

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

Machine learning aplicado à engenharia civil: análise preditiva da resistência à compressão do concreto*Machine learning applied to civil engineering: predictive modeling of concrete compressive strength*

Aprendizaje automático aplicado a la ingeniería civil: análisis predictivo de la resistencia a la compresión del hormigón

Luciana Andrade Villar¹**Milton Augusto Pinotti²**

RESUMO: Observa-se um avanço significativo no processo de digitalização da engenharia civil, impulsionado pelo desenvolvimento de tecnologias computacionais e pelo aumento da disponibilidade de dados técnicos provenientes de processos produtivos. Nesse cenário, técnicas de Inteligência Artificial (IA), como Machine Learning (ML) e Deep Learning (DL), aplicam-se à engenharia civil para analisar grandes volumes de dados e identificar padrões complexos em variáveis de dosagem e de resistência. Na área da tecnologia do concreto, essas técnicas têm sido investigadas como ferramentas promissoras para estimar propriedades mecânicas do material, como a resistência à compressão, a partir de parâmetros relacionados à dosagem e às condições de produção. O presente artigo apresenta a aplicação e a avaliação de modelos de Machine Learning para a análise preditiva da resistência à compressão do concreto (f_{ck}). O objetivo principal é aplicar algoritmos de regressão supervisionada para estimar o f_{ck} a partir de parâmetros de dosagem, contribuindo para o controle tecnológico e para a tomada de decisão em projetos estruturais. A metodologia adotada incluiu a coleta de 300 amostras de traços reais de uma concreteira localizada em Brusque/SC, abrangendo dez atributos relacionados à composição do concreto. Realizou-se uma seleção preliminar de algoritmos, incluindo Regressão Linear, Árvore de Decisão, Random Forest, SVR e K-vizinhos mais próximos (KNN), sendo o SVR selecionado como o modelo de melhor desempenho, com base na similaridade e na proximidade dos dados. A validação do modelo foi planejada com dados históricos e experimentais, utilizando validação cruzada k-fold. Conclui-se que a abordagem baseada em SVR oferece uma ferramenta promissora para a predição do f_{ck} , com potencial de aplicação prática no controle de qualidade do concreto e na otimização de dosagens em centrais dosadoras.

Palavras-chave: resistência à compressão do concreto; Machine Learning aplicado à resistência do concreto; K vizinhos mais próximos; SVR; predição.

1 INTRODUÇÃO

As inovações tecnológicas têm permeado diversos segmentos da sociedade contemporânea, promovendo transformações significativas nos processos produtivos, na gestão da informação e na tomada de decisões em diferentes áreas do conhecimento. No campo da engenharia civil, esse cenário não é diferente, uma vez que o avanço das tecnologias computacionais e das ferramentas de análise de dados tem contribuído para o aprimoramento de métodos tradicionais

¹ Acadêmica do curso de Eng. Civil da UNIFEFE. E-mail: luciana.villar@unifebe.edu.br

² Professor orientador. Mestre em Engenharia Elétrica. E-mail: pinotti@unifebe.edu.br

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

de projeto, execução e controle de obras.

Essa evolução digital permite uma análise mais criteriosa das características dos materiais empregados na construção. Entre as características mecânicas, a resistência à compressão (f_{ck} , resistência característica aos 28 dias, NBR 5738) é um parâmetro-chave para o desempenho estrutural e o controle de qualidade, conforme as normas técnicas.

A resistência à compressão do concreto (f_{ck}) relaciona-se diretamente à composição do traço e às condições de produção, incluindo a proporção cimento-agregados-água, a relação água/cimento (a/c), aditivos e as condições de mistura, lançamento, adensamento e cura. A influência simultânea dessas variáveis gera complexidade, o que desafia previsões precisas com métodos tradicionais (Neville, 2011; Helene e Terzian, 1992).

Nesse contexto, o controle tecnológico do concreto desempenha um papel fundamental na garantia da qualidade e do desempenho estrutural das construções. Esse processo envolve a verificação das propriedades dos materiais constituintes, o acompanhamento das etapas de produção e a realização de ensaios laboratoriais, sendo o ensaio de resistência à compressão um dos principais métodos utilizados para avaliar a qualidade do concreto produzido. Entretanto, tais procedimentos são realizados após a produção do material, o que limita a possibilidade de antecipar o desempenho do concreto antes de sua aplicação em obra (Helene; Terzian, 1992).

Paralelamente, observa-se um avanço significativo no processo de digitalização da engenharia civil, impulsionado pelo desenvolvimento de tecnologias computacionais e pelo aumento da disponibilidade de dados técnicos provenientes dos processos produtivos. Nesse cenário, técnicas de Inteligência Artificial (IA), como Machine Learning (ML) e Deep Learning (DL), aplicam-se à engenharia civil para analisar grandes volumes de dados e identificar padrões complexos em variáveis de dosagem e de resistência (Goodfellow; Bengio; Courville, 2016). O Machine Learning (Aprendizado de Máquina), em particular, destaca-se por possibilitar o desenvolvimento de modelos preditivos capazes de aprender a partir de dados históricos, identificar relações entre variáveis e realizar previsões com base em padrões observados. Diferentemente dos sistemas computacionais tradicionais, baseados em regras previamente programadas, os modelos de aprendizado de máquina são capazes de extrair padrões diretamente dos dados, permitindo a resolução de problemas complexos por meio de processos de aprendizagem automatizada (Goodfellow; Bengio; Courville, 2016).

Na área da tecnologia do concreto, essas técnicas têm sido investigadas como ferramentas promissoras para estimar propriedades mecânicas do material, como a resistência à

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

compressão, a partir de parâmetros relacionados à dosagem e às condições de produção.

Diante desse contexto, o presente trabalho consiste em uma revisão bibliográfica sobre a aplicação de técnicas de Machine Learning para a análise e a previsão da resistência à compressão do concreto. A pesquisa tem como foco a fundamentação teórica dos principais algoritmos de regressão utilizados na literatura, com ênfase nos modelos SVR (Support Vector Regression) e Decision Tree Regressor, bem como a descrição dos procedimentos metodológicos e das métricas de avaliação empregadas em estudos correlatos. Os testes experimentais com os modelos preditivos, incluindo a etapa de otimização de hiperparâmetros, a validação com dados reais e a análise comparativa dos resultados, serão realizados em etapa posterior da investigação, o que configura a continuidade natural deste estudo.

