

PREDIÇÃO DA MANUTENÇÃO DE MÁQUINAS FLORESTAIS AUTOPROPELIDAS EM SISTEMA *FULL TREE* UTILIZANDO *MACHINE LEARNING*

SILVA, T.¹; SASSO JUNIOR, V. A.¹; PANDOLFO, P. T.¹; ALMEIDA, R. O.²; SIMÕES,
D.¹

¹Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrônômicas, Botucatu;

²Instituto Federal de Ensino, Ciência e Tecnologia do Sudeste de Minas Gerais, Campus
Muriaé.

RESUMO

A colheita mecanizada de madeira de *Eucalyptus* pode ser realizada pelo sistema *full tree*, no qual utiliza o *feller buncher* para o corte e a derrubada de árvores. A utilização de máquinas florestais autopropelidas ressalta a importância da correta previsão referente à manutenção. Em vista disso, aplica-se o *machine learning* visando encontrar resultados interpretativos para tomada de decisões estratégicas e redução dos custos de manutenção. A partir deste contexto, o objetivo foi verificar se é possível prever a manutenção de *feller bunchers* empregados na derrubada de árvores por meio do *machine learning*. O estudo foi conduzido em florestas plantadas de *Eucalyptus*, localizadas no Centro-Oeste do estado de São Paulo. O banco de dados estruturado foi produzido por meio de dois atributos: horímetro e tipo de manutenção, sendo esse último o atributo alvo da predição. Após o processo de *data wrangling*, foi utilizado o aprendizado supervisionado, com balanceamento de dados via *oversampling*. Nesse caso, o banco de dados foi dividido em conjunto de treinamento e teste. Em sequência a aplicação dos 14 algoritmos de aprendizagem, observou-se uma melhor performance em três modelos: *CatBoost Classifier*, *Linear Discriminant Analysis* e *Naïve Bayes*. Durante o processo de treinamento foi aplicado o *cross-validation*. Assim, a partir do algoritmo *CatBoost Classifier*, em modo *Default*, encontra-se o melhor modelo gerado para predição do tipo de manutenção de máquinas florestais autopropelidas, com performance de 0,46, 0,60, e 0,74 para as métricas Coeficiente de Correlação de Matthews, Precisão e Área Sob a Curva, respectivamente.

Palavras-chave: *CatBoost Classifier*; colheita de madeira; *Eucalyptus*.

INTRODUÇÃO

O gênero *Eucalyptus* é o mais cultivado pelo setor florestal, visto que, tem sido amplamente estudado para maximizar a produção da planta por meio do uso das suas espécies e combinações híbridas. No Brasil, o eucalipto destaca-se por apresentar sinergia diante do melhoramento genético, das características florestais, das condições climáticas favoráveis, da mecanização no processo da colheita de madeira e do tempo de rotação da cultura (TRUEMAN, et al., 2018; SPINELLI e MOURA, 2019).

Consequente, a colheita mecanizada de madeira de *Eucalyptus* pode ser realizada pelo sistema de árvores inteiras (*full tree*), que concerne na remoção das árvores do interior do talhão para um pátio intermediário ou para as bordas das estradas florestais. Destarte, utiliza-se, comumente, esse sistema para o arraste de árvores de maior volume médio individual.

Vale destacar que, no *full tree* empregam-se algumas máquinas florestais autopropelidas, como: *feller buncher*, *skidder* e *harvester* (SCORUPSKI, et al., 2017; SOUZA, et al., 2018; SIMÕES, et al., 2022).

Dentre essas máquinas utilizadas no sistema *full tree*, tem-se o *feller buncher*, no qual possui na parte frontal o cabeçote de corte com disco, sendo responsável por exercer o corte e a derrubada de árvores. Nesse caso, destaca-se que essas atividades acontecem próximo ao nível do solo. Em razão disso, com intuito de aumentar a estabilidade e o conforto dos operadores, o *feller buncher* é equipado com um sistema de nivelamento (ANDRONOV, et al., 2020; PEREIRA, et al., 2020).

No setor florestal, atribui-se importância à manutenção das máquinas florestais autopropelidas dentro da atividade de colheita de madeira, dado que, a manutenção pode afetar o tempo de operação e a eficiência. Assim, deve-se identificar e resolver as falhas das máquinas a fim de evitar o desligamento nos processos de produção. Isto posto, a disponibilidade mecânica da máquina florestal autopropelida depende diretamente da manutenção preventiva, sendo que, a realização da mesma pode eliminar ou reduzir intervenções corretivas (PACOLLA, 2017; WAN, et al., 2017; CARVALHO, et al., 2019).

