**Predição do ruído ocupacional emitido na colheita mecanizada de madeira**

**Rafaele Almeida Munis1, Diego Aparecido Camargo1, Danilo Simões2**

1 Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agronômicas, Botucatu, São Paulo; 2 Universidade Estadual Paulista (Unesp), Câmpus Experimental de Itapeva, Itapeva, São Paulo. (rafaele.munis@gmail.com).

**RESUMO:** A predição do ruído ocupacional contribui para o gerenciamento florestal, sobretudo, com vistas à proteção da saúde e segurança dos operadores. Deste modo, foi avaliado se o algoritmo de florestas aleatórias é adequado na predição do ruído ocupacional emitido em operações de colheita mecanizada de madeira. Para isto, foram utilizadas as dosimetrias de máquinas florestais autopropelidas empregadas no corte, processamento e arraste da madeira em um sistema de árvores inteiras. Os dados foram particionados em treinamento e teste, para avaliação da performance do algoritmo de florestas aleatórias por meio do coeficiente de determinação ajustado ao risco e estimativa do erro quadrático médio. A performance do modelo de treinamento resultou em um coeficiente de determinação ajustado ao risco de 56,98%, contudo, quando avaliada os dados de teste, foi de 3,68%, ocorrendo *overfitting* do algoritmo de floresta aleatória aos dados de treinamento, não sendo viável na predição do ruído ocupacional.

**Palavras-chave:** aprendizado de máquinas, *Eucalyptus*, florestas aleatórias, *overfitting*

1. **INTRODUÇÃO**

Comumente os gestores florestais defrontam-se com situações que podem comprometer a saúde e segurança ocupacional de seus colaboradores, mais especificamente, operadores que atuam na colheita mecanizada de madeira. Exemplo disto, a dissipação de energia que ocorre por meio da geração de ruído, das fontes emissoras, isto é, das máquinas florestais autopropelidas (MARZANO et al., 2017).

Logo, a assertividade nas tomadas de decisões se faz necessária no cotidiano de gestores florestais. Deste modo, a aprendizagem de máquinas, apresenta-se como uma alternativa eficiente no processo de informatização dos dados (BAO; DING, 2020), conseguinte, predição de variáveis *target*, como o ruído ocupacional.

Dentre os algoritmos desenvolvidos para a predição de dados, denominado florestas aleatórias, destaca-se por ser um dos mais importantes e possibilitar a combinação de inúmeras árvores de decisão, com a relação entrada-saída aprendida em certos intervalos de confiança com a ajuda de dados experimentais (YEŞILKANAT, 2020).

Com vistas à proteção da saúde e segurança dos operadores de máquinas florestais autopropelidas, o emprego do algoritmo de aprendizagem de máquinas por floresta aleatória justifica-se por proporcionar a predição do ruído ocupacional aos quais os operadores são expostos e, conseguinte, adoção de medidas preventivas ou corretivas. Desta maneira, foi avaliado se o algoritmo de florestas aleatórias é adequado na predição do ruído ocupacional emitido em operações de colheita mecanizada de madeira.

1. **MATERIAL E MÉTODOS**

*Caracterização dos dados*

O estudo foi conduzido em operações da colheita mecanizada de madeira de *Eucalyptus*, ponderando 17 máquinas florestais autopropelidas empregadas no sistema de árvores inteiras, as quais dividem-se em:

Cinco *feller-bunchers* das marcas *Tigercat* e *John Deere*, modelos L 870 C e 953 MH com média de horas de uso acumulada de 11.978 ±5,2 horas; Quatro *grapple skidder* da marca *Tigercat*, modelos 635 D e 635 E com média de horas de uso acumulada de 12.258 ±8,9 horas; Cinco *grapple processor* da marca *John Deere*, modelos 351 G, 352 G, 903 k, 953 MH com média de horas de uso acumulada de 12.248 ±3,7 horas; Três *grapple saw* da marca *John Deere*, modelos 350 G e 909 k com média de horas de uso acumulada de 10.971 ±4,5 horas.

Destarte, o estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual Paulista "Júlio de Mesquita Filho" - Faculdade de Medicina - Câmpus de Botucatu, com o número do parecer 3.492.969.

As dosimetrias foram coletadas por meio de dois medidores integradores de uso pessoal da marca *Instrutherm*, modelo DOS-500 e modelo DOS-600. O período de observação foi de uma jornada de trabalho de oito horas por máquina, obtendo assim como variável resposta o nível médio de exposição diária.

