ENTAC 2024
XX ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO
Maceió, Brasil, 9 a 11 de outubro de 2024

Modelo de classificação das perdas por *making-do* com uso de aprendizado de máquina

Making-do waste classification model using machine learning

Tatiana Gondim do Amaral
Universidade Federal de Goiás | Goiânia | Brasil | tatianagondim@ufg.br
Caio César Medeiros Maciel
Universidade Federal de Goiás | Aparecida de Goiânia | Brasil | caiocesar.eng@hotmail.com
Marcos Paulino Roriz Junior
Universidade Federal de Goiás | Aparecida de Goiânia | Brasil | marcosroriz@ufg.br
Gabriella Soares de Paula
Universidade Federal de Goiás | Goiânia | Brasil | gabriella_s2_soares@discente.ufg.br

Resumo

O presente artigo tem o objetivo de apresentar a aplicação de aprendizado de máquina para a classificação de perdas por *making-do*. Após a compreensão da tipologia dos dados de entrada para classificação das perdas foi feito um levantamento dos possíveis modelos de algoritmos que poderiam classificar as perdas. Posteriormente foi utilizado um banco de dados do *QuizQuality* já classificados para treinar os algoritmos e realizar a classificação com diversos modelos de aprendizado de máquina e os que tiveram melhor capacidade de categorização foi a rede neural, modelo KNN e o empilhamento das duas. Os modelos foram capazes de classificar as perdas com um *recall* de 95%. A aplicação de aprendizado de máquina para as análises e classificação das perdas possui relevância, pois a classificação é laboriosa e a utilização de modelos de inteligência artificial pode agilizar este processo.

Palavras-chave: Perdas. *Making-do*. Aprendizado de máquina. Banco de dados. Modelo de aprendizado de máquina.

Abstract

This article aims to present the application of machine learning for the classification of making-do losses. After understanding the typology of input data for loss classification, a survey of potential algorithm models capable of classifying the losses was conducted. Subsequently, a QuizQuality database, already classified, was used to train the algorithms and classify them using various machine learning models. The models that showed the best categorization ability were neural networks, KNN models, and a combination of the two. The models successfully classified the losses with a recall rate of 95%. The application of machine learning for loss

Como citar:
AMARAL, T.; MACIEL, C.; RORIZ JUNIOR, M.; PAULA, G. Modelo de classificação das perdas por *making-do* com uso de aprendizado de máquina. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 20., 2024, Maceió. **Anais**... Maceió: ANTAC, 2024.

analysis is highly relevant, as the conventional process carried out by researchers is labor-intensive, and the use of artificial intelligence models can expedite the analysis process.

Keywords: Waste. Making-do. Machine Learning. Database. Machine learning model.

INTRODUÇÃO

Os dados em um estudo de classificação de perdas por *making-do* dependem dos objetivos da pesquisa, das práticas de gestão da obra e das informações disponíveis que são predominantemente nominais (etapa da obra, atividade, mão de obra), ordinais (aprovação da tarefa) e textuais (descrição da perda, detalhamento da ação de mitigação) o que torna o processo de classificação laborioso depende da análise humana [1][2][3][4][5]. Esses dados desempenham um papel fundamental na identificação, categorização e análise das perdas, ajudando a compreender as causas subjacentes e os impactos dessas perdas no contexto da construção civil.

O aprendizado de máquina possibilita que sistemas computacionais aprendam a classificar dados sem ser programados explicitamente, isto é, sem definir explicitamente esses limites [6]. A técnica possibilita aprender uma função a partir de dados de entrada e de dados de saída. Neste caso, os dados de entrada seriam as não-conformidades das empresas do banco de dados banco de dados da Plataforma de Gestão *QuizQuality*, enquanto a saída seria o pré-requisito, a categoria e impacto da perda por *making-do*.

Há diversos métodos de aprendizado de máquina, desde métodos clássicos, como Regressão Linear e Regressão Logística, até métodos mais sofisticados, como Redes Neurais Profundas e Máquinas de Vetores de Suporte. Em termos práticos, a principal diferença entre os métodos se dá nos algoritmos utilizados para aprender a função, a partir dos dados de entrada e saída, e, consequentemente, na precisão de previsão [7]. Além do desempenho em predizer a saída, os métodos se diferenciam no fator de explicabilidade, isto é, de conseguir explicar quais são os parâmetros que levam a predizer uma determinada saída [8].

Para este trabalho optou-se por explorar o programa *Orange* por ser um *software* de fácil parametrização. Ademais, ele inclui diversos métodos para criação de modelos de aprendizado de máquina.

REFERENCIAL TEÓRICO

Conforme [9], a perda por *making-do* ocorre quando uma atividade no canteiro se inicia sem todos os pré-requisitos necessários, ou quando a execução continua mesmo na ausência de um ou mais pré-requisitos. [10] refere-se a esse conjunto de pré-requisitos como *kit* completo, destacando suas duas consequências técnicas: aumento do tempo do processo e da variabilidade, causando trabalho inacabado, aumento do tempo de espera e redução da qualidade, gerando retrabalho; e comportamentais, como redução da motivação dos trabalhadores e diminuição do esforço para garantir os itens ausentes do *kit*. Adicionalmente, [9] introduz a perda de segurança resultante do *making-do*, que surge de condições anormais da produção.

