

CUSTOS DO TRANSPORTE RODOVIÁRIO DE MADEIRA DE *Eucalyptus*: ABORDAGEM EM APRENDIZAGEM DE MÁQUINA

SILVA, T.¹; SASSO JUNIOR, V. A.¹; ALMEIDA, R. O.²; MUNIS, R. A.¹; 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

O transporte de madeira, majoritariamente rodoviário, apresenta-se como um dos segmentos mais onerosos dentre as atividades florestais. Entretanto, a aplicação do *machine learning* pode resultar em ganhos de precisão e redução de custos na produção. Diante do exposto, o objetivo foi prever por meio do *machine learning*, via análise de regressão, os custos do transporte rodoviário de madeira, o qual foi realizado por dois caminhões com diferentes características técnicas. O estudo foi conduzido em florestas plantadas de *Eucalyptus grandis*, cujo processamento dessa matéria-prima foi realizado no estado Rio Negro, no Uruguai. A análise foi pautada na construção de um banco de dados estruturado a partir da seleção dos atributos. Destaca-se que os custos do modal rodoviário constituíram-se como o atributo alvo da predição. Após o processo de *data wrangling*, os dados foram balanceados por *oversampling* e divididos em conjunto de treinamento e teste. A partir da aplicação de 19 algoritmos de aprendizagem, modo *Default*, constatou-se que os modelos *Bayesian Ridge*, *Gradient Boosting Regressor* e *Random Forest Regressor* obtiveram melhor performance, sendo R^2 treinamento igual a 0,55, 0,53 e 0,52, respectivamente. Em todo processo de treinamento foi aplicado o *cross-validation*. Desse modo, o melhor modelo, utilizando o algoritmo *Random Forest Regressor*, em modo *Default*, atinge uma performance de 0,89, sendo um modelo de alta performance na predição de custos no transporte rodoviário de madeira de *Eucalyptus*.

Palavras-chave: florestas plantadas, operações florestais, *Random Forest*.

INTRODUÇÃO

Considera-se as florestas plantadas do gênero *Eucalyptus*, pertencente à família *Myrtaceae*, uma importante matéria-prima para diversos setores, como a indústria de papel e celulose. Embora as espécies de *Eucalyptus* sejam originalmente da Austrália e Tasmânia, nota-se que o gênero apresenta uma boa adaptabilidade climática e um rápido crescimento em variadas regiões tropicais e subtropicais. Destarte, ressalta-se que a colheita em florestas plantadas e o transporte de madeira apresentam grande importância econômica perante a produção florestal (MOREIRA, et al., 2017; SALEHI, et al., 2019; VIEIRA, et al., 2021).

Dentre os meios de transporte de madeira, os quais podem ser: aquaviário, ferroviário ou rodoviário, tem-se que cada um difere nos custos e nos fatores de uso. A partir disso, utiliza-se mais comumente o modal rodoviário no setor florestal, sendo que, a movimentação de madeira pode ser feita por diferentes tipos e configurações de caminhão e reboque, cuja escolha relaciona-se ao tipo de matéria-prima a ser transportada (KOIRALA, et al., 2017; MOKHIREV, et al., 2018).

O transporte rodoviário de madeira caracteriza-se pela movimentação da matéria-prima entre o plantio ou indústria florestal até o local de consumo, destaca-se o uso de estradas durante a operação. Consequente, alguns fatores podem resultar no potencial aumento dos custos de transporte rodoviário de madeira. Posto isto, esses fatores são: inclinação, comprimento, tipo e condição da estrada, além da carga e capacidade dos caminhões (LACHINI, et al., 2018; AKAY, et al., 2021).

O transporte rodoviário de madeira influencia o custo total do material lenhoso entregue as indústrias florestais, visto que, a distância entre os locais de abastecimento e o consumidor pode definir a viabilidade econômica do investimento. Segundo Trzcinski, et al. (2018), estima-se que a proporção dos custos relativos do transporte de madeira seja de, aproximadamente, 17% ao comparar com custos totais de atividades florestais, possuindo uma participação considerável dentre os outros setores da economia (FERNANDES, et al., 2020).

Tratando-se do avanço dos métodos computacionais, notou-se o uso de aprendizagem de máquina (*machine learning*) em diversas áreas do conhecimento, como em aplicações florestais, no qual resultou em ganhos de precisão e redução de custos na produção. Dessa forma, emprega-se o *machine learning* visando o uso mais eficiente dos dados pelas máquinas, de tal modo que, tornam-se necessários diferentes algoritmos para resolver problemas de dados (VENDRUSCOLO, et al., 2017; MAHESH, 2020).