*Predição do ruído ocupacional*

A fim de predizer o ruído emitido pelas máquinas florestais autopropelidas, foram consideradas como variáveis preditoras: horas de uso acumulada das máquinas (HM), tempo de experiência dos operadores (TE) e a máquina florestal autopropelida operada (MF). A aplicação do algoritmo de florestas aleatórias, foi por meio do *software* R (R CORE TEAM, 2019), pacote *randomForest* e função *randomForest*.

Destarte, os dados foram previamente particionados em 70% para treinamento e 30% para teste, sendo a performance avaliada por meio do coeficiente de determinação ajustado ao risco e a importância relativa das variáveis avaliadas em função do erro quadrático médio.

1. **RESULTADOS E DISCUSSÃO**

Na identificação entre as variáveis preditoras relacionadas, a que exercia maior influência no ruído ocupacional emitido, foi possível perceber que a MF foi responsável pela maior diminuição do erro médio quadrático (Figura 1).

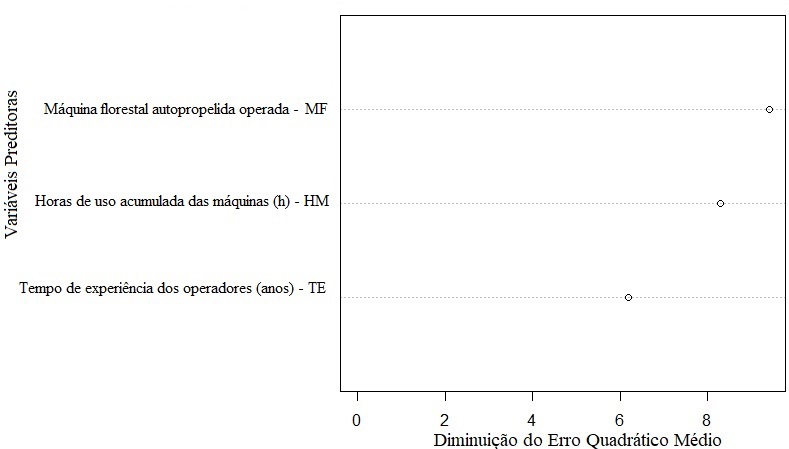


Figura 1. Importância relativa das variáveis preditoras do ruído ocupacional das máquinas florestais autopropelidas.

Na modelagem do ruído ocupacional por meio do algoritmo de florestas aleatórias, com a partição dos dados de treinamento, percebeu-se um coeficiente de determinação ajustado ao risco de 56,98%.

Contudo, de acordo com Yeşilkanat (2020), o nível de sucesso do modelo de estimativa é determinado testando os dados de validação após o aprendizado suficiente ser fornecido pela máquina. Deste modo, o coeficiente de determinação ajustado ao risco foi de 3,68%.

Este coeficiente de determinação reflete *overfitting* dos dados, uma vez que o algoritmo ficou sobre ajustado nos dados de treinamento, uma vez que a dosimetria das 17 máquinas florestais autopropelida utilizadas na predição do ruído ocupacional não foram suficientes para o aprendizado correto do algoritmo.

Ahmadi e Rezghi (2020) explicam que em aprendizado de máquina, há uma compensação entre o número de variáveis e a ocorrência de *overfitting*. Ademais, Ozyurt (2020) corrobora que isto também está associado ao tamanho apropriado do conjunto de dados para o treinamento.

1. **CONCLUSÕES**

A variável preditora no modelo ajustado com a partição dos dados de treinamento, com maior importância relativa é a máquina florestal autopropelida que está sendo operada.

O algoritmo de floresta aleatória não é adequado para predição do ruído ocupacional emitido em operações de colheita mecanizada de madeira em decorrência do *overfitting* dos dados.

1. **REFERÊNCIAS**

AHMADI, S.; REZGHI, M. Generalized low-rank approximation of matrices based on multiple transformation pairs. **Pattern Recognition**, v. 108, e. 107545, 2020.

BAO, K.; DING, Y. Network security analysis using big data technology and improved neural network. **Journal of Ambient Intelligence and Humanized Computing**, 2020. Disponível em: <https://doi.org/10.1007/s12652-020-02080-1> Acesso em: 26/09/2020.

MARZANO, F. L. C.; SOUZA, A. P.; MINETTE, L. J. Proposal for an ergonomic conformity index for evaluation of harvesters and forwarders. **Revista Árvore**, v. 41, n. 4, e. 410401, 2017.

OZYURT, F. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. **Journal of Supercomputing**, v. 6, n. 11, p. 8413-8431, 2020.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019. Disponível em: <https://www.R-project.org/> Acesso em: 26/09/2020.

YEŞILKANAT, C. M. Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. **Chaos, Solitons & Fractals**, v. 140, e. 110210, 2020.