

## **TRANSPORTE RODOVIÁRIO DE MADEIRA: DECISÕES ESTRATÉGICAS A PARTIR DE *MACHINE LEARNING***

SASSO JÚNIOR, V. A.<sup>1</sup>; SILVA, T.<sup>1</sup>; ALMEIDA, R. O.<sup>2</sup>; MUNIS, R. A.<sup>1</sup>; SIMÕES, D.<sup>1</sup>

<sup>1</sup>Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrônômicas, Botucatu;

<sup>2</sup>Instituto Federal de Ensino, Ciência e Tecnologia do Sudeste de Minas Gerais, Câmpus Muriaé.

### **RESUMO**

No ciclo produtivo das florestas plantadas de *Eucalyptus*, vem sendo buscado otimizar suas etapas e reduzir seus custos. Dentre elas, destaca-se o transporte rodoviário de madeira, com dispêndios representando cerca de 45% dos custos totais. Neste sentido, o volume das madeiras possui importante impacto, sendo fundamental sua correta estimativa. Isto posto, foi verificado se há precisão na estimativa do volume de madeira de *Eucalyptus* transportadas por caminhão trator em rodovias, por meio de aprendizagem de máquina. Realizou-se o processo de ajuste dos dados das variáveis ponderadas por meio da linguagem de programação R. Após a execução do *data wrangling*, utilizou-se o aprendizado supervisionado, via linguagem de programação *Python*, com balanceamento de dados via *oversampling*, e divisão das instâncias em conjuntos de treinamento e teste. Posterior a modelagem, à seleção dos modelos com melhor performance em coeficiente de determinação, erro absoluto médio e erro quadrático médio, e da realização do treinamento e teste, pode-se confirmar que o modelo obteve alta performance na estimativa do volume de madeiras, sendo que o modelo de combinação gerado a partir de *AdaBoost Regressor*, *Bayesian Ridge* e *Gradient Boosting Regressor*, demonstrou a melhor performance para predição.

**Palavras-Chave:** predição de volume, modelagem de dados, modelos combinados.

### **INTRODUÇÃO**

O gênero *Eucalyptus*, um dos principais da família *Myrtaceae*, é a folhosa mais plantada no mundo, contando com uma área de aproximadamente 25 milhões de hectares. Isto deve-se à elevada resistência a estresses bióticos, ao rápido crescimento de sua madeira, a qual pode ter diversas finalidades. Dessarte, o gênero vem ganhando considerável importância econômica e valor comercial (ABIRI, et al., 2020; ELLI, et al., 2020; FREITAS, et al., 2020; FERNANDES, et al., 2022).

Dentre as etapas de produção em florestas plantadas, o transporte de madeira consiste em movimentar os feixes para unidades consumidoras ou estaleiros. Na conexão entre floresta plantada e as unidades fabris, destaca-se o modal rodoviário com deslocamento por caminhões, em virtude da variada oferta de veículos e ao curto período de deslocamento. Este processo representa cerca de 45% dos dispêndios totais da cadeia produtiva (MALLADI e SOWLATI, 2017; MÜLLER, et al., 2019).

Decorrente do contexto, um sistema de transporte inadequado pode proporcionar inviabilidade econômica. Por conseguinte, busca-se otimizá-lo e reduzir os custos, para maximizar os ganhos na produção. Uma variável que possui importante impacto nos dispêndios

do transporte é o volume das madeiras (LACHINI, et al., 2018; TYMENDORF e TRZCIŃSKI, 2020).

O volume relaciona-se com a densidade aparente e a umidade da madeira, gerando um aumento em seu valor e elevando o peso da tara do caminhão, podendo alcançar rapidamente o limite de carregamento e, conseqüente, diminuir o lucro. Portanto, o volume deve ser corretamente estimado, visando aprimorar a etapa de transporte rodoviário de madeira (KOIRALA, et al., 2017; DANTAS, et al., 2020).

A aprendizagem de máquinas, ou *machine learning*, é definido como a capacidade de máquinas aprenderem a lidar com dados eficientemente, sem programação estrita. O processo é feito, principalmente, de forma supervisionada, não supervisionada ou semi-supervisionada, efetuando análises como classificação, regressão, regras de associação, clusterização, séries temporais e detecção de anomalias. Dentre seus usos, há a predição do volume de madeiras transportadas (LIAKOS, et al., 2018; UCCI, et al., 2019; FLUKE e JACOBS, 2020; MAHESH, 2020).

