



X Encontro Brasileiro de Administração Pública.
ISSN: 2594-5688
secretaria@sbap.org.br
Sociedade Brasileira de Administração Pública

Identificação automática de conluio em licitações do Comprasnet com Aprendizado de Máquina

Rodrigo Vilela Fonseca De Souza, Alexandre Solon Nery, Fábio Lúcio Lopes De Mendonça

[ARTIGO] GT 22 Experiências e desafios da Inteligência Artificial no Setor Público

Identificação automática de conluio em licitações do Comprasnet com Aprendizado de Máquina

Resumo:

O Governo Federal Brasileiro, somente entre 2018 e 2021, executou mais de 122 mil processos da modalidade de licitação pregão, através do Portal de Compras Comprasnet, totalizando R\$ 144 bilhões. A auditoria destes certames é uma das atribuições da Controladoria Geral da União - CGU, que desenvolveu ferramentas para apoiar esta atividade, possibilitando que mais de R\$ 6 bilhões em pregões fossem revogados, suspensos ou ajustados. Buscando evoluir estas ferramentas, a Inteligência Artificial aplicada à análise de dados, a partir de algoritmos de aprendizado de máquina, se apresenta como promissora na indicação de conluio entre os participantes de uma licitação. Neste trabalho foi realizado um estudo de um conjunto de algoritmos de aprendizado de máquina, aplicados em 4 cenários com variáveis extraídas do Comprasnet para a identificação de conluio. Considerando todas as métricas adotadas no experimento, os algoritmos de melhor desempenho foram: *Extra Trees*, *Random Forest* e *Ada Boost*.

Palavras-chave: Licitação. Pregão eletrônico. Conluio. Inteligência artificial.

Introdução:

A modalidade de licitação para compras públicas denominada pregão, foi instituída no Brasil pela Lei 10.520/2002 e regulamentada, na forma de realização eletrônica, no âmbito da administração pública federal, pelo Decreto 10.024/2019 (BRASIL, 2002; BRASIL, 2019a). A execução dos pregões eletrônicos, com exceções admitidas pelo Decreto, ocorre pelo Portal de Compras do Governo Federal, também conhecido como Comprasnet (BRASIL, 2019a). O governo federal brasileiro, entre os anos de 2018 e 2021, somente com esta modalidade, empregou aproximadamente R\$ 144 bilhões, executados em 122 mil processos distintos (BRASIL, 2022).

A Controladoria-Geral da União (CGU), conforme estipulado na Lei 13.844/2019, atua na defesa do patrimônio público, no controle interno, na auditoria pública, na prevenção e no combate à corrupção (BRASIL, 2019b). Dessa forma, auditar as compras públicas está entre as competências da CGU. Isto inclui, o acompanhamento de procedimentos e processos administrativos em curso em órgãos ou entidades da administração pública federal. Bem como, a realização de inspeções e avocação de procedimentos e processos em curso na administração pública federal, para exame de sua regularidade, e proposição de providências ou correção de falhas (BRASIL, 2019b).

Neste sentido, a CGU, na busca de respostas tempestivas do exercício de suas atribuições frente aos processos licitatórios publicados, desenvolveu uma ferramenta de análise de editais e licitações (BRASIL, 2020). O Analisador de Licitações e Editais, batizado de ALICE, é um algoritmo autônomo que acessa diariamente as publicações no Comprasnet para (DANTAS; MARTINS, 2022): 1 - capturar as novas licitações (editais e anexos); 2 - determinar o valor estimado e classificar as licitações por tema; 3 - aplicar

trilhas de auditoria para detecção de possíveis irregularidades; 4 - enviar alertas aos auditores para avaliação das licitações selecionadas.

Com os alertas gerados pela ferramenta ALICE, os auditores da CGU executam auditorias preventivas direcionadas em um universo médio diário de 250 novas licitações. Somente entre 2019 e 2020, os alertas gerados resultaram em um total de R\$ 6,79 bilhões em licitações que foram auditadas preventivamente e demandaram ações da Controladoria (ROCHA, 2019; BRASIL, 2020).

Apesar da reconhecida eficiência da ALICE, inclusive pelo uso de modelos de Inteligência Artificial (IA), ainda não existem trilhas que analisem individualmente item a item do pregão para identificar possível conluio entre os participantes (DANTAS; MARTINS, 2022; PANIS, 2020). Esse tipo de análise é de difícil execução, pois os auditores devem avaliar, em tempo hábil, muitas variáveis e um grande volume de informações neste universo de novas licitações publicadas diariamente (ROCHA, 2019).

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) publicou, em um de seus relatórios, baseado em uma pesquisa não científica, uma estimativa de sobrepreço entre 10 e 20% gerado pela atuação de cartéis, se comparado ao preço em mercados competitivos (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT., 2002). Em alguns casos, o sobrepreço pode chegar a 50%, causando perdas anuais de centenas de bilhões de reais aos consumidores. Dois casos que podemos citar no Brasil são a Operação Lava Jato, que identificou prejuízo de bilhões de reais (AGUILERA, 2019), e o Cartel de Portas de Segurança Giratórias, onde a colusão gerou sobrepreço de 25% (CADE, 2019).

É neste ponto que propomos a criação de uma trilha de auditoria que apoie os auditores nestas análises individualizadas, indicando um possível conluio para formação de cartel entre os participantes do pregão auditado pela ALICE. Em contextos correlatos, Aprendizado de Máquina - AM, um dos subcampos da IA, aplicada à análise de dados, vem sendo aplicadas em outras instituições, obtendo resultados animadores e aumentando a eficiência nos gastos públicos. O Tribunal de Contas da União, desde 2017, utiliza também uma versão da ferramenta ALICE e desenvolveu a Mônica, um painel onde todas as compras públicas podem ser visualizadas, incluindo as que são ignoradas pela ALICE, como contratações diretas e inexigibilidades de licitação (DESORDI; BONA, 2020).

