

Utilização de métodos de Machine Learning e BIM para executar previsões de volumes de Resíduos da Construção e Demolição (RCD)

DOI: <http://dx.doi.org/10.55449/conresol.8.25.VII-018>

Jorge González Fariás (*), Laryssa Franco de Carvalho Willcox, Alex Ximenes Naves, Assed Naked Haddad

* Universidade Federal do Rio de Janeiro, jorgegonzalez@poli.ufrj.br

RESUMO

Com cada vez mais resíduos sólidos no mundo, e estratégias de sustentabilidades mais populares e sendo aplicadas, o correto gerenciamento dos Resíduos da Construção e Demolição (RCD) faz-se necessário na indústria da construção. Este trabalho tem como objetivo propor um método de prever o volume de RCD que pode surgir de uma edificação na sua fase de demolição ao utilizar modelos de Machine Learning (ML) alimentados por bancos de dados de quantitativos de material obtidos de modelos BIM de edificações. Para isso, foi feita uma revisão bibliográfica simples nos bancos de dados da Scopus e Google Scholar para entender quais são os modelos de Machine Learning mais apropriados para fazer previsões de volume de RCD; para determinar os coeficientes de geração de volumes de RCD; e para entender como é feito o transpasso de informações do ambiente BIM até o ambiente Python de programação. Foi modelada uma edificação utilizando o software Autodesk Revit para entender como podem ser criadas as tabelas de quantitativos; e foi utilizado o ambiente de programação Python para aplicar os modelos de ML nas previsões. Como resultado, conseguiu-se propor o método de previsão de volumes de RCD utilizando esta ligação BIM-ML, que consta de três partes fundamentais: construção do banco de dados; geração dos modelos de previsão no ambiente Python; e avaliação de desempenho dos modelos de previsão. O processo de gerenciamento de resíduos necessita de estimativas precisas da quantidade de RCD para otimizar sua coleta, reciclagem e reutilização. Além disso, a indústria da construção deve conhecer estes volumes de resíduos para sua posterior redução. Como contribuição desta pesquisa temos que a aplicação de algoritmos de ML pode atender a essa demanda, devido à sua capacidade de aprendizagem; a principal inovação consiste em aperfeiçoar a utilização destes modelos para estimar com maior precisão os volumes de geração de RCD na indústria da construção civil.

PALAVRAS-CHAVE: Aprendizado de Máquinas, BIM, RCD, Previsão, Construção.

ABSTRACT

With more and more solid waste in the world, and sustainability strategies becoming more popular and being applied, the correct management of Construction and Demolition Waste (CDW) is necessary in the construction industry. This work aims to propose a method to predict the volume of CDW that may arise from a building in its demolition phase by using Machine Learning (ML) models fed by quantitative databases of material obtained from BIM models of buildings. For this, a simple bibliographic review was carried out in the Scopus and Google Scholar databases to understand which Machine Learning models are most suitable for making variations in CDW volume; to determine the coefficients for generating CDW volumes; and to understand how the transfer of information from the BIM environment to the Python programming environment is done. A building was modeled using Autodesk Revit software to understand how the quantitative tables can be created; and the Python programming environment was used to apply the ML models to the variations. As a result, we were able to propose a method for predicting CDW volumes using this BIM-ML interface, which consists of three fundamental parts: building the database; generating prediction models in the Python environment; and evaluating the performance of the prediction models. The waste management process requires accurate estimates of the amount of CDW to improve its collection, recycling, and reuse. In addition, the construction industry must know these waste volumes for their subsequent reduction. As a contribution of this research, we have that an application of ML algorithms can meet this demand, due to its learning capacity; the main innovation consists in improving the use of these models to more accurately estimate CDW generation volumes in the construction industry.

KEY WORDS: Machine Learning, BIM, RCD, Previsions, Construction.

INTRODUÇÃO

O crescimento populacional e a aceleração da urbanização levaram a um aumento acentuado na produção de resíduos sólidos urbanos, levando aos pesquisadores a procurar tecnologias avançadas para resolver esse problema. Com a rápida

urbanização, desenvolvimento econômico e crescimento populacional, a quantidade de resíduos sólidos urbanos aumentou significativamente, e é esperado que no mundo aumente para 2,2 bilhões de toneladas por ano entre 2025 e 2030. A grande quantidade de resíduos sólidos se tornou uma séria ameaça à cidade e ao ambiente ecológico ao redor, o que causa muitos problemas, como despejo ilegal e poluição (XIA et al., 2022).

O setor da construção, conhecido por seu consumo substancial de materiais, contribui significativamente para a produção industrial. Consequentemente, as matérias-primas são mais consumidas, enquanto uma quantidade significativa de resíduos é gerada devido à extensa destruição e reconstrução de edifícios. Isto gera um volume substancial de resíduos sólidos, muitas vezes destinados a aterros sanitários, causando poluição ambiental significativa (SIRIMEWAN et al., 2024).

Estes resíduos sólidos são chamados de resíduos de construção civil (RCC), um termo frequentemente usado de forma intercambiável com resíduos de construção e demolição (RCD), que são gerados por atividades de construção, renovação ou demolição. Compreende materiais inertes e não inertes, incluindo concreto, aço, chorume, madeira, vidro, tijolos, agregados, entulho e pedras, solo e areia, asfalto, plástico, gesso, cerâmica, papelão, tecido e matéria orgânica etc. (LU et al., 2021). Esses materiais são frequentemente misturados de forma heterogênea, tornando o processamento e o gerenciamento de resíduos mais desafiadores (SIRIMEWAN et al., 2024).

