. **Decreto-Lei nº 1.881, de 27 de Agosto de 1981**. Altera a Lei nº 5.172, de 25 de outubro de 1966, cria a Reserva do Fundo de Participação dos Municípios - FPM e dá outras providências. Brasília, 1981.

_____. **Lei nº 7.453, de 27 de dezembro de 1985**. Modifica o artigo 27 e seus parágrafos da Lei nº 2.004, de 3 de outubro de 1953, alterada pela Lei nº 3.257, de 2 de setembro de 1957, que "dispõe sobre a Política Nacional do Petróleo e define as atribuições do Conselho Nacional do Petróleo, institui a Sociedade por Ações Petróleo Brasileiro Sociedade Anônima e dá outras providências", Brasília, 1985.

_____. **Constituição Federal**. [S.l.]. 1988.

_____. **Lei Complementar nº 61, de 26 de dezembro de 1989**. Estabelece normas para a participação dos Estados e do Distrito Federal no produto da arrecadação do Imposto sobre Produtos Industrializados IPI, relativamente às exportações. Brasília, 1989.

_____. **Lei Complementar nº 62, de 28 de dezembro de 1989**. Estabelece normas sobre o cálculo, a entrega e o controle das liberações dos recursos dos Fundos de Participação e dá outras providências. Brasília, 1989a.

_____. **Lei nº 7.990 de 28 de dezembro de 1989.** Institui, para os Estados, Distrito Federal e Municípios, compensação financeira pelo resultado da exploração de petróleo ou gás natural, de recursos hídricos para fins de geração de energia elétrica, de recursos, minerais em seus respectivos territórios plataformas continental, mar territorial ou zona econômica exclusiva, e dá outras providências. (Art. 21, XIX da CF). Brasília, 1989b.

_____. **Lei Complementar Nº 63, de 11 de janeiro de 1990.** Dispõe sobre critérios e prazos de crédito das parcelas do produto da arrecadação de impostos de competência dos Estados e de transferências por estes recebidos, pertencentes aos Municípios, e dá outras providências. Brasília-DF, 11 Janeiro 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp63.htm.

_____. **Lei Complementar nº 101.** Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, Brasília, 4 Maio 2000.

_____. **Lei nº 11.638, de 28 de dezembro de 2007.** Altera e revoga dispositivos da Lei no 6.404, de 15 de dezembro de 1976, e da Lei no 6.385, de 7 de dezembro de 1976, e estende às sociedades de grande porte disposições relativas à elaboração e divulgação de demonstrações financeiras. Brasília, 2007.

_____. **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, informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios. Brasília, DF, 2009. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp131.htm. Acesso em: 15 Novembro 2021.

_____. **Lei nº 12.527, de 18 de dezembro de 2011.** Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília/DF, 18 Novembro 2011.

_____. **Lei nº 12.734, de 30 de novembro de 2012.** Modifica as Leis nº 9.478, de 6 de agosto de 1997, e nº 12.351, de 22 de dezembro de 2010, para determinar novas regras de distribuição entre os entes da Federação dos royalties e da participação especial devidos em função da exploração de petróleo, gás natural e outros hidrocarbonetos fluidos, e para aprimorar o marco regulatório sobre a exploração desses recursos no regime de partilha.

Brasília, 2012. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112734.htm.

_____. **Lei nº 13.540, de 18 de dezembro de 2017.** Altera as Leis n.º 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). Brasília, Dezembro 2017.

_____. **Decreto nº 9.407, de 12 de junho de 2018.** Regulamenta o disposto no inciso VII do § 2º e no § 5º do art. 2º da Lei nº 8.001, de 13 de março de 1990. Brasília, 2018.

_____. **Emenda Constitucional nº 108, de 26 de agosto de 2020.** Altera a Constituição Federal para estabelecer critérios de distribuição da cota municipal do Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS), para disciplinar a disponibilização de dados contábeis pelos entes federados, para tratar do planejamento na ordem social e para dispor sobre o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb); altera o Ato das Disposições Constitucionais Transitórias; e dá outras providências. Brasília, 2020. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc108.htm.

_____. **Nota Técnica Conjunta nº 03: Subsídios à apreciação do Projeto de Lei de Diretrizes Orçamentárias para 2022.** Congresso Nacional. Brasília. 2021c.

_____. Transferências a Estados e Municípios. **Tesouro Nacional Transparente**, s.d. Disponível em: <https://www.tesourotransparente.gov.br/temas/estados-e-municipios/transferencias-a-estados-e-municipios>. Acesso em: 22 Dezembro 2021.

_____, CGU. Orçamento público. **Portal da Transparência**, s.d. Disponível em: <https://www.portaltransparencia.gov.br/entenda-a-gestao-publica/orcamento-publico>. Acesso em: 26 Novembro 2021.

_____, MINISTÉRIO DA FAZENDA. **Receitas Públicas:** manual de procedimentos aplicados à União, aos Estados, ao Distrito Federal e aos Municípios. 4. ed. Brasília: STN/Coordenação Geral de Contabilidade, 2007. 233 p.

_____, _____. **Princípios Básicos:** o que você precisa saber sobre as transferências fiscais da União. Brasília, DF: Presidência da República, mar. 2016. 22 p. Disponível em: https://cdn.tesouro.gov.br/sistemas-internos///apex//producao//sistemas//thot//arquivos//publicacoes/28549_909191/anexos/4540_910628///pge_cartilha_principios_basicos.pdf?v=1281. Acesso em: 2022 Janeiro 27.

- _____, SOF. Classificação da Receita. In: _____ **Manual Técnico do Orçamento - MTO Online**. 14. ed. Brasília: [s.n.], 2022. Cap. 3, p. s.n. Disponível em: https://www1.siof.planejamento.gov.br/mto/doku.php/mto2022:cap3#dokuwiki_top. Acesso em: 20 Junho 2022.
- _____, _____. CLASSIFICAÇÃO DA RECEITA. In: BRASIL; FEDERAL, S. D. O. **Manual Técnico de Orçamento**. [Online]: [s.n.], 2021a. Cap. 3, p. n.p. Disponível em: <https://www1.siof.planejamento.gov.br/mto/doku.php/mto2021:cap3#previsao>. Acesso em: 20 Novembro 2021.
- _____; _____. ESTIMATIVAS DE RECEITAS ORÇAMENTÁRIAS. In: BRASIL; FEDERAL, S. D. O. **Manual Técnico de Orçamento**. [Online]: [s.n.], 2021b. Cap. 7, p. n.p. Disponível em: <https://www1.siof.planejamento.gov.br/mto/doku.php/mto2021:cap7>. Acesso em: 26 Novembro 2021.
- _____, TCU. **Política pública em dez passos**. Tribunal de Contas da União. [S.l.]. 2021a.
- _____, _____. Política pública em dez passos, 2021. Disponível em: https://portal.tcu.gov.br/data/files/1E/D0/D4/DF/12F99710D5C6CE87F18818A8/Politica%20Publica%20em%20Dez%20Passos_web.pdf. Acesso em: 26 Novembro 2021.
- _____; IPEA. Taxa de câmbio comercial para compra: real (R\$) / dólar americano (US\$) - média. **Instituto de Pesquisa Econômica Aplicada**, 2022. Disponível em: <http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M>. Acesso em: 20 Julho 2022.
- BROWNLEE, J. **Long Short-Term Memory Networks With Python: Develop Sequence Prediction Models With Deep Learning**. ebook. ed. [S.l.]: Machine Learning Mastery, 2017.
- _____, J. **Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python**. [S.l.]: Machine Learning Mastery, 2019.
- _____, J. B. **Introduction to Time Series Forecasting with Python: How to prepare data and develop models to predict the future**. ebook. ed. [S.l.]: Machine Learning Mastery, 2020.
- BRUCE, P.; BRUCE, A. **Estatística Prática para Cientistas de Dados: 50 conceitos essenciais**. Tradução de Luciana Ferraz. Rio de Janeiro-RJ: Alta Books, 2019. 320 p. ISBN ISBN978-85-508-1300-4 E-Book.
- CALVO, B.; SANTAFÉ, G. scmap: Statistical Comparison of Multiple Algorithms in Multiple Problems. **The R Journal**, 8, n. 1, Agosto 2016. 248-256. Disponível em: <https://journal.r-project.org/archive/2016/RJ-2016-017/index.html>. Acesso em: 15 dezembro 2022.

- CARMO, M. M. D.; BOLDT, F. D. A.; KOMATI, K. S. **Previsão de receitas de ICMS do estado do Espírito Santo através de seleção de características em cascata e técnicas de Aprendizado de Máquina**. Anais do XVI Encontro Nacional de Inteligência Artificial de Aprendizagem de Máquina. Salvador/BA: [s.n.]. 2019. p. 118-129.
- CHAPMAN, P. et al. **CRISP-DM 1.0: Step-by-step data mining guide**. [S.l.]: [s.n.], 2000. Disponível em: <https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfbd2d5a72>. Acesso em: 10 Julho 2022.
- CHOLLET, F. Fundamentals of machine learning. In: CHOLLET, F. **Deep Learning with Python**: MEAP04. 2ª. ed. Shelter Island, NY: Manning, 2021. Cap. 5, p. 130-162. Disponível em: <https://livebook.manning.com/#!/book/deep-learning-with-python-second-edition/discussion>. Acesso em: 18 Agosto 2022. To comment go to liveBook.
- CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS (CNM). **Coletânea Gestão Pública Municipal**. [S.l.]: [Online], v. 1, 2016. Disponível em: www.cnm.org.br. Acesso em: 30 Outubro 2021.
- COSTA, M. M. **Execução Orçamentária, Financeira e Contábil**. Curso Integrado em Administração Orçamentária, Financeira, Contábil e Controle no Serviço Público. São Luís: EsafiEscola. 2019. p. 222.
- DATA SCIENCE ACADEMY. Machine Learning – Guia Definitivo – Parte 3. In: ACADEMY, D. S. **Deep Learning Book**. [S.l.]: [s.n.], 2022. Cap. 93. Disponível em: <https://www.deeplearningbook.com.br/machine-learning-guia-definitivo-parte-3/>. Acesso em: 20 Agosto 2022. ebook on-line.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **The Journal of Machine learning research**, Ljubljana, Slovenia, 2006. 1-30. Disponível em: <https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf>. Acesso em: 26 Fevereiro 2022.
- DUARTE, G. J. et al. **Guia brasileiro de análise de dados: armadilhas & soluções**. Brasília/DF: Enap, 2021. 251 p. ISBN ISBN: 978-65-87791-25-8.
- DUBOUE, P. **The art of feature engineering: essentials for machine learning**. New York, NY: Cambridge University Press, 2020. ISBN ISBN 9781108671682 (epub).
- FÁVERO, L. P.; BELFIORE, P. Modelos de regressão simples e múltipla. In: _____ **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata**. 1. ed. Rio de Janeiro: Elsevier, 2017. Cap. 12, p. 511-609.

- FIIRST, C. et al. 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. Disponível em: <https://online.unisc.br/seer/index.php/cepe/article/view/12407>. Acesso em: 7 Outubro 2021.
- GÉRON, A. **Mãos à Obra Aprendizado de Máquina com Scikit-Learn & TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes**. Tradução de Rafael Contatori. Rio de Janeiro: Alta Books, 2019. 576 p.
- _____, _____. _____. _____. _____. 2. ed. Rio de Janeiro: Alta Books, 2021. 576 p.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo-SP: Atlas, 2017.
- GONTIJO, V.; GREGGIANIN, E.; MENDES, G. R. Orçamento da União: princípios orçamentários. **Câmara dos Deputados - Curso sobre Processo Orçamentário**, s.d. Disponível em: <https://www2.camara.leg.br/orcamento-da-uniao/cidadao/entenda/cursopo/principios.html>. Acesso em: 10 Maio 2022.
- HARRISON, M. **Machine Learning guia de referência rápida: trabalhando com dados estruturados em Python**. Tradução de Lúcia A. Kinoshita. São Paulo/SP: Novatec, 2020. ISBN ISBN: 978-85-7522-818-0.
- HART, M. et al. Conceitos básicos para consumidores do serviço do Power BI. **Documentação do Power BI**, 14 Julho 2022. Disponível em: <https://docs.microsoft.com/pt-br/power-bi/consumer/end-user-basic-concepts>. Acesso em: 20 Julho 2022.
- HE, X. et al. **Practical Lessons from Predicting Clicks on Ads at Facebook**. Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (ADKDD'14). New York, NY, USA: Association for Computing Machinery. 2014. p. 1-9.
- HIGGINBOTHAM, J. **Principles of Web API Design: delivering value with APIs and microservices**. Boston: Pearson Addison-Wesley, 2022.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 2. ed. Melbourne, Australia: OTexts, 2018. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 2021 Dezembro 2021.
- IZBICKI, R.; SANTOS, T. M. D. I Regressão: Introdução. In: IZBICKI, R.; SANTOS, T. M. D. **Aprendizado de máquina: uma abordagem estatística (e-book)**. São Carlos, SP: Câmara Brasileira do Livro, 2020. Cap. 1, p. 3-24. ISBN ISBN 978-65-00-02410-4.
- KATSIKOPOULOS, K. V.; DURBACH, I. N.; STEWART, T. J. When should we use simple decision models? A synthesis of various research strands. **Omega**, Dezembro 2018. 17-25.

- KORSTANJE, J. **Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR**. Maisons Alfort, France: Apress, 2021. 294 p.
- _____. Model Evaluation for Forecasting. In: KORSTANJE, J. **Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR**. Maisons Alfort, France: Apress, 2021. Cap. 2, p. 21-42.
- LAGO, K.; ALVES, L. Fundamentos do Business Intelligence, Introdução ao Power BI. In: LAGO, K.; ALVES, L. **Dominando o Power BI**. 2. ed. São Paulo: DATAB Inteligência e Estratégia, 2019. Cap. 1-2, p. 20-25.
- LAKSHMANAN, V.; ROBINSON, S.; MUNN, M. Reproducibility Design Patterns. In: LAKSHMANAN, V.; ROBINSON, S.; MUNN, M. **Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps**. 2. ed. Sebastopol, CA: O'Reilly, 2021. Cap. 6, p. 249-318.
- LIMA, L. L. et al. Planejamento governamental nos municípios brasileiros: em direção a uma agenda de pesquisa. **Cadernos EBAPE.BR**, Rio de Janeiro/RJ, 18, Abril-Junho 2020. 323-335. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/cadernosebape/article/view/77989>. Acesso em: 28 Novembro 2021.
- LOSSO, R. D. Deflacionamento: Nota Técnica. **Informações fipec**, julho 2020. 18-24. Disponível em: <https://www.fipec.org.br/pt-br/publicacoes/bif>. Acesso em: 23 outubro 2022.
- MAKRIDAKIS, S.; SPILOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: 100,000 time series and 61 forecasting methods. **International Journal of Forecasting**, 36, January-March 2020. 54-74. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207019301128>. Acesso em: 16 Junho 2022.
- _____. M5 accuracy competition: Results, findings, and conclusions. In: **International Journal of Forecasting**, 11 January 2022. 1-19. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207021001874>. Acesso em: 15 Maio 2022.
- MARANHÃO, GOVERNO DO ESTADO DO. **Lei Ordinária Estadual nº 5.599 de 24 de dezembro de 1992**. Dispõe sobre a distribuição das parcelas do Imposto sobre Circulação de Mercadoria e sobre Prestações de Serviços de Transportes Interestadual e Intermunicipal de Comunicações ICMS, pertencentes aos Municípios, e dá outras providências. São Luís, 1992.
- MARTÍNEZ-PLUMED, F. et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. **IEEE Transactions on Knowledge and Data**

- Engineering**, v. 33, p. 3048-3061, 01 Agosto 2021. ISSN 8. Disponível em: <https://ieeexplore.ieee.org/document/8943998>. Acesso em: 08 Junho 2022.
- MIKESELL, J. L. Often Wrong, Never Uncertain: Lessons from 40 Years of State Revenue Forecasting, 28 Maio 2018. Disponível em: <https://doi.org/10.1111/puar.12954>. Acesso em: 07 Outubro 2021.
- MINITAB. Análise de regressão múltipla: Use o R-quadrado ajustado e o R-quadrado predito para incluir o número correto de variáveis: Editor Minitab. **Blog da Minitab**, 07 maio 2019. Disponível em: <https://blog.minitab.com/pt/analise-de-regressao-multipla-use-o-r-quadrado-ajustado-e-o-r-quadrado-predito-para-incluir-o-numero-correto-de-variaveis>. Acesso em: 3 janeiro 2023.
- NAIBERT, N. T. **Orçamento público: metodologia alternativa para previsão de receitas municipais**. Dissertação (Mestrado em Administração Pública) - FGV. Rio de Janeiro, p. 66. 2021.
- NIELSEN, A. Financial Applications. In: NIELSEN, A. **Practical Time Series Analysis: Prediction with Statistics and Machine Learning**. 2. ed. Sebastopol, CA: O'Reilly, 2020. Cap. 14, p. 403-424.
- NOCH, M. Y.; PATTIASINA, V.; PURWANTO, H. A. Elasticity Measurement of Local Taxes and Charges In Forecast Of Own-Source Revenue (PAD) Of Provincial Government In Indonesia, 11 Janeiro 2018. Disponível em: <https://ojs.unud.ac.id/index.php/Akuntansi/article/view/35361>. Acesso em: 07 Outubro 2021.
- PAMPLONA, E. et al. Desempenho do Modelo ARMA na Previsão das Receitas Orçamentárias dos Municípios do Estado do Paraná, Jan-mar 2019. Disponível em: <https://periodicos.ufv.br/apgs/article/view/5257>. Acesso em: 7 Outubro 2021.
- PARMEZAN, A. R. S.; SOUZA, V. M. A.; BATISTA, G. 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. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025519300945>. Acesso em: 22 Junho 2022.
- PAVLYUK, D. Random Forest Variable Selection for Sparse Vector Autoregressive Models. In: VALENZUELA, O., et al. **Theory and Applications of Time Series Analysis**. [S.l.]: Springer, 2020. p. 3-17.
- PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python - 1.13. Seleção de recursos. **Journal of Machine Learning Research**, 2007. ISSN 12. Disponível em: <https://scikit->

