

# Estimating the Risk of Failure in Government Auctions in Brazil

João B. R. Fonseca<sup>1,2</sup>, Luis F. P. Garcia<sup>1</sup>, Marcos F. Caetano<sup>1</sup>

<sup>1</sup>Departamento de Ciências da Computação – Universidade de Brasília (UnB)  
Brasília - DF - Brasil

<sup>2</sup>Tribunal de Contas da União - TCU  
Brasília - DF - Brasil

fonsecajb@tcu.gov.br, luis.garcia@unb.br, mfcaetano@unb.br

**Abstract.** *The Brazilian federal government conducts approximately 63 thousand electronic auctions annually, spending around R\$190 billion. Oversight agencies face challenges in monitoring all procurement due to the high volume and cost. This paper proposes a model to rank auctions based on the risk of failures. The results indicate that the machine learning model based on the XG-Boost algorithm is effective in identifying auction items with a higher risk of serious failures, with a recall of 84% for the class of interest. The model has the potential to optimize oversight resources, directing efforts towards the most critical auctions, increasing the probability of detecting failures, and enhancing the control expectations of government agencies.*

**Resumo.** *O governo federal brasileiro realiza aproximadamente 63 mil pregões eletrônicos por ano, com despesas de cerca de R\$190 bilhões. Agências de fiscalização enfrentam desafios para monitorar essas aquisições devido ao alto volume e custo. Este artigo propõe um modelo para classificar os pregões com base no risco de irregularidades. Os resultados obtidos indicam que o modelo de aprendizagem de máquina baseado no algoritmo XGBoost é eficaz na identificação de itens de pregão com maior risco de irregularidades graves, com um recall de 84% para a classe de interesse. O modelo tem o potencial de otimizar os recursos de fiscalização, direcionando os esforços para os pregões mais críticos, aumentando a probabilidade de detectar irregularidades e a expectativa de controle dos órgãos governamentais fiscalizados.*

## 1. Introdução

Para aquisição de bens e serviços o governo brasileiro utiliza leilões reversos, sendo vencedor o vendedor que ofertar o menor preço entre os concorrentes após uma sucessão de lances. Anualmente, o governo federal brasileiro conduz cerca de 63 mil leilões dessa natureza, denominados pregões, com gasto anual da ordem R\$190 bilhões [Brasil 2024].

Em termos teóricos, a competição entre fornecedores garantiria a aquisição de produtos e serviços pelo menor preço, mas esse objetivo pode ser frustrado se existir conluio entre licitantes, conluio entre comprador e licitante ou uma combinação de ambos. Em todos os casos, o efeito é a aquisição de produtos e serviços a um preço superior ao que seria obtido caso existisse efetiva competição entre fornecedores [Mariotto 1991].

As agências de fiscalização brasileiras – TCU, CGU, por exemplo – têm a finalidade de identificar e punir ilícitos. Para além dos aspectos punitivos, a fiscalização tem caráter preventivo e pedagógico, gerando expectativa de controle nos jurisdicionados [Carvalho Neto et al. 2011].

Embora não existam medidas oficiais, a opinião dos especialistas em compras públicas consultados é de que irregularidades em pregões são exceção, sendo provavelmente inferior a 1% a proporção de pregões com irregularidades graves. Selecionar tais pregões influenciaria a percepção de efetividade da fiscalização entre os jurisdicionados, aumentando a expectativa de controle dos órgãos governamentais. Em outros termos, o interesse na introdução deliberada de irregularidades em pregões seria diminuído, pois, em termos de expectativa de ganhos econômicos, a propensão a desvios éticos por servidores públicos é proporcional à probabilidade de punição [Cunha and Soares 2002].

Em regra, a fiscalização é realizada sobre uma pequena amostra da população de interesse, pois o custo de fiscalizar não pode ser superior ao seu benefício. Para lidar com o volume diário de pregões, foram criadas verificações automáticas da ocorrência de padrões de irregularidades. A organização dos dados simplificou a fiscalização ao reduzir a assimetria informacional entre os gestores e a agência fiscalizadora. Apesar disso, não existe mecanismo de seleção de pregões para fiscalização com base em risco; a seleção é feita por julgamento profissional e a escolha reflete o conhecimento prévio sobre o assunto ou heurísticas pessoais do responsável pela seleção. Nesse contexto, irregularidades fora dos padrões previamente conhecidos não serão identificadas.