Portanto, o gerenciamento eficiente de RCD pode reduzir resíduos enviados para aterros, minimizar impactos ambientais e garantir o gerenciamento de recursos naturais. A reciclagem de resíduos e a recuperação de recursos são amplamente reconhecidas na economia circular como abordagens práticas para mitigar o impacto da geração de resíduos no meio ambiente (SIRIMEWAN et al., 2024).

A gestão de RCD são as atividades e ações necessárias desde o início até o descarte deles. Isso envolve coleta, transporte, processamento e descarte de lixo e monitoramento e controle da gestão de resíduos. A gestão de resíduos visa principalmente reduzir a quantidade de material não utilizado e prevenir potenciais riscos à saúde e ao meio ambiente, desde a coleta de resíduos até a reciclagem e o monitoramento (CHEN, 2022).

Para regiões onde os dados de geração de resíduos são regularmente coletados e divulgados por sistemas de registro oficiais, a estimativa é desnecessária. No entanto, regiões emergentes, como no Brasil, geralmente não têm esses sistemas em vigor. É reconhecido que dados confiáveis sobre a geração de resíduos de construção são um pré-requisito para qualquer esforço de gerenciamento de resíduos baseado em evidências, mas tais dados permanecem escassos em muitas economias em desenvolvimento devido aos seus sistemas de registro rudimentares (LU et al., 2021).

Há diversos de estudos sobre quantificação de resíduos sólidos na ausência de dados diretos, onde outras estatísticas ou sinais, como população, crescimento econômico, despesas de construção, renda, decadência urbana e níveis de reciclagem de resíduos, entre outros fatores demográficos, são analisados para informar a geração de resíduos (GU et al., 2015; KANNANGARA et al., 2018; SINGH; UPPALURI, 2023). Tais estudos frequentemente adotam algoritmos complicados para estimar a geração de resíduos, mas podem exibir overfitting onde os modelos relatam resultados de ajuste próximo ou exato em conjuntos de dados de treinamento, mas resultados ruins em conjuntos de dados de teste. Além disso, poucos estudos foram além da definição dos fatores que podem prever a geração de RCD para entender o quanto cada fator contribui para a previsão (LU et al., 2021).

Para fazer as previsões, fatores relacionados à construção incluem: volume total de materiais de construção, área de piso pronta em edificações em construção, área de piso em edificações concluídas, entre outros. Embora seja impossível obter a quantidade direta de geração de resíduos de RCD, esses fatores com disponibilidade razoável de dados e análises adequadas podem produzir uma estimativa satisfatória da geração de resíduos de RCD (LU et al., 2021).

Outros autores têm definido outras variáveis relacionadas à construção incluem o espaço de piso de edifícios concluídos (novos e demolidos), a produção total de resíduos do processo de construção, e outras características relacionadas ao espaço da construção (por exemplo, áreas vagas). Embora seja desafiador quantificar precisamente a geração de RCD, a disponibilidade de registros apropriados e suficientes para essas variáveis pode produzir uma estimativa confiável da geração de RCD (MAGED; ELSHABOURY; AKANBI, 2024). Com os fatores potenciais conhecidos, vários métodos foram propostos para estimar a geração de resíduos de construção, alguns deles são: visita ao local, cálculo da taxa de geração, análise de ciclo de vida, acumulação do sistema de classificação, modelagem de fatores, e outros (por exemplo, estimativa automatizada baseada em dados do modelo BIM) (LU et al., 2021).

O método de visita ao local exige que o investigador conduza pesquisas no local, incluindo medição direta por meio do levantamento do peso ou volume e medição indireta pela adoção de outros indicadores facilmente acessíveis, como tíquetes de transporte. Para o método de taxa de geração, o volume total de resíduos pode ser calculado multiplicando a quantidade de uma unidade específica por sua taxa de geração correspondente, por exemplo, cálculo baseado em área. O método de análise de ciclo de vida assume que todos os edifícios devem ser desmontados após um certo período de vida útil e os resíduos de RCD podem ser deduzidos do cálculo da soma da massa a ser removida na expiração. O método de acumulação do sistema de classificação combina o método de taxa de geração com uma classificação de resíduos para quantificar cada material específico. O método de modelagem de fatores projeta modelos de previsão com base em variáveis acessíveis, como modelos de regressão linear (LU et al., 2021).

Algoritmos de aprendizado de máquina (ML) são bons em modelar processos não lineares complexos e têm sido gradualmente adotados para promover o gerenciamento de resíduos sólidos urbanos e ajudar o desenvolvimento sustentável do meio ambiente nos últimos anos (BADUGE et al., 2022; XIA et al., 2022). ML usa dados para "treinar" e aprende como concluir tarefas a partir dos dados por meio de vários algoritmos. Esses algoritmos tentam extrair informações ocultas de uma grande quantidade de dados históricos e usá-los para regressão ou classificação (BADUGE et al., 2022; XIA et al., 2022).

O aprendizado de máquina (ML) desempenha um papel crucial nas técnicas de triagem automatizadas, o que leva a uma maior precisão e melhor qualidade de separação de resíduos, aumentando toda a eficiência da gestão de resíduos, melhorando a coleta de resíduos e avançando o programa (CHEN, 2022).

Em relação com os modelos de aprendizado de máquina (ML), vários algoritmos e modelos estão surgindo para estimar a geração de resíduos, entre eles se encontram: regressão linear múltipla (MLR), árvore de decisão (DT), modelos cinza (GM) e redes neurais artificiais (ANN) (Lu et al., 2021). Também máquina de vetores de suporte (SVM), K-Nearest-Neighbor (KNN), floresta aleatória (RF) e sistema de inferência fuzzy de rede adaptável (ANFIS). Entre todos eles, ANN é o método mais comumente usados, mas seu desempenho é afetado por sobre ajuste de dados, mínimos locais e generalização ruim (XIA et al., 2022).

