

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**INSTITUTO DE CIÊNCIA EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM**  
**COMPUTACIONAL**

**Geraldo Furtado Neto**

**Metamodelo com arquitetura em duas camadas para a previsão de**  
**propriedades mecânicas de misturas de concreto**

Juiz de Fora

2020



Geraldo Furtado Neto

Metamodelo com arquitetura em duas camadas para a previsão de  
propriedades mecânicas de misturas de concreto

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração em Sistemas Computacionais, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional.

Orientador: Prof. Dr. Leonardo Goliatt da Fonseca

Juiz de Fora

2020

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Furtado Neto, Geraldo.

Metamodelo com arquitetura em duas camadas para a previsão de propriedades mecânicas de misturas de concreto / Geraldo Furtado Neto. – 2020.

77 f. : il.

Orientador: Prof. Dr. Leonardo Goliatt da Fonseca

Dissertação – Universidade Federal de Juiz de Fora, Instituto de Ciência Exatas. Programa de Pós-graduação em Modelagem Computacional, 2020.

1. Inteligência computacional. 2. Concreto. 3. Otimização. 4. Stacking.  
I. Fonseca, Leonardo Goliatt da, orient. II. Título.

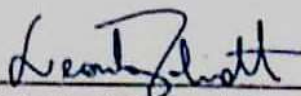
**Geraldo Furtado Neto**

**Metamodelo com arquitetura em duas camadas para a previsão de propriedades mecânicas de misturas de concretos**

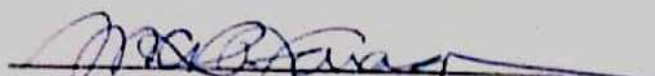
Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração Interdisciplinar, como requisito parcial à obtenção do grau de Mestre em Modelagem Computacional.

Aprovada em 03 de março de 2020

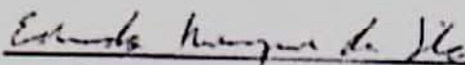
**BANCA EXAMINADORA**




Prof. Dr. Leonardo Goliatt da Fonseca - Orientador  
Universidade Federal de Juiz de Fora



Prof.ª Dr. Michèle Cristina Resende Farage  
Universidade Federal de Juiz de Fora



Prof. Dr. Eduardo Krempser da Silva  
Fundação Oswaldo Cruz - FIOCRUZ



Prof. Dr. Francisco Augusto Lima Manfrini  
Instituto Federal do Sudeste de Minas Gerais



## AGRADECIMENTOS

A Deus por ter me dado a capacidade de seguir em frente e coragem para vencer mais esse grande desafio.

Aos meus pais, Neusa e Geraldo, e minha irmã, Lygia, pelo amor, confiança e apoio em todos os momentos da minha vida.

Aos meus amigos mais próximos pelo apoio ao longo deste trabalho, lembrando-me sempre dos meus objetivos.

Ao meu orientador, Prof. Leonardo, pela dedicação, empenho e paciência ao longo do desenvolvimento deste trabalho.

Aos demais professores que me acompanharam ao longo da minha jornada na Pós-Graduação, pelo comprometimento com o ensino de qualidade.

A todos os funcionários do programa por estarem sempre dispostos a ajudar em todas as situações.





## RESUMO

As propriedades mecânicas do concreto, como módulo de resiliência e resistência à compressão, são de grande importância na etapa de projeto. Determinar estas propriedades através de informações sobre os componentes da mistura e características físicas do concreto, não é trivial. Essa tarefa torna-se complexa pois as relações entre as quantidades e tipos dos componentes da mistura do concreto e suas propriedades mecânicas possuem um alto grau de não-linearidades. O objetivo do presente trabalho é empregar métodos de aprendizado de máquina para criar modelos de regressão, que sejam capazes de prever as propriedades mecânicas de uma mistura de concreto. Visando investigar a possibilidade de melhora dos resultados já obtidos com o uso de técnicas de aprendizado de máquina, será utilizada uma técnica de aprendizado em conjunto, o *Stacking*. A técnica consiste em um modelo de camadas, onde a primeira é composta por modelos individuais e a segunda por um metamodelo que tem a função de utilizar as previsões realizadas pelos modelos de primeira camada como dados de entrada, objetivando realizar uma previsão final. Foram usadas quatro diferentes bases de dados experimentais obtidas da literatura, cada base possuindo quantidade de amostras, número e tipos de atributos, distintos, permitindo verificar a capacidade de generalização da arquitetura criada. Com o uso desta técnica pretende-se que o metamodelo apresente uma capacidade de previsão melhor ou similar a apresentada pelo uso individual dos métodos. Todo o arcabouço computacional foi implementado utilizando a linguagem Python e um conjunto de bibliotecas. Foram utilizados testes estatísticos para avaliar se existia diferença significativa entre o desempenho do *Stacking* e dos modelos individuais. Ao final das análises dos resultados das métricas de avaliação e dos testes estatísticos foi possível perceber que os resultados do *Stacking* são melhores ou no mínimo semelhantes aos apresentados pelos modelos individuais com melhor desempenho.

Palavras-chave: Inteligência computacional. Concreto. Otimização. Stacking.



## ABSTRACT

The mechanical properties of concrete, such as modulus of resilience and compressive strength, are of great importance in the design stage. Determining these properties through information about the mix components and physical characteristics of the concrete is not trivial. This task becomes complex because the relationships between the quantities and types of concrete mix components and their mechanical properties have nonlinearities. The objective of the present work is to employ machine learning methods to create regression models that are able to predict the mechanical properties of a concrete mix. In order to investigate the possibility of improving the results already obtained with the use of machine learning techniques, a joint learning technique, Stacking, will be used. This technique consists of a layer model, where the first is composed of individual models and the second by a metamodel that has the function of using the predictions made by the first layer models as input data, aiming to make a final forecast. Four different experimental databases obtained from the literature were used, each database having different samples, number and types of attributes, allowing to verify the generalizability of the created architecture. With the use of this technique it is intended that the metamodel presents a better predictive capacity than the one presented by the individual methods. The entire computational framework was implemented using the Python language and a set of libraries. Statistical tests were used to assess whether there was a significant difference between Stacking performance and individual models. At the end of the analysis of the results of the evaluation metrics and the statistical tests it was possible to realize that the results of Stacking are better or at least similar to those presented by the best performing individual models.

Key-words: Computational Intelligence. Concrete. Optimization. Stacking.



## LISTA DE ILUSTRAÇÕES

Figura 1	- Modelo esquemático da arquitetura do <i>Stacking</i> . . . . .	19
Figura 2	- Diagrama tensão-deformação para o concreto . . . . .	25
Gráfico 1	- Associação superplastificante e resistência à compressão. . . . .	32
Gráfico 2	- Associação Redutor de água e resistência à compressão. . . . .	34
Gráfico 3	- Associação Água/Cimento e resistência à compressão. . . . .	35
Gráfico 4	- Associação Quantidade de cimento e resistência à compressão. . . . .	36
Figura 3	- Exemplo de uma curva de regressão treinada pelo SVR . . . . .	38
Figura 4	- Modelo de um neurônio artificial . . . . .	39
Figura 5	- Arquitetura de camadas – <i>Stacking</i> . . . . .	44
Figura 6	- Atualização de posição - $\phi_p$ , $\phi_g$ , $r_1$ e $r_2$ igual a 1 . . . . .	46
Figura 7	- Validação cruzada - $k=5$ . . . . .	46
Figura 8	- Topologia do <i>Stacking</i> - Metodologias . . . . .	49
Gráfico 5	- Boxplots Métricas - Base de dados 1 - Primeira Metodologia . . . . .	61
Gráfico 6	- Boxplots Métricas - Base de dados 2 - Primeira Metodologia . . . . .	61
Gráfico 7	- Boxplots Métricas - Base de dados 3 - Primeira Metodologia . . . . .	62
Gráfico 8	- Boxplots Métricas - Base de dados 4 - Primeira Metodologia . . . . .	62
Gráfico 9	- Boxplots Métricas - Base de dados 1 - Segunda Metodologia . . . . .	67
Gráfico 10	- Boxplots Métricas - Base de dados 2 - Segunda Metodologia . . . . .	68
Gráfico 11	- Boxplots Métricas - Base de dados 3 - Segunda Metodologia . . . . .	68
Gráfico 12	- Boxplots Métricas - Base de dados 4 - Segunda Metodologia . . . . .	69



## LISTA DE ABREVIATURAS E SIGLAS

ANFIS	Sistema de inferência neural-fuzzy adaptativo
CART	Árvore de classificação e regressão
DT	Árvore de decisão
ELM	Máquina de aprendizado extremo
KNN	K-vizinhos mais próximos
L-BFGS	Memória limitada <i>Broyden-Fletcher-Goldfarb-Shanno</i>
LR	Regressão linear
MAPE	Erro percentual absoluto médio
MLP	Rede neural de múltiplas camadas
MSE	Erro médio quadrático
NBR	Norma Técnica Brasileira
PSO	Algoritmo de otimização por enxame de partículas
$R^2$	Coefficiente de correlação
RBF	Função de base radial
RMSE	Raiz do erro quadrático médio
RNA	Rede neural artificial
SVM	Máquina de vetores suporte
SVR	Máquina de vetores suporte para regressão





## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>2</b>	<b>CONCEITOS FUNDAMENTAIS E REVISÃO BIBLIOGRÁFICA</b>	<b>21</b>
2.1	CONCRETO	21
2.1.1	<b>Tipos de concreto</b>	22
2.1.1.1	<i>Concreto de agregado leve</i>	22
2.1.1.2	<i>Concreto autocompactável</i>	23
2.1.1.3	<i>Concreto de alto desempenho</i>	23
2.1.2	<b>Propriedades mecânicas</b>	24
2.1.2.1	<i>Resistência a compressão</i>	25
2.2	PREVISÃO DE PROPRIEDADES MECÂNICAS DO CONCRETO	26
2.3	OTIMIZAÇÃO DE PARÂMETROS DE MÉTODOS DE PREVISÃO	28
2.4	ENSEMBLE	29
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>31</b>
3.1	BASE DE DADOS	31
3.1.1	<b>Base de dados 1</b>	31
3.1.2	<b>Base de dados 2</b>	32
3.1.3	<b>Base de dados 3</b>	33
3.1.4	<b>Base de dados 4</b>	35
3.2	MÉTODOS DE REGRESSÃO	36
3.2.1	<b>Máquina de Vetores Suporte (SVM)</b>	37
3.2.2	<b>Rede Neural Artificial (RNA)</b>	39
3.2.3	<b>Máquina de Aprendizado Extremo (ELM)</b>	40
3.2.4	<b>Árvore de decisão (DT)</b>	41
3.2.5	<b><i>K</i>-vizinhos mais próximos (KNN)</b>	42
3.3	STACKING	43
3.4	MÉTODO DE OTIMIZAÇÃO	44
3.5	VALIDAÇÃO CRUZADA	46
<b>4</b>	<b>METODOLOGIA</b>	<b>49</b>
4.1	METODOLOGIAS PROPOSTAS	49
4.1.1	<b>Primeira metodologia</b>	50
4.1.2	<b>Segunda metodologia</b>	50
4.2	MÉTRICAS DE AVALIAÇÃO	52
4.3	TESTES ESTATÍSTICOS	53
4.4	CÓDIGO COMPUTACIONAL	54
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>57</b>

5.1	RESULTADOS EXPERIMENTAIS . . . . .	57
5.1.1	Resultados da primeira metodologia . . . . .	57
5.1.2	Resultados da segunda metodologia . . . . .	63
5.1.3	Comparação entre as metodologias propostas . . . . .	70
5.1.4	Comparação com outras abordagens propostas na literatura . .	71
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>73</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>75</b>

## 1 INTRODUÇÃO

O concreto é um dos materiais mais utilizado na construção civil em todo o mundo. A principal matéria prima do concreto, o cimento, teve uma produção no ano de 2018 de 4,1 bilhões de toneladas, sendo a China o maior mercado, representando 57,8% (USGS, 2019). Este fato está atrelado a algumas características que fazem do concreto um recurso estrutural atrativo. A capacidade de se adaptar a diferentes tipos de formas, elevado tempo de vida útil, baixa demanda de manutenção e boa resistência a intempéries, são algumas das características que fazem do concreto um material tão utilizado.

O aumento do uso do concreto ao longo do anos pode ser explicado pela alta taxa de urbanização, desenvolvimento econômico e o crescimento populacional, estes fatores fazem com que a demanda pelo principal componente do concreto, o cimento, tenha aumentado nos últimos anos (GAO *et al.*, 2017; SHEN *et al.*, 2017).

As propriedades mecânicas e físicas do concreto são informações importantes durante a fase de projeto estrutural, uma vez que cada tipo de projeto possui requisitos próprios de resistência a compressão, durabilidade, peso específico, dentre outros. Existem inúmeros tipos de concreto, cada tipo possui uma aplicação específica, com características que se adaptam aos requisitos de projeto. As propriedades mecânicas estão ligadas aos componentes da mistura do concreto e às quantidades de cada componente. A mistura do concreto tem como componentes básicos: agregados, finos ou graúdos, água e cimento. Adições e aditivos também são incluídos nas misturas. As adições são normalmente materiais minerais com propriedades cimentantes, ou pozolânicas, e os aditivos são produtos químicos que podem, por exemplo, serem utilizados para reduzir a relação água/cimento (NEVILLE, 2015).

A definição das propriedades físicas do concreto é normalmente realizada através de testes laboratoriais. Estes testes são normatizados e permitem definir propriedades como a resistência a compressão, módulo de elasticidade, resistência a tração, dentre outras. Os testes realizados para definir as propriedades mecânicas do concreto são destrutivos, como os definidos pela NBR 5739 (ABNT, 2018), onde corpos de prova são submetidos a cargas até o rompimento dos mesmos.

Uma das propriedades mecânicas mais importantes a ser definida no momento do projeto estrutural é a resistência a compressão, que um dado concreto tem que possuir, visando suportar a carga da estrutura e as demais cargas ao qual o mesmo será submetido.

Uma possibilidade para reduzir a necessidade de testes laboratoriais é estabelecer uma relação entre a quantidade de cada componente da mistura com as propriedades mecânicas do concreto. Isto permitiria definir os parâmetros da mistura a partir de um

valor estabelecido em projeto para uma determinada propriedade mecânica, como por exemplo a resistência a compressão. Porém, definir esta relação não é uma tarefa trivial, uma vez que existe um alto grau de não linearidade na mesma. No entanto esta dificuldade vem sendo superada com o uso de métodos de aprendizado de máquina.

Os métodos de aprendizado de máquina vêm sendo empregados para solucionar problemas em diferentes áreas de pesquisa. Estes métodos baseiam-se na capacidade de um algoritmo realizar o ajuste de pesos ou parâmetros para criar um modelo que seja capaz de realizar a previsão de uma informação. Este ajuste é chamado de aprendizado, o mesmo pode acontecer basicamente de duas formas, supervisionado ou não (WEISS; KULIKOWSKI, 1991). O aprendizado supervisionado acontece quando o algoritmo tem acesso a um conjunto de dados que contenha o resultado real do fenômeno a ser modelado. Cada amostra deste conjunto é formada por um subconjunto de informações que são utilizadas como parâmetros de entrada do modelo e uma ou mais informações que representam os parâmetros de saída do modelo. No caso da previsão da resistência a compressão do concreto, os parâmetros de entrada são informações referentes aos componentes da mistura e outras informações extras, como, por exemplo, o tempo de cura do corpo de prova, e o parâmetro de saída seria a resistência a compressão.

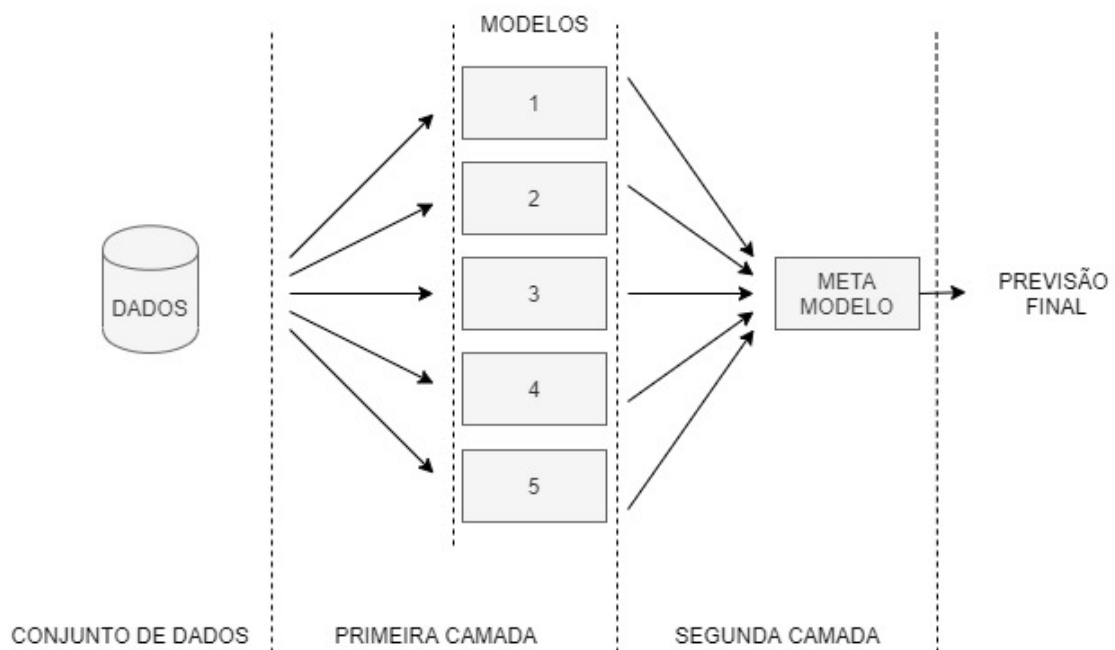
Na literatura podemos encontrar um número significativo de trabalhos que utilizaram métodos como, rede neural artificial, máquina de vetores suporte, árvores de decisão dentre outros para realizar a previsão das propriedades mecânicas do concreto (ABD A.; ABD M., 2017; AHMADU-NEDUSHAN, 2012; KASPERKIEWICZ; RACZ; DUBRAWSKI, 1995; NI; WANG, 2000; TSAI; LIN, 2011). Alguns autores alinharam o uso das técnicas de aprendizado de máquina com o uso de algoritmos de otimização para melhorar o ajuste dos modelos de previsão através da otimização dos seus parâmetros, técnicas como algoritmos genéticos, modelos Bayesianos e enxames de partículas, já foram utilizadas (HUANG; DUN, 2008; LIN *et al.*, 2008; OLIVEIRA *et al.*, 2010).

Visando conseguir um desempenho de previsão melhor que o alcançado com o uso individual das técnicas de aprendizado de máquina, foram desenvolvidas estratégias que permitem utilizar um conjunto destas técnicas para realizar a previsão, estas estratégias são chamadas de *Ensemble*. Existem vários algoritmos que podem ser classificados como *Ensemble*, ou seja, que trabalham com aprendizado em conjunto. Algoritmos do tipo *boosting* (reforço) ou *bagging* (ensacamento), são alguns exemplos de algoritmos que utilizam um conjunto de preditores para realizar uma previsão. Estas técnicas vêm sendo utilizadas em diferentes área, porém são poucos os trabalhos que exploram o uso das mesmas para a previsão das propriedades mecânicas do concreto.

O presente trabalho busca unir estratégias já consolidadas a uma estratégia ainda

pouco explorada, para realizar a previsão da resistência a compressão do concreto. Será utilizada a técnica de aprendizado em conjunto chamada *Stacking*. O *Stacking* é uma estratégia de aprendizado em camadas onde, na primeira camada está presente um conjunto de algoritmos de aprendizado de máquina que são treinados individualmente com os dados disponíveis, na segunda camada um metamodelo, que pode ser uma técnica de aprendizado de máquina ou por exemplo uma regressão linear, encarrega-se de realizar a previsão final tomando como dados de entrada as previsões realizadas pelos modelos de primeira camada. Na Figura 1 pode ser observado um modelo esquemático. O *Stacking* pode apresentar um número maior de camadas, porém neste trabalho serão utilizadas apenas duas.

