

O Uso De Aprendizado De Máquina Na Estimativa Do Déficit Habitacional: Uma Revisão Sistemática Da Literatura

Lina Sun Young Park

Graduanda em Estatística, UFSCar, Brasil

linapark@estudante.ufscar.br

ORCID: 0009-0005-4893-2784

Pedro Henrique Lopes de Almeida

Graduando em Ciências da Computação, UFV, Brasil

pedro.almeida3@ufv.br

ORCID: 0009-0008-5070-3982

Leonardo Cavalcante da Silva

Graduando em Ciências da Computação, UFSCar, Brasil

leonardo.silva@estudante.ufscar.br

ORCID: 0009-0003-4006-6637

Guilherme Henrique de Carvalho

Graduando em Matemática, UFSCar, Brasil

guilherme.carvalho@estudante.ufscar.br

ORCID: 0009-0000-8487-1118

Priscila Kauana Barelli Forcel

Mestra em Engenharia Urbana, UFSCar, Brasil

priscilaforcel@ufscar.br

ORCID: 0000-0002-1321-4716

Tatiane Ferreira Olivatto

Doutoranda em Engenharia Urbana, UFSCar, Brasil

tatianeolivatto@ufscar.br

ORCID: 0000-0002-5770-7088

Rafael de Paula Garcia

Professor Doutor, UFV, Brasil

rafael.pgarcia@ufv.br

ORCID: 0000-0002-1839-1842

Elza Luli Miyasaka

Professora Doutora, UFSCar, Brasil

elza.miyasaka@ufscar.br

ORCID: 0000-0003-4480-9672

O Uso De Aprendizado De Máquina Na Estimativa Do Déficit Habitacional: Uma Revisão Sistemática Da Literatura

RESUMO

Objetivo - Identificar estudos que aplicam técnicas vinculadas à Inteligência Artificial, incluindo aprendizado de máquina, aprendizado profundo e mineração de dados, à análise do déficit habitacional ou a tarefas correlatas fundamentadas em dados censitários.

Metodologia - A identificação dos estudos ocorre pelas etapas de revisão bibliométrica e sistemática da literatura científica.

Originalidade/relevância - Com os avanços recentes em métodos computacionais, especialmente em algoritmos de aprendizado de máquina, surge o interesse em avaliar seu potencial para lidar com a complexidade do fenômeno do déficit habitacional e subsidiar a formulação de políticas urbanas e habitacionais mais eficazes.

Resultados - A partir da base de dados *Scopus*, foram identificados 1.528 documentos publicados entre 1985 e 2024 que articulam métodos computacionais e dados censitários, dos quais 18, publicados entre 2015 e 2024, abordam diretamente questões relacionadas ao déficit habitacional.

Contribuições teóricas/metodológicas - Os resultados indicam a predominância de algoritmos baseados em árvores, com destaque para o método de Floresta Aleatória.

Contribuições sociais e ambientais - A diversidade de áreas de aplicação e de tipos de dados utilizados limita a definição de um modelo ideal, evidenciando a necessidade de aprofundamento nas abordagens e contextos analisados.

PALAVRAS-CHAVE: Censo, Dados Demográficos, Estudos Urbanos, Inteligência Artificial, Revisão Bibliométrica.

The Use Of Machine Learning In Estimating The Housing Deficit: A Systematic Literature Review

ABSTRACT

Objective – Identify studies that apply techniques related to Artificial Intelligence — including machine learning, deep learning, and data mining — to the analysis of the housing deficit or similar tasks based on census data.

Methodology – The studies are identified through bibliometric and systematic reviews of the scientific literature.

Originality/Relevance – With recent advances in computational methods, especially machine learning algorithms, there is interest in evaluating their potential to deal with the complexity of the housing deficit phenomenon and support the formulation of more effective urban and housing policies.

Results – Using the Scopus database, 1,528 documents published between 1985 and 2024 were identified that combine computational methods and census data, of which 18, published between 2015 and 2024, directly address issues related to the housing deficit.

Theoretical/Methodological Contributions – The results indicate the predominance of tree-based algorithms, particularly the Random Forest method.

Social and Environmental Contributions – the diversity of application areas and data types used limits the identification of an ideal model, highlighting the need for further exploration of the approaches and contexts analyzed.

KEYWORDS: Census, Demographic Data, Urban Studies, Artificial Intelligence, Bibliometric Review.

El Uso De Aprendizaje Automático En La Estimación Del Déficit Habitacional: Una Revisión Sistemática De La Literatura

RESUMEN

Objetivo – Identificar estudios que apliquen técnicas relacionadas con la Inteligencia Artificial —incluyendo aprendizaje automático, aprendizaje profundo y minería de datos— al análisis del déficit habitacional o a tareas similares basadas en datos censales.

Metodología – La identificación de los estudios se realiza mediante etapas de revisión bibliométrica y sistemática de la literatura científica.

Originalidad/Relevancia – Con los recientes avances en métodos computacionales, especialmente en algoritmos de aprendizaje automático, hay interés en evaluar su potencial para tratar la complejidad del fenómeno del déficit habitacional y apoyar el desarrollo de políticas urbanas y habitacionales más eficaces.

Resultados – A partir de la base de datos Scopus, se identificaron 1.528 documentos publicados entre 1985 y 2024 que combinan métodos computacionales y datos censales, de los cuales 18, publicados entre 2015 y 2024, abordan directamente cuestiones relacionadas con el déficit habitacional.

Contribuciones Teóricas/Metodológicas – Los resultados indican la predominancia de algoritmos basados en árboles, destacándose el método de Bosque Aleatorio.

Contribuciones Sociales y Ambientales – La diversidad de áreas de aplicación y tipos de datos utilizados limita la identificación de un modelo ideal, lo que evidencia la necesidad de profundizar en los enfoques y contextos analizados.

PALABRAS CLAVE: Censo, Datos Demográficos, Estudios Urbanos, Inteligencia Artificial, Revisión Bibliométrica.

1 INTRODUÇÃO

O processo de urbanização brasileira, marcado pela estreita relação entre o capital industrial e a produção da cidade, gerou múltiplas dimensões morfológicas e sociais que refletem profundas desigualdades territoriais (Maricato, 2002; Santos, 1980; Villaça, 1999; Feldman, 2005). Desde a década de 1960, o crescimento da população urbana impulsionou demandas por moradia, transporte e infraestrutura, e expôs as insuficiências das políticas públicas para atender à população crescente (Maricato, 2002).

Essa dinâmica resultou na formação de vastas áreas urbanas irregulares, onde a população de baixa renda, muitas vezes excluída do mercado formal de habitação, ocupa territórios periféricos e desvalorizados. Tal fenômeno evidencia o caráter estrutural e excludente da produção do espaço urbano (Retto Júnior; Maricato, 2021; Villaça, 2012; Bonduki, 1988).

O déficit habitacional no Brasil permanece um desafio grave e multifacetado, que ultrapassa a marca de 6 milhões de moradias (FJP, 2024). Tradicionalmente tratado de forma quantitativa, o déficit deve ser compreendido como um fenômeno complexo, atravessado por dimensões sociais, territoriais e políticas que afetam diretamente o direito à moradia digna e à justiça urbana (Arantes *et al.*, 2002; Prieto; Laczynski, 2020; Rolnik, 2015).

