ESTIMATIVA DE CUSTO DE OBRAS PÚBLICAS DE PRESÍDIOS: USO DE TÉCNICAS DE MACHINE LEARNING

Cost estimation of public prison buildings: using machine learning techniques

Andrei Benício Barbosa Silva
Universidade Federal do Ceará | Fortaleza, Ceará | andrei-benicio@hotmail.com

Camily Vasconcelos Barbosa
Universidade Federal do Ceará | Fortaleza, Ceará | vasconcelosmily@alu.ufc.br

Marcelo Silva Santos Segundo
Universidade Federal do Ceará | Fortaleza, Ceará | marcelosegundo@alu.ufc.br

José de Paula Barros Neto
Universidade Federal do Ceará | Fortaleza, Ceará | barrosneto@gercon.ufc.br

RESUMO

Estimar custos de construção iniciais para diferentes tipologias de obras é uma tarefa desafiadora, especialmente no caso de presídios públicos, que possuem características únicas e complexidades não abrangidas por métodos tradicionais de estimativa. Este estudo propõe o uso de técnicas de *machine learning* para a estimativa parametrizada de custos de obras públicas de presídios. Foram desenvolvidos quatro modelos de *machine learning* com o objetivo de verificar qual o modelo mais adequado para alcançar estimativas mais precisas para essa tipologia. Entre os modelos avaliados, o de árvores de decisão apresentou o melhor desempenho, com erro percentual médio absoluto (MAPE) de 16,47%, seguido pelo modelo de *support vector machine*, que obteve MAPE de 21,19%. Os outros modelos apresentaram erros mais significativos e foram considerados inadequados para estimar custos. Ao comparar os resultados com a regressão linear simples, observou-se a superioridade dos modelos desenvolvidos em termos de precisão. Os achados indicam que o uso de *machine learning* é uma ferramenta eficaz para lidar com as complexidades da estimativa de custos de presídios públicos, proporcionando maior exatidão do que os métodos tradicionais. Esses resultados ressaltam o potencial dessas técnicas para otimizar o planejamento de obras dessa natureza, tornando-as uma alternativa promissora para superar desafios existentes.

Palavras-chave: Orçamento paramétrico; Presídios públicos; Machine learning.

ABSTRACT

Estimating initial construction costs for different types of projects is a challenging task, especially in the case of public prisons, which have unique characteristics and complexities not addressed by traditional estimation methods. This study proposes the use of machine learning techniques for the parametric cost estimation of public prison construction projects. Four machine learning models were developed to determine which model is most suitable for achieving more accurate estimates for this type of project. Among the evaluated models, the decision tree model showed the best performance, with a mean absolute percentage error (MAPE) of 16.47%, followed by the support vector machine model, which achieved a MAPE of 21.19%. The other models exhibited higher error rates and were deemed unsuitable for cost estimation. When comparing the results with simple linear regression, the developed models demonstrated superior accuracy. The findings indicate that machine learning is an effective tool for addressing the complexities of cost estimation for public prisons, offering greater precision than traditional methods. These results highlight the potential of these techniques to optimize the planning of such projects, making them a promising alternative to overcome existing challenges.

Keywords: Parametric Budgeting; Public Prisons; Machine Learning.

1 INTRODUÇÃO

A determinação precisa dos custos desempenha um papel fundamental na construção civil, sendo uma ferramenta essencial para o gerenciamento de obras (Otero, 2000); portanto, é um elemento crucial em qualquer projeto de construção, seja ele de natureza pública ou privada. Segundo Barros e Falcão (2016), é evidente a importância da acurácia e da precisão em orçamentos paramétricos para o gerenciamento e acompanhamento dos projetos. No entanto, essa tarefa se revela particularmente desafiadora nas fases iniciais do processo de *design*, quando as informações disponíveis são escassas e imprecisas (An; Kang, 2005). Yun (2022) descreve que os custos de construção por unidade de área construída são frequentemente utilizados como o critério mais comum para prever os custos de construção com base na análise de regressão. Para Yang *et al*. (2022), ao estimar o custo de construção por unidade de área, as abordagens existentes não consideram outros fatores além de informações relacionadas à área, causando erros de estimativa. No cenário brasileiro, dois indicadores frequentemente utilizados para estimar custos em projetos de edificação são o Custo Unitário Básico de Construção (CUB) e o Custo Médio por Metro Quadrado na Construção Civil. Contudo, esses indicadores demonstram limitações ao abordar a formação de custos em tipologias construtivas altamente específicas, como é o caso das edificações penais, uma vez que os empreendimentos da construção possuem características únicas que devem ser consideradas durante as atividades de estimativa de custos e planejamento.