Em geral, todos os métodos para estimar a geração de resíduos de RCD têm seus próprios pontos fortes e fracos. Embora os modelos de ML possam prever com precisão a geração de RSU, o maior desafio é a falta de dados históricos, o que pode ser atribuído à gestão ineficaz de resíduos devido à não confiável previsão de volumes de RCD. Com dados suficientes, os métodos de ML podem atingir maior precisão na previsão da geração de RSU (XIA et al., 2022).

Também devido à insuficiência de dados das edificações, modelos de ML precisos relacionados a RCD são difíceis de construir. Além disso, quase todos os conjuntos de dados usados para treinamento não estão disponíveis publicamente e, sem um conjunto de dados de referência aberto, dificultarão a aplicação de algoritmos de ML para RCD (XIA et al., 2022). Atualmente, existem várias maneiras de resolver o problema de dados pequenos (por exemplo, na classificação de resíduos, técnicas de aumento de dados são usadas para aumentar os dados da amostra para melhor aplicar o algoritmo DL) (XIA et al., 2022).

Porém, faz-se necessária a utilização de dados reais e representativos para alimentar estes modelos e considerar as previsões como corretas. Por esta razão surge a seguinte pergunta de pesquisa: é possível criar uma previsão correta da estimativa de RCD que produzirá uma edificação na sua fase de demolição? Estes modelos de ML para realizar previsões de volumes de RCD resultam confiáveis?

De acordo com as pesquisas atuais, os modelos de ML conseguem emitir resultados de previsão corretos. Para isso, faz-se necessário selecionar o modelo de ML mais apropriado com a situação a ser prevista, e alimentar o modelo com um banco de dados sólido e representativo torna-se imprescindível; isto garante resultados certos e precisos para estimativas. Estes modelos de ML conseguem estimar volumes de RCD em edificações de forma fácil, adequada, e prática, tanto nas etapas de desenho, quanto em etapas posteriores do ciclo de vida das construções.

Esta pesquisa resulta em diversas implicações, para resolver problemas de RCD e promover a sustentabilidade. Em relação com as implicações teóricas, esta pesquisa contribui para enriquecer a implementação de ferramentas tecnológicas da Indústria 4.0, como é a utilização de modelos de ML; enquanto mais se utilizam estes algoritmos, mais se aperfeiçoam as técnicas, e mais dados são obtidos em prol de implementar soluções que permitam a reciclagem, reutilização e redução destes resíduos. Em relação com as implicações práticas, o aporte desta pesquisa baseia-se na

possibilidade de obter de forma fácil e prática, estimativas significativas de volumes de RCD para prédios, sempre que houver informações sobre a quantidade de materiais que constituem estes.

O restante do artigo fica organizado da seguinte forma: na seção 2 é mostrado o objetivo da pesquisa, na seção 3 explica-se o processo da pesquisa até a obtenção dos resultados incluindo os materiais e software utilizados, na seção 4 mostra os resultados dos modelos de previsão, e na seção 5 é possível ver as conclusões deste trabalho.

OBJETIVO

O objetivo desta pesquisa baseia-se em propor um método de utilização de dois modelos de Machine Learning, chamados Regressão Linear Múltipla e Árvores de Decisão, para prever a emissão de volumes de RCD provenientes de modelos BIM de edificações em 3D, e posteriormente propor como comparar qual deles é o que tem o melhor desempenho na previsão de volumes de RCD.

METODOLOGIA

Para a realização desta pesquisa, o processo foi dividido em três etapas: definição dos modelos 3D, pesquisa bibliográfica, e a proposta de aplicação dos métodos de Machine Learning. Para melhor compreensão do processo, este pode ser visualizado na Figura 1. Na etapa 1, os resultados das previsões que poderão ser obtidos no futuro ao começar pela modelagem das edificações. Como único exemplo, se tem a modelagem de uma edificação de 9 andares no software Autodesk Revit, como pode ser visualizado na Figura 2. Qualquer modelo BIM de edificação pode ser utilizado pois o interessante desta metodologia é a capacidade de gerar quantitativos de material automaticamente. Os materiais da edificação que podem ser incluídos são mostrados na Figura 3, que é uma captura de uma tabela que serve como exemplo; eles incluem tijolo, cerâmico, concreto, gesso, tinta, aço, vidro, entre outros.

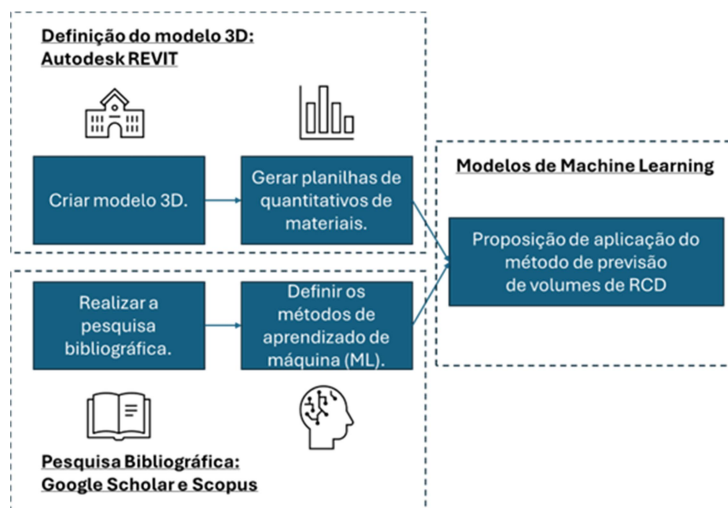


Figura 1: Fluxograma da pesquisa. Fonte: Própria.

