



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA CIVIL**



Trabalho de Conclusão de Curso

**Otimização Multiobjetivo de Misturas de Concreto Utilizando
XGBoost e Algoritmo genético: Uma abordagem Sustentável
Baseada em IA.**

ISAAC MORAES DOS SANTOS – 201606740105

**Belém - PA
16 de setembro de 2025**

**ISAAC MORAES DOS SANTOS
EDILSON MORAIS LIMA E SILVA**

**Otimização Multiobjetivo de Misturas de Concreto Utilizando
XGBoost e Algoritmo genético: Uma abordagem Sustentável
Baseada em IA.**

Artigo apresentado como requisito para conclusão do curso de Bacharelado em Engenharia Civil, pela Faculdade de Engenharia Civil do Instituto de Tecnologia – UFPA.

Orientador (a): Prof. Dr. Edilson Morais Lima e Silva

**Belém - PA
16 de setembro de 2025**

**ISAAC MORAES DOS SANTOS
EDILSON MORAIS LIMA E SILVA**

**Otimização Multiobjetivo de Misturas de Concreto Utilizando
XGBoost e Algoritmo genético: Uma abordagem Sustentável
Baseada em IA.**

Monografia apresentado como requisito para
conclusão do curso de Bacharelado em Engenharia
Civil, pela Faculdade de Engenharia Civil do
instituto de tecnologia – UFPA.

Belém-PA, terça-feira, 16 de setembro de 2025.

Banca Examinadora

PROF. DR. EDILSON MORAIS LIMA E SILVA

Orientador

PROF(A). DRA. LUCIANA DE NAZARÉ PINHEIRO CORDEIRO

Examinador

PROF(A). DRA. NÍVEA GABRIELA BENEVIDES DE ALBURQUERQUE

Examinador

PROF. DR. SANDOVAL JOSE RODRIGUES JUNIOR

Examinador

Dedico este trabalho a todos que, com seu apoio, tornaram este sonho possível. A caminhada exigiu sacrifícios, mas hoje entendo perfeitamente que **é justo que muito custe, o que muito vale.**

AGRADECIMENTOS

Ao encerrar este ciclo tão importante, meu coração se enche de gratidão e é impossível não olhar para trás e refletir sobre a jornada que me trouxe até aqui. O caminho foi árduo, marcado por longas jornadas que se dividiam entre o trabalho e os estudos, um desafio que testou meus limites e exigiu resiliência a cada passo. Em meio às dificuldades, o que me manteve firme foi a certeza de um sonho de infância: ser Engenheiro Civil. O amor pela engenharia nunca arrefeceu; ao contrário, tornou-se o combustível que alimentou minha determinação. Sabia que não seria fácil, mas hoje, com este trabalho em mãos, posso afirmar com orgulho: eu consegui.

Esta conquista não é apenas minha. Carrego comigo o apoio de pessoas essenciais, a começar pela minha família. Aos meus pais, minha eterna gratidão por todo o suporte, incentivo e por terem sido a base fundamental na primeira metade deste curso. Mesmo após a distância imposta pela mudança de estado, o amor e os valores que me ensinaram permaneceram como meu alicerce.

Quando a saudade apertou, a vida me presenteou com amigos que se tornaram minha família aqui em Belém. A vocês, meus amigos, agradeço por cada palavra de incentivo, pela companhia nos momentos de estudo e, principalmente, por terem preenchido o vazio e tornado esta caminhada mais leve e feliz.

De forma especial, dedico esta vitória à minha namorada e à nossa filha. Vocês são o meu lugar seguro, meu refúgio e a fonte de todo o meu conforto. Nos dias mais exaustivos, era no amor de vocês que eu encontrava a força para continuar. Obrigado pela paciência, pelo carinho e por serem a razão maior de toda a minha dedicação.

Por fim, agradeço aos meus professores e mestres. O conhecimento que adquiri vai muito além dos livros e cálculos; aprendi com a paixão e a dedicação de cada um de vocês. Obrigado por compartilharem sua sabedoria e por me ajudarem a construir o conhecimento necessário para me tornar o Engenheiro Civil que sempre sonhei ser.

A todos vocês, meu muito obrigado.

RESUMO

Este trabalho propõe uma abordagem inovadora para a otimização multiobjetivo de misturas de concreto, visando conciliar resistência mecânica, redução de custos e diminuição da pegada de carbono (CO₂). A metodologia integra o algoritmo preditivo XGBoost, com quantificação de incertezas, e o algoritmo genético NSGA-II para otimização. O XGBoost foi refinado para prever a resistência do concreto, enquanto o NSGA-II explorou soluções que equilibram os múltiplos objetivos de desempenho e sustentabilidade.

Os resultados demonstram a eficácia do modelo de machine learning na otimização de misturas de concreto, validando estatística e matematicamente sua precisão e robustez em comparação com métodos tradicionais. Este estudo destaca o potencial da inteligência artificial para aprimorar a engenharia civil, oferecendo ferramentas para decisões mais informadas e sustentáveis. Futuras etapas incluem a realização de ensaios laboratoriais para validação prática dos resultados computacionais, consolidando a aplicabilidade da abordagem em um contexto real.

Palavra-chave: Otimização Multiobjetivo; Concreto; Machine Learning; XGBoost; Algoritmo Genético; Sustentabilidade; Resistência a Compressão; Traço de Concreto.

ABSTRACT

This work proposes an innovative approach for the multi-objective optimization of concrete mixtures, aiming to reconcile mechanical strength, cost reduction, and carbon footprint (CO₂) decrease. The methodology integrates the predictive XGBoost algorithm, with uncertainty quantification, and the NSGA-II genetic algorithm for optimization. XGBoost was refined to predict concrete strength, while NSGA-II explored solutions that balance multiple performance and sustainability objectives.

The results demonstrate the effectiveness of the machine learning model in optimizing concrete mixtures, statistically and mathematically validating its precision and robustness compared to traditional methods. This study highlights the potential of artificial intelligence to enhance civil engineering, offering tools for more informed and sustainable decisions. Future steps include laboratory testing for practical validation of computational results, consolidating the applicability of the proposed approach in a real context.

Keywords: Multi-objective Optimization; Concrete; Machine Learning; XGBoost; Genetic Algorithm; Sustainability; Compressive Strength; Concrete Mix Design.

LISTA DE ILUSTRAÇÕES

Figura 1 - Fluxograma da 2º etapa	34
Figura 2 - Fluxograma Geral de treinamento de Otimização Multiobjetiva de Mistura de Concreto.	37
Figura 3 - Matriz de Correlação	50

LISTA DE TABELAS

Tabela 1 - Fatores Determinantes da Resistência à Compressão do Concreto	20
Tabela 2 - Resultados Consolidados da Análise Multimodal (Média da Validação Cruzada K=5).	36
Tabela 3 - Componentes e Desempenho da Dosagem Otimizada 3.	53
Tabela 4 - Evolução da resistência (MPa) Mistura Otimizada 1	53
Tabela 5 - Componentes e Desempenho da Dosagem Otimizada 2.	54
Tabela 6- Evolução da resistência (MPa) Mistura Otimizada 2	54
Tabela 7 - Componentes e Desempenho da Dosagem Otimizada 3.	55
Tabela 8 - Evolução da resistência (MPa) Mistura Otimizada 3	55
Tabela 9 - Componentes e Desempenho das misturas tradicionais 1	57
Tabela 10 -Evolução da resistência (MPa) Misturas Tradicionais 1	57
Tabela 11 - Tabela 6 - Componentes e Desempenho das misturas tradicionais 2	58
Tabela 12 - Evolução da resistência (MPa) Misturas Tradicionais 2	59
Tabela 13 - Componentes e Desempenho das misturas tradicionais 3	60
Tabela 14 - Evolução da resistência (MPa) Misturas Tradicionais 3.	60

LISTA DE GRÁFICO

gráfico 1 - Distribuição da Resistência por Idade de Cura. _____	51
gráfico 2 - Distribuição dos Resíduos. _____	52
gráfico 3 - Gráfico de Resíduos Versus Valores Preditos. _____	52
gráfico 4 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 25 MPa. _____	53
gráfico 5 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 40 MPa. _____	54
gráfico 6 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 50 MPa. _____	55
gráfico 7 - Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 1. _____	58
gráfico 8 -Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 2. _____	59
gráfico 9 - Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 3. _____	60
gráfico 10 - Comparação de Componentes por método - 25 MPa. _____	62
gráfico 11 - Evolução da Resistência à compressão - 25 MPa. _____	63
gráfico 12 - Gráfico Radar - Comparativo de Métodos de Dosagem - 25 MPa. _____	64
gráfico 13 - Comparação de Componentes por método - 40 MPa _____	65
gráfico 14 - Evolução da Resistência à compressão - 40 MPa. _____	66
gráfico 15 - Gráfico Radar - Comparativo de Métodos de Dosagem - 40 MPa. _____	67
gráfico 16 - Comparação de Componentes por método - 50 MPa _____	68
gráfico 17 - Comparação de Componentes por método - 50 MPa. _____	68
gráfico 18 - Gráfico Radar - Comparativo de Métodos de Dosagem - 50 MPa. _____	69

LISTA DE ABREVIATURAS E SIGLAS

CCP	Concreto de Cimento Portland
ML	Machine Learning
NSGA-II	Non-dominated Sorting Genetic Algorithm II
C ₃ S	Silicato Tricálcico
C ₂ S	Silicato Bicálcico
C ₃ A	Aluminato Tricálcico
C ₄ AF	Ferroaluminato Tetracálcico
CSH	Silicato de Cálcio Hidratado
ABCP	Associação Brasileira de Cimento Portland
ACI	American Concrete Institute
IS	Indian Standard
BS	British Standard
EN	European Norm
MCS	Materiais Cimentícios Suplementares
RCD	Resíduos de Construção e Demolição
CCUS	Carbon Capture, Utilization and Storage
XAI	Explainable Artificial Intelligence - Inteligência Artificial Explicável
SVR	Support Vector Regression
MLP	Multi-Layer Perceptron
GPR	Gaussian Process Regressor
ANN	Redes Neurais Artificiais
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error

1. INTRODUÇÃO	14
2. CONCRETO DE CIMENTO PORTLAND	15
2.1. PROPRIEDADES DO CONCRETO NO ESTADO FRESCO	17
2.2. PRODUTOS DE HIDRATAÇÃO: FUNÇÕES E IMPORTÂNCIAS	18
2.3. RESISTÊNCIA A COMPRESSÃO	20
2.4. MÉTODOS TRADICIONAIS DE DOSAGEM	21
2.5. EMISSÕES DE CO ₂ : O IMPACTO AMBIENTAL DO USO DO CONCRETO DE CIMENTO PORTLAND.	23
3. ANÁLISE DE RISCO E INCERTEZAS NA DOSAGEM DE CONCRETO DE CIMENTO PORTLAND.	24
3.1. ANÁLISE DE RISCO	24
3.1.1. CONCEITO E CONCEPÇÕES DE RISCO	25
4. DESCONSTRUINDO AS BARREIRAS DA APRENDIZAGEM DE MÁQUINA: AVALIANDO MODELOS DE MACHINE LEARNING PARA PREVER A RESISTENCIA DO CONCRETO.	25
4.1. O ESPECTRO DE MODELOS PREDITIVOS	26
4.2. METODOLOGIA DE VALIDAÇÃO: ASSEGURANDO A ROBUSTEZ PREDITIVA.	28
5. MACHINE LEARNING	29
5.1. HIPERPARAMETRO EM UM MODELO DE ML	29
5.2. TIPOS DE ALGORITMO DE MACHINE LEARNING	29
5.3. MACHINE LEARNING APLICADO A CONCRETOS	30
5.4. PRINCIPAIS DESAFIOS DA ANÁLISE DE DADOS COM MACHINE LEARNING	31
6. METODOLOGIA	32
6.1. ETAPA 01- SELEÇÃO E PRÉ-PROCESSAMENTO DO BANCO DE DADOS.	32
6.2. ETAPA 02- ANÁLISE MULTIMODAL E SELEÇÃO DO MODELO DE MACHINE LEARNING.	34
6.2.1. RESULTADOS E DISCUSSÃO	35
6.3. ETAPA 03 - OTIMIZAÇÃO DE MISTURAS DE CONCRETO COM IA	36
6.3.1. Estrutura do Script e Bibliotecas Utilizadas	38
6.3.2. Pré-processamento de Dados	39
6.3.3. Engenharia de Features (Atributos)	40
6.3.4. Treinamento de Modelos: XGBoost Quantile	41
6.3.5. Predição de Resistência com Correção de Cruzamento de Quantis	43
6.3.6. Otimização Multiobjetivo	44
6.3.7. Visualização de Dados	45
6.3.8. Função Principal (main())	46
7. RESULTADOS	47
7.1. ANÁLISE MULTIMODAL E SELEÇÃO DO MODELO PREDITIVO.	48
7.2. OTIMIZAÇÃO MULTI OBJETIVO DAS MISTURAS DE CONCRETO.	48
7.2.1. MISTURA OTIMIZADA PARA 25 MPA AOS 28 DIAS	52
7.2.2. MISTURA OTIMIZADA PARA 40 MPA AOS 28 DIAS	53
7.2.3. MISTURA OTIMIZADA PARA 50 MPA AOS 28 DIAS	54

7.3.	MISTURA A PARTIR DE MODELOS TRADICIONAIS.	56
7.3.1.	MISTURA TRADICIONAL PARA 25 MPA AOS 28 DIAS	57
7.3.2.	MISTURA TRADICIONAL PARA 40 MPA AOS 28 DIAS	58
7.3.3.	MISTURA TRADICIONAL PARA 50 MPA AOS 28 DIAS	59
7.4.	COMPARAÇÃO: DOSAGEM OTIMIZADA COM ML VS MÉTODOS DE DOSAGEM TRADICIONAIS.	61
7.4.1.	COMPARAÇÃO DAS MISTURA PARA 25 MPA AOS 28 DIAS	61
7.4.2.	COMPARAÇÃO DAS MISTURA PARA 40 MPA AOS 28 DIAS	64
7.4.3.	COMPARAÇÃO DAS MISTURA PARA 50 MPA AOS 28 DIAS	67
8.	AVALIAÇÃO E CONCLUSÃO	70
9.	REFERÊNCIAS BIBLIOGRÁFICAS	71

1. INTRODUÇÃO

Há uma relação direta entre população e urbanização. No último século, a população mundial experimentou um crescimento enorme, indo de 1,5 bilhão para 7 bilhões de habitantes, com maior concentração nos grandes centros urbanos. Atualmente, aproximadamente 43% dessa população reside em cidades e suas regiões metropolitanas. Vale frisar que 17 das 20 megacidades com 10 milhões ou mais de habitantes, estão localizadas em países em desenvolvimento (MEHTA; MONTEIRO, 2014). Para suportar essa concentração populacional é impulsionada uma demanda massiva por materiais de construção, essenciais para a edificação de moradias, instalações industriais e comerciais, sistemas de água potável e saneamento, barragens e canais, além de infraestruturas de transporte como estradas, pontes e túneis. Nessa conjuntura de urbanização e desenvolvimento, o Concreto de Cimento Portland (CCP) emerge como o principal material de construção utilizado mundialmente, sendo, em volume, o produto de maior produção no mundo contemporâneo (MEHTA; MONTEIRO, 2014).

Embora a resistência do concreto seja frequentemente apontada como sua propriedade de maior valor, em diversas aplicações práticas, outras características como a durabilidade, a impermeabilidade e a estabilidade volumétrica podem ser consideradas de maior relevância. Contudo, a resistência ainda oferece uma visão abrangente da qualidade do concreto, pois sua magnitude está intrinsecamente ligada à microestrutura da pasta de cimento (NEVILLE; BROOKS, 2010). A partir dessa visão, é possível presumir junto ao crescimento urbano, a importância de garantir que as estruturas de concreto atendam as demandas de resistências e segurança para as quais foram projetadas, e a capacidade de prever com precisão a resistência a compressão do concreto é um determinante essencial no desempenho estrutural.

Segundo MALHOTRA et al. (2017) o ensaio de compressão, é um procedimento padronizado, que avalia a capacidade de resistência do concreto, considerando diferentes proporções de seus componentes: cimento, agregado miúdo e agregado graúdo, além da quantidade de água utilizada na mistura. Tradicionalmente, a avaliação da resistência do concreto depende de métodos experimentais e ensaios laboratoriais, processos que demandam tempo e recursos significativos. Consequentemente, a dosagem experimental, sendo uma abordagem empírica, não resulta na formulação de um modelo matemático preciso, principalmente devido à natureza heterogênea do concreto. (OLIVEIRA et al. 2007). Essa limitação tem motivado a crescente adoção de técnicas computacionais e, particularmente, de *Machine Learning* (ML) para a previsão das propriedades dos materiais de construção.

Diante dos avanços recentes na aplicação de aprendizado de máquina na engenharia civil, este trabalho propõe uma abordagem estruturada e explicável voltada à **otimização de misturas de concreto**. Inicialmente, realiza-se uma **análise comparativa multimodal de modelos preditivos de ML**, a fim de identificar aquele com melhor desempenho diante do conjunto de dados adotado. Tal prática é fundamental, dado que não existe um algoritmo universalmente superior — sendo necessário testar diferentes modelos e ajustá-los conforme o problema (DOMINGOS, 2012).

Após a seleção do modelo mais promissor, procede-se ao **refinamento preditivo utilizando o algoritmo XGBoost**, com três variações baseadas em regressão por quantis: o modelo central (quantil 0.5), para **previsão pontual da resistência**, e dois modelos auxiliares (quantis 0.05 e 0.9), com a finalidade de **quantificar incertezas** e fornecer **intervalos de confiança calibrados** (TAGASOVSKA; LOPEZ-PAZ, 2018/2019; CHEN; GUESTRIN, 2023). Essa abordagem contribui para aumentar a **confiabilidade nas decisões baseadas em modelos de ML** no contexto da engenharia de materiais.

Posteriormente, a estrutura metodológica é complementada com uma **otimização multiobjetiva baseada no algoritmo genético NSGA-II** (Non-dominated Sorting Genetic Algorithm II), o qual é amplamente reconhecido por sua eficiência a manipulação de grandes populações e pela capacidade de manter diversidade entre as soluções. Isso é possível graças à sua estrutura baseada em **ordenamento rápido por não dominância, estratégia elitista** e ao uso do **mecanismo de distância de aglomeração (crowding distance)**, que assegura uma distribuição adequada da **fronteira de Pareto** (DEB et al., 2002). O objetivo é **conciliar resistência mecânica, custo da mistura e pegada de carbono (CO₂)**, promovendo soluções tecnicamente viáveis e ambientalmente sustentáveis.

A proposta visa atender às necessidades reais de projetos em engenharia civil, integrando eficiência técnica, sustentabilidade e explicabilidade. Por fim, a eficácia do modelo será validada com um **conjunto de dados estruturado**, representativo de aplicações concretas no setor da construção civil, reforçando sua aplicabilidade prática.

2. CONCRETO DE CIMENTO PORTLAND

O concreto de cimento Portland constitui o alicerce da infraestrutura moderna, sendo o material de construção mais consumido globalmente. Sua versatilidade e desempenho derivam de seu principal componente ativo: o cimento Portland. Este é definido como um aglomerante hidráulico, produzido a partir da calcinação de matérias-primas como calcário e argila em fornos rotativos a temperaturas de aproximadamente 1400 °C, o que resulta em um produto intermediário granular denominado clínquer. Subsequentemente, o clínquer é resfriado e moído finamente com uma pequena quantidade de gipsita (sulfato de cálcio), que atua como reguladora do tempo de pega. Ao ser misturado com água, o cimento forma uma pasta com propriedades ligantes que endurece e desenvolve resistência mecânica, mesmo quando submerso, o que caracteriza sua essencial natureza hidráulica para a engenharia civil (MEHTA; MONTEIRO, 2014).

A reatividade e, conseqüentemente, as propriedades do cimento são determinadas pela composição química do clínquer. Durante a calcinação, formam-se quatro compostos mineralógicos principais: o silicato tricálcico ($3\text{CaO}\cdot\text{SiO}_2$, ou C_3S na notação cimenteira), o silicato bicálcico ($2\text{CaO}\cdot\text{SiO}_2$, ou C_2S), o aluminato tricálcico ($3\text{CaO}\cdot\text{Al}_2\text{O}_3$, ou C_3A) e o ferroaluminato tetracálcico ($4\text{CaO}\cdot\text{Al}_2\text{O}_3\cdot\text{Fe}_2\text{O}_3$, ou C_4AF) (NEVILLE, 2011). A proporção desses compostos é controlada para produzir diferentes tipos de cimento com características específicas. Por exemplo, um cimento de alta resistência inicial (CP V-ARI) é formulado para ter um teor elevado de C_3S , principal responsável pelo desenvolvimento de resistência nas primeiras idades (MEHTA; MONTEIRO, 2014). Em contrapartida, o C_2S hidrata-se mais lentamente, contribuindo significativamente para o ganho de resistência em idades mais avançadas (TAYLOR, 1997). Já o C_3A reage muito rapidamente e contribui para a resistência no primeiro dia, sendo também o principal gerador de calor de hidratação, um fator crítico em concretagens de grandes volumes (NEVILLE, 2011; MEHTA; MONTEIRO, 2014).

O processo de endurecimento do cimento é denominado hidratação, uma série de reações químicas exotérmicas que ocorrem quando os compostos do clínquer entram em contato com a água. O produto mais importante e abundante dessas reações é o **silicato de cálcio hidratado (C-S-H)**, um gel amorfo que atua como o principal agente ligante da matriz cimentícia, envolvendo os agregados e preenchendo os vazios. A microestrutura e a quantidade de C-S-H formado são os fatores que governam, em grande parte, as propriedades mecânicas e de durabilidade do concreto endurecido. Um subproduto

significativo da hidratação do C_3S e do C_2S é o hidróxido de cálcio ($Ca(OH)_2$), também conhecido como portlandita, que contribui para a alcalinidade da pasta, essencial para a proteção da armadura de aço contra a corrosão. A engenharia de propriedades do concreto, portanto, começa no nível microestrutural, com a seleção de um tipo de cimento cujas características químicas e de reatividade são otimizadas para a aplicação final. A existência de diversos tipos de cimento, como o Comum (CP I), Composto (CP II, com adições de escória, pozolana ou filler), de Alto-Forno (CP III) e Pozolânico (CP IV), reflete essa engenharia deliberada, onde as adições não apenas substituem parte do clínquer por razões econômicas e ambientais, mas também conferem propriedades específicas como maior impermeabilidade, durabilidade em ambientes agressivos e menor calor de hidratação.

Embora o cimento seja o componente quimicamente ativo, o concreto é, por definição, um material compósito. Os agregados e a água, embora frequentemente vistos como constituintes mais passivos, desempenham papéis volumétricos, econômicos e funcionais que são igualmente críticos para o desempenho final da mistura. Os agregados, materiais minerais inertes como areia e brita, compõem a maior parte do volume do concreto, tipicamente entre 60% e 80% do seu peso. Sua função principal é fornecer um esqueleto granular estável, conferindo "corpo" à mistura, o que reduz significativamente o custo total (dado que o cimento é o componente mais caro) e mitiga os efeitos da retração da pasta de cimento durante o endurecimento.

Os agregados são classificados com base em sua granulometria em miúdos (areia), que passam pela peneira de malha de 4,8 mm, e grãos (brita ou seixo), que ficam retidos nela. As características desses materiais, como a distribuição granulométrica (curva granulométrica), a forma e a textura das partículas, e a ausência de impurezas (como argila e matéria orgânica), exercem uma influência profunda nas propriedades do concreto, tanto no estado fresco quanto no endurecido. Uma curva granulométrica bem distribuída, com uma variedade de tamanhos de partículas, minimiza o volume de vazios entre os agregados, o que, por sua vez, reduz a quantidade de pasta de cimento necessária para preencher esses vazios, resultando em um concreto mais denso, resistente e econômico.

A água, por sua vez, desempenha uma dupla função essencial e, em certos aspectos, conflitante. Primeiramente, é o reagente indispensável para a hidratação do cimento, o processo químico que confere ao concreto suas propriedades de resistência e durabilidade. Sem água suficiente, as reações de hidratação cessam prematuramente, comprometendo o desenvolvimento do potencial máximo do material. Em segundo lugar, a água confere fluidez e plasticidade à mistura fresca, uma propriedade conhecida como trabalhabilidade. Essa fluidez é necessária para que o concreto possa ser transportado, lançado nas fôrmas, adensado em torno das armaduras e acabado adequadamente, sem a formação de vazios ou segregação.

É na interação entre essas duas funções que reside o dilema fundamental da dosagem de concreto. Para o estado fresco, um maior teor de água geralmente melhora a trabalhabilidade, facilitando a aplicação. No entanto, para o estado endurecido, qualquer água que não seja consumida na hidratação evapora, abandonando uma rede de poros capilares. Essa porosidade aumentada enfraquece a matriz, reduzindo drasticamente a resistência à compressão e a durabilidade do concreto. Essa relação crítica é quantificada pela relação água/cimento (a/c), definida como a razão em massa entre a água e o cimento na mistura. A Lei de Abrams, um dos princípios mais fundamentais da tecnologia do concreto, estabelece que, para um dado conjunto de materiais e condições de cura, a resistência do concreto é inversamente proporcional à sua relação a/c . Portanto, a busca por um concreto de alto desempenho envolve a minimização da relação a/c .

Este conflito intrínseco — a necessidade de água para a trabalhabilidade versus os prejuízos do excesso de água para a resistência — é o cerne do desafio da dosagem. A solução tradicional para este dilema tem sido o uso de aditivos químicos, especialmente os plastificantes ou redutores de água. Esses aditivos atuam dispersando as partículas de cimento, o que aumenta a fluidez da mistura sem a necessidade de adicionar mais água, permitindo assim a produção de concretos trabalháveis com baixas relações a/c. A otimização da dosagem, portanto, não é apenas uma questão de misturar ingredientes, mas de gerenciar esse delicado equilíbrio para atender simultaneamente aos requisitos de desempenho no estado fresco e no estado endurecido.

2.1. PROPRIEDADES DO CONCRETO NO ESTADO FRESCO

As propriedades do concreto no seu estado fresco, ou seja, no período entre a mistura e o início do endurecimento, são determinantes para a qualidade final da estrutura. Um concreto que não pode ser transportado, lançado e adensado de forma eficiente e homogênea resultará inevitavelmente em um produto endurecido com defeitos, comprometendo seu desempenho mecânico e sua durabilidade. As principais propriedades a serem controladas nesta fase são a trabalhabilidade, a consistência, e a resistência a fenômenos patológicos como a segregação e a exsudação.

A trabalhabilidade é um termo qualitativo que descreve a facilidade com que o concreto pode ser misturado, transportado, lançado e adensado com o mínimo de perda de homogeneidade. É uma propriedade complexa influenciada por múltiplos fatores, incluindo a relação água/cimento, as características dos agregados (granulometria, forma e textura), a finura do cimento e o uso de aditivos. A consistência é a principal medida quantitativa da trabalhabilidade e refere-se à fluidez ou mobilidade da massa de concreto. O método padrão para avaliar a consistência em campo e em laboratório é o ensaio de abatimento do tronco de cone (Slump Test), normatizado no Brasil pela ABNT NBR NM 67. O procedimento consiste em preencher um molde em forma de tronco de cone com concreto fresco em três camadas de igual altura. Cada camada é compactada com 25 golpes de uma haste de aço padronizada. Após o preenchimento e o rasamento da superfície, o molde é removido verticalmente, e a diferença de altura entre o molde e a massa de concreto assentada é medida. Esse valor, expresso em milímetros, é o abatimento ou "slump". Um slump maior indica um concreto mais fluido e, geralmente, mais trabalhável.