No país, a metodologia criada pela Fundação João Pinheiro (FJP), em 1995, é tida como principal referência na estimativa do déficit habitacional. Segundo a FJP (2013), o conceito divide-se em dois aspectos: o déficit propriamente dito, que corresponde à demanda por novas unidades habitacionais, e a inadequação habitacional, que se refere à necessidade de melhorias nas moradias existentes. Essa distinção diferencia o déficit quantitativo, ligado à insuficiência de unidades habitacionais, do déficit qualitativo, relacionado às condições precárias das moradias e à ausência de infraestrutura adequada (Genevois; Costa, 2001; FJP, 2013).

Em outros países, como o Canadá, o déficit considera tanto aspectos habitacionais quanto financeiros, e asseguram que as moradias estejam localizadas em áreas adequadas e acessíveis, e que as famílias tenham condições de acessá-las (Santos, 2022). Na América Latina, países como Argentina, Chile e Panamá incluem entre os critérios de déficit habitacional as moradias irrecuperáveis (Arriagada Luco, 2005).

A mensuração do déficit habitacional revela-se, portanto, como uma tarefa complexa, cujos parâmetros e métodos utilizados variam de acordo com diferentes aspectos regionais, socioculturais e temporais, além das limitações dos dados disponíveis. Embora o método da FJP predomine no Brasil, diferentes abordagens complementares ampliam a compreensão do fenômeno (Magnabosco; Cunha; Garcia, 2012; Genevois; Costa, 2001). A existência dessas divergências, ainda de necessária para o avanço do campo, torna o tema mais complexo e impacta diretamente a formulação de políticas urbanas e habitacionais.

Diante da crescente complexidade dos fenômenos urbanos e da diversidade de fontes e formatos de dados, o campo da habitação tem incorporado progressivamente ferramentas computacionais avançadas oriundas da ciência de dados. Entre elas, destacam-se os conceitos de Inteligência Artificial (IA), que envolvem sistemas capazes de executar tarefas que requerem raciocínio, percepção e aprendizado, antes restritos à inteligência humana (Russell; Norvig, 2020). Uma de suas subáreas mais relevantes é o Aprendizado de Máquina (AM), que se refere a métodos computacionais capazes de identificar padrões e realizar previsões a partir de grandes volumes de dados, sem programação explícita (Mitchell, 1997; Bishop, 2006).

Associada a esses métodos, a Mineração de Dados extrai informações relevantes e estruturadas a partir de grandes bases de dados, e é frequentemente aplicada em estudos urbanos e habitacionais (Han; Kamber; Pei, 2011). Já o Aprendizado Profundo, uma vertente mais recente do AM, utiliza redes neurais artificiais com múltiplas camadas para modelar relações complexas entre variáveis, como destaque para reconhecimento de imagens, classificação de padrões e análise espacial (Goodfellow; Bengio; Courville, 2016). A articulação dessas abordagens mostra-se promissora para enfrentar desafios analíticos em habitação e planejamento urbano, dada sua capacidade de lidar com dados de alta dimensionalidade, heterogeneidade e incerteza.

Nesse contexto, os avanços na área de IA oferecem ferramentas para enfrentar a complexidade e a variabilidade de dados habitacionais. O uso e o ajuste dessas técnicas permitem trabalhar com diferentes parâmetros e conjuntos de dados, e reduzir a complexidade da mensuração a fim de favorecer a identificação de padrões antes ocultos.

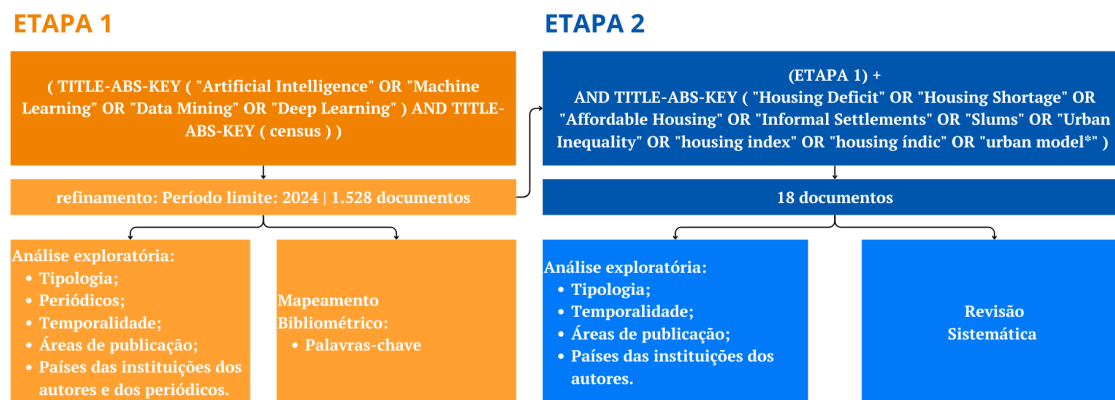
Portanto, este estudo apresenta, por meio de uma revisão bibliométrica e sistemática, um panorama da aplicação de métodos de AM na identificação do déficit habitacional e em trabalhos correlatos com dados demográficos e censitários. Para isso, foram analisados documentos indexados na base de dados *Scopus*, reconhecida pela abrangência e qualidade científica. A análise bibliométrica representa uma técnica de grande impacto para mapear aspectos acadêmicos e identificar tendências de pesquisa, especialmente em produções interdisciplinares (Ellegaard; Wallin, 2015; Medeiros; Vitoriano, 2015).

2 MATERIAIS E MÉTODOS

Esta pesquisa adotou uma abordagem metodológica combinada em duas etapas: análise bibliométrica e revisão sistemática da literatura. Ambas com o objetivo de identificar como técnicas de Aprendizado de Máquina são aplicadas em estudos voltados à mensuração do déficit habitacional ou em análises correlatas baseadas em dados censitários e demográficos.

As fases metodológicas representadas na Figura 1 foram estruturadas a partir de documentos indexados na base *Scopus*, coletados em 28 de janeiro de 2025. Para assegurar a robustez da revisão, a etapa inicial definiu claramente os termos de busca e os critérios de inclusão e exclusão (Silva *et al.*, 2021). A investigação foi organizada em duas etapas complementares:

Figura 1 - Fluxograma dos procedimentos metodológicos



Fonte: elaborado pelos autores (2025)

A etapa 1 consistiu na identificação de trabalhos que aplicaram métodos de inteligência artificial ou mineração de dados em tarefas com dados censitários. A *string* de busca foi composta por dois grupos de termos em inglês: o primeiro "*Artificial Intelligence*", "*Machine Learning*", "*Data Mining*" e "*Deep Learning*", que representam os métodos computacionais de interesse; no segundo o termo "*census*", relacionado à natureza dos dados utilizados. Os grupos foram combinados através dos operadores booleanos *AND* e *OR* na *Scopus*. E resultou em 1.545 documentos, dos quais 1.528 correspondem ao limite temporal até 2024. Os resultados desta etapa fundamentam a análise bibliométrica.

Na sequência, a busca foi refinada com a inclusão de um terceiro grupo de termos específicos relacionados ao déficit habitacional: "*Housing Deficit*", "*Housing Shortage*", "*Affordable Housing*", "*Informal Settlements*", "*Slums*", "*Urban Inequality*", "*Housing Index*", "*Housing Indic*" e "*Urban Model*". A combinação com os anteriores limitou os resultados a estudos que articulam técnicas computacionais com temáticas habitacionais. Essa triagem identificou 18 artigos, que serviram de base para a análise exploratória e a revisão sistemática apresentadas neste artigo.

O conjunto dessas etapas permitiu identificar pesquisas que aplicam aprendizado de máquina às análises do déficit habitacional, da vulnerabilidade urbana e da caracterização de assentamentos precários. Esses estudos oferecem subsídios para o aprimoramento de diagnósticos territoriais e para o desenvolvimento de políticas públicas habitacionais.

