

UNIVERSIDADE FEDERAL DO PARANÁ

GUSTAVO FRIEDERICHS

USO DA INTELIGÊNCIA ARTIFICIAL PARA ESTIMATIVA DA PRODUTIVIDADE DE
HARVESTER EM OPERAÇÕES DE COLHEITA FLORESTAL EM SISTEMA DE
TORAS CURTAS

CURITIBA

2016

GUSTAVO FRIEDERICHS

**USO DA INTELIGÊNCIA ARTIFICIAL PARA ESTIMATIVA DA PRODUTIVIDADE
DE HARVESTER EM OPERAÇÕES DE COLHEITA FLORESTAL EM SISTEMA
DE TORAS CURTAS**

Trabalho de Conclusão de Curso apresentado ao Curso de MBA em Manejo Florestal de Precisão do Programa de Educação Continuada em Ciências Agrárias, da Universidade Federal do Paraná, como requisito para a obtenção do título de Especialista.

Orientador: Prof. Dr. Jaime Wojciechowski

Coorientador: Prof. Dr. Jean A. Sampietro

CURITIBA

2016

DEDICATÓRIA

A Deus, pela saúde e força. À minha família, pelo apoio e incentivo durante a minha caminhada,

Dedico.

AGRADECIMENTOS

Inicialmente, eu agradeço à minha família, sempre preocupada com o hoje e o amanhã, ensinando a aprender com o ontem.

À Nathalia, que além de companheira e fiel confidente, foi incentivadora da minha formação.

À Lwarcel Celulose por proporcionar experiência profissional e o desenvolvimento do presente estudo.

Aos orientadores, Professores Jaime Wojciechowski e Jean Sampietro, que além do suporte, dedicaram um tempo especial na construção do presente estudo.

Aos professores do presente MBA pelos ensinamentos repassados. Aos alunos do curso e companheiros do encontro presencial, que atuam em diferentes realidades e proporcionaram agregação de conhecimentos.

À Engenharia Florestal, por me deixar ver com outros olhos a riqueza das florestas e recursos naturais, tornando-me um agente promotor da sustentabilidade.

A todos envolvidos direta e indiretamente,
Eu agradeço.

*The way we do small things
determines the way we do
everything.*

Robin Sharma

Resumo

Uso da inteligência artificial para estimativa da produtividade de *Harvester* em operações de colheita florestal em sistema de toras curtas

O uso de técnicas de inteligência artificial tem ganhado maior atenção de estudiosos da ciência florestal nos últimos anos. As áreas de manejo florestal, biometria, operações florestais e sistemas de informações geográficas possuem estudos exploratórios e aplicados com essas técnicas, em especial com as redes neurais artificiais. O objetivo do presente estudo foi estimar a produtividade da máquina *Harvester* em sistema de colheita *cut-to-length*, através da aplicação de algoritmos de inteligência artificial, e comparação com o método de regressão. Os dados operacionais foram coletados na colheita de povoamentos florestais na região Centro-oeste de São Paulo. Os algoritmos de inteligência artificial avaliados foram *Random Forests*, *Data Mining* (k-vizinhos mais próximos-KNN) e Redes Neurais Artificiais, os quais foram avaliadas quanto à qualidade estatística dos resultados de produtividade estimada. As técnicas de inteligência artificial apresentaram melhor qualidade estatística em comparação ao método convencional, em destaque para Redes Neurais Artificiais e *Random forests* com os menores erros de predição e distribuição de resíduos.

Palavras-chaves: operações florestais, algoritmos, planejamento operacional.

Abstract

Artificial intelligence techniques for predicting the productivity of *Harvester* in cut-to-length timber harvesting system

Artificial intelligence techniques have been studying and drawing attention from researches of forest science in the last years. Forest management, dendrometry, forest operations and geographic information systems have exploratory and applied projects on these techniques, especially artificial neural networks. This study aims to estimate the productivity of *Harvester* machine in cut-to-length timber harvesting system through artificial intelligence algorithms and comparing with regression method. Data were collected from the harvesting process in planted forest stands in Midwest of State of São Paulo, Brazil. The artificial intelligence algorithms applied were *Random forests*, *Data mining* (k-nearest neighbor) and Artificial neural networks and evaluated on its statistics quality of estimation of the *Harvester* productivity. Artificial intelligence showed better statistical quality than equation, especially artificial neural networks and random forests with lower prediction error and residue distributions.

Keywords: forest operations, algorithms, operational planning.

SUMÁRIO

1. INTRODUÇÃO.....	7
1.1 Contextualização.....	7
1.1.1 A colheita florestal e o planejamento.....	7
1.1.2 Inteligência artificial.....	8
1.2 Justificativa.....	10
1.3 Objetivos.....	10
1.3.1 Geral.....	10
1.3.2 Específicos.....	10
2. MATERIAL E MÉTODOS.....	11
2.1. Área de estudo.....	11
2.2. Sistema de colheita.....	12
2.3. Dados de estudo.....	13
2.3.1. Floresta.....	14
2.3.2. Máquina.....	15
2.3.3. Planejamento.....	15
2.4. Estimativas de produtividade.....	16
2.4.1. Método Convencional (EQ).....	16
2.4.2. Data Mining (DM).....	17
2.4.3. Random forests (RF).....	17
2.4.4. Redes neurais artificiais (RNA).....	18
2.5. Processamento dos dados.....	18
2.6. Análise dos critérios estatísticos.....	19
3. RESULTADOS E DISCUSSÃO.....	21
4. CONCLUSÕES.....	33
5. CONSIDERAÇÕES FINAIS.....	34
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	35

1. INTRODUÇÃO

1.1 Contextualização

1.1.1 A colheita florestal e o planejamento

O Brasil é um dos maiores produtores de florestas plantadas do mundo, apresentando ainda a maior produtividade e um dos menores custos de produção da madeira. Do total de 7,8 milhões de hectares plantados, cerca de 54% são cultivados por empresas de celulose e papel, painéis e chapas, e carvão vegetal (INDÚSTRIA BRASILEIRA DE ÁRVORES – IBÁ, 2016).

Segundo Silva (2015), uma parcela significativa do custo total da madeira entregue no pátio das empresas está associada às atividades e operações da colheita e transporte florestal. As operações da colheita florestal têm início no corte da floresta e prosseguem até o carregamento da madeira nos caminhões de transporte. Dentre as máquinas e equipamentos que realizam as operações de colheita, se destaca o *Harvester*, o qual é capaz de realizar as etapas de derrubada, desgalhamento, descascamento, traçamento e pré-extração ou empilhamento da madeira, podendo ser considerada a principal máquina da colheita florestal em sistemas *cut-to-length* (toras curtas).

De acordo com Malinovski et al. (2006), cerca de 35 variáveis influenciam no rendimento operacional (medido pela variável produtividade efetiva) de *Harvesters* durante a operação de colheita florestal, estando relacionadas ao povoamento, terreno e planejamento das operações. Nesse sentido, o planejamento da colheita florestal busca calcular todas as variáveis que influenciam nas operações, resultando em mapas de microplanejamento, dimensionamento de máquinas, tempos e equipes.

Assim, com o planejamento é possível determinar o desembolso financeiro para as operações, visando a redução de custos, meta fundamental para qualquer empresa florestal que visa manter ou ganhar competitividade. Atualmente, um dos grandes desafios está em otimizar as operações da colheita florestal, racionalizando os recursos envolvidos e reduzindo os impactos ambientais.

Nesse sentido, os sistemas computacionais são importantes ferramentas

para auxiliar no planejamento e na redução dos custos na colheita florestal. O investimento elevado em máquinas na colheita da madeira sem um planejamento adequado da operação e sem a aplicação de ferramentas tecnológicas, pode resultar em um aumento nos custos e problemas de gestão das operações (MALINOVSKI et al., 2008).

1.1.2 Inteligência artificial

Inteligência artificial refere-se ao o estudo e projeto de agentes inteligentes, em que o agente inteligente é um sistema que entende o se ambiente e toma decisões baseadas neste e maximiza suas chances de resolver problemas (MCCARTHY, 2002). As ramificações da inteligência artificial são apresentadas pela Figura 1.

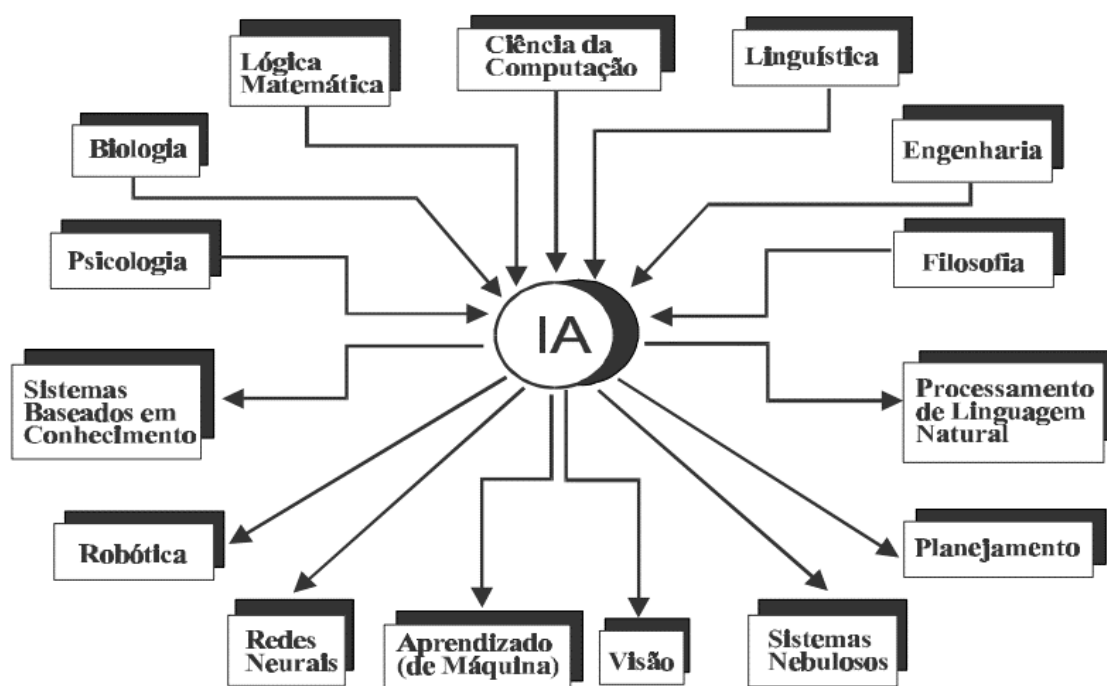


Figura 1. Áreas e ramificações da inteligência artificial segundo MONARD E BARANAUKAS (2000).