De acordo com Elmousalami (2019), as maiores dificuldades para elaboração de uma boa estimativa de custo em etapas iniciais é a falta de informações preliminares disponíveis e a incerteza. Para Elfaki, Alatawi e Abushandi (2014), elaborar uma estimativa é uma tarefa que depende da expertise do profissional, porém, essa expertise não pode ser padronizada ou replicada, ficando dependente da experiência individual de cada profissional. A ausência de critérios padronizados para a elaboração de uma estimativa abre caminho para que autores trabalhem com diferentes abordagens na realização de estimativas. Elfaki, Alatawi e Abushandi (2014) ainda reforçam que os engenheiros necessitam de vários anos para desenvolver os conhecimentos necessários para conduzir o processo de estimativa de custos, conhecimentos esses que não podem ser documentados ou autenticados.

Uma das formas de mitigar os problemas apresentados no âmbito da elaboração de estimativas de custos em etapas iniciais de projetos, consiste na aplicação de modelos de aprendizado de máquina. Segundo Yun (2022), com o desenvolvimento da tecnologia de Machine Learning, tornou-se possível prever facilmente os custos de construção considerando vários fatores de influência dependendo do nível de aquisição de dados. A aplicação de técnicas de Machine Learning pode identificar tendências e padrões em dados que os humanos e os métodos tradicionais de previsão não conseguem detectar, permitindo previsões e tomadas de decisão mais precisas (Dataflair, 2022). Segundo Kumar, Thakur e Gupta (2021), o *machine learning* permite que sistemas aprendam com dados e tomem decisões com mínima intervenção humana. É fundamental desenvolver referências de custos sólidas para embasar estimativas iniciais do projeto, melhorando a gestão de custos, agilizando projeções orçamentárias e reduzindo riscos de inviabilidade por falhas no planejamento (Tanim; Ahmad, 2025).

Diante desse cenário, este estudo propõe a aplicação de técnicas de *machine learning* para desenvolver modelos de estimativa de custo voltadas para obras públicas de presídios. Objetiva-se com o uso de técnicas de ML, elaborar um processo de estimativa padronizado, claro e que auxilie o processo de tomada de decisão no desenvolvimento de projetos em fases iniciais.

São propostos quatro modelos: regressão Lasso, redes neurais artificiais, árvores de decisão e *support vector machine*, avaliados por visualizações e métricas. Os resultados deste estudo têm o potencial de impactar positivamente a gestão de custos em obras públicas de presídios, facilitando a elaboração de estimativas de custo em etapas iniciais de projetos e, consequentemente, contribuindo para a elucidação de tomadas de decisões mais precisas.

2 REFERENCIAL TEÓRICO

Enquadradas na área de conhecimento da gestão de custos, as estimativas fornecem informações essenciais para a tomada de decisão no desenvolvimento de projetos (PMI, 2016). O principal objetivo da prática de estimar custos, desde o seu estabelecimento dentro da disciplina de levantamento quantitativo na década iniciada em 1950, tem sido fornecer uma base para controlar os custos do projeto com a elaboração de estimativas de custos (Ashworth; Perera, 2015).

Miranda *et al*. (2022) identificam três abordagens para avaliação de técnicas de estimativa de custo: (i) uso de uma única técnica para desenvolver o modelo, de forma restritiva; (ii) busca de alternativas para aprimorar uma técnica específica, visando sua melhor performance; e (iii) comparação entre diferentes técnicas, sem foco exclusivo na otimização individual, para identificar a mais adequada ao objetivo da predição. Além das abordagens existentes, deve-se avaliar quais modelos podem ser utilizados em estimativas de custos e suas formas de aplicação.

Jin *et al*. (2012) desenvolveram um modelo de previsão de custos baseado em Raciocínio Baseado em Casos (RBC), aprimorado com Análise de Regressão Múltipla (ARM) na fase de revisão. O modelo melhorou a previsão de custos em 17,23% para instalações comerciais e 4,39% para habitações multifamiliares em relação ao RBC tradicional.

Al-Tawal, Arafah e Sweis (2020) avaliaram a eficácia de Redes Neurais Artificiais (RNA) na estimativa de custos nas fases iniciais do projeto de edifícios. Desenvolveram três modelos, um para cada fase do projeto (conceitual, esquemático e detalhado), alcançando precisões médias de 97%, 98% e 98%, respectivamente.

Creese e Li (1995) comprovaram a superioridade das redes neurais artificiais (RNAs) sobre a regressão linear na estimativa de custos de pontes de madeira. Lee *et al*. (2022) também destacam o uso de RNA para estimativas com dados limitados, dada a maior precisão e estabilidade.

Omer (2022) salienta que a Análise de Regressão Múltipla não é um modelo único de aplicação, e desenvolve duas formas de implementação, a regressão linear Ridge e regressão linear Lasso. O objetivo desse modelo de regressão é representar matematicamente os dados com o mínimo de erro nas previsões.