3 RESULTADOS E DISCUSSÕES

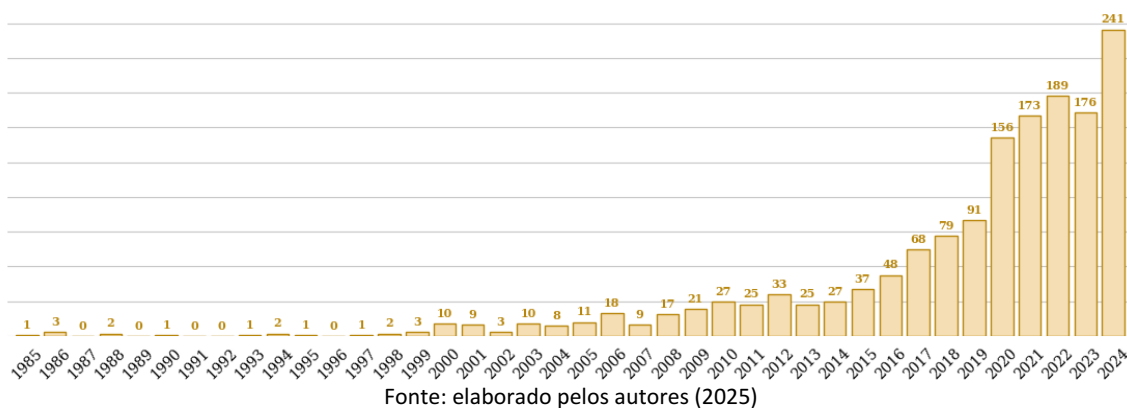
3.1 Etapa 1 | Aprendizado de Máquina aplicado a dados censitários

A busca dos termos "*Artificial Intelligence*", "*Machine Learning*", "*Data Mining*" e "*Deep Learning*", em conjunto com "*census*", resultou em 1.528 documentos. O primeiro documento indexado é um artigo de conferência de 1985, da Universidade de Missouri (Estados Unidos), que apresenta uma análise exploratória de dados por meio de um sistema especialista que combina raciocínio simbólico e computação numérica para avaliar incertezas. Para testar a

robustez do modelo, o autor utilizou uma matriz de dados extensa e complexa, comparável a um conjunto de dados censitários.

Quanto à tipologia de documentos, predominam os artigos científicos (881 registros, 57,65%). Em seguida, aparecem os artigos de conferência (566 publicações, 37,04%), o que evidencia a relevância dos eventos científicos como espaços para divulgação de pesquisas em desenvolvimento. Foram identificados ainda 33 capítulos de livro, 29 artigos de revisão e 19 registros referentes a notas, editoriais e erratas. Essa distribuição contribui para delinear um panorama científico do tema analisado.

Figura 2 - Distribuição anual dos documentos acadêmicos



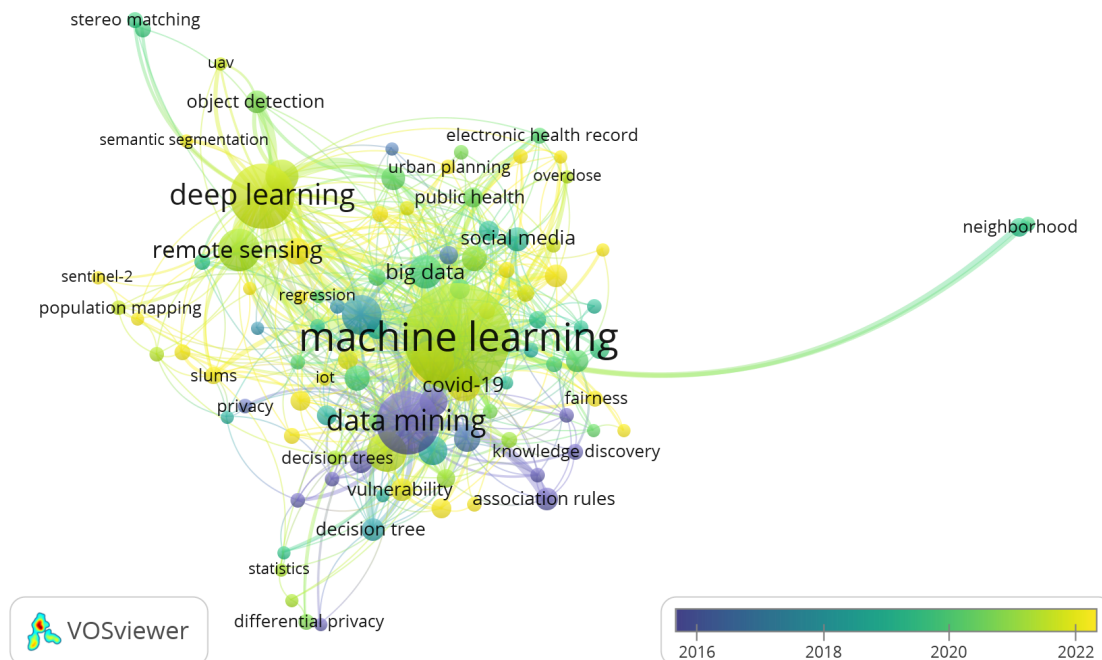
A evolução temporal das publicações (Figura 2) indica um crescimento expressivo a partir de 2000, com dez publicações. Em 2006, registraram-se dezoito publicações, seguidas de uma queda. A partir de 2015, a produção cresceu de forma constante, e alcançou a marca de 91 documentos em 2019, 156 em 2020, 189 em 2021, 176 em 2023 e 241 em 2024. A crescente entre os anos de 2019 e 2021 pode estar associada à intensificação da análise de dados censitários durante a pandemia de COVID-19, tema recorrente nas palavras-chave dos artigos (ver Figura 3).

Quanto às áreas do conhecimento, destacam-se Ciência da Computação (25,4%), Engenharias (11,9%) e Ciências Sociais (10,2%). Os Estados Unidos lideram a produção com 643 documentos, seguidos por China (175) e Reino Unido (116). O Brasil aparece em 10º lugar, com 44 publicações.

Para aprofundar a análise, foi utilizado o *software VOSviewer*, que mapeou 3.841 palavras-chave indexadas pelos autores. Após aplicar o filtro aos termos com ao menos cinco ocorrências, restaram 111 palavras-chave. Com o uso da função *thesaurus*, que agrupou termos semelhantes - “*census data*” e “*census*”, “*ai*” e “*artificial intelligence*”, “*social vulnerability*” e “*vulnerability*”, obteve-se 101 palavras mapeadas.

O mapeamento bibliométrico (Figura 3) revela a centralidade do termo “*machine learning*”, associado principalmente a “*remote sensing*”, “*random forest*”, “*prediction*”, “*artificial intelligence*” e “*covid-19*”. Isso evidencia a aplicação de ferramentas computacionais e estatísticas a bases de dados censitários e georreferenciados. “*Machine learning*” é o termo mais recorrente, com 319 ocorrências, seguido por “*deep learning*” (118), “*data mining*” (111) e “*artificial intelligence*” (53). O termo “*census*” aparece 46 vezes, conectado aos termos anteriores, além de “*remote sensing*”, “*random forest*” e “*big data*”.

Figura 3 - Mapa bibliométrico das palavras-chave mencionadas pelos autores



Fonte: elaborado pelos autores, gerado pelo VOSviewer (2025)

A distribuição temporal dos termos evidencia maior ênfase de “*machine learning*” em 2021. “*Data mining*”, uma ferramenta estatística clássica, teve maior uso em 2016, associado a “*association rules*”. “*Remote sensing*” alcançou destaque em 2021, enquanto “*deep learning*” ganhou relevância ao final do mesmo ano.