Assim, tem-se como objetivo geral deste trabalho realizar uma revisão bibliográfica sobre modelos preditivos baseados em técnicas de Machine Learning voltados à previsão da resistência à compressão do concreto, a partir de parâmetros de dosagem e de dados históricos de produção. A relevância desta pesquisa está relacionada à possibilidade de integrar métodos tradicionais da tecnologia do concreto a ferramentas avançadas de análise de dados, contribuindo para o aprimoramento dos processos de previsão e controle da qualidade do material na engenharia civil.

2 FUNDAMENTAÇÃO TEÓRICA

Este estudo abrange os princípios da tecnologia do concreto e as bases computacionais do Machine Learning, estabelecendo a conexão entre as variáveis de dosagem e a capacidade preditiva dos algoritmos supervisionados.

2.1 TECNOLOGIA DO CONCRETO

O concreto constitui um material compósito formado por pasta ligante (cimento Portland e água) e agregados (miúdos e grãos), cuja performance resulta da interação química e física na zona de transição interfacial entre a pasta e as partículas (Mehta; Monteiro, 2014). Essa interação determina propriedades cruciais, como a trabalhabilidade no estado fresco e a resistência no estado endurecido. Sua ampla aplicação na engenharia civil decorre de atributos como elevada resistência mecânica, durabilidade sob diversas condições ambientais, versatilidade na moldagem e custo econômico relativo, o que o consolida como insumo

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

primordial em edificações e obras de infraestrutura (Neville, 2011).

A complexidade das interações entre os componentes do concreto e os fatores ambientais introduz uma variabilidade inerente ao material. Para monitorar esse comportamento, utilizam-se protocolos de ensaio padronizados: enquanto o ensaio de abatimento (slump test) avalia a trabalhabilidade e a consistência no estado fresco, a compressão axial é empregada para quantificar a resistência característica (fck) aos 28 dias. Nesse cenário, a acentuada não linearidade das interações entre os insumos, evidenciada pela correlação moderada entre o consumo de cimento e o fck ($r=0,52$), justifica a adoção de técnicas de Machine Learning para uma predição mais robusta e assertiva (Hoefelmann, 2021).

2.1.1 COMPOSIÇÃO

A composição do concreto é o fator primordial que define seu desempenho. O cimento Portland, ao reagir com a água em um processo de hidratação, forma o gel C-S-H (silicato de cálcio hidratado), que atua como a matriz ligante, aglutinando os agregados.

A qualidade e a quantidade dos agregados (miúdos e graúdos) são igualmente cruciais, influenciando a trabalhabilidade, a densidade e a resistência final do concreto, sendo a granulometria um aspecto fundamental a ser controlado (Carpinteiro, 2005). A relação água/cimento (%) é um dos parâmetros mais críticos, pois governa diretamente a porosidade da pasta de cimento e, conseqüentemente, a resistência e a durabilidade do material.

Em um contexto regional, Resner (2021) analisou traços de concreto em usinas de Santa Catarina, identificando padrões de dosagem que se alinham às diretrizes da NBR 12655 (ABNT, 2015) para concreto dosado em canteiro. A Tabela 1 apresenta uma síntese dos traços típicos observados, destacando a faixa de variação dos componentes e sua influência sobre a resistência à compressão.

Tabela 1 – Traços típicos

Componente	Traço médio (kg/m ³)	Traço SC (kg/m ³)	Influência fck
Cimento	350	300-450	Alta ($r=0,52$)
Água	175	150-200	Negativa (relação % 0,45-0,55)
Agregado Miúdo	700	650-750	Média (granulometria)
Agregado Graúdo	1050	1000-1100	Média (forma e resistência)
Idade Cura	28 dias	7-28 dias	Alta ($r=0,48$)

Fonte: Autor, com base em adaptação de Resner (2021).

2.1.2 FATORES INFLUENCIADORES E VARIABILIDADE

As variáveis que influenciam a resistência à compressão (f_{ck}) do concreto são diversas, sendo a relação água/cimento o fator determinante para a qualidade da mistura. Conforme os princípios estabelecidos por Neville (2011), uma menor proporção de água em relação ao volume de cimento reduz a porosidade da pasta endurecida, elevando a resistência mecânica, desde que seja garantida a hidratação completa das partículas de cimento.

Para equilibrar essa relação sem comprometer a trabalhabilidade, utilizam-se aditivos superplastificantes, que permitem reduzir o consumo de água, mantendo a fluidez necessária ao adensamento. O monitoramento dessa consistência no estado fresco é realizado por meio do ensaio de abatimento do tronco de cone (slump test), que assegura a conformidade da mistura antes da moldagem. O desenvolvimento da resistência é então monitorado durante o período de cura, sob condições controladas de temperatura e umidade, tendo a idade de 28 dias como o marco normativo para a validação dos resultados de f_{ck} , seguindo os critérios de controle tecnológico vigentes.

2.2 RESISTÊNCIA À COMPRESSÃO DO CONCRETO

A resistência característica à compressão (f_{ck} , MPa) aos 28 dias define o dimensionamento estrutural (NBR 6118:2023).

Essa propriedade é validada conforme demonstram os ensaios em cilindros, que calculam f_{ck} como o percentil 5% (95% de confiabilidade). Nesse contexto, para concreto usinado em SC, f_{ck} de 25-30 MPa equilibra custo, durabilidade e produtividade.

2.2.1 DOSAGEM E INFLUÊNCIAS PRINCIPAIS

A relação água/cimento (a/c) é o fator determinante da resistência à compressão (f_{ck}). A utilização de índices reduzidos (como a faixa de 0,45 a 0,55) diminui a porosidade da pasta de cimento endurecida, resultando no aumento da resistência mecânica conforme os princípios da Lei de Abrams (NEVILLE, 2011). Essa dosagem assegura o equilíbrio entre a hidratação completa das partículas de cimento e a trabalhabilidade necessária ao adensamento, o que mitiga a formação de fissuras iniciais e atende aos requisitos de durabilidade previstos na NBR 6118, item 7.4.2.

