

PREVISÃO DO SUCESSO DE PROCESSOS LICITATÓRIOS: UM ESTUDO DE CLASSIFICAÇÃO POR REGRESSÃO LOGÍSTICA

Discente: Gabriel Gonçalves Gomes
Professor: Dr. Arturo Alejandro Zavala Zavala
Programa de Pós-graduação em: Economia

Resumo: O presente estudo analisa o poder preditivo de um modelo de regressão logística aplicado à classificação do resultado de processos licitatórios para aquisição de açúcar cristal no período de 2023 a 2025. A pesquisa utiliza como base de dados o Portal Nacional de Contratações Públicas (PNCP), por meio de sua API de dados abertos, resultando em uma amostra final de 2.462 observações. As variáveis independentes selecionadas incluem o valor unitário estimado, a quantidade adquirida, além de variáveis categóricas representando o tipo de embalagem e a modalidade de contratação (pregão ou dispensa). O modelo de regressão logística, estimado com ajuste de pesos balanceados para tratar o desbalanceamento da amostra (2.150 homologados e 350 frustrados), alcançou acurácia global de 68% e área sob a curva ROC (AUC) de 0,76. Destaca-se a sensibilidade (*recall*) de 72% para a classe minoritária (processos frustrados), o que demonstra sua utilidade como ferramenta de alerta preliminar de risco. Entretanto, a baixa precisão para essa classe (26%) indica a necessidade de aprimoramento do modelo, mediante inclusão de novas variáveis e metodologias. Os resultados evidenciam que fatores simples, como valor unitário, quantidade adquirida e modalidade de pregão, já oferecem capacidade de sinalizar probabilidades de sucesso ou fracasso. Conclui-se que o modelo pode subsidiar o planejamento das compras públicas, contribuindo para o aprimoramento da eficiência na gestão das contratações e evitando custos com processos frustrados.

Abstract: This study analyzes the predictive power of a logistic regression model applied to the classification of bidding outcomes for crystal sugar procurement between 2023 and 2025. The research is based on data from the National Public Procurement Portal (PNCP) through its open data API, resulting in a final sample of 2,462 observations. The independent variables include the estimated unit price, the purchased quantity, and categorical variables representing the type of packaging and procurement modality (public bidding or direct purchase). The logistic regression model, estimated with balanced class weights to address sample imbalance (2,150 awarded and 350 failed bids), achieved an overall accuracy of 68% and an area under the ROC curve (AUC) of 0.76. The model's sensitivity (*recall*) for the minority class (failed processes) reached 72%, demonstrating its usefulness as a preliminary risk alert tool. However, the low precision for this class (26%) indicates the need for model improvement through the inclusion of new variables and methodologies. The results show that simple factors—such as unit price, purchased quantity, and bidding modality—already provide meaningful predictive capacity to signal probabilities of success or failure. It is concluded that the model can support public procurement planning, contributing to improved efficiency in contract management and the reduction of costs associated with failed processes.

Palavras-chave: Licitações. Ciência de Dados. Modelos de Classificação.

Keywords: Public Procurement. Data Science. Classification Models.

Classificação JEL: H57, C21, C55, D73.

1. INTRODUÇÃO

A importância do mercado de compras governamentais pode ser evidenciada por sua dimensão em relação ao PIB do país. Segundo o estudo de Garcia Ribeiro e Inácio (2019), no Brasil, o tamanho médio desse mercado na última década foi de 12,5% do Produto Interno Bruto (PIB), em linha com a média de 13,2% nos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Além disso, conforme apontado por Marinho (2024), as compras públicas representam cerca de 23% dos gastos governamentais totais no Brasil. Essa dimensão, tanto no âmbito nacional quanto no global, e a consequente relevância para a economia em geral e para o orçamento dos governos, evidenciam a necessidade de estudos que analisem suas características.