- [learn.org/stable/modules/feature_selection.html#feature-selection](https://scikit-learn.org/stable/modules/feature_selection.html#feature-selection). Acesso em: 25 Outubro 2021.
- _____. _____. - 1.11. Ensemble methods. **Journal of Machine Learning Research**, 2011. 2825-2830. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html#ensemble>. Acesso em: 25 Outubro 2021.
- _____. _____. Regression Metrics. **Journal of Machine Learning Research**, 2011b. 2825--2830. Disponível em: https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics. Acesso em: 25 Dezembro 2021.
- PETROPOULOS, F. et al. Forecasting: theory and practice. **International Journal of Forecasting**, 38, 2022. 705-871. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207021001758>. Acesso em: 12 Agosto 2022. ISSN: 0169-2070.
- PLOTNIKOVA, V.; DUMAS, M.; MILANI, F. P. Adapting the CRISP-DM Data Mining Process: A Case Study in the Financial Services Domain. **Research Challenges in Information Science. RCIS 2021. Lecture Notes in Business Information Processing**, Limassol, Cyprus, 415, 08 Maio 2021. 55-71. Disponível em: <https://www.semanticscholar.org/paper/Adapting-the-CRISP-DM-Data-Mining-Process%3A-A-Case-Plotnikova-Dumas/a4e44d5ce058e6b815631947db23044a22f8b114>. Acesso em: 10 Julho 2022.
- PROQUE, A. L. et al. Efeitos econômicos da nova política tributária sobre a demanda de combustíveis para a economia brasileira. **XXVI Encontro Regional de Economia**, online, 27 e 28 Outubro 2021. Disponível em: https://www.anpec.org.br/nordeste/2021/submissao/arquivos_identificados/081-21edd10b419ba536b9c853236f7c88ec.pdf. Acesso em: 22 Março 2022.
- RAFFERTY, G. The History and Development of Time Series Forecasting. In: RAFFERTY, G. **Forecasting Time Series Data with Facebook Prophet: Build, improve, and optimize time series forecasting models using the advanced forecasting tool**. 2ª. ed. Birmingham-Mumbai: Packt, 2021. Cap. 1, p. 15-27.
- RODRIGUES, R. R. F. et al. **Séries Temporais de Holt-Winters na análise preditiva da arrecadação do ICMS no estado do Rio Grande do Norte**. Anais do XXVII Congresso Brasileiro de Custos - ABC. [S.l.]: [s.n.]. 9-11 Novembro 2020. p. 14.
- SAMPAIO, I. G. et al. Avaliação de modelos de predição e previsão construídos por algoritmos de aprendizado de máquina em problemas de Cidades Inteligentes. In: DAVID, J. M. N., et al. **Tópicos em sistemas de informação [recurso eletrônico]: minicursos do XIV**

Simpósio Brasileiro de Sistemas de Informação. 15. ed. Aracajú/SE: Sociedade Brasileira de Computação, 2019. Cap. 4, p. 81-113.

SÃO LUÍS (MA). **Lei Complementar nº 6.878**. Dispõe sobre a estimativa de receita e fixa a despesa do Município de São Luís para o Exercício Financeiro de 2021, e dá outras providências. São Luís, MA, 31 Dezembro 2020. 110. Disponível em: <https://transparencia.saoluis.ma.gov.br/pagina/3043/>. Acesso em: 15 Julho 2022.

_____. **Decreto Municipal nº 56.835**. Estabelece Normas para a Programação e Execução Orçamentária e Financeira dos Orçamentos Fiscal e da Seguridade Social do exercício 2021, e dá outras providências. São Luís, 4 Janeiro 2021a. Disponível em: <http://sistemas.semاد.saoluis.ma.gov.br/easysearch/cachedownloader?collection=default&docId=ec69d2c164d0ccc69a4342ab3d463c9824f5ff02&fieldName=Visualizar&extension=bht.zip#q=execu%C3%A7%C3%A3o%20execu%C3%A7%C3%A3o%20financeira%202021>. Acesso em: 22 Outubro 2021.

_____. Receitas Total e detalhada. **Portal da Transparência**, 2021b. Disponível em: https://saoluis.giap.com.br/apex/saoluis/wwv_flow.accept. Acesso em: 15 Setembro 2021.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A Systematic Literature Review on Applying CRISP-DM Process Model. **Procedia Computer Science**, 181, 2021. 526-534. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>. Acesso em: 8 Junho 2022.

SEKAR, M. **Machine Learning for Auditors: Automating Fraud Investigations Through Artificial Intelligence**. Calgary, AB, Canada: Apress, 2022. 241 p.

SILVA, A. C. B. D. et al. Utilização de Séries Temporais na Previsão de Arrecadação de ICMS no Estado do Pará, 2016. Disponível em: <https://periodicos.ufpe.br/revistas/SUCC/article/view/22988>. Acesso em: 30 Setembro 2021.

SILVA, A. M. D. et al. Orçamento Público, [Como Prever as Receitas?], 19 a 31 Outubro 2018. Disponível em: <http://www.poncedaher.net.br/egen/sites/default/files/Or%C3%A7amento%20P%C3%BAblico%2C%20Como%20Prever%20as%20Receitas.pdf>. Acesso em: 7 Outubro 2021.

SILVA, L. F. D. et al. Correlação das Variáveis Socioeconômicas e Ambientais com royalties Petrolíferos e CFEM Municipais, 28 Setembro 2017. Disponível em: <https://www.scielo.br/j/floram/a/DqydgD5hL8pCMdFW4CHdLDg/?lang=pt>. Acesso em: 28 Setembro 2021.

- SOUSA, K. R. et al. Previsão de Receitas tributárias: uma análise do icms no ceará por meio de séries temporais, Junho 2019. Disponível em: https://doi.org/http://dx.doi.org/10.26853/Refas_ISSN-2359-182X_v05n05_00. Acesso em: 6 Outubro 2021.
- TERPILOWSKI, M. Docs: Introduction. **Posthocs Scikit**, 2021. Disponível em: <https://scikit-posthocs.readthedocs.io/en/latest/intro/>. Acesso em: 12 Agosto 2022.
- TSAY, R. S. **Multivariate Time Series Analysis: with R and Financial Applications**. 1. ed. Chicago: Wiley, 2014.
- WADE, C. **Hands-On Gradient Boosting with XGBoost and scikit-learn: perform accessible machine learning and extreme gradient boosting with Python**. Birmingham - Mumbai: Packt, 2020. ISBN ISBN 978-1-83921-835-4. ebook.
- WAZLAWICK, R. S. **Metodologia de pesquisa para ciência da computação**. 3. ed. Rio de Janeiro-RJ: LTC, 2021.
- WIRTH, R.; HIPPI, J. Crisp-dm: towards a standard process model for data mining, 2000. 11. Disponível em: <https://www.semanticscholar.org/paper/Crisp-dm%3A-towards-a-standard-process-model-for-Wirth-Hipp/48b9293cfd4297f855867ca278f7069abc6a9c24>. Acesso em: 25 Outubro 2021.
- XGBOOST DOCUMENTATION. Python API reference. **XGBoost Python Package**, 2021. ISSN Revision 16bca5d4. Disponível em: https://xgboost.readthedocs.io/en/latest/python/python_api.html. Acesso em: 25 Outubro 2021.

APÊNDICES

APÊNDICE A: QUESTIONÁRIO DE DISSERTAÇÃO DE MESTRADO APLICADO ÀS CAPITAIS DA ABRASF

13/06/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google



Questionário Dissertação de Mestrado

Perguntas

Respostas

7

Condições

Seção 1 de 2

Questionário Dissertação de Mestrado

O presente questionário trata da investigação, no âmbito da Dissertação de Mestrado Profissional em Engenharia da Computação e Sistemas, do Programa de Pós-Graduação em Engenharia da Computação e Sistemas, da Universidade Estadual do Maranhão (UEMA) realizada por Cláudia Patrícia Silva Pimentel. Cumpre-se registrar que os resultados obtidos serão utilizados apenas para fins acadêmicos e seu resultado será encaminhado ao seu e-mail.

E-mail *

E-mail válido

Este formulário está coletando e-mails. [Alterar configurações](#)

Qual é a Cidade em que você trabalha para o Governo Municipal? *

1. Aracaju (SE)
2. Belém (PA)
3. Belo Horizonte (MG)
4. Boa Vista (RR)
5. Brasília (DF)
6. Campo Grande (MS)
7. Cuiabá (MT)
8. Curitiba (PR)
9. Florianópolis (SC)
10. Fortaleza (CE)

https://docs.google.com/forms/d/1vKJqUSTolyzWVul1NpsBSzmMbcJRtp_qP_ZeZjMeEIE/edit

1/5

13/05/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google

11. Goiânia (GO)
12. João Pessoa (PB)
13. Macapá (AP)
14. Maceió (AL)
15. Manaus (AM)
16. Natal (RN)
17. Palmas (TO)
18. Porto Alegre (RS)
19. Porto Velho (RO)
20. Recife (PE)
21. Rio Branco (AC)
22. Rio de Janeiro (RJ)
23. Salvador (BA)
24. São Luís (MA)
25. São Paulo (SP)
26. Teresina (PI)
27. Vitória (ES)

Qual o seu vínculo com o Governo Municipal? *

- ☐ Efetivo
- ☐ Comissionado

13/05/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google

☐ Serviço Prestado☐ Terceirizado**Há quantos anos trabalha no Governo Municipal?**

1. Há menos de 1 ano
2. De 1 a 3 anos
3. De 4 a 7 anos
4. De 8 a 10 anos
5. mais de 11 anos

Qual o seu curso de formação?

- ☐ Ciências Contábeis
- ☐ Economia
- ☐ Administração
- ☐ Direito
- ☐ Engenharia (e variações)
- ☐ Computação (variações em Tecnologia)
- ☐ Outros...

Em sua capital existe algum setor, assessoria ou departamento destinado ao monitoramento e previsão das Receitas de Transferências Constitucionais e legais? *

☐ Sim☐ Não

13/05/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google

Após a seção 1 Continuar para a próxima seção

Seção 2 de 2

Previsão de Receitas

Nesta seção as questões buscam identificar informações de como são feitas as previsões das receitas, especialmente as Transferências Constitucionais e legais em atendimento ao que prescreve o art. 12 da LRF, "ipsis litteris".

"Art. 12. As previsões de receita observarão as normas técnicas e legais, considerarão os efeitos das alterações na legislação, da variação do índice de preços, do crescimento econômico ou de qualquer outro fator relevante e serão acompanhadas de demonstrativo de sua evolução nos últimos três anos, da projeção para os dois seguintes àquele a que se referirem, e da metodologia de cálculo e premissas utilizadas.

[...]

§ 3º O Poder Executivo de cada ente colocará à disposição dos demais Poderes e do Ministério Público, no mínimo trinta dias antes do prazo final para encaminhamento de suas propostas orçamentárias, os estudos e as estimativas das receitas para o exercício subsequente, inclusive da

Seu município tem Convênio com Universidades para capacitação ou cooperação técnica em previsão de receitas?

☐ Sim

☐ Não

Em seu Governo Municipal são oferecidos cursos de capacitação relacionados à Previsão de Receitas Públicas aos colaboradores responsáveis por esta tarefa?

☐ Sim

☐ Não

Caso a resposta da questão anterior seja "sim": Qual a frequência da oferta destes cursos

13/05/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google

☐ Anual

☐ Semestral

☐ não ofertado

Seguindo: São ofertados cursos de capacitação em estatística aos colaboradores?

☐ Sim

☐ Não

Para a tarefa de Previsão de Receitas Públicas, seus técnicos elaboram os estudos e memórias de cálculos do art. 12 da LRF com base em que técnicas: *

☐ Orientações do MCASP da STN

☐ Estatísticas

☐ Aprendizagem de Máquina (Inteligência Artificial)

☐ Terceirizam esta tarefa

Complementando a questão anterior: A mesma técnica é utilizada para todas as *

☐ Sim

☐ Não, utilizam-se técnicas diferentes para previsão de cada receita.

Há interesse da área técnica de previsão de receitas quanto à aplicação de técnicas de Inteligência Artificial? *

☐ Sim

☐ Não

13/05/2022 11:21

Questionário Dissertação de Mestrado - Formulários Google

Seu Município tem algum sistema que faz previsão de receitas?

☐ Sim

☐ Não

Caso tenham algum sistema que faz previsão de receitas, sua Capital disponibiliza-o para os demais Municípios?

☐ Sim

☐ Não

Obrigada por sua participação.

Descrição (opcional)



APÊNDICE B: Resumo geral das receitas

Receita	Descrição/Tipo repasse	Legislação principal	Cálculo do Índice		Cálculo da distribuição	* Periodicidade	
			Índice	Vigência		Tipo	Arrec. Crédito
FPM	Fundo de Participação dos Municípios, dividido em FPM principal, distribuído todos os meses, Emenda Constitucional 55 (EC-55) depositada em dezembro, EC-84 paga em julho e EC-112, em setembro (iniciando em 2022); Tipo: repasse constitucional;	Artigo 159, I, “b”, “d” e “c” da CF	$CIFPM = \frac{Pop_{cap} * \frac{1}{RdPer_{UF}}}{\sum_{Cap=1}^{27} Pop_{cap} * \frac{1}{RdPer_{UF}}}$	Janeiro	$AL_m = (AB_{IR} + AB_{PI}) - (D_{IR} + D_{PI})$ $VP_{FPM} = 22,5\% * AL_m$ $V_{cp} = 10\% * VP_{FPM} * CIFPM_{at}$ $V_{EC-55} = 10\% \left[\left(\sum_{m=1}^{11} AL_m \right) CIFPM_{at} 1\% + (AL_{dez} CIFPM_{an} 1\%) \right]$ $V_{EC-84} = 10\% \left[\left(\sum_{m=7}^{12} AL_m \right) CIFPM_{an} 1\% + \left(\sum_{m=1}^6 AL_m \right) CIFPM_{at} 1\% \right]$ $V_{ate6} = \left(\sum_{m=1}^6 AL_m \right) CIFPM_{at} 1\%$	Decendial (dias 10, 20 e 30)	10 dias 10 depois
IPI	Imposto sobre Produtos Industrializados de exportação; Tipo: repasse constitucional;	Artigo 159, II, §3º da CF	Usa o CIP _{UF} calculado pelo TCU IPM do ICMS		$AL_{IPIm} = AB_{PI} - D_{PI}$ $IP_{IUF} = 10\% * (AL_{IPIm})$ $IP_{Im} = IP_{IUF} * (CIP_{IUF} * 25\% * IPM)$		
ICMS	Imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual, intermunicipal e de comunicação; Tipo: repasse constitucional;	Artigo 158, “IV”, da CF	**IPM – calculado em junho p/faixa populacional e o VA declarado ao Estado ***IPM = 65%*VA + (25%*Lei Estadual + 10%*Educação)	Janeiro	$AL_m = AB_m - D_m$ $Arr_{dm} = 25\% * AL_m$ $ICMS_{dm} = Arr_{dm} * IPM$	Semanal (Toda Terça)	7 dias 7 dias depois
IPVA	Imposto sobre a Propriedade de Veículos Automotores; Tipo: repasse constitucional;	Artigo 158, inciso “III” da CF	50% da arrecadação líquida do IPVA de domicílio da placa do veículo automotor	não se aplica	$AL_{IPVAm} = AB_{IPVAm} - D_{IPVAm}$ $IPVAm = 50\% * AL_{IPVAm}$		
FEP	Fundo Especial de Petróleo; Tipo: repasse legal	Lei 7.453/19850	CIFPM, mesmo critério do FPM	Janeiro	FEP = 80%*Pa* CIFPM	Mensal	Mês anterior fim do mês
CFEM	Compensação Financeira pela Exploração Mineral; Tipo: repasse constitucional.	Artigo 20, § 1º, da CF	Índice calculado pela ANM em abril Proporção da produção por atividades do município	Maior	60% Produtores e 15% municípios afetados pelo Porto, Ferrovia, Dutos e Rejeitos, conforme norma.	Mensal	Mês anterior Até 2º decênio depois

Legendas:

$\frac{Pop_{cap}}{\sum Pop_{cps}}$ é a relação entre a população da Capital de Estado pelo total de todas as capitais;

$\frac{1}{RdPer_{UF}}$ é o inverso da renda per capita do Estado;