A partir de 2022, surgem termos diretamente relacionados ao déficit habitacional, como “*slums*”, “*poverty*”, “*housing*” e “*vulnerability*”. Observa-se que “*remote sensing*”, conectado a “*deep learning*” e “*machine learning*”, apresenta conexões relevantes com “*satellite images*”, “*slums*” e “*housing*”, o que reforça a tendência de aplicações urbanas e territoriais no uso de dados censitários.

Por fim, a análise dos periódicos mostra que os 1.528 documentos estão publicados em 138 fontes de publicações, sendo que cinco delas reúnem mais de 14 publicações cada, que somam 155 artigos (10,14% do total). A Tabela 1 apresenta os cinco periódicos com maior número de artigos sobre o tema, acompanhados de seus respectivos número de citações e *CiteScore*.

Tabela 1 - Periódicos com mais publicações

PERIÓDICO	Nº DE PUBLICAÇÕES	Nº DE CITAÇÕES	CITE SCORE (2023)
Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	81	1886	2.6
Plos One	26	776	6.2
Remote Sensing	17	204	8.3
Science Of the Total Environment	16	701	17.6
Ceur Workshop Proceedings	15	17	1.1

Fonte: elaborado pelos autores (2025)

O *CiteScore*¹, métrica que mede o impacto de periódicos científicos, evidencia diferenças significativas entre produtividade e relevância. Por exemplo, apesar de o *“Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics”* liderar em número de publicações, apresenta *CiteScore* 2,6. Já o *“Science of the Total Environment”* alcança 17,6, o maior *CiteScore* entre os periódicos analisados, mesmo com menor volume de artigos.

O *“Science Of The Total Environment”* apresenta o maior *CiteScore* (17,6), com 16 documentos e 701 citações. Entre eles, destaca-se o artigo *“Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model”*, publicado em 2019, citado 218 vezes. O artigo de conferência mais citado é *“Non-parametric local transforms for computing visual correspondence”*, publicado em 1994 no volume 801 da série *“Lecture Notes In Computer Science”*, com 1371 citações.

Apesar das diferenças entre volume de publicações e impacto médio medido pelo *CiteScore*, o desempenho individual dos artigos merece destaque. Alguns estudos publicados em periódicos de menor impacto obtiveram repercussão na comunidade científica, o que demonstra a relevância de estudos específicos no campo. A Tabela 2 apresenta os dez artigos mais citados, com informações sobre o tipo de documento, ano de publicação, número de citações acumuladas e o *CiteScore* do respectivo periódico. Também indica os países das instituições dos autores e países-sede dos periódicos.

¹ Calculado anualmente, representa a média de citações recebidas por artigos publicados em um periódico nos últimos quatro anos. Seu cálculo consiste em dividir o número de citações recebidas pelo número de documentos publicados nesse período.

Tabela 2 - Artigos mais citados

TÍTULO (país das instituições dos autores)	PERIÓDICO (país sede)	TIPO DO DOCUMENTO	ANO DE PUBLICAÇÃO	NÚMERO DE CITAÇÕES	CITESCORE DO PERIÓDICO (2023)
Non-parametric local transforms for computing visual correspondence (Estados Unidos)	Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics (Alemanha)	Artigo de Conferência	1994	1371	2.6
Beyond Market Baskets: Generalizing Association Rules to Correlations (Estados Unidos)	SIGMOD Record (ACM Special Interest Group on Management of Data) (Estados Unidos)	Artigo	1997	1071	3.1
Mitigating Unwanted Biases with Adversarial Learning (Estados Unidos)	AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (Estados Unidos)	Artigo de Conferência	2018	829	-
Evaluation of stereo matching costs on images with radiometric differences (Alemanha e Estados Unidos)	IEEE Transactions on Pattern Analysis and Machine Intelligence (Estados Unidos)	Artigo	2009	668	28.4
Word embeddings quantify 100 years of gender and ethnic stereotypes (Estados Unidos)	Proceedings of the National Academy of Sciences of the United States of America (Estados Unidos)	Artigo	2018	661	19.0
Efficient mining of association rules using closed itemset lattices (França)	Information Systems (Reino Unido)	Artigo	1999	599	9.4
Origin-destination trips by purpose and time of day inferred from mobile phone data (Estados Unidos)	Transportation Research Part C: Emerging Technologies (Reino Unido)	Artigo	2015	490	15.8
Crowd analysis: A survey (Reino Unido)	Machine Vision and Applications (Alemanha)	Artigo	2008	486	6.3

UnFlow: Unsupervised learning of optical flow with a bidirectional census loss (Alemanha)	32nd AAAI Conference on Artificial Intelligence, AAAI 2018 (Estados Unidos)	Artigo de Conferência	2018	392	-
Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium (Estados Unidos, Sudão, Sri Lanka, Índia, Austrália)	International Journal of Remote Sensing (Reino Unido)	Artigo	2009	376	7.0

Fonte: elaborada pelos autores (2025)

Observa-se que o periódico *“Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics”* apesar do *CiteScore* ser relativamente baixo (2.6), aparece tanto entre os artigos mais citados quanto entre os que mais publicaram sobre o tema, possivelmente em razão de sua ampla difusão em conferências da área. Em contraste, o periódico com maior *CiteScore* é o *“IEEE Transactions on Pattern Analysis and Machine Intelligence”*, com 28.4 em 2023, mesmo com menor número de publicações relacionadas ao recorte desta pesquisa.

Os artigos *“Mitigating Unwanted Biases with Adversarial Learning”* (AIES, 2018) e *“UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss”* (AAAI, 2018) não possuem *CiteScore* atribuído na base *Scopus* por se tratarem de anais de conferências. Apesar disso, a AAAI é de grande relevância, tendo apresentado, em 2020, uma média de 9.662 citações por artigo, indicando elevado impacto acadêmico. Já os anais da AIES (Conference on AI, Ethics, and Society) apresentaram, também em 2020, uma média de 3.600 citações por documento.

A análise dos países de origem dos artigos mais citados confirma a predominância dos Estados Unidos na autoria dos trabalhos mais citados, embora Alemanha e Reino Unido também se destaquem. O décimo artigo mais citado apresenta autoria internacional diversificada, com instituições de Estados Unidos, Sudão, Sri Lanka, Índia e Austrália, o que evidencia o caráter global das pesquisas que envolvem dados censitários e inteligência artificial.

As análises da Etapa 1 fornecem um panorama da evolução do uso de inteligência artificial, aprendizado de máquina e outras técnicas computacionais aplicadas a dados censitários ao longo das últimas quatro décadas. O volume de artigos e a diversidade de temas demonstram o amadurecimento desse campo de investigação, que integra tanto as Ciências da Computação quanto as Ciências Sociais.

Diante do objetivo deste estudo, mapear como essas abordagens podem contribuir para a estimativa do déficit habitacional, os achados desta primeira etapa são fundamentais para subsidiar a escolha de abordagens computacionais mais eficazes. Na Etapa 2, a análise será aprofundada com base em publicações que articulam diretamente esses métodos com a temática habitacional, especialmente no contexto do déficit e da vulnerabilidade urbana. Entre

os artigos filtrados, destacam-se estudos que utilizam microdados censitários para investigar tendências de coabitação intergeracional em países em desenvolvimento.