Figura 1 - Modelo esquemático da arquitetura do *Stacking*



Fonte: Elaborado pelo autor (2020)

Será explorada também a forma de otimização para os modelos pertencentes à primeira camada do *Stacking* e para o metamodelo. O uso dos métodos de otimização tem como objetivo a escolha otimizada dos parâmetros de cada método buscando reduzir o erro de predição. Podemos observar na literatura o uso de diferentes métodos de otimização, como os algoritmos evolutivos (HUANG; DUN, 2008; OLIVEIRA *et al.*, 2010), modelos Bayesianos (SNOEK; LAROCHELLE; ADAMS, 2012) e outros. O algoritmo de otimização PSO (Particle Swarm Optimization) será utilizado, pois possui facilidade de implementação e já se mostrou eficiente pela literatura. Duas metodologias de otimização serão avaliadas, uma em que serão otimizados em conjuntos os parâmetros dos modelos de primeira camada e do metamodelo e outro onde os parâmetros de cada modelo serão

otimizados individualmente.

Os objetivos que se esperam alcançar com este trabalho são: investigar a eficiência do uso do *Stacking* como técnica para previsão da resistência a compressão de misturas de concreto, verificar se os resultados obtidos com *Stacking* são no mínimo similares aos obtidos com o uso individual de técnicas de inteligência computacional e investigar a melhor forma de aplicação do PSO como ferramenta de otimização dos parâmetros dos algoritmos de inteligência computacional utilizados.

A divisão deste trabalho se dará da forma apresentada a seguir. O capítulo 2 trará um embasamento teórico sobre os temas pertinentes ao desenvolvimento do trabalho bem como uma revisão bibliográfica. No capítulo 3 serão apresentadas as bases de dados que serão utilizadas para testar a eficiência das metodologias propostas. Neste capítulo também estão detalhados os métodos de aprendizado de máquina que serão utilizados, assim como as estratégias de aprendizado conjunto e otimização. As metodologias propostas, as métricas e os testes estatísticos utilizados na avaliação dos resultados serão apresentados no capítulo 4. O capítulo 5 apresentará os resultados alcançados e a análise destes. Por fim no capítulo 6 serão apresentadas as conclusões obtidas ao longo do desenvolvimento do trabalho e também os possíveis desdobramentos desta pesquisa.

## 2 CONCEITOS FUNDAMENTAIS E REVISÃO BIBLIOGRÁFICA

Este capítulo irá realizar um alinhamento sobre os temas principais abordados neste trabalho, apresentando os conceitos necessários para que o leitor possa compreender o desenvolvimento realizado. Será explorados também neste capítulo a literatura referência para o desenvolvimento do trabalho.

### 2.1 CONCRETO

O concreto vem sendo ao longo dos anos um dos materiais mais utilizado para construções. Motivos que justificam este uso frente a outros materiais e técnicas são a durabilidade, baixa manutenção em condições de serviço convencionais e a resistência. A mistura base do concreto é composta por cimento e água, meio cimentante, a esta mistura normalmente são acrescentados, agregados, aditivos e adições.

A composição de uma mistura de concreto possui diferentes proporções dos materiais que a constituem, além das proporções, os tipos de, agregados, aditivos e adições, podem diferenciar os tipos de concreto.

Os agregados utilizados nas misturas podem ser classificados, por exemplo, quanto a sua granulometria, peso específico e composição. Quanto à granulometria, os agregados podem ser graúdos, obtidos geralmente através da britagem de rochas, ou miúdos, como a areia. Os agregados representam cerca de 75% do volume da mistura, isso se justifica devido ao custo reduzido dos agregados frente ao do cimento, além de melhorar a estabilidade e durabilidade do concreto. Uma outra forma de diferenciar os agregados é através do seu peso específico, característica essa, muito importante, pois influencia diretamente no peso das estruturas de concreto. Os agregados leves, como a argila expandida, substituem rochas britadas nas misturas, possibilitando uma redução significativa no peso do concreto e reduzem a condutividade térmica, embora tenham, de modo geral, efeito redutor sobre a resistência mecânica.

O uso de aditivos para alterar as propriedades físicas da mistura do concreto é uma técnica utilizada desde as sociedades pré-modernas. Aditivos podem ser definidos por produtos químicos que são adicionados à mistura na porcentagem máxima de 5% em relação à massa de cimento, exceto em casos especiais (NEVILLE, 2015). Os tipos de aditivo são definidos de acordo com o efeito que o mesmo causa na mistura do concreto. Alguns tipos principais são: os redutores de água ou plastificantes, retardadores, aceleradores e incorporadores de ar. Os aditivos plastificantes permitem reduzir a razão água/cimento sem prejuízo à trabalhabilidade e à resistência mecânica do material. Os retardadores e aceleradores, influenciam no tempo de pega, período de ganho inicial de endurecimento e

ganho de resistência do material. Podem ser utilizados, dentre outros casos, para adequar a mistura às temperaturas na qual a mesma será trabalhada. Os aditivos incorporadores de ar são utilizados para tornar o concreto mais coesivo, reduzir a permeabilidade mantendo a resistência mecânica, além de melhorar a trabalhabilidade da mistura.

As adições são materiais utilizados para modificar as características físicas do concreto de forma permanente. Esses materiais normalmente são incorporados em quantidades superiores a 5% em relação à massa de cimento. As adições mais comumente utilizadas são materiais minerais com propriedades pozolânicas, como a sílica ativa e as cinzas volantes. A sílica ativa, obtida como um subproduto da fabricação de silício metálico, quando adicionada ao concreto garante ganho de resistência, aumento de impermeabilidade e por consequência maior durabilidade do concreto. As cinzas volantes são um subproduto da queima do carvão mineral e possuem uma baixíssima granulometria, o que garante que quando adicionada ao concreto permite uma redução significativa da permeabilidade e ganho de durabilidade. As cinzas volantes afetam pouco a resistência do concreto.

### 2.1.1 Tipos de concreto

As aplicações do concreto na construção civil abrangem diversas áreas e demandam diferentes tipos de misturas, com diferentes características físicas.

Os tipos de concreto são diversos, variando a resistência, tempo de pega, peso específico dentre outras características. Neste trabalho serão exploradas bases de dados de três tipos diferentes de concreto.

#### 2.1.1.1 *Concreto de agregado leve*

O concreto de agregado leve é produzido a partir de agregados com baixo peso específico. Este tipo de concreto apresenta redução do peso específico de aproximadamente 30% quando comparado com concretos normais. A NBR 8953 (ABNT, 2015) define concreto leve como sendo o concreto que possui massa específica seca inferior a 2000 kg/m<sup>3</sup>.

O uso do concreto de agregado leve se torna atrativo uma vez que reduz a carga estrutural sobre as fundações, possibilitando que as mesmas sejam menores, consumindo menos material e reduzindo custos. Este tipo de concreto possui boas características de isolamento térmico. O concreto de agregado leve também possibilita uma redução do consumo de material, melhorando a produtividade.

A origem dos agregados leves permite separá-los em dois grupos distintos:

a) Naturais: diatomita, pedra-pomes, escória, cinzas vulcânicas e tufos. Todos



são de origem vulcânica, exceto a diatomita. Estes agregados não são muito utilizados, pois são encontrados apenas em alguns poucos locais. A pedra-pomes produz um concreto com peso específico que varia de 700 a 1400 kg/m<sup>3</sup>. O concreto com pedra-pomes foi utilizado na construção do templo grego, Panteão;

- b) Artificiais: são produzidos por processos industriais onde ocorre a expansão de materiais como a argila, cinzas volantes e escória de alto-forno. Os processos industriais geram em sua maioria agregados com formatos esféricos que podem sofrer ou não britagem de acordo com a granulometria desejada. Os concretos produzidos com esses agregados têm massa específica variando de 1400 a 1800 kg/m<sup>3</sup>.

#### 2.1.1.2 *Concreto autocompactável*

O concreto autocompactável ou autoadensável, é um tipo de concreto que tem como principal característica a fluidez que permite ao mesmo se adensar apenas pelo seu próprio peso, como define a NBR 15823-1:2010 (ABNT, 2010).

O concreto autocompactável é utilizado principalmente em situações onde a estrutura possua uma grande densidade de armaduras ou outros obstáculos, além de estruturas compostas com formas esguias e não convencionais que possuem áreas de difícil acesso. A vantagem da utilização deste tipo de concreto nas situações citadas se dá pela possibilidade de dispensa do uso de vibração para auxiliar na compactação do concreto. A utilização de vibração ocasiona ruídos que geram incômodo à vizinhança e aos funcionários. Além do ruído, o uso da vibração pode acarretar doenças ocupacionais nos operadores de vibradores.

Existem diferentes formas de obter o concreto autocompactável, como o uso de agregados mais finos que o usual, com a alteração da relação água/cimento, ou ainda com o uso de aditivos superplastificantes. Os aditivos superplastificantes, são aditivos químicos comumente a base de naftaleno e melamina. Estes aditivos permitem que ocorra uma maior dispersão das partículas de cimento, aumentando a trabalhabilidade da mistura. O uso destes aditivos pode gerar concreto com abatimentos na ordem de 200 mm (NEVILLE, 2015). No Brasil o conjunto das normas NBR 15823-1 a 6:2010, classifica, estabelece o controle e aceitação, bem como apresenta as diferentes metodologias de testes a serem realizados para a classificação do concreto.

#### 2.1.1.3 *Concreto de alto desempenho*

O concreto de alto desempenho reuni diferentes características como a elevada durabilidade e elevado módulo de elasticidade, mas os altos valores de resistência a compressão representam a característica mais importante deste tipo de concreto. Valores

de resistência acima de 50 MPa já podem ser considerados altos quando comparados com concretos comuns, que alcançam em média 25 MPa, Neville estabelece, porém, que o alto desempenho quanto a resistência deve ser considerado para valores acima de 80MPa (NEVILLE, 2015).

Este concreto reuni em sua composição aditivos e adições, que são utilizados em outros tipos de concreto, com o intuito de reunir as diferentes características proporcionadas pelos mesmos. Na mistura do concreto de alto desempenho sempre vai estar presente um aditivo superplastificante, que possui a função de reduzir a relação água/cimento na mistura, o que permite um ganho de resistência. A relação água/cimento em concretos de alto desempenho fica em torno de 0,25. O uso de adições como sílica ativa e cinzas volantes ocorre sempre em proporções maiores do que em outros tipos de concreto, podendo chegar a 15% da massa de materiais cimentícios.

O uso do concreto de alto desempenho tende a aumentar, pois este material reuni características que estão cada vez mais sendo exploradas nos projetos estruturais. A alta resistência permite a redução das seções e a alta durabilidade garante maior vida útil as estruturas (AÏTICIN, 2000).

### 2.1.2 Propriedades mecânicas

As propriedades mecânicas de um material são um conjunto de informações que permitem caracterizá-lo quanto a reação do mesmo quando submetido a esforços externos. O conhecimento dessas propriedades é de suma importância no momento de projetos de engenharia. Estas propriedades permitem identificar, por exemplo, a deformação ou ainda a carga suportada por um certo tipo de material.

Para definir as propriedades mecânicas de um material são realizados experimentos laboratoriais. Nestes experimentos os materiais, representados por corpos de prova, que são amostras com tamanho e forma predefinidas por normas específicas para cada tipo de experimento, são submetidos a condições de carregamento que simulam as reais condições a que o material será submetido. Além dos tipos de esforços, tração, compressão ou cisalhamento, o material estará submetido a condições específicas de temperatura, umidade e outras condições, visando simular as condições ambientais a que o material estará sujeito durante o seu uso.

Existem diferentes propriedades que podem caracterizar um material, como por exemplo o módulo de elasticidade, a resistência a tração ou compressão, dureza, tenacidade, dentre outras. Neste trabalho será analisada a resistência a compressão. Para o concreto a resistência a compressão é uma das principais propriedades a ser considerada no momento do projeto de estruturas. Esta propriedade está relacionada à capacidade do material em

distribuir a carga aplicada sobre a sua seção transversal.

### 2.1.2.1 Resistência a compressão

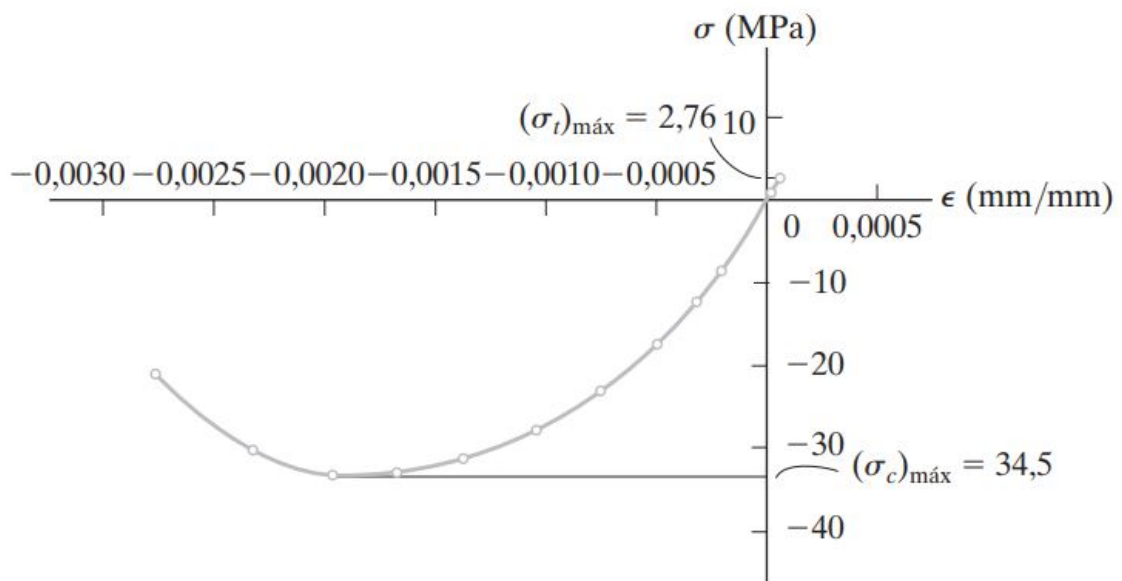
O valor de resistência a compressão de um corpo de prova de concreto é obtido através do ensaio de compressão. Neste ensaio um corpo de prova cilíndrico é submetido a um carregamento, que aumenta de forma gradual e controlada, até que o corpo de prova rompa.

O valor da resistência a compressão é expresso pelo valor da tensão máxima de compressão suportada pelo corpo de prova momentos antes da ruptura. A tensão é a força por unidade de área. No caso deste ensaio a força age na direção perpendicular à área da seção transversal e a tensão é chamada de normal. A tensão normal média é dada pela Equação 2.1

$$\sigma = \frac{P}{A} \quad (2.1)$$

O concreto possui maior valor de resistência a compressão do que de resistência a tração, como pode ser observado na Figura 2, onde é apresentado o diagrama de tensão por deformação do concreto. Para aumentar a resistência a tração do concreto são utilizadas armaduras de aço nas estruturas.

Figura 2 - Diagrama tensão-deformação para o concreto



Fonte: Adaptado de Hibbeler (2010)

## 2.2 PREVISÃO DE PROPRIEDADES MECÂNICAS DO CONCRETO

As propriedades mecânicas de concreto são extremamente importantes para a fase de projeto de engenharia. Comumente para identificar estas propriedades são realizados ensaios, como os já citados anteriormente neste trabalho, em corpos de prova de diferentes misturas de concreto. A partir dos resultados dos ensaios, são escolhidas as misturas a serem utilizadas nos projetos. O fato é que esta metodologia demanda tempo e recursos e ainda fica restrita a um número limitado de misturas possíveis. Uma forma de reduzir a necessidade de ensaios para determinar as propriedades mecânicas de misturas de concreto é conseguir prever estas propriedades através da composição da mistura. O fato é que esta previsão não é trivial uma vez que a relação entre as proporções dos componentes da mistura e as propriedades mecânicas é não linear. Desta forma o uso de técnicas de aprendizado de máquina vem sendo utilizada para modelar a relação entre as propriedades mecânicas e os componentes do concreto.

No trabalho de Abd A. e Abd M. (2017), foram utilizados os métodos de regressão não linear multivariada e a máquina de vetores suporte (SVM), para prever a resistência a compressão. O conjunto de dados utilizado neste trabalho provém de misturas de concreto leve com espuma, ou concreto celular, e contém 150 amostras. Foram utilizados como dados de entrada para os modelos 8 parâmetros relacionados aos componentes da mistura do concreto, como razão água/cimento e granulometria do agregado fino. Também foi utilizado como dado de entrada nos modelos o valor da resistência a compressão para 7 dias de cura. O modelo de regressão não linear multivariada, utilizando como função de perda mínimos quadrados, obteve coeficiente de correlação de 0.9581326, o que representa uma boa correlação entre os valores reais e previstos. Já o modelo SVM utilizando as funções de kernel, base radial, linear, polinomial e sigmoideal, obteve respectivamente os seguintes coeficientes de correlação: 0.986, 0.951, 0.976 e 0.851. Após os testes e análise dos resultados foi possível perceber que os modelos apresentam-se como boas ferramentas para previsão da resistência a compressão, sendo o modelo SVM com função de kernel de base radial, o que apresentou melhores resultados dentre os testados.

O trabalho de Ahmadu-Nedushan (2012) verificou a capacidade de previsão de resistência a compressão de concretos, do algoritmo k-vizinhos mais próximos (KNN), de quatro variações do algoritmo KNN, de redes neurais de regularização generalizada, de um regressão gradual e da rede neural modular. O autor implementou todos os algoritmos citados anteriormente com exceção apenas da rede neural modular, para este modelo o autor usou os resultados apresentados no trabalho de Tsai e Lin (2011). Foi utilizado um algoritmo evolutivo, evolução diferencial, com o intuito de otimizar os parâmetros dos métodos implementados. A base de dados utilizada é composta de 104 amostras de

misturas de concreto de alto desempenho, cada amostra de dados é composta de sete dados de entrada, relacionados aos elementos da mistura, e da resistência a compressão do concreto aos 28 dias. Dos algoritmos derivados do KNN implementados pelo autor destaca-se o que utiliza um conjunto de pesos para ponderar a significância de cada atributo de entrada. Neste algoritmo foram otimizados os valores dos pesos, o número de  $k$  vizinhos e um valor  $p$  relacionado a distância de Minkowski. Este algoritmo apresentou os melhores resultados dentre os outros avaliados, chegando a um valor de 0.9844 de coeficiente de correlação e 1.1739 de RMSE.

A previsão da resistência a compressão para concretos de alto desempenho também foi abordada no trabalho de Al-Shamiri (2019), neste trabalho foi realizada uma comparação entre um rede neural treinada com o algoritmo *Back propagation* e uma maquina de aprendizado extremo (ELM). A ELM é uma rede neural cuje os pesos das sinapses são definidos de forma randômica, o que garante um treinamento rápido e comprovadamente bons resultados. A base de dados utilizada contia 324 amostras, formadas por cinco parâmetros da mistura e pela resistência a compressão aos 28 dias. Os testes realizados com as duas metodologias obtiveram resultados satisfatórios com coeficientes de correlação no ordem de 0.99. Os resultados indicam a eficiências destas técnicas na previsão da resistência a compressão.

Um algoritmo baseado em árvores de decisão foi utilizado por Behnood *et al.* (2017) para prever a resistência a compressão de concretos normais e de alto desempenho. Os algoritmos baseados em árvore de decisão permitem a melhor manipulação de base de dados com um número expressivo de amostras e fornecem informações com maior clareza sobre as equações matemáticas geradas pelo modelo. O autor utilizou o algoritmo M5P, que é uma modificação do algoritmo M5, possuindo uma etapa de suavização que garante maior confiabilidade à predição. O valor de coeficiente de correlação obtido pelo autor foi de 0.900, denotando que estratégias utilizando árvores são indicadas para este tipo de problema.

No trabalho desenvolvido por Bonifácio (2017), o autor utilizou em conjunto as técnicas de maquinas de vetores suportes para regressão (SVR) e redes neurais artificiais (RNA) para realizar a previsão das propriedades mecânicas de concreto de agregado leve. O autor obteve resultados satisfatórios com o uso das técnicas de inteligência computacional.