No estado fresco, o concreto está suscetível a dois fenômenos indesejáveis: a segregação e a exsudação. A segregação refere-se à separação dos componentes do concreto devido a diferenças de massa específica e tamanho, onde os agregados graúdos tendem a se assentar e a argamassa (cimento, areia e água) ascende. Esse fenômeno é intensificado por misturas com baixa coesão, excesso de água, vibração excessiva ou transporte inadequado, resultando em regiões não homogêneas na estrutura, como "ninhos de concretagem" ou "bicheiras", que comprometem a resistência e a durabilidade.

A exsudação, ou sangramento (bleeding), é uma forma de segregação em que a água da mistura migra para a superfície do concreto recém-lançado. Esse movimento ascendente pode arrastar partículas finas de cimento, formando uma camada superficial de nata de cimento. A exsudação excessiva é prejudicial, pois a camada superficial torna-se fraca, porosa e pouco resistente à abrasão. Além disso, a água aprisionada sob agregados graúdos e armaduras pode criar vazios, enfraquecendo a zona de transição pasta-agregado e a aderência entre o concreto e o aço.

A interconexão entre essas propriedades estabelece uma clara cadeia causal: uma dosagem inadequada, como uma relação a/c excessiva ou uma granulometria deficiente, leva a uma baixa coesão da mistura. Essa baixa coesão promove a segregação e a

exsudação durante o manuseio e o adensamento. Esses fenômenos, por sua vez, resultam em uma microestrutura final heterogênea, com porosidade elevada, zonas de transição fracas e aderência comprometida. O resultado é um concreto com resistência e durabilidade significativamente inferiores às projetadas. Portanto, o controle das propriedades no estado fresco não é meramente uma questão de conveniência executiva; é um pré-requisito fundamental para garantir a integridade e a vida útil da estrutura. Qualquer modelo de otimização de concreto deve, obrigatoriamente, tratar a trabalhabilidade (medida pelo slump) como uma restrição crítica de desempenho, e não apenas como uma preferência.

2.2. PRODUTOS DE HIDRATAÇÃO: FUNÇÕES E IMPORTÂNCIAS

A hidratação do cimento Portland é um processo complexo de reações exotérmicas entre a água e os compostos mineralógicos do clínquer, principalmente o silicato tricálcico (C₃S), o silicato bicálcico (C₂S), o aluminato tricálcico (C₃A) e o ferroaluminato tetracálcico (C₄AF). Essas reações dão origem a uma matriz sólida e resistente, cujas propriedades são definidas pelos produtos formados, notadamente o silicato de cálcio hidratado (C-S-H) e o hidróxido de cálcio (portlandita) (NEVILLE, 2011). A seguir, detalha-se a contribuição dos principais compostos do clínquer para a formação desses produtos e, conseqüentemente, para o desempenho do concreto.

O hidróxido de cálcio, também conhecido como portlandita, é um dos primeiros produtos da hidratação do cimento em solução aquosa. Sua presença confere basicidade ao material, fundamental para a proteção contra corrosão. Em contrapartida, seu excesso pode ocasionar patologias como a lixiviação, que é a dissolução do hidróxido de cálcio por meio de água que infiltra na estrutura e transporta esses sais para a superfície (efeito: eflorescências, perda da resistência mecânica devido à perda de sólidos do concreto, além de expor a armadura e o próprio concreto/argamassa a agentes agressivos). No entanto, a portlandita tem grande importância, pois reage com as escórias de alto-forno e os materiais pozzolânicos presentes em alguns cimentos, conferindo propriedades e características específicas a esses materiais.

- **SILICATO TRICÁLCICO (C₃S OU ALITA)**

O silicato tricálcico (3CaO·SiO₂), ou alita, é o composto presente em maior proporção nos cimentos de uso geral (tipicamente entre 45% e 65%) e o principal responsável pelo desenvolvimento da resistência nas idades iniciais, especialmente na primeira semana (MEHTA; MONTEIRO, 2014). Sua reação de hidratação é relativamente rápida e libera uma quantidade significativa de calor. A reação produz o silicato de cálcio hidratado (C-S-H), que é o principal responsável pela resistência, e uma quantidade considerável de hidróxido de cálcio [Ca(OH)₂], conforme a equação:

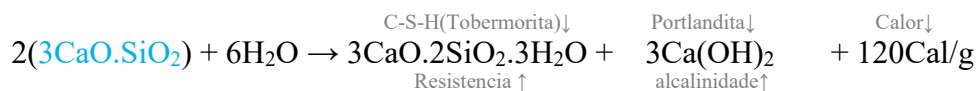


Figura 1 – reação hidratação do silicato tricálcico

A grande quantidade de hidróxido de cálcio liberada eleva o pH da pasta para valores superiores a 12.5, criando um ambiente alcalino que é fundamental para a passivação e proteção da armadura de aço contra a corrosão. Por outro lado, o Ca(OH)₂ é um composto solúvel e vulnerável à lixiviação, um processo em que a percolação de água

dissolve e carrega o hidróxido para a superfície, podendo causar eflorescências, aumento da porosidade e redução da durabilidade (TAYLOR, 1997).

- **SILICATO BICÁLCICO (C₂S OU BELITA)**

O silicato bicálcico (2CaO·SiO₂), ou belita, compõe entre 15% e 35% do cimento Portland comum. Sua hidratação é marcadamente mais lenta que a do C₃S, iniciando-se de forma significativa apenas após vários dias. Conseqüentemente, a belita contribui pouco para a resistência inicial, mas é a principal responsável pelo ganho de resistência em idades avançadas (após 28 dias e ao longo de anos) (NEVILLE, 2011). A reação gera proporcionalmente mais C-S-H e menos hidróxido de cálcio que a do C₃S, além de liberar menos da metade do calor de hidratação.

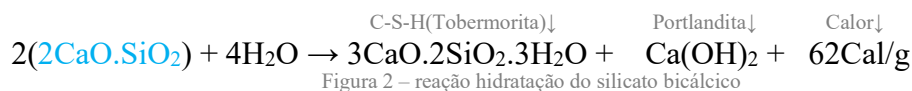


Figura 2 – reação hidratação do silicato bicálcico

Devido ao baixo calor de hidratação, cimentos ricos em C₂S são ideais para estruturas de grande volume, minimizando o risco de fissuração térmica. A menor produção de Ca(OH)₂ torna o concreto mais resistente a ataques químicos, como o de sulfatos, e menos propenso à lixiviação, resultando em uma matriz mais densa e durável a longo prazo (MEHTA; MONTEIRO, 2014).

- **ALUMINATO TRICÁLCICO (C₃A) E FERROALUMINATO TETRACÁLCICO (C₄AF)**

O aluminato tricálcico (3CaO·Al₂O₃), presente em teores de 7% a 15%, é o composto mais reativo do cimento. Sua hidratação é quase instantânea e libera uma grande quantidade de calor. Para evitar a "pega relâmpago" (endurecimento imediato), adiciona-se gesso (CaSO₄·2H₂O) durante a moagem do clínquer. O gesso reage com o C₃A para formar uma camada de cristais de etringita sobre as partículas, controlando sua reatividade e regulando o tempo de pega (TAYLOR, 1997). Embora contribua para a resistência no primeiro dia, sua presença em excesso pode ser prejudicial à durabilidade, especialmente em ambientes com sulfatos, onde pode formar etringita tardia, causando expansão e fissuração do concreto. O ferroaluminato tetracálcico (4CaO·Al₂O₃·Fe₂O₃ ou C₄AF), presente em teores de 6% a 10%, é o menos reativo dos compostos principais. Sua hidratação é lenta e contribui pouco para a resistência inicial, mas confere a coloração cinza ao cimento. Sua importância reside principalmente na sua capacidade de reagir com o gesso para formar sulfoaluminato de cálcio (CSA), que contribuem para a resistência e durabilidade do concreto (NEVILLE, 2011).

- **PRODUTOS SECUNDÁRIOS E IMPUREZAS**

Além dos produtos principais, a hidratação gera subprodutos e lida com impurezas do clínquer. Os óxidos de sódio (Na₂O) e potássio (K₂O), conhecidos como álcalis, são as impurezas mais relevantes. Em contato com a água, formam hidróxidos altamente alcalinos que podem reagir com agregados reativos (reação álcali-agregado), gerando um gel expansivo que causa microfissuração interna e compromete a integridade do concreto. Por essa razão, as normas técnicas limitam o teor de álcalis nos cimentos, especialmente naqueles destinados a obras de grande responsabilidade (Neville, 2011).

2.3. RESISTÊNCIA A COMPRESSÃO

Após o período de cura, o concreto transita do estado fresco para o endurecido, desenvolvendo as propriedades mecânicas que o tornam um material estrutural por excelência. Dentre essas propriedades, a resistência à compressão (f_c) é universalmente reconhecida como a mais importante e fundamental. Ela não apenas serve como o principal parâmetro para o dimensionamento de elementos estruturais de concreto armado e protendido, mas também funciona como um indicador geral da qualidade do concreto, pois está intrinsecamente correlacionada com outras propriedades importantes, como a durabilidade, o módulo de elasticidade e a resistência à abrasão.

A resistência à compressão de um composto cimentício é, em sua essência, uma função inversa de sua porosidade total. A matriz de pasta de cimento endurecida contém poros de gel, poros capilares e, potencialmente, vazios de ar aprisionado. Quanto maior o volume total desses vazios, menor será a área sólida efetiva para resistir às tensões aplicadas, resultando em menor resistência. Diversos fatores interligados influenciam a porosidade e, por conseguinte, a resistência final do concreto, conforme detalhado na Tabela 1.

Tabela 1 - Fatores Determinantes da Resistência à Compressão do Concreto

Fator	Mecanismo de Influência
Relação Água/Cimento (a/c)	Controla a porosidade da pasta de cimento. Menor a/c resulta em menor porosidade e maior resistência (Lei de Abrams).
Porosidade (Ar aprisionado)	Vazios não contribuem para a resistência, atuando como concentradores de tensão.
Grau de Hidratação (Idade/Cura)	O desenvolvimento do gel C-S-H, responsável pela resistência, depende da progressão das reações de hidratação. A cura adequada garante água para essas reações.
Tipo e Finura do Cimento	A composição química (C3S, C2S) e a finura do cimento determinam a velocidade e o grau das reações de hidratação e, conseqüentemente, a taxa de ganho de resistência.
Qualidade do Agregado	A resistência intrínseca do agregado pode ser o elo mais fraco em concretos de alta resistência. A aderência pasta-agregado também é crucial.
Adensamento	A compactação adequada da mistura remove o ar aprisionado, reduzindo a porosidade e aumentando a densidade e a resistência.

O fator mais dominante é, sem dúvida, a relação água/cimento (a/c). A Lei de Abrams formaliza essa relação inversa, estabelecendo que, para concretos plásticos e bem adensados, a resistência é determinada quase que exclusivamente pela relação a/c. Uma relação a/c baixa leva a uma pasta de cimento mais densa, com menor volume de poros capilares, resultando em um concreto mais resistente e durável. O grau de hidratação também é crucial; a resistência aumenta com o tempo (idade do concreto) à medida que mais produtos de hidratação são formados. Esse processo, no entanto, requer a presença contínua de água. A cura — o processo de manter o concreto úmido e a uma temperatura favorável, especialmente nas primeiras idades — é, portanto, essencial para permitir que a hidratação prossiga e a resistência potencial seja alcançada. Uma cura inadequada pode

levar à evaporação prematura da água, interrompendo a hidratação e resultando em uma resistência significativamente menor do que a projetada.

A determinação da resistência à compressão é realizada por meio de um ensaio padronizado, descrito no Brasil pela ABNT NBR 5739. O procedimento envolve a moldagem de corpos de prova cilíndricos, que são submetidos a um processo de cura controlada até a idade de ensaio (tipicamente 28 dias). No ensaio, o corpo de prova é posicionado em uma prensa hidráulica e submetido a uma carga de compressão axial, aplicada de forma contínua e a uma taxa constante, até a sua ruptura. A resistência à compressão é então calculada dividindo-se a carga máxima suportada pelo corpo de prova pela área de sua seção transversal. Este valor, conhecido como resistência característica à compressão (f_{ck}), é o parâmetro utilizado nos cálculos de projeto estrutural.

2.4. MÉTODOS TRADICIONAIS DE DOSAGEM

A dosagem do concreto é o procedimento técnico de selecionar e proporcionar os materiais constituintes — cimento, agregados, água e aditivos — de modo a obter uma mistura que atenda a requisitos predeterminados no estado fresco e no estado endurecido. Entre os métodos tradicionais amplamente utilizados, destacam-se as práticas orientadas pela **ABCP**, o guia **ACI 211.1**, o **IS 10262** (guia indiano de dosagem) e as orientações práticas do **BS 8500 / EN 206** (norma europeia complementada pelo código britânico). Estes métodos compartilham uma base conceitual comum: partem da definição de uma resistência alvo (f_{ck}) e de uma trabalhabilidade desejada ($slump$) e, a partir daí, utilizam tabelas, curvas e procedimentos empíricos para estimar a relação água/cimento inicial, o volume de água necessário e as proporções dos agregados.

No método prático adotado pela **ABCP**, as estimativas iniciais são calibradas segundo parâmetros locais (consumo de água para determinada consistência, fração entre finos e graúdos, módulo de finura) e confirmadas por meio de ensaios-piloto em laboratório. A **ABCP** representa a prática consolidada no Brasil e costuma ser a referência operacional para usinas e canteiros que dependem de materiais regionais e rotinas estabelecidas.

O **ACI 211.1** fornece um guia estruturado e amplamente referenciado internacionalmente para a proporção de traços de concreto. O **ACI** descreve passo a passo a estimativa da relação a/c , do conteúdo de água e do teor de cimento, além de orientar a seleção granulométrica e os procedimentos de verificação laboratorial. Sua força é a clareza e a ampla base experimental incorporada às recomendações.

O **IS 10262** é um guia prático utilizado na Índia para proporção de misturas; sua metodologia é análoga às do **ACI** ao oferecer procedimentos passo a passo para estimar relação a/c , consumo de água e proporções de agregados, com parâmetros ajustados à realidade dos materiais e práticas locais na região em que é aplicado. Por isso é um método internacionalmente relevante para comparação e aplicação em contextos que empreguem práticas ou materiais similares.

A combinação **BS 8500 / EN 206** representa a abordagem europeia para especificação e orientação prática do concreto. A **EN 206** estabelece requisitos de especificação, produção e conformidade do concreto (incluindo classes de exposição e limites de desempenho), enquanto o **BS 8500** — quando presente — fornece orientações práticas e nacionais complementares para dosagem e aplicação no Reino Unido. Juntas, essas normas suportam tanto a prescrição normativa quanto recomendações de dosagem adaptadas a condições europeias.

Em termos operacionais, o fluxo típico desses métodos segue passos bem estabelecidos: (i) fixação de f_{ck} e $slump$; (ii) consulta a tabelas/curvas para estimativa

inicial de relação a/c e consumo de água; (iii) dimensionamento das frações de agregado com base em conceitos de granulometria e volume de vazios; (iv) produção de misturas experimentais para verificação e ajustes. Esse processo é eficiente para concretos convencionais produzidos com materiais cuja caracterização está alinhada às premissas dos métodos.

Todavia, a crescente demanda por redução do teor de clínquer (uso de MCS - Materiais Cimentícios Suplementares), adoção de agregados reciclados e metas de diminuição das emissões de CO₂ impõe variabilidade e desafios que aumentam a incerteza das estimativas obtidas por procedimentos empíricos. Por esse motivo, e sem desconsiderar a utilidade dos métodos tradicionais, este trabalho opta por manter ABCP, ACI, IS 10262 e BS 8500/EN 206 como referências técnicas e comparativas, utilizando-as como base de comparação para abordagens de otimização suportadas por Machine Learning, que serão apresentadas e validadas nas seções seguintes.

A indústria da construção civil enfrenta uma pressão crescente para mitigar seu significativo impacto ambiental. A produção de cimento Portland, em particular, é um processo intensivo em energia e responsável por uma parcela substancial das emissões globais de dióxido de carbono (CO₂). Nesse contexto, o paradigma da sustentabilidade deixou de ser uma consideração secundária para se tornar um requisito central no desenvolvimento de novos materiais de construção, incluindo o concreto.

A busca por um concreto mais sustentável se manifesta principalmente em duas estratégias: a redução do consumo de clínquer de cimento e a substituição de agregados naturais por materiais reciclados. A primeira estratégia é alcançada através do uso de Materiais Cimentícios Suplementares (MCS), como a cinza volante (resíduo da queima de carvão em termoelétricas), a escória granulada de alto-forno (subproduto da indústria siderúrgica) e o metacaulim. Esses materiais não apenas substituem uma parte do cimento, reduzindo o custo e a pegada de carbono do concreto, mas também podem melhorar suas propriedades de longo prazo, como a durabilidade e a resistência a ataques químicos, devido às suas reações pozolânicas ou cimentícias. A segunda estratégia envolve o uso de agregados reciclados, provenientes da britagem de Resíduos de Construção e Demolição (RCD), como uma alternativa aos agregados virgens extraídos de fontes naturais.

A incorporação desses materiais alternativos, no entanto, introduz um nível de complexidade e variabilidade que os métodos de dosagem convencionais não foram projetados para gerenciar. Os agregados reciclados, por exemplo, geralmente possuem uma pasta de cimento antiga aderida à sua superfície, o que os torna mais porosos e com maior capacidade de absorção de água em comparação com os agregados naturais. Essa característica afeta diretamente a trabalhabilidade da mistura e a relação água/cimento efetiva, tornando o controle do concreto mais desafiador. Da mesma forma, os diferentes tipos de MCS possuem cinéticas de reação distintas, influenciando a taxa de ganho de resistência, o calor de hidratação e o desenvolvimento da microestrutura de maneiras complexas.

Essa complexidade transforma fundamentalmente o problema da dosagem. Não se trata mais de atingir uma única meta de resistência (fck). Em vez disso, a dosagem de um concreto sustentável torna-se um problema de otimização multiobjetivo, no qual é necessário encontrar um equilíbrio entre múltiplos objetivos, muitas vezes conflitantes. Os objetivos a serem otimizados simultaneamente incluem:

- **Minimização do Custo:** Reduzir o consumo de cimento e utilizar agregados reciclados de menor custo.

- **Minimização do Impacto Ambiental:** Diminuir as emissões de CO₂ incorporado, principalmente pela redução do teor de clínquer.
- **Maximização do Desempenho Mecânico:** Atingir a resistência à compressão e outras propriedades mecânicas especificadas em projeto.

O conflito entre esses objetivos é evidente. Por exemplo, a substituição agressiva de cimento por MCS para reduzir custos e emissões pode levar a um ganho de resistência inicial mais lento, impactando os cronogramas da obra. O uso de agregados reciclados para promover a circularidade pode exigir mais água ou aditivos para manter a trabalhabilidade, o que pode afetar a resistência e o custo final. Os métodos de dosagem convencionais, por serem uniojetivo e reativos, são incapazes de navegar eficientemente neste espaço de busca multidimensional e complexo. Eles não conseguem identificar o conjunto de soluções ótimas de compromisso, conhecido como a "Fronteira de Pareto". Essa lacuna metodológica cria a necessidade imperativa de abordagens mais avançadas e inteligentes, capazes de lidar com a otimização multiobjetivo de forma sistemática e eficiente, justificando plenamente a aplicação de técnicas de inteligência artificial como a proposta neste trabalho.

2.5. EMISSÕES DE CO₂: O IMPACTO AMBIENTAL DO USO DO CONCRETO DE CIMENTO PORTLAND.

O concreto de cimento Portland é o material de construção mais consumido no mundo, superado em volume apenas pela água. Sua onipresença na infraestrutura global, de edifícios a barragens e estradas, o torna um pilar da sociedade moderna. Contudo, essa produção em larga escala acarreta um impacto ambiental significativo, notadamente através da sua elevada pegada de carbono, que está intrinsecamente ligada ao seu principal componente aglomerante: o cimento Portland (Mehta & Monteiro, 2014).

A intensidade carbônica do cimento deriva de dois fatores principais em sua produção. O primeiro, e mais significativo, é o processo químico de descarbonatação (ou calcinação) da matéria-prima, a pedra calcária (CaCO₃). Para produzir o óxido de cálcio (CaO), um componente essencial do clínquer, o calcário é aquecido a altas temperaturas, liberando dióxido de carbono diretamente na atmosfera como um subproduto da reação (CaCO₃ → CaO + CO₂). Este processo é responsável por aproximadamente 60% das emissões totais da fabricação do cimento. Os 40% restantes são, em sua maioria, provenientes da queima de combustíveis fósseis para aquecer os fornos rotativos a temperaturas que alcançam 1450 °C, necessárias para a clínquerização (IEA, 2023).

Para quantificar essa pegada de carbono, dados de inventários de ciclo de vida (ICV) são ferramentas essenciais. A base de dados *Inventory of Carbon and Energy* (ICE), mantida pela Circular Ecology, é uma referência internacionalmente reconhecida para o carbono incorporado em materiais de construção. De acordo com a base de dados ICE, o cimento Portland comum (tipo CEM I) possui um valor médio de carbono incorporado de **0,83 kg de CO₂ equivalente (kgCO₂e) por quilograma de cimento** (Hammond & Jones, 2019).

Este alto valor do aglomerante se reflete diretamente na pegada de carbono do concreto. Embora os agregados e a água tenham um impacto consideravelmente menor, a fração de cimento, mesmo correspondendo a apenas 10-15% da massa do concreto, domina seu impacto ambiental. Utilizando os mesmos dados de referência, um concreto de resistência convencional (ex: C20/25) apresenta um valor de carbono incorporado de aproximadamente **0,138 kgCO₂e por quilograma de concreto**. Em termos de volume,

uma unidade funcional mais prática para o material, isso se traduz em cerca de **330 kgCO₂ e por metro cúbico (m³)** de concreto (Hammond & Jones, 2019).

Quando esses números são extrapolados para a escala global, o impacto da indústria cimenteira torna-se evidente. Estima-se que a produção de cimento seja responsável por cerca de **7% a 8% do total das emissões antropogênicas de CO₂** no planeta, posicionando-a como uma das maiores fontes industriais de gases de efeito estufa (IEA, 2023). Portanto, a busca por estratégias de mitigação, como o uso de cimentos com adições (CP II, CP III, CP IV), a otimização de traços de concreto, o desenvolvimento de novos aglomerantes de baixo carbono e a implementação de tecnologias de captura e utilização de carbono (CCUS), é um campo de pesquisa e inovação fundamental para a sustentabilidade da construção civil.

3. ANÁLISE DE RISCO E INCERTEZAS NA DOSAGEM DE CONCRETO DE CIMENTO PORTLAND.

Os projetos estruturais de engenharia civil apresentam diversos desafios na sua elaboração, dentre eles, as incertezas inerentes aos empreendimentos de engenharia. Materiais podem ter propriedades ligeiramente diferentes do esperado, cargas podem variar ao longo do tempo e condições ambientais podem ser imprevisíveis. A análise probabilística é uma ferramenta que nos permite quantificar essas incertezas e avaliar o risco de falha de uma estrutura.

Imagine que a resistência de um pilar se comporta como uma roleta. Mesmo seguindo todos os parâmetros normativos, não é possível determinar exatamente onde a bolinha vai parar, mas podemos estimar a probabilidade de ela cair em um determinado número. Intrínseco a isso, a segurança dos sistemas estruturais de um empreendimento de construção civil depende abalizadamente da quantificação das incertezas que as ações, resistências, solicitações e os modelos numéricos podem apresentar. A partir da década de 70 os primeiros estudos focados nos conceitos de confiabilidade estrutural foram publicados, tais quais iriam atingir a maturidade científica nos anos 90, somente a partir desse conceito foi possível realizar o projeto probabilístico em estruturas. O índice de confiabilidade é como uma medida dessa probabilidade. Quanto maior o índice, menor a chance de a estrutura falhar. Anexo-a isso, foi possível a elaboração de projetos mais econômicos e eficientes, sem perder os parâmetros de funcionalidade e segurança que a norma exige.

Ao contrário das abordagens tradicionais, que assumem valores característicos de resistência, valores nominais dos carregamentos, e de valores determinísticos para os coeficientes de segurança especificados em norma. Em contrapartida, a análise probabilística reconhece que a realidade é mais complexa e que as variáveis envolvidas em um projeto possuem uma certa variabilidade. Essa abordagem nos permite tomar decisões de projeto mais informadas e seguras.

3.1. ANÁLISE DE RISCO

Para a realização de uma análise de risco coerente, é preciso compreender as variáveis do risco de forma abrangente desde a identificação do perigo, avaliação da debilidade, até a caracterização do próprio risco. Neste aspecto, a fim de cumprir os requisitos básicos para a realização de um estudo de análise de risco, é imprescindível considerar a probabilidade de ocorrência de eventos atípicos em um ambiente suscetível. A literatura sobre o tema remonta às primeiras décadas do século XX, com destaque para o trabalho de Knight (1921) intitulado: “Risk, Uncertainty and Profit”, que, segundo ele,

o risco ocorre quando as probabilidades de um evento são conhecidas, enquanto a incerteza se manifesta quando essas probabilidades são desconhecidas ou incalculáveis

3.1.1. CONCEITO E CONCEPÇÕES DE RISCO

O conceito de risco é multifacetado e pode ser abordado sob diferentes perspectivas. Embora existam nuances entre as diversas definições, todas convergem para a ideia de uma ameaça potencial ou de um evento negativo. Ressaltar que, os desafios que as múltiplas vertentes da etimologia da palavra "risco" apresentam retomam o século XIV onde o primeiro registro da palavra 'risco' surge, em castelhano (Riesgo), tendo inicialmente um significado relacionado a desafios e incertezas, sem a conotação clara de perigo potencial (Corominas, 1980).

Sobre outra ótica, estudos etimológicos apontam que a palavra 'risco' tem suas raízes no latim 'resecum', "o que corta", termo utilizado para descrever os perigos ocultos no mar como os enfrentados por navegadores (Gamba e Santos, 2006). Com o tempo, o conceito de risco se ampliou, englobando a ideia de possibilidade de ocorrência de um evento. Na área da engenharia, o risco é quantificado de forma mais precisa, sendo definido como a probabilidade de um determinado evento acontecer, somada à magnitude de suas consequências, sejam elas positivas ou negativas.

Para Lourenço e Amaro (2018), a definição de risco pode ser resumida a um aspecto semântico, considerando que, na língua portuguesa, termos como risco, perigo, ameaça e incerteza são frequentemente utilizados como sinônimos, o que pode gerar certa ambiguidade. No entanto, os autores destacam que cada um desses termos carrega nuances específicas e pode ser mais adequado para determinados contextos. Além disso, a diversidade linguística reflete essa complexidade: enquanto a cultura anglo-saxônica enfatiza o caráter "acaso" ou "casualidade" (hazard), a cultura francesa se concentra na ideia de algo "aleatório" (aléa). Já as línguas espanhola e italiana, com suas raízes comuns, compartilham a ideia de perigo como núcleo do conceito de risco (peligrosidad, pericolosità).