$\frac{PopCap}{\sum_{Cap=1}^{27} PopCap} * \frac{1}{RdPer_{UF}}$ é coeficiente da Capital de Estado obtido pela multiplicação dos 2 anteriores;

$\sum_{Cap=1}^{27} \left(\frac{PopCap}{\sum_{Cap=1}^{27} PopCap} * \frac{1}{RdPer_{UF}} \right)$ é o somatório dos coeficientes de todas as capitais;

AL_m é arrecadação líquida no mês;

AB_{IR} é a arrecadação bruta de Imposto de Renda (IR) no mês;

AB_{IPI} é a arrecadação bruta de IPI no mês;

D_{IR} são as deduções do IR no mês;

D_{IPI} são as deduções de IPI no mês;

VP_{FPM} é o valor dos 10% do FPM principal para ser distribuídos a todos os Municípios no mês;

$CIFPM_{at}$ é o coeficiente individual de participação na distribuição do FPM calculado pelo TCU para ano atual e a $CIFPM_{an}$ é do ano anterior;

V_{cp} é o valor a receber por Capital de Estado no FPM principal, resultado da multiplicação de 10% do VP_{FPM} com $CIFPM_{at}$;

$(\sum_{m=1}^{11} AL_m) CIFPM_{at} 1\%$ é o cálculo da cota no ano atual da EC-55 sob 1% da soma da AL_m de janeiro a novembro do ano atual multiplicado pelo $CIFPM_{at}$;

$AL_{dez} CIFPM_{an} 1\%$ é da AL_{dez} de dezembro do ano anterior multiplicado por 1%;

V_{EC-55} é valor da cota de 10% de cada Capital de Estado da EC-55 depositada em dezembro;

V_{at6} é 1% da soma da AL_m de janeiro a junho do ano atual;

e V_{EC-84} é o da EC-84 paga em julho;

AL_{IPM} é a arrecadação líquida do IPI no mês obtida pela subtração do AB_{IPI} pelo D_{IPI} ;

IPI_{UF} é o valor total a ser distribuídos entre os Estados sob 10% da AL_{IPM} ;

IPI_m é a cota de cada Cidade obtida com o cálculo de 25% do IPI_{UF} multiplicado pelo Coeficiente do Estado ($CIFIUF$) com o IPM do referido Município;

Arr_{dm} é o montante de distribuição da arrecadação do ICMS pelo cálculo de 25% da arrecadação líquida (AL_m);

$ICMS_{dm}$ é o crédito da cota do ICMS municipal, calculado pela multiplicação do Arr_{dm} com o IPM da Cidade;

AL_{IPVAm} é a arrecadação líquida de IPVA no mês obtida com a subtração da arrecadação bruta (AB_{IPVAm}) pelas deduções do imposto (D_{IPVAm});

$IPVAm$ é o repasse do IPVA de cada Município retornando 50% da AL_{IPVAm} ;

FEP é o montante resultante do 80% da parcela (Pa) que a legislação dispõe para este fundo (ver legislação) pelo $CIFPM$;

A CFEM tem um cálculo mais complexo, pois depende das normas publicados pela Agência Nacional de Mineração;

* A periodicidade, para os repasses dos Governos Estaduais, pode variar em cada Unidade da Federação, definido por lei ou norma;

** O cálculo do IPM é definido por cada Estado, mas a Emenda Constitucional 108/2020, alterou a legislação, exigindo adequação em todas as Unidades da Federação;

*** O IPM, dependendo da nova legislação, calculará 65% do Valor Adicional (VA), mais 35% da distribuição definida em Lei Estadual e, mas sendo obrigatório o mínimo de 10% para índices educacionais;

APÊNDICE C: Relatório inicial de coleta de dados da primeira iteração com a Capital São Luís

Trata o presente relatório do resumo da tarefa de “Coleta inicial dos dados” da Metodologia CRISP-DM da Previsão das Receitas Públicas. Todos os dados obtidos atendem aos ditames da Lei Nº 12.527, de 18 de novembro de 2011, Lei de acesso à informação no setor público brasileiro, já que elas estão disponíveis nos sites de Transparência de Entidades do setor público (BRASIL, 2011).

FERRAMENTA

Para a coleta e unificação do *Dataset* foi utilizado o *software Power BI Desktop*, versão: 2.96.901.0 64-bit, ferramenta gratuita da *Microsoft*. Incluindo os *scripts* em *Python* para algumas tabelas, mas todos incluídos na ferramenta.

CONFIGURAÇÕES INICIAIS

Para acessar a API dos dados abertos do Portal da Transparência do Governo Federal será necessário fazer um cadastro no sistema do Gov.br, autorizando uma chave de acesso (token), a qual é fornecida pelo e-mail cadastrado na conta, melhores esclarecimento no site⁴⁶.

Quadro 26- Lista de variáveis globais

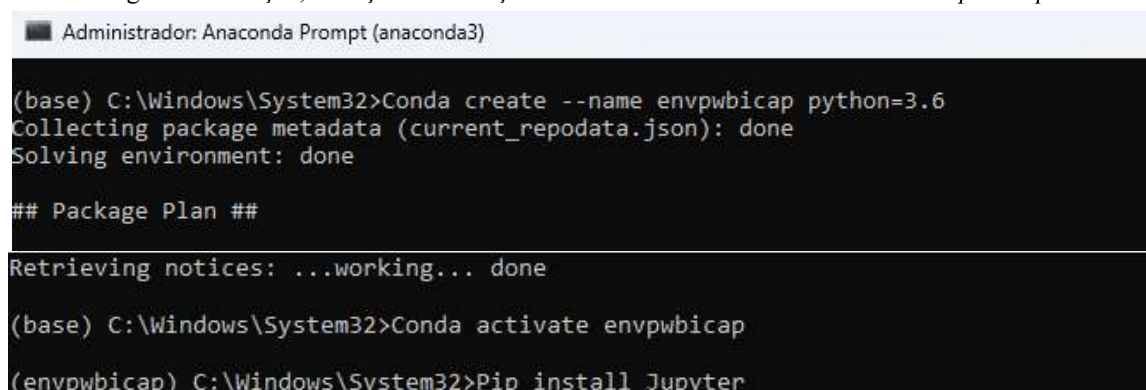
Denominação	Tipo	Descrição da variável global	Tabelas dependentes
Caminho	Text	Determina o caminho de um arquivo físico que por algum motivo não está sendo adquirido pela internet.	
MenorData	Date	Pode ser incluída manualmente ou por código, o qual verifica nas receitas das capitais a maior data entre as menores datas dos dados disponíveis.	d_Calendario, f_Dolar_Comercial, f_SistNacPesqCustIndConstCivil
DataAtual	Text	Calcula a data atual do sistema operacional convertida para texto, formato “MM-dd-yyyy”.	MaiorData, f_Dolar_Comercial, f_SistNacPesqCustIndConstCivil
MaiorData	Date	Também pode ser determina a maior data, configurada para o dia 15 de dezembro do ano da variável <i>DataAtual</i> . Podendo ser alterada manualmente.	d_Calendario, Mese_MenorAno
VarPerCalendario	Text	Em formato de texto cria uma lista para as consultas das API do IBGE por Mês e Ano, utilizando a lista da coluna “[AnoMes]” da tabela oculta “VarPerCalendario”. Fica oculta para o Painel.	f_SistNacPesqCustIndConstCivil
AnoMesAtual	Text	Pela data atual do sistema extrai ano e mês anterior ao atual para utilizar nas buscas nas APIs.	VarPerCalendario, f_SistNacPesqCustIndConstCivil

Fonte: Elaborado pela Autora (2023).

⁴⁶ Portal da Transparência. API de dados – Cadastro. Disponível em: <https://www.portaldatransparencia.gov.br/api-de-dados/cadastrar-email>. Acessado em 11 outubro de 2022. Um exemplo, na linguagem *Python* pode ser encontrado no *Jupyter Notebook* disponível em: <https://colab.research.google.com/drive/1IRbpw9NfxiTrGTecSuh4QNzX0hUjJV8c?usp=sharing#scrollTo=WII6A77Z2QMH>.

Para as configurações iniciais foram criadas as variáveis globais que serão utilizadas em várias tabelas, por isso foram concentradas no grupo de configurações, sendo elas elencadas no Quadro 26.

Figura 7 - Criação, ativação e instalação de bibliotecas no ambiente virtual “envpwbicap”



```

Administrador: Anaconda Prompt (anaconda3)

(base) C:\Windows\System32>Conda create --name envpwbicap python=3.6
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

Retrieving notices: ...working... done

(base) C:\Windows\System32>Conda activate envpwbicap
(envpwbicap) C:\Windows\System32>Pip install Jupyter
  
```

Uma configuração muito importante a ser feita é a criação de um ambiente virtual para o *Python*, pois os *web scraping* e *crawling* foram executados com *scripts* dessa linguagem, uma vez que na linguagem “M” do *Power Query* estava sendo consumida muita memória, tornando o processo mais demorado. Neste sentido, sugere-se criar um ambiente virtual no *prompt* do Anaconda, denominado de “envpwbicap”, além de configurar o *Power BI Desktop* para utilizar este ambiente⁴⁷, conforme Figura 7.

Devem ser instaladas neste ambiente, inicialmente, os pacotes:

- “Jupyter” foi a IDE definida para utilização neste trabalho;
- “pandas” (versão 1.3.5), contém ferramentas de estruturas de dados;
- “numpy” (versão 1.21.6), o qual fornece um grupo de funções e operações para executar cálculos numéricos e é instalada junto com o “pandas”;
- “matplotlib” (versão 3.2.2) um pacote para criação e manipulação de gráficos que também é instalado com o “pandas”;
- “HTMLParser” (versão 0.0.2) faz a identificação de *tags* ou dados nos códigos HTML das páginas da *internet*;
- “html5lib” (versão 1.1) é uma biblioteca *Python* para análise de HTML;
- “request” (versão 2.27.1) é uma biblioteca para manipulação de HTTP para *Python*;

⁴⁷ Detalhes para instalação, criação e execução de ambientes virtuais no Power BI Desktop consultar o site da Microsoft, disponível em: A questão das credenciais: <https://learn.microsoft.com/pt-br/power-bi/connect-data/desktop-python-scripts>. Acessado em 08 de outubro de 2022.

- “*collections-extended*” (versão 2.0.2) módulo que implementa tipos de dados de coleções para suporte aos que já estão embutidos no *Python*;
- “*os*” é o pacote que manipula arquivos do sistema operacional;
- “*zipfile36*” (versão 36-0.1.3) disponibiliza ferramentas para trabalhar com arquivos compactados;
- “*openpyxl*” (versão 3.0.10) é uma biblioteca para escrever e ler arquivos no formato Office Open XML e Excel2010;
- “*beautifulsoup4*” (versão 4.11.2) é um pacote que auxilia na manipulação dos códigos de HTML das páginas consultadas;
- “*wget*” (versão 3.2) para fazer download de dados de páginas da *internet*;
- “*certifi*” (versão 2022.9.24) possui uma coleção de certificados raiz do Mozilla para validar a confiabilidade dos certificados SSL;
- “*urllib3*” (versão 1.24.3) é uma biblioteca que possui recursos de cliente HTTP para conectar a um website e recuperar o conteúdo dele;

COLETA DOS DADOS

A primeira coleta foi iniciada em setembro de 2021, com as receitas do APÊNDICE B, que foram codificadas como: 1=CFEM, 2=IPI, 4=FEP, 5=FPM, 6=ICMS, 7=IPVA e 8=CFEM AFETADOS. Destas, apenas o ITR foi descartado, pela sua menor relevância financeira. Devido às inúmeras fontes de dados e formatos, elas foram transformadas em 27 tabelas, quando do seu carregamento na ferramenta, com suas respectivas variáveis resumidas no Quadro 27.

Quadro 27 - Resumo descritivo das tabelas por fonte, formato e variáveis da coleta inicial de dados

Nº	Tabela	Descrição	Formato	Local	Variáveis	Fonte de Coleta
01	f_Programaca _Financeira	Contém os dados da execução financeira do Município, da arrecadação em sua publicação contábil, publicados no portal da transparência do Governo Municipal, desta tabela retira-se o Valor Executado, a variável alvo.	Arquivo “PDF” e “CSV”	São Luís	Data, DATA; Cod_Nivel04, DESC_NV_04, Classif/Codigo, Espec_Receita, Fonte, “%”, Tipo, RCL, Valor, Ano, Mes, RECEITA, RECEITA_CODIGO, NOVA_CLASSIF, RECEITA_LRF, DUODECIMO, Nivel_RCL, CodigoRCL, Receita_RCL_STN, Rubrica, NrMes, Destinacao, CodNivelRCL, Cód_Desc	Portal Transparência São Luís
02	f_Dolar_Come rcial	Contém os dados taxa de câmbio é o preço de uma moeda estrangeira medido em unidades ou frações (centavos) da moeda nacional, refletindo o custo de uma moeda em relação à outra.	Planilha eletrônica “XLS”	Aplicado ao Brasil	Data; DolarComercial	Banco Central do Brasil
03	f_Arrec_Gov_ Estado	Reúne os dados da arrecadação total do Governo do Estado, neste caso. Este conjunto de dados foi carregado, limpo e transformado em um Relatório de BI exclusivo.	Site, páginas “HTML”	Estado MA	DATA, ICMS, ST_EXTERNO, ST_INTERNO, ICMS_SIMP_NACIO, ICMS_DIV_ATIV_TRIB, IPVA, ICMS_TOTAL	Portal Transparência Governo do Estado do Maranhão
04	f_Arrec_Gov_ MA_Municipi os	Agrupamento das pesquisas feitas no site do Governo do Estado, Portal da Transparência para a arrecadação de tributos por Município. Este conjunto de dados foi carregado, limpo e transformado em um Relatório de BI exclusivo.	Arquivo “CSV”	São Luís	Data”, Municípios, ICMS_DvAtTrb_Muns, IPVA_Muns, TOTAL_ArrecMun, ICMS_Muns	Portal Transparência Governo do Estado do Maranhão
05	f_IndGeralPre coMercado	Recolhe o Índice Geral de Preços - Mercado. Indicador mensal do nível de atividade econômica do país, englobando seus principais setores.	Site, páginas “HTML”	Aplicado ao Brasil	DATA, IGPM_ACUMULADO, IGPM	Fundação Getúlio Vargas
06	f_Importacao_ Exportacao_B rasil	Tabelas com as séries históricas dos Totais mensais dos valores das exportações e importações com os valores em dólar.	Planilha eletrônica “XLS”	Brasil	DATA, Exportacao, Importacao, Saldo, Corrente, Var_Export_MesAnoAnt, Var_Import_MesAnoAnt,	Portal da Transparência -

Nº	Tabela	Descrição	Formato	Local	Variáveis	Fonte de Coleta
					Var_Saldo_MesAnoAnt, Var_Corrente_MesAnoAnt	Dados Abertos ou COMEXSTAT
07	f_IndNacPrec oConsumidor	Tabela do IBGE do Índice Nacional de Preços ao Consumidor nº 1736- INPC - Série histórica com número-índice, variação mensal e variações acumuladas em 3 meses, em 6 meses, no ano e em 12 meses (a partir de abril/1979).	JSON da API do IBGE	Brasil	DATA, INPC, INPC_Nr_Indice, INPC_AC3M, INPC_AC6M, INPC_AC12M, INPC_AC_ANO	IBGE - API
08	f_IndPrecoProdutor	Tabela do IBGE do Índice de Preços ao Produtor nº 6903- Índice de Preços ao Produtor, por tipo de índice, indústria geral, indústrias extrativas e indústrias de transformação e atividades (dezembro de 2018 = 100)	JSON da API do IBGE	Brasil	DATA, IPP, IPP_Nr_Indice, IPP_M_M-12	IBGE - API
09	f_PIB_Vlr_Corrente_Brasil_Milhoes	Unificação dos dados da Tabela 5938 - Produto interno bruto a preços correntes, impostos, líquidos de subsídios, sobre produtos a preços correntes e valor adicionado bruto a preços correntes total e por atividade econômica, e respectivas participações - Referência 2010 e as variações anual, trimestral e mensal da Tabela do PIB. Esta tabela foi elaborada em um relatório do <i>Power BI Desktop</i> exclusiva para depois ser juntada na geral	Planilha eletrônica “XLS”, Arquivo “CSV” e JSON da API do IBGE	Brasil, Estados e Municípios	DATA, PIB_MensalBrasil, Ano, PIB_Muni, RndPrCap_Mun, VA_BrtAdm, VA_BrtAgro, VA_BrtIndust, VA_BrtServExclAdm, VA_BrtTotal, VA_ImpostosProd, PrclnvRndPrCpMn, FtInvRndPrCap_Mun, PIB_CorrTotal, PercPIB_Total, PIB_Municipal, RndPrCap_Mun, VA_BrutoAdministracao, VA_BrutoAgropecuaria, VA_BrutoIndustria, VA_BrutoServicosExcluiAdm, VA_BrutoTotal, VA_ImpostosProdutos	IBGE IBGE - SIDRA
10	f_SELIC	Histórico das taxas de juros fixadas pelo Copom e evolução que são registradas diariamente no Sistema Especial de Liquidação e de Custódia (Selic), consiste na taxa média cobrada em negociações com títulos emitidos pelo Tesouro Nacional, uma taxa de juros de referência para os demais juros da economia.	Site, páginas “HTML”	Brasil	Data, Taxa_SELIC_Periodo,	Banco Central do Brasil
11	d_Populacao_Estimativa	Dados do Censo e Estimativa da população do IBGE. Esta tabela foi elaborada em um relatório do <i>Power BI Desktop</i> exclusiva para posterior agregação a base geral	JSON da API do IBGE	Brasil, Estados e Municípios	Data, PopulacaoCapitais, FtPopulacao, POPULACAO, Per_EntreCapitais, PerPopulacaoUF, DataCIFPM	IBGE IBGE - API
12	d_IPCA	Agrupar o Índice Nacional de Preços ao Consumidor Amplo 15 (IPCA) da Tabela 3065 - IPCA15 - Série histórica com número-índice,	Arquivo “CSV”	Brasil	Data, NR_MÊS, INDICE_IPCA, IPCA_ANO, IPCA_MES, IPCA_12_MESES, IPCA_SEM, IPCA_TRIM, Ano, TaxaIPCA	IBGE - SIDRA

