

APRENDIZADO DE MÁQUINA APLICADO NA PREDIÇÃO DA ATIVIDADE DE MANUTENÇÃO DE MÁQUINAS FLORESTAIS

SASSO JUNIOR, V. A.¹; SILVA, T.¹; ALMEIDA, R. O.²; ROCHA, Q. S.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, Câmpus Muriaé.

RESUMO

A atividade de manutenção é importante para garantir a vida útil das máquinas florestais e interferem diretamente na produtividade da operação de colheita mecanizada de madeira. Diante do exposto, o objetivo foi verificar se a aprendizagem de máquina alcançou boa performance na predição do tipo de atividade de manutenção da máquina florestal autopropelida *feller buncher*. Com os dados coletados em uma floresta plantada de *Eucalyptus* no sistema de colheita da madeira *full tree*, ponderamos os atributos de data, horímetro e tipo de atividade de manutenção do *feller buncher*. Foi aplicado 14 algoritmos de aprendizagem de máquina utilizando as linguagens de programação R e Python. Os modelos *Linear Discriminant Analysis* e *Logistic Regression*, via *ensembled* e *default*, respectivamente, alcançam razoável performance de predição, ambos com *Matthew's Correlation Coefficient* igual a 0,29 e precisão de 0,51, para o tipo de atividade de manutenção da máquina florestal autopropelida *feller buncher*, por meio do aprendizado de máquina.

Palavras-Chave: algoritmos de aprendizagem; colheita de madeira; *feller buncher*.

INTRODUÇÃO

As florestas plantadas representam cerca de 7% da cobertura florestal mundial, representando aproximadamente metade da produção global de madeira. Neste sentido, há crescente procura pelo gênero *Eucalyptus*, visto seu rendimento na produção de matéria-prima florestal, crescimento rápido e adaptabilidade em diversas condições edafoclimáticas. Em vista disto, o gênero vem ganhando importância econômica e valor comercial (SABIU et al., 2017; SPINELLI, et al., 2018; ANDRADE et al., 2019; NÓIA JÚNIOR, et al., 2019).

A colheita de madeira pode ser conduzida mediante diferentes métodos e sistemas. Dentre estes, destaca-se o método mecanizado e o sistema de colheita de árvores inteiras ou *full tree*, composto pelas atividades de derrubada acumulada de árvores, executada pelo *feller buncher*, arraste dos feixes de madeira para fora do talhão por meio do *grapple skidder* e processamento destes feixes na margem da estrada realizada pelo *grapple saw* (BILICI et al., 2019, TASSINARI et al., 2019; LIMA, et al., 2020; MIYAJIMA, et al., 2020; GÜLCI et al., 2021).

O *feller buncher*, máquina florestal autopropelida com rodados pneumáticos ou esteiras, é equipado com cabeçote com disco de corte. Fatores como topografia, volume por árvore, tamanho da área de corte, entre outros, podem afetar sua produtividade e interferir nos custos

de operação. Além destes, a manutenção da máquina é de grande importância, pois representa 60% dos dispêndios (SIMÕES et al., 2018; DINIZ, et al., 2019; SOMAN, et al., 2019; STRANDGARD, et al., 2019).

A manutenção, definida como combinação de técnicas administrativas, de retenção ou restauração de um item para o estado em que realiza a função requerida, influi no tempo de operação e na eficiência da máquina. Para reduzir essa inatividade e os custos, vem buscando-se prever a manutenção por meio de aprendizagem de máquinas. (CARVALO, et al., 2019; WIJESINGHE e MALLAWARACHCHI, 2019; ÇINAR, et al., 2020).

Com o desenvolvimento da inteligência artificial, a aprendizagem de máquinas é amplamente aplicada em diversas áreas do conhecimento. Define-se aprendizagem de máquina como um processo computacional que, a partir de dados de entrada, executa uma tarefa sem ser estritamente programado. Busca reconhecer padrões nos dados que informam como os problemas invisíveis são tratados (LIAKOS, et al., 2018; CARLEO, et al., 2019; WEI, et al., 2019;).

Os algoritmos de aprendizagem de máquinas são subdivididos em três categorias. O aprendizado supervisionado utiliza dados de entrada e saída conhecidos para treinar o algoritmo. O não supervisionado trata do reconhecimento de padrões existentes não identificados dos dados. Já o aprendizado semi-supervisionado combina a função de ambos, podendo melhorar o desempenho em uma das tarefas (ALZUBI, et al., 2018; ANGELOPOULOS, et al., 2020; ENGELEN e HOOS, 2020; GAMBELLA, et al., 2021).

