

**ESTIMATIVA DA ALTURA TOTAL DE ÁRVORES UTILIZANDO DIFERENTES TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

**ESTIMATION OF TOTAL TREE HEIGHT USING DIFFERENT ARTIFICIAL INTELLIGENCE TECHNIQUES**

Thais Sales Gonçalves<sup>1</sup> Gabriela Letícia Ramos Carvalho Paulo Ricardo Santos Miranda  
Emanuelly Canabrava Magalhães Sthefany Mendes Zuba Carlos Alberto Araújo Junior

**RESUMO**

O presente trabalho teve como objetivo avaliar o desempenho de quatro técnicas de inteligência artificial para estimativa da altura total de árvores adotando como variáveis independentes Idade, Área útil por planta e diâmetro a 1,30 m do solo (Dap). As técnicas testadas foram *Random Forest*, *Neuro Fuzzy*, *Support Vector Machine* e Redes Neurais Artificiais, sendo o processamento dos dados possibilitado pelo uso dos softwares Statística, MatLab e NeuroForest. Foi realizado dois tipos de treinamento; para o primeiro os dados referentes ao último ano do inventário florestal não foram considerados, e para o segundo foi utilizado os dados de todos os anos sendo estes aleatorizados. Após o processamento, foram calculadas as estatísticas Bias, RQME, correlação e erro percentual médio; gerados gráficos de dispersão, considerando os valores observados; e histograma de resíduos. Constatou-se que o processamento com os dados aleatorizados apresentou os melhores índices estatísticos e a técnica *Support Vector Machine* apresentou os melhores resultados quando comparada às demais.

**Palavras-chave:** Inventário florestal; Dendrometria; Análise estatística.

**ABSTRACT**

The objective of this study was to evaluate the performance of four artificial intelligence techniques to estimate the total height of trees, using as independent variables Age, Area per plant and diameter at 1,30 m of soil (Dap). The techniques tested were Random Forest, Neuro Fuzzy, Support Vector Machine and Artificial Neural Networks, being the data processing made possible by the use of Statística, MatLab and NeuroForest software. Two types of training were performed; for the first one the data referring to the last year of the forest inventory were not considered, and for the second one the data of all the years were used being these randomized. After the processing, the statistics Bias, RQME, correlation and average percentage error were calculated; scatter plots were generated, considering the observed values; and histogram of residues. It was verified that the processing with the randomized data presented the best statistical indices and the Support Vector Machine technique presented the best results when compared to the others.

**Keywords:** Forest Inventory; Dendrometry; Statistical analysis.

**INTRODUÇÃO**

A mensuração florestal é uma atividade que fornece informações sobre um povoamento florestal, a fim de se obter o conhecimento das potencialidades produtivas, além de atender diversas finalidades, as quais visam estabelecer metas comerciais, de manejo, ordenação florestal e de pesquisa (JESUS et al., 2012). As variáveis dendrométricas que podem ser mensuradas são diâmetro, área basal, altura, volume, entre outras (DE JESUS et al., 2015).

Dentre as variáveis mensuráveis, a de maior destaque e mais recorrente em inventários florestais, é a altura, sendo de fundamental importância na determinação do volume produtivo, norteando o planejamento, estruturação e direcionamento das ações. A altura é estabelecida por Machado & Figueiredo Filho (2014) como sendo a distância linear ao longo do eixo principal da árvore, partindo do solo até o topo ou até outro ponto referencial.

A medição de tal variável é uma atividade onerosa, tornando inviável a obtenção da altura de todas as árvores que compõem as parcelas nos povoamentos florestais (DE JESUS et. al., 2015.), sendo estabelecidas portanto equações matemáticas, comumente referenciadas como relações hipsométricas que possibilitem estimar a altura das árvores contidas em uma determinada parcela, tendo como base os indivíduos amostrados.