Nº	Tabela	Descrição	Formato	Local	Variáveis	Fonte de Coleta
		variação mensal e variações acumuladas em 3 meses, em 6 meses, no ano e em 12 meses (a partir de maio/2000) do IBGE				
13	f_Petroleo_Gasolina	Série Histórica de Preços de Combustíveis, com base na pesquisa de preços da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, no formato de dados abertos.	Arquivo "CSV"	Brasil	Data, Mês, Petróleo_Brent_Brasil, Gasolina_Brasil	ANP - Dados Abertos
14	f_ReservasInternacionais	Série histórica da tabela nº 3546 de Reservas internacionais - Total - mensal que, segundo o Banco Central do Brasil, são "Ativos externos do Brasil, geridos pelo BC. São recursos para utilização imediata, se eventualmente for necessário financiar saldos negativos das contas externas e/ou estabilizar o mercado de câmbio"	Arquivo "CSV"	Brasil	Data, ResvInternac	Banco Central do Brasil Banco Central do Brasil - Séries Históricas
15	f_Exportacao_SaoLuis	Base de dados do <i>Comex Stat</i> de Exportações de 2. Base de dados detalhada por Município da empresa exportadora/importadora e Posição do Sistema Harmonizado (SH4).	Arquivo "CSV"	Estado MA e São Luís	Data, "TransacoesEspeciais", "Prod_Minerais", "MetaisComuns", "MadCarvVegObrMad", "EXPT_SLZ_ProdIndQuim", "Prod_InduAlim", "Prod_ReinoVegetal", "PedrasPreciosas", "InstrAparFoto", "MaqAparEletr", "MatTexteis", "MeracdProdDiv", "MatTransporte", "PastMadCelulo", "AnimaisVivos", "GordOleosAnimVeget", "PlastBorracha", "ObrPedGesCimCerVid", "CalcChapArtFloresArtif", "PelCourBolsas", "EXPORT_SLZ_TOTAL"	COMEXSTAT Portal Transparência - Dados Abertos
16	f_ResultadosEstado_Municipios	Diferença entre o que o governo arrecada e gasta, serve para abater a dívida, mas exclui os juros. No caso de ocorrência de superavit, o governo abate sua dívida. Já o resultado nominal é a soma do resultado primário com os juros a pagar. Assim foram unificadas as tabelas do Banco Central do Brasil de: Necessidades de financiamento (14821, 14847 e 14873), Outros fluxos (ajustes, 14899), Dívida líquida (15125, 15538 e 15668), Receita dos estados e municípios (4336 e 4363) e Dívida líquida (5798). Todas terão que ser revisadas pois foram descontinuadas pelo Banco Central.	Arquivo "CSV"	Estado MA e São Luís	DataComp, Trimestre, FinMA_MnRestPrim, FinMA_MnRestPrim, FinMA_Mn_JrNom, Ajustes_MA_Munic, DvLq_MA_Munic, DvLq_Gov_MA, DvLq_MA_Munic, Receita_Arrec_ICMS_MA, Receita_Transf_Uniao_MA, DvLq_MA_ExcCapital	Banco Central do Brasil - Séries Históricas

Nº	Tabela	Descrição	Formato	Local	Variáveis	Fonte de Coleta
17	F FROTA Municipal	Dados do Departamento Nacional de Trânsito (DENATRAN) do Ministério da Infraestrutura contendo a quantidade de veículos da frota de veículos do Município, advindos da tabela de Frota Nacional, 02 - Frota por Município e Tipo.	Arquivo "CSV"	São Luís	DATA, Veic_PgtoIPVA	SENATRAN
18	f_Arrecadação_CFEM	Resgate da divulgação no site da Agência Nacional de Mineração (ANM) da arrecadação dos Estados e Municípios Produtores de bens minerais.	Site, páginas "HTML"	Estado MA e São Luís	DATA, CFEM_MA, CFEM_Total, CFEM_Município, DataRepasso	ANM
19	f_SalárioMínimo	Define o limite mínimo para os salários pagos na produção de bens e serviços, assim como benefícios previdenciários.	Site, páginas "HTML"	Brasil	Data, Dispositivo legal, SimboloMoeda, SalarioMinimo, DATAPublicacao, DataVigencia	IPEADATA e
20	f_TaxaDesocupadosBrasil	Total de pessoas sem trabalho, procurando emprego, de 14 anos ou mais de idade. Pesquisa Nacional por Amostra de Domicílios Contínua mensal, Tabela 4092 - Pessoas de 14 anos ou mais de idade, por condição em relação à força de trabalho e condição de ocupação.	Arquivo "CSV"	Brasil	DATA, TxDesocupados	IBGE - SIDRA
21	f_Pesq_Indus_Mensal_Prod Física	Tabela da Pesquisa Industrial Mensal - Produção Física, disponível pela Tabela 3653 - Produção Física Industrial, por seções e atividades industriais (série encerrada em janeiro de 2022, vai precisar ser revisada)	Arquivo "CSV"	Brasil	DATA, PIM_PF	IBGE - SIDRA
22	f_SistNacPesqCustIndConstCivil	De acordo com o IBGE, "o Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil - SINAPI tem por objetivo a produção de séries mensais de custos e índices para o setor habitacional, e de séries mensais de salários medianos de mão de obra e preços medianos de materiais, máquinas e equipamentos e serviços da construção para os setores de saneamento básico, infraestrutura e habitação." Aqui foi utilizada a Tabela nº 647 - Custo de projeto m², por tipo de projeto e padrão de acabamento	Arquivo "CSV" ou JSON da API do IBGE	Estado do MA	DATA, SINAP_Alto, SINAP_Normal, SINAP_Baixo, SINAP_Mínimo	IBGE - SIDRA ou IBGE IBGE - API
23	f_ResultadoPerc_Principal	Mesma tabela f_Resultados_Estado_Municípios, mas esta é a nível Nacional os resultados do Governo Federal.	Arquivo "TXT"	Brasil	Data, DataVigencia, PercResSP_Nom, PercResSP_Prim	Banco Central do Brasil

Nº	Tabela	Descrição	Formato	Local	Variáveis	Fonte de Coleta
24	f_IBC_BR_AtivEcon_Per	Na descrição do Banco Central esta tabela reúne “O Índice de Atividade Econômica do Banco Central (IBC-Br) reflete as atividades da indústria, dos serviços e da agropecuária.” Foi utilizada a tabela nº 24364 - Índice de Atividade Econômica do Banco Central (IBC-Br) - com ajuste sazonal.	Arquivo “TXT”	Brasil	Data, IBC_Br, DateTime	Banco Central do Brasil e SGS - Sistema Gerenciador de Séries Temporais (bcb.gov.br)
25	d_Coeficientes CIFPM	Tabela com o agrupamento dos coeficientes individuais de distribuição do FPM para os municípios, sendo isolado apenas o de São Luís	Planilha eletrônica “XLS”, Arquivo “PDF”	Municípios de Brasil (São Luís)	Data, Coef_CIFPM_SLZ	Tribunal de Contas da União
26	d_Coeficiente_IPM	Tabela com unificação do coeficiente do Índice de Participação dos Municípios para distribuição no Estado da receita do ICMS, IPI e 50% da CIDE, separado apenas o Município de São Luís	Arquivo “PDF”	Municípios de MA (São Luís)	Data, Coef_IPM	Portal Transparência Governo do Estado do Maranhão
27	d_Calendarario	Tabela criada por <i>procedure</i> em linguagem “M” para criação dos relacionamentos entre as tabelas pela chave primária “Data”. Depende de dois parâmetros que recebem 2 variáveis globais: “MenorData” e “MaiorData”.	Não se aplica	Não se aplica	Data, Ano, MesNr, Nome_Mes, Trimestre, Quinzena, Período Quinzena, BimNr, Bimestre, MES_ANO	Função que verifica a Maior e a Menor data entre as tabelas elencadas.

Fonte: Elaborado pela Autora (2023).

ETAPAS DE EXTRAÇÃO, CARREGAMENTO E TRANSFORMAÇÃO DOS DADOS

Tabela 01 - f_Programaca_Financeira

Descrição: Contém os dados da execução financeira do Município, da arrecadação em sua publicação contábil, publicados no portal da transparência do Governo Municipal.

Fonte de coleta: https://saoluis.giap.com.br/apex/saoluis/www_flow.accept.

Formato da fonte: Parte dos arquivos foram reutilizados do Trabalho de conclusão de curso, os quais estavam em formato “PDF” e os demais foram baixados no portal do Governo Municipal de São Luís em formato “CSV”. A Figura 8 apresenta um exemplo do arquivo em “PDF” do Balancete da Receita do Governo Municipal de São Luís.

Figura 8 - Visualização do Balancete da Receita

MUNICÍPIO DE SÃO LUIS - CONSOLIDADO GERAL											
Balancete da Receita referente ao Período de 01/01/2019 a 31/01/2019											
Exercício de 2019											
				RUBR.	%	APLIC	CLASSIFICAÇÃO	ESPECIFICAÇÃO DA RECEITA		RECEITA	RECETA ORÇADA
										NO MÊS	NO ANO
BalRec2019-01 - Por Aplicação.pdf	13/03/2021		100000000000					RECEITAS CORRENTES		267.757.812,34	3.156.219.085,27
BalRec2019-02 - Por Aplicação.pdf	13/03/2021		110000000000					IMPOSTOS, TAXAS E CONTRIBUIÇÕES DE MELHORIA		55.478.493,35	760.179.393,00
BalRec2019-03 - Por Aplicação.pdf	13/03/2021		110000000000					IMPOSTOS, TAXAS E CONTRIBUIÇÕES DE MELHORIAS		6.082,51	3.722.886,00
BalRec2019-04 - Por Aplicação.pdf	13/03/2021		110000000000					IMPOSTOS		51.841.884,96	725.904.802,00
BalRec2019-05 - Por Aplicação.pdf	13/03/2021		111000000000					IMPOSTOS SOBRE A RENDA E PROVENTOS DE QUALQUER NATUREZA		2.540.210,40	75.221.610,00
BalRec2019-06 - Por Aplicação.pdf	13/03/2021		111300000000					IMPOSTO SOBRE A RENDA - RETIDO NA FONTE		2.540.210,40	75.221.610,00
BalRec2019-07 - Por Aplicação.pdf	13/03/2021		111303000000					IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO		2.540.210,40	75.221.610,00
BalRec2019-08 - Por Aplicação.pdf	13/03/2021		111303100000					IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO - PRINCIPAL		14.206,35	7.522.161,00
BalRec2019-09 - Por Aplicação.pdf	13/03/2021		111303110001	10	0001	100000119		IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO - PRINCIPAL		21.309,53	11.283.241,50
BalRec2019-10 - Por Aplicação.pdf	13/03/2021		111303110001	15	0001	102000000		IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO - PRINCIPAL		35.515,88	18.805.402,50
BalRec2019-11 - Por Aplicação.pdf	13/03/2021		111303110001	25	0001	101000000		IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO - PRINCIPAL		71.031,75	37.610.805,00
BalRec2019-12 - Por Aplicação.pdf	13/03/2021		111303110001	50	0001	100000000		IMPOSTO SOBRE A RENDA - RETIDO NA FONTE - TRABALHO - PRINCIPAL		2.398.146,90	0,00
			111303110002	100	0176	100000000		OUTROS PENDIMENTOS			

Fonte: Balancete da Receita (SÃO LUÍS-MA, 2021b).

Periodicidade da frequência dos dados: Anual e bimestral.

Periodicidade de Publicação: ao final de cada bimestre.

Limpeza e Transformação: Os procedimentos efetuados para ajustar os dados carregados foram efetivados no Relatório Principal do Power BI Desktop, no arquivo denominado V6_BI_RECEITAS_TRANSF_DADOS_CAPITAIS, em que serão unificadas todas as informações das Capitais. Por intermédio da Figura 9 é possível ler o roteiro das “ETAPAS APLICADAS”, as estatísticas descritivas de cada variável da tabela, quantidade de erros, valores nulos e ausentes.



Tabela 02 - $f_{\text{Dolar_Comercial}}$

Descrição: De acordo com o Instituto de Pesquisa Econômica Aplicada (IPEA), “a taxa de câmbio é o preço de uma moeda estrangeira medido em unidades ou frações (centavos) da moeda nacional, refletindo o custo de uma moeda em relação à outra. [...] Também conhecida como PTAX, esta cotação corresponde à média aritmética das taxas de compra das consultas realizadas diariamente” (BRASIL; IPEA, 2022).

Ajuste dos dados para mensal Média entre o maior e o menor valor cotado do dólar comercial no mês. Cotação do Dólar dos EUA. Cotações de fechamento de uma moeda em um período.

Fonte de coleta: <https://www.bcb.gov.br/estabilidadefinanceira/historicocotacoes>.

Formato da fonte: página na web e planilhas eletrônicas do tipo “XLS”.

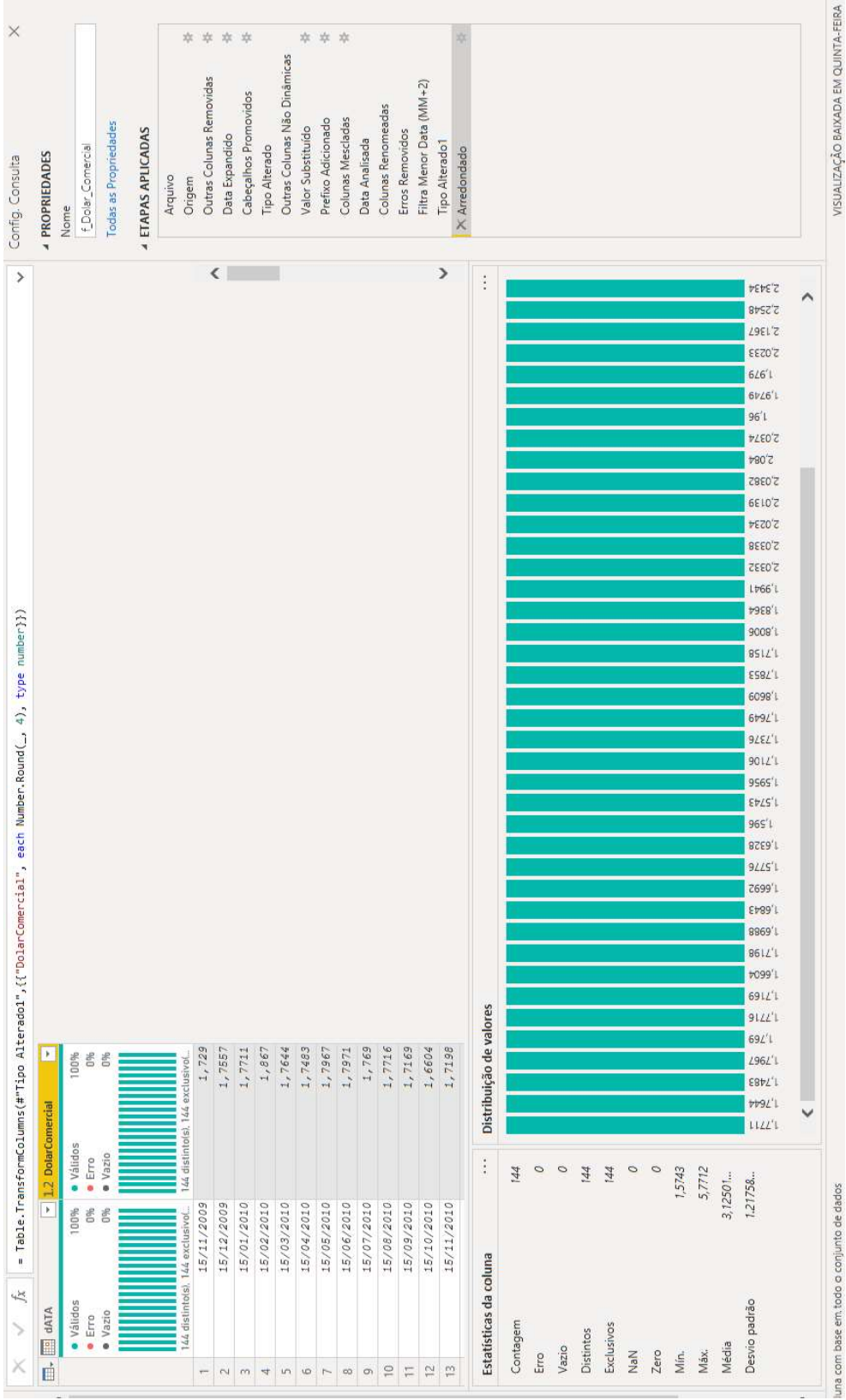
Periodicidade da frequência dos dados: Diário.