Diante disso, pode-se aplicar a aprendizagem de máquina (*machine learning*) visando encontrar resultados interpretativos para tomada de decisões estratégicas. Dessa forma, o *machine learning* pode solucionar problemas acerca da manutenção das máquinas florestais autopropelidas e proporcionar vantagens, como: redução de custos da manutenção, redução de falhas de máquinas, redução da parada de reparo, reposição do estoque de peças, aumento da vida útil das peças, aumento da produção, melhoria na segurança do operador, verificação da necessidade de reparos, entre outros (PERES, et al., 2018; ÇINAR, et al., 2020).

As técnicas de *machine learning* constituem um amplo e bem consolidado subcampo da inteligência artificial, cujas análises estatísticas são utilizadas com o intuito de extrair padrões a partir de um conjunto de dados, possibilitando algoritmos específicos para cada situação. Desse modo, geralmente, torna-se necessário um volume maior de dados, uma vez que, os problemas a serem solucionados são mais complexos, do ponto de vista computacional (NWANGANGA e CHAPPLE, 2020; BORGES, et al., 2021).

Diante deste contexto, justifica-se o uso do *machine learning* a fim de prever a manutenção de máquinas florestais autopropelidas visando uma correta tomada de decisão e a redução dos custos de manutenção. Destarte, o objetivo foi verificar se é possível prever a manutenção de *feller bunchers* empregados na derrubada de árvores de *Eucalyptus* por meio do *machine learning*.

MATERIAL E MÉTODOS

Foram utilizados dados provenientes de áreas de colheita da madeira com florestas plantadas de *Eucalyptus*, localizadas no Centro-Oeste do estado de São Paulo. A atividade de corte e derrubada de árvores foi realizada por três máquinas florestais autopropelidas, marca *John Deere*, caracterizadas como *feller buncher*, a saber:

Feller buncher 1: modelo 903K, com potência nominal de 199 kW, a grua com alcance máximo de 7,01 m, massa aproximada de 29.630 kg e sistema de rodados compostos por esteiras. O modelo do cabeçote de corte era FR22B, composto por disco de corte, com capacidade de corte de 559 mm e capacidade de acúmulo de 0,48 m³. No início da coleta de dados o horímetro apresentava 35.842 horas de uso acumuladas.

Feller buncher 2: modelo 903M, com potência nominal de 213 kW, a grua com alcance máximo de 7,01 m, massa aproximada de 31.590 kg e sistema de rodados compostos por esteiras. O modelo do cabeçote de corte era FR22B, composto por disco de corte, com capacidade de corte de 559 mm e capacidade de acúmulo de 0,48 m³. No início da coleta de dados o horímetro apresentava 24.857 horas de uso acumuladas.

Feller buncher 3: modelo 903M, com potência nominal de 213 kW, a grua com alcance máximo de 7,01 m, massa aproximada de 31.590 kg e sistema de rodados compostos por esteiras. O modelo do cabeçote de corte era FR22B, composto por disco de corte, com capacidade de corte de 559 mm e capacidade de acúmulo de 0,48 m³. No início da coleta de dados o horímetro apresentava 13.464 horas de uso acumuladas.

O banco de dados estruturado foi produzido por meio das informações da empresa do segmento de chapas de fibra de madeira, cujo foco era a manutenção de máquinas florestais autopropelidas. Destes dados, foram selecionados dois atributos, os quais eram um do tipo numérico e um do tipo categórico, a saber: horímetro e tipo de manutenção, respectivamente. Destaca-se que o tipo de manutenção era o atributo alvo da predição.

Isto posto, foram constatadas 225 instâncias. Ademais, realizou-se o processo de ajuste dos dados por meio da linguagem de programação R, sob contribuição dos pacotes *multidplyr*, *tidyverse*, *doFuture*, *foreach*, entre outros (BENGTSSON, 2021; WALLIG, 2021; WICKHAM, 2021a; WICKHAM, 2021b; RStudio Team, 2022).

Posteriormente ao processo de *data wrangling*, foi utilizado o aprendizado supervisionado, com balanceamento de dados via *oversampling* (SMOTE). Sendo que, o banco de dados foi dividido da seguinte forma: conjunto de treinamento (90% das instâncias) e conjunto de teste (10% das instâncias) (NASTESKI, 2017; ELREEDY e ATIYA, 2019).

