

Uso de Redes Neurais Artificiais (RNA) na estimativa volumétrica de espécies da Caatinga arbórea

| **Loic Aymard Mfoumbyt**

Universidade Federal de Minas Gerais (UFMG)

| **Renato Dourado Maia**

Universidade Federal de Minas Gerais (UFMG)

| **Christian Dias Cabacinha**

Universidade Federal de Minas Gerais (UFMG)

RESUMO

A Caatinga Arbórea é uma tipologia vegetacional da Caatinga, bioma exclusivamente brasileiro, composta por diversas espécies de interesse medicinal, alimentício (humano e animal) e econômico. O conhecimento do seu estoque volumétrico é fundamental na elaboração de planos de manejo e o uso das Redes Neurais Artificiais é um método que se destaca por sua capacidade de produzir excelentes resultados. Assim, o objetivo deste trabalho foi definir o número mínimo de indivíduos para obter redes satisfatórias a partir do treinamento de Redes Neurais Artificiais com algoritmo genético NEAT para a estimativa do volume comercial da Caatinga arbórea. Dados de volume comercial com casca de 300 indivíduos foram obtidos para compor a base de dados para o treinamento e validação das Redes Neurais Artificiais. O processamento foi composto por duas fases, na qual a primeira continha redes com 10 indivíduos até 100 em intervalos de 10 indivíduos e a segunda, redes de 100 indivíduos até 1000 em intervalos de 100 indivíduos. Realizou-se uma análise das estatísticas correlação e raiz quadrada do erro médio para cada fase, A situação que gerou os melhores resultados de treinamento foi comparada a partir do teste de sinais para 5%. As redes com número de indivíduos inferior a 300 foram descartadas e a partir de 400 indivíduos, com densidade populacional de 0,1, apresentaram as maiores correlações e as menores raízes quadradas do erro médio no treinamento das redes. Assim, concluiu-se que durante o treinamento de redes neurais artificiais para estimar o volume comercial com casca utilizando como variáveis de entrada o DAP e a HT com o algoritmo NEAT, recomenda-se utilizar pelo menos 400 indivíduos com densidade populacional de 0,1.

Palavras-chave: Caatinga, Algoritmo Genético, Produção Volumétrica, Tamanho Populacional.

■ INTRODUÇÃO

A Caatinga é um bioma exclusivamente brasileiro que ocupa uma área de 844.453 km², 10% do território do país, distribuída nos estados do Ceará, Alagoas, Minas Gerais, Bahia, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Subdivisão deste bioma, a Caatinga Arbórea é composta por diversas espécies de interesse medicinal, alimentício (humano e animal) e econômico (RIZZINI, 1997), além da utilização da sua madeira para serraria e carvão, como exemplo da *Jatropha molíssima* (Pohl) Baill. (Pinhão), a *Ziziphus joazeiro* Mart (Juazeiro) e *Myracrodruon urundeuva* Allemão (Aroeira), comumente destacadas em planos de manejo na região.

Diante da importância e uso da Caatinga Arbórea, faz-se necessário realizar estudos para obtenção do estoque volumétrico presente, realizando a supressão dos indivíduos para valoração e quantificação do que será explorado (SANQUETTA *et al.*, 2014). Comumente, usa-se equações volumétricas ajustadas para estimar o volume da vegetação, porém os modelos matemáticos encontrados na literatura nem sempre conseguem explicar o comportamento dessa variável de forma satisfatória devido à alta variabilidade que envolve algumas variáveis como solo, espécie e bioma interferindo diretamente no volume das árvores (REZENDE *et al.*, 2006). Como saída, pode-se fazer uso de outras ferramentas como as Redes Neurais Artificiais (RNA).

As Redes Neurais Artificiais (RNAs) são sistemas complexos que, através da entrada de dados, detectam padrões e funcionam similarmente a um cérebro humano, capazes de fazer interpolações e extrapolações de forma linear e geram cenários distintos em diferentes arquiteturas que são algoritmos ligadas aos neurônios (HAYKIN, 2001; BRAGA *et al.*, 2014). Essa metodologia tem se destacado pela sua capacidade de produzir excelentes resultados e pela tendência em gerar erros mínimos.

Um dos algoritmos que pode ser usado para o treinamento de redes neurais é o Neuroevolution of Augmenting Topologies (NEAT). O Neuroevolution of Augmenting Topologies, é um algoritmo genético de otimização e busca, baseado na teoria da seleção natural e genética de Darwin, que a partir de uma técnica de um algoritmo evolutivo ajusta simultaneamente os pesos sinápticos e a topologia de uma Rede Neural Artificial.