Para Hashemi, Ebadati e Kaur (2020) árvores de regressão são uma boa substituição para métodos básicos de regressão, no entanto é o método menos adotado para prever custos. O modelo foi utilizado por Cho e Chun (2015) para estimar a quantidade de barras de armadura a fim de prever o custo de estruturas por meio do levantamento de quantidades.

Son, Kim e Kim (2012) compararam o método AD com o modelo de *support vector machine*, obtendo SVM como melhor preditor. SVM é uma técnica de aprendizado estatístico supervisionado não paramétrico que foi introduzida por Vapnik (1979). A principal vantagem do SVM é a capacidade de autoaprendizagem e alto desempenho na generalização (SHIN *et al*., 2009). Alguns estudos concluem a superioridade desse método em comparação aos demais (Son; Kim; Kim, 2012; Petruseva *et al*., 2017; Jaafari; Pazhouhan; Bettinger, 2021).

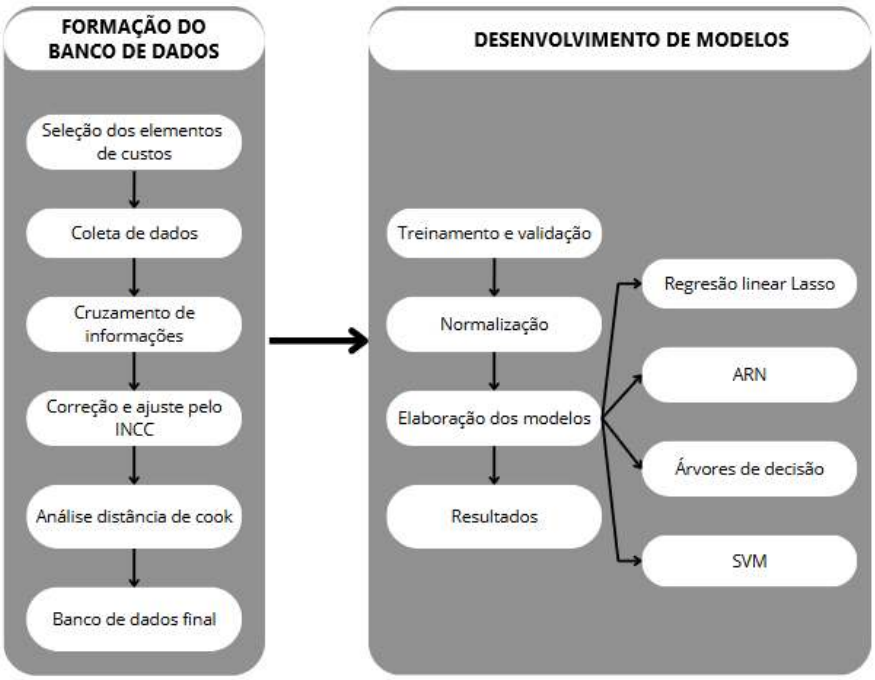
As técnicas de regressão se mostram como uma ferramenta estatística poderosa, capaz de prever custos de diferentes construções (Ratner, 2010). Dessa forma, Elmousalami (2019) observa que embora várias técnicas tenham seus prós e contras, ainda são escassos os estudos que buscam identificar, de forma comparativa, a técnica com melhor desempenho preditivo em estimativas de custo.

3 METODOLOGIA

Segundo Elmousalami (2021), a modelagem preditiva consiste em dois estágios principais: engenharia de recursos e desenvolvimento do modelo. A engenharia de recursos é a seleção dos principais direcionadores de custo como entrada para calcular o custo do projeto, enquanto o desenvolvimento do modelo é aplicar técnicas de aprendizado estatístico para entender o padrão entre os principais direcionadores de custo e o custo final do projeto.

Com essas considerações, a metodologia proposta para estimar os custos de construção de presídios consistiu em duas etapas principais: (1) Formação do banco de dados e (2) Desenvolvimento de modelos (Figura 1).

Figura 1: Fluxo de dados e processos de desenvolvimento de pesquisa.



Fonte: Arquivo pessoal.

3.1 FORMAÇÃO DO BANCO DE DADOS

A escolha das variáveis independentes (elementos de custo) utilizadas neste trabalho é baseada nas variáveis finais escolhidas por Beltrão *et al.* (2022). Os autores utilizaram-se do indicador de Fator de Influência de Variável (FIV) para seleção das variáveis independentes mais relevantes e aplicaram a um questionário a profissionais com experiência na formação de custos da tipologia de obra modelada. Ao final, foram selecionadas cinco variáveis de custo que são consideradas importantes por afetarem o custo de construção de presídios de forma independente, linear e com pouca ou nenhuma correlação entre si (Quadro 1).

Quadro 1: Relação de variáveis Independentes

VARIÁVEL	MÉTRICA
Área construída	Metro quadrado
Quantidade de vagas	Unidade
Área total do terreno	Metro quadrado
Número de celas	Unidade
BDI	Porcentagem

Fonte: Adaptado de Beltrão *et al*. (2022).