Após, aplicou-se 14 algoritmos de aprendizagem, de modo a selecionar aqueles de melhor performance. Consequente, os algoritmos selecionados passaram pelo processo de *tune*, *ensemble*, *blend* e *stack*, visando à obtenção de melhores performances de predição. Ressalta-se que durante o processo de treinamento foi aplicado o *cross-validation*.

Em relação às métricas utilizadas, tem-se que as escolhidas foram, principalmente, o Coeficiente de Correlação de Matthews (MCC), Precisão e Área Sob a Curva (AUC). Logo, realizou-se o *machine learning* por meio da linguagem de programação *Python*, o qual utilizou as bibliotecas *SciPy*, *Scikit-learn*, *pycaret*, *NumPy*, *Pandas* e *Matplotlib* (CHEN, 2017; RAHMATI, et al., 2019; RASCHKA e MIRJALILI, 2019; TAREEN e KINNEY, 2020; BONALD, et al., 2020; GAIN e HOTTI, 2020; HARRIS, et al., 2020).

RESULTADOS E DISCUSSÃO

As análises exploratórias demonstraram o comportamento inicial dos dados, assim, ao considerar as 225 instâncias tem-se que o atributo horímetro apresentou três grupos de valores distintos (Figura1).

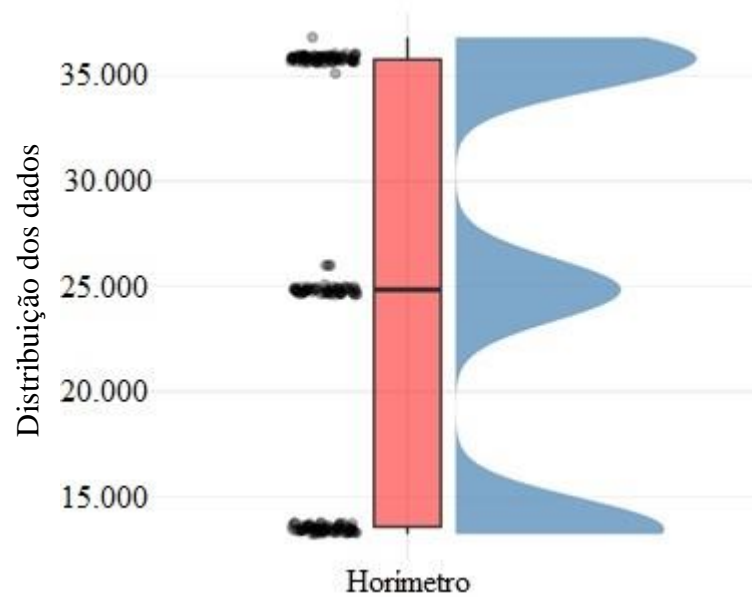


Figura 1. Distribuição dos dados do atributo numérico horímetro.

Em sequência a aplicação dos 14 algoritmos de aprendizagem, observou-se uma melhor performance em três modelos, ponderando as métricas de tratamento MCC, Precisão e AUC. Dessa forma, tem-se o modelo *CatBoost Classifier* (CBC) que apresentou um MCC de 0,31, uma Precisão de 0,59 e uma AUC de 0,71. O *Linear Discriminant Analysis* (LDA) com o MCC de 0,35, a Precisão de 0,61 e a AUC de 0,72. Por fim, o *Naive Bayes* (NB) com o MCC de 0,35, a Precisão de 0,61 e a AUC de 0,71 (Figura 2).