Diante disso, o presente trabalho teve por objetivo definir o número mínimo de indivíduos para obter redes satisfatórias a partir do treinamento de Redes Neurais Artificiais com algoritmo genético NEAT para a estimativa do volume comercial da Caatinga arbórea.

■ MÉTODOS

Caracterização da área de estudo

O estudo foi realizado na Floresta Nacional Contendas do Sincorá (FLONA) localizada no município de Contendas do Sincorá-BA. A FLONA é uma unidade de conservação introduzida no grupo de uso sustentável, segundo a Lei nº 9.985, que determina o Sistema Nacional de Unidades de Conservação (BRASIL, 2000). A Floresta Nacional Contendas do Sincorá possui área de 11.215,78 hectares e está situada nas coordenadas 13°55'14,51"S e 41°6'54,79"O. O solo predominante na área é o Latossolo Vermelho-Amarelo, com expressivas faixas de Neossolos Litólicos a oeste e leste do município de Contendas do Sincorá (INEMA, 2014). O clima é semiárido quente do tipo BSw^h, conforme a classificação de Köppen com valores pluviométricos que chegam entre 500 a 1.000 mm anuais, temperatura de 21 a 28 °C e umidade relativa de 60 a 70%.

Área amostral e critérios de inclusão

Para seleção de indivíduos foram utilizados os dados de inventário, ou seja, distribuiu-se em uma área de Caatinga arbórea arbustiva três blocos com 16 parcelas cada, sendo cada parcela subdividida em parcelas de 400 m² (20m x 20m). Quatro tratamentos e a testemunha. Na comunidade, foram considerados todos os indivíduos com diâmetro a toros repartidos ao acaso foram implantados, sendo estes atribuídos baseando-se no corte raso, corte seletivo das espécies de relevância econômica, com base no diâmetro de inclusão m 1,30 m do solo (DAP) superior ou igual a 5 cm, usando a suta para medição e estimativa visual para determinar a variável altura em todas as parcelas.

Cubagem Rigorosa

A distribuição por classe diamétrica foi o critério adotado para obter-se uma amostragem significativa do volume comercial com casca (VC_{cc}, em m³), considerando uma amplitude de 5 cm por classes, com amostragem total de 300 indivíduos, conforme métodos utilizados para florestas nativas em estágio inicial ou secundário de regeneração (SOUZA; SOARES, 2013).

Foram mensurados diâmetros das diferentes espécies com suta, utilizando o método de Smalian, que considera as medições da árvore com seções ao longo do fuste, nas posições 0,1; 0,3; 0,5; 0,7; 0,9; 1,1; 1,3; 2,0; metros e assim sucessivamente, até a altura total (CAMPOS; LEITE, 2013; MACHADO; FIGUEIREDO FILHO, 2009; SOARES *et al.*, 2006). Desse modo, foi estabelecido um limite mínimo de 3 cm de diâmetro ao medir cada seção, bem como realizadas avaliações de galhos com circunferência superior à 3 cm de diâmetro.

Base de Dados

Os dados de volume resultantes da cubagem rigorosa, pertencem a 24 espécies identificadas e 5 espécies identificadas a nível de gênero, 63 morfo espécies não foram identificadas devido a perda de folhas decorrente à época da coleta de dados quando registriu-se uma precipitação muito baixa. As seguintes variáveis quantitativas foram consideradas: diâmetro com casca a 1,3 m de altura (DAP), altura total – em metros (HT) e o volume comercial com casca (VCcc, em m³).

Tabela 1. Valores globais das variáveis quantitativas mensuradas na cubagem.

Variável	Mínimo	Máximo	Médio	Desvio Padrão
DAP (cm)	5,00	52,60	10,50	6,96
HT (m)	3,50	11,50	6,74	1,41
VCcc (m ³)	0,00354	1,07307	0,05743	0,11392

Fonte: Do autor, 2021.

Redes neurais artificiais

No setor florestal, as arquiteturas *Radial Basis Function* (RBF) e redes *Multilayer Perceptron* (MLP) são as mais usadas para aplicar as redes neurais artificiais (BRAGA *et al.*, 2014). Neste estudo, utilizamos a rede do tipo MLP que apresenta menos complexidades em razão da composição da sua arquitetura, as quais formam uma camada de entrada de informações a ser tratadas, uma camada intermediária com a função de processar os dados e uma camada de saída com finalidade de dar resultados do treinamento (BINOTI *et al.*, 2013; 2014). Assim, as redes do tipo MLP foram testadas usando o algoritmo *Neuroevolution of Augmenting Topologies* (NEAT).