Diante do contexto, a predição da atividade de manutenção é importante na redução de custos e aumento da produtividade das máquinas florestais. Destarte, objetivou-se verificar se a classificação dos dados por aprendizagem de máquina alcançou boa performance na predição do tipo de atividade de manutenção da máquina florestal autopropelida *feller buncher*.

MATERIAL E MÉTODOS

Desenvolveu-se o estudo a partir de dados coletados provenientes de florestas plantadas de *Eucalyptus*, localizadas na região Centro-Oeste do estado de São Paulo, Brasil. O sistema de colheita da madeira utilizado foi o *full tree*, cuja operação de derrubada foi feita de forma mecanizada, por meio do *feller buncher*. Foram estudados três *feller bunchers* (FB) da marca *John Deere*, sendo o FB1 do modelo 903K e o FB2 e FB3 do modelo 903M, cuja características estão expostas a seguir:

O FB1 apresentava motor com potência nominal de 199 kW, massa aproximada de 29.630 kg, rodados de esteira, alcance máximo da grua de 7,01 m e horímetro com 35.842 horas de uso acumuladas. Já o FB2 apresentava motor com potência nominal de 213 kW, massa aproximada de 31.590 kg, rodados de esteira, alcance máximo da grua de 7,01 m e horímetro com 24.857 horas de uso acumuladas. O FB3 apresentava motor com potência nominal de 213 kW, massa aproximada de 31.590 kg, rodados de esteira, alcance máximo da grua de 7,01 m e horímetro com 13.464 horas de uso acumuladas. Dessarte, os três *feller bunchers* apresentavam

o cabeçote de corte do modelo FR22B, composto por um disco de corte, com capacidade de corte de 559 mm e capacidade de acúmulo de 0,48 m³.

Ponderou-se os dados coletados a partir das instâncias de horímetro e tipo de atividade de manutenção, a saber: manutenção hidráulica, elétrica, mecânica e preventiva; sendo estas o alvo da predição. Realizou-se o ajuste dos dados fazendo uso da linguagem de programação R, a partir de pacotes como o *doFuture*, *foreach*, *multidplyr*, *tidyverse* (WICKHAM, et al., 2019; ZHANG, et al., 2019; TOPÇUOĞLU, et al., 2021; RSTUDIO TEAM, 2022).

Ulterior a execução do *data wrangling*, utilizou-se o aprendizado supervisionado, com balanceamento de dados por *oversampling – SMOTE*, além do banco de dados dividido em conjunto de teste e de treinamento com, respectivamente, 10% e 90% das instâncias (LUO, et al., 2019; PARSONS, et al., 2021).

Aplicou-se 14 algoritmos de aprendizagem de máquina, modo *default*, para posterior seleção dos algoritmos com melhor performance. Os selecionados passaram pelos processos de *tune*, *ensemble*, *blend*, e *stack*, utilizando *cross-validation*, visando alcançar performances de predição melhores.

As principais métricas utilizadas foram o coeficiente de correlação de Matthews ou *Matthew's Correlation Coefficient* (MCC) e a Precisão. O MCC varia nos intervalos de +1 e -1, como acerto e erro de classificação, respectivamente. Ademais, quando o valor margeia 0, a métrica indica aleatoriedade (CHICCO e JURMAN, 2020).

Foi conduzida a realização do aprendizado de máquina pela linguagem de programação *Python*, utilizando as seguintes bibliotecas: *Matplotlib*, *NumPy*, *Pandas*, *pycaret*, *Scikit-learn* e *SciPy* (JOLLY, 2018; LEMENKOVA, 2019; DOMINGUEZ, et al., 2021; PURCELL, et al., 2022).

RESULTADOS E DISCUSSÃO

Na análise exploratória, a base de dados utilizada continha o atributo horímetro e o atributo alvo tipo de atividade de manutenção, não apresentou presença de outliers. Dessarte, com a aplicação dos algoritmos de aprendizagem aos dados de treinamento (Tabela 1), modo *default*, foram selecionados três diferentes algoritmos, a saber: *Ada Boost Classifier*, *Linear Discriminant Analysis*, *Logistic Regression*.