No Brasil, diversas normas regem o processo de compras públicas. A Constituição Federal determina que, em regra, elas devem ser precedidas de processo licitatório (BRASIL, 1988) e, por meio da Lei nº 14.133/21 (Lei de Licitações e Contratos Administrativos - NLLC), são estabelecidas as normas gerais que regem este procedimento. A modalidade mais utilizada e que propicia maior eficiência é o pregão eletrônico (Faria *et al.*, 2006). Em 2024, apenas por meio da plataforma Compras.gov.br foram realizados 98.317 processos nessa modalidade, que é obrigatória no caso de compras e serviços comuns (BRASIL, 2024). Além de reduzir custos e prazos para fornecedores e Administração, essa modalidade amplia a transparência das informações e a competitividade.

Nesse âmbito, entre os princípios explícitos da Lei de Licitações, está a economicidade, que, segundo Alexandrino e Paulo (2017), exige da Administração Pública a realização de suas atividades com o menor custo possível e a máxima efetividade, buscando a melhor relação custo-benefício. Diante disso e considerando a crescente disponibilidade de dados, resultante dos procedimentos eletrônicos, pesquisas quantitativas podem contribuir para o cumprimento dessa diretriz.

Sob essa ótica, nos últimos anos, a Administração Pública tem buscado fomentar a divulgação de dados à sociedade. Desde 2011, a Lei nº 12.527/2011 (Lei de Acesso à Informação – LAI) estabelece a obrigação de transparência ativa, pela qual devem ser divulgadas informações independentemente de solicitação, em formato eletrônico, aberto, estruturado e legível por máquina. O Decreto nº 8.777/2016 deu mais um passo nesse sentido, instituindo a política de dados abertos do Governo Federal. Finalmente, no contexto de compras públicas, essa orientação culminou na criação do Portal Nacional de Contratações Públicas (PNCP) pela NLLC, que tornou obrigatória a centralização e divulgação de todos os atos e informações de compras públicas realizadas no país nesse sítio eletrônico em formato de dados abertos. No final de 2023, foi disponibilizada a API (Interface de Programação de Aplicações), ferramenta que materializou a possibilidade de consulta e a utilização dessas informações em estudos voltados ao aprimoramento das políticas de compras públicas.

Essa realidade tem ampliado a aplicação de técnicas de ciência de dados e aprendizado de máquina (*machine learning*) em análises das grandes bases de dados disponibilizadas. A ciência de dados, entendida como um campo interdisciplinar que integra estatística, programação e análise de informação, busca extrair padrões e conhecimento útil a partir de grandes volumes de dados, enquanto o aprendizado de máquina fornece algoritmos que utilizam a experiência para melhorar o desempenho e realizar previsões mais precisas (HAN; KAMBER; PEI, 2012; MOHRI *et al.*, 2018).

Nos últimos anos, diversos estudos voltaram-se à análise dos determinantes da economicidade em compras públicas, usualmente tomando como variável dependente a diferença percentual entre o preço estimado e o valor final contratado. O presente estudo, embora inspirado nesse referencial, propõe uma abordagem distinta: avaliar o poder preditivo de um modelo de classificação, baseado em regressão logística, para estimar a probabilidade de

sucesso ou fracasso em processos licitatórios. A investigação utiliza como base empírica as aquisições de açúcar cristal realizadas entre 2023 e 2025, permitindo não apenas testar a acurácia e as métricas de desempenho do modelo, mas também discutir suas implicações práticas para o planejamento e a gestão das compras governamentais.

2. REFERENCIAL TEÓRICO

A primeira etapa do trabalho consistiu na revisão de estudos econométricos sobre compras públicas. A maioria das pesquisas utilizavam metodologias de regressão, e, embora o presente trabalho trate de uma análise de classificação, a revisão de estudos anteriores possibilitou compreender as variáveis de interesse, os referenciais teóricos aplicados e as principais conclusões obtidas. Com esse intuito, foram analisadas três pesquisas cujo objetivo era identificar as variáveis que influenciam a economia das compras públicas, utilizando a regressão linear múltipla.

A pesquisa de Faria et al. (2010) testou a relação da economicidade com nove variáveis independentes, a partir de modelos de regressão linear múltipla apoiados na Teoria dos Leilões e na Economia Institucional. Com base em dados do SIASG, ComprasNet e questionários aplicados a fornecedores, foram analisados 195 pregões eletrônicos de 2008. O estudo identificou que o número de fornecedores, a especificidade dos ativos, a quantidade adquirida e a frequência das transações explicavam 65,7% da variação da economicidade, reforçando a importância de estratégias voltadas à ampliação da concorrência, planejamento de compras conjuntas e fortalecimento da relação com fornecedores.