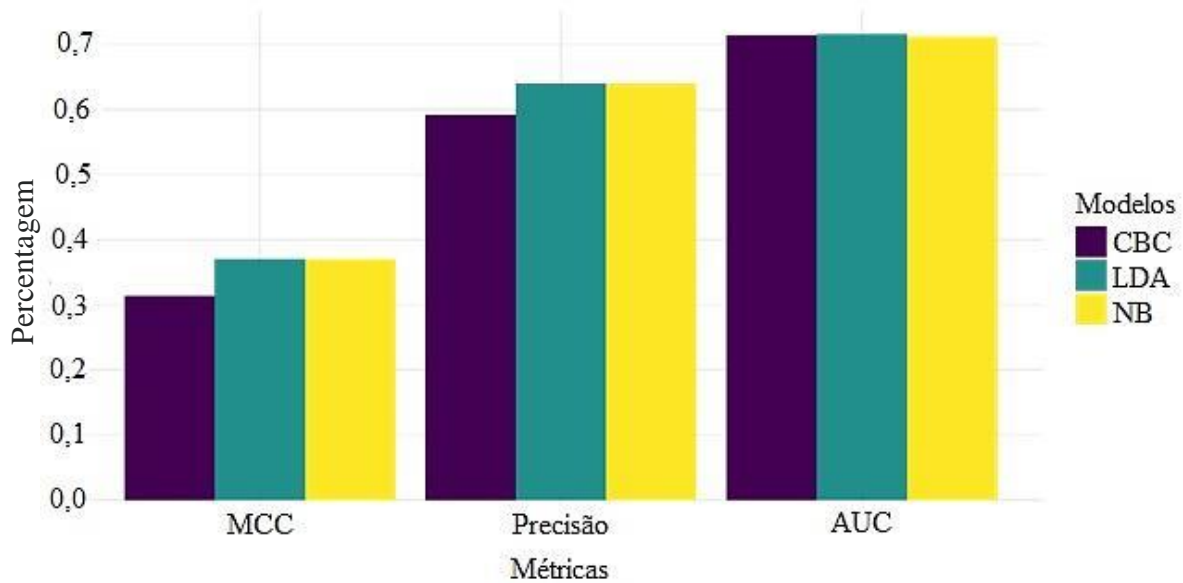


Figura 2. Comparação dos modelos selecionados a partir das métricas de tratamento MCC, Precisão e AUC.

Após a execução do processo de *tune*, *ensemble*, *blend* e *stack*, identificou-se o modelo com maior performance por meio de um comparativo entre todos os modelos criados. Consequente, o modelo *CatBoost Classifier*, em modo *Default*, alcançou a performance de 0,46, 0,60 e 0,74 para as métricas MCC, Precisão e AUC, respectivamente (Tabela 1). Sendo que, os outros modelos, isto é, *Blended*, *Stacked*, *Linear Discriminant Analysis* e *Naive Bayes* apresentaram o mesma performance diante das métricas teste.

Tabela 1. Comparação dos modelos a partir das métricas de treinamento e teste.

Modelos	Modos	Métricas treinamento			Métricas teste		
		MCC	Precisão	AUC	MCC	Precisão	AUC
<i>Blended</i>	-	0,37	0,64	0,71	0,40	0,55	0,71
<i>Stacked</i>	-	0,37	0,64	0,71	0,40	0,55	0,71
CBC	<i>Default</i>	0,31	0,59	0,71	0,46	0,60	0,74
	<i>Ensembled</i>	0,34	0,60	0,73	0,40	0,55	0,71
	<i>Tuned</i>	0,37	0,64	0,70	0,40	0,55	0,71
LDA	<i>Default</i>	0,35	0,61	0,72	0,40	0,55	0,71
	<i>Ensembled</i>	0,37	0,64	0,71	0,40	0,55	0,71
	<i>Tuned</i>	0,35	0,61	0,72	0,40	0,55	0,71
NB	<i>Default</i>	0,35	0,61	0,71	0,40	0,55	0,71
	<i>Ensembled</i>	0,37	0,64	0,71	0,40	0,55	0,71
	<i>Tuned</i>	0,35	0,61	0,71	0,40	0,55	0,71

Segundo Prokhorenkova, et al. (2018), o algoritmo *CatBoost Classifier* caracteriza-se como sendo uma implementação de aumento de gradiente, no qual utiliza as árvores de decisão binárias como preditores de base. Diante disso, ao avaliar o algoritmo, Ibrahim, et al. (2020) constataram em seu estudo a eficiência e precisão do *CatBoost Classifier* ao prever os recursos categóricos, logo, recomenda-se o uso do algoritmo em modelos preditivos.

Ademais, notou-se que o uso do *CatBoost Classifier* proporciona a economia de tempo computacional e de recursos consideráveis, visto que, o algoritmo utiliza estatísticas baseadas em alvos modificados visando manipular adequadamente os recursos categóricos durante o tempo de treinamento. Vale ressaltar que, o *CatBoost Classifier* apresenta uma boa solução diante de problemas que envolvem dados heterogêneos (situação semelhante a este estudo), todavia o mesmo algoritmo pode não ser recomendado para problemas envolvendo dados homogêneos (HANCOCK e KHOSHGOFTAAR, 2020; HUSSAIN, et al., 2021).

