Detecção de Fraudes em Licitações Públicas: Uma Comparação de Modelos de Detecção de Anomalias

Breno M. de Abreu[1], **Thomaz H.S. Pereira**[1], **Luiz Gomes-Jr**[1]

[1]DAINF - Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba - Brasil

{brenoabreu,thomazhugo}@alunos.utfpr.edu.br, lcjunior@utfpr.edu.br

Abstract. *Outlier detection in datasets is of great relevance to the accounting field where anomalous records can indicate fraud instances. This type of analysis can be used to mitigate risks and avoid financial loss, especially in the public sector. The goal of this project is to autonomously detect anomalies in real invoice data of biddings from the brazilian public sector, performing a comparative analysis between the algorithms Local Outlier Factor, Isolation Forest and Self-Organizing Maps regarding their ability to locate possible cases of fraud. The results showed that Isolation Forest was more efficient in detecting fraud compared to the other algorithms.*

Resumo. *A detecção de anomalias em conjuntos de dados é de grande relevância para a área contábil onde registros irregulares podem ser considerados indícios de fraude. Este tipo de análise pode ser usado para mitigar riscos e evitar perdas financeiras, sobretudo no setor público. Este trabalho tem como objetivo a detecção de anomalias em dados de notas fiscais reais de licitações do setor público brasileiro de forma autônoma, realizando uma análise comparativa entre os algoritmos Local Outlier Factor, Isolation Forest e Self-Organizing Maps em relação à sua capacidade de localizar possíveis casos de fraude. Os resultados mostraram que o Isolation Forest foi mais efetivo em detectar fraudes em comparação aos demais algoritmos.*

1. Introdução

Conforme estabelecido pela legislação brasileira, qualquer ação que utilize intencionalmente de manipulação da contabilidade com o objetivo de burlar os procedimentos de licitação favorecendo entidades específicas e assim prejudicando a Administração Pública, é considerada uma fraude em licitação [Gomes 2017].Um levantamento feito pela ONG Transparência Brasil detectou possíveis fraudes em contratos públicos entre fevereiro de 2020 e outubro de 2022 (período da pandemia de COVID-19), que foram firmados com empresas que não são do mesmo ramo do produto adquirido, fornecedores inscritos no Cadastro de Empresas Inidôneas e Suspensas (CEIS), empresas com faturamento suspeito e empresas abertas 30 dias ou menos antes do acordo firmado. Esses contratos suspeitos totalizam cerca de 2 bilhões de reais [Oliveira 2022], e ao ano o Instituto Ética Saúde estima um prejuízo anual de 22,54 bilhões de reais na área da saúde [Puente and Ameida 2021].

Dentre as várias formas de se fraudar uma licitação pública destaca-se o sobrepreço que, de acordo com a Lei Federal 14.133/21 Art. 6, ocorre quando o valor de

uma mercadoria ou serviço é "expressivamente superior aos preços referenciais de mercado", podendo ser considerada para o valor unitário ou global do objeto [Brasil 2021]. Pode-se ainda estabelecer o ato de superfaturamento por superdimensionamento que tem o objetivo de inflar o custo total da mercadoria, aumentando desnecessariamente a quantidade de produtos de forma a resultar no sobrepreço do contrato [Lopes et al. 2021].

A detecção de fraudes e sua prevenção é uma responsabilidade do setor administrativo da entidade, portanto, a atividade de auditoria se faz necessária para a identificação de fraudes e erros contábeis [Schwindt and Corazza 2008]. O processo de detecção envolve a busca por indícios de forma manual que muitas vezes dependem de denúncias para serem encontrados, culminando em uma porcentagem pequena de agentes públicos e empresas mal-intencionadas envolvidos em fraudes que são levados a julgamento. Entretanto, há a possibilidade de otimizar este processo através do uso de técnicas de aprendizado de máquina para detecção de *outliers* [Hamelers 2021] que possibilitam a identificação automática de anomalias nas bases de dados.

O objetivo deste estudo é, a partir dos dados de notas fiscais reais de licitações disponibilizados em um conjunto de dados estruturados e sem a categorização dos documentos em normais e fraudulentos, empregar técnicas de aprendizado de máquina não-supervisionado para detectar anomalias visando a aplicação de uma metodologia capaz de detectar fraudes. Para tal, esta pesquisa explora três algoritmos para detecção de outliers: *Local Outlier Factor*, *Isolation Forest* e *Self-Organizing Maps*.