Desta forma, o algoritmo genético (AG) proporcionou o treinamento e a validação das redes, isto é, 70% dos dados foram usados para treinamento e 30% para validação (ARAÚJO, 2015b). Foi considerado uma densidade populacional de 0,1, com um número de indivíduos que variou de 10 até 100 em um intervalo de 10 para a primeira fase e de 100 a 1000 com intervalo de 100 para segunda fase de treinamento.

CrITÉRIOS estatísticos para avaliação do desempenho da RNA

Para estudo do comportamento das redes, foram avaliados o coeficiente de correlação entre os valores observados e estimados do volume ($r_{y\hat{y}}$) e a raiz do quadrado médio do erro (RQME%) para análise estatística, assim como para avaliação das dispersões gráficas dos resíduos percentuais foram considerados os gráficos do tipo *boxplot* para comparação das

redes. A situação que gerou os melhores resultados de treinamento foi comparada a partir do teste de sinais para 5%.

■ RESULTADOS

As redes neurais artificiais foram treinadas e validadas em duas fases, a primeira com 10 até 100 indivíduos e a segunda de 100 a 1000, as estatísticas das redes selecionadas podem ser observadas nas Tabelas 2 e 3, respectivamente.

Tabela 2. Estatísticas das redes neurais artificiais com uso do algoritmo *Neuroevolution of Augmenting Topologies* (NEAT) para estimativa do volume comercial da casca (VCcc), considerando população de 10 até 100 indivíduos.

Tamanho da população	Treinamento		Validação	
	$r_{\hat{y}y}$	RQME%	$r_{\hat{y}y}$	RQME%
10	95,78	3,18	98,55	2,98
20	96,19	3,00	98,71	2,65
30	96,39	2,92	98,90	2,48
40	96,26	2,97	98,84	2,55
50	96,46	2,89	98,94	2,42
60	96,24	2,98	98,81	2,61
70	96,45	2,89	98,92	2,44
80	96,50	2,87	98,97	2,38
90	96,58	2,83	97,78	2,61
100	96,48	2,88	98,95	2,42

Fonte: Do autor, 2021.

Tabela 3. Estatísticas das redes neurais artificiais com uso do algoritmo *Neuroevolution of Augmenting Topologies* (NEAT) para estimativa do volume comercial da casca (VCcc), considerando população de 10 até 100 indivíduos.

Tamanho da população	Treinamento		Validação	
	rr'%	RQME%	rr'%	RQME%
100	96,48	2,88	98,96	2,42
200	96,74	2,76	99,18	2,27
300	96,80	2,73	98,94	2,38
400	97,22	2,54	97,16	2,65
500	97,38	2,48	96,81	3,23
600	97,45	2,44	91,33	4,31
700	97,68	2,33	93,36	4,94
800	97,74	2,30	89,67	4,36
900	97,60	2,36	94,32	4,09
1000	97,48	2,43	93,65	4,65