O banco de dados de Beltrão *et al*. (2022) preserva a privacidade das cadeias, presídios e valores orçamentários. Para suprir essas lacunas, este estudo cruzou informações sobre localização, dimensões e características dos presídios com orçamentos públicos divulgados por sites oficiais de governos estaduais. O valor de custo de construção referente a um dos presídios não foi encontrado, sendo descartado do banco de dados. Com objetivo de manter a confidencialidade, os dados referentes aos preços também não puderam ser divulgados no corpo deste trabalho, assim como os nomes das cadeias e penitenciárias.

Uma vez realizada a coleta, os preços foram reajustados para um mesmo mês-base por meio do Índice Nacional de Custo da Construção (INCC-DI), visando normalizar os efeitos decorrentes da inflação. Para tanto, utilizou-se a Equação 1:

$$P_{reaj} = P_{orça} \times [1 + (\frac{I1-I0}{I0})]$$

Onde:

Preaj: preço reajustado, em reais;

Porça: preço do orçamento, em reais;

I1: índice publicado para o mês-base de reajuste (dezembro de 2024); e

I0: índice publicado para o mês-base do orçamento.

Para o reajuste e normalização dos preços, adotou-se o mês-base de dezembro de 2024, que é a referência mais recente ao desenvolvimento da modelagem.

Os dados do banco apresentam características próprias que precisam ser avaliadas quanto à coesão. Inicialmente, verifica-se a presença de dados incorretos, repetidos, incompletos ou ausentes. Após verificar as informações, deve-se identificar dados discrepantes que podem prejudicar a análise (outliers). Utiliza-se a distância de Cook para medir o impacto de um caso no modelo de regressão (STEVENS, 2002). Os valores inferiores a 1 não exigem exclusão, enquanto valores superiores a 1 indicam a necessidade de exclusão.

O banco de dados final ficou composto conforme a Tabela 1.

Tabela 1: Banco de dados

ESTADO	CONSTRUÇÃO	ÁREA CONSTRUÍDA (M²)	Nº DE VAGAS	ÁREA TOTAL (M²)	Nº DE CELAS	BDI (%)	MÊS-BASE (SINAPI)
AP	1	8254,62	422	37899,67	64	25,22	08/19
AM	2	6892,05	286	15780,85	52	28,35	07/17
CE	3	6763,62	168	20357,12	104	24,98	12/17
GO	4	6982,05	388	15780,85	66	26,35	06/14
MG	5	6982,05	388	15780,85	66	25,00	10/18
PA	6	2678,96	306	10038,95	48	25,22	11/17
PB	7	13964,10	748	31561,70	134	28,63	01/18
RS	8	6982,05	388	15780,85	66	28,75	05/19
PI	9	6359,94	336	26096,65	105	26,43	04/19
RN	10	6942,92	420	62663,00	69	30,00	08/18
SC	11	10862,91	364	22511,24	79	25,00	09/18
SE	12	15902,85	632	43005,28	194	24,99	06/17
SP	13	11132,27	768	90000,00	64	27,55	07/14

Fonte: Adaptado de Beltrão *et al*. (2022).

3.2 DESENVOLVIMENTO DE MODELOS

Os modelos serão desenvolvidos de acordo com os seguintes componentes: separação de treino e teste, normalização da elaboração dos modelos e avaliação dos resultados.

Para treinamento e validação do modelo, os dados coletados nas medições de campo são divididos aleatoriamente em dois conjuntos. De 13 amostras, 70% foram usadas como conjunto de dados de treinamento e os 30 % restantes foram reservados para teste de modelo. Segundo Géron (2019), a divisão dos dados em conjuntos de treinamento e teste, comumente na proporção de 70/30, é uma prática habitual que visa garantir uma avaliação justa do desempenho do modelo.

Na presença de dados com diferentes intervalos das variáveis, recomenda-se que os conjuntos de dados sejam normalizados. Para evitar que variáveis com diferentes escalas dominem a análise, recomenda-se a normalização dos dados (Han, Pei e Kamber, 2011). Assim, valores discrepantes podem ser trabalhados sem distorções. Neste estudo, os dados foram normalizados entre 0 e 1 usando a normalização Standard do Python.

Os modelos escolhidos para desenvolvimento da pesquisa foram regressão Lasso (RL), redes neurais artificiais (RNA), árvores de decisão (AD) e *support vector machine* (SVM). Os modelos foram desenvolvidos usando a linguagem Python por meio da plataforma *Jupyter Notebook*.

O desempenho dos modelos foi avaliado usando o MAPE, que mede a magnitude média do erro. O MAPE deve ser o mais próximo possível de zero (ou seja, sem erro entre os custos reais e previstos) para indicar um excelente desempenho do modelo. Segundo Esteves *et al.* (2018), o MAPE avalia a precisão de modelos preditivos, pois sua interpretação percentual facilita a tomada de decisões.