O restante do artigo está organizado da seguinte forma: A Seção 2 apresenta os fundamentos e trabalhos relacionados. A Seção 3 descreve a metodologia usada. A Seção 4 apresenta os resultados enquanto as conclusões são descritas na Seção 5.

2. Fundamentos e Trabalhos Relacionados

Anomalias, também chamadas de *outliers*, são instâncias que exibem um comportamento divergente dos demais registros da amostra; ou seja, que apresentam atributos comportamentais inconformes com os esperados dado um determinado contexto. Assim, a detecção de *outliers* atua sobre um conjunto de registros onde cada indivíduo é analisado e, comparando seu comportamento com os demais, é categorizado como normal ou anômalo, obtendo uma delineação clara entre as duas classes.

Para realizar a detecção de *outliers* em dados estruturados é necessário considerar dois principais desafios: (i) raramente há uma diferenciação rotulada entre quais registros são normais e quais são anômalos, e (ii) há múltiplos contextos diferentes presentes nas amostras, o que dificulta uma análise global dos registros, assim, cada indivíduo deve ser analisado levando em consideração o contexto ao qual pertence.

Para o primeiro problema (i), como as bases de dados que serão analisadas raramente apresentam a distinção entre instâncias normais e anômalas, impõe-se a necessidade de utilizar o aprendizado não-supervisionado que irá identificar registros que fogem à normalidade em relação aos outros indivíduos sem a necessidade de haver uma classificação prévia do estado de cada entrada [Hamelers 2021][Paula et al. 2016]. Assume-se, neste caso, que os dados que não apresentam anomalias são mais frequentes que os que *outliers*, ou seja, há uma diferença de proporção considerável entre as duas classes e só assim é possível realizar a detecção destas anomalias para o tipo de dados em questão [Chandola et al. 2009].

Para o segundo problema (ii) caracteriza-se um atributo contextual aquele que identifica o contexto, ou vizinhança, ao qual um determinado registro pertence [Chandola et al. 2009]. Dessa forma, haverá diversas vizinhanças distribuídas pelo espaço de características e o que pode ser considerado uma anomalia para uma vizinhança pode não ser para outra dificultando, assim, a detecção de *outliers*.

Nas subseções a seguir serão descritos os algoritmos não-supervisionados Local Outlier Factor, Isolation Forest e Self-Organizing Maps, os métodos mais promissores encontrados para a detecção de fraudes no contexto da pesquisa.

2.0.1. Local Outlier Factor

Local Outlier Factor (LOF) é uma métrica aplicada para cada ponto em um conjunto de dados que visa calcular o grau de divergência deste ponto em relação aos seus k-vizinhos mais próximos. Para tal, compara-se a densidade local de uma entrada com a densidade local dos seus vizinhos; havendo uma disparidade significativa entre os resultados em que a densidade da entrada analisada é menor que a de seus vizinhos, pode-se classificar o registro como anômalo [Breunig et al. 2000].

Esta métrica permite encontrar *outliers* em contextos específicos, dado que a forma em que é calculada apenas leva em consideração os vizinhos mais próximos de uma instância. Assim, é ideal para identificar anomalias a partir de *clusters* de dados que apresentam densidades diferentes entre si. O método em questão determina que quanto maior o LOF de um ponto, mais provável de ser uma anomalia. Dessa forma, calcula-se o LOF para cada instância de um conjunto de dados a partir de uma quantidade predeterminada de vizinhos (k), e as instâncias com maior pontuação, dado um certo valor de contaminação, são compreendidos como *outliers*. O método também pode ser utilizado para localizar *clusters* de anomalias dependendo do valor de k determinado.

2.0.2. Isolation Forests

O Isolation Forest, também chamado apenas de iForest, é um algoritmo utilizado especificamente para a detecção de *outliers* e parte do princípio que como as anomalias são escassas e divergem das instâncias normais, há uma alta probabilidade de que com a construção de uma árvore de decisão estas sejam encontradas a uma distância menor da raiz da árvore [Liu et al. 2008]. Uma iForest apresenta as seguintes definições:

