

Modelos de prognóstico de produtividade na construção civil: caracterização e análise crítica comparativa em estudo de caso

Construction productivity forecasting modelling: characterization and comparative critical analysis on a case study

Marcelo Inocência Ferreira Correa 

Maria Carolina Gomes de Oliveira Brandstetter 

Larsson Diogo Seabra Coelho Romagnoli 

Resumo

As taxas de produtividade utilizadas como referência por construtoras, de maneira geral, são obtidas de forma empírica, por meio de banco de dados de projetos anteriores, ou com base em índices de referência de manuais de orçamentação. Entretanto, a prática de indicadores médios de produtividade significa uma postura simplificada demais ante as atuais necessidades de entendimento mais aprofundado das atividades de construção, haja vista a grande quantidade de fatores de conteúdo e contexto que podem influenciar os serviços. Uma alternativa de prever a produtividade é por meio de modelos de previsão ou prognóstico, abordagens sistemáticas utilizadas para desenvolver representações matemáticas ou computacionais que descrevem a realidade de um sistema, processo ou fenômeno. Assim, este trabalho tem como objetivo aplicar e comparar quatro diferentes técnicas de modelagem para prognóstico da produtividade, sendo dois modelos estatísticos e dois modelos de inteligência artificial. Com base em nove fatores de entrada de conteúdo e contexto considerados significativos para serviços de execução de formas de concreto, o prognóstico da produtividade foi realizado e os diferentes modelos empregados foram avaliados e comparados. Os resultados demonstram que nem sempre é possível encontrar em um único modelo os melhores parâmetros de assertividade.

Palavras-chave: Produtividade. Mão de obra. Prognóstico. Inteligência Artificial.

Abstract

The productivity rates used as references by construction companies are generally obtained empirically, either through databases of previous projects or based on reference indices from budgeting manuals. However, the use of average productivity indicators represents an overly simplistic approach considering the current need for a deeper understanding of construction activities, given the large number of content and context factors that can influence services. An alternative for predicting productivity lies in forecasting models, which are systematic approaches used to develop mathematical or computational representations that describe the reality of a system, process, or phenomenon. Thus, this study aims to apply and compare four different modeling techniques for productivity forecasting, including two statistical models and two artificial intelligence models. The productivity forecasting was carried out based on nine content and context input factors deemed significant for concrete formwork execution services. The different models employed were evaluated. The results demonstrate that it is not always possible to find the best accuracy parameters within a single model.

Keywords: Productivity. Labor. Forecasting. Artificial Intelligence.

¹Marcelo Inocência Ferreira
Correa

¹Universidade Federal de Goiás
Goiânia - GO - Brasil

²Maria Carolina Gomes de
Oliveira Brandstetter

²Universidade Federal de Goiás
Goiânia - GO - Brasil

³Larsson Diogo Seabra Coelho
Romagnoli

³Universidade Federal de Goiás
Goiânia - GO - Brasil

Recebido em 17/01/25

Aceito em 27/05/25

Introdução

A produtividade pode ser conceituada como a capacidade de transformação da força de trabalho em produto construído (Souza; Morasco; Ribeiro, 2017) ou em outras palavras, produtividade é a eficiência com a qual os insumos em um canteiro de obras são transformados, considerando os fatores de conteúdo e contexto de cada situação.

O controle e o acompanhamento da produtividade pode ser um instrumento importante para a gestão da mão de obra (Thomas; Yakoumis, 1987; Sanders; Thomas, 1993), fornecendo meios de subsidiar práticas para redução de custos e aumento da motivação e da eficiência no trabalho, em especial na construção civil, tida por muitos estudos como a indústria de menor evolução da produtividade desde a sua existência (Dixit *et al.*, 2019).

Em geral, as taxas de produção por homem hora utilizadas como referência por construtores são empíricas, podendo requerer ajustes para se adequar conforme o projeto da construção. Estimativas também podem ser obtidas por meio de banco de dados de projetos anteriores, podendo ser extrapoladas para empreendimentos futuros, ou com base em índices de referência obtidos de manuais de orçamentação (Souza; Marchioro; Araujo, 2005).

De acordo com Flaten *et al.* (2018), dados de produtividade insatisfatórios afetam a precisão das durações das atividades previstas, e durações imprecisas das atividades impossibilitam a produção de um cronograma de construção confiável. Ademais, a prática de adoção de valores médios de produtividade para um serviço significa uma postura simplificada demais ante as atuais necessidades de entendimento mais aprofundado das atividades de construção, que nem sempre são adequadas às características de conteúdo e de contexto da obra a ser construída (Dornelas, 2013; Marchiori, 2009; Souza; Marchioro; Araujo, 2005).

Dessa forma, a produtividade pode ser alternativamente prevista por meio de modelos de previsão ou prognóstico, cujo desempenho requer a escolha de uma técnica ou ferramenta apropriada e de dados de entrada condizentes com a realidade dos canteiros de obras (Aziz; Eid; Amer, 2020). Justifica-se, portanto, o desenvolvimento e o aprimoramento desses modelos.

Produzir prognósticos confiáveis pode ser uma tarefa complexa pelos muitos fatores que podem influenciar um projeto de construção (Flaten *et al.*, 2018). A maioria desses fatores são dependentes uns dos outros, o que torna mais desafiador estudar seu efeito sobre a produtividade do trabalho.

Neste contexto, este trabalho tem como objetivo aplicar e comparar quatro diferentes técnicas de modelagem para prognóstico da produtividade, com uso de dados de entrada de fatores de conteúdo e contexto que impactam serviços de construção civil, avaliando os diferentes modelos selecionados quanto às características e aos desempenhos dos resultados.

Referencial teórico

A modelagem é uma abordagem sistemática utilizada para desenvolver representações matemáticas ou computacionais que descrevem a realidade de um sistema, processo ou fenômeno (Dixit *et al.*, 2019). Essa técnica oferece estruturas que possibilitam compreender as relações entre variáveis, identificar padrões, realizar previsões e fundamentar a tomada de decisões, apoiando-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento no futuro.

Segundo Ebrahimi, Fayek e Sumati (2021), os modelos de prognóstico de produtividade podem ser classificados em três categorias principais, com base na técnica de modelagem utilizada: modelos estatísticos, modelos baseados em inteligência artificial e modelos de simulação. Cada categoria apresenta abordagens e características específicas, que devem ser escolhidas conforme a natureza dos dados, o contexto de aplicação e os objetivos do prognóstico.

Modelos estatísticos

Os modelos estatísticos referem-se a um conjunto de técnicas matemáticas para representar a realidade de forma probabilística, buscando estabelecer uma relação matemática entre as variáveis de entrada e saída por meio de estatísticas tradicionais, como técnicas de regressão, análise de tendências e séries temporais (Farias; Soares; César, 2008; Montgomery; Jennings; Kulahci, 2015).

De acordo com apontamento de vários autores (Ebrahimi; Fayek; Sumati, 2021; Mohamed; Srinavin, 2005; Mohsenijam; Lu, 2019; Sonmez; Rowings, 1998; Thomas; Yiakoumis, 1987; Zeviani; Ribeiro Júnior; Bonat,

2013), a análise de regressão permite o estudo de uma série de problemas práticos e nas mais diversas áreas, o que faz com que os modelos de regressão sejam largamente adotados e, provavelmente, representem a mais utilizada dentre as metodologias estatísticas para prognóstico.

Zeviani, Ribeiro Júnior e Bonat (2013) explicam que os modelos de regressão são usados para descrever o comportamento de uma variável, a variável dependente (ou variável resposta), em relação a uma ou mais variáveis, as variáveis explicativas (ou variáveis independentes, covariáveis, estímulos), com o objetivo de estimar e/ou prever o valor médio da variável dependente em termo dos valores conhecidos ou fixos das variáveis explicativas. Com tais modelos busca-se encontrar e descrever padrões de homogeneidade dentre a heterogeneidade observada em um conjunto de observações.

Modelo estatístico de regressão linear

O modelo de regressão é classificado como linear se a variável de interesse é função linear dos parâmetros de entrada. Caso contrário, a regressão é não linear. Quando há múltiplos relacionamentos entre a variável dependente e as variáveis independentes descritos através da combinação linear de coeficientes de regressão, o modelo é ainda classificado como regressão linear múltipla (FARIAS *et al.*, 2008).

Segundo Montgomery, Jennings e Kulahci (2015), a regressão linear múltipla é descrita genericamente pela Equação 1.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad \text{Eq. 1}$$

Em que y é a variável de resposta, $x_1, x_2 \dots x_k$ são as k variáveis de entrada, $\beta_0, \beta_1, \beta_2 \dots \beta_k$ são os coeficientes de regressão e ε é a componente aleatória, denominada resíduo, que não pode ser estimada. Para que o modelo seja válido, a distribuição estatística do resíduo deve obedecer a uma distribuição gaussiana de média zero e variância σ^2 e, além disso, os resíduos devem possuir covariância zero, o que significa que devem ser independentes entre si.

No desenvolvimento de um modelo de previsão baseado na regressão linear, é necessário verificar se pelo menos uma das variáveis de entrada possui contribuição significativa para o valor da resposta, ou se a variação das entradas tem um impacto mínimo ou nulo no valor de saída devido à predominância de componentes aleatórias (Farias; Soares; César, 2008). Para avaliar essa significância, utiliza-se um teste de hipóteses com as premissas das Equações 2 e 3:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_k = 0 \quad \text{Eq. 2}$$

$$H_1: \beta_j \neq 0 \quad \text{para no mínimo um } j \quad \text{Eq. 3}$$

Em que:

H_0 é a hipótese nula; e

H_1 é a hipótese alternativa.

Para se determinar qual hipótese é válida, considera-se H_0 verdadeira e calcula-se a sua probabilidade, denominada de *valor-P*. Este valor é comparado com valores críticos que são determinados pelo nível de significância adotado. O nível de significância é definido como a probabilidade de rejeitar incorretamente a hipótese H_0 . Resultados de *valor-P* menores que o nível de significância determinam a rejeição da hipótese H_0 , indicando que ao menos uma das variáveis dependentes contribui para a variabilidade da variável resposta.