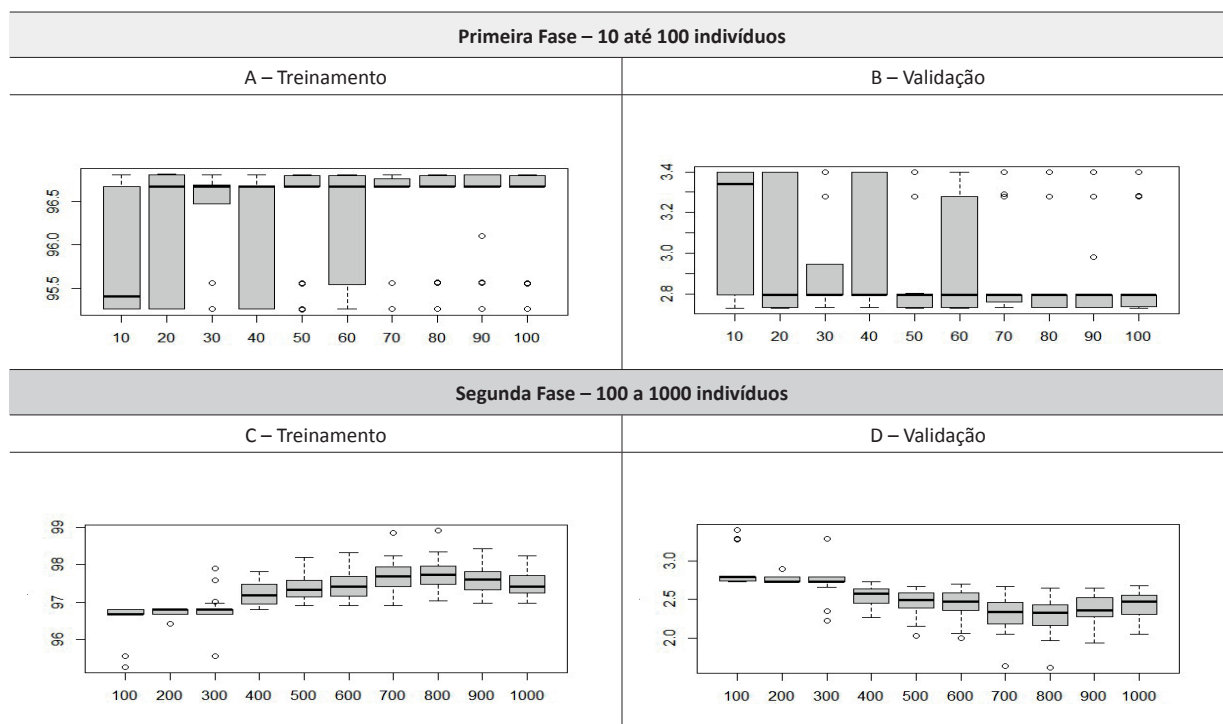
Fonte: Do autor, 2021.

As Figuras 1A e 1B, mostram os boxplot da fase 1 de processamento. Verifica-se tanto para o treinamento como para a validação, uma grande variação nos comportamentos das redes neurais.

As Figuras 1C e 1D, mostram os boxplot da fase 2 de processamento, onde verificou-se resultados mais satisfatórios, exceto para as configurações considerando 100, 200

e 300 indivíduos que apresentaram comportamento semelhantes aos das redes da fase 1 de processamento.

Figura 1. Boxplot para a correlação (a) e RQME% (b) para a fase 1 de processamento dos dados e para a correlação (c) e RQME% (d) para a fase 2 de processamento dos dados.



Fonte: Do autor, 2021.

Após avaliar os resultados das estatísticas e gráficos boxplot, verificou-se que as redes a partir de 400 indivíduos (fase 2) apresentaram os melhores resultados e estas foram comparadas pelo teste de sinal. O resultado dessa comparação pode ser observado na Tabela 4.

Tabela 4. Resultados das comparações de redes da fase de treinamento para RQME do teste de sinais de estimativa do volume comercial com casca (VCcc).

	400	500	600	700	800	900	1000
400		•	•	•	•	•	•
500			ns	•	•	•	ns
600				ns	•	•	ns
700					•	ns	ns
800						ns	•
900							ns
1000							

Onde: “•” significativo para $\alpha=0,05$; ns = não significativo.

Fonte: Do autor, 2021.

■ DISCUSSÃO

Observou-se pelos os resultados das Tabelas 2 e 3, para as estatísticas correlação e raiz do quadrado médio do erro que não houveram grandes diferenças para as duas estatísticas, tanto na fase de treinamento como na validação para as duas fases analisadas. Contudo, verifica-se que os resultados de treinamento da segunda fase são ligeiramente melhores (correlações mais altas e raízes do quadrado médio do erro menores) aos da primeira fase. Já na validação os resultados da primeira fase se mostraram ligeiramente melhores.

Analisando as Figuras 1A e 1B, observa-se que a mediana para a correlação aumentou a partir da rede com 20 indivíduos e praticamente não se alterou até a rede com 100 indivíduos. Já para o RQME% a rede 10 indivíduos apresentou a maior mediana, a partir de 20 indivíduos a mediana diminuiu e praticamente não se alterou até a rede com 100 indivíduos. Com relação a dispersão, verificou-se grande dispersão e uma forte assimetria nos resultados de correlação e RQME% nas redes com 10, 20, 40 e 60 indivíduos. Para a rede 10, a assimetria foi positiva, para as demais a assimetria foi negativa.

