UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

FCA
FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

RAFAELE ALMEIDA MUNIS

PREDIÇÃO DA PRODUTIVIDADE DE *HARVESTER* E *FORWARDER* POR ABORDAGEM DE *MACHINE LEARNING*

Botucatu

2023

RAFAELE ALMEIDA MUNIS

PREDIÇÃO DA PRODUTIVIDADE DE *HARVESTER* E *FORWARDER* POR ABORDAGEM DE *MACHINE LEARNING*

Tese de Doutorado apresentada à Faculdade de Ciências Agronômicas da Unesp Câmpus de Botucatu, para obtenção do título de Doutora em Ciência Florestal.

Orientador: Danilo Simões

Coorientador: Jaime Wojciechowski

Botucatu

2023

M966p

Munis, Rafaele Almeida

Predição da produtividade de harvester e forwarder por abordagem de machine learning / Rafaele Almeida Munis. -- Botucatu, 2023

103 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agronômicas, Botucatu

Orientador: Danilo Simões

Coorientador: Jaime Wojciechowski

1. Eucalyptus. 2. Colheita florestal. 3. Machine learning. 4. Produtividade. 5. Data mining. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: PREDIÇÃO DA PRODUTIVIDADE DE HARVESTER E FORWARDER POR ABORDAGEM DE MACHINE LEARNING

AUTORA: RAFAELE ALMEIDA MUNIS
ORIENTADOR: DANILO SIMOES
COORIENTADOR: JAIME WOJCIECHOWSKI

Aprovada como parte das exigências para obtenção do Título de Doutora em Ciência Florestal, pela Comissão Examinadora:

Prof. Dr. DANILO SIMOES (Participaçao Presencial)
Departamento de Ciencia Florestal Solos e Ambiente / FCAUnesp Botucatu

Prof. Dr. EDUARDO DA SILVA LOPES (Participaçao Virtual)
Ciência Florestal / Universidade Estadual do Centro Oeste

Prof. Dr. RODRIGO DE OLIVEIRA ALMEIDA (Participaçao Virtual)
Educação, Ciência e Tecnologia / Instituto Federal do Sudeste de Minas Gerais

Prof. Dr. RAFAEL PLANA SIMÕES (Participaçao Presencial)
Bioprocessos e Biotecnologia / Faculdade de Ciencias Agronomicas de Botucatu - UNESP

Prof. Dr. RICHARDSON BARBOSA GOMES DA SILVA (Participaçao Presencial) Richardson Barbosa Gomes da Silva
Pós-Doutorando - Departamento de Ciência Florestal, Solos e Ambiente / Faculdade de Ciências Agronômicas de Botucatu

Botucatu, 15 de dezembro de 2023

Aos meus amados avós,

Maria Lourdes, Rute, José Elias e Dorival,

dedico

AGRADECIMENTOS

A Deus.

Aos meus amados pais, Ana Andréa e Daniel Munis, por serem minha inspiração e exemplos de seres humanos a seguir, além de todo apoio e amor incondicional.

Aos meus queridos e amados irmãos, Aline e Alisson, pelo companheirismo, afeto e amor.

Ao meu amado companheiro, Diego, pela motivação, apoio e amor em todo percurso.

Ao Prof. Dr. Danilo Simões, pela orientação, ensinamentos, paciência e exemplo de pesquisador e educador.

Ao Prof. Dr. Jaime Wojciechowski pela co-orientação.

Aos professores que compõe a banca pelos ensinamentos compartilhados para o desenvolvimento do estudo: Prof. Dr. Rodrigo de Oliveira Almeida, Prof. Dr. Richardson Barbosa, Prof. Dr. Rafael Plana Simões, Prof. Dr. Eduardo da Silva Lopes.

Aos meus amigos de departamento pelo auxílio diário, em especial, ao Richardson, Rodrigo e Diego: vocês foram fundamentais para essa conclusão!

A Laryssa Cavalheiro e a Maria Eugênia, pela confiança e indicação: vocês proporcionaram o início da minha carreira profissional.

Aos meus amigos e companheiros de trabalho, em especial, a *squad* jurídico do ID, Bradesco.

A todos meus amigos e familiares que me auxiliaram e apoiaram durante o desenvolvimento do estudo.

A Faculdade de Ciências Agrárias da Universidade Estadual Paulista “Júlio de Mesquita Filho”, ao departamento de Ciência Florestal, Solos e Ambiente, ao Programa de Pós-Graduação em Ciência Florestal.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil – CAPES – Código de financiamento 001.

Este trabalho também foi realizado com o apoio da Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Portanto dele, por Ele e para Ele são todas as coisas. A Ele, pois, (cabe) a glória eternamente! Amém.”

Novo testamento: BÍBLIA, E. P. Romanos. In: Bíblia Sagrada. Tradução de Ivo Stornilo. Edição Pastoral. Brasília - SP: Editora Paulus, 1991.

RESUMO

Modelos de *machine learning* envolvendo algoritmos de regressão podem ser utilizados por gestores florestais na otimização dos processos em que empreendem e como suporte às tomadas de decisões. A excelente performance preditiva dos modelos de *machine learning* pode contribuir para o gerenciamento eficiente da cadeia de suprimentos, redução de desperdícios e, otimização do tempo, máquinas e equipamentos. Na operação de colheita mecanizada de madeira, a incorporação de ferramentas preditivas, como a técnica de *machine learning*, auxilia nos planejamentos estratégico, tático e operacional, reduzindo a complexidade e aproveitando a diversidade dos dados. Assim, como o planejamento pode ser associado a provisões, verificamos se a produtividade de máquinas florestais autopropelidas pode ser obtida por meio da implementação de algoritmos de *machine learning* no sistema *cut-to-length*. Foram avaliados dados da atividade de corte e processamento de madeira provenientes de florestas plantadas de *Pinus* e *Eucalyptus* no Brasil, realizados por 21 *harvesters* e 23 *forwarders* em sistemas *cut-to-length*. Métodos de aprendizado por *blending* e *stacking* foram testados, incorporando a permutação de 18 e 22 algoritmos na predição da produtividade de *harvesters* e *forwarders*, respectivamente. Os métodos de aprendizado e os algoritmos foram comparados a partir de métricas de performance. Mesmo com o comprometimento da qualidade da base, métricas de desempenho satisfatórias foram obtidas pelo *default* dos algoritmos. Na ingestão de dados aptos por algoritmos de *machine learning*, alcança-se coeficiente de determinação de 0,74 na estimativa da produtividade de *harvesters* por meio do *default* do algoritmo *extra trees regressor* e o modo *default* do *light gradient boosting* de 0,77 na estimativa da produtividade na estimativa da produtividade do *forwarder*.

Palavras-chave: qualidade de dados*; Eucalyptus;* florestas plantadas*; machine learning*; máquinas florestais autopropelidas*;* operações florestais*.*

ABSTRACT

Machine learning models involving regression algorithms can be used by forest managers to optimize the processes they undertake and provide support for decision-making. The excellent predictive performance of machine learning models can contribute to efficient supply chain management, waste reduction, and the optimization of time, machinery, and equipment. In the mechanized wood harvesting operation, the incorporation of predictive tools, such as machine learning, assists in strategic, tactical, and operational planning, reducing complexity and leveraging data diversity. Just as planning can be associated with provisions, we investigate whether the productivity of self-propelled forestry machines can be achieved through the implementation of machine learning algorithms in the *cut-to-length* system. Data from the cutting and wood processing activities in planted *Pinus* and *Eucalyptus* forests in Brazil were evaluated, carried out by 21 *harvesters* and 23 *forwarders* in *cut-to-length* systems. Blending and stacking learning methods were tested, incorporating the permutation of 18 and 22 algorithms in predicting the productivity of *harvesters* and *forwarders*, respectively. Learning methods and algorithms were compared based on performance metrics. Even with the compromise of the database quality, satisfactory performance metrics were obtained using default algorithms. When processing suitable data through machine learning algorithms, a coefficient of determination of 0.74 was achieved in estimating *harvester* productivity through the default extra trees regressor algorithm and a default light gradient boosting mode with a coefficient of determination of 0.77 in estimating *forwarder* productivity.

Keywords: data quality; *Eucalyptus*; forest operations; machine learning; planted forests; self-propelled forest machines.

LISTA DE ILUSTRAÇÕES

Capítulo 1 – Métodos de aprendizado de máquina na estimativa da produtividade de *Harvesters*: colheita mecanizada de madeira no Brasil

Figura 1 - Valores médios dos coeficientes de determinação dos modelos empregados..... 52

Figura 2 – Comparação dos valores do gráfico de dispersão dos modelos empregados..... 53

Figura 3 – Distribuição de densidade dos modelos empregados..... 53

Capítulo 2 – Planejamento tático de *forwarders*: uma abordagem baseada em dados do baldeio de madeira no Brasil

Figura 1 - Coeficiente de determinação dos 22 modelos ajustados em modo *default*..... 63

Figura 2 – Métricas quantitativas da performance do modelo eleito *Light Gradient Boosting*..... 65

LISTA DE TABELAS

Capítulo 1 – Métodos de aprendizado de máquina na estimativa da produtividade de *Harvesters*: colheita mecanizada de madeira no Brasil

Tabela 1 – Especificações técnica das máquinas florestais autopropelidas empregadas no corte e processamento das árvores..... 48

Tabela 2 – Procedimentos de *data wrangling* aplicado ao *dataset* proveniente da operação de colheita mecanizada de madeira..... 49

Tabela 3 – Média, desvio padrão e mediana do *dataset* proveniente da operação de colheita mecanizada de madeira, após processo de remoção de *outliers* dos três atributos iniciais..... 50

Tabela 4 – Métrica de avaliação dos modelos baseados em *decision tree*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira..... 50

Tabela 5 – Métrica de avaliação dos modelos baseados em *gradient boosted machine*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira..... 50

Tabela 6 – Métrica de avaliação dos modelos baseados em *linear regression*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira..... 51

Tabela 7 – Métrica de avaliação de modelos baseados em *k-nearest neighbor, multi-layer perception regressor, random sample consensus, support vector regression, dummy regressor e passive-aggressive regressor* aplicado ao conjunto de treinamento da operação mecanizada de colheita de madeira..... 51

Tabela 8 – Desempenho obtido no conjunto de treinamento, em modo *default, stacking ensemble learning e blending ensemble learning*..... 52

Tabela 9 – Resultados obtidos no conjunto de dados de teste, em modo default, stacking ensemble learning e blending ensemble learning..... 52

Capítulo 2 – Planejamento tático de *forwarders*: uma abordagem baseada em dados do baldeio de madeira no Brasil

Tabela 1 - Perfil dos dados do *dataset* utilizado para ajustar os modelos de previsão de produtividade do *forwarder*..... 61
Tabela 2 – Estimativa dos erros gerados na predição a partir dos modelos em modo default pelo Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE)..... 64
Tabela 3 - Desempenho dos melhores modelos no modo default após ajuste do coeficiente de determinação do processo (R^2), Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE)..... 65
Tabela 4 - Estimativa dos erros gerados na predição a partir dos modelos em modo Voting pelo Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE)..... 55

LISTA DE ABREVIATURAS E SIGLAS

CTL	*Cut-to-length*
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MSE	Mean Squared Error
RMSLE	Root Mean Square Logarithmic
SMOTE	Synthetic Minority Over-sampling

SUMÁRIO

INTRODUÇÃO GERAL..... 23

REVISÃO DE LITERATURA..... 25

Florestas Plantadas..... 25

Colheita mecanizada de madeira..... 27

Sistema cut-to-length..... 28

Gerenciamento florestal..... 29

Machine learning Aplicada em Atividades Florestais..... 30

Data mining..... 32

Machine learning..... 33

Aprendizado supervisionado..... 34

Aprendizado não supervisionado..... 36

Aprendizado semi-supervisionado..... 37

Aprendizado por reforço..... 37

Métodos de aprendizado em conjunto..... 38

Árvore de decisão..... 39

Random forest..... 40

Gradient boosting..... 41

Regressão linear..... 42

K-nearest neighbor..... 42

Support vector machine..... 43

Rede neural artificial..... 44

Capítulo 1 - Métodos de aprendizado de máquina na estimativa da produtividade de *harvesters*: colheita mecanizada de madeira no Brasil..... 46

1.1 Introdução..... 46

1.2 Material e Métodos..... 47

1.2.1 *Dataset*..... 47

1.2.2 *Diferentes métodos de aprendizado e abordagens de algoritmos*..... 48

1.3 Resultados..... 49

1.3.1 Qualidade do *dataset*..... 49

1.3.2 Diferentes métodos de aprendizado e abordagens de algoritmos..... 50

1.4 Discussão..... 54

1.5 Conclusões..... 55
Referências..... 55
Capítulo 2 - Planejamento tático de *forwarders*: uma abordagem baseada em dados do baldeio de madeira no Brasil..... 59
2.1 Introdução..... 59
2.2 Material e Métodos..... 60
2.2.1 *Origem dos dados.*..... 60
2.2.2 *Variáveis dependentes e target no conjunto de dados.*..... 61
2.2.3 Modelagem da produtividade do *forwarder*..... 62
2.2.4 *Escolhendo o melhor modelo*..... 63
2.3 Resultados..... 63
2.4 Discussão..... 66
2.5 Conclusões..... 67
Referências..... 67
CONSIDERAÇÕES FINAIS..... 71
REFERÊNCIAS..... 73

INTRODUÇÃO GERAL

Florestas plantadas com espécies de rápido crescimento e variedade comercial, como *Eucalyptus* e *Pinus*, atuam como fonte de matéria-prima para diversos setores industriais, incluindo a fabricação de painéis de madeira. As características específicas da madeira, como densidade, resistência mecânica e composição química, são determinantes para sua adequação a diferentes aplicações industriais.

Além disso, condições edafoclimáticas favoráveis, aliadas ao uso de tecnologias avançadas em colheita mecanizada, não apenas asseguram a produção de madeira com qualidade, mas também impulsionam a produtividade dessas florestas plantadas. A colheita mecanizada da madeira, etapa essencial da cadeia de florestas plantadas produtivas, engloba o corte de árvores, o transporte interno na área de manejo e a remoção da madeira para processamento industrial.

A integração de tecnologias avançadas no planejamento da colheita mecanizada da madeira otimiza a precisão das operações, reduz os impactos ambientais e promove maior segurança para os trabalhadores envolvidos. Quando mecanizada, normalmente o sistema *cut-to-length* (CTL) é composto por duas máquinas florestais autopropelidas, sendo elas, o *harvester* e o *forwarder*. O *harvester* realiza o corte, o desgalhamento e seccionamento das toras em medidas específicas diretamente no campo. Já o *forwarder* é responsável por transportar os troncos processados até o pátio intermediários.

Não obstante de outros setores, na colheita mecanizada de madeira, a disponibilização e o acesso às informações é uma realidade. Com a incorporação de técnicas de *data wrangling*, isto é, preparação dos dados, associadas a técnica de *machine learning*, implementa-se a extração de conhecimento destas informações. Assim, gestores, dos diversos segmentos, sobretudo do setor florestal, são inseridos em uma nova configuração, que os atribui autonomia e segurança no exercício de empreender às tomadas de decisões.

Deste modo, gestores florestais, mais especificamente os que atuam no gerenciamento de sistemas de colheita de madeira em florestas plantadas, acessam cluster específicos de informações. Exemplo disto, relatórios de inventário florestal, interfaces tecnológicas das máquinas autopropelidas empregadas na operação,

sensores adaptados aos sistemas, satélites de sensoriamento remoto, ou ainda, na concatenação de diversos métodos de entrada de dados.

Por demandar aporte financeiro, a exequibilidade das estratégias de gerenciamento florestal empregadas aos sistemas de colheita de madeira demanda a construção de modelos preditivos. Estes modelos, propiciam à prévia classificação das condições aos quais estarão expostos ao adotar, por exemplo, um tipo de sistema de colheita. As consequências destas escolhas, atribuirão, inclusive, a magnitude da produtividade a partir da adoção de um determinado sistema.

Assim, gestores florestais podem aliar o *data wrangling* com algoritmos de *machine learning* e estabelecer modelos de predição com as informações das distintas bases as quais possuem acesso. Se unificadas, podem aumentar a acurácia de suas decisões, sobretudo, no gerenciamento da colheita mecanizada de madeira que permitam otimizar a alocação de recursos, reduzir custos e aumentar a eficiência do processo produtivo.

Como o planejamento pode ser associado a provisões, o objetivo geral deste estudo foi verificar se a produtividade de máquinas florestais autopropelidas pode ser predita por meio de algoritmos de *machine learning* no sistema de colheita mecanizado de madeira *cut-to-length*.

No primeiro capítulo, desenvolvendo e validando modelos de machine learning capazes de estimar a produtividade de *harvesters* em função de variáveis operacionais, ambientais e características da floresta. E no segundo capítulo, construindo e avaliando modelos de machine learning para prever a produtividade de *forwarders*, considerando fatores como distância de transporte, topografia do terreno e características das toras.

REVISÃO DE LITERATURA

Florestas Plantadas

Comparadas às florestas naturais, as florestas plantadas geralmente possuem menor diversidade de espécies e estruturas florestais mais simples. Além de atuarem na regulação de gases de efeito estufa, destacam-se na economia global por fornecerem matéria-prima para diversas indústrias, como a de papel e celulose, construção civil e bioenergia (Aerts; Honnay, 2011; Zemp *et al.*, 2019; Dymond; Giles-Hansen; Asante, 2020; Getachew *et al.*, 2024; Patil *et al.*, 2024; Xue *et al.*, 2024).

O potencial produtivo das florestas plantadas pode ser influenciado por diversos fatores, como solo, luz, altura dominante das árvores e a idade em um local específico. Por meio desses fatores, o planejamento e a previsibilidade da produtividade dessas florestas plantadas ajudam a estabilizar os mercados de produtos florestais (Lemos *et al.*, 2023; Munis *et al.*, 2023; Lv; Duan; Zhang, 2024; Stefanello *et al.*, 2024).

Em consonância com o benefício comercial, práticas de manejo sustentável auxiliam no aumento da produtividade das florestas plantadas sem comprometer o meio ambiente. Técnicas como rotação de culturas e manejo integrado de pragas são algumas das abordagens que podem ser adotadas para promover a sustentabilidade e produtividade dessas florestas (Forrester; Bauhus, 2016; Munis *et al.*, 2022a Negi *et al.*, 2024).

Neste contexto, políticas de incentivo oferecem oportunidades para o desenvolvimento futuro e a maximização dos benefícios de florestas plantadas. Associada à adaptabilidade e produtividade dessas florestas, condições agroecológicas específicas, faz com que gestores florestais optem pelo manejo, inclusive em áreas remotas e de difícil acesso (Brockerhoff *et al.*, 2013; Camargo *et al.*, 2022a; Welfle; Röder, 2022; Alemayehu; Melka; Eshete, 2024).

Comercialmente, em diferentes regiões do mundo, as florestas plantadas foram implementadas com sucesso, corroborando melhores práticas que podem ser aplicadas em outras áreas para alcançar resultados semelhantes. Exemplo disso são países como China, Índia, Brasil, África do Sul, Etiópia e outros, que tiveram o cultivo extensivo de florestas plantadas, onde, por exemplo, o *Eucalyptus* é o gênero de árvore mais plantado (Payn *et al.*, 2015; Daba, 2016; Camargo *et al.*, 2022b; Cheng *et al.*, 2024; Getachew *et al.*, 2024).

Em suma, as florestas plantadas desempenham um papel vital tanto para a economia global quanto para o meio ambiente. Embora apresentem menor diversidade de espécies e estruturas mais simples em comparação com as florestas naturais, sua capacidade de fornecer matéria-prima de forma previsível e eficiente é indispensável para a economia do setor florestal (Rollan *et al.*, 2018; Bhatt *et al.*, 2021; Munis *et al.*, 2022b; Seipp *et al.*, 2023).

Somada a uma variedade de híbridos, o gênero *Eucalyptus* inclui mais de 900 espécies, o que o torna amplamente dominante em várias partes do mundo, apesar de sua origem australiana. Sua cadeia de valor foi fundamental para garantir o desenvolvimento de florestas plantadas em rápida evolução. Com início em viveiros especializados na produção de mudas de *Eucalyptus*, sequencia-se sua cadeia de valor com o cultivo e manejo da floresta plantada. Quando atingem a maturidade adequada para a colheita, as árvores são cortadas e preparadas para transporte e, conseguinte, passam pelo processamento industrial (Oikari *et al.*, 2010; Zubizarreta-Gerendiain; Pukkala; Peltola, 2016; Cataldo *et al.*, 2020; Florêncio; Martins; Fagundes, 2022; Lopes *et al.*, 2022; MIchelon *et al.*, 2024).

Podendo ser transformada em uma variedade de produto, alimentam redes de distribuição estabelecidas. Em atendimento à agenda *Environmental Social Governance* (ESG), parte da cadeia de valor também inclui a reciclagem e reutilização de resíduos de *Eucalyptus*, como cascas e serragem, que podem ser utilizados para produção de energia, compostagem ou outros usos industriais. Em todo o mundo, as plantações de *Eucalyptus* são consideradas significativas, pois fornecem matérias-primas valiosas para as indústrias madeireiras. A gestão racional das florestas plantadas de *Eucalyptus* é, portanto, imperativa para garantir a produção sustentada e o fornecimento de matérias-primas ao negócio (Gumus; Acar; Toksoy, 2008; Adewuyi; Olofin, 2015; Eker, 2020; Holm *et al.*, 2020; Iglesias *et al.*, 2021; Chiang; Castro; Molina, 2023; Chu *et al.*, 2023; Bellink; Verburg, 2023).

Além de sua contribuição econômica e ambiental, as plantações de *Eucalyptus* também desempenham um importante papel na promoção do desenvolvimento socioeconômico local em muitas regiões do mundo. A criação de empregos diretos e indiretos nas etapas de cultivo, colheita e processamento, proporcionam oportunidades de trabalho e impulsionam a economia local (Nabuurs *et al.*, 2014; Ribeiro; Azevedo-Ramos; Santos, 2020; Parajuli *et al.*, 2022; Temu *et al.*, 2022; Xu; Cai; Zhao, 2024).

Colheita mecanizada de madeira

A seleção do sistema de colheita nas operações florestais é uma componente integral do manejo florestal aplicado. Em operações florestais, o uso de máquinas florestais autopropelidas para o corte e baldeio de madeira é prática comum em todo o mundo. Assim, o desempenho de uma equipe de colheita em termos de sua capacidade de transformar insumos em produtos é influenciado por fatores discricionários sob o controle da unidade, como a seleção de máquinas e operadores (Becker *et al.*, 2018; Cudzik *et al.*, 2018; Norihiro *et al.*, 2018; Obi; Visser, 2018; Ceccherini *et al.*, 2020).

Com os avanços tecnológicos, máquinas florestais autopropelidas para corte e baldeio da madeira passaram a ser empregadas. Nesse sentido, especialmente o planejamento do baldeio de madeira envolvem esforços com múltiplos propósitos. Sendo assim, essas máquinas devem ser planejadas e dimensionadas, para as operações serem fisicamente e economicamente viáveis (Ilintsev *et al.*, 2018; Haas; Kirchner; Lang, 2020; Zhang; Barten, 2009; Ezzati; Najafi; Bettinger, 2016; Daugherty *et al.*, 2018; Jayasuriya; Germain; Wagner, 2020).

De acordo com Lopes *et al.* (2017), faz-se necessário conhecer as variáveis que exercem influência na operação de colheita mecanizada de madeira. Miyajima *et al.* (2020) corroboram a necessidade de otimização dos recursos aplicados nessa operação. Gestores florestais precisam entender os benefícios e as desvantagens de diferentes estratégias do manejo de sistemas de colheita mecanizada de madeira e, por conseguinte, permitir previsões eficientes de planejamento e produtividade (Marčeta *et al.*, 2020; Mederski *et al.*, 2021; Raulerson *et al.*, 2020).

Acompanhando avanços tecnológicos que permitem maior eficiência e adaptabilidade, a utilização de sistemas mecanizados na colheita de madeira proporcionou aumento da qualidade do produto, conforto e segurança aos operadores e redução dos custos de produção. Além disso, a automação e a digitalização desses sistemas têm facilitado o monitoramento em tempo real das operações, possibilitando uma gestão mais precisa e integrada dos processos florestais (Pereira *et al.*, 2015; Rodrigues *et al.*, 2019; Silva *et al.*, 2011).

Sistema *cut-to-length*

O sistema *cut-to-length* (CTL) pode ser realizado por meio de máquinas florestais autopropelidas, ou seja, um sistema mecanizado especializado para o corte e processamento da madeira ainda no interior dos talhões. Esse sistema é amplamente utilizado em operações florestais devido à sua eficiência e menor impacto ambiental. Portanto, quando mecanizado, as máquinas florestais autopropelidas utilizadas no sistema CTL são o *harvester* e *forwarder* (Chen *et al.*, 2017; Eriksson; Lindroos, 2014; Laina *et al.*, 2013; Prinz *et al.*, 2018).

O *harvester* é uma máquina multifuncional que derruba, desgalha, descasca e corta a árvore em troncos de comprimentos predeterminados, isto é, realiza o processamento completo da árvore. Já o *forwarder* é responsável pelo baldeio das toras, que foram seccionadas pelo *harvester*, para o local de carregamento, onde a madeira é preparada para o transporte ou em um pátio de armazenamento temporário (Almeida *et al.*, 2024; Polowy; Rutkowski, 2024; Wester; Eliasson, 2003).

De acordo com Bačić *et al*. (2024) diferentes autores abordam o sistema CTL sobre diversos aspectos, desde a eficiência energética e ambiental até a segurança e os impactos na regeneração florestal e biodiversidade. Os benefícios da escolha do CTL na realização da colheita mecanizada da madeira, destacando que, ao realizar o processamento da madeira no próprio talhão, reduz o número de máquinas necessárias nas operações e, consequentemente, diminui o tráfego pesado e a compactação do solo (Palander *et al.*, 2024; Sagar *et al.*, 2024; Sirén; Aaltio, 2003).

Outro aspecto importante é a flexibilidade operacional que o sistema CTL proporciona. O *harvester* com o *forwarder* podem ser configurados para operar em uma ampla gama de terrenos, desde planícies até áreas com relevo acidentado (Garren *et al.*, 2019; Kormanek *et al.*, 2023). Apesar de aumentar a precisão das operações, reduzindo o desperdício de madeira e maximizando a produtividade por hora trabalhada, o sistema CTL ainda promove melhorias significativas na segurança dos trabalhadores florestais. As cabines dos *harvesters* e *forwarders* são projetadas para oferecer maior proteção e conforto, possibilitando um ambiente de trabalho mais seguro e ergonômico (Camargo *et al.*, 2021; Gerasimov; Sokolov, 2014; Synwoldt; Gellerstedt, 2003).

A produtividade do *harvester* pode ser influenciada por diversos fatores relacionados à floresta plantada e ao terreno. A idade da floresta plantada

desempenha um papel importante, florestas plantadas com ciclos mais longos tendem a ter árvores com troncos maiores, aumentando a eficiência do corte e processamento. O índice de sítio, que reflete a qualidade do solo e as condições de crescimento, também impacta a produtividade, já que locais com melhor crescimento resultam em árvores mais robustas (Talbot *et al.*, 2003; Leite *et al*., 2014; Fernandes *et al*., 2013).