Periodicidade de Publicação: ao final de cada dia.

Limpeza e Transformação: Os procedimentos efetuados para ajustar os dados carregados foram efetivados no Relatório Principal do Power BI Desktop, no arquivo denominado V6_BI_RECEITAS_TRANSF_DADOS_CAPITAIS, em que serão unificadas todas as informações das Capitais. Por intermédio da Figura 10 é possível ler o roteiro das “ETAPAS APLICADAS”, as estatísticas descritivas de cada variável da tabela, quantidade de erros, valores nulos e ausentes.

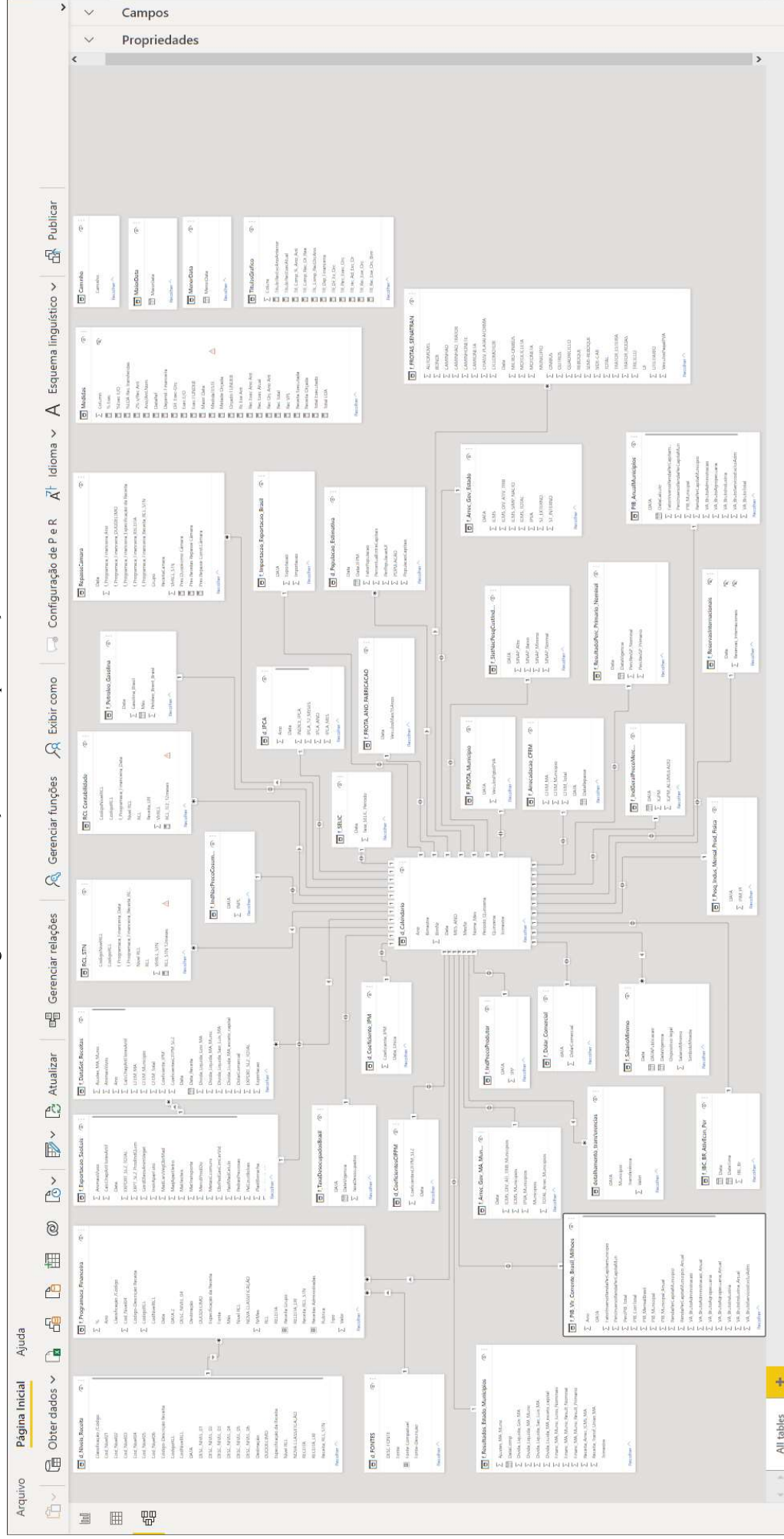
As duas tabelas são apenas exemplos escolhidos para ilustrar as demais elencadas na Quadro 27. Concluída a coleta dos dados, todas as tabelas foram interligadas por intermédio do campo “Data” da tabela “ $d_Calendario$ ” e suas respectivas colunas de datas na cardinalidade de 1 para muitos, utilizando a topologia estrela, conforme ilustração da Figura 11.

Figura 10 - Limpeza e Transformação das variáveis da tabela “f_Dolar_Comercial”



Fonte: Imagem do componente “Power Query Editor” da ferramenta Power BI Desktop.

Figura 11 - Unificação das tabelas para criação do Dataset



Fonte: Imagem do componente “Modelagem” da ferramenta Power BI Desktop.

APÊNDICE D: Relatório de descrição dos dados da primeira iteração com a Capital São Luís

No presente relatório, relacionam-se as possíveis variáveis que influenciam as receitas, incluindo sua descrição, periodicidade de publicação e frequências que são disponibilizadas. Ao término da tarefa de extração, carregamento e transformação dos dados de cada tabela elas foram agregadas na tabela “*f_DataSet_Receitas*”, criando o arquivo do tipo “CSV” para a fase seguinte de Preparação dos dados.

Quadro 28 - Descrição dos dados por variável, relação com cada receita e periodicidade

Nr	Tabela	Tipo	Variável	FPM	IP	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
0	d_Calendario	datetime	Data_Receita	X	X	X	X	X	X	Data utilizada como chave primária em todas as tabelas, unificando os dados de forma mensal, mas a tabela é diária.	Diário	Início do dia	(procedure linguagem "M")
1	d_Coeficient e_IPM	float64	Coef_IPM		X			X		Coefficiente de distribuição Estadual da receita do ICMS de São Luís.	Anual	Final ano	Portal SEFA-MA
2	d_Coeficient esCIFPM	float64	Coef_CIFPM_SLZ	X			X			Coefficiente individual de distribuição do FPM para o município do São Luís.	Anual	Final ano	Portal TCU
3	d_IPCA	float64	TxIPCA	X	X	X	X	X	X	Índice de Preço ao Consumidor Amplo.	Mensal	Mês depois	Sidra-IBGE
4	d_Populacao _Estimativa	int64	POPULACAO	X	X		X	X		População Municipal série histórica do IBGE. Utiliza-se a do ano anterior.	Anual	A cada mês de julho, divulga o ano anterior	IBGE e API-IBGE.
5		float64	FtPOPULACAO	X			X			Cálculo da população relativa ao montante da soma de todas as capitais, conforme método do TCU.			
6		float64	Per_EntreCapitais	X	X		X			Percentual da População de São Luís em relação a todas as Capitais.			
7		float64	PerPopulacaoUF	X	X		X			Percentual da População de São Luís em relação aos Municípios do Estado.			
8	f_Arrec_Gov _Estado	float64	ICMS					X		Arrecadação do ICMS no Estado do MA.	Diária	Tempo real	Portal SEFA-MA
9		float64	ICMS_DvAtTrb					X		Valor da dívida ativa tributária paga no Estado do MA.			
10		float64	ICMS_TOTAL					X		Soma do ICMS com Simples Nacional, mais as Substituições Tributárias Internas e Externas da arrecadação do Governo do Estado.	Diária	Tempo real	Portal SEFAZ-MA
11		float64	ST_EXTERNO					X		Modalidade de Substituição Tributária de movimento externa no Estado.			
12	f_Arrec_Gov _Estado	float64	ST_INTERNO					X		Arrecadação de Substituição Tributária de movimento interna no Estado.	Diária	Tempo real	Portal SEFAZ-MA
13		float64	ICMS_SIMP_NACIO					X		Arrecadação do ICMS recolhido no Simples Nacional do Governo do Estado.			

Nr	Tabela	Tipo	Variável	FPM	IPI	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
14		float64	IPVA						X	Arrecadação do IPVA recolhido no Estado.			
15	f_Arrec_Gov _MA_Municipios	float64	ICMS_DvAtTrb_Muns					X	X	Arrecadação do ICMS recolhido no Município do Estado do Maranhão.			
16		float64	ICMS_Muns					X		Arrecadação do ICMS recolhido no Município do Estado do Maranhão.			
17		float64	IPVA_Muns						X	Arrecadação do IPVA recolhido no Município do Estado do Maranhão.			
18		float64	TOTAL_ArrecMun					X		Total da arrecadação ocorrida em cada Município do Estado.			
19	f_Arrecadacao_CFEM	float64	CFEM_MA			X				Arrecadação da CFEM de cada Estado somando todos os Municípios produtores, divulgado pela ANM.			
20		float64	CFEM_Total			X				Arrecadação da CFEM de todos os Estado que possuem Município produtores.			
21		float64	CFEM_Mun			X				Arrecadação por Município produtor, valores da CFEM recolhido bruto, divulgado pela ANM.			
22	f_Dolar_Comercial	float64	DolarComercial	X	X	X	X			Média entre o maior e o menor valor cotado do dólar comercial no mês. Cotação do Dólar dos EUA. Cotações de fechamento de uma moeda em um período.	Diário	Final do dia	Banco Central do Brasil
23	F_FROTA_Municipio	float64	Veic_PgtolIPVA						X	Quantidade de veículos com registro de emplacamento do Município.	Mensal	Mês depois	Portal SENATRAN
24	f_IBC_BR_AtivEcon_Per	float64	IBC_Br	X	X			X		Índice de Atividade Econômica do Banco Central (IBC-Br) reflete as atividades da indústria, dos serviços e da agropecuária.	Mensal	Mês depois	Banco Central do Brasil
25	f_Importacao_Exportacao_Brasil	float64	Importacao		X			X		Importação de bens: corresponde ao valor total de ingresso de mercadorias originárias do exterior para a unidade da federação especificada.		2ª quinzena após fim do mês de referência	Portal do Comércio Exterior
26		float64	Exportacao		X			X		Exportação de bens: corresponde ao valor total das mercadorias vendidas ao exterior pela unidade da federação especificada.			
27	f_IndGeralPrecoMercado	float64	IGPM		X			X		Índice Geral de Preços - Mercado. Indicador mensal do nível de atividade econômica do país, englobando seus principais setores. apura informações sobre a variação de preços do dia 21 do mês anterior ao dia 20 do mês de coleta.	Mensal	Final do mês	Portal FGV
28	f_IndNacPrecoConsumidor	float64	INPC		X	X	X	X	X	Índice Nacional de Preços ao Consumidor variação mensal.	Mensal	Mês depois	API-IBGE
29	f_IndPrecoProdutor	float64	IPP		X		X	X		Índice de Preços ao Produtor, focado na variável da indústria geral.	Mensal	Bimestral	
30	f_Petroleo_Gasolina	float64	Petroleo_Brent_Brasil		X		X	X		Preço médio do petróleo Brent. referência para formação de preço pela Petrobras. Cálculo da média entre a maior e menor cotação do mês, multiplicado pelo dólar.			
31		float64	Gasolina_Brasil		X		X	X		Preço médio da Gasolina comum divulgado pela Agência Nacional de Petróleo	Mensal	Fim do mês	Banco Central do Brasil

Nr	Tabela	Tipo	Variável	FPM	IPI	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
32	f_PIB_Vlr_Corrente_Brasil_Milhoes	float64	PreInvRndPrCpMn	X			X			Cálculo feito para o CIFPM. O valor é 1 dividido pela “RendaPerCapitaMunicípio” e o resultado dividido pelo total das Capitais multiplicado por 100	Anual (cálculo p/mensal)	Julho para 2 anos anteriores	IBGE
33		float64	PIB_Mun	X	X		X	X		Derivação do cálculo da renda per capita, incluir o valor para o exercício, Produto Interno Bruto, a preços correntes, utilizou-se a proporção da sazonalidade do PIB Brasil Mensal para aplicar no PIB Municipal.			
34		float64	PIB_MenBR	X	X	X	X	X	X	Produto Interno Bruto do Brasil, desmembrado da série trimestral pelo BACEN (método de <i>Denton-Chollette</i>). Decorrente da influência na renda será incluído para o mesmo exercício.	Mensal	2 meses anteriores	IPEA
35		float64	RndPrCap_Mun	X			X			Derivação do cálculo da renda per capita, incluir o valor para o exercício, Produto Interno Bruto per capita, a preços correntes.			
36		float64	VA_BrtAdm	X	X		X	X		Valor adicionado bruto da Administração, defesa, educação e saúde públicas e seguridade social, a preços correntes do Município.			
37		float64	VA_BrtAgro	X	X		X	X		Valor adicionado bruto da Agropecuária, a preços correntes do Município.			
38		float64	VA_BrtIndust	X	X		X	X		Valor adicionado bruto da Indústria, a preços correntes do Município.			
39		float64	VA_BrtServExclAdm	X	X		X	X		Valor adicionado bruto dos Serviços, a preços correntes – exceto Administração, defesa, educação e saúde públicas e seguridade social do Município.	Anual (cálculo p/mensal)	Julho para 2 anos anteriores	IBGE
40		float64	VA_BrtTotal	X	X		X	X		Valor adicionado bruto total, a preços correntes do Município.			
41		float64	VA_ImpostosProd_	X	X		X	X		Valor adicionado de Impostos, líquidos de subsídios, sobre produtos, a preços correntes do Município.			
42		float64	FInvRndPrCap_Mun	X			X			Mesmo cálculo feito para CIFPM, mas em formato de índice.			
43	f_Programacao_Financieira	object	RECEITA	X	X	X	X	X	X	Categoria da Série Temporal, o Tipo da Receita, com a descrição popular da sigla da Receita: CFEM, FEP, CFEM AFETADOS, FPM, IPVA, ICMS e IPI.	Mensal	Bimestral	Portal Gov. São Luís/MA
44		float64	Realizado	X	X	X	X	X	X	Variável “target”. Valor Bruto da Receita executada nos Balançetes da Receita.		Bimestral	
45	f_ReservasInternacionais	float64	ResvInternac	X	X	X	X	X		Reservas Internacionais, utilização em eventuais financiamentos de saldos negativos das contas externas e/ou estabilizar o mercado de câmbio.	Mensal	Final de 2 meses Anteriores	Banco Central do Brasil
46	f_ResultadoPerc_PrimaryNominal	float64	PercResSP_Nom	X	X			X		Percentual do resultado nominal, soma do resultado primário com os juros a paga.			
47		float64	PercResSP_Prim	X	X			X		Percentual do resultado primário, diferença do quanto o governo gasta e arrecada.	Mensal	Mês depois	Banco Central do Brasil
48		float64	EXPT_SLZ_TOTAL		X			X		Valor de todas as categorias de produtos exportados pelo município de São Luís.	Mensal	2ª quinzena	ComexStat

Nr	Tabela	Tipo	Variável	FPM	IPI	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
49	f_Resultados _Estado_Mu nicipios	float64	EXPT_SLZ_ProdIndQuim		X			X		Valores dos produtos das indústrias químicas ou indústrias conexas que são exportados pelo município de São Luís.	após fim do mês	ComexStat	
50		int64	CalcChapArtFloresArtif		X			X		Exportações de calçado, chapéus e artefatos de uso semelhante, guarda-chuvas, guarda-sois, bengalas, chicotes e suas partes; penas preparadas e suas obras; flores artificiais; obras de cabelo.			
51		float64	AnimaisVivos		X			X		Exportação de animais vivos e produtos do reino animal: animais vivos, carnes e miudezas, comestíveis, peixes e crustáceos, moluscos e outros invertebrados aquáticos, leite e laticínios; ovos de aves; mel natural; produtos comestíveis de origem animal, não especificados nem compreendidos noutros capítulos; e outros produtos de origem animal, não especificados nem compreendidos noutros capítulos.			
52		float64	GordOleosAnimVeget		X			X		Exportações de gorduras e óleos animais ou vegetais; produtos da sua dissociação; gorduras alimentares elaboradas; ceras de origem animal ou vegetal.			
53		float64	InstrAparFoto		X			X		Exportação de instrumentos e aparelhos de ótica, fotografia ou cinematografia, medida, controle ou de precisão; Instrumentos e aparelhos médico-cirúrgicos; Relógios e aparelhos semelhantes; Instrumentos musicais; suas partes e acessórios.			
54		float64	MadCarvVegObrMad		X			X		Exportações de madeira, carvão vegetal e obras de madeira: cortiça e suas obras; obras de espartaria ou de cestaria.			
55		float64	MaqAparEletr		X			X		Exportação de máquinas e aparelhos, material elétrico e suas partes; aparelhos de gravação ou reprodução de som, aparelhos de gravação ou reprodução de imagens e de som em televisão, e suas partes e acessórios.			
56		float64	MatTransporte		X				X				Exportação de material de transporte.
57		float64	MatTexteis		X				X				Exportação de matérias têxteis e suas obras.
58		float64	MercdProdDiv		X				X				Exportação de mercadorias e produtos diversos.
59		float64	ObrPedGesCimCerVid		X				X				Exportação de obras de pedra, gesso, cimento, amianto, mica ou de matérias semelhantes; produtos cerâmicos; Vidro e suas obras.
60		float64	MetaisCcomuns		X	X			X				Exportação de metais comuns e suas obras.
61		float64	PastMadCelulo		X				X				Exportação de pastas de madeira ou de outras matérias fibrosas celulósicas; papel ou cartão para reciclar (desperdícios e aparas); papel e suas obras.
62		float64	PelCourBolsas		X				X				Exportação de peles, couros, peles com pelo e obras destas matérias; artigos de correio ou de seleiro; artigos de viagem, bolsas e artefatos semelhantes; obras de tripa.
63		float64	PlastBorracha		X				X				Exportação de plásticos e suas obras; borracha e suas obras.
64		float64	Prod_InduAlim		X				X				Exportação de produtos das industriais alimentares; bebidas, líquidos alcoólicos e vinares; Tabaco e seus sucedâneos manufaturados.