3.2 Etapa 2 | Técnicas Computacionais Aplicadas à Temática Habitacional

Nesta etapa, a estratégia de busca foi refinada com a introdução de termos relacionados diretamente à temática habitacional e urbana: "*Housing Deficit*", "*Housing Shortage*", "*Affordable Housing*", "*Informal Settlements*", "*Slums*", "*Urban Inequality*", "*Housing Index*", "*Housing Indic*" e "*Urban Model**". A combinação desses termos com os anteriores permitiu identificar publicações que, além de empregar técnicas estatísticas e computacionais com dados censitários, tratam de forma explícita o déficit habitacional ou termos correlatos, como moradia, informalidade e desigualdade urbana. A busca resultou em 18 documentos, distribuídos entre 11 artigos científicos, 4 artigos de conferência e 3 capítulos de livros. A distribuição temporal concentra-se entre 2015 a 2024, com destaque para 2024, que registrou o maior número de publicações - 5 documentos, seguido por 2019 e 2023, com 4 publicações cada.

As áreas de conhecimento evidenciam o caráter interdisciplinar do tema. As Ciências Sociais lideram com 24,4% das publicações (11 documentos), seguida por Ciências da Computação 22,2% (10), Ciências Ambientais com 15,6% (7), Ciências da Terra e Planetárias 11,1% (5). Há contribuições pontuais em Física e Astronomia, Engenharia, Agricultura e Ciências Biológicas, Medicina, Matemática, Energia, Economia, Econometria e Finanças e Negócios, Gestão e Contabilidade. Esse panorama revela o potencial integrador do tema, ao cruzar fronteiras disciplinares entre ciências exatas, sociais e aplicadas.

Quanto à distribuição geográfica, os Estados Unidos lideram com 7 publicações, seguidos por Holanda (5), Bélgica e Espanha (com 4 documentos cada), Reino Unido (3) e Brasil (2), acompanhado por Quênia e Suécia. Essa diversidade indica o crescente interesse internacional por tecnologias aplicadas ao planejamento urbano e à habitação, especialmente em contextos urbanos heterogêneos e de alta vulnerabilidade social.

Após a leitura integral dos documentos, realizou-se uma revisão sistemática descritiva e analítica, com o objetivo de compreender como os estudos empregam técnicas de aprendizado de máquina na análise do déficit habitacional, vulnerabilidade urbana e assentamentos precários. As metodologias envolvem dados censitários, sensoriamento remoto e fontes complementares. Cada estudo foi analisado quanto ao contexto territorial, à abordagem metodológica, aos algoritmos utilizados, às limitações identificadas e às recomendações para pesquisas futuras. Essa sistematização fornece subsídios metodológicos para a aplicação de modelos adaptados à realidade brasileira, especialmente na desagregação de dados censitários para estimar o déficit habitacional em diferentes escalas.

Mazzei e Palma (2015), no estudo "*A decision support system for the analysis of mobility*", analisam padrões de mobilidade urbana na Itália a partir de modelos de interação espacial e sistemas de suporte à decisão. O objetivo central é compreender os meios de transporte utilizados pela população e os fatores que influenciam seus deslocamentos, incluindo os impactos na poluição do ar. Os dados utilizados são provenientes do censo italiano de 2011, ainda que limitados de maneira intramunicipal. Embora o estudo não aplique técnicas de aprendizado de máquina, a modelagem é implementada por meio do *software Decision Support System* (DSS2000).

Mahabir *et al.* (2018), em *“Detecting and mapping slums using open data: a case study in Kenya”*, propõem uma abordagem crítica à confiabilidade dos dados sobre favelas, com foco em contextos de países em desenvolvimento. A metodologia combina sensoriamento remoto, dados de plataformas abertas - como *Google Maps* e registros de imobiliárias, e aplicam regressão logística e árvores de decisão. Os principais desafios envolvem a definição do que constitui uma favela e a acurácia dos dados disponíveis. O estudo sugere a incorporação de variáveis socioeconômicas em análises futuras.

Abarca-Alvarez, Reinoso-Bellido e Campos-Sánchez (2019), em *“Decision Model to Predict Social Vulnerability Using Artificial Intelligence”*, desenvolvem um modelo preditivo baseado em Redes Neurais e Árvores de Decisão, com foco na vulnerabilidade social da população da Andaluzia, Espanha e utilizam dados do censo de 2001. O estudo destaca como principais limitações a desatualização dos dados e a necessidade de limpeza e integração dos registros. Os autores recomendam replicações do modelo com bases atualizadas e em diferentes contextos urbanos.

Kuffer *et al.* (2019), no estudo *“Do we underestimate the global slum population?”*, analisam a acurácia das estimativas populacionais em áreas de favela na Tanzânia. A pesquisa integra dados de sensoriamento remoto, amostragens locais e registros de ONGs. O objetivo é evidenciar o potencial do sensoriamento remoto como ferramenta de validação e atualização das bases censitárias. Entre os principais obstáculos, destaca-se a dificuldade de identificação das moradias em áreas densas e irregulares. O uso de Redes Neurais Convolucionais é apontado como uma possibilidade metodológica para estudos futuros.

Engstrom *et al.* (2019), em *“Evaluating the Relationship Between Contextual Features Derived from Very High Spatial Resolution Imagery and Urban Attributes: A Case Study in Sri Lanka”*, analisam a relação entre atributos físicos urbanos (como construções e vias) e características contextuais extraídas de imagens com altíssima resolução espacial (VHSR – *Very High Spatial Resolution*). Utilizam Regressão Linear Múltipla com dados de satélite e do *OpenStreetMap* para avaliar padrões urbanos no Sri Lanka. Entre as limitações, destacam-se o alto custo e a escassez dessas imagens. Os autores sugerem o uso de imagens mais acessíveis e replicação da metodologia em outros países.

Engstrom *et al.* (2019), em *“Mapping Poverty and Slums Using Multiple Methodologies in Accra, Ghana”*, combinam dados censitários, sensoriamento remoto e pesquisas domiciliares para mapear a pobreza e a localização de favelas em Gana. O estudo utiliza o algoritmo de Floresta Aleatória para classificar áreas segundo características domiciliares e acesso a saneamento. As principais limitações são o tamanho da amostra e as variações conceituais sobre o que caracteriza uma favela. Os autores recomendam a inclusão de novas variáveis, análises mais aprofundadas das causas da pobreza e avaliações do impacto de políticas públicas.

Thomson *et al.* (2020), no artigo *“Need for an Integrated Deprived Area “Slum” Mapping System (IDEAMAPS) in Low- and Middle-Income Countries (LMICs)”*, argumentam pela criação de um sistema integrado de mapeamento de áreas urbanas precárias, com base em algoritmos de aprendizado de máquina. O foco é apoiar a meta 11 dos ODS da ONU, que trata de cidades e comunidades sustentáveis. O estudo baseia-se em dados globais, imagens de satélite e registros locais da Índia. Como limitações, apontam a escassez de dados em regiões vulneráveis e a falta de apoio institucional. Para o futuro, propõem a consolidação do sistema e maior envolvimento dos governos locais.

Merodio Gómez *et al.* (2021), em *“Earth Observations and Statistics: Unlocking Sociodemographic Knowledge through the Power of Satellite Images”*, aplicam Redes Neurais Convolucionais (CNNs) e algoritmos como Floresta Aleatória para obter informações socioeconômicas por meio de imagens de satélite em cidades como México, Nairobi e Bangalore. A proposta é integrar observações da Terra e inteligência artificial para classificar áreas privadas e gerar indicadores sociodemográficos. Os principais desafios envolvem a necessidade de grandes volumes de dados, questões éticas e a complexidade dos modelos utilizados. Como recomendação, os autores propõem análises temporais mais robustas, aprimoramento dos modelos de IA e maior participação das comunidades locais nas etapas de coleta.