No mesmo sentido, Melo e Diniz Filho (2018) investigaram pregões eletrônicos realizados em um órgão público do Maranhão no exercício de 2017. Foram estimados modelos de regressão linear múltipla com erros-padrão robustos para testar o impacto do número de fornecedores, número de lances e quantidade adquirida sobre os preços finais. Embora o modelo tenha apresentado poder explicativo de 48% (R^2 ajustado), apenas a variável número de fornecedores foi estatisticamente significativa, com efeito negativo sobre os preços. Os autores concluíram que a participação mais ampla de licitantes constitui fator determinante para reduzir custos, indicando que medidas de divulgação e redução de barreiras de entrada devem ser priorizadas pela Administração Pública.

Por fim, Ferreira (2023) analisou 79 pregões eletrônicos da Universidade Federal de Pernambuco realizados entre 2019 e 2022, resultando em uma amostra de 2.082 itens. Utilizando regressão linear múltipla, a autora investigou variáveis como número de licitantes, número de lances, quantidade, forma de divulgação do edital, tipo de objeto e tipo de contratação (SRP ou tradicional), tendo como variável dependente a economia percentual obtida. A análise mostrou economia média de 26,3% e confirmou a relevância de práticas como planejamento de compras conjuntas, capacitação dos pregoeiros em técnicas de negociação e realização de pesquisas de preços de qualidade. As conclusões destacam o potencial da gestão ativa das variáveis institucionais para aumentar a competitividade e reduzir gastos públicos, além de sugerirem a replicação do modelo em diferentes órgãos e contextos.

Em síntese, o Quadro 1 apresenta as variáveis explicativas avaliadas nos trabalhos analisados e o resultado obtido, relacionado à influência sobre a variável explicada (direta ou inversa).

Quadro 1 – Resumo de variáveis analisadas em estudos anteriores

Estudo	Variáveis (tipo de influência)	Método
Faria <i>et al.</i> (2010)	1) Número de licitantes (direta); 2) Especificidade dos ativos (inversa); 3) Quantidade comprada (direta);	Regressão linear múltipla

	4) Frequência das transações (direta).	
Melo e Diniz Filho (2018)	1) Número de licitantes (direta); 2) Número de lances dados na disputa; 3) Quantidade comprada.	Regressão linear múltipla
Ferreira (2023)	1) Número de licitantes (direta); 2) Número de lances dados na disputa (direta); 3) Quantidade comprada (direta); 4) Forma de divulgação do edital (inversa); 5) Tipo de objeto (inversa); 6) Forma de contratação (direta).	Regressão linear múltipla

Fonte: adaptado de Ferreira (2023).

3. METODOLOGIA

3.1 Escolha e Coleta dos Dados

Após a revisão teórica de estudos sobre licitações públicas e análise de variáveis relevantes, foram coletados dados de processos de aquisição de açúcar realizados entre 2023 e 2025. A base utilizada foi o Portal Nacional de Contratações Públicas (PNCP), por meio de sua API de dados abertos, complementada com registros de pesquisa de preços disponíveis no portal. O produto foi escolhido por sua elevada representatividade nas compras governamentais e pela padronização recente no Catálogo Eletrônico, que garante uniformidade de especificações. A partir das bases consultadas, obteve-se inicialmente um conjunto de 3.143 itens, o qual, após tratamento, limpeza e exclusão de inconsistências, resultou em uma amostra final de 2.462 observações.

3.2 Definição das Variáveis e Pré-processamento

A variável dependente do estudo foi a Situação do processo licitatório, categorizada de forma binária em homologado (1) ou frustrado (0). As variáveis independentes selecionadas incluíram o preço unitário estimado (R\$/kg), a quantidade adquirida (transformada em logaritmo natural para estabilizar a variância e reduzir a heterogeneidade da escala), além de variáveis *dummies* que representaram diferentes modalidades de contratação (ex.: pregão ou dispensa) e tipos de embalagem (0,5 kg, 2 kg, 5 kg e 30 kg). Para reduzir distorções, os dados numéricos foram padronizados com `StandardScaler`, e para lidar com o desbalanceamento da amostra (aprox. 2.150 homologados e 350 frustrados), foi aplicada a técnica de pesos balanceados na regressão logística, ajustando a importância relativa da classe minoritária.