Especificamente sobre a aplicação de IA na identificação de problemas em

licitações, Maia *et al.* (2019) definiram um conjunto de variáveis que representam anomalias na contratação. Entre as anomalias identificadas, o sobrepreço foi definido como o somatório da diferença entre o valor oferecido pela primeira colocada e o valor real de celebração do contrato. Esta simplificação possibilita a análise, mas caso os participantes da disputa estejam em uma ação combinada para fraudar a licitação, onde o primeiro colocado seja efetivamente a empresa que celebrará o contrato, o método não identificará o sobrepreço resultante do cartel estabelecido (MAIA *et al.*, 2019).

Faria *et al.* (2010), baseados em um modelo de variáveis quantitativas, demonstraram que alguns fatores são preponderantes na variação dos preços dos produtos contratados por pregão eletrônico. As variáveis número de fornecedores de um item, especificidade do item, quantidade e frequência deste item nas transações, apresentam, com o método dos mínimos quadrados ordinários, um poder de explicação conjunta de 67,4% da variação do preço na licitação (FARIA *et al.*, 2010).

García Rodríguez *et al.* (2022) testaram e compararam a precisão de algoritmos de AM, listados no Quadro 1, com capacidade de detecção de conluio entre os participantes de uma licitação, classificando os lances como colusivos ou competitivos. Os testes ocorreram sobre conjuntos de dados colusivos, resultantes de investigações criminais por formação de cartéis na disputa de licitações em cinco países. Além disso, o referido estudo, verificou a efetividade da aplicação de índices estatísticos calculados a partir dos próprios *datasets* (García Rodríguez *et al.*, 2022).

Em primeira análise, seria possível a indicação de conluio para formação de cartel entre as empresas participantes de pregões eletrônicos realizados no Comprasnet. Bem como, de forma experimental, verificar a relevância das variáveis relacionadas à variação de preços indicadas por Faria *et al.* (2010) na caracterização de conluio a partir de dados extraídos do Comprasnet. Assim, de forma geral, o objetivo que buscamos atingir é identificar conluio em pregões eletrônicos do governo federal no Comprasnet.

Este artigo, inicia com uma revisão da literatura sobre a legislação atual que caracteriza conluio, bem como os trabalhos que abordaram a aplicação de AM nesse contexto, tanto de forma geral, quanto aplicado aos pregões realizados no Comprasnet e as lacunas identificadas (1). Posteriormente, apresentamos como o trabalho foi desenvolvido, com os procedimentos necessários para a geração do *dataset* e a aplicação dos algoritmos de AM selecionados (2). Na sequência, os resultados do experimento são

apresentados e analisados por meio da comparação com outros experimentos realizados (3). Concluindo, apresentamos os avanços obtidos e possíveis futuras evoluções.

1. Revisão da literatura

O Conselho Administrativo de Defesa Econômica (CADE), define a formação de cartéis em licitações como o conluio entre agentes econômicos para restringir ou eliminar a concorrência do processo de compra de bens ou contratação de serviços pela Administração Pública. Esta prática traz prejuízos ao Estado, pois limita a efetiva concorrência natural, acarretando preços mais elevados, produtos e serviços de menor qualidade ou aquisição de quantidade menor do que a necessária (CADE, 2019).

No ordenamento jurídico brasileiro, previsto na Lei nº 12.529/2011, em seu artigo 36, § 3º, inciso I, alínea “d”, a prática de cartel em licitação configura ilícito antitruste (BRASIL, 2011). Esta conduta anticompetitiva, que fere a livre concorrência, é investigada e julgada pelo CADE, que pode aplicar sanções às empresas e pessoas físicas, como multas ou impedi-las de participar de licitações, além de outras penalidades prevista na Lei (CADE, 2019).

Nai *et al.* (2022), em uma revisão sistemática de 15 trabalhos relacionados à identificação de fraude em licitações públicas, publicados entre 2016 e 2021, identificaram que a maioria dos artigos adotou métodos típicos de AM, enquanto dois grupos menores aplicaram principalmente redes neurais e análise de redes.

A identificação de conluio em pregões eletrônicos do Comprasnet foi objeto de estudo de Maia *et al.* (2019), propondo uma metodologia para detecção de anomalias, buscando possíveis conluios entre empresas licitantes. O modelo é baseado na utilização de mineração de padrões frequentes, correlação de series temporais multivariadas e análise conjugada de multicritérios, chamadas microexpressões (MAIA *et al.*, 2019). No entanto, nem todas as microexpressões propostas são baseadas em dados disponíveis para cálculo de forma abrangente, como o motivo de desclassificação de uma empresa que venceu uma licitação, limitando sua aplicação.

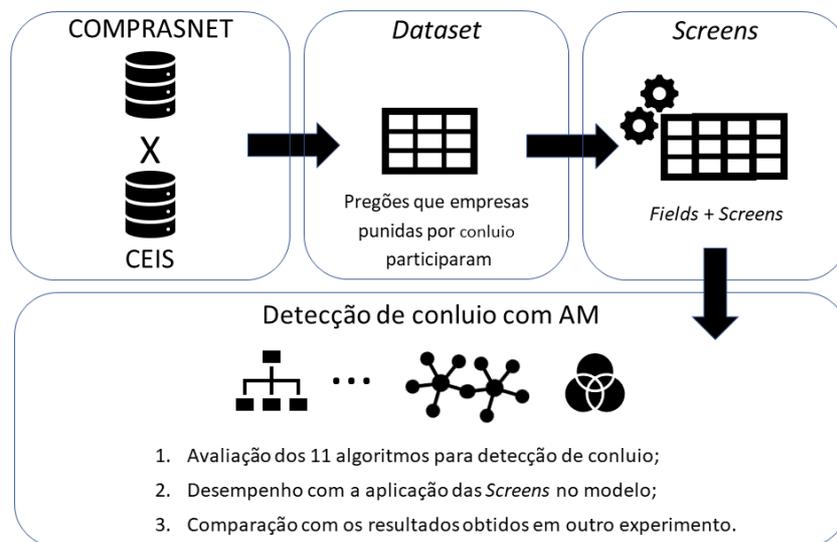
No experimento de García Rodríguez *et al.* (2022), 11 algoritmos, listados no Quadro 1, foram massivamente testados em *datasets* que caracterizam a prática de conluio entre os participantes de licitações públicas, oriundos de investigações criminais no Brasil, Itália, Japão, Suíça e Estados Unidos. Os cenários de teste envolveram a

variação dos dados disponíveis nos *datasets*, bem como, a inclusão de índices estatísticos calculados sobre os próprios *datasets*. De acordo com os autores, os 3 melhores algoritmos foram: *Extra Trees*, *Random Forest* e *Ada Boost*. Considerando o cenário onde todos os dados dos *datasets* estavam disponíveis, as acurácias observadas para estes algoritmos variaram entre 81% e 95%, com acurácias balanceadas geralmente acima de 73% (exceto para o *dataset* dos Estados Unidos) (García Rodríguez *et al.*, 2022).