Além disso, a densidade da floresta e a espécie das árvores podem afetar a facilidade de operação, sendo que florestas muito densas ou espécies *hardwoods* exigem maior esforço. O tipo e a declividade do terreno são outro fator que impacta nas operações do *harvester*, áreas mais acidentadas dificultam sua movimentação e manuseio (Kärhä *et al*., 2004; Holzleitner *et al*., 2019).

A declividade do terreno também é um dos elementos que impactam a produtividade do *forwarder*, pois terrenos inclinados podem reduzir a velocidade de deslocamento e aumentar o esforço necessário. Além disso, a distância entre os locais de carregamento e descarregamento exercem um impacto negativo, uma vez que distâncias mais longas diminuem a produtividade (Cadei *et al*., 2020; Marcu; Borz, 2024).

Outro fator que impacta a produtividade do *forwarder*, é a capacidade de considerada determinante, uma vez que modelos com maior capacidade podem transportar mais madeira por viagem, melhorando o desempenho geral. Também deve-se considerar as condições do solo, como umidade excessiva, podem reduzir a tração e dificultar o deslocamento (Väätäinen *et al.*, 2021; West *et al.*, 2022; Bumber *et al*., 2023).

Gerenciamento florestal

O gerenciamento florestal consiste no planejamento da implantação e manutenção das florestas plantadas, envolvendo a operação de colheita, o abastecimento contínuo das indústrias, o cumprimento de legislações ambientais e trabalhistas e tudo o que envolve a cadeia de suprimentos. O gerenciamento inclui, impreterivelmente, implementar e executar normas de saúde e segurança do trabalho na gestão de todos os colaboradores que fazem parte dos processos e operações (Camargo; Munis; Simões, 2021; Alayet *et al.*, 2016 Bligård; Häggström, 2019).

Conforme a complexidade das decisões, torna-se mais difícil para os tomadores de decisão identificar uma alternativa de gerenciamento que maximize todos os critérios de decisão. Portanto, o gerenciamento florestal deve ser pautado em medidas técnicas que examinem a natureza dos problemas abordados e a incorporação do risco no manejo florestal e na tomada de decisões (Pukkala; Kangas, 1993; Ananda; Herath, 2009).

As informações das florestas servem de base para o planejamento florestal, permitindo o uso de modelos de crescimento e rendimento para prever condições futuras com possibilidade de diferentes opções de gerenciamento. Dados coletados em inventário florestal, avaliações das medições em nível de talhão, manejo florestal auxiliam e viabilizam a gestão (Mcroberts; Tomppo, 2007; Liang *et al.*, 2016; Eyvindson; Saad; Eriksson, 2017).

Pesquisadores e profissionais estão utilizando ferramentas tecnológicas que auxiliam nas tomadas de decisões para tornar a gestão florestal mais eficiente. Com base em indicadores de qualidade, critérios de avaliação, métodos de diagnóstico para a medição, técnicas de análise de risco aprimoram as estruturas que apoiam os tomadores de decisão em seus esforços (Mendoza; Prabhu, 2000; Hunt; Haider, 2001; Pollard *et al.*, 2010; Bai; Sarkis, 2019; Welch *et al.*, 2019).

Machine learning Aplicada em Atividades Florestais

O gerenciamento correto das informações obtidas no cumprimento das atividades florestais, auxilia na formação de estratégias. Associar essas informações com análises econômicas podem ajudar os proprietários de terras a tomar decisões informadas. A economia dos sistemas de recursos florestais deve ser estudada ponderando os custos totais das cadeias de suprimento, sobretudo, os custos a colheita da madeira (Wagner, 2012; Karttunen; Laitila, 2015; Hitchner *et al.*, 2021).

A gestão deve procurar garantir que as várias perspectivas possíveis sejam consideradas ainda no início do processo de tomada de decisão estratégica. Contudo, quando os gestores aliam o conhecimento com o acesso a fontes confiáveis de informações, se beneficiam de decisões mais informadas e consistentes (Smith *et al.*, 2008; Bozhesku, 2011; Osuszek; Stanek; Twardowski, 2016).

A abordagem da técnica de *machine learning* é fundamental para a tomada de decisão e controle das interferências humanas em ecossistemas naturais,

sobretudo, em florestas. A inovação e a tecnologia se tornaram a tendência do desenvolvimento global. Técnicas de *machine learning* são amplamente recomendadas para aplicações ecológicas, dada suas precisões na classificação e acurácia na detecção de alterações dinâmicas (Elith; Leathwick; Hastie, 2008; Yu *et al.*, 2020; Shabani; Pourghasemi; Blaschke, 2020; Gao; Su, 2020; Švač; Cagáňová, 2020).

Pourghasemi *et al.* (2020), por exemplo, utilizaram-se da abordagem de *machine learning* para mapear os riscos ambientais, como incêndios florestais. Do mesmo modo, Hararuk, Kurz e Didion (2020) combinaram observações de inventário florestal com técnicas de *machine learning* para quantificar as taxas de decaimento de árvores abatidas em florestas. Logan *et al.* (2020) utilizaram-se da abordagem para examinar a influência das condições micro ambientais nas características das plantas lenhosas em camadas de regeneração.

Zhou *et al.* (2020b) verificaram que métodos de aprendizagem de máquinas são eficazes no monitoramento da dinâmica de florestas plantadas. Soma-se a capacidade de algoritmos de *machine learning* em explicar fatores espaciais da composição das espécies arbóreas de dossel em função da cobertura florestal. A fusão de dados do sistema de comunicação que conecta os diferentes componentes de uma máquina florestal, como sensores, atuadores e unidades de controle eletrônico com as informações de inventário florestal permite que métodos de *machine learning* descubram relações entre a máquina florestal e variáveis do ambiente captadas pelos sensores como localização via GPS e inclinação do terreno (Melander; Einola; Ritala, 2020; Balzotti *et al.*, 2020).

Métodos de *machine learning* foram utilizados para construir um modelo capaz de prever a classe de pedregosidade de um determinado local, a fim de viabilizar a operação de colheita mecanizada de madeira (Melander; Ritala; Strandström, 2019). Algoritmos de *machine learning* fornecem suporte no monitoramento da qualidade da madeira no momento da colheita mecanizada de madeira e na estimaram as probabilidades de ocorrência de queimada florestal por meio de classificação de imagens (Palander *et al.*, 2018; Hawbaker *et al.* 2020).

Além disso, métodos de machine learning têm sido aplicados para otimizar a eficiência operacional das máquinas de colheita mecanizada de madeira. Por exemplo, Almeida *et al*. (2024) desenvolveram um modelo preditivo que utiliza dados

de sensores instalados em *harvesters* para criar um modelo de alto desempenho que preveja com precisão sua produtividade em florestas plantadas de *Eucalyptus*.

Data mining

Com o avanço da tecnologia da informação, grandes quantidades de dados estão disponibilizadas para o acesso de pesquisadores e gestores. A técnica de mineração de dados, da língua inglesa *data mining*, foi reconhecida como uma das ferramentas mais promissoras para extração de informação de conjuntos de dados. O *data mining* pode ser utilizado para extrair as informações mais relevantes em um conjunto de dados, bem como, para aumentar a precisão e a qualidade dos resultados na estimativa de parâmetros críticos. A utilização de algoritmos de *data mining* fornece poder preditivo (Kim *et al.*, 2018; Xu *et al.*, 2020; Zhao *et al.*, 2020; Khosravi *et al.*, 2020; Schuh; Prote; Hünnekes, 2020; Zonouzi; Kargari, 2020).

Os algoritmos de *data mining*, extraem informações quantitativas e qualitativas, de modo que os fatores de influência possam ser reconhecidos e categorizados. Por meio de técnicas de *data mining* é possível ainda, investigar associações entre variáveis e sustentar hipóteses pautadas nos conjuntos de dados. Portanto, as técnicas de *data mining* são amplamente utilizadas para reduzir a dimensionalidade de problemas (Hallberg; Salimi, 2020; Ren; Zhang; Liu, 2020; Cho *et al.*, 2020; Maia; Plastino; Penna, 2020).

Os conhecimentos gerados, decorrentes da aplicação da mineração dos dados, desempenham um papel fundamental na tomada de decisões inteligente. Relaciona-se ainda, na melhoria do desempenho de sistemas e no gerenciamento de recursos e serviços. A introdução de modernos sistemas e tecnologias de informação nos processos de negócios torna possível a utilização de grandes quantidades de dados (Hunt, 2021; Shaikh *et al.*, 2021; Mitroshin *et al.*, 2022; Huang, 2020; Sunhare; Chowdhary; Chattopadhyay, 2020; Li *et al.*, 2021; Fonseca-Galindo *et al.*, 2022).

Nesse contexto, a mineração de dados promove a análise de comportamento de um espectro em massa com tendências ou características comuns. Assim, um banco de dados associado a programas de mineração de dados resulta em dados orientados à uma variável resposta conhecida. Além disso, auxilia na análise de padrões e características desconhecidas, permitindo que as empresas se informem e tomem decisões (Kamath *et al.*, 2021; He *et al.*, 2022; Mao *et al.*, 2022).

As rotinas de integração e gerenciamento para a análise de dados auxiliam no diagnóstico de problemas de qualidade de dados, além de suportar hipóteses. O *data wrangling* caracteriza-se pela transformação e preparação dos dados por meio da identificação, extração, filtragem e integração dos conjuntos de dados adequadas ao tipo de análise (Kandel *et al.*, 2011; Endel; Piringer, 2015; Furche *et al.*, 2016; Bellomarini *et al.*, 2022).

A mineração de dados demonstra ser um método confiável para analisar e descobrir conhecimento útil. Além disto, algumas ferramentas de mineração de dados possuem recursos de *data wrangling* que automatizam o processo de filtragem de dados. A transformação de dados permite processamento rápido de consultas de negócios com soluções certas para analistas e usuários consumidores (Santos-Pereira; Gruenwald; Bernardino, 2021; Shakorshahabi *et al.*, 2021; Costa *et al.*, 2022; Konstantinou *et al.*, 2017; Patil; Hiremath, 2018).

Os processos de *data wrangling* precisam garantir que as etapas de transformação de dados se beneficiem da automação sempre que possível. A técnica de *data wrangling* permite ao usuário explorar o fluxo de dados e distribuição de problemas em todo o conjunto de dados para determinar padrões de erro. Destaca-se ainda, como técnica que propicia o desenvolvimento da qualidade ao longo do tempo em uma sequência de operações (Koehler *et al.*, 2017; Bors; Gschwandtner; Miksch, 2019; Bellomarini *et al.*, 2022; Yang *et al.*, 2022).

Machine learning

A obtenção de dados de observações, medições ou leituras de instrumentos, podem ser quantificados de forma a atuarem na classificação ou predição de dados, por meio do aprendizado de máquina, da língua inglesa *machine learning* (Liu *et al.*, 2020a; Xu *et al.*, 2020). A combinação da técnica de *machine learning* (ML) com técnicas de *data mining* podem melhorar a capacidade de mensuração de riscos. Em algumas aplicações, aprimora os parâmetros de interesse como minimizadores de uma função de risco, separados de modelos estatísticos (Wang; Martin, 2020; Zhang, 2020).

Essa seleção é fundamental para facilitar a interpretabilidade dos dados de modo a revelar relacionamentos em dados estruturados e fornecer melhor poder preditivo ao modelo. Além de atuar na modelagem preditiva, a técnica de *machine*

learning fornece subsídios quantitativos ou qualitativos na detecção de falhas operacionais ou fraudes. A disponibilidade de recursos de dados ajuda as organizações a extraírem informações úteis e a tomar decisões mais bem fundamentadas (Ta, 2019; Bozdağ; Dokuz; Gökçek, 2020; Wei; Yang; Sun, 2020; Won; Manzour; Chaovalitwongse, 2020).

O desenvolvimento de um modelo para predizer o comportamento dinâmico, sob diferentes configurações, facilita a otimização de processos. A técnica de ML, quando bem ajustada, pode fornecer resultados validados, quando se tem um conjunto de dados de validação, e precisos para expandir o conhecimento das características que compõem os processos. Salienta-se que existem vários parâmetros dos algoritmos de treinamento que precisam ser selecionados para os diferentes algoritmos de ML (Kudyba, 2018; Jaegher *et al.*, 2020; Liu *et al.*, 2020b; Escrig *et al.*, 2020; Mccarthy *et al.*, 2020; Shamsara; Schüürmann, 2020).

Quando há requerimento de alta precisão em análises, técnicas de *machine learning* são amplamente utilizadas para criar métodos de análise que identificam padrões ou tendências. Nesses casos, técnicas de ML são empregadas na resolução de funções críticas envolvendo *data mining*, mineração de texto, análise preditiva e sistemas de tomada de decisão. A utilização da técnica de *machine learning* pode dar-se por aprendizado supervisionado e não supervisionado. Contudo, outros tipos de algoritmos de *machine learning* são obtidos para aprendizados semi-supervisionado e por reforço (Chowdhury *et al.*, 2020; Ma *et al.*, 2020; Orozco *et al.*, 2020; Hwang; Jun, 2014; Camastra; Vinciarelli, 2015; Wang *et al.*, 2015; Dornheim; Link; Gumbsch, 2020; Luo *et al.*, 2020).

Aprendizado supervisionado

Neste tipo de aprendizado, o algoritmo de treinamento supervisionado, aprende com base em dados de limites de decisão entre infinitas categorias (ou classes). Supõe-se que um classificador supervisionado possa anotar um conjunto de dados quando um especialista fornece amostras rotuladas de todas as classes, conseguinte, validação (Fu; Kara, 2011; Amorim *et al.*, 2019; Hussain *et al.*, 2020).

O aprendizado supervisionado só é possível de ser empregado em amostras rotuladas (*labeled sample*). Amostras rotuladas caracterizam-se por conter rótulos ou classes, isto é, informações importantes sobre a distribuição de dados, o que é

essencial para produzir um classificador ajustado. Em uma estrutura de aprendizado supervisionado, as amostras rotuladas são dispostas aleatoriamente no conjunto de treinamento, a fim de atualizar os parâmetros (Liu; Xiang, 2017; Aydemir; Bilgin, 2020; Shi *et al.*, 2020; Tian *et al.*, 2020; Chu; Torre; Cohn, 2019; Frank *et al.*, 2020; Wei *et al.*, 2020).

Como o aprendizado supervisionado é baseado em grandes quantidades de dados rotulados, isto requer tempo e recursos extensivos. A coleta de amostras rotuladas, por vezes, pode ser cara e onerosa, limitando o espaço amostral de treinamento e apresentando um desafio para a classificação supervisionada (He; Hu; Hang, 2018; Rezagholiradeh; Haidar, 2018; Zhang; Yu, 2020).

No aprendizado supervisionado, várias métricas de classe local e uma métrica global são treinadas em conjunto. É um método de treinamento no qual o número de dados de entrada e o número de rótulos devem ser iguais. Ressalta-se que se deve explorar o máximo possível de informações de dados para melhorar a precisão do aprendizado da distribuição de rótulos (Jia *et al.*, 2021; Shim *et al.*, 2022; Sun; Yang, 2022).

Dentre as diversas abordagens para a análise de conteúdo automatizada, os métodos de aprendizado supervisionado são utilizados. Eles podem ser empregados para classificar eventos automaticamente com base em um sistema de classificação pré-existente. Este tipo de algoritmo utiliza de informações explícitas de categorização. Portanto, nos casos de aprendizado supervisionado, exemplos são mostrados e testados para que um diagnóstico da verdade básica seja reconhecido ou não (Robins, 1992; Collingwood; Wilkerson, 2012; Wey; Kosorok, 2013; Dun; Soroka; Wlezien, 2021).

Algoritmos supervisionados são mais sensíveis a composição dos dados e auxiliam na detecção de mudanças na distribuição das variáveis do processo. Eles caracterizam-se por conter um conjunto de amostras de treinamento rotuladas. A aplicação de algoritmos de aprendizado supervisionado para previsão promove a detecção precoce de problemas em contextos específicos. Além disto, o aprendizado supervisionado para classificação de dados, potencializa a discriminação de mudanças e a detecção de limites. Inclusive, por meio da abordagem de aprendizado supervisionado hierárquico, pode-se aumentar a probabilidade da detecção de anomalias (Li; Runger; Tuv, 2007; Vatsavai; Shekhar; Burk, 2007; Okori; Obua, 2011).

Contudo, para aplicação destes algoritmos, a falta de dados rotulados é um desafio crítico que deve ser superado para treinar modelos de aprendizagem supervisionada. A partir de *benchmarking* baseado em critérios de avaliação claramente definidos, o aprendizado supervisionado pode ser implementado com êxito. São exemplos de algoritmos por aprendizado supervisionado o *Random Forest*, *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-nearest neighbor* (KNN), *Perceptron* Multicamadas (MP) e regressão logística (Chen; Hsu; Su, 2017; Zhou; Li, 2017; Kumar *et al.*, 2018; (Yakimovich *et al.*, 2021; Alrfou *et al.*, 2022; Mehari; Strodthoff, 2022).

Aprendizado não supervisionado

Em um *machine learning* não supervisionado, o algoritmo consiste em um codificador automático, que obtém a distribuição de recursos desejada, alterando a função objetivo. Neste tipo de aprendizado, a identificação de *clusters* na presença de diferentes classes, é implementada. No aprendizado não supervisionado, os rótulos dos dados de treinamento são desconhecidos (Gao; Su, 2020; Mondal; Barman, 2020; Zhang *et al.*, 2020a).

Se os dados forem desconhecidos e o contexto não for facilmente entendido, técnicas não supervisionadas são empregadas. Dentre as vantagens do aprendizado não supervisionado está a viabilidade em descobrir o número natural de *clusters* no conjunto de dados. A definição de *clusters* viabiliza a obtenção de grupos derivados de um grande conjunto de dados (Bhopale; Tiwari, 2020; Morette *et al.*, 2020; Richhariya; Tanveer, 2020).

Essa divisão por *clusters* auxilia na expectativa de acesso rápido e preciso ao conhecimento estruturado. A amostragem não supervisionada permite a coleta de dados sob condições reais de operação, contrário de amostras geradas artificialmente a partir do aumento de dados. Essa técnica de clusterização auxilia na organização e categorização de grandes volumes de dados, permitindo um acesso mais rápido e preciso a informações relevantes (Parisi *et al.*, 2014; Li; Li; Ma, 2020).

Por meio de alguns algoritmos, é possível promover o agrupamento e a classificação dos dados. Embora o método de aprendizado não supervisionado permita maior acurácia na predição dos dados, face ao aprendizado supervisionado,

a convergência do modelo pode não ser alcançada e, conseguinte os resultados podem não ser precisos (Murray; Perera, 2020; Tseng *et al.*, 2020).

Aprendizado semi-supervisionado

O aprendizado semi-supervisionado consiste em refinar a classificação de um conjunto de dados. Este processo ocorre por meio da concatenação de dados não rotulados, facilmente acessíveis e com amplo volume de informação, com dados rotulados, porém em número limitado. A partir do aprendizado semi-supervisionado, um conjunto de testes pode ser construído, com dados rotulados, para avaliar o desempenho da classificação do algoritmo (Khan; Lee, 2019; Luo *et al.*, 2020; Roopa; Nagarajan, 2020).

Quando aplicado, o aprendizado semi-supervisionado, reduz a complexidade da validação de resultados no aprendizado não supervisionado. Conseguinte, o aprendizado semi-supervisionado pode ser realizado automaticamente, com monitoramento técnico de especialistas. O benefício principal é a melhoria da capacidade de generalização do modelo, aproveitando a abundância de dados não rotulados para refinar as previsões com um número reduzido de amostras rotuladas (Okaro *et al.*, 2019; Alipour; Harris, 2020; Almagro *et al.*, 2020; Morette *et al.*, 2020; Zhang *et al.*, 2020b).

No aprendizado semi-supervisionado, uma correlação entre os recursos pode ser alcançada na aplicação de amostras rotuladas e não-rotuladas. A interação humana durante o treinamento pode ser minimizada na estrutura de aprendizado semi-supervisionado e a eficiência do diagnóstico melhorada. Ao combinar dados rotulados e não-rotulados, os modelos podem alcançar melhor desempenho em tarefas de classificação e previsão, aproveitando a grande quantidade de dados não rotulados disponíveis (Li; Li; Ma, 2020; Sheikhpour *et al.*, 2020).

Aprendizado por reforço

Na técnica de ML por reforço um agente aprende por tentativa e erro atuando em ambientes dinâmicos. Os agentes têm a capacidade de interagir entre si de maneira distribuída e encontrar a melhor estratégia em ambientes competitivos. Com

a abordagem de aprendizado por reforço, o agente aprende uma política de comportamento baseado nas observações ao longo do tempo (Monteiro; Ribeiro, 2004; Glatt *et al.*, 2020; Samadi; Badri; Ebrahimpour, 2020).

Assim, cada ação executada pelo agente é associada a uma recompensa, a fim de obter uma política ótima. O modelo de aprendizagem por reforço aprende a executar ações ideais, para então ajustar a região de interesse inferida por meio do quadro anterior, conseguinte, atualização do modelo. Isto é válido, mesmo que a política ideal seja a que adéqua como melhor decisão ir contra os comportamentos de outros agentes (Soleymani; Paquet, 2020; Wang; Deng; Pan, 2020).

Com o agente de aprendizado por reforço adaptado à região de interesse, ocorre atuações, em tempo real, de forma inteligente. A aprendizagem por reforço, pertence a divisão de algoritmos de aprendizagem semi-supervisionados, porém acrescidos de uma exploração profunda e dinâmica das melhores ações, visando recompensas (Ghorbani; Afsharchi; Derhami, 2020; Kang *et al.*, 2020; Mhaisen *et al.*, 2020; Sun *et al.*, 2020; Wen; Chen; Li, 2020).

Métodos de aprendizado em conjunto

O método de aprendizado por *blending ensemble learning* é uma técnica de *machine learning* em conjunto que usa um meta-modelo para combinar os resultados preditivos, a partir de modelos de base heterogêneos dispostos em pelo menos uma camada. A variação do conjunto de combinação de predição é baseada em um conjunto de validação de dados, que foi utilizado para ajustar o meta-modelo (Chatzimparmpas *et al.*, 2021; Hansrajh; Adeliyi; Wing, 2021).

A *blending ensemble learning* é aplicada em modelo para aprendizado robusto, que podem prever várias tarefas ao mesmo tempo com menor parâmetro e melhorar a sensibilidade do modelo de reconhecimento, compartilhando informações entre diferentes tarefas. Essa técnica permite reconhecer e treinar algoritmo de segundo nível para combinar o modelo de predições de forma otimizada e formar um conjunto predições (Malik *et al.*, 2019; Fang *et al.*, 2020; Hao *et al.*, 2020).

A abordagem de *machine learning* reduz o erro e produz resultados mais precisos em comparação com as abordagens convencionais de modelos com base na média simplista ponderados igualmente ou de forma desigual. Os resultados da técnica de *machine learning* combinado com vários modelos permitem melhorar o

desempenho preditivo em comparação com um único modelo (Lu *et al.*, 2015; Chen *et al.*, 2020; Chen *et al.* 2020, Wu *et al.*, 2021).

Já o método de aprendizado por *stacking* ensemble *learning*, é um método de aprendizagem de conjunto integrado, o qual combina vários modelos de classificação ou regressão por meio de um meta-modelo. O modelo é treinado no conjunto de dados de treinamento completo e os resultados aplicados para o treinamento do novo modelo. O princípio fundamental é integrar as vantagens de diferentes algoritmos para obter um desempenho superior em comparação com modelos individuais (Divina *et al.*, 2018; Shen *et al.*, 2019).

O método é caracterizado por utilizar vários modelos para efetuar previsões combinando os valores previstos. A combinação dos valores pode ser obtida por meio da média dos resultados ou por modelos adicionais que agreguem à previsão final as previsões de todos os modelos individuais. A classificação com base neste método consiste em duas etapas em que a saída de um nível de classificador no primeiro estágio é usada como entrada para o segundo nível (Sun; Trevor, 2018; Lee; Kim; Ko, 2019).

Diante disso, o método de aprendizagem reduz o erro de generalização, aumentando a heterogeneidade dos dados básicos e utilizando diferentes algoritmos de aprendizado para combinar os resultados previstos pelo *database*. O *stacking* oferece a vantagem de integrar diferentes perspectivas e capacidades dos modelos individuais, resultando em um desempenho mais robusto e uma maior capacidade de generalização (Cui *et al.*, 2021; Li *et al.*, 2021; Shabankareh *et al.*, 2021).

Árvore de decisão

A árvore de decisão é um tipo de algoritmo de *machine learning* supervisionado, amplamente empregado na classificação de dados. Em sua estrutura, quando há situações de classificação, o operador separa valores pertencentes a diferentes classes, ou então, em caso de regressão, separa-os para reduzir o erro do parâmetro de critério selecionado (Ahmed *et al.*, 2020; Chowdhury *et al.*, 2020; Meenachi; Ramakrishnan, 2020).

Sendo um método não paramétrico de *data mining*, a árvore de decisão é capaz de lidar com relacionamentos não lineares entre contribuidores e a probabilidade de falha, além de possuir capacidade de detectar padrões em um

conjunto de *big data*. As árvores de decisão atuam como uma ferramenta de previsão, que divide um conjunto de dados em subconjuntos cada vez menores. Sendo assim, ao dividir os dados em subgrupos homogêneos, considerando suas propriedades, uma árvore é ajustada aos dados de calibração (Zhou *et al.*, 2020a; Faiz *et al.*, 2020; Jafari; Rajaee; Kisi, 2020; Wolff; O'donncha; Chen, 2020).

A seleção de um atributo para o posicionamento da árvore de decisão em uma posição apropriada, por exemplo, raiz da árvore, é uma decisão importante. Nesta partição da estrutura de uma árvore de decisão, a entropia de Shannon desempenha um papel fundamental, uma vez que por meio dela são feitas as seleções dos atributos (Shannon, 1948; Shaheen; Zafar; Khan, 2019; Singer; Anuar; Ben-Gal, 2020).

A árvore de decisão como classificador é usada para treinar o modelo com base no rótulo categórico e a árvore de decisão como regressor é usada para treinar o modelo com base em um rótulo não categórico. As árvores dividem recursivamente o espaço de partição nas regiões correspondentes às classes, escolhendo um recurso que forneça o maior ganho de informações (Friedman, 2001; Kushwah *et al.*, 2021).

A técnica de *machine learning*, como principal categoria de seu tipo, compõe diversos algoritmos, entretanto, modelos de árvore de decisão geralmente são preferíveis. Aliar informações mais precisas, como melhores previsões, a partir da combinação de algoritmos de árvores de decisão e otimização combinatória, promovem decisões mais bem informadas (Liu *et al.*, 2022; Ramos *et al.*, 2022; Gumuskaya *et al.,* 2021; He *et al.*, 2021; Cunha *et al.*, 2022).