4 RESULTADOS E DISCUSSÕES

Os resultados encontrados pela aplicação da metodologia ao banco de dados são apresentados na Tabela 2.

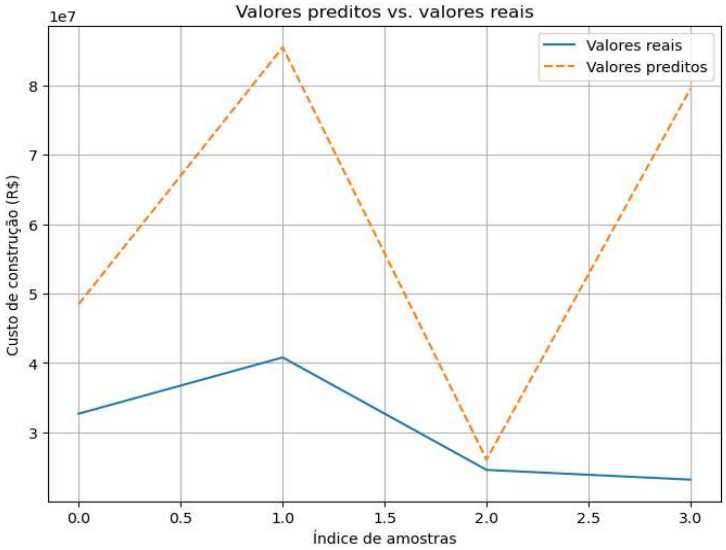
Tabela 2: Resultado de valores MAPE dos modelos

MODELO	MAPE
Regressão Lasso	97,53%
RNA	172,35%
Árvores de decisão	16,47%
SVM	21,19%

Fonte: Arquivo pessoal.

Os valores de MAPE quantificam a média percentual das diferenças absolutas entre as previsões do modelo e os valores reais. Valores menores de MAPE indicam que o modelo está fazendo previsões mais precisas. Dessa forma, percebe-se que os modelos regressão Lasso e redes neurais artificiais demonstram não se adequar apropriadamente aos dados, com valores de 97,53% e 172,35%, respectivamente. A baixa capacidade de predição do modelo de Redes Neurais Artificiais tende a ocorrer pela limitação na quantidade de dados trabalhados nesta pesquisa. Assim, escassa disponibilidade de informações de dados afeta a capacidade dos modelos de desenvolver previsões mais acuradas e como consequência, ocorrer *overfitting* (DELGADO, OYEDELE, 2021). O modelo de regressão Lasso tende a os valores dos dados encolherem para o centro ou média, a fim de evitar o sobreajuste. Dessa forma, é esperado alto valor MAPE encontrado para o modelo em virtude da baixa quantidade de dados utilizados para treino e teste.

Gráfico 1: Relação de valores previstos e reais para o modelo de Regressão Lasso.

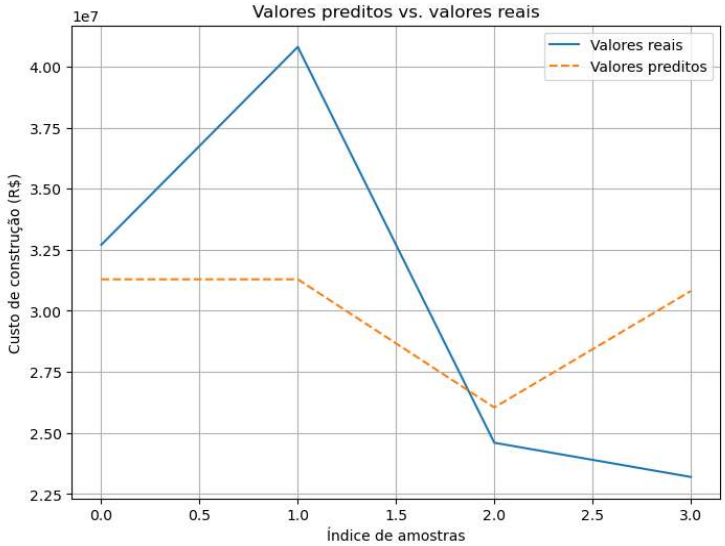


Fonte: Arquivo pessoal.

No Gráfico 1 pode-se perceber a inexistência do acompanhamento dos valores reais com os valores previstos. Isso demonstra que o modelo não consegue desempenhar um papel salutar na predição de custos para esses dados. Segundo Lewis (1982), o MAPE classificado como fornecedor de previsão imprecisa se o valor de MAPE for maior que 50%. Os valores obtidos pelos modelos Lasso e RNA os qualificam como previsões imprecisas, que para este estudo serão descartados seu caráter de validação para estimativa de custos.