2. Metodologia proposta

Buscando replicar o experimento de García Rodríguez *et al.* (2022) para o contexto do Comprasnet, visando a identificação de conluio em pregões eletrônicos, bem como, estendê-lo com a inclusão no *dataset* das variáveis identificadas pela variação de preços nestes pregões, definidas por Faria *et al.* (2010), os procedimentos executados neste trabalho estão resumidos na Figura 1. As etapas do experimento serão descritas em maiores detalhes nesta seção.

Figura 1 – Metodologia para extração do *dataset* e análise dos algoritmos.



Fonte: Elaborado pelos autores

O Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS) consiste em uma relação de empresas e pessoas físicas que sofreram sanções que implicaram a restrição de participar de licitações ou de celebrar contratos com a Administração Pública. Este cadastro, mantido pela CGU e de acesso público pelo Portal da Transparência, possui entre outros tipos de sanções, a relação de entes punidos pela prática de conluio, enquadrados na Lei nº 12.529/2011. Dessa forma, aplicando os filtros categoria da sanção

do tipo “Impedimento” e órgão sancionador como “CADE” no CEIS, identificamos o CNPJ das 24 empresas com impedimento de contratação com base na Lei Antitruste (BRASIL, 2022).

Tais empresas estão proibidas de contratar com instituições financeiras oficiais e participar de licitação tendo por objeto aquisições, alienações, realização de obras e serviços, concessão de serviços públicos, na administração pública federal, estadual, municipal e do Distrito Federal, bem como em entidades da administração indireta, por prazo não inferior a 5 (cinco) anos, conforme o inciso II, artigo 38, da Lei 12.529/2011, pelos seguintes períodos identificados nos dados extraídos do CEIS: 23/12/2020 a 23/12/2025, 18/08/2021 a 18/08/2026 e 01/09/2021 a 01/09/2026 (BRASIL, 2022).

A partir do CNPJ das empresas com a referida restrição no CEIS, realizamos um levantamento na base do Comprasnet, dos pregões eletrônicos que tais empresas participaram previamente à punição recebida. Estes pregões, bem como seus itens e lances ofertados pelas empresas participantes, foram extraídos direto da base de dados do Comprasnet, via Acordo de Cooperação entre CGU e Ministério da Economia, para a construção de uma base de licitações disputadas por empresas punidas pela prática de conluio (BRASIL; BRASIL, 2021). Os lances ofertados pelas empresas punidas por conluio foram previamente classificados como colusivos. Já os lances ofertados pelas demais empresas participantes dos certames, foram classificados como competitivos. Dessa forma, geramos o *dataset* rotulado e formatado como os *datasets* do trabalho de García Rodríguez *et al.* (2022).

O critério de marcação de conluio dos lances apresentados pelas empresas punidas com base na Lei Antitruste tem fins práticos na formação do *dataset*, buscando identificar este comportamento anticompetitivo. Estratégias semelhantes foram adotadas por García Rodríguez *et al.* (2022) na formação dos 6 *datasets* do estudo, ao marcar um item licitado como alvo de conluio quando ao menos um licitante identificado como colusivo ofertou lances naquele item.

Os pregões identificados no Comprasnet que obtiveram a participação de pelo menos uma das empresas punidas, compreendem certames realizados entre 2006 e 2015, onde foram licitados 445 itens, que geraram 11.049 lances individuais. Apenas 6,26% desses lances, o que corresponde a 692, foram assumidos como colusivos, seguindo a estratégia de marcação descrita acima. Os lances foram realizados por 147 empresas diferentes, entre elas, apenas 7 estavam entre as 24 identificadas no CEIS.

Os resultados obtidos neste experimento serão comparados com os resultados obtidos no experimento original, restrito ao *dataset* oriundo do Japão. Este *dataset* é o que mais se parece com o obtido neste estudo, tanto em relação à quantidade de itens licitados e número de lances, quanto em relação ao percentual de itens colusivos e competitivos.

2.1 Screens

A identificação de conluio a partir de algoritmos de AM é potencializada pela inclusão de *Screening Variables*, ou simplesmente *screens*. As *screens* são índices estatísticos calculados a partir dos lances ofertados pelos licitantes do certame (HUBER; IMHOF, 2019). No experimento realizado por García Rodríguez *et al.* (2022), as *screens* também melhoraram as métricas utilizadas para avaliar o desempenho dos algoritmos.

Assim, a partir do *dataset* gerado neste experimento, o próximo passo foi o cálculo das seguintes *screens* para cada item licitado, com base no mesmo mecanismo utilizado do experimento precursor, cujo código fonte está disponível nos apêndices:

- *Coefficient of Variation (CV)*: representa a razão entre o desvio padrão dos lances do item e a média dos lances deste item;
- *Spread (SPD)*: dispersão dos lances de um item;
- *Difference between the two lowest bids (DIFFP)*: mede a diferença relativa entre os dois lances mais baixos do item;
- *Relative Distance (RD)*: índice alternativo ao *DIFFP*, substituindo o termo no denominador pelo desvio padrão dos lances perdedores do item;
- *Skewness (SKEW)*: possibilita verificar a distribuição assimétrica dos lances;
- *Excess Kurtosis (KURT)*: a condensação de valores dos lances muito próximos (ou muito distantes) da média de lances do item - requer pelo menos 4 lances;
- *Kolmogorov-Smirnov test (KSTEST)*: representa a semelhança dos valores dos lances para uma distribuição uniforme.