3.3 Regressão Logística e Avaliação do Modelo

Com a base estruturada, aplicou-se a Regressão Logística binária como modelo de classificação, escolhida pela sua robustez, interpretabilidade e uso consagrado em estudos de análise preditiva (HOSMER; LEMESHOW, 2013). O desempenho do modelo foi avaliado a partir de métricas de classificação, incluindo acurácia, precisão, *recall* e *F1-score*, com ênfase no equilíbrio entre classes. Também foram estimadas curvas ROC e *Precision-Recall*, bem como as respectivas áreas sob a curva (AUC), a fim de medir a capacidade discriminatória do modelo. A implementação foi realizada em Python, utilizando as bibliotecas `scikit-learn`, `pandas`, `numpy` e `matplotlib`.

A regressão logística foi escolhida por sua adequação a problemas de classificação binária, oferecendo resultados facilmente interpretáveis e métricas consolidadas de avaliação.

Diferentemente de algoritmos mais complexos, como *Random Forest* ou *Support Vector Machines*, a regressão logística permite compreender diretamente o efeito marginal de cada variável explicativa sobre a probabilidade de ocorrência do evento de interesse, o que é particularmente relevante em estudos aplicados à administração pública, em que a transparência e a clareza da interpretação dos resultados são fundamentais (HOSMER; LEMESHOW, 2013). Além disso, o método é robusto mesmo em bases de tamanho moderado e com certo desbalanceamento entre classes, desde que tratadas adequadamente com técnicas de ponderação.

4. RESULTADOS E DISCUSSÕES

4.1 Tratamento dos Dados

O pré-processamento incluiu a transformação da variável quantidade pelo logaritmo natural, de forma a reduzir assimetrias e tornar a distribuição mais adequada para o modelo, além da padronização das variáveis contínuas (valor unitário e quantidade). As variáveis categóricas relacionadas ao tipo de embalagem e à modalidade de contratação foram representadas por meio de variáveis *dummy*. Após esse tratamento, a amostra final foi composta por 2.462 observações, das quais 2.150 correspondem a processos homologados e 350 a processos frustrados, evidenciando o desbalanceamento natural da variável resposta.

Quadro 2 – Variáveis utilizadas no modelo de classificação

Variável	Descrição	Tipo
Situação	Variável dependente (0=homologado, 1=frustrado)	Binária
valor	Valor unitário estimado (R\$/kg)	Contínua
ln(quantidade)	Logaritmo natural da quantidade adquirida (kg)	Contínua
<i>Dummies</i> de embalagem	Pacotes de 0,5 kg, 2 kg, 5 kg, 30 kg	<i>Dummy</i>
D - pregão	Modalidade pregão eletrônico (1=sim; 0=dispensa)	<i>Dummy</i>

Fonte: elaboração própria.

4.2 Ajuste do Modelo de Regressão Logística

O modelo de regressão logística foi estimado com ajuste de pesos balanceados, de modo a compensar o desbalanceamento entre classes. A Tabela 1 apresenta o resumo das métricas de desempenho na base de teste (20% da amostra). Observa-se que a acurácia global do modelo foi de 68%, com *recall* de 72% para a classe minoritária (frustrado), indicando que o modelo conseguiu capturar boa parte dos casos relevantes, ainda que à custa de uma menor precisão (26%). Para a classe majoritária (homologado), os indicadores apresentaram precisão elevada (94%), mas *recall* moderado (67%).