Nos trabalhos apresentados nesta seção é possível verificar o uso de distintas técnicas de aprendizado de máquina sendo utilizadas na previsão da resistência a compressão de diferentes tipos e composições de concreto. Vale ressaltar que em todas as pesquisas apresentadas, os resultados indicam que as técnicas utilizadas são eficientes na predição para o problema da resistência a compressão. Este fato respalda a intenção deste estudo

em verificar se o uso de técnicas que combinam a capacidade de predição destes diferentes modelos pode gerar uma melhora na predição da resistência a compressão de concretos.

### 2.3 OTIMIZAÇÃO DE PARÂMETROS DE MÉTODOS DE PREVISÃO

A escolha dos parâmetros utilizados pelas técnicas de aprendizado de máquina constituiu um ponto importante para o sucesso do modelo a ser construído. A determinação dos valores destes parâmetros não é uma tarefa trivial, uma vez que as mudanças em um determinado parâmetro não geram alterações claras na resposta final do modelo. O uso de técnicas que automatizam a escolha destes parâmetros é muito explorado. As técnicas de otimização destes parâmetros visa melhorar o desempenho preditivo dos algoritmos.

No trabalho desenvolvido por Lerman (1980), o autor apresenta uma das técnicas mais simples utilizadas na otimização de parâmetros, a busca em grade ou *grid-search*. A técnica consiste em uma busca exaustiva em um conjunto de parâmetros a serem testados. Esta técnica pode acarretar em um alto custo computacional se o número de parâmetros a serem otimizados for elevado e com um grande intervalo de possibilidades. Uma outra desvantagem desta técnica é que os parâmetros só podem ser representados por valores discretos, o que algumas das vezes restringe a busca realizada pelo algoritmo.

O uso de técnicas bio-inspiradas permite sanar alguns dos problemas que a busca em grade possui. Estas técnicas são inspiradas em comportamentos de agentes naturais que possuem o que pode ser chamado de inteligência coletiva ou inteligência de enxames. O PSO (*Particle Swarm Optimization*), é um algoritmo baseado em enxames de partículas (EBERHART; KENNEDY, 1995) e consiste em uma das técnicas de otimização bio-inspirada mais explorada.

No trabalho de Gilan, Jovein e Ramezaniyanpour (2012) foi utilizado o PSO em conjunto com o SVM para prever a resistência a compressão de misturas de concreto contendo metacaulim. Foi realizada a comparação do modelo híbrido PSO-SVM com um modelo baseado em um sistema de inferência neural-fuzzy adaptativo (ANFIS) visando verificar se o modelo híbrido possui uma melhor capacidade de previsão. Os resultados obtidos pelos autores indicam que o uso do modelo híbrido PSO-SVM, garante maior capacidade de previsão da resistência a compressão das misturas de concreto.

Já no trabalho de Qi, Fourie e Chen (2018) o PSO foi utilizado para otimizar os parâmetros da arquitetura de uma rede neural artificial (RNA). Neste trabalho o PSO em conjunto com a RNA foram utilizados para a previsão da resistência a compressão de pasta de cimento utilizada no projeto de estacas. Os resultados alcançados pelos autores indicam que o uso do PSO em conjunto com a RNA garantiu uma boa qualidade de previsão uma vez que os valores previstos pelo modelo se aproximaram dos valores experimentais.

O uso do PSO no trabalho de Chatterjee *et al.* (2017) difere-se dos apresentados anteriormente pelo fato que neste trabalho o autor utilizou-o para otimizar os pesos sinápticos da rede. Neste trabalho o modelo foi utilizado para prever falhas estruturais em edifícios de vários andares construídos com concreto armado. O modelo proposto foi comparado a outros modelos de rede neural, e apresentou resultados mais satisfatórios que os demais modelos.

Neste trabalho será utilizado o PSO como algoritmo de otimização dos parâmetros dos métodos de aprendizado de máquina que serão utilizados. Esta escolha pode ser justificada com base nos resultados dos trabalhos apresentados anteriormente e por muitos outros presentes na literatura. Um outro ponto chave para a escolha do PSO é sua simplicidade de implementação, o que garante boa replicabilidade da metodologia apresentada neste trabalho.

## 2.4 ENSEMBLE

O uso de técnicas de aprendizado de máquina para modelar problemas complexos já é uma metodologia consolidada. Uma vez que cada uma das técnicas apresenta diferentes características, a resposta a modelagem de um mesmo problema pode ser diferente de técnica para técnica. Além da diferença entre as técnicas, modelos gerados por uma mesma técnica podem apresentar diferenças entre si. Visando ampliar a capacidade de previsão, uma metodologia que vem sendo cada vez mais explorada é a utilização de diferentes algoritmos de aprendizado de máquina para realizar uma previsão conjunta, técnica denominada *Ensemble*.

Existem diferentes técnicas de *Ensemble* que podem ser utilizadas como:

- a) *Bagging*: ou ensacamento, que consistem geralmente em treinar modelos homogêneos paralelamente e combinar os seus resultados através de uma técnica determinística para gerar a previsão final;
- b) *Boosting*: ou reforço, treina modelos homogêneo sequencialmente. A soma ponderada dos modelos treinados resulta em um modelo final que engloba as características tratadas em cada modelo individualmente;
- c) *Stacking*: ou empilhamento, é uma técnica que utiliza um conjunto de modelos normalmente heterogêneos, treinados paralelamente e um meta-modelo que é treinado com o conjunto de previsões realizadas pelos modelos de primeira camada e realiza a previsão final do problema.

Neste trabalho daremos enfoque ao uso do *Stacking*. Esta técnica foi apresentada por Wolpert (1992), onde é definida como um recurso para reduzir a taxa de erro de

generalização de modelos individuais. O uso do *Stacking* em pesquisa voltada para a previsão de propriedades mecânicas do concreto foi apresentado por Chou *et al.* (2014) onde se buscava a previsão da resistência a compressão. O autor realizou a comparação da capacidade de previsão individual de técnicas como a rede neural de perceptron multicamada (MLP), máquina de vetores de suporte (SVM), árvore de classificação e regressão (CART) e regressão linear (LR), com o *Stacking* formado por conjuntos das técnicas citadas. Ao final da pesquisa foi constatado que o modelo que tinha na primeira camada os algoritmos, MLP, CART, SVM e LR e como meta-modelo o SVM foi o que apresentou os melhores resultados de previsão.



### 3 MATERIAIS E MÉTODOS

Serão apresentados neste capítulo as bases de dados que serão utilizadas, assim como os métodos computacionais utilizados e as estratégias que forma propostas.

#### 3.1 BASE DE DADOS

No presente trabalho serão utilizadas quatro bases de dados, visando analisar a capacidade de previsão do modelo proposto. As bases de dados utilizadas são conjunto de dados experimentais de diferentes tipos de misturas de concreto. A seguir serão apresentadas as bases utilizadas neste trabalho.

##### 3.1.1 Base de dados 1

Esta base de dados foi obtida do trabalho desenvolvido por Lim, Yoon e Kim (2004). A base dados contém informações extraídas de 104 amostras de concreto de alto desempenho. Os corpos de prova cilíndricos (100 x 200 mm) foram retirados dos moldes depois de 24h e foram curados durante 28 dias e então submetidos a teste para verificar a resistência a compressão. Na Tabela 1 são apresentados os parâmetros das misturas e os seus respectivos valores máximos e mínimos e também os valores para resistência a compressão.

Tabela 1 – Componentes da Mistura Base de Dados 1

Parâmetros	Mínimo	Máximo
Água/cimento (a/c)(%)	30	45
Água (W) (kg/m <sup>3</sup> )	160	180
Agregado miúdo (s/a) (%)	37	53
Cinzas volantes (FA) (%)	0	20
Superplastificante (SP) (kg/m <sup>3</sup> )	1.89	8.5
Incorporador de Ar (AE) (kg/m <sup>3</sup> )	36	78
Resistência à compressão (MPa)	38	74

Fonte: Elaborado pelo autor (2020)

Na Tabela 2 são apresentados os valores do coeficiente de correlação de Pearson ( $\rho$  de Pearson), que apresenta a correlação entre os parâmetros da mistura e a resistência à compressão. Quanto mais próximo de um maior o grau de correlação e quando negativo apresenta um correlação inversa.

No Gráfico 1 podemos observar o gráfico de associação entre a quantidade de Superplastificante (SP), parâmetro que apresentou maior  $\rho$  de Pearson, e a resistência à

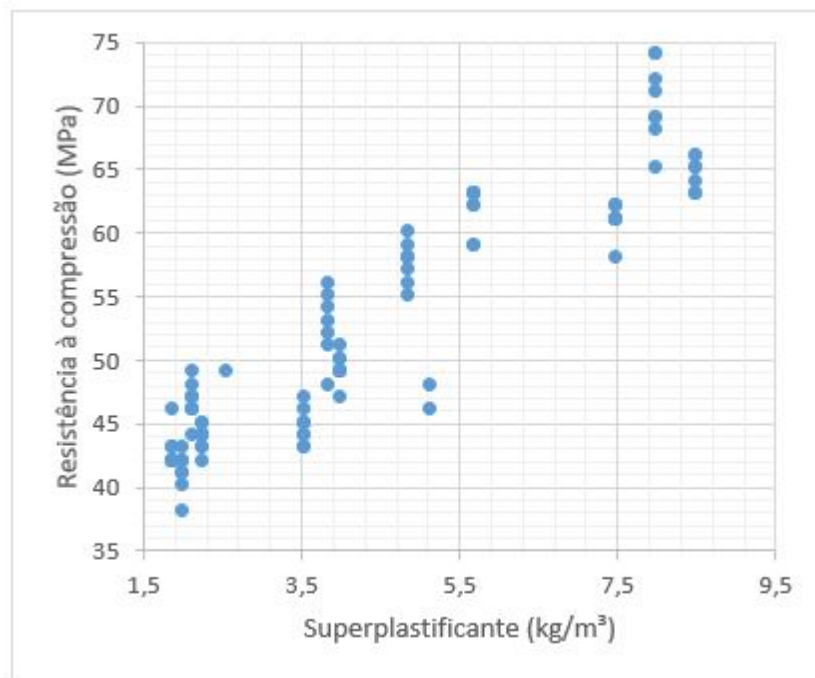
Tabela 2 – Coeficiente de correlação de Pearson: Base de Dados 1

Parâmetros	a/c	W	s/a	FA	SP	AE
$\rho$ de Pearson	-0.909	-0.286	-0.333	-0.068	0.922	0.841

Fonte: Elaborado pelo autor (2020)

compressão. Vale ressaltar que mesmo com um alto correlação não é possível estabelecer o valor da resistência à compressão do concreto utilizando apenas a quantidade de Superplastificante na mistura, uma vez que a participação dos outros componentes influencia nesta propriedade.

Gráfico 1 - Associação superplastificante e resistência à compressão.



Fonte: Elaborado pelo autor (2020)

### 3.1.2 Base de dados 2

A segunda base de dados foi retirada do trabalho de Lam, Wong e Poon (1998). O intuito do trabalho era investigar a influência das cinzas volantes e sílica ativa na resistência a compressão do concreto de alto desempenho. Foram desenvolvidas 24 misturas diferentes, variando a quantidade dos componentes apresentados na Tabela 3. Os teste de compressão ocorreram em seis diferentes períodos de cura do corpos de prova: 3, 7, 28, 56, 90 e 180 dias. No total foram realizados 144 testes, compondo assim a base de dados utilizada

neste trabalho. Os coeficientes  $\rho$  de Pearson referentes aos parâmetros da Base de Dados 2 podem ser observados na Tabela 4.

Tabela 3 – Componentes da Mistura Base de Dados 2

Parâmetros	Mínimo	Máximo
Água (W) (kg/m <sup>3</sup> )	150	205
Agregado miúdo (sa) (kg/m <sup>3</sup> )	536	724
Cinzas Volante (FA) (%)	0	55
Sílica Ativa (SF) (%)	0	5
Materiais cimenticos totais (TCM) (kg/m <sup>3</sup> )	400	500
Agregado graúdo (ca) (kg/m <sup>3</sup> )	1086	1157
Mistura redutora de água a base de naftaleno (HRWRA) (L/m <sup>3</sup> )	0	13
Dias de cura (D)	3	180
Resistência à compressão (MPa)	24	107.8

Fonte: Elaborado pelo autor (2020)

Tabela 4 – Coeficiente de correlação de Pearson Base de Dados 2

Parâmetros	W	sa	FA	SF	TCM	ca	HRWRA	D
$\rho$ de Pearson	-0.435	0.335	-0.188	0.513	0.463	-0.386	0.583	0.467

Fonte: Elaborado pelo autor (2020)

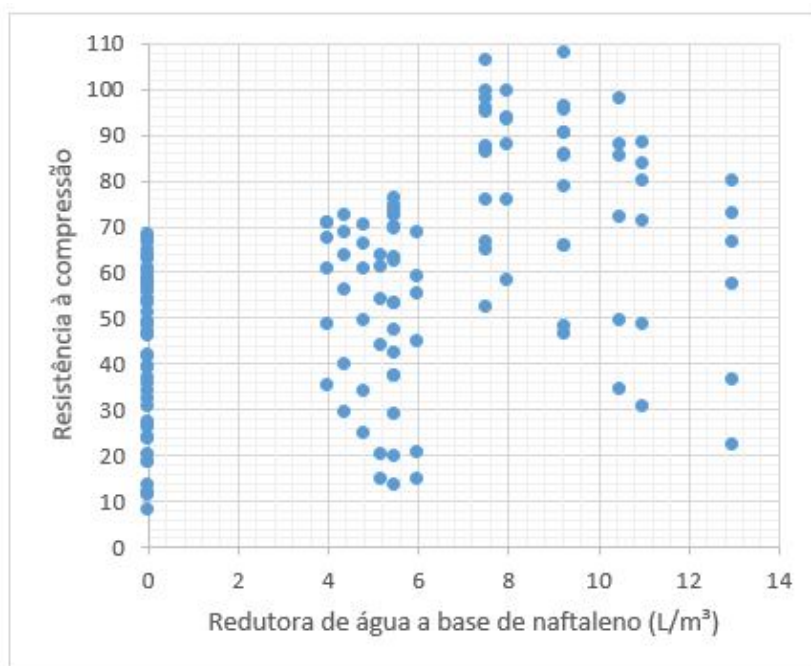
O componente da mistura que apresenta maior coeficiente de correlação é a mistura redutora de água a base de naftaleno, que é um Superplastificante assim como o apresentado na Base de dados 1. No Gráfico 2 podemos observar o gráfico de associação deste componente com a resistência à compressão.

### 3.1.3 Base de dados 3

A Base de Dados 3 foi retirada do trabalho de Siddique, Aggarwal P. e Aggarwal Y. (2011). O trabalho buscava investigar a capacidade de redes neurais artificiais na predição da resistência à compressão de concreto auto compactável contendo cinzas volantes. Os dados obtidos neste trabalho foram coletados pelo autor na literatura. Foram utilizadas bases de dados compatíveis em relação aos materiais presentes nas misturas. A Base de Dados 3 contém 80 amostras e os componentes das misturas estão apresentados na Tabela 5.

Na Tabela 6 são apresentados os  $\rho$  de Pearson referentes aos componentes das misturas da Base de dados 3 e a resistência à compressão.

Gráfico 2 - Associação Redutor de água e resistência à compressão.



Fonte: Elaborado pelo autor (2020)

Tabela 5 – Componentes da Mistura Base de Dados 3

Parâmetros	Mínimo	Máximo
Cimento ( $\text{kg}/\text{m}^3$ )	160	427
Cinzas volantes (FA) ( $\text{kg}/\text{m}^3$ )	0	261
Água/Cimento (a/c)	0.33	0.87
Superplastificante (SP) (%)	0	1
Areia ( $\text{kg}/\text{m}^3$ )	478	1079
Agregado graúdo (CA) ( $\text{kg}/\text{m}^3$ )	621	923
Resistência à compressão (MPa)	10.2	73.5

Fonte: Elaborado pelo autor (2020)

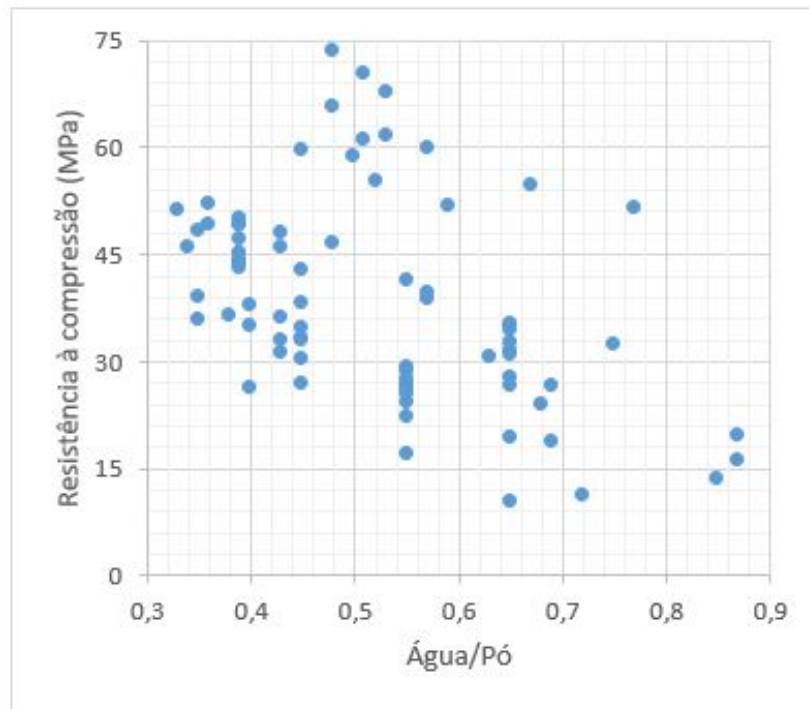
Tabela 6 – Parâmetros Base de Dados 3

Parâmetros	Cimento	FA	a/c	SP	Areia	CA
$\rho$ de Pearson	0.288	0.214	-0.466	-0.442	0.375	-0.027

Fonte: Elaborado pelo autor (2020)

O coeficiente  $\rho$  de Pearson de maior valor é o da relação Água/Cimento da mistura, no Gráfico 3 podemos observar o gráfico de associação entre este componente da mistura e a resistência à compressão.

Gráfico 3 - Associação Água/Cimento e resistência à compressão.



Fonte: Elaborado pelo autor (2020)

#### 3.1.4 Base de dados 4

Esta base de dados foi extraída do trabalho desenvolvido por Ke (2008). O estudo visava entender a influência do tipo de agregado nas propriedades mecânicas de concretos de agregado leve. A Base de dados conta com 75 amostras, cada amostra é composta por quatro parâmetros da mistura e a resistência à compressão com 28 dias. Os parâmetros da base de dados são apresentados na Tabela 7, bem como os seus valores máximos e mínimos. Os  $\rho$  de Pearson são apresentados na Tabela 8.