Buscando testar outras *screens* na identificação de conluio, adicionamos novas variáveis ao *dataset*. Faria *et al.* (2010) demonstraram que as variáveis a seguir

apresentaram um poder de explicação conjunta de 67,4% das variações dos preços em pregões eletrônicos:

- Total de lances (LANC): quantidade de lances apresentados na disputa de um item do pregão, com interferência positiva, pois quanto mais lances um item recebe, maior a probabilidade do preço cair;
- Quantidade (QUANT): quantidade total licitada do item no pregão, com interferência positiva, pois quanto maior a quantidade a ser comprada, maior a probabilidade do custo unitário ser menor;
- Escore de especificidade (ESPECIF): o escore de especificidade tem uma interferência negativa, pois quanto maior a especificidade do item comprado, menor será o número de empresas concorrentes e, por isso, menor será a concorrência na licitação;
- Escore de frequência (FREQ): o escore de frequência da empresa vencedora de um item do pregão possui uma interferência positiva, pois fornecedores que prezam relacionamentos mais duradouros fazem maiores esforços para continuar o fornecimento e, para isso, são obrigados a reduzir seus preços.

Tais variáveis foram calculadas a partir da base de dados do Comprasnet e adicionadas ao *dataset*, buscando dotar o modelo de características que indiquem a variação de preço de um item licitado. Ou seja, quanto menos variou o preço de um item, maior a probabilidade da existência de conluio na licitação, já que o objetivo dessa prática é garantir que o preço varie pouco, sem competição entre os participantes (CADE, 2019).

2.2 Algoritmos e cenários de teste

Os algoritmos de AM testados no experimento, listados na Quadro 1, assim como realizado por García Rodríguez *et al.* (2022), buscam resolver a classificação binária dos lances das licitações do Comprasnet entre colusivos e competitivos (ou não colusivos). Os algoritmos baseados em *Ensemble methods* (*Extra Trees*, *Random Forest*, *Ada Boost* e *Gradient Boosting*) obtiveram os melhores resultados no referido experimento.

Os algoritmos foram testados em 4 cenários diferentes com o *dataset* gerado a partir do Comprasnet, variando os campos selecionados em cada cenário da seguinte forma:

- Cenário 1 (todos os campos). Este cenário inclui todos os campos disponíveis no *dataset*, incluindo as *screens* originárias do estudo de Faria *et al.* (2010). Uma exceção é o campo que identifica os licitantes, buscando evitar um possível direcionamento dos algoritmos com esses dados.
- Cenário 2 (todos os campos + *screens*). Os campos do cenário 1 mais as *screens* (CV, SPD, DIFFP, RD, SKEW, KURT e KSTEST).
- Cenário 3 (campos comuns). Somente os campos comuns a todos os *datasets* do estudo original: código da licitação, os valores dos lances, o lance vencedor e o número de lances por item licitado.
- Cenário 4 (campos comuns + *screens*). Os campos do cenário 3 mais as 7 *screens* iniciais adicionadas ao *dataset*.

Quadro 1 – Algoritmos testados no experimento.

| MÉTODO | ALGORITMO |
|--------------------------------|---|
| <i>Linear Model</i> | <i>SGD (Stochastic Gradient Descent)</i> (ZHANG, 2004) |
| <i>Ensemble methods</i> | <i>Extra Trees</i> (GEURTS; ERNST; WEHENKEL, 2006) |
| | <i>Random Forest</i> (BREIMAN, 2001) |
| | <i>Ada Boost</i> (FREUND; SCHAPIRE, 1997) |
| | <i>Gradient Boosting</i> (FRIEDMAN, 2001) |
| <i>Support Vector Machines</i> | <i>SVC</i> (CORTES; VAPNIK, 1995) |
| <i>Nearest Neighbors</i> | <i>K Neighbors</i> (ALTMAN, 1992) |
| <i>Neural network models</i> | <i>MLP (Multi-Layer Perceptron)</i> (HASTIE <i>et al.</i> , 2009) |
| <i>Naive Bayes</i> | <i>Bernoulli Naive Bayes</i> (HASTIE <i>et al.</i> , 2009) |
| | <i>Gaussian Naive Bayes</i> (HASTIE <i>et al.</i> , 2009) |
| <i>Gaussian Process</i> | <i>Gaussian Process</i> (RASMUSSEN; WILLIAMS, 2005) |

Fonte: Adaptado de García Rodríguez *et al.* (2022)

Os cenários 1 e 3 são diferenciados apenas pelas *screens* propostas por Faria *et al.* (2010): total de lances, quantidade, escore de especificidade e escore de frequência. Dessa forma, a comparação do resultado dos algoritmos nestes dois cenários evidenciará a relevância delas na identificação de lances colusivos. No entanto, o total de lances

recebidos por um item de pregão já faz parte do *dataset* no cenário 1. Desta forma, apenas as outras três diferenciam os cenários efetivamente.

2.3 Métricas

A comparação do desempenho dos algoritmos do experimento levou em consideração 5 métricas de erro, definidas para este contexto de classificação:

- *Accuracy*: percentual de classificações corretas, tanto de itens colusivos quanto de competitivos, em relação ao total de lances;
- *Precision*: habilidade do classificador de não determinar como colusivo um lance competitivo;
- *Recall*: habilidade do classificador de identificar todos os lances colusivos;
- *Balanced Accuracy*: bom indicador para *datasets* desbalanceados, como no *dataset* gerado, indicando a média de identificação positiva de lances de cada classe (colusivos/competitivos);
- *F1 score*: média harmônica entre *Precision* e *Recall*, onde o score mais próximo de 1 representa o melhor valor e mais próximo de 0, o pior.

O objetivo final da classificação é realizar a distinção entre lances colusivos e competitivos entre os n lances do *dataset*. Assim, classificar um lance corretamente como colusivo consiste em realizar uma identificação *True Positive (TP)*. Classificá-lo como competitivo corretamente significa realizar uma identificação *True Negative (TN)*. Por outro lado, quando um lance competitivo é classificado como colusivo, temos uma identificação *False Positive (FP)*. E por fim, quando um lance é colusivo, mas classificado como competitivo, temos uma identificação *False Negative (FN)*.