A análise dos valores de MAPE dos modelos AD e SVM mostraram-se eficazes quanto a sua adequação aos dados trabalhados. À luz de Lewis (1982), o MAPE classificado como fornecedor de boa previsão se o valor de MAPE estiver entre 10% e 20% e entre 20% e 50% MAPE gera previsões aceitáveis. Os valores de 16,47% e 21,19%, respectivamente, demonstram que os modelos podem carregar consigo maior poder de predição com boa confiabilidade. O valor obtido pelo modelo AD, de boa previsão, se enquadra como sendo o modelo que melhor se adequa aos dados trabalhados. AD é um modelo de aprendizado que pode produzir desempenho preciso sem problemas de *overfitting* (BREIMAN *et al.*, 1984). Isso pode ser visualizado no Gráfico 2, quando comparado ao Gráfico 2, apresenta uma disposição de variação dos valores reais mais próxima dos valores preditos.

Gráfico 2: Relação de valores previstos e reais para o modelo de Arvores de decisão.



Fonte: Arquivo pessoal.

Destaca-se que os resultados obtidos pelo uso dos modelos Árvores de Decisão e SVM, apresentam boa confiabilidade quando comparados ao estudo de onde foram coletadas a base de dados e os parâmetros escolhidos de Beltrão *et al.* (2022). O estudo de Beltrão *et al.* (2022) obteve resultados para MAPE, respectivamente de 23% e 25%, nos dois modelos válidos com performance adequada. Ambos resultados se enquadram como geradores de previsões adequadas, porém, destaca-se que cada modelo desenvolvido pelos autores mencionados trabalha apenas com apenas uma variável independente (regressão linear simples). Os modelos desenvolvidos nesta pesquisa se diferenciam, pois utilizam-se de todas as variáveis independentes selecionadas para determinação do valor predito, apresentando maior representatividade dos dados gerais coletados para geração de estimativas.

5 CONSIDERAÇÕES FINAIS

Este estudo investigou a viabilidade e a eficácia das técnicas de *machine learning* na criação de um referencial de custos específico para estimativas em obras públicas de presídios. Para isso, foram desenvolvidos e comparados quatro modelos de *machine learning*: regressão Lasso, redes neurais artificiais, árvores de decisão e *support vector machine*. Para determinação das principais variáveis independentes que impactam o custo das obras e formação do banco de dados, utilizou-se como referência o trabalho desenvolvido por Beltrão *et al.* (2022). Os modelos aqui desenvolvidos foram construídos utilizando linguagem de programação Python, treinados e validados de forma aleatória com base em dados de 13 obras de construção de presídios no Brasil.

Os resultados indicaram que o modelo de árvores de decisão apresentou o melhor desempenho na estimativa de custos de construção de presídios, com um MAPE de 16,47%. O modelo de SVM também apresentou um MAPE satisfatório de 21%. No entanto, os modelos de RNA e Regressão Lasso não obtiveram bons resultados, com MAPEs próximos a 100%, indicando que esses modelos não são indicados para predição do custo de obras de presídios. Percebeu-se que a não adequação dos modelos não considerados válidos deve-se à pequena quantidade de dados trabalhados no banco de dados. Os modelos de AD e SVM apresentaram melhorias quando comparados ao modelo desenvolvido por Beltrão *et al.* (2022) que utiliza para predição dois modelos válidos baseados em regressão linear simples, sendo que nenhum dos modelos foi considerado boa previsão. Com base nos resultados, podemos concluir que as técnicas de ML são ferramentas eficazes para a estimativa de custos de construção de presídios, sendo capazes de estimar os custos com alta precisão, superando as aplicações de estimativas de regressão linear simples. Ao fornecer estimativas de custos mais precisas, os modelos podem ajudar a reduzir desperdícios e aumentar a eficiência do processo de construção. O uso de ferramentas computacionais nas fases iniciais dos empreendimentos contribui para escolhas de planejamento mais precisas e redução da imprevisibilidade de custos.

Destaca-se também, a ótima viabilidade na utilização de bases de dados secundárias de trabalhos para novas aplicações dos dados, incremento de técnicas e análise de resultados. Os dados de outros trabalhos podem representar uma comparação direta entre os resultados do uso de diferentes técnicas para um mesmo conjunto de informações.

Recomendações para trabalhos futuros:

• Realizar estudos com um conjunto de dados maior e mais diversificado. Isso pode ajudar a melhorar ainda mais a precisão dos modelos de *machine learning*, bem como possibilitar a utilização de diferentes modelos que se ajustem melhor à quantidade de dados disponível.

• Investigar outras técnicas de ML para a estimativa de custos de construção de presídios. Novas técnicas podem ser ainda mais eficazes do que as utilizadas neste estudo.

• Aplicar os modelos de *machine learning* em projetos de edificações com outras tipologias específicas. Isso permitirá avaliar a viabilidade e a eficácia dos modelos na prática.