A técnica de *machine learning*, incluindo algoritmos de árvore de decisão, geralmente apresenta alta confiabilidade e baixo custo computacional. Entretanto, toda árvore de decisão não é independente, porque uma nova árvore de decisão é adicionada, aumentando a ênfase nas amostras classificadas incorretamente. A dependência presente na estrutura das árvores de decisão, faz com que ela seja extremamente sensível às alterações nos dados de base, isto é, no efeito longitudinal (Čukić *et al.*, 2020; Liang *et al.*, 2020; Martin *et al.*, 2020).

Random forest

O algoritmo de *random forest*, consiste em uma técnica de *machine learning* em conjunto, uma vez que, geram muitos classificadores ou regressores, agregando seus resultados. O algoritmo de *random forest* é uma combinação de preditores de

árvores. Este tipo de algoritmo permite que a combinação de muitos classificadores únicos forneça uma representação mais diversificada dos dados (Breiman, 2001; Iranitalab; Khattak, 2020; Breiman, 1999; Maman *et al.*, 2020).

Random forest é um algoritmo, geralmente, eficiente e que reduz distorções e variações na classificação e regressão de dados. Por ser um método de aprendizado em conjunto, destinado a combinar vários resultados previstos de aprendizado, melhora a capacidade de generalização do aprendizado individual e a robustez (Ahmed *et al.*, 2020; Jaber *et al.*, 2020; Pei *et al.*, 2020).

Destarte, cada classificador da árvore é criado por um vetor aleatório amostrado independentemente do vetor de entrada. Como este algoritmo é definido por uma sequência de regras que dividem os dados, ele reduz a variação da melhor maneira possível. O método de classificação por algoritmos de *random forest*, permite inferir a relevância dos fatores preditivos nas variáveis *target* (Baidai *et al.*, 2020; Marzouq *et al.*, 2020; Zhi; Yang; Fu, 2020).

Gradient Boosting

Desenvolvido por Friedman (2001), o *Gradient Boosting* ou regressão por *Gradient Boosting* é um algoritmo de aprendizado de conjunto ajustado por meio de uma série de cálculos iterativos. O algoritmo do *Gradient Boosting* foi motivado como um método de descida de gradiente no espaço funcional, capaz de ajustar modelos preditivos não paramétricos genéricos (Ma; Fildes, 2020; Peng *et al.*, 2020).

Este algoritmo de aprendizado desenvolve as árvores subsequentes, pautado nos resíduos da árvore anterior, sob ajuste de derivativos de primeira ordem. Os algoritmos árvore de decisão, *Random Forest* e *Gradient Boosting* são métodos lógicos baseados em regras que podem detectar relacionamentos não lineares e identificar a importância da variável, mesmo com valores ausentes (Rahman *et al.*, 2020; Truong; Choi, 2020).

Dentre os algoritmos de classificação mais utilizados está o *Gradient Boosting*, haja vista seu desempenho superior em tarefas que dependem da engenharia de recursos. Soma-se que o *Gradient Boosting* pode fornecer informações abundantes sobre os efeitos não lineares para variáveis independentes, desenhando os efeitos parciais. Contrário ao algoritmo de *Random Forest*, os modelos individuais no *Gradient Boosting* não são produzidos com base na seleção aleatória de dados,

mas sequencialmente, além de ser propenso a sobreajuste (Ville; Neville, 2013; Das; Khan; Ahmed, 2020; Gilbert; Mateu; Planes, 2020; Li *et al.*, 2020).

Regressão Linear

A regressão linear consiste em uma análise estatística para estimar a relação entre duas variáveis, ou seja, avalia valores de uma variável dependente a partir de uma variável independente. É uma técnica de modelagem em que uma variável dependente é prevista com base em uma ou mais variáveis independentes (Ludbrook, 2010; Uyanik; Güler, 2013; Kumari; Yadav, 2018).

Na regressão linear, os dados são modelados para se ajustar a uma linha reta ou ajustar duas variáveis para que um atributo possa ser usado para prever o outro. A regressão linear simples é utilizada para estimar os valores da função densidade e sua segunda derivada em um ponto. Deste modo, uma variável de resposta pode ser modelada como uma função linear de outra variável preditora (Hino; Koshijima; Murata, 2015; Ilic *et al.*, 2021; Loftus, 2022).

A regressão linear simples é uma suposição subentendida da população homogênea, que estima os parâmetros de regressão capturando as características da população representadas na amostra. No entanto, em alguns eventos, a suposição da amostra pode conter várias subpopulações, não sendo realista (Águila; Benítez-Parejo, 2011; Bashir; Carter, 2012).

A regressão linear múltipla é uma extensão da regressão linear, onde mais de dois atributos estão envolvidos e os dados são ajustados a uma superfície multidimensional. Esses coeficientes podem ser resolvidos pelo método dos mínimos quadrados, que minimiza o erro entre a linha real que separa os dados e a estimativa da linha (Han; Kamber; Pei, 2012; Olive, 2017; Bertsimas; Li, 2020; Sarker, 2021).

K-nearest neighbor

O método *K-nearest neighbor* é aplicado na classificação de objetos com base em dados de aprendizagem que estão mais próximos do objeto, a partir da comparação entre os dados anteriores e os dados atuais. Essa técnica é conhecida por ser robusta quando utilizada em várias aplicações de mineração de dados como

um modelo de classificação com grandes conjuntos de dados (Jiang *et al.*, 2007; Kramer, 2013; Lubis; Lubis; Al-Khowarizmi, 2020).

O método *K-nearest neighbor* é uma das primeiras escolhas para um estudo de classificação quando há pouco ou nenhum conhecimento prévio sobre a distribuição dos dados. Caracterizado como uma técnica de mineração de dados simples, mas eficaz. Parâmetros de controle adaptativo, incluindo coeficientes de aceleração variáveis e peso de inércia variável no tempo, são empregados (Ling; Mei, 2009; Triguero *et al.*, 2018).

Esses parâmetros auxiliam controlar, de forma eficiente, a capacidade de busca local e global do algoritmo. As análises *K-nearest neighbor* são não paramétricas, as quais podem ser desenvolvidas com variáveis de resposta categóricas para as classificações dos dados. Contudo, as técnicas podem ser empregadas a partir de variáveis de resposta contínuas para a previsão, e serem utilizadas para predição univariada e multivariada (Fayed; Atiya, 2009; Chen *et al.*, 2011, Mangalova; Agafonov, 2014; Ferreira *et al.*, 2015; Mcroberts, 2012; Xia *et al.*, 2015).

Quando existe dois ou mais grupos de dados no algoritmo classificado em distintas classes, pondera-se as distâncias entre os objetos. A classificação *K-nearest neighbor* pode ser desenvolvida a partir da representação de escolhas múltiplas de dois parâmetros-chave. Assim, a partir de um é aplicada a definição da função de pertinência e do outro, o cálculo da regra de votação (Derrac *et al.*, 2016; Hoang *et al.*, 2018).

Support vector machine

O s*upport vector machine* é um algoritmo de aprendizagem supervisionada, aplicável tanto a problemas de classificação como de regressão. Envolve técnicas estatísticas e de *machine learning* com o objetivo principal de previsão. Pode ser aplicado em dados contínuos, resultados binários e categóricos análogos a gaussianos, regressões logísticas e multinominais (Shmilovici, 2010; Guenther; Schonlau, 2016).

O *support vector machine* é uma ferramenta de modelagem e previsão de sucesso para uma variedade de aplicações com tratamento rigoroso dos resultados de última geração. Capaz de resolver problemas de classificação binários e ainda

pode ser aplicada em conjunto com técnicas para as soluções de problemas de análise de dados (Lorena; Carvalho, 2007; Steinwart; Christmann, 2008).

O support vector machine aprenderá com um conjunto de dados de treinamento e tentará generalizar e fazer previsões corretas sobre novos dados. Esta formulação visa combinar observações de gradiente com as observações de rótulos inteiros em problemas de classificação e valores reais na regressão dentro uma única estrutura de otimização (Mammone *et al.*, 2009; Campbell; Ying, 2011; Shukla; Billard, 2014).

A vantagem da abordagem do *support vector machine* é que soluções esparsas para problemas de classificação e regressão geralmente são obtidas em apenas algumas amostras. Estas amostras são determinantes na obtenção das funções de classificação ou regressão. Esse conceito de soluções esparsas torna o *support vector machine* especialmente eficiente, pois, ao usar apenas as amostras mais influentes, o modelo consegue evitar o sobreajuste e manter a complexidade do modelo relativamente baixa, sem sacrificar a precisão (Moguerza; Muñoz, 2006; Zhang, 2020).

Rede neural artificial

A rede neural artificial é uma técnica de *machine learning* desenvolvido a partir da ideia de simular o cérebro humano. Uma rede neural artificial pode ser criada simulando uma rede de modelos de neurônios em um computador. A rede neural artificial pode identificar e aprender padrões entre conjuntos de dados de entrada e valores de destino correspondentes (Krogh, 2008; Zou; Han; So, 2008; Sharma; Chopra, 2013).

Os modelos de rede neural são capazes de aprender qualquer mapeamento não linear complexo e aproximar qualquer função contínua. As redes neurais podem fornecer uma gama de novas técnicas poderosas para resolver problemas nas análises de dados, detecção de falhas, identificação de processos controle, previsão, modelagem, avaliação e diagnóstico de falhas em sistemas (Hakimpoor *et al.*, 2011; Pirdashti, *et al.*, 2013; ATA, 2015).

A classificação da rede é determinada pelas conexões que interconectam os neurônios. As conexões são ajustadas ou treinadas para atingir um comportamento geral desejado da rede e governada por seu algoritmo de aprendizagem. O aspecto

do projeto de rede neural depende da seleção de um método de treinamento adequado e eficiente a partir de muitos algoritmos de treinamento possíveis e de parâmetros correspondentes (Pham; Packianather; Afify, 2007; Pirdashti *et al.*, 2013; Kubat, 2021).

O uso de modelos de redes neurais artificiais para enriquecer as capacidades analíticas e preditivas dos sistemas de apoio à decisão tem aumentado. Em complemento, diversas métricas podem compor uma análise sistemática para validação do modelo de rede neural artificial obtido, por exemplo, considerando o erro quadrático médio e o coeficiente de determinação. Assim, o modelo de rede neural pode ser verificado e testado a partir de dados experimentais (Mumali, 2022; Peng *et al.*, 2022; Skrypnik *et al.*, 2022; Tam *et al.*, 2022).

CAPÍTULO 1

Métodos de aprendizado de máquina na estimativa da produtividade de *Harvesters*: colheita mecanizada de madeira no Brasil [1]

Resumo: A correta captação de informações das operações florestais realizadas em florestas plantadas pode auxiliar no manejo da colheita mecanizada de madeira. A gestão adequada deve ser capaz de dimensionar recursos e ferramentas necessárias ao cumprimento das operações e auxiliar no planejamento estratégico, tático e operacional. Com o objetivo de facilitar a tomada de decisão dos gestores florestais, este trabalho teve como objetivo analisar o desempenho de modelos de *machine learning* na estimativa da produtividade de *harvesters*. Como preditores de produtividade, utilizamos a disponibilidade de horas de uso da máquina, volumes médios individuais de árvores e declividades do terreno. O conjunto de dados foi composto por 144.973 registros, obtidos em um período de 28 meses. Testamos o desempenho preditivo de 24 algoritmos de *machine learning* no modo *default*. Além disso, testamos o desempenho de métodos de aprendizagem conjunta *blending* e *stacking*. Avaliamos o ajuste do modelo usando a raiz do erro quadrático médio, erro médio absoluto, erro percentual médio absoluto e coeficiente de determinação. Após a limpeza inicial do banco de dados, apenas 1,12% puderam ser utilizados para construir o modelo. A aprendizagem por blending ensemble destacou-se com coeficiente de determinação de 0,77 e erro percentual médio absoluto de 15%. A partir da utilização de dados de algoritmos de *machine learning*, tornou-se possível prever a produtividade dos *harvesters*. Testar uma variedade de modelos de *machine learning* com dinâmicas diferentes contribuiu para a técnica de *machine learning* que nos ajudou a atingir nosso objetivo: maximizar o desempenho do modelo por meio da realização de experimentação.

Palavras-chave: volume médio individual**;** *blending ensemble learning*; tomada de decisão; declividade do terreno; florestas plantadas; *stacking ensemble learning*.

1.1 Introdução

A gestão integra a rotina dos gestores florestais responsáveis por orientar e implementar operações madeireiras mecanizadas. A otimização do tempo e dos ativos biológicos capitalizados nas florestas plantadas, quando exaustos por *harvesters*, afeta o sucesso da operação. Assim, é necessário conhecer as variáveis que influenciam a colheita mecanizada de madeira, permitindo um planejamento mais eficaz.

Indicadores de qualidade, critérios de avaliação e técnicas de análise de risco melhoram as estruturas que apoiam os tomadores de decisão. Ao fazer isso, os dados coletados no inventário florestal, medições no nível do povoamento, manejo florestal operacional e computadores de bordo em máquinas florestais ajudam e permitem os procedimentos de gerenciamento para operações de colheita de madeira [1-6].

A manipulação e reutilização dessas informações promovem a sua utilização no próprio desenvolvimento da gestão e auxiliam na identificação de oportunidades que podem ser previstas. No entanto, o escopo do planejamento de uma operação de colheita mecanizada de madeira requer necessariamente uma base de dados quantitativa, robusta e confiável [7,8].

Diariamente, uma grande quantidade de dados é gerada a partir do registro das atividades das máquinas que compõem os diferentes sistemas de colheita. No Brasil, o sistema mecanizado *cut-to-length* (CTL), composto por um *harvester* e um *forwarder*, é normalmente utilizado para fornecer matéria-prima para indústrias de base florestal. Devido à utilização do *harvester* na configuração do sistema CTL, as árvores são cortadas e seccionadas em toras processadas. Além disso, a madeira sem resíduos de colheita, própria para processos industriais, é preparada para extração por meio de *forwarder* [9-16].

As fontes e estimativas comunicadas pelos operadores promovem o nível de qualidade do conjunto de dados. Ao fornecer alertas antecipados, o *machine learning* fornece subsídios qualitativos na detecção de associações entre variáveis, além de apoiar hipóteses. No entanto, os algoritmos de *machine learning* devem ser associados a *data wrangling* para destacar valores discrepantes, reduzir o tamanho e facilitar a filtragem do conjunto de dados [17-21].

[1] Capítulo publicado e redigido de acordo com as normas da revista Forests

Além do *data wrangling*, a preparação dos dados se adapta a diferentes análises por meio da identificação, extração, limpeza, transformação e integração de conjuntos de dados. A organização dos dados permite ao usuário explorar o fluxo de dados e a distribuição de problemas ao longo do conjunto de dados para detectar padrões de erro. Esta técnica propícia um desenvolvimento com qualidade na observação das operações ao longo do tempo [22-27].

Os gestores florestais, por meio de conjuntos de dados, são capazes de extrair informações úteis e tomar decisões mais bem informadas que incluem, por exemplo, estimar a produtividade de *harvesters*. Desenvolver modelos para previsão do comportamento dinâmico da produtividade de máquinas, sob diferentes configurações, pode contribuir para otimização da operação [28-31].

Associar o conhecimento gerado a partir da aplicação do *data wrangling* em um conjunto de dados com algoritmos de *machine learning* pode promover uma tomada de decisão inteligente. A união de técnicas de *machine learning* pode melhorar o desempenho do sistema *cut-to-length*, e a gestão de recursos florestais e serviços [32-35].

Alguns estudos investigaram e propuseram maneiras de melhorar essa associação, por exemplo, métodos de *machine learning* em conjuntos usam um meta-modelo para combinar resultados preditivos de modelos básicos organizados em pelo menos uma camada. Assim, a variação nas combinações de previsões é baseada em um conjunto de validação de dados [36-39].

Os métodos de aprendizagem por *blending ensemble learning* e *stacking ensemble learning* podem promover modelos mais robustos e precisos porque diminuem exponencialmente a taxa de erro da aprendizagem conjunta. O método *blending ensemble learning* permite reconhecer e treinar algoritmos de segundo nível para combinar modelos de previsão de maneira ideal, enquanto o *stacking ensemble learning* combina valores por média ou por modelos adicionais que se somam às previsões parciais ou finais de todos os modelos individuais [40-46].

Os modelos de *machine learning* são cada vez mais utilizados para prever atributos porque, por não estarem sujeitos às suposições estatísticas tradicionais, podem incorporar variáveis e encontrar padrões não lineares e complexos nos dados. O *machine learning* inclui uma variedade de algoritmos para aprender regras preditivas a partir de dados históricos e construir modelos que possam prever valores futuros [47-51].

O *machine learning* tornou-se uma ferramenta eficaz de análise preditiva para grandes volumes de dados, com aplicações na medicina, nas finanças e, recentemente, no setor florestal. Foi comprovado que a automação de processos florestais aumenta a produtividade e a qualidade do trabalho. No entanto, a operação mecanizada de colheita de madeira é desafiadora e complexa [52-55].

Tradicionalmente, a produtividade é estimada por meio do estudo de tempo, que consiste em analisar os elementos de trabalho que compõem o ciclo operacional do *harvester*, individualmente, associando o tempo necessário e o volume de madeira. No entanto, os requisitos de mão de obra e a quantidade mínima de amostra necessária torna-o caro para as indústrias de base florestal.

Diante deste contexto de interações complexas, justifica-se aos gestores florestais a utilização de ferramentas analíticas eficientes, como o *machine learning*, para auxiliá-los na tomada de decisão. Assim, verificamos se a produtividade de máquinas florestais autopropelidas pode ser estimada por meio de algoritmos de *machine learning* por diferentes métodos de aprendizado.

1.2 Material e Métodos

1.2.1 Dataset

Foram utilizados dados estruturados de produção e operação de colheita mecanizada de madeira, em florestas plantadas de *Eucalyptus* e *Pinus*, realizadas por sistemas *cut-to-length*. As florestas plantadas com *Eucalyptus* apresentaram espaçamento de 3,3 m × 1,8 m e idade média de 14 ± 9,87 anos. A floresta de *Pinus* apresentou espaçamento de 3,3 m x 1,8 m e idade média de 22 ± 9,09 anos. A madeira dessas florestas foi utilizada como matéria-prima para a produção de celulose e papel.

As condições meteorológicas médias na região de estudo, segundo o Instituto Nacional de Meteorologia [56], foram umidade relativa de 69,24%, velocidade do vento de 4,29 m s^{-1} e temperatura do ar de 289,3 K. As operações ocorreram no Brasil, em região com gradiente de declividade de 7,32% a 35,06%. Os intervalos foram categorizados em relevo suave (3% a 10%), moderado (10% a 32%) e íngreme (32% a 56%), de acordo com Speight [57].

Esta pesquisa foi baseada em dados empíricos e insumos silviculturais; portanto, os dados faziam parte dos registros diários coletados em campo pelos computadores de bordo dos *harvesters*. Apesar de considerarmos todos os registros na análise inicial, empregamos uma série de controles compensatórios a partir do *data wrangling*.

Nos dois turnos diários de 10 horas, a disponibilidade de máquinas, os volumes médios individuais das árvores e a declividade do terreno somaram 144.973 ocorrências ocorridas no período de 28 meses. As bases foram rotuladas, unidas e manipuladas usando a linguagem de programação R [58].

Os tempos reais gastos nas atividades foram registrados nos computadores de bordo dos *harvesters*. Dessa forma, estimamos a produtividade a partir da razão entre o volume de madeira extraído por um *harvester*, em metros cúbicos, e o tempo efetivo de operação, em segundos [59-60]. O mesmo operador poderia operar diferentes marcas e modelos de colheitadeiras de madeira; entretanto, esta variável não foi adicionada na construção do modelo, devido à dificuldade de rastreamento desses dados no banco de dados. Ao todo foram utilizados os registros operacionais de 21 *harvesters* (Tabela 1).

Tabela 1. Especificações técnica das máquinas florestais autopropelidas empregadas no corte e processamento das árvores.

Máquina florestal autopropelida			Quantidades de máquinas
Marca	Modelo	Potência (KW)	florestais autopropelidas
Harvester Ponsse	ERGO 8W	205	17
Harvester John Deere	1270E	170	2
Harvester Ponsse	BEAVER	150	1
Harvester John Deere	1270G	200	1

Por meio da linguagem de programação R, ao implementar rotinas de aprendizado de máquina para planejamento de gestão e detecção de qualidade de dados, consideramos, segundo Konstantinou e Paton [61], procedimentos de transformação, limpeza e fusão de diferentes fontes. Construímos uma rotina de data wrangling, na qual foram removidos os outliers e variáveis potencialmente correlacionadas.

As instâncias passaram pelo processo de data wrangling, que foi realizado para transformar e coletar adequadamente os dados adquiridos. Adicionalmente, por meio do intervalo interquartil, definido conceitualmente pelo intervalo de Tukey [62], retiramos os outliers e, por meio da correlação de Spearman, verificamos as correlações entre os atributos ($p < 0,05$).

Além disso, foi realizado o balanceamento de dados do SMOTE. O SMOTE foi adotado por ser um algoritmo de referência para resolver o problema de aprendizagem por desequilíbrio de classes [63]. O algoritmo SMOTE tem a dinâmica de gerar novos exemplos sintéticos na vizinhança de pequenos grupos de instâncias próximas, utilizando o k-vizinho mais próximo [64]. A função foi implementada a partir do pacote smotefamily.

1.2.2 Diferentes métodos de aprendizado e abordagens de algoritmos

Usando um único conjunto de dados, comparamos o desempenho preditivo de 24 algoritmos de machine learning para estimar a produtividade de *harvesters*. Esses algoritmos foram baseados em uma árvore de decisão, gradient boosting machine, regressão linear, k-nearest neighbors, support vector machine e artificial neural network.

Adotou-se o gradiente (5, 10, 15, 20 e 25) para validação cruzada, em que os hiperparâmetros foram otimizados automaticamente. Por fim, implementamos os métodos de aprendizagem por *stacking ensemble* e *blending ensemble*. Ordenamos o método *stacking ensemble learning* em uma estrutura de dados hierárquica. Em cada conjunto de dobras, aplicamos validação cruzada *k-fold*.

O desempenho preditivo dos modelos em relação aos dados não observados foi maximizado pelo coeficiente de determinação (R^2), minimizando o teste RMSE, que foi determinado a partir da média amostral aleatória gerada ($n = 80$). Implementamos a regressão pelo aprendizado supervisionado utilizando a biblioteca *PyCaret* da linguagem de programação Python [65] para automatizar o fluxo de trabalho de *machine learning* e o desenvolvimento de modelos. Das instâncias de dados, agrupamos 90%

(n = 1.466) como conjunto de treinamento e 10% (n = 163) como conjunto de teste. Testamos algoritmos de *machine learning* e os selecionamos de acordo com seu desempenho na previsão da produtividade de *harvesters*, usando métricas estatísticas universais para avaliar o desempenho dos modelos [66], como raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE), erro percentual médio absoluto (MAPE) e coeficiente de determinação (R^2).

Assim, submetemos os algoritmos a diferentes métodos de *machine learning*. Primeiramente, verificamos o desempenho dos algoritmos em modo *default*. Para melhorar o desempenho, ajustamos os hiperparâmetros dos algoritmos selecionados. Combinamos os dados validados e formamos o conjunto de meta-recursos, os dados de teste e o conjunto de destino. Novamente, combinamos os conjuntos usando novos conjuntos de meta-recursos, criando um novo conjunto de meta-treinamento. Os novos conjuntos de metas formaram um novo meta-teste. Geramos previsões finais por meta-aluno de nível um, a partir do treinamento com o conjunto de meta-treinamento.

O método de aprendizado combinado consistiu em combinar algoritmos de aprendizagem de máquina para minimizar as taxas de erro de previsão. Para isso, dividimos o conjunto de dados em treinamento e teste, além de implementar algoritmos de camada zero, que geraram conjuntos de validação e teste. Combinamos os respectivos conjuntos com novos conjuntos de meta-treinamento e meta-teste [67] e geramos previsões finais.

1.3 Resultados

1.3.1. Qualidade do dataset

A manipulação do conjunto de dados com registros diários da operação mecanizada de colheita de madeira resultou em uma amostra de 144.973 ocorrências. Porém, por ter sido consolidado a partir da unidade de diferentes fontes, incluindo notas manuais, a qualidade do conjunto de dados ficou parcialmente comprometida. Assim, removemos instâncias duplicadas e instâncias com informações ausentes.

Com uma rotina de *data wrangling*, além de filtrar e transformar os dados, realizamos um exame de qualidade dos dados, excluindo *outliers* e promovendo balanceamento. Vale ressaltar que, apesar dos *harvesters* possuírem computadores de bordo, o processo de registro dos dados ainda exigia interações manuais. Consequentemente, implementamos os modelos com algoritmos de *machine learning* em 1,12% do conjunto de dados (Tabela 2).

Tabela 2. Procedimentos de *data wrangling* aplicado ao *dataset* proveniente da operação de colheita mecanizada de madeira.

Procedimento de *data wrangling*	*Input*	*Output*
Instâncias duplicadas	144.973	144.781
Remoção de dados faltantes	144. 781	98.459
Erros de anotações	98.459	1.703
Remoção de *outliers*	1.703	1.629

Os atributos selecionados para a construção do modelo foram volumes médios individuais de árvores, declividade do terreno e disponibilidade de horas de uso da máquina. Mais detalhes sobre a média, desvio padrão e mediana do conjunto de dados da operação mecanizada de colheita de madeira para os atributos em estudo são apresentados na Tabela 3.

Tabela 3. Média, desvio padrão e mediana do *dataset* proveniente da operação de colheita mecanizada de madeira, após processo de remoção de *outliers* dos três atributos iniciais.

Atributo	Mínimo	Máximo	Média	Desvio padrão	Mediana	Assimetria	Curtose
Volume Médio Individual [m³]	0,11	0,62	0,35	0,09	0,36	0,09	2,83
Declividade [%]	2,86	35,06	12,37	6,46	11,40	1,62	6,26
Disponibilidade de Horas de Uso [h]	0,15	1,33	0,69	0,12	0,70	-0,55	5,19

1.3.2. Diferentes métodos de aprendizado e abordagens de algoritmos

Primeiro, analisamos o desempenho preditivo de 24 algoritmos, individualmente, com base em métricas de ajuste de modelo. Dos três algoritmos treinados, baseados na árvore de decisão, destacou-se o coeficiente de determinação das árvores extras, que foi 0,01 superior ao coeficiente de determinação da floresta aleatória e 0,22 superior ao da árvore de decisão (Tabela 4).

Tabela 4. Métrica de avaliação dos modelos baseados em *decision tree*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira.

Modelo	MAE	RMSE	R^2	MAPE	Time
Extra Trees Regressor	3,21	4,68	0,72	17	0,12
Random Forest Regressor	3,36	4,72	0,71	18	0,13
Decision Tree Regressor	4,26	6,22	0,50	22	0,01

Ao analisar os quatro algoritmos baseados em *gradient-boosted machines*, o melhor coeficiente de determinação foi obtido pelo *CatBoost Regressor*, que foi 0,04 maior que o *Gradient Boosting Regressor* e 0,20 maior que o *AdaBoost Regressor* (Tabela 5).

Tabela 5. Métrica de avaliação dos modelos baseados em *gradient boosted machine*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira.