Ponderando a importância das corretas tomadas de decisões no transporte de madeira, torna-se essencial aprender o padrão latente de dados históricos a partir do *machine learning*, a fim de modelar o comportamento de um sistema e automatizar a construção do modelo analítico. Isto posto, a capacidade preditiva de encontrar o melhor modelo utiliza-se de ferramentas específicas, assim, avalia-se a capacidade do modelo segundo a previsão de dados inéditos e a sua flexibilidade, tendo em vista a estrutura do modelo subjacente (TIZGHADAM, et al., 2019; LISKI, et al., 2020).

Diante deste contexto, objetivou-se prever por meio do *machine learning*, via análise de regressão, os custos do transporte rodoviário de madeira de florestas plantadas de *Eucalyptus*, o qual foi realizado por dois caminhões com diferentes características técnicas.

MATERIAL E MÉTODOS

A madeira alvo de estudo foi proveniente de florestas plantadas de *Eucalyptus grandis*, com idade média de 161 meses, pertencentes a uma empresa do segmento de celulose branqueada de fibra curta de eucalipto, localizada no estado Rio Negro, no Uruguai. Realizou-se a colheita mecanizada de madeira por meio do sistema *cut-to-length* mecanizado, portanto, com o processamento da madeira no local do corte de derrubada e em regime de manejo de corte raso.

O baldeio da madeira consistiu na remoção de toras com comprimento de 7,2m do interior do maciço florestal e na alocação em pilhas junto às estradas florestais. A densidade média da madeira era de $0,77 \text{ g cm}^{-3}$. O transporte de madeira foi realizado em rodovias,

predominantemente, de geometria plana e com pavimentação asfáltica. Ademais, possuía boas condições de conservação e pistas de rodagem simples. O fluxo do transporte deu-se a partir das viagens realizadas durante um período de 35 meses.

Foram considerados dois tipos de Combinação de Veículos para Transporte de Carga (CVC), caracterizadas como: CVC1: caminhão trucado contendo um eixo simples e um conjunto de eixos em *tandem* duplo, um reboque com dois eixos duplos, sendo o comprimento total de 20 m e o Peso Bruto Total Combinado (PBTC) de 45 t; CVC2: caminhão trator trucado contendo um eixo simples e um conjunto de eixos em *tandem* duplo, um semirreboque que dispunha de um conjunto de eixos em *tandem* triplo, comprimento total de 18,60 m e PBTC de 49,5 t.

Construiu-se o banco de dados estruturado a partir da seleção dos atributos CVC1 e CVC2 utilizados no transporte rodoviário de madeira. Ademais, adotou-se o uso de outros atributos (AT), a saber: idade do povoamento (AT1), densidade da madeira (AT2), quantidade de dias após a colheita da madeira (AT3), tempo do ciclo operacional (AT4), velocidade da CVC com carga (AT5), valor do frete por quilômetro percorrido (AT6) e custo do transporte rodoviário de madeira por metro cúbico (AT7). Sendo AT7, o atributo alvo da predição.

Ao longo do processo de ajuste dos dados, utilizou-se a linguagem de programação R, com auxílio dos pacotes *multidplyr*, *tidyverse*, *doFuture*, *foreach*, entre outros. Após o *data wrangling*, os dados foram balanceados por *oversampling* (SMOTE). Além disso, dividiram-se os dados da seguinte forma: conjunto de treinamento com 90% das instâncias e conjunto de teste com 10% das instâncias (MALDONADO, et al., 2019; BENGTTSSON, 2021; WALLIG, 2021; WICKHAM, 2021a; WICKHAM, 2021b; RStudio Team, 2022).

A partir da aplicação de 19 algoritmos de aprendizagem (modo *Default*), foram selecionados três algoritmos com melhor performance. Nesse caso, esses algoritmos foram submetidos ao processo de *tune*, *ensemble*, *blend* e *stack*, de modo que, em todo processo de treinamento foi aplicado o *cross-validation*. Ao realizar a comparação dos modelos, utilizou-se, principalmente, a métrica R^2 (treinamento ou teste), além do erro absoluto médio (MAE) e erro percentual médio absoluto (MAPE).

Por fim, ressalta-se o uso da linguagem de programação *Python* durante o processo de *machine learning*, com o auxílio das bibliotecas *SciPy*, *Scikit-learn*, *Pycaret*, *NumPy*, *Pandas* e *Matplotlib* (KADIYALA e KUMAR, 2017; LEMENKOVA, 2019; PYCARET, 2020; VAN ROSSUM e DRAKE, 2020; VIRTANEN, et al., 2020).