Tabela 1. Performance obtida com a aplicação dos 14 algoritmos de aprendizagem de máquina sobre o conjunto de treinamento.

Modelo	MCC	Precisão
<i>Logistic Regression</i>	0,22	0,52
<i>Ridge Classifier</i>	0,22	0,52

<i>Linear Discriminant Analysis</i>	0,21	0,52
<i>Ada Boost Classifier</i>	0,14	0,47
<i>Naive Bayes</i>	0,13	0,23
<i>SVM - Linear Kernel</i>	0,10	0,34
<i>CatBoost Classifier</i>	0,10	0,47
<i>Gradient Boosting Classifier</i>	0,08	0,46
<i>K Neighbors Classifier</i>	0,06	0,39
<i>Random Forest Classifier</i>	0,06	0,43
<i>Quadratic Discriminant Analysis</i>	0,06	0,27
<i>Extra Trees Classifier</i>	0,05	0,43
<i>Decision Tree Classifier</i>	0,03	0,40
<i>Dummy Classifier</i>	0,00	0,10

No conjunto de teste (Tabela 2), destacaram-se tanto o modelo *Linear Discriminant Analysis* no processo *tuned*, quanto o *Logistic Regression* no *default*.

Tabela 2. Performance obtida no conjunto de teste, a partir da aplicação dos modelos sobre o conjunto de dados desconhecidos, nos processos *Blend*, *Stack*, *Default*, *Ensembled*, *Tuned*.

Modelo	Processo	MCC	Precisão
Modelo Combinado – <i>Blend</i>	-	0,28	0,23
Modelo Combinado – <i>Stacked</i>	-	0,28	0,23
<i>Ada Boost Classifier</i>	<i>Default</i>	0,16	0,39
<i>Ada Boost Classifier</i>	<i>Ensembled</i>	0,28	0,23
<i>Ada Boost Classifier</i>	<i>Tuned</i>	0,28	0,23
<i>Linear Discriminant Analysis</i>	<i>Default</i>	0,20	0,45
<i>Linear Discriminant Analysis</i>	<i>Ensembled</i>	0,29	0,51

<i>Linear Discriminant Analysis</i>	<i>Tuned</i>	0,29	0,51
<i>Logistic Regression</i>	<i>Default</i>	0,29	0,51
<i>Logistic Regression</i>	<i>Ensembled</i>	0,29	0,51
<i>Logistic Regression</i>	<i>Tuned</i>	0,29	0,51

O modelo *Linear Discriminant Analysis* no processo *tune* e o modelo *Logistic Regression* no processo *default*, apresentaram maior performance, sendo estes com menor complexidade. Apesar dos resultados apresentarem os maiores valores para as métricas MCC e Precisão, tais modelos necessitam de ajustes para alcançar melhores performances de predição.

O principal ponto que influenciou para esta performance foi a baixa quantidade de dados, comprometendo assim os modelos gerados. A base de dados continha apenas três *feller bunchers* com tempo de uso distintos, de 13.464, 24.857 e 35.842 horas, em média, formando *gaps* no atributo horímetro e, conseqüentemente, limitando a performance preditiva dos modelos gerados.

Na aprendizagem de máquina, são normalmente utilizados diversos atributos para predição do atributo alvo, de modo a alcançar altas performances de predição (CHENG, et al., 2020; KAPARTHI e BUMBLAUSKAS, 2020; AYVAZ; ALPAY, 2021). No estudo, foi utilizado somente um atributo, horímetro, para a predição do atributo alvo, tipo de atividade de manutenção, o que pode ter comprometido a performance de predição.

CONCLUSÕES

A predição do tipo de atividade de manutenção da máquina florestal autopropelida *feller buncher* por aprendizagem de máquina atingiu baixa performance.

Dos modelos testados, o *Linear Discriminant Analysis* e o *Logistic Regression*, via *tune* e *default*, apresentaram melhor performance.

A quantidade de dados não foi suficiente para gerar uma boa performance nos modelos, necessitando de ajustes tanto na base de dados quanto nos hiperparâmetros dos algoritmos de aprendizagem, a fim de alcançarem melhores resultados.