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

2.2.2 FATORES ESPECÍFICOS DA RESISTÊNCIA FCK E CONTROLE DE VARIABILIDADE

Diversos fatores modulam a resistência final à compressão (f_{ck}). Além da relação água-cimento, destacam-se o tipo de cimento, a qualidade dos agregados e a eficiência da cura (Neville, 2011). Além disso, variáveis ambientais, como temperatura e umidade relativa, desempenham um papel crítico na cinética de hidratação do cimento. Segundo Mehta e Monteiro (2014), o monitoramento contínuo dessas condições é essencial e deve seguir as diretrizes da NBR 14931, especialmente em elementos pré-moldados, nos quais o controle rigoroso da temperatura e da consistência no estado fresco assegura a integridade estrutural. Essa dependência de múltiplos fatores reflete-se na variabilidade produtiva do concreto, que comumente apresenta dispersão, com coeficientes de variação entre 10% e 15% (NBR 12655). Tais oscilações decorrem de flutuações inevitáveis nas etapas de mistura, transporte e cura. Uma limitação relevante no processo convencional é que os ensaios destrutivos, realizados conforme a NBR 5739 aos 28 dias, fornecem resultados tardios, o que limita intervenções imediatas. Portanto, a compreensão dessa variabilidade reforça a importância de métodos de controle mais ágeis e rigorosos, fundamentados nas normas técnicas vigentes.

2.3 MACHINE LEARNING E MODELOS PREDICTIVOS

O Machine Learning (ML) é um campo da inteligência artificial que capacita sistemas a aprender e aprimorar seu desempenho a partir de dados, sem serem explicitamente programados para cada tarefa específica (Goodfellow et al., 2016). Essa capacidade de identificar padrões complexos e realizar previsões torna o ML uma ferramenta valiosa na engenharia civil, especialmente na modelagem de materiais como o concreto.

A previsão da resistência à compressão (f_{ck}) do concreto é um desafio notável, dada a natureza não-linear das interações entre seus componentes, como a relação água/cimento ($\frac{a}{c}$), o tipo de cimento, a proporção de agregados e a idade de cura.

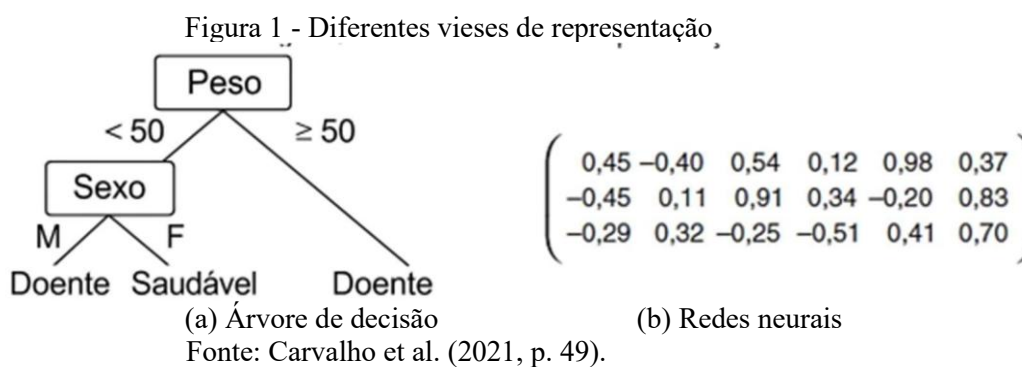
Os métodos tradicionais frequentemente simplificam essas relações, enquanto o ML oferece uma abordagem robusta para capturar essa complexidade e otimizar o controle de qualidade.

2.3.1 ALGORITMOS DE ML

Um algoritmo de ML é um conjunto de instruções para resolver problemas por meio do

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

aprendizado de dados, sem programação explícita. No contexto de predição de *fck* do concreto, constrói-se hipóteses de atributos (α , cura) para mapear relações não-lineares, em que cada algoritmo usa um viés de representação específico, como árvores de decisão (hierarquia) ou redes neurais (pesos), limitando as hipóteses para a generalização (Figura 1).



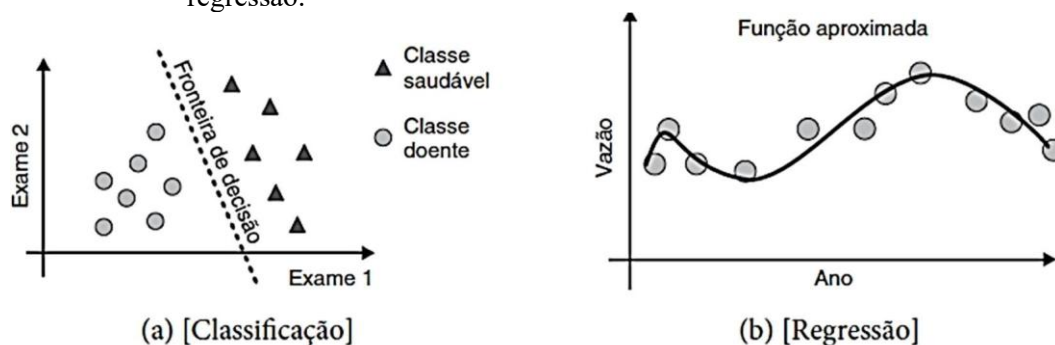
El sesgo de búsqueda explora modelos de forma eficiente. Essencial para predições robustas em dados de concreto.

2.3.2 REGRESSÃO E CLASSIFICAÇÃO

Um algoritmo preditivo de Machine Learning (ML) é uma função que, a partir de um conjunto de exemplos rotulados, constrói um estimador. O atributo assume valores dentro de um domínio previamente conhecido e, quando esse domínio é composto por valores nominais, o problema é classificado como de classificação (ou aprendizado de conceitos) e o estimador gerado é chamado de classificador. Por outro lado, se o domínio é um conjunto infinito e ordenado de valores, trata-se de um problema de regressão e o estimador é conhecido como regressor. (Dietterich, 1998).

Tanto o classificador quanto o regressor são funções que recebem um exemplo não rotulado e produzem uma saída: no caso do classificador, atribui o exemplo a uma das classes possíveis; no caso do regressor, estima um valor real correspondente ao exemplo apresentado. A Figura 2 exemplifica a definição de classificação e de regressão. (Carvalho et al., 2021).