Em virtude da abundância de dados na atualidade, a demanda por *machine learning* está se avolumando, com aplicação em diversas áreas, como na medicina, farmacologia, educação, comunicação, cibersegurança e, sobretudo, na silvicultura. Sua aplicação na silvicultura dá-se para predição de variáveis da estrutura florestal, no estoque de carbono, aspectos ecológicos, no transporte de madeira, entre outros (CÍHAN, et al., 2017; CARLEO, et al., 2019; ARUMÄE, et al., 2022).

Diante deste contexto, objetivou verificar se a aprendizagem de máquina, por regressão, atingirá boa performance na predição do volume de madeiras de *Eucalyptus* transportadas pelo modal rodoviário.

## MATERIAL E MÉTODOS

O estudo foi desenvolvido a partir de dados provenientes de florestas plantadas de *Eucalyptus grandis*, localizados na região Oeste do Uruguai. A idade média da floresta plantada era de 172 meses, com densidade média de  $0,87 \text{ g cm}^{-3}$ . O comprimento das toras transportadas correspondia a 7,2 m.

A Combinação de Veículos para Transporte de Carga, caminhão trator trucado com semirreboque, utilizado para o transporte das madeiras possuía um eixo simples e um conjunto de eixos em *tandem* duplo, juntamente a um semirreboque com conjunto de eixos em *tandem* triplo com distância maior que 1,20 m e menor que 2,40 m entre eles, apresentando comprimento de 18,60 m.

As rodovias utilizadas para o transporte de madeira tinham predominância de geometria plana, pavimentação asfáltica, boas condições de conservação e eram abrangidas, em sua maioria, com pistas de rodagem simples.

Foi ponderado o transporte de madeira realizado em um período de 35 meses, por meio de um modelo de Combinação de Veículos para Transporte de Carga (CVC), caminhão trator trucado com semirreboque, cujo Peso Bruto Total Combinado (PBTC) correspondia à 49,5

toneladas. Posto isto, estruturou-se o banco de dados com os atributos de idade da floresta, densidade da madeira, quantidade de dias que a madeira permaneceu no campo após a derrubada, tempo do ciclo operacional, velocidade da CVC com carga e volume da madeira, sendo este o alvo de predição. O processo de ajuste dos dados foi realizado por meio da linguagem de programação R, a partir dos pacotes *doFuture*, *foreach*, *multidplyr* e *tidyverse* (BENGTSSON, 2021; WALLIG, 2021; WICKHAM, 2021a; WICKHAM, 2021b; RSTUDIO TEAM, 2022).

Sequente a execução da análise exploratória e do *data wrangling*, foi utilizado a aprendizagem supervisionada, com balanceamento de dados via *oversampling – SMOTE*, dividindo o banco de dados em conjunto de treinamento e teste, com 90% e 10% das instâncias, respectivamente (LEMAÎTRE, et al., 2017; SUHAIMI e ABAS, 2020).

Foram aplicados 16 algoritmos de aprendizagem, modo *default*, a fim de selecionar os de melhor performance. Os selecionados passaram pelos processos de *tune*, *ensemble*, *blend*, e *stack*, todos com *cross-validation*, para alcançar melhor performance de predição. Para avaliação da performance dos modelos, as métricas utilizadas foram o coeficiente de determinação ( $R^2$ ), o erro absoluto médio ou *mean absolut error* (MAE) e o erro quadrático médio ou *mean square error* (MSE).

O aprendizado de máquina foi conduzido por meio da linguagem de programação *Python*, utilizando as bibliotecas *Matplotlib*, *NumPy*, *Pandas*, *pycaret*, *Scikit-learn* e *SciPy* (MCKINNEY, 2017; HARRIS, et al., 2020; NELLI, 2018; PYCARET, 2020; VIRTANEN et al., 2020; GAUR, et al., 2021).

## RESULTADOS E DISCUSSÃO

Observando o comportamento das 484 instâncias na análise exploratória, notou-se presença de *outliers* nos atributos de tempo do ciclo operacional e velocidade da CVC com carga. Após o processo de ajuste e aplicação do *data wrangling*, as instâncias foram reduzidas para 464 (Tabela 1), separando aleatoriamente 90% para o conjunto de treinamento e 10% para o de teste.

Tabela 1. Comportamento do Banco de dados estruturados com as instâncias, comparando o mesmo antes e após a remoção dos *outliers*, na análise exploratória.