<sup>1</sup> Técnico (a) em Informática. Discente de Engenharia Florestal, Universidade Federal de Minas Gerais, 39.404-547, Montes Claros, MG, Brasil. E-mail: thaiswj@gmail.com

Assim, a relação hipsométrica é a relação existente entre a altura e o diâmetro das árvores de um povoamento, podendo ser representada por um ajuste matemático de uma regressão (SCHIMIDT, 1977) ou utilizando técnicas de inteligência artificial. Dentre as técnicas de inteligência artificial existentes, aqui propõe-se um comparativo entre quatro ferramentas matemáticas distintas: *Random Forest*, *Neuro Fuzzy*, *Support Vector Machine* e Redes Neurais Artificiais.

Conforme definido por Breiman (2001), *Random Forest* consiste em uma combinação de preditores de árvores, de tal forma que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta. *Support Vector Machine* é uma técnica de aprendizado de máquina, embasada pela teoria de aprendizado estatístico, desenvolvida por Vapnik (1995) que estabelece uma série de princípios a serem seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu (LORENA & CARVALHO, 2017).

Redes Neurais Artificiais (RNAs), são definidas como um sistema massivamente paralelo e distribuído, composto por unidades de processamento simples, inspiradas no sistema neural, que possuem uma capacidade natural de armazenar e utilizar conhecimento (HAYKIN, 1999). *Neuro Fuzzy* são sistemas híbridos que usam estratégias locais de aprendizado para aprender conjuntos difusos e regras difusas fazendo a junção de dois importantes paradigmas: redes neurais artificiais, elucidado anteriormente, e sistemas de inferência *fuzzy*, que, resumidamente, permite trabalhar com imprecisões (ARAÚJO JÚNIOR et. al., 2016).

Por conseguinte, o presente trabalho objetiva, através do uso de técnicas de inteligência artificial, submeter uma base de dados amostrais conhecidos a diferentes ferramentas matemáticas aqui abordadas para estimar a altura total de indivíduos de uma determinada parcela e analisar o resultado obtido em cada processamento realizado tendo como variáveis de entrada para análise e suporte no cálculo da altura total, o diâmetro a 1,30 m do solo (Dap), área e idade.

## MATERIAIS E MÉTODOS

O banco de dados utilizado no processamento foi obtido de inventários florestais contínuos, realizados no período de 2013 a 2016 em plantios florestais na região do norte do estado de Minas Gerais e dividida em treinamento e validação equivalentes 11929 e 202 dados, respectivamente. O banco de dados foi separado em dados de treinamento e de validação considerando duas abordagens: Abordagem 1, que consiste em dados de treinamento referentes aos anos de 2013, 2014 e 2015 e para validação do modelo dados referentes ao ano de 2016; e Abordagem 2, com escolha aleatória dos dados para treinamento e validação.

Foram avaliadas quatro ferramentas de inteligência artificial para estimar valores de altura total a partir de dados de Idade (Id), Área útil por planta (Ap) e Dap, sendo elas: sistema de inferência neuro fuzzy (NF), redes neurais artificiais (RNA), random forest (RF) e support vector machine (SVM). Para tal, empregou-se o uso de três softwares, o NeuroForest para o processamento dos dados utilizando as redes neurais artificiais, o Statistica versão 13 para a aplicação das técnicas Random Forest e Support Vector Machine e o MatLab para Neuro fuzzy.

As parametrizações consideradas para cada ferramenta foram as seguintes: 500 árvores e 2 valores de predição para Random Forest; Gamma igual a 0,330, Epsilon igual a 0,2 e Capacity igual a 0,6 para Support Vector Machine; 500 ciclos como critério de parada, 8 neurônios na camada intermediária, função sigmoideal e algoritmo de aprendizagem Resilient Propagation (RPROP) para as Redes Neurais Artificiais, considerando 30 repetições; e 5 épocas de treinamento e 3 funções de pertinência para o sistema de inferência neuro-fuzzy.