O método de identificação das perdas por *making-do*, proposto por [11] e complementado por [2][12], compreende três grupos distintos. O primeiro é empregado para identificar pré-requisitos ausentes nos pacotes de trabalho, como informação, materiais, mão de obra, ferramentas/equipamentos, espaço, serviços interdependentes, condições externas e instalações temporárias; o segundo para determinar categorias de perdas mais impactadas, como acesso/mobilidade, ajuste de componentes, área de trabalho, armazenamento, equipamentos/ferramentas, instalações provisórias, proteção e sequenciamento; e o último para avaliar os efeitos dessas perdas ou impactos, como diminuição da produtividade, redução da qualidade, retrabalho, perdas de material, redução da segurança, desmotivação e falta de terminalidade.

MÉTODO DE CLASSIFICAÇÃO

O estudo [13] apresentou o conceito de atividades facilitadoras que, por meio da remoção de restrições ou por meio de antecipação gerencial, trazem melhorias a um processo. Tais atividades referem-se a ações de aprimoramento implementadas em setores específicos para otimizar a eficiência e eficácia dos processos. Essa abordagem está intrinsecamente ligada ao conceito de melhoria contínua e resulta em um fluxo de produção contínuo, minimizando as chances de ocorrência de perdas por *making-do* [14].

Dessa maneira, baseado na classificação dos fluxos de entrada em processos de construção, desenvolveu um método para identificar o *making-do* em canteiros de obras [11]. Posteriormente, esse método foi refinado [12] incluindo a categoria "Sequenciamento" proposta por [2], que foi incorporada ao mapa conceitual e aplicada por [3].

Adicionalmente, [2] também identificaram uma correlação entre os pacotes de trabalho informais e o aumento de perdas por *making-do*. Sendo pacote formal o pacote planejado e executado na semana e pacote informal é definido como pacote que não foi planejado para a semana e é executado. Assim, essa subdivisão se torna relevante, uma vez que evidencia o grau de planejamento e a relação direta entre o aumento de perdas e a informalidade dos pacotes de trabalho. Dentre as várias metodologias de coleta de dados para classificação de perdas, diferentes abordagens foram desenvolvidas por vários autores ao longo do tempo. Nota-se que em sua grande maioria os dados são qualitativos e descritivos como causa, descrição, decisões.

É correto observar que, em muitos estudos de classificação de perdas por *making-do*, os dados de entrada predominantes são nominais e ordinais. Esses dados são essenciais para categorizar e descrever as diferentes características das perdas identificadas em um projeto de construção ou em um canteiro de obras. Também existem alguns dados numéricos e datas.

Portanto, os dados em um estudo de classificação de perdas por *making-do* como pode-se observar nos trabalhos de [1][2][3][4][5][15][16] dependem dos objetivos da pesquisa, das práticas de gestão da obra e das informações disponíveis que são

predominantemente nominais (etapa da obra, atividade, mão de obra), ordinais (aprovação da tarefa) e textuais (descrição da perda, detalhamento da ação de mitigação), tornando o processo de classificação laborioso e dependente da análise humana. Esses dados desempenham um papel fundamental na identificação, categorização e análise das perdas, ajudando a compreender melhor as causas subjacentes e os impactos dessas perdas no contexto da construção civil.

APRENDIZADO DE MÁQUINA

O aprendizado de máquina visa permitir que sistemas computacionais aprendam a classificar dados sem serem explicitamente programados, ou seja, sem definir esses limites de forma explícita [6]. A técnica envolve aprender uma função a partir de dados de entrada e saída. Neste caso, os dados de entrada seriam as não-conformidades das empresas do banco de dados do QuizQuality, enquanto a saída seria o pré-requisito, a categoria e o impacto da perda por *making-do*.

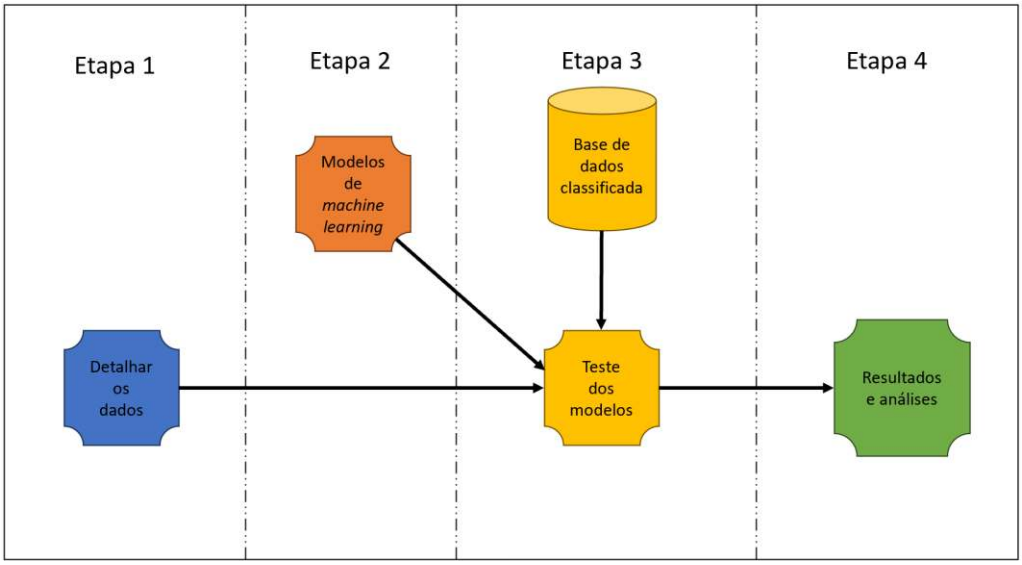
Existem diversos métodos de aprendizado de máquina, desde métodos clássicos, como Regressão Linear e Regressão Logística, até métodos mais avançados, como Redes Neurais Profundas e Máquinas de Vetores de Suporte. A principal diferença prática entre os métodos está nos algoritmos usados para aprender a função, o que influencia na precisão da previsão [17]. Além disso, os métodos variam em sua capacidade de explicar os parâmetros que levam a uma determinada previsão [8].