Atributos	<i>Outliers</i> Presentes		<i>Outliers</i> Ausentes	
	Média	Desvio Padrão	Média	Desvio Padrão
Idade da floresta (meses)	171,89	13,71	172,45	13,52
Densidade da madeira ( $\text{g cm}^{-3}$ )	0,87	0,13	0,87	0,13
Tempo de permanência da madeira no campo após derrubada (dias)	69,99	61,81	69,76	62,12
Tempo do ciclo operacional (hora)	6,19	1,09	6,21	1,03

Velocidade da CVC com carga (km h <sup>-1</sup> )	41,32	6,76	40,68	5,65
Volume (cm <sup>3</sup> )	35,53	5,38	35,38	5,37

Em seguida, com a aplicação dos modelos aos dados de treinamento (Tabela 2), modo *default*, foram selecionados três diferentes algoritmos de aprendizagem, a saber: *AdaBoost Regressor*, *Bayesian Ridge* e *Gradient Boosting Regressor*.

Tabela 2. Performance obtida com os 16 algoritmos aplicados sobre o conjunto de treinamento.

Modelo	MAE	MSE	R <sup>2</sup>
<i>Linear Regression</i>	1,41	3,31	0,89
<i>Bayesian Ridge</i>	1,42	3,32	0,89
<i>Ridge Regression</i>	1,42	3,31	0,89
<i>AdaBoost Regressor</i>	1,48	3,47	0,88
<i>Gradient Boosting Regressor</i>	1,45	3,38	0,88
<i>CatBoost Regressor</i>	1,48	3,47	0,88
<i>Extra Trees Regressor</i>	1,44	3,48	0,88
<i>Extra Trees Regressor</i>	1,46	3,42	0,88
<i>Orthogonal Matching Pursuit</i>	1,45	3,51	0,88
<i>Extreme Gradient Boosting</i>	1,53	3,74	0,87
<i>Least Angle Regression</i>	1,53	4,03	0,86
<i>Huber Regressor</i>	1,65	4,43	0,85
<i>Lasso Regression</i>	1,68	4,69	0,84
<i>Decision Tree Regressor</i>	1,79	5,37	0,81
<i>K Neighbors Regressor</i>	1,81	5,96	0,80
<i>Elastic Net</i>	2,07	6,70	0,77

Desta forma, no conjunto de treinamento, o modelo *Bayesian Ridge* obteve a melhor performance com R<sup>2</sup> de 0,89, seguido pelo *AdaBoost Regressor* com R<sup>2</sup> igual a 0,88 e o *Gradient Boosting Regressor*, salientando-se com R<sup>2</sup> de 0,88. Após o processo de *tune*, *ensemble*, *blend*, e *stack*, todos os modelos gerados foram aplicados sobre o conjunto de teste (Tabela 3), destacando-se o modelo combinado, via *blend*, dos três algoritmos utilizados, alcançando um valor de R<sup>2</sup> de 0,79.

Tabela 3. Performance obtida nos testes, a partir da aplicação dos modelos sobre o conjunto de dados desconhecidos, nos processos *Blend*, *Stack*, *Default*, *Ensembled*, *Tuned*.

Modelo	Processo	R <sup>2</sup>	MAE	MSE
Modelo Combinado - <i>Blend</i>	-	0,79	1,73	4,85
Modelo Combinado - <i>Stacked</i>	-	0,78	1,76	5,08
<i>AdaBoost Regressor</i>	<i>Default</i>	0,76	1,92	5,53
<i>AdaBoost Regressor</i>	<i>Ensembled</i>	0,77	1,84	5,11
<i>AdaBoost Regressor</i>	<i>Tuned</i>	0,76	1,90	5,33
<i>Bayesian Ridge</i>	<i>Default</i>	0,77	1,78	5,20
<i>Bayesian Ridge</i>	<i>Ensembled</i>	0,78	1,74	5,00
<i>Bayesian Ridge</i>	<i>Tuned</i>	0,78	1,76	5,07
<i>Gradient Boosting Regressor</i>	<i>Default</i>	0,75	1,89	5,64
<i>Gradient Boosting Regressor</i>	<i>Ensembled</i>	0,78	1,76	5,08
<i>Gradient Boosting Regressor</i>	<i>Tuned</i>	0,77	1,82	5,24

O modelo combinado, via *blend*, supera os três modelos selecionados previamente durante o treinamento. Isto pode ter ocorrido devido a sua característica de combinar modelos, aproveitando a complementaridade de todos, a fim de fornecer previsões aprimoradas (CANABARRO, 2019).

