

APLICANDO CONCEITOS DE MACHINE LEARNING EM ÓRGÃOS PÚBLICOS UTILIZANDO SÉRIES TEMPORAIS: UMA ANÁLISE FINANCEIRA DA PREFEITURA DE ADAMANTINA

GUILHERME CHUMAN TRONCON¹
FABRÍCIO DA SILVA MALACHIAS¹
ELIANE VENDRAMINI DE OLIVEIRA¹

RESUMO

A análise de dados envolve a manipulação de registros que, à primeira vista, podem parecer confusos ou carentes de contexto. Contudo, por meio de um processamento adequado e uma observação minuciosa, torna-se possível gerar informações que ajudam gestores e proprietários de organizações a tomar decisões mais embasadas. Esses procedimentos são conduzidos pelo cientista de dados, um profissional responsável por analisar registros no ambiente, compreender o seu funcionamento e manipulá-los para gerar insights e soluções de mercado inteligentes para o futuro das organizações. Os dados disponíveis podem ser analisados de diversas formas, desde a detecção de padrões até o rastreamento de tendências, tudo dependendo do contexto e do ambiente nos quais os dados estão inseridos. Aplicando técnicas e conceitos adequados, o cientista de dados é capaz até mesmo de "prever o futuro", mesmo em meio à incerteza inerente ao universo. Buscando demonstrar o impacto e a relevância do cientista de dados, o presente artigo apresenta um estudo de caso realizado na prefeitura de Adamantina, onde, após a coleta dos registros de despesas orçamentárias, pretende-se realizar uma análise financeira para entender o histórico de gastos do órgão público e apresentá-lo de forma clara e objetiva tanto para os membros da administração pública quanto para o cidadão comum. Além disso, espera-se que, por meio desse entendimento, seja possível realizar projeções para o futuro.

Palavras-chave: Análise financeira; Administração pública; Séries temporais.

ABSTRACT

Data analysis involves manipulating records that, at first glance, may seem confusing or lacking context. However, through proper processing and careful observation, it becomes possible to generate information that helps managers and owners of organizations make more informed decisions. These procedures are conducted by the data scientist, a professional responsible for analyzing records in the environment, understanding their functioning, and manipulating them to generate insights and intelligent market solutions for the future of organizations. The available data can be analyzed in various ways, from pattern detection to trend tracking, all depending on the context and environment in which the data is embedded. By applying appropriate techniques and concepts, the data scientist is even able to "predict the future," even amidst the uncertainty inherent in the universe. Seeking to demonstrate the impact and relevance of the data scientist, this article presents a case study conducted at the Adamantina city hall, where, after collecting records of budgetary expenses, the aim is to carry out a financial analysis to understand the history of public spending and present it in a clear and objective way to both members of the public administration and the average citizen. In addition, it is expected that, through this understanding, it will be possible to make projections for the future.

Keywords: Financial analysis; Public administration; Time series.

¹ Faculdade de Tecnologia de Adamantina (Fatec), Adamantina-SP, Brasil

1 INTRODUÇÃO

A tecnologia está cada vez mais integrada às organizações, e o domínio de ferramentas tecnológicas torna-se fundamental para se adaptar a essa nova realidade. Maestri et al. (2021) realizou um estudo comparativo entre as etapas da revolução industrial com foco na 3.5, 4.0 e 5.0 e é notável a presença de termos galgados pela computação como “Inteligência Artificial” e “Big Data” ainda na indústria 3.5 cujo artigo sugere datar do ano de 2017. Logo é nítido o interesse por automatizar e otimizar ainda mais as tarefas feitas pelo homem. Um recurso amplamente utilizado para alcançar tal objetivo é a análise de dados por meio do aprendizado de máquina.

O aprendizado de máquina teve início nos anos 60 como um campo da inteligência artificial, mas ao longo dos anos, especializou-se, tornando-se um campo próprio com o objetivo principal de identificar padrões no universo e aprender a partir dos dados. Existem inúmeras aplicações para o aprendizado de máquina em diversas áreas do conhecimento, como medicina, contabilidade e estatística.

Uma aplicação notável dos algoritmos de aprendizado de máquina é na análise de séries temporais. Sezer, Gudelek e Ozbayoglu (2020) corroboram com essa afirmação por meio de sua análise literária ocorrida no período de 2005 a 2019 onde afirmam o entusiasmo do cenário financeiro pela previsão de séries temporais por meio de algoritmos de aprendizado de máquina. Por meio de modelos matemáticos, é possível utilizar dados do passado para prever o futuro. Em um mundo onde a tecnologia promove mudanças constantes, o uso dessas informações é essencial para tomadas de decisão assertivas em prol das organizações.