Para este trabalho optou-se por explorar o programa Orange por ser um *software* eficaz e de fácil parametrização. Além disso, ele inclui diversos métodos para criação de modelos de aprendizado de máquina, como Indução de regra CN2, Modelo KNN, Árvore de decisão, Floresta Aleatória, Boosting Gradiente, Máquina de Vetores de Suporte (SVM), Regressão Linear, Regressão Logística, Naive Bayes, AdaBoost, Rede Neural, entre outros.

MÉTODO

A pesquisa possui uma abordagem de modelagem quantitativa empírica descritiva, combinando elementos das pesquisas exploratórias e das pesquisas experimentais. Seu objetivo é investigar fenômenos pouco explorados ou pouco estudados, utilizando métodos experimentais [18]. Nesse contexto, busca-se obter percepções e hipóteses preliminares sobre um tópico específico, ao mesmo tempo em que se viabiliza a manipulação de dados para se observar e se compreender relações de causa e efeito. O estudo foi dividido em quatro etapas distintas, conforme ilustra a Figura 1.

Figura 1: Etapas da pesquisa.



Fonte: Autores, 2024.

Etapa 1: Detalhamento dos Dados para Avaliação das Perdas

Foi conduzida uma análise bibliográfica abrangente das classificações de perdas de produção. A partir dessa análise, foram identificados os principais tipos de dados que servem como base para a classificação das perdas, por meio do conceito de *making-do.*

Etapa 2: Ferramentas de Aprendizado de Máquina

Para a seleção das ferramentas, levou-se em consideração o detalhamento dos tipos de dados para avaliação das perdas. Esta análise constatou que os dados consistiam em elementos qualitativos, os quais se repetiam, como bloco, empresa (cuja quantidade de opções poderia ser limitada) ou dados nominais, bem como elementos descritivos (Quadro 1), ou seja, textos independentes que não necessariamente se repetiam e podiam ocorrer em múltiplas instâncias. Os dados do Quadro 1 são aqueles que influenciaram na classificação. Como bloco, empresa, andar não influenciam na classificação da perda não entraram como variável da perda.

No Quadro 1 são apresentados dois exemplos de caso de perda. A coluna “característica/descrição” traz a descrição e as características da tarefa/atividade; a coluna “problema” é utilizada para compreender o erro cometido que levou a perda e a coluna “ação contenção” é ação de mitigação determinada para perda que ocorreu. A coluna “junção” é a junção das outras 3 colunas. Essa coluna foi criada, pois ela que é utilizada no processo de aprendizado de máquina, pois contém todos os textos nas ocorrências analisadas.

Quadro 1: Exemplo de dado de entrada

ITEM	CARACTERÍSTICA/DESCRIÇÃO	PROBLEMA	AÇÃO CONTENÇÃO	JUNÇÃO
33	A altura dos cabos de protenção estão de acordo com o estabelecido em projeto?	Evidenciado que os cabos de protensão não estão na altura.	Colocar os cabos na altura conforme o projeto.	A altura dos cabos de protenção estão de acordo com o estabelecido em projeto? Evidenciado que os cabos de protensão não estão na altura. Colocar os cabos na altura conforme o projeto.
493	As dimensões da armadura e cabos estão de acordo com o estabelecido em projeto?	Estribos está em desacordo com o projeto.	Refazer todo o estribo	As dimensões da armadura e cabos estão de acordo com o estabelecido em projeto? Se os estribos estão em desacordo com o projeto, refazer todo o estribo.

Fonte: Autores, 2024.

No Quadro 2 pode-se verificar a classificação das perdas referente ao Quadro 1. Observa-se que, tanto no item 33 e 493 a tarefa não estava de acordo com o projeto, concluindo-se assim que o pré-requisito faltante foi a mão de obra. A categoria foi ajustes de componentes, pois será necessário uma correção inesperada e o impacto foi retrabalho, pois a atividade deverá ser retrabalhada.

Quadro 2: Exemplo de dados de saída

EQUIPE	ETAPA	SUBETAPA	ATIVIDADE	PRÉ-REQUISITO	CATEGORIA	IMPACTO
Armador	Estrutura	Superestrutura	Armação da superestrutura	Mão de Obra	Ajustes de componentes	Retrabalho
Armador	Estrutura	Superestrutura	Armação da superestrutura	Mão de Obra	Ajustes de componentes	Retrabalho

Fonte: Autores, 2024.