Este artigo apresenta um modelo para ordenação de pregões segundo critérios de risco, gerando uma lista ordenada de pregões segundo a probabilidade de irregularidade. A premissa é de que pregões com irregularidades deliberadamente introduzidas possuem padrões distintos de pregões regularmente conduzidos, sendo passíveis de classificação por meio de aprendizagem de máquina. O objetivo é classificar itens de pregão segundo o risco de irregularidade e comparar a performance de classificadores de acordo com a métrica de avaliação escolhida (recall).

## **2. Trabalhos Correlatos**

A análise de dados referentes a licitações e contratos tem como desafio inicial a coleta e processamento dos dados [Menezes et al. 2023, Brandão et al. 2024, Maia et al. 2020]. As dificuldades de tratamento e rotulagem dos dados obrigam o uso de recortes específicos, normalmente uma combinação de critérios geográficos, temporais, temáticos ou de valor. Para enfrentar esse problema, [Brandão et al. 2024] propõe um pipeline semiautomatizado formado por um meta-classificador heurístico e um módulo de qualidade de dados e o utiliza para detecção de fraudes e superfaturamento em compras públicas.

A priorização de esforços de auditoria em licitações e contratos por meio de estimação automática de risco por aprendizagem de máquina é tratada em [Menezes et al. 2023, Maia et al. 2020, Henrique et al. 2020, Sun and Sales 2018]. [Maia et al. 2020] propõe metodologia de detecção de anomalias para identificar fraudes e conluio entre licitantes mediante uso de regras de associação, correlação temporal e análise multicritério, tendo sido aplicada na análise de pregões eletrônicos de agências federais no estado de Minas Gerais, detectando possível conluio entre cinco fornecedores. [Sun and Sales 2018] usa redes neurais artificiais tradicionais e profun-

das para previsão de irregularidades em compras públicas e triagem inicial de contratos. [Henrique et al. 2020] classifica contratos públicos em relação ao risco de não cumprimento de seus termos de entrega usando modelos estatísticos e de aprendizado de máquina. [Menezes et al. 2023] estima o risco de falha em contratos públicos com uso de aprendizado de máquina em nível municipal. Seus resultados indicam pouca eficácia de uso de variáveis ao nível de fornecedores, sendo considerada promissora a abordagem ao nível de contrato, em particular quando associada com indicadores econômicos para caracterização dos municípios compradores.

Em termos de avaliação de algoritmos para tarefas de classificação, [Shwartz-Ziv and Armon 2022] compara os resultados de classificação usando XGBoost e redes neurais profundas em dados tabulares. Em todos os cenários testados, os classificadores baseados no algoritmo XGBoost têm desempenho superior aos classificadores baseados em redes neurais profundas.

Diferentemente dos trabalhos de detecção de anomalias e classificação citados, cujos objetos foram recortes específicos do universo de compras e contratos, neste artigo é apresentado um classificador binário de itens de pregão com aplicação universal, baseado em atributos do item de compra, do comprador, do fornecedor vencedor e da disputa entre os concorrentes em cada item.

### 3. Metodologia

A metodologia adotada seguiu o modelo de referência CRISP-DM e o fluxo de trabalho representado na figura 1.

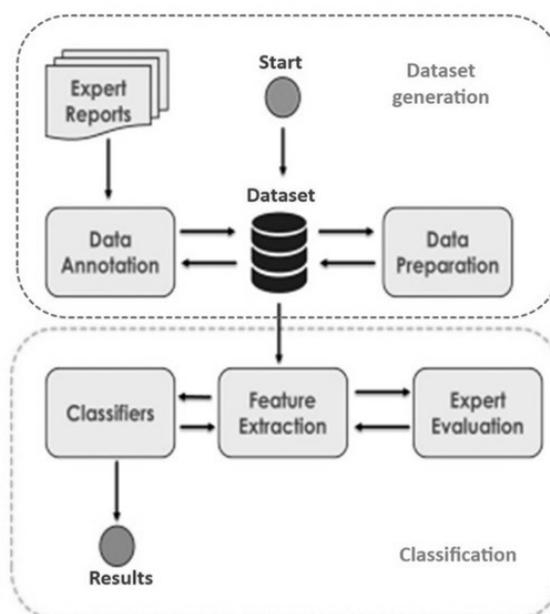


Figura 1. Fluxo de trabalho metodológico

#### 3.1. Anotação dos Dados

Utilizando os dados de pregões constantes da base SIASG e com auxílio de especialista em fiscalização de compras e contratos do TCU, foram selecionados os itens de pregão que atendiam simultaneamente aos seguintes critérios:

1. pregões realizados pelo governo federal entre janeiro de 2016 e março de 2024;
2. itens de pregão com vencedores conhecidos;
3. itens de pregão com valor homologado maior ou igual a cem mil reais, valor considerado o mínimo viável para justificar o exercício da fiscalização pelo TCU, pois os custos de fiscalização devem ser inferiores aos benefícios esperados.