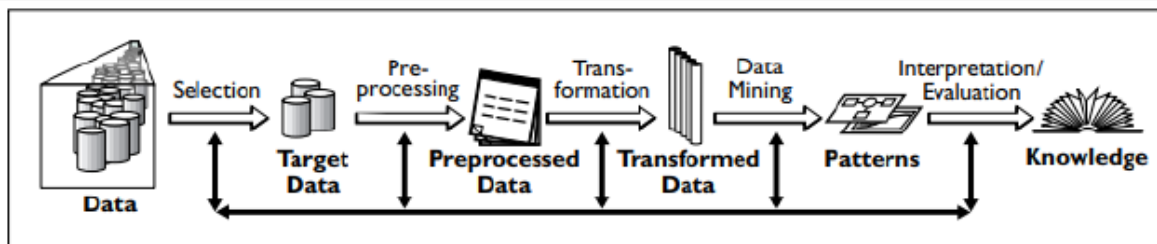
Nesse contexto, este trabalho tem como objetivo principal utilizar algoritmos de aprendizado de máquina para analisar as contas públicas da Prefeitura de Adamantina no período de 2017 a 2023, onde propõe-se a construção de séries temporais geradas a partir das mesmas com o objetivo de fornecer uma compreensão clara sobre o funcionamento e a evolução de tais dados e, com base nos mesmos, realizar projeções futuras para o final do ano de 2023.

2 REFERENCIAL TEÓRICO

2.1 KDD (Knowledge Discovery in Databases)

O KDD (*Knowledge Discovery in Databases*) foi uma metodologia sugerida por Fayyad et al em 1996 e consiste em um processo de cinco etapas para extração de conhecimento: Seleção, Pré-processamento, Transformação, Mineração de Dados e Interpretação/Avaliação.

Figura 1: Visão geral das etapas que constituem o processo KDD



Fonte: Fayyad et al, 1996

As etapas da Seleção a Transformação consistem no tratamento dos dados e sua preparação para ser submetida aos métodos de processamento desejados, Steiner et al (2006) definiu as três primeiras etapas do KDD como a parte de análise exploratória enquanto a etapa de Mineração é de fato a aplicação dos algoritmos para enfim chegar à descoberta de conhecimento que a metodologia sugere à partir da etapa de Interpretação/Avaliação.

2.2 Aprendizado de Máquina

De acordo com Izbicki (2020) o aprendizado de máquina nasceu na década de 60 como um campo da inteligência artificial e por isso estava relacionado exclusivamente com o meio computacional, porém com o passar dos anos e o refinamento das técnicas, o aprendizado de máquina ganhou seu espaço como um estudo em si para além da inteligência artificial e possui as mais diversas aplicações inclusive no ramo da estatística.

Mahesh (2018) define o aprendizado de máquina como uma ferramenta utilizada para ensinar a máquina a lidar com os dados de forma mais eficiente. O propósito do aprendizado de máquina é aprender a partir dos dados. Mahesh também dividiu os algoritmos de aprendizagem de máquina em oito tipos:

Supervisionados, Não-Supervisionados, Semi-supervisionados, Por Reforço, Multitarefa, Em Conjunto, Redes Neurais e Baseado em Instâncias, porém existem autores que trabalham com uma visão mais sucinta dos tipos de aprendizado de máquina os dividindo apenas em três categorias: Supervisionados, Não-Supervisionados e Por Reforço.

Polizeli (2019) descreve os algoritmos Supervisionados como aqueles onde há a disponibilidade de um conjunto de dados com classes definidas para treinamento, já os algoritmos Não Supervisionados não contam com tais classes pré-definidas e realiza a classificação dos dados identificando padrões e tendências seguindo parâmetros pré-definidos. Quanto aos algoritmos de aprendizado por reforço, Fontana (2020) estabelece um aprendizado baseado em recompensa onde há um agente observador que toma decisões baseado em tal observação e recebe um retorno do ambiente, de acordo com o retorno obtido o agente então toma novas decisões.