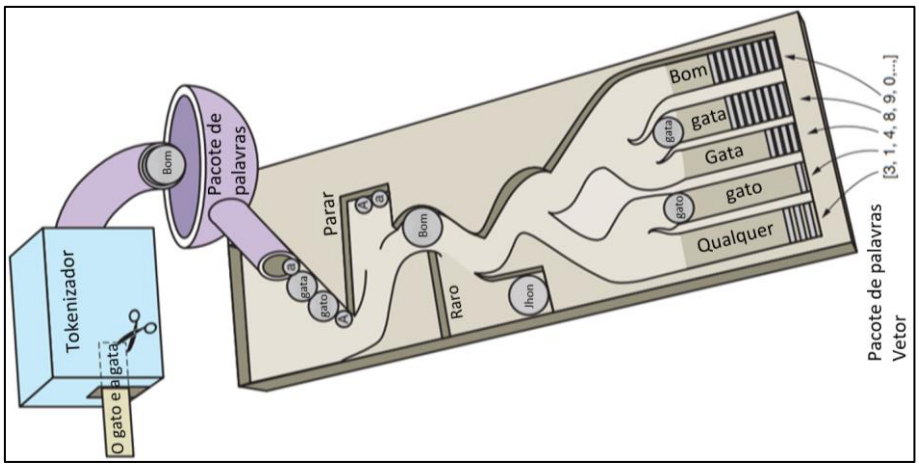
Assim, embasados na análise prévia da literatura e na exploração das ferramentas disponíveis no *Software Orange* (programa selecionado para conduzir as análises), foram identificadas as principais ferramentas com a capacidade de analisar dados qualitativos e descritivos (textos).

ETAPA 3: TESTE E AVALIAÇÃO COM BANCO DE DADOS CLASSIFICADO

Após a seleção das ferramentas, procedeu-se com a realização de testes utilizando o banco de dados do *QuizQuality*, o qual continha 6.339 já classificados de obras de multipavimentos. Nesse contexto, 80% dos dados foram empregados no treinamento das ferramentas, permitindo assim compreender o comportamento do aprendizado de máquina. Na etapa subsequente, os 20% restantes dos dados foram submetidos à classificação pelas ferramentas. Por meio desse processo, foi possível determinar as acurácias das ferramentas e, por conseguinte, selecionar aquela que demonstrou o melhor desempenho.

Para a análise dos dados de textos foram necessárias seu processamento em vetores que é similar a Figura 2, esse processamento foi necessário, pois os dados de entrada (características/descrição, problema e ação de contenção) não são nominais e sim textos de linguagem natural.

Figura 2: Sistema de classificação de texto.



Fonte: Adaptado de [19].

Para isso foi feito o procedimento conforme a Figura 3. Seguindo as etapas descritas abaixo:

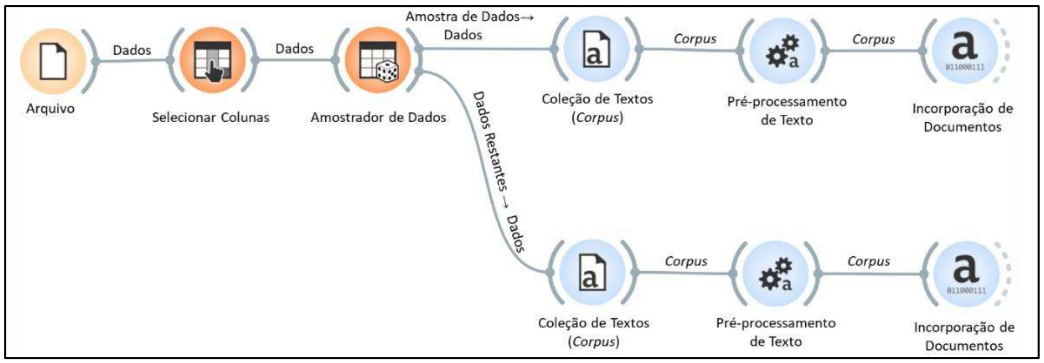
I. Arquivo (*File*): Arquivo contendo o banco de dados foi adicionado pela função Arquivo (*File*);

II. Selecionar coluna (*Select Columns*): Função para retirar dados não influentes na classificação da perda, como número de identificação;

III. Amostrador de dados (*Data sampler*): Essa função separa o banco de dados em dois bancos, o primeiro (80% dos dados) é usado para treinar o modelo e o segundo (20%) é usado para testar os modelos;

IV. Coleção de Textos (*Corpus*): Essa função contabiliza os dados textuais, cada palavra é contabilizada quantas vezes é repetida em cada texto conforme a Figura 4.

Figura 3: Fluxograma da conversão de texto em vetores.



Fonte: Autores, 2024.

Figura 4: Contabilização das palavras.

	PRÉ-REQUISITO	Junção	{...}
1	Materiais e componente	TONALIDADE E COR DO REJU...	COR=1, REJUNTE=1, TONALIDADE=1, algumas=1, camada...
2	Mão de Obra	Fixação das Tubulações e Con...	Conexões=1, Fixação=1, Tubulações=1, apto=1, bho=1, ca...
3	Mão de Obra	Juntas superfície irregular difi...	Juntas=1, dificultando=1, emboço=1, impermeabilização=...
4	Materiais e componente	Revestimentos de Pisos - Man...	Manchas=1, Pisos=1, Revestimentos=1, Sala=1, mancha=1,...
5	Mão de Obra	Esquadria de Madeira e Prend...	Esquadria=1, Funcionamento=1, Madeira=1, Porta=1, Pren...
6	Mão de Obra	Paredes - Pinturas e Armadore...	Acabamentos=1, Armadores=1, Paredes=1, Pinturas=1, aca...
7	Mão de Obra	Paredes / Forro Pinturas em T...	Forro=1, Latex=1, Manchas=1, Paredes=1, Pinturas=1, Text...
8	Materiais e componente	Rodapés - Integridade rodapé...	Integridade=1, Rodapés=1, danificado=1, pintura=1, quart...
9	Mão de Obra	Fixação de tubulações, caixas ...	Fixação=1, Limpeza=1, caixas=1, elétricos=1, limpar=1, qu...
10	Materiais e componente	Esquadrias de Alumínio - Ved...	Alumínio=1, Aplicação=1, Esquadrias=1, Falta=1, Vedação...