Para analisar as alturas estimadas considerou-se, para cada técnica, a elaboração de gráficos de dispersão e histogramas de resíduos. Ainda, foram calculados as estatísticas Bias (equação 1), raiz quadrada do erro médio quadrático em porcentagem (Equação 2), correlação entre valores observados e estimados (equação 3) e a média do erro percentual (Equação 4).

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (\text{EQUAÇÃO 1})$$

$$\text{RQME (\%)} = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{EQUAÇÃO 2})$$

$$r = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{EQUAÇÃO 3})$$

$$EMP = \frac{1}{n} \sum_{i=1}^n \left[ 100 \cdot \frac{(y_i - \hat{y}_i)}{y_i} \right] \quad (\text{EQUAÇÃO 4})$$

## RESULTADOS E DISCUSSÃO

De maneira geral, a Abordagem 2 apresentou resultados melhores que a Abordagem 1 para os dados de validação (Tabela 1), o que pode ser explicado pelo fato de que os valores observados para os dados de validação se concentraram no extremo superior da distribuição dos valores observados nos dados de treinamento. Tal porção não é tão representativa para o treinamento quanto os valores das alturas medianas e menores (Figura 1).

Dentre as ferramentas utilizadas e considerando a Abordagem 1, a que se destaca é a Random Forest, apresentando os melhores valores de correlação, bias, raiz quadrada do erro médio quadrático e erro médio percentual. Já para a Abordagem 2, as melhores ferramentas foram a Support Vector Machine e a Random Forest. Tais técnicas são pouco exploradas para obtenção de tais estimativas, o que significa que é possível realizar estudos mais aprofundados e que consigam gerar resultados melhores que os encontrados aqui.

TABELA 1: Resultados para as estatísticas de avaliação das técnicas de inteligência artificial empregadas.

TABLE 1: Evaluation statistics results for the artificial intelligence techniques used.

Abordagem	Estatística	RNA <sub>t</sub>	RNA <sub>v</sub>	SVM <sub>t</sub>	SVM <sub>v</sub>	RF <sub>t</sub>	RF <sub>v</sub>	NF <sub>t</sub>	NF <sub>v</sub>
1	Bias	0.00	-1.24	0.04	-1.44	0.00	-0.64	0.00	-1.36
	RQME%	11.46	7.77	11.56	8.67	10.39	7.34	11.23	8.58
	EMP	1.77	-6.90	2.81	-8.13	1.66	-3.46	1.60	-7.46
	r	0.94	0.71	0.94	0.67	0.95	0.72	0.94	0.64
2	Bias	-0.01	0.14	-0.07	0.02	0.00	0.11	0.00	0.13
	RQME%	11.33	11.18	11.40	11.24	10.33	10.74	11.16	11.12
	EMP	1.62	2.33	1.39	1.88	1.66	2.22	1.58	2.25
	r	0.94	0.94	0.94	0.94	0.95	0.95	0.94	0.94

Em que, RNA<sub>t</sub> = Rede Neural Artificial treino; RNA<sub>v</sub> = Rede Neural Artificial validação; SVM<sub>t</sub> = *Support Vector Machine* treino; SVM<sub>v</sub> = *Support Vector Machine* validação; RF<sub>t</sub> = *Random Forest* treino; RF<sub>v</sub> = *Random Forest* validação; NF<sub>t</sub> = *Neuro Fuzzy* treino; NF<sub>v</sub> = *Neuro Fuzzy* validação.

Os gráficos de dispersão revelam que a Abordagem 1 proporcionou uma base de dados de validação que se concentrou em árvores com alturas acima de 23m (Figura 1), ao passo que os dados de treinamento possuem uma maior quantidade de alturas abaixo de tal valor. Já para a Abordagem 2, houve uma melhor distribuição dos dados de validação.