Nr	Tabela	Tipo	Variável	FPM	IPI	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
65		float64	Prod_Minerais		X	X		X		Exportação de produtos minerais.			
66		float64	Prod_ReinoVegetal		X			X		Exportação de produtos do reino vegetal.			
67		float64	TransacoesEspeciais		X			X		Exportação de produtos transações especiais.			
68		float64	Ajustes_MA_Munic	X						Outros fluxos (ajustes) do Estado do MA e municípios, valores em R\$ (milhões) proveniente da Tabela nº 14899 - Outros fluxos (ajustes) do Estado do MA e municípios - Fluxo no trimestre - R\$ (milhões)			
69		float64	DvLq_Gov_MA	X						Dívida líquida do Governo do Estado do Maranhão proveniente da Tabela nº 15538 - Dívida líquida do Governo do Estado do Maranhão - R\$ (milhões)			
70		float64	DvLq_MA_Munic	X						Dívida líquida de todos os municípios do Estado do Maranhão, incluindo a capital proveniente da Tabela nº 15125 - Dívida líquida do Estado do Maranhão e municípios - R\$ (milhões)			
71		float64	DvLq_Sao_Luis_MA	X						Dívida líquida do município de São Luís do Estado do Maranhão proveniente da Tabela nº 15668 - Dívida líquida do Município de São Luís (MA) - R\$ (milhões)			
72		float64	DvLq_MA_ExcCapital	X						Dívida líquida dos municípios do Estado do Maranhão, exceto a capital proveniente da Tabela nº 15798 - Dívida líquida dos municípios do Estado do Maranhão, exceto a capital - R\$ (milhões)	Trim.	30 dias	Banco Central do Brasil
73		float64	FinMA_Mn_JrNom	X						Fluxo dos juros, sobre a dívida interna e externa valores nominais do Município proveniente da Tabela nº 14873 - Necessidades de financiamento do Estado do MA e municípios - Juros nominais - Fluxo no trimestre - R\$ (milhões)			
74		float64	FinMA_MnRestPrim	X						Resultados primários, componente não financeiros do resultado fiscal da dívida do Município, resultado da subtração do resultado com os juros, ambos nominais, proveniente da Tabela nº 14821 - Necessidades de financiamento do Estado do MA e municípios - Resultado primário - Fluxo no trimestre - R\$ (milhões)			
75		float64	FinMA_MnRestNom	X						Corresponde a variação nominal dos saldos da dívida líquida, é composta pelos juros reais, dívida atualizada e resultado primário do Município proveniente da Tabela nº 14847 - Necessidades de financiamento do Estado do MA e municípios - Resultado nominal - Fluxo no trimestre - R\$ (milhões).			
76	f_SalarioMinimo	float64	SalarioMinimo	X						Define o limite mínimo para os salários pagos na produção de bens e serviços, assim como benefícios previdenciários.	Anual	Até final do ano anterior	IPEA
77	f_SELIC	float64	Tx_SELIC_Perd	X	X	X	X	X	X	Taxa que o governo define o "custo do dinheiro" na economia brasileira, baliza as taxas das instituições financeiras.	8 reuniões	Dia seguinte	Banco Central do Brasil

Nr	Tabela	Tipo	Variável	FPM	IPI	CFEM	FEP	ICMS	IPVA	Descrição	Periodicidade		Site para download
											Frequência	Publicação	
78	f_SistNacPes qCustIndCon stCivil	float64	SINAP_Alto	X	X			X		Custo de projeto por m², por tipo de projeto e padrão de acabamento Alto do Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil (SINAPI), tabela nº 647.	Mensal	Mês depois	SIDRA-IBGE
79		float64	SINAP_Baixo	X	X			X		Custo de projeto por m², por tipo de projeto e padrão de acabamento Baixo do SINAPI, tabela nº 647.			
80		float64	SINAP_Mínimo	X	X			X		Custo de projeto por m², por tipo de projeto e padrão de acabamento Mínimo do SINAPI, tabela nº 647.			
81		float64	SINAP_Normal	X	X			X		Custo de projeto por m², por tipo de projeto e padrão de acabamento Normal do SINAPI, tabela nº 647.			
82	f_TaxaDesocupadosBrasil	float64	TxDesocupados	X				X		Total de pessoas sem trabalho, procurando emprego, de 14 anos ou mais de idade.	Mensal	Mês depois	SIDRA-IBGE

Fonte: Elaborado pela Autora (2023).

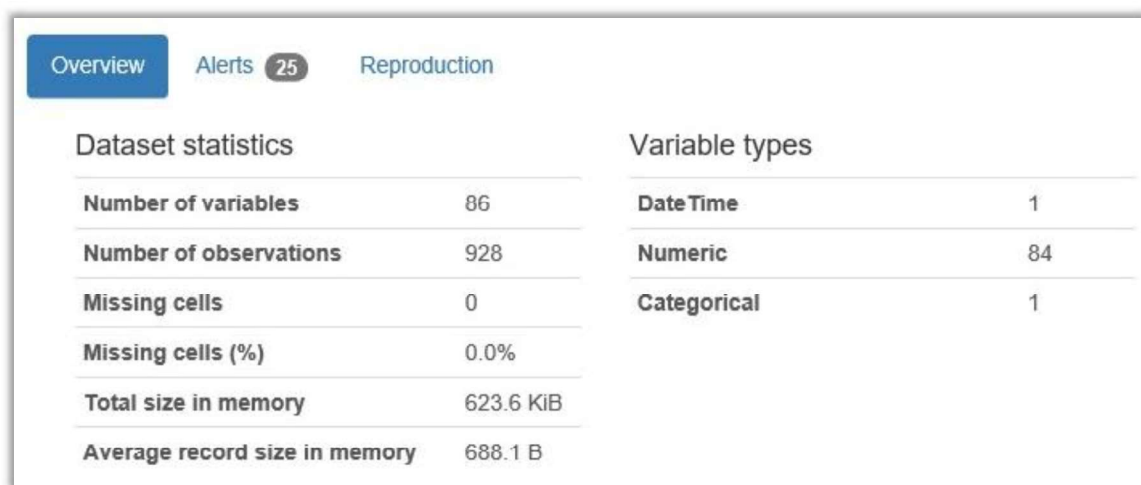
APÊNDICE E: Relatório de exploração de dados com o *pandas profiling* de São Luís

O presente relatório versa sobre a exploração dos dados do *Dataset* criado para o Município de São Luís, cujas variáveis estão elencadas no APÊNDICE D, utilizando o pacote “*pandas profiling*” da linguagem *Python*, executado no *Jupyter Notebook* e salvo em um arquivo “HTML”.

VISÃO GERAL

Iniciando com o resumo geral, listado no Quadro 29, o *Dataset*, em comento, possui 86 variáveis com 928 observações, sem valores ausentes, com 1 variável do tipo data, 1 categórica e 84 numéricas.

Quadro 29 - Visão geral dos dados do *Dataset* pelo relatório do *pandas profiling*



The screenshot shows the 'Overview' tab of the pandas profiling report. It features two tables: 'Dataset statistics' and 'Variable types'. The 'Dataset statistics' table lists metrics such as the number of variables (86), observations (928), missing cells (0), and memory usage. The 'Variable types' table shows the distribution of data types: 1 DateTime, 84 Numeric, and 1 Categorical.

Dataset statistics		Variable types	
Number of variables	86	DateTime	1
Number of observations	928	Numeric	84
Missing cells	0	Categorical	1
Missing cells (%)	0.0%		
Total size in memory	623.6 KiB		
Average record size in memory	688.1 B		

Fonte: Relatório gerado pelo pacote “*pandas profiling*” (2023).

O pacote disponibiliza, ainda na visão geral, os principais alertas que encontrou, os quais foram relacionados no Quadro 30, no qual identificam-se 19 variáveis com percentuais relevantes de valores com “ZEROS”. As 4 variáveis com valores constantes, já eram esperadas, pois fazem parte do cálculo do coeficiente da recita do FPM, sendo única para cada Capital, como o *Dataset* é apenas do Município de São Luís, foi classificada como alerta. As 2 variáveis únicas, “Data_Receita” e “Realizado” também são esperadas, uma vez que a primeira será utilizada como índice e a segunda é a variável alvo.

Quadro 30 - Relação dos alertas do relatório de exploração dos dados do *pandas profiling*

Overview Alerts 25 Reproduction		
Alerts		
FtPOPULACAO	has constant value "2.5"	Constant
PrcInvRndPrCpMn	has constant value "3.704"	Constant
CalcChapArtFloresArtif	has constant value "2"	Constant
FtInvRndPrCap_Mun	has constant value "0.4"	Constant
Data_Receita	has unique values	Unique
Realizado	has unique values	Unique
Veic_PgtoIPVA	has 14 (1.5%) zeros	Zeros
AnimaisVivos	has 706 (76.1%) zeros	Zeros
GordOleosAnimVeget	has 714 (76.9%) zeros	Zeros
InstrAparFoto	has 760 (81.9%) zeros	Zeros
MadCarvVegObrMad	has 721 (77.7%) zeros	Zeros
MaqAparEleetro	has 657 (70.8%) zeros	Zeros
MatTransporte	has 851 (91.7%) zeros	Zeros
MatTexteis	has 707 (76.2%) zeros	Zeros
MercdProdDiv	has 721 (77.7%) zeros	Zeros
ObrPedGesCimCerVid	has 738 (79.5%) zeros	Zeros
MetaisCcomuns	has 378 (40.7%) zeros	Zeros
Orçado	has 410 (44.2%) zeros	Zeros
PastMadCelulo	has 714 (76.9%) zeros	Zeros
PelCourBolsas	has 859 (92.6%) zeros	Zeros
PlastBorracha	has 727 (78.3%) zeros	Zeros
Prod_InduAlim	has 702 (75.6%) zeros	Zeros
Prod_Minerais	has 379 (40.8%) zeros	Zeros
Prod_ReinoVegetal	has 562 (60.6%) zeros	Zeros
TransacoesEspeciais	has 200 (21.6%) zeros	Zeros

Fonte: Relatório gerado pelo pacote "*pandas profiling*" (2023).

Adiantando os afazeres da fase de "Preparação dos dados", decidiu-se por eliminar as variáveis que possuem valores "ZEROS" acima de 40 %. A variável "TransacoesEspeciais" também foi suprimida, pois notou-se que São Luís não tem grande relevância para exportação. Quanto a "Veic_PgtoIPVA" também foi suprimida, pelo fato de ter sua série temporal com publicação atrasada, no Governo Federal (caso raro), devido ao período da Pandemia da COVID-19, quando voltaram as publicações normais já havia sido executada a 1ª iteração, a série histórica do Governo do Estado do Maranhão está com 3 anos de atraso e ainda não voltou até a elaboração deste relatório.

ANALISANDO AS VARIÁVEIS

Ao se explorar as variáveis do *Dataset*, três estão intimamente interligadas: “Data_Receita”, “Realizado” e “RECEITA”. Ao verificar a “Data_Receita” os resumos da Figura 12 obtém-se uma série temporal de 11 anos, com 928 observações, sem valores ausentes, 100% distintas, tudo certo para um atributo que será o índice. Contudo, o histograma não apresenta uma uniformização na distribuição das datas.

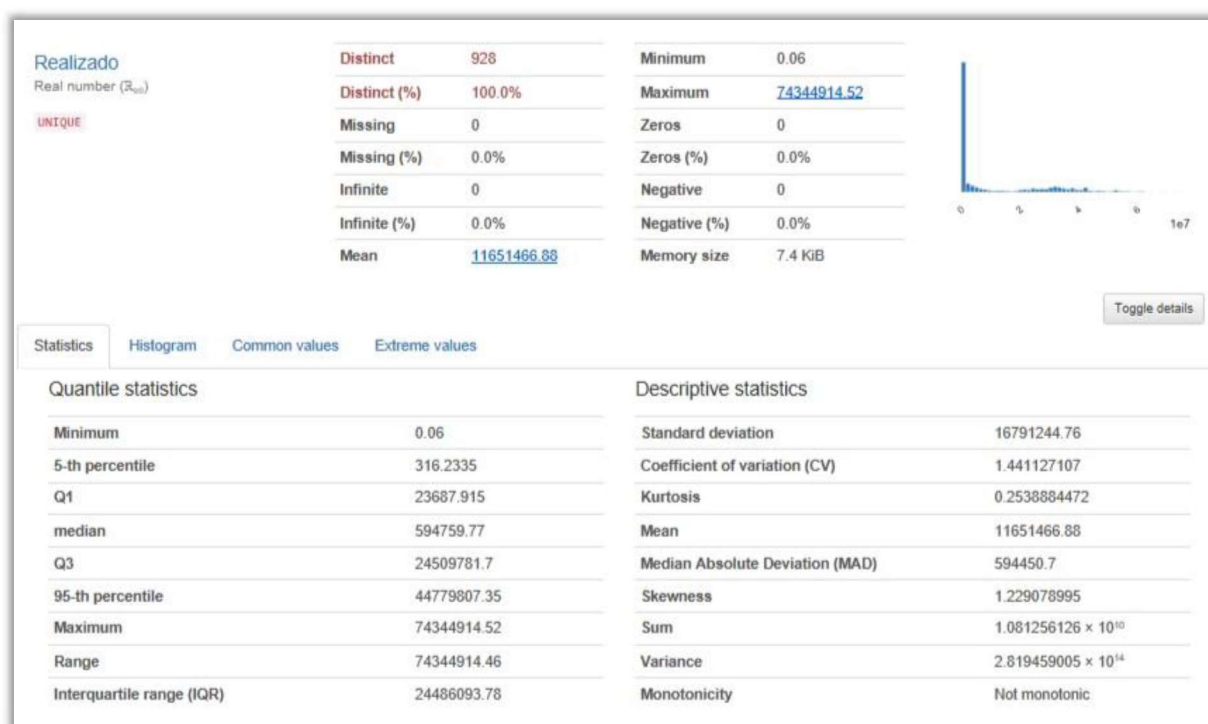
Figura 12 - Variável "Data_Receita" quadro resumo das estatísticas e histograma



Fonte: Relatório gerado pelo pacote “*pandas profiling*” (2023).

Seguindo-se a exploração, verificou-se a distribuição e o quadro resumo do principal atributo do estudo: “Realizado”. Na Figura 13, de imediato, a ausência de normalidade e os outlier são evidentes, a mediana e a média muito discrepantes.

Figura 13 - Variável " Realizado " quadro resumo das estatísticas e histograma



Fonte: Relatório gerado pelo pacote “*pandas profiling*” (2023).

Mesmo observando os valores extremos, conforme Quadro 31, com os 5 maiores e os menores, não é possível avaliar as questões de *outlier*, pois as 928 observações são de todas as receitas de transferências constitucionais.

Quadro 31 - Valores extremos, entre 5 maiores e menores, da Variável "Realizado"

Minimum 5 values			Minimum 5 values			Maximum 5 values		
Value	Count	Frequency (%)	Value	Count	Frequency (%)	Value	Count	Frequency (%)
0.06	1	0.1%	74344914.52	1	0.1%			
6.25	1	0.1%	69333156.28	1	0.1%			
6.91	1	0.1%	64384961.39	1	0.1%			
7.5	1	0.1%	62094000.13	1	0.1%			
14.92	1	0.1%	62091899.38	1	0.1%			

Fonte: Relatório gerado pelo pacote “*pandas profiling*” (2023).

Para dirimir tal situação, fez-se mister do auxílio da variável “RECEITA” para investigá-la, uma vez que esta faz a identificação dos 10 subconjuntos, por intermédio de suas categorias, para os quais se pretende prever os valores do “Realizado”. E de acordo com a Figura 14, consegue compreender, inicialmente, que os dados não estão com suas séries temporais consolidadas corretamente, pois todas as categorias deste atributo deveriam ter o mesmo número de observações 132. Sendo conveniente registrar a exceção da quantidade de observações das 3 últimas categorias, CFEM AFETADOS e as Emendas Constitucionais do FPM.