Modelo	MAE	RMSE [$m^3\ h^{-1}$]	R^2	MAPE [%]	Time [s]
CatBoost Regressor	3,35	4,55	0,73	18	1,40
Gradient Boosting Regressor	3,68	4,91	0,69	20	0,04
AdaBoost Regressor	4,82	6,04	0,53	18	0,03

A disponibilidade do algoritmo baseado em regressão linear contribuiu para a aplicação dos 12 algoritmos treinados (Tabela 6).

Tabela 6. Métrica de avaliação dos modelos baseados em *linear regression*, aplicados ao *dataset* proveniente da operação de colheita mecanizada de madeira.

Modelo	MAE	RMSE [$m^3\ h^{-1}$]	R^2	MAPE [%]	Time [s]
Automatic Relevance Determination	4,62	6,02	0,53	25	0,03
Kernel Ridge	4,65	6,04	0,53	25	0,02
Linear Regression	4,62	6,03	0,53	25	0,30
Huber Regressor	4,58	6,05	0,53	25	0,02
Ridge Regression	4,64	6,04	0,53	25	0,20
Bayesian Ridge	4,62	6,03	0,53	25	0,02
TheilSen Regressor	4,72	6,19	0,51	26	0,51
Least Angle Regression	4,87	6,38	0,47	26	0,01
Orthogonal Matching Pursuit	5,49	6,87	0,4	29	0,01
Lasso Regression	6,68	8,35	0,11	40	0,28
Elastic Net	6,68	8,34	0,11	40	0,02
Lasso Least Angle Regression	7,32	8,86	0	44	0,02

Verificou-se que os algoritmos *Automatic Relevance Determination*, *Kernel Ridge*, *Linear Regression*, *Huber Regression*, *Ridge Regression* e *Bayesian Ridge* apresentaram o mesmo coeficiente de determinação, que foi 0,02 maior que o do *Regressor TheilSen*, 0,06 maior que a *Least Angle Regression*, 0,13 maior do que *Orthogonal Matching Pursuit* e 0,42 maior que *Lasso Regression* e *Elastic Net*.

Apesar de apresentarem dinâmicas diferentes, o melhor coeficiente de determinação foi obtido pelo k-neighbors regressor, que foi 0,05 superior ao *multi-layer perceptron regressor*, 0,18 superior ao *Random Sample Consensus* e 0,45 superior à *Support Vector Regression* (Tabela 7).

Tabela 7. Métrica de avaliação de modelos baseados em *k-nearest neighbor*, *multi-layer perception regressor*, *random sample consensus*, *support vector regression*, *dummy regressor* e *passive-aggressive regressor* aplicado ao conjunto de treinamento da operação mecanizada de colheita de madeira.

Modelo	MAE	RMSE [$m^3\ h^{-1}$]	R^2	MAPE [%]	Time [s]
K-Neighbors Regressor	4,09	5,59	0,6	24	0,02
Multilayer Perceptron Regressor	4,57	5,94	0,55	25	0,40
Random Sample Consensus	5,10	6,73	0,42	27	0,02
Support Vector Regression	6,35	8,18	0,15	37	0,03
Dummy Regressor	7,32	8,86	0	44	0,01
Passive-Aggressive Regressor	8,25	10,02	−0,49	52	0,02

Entre os modelos aplicados que apresentaram melhores coeficientes de determinação estão o *blending ensemble* e o *stacking ensemble*. A seguir, os algoritmos foram testados em modo *default*, com destaque para o *Extra Trees Regressor* (Tabela 8).

Tabela 8. Desempenho obtido no conjunto de treinamento, em modo *default*, stacking ensemble learning e blending ensemble learning.

Modelo	Processo	MAE	RMSE [$m^3 h^{-1}$]	R^2	MAPE [%]
Extra Trees Regressor (A)	*Default*	3,31	4,49	0,74	18
CatBoost Regressor (B)	*Default*	3,13	4,58	0,73	17
K-Neighbors Regressor (C)	*Default*	4,06	5,54	0,61	24
A, B e C	*Blending ensemble*	3,18	4,43	0,75	18
A, B e C–A	*Stacking ensemble*	3,21	4,43	0,75	18

Ao analisar as métricas no teste de conjunto de dados, o modelo *blending ensemble* foi confirmado como o melhor preditor de produtividade dos *harvesters* (Tabela 9).

Tabela 9. Resultados obtidos no conjunto de dados de teste, em modo *default*, *stacking ensemble learning* e *blending ensemble learning*.

Modelo	Processo	R^2	MAPE [%]
A, B e C	*Blending ensemble*	0,77	15
A, B e C–A	*Stacking ensemble*	0,76	14
Extra Trees Regressor	*Default*	0,74	14
CatBoost Regressor	*Default*	0,74	15
K-Neighbors Regressor	*Default*	0,67	19

Além disso, como avaliação do desempenho geral do modelo, verificamos as 80 instâncias de dados do conjunto de testes, com a resposta. Assim, ficou evidente que o *blending ensemble*, seguido pelo *stacking ensemble*, produziu valores médios de R^2 relativamente mais elevados (Figura 1) e um menor grau de dispersão.

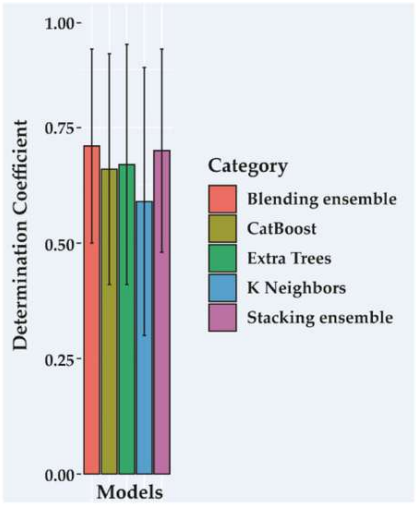


Figura 1. Valores médios dos coeficientes de determinação dos modelos empregados.

A Figura 2 ilustra o desempenho dos principais algoritmos utilizados na construção de modelos para previsão de produtividade, em relação aos valores observados.

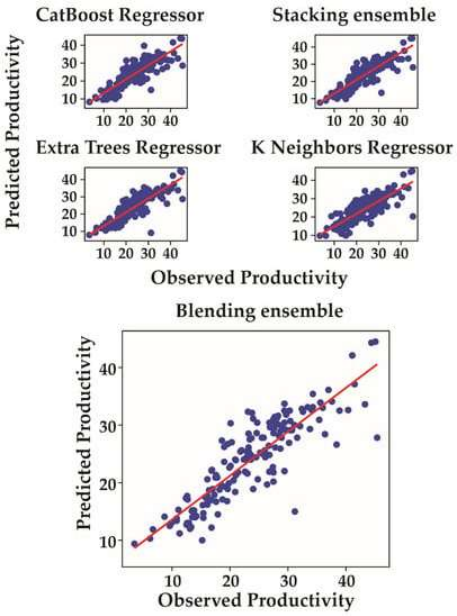


Figura 2. Comparação dos valores do gráfico de dispersão dos modelos empregados.

Quando selecionamos o algoritmo caixa preta, o aumento no desempenho comprometeu a interpretação das relações entre as variáveis preditoras e a variável alvo. Funções matemáticas complexas dificultavam inferências de especialistas técnicos. Porém, ao visualizar as distribuições das variáveis preditoras em cada quartil da variável resposta foi possível inferir que maior produtividade estava associada a maior disponibilidade de máquinas e menores níveis de inclinação.

Embora os algoritmos utilizados na construção dos modelos não permitam interpretabilidade, conforme evidenciado pelos quartis de produtividade do conjunto de testes, foi determinada a distribuição de densidade das variáveis preditoras dos volumes médios individuais das árvores, da inclinação do terreno e da disponibilidade das máquinas (Figura 3).

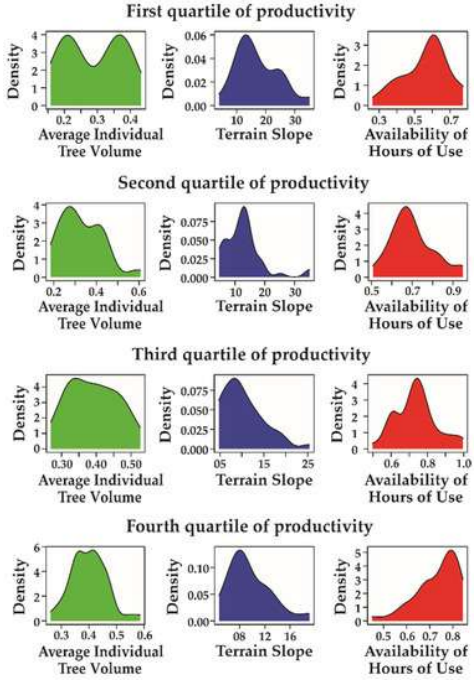


Figura 3. Distribuição de densidade dos modelos empregados.

1.4 Discussão

Incorporar modelos de *machine learning* às rotinas de gestão das operações florestais permite aos gestores inferir ajustes táticos e operacionais, com agilidade na tomada de decisões e prognósticos precisos. Na colheita mecanizada de madeira, a dinâmica do cenário e as influências externas que impactam as atividades exigem essa adaptabilidade juntamente com capacidades de previsão.

No entanto, a condução e o monitoramento do desempenho das operações mecanizadas de colheita de madeira, utilizando ferramentas analíticas como *machine learning*, são restritos devido à quantidade e qualidade dos dados disponíveis. Liski, *et al.* [31] e Maktoubian *et al.* [68]. Demirci *et al.* [69] e Abbasi *et al.* [70] relatam que a descentralização e a falta de gestão de dados em ambientes florestais reduzem a obtenção de resultados significativos.

Os *harvesters* que atuaram como fontes de dados possuíam tecnologia embarcada, com registros de saída das atividades de corte e seccionamento de madeira. A falta de interoperabilidade entre os dispositivos eletrônicos tornou suscetível a comunicação e a transferência de dados, comprometendo a possibilidade de correções instantâneas e a percepção de desvios nas notas. Além disso, Buccafurri *et al.* [71] e Shi *et al.* [72] apontam que a qualidade das instâncias geradas faz parte de um processo cooperativo, que exige participação e, portanto, nivelamento de todos os envolvidos. A gestão de dados deve estar alinhada com a organização das operações, o que torna responsabilidade dos gestores florestais coordenar esses esforços.

O volume residual de dados, após a execução dos processos de *data wrangling*, ainda foi suficiente para verificar o desempenho dos algoritmos de *machine learning* na modelagem da produtividade da colheita mecanizada de madeira. Dos grupos de algoritmos aplicados na modelagem, no processo padrão, os que tiveram melhor desempenho foram aqueles baseados em *decision tree*, *gradient-boosted machine* e *k-nearest neighbor*. Portanto, os algoritmos de melhor desempenho individual foram, respectivamente: *extract trees*, *gradient-boosted* e *k-nearest neighbor*.

Existem muitos tipos de árvores de decisão que têm como núcleo a entropia da informação. Segundo An e Zhou [73], no processo de análise específica, as informações obtidas para cada atributo são classificadas e ordenadas. Dentre os algoritmos de árvore de decisão avaliados, o que apresentou melhor desempenho foi o algoritmo de *extremely randomized trees* ou *extra trees*. Este algoritmo foi desenvolvido por Geurts *et al.*, [74] e utiliza o mesmo princípio do *random forests*. No entanto, conforme apoiado por Ahmad *et al.* [75], o *extra trees* pode ter se diferenciado usando todo o conjunto de dados de treinamento para treinar cada árvore de regressão e não apenas uma réplica de *bootstrap*.

Dos algoritmos baseados em *gradient-boosted machines*, o *CatBoost Regressor* apresentou o modelo que melhor se ajusta aos dados. Este algoritmo desenvolvido por Prokhorenkova *et al.* [76] é um aprimoramento do *gradient boosting*, projetado para evitar dependência de atributos e melhorar a precisão da previsão em pequenos conjuntos de dados. Por se tratar de um algoritmo não paramétrico que, segundo Ortiz-Bejar *et al*. [77], armazena todas as observações conhecidas e as utiliza na previsão baseada em funções de similaridade, o terceiro melhor desempenho foi do modelo baseado em *k-nearest neighbor*.

Como forma de melhorar a previsão, foram realizados testes com os métodos de aprendizagem *blending ensemble* e *stacking ensemble learning*, utilizando aprendizagem combinada. Esses aprendizados foram combinados a partir dos três algoritmos, em modo *default*, que apresentaram os melhores desempenhos. As previsões obtidas por ambos os métodos foram superiores às obtidas pelos algoritmos em modo *default*. Jong *et al.* [78] e Jordan e Mitchell [79] apontam que, em geral, métodos de aprendizagem combinados aumentam o desempenho de modelos construídos com *machine learning*.

Associar o uso do *blending ensemble* à possibilidade de pré-determinar a produtividade, com base em atributos de volumes médios individuais das árvores, declividade do terreno e disponibilidade de horas de uso das máquinas, promove dinamismo no planejamento dos gestores, especialmente no planejamento operacional, que exige respostas ágeis em condições operacionais adversas. Isto corrobora as limitações do método tradicional de estimativa de produtividade por meio do estudo de tempos.

Além disso, a comparação por meio dos valores dos diagramas de dispersão dos modelos empregados demonstrou os efeitos das variáveis preditoras sobre a produtividade. Nos quartis superiores, em condições de operação com menores declividades e maior disponibilidade da máquina, seus efeitos aumentam consideravelmente a produtividade.

A construção de modelos envolvendo algoritmos de *machine learning*, além de fornecer previsão da produtividade dos *harvesters* na operação mecanizada de colheita de madeira, permitiu olhar para as bases que orientam as decisões estratégicas das operações em florestas plantadas. Esta oportunidade mostrou que, apesar da qualidade, dados adequados promovem a extração de conhecimento, principalmente de atributos não correlacionados com a produtividade.

1.5 Conclusões

A partir da utilização de dados ajustados com algoritmos de *machine learning* é possível prever a produtividade das colheitadeiras de madeira.

Entre os atributos que compõem os conjuntos de dados das atividades mecanizadas de colheita de madeira, o volume médio individual das árvores, a inclinação do terreno e a disponibilidade de máquinas são os principais fatores que impactam a estimativa da produtividade dos *harvesters*.

Testar diversos algoritmos de *machine learning* com diferentes dinâmicas contribuiu para o desenvolvimento de uma técnica de *machine learning* que possibilitou o que se propõe, ou seja, a experimentação e o bom desempenho dos modelos. Assim, a escolha pelo *blending ensemble learning* foi orientada pela comparação de métricas estatísticas de ajuste do modelo.

Entre os métodos de aprendizagem por *blending ensemble*, *stacking ensemble* e algoritmos, no modo *default*, o *blending ensemble* apresentou coeficiente de determinação de 0,77 e erro percentual absoluto médio de 15%.

Agradecimento

Este estudo foi realizado com o apoio da Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

1. Pollard, S.J.T.; Brookes, A.; Earl, N.; Lowe, J.; Kearney, T.; Nathanail, C.P. Integrating decision tools for the sustainable management of land contamination. *Sci. Total Environ.* **2004**, *325*, 15–28.
2. Bai, C.; Sarkis, J. Integrating and extending data and decision tools for sustainable third-party reverse logistics provider selection. *Comput. Oper. Res.* **2019**, *110*, 188–207.
3. Welch, H.; Brodie, S.; Jacox, M.G.; Bograd, S.J.; Hazen, E.L. Decision-support tools for dynamic management. *Conserv. Biol.* **2020**, *34*, 589–599.
4. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419.
5. Liang, X.; Kankare, V.; Hyyppä, J.; Wang, Y.; Kukko, A.; Haggrén, H.; Yu, X.; Kaartinen, H.; Jaakkola, A.; Guan, F.; *et al.* Terrestrial laser scanning in forest inventories. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 63–77.
6. Eyvindson, K.; Saad, R.; Eriksson, L.O. Incorporating stand level risk management options into forest decision support systems. *For. Syst.* **2017**, *26*, e013.
7. Wagner, J.E. Misinterpreting the internal rate of return in sustainable forest management planning and economic analysis. *J. Sustain. For.* **2012**, *31*, 239–266.
8. Karttunen, K.; Laitila, J. Forest management regime options for integrated small-diameter wood harvesting and supply chain from young Scots pine (*Pinus sylvestris* L.) stands. *Int. J. For. Eng.* **2015**, *26*, 124–138.
9. Camargo, D.A.; Munis, R.A.; Simões, D. Investigation of exposure to occupational noise among forestry machine operators: A case study in Brazil. *Forests* **2021**, *12*, 299.
10. Fernandez-Lacruz, R.; Edlund, M.; Bergström, D.; Lindroos, O. Productivity and profitability of harvesting overgrown roadside verges—A Swedish case study. *Int. J. For. Eng.* **2021**, *32*, 19–28.
11. Visser, R.; Berkett, H. Effect of terrain steepness on machine slope when harvesting. *Int. J. For. Eng.* **2015**, *26*, 1–9.
12. Sherwin, L.M.; Owende, P.M.O.; Kanali, C.L.; Lyons, J.; Ward, S.M. Influence of tyre inflation pressure on whole-body vibrations transmitted to the operator in a *cut-to-length* timber *harvester*. *Appl. Ergon.* **2004**, *35*, 253–261.
13. Ovaskainen, H.; Heikkilä, M. Visuospatial cognitive abilities in *cut-to-length* single-grip timber *harvester* work. *Int. J. Ind. Ergon.* **2007**, *37*, 771–780.

14. Walsh, D.; Strandgard, M. ScienceDirect Productivity and cost of harvesting a stemwood biomass product from integrated *cut-to-length* harvest operations in Australian Pinus radiata plantations. *Biomass Bioenergy* **2014**, *738*, 93–102.
15. Hera, P.L.; Morales, D.O.; Mendoza-Trejo, O. A study case of Dynamic Motion Primitives as a motion planning method to automate the work of forestry cranes. *Comput. Electron. Agric.* **2021**, *183*, 106037.
16. Huang, X. Application analysis of AI reasoning engine in microblog culture industry. *Pers. Ubiquitous Comput.* **2020**, *24*, 393–403.
17. Cho, G.; Park, H.-M.; Jung, W.-M.; Cha, W.-S.; Lee, D.; Chae, Y. Identification of candidate medicinal herbs for skincare via data mining of the classic Donguibogam text on Korean medicine. *Integr. Med. Res.* **2020**, *9*, 100436.
18. Rodrigues de Holanda Maia, M.; Plastino, A.; Penna, P.H.V. MineReduce: An approach based on data mining for problem size reduction. *Comput. Oper. Res.* **2020**, *122*, 104995.
19. Xu, Z.; Cheng, X.; Wang, K.; Yang, S. Analysis of the environmental trend of network finance and its influence on traditional commercial banks. *J. Comput. Appl. Math.* **2020**, *379*, 112907.
20. da Silva, A.K.V.; Borges, M.V.V.; Batista, T.S.; da Junior, C.A.S.; Furuya, D.E.G.; Osco, L.P.; Teodoro, L.P.R.; Baio, F.H.R.; Ramos, A.P.M.; Gonçalves, W.N.; *et al.* Predicting *Eucalyptus* diameter at breast height and total height with uav-based spectral indices and machine learning. *Forests* **2021**, *12*, 582.
21. Giannetti, F.; Pecchi, M.; Travaglini, D.; Francini, S.; D'amico, G.; Vangi, E.; Cocozza, C.; Chirici, G. Estimating vaia windstorm damaged forest area in italy using time series sentinel-2 imagery and continuous change detection algorithms. *Forests* **2021**, *12*, 680.
22. Kandel, S.; Heer, J.; Plaisant, C.; Kennedy, J.; Van Ham, F.; Riche, N.H.; Weaver, C.; Lee, B.; Brodbeck, D.; Buono, P. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Inf. Vis.* **2011**, *10*, 271–288.
23. Endel, F.; Piringer, H. Data Wrangling: Making data useful again. *IFAC-PapersOnLine* **2015**, *28*, 111–112.
24. Furche, T.; Gottlob, G.; Libkin, L.; Orsi, G.; Paton, N.W. Data wrangling for big data: Challenges and opportunities. *Adv. Database Technol.-EDBT* **2016**, *2016*, 473–478.
25. Bellomarini, L.; Fayzrakhmanov, R.R.; Gottlob, G.; Kravchenko, A.; Laurenza, E.; Nenov, Y.; Reissfelder, S.; Sallinger, E.; Sherkhonov, E.; Vahdati, S.; *et al.* Data science with Vadalog: Knowledge Graphs with machine learning and reasoning in practice. *Futur. Gener. Comput. Syst.* **2022**, *129*, 407–422.
26. Bors, C.; Gschwandtner, T.; Miksch, S. Capturing and visualizing provenance from data wrangling. *IEEE Comput. Graph. Appl.* **2019**, *39*, 61–75.
27. Yang, J.; Li, R.; Chen, L.; Hu, Y.; Dou, Z. Research on equipment corrosion diagnosis method and prediction model driven by data. *Process Saf. Environ. Prot.* **2022**, *158*, 418–431.
28. De Jaegher, B.; Larumbe, E.; De Schepper, W.; Verliefde, A.; Nopens, I. Colloidal fouling in electrodialysis: A neural differential equations model. *Sep. Purif. Technol.* **2020**, *249*, 116939.
29. Kudyba, S. A hybrid analytic approach for understanding patient demand for mental health services. *Netw. Model. Anal. Health Inform. Bioinforma.* **2018**, *7*, 3.
30. Liu, J.; Kadziński, M.; Liao, X.; Mao, X.; Wang, Y. A preference learning framework for multiple criteria sorting with diverse additive value models and valued assignment examples. *Eur. J. Oper. Res.* **2020**, *286*, 963–985.
31. Liski, E.; Jounela, P.; Korpunen, H.; Sosa, A.; Lindroos, O.; Jylhä, P. Modeling the productivity of mechanized CTL harvesting with statistical machine learning methods. *Int. J. For. Eng.* **2020**, *31*, 253–262.
32. Sunhare, P.; Chowdhary, R.R.; Chattopadhyay, M.K. Internet of things and data mining: An application oriented survey. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 3569–3590.
33. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. *Forests* **2021**, *12*, 216.
34. Koreň, M.; Jakuš, R.; Zápotocký, M.; Barka, I.; Holuša, J.; Ďuračiová, R.; Blaženec, M. Assessment of machine learning algorithms for modeling the spatial distribution of bark beetle infestation. *Forests* **2021**, *12*, 395.
35. Kusiak, A. Data mining: Manufacturing and service applications. *Int. J. Prod. Res.* **2006**, *44*, 4175–4191.
36. Maxwell, A.E.; Warner, T.A.; Strager, M.P.; Conley, J.F.; Sharp, A.L. Assessing machine-learning algorithms and image- and lidar-derived variables for GEOBIA classification of mining and mine reclamation. *Int. J. Remote Sens.* **2015**, *36*, 954–978.
37. Taylor, P.; Griffiths, N.; Bhalerao, A.; Anand, S.; Popham, T.; Xu, Z.; Gelencser, A. Data Mining for Vehicle Telemetry. *Appl. Artif. Intell.* **2016**, *30*, 233–256.
38. Chatzimparmpas, A.; Martins, R.M.; Kucher, K.; Kerren, A. Empirical Study: Visual Analytics for Comparing Stacking to Blending Ensemble Learning. In Proceedings of the 2021 23rd International

Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, 26–28 May 2021; pp. 1–8.
39. Hansrajh, A.; Adeliyi, T.T.; Wing, J. Detection of Online Fake News Using Blending Ensemble Learning. *Sci. Program.* **2021**, *2021*, 3434458.
40. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Alhaj Hamoud, Y. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric.* **2021**, *184*, 106039.
41. Divina, F.; Gilson, A.; Goméz-Vela, F.; Torres, M.G.; Torres, J.F. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* **2018**, *11*, 949.
42. Malik, F.A.; Ye, W.; Chen, Q.; Li, D. Recommendation algorithm based on blending learning. In Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference, Guangzhou, China, 22–24 June 2019; pp. 113–117.
43. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 321–347.
44. Hao, M.; Cao, W.H.; Liu, Z.T.; Wu, M.; Xiao, P. Visual-audio emotion recognition based on multi-task and ensemble learning with multiple features. *Neurocomputing* **2020**, *391*, 42–51.
45. Sun, W.; Trevor, B. A stacking ensemble learning framework for annual river ice breakup dates. *J. Hydrol.* **2018**, *561*, 636–650.
46. Lee, J.; Kim, J.; Ko, W. Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method. *Appl. Sci.* **2019**, *9*, 1231.
47. Ünver-Okan, S. Modelling of work efficiency in cable traction with tractor implementing the least-squares methods and robust regression. *Croat. J. For. Eng.* **2020**, *41*, 109–117.
48. Arumugam, K.; Swathi, Y.; Sanchez, D.T.; Mustafa, M.; Phoemchalard, C.; Phasinam, K.; Okoronkwo, E. Towards applicability of machine learning techniques in agriculture and energy sector. *Mater. Today Proc.* **2022**, *51*, 2260–2263.
49. Morera, A.; Martínez de Aragón, J.; De Cáceres, M.; Bonet, J.A.; De-Miguel, S. Historical and future spatially-explicit climate change impacts on mycorrhizal and saprotrophic macrofungal productivity in Mediterranean pine forests. *Agric. For. Meteorol.* **2022**, *319*, 108918.
50. Jin, Z.; Shang, J.; Zhu, Q.; Ling, C.; Xie, W.; Qiang, B. RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In *Web Information Systems Engineering—WISE 2020*; Lecture Notes in Computer Science; Huang, Z., Beek, W., Wang, H., Zhou, R., Zhang, Y., Eds.; Springer: Cham, Switzerland, 2020; Volume 123, pp. 503–515.
51. Chen, Y.; Dong, C.; Wu, B. Crown Profile Modeling and Prediction Based on Ensemble Learning. *Forests* **2022**, *13*, 410.
52. Palonen, T.; Hyyti, H.; Visala, A. Augmented Reality in Forest Machine Cabin. *IFAC-PapersOnLine* **2017**, *50*, 5410–5417.
53. Marčeta, D.; Petković, V.; Ljubojević, D.; Potočnik, I. Harvesting System Suitability as Decision Support in Selection Cutting Forest Management in Northwest Bosnia and Herzegovina. *Croat. J. For. Eng.* **2020**, *41*, 251–265.
54. Aworka, R.; Cedric, L.S.; Adoni, W.Y.H.; Zoueu, J.T.; Mutombo, F.K.; Kimpolo, C.L.M.; Nahhal, T.; Krichen, M. Agricultural decision system based on advanced machine learning models for yield prediction: Case of East African countries. *Smart Agric. Technol. J* **2022**, *2*, 100048.
55. Kamarulzaman, A.M.M.; Jaafar, W.S.W.M.; Maulud, K.N.A.; Saad, S.N.M.; Omar, H.; Mohan, M. Integrated Segmentation Approach with Machine Learning Classifier in Detecting and Mapping Post Selective Logging Impacts Using UAV Imagery. *Forests* **2022**, *13*, 48.
56. National Institute of Meteorology. Available online: https//portal.inmet.gov.br/dadoshistoricos (accessed on 18 February 2021).
57. Speight, J.G.; Isbell, R.F. Soil Profiles. In *Australian Soil and Land Survey Field Handbook*; CSIRO: Canberra, Australia, 2009; pp. 132–133. ISBN 9780643093959.
58. R Development Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Computing Statistical: Vienna, Austria, 2021; ISBN 3-900051-07-0.
59. Suzuki, Y.; Yoshimura, T. Assessment of broad-leaved forest stand management: Stock densities, thinning costs and profits over a 60-year rotation period. *Croat. J. For. Eng.* **2019**, *40*, 365–375.
60. Lemm, R.; Blattert, C.; Holm, S.; Bont, L.; Thees, O. Improving economic management decisions in forestry with the sorsim assortment model. *Croat. J. For. Eng.* **2020**, *41*, 71–83.
61. Konstantinou, N.; Paton, N.W. Feedback driven improvement of data preparation pipelines. *Inf. Syst.* **2020**, *92*, 101480.