2.3 Séries temporais

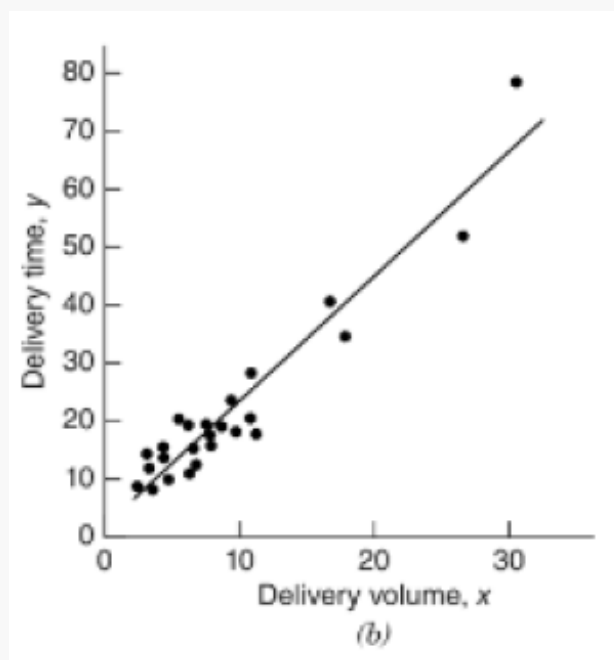
Uma das aplicações para o aprendizado de máquina são as séries temporais que, de acordo com Morettin (2018), são definidas por um conjunto de observações ordenadas no tempo, ou seja, são registros espaçados e constantes que contam uma história e por meio da observação de padrões é possível prever os próximos registros dessa história.

Oliveira (2007) afirma que para analisar séries temporais, estas são submetidas a modelos matemáticos para identificar as variações dentro da série de dados analisada detectando padrões a fim de quantificar e compreender o fenômeno de tal variação temporal e assim entender o passado para prever o futuro. Oliveira segue afirmando que dentre as principais informações a serem extraídas de uma série temporal estão as tendências, ciclos, estações, irregularidades e relações. As tendências dizem respeito ao padrão detectado que são utilizadas como base para previsão do comportamento dos dados, porém não se pode reduzir a análise a meios tão simplistas. A identificação de ciclos e estações são fundamentais para entender alguns agrupamentos de dados que se repetem em determinados momentos formando os chamados períodos sazonais ou sazonalidades. Já as irregularidades tratam-se de *outliers* que devem ser levados em conta na hora de se realizar a previsão e por fim as relações entre os dados identificam como

as demais variáveis podem influenciar diretamente nos resultados.

Uma das maneiras de se trabalhar com séries temporais porém não tão popular é a regressão linear simples, essa técnica estatística consiste em verificar a relação de dependência direta entre duas variáveis, Montgomery et. al (2021) apresenta o modelo da regressão linear por uma função afim: $y = \beta_0 + \beta_1 x$ onde β_0 corresponde à variável independente, ou seja, os valores previsores que serão inseridos, y corresponde à variável dependente, ou seja, o valor inserido em x determina o valor que y terá sendo este o nosso valor previsto, já β_0 e β_1 são valores constantes que determinam a altura e a inclinação da reta, β_0 é o valor de intercepção entre todos os valores e conhecidos e diz o ponto onde a reta da função passa pelo eixo y , já β_1 é o valor de declive entre x e y e define a inclinação que a reta terá, a imagem a seguir apresenta um exemplo prático de como a função funciona na prática:

Figura 2: Exemplo de modelo linear representado por uma função $y = \beta_0 + \beta_1 x$



Fonte: Adaptado de Montgomery et. al (2021)

O modelo não é tão popular pois não lida bem com *outliers*, valores muito dispersos que podem comprometer a tendência observada, as sazonalidades também podem ser um problema. Existem diversas alternativas para lidar com tais problemas como o ARIMA (Auto-Regressive Integrated Moving Averages) que trabalha com médias móveis, o algoritmo do Profeta do Facebook que se baseia em um modelo de regressão aditiva podendo utilizar crescimentos aditivos ou logísticos e até mesmo o uso de Redes Neurais, Ning, Kazemi e

Tahmasebi (2022) realizaram um estudo comparativo entre esses três métodos na previsão de produção de óleo na bacia de Denver-Julesburg, embora seus resultados apontem para uma melhor performance do ARIMA e das redes neurais LSTM (*Long-Short-Term Memory*), os autores afirmam que todos os modelos foram capazes de prever a produção e que, no caso do Profeta, talvez a base não dependa tão somente da sazonalidade algo que conta muito para este algoritmo.