Este estudo contribui para a literatura sobre o uso de ML na estimativa de custos de construção. Os resultados demonstram que as técnicas de *machine learning* podem ser ferramentas valiosas para a gestão de custos em obras públicas de presídios ainda nas fases iniciais.

6 AGRADECIMENTOS

Agradeço ao Grupo de Pesquisa e Assessoria em Gerenciamento na Construção Civil (GERCON) pelo apoio e ensinamentos nas discussões de assuntos relevantes para elaboração deste trabalho. Agradeço também ao apoio oferecido pela FUNCAP por meio da bolsa de auxílio da instituição.

REFERÊNCIAS

AL-TAWAL, D.R.; ARAFAH, M.; SWEIS, G.J. A model utilizing the artificial neural network in cost estimation of construction projects in Jordan. Engineering, Construction and Architectural Management, Vol. 28 No. 9, 2020, pp. 2466-2488. https://doi.org/10.1108/ECAM-06-2020-0402.

AN, S.; KANG, K. A study on predicting construction cost of apartment housing using experts' knowledge at the early stage of projects. Journal of the Architectural Institute of Korea, v. 21, n. 6, p. 81–88, 2005.

ASHWORTH, A.; PERERA, S. Cost studies of buildings. 6. ed. Abingdon, Oxon, UK; New York, NY, USA: Routledge, 2015.

BARROS, A. C. S.; FALCÃO, D. F. Orçamento paramétrico como ferramenta de controle de custos na construção civil. In: CONFERÊNCIA INTERNACIONAL DA LATIN AMERICAN REAL ESTATE SOCIETY – LARES, 16., 2016, São Paulo. Anais [...]. São Paulo: LARES, 2016.

BELTRÃO, L. M. P.; CARVALHO, M. T. M.; BLUMENSCHEIN, R. N.; PAIVA, Á. T. DE; FREITAS, M. V. R. DE. Modelos para estimativa de custos com o uso de regressão linear: modelagem com obras penitenciárias. *Ambiente Construído*, v. 22, n. 3, p. 193–211, 2022. DOI: https://doi.org/10.1590/s1678-86212022000300615.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R. A.; STONE, C. J. Classification and regression trees. 1. ed. New York: Chapman and Hall/CRC, 1984. Disponível em: https://doi.org/10.1201/9781315139470.

CHO, J.; CHUN, J. Cost estimating methods for RC structures by quantity takeoff and quantity prediction in the design development stage. *Journal of Asian Architecture and Building Engineering*, v. 14, n. 1, p. 65–72, 2015. DOI: https://doi.org/10.3130/jaabe.14.65.

CREESE, R.; LI, L. Cost estimation of timber bridge using neural networks. *Cost Engineering*, v. 37, n. 5, p. 17–22, 1995.

DATAFLAIR. Advantages and disadvantages of machine learning language. 2022. Disponível em: https://data-flair.training/blogs/advantages-and-disadvantages-of-machine-learning/. Acesso em: 04 jun. 2024.

DELGADO, A.; OYEDELE, L. Deep learning with small datasets: using autoencoders to address limited datasets in construction management. *Applied Soft Computing*, v. 112, p. 107836, 2021. DOI: https://doi.org/10.1016/j.asoc.2021.107836.

ELFAKI, A. O.; ALATAWI, S.; ABUSHANDI, E. Using intelligent techniques in construction project cost estimation: 10-year survey. Advances in Civil Engineering, p. 107926, 2014. DOI: https://doi.org/10.1155/2014/107926.

ELMOUSALAMI, H. H. Artificial intelligence and parametric construction cost estimate modeling: state-of-the-art review. Journal of Construction Engineering and Management, v. 146, n. 1, 2019. DOI: https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678.

ELMOUSALAMI, H. H. Comparison of artificial intelligence techniques for project conceptual cost prediction: a case study and comparative analysis. *IEEE Transactions on Engineering Management*, v. 68, n. 1, p. 183–196, 2021. DOI: https://doi.org/10.1109/TEM.2020.2972078.

ESTEVES, A. G. L.; MEDEIROS, L. M.; MEDEIROS, F. M. M.; PERKUSICH, M. Avaliação de modelos preditivos de regressão para estimar esforço de software. Revista de Sistemas e Computação, Salvador, v. 8, n. 2, p. 592–604, jul./dez. 2018. Disponível em: https://www.researchgate.net/publication/333001575_Avaliacao_de_Modelos_de_Preditivos_de_Regressao_para_Estimar_Esforco_de_Software. Acesso em: 16 abr. 2025.

GÉRON, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. 2. ed. Sebastopol: O'Reilly Media, 2019.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. *Data mining: concepts and techniques*. 3. ed. San Francisco: Morgan Kaufmann, 2011.