62. Tukey, J.W. *Exploratory Data Analysis*, 1st ed.; Pearson: London, UK, 1977; ISBN 978-0201076165.
63. Zhang, A.; Yu, H.; Zhou, S.; Huan, Z.; Yang, X. Instance weighted SMOTE by indirectly exploring the data distribution. *Knowl.-Based Syst.* **2022**, *249*, 108919.
64. Juez-Gil, M.; Arnaiz-González, Á.; Rodríguez, J.J.; López-Nozal, C.; García-Osorio, C. Approx-SMOTE: Fast SMOTE for Big Data on Apache Spark. *Neurocomputing* **2021**, *464*, 432–437.
65. PyCaret Org. 2021. Available online: https://pycaret.org (accessed on 18 February 2021).
66. Mitchell, T.M. *Machine Learning*, 1st ed.; McGraw-Hill Science/Engineering/Math: New York, NY, USA, 1977; ISBN 978-0070428072.
67. Lin, X.; Wu, J.; Wei, Y. An ensemble learning velocity prediction-based energy management strategy for a plug-in hybrid electric vehicle considering driving pattern adaptive reference SOC. *Energy* **2021**, *234*, 121308.
68. Maktoubian, J.; Taskhiri, M.S.; Turner, P. Intelligent predictive maintenance (Ipdm) in forestry: A review of challenges and opportunities. *Forests* **2021**, *12*, 1495.
69. Demirci, M.; Yesil, A.; Bettinger, P. Introducing a New Approach in Stand Tending Planning and Thinning Block Designation by Using Mixed Integer Goal Programming. *Croat. J. For. Eng.* **2022**, *43*, 134–151.
70. Abbasi, R.; Martinez, P.; Ahmad, R. The digitization of agricultural industry—A systematic literature review on agriculture 4.0. *Smart Agric. Technol.* **2022**, *2*, 100042.
71. Buccafurri, F.; De Meo, P.; Fugini, M.; Furnari, R.; Goy, A.; Lax, G.; Lops, P.; Modafferi, S.; Pernici, B.; Redavid, D.; *et al.* Analysis of QoS in cooperative services for real time applications. *Data Knowl. Eng.* **2008**, *67*, 463–484.
72. Shi, M.; Xu, J.; Liu, S.; Xu, Z. Productivity-Based Land Suitability and Management Sensitivity Analysis: The *Eucalyptus* E. urophylla × E. grandis Case. *Forests* **2022**, *13*, 340.
73. An, Y.; Zhou, H. Short term effect evaluation model of rural energy construction revitalization based on ID3 decision tree algorithm. *Energy Rep.* **2022**, *8*, 1004–1012.
74. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42.
75. Ahmad, M.W.; Reynolds, J.; Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **2018**, *203*, 810–821.
76. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *2018*, 6638–6648.
77. Ortiz-Bejar, J.; Graff, M.; Tellez, E.S.; Ortiz-Bejar, J.; Jacobo, J.C. κ-Nearest neighbor regressors optimized by using random search. In Proceedings of the 2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 14–16 November 2018.
78. Jong, N.; Krumeich, J.S.M.; Verstegen, D.M.L. To what extent can PBL principles be applied in blended learning: Lessons learned from health master programs. *Med. Teach.* **2017**, *39*, 203–211.
79. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260.

CAPÍTULO 2

Planejamento tático de *forwarders*: uma abordagem baseada em dados do baldeio de madeira no Brasil[2]

Resumo: O planejamento tático na colheita de madeira envolve aspectos relacionados ao macroplanejamento florestal e, particularmente, à alocação de recursos e ao sequenciamento das atividades, todos os quais afetam a alocação de madeira em pátios e estradas florestais, sobretudo, a produtividade das máquinas florestais. As abordagens baseadas em dados incentivam a utilização de informações obtidas a partir de dados para aumentar a eficiência da tomada de decisões e apoiar o desenvolvimento de estratégias de curto prazo. Portanto, nossa investigação teve como objetivo determinar se uma abordagem baseada em dados pode gerar subsídios suficientes para modelar a produtividade dos *forwarders* no baldeio de madeira em florestas plantadas de *Pinus* e *Eucalyptus*, para apoiar o planejamento tático. Utilizamos 3.812 instâncias de dados brutos que foram gerados ao longo de um período de 36 meses. Os dados foram coletados de 23 *forwarders* que atuavam em florestas plantadas de *Pinus* e *Eucalyptus*. Aplicamos 22 algoritmos de regressão, por um método de aprendizado supervisionado, a partir de uma abordagem experimental de *machine learning* às instâncias de dados. Avaliamos os modelos ajustados usando três métricas de desempenho. Dos algoritmos testados, o modo *default* do *light gradient boosting* produziu um erro quadrático médio de 14,80 m^3 h^{-1}, um erro médio absoluto de 2,70 e um coeficiente de determinação de 0,77. Portanto, os métodos baseados em dados apoiam adequadamente a modelagem da produtividade dos *forwarders* no baldeio de madeira em florestas plantadas e auxiliam os gestores florestais com o planejamento tático.

Palavras-chave: análise preditiva; *cut-to-length*; operações florestais; produtividade florestal; gestão florestal inteligente.

2.1 Introdução

As técnicas de manipulação de dados, como o *machine learning*, podem permitir o planeamento tático integrado da colheita mecanizada, do baldeio de madeira para pátios de armazenamento e de estradas florestais. A integração desses processos promove a gestão racional dos recursos, alocando de forma eficiente as matérias-primas para atender às necessidades das indústrias de base florestal. Os gestores florestais devem não apenas compreender a cadeia produtiva das florestas plantadas, mas também estar atentos aos incrementos nos padrões de desempenho do sistema homem-máquina. Avaliar as interações entre o sistema e o ambiente externo é crucial para identificar e aprimorar as funções do operador durante o processo de integração de uma máquina florestal.

Economicamente, as receitas devem compensar os custos de oportunidade incorridos e cobrir as despesas dos sistemas mecanizados escolhidos para garantir a viabilidade de um projeto [1]. Os custos financeiros envolvidos na implementação de estratégias de manejo florestal em sistemas de colheita de madeira podem ser avaliados por meio da construção de modelos preditivos para determinar sua viabilidade. Por exemplo, ao adotar um sistema de colheita, esses modelos oferecem uma classificação avançada das condições que enfrentará.

O sistema de colheita *cut-to-length* tornou-se o método mais popular na indústria de papel. Este sistema normalmente envolve *harvesters* e *forwarders* que processam e baldeiam as árvores, respectivamente [2-3]. Mapear a produtividade das máquinas de colheita florestal, como o *forwarder*, em sistemas *cut-to-length* pode ser uma fonte valiosa de dados para o gerenciamento florestal inteligente. Esses dados contabilizam todos os eventos e intervenções durante a colheita de madeira e podem ser utilizados para informar estratégias de gestão florestal [4].

O mapeamento da produtividade dos *forwarders* fornece aos gestores florestais dados quantitativos que podem ser utilizados de forma conveniente para o planejamento tático. A utilização de máquinas autopropelidas para a colheita de madeira é uma prática comum em todo o mundo nas operações florestais. Portanto, o desempenho de uma equipe de colheita depende de fatores discricionários controlados pela unidade, como a seleção de máquinas e operadores, que afetam sua capacidade de converter insumos em resultados [5-7].

[2] Capítulo publicado e redigido de acordo com as normas da revista Forests

Os gestores florestais podem antecipar desafios, alocar recursos e estabelecer objetivos utilizando os três níveis de tomada de decisão: tomada de decisão estratégica, táctica e operacional. No entanto, a nível tático, os gestores florestais avaliam as decisões de médio prazo necessárias para tomar decisões sustentáveis, garantir o fornecimento eficaz de matérias-primas e executar planos operacionais [8-11].

Os gestores florestais podem criar estratégias de colheita para cumprir objetivos unitários específicos dentro de uma determinada escala de tempo, utilizando abordagens baseadas em dados – como a coleta de dados em resposta a uma procura específica – ao dimensionar a frota de *forwarders*, por exemplo. Neste cenário, o uso de abordagens baseadas em dados pode reduzir ou eliminar a subjetividade na atribuição de parâmetros, levando a previsões confiáveis e eficientes [12-14].

As técnicas de *machine learning* utilizam uma abordagem baseada em dados para gerar informações confiáveis e acelerar o processo de obtenção de métricas de desempenho para atividades de *forwarders* em florestas plantadas. Infelizmente, as abordagens tradicionais para obter informações sobre a produtividade das máquinas de colheita florestal podem ser subjetivas; isso ocorre principalmente porque eles não são generalizáveis. Quando se trata de eficiência e precisão, os procedimentos baseados em dados têm uma vantagem sobre as abordagens tradicionais baseadas na experiência [15-17].

Para obter alta precisão durante a análise, técnicas de *machine learning* são comumente empregadas para criar metodologias de análise que identificam padrões ou tendências. Essas técnicas são utilizadas para executar funções críticas em mineração de dados, mineração de texto, análise preditiva e sistemas de tomada de decisão [18-20]. Um tipo de aprendizado de máquina é a abordagem supervisionada. Ao usar a abordagem supervisionada, os algoritmos são treinados usando dados rotulados para fazer previsões. A abordagem supervisionada apresenta diversas vantagens em cenários de análise de dados, pois permite previsões altamente precisas, adaptação a problemas complexos e otimização de decisões. No aprendizado supervisionado, o modelo ajusta seus parâmetros iterativamente para corresponder à saída das previsões com base na entrada, garantindo a precisão [21-22].

A literatura tem discutido o uso de aprendizado de máquina para prever a produtividade de máquinas florestais utilizadas na colheita mecanizada de madeira. Eriksson e Lindroos [23] treinaram modelos de aprendizado de máquina para estimar a produtividade de *harvesters* e *forwarders*. A precisão da estimativa foi maior para os *harvesters*. Guerra *et al.* [24] estimaram a produtividade de *harvesters* que operam em florestas de *Pinus* e *Eucalyptus*. Ghaffariyan *et al.* [25] previram a produtividade de *harvesters* e *forwarders* utilizando modelos de regressão múltipla. O foco estava no poder explicativo das variáveis preditoras.

No entanto, a previsão e o monitoramento da produtividade dos *forwarders* podem fornecer aos gestores florestais *insights* sobre previsão de demanda, rotas, análise comportamental, planejamento de instalações e, o mais importante, previsão de tráfego de caminhões [26]. Foi detectada uma lacuna na literatura em relação à utilização de técnicas de aprendizado de máquina e dados empíricos relativos a máquinas de colheita florestal, florestas plantadas e turnos de trabalho na estimativa da produtividade dos *forwarders*. À medida que os gestores florestais planejam a colheita de florestas plantadas, espera-se que um modelo preditivo de produtividade dos *forwarders* apoie o planejamento tático das operações, incluindo o dimensionamento do cronograma de atividades, os turnos de operações e o número de *forwarders* necessários.

Nós avaliamos se uma abordagem baseada em dados, especificamente aprendizado de máquina, pode fornecer informações suficientes para prever a produtividade no baldeio de madeira em florestas plantadas de *Pinus* e *Eucalyptus*. Nossa investigação teve como objetivo determinar se uma abordagem baseada em dados pode gerar informações suficientes para modelar a produtividade dos *forwarders* no transporte de madeira em florestas plantadas de *Pinus* e *Eucalyptus* para apoiar o planejamento tático.

2.2 Material e Métodos

2.2.1 Origem dos dados

O estudo coletou dados empíricos da e *controller area network* (CAN) da atividade de baldeio realizada por *forwarders* após a colheita de madeira ter sido realizada por *harvesters* em florestas plantadas de *Pinus* e *Eucalyptus*, no Brasil. As áreas totais dessas florestas foram de 325 ha e 290 ha para

Pinus e *Eucalyptus*, respectivamente. As declividades da região variam de 7,32% a 35,06%, e os intervalos foram classificados por Speight e Isbell [27] como suaves (3% a 10%), moderados (10% a 32%) e íngremes (32% a 56 %).

As florestas plantadas com *Pinus* tinham espaçamento de 3,3 m x 1,8 m, idade média de 22 anos e volume total colhido de 146.680 m^3. As florestas plantadas com *Eucalyptus* também tiveram espaçamento de 3,3 m x 1,8 m, mas a idade média foi de 14 anos e o volume total colhido foi de 235.720 m^3. A distância média para transporte de madeira da floresta de *Pinus* foi de 273 m, e para a floresta de *Eucalyptus* foi de 285 m. Após a colheita, a madeira era colocada nas bordas dos talhões, próximo às estradas florestais, para posterior carregamento e transporte até a indústria de base florestal para produção de papel.

Analisamos os dados operacionais de 21 *forwarders* Elephant King (modelo fabricado pela Ponsse Plc, Vieremä, Finlândia) em condições normais de trabalho. Os *forwarders* tinham potência nominal de 205 kW, capacidade de carga de 20.000 kg e estavam equipados com sistemas de rodados com pneus de 0,36 m^2 e tração 8×8. Além disso, analisamos dois *forwarders* Ponsse Buffalo (modelo fabricado pela Ponsse Plc, Vieremä, Finlândia) com potência de 210 kW e capacidade de 15.000 kg. Os *forwarders* eram equipados com garra de 0,36 m^2, sistemas de rodas com pneus e tração 8×8.

2.2.2 Variáveis dependentes e target no conjunto de dados

Utilizamos 3.812 instâncias de dados brutos gerados ao longo de um período de 36 meses. Como a previsão de rendimento pode ser influenciada por diferentes atributos do *forwarder* [28], o conjunto de dados avaliado consistiu em informações relacionadas aos *forwarders*, florestas plantadas e turnos de trabalho. A análise utilizou seis variáveis, sendo quatro contínuas (Tabela 1) e duas categóricas. A variável *target* foi as horas produtivas da máquina ($m^3\ h^{-1}$). Ressalta-se que o horímetro do *forwarder* foi incluído nos ajustes do modelo, mas acabou sendo removido por aumentar a complexidade do ajuste sem contribuir para a minimização do erro.

Tabela 1. Perfil dos dados do *dataset* utilizado para ajustar os modelos de previsão de produtividade do *forwarder*.

Variáveis	Média	Desvio Padrão	Valor Mínimo	Mediana	Valor Máximo
Inclinação do terreno (%)	12,85	5,81	4,49	11,84	37,46
Área do talhão (m^2)	19,48	19,20	0,11	12,42	76,71
Distância total (m)	5.387,78	2.125,68	-	5.380,50	17.269,00
Distância média (m)	281,41	170,68	-	231,79	1.489,50

Entre as variáveis dependentes relacionadas aos *forwarders*, avaliamos a distância total e média de transporte de madeira (expressa em metros). Os atributos relacionados às florestas plantadas que avaliamos incluíram declividade do terreno (%), área do talhão (m^2) e espécies. Para atender à demanda industrial por madeira no Brasil, normalmente são necessários dois turnos diários de trabalho. Em nossa análise também consideramos os turnos de trabalho, que transformamos em uma variável dummy com valor zero ou um para indicar a ausência ou presença do atributo. Isso nos permitiu classificar os dados em categorias mutuamente exclusivas, cada uma consistindo de 10 horas por dia.

A produtividade do *forwarder* foi estimada como o produto da capacidade máxima de madeira transportada pelo *forwarder* e o número reportado de cargas transportadas, dividido pelo tempo efetivo da atividade. Assim, a produtividade (Equação 1) é calculada dividindo-se o total de madeira encaminhada (υ) em m^3 pela soma correspondente de horas produtivas da máquina (h).

$$Produtividade = \frac{\upsilon}{h'} \quad (1)$$

Após obter os dados brutos, utilizamos o R para limpar e transformar o conjunto de dados, um processo conhecido como *data wrangling* [29]. A maioria das variáveis foi utilizada no modelo no seu estado original, com exceção da variável 'turnos de trabalho', que foi convertida em variáveis *dummy* para avaliar o seu impacto. Além disso, os *outliers* foram eliminados por meio da aplicação do método

do intervalo interquartil, delineado conceitualmente pelo intervalo de *Tukey*, e também foram removidos os valores *missings* das variáveis numéricas.

2.2.3 Modelagem da produtividade do forwarder

Na primeira etapa de modelagem, os algoritmos supervisionados são aplicados ao conjunto de dados de treinamento no modo *default*. Das instâncias de dados, atribuímos 80% (n = 3.049) como conjunto de treinamento e 20% (n = 763) como conjunto de teste.

Ressaltamos que a técnica de aprendizado de máquina supervisionado compreende uma série de etapas ordenadas necessárias para realizar o modelo de previsão. Essas etapas incluem: determinar o tipo de estudo – classificação, regressão ou detecção de anomalias; realizar etapas de limpeza e processamento de dados e transformar os dados para garantir que estejam preparados para análise; dividir os dados em conjuntos de treinamento, validação e teste; identificar as variáveis do conjunto de dados que servirão de entrada para o modelo; selecionar um ou mais algoritmos de aprendizado de máquina adequados para o problema; ajustar os hiperparâmetros do modelo usando o conjunto de validação (se aplicável) e avaliar o desempenho em dados desconhecidos; utilizar métricas de avaliação apropriadas para medir o desempenho do modelo nos dados de teste.

Os pesos atribuídos aos modelos de regressão são determinados com base na sua precisão. A previsão final é obtida iterativamente por meio de um fluxo de trabalho que consolida as previsões dos modelos de regressão aplicados ao conjunto de dados de teste [30-31]. Em nossa estrutura de aprendizagem supervisionada, os dados rotulados foram organizados aleatoriamente no conjunto de treinamento para atualizar os parâmetros. Como resultado, os rótulos foram exibidos durante o treinamento e mascarados durante os testes.

Um modelo de previsão de produtividade do *forwarder* foi construído usando a biblioteca scikit-learn [32] em *Python*. Algoritmos de aprendizado de máquina foram aplicados aos dados por meio de aprendizado supervisionado [33-34]. O ajuste de hiperparâmetros foi avaliado para otimizar os *folds*, iterações e estimador do modelo com diferentes métodos de aprendizado, como modo *default*, modo *tuned*, modo *voting* e modo de *stacking* [35-38].

Por *default*, os algoritmos têm impacto limitado porque não otimizam suas configurações de hiperparâmetros [39]. Para superar esse problema, os parâmetros do modelo são ajustados – por exemplo, com *Tuned* – para otimizar seu desempenho. Este processo envolve encontrar a combinação ideal, o que leva a um melhor desempenho.

Para desenvolver modelos capazes de prever a produtividade do *forwarder* e otimizar sua produtividade, é necessário o ajuste de hiperparâmetros. O ajuste de hiperparâmetros é a otimização automática dos parâmetros de um modelo de aprendizado de máquina [36].

No entanto, diversas técnicas avançadas envolvendo aprendizagem em conjunto são avaliadas, incluindo *voting* e *stacking*. O sistema de ponderação por *voting* permite que dois ou três métodos de aprendizado de máquina se complementem [37,40]. Um modelo *stacking* permite a combinação de vários algoritmos de aprendizado de máquina sob a supervisão do meta-modelo, alcançando, em última análise, a maior precisão possível [41].

O aprendizado de máquina é um processo experimental. Em outras palavras, quando construímos um modelo de aprendizado de máquina, permitimos que programas de computador aprendam com os dados fornecidos e façam ajustes nos hiperparâmetros com base nos mesmos dados [42]. Avaliamos vários algoritmos, hiperparâmetros e métodos de aprendizagem para identificar a solução mais eficaz.

Testamos os 22 algoritmos de aprendizado de máquina a seguir, com diferentes bases metodológicas para determinar o melhor ajuste para estimar a produtividade do *forwarder*, conforme sugerido por Souza *et al.* [43]. Este estudo utilizou um total de 22 algoritmos, escolhidos por sua facilidade de implementação e capacidade de executar processamento paralelo (*multithreading*) usando as bibliotecas *Python scikit-learn, CatBoost, XGBoost e SymbolicRegressor* [32,44-45]. Esses algoritmos pertencem à categoria de métodos lineares (Automatic Relevance Determination, Bayesian Ridge, ElasticNet, Huber, Lasso, LassoLars, Linear Regression, Orthogonal Matching Pursuit, Ridge), árvore de decisão (CART), nearest neighbors (Nearest Neighbors), support vector machine (Support Vector Machine Linear Kernel, SVM rbf Kernel), neural artificial network (Multilayer Perceptron), ensemble methods (AdaBoost, Extremely Randomized Trees, CatBoost, Extreme Gradient Boosting, Gradient Boosting, Light Gradient Boosting, Random Forest) e algoritmos genéticos (Symbolic Regressor).

2.2.4 Escolhendo o melhor modelo

Para avaliar o desempenho dos modelos ajustados, utilizamos três métodos quantitativos comumente recomendados [46,47]. Utilizamos o coeficiente de determinação (R2) dos valores previstos (Pi) e observados (Si) para expressar a variação da variável resposta sobre a variação das variáveis preditoras (Equação 2) [48]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (P_i - S_i)^2}{\sum_{i-1}^{n} (P_i - \bar{S})^2}, \quad (2)$$

A Root Mean Squared Error ($RMSE$) foi empregada para medir a distância euclidiana entre os valores reais dos valores preditos (Equação 3) [49]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - S_i)^2}{n}}, \quad (3)$$

Utilizamos a Mean Absolute Error (MAE) para medir o erro entre observações pareadas conforme a Equação 4 [50]:

$$MAE = \frac{\sum_{i=1}^{n} |P_i - S_i|}{n}, \quad (4)$$

2.3 Resultados

O desempenho de cada modelo foi avaliado comparando-se os resultados preditos por eles com os valores reais de produtividade. Conforme a Figure 1, os modelos, em modo *default*, produziram variações média de 0,56 ± 0,22. Destacando–se o Light Gradient Boosting com R² de 0,21 acima da média geral.

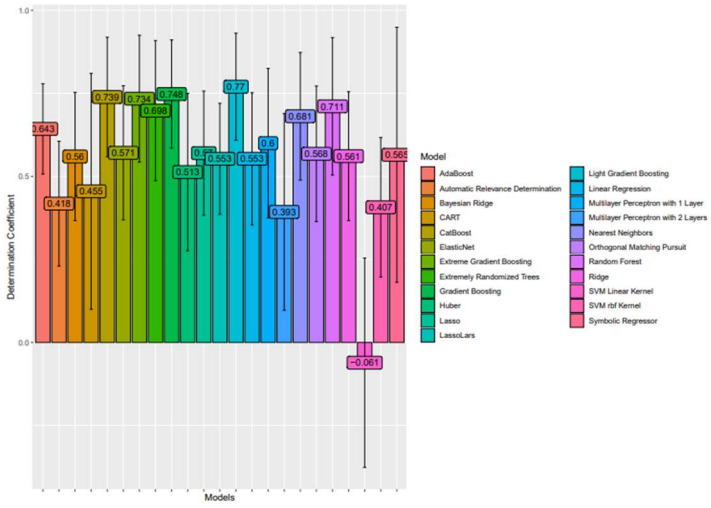


Figura 1. Coeficiente de determinação dos 22 modelos ajustados em modo *default*.

Em relação aos erros gerados na predição da produtividade (Tabela 2), em modo *default*, os modelos que produziram os menores ruídos foram quatro algoritmos de boosting (Light Gradient Boosting), dois de árvore de decisão (Random Forest) e um de agrupamento (Nearest Neighbors).

Tabela 2. Estimativa dos erros gerados na predição a partir dos modelos em modo default pelo Root Mean Squared Error ($RMSE$) e Mean Absolute Error (MAE).

Modelo	Média ± desvio padrão (RMSE)	Média ± desvio padrão (MAE)
Light Gradient Boosting	14,80 ± 10,55	2,70 ± 0,86
Gradient Boosting	16,11 ± 10,55	2,81 ± 0,87
CatBoost	16,89 ± 11,64	2,83 ± 0,87
Extreme Gradient Boosting	16,97 ± 11,99	2,89 ± 0,87
Random Forest	18,37 ± 13,10	3,00 ± 0,95
Extremely Randomized Trees	19,23 ± 13,39	3,01 ± 1,04
Nearest Neighbors	20,34 ± 12,31	3,33 ± 1,01
AdaBoost	23,11 ± 8,95	3,83 ± 0,79
Multilayer Perceptron with 1 Layer	25,82 ± 14,69	3,64 ± 0,99
Lasso	27,21 ± 11,39	4,16 ± 0,85
Symbolic Regressor	27,49 ± 22,90	3,28 ± 1,25
ElasticNet	27,57 ± 12,72	4,21 ± 0,89
Orthogonal Matching Pursuit	27,74 ± 13,20	4,20 ± 0,91
Bayesian Ridge	28,20 ± 12,36	4,28 ± 0,86
Ridge	28,27 ± 12,63	4,27 ± 0,96
LassoLars	28,79 ± 10,80	4,33 ± 0,82
Linear Regression	28,82 ± 12,88	4,29 ± 0,90
Huber	31,24 ± 12,50	4,27 ± 0,92
CART	34,43 ± 21,89	4,02 ± 1,27
Automatic Relevance Determination	37,70 ± 11,50	5,02 ± 0,80
SVM rbf Kernel	38,02 ± 13,38	4,97 ± 0,87
Multilayer Perceptron with 2 Layers	38,86 ± 18,38	4,77 ± 1,11
SVM Linear Kernel	68,01 ± 20,29	7,15 ± 1,20

Controlamos a força dos parâmetros de penalidade dos algoritmos *Ridge Regression* e *Lasso Regression*, que apresentaram melhor desempenho no modo *default*, por meio de um processo ajustado (Tabela 3).

Tabela 3. Desempenho dos melhores modelos no modo *default* após ajuste do coeficiente de determinação do processo (R^2), Root Mean Squared Error ($RMSE$) e Mean Absolute Error (MAE).

Modelo	Média ± desvio padrão (R^2)	Média ± desvio padrão (RMSE)	Média ± desvio padrão (MAE)
Light Gradient Boosting	0,74 ± 0,18	16,63 ± 12,16	2,87 ± 0,83
earest Neighbors	0,72 ± 0,18	17,68 ± 11,39	3,13 ± 0,85
Random Forest	0,69 ± 0,17	19,67 ± 11,34	3,32 ± 0,94

Considerando o modo *Stacking* dos algoritmos *Light Gradient Boosting, Nearest Neighbors* e o *Linear Regression*, nenhum dos nove arranjos elevaram a capacidade preditiva, obtendo um R^2 máximo de 0,70 ± 0,04, e mínimos RMSE de 19,03 ± 2,23 e MAE de 3,18 ± 0,18. Do mesmo modo, ao ajustar o modo *Stacking* com quatro camadas, portanto, 16 arranjos, alcançamos um R^2 máximo de 0.72 ± 0.03, e mínimos RMSE de 17,27 ± 2,14 e MAE de 2,99 ± 0,18. Apesar do aprendizado em modo *voting* ter se destacado entre as três variações de *stacking* (Tabela 4), não superou a performance do *Light Gradient Boosting* em modo *default*.