Dentro do software Autodesk Revit em particular, os quantitativos de materiais podem ser gerados automaticamente ao mostrar “Tabelas”, e selecionar as opções de “Fazer novo levantamento de material”, selecionar “Todos os Sistemas”, da fase já “Construída”, e selecionar as opções das características de material, como nome, tipo, volume, peso, entre outros. Depois, a tabela de quantitativos de material é exportada para um arquivo em formato .csv, que pode ser lido dentro do Microsoft Excel para criar um arquivo em formato .xlsx. Uma vez obtidos os quantitativos de materiais existentes neste último formato, estes são passados para o ambiente Python, onde podem ser feitas previsões de quanto de todo o material virará resíduo de demolição.

Na etapa 2, consultaram-se os bancos de dados bibliográficos da Scopus e de Google Scholar, considerando a produção acadêmica dos últimos 4 anos (2021 – 2024), utilizando as fórmulas de pesquisa “previsão resíduos da Construção Machine Learning”, “Árvores de Decisão”, “Regressão Linear Múltipla”, “Machine Learning Prevision Construction and Demolition Waste”, “Machine Learning previsão entulho construção”, “Machine Learning previsão resíduos

construção”, “Machine Learning previsão resíduos construção civil”, e “Machine Learning previsão RCC RCD”. Foram considerados apenas resultados nas línguas inglesa e portuguesa. Só foram selecionados os trabalhos abertos, e relacionados com a utilização de modelos de ML na construção civil.

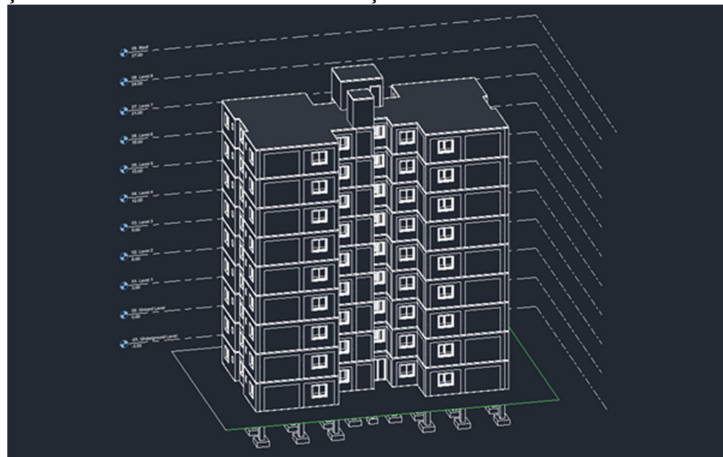


Figura 2: Edificação considerada no estudo. Fonte: Própria.

Levantamento de material				
Família e Tipo	Material: Nome	Material: Peso da unidade (kN/m³)	Material: Volume (m³)	Material: Área (m²)
Basic Wat: Wall 15cm	Ceramic Tile	19,6	208,14	2313
Basic Wat: Wall 15cm	Concrete CastIn Situ	23,6	92,51	4625
Basic Wat: Wall 15cm	Paint	0	4,63	4625
Basic Wat: Wall 15cm	Plaster	11	41,63	4625
Basic Wat: Wall 15cm Aço	Stainless Steel	79,2	10,83	72
Basic Wat: Wall 30cm	Air	0	150,53	2509
Basic Wat: Wall 30cm	Ceramic Tile	19,6	451,6	5018
Basic Wat: Wall 30cm	Concrete CastIn Situ	23,6	100,35	5018
Basic Wat: Wall 30cm	Paint	0	5,02	5018
Basic Wat: Wall 30cm	Plaster	11	45,16	5018
Floor: Ceramic floor 5cm	Ceramic Tile	19,6	18,94	2691
Floor: Ceramic floor 5cm	Concrete CastIn Situ	23,6	107,94	2691
Floor: Ceramic floor 5cm	Concrete Lightweight	17,3	8,07	2691
Floor: Concrete 10cm 25MPa	Concrete Cast-in-Place - C25	23,6	270,37	2704
M Concrete-Rectangular Beam: Beam 30 x 30	Concrete Cast-in-Place gray	23,6	7,9	114
M Concrete-Rectangular Beam: Beam 30 x 40	Concrete Cast-in-Place gray	23,6	173,88	2129
M Concrete-Rectangular Column: 30 x 45	Concrete Cast-in-Place gray	23,6	11,55	136
M Concrete-Rectangular Column: 45 x 60	Concrete Cast-in-Place gray	23,6	215,46	1827
M Fixed: Window 40 x 60	Glass	23,7	0,06	10
M Fixed: Window 40 x 60	Sash	0	0,71	58
M Footing-Rectangular: 100 x 80 x 50	Concrete Cast-in-Place gray	23,6	1,2	10
M Footing-Rectangular: 160 x 120 x 50	Concrete Cast-in-Place gray	23,6	26,88	186
M Single-Flush: Door 70 x 210	Door - Frame	0	1,08	115
M Single-Flush: Door 70 x 210	Door - Panel	0	4,2	181
M Single-Flush: Door 80 x 210	Door - Frame	0	2,47	263
M Single-Flush: Door 80 x 210	Door - Panel	0	10,8	461
M Single-Flush: Door 90 x 210	Door - Frame	0	0,74	79
M Single-Flush: Door 90 x 210	Door - Panel	0	3,57	151
Slider with Trim: Window 120 x 120	Glass	23,7	0,62	101
Slider with Trim: Window 120 x 120	Metal - Painted - Gray	72,5	2,94	285
Slider with Trim: Window 140 x 120	Glass	23,7	1,81	294
Slider with Trim: Window 140 x 120	Metal - Painted - Gray	72,5	7,77	753

Figura 3. Materiais da edificação que podem ser considerados. Fonte: Própria.