Figura 2 – Gráfico representativo da classificação e da regressão.



Fonte: Carvalho et al. (2021), p 49.

2.3.3 ALGORITMOS DE REGRESSÃO

A Árvore de Decisão CART (Classification and Regression Trees) é um algoritmo de regressão que constrói uma estrutura hierárquica de decisões baseada em divisões recursivas dos dados, utilizando critérios como Gini ou Entropia para minimizar a impureza, sendo interpretável e robusta a outliers (Breiman et al., 1984).

A Floresta Aleatória (Random Forest, RF) estende o CART por meio de um ensemble de múltiplas árvores (bagging + feature randomness), reduzindo a variância e o overfitting, com predições agregadas pela média (Breiman, 2001).

A robustez desse método sustenta sua viabilidade técnica na engenharia civil, visto que estudos recentes validam a aplicação de CART e Random Forest na predição de f_{ck} , conforme NBR 6118 (2023) e NBR 12655 (2022), que enfatizam o controle de variáveis como a razão $\frac{a}{c}$ e a cura.

Silva et al. (2023) analisaram concretos brasileiros com Random Forest, obtendo $R^2 = 0,95$ e destacando $\frac{a}{c}$ e agregados como preditores principais. Omotayo, Arum e Ikumapayi (2024), por sua vez, compararam métodos de ML, com o RF superando o CART (RMSE = 3.2 MPa vs. 4.8 Mpa). Xu et al. (2021), por outro lado, propuseram um modelo ensemble para concreto pronto, alcançando $R^2 = 0.92$ em traços variáveis.

O K-Vizinhos Mais Próximos (KNN) é um algoritmo de aprendizado baseado em instâncias que realiza predições com base na similaridade entre amostras no espaço de atributos. Diferentemente dos modelos baseados em árvores, o KNN não constrói uma função de mapeamento explícita durante o treinamento, mas armazena todo o conjunto de dados para realizar inferências no momento da predição.

2.4 MÉTRICAS DE AVALIAÇÃO DE MODELOS DE REGRESSÃO

A avaliação da performance de modelos de Machine Learning, especialmente em Tarefas de regressão, como a predição da resistência característica à compressão (f_{ck}) do concreto, são cruciais para determinar sua acurácia e generalização.

Métricas estatísticas permitem quantificar a discrepância entre os valores preditos pelo modelo e os valores reais observados, fornecendo uma base objetiva para a comparação e seleção de algoritmos (Harrison, 2020; Carvalho, 2021). Sendo assim, a escolha adequada das métricas é fundamental para compreender as características dos erros do modelo e sua adequação ao problema em questão.

2.4.1 COEFICIENTE DE DETERMINAÇÃO (R^2)

O Coeficiente de Determinação, ou R^2 , é uma métrica amplamente utilizada para avaliar a proporção da variância da variável dependente (f_{ck}) explicada pelas variáveis independentes do modelo. Seu valor varia de 0 a 1, em que valores mais próximos de 1 indicam que o modelo explica uma parcela maior da variabilidade dos dados, ou seja, apresenta um melhor ajuste. É calculado conforme a Equação 1:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

Onde

R = Coeficiente de determinação

f = função preditiva y_i = Atributo alvo conhecido. Representa o valor real da f_{ck} para a i -ésima amostra.

x_i = valor predito pelo modelo para a i -ésima amostra.

\bar{y} Média dos valores reais de f_{ck} .

n = número total de amostras.

Para modelos de Florestas Aleatórias aplicados à predição de f_{ck} , espera-se um R^2 entre 0,92 e 0,98, indicando alta capacidade explicativa.

2.4.2 ERRO QUADRÁTICO MÉDIO (MSE) E RAIZ DO ERRO QUADRÁTICO MÉDIO (RMSE)

O Erro Quadrático Médio (MSE) mede a média dos quadrados dos erros, ou seja, a diferença

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

entre os valores preditos e os valores reais. É uma métrica que penaliza erros maiores de forma mais significativa devido à elevação ao quadrado, tornando-a sensível a outliers. O RMSE é a mesma métrica da MSE, porém normalizado para a mesma unidade da variável-alvo, tornando a análise e a interpretação do erro mais intuitivas. O MSE é definido conforme a equação 2 e o RMSE conforme a equação 3:

n

$$MSE(f) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2)$$

Onde:

MSE = Erro quadrático médio

$$RMSE(f) = \sqrt{MSE} \quad (3)$$

Onde:

$RMSE$ = Raiz do erro quadrático médio

2.4.3 ERRO ABSOLUTO MÉDIO (MAE)

O erro absoluto Médio (MAE) calcula a média dos valores absolutos dos erros. Diferentemente do MSE/RMSE, o MAE não eleva os erros ao quadrado, tornando-o menos sensível a outliers e mais robusto em distribuições de erro com valores extremos. O MAE é expresso pela Equação 4:

n

1

$$MAE(f) = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (4)$$

Onde:

MAE = Erro Absoluto médio

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

2.4.4 ANÁLISE DE RESÍDUOS E FLUXO DE AVALIAÇÃO

Além das métricas quantitativas, a análise gráfica dos resíduos (diferenças entre os valores observados e os preditos) é fundamental (Hoefelmann, 2021). Essa avaliação permite identificar as características de um bom modelo: resíduos distribuídos simetricamente em torno de zero, sem padrões discerníveis, o que indica que os erros são aleatórios e que não há viés sistemático nas predições (Harrison, 2020).

Gráficos de resíduos versus valores preditos ou versus variáveis independentes podem revelar problemas como heterocedasticidade ou variáveis omitidas. Dessa forma, o fluxo de avaliação de modelos de Machine Learning geralmente envolve a divisão do conjunto de dados em subconjuntos de treinamento e de teste (`train_test_split`) para garantir que o modelo seja avaliado em dados não vistos durante o treinamento.

Para uma avaliação mais robusta e para mitigar a dependência da divisão específica dos dados, técnicas como a validação cruzada *k*-fold (*k*-fold Cross-Validation) são empregadas. Nela, o conjunto de dados é dividido em '*k*' partes, e o modelo é treinado e testado '*k*' vezes, utilizando uma parte diferente para teste em cada iteração; as métricas são então calculadas como a média dos '*k*' resultados.