Tabela 4. Estimativa dos erros gerados na predição a partir dos modelos em modo *Voting* pelo *Root Mean Squared Error (RMSE) e Mean Absolute Error (MAE).*

Modelo	Modo	Média ± desvio padrão (R^2)	Média ± desvio padrão (RMSE)	Média ± desvio padrão (MAE)
Random Forest x Light Gradient Boosting	Voting	0,74 ± 0,03	16,20 ± 2,00	2,90 ± 0,16
Light Gradient Boosting x Nearest Neighbors	Voting	0,74 ± 0,03	16,20 ± 2,00	2,90 ± 0,16
Nearest Neighbors x Random Forest	Voting	0,74 ± 0,03	16,20 ± 2,00	2,90 ± 0,16
Nearest Neighbors x Light Gradient Boosting x Random Forest	Voting	0,75 ± 0,03	15,34 ± 2,02	2,81 ± 0,17

Neste contexto, mesmo com os diferentes métodos de aprendizado e com o ajuste dos hiperparâmetros, a melhor performance da predição da produtividade foi obtida por meio do algoritmo *Light Gradient Boosting*, em modo *default* (Figura 2).

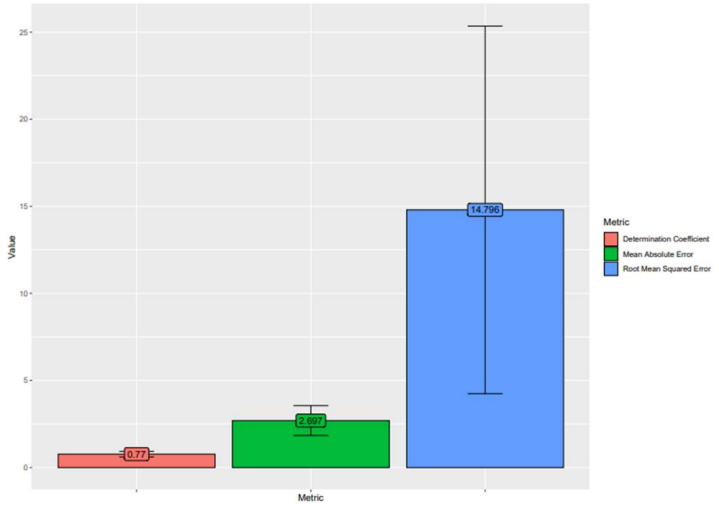


Figura 2. Métricas quantitativas da performance do modelo eleito Light Gradient Boosting.

2.4 Discussão

Cada algoritmo possuí uma base metodológica que confere pontos fortes e fracos, os quais impactam na performance do modelo em resposta à resolução do problema de predição da produtividade de *forwarders*. O desempenho de cada modelo, em modo *default*, comparando-se os resultados preditos por eles com os valores reais de produtividade, oscilou de -0,06 a 0,77 do *SVM Linear Kernel* ao *Light Gradient Boosting*, respectivamente. Isto pode estar relacionado ao tamanho do *dataset*, ou também pode estar relacionado à estrutura do conjunto de dados. Portanto, segundo Goliatt *et al.* [51] e Altin *et al.* [52], é fundamental escolher o algoritmo correto para um problema específico.

Vários fatores influenciaram o padrão de diferenciação entre os algoritmos. Conforme observado por Cipolla e Gondzio [53], o desempenho do *SVM Linear Kernel* está relacionado aos altos requisitos de armazenamento e manipulação de matrizes de *kernel* densas e não estruturadas. O desempenho da previsibilidade foi melhorado pelo *Light Gradient Boosting*, que foi desenvolvido para reduzir o consumo de memória [54]. O estudo preditivo altamente preciso de Sun *et al.* [55] confirmou que o desempenho do *Light Gradient Boosting* é mais robusto em comparação com o do SVM. Como resultado, corroboramos que os algoritmos baseados em *boosting* foram mais eficientes na previsão da produtividade do *forwarder*.

Além do desempenho dos algoritmos, os resultados da análise residual, ou seja, o RMSE e o MAE, apoiam a nossa escolha do modelo de melhor ajuste. Nesta análise, mesmo com penalidade, o *Light Gradient Boosting* se destacou entre os demais algoritmos testados. Segundo Saaidi *et al.* [56], esses ajustes são garantidos por um processo de otimização iterativo para minimizar pseudoresíduos. Porém, além de confirmar o ajuste do modelo pelo R^2, relacionamos o desempenho do modelo à minimização da função de erro entre os valores reais de produtividade e os valores de produtividade previstos.

A integração do ajuste de hiperparâmetros com a estratégia de aprendizagem visa melhorar o desempenho preditivo dos modelos gerados. Uma comparação entre os valores de produtividade previstos pelo *Light Gradient Boosting* e aqueles obtidos após o processo de ajuste mostrou uma redução de 0,03 no R^2 e incrementos de 1,83 e 0,17 no RMSE e MAE, respectivamente. Da mesma forma, a aplicação do método combinado de aprendizado de *voting* e *stacking* com três e quatro camadas de algoritmos resultou em um ligeiro declínio no desempenho do modelo. No entanto, no modo *default*, o *Light Gradient Boosting* persistiu como o melhor desempenho.

Li *et al.* [57] sugeriram que o uso de uma estratégia de aprendizado integrado deveria ser limitado a problemas de previsão com múltiplos estágios, onde diferentes fatores influenciam cada estágio. Para prever a produtividade dos *forwarders*, o algoritmo *Light Gradient Boosting* em modo *default* foi suficiente, visto que os fatores de influência foram generalizados em todas as etapas, e garantiu praticidade e capacidade preditiva abrangente.

Tiernan *et al.* [58] observaram que o uso de fatores influentes para prever a produtividade do *forwarder* no baldeio de madeira fornece um meio conveniente e transparente para negociar contratos florestais. Propor a produtividade esperada do *forwarder* no planejamento tático com uma explicação da variação das variáveis preditoras aumentou a autonomia dos gestores florestais e, mais importante, reduziu o otimizou o tempo de se dedicar a abordagens tradicionais, como estudos de tempo.

O modelo de aprendizado de máquina desenvolvido pode auxiliar os gestores florestais em suas atividades, fornecendo estimativas precisas do volume de madeira esperado, atingindo um R^2 de 0,77. Para obter uma estimativa mais precisa, os gestores florestais podem otimizar a logística de colheita e planejar taticamente as operações de acordo com as projeções de produção.

Malinovski *et al.* [59] constataram que a produtividade dos *forwarders* é influenciada pela distância de transporte devido à necessidade de possíveis viagens. Cadei *et al.* [60] confirmaram que a produtividade de um *forwarder* é significativamente afetada pela distância percorrida, volume de carga e inclinação do terreno durante o movimento. Levar em conta essas fontes de incerteza ao modelar a produtividade de um *forwarder* dá ao gestor florestal confiança na programação das operações, na alocação de equipes e máquinas e na definição de rotas em tempo hábil com base em decisões informadas.

A utilização de recursos orientados por dados promove a gestão florestal em gestão florestal inteligente. Os resultados dos eventos e ações executadas durante a manipulação das toras de madeira são indicadores da produtividade do *forwarder*. No entanto, um processo baseado em dados não pode depender exclusivamente da criação de um modelo de aprendizado automático para manter a

inferência. A observação é um passo subsequente crucial na incorporação e utilização deste recurso no planejamento tático das operações.

Feng *et al.* [61] relataram o desenvolvimento de diversas bibliotecas em diversas linguagens de programação que poderiam servir a esse propósito. Klass *et al.* [62] reforçam que as probabilidades relatadas a partir de modelos de aprendizado de máquina podem ser enganosas se as variáveis não forem monitoradas de forma agnóstica.

Devido à natureza em constante mudança do setor florestal, o modelo foi treinado numa base de dados transitória. Como resultado, os gestores florestais podem avaliar o modelo desenvolvendo métodos para demonstrar a sua capacidade de detectar até mesmo pequenos desvios no monitoramento, incluindo dimensões técnicas e empresariais, e comparando-o com dados recentes.

2.5 Conclusões

A produtividade dos *forwarders* no transporte de madeira em florestas plantadas de *Pinus* e *Eucalyptus* pode ser modelada usando abordagens baseadas em dados. Essa modelagem auxilia os gestores florestais no planejamento tático.

O modo *default* de otimização iterativa para minimizar pseudoresíduos usando *Light Gradient Boosting* alcança excelente desempenho em termos de estimativa de produtividade, com coeficiente de determinação de 0,77.

Os fatores que afetam a produtividade do *forwarder*, como distância de transporte de madeira, inclinação do terreno, área do talhão, espécie e turno de trabalho, são generalizados ao longo do processo de construção do modelo, dispensando a utilização de métodos de aprendizado integrado.

Agradecimento

Este estudo foi realizado com o apoio da Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

1. Rossi, V.; Claeys, F.; Bastin, D.; Gourlet-Fleury, S.; Guizol, P.; Eba'a-Atyi, R.; Sonwa, D.J.; Lescuyer, G.; Picard, N. Could REDD+ mechanisms induce logging companies to reduce forest degradation in Central Africa? *J. For. Econ.* **2017**, *29*, 107–117.
2. Landhäusser, S.M. Impact of slash removal, drag scarification, and mounding on lodgepole pine cone distribution and seedling regeneration after *cut-to-length* harvesting on high elevation sites. *For. Ecol. Manag.* **2009**, *258*, 43–49.
3. Prinz, R.; Spinelli, R.; Magagnotti, N.; Routa, J.; Asikainen, A. Modifying the settings of CTL timber harvesting machines to reduce fuel consumption and CO_2 emissions. *J. Clean. Prod.* **2018**, *197*, 208–217.
4. Felipe Maldaner, L.; de Paula Corrêdo, L.; Fernanda Canata, T.; Paulo Molin, J. Predicting the sugarcane yield in real-time by *harvester* engine parameters and machine learning approaches. *Comput. Electron. Agric.* **2021**, *181*, 105945.
5. Cudzik, A.; Brennensthul, M.; Białczyk, W.; Czarnecki, J. Tractive performance of tyres in forest conditions–Impact assessment of ground and tyres parameters. *Croat. J. For. Eng.* **2018**, *39*, 85–96.
6. Norihiro, J.; Ackerman, P.; Spong, B.D.; Längin, D. Productivity model for *cut-to-length harvester* operation in South African *Eucalyptus* pulpwood plantations. *Croat. J. For. Eng.* **2018**, *39*, 1–13.
7. Obi, O.F.; Visser, R. Including exogenous factors in the evaluation of harvesting crew technical efficiency using a multi-step data envelopment analysis procedure. *Croat. J. For. Eng.* **2018**, *39*, 153–162.
8. Martell, D.L.; Gunn, E.A.; Weintraub, A. Forest management challenges for operational researchers. *Eur. J. Oper. Res.* **1998**, *104*, 1–17.
9. Andersson, D.; Eriksson, L.O. Effects of temporal aggregation in integrated strategic/tactical and strategic forest planning. *For. Policy Econ.* **2007**, *9*, 965–981.
10. Lin, T.; Rodríguez, L.F.; Shastri, Y.N.; Hansen, A.C.; Ting, K.C. Integrated strategic and tactical biomass–biofuel supply chain optimization. *Bioresour. Technol.* **2014**, *156*, 256–266.
11. Fuentealba, S.; Pradenas, L.; Linfati, R.; Ferland, J.A. Forest harvest and sawmills: An integrated tactical planning model. *Comput. Electron. Agric.* **2019**, *156*, 275–281.
12. Nasiri, S.; Khosravani, M.R. Applications of data-driven approaches in prediction of fatigue and fracture. *Mater. Today Commun.* **2022**, *33*, 104437.

13. Wan, X.; Doherty, J. A data-driven approach for forecasting embankment settlement accounting for multi-stage construction. *Comput. Geotech.* **2022**, *152*, 105001.
14. Marenić, E.; Seychal, G.; Passieux, J.-C. Data driven approach in multiphysics framework: Application to coupled electro-mechanical problems. *Comput. Methods Appl. Mech. Eng.* **2022**, *395*, 114959.
15. Hasti, V.R.; Navarkar, A.; Gore, J.P. A data-driven approach using machine learning for early detection of the lean blowout. *Energy AI* **2021**, *5*, 100099.
16. Tinmitondé, S.; He, X.; Yan, L.; Hounye, A.H. Data-driven prediction of critical flutter velocity of long-span suspension bridges using a probabilistic machine learning approach. *Comput. Struct.* **2023**, *280*, 107002.
17. Meng, J.; Zhou, Y.-J.; Ye, T.-R.; Xiao, Y.-T.; Lu, Y.-Q.; Zheng, A.-W.; Liang, B. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. *Pet. Sci.* **2023**, *20*, 277–294.
18. Sunhare, P.; Chowdhary, R.R.; Chattopadhyay, M.K. Internet of things and data mining: An application oriented survey. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 3569–3590.
19. Ma, S.; Fildes, R. Forecasting third-party mobile payments with implications for customer flow prediction. *Int. J. Forecast.* **2020**, *36*, 739–760.
20. Sandoval Orozco, A.L.; Quinto Huamán, C.; Povedano Álvarez, D.; García Villalba, L.J. A machine learning forensics technique to detect post-processing in digital videos. *Futur. Gener. Comput. Syst.* **2020**, *111*, 199–212.
21. Lim, P.K.; Julca, I.; Mutwil, M. Redesigning plant specialized metabolism with supervised machine learning using publicly available reactome data. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1639–1650.
22. Singh, P. Systematic review of data-centric approaches in artificial intelligence and machine learning. *Data Sci. Manag.* **2023**, *6*, 144–157.
23. Eriksson, M.; Lindroos, O. Productivity of *harvesters* and *forwarders* in CTL operations in northern Sweden based on large follow-up datasets. *Int. J. For. Eng.* **2014**, *25*, 179–200.
24. Guerra, S.P.S.; Oguri, G.; Spinelli, R. Harvesting *Eucalyptus* energy plantations in Brazil with a modified New Holland forage *harvester*. *Biomass Bioenergy* **2016**, *86*, 21–27.
25. Ghaffariyan, M.R.; Acuna, M.; Brown, M. Machine Productivity Evaluation for *Harvesters* and *Forwarders* in Thinning Operations in Australia. *Silva Balc.* **2019**, *20*, 13–25.
26. Peng, W.; Karimi Sadaghiani, O. Enhancement of quality and quantity of woody biomass produced in forests using machine learning algorithms. *Biomass Bioenergy* **2023**, *175*, 106884.
27. Speight, J.G.; Isbell, R.F. Soil Profiles. In *Australian Soil and Land Survey Field Handbook*; CSIRO: Canberra, Australia, 2009; pp. 132–133. ISBN 9780643093959.
28. Sotomayor, L.N.; Cracknell, M.J.; Musk, R. Supervised machine learning for predicting and interpreting dynamic drivers of plantation forest productivity in northern Tasmania, Australia. *Comput. Electron. Agric.* **2023**, *209*, 107804.
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Computing Statistical: Vienna, Austria, 2022; ISBN 3-900051-07-0.
30. Naruei, I.; Keynia, F.; Sabbagh Molahosseini, A. Hunter–prey optimization: Algorithm and applications. *Soft Comput.* **2022**, *26*, 1279–1314.
31. Obaidullah, A.J. Implementing and tuning machine learning-based models for description of solubility variations of nanomedicine in supercritical solvent for development of green processing. *Case Stud. Therm. Eng.* **2023**, *49*, 103200.
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. Dong, M.; Sun, M.; Song, D.; Huang, L.; Yang, J.; Joo, Y.H. Real-time detection of wind power abnormal data based on semi-supervised learning Robust Random Cut Forest. *Energy* **2022**, *257*, 124761.
34. Amati, M.; Tiede, J.; Sun, Q.; Deilami, K.; Hurley, J.; Fox, A.; Dickson, J. Using machine learning to identify urban forest crown bounding boxes (CBB): Exploring a new method to develop urban forest policy. *Urban For. Urban Green.* **2023**, *85*, 127943.
35. Xie, Y.; Zhu, C.; Zhou, W.; Li, Z.; Liu, X.; Tu, M. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *J. Pet. Sci. Eng.* **2018**, *160*, 182–193.
36. Jamaleddyn, I.; El ayachi, R.; Biniz, M. An improved approach to Arabic news classification based on hyperparameter tuning of machine learning algorithms. *J. Eng. Res.* **2023**, *11*, 100061.

37. Kim, S.; Yoon, H.C.; Lim, J.-T.; Jeong, D.; Kim, K.H. Productivity prediction in the Wolfcamp A and B using weighted voting ensemble machine learning method. *Gas Sci. Eng.* **2023**, *111*, 204916.
38. Ghasemieh, A.; Lloyed, A.; Bahrami, P.; Vajar, P.; Kashef, R. A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. *Decis. Anal. J.* **2023**, *7*, 100242.
39. Nadirgil, O. Carbon price prediction using multiple hybrid machine learning models optimized by genetic algorithm. *J. Environ. Manag.* **2023**, *342*, 118061.
40. Tavana, P.; Akraminia, M.; Koochari, A.; Bagherifard, A. An efficient ensemble method for detecting spinal curvature type using deep transfer learning and soft voting classifier. *Expert Syst. Appl.* **2023**, *213*, 119290.
41. Shafighfard, T.; Bagherzadeh, F.; Rizi, R.A.; Yoo, D.-Y. Data-driven compressive strength prediction of steel fiber reinforced concrete (SFRC) subjected to elevated temperatures using stacked machine learning algorithms. *J. Mater. Res. Technol.* **2022**, *21*, 3777–3794.
42. Arboretti, R.; Ceccato, R.; Pegoraro, L.; Salmaso, L. Design of Experiments and machine learning for product innovation: A systematic literature review. *Qual. Reliab. Eng. Int.* **2022**, *38*, 1131–1156.
43. de Souza, L.P.B.; Ferreira, G.R.B.; Camerini, I.G.; de Correia, T.M.; de Ribeiro, M.G.C.; Hidalgo, J.A.S.; de São João, B.L.D.; Llerena, R.W.A.; Kubrusly, A.C.; Ayala, H.V.H.; *et al.* Machine learning-based cement integrity evaluation with a through-tubing logging experimental setup. *Geoenergy Sci. Eng.* **2023**, *227*, 211882.
44. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *2018*, 6638–6648.
45. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
46. Begum, M.Y. Advanced modeling based on machine learning for evaluation of drug nanoparticle preparation via green technology: Theoretical assessment of solubility variations. *Case Stud. Therm. Eng.* **2023**, *45*, 103029.
47. Brugere, L.; Kwon, Y.; Frazier, A.E.; Kedron, P. Improved prediction of tree species richness and interpretability of environmental drivers using a machine learning approach. *For. Ecol. Manag.* **2023**, *539*, 120972.
48. Jayasinghe, T.; wei Chen, B.; Zhang, Z.; Meng, X.; Li, Y.; Gunawardena, T.; Mangalathu, S.; Mendis, P. Data-driven shear strength predictions of recycled aggregate concrete beams with/without shear reinforcement by applying machine learning approaches. *Constr. Build. Mater.* **2023**, *387*, 131604.
49. Liemohn, M.W.; Shane, A.D.; Azari, A.R.; Petersen, A.K.; Swiger, B.M.; Mukhopadhyay, A. RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *J. Atmos. Solar-Terrestrial Phys.* **2021**, *218*, 105624.
50. Sathishkumar, V.E.; Ramu, A.G.; Cho, J. Machine learning algorithms to predict the catalytic reduction performance of eco-toxic nitrophenols and azo dyes contaminants (Invited Article). *Alex. Eng. J.* **2023**, *72*, 673–693.
51. Goliatt, L.; Saporetti, C.M.; Oliveira, L.C.; Pereira, E. Performance of evolutionary optimized machine learning for modeling total organic carbon in core samples of shale gas fields. *Petroleum*, 2023; *in press*.
52. Altin, F.G.; Budak, İ.; Özcan, F. Predicting the amount of medical waste using kernel-based SVM and deep learning methods for a private hospital in Turkey. *Sustain. Chem. Pharm.* **2023**, *33*, 101060.
53. Cipolla, S.; Gondzio, J. Training very large scale nonlinear SVMs using Alternating Direction Method of Multipliers coupled with the Hierarchically Semi-Separable kernel approximations. *EURO J. Comput. Optim.* **2022**, *10*, 100046.
54. Thongthammachart, T.; Araki, S.; Shimadera, H.; Matsuo, T.; Kondo, A. Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO2 and PM2.5 levels in Kansai region, Japan. *Environ. Model. Softw.* **2022**, *155*, 105447.
55. Sun, X.; Liu, M.; Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Financ. Res. Lett.* **2020**, *32*, 101084.
56. Saaidi, A.; Bichri, A.; Abderafi, S. Efficient machine learning model to predict dynamic viscosity in phosphoric acid production. *Results Eng.* **2023**, *18*, 101024.
57. Li, N.; Wu, Y.; Wang, Q.; Ye, H.; Wang, L.; Jia, M.; Zhao, S. Underground mine truck travel time prediction based on stacking integrated learning. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105873.
58. Tiernan, D.; Zeleke, G.; Owende, P.M.; Kanali, C.; Lyons, J.; Ward, S. Effect of Working Conditions on *Forwarder* Productivity in *Cut-to-length* Timber Harvesting on Sensitive Forest Sites in Ireland. *Biosyst. Eng.* **2004**, *87*, 167–177.

59. Malinovski, R.A.; Malinovski, R.A.; Malinovski, J.R. Análise das variáveis de influência na produtividade das máquinas de colheita de madeira em função das características físicas do terreno, do povoamento e do planejamento operacional florestal. *Floresta* **2006**, *36*, 169–182.
60. Cadei, A.; Mologni, O.; Röser, D.; Cavalli, R.; Grigolato, S. *Forwarder* Productivity in Salvage Logging Operations in Difficult Terrain. *Forests* **2020**, *11*, 341.
61. Feng, X.; Wang, D.; Hou, B.; Yan, T. Interpretable federated learning for machine condition monitoring: Interpretable average global model as a fault feature library. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106632.
62. Klass, L.; Kabza, A.; Sehnke, F.; Strecker, K.; Hölzle, M. Lifelong performance monitoring of PEM fuel cells using machine learning models. *J. Power Sources* **2023**, *580*, 233308.

CONSIDERAÇÕES FINAIS

A correta captação de informações nas operações florestais, aliada à gestão eficiente, pode maximizar a produtividade na colheita mecanizada de madeira. O uso de modelos de *machine learning*, como analisado nos dois capítulos, proporciona uma ferramenta valiosa para a predição e otimização da produtividade de *harvesters* e *forwarders*.

A incorporação de dados como disponibilidade de horas de uso, volumes médios de árvores e declividades do terreno destaca a importância da variedade de preditores na modelagem das produtividades de máquinas florestais autopropelidas. Os resultados da aprendizagem por *blending* e do modo *default* do *light gradient boosting* destacaram-se na estimativa da produtividade de *harvesters* e *forwarders*, respectivamente.

A aplicação de 24 algoritmos de *machine learning* e 22 algoritmos de regressão em diferentes contextos contribui para uma abordagem experimental robusta e que capta a dinâmica de cada modelo ajustado. A abordagem baseada em dados revelou-se eficaz para apoiar o planejamento tático na colheita de madeira, fornecendo subsídios suficientes para a modelagem da produtividade das máquinas florestais autopropelidas.

A consideração da diversidade dos dados e a aplicação de técnicas de *machine learning* contribuíram para a maximização do desempenho dos modelos, facilitando a experimentação e a tomada de decisões. Essas considerações e aplicações práticas destacam o papel crucial que a tecnologia de *machine learning* desempenha na otimização das operações florestais, proporcionando *insights* valiosos para os gestores florestais e contribuindo para uma abordagem mais eficiente e estratégica no setor.

REFERÊNCIAS

AERTS, R.; HONNAY, O**.** Forest restoration, biodiversity and ecosystem functioning. **BMC Ecology**, London, v. 11, p. 29, 2011. Disponível em: https://bmcecol.biomedcentral.com/articles/10.1186/1472-6785-11-29. Acesso em: 30 de jan. 2024.

ÁGUILA, M. M. R.; BENÍTEZ-PAREJO, N. Simple linear and multivariate regression models. **Allergologia et Immunopathologia**, Madrid, v. 39, n. 3, p. 159-173. Disponível em: https://doi.org/10.1016/j.aller.2011.02.001. Acesso em: 30 de dez. 2021.

AHMED, H.; YOUNIS, E. M. G.; HENDAWI, A.; ALI, A. A. Heart disease identification from patients' social posts, machine learning solution on Spark. **Future Generation Computer Systems**, Amsterdam, v. 111, p. 714-722, 2020. Disponível em: https://doi.org/10.1016/j.future.2019.09.056. Acesso em: 12 de jul. 2020.

ALAYET, C.; LEHOUX, N.; LEBEL, L.; BOUCHARD, M. Centralized supply chain planning model for multiple forest companies. **Information Systems and Operational Research**, Toronto, v. 54, n. 3, p. 171-191, 2016. Disponível em: https://doi.org/10.1080/03155986.2016.1197544. Acesso em: 03 de mar. 2022.

ALIPOUR, M.; HARRIS, D. K. A big data analytics strategy for scalable urban infrastructure condition assessment using semi-supervised multi-transform self-training. **Journal of Civil Structural Health Monitoring**, Cham, v. 10, p. 313-332, 2020. Disponível em: https://doi.org/10.1007/s13349-020-00386-4. Acesso em: 11 de jul. 2020.

ALMAGRO, G. G.; LUENGO, J.; CANO, J. R.; GARCÍA, S. DILS: Constrained clustering through dual iterative local search. **Computers & Operations Research**, Amsterdam, v. 121, n. 104979, 2020. Disponível em: https://doi.org/10.1016/j.cor.2020.104979. Acesso em: 10 de jul. 2020.

ALMEIDA, R. O.; DA SILVA, R. B. G.; SIMÕES, D. Cut-to-length harvesting prediction tool: machine learning model based on harvest and weather features. **Forests**, Basel, v. 15, n. 8, p. 1398, 2024. Disponível em: https://doi.org/10.3390/f15081398. Acesso em: 30 de jan. 2024.

ALRFOU, K.; KORDIJAZI, A.; ROHATGI, P.; ZHAO, T. Synergy of unsupervised and supervised machine learning methods for the segmentation of the graphite particles in the microstructure of ductile iron. **Materials Today Communications**, Oxford, v. 30, n. 103174, 2022. Disponível em: https://doi.org/10.1016/j.mtcomm.2022.103174. Acesso em: 10 de fev. 2022.

AMORIM, W. P.; TETILA, E. C.; PISTORI, H.; PAPA, J. P. Semi-supervised learning with convolutional neural networks for UAV images automatic recognition. **Computers and Electronics in Agriculture**, Amsterdam, v. 164, n. 104932, 2019. Disponível em: https://doi.org/10.1016/j.compag.2019.104932. Acesso em: 09 de jul. 2020.