Portanto, podemos calcular as métricas relacionadas da seguinte forma:

$$Accuracy = \frac{TP + TN}{n}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\text{Balanced Accuracy} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.4 Treinamento e execução

Os algoritmos de AM foram treinados e testados para detectar conluio no *dataset* gerado do Comprasnet nos 4 cenários apresentados. Cada algoritmo foi executado 50 vezes, aleatoriamente alterando o subconjunto de treinamento, composto por 80% das licitações, e o subconjunto de teste, pelos 20% restantes. Os parâmetros utilizados em cada algoritmo e todo o processamento do *dataset* estão disponíveis no apêndice.

A cada iteração dos algoritmos, as métricas de erros foram calculadas e armazenadas. Dessa forma, as métricas apresentadas abaixo correspondem às médias obtidas ao final dessas iterações.

3. Resultados e discussão

Entre as métricas obtidas no experimento, *Accuracy*, *FP*, *FN* e *Balanced Accuracy*, assim como no estudo de García Rodríguez *et al.* (2022), nos permitem analisar o desempenho dos algoritmos nos 4 cenários apresentados para o *dataset* extraído do Comprasnet. Além da comparação entre os cenários em si, a inclusão das métricas do *dataset* do Japão, obtidas no experimento precursor, possibilita a análise dos resultados entre os dois *datasets*. A Figura 2 resume os dados obtidos nos cenários 1 e 2. Já a Figura 3 contempla os cenários 3 e 4.

Nestas figuras, as setas verdes indicam um melhor desempenho do algoritmo considerando a métrica no cenário, comparando-se o *dataset* do Comprasnet com o do Japão. Já as setas vermelhas, indicam uma piora nos índices em relação ao Comprasnet. O círculo verde indica métricas iguais em ambos os *datasets*.

De modo geral, considerando todos os cenários, os melhores algoritmos foram: *Extra Trees*, *Random Forest*, *Ada Boost* e *Gradient Boosting*. Mas em cenários isolados, outros algoritmos obtiveram melhores resultados, como o caso do *Gaussian Process* no cenário 4. Em menos da metade dos casos as *screens* propiciaram um melhor desempenho dos algoritmos, tanto na acurácia quanto na taxa de *FP* e *FN*.

Em quase todos os cenários, podemos observar uma elevação dos índices de acurácia e melhora da taxa de *FP* na comparação do *dataset* do Comprasnet com o do Japão. O mesmo não ocorre quando analisamos a taxa de *FN* e a acurácia balanceada.

Mesmo assim, para alguns cenários e algoritmos, identificamos a melhora ou estabilização de todas as métricas, como no caso do algoritmo MLP no cenário 4.

Figura 2 – Métricas do experimento para os cenários 1 e 2.

Cenário 1 (todos os campos) e 2 (todos os campos + screens) - Algoritmos

| Métricas | Dataset\Cenário | SGD | | Extra Trees | | Random Forest | | Ada Boost | | Gradient Boosting | | SVC | |
|--------------------------|-----------------|------|------|-------------|------|---------------|------|-----------|------|-------------------|------|------|------|
| | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Accuracy(%) | Comprasnet | 84,4 | 88 | 94,6 | 94,8 | 94,2 | 94,2 | 93,6 | 92 | 94,2 | 92,8 | 76,8 | 77,7 |
| | Japão | 87,8 | 87,8 | 94,7 | 94,5 | 93,1 | 93 | 93,5 | 93,1 | 90,5 | 89,2 | 87,8 | 87,9 |
| False Positives (FP) (%) | Comprasnet | 9,5 | 6,2 | 0,3 | 0,3 | 1 | 0,8 | 1,2 | 3,1 | 0,3 | 0,6 | 19,4 | 18,8 |
| | Japão | 5,6 | 5,6 | 0,9 | 0,8 | 2,6 | 2,6 | 2,4 | 2,5 | 4,8 | 5,8 | 9,4 | 9,3 |
| False Negatives (FN) (%) | Comprasnet | 6,1 | 5,8 | 5 | 5 | 4,8 | 5 | 5,3 | 5 | 5,5 | 6,6 | 3,8 | 3,5 |
| | Japão | 6,6 | 6,6 | 4,4 | 4,7 | 4,3 | 4,5 | 4,1 | 4,4 | 4,7 | 5 | 2,8 | 2,7 |
| Balanced accuracy (%) | Comprasnet | 49,7 | 49,5 | 61,1 | 61 | 61,1 | 60 | 58,1 | 57,7 | 62,3 | 59,8 | 49,2 | 67 |
| | Japão | 67,6 | 67,9 | 79,8 | 78,7 | 79,3 | 78,6 | 80,4 | 79,2 | 76,3 | 75,3 | 82,9 | 83,1 |

Cenário 1 (todos os campos) e 2 (todos os campos + screens) - Algoritmos

| Métricas | Dataset\Cenário | K Neighbors | | MLP | | Bernoulli Naive Bayes | | Gaussian Naive Bayes | | Gaussian Process | |
|--------------------------|-----------------|-------------|------|------|------|-----------------------|------|----------------------|------|------------------|------|
| | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Accuracy(%) | Comprasnet | 94,5 | 94,5 | 93,9 | 93,6 | 93,6 | 91,8 | 84,3 | 69 | 92,4 | 93,1 |
| | Japão | 92,5 | 92,5 | 88,7 | 88,8 | 88,7 | 88,6 | 94,6 | 94,6 | 89,5 | 88,9 |
| False Positives (FP) (%) | Comprasnet | 0,4 | 0,3 | 0 | 0 | 0 | 2 | 10,3 | 26,6 | 0,1 | 0,2 |
| | Japão | 3,7 | 3,6 | 0,1 | 0 | 0 | 0 | 0,7 | 0,7 | 0 | 0 |
| False Negatives (FN) (%) | Comprasnet | 5,1 | 5,2 | 6,1 | 6,4 | 6,4 | 6,2 | 5,5 | 4,4 | 7,4 | 6,7 |
| | Japão | 3,8 | 3,9 | 11,3 | 11,2 | 11,3 | 11,4 | 4,7 | 4,7 | 10,5 | 11,1 |
| Balanced accuracy (%) | Comprasnet | 59,7 | 58,6 | 50 | 50 | 50 | 50,4 | 49,5 | 50,2 | 51,2 | 51,2 |
| | Japão | 80,7 | 80,7 | 50 | 50,1 | 50 | 50,1 | 78,3 | 78,7 | 50 | 50 |

Fonte: Elaborado pelos autores

Figura 3 – Métricas do experimento para os cenários 3 e 4.