Outros indicadores relevantes desta técnica são o indicador estatístico F , o coeficiente de determinação R^2 , o coeficiente de determinação ajustado (R^2_a) e o erro quadrado médio (EQM). O valor F na regressão linear avalia a significância global do modelo, verificando se pelo menos uma das variáveis independentes explica a variável dependente de forma estatisticamente significativa. O coeficiente de determinação R^2 indica a precisão geral do modelo medindo quanto da variabilidade variável de resposta é explicada pelo modelo (Cottrell, 2006).

Modelo estatístico de regressão logística

Conforme explanado por Gurmu (2019), outra técnica estatística com grande potencial de aplicação na construção civil é a regressão logística, utilizada para criar um modelo que permite a previsão de valores de uma variável categórica, geralmente binária, com base em uma ou mais variáveis independentes. O objetivo é compreender e quantificar a relação entre as variáveis independentes e a variável dependente categórica, permitindo a tomada de decisões preditivas ou inferenciais. A partir do modelo gerado é possível calcular ou prever a probabilidade de um evento ocorrer, dado um conjunto de observações aleatórias.

De acordo com Batista (2015), a regressão logística permite estimar a probabilidade P de ocorrência de um evento. No caso da variável dependente y assumir apenas dois possíveis estados (1 ou 0) e haver um conjunto de k variáveis independentes $x_1, x_2 \dots x_k$, o modelo de regressão logística pode ser escrito conforme Equações 4 e 5:

$$P(y = 1) = \frac{1}{1 + e^{-g(x)}} \quad \text{Eq. 4}$$

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad \text{Eq. 5}$$

Em que $\beta_0, \beta_1, \beta_2 \dots \beta_k$ são os coeficientes de regressão, estimados a partir do conjunto de dados, pelo método da máxima verossimilhança, que encontra uma combinação de coeficientes que maximiza a probabilidade de a amostra ter sido observada. A função $g(x)$ representa a transformação dos preditores para garantir que a saída seja uma probabilidade entre 0 e 1. Essa função é a função sigmoide (ou logística) e, quando há múltiplas variáveis independentes, x é um vetor.

Assim como é possível estimar diretamente a probabilidade de ocorrência de um evento, é factível calcular a probabilidade de não ocorrência pela diferença, conforme Equação 6. Na regressão logística, a principal suposição é a de que o logaritmo da razão entre as probabilidades de ocorrência e não ocorrência do evento é linear (Equações 7 e 8).

$$P(y = 0) = 1 - P(y = 1) \quad \text{Eq. 6}$$

$$\frac{P(y=1)}{P(y=0)} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k} \quad \text{Eq. 7}$$

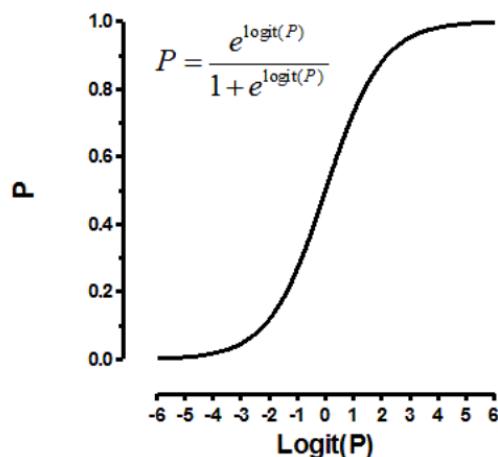
$$\ln \frac{P(y=1)}{P(y=0)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad \text{Eq. 8}$$

Por esse motivo, ao interpretar os coeficientes da regressão logística, opta-se pela interpretação de e^β não diretamente de β . Contudo, quando se utiliza o modelo logístico do ponto de vista de discriminação entre grupos, não há grande interesse na interpretação dos coeficientes (Batista, 2015). A curva do modelo tem comportamento probabilístico no formato da letra S, o que é característica da regressão logística, conforme apresentado na Figura 1.

Na opinião de Gurmu (2019), a grande vantagem do modelo de regressão logística é a expressão do resultado de forma probabilística, que proporciona uma visão mais ampla e realista da produtividade na indústria da construção ao refletir a natureza incerta e variável dos processos construtivos. Entretanto, a interpretação dos resultados probabilísticos pode ser mais difícil em comparação à regressão linear, pois exige uma compreensão das distribuições de probabilidade e do conceito de incerteza.

De acordo com Song e Abourizk (2008), não apenas as técnicas de regressão, mas os modelos estatísticos em geral, são limitados pelo número de fatores de influência que podem ser incluídos e sua capacidade de medir o efeito combinado destes fatores. Em contrapartida, esses modelos são de uso comum na engenharia, o que facilita sua disseminação, e exigem menos recursos computacionais em comparação com modelos mais complexos, como os de inteligência artificial.

Figura 1 - Curva do modelo de regressão logística (Batista, 2015)



Modelos de inteligência artificial

Os modelos de inteligência artificial (IA) utilizam algoritmos e técnicas de aprendizado de máquina para analisar grandes volumes de dados e identificar padrões complexos que podem ser usados para fazer prognósticos. Estes modelos são caracterizados por sua capacidade de aprendizado a partir de dados disponíveis, processamentos não lineares e flexibilidade quanto ao tipo de dados de entrada (Bhilwade *et al.*, 2022).

Segundo Oral *et al.* (2012), os métodos de aprendizado de máquina podem ser divididos em duas categorias principais: supervisionado e não supervisionado. Na primeira, o modelo é treinado usando um conjunto de dados cuja variável de resposta é pré-determinada. No aprendizado não supervisionado, o modelo é treinado em um conjunto de dados que não possui rótulos ou informações de saída conhecidas. O objetivo é que o modelo encontre padrões, estruturas ou relações ocultas nos dados por conta própria, sem orientação externa. Essas categorias se diferem teoricamente, portanto, quanto à estrutura causal dos processos de aprendizagem.

Os modelos de inteligência artificial, de uma maneira geral, têm como principal característica a capacidade de aprender com a experiência para melhorar seu desempenho, adaptar-se a mudanças e encontrar padrões entre conjuntos de dados, o que faz com que sejam os métodos de previsão mais eficazes, segundo Ebrahimi, Fayek e Sumati (2021). Os autores relatam ainda que modelar a produtividade na construção é desafiador porque requer a avaliação do impacto de vários fatores simultaneamente e, nesse sentido, várias técnicas de IA foram aplicadas com relativo sucesso para esta natureza de prognóstico.

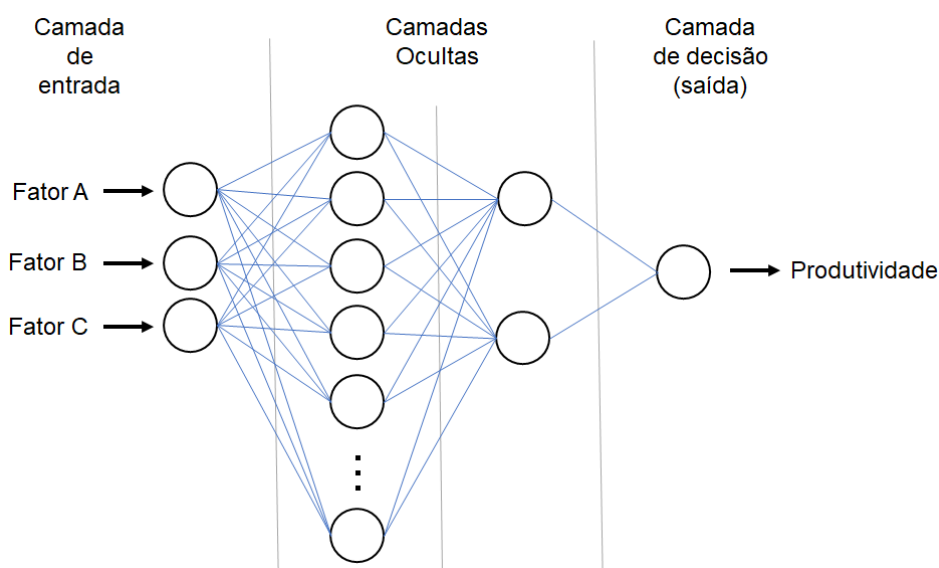
De acordo com a revisão sistemática de literatura realizada por Correa (2023), as técnicas de aprendizado de máquina utilizadas em estudos de prognóstico da produtividade na construção civil são: rede neural artificial, floresta aleatória, sistema de inferência fuzzy, algoritmo genético, máquina de vetores de suporte, algoritmo de otimização do gafanhoto, inferência bayesiana, k-ésimo vizinho mais próximo, sendo as duas primeiras aprofundadas e aplicadas nestas pesquisas.

Modelo de inteligência artificial de rede neural artificial

A rede neural artificial (RNA) é um sistema de computação inspirado no funcionamento do cérebro humano. O adjetivo “neural” é usado porque que uma RNA é composta por neurônios artificiais, que são módulos interconectados descritos por funções matemáticas. Cada neurônio recebe uma ou mais entradas, realiza um cálculo ponderado dessas entradas e aplica uma função de ativação para produzir uma saída (Sonmez; Rowings, 1998).

Uma RNA é organizada em camadas, sendo a camada de entrada responsável por receber os dados de entrada, as camadas intermediárias (também chamadas de camadas ocultas) processam as informações e a camada de saída produz o resultado final como, por exemplo, a produtividade, conforme apresentado na Figura 2.

Figura 2 - Diagrama esquemático de um modelo de rede neural artificial



Fonte: adaptado de Oral, Oral e Aydin (2012).

O funcionamento de uma RNA ocorre em duas fases principais: a fase de treinamento e a fase de inferência. Na fase de treinamento, a RNA é exposta a um conjunto de dados composto por pares de entradas e saídas desejadas. Utilizando um algoritmo de aprendizado, os pesos sinápticos entre os neurônios são ajustados de forma iterativa para minimizar a diferença entre as saídas produzidas pela RNA e as saídas desejadas. Após o treinamento, na fase de inferência, a RNA é capaz de receber novos dados de entrada e produzir saídas com base nos padrões aprendidos durante o treinamento. Cada neurônio realiza seu cálculo ponderado e aplica sua função de ativação, propagando as informações ao longo das camadas até chegar à camada de saída (Muqem *et al.*, 2011).

Segundo Khan (2005), o processo de aprendizado em uma RNA permite que ela seja capaz de reconhecer padrões complexos nos dados e realizar tarefas como classificação, regressão, reconhecimento de padrões, entre outras, dependendo da arquitetura e configuração da rede. Uma das principais vantagens da RNA é sua capacidade de lidar com dados não lineares e encontrar soluções aproximadas para problemas complexos.