ANANDA, J.; HERATH, G. A critical review of multi-criteria decision making methods with special reference to forest management and planning. **Ecological Economics**, Amsterdam, v. 68, n. 10, p. 2535-2548, 2009. Disponível em: https://doi.org/10.1016/j.ecolecon.2009.05.010. Acesso em: 21 de dez. 2021.

ARAÚJO JÚNIOR, C. A.; LEITE, H. G.; SOARES, C. P. B.; BINOTI, D. H. B.; SOUZA, A. P.; SANTANA, A. F.; TORRE, C. M. M. E. A multi-agent system for forest transport activity planning. **Revista Cerne**, Lavras, v. 23, n. 3, p. 329-337, 2017. Disponível em: https://doi.org/10.1590/01047760201723032335. Acesso em: 03 de ago. 2020.

ATA, R. Artificial neural networks applications in wind energy systems: a review. **Renewable and Sustainable Energy Reviews**, Amsterdam, v. 49, p. 534-562, 2015. Disponível em: https://doi.org/10.1016/j.rser.2015.04.166. Acesso em: 05 de jan. 2022.

AYDEMIR, M. S.; BILGIN, G. Semi-supervised sparse representation classifier (S3RC) with deep features on small sample sized hyperspectral images. **Neurocomputing**, Amsterdam, v. 399, p. 213-226, 2020. Disponível em: https://doi.org/10.1016/j.neucom.2020.02.092. Acesso em: 11 de jul. 2020.

BAI, C.; SARKIS, J. Integrating and extending data and decision tools for sustainable third-party reverse logistics provider selection. **Computers & Operations Research**, Amsterdam, v. 110, p. 188-207, 2019. Disponível em: https://doi.org/10.1016/j.cor.2018.06.005. Acesso em: 22 de dez. 2021.

BAIDAI, Y.; DAGORN, L.; AMANDE, M. J.; GAERTNER, D.; CAPELLO, M. Machine learning for characterizing tropical tuna aggregations under Drifting Fish Aggregating Devices (DFADs) from commercial echosounder buoys data. **Fisheries Research**, Amsterdam, v. 229, n. 105613, 2020. Disponível em: https://doi.org/10.1016/j.fishres.2020.105613. Acesso em: 13 de jul. 2020.

BALZOTTI, C. S.; ASNER, G. P.; ADKINS, E. D.; PARSONS, E. W. Spatial drivers of composition and connectivity across endangered tropical dry forests. **Journal of Applied Ecology**, London, p. 1-12. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1111/1365-2664.13632. Acesso em: 14 de jul. 2020.

BASHIR, S.; CARTER, E. M. Robust mixture of linear regression models. **Communications in Statistics - Theory and Methods**, New York, v. 41, n. 18, p. 3371-3388. Disponível em: https://doi.org/10.1080/03610926.2011.558655. Acesso em: 29 de dez. 2021.

BECKER, R. M.; KEEFE, R. F.; ANDERSON, N. M.; EITEL, J. U. H. Use of lidar-derived landscape parameters to characterize alternative harvest system options in the Inland Northwest. **International Journal of Forest Engineering**, London, v. 29, n. 3, p. 179-191, 2018. Disponível em: https://doi.org/10.1080/14942119.2018.1497255. Acesso em: 18 de jul. 2020.

BELLOMARINI, L.; FAYZRAKHMANOV, R. E.; GOTTLOB, G.; KRAVCHENKO, A.; LAURENZA, E.; NENOV, Y.; REISSFELDER, S.; SALLINGER, E.; SHERKHONOV, E.; VAHDATI, S.; WU, L. Data science with vadalog: knowledge graphs with machine learning and reasoning in practice. **Future Generation Computer Systems**, Amsterdam, v.129, p. 407-422, 2022. Disponível em: https://doi.org/10.1016/j.future.2021.10.021. Acesso em: 29 de nov. 2021.

BERTSIMAS, D.; LI, M. L. Scalable holistic linear regression. **Operations Research Letters**, Amsterdam, v. 48, n. 3. p. 203-208, 2020. Disponível em: https://doi.org/10.1016/j.orl.2020.02.008. Acesso em: 30 de nov. 2021.

BHOPALE, A. P.; TIWARI, A. Swarm optimized cluster based framework for information retrieval. **Expert Systems with Applications**, Amsterdam, v. 154, n. 113441, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113441. Acesso em: 10 de jul. 2020.

BLIGÅRD, L. O.; HÄGGSTRÖM, C. CCPE – the use of an analytical method to evaluate safety and ergonomics in maintenance of forest machinery. **Australian Forestry**, Abingdon, v. 82, n. 1, p. 29-34, 2019. Disponível em: https://doi.org/10.1080/00049158.2019.1574540. Acesso em: 03 de mar. 2022.

BÓRDON, M. R.; MONTAGNA, J. M.; CORSANO, G. Mixed integer linear programming approaches for solving the raw material allocation, routing and scheduling problems in the forest industry. **International Journal of Industrial Engineering Computations**, Toronto, v. 11, p. 525-548, 2020. Disponível em: http://www.growingscience.com/ijiec/Vol11/IJIEC_2020_13.pdf. Acesso em: 03 de ago. 2020.

BORS, C.; GSCHWANDTNER, T.; MIKSCH, S. Capturing and visualizing provenance from data wrangling. **IEEE Computer Graphics and Applications**, Los Alamitos, v. 39, n. 6, p. 61-75, 2019. Disponível em: https://doi.org/10.1109/MCG.2019.2941856. Acesso em: 04 de jan. 2022.

BOZDAĞ, A.; DOKUZ, Y.; GÖKÇEK, O. B. Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey. **Environmental Pollution**, Amsterdam, v. 263, n. 114635, 2020. Disponível em: https://doi.org/10.1016/j.envpol.2020.114635. Acesso em: 10 de jul. 2020.

BOZHESKU, M. G. Towards improving strategic environmental assessment follow-up through stakeholder participation: a case of the Pasquia-Porcupine Forest Management Plan, Saskatchewan, Canada. **Journal of Environmental Planning and Management**, Abingdon, v. 55, n. 8, p. 1058-1074, 2011. Disponível em: https://doi.org/10.1080/09640568.2011.636544. Acesso em: 03 de mar. 2022.

BREIMAN, L. **Random forests-random features**. Technical Report 567, Statistics Department, University of California, Berkeley, 1999. Disponível em: https://www.stat.berkeley.edu/~breiman/random-forests.pdf. Acesso em: 13 de jul. 2020.

BREIMAN, L. **Random Forests. Machine Learning**, v. 45, p. 5-32, 2001. Disponível em: https://doi.org/10.1023/A:1010933404324. Acesso em: 13 de jul. 2020.

BUMBER, Z.; ĐUKA, A.; PANDUR, Z.; PORŠINSKY, T. Gradeability of a *forwarder* based on traction performance. **Forests**, Basel, v. 14, n. 1, p. 103, 2023. Disponível em: https://doi.org/10.3390/f14010103. Acesso em: 05 de nov. 2024.

CADEI, A.; MOLOGNI, O.; RÖSER, D.; CAVALLI, R.; GRIGOLATO, S. Forwarder productivity in salvage logging operations in difficult terrain. **Forests**, Basel, v. 11, n. 3, p. 341, 2020. Disponível em: https://doi.org/10.3390/f11030341. Acesso em: 05 de nov. 2024.

CAMARGO, D. A.; RAFAELE, A. M.; SIMÕES, D. Investigation of exposure to occupational noise among forestry machine operators: a case study in Brazil. **Forests**, Basel, v. 2, n. 3, p. 299. Disponível em: https://doi.org/10.3390/f12030299. Acesso em: 15 de fev. 2022.

CAMASTRA F., VINCIARELLI A. **Machine Learning for Audio, Image and Video Analysis** 978-1-4471-6734-1, Springer, London, UK (2015).

CAMPBELL, C.; YING, Y. Learning with Support Vector Machines. **Synthesis Lectures on Artificial Intelligence and Machine Learning**, San Rafael, v. 5, n. 1, p. 1-95. Disponível em: https://doi.org/10.2200/s00324ed1v01y201102aim010. Acesso em: 03 de jan. 2021.

CECCHERINI, G.; DUVEILLER, G.; GRASSI, G.; LEMOINE, G.; AVITABILE, V.; PILLI, R.; CESCATTI, A. Abrupt increase in harvested forest area over Europe after 2015. **Nature**, London, v. 583, p. 72-77, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1038/s41586-020-2438-y. Acesso em: 18 de jul. 2020.

CHATZIMPARMPAS, A.; MARTINS, R. M.; KUCHER, K.; KERREN, A. Empirical study: visual analytics for comparing stacking to blending ensemble learning. In: *International Conference on Control Systems and Computer Science*, 24., 2021, Bucharest. Proceedings [...]. *Piscataway*: IEEE, 2021. p. 1-8. DOI: https://doi.org/10.1109/CSCS52396.2021.00008. Acesso em: 23 dez. 2021.

CHEN, L. H.; HSU, B. C.; SU, C. W. A supervised learning approach to flashlight detection. **Cybernetics and Systems**, Philadelphia, v. 48, p. 1-12, 2016. Disponível em: https://doi.org/10.1080/01969722.2016.1243400. Acesso em: 26 de fev. 2022.

CHEN, C.; NARANI, A.; DANIYAR, A.; MCCAULEY, J.; BROWN, S.; PRAY, T.; TANJORE, D. Ensemble models of feedstock blend ratios to minimize supply chain risk in bio-based manufacturing. **Biochemical Engineering Journal**, Amsterdam, In Press, 2020. Disponível em: https://doi.org/10.1016/j.bej.2020.107896. Acesso em: 23 de dez. 2021.

CHEN, C.; TANAKA, K.; KOTERA, M.; FUNATSU, K. Comparison and improvement of the predictability and interpretability with ensemble learning models in QSPR applications. **Journal of Cheminformatics**, London, v. 12, n. 19, p. 1-16, 2020.

Disponível em: https://doi.org/10.1186/s13321-020-0417-9. Acesso em: 23 de dez. 2021.

CHEN, H.; YANG, B.; WANG, G.; LIU, J.; XU, X.; WANG, S.; LIU, D. A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. **Knowledge-Based Systems**, Amsterdam, v. 24, n. 8, p. 1348-1359, 2011. Disponível em: https://doi.org/10.1016/j.knosys.2011.06.008. Acesso em: 03 de jan. 2021.

CHO, G.; PARK, H. M.; JUNG, W. M.; CHA, W. S.; LEE, D.; CHAE, Y. Identification of candidate medicinal herbs for skincare via data mining of the classic donguibogam Text on Korean Medicine. **Integrative Medicine Research**, Seoul, n. 100436. In Press, 2020. Disponível em: https://doi.org/10.1016/j.imr.2020.100436. Acesso em: 09 de jul. 2020.

CHOWDHURY, R.; MAHDY, M. R. C.; ALAM, T. N.; QUADERI, G. D. A.; RAHMAN, M. A. Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model. **Physica A: Statistical Mechanics and its Applications**, Amsterdam, v. 555, n. 124444, 2020. Disponível em: https://doi.org/10.1016/j.physa.2020.124444. Acesso em: 12 de jul. 2020.

CHU, W. S.; TORRE, F. D. L.; COHN, J. F. Learning facial action units with spatiotemporal cues and multi-label sampling. **Image and Vision Computing**, Amsterdam, v. 81, p. 1-14, 2019. Disponível em: https://doi.org/10.1016/j.imavis.2018.10.002. Acesso em: 09 de jul. 2020.

COLLINGWOOD, L.; WILKERSON, J. Tradeoffs in accuracy and efficiency in supervised learning methods. **Journal of Information Technology & Politics**, Philadelphia, v. 9, p. 298-318, 2012. Disponível em: https://doi.org/10.1080/19331681.2012.669191. Acesso em: 26 de fev. 2022.

COSTA, N. L.; ALVES, M. S.; RODRIGUES, N. S.; BANDEIRA, C. M.; ALVES, M. G. O.; MENDES, M. A.; ALVES, L. A. C.; ALMEIDA, J. D.; BARBOSA, R. Finding the combination of multiple biomarkers to diagnose oral squamous cell carcinoma - A data mining approach. **Computers in Biology and Medicine**, Amsterdam, v. 143, n. 105296. Disponível em: https://doi.org/10.1016/j.compbiomed.2022.105296. Acesso em: 13 de fev. 2020.

CUDZIK, A.; BRENNENSTHUL, M.; BIAŁCZYK, W.; CZARNECKI, J. Tractive performance of tyres in forest conditions – impact assessment of ground and tyres parameters. **Croatian Journal of Forest Engineering**, Zagreb, v. 39, n. 1, p. 85-96, 2018. Disponível em: http://www.crojfe.com/site/assets/files/4196/cudzik.pdf. Acesso em: 18 de jul. 2020.

CUI, S.; YIN, Y.; WANG, D.; LI, Z.; WANG, Y. A stacking-based ensemble learning method for earthquake casualty prediction. **Applied Soft Computing**, Amsterdam, v. 101, In Press, 2021. Disponível em: https://doi.org/10.1016/j.asoc.2020.107038. Acesso em: 27 de dez. 2021.

CUNHA, G. L.; FERNANDES, R. A. S.; FERNANDES, T. C. C. Small-signal stability analysis in smart grids: An approach based on distributed decision trees. **Electric Power Systems Research**, Amsterdam, v. 203, n. 107651, 2022. Disponível em: https://doi.org/10.1016/j.epsr.2021.107651. Acesso em: 13 de fev. 2022.

ČUKIĆ, M.; STOKIĆ, M.; SIMIĆ, S.; POKRAJAC, D. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. **Cognitive Neurodynamics**, Berlin, v. 14, p 443-455, 2020. Disponível em: https://doi.org/10.1007/s11571-020-09581-x. Acesso em: 12 de jul. 2020.

DAS, A.; KHAN, M. N.; AHMED, M. M. Detecting lane change maneuvers using SHRP2 naturalistic driving data: A comparative study machine learning techniques. **Accident Analysis & Prevention**, Amsterdam, v. 142, n. 105578, 2020. Disponível em: https://doi.org/10.1016/j.aap.2020.105578. Acesso em: 13 de jul. 2020.

DAUGHERTY, B.; SESSIONS, J.; CRISTALES, R. Z.; WING, M. G. Improving large trailer access for biomass recovery in steep terrain**. Forest Science**, Bethesda, v. 64, n. 4, p. 429-441, 2018. Disponível em: https://doi.org/10.1093/forsci/fxx020. Acesso em: 18 de jul. 2020.

DERRAC, J.; CHICLANA, F.; GARCÍA, S.; HERRERA, F.; Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets. **Information Sciences**, Amsterdam, v. 329, p. 144-163, 2016. Disponível em: https://doi.org/10.1016/j.ins.2015.09.007. Acesso em: 03 de jan. 2021.

DIVINA, F.; GILSON, A.; GOMÉZ-VELA, F.; TORRES, M. G.; TORRES, J. F. Stacking ensemble learning for short-term electricity consumption forecasting. **Energies**, Basel, v. 11, n. 4, p. 949. Disponível em: https://doi.org/10.3390/en11040949. Acesso em: 27 de dez. 2021.

DORNHEIM, J.; LINK, N.; GUMBSCH, P. Model-free adaptive optimal control of episodic fixed-horizon manufacturing processes using reinforcement learning. **International Journal of Control, Automation and Systems**, Seoul, v. 18, p. 1593-1604, 2020. Disponível em: http://dx.doi.org/10.1007/s12555-019-0120-7. Acesso em: 10 de jul. 2020.

DUN, L.; SOROKA, S.; WLEZIEN, C. Dictionaries, supervised learning, and media coverage of public policy. **Political Communication**, Philadelphia, v. 38, p. 140-158. Disponível em: https://doi.org/10.1080/10584609.2020.1763529. Acesso em: 26 de fev. 2022.

ELITH, J.; LEATHWICK, J. R.; HASTIE, T. A working guide to boosted regression trees. **The Journal of Animal Ecology**, London, v. 77, n. 4, p. 802-813, 2008. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1111/j.1365-2656.2008.01390.x. Acesso em: 14 de jul. 2020.

ENDEL, F.; PIRINGER, H. Data wrangling: making data useful again. **IFAC-PapersOnLine**, Amsterdam, v. 48-1, n. 1, p. 111-112, 2015. Disponível em:

https://www.sciencedirect.com/science/article/pii/S2405896315001986#!. Acesso em 29 de nov. 2021.

ESCRIGA, J.; WOOLLEY, E.; SIMEONE, A.; WATSONA, N. J. Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning. **Food Control**, Amsterdam, v. 116, n. 107309, 2020. Disponível em: https://doi.org/10.1016/j.foodcont.2020.107309. Acesso em 08 de jul. 2020.

EYVINDSON, K.; SAAD, R.; ERIKSSON, L. O. Incorporating stand level risk management options into forest decision support systems. **Forest Systems**, Madrid, v. 26, p. e013, 2017. Disponível em: https://revistas.inia.es/index.php/fs/article/view/10445. Acesso em: 20 de dez. 2020.

EZZATI, S.; NAJAFI, A.; BETTINGER, P. Finding feasible harvest zones in mountainous areas using integrated spatial multi-criteria decision analysis. **Land Use Policy**, Amsterdam, v. 59, p. 478-491, 2016. Disponível em: https://doi.org/10.1016/j.landusepol.2016.09.020. Acesso em: 18 de jul. 2020.

FAIZ, M. F.; ARSHAD, J.; ALAZAB, M.; SHALAGINOV, A. Predicting likelihood of legitimate data loss in email DLP. **Future Generation Computer Systems**, Amsterdam, v. 110, p. 744-757, 2020. Disponível em: https://doi.org/10.1016/j.future.2019.11.004. Acesso em: 12 de jul. 2020.

FANG, Z.; WANG, Y.; PENG, L.; HONG, H. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. **International Journal of Geographical Information Science**, Abingdon, v. 35, n. 2, p. 321-347,2021. Disponível em: https://doi.org/10.1080/13658816.2020.1808897. Acesso em: 23 de dez. 2021.

FAYED, H. A.; ATIYA, A. F. A novel template reduction approach for the -nearest neighbor met. **IEEE Transactions on Neural Networks**, Piscataway, v. 20, n. 5, p. 890-896, 2009. Disponível em: https://doi.org/10.1109/tnn.2009.2018547. Acesso em: 03 de jan. 2021.

FERNANDES, H. C.; BURLA, E. R.; LEITE, E. DA S.; MINETTE, L. J. Avaliação técnica e econômica de um “Harvester” em diferentes condições de terreno e produtividade da Floresta. **Scientia Forestalis**, Piracicaba, v. 41, n. 97, p. 145-151, mar. 2013. Disponível em: https://www.ipef.br/publicacoes/scientia/nr97/cap15.pdf. Acesso em: 05 de jan. 2022.

FERREIRA, V. J. E.; COSTA, C. H. S.; DE MIRANDA, R. M.; FIGUEIREDO, A. F. The use of the k nearest neighbor method to classify the representative elements. **Educación Química**, Cidade do México, v. 26, n. 3, p. 195-201. Disponível em: https://doi.org/10.1016/j.eq.2015.05.004. Acesso em: 03 de jan. 2021.

FONSECA-GALINDO, J. C.; SURITA, G. C.; MAIA NETO, J.; CASTRO, C. L.; LEMOS, A. P. A multi-agent system for solving the dynamic capacitated vehicle routing problem with stochastic customers using trajectory data mining. **Expert Systems with Applications**, Amsterdam, n. 116602, 2022. Disponível em: https://doi.org/10.1016/j.eswa.2022.116602. Acesso em: 13 de fev. 2022.

FRANK, S. M.; QI, A.; RAVASIO, D.; SASAKI, Y.; ROSEN, E. L.; WATANABE, T. Supervised learning occurs in visual perceptual learning of complex natural images. **Current Biology**, Cambridge, In Press, 2020. Disponível em: https://doi.org/10.1016/j.cub.2020.05.050. Acesso em: 09 de jul. 2020.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, Beachwood, v. 29, p. 1189-1232, 2001. Disponível em: https://www.jstor.org/stable/2699986. Acesso em: 12 de jul. 2020.

FU, L.; KARA, L. B. Neural network-based symbol recognition using a few labeled samples. **Computers & Graphics**, Amsterdam, v. 35, n. 5, p. 955-966, 2011. Disponível em: https://doi.org/10.1016/j.cag.2011.07.001. Acesso em: 09 de jul. 2020.

FURCHE, T.; GOTTLOB, G.; LIBKIN, L.; ORSI, G.; PATON, N. W. Data wrangling for big data: challenges and opportunities. In: International Conference on Extending Database Technology, 19., 2016, Bordeaux. **Proceedings of the 19th International Conference on Extending Database Technology. Konstanz**: OpenProceedings.org, 2016. p. 473-478. Disponível em: https://doi.org/10.5441/002/edbt.2016.44. Acesso em: 29 nov. 2021.

GAO, W.; SU, C. Analysis on block chain financial transaction under artificial neural network of deep learning. **Journal of Computational and Applied Mathematics**, Amsterdam, v. 380, n. 112991, 2020. Disponível em: https://doi.org/10.1016/j.cam.2020.112991. Acesso em: 08 de jul. 2020.

GARRETT, L. G.; BEETS, P. N.; CLINTON, P. W.; SMAILL, S. J. National series of long-term intensive harvesting trials in Pinus radiata stands in New Zealand: Initial biomass, carbon and nutrient pool data. **Data in Brief**, Amsterdam, v. 27, n. 104757, 2019. Disponível em: https://doi.org/10.1016/j.dib.2019.104757. Acesso em: 18 de jul. 2020.

GHORBANI, F.; AFSHARCHI, M.; DERHAMI, V. learning with policy prediction in continuous state-action multi-agent decision processes. **Soft Computing**, Berlin, v. 24, p. 901-918, 2020. Disponível em: https://doi.org/10.1007/s00500-019-04600-4. Acesso em: 11 de jul. 2020.

GILBERT, D.; MATEU, C.; PLANES, J. HYDRA: A multimodal deep learning framework for malware classification. **Computers & Security**, Amsterdam, v. 95, n. 101873, 2020. Disponível em: https://doi.org/10.1016/j.cose.2020.101873. Acesso em: 13 de jul. 2020.

GLATT, R.; SILVA, F. L.; BIANCHI, R. A. C.; COSTA, A. H. R. DECAF: Deep case-based policy inference for knowledge transfer in reinforcement learning. **Expert Systems with Applications**, Amsterdam, v. 156, n. 113420, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113420. Acesso em: 10 de jul. 2020.

GUENTHER, N.; SCHONLAU, M. Support vector machines. **The Stata Journal**, College Station, v. 16, n, 4, p. 917-937, 2016. Disponível em:

https://journals.sagepub.com/doi/pdf/10.1177/1536867X1601600407. Acesso em: 03 de jan. 2022.

GUMUSKAYA, V.; JAARSVELD, W. V.; DIJKMAN, R.; GREFEN, P.; VEENSTRA, A. Integrating stochastic programs and decision trees in capacitated barge planning with uncertain container arrivals. **Transportation Research Part C: Emerging Technologies**, Amsterdam, v. 132, n. 103383, 2021. Disponível em: https://doi.org/10.1016/j.trc.2021.103383. Acesso em: 13 de fev. 2022.

HAAS, J.; KIRCHNER, H. S.; LANG, F. Modeling soil erosion after mechanized logging operations on steep terrain in the Northern Black Forest, Germany. **European Journal of Forest Research**, Berlin, v. 139, p. 549-565, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s10342-020-01269-5. Acesso em: 18 de jul. 2020.

HAKIMPOOR, H.; ARSHAD, K. A. B. A.; TAT, H. H.; KHANI, N.; RAHMANDOUST, M. Artificial neural networks' applications in management. **World Applied Sciences Journal**, Faisalabad, v. 14, n. 7, p. 1008-1019, 2011. Disponível em: https://www.idosi.org/wasj/wasj14(7)2011.htm. Acesso em: 05 de jan. 2022.

HALLBERG, D.; SALIMI, N. Qualitative and quantitative analysis of definitions of e-health and m-health. **Healthcare Informatics Research**, Seoul, v. 6, n. 2, p. 119-128, 2020. Disponível em: https://doi.org/10.4258/hir.2020.26.2.119. Acesso em: 10 de jul. 2020.

HAN, J.; KAMBER, M.; PEI, J*. Data mining: concepts and techniques*. 3. ed. Amsterdam: Elsevier, 2012.

HANSRAJH, A.; ADELIYI, T. T.; WING, J. Detection of online fake news using blending ensemble learning. **Scientific Programming**, Amsterdam, v. 2021, ID 3434458, p. 2021, 2021. Disponível em: https://doi.org/10.1155/2021/3434458. Acesso em: 23 de dez. 2021.

HAO, M.; CAO, W.; LIU, Z.; WU, M.; XIAO, P. Visual-audio emotion recognition based on multi-task and ensemble learning with multiple features. **Neurocomputing**, Amsterdam, v. 391, p. 42-51, 2020. Disponível em: https://doi.org/10.1016/j.neucom.2020.01.048. Acesso em: 23 de dez. 2021.

HARARUK, O.; KURZ, W. A.; DIDION, M. Dynamics of dead wood decay in Swiss forests. **Forest Ecosystems**, London, v. 7, n. 36, p. 1-16. Disponível em: https://doi.org/10.1186/s40663-020-00248-x. Acesso em: 14 de jul. 2020.

HAWBAKER, T. J.; VANDERHOOF, M. K.; SCHMIDT, G. L.; BEAL, Y. J.; PICOTTE, J. J.; TAKACS, J. D.; FALGOUT, J. T.; DWYER, J. L. The Landsat Burned Area algorithm and products for the conterminous United States. **Remote Sensing of Environment**, Amsterdam, v. 244, n. 111801, 2020. Disponível em: https://doi.org/10.1016/j.rse.2020.111801. Acesso em: 14 de jul. 2020.

HE, Y.; CHU, Y.; SONG, Y.; LIU, M.; SHI, S.; CHEN, X. Analysis of design strategy of energy efficient buildings based on databases by using data mining and statistical

metrics approach. **Energy and Buildings**, Amsterdam, v. 258, n. 111811, 2022. Disponível em: https://doi.org/10.1016/j.enbuild.2021.111811. Acesso em: 13 de fev. 2022.

HE, Z.; HU, J.; WANG, Y. Low-rank tensor learning for classification of hyperspectral image with limited labeled samples. **Signal Processing**, Amsterdam, v. 145, p. 12-25, 2018. Disponível em: https://doi.org/10.1016/j.sigpro.2017.11.007. Acesso em: 09 de jul. 2020.

HE, W.; WANG, Y.; ZHOU, M.; WANG, B. A novel parameters correction and multivariable decision tree method for edge computing enabled HGR system. **Neurocomputing**, Amsterdam, In Press, 2021. Disponível em: https://doi.org/10.1016/j.neucom.2021.08.147. Acesso em: 13 de fev. 2022.