AGRADECIMENTOS

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

REFERÊNCIAS

- ALZUBI, J.; NAYYAR, A. KUMAR, A. Machine Learning from Theory to Algorithms: an overview. **Journal of Physics**, v. 1142, n. 5, p.012012, 2018.
- ANDRADE, J. R.; MAIA-JÚNIOR, S. O.; SANTOS, A. F. S.; SILVA, V. M.; BEZERRA, L. T.; SILVA, J. R. R.; SANTOS, C. M.; FERREIRA, V. M.; ENDRES, L. Photosynthetic performance in Eucalyptus clones cultivated in saline soil. **Emirates Journal of Food and Agriculture**, v. 31, n. 5, p.368-379, 2019.
- ANGELOPOULOS, A.; MICHAILIDIS, E. T.; NOMIKOS, N.; TRAKADAS, P.; HATZIEFREMIDIS, A.; VOLIOTIS, S. ZAHARIADIS, T. Tackling faults in the industry 4.0 era: a survey of machine-learning solutions and key aspects. **Sensors**, v. 20, n. 1, p.109, 2020.
- AYVAZ, S.; ALPAY, K. Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time. **Expert Systems With Applications**, v. 173, n. 11, p.114598, 2021.
- BILICI, E.; AKAY, A. E.; ABBAS, D. Assessing the effects of site factors on the productivity of a feller buncher: a time and motion analysis. **Journal of Forestry Research**, v. 30, n. 4, p.1471-1478, 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.045002, 2019.
- CARVALO, 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. 11, p.106024, 2019.
- CHENG, J. C. P.; CHEN, W.; CHEN, K.; WANG, Q. Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. **Automation in Construction**, v. 112, n. 4, p.103087, 2020.
- CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, v. 21, n. 6, p.1-13, 2020.
- 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, 2020.
- DINIZ, C. C. C.; LOPES, E. S.; MIRANDA, G. M.; KOEHLER, H. S.; SOUZA, E. K. C. Analysis of indicators and cost of world class maintenance (wcm) in forest machines. **Floresta**, v. 49, n. 3, p.533 -542, 2019.
- DOMINGUEZ, D. A.; EJEH, J.; DUNBAR, A. D. F.; BROWN, S. F. Machine learning approach for electric vehicle availability forecast to provide vehicle-to-home services. **Energy Reports**, v. 7, n. 2, p.71-80, 2021.

ENGELEN, J. E. V.; HOOS, H. H. A survey on semi-supervised learning. **Machine Learning**, v. 109, n. 2, p.373–440, 2020.

GAMBELLA, C.; GHADDAR, B.; NAOUM-SAWAYA, J. Optimization problems for machine learning: a survey. **European Journal of Operational Research**, v. 290, n. 2, p.807–828, 2021.

GÜLCI, N.; YÜKSEL, K; GÜLCI, S.; SERIN, H.; BILICI, E.; AKAY, A. E. Analysis of a feller-buncher productivity: a case study of whole-tree harvesting from Marmara region, Turkey. **Annals of Forest Research**, v. 64, n. 1, p.99-110, 2021.

JOLLY, E. Pymer4: connecting r and python for linear mixed modeling. **Journal of Open Source Software**, v. 3, n. 31, p.862, 2018.

KAPARTHI, S.; BUMBLAUSKAS, D. Designing predictive maintenance systems using decision tree-based machine learning techniques. **International Journal of Quality & Reliability Management**, v. 37, n. 4, p.659-686, 2020.

LEMENKOVA, P. Processing oceanographic data by python libraries numpy, scipy and pandas. **Aquatic Research**, v. 2, n. 2, p.73-91, 2019.

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.

LIMA, L. F.; PELISSARI, A. L.; RODRIGUES, C. K.; SOUSA, N. J.; CORTE, A. P. D. Quality assessment of pine wood harvesting by residue inventory using line intercept cluster sampling. **International Journal of Forest Engineering**, v. 31, n. 3, p.205-210, 2020.

LUO, M.; WANG, K.; CAI, Z.; LIU, A.; LI, Y.; CHEANG, C. F. Using imbalanced triangle synthetic data for machine learning anomaly detection. **Computers, Materials & Continua**, v. 58, n. 1, p.15-26, 2019.

MIYAJIMA, R. H.; FENNER, P. T.; BATISTELA, G. C.; SIMÕES, D. Technical-economic analysis of grapple saw: a stochastic approach. **Croatian Journal of Forest Engineering**, v. 41, n. 2, p.219-229, 2020.