Tabela 7 – Componentes da Mistura Base de Dados 4

Parâmetros	Mínimo	Máximo
Água/Cimento (W/B) (%)	29.0	44.6
Quantidade de cimento (B) (kg/m <sup>3</sup> )	415	826
Volume de Agregado (VA) (%)	0	45
Peso específico do agregado (AD) (kg/m <sup>3</sup> )	3.18x10 <sup>6</sup>	3.60x10 <sup>7</sup>
Resistência à compressão (MPa)	24.9	86

Fonte: Elaborado pelo autor (2020)

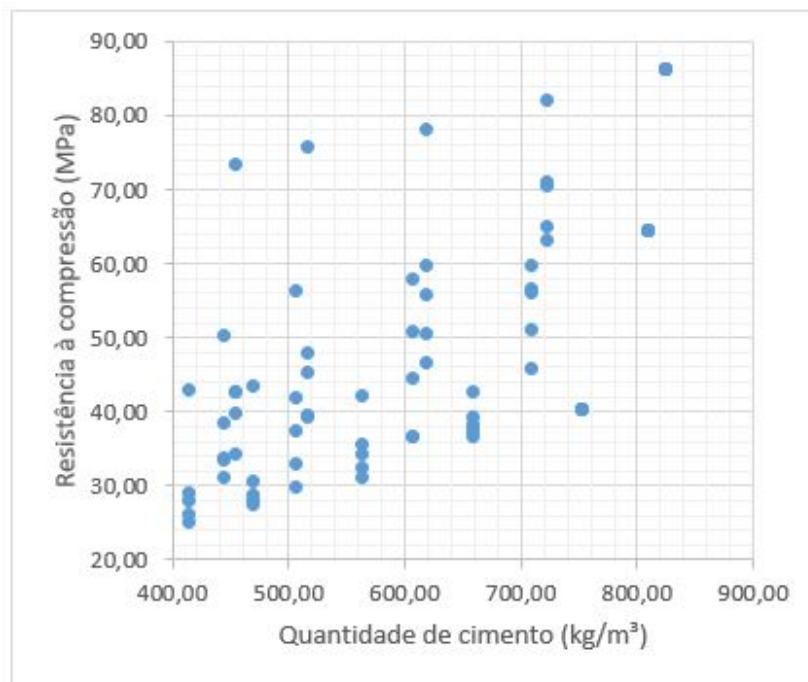
Tabela 8 – Coeficiente de correlação de Pearson Base de Dados 4

Parâmetros	W/B	B	VA	AD
$\rho$ de Pearson	-0.641	0.662	-0.552	0.341

Fonte: Elaborado pelo autor (2020)

No Gráfico 4 podemos observar o gráfico de associação entre a quantidade de cimento e a resistência à compressão, uma vez que estes dois parâmetros apresentam o maior valor de coeficiente de correlação.

Gráfico 4 - Associação Quantidade de cimento e resistência à compressão.



Fonte: Elaborado pelo autor (2020)

### 3.2 MÉTODOS DE REGRESSÃO

Neste trabalho, como já anteriormente citado, serão utilizados algoritmos de aprendizado de máquina para auxiliar na previsão de característica mecânica de misturas de concreto. Estes algoritmos estão inseridos na grande área da Inteligência Computacional. Os algoritmos pertencentes a este campo de investigação possuem a capacidade de através da análise de um conjunto de dados criar modelos de previsão.

Para que os algoritmos sejam capazes de criar os modelos de previsão, os mesmos possuem uma etapa em comum, o treinamento. Nesta etapa é onde os dados são apresentados aos algoritmos e os mesmos realizam o processo de aprendizado, extraindo informações dos dados, ajustando os parâmetros dos modelos e testando a eficácia do processo de previsão. Esta metodologia de aprendizagem é denominada de supervisionada, uma vez que são apresentados durante o processo as amostras e os resultados esperados.

As bases de dados utilizadas neste trabalho podem ser classificadas como contínuas quanto às características da variável de saída (a resistência à compressão). Esta característica induz que sejam utilizados métodos de regressão. Estes métodos buscam encontrar uma função  $f(X)$  contínua que tenha como imagem o conjunto dos resultados esperados para as amostras apresentadas durante o treinamento. O processo de treinamento dos métodos de regressão busca resolver um problema de minimização do erro,  $E(X, Y) = f(X) - Y$ , onde  $E$  representa o erro entre a previsão realizada pelo modelo e o resultado esperado.

No desenvolvimento deste trabalho foram utilizados cinco métodos de regressão baseados em inteligência computacional. Os métodos de inteligência computacional foram escolhidos devido à vasta utilização dos mesmos na modelagem de problemas semelhantes ao deste trabalho e à disponibilidade dos mesmos em bibliotecas de aprendizado de máquina em diferentes linguagens de programação. A seguir serão apresentados os métodos.

### 3.2.1 Máquina de Vetores Suporte (SVM)

As máquinas de Vetores Suporte (*Support Vector Machines*, SVM) são uma técnica de aprendizado de máquina desenvolvida por Vapnik em 1995 (VAPNIK, 1999). A técnica se baseia na teoria do aprendizado estático, que tem como objetivo a redução do erro de generalização (VAPNIK, 1998). O SVM vem sendo utilizado em diferentes campos de atuação como o reconhecimento de imagens, categorização de textos e bioinformática, apresentando resultados comparáveis ou algumas vezes superiores a técnicas como Redes Neurais Artificiais. Esse fato pode ser justificado devido à capacidade do SVM em lidar com conjuntos de dados de grandes dimensões. Neste trabalho será utilizado o algoritmo de Vetores Suporte para Regressão (Support Vector Regression, SVR). Essa versão do SVM é capaz de trabalhar com valores contínuos para a saída.

O SVR parte do princípio de uma máquina linear que tem como objetivo mapear os valores de entrada enquanto minimiza o erro de generalização. Esta máquina é definida por

$$\hat{y} = (w \cdot x) + b \quad (3.1)$$

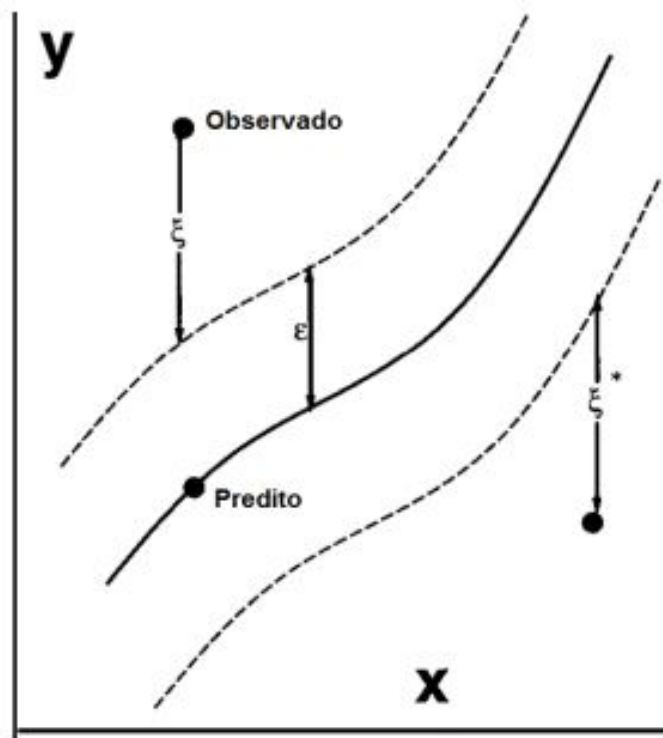
onde  $\hat{y}$  é a previsão baseada na entrada  $x$ ,  $b$  representa a bias e  $(\cdot)$  é o produto interno.

No problema de otimização do SVR são incluídas duas variáveis de folga ( $\xi$  e  $\xi^*$ ), a formulação do problema é

$$\begin{aligned} & \text{mim} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{sujeito} \begin{cases} y_i - \hat{y}_i \leq \varepsilon + \xi_i^* \\ \hat{y}_i - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n \end{cases} \end{aligned} \quad (3.2)$$

o parâmetro  $C$  é utilizado para regular a folga ( $\xi$ ) tolerada. A escolha de  $C$  também influi na complexidade da obtenção do modelo. Na Figura 3 podemos observar uma entrada  $x$  e os parâmetros do SVR.

Figura 3 - Exemplo de uma curva de regressão treinada pelo SVR



Fonte: (DRUCKER *et al.*, 1997)

O problema de regressão pode ser resolvido na sua forma dual, onde  $w$  pode ser substituído por  $w = \sum_{i,j=1}^n (\alpha_i + \alpha_i^*) x_i$ . A saída do SVR é dada pela equação 3.3.

$$\hat{y}(x) = \sum_{i,j=1}^n (\alpha_i + \alpha_i^*) (x_i \cdot x) + b \quad (3.3)$$



Para realizar a regressão não linear é definida uma função  $K(x, x') = \phi(x) \cdot \phi(x')$ , onde  $\phi(x)$  é uma transformação não linear. Esta função é chamada de Kernel. Desta forma a saída do SVR não linear será dada pela equação 3.4.

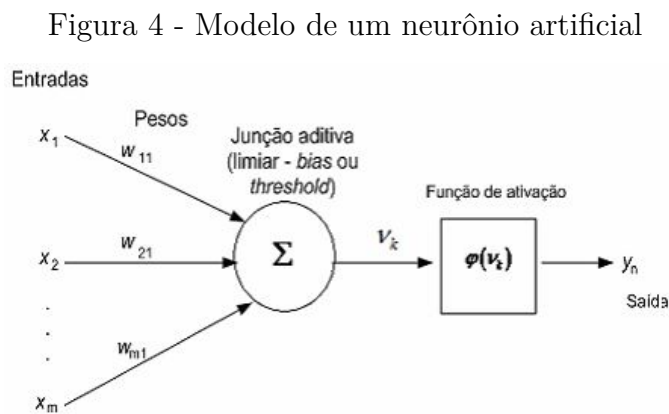
$$\hat{y}(x) = \sum_{i,j=1}^n (\alpha_i + \alpha_i^*) K(x_i \cdot x) + b \quad (3.4)$$

Neste trabalho a função Kernel utilizada será  $K(x_i, x) = e^{-\gamma|x-x'|^2}$ , que é uma função de base radial (RBF). O parâmetro  $\gamma$  é um coeficiente a ser definido para o kernel, neste trabalho será definido através de um processo de otimização.

### 3.2.2 Rede Neural Artificial (RNA)

As Redes Neurais Artificiais são algoritmos de aprendizado de máquina baseados no sistema neural do cérebro de animais. As RNA's são capazes de aprender, generalizar e organizar dados (KRÖSE *et al.*, 1993).

A menor unidade de uma RNA são os neurônios artificiais. A união de diversos neurônios em camadas forma uma rede. Nesta rede os neurônios se comunicam através das sinapses. A Figura 4 apresenta um modelo de um neurônio artificial.



Fonte: (HAYKIN, 1999)

O processo de aprendizado de uma RNA consiste no ajuste dos pesos ( $w_i$ ) buscando a minimização do erro de predição da rede. Existem na literatura vários algoritmos capazes de realizar o treinamento das redes. Neste trabalho iremos utilizar o L-BFGS, que é um algoritmo de otimização, com capacidade de solucionar problemas com um grande número de variáveis controlando a quantidade de memória utilizada.

As Redes Neurais Perceptron Multicamadas (MLP, *Multilayer Perceptron*) são redes que possuem no mínimo uma camada interna. Este tipo de rede é muito utilizado, pois

como pode possuir mais de uma camada interna é indicada para resolver problemas não lineares. Neste trabalho serão utilizadas redes que possuem até três camadas internas, cada uma sendo composta por até cem neurônios. A topologia que será utilizada é *feed-forward*, unidirecional simples.

A saída de cada neurônio é obtida através da função de ativação. Neste trabalho serão utilizadas três funções de ativação, apresentadas na Tabela 9.

Tabela 9 – Funções de ativação - MLP

Nome	Função
Linear	$\phi(v) = v$
Linear retificada	$\phi(v) = \max(0, x)$
Logística	$\phi(v) = \frac{1}{1+e^{-v}}$

Fonte: Elaborado pelo autor (2020)

### 3.2.3 Máquina de Aprendizado Extremo (ELM)

A máquina de Aprendizado Extremo (ELM, *Extreme Learning Machine*) nada mais é que uma rede MLP com apenas uma camada interna. O que difere este tipo de rede das demais é sua estratégia de ajuste dos pesos (ZHU *et al.*, 2005). Nesta estratégia os pesos da camada interna são atribuídos aleatoriamente, enquanto os pesos da camada de saída são obtidos de forma analítica.

A saída da ELM é definida por

$$\hat{y}(x) = \sum_{i=1}^N H(x)w_i \quad (3.5)$$

onde  $N$  é o número de neurônios da camada oculta,  $w$  os pesos da camada de saída e  $H(x)$  é igual à saída da função de ativação dos neurônios da camada oculta. Neste trabalho serão utilizadas as funções de ativação descritas na Tabela 10.

No processo de treinamento da ELM os parâmetros,  $a$ , referente aos pesos da camada oculta, e  $b$ , referente a *bias*, são definidos de forma aleatória, o que garante uma redução significativa no tempo de treinamento desse tipo de rede. Já os pesos  $w$  são definidos pela minimização do erro, conforme a equação 3.6.

$$\min \|H(x)w - y(x)\| \quad (3.6)$$

que é encontrada solucionando o problema  $Hw=y$ , escrito na forma matricial na Equação 3.7.

$$\begin{bmatrix} h_1(x_1) & \cdots & h_N(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_n) & \cdots & h_N(x_n) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad (3.7)$$

Tabela 10 – Funções de ativação - MLP

Nome	Função
Linear	$H(x) = ax + b$
Linear retificada	$H(x) = \begin{cases} 0 & \text{se } x < 0 \\ ax + b & \text{se } x \geq 0 \end{cases}$
Logística	$H(x) = \frac{1}{1+e^{-(ax+b)}}$
Gaussiana	$H(x) = e^{-(ax+b)^2}$
Multiquádrica	$H(x) = (\ a - x\  + b^2)^{\frac{1}{2}}$
Multiquádrica inversa	$H(x) = 1/(\ a - x\  + b^2)^{\frac{1}{2}}$

Fonte: Elaborado pelo autor (2020)

Os pesos que minimizam o erro são encontrados resolvendo  $w = H^\dagger y$ , onde  $H^\dagger$  é a matriz generalizada inversa de Moore-Penrose.

### 3.2.4 Árvore de decisão (DT)

As Árvores de Decisão (DT, *Decision Tree*) são algoritmos de aprendizado de máquina capazes de gerar sistemas especialistas para solucionar problemas de classificação (discretos) e regressão (contínuos).

As Árvores de Decisão são construídas a partir de um conjunto de testes que são realizados nos dados de entrada. Os nós internos da Árvore são as representações dos testes, no caso de problemas de regressão estes testes são quantitativos, realizados com o valor de saída definido para o conjunto de atributos em comparação a um valor que gera uma divisão. Os testes são normalmente realizados verificando se um valor é maior ou menor que o valor de divisão. Os nós folhas da Árvore armazenam um valor de retorno, referente a uma dada entrada. O valor de retorno de um nó folha é definido pela média dos valores de saída referentes a todos os conjuntos de dados de teste que chegaram até aquele nó.

Neste trabalho o algoritmo utilizado na indução das Árvores é o CART (BREIMAN *et al.*, 1984). O CART induz árvores binárias utilizando as variáveis de entrada e um limiar, objetivando conseguir o maior ganho de informação em cada nó, reduzindo assim a impureza do nó. Nos problemas de regressão a impureza de um nó é definida por uma função que deverá ter seu valor minimizado. No caso do CART a função é o erro quadrático

médio (MSE) que é definido por

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (3.8)$$

O algoritmo CART decide sobre quais as divisões serão criadas e também sobre a forma (topologia) da árvore (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Sabendo que os dados de entrada são constituídos de  $j$  variáveis e uma saída, ou seja,  $(x_i, y_i)$   $i = 1, 2, \dots, N$ , com  $x_i = (v_{i1}, v_{i2}, \dots, v_{ij})$ , que serão geradas  $K$  partições em regiões,  $R_1, R_2, \dots, R_K$  e que a saída será dada por uma constante  $c_m$  em cada região, podemos escrever a saída como

$$\hat{y}(x) = \sum_{m=1}^k c_m I(y|x \in R_m). \quad (3.9)$$

Como o critério de minimização adotado é o erro quadrático médio, podemos verificar que o  $c_m$  ótimo é a média dos  $y_i$  pertencentes à região  $R_m$ .

$$c_m = \text{média}(y|x \in R_m). \quad (3.10)$$

Então para obter a variável de divisão  $h$  e o ponto de divisão  $s$  é resolvido o problema de minimização dado pela equação 3.11.

$$\min_{j,s} \left[ \min_{c_1} \frac{1}{N} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \frac{1}{N} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (3.11)$$

### 3.2.5 **K-vizinhos mais próximos (KNN)**

O método  $K$ -vizinhos mais próximos (KNN, *K-nearest neighbors*) é um algoritmo baseado em instância, que realiza a previsão através da comparação entre conjuntos de atributos que possuem saídas semelhantes (AHA; KIBLER; ALBERT, 1991). O KNN parte do princípio que um conjunto de dados que é próximo no espaço de atributos também será próximo no espaço de respostas. Partindo deste ponto o algoritmo busca encontrar um número  $k$  de vizinhos de um conjunto de atributos e através do valor da variável de resposta realizar a previsão.

Para encontrar os  $k$ -vizinhos mais próximos são utilizadas diferentes estratégias, dentre as quais a mais simples é a distância euclidiana. Neste trabalho serão utilizados algoritmos baseados em árvores de decisão como o *KD-Tree* e *BallTree*, que podem otimizar o processo de busca dos vizinhos. Definido o conjunto de vizinhos mais próximos a previsão para um conjunto de atributos pode ser dada pela média do valor de resposta dos  $k$ -vizinhos mais próximos

$$\hat{y} = \frac{\sum_{i=1}^k f(X_k)}{k} \quad (3.12)$$

onde  $(X_1, X_2, \dots, X_k)$  é o conjunto de atributos dos  $k$ -vizinhos mais próximos e  $f(X_i)$  é o valor de resposta de cada conjunto de atributos.

Para realizar uma ponderação nos valores de resposta do vizinhos, pode ser adicionado na Equação 3.12 um peso ( $w_i$ ) que terá como objetivo valorizar os vizinhos mais próximos.

$$\hat{y} = \frac{\sum_{i=1}^k w_i f(X_k)}{k}. \quad (3.13)$$

Uma forma de definir o valor do peso é calculando o inverso da distância entre as instâncias comparadas. Neste trabalho os pesos  $w_i$  serão uniformes para todos os  $k$ -vizinhos.

### 3.3 STACKING

Técnicas que combinam a capacidade de predição de modelos, gerados por algoritmos de aprendizagem, visando alcançar melhores resultados do que os modelos individualmente, são conhecidas como *ensemble*.

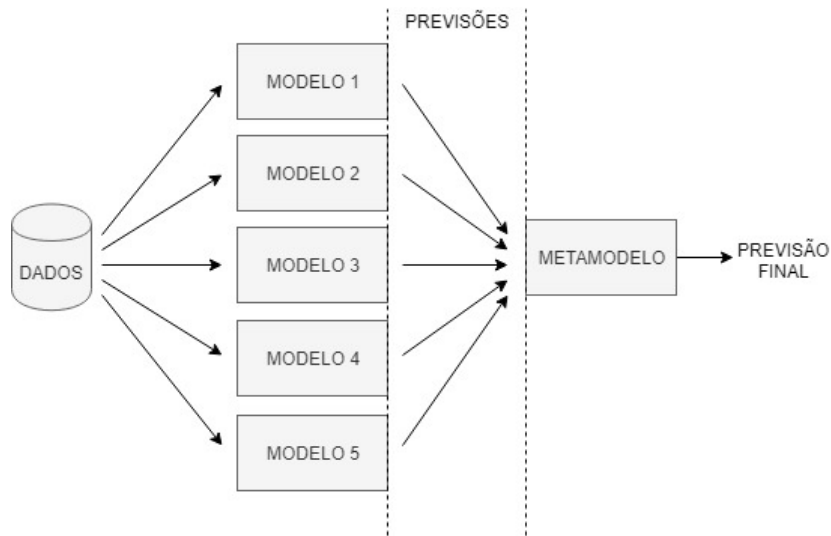
Existem diferentes técnicas de ensemble, que podem ser divididas em categorias quanto a: seus objetivos na melhoria dos resultados de predição, estratégias utilizadas para combinação dos resultados individuais e quanto aos tipos de algoritmos individuais utilizados. Esta última categoria pode ser subdividida entre combinações heterogêneas e homogêneas. Técnicas como *bagging* (BREIMAN *et al.*, 1984) e *boosting* (FREUND, 1995) são exemplos de estratégias de combinação que utilizam algoritmos homogêneos. Neste trabalho a estratégia de *ensemble* que será utilizada será o *stacking* que utiliza algoritmos heterogêneos.

O *stacking* foi definido por Wolpert (1992), quando propôs uma arquitetura de camadas (Figura 5), onde os algoritmos pertencentes ao primeiro nível (nível-0) seriam treinados com um conjunto de amostras e gerariam predições que seriam utilizadas como conjunto de treinamento para o algoritmo do segundo nível (nível-1), também chamado de metamodelo, que por fim geraria as predições finais.