As Redes Neurais Artificiais (RNA) e o Aprendizado de Máquina (AM) são duas grandes ramificações da Inteligência Artificial.

As RNA são algoritmos baseados em redes biológicas de pensamento do ser humano. De acordo com Bullinaria (2014), a rede neural é composta por diversas unidades de processamento simples e se formam em camadas de redes. De acordo com Eraslan (2009), durante o treinamento, a RNA otimiza as ligações entre as unidades até a minimização do erro das previsões, atingindo a precisão especificada previamente.

Segundo Santos (2001) e Silva (2015), os algoritmos de RNAs são muito úteis e possuem grande eficiência, principalmente para problemas em que os seus processos são de difícil estimação através das equações físicas. Para Leite et al. (2015), a aplicação de RNA em empresas florestais tem se tornado cada vez maior e tornando-se uma ferramenta estratégica no planejamento. Diversos trabalhos têm avaliado a RNA na melhoria de precisão de estimativas e redução de custos na ciência florestal, como os autores Carvalho (2005), Soares (2011), Oliveira (2013), Binoti et al. (2014a), Binoti et al. (2014b), Silveira (2014), Miguel (2015), Silva (2015) e Schikowski (2016), os quais realizaram estudos com RNA em manejo florestal, biometria, operações florestais e sistemas de informações geográficas.

O Aprendizado de Máquina (AM) inclui técnicas capazes de realizar a estimação de variáveis através de experiências passadas, baseado em funções de proximidade ou hipóteses (Faceli et al., 2001; Schikowski, 2016). Diversas são as técnicas de AM disponíveis, dentre elas a *Data mining* ou mineração de dados e *Random forests* ou florestas aleatórias.

Data Mining refere-se ao processo de exploração de informações em um banco de dados. De acordo com Tan et al. (2009), a mineração de dados identifica padrões desconhecidos e descobre relações entre as variáveis do banco de dados. Os autores Wojciechowski (2015) e Schikowski (2016) realizaram estudos voltados a determinação de estoque de madeira e carbono em povoamentos e modelagem individual de árvores.

As *Random forests* são árvores de tomada de decisão baseadas na divisão binária das variáveis de entrada. Cada galho da árvore é denominado de nó, que treinado objetiva formar outros nós, resultando em estimativas da variável dependente em função da sua correlação com o banco de dados. De acordo com Schikowski (2016), as *Random forests* consistem de um conjunto de árvores de

regressão que são treinadas randomicamente e cada árvore indica sua decisão sobre o valor do objeto.

1.2 Justificativa

Sabendo dos resultados promissores da aplicação da inteligência artificial na ciência florestal, em especial para biometria e manejo das florestas, e da importância do planejamento da colheita florestal, as técnicas de inteligência artificial podem e devem ser avaliadas no âmbito de estimativa de variáveis das operações florestais.

Diferentes técnicas de inteligência artificial podem ser aplicadas, tais como Redes Neurais Artificiais, *Data Mining* (mineração de dados) e *Random Forests* (florestas aleatórias).

1.3 Objetivos

1.3.1 Geral

Estimar a produtividade de cabeçote *Harvester* em sistema de colheita *cut-to-length* através da aplicação de técnicas de inteligência artificial e regressão linear.

1.3.2 Específicos

Aplicar três técnicas de inteligência artificial para a estimação da produtividade de *Harvester* na colheita de toras curtas de Eucalipto;

Comparar as técnicas de inteligência artificial com a metodologia convencional (regressão linear) quanto à qualidade estatística;

Identificar e definir a(s) técnica(s) com melhor qualidade estatística na estimativa da produtividade de *Harvester*; e

Aplicar as técnicas avaliadas em três talhões florestais, observando o *gap* existente entre o tempo planejado e observado.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

O presente estudo foi conduzido em uma empresa florestal no Centro-oeste do estado de São Paulo. Os dados de colheita florestal foram coletados em operações de colheita em talhões florestais nos municípios paulistas de Duartina, Lençóis Paulista, Pratânia, Piratininga, Avaí e Cabrália Paulista, localizados geograficamente pela Figura 1.

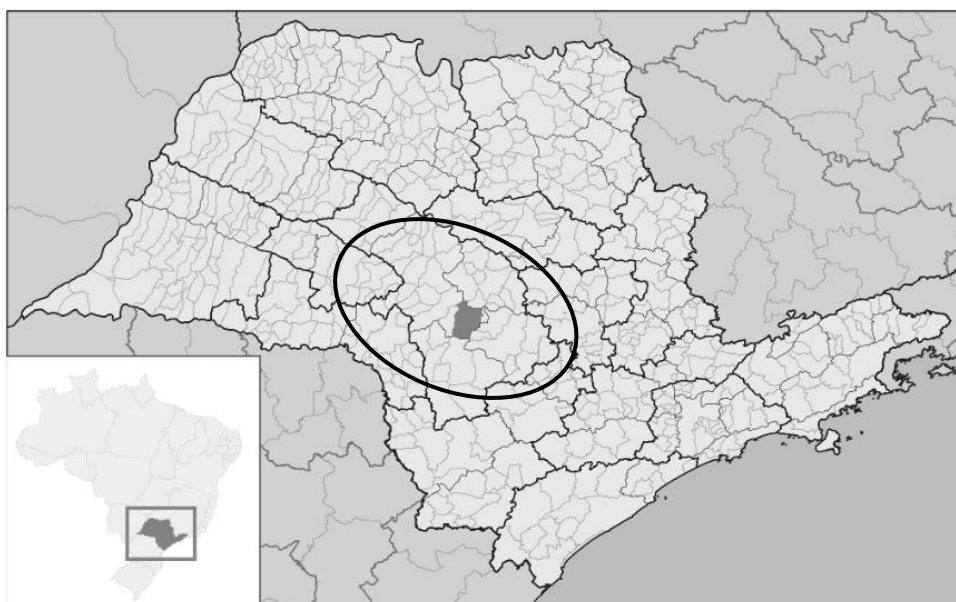


Figura 2. Localização da cidade de Lençóis Paulista e região, Centro-oeste do Estado de São Paulo.

Fonte: Abreu (2006).

Os municípios alvos do estudo apresentam precipitação média anual de 1.357 mm, temperaturas média mínima de 16,3 °C e média máxima de 22,8 °C (Figura 3).

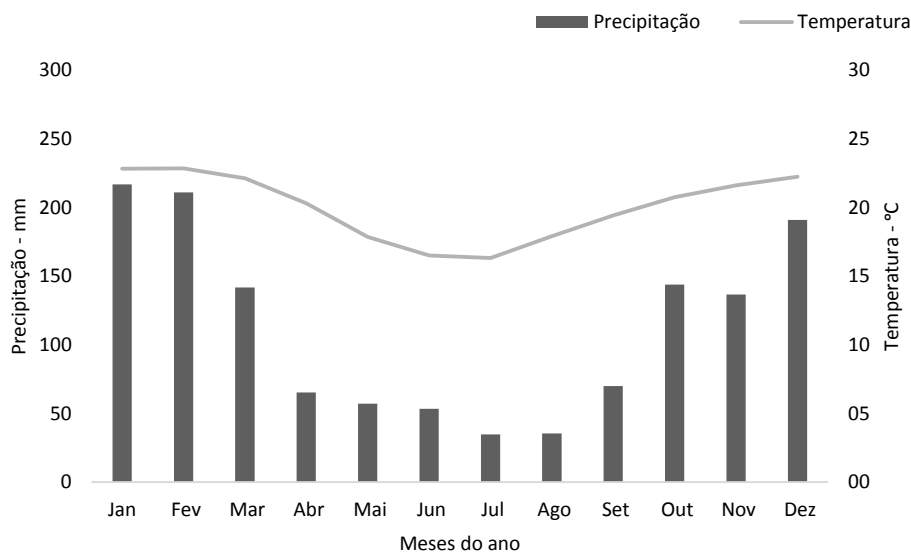


Figura 3. Precipitação e temperatura média da região de estudo.

Fonte: Adaptado de Alvares et al. (2013).

A espécie estudada foi o *Eucalyptus* spp., com plantios florestais largamente cultivados na região para diversos fins. As áreas em estudo possuíam plantios da espécie voltados à produção de celulose de fibra curta branqueada.

2.2. Sistema de colheita

A colheita florestal estudada apresenta operação em módulo máquinas *Harvester* e *Forwarder*. O sistema de colheita estudado foi o *cut-to-length*. Nesse sistema, a máquina *Harvester* realiza as atividades de derrubada das árvores com posterior desgalhamento, descascamento, destopamento, traçamento e empilhamento das toras. As operações de carregamento, baldeio e descarregamento da madeira até a margem do talhão foram realizadas pelo o trator florestal *Forwarder*. A composição de máquinas do sistema de colheita é apresentada na Figura 4.