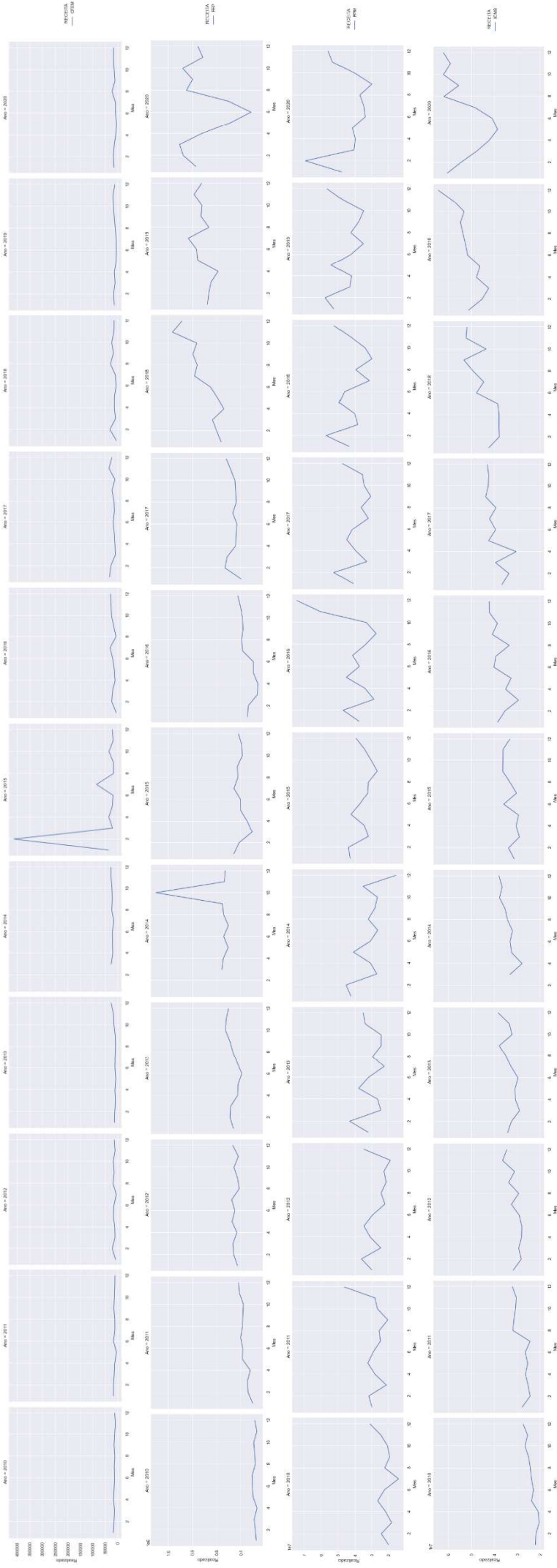
Figura 14 - Variável "RECEITA" quadro resumo das estatísticas e percentual de distribuição

RECEITA Categorical		Distinct 10 Distinct (%) 1.1% Missing 0 Missing (%) 0.0% Memory size 7.4 KiB	FPM 132 FEP 130 ICMS 130 IPVA 130 CFEM 126 Other values (5) 280
Common Values			
Value	Count	Frequency (%)	
FPM	132	14.2%	
FEP	130	14.0%	
ICMS	130	14.0%	
IPVA	130	14.0%	
CFEM	126	13.6%	
IPI	124	13.4%	
ITR	123	13.3%	
CFEM AFETADOS	18	1.9%	
FPM EC 55/07	12	1.3%	

Fonte: Relatório gerado pelo pacote “*pandas profiling*” (2023).

Quando o *Dataset* é separado por subconjuntos, conforme as categorias do atributo “RECEITA” com suas respectivas observações do “Realizado” é possível identificar que há dados ausentes, meses que estão faltando dentro das séries temporais, bem como *outliers*, conforme Figura 15.

Figura 15 - Plotagem em grid das séries temporais da variável “Realizado” por rótulos da categoria da “RECEITA”



APÊNDICE F: Comprovante do depósito de IPI no Agente Financeiro

Tendo sido descoberto que havia divergências nas transferências do IPI-Exportação, foi encaminhado um memorando informando o fato e sugerindo a solicitação de explicações por parte do Agente Financeiro e ulteriormente envio de Ofício ao Governo do Estado do Maranhão para elucidar a ocorrência (ver ANEXOS A, B, C e D). Sendo importante registrar que os valores de anos anteriores, mencionados no APÊNDICE G, foram encaminhados à Procuradoria Geral do Município e à Controladoria Geral do Municípios para as devidas providências. Todavia os repasses ausentes referente ao exercício financeiro atual, dos dias 10 de janeiro de 2022 (ANEXO C) e 10 de junho de 2022 (ANEXO D), especificado na Tabela 15 e Tabela 16, baseados nos depósitos efetivados ao Governo do Estado do Maranhão pelo Federal dos meses de janeiro (ANEXO A) e de junho (ANEXO B), foram cobrado das referidas Entidades.

Tabela 15 - Demonstrativo dos Cálculo do IP do dia 10/01/2022

Valores distribuídos, conforme demonstrativo do ANEXO A

Data Oficial	IPI Municípios	FUNDEB Municípios	PASEP Municípios
10/01/2022	1.438.337,89	287.667,57	11.506,70

Valores distribuídos para o Município de São Luís, conforme IPM do exercício de 2021

Valor Líquido	IPI Bruto SLZ	FUNDEB Municípios	PASEP Municípios	Coeficiente IPM 2021
374.402,96	472.731,00	94.546,20	3.781,85	32,86647790

Valores distribuídos para o Município de São Luís, conforme IPM do exercício de 2022

Valor Líquido	Valor Bruto	FUNDEB	PASEP	Coeficiente IPM 2022
364.294,86	459.968,16	91.993,60	3.679,70	31,97914500
Difereça da utilização dos coeficientes do IPM			10.108,10	

Fonte: Elaborado pela Autora (2023) com base nos depósitos dos ANEXOS A e D.

Vale ressaltar que, da aludida tabela, foi cobrado o valor com o coeficiente do IPM de 2021, mas o depósito foi efetivado com o de 2022, conforme consulta pública do ANEXO D, deixado a cargo do Município de São Luís os tramites legais para novas averiguações da diferença da Tabela 15. No ANEXO D é possível identificar os valores da Tabela 15 e Tabela 16, nos dias 21 e 22 de junho, respectivamente.

Tabela 16 - Demonstrativo dos Cálculo do IP do dia 10/06/2022

Valores distribuídos, conforme demonstrativo do ANEXO B

Data Oficial	IPI Municípios	FUNDEB Municípios	PASEP Municípios
10/06/2022	944.535,41	188.907,08	7.556,28

Valores distribuídos para o Município de São Luís, conforme IPM do exercício de 2021

Valor Líquido	IPI Bruto SLZ	FUNDEB Municípios	PASEP Municípios	Coeficiente IPM 2022
239.227,05	302.054,35	60.410,87	2.416,43	31,97914500

Fonte: Elaborado pela Autora (2023) com base nos depósitos dos ANEXOS B e D.

APÊNDICE G: Relatório de limpeza dos dados de São Luís

Norteando-se pelo Relatório de exploração dos dados do APÊNDICE E, o presente relatório tem o intuito de registrar as principais atividade executada na etapa de limpeza dos dados da fase de “Preparação dos dados” do CRISP-DM.

Ajustando os valores “ZEROS”

No Quadro 30 do APÊNDICE E foram elencados os atributos com elevado percentual de valores com ZEROS, ao investigar cada variável no APÊNDICE D: Relatório de descrição dos dados da primeira iteração com a Capital São Luís, inferiu-se que elas tratam da desagregação das exportações do Município, podendo ser removidas, haja vista que há uma variável (“EXPT_SLZ_TOTAL”) para o total das exportações, cuja série não possui valores “ZEROS”.

Ajustando os valores constantes

O apêndice em comento apresentou, no mesmo quadro, 3 valores como constantes:

- 1) - FatorPopulacao uma constante com valor de "2.5";
- 2) - PercInversoRendaPerCapitaMun constante com valor de "3.704";
- 3) - FatorInversoRendaPerCapitamunicipio outra constante com valor de "0.4";

Estes valores são utilizados no cálculo do coeficiente do FPM, o CIFPM, que somente é divulgado no final do ano, em meados de novembro para o exercício seguinte e o do IPM, publicado o provisório em julho e o definitivo até dezembro. Desta forma, inicialmente, pensou-se em fazer uma transformação do coeficiente anterior com a multiplicação destas constantes, mas optou-se por retirá-los por enquanto, uma vez que a primeira iteração estão apenas os dados do Município de São Luís e foi elaborado para uma previsão trimestral. Contudo, quando da inclusão das demais capitais eles deverão permanecer para averiguar o comportamento.

Ajustando a quantidade de registros das séries

Logo no início do Relatório do APÊNDICE E, foi verificado que os subconjuntos das Receitas não estão com suas séries completas, pois as observações deveriam ter a mesma

quantidade. Para dirimir tal fato, foram construídos agrupamento por ano e mês para cada receita, como o exemplo da Tabela 17 cuja visualização dos valores ausente ficou mais evidente.

Tabela 17 - Mensuração do IPI em uma *pivot table* com disposição anual, mês a mês da publicação contábil

Ano / Mês	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	324.774,33	464.345,80	179.210,76	NaN	NaN	NaN	458.609,09	388.318,24	457.341,78	526.239,00	482.262,65	687.264,52
2	287.851,66	372.337,58	380.934,87	692.316,35	623.983,05	NaN	355.391,64	424.136,95	468.640,16	569.535,89	521.199,64	558.950,13
3	255.448,07	312.539,63	336.757,58	263.070,23	278.861,68	NaN	346.974,25	335.542,95	335.294,54	NaN	480.991,45	663.711,26
4	270.447,90	389.353,56	352.702,86	308.950,36	304.074,38	NaN	356.024,87	402.287,66	472.876,63	518.565,73	161.614,16	974.688,23
5	308.042,22	330.190,24	404.440,25	343.625,39	330.730,93	796.392,39	404.294,90	380.161,85	425.299,56	546.038,09	67.235,40	515.333,88
6	295.901,57	365.029,65	361.310,65	357.895,08	315.003,70	252.060,52	240.573,82	382.228,69	389.897,85	572.639,19	331.041,36	718.284,99
7	326.064,00	348.867,32	336.936,02	352.535,58	323.606,89	224.504,65	326.872,95	423.729,09	470.170,32	579.695,77	464.941,94	681.692,47
8	365.904,76	359.982,33	372.420,14	377.215,40	333.567,14	238.883,42	330.302,02	3.035.996,80	425.064,95	518.832,80	537.712,98	573.245,53
9	295.425,82	333.713,02	376.478,63	289.678,74	313.903,97	91.632,94	361.700,51	429.175,71	425.004,51	632.470,13	631.608,48	682.070,39
10	333.801,52	465.084,63	364.017,53	420.251,84	336.164,79	287.326,12	377.501,27	-2.092.561,35	431.376,09	559.728,86	644.985,83	716.718,83
11	371.729,34	442.117,79	368.524,21	406.881,35	386.017,34	265.540,18	368.748,69	438.002,00	414.146,66	205.501,62	870.233,46	672.568,80
12	414.046,87	500.156,26	368.524,21	434.868,38	380.886,83	262.204,70	455.847,10	604.315,53	499.031,95	690.900,60	648.137,50	529.333,59
Total	3.849.438,06	4.683.717,81	4.202.257,71	4.247.288,70	3.926.800,70	2.418.544,92	4.382.841,11	5.151.334,12	5.214.145,00	5.920.147,68	5.841.964,85	7.973.862,62

Fonte: Elaborado pela Autora (2023).

Após uma pesquisa no portal da transparência do Governo Federal, os valores discrepantes ou intervalos faltantes foram identificados como segue:

- Em dezembro de 2011, foi registrado R\$ 500.156,34, mas o valor bruto⁴⁸ correto era R\$ 407.310,73 ⁴⁹;
- Em janeiro de 2012, o valor de R\$ 179.210,76 publicado no balancete deveria ser R\$ 184.612,98, pois janeiro e fevereiro foram pagos no mesmo dia de fevereiro⁴⁹;
- Em fevereiro de 2012, o montante de R\$ 380.934,87 também foi corrigido para R\$ 383.125,64 ⁴⁹;
- Em fevereiro de 2013, a importância de R\$ 682.316,35 representa a soma dos meses de janeiro e fevereiro. Assim, os valores corretos foram: para janeiro de R\$ 411.150,22 e o de fevereiro R\$ 277.901,61 ⁴⁹;
- Em fevereiro de 2014, o valor de R\$ 623.983,05, apresentou a mesma situação. sendo substituídos: em janeiro por R\$ 337.090,23 e fevereiro por R\$ 285.388,80 ⁴⁹;
- Em maio de 2015, a importância de R\$ 796.392,39, repete-se, desmembrando-se em janeiro com R\$ 413.358,15, em fevereiro com R\$ 232.871,93, em março com R\$ 228.423,18, em abril com R\$ 239.245,73 e em maio com R\$ 269.895,91. Neste caso,

⁴⁸ Trata-se do princípio do orçamento bruto, o qual recomenda o registro do ingresso das receitas sem deduções com seu valor total, prescrito no art. 6º da Lei nº 4.320, de 1964.

⁴⁹ Conforme legislação vigente, a receita deve ser registrada em seu valor bruto. Logo, o montante da divulgação do Demonstrativo de Distribuição da Arrecadação deve ser acrescido dos percentuais da legislação em cada ano do FUNDEB e o 1% do PASEP. Disponível em: <https://www42.bb.com.br/portalbb/daf/beneficiario,802,4647,4652,0,1,bbx>.

- diferente dos anteriores os valores publicados não conferem com o montante que ingressou em maio, ficando o mês de abril, sem indícios de ingresso de recursos⁴⁹;
- g) Em setembro de 2015, o valor de R\$ 91.632,94, não confere com o repasse divulgado de R\$ 188.453,12;
- h) Em 2017, nos meses de agosto ocorreu o ingresso de R\$ 2.092.561,35, o qual entende-se que foi algum tipo de erro, pois em outubro o mesmo valor foi estornado;
- i) Em março de 2019, também não foi possível encontrar registro de ingresso do repasse, mas foi colocado o valor divulgado, evitando a ausência de R\$ 550.047,33⁴⁹;
- j) Em abril de 2019, registrou-se R\$ 518.565,73, contudo o divulgado foi de R\$ 557.644,28⁴⁹;
- k) Em novembro de 2019, a série histórica exibiu R\$ 205.501,75, todavia o divulgado foi R\$ 587.796,59⁵⁰;
- l) Em abril de 2020, o montante, no balancete é de R\$ 161.614,16, contudo a publicação foi de R\$ 431.650,04⁵⁰;
- m) Em maio de 2020, encontrou-se o valor de R\$ 67.235,40, mas exposto foi de R\$ 430.382,25⁵⁰;
- n) Em junho de 2020, o total foi de R\$ 331.041,36, porém o divulgado foi de R\$ 387.039,21⁵⁰;
- o) Em dezembro de 2020, a soma do mês foi de R\$ 648.137,50, mas o valor publicado foi de R\$ 833.366,43⁵⁰.

Ao final dos ajustes a Tabela 17 foi substituída pela listagem da Tabela 18.

Tabela 18 – Ajuste da série temporal do IPI com disposição anual, mês a mês da publicação contábil

Ano / Mês	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	307.612,85	396.546,98	85.373,73	426.845,02	344.321,07	356.907,90	452.627,06	387.191,25	449.568,37	515.131,92	483.970,83	702.313,98
2	288.466,85	394.851,31	474.771,90	315.691,97	285.388,80	232.871,93	355.391,64	409.265,09	468.640,16	569.535,89	521.199,64	558.950,13
3	263.728,72	290.025,90	336.757,58	278.688,47	278.861,68	228.423,18	346.974,25	332.303,92	390.868,28	550.047,33	480.991,45	663.711,26
4	267.808,10	380.066,48	352.702,86	255.541,76	304.074,55	239.245,73	356.024,87	387.538,57	417.302,89	557.644,28	431.650,04	665.671,44
5	273.881,47	330.190,24	404.440,25	343.625,39	330.730,93	256.847,68	404.294,90	376.492,12	425.299,56	546.038,09	430.382,25	643.409,17
6	320.008,52	365.029,65	361.310,65	357.895,08	315.003,68	232.962,85	221.913,74	378.538,92	467.036,03	572.639,19	387.039,21	590.209,70
7	340.709,89	311.294,84	336.936,02	352.535,58	323.606,89	177.614,38	326.872,95	398.040,75	393.026,29	579.695,77	464.941,94	681.692,47
8	317.541,15	359.982,33	372.420,14	377.215,40	333.567,14	188.990,00	330.302,02	397.927,01	425.064,95	518.832,80	537.712,98	573.245,53
9	328.883,02	333.713,02	376.478,63	343.126,19	313.903,97	188.453,12	361.700,51	425.032,75	425.004,51	632.470,13	631.608,48	682.070,39
10	342.274,53	440.955,30	364.017,53	366.804,39	336.164,79	271.582,33	368.850,13	481.966,71	431.376,09	559.728,86	722.850,62	772.512,79
11	364.617,43	442.005,16	368.524,21	406.881,35	386.017,34	265.540,18	399.532,06	438.002,00	414.146,66	587.796,59	792.368,67	672.568,80
12	407.310,73	549.585,97	422.685,91	434.868,38	380.886,83	262.204,70	425.063,73	604.315,53	499.031,95	690.900,60	833.366,43	687.603,04
Total	3.822.843,26	4.594.247,18	4.256.419,41	4.259.718,98	3.932.527,67	2.901.643,98	4.349.547,86	5.016.614,62	5.206.365,74	6.880.461,45	6.718.082,54	7.893.958,70

Fonte: Elaborado pela Autora (2023).