Fonte: Autores, 2024.

V. Pré-processamento de texto (*Prepocess Text*): Nessa etapa o texto passa por um pré-processamento para retirar possíveis símbolos dos textos que não tem significância na classificação das perdas, por exemplo um ponto de exclamação, artigos e preposições;

VI. Incorporação de Documentos (*Document Embedding*): Nessa etapa cada texto é transformado em 384 vetores que representam o texto original (Figura 5), possibilitando assim o processamento pelos modelos.

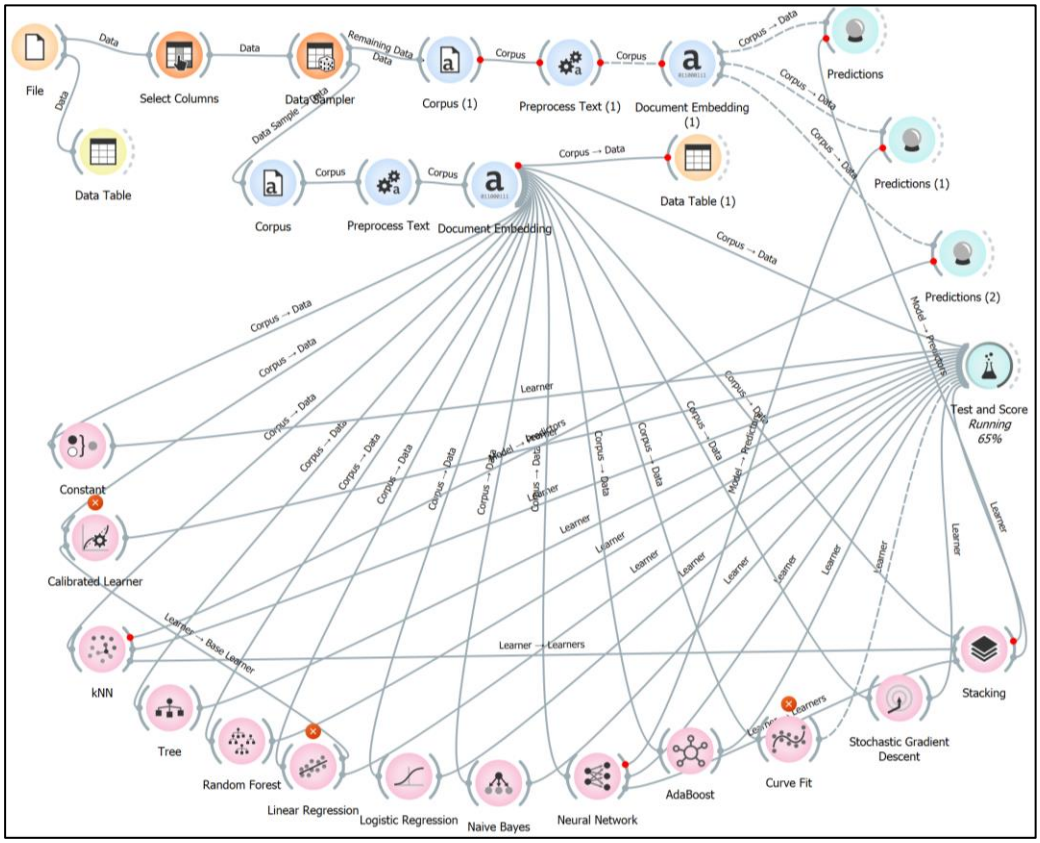
Figura 5: Exemplos de textos transformados em vetores.