Os registros de ocorrência de penalizações graves de fornecedores por irregularidades em contratos foram associadas aos itens de pregão de onde originaram-se tais contratos. Esses itens foram rotulados com a classe 1, classe positiva para ocorrência de irregularidades, gerando uma base de 3.651 exemplos.

A classe 0, classe negativa para ocorrência de irregularidades, foi rotulada com auxílio de especialista. Observe-se que a inexistência de irregularidade conhecida não é suficiente para caracterizar o item como sendo da classe 0, pois somente após avaliação de um especialista pode-se afirmar se existem ou não irregularidades. O especialista listou hipóteses sugestivas de irregularidades e isso foi utilizado como critério de exclusão de itens de pregão, de modo que, existindo qualquer indício de irregularidade associado a determinado item de pregão, ele não integraria a classe 0.

Critérios para utilizados para exclusão dos itens de pregão:

1. fornecedor vencedor com histórico de punição por irregularidade contratual;
2. registro de falha na execução de contrato derivado do item de pregão;
3. existência de apontamento sobre o pregão nos sistemas de análise automatizada utilizados pelo TCU.

Da base original de itens de pregão, utilizando o conjunto de regras de exclusão listados, 32.147 itens de pregão foram rotulados como classe 0. Desse total, foram escolhidos aleatoriamente 3.651 exemplos, gerando uma base de dados balanceada com 50% de exemplos de cada classe<sup>1</sup>.

### 3.2. Seleção de Atributos

Foram escolhidas 43 variáveis para caracterização dos itens de pregão – a lista de variáveis e as descrições estão disponíveis no apêndice deste artigo. Basicamente, as variáveis escolhidas caracterizam nos aspectos essenciais as seguintes entidades:

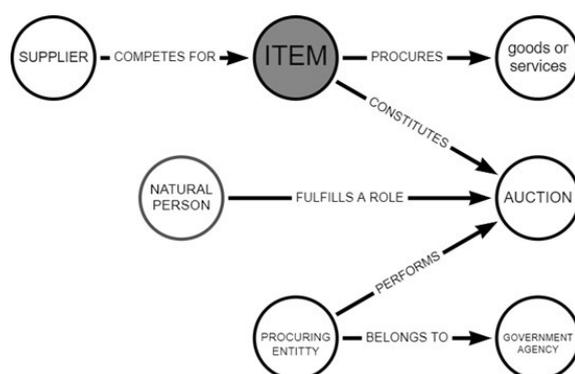
1. Item: objeto da disputa entre os fornecedores concorrentes;
2. Material ou serviço: material ou serviço adquirido em cada item de pregão.
3. Pessoa: pessoas físicas responsáveis pelo pregão, na condição de pregoeiro ou homologador;
4. Fornecedor: fornecedor vencedor do item de pregão;
5. Pregão: unidade agregadora de itens de pregão; : unidade governamental responsável pela realização de pregões;
6. Órgão: unidade governamental agregadora de unidades compradoras.

A figura 2 reproduz as relações entre as entidades.

Em termos gerais, cada item de pregão é um leilão em particular para aquisição de produtos ou serviços. Um pregão é realizado por uma unidade compradora, pode ter um ou mais itens. Cada unidade compradora pertence um órgão e um órgão pode ter mais de uma unidade compradora.

---

<sup>1</sup>O conjunto de dados rotulados pode ser consultado em <https://github.com/fonsecajb/pregao>



**Figura 2. Relações entre entidades**

### 3.3. Preparação dos Dados

A variável categórica ID\_LCAL\_UF foi dividida em variáveis binárias por meio de one-hot-encoding. Os valores nulos nas variáveis QTD\_LANCES\_VENCEDOR, VL\_NEGOC\_SESSAO\_PUBL, VL\_UNIT\_PROPOSTO, VL\_TOTAL\_PROPOSTO foram substituídos por zero.

O conjunto de variáveis foi padronizado usando z-score, com média 0 e variância 1. O conjunto de dados foi separado em conjunto de treinamento (75%) e teste (25%), com registros escolhidos aleatoriamente.