Em termos de aplicação das técnicas para estimativas dentro do intervalo de dados utilizados no treinamento, como é o caso da Abordagem 2, todas as técnicas apresentaram ótimos resultados, podendo ser aplicadas em outros trabalhos. Entretanto, considerando os custos para obtenção de dados de inventário florestal, é interessante desenvolver técnicas que permitam estimar valores fora do intervalo definido para os dados de treinamento, o que não foi satisfatoriamente alcançado neste trabalho. Nesse sentido, é importante a sequência de trabalhos que possam esclarecer quais os principais pontos que devem ser considerados para que essa finalidade seja alcançada.

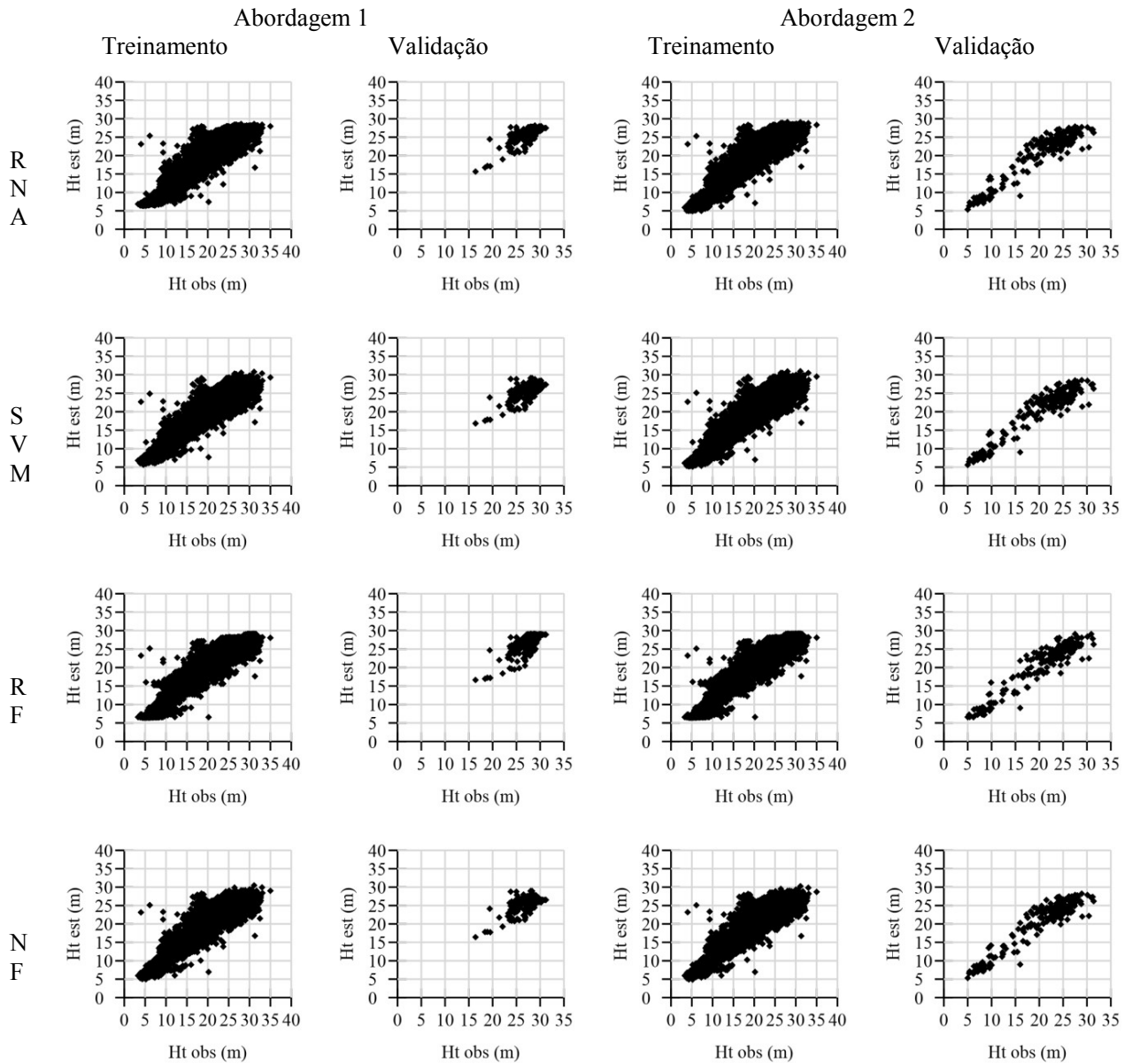


FIGURA 1: Gráficos de dispersão considerando os valores estimados e observados para cada conjunto avaliado.  
 FIGURE 1: Scatter plots considering the estimated and observed values for each evaluated set.

Os resíduos das estimativas se concentraram entre as classes de -25% e 25% para os dados de treinamento em ambas as abordagens. Para os dados de validação, os erros se concentraram nas classes entre -25 e 35%. O que mostra que mesmo apresentando resultados inferiores, a Abordagem 1 pode ser considerada em estudos posteriores que permitam uma melhoria nas estimativas, para todas as ferramentas de inteligência artificial empregadas.