	PRÉ-REQUISITO	Junção	Dim1 True True	Dim2 True True	Dim3 True True	Dim4 True True	Dim5 True True	Dim6 True True
1	Equipamentos ...	Locação das est...	0.132742	-0.0371847	0.0274826	-0.173142	-0.158087	-0.152429
2	Materiais e com...	Rodapés - Inte...	0.0128997	-0.00647336	0.0705915	0.0337555	0.162191	-0.266895
3	Mão de Obra	Espaço e posiçã...	0.214169	0.341076	-0.155635	-0.0683022	-0.208024	-0.0133135
4	Mão de Obra	Esquadro e pru...	-0.18941	0.0629965	-0.0557362	0.00640846	0.00637953	-0.20828
5	Mão de Obra	Paredes - Pintur...	0.177171	0.226419	-0.104572	0.305105	0.205609	0.178991
6	Mão de Obra	Planicidade e p...	-0.00790298	0.261654	0.131579	0.284264	-0.0103033	-0.204494
7	Materiais e com...	INTEGRIDADE ...	-0.00475314	0.0187142	-0.0830805	-0.0629294	-0.0804797	-0.0425104
8	Condições exte...	Revestimentos ...	-0.0197566	0.0949641	0.174285	0.111969	0.0875384	-0.0546769
9	Materiais e com...	INTEGRIDADE ...	0.0875906	0.208024	-0.104713	-0.108189	-0.0708916	-0.0958154
10	Materiais e com...	Esquadria de M...	0.0973974	0.0125606	0.165848	-0.043096	0.0658446	-0.00551703
11	Tarefas interde...	Bancada, Peitor...	0.0942781	-0.0978093	0.108858	0.118229	-0.00238385	0.0334589
12	Mão de Obra	Limpeza e Outr...	-0.130666	0.0910244	0.190885	0.183462	-0.114602	0.133294
13	Materiais e com...	INTEGRIDADE ...	-0.142334	0.0694812	-0.108799	-0.0280661	-0.0275095	0.0608332
14	Mão de Obra	Paredes - Pintur...	0.273295	0.152194	0.0302904	0.22708	0.0611749	0.297659
15	Tarefas interde...	Varanda Facha...	0.246181	0.18756	0.0378175	0.367517	-0.0637718	0.0605816
16	Mão de Obra	Limpeza e Outr...	0.0774662	-0.0208296	0.102481	0.170505	0.118889	-0.0816948
17	Mão de Obra	As quantidades...	-0.187081	0.206703	-0.128539	-0.044847	0.0345352	0.0110977

Fonte: Autores.

Após adequação dos dados, 16 modelos foram utilizados para descrever os pré-requisitos ausentes (Figura 6) por meio dos dados de entrada exemplificados no Quadro 1. Entretanto, 3 modelos não funcionaram (Indução CN2, Descida de Gradiente Estocástica e *Boosting* Gradiente), outros 3 não foram capazes de criar um modelo, pois são usados apenas para dados numéricos (Regressão linear, *Curve Fit*) ou para classificação binária (Modelo Calibrado). Vale ressaltar que os modelos funcionaram nas configurações padrões do *software Orange*. Por último, os modelos que obtiveram melhor desempenho foram utilizados para predição dos 20% dos dados restantes que foram separados para teste, como pode ser visto na Figura 6 por meio da função *Prediction* (predição).

Figura 6: Modelos aplicados aos dados.



Fonte: Autores, 2024.

RESULTADOS E DISCUSSÕES

Para análise dos dados foi considerado dois parâmetros: a precisão e o *recall*. Por exemplo, se em uma coleção de documentos tivessem 40 documentos relevantes e um algoritmo selecionasse 10 documentos relevantes sendo apenas 8 de fato relevantes, então o "recall em 10" se referiria a recuperação de todos os dados relevantes, ou seja, 8/40=20%; enquanto "precisão em 10" seria 8/10=80% [18]. De maneira direta, a alta precisão indica que o algoritmo trouxe mais resultados relevantes do que irrelevantes, ao passo que o alto *recall* indica que o algoritmo recuperou a maioria dos resultados relevantes.

AVALIAÇÃO DOS MODELOS CRIADOS

A Figura 8 ilustra alguns exemplos de modelos de rede neural de predição. O erro da coluna mostra a imprecisão da previsão no caso específico avaliado. Nesse exemplo, apenas o caso 4 foi predito incorretamente.

Figura 7: Exemplo de predição usando rede neural.