1. *Isolation Tree*: uma árvore de isolação onde os nós-folhas representam as saídas esperadas dada a combinação dos valores de cada atributo de um registro de entrada, e os demais são nós de decisão onde a partir do valor de uma determinada dimensão, separam os dados em dois grupos; ou seja, cada nó de decisão determina para qual o seguinte nó de decisão o algoritmo irá. No final dos testes o algoritmo indicará um nó-folha com o resultado esperado.
2. Distância do Caminho: é a distância entre um nó-folha e a raiz da árvore e é medido pela quantidade de arestas entre os nós.
3. *Anomaly Score*: é uma pontuação derivada da distância do caminho utilizada para determinar se um nó é ou não uma anomalia. No caso das Isolation Forests, um valor próximo de 1 indica uma anomalia; um valor menor que 0.5 quase certamente

indica uma instância normal; e valores acima de 0.6 indicam uma provável anomalia. Sendo assim, quando um conjunto de dados apresenta pontuações próximas de 0.5 para todas as instâncias pode-se concluir que não há anomalias significativas na base. O score é encontrado a partir do cálculo das distâncias entre um nó-folha e a raiz de múltiplas árvores de isolação, que formam uma floresta de isolação.

2.0.3. Self-Organizing Maps

Self-Organizing Maps (SOM) é um algoritmo utilizado para agrupamento e redução de dimensionalidade que pode também ser utilizado para a detecção de *outliers*. Seu funcionamento permite realizar a projeção dos dados em um espaço de dimensionalidade reduzida e em que as anomalias são encontradas através da análise da distância entre uma instância e seu neurônio mais próximo, chamado *Best Matching Unit* (BMU). O SOM faz uso de uma matriz de neurônios, normalmente com duas ou três dimensões, que com o processo de aprendizado tendem a acompanhar a topologia dos dados originais. Ou seja, a posição espacial dos neurônios após a etapa de aprendizagem se assemelha à posição espacial das instâncias do conjunto de dados mas em um espaço com um número menor de dimensões. Com isso, é possível deduzir que anomalias são encontradas a uma distância relativamente distante dos neurônios em comparação com as instâncias normais, ou em pequenos grupos que destoam dos demais possibilitando, assim, detectar registros que possuem um comportamento atípico [Brzezinska and Horyn 2022].

2.0.4. Trabalhos Relacionados

O uso do cálculo do Local Outlier Factor (LOF) para detecção de anomalias demonstrou resultados positivos no contexto de saúde pública [Shan et al. 2009]. A pesquisa concluiu que o LOF é um método efetivo para encontrar anomalias em faturamentos tendo sido validado com a ajuda de especialistas da área contábil. Os resultados são comparáveis com a análise aprofundada dos dados realizada por especialistas.

No contexto de detecção de fraudes financeiras em faturas, a utilização de iForests em comparação aos algoritmos Local Outlier Factor (LOF) e One Class Support Vector Machine (OCSVM) se mostra superior não somente na detecção de anomalias em si mas também na performance e interpretabilidade dos resultados [Hamelers 2021].

Também é possível utilizar abordagens que utilizam versões aprimoradas do SOM para permitir a detecção de registros fraudulentos com desvios mais sutis. Uma delas é a criação de um Growing Hierarchical Self-Organizing Map (GHSOM) que permite criar estruturas mais complexas de SOMs progressivamente construindo múltiplos mapas e os organizando em uma estrutura hierárquica em diferentes níveis, possibilitando lidar melhor com registros com alta dimensionalidade. O modelo apresentado facilita a detecção de de pequenos grupos de instâncias anômalas assim como de registros que se encontram relativamente distantes de seus respectivos neurônios mais próximos [Huang et al. 2014].

3. Metodologia

Para este trabalho foi disponibilizada uma base de dados estruturados contendo informações sobre documentos de licitação provindos do Sistema de Saúde do Estado

da Paraíba referentes ao ano 2016. A base apresenta 2,089,317 registros e os valores são distribuídos em 62 colunas. As linhas representam itens de uma nota fiscal, sendo que uma nota pode apresentar mais de um item. Sendo assim, existe uma relação 1 para N entre a nota fiscal e seus itens. Os dados são referentes a produtos farmacêuticos apenas e não há a classificação de cada registro que o identifique como fraude ou não. A base de dados em questão foi selecionada por ser a única fonte acessível no momento que apresenta os dados das notas fiscais de licitações de maneira tabular, sendo um formato conveniente para ser aplicado nos modelos de detecção de anomalias.