Diversos autores têm pesquisado sobre a utilização de modelos de previsão para determinar volumes de RCD nas construções, alguns deles chegando até a incluir conceitos de economia circular, outros comparando diversos modelos, e outros classificando os materiais. No Brasil, o tema ainda está nas suas primeiras fases de desenvolvimento, sugerindo a adoção de metodologias externas para adequá-las à realidade nacional. Alguns destes trabalhos, os mais relacionados com esta pesquisa, se mostram à continuação.

(CATARINA; DA LUZ; PEREZ PALHA, 2024) em 2024 investigaram de forma detalhada os RCD ao longo do ciclo de vida de uma edificação de 8 andares no Brasil. Para isso, utilizaram o software Autodesk Revit, a sua interface Dynamo, e a linguagem de programação Python, para desenvolver um algoritmo de quantificação. Eles consideraram apenas as etapas de construção e demolição do empreendimento, e os elementos estudados foram lajes de concreto, pintura, tijolos, azulejos, gesso, portas de madeira, forro e pisos.

(CHA; MOON; KIM, 2022) conduziram em 2022 uma pesquisa na Coreia utilizando diferentes algoritmos de ML (Support Vector Machine, Artificial Neural Networks, K-Nearest Neighbor, Linear Regression, e Random Forest) para desenvolver um modelo preditivo otimizado da taxa de geração de resíduos de demolição. Compararam os coeficientes de determinação e percentual de desvio, apontando que as Artificial Neural Networks e Support Vector Machine tiveram o melhor desempenho.

(LIN et al., 2023) pesquisaram em 2022 um método altamente eficiente para selecionar o RCD visando promover sistemas de reciclagem. O executaram ao classificar dez tipos de RCD usando dois modelos de aprendizado de máquina (C&DNet e ResNet) baseado em transferência de conhecimento e aprendizado cíclico. No final, resultou o modelo C&DNet com a maior precisão.

(CHA et al., 2023) em 2022 desenvolveram modelos predictivos híbridos de Machine Learning para melhorar o desempenho destes. Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Support Vector Machine Regression (SVMR), e Categorical Principal Components Analysis (CATPCA) foram analisados. Eles encontraram que o modelo com o melhor desempenho foi o híbrido de CATPCA-SVMR, identificando neste modelo uma ótima ferramenta para melhorar a previsão de volumes de RCD.

(SAKA et al., 2023) em 2023 procuraram quantificar resíduos de demolição usando Variable Modelling (VM) e Machine Learning (ML) considerando como banco de dados 2280 projetos. O modelo que utilizou XGBoost resultou o mais certo, ao relacionar a metodologia BIM com modelos de ML de uma forma otimizada. Com isso, foi possível prever volumes de RCD que podem ser utilizados dentro da economia circular, e melhorar a gestão destes resíduos.

(JAFARI; MOUSAVI, 2024) em 2024 desenvolveram em Irã um modelo de previsão mensal de RCD para essa cidade, usando Multilayer Perceptron Neural Network, Radial Basis Function Neural Network, Support Vector Machines, e Adaptive Neuro-Fuzzy Interference System. Este último resultou o modelo que melhor representou os valores observados de geração de RCD. Com isso, é possível empregar esta ferramenta para criação de políticas públicas e auxiliar o processo de tomada de decisão na gestão de RCD.

(SOLÍS-GUZMÁN et al., 2009) em 2009 desenvolveu um modelo de gestão de RCD num conjunto habitacional em Sevilha, Espanha, para prever volumes estimados deste material. Este modelo foi baseado nos registros de mais de 100 projetos de habitação. No final, estimou os coeficientes aplicáveis para volume de RCD previsto.

Nesta pesquisa, será proposta a utilização de dois modelos de Machine Learning chamados Regressão Linear Múltipla e Árvores de Decisão; porém, pode ser qualquer outro modelo de Machine Learning que consiga fazer previsões. O modelo de Regressão Linear Múltipla foi escolhido pois este permite avaliar mais de uma variável independente que pode influenciar a variável dependente, emite em certas circunstâncias previsões mais precisas, realistas e complexas, o que contribui a melhorias na gestão de processos. O modelo de Árvore de Decisão foi escolhido devido a que permite que mais de uma variável possa ser aplicada no processo de análise da máquina, e como todo o modelo conta com vários ramos, a árvore pode gerar resultados bem diferentes para cada variável, o que permite uma perspectiva muito mais ampla das possibilidades de previsão.

A regressão linear múltipla avalia a relação entre mais de duas variáveis para assim entender o efeito de uma sobre as outras. Esta técnica surge pela necessidade de obter melhores modelos que consigam explicar a variável dependente, quer dizer, quais são todas as variáveis explicativas para o modelo. Esta técnica é bastante difundida pela literatura acadêmica devido à facilidade e simplicidade de prever os dados, e versatilidade na modelagem das funções (FELIPE; MOURA; SOPCHAKI, 2023; TEIXEIRA; DIONÍSIO BASÍLIO, 2023).