HASHEMI, S. T.; EBADATI, O. M.; KAUR, H. Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. SN Applied Sciences, v. 2, n. 10, p. 1–27, 2020. DOI: https://doi.org/10.1007/s42452-020-03497-1.

JAAFARI, A; PAZHOUHAN, I; BETTINGER, P. Machine Learning Modeling of Forest Road Construction Costs. Forests, [S.L.], v. 12, n. 9, p. 1169, 28 ago. 2021. MDPI AG. http://dx.doi.org/10.3390/f12091169.

JIN, R. et al. MRA-based revised CBR model for cost prediction in the early stage of construction projects, **Expert Syst**. **Appl**. 39 (5) (2012) 5214–5222, 4. 2012.

KUMAR, A.; THAKUR, M.; GUPTA, A. Machine Learning and Deep Learning. 2021. Disponível em: https://arxiv.org/abs/2104.05314. Acesso em: 23 abr. 2025.

LEE, J. W. et al. Privacy-Preserving Machine Learning With Fully Homomorphic Encryption for Deep Neural Network. in IEEE Access, vol. 10, pp. 30039-30054, 2022. doi: 10.1109/ACCESS.2022.3159694.

LEWIS, C. D. 1982. Industrial and business forecasting methods. London: Butterworth.

MIRANDA, C S.L.; CASTILLO, E. D. R.; GONZALEZ, V.; ADAFIN, J. Predictive Analytics for Early-Stage Construction Costs Estimation. **Buildings**, 12, 1043, 2022. https:// doi.org/10.3390/buildings12071043.

OMER, P. A. Improving prediction accuracy of Lasso and Ridge regression as an alternative to LS regression to identify variable selection problems. Zanco Journal of Pure and Applied Sciences, v. 34, supl. 6, p. 33–45, 2022. DOI: 10.21271/ZJPAS.34.s6.5. Disponível em: https://www.researchgate.net/publication/385137550. Acesso em: 5 maio 2025.

OTERO, J.A. Análise paramétrica de dados orçamentários para estimativas de custo na construção de edifícios: estudo de caso voltado para a questão da variabilidade. 214f. 2000. Dissertação (Mestrado em Engenharia de Produção). Universidade Federal de Santa Catarina, Florianópolis, 2000. Disponível em: https://repositorio.ufsc.br/xmlui/bitstream/123456789/78481/1/170830.pdf

PETRUSEVA, S.; ZILESKA P.; ŽUJO V. BRKAN V. Previsão de custos de construção: comparação da precisão da regressão linear e modelos de máquinas de vetores de suporte. Tehnicki Vjesnik-Technical Gaz. 24 (5): 1431–1438, 2017.

PROJECT MANAGEMENT INSTITUTE (PMI). Construction extension to the PMBOK® Guide. 2. ed. Newtown Square: Project Management Institute, 2016.

RATNER, B. Variable selection methods in regression: ignorable problem, outing notable solution. J**ournal of Targeting, Measurement and Analysis for Marketing**, v. 18, n. 1, p. 65–75, 2010. Disponível em: https://doi.org/10.1057/jt.2009.26.

SHIN, Y.; KIM, D. W.; KIM, J. Y.; KANG, K. I.; CHO, M.-Y.; CHO, H.-H. Application of Adaboost to the Retaining Wall Method Selection in Construction, **Journal of Computing in Civil Engineering**, Vol. 23, No. 3, 2009, pp. 188-192. doi:10.1061/(ASCE)CP.1943-5487.0000001.

SON, H.; KIM, C.; KIM, C. Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. Automation in Construction, v. 27, p. 60–66, 2012.

STEVENS, J. P. Applied multivariate statistics for the social sciences. 4th ed. Hillsdale, NJ: Erlbaum, 2002.

TANIM, SAKHAWAT HUSSAIN; AHMAD, MD SABBIR. AI driven strategic decision-making in IT project management: Enhancing risk assessment, cost control, and efficiency. World Journal of Advanced Research and Reviews, v. 25, n. 2, p. 247–268, 2025. Disponível em: https://journalwjarr.com/content/ai-driven-strategic-decision-making-it-project-management-enhancing-risk-assessment-cost. Acesso em: 23 mai. 2025.

VAPNIK, V. Estimation of Dependences Based on Empirical Data (in Russian). Moscow: Nauka. 1979.

YANG, S.W.; MOON, S.W.; JANG, H.; CHOO, S.; KIM, S.A. Parametric Method and Building Information Modeling-Based Cost Estimation Model for Construction Cost Prediction in Architectural Planning. Appl. Sci. 2022, 12, 9553. https://doi.org/10.3390/app12199553.

YUN, S. Predictive Performance of Building Construction Estimation: An Analysis based on ANN Model. **Journal of System and Management Sciences.** Vol. 12. No. 2, pp. 325-335, 2022. Doi: 10.33168/JSMS.2022.0216.