Essas métricas de avaliação (R^2 , RMSE, MAE) foram aplicadas em contextos reais, alinhados à NBR 12655 para ensaios de f_{ck} . Os estudos de Hoefelmann (2021) e Chou et al. (2011) confirmam a robustez do RF em capturar não-linearidades de cura e de agregados, com resíduos aleatórios indicando boa generalização.

3 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta o percurso metodológico estabelecido para a execução do trabalho, compreendendo as fases que partem da coleta de informações até a verificação do modelo preditivo, visto o propósito de criar um sistema de Machine Learning capaz de estimar a resistência à compressão (f_{ck}) do concreto baseando-se em sua composição.

O estudo inicia com um levantamento bibliográfico para fundamentar tecnicamente a pesquisa, seguido pela utilização de uma base de dados de uma usina de concreto local.

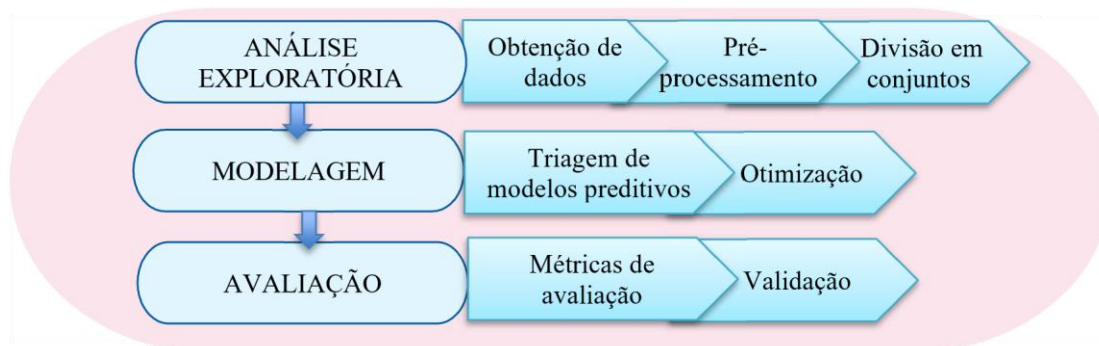
Definido o conjunto de dados, iniciou-se a etapa de análise, composta pela extração das variáveis, seu pré-processamento e a partição posterior em amostras de treinamento e teste.

Na fase de modelagem, executou-se uma triagem inicial entre diversos algoritmos de regressão

para comparação, progredindo para a otimização daquele com melhor desempenho.

Por fim, a etapa de avaliação descreve os indicadores de desempenho e o processo de validação da ferramenta desenvolvida. A Figura 3 apresenta o fluxograma metodológico que sintetiza o processo de trabalho.

Figura 3 – Fluxograma metodológico



Fonte: Os autores (2026).

O tratamento das informações e a construção do algoritmo foram realizados em linguagem Python, utilizando a distribuição Anaconda e o aplicativo Jupyter.

3.1 CONJUNTO DE DADOS (DATASET)

Visando assegurar a integridade, a pertinência e a conformidade das informações com o escopo desta pesquisa, optou-se por um conjunto de dados que abrange os parâmetros de dosagem e os resultados de resistência à compressão. Este dataset é composto por registros reais coletados diretamente em uma usina dosadora de concreto localizada no município de Brusque/SC. Os dados foram extraídos de relatórios técnicos de produção e ensaios laboratoriais, contemplando uma amostragem robusta que reflete as condições práticas de fabricação e controle tecnológico da região.

Tabela 2 – Parâmetros de traço

Componente	Cimento	Areia Artificial	Areia natural	Brita 1	Brita 0	Água	Aditivo	Superplastificante
Traço 1	277	522	352	655	276	185	2,2	0,0
Traço 2	360	449	454	488	320	195	1,08	1,08

Fonte: Os autores (2026).

Com base na composição dos traços apresentados na Tabela 2, foram realizados os

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

procedimentos de moldagem e cura dos corpos de prova cilíndricos para cada traço experimental. Para garantir a representatividade da amostra e a confiabilidade dos resultados, realizaram-se moldagens em série para cada dosagem, seguindo rigorosamente os critérios de adensamento e de cura úmida. As leituras de resistência à compressão foram obtidas por meio do rompimento dos corpos de prova em prensa hidráulica devidamente calibrada, com registros realizados aos 28 dias de idade. O conjunto de dados resultante, consolidado a partir dos relatórios técnicos da usina regional, serviu de base para o treinamento e a validação dos modelos preditivos.

3.1 PRÉ-PROCESSAMENTO

3.1.1 INTEGRAÇÃO

O conjunto de dados (dataset) utilizado nesta pesquisa foi consolidado a partir dos Registros técnicos da usina regional, integrando as variáveis de dosagem e os resultados de resistência em uma base de dados unificada. Esta base é composta por 10 colunas de atributos, abrangendo os componentes da mistura (cimento, areia artificial, areia natural, brita 1, brita 0, água, aditivo e superplastificante) e os resultados de desempenho mecânico (R7 e R28). Ao todo, foram processados 300 registros de corpos de prova reais. Após a etapa de integração e estruturação, os dados foram exportados para o ambiente Jupyter no formato .csv, permitindo o início dos procedimentos de pré-processamento e de análise preditiva.

3.1.2 AVALIAÇÃO DE ATRIBUTOS

A etapa de avaliação de atributos foi conduzida com o intuito de selecionar as variáveis de maior poder preditivo e garantir a integridade estatística do modelo. Foram removidos os registros de identificação dos corpos de prova (ID CP), bem como as colunas que apresentavam valores constantes ou não apresentavam variabilidade relevante para o aprendizado do algoritmo.

Diferentemente dos identificadores, o resultado da resistência aos 7 dias (R7) foi mantido no conjunto de dados. A permanência desse atributo permite ao modelo captar a cinética de endurecimento inicial do concreto, servindo como um indicador técnico valioso para a predição da resistência aos 28 dias (R28). Essa integração permite que o algoritmo identifique padrões de ganho de resistência que dependem da interação química entre o cimento e os aditivos ao

longo do tempo.