Referente ao desempenho do modelo combinado, Zeng, et al. (2021), indica que em sua análise do desenvolvimento e validação do modelo, ele superou os escores convencionais na previsão, mostrando-se flexível na integração de diferentes modelos de aprendizado de máquinas. Em consonância, este modelo evidenciou, também, maior adequação para a previsão de volume da madeira de *Eucalyptus*.

Técnicas estatísticas clássicas, como correlação e regressão não linear possuem dificuldades para modelagem de dados biológicos, principalmente por retornarem *overfitting*, conforme Reis, et al. (2018). Assim, a modelagem feita por *machine learning*, alcançou boa performance na previsão, sendo capaz de atender às necessidades para esse conjunto de dados.

## CONCLUSÕES

O modelo combinado, via *blend*, gerado a partir dos algoritmos *AdaBoost Regressor*, *Bayesian Ridge* e *Gradient Boosting Regressor*, alcançou alta performance de previsão, R<sup>2</sup> de 0,79, para o volume de madeiras de *Eucalyptus* transportadas pelo modal rodoviário, por meio do aprendizado de máquina.

## AGRADECIMENTOS

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

## REFERÊNCIAS

- ABIRI, R.; ATABAKI, N.; ABDUL-HAMID, H.; SANUSI, R.; SHUKOR, N. A. A.; SHAHARUDDIN, N. A.; AHMAD, S. A.; MALIK, S. The prospect of physiological events associated with the micropropagation of *Eucalyptus* sp. **Forests**, v. 11, n. 11, p.1211, 2020.
- ARUMÄE, T.; LANG, M.; SIMS, A.; LAARMANN, D. planning of commercial thinnings using machine learning and airborne lidar data. **Forests**, v. 13, n. 2, p.206, 2022.
- 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>>. 2022.
- CANABARRO, A.; BRITO, S.; CHAVES, R. Machine learning nonlocal correlations. **Physical Review Letters**, v. 122, n. 20, p.200401, 2019.
- CARLEO, G.; CIRAC, I.; CRANMER, K.; DAUDET, L.; SCHULD, M.; TISHBY, N.; VOGT-MARANTO, L.; ZDEBOROVÁ, L. Machine learning and the physical sciences. **Reviews of Modern Physics**, v. 91, n. 4, p.1-39, 2019.
- ÇIHAN, P.; GÖKÇE, E.; KALIPSIZ, O. A review of machine learning applications in veterinary field. **Kafkas Üniversitesi Veteriner Fakültesi Dergisi**, v. 23, n. 4, p.673-680, 2017.
- DANTAS, D.; CALEGARIO, N.; ACERBI JÚNIOR, W. A.; CARVALHO, S. P. C.; ISAAC JÚNIOR, M. A.; MELO, E. A. Multilevel nonlinear mixed-effects model and machine learning for predicting the volume of *Eucalyptus* spp. trees. **Cerne**, v. 26, n. 1, p.48-57, 2020.
- ELLI, E. F.; SENTELHAS, P. C.; CARNEIRO, R. L.; ALVARES, C. A. Gauging the effects of climate variability on *Eucalyptus* plantations productivity across Brazil: a process-based modelling approach. **Ecological Indicators**, v. 114, n. 7, p.106325, 2020.
- FERNANDES, B. C. C.; MENDES, K. F.; TORNISIELO, V. L.; TEÓFILO, T. M. S.; TAKESHITA, V.; CHAGAS, P. S. F.; LINS, H. A.; SOUZA, M. F.; SILVA, D. V. Effect of pyrolysis temperature on eucalyptus wood residues biochar on availability and transport of hexazinone in soil. **International Journal of Environmental Science and Technology**, v. 19, n. 1, p.499-514, 2022.
- FLUKE, C. J.; JACOBS, C. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. **WIREs Data Mining and Knowledge Discovery**, v. 10, n. 2, e1349, 2020.
- FREITAS, E. C. S.; PAIVA, H. N.; NEVES, J. C. L.; MARCATTI, G. E.; LEITE, H. G. Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops & Products**, v. 146, n. 4, p.112149, 2020.

GAUR, H.; DAKSSA, L.; DAWOOD, M.; SAMAIYA, N. K. A novel stress-based formulation of finite element analysis<sup>#</sup>. **Journal of Zhejiang University-SCIENCE A**, v. 22, n. 6, p.481-491, 2021.