NÓIA JÚNIOR, R. S.; SCHWERS, F.; SAFANELLI, J. L.; RODRIGUES, J. C.; SENTELHAS, P. C. Eucalyptus rust climatic risk as affected by topography and ENSO phenomenon. **Australasian Plant Pathology**, v. 48, n. 2, p.131-141, 2019.

PARSONS, S.; AMANI, M.; MOGHIMI, A. Ocean colour mapping using remote sensing technology and an unsupervised machine learning algorithm. **Journal of Ocean Technology**, v. 16, n. 3, p.103-115, 2021.

PURCELL, T. A. R.; SCHEFFLER, M.; CARBOGNO, C.; GHIRINGHELLI, L. M. SISSO++: a C++ implementation of the sure-independence screening and sparsifying operator approach. **Journal of Open Source Software**, v. 7, n. 71, p.3960, 2022.

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

SABIU, S.; AJANI, E. O.; SUNMONU, T. O.; BALOGUN, F. O.; ASHAFI, A. O. T.; OTHMAN, R. B.; OLOWA, S. K. Mechanism of hepatoprotective potential of aqueous leaves extract of *Eucalyptus obliqua* (Myrtaceae) in carbon tetrachloride intoxicated Wistar rats. **Journal of Applied Pharmaceutical Science**, v. 7, n. 8, p.183-190, 2017.

SIMÕES, D.; MIYAJIMA, R. H.; TONIN, R. P.; FENNER, P. T.; BATISTELA, G. C. Incorporation of uncertainty in technical and economic analysis of a feller-buncher. **Floresta**, v. 48, n. 3, p.403-412, 2018.

SOMAN, H.; KIZHA, A. R.; ROTH, B. E. Impacts of silvicultural prescriptions and implementation of best management practices on timber harvesting costs. **International Journal of Forest Engineering**, v. 30, n. 1, p.14-25, 2019.

SPINELLI, R.; MOURA, A. C. A.; SILVA, P. M. Decreasing the diesel fuel consumption and CO₂ emissions of industrial in-field chipping operations. **Journal of Cleaner Production**, v. 172, n. 3, p.2174-2181, 2018.

STRANDGARD, M.; MITCHELL, R.; WIEDEMANN, J. Comparison of productivity, cost and chip quality of four balanced harvest systems operating in a eucalyptus globulus plantation in western australia. **Croatian Journal of Forest Engineering**, v. 40, n. 1, p.39-48, 2019.

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.

TASSINARI, D.; ANDRADE, M. L. C.; DIAS-JUNIOR, M. S.; MARTINS, R. P.; ROCHA, W. W.; PAIS, P. S. A. M.; SOUZA, Z. R. Soil compaction caused by harvesting, skidding and wood processing in eucalyptus forests on coarse-textured tropical soils. **Soil Use and Management**, v. 32, n. 3, p.400-411, 2019.

TOPÇUOĞLU, B. D.; LAPP, Z.; SOVACOOOL, K. L.; SNITKIN, E.; WIENS, J.; SCHLOSS, P. D. Mikropml: user-friendly R package for supervised machine learning pipelines. **Journal of Open Source Software**, v. 6, n. 61, p.3073, 2021.

WEI, J.; CHU, X.; SUN, XY.; XU, K.; DENG, HX.; CHEN, J.; WEI, Z.; LEI, M. Machine learning in materials science. **InfoMat**, v. 1, n. 3, p.338-358, 2019.

WICKHAM, H.; AVERICK, M.; BRYAN, J.; CHANG, W.; MCGOWAN, L. A.; FRANÇOIS, R.; GROLEMUND, G.; HAYES, A.; HENRY, L.; HESTER, J.; KUHN, M.; PEDERSEN, T. L.; MILLER, E.; BACHE, S. M.; MÜLLER, K.; OOMS, J.; ROBINSON, D.; SEIDEL, D. P.; SPINU, V.; TAKAHASHI, K.; VAUGHAN, D.; WILKE, C.; WOO, K.; YUTANI, H. Welcome to the Tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p.1686, 2019.

WIJESINGHE, D.; MALLAWARACHCHI, H. A systematic approach for maintenance performance measurement: apparel industry in Sri Lanka. **Journal of Quality in Maintenance Engineering**, v. 25, n. 1, p.41-53, 2019.

ZHANG, X.; NIEUWDORP, M.; GROEN, A. K.; ZWINDERMAN, A. H. Statistical evaluation of diet-microbe associations. **BMC Microbiology**, v. 19, n. 90, p.1-10, 2019.