A exclusão dessas variáveis foi realizada por meio da função `.drop` da biblioteca Pandas, o que permitiu otimizar o conjunto de dados. Com esse procedimento, a estrutura do dataset foi refinada, mantendo apenas os atributos essenciais para a modelagem, o que reduz a complexidade computacional e mitiga o risco de sobreajuste (overfitting). A Figura 4 apresenta o resumo estatístico do conjunto de dados

Figura 4 – Dados estatísticos dos atributos.

	Cimento	Areia Artificial	Areia Natural	Brita 1	Brita 0	Água	Aditivo	Superplastificante	R7 (Mpa)	R28 (Mpa)
count	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000
mean	304.666667	497.666667	386.000000	599.333333	290.666667	188.333333	2.140000	0.360000	27.845667	33.617667
std	39.191950	34.470028	48.163601	78.856091	20.776455	4.721922	0.113326	0.509968	3.249906	2.726052
min	277.000000	449.000000	352.000000	488.000000	276.000000	185.000000	1.980000	0.000000	20.100000	25.800000
25%	277.000000	449.000000	352.000000	488.000000	276.000000	185.000000	1.980000	0.000000	25.675000	31.600000
50%	277.000000	522.000000	352.000000	655.000000	276.000000	185.000000	2.220000	0.000000	27.600000	34.000000
75%	360.000000	522.000000	454.000000	655.000000	320.000000	195.000000	2.220000	1.080000	30.000000	35.500000
max	360.000000	522.000000	454.000000	655.000000	320.000000	195.000000	2.220000	1.080000	35.000000	41.200000

Fonte: Os autores (2026).

3.1.3 PADRONIZAÇÃO

A padronização dos dados é uma etapa de pré-processamento que visa normalizar as escalas das variáveis, garantindo que todas possuam média zero e desvio padrão unitário. Este procedimento é fundamental para o desempenho de algoritmos de aprendizado de máquina, pois evita que atributos com ordens de grandeza superiores exerçam influência desproporcional durante o treinamento dos modelos.

No contexto desta pesquisa, a aplicação da padronização justifica-se pela disparidade numérica entre as variáveis de entrada, como o consumo de cimento (em kg/m³) e o teor de aditivos. Para a execução desta tarefa, utilizou-se a função `StandardScaler`, disponível no módulo `preprocessing` da biblioteca `Scikit-learn`.

3.2 DIVISÃO DOS CONJUNTOS DE DADOS

O estágio inicial da construção da arquitetura preditiva consiste no treinamento do algoritmo com a base de dados consolidada. Para essa finalidade, empregou-se a biblioteca `scikit-learn` (`sklearn`), especificamente por meio da função `train_test_split`. Esse recurso permite a partição aleatória das informações em dois grupos distintos: um voltado ao aprendizado e outro

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

dedicado à validação dos resultados. Nesta pesquisa, definiu-se que 80% dos registros seriam destinados ao treinamento, e os 20% restantes, à fase de testes.

Para assegurar uma avaliação imparcial do desempenho do sistema, os dados foram divididos em conjuntos de treinamento e de teste. Essa distribuição garante que a eficácia do modelo seja validada em dados não vistos durante a fase de aprendizado. A organização das variáveis (atributos) foi estruturada conforme segue:

- X_{train} : compreende os atributos preditores (dosagens) utilizados para o aprendizado do modelo;
- y_{train} : armazena a variável alvo correspondente à resistência à compressão (fck) para o treinamento;
- X_{test} : engloba as variáveis de entrada que servirão para testar a acurácia do modelo;
- y_{test} : contém os valores reais de resistência associados ao X_{test} , permitindo a verificação estatística das previsões.

Essa metodologia de partição de dados é imprescindível para atestar o potencial de generalização do modelo perante novas dosagens, mitigando o risco de sobreajuste aos dados originais.

3.3 AVALIAÇÃO PRELIMINAR DE ALGORITMOS DE REGRESSÃO

Com o intuito de identificar o algoritmo com maior capacidade de generalização para a estimativa da resistência do concreto, realizou-se uma triagem inicial comparativa entre os modelos de regressão disponíveis na biblioteca scikit-learn. Para assegurar a escolha do modelo mais eficiente e robusto, foram avaliados os seguintes algoritmos:

- Regressão Linear (sklearn.linear_model): modelo estatístico clássico para análise de relações lineares;
- Regressor de Árvore de Decisão — CART (sklearn.tree): algoritmo que subdivide os dados em estruturas hierárquicas;
- Regressor Floresta Aleatória — Random Forest (sklearn.ensemble): técnica de conjunto (ensemble) que combina múltiplas árvores de decisão;
- Regressor de Vetores de Suporte — SVR (sklearn.svm): algoritmo que busca encontrar o melhor hiperplano para a regressão.

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

- K vizinhos mais próximos — KNN (sklearn.neighbors): modelo baseado na similaridade e proximidade dos dados;

O treinamento de cada modelo foi realizado por meio da função fit (X_{train} , y_{train}), que submete os parâmetros de dosagem e seus respectivos resultados de resistência ao processamento dos algoritmos. Esse procedimento permite que os algoritmos identifiquem padrões e interdependências presentes no dataset. Após a etapa de treinamento, utilizou-se a função predict(X_{test}) para que os modelos processassem as variáveis do conjunto de teste, gerando o vetor y_{pred} . Este conjunto armazena as estimativas produzidas por cada algoritmo de regressão (listados acima), possibilitando a análise estatística posterior em relação aos valores reais medidos em laboratório.

3.4 AVALIAÇÃO DOS MODELOS

A avaliação da eficácia dos modelos detalhados na seção anterior foi realizada com base em quatro indicadores estatísticos de regressão: R^2 , MAE, MSE e RMSE. Tais métricas permitem comparar os resultados reais obtidos no conjunto de teste (y_{test}) com as estimativas geradas pelos algoritmos (y_{pred}), possibilitando a quantificação da precisão das predições.

A implementação desses indicadores foi realizada por meio de funções específicas da biblioteca scikit-learn. O coeficiente R^2 permitiu verificar o percentual de variância explicado pelo modelo, enquanto o MAE e o RMSE forneceram a magnitude do erro médio e a sensibilidade a desvios maiores, respectivamente.