As principais informações presentes na base são: número identificador da nota; data de emissão; identificador das entidades emissora e destinatária; localização física das entidades emissora e destinatária; valor total da nota; outros valores que constituem o valor total da nota (valor da base de cálculo do Imposto Sobre Circulação de Mercadorias e Serviços (ICMS), do frete, entre outros); descrição de cada produto e seu código NCM [1]; valor unitário de cada produto; quantidade do produto; e outros valores que constituem o valor total do produto (valor do Imposto Sobre Circulação de Mercadorias e Serviços (ICMS), do Imposto Sobre Produtos Industrializados (IPI), entre outros).

A partir da base de dados original foram, aplicadas técnicas de limpeza de dados mantendo apenas as características mais relevantes. As variáveis excluídas foram retiradas por: apresentar interações com outras características (como o valor total da nota e valores de impostos que são calculados a partir da quantidade e do valor unitário do produto); ou conter quantidades expressivas de valores vazios; ou incluir textos que não são relevantes para auxiliar na localização contextual de uma instância (como o nome do bairro do destinatário ou descrições adicionais do produto). Ainda foram incluídos dois novos tipos de dados: o tempo em segundos a partir da Era Unix de quando a nota foi emitida e os valores da longitude e latitude das cidades. O conjunto de dados utilizados na etapa de aprendizado dos modelos contém as seguintes características:

- Tempo, em segundos a partir da Era Unix, da data de emissão da nota fiscal
- Latitude do emitente
- Longitude do emitente
- Latitude do destinatário
- Longitude do destinatário
- Valor do NCM
- Quantidade de itens
- Valor unitário do item

Após o processo de limpeza, transformação e normalização de dados, uma nova base foi criada contendo apenas os dados com o valor do NCM mais comum (código 3004.90.99). Essa decisão foi tomada para que fosse possível aplicar modelos de aprendizado de máquina, como o SOM, que demandam uma quantidade mais alta de memória em comparação a algoritmos como o iForest e o LOF e não poderiam ser aplicados com os dados completos. Com isso, além da quantidade reduzida de dados (totalizando 76,457 entradas), também houve a redução de dimensionalidade permitindo a melhor aplicação

[1] A Nomenclatura Comum do Mercosul (NCM) diz respeito é um código que "[...] permite, pela aplicação de regras e procedimentos próprios, determinar um único código numérico para uma dada mercadoria.". Dessa forma, cada tipo de produto possui um código NCM associado a ele e usado, fundamentalmente, na aplicação de tributos em operações de comércio exterior [Brasil 2019].

Tabela 1. Métodos e parâmetros utilizados.

Método	Biblioteca	Parâmetros
SOM	Minisom [Vettigli 2018]	contamination = 0.01 map size = 40x40 sigma = 3 learning rate = 0.5 neighborhood function = triangle random seed = 26 training iterations = 1,000,000
iForest	Scikit IsolationForest [Pedregosa et al. 2011]	contamination = 0.01 random state = 26 number of estimators = 100 max features = 1.0 max samples = auto
LOF	Scikit LocalOutlierFactor [Pedregosa et al. 2011]	contamination = 0.01 number of neighbors (k) = 10 algorithm = auto leaf size = 30 metric = minkowski

dos algoritmos de detecção e facilitando a análise dos resultados encontrados visto que todos os itens pertencem à mesma categoria.

3.1. Implementação dos Métodos

Três algoritmos de aprendizado de máquina para detecção de *outliers* foram aplicados sobre a base de dados: Self-Organizing Maps (SOM), Isolation Forest (iForest) e Local Outlier Factor (LOF). A Tabela 1 apresenta o valor dos parâmetros relevantes usados em cada modelo. Como não há *Gold Standard* (classificação de fraude) para os dados, não foi possível realizar uma etapa de otimização de parâmetros. Os parâmetros utilizados foram escolhidos por serem valores padrão ou valores que demonstraram boa razão entre precisão e eficiência.

As instâncias anômalas são descobertas analisando o seu *anomaly score* que pode ser inferido de diferentes maneiras para cada algoritmo. No caso do SOM, a pontuação é definida a partir do cálculo da distância de cada instância em relação a seu BMU. Os pontos com distâncias mais altas são considerados anomalias. Para o iForest a pontuação é encontrada a partir do cálculo da distância em que um nó-folha que representa uma instância está da raiz. Distâncias pequenas indicam prováveis *outliers*. E para o LOF a pontuação é o próprio valor do Local Outlier Factor que é calculado para cada instância. Quanto maior o valor, mais alta a probabilidade do registro ser anômalo.