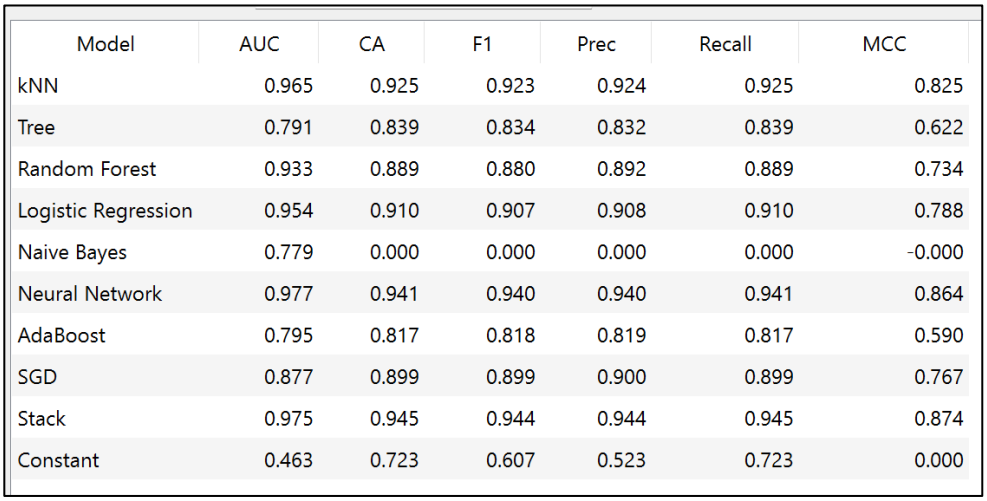


	Neural Network	error	PRÉ-REQUISITO	Junção	Dim1	Dim2
1	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 → Mão de Obra	0.000	Mão de Obra	O AMBIENTE D...	0.10777	-0.0142492
2	0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.01 : 0.00 → Equipamentos ou Ferramentas	0.011	Equipamentos ou Ferrame...	Profundidade d...	0.0675948	0.0498018
3	0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 → Materiais e componente	0.001	Materiais e componente	Ralos e Descarg...	0.00764766	0.0712124
4	0.02 : 0.00 : 0.04 : 0.00 : 0.26 : 0.68 : 0.01 → Mão de Obra	0.745	Materiais e componente	Rodapés - Inte...	-0.139698	0.159239
5	0.00 : 0.00 : 0.00 : 0.00 : 0.68 : 0.32 : 0.00 → Materiais e componente	0.318	Materiais e componente	Ralos e Descarg...	0.097311	-0.0109719
6	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 → Mão de Obra	0.000	Mão de Obra	Paredes Pintur...	0.170618	0.320149
7	0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 → Materiais e componente	0.000	Materiais e componente	Esquadrias de A...	-0.0547868	0.304647
8	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 → Mão de Obra	0.000	Mão de Obra	As quantidades...	0.146353	0.236687

Fonte: Autores, 2024.

Na Figura 8 se encontram a avalição de todos os modelos aplicados a classificação das categorias para determinar o mais adequado e os com maiores *recall* (Coluna *Recall* na Figura 8 foram o método KNN (92,5%) e rede neural (94,1%), como os dois obtiveram maiores acertos foram usados no método de empilhamento alcançando um *recall* de 94,4% que foi o melhor resultado.

Figura 8: Resultados dos modelos de aprendizado de máquina para classificação da categoria.



Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.965	0.925	0.923	0.924	0.925	0.825
Tree	0.791	0.839	0.834	0.832	0.839	0.622
Random Forest	0.933	0.889	0.880	0.892	0.889	0.734
Logistic Regression	0.954	0.910	0.907	0.908	0.910	0.788
Naive Bayes	0.779	0.000	0.000	0.000	0.000	-0.000
Neural Network	0.977	0.941	0.940	0.940	0.941	0.864
AdaBoost	0.795	0.817	0.818	0.819	0.817	0.590
SGD	0.877	0.899	0.899	0.900	0.899	0.767
Stack	0.975	0.945	0.944	0.944	0.945	0.874
Constant	0.463	0.723	0.607	0.523	0.723	0.000

Fonte: Autores, 2024.

O Quadro 3 apresenta qual modelo foi mais eficiente para predizer cada dado de saída, em sua grande maioria, o modelo mais eficiente foi o empilhamento que é a junção do modelo de rede neural e o KNN e o segundo modelo mais eficiente foi a rede neural. Como o empilhamento não é um modelo em si e sim a combinação de outros modelos, o modelo mais eficiente para a classificação das perdas por *making-do* é a rede neural.

Quadro 3: Modelos mais eficientes para cada dado de saída.

DADOS DE SAÍDA	MELHOR MODELO	RECALL	SEGUNDO MODELO	RECALL
Etapa	Empilhamento	94,2%	Rede Neural	93,8%
Subetapa	Empilhamento	90,8%	Rede Neural	90,5%
Atividade	Empilhamento	89,1%	Rede Neural	88,9%
Equipe	Rede Neural	92,1%	Empilhamento	92,1%
Pré-requisito	Empilhamento	91,4%	Rede Neural	91,3%
Categoria	Empilhamento	94,5%	Rede Neural	94,1%
Impacto	Empilhamento	90,6%	Rede Neural	90,0%

Fonte: Autores, 2024.

A avaliação da capacidade dos 3 métodos predizerem os dados que foram separados para serem amostras está apresentada no Quadro 4. O *recall* do empilhamento e da rede neural foram próximas respectivamente 93,6% e 94,4% enquanto o modelo KNN foi de 92,3%. O *recall* mostra que o algoritmo consegue retornar a maioria dos resultados relevantes. Considerando ser um método qualitativo um *recall* acima de 90% é muito assertivo.

Quadro 4: Resultados dos modelos aplicados a base de dados separada.

MODELO	RECALL
Empilhamento	94,40%
Rede Neural	93,60%
KNN	92,30%

Fonte: Autores, 2024.

No Quadro 5, apresenta-se o modelo que apresentou melhor resultado para cada dado de saída, considerando os dados separados para a amostra. Vale ressaltar que os únicos dados de saída que obtiveram *recall* abaixo de 90% foram atividade e impacto. A atividade quando classificada, tem como dado de entrada a etapa e, como neste caso, não foi utilizado como dado de entrada, poderia ser a causa do *recall* abaixo de 90%. Já o impacto provavelmente obteve um *recall* abaixo de 90%, pois em uma atividade pode ocorrer mais de um impacto e o pesquisador classifica o mais relevante e como esta definição é feita pelos pesquisadores se torna uma decisão com maior subjetividade.

Quadro 5: *Recall* da predição do modelo com a amostra de dados.

Dados de saída	Modelo para predição	Recall	Modelo para predição	Recall
Etapa	Empilhamento	93,60%	Rede Neural	93,90%
Subetapa	Empilhamento	91,30%	Rede Neural	91,30%
Atividade	Empilhamento	88,20%	Rede Neural	88,30%
Equipe	Empilhamento	91,80%	Rede Neural	91,80%
Pré-requisito	Empilhamento	91,40%	Rede Neural	90,60%
Categoria	Empilhamento	94,40%	Rede Neural	93,60%
Impacto	Empilhamento	89,30%	Rede Neural	88,50%

Fonte: Autores, 2024.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Os modelos de aprendizado de máquina têm demonstrado eficácia na classificação das perdas por *making-do*, alcançando um *recall* de quase 95% para variáveis qualitativas. Esse resultado é significativo, pois pode diminuir a laboriosidade da classificação dos dados em classificações futuras.