A árvore de decisão consiste em uma técnica de aprendizado de máquina supervisionada em que são selecionadas variáveis preditoras e uma variável-alvo. Elas são, dentre outros modelos de aprendizagem supervisionada, modelos preditivos com maior precisão e de simples compreensão (FREIRE; SOUZA FILHO, 2022; MOTA GONÇALVES, 2022). Ao utilizá-las como classificador ou regressor, esse modelo permite entender o caminho gerador da predição via inspeção visual (SILVA SANTOS, 2022).

Na etapa 3, o fluxo de trabalho de aplicação dos métodos tradicionais de ML geralmente consiste em três etapas: (1) processamento de dados e extração de recursos, (2) escolha dos algoritmos e parâmetros de ML adequados e (3) teste e avaliação de desempenho. O pré-processamento de dados é uma etapa importante a se fazer, pois se os dados são processados corretamente tem um grande impacto nos resultados de treinamento e previsão. Além disso, a seleção apropriada das variáveis de entrada pode influenciar o desempenho do processo de aprendizagem e a robustez dos modelos de ML. A seleção adequada do modelo melhoraria o desempenho e diminuiria a complexidade da modelagem (CHEN, 2022).

Uma etapa importante na avaliação da eficácia dos modelos de ML é conduzir experimentos de comparação para validar a qualidade e desempenho dos dados obtidos ao avaliar (XIA et al., 2022). Os mais comuns são o valor do Erro

Médio Absoluto (MAE), Erro Médio Quadrático (MSE), e o Coeficiente de Determinação (R^2), os quais são propostos para analisar o desempenho dos modelos desta pesquisa.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - p_i| \quad \text{Equação (1)}$$

MAE = Erro médio absoluto. n = número de amostras. Y_i = valor observado para cada amostra. p_i = valor previsto para cada amostra.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad \text{Equação (2)}$$

MSE = Erro médio quadrático. n = número de amostras. Y_i = valores observados. $\hat{Y}_i$ = valores esperados.

$$R^2 = 1 - \frac{RSS}{TSS} \quad \text{Equação (3)}$$

R^2 = coeficiente de determinação. RSS = soma dos quadrados dos resíduos. TSS = soma total dos quadrados.

RESULTADOS

O resultado desta pesquisa baseia-se na proposição de método de previsão de volumes de RCD usando dois modelos de Machine Learning denominados Regressão Linear Múltipla e Árvores de Decisão. Por falta de um banco de dados sólido e representativo, não foi possível validar o modelo de previsão, porém, este é proposto nos próximos parágrafos.

O banco de dados deve ser construído considerando diversos modelos BIM de edificações. Os modelos de Machine Learning precisam de bancos de dados de mais de 500 registros de volumes de material para conseguir treinar os modelos e emitir resultados certos. Por esta razão, e a diferença do trabalho de (Solís Guzman et. Al., 2009), mais edificações deverão ser consideradas para obter mais volumes de resíduos que sirvam como parâmetro para as previsões.

Depois, no banco de dados geral de volumes, deverão ser aplicados os coeficientes de (Solís-Guzmán et al., 2009), para obter os valores de volume de RCD previsto, como pode ser visualizado na Tabela 1 a continuação. Estes coeficientes deverão multiplicar, dentro do Microsoft Excel, os quantitativos de cada registro de material para prever quanto de cada material é produzido por área construída. Uma vez obtidos os registros de mais de 500 valores de volume de RCD previsto, o banco de dados está pronto para realizar as previsões.

Tabela 1. Percentuais de volumes de RCD previstos de acordo com o material. Fonte: (Catarina et al., 2024; Solís-Guzmán et al., 2009).

Elemento	Volume de Demolição (%)
Laje de Concreto	34,24
Concreto Armado	10,26
Tijolo exterior	11,69
Tijolo Interior	4,31
Azulejo	1,53
Gesso	3,47
Piso	0,62
Teto	0,48
Pintura exterior	0,10
Pintura interior	0,26
Portas de madeira	0,68

Vidro	0,11
Caixonetes metálicos	0,26

Uma vez que o arquivo no Microsoft Excel, com já se encontra no ambiente Python, é estabelecido o modelo de Regressão Linear Múltipla que aplica algoritmos de Machine Learning, o qual baseia-se no script que se mostra na Figura 4. Mesma coisa acontece com o modelo de Árvore de Decisão, que segue a ordem do script como mostrado na Figura 5. Na sequência, poderão ser emitidos os gráficos e figuras das previsões de volume de RCD.

Após obter os resultados das duas funções dos modelos, a performance destes é avaliado ao comparar os resultados dos valores de erro médio absoluto, erro médio quadrático, e o coeficiente de determinação. Os valores de Erro Médio Absoluto e Erro Médio Quadrático, enquanto menores são, mais confiável é o modelo. Para os resultados de Coeficiente de Determinação, enquanto o valor deste estiver mais perto de 1, mais acertado será o modelo, e com isso, mais verdadeira será a previsão. Um fluxograma da proposta do modelo é mostrado à continuação na Figura 6, mostrando a série de passos que devem ser seguidos para obter resultados confiáveis nas previsões de volumes de RCD para cada metro quadrado construído.

Se comparado este método proposto, com outros métodos que utilizam modelos de Machine Learning, os procedimentos e as ferramentas são basicamente as mesmas: criar um banco de dados representativo, o suficientemente rico, treinar os bancos de dados, e gerar o modelo de previsão. Os trabalhos de (CHA; MOON; KIM, 2022; KANNANGARA et al., 2018; SINGH; UPPALURI, 2023) adicionam ou retiram passos intermediários, mas o princípio é o mesmo.