3.5 OTIMIZAÇÃO DE ALGORITMOS

O desempenho do modelo preditivo foi aprimorado por meio da configuração de hiperparâmetros, ajustando os parâmetros internos dos algoritmos para elevar a precisão e a robustez das estimativas. Para essa finalidade, empregou-se a técnica de busca aleatória, com validação cruzada, por meio da função RandomizedSearchCV da biblioteca scikit-learn. Essa abordagem possibilitou a exploração eficiente de diversas combinações de parâmetros pré-definidos.

Os hiperparâmetros e seus respectivos intervalos de variação foram definidos conforme as especificidades do modelo K-Vizinhos Mais Próximos (KNN). A busca foi configurada para processar um número determinado de combinações aleatórias e, ao término do procedimento,

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

a configuração que apresentou o melhor desempenho médio na validação foi selecionada para compor a versão final do modelo preditivo.

3.6 VALIDAÇÃO DO MODELO

Para validar o algoritmo desenvolvido nesta pesquisa, foram adotados dois procedimentos distintos. O primeiro consistiu na validação cruzada dos dados históricos da usina de Brusque, integrada ao processo padrão de treinamento de Machine Learning. O segundo procedimento envolveu a validação com dados experimentais coletados na própria usina, com o propósito de verificar a capacidade de generalização do modelo perante dosagens inéditas que não foram apresentadas durante a fase de aprendizado.

3.6.1 VALIDAÇÃO COM DADOS EXPERIMENTAIS

Para validar o algoritmo desenvolvido, foi realizada uma etapa de verificação com dados experimentais coletados diretamente na usina dosadora. Esse procedimento teve como objetivo testar a capacidade de generalização do modelo perante dosagens reais que não integraram o conjunto de treinamento inicial. Os traços foram selecionados de modo a replicar as condições operacionais da base de dados, mantendo as mesmas especificações de insumos e as relações água/cimento apresentadas na Tabela 2.

A coleta das informações seguiu os padrões laboratoriais da usina, em que os resultados de resistência à compressão foram medidos por meio de prensas hidráulicas calibradas, conforme as normas NBR 5738 e NBR 5739. Diferentemente do método de usinagem e medição de rugosidade proposto por Canal (2022), esta validação fundamentou-se no controle tecnológico do concreto endurecido aos 28 dias. Os novos valores foram devidamente tabelados e importados para o ambiente Jupyter, no qual, por meio da função `predict(X_valid)`, foi gerado o conjunto de previsões `y_pred_valid`.

Ao final, os resultados reais das dosagens experimentais foram confrontados com as estimativas geradas pelo algoritmo. A performance foi avaliada por meio das métricas R^2 , MAE, MSE e RMSE, além da análise de resíduos, o que atesta a confiabilidade da ferramenta para aplicações práticas no monitoramento da resistência do concreto regional.

4 ANÁLISE DOS RESULTADOS

Nesta seção, apresentam-se os principais achados decorrentes da aplicação dos modelos de regressão aos dados experimentais, com ênfase na variável dependente resistência à compressão do concreto (f_{ck}). Primeiramente, descrevem-se as métricas de desempenho obtidas na etapa de avaliação preliminar dos modelos, seguidas pela comparação entre os diferentes algoritmos testados, destacando-se o SVR e o Decision Tree regressor, que apresentaram desempenhos semelhantes e os melhores resultados nesta fase inicial.

Posteriormente, examina-se o impacto dos hiperparâmetros no comportamento do modelo. A interpretação das métricas de erro e da distribuição dos resíduos é realizada, o que possibilita uma avaliação confiável dos resultados. Por último, apresentam-se os testes executados com dados do experimento, com o objetivo de avaliar a capacidade do modelo de generalizar a dados não observados durante o treinamento.

4.1 RESULTADOS DA AVALIAÇÃO PRELIMINAR DOS ALGORITMOS

A Tabela 3 apresenta os indicadores de desempenho dos algoritmos avaliados na etapa preliminar. O algoritmo SVR (Support Vector Regression) destacou-se nesta fase, apresentando o maior coeficiente de determinação (R^2).

Tabela 3 – Resultados da avaliação preliminar

Algoritmo	R2	MÃE	MSE	RMSE
SVR	0,76	0,48	0,63	0,76
Decision Tree	0,75	0,48	0,65	0,81
KNNeighbors	0,70	0,61	0,79	0,89
Random Forest	0,63	0,69	0,96	0,98
Linear Regression	0,63	0,69	0,96	0,98

Fonte: Os autores (2026).

4.2 OTIMIZAÇÃO DO ALGORITMO SVR

Os algoritmos SVR (Support Vector Regression) e Decision Tree Regressor foram submetidos a um processo de ajuste de hiperparâmetros utilizando a função `RandomizedSearchCV`, com o propósito de identificar a melhor configuração para o conjunto de dados em questão.

A busca foi parametrizada para testar 50 combinações aleatórias de hiperparâmetros dentro de

Ano VII, v.1 2026 | **submissão: 23/05/2026** | **aceito: 26/05/2026** | **publicação: 29/05/2026**

intervalos previamente estabelecidos para cada algoritmo. Após a conclusão da execução, a combinação que apresentou o melhor desempenho no conjunto de validação foi selecionada como a configuração definitiva de cada modelo. Os hiperparâmetros submetidos a essa busca, seus respectivos intervalos de variação e os valores ótimos identificados constam das Tabelas 4 e 5.

Tabela 4 – Seleção de Hiperparâmetros - SVR

Algoritmo	Intervalos	Melhor parâmetro
Kernel	[linear, rbf, poly, sigmoid]	A definir
C	[0.1, 1, 10, 100]	A definir
Epsilon	[0.01, 0.1, 0.2, 0.5]	A definir
Gamma	[scale, auto, 0.01, 0.1]	A definir
Degree	[2, 3, 4]	A definir
Tol	[1e-3, 1e-4, 1e-5]	A definir

Fonte: Os autores (2026).

Tabela 5 – Seleção de Hiperparâmetros – Decision Tree Regressor

Algoritmo	Intervalos	Melhor parâmetro
Criterion	[squared error, friedman mse, absolute error, poisson]	A definir
max_depth	[None, 10, 20, 30]	A definir
min_samples_split	[2, 5, 10]	A definir
min_samples_leaf	[1, 2, 4]	A definir
max_features	[sqrt, log2, None]	A definir
ccp_alpha	[0.0, 0.01, 0.1]	A definir

Fonte: Os autores (2026).