O objetivo do *stacking* é reduzir o erro de generalização através do uso de modelos em cascata. Parte-se da premissa que cada modelo de forma individual é menos capaz de realizar uma boa predição do que o conjunto destes modelos.

Uma vantagem, ou no mínimo diferença, entre o *stacking* e outras técnicas de ensemble é a forma como ocorre a combinação entre as predições dos modelos de nível-0. Enquanto técnicas como o *bagging* e *boosting* utilizam formas mais simples para realizar a combinação, como por exemplo o uso da média das previsões individuais como previsão

Figura 5 - Arquitetura de camadas – Stacking



Fonte: Elaborado pelo autor (2020)

final, o stacking utiliza modelos mais robustos para realizar a previsão final, como exemplo a regressão linear ou até mesmo algoritmos de aprendizagem.

Neste trabalho será utilizada a regressão linear como metamodelo. Justifica-se o uso deste algoritmo por se tratar de uma técnica onde é possível avaliar com maior clareza a participação das predições dos modelos de nível-0 na predição final.

### 3.4 MÉTODO DE OTIMIZAÇÃO

Os métodos de otimização são técnicas que buscam otimizar a solução de um problema avaliado através de uma função objetivo. Existem diferentes métodos de otimização na literatura, os quais podem ser classificados em métodos exatos e aproximados. Os métodos exatos são capazes de encontrar a solução ótima de um problema. Porém alguns problemas são complexos e não podem ter suas soluções ótimas obtidas facilmente, então são utilizados os métodos aproximados, que são capazes de encontrar soluções próximas às ótimas. Os métodos aproximados, também chamados de heurísticos, são inspirados em fenômenos naturais como, o voo de pássaros, evolução de espécies, dentre outros.

Um dos algoritmos de otimização encontrado na literatura é o baseado em enxames de partículas (PSO, *Particle Swarm Optimization*). O PSO é um dos algoritmos de otimização evolucionários mais simples. Foi desenvolvido por Eberhart e Kennedy (1995) e ainda é amplamente utilizado atualmente.

O PSO é inspirado no comportamento de conjuntos de animais, como por exemplo

os pássaros. O algoritmo é estocástico e se baseia em um conceito muito simples: cada indivíduo do conjunto se desloca sobre o espaço de busca com uma velocidade que é ajustada através das suas experiências e das experiências do conjunto. Neste cenário podemos definir conjunto de indivíduos como a população e definir que cada indivíduo representa uma solução para o problema de minimização. O que define se a posição do indivíduo é boa, mais próximo da ótima, é o seu valor avaliado pela função objetivo.

A função objetivo é definida de acordo com o problema a ser solucionado. Em problemas de otimização de parâmetros de um modelo de regressão pode ser utilizada uma função de erro referente à saída do modelo. Para delimitar o espaço de busca são definidos valores máximos e mínimos de acordo com o problema a ser solucionado.

Os indivíduos da população são dotados de uma memória que registra a melhor posição do indivíduo, chamada de  $pBest$ , e de uma memória coletiva que registra a melhor posição já atingida pelo conjunto de indivíduos, chamada de  $gBest$ . Estas características permitem que os indivíduos avaliem sua próxima posição dentro do espaço de busca visando alcançar pontos que representem melhores soluções para o problema. O processo de aprendizado do algoritmo é definido por esta capacidade de memória dos indivíduos, o que permite mudar a velocidade ( $v$ ) e a direção ( $x$ ).

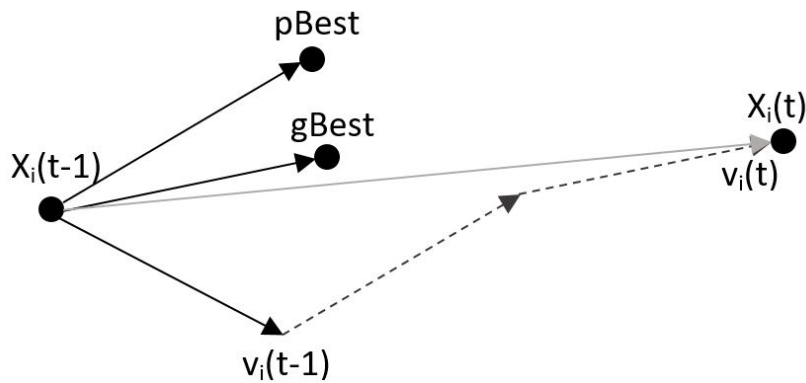
$$v_i(t) = v_i(t - 1) + \phi_p \cdot r_1(x_{pBest} - x_i(t - 1)) + \phi_g \cdot r_2(x_{gBest} - x_i(t - 1)) \quad (3.14)$$

$$x_i(t) = x_i(t - 1) + v_i(t) \quad (3.15)$$

Na Equação 3.14 as constantes  $\phi_p$  e  $\phi_g$  representam a taxa de velocidade na direção da melhor posição individual e na melhor posição global, respectivamente. As variáveis  $r_1$  e  $r_2$  são introduzidas na Equação 3.14 para gerar aleatoriedade no processo de aprendizagem do algoritmo. As variáveis são geradas aleatoriamente e ficam compreendidas no intervalo de 0 a 1. Esta aleatoriedade garante uma exploração mais completa do espaço de busca. Na Figura 6 pode ser observado o deslocamento de um indivíduo no espaço de busca.

O PSO pode ser utilizado para espaços de buscas discretos, numéricos ou não numéricos, como por exemplo no caso da otimização dos parâmetros de uma rede neural MLP, onde podem ser otimizados o número de camadas ocultas e a função de ativação. Nos casos de parâmetros discretos numéricos a resposta do PSO é convertida para um valor inteiro, porém nos casos de parâmetros discretos não numéricos uma estratégia deve ser utilizada. Uma destas estratégias é atribuir um valor inteiro a cada uma das opções parâmetro e desta forma utilizar o mesmo procedimento utilizado para parâmetros discretos numéricos para definir a opção.

Figura 6 - Atualização de posição -  $\phi_p$ ,  $\phi_g$ ,  $r_1$  e  $r_2$  igual a 1



Fonte: Elaborado pelo autor (2020)

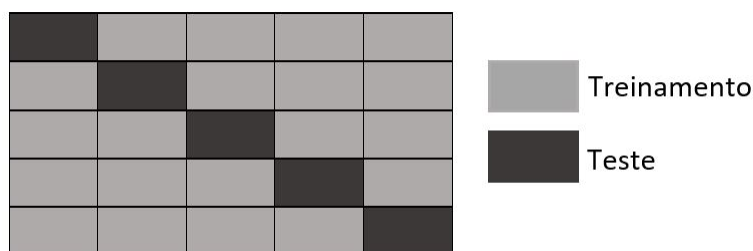
O algoritmo pode ter como critério de parada o número máximo de interações ou um valor de tolerância para a variação do valor da função objetivo. Neste trabalho iremos utilizar o número máximo de interações como critério de parada.

### 3.5 VALIDAÇÃO CRUZADA

O método de validação cruzada é uma técnica utilizada para avaliar o desempenho de estimadores utilizando a totalidade dos dados do conjunto. Este método é utilizado para minimizar os erros de previsão ocasionados por *overfitting* e também é indicado para validação de bases que possuem um número reduzido de dados amostrais.

A técnica de validação cruzada que será utilizada é o *k-fold*, que consiste na divisão da base de dados em  $k$  conjuntos de tamanhos iguais, e realiza o ajuste do modelo em  $k-1$  conjuntos e a validação no conjunto restante.

Figura 7 - Validação cruzada -  $k=5$



Fonte: Elaborado pelo autor (2020)



Este processo é realizado  $k$  vezes e modelo e validado para cada parte do conjunto de dados. A Figura 7 apresenta uma divisão de um conjunto de dados em  $k=5$  partes.

O ajuste do parâmetro  $k$  deve ser realizado de forma adequada para não impactar negativamente no resultado final. O valor de  $k$  é normalmente escolhido entre 5 e 10. A escolha deste valor depende diretamente do tamanho da base de dados que será utilizada, uma vez que a escolha de um valor grande para  $k$  pode gerar um conjunto de treinamento que não reúna todas as características da base de dados.



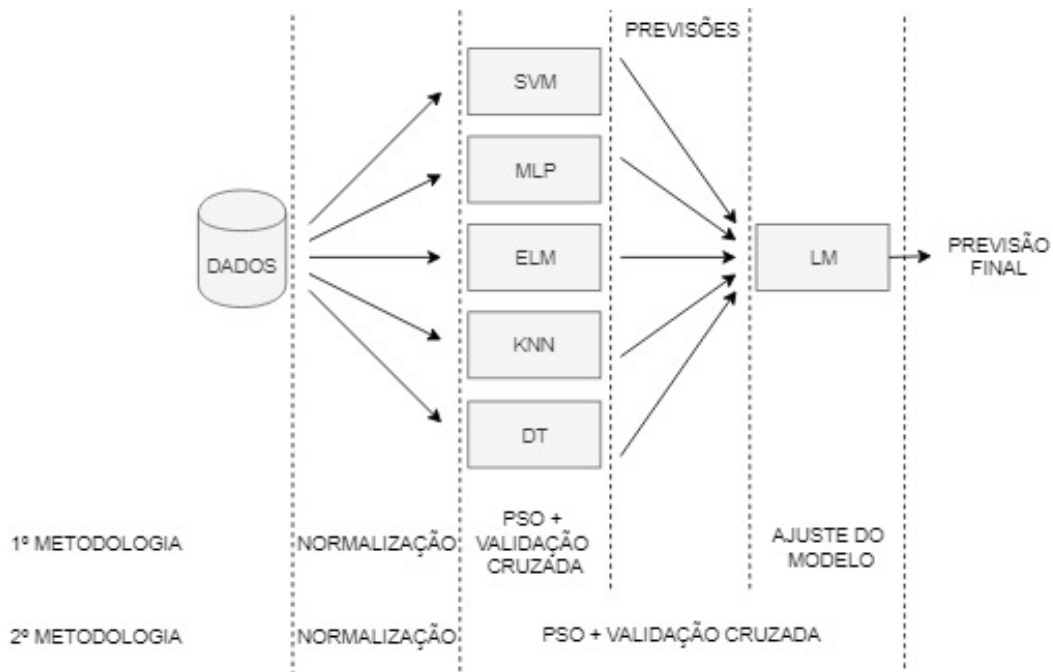
## 4 METODOLOGIA

Para o desenvolvimento deste trabalho serão propostas duas metodologias para que serão comparadas em relação a sua capacidade de previsão da resistência a compressão do concreto.

### 4.1 METODOLOGIAS PROPOSTAS

Serão propostas duas metodologias para realizar a previsão da resistência à compressão do concreto. A topologia utilizada pelo *Stacking* nas duas metodologias será a de dois níveis. O primeiro nível do *Stacking* será composto de cinco métodos de inteligência computacional, apresentados na Seção 3.2, e o segundo nível será o metamodelo, um algoritmo de regressão linear. Um modelo gráfico da topologia utilizada é apresentado na Figura 8.

Figura 8 - Topologia do *Stacking* - Metodologias



Fonte: Elaborado pelo autor (2020)

A diferença principal entre as duas metodologias está na forma de utilização do PSO para otimização dos parâmetros dos modelos de primeiro nível e o ajuste dos meta-parâmetros da regressão linear. Como observado na Figura 12 na primeira metodologia os modelos da primeira camada serão otimizados individualmente utilizando o PSO, as previsões dos modelos otimizados servirão como dados de entrada para o metamodelo que

terá seu parâmetros ajustados pelo método dos mínimos quadrados, já na segunda metodologia os modelos de primeira camada serão otimizados juntamente com os parâmetros dos metamodelos, tendo como função objetivo da otimização o erro médio quadrático da previsão final.

#### 4.1.1 Primeira metodologia

Na primeira metodologia o algoritmo de otimização, PSO, é executado para cada método da primeira camada. A otimização dos parâmetros do método nesta metodologia, tem como objetivo a minimização do erro médio quadrático de cada modelo de forma individual.

Os parâmetros utilizados pelo PSO são apresentados na Tabela 11.

Tabela 11 – Parâmetros PSO - Primeira metodologia

Parâmetros	Valor				
População	SVM	ELM	MLP	KNN	DT
	40	20	15	15	15
Nº máximo de iterações	35				
$\omega$	0,6				
$\phi_g$	0,5				
$\phi_p$	0,5				
Função objetivo	MSE				

Fonte: Elaborado pelo autor (2020)

Cada método presente na primeira camada do *Stacking* possui um conjunto de parâmetros cujo ajuste influencia diretamente na qualidade de previsão do método. Na Tabela 12 são apresentados os parâmetros que serão otimizados pelo PSO em cada método. Os valores apresentados na Tabela 12 foram obtidos da literatura e de experiências prévias.

Durante o processo de obtenção dos parâmetros otimizados, os modelos de primeira camada são treinados utilizando a estratégia de validação cruzada. O uso desta estratégia se justifica principalmente devido ao número reduzido de amostras nas bases de dados estudadas.

O metamodelo como já citado anteriormente é um método de regressão linear, ajustado pela abordagem dos mínimos quadrados. O modelo é ajustado sobre as previsões realizadas pelos modelos do primeiro nível, obtidos através do processo de otimização.

#### 4.1.2 Segunda metodologia

Nesta segunda metodologia, a forma de otimização dos parâmetros dos modelos de primeira camada e dos meta parâmetros do modelo de regressão linear, metamodelo,

Tabela 12 – Parâmetros a serem ajustados pelo PSO

Modelo	Parâmetro	Variação
SVM	Precisão da regressão ( $\epsilon$ )	$[10^{-6}, 10^{-4}]$
	Parâmetro de regularização (C)	$[1, 10^2]$
	Coefficiente do Kernel ( $\gamma$ )	$[1^{-2}, 10]$
MLP	Taxa de aprendizagem ( $\eta$ )	$[10^{-5}, 1]$
	Termo de regularização( $\alpha$ )	$[10^{-5}, 10^{-2}]$
	Função de ativação	Linear, Linear retificada e Logística
	Nº de Camadas ocultas	1,2 ou 3
	Nº de neurônios em cada camada	$[1, 100]$
ELM	Nº de neurônios na camada oculta	$[1,50]$
	Função de ativação	Linear, Linear retificada, Logística, Gaussiana, Multiquádrica e Inversa Multiquádrica
DT	Profundidade máxima	$[5,50]$
	Nº mínimo para divisão de nó	$[2,3]$
	Nº mínimo de amostras por folha	$[1,4]$
KNN	Nº de vizinhos	$[3,10]$

Fonte: Elaborado pelo autor (2020)

difere-se da primeira metodologia, pois são realizadas em conjunto.

A otimização nesta metodologia tem como objetivo garantir a minimização do erro médio quadrático da predição final realizada pelo metamodelo.

O algoritmo de otimização, PSO, é aplicado no *Stacking*, buscando otimizar os parâmetros dos modelos do primeiro nível e ajustar o metamodelo. Os parâmetros do PSO para esta metodologia são apresentado na Tabela 13.

Tabela 13 – Parâmetros PSO - Segunda metodologia

Parâmetros	Valor
População	100
Nº máximo de interações	70
$\omega$	0,6
$\phi_g$	0,7
$\phi_p$	0,7
Função objetivo	MSE

Fonte: Elaborado pelo autor (2020)

A saída do meta modelo, e por consequência do *Stacking*, é dada pela Equação 4.1.

$$\hat{y}_{Stacking} = \left( \sum_{i=1}^N \alpha_i \hat{y}_i \right) + b \quad (4.1)$$

onde  $N$  é o número de modelos do primeiro nível,  $\hat{y}_i$  são as previsões dos modelos do primeiro nível,  $\alpha_i$  são os meta parâmetros do metamodelo de regressão linear e  $b$  é a *bias*.

No processo de otimização os parâmetros  $\alpha_i$  podem assumir valores entre 0 e 1. Já o  $b$  pode assumir valores entre -10 e 10. Os parâmetros dos métodos de primeira camada podem assumir valores de acordo com a Tabela 12. O PSO nesta estratégia possui um espaço de busca de vinte e duas dimensões, o que justifica o aumento do tamanho da população e do número máximo de interações.

Visando possibilitar a análise da influência da previsão de cada modelo de primeiro nível no resultado final do *Stacking*, foi utilizada uma restrição no PSO (Equação 4.2) para que a soma dos parâmetros  $\alpha_i$  seja 1 com uma tolerância ( $t$ ) de 0,05.

$$1 - t \leq \left( \sum_{i=1}^N \alpha_i \right) \leq 1 + t \quad (4.2)$$

## 4.2 MÉTRICAS DE AVALIAÇÃO

Para avaliar a qualidade de predição das metodologias utilizadas neste trabalho serão usadas métricas de avaliação estatísticas baseadas no erro de predição. As métricas podem ser definidas de acordo com o tipo de problema a ser avaliado. Os resultados das métricas são interpretados, seguindo regras preestabelecidas, para concluir a qualidade de um método.

No presente trabalho serão utilizadas três métricas, que têm seu uso difundido na literatura, são elas: o coeficiente de determinação ( $R^2$ ), a raiz do erro médio quadrático (RMSE, *Root Mean Squared Error*) e o erro percentual absoluto médio (MAPE, *Mean Percentual Error*). A escolha destas métricas se justifica pela sua fácil interpretação e pela possibilidade de comparação com outros trabalhos semelhantes.

Estabelecendo  $\hat{y}$  como a saída estimada,  $y$  como o rótulo das amostras,  $\bar{y}$  como a média dos rótulos das amostras e  $N$  o número de amostras, as métricas são definidas por:

O coeficiente de determinação ( $R^2$ ) é definido pela equação 4.3.

$$R^2(y, \hat{y}) = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4.3)$$

onde ( $R^2$ ) varia de 0 a 1, sendo que quanto mais próximo de 1 melhor a qualidade de generalização do modelo;

O RMSE é obtido através da equação 4.4.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4.4)$$

O RMSE é definido na mesma unidade que a saída do método, o que facilita a sua interpretação. A métrica pode ser definida como o desvio padrão entre os valores previstos e amostrais.

O MAPE é definido pela equação 4.5.

$$MAPE(y, \hat{y}) = \frac{1}{N} \left( \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \right) 100 \quad (4.5)$$

O valor de MAPE é dado em porcentagem, o que facilita a comparação da qualidade de predição do método, mesmo quando se trabalha com problemas distintos.

### 4.3 TESTES ESTATÍSTICOS

Foram utilizados testes estatísticos com o objetivo principal de identificar a existência de semelhança significativa entre os modelos de primeira camada e os resultados do *Stacking*.

Para a determinação da existência de semelhança entre três ou mais grupos são utilizados métodos paramétricos ou não paramétricos, dependendo se as amostras apresentam normalidade na sua distribuição. O teste de Shapiro-Wilk é utilizado para responder se a hipótese de normalidade é verdadeira para uma dada amostra (GHASEMI; ZAHEDIASL, 2012). O teste apresenta como resultado um p-valor que indica se a amostra pode ser considerada como oriunda de uma distribuição normal. Caso o p-valor seja menor que 0,05 podemos dizer que a hipótese de normalidade foi rejeitada.

O teste de Lilliefors, também é utilizado para indicar a normalidade de uma dada amostra. Este teste é adaptado do teste de Kolmogorov-Smirnov e possui a mesma estatística deste (WHITNALL; OSWALD; MATHER, 2011), que é a diferença máxima entre a função de distribuição empírica e a função de distribuição cumulativa teórica. A hipótese nula para este teste é que a amostra provém de uma distribuição normal, que é confirmada caso o p-valor, resultado do teste, seja maior que 0,05.

O uso de testes paramétricos ou não paramétricos será definido a partir dos resultados dos testes de Shapiro-Wilk e Lilliefors. Os testes paramétricos são utilizados para os casos onde as amostras são originadas de distribuições que apresentam normalidade, permitindo a inferência sobre parâmetros que caracterizam a distribuição de origem da amostra. O uso dos testes não paramétricos se dá quando as distribuições de origem das amostras não são determinadas, neste caso é feita inferência sobre o centro da distribuição.