2.4 Despesas públicas

O termo “Despesa Pública” consta no glossário de termos orçamentários do congresso nacional como: “Aplicação de recursos públicos com o fim de atender a uma necessidade do ente federado, podendo ser de natureza extraorçamentária ou orçamentária” (CONGRESSO NACIONAL, 2023). Jund Fo (2008) aprofundou esse termo como um conjunto de despesas oriundas do Estados ou de qualquer pessoa de Direito Público sendo assim parte do orçamento.

Conforme foi observado, as despesas públicas, ainda que fazendo parte do orçamento público, são divididas em “Despesas Orçamentárias” e “Despesas Extraorçamentárias”. Ambos os termos também constam no glossário de termos orçamentários do congresso nacional (2023), tratando as despesas orçamentárias como as despesas previstas na Lei Orçamentária Anual (LOA), uma lei padrão que é publicada todo ano um pouco antes do início do próximo exercício e que contém os dispêndios essenciais para manutenção dos entes públicos tais como recursos humanos, custeio, manutenção e ampliação dos serviços públicos prestados à sociedade. Já as despesas extraorçamentárias são aquelas que não exigem autorização, pois são dívidas imprevisíveis como devoluções e restos a pagar.

Dada a definição apresentada, entende-se que o congresso nacional planeja os gastos orçamentários para o exercício seguinte. Logo, ainda que a organização esteja sujeita a diversas variáveis imprevisíveis que impactam diretamente nos gastos públicos, entende-se que certos gastos ordinários fazem parte de uma rotina financeira consideravelmente previsível e desta forma as despesas orçamentárias podem ser interpretadas como um registro consistente. Assim é elegível a ser interpretado na forma de séries temporais para ser aplicada a modelos de aprendizado de máquina.

3 METODOLOGIA

Ao longo da história inúmeros autores se utilizaram do KDD como metodologia fundamental para trabalhar com aprendizado de máquina. Costa (2020) utilizou-se do KDD para avaliar a utilização de informações contábeis como geradoras de conhecimento para as empresas na qual o autor conclui que o uso de metodologias como o KDD é promissor na geração de conhecimento aliado ao manuseio de novas tecnologias para automatizar e agilizar o processo de análise dos dados. Da mesma forma, o KDD foi utilizado dentro do presente trabalho para a seleção da base adequada, bem como para o tratamento dos mesmos até o ponto que estivessem prontos para passarem pelos algoritmos indicados e se então avaliar os resultados obtidos.

A extração dos dados se deu por meio de consultas ao portal de transparência da prefeitura de Adamantina disponível em <<http://transparenciadamantina.sp.gov.br>> até a data deste artigo, foi necessário realizar os downloads dos arquivos mês a mês, exercício por exercício, foram baixados um total de 79 arquivos em formato .CSV (Comma-separated values) que correspondem aos lançamentos mensais do período de janeiro de 2017 até julho de 2023, abaixo a imagem demonstra parte desse processo:

Figura 3: Fonte de extração dos dados da prefeitura de Adamantina



Fonte: Autoria Própria

Após o download dos arquivos for concluído, uma análise inicial foi realizada, visualizando os dados por meio do aplicativo Google Planilhas, observa-se uma formatação peculiar com várias linhas em branco e colunas dentro de colunas, acredita-se que o formato ocorre de tal maneira para ser recebido por geradores

de PDF que traduzir os dados no formato de documento formal que se costuma ver em organizações tanto de domínio público quanto privado.

Figura 4: Primeira visualização dos dados realizada no Google Planilhas

	A	B	C	D	E	F	G	H	I	J	K	L
1	Despesa	Descrição	Fun. Res.	Ficha	Exercício	Atencião	Atual	Saldo Atual	Anterior	No Mes	Total	A Propos
2	EXECUTIVO MUNICIPAL											
3	2 EXECUTIVO MUNICIPAL											
4	2.01 SECRETARIA DE GABINETE											
5	02.01.01											
6	04.122.9011.1009	MELHORIA CONTINUA DE EQUIPAMENTOS VEICULOS, MAQUINAS E UTENSILIOS										
7	4.4.90.52	EQUIPAMENTOS E	1	1	1000	-1000	0	0	0	0	0	0
8	04.122.9011.1006	PLANO DE DESENVOLVIMENTO DA TECNOLOGIA DE INFORMACAO										
9	4.4.90.52	EQUIPAMENTOS E	1	2	1000	-1000	0	0	0	0	0	0
10	04.122.9011.2701	PUBLICACOES E DIVULGACOES										
11	3.3.90.39	OUTROS SERVICOS	1	3	18000	38200	215200	24639.93	216920.41	-24460	193960.41	
12	1 ASSINATURAS DE PERIODICOS E ANUIDADES								11200	0	11200	
13	47 SERVICOS DE COMUNICACAO EM GERAL								18213.2	0	18213.2	
14	59 SERVICOS DE ALUGU VIDEOS E FOTO								790	0	790	
15	88 SERVICOS DE PUBLICIDADE E PROPAGANDA								115760	-24460	91300	
16	90 SERVICOS DE PUBLICIDADE LEGAL								72987.21	0	72987.21	
17	04.122.9011.2592	DESPESAS SOB REGIME DE ADIANTAMENTOS										
18	3.3.90.14	DIARIAS - PESSOAL	1	4	60000	20000	80000	4235.73	76100.65	-336.18	75764.27	