## 4. Experimento

Inicialmente foram treinados modelos utilizando os algoritmos Decision Tree, Light GBM, Random Forest, Regressão Logística, SVC e XGBoost, usando a configuração padrão da biblioteca scikit-learn. Em todos os casos foram utilizadas dez partições estratificadas para validação cruzada.

Para avaliação dos resultados, foi utilizado o recall como métrica. O recall é assim definido:

$$Recall = \frac{Verdadeiro\ Positivo}{Verdadeiro\ Positivo + Falso\ Negativo}$$

O recall é adequado para seleção de itens porque um valor alto para essa métrica indica que o modelo é capaz de identificar a maioria dos casos positivos, embora possa produzir números maiores de falsos positivos.

Os resultados obtidos com o conjunto de treinamento estão representados na tabela 1.

Por ter obtido o melhor resultado na seleção inicial, o algoritmo XGBoost foi escolhido para refinamento do modelo, mediante ajuste de hiperparâmetros. Em sua versão final, foram adotados os seguintes hiperparâmetros: learning\_rate=0.01, n\_estimators=70, max\_depth=3, min\_child\_weight=1, gamma=0.0.

Os resultados estão exibidos na figura 3.

O recall de 0,8434 indica que o modelo consegue identificar aproximadamente corretamente 84,34% das instâncias positivas.

**Tabela 1. Recall dos Classificadores**

<b>Classificador</b>	<b>recall</b>
XGBoost	0.9624
Light GBM	0.959
Random Forest	0.925
Decision Tree	0.894
KNeighbors	0.876
SVC	0.847
Logistic Regression	0.785

Evaluation Metrics Report

---

	precision	recall	f1-score	support
0	0.8210	0.7185	0.7664	913
1	0.7498	0.8434	0.7938	913
accuracy			0.7809	1826
macro avg	0.7854	0.7809	0.7801	1826
weighted avg	0.7854	0.7809	0.7801	1826

**Figura 3. Métricas de Avaliação**

A métrica precisão mede a proporção de verdadeiros positivos em relação ao total de previsões positivas feitas pelo modelo. Ela indica a exatidão das previsões positivas feitas pelo modelo.

$$Precisao = \frac{Verdadeiro\ Positivo}{Verdadeiro\ Positivo + Falso\ Positivo}$$

Uma precisão de 0,7498 indica que em 74,98% das vezes que o modelo previu uma instância como positiva, ela realmente era positiva.

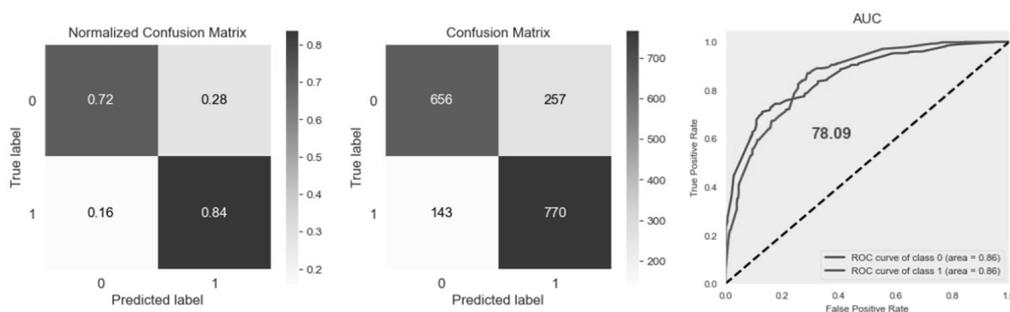
Por fim, a métrica f1-score é a média harmônica entre a precisão e o recall, ponderando a capacidade de encontrar instâncias positivas e a precisão das previsões realizadas. Um f1-score de 0,7938 indica equilíbrio entre a correção das previsões de verdadeiros positivos e a perda de instâncias positivas.

A figura 4 mostra a matriz de confusão e as curvas ROC para as classes 0 e 1.

O valor da AUC = 78,09 indica que há uma probabilidade de 78,09% de que o modelo classificará corretamente uma instância escolhida aleatoriamente, sugerindo um bom desempenho geral do modelo na distinção entre classes.

Especificamente para a classe 1, tem-se uma AUC = 0,86, o que indica uma probabilidade de 86% de que o modelo escolherá corretamente uma instância da classe 1 em comparação com uma instância da classe 0, ambas escolhidas aleatoriamente.