Modelo de inteligência artificial de floresta aleatória

A Floresta Aleatória (FA) é um algoritmo de aprendizado de máquina que opera construindo uma coleção de árvores de decisão durante o treinamento (Momade *et al.*, 2020). Cada árvore de decisão é construída de forma independente, utilizando uma amostra aleatória do conjunto de dados de treinamento e um subconjunto aleatório das características (variáveis).

Cutler, Cutler e Stevens (2012) explicam que a unidade básica da FA é uma árvore binária construída utilizando a técnica de classificação e regressão denominada CART (sigla oriunda do termo inglês *Classification and Regression Tree*), no qual divisões binárias particionam recursivamente a árvore em nós terminais homogêneos ou quase homogêneos (as extremidades da árvore). Uma divisão binária adequada direciona os dados entre os nós, de modo que a homogeneidade resultante em nós primários seja aprimorada em nós seguintes. A FA tem como princípio a geração de uma coleção de várias árvores, onde cada árvore é desenvolvida usando uma amostra aleatória dos dados originais. Após a construção das árvores, a FA utiliza um processo de votação para fazer previsões. Para um novo exemplo de entrada, cada árvore de decisão faz uma previsão individual e o resultado final é determinado pela maioria das previsões.

Método

Para atingir os objetivos propostos, adotou-se como estratégia a divisão desta pesquisa em etapas, denominadas de Etapa 1 – Compreensão do tema; Etapa 2 – Seleção dos modelos e dos critérios de comparação; Etapa 3 – Aplicação dos modelos; Etapa 4 – Avaliação e comparação dos modelos; e Etapa 5 - Reflexão e contribuições. Na Figura 3 é apresentada a sequência destas etapas, com respectivas saídas.

Figura 3 - Fluxo de etapas da pesquisa

Etapas da pesquisa	Descrição	Saídas
Etapa 1	Compreensão do tema	Referencial teórico
Etapa 2	Seleção dos modelos e dos critérios de comparação	Seleção dos modelos e definição dos critérios de comparação
Etapa 3	Aplicação dos modelos	Resultados da aplicação dos modelos com dados de entrada
Etapa 4	Avaliação e comparação dos modelos	Análise comparativa dos modelos selecionados para estudo
Etapa 5	Reflexão e contribuições	Considerações finais sobre os resultados dos modelos aplicados

As informações para modelagem utilizadas neste estudo foram obtidas de uma pesquisa de campo realizada por Khan (2005). A escolha dos dados deste trabalho se deu pela sua disponibilidade de forma pública e pelo fato de o banco de dados conter uma amostra expressiva de pontos de observação, coletados de forma diária, além das variáveis passíveis de aplicação dos modelos estudados como os fatores que impactam a produtividade.

Foram realizadas observações de campo e coletas diárias de dados em dois empreendimentos de edifícios multipavimentos localizados na cidade de Montreal, Canadá. O período de observação foi de 18 meses e o objeto de estudo foi a atividade de montagem de formas, sendo coletados um total 221 pontos de dados. O levantamento de campo incluiu a medição de dados diários de produtividade, cuja métrica foi expressa pela relação de unidades de produção por horas trabalhadas (m^2/Hh). O estudo também realizou a medição de nove fatores considerados relevantes pelo autor em termos de impacto na produtividade, dividido em três grupos: fatores de clima, fatores de equipe e fatores de projeto. A descrição e unidade de medida destes fatores são apresentadas no Quadro 1. Os critérios para escolha destes fatores podem ser consultados com maior profundidade no estudo de Khan (2005).

Tendo em vista o uso mais frequente da Razão Unitária de Produção (RUP) no Brasil, foi realizada a conversão da métrica de produtividade para a relação de horas trabalhadas por unidades de produção por (Hh/m^2).

Para formulação e validação dos modelos, o conjunto de 221 dados existentes foi dividido em duas partes, sendo uma para treinamento e outra para teste. Conforme explicado por Flaten *et al.* (2018), é impossível acessar dados do futuro antes que eles ocorram, mas é possível reservar alguns dos dados atualmente disponíveis e tratá-los como se fossem dados posteriores. Com essa abordagem, o conjunto de dados original foi dividido aleatoriamente, de forma que 80% dos dados, correspondente a 177 observações, foram utilizados para treinamento e os 20% restantes como conjunto de teste. Os dados do conjunto de treinamento foram usados para regenerar modelos e os dados do conjunto de teste foram usados para avaliação comparando os valores observados em relação aos valores previstos.

Conforme descrito na seção anterior, os modelos de prognóstico podem ser classificados em três grandes grupos: modelos estatísticos, modelos de inteligência artificial e modelos de simulação. Dentre os modelos estatísticos, foram selecionados para este estudo as técnicas de regressão linear e regressão logística. De acordo com a literatura consultada e com os resultados desta pesquisa, a análise de regressão é o método mais comum utilizado em modelos estatísticos e, por esta razão, estas técnicas foram selecionadas. Além disso, as técnicas de análise de regressão podem ser aplicadas com o uso de planilhas eletrônicas comerciais e de fácil acessibilidade.

Quadro 1 - Fatores de produtividade do estudo

Fator	Descrição	Unidade
Temperatura	Temperatura média de oito horas diárias de trabalho	°C
Umidade relativa	Umidade relativa média de oito horas diárias de trabalho	%
Velocidade do vento	Velocidade de vento média de oito horas diárias de trabalho	km/h
Precipitação	Valores numéricos atribuídos (sem precipitação = 0; chuva leve = 1; neve = 2; chuva = 3)	--
Tamanho da Equipe	Número de pessoas na equipe	Unidade
Porcentagem de mão de obra sem qualificação	Relação entre número de profissionais não qualificados e número de pessoas na equipe	%
Tipo de trabalho	Valores numéricos atribuídos ao elemento da forma (lajes = 1; paredes = 2; pilares = 3)	--
Nível do piso (andar)	Valor numérico do andar de trabalho	--
Método de trabalho	Valores numéricos atribuídos à técnica de trabalho em função do elemento (forma tradicional = 1; forma suspensa = 2)	--

No grupo de modelos de inteligência artificial, foram selecionadas técnicas de aprendizado de máquina supervisionados, ou seja, aqueles em que variável de resposta é pré-definida e os algoritmos desenvolvem um modelo de função preditiva que relacionada os dados de entrada com o objetivo previsto (variável de resposta). No caso específico deste estudo, a produtividade foi a variável definida a ser prevista em função dos fatores de conteúdo e contexto. Dessa forma, os seguintes algoritmos de aprendizado de máquina foram utilizados: rede neural artificial e floresta aleatória.

Os modelos de simulação não foram utilizados para comparação neste trabalho, pois requerem uma coleta de dados de campos diferente da realizada nos dados de entrada deste estudo e são aplicados por meio de ferramentas específicas de modelagem.

A aplicação do modelo estatístico de regressão linear ocorreu com a utilização do software *Microsoft Excel*, haja vista se tratar de uma ferramenta comercial de planilha eletrônica de fácil acessibilidade. Essa ferramenta possui uma função de análise de dados que realiza a análise de regressão linear com base nos dados fornecidos e exibe os resultados em uma determinada célula especificada. Como resultados, são exibidos os dados de estatística de regressão, os coeficientes da equação, a análise de variância (ANOVA), a análise da influência de variáveis com base no *valor-P* e o histograma de resíduos dos dados do modelo (Correa, 2023).

Para aplicação dos modelos de inteligência artificial e de regressão logística, foi utilizada a plataforma de aprendizado de máquina *BigML* (BIGML, 2023), cuja utilização é realizada por meio de qualquer navegador de *internet*. Os cálculos foram realizados nos próprios servidores da plataforma e os dados armazenados em nuvem. Esta ferramenta foi selecionada para a pesquisa devido a sua praticidade de uso, além de ser gratuita para o tamanho dos dados utilizados nas modelagens deste estudo.

Resultados e discussões

Caracterização e preparação dos dados de entrada

Os dados utilizados neste estudo são compostos por um total 221 pontos de coletas diárias. Os dados de entrada de precipitação, tipo de trabalho e método de trabalho foram submetidos a um critério para conversão numérica, conforme indicado no Quadro 1, pois os modelos estatísticos admitem como entrada apenas valores quantitativos. Nos modelos de inteligência artificial foram mantidas as formas originais, visto que possuem a capacidade de processar qualquer tipo de dado.

Na Tabela 1 é apresentada a estatística descritiva dos dados utilizados, que oferecem uma síntese do conjunto de estudo. Observa-se que os valores apresentam uma variabilidade ampla para todos os fatores, traduzindo as condições dinâmicas e as relações complexas presentes em canteiros de obras da construção civil.

Modelo estatístico de regressão linear

Os dados selecionados para este estudo foram utilizados para o prognóstico da produtividade com o uso de modelo estatístico de regressão linear. Na Tabela 2 são apresentados o coeficiente de determinação (R^2) e coeficiente de determinação ajustado (R_a^2), que indicam a precisão geral do modelo, ou seja, o quanto o modelo é capaz de explicar a variabilidade da variável independente (produtividade) em função dos fatores de entrada. Para os dados em questão, o valor de R^2 do modelo é de 53,66% e o de R_a^2 igual a 51,16%.

Tabela 1 - Estatística descritiva dos dados de produtividade dos serviços de formas

Variável	Média	Erro padrão	Desvio padrão	Mediana	Mínimo	Máximo	Intervalo	Nº de dados
Temperatura	4,1	0,8	12,0	3,0	-26,0	25,0	51,0	221
Umidade Relativa	66,3	1,1	15,7	67,0	18,0	97,0	79,0	221
Velocidade de vento	15,4	0,6	8,5	14,0	3,0	43,0	40,0	221
Tamanho da equipe	16,0	0,3	5,1	18,0	8,0	24,0	16,0	221
MO sem qualificação	35,5	0,3	3,8	36,0	29,0	47,0	18,0	221
Precipitação	--	--	--	--	0,0	3,0	3,0	221
Tipo de trabalho	--	--	--	--	1,0	3,0	2,0	221
Andar	--	--	--	--	1,0	17,0	16,0	221
Método de trabalho	--	--	--	--	1,0	2,0	1,0	221
Produtividade (RUP)	0,67	0,01	0,15	0,66	0,40	1,22	0,82	221

Tabela 2 - Estatística de regressão dos dados do modelo estatístico de regressão linear

Descrição	Coefficiente
Coefficiente de determinação (R^2)	0,5366
Coefficiente de determinação ajustado (R_a^2)	0,5116
Erro padrão	0,1057
Observações	177

O valor F de significação, demonstrado na Tabela 3, indicou resultado menor do que 5%, o que significa que pelo menos uma variável independente influencia a variável resposta, para um nível de confiança de 95%.