Figura 4. Modal de colheita florestal. A. Máquina *Harvester* sobre esteiras, modelo 2154D. B. Trator florestal de baldeio da madeira colhida, modelo 1910E.

Fonte: Adaptado de John Deere (2016a) e John Deere (2016b).

O produto final do sistema de colheita de toras curtas é a madeira disposta na forma de pilhas na margem do talhão, para posterior carregamento em veículos para realização do transporte principal.

Durante o período de avaliação, as máquinas *Harvester* operaram com entrada no talhão para realizar o corte em eitos de 3 a 4 linhas de árvores, movimentando-se para frente sobre a segunda linha do eito. A disposição das pilhas de toras era realizada ao lado esquerdo da máquina visando facilitar o trabalho a operações subsequente de extração da madeira. Esta era realizada por tratores florestais *Forwarder* sobre a linha que possui os resíduos de copa e descascamento da madeira colhida, minimizando os impactos sobre solo e, conseqüente, compactação do solo. O intervalo de até 30 dias ocorre entre a operação das máquinas *Harvester* e *Forwarder*.

2.3. Dados de estudo

As variáveis utilizadas para as estimativas foram definidas com base na sua influência na produtividade de *Harvester* conforme Malinovski et al. (2006).

Com base na classificação das variáveis de influência na operação de *Harvester*, as classes de variáveis no presente estudo foram: variáveis da Floresta (Idade, Volume médio individual e Densidade de árvores), Máquina (código da máquina na Empresa) e Planejamento (Bimestre de colheita, Equipe e Sortimento).

A variável dependente a estimada foi a produtividade das máquinas *Harvesters*. A Tabela 1 apresenta o exemplo do cabeçalho do banco de dados do presente estudo.

Tabela 1. Variáveis utilizadas para a formação do banco de dados.

Talhão	Floresta			Máquina	Planejamento			Produtividade observada (m ³ /he)
	Idade (anos)	VMI (m ³ /árvore)	N (árvores/ha)	Código da máquina	Bimestre (1 a 6)	Equipe (turno)	Sortimento (B, C)	
451109	7,2	0,323	1.062	0850	2	D	C	33,16
613051	8,2	0,336	1.065	1120	4	A	C	37,06
...

Onde: VMI: volume médio individual, m³/árvore; N: densidade, árvores/hectare; Bimestre: bimestre de colheita da floresta, em que 1 (Jan, Fev), 2 (Mar, Abr), 3 (Mai, Jun), 4 (Jul, Ago), 5 (Set, Out) e 6 (Nov, Dez); Sortimento: dimensões e destino da madeira colhida, B (biomassa) e C (celulose).

As variáveis da classe Floresta foram obtidas por meio de inventários florestais pré-corte e cadastro florestal. A área basal média dos talhões florestais foi 27,3 m²/ha, volume médio total de 412 m³/ha.

As variáveis das classes Máquina, Planejamento e Produtividade observada foram obtidas através dos computadores de bordo das máquinas avaliadas. O computador de bordo registra diversas informações operacionais e de rendimento.

2.3.1. Floresta

As variáveis utilizadas no estudo quanto na classe Floresta referem-se aos povoamentos, como a idade (do plantio a colheita), o volume médio individual das árvores (VMI) e a densidade de árvores (N). As variáveis foram definidas após uma revisão bibliográfica do estudo realizado por Malinovski et al. (2006), em que são definidas as principais variáveis de influência na produtividade da colheita da madeira.

De acordo com a idade da floresta tem relação indireta com a produtividade da colheita florestal, uma vez que quanto mais velho o povoamento, maior será o VMI, menor será a densidade de árvores (em decorrência de mortalidades) e maior será a probabilidade de ocorrer sub-bosque, dificultando a operação. O volume médio individual tem relação direta e positiva com a produtividade da máquina *Harvester*, conforme observado por Bramucci & Seixas (2002), em que a variável VMI explica em grande parte a produtividade da operação. Ainda os mesmos autores destacam que a densidade de árvores tem relação direta com a produtividade de *Harvester*, e que implica na redução do VMI das árvores, porém há uma faixa ótima de densidade (800 a 1.200 árvores por hectare).

2.3.2. Máquina

Conforme Bramucci & Seixas (2002) o modelo das máquinas, características construtivas e vida útil influenciam na produtividade de *Harvester*. No presente estudo foi considerada apenas uma variável, o código de identificação da máquina.

2.3.3. Planejamento

As variáveis relacionadas ao planejamento aplicadas foram o bimestre de colheita de cada talhão, a equipe (turno) de colheita e o sortimento planejado para cada talhão.

De acordo com Malinovski et al. (2006), a precipitação tem influência sobre o desempenho das máquinas. O efeito da precipitação foi englobado em classes de bimestre de colheita, onde ocorrem influências negativas de períodos mais chuvosos sobre a produtividade das máquinas *Harvester*, logo que os meses de maior precipitação apresentam maiores restrições para deslocamento das máquinas, porém, maior facilidade no descascamento da madeira.

Equipes com mais experiência possuem melhor desempenho na operação de *Harvester*. Para o presente estudo a experiência dos operadores foi englobada em equipes de colheita.

O sortimento planejado para a madeira possui influência direta com a produtividade das máquinas *Harvester*, como exemplo, a produção de sortimento

para uso energético (biomassa) possui uma etapa a menos na colheita com *Harvester*. A não necessidade da etapa de descascamento implica na elevação da produtividade, conforme observado por Bramucci & Seixas (2002).

2.4. Estimativas de produtividade

De posse do banco de dados com todas as variáveis de estudo, foram realizadas as estimativas de produtividade, ou também conhecido por rendimento operacional da máquina *Harvester*.

As estimativas de produtividade foram realizadas por meio de regressão e três técnicas de inteligência artificial: *Data mining*, *Random Forests* e Redes Neurais Artificiais.

Os dados foram aplicados em todas as técnicas de estimação em dois conjuntos, sendo um de treinamento (70% do volume de dados) e de validação (30% do volume de dados). Os dados pertencentes a cada conjunto de dados foram definidos aleatoriamente na base de dados inicial (100% dos dados) e aplicados sistematicamente para as técnicas, ou seja, os mesmos dados do conjunto de treinamento foram aplicados em todos os algoritmos e regressão.

2.4.1. Método Convencional (EQ)

A produtividade de máquina *Harvester* foi determinada através da Equação 1.

$$Pr = \frac{V}{He} \quad (1)$$

Onde: Pr: produtividade efetiva, m³/hora efetiva; V: volume de madeira colhida e processada, m³; He: horas efetivas de trabalho por turno, horas.

A metodologia convencional considerada no presente estudo foi a técnica de regressão linear. O modelo escolhido para o ajuste dos dados é apresentado pela Equação 2.

$$Pr = b_0 + b_1 * I + b_2 * VMI + b_3 * N \quad (2)$$

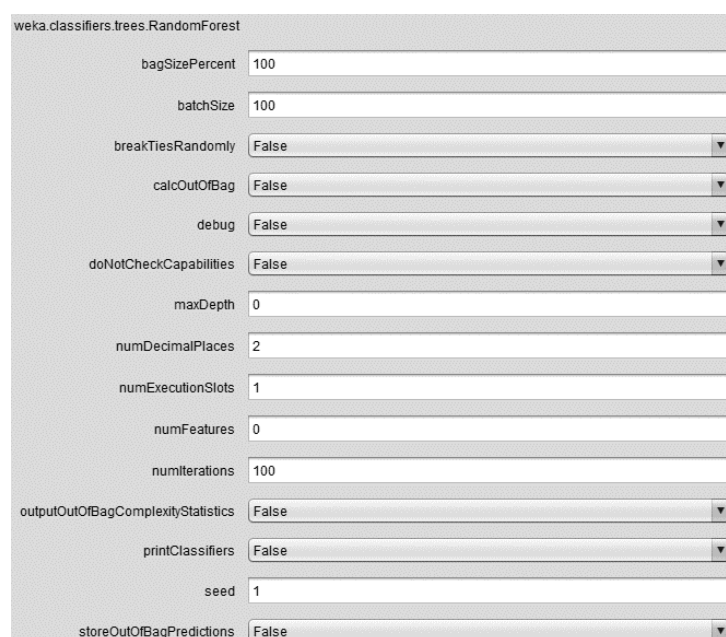
Onde: Pr: produtividade efetiva, m³/hora efetiva; I: idade da floresta, anos; VMI: volume médio individual, m³/árvore; N: densidade de árvores, árvores/hectare.

2.4.2. Data Mining (DM)

No presente estudo foi aplicado e avaliado o algoritmo k-NN (*k-nearest neighbor*) na estimação da variável produtividade *Harvester*. O software utilizado foi o JCarbon e as configurações foram: 3 vizinhos mais próximos e ponderação da distância euclidiana ($1/d^5$). Além das configurações definidas, foram avaliadas outras configurações de distâncias (Euclidiana, Quadrática, Manhathan e Chebychev), vizinhos k-NN (1, 3, 5, 7, 9 e 11) e ponderação (sem, $1/d$, $1/d^2$, $1/d^3$, $1/d^4$ e $1/d^5$) e suas combinações, totalizando 104 simulações e definida a de melhor qualidade estatística.

2.4.3. Random forests (RF)

O algoritmo *Random forests* foi aplicado e avaliado no presente estudo com as configurações básicas do software Weka utilizado, conforme a Figura 5. As configurações abaixo representaram a única simulação realizada para o algoritmo.