HARRIS, C. R.; MILLMAN, K. J.; WALT, S. J. V. D.; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; KERKWIJK, M. H. V.; BRETT, M.; HALDANE, A.; RÍO, J. F.; WIEBE, M.; PETERSON, P.; GÉRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. **Nature** v. 585, n. 9, p.357–362, 2020.

KOIRALA, A.; KIZHA, A. R.; ROTH, B. E. Perceiving major problems in forest products transportation by trucks and trailers: a cross-sectional survey. **European Journal of Forest Engineering**, v. 3, n. 1, p.23-34, 2017.

LACHINI, E.; FIEDLER, N.; SILVA, G.; PINHEIRO, C.; CARMO, F. Operational analysis of forestry transportation using self-loading trucks in a mountainous region. **Floresta e Ambiente**, v. 25, n. 4, e20150060, 2018.

LEMAÎTRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. **Journal of Machine Learning Research**, v. 18, n. 1, p.1-5, 2017.

LIAKOS, K. G.; BUSATO, P.; MOSHOU, D.; PEARSON, S.; BOCHTIS, D. Machine learning in agriculture: a review. **Sensors**, v. 18, n. 8, p.2674, 2018.

MALLADI, K. T.; SOWLATI, T. Optimization of operational level transportation planning in forestry: a review. **International Journal of Forest Engineering**, v. 28, n. 3, p.198–210, 2017.

MAHESH, B. Machine learning algorithms: a review. **International Journal of Science and Research**, v. 9, n. 1, p.381-386, 2020.

MCKINNEY, W. **Python for data analysis: data wrangling with Pandas, Numpy, and IPython**. Massachusetts: O'Reilly Media; 2017. v. 2. 550.

MÜLLER, F.; JAEGER, D.; HANEWINKEL, M. Digitization in wood supply: a review on how Industry 4.0 will change the forest value chain. **Computers and Electronics in Agriculture**, v. 162, n. 7, p.206-218, 2019.

NELLI, F. Machine learning with Scikit-Learn. In: NELLI, F. **Python Data Analytics**. Berkeley: Apress; 2018. v. 2. 569 p.

PYCARET (2020). **Low-code machine learning**. PyCaret version 2.3.4. URL <https://pycaret.org/>.

REIS, A. A.; CARVALHO, M. C.; MELLO, J. M.; GOMIDE, L. R.; FERRAZ FILHO, A. C.; ACERBI JUNIOR, F. W. Spatial prediction of basal area and volume in *Eucalyptus* stands using Landsat TM data: an assessment of prediction methods. **New Zealand Journal of Forestry Science**, v. 48, n. 1, p.1-17, 2018.

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

SUHAIMI, N. A. D.; ABAS, H. A systematic literature review on supervised machine learning algorithms. **PERINTIS eJournal**, v. 10, n. 1, p.1-24, 2020.

TYMENDORF, L.; TRZCIŃSKI, G.. Multi-factorial load analysis of pine sawlogs in transport to sawmill. **Forests**, v. 11, n. 4, p.366, 2020.

UCCI, D.; ANIELLO, L.; BALDONI, R. Survey of machine learning techniques for malware analysis. **Computers & Security**, v. 81, n. 2, p.123-147, 2019.

VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E.; HABERLAND, M.; REDDY, T.; COURNAPEAU, D.; BUROVSKI, E.; PETERSON, P.; WECKESSER, W.; BRIGHT, J.; VAN DER WALT, S. T.; BRETT, M.; WILSON, J.; MILLMAN, K. J.; MAYOROV, N.; NELSON, A. R. J.; JONES, E.; KERN, R.; LARSON, E.; CAREY, C. J.; POLAT, I.; FENG, Y.; MOORE, E. W.; VANDERPLAS, J.; LAXALDE, D.; PERKTOLD, J.; CIMRMAN, R.; HENRIKSEN, I.; QUINTERO, E. A.; HARRIS, C. R.; ARCHIBALD, A. M.; RIBEIRO, A. H.; PEDREGOSA, F.; VAN MULBREGT, P. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 1, p.261–272, 2020.

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

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.

ZENG, Z.; YAO, S.; ZHENG, J.; GONG X. Development and validation of a novel blending machine learning model for hospital mortality prediction in ICU patients with Sepsis. **BioData Mining**, v. 14, n. 40, p.1-15, 2021.