Fonte: Autoria Própria

A fim de tratar a informação corretamente, os arquivos foram tratados no editor de código Visual Studio Code onde foram removidas todas as linhas em branco e colunas que não faziam parte do objeto de estudo, é importante destacar que, para essa análise, foram utilizados exclusivamente os registros de despesas orçamentárias, uma vez que se trata de um ambiente mais controlado. Reconhece-se que o mercado financeiro é extremamente complexo, e assuntos relacionados a finanças, como gastos e custos, tendem a ser influenciados de maneira semelhante. Assim, a escolha pelos lançamentos orçamentários foi uma alternativa para lidar com as incertezas inerentes ao objeto de estudo, logo restaram apenas as colunas: “Despesa”, que corresponde ao código da despesa, “Descrição”, que corresponde ao setor responsável pela despesa ou a descrição da despesa em si, “Exercício”, dado este que não constava no arquivo original pois o mesmo já correspondia a um período específico sendo assim adicionado para que os arquivos fossem combinados futuramente sem se perder o registro do tempo em que ocorreu o gasto e por fim “Total” que corresponde ao valor gasto, é possível conferir o resultado final do arquivo formatado na figura a seguir:

Figura 5: Resultado da formatação e limpeza dos dados iniciais

Despesa, Descrição, Exercício, Total
EXECUTIVO MUNICIPAL,, 01/21,
02, EXECUTIVO MUNICIPAL, 01/21,
02.01, SECRETARIA DE GABINETE, 01/21,
02.01.00,, 01/21,
04.122.0011.1009, "RENOVACAO DA FROTA MUNICIPAL", 01/21,
4.4.90.52, EQUIPAMENTOS E MATERIAL PERMANENTE, 01/21, 0.00
04.122.0011.1036, PLANO DE DESENVOLVIMENTO DA TECNOLOGIA DE INFORMACAO, 01/21,
4.4.90.52, EQUIPAMENTOS E MATERIAL PERMANENTE, 01/21, 0.00
04.122.0011.2101, PUBLICACOES E DIVULGACOES, 01/21,
3.3.90.39, OUTROS SERVICOS DE TERCEIROS - PESSOA JURIDICA, 01/21, 5269.18
47, SERVICOS DE COMUNICACAO EM GERAL, 01/21, 3469.18
88, SERVICOS DE PUBLICIDADE E PROPAGANDA, 01/21, 1000.00
90, SERVICOS DE PUBLICIDADE LEGAL, 01/21, 0.00
04.122.0011.2102, DESPESAS SOB REGIME DE ADIANTAMENTOS, 01/21,
3.3.90.14, DIARIAS - PESSOAL CIVIL, 01/21, 0.00

Fonte: Autoria Própria

Por fim, os arquivos foram enviados ao ambiente de desenvolvimento Google Colab, um software online livre para construção do Dataframe que seria submetido aos algoritmos, para tal construção utilizou-se de laços de repetição que percorrer todos os registros identificando padrões nos registros, percebeu-se que os setores responsáveis continham o dígito “02.” e as linhas subsequentes continham os gastos dos mesmos, logo essa verificação foi utilizada para unificar as despesas por setor e por ano e assim foi construído o Dataframe final conforme a figura abaixo:

Figura 6: Construção do Dataframe para análise

```
for key, values in data_by_month_and_sector_lite.items():
    for k, v in values.items():
        s = 0
        for i in v:
            s += i
        data_by_month_and_sector_lite[key][k] = float("%.2f" % s)
print(data_by_month_and_sector_lite)
```

```
df = pd.DataFrame(data_by_month_and_sector_lite)
df2 = df.drop(columns=['08/22', '09/22', '10/22', '11/22', '12/22'])
df2.head()
```