⁵⁰ *Ibi Id.*

APÊNDICE H: Resumo do ranking de importância das variáveis do conjunto de dados de São Luís por receita

Variável exógena	Ranking das Variáveis							Descrição	Relação com a Receita						
	1	2	4	5	6	7	5		2	1	4	6	7		
PrclnvRndPrCpMn			31					Cálculo feito para o CIFPM. O valor é 1 dividido pela “RendaPerCapitaMunicípio” e o resultado dividido pelo total das Capitais multiplicado por 100	x		x				
PIB_MenBR	16	14	5	5	33			Produto Interno Bruto do Brasil, desmembrado da série trimestral pelo BACEN (método de Denton-Cholette).	x	x		x	x		
PIB_Mun	32	45			16			Produto Interno Bruto, a preços correntes, utilizou-se a proporção da sazonalidade do PIB Brasil Mensal para aplicar no PIB Municipal.	x	x		x	x		
RndPrCap_Mun	33	47			14			Derivação do cálculo da renda per capita do município, a preços correntes.	x			x			
VA_BrtAgro		1						Valor adicionado bruto da Agropecuária, a preços correntes do Município.	x	x		x	x		
VA_BrtIndust	38	54			12			Valor adicionado bruto da Indústria, a preços correntes do Município.	x	x		x	x		
VA_BrtServExclAdm	23	25			18			Valor adicionado bruto dos Serviços, a preços correntes – exceto Administração, defesa, educação e saúde públicas e seguridade social do Município.	x	x		x	x		
VA_BrtAdm	19	49			19			Valor adicionado bruto da Administração, defesa, educação e saúde públicas e seguridade social. a preços correntes do Município.	x	x		x	x		
VA_BrtTotal	31	37			17			Valor adicionado bruto total, a preços correntes do Município.	x	x		x	x		
VA_ImpostosProd	27	53			15			Impostos, líquidos de subsídios, sobre produtos, a preços correntes do Município.	x			x	x		
SalarioMinimo	14		16		28	6		Define o limite mínimo para os salários pagos na produção de bens e serviços, assim como benefícios previdenciários.	x				x		
DolarComercial	20	46	29		20	18		Média entre o maior e o menor valor cotado do dólar comercia no mês. Cotação do Dólar dos EUA. Cotações de fechamento de uma moeda em um período.	x	x		x	x		
PetoleoBrentBR		35	7	10				Preço médio do petróleo Brent, referência para formação de preço pela Petrobras. Cálculo da média entre a maior e menor cotação do mês, multiplicado pelo dólar.		x		x	x		
Gasolina_BR	8		6	22				Preço médio da Gasolina comum divulgado pela Agência Nacional de Petróleo (ANPM).		x		x	x		
ResvInternac	15	36	2	27	5			Reservas Internacionais, utilização em eventuais financiamentos de saldos negativos das contas externas e/ou estabilizar o mercado de câmbio.		x		x	x		
Importacao	5		18	12		7		Importação de bens: corresponde ao valor total de ingresso de mercadorias originárias do exterior para a unidade da federação especificada.				x	x		
EXPT_SLZ_ProdIndQuim	3	51	3	9				Valores dos Produtos das indústrias químicas ou indústrias conexas que são exportados pelo município de São Luís.	x	x			x		
EXPT_SLZ_TOTAL	9	24	4					Valor de todas as categorias de produtos exportados pelo município de São Luís.	x	x			x		
MetaisCoomuns	27	33			8	8		Exportação de metais comuns e suas obras.		x	x		x		
Prod_Minerais	25	8		11	17			Exportação de produtos minerais.		x	x		x		
TransacoesEspeciais		23		13	16			Exportação de produtos transações especiais.		x			x		
Exportacao	36	5	19		15			Exportação de bens: corresponde ao valor total das mercadorias vendidas ao exterior pela unidade da federação especificada.		x		x	x		
ICMS		11	28		32	12		Recolhimento do ICMS no Estado.					x		
ICMS_DvAtTrib		18	27		2	20		Valor da divida ativa tributária paga no Estado.					x		
ICMS_TOTAL	13	17	26		35	13		Soma do ICMS com Simples Nacional, mais as Substituições Tributárias Internas e Externas da arrecadação do Governo do Estado.					x		
ICMS_SimpNacio	4	16	23		29	14		Arrecadação do ICMS recolhido no Simples Nacional do Governo do Estado.					x		
IPVA	2	52	22	1				Arrecadação do IPVA recolhido no Estado.					x		
ST_INTERNO		50	24					Arrecadação de Substituição Tributária de movimento interna no Estado.					x		
ST_EXTERNO	12	4	25		34			Modalidade de Substituição Tributária de movimento externa no Estado.					x		
CFEM_Total		44		4				Arrecadação da CFEM de todos os Estado que possuem Município produtores.				x			
CFEM_MA	35	6	14					Arrecadação da CFEM de cada Estado somando todos os Municípios produtores.					x		
CFEM_Mun	11	29						Arrecadação por Município produtor, valores da CFEM recolhido bruto.					x		

Variável exógena	Ranking das Variáveis							Descrição	Relação com a Receita						
	1	2	4	5	6	7			5	2	1	4	6	7	
ICMS_DvA(Trb_Muns		37		2				Arrecadação do Governo do Estado efetivada em cada Município.					X		
IPVA_Muns		40						Arrecadação do IPVA recolhido no Município do Estado do Maranhão.						X	
ICMS_Muns	18	12		7	3			Arrecadação do ICMS recolhido no Município do Estado do Maranhão.						X	
TOTAL_ArrecMun	17	13		8	4			Total da arrecadação ocorrida em cada Município do Estado.						X	
Veic_Pgt(IPVA)		32						Quantidade de veículos emplacados no Município.							X
TxDescupados		31	9	19	13			Total de pessoas sem trabalho, procurando emprego, de 14 anos ou mais de idade.							X
TxIPCA	6	21	8	9				Índice de Preço ao Consumidor Amplo.							X
Tx_SELIC_Perd	10	55	15			4		Taxa que o governo define o “custo do dinheiro” na economia brasileira, baliza as taxas das instituições financeiras.	X	X	X	X	X	X	X
IGPM		30	21	15	5	10		Índice Geral de Preços – Mercado.							X
INPC		15	14			3		Índice Nacional de Preços ao Consumidor variação mensal.							X
IBC_Br		34	20	20		9		Índice de Atividade Econômica do Banco Central (IBC-Br) reflete as atividades da indústria, dos serviços e da agropecuária.	X	X	X				X
IPP	7	28	6					Índice de Preços ao Produtor, indústria geral, extrativas e de transformação.							X
PereResSP_Nom	29				31			Percentual do resultado nominal, soma do resultado primário com os juros a paga.							
SINAP_Alto	26	43	13		26			Custo de projeto por m², por tipo de projeto e padrão de acabamento Alto.	X	X					X
SINAP_Baixo	21	41	12		25			Custo de projeto por m², por tipo de projeto e padrão de acabamento Baixo.	X	X					X
SINAP_Mínimo	22	38	11		23			Custo de projeto por m², por tipo de projeto e padrão de acabamento Mínimo.	X	X	X				X
SINAP_Normal	24	42	10		24			Custo de projeto por m², por tipo de projeto e padrão de acabamento Normal.	X	X	X				X
Coef_CIFPM_SLZ		3			10			Coefficiente individual de distribuição do FPM para o município do São Luis.	X				X		
Coef_IPM	34	10			21			Coefficiente de distribuição Estadual da receita do ICMS de São Luis.		X					X
FinMA_MnRestPrim		19			17			Resultados primários, componente não financeiros do resultado fiscal da dívida do Município, resultado da subtração do resultado com os juros, ambos nominais.	X						
FinMA_Mn_JrNom		26						Fluxo dos juros, sobre a dívida interna e externa valores nominais do Município.							
FinMA_MnRestNom		20			18			Corresponde à variação nominal dos saldos da dívida líquida, é composta pelos juros reais, dívida atualizada e resultado primário do Município.	X						
Ajustes_MA_Munic	1	48						Outros fluxos (ajustes) do Estado do MA e municípios, valores em R\$ (milhões)	X						
DvLq_Gov_MA	30	22			30			Dívida líquida do Governo do Estado do Maranhão	X						
DvLq_MA_ExcCapital		9		3				Dívida líquida dos municípios do Estado do Maranhão, exceto a capital	X						
DvLq_Sao_Luis_MA		7		11	6			Dívida líquida do município de São Luis do Estado do Maranhão	X						
RECEITA								Categoria da Série Temporal, o Tipo da Receita: CFEM, FEP, ITR, e outras,	X	X	X	X	X	X	X
x_1			30	2	7	19		Valor da Receita do atributo “Realizado” no mês anterior,	X	X	X	X	X	X	X
x_2			2	17		2		Valor da Receita do atributo “Realizado” de 2 meses anteriores,	X	X	X	X	X	X	X
x_3	28	27	1	16		11		Valor da Receita do atributo “Realizado” de 3 meses anteriores,	X	X	X	X	X	X	X
Realizado								Variável “target”. Valor Bruto da Receita executada nos Balançetes da Receita.	X	X	X	X	X	X	X

Observações:

- 1) A coluna “Ranking das variáveis” especifica a posição da importância da variável para a Receita.
- 2) A coluna “Relação com a Receita” identifica que a variável foi incluída por ter algum tipo de relação com a formação da receita.
- 3) As receitas foram codificadas para reduzir o tamanho da coluna, logo: 1=CFEM, 2=IPI, 4=FEP, 5=FPM, 6=ICMS e 7=IPVA.

ANEXOS

ANEXO A - Demonstrativo de distribuição da Arrecadação do Governo do Estado do Maranhão do mês de janeiro de 2022

27/07/2022 11:10

[bb.com.br]



(http://www.bb.com.br)

DEMONSTRATIVO DE DISTRIBUIÇÃO DA ARRECAÇÃO

27/07/2022

SISBB - Sistema de Informações Banco do Brasil

10:55:13

MARANHÃO - MA

IPI - IPI - EXPORTACAO

DATA	PARCELA	VALOR DISTRIBUIDO
10.01.2022	IPI - ESTADO	R\$ 4.315.013,66 C
	IPI-MUNICIPIOS	R\$ 1.438.337,89 C
	PASEP ESTADO	R\$ 34.520,10 D
	PASEP MUNICIPIO	R\$ 11.506,70 D
	FUNDEB ESTADO	R\$ 863.002,73 D
	FUNDEB MUNICIP	R\$ 287.667,57 D
	TOTAL:	R\$ 4.556.654,45 C
20.01.2022	IPI - ESTADO	R\$ 1.811.746,74 C
	IPI-MUNICIPIOS	R\$ 603.915,58 C
	PASEP ESTADO	R\$ 14.493,97 D
	PASEP MUNICIPIO	R\$ 4.831,32 D
	FUNDEB ESTADO	R\$ 362.349,34 D
	FUNDEB MUNICIP	R\$ 120.783,11 D
	TOTAL:	R\$ 1.913.204,58 C
28.01.2022	IPI - ESTADO	R\$ 219.741,00 C
	IPI-MUNICIPIOS	R\$ 73.247,00 C
	PASEP ESTADO	R\$ 1.757,92 D
	PASEP MUNICIPIO	R\$ 585,97 D
	FUNDEB ESTADO	R\$ 43.948,20 D
	FUNDEB MUNICIP	R\$ 14.649,40 D
	TOTAL:	R\$ 232.046,51 C
TOTAIS	IPI - ESTADO	R\$ 6.346.501,40 C
	IPI-MUNICIPIOS	R\$ 2.115.500,47 C
	PASEP ESTADO	R\$ 50.771,99 D
	PASEP MUNICIPIO	R\$ 16.923,99 D
	FUNDEB ESTADO	R\$ 1.269.300,27 D
	FUNDEB MUNICIP	R\$ 423.100,08 D
	DEBITO FUNDO	R\$ 1.760.096,33 D

ANEXO B - Demonstrativo de Distribuição da Arrecadação do Governo do Estado do Maranhão do mês de junho de 2022

27/07/2022 11:12

[bb.com.br]



(http://www.bb.com.br)

DEMONSTRATIVO DE DISTRIBUIÇÃO DA ARRECAÇÃO

27/07/2022

SISBB - Sistema de Informações Banco do Brasil

11:04:05

MARANHÃO - MA

IPI - IPI - EXPORTACAO		
DATA	PARCELA	VALOR DISTRIBUIDO
10.06.2022	IPI - ESTADO	R\$ 2.833.606,22 C
	IPI-MUNICIPIOS	R\$ 944.535,41 C
	PASEP ESTADO	R\$ 22.668,84 D
	PASEP MUNICIPIO	R\$ 7.556,28 D
	FUNDEB ESTADO	R\$ 566.721,24 D
	FUNDEB MUNICIP	R\$ 188.907,08 D
	TOTAL:	R\$ 2.992.288,19 C
20.06.2022	IPI - ESTADO	R\$ 1.372.190,95 C
	IPI-MUNICIPIOS	R\$ 457.396,98 C
	PASEP ESTADO	R\$ 10.977,52 D
	PASEP MUNICIPIO	R\$ 3.659,17 D
	FUNDEB ESTADO	R\$ 274.438,19 D
	FUNDEB MUNICIP	R\$ 91.479,39 D
	TOTAL:	R\$ 1.449.033,66 C
30.06.2022	IPI - ESTADO	R\$ 698.568,29 C
	IPI-MUNICIPIOS	R\$ 232.856,10 C
	PASEP ESTADO	R\$ 5.588,54 D
	PASEP MUNICIPIO	R\$ 1.862,84 D
	FUNDEB ESTADO	R\$ 139.713,65 D
	FUNDEB MUNICIP	R\$ 46.571,22 D
	TOTAL:	R\$ 737.688,14 C
TOTAIS	IPI - ESTADO	R\$ 4.904.365,46 C
	IPI-MUNICIPIOS	R\$ 1.634.788,49 C
	PASEP ESTADO	R\$ 39.234,90 D
	PASEP MUNICIPIO	R\$ 13.078,29 D
	FUNDEB ESTADO	R\$ 960.873,08 D
	FUNDEB MUNICIP	R\$ 326.957,69 D
	DEBITO FUNDO	R\$ 1.360.143,96 D

https://www42.bb.com.br/portaibb/daf/demonstrativo,802,4647,4652,0,1,1,1.bbx?cid=1863041

1/2

ANEXO C - Demonstrativo de Distribuição da Arrecadação do Governo Do Municipal de São Luís do mês de janeiro de 2022

27/07/2022 11:12

[bb.com.br]



(http://www.bb.com.br)

DEMONSTRATIVO DE DISTRIBUIÇÃO DA ARRECAÇÃO

27/07/2022

SISBB - Sistema de Informações Banco do Brasil

11:05:12

SAO LUIS - MA

IPM - IPI EXPORTACAO - COTA MUNICIPIO		
DATA	PARCELA	VALOR DISTRIBUIDO
20.01.2022	IPM IPI-MUNIC	R\$ 152.956,59 C
28.01.2022	IPM IPI-MUNIC	R\$ 18.551,62 C
TOTAIS	IPM IPI-MUNIC	R\$ 171.508,21 C
	DEBITO FUNDO	R\$ 0,00 D
	CREDITO FUNDO	R\$ 171.508,21 C
TOTAL DOS REPASSES NO PERIODO		
	DEBITO BENEF.	R\$ 0,00 D
	CREDITO BENEF.	R\$ 171.508,21 C

ANEXO D - Demonstrativo de Distribuição da Arrecadação do Governo Do Municipal de São Luís do mês de junho de 2022

27/07/2022 11:13

[bb.com.br]



(http://www.bb.com.br)

DEMONSTRATIVO DE DISTRIBUIÇÃO DA ARRECAÇÃO

27/07/2022

SISBB - Sistema de Informações Banco do Brasil

11:06:07

SAO LUIS - MA

IPM - IPI EXPORTACAO - COTA MUNICIPIO		
DATA	PARCELA	VALOR DISTRIBUIDO
20.06.2022	IPM IPI-MUNIC	R\$ 115.847,12 C
21.06.2022	IPM IPI-MUNIC	R\$ 364.294,86 C
22.06.2022	IPM IPI-MUNIC	R\$ 239.227,05 C
30.06.2022	IPM IPI-MUNIC	R\$ 58.976,62 C
TOTAIS	IPM IPI-MUNIC	R\$ 778.345,65 C
	DEBITO FUNDO	R\$ 0,00 D
	CREDITO FUNDO	R\$ 778.345,65 C
TOTAL DOS REPASSES NO PERIODO		
	DEBITO BENEF.	R\$ 0,00 D
	CREDITO BENEF.	R\$ 778.345,65 C