RESULTADOS E DISCUSSÃO

Durante as análises exploratórias dos dados, foi detectada a presença de *outliers* nos atributos: tempo de operação (AT4) e velocidade da CVC com carga (AT5) (Tabela 1). Dessa forma, foram constatadas, inicialmente, cerca de 1.065 instâncias, contudo, após a retirada dos *outliers*, totalizaram 1.019 instâncias.

Tabela 1. Média, desvio padrão, assimetria e curtose dos atributos, com e sem a presença de *outliers*.

Atributos	Com <i>outliers</i>			Sem <i>outliers</i>		
	Média	Assimetria	Curtose	Média	Assimetria	Curtose
AT1 (meses)	161,20 (12,82)	0,44	-0,45	161,40 (12,84)	0,44	-0,49
AT2 (g cm ⁻³)	0,77 (0,06)	0,18	-0,43	0,77 (0,06)	0,18	-0,43
AT3 (dias)	80,45 (75,64)	2,83	10,25	81,01 (76,76)	2,81	9,96
AT4 (hora)	6,20 (1,17)	0,89	1,55	6,13 (1,01)	0,42	-0,23
AT5 (km h ⁻¹)	42,62 (6,99)	0,53	1,94	42,54 (6,19)	0,14	-0,19
AT6 (USD km ⁻¹)	3,19 (0,26)	0,86	1,19	3,18 (0,23)	0,47	-0,18
AT7 (USD m ⁻³)	0,11 (0,01)	0,63	0,92	0,11 (0,01)	0,43	0,19

Após a aplicação de 19 algoritmos de aprendizagem, constatou-se que os modelos *Bayesian Ridge* (BR), *Gradient Boosting Regressor* (GBR) e *Random Forest Regressor* (RF) obtiveram melhor performance, sendo R^2 treinamento igual a 0,55, 0,53 e 0,52, respectivamente (Figura 1). Vale destacar que, além do R^2 , ponderaram-se também as métricas MAE e MAPE.

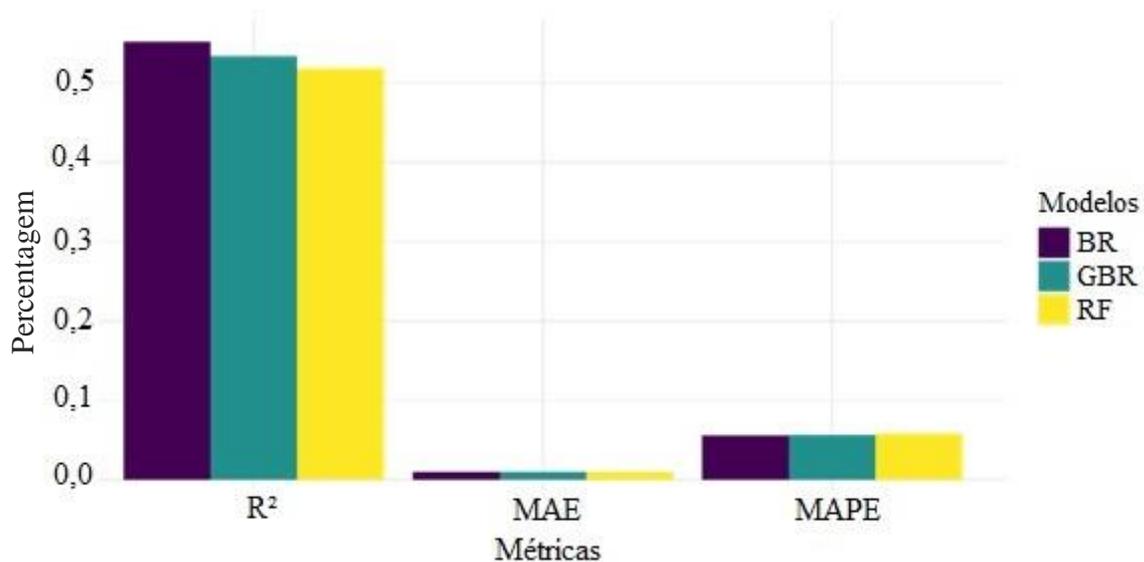


Figura 1. Comparação dos algoritmos aplicados a partir das métricas R^2 treinamento, MAE e MAPE.

Posteriormente ao processo de *tune*, *ensemble*, *blend*, e *stack*, comparou-se todos os modelos criados, de forma a selecionar o modelo de maior performance. Logo, o modelo *Random Forest Regressor* (modo *Default*) atingiu a performance do R^2 teste de 0,886 (Tabela 2). Torna-se importante ressaltar que, ao analisar os modelos, isto é, *Blended*, *Stacked*, *Bayesian Ridge*, *Gradient Boosting Regressor* e *Random Forest Regressor*, nota-se que somente o modelo RF (modo *Default*) alcançou um valor de R^2 teste superior a 0,80.