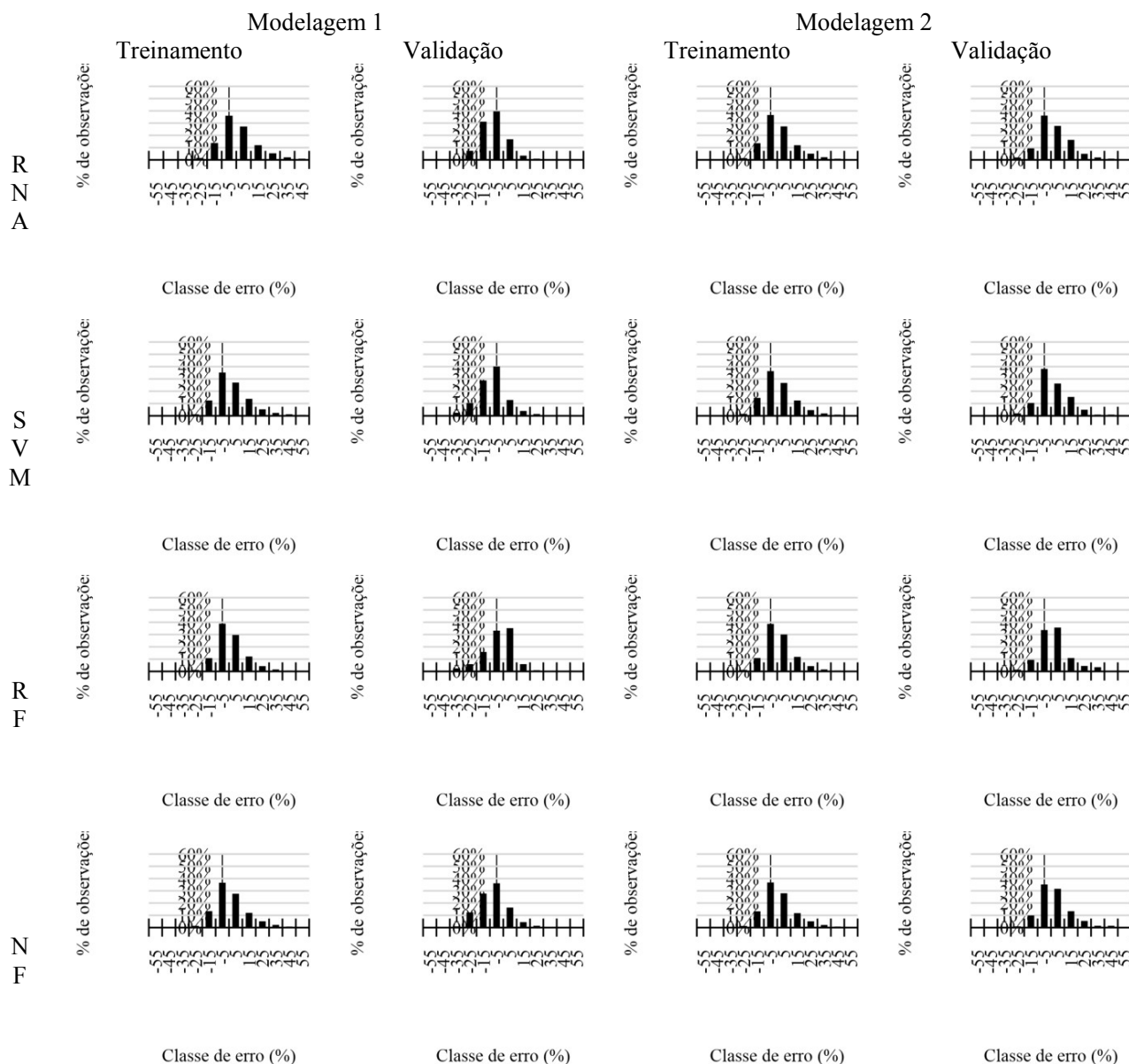


FIGURA 2: Histogramas de resíduos para cada conjunto avaliado.  
 FIGURE 2: Histograms of residues for each evaluated set.

**CONCLUSÕES**

Todas as técnicas avaliadas podem ser utilizadas para estimativa de altura total de árvores de eucalipto. A abordagem de treinamento das ferramentas que considera apenas dados históricos necessita ser estudada de maneira mais aprofundada.

**AGRADECIMENTOS**

Os presentes autores expressam seus agradecimentos à FAPEMIG e ao CNPq pelo auxílio financeiro e ao Instituto de Ciências Agrárias da Universidade Federal de Minas Gerais pelo suporte científico.

**REFERÊNCIAS BIBLIOGRÁFICAS**

ARAÚJO JÚNIOR, C. A. et al. Modelling and forecast of charcoal prices using a neuro-fuzzy system. *Cerne*, v. 22, n. 2, p. 151-158, 2016.

- BINOTI, M.L.M.S. **Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto**. Dissertação (Mestrado em Ciência Florestal), Universidade Federal de Viçosa, Viçosa, 54 f. 2010
- BREIMAN, L. "Random forests." *Machine learning* 45.1 (2001): 5-32.