	01/17	02/17	03/17	04/17	05/17	06/17	07/17	08/17	09/17	10/17	...	10/22	11/22	12
02.01 - SECRETARIA DE GABINETE	5270.71	53734.02	125724.62	185083.96	249509.60	311340.71	371957.25	444257.73	507729.84	601197.63	...	1033188.77	1216606.15	1433322
02.02 - SECRETARIA DE ASSUNTOS JURIDICOS	125063.45	341040.29	547265.78	715096.67	913014.08	1164008.12	1379250.09	1628826.46	1937161.22	2164581.89	...	22758.54	22758.54	2275
02.03 - SECRETARIA DE PLANEJAMENTO	17326.93	114346.36	232720.96	317833.82	400989.65	490824.07	595996.58	704145.11	790336.35	875896.96	...	1963109.09	2217006.04	261954
02.04 - SECRETARIA DE FINANÇAS	95931.61	348939.60	576850.02	795071.91	1048600.04	1245251.01	1436922.94	1660249.36	1902296.58	2234829.88	...	4030061.24	4880078.22	544037
02.05 - SECRETARIA DE ADMINISTRACAO	121620.72	782813.29	1435088.45	2106349.89	2804095.77	3495839.45	4173205.63	4889465.76	5674963.46	6293334.19	...	7784574.53	8962428.53	1109164

25 rows x 15 columns

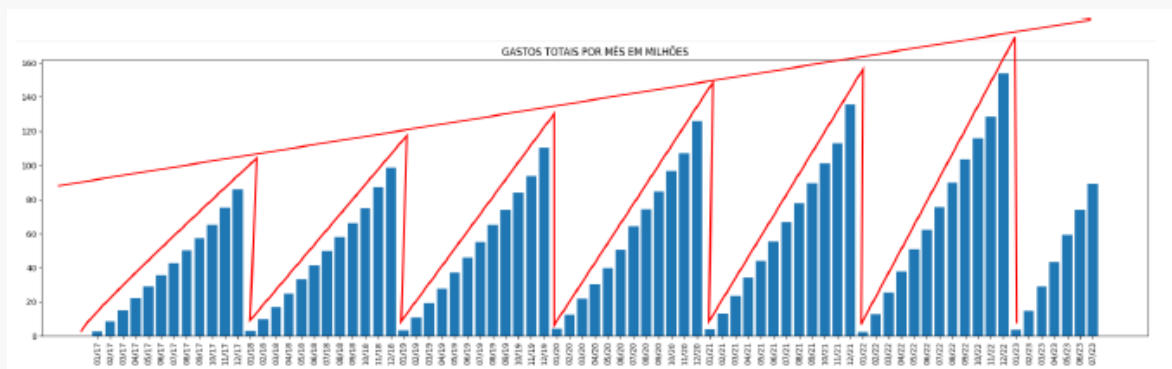
Fonte: Autoria Própria

Já a etapa de mineração passou por 2 fases, tabulação e processamento. Durante a etapa de tabulação os dados foram organizados e dispostos de forma gráfica para tentar identificar padrões e demonstrar a evolução das contas públicas e geração das séries temporais em forma de listas. Na segunda etapa as séries temporais extraídas foram submetidas aos algoritmos de Regressão Linear e Profeta do Facebook para realização de prospecções.

4 RESULTADOS

A primeira análise realizada foi referente aos histórico de gastos totais mês a mês, inicialmente é possível ver que os gastos se comportam de forma cíclica, crescendo gradualmente ao longo do ano e depois retornam ao seu ponto inicial no começo do próximo ano, trazendo um formato semelhante ao de um “serrote” que foi marcado em destaque na Figura 7 para melhor visualização, observando os dados mais atentamente, percebe-se que os valores crescem proporcionalmente de um ano para o outro, demonstrando o aumento natural dos preços segundo a inflação que impacta nos gastos orçados.