É importante ressaltar que a revisão do desempenho dos modelos é relevante para determinar a confiabilidade deles, determinando se estes realmente refletem um comportamento similar ao mostrado pelos casos de quantificação de volumes reais. Como exemplo, a pesquisas de (JAFARI; MOUSAVI, 2024) também avalia o valor do Coeficiente de Determinação como métrica; no caso (SINGH; UPPALURI, 2023) avaliou o Coeficiente de Determinação e o Erro Médio Absoluto; e (CHA; MOON; KIM, 2022), como neste estudo, avalia os resultados do MAE, MSE e R^2 para analisar o desempenho dos modelos.

```
# Multiple Linear Regression

# Importing the libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Importing the dataset
dataset = pd.read_csv('50_Startups.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
print(X)

# Encoding categorical data
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [3])], remainder='passthrough')
X = np.array(ct.fit_transform(X))
print(X)

# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

# Training the Multiple Linear Regression model on the Training set
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)

# Predicting the Test set results
y_pred = regressor.predict(X_test)
np.set_printoptions(precision=2)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))
```

Figura 4: Script do modelo de ML Regressão Linear Múltipla. Fonte: Própria.



```
# Decision Tree Regression

# Importing the libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Importing the dataset
dataset = pd.read_csv('RCD.csv')
X = dataset.iloc[:, 1:-1].values
y = dataset.iloc[:, -1].values

# Training the Decision Tree Regression model on the whole dataset
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X, y)

# Predicting a new result
regressor.predict([[6.5]])

# Visualising the Decision Tree Regression results (higher resolution)
X_grid = np.arange(min(X), max(X), 0.01)
X_grid = X_grid.reshape((len(X_grid), 1))
plt.scatter(X, y, color = 'red')
plt.plot(X_grid, regressor.predict(X_grid), color = 'blue')
plt.title('Truth or Bluff (Decision Tree Regression)')
plt.xlabel('Material')
plt.ylabel('Resíduos')
plt.show()
```

Figura 5: Script do modelo de ML Árvore de Decisão.

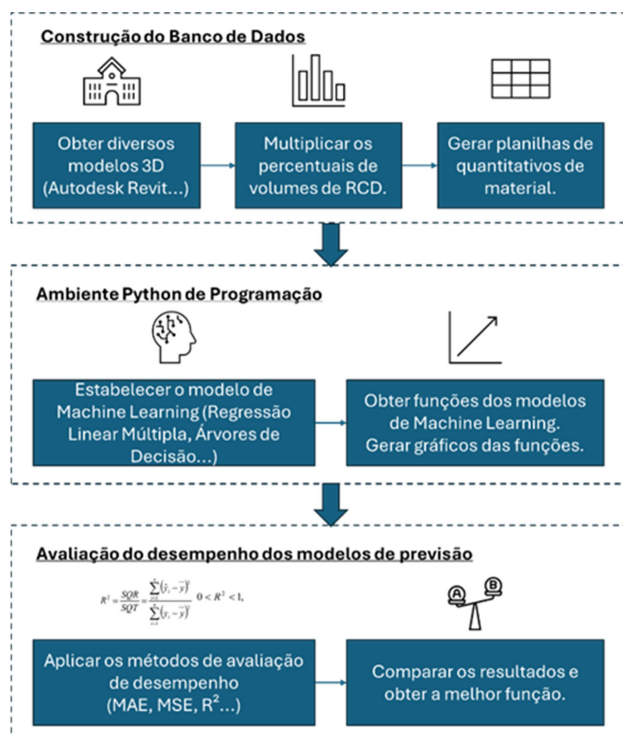


Figura 6: Proposta do modelo de previsão. Fonte: Própria.

CONCLUSÕES

Pelo fato de a quantidade de resíduos sólidos estarem aumentando hoje em dia, surge a necessidade de entender e prever qual o volume destes para conseguir realizar um melhor processo de gestão. Nas edificações, os volumes de resíduos da construção e demolição representam um problema para gerenciar de forma eficiente, procurando que estes que não causem dificuldades no canteiro de obras, ao mesmo tempo que se garante seu manejo sustentável.

Esta pesquisa conseguiu alcançar seu objetivo, que foi propor um método de previsão de volume de RCD ao utilizar modelos de Machine Learning e considerando os quantitativos de material de edificações de modelos BIM. Neste caso, os dois modelos de Machine Learning considerados no processo são Regressão Linear Múltipla e Árvores de Decisão porque são razoavelmente fáceis de utilizar, e consideram eficientemente as variáveis independentes que geram um efeito na variável dependente. Para os quantitativos de material, o software BIM proposto foi Autodesk Revit, devido à sua facilidade e popularidade de utilização na indústria da construção.

Porém, qualquer método de Machine Learning pode ser utilizado sempre que cumpra com a tarefa de realizar previsões, ao mesmo tempo que qualquer software BIM pode ser empregado. O mais relevante a ter em consideração é que o banco de dados que seja utilizado para treinar o modelo de ML e executar as previsões, deve ser sólido e representativo o suficiente para gerar resultados corretos.

Com esta pesquisa, a lacuna de conhecimento na ciência vai ficando preenchida aos poucos, pois as previsões de volumes de RCD utilizando modelos de ML são uma novidade que ainda é explorada e aperfeiçoada na bibliografia, especialmente nos dias atuais. Ao realizar estes trabalhos, diversas implicações positivas surgem e se implementam tanto na ciência quanto no âmbito profissional, pois os modelos de ML são mais bem treinados, mais dados são obtidos e considerados, e com isso, as previsões de volumes de RCD ficam mais certeiras.