Embora as demais configurações tenham apresentado uma dispersão menor, verificou-se ainda assimetria, sendo para a rede com 30 indivíduos negativa e as demais configurações uma assimetria positiva, sendo que para as redes 50, 70, 80, 90 e 100 a mediana foi igual ao primeiro quartil, resultado que pode estar associado ao grande número de valores discrepantes identificados na análise do boxplot. Diante destes resultados verificados as redes treinadas durante a fase 1 de processamento foram consideradas insatisfatórias e descartadas para a estimativa de volume comercial para Caatinga arbórea.

Esse comportamento era esperado, embora as redes com o maior número de indivíduos o custo computacional na etapa de treinamento seja maior, populações maiores toleram maior número de espécies potenciais, ou seja, maior quantidade de RNA testadas com alta diversidade topológica, aumentando as chances de obter bons resultados (STANLEY; MIIKKULAIEN, 2002).

Observando as Figuras 1C e 1D, verificou-se que a partir de 400 indivíduos a mediana da correlação aumentou até a rede com 800 e voltou a diminuir para as redes 900 e 1000 indivíduos. Para o RQME%, como esperado, verificou-se o contrário, a partir de 400 indivíduos os valores das medianas decresceram até o mínimo que ocorreu na rede com 800, voltando a subir para as redes 900 e 1000 indivíduos.

Araújo (2015) quando avaliou redes NEAT para estimar o volume de formações florestais no estado de Minas Gerais, testou as seguintes configurações para o treinamento: número de indivíduos da população variando em 100, 300, 500, 600, 800 e 1000 e densidades de conexões iniciais de 0,1 e 0,5. A autora constatou que somente a partir de populações maiores que 600 foram geradas redes com bom desempenho e concluiu que a partir das

duas densidades de conexões iniciais analisadas (0,1 e 0,5), é possível obter estimativas volumétricas com bom grau de exatidão para as onze tipologias florestais estudadas e diferentemente ao verificado neste estudo, recomendou populações iniciais com número de indivíduos maiores que 600, já que com populações inferiores não foram encontradas redes com bom desempenho para essa finalidade.

Os resultados do teste de sinais (Tabela 4) evidenciaram que a rede de 800 indivíduos (que apresentou menor mediana para a RQME%) não se diferenciou estatisticamente apenas da rede com 900 indivíduos. Contudo, as diferenças entre a rede de 800 indivíduos e demais configurações comparadas não são na prática diferenças a serem consideradas para o treinamento ao se ponderar os ganhos de estimativa do volume e o tempo de processamento das configurações avaliadas.

Neste sentido, verificou-se que a partir de 400 indivíduos as redes geradas estimavam o volume satisfatoriamente com um custo computacional bem inferior às redes de 700, 800 e ou 900 indivíduos.

■ CONCLUSÃO

Os resultados desse trabalho evidenciam a potencialidade do uso de rede neurais treinadas com o algoritmo *Neuroevolution of Augmenting Topologies* (NEAT) para estimar o volume na Caatinga arbórea. Assim como, permitem concluir que durante o treinamento de redes para estimar o volume comercial com casca utilizando como variáveis de entrada o DAP e a HT com o algoritmo NEAT, recomendar-se utilizar pelo menos 400 indivíduos com densidade populacional de 0,1.

■ REFERÊNCIAS

1. ARAÚJO, B. H. P. de. **Modelagem da altura, volume e afilamento de *Calycophyllum spruceanum* Benth. empregando regressão e redes neurais artificiais.** Manaus-AM: INPA, 2015b, 53 p. Dissertação (Mestrado em Ciências de Florestas Tropicais) – Instituto Nacional de Pesquisas da Amazônia.
2. BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira**, v. 5, n. 1, p. 58-67, 2014.
3. BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639-645, 2013