4. DESCONSTRUINDO AS BARREIRAS DA APRENDIZAGEM DE MÁQUINA: AVALIANDO MODELOS DE MACHINE LEARNING PARA PREVER A RESISTÊNCIA DO CONCRETO.

A previsão precisa da resistência à compressão do concreto é um pilar para a segurança e a economia na construção civil. Nas últimas décadas, a aplicação de técnicas de Aprendizagem de Máquina (*Machine Learning* - ML) emergiu como uma alternativa promissora aos métodos de dosagem tradicionais, oferecendo o potencial de modelar as complexas interações entre os componentes do concreto e prever seu desempenho com alta acurácia. No entanto, apesar do sucesso demonstrado em ambientes acadêmicos, a adoção generalizada dessas ferramentas na indústria da construção ainda enfrenta barreiras significativas. A principal delas é a desconfiança em modelos que operam como uma "caixa-preta" (*black box*), onde os processos de decisão são opacos para o engenheiro, gerando incerteza sobre a confiabilidade das previsões. Adicionalmente, a vasta gama de algoritmos disponíveis cria uma barreira de escolha, dificultando a identificação do modelo mais robusto e adequado para a tarefa específica.

Este estudo se propõe a desconstruir sistematicamente essas barreiras. A primeira etapa para transpor a desconfiança é substituir a incerteza pela evidência. Para isso, realizamos uma análise comparativa rigorosa de seis modelos de ML, representando diferentes paradigmas da inteligência artificial. O objetivo não é apenas identificar o modelo com a maior precisão, mas também entender as forças e fraquezas de cada

abordagem. Subsequentemente, discutimos como a aplicação de técnicas de Inteligência Artificial Explicável (XAI) pode iluminar o funcionamento interno do modelo de melhor desempenho, transformando a "caixa-preta" em uma ferramenta transparente e auditável, cujo raciocínio pode ser validado pelo conhecimento de domínio da engenharia.

4.1. O ESPECTRO DE MODELOS PREDITIVOS

Para estabelecer uma base de confiança, é imperativo avaliar de forma sistemática um espectro diversificado de algoritmos. A seleção de modelos para este estudo foi deliberadamente abrangente, incluindo desde abordagens lineares clássicas até complexos métodos de *ensemble*, permitindo uma visão panorâmica do cenário preditivo.

Para conduzir uma análise comparativa robusta, foram selecionados seis modelos de regressão que abrangem diferentes paradigmas fundamentais da aprendizagem de máquina: um modelo linear de base, um modelo baseado em kernel, dois métodos de *ensemble* (um baseado em *bagging* e outro em *boosting*), uma rede neural artificial e um modelo probabilístico bayesiano. A seguir, detalha-se o princípio de funcionamento de cada modelo(model_selection_rationale.md):

- **Regressão Linear**

A Regressão Linear é um modelo estatístico fundamental que busca estabelecer uma relação de linearidade entre uma variável dependente (alvo) e uma ou mais variáveis independentes (features). Devido à sua simplicidade e interpretabilidade, foi utilizada como um modelo de *baseline* para aferir o ganho de desempenho proporcionado por abordagens mais complexas.

O modelo assume que a variável de saída Y pode ser expressa como uma combinação linear das variáveis de entrada X_i , conforme a equação:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Onde β_0 é o intercepto, $\beta_{1..n}$ são os coeficientes que representam o peso de cada feature, e ϵ é o termo de erro. O treinamento consiste em encontrar os coeficientes β que minimizam a soma dos erros quadráticos entre os valores previstos e os observados (James et al., 2013). Sua principal limitação é a premissa de linearidade, que pode não capturar adequadamente as interações complexas entre os componentes do concreto e sua resistência.

- **Support Vector Regression (SVR)**

O *Support Vector Regression* (SVR) é uma adaptação do algoritmo *Support Vector Machine* (SVM) para problemas de regressão. Diferentemente da Regressão Linear, que minimiza o erro para todos os pontos, o SVR busca ajustar um hiperplano de regressão que maximize a quantidade de pontos de dados dentro de uma margem de tolerância, definida por uma função de perda épsilon-insensível (ϵ) (Drucker et al., 1997). Pontos fora dessa margem são penalizados, e os pontos que definem os limites da margem são chamados de vetores de suporte. A principal vantagem do SVR é o uso do “truque do kernel” (*kernel trick*), que permite mapear as features para um espaço de dimensão superior, possibilitando a modelagem de relações não-lineares complexas de forma computacionalmente eficiente.

- **Random Forest Regressor**

O *Random Forest* é um método de *ensemble* que opera através da construção de múltiplas árvores de decisão durante o treinamento e agrega suas previsões para obter um resultado. A premissa fundamental é que a combinação de múltiplos modelos (a “floresta”) produz um regressor mais robusto e preciso do que uma única árvore de decisão, que é suscetível a sobreajuste (*overfitting*) (Breiman, 2001). Seu funcionamento baseia-se em dois princípios de aleatoriedade:

- **Bagging (*Bootstrap Aggregating*):** Cada árvore é treinada com uma amostra aleatória dos dados de treinamento, com reposição.
- **Subespaço de Features Aleatório:** Em cada nó da árvore, a decisão de divisão é tomada considerando apenas um subconjunto aleatório das features disponíveis.

Neste estudo, o modelo foi configurado com 100 árvores. A previsão final é calculada pela média das previsões de todas as árvores individuais da floresta.

- **MLP Regressor (Multi-layer Perceptron)**

O MLP é um tipo de rede neural artificial *feedforward* utilizado para tarefas de regressão. Sua arquitetura é inspirada no cérebro humano e consiste em camadas de neurônios interconectados: uma camada de entrada que recebe as features, uma ou mais camadas ocultas que realizam transformações não-lineares, e uma camada de saída que produz a predição.

Neste estudo, utilizou-se uma arquitetura com *uma camada oculta de 100 neurônios*. A não-linearidade, essencial para aprender as complexas relações nos dados do concreto, foi introduzida pela *função de ativação ReLU (Rectified Linear Unit)*. O ajuste dos pesos sinápticos para minimizar o erro de previsão foi realizado pelo *otimizador Adam*, um algoritmo de otimização estocástica eficiente, com um máximo de *500 iterações* (Goodfellow et al., 2016).

- **Gaussian Process Regressor (GPR)**

O *Gaussian Process Regressor* (GPR) é um modelo não-paramétrico que se destaca por sua abordagem probabilística (Bayesiana). Em vez de produzir uma única estimativa, o GPR modela uma distribuição de probabilidade (especificamente, uma distribuição Gaussiana) sobre todas as funções possíveis que se ajustam aos dados.

Como resultado, cada previsão é acompanhada de uma medida de incerteza (a variância), o que é particularmente valioso em aplicações de engenharia. O comportamento do GPR é definido por um kernel (função de covariância), que mede a similaridade entre os pontos de dados. Para este trabalho, foi utilizada uma combinação de kernels (*RBF, Constant e WhiteKernel*) para modelar a função subjacente, sua escala e o ruído nos dados. O modelo foi otimizado com *9 reinicializações do otimizador* para encontrar os hiperparâmetros ótimos do kernel, aumentando a robustez da solução encontrada (Rasmussen & Williams, 2006).

- **XGBoost Regressor:**

O XGBoost é uma implementação altamente otimizada e escalável do algoritmo de *Gradient Boosting*. Trata-se de um método de *ensemble* que constrói árvores de

decisão de forma sequencial, onde cada nova árvore é treinada para corrigir os erros residuais da sequência de árvores anterior (Chen & Guestrin, 2016).

Diferente do Random Forest, que constrói árvores em paralelo, essa abordagem iterativa permite que o modelo se concentre progressivamente nos exemplos mais difíceis de prever. O termo "eXtreme" refere-se às suas otimizações de sistema, como paralelismo e regularização integrada (L1 e L2) para prevenir o *overfitting*, que o tornaram um dos algoritmos de melhor desempenho para dados tabulares. No treinamento, foi utilizada a função objetivo padrão para regressão, *reg:squarederror*, que instrui o modelo a minimizar o erro quadrático médio.

Todos os modelos foram implementados utilizando as bibliotecas Scikit-learn e XGBoost em Python.

4.2. METODOLOGIA DE VALIDAÇÃO: ASSEGURANDO A ROBUSTEZ PREDITIVA.

A confiança em um modelo preditivo depende diretamente da robustez de sua validação. Para garantir que os resultados desta análise comparativa sejam generalizáveis e não meros artefatos do conjunto de dados específico, foi empregada uma metodologia rigorosa.

A validação cruzada K-Fold, com $K=5$, foi utilizada para avaliar o desempenho de cada modelo. Este procedimento consiste em dividir o *dataset* em cinco subconjuntos (*folds*). Em cada uma das cinco iterações, um *fold* é reservado para teste, enquanto os outros quatro são utilizados para treinamento. Ao final, a performance do modelo é a média dos resultados obtidos nas cinco iterações, garantindo que todos os dados tenham sido utilizados tanto para treino quanto para teste, o que fornece uma estimativa mais estável e confiável do desempenho do modelo em dados não vistos. O desempenho foi quantificado por meio de três métricas padrões para problemas de regressão:

- **Coefficiente de Determinação (R^2):** Mede a proporção da variância na resistência do concreto que é explicada pelos componentes da mistura. Valores mais próximos de 1 indicam um ajuste superior do modelo.
- **Erro Médio Absoluto (MAE):** Representa a média das diferenças absolutas entre os valores previstos e os reais, fornecendo uma medida direta do erro médio em MPa.
- **Raiz do Erro Quadrático Médio (RMSE):** Similar ao MAE, mas penaliza erros maiores com mais intensidade devido à elevação ao quadrado. Também é expresso em MPa.

Para garantir a integridade do processo e evitar o vazamento de informações do conjunto de teste para o de treino, um *Pipeline* do Scikit-learn foi utilizado para encadear o pré-processamento (com *StandardScaler* para normalização das *features*) e o treinamento do modelo em cada *fold* da validação cruzada.

Para **MAE** e **RMSE**, valores menores indicam melhor desempenho. As métricas foram calculadas para cada *fold* da validação cruzada, e os resultados médios foram reportados.

Um *Pipeline* do Scikit-learn foi utilizado para encadear o pré-processamento (*StandardScaler*) e o treinamento do modelo, garantindo que o escalonamento fosse ajustado apenas nos dados de treinamento de cada *fold*, evitando vazamento de informação do conjunto de teste.

5. MACHINE LEARNING

O Machine Learning (ML), ou Aprendizado de Máquina, é um subcampo da inteligência artificial focado no desenvolvimento de algoritmos que permitem aos computadores aprenderem a partir de dados e melhorar seu desempenho em uma determinada tarefa sem serem explicitamente programados para ela (SAMUEL, 1959). Em vez de seguir instruções estáticas, os modelos de ML são treinados em grandes conjuntos de dados para identificar padrões, fazer previsões e tomar decisões.

Na engenharia civil, e especificamente na tecnologia do concreto, a aplicação de ML tem se consolidado como uma ferramenta poderosa para modelar relações complexas e não lineares entre os componentes do material e suas propriedades mecânicas e de durabilidade. Fenômenos como a hidratação do cimento e a interação entre agregados criam um sistema multifatorial cuja previsão de desempenho por meio de modelos puramente analíticos é um desafio significativo (KADDOUR et al., 2023). Nesse contexto, os algoritmos de ML oferecem uma abordagem orientada por dados (data-driven) capaz de prever com alta acurácia propriedades como a resistência à compressão, otimizando traços para atender a critérios de custo, sustentabilidade e desempenho.

5.1. HIPERPARAMETRO EM UM MODELO DE ML

Diferentemente dos parâmetros de um modelo, que são aprendidos diretamente a partir dos dados durante o processo de treinamento (como os pesos em uma regressão linear), os hiperparâmetros são configurações externas definidas pelo desenvolvedor antes do início do treinamento. Eles governam o processo de aprendizado e têm um impacto direto na performance, complexidade e capacidade de generalização do modelo (BERGSTRÄ; BENGIO, 2012). A escolha inadequada de hiperparâmetros pode levar a dois problemas comuns: *overfitting*, onde o modelo se ajusta excessivamente aos dados de treino e perde a capacidade de generalizar para novos dados, ou *underfitting*, onde o modelo é simples demais para capturar os padrões subjacentes nos dados.

O processo de encontrar a combinação ótima de hiperparâmetros é conhecido como "ajuste de hiperparâmetros" ou *hyperparameter tuning*. Existem diversas estratégias para realizar essa tarefa, sendo as mais comuns a busca em grade (*Grid Search*) e a busca aleatória (*Randomized Search*). Enquanto a busca em grade testa exaustivamente todas as combinações possíveis de um conjunto pré-definido de valores, a busca aleatória seleciona um número fixo de combinações a partir de distribuições estatísticas, o que se mostra mais eficiente em espaços de busca de alta dimensão (BERGSTRÄ; BENGIO, 2012).

No presente trabalho, a otimização de hiperparâmetros foi uma etapa crucial para garantir a robustez do modelo preditivo. Adotou-se a estratégia de busca aleatória por meio da função `RandomizedSearchCV` da biblioteca `Scikit-learn`, que, associada à validação cruzada (*K-Fold* com 5 partições), assegura que a performance do modelo seja avaliada de forma consistente em diferentes subconjuntos de dados, evitando vieses. Foram realizadas 150 iterações de combinações de hiperparâmetros para o algoritmo `XGBoost`, explorando um espaço de busca que incluía `n_estimators` (número de árvores), `max_depth` (profundidade máxima das árvores), `learning_rate` (taxa de aprendizado), entre outros, garantindo um modelo final calibrado e com alta capacidade preditiva.

5.2. TIPOS DE ALGORITMO DE MACHINE LEARNING

Os algoritmos de Machine Learning são categorizados principalmente com base no tipo de aprendizado. A modalidade mais comum na engenharia de materiais é o **aprendizado supervisionado**, no qual o algoritmo é treinado com um conjunto de dados rotulado, contendo tanto as variáveis de entrada (features) quanto a variável de saída desejada (target). O objetivo é aprender uma função que mapeie as entradas às saídas. O aprendizado supervisionado se divide em duas categorias principais de tarefas:

1. **Regressão:** A variável de saída é um valor contínuo. Exemplos incluem a previsão da resistência à compressão do concreto (em MPa), o consumo de cimento ou o custo do traço.
2. **Classificação:** A variável de saída é uma categoria discreta, como classificar um concreto como de "alta resistência" ou "resistência convencional".

Dentro da regressão, destacam-se os algoritmos de *ensemble*, que combinam múltiplos modelos mais fracos para criar um modelo final mais robusto e preciso. Uma das técnicas de *ensemble* mais eficazes é o *Gradient Boosting*, que constrói modelos de forma sequencial, onde cada novo modelo tenta corrigir os erros do anterior.

O **XGBoost (Extreme Gradient Boosting)** é uma implementação otimizada e escalável do *Gradient Boosting*, desenvolvida por Chen e Guestrin (2016). Ele se tornou o algoritmo de escolha para muitas competições e aplicações industriais devido à sua alta performance, velocidade de treinamento e controle de *overfitting* por meio de técnicas de regularização.

Neste estudo, o problema de prever a resistência do concreto foi enquadrado como uma tarefa de regressão. Optou-se pelo uso do **XGBoost na sua modalidade de Regressão Quantílica**. Diferentemente da regressão tradicional, que prevê apenas o valor médio (ou mediano), a regressão quantílica permite estimar diferentes quantis da distribuição da variável de saída. No contexto deste trabalho, foram treinados três modelos distintos para prever o 5º percentil (limite inferior), o 50º percentil (mediana, a previsão mais provável) e o 95º percentil (limite superior). Essa abordagem é particularmente valiosa, pois fornece não apenas uma previsão pontual, mas um intervalo de confiança para a resistência do concreto, quantificando a incerteza associada às previsões do modelo, o que é de extrema importância para a tomada de decisão em engenharia.

5.3. MACHINE LEARNING APLICADO A CONCRETOS

A aplicação de Machine Learning para modelar as propriedades do concreto tem sido extensivamente explorada na literatura científica, com o objetivo de superar as limitações dos modelos empíricos tradicionais. Estudos demonstram a eficácia de algoritmos como Redes Neurais Artificiais (ANNs), Máquinas de Vetores de Suporte (SVMs) e, mais recentemente, algoritmos baseados em árvores de decisão como o XGBoost, para prever a resistência à compressão com base na composição do traço (CHOPRA et al., 2016; AHMAD et al., 2021).

Uma etapa fundamental para o sucesso desses modelos é a **engenharia de características (feature engineering)**, que consiste em criar variáveis de entrada a partir das existentes, utilizando o conhecimento de domínio para fornecer informações mais relevantes ao algoritmo. No caso do concreto, relações conhecidas como a razão água/cimento são determinantes para a resistência.

Alinhado a essa prática, este trabalho implementou uma etapa de engenharia de características, criando variáveis como "ligante" (soma de cimento e adições como sílica

ativa e cinza volante) e "relacao_água_ligante". Essa transformação permite que o modelo aprenda diretamente a partir de relações já consagradas pela ciência dos materiais, melhorando sua acurácia e capacidade de generalização.

Além da previsão de propriedades, o ML tem sido acoplado a algoritmos de otimização para encontrar traços de concreto que atendam a múltiplos critérios conflitantes, como maximizar a resistência, minimizar o custo e reduzir o impacto ambiental (emissão de CO₂). Para isso, são utilizados algoritmos de **otimização multiobjetivo**, como o NSGA-II (*Non-dominated Sorting Genetic Algorithm II*), proposto por Deb et al. (2002). Esse algoritmo evolutivo é capaz de encontrar um conjunto de soluções ótimas (conhecido como fronteira de Pareto), que representam os melhores compromissos possíveis entre os objetivos conflitantes.

A metodologia deste estudo integrou o modelo preditivo de XGBoost com o otimizador NSGA-II. Após o treinamento, o modelo de ML atuou como uma "função de avaliação" rápida e precisa, estimando a resistência, o custo e a emissão de CO₂ para milhares de traços virtuais gerados pelo algoritmo genético. O resultado é uma fronteira de Pareto que oferece ao engenheiro um portfólio de traços otimizados, permitindo uma escolha informada que equilibra desempenho técnico, viabilidade econômica e sustentabilidade.

5.4. PRINCIPAIS DESAFIOS DA ANÁLISE DE DADOS COM MACHINE LEARNING

Apesar de seu grande potencial, a aplicação de Machine Learning enfrenta desafios práticos que devem ser cuidadosamente gerenciados para garantir resultados confiáveis. Kelleher et al. (2020) destacam que a qualidade do modelo final é diretamente dependente da qualidade dos dados de entrada. Os principais desafios incluem:

- **Qualidade e Quantidade de Dados:** Modelos de ML requerem grandes volumes de dados de alta qualidade. Dados escassos, ruidosos, ou com erros de medição podem levar a modelos com baixo desempenho e conclusões equivocadas. Na engenharia civil, a obtenção de dados de ensaios laboratoriais pode ser cara e demorada.
- **Pré-processamento e Limpeza de Dados:** Dados brutos raramente estão prontos para serem utilizados. Eles podem conter valores ausentes, inconsistências (como nomes de colunas diferentes) e *outliers* (valores atípicos que se desviam significativamente do padrão). A falha em tratar adequadamente esses problemas pode enviesar o treinamento do modelo. No presente estudo, foi implementada uma rotina de limpeza que padronizou os nomes das colunas e um método de detecção de *outliers* baseado na análise dos resíduos de um modelo preliminar, removendo pontos que poderiam comprometer a integridade do treinamento final.
- **Interpretabilidade do Modelo ("*Black Box*" Problem):** Modelos complexos como XGBoost e redes neurais, embora muito precisos, são muitas vezes considerados "caixas-pretas", pois seu processo de decisão interna não é facilmente compreendido por humanos. Isso pode ser uma barreira para a aceitação em áreas críticas como a engenharia, onde a compreensão do "porquê" por trás de uma decisão é fundamental.
- **Validação e Generalização:** É crucial garantir que o modelo não apenas memorize os dados de treinamento, mas que seja capaz de generalizar para novos dados não vistos. Conforme discutido, técnicas como a validação cruzada são essenciais para avaliar a verdadeira capacidade de generalização do modelo.

- **Incerteza do Modelo:** Modelos de regressão padrão fornecem uma previsão pontual sem indicar o nível de confiança associado. Como abordado neste trabalho através da regressão quantílica, quantificar a incerteza é um passo importante para tornar as previsões de ML mais úteis e seguras para aplicações práticas na engenharia. Adicionalmente, foi implementada uma etapa de correção para o cruzamento de quantis, um problema comum em que o limite inferior previsto pode exceder o superior, garantindo a consistência lógica dos intervalos de confiança gerados.

6. METODOLOGIA

6.1. ETAPA 01- SELEÇÃO E PRÉ-PROCESSAMENTO DO BANCO DE DADOS.

Para o desenvolvimento deste estudo, utilizou-se o renomado banco de dados de Resistência à Compressão do Concreto (*Concrete Compressive Strength*), disponível no repositório da Universidade da Califórnia, Irvine (UCI Machine Learning Repository) sob o ID 165. Este dataset, amplamente validado na literatura científica (YEH, 1998; YOUNG et al., 2019), é composto por 1030 ensaios de compressão e abrange oito variáveis preditoras quantitativas e uma variável alvo:

- Cimento (kg/m^3)
- Escória de Alto Forno (kg/m^3)
- Cinza Volante (kg/m^3)
- Água (kg/m^3)
- Superplastificante (kg/m^3)
- Agregado Grosso (kg/m^3)
- Agregado Fino (kg/m^3)
- Idade do Concreto (dias)
- Resistência à Compressão (MPa) - variável alvo

A aquisição dos dados foi automatizada via um script em linguagem Python, utilizando a função `fetch_ucirepo` da biblioteca `ucimlrepo`. Após o carregamento, a primeira etapa do pré-processamento consistiu na renomeação das colunas para termos em português, a fim de facilitar a manipulação e a interpretação subsequente.

Visando aprimorar a qualidade do conjunto de treinamento e a robustez dos modelos, foi implementada uma etapa criteriosa de detecção e remoção de *outliers*. Para tal, um modelo `RandomForestRegressor` foi treinado temporariamente para estimar as predições. Instâncias cujo valor absoluto do resíduo (diferença entre o valor real e o previsto) excedeu um limiar empírico de 2.43 MPa foram consideradas anômalas e removidas. Este procedimento é fundamental para mitigar ruídos e vieses, assegurando que o treinamento seja realizado em um conjunto de dados mais coeso e representativo.

- **DADOS REFERENTES AOS PREÇOS DOS INSUMOS**

A apuração dos custos dos componentes do concreto foi realizada por meio de uma metodologia que prioriza fontes de dados oficiais e de alta credibilidade, garantindo a rastreabilidade e a precisão necessárias para a pesquisa acadêmica. Para os insumos padronizados, como cimento e agregados (miúdo e graúdo), a principal referência utilizada foi o Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil (SINAPI), gerenciado pela Caixa Econômica Federal (CAIXA) em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE). A confiabilidade desta fonte é atestada pelo Decreto nº 7.983/2013, que a estabelece como referência obrigatória para orçamentos de obras contratadas com recursos da União (BRASIL, 2013).

Adicionalmente, o custo da água foi obtido diretamente da tabela tarifária oficial da categoria industrial, publicada pela concessionária local, a Companhia de Saneamento do Pará (COSANPA, 2019).

Para os materiais que não constam nas tabelas do SINAPI, como subprodutos industriais e aditivos químicos, foram adotadas abordagens metodológicas específicas. O custo de adições cimentícias, como a escória de alto-forno (30% do custo do cimento) e as cinzas volantes (25% do custo do cimento), foi estimado como uma fração do custo do cimento Portland, uma prática alinhada à literatura acadêmica que analisa a viabilidade econômica da substituição de clínquer (LUGARESI, 2023). No caso do aditivo superplastificante, um produto de marca, o valor foi determinado por meio de pesquisa de mercado em fornecedores e o seu custo efetivo por quilograma foi calculado a partir de dados da ficha técnica do produto, que informa sua densidade (SIKA, [s.d.]).

A adoção deste protocolo de pesquisa multifacetado assegura um elevado grau de confiabilidade e reprodutibilidade dos dados de entrada. Ao substituir estimativas genéricas por valores extraídos de publicações governamentais, tarifas oficiais e metodologias documentadas, a análise de custos se fundamenta em uma base de dados verificável e geograficamente específica para o estado do Pará. Essa abordagem rigorosa confere validade às análises econômicas subsequentes, distanciando-se da imprecisão de buscas informais e alinhando-se às exigências do método científico.

• DADOS REFERENTES AOS PREÇOS DOS INSUMOS

A quantificação dos fatores de emissão de dióxido de carbono para os insumos da construção civil foi baseada no Inventário de Carbono e Energia (ICE), versão 3.0, uma base de dados de referência internacional. Desenvolvida por pesquisadores da Universidade de Bath, no Reino Unido, a base de dados ICE é o resultado de uma extensa meta-análise da literatura científica, compilando e avaliando dados de centenas de fontes secundárias, como artigos acadêmicos, relatórios setoriais e outros estudos de Avaliação do Ciclo de Vida (ACV) (HAMMOND; JONES, 2019). O objetivo desta compilação é fornecer um inventário abrangente e de livre acesso com coeficientes de carbono incorporado para uma vasta gama de materiais de construção, permitindo a realização de análises ambientais de forma padronizada.

A metodologia empregada na base de dados ICE adota, predominantemente, uma fronteira de sistema "do berço ao portão" (*cradle-to-gate*). Isso significa que os valores de emissão englobam todos os processos desde a extração da matéria-prima, passando pelo transporte e processamento, até o momento em que o produto deixa o portão da fábrica (HAMMOND; JONES, 2019). Os fatores são expressos em quilogramas de dióxido de carbono equivalente (kg CO₂e), uma métrica que considera não apenas o CO₂, mas também outros gases de efeito estufa, convertidos com base em seu potencial de aquecimento global. Essa abordagem sistemática garante que os dados sejam comparáveis entre diferentes materiais dentro do escopo definido.

A base de dados ICE é uma fonte de alta confiabilidade e amplamente reconhecida no meio acadêmico e profissional, sendo utilizada em milhares de projetos e publicações ao redor do mundo. Sua credibilidade é reforçada pela transparência metodológica, que inclui a referência às fontes originais para cada dado e a aplicação de uma matriz de qualidade para avaliar a robustez das informações. Embora os dados representem, em muitos casos, médias da indústria europeia ou global, a sua abrangência e o rigor na compilação a estabelecem como uma ferramenta fundamental e fidedigna para a estimativa da pegada de carbono na construção civil, especialmente em estudos que

buscam um benchmark internacional ou na ausência de dados primários locais verificados.