The image shows a screenshot of the Weka software interface for configuring the Random Forest classifier. The window title is "weka.classifiers.trees.RandomForest". The configuration parameters are as follows:

Parameter	Value
bagSizePercent	100
batchSize	100
breakTiesRandomly	False
calcOutOfBag	False
debug	False
doNotCheckCapabilities	False
maxDepth	0
numDecimalPlaces	2
numExecutionSlots	1
numFeatures	0
numIterations	100
outputOutOfBagComplexityStatistics	False
printClassifiers	False
seed	1
storeOutOfBagPredictions	False

Figura 5. Configurações de software *Weka* para o algoritmo *Random forests*.

2.4.4. Redes neurais artificiais (RNA)

A RNA aplicada e avaliada no presente estudo foi a MLP (*Multilayer Perceptron*), com o algoritmo *Resilient propagation* treinado e validado na camada intermediária de 10 neurônios e função sigmoideal, e com a camada de saída na função sigmoideal. Além das configurações definidas, foram realizadas simulações alterando configurações do software, tais como o tipo de treinamento (*resilient propagation*, *backpropagation* e RBF), camada intermediária (sigmoideal), neurônios (7, 8, 10 e 14) e camada de saída (sigmoideal e hiperbólica), sendo escolhida a de melhor qualidade estatística.

O software utilizado para o processamento e aplicação da RNA foi o Neuro4, de origem Brasileira (Figura 6).



Figura 6. Software Neuro4 de processamento de Redes Neurais Artificias.

2.5. Processamento dos dados

Para as estimativas, o banco de dados apresentou apenas dados consistentes ao objetivo das estimativas, sendo excluídos aqueles com comportamento fora do esperado (*outliers*). Paralelamente foi realizada a estatística descritiva da base de dados.

O processamento dos dados foi realizado através de *softwares* específicos para cada algoritmo, conforme abaixo:

- *Microsoft Excel*: compilação das informações, formação do banco de dados e análise das variáveis. O *Microsoft Excel* é um editor eletrônico de planilhas, que inclui a possibilidade da geração de cálculos e gráficos (MICROSOFT, 2012). A versão utilizada foi 2013.
- *JCarbon*: processamento de dados em algoritmos *data mining*. O *software JCarbon* foi desenvolvido na Universidade Federal do Paraná e disponível na plataforma web (www.jcarbon.ufpr.br), com um banco de dados de locais cadastrados com informações florestais, equações e modelos e propôs diversas alternativas para a realização de estimativas de volume, biomassa e carbono, cobrindo as principais necessidades do usuário (WOJCIECHOWSKI, 2015). A versão utilizada do *software* foi JCarbon 1.0.
- *Weka*: processamento de dados em algoritmos *Random forests*. O *Weka* é uma coleção de algoritmos de aprendizado de máquina para a mineração de dados, além de permitir o desenvolvimento de novos aprendizados. Desenvolvido na Universidade de *Waikato*, Nova Zelândia, atualmente é mantida por uma comunidade de entusiastas por ser um software livre disponível sobre a licença GPL (HALL et al., 2009). A versão utilizada do *software* foi *Weka* 3.8.
- *Neuro4*: processamento de dados em algoritmos de redes neurais artificiais. O *Neuro4* é a atualização do *NeuroForest*, *software* Brasileiro e desenvolvido para a geração e aplicação de RNA em diversos treinamentos e estruturas de algoritmos desejados. A versão utilizada do *software* foi *NeuroForest* 4.0.2.

2.6. Análise dos critérios estatísticos

As estimativas geradas através das técnicas de estimação da produtividade foram avaliadas quanto à qualidade estatística. A qualidade foi determinada através dos critérios estatísticos, sendo: erro quadrático médio absoluto e relativo da variável estimada (EQUAÇÃO 3), correlação de *Pearson* entre os dados estimados e observados (EQUAÇÃO 4), resíduos percentuais das estimativas (EQUAÇÃO 5) e análise gráfica de resíduos e ajustes.

$$RMSE\% = \frac{100}{\bar{y}} * \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (3)$$

Onde: RMSE%: erro quadrático médio, %; \bar{Y} : média observada; Y_i : valores observados; \hat{Y}_i : valores estimados; n: número de observações.

$$r = \frac{cov(Y, \hat{Y})}{\sqrt{S^2(Y) S^2(\hat{Y})}} \quad (4)$$

Onde: r: correlação de *Pearson*; cov (Y, \hat{Y}): covariância. S²: variância.

$$Erro\% = \frac{(\hat{Y} - Y)}{Y} * 100 \quad (5)$$

Onde: Erro%: erro médio, %; Y: valores observados; \hat{Y} : valores estimados.

Os ajustes foram avaliados graficamente conforme a distribuição dos resíduos em relação à variável observada, relação entre os dados estimados e observados, gráfico de distribuição de frequência de resíduos por classes.

3. RESULTADOS E DISCUSSÃO

A estatística descritiva das variáveis discretas está apresentada pela Tabela 2.

Tabela 2. Estatística descritiva das variáveis contínuas e discretas do estudo.

Estatística	Idade (anos)		VMI (m ³ /árvore)		N (árvores/ha)		Produtividade (m ³ /he)	
	T	V	T	V	T	V	T	V
Média	7,4	7,4	0,293	0,295	1.093	1.090	29,2	29,3
Mínimo	6,7	6,7	0,225	0,227	803	788	17,4	17,0
Máximo	8,6	8,5	0,397	0,400	1.277	1.282	44,2	43,0
Amplitude	5,8	5,8	0,389	0,389	1.160	1.160	77,4	52,1
Desvio Padrão	0,7	0,7	0,050	0,049	135	139	7,5	7,3
CV (%)	9,5	8,9	17,0	16,5	12,4	12,8	25,9	24,9
Observações	2.997	1.298	2.997	1.298	2.997	1.298	2.997	1.298

Onde: VMI: volume médio individual, m³/árvore; N: densidade, árvores/hectare; T: dados de treinamento, 70% do banco de dados; V: dados de validação, 30% do banco de dados; CV: coeficiente de variação, %. Mínimo: média dos 10% menores dados observados; Máximo: média dos 10% maiores dados observados; Amplitude: diferença entre o maior e menor valor observado.

Observa-se que a média e as medidas de dispersão da idade, volume médio individual, densidade de árvores e produtividade (todos dados numéricos), apresentaram-se de forma semelhante para o conjunto de dados de treinamento e validação. O número de dados utilizados nas estimativas do presente estudo foi de 4.295, quantidade suficiente para abranger a amplitude de dados da variável estimada no estudo.

A idade base dos talhões florestais colhidos na Empresa de estudo é de 7 anos, havendo ainda talhões com idades mais avançadas, logo os métodos de estimação também foram aplicados para os talhões mais velhos, não inviabilizando as estimativas desses casos. O volume médio individual (VMI) apresentou elevada

variação, fato esse esperado para o elevado número de observações e sítios florestais existentes na região.

Observa-se que todas as variáveis discretas apresentaram elevado desvio padrão e erro padrão da média, caracterizando a elevada amplitude dos dados, tornando as estimativas realizadas com boa capacidade de predição para os valores extremos.

Os critérios estatísticos resultantes das estimativas de produtividade são apresentados pela Tabela 3, para as quatro técnicas de estimação.

Tabela 3. Resultados estatísticos das estimativas de produtividade por inteligência artificial e regressão.

Estatística	RNA		RF		DM		EQ	
	T	V	T	V	T	V	T	V
RMSE (m ³ /he)	5,3	6,0	3,6	5,6	11,7	7,4	6,7	6,6
RMSE (%)	18%	20%	12%	19%	40%	25%	23%	23%
Erro (%)	-3,8%	-4,6%	-2,1%	-4,0%	28,6%	10,5%	-5,7%	-5,7%
Correlação de <i>Pearson</i>	0,69	0,58	0,88	0,65	0,49	0,57	0,43	0,42

Onde: RMSE: erro quadrático médio; Correlação: correlação de *Pearson*; RNA: algoritmo de Rede Neural Artificial; RF: algoritmo de *Random Forests*; DM-kNN: algoritmo de *Data Mining* k-NN; EQ: equação ajustada por regressão; T: dados de treinamento; V: dados de validação.

Em que: $EQ = 23,999 + 56,449 * VMI - 0,006 * N - 0,678 * I$

Relacionando-se os resultados acima à qualidade estatística das estimativas, espera-se que os resultados de RMSE absoluto/relativo e Erro sejam menores possíveis, e a Correlação de *Pearson* seja maior.

Observa-se que o algoritmo de *Random forests* apresentou os menores valores de RMSE e Erro para o treinamento e validação das estimativas, tendo também os melhores valores de correlação de *Pearson*. Esses resultados indicam a melhor qualidade estatística do algoritmo para estimativas da produtividade da máquina *Harvester*.

As estimativas realizadas através de Redes Neurais Artificiais (RNA) apresentaram-se com qualidade estatística ligeiramente inferior àquelas realizadas por *Random forests*. Embora apresente menor qualidade estatística, os critérios estatísticos apresentados demonstram uma boa capacidade de predição da produtividade da máquina *Harvester*, com erro médio de até -5%.

Silva (2015) aplicando diversas RNAs na colheita florestal encontrou valores de RMSE% próximos ao do presente estudo, para a estimativa da produtividade da máquina *Forwarder* (10% a 20% para o treinamento e 17% a 21% para a validação das RNAs).

Destacam-se os valores de correlação para os algoritmos de *Random forests* e RNA, indicando uma boa relação entre os dados observados e estimados para esses algoritmos em treinamento.