O teste não paramétrico utilizado no decorrer deste trabalho será o de Kruskal-Wallis. O teste é utilizado para definir se existe diferença significativa entre as medianas das distribuições de dois ou mais grupos de uma variável independente, contínua ou ordinária. Este teste é uma alternativa ao uso do teste paramétrico ANOVA (VARGHA; DELANEY, 1998). O teste ANOVA tem como pré-requisito para o seu uso a normalidade e homoscedasticidade, igual distribuição de variância, das amostras. No teste de Kruskal-Wallis, como em outros testes, uma estatística é calculada e comparada com um ponto de corte, definido pelo nível de significância que normalmente é de 0,05. As hipóteses a serem testadas são:  $H_0$  as medianas das populações são iguais;  $H_1$  as medianas das populações são diferentes.

O uso de testes como o de Kruskal-Wallis e ANOVA possibilita indicar apenas se existe diferença estatisticamente significativa entre os grupos de amostras testados. A definição de quais grupos de amostras apresentam diferenças significativas entre si é obtida com o uso de testes *Post Hoc*.

O teste de Dunn é um teste estatístico *Post Hoc* não paramétrico utilizado para comparar pares de grupos de amostras e identificar se existe diferença significativa entre os pares (DINNO, 2015). Este teste é utilizado após o teste de Kruskal-Wallis apresentar a existência de diferença significativa entre os grupos comparados. O teste apresenta um p-valor para cada par de amostras comparadas, se o p-valor for menor que 0,05 indica que as amostras comparadas não apresentam semelhança significativa.

#### 4.4 CÓDIGO COMPUTACIONAL

O desenvolvimento do código computacional deste trabalho foi realizado na linguagem Python. Para realizar a implementação das técnicas de aprendizado de máquina foi utilizada a biblioteca *Scikit-learn* (PEDREGOSA *et al.*, 2011), e para a preparação e análise dos dados foi utilizada a biblioteca *Pandas* (MCKINNEY, 2010). Ambas as bibliotecas são amplamente utilizadas em trabalhos semelhantes a esse, o que credita credibilidade ao seu uso.

A rotina desenvolvida inicia-se com a leitura das bases de dados e a preparação das mesmas para serem utilizadas pelas técnicas de aprendizagem de máquina. Uma parte importante desta preparação é a normalização dos dados, processo importante pois leva a melhores resultados. Para realizar este procedimento foi estabelecido um limite mínimo e um máximo para os dados, respectivamente 0,15 e 0,85. Após a escolha dos limites, foi realizada a normalização de acordo com as Equações 4.6 e 4.7.



$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.6)$$

$$Y_n = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (4.7)$$

onde  $X_{min}$  e  $Y_{min}$  são os valores mínimos,  $X_{max}$  e  $Y_{max}$  são os valores máximos e  $X_n$  e  $Y_n$  são os valores normalizados.

Com os dados normalizados é iniciada a etapa de otimização dos parâmetros das técnicas de aprendizagem. Essa otimização é realizada pelo PSO com auxílio da técnica de validação cruzada. Obtidas as previsões dos modelos de primeira camada otimizados, é possível realizar o ajuste do metamodelo, assim obtendo a previsão final do *Stacking*. Sabendo que todo o processo descrito neste parágrafo possui uma parcela de aleatoriedade, são realizadas 35 execuções do mesmo visando garantir uma amostra significativa de resultados.

As métricas apresentadas na seção 4.2, são calculadas para os resultados das execuções do algoritmo. Os valores de média e desvio padrão de cada métrica são calculados, para possibilitar a análise da capacidade de predição de cada técnica de aprendizagem e do *Stacking*.



## 5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados das simulações das duas metodologias aplicadas às quatro base de dados. Os resultados serão apresentados através das métricas de avaliação e de gráficos que auxiliaram na análise dos resultados das métricas.

### 5.1 RESULTADOS EXPERIMENTAIS

Os resultados das métricas serão apresentados através da média e do desvio padrão para as trinta e cinco execuções realizadas. Serão apresentados os resultados da primeira metodologia aplicada às quatro base de dados e em seguida da segunda metodologia.

#### 5.1.1 Resultados da primeira metodologia

Os resultados das métricas para as simulações utilizando a primeira metodologia, Seção 4.1, referente às quatro base de dados, são apresentados nas Tabelas 14 a 17. São apresentados os valores da média e do desvio padrão (valor apresentado entre parênteses). Estão destacados nas tabelas os melhores resultados, de média, dentre os métodos de primeira camada e o *Stacking*.

Tabela 14 – Resultados métricas de avaliação - Base de dados 1 - Primeira Metodologia

Output	MAPE	RMSE	$R^2$
SVM	2.021 (0.141)	<b>1.302</b> (0.105)	<b>0.981</b> (0.003)
MLP	2.135 (0.179)	1.503 (0.130)	0.974 (0.005)
ELM	2.506 (0.199)	1.714 (0.162)	0.966 (0.007)
KNN	3.608 (0.212)	2.440 (0.163)	0.932 (0.009)
DT	3.120 (0.284)	2.286 (0.266)	0.940 (0.014)
Stacking	<b>1.957</b> (0.110)	1.306 (0.091)	<b>0.981</b> (0.003)

Fonte: Elaborado pelo autor (2020)

Observando os valores da média das métricas de avaliação pode-se perceber que o *Stacking* não apresenta o melhor resultado para todas as bases de dados, porém sempre apresenta-se como, no mínimo, o segundo melhor resultado. Quando se observa o desvio padrão a mesma situação ocorre.

Visando identificar se existe diferença significativa entre os resultados das métricas de avaliação do *Stacking* e dos demais métodos de aprendizagem de máquina serão utilizados testes estatísticos. Para definir se serão utilizados testes paramétricos ou não paramétricos, foram realizados os testes de Shapiro-Wilk e Lilliefors, que indicam se os resultados das

Tabela 15 – Resultados métricas de avaliação - Base de dados  
2 - Primeira Metodologia

Output	MAPE	RMSE	$R^2$
SVM	11.288 (0.713)	6.379 (0.462)	0.927 (0.011)
MLP	8.703 (0.553)	4.567 (0.309)	0.963 (0.005)
ELM	13.376 (0.776)	7.695 (0.763)	0.893 (0.022)
KNN	25.543 (1.440)	11.990 (0.716)	0.743 (0.031)
DT	13.730 (1.268)	7.551 (0.569)	0.898 (0.016)
Stacking	<b>8.137</b> (0.559)	<b>4.344</b> (0.279)	<b>0.966</b> (0.004)

Fonte: Elaborado pelo autor (2020)

Tabela 16 – Resultados métricas de avaliação - Base de dados  
3 - Primeira Metodologia

Output	MAPE	RMSE	$R^2$
SVM	<b>11.360</b> (0.919)	<b>5.209</b> (0.369)	<b>0.862 (0.020)</b>
MLP	13.427 (1.002)	6.159 (0.664)	0.805 (0.044)
ELM	13.354 (1.358)	6.101 (0.760)	0.808 (0.051)
KNN	22.745 (1.156)	8.337 (0.475)	0.646 (0.041)
DT	26.019 (3.875)	11.008 (1.636)	0.372 (0.200)
Stacking	11.766 (0.813)	5.343 (0.362)	0.855 (0.020)

Fonte: Elaborado pelo autor (2020)

Tabela 17 – Resultados métricas de avaliação - Base de dados  
4 - Primeira Metodologia

Output	MAPE	RMSE	$R^2$
SVM	<b>5.304</b> (0.437)	<b>3.320</b> (0.316)	<b>0.961</b> (0.008)
MLP	5.846 (0.590)	3.595 (0.346)	0.954 (0.009)
ELM	5.865 (0.494)	3.561 (0.495)	0.955 (0.014)
KNN	9.467 (0.758)	6.314 (0.642)	0.859 (0.029)
DT	9.479 (1.977)	6.520 (1.649)	0.842 (0.085)
Stacking	5.504 (0.258)	3.350 (0.310)	0.960 (0.009)

Fonte: Elaborado pelo autor (2020)

métricas de avaliação apresentam normalidade na distribuição. Nas Tabelas 18 e 19 estão apresentados os *p-valores* para os testes de normalidade.

Observando os resultados apresentados nas Tabelas 18 e 19 percebe-se que nenhuma das métricas apresentou, em todos os grupos de amostras, distribuição normal. Sabendo que as métricas não apresentam distribuição normal em suas amostras, é necessária a aplicação de um teste não paramétrico para identificar se existe diferença significativa entre os métodos. O teste utilizado será o de *Kruskal-Wallis*.

Os resultados apresentados na Tabela 20 indicam que as medianas das métricas dos

Tabela 18 – P-valores - Teste de *Shapiro-Wilk* - Primeira metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.243</b>	<b>0.458</b>	<b>0.272</b>	0.009	0.004	0.001
MLP	0.028	<b>0.061</b>	0.009	<b>0.659</b>	<b>0.197</b>	<b>0.097</b>
ELM	0.023	0.002	0.000	<b>0.930</b>	0.014	0.002
KNN	<b>0.881</b>	<b>0.296</b>	<b>0.363</b>	<b>0.605</b>	<b>0.287</b>	<b>0.141</b>
DT	<b>0.544</b>	<b>0.618</b>	<b>0.229</b>	0.000	0.001	0.000
<i>Stacking</i>	<b>0.647</b>	<b>0.560</b>	<b>0.505</b>	<b>0.127</b>	<b>0.421</b>	<b>0.187</b>
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.279</b>	<b>0.061</b>	0.017	<b>0.771</b>	0.028	0.004
MLP	<b>0.708</b>	0.003	0.000	<b>0.123</b>	<b>0.240</b>	0.042
ELM	0.003	0.000	0.000	0.002	0.000	0.000
KNN	<b>0.979</b>	<b>0.462</b>	<b>0.246</b>	<b>0.219</b>	<b>0.165</b>	0.039
DT	<b>0.084</b>	0.013	0.000	0.031	0.022	0.000
<i>Stacking</i>	<b>0.892</b>	<b>0.373</b>	<b>0.162</b>	<b>0.142</b>	0.000	0.000

Fonte: Elaborado pelo autor (2020)

Tabela 19 – P-valores - Teste de *Lilliefors* - Primeira metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.110</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	0.010	0.003
MLP	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.069</b>	<b>0.137</b>
ELM	0.047	0.049	0.015	<b>0.200</b>	<b>0.109</b>	0.028
KNN	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>
DT	<b>0.200</b>	<b>0.133</b>	0.036	0.014	0.076	0.026
<i>Stacking</i>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.049	<b>0.200</b>	<b>0.166</b>	<b>0.200</b>	<b>0.200</b>	<b>0.118</b>
MLP	<b>0.200</b>	<b>0.038</b>	<b>0.022</b>	<b>0.127</b>	<b>0.062</b>	<b>0.037</b>
ELM	0.024	0.003	0.000	0.002	0.000	0.000
KNN	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	0.039	0.012
DT	<b>0.200</b>	<b>0.200</b>	<b>0.120</b>	0.033	<b>0.124</b>	0.005
<i>Stacking</i>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	0.000	0.000

Fonte: Elaborado pelo autor (2020)

modelos de primeira camada e do *Stacking* apresentam diferença significativa, uma vez que os *p-valores* para todos os testes foram menores que 0,05. Uma vez verificado que existe

Tabela 20 – P-valores - Teste de *Kruskal-Wallis* - Primeira metodologia

Base de dados	MAPE	RMSE	$R^2$
1	<0.0001	<0.0001	<0.0001
2	<0.0001	<0.0001	<0.0001
3	<0.0001	<0.0001	<0.0001
4	<0.0001	<0.0001	<0.0001

Fonte: Elaborado pelo autor (2020)

diferença significativa entres os métodos, será utilizado o teste *Post Hoc* de Dunn com o intuito de verificar onde está localizada a diferença. Como neste trabalho o principal interesse é verificar se os resultados obtidos pelo *Stacking* apresentam diferença significativa com os modelos de primeira camada, serão apresentadas apenas estas comparações, e suprimidas as comparações entre os modelos de primeira camada.

Tabela 21 – P-valores - Teste de *Dunn* - *Stacking* x Modelos de primeira camada -Primeira metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.363</b>	<b>0.948</b>	<b>0.948</b>	0.000	0.000	0.000
MLP	0.015	0.001	0.000	<b>0.196</b>	<b>0.302</b>	<b>0.302</b>
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	0.000	0.000	0.000	0.000	0.000	0.000
DT	0.000	0.000	0.000	0.000	0.000	0.000
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.346</b>	<b>0.428</b>	<b>0.428</b>	<b>0.221</b>	<b>0.847</b>	<b>0.847</b>
MLP	0.000	0.000	0.000	<b>0.109</b>	0.023	0.023
ELM	0.002	0.001	0.001	0.047	<b>0.116</b>	<b>0.116</b>
KNN	0.000	0.000	0.000	0.000	0.000	0.000
DT	0.000	0.000	0.000	0.000	0.000	0.000

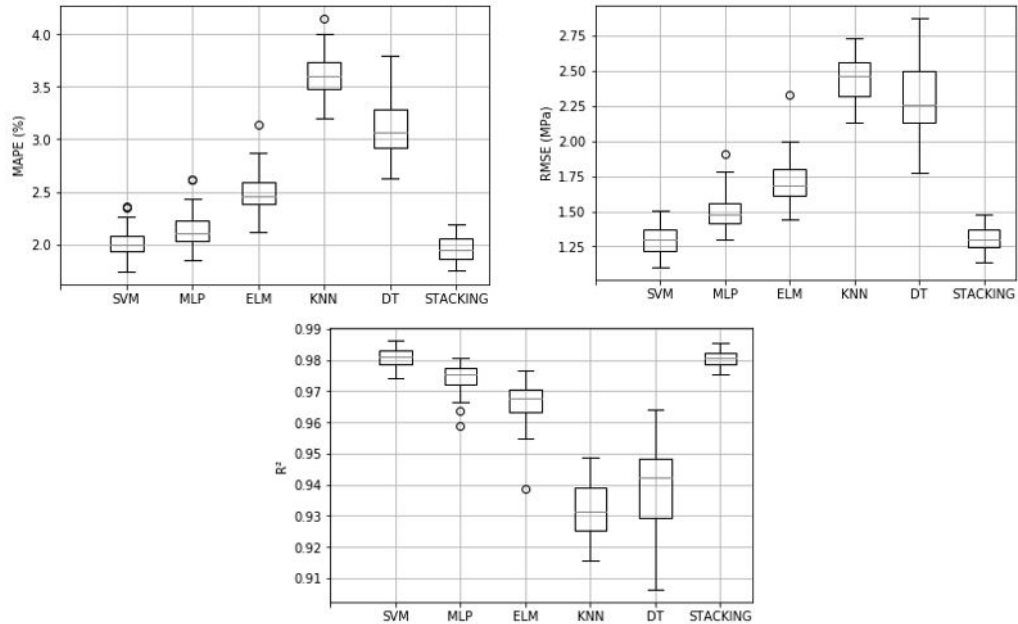
Fonte: Elaborado pelo autor (2020)

Comparando os resultados apresentados na Tabela 21 e das Tabelas 14 a 17, podemos observar que o *Stacking* apresenta resultados das métricas estatisticamente semelhantes ao modelo de primeira camada com melhores resultados.

Nas Tabelas 14 a 17 um outro ponto que deve ser observado é a variância das médias das métricas de avaliação. Observa-se que para as quatro bases de dados a variância para os resultados do *Stacking* apresenta-se como a menor ou então como a segunda menor dentre os métodos. Para auxiliar na análise da variância, podem ser observadas as Gráficos

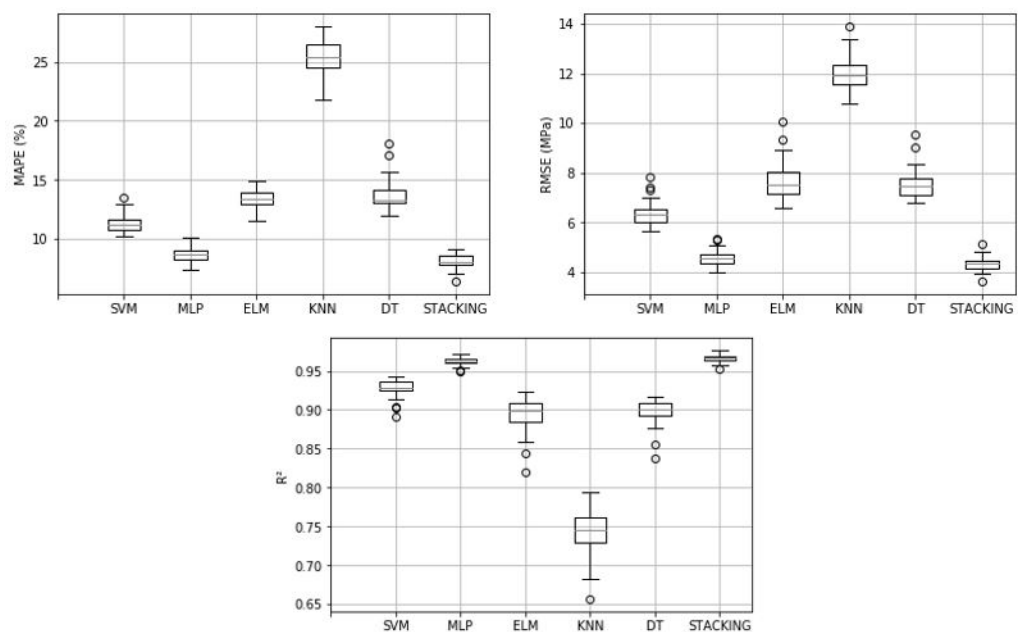
5 a 8, que apresentam gráficos de *boxplot* dos resultados das métricas para cada Base de dados.

Gráfico 5 - Boxplots Métricas - Base de dados 1 - Primeira Metodologia



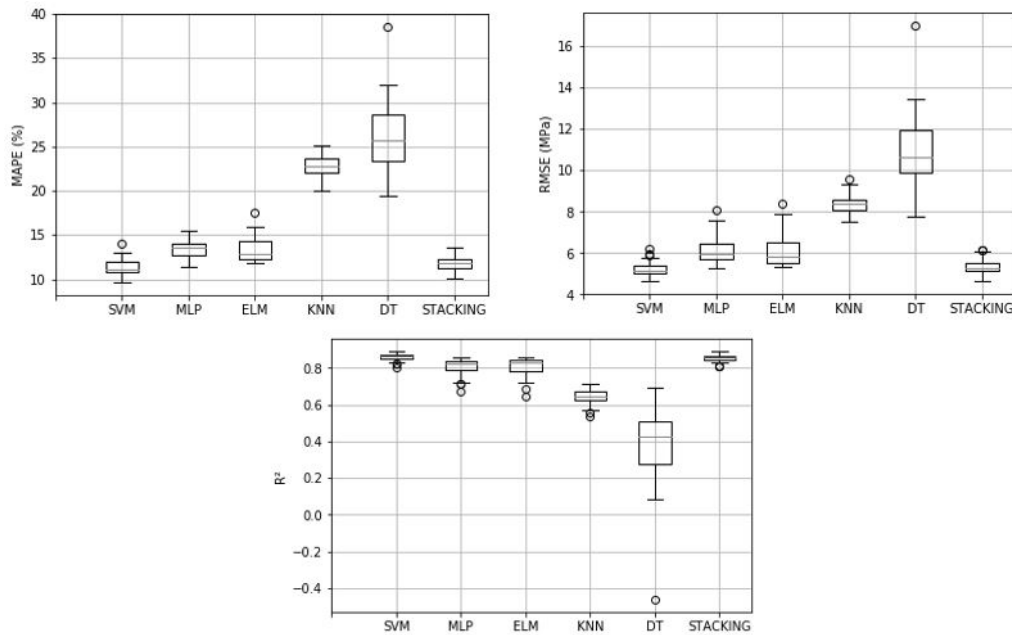
Fonte: Elaborado pelo autor (2020)

Gráfico 6 - Boxplots Métricas - Base de dados 2 - Primeira Metodologia



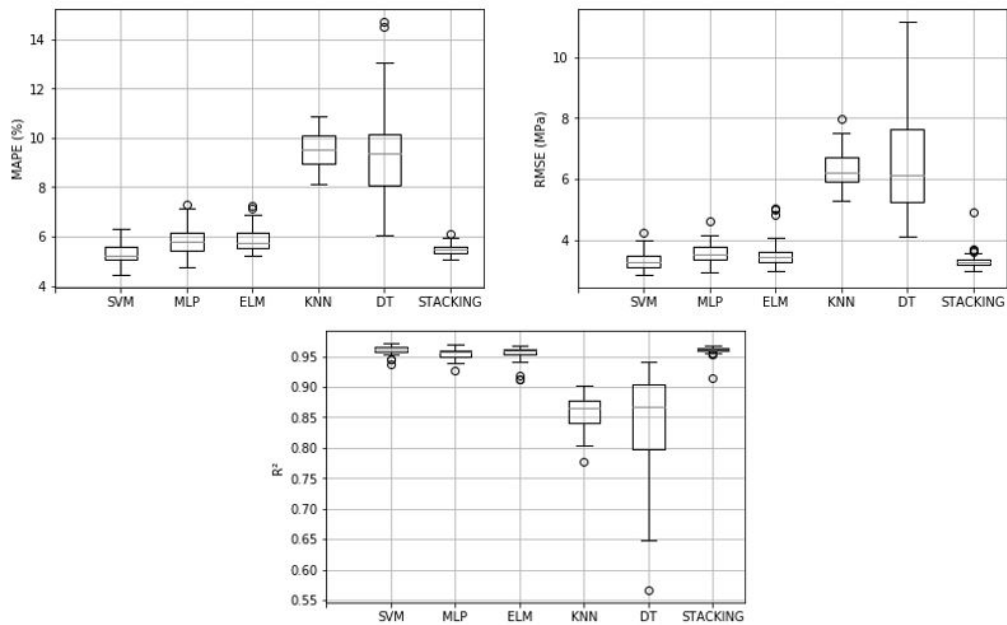
Fonte: Elaborado pelo autor (2020)

Gráfico 7 - Boxplots Métricas - Base de dados 3 - Primeira Metodologia



Fonte: Elaborado pelo autor (2020)

Gráfico 8 - Boxplots Métricas - Base de dados 4 - Primeira Metodologia



Fonte: Elaborado pelo autor (2020)

Através dos Gráficos 5 a 8 pode-se confirmar o que foi observado nos resultados: o *Stacking* apresenta baixa variância, o que garante confiabilidade de previsão ao método.