Cenário 3 (campos comuns) e 4 (campos comuns + screens) - Algoritmos

| Métricas | Dataset\Cenário | SGD | | Extra Trees | | Random Forest | | Ada Boost | | Gradient Boosting | | SVC | |
|--------------------------|-----------------|------|------|-------------|------|---------------|------|-----------|------|-------------------|------|------|------|
| | | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 |
| Accuracy(%) | Comprasnet | 93,4 | 88,1 | 94,8 | 94,6 | 94,2 | 93,8 | 94,4 | 93,5 | 93 | 93,6 | 90,7 | 79,2 |
| | Japão | 83,9 | 83,7 | 94,5 | 94,5 | 93,2 | 93,4 | 93,3 | 92,3 | 90,7 | 87,9 | 85,5 | 82,5 |
| False Positives (FP) (%) | Comprasnet | 0,2 | 5,6 | 0,3 | 0,2 | 0,8 | 0,8 | 0,6 | 1,2 | 0,4 | 1,3 | 5,4 | 17,3 |
| | Japão | 8,2 | 8,2 | 1,1 | 0,6 | 2,5 | 2,2 | 2,2 | 3 | 5,3 | 7,7 | 11,5 | 14,5 |
| False Negatives (FN) (%) | Comprasnet | 6,4 | 6,2 | 4,9 | 5,2 | 5 | 5,4 | 4,9 | 5,3 | 6,6 | 5,1 | 3,9 | 3,5 |
| | Japão | 7,9 | 8,1 | 4,5 | 4,9 | 4,3 | 4,3 | 4,5 | 4,7 | 4 | 4,4 | 3 | 3 |
| Balanced accuracy (%) | Comprasnet | 49,9 | 50 | 59,1 | 58,4 | 59,5 | 59 | 58,9 | 58,5 | 58,3 | 61,3 | 59,6 | 65,2 |
| | Japão | 60,5 | 59,6 | 79,6 | 78,2 | 79,5 | 79,4 | 78,9 | 77,3 | 77,4 | 75,8 | 80,2 | 78,4 |

Cenário 3 (campos comuns) e 4 (campos comuns + screens) - Algoritmos

| Métricas | Dataset\Cenário | K Neighbors | | MLP | | Bernoulli Naive Bayes | | Gaussian Naive Bayes | | Gaussian Process | |
|--------------------------|-----------------|-------------|------|------|------|-----------------------|------|----------------------|------|------------------|------|
| | | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 |
| Accuracy(%) | Comprasnet | 94,4 | 94,7 | 93,7 | 93,7 | 93,9 | 92,9 | 93,4 | 66,6 | 92,3 | 95,5 |
| | Japão | 92,3 | 92,4 | 88,7 | 88,7 | 88,8 | 88,8 | 94 | 94,3 | 88,7 | 88,9 |
| False Positives (FP) (%) | Comprasnet | 0,4 | 0,4 | 0 | 0 | 0 | 1 | 0,8 | 29,3 | 0,3 | 0,2 |
| | Japão | 3,7 | 3,6 | 0 | 0 | 0 | 0 | 0,3 | 0,4 | 0 | 0 |
| False Negatives (FN) (%) | Comprasnet | 5,2 | 4,9 | 6,3 | 6,3 | 6,1 | 6,1 | 5,8 | 4,1 | 7,5 | 4,2 |
| | Japão | 4,1 | 3,9 | 11,2 | 11,3 | 11,2 | 11,2 | 5,6 | 5,4 | 11,3 | 11,1 |
| Balanced accuracy (%) | Comprasnet | 59 | 59,3 | 50 | 50 | 50 | 50,8 | 49,6 | 51,5 | 53,5 | 53,7 |
| | Japão | 80,2 | 80,5 | 50,1 | 50 | 50 | 50 | 74,4 | 75,7 | 50,2 | 50 |

Fonte: Elaborado pelos autores

Especificamente em relação às *screens* propostas por Faria *et al.* (2010), ao compararmos os cenários 1 e 3, em menos da metade dos algoritmos as métricas foram melhoradas por elas. De forma objetiva, a contribuição dessas variáveis indica um menor impacto na detecção de conluio. No entanto, a ausência de dados refinados na base do Comprasnet para o cálculo preciso destes índices pode ter interferido na sua verificação. O cálculo da especificidade e da frequência dos itens licitados não foi suficientemente preciso nos dados acessados, pois os materiais licitados estavam indicados de maneira

abrangente, sem indicar o tipo de material na sua especificação mais detalhada.

Analisando a acurácia balanceada, observamos que os melhores resultados para o *dataset* do Japão estão em torno de 80%. Já para o *dataset* do Comprasnet, os melhores resultados são consideravelmente inferiores, em torno de 60%. Esta diferença acentuada pode ter origem no forte desbalanceamento do *dataset* do Comprasnet, com apenas 6,26% de lances colusivos, contra 8,09% do *dataset* do Japão. García Rodríguez *et al.* (2022) também obteve valores semelhantes para esta métrica no *dataset* dos Estados Unidos, onde apenas 9,17% das empresas eram colusivas.