Lynge *et al.* (2022), em *“Developing neighbourhood typologies and understanding urban inequality: a data-driven approach”*, utilizam o algoritmo K-médias para construir tipologias de bairros na África do Sul, com base em dados do censo de 2011. A abordagem tem caráter indutivo e visa capturar desigualdades urbanas a partir da estrutura socioespacial das vizinhanças. As limitações apontadas incluem a necessidade de interpretações subjetivas e o uso de dados censitários desatualizados. O estudo pretende ser replicado com os dados do censo de 2022.

El Moudden *et al.* (2022), em *“Slum image detection and localization using transfer learning: a case study in Northern Morocco”*, investigam a detecção e localização de favelas no norte do Marrocos a partir de imagens de satélite e técnicas de aprendizado profundo. Utilizam Redes Neurais Convolucionais (CNNs), com abordagem parcial baseada em imagens classificatórias. A principal dificuldade é o redimensionamento das imagens, que pode comprometer a qualidade da análise. Como proposta futura, destacam o uso completo das CNNs com camadas do tipo backbone para refinar a segmentação e a qualidade da classificação.

Tadesse *et al.* (2023), no artigo *“Association among household water, sanitation, and hygiene (WASH) status and typhoid fever risk in urban slums: a prospective cohort study in Bangladesh”*, analisam a relação entre os serviços de básicos de saneamento (WASH - *Water, sanitation and hygiene*) e os riscos de febre tifóide em favelas de Bangladesh. O estudo aplica o algoritmo de Árvore de Decisão para classificar os dados, divididos em 60% para treinamento e 40% para validação. O uso de dados espaciais é destacado, com a perspectiva de aplicação futura em outros contextos urbanos.

Wang *et al.* (2023), em *“Landscape of multiculturalism in Australia: Tracking ethnic diversity and its relation with neighbourhood features in 2001–2021”*, aplicam o método de Floresta Aleatória para investigar a diversidade étnica na Austrália e sua relação com características dos bairros. Utilizam dados censitários de duas décadas (2001 a 2021), e combinam variáveis socioeconômicas, étnicas e habitacionais. Entre as limitações, estão o uso de um único índice de diversidade e o número reduzido de variáveis habitacionais. Como proposta futura, sugerem expandir os dados sobre migração e aprofundar a análise sobre a mobilidade populacional e sua influência na diversidade urbana.

Stacy *et al.* (2023), no artigo *“Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?”*, investigam o impacto das reformas de uso do solo no acesso à moradia nos Estados Unidos. Utilizam algoritmos de aprendizado de máquina para minerar textos jornalísticos, identificam menções a políticas de adensamento urbano, sua localização e efeitos. A base de dados inclui o censo dos EUA e informações do serviço postal nacional. O estudo enfrenta limitações devido à dependência de registros informais (jornais) e à

restrição geográfica da amostra. Como continuidade, os autores sugerem ampliar o escopo da análise para mais cidades e variáveis políticas.

Sagi *et al.* (2024), em *“An Unsupervised Machine Learning Approach for Spatial Analysis of Urban Systems Through Neighborhood Dynamics”*, investigam as dinâmicas territoriais na Inglaterra a partir de transações imobiliárias ao longo do tempo. Utilizam o algoritmo K-médias, uma técnica de aprendizado não supervisionado, para identificar padrões de transformação nos bairros. A análise é baseada em dados detalhados de transações e localização censitária. Um dos principais desafios metodológicos diz respeito à validação dos agrupamentos, que depende da interpretação especializada de profissionais da área urbana.

Khalil, Fatmi e Orvin (2024), em *“Developing and microsimulating demographic dynamics for an integrated urban model: a comparison between logistic regression and machine learning techniques”*, comparam o desempenho de diferentes técnicas de aprendizado de máquina - Floresta Aleatória, eXtreme Gradient Boosting (XGBoost), Redes Neurais e Regressão Logística - em modelos de dinâmica demográfica urbana no Canadá. A análise é feita em Python utilizando bases de dados específicas canadenses. A principal limitação refere-se à representação da variável renda. Como perspectiva futura, os autores propõem o uso de séries históricas mais longas para aprimorar os modelos preditivos.

Abascal *et al.* (2024), no estudo *“Making Urban Slum Population Visible: Citizens and Satellites to Reinforce Slum Censuses”*, propõem um modelo de estimativa populacional de favelas no Quênia por meio da integração entre sensoriamento remoto e dados gerados por moradores, em uma abordagem de ciência cidadã. Utilizam o algoritmo de Floresta Aleatória para relacionar a morfologia urbana e a densidade populacional. A base inclui dados do censo de 2019, imagens de satélite e registros comunitários. Os principais desafios referem-se à precisão dos dados de campo. A expectativa é aplicar a metodologia em outras cidades e incorporar indicadores socioeconômicos para ampliar sua robustez.

Vanhuyse *et al.* (2024), em *“Putting the Invisible on the Map: Low-Cost Earth Observation for Mapping and Characterizing Deprived Urban Areas (Slums)”*, utilizam métodos de observação da Terra para mapear e caracterizar áreas de privação urbana no Quênia. Aplicam estratégias de agrupamento e classificação baseadas em imagens de satélite e dados de cobertura do solo. Entre os desafios estão a baixa disponibilidade de dados e a dificuldade de generalizar os resultados. Como evolução do estudo, os autores sugerem empregar técnicas de aprendizado profundo, utilizar imagens em alta resolução e integrar dados de diversas fontes.

Pedrassoli *et al.* (2024), em *“Reconstructing 36 Years of Spatiotemporal Dynamics of Slums in Brazil by Integrating EO and Census Data”*, mapeiam a evolução espacial e temporal das favelas brasileiras entre 1985 e 2021. Com base em dados do IBGE e imagens de satélite, aplicam classificadores de imagem desenvolvidos pela plataforma *MapBiomas*, com técnicas de aprendizado de máquina não especificadas. A principal limitação refere-se à dependência dos dados do censo brasileiro. O estudo recomenda incorporar variáveis socioeconômicas, ambientais e novas fontes de dados para refinar os modelos.

A revisão evidencia a relevância crescente de métodos supervisionados, principalmente os baseados em árvores, como a Floresta Aleatória, na construção de modelos preditivos robustos. Desafios comuns incluem baixa qualidade e desatualização dos dados, além da dificuldade de acesso a imagens de alta resolução. Esses obstáculos refletem a realidade de muitos países do Sul Global, incluindo o Brasil.

Ainda assim, os estudos analisados oferecem subsídios metodológicos sólidos para aplicação de aprendizado de máquina na desagregação do déficit habitacional em escala intramunicipal. Ao reunir e sistematizar essas experiências, esta pesquisa fortalece o campo da habitação e aponta caminhos concretos para diagnósticos mais precisos e sensíveis às desigualdades territoriais.

4 CONSIDERAÇÕES FINAIS

A análise bibliométrica conduzida na Etapa 1 evidenciou o crescimento das publicações que aplicam inteligência artificial e mineração de dados a informações censitárias. Observou-se a consolidação de um campo interdisciplinar em expansão, especialmente a partir de 2019, impulsionado no contexto da pandemia de Covid-19. A identificação dos países com maior produção - com destaque para os Estados Unidos, bem como a caracterização dos tipos de documentos publicados e das principais áreas do conhecimento envolvidas, permitiu construir um panorama abrangente das pesquisas. A centralidade de termos como *“Machine learning”*, *“Remote sensing”* e *“Deep learning”*, e suas conexões com questões urbanas e territoriais, confirma a tendência consolidada de aplicação dessas tecnologias a problemas sociais complexos. Esse cenário oferece uma base teórica e metodológica robusta para o avanço das investigações sobre déficit habitacional e vulnerabilidades urbanas.