Como o uso previsto do modelo é a seleção para fiscalização de um número arbitrário relativamente pequeno de itens de pregão, cabe avaliar o resultado do modelo



**Figura 4. Matriz de Confusão e AUC**

considerando a probabilidade atribuída pelo modelo de serem da classe 1, ordenadas de forma decrescente segundo a probabilidade. Para as cem primeiras instâncias da classe 1, a acurácia, a precisão, o recall e a f1-score foi igual a 1, pois todas as instâncias foram classificadas corretamente. Para as duzentas primeiras instâncias, 191 foram classificadas corretamente, tendo gerado a seguintes métricas: acurácia = 0,955; recall = 1; precisão = 0,955, f1-score = 0,977.

## 5. Conclusão e Trabalhos Futuros

O classificador XGBoost é promissor para classificação de risco de irregularidades graves em itens de pregão. O valor de recall obtido indica a possibilidade de uso do emprego na seleção de pregão para fiscalização em situações de uso real. A possibilidade de ordenar os itens de pregão segundo a probabilidade de serem irregulares permite ao auditor a seleção conforme um limiar definido de acordo com suas necessidades. Nos dados de teste, considerando uma lista ordenada de forma decrescente de probabilidade das instâncias classificadas pertencerem à classe 1, considerando as 200 primeiras instâncias, os resultados indicam valores iguais ou próximos a 1 para as métricas acurácia, recall, precisão e f1-score.

Para trabalhos futuros, pretende-se criar modelos multiclases, com previsão de classes de irregularidades em itens de pregão, e criar modelos de previsão de irregularidades em pregões aplicáveis em contextos específicos, como saúde, educação, compras militares, obras, por exemplo.

## Referências

- Brandão, M. A., Reis, A. P. G., Mendes, B. M. A., Almeida, C. A. B. D., Oliveira, G. P., Hott, H., Gomide, L. D., Costa, L. L., Silva, M. O., Lacerda, A., and Pappa, G. L. (2024). PLUS: A Semi-automated Pipeline for Fraud Detection in Public Bids. *Digital Government: Research and Practice*, 5(1), 1-16.
- Brasil. (2024). Painel de Compras. Ministério da Gestão e da Inovação em Serviços Públicos. Disponível em: <http://paineldecompras.economia.gov.br>. Consulta em 11/7/2024.
- Carvalho Neto, A., Freitas, C., Cruz, I., and Akutsu, L. (2011). Auditoria governamental. Brasília: TCU, Instituto Serzedello Correa.

- Cunha, M., and Soares, M. (2002). Um estudo introdutório para mensurar o grau de exposição dos órgãos governamentais ao risco do desvio ético. In BRASIL. Desvios éticos: risco institucional, pages 13-23. Brasília: Comissão de Ética Pública.
- Domingos, S. L., Carvalho, R. N., and Ramos, G. N. (2016). Identifying it purchases anomalies in the brazilian government procurement system using deep learning. In *15th IEEE ICMLA*, pages 722-727.
- Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2020). Contracting in brazilian public administration: A machine learning approach. *Expert Systems*, 37(5).
- Lyra, M. S., Damásio, B., Pinheiro, F. L., and Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science*, 7(1).
- Maia, P., Meira Jr., W., Cerqueira, B., and Cruz, G. (2020). Auditing Government Purchases with a Multicriteria Anomaly Detection Strategy. *Journal of Information and Data Management*, 11(1), 50-65.
- Mariotto, F. L. (1991) O conceito de competitividade da empresa: uma análise crítica. *Revista de Administração de Empresas*, 31(2), 37-52.
- Menezes, T. L. de, Andrade, N. F. de, and Morais, F. J. A. (2023). The Effectiveness of Machine Learning to Estimate the Risk of Failure in Brazilian Public Contracts. In *2023 International Conference on Machine Learning and Applications (ICMLA)*, pages 2071-2078.
- Shwartz-Ziv, R., and Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90.
- Sun, T., and Sales, L. J. (2018). Predicting Public Procurement Irregularity: An Application of Neural Networks. *Journal of Emerging Technologies in Accounting*, 15(1), 141-154.