6.2. ETAPA 02- ANÁLISE MULTIMODAL E SELEÇÃO DO MODELO DE MACHINE LEARNING.

Na segunda etapa, foi realizada uma análise multimodal para identificar o modelo de aprendizado de máquina mais adequado para a previsão da resistência do concreto. Esta fase envolveu a experimentação com diferentes algoritmos de Machine Learning, conforme implementado no script `ANALISE_MULTIMODAL_DO_CONCRETO.py`. O objetivo foi comparar o desempenho de diversos modelos preditivos em relação à acurácia, robustez e capacidade de generalização, utilizando métricas de avaliação apropriadas. A escolha do modelo mais promissor foi baseada em critérios que consideraram não apenas a performance preditiva, mas também a interpretabilidade e a aplicabilidade no contexto da engenharia civil. A Figura 01 apresenta um fluxograma resumido para escolha do modelo de machine learning a ser utilizado na otimização da dosagem de concreto.

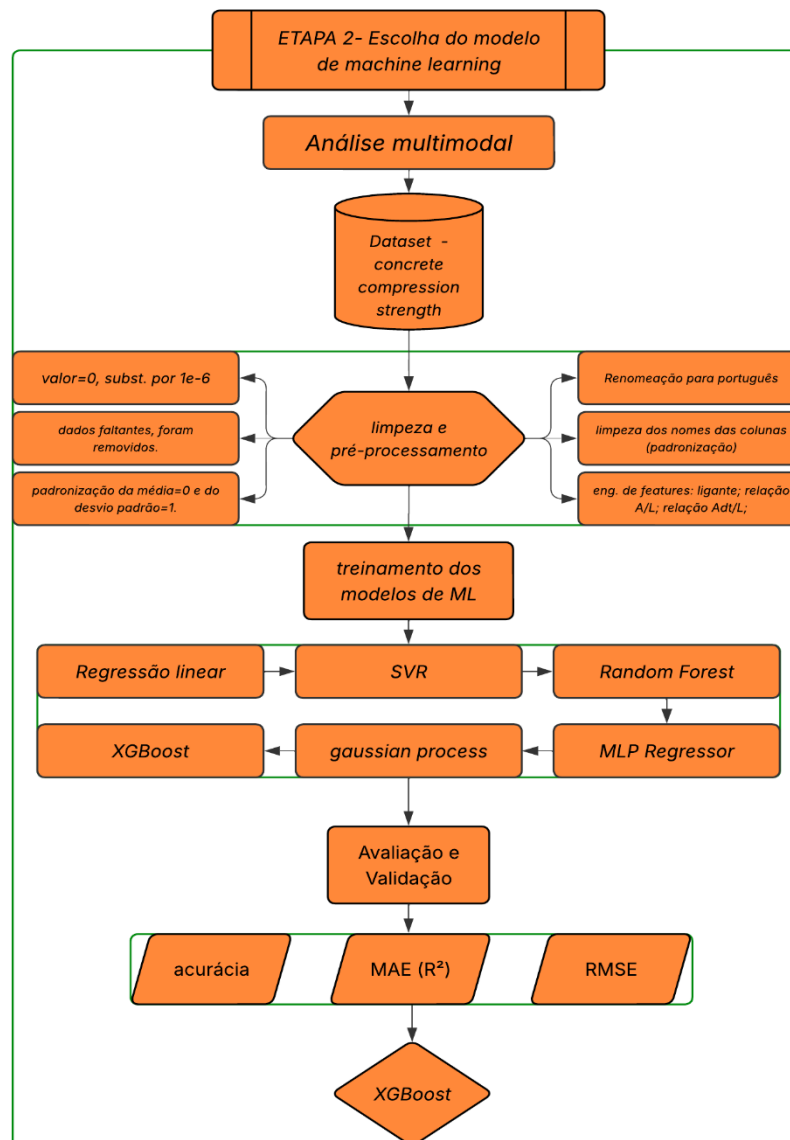


Figura 1 - Fluxograma da 2ª etapa

Esta etapa do estudo realiza uma análise multimodal comparativa de seis modelos de aprendizado de máquina (**Regressão Linear, Support Vector Regression - SVR, Random Forest, Multi-layer Perceptron - MLP, Gaussian Process Regression - GPR e Extreme Gradient Boosting - XGBoost**) aplicados ao dataset de Resistência à Compressão do Concreto do repositório UCI. O objetivo foi identificar o modelo com melhor desempenho preditivo para este problema específico, utilizando validação cruzada (**K-Fold com 5 folds**) e métricas padrões de regressão (**R^2 , MAE, RMSE**). *Os resultados indicaram que modelos baseados em árvores*, particularmente **XGBoost ($R^2=0.930$, RMSE=4.391 MPa)** e **Random Forest ($R^2=0.911$, RMSE=4.981 MPa)**, apresentaram o melhor desempenho, **superando significativamente modelos lineares e SVR**. **Gaussian Process** também mostrou forte desempenho (**$R^2=0.885$, RMSE=5.645 MPa**), enquanto **MLP** obteve resultados intermediários (**$R^2=0.818$, RMSE=7.098 MPa**). *A análise demonstra a superioridade de modelos não-lineares complexos para capturar as relações intrincadas entre os componentes do concreto e sua resistência*, fornecendo uma base sólida para a seleção de modelos em aplicações práticas e futuras otimizações de misturas de concreto.

6.2.1. RESULTADOS E DISCUSSÃO

Para obter uma estimativa robusta do desempenho de generalização de cada modelo, foi empregada a avaliação cruzada com $k=5$ folds. O dataset foi dividido em 5 partes (folds); a cada iteração, 4 folds foram usados para treinamento e 1 fold para teste. O processo foi repetido 5 vezes, garantindo que cada fold fosse usado como conjunto de teste uma vez.

O desempenho foi avaliado utilizando três métricas padrões para problemas de regressão:

- **Coefficiente de Determinação (R^2):** Indica a proporção da variância na variável dependente que é previsível a partir das variáveis independentes. Valores mais próximos de 1 indicam melhor ajuste.
- **Erro Médio Absoluto (MAE):** Média das diferenças absolutas entre os valores preditos e reais. Fornece o erro médio na unidade da variável target (MPa).
- **Raiz do Erro Quadrático Médio (RMSE):** Raiz quadrada da média dos erros quadráticos. Penaliza erros maiores mais fortemente que o MAE e está na unidade da variável target (MPa).

Para MAE e RMSE, valores menores indicam melhor desempenho. As métricas foram calculadas para cada fold da validação cruzada, e os resultados médios foram reportados.

Um pipeline do scikit-learn foi utilizado para encadear o pré-processamento (StandardScaler) e o treinamento do modelo, garantindo que o escalonamento fosse ajustado apenas nos dados de treinamento de cada fold, evitando vazamento de informação do conjunto de teste.

A execução da análise multimodal foi executada conforme descrito anteriormente e produziu resultados claros, consolidados na Tabela 1, que apresenta os resultados médios das métricas de desempenho (R^2 , MAE, RMSE) e o tempo médio de execução para cada um dos seis modelos avaliados, obtidos através da validação cruzada de 5 folds.

Tabela 2 - Resultados Consolidados da Análise Multimodal (Média da Validação Cruzada K=5).

Modelo	R ² médio (Acurácia)	MAE Médio (MPa)	RMSE Médio (MPa)	Tempo Médio (s)
XGBoost	0.930	2.824	4.391	0.31
Random Forest	0.911	3.419	4.981	1.14
Gaussian Process	0.885	4.087	5.645	84.62
MLP Regressor	0.818	5.348	7.098	2.34
SVR	0.652	7.716	9.833	0.13
Regressão Linear	0.604	8.278	10.460	1.69

Nota: Os resultados foram ordenados pelo R² médio em ordem decrescente.

Os resultados demonstram inequivocamente a superioridade dos modelos de *ensemble* baseados em árvores. O **XGBoost Regressor** se destacou como o modelo de melhor desempenho, alcançando o maior R² (0.930) e os menores erros (MAE de 2.824 MPa e RMSE de 4.391 MPa), indicando a mais alta capacidade de generalização. O *Random Forest* também apresentou um desempenho excelente, validando a eficácia de abordagens de *ensemble* para este problema.

O *Gaussian Process Regressor* mostrou-se uma alternativa robusta, embora seu custo computacional elevado possa ser um impeditivo prático. Por outro lado, o desempenho modesto da Regressão Linear (R² de 0.604) confirma sua incapacidade de capturar as relações não-lineares intrínsecas à ciência dos materiais do concreto.

Estes resultados, por si só, já começam a desconstruir uma das barreiras: a incerteza sobre qual ferramenta utilizar. A evidência aponta fortemente para algoritmos de *boosting* como o XGBoost. Contudo, a alta performance desses modelos está atrelada a uma maior complexidade, o que nos leva diretamente à barreira da "caixa-preta". Um engenheiro pode hesitar em confiar em uma previsão, por mais precisa que seja, se não puder compreender sua lógica subjacente.

É neste ponto que a **Inteligência Artificial Explicável (XAI)** se torna não apenas útil, mas essencial. Técnicas como o SHAP (*SHapley Additive exPlanations*) permitem "abrir" a caixa-preta e quantificar a contribuição de cada variável de entrada (cimento, água, agregados, idade etc.) para uma previsão específica. Ao aplicar XAI ao modelo XGBoost treinado, é possível verificar se o modelo aprendeu relações que são consistentes com os princípios fundamentais da tecnologia do concreto — por exemplo, a importância dominante da relação água/cimento.

Portanto, a análise preditiva aqui apresentada é o primeiro passo fundamental. Ela identifica a ferramenta mais poderosa. O passo seguinte, e crucial para a adoção na prática, é usar a XAI para traduzir o funcionamento dessa ferramenta em insights compreensíveis e confiáveis para o engenheiro, alinhando o poder preditivo da máquina com o conhecimento e a experiência humana. Esta abordagem híbrida, que une predição rigorosa e explicação transparente, oferece um caminho claro para superar as barreiras e integrar de forma segura e eficaz a aprendizagem de máquina na engenharia de concreto.

6.3. ETAPA 03 - OTIMIZAÇÃO DE MISTURAS DE CONCRETO COM IA

O script anterior foi uma ferramenta de decisão para a construção desse script, ele foi fundamental para a tomada de decisão da escolha do modelo de machine learning que usamos no modelo principal de aprendizagem deste estudo.

O script *CONCRETE_AI.PY* é uma ferramenta robusta desenvolvida para otimizar misturas de concreto que foi construída a partir de diversas etapas de tratamento e processamento de dados, até chegar em um modelo final bem ajustado, utilizando técnicas avançadas de aprendizado de máquina e otimização multiobjetivo. Seu principal objetivo é prever a resistência à compressão do concreto e, em seguida, otimizar a composição da mistura para atender a uma resistência alvo, considerando também custos e emissões de CO2. Este tópico detalha a estrutura, as funcionalidades e a lógica por trás de cada componente do script. A fim de melhorar o entendimento do fluxo de treinamento, abaixo destaca-se um fluxograma que ilustra esse treinamento.

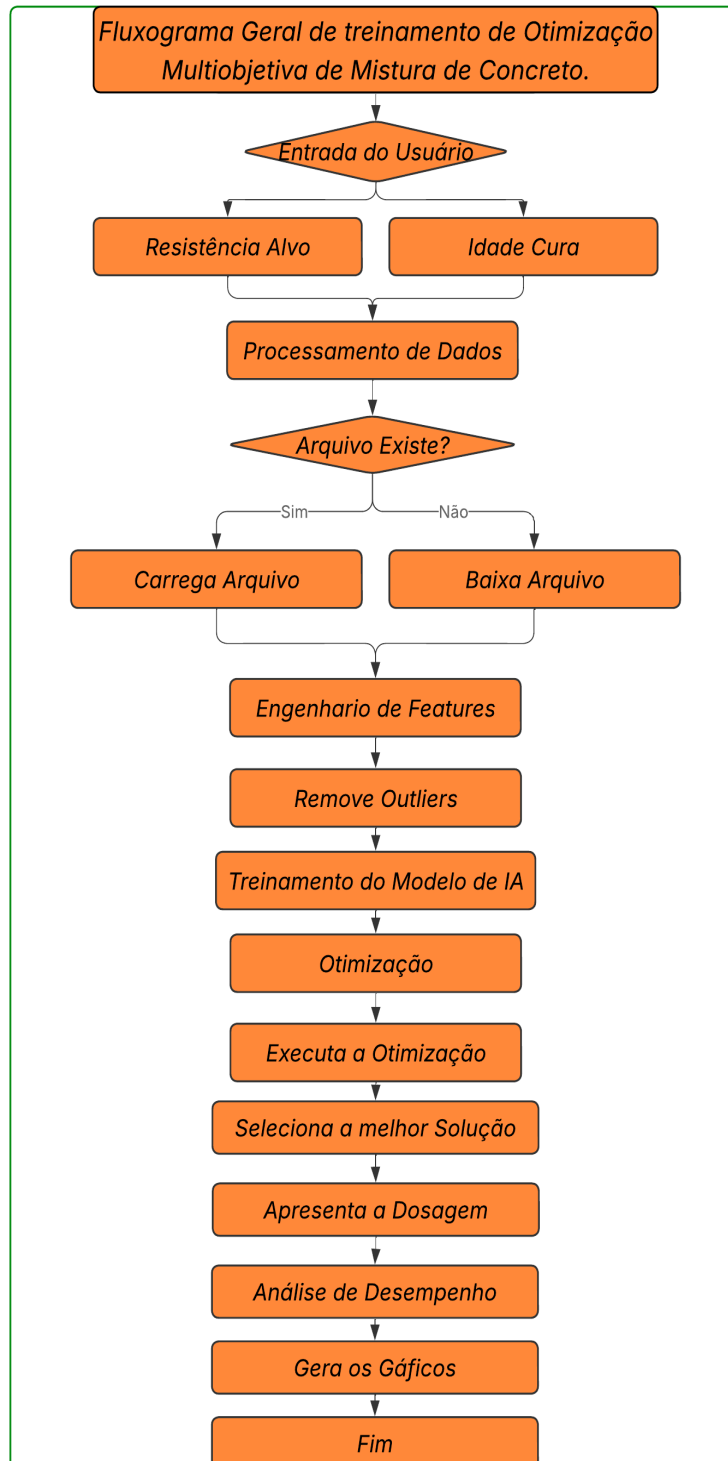


Figura 2 - Fluxograma Geral de treinamento de Otimização Multiobjetiva de Mistura de Concreto.

O concreto, sendo o material de construção mais utilizado globalmente, tem sua qualidade e desempenho intrinsecamente ligados à sua resistência à compressão. A determinação e otimização dessa resistência são cruciais para a segurança e a sustentabilidade na engenharia civil. O `CONCRETE_AI.py` aborda essa complexidade através de uma abordagem baseada em dados, aproveitando o poder do aprendizado de máquina para modelar as relações não-lineares entre os componentes do concreto e sua resistência.

O script integra várias etapas essenciais de um projeto de ciência de dados e engenharia, incluindo:

- **Carregamento e Pré-processamento de Dados:** Aquisição de dados de um repositório público (UCI) e preparação para análise.
- **Engenharia de Features:** Criação de novas variáveis que capturam relações importantes entre os componentes do concreto.
- **Treinamento de Modelos de Machine Learning:** Utilização de modelos XGBoost Quantile para prever a resistência à compressão e suas faixas de incerteza.
- **Otimização Multiobjetivo:** Emprego do algoritmo NSGA-II para encontrar misturas de concreto que equilibrem resistência, custo e emissões.
- **Visualização de Dados:** Geração de gráficos para entender o comportamento do dataset e a evolução da resistência.

Este tópico servirá como um guia para entender como cada uma dessas partes funciona e se interliga para formar uma solução completa para a otimização de misturas de concreto.

6.3.1. Estrutura do Script e Bibliotecas Utilizadas

O script `CONCRETE_AI.py` é modular, organizado em funções e classes que encapsulam diferentes aspectos do processo de otimização. As principais bibliotecas Python utilizadas são:

- **pandas** e **numpy**: Para manipulação e operações numéricas com dados.
- **joblib** e **os**: Para persistência de modelos e operações de sistema de arquivos.
- **warnings**: Para gerenciar avisos durante a execução.
- **ucimlrepo**: Para buscar o dataset de Resistência à Compressão do Concreto diretamente do repositório UCI.
- **sklearn (Scikit-learn)**: Um pilar para o aprendizado de máquina, fornecendo ferramentas para pré-processamento (`StandardScaler`, `ColumnTransformer`), modelos (`RandomForestRegressor`, `Pipeline`) e validação de modelos (`KFold`, `RandomizedSearchCV`).
- **XGBoost**: Uma biblioteca otimizada para algoritmos de gradient boosting, utilizada aqui para treinamento de modelos de regressão por quantil.
- **pymoo**: Uma estrutura de código aberto para otimização multiobjetivo, especificamente para a implementação do algoritmo NSGA-II.
- **matplotlib.pyplot** e **seaborn**: Para a criação de visualizações de dados de alta qualidade.

Além das bibliotecas, o script define constantes importantes para o cálculo de custos e emissões de CO2 associadas aos componentes do concreto (*COMPONENT_COSTS* e *CO2_FACTORS*). Essas constantes são fundamentais para a função de otimização, permitindo que o modelo considere não apenas a resistência, mas também os impactos econômicos e ambientais da mistura.

6.3.2. Pré-processamento de Dados

O pré-processamento de dados é uma etapa crucial para garantir a qualidade e a consistência dos dados de entrada para os modelos de aprendizado de máquina. O script *CONCRETE_AI.py* implementa as seguintes funções e métodos para essa finalidade:

6.3.2.1. Limpeza e Padronização das Colunas

Esta função utilitária é responsável por padronizar os nomes das colunas de um DataFrame. Ela realiza as seguintes operações:

- Remove espaços em branco no início e no final dos nomes das colunas (*str.strip()*).
- Converte todos os caracteres para minúsculas (*str.lower()*).
- Substitui caracteres não alfanuméricos (exceto números) por underscores (`_`) usando uma expressão regular (*str.replace*(`[/^[a-z0-9]/, '_ ', regex=True)`).
- Substitui múltiplos underscores consecutivos por um único underscore (*str.replace*(`['_ +]', '_ ', regex=True)`).

Essa padronização é essencial para evitar erros de digitação e garantir que as colunas sejam referenciadas de forma consistente em todo o script.

6.3.2.2. Carregamento do banco de dados e pré-processamento inicial

Este método, parte da classe *ConcreteOptimizer*, é responsável por carregar o dataset e aplicar um pré-processamento inicial. Ele verifica se um arquivo CSV local (*concrete_data.csv*) já existe. Se sim, carrega os dados desse arquivo. Caso contrário, ele busca o dataset de Resistência à Compressão do Concreto (ID: 165) diretamente do repositório UCI Machine Learning usando a biblioteca *ucimlrepo*. Após o carregamento, as colunas são renomeadas para termos em português, conforme um mapeamento predefinido, tornando o dataset mais intuitivo para usuários de língua portuguesa. Os dados são então salvos localmente para futuras execuções, evitando a necessidade de baixar o dataset repetidamente.

Exemplo do mapeamento de colunas:

```
Python
mapping = {
    "cement": "cimento",
    "blast_furnace_slag": "escoria",
    "fly_ash": "cinzas",
    "water": "agua",
    "superplasticizer": "aditivo",
    "coarse aggregate": "agregado grosso",
    "fine aggregate": "agregado miudo",
    "age": "idade cura",
    "concrete compressive strength": "resistencia mpa"
}
```

Este método garante que o dataset esteja sempre disponível e com um formato padronizado para as etapas subsequentes de engenharia de features e treinamento de modelos.

6.3.3. Engenharia de Features (Atributos)

A engenharia de features é o processo de usar o conhecimento de domínio para criar variáveis a partir dos dados existentes, com o objetivo de melhorar o desempenho dos modelos de aprendizado de máquina. No script `CONCRETE_AI.py`, essa etapa é realizada pelo método `ConcreteOptimizer.engineer_features(df)`.

Este método calcula e adiciona cinco novas features ao DataFrame, que são cruciais para capturar relações mais complexas e significativas entre os componentes do concreto e sua resistência. As features criadas são:

ligante: Representa a massa total de materiais ligantes na mistura. É calculada como a soma das massas de cimento, escória de alto-forno e cinzas volantes. Esta feature é importante porque a quantidade total de ligantes afeta diretamente a formação da pasta de cimento e, conseqüentemente, a resistência do concreto.

```
Python
df["ligante"] = df["cimento"] + df["escoria"] + df["cinzas"]
```

relacao_água_ligante: É a razão entre a massa de água e a massa total de ligantes. Esta é uma das relações mais críticas na tecnologia do concreto, pois uma menor relação água/ligante geralmente resulta em maior resistência, desde que a trabalhabilidade seja mantida. Para evitar divisão por zero, um pequeno valor `epsilon` (1e-6) é adicionado ao denominador.

```
Python
df["relacao_agua_ligante"] = df["agua"] / df["ligante"].replace(0, epsilon)
```

relacao_aditivo_ligante: Representa a razão entre a massa de superplastificante (aditivo) e a massa total de ligantes. Aditivos superplastificantes são usados para reduzir a quantidade de água necessária para uma dada trabalhabilidade, o que pode levar a um aumento da resistência. Novamente, `epsilon` é usado para evitar divisão por zero.

```
Python
df["relacao_aditivo_ligante"] = df["aditivo"] / df["ligante"].replace(0, epsilon)
```

relacao_agregado_fino_grosso: É a razão entre a massa de agregado miúdo (areia) e a massa de agregado graúdo (brita). A proporção entre agregados finos e grossos afeta a compactidade da mistura e a trabalhabilidade, influenciando indiretamente a resistência. *epsilon* é usado para evitar divisão por zero.

```
Python
df["relacao_agregado_fino_grosso"] = df["agregado_miudo"].replace(0, epsilon) /
df["agregado_grosso"].replace(0, epsilon)
```

proporcao_pasta: Calcula a proporção da pasta de cimento (água + ligante) em relação à massa total de todos os componentes da mistura (pasta + agregados). A quantidade de pasta é fundamental para preencher os vazios entre os agregados e

envolver as partículas, afetando a coesão e a resistência do concreto. `epsilon` é usado para evitar divisão por zero.

Python

```
pasta = df["agua"] + df["ligante"]
total_solidos = pasta + df["agregado_miudo"] + df["agregado_grosso"]
df["proporcao_pasta"] = pasta / total_solidos.replace(0, epsilon)
```

Após a criação dessas novas features, o método também lida com quaisquer valores infinitos (`np.inf`, `-np.inf`) ou `NaN` que possam ter surgido devido às operações de divisão ou outros fatores. Esses valores são substituídos pela mediana da respectiva coluna (`df.fillna(df.median(), inplace=True)`). Este tratamento garante que o dataset esteja limpo e pronto para o treinamento dos modelos, sem valores ausentes ou problemáticos que possam afetar a performance do aprendizado de máquina.

A engenharia de features é um passo vital, pois permite que os modelos capturem a complexidade do problema de forma mais eficaz, transformando dados brutos em informações mais significativas e preditivas. As novas features adicionam uma camada de conhecimento de engenharia ao dataset, o que é frequentemente mais poderoso do que depender apenas das features originais.

6.3.4. Treinamento de Modelos: XGBoost Quantile

O método `ConcreteOptimizer.train_model()` é o coração do script no que diz respeito ao aprendizado de máquina. Ele é responsável por treinar modelos de regressão por quantil utilizando o algoritmo XGBoost. A abordagem de regressão por quantil permite prever não apenas a mediana da resistência à compressão, mas também faixas de incerteza (quantis inferior e superior), o que é extremamente valioso em aplicações de engenharia onde a confiabilidade das previsões é crítica.

6.3.4.1. Detecção e Remoção de Outliers

Antes do treinamento principal dos modelos de quantil, o script realiza uma etapa de detecção e remoção de outliers. Esta é uma prática importante para melhorar a robustez e o desempenho dos modelos, pois outliers podem distorcer o processo de aprendizado. A metodologia empregada é a seguinte:

- **Pré-processamento Temporário:** Um `ColumnTransformer` com `StandardScaler` é aplicado às features para padronizá-las. Isso é feito dentro de um `Pipeline` temporário para garantir que o escalonamento seja consistente.
- **Modelo Temporário:** Um `RandomForestRegressor` simples (com 50 estimadores) é treinado nos dados pré-processados. Este modelo é usado para obter uma estimativa inicial das previsões.
- **Cálculo de Resíduos:** Os resíduos são calculados como a diferença entre os valores reais de resistência (\hat{y}) e as previsões do modelo temporário.
- **Identificação de Outliers:** Instâncias onde o valor absoluto do resíduo excede um `error_threshold` de 2.43 MPa são identificadas como outliers. Este limite é um valor empírico que pode ser ajustado.

- **Remoção de Outliers:** As linhas correspondentes aos outliers são removidas do DataFrame original, resultando em um `df_cleaned` que será usado para o treinamento dos modelos de quantil.

Esta etapa de limpeza de dados é crucial para garantir que os modelos sejam treinados em um conjunto de dados mais representativo e menos ruidoso, levando a previsões mais precisas e confiáveis.

6.3.4.2. Treinamento do Modelo da Mediana (q=0.5) com Otimização de Hiperparâmetros

A primeira e mais crítica etapa do treinamento é a otimização e o treinamento do modelo XGBoost para o quantil da mediana (q=0.5). Isso é feito utilizando `RandomizedSearchCV` para encontrar os melhores hiperparâmetros para o `XGBRegressor`.

- **Pipeline de Busca:** Um `Pipeline` é construído, encadeando o `StandardScaler` (para pré-processamento das features) e o `'XGBRegressor'` (configurado para `objective='reg:quantileerror'` e `quantile_alpha=0.5`).
- **Espaço de Busca de Hiperparâmetros (`param_dist`):** Um dicionário define as distribuições de possíveis valores para os hiperparâmetros do XGBoost, como `n_estimators`, `learning_rate`, `max_depth`, `subsample`, `colsample_bytree` e `gamma`. A `RandomizedSearchCV` explorará combinações desses parâmetros.
- **Validação Cruzada:** `KFold` com 10 splits e `shuffle=True` é usado para a validação cruzada, garantindo uma avaliação robusta do desempenho do modelo.
- **Métrica de Avaliação:** O `scoring` é definido como `neg_root_mean_squared_error`, que é o negativo do RMSE. A `RandomizedSearchCV` busca maximizar essa métrica (ou seja, minimizar o RMSE).
- **Execução da Busca:** A `RandomizedSearchCV` é executada com 150 iterações (`n_iter=150`), buscando a melhor combinação de hiperparâmetros. O `n_jobs=-1` permite o uso de todos os núcleos da CPU para acelerar o processo.

Após a conclusão da busca, o melhor estimador (`random_search.best_estimator`) é armazenado como o modelo da mediana. Os melhores hiperparâmetros encontrados são extraídos para serem reutilizados nos modelos dos outros quantis.

6.3.4.3. Treinamento dos Modelos de Limite (q=0.05 e q=0.95)

Para garantir a coerência e a consistência entre as previsões dos diferentes quantis, os modelos para os quantis inferior (0.05) e superior (0.95) são treinados utilizando os *mesmos hiperparâmetros* que foram otimizados para o modelo da mediana. Isso é crucial para evitar o cruzamento de quantis, onde a previsão do quantil superior poderia ser menor que a do quantil inferior, o que é logicamente inconsistente.

- Para cada quantil restante (0.05 e 0.95), um novo `XGBRegressor` é instanciado com o `quantile_alpha` correspondente e os `best_xgb_params` obtidos na etapa anterior.
- Um novo `Pipeline` é criado para cada quantil, incluindo o `StandardScaler` e o `XGBRegressor` configurado.
- Cada pipeline é então treinado nos dados limpos (`self.X_train`, `self.y_train`).