A função do *machine learning* diante da manutenção das máquinas florestais autopropelidas baseia-se no prognóstico e previsão de falhas, além de também ser usado para diagnosticar as falhas. Em vista disso, utilizou-se neste estudo o *machine learning* para a predição do tipo de manutenção. Logo, nota-se que desenvolver e implementar uma estratégia de manutenção inteligente permite prever quando a manutenção deverá ser realizada, reduzindo os custos de manutenção e aumentando a produtividades dos *feller bunchers* (WU, et al., 2017; SCHMIDT e WANG, 2018).

CONCLUSÃO

A partir do algoritmo *CatBoost Classifier*, em modo *Default*, encontra-se o melhor modelo gerado para predição do tipo de manutenção de máquinas florestais autopropelidas. O modelo apresenta performance de 0,46, 0,60, e 0,74 para as métricas MCC, Precisão e AUC, respectivamente.

AGRADECIMENTOS

O estudo foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

REFERÊNCIAS

ANDRONOV, A. V.; TARADIN, G. S.; ZVEREV, I. A. Models for the determination of basic parameters of tracked feller buncher machines. **IOP Conference Series: Materials Science and Engineering**, v. 817, n. 1, p.1-7, 2020.

BENGTSSON, H. DoFuture: a universal foreach parallel adapter using the future api of the 'future' package. **R package version 0.12.0**. URL <<https://cloud.r-project.org>>. 2021.

BONALD, T.; LARA, N.; LUTZ, Q.; CHARPENTIER, B. Scikit-network: graph analysis in Python. **Journal of Machine Learning Research**, v. 21, n. 185, p.1-6, 2020.

BORGES, A. F. S.; LAURINDO, F. J. B.; SPÍNOLA, M. M.; GONÇALVES, R. F.; MATTOS, C. A. The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. **International Journal Of Information Management**, v. 57, n. 1, p.1-16, 2021.

CARVALHO, T. P.; SOARES, F. A. A. M. N.; VITA, R.; FRANCISCO, R. P.; BASTO, J. P.; ALCALÁ, S. G. S. A systematic literature review of machine learning methods applied to predictive maintenance. **Computers & Industrial Engineering**, v. 137, n. 1, p.1-10, 2019.

CHEN, D. Y. **Pandas for Everyone: Python Data Analysis**. Boston: Addison-Wesley Professional; 2017. v. 1. 406 p.

ÇINAR, Z. M.; NUHU, A. A.; ZEESHAN, Q.; KORHAN, O.; ASMAEL, M.; SAFAEI, B. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. **Sustainability**, v. 12, n. 19, p.8211-8253, 2020.

ELREEDY, D.; ATIYA, A. F. A comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. **Information Sciences**, v. 505, n. 1, p.32-64, 2019.

GAIN, U.; HOTTI, V. Low-code AutoML-augmented data pipeline – a review and experiments. **Journal of Physics: Conference Series**, v. 1828, n. 1, p.2-8, 2020.

HANCOCK, J. T.; KHOSHGOFTAAR, T. M. CatBoost for big data: an interdisciplinary review. **Journal of Big Data**, v. 7, n. 1, p.1-45, 2020.

HARRIS, C. R.; MILLMAN, K. J.; VAN DER WALT, S. J.; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; VAN KERKWIJK, M. H.; BRETT, M.; HALDANE, A.; DEL RÍO,

J. F.; WIEBE, M.; PETERSON, P.; MARCHANT, P. G.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. **Nature**, v. 585, n. 7825, p.357-362, 2020.

HUSSAIN, S.; MUSTAFA, M. W.; JUMANI, T. A.; BALOCH, S. K.; ALOTAIBI, H.; KHAN, I.; KHAN, A. A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. **Energy Reports**, v. 7, n. 1, p.4425-4436, 2021.

IBRAHIM, A. A.; RIDWAN, R. L.; MUHAMMED, M. M.; ABDULAZIZ, R. O.; SAHEED, G. A. Comparison of the CatBoost Classifier with other machine learning methods. **International Journal of Advanced Computer Science and Applications**, v. 11, n. 11, p.738-748, 2020.

NASTESKI, V. An overview of the supervised machine learning methods. **Horizons**, v. 4, n. 1, p.51-62, 2017.

NWANGANGA, F.; CHAPPLE, M. **Practical Machine Learning in R**. Hoboken: Wiley; 2020. v. 1. 464 p.

PACOLLA, J. E. **Manutenção e operação de equipamentos móveis**. São José dos Campos: JAC; 2017. v. 2. 270 p.