A partir do *anomaly score* de cada instância, os resultados de cada algoritmo foram ordenados de forma que os registros com maior probabilidade de serem anomalias fossem encontrados no topo da lista. Então, as 10 instâncias com pontuações maiores foram escolhidas a fim de serem analisadas manualmente pelos desenvolvedores que, ao realizar comparações entre as instâncias categorizadas como anômalas e os demais registros relevantes, puderam avaliar os resultados encontrados a partir dessa amostra.

Um exemplo de análise realizada pode ser observada na Figura 1 em que o modelo gerado pelo iForest apontou duas instâncias anômalas para o mesmo produto. O primeiro passo para a análise foi identificar as características relevantes sobre o registro categorizado como *outlier* pelo algoritmo; no caso do exemplo, foram utilizadas as informações apresentadas pelo campo de descrição do produto onde são informados seu nome e dosagem. A partir dessa descrição, foram escolhidos os termos ”sinvastatina” e ”40mg” para realizar uma pesquisa na base de dados e permitir a comparação entre os registros. Como é possível perceber na tabela resultante da pesquisa, as instâncias com índice 71968 e 71970 (precisamente os registros apontados pelo iForest), apresentam um valor unitário relativamente mais baixo que as demais instâncias, assim como quantidades expressivamente mais altas. Esses registros foram identificados como verdadeiras anomalias pelo desenvolvedor e dois possíveis casos de superdimensionamento.

É importante, porém, ressaltar que os valores encontrados podem ser provenientes de diferentes contextos como a localização da entidade destinatária ou situações anômalas mas legítimas que podem justificar a compra de quantidades altas do produto. Sendo assim, em trabalhos futuros, é importante incluir mais dados de contextualização como variáveis socioeconômicas dos lugares que realizaram a compra dos produtos, e realizar uma análise mais minuciosa para procurar razões que podem justificar os valores encontrados nas notas fiscais.

Figura 1. Exemplo de comparação de instâncias anômalas (índice 71968 e 71970) com registros similares (parte das linhas e colunas foram omitidas para facilitar a clareza e legibilidade).

	prod_ncm	prod_desc	prod_quant	prod_valor_unit	prod_valor_total
18896	30049099	sinvastatina 40mg (sanval) sinvaston	30.0	0.28	8.4
18924	30049099	sinvastatina 40mg (sanval) sinvaston - lote: at412 31/05/2017	100.0	0.35	35.0
19123	30049099	sinvastatina 40mg (sanval) sinvaston - lote: at412 31/05/2017	500.0	0.19	95.0
20331	30049099	sinvastatina 40mg comprimidos	500.0	0.37	185.0
20336	30049099	sinvastatina 40mg comprimidos	1020.0	0.37	377.4
21195	30049099	sinvastatina 40mg comprimidos	2100.0	0.21	441.0
23443	30049099	sinvastatina 40mg (sanval) sinvaston	300.0	0.35	105.0
23913	30049099	sinvastatina 40mg comprimidos	2100.0	0.21	441.0
26088	30049099	sinvastatina 40mg (sanval) sinvaston	500.0	0.35	175.0
26491	30049099	sinvastatina 40mg comprimidos	2100.0	0.21	441.0
27046	30049099	sinvastatina 40mg (sanval) sinvaston	600.0	0.19	114.0
27606	30049099	sinvastatina 40mg comprimidos	394.0	0.40	157.6
28264	30049099	sinvastatina 40mg comprimidos	500.0	0.37	185.0
71968	30049099	sinvastatina 40mg comprimido multilab	150000.0	0.12	18000.0
71970	30049099	sinvastatina 40mg comprimido multilab	60000.0	0.13	7800.0

4. Resultados

Para examinar os resultados foi necessário avaliá-los manualmente já que, dado que os métodos utilizados tratam de aprendizado não-supervisionado, não é possível avaliá-los utilizando as ferramentas convencionais usadas na avaliação de modelos de aprendizado supervisionado e semi-supervisionado. Além disso, os autores não contaram com o

auxílio de um especialista na área que poderia indicar se os casos apontados pelos algoritmos são, de fato, casos de fraude. Sendo assim, foi utilizado um método de avaliação que parte da análise manual dos resultados encontrados pelos algoritmos e sua categorização em quatro grupos com características distintas, como explicado abaixo. A classificação foi realizada de forma subjetiva, baseada na interpretação do avaliador ao observar uma instância e compará-la com outros registros referentes ao mesmo produto.