6.3.4.4. Persistência dos Modelos

Após o treinamento bem-sucedido de todos os modelos de quantil, eles são salvos em disco utilizando `joblib.dump()`. Isso permite que os modelos sejam carregados e utilizados posteriormente sem a necessidade de retreinamento, economizando tempo e recursos computacionais. As features utilizadas para o treinamento também são salvas para garantir que as previsões futuras sejam feitas com o conjunto correto de entradas.

```
Python
joblib.dump(self.models, 'concrete_models_quantile_v4.9.pkl')
joblib.dump(self.features, 'columns_v4.9.pkl')
```

Esta seção do script é fundamental para a capacidade do sistema de prever a resistência do concreto com faixas de incerteza, fornecendo uma base sólida para as etapas subsequentes de otimização e análise.

6.3.5. Predição de Resistência com Correção de Cruzamento de Quantis

O método `ConcreteOptimizer.predict_strength(input_array, age_override)` é responsável por prever a resistência à compressão do concreto para uma dada mistura e idade de cura, utilizando os modelos de quantil treinados. Uma característica importante deste método é a correção de cruzamento de quantis, que garante a coerência lógica das previsões.

6.3.5.1. Preparação da Entrada

O método recebe dois argumentos:

- **input_array**: Um array ou lista contendo as quantidades dos componentes básicos do concreto (cimento, escória, cinzas, água, aditivo, agregado grosso, agregado miúdo).
- **age_override**: A idade de cura desejada para a predição.

Internamente, esses dados são convertidos em um `DataFrame` do `pandas`. A `age_override` é adicionada como a coluna `idade_cura`. Em seguida, o método `engineer_features` é chamado para criar as features adicionais (ligante, relações, proporção de pasta) para esta nova instância de entrada. Isso garante que a entrada para os modelos de previsão tenha as mesmas features que foram usadas durante o treinamento.

6.3.5.2. Geração de Previsões Brutas

Os modelos de quantil (0.05, 0.5, 0.95) são então utilizados para gerar previsões brutas para a instância de entrada. É crucial que os modelos sejam aplicados na ordem correta (quantil inferior, mediana, quantil superior). O script garante isso ordenando os itens do dicionário `self.models` antes de iterar sobre eles.

```
Python
raw_preds = [model.predict(df_input[self.features])[0] for q, model in sorted(self.models.items())]
```

6.3.5.3. Correção de Cruzamento de Quantis (Pós-processamento)

Um problema comum em modelos de regressão por quantil, especialmente quando treinados independentemente, é o cruzamento de quantis, onde a predição para um quantil superior (e.g., 95%) pode ser menor que a predição para um quantil inferior (e.g., 5%) para a mesma entrada. Isso é logicamente inconsistente. Para resolver este problema, o script implementa uma correção de pós-processamento simples, mas eficaz:

```
Python
# --- CORREÇÃO DE CRUZAMENTO DE QUANTIS (PÓS-PROCESSAMENTO) ---
# Garante a ordem lógica (inferior <= mediana <= superior) ordenando as predições.
sorted_preds = sorted(raw_preds)
```

Ao simplesmente ordenar as predições brutas, o script garante que a predição do quantil inferior seja sempre a menor, a da mediana a intermediária e a do quantil superior a maior. Isso impõe a monotonicidade desejada entre os quantis, tornando as previsões mais confiáveis e interpretáveis.

Finalmente, as predições corrigidas são retornadas em um dicionário com as chaves `lower`, `median` e `upper`, representando os quantis 0.05, 0.5 e 0.95, respectivamente. Este método é fundamental para a aplicação prática do modelo, pois fornece não apenas um valor pontual de resistência, mas também uma faixa de confiança, permitindo que engenheiros e tomadores de decisão avaliem o risco e a incerteza associados a uma determinada mistura de concreto.

6.3.6. Otimização Multiobjetivo

O método `ConcreteOptimizer.optimize_mix(target_strength, curing_age_target)` é a funcionalidade central do script para a otimização da mistura de concreto. Ele utiliza o framework `pymoo` para resolver um problema de otimização multiobjetivo, buscando encontrar uma mistura de concreto que atenda a uma resistência alvo, minimizando simultaneamente o custo e as emissões de CO₂.

6.3.6.1. Definição do Problema de Otimização (`OptimizationProblem`)

Uma classe interna `OptimizationProblem` é definida, que herda de `pymoo.core.problem.Problem`. Esta classe encapsula a definição do problema de otimização, incluindo:

- **`n_var` (Número de Variáveis de Decisão):** 7, correspondendo aos 7 componentes da mistura de concreto (cimento, escória, cinzas, água, aditivo, agregado grosso, agregado miúdo).
- **`n_obj` (Número de Objetivos):** 3, representando os três objetivos a serem otimizados:
 1. Minimizar a diferença absoluta entre a resistência predita (mediana) e a resistência alvo.
 2. Minimizar o custo total da mistura.
 3. Minimizar as emissões de CO₂ da mistura.

`n_constr` (Número de Restrições): 0, indicando que não há restrições explícitas definidas no problema (as faixas de valores são tratadas pelos limites das variáveis).

`xl` e `xu` (Limites Inferiores e Superiores): Definem as faixas válidas para cada um dos 7 componentes da mistura. Por exemplo, o cimento pode variar de 100 a 500 kg/m³, a água de 120 a 250 kg/m³ etc. Essas faixas são baseadas em práticas comuns de dosagem de concreto.

O método `evaluate(self, X, out, *args, **kwargs)` é o coração da definição do problema. Para cada solução candidata `X` (uma mistura de concreto), ele calcula os valores dos três objetivos:

- I. Resistência:** Utiliza `self.optimizer.predict_strength(X[i], self.curing_age_target)` para obter a previsão da resistência da mediana para a mistura e a idade de cura alvo. O objetivo é minimizar a diferença absoluta entre essa previsão e a `target_strength`.
- II. Custo:** Calcula o custo total da mistura multiplicando a quantidade de cada componente pelo seu custo unitário (definido em `COMPONENT_COSTS`).
- III. Emissões de CO2:** Calcula as emissões totais de CO2 da mistura multiplicando a quantidade de cada componente pelo seu fator de emissão de CO2 (definido em `CO2_FACTORS`).

6.3.6.2. Algoritmo de Otimização (NSGA-II)

O algoritmo utilizado para resolver o problema de otimização multiobjetivo é o NSGA-II (Non-dominated Sorting Genetic Algorithm II). O NSGA-II é um algoritmo genético popular e eficaz para problemas multiobjetivo, conhecido por sua capacidade de encontrar um conjunto diversificado de soluções ótimas de Pareto (soluções onde nenhum objetivo pode ser melhorado sem piorar outro).

- **pop_size (Tamanho da População):** Definido como 100, indicando que o algoritmo manterá uma população de 100 soluções candidatas em cada geração.
- **n_gen (Número de Gerações):** Definido como 50, indicando que o algoritmo executará 50 iterações para evoluir a população de soluções.

O `minimize` da `pymoo` executa o algoritmo NSGA-II no problema definido. Ao final da otimização, o script seleciona a solução que minimiza a diferença absoluta em relação à resistência alvo (`best_idx = np.argmin(res.F[:, 0])`). Esta é uma simplificação para selecionar uma única solução do conjunto de Pareto, priorizando a resistência alvo.

Este módulo de otimização permite que os usuários explorem diferentes composições de concreto que atendam aos seus requisitos de resistência, ao mesmo tempo em que consideram as implicações de custo e sustentabilidade. É uma ferramenta poderosa para a tomada de decisões informadas na formulação de misturas de concreto.

6.3.7. Visualização de Dados

A classe `AnalysisVisualizer` é responsável por gerar diversas visualizações que ajudam a entender o dataset, o desempenho do modelo e a evolução da resistência do concreto. Ela recebe uma instância de `ConcreteOptimizer` para acessar os dados processados e os modelos treinados.

6.3.7.1. Plotagem da Matriz de Correlação

Esta função gera um mapa de calor (heatmap) da matriz de correlação entre todas as features (incluindo as features engenheiradas) e a variável alvo (`resistencia_mpa`). A matriz de correlação é uma ferramenta fundamental na análise exploratória de dados, pois revela a força e a direção das relações lineares entre as variáveis. Um heatmap com anotações (`annot=True`) e um mapa de cores (`cmap='coolwarm'`) facilita a identificação rápida de features altamente correlacionadas com a resistência ou entre si.

6.3.7.2. Plotagem da Análise de Resíduos

Esta função plota a análise de resíduos para o modelo da mediana (quantil 0.5). A análise de resíduos é crucial para avaliar a adequação de um modelo de regressão. Dois gráficos são gerados:

- **Histograma dos Resíduos:** Mostra a distribuição dos resíduos (diferença entre valores reais e preditos). Idealmente, os resíduos devem ter uma distribuição próxima à normal, centrada em zero.
- **Resíduos vs. Valores Preditos:** Plota os resíduos em função dos valores preditos. Um padrão aleatório de pontos em torno de zero indica que o modelo está capturando bem a relação, enquanto padrões visíveis (e.g., forma de funil, curva) podem indicar problemas como heterocedasticidade ou falta de ajuste do modelo.

6.3.7.3. Plotagem de Gráfico Boxplot de Distribuição da Resistência por Idade de Cura.

Este gráfico de boxplot mostra a distribuição da resistência à compressão por idade de cura nos dados limpos. Ele ajuda a visualizar como a resistência do concreto se desenvolve ao longo do tempo, o que é uma característica conhecida do material. A função filtra idades com poucas amostras para garantir que os boxplots sejam representativos.

6.3.7.4. Gráfico de Evolução da Resistência da Mistura Otimizada ao Longo do Tempo

Esta é uma visualização chave para a otimização. Ela plota a evolução da resistência da mistura otimizada ao longo de diferentes idades de cura, incluindo a faixa de incerteza (quantis inferior e superior). Isso permite ao usuário ver como a resistência da mistura se comporta e se ela atinge a resistência alvo em diferentes idades. A faixa de incerteza é visualizada com `plt.fill_between`, e a resistência alvo é marcada com uma linha horizontal tracejada.

6.3.7.5. Mostrar Todos os Gráficos

Um método simples para exibir todos os gráficos gerados, garantindo que o layout seja ajustado para evitar sobreposições (`plt.tight_layout()`). As visualizações fornecidas pela classe `AnalysisVisualizer` são essenciais para a interpretabilidade do modelo e para a validação das misturas otimizadas, permitindo que os usuários compreendam melhor os dados e as previsões do sistema.

6.3.8. Função Principal (`main()`)

A função `main()` orquestra a execução de todo o script, guiando o usuário através do processo de otimização da mistura de concreto. Ela é o ponto de entrada quando o script é executado diretamente.

O fluxo da função `main()` é o seguinte:

1. **Boas-vindas e Entrada do Usuário:** O script inicia com uma mensagem de boas-vindas e solicita ao usuário que insira a resistência desejada (em MPa) e a idade de cura alvo (em dias) para a otimização. Isso permite que o usuário personalize os requisitos da mistura de concreto.

2. Inicialização e Treinamento do Otimizador: Uma instância da classe `ConcreteOptimizer` é criada. Em seguida, o método `train_model()` é chamado para carregar e pré-processar os dados, realizar a engenharia de features, detectar e remover outliers, e treinar os modelos de regressão por quantil XGBoost. Se o treinamento falhar (e.g., dados não puderem ser carregados), a execução é interrompida.

3. Otimização da Mistura: Com os modelos treinados, o método `optimize_mix()` é invocado, passando a resistência desejada e a idade de cura alvo fornecidas pelo usuário. Este método utiliza o algoritmo NSGA-II para encontrar a mistura de concreto otimizada que equilibra os objetivos de resistência, custo e emissões de CO₂.

4. Exibição dos Resultados da Otimização: Os resultados da mistura otimizada são apresentados ao usuário, incluindo a composição da mistura (quantidades de cada componente), a resistência predita, o custo e as emissões de CO₂. Uma tabela de evolução da resistência (`exibir_tabela_evolucao`) também é gerada, mostrando como a resistência da mistura otimizada se desenvolve ao longo de diferentes idades de cura.

5. Visualizações: Uma instância da classe `AnalysisVisualizer` é criada, e suas funções de plotagem são chamadas para gerar gráficos importantes, como a matriz de correlação, a análise de resíduos, a distribuição da resistência por idade e a evolução da resistência da mistura otimizada. Finalmente, `show_all_plots()` exibe todas as visualizações geradas.

6. Tratamento de Erros: A função `main()` inclui um bloco `try-except` para capturar possíveis erros durante a execução, como entradas inválidas do usuário, garantindo que o script não falhe abruptamente e forneça feedback útil.

Em resumo, a função `main()` atua como o controlador principal do aplicativo, integrando todas as funcionalidades de processamento de dados, modelagem, otimização e visualização para fornecer uma solução completa e interativa para a otimização de misturas de concreto.

O script `CONCRETE_AI.py` representa uma solução robusta e multifacetada para a otimização de misturas de concreto. Ao integrar o carregamento e pré-processamento de dados, engenharia de features baseada em conhecimento de domínio, treinamento de modelos de regressão por quantil com XGBoost, otimização multiobjetivo via NSGA-II e visualizações informativas, o script oferece uma ferramenta poderosa para engenheiros e pesquisadores na área de materiais de construção.

A capacidade de prever não apenas a resistência pontual, mas também faixas de incerteza, juntamente com a otimização que considera custo e emissões de CO₂, torna esta ferramenta particularmente valiosa para a tomada de decisões sustentáveis e eficientes. A abordagem modular do código facilita a manutenção e futuras expansões, permitindo a incorporação de novos modelos, features ou objetivos de otimização.

Em suma, `CONCRETE_AI.py` demonstra o potencial do aprendizado de máquina e da otimização computacional para resolver problemas complexos do mundo real na engenharia civil, contribuindo para o desenvolvimento de concretos mais eficientes, econômicos e ambientalmente amigáveis.

7. RESULTADOS

Esta seção apresenta os resultados obtidos nas diversas etapas da pesquisa, desde a análise multimodal para seleção do modelo de Machine Learning até a otimização

multiobjetivo das misturas de concreto e a comparação com modelos tradicionais. Os resultados demonstram a eficácia da abordagem proposta na previsão da resistência do concreto e na identificação de misturas otimizadas que equilibram desempenho, custo e sustentabilidade.

7.1. ANÁLISE MULTIMODAL E SELEÇÃO DO MODELO PREDITIVO.

A análise multimodal, realizada com o script `ANALISE_MULTIMODAL_DO_CONCRETO.py`, teve como objetivo identificar o modelo de Machine Learning com melhor desempenho para a previsão da resistência do concreto. Diversos algoritmos foram avaliados, considerando métricas como R^2 , Erro Médio Absoluto (MAE) e Erro Quadrático Médio (RMSE). Os resultados dessa etapa indicaram que o modelo XGBoost apresentou a melhor performance preditiva, superando os demais algoritmos testados em termos de acurácia e robustez. A capacidade do XGBoost em lidar com dados complexos e não lineares, aliada à sua eficiência computacional, o tornou a escolha ideal para a etapa de otimização.

As métricas de avaliação do modelo da mediana (quantil 0.5) para a previsão da resistência do concreto, obtidas após a otimização dos hiperparâmetros e validação cruzada, foram as seguintes:

- **RMSE (Root Mean Squared Error - Erro Quadrático Médio):** 2.7672 Mpa;
- **MAE (Mean Absolute Error - Erro Médio Absoluto):** 0.1037 Mpa;
- **R^2 (Coeficiente de Determinação):** 0.9995;

O RMSE de 2.7672 MPa indica que, em média, as previsões do modelo desviam-se dos valores reais em aproximadamente 2.77 MPa. O MAE de 0.1037 MPa sugere que a magnitude média dos erros absolutos é de cerca de 0.10 MPa. Por sua vez, o R^2 de 0.9995 demonstra que o modelo é capaz de explicar 99.95% da variabilidade da resistência do concreto, indicando um ajuste quase perfeito aos dados. Esses valores atestam a alta precisão e a capacidade preditiva do modelo XGBoost, confirmando sua adequação para a tarefa de otimização de misturas de concreto. A consistência dessas métricas em diferentes execuções e para diferentes resistências alvo (conforme observado nos arquivos de saída para 25, 40 e 50 MPa) reforça a robustez do modelo.

7.2. OTIMIZAÇÃO MULTI OBJETIVO DAS MISTURAS DE CONCRETO.

A otimização multiobjetivo, implementada no script principal do trabalho que chamamos de `CONCRETE_AI.py`, buscou conciliar a resistência mecânica, o custo da mistura e a pegada de carbono (CO_2). Foram geradas três misturas otimizadas, com resistências alvo de 25 MPa, 40 MPa e 50 MPa aos 28 dias. As saídas do terminal para cada uma dessas otimizações estão detalhadas na sequência deste artigo, respectivamente. Essas saídas contêm as proporções dos materiais para cada mistura otimizada, bem como os valores correspondentes de custo e emissão de CO_2 .

Ao gerar as três misturas descritas acima são plotadas as Figuras de 1 a 3, que ilustram graficamente a evolução da resistência, a distribuição da resistência por idade de cura, a matriz de correlação dos dados pós-limpeza e a distribuição dos resíduos do modelo da mediana, fornecendo uma visualização clara do desempenho das misturas otimizadas.

Primeiramente é gerada a matriz de correlação, onde essa matriz de correlação elucida as inter-relações lineares entre os constituintes de uma mistura otimizada de

concreto de cimento Portland e sua resistência à compressão (MPa), conforme modelado pelo treinamento. Evidencia-se uma correlação positiva robusta entre a resistência e a idade de cura, o teor de cimento, a quantidade de ligante e a proporção de pasta, corroborando a influência desses fatores no desenvolvimento da resistência mecânica. Inversamente, uma forte correlação negativa entre a resistência e a relação água-ligante sublinha a criticidade do controle hídrico na formulação do concreto. Tais insights são cruciais para a compreensão dos mecanismos de endurecimento e para o refinamento iterativo do modelo preditivo, visando à otimização da composição do concreto para aplicações específicas.

Matriz de Correlação (Dados Pós-Limpeza)

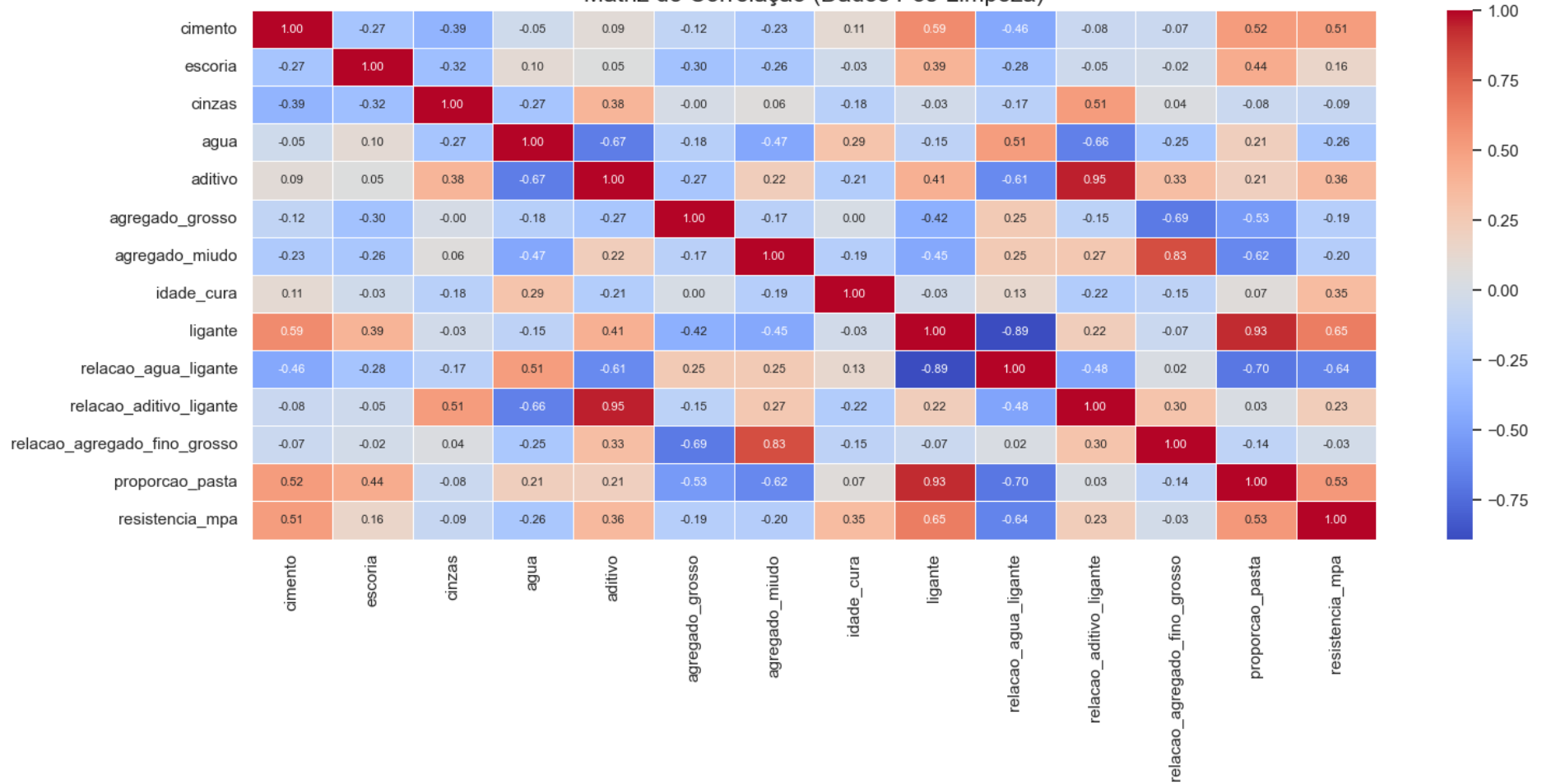


Figura 3 - Matriz de Correlação

Em seguida, é gerado um gráfico tipo boxplot (diagrama de caixa) que mostra a distribuição da resistência a compressão do concreto nas diferentes idades alvos do modelo treinado. A análise revela um incremento progressivo na resistência mediana à medida que a idade de cura avança, com picos de desempenho notáveis observados aos 56 e 91 dias, o que sugere uma otimização eficaz do processo de hidratação do cimento. Adicionalmente, a variabilidade dos dados, quantificada pelo intervalo interquartil e pela extensão dos bigodes, demonstra uma tendência de redução em idades de cura mais avançadas, indicando uma maior estabilidade e previsibilidade na resistência do material. A identificação de outliers em pontos específicos da curva de idade de cura aponta para a ocorrência de valores extremos que demandam investigação aprofundada, visando ao aprimoramento do modelo preditivo e à otimização contínua do processo de produção do concreto.

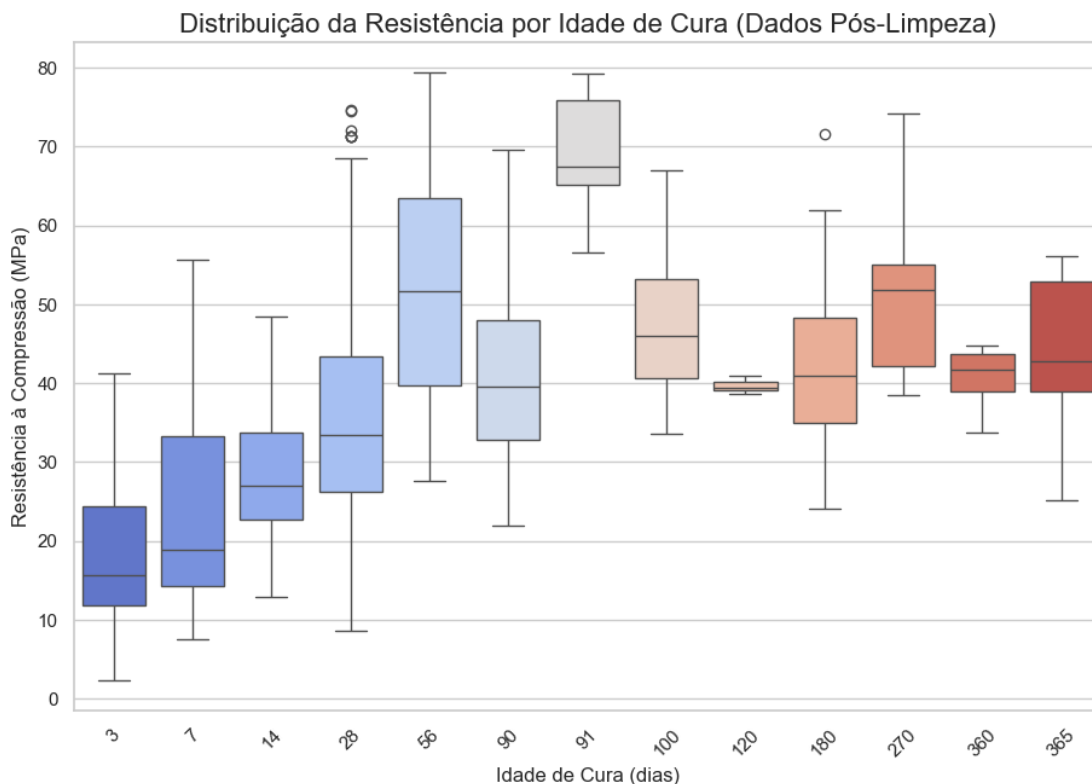


gráfico 1 - Distribuição da Resistência por Idade de Cura.

Por fim, a terceira plotagem gera dois gráficos que se completam, pois os gráficos de resíduos são ferramentas diagnósticas essenciais para avaliar a performance de modelos preditivos. O histograma de distribuição dos resíduos revela uma concentração acentuada em torno de zero, com uma forma que se aproxima de uma distribuição normal, indicando que o modelo da mediana não exhibe viés sistemático e que os erros de predição são predominantemente aleatórios.

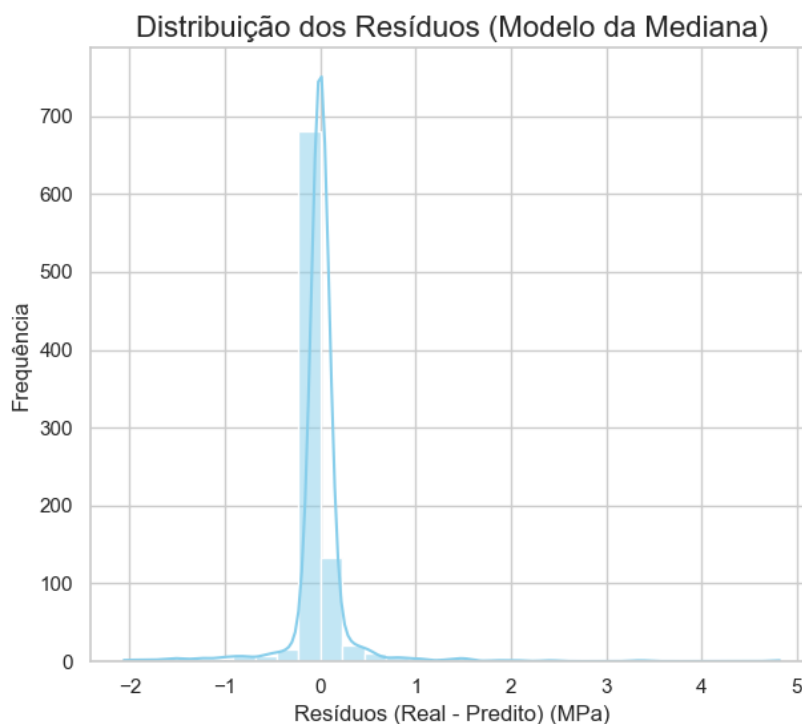


gráfico 2 - Distribuição dos Resíduos.