Na Figura 4 é apresentado o histograma dos resíduos do modelo. Valores de resíduos padronizados próximos a zero indicam que os pontos são distribuídos aleatoriamente e a suposição de homoscedasticidade é observada e, portanto, as premissas da análise de regressão linear estão atendidas.

Na Tabela 4 são apresentados os coeficientes da equação de regressão linear gerada. Os sinais dos coeficientes dos fatores indicam uma relação direta ou inversa com a variável de resposta. Quanto ao *valor-P* de significação, o resultado menor que 5% indica que a variável independente em questão influencia a variável resposta, para um nível de confiança de 95%. Assim, segundo este modelo e para os dados em questão, observa-se que apenas as variáveis temperatura, velocidade do vento e tamanho da equipe tem influência significativa na produtividade.

A análise de sensibilidade das variáveis do modelo foi realizada a partir dos coeficientes dos fatores, conforme apresentado na Figura 5. Nesta forma de apresentação, a variação de um dos fatores sobre a produtividade é verificada quando os demais são mantidos constantes. As inclinações das retas dos gráficos retratam a influência de cada uma das variáveis, sendo as inclinações mais acentuadas aquelas de maior influência sobre a produtividade. Como cada variável possui sua própria unidade, o gráfico é representado de forma normalizada, sendo o eixo horizontal traçado por meio dos valores de 0,01 a 1. Observa-se que os fatores indicados com influência significativa na produtividade (temperatura, velocidade do vento e tamanho da equipe) apresentam curvas mais inclinadas, enquanto nos demais as retas têm configuração horizontal ou mais suaves.

Modelo estatístico de regressão logística

Conforme apresentado na seção anterior, o modelo gerado pela regressão logística tem por função calcular ou prever a probabilidade de um evento ocorrer, dado um conjunto de fatores de entrada. Neste tipo de modelo, a variável de resposta é binária ou dicotômica e, portanto, a forma de aplicação é diferente dos demais modelos apresentados neste estudo.

Visto que a produtividade é uma variável numérica contínua, uma forma de conversão da variável de resposta para aplicação nesse modelo pode ser feita a partir do estabelecimento de um valor de referência para a Razão Unitária de Produção (RUP), abaixo do qual a produtividade é considerada satisfatória e acima representa uma condição imprópria. Para este estudo, o valor de referência atribuído para a RUP foi de 0,7 Hh/m², sendo a produtividade observada abaixo desse valor classificada como “SIM” (ou seja, o valor da produtividade está dentro dos padrões esperados) e, acima desta referência, classificados como “NÃO” (ou seja, a RUP ultrapassa o valor limite de referência e não corresponde ao esperado).

Este modelo tem como resultado principal a probabilidade de ocorrência de um evento a partir de parâmetros de entrada estabelecidos, conforme apresentado na Figura 6(a), sendo representada no eixo “x” a temperatura de trabalho e no eixo “y” a probabilidade de ocorrência do evento. Assim, observa-se que a probabilidade de a RUP ser menor do que 0,7 Hh/m² para um cenário de temperatura próximo de 5 °C é de 80,94%, conforme apresentado na Figura 6(b).

Tabela 3 - Análise de variância (ANOVA) dos dados do modelo estatístico de regressão linear

Dimensão	gl	SQ	MQ	F	F de significação
Regressão	9	2,1594	0,2399	21,4845	7,5723 ⁻²⁴
Resíduo	167	1,8650	0,0112	–	–
Total	176	4,0243	–	–	–

Figura 4 - Histograma de resíduos dos dados do modelo

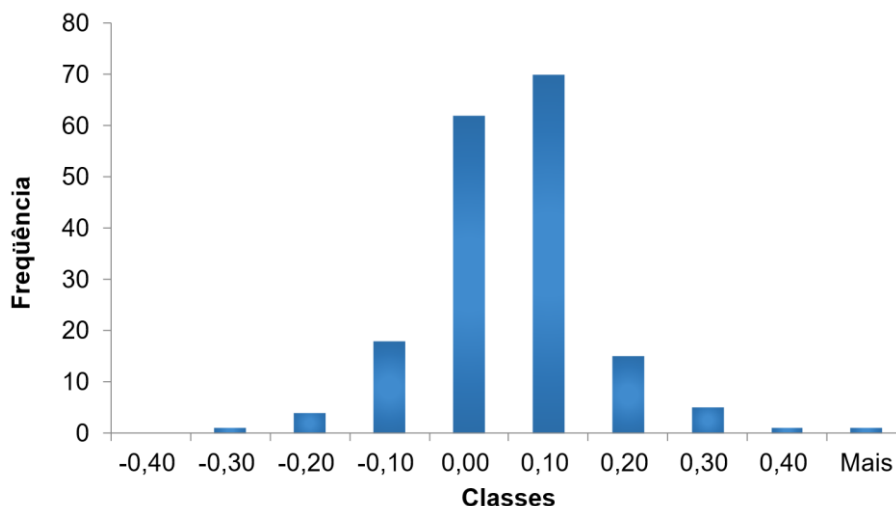


Tabela 4 - Coeficientes da regressão linear e análise da influência das variáveis com base no valor-P

Variável	Coefficiente	Erro padrão	Estatística t	valor-P
Interseção	0,3853	0,1378	2,7967	0,6%
Temperatura (°C)	-0,0088	0,0009	-9,5397	0,0%
Umidade Relativa (%)	0,0010	0,0006	1,6736	9,6%
Precipitação	0,0071	0,0158	0,4501	65,3%
Velocidade de vento (km/h)	0,0048	0,0011	4,4989	0,0%
Tamanho da equipe	0,0078	0,0024	3,2215	0,2%
MO sem qualificação (%)	0,0003	0,0023	0,1512	88,0%
Tipo de trabalho	0,0479	0,0268	1,7895	7,5%
Andar	0,0025	0,0030	0,8312	40,7%
Método de trabalho	-0,0389	0,0270	-1,4397	15,2%

Figura 5 - Análise de sensibilidade normalizada das variáveis de entrada para o modelo de regressão linear

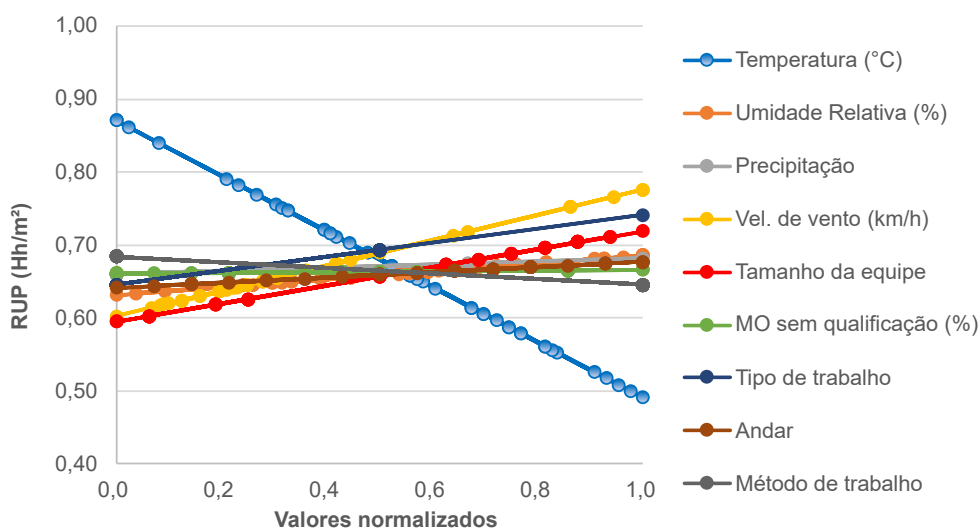
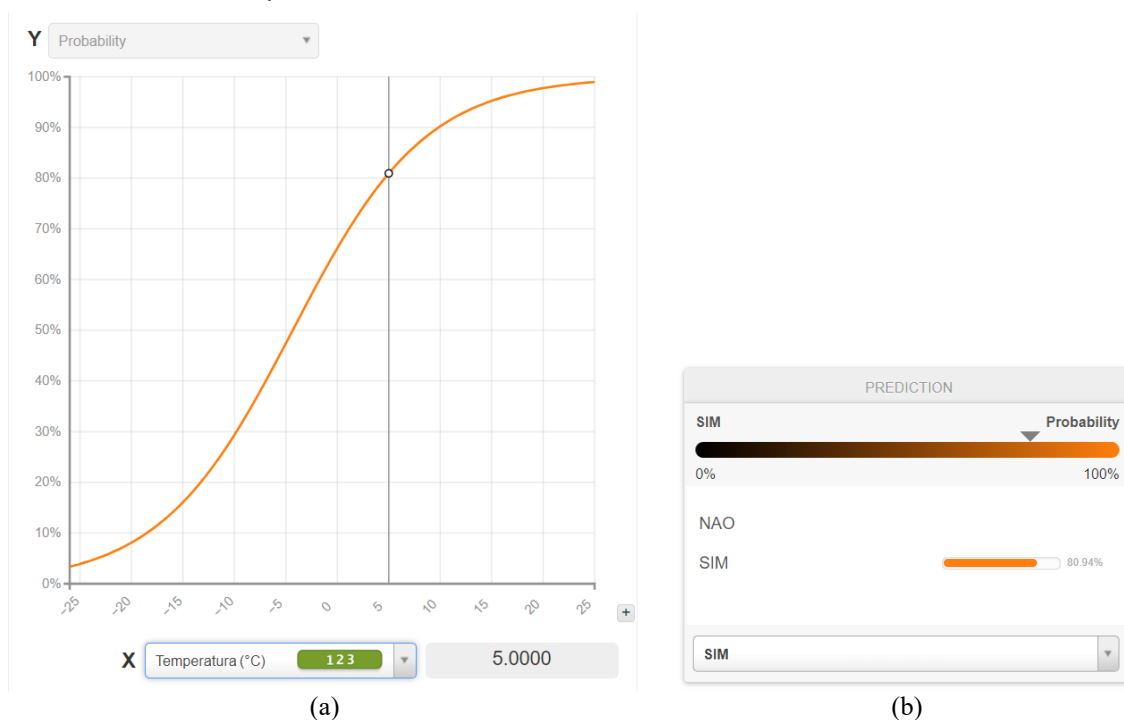


Figura 6 - (a) Representação gráfica do modelo de regressão logística; (b) probabilidade de ocorrência do evento de referência.