Além disso, um alto *recall* aumenta a confiabilidade do método, sugerindo a utilização de redes neurais ou a combinação de redes com outros métodos eficazes, como o KNN, em pesquisas futuras. A aplicação de aprendizado de máquina não se restringe à

classificação de perdas por *making-do,* podendo ser estendida a diversas outras áreas da construção civil.

Como perspectiva de estudo futuro, sugere-se a aplicação do aprendizado de máquina para classificação de novos bancos de dados para avaliar a sua aplicabilidade. Além disso, sugere-se a utilização da classificação com uso de aprendizado de máquina para outras obras e tipologias, considerando que o presente estudo se concentrou em obras de edificações de multipavimentos.

REFERÊNCIAS

[1] PIKAS, E.; SACKS, R.; PRIVEN, V. (2012**). Go or No-Go Decisions at the Construction Workface: Uncertainty, Perceptions of Readiness, Making Ready and Making-do.** 20th Annual Conference of the International Group for Lean Construction. San Diego, USA, 18-20 Jul 2012.

[2] FIREMAN, M. C. T.; FORMOSO, C. T. **Integrating production and quality control: monitoring making-do and unfinished work.** Annual Conference of the International Group for Lean Construction, 21th, Fortaleza, 2013. Proceedings… Fortaleza, 2013.

[3] LEÃO, C.F.; FORMOSO, C.T.; ISATTO, E. L. **Integrating Production and Quality Control with the Support of Information Technology.** Annual Conference of The International Group for Lean Construction, 22, IGLC, Oslo, 2014.

[4] LEÃO, C. F.; ISATTO, E. L.; FORMOSO, C. T. **Proposta de modelo para controle integrado da produção e da qualidade com apoio da computação móvel.** Ambiente Construído, Porto Alegre, v. 16, n. 4, p. 109-124. 2016.

[5] AMARAL, T. G.; BRAGA, P. B.; BARROS NETO, J. **Application of Dynamic Spreadsheets in the Analysis of Waste by Making-do.** Annual Conference of the International Group for Lean Construction, 28th, Berkeley, IGLC, Berkeley, 2020.

[6] FLACH, P. Machine learning**: The art and science of algorithms that make sense of data.** Cambridge University Press, 2012.

[7] ZHANG, C.; LIU, C.; ZHANG, X.; ALMPANIDIS, G. **An up-to-date comparison of state-of-the-art classification algorithms.** Expert Systems with Applications, v. 82, p. 128–150, 2017.

[8] LOYOLA-GONZALEZ, Octavio. **Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View.** IEEE Access, v. 7, p. 154096–154113, 2019.

[9] KOSKELA, L. Making-do – **The Eighth Category of Waste.** Annual Conference of the International Group for Lean Construction, 12, 2004, Dinamarca, 2004.

[10] RONEN, B. **The Complete Kit Concept.** International Journal of Production, v. 30, n. 10, p. 2457–2466, 1992.

[11] SOMMER, L. **Contribuições Para Um Método de Identificação de Perdas Por improvisação em Canteiros de Obras.** Porto Alegre, 2010. 150 f. Dissertação (Mestrado em Engenharia Civil) – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

[12] SANTOS, P. R. R.; SANTOS, D. DE G. **Investigação de perdas devido ao trabalho inacabado e o seu impacto no tempo de ciclo dos processos construtivos.** Ambiente Construído, Porto Alegre, v. 17, n. 2, p. 39-52, abr./jun. 2017.

[13] SANTOS, D. G. **Modelo de gestão de processos na construção civil para identificação de atividades facilitadoras.** Doutora Dissertação. Pós-Graduação no Programa de Engenharia de Produção, Universidade Federal de Santa Catarina, 2004.

[14] KOSKELA, L. **An exploration towards a production theory and its application to construction.** Thesis (Ph.D) - Technical Research Centre of Finland, Espoo, 2000.

[15] FORMOSO, C. T.; SOMMER, L.; KOSKELA, L.; ISATTO, E. L. **The identification and analysis of making-do waste: insights from two Brazilian construction sites.** Ambiente Construído, v. 17, n. 3, p. 183-197, jul. /set. 2017.

[16] SANTOS, E. M.; FONTENELE, A. D.; MACHADO, M. L.; BARROS NETO, J. P.; AMARAL, T. G. **Analysis of Making-do Waste at Site in Fortaleza, Ceará, Brazil.** Annual Conference of the International Group for Lean Construction, 28th, Berkeley, IGLC, Berkeley, 2020.

[17] ZHANG, C.; LIU, C.; ZHANG, X.; ALMPANIDIS, G. **An up-to-date comparison of state-of-the-art classification algorithms.** Expert Systems with Applications, v. 82, p. 128–150, 2017.

[18] BERTRAND J. W. M.; FRASOO J.C**. Modelling and simulation. Operations managemant research methodologies using quantitative modeling.** International Journal of Operartions& Production Management. Vol. 22, No 22, 2002. P. 241-264.

[19] LANE, H.; HOWARD, C.; HAPKE, H. M. **Natural Language Processing in Action: Understanding, analyzing and Generation text with Python.** Manning, Shelter, NY, 2019.