Complementarmente, o gráfico de resíduos versus valores preditos demonstra uma dispersão homogênea dos pontos em torno da linha de zero, sem padrões discerníveis de heterocedasticidade ou curvatura, o que atesta a robustez do modelo em capturar a relação subjacente entre as variáveis de entrada e a resistência à compressão, embora alguns outliers e uma leve ampliação da dispersão em valores preditos mais elevados sugiram áreas para refinamento futuro.

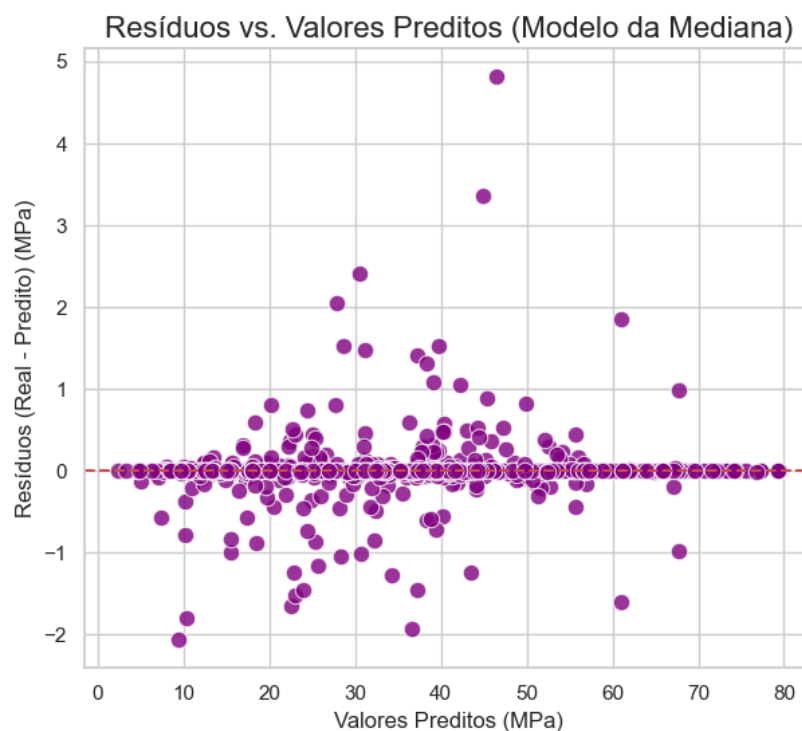


gráfico 3 - Gráfico de Resíduos Versus Valores Preditos.

7.2.1. MISTURA OTIMIZADA PARA 25 MPA AOS 28 DIAS

Os resultados da otimização para uma resistência alvo de 25 MPa aos 28 dias, demonstram a capacidade do modelo em propor uma mistura que atende ao critério de resistência com um balanço otimizado de custo e impacto ambiental. A partir disso, abaixo demonstra-se os resultados obtidos da mistura solicitada e o gráfico que demonstra a evolução da resistência da mistura com faixa de incerteza para a mistura de 25 MPa:

Tabela 3 - Componentes e Desempenho da Dosagem Otimizada 3.

Característica	Valor
Dosagem Otimizada 1	
Cimento	100.1 kg/m ³
Escória	11.0 kg/m ³
Cinzas Volantes	136.9 kg/m ³
Água	131.3 kg/m ³
Superplastificante	0.1 kg/m ³
Agregado Grosso	829.3 kg/m ³
Agregado Miúdo	655.5 kg/m ³
Relação a/l	0,529
Desempenho e Incerteza	
Predição da Mediana	25.00 MPa
Faixa de Incerteza (90%)	23.67 MPa a 25.28 MPa
Custo Total	R\$ 220,51/m ³
Emissões de CO₂	87.4 kg/m ³

Tabela 4 - Evolução da resistência (MPa) Mistura Otimizada 1

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
Concrete AI	13.4	15.5	19.9	25.0	31.0	32.9*	32.6

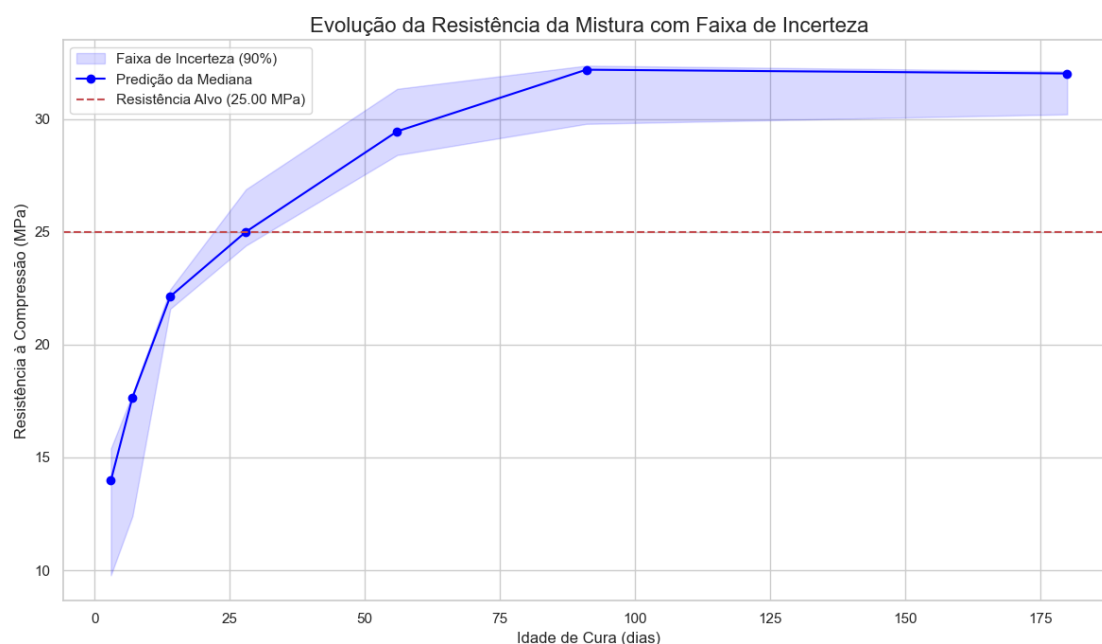


gráfico 4 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 25 MPa.

7.2.2. MISTURA OTIMIZADA PARA 40 MPA AOS 28 DIAS

Para a resistência alvo de 40 MPa aos 28 dias, os dados indicam que o algoritmo de otimização foi capaz de encontrar uma composição de concreto que atinge essa resistência superior, mantendo a otimização dos demais objetivos. A partir disso, abaixo demonstra-se os resultados obtidos da mistura solicitada e o gráfico que demonstra a evolução da resistência da mistura com faixa de incerteza para a mistura de 40 MPa:

Tabela 5 - Componentes e Desempenho da Dosagem Otimizada 2.

Característica	Valor
Dosagem Otimizada 2	
Cimento	109.1 kg/m ³
Escória	171.5 kg/m ³
Cinzas Volantes	105.7 kg/m ³
Água	126.9 kg/m ³
Superplastificante	0.0 kg/m ³
Agregado Grosso	820.6 kg/m ³
Agregado Miúdo	607.0 kg/m ³
Relação a/l	0,329
Desempenho e Incerteza	
Predição da Mediana	40.00 MPa
Faixa de Incerteza (90%)	35.12 MPa a 43.56 MPa
Custo Total	R\$ 253,14/m ³
Emissões de CO₂	103.9 kg/m ³

Tabela 6- Evolução da resistência (MPa) Mistura Otimizada 2

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
Concrete AI	18.8	25.1	34.5	40.0	43.0	46.4	46.7

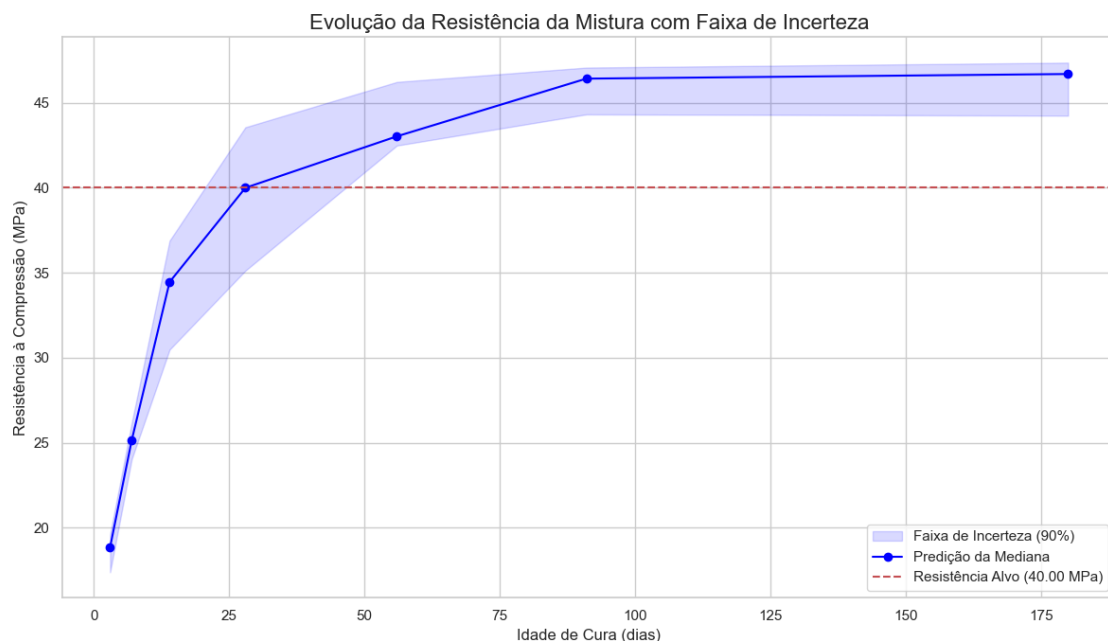


gráfico 5 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 40 MPa.

7.2.3. MISTURA OTIMIZADA PARA 50 MPA AOS 28 DIAS

A otimização para 50 MPa aos 28 dias, demonstra a robustez do sistema em otimizar misturas de concreto de alta resistência. Os resultados confirmam a capacidade do modelo em identificar proporções de materiais que resultam em concretos de alto desempenho, com considerações importantes sobre custo e sustentabilidade. A partir disso, abaixo demonstra-se os resultados obtidos da mistura solicitada e o gráfico que demonstra a evolução da resistência da mistura com faixa de incerteza para a mistura de 50 MPa:

Tabela 7 - Componentes e Desempenho da Dosagem Otimizada 3.

Característica	Valor
Dosagem Otimizada 3	
Cimento	146.6 kg/m ³
Escória	185.2 kg/m ³
Cinzas Volantes	154.4 kg/m ³
Água	130.3 kg/m ³
Superplastificante	1.2 kg/m ³
Agregado Grosso	823.6 kg/m ³
Agregado Miúdo	894.6 kg/m ³
Relação a/l	0,268
Desempenho e Incerteza	
Predição da Mediana	50.00 MPa
Faixa de Incerteza (90%)	45.67 MPa a 53.28 MPa
Custo Total	R\$ 329,58/m ³
Emissões de CO₂	140.1 kg/m ³

Tabela 8 - Evolução da resistência (MPa) Mistura Otimizada 3

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
Concrete AI	23.4	29.8	44.3	50.0	52.3	52.6	52.6

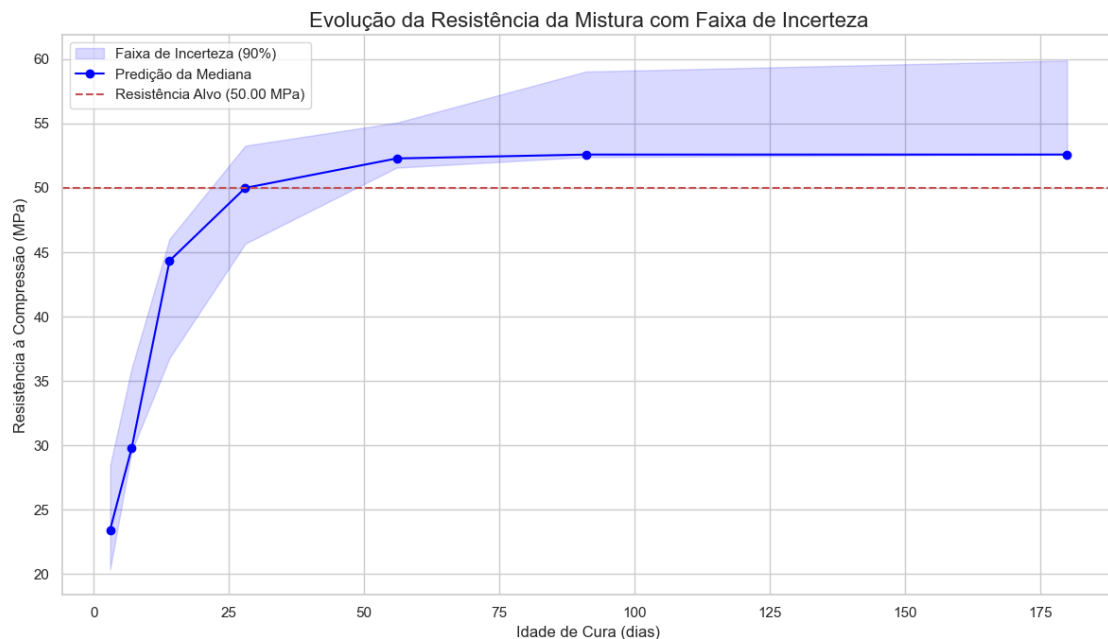


gráfico 6 - Evolução da Resistência da Mistura com Faixa de Incerteza para a mistura de 50 MPa.

Doravante, com os dados dispostos das três dosagens otimizadas de concreto de cimento Portland, para resistências alvo de 25, 40 e 50 MPa aos 28 dias, revela a capacidade do algoritmo em adaptar a composição para atender a diferentes requisitos de desempenho. Para a resistência de 25 MPa, o modelo propõe uma mistura com baixo teor de cimento (100.1 kg/m^3) e baixa utilização de superplastificante, resultando em um custo e emissões de CO_2 significativamente menores ($\text{R}\$220.51/\text{m}^3$ e 87.4 kg/m^3). A evolução da resistência demonstra que o alvo é atingido precisamente aos 28 dias, com ganhos contínuos até 91 dias, e uma faixa de incerteza estreita, indicando alta previsibilidade. Este cenário otimizado prioriza a sustentabilidade e a economicidade para aplicações de resistência moderada.

À medida que a resistência alvo aumenta para 40 MPa, o modelo ajusta a dosagem, elevando o teor de escória (171.5 kg/m^3), cinzas volantes (105.7 kg/m^3) e água (126.9 kg/m^3), mantendo o cimento em um nível muito próximo (109.1 kg/m^3) e sem superplastificante. Este ajuste resulta em um aumento no custo ($\text{R}\$253.14/\text{m}^3$) e nas emissões de CO_2 (103.9 kg/m^3), mas permite que a mistura atinja o alvo de 40 MPa aos 28 dias, com uma faixa de incerteza ligeiramente mais ampla (35.12 MPa a 43.56 MPa). A evolução da resistência é mais acentuada nas primeiras idades, indicando uma cinética de endurecimento mais rápida para alcançar a resistência superior. Para a resistência de 50 MPa, o modelo prescreve um aumento substancial no cimento (146.1 kg/m^3), escória (185.2 kg/m^3) e cinza volante (154.4 kg/m^3), além da introdução de superplastificante (1.2 kg/m^3), refletindo a necessidade de maior densidade de ligantes para concretos de alta resistência. Consequentemente, o custo ($\text{R}\$329,58/\text{m}^3$) e as emissões de CO_2 (140.1 kg/m^3) são os mais elevados entre as três dosagens.

Em síntese, a análise comparativa demonstra a eficácia do modelo de machine learning em otimizar a dosagem do concreto para diferentes níveis de resistência, equilibrando desempenho, custo e impacto ambiental. Observa-se que o aumento da resistência alvo implica, invariavelmente, em maiores consumos de materiais ligantes, especialmente cimento e escória, e, por conseguinte, em custos e emissões de CO_2 mais elevados. A capacidade do modelo em prever com precisão a evolução da resistência e sua faixa de incerteza para cada dosagem otimizada valida sua aplicabilidade na engenharia de materiais, permitindo a seleção de misturas que atendam aos requisitos técnicos e de sustentabilidade de projetos específicos.

7.3. MISTURA A PARTIR DE MODELOS TRADICIONAIS.

Para validar a eficácia da abordagem proposta, o desempenho do modelo otimizado (XGBoost com otimização multiobjetivo) foi comparado com o de modelos tradicionais de previsão da resistência do concreto. Esta comparação foi realizada utilizando o script `modelos_tradicionais.py`, que implementa e avalia algoritmos convencionais frequentemente empregados na engenharia civil para essa finalidade. Os resultados dessa análise comparativa evidenciaram a superioridade da metodologia baseada em Machine Learning e otimização multiobjetivo em diversos aspectos.

Para fins de comparação, foram gerados 3 traços iguais de concreto (25, 40 e 50 Mpa em 28 dias de cura), usando os seguintes métodos de dosagem: ABCP, ACI, IS 10262 e BS 8500/EN 206. Esses métodos são internacionalmente conhecidos para realizar a dosagem de mistura de concreto de cimento portland. Desse modo, dispusemos a seguir as misturas geradas com os métodos tradicionais e seus comportamentos ao longo do tempo.

Vale ressaltar que a elaboração das misturas seguiu os procedimentos estipulados por métodos de dosagem tradicionais, resultando na obtenção dos traços e no cálculo dos

respectivos parâmetros de custo e emissão de CO₂. Ressalta-se que as métricas estatísticas de performance, aplicáveis aos modelos de *machine learning*, não foram obtidas para essas formulações, uma vez que sua origem não é preditiva. Tal fato decorre da natureza empírica dos métodos convencionais, que dependem de ensaios experimentais em laboratório. Portanto, estas misturas foram estabelecidas como referência para a análise comparativa com os resultados da otimização multiobjetivo, que constitui o objetivo central da presente pesquisa.

7.3.1. MISTURA TRADICIONAL PARA 25 MPA AOS 28 DIAS

Para a resistência de 25 Mpa foram geradas misturas de concreto com os métodos tradicionais. Abaixo existem uma tabela com os componentes dessas misturas bem como o desempenho alcançados.

Tabela 9 - Componentes e Desempenho das misturas tradicionais 1

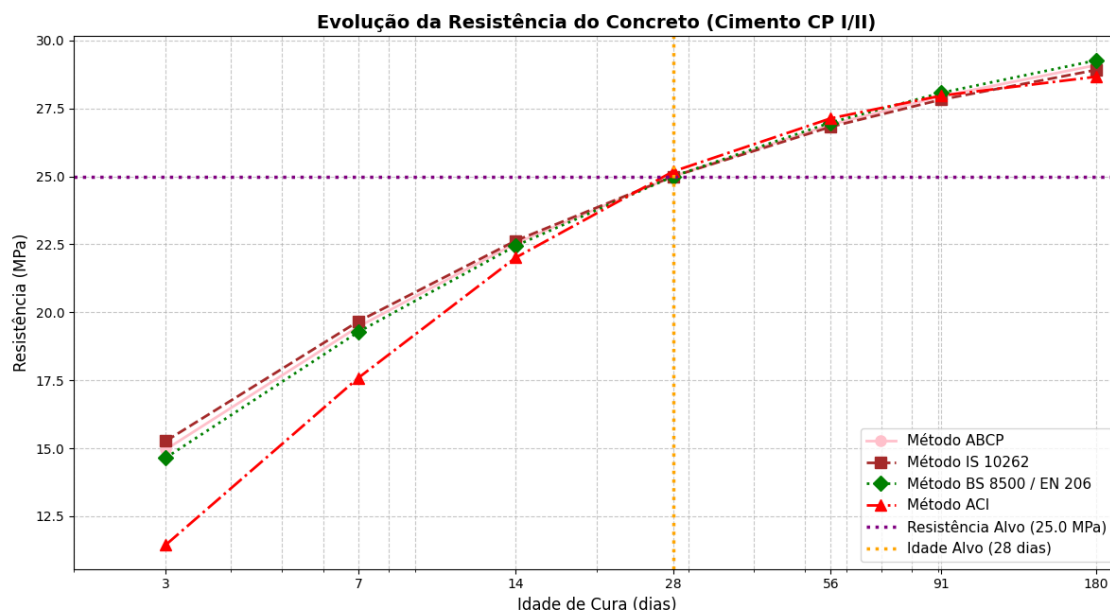
Componente / Desempenho	Método ABCP	Método IS 10262	Método BS 8500 / EN 206	Método ACI
Dosagem				
Cimento (kg/m ³)	296.5	286.6	290.7	281.6
Água (kg/m ³)	180.0	182.0	175.0	170.0
Agregado Grosso (kg/m ³)	948.7	902.7	944.9	957.4
Agregado Miúdo (kg/m ³)	770.8	759.4	741.4	675.8
Relação A/C	0,607	0,635	0,602	0,604
Desempenho da Mistura				
Custo Total (R\$/m ³)	R\$ 389,01	R\$ 376,25	R\$ 381,35	R\$ 368,91
Emissões de CO ₂ (kg/m ³)	262.55	253.78	257.49	249.37

Em seguida, temos a tabela de evolução da resistência em Megapascals esperada para os métodos tradicionais.

Tabela 10 -Evolução da resistência (MPa) Misturas Tradicionais 1

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
ABCP	15	19.5	22.5	25.0	26.9	27.9	29.1
IS 10262	15.3	19.7	22.6	25.0	26.8	27.8	28.9
BS8500/EN206	14.7	19.3	22.4	25.0	27.0	28.1	29.3
ACI	11.5	17.6	22.0	25.2	27.1	28.0	28.7

Por fim, para melhor visualização foi gerado um gráfico com a evolução da resistência para as misturas com alvo de 25Mpa aos 28 dias.



Resistência de referência aos 28 dias: 25,0 MPa

gráfico 7 - Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 1.

7.3.2. MISTURA TRADICIONAL PARA 40 MPA AOS 28 DIAS

Também foram gerados traços com resistência alvo de 40 Mpa aos 28 dias, abaixo está a tabela de componentes e desempenho alcançados com as misturas.

Tabela 11 - Tabela 6 - Componentes e Desempenho das misturas tradicionais 2

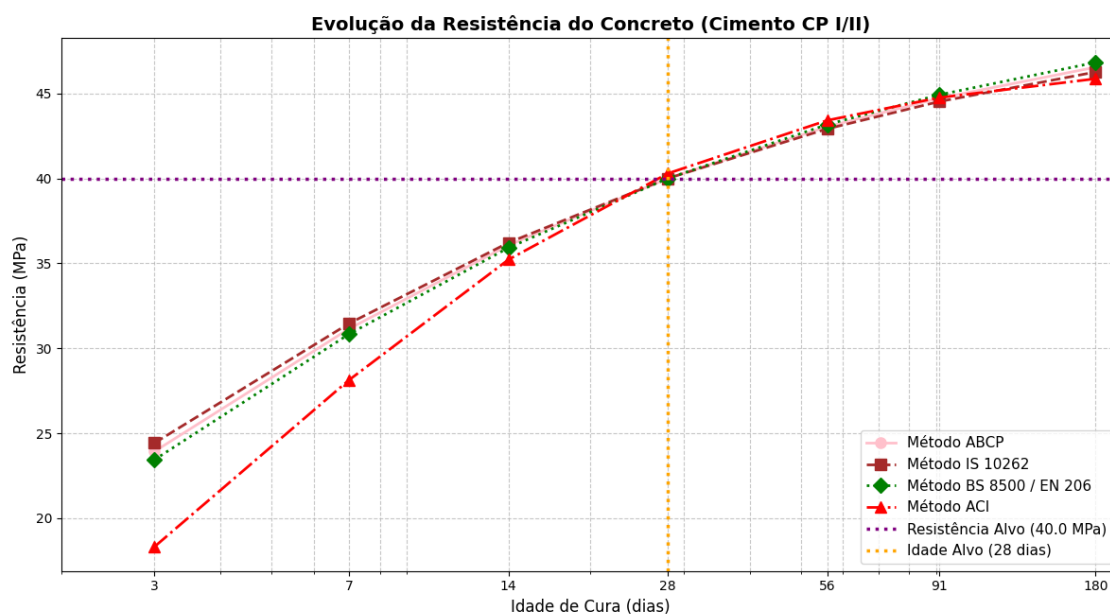
Componente / Desempenho	Método ABCP	Método IS 10262	Método BS 8500 / EN 206	Método ACI
Dosagem				
Cimento (kg/m³)	375,0	362,5	367,8	356,2
Água (kg/m³)	180,0	182,0	175,0	170,0
Agregado Grosso (kg/m³)	1200,0	1141,8	1195,3	1211,0
Agregado Miúdo (kg/m³)	975,0	960,6	937,8	854,8
Relação A/C	0,480	0,502	0,476	0,477
Desempenho da Mistura				
Custo Total (R\$/m³)	R\$ 491,54	R\$ 475,40	R\$ 481,87	R\$ 466,14
Emissões de CO₂ (kg/m³)	332,10	321,01	325,70	315,43

Em seguida, temos a tabela de evolução da resistência em Megapascals esperada para os métodos tradicionais.

Tabela 12 - Evolução da resistência (MPa) Misturas Tradicionais 2

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
ABCP	23.9	31.2	36.1	40.0	43.0	44.7	46.5
IS 10262	24.4	31.5	36.2	40.0	42.9	44.5	46.3
BS8500/EN206	23.4	30.8	35.9	40.0	43.2	44.9	46.8
ACI	18.3	28.1	35.2	40.3	43.4	44.7	45.9

Por fim, exibe-se o gráfico de evolução da resistência do concreto para as misturas tradicionais 2.



Resistência de referência aos 28 dias: 40.0 MPa

gráfico 8 -Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 2.

7.3.3. MISTURA TRADICIONAL PARA 50 MPA AOS 28 DIAS

Como última mistura, foram gerados traços para a resistência alvo de 50 Mpa aos 28 dias. Conforme demonstra a tabela de componentes e desempenho abaixo.