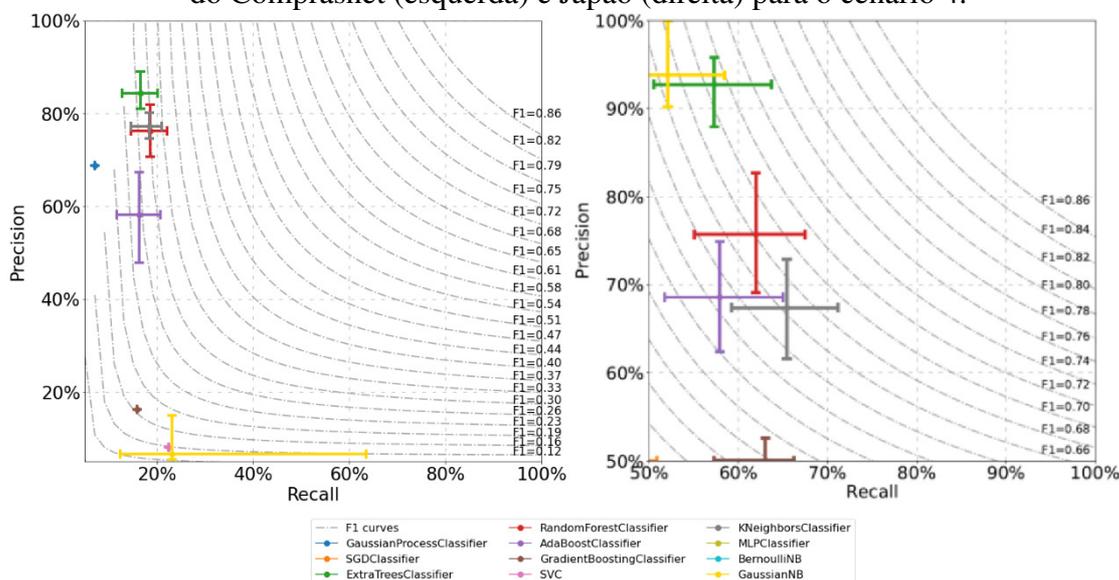
Diferentemente da estratégia de classificação adotada nos *datasets* do estudo de García Rodríguez *et al.* (2022), onde todos os lances de um item licitado foram marcados como colusivos se um determinado percentual de lances fosse efetivamente identificado como tal nas investigações realizadas, somente marcamos como colusivos os lances das empresas punidas por conluio. Esta estratégia de marcação restringiu o número de lances assumidos como colusivos, o que pode justificar a acentuada diferença da acurácia balanceada obtida neste experimento.

Com relação às métricas *Precision*, *Recall* e *F1 score*, a Figura 4 possibilita a comparação do desempenho dos melhores algoritmos do cenário 4, onde os campos comuns aos *datasets* e *screens* estão presentes, para os *datasets* do Comprasnet e Japão. As métricas foram representadas por uma cruz para cada algoritmo, onde o ponto de encontro representa a mediana da *Precision* e *Recall*. As extremidades das cruzes representam os mínimos e máximos obtidos para estas métricas em cada uma das 50 iterações dos algoritmos no experimento. Por fim, os valores das 3 métricas relacionadas no gráfico, pertencem ao interior do retângulo formado pela cruz de cada algoritmo. Os algoritmos que não geraram uma cruz no gráfico, foram suprimidos por gerarem valores baixos para as métricas nos *datasets* do Comprasnet (<5%) e Japão (< 50%).

Identificamos que os algoritmos *Extra Trees*, *Random Forest* e *Ada Boost*, baseados em *ensemble methods*, considerando a métrica *Precision*, obtiveram valores entre 50% e 90% aproximadamente, mais próximos aos obtidos por García Rodríguez *et al.* (2022). Isso não foi observado para as métricas *Recall* e, conseqüentemente, *F1 score*. Estes indicadores com resultados piores podem estar associados à estratégia de marcação de conluio utilizada no *dataset* gerado. Outra possível justificativa para os índices piores neste caso, assim como no *dataset* dos Estados Unidos, é o baixo percentual de empresas

colusivas obtido no experimento: 4,76% (7 colusivas em 147 empresas).

Figura 4 – Métricas *Precision*, *Recall* e *F1 score* para o *dataset* do Comprasnet (esquerda) e Japão (direita) para o cenário 4.



Fonte: Elaborado pelos autores com adaptações de García Rodríguez *et al.* (2022)

Considerações finais

Detectar a atuação de cartéis em licitações é importante pelo seu potencial de prejudicar a competitividade e a transparência do processo de licitação. Isso pode resultar em prejuízos para o setor público e para a sociedade em geral, já que os cartéis podem levar a maiores custos e menor qualidade nos produtos e serviços adquiridos.

É importante registrar que o uso do aprendizado de máquina para detectar cartéis em licitações é apenas uma ferramenta que pode ser usada como parte de um processo mais amplo de investigação e análise. A escassez de dados confiáveis e completos, a dificuldade em identificar padrões que indiquem a existência de um cartel e a complexidade do processo de licitação, tornam a detecção de cartel em licitações um grande desafio. Dessa forma, a análise final de um auditor para confirmar ou descartar qualquer apontamento de conluio realizado pelos modelos de AM é fundamental.

No entanto, os algoritmos de AM baseados em *ensemble methods* obtiveram um desempenho expressivo neste experimento, indicando como promissora a identificação de conluio desejada. A acurácia alcançada nos 4 cenários foi maior que 90%, além da baixa taxa de *FP*, próxima a 1%, e *FN*, próxima a 5%. Estes resultados são compatíveis com os obtidos por García Rodríguez *et al.* (2022) para este grupo de algoritmos.

Também alinhado aos resultados obtidos por García Rodríguez *et al.* (2022), considerando as métricas *Precision*, *Recall* e *F1 score*, onde o algoritmo *Gradient Boosting*, obteve baixo desempenho, destacamos os três algoritmos com melhor desempenho, considerando todas as métricas adotadas: *Extra Trees*, *Random Forest* e *Ada Boost*. Dessa forma, concluímos que estes são os algoritmos mais promissores para a implantação da trilha de auditoria que identifica possível conluio em pregões do Comprasnet analisados pela ALICE, fortalecendo o trabalho de auditoria da CGU.

As *screens* não influenciaram significativamente as métricas obtidas entre os cenários. Adicionalmente, as variáveis Total de lances, Quantidade, Escore de especificidade e Escore de frequência, propostas no trabalho de Faria *et al.* (2010), tratadas neste experimento como possíveis novas *screens* para aumentar a capacidade de detecção de conluio pelos algoritmos testados, também não produziram efeitos significativos. Mesmo assim, consideramos a inclusão destas variáveis no experimento como uma contribuição obtida. Em trabalhos futuros, estas variáveis podem ter o seu cálculo aprimorado, aumentando a precisão da identificação de materiais equivalentes em outras licitações do Comprasnet.