A revisão sistemática evidenciou forte dependência de dados censitários, frequentemente desatualizados e insuficientes para a complexidade das análises. Essa limitação demanda o uso de fontes complementares, muitas vezes de difícil acesso ou com alto custo, como imagens de satélite. Em países do Sul Global, a escassez de recursos financeiros e tecnológicos agrava ainda mais essas dificuldades, limitando o alcance e a qualidade das análises.

Nos estudos que utilizaram bases fotográficas ou imagens de satélite, destacaram-se desafios relacionados à resolução e à qualidade dos dados. A ausência de equipamentos adequados e o alto custo de imagens de alta definição dificultam a validação dos modelos, especialmente em contextos urbanos periféricos ou informais.

Entre os métodos de aprendizado de máquina identificados, os algoritmos baseados em árvores - com destaque para a Floresta Aleatória - foram os mais recorrentes, principalmente em estudos voltados à caracterização de favelas e áreas de vulnerabilidade. No entanto, não se observou uma convergência metodológica entre os artigos, devido às variações nos tipos de dados, nos objetivos das pesquisas e nos contextos territoriais.

A revisão sistemática demonstrou que, apesar dos desafios, o uso de aprendizado de máquina na análise do déficit habitacional apresenta grande potencial. Os estudos evidenciam a capacidade dessas técnicas para inferir padrões espaciais, classificar assentamentos precários e identificar fatores socioeconômicos relacionados à moradia inadequada. Ao reunir e analisar criticamente essas experiências, este artigo consolida uma base metodológica que pode orientar futuras investigações aplicadas ao contexto brasileiro, especialmente para desagregar o déficit habitacional em escalas territoriais mais detalhadas a partir de dados censitários.

5 AGRADECIMENTOS

Este trabalho contou com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001, e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - processo nº 2023/11899-6, cujos auxílios foram fundamentais para o desenvolvimento desta pesquisa.

REFERÊNCIAS

- ABARCA-ALVAREZ, F. J.; REINOSO-BELLIDO, R.; CAMPOS-SÁNCHEZ, F. S.. Decision model for predicting social vulnerability using artificial intelligence. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 575, 2019. DOI: <https://doi.org/10.3390/ijgi8120575>
- ABASCAL, A.; GEORGANOS, S.; KUFFER, M.; VANHUYSE, S.; THOMSON, D.; WANG, J.; MANYASI, L.; OTUNGA, D. M.; OCHIENG, B.; OCHIENG, T.; KLINNERT, J.; WOLFF, E.. Making urban slum population visible: citizens and satellites to reinforce slum censuses. In: **Urban inequalities from space: Earth observation applications in the majority world**. Cham: Springer International Publishing, 2024. p. 287-302. DOI: https://doi.org/10.1007/978-3-031-49183-2_14
- ARRIAGADA LUCO, C.. **El déficit habitacional en Brasil y México y sus dos megaciudades globales: estudio con los censos de 1990 y 2000**. Naciones Unidas, CEPAL, 2005.
- ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE. Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018. Proceedings. Palo Alto: AAAI Press, 2018. Disponível em: <https://ojs.aaai.org/index.php/AAAI/issue/view/301>. Acesso em: 23 maio 2025.
- BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. New York: springer, 2006. Disponível em: <https://link.springer.com/book/10.1007/978-0-387-45528-0>. Acesso em: 21 maio 2025.
- DE MEDEIROS, J. M. G.; VITORIANO, M. A. V.. A evolução da bibliometria e sua interdisciplinaridade na produção científica brasileira. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 13, n. 3, p. 491-503, 2015.
- EL MOUDDEN, T.; DAHMANI, R.; AMNAL, M.; FORA, A. A.. Slum image detection and localization using transfer learning: a case study in Northern Morocco. **International Journal of Electrical and Computer Engineering (IJECE)**, v. 13, n. 3, p. 3299-3310, 2023. DOI: <http://doi.org/10.11591/ijece.v13i3.pp3299-3310>
- ELLEGAARD, Ole; WALLIN, Johan A. The bibliometric analysis of scholarly production: How great is the impact?. **Scientometrics**, v. 105, n. 3, p. 1809-1831, 2015.
- ENGSTROM, R.; HARRISON, R.; MANN, M.; FLETCHER, A.. Evaluating the relationship between contextual features derived from very high spatial resolution imagery and urban attributes: A case study in Sri Lanka. In: **2019 Joint Urban Remote Sensing Event (JURSE)**. IEEE, 2019. p. 1-4. DOI: 10.1109/JURSE.2019.8809041
- ENGSTROM, R.; PAVELESKU, D.; TANAKA, T.; WAMBILE, A.. Mapping poverty and slums using multiple methodologies in Accra, Ghana. In: **2019 Joint Urban Remote Sensing Event (JURSE)**. IEEE, 2019. p. 1-4. DOI: 10.1109/JURSE.2019.8809052.
- GENEVOIS, M. L. B. P.; COSTA, O. V.. Carência habitacional e déficit de moradias: questões metodológicas. **São Paulo em perspectiva**, v. 15, p. 73-84, 2001.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge: MIT press, 2016. Disponível em: <https://www.deeplearningbook.org/>. Acesso em: 21 maio 2025.
- HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques. 3rd Edition**, University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University, 2011. Disponível em: <https://www.elsevier.com/books/data-mining/han/978-0-12-381479-1>. Acesso em: 21 maio 2025.
- KHALIL, M. A.; FATMI, M. R.; ORVIN, M.. Developing and microsimulating demographic dynamics for an integrated urban model: a comparison between logistic regression and machine learning techniques. **Transportation**, p. 1-35, 2024.
- KUFFER, M.; PERSELLO, C.; PFEFFER, K.; SLIUZAS, R.; RAO, V.. Do we underestimate the global slum population?. In: **2019 Joint Urban Remote Sensing Event (JURSE)**. IEEE, 2019. p. 1-4. DOI: 10.1109/JURSE.2019.8809066
- LEARNING, Mitchell Tom Machine. McGraw-Hill New York. 1997. Disponível em: <https://www.cs.cmu.edu/~tom/mlbook.html>. Acesso em: 21 maio 2025.

LYNGE, H.; VISAGIE, J.; SCHEBA, A.; TUROK, I.; EVERATT, D.; ABRAHAMS, C.. Developing neighbourhood typologies and understanding urban inequality: a data-driven approach. **Regional studies, regional science**, v. 9, n. 1, p. 618-640, 2022. DOI: 10.1080/21681376.2022.2132180

MAHABIR, R.; AGOURIS, P.; STEFANIDIS, A.; CROITORU, A.; CROOKS, A. T.. Detecting and mapping slums using open data: A case study in Kenya. **International Journal of Digital Earth**, v. 13, n. 6, p. 683-707, 2018. DOI: 10.1080/17538947.2018.1554010.

MAGNABOSCO, A. L.; CUNHA, P. H. F.; GARCIA, F.. Metodologias de mensuração do déficit habitacional no Brasil: uma comparação conceitual e empírica—2001 a 2009. **Pesquisa & Debate Revista do Programa de Estudos Pós-Graduados em Economia Política**, v. 23, n. 2 (42), 2012. Disponível em: <https://revistas.pucsp.br/index.php/rpe/article/view/13072>.