Os conjuntos de dados de treinamento e validação apresentaram diferentes resultados para os algoritmos RNA e *Random forests*, havendo redução da qualidade estatística para todos os critérios estatísticos, quando o algoritmo foi validado com uma base de dados diferente. Esse comportamento também foi observado por Silva (2015) analisando a produtividade da máquina *Forwarder* e é justificado por Heaton (2008), onde o algoritmo sofre *over training*, causando a redução da precisão, pois o algoritmo se especializa demais no conjunto de treinamento.

O algoritmo do vizinho mais próximo em *Data mining* apresentou bons valores de correlação para o conjunto de treinamento e validação dos dados, porém elevados valores de RMSE absoluto/relativo e Erro médio. Destaca-se o fato de que o conjunto de validação ter apresentado resultados estatísticos de melhor qualidade, quando comparado do seu conjunto de treinamento. Esse comportamento pode ser explicado pela presença de dados mais próximos no conjunto de validação do que em treinamento, levando a melhores estimativas.

A equação ajustada através da técnica de regressão apresentou uma baixa qualidade estatística no estudo. Por apresentar a capacidade englobar variáveis discretas (Idade, VMI e N) a equação ajustada por regressão apresentou-se na terceira posição em relação qualidade estatística.

Os resultados de RMSE, Erro e Correlação para a equação ajustada tiveram pouca diferença entre os conjuntos de treinamento e validação. Esse

comportamento indica a capacidade do modelo utilizado em generalizar as estimativas com novos dados e qualidade estatística semelhante ao treinamento.

As estatísticas descritivas das estimativas da produtividade são apresentadas na Tabela 4, com valores das medidas de posição e dispersão. A estatística descritiva representa em medidas de posição e dispersão o comportamento dos dados observados e estimados.

Tabela 4. Estatística descritiva dos dados de produtividade observada e estimada (conjunto de validação).

Técnica	Média (m ³ /he)	Mínimo (m ³ /he)	Máximo (m ³ /he)	Amplitude (m ³ /he)	Desvio padrão (m ³ /he)	CV (%)
OBS	29,3	17,1	43,0	25,9	7,3	24,9%
RNA	29,3	25,8	35,9	10,1	4,9	16,7%
RF	29,3	25,0	38,0	13,0	5,8	19,8%
DM	25,3	20,6	32,5	11,9	5,9	23,3%
EQ	29,2	27,6	32,4	4,9	3,1	10,6%

Onde: OBS: dados observados ou reais; RNA: redes neurais artificiais; RF: *Random forests*; DM: *Data mining*; EQ: equação ajustada por regressão; Mínimo: média dos 10% menores dados observados; Máximo: média dos 10% maiores dados observados; CV: coeficiente de variação. O número total de estimativas realizadas foram de 1.298.

O conjunto de validação dos dados representa a aplicação das técnicas avaliadas em dados não utilizados para o ajuste, simulando a operacionalização dos algoritmos. Observa-se que esse conjunto apresenta elevada amplitude e erro padrão da média dos dados observados, indicando que o total de 1.299 dados é capaz de representar a variação da produtividade da máquina *Harvester*.

Os algoritmos avaliados apresentaram valores médios de produtividade estimada próximos da média observada, à exceção de *Data mining* (25,3 m³/he).

De forma prática, o algoritmo que apresentar valores das medidas de posição e dispersão mais próximos dos dados observados, tende a possuir uma maior capacidade de estima-los. O desvio padrão e coeficiente padrão indicam que

os algoritmos *Random forests* e *Data mining* são os mais próximos dos dados observados, com melhor capacidade de estima-los.

A Figura 7 apresenta os gráficos de dispersão dos dados estimados para o conjunto de treinamento. Já a Figura 8 apresenta a dispersão das estimativas para o conjunto de validação.

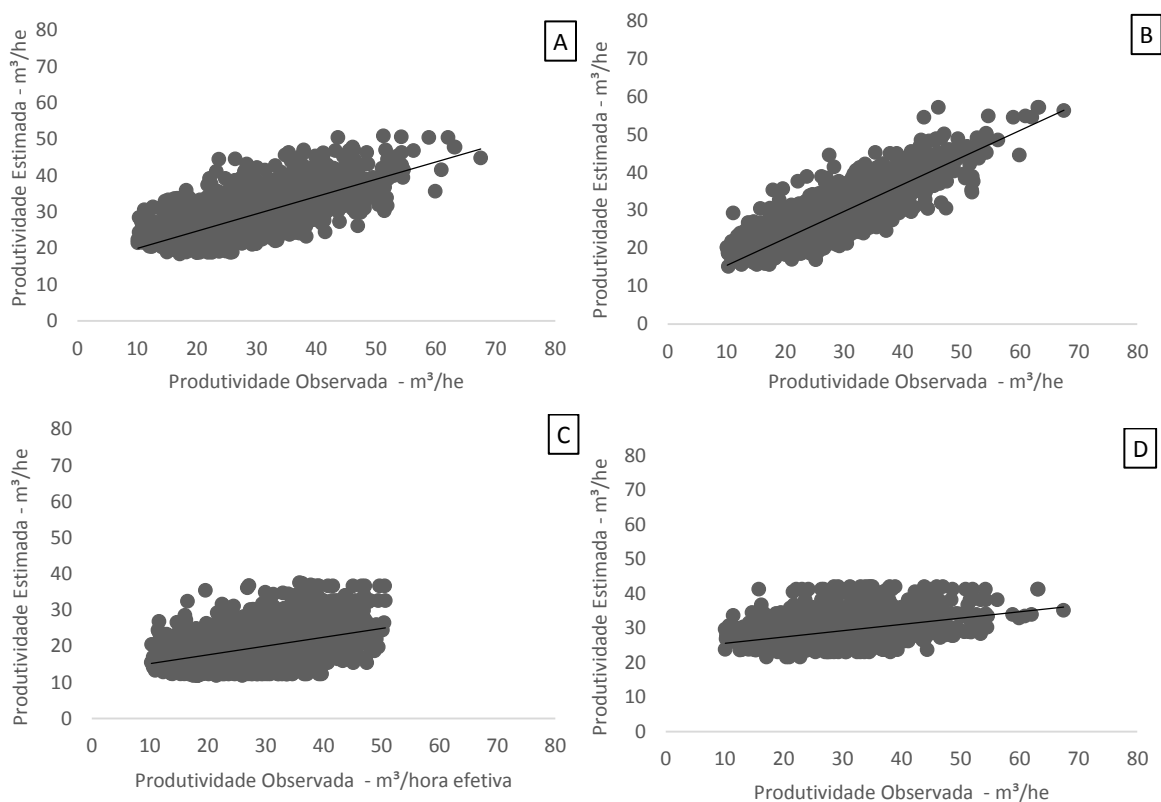


Figura 7. Gráficos de dispersão das estimativas em relação aos dados reais de treinamento (70% da massa de dados). A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

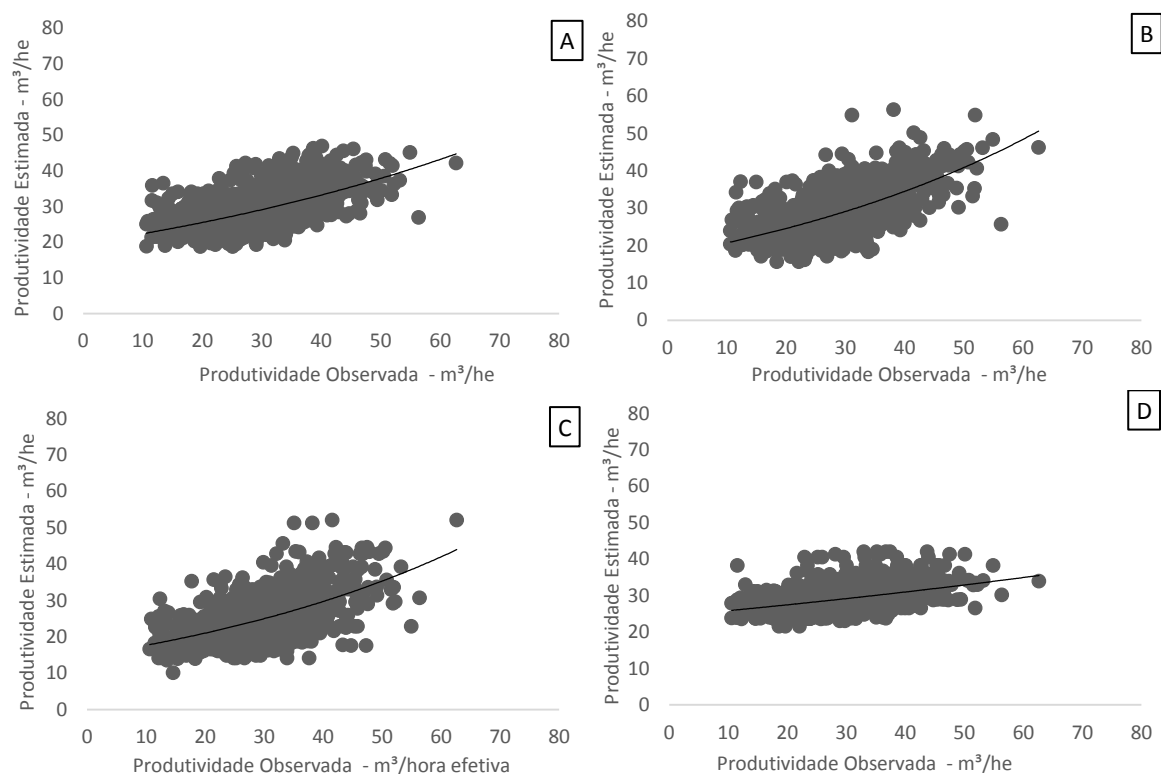


Figura 8. Gráficos de dispersão das estimativas em relação aos dados reais de validação (30% da massa de dados). A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

O conjunto de dados de treinamento de *Random forests* apresentou estimativas com dispersão mais adequadas em comparação às demais técnicas. Já o conjunto de validação não apresentou dispersão dos dados estimados com a mesma qualidade.