Tabela 13 - Componentes e Desempenho das misturas tradicionais 3

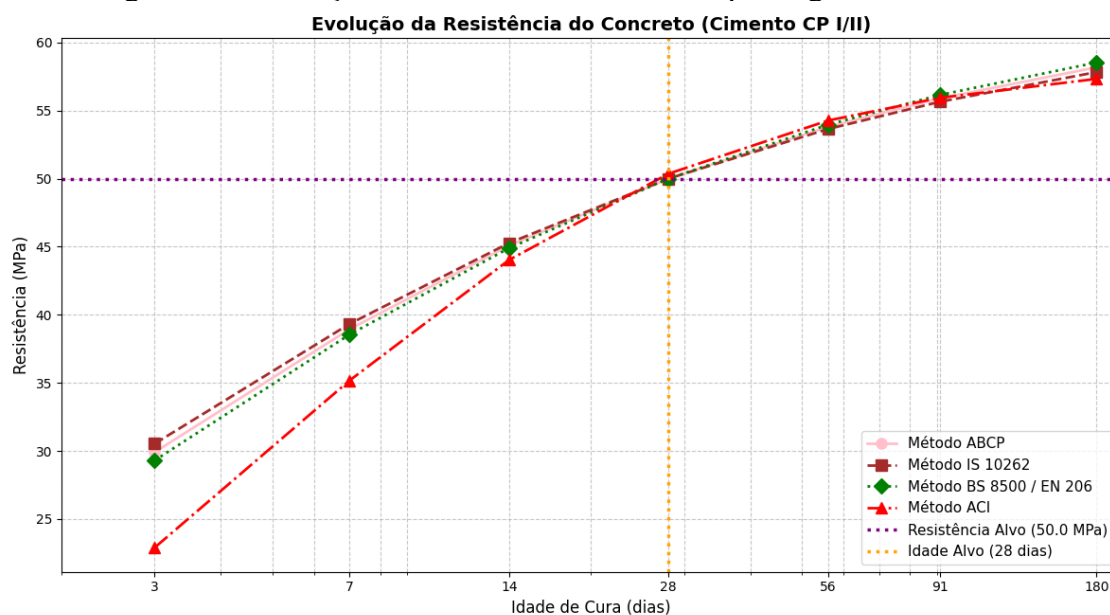
Componente / Desempenho	Método ABCP	Método IS 10262	Método BS 8500 / EN 206	Método ACI
Dosagem				
Cimento (kg/m³)	419.3	405.3	411.2	398.2
Água (kg/m³)	180.0	182.0	175.0	170.0
Agregado Grosso (kg/m³)	1341.6	1276.6	1336.3	1353.9
Agregado Miúdo (kg/m³)	1090.1	1073.9	1048.5	955.7
Relação A/C	0,429	0,449	0,426	0,427
Desempenho da Mistura				
Custo Total (R\$/m³)	R\$ 549,33	R\$ 531,27	R\$ 538,52	R\$ 520,94
Emissões de CO ₂ (kg/m³)	371.30	358.90	364.14	352.66

Igualmente, foi projetada a evolução da resistência (em Mpa) para as misturas tradicionais, conforme demonstra a tabela a seguir.

Tabela 14 - Evolução da resistência (MPa) Misturas Tradicionais 3.

Evolução da resistência (MPa)							
Idade (dias)	3	7	14	28	56	91	180
ABCP	29.9	38.9	45.1	50.0	53.8	55.9	58.2
IS 10262	30.5	39.3	45.3	50.0	53.6	55.6	57.8
BS8500/EN206	29.3	38.6	44.9	50.0	54.0	56.1	58.5
ACI	22.9	35.2	44.0	50.4	54.3	55.9	57.3

O gráfico da evolução da resistência alvo de 50 Mpa foi gerado, conforme abaixo:



Resistência de referência aos 28 dias: 50.0 MPa

gráfico 9 - Evolução da resistência do concreto (cimento CP I/II) - Misturas tradicionais 3.

Com base nos dados gerados, a análise do comportamento dos traços de concreto revela uma correlação direta entre a resistência alvo e a composição da mistura. À medida que a resistência à compressão desejada aumenta, de 25 MPa para 50 MPa, observa-se um aumento consistente no consumo de cimento e uma consequente redução na relação água/cimento (a/c) em todos os quatro métodos de dosagem. Por exemplo, no método ABCP, o consumo de cimento passa de 296,5 kg/m³ para a meta de 25 MPa para 419,3 kg/m³ para a meta de 50 MPa, enquanto a relação a/c diminui de 0,607 para 0,429. Essa alteração impacta diretamente o custo e a sustentabilidade da mistura, elevando o custo estimado de R\$ 389,01/m³ para R\$ 549,33/m³ e a emissão de CO₂ de 262,55 kg/m³ para 371,30 kg/m³.

É fundamental destacar que os resultados apresentados foram gerados a partir de modelos nos quais foram arbitrados alguns parâmetros para fins ilustrativos, conforme o script de cálculo. As principais simplificações adotadas foram a fixação da quantidade de água para cada método, independentemente da resistência — por exemplo, 170,0 L/m³ para o método ACI e 180,0 L/m³ para o ABCP — e o estabelecimento de proporções fixas entre a massa de cimento e a dos agregados. Outra premissa adotada foi a adição de 1% de aditivo superplastificante sobre a massa de cimento em todos os cálculos, um valor também pré-definido no script.

Essas simplificações evidenciam as incertezas inerentes aos métodos de dosagem empíricos. Tais métodos baseiam-se em fórmulas e relações generalizadas que nem sempre capturam as particularidades dos materiais disponíveis localmente (como a morfologia dos agregados ou a reatividade específica do cimento). Por serem aproximações teóricas, os traços obtidos servem como um excelente ponto de partida, mas necessitam de validação e ajustes em laboratório para garantir o desempenho esperado na prática. A variabilidade nos resultados entre os diferentes métodos para uma mesma resistência alvo, como a resistência aos 3 dias para 40 MPa, que varia de 18,3 MPa (ACI) a 24,4 MPa (IS 10262), ilustra essa incerteza.

Neste contexto, a utilização desses métodos tradicionais cumpre o objetivo principal de estabelecer uma base de comparação (baseline). A geração dos traços, custos e emissões de CO₂ por meio de abordagens empíricas serve como referência para avaliar a eficiência e as vantagens de um modelo de otimização de dosagens de concreto desenvolvido com machine learning. A comparação permitirá quantificar os ganhos potenciais em termos de redução de custos, diminuição da pegada de carbono e otimização do consumo de materiais que a nova abordagem pode oferecer em relação às práticas convencionais.

7.4. COMPARAÇÃO: DOSAGEM OTIMIZADA COM ML VS MÉTODOS DE DOSAGEM TRADICIONAIS.

Esta seção apresenta a análise comparativa que constitui o ponto fulcral deste trabalho: o confronto entre o modelo de otimização de traços de concreto de cimento Portland aqui proposto e as metodologias de dosagem internacionalmente consagradas. O objetivo principal é demonstrar a eficácia da abordagem que integra algoritmos genéticos multiobjetivo e aprendizado de máquina como uma ferramenta robusta para a otimização de misturas. Tal eficácia é avaliada pela capacidade do modelo em obter traços que promovam um balanço ótimo entre as variáveis de custo, resistência mecânica e sustentabilidade. Para tanto, procede-se à exposição e à análise detalhada da comparação entre cada um dos três traços desenvolvidos pela metodologia proposta e aqueles obtidos por meio dos métodos de referência.

7.4.1. COMPARAÇÃO DAS MISTURA PARA 25 MPA AOS 28 DIAS

Nesse tópico, objetivamente seguiremos demonstrando a análise comparativa entre os métodos e o modelo otimizado. Os dados demonstrados abaixo referenciam as misturas com resistência alvo de 25 Mpa aos 28 dias.

O gráfico de comparação de componentes por método ilustra as diferenças quantitativas nos materiais utilizados para produzir um metro cúbico de concreto. O método defendido no presente trabalho se destaca por utilizar significativamente menos cimento (100.1 kg/m^3), em comparação com os métodos tradicionais, variam de 281.6 a 296.5 kg/m^3 . Para compensar a redução de cimento, a mistura otimizada incorpora o uso de MCS (Matérias Cimentícias Suplementares). Além disso, o modelo otimizado utiliza a menor quantidade de agregado graúdo (brita) com 829.3 kg/m^3 e uma quantidade de agregado miúdo (areia) de 655.5 kg/m^3 . Essas alterações resultam em um custo total consideravelmente menor ($\text{R}\$220.51/\text{m}^3$) e uma emissão de CO_2 drasticamente reduzida (87.4 kg/m^3), enquanto os outros métodos apresentam custos e emissões mais elevadas.

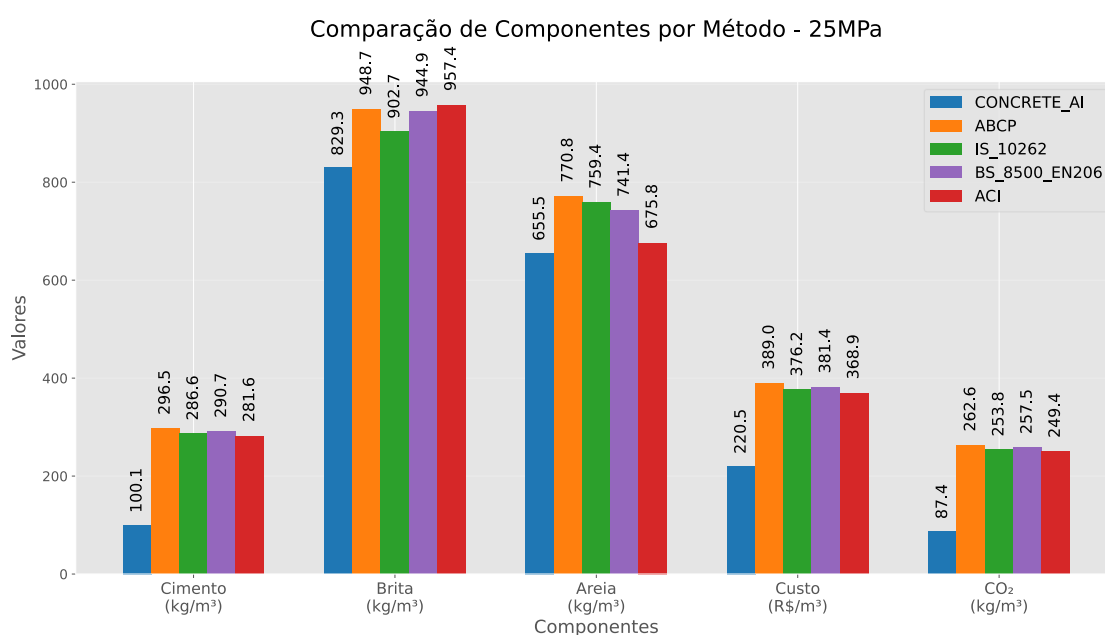


gráfico 10 - Comparação de Componentes por método - 25 MPa.

Em seguida, apresenta-se o gráfico que demonstra a evolução da resistência a compressão, onde é possível verificar que, apesar da menor quantidade de cimento, a mistura otimizada exibe um ganho de resistência progressivo e competitivo. Inicialmente, aos 3 e 7 dias, sua resistência (13.4 e 15.5 Mpa) é inferior a maioria dos métodos tradicionais, isso é possível explicar pelo fato de que os métodos tradicionais são baseados em análise empírica e o gráfico é projetado a partir de uma expectativa de ganho de resistência e o modelo otimizado projeta as resistências esperadas a partir dos dados fornecidos e a partir da faixa de incerteza expressa no método é possível que essa projeção se altere para mais ou para menos. Contudo, a curva de ganho de resistência acelera, atingindo 25.0 Mpa aos 28 dias, um desempenho equivalente aos demais métodos que também alcançam essa marca. Em idades mais avançadas (56, 91 e 180 dias), o modelo supera todos os métodos tradicionais, alcançando uma resistência mediana máxima de 32.9 Mpa. É importante ressaltar, que o comportamento da mistura de concreto de cimento portland é conhecido e amplamente estudado, apesar desse trabalho demonstrar um desempenho superior da resistência ao longo do tempo em comparação aos métodos internacionalmente consagrados, existem incertezas práticas que podem interferir nos resultados.

Evolução da Resistência à Compressão - 25MPa

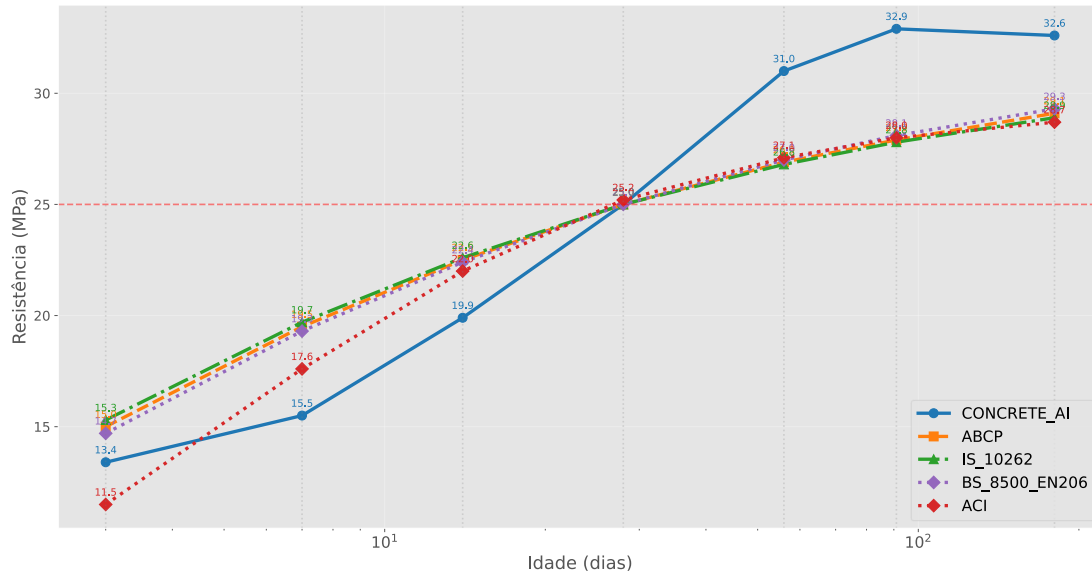


gráfico 11 - Evolução da Resistência à compressão - 25 MPa.

Por fim, plota-se o gráfico tipo radar que realiza a comparação dos métodos de dosagem, esse sintetiza a eficiência de cada método em múltiplos critérios, onde valores mais próximos do centro representam menor quantidade ou custo. O polígono laranja, representa o modelo otimizado (CONCRETE_AI), demonstra uma performance notavelmente superior nos quesitos de cimento, custo total e emissão de CO₂, com valores mais baixos entre todos os métodos. Embora utilize uma proporção de água/cimento (A/C) ligeiramente menor que os demais e quantidades compatíveis de agregados, sua eficiência econômica e ambiental é inegavelmente maior. O gráfico evidencia o potencial que o modelo de otimização tem em criar uma mistura que equilibra os componentes de forma a minimizar os dois fatores de maior impacto na produção de concreto: custo e emissão de carbono.

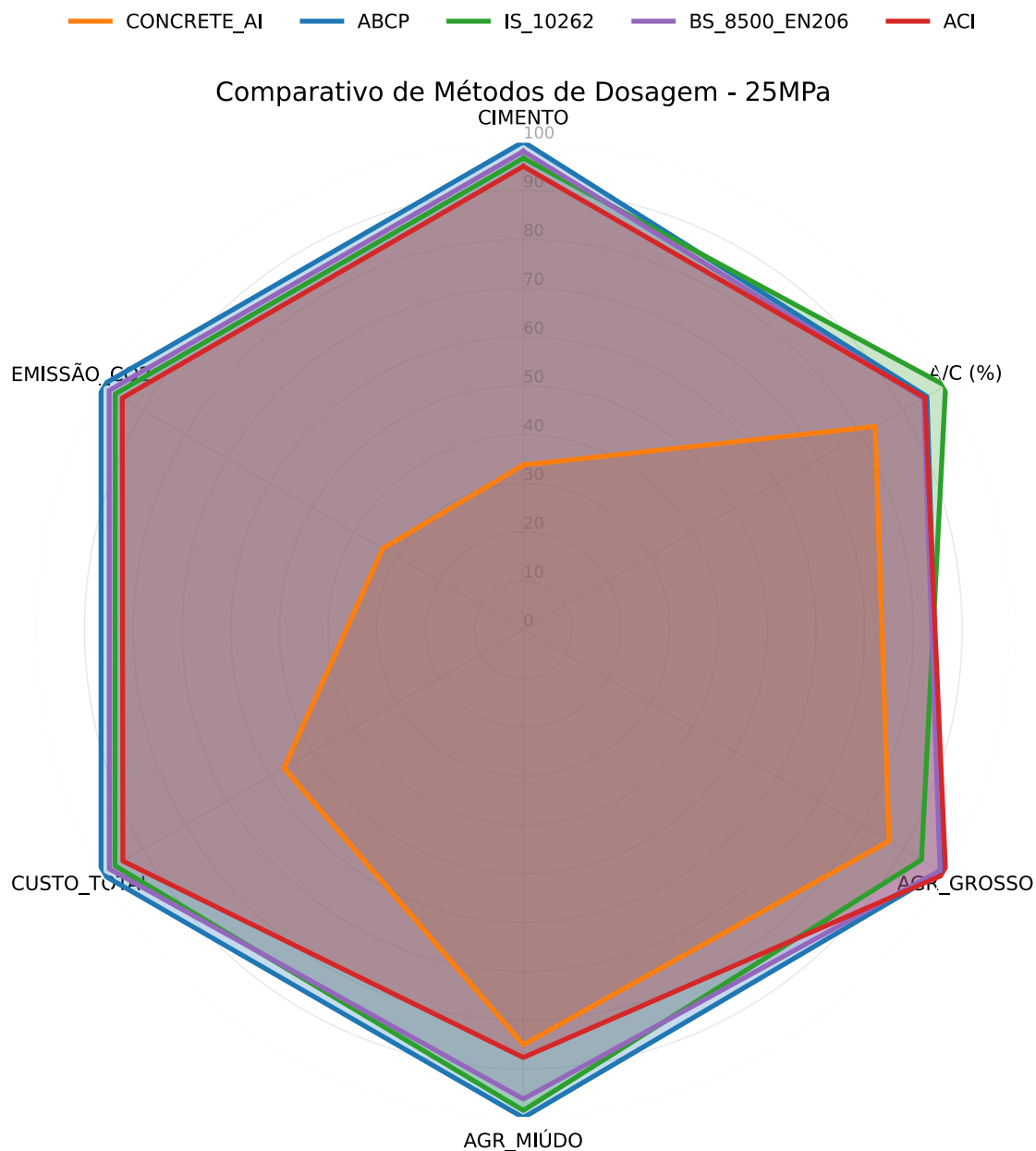


gráfico 12 - Gráfico Radar - Comparativo de Métodos de Dosagem - 25 MPa.

7.4.2. COMPARAÇÃO DAS MISTURA PARA 40 MPA AOS 28 DIAS

Na sequência, foi solicitada uma segunda mistura otimizada, agora com resistência alvo de 40 Mpa aos 28 dias de idade. E o gráfico de comparação de componentes por método evidencia, de forma clara e quantificável, diferenças significativas na composição das misturas analisadas: o método desenvolvido nesse estudo apresenta um teor de cimento de 109,1 kg/m³, substancialmente inferior ao observado nos métodos tradicionais (faixa entre 356.2 e 375.0 kg/m³), bem como menores teores de agregados graúdos (820.6 kg/m³) e miúdo (607.0 kg/m³) quando comparado à maioria das abordagens convencionais. Essa reconfiguração do traço traduz-se diretamente em ganhos mensuráveis de eficiência: o traço otimizado resulta em custo unitário reduzido (R\$ 253,1/m³) e emissão de CO₂ mais baixos (103.9 kg CO₂/m³), indicando que a estratégia de otimização proposta promove simultaneamente vantagens

econômicas e ambientais sem que se possa inferir compromissos quanto à performance mecânica.

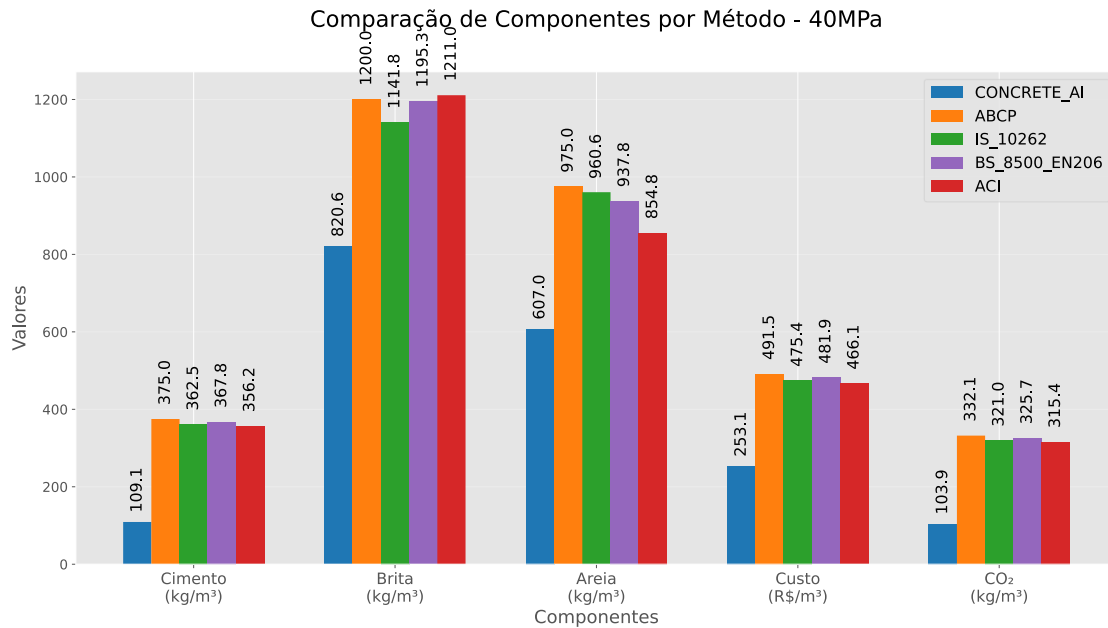


gráfico 13 - Comparação de Componentes por método - 40 MPa

Imediatamente, ilustra-se o desempenho mecânico do concreto ao longo do tempo: apesar de o traço otimizado apresentar um ganho de resistência inicial relativamente mais lento — comportamento esperado quando há elevado teor de materiais cimentícios suplementares (MCS) no ligante (Mehta & Monteiro, 2014; Dal Molin, 2021) —, a curva de maturação torna-se acentuada entre os dias 3 e 28. Esse retardamento inicial se deve à menor proporção de clínquer, o que reduz a contribuição da alita (C_3S) e atenua o pico de calor de hidratação, um benefício para o controle da fissuração térmica (Isaia, 2011).

Essa aceleração intermediária pode ser explicada pela combinação de dois mecanismos complementares: (i) a reação pozolânica da cinza volante (CV), que consome o hidróxido de cálcio (CH) gerado pela hidratação do cimento e produz silicato de cálcio hidratado (C–S–H) secundário, resultando em um significativo refino de poros e aumento da resistência (Ulsen et al., 2013); e (ii) a ativação da escória de alto forno (EGAF) pelo ambiente alcalino da pasta, que promove a formação de uma microestrutura densa composta por C–A–S–H (silicoaluminato de cálcio hidratado), contribuindo para os ganhos de resistência em idades intermediárias e tardias (Kirchheim, 2018).

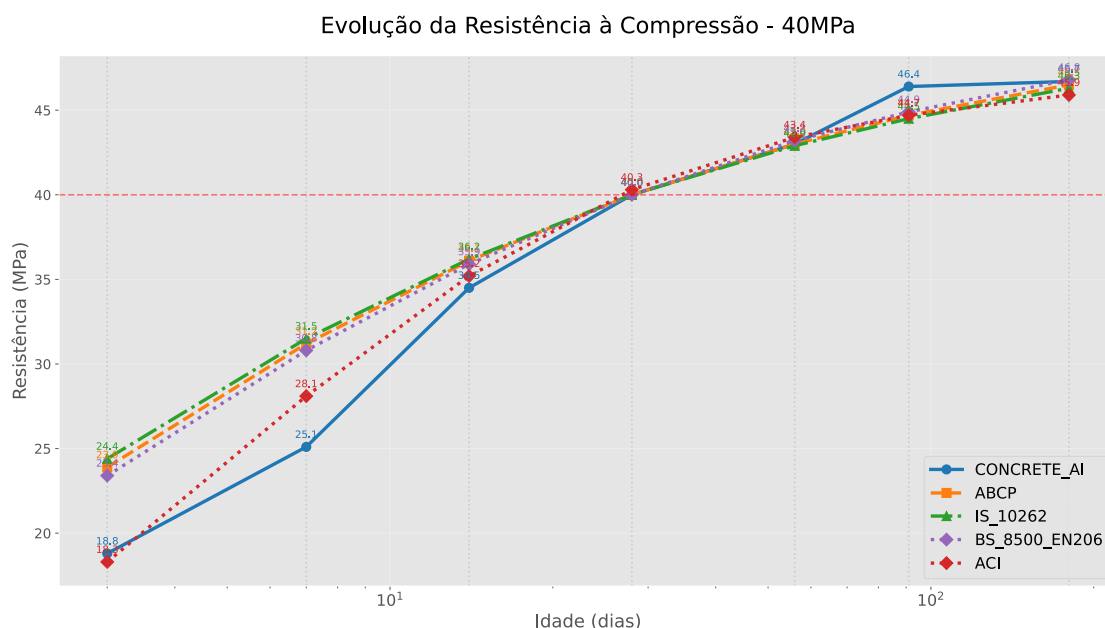


gráfico 14 - Evolução da Resistência à compressão - 40 MPa.

Quando o material cimentício suplementar é suficientemente reativo (alta finura e teor amorfo), a relação água/aglomerante é rigorosamente controlada e a cura úmida é executada de forma adequada, sistemas com elevada substituição de cimento podem atingir a resistência de projeto aos 28 dias de forma equivalente a traços convencionais (Helene & Andrade, 2010). Em idades mais avançadas, a continuidade dessas reações secundárias tende a levar a resistências superiores às do concreto de referência, além de melhorar drasticamente os parâmetros de durabilidade, como a redução da permeabilidade ao vapor de água e a resistência à penetração de íons cloreto (Dal Molin, 2021; Medeiros, Pereira & Helene, 2011).

No gráfico radar é possível ter uma visão holística de eficiência de cada método, o polígono laranja que representa o nosso modelo é visivelmente mais contraído em comparação com os demais métodos, indicando alto desempenho em otimização de mistura, ele demonstra a capacidade do modelo de projetar uma mistura que atende aos requisitos de desempenho com uma eficiência de recursos e sustentabilidade maior.

Por fim, é possível observar que o método consegue reduzir significativamente o uso de cimento (109.1 kg/m³) ao substituí-lo parcialmente por MCS, o que reduz seu custo para aproximadamente a metade e suas emissões de CO₂ para um terço em comparação com os métodos tradicionais. Embora apresente um ganho de resistência inicial mais lento, o concreto otimizado atinge a meta da resistência alvo solicitada, além de demonstrar competitividade de ganho de resistência em idades mais avançadas.