Para avaliar a eficácia dos métodos LOF, iForest e SOM foram escolhidos os 10 *outliers* com maior pontuação de anomalia encontrados por cada modelo. Os resultados foram analisados manualmente pelos desenvolvedores e classificados levando em consideração a potencialidade do registro ser fraudulento.

As potenciais anomalias foram classificadas em quatro categorias:

- Categoria A - Potencial fraude (significativo): quando há fortes indícios de que a instância é uma fraude, apresentando valores discrepantes (como quantidades e valor unitário do item) em relação aos demais registros;
- Categoria B - Potencial fraude (moderado): quando há evidências que a instância é uma possível fraude mas não há dados para comparação suficientes para realizar uma afirmação segura;
- Categoria C - Potencial não-fraude: quando não há evidências que a instância é uma fraude e os valores encontrados são condizentes com os dos demais registros, sendo assim um falso-positivo;
- Categoria D - Indeterminado: quando a instância está isolada e não permite uma comparação com os demais registros.

A Figura 2(a) apresenta os resultados encontrados para cada grupo em relação a porcentagem de instâncias que, de fato, são anomalias. Para tal, considera-se o número de registros classificados como tipo A, B ou D para Anomalias (verdadeiros-positivos) e do tipo C como Normais (falso-positivos). A Figura 2(b) apresenta os resultados encontrados para cada grupo e para cada categoria individual.

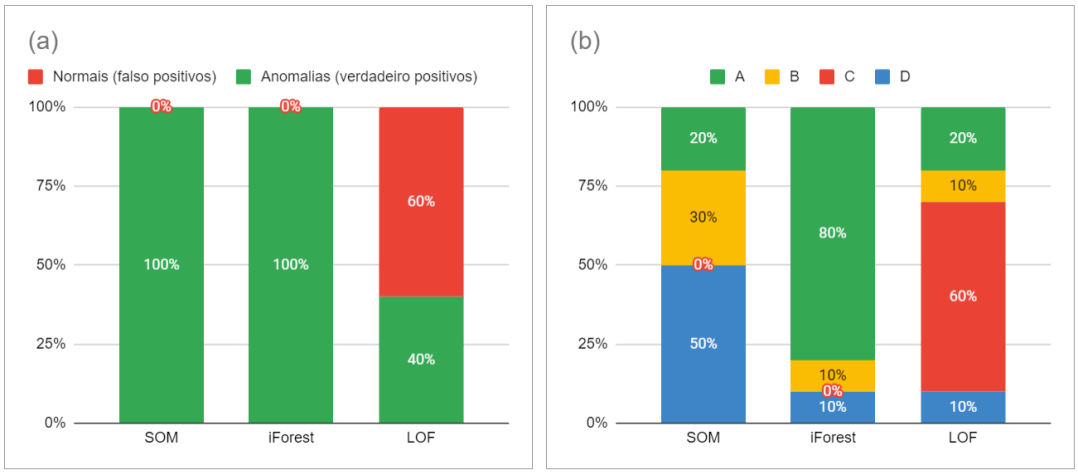

Figura 2. (a) Proporção de instâncias anômalas e normais para cada grupo. (b) Proporção de instâncias pertencentes a cada categoria para cada grupo.

Percebe-se que os 10 principais *outliers* encontrados para o SOM e para o iForest são verdadeiras anomalias. Porém, o iForest apresentou resultados que possuem uma pro-

babilidade mais alta de apresentarem algum tipo de fraude, enquanto que o SOM indicou mais instâncias isoladas ou que não apresentam indícios fortes de serem fraudulentas. No caso do LOF, a maior parte dos resultados foram falso-positivos, ou seja, instâncias que não apresentam nenhuma evidência de serem nem anomalias, nem possíveis fraudes.

Ainda, é necessário apontar que as instâncias classificadas como sendo do grupo A em geral apresentaram valores unitários mais baixos em comparação com as demais instâncias mas uma quantidade significativamente mais alta, o que pode indicar casos de superdimensionamento. Casos de sobrepreço foram raramente encontrados nas amostras analisadas.