O modelo de regressão logística também permite uma análise de sensibilidade de suas variáveis, conforme apresentado na Figura 7. Nesta forma de apresentação, a probabilidade de ocorrência do evento de referência é demonstrada em função da variação de um dos fatores, enquanto os demais são mantidos constantes. As curvas de probabilidade dos gráficos retratam a influência de cada uma das variáveis, sendo as inclinações mais acentuadas aquelas de maior influência sobre a produtividade. Observa-se que os fatores de temperatura e velocidade do vento apresentam curvas mais inclinadas, enquanto nos demais as curvas têm configuração horizontal ou são mais suaves. Os gráficos também possibilitam a interpretação quanto à correlação positiva ou negativa de um fator com a variável independente.

Como a variável de resposta neste caso é binária e não numérica, não é aplicável o cálculo dos parâmetros de desempenho dos outros modelos estudados. O indicador de desempenho pertinente a esta situação é o percentual de acertos (índice calculado pela relação do número de valores previstos corretamente pelo modelo em relação ao número total de dados de teste), conforme apresentado na Figura 8. Para os dados do estudo em questão, observa-se um alto grau de assertividade do modelo, com valor acima de 80% de acerto.

Modelo de inteligência artificial de rede neural artificial

Os dados selecionados para este estudo foram utilizados para o prognóstico da produtividade com o uso de modelo de inteligência artificial de rede neural artificial. Uma informação de saída relevante resultante deste modelo é o nível de influência das variáveis independentes sobre a variável de resposta, conforme apresentado na Figura 9. Para os dados do estudo em questão, o modelo apontou as variáveis temperatura e velocidade como as de maior grau de influência sobre a produtividade.

Figura 7 - Análise de sensibilidade das variáveis de entrada para o modelo de regressão logística

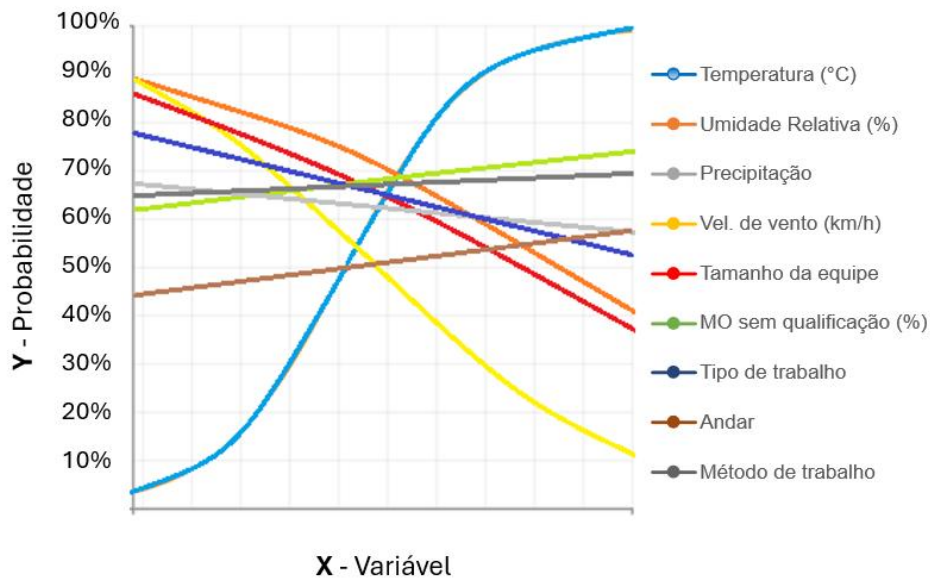


Figura 8 - Porcentagem de acertos e erros do modelo de regressão logística

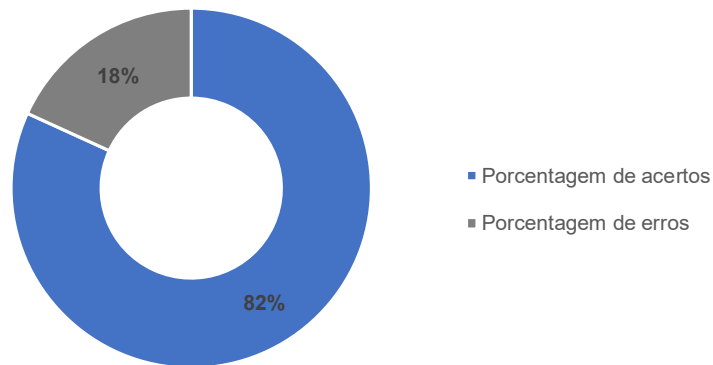
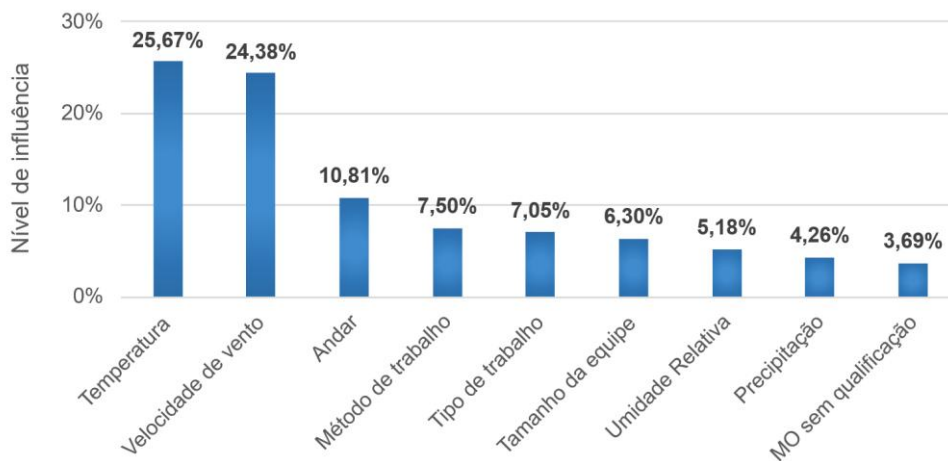


Figura 9 - Nível de influência das variáveis independentes apontadas pelo modelo de inteligência artificial de rede neural artificial



Assim como nos modelos estatísticos, esta ferramenta utilizada para modelagem de rede neural artificial também permite a análise de sensibilidade de suas variáveis, com visualização em pares. Na Figura 10(a) é representada a sensibilidade do modelo para as duas variáveis independentes apontadas como de maior influência sobre a variável de resposta – de forma bidimensional, as cores azul e verde representam os valores extremos. Na Figura 10(b) é indicado no lado direito o prognóstico da produtividade para os valores de temperatura e velocidade do vento estabelecidos.

Modelo de inteligência artificial de floresta aleatória

Esta técnica utiliza um método de aprendizado conjunto para prognóstico a partir do desenvolvimento de uma grande quantidade de árvores de decisões no momento do treinamento. Para os dados analisados em questão, a ferramenta realizou a construção de 59 configurações de árvores de decisões. Assim, a partir dos parâmetros de entrada estabelecidos, cada árvore de decisão faz uma previsão individual e o resultado final é determinado pela maioria das previsões. Na Figura 11 é apresentado um exemplo das 59 configurações geradas.

Assim como no modelo de redes neurais, a ferramenta realiza o cálculo do nível de influência das variáveis independentes sobre a variável de resposta, conforme apresentado nas Figuras 12 e 13. Para os dados do estudo em questão, o modelo apontou a variável temperatura como a de maior grau de influência sobre a produtividade. Conforme representado na Figura 13(a), também é possível a análise de sensibilidade de suas variáveis, com visualização em pares de forma bidimensional, em que as cores azul e verde representam os valores extremos. Na Figura 13(b) é demonstrado o prognóstico da produtividade quando utilizada forma de concreto suspensa e em local de temperatura ambiente de 23 °C.

Figura 10 - (a) Análise de sensibilidade das variáveis de entrada temperatura e velocidade do vento para o modelo de rede neural artificial; (b) prognóstico de produtividade para os valores de temperatura e velocidade do vento estabelecidos.