Quadro 3 – Desempenho do modelo de regressão logística

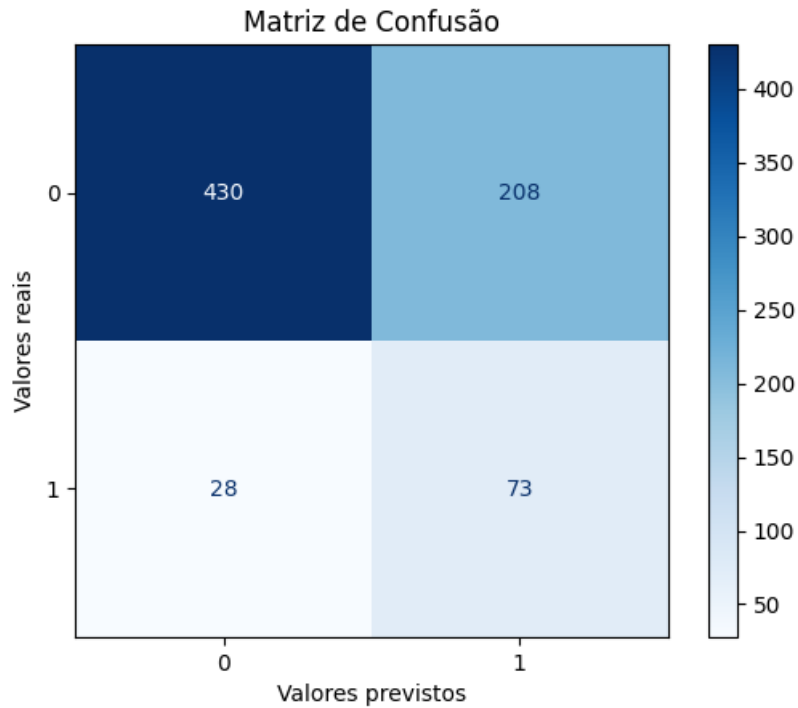
Classe	Precisão	<i>Recall</i>	F1-score	Suporte
Homologado	0,94	0,67	0,78	638
Frustrado	0,26	0,72	0,38	101
Acurácia			0,68	739

Fonte: resultados da pesquisa.

4.3 Curva ROC e Matriz de Confusão

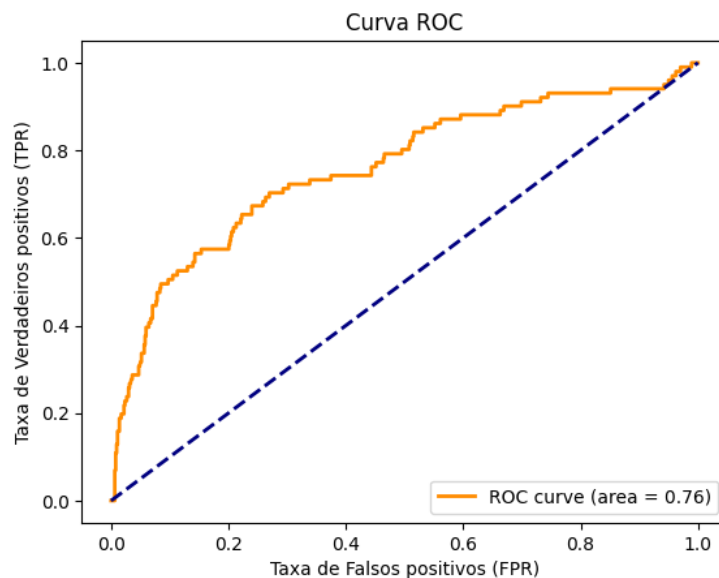
A Figura 1 apresenta a matriz de confusão, que evidencia que 73 casos frustrados foram corretamente classificados, contra 28 incorretos. Por outro lado, 430 casos homologados foram corretamente previstos, mas houve 208 falsos positivos. A Figura 2 mostra a curva ROC, com $AUC=0,76$, indicando desempenho satisfatório do modelo na discriminação entre processos homologados e frustrados. Ainda que não atinja patamares considerados excelentes ($AUC > 0,85$), o resultado demonstra que a regressão logística é capaz de capturar padrões significativos nos dados.

Figura 1 – Matriz de Confusão



Fonte: resultados da pesquisa.

Figura 2 – Curva ROC com AUC



Fonte: resultados da pesquisa.