## Apêndice - Lista de Variáveis

1. RP: indicador de uso de registro de preço – pregão para aquisição futura.
2. ID\_UNDD\_RESP\_COMPRA: unidade governamental compradora.
3. ID\_UNDD\_ORGAO: órgão comprador, agregador de unidades compradoras
4. ID\_UNDD\_ORGAO\_SUP: órgão superior, agregador de órgãos compradores.
5. ID\_LCAL\_UF: unidade da federação (estado) onde foi realizado o pregão.
6. ID\_CMPR\_CPF\_PREGOEIRO: identificador numérico da pessoa responsável pela condução do pregão.
7. ID\_CMPR\_CPF\_HOMOLOGADOR: identificador numérico da pessoa responsável pela validação do resultado do pregão.
8. QTD\_MESES\_VALIDADE\_ATA: número de meses de validade dos preços para aquisição futura em caso de uso de registro de preços.
9. QT\_ITCP\_LANCES\_ITEM: número de lances recebidos na disputa pelo item.
10. ID\_ITCP\_CLASSE\_MAT\_SERV: classe do material ou serviço adquirido
11. ID\_ITCP\_PADRAO\_DESC\_MAT: descrição do material ou serviço
12. ID\_ITCP\_TP\_MATERIAL\_SERVICO: tipo de material ou serviço
13. ID\_ITCP\_SIT\_ATUAL\_MAT\_SERV: situação do material ou serviço
14. ID\_ITCP\_GRUPO\_MATERIAL: grupo do material
15. ID\_ITCP\_GRUPO\_SERVICO: grupo do serviço
16. FORNECEDOR: identificador do fornecedor vencedor da disputa no item do pregão
17. QT\_OFERTADA: quantidade ofertada do material ou serviço pelo fornecedor
18. VL\_UNIT\_PROPOSTO: valor unitário proposto pelo fornecedor no lance inicial
19. VL\_TOTAL\_PROPOSTO: valor total proposto pelo fornecedor no lance inicial
20. VL\_ULT\_LANCE: valor do último lance do fornecedor na disputa pelo item
21. VL\_NEGOC\_SESSAO\_PUBL: valor negociado pelo comprador como fornecedor após a definição do vencedor do pregão
22. VL\_ACEITO: valor pactuado entre o comprador e o fornecedor
23. VL\_ULT\_RENEG\_ATA\_SRP: valor pactuado entre o comprador e o fornecedor na última negociação de preços em pregão para aquisição futura.
24. VL\_PRECO\_UNIT\_HOMOLOG: valor a ser pago ao fornecedor por unidade de material ou serviço
25. VL\_PRECO\_TOTAL\_HOMOLOG: valor a ser pago ao fornecedor pelo fornecimento de todas as unidades de material ou serviço previstas
26. PRAZO\_ABERTURA\_SESSAO: número de dias entre o anúncio do pregão e a data da disputa.
27. DURACAO\_SESSAO: número de dias de duração da sessão de disputa entre fornecedores concorrentes.
28. PRAZO\_HOMOLOGACAO: número de dias entre a definição do vencedor do item de pregão e a validação do resultado.
29. PRAZO\_RESULTADO: número de dias para publicação do resultado após o início da sessão de disputa entre fornecedores.
30. PRAZO\_TOTAL: número de dias entre a publicação do edital e a validação do resultado.
31. PRAZO\_PROPOSTA: número de dias entre o anúncio do pregão e a data de entrega da proposta inicial pelos concorrentes.
32. PRAZO\_DISPUTA: número de dias entre

33. PRAZO\_ULTIMO\_LANCE: número de dias entre o primeiro e o último lance do fornecedor
34. DIA\_DISPUTA: dia do mês em que a disputa pelo item do pregão foi iniciada.
35. SEMANA\_DISPUTA : número da semana no ano em que a disputa pelo item do pregão foi iniciada.
36. DIA\_SEMANA\_DISPUTA: número do dia da semana em que a disputa pelo item do pregão foi iniciada.
37. REDUCAO\_DISPUTA: percentual de redução do preço do fornecedor vencedor em relação ao seu lance inicial.
38. REDUCAO\_NEGOCIACAO: percentual de redução do preço do fornecedor vencedor em relação ao seu último lance.
39. ALERTA: indicador da existência de alerta de indício de irregularidade emitido por ferramentas de verificação automática.
40. QTD\_LANCES\_VENCEDOR: número de lances realizados pelo fornecedor vencedor.
41. QTD\_CONCORRENTES: número de concorrentes na disputa pelo item.
42. QTD\_MEDIA\_LANCES: número médio de lances por concorrente no item.
43. CLASSE: variável alvo binária, assumindo os valores 0 e 1, conforme a anotação de ocorrência ou não de irregularidade.