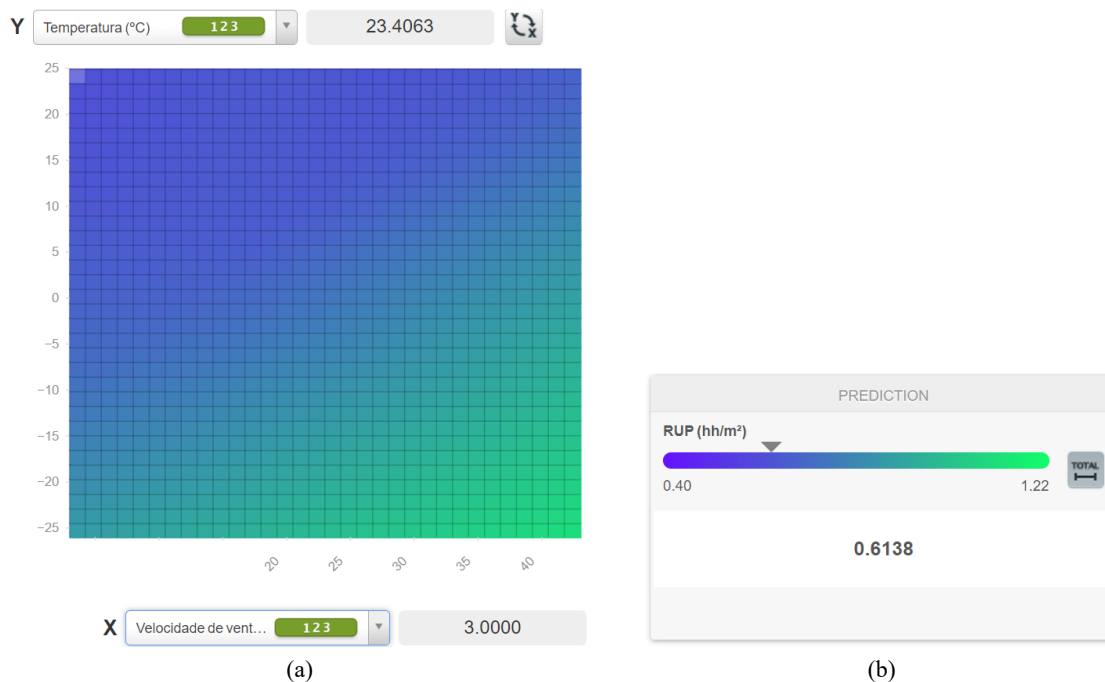


Figura 11 - Configuração de uma árvore de decisões gerada pelo modelo

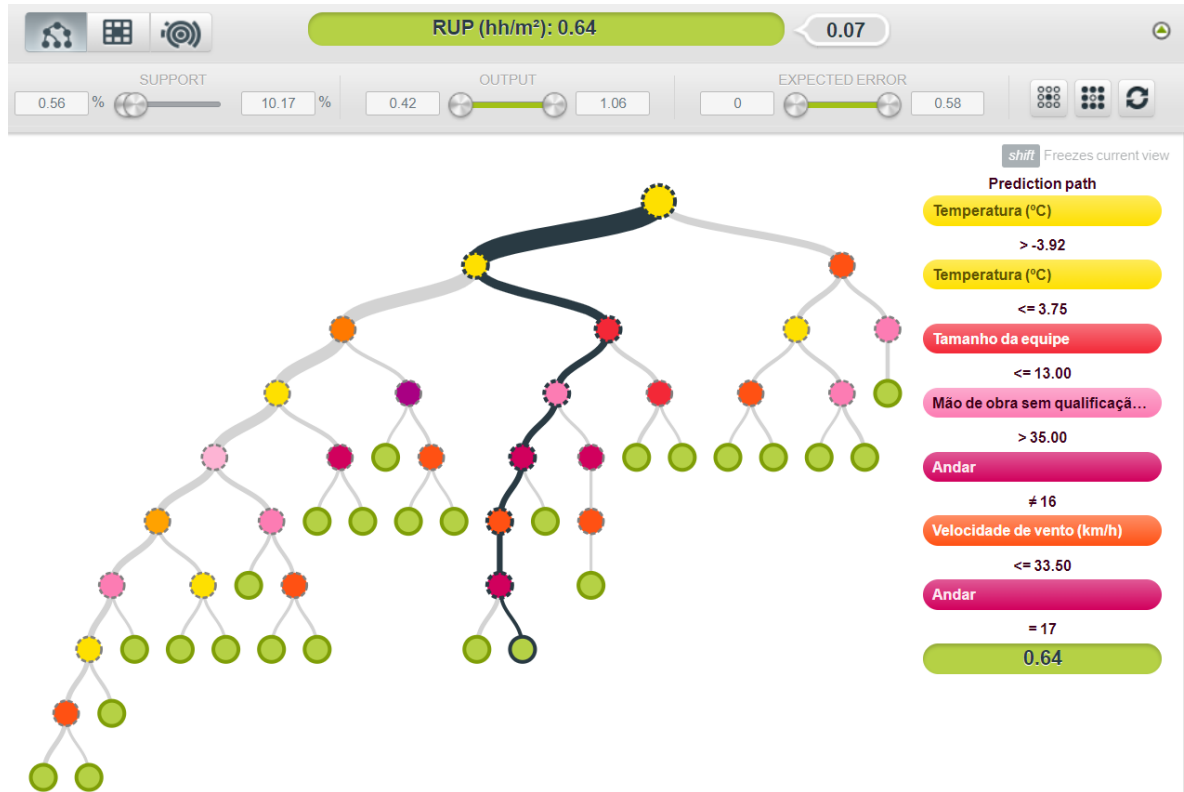


Figura 12 - Nível de influência das variáveis independentes apontadas pelo modelo de inteligência artificial de floresta aleatória

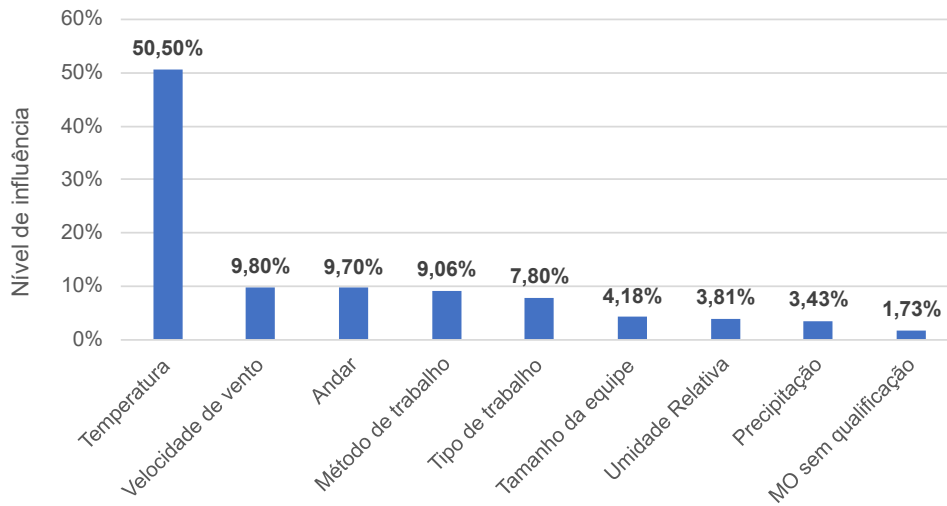
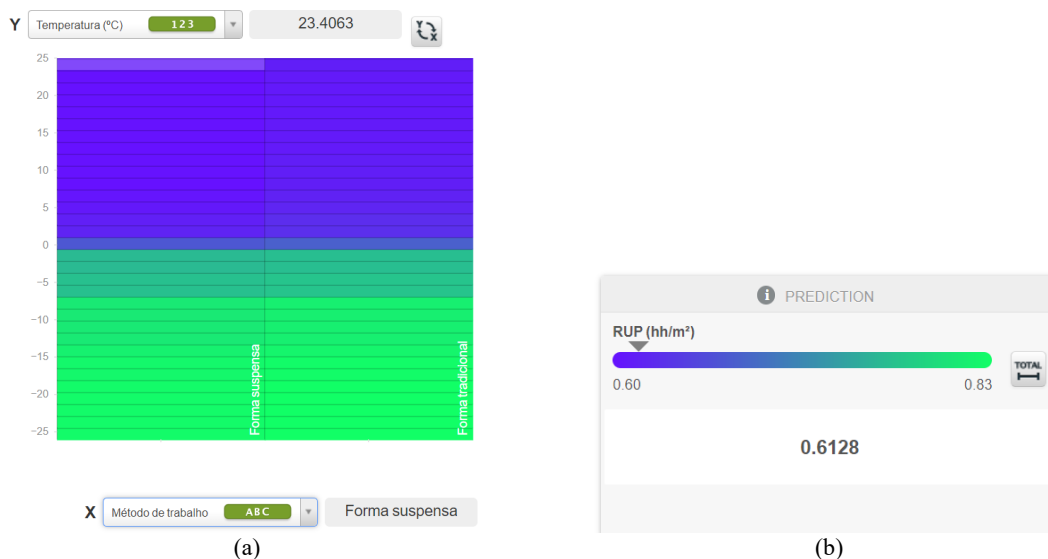


Figura 13 - (a) Análise de sensibilidade das variáveis de entrada temperatura e método de trabalho para o modelo de floresta aleatória; (b) prognóstico de produtividade para os valores de temperatura e método de trabalho estabelecidos.



Comparação dos modelos quanto à precisão de prognóstico

Para comparação de desempenho do modelo, o prognóstico dos valores de produtividade foi realizado utilizando-se os 20% dos dados reservados para teste. Dessa forma, os modelos aplicados neste estudo foram comparados quanto a sua precisão.

O principal aspecto de desempenho de um modelo de prognóstico está relacionado à sua capacidade de fornecer previsões precisas e consistentes. Isso implica em verificar se o modelo consegue capturar as tendências e padrões presentes nos dados de entrada e gerar resultados coerentes com os valores reais observados. Na Tabela 5 são apresentados os parâmetros de desempenhos dos modelos, destacando-se os valores de melhor performance, e, na Figura 14, o percentual de assertividade dos modelos considerando cinco faixas de margens de erro estipuladas.

Observa-se que alguns parâmetros de desempenho são específicos para determinados modelos, enquanto outros são comuns. Para os modelos selecionados, não foi possível encontrar nenhum parâmetro em comum para uma avaliação simultânea, o que pode dificultar a identificação daquele de melhor desempenho.

Para contornar esta situação, algumas alternativas de comparação são possíveis. A primeira alternativa é realizar a comparação aos pares, buscando avaliar apenas modelos que possuem parâmetros em comum. Assim, observa-se que os modelos de inteligência artificial apresentam indicadores melhores do que o modelo de regressão linear. De forma análoga, dentre os modelos de inteligência artificial, observa-se que o modelo de floresta aleatória apresenta indicadores melhores do que o modelo de rede neural artificial.

Uma segunda opção é selecionar uma única métrica de desempenho que melhor atende às necessidades específicas de uma situação. Por exemplo, caso seja considerado que o melhor modelo seja aquele com maiores acertos de previsão para uma margem de erro de até 10% (o que seria importante para minimizar problemas de orçamento), o modelo mais preciso é o de floresta aleatória, pois possui índice de assertividade de 64% para este conjunto de dados, conforme apresentado na Tabela 5 e na Figura 14.