O SVM para as Bases 1, 3 e 4 é o método que apresenta variância mais próxima a do *Stacking*, já na Base 2 a MLP é o que apresenta maior proximidade de variância com o *Stacking*.

Buscando avaliar a participação de cada modelo de primeira camada na previsão final do *Stacking*, foi utilizada a regressão linear como metamodelo, como explicado anteriormente. Na Tabela 22 são apresentadas as médias dos coeficientes de regressão ( $\alpha$ ) associados a cada método de primeira camada e o termo de interceptação ( $\beta$ ), para as execuções realizadas para cada base de dados.

Tabela 22 – Coeficientes do metamodelo - Primeira metodologia

	Base 1	Base 2	Base 3	Base 4
$\alpha_{SVM}$	<u>0.582</u>	0.141	<u>0.575</u>	<u>0.327</u>
$\alpha_{MLP}$	0.235	<u>0.646</u>	0.103	<u>0.290</u>
$\alpha_{ELM}$	0.107	0.045	0.223	<u>0.318</u>
$\alpha_{KNN}$	0.032	0.006	0.083	0.012
$\alpha_{DT}$	0.041	0.173	0.048	0.046
$\beta$	0.165	-0.788	-1.537	0.386

Fonte: Elaborado pelo autor (2020)

Observando os resultados apresentados na Tabela 22 pode-se perceber que existe uma relação direta entre a qualidade da previsão individual do método de primeira camada e a sua participação na previsão feita pelo metamodelo. Tomando como exemplo os resultados para a Base de dados 1, observamos que o SVM apresenta a maior participação (58,2%) na previsão final, o que é justificado uma vez que foi o modelo com melhor resultado individual para esta base. A participação da MLP (23,5%) e a da ELM (10,7%) apresentam também proporcionalidade ao desempenho dos modelos, assim como o KNN (3,2%) e a DT (4,1%).

Os métodos KNN e DT apresentam nas quatro bases de dados baixa participação na previsão final do *Stacking*, não chegando a impactar nem em 10% em média a previsão final. Apesar dessa visão geral podemos observar que na Base de dados 2 o DT possui mais representatividade na previsão final do que o SVM e na Base de dados 3 o KNN apresenta representatividade próxima a obtida pelo MLP.

### 5.1.2 Resultados da segunda metodologia

Os resultados das métricas para as simulações utilizando a segunda metodologia, Seção 4.1, são apresentados nas Tabelas 23 a 26. A apresentação dos dados nas tabelas segue o mesmo modelo utilizado para a primeira metodologia: média para as 35 execuções e desvio padrão.

Tabela 23 – Resultados métricas de avaliação - Base de dados  
1 - Segunda Metodologia

Output	MAPE	RMSE	$R^2$
SVM	2.788 (0.378)	1.880 (0.267)	0.959 (0.012)
MLP	2.643 (0.999)	1.838 (0.740)	0.955 (0.059)
ELM	3.213 (1.557)	2.254 (1.126)	0.928 (0.111)
KNN	3.594 (0.194)	2.512 (0.149)	0.928 (0.008)
DT	3.433 (0.333)	2.556 (0.301)	0.925 (0.018)
Stacking	<b>2.349</b> (0.269)	<b>1.660</b> (0.178)	<b>0.968</b> (0.007)

Fonte: Elaborado pelo autor (2020)

Tabela 24 – Resultados métricas de avaliação - Base de dados  
2 - Segunda Metodologia

Output	MAPE	RMSE	$R^2$
SVM	14.484 (2.316)	11.713 (3.907)	0.728 (0.178)
MLP	<b>12.242</b> (7.050)	<b>6.161</b> (3.066)	<b>0.916</b> (0.110)
ELM	21.028 (12.165)	10.805 (6.991)	0.705 (0.717)
KNN	32.995 (4.378)	14.099 (1.409)	0.642 (0.069)
DT	15.388 (1.815)	8.288 (0.772)	0.877 (0.023)
Stacking	13.371 (2.749)	7.084 (1.605)	0.906 (0.047)

Fonte: Elaborado pelo autor (2020)

Tabela 25 – Resultados métricas de avaliação - Base de dados  
3 - Segunda Metodologia

Output	MAPE	RMSE	$R^2$
SVM	18.943 (3.674)	8.540 (1.967)	0.610 (0.180)
MLP	<b>15.062</b> (4.650)	6.666 (1.644)	0.761 (0.157)
ELM	20.378 (12.290)	9.590 (5.285)	0.392 (0.933)
KNN	23.523 (1.643)	8.557 (0.582)	0.627 (0.051)
DT	25.563 (2.413)	10.682 (1.119)	0.415 (0.126)
Stacking	15.255 (2.535)	<b>6.363</b> (0.775)	<b>0.792</b> (0.057)

Fonte: Elaborado pelo autor (2020)

Avaliando apenas os valores médios das métricas, apresentadas nas Tabelas 23 a 26, pode-se observar que o *Stacking* não apresenta melhor resultado em todos os casos. Porém o *Stacking* gera no mínimo o segundo melhor resultado quando comparado com os modelos de primeira camada.

Como realizado para os resultados da primeira metodologia, serão utilizados testes estatísticos com o intuito de identificar se existe diferença significativa entre os resultados das métricas de avaliação do *Stacking* e dos demais métodos de aprendizagem de máquina. Foram realizados os teste de Shapiro-Wilk e Lilliefors, para definir se seriam utilizados testes

Tabela 26 – Resultados métricas de avaliação - Base de dados 4 - Segunda Metodologia

Output	MAPE	RMSE	$R^2$
SVM	8.645 (2.143)	5.660 (1.601)	0.879 (0.066)
MLP	7.947 (3.040)	4.839 (1.757)	0.907 (0.068)
ELM	10.432 (7.790)	6.829 (4.848)	0.755 (0.444)
KNN	10.090 (1.006)	6.754 (0.632)	0.839 (0.030)
DT	12.000 (2.397)	8.046 (1.501)	0.766 (0.081)
Stacking	<b>7.627</b> (1.478)	<b>4.732</b> (1.037)	<b>0.918</b> (0.037)

Fonte: Elaborado pelo autor (2020)

paramétricos ou não paramétricos. Os resultados destes teste indicam se os resultados das métricas de avaliação apresentam normalidade na distribuição. Nas Tabelas 27 e 28 estão apresentados os *p-valores* para os testes de normalidade.

Tabela 27 – P-valores - Teste de *Shapiro-Wilk* - Segunda metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	<b>0.089</b>	<b>0.150</b>	0.016	0.000	0.030	0.001
MLP	0.000	0.000	0.000	0.000	0.000	0.000
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	<b>0.070</b>	<b>0.802</b>	<b>0.864</b>	<b>0.067</b>	<b>0.704</b>	<b>0.971</b>
DT	<b>0.785</b>	<b>0.396</b>	<b>0.211</b>	<b>0.272</b>	<b>0.153</b>	0.044
<i>Stacking</i>	0.016	0.026	0.005	<b>0.051</b>	0.001	0.000
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.001	0.013	0.001	<b>0.102</b>	0.024	0.004
MLP	0.000	0.000	0.000	0.000	0.000	0.000
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	<b>0.616</b>	<b>0.646</b>	<b>0.259</b>	<b>0.257</b>	<b>0.474</b>	<b>0.313</b>
DT	<b>0.978</b>	<b>0.071</b>	0.008	<b>0.737</b>	<b>0.194</b>	<b>0.466</b>
<i>Stacking</i>	0.015	0.000	0.000	0.015	<b>0.053</b>	0.006

Fonte: Elaborado pelo autor (2020)

Os resultados apresentados nas Tabelas 27 e 28 indicam que nenhuma das métricas apresentou distribuição normal em todas as amostras. A partir destes resultados indica-se a necessidade de aplicação de testes não paramétricos. Como realizado na primeira metodologia será aplicado o teste de *Kruskal-Wallis*, para identificar se existe diferença significativa entre os modelos de primeira camada e o *Stacking*. Na Tabela 29 são apresentados os *p-valores* para o testes.

Tabela 28 – P-valores - Teste de *Lilliefors* - Segunda metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.035	0.007	0.001	0.000	<b>0.056</b>	0.033
MLP	0.000	0.000	0.000	0.000	0.000	0.000
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	<b>0.166</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>
DT	<b>0.200</b>	0.020	0.004	<b>0.200</b>	0.014	0.004
<i>Stacking</i>	<b>0.200</b>	0.012	0.005	<b>0.164</b>	0.033	0.001
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.001	0.028	0.015	<b>0.122</b>	0.038	0.008
MLP	0.000	0.000	0.000	0.000	0.000	0.000
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.200</b>	<b>0.168</b>
DT	<b>0.200</b>	<b>0.135</b>	0.037	<b>0.200</b>	<b>0.190</b>	<b>0.200</b>
<i>Stacking</i>	<b>0.200</b>	<b>0.107</b>	<b>0.107</b>	<b>0.200</b>	<b>0.181</b>	<b>0.182</b>

Fonte: Elaborado pelo autor (2020)

Tabela 29 – P-valores - Teste de *Kruskal-Wallis* - Segunda metodologia

Base de dados	MAPE	RMSE	$R^2$
1	<0.0001	<0.0001	<0.0001
2	<0.0001	<0.0001	<0.0001
3	<0.0001	<0.0001	<0.0001
4	<0.0001	<0.0001	<0.0001

Fonte: Elaborado pelo autor (2020)

Os resultados na Tabela 29 indicam que existe diferença significativa entre os modelos de primeira camada e o *Stacking*. Sabendo que existe diferença entre os métodos faz-se necessário identificar entre quais métodos existe diferença. Para isso, como realizado na análise dos resultados da primeira metodologia, será utilizado o teste *Post Hoc* de *Dunn*. Na Tabela 30 são apresentadas as comparações entre os modelos de primeira camada e o *Stacking*.

Observando os resultados apresentados na Tabela 30 e comparando com os resultados apresentados nas Tabelas 23 a 26, verificamos que o *Stacking* apresenta semelhança significativa com os modelos de primeira camada que apresentam os melhores resultados individuais. As Bases de dados 2 e 4 apresentaram semelhança entre o *Stacking* e mais de um modelos de primeira camada, o que pode se dever aos altos valores de variância apresentados pelas métricas nesta metodologia.

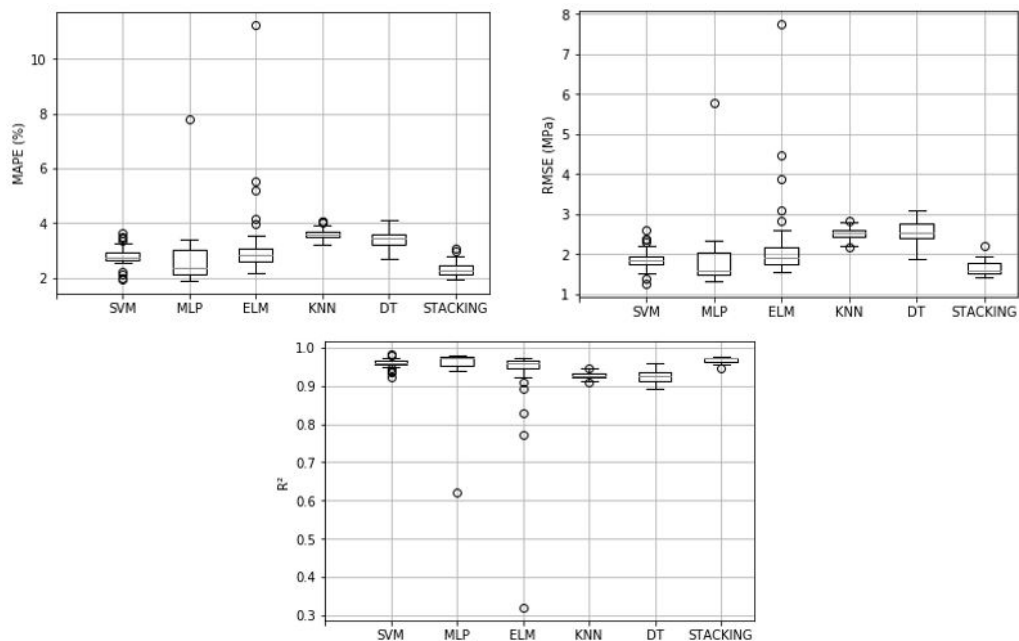
Tabela 30 – P-valores - Teste de *Dunn* - *Stacking* x Modelos de primeira camada -Segunda metodologia

Método	<i>p</i> -valor					
	Base 1			Base 2		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.002	0.015	0.015	<b>0.331</b>	0.000	0.000
MLP	<b>0.184</b>	<b>0.375</b>	<b>0.375</b>	0.041	<b>0.115</b>	<b>0.115</b>
ELM	0.000	0.000	0.000	0.000	0.000	0.000
KNN	0.000	0.000	0.000	0.000	0.000	0.000
DT	0.000	0.000	0.000	0.025	0.045	0.045
	Base 3			Base 4		
	MAPE	RMSE	$R^2$	MAPE	RMSE	$R^2$
SVM	0.008	0.000	0.000	<b>0.221</b>	<b>0.847</b>	<b>0.847</b>
MLP	<b>0.511</b>	<b>0.704</b>	<b>0.704</b>	<b>0.101</b>	<b>0.903</b>	<b>0.903</b>
ELM	0.036	0.000	0.000	<b>0.132</b>	0.043	0.043
KNN	0.000	0.000	0.000	0.000	0.000	0.000
DT	0.000	0.000	0.000	0.000	0.000	0.000

Fonte: Elaborado pelo autor (2020)

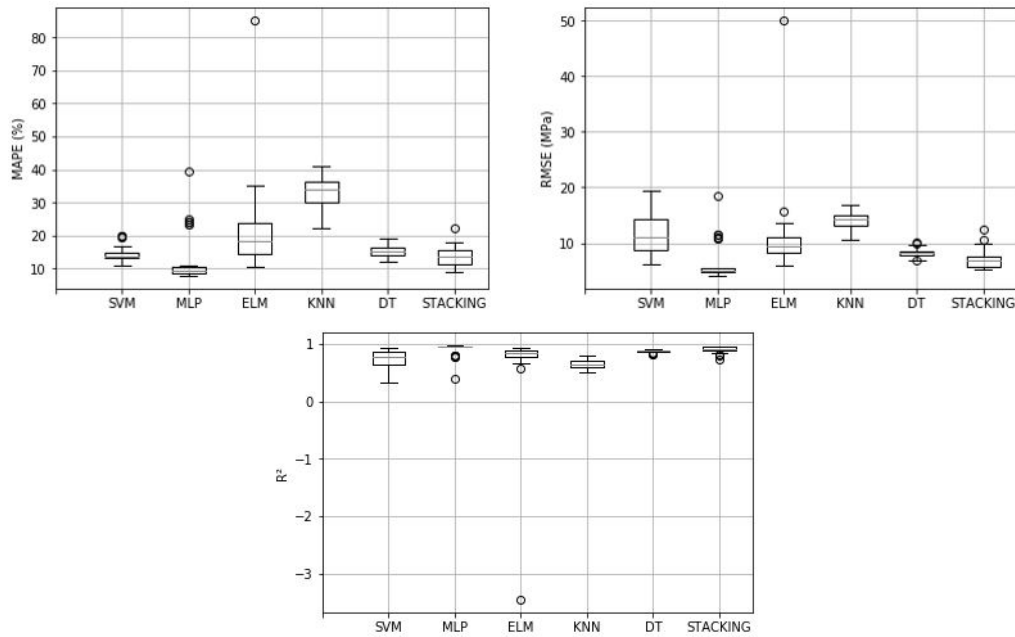
Para avaliar a variância dos resultados das métricas de avaliação para os modelos de primeira camada e para o *Stacking* foram utilizados gráficos de *boxplot*, apresentados nas Gráficos 9 a 12.