Tabela 2. Comparação dos modelos por meio da métrica R^2 treinamento e R^2 teste.

Modelos	Modos	R^2 treinamento	R^2 teste
<i>Blended</i>	-	0,602	0,588
<i>Stacked</i>	-	0,596	0,588
<i>Bayesian Ridge</i>	<i>Default</i>	0,552	0,563
	<i>Ensembled</i>	0,598	0,567
	<i>Tuned</i>	0,552	0,566
<i>Gradient Boosting Regressor</i>	<i>Default</i>	0,534	0,660
	<i>Ensembled</i>	0,588	0,571
	<i>Tuned</i>	0,545	0,571
<i>Random Forest Regressor</i>	<i>Default</i>	0,519	0,886
	<i>Ensembled</i>	0,586	0,596
	<i>Tuned</i>	0,552	0,596

Conforme Lopes, et al. (2021), o *Random Forest Regressor* caracteriza-se como um método de *machine learning* não paramétrico, sendo indicado na produção de modelos, pois otimiza a predição por meio do ajuste de um conjunto de árvores, as quais são formadas por variáveis e uma amostra aleatória. Assim, Tribby, et al. (2017) apontam que o *Random Forest Regressor* faz a divisão dos dados de entrada em dois conjuntos, sendo eles: dados de treinamento e dados de teste.

Ao utilizar o modelo *Random Forest Regressor* em seus estudos, Pereira, et al. (2021) observou que o desempenho da métrica R^2 foi superior as demais métricas testadas, sendo que, as métricas em comum entre ambos os estudos eram: R^2 e MAE. Logo, utiliza-se o *Random Forest Regressor* tanto para tarefas de classificação e regressão, devido à flexibilidade do modelo, visto que, geralmente entrega ótimos resultados, do ponto de vista computacional, mesmo sem realizar o ajuste dos hiperparâmetros (ABDULKAREEM e ABDULAZEEZ, 2021).

Por conseguinte, o uso do *machine learning* no setor florestal é tido como uma tarefa interdisciplinar promissora, uma vez que, permite reduzir os custos das empresas florestais.

Diante disso, torna-se vantajoso para o transporte rodoviário de madeira a implementação desses algoritmos de aprendizagem, dado que, o modal rodoviário é considerado oneroso dentro do setor florestal (KORCHAGIN, et al., 2020).

CONCLUSÃO

A aprendizagem de máquina demonstra um significativo potencial ao ser aplicado na modelagem florestal. Assim, o melhor modelo gerado na predição de custos no transporte rodoviário de madeira de *Eucalyptus* é a partir do algoritmo *Random Forest Regressor*, em modo *Default*, com performance de 0,89. O modelo apresenta alta performance e a predição dos custos mostra-se próxima dos valores reais, isto é, conjunto de teste.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

REFERÊNCIAS

ABDULKAREEM, N. M.; ABDULAZEEZ, A. M. Machine learning classification based on Radom Forest algorithm: a review. **International Journal of Science and Business**, v. 5, n. 2, p.128-142, 2021.

AKAY, A. E.; SERIN, H.; SESSIONS, J.; BILICI, E.; PAK, M. Evaluating the effects of improving forest road standards on economic value of forest products. **Croatian Journal of Forest Engineering**, v. 42, n. 2, p.245-258, 2021.

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.

FERNANDES, D. L.; MATOS, L. M. A.; MAGALHÃES, E. C.; CABACINHA, C. D.; ASSIS, A. L.; ARAÚJO JÚNIOR, C. A. Influence of wood stacking location on forest transport costs. **Revista Floresta**, v. 50, n. 1, p.1047-1052, 2020.

KADIYALA, A.; KUMAR, A. Applications of Python to evaluate environmental data science problems. **Environmental Progress e Sustainable Energy**, v. 36, n. 6, p. 1580-1586, 2017.

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.

KORCHAGIN, S.; SERDECHNY, D.; KIM, R.; TERIN, D.; BEY, M. The use of machine learning methods in the diagnosis of diseases of crops. **E3S Web of Conferences**, v. 176, n. 1, p.1-4, 2020.

LACHINI, E.; FIEDLER, N. ; SILVA, G.; PINHEIRO, C.; CARMO, F. Operational analysis of forestry transportation using self-loading trucks in a mountainous region. **Floram - Floresta e Ambiente**, v. 25, n. 4, p.1-7, 2018.