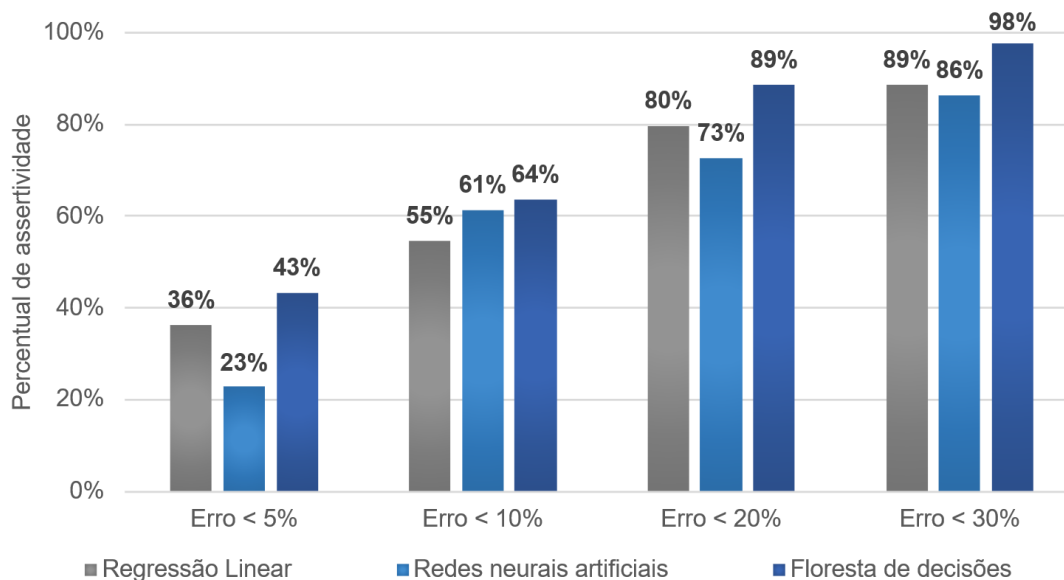
Outra possibilidade é comparar o desempenho de um tipo de modelo com valores de referência da literatura. Segundo Amr, Mona e Hesham (2019), valores R^2 ajustado maiores do que 70% denotam que o modelo de regressão linear apresenta uma forte correlação entre a variável dependente e as variáveis independentes. Para modelos de inteligência artificial, são desejáveis valores do erro absoluto médio (EAM) e erro quadrado médio (EQM) próximos a zero (Khan, 2005). Entretanto, é importante ressaltar que os valores de referência para os indicadores de desempenho podem variar de acordo com a natureza do problema, o conjunto de dados utilizado e outros fatores de contexto específicos.

Tabela 5 - Resumo dos parâmetros quantitativos de desempenho dos modelos deste estudo, destacado em cinza os valores de melhor performance

Parâmetros de desempenho	Regressão linear	Regressão logística	Rede neural artificial	Floresta aleatória
R-Quadrado	0,5366	NA	NA	NA
R-Quadrado Ajustado	0,5116	NA	NA	NA
Erro absoluto médio	0,0710	NA	0,0795	0,0569
Erro absoluto percentual médio	12,34%	NA	13,47%	8,93%
Erro quadrado médio	0,0090	NA	0,0107	0,0067
Raiz do erro quadrado médio	0,0947	NA	0,1000	0,0800
Erro mínimo absoluto	0,0002	NA	0,0008	0,0016
Erro máximo absoluto	0,2634	NA	0,2552	0,3204
Erro relativo até 5%	36%	NA	23%	43%
Erro relativo até 10%	55%	NA	61%	64%
Erro relativo até 20%	80%	NA	73%	89%
Erro relativo até 30%	89%	NA	86%	98%
Porcentagem de acertos	NA	82%	NA	NA

Nota: NA - Não aplicável.

Figura 14 - Percentual de assertividade dos modelos considerando faixas de margens de erro

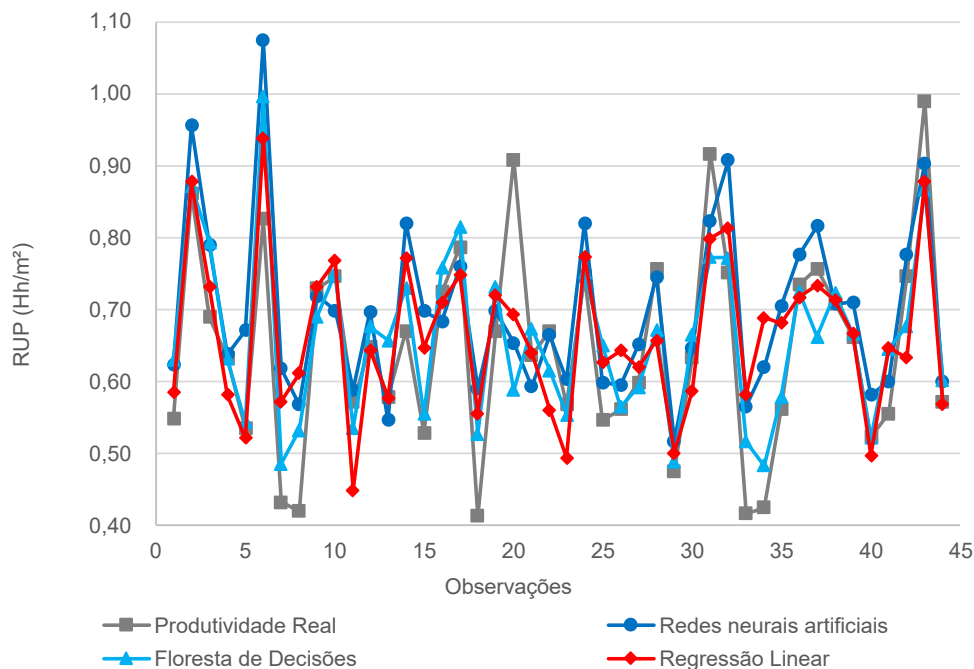


A análise gráfica também pode auxiliar na comparação dos modelos. Na Figura 15 estão representados os valores reais em relação aos previstos pelos modelos aplicados neste estudo. Observa-se que não há nenhuma discrepância significativa que possa provocar algum ponto de atenção dentre os modelos. No entanto, percebe-se que, todos os modelos têm como característica a dificuldade de realizar prognósticos precisos quando os valores reais apresentam valores extremos. Este fato pode ser constatado pelo descolamento dos pontos em cinza (valores reais) quando comparados pontos coloridos (previstos pelos modelos) nos casos em que a produtividade alcança valores máximos e mínimos.

Comparação dos modelos quanto à influência e sensibilidade das variáveis

Compreender o nível de influência das variáveis de entrada na produtividade e identificar quais fatores têm maior impacto nos resultados são aspectos fundamentais em um modelo. Todos os modelos analisados neste estudo apresentam uma avaliação quanto ao nível de influência, porém com formas diferentes de expressão.

Figura 15 - Valores reais x previstos pelos modelos aplicados neste estudo



No modelo de regressão linear, é apresentado o resultado de *valor-P* de significação, cujo valor menor que 5% indica que a variável independente em questão influencia a variável resposta, para um nível de confiança de 95% (Tabela 4). Nos modelos de inteligência artificial, é possível realizar o cálculo do nível de influência das variáveis independentes sobre a variável de resposta, conforme apresentados nas Figuras 9 e 12. No modelo de regressão logística, o grau de influência é representado pelas curvas de probabilidade de ocorrência da variável de resposta de referência, conforme Figura 6.

Por outro lado, os resultados do nível de influência dos fatores podem variar de acordo com o modelo adotado, conforme observado neste trabalho. O modelo de regressão linear indicou como principais influentes as variáveis de temperatura, velocidade do vento e tamanho da equipe; no modelo de regressão logística e no modelo de rede neural artificial foram apontadas temperatura e velocidade do vento; enquanto no modelo de floresta aleatória, apenas a temperatura. Os valores dos níveis de influência encontrados também são diferentes entre os modelos estudados. Assim, é importante avaliar se os resultados do modelo são consistentes com a experiência prática ou com outros estudos semelhantes.

Além de identificação das variáveis mais influentes, também deve ser considerada a direção e magnitude do efeito dos fatores de entrada. Essa análise permite compreender como cada variável influencia a produtividade e em que medida. A direção do efeito das variáveis indica se elas têm uma relação positiva ou negativa com a produtividade. Uma variável com efeito positivo aumenta a produtividade quando seus valores aumentam, enquanto uma variável com efeito negativo reduz a produtividade à medida que seus valores aumentam.

Os modelos estatísticos, tanto de regressão linear quanto de regressão logística, apresentam essa característica bem evidente. Os sinais (positivo ou negativo) e os valores dos coeficientes da regressão linear indicam a direção e magnitude do efeito dos fatores de entrada. As direções das curvas e as probabilidades associadas fazem este papel na regressão logística. Nos modelos de inteligência artificial, não é possível identificar diretamente de que forma as variáveis de entrada influenciam nos resultados.

A análise de sensibilidade das variáveis de entrada visa compreender o impacto e a influência de cada variável na previsão da produtividade e ajuda a identificar quais variáveis têm maior ou menor efeito nos resultados. Neste estudo, essa análise foi realizada para todos os modelos, conforme Figuras 5, 7, 10 e 13. As análises realizadas para o modelo de regressão linear foram univariadas, ou seja, cada variável foi alterada individualmente enquanto as demais se mantiveram constantes, permitindo identificar a magnitude do efeito de um fator isoladamente. Nos demais modelos foi possível a análise bivariada, ou seja, com avaliação dos fatores em pares.

Tanto a análise do nível de influência quanto da sensibilidade das variáveis pode ser útil para simplificação dos modelos, uma vez que é possível ocultar ou mesmo eliminar fatores considerados não significativos para a modelagem. Além de simplificar o modelo, esta abordagem pode resultar em uma menor necessidade de recursos computacionais para processamento dos dados e na redução de esforços e de recursos necessários para levantamento dos dados de campo. Contudo, este processo de simplificação deve ser realizado com base em critérios estatísticos claramente definidos (Farias; Soares; César, 2008).

Embora os modelos de prognóstico representem uma importante ferramenta para a estimativa de produtividade na construção civil, seu uso isolado não constitui, por si só, uma solução definitiva para os desafios presentes nesse setor. A eficácia dos modelos depende fortemente da qualidade dos dados de entrada, da representatividade da amostra e da escolha criteriosa dos fatores considerados. É fundamental que os fatores selecionados apresentem relação causal coerente com a produtividade e que sejam, na medida do possível, estatisticamente independentes, a fim de evitar distorções nos resultados devido a interações entre variáveis.

Além disso, a utilização de modelos preditivos não substitui uma análise crítica e aprofundada dos elementos que influenciam a produtividade. A interpretação dos resultados deve ser feita com cautela, preferencialmente por especialistas, uma vez que conclusões equivocadas podem ser geradas se os modelos forem aplicados de forma acrítica. Portanto, é necessário destacar as limitações inerentes ao uso dessas ferramentas e ressaltar que sua aplicação exige não apenas domínio técnico, mas também discernimento na análise dos resultados gerados.