Os gráficos de estimativas de RNA e *Random forests* apresentaram-se com dispersão similar, com ascendência dos pontos de dispersão à medida que a variável observada cresce. Esse comportamento não foi observado para o algoritmo de *Data mining* e Equação, em que a dispersão ocorre no sentido horizontal sem ascendência dos pontos.

A ascendência das estimativas observadas graficamente indica a capacidade dos algoritmos de RNA e *Random forests* estimarem a produtividade com menores erros. Esse comportamento é observado para os dois conjuntos de dados avaliados.

Os gráficos de resíduos das estimativas do conjunto de treinamento são apresentados pela Figura 9. A Figura 10 apresenta a dispersão dos resíduos do conjunto de validação.

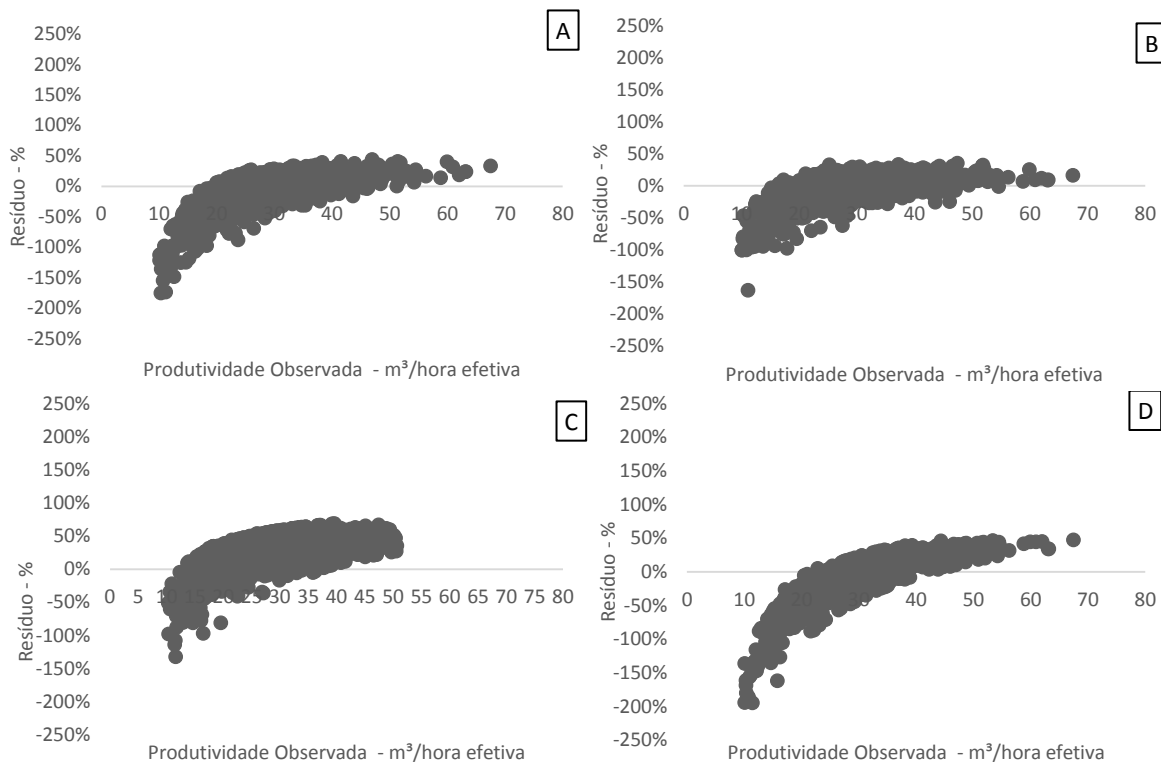


Figura 9. Dispersão dos resíduos das estimativas em função dos dados reais do treinamento (70% da massa de dados). A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

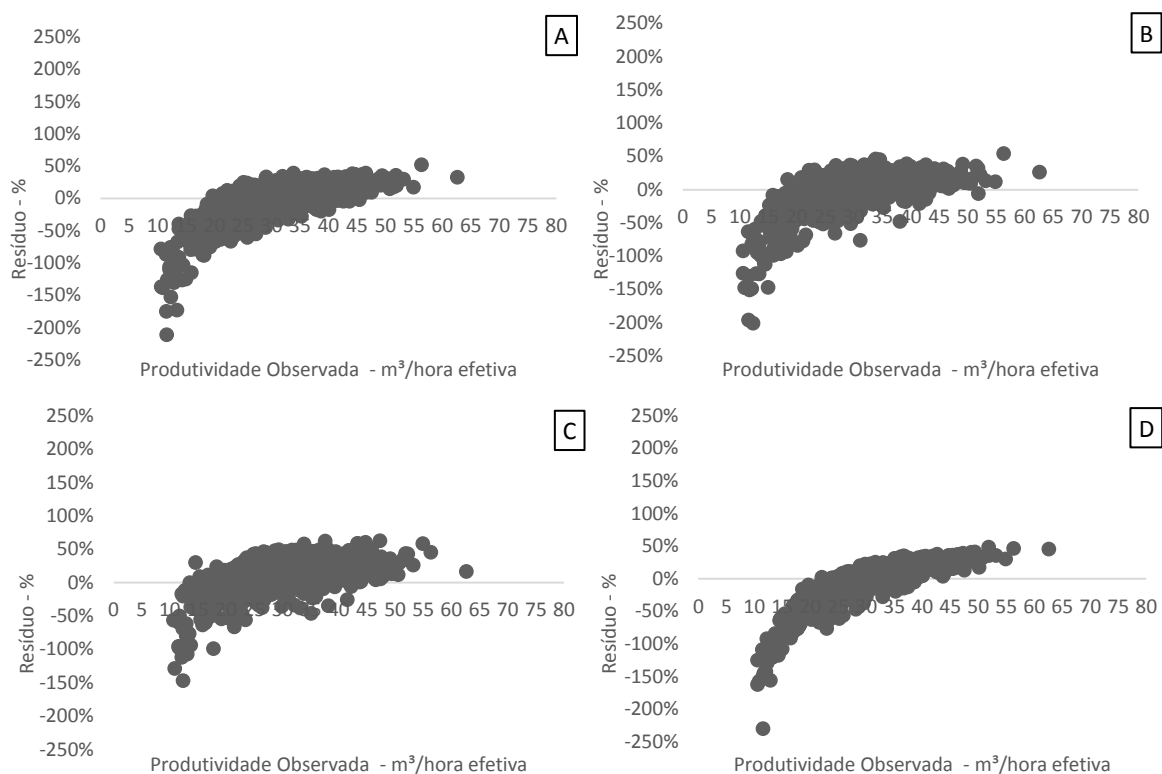


Figura 10. Dispersão dos resíduos das estimativas em função dos dados reais da validação (30% da massa de dados). A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

Observa-se para o conjunto de treinamento que os resíduos mais extremos ocorreram para os dados abaixo de 20 m³/he. Destaca-se que a faixa entre 20 e 60 m³/he apresenta os menores resíduos para RNA e *Random forests*. Já para *Data mining* a faixa de valores com menores resíduos ocorre entre 20 e 35 m³/he, tornando o algoritmo mais limitado para valores elevados de produtividade. A Equação avaliada apresenta como faixa de menores resíduos as produtividades entre 20 e 45 m³/he.

Avaliando a aplicação das técnicas no conjunto de validação, todas apresentaram uma faixa ótima entre 20 e 40 m³/he.

Os gráficos apresentados pela Figura 11 e Figura 12 representam a distribuição de frequência dos resíduos das estimativas realizadas pelas quatro técnicas.