MAZZEI, M.; PALMA, A. L.. A Decision support system for the analysis of mobility. In: **International Conference on Computational Science and Its Applications**. Cham: Springer International Publishing, 2015. p. 390-402. DOI: https://doi.org/10.1007/978-3-319-21470-2_28

MERODIO GÓMEZ, P.; JUAREZ CARRILLO, O. J.; KUFFER, M.; THOMSON, D. R.; OLARTE QUIROZ, J. L.; VILLASEÑOR GARCÍA, E.; VANHUYSE, S.; ABASCAL, Á.; OLUOCH, I.; NAGENBORG, M.; PERSELLO, C.; BRITO, P. L. Earth observations and statistics: Unlocking sociodemographic knowledge through the power of satellite images. **Sustainability**, v. 13, n. 22, p. 12640, 2021. DOI: <https://doi.org/10.3390/su132212640>

PEDRASSOLI, J. C.; GOMES, J. G.; MELO, B. M.; SANTOS JUNIOR, E. R.; JUSTINIANO, E. F.; KAWAKUBO, F. S.; FANTIN, M.; MARTINES, M. R.; MORATO, R. G.. Reconstructing 36 Years of Spatiotemporal Dynamics of Slums in Brazil by Integrating EO and Census Data. In: **Urban Inequalities from Space: Earth Observation Applications in the Majority World**. Cham: Springer International Publishing, 2024. p. 191-215. DOI: https://doi.org/10.1007/978-3-031-49183-2_10

RUSSELL, Stuart; NORVIG Peter. Artificial Intelligence: A Modern Approach (4th ed.). Pearson, 2020. Disponível em: <https://aima.cs.berkeley.edu/>. Acesso em: 21 maio 2025.

SAGI, A.; GAL, A.; BROITMAN, D.; CZAMANSKI, D.. An unsupervised machine learning approach to the spatial analysis of urban systems through neighbourhoods' dynamics. **Land Use Policy**, v. 144, p. 107235, 2024. DOI: <https://doi.org/10.1016/j.landusepol.2024.107235>

SANTOS, E. C.. Ensaios e discussões sobre o déficit habitacional no Brasil. **Belo Horizonte: Fundação João Pinheiro**, 2022.

SCIMAGO LAB. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. United States. Disponível em: <https://www.scimagojr.com/journalsearch.php?q=21100896173&tip=sid&clean=0>. Acesso em: 23 maio 2025.

SILVA, F. X.; BARBOSA, E. P.; MENDES, V. R.; SANTOS, A. L. F. dos. Três tipos de estudos de revisão nas pesquisas educacionais: caracterização e análise. **SciELO Preprints**, 2021. Disponível em: <https://doi.org/10.1590/SciELOPreprints.2897>. Acesso em: 29 jul. 2025.

STACY, C.; Davis, C.; Freemark, Y. S.; Lo, L.; MacDonald, G.; Zheng, V.; Pendall, R. Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?. **Urban Studies**, v. 60, n. 14, p. 2919-2940, 2023. DOI: <https://doi-org.ez31.periodicos.capes.gov.br/10.1177/00420980231159500>

TADESSE, B.; KHANAM F.; AHMMED F.; LIU X.; ISLAM M.; KIM D.; KANG S.; IM J.; CHOWDHURY F.; AHMED T.; AZIZ A.; HOQUE M.; PARK J.; PAK G.; JEON H.; ZAMAN K.; KHAN A.; KIM J.; MARKS F.; QADRI F.; CLEMENS J.. Association among household water, sanitation, and hygiene (WASH) status and typhoid risk in urban slums: Prospective cohort study in Bangladesh. **JMIR public health and surveillance**, v. 9, n. 1, p. e41207, 2023. DOI: 10.2196/41207

THOMSON, D. R.; KUFFER, M.; BOO, G.; HATI, B.; GRIPPA, T.; ELSEY, H.; LINARD, C.; MAHABIR, R.; KYOBUTUNGI, C.; MAVITI, J.; MWANIKI, D.; NDUGWA, R.; MAKAU, J.; SLIUZAS, R.; CHERUIYOT, S.; NYAMBUGA, K.; MBOGA, N.; KIMANI, N. W.; DE ALBUQUERQUE, J. P.; KABARIA, C.. Need for an integrated deprived area “slum” mapping system (IDEAMAPS) in low-and middle-income countries (LMICs). **Social Sciences**, v. 9, n. 5, p. 80, 2020. DOI: <https://doi.org/10.3390/socsci9050080>

Revista Científica ANAP Brasil

ISSN 1984-3240 – Volume 18, número 46, 2025

VANHUYSE, S. KUFFER, M.; GEORGANOS, S.; WANG, J.; ABASCAL, A.; GRIPPA, T.; WOLFF, E.. Putting the Invisible on the Map: Low-Cost Earth Observation for Mapping and Characterizing Deprived Urban Areas (Slums). In: **Urban Inequalities from Space: Earth Observation Applications in the Majority World**. Cham: Springer International Publishing, 2024. p. 119-137. DOI: https://doi.org/10.1007/978-3-031-49183-2_7

WANG, S.; CAI, W.; SUN, Q.; MARTIN, C.; TEWARI, S.; HURLEY, J.; AMATI, M.; DUCKHAM, M.; CHOY, S.. Landscape of multiculturalism in Australia: Tracking ethnic diversity and its relation with neighbourhood features in 2001–2021. **Applied geography**, v. 160, p. 103114, 2023. DOI: <https://doi.org/10.1016/j.apgeog.2023.103114>

CONTRIBUIÇÃO DE CADA AUTOR

- **Concepção e Design do Estudo:** Lina, Pedro, Leonardo, Guilherme, Priscila, Tatiane, Elza, Rafael.
- **Curadoria de Dados:** Lina, Pedro, Leonardo, Guilherme, Priscila, Tatiane.
- **Análise Formal:** Lina, Pedro, Leonardo, Guilherme.
- **Aquisição de Financiamento:** Priscila, Tatiane, Elza, Rafael.
- **Investigação:** Lina, Pedro, Leonardo, Guilherme.
- **Metodologia:** Priscila, Tatiane.
- **Redação - Rascunho Inicial:** Lina, Pedro, Leonardo, Guilherme.
- **Redação - Revisão Crítica:** Priscila, Tatiane, Elza, Rafael.
- **Revisão e Edição Final:** Leonardo, Priscila.
- **Supervisão:** Priscila, Tatiane, Elza, Rafael.

DECLARAÇÃO DE CONFLITOS DE INTERESSE

Nós, Lina Sun Young Park, Pedro Henrique Lopes de Almeida, Leonardo Cavalcante da Silva, Guilherme Henrique Carvalho, Priscila Kauana Barelli Forcel, Tatiane Ferreira Olivatto, Rafael de Paula Garcia e Elza Luli Miyasaka, declaramos que o manuscrito intitulado "**O USO DE APRENDIZADO DE MÁQUINA NA ESTIMATIVA DO DÉFICIT HABITACIONAL: UMA REVISÃO SISTEMÁTICA DA LITERATURA**":

1. **Vínculos Financeiros:** Não possui vínculos financeiros que possam influenciar os resultados ou interpretação do trabalho. Este trabalho contou com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001, e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - processo nº 2023/11899-6.
2. **Relações Profissionais:** Não possui relações profissionais que possam impactar na análise, interpretação ou apresentação dos resultados. Nenhuma relação profissional relevante ao conteúdo deste manuscrito foi estabelecida.
3. **Conflitos Pessoais:** Não possui conflitos de interesse pessoais relacionados ao conteúdo do manuscrito. Nenhum conflito pessoal relacionado ao conteúdo foi identificado.