4. BINOTI, M. L. M. **Redes neurais artificiais para prognose da produção de povoa-mentos não desbastados de eucalipto**. Viçosa-MG: UFV, 2010, 54 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Viçosa
5. BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed.- [Reimpr.]. LTC, Rio de Janeiro, 2014, 26 p.
6. CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 4.ed., Viçosa: Editora UFV, 2013. 605p.
7. COULSON, R.N.; FOLSE, L.J.; LOH, D. K. Artificial intelligence and natural resource management. **Science**, v. 237, n. 4812, p. 262-267, 1987.
8. FERNANDES, Moabe Ferreira; QUEIROZ, Luciano Paganucci de. Vegetação e flora da Caatinga. **Ciência e cultura**, v.70, n.4, p.51-56, 2018.
9. FIGUEIREDO, E. O.; SCHROEDER, R.; PAPA, D. de A. Fatores de Forma para 20 Espécies Florestais comerciais da Amazônia. **Comunicado Técnico**, n. 173. Embrapa, Rio Branco, AC, 2009.
10. FIGUEIRÔA, J. M. et al., Madeireiras, In: SAMPAIO, E. V. S. B.; et al. **Espécies da Flora Nordestina de Importância Econômica Potencial**. Recife: APNE, 2005, p, 101-133.
11. GORGENS, E. B.; LEITE, H. G.; SANTOS, H. do N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, v. 33, n. 6, p.1141-1147, 2009.
12. HAYKIN, S. **Neural networks and learning machines**. 3. ed. Prentice Hall, New York, 2009, 906 p.
13. HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Bookman, Porto Alegre, 2001, 900 p.
14. IBAMA – Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis. **Plano de Manejo: Floresta Nacional Contendas do Sincorá, Volume I**. MMA, Brasília, 2006.
15. KRAMER, O. **Genetic Algorithms Essentials**. International publishing, 2017
16. LEITE, H. G. et al. Modelo de afilamento de cerne de *Tectona grandis* L.f. Revista Scientia Forestalis, **Piracicaba**, v. 39, n. 89, p.53-59, mar. 2011.
17. MACHADO, S.A.; FIGUEIREDO FILHO, A. **Dendrometria**, 2^a.ed., 1.^a reimpressão, Guarapuava:UNICENTRO, 2009. 309p.
18. REZENDE, A. V.; VALE, A. T.; SANQUETTA, C. R.; FIGUEIREDO FILHO, A.; FELFILI, J. M. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. **Scientia Forestalis**, n. 71, p. 65-76, 2006.
19. RIZZINI, C.T. **Tratado de Fitogeografia do Brasil: Aspectos ecológicos, sociológicos e florísticos**. 2. ed. Âmbito Cultural, Rio de Janeiro, Volume único ,1997, 747 p.

20. SANQUETTA, C. R.; DALLA CORTE, A. P.; RODRIGUES, A. L.; WATZLAWICK, L. F. **Inventários Florestais**: planejamento e execução. 3. ed. Multi-graphic gráfica e editora, Curitiba, 2014, 406 p.
21. SANQUETTA, C. R.; DALLA CORTE, A. P.; JACON, A. D. Crown area and trunk diameter relationship for tree species at a mixed-araucaria natural forest in the midsouthern Parana state, Brazil. **Floresta**, v. 41, n. 1, p. 63-72, 2011.
22. SANTOS, A. T. dos; MATTOS, P. P. de; BRAZ, E. M.; ROSOT, N. C. Equação de volume e relação hipsométrica em plantio de *Ocotea porosa*. **Pesquisa florestal brasileira**, v. 32, n. 69, p. 13-21, 2012.
23. SCOLFORO, J. R.; RUFINI, A. L.; MELLO, J. M. de; OLIVEIRA, A. D. de; CASTRO SILVA, C. P. de; TRUGILHO; P. F. Equações de volume, peso de matéria de seca e produção de tanino do angico-vermelho (*Anandenanthera colubrina* (Benth) Brenan), em Minas Gerais. In: SCOLFORO, J. R.; OLIVEIRA, A. D.; ACERBI JÚNIOR, F. W. (Ed.). **Inventário Florestal de Minas Gerais - Equações de Volume, Peso de Matéria Seca e Carbono para Diferentes Fisionomias da Flora Nativa**. Lavras: UFLA, 2008, p. 117-128.
24. SILVA, I. N. d. et al. *Artificial Neural Networks : A Practical Course*. 1. ed. [S.l.]: Springer International Publishing, 2017. ISBN 978-3-319-43162-8,978-3-319-43161-1. Citado 3 vezes nas páginas 29, 30 e 36.
25. SOARES, C. P. B.; NETO, F. de P.; SOUZA, A. L. de (Ed.). **Dendrometria e Inventário Florestal**. 2. ed. UFV, Viçosa, 2011a, 272 p
26. SOUZA, A. L. de; SOARES, C. P. B. **Florestas nativas**: estrutura dinâmica e manejo. UFV, Viçosa, 2013, 322 p.
27. SOUZA, H. S. et al. Comparação de métodos de cubagem para eucalipto. *Rev. Bras. Biom.*, Lavras, v.35, n.1, p.17-26, 2017.
28. STANLEY, K. O.; MIIKKULAINEN, R. Evolving Neural Networks through augmenting Topologies. **Evolutionary Computation**, v. 10, n. 2, p. 99-127, 2002.