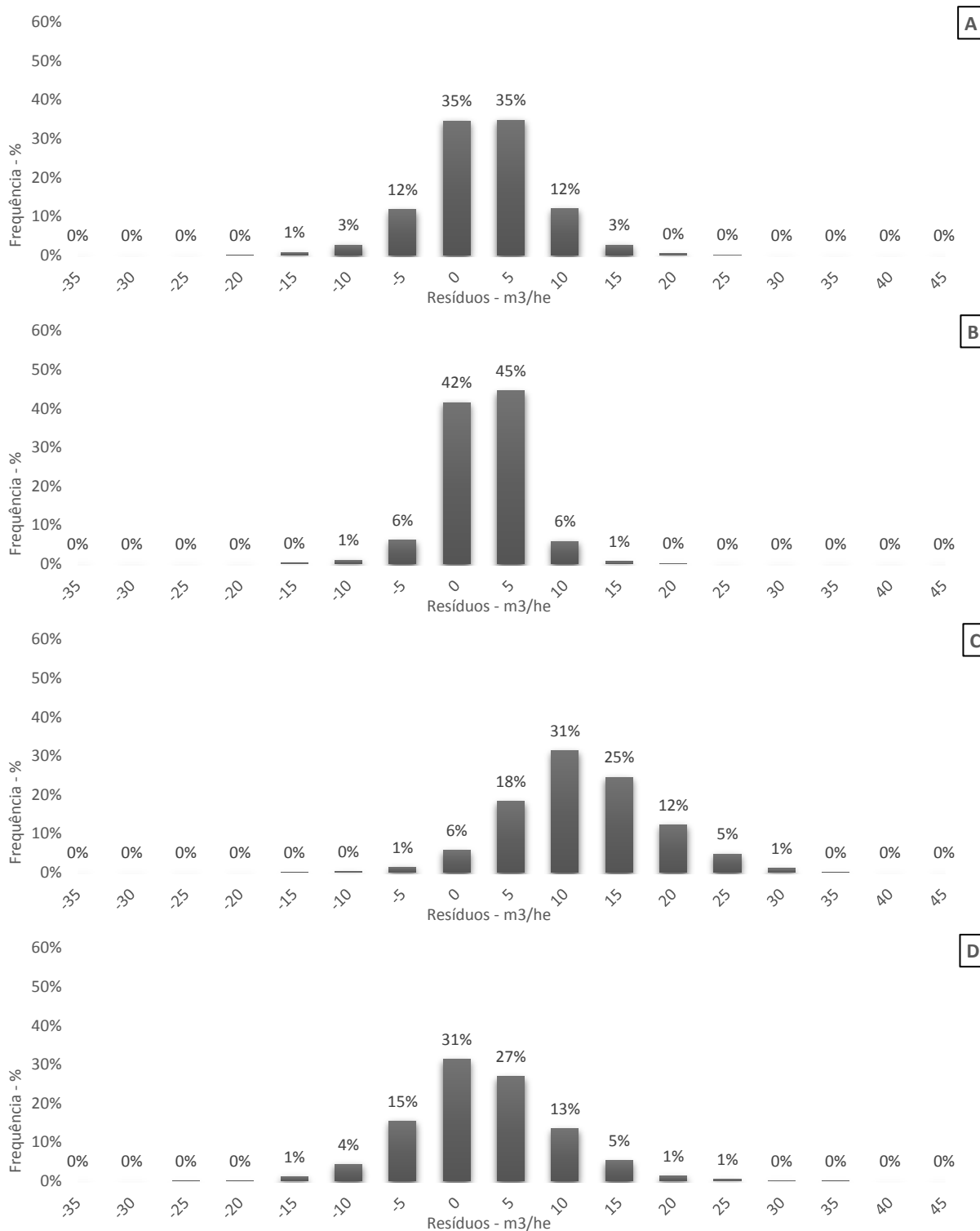


Figura 11. Resíduos e frequência de ocorrência no treinamento das estimativas de produtividade de máquina *Harvester*. A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

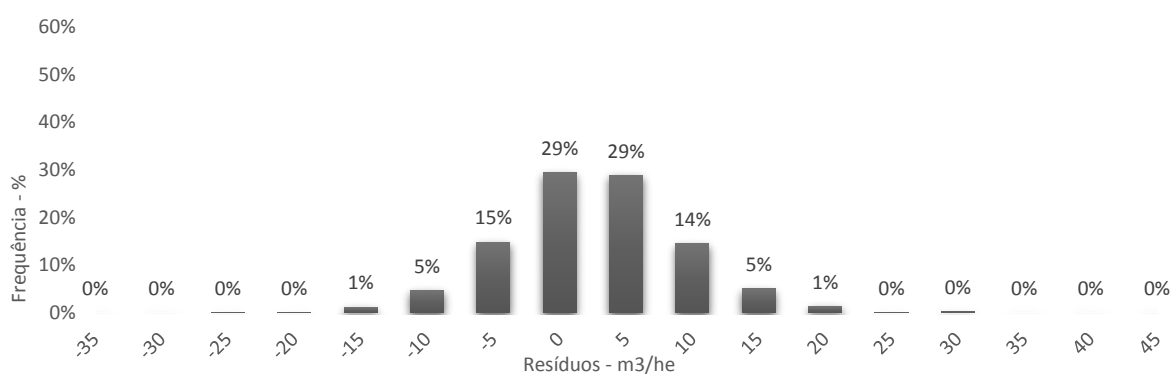
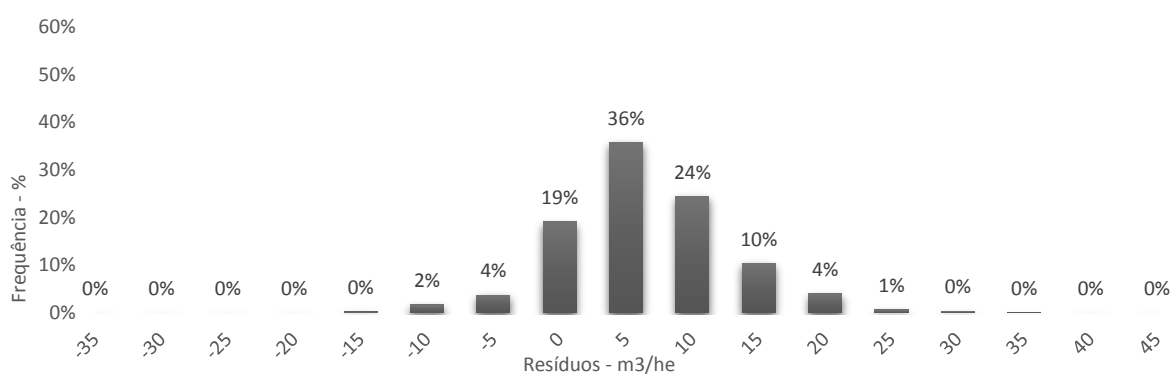
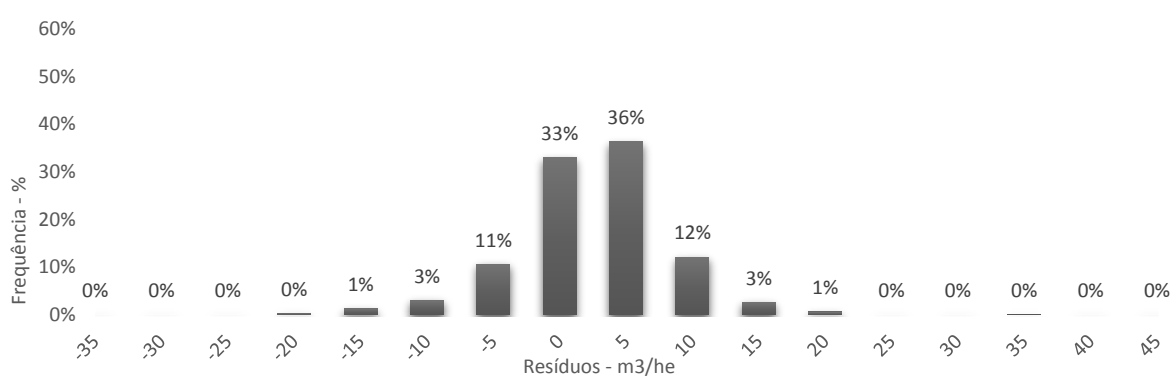
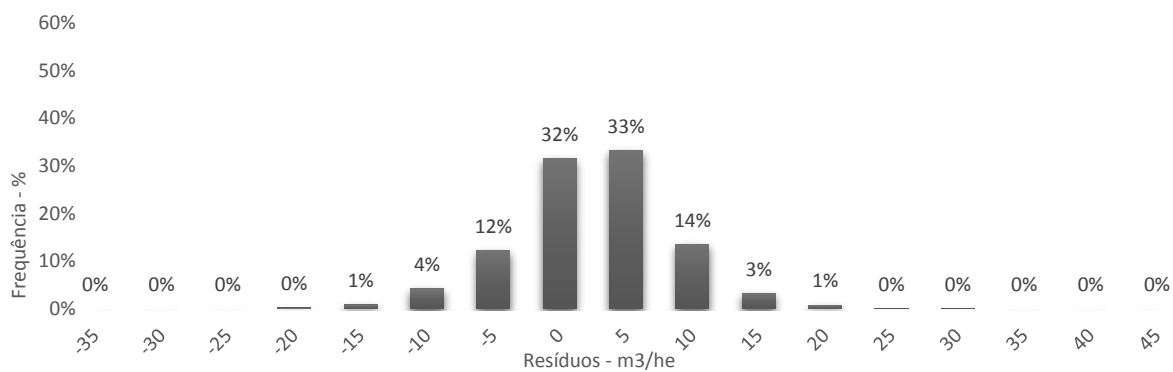


Figura 12. Resíduos e frequência de ocorrência na validação das estimativas de produtividade de máquina *Harvester*. A. Redes neurais artificiais. B. *Random forests*. C. *Data mining*. D. Equação ajustada por regressão.

Observa-se pelos gráficos que a distribuição da frequência de resíduos se comporta de forma similar para os conjuntos de treinamento e validação em todas as técnicas avaliadas. Novamente os algoritmos de RNA e *Random forests* apresentam-se mais semelhantes na distribuição dos resíduos, com maior concentração nas classes de 12,5 a 37,5%.

A técnica de *Data mining* apresentou maiores frequências de resíduos na classe de 62,5% para o conjunto de treinamento, e 37,5% para o conjunto de validação. As classes de ocorrência da maior parte dos resíduos é de 12,5 a 97,5%.

A equação avaliada apresentou resíduos mais distribuídos e em 3 classes, sendo um comportamento não desejado. Espera-se que os resíduos se comportem de forma mais concentrada e em poucas classes.

As técnicas de inteligência artificial aplicadas e avaliadas apresentaram melhor qualidade estatística do que a metodologia convencional (regressão) para a correlação de *Pearson*. A metodologia convencional apresentou maior RMSE absoluto/relativo e Erro relativo do que o algoritmo testado em *Data mining*.

A operacionalização das técnicas de estimação estudados foi realizada através da aplicação dos valores estimados no cálculo do tempo estimado para a colheita de três talhões florestais da empresa. A Figura 13 apresenta o atendimento percentual do tempo de operação planejado para os três talhões florestais.