LEMENKOVA, P. Processing oceanographic data by Python libraries Numpy, Scipy and Pandas. **Aquatic Research**, v. 2, n. 2, p.73-91, 2019.

LISKI, E.; JOUNELA, P.; KORPUNEN, H.; SOSA, A.; LINDROOS, O.; JYLHA, P. Modeling the productivity of mechanized CTL harvesting with statistical machine learning methods. **International Journal of Forest Engineering**, v. 31, n. 3, p.253-262, 2020.

LOPES, I. L.; ARAÚJO, L. A.; MIRANDA, E. N.; ABREU, V. S.; GOMES, V. S.; ALMEIDA, B. C.; GONÇALVES, A. F. A.; BARBOSA, L. O.; GOMIDE, L. R. **Aplicação de técnicas de regressão linear e aprendizagem de máquinas na predição da altura total de árvores de *Eucalyptus spp.*** Minas Gerais: Editora Científica Digital; 2021. v. 1. 440p.

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

MALDONADO, S.; LÓPEZ, J.; VAIRETTI, C. An alternative SMOTE oversampling strategy for high-dimensional datasets. **Applied Soft Computing Journal**, v. 76, n. 1, p.380–389, 2019.

MOKHIREV, A. P.; POZDNYAKOVA, M. O.; MEDVEDEV, S. O.; MAMMATOV, V. O. Assessment of availability of wood resources using geographic information and analytical systems (the Krasnoyarsk territory as a case study). **Journal of Applied Engineering Science**, v. 16, n. 3, p.313-319, 2018.

MOREIRA, J. M. M. A. P.; SIMIONI, F. J.; OLIVEIRA, E. B. Importância e desempenho das florestas plantadas no contexto do agronegócio brasileiro. **Revista Floresta**, v. 47, n. 1, p.85-94, 2017.

PEREIRA, K. D.; CARNEIRO, A. P. S.; SANTOS, G. R.; CARNEIRO, A. C. O.; LEITE, H. G.; BORGES, F. P. Study of the influence of wood properties on the charcoal production: applying the random forest algorithm. **Revista Árvore**, v. 45, n. 99, p.1-10, 2021.

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

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

SALEHI, B.; SHARIFI-RAD, J.; QUISPE, C.; LLAIQUE, H.; VILLALOBOS, M.; SMERIGLIO, A.; TROMBETTA, D.; EZZAT, S. M.; SALEM, M. A.; ZAYED, A.; CASTILLO, C. M. S.; YAZDI, S. E.; SEN, S.; ACHARYA, K.; SHAROPOV, F.; MARTINS, N. Insights into *Eucalyptus* genus chemical constituents, biological activities and health-promoting effects. **Trends in Food Science e Technology**, v. 91, n. 1, p.609-624, 2019.

TIZGHADAM, A.; KHAZAEI, H.; MOGHADDAM, M. H. Y.; HASSAN, Y. Machine learning in transportation. **Journal of Advanced Transportation**, v. 2019, n. 1, p.1-4, 2019.

TRIBBY, C. P.; MILLER, H. J.; BROWN, B. B.; WERNER, C. M.; SMITH, K. R. Analyzing walking route choice through built environments using random forests and discrete choice techniques. **Environment and Planning B: Urban Analytics and City Science**, v. 44, n. 6, p.1145-1167, 2017.

TRZCINSKI, G.; MOSKALIK, T.; WOJTAN, R. Total weight and axle loads of truck units in the transport of timber depending on the timber cargo. **Forests**, v. 9, n. 4, p. 164-176, 2018.

VAN ROSSUM, G.; DRAKE, F. L. Extending and embedding Python. **Release 3.10.4**. URL <https://fossies.org/linux/python-docs-pdf-a4/extending.pdf>. 2020.

VENDRUSCOLO, D. G. S.; CHAVES, A. G. S.; MEDEIROS, R. A.; SILVA, R. S.; SOUZA, H. S.; DRESCHER, R.; LEITE, H. G. Estimativa da altura de árvores de *Tectona grandis* L.f. utilizando regressão e redes neurais artificiais. **Nativa**, v. 5, n. 1, p.52-58, 2017.

VIEIRA, T. A. S.; ARRIEL, T. G.; ZANUNCIO, A. J. V.; CARVALHO, A. G.; BRANCO-VIEIRA, M.; CARABINEIRO, S. A. C.; TRUGILHO, P. F. Determination of the chemical composition of *Eucalyptus* spp. for cellulosic pulp production. **Forests**, v. 12, n. 12, p.1649-1664, 2021.

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.