4.4 Implicações dos Resultados

Os resultados sugerem que a classificação do sucesso de processos licitatórios apresenta relação com variáveis simples, como o valor unitário, a quantidade adquirida e a modalidade de pregão. O bom recall obtido para a classe frustrada indica que o modelo pode ser útil como ferramenta de alerta para gestores, sinalizando processos com maior probabilidade de insucesso. Contudo, a baixa precisão nessa classe evidencia que parte dos alertas seriam falsos positivos, o que limita sua utilização direta como instrumento decisório. Assim, a principal contribuição prática do modelo é servir como sinalizador preliminar de risco, devendo ser complementado por análises qualitativas e outras variáveis contextuais não incluídas no presente estudo.

5. CONCLUSÃO

Considerando os resultados obtidos, conclui-se que o modelo de regressão logística apresentou desempenho satisfatório para a classificação de processos licitatórios de aquisição de açúcar cristal entre homologados e frustrados. Apesar das limitações inerentes ao desbalanceamento da base de dados, o modelo alcançou acurácia de 68% e AUC de 0,74, com destaque para o bom recall da classe minoritária (frustrados), o que indica utilidade prática como ferramenta de alerta preliminar. Esses resultados sugerem que variáveis simples, como valor unitário, quantidade adquirida e modalidade de pregão, possuem capacidade de sinalizar riscos de insucesso, ainda que a precisão para a classe minoritária permaneça limitada.

Além disso, observa-se que o modelo pode servir de subsídio ao planejamento das contratações, permitindo que gestores identifiquem previamente processos mais propensos ao fracasso e adotem medidas preventivas. No entanto, o baixo poder discriminatório em termos de precisão reforça a necessidade de aprimoramento do modelo, com a inclusão de variáveis adicionais relacionadas à competição, prazos e exigências editalícias. Recomenda-se, portanto, que futuras pesquisas explorem a incorporação de novas fontes de dados e diferentes algoritmos de classificação, de modo a aumentar a confiabilidade dos resultados e oferecer suporte mais robusto às decisões estratégicas na gestão das compras públicas.

Referências Bibliográficas:

ALEXANDRINO, M.; PAULO, V. **Direito administrativo descomplicado**. 26. ed. São Paulo: Método, 2017.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

BRASIL. **Lei nº 14.133, de 1º de abril de 2021**. Lei de Licitações e Contratos Administrativos. Brasília, DF: Presidência da República, 2021. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14133.htm.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Lei de Acesso à Informação. Brasília, DF: Presidência da República, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/L12527.htm.

BRASIL. **Decreto nº 8.777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo Federal. Brasília, DF: Presidência da República, 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/D8777.htm.

BRASIL. **Compras.gov.br**. Estatísticas de pregões eletrônicos. Brasília, DF: Ministério da Gestão e da Inovação em Serviços Públicos, 2024. Disponível em: <https://www.compras.gov.br>.

FARIA, A. F. et al. **Pregão eletrônico: estudo sobre sua economicidade em compras públicas**. Revista de Administração Pública, Rio de Janeiro, v. 40, n. 6, p. 1065-1088, nov./dez. 2006.

FARIA, A. F. **Determinantes da economicidade em licitações: uma análise empírica do pregão eletrônico**. 2010. Tese (Doutorado em Administração) – Universidade Federal de Viçosa, Viçosa, 2010.

FERREIRA, L. M. **Fatores determinantes da economicidade em pregões eletrônicos: um estudo na Universidade Federal de Pernambuco**. Dissertação (Mestrado em Administração) – Universidade Federal de Pernambuco, Recife, 2023.

GARCIA RIBEIRO, C.; INÁCIO, J. C. **Mercado de compras públicas: panorama e desafios**. Revista de Administração Pública, Rio de Janeiro, v. 53, n. 4, p. 803-824, 2019.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Amsterdam: Elsevier, 2012.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. 3. ed. New York: Wiley, 2013.

MARINHO, A. **Panorama recente das compras públicas no Brasil**. Revista do Serviço Público, Brasília, v. 75, n. 1, p. 45-62, 2024.

MELO, E. R. S.; DINIZ FILHO, J. W. F. **Elementos impactantes na variação dos preços em pregões eletrônicos**. Revista de Administração, Gestão e Contabilidade, v. 6, n. 25, p. 162-178, 2018.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. 2. ed. Cambridge: MIT Press, 2018.