O *dataset* com lances colusivos extraídos do Comprasnet, apresentados por empresas punidas com base na Lei 12.529/2011, também pode ser considerado uma contribuição obtida com o experimento, dada a escassez de *datasets* para este tipo de estudo (García Rodríguez *et al.*, 2022). Mesmo com resultados equivalentes aos obtidos em *datasets* do experimento de García Rodríguez *et al.* (2022), a estratégia de marcação dos lances colusivos é uma possível fragilidade e demanda a confirmação da sua capacidade de treinar um modelo para a detecção efetiva de indícios de conluio entre os participantes de outros pregões realizados no Comprasnet. Assim, o desenvolvimento e aplicação de testes na base completa do Comprasnet para a identificação de novos casos de conluio a partir do algoritmo treinado fica como trabalho futuro.

Outra possível linha de investigação para trabalhos futuros é a mudança na estratégia de marcação de conluio na geração do *dataset*, assumindo todos os lances de um item como colusivos quando um licitante punido estiver na disputa. Esta estratégia busca diminuir o desbalanceamento do *dataset* e, conseqüentemente, aumentar a acurácia balanceada obtida neste experimento.

Referências:

AGUILERA, R. C. A. O. de. Corruption perception index: Reflections in brazil before and during operation car wash. *Contextualizaciones Latinoamericanas*, v. 2, n. 21, 2019.

ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Taylor & Francis, v. 46, n. 3, p. 175–185, 1992.

BRASIL. Lei nº 10.520, de 17 de julho de 2002. *Diário Oficial da União*, Brasília, DF, 2002. ISSN 1677-7042. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/110520.htm.

BRASIL. Lei nº 12.529, de 30 de novembro de 2011. *Diário Oficial da União*, Brasília, DF, 2011. ISSN 1677-7042. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112529.htm.

BRASIL. Decreto nº 10.024, de 20 de setembro de 2019. *Diário Oficial da União*, Brasília, DF, 2019. ISSN 1677-7042. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10024.htm.

BRASIL. Lei nº 13.844, de 18 de junho de 2019. *Diário Oficial da União*, Brasília, DF, 2019. ISSN 1677-7042. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Lei/L13844.htm.

BRASIL, C. G. d. U. *Relatório de Gestão: exercício 2020*. 2020. Disponível em: <https://repositorio.cgu.gov.br/handle/1/65266>. Acesso em: 07 dezembro 2022.

BRASIL, C. G. d. U. *Detalhamento das Sanções Vigentes*. 2022. Disponível em: <https://portaldatransparencia.gov.br/sancoes/consulta>. Acesso em: 20 maio 2022.

BRASIL, C. G. d. U.; BRASIL, M. d. E. M. *Acordo de Cooperação Técnica CGU/SEGES-ME Nº 135/2021*. 2021. Disponível em: <https://repositorio.cgu.gov.br/handle/1/66596>. Acesso em: 12 dezembro 2022.

BRASIL, C.-G. d. U. C. *Licitações com contratação realizada*. 2022. Disponível em: <https://www.portaltransparencia.gov.br/licitacoes>. Acesso em: 17 novembro 2022.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

CADE. *Guia de Combate a cartéis em licitação*. [S.l.], 2019. Disponível em:

<https://cdn.cade.gov.br/Portal/centrais-de-conteudo/publicacoes/guias-do-cade/guia-de-combate-a-carteis-em-licitacao-versao-final-1.pdf>.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

DANTAS, D. d. Q.; MARTINS, L. B. Ferramenta alice: estudo sobre sua eficácia e eficiência no uso dessa ferramenta como fundamento para a prevenção e o combate à corrupção no âmbito da controladoria-geral da união. Universidade Católica de Brasília, 2022. Disponível em: <https://repositorio.cgu.gov.br/handle/1/68888>.

DESORDI, D.; BONA, C. D. A inteligência artificial e a eficiência na administração pública. *Revista de Direito*, Universidade Federal Viçosa, v. 12, n. 2, p. 1–22, 2020.

FARIA, E. R. d. *et al.* Fatores determinantes na variação dos preços dos produtos contratados por pregão eletrônico. *Revista de Administração Pública*, SciELO Brasil, v. 44, n. 6, p. 1405–1428, 2010.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, Elsevier, v. 55, n. 1, p. 119–139, 1997.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

García Rodríguez, M. J. *et al.* Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*, v. 133, p. 104047, 2022. ISSN 0926-5805. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0926580521004982>.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Machine learning*, Springer, v. 63, n. 1, p. 3–42, 2006.

HASTIE, T. *et al.* *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2.

HUBER, M.; IMHOF, D. Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, v. 65, p. 277–301, 2019. ISSN 0167-7187. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0167718719300219>.

MAIA, P. *et al.* Multicriteria anomaly detection in government purchases. In: SBC. *Anais do VII Symposium on Knowledge Discovery, Mining and Learning*. [S.l.], 2019. p. 97–104.

NAI, R. *et al.* Public procurement fraud detection and artificial intelligence techniques: a literature review. 2022.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *Fighting Hard-core Cartels Harm, Effective Sanctions and Leniency Programmes*. Paris: OECD Publishing, 2002. ISBN 1-280-08111-2. Disponível em: <https://www.oecd.org/competition/cartels/1841891.pdf>.

PANIS, A. d. C. Inovação em compras públicas: estudo de caso do robô alice da controladoria-geral da união (CGU). 2020.

RASMUSSEN, C. E.; WILLIAMS, C. K. I. *Gaussian processes for machine learning*. London, England: MIT Press, 2005. (Adaptive Computation and Machine Learning series).

ROCHA, A. L. M. d. *Ferramenta Alice: Auditoria Preventiva em Licitações. Fórum: O Controle no Combate à Corrupção/2019*. 2019. Disponível em: <https://repositorio.cgu.gov.br/handle/1/43580>. Acesso em: 17 junho 2022.

ZHANG, T. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In: *Proceedings of the twenty-first international conference on Machine learning*. [S.l.: s.n.], 2004. p. 116.

Apêndice A

Os scripts utilizados neste experimento, adaptados do trabalho de García Rodríguez *et al.* (2022), bem como o dataset gerado, estão disponíveis no github do autor e <https://anonymfile.com/NV3nn/extra.zip>.