Figura 7: Gastos totais por mês

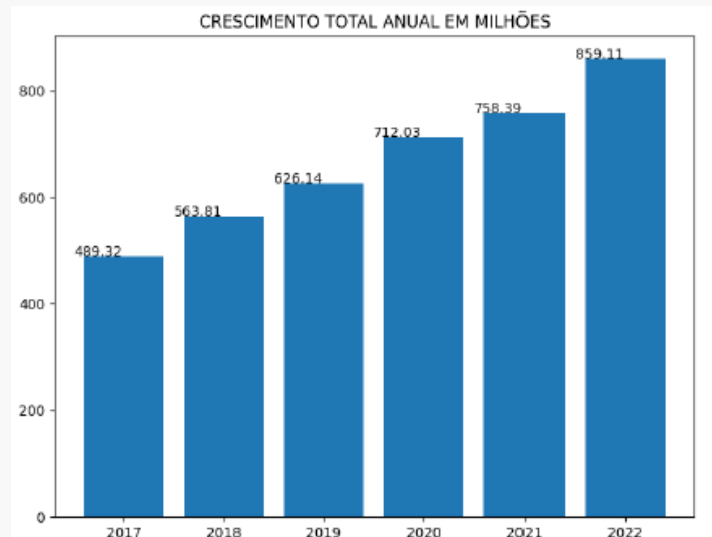


Fonte: Autoria Própria

Por conta do formato dos dados apresentado, a Regressão Linear simples não obteve um resultado satisfatório, obtendo apenas 10% de coeficiente de determinação, valor este responsável por indicar o quanto as variáveis estão relacionadas, o valor extremamente baixo já era esperado uma vez que o algoritmo não trabalha bem com sazonalidades e traça uma linha de tendência completamente desconexa do contexto averiguado na prefeitura, para contornar este problema foi calculada a média de gastos anuais, gerando uma

nova série de dados apresentada na Figura 8:

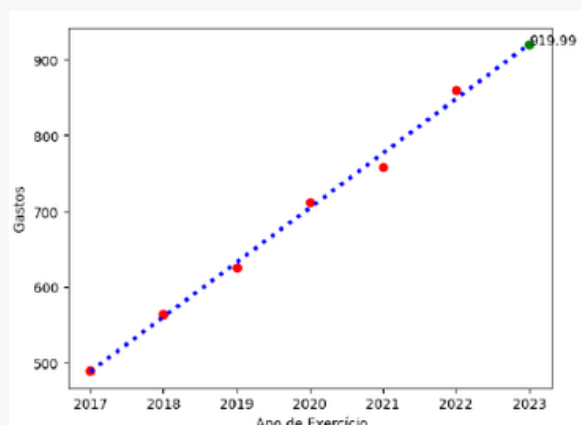
Figura 8: Gastos totais por ano em milhões.



Fonte: Autoria Própria

A nova série de dados sendo mais constante e estável quando aplicada ao modelo de regressão linear gerou um coeficiente de determinação de 99%, o aumento espantoso do coeficiente chega a ser suspeito, pois valores muito próximos ou acima de 100% de correlação costumam ser chamados de “*overfit*” e sinalizam algum erro nos cálculos ou manipulação dos dados, porém ao observar os valores encontrados em relação à previsão realizada percebe-se que de fato os dados são extremamente próximos o que justifica a taxa de correlação tão alta.

Figura 9: Tendência de gastos anuais em milhões



Fonte: Autoria Própria

Tendo conhecimento da previsão para o final do ano, basta agora saber se a prefeitura está gastando dentro do previsto, para checar essa informação foi considerado o período de janeiro à julho(mês em que se encerrou a coleta de dados) em todos os exercícios desde 2017, ao calcular o percentual da somatória do período observado em relação ao exercício a qual o período pertence, encontrou-se um valor constante de aproximadamente 30% em todos os anos inclusive no ano de 2023 que apresenta 34% dos gastos em cima da previsão apontada para o exercício de 2023 segundo a regressão linear conforme apontado na Figura 10:

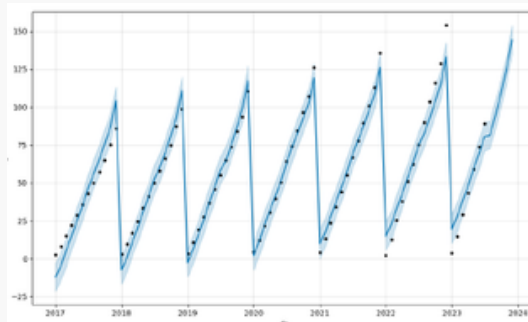
Figura 10: Taxas percentuais de gastos orçamentários até julho em relação ao exercício



Fonte: Autoria Própria

A leve elevação de 3 a 4 pontos percentuais apontada pelos gráficos demonstram que o modelo pode ter estimado valores um pouco abaixo do gasto esperado mas ainda sim permanece bem próximo do que foi encontrado, para dar mais sustentação aos cálculos decidiu-se aplicar também o algoritmo do profeta para complementar a análise.