PEREIRA, G.; FENNER, P. T.; BATISTELA, G. C.; SIMÕES, D. Análise técnica-econômica da derrubada de *Eucalyptus sp.* com *feller-buncher*: uma abordagem estocástica. **Scientia Forestalis**, v. 48, n. 126, p.1-13, 2020.

PERES, R. S.; ROCHA, A. D.; LEITÃO, P.; BARATA, J. IDARTS – towards intelligent data analysis and real-time supervision for industry 4.0. **Computers in Industry**, v. 101, n. 1, p.138-146, 2018.

PROKHORENKOVA, L.; GUSEV, G.; VOROBEV, A.; DOROGUSH, A. V.; GULIN, A. CatBoost: unbiased boosting with categorical features. **Advances In Neural Information Processing Systems**, v. 31, n. 1, p.6638-6648, 2018.

RAHMATI, O.; MOGHADDAM, D. D.; MOOSAVI, V.; KALANTARI, Z.; SAMADI, M.; LEE, S.; BUI, D. T. An automated Python Language-Based tool for creating absence samples in groundwater potential mapping. **Remote Sensing**, v. 11, n. 11, p.1375-1397, 2019.

RASCHKA, S.; MIRJALILI, V. **Python machine learning: machine learning and deep learning with Python, Scikit-learn, and TensorFlow**. Birmingham: Packt Publishing; 2019. v. 3. 770 p.

RStudio Team (2022). **RStudio: Integrated Development for R**. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

SCHMIDT, B., WANG, L. Predictive maintenance of machine tool linear axes: a case from manufacturing industry. **Procedia Manufacturing**, v. 17, n. 1, p.118–125, 2018.

SCORUPSKI, A. J.; OLIVEIRA, F. M.; DINIZ, C. C. C.; LOPES, E. S. Produtividade do processamento mecanizado da madeira de pinus nos sistemas *full tree* e *cut to length* em diferentes volumes individuais. **Biofix Scientific Journal**, v. 2, n. 1, p.12-15, 2017.

SIMÕES, D.; AVELINO, L. T.; MUNIS, R. A.; BATISTELA, G. C.; MIYAJIMA, R. H. Grapple saw's operating conditions influence on the productivity and cost of processing felled trees. **Revista Floresta**, v. 52, n. 1, p.64-73, 2022.

SOUZA, F. L.; SAMPIETRO, J. A.; DACOREGIO, H. M.; SOARES, P. R. C.; LOPES, E. S.; QUADROS, D. S. Densidade ótima e aceitável de estradas na colheita de pinus no sistema de toras curtas e árvores inteiras. **Scientia Forestalis**, v. 46, n. 118, p.189-198, 2018.

SPINELLI, R.; MOURA, A. C. A. Decreasing the fuel consumption and CO₂ emissions of excavator-based harvesters with a machine control system. **Forests**, v. 10, n. 43, p.3-12, 2019.

TAREEN, A.; KINNEY, J. B. Logomaker: beautiful sequence logos in Python. **Bioinformatics**, v. 36, n. 7, p.2272-2274, 2020.

TRUEMAN, S. J.; HUNG, C. D.; WENDLING, I. Tissue culture of *Corymbia* and *Eucalyptus*. **Forests**, v. 9, n. 2, p.1-42, 2018.

WALLIG, M. Foreach: provides foreach looping construct. **R package version 1.5.2**. URL <<https://cloud.r-project.org>>. 2021.

WAN, J.; TANG, S.; LI, D.; WANG, S.; LIU, C.; ABBAS, H.; VASILAKOS, A. V. A manufacturing Big Data solution for active preventive maintenance. **IEEE Transactions on Industrial Informatics**, v. 13, n. 4, p.2039–2047, 2017.

WICKHAM, H. Multidplyr: a multi-process 'dplyr' backend. **R package version 0.1.1**. URL <<https://cloud.r-project.org>>. 2021a.

WICKHAM, H. Tidyverse: easily install and load the 'tidyverse'. **R package version 2.0.1**. URL <<https://cloud.r-project.org>>. 2021b.

WU, D.; JENNINGS, C.; TERPENNY, J.; GAO, R. X.; KUMARA, S. A comparative study on Machine Learning algorithms for smart manufacturing: tool wear prediction using Random Forests. **Journal of Manufacturing Science and Engineering**, v. 139, n. 7, p.1-9, 2017.