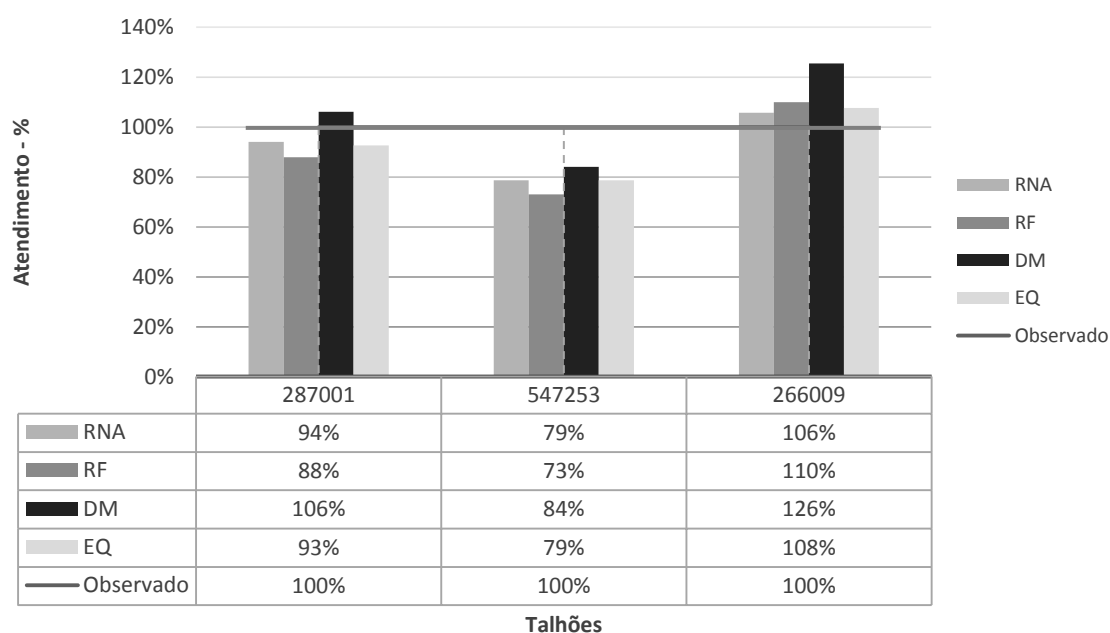


Figura 13. Resíduo do atendimento do tempo planejado para três talhões florestais.

Observa-se que a RNA e EQ ajustada tiveram comportamento semelhante para os três talhões florestais simulados. O algoritmo *Data mining* apresentou tempo planejado superior às demais técnicas para todos os talhões, com tendência a superestimar o tempo planejado para a operação de *Harvester* na colheita florestal.

Analisando a operacionalização dos algoritmos, as técnicas RNA e EQ apresentaram o menor coeficiente de variação para o tempo planejado entre os talhões analisados (abaixo de 16%).

4. CONCLUSÕES

As técnicas de inteligência artificial foram capazes de estimar a produtividade da máquina *Harvester* e possuem qualidade estatística superior método convencional (Equação).

Os algoritmos Redes Neurais Artificiais e *Random* são mais indicados para a estimativa da produtividade de *Harvester*, frente ao conjunto de dados utilizado e configurações de ajuste. Destaca-se que os mesmos apresentaram erros médios inferiores a -5%.

O algoritmo de *Data mining* tende a superestimar a produtividade de *Harvester*, conforme observado na análise de resíduos e operacionalização da técnica, com erro médio superior a +10%.

5. CONSIDERAÇÕES FINAIS

O presente estudo foi desenvolvido durante o programa de MBA em Manejo Florestal de Precisão na Universidade Federal do Paraná. A análise de dados e compilação de resultados demandaram muito estudo e dedicação para o entendimento das técnicas de inteligência artificial avaliadas. Embora a área de inteligência artificial seja nova e aparentemente complexa, pesquisadores das diversas áreas que atuam com modelagem podem buscar aprender e aplicar as técnicas nas suas realidades.

Recomenda-se para trabalhos derivados do presente estudo maior aprofundamento em configurações específicas dos algoritmos com melhor qualidade estatística estudados (Redes Neurais Artificiais e *Random forests*).

O uso das variáveis de máquina e planejamento (equipe e bimestre de colheita) dependem da pré-existência de um sistema de agendamento otimizado que possibilite planejar antecipadamente essas informações. Na inexistência de um sistema de agendamento que suporte dessa forma, recomenda-se aplicar apenas variáveis que não dependam da máquina ou de planejamento operacional.

Recomendam-se que os pesquisadores apliquem os algoritmos aqui avaliados em suas situações específicas, visto à qualidade estatística estar relacionada ao conjunto de dados avaliados em cada caso.

De forma geral, poucos estudos relacionados à modelagem de operações florestais são encontrados, principalmente quanto ao uso da inteligência artificial.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, R. L. **Map locator of São Paulo's Lençóis Paulista city**. 2006. Disponível em: https://commons.wikimedia.org/wiki/File:SaoPaulo_MesoMicroMunicip.svg. Acesso em: 10/10/2016.

ALVARES, C. A.; et. al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

BINOTI, M. L. M. S.; et. al. Redes Neurais Artificiais para a estimação do volume de árvores. **Revista Árvore**, v. 38, n. 2, p. 283-288, 2014a.

BINOTI, M. L. M. S.; et. al. Utilização de redes neurais artificiais para a projeção da distribuição diamétrica de povoamentos equiâneos. **Revista Árvore**, v. 38, n. 4, p. 747-754, 2014b.

BRAMUCCI, M.; SEIXAS, F. Determinação e quantificação de fatores de influência sobre a produtividade de “*harvesters*” na colheita florestal. **Scientia Forestalis**, Piracicaba, v. 62, p. 62-74, 2002.

BULLINARIA, J. A. **Introduction to Neural Computation - Notas de Aula**. 2014. Disponível em: <http://www.cs.bham.ac.uk/~jxb/inc.html>. Acesso em: 10/10/2016.

CARVALHO, P. A. C. F. **Modelagem do risco de incêndio florestal com redes neuronais artificiais: aplicação ao Parque Natural do Montesinho**. 197p. Dissertação (Mestrado em Ciência e Sistemas de Informação Geográfica) – Instituto Superior de Estatística e Gestão de Informação, Lisboa, Portugal, 2005.

ERASLAN, E. The Estimation of Product Standard Time by Artificial Neural Networks in the Molding Industry. **Mathematical Problems in Engineering**. 12p. 2009.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 378p., 2011.

HALL, M.; et. al. **The WEKA Data Mining Software: An Update**. SIGKDD Explorations, v. 11, n. 1, 2009.

HEATON, J. **Introduction to neural networks with Java**. Clarkson: Heaton Reserarch, 2ª edição. 2008.

IBA – INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório IBÁ 2016**. São Paulo: IBÁ, 2016.

JOHN DEERE. **Máquinas florestais: Forwarder 1910E**. 2016. Disponível em: https://www.deere.com.br/pt_BR/products/equipment/forestry_swing_machines/2154d/2154d.page. Acesso em: 30/08/2016.

JOHN DEERE. **Máquinas florestais: Máquina Swing Machine Florestal 2154D**. 2016. Disponível em: https://www.deere.com.br/pt_BR/products/equipment/forestry_swing_machines/2154d/2154d.page. Acesso em: 30/08/2016.

MALINOVSKI, R. A.; et. al. Otimização da distância de extração de madeira com forwarder. **Scientia Forestalis**, v. 36, n. 79, p. 171-179, 2008.

MALINOVSKI, R. A.; et. al. Análise das variáveis de influência na produtividade das máquinas de colheita de madeira em função das características físicas do terreno, do povoamento e do planejamento operacional florestal. **Floresta**, v. 36, n. 2, p. 166-182, 2006.

MCCARTHY, J. **Actions and other events in situation calculus**. Proceedings of KR-2002, 2002.

MICROSOFT, **Microsoft Excel 2013, Versão 2013**. Redmond, Washington: Microsoft, 2012. (*Software*).

MIGUEL, E. P.; et. al. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. **Pesquisa Agropecuária Brasileira**, v. 50, n. 9, p. 829-839, 2015.

MONARD, M. C.; BARANAUKAS, J. A. **Aplicações de Inteligência Artificial: Uma Visão Geral**. Instituto de Ciências Matemáticas e de Computação de São Carlos, 2000.

OLIVEIRA, R. J.; et. al. Análise de Processo Hierárquico para Gerir as Estradas Florestais. **Floresta e Ambiente**, v. 20, n. 1, p. 38-44, 2013.

SANTOS, C. C. **Modelagem de bacias urbanas com redes neurais artificiais**. 105p. Dissertação (Mestrado em Engenharia Hidráulica) - Escola Politécnica, Universidade de São Paulo - USP, São Paulo, 2001.

SCHIKOWSKI, A. B. **Estimativa do volume e da forma do fuste utilizando técnicas de aprendizado de máquina**. 73p. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal do Paraná, PR, 2016.

SILVA, F. **Aplicação de redes neurais artificiais para estimar a produtividade do forwarder na extração de madeira**. 63p. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, MG, 2015.

SILVEIRA, D. P. **Estimação do volume de madeira empilhada utilizando imagens digitais e redes neurais**. 66p. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, MG, 2014.

SOARES, F. A. A. M. N. **Predição recursiva de diâmetros de clones de eucalipto utilizando rede Perceptron de múltiplas camadas para o cálculo de volume**.

111p. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Uberlândia, Uberlândia, MG, 2011.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining – Mineração**. Ciência Moderna, 2009.

WOJCIECHOWSKI, J. **JCARBON - software na web com data mining para estimativas de volume, biomassa e carbono em florestas**. 174p. Tese (Doutorado em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, PR, 2015.