Conclusões

Neste estudo foram aplicadas e avaliadas quatro técnicas de modelagem com vistas a prognóstico de produtividade, sendo dois modelos estatísticos e dois modelos de inteligência artificial. Com base em fatores de entrada de conteúdo e contexto considerados significativos para serviços de execução de formas de concreto em canteiros de obras, o prognóstico da produtividade foi realizado e os diferentes modelos empregados foram avaliados e comparados quanto aos seus desempenhos.

A caracterização inicial dos dados de entrada demonstrou a alta variabilidade dessas grandezas, refletindo as condições dinâmicas e as relações complexas presentes em canteiros de obras da construção civil. Ressalta-se, ainda, que esses fatores representam apenas uma parte de um conjunto de situações que podem influenciar na produtividade. Cada obra de construção civil é única, com diferentes condições geográficas, climáticas, econômicas e sociais. Fatores definidos pela própria empresa também são essenciais, como condições de trabalho, logística e gestão de recursos.

Importante ressaltar que a comparação do desempenho e das propriedades dos modelos de prognóstico de produtividade foi realizada apenas com base na aplicação dos dados deste estudo; portanto, limitam-se aos valores dos parâmetros de entrada utilizados e aos modelos selecionados, sem o propósito de testar e avaliar todos os modelos existentes, tampouco exaurir um assunto tão extenso e complexo. Da mesma forma, os dados de entrada utilizados nos modelos consideraram apenas fatores específicos de conteúdo e de contexto. Não foram considerados aspectos de gestão e de logística próprios da empresa construtora que realizou as atividades. Além disso, tratando-se de uma pesquisa conduzida em outro país, os dados não refletem necessariamente a realidade da construção civil no Brasil.

A compreensão do grau de influência das variáveis de entrada na produtividade e a identificação dos fatores que têm maior impacto nos resultados são elementos essenciais em um modelo. Todos os modelos utilizados neste estudo possuem essa característica, embora se manifestem de maneiras distintas. A análise tanto do nível de influência quanto da sensibilidade das variáveis pode ser benéfica para simplificar os modelos, uma vez que é possível ocultar ou até mesmo eliminar fatores considerados não significativos para a modelagem. Além de simplificar o modelo, essa abordagem pode resultar em uma redução na necessidade de recursos computacionais para processar os dados, bem como diminuir os esforços e recursos necessários para coletar dados de campo.

Finalmente, a precisão do modelo é uma aferição de quão próximo está o valor previsto do valor real e quão consistentes são os prognósticos realizados. Neste trabalho foi realizado o cálculo de indicadores de desempenho encontrados na literatura para fins de avaliação e comparação. Os resultados demonstram que nem sempre é possível encontrar em um único modelo os melhores parâmetros de assertividade, o que pode implicar na necessidade de seleção de uma ou mais métricas que melhor representem os objetivos específicos de um projeto. Também foram destacadas limitações dessas ferramentas e da importância de interpretações embasadas dos resultados, preferencialmente com apoio de especialistas, a fim de evitar conclusões equivocadas.

De forma resumida, as principais contribuições desta pesquisa foram a comparação prática de diferentes técnicas e a consolidação de informações fragmentadas existentes na literatura, preenchendo uma lacuna importante para os estudos de produtividade na construção civil.

Referência

- AMR, F. H.; MONA, M. A.; HESHAM, M. O. Optimizing labor productivity in Egypt using regression prediction models. In: CSCE ANNUAL CONFERENCE, Laval, 2019. **Proceedings [...]** Laval, 2019.
- AZIZ, M. A. E.; EID, M. S.; AMER, N. Concrete pouring production rate estimation: a bayesian network approach. Em: CONSTRUCTION RESEARCH CONGRESS, Tempe, 2020. **Proceedings [...]** Tempe: American Society of Civil Engineers, 2020.
- BATISTA, A. S. **Regressão logística: uma introdução ao modelo estatístico**. Porto: Vida Economica Editorial, 2015.
- BHILWADE, V. *et al.* Identification of factors significantly influencing the labour productivity of formwork systems using artificial neural network. In: INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ADVANCEMENTS IN COMPUTATIONAL SCIENCES, 2., Tashkent, 2022. **Proceedings [...]** Tashkent: IEEE, 2022.
- BIGML. **Comprehensive Machine Learning Platform**. Disponível em: <http://bigml.com/>. Acesso em: 06 jul. 2023.
- CORREA, M. I. F. **Comparação de modelos de prognóstico de produtividade aplicados na construção civil**. Goiânia, 2023. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal de Goiás, Goiânia, 2023.
- COTTRELL, D. S. Contractor process improvement for enhancing construction productivity. **Journal of Construction Engineering and Management**, v. 132, n. 2, p. 189–196, 2006.
- CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random forests. **Ensemble Machine Learning**, p. 157–175, 2012.
- DIXIT, S. *et al.* Evolution of studies in construction productivity: a systematic literature review (2006–2017). **Ain Shams Engineering Journal**, v. 10, n. 3, p. 555–564, 2019.
- DORNELAS, R. C. **Estudo de métodos para prognóstico da produtividade na execução de rodovias: terraplenagem e pavimentação asfáltica, uma nova abordagem**. São Paulo, 2013. Tese (Doutorado em Engenharia Civil) - Universidade de São Paulo, São Paulo, 2013.
- EBRAHIMI, S.; FAYEK, A. R.; SUMATI, V. Hybrid Artificial Intelligence HFS-RF-PSO Model for Construction Labor Productivity Prediction and Optimization. **Algorithms**, v. 14, n. 7, p. 18, 2021.
- FARIAS, A. A.; SOARES, J. F.; CÉSAR, C. C. **Introdução à Estatística**. São Paulo: LTC, 2008.
- FLATEN, B. T. *et al.* Estimating productivity rates for establishing contract time. In: CONSTRUCTION RESEARCH CONGRESS, New Orleans, 2018. **Proceedings [...]** New Orleans: American Society of Civil Engineers, 2018.
- GURMU, A. T. Tools for measuring construction materials management practices and predicting labor productivity in multistory building projects. **Journal of Construction Engineering and Management**, v. 145, n. 2, p. 04018139, 2019.
- KHAN, Z. **Modeling and parameter ranking of construction labor productivity**. Montreal, 2005. Thesis (Masters) - Concordia University, Montreal, 2005.
- MARCHIORI, F. F. **Desenvolvimento de um método para elaboração de redes de composições de custo para orçamentação de obras de edificações**. São Paulo, 2009. Tese (Doutorado em engenharia Civil) - Universidade de São Paulo, São Paulo, 2009.
- MOHAMED, S.; SRINAVIN, K. Forecasting labor productivity changes in construction using the PMV index. **International Journal of Industrial Ergonomics**, v. 35, n. 4, p. 345–351, 2005.
- MOHSENIJAM, A.; LU, M. Framework for developing labour-hour prediction models from project design features: case study in structural steel fabrication. **Canadian Journal of Civil Engineering**, v. 46, n. 10, p. 871–880, 2019.

- MOMADE, M. H. *et al.* Modelling labour productivity using SVM and RF: a comparative study on classifiers performance. **International Journal of Construction Management**, p. 1–11, 2020.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. New York: John Wiley & Sons, 2015.
- MUQEEM, S. *et al.* Construction labor production rates modeling using artificial neural network. **Journal of Information Technology in Construction**, v. 16, p. 713–726, 2011.
- ORAL, M.; ORAL, E. L.; AYDIN, A. Supervised vs. unsupervised learning for construction crew productivity prediction. **Automation in Construction**, v. 22, p. 271–276, 2012.
- SANDERS, S. R.; THOMAS, H. R. Masonry productivity forecasting model. **Journal of Construction Engineering and Management**, v. 119, n. 1, p. 163–179, 1993.
- SONG, L.; ABOURIZK, S. Measuring and modeling labor productivity using historical data. **Journal of Construction Engineering and Management**, v. 134, n. 10, p. 786–794, 2008.
- SONMEZ, R.; ROWINGS, J. E. Construction labor productivity modeling with neural networks. **Journal of Construction Engineering and Management**, v. 124, n. 6, p. 498–504, 1998.
- SOUZA, U. E. L.; MARCHIORI, F. F.; ARAUJO, L. O. C. Variable labor productivity unit rate: evaluation by professionals. In: CIB INTERNATIONAL SYMPOSIUM, Helsinki, 2005. **Proceedings [...]** Helsinki, 2005.
- SOUZA, U. E. L.; MORASCO, F. G.; RIBEIRO, G. N. B. **Manual básico de indicadores de produtividade na construção civil**. Brasília: CBIC, 2017. v. 1
- THOMAS, H. R.; YIAKOUMIS, I. Factor model of construction productivity. **Journal of Construction Engineering and Management**, v. 113, n. 4, p. 623–639, 1987.
- ZEVIANI, W. M.; RIBEIRO JÚNIOR, P. J.; BONAT, W. H. Modelos de regressão não linear. In: CONGRESSO DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 58, Campina Grande, 2013. **Anais [...]** Campina Grande, 2013.

Declaração de Disponibilidade de Dados

Os dados de pesquisa só estão disponíveis mediante solicitação

Marcelo Inocêncio Ferreira Correa

Escola de Engenharia Civil e Ambiental | Universidade Federal de Goiás | Av. Universitária, Quadra 86, Lote Área 1488, Setor Leste Universitário | Goiânia - GO - Brasil | CEP 74605-220 | Tel.: (62) 98643-6499 | E-mail: marcelo_eec@yahoo.com.br

Maria Carolina Gomes de Oliveira Brandstetter

Escola de Engenharia Civil e Ambiental | Universidade Federal de Goiás | Tel.: (62) 98299-1441 | E-mail: maria_carolina_brandstetter@ufg.br

Larsson Diogo Seabra Coelho Romagnoli

Escola de Engenharia Civil e Ambiental | Universidade Federal de Goiás | Tel.: (62) 985638791 | E-mail: larssoncoelho@discente.ufg.br

Editor: Ariovaldo Denis Granja

Ambiente Construído

Revista da Associação Nacional de Tecnologia do Ambiente Construído

Av. Osvaldo Aranha, 99 - 3º andar, Centro

Porto Alegre - RS - Brasil

CEP 90035-190

Telefone: +55 (51) 3308-4084

www.seer.ufrgs.br/ambienteconstruido

www.scielo.br/ac

E-mail: ambienteconstruido@ufrgs.br



This is an open-access article distributed under the terms of the Creative Commons Attribution License.