Figura 11: Evolução dos Gastos em milhões segundo algoritmo do profeta



Fonte: Autoria Própria

O mesmo formato de dados observado que inviabilizou inicialmente o uso da Regressão Linear Simples chama atenção para o uso do algoritmo do Profeta do Facebook, embora a baixa quantidade de dados apontem para uma aplicação mais otimizada do ARIMA, acredita-se que o Profeta faça bom uso da sazonalidade detectada. O coeficiente de determinação obtido para esse modelo foi de 96%, ambos os algoritmos previram um montante de aproximadamente R\$900 milhões em gastos orçamentários até o fim do ano. É interessante ressaltar que o modo que o algoritmo do profeta funciona nos permite pular etapas, não necessitando de calcular as médias anuais e nem os percentuais mensais para entender o crescimento dos gastos que se comportou de forma muito semelhante em ambos algoritmos.

5 CONSIDERAÇÕES FINAIS

Por meio dos padrões identificados percebe-se que as contas da prefeitura se mantêm estáveis a mais de sete anos apesar da lamentável pandemia do vírus COVID-19 o que aponta uma gestão saudável e controlada. Como foi relatado no decorrer do artigo, a pesquisa foi feita com um conjunto de dados muito específico, em meio a essa decisão inúmeros insights e cenários observados foram descartados porque permeiam uma seara muito mais complexa assim como já foi dito, tais decisões infelizmente podem tornar a pesquisa consideravelmente breve e limitada porém a análise demonstra o potencial das ferramentas para as organizações, também é válido ressaltar que outras técnicas citadas no artigo podem ser aplicadas como o ARIMA e as Redes Neurais LSTM logo o presente trabalho está aberto a contribuições que o deixem mais completo e útil para a prefeitura e seus cidadãos.

REFERÊNCIAS

CONGRESSO NACIONAL. **Glossário Orçamentário**. Termo: Despesa Pública. Disponível em:
<https://www.congressonacional.leg.br/legislacao/publicacoes/glossario-orcamentario/-/orcamentario/termo/despesa_publica>. Acesso em: 13/08/2023.

COSTA, Igor Cordeiro da. **Descoberta do conhecimento e utilidade da informação contábil**. 2020.

DE OLIVEIRA, Pedro Carvalho. **Séries Temporais**: analisar o passado, prever o futuro. 2007.

FAYYAD, Usama M. et al. **Knowledge Discovery and Data Mining**: towards a unifying framework. In: KDD. p. 82-88. 1996.

FAYYAD, Usama; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. **The KDD process for extracting useful knowledge from volumes of data**. Communications of the ACM, v. 39, n. 11, p. 27-34, 1996.

FONTANA, Éliton. **Introdução aos algoritmos de aprendizagem supervisionada**. Departamento de Engenharia Química, Universidade Federal do Paraná, 2020.

IZBICKI, Rafael; DOS SANTOS, Tiago Mendonça. **Aprendizado de máquina**: uma abordagem estatística. Rafael Izbicki, 2020.

JUNDO, SERGIO. **AFO-Administração Financeira e Orçamentária**. Elsevier Brasil, 2008.

MAESTRI, Gabriela et al. Revoluções tecnológicas e a relação com o setor têxtil: perspectivas baseadas em Indústria 3.5, Indústria 4.0 e Indústria 5.0. **Revista Eletrônica Perspectivas da Ciência e Tecnologia**-ISSN: 1984-5693, v. 13, 2021.

MAHESH, Batta. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**. [Internet], v. 9, n. 1, p. 381-386, 2020.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis**. John Wiley & Sons, 2021.

MORETTIN, Pedro A.; TOLOI, Clélia MC. **Análise de séries temporais**: modelos lineares univariados. Editora Blucher, 2018.

NING, Yanrui; KAZEMI, Hossein; TAHMASEBI, Pejman. **A comparative machine learning study for time series oil production forecasting**: ARIMA, LSTM, and Prophet. Computers & Geosciences, v. 164, p. 105126, 2022.

POLIZELI, Mateus Vendramini. **Aplicação de algoritmos não supervisionados em dados eleitorais**. 2019.

SEZER, Omer Berat; GUDELEK, Mehmet Ugur; OZBAYOGLU, Ahmet Murat. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. **Applied soft computing**, v. 90, p. 106181, 2020.

STEINER, Maria Teresinha Arns et al. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. **Gestão & Produção**, v. 13, p. 325-337, 2006.