Gráfico 9 - Boxplots Métricas - Base de dados 1 - Segunda Metodologia



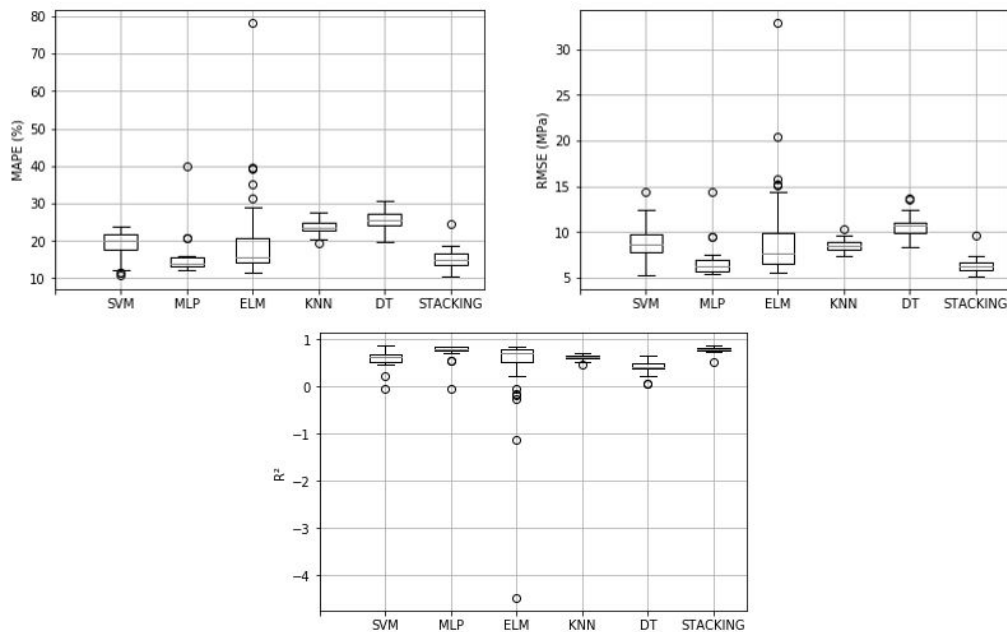
Fonte: Elaborado pelo autor (2020)

Gráfico 10 - Boxplots Métricas - Base de dados 2 - Segunda Metodologia



Fonte: Elaborado pelo autor (2020)

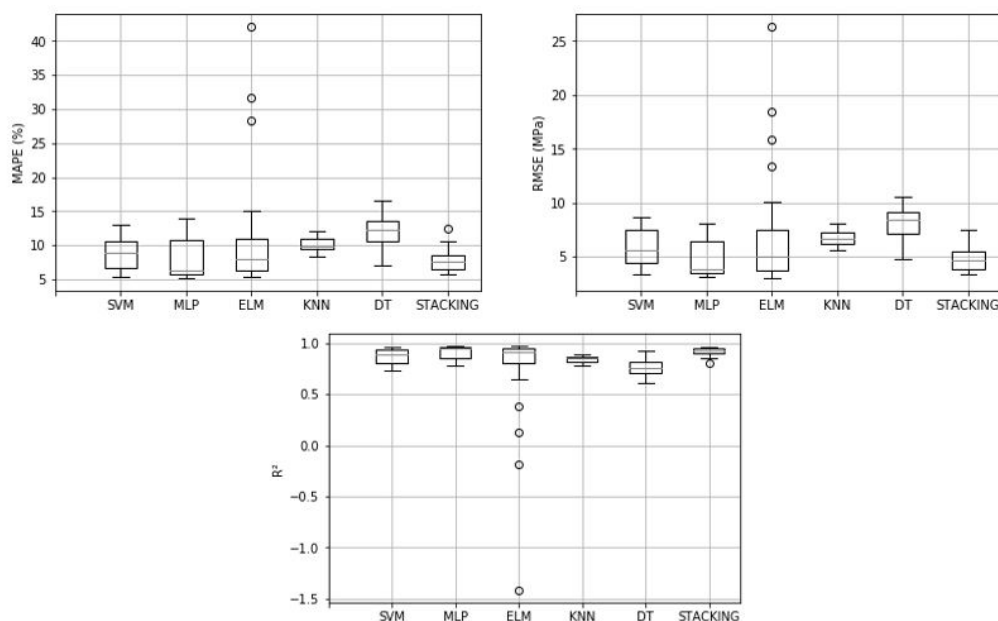
Gráfico 11 - Boxplots Métricas - Base de dados 3 - Segunda Metodologia



Fonte: Elaborado pelo autor (2020)

Nas execuções realizadas para a segunda metodologia com as quatro base de dados, quando observada a variância dos resultados das métricas de avaliação, pode-se perceber

Gráfico 12 - Boxplots Métricas - Base de dados 4 - Segunda Metodologia



Fonte: Elaborado pelo autor (2020)

que os modelos de primeira camada MLP e ELM possuem valores de variância mais elevados. Observando os Gráficos 9 a 12 percebe-se que o MLP e ELM apresentam em seus *boxplots* pontos de *outliers*, o que mostra que estes modelos apresentaram baixa precisão. As variâncias das métricas para os resultados do *Stacking* apresentam valores baixos para as execuções com as Bases 1, 3 e 4 e para a Base 2 valores mais elevados, tomando como referência a menor variância dentre os modelos. Os modelos SVM, KNN e DT apresentaram menores valores de variância.

Para realizar a análise da influência de cada método na previsão final do *Stacking* são apresentadas na Tabela 31 as médias dos coeficientes de regressão ( $\alpha$ ) associados a cada método de primeira camada e o termo de interceptação ( $\beta$ ), para as execuções realizadas para cada base de dados.

Os coeficientes do metamodelo nesta metodologia foram obtidos através da otimização utilizando o algoritmo PSO, como explicado na seção 4.1.2. Realizando a comparação entre os resultados apresentados na Tabela 31 e os resultados das Tabelas 23 a 26, podemos verificar que não existe uma relação direta entre a participação dos modelos de primeira camada, na predição do *Stacking*, e os resultados das métricas de avaliação de cada modelo. Tomando como exemplo os resultados para a Base de dados 1, quando observamos a Tabela 23 podemos perceber que o modelo SVM apresenta os melhores resultados individuais, no entanto a sua participação na previsão do *Stacking* é menor do que a da ELM, que por

Tabela 31 – Coeficientes do metamodelo - Segunda metodologia

	Base 1	Base 2	Base 3	Base 4
$\alpha_{SVM}$	0.237	0.245	0.249	0.147
$\alpha_{MLP}$	0.156	0.162	0.258	0.225
$\alpha_{ELM}$	0.243	0.211	0.133	0.189
$\alpha_{KNN}$	0.225	0.172	0.205	0.195
$\alpha_{DT}$	0.152	0.236	0.120	0.246
$\beta$	-0.004	-0.011	0.013	0.004

Fonte: Elaborado pelo autor (2020)

suas vez é apenas o terceiro melhor modelo individual quando comparamos as métricas de avaliação.

### 5.1.3 Comparação entre as metodologias propostas

Esta seção tem como intuito realizar a comparação dos resultados obtidos pelas duas metodologias, objetivando criar pontos de observações que permitam discernir sobre qual metodologia apresenta melhor resultado, seja geral ou em aspectos pontuais.

Iniciando a comparação das metodologias pela análise dos resultados das métricas de avaliação dos modelos de primeira camada e do *Stacking* utilizando os dados apresentados nas Tabelas 14 a 17 e 23 a 26, podemos observar que de forma geral os resultados apresentados pela primeira metodologia são melhores tanto em média como em desvio padrão. Existem apenas duas exceções a afirmação anterior, o modelo KNN na segunda metodologia, para simulações com a Base de dados 1, apresenta resultado da métrica MAPE melhor do que na primeira metodologia, fato semelhante ao que acontece para simulações com a Base de dados 3, onde o modelo DT apresenta melhores resultados das métricas MAPE e RMSE. Observando ainda as Tabelas 14 a 17 e 23 a 26, podemos perceber que ambas as metodologias apresentam resultados satisfatórios no uso do *Stacking*, uma vez que os resultados das métricas de avaliação para as previsões do *Stacking* ou se apresentam como os melhores, em comparação com as dos modelos individuais, ou estatisticamente iguais a do modelo de primeira camada, que apresenta os melhores resultados para as métricas de avaliação.

Um ponto que também pode ser comparado entre as metodologias é a capacidade de análise da participação dos modelos de primeira camada na previsão final do *Stacking*. A primeira metodologia permitiu visualizar que os resultados individuais dos modelos de primeira camada estavam diretamente relacionados à sua participação na previsão final do *Stacking*, uma vez que quanto melhor o resultado do modelo maior a sua participação, o que nos permite induzir que o ajuste do metamodelo na primeira metodologia foi satisfatório.



Já na segunda metodologia percebe-se uma ausência de relação direta, ou de qualquer outra relação identificável, entre os resultados individuais dos modelos de primeira camada e a sua participação no resultado final do *Stacking*. Através desta observação não se pode definir o ajuste do metamodelo como não satisfatório, mas pode ser encarado como um indicador da maior complexidade de ajuste do metamodelo na segunda metodologia.

#### 5.1.4 Comparação com outras abordagens propostas na literatura

Na Tabela 32 é apresentada uma comparação entre os resultados obtidos pelos trabalhos apresentados na Seção 2.2 e os resultados obtidos pelo *Stacking* para a Base de Dados 1 na primeira e segunda metodologia.

Tabela 32 – Comparação dos Resultados do Stacking e outros trabalhos

	MAPE	RMSE	$R^2$
Stacking 1º Metodologia	1.957	1.306	0.981
Stacking 2º Metodologia	2.349	1.660	0.968
(ABD A.; ABD M., 2017)	-	1.929	0.986
(AHMADU-NEDUSHAN, 2012)	-	1.174	0.984
(AL-SHAMIRI <i>et al.</i> , 2019)	1.818	1.134	0.994
(BEHNOOD <i>et al.</i> , 2017)	13.160	6.178	0.900

Fonte: Elaborado pelo autor (2020)

Ao analisarmos a Tabela 32 podemos perceber que os resultados obtidos pelos *Stacking* estão semelhantes aos resultados obtidos por outras metodologias. Este fato indica que o *Stacking* é uma metodologia promissora quando se trata da previsão das propriedades mecânicas do concreto. Vale ressaltar que os trabalhos utilizados para a comparação focaram no uso de apenas uma metodologia e exploraram ao máximo o potencial da mesma, o que permite criar uma hipótese, que pode ser explorada em trabalhos futuros, que ao melhorar o desempenho dos modelos de primeira camada do *Stacking* podemos melhorar os resultados apresentados pelo metamodelo da segunda camada.



## 6 CONCLUSÃO

O presente trabalho buscou investigar a capacidade do método de aprendizado em conjunto, *Stacking*, na previsão da resistência à compressão de concretos com diferentes características. O *Stacking* foi utilizado em conjunto com o algoritmo de otimização, PSO, com o propósito de otimizar o desempenho dos métodos de aprendizado de máquina presentes na primeira camada do *Stacking*, bem como melhorar o desempenho do método de regressão linear utilizado como metamodelo na segunda camada. Foram utilizados cinco diferentes métodos de aprendizado de máquina na primeira camada, as redes neurais artificiais de múltiplas camadas (MLP), máquinas de vetores suporte (SVM), máquinas de aprendizado extremo (ELM), k-vizinhos mais próximos (KNN) e árvores de decisão (DT). Os modelos foram treinados e validados usando o método de validação cruzada, k-fold. O processo de treinamento, validação e cálculo das métricas de avaliação foi realizado trinta e cinco vezes visando possibilitar uma análise estatística dos resultados.

Foram propostas neste trabalho duas metodologias para realizar a otimização dos parâmetros dos métodos de primeira camada e o ajuste dos parâmetros do metamodelo.

Os resultados alcançados na primeira metodologia indicaram um bom desempenho do *Stacking* na previsão da resistência a compressão. O *Stacking* apresentou resultados médios no MAPE (erro percentual absoluto médio) de aproximadamente 11% no pior caso, sendo este resultado para a Base de dados 3, no entanto para a Base de dados 1 apresentou resultado 2%. Quando observamos os resultados do coeficiente de correlação ( $R^2$ ) o *Stacking* apresentou resultados de aproximadamente 0,85 no pior caso, novamente para Base de dados 3, para as demais bases os valores ficaram acima de 0,95. Os testes estatísticos realizados, utilizando os resultados das métricas de avaliação, indicaram que o *Stacking* apresenta resultados tão bons quanto o melhor método de primeira camada. As análises gráficas dos *Boxplots* indicam que o *Stacking* apresenta bons resultados de desvio padrão, sendo mais exato que os métodos de primeira camada.

Para a segunda metodologia podemos observar resultados que diferem de forma quantitativa da primeira, porém do ponto de vista qualitativo as duas metodologias apresentam resultados semelhantes.

Nos resultados para a segunda metodologia o *Stacking* apresentou no pior caso resultado médio do MAPE de aproximadamente 15%, neste caso para a Base de dados 3. O melhor resultado foi de 2,4% obtido para a Base de dados 1. O resultado para o coeficiente de correlação foi no pior caso de 0,79, para a Base de dados 3, para as demais bases os resultados superaram o valor de 0,9. Para esta metodologia os resultados do *Stacking* também foram estatisticamente semelhantes aos resultados obtidos pelo melhor

modelo de primeira camada. Analisando a variância dos resultados das métricas podemos verificar o mesmo que aconteceu na primeira metodologia, o *Stacking* apresenta melhor precisão que os modelos de primeira camada.

Os modelos de primeira camada que apresentaram melhores resultados foram o SVM e a MLP, os mesmos apresentaram resultados estatisticamente semelhantes ao do *Stacking*, na primeira e segunda metodologia, respectivamente. Os resultados deste trabalho, em comparação com resultados apresentados nos artigos citados na Seção 2.2, permitem concluir que o objetivo central foi alcançado, o *Stacking* apresentou-se como uma ferramenta eficaz para realizar a previsão da resistência à compressão do concreto. O arcabouço computacional criado, formado pelo *Stacking* e o PSO, provou-se eficiente e promissor como campo de estudo a ser explorado.

Podem ser considerados como trabalhos que continuaram a explorar o proposto neste trabalho, os seguintes temas:

- a) Explorar o uso de outras técnicas de otimização, como estratégias evolutivas em substituição do algoritmo de enxame de partículas;
- b) Utilizar algoritmos de inteligência artificial mais robustos como metamodelo e como modelos de primeira camada;
- c) Aplicar um procedimento de seleção de características no processo de otimização de modo que o algoritmo evolutivo encontre também as características mais relevantes.

## REFERÊNCIAS

- ABD A., Abbas; ABD M., Suhad. Modelling the strength of lightweight foamed concrete using support vector machine (svm). **Case Studies in Construction Materials**, Elsevier, v. 6, p. 8–15, 2017.
- AHA, David W; KIBLER, Dennis; ALBERT, Marc K. Instance-based learning algorithms. **Machine learning**, Springer, v. 6, n. 1, p. 37–66, 1991.
- AHMADU-NEDUSHAN, Behrouz. An optimized instance based learning algorithm for estimation of compressive strength of concrete. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 25, n. 5, p. 1073–1081, 2012.
- AÍTICIN, PC. **Concreto de alto desempenho**. Pini, São Paulo, v. 5, p. 104–105, 2000.
- AL-SHAMIRI, Abobakr Khalil; KIM, Joong Hoon; YUAN, Tian-Feng; YOON, Young Soo. Modeling the compressive strength of high-strength concrete: An extreme learning approach. **Construction and Building Materials**, Elsevier, v. 208, p. 204–219, 2019.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 15823**: Concreto autoadensável. parte 1: Classificação, controle e recebimento no estado fresco. Rio de Janeiro. ABNT, 2010.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 8953**: Concreto para fins estruturais-classificação pela massa específica, por grupos de resistência e consistência. Rio de Janeiro. ABNT, 2015.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 5739**: Concreto - ensaio de compressão de corpos de prova cilíndricos. Rio de Janeiro. ABNT, 2018.
- BEHNOOD, Ali; BEHNOOD, Venous; GHAREHVERAN, Mahsa Modiri; ALYAMAC, Kursat Esat. Prediction of the compressive strength of normal and high-performance concretes using m5p model tree algorithm. **Construction and Building Materials**, Elsevier, v. 142, p. 199–207, 2017.
- BREIMAN, Leo; FRIEDMAN, Jerome; STONE, Charles J; OLSHEN, Richard A. **Classification and regression trees**. [S.l.]: CRC press, 1984.
- DINNO, Alexis. Nonparametric pairwise multiple comparisons in independent groups using dunn’s test. **The Stata Journal**, SAGE Publications Sage CA: Los Angeles, CA, v. 15, n. 1, p. 292–300, 2015.
- DRUCKER, Harris; BURGESS, Christopher JC; KAUFMAN, Linda; SMOLA, Alex J; VAPNIK, Vladimir. Support vector regression machines: Advances in neural information processing systems. **MIT Press Denver**, USA, p. 155–161, 1997.
- EBERHART, Russell; KENNEDY, James. A new optimizer using particle swarm theory. **Proceedings of the Sixth International Symposium on Micro Machine and Human Science**, p. 39–43, 1995.
- FREUND, Yoav. Boosting a weak learning algorithm by majority. **Information and computation**, Elsevier, v. 121, n. 2, p. 256–285, 1995.

GAO, Tianming; SHEN, Lei; SHEN, Ming; LIU, Litao; CHEN, Fengnan; GAO, Li. Evolution and projection of co2 emissions for china's cement industry from 1980 to 2020. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 74, p. 522–537, 2017.

GHASEMI, Asghar; ZAHEDIASL, Saleh. Normality tests for statistical analysis: a guide for non-statisticians. **International Journal of Endocrinology and Metabolism**, Kowsar Medical Institute, v. 10, n. 2, p. 486, 2012.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning**: data mining, inference, and prediction. [S.l.]: Springer Science & Business Media, 2009.

HAYKIN, Simon. **Neural networks**: a comprehensive foundation. [S.l.]: New Jersey: Printice-hall., 1999.

HUANG, Cheng-Lung; DUN, Jian-Fan. A distributed pso–svm hybrid system with feature selection and parameter optimization. **Applied soft computing**, Elsevier, v. 8, n. 4, p. 1381–1391, 2008.

KASPERKIEWICZ, Janusz; RACZ, Janusz; DUBRAWSKI, Artur. Hpc strength prediction using artificial neural network. **Journal of Computing in Civil Engineering**, American Society of Civil Engineers, v. 9, n. 4, p. 279–284, 1995.

KRÖSE, Ben; KROSE, Ben; SMAGT, Patrick VAN DER; SMAGT, Patrick. An introduction to neural networks. **Citeseer**, 1993.

LIN, Shih-Wei; YING, Kuo-Ching; CHEN, Shih-Chieh; LEE, Zne-Jung. Particle swarm optimization for parameter determination and feature selection of support vector machines. **Expert systems with applications**, Elsevier, v. 35, n. 4, p. 1817–1824, 2008.

MCKINNEY, Wes. Data structures for statistical computing in python. **Proceedings of the 9th Python in Science Conference**, Austin-TX, v. 445, p. 51–56, 2010.

NEVILLE, Adam M. **Propriedades do Concreto**. 5. ed. [S.l.]: Bookman Editora, 2015.

NI, Hong-Guang; WANG, Ji-Zong. Prediction of compressive strength of concrete by neural networks. **Cement and Concrete Research**, Elsevier, v. 30, n. 8, p. 1245–1250, 2000.

OLIVEIRA, Adriano LI; BRAGA, Petronio L; LIMA, Ricardo MF; CORNÉLIO, Márcio L. Ga-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation. **Information and Software Technology**, Elsevier, v. 52, n. 11, p. 1155–1166, 2010.

PEDREGOSA, Fabian; VAROQUAUX, Gaël; GRAMFORT, Alexandre; MICHEL, Vincent; THIRION, Bertrand; GRISEL, Olivier; BLONDEL, Mathieu; PRETTENHOFER, Peter; WEISS, Ron; DUBOURG, Vincent. Scikit-learn: Machine learning in python. **The Journal of Machine Learning Research**, JMLR. org, v. 12, p. 2825–2830, 2011.

SHEN, Weiguo; LIU, Yi; YAN, Bilan; WANG, Jing; HE, Pengtao; ZHOU, Congcong; HUO, Xujia; ZHANG, Wuzong; XU, Gelong; DING, Qingjun. Cement industry of china: driving force, environment impact and sustainable development. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 75, p. 618–628, 2017.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. Practical bayesian optimization of machine learning algorithms. **Advances in Neural Information Processing Systems**, p. 2951–2959, 2012.

TSAI, Hsing-Chih; LIN, Yong-Huang. Modular neural network programming with genetic optimization. **Expert Systems with Applications**, Elsevier, v. 38, n. 9, p. 11032–11039, 2011.

USGS. **Mineral commodity summaries**. 2019.

VAPNIK, Vladimir. **Statistical learning theory**. [S.l.]: Wiley, New York, 1998.

VAPNIK, Vladimir. **The Nature of Statistical Learning Theory**. 2. ed. [S.l.]: Springer, 1999.

VARGHA, András; DELANEY, Harold D. The kruskal-wallis test and stochastic homogeneity. **Journal of Educational and behavioral Statistics**, SAGE Publications, Los Angeles-CA, v. 23, n. 2, p. 170–192, 1998.

WEISS, Sholom M; KULIKOWSKI, Casimir A. **Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems**. [S.l.]: Morgan Kaufmann Publishers Inc., 1991.

WHITNALL, Carolyn; OSWALD, Elisabeth; MATHER, Luke. An exploration of the kolmogorov-smirnov test as a competitor to mutual information analysis. **International Conference on Smart Card Research and Advanced Applications**, Springer, p. 234–251, 2011.

ZHU, Qin-Yu; QIN, A Kai; SUGANTHAN, Ponnuthurai N; HUANG, Guang-Bin. Evolutionary extreme learning machine. **Pattern recognition**, Elsevier, v. 38, n. 10, p. 1759–1763, 2005.