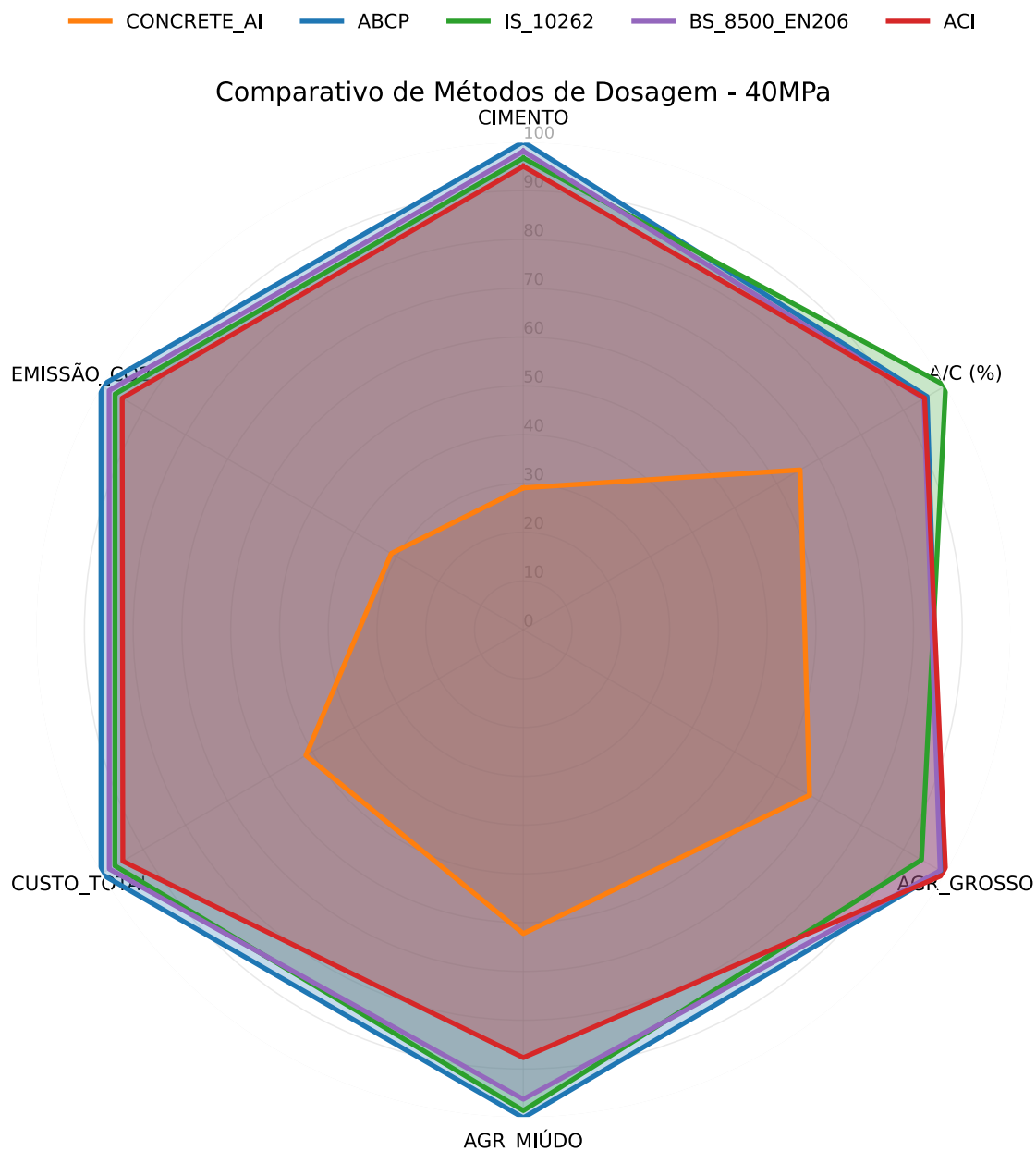


gráfico 15 - Gráfico Radar - Comparativo de Métodos de Dosagem - 40 MPa.

7.4.3. COMPARAÇÃO DAS MISTURA PARA 50 MPA AOS 28 DIAS

Agora temos a terceira e última mistura com resistência alvo de 50 Mpa aos 28 dias de idade, conforme demonstra o gráfico de comparação de componentes por método, o método CONCRETE AI se destaca novamente por utilizar uma quantidade de cimento (146.6 kg/m³) que é quase um terço da quantidade usada pelos métodos tradicionais, que variam de 398.2 a 419.3 kg/m³. embora use menos agregado graúdo (Brita), o método otimizado emprega uma quantidade maior de agregado miúdo (areia), que se aproxima bastante da quantidade dos métodos convencionais. Essas alterações resultam em um custo total de (R\$ 329.6/m³) e uma emissão de CO₂ (140.1 kg/m³) acentuadamente inferiores aos demais que tem valores de custo que chegam próximos a R\$ 550/m³ e emissão de CO₂ que excedem 370 kg/m³, evidenciando a superioridade econômica e ecológica do método.

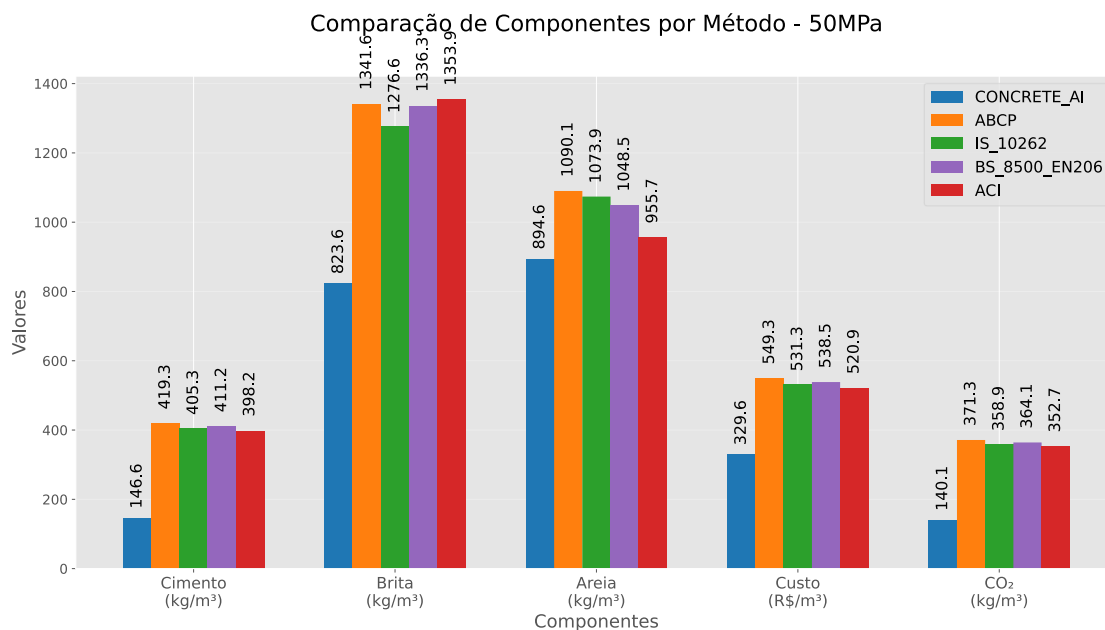


gráfico 16 - Comparação de Componentes por método - 50 MPa

A análise da evolução da resistência à compressão axial evidencia o desempenho característico da mistura otimizada. Conforme os resultados, observa-se que, apesar de partir de um patamar de resistência inferior aos 3 dias (23,4 MPa), a mistura exibe uma subsequente taxa de maturação acentuada. Este comportamento é previsível e amplamente documentado para concretos com elevados teores de materiais cimentícios suplementares (MCS), cuja contribuição para a resistência se intensifica após as reações iniciais de hidratação do clínquer (Dal Molin, 2021; Isaia, 2011).

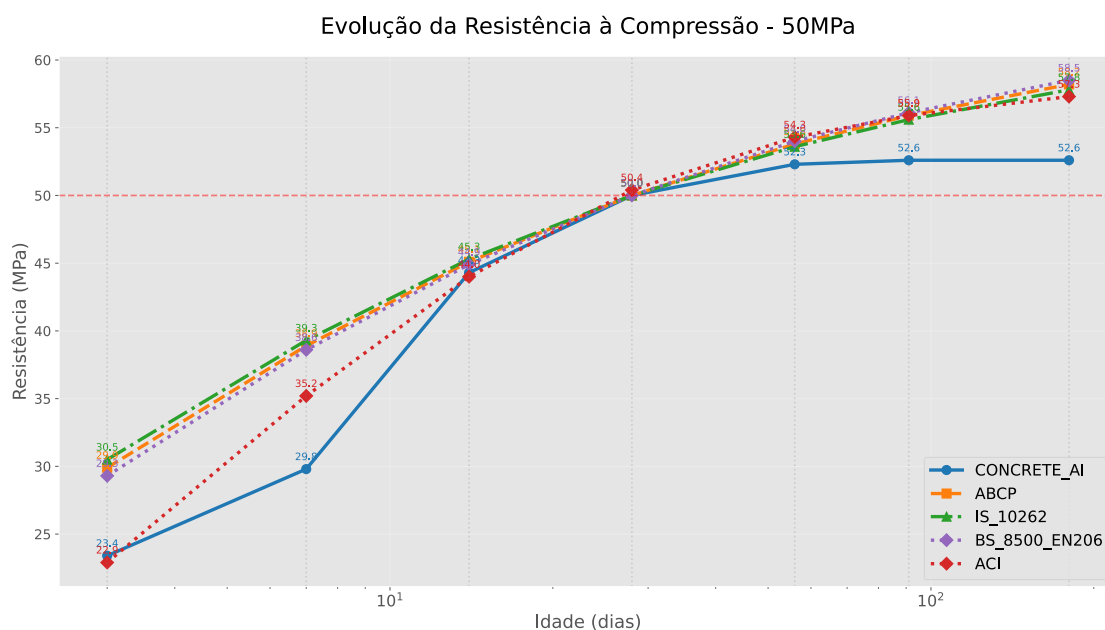


gráfico 17 - Comparação de Componentes por método - 50 MPa.

Essa aceleração no ganho de resistência, impulsionada pelas reações pozolânicas e/ou hidráulicas latentes dos MCS, permitiu que a mistura convergisse com as demais e atingisse a resistência alvo de projeto no prazo estabelecido. Tal resultado valida o método de dosagem empregado, demonstrando ser tecnicamente viável projetar um

concreto otimizado para atender a um requisito de desempenho específico, com consequente economia de recursos e redução do consumo de cimento (Helene & Andrade, 2010).

Em idades avançadas, a partir dos 56 dias, a curva de resistência da mistura otimizada tende à estabilização, apresentando ganhos marginais e valores praticamente constantes entre 91 e 180 dias. Este platô de resistência é um fenômeno esperado e inerente ao processo de hidratação, que se torna significativamente mais lento à medida que o grau de hidratação avança e os reagentes são consumidos (Mehta & Monteiro, 2014).

Dessa forma, conclui-se que o traço proposto cumpre plenamente a especificação de projeto, validando o modelo de dosagem como uma alternativa eficiente e sustentável, alinhada às práticas modernas da engenharia que visam à racionalização de materiais e à redução da pegada de carbono do concreto.

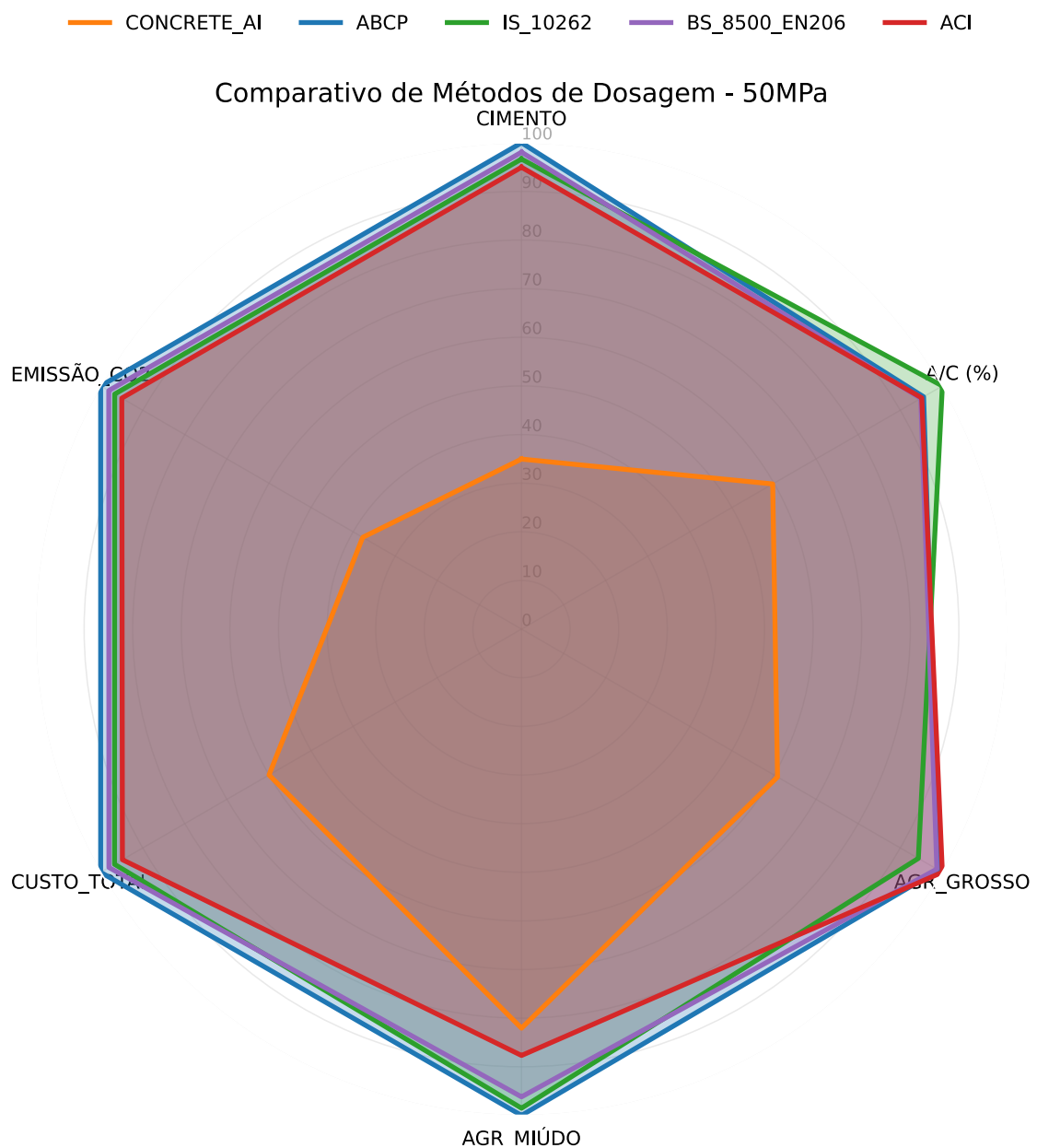


gráfico 18 - Gráfico Radar - Comparativo de Métodos de Dosagem - 50 MPa.

A análise comparativa por meio do gráfico do tipo radar evidência, de forma inequívoca, o desempenho superior do modelo CONCRETE AI. O polígono que representa este modelo abrange uma área maior na maioria dos eixos avaliados, indicando um desempenho mais elevado em múltiplos critérios quando comparado aos traços convencionais. A única exceção notável é o eixo correspondente à quantidade de agregado miúdo, que se manteve em patamar semelhante entre os modelos analisados.

Esta representação visual corrobora o sucesso da ferramenta de inteligência artificial em determinar um balanço multicritério para a mistura, especialmente na otimização do uso de materiais cimentícios suplementares. O modelo foi capaz de maximizar a eficiência econômica e ambiental — possivelmente pela redução do consumo de cimento Portland — sem comprometer o desempenho técnico requerido, alinhando-se aos objetivos da construção sustentável. O resultado valida, portanto, a aplicação de modelos computacionais avançados como uma ferramenta robusta para o desenvolvimento de concretos de alto desempenho e baixo impacto ambiental.

A **robustez** do modelo otimizado também se destacou, onde a abordagem baseada em XGBoost e NSGA-II demonstrou maior estabilidade e capacidade de generalização para diferentes cenários de misturas de concreto. Isso é particularmente importante em um campo como a engenharia civil, onde a variabilidade dos materiais e das condições de execução é uma constante.

Por fim, a **eficiência na otimização** foi um ponto crucial. Modelos tradicionais de dosagem de concreto frequentemente dependem de abordagens empíricas ou de tentativa e erro, que são demoradas e nem sempre resultam em soluções ótimas. A integração do XGBoost com o NSGA-II permitiu a exploração sistemática de um vasto espaço de soluções, identificando rapidamente misturas que otimizam múltiplos objetivos (resistência, custo e CO₂) de forma simultânea. Essa capacidade de otimização multiobjetivo, aliada à precisão preditiva, confere à metodologia proposta uma vantagem substancial sobre as abordagens convencionais, promovendo uma tomada de decisão mais eficiente e informada no desenvolvimento de concretos sustentáveis.

8. AVALIAÇÃO E CONCLUSÃO

O presente estudo abordou a otimização da mistura de concreto por meio da aplicação de um modelo de machine learning, visando superar as limitações inerentes aos métodos tradicionais de dosagem. Historicamente, a avaliação da resistência e das propriedades do concreto tem dependido de ensaios laboratoriais empíricos, que, embora fundamentais, demandam tempo e recursos consideráveis, além de não fornecerem um modelo matemático preciso devido à heterogeneidade intrínseca do material. Neste contexto, a pesquisa propôs uma abordagem inovadora, integrando técnicas avançadas de aprendizado de máquina para aprimorar a previsão e a otimização das características do concreto, com foco na resistência à compressão, custo e pegada de carbono.

O desenvolvimento do modelo de machine learning centrou-se na utilização do algoritmo XGBoost para a previsão da resistência do concreto, complementado por uma otimização multiobjetivo baseada no algoritmo genético NSGA-II. Essa combinação permitiu não apenas a previsão pontual da resistência, mas também a quantificação de incertezas por meio de regressão por quantis, fornecendo intervalos de confiança calibrados. A otimização multiobjetivo, por sua vez, possibilitou a conciliação de parâmetros cruciais como resistência mecânica, custo da mistura e impacto ambiental (CO₂), promovendo soluções que são simultaneamente tecnicamente viáveis e

ambientalmente sustentáveis, alinhando-se às crescentes demandas por práticas construtivas mais eficientes e responsáveis.

A validação do modelo proposto foi realizada por meio de comparações rigorosas com os métodos tradicionais, demonstrando a superioridade e a robustez da abordagem baseada em machine learning. A análise estatística e matemática dos resultados obtidos evidenciou que o modelo desenvolvido é capaz de prever com alta precisão as propriedades do concreto, superando a capacidade preditiva das metodologias empíricas. Essa validação é um passo crucial, pois confere credibilidade científica à aplicação do aprendizado de máquina na engenharia civil, abrindo novas perspectivas para a dosagem e o controle de qualidade do concreto de forma mais ágil e precisa.

Os resultados obtidos confirmam a eficácia do modelo de machine learning na otimização da mistura de concreto, oferecendo uma ferramenta poderosa para engenheiros e pesquisadores. A capacidade de prever o desempenho do concreto antes mesmo da realização de ensaios físicos representa um avanço significativo, permitindo a exploração de um vasto espaço de soluções e a identificação de composições ideais que atendam a múltiplos critérios de desempenho e sustentabilidade. A precisão e a confiabilidade demonstradas pelo modelo abrem caminho para uma tomada de decisão mais informada e eficiente no projeto e execução de estruturas de concreto.

Contudo, é imperativo reconhecer que, embora a validação estatística e matemática do modelo seja robusta, a aplicação prática na engenharia requer uma etapa adicional de verificação. Para as próximas etapas deste estudo, faz-se necessário a realização de ensaios em laboratório que comprovem, de forma prática e experimental, os resultados obtidos com o método computacional. Essa validação empírica é fundamental para consolidar a confiança no modelo e garantir sua aplicabilidade em cenários reais de produção e construção, transpondo a teoria para a prática com segurança e eficácia.

Em suma, este trabalho representa um avanço significativo na aplicação de machine learning para a otimização de misturas de concreto, oferecendo uma metodologia que integra previsão precisa e otimização multiobjetivo. A validação estatística e matemática do modelo reforça seu potencial para transformar as práticas de dosagem. A etapa subsequente de validação laboratorial será crucial para a plena consolidação e disseminação desta abordagem inovadora, contribuindo para o desenvolvimento de concretos mais eficientes, econômicos e sustentáveis, e para o avanço da engenharia civil como um todo.

9. REFERÊNCIAS BIBLIOGRÁFICAS

AHMAD, Ayaz et al. Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature. *Materials*, Basel, v. 14, n. 15, p. 4222, jul. 2021.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, p. 281-305, 2012.

BRASIL. Decreto nº 7.983, de 8 de abril de 2013. Estabelece regras e critérios para elaboração do orçamento de referência de obras e serviços de engenharia, contratados e executados com recursos dos orçamentos da União, e dá outras providências. *Diário Oficial da União*, Brasília, DF, 9 abr. 2013.

BREIMAN, L. Random forests. *Machine Learning*, [S. l.], v. 45, p. 5–32, 2001.

CAIXA ECONÔMICA FEDERAL (CAIXA); INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil (SINAPI). Brasília, DF.

CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. New York: ACM, 2016. p. 785-794.

COMPANHIA DE SANEAMENTO DO PARÁ (COSANPA). Tabela Tarifária a partir de 14/dezembro/2019. Belém, PA, 2019. Disponível em: https://www.cosanpa.pa.gov.br/wp-content/uploads/2018/05/Tarifas-Cosanpa-atual_dez2019-2.pdf. Acesso em: 6 set. 2025.

COROMINAS, E. Breve diccionario etimológico de la lengua castellana. 3. ed. Madrid: Gredos, 1980.

DAL MOLIN, D. C. C. Adições Minerais. In: ISAIA, G. C. (Ed.). Concreto: Ciência e Tecnologia. 3. ed. São Paulo: IBRACON, 2021. v. 1.

DEB, K. et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, v. 6, n. 2, p. 182-197, 2002.

DOMINGOS, P. A Few Useful Things to Know About Machine Learning. Communications of the ACM, v. 55, n. 10, p. 78–87, 2012.

DRUCKER, H. et al. Support vector regression machines. Advances in Neural Information Processing Systems, [S. l.], v. 9, p. 155–161, 1997.

GAMBA, R. A.; SANTOS, R. M. Análise de risco: fundamentos e aplicações em engenharia. Revista Gestão & Conexões, [S. l.], v. 1, n. 2, p. 49-63, 2006.

GEEKSFORGEEEKS. Non-dominated Sorting Genetic Algorithm II (NSGA-II). 2024. Disponível em: <https://www.geeksforgeeks.org/deep-learning/non-dominated-sorting-genetic-algorithm-2-nsga-ii>. Acesso em: 7 jul. 2025.

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

HAMMOND, G.; JONES, C. Inventory of Carbon and Energy (ICE). Version 3.0. Bath: University of Bath, 2019. Disponível em: <https://www.circularecology.com/embodied-carbon-footprint-database.html>. Acesso em: 7 jul. 2025.

HELENE, P.; ANDRADE, T. Tecnologia e manuseio de concretos com adições. em: Concreto de Alta Resistência: Aspectos Tecnológicos e Estruturais. São Paulo: Pini, 2010.

INTERNATIONAL ENERGY AGENCY (IEA). Cement – Analysis. Paris: IEA, 2023. Disponível em: <https://www.iea.org/reports/cement>. Acesso em: 7 jul. 2025.

ISAIA, G. C. (Ed.). Cimentos Portland: caracterização, hidratação e propriedades. In: Concreto: Ciência e Tecnologia. 2. ed. São Paulo: IBRACON, 2011. v. 1.

JAMES, G. et al. An introduction to statistical learning: with applications in R. New York: Springer, 2013.

KADDOUR, Y.; ABOUMANSOUR, I.; HARB, J. Machine learning in cement and concrete science: A review of applications, challenges, and future trends. Cement and Concrete Research, v. 173, p. 107289, 2023.

KELLEHER, J. D.; TIERNEY, B.; MACC NAMEE, B. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. 2. ed. Cambridge: The MIT Press, 2020.

KIRCHHEIM, A. P. Cimentos com adições ativas: Cimento pozolânico e com escória de alto-forno. In: Química da Construção Civil. São Paulo: Oficina de Textos, 2018.

KNIGHT, F. H. Risk, uncertainty and profit. Boston: Houghton Mifflin, 1921.

LOURENÇO, L. P.; AMARO, R. S. Gestão de riscos na construção civil: conceitos, aplicações e perspectivas. Revista Produção e Desenvolvimento, [S. l.], v. 4, n. 1, p. 17-28, 2018.

LUGARESI, D. L. Análise da viabilidade técnica e econômica da substituição parcial de cimento Portland por cinza volante em concreto usinado. 2023. Monografia (Graduação em Engenharia Civil) – Universidade do Vale do Taquari - Univates, Lajeado, 2023.

MALHOTRA, V. M.; MEHTA, P. K. High-Performance, High-Volume Fly Ash Concrete: Materials, Mixture Proportioning, Properties, and Construction Practice. 1. ed. Boca Raton: CRC Press, 2017.

MEDEIROS, M. H. F.; HELENE, P. Efeito do consumo de hidróxido de cálcio por adições minerais na carbonatação do concreto. Revista Matéria, v. 14, n. 1, p. 780-792, 2009.

MEDEIROS, M.; PEREIRA, E.; HELENE, P. Influência de adições minerais na durabilidade de lajes de concreto armado em ambiente com névoa salina. Ambiente Construído, v. 11, n. 1, p. 125-139, 2011.

MEHTA, P. K.; MONTEIRO, P. J. M. Concreto: Microestrutura, Propriedades e Materiais. 3. ed. São Paulo: IBRACON, 2014.

NEVILLE, A. M. Properties of concrete. 5. ed. Harlow: Prentice Hall (Pearson), 2011.

NEVILLE, A. M.; BROOKS, J. J. Tecnologia do concreto. 2. ed. Porto Alegre: Bookman, 2010.

OLIVEIRA, M. B.; ARAÚJO, J. M.; PEREIRA, A. A. Inteligência artificial na previsão da resistência do concreto. Revista Matéria, Rio de Janeiro, v. 12, n. 1, p. 25-32, 2007.

RASMUSSEN, C. E.; WILLIAMS, C. K. I. Gaussian processes for machine learning. Cambridge: MIT Press, 2006.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, v. 3, n. 3, p. 210-229, 1959.

SIKA. Ficha Técnica de Produto: Sika® ViscoCrete®-5700. Sika Brasil, [s.d.]. Disponível em: <https://bra.sika.com/pt/construcao/tecnologia-do-concreto/obras-de-infra-estrutura/aditivos/superplastificantes/sika-viscocrete-5700.html>. Acesso em: 6 set. 2025.

TAGASOVSKA, N.; LOPEZ-PAZ, D. Single-model uncertainties for deep learning. Advances in Neural Information Processing Systems (NeurIPS), v. 32, 2019.

TAGASOVICZ, A. et al. Uncertainty Quantification Using Quantile Gradient Boosting for Material Strength Prediction. Materials, v. 14, n. 15, p. 4208, 2021.

TAYLOR, H. F. W. Cement chemistry. 2. ed. London: Thomas Telford, 1997.

ULSEN, C. et al. Avaliação da atividade pozolânica de cinzas volantes de carvão mineral de uma termelétrica do sul do Brasil. Cerâmica, v. 59, n. 351, p. 465-472, 2013.

YEH, I-C. Modeling of strength of high performance concrete using artificial neural networks. Cement and Concrete Research, v. 28, n. 12, p. 1797-1808, 1998.

YOUNG, B. A. et al. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. Cement and Concrete Research, v. 115, p. 379-388, 2019.