Diversas limitações foram encontradas nesta pesquisa: a primeira foi não ter considerado a validação da metodologia, isso por causa de ter um banco de dados o suficientemente sólido e nutrido; para isso seriam necessários os dados de mais de 500 modelos 3D de edificações, que conseguissem treinar os modelos de ML. Outra limitação foi ter considerado apenas uma fonte de coeficientes de volumes de demolição, que multiplicam os dados dos volumes de material provenientes dos modelos de edificações. Por estas razões, recomenda-se aplicar esta metodologia em futuros trabalhos, contemplando mais dados de edificações, e mais dados de coeficientes, para garantir modelos de ML mais bem treinados, capazes de oferecer previsões corretas.

Ao final do texto deverão aparecer as **REFERÊNCIAS BIBLIOGRÁFICAS**, utilizando fonte Times New Roman, **tamanho 10**, alinhamento de parágrafo justificado e espaçamento de parágrafo de 6 pontos (depois). No início de cada item bibliográfico deverá ser usado um marcador de numeração crescente, com **Posição do Número**: esquerdo, **Alinhado em**: 0 cm e distância de recuo de texto **Recuar em**: 0,6 cm. Inserir no máximo 15 Referências Bibliográficas. Exemplo:

2 linhas em branco, fonte Times New Roman, tamanho 10

REFERÊNCIAS BIBLIOGRÁFICAS

1. BADUGE, S. K. et al. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. **Automation in Construction**, v. 141, p. 104440, set. 2022.
2. CATARINA, F.; DA LUZ, R.; PEREZ PALHA, R. Uso do sistema BIM na quantificação dos resíduos gerados na construção civil: utilização do Revit e Dynamo aliados a linguagens de programação Python. **ENTAC**, 2024.
3. CHA, G. W. et al. Development of Machine Learning Model for Prediction of Demolition Waste Generation Rate of Buildings in Redevelopment Areas. **International Journal of Environmental Research and Public Health**, v. 20, n. 1, 1 jan. 2023.
4. CHA, G. W.; MOON, H. J.; KIM, Y. C. A hybrid machine-learning model for predicting the waste generation rate of building demolition projects. **Journal of Cleaner Production**, v. 375, 15 nov. 2022.
5. CHEN, X. **Machine learning approach for a circular economy with waste recycling in smart cities**. **Energy Reports** Elsevier Ltd, , 1 nov. 2022.

6. FELIPE, L.; MOURA, S.; SOPCHAKI, C. H. DESEMPENHO DO MÉTODO ESTATÍSTICO DE REGRESSÃO LINEAR MÚLTIPLA NO PREENCHIMENTO DE FALHAS EM DADOS PLUVIOMÉTRICOS. **Revista Contexto Geográfico**, p. 1–13, 10 set. 2023.
7. FREIRE, L. L.; SOUZA FILHO, F. DE A. Interrelationships between eutrophication predictors in Brazilian semiarid reservoirs: how to measure? A decision tree machine learning application. **Engenharia Sanitaria e Ambiental**, v. 27, n. 6, p. 1155–1165, 2022.
8. GU, B. et al. Characterization, quantification and management of household solid waste: A case study in China. **Resources, Conservation and Recycling**, v. 98, p. 67–75, 1 maio 2015.
9. JAFARI, M.; MOUSAVI, E. Machine learning-based prediction of construction and demolition waste generation in developing countries: a case study. **Environmental Science and Pollution Research**, 2024.
10. KANNANGARA, M. et al. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. **Waste Management**, v. 74, p. 3–15, 1 abr. 2018.
11. LIN, K. et al. Applying machine learning to fine classify construction and demolition waste based on deep residual network and knowledge transfer. **Environment, Development and Sustainability**, v. 25, n. 8, p. 8819–8836, 1 ago. 2023.
12. LU, W. et al. Estimating construction waste generation in the Greater Bay Area, China using machine learning. **Waste Management**, v. 134, p. 78–88, 1 out. 2021.
13. MAGED, A.; ELSHABOURY, N.; AKANBI, L. Data-driven prediction of construction and demolition waste generation using limited datasets in developing countries: an optimized extreme gradient boosting approach. **Environment, Development and Sustainability**, 2024.
14. MOTA GONÇALVES, M. **PRECIFICAÇÃO DE IMÓVEIS UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA E ÁRVORES DE DECISÃO**. Florianópolis: [s.n.].
15. SAKA, A. et al. **Integrated BIM and Machine Learning System for Circularity Prediction of Construction Demolition Waste**. [s.l: s.n.].
16. SILVA SANTOS, S. **Indução de Árvores de Decisão Oblíquas como Explicadores de Predições por Modelos de Aprendizado de Máquina**. Belo Horizonte - Minas Gerais: [s.n.].
17. SINGH, T.; UPPALURI, R. V. S. Machine learning tool-based prediction and forecasting of municipal solid waste generation rate: a case study in Guwahati, Assam, India. **International Journal of Environmental Science and Technology**, v. 20, n. 11, p. 12207–12230, 1 nov. 2023.

18. SIRIMEWAN, D. et al. Deep learning-based models for environmental management: Recognizing construction, renovation, and demolition waste in-the-wild. **Journal of Environmental Management**, v. 351, 1 fev. 2024.
19. SOLÍS-GUZMÁN, J. et al. A Spanish model for quantification and management of construction waste. **Waste Management**, v. 29, n. 9, p. 2542–2548, set. 2009.
20. TEIXEIRA, A.; DIONÍSIO BASÍLIO, M. **Regressão linear múltipla**. [s.l: s.n.].
21. XIA, W. et al. **Application of machine learning algorithms in municipal solid waste management: A mini review**. **Waste Management and Research** SAGE Publications Ltd, 1 jun. 2022.