5. Conclusão e Trabalhos Futuros

Os resultados obtidos ao longo deste trabalho elucidam a suposição que, a partir do uso de métodos de aprendizado de máquina, é possível descobrir possíveis casos de fraude de forma a agilizar seu processo de detecção juntamente a um auditor.

O objetivo de avaliar a eficácia dos modelos estudados em sua capacidade de detectar anomalias foi alcançado, possibilitando localizar algoritmos úteis para a resolução do problema e determinar a qualidade de seus resultados. Dos métodos utilizados, destaca-se o iForest, que se mostrou mais promissor na detecção de possíveis instâncias fraudulentas em comparação aos demais métodos apresentados aqui.

Conclui-se que estes métodos têm a capacidade de detectar anomalias no contexto financeiro e podem ser utilizados em organizações como uma ferramenta auxiliar na detecção de fraudes. Porém, é necessário ressaltar a importância da análise humana, propriamente de um auditor especializado, para validar os resultados encontrados pelos modelos. Os algoritmos servem como instrumentos para agilizar o processo e detectar divergências possivelmente imperceptíveis, mas para determinar a veracidade dos resultados é fundamental analisar os contextos a partir do ponto de vista humano, levando em conta todas as sutilezas e complexidades contextuais do sistema financeiro.

Entre possíveis trabalhos futuros destacamos:

1. Encontrar uma forma melhor de identificar os tipos diferentes de produtos na base de dados permitindo um maior detalhamento e padronização da descrição dos produtos;
2. Ampliar as variáveis de contextualização adicionando dados socieconômicos;
3. Compreender mais profundamente os dados originais com o auxílio de especialistas na área possibilitando uma melhor preparação dos dados;
4. Utilizar um número maior de amostras na etapa de avaliação;
5. Avaliar os resultados com o auxílio de especialistas na área.

Referências

Brasil (2019). Ncm.

Brasil (2021). Lei federal nº 14.133/21 art. 90, de 1 de abril de 1993. *Diário Oficial [da] República Federativa do Brasil*. Acessado em: 25 Outubro 2023.

Breunig, M., Kriegel, H.-P., Ng, R., and Sander, J. (2000). Lof: Identifying density-based local outliers. *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. Acessado em: 15 Outubro 2023.

Brzezinska, A. and Horyn, C. (2022). Self-organizing map algorithm as a tool for outlier detection. *Procedia Computer Science*. Acessado em: 27 Maio 2023.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3). Acessado em: 27 Maio 2023.

Gomes, Y. (2017). Fraude em licitação no regime militar de 1964. *Anais do IV Simpósio de História do Direito*. Acessado em: 03 Junho 2023.

Hamelers, L. (2021). Detecting and explaining potential financial fraud cases in invoice data with machine learning. *University of Twente*. Acessado em: 27 Maio 2023.

Huang, S.-Y., Tsaih, R.-H., and Yu, F. (2014). Topological pattern discovery and feature extraction for fraudulent financial reporting. *Expert Systems with Applications*. Acessado em: 27 Maio 2023.

Liu, F., Ting, K., and Zhou, Z.-H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*. Acessado em: 27 Maio 2023.

Lopes, A., Raupp, A., Magro, R., and signor, R. (2021). O superfaturamento está definido na lei nº 14.133/2021, e agora? Acessado em: 15 Outubro 2023.

Oliveira, C. (2022). Governo bolsonaro: possíveis fraudes durante pandemia de covid-19 somam r$ 2 bilhões: Transparência brasil analisou 248 compras e contratações de serviços firmadas entre fevereiro de 2020 e outubro de 2022. Acessado em: 04 maio 2023.

Paula, E., Ladeira, M., Carvalho, R., and Marzagão, T. (2016). Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering. *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. Acessado em: 27 Maio 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Puente, B. and Ameida, P. (2021). Brasil pode perder mais de r$ 20 bilhões por ano com desvios na saúde. cnn brasil, rio de janeiro. 2021. Acessado em: 04 maio 2023.

Schwindt, C. and Corazza, H. (2008). *Princípios Fundamentais e Normas Brasileiras de Contabilidade*. CFC, Brasília.

Shan, Y., Murray, D., and Sutinen, A. (2009). Discovering inappropriate billings with local density based outlier detection method. *AusDM '09: Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*. Acessado em: 15 Outubro 2023.

Vettigli, G. (2018). Minisom: minimalistic and numpy-based implementation of the self organizing map. Acessado em: 15 Outubro 2023.