- DE CASTRO, F. C. C.; DE CASTRO, M. C. F. **Redes neurais artificiais**. Porto Alegre, RS: **Pontifícia Universidade Católica do Rio Grande do Sul**, 2001.
- DE JESUS, C. M. et al. Modelagem hipsométrica em povoamento clonal de *Eucalyptus urophylla* x *Eucalyptus grandis* no Distrito Federal. 2015.
- HAYKIN, S.. *Neural Networks and Learning Machines*. Rev. ed of: **Neural networks**. 2nd ed., 1999.
- RIBEIRO, A, et al. Estratégias e Metodologias de Ajuste de Modelos Hipsométricos em plantios de *Eucalyptus* sp.. **CERNE** [en linea] 2010, 16 (Enero-Marzo) : [Fecha de consulta: 13 de julio de 2018] Disponible en:<<http://www.redalyc.org/articulo.oa?id=74415015003>> ISSN 0104-7760.
- SCHIMIDT, P. B. Determinação indireta da relação hipsométrica para povoamentos de *Pinus taeda*. L. Floresta, v.8, n.8 p.24-27, 1977.
- STEINWART, I.; CHRISTMANN, A. Support vector machines. **Springer Science & Business Media**, 2008.
- VAPNIK , V. N. and CHERVONENKIS ,A. Y. On the uniform convergence of relative frequencies of events to their probabilities. **Theory of Probability and its Applications**, 16(2):283–305, 1971.
- VAPNIK ,V. N.. **Statistical Learning Theory**. John Wiley and Sons, 1998.