4.3 IMPORTÂNCIA DE ATRIBUTOS (Feature Importance)

A análise de importância de atributos, que quantifica a contribuição de cada variável nas previsões do modelo, evidenciou que R7 e R28 (resistências do concreto aos 7 e 28 dias) são os fatores mais determinantes para a predição. Esses resultados são coerentes, uma vez que as resistências iniciais e finais do concreto refletem diretamente a qualidade da mistura e sua capacidade de desenvolver resistência. Os demais parâmetros do traço também apresentaram contribuições significativas. Identificou-se, contudo, que o traço 2 gerava valores atípicos (outliers) que impactavam o desempenho do modelo.

4.4 TRABALHOS FUTUROS: TESTES EXPERIMENTAIS COM SVR E DECISION TREE

O presente trabalho consistiu em uma revisão bibliográfica sobre a aplicação de algoritmos de Machine Learning para a predição da resistência à compressão do concreto (f_{ck}), com ênfase nos modelos SVR (Support Vector Regression) e Decision Tree Regressor. A fundamentação teórica apresentada, juntamente com a descrição dos procedimentos metodológicos e das métricas de avaliação, forneceu a base conceitual necessária para a continuidade da investigação.

Como etapa futura, prevê-se a realização de testes experimentais com os algoritmos SVR e Decision Tree Regressor, utilizando o conjunto de dados descrito na seção 3.1. Ambos os modelos serão submetidos ao processo de otimização de hiperparâmetros por meio da função RandomizedSearchCV, nos intervalos previamente estabelecidos, com o objetivo de identificar a configuração com o melhor desempenho preditivo.

Após a etapa de otimização, os modelos serão validados com dados experimentais e avaliados com base nas métricas R^2 , MAE, MSE e RMSE. A análise comparativa dos resultados permitirá identificar qual dos dois algoritmos apresenta o melhor equilíbrio entre precisão e capacidade de generalização para a predição do f_{ck} , contribuindo para o aprimoramento do controle tecnológico do concreto na engenharia civil.

CONSIDERAÇÕES FINAIS

O presente trabalho apresentou uma revisão bibliográfica sobre a aplicação de técnicas de Machine Learning para a predição da resistência à compressão do concreto (f_{ck}), com ênfase nos algoritmos SVR (Support Vector Regression) e Decision Tree Regressor. A fundamentação teórica abordou os principais conceitos relacionados à tecnologia do concreto, aos fatores que influenciam sua resistência e aos fundamentos dos modelos preditivos, bem como as métricas de avaliação empregadas na literatura especializada.

A partir da revisão realizada, foi possível constatar que a utilização de algoritmos de regressão tem se mostrado uma abordagem promissora para estimar o f_{ck} com base em parâmetros de dosagem e em dados históricos de produção. A literatura consultada indica que tanto o SVR quanto a Árvore de Decisão têm potencial para capturar relações não lineares entre as variáveis de entrada e a resistência do concreto, podendo auxiliar no controle tecnológico e na

Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

otimização de traços.

A análise dos procedimentos metodológicos descritos em estudos correlatos permitiu identificar as etapas fundamentais para o desenvolvimento de modelos preditivos, incluindo o pré-processamento de dados, a divisão dos conjuntos de treinamento e de teste, a avaliação preliminar dos algoritmos e a otimização de hiperparâmetros. As métricas R^2 , MAE, MSE e RMSE mostraram-se adequadas para a avaliação comparativa do desempenho dos modelos.

Ressalta-se que este trabalho constitui uma etapa inicial de investigação, com como principais contribuições a sistematização do referencial teórico e a definição da metodologia a ser aplicada. Os testes experimentais com os algoritmos SVR e Decision Tree Regressor, incluindo a otimização de hiperparâmetros via RandomizedSearchCV e a validação com dados reais, serão realizados em etapa posterior, o que configura a continuidade natural desta pesquisa.

Espera-se que os resultados futuros contribuam para o aprimoramento do controle tecnológico do concreto na engenharia civil, integrando métodos tradicionais a ferramentas avançadas de análise de dados e possibilitando estimativas mais rápidas e eficientes da resistência do material antes mesmo da realização dos ensaios aos 28 dias.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 5738: Concreto — Procedimento para moldagem e cura de corpos de prova. Rio de Janeiro, 2015.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 5739: Concreto — Ensaio de compressão de corpos de prova cilíndricos. Rio de Janeiro, 2018.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6118: Projeto de estruturas de concreto — Procedimento. Rio de Janeiro, 2023.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 12655: Concreto de cimento Portland — Preparo, controle, recebimento e aceitação — Procedimento. Rio de Janeiro, 2022.

BISHOP, C. M. Pattern recognition and machine learning. New York: Springer, 2006.

BREIMAN, L. Random forests. Machine Learning, v. 45, n. 1, p. 5–32, 2001.

BREIMAN, L. et al. Classification and regression trees. New York: Chapman & Hall, 1984.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge: MIT Press, 2016.

HELENE, P.; TERZIAN, P. Manual de dosagem e controle do concreto. São Paulo: Pini, 1992.



Ano VII, v.1 2026 | submissão: 23/05/2026 | aceito: 26/05/2026 | publicação: 29/05/2026

MEHTA, P. K.; MONTEIRO, P. J. M. Concreto: microestrutura, propriedades e materiais. 2. ed. São Paulo: IBRACON, 2014.

NEVILLE, A. M. Properties of concrete. 5. ed. Londres: Pearson, 2011.

OMOTAYO, T. S.; ARUM, C.; IKUMAFAYI, C. M. Comparative analysis of machine learning models for predicting concrete compressive strength. Asian Journal of Civil Engineering, v. 25, p. 1–12, 2024.

SILVA, R. et al. Previsão da resistência à compressão do concreto com Random Forest. Revista Ibracon de Estruturas e Materiais, v. 16, n. 2, p. 1-15, 2023.

XU, J. et al. Ensemble learning models for predicting concrete compressive strength—Construction and Building Materials, v. 303, p. 124-132, 2021.