HINO, H.; KOSHIJIMA, K.; MURATA, N. Non-parametric entropy estimators based on simple linear regression. **Computational Statistics & Data Analysis**, Amsterdam, v. 89, p. 72-84, 2015. Disponível em: https://doi.org/10.1016/j.csda.2015.03.011. Acesso em: 30 de dez. 2021.

HITCHNER, S.; GOYKE, N.; THOMAS, M.; SCHELHAS, J.; DWIVEDI, P. Beyond the Math: Case Studies of Black Forest Landowners in Georgia, United States. **Journal of Sustainable Forestry**, London, 2021. Disponível em: https://doi.org/10.1080/10549811.2021.2010573. Acesso em: 03 de mar. 2022.

HOANG, M. T.; ZHU, Y.; YUEN, B.; REESE, T.; DONG, X.; LU, T.; WESTENDORP, R.; XIE, M. Soft range limited k-nearest neighbours algorithm for indoor localization enhancement. **IEEE Sensors Journal**, New York, v. 18, n. 24, p. 1-1, 2018. Disponível em: https://doi.org/10.1109/JSEN.2018.2874453. Acesso em: 03 de jan. 2021.

HOLZLEITNER, F.; LANGMAIER, M.; HOCHBICHLER, E.; OBERMAYER, B.; STAMPFER, K.; KANZIAN, C. Effect of prior tree marking, thinning method and topping diameter on harvester performance in a first thinning operation – a field experiment. **Silva Fennica**, Helsinki, v. 53, n. 3, p. e10178, 2019. Disponível em: https://doi.org/10.14214/sf.10178. Acesso em: 05 de nov. 2024.

HUANG, X. Application analysis of AI reasoning engine in microblog culture industry. **Personal and Ubiquitous Computing**, London, v. 24, p. 393-403, 2020. Disponível em: https://doi.org/10.1007/s00779-019-01338-6. Acesso em: 10 de jul. 2020.

HULTQVIST, D.; OLSSON, L. Optimization of raw material procurement at pulp or paper mills – the influence of weather-related risks**. International Journal of Systems Science**, London, v. 37, n. 4, p. 253-269, 2006. Disponível em: https://doi.org/10.1080/00207720600568194. Acesso em: 03 de ago. 2020.

HUNT, I. In-sample tests of predictability are superior to pseudo-out-of-sample tests, even when data mining. **International Journal of Forecasting**, Amsterdam, In Press, 2021. Disponível em: https://doi.org/10.1016/j.ijforecast.2021.05.006. Acesso em: 13 de fev. 2022.

HUNT, L.; HAIDER, W. Fair and effective decision making in forest management planning. **Society & Natural Resources**, London, v. 14, n. 10, p. 873-887, 2001. Disponível em: https://doi.org/10.1080/089419201753242788. Acesso em: 21 de dez. 2021.

HUSSAIN, F.; BASU, S.; HENG, J. J. H.; LOO, L. H.; ZINK, D. Predicting direct hepatocyte toxicity in humans by combining high-throughput imaging of HepaRG cells and machine learning-based phenotypic profiling. **Archives of Toxicology**, Berlin. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s00204-020-02778-3. Acesso em: 11 de jul. 2020.

HWANG, W. Y.; JUN, C. H. Supervised learning-based collaborative filtering using market basket data for the cold-start problem. **Industrial Engineering & Management Systems**, Seoul, v. 13, n. 4, p. 421-431, 2014. Disponível em: https://doi.org/10.7232/iems.2014.13.4.421. Acesso em: 10 de jul. 2020.

IGLESIAS, S.P.; MIYAZAKI, M.R.; MARIANO, A.P.; FRANCO, T.T. Techno-economic assessment of bio-oil produced from Eucalyptus forestry residues, **Industrial Crops & Products**, Amsterdam, v. 171, p.1-10, 2021. Disponível em: https://doi.org/10.1016/j.indcrop.2021.113936. Acesso em: 10 de jul. 2020.

ILIC, I.; GÖRGÜLÜ, B.; CEVIK, M.; BAYDOĞAN, M. G. Explainable boosted linear regression for time series forecasting. **Pattern Recognition**, Amsterdam, v. 120, In Press, 2021. Disponível em: https://doi.org/10.1016/j.patcog.2021.108144. Acesso em: 30 de dez. 2021.

ILINTSEV, A.; NAKVASINA, E.; ALEYNIKOV, A.; TRETYAKOV, S.; KOPTEV, S.; BOGDANOV, A. Middle-Term Changes in Topsoils Properties on skidding trails and cutting strips after long-gradual cutting: a case study in the boreal forest of the North-East of Russia. **Croatian Journal of Forest Engineering**, Zagreb, v. 39, n. 1, p. 71-83, 2018. Disponível em: http://www.crojfe.com/site/assets/files/4189/ilintsev.pdf. Acesso em: 18 de jul. 2020.

IRANITALAB, A.; KHATTAK, A. Probabilistic classification of hazardous materials release events in train incidents and cargo tank truck crashes. **Reliability Engineering & System Safety**, London, v. 199, n. 106914, 2020. Disponível em: https://doi.org/10.1016/j.ress.2020.106914. Acesso em: 03 de ago. 2020.

JABER, M. M.; ABD, S. K.; SHAKEEL, P. M.; BURHANUDDIN, M. A.; MOHAMMED, M. A.; YUSSOF, S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. **Measurement**, Amsterdam, v. 162, In Press, 2020. Disponível em: https://doi.org/10.1016/j.measurement.2020.107883. Acesso em: 13 de jul. 2020.

JAEGHER, B. D.; LARUMBE, E.; SCHEPPER, W. D.; VERLIEFDE, A.; NOPENS, I. Colloidal fouling in electrodialysis: A neural differential equations model. **Separation and Purification Technology**, Amsterdam, v. 249, n. 116939, 2020. Disponível em: https://doi.org/10.1016/j.seppur.2020.116939. Acesso em: 08 de jul. 2020.

JAFARI, H.; RAJAEE, T.; KISI, O. Improved water quality prediction with hybrid wavelet-genetic programming model and shannon entropy**. Natural Resources Research**, New York. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s11053-020-09702-7. Acesso em: 12 de jul. 2020.

JAYASURIYA, M. T.; GERMAIN, R. H.; WAGNER, J. E. Protecting timberland RMZs through carbon markets: A protocol for riparian carbon offsets. **Forest Policy and Economics**, Amsterdam, v. 111, n. 102084, 2020. Disponível em: https://doi.org/10.1016/j.forpol.2019.102084. Acesso em: 18 de jul. 2020.

JIA, X.; WEN, T.; DING, W.; LI, H.; LI, W. Semi-supervised label distribution learning via projection graph embedding. **Information Sciences**, Amsterdam, v. 581, p. 840-855, 2021. Disponível em: https://doi.org/10.1016/j.ins.2021.10.009. Acesso em: 10 de fev. 2022.

JIANG, L.; CAI, Z.; WANG, D.; JIANG, S. Survey of improving k-nearest-neighbor for classification. *In*: Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 4, 2007, Haikou. **Conference** […] China, 2007, p. 679-683. Disponível em: https://doi.org/10.1109/FSKD.2007.552. Acesso em: 30 de dez. 2021.

KAMATH, P.; PATIL, P.; SUSHMA, S. S.; SOWMYA, S. Crop yield forecasting using data mining. **Global Transitions Proceedings**, China, v. 2, p. 402-407, 2021. Disponível em: https://doi.org/10.1016/j.gltp.2021.08.008. Acesso em: 13 de fev. 2022.

KANDEL, S.; HEER, J.; PLAISANT, C.; KENNEDY, J.; VAN HAM, F.; RICHE, N. H.; WEAVER, C.; LEE, B.; BRODBECK, D.; BUONO, P. Research directions in data wrangling: Visualizations and transformations for usable and credible data. **Information Visualization**, London, v. 10, n. 4, p. 271-288, 2011. Disponível em: https://doi.org/10.1177/1473871611415994. Acesso em: 29 de nov. 2021.

KANG, C.; HUANG, J.; ZHANG, Z.; LIU, Q.; XIANG, W.; ZHAO, Z.; LIU, X.; CHONG, L. An automatic algorithm of identifying vulnerable spots of internet data center power systems based on reinforcement learning. **International Journal of Electrical Power & Energy Systems**, London, v. 121, n. 106145, 2020. Disponível em: https://doi.org/10.1016/j.ijepes.2020.106145. Acesso em: 10 de jul. 2020.

KÄRHÄ, K.; RÖNKKÖ, E.; GUMSE, S.-I. Productivity and cutting costs of thinning harvesters**. International Journal of Forest Engineering**, London, v. 15, n. 2, 2004. Disponível em: https://doi.org/10.1080/14942119.2004.10702496. Acesso em: 05 de nov. 2024.

KARTTUNEN, K.; LAITILA, J. Forest management regime options for integrated small-diameter wood harvesting and supply chain from young Scots pine (Pinus sylvestris L.) stands. **International Journal of Forest Engineering**, London, v. 26, n. 2, p.124-138, 2015. Disponível em: https://doi.org/10.1080/14942119.2015.1083749. Acesso em: 03 de mar. 2022.

KHAN, J.; LEE, Y. K. LeSSA: A unified framework based on lexicons and semi-supervised learning approaches for textual sentiment classification. **Applied Sciences**, Basel, v. 9, n. 24, p. 5562-5590, 2019. Disponível em: https://doi.org/10.3390/app9245562. Acesso em: 11 de jul. 2020.

KHOSRAVI, K.; COOPER, J. R.; DAGGUPATI, P.; PHAM, B. T.; BUI, D. T.; Bedload transport rate prediction: Application of novel hybrid data mining techniques. **Journal of Hydrology**, Amsterdam, v. 585, n. 124774, 2020. Disponível em: https://doi.org/10.1016/j.jhydrol.2020.124774. Acesso em: 09 de jul. 2020.

KIM, T. J.; LEE, J. S.; KIM, J. W.; OH, M. S.; MO, H.; LEE, C. H.; JEONG, H. Y.; JUNG, K. H.; LIM, J. S.; KO, S. B.; YU, K. H.; LEE, B. C.; YOON, B. W. Building linked big data for stroke in Korea: linkage of stroke registry and national health insurance claims data. **Journal of korean Medical Science**, Seoul, v. 33, n. 53, e. 343, 2018. Disponível em: https://doi.org/10.3346/jkms.2018.33.e343. Acesso em: 10 de jul. 2020.

KOEHLER, M.; BOGATU, A.; CIVILI, C.; KONSTANTINOU, N.; ABEL, E.; FERNANDES, A. A. A.; KEANE, J.; LIBKIN, L.; PATON, N. W. Data context informed data wrangling. *In*: IEEE International Conference on Big Data (Big Data), 2017, Boston. **Conference** […] MA, USA: IEEE, 2017. p. 956-963. Disponível em: https://doi.org/ 10.1109/BigData.2017.8258015. Acesso em: 04 de jan. 2022.

KONSTANTINOU, N.; KOEHLER, M.; ABEL, E.; CIVILI, C.; NEUMAYR, B.; SALLINGER, E.; FERNANDES, A. A. A.; GOTTLOB, G.; KEANE, J. A.; LIBKIN, L.; PATON, N. W. The VADA Architecture for Cost-Effective Data Wrangling. *In*: International Conference on Management of Data, 2017, Chicago. **Conference** […] IIlinois, USA: SIGMOD'17, 2017. p. 1599-1602. Disponível em: https://doi.org/10.1145/3035918.3058730. Acesso em: 04 de jan. 2022.

KRAMER O. K-Nearest Neighbors. *In: Dimensionality reduction with unsupervised nearest neighbors*. intelligent systems reference library, Springer, Berlin: Heidelberg, 2013. cap. 11, p. 13-23. Disponível em: https://doi.org/10.1007/978-3-642-38652-7_2. Acesso em: 30 de dez. 2021.

KROGH, A. What are artificial neural networks? **Nature Biotechnology**, London States, v. 26, n. 2, p. 195-197, 2008. Disponível em: https://doi.org/10.1038/nbt1386. Acesso em: 05 de jan. 2022.

KUBAT, M. Artificial Neural Networks. *In*: *An introduction to machine learning*. Cham: Springer, 2021. p. 117-143. Disponível em: https://doi.org/10.1007/978-3-030-81935-4_6. Acesso em: 05 de jan. 2022.

KUDYBA, S. A hybrid analytic approach for understanding patient demand for mental health services. **Network Modeling Analysis in Health Informatics and Bioinformatics**, Berlin, v. 7, n. 3, p. 1-8, 2018. Disponível em: https://doi.org/10.1007/s13721-018-0164-2. Acesso em: 10 de jul. 2020.

KUHN, M.; WICKHAM, H. **recipes: Preprocessing Tools to Create Design Matrices**. R package version 0.1.13, 2020. Disponível em: https://CRAN.R-project.org/package=recipes. Acesso em: 10 de ago. 2020.

KUMAR, N.; VENUGOPAL, D.; QIU, L.; KUMAR, S. Detecting review manipulation on online platforms with hierarchical supervised learning. **Journal of Management Information Systems**, London, v. 35, p. 350-380, 2018. Disponível em: https://doi.org/10.1080/07421222.2018.1440758. Acesso em: 26 de fev. 2022.

KUMARI, K.; YADAV, S. Linear regression analysis study. **Journal of the Practice of Cardiovascular Sciences**, Philadelphia, v. 4, n. 1, p. 33-36, 2018. Disponível em: https://doi.org/ 10.4103/jpcs.jpcs_8_18. Acesso em: 29 de dez. 2021.

KUSHWAH, J. S.; KUMAR, A.; PATEL, S.; SONI, R.; GAWANDE, A.; GUPTA, S. Comparative study of regressor and classifier with decision tree using modern tools. **Materials Today: Proceedings**, Amsterdam, 2021. Disponível em: https://doi.org/10.1016/j.matpr.2021.11.635. Acesso em: 13 de fev. 2022.

LEE, J.; KIM, J.; KO, W. Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method. **Applied Sciences**, Basel, v. 9, p. 1231, 2019. Disponível em: https://doi.org/10.3390/app9061231. Acesso em: 27 de dez. 2021.

LEITE, E. D. A S.; MINETTE, L. J.; FERNANDES, H. C.; SOUZA, A. P. DE; AMARAL, E. J. DO; LACERDA, E. DAS G. Desempenho do *harvester* na colheita de eucalipto em diferentes espaçamentos e declividades. **Revista Árvore**, Viçosa, v. 38, n. 1, p. 000-000, 2014. Disponível em: https://www.scielo.br/j/rarv/a/NtsLy4Ns7WhgnkHZJ86V4Gg/. Acesso em: 05 de jan. 2022.

LI, D.; LIANG, W.; FENG, X.; RUAN, T.; JIANG, G. Recent advances in data-mining techniques for measuring transformation products by high-resolution mass spectrometry. **TrAC Trends in Analytical Chemistry**, Amsterdam, v. 143, n.116409, 2021. Disponível em: https://doi.org/10.1016/j.trac.2021.116409. Acesso em: 13 de fev. 2022.

LI, F.; RUNGER, G. C.; TUV, E. Supervised learning for change-point detection. **International Journal of Production Research**, London, v. 44, p. 2853-2868, 2007. Disponível em: https://doi.org/10.1080/00207540600669846. Acesso em: 26 de fev. 2022.

LI, G.; FANG, S.; MA, J.; CHENG, J. Modeling merging acceleration and deceleration behavior based on gradient-boosting decision tree. **Journal of Transportation Engineering**, Part A: Systems, Reston, v. 146, n. 7, 2020. Disponível em: https://doi.org/10.1061/JTEPBS.0000386. Acesso em: 13 de jul. 2020.

LI, H.; LI, X.; LIU, X.; BU, X.; LI, H.; LYU, Q. Prediction of blast furnace parameters using feature engineering and Stacking algorithm. **Ironmaking & Steelmaking**, London, 2021. Disponível em: https://doi.org/10.1080/03019233.2021.1992816. Acesso em: 29 de dez. 2021.

LI, X.; LI, X.; MA, H. Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. **Mechanical Systems and Signal Processing**, London, v. 143, n. 106825, 2020. Disponível em: https://doi.org/10.1016/j.ymssp.2020.106825. Acesso em: 10 de jul. 2020.

LIANG, X.; KANKARE, V.; HYYPPÄ, J.; WANG, Y.; KUKKO, A.; HAGGRÉN, H.; YU, X.; KAARTINEN, H.; JAAKKOLA, A., GUAN, F.; HOLOPAINEN, M.; VASTARANTA, M. Terrestrial laser scanning in forest inventories. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 115, p. 63-77, 2016. Disponível em: https://www.sciencedirect.com/science/article/pii/S0924271616000204#!. Acesso em: 20 de dez. 2021.

LIANG, W.; LUO, S.; ZHAO, G.; WU, H. Predicting hard rock pillar stability using gbdt, xgboost, and lightgbm algorithms. **Mathematics**, Basel, v. 8, n. 5, p. 765- , 2020. Disponível em: https://doi.org/10.3390/math8050765. Acesso em: 12 de jul. 2020.

LING, W.; MEI, F. D. Estimation of missing values using a weighted k-nearest neighbors algorithm. **IEEE International Conference on Environmental Science and Information Application Technology**, Wuhan, China, p. 660-663. Disponível em: https://doi.org/10.1109/esiat.2009.206. Acesso em: 03 de jan. 2020.

LIU, J.; GU, J.; LI, H.; CARLSON, K. H. Machine learning and transport simulations for groundwater anomaly detection. **Journal of Computational and Applied Mathematics**, Amsterdam, v. 380, n. 112982, 2020a. Disponível em: https://doi.org/10.1016/j.cam.2020.112982. Acesso em: 08 de jul. 2020.

LIU, J.; KADZIŃSKI, M.; LIAO, X.; MAO, X.; WANG, Y. A preference learning framework for multiple criteria sorting with diverse additive value models and valued assignment examples. **European Journal of Operational Research**, Amsterdam, v. 286, n. 3, p. 963-985, 2020b. Disponível em: https://doi.org/10.1016/j.ejor.2020.04.013. Acesso em: 08 de jul. 2020.

LIU, Y.; XIANG, C. Hybrid learning network: a novel architecture for fast learning. **Procedia Computer Science**, Amsterdam, v. 122, p. 622-628, 2017. Disponível em: https://doi.org/10.1016/j.procs.2017.11.415. Acesso em: 09 de jul. 2020.

LIU, X.; LIU, T.; FENG, P. Long-term performance prediction framework based on XGBoost decision tree for pultruded FRP composites exposed to water, humidity and alkaline solution. **Composite Structures**, Amsterdam, v. 248, n. 115184, 2022. Disponível em: https://doi.org/10.1016/j.compstruct.2022.115184. Acesso em: 13 de fev. 2022.

LOFTUS, S. C. **Basic Statistics with R**: Reaching Decisions with Data. Chapter 19 - Simple linear regression. Amsterdam: Elsevier, 2022. Disponível em: https://doi.org/10.1016/B978-0-12-820788-8.00032-8. Acesso em: 30 de dez. 2021.

LOGAN, A. T.; GOODE, J. D.; KEELLINGS, D. J.; HART, J. L. Microsite Influence on Woody Plant Regeneration in a Pinus palustris Woodland Following Catastrophic

Disturbance. **Forests**, Basel, v. 11, n. 5, p. 588-607, 2020. Disponível em: https://doi.org/10.3390/f11050588. Acesso em 06 de jul. 2020.

LOPES, E. S.; OLIVEIRA, D.; RODRIGUES, C. K.; DRINKO, C. H. Variables influencing working time and skidder productivity in wood extraction. **Nativa**, Sinop, v. 5, n. 4, p. 298-302, 2017. Disponível em: https://doi.org/10.31413/nativa.v5i4.4600. Acesso em: 18 de jul. 2020.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às *support vector machines*. **Revista de Informática Teórica e Aplicada**, Rio Grande do Sul, v. 14, n. 2, p. 43-67, 2007. Disponível em: https://doi.org/10.22456/2175-2745.5690. Acesso em: 03 de jan. 2022.

LU, S.; HWANG, Y.; KHABIBRAKHMANOV, I.; MARIANNO, F. J.; SHAO, X.; ZHANG, J.; HODGE, B.; HAMANN, H. F. Machine learning based multi-physical-model blending for enhancing renewable energy forecast: improvement via situation dependent error correction. In: **EUROPEAN CONTROL CONFERENCE (ECC)**, 2015, Linz. Proceedings… Linz: IEEE, 2015. Disponível em: https://doi.org/10.1109/ECC.2015.7330558. Acesso em: 23 dez. 2021.

LUBIS, A. R.; LUBIS, M.; AL-KHOWARIZMI. Optimization of distance formula in K-Nearest Neighbor method. **Bulletin of Electrical Engineering and Informatics**, Yogyakarta, v. 9, n. 1, p. 326-338. Disponível em: https://doi.org/10.11591/eei.v9i1.1464. Acesso em: 30 de dez. 2021.

LUDBROOK, J. Linear regression analysis for comparing two measurers or methods ofmeasurement: But which regression? **Clinical and Experimental Pharmacology and Physiology**, Hoboken, v.37, p. 692-699, 2010. Disponível em: https://doi.org/10.1111/j.1440-1681.2010.05376.x. Acesso em: 29 de dez. 2021.

LUO, Y.; JI, R.; GUAN, T.; YU, J.; LIU, P.; YANG, Y. Every node counts: Self-ensembling graph convolutional networks for semi-supervised learning. **Pattern Recognition**, Amsterdam, v. 106, n. 107451, 2020. Disponível em: https://doi.org/10.1016/j.patcog.2020.107451. Acesso em: 10 de jul. 2020.

MA, S.; FILDES, R. Forecasting third-party mobile payments with implications for customer flow prediction. **International Journal of Forecasting**, Amsterdam, v. 36, n. 3, p. 739-760, 2020. Disponível em: https://doi.org/10.1016/j.ijforecast.2019.08.012. Acesso em: 13 de jul. 2020.

MA, F.; SUN, T.; LIU, L.; JING, H. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. **Future Generation Computer Systems**, Amsterdam, v. 111, p. 17-26, 2020. Disponível em: https://doi.org/10.1016/j.future.2020.04.036. Acesso em: 08 de jul. 2020.

MAIA, M. R. H.; PLASTINO, A.; PENNA, P. H. V. MineReduce: An approach based on data mining for problem size reduction. **Computers & Operations Research**, Amsterdam, v. 122, n. 104995, 2020. Disponível em: https://doi.org/10.1016/j.cor.2020.104995. Acesso em: 09 de jul. 2020.

MALLADI, K. T.; QUIRION-BLAIS, O.; SOWLATI, T. Development of a decision support tool for optimizing the short-term logistics of forest-based biomass. **Applied Energy**, Amsterdam, v. 216, p. 662-677, 2018. Disponível em: https://doi.org/10.1016/j.apenergy.2018.02.027. Acesso em: 03 de ago. 2020.

MALIK, F. A.; YE, W.; CHEN, Q.; LI, D. Recommendation algorithm based on blending learning. In**: HIGH PERFORMANCE COMPUTING AND CLUSTER TECHNOLOGIES CONFERENCE (HPCCT),** 3., 2019. *Proceedings…* New York: ACM, 2019. Disponível em: https://doi.org/10.1145/3341069.3342983. Acesso em: 23 dez. 2021.

MALOVRH, S. P.; KUMER, P.; GLAVONJIC, P.; NONIC, D.; NEDELJKOVIC, J.; KISIN, B.; AVDIBEGOVIC, M. Different organizational models of private forest owners as a possibility to increase wood mobilization in slovenia and serbia. **Croatian journal of forest engineering**, Zagreb, v. 38, n. 1, p. 127-140, 2017. Disponível em: http://www.crojfe.com/site/assets/files/4073/pezdevsek.pdf. Acesso em: 03 de ago. 2020.

MAMAN, Z. S.; CHEN, Y. J.; BAGHDADI, A.; LOMBARDO, S.; CAVUOTO, L. A.; MEGAHED, F. M. A data analytic framework for physical fatigue management using wearable sensors. **Expert Systems with Applications**, Amsterdam, v. 155, n. 113405, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113405. Acesso em: 13 de jul. 2020.

MANGALOVA, E.; AGAFONOV, E. Wind power forecasting using the k-nearest neighbors algorithm. **International Journal of Forecasting**, Amsterdam, v. 30, n. 2, p. 402-406, 2014. Disponível em: https://doi.org/10.1016/j.ijforecast.2013.07.008. Acesso em: 03 de jan. 2022.

MAMMONE, A.; TURCHI, M.; CRISTIANINI, N. Support vector machines. Wiley Interdisciplinary Reviews: **Computational Statistics**, Berlin, v. 1, n. 3, p. 283-289. Disponível em: https://doi.org/10.1002/wics.49. Acesso em: 03 de jan. 2022.

MAO, X.; XIA, L.; YANG, L.; YOU, Y.; LUO, P.; LI, Y.; WU, Y.; JIANG, G. Data mining of natural hazard biomarkers and metabolites with integrated metabolomic tools. **Journal of Hazardous Materials**, Amsterdam, v. 427, n. 127912, 2022. Disponível em: https://doi.org/10.1016/j.jhazmat.2021.127912. Acesso em: 13 de fev. 2022.

MARČETA, D.; PETKOVIĆ, V.; LJUBOJEVIĆ, D.; POTOČNIK, I. Harvesting system suitability as decision support in selection cutting forest management in Northwest Bosnia and Herzegovina. **Croatian Journal of Forest Engineering**, Zagreb, v. 41, n. 2, p. 251-265, 2020. Disponível em: https://doi.org/10.5552/crojfe.2020.744. Acesso em: 07 de abr. 2022.

MARCU, M. V.; BORZ, S. A. Effect of trail condition, slope, and direction of extraction on forwarding performance: Insights from a controlled comparative study. **Forests**, Basel, v. 15, n. 10, p. 1790, 2024. Disponível em: https://doi.org/10.3390/f15101790. Acesso em: 05 de nov. 2024.

MARTIN, M. R. L.; HUTTON, C. P.; SPARKS, T.; SNOWDEN, T.; CHRISTIE, B. R. Multiple object tracking scores predict post-concussion status years after mild traumatic brain injury. **Journal of Neurotrauma**, New Rochelle, p. 1-11. Disponível em: https://doi.org/10.1089/neu.2019.6842. Acesso em: 18 de jul. 2020.

MARZOUQ, M.; FADILI, H. E.; ZENKOUAR, K.; LAKHLIAI, Z.; AMOUZG, M. Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data. **Renewable Energy**, Amsterdam, v. 157, p. 214-231, 2020. Disponível em: https://doi.org/10.1016/j.renene.2020.04.133. Acesso em: 13 de jul. 2020.

MCCARTHY, M. J.; JESSEN, B.; BARRY, M. J.; FIGUEROA, M.; MCINTOSH, J.; MURRAY, T.; SCHMID, J.; KARGER, F. E. M. Mapping hurricane damage: A comparative analysis of satellite monitoring methods. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 91, n. 102134, 2020. Disponível em: https://doi.org/10.1016/j.jag.2020.102134. Acesso em: 10 de jul. 2020.

MCROBERTS, R. E.; TOMPPO, E. O. Remote sensing support for national forest inventories. **Remote Sensing of Environment**, Amsterdam, v. 110, n. 4, p. 412-419, 2007. Disponível em: https://doi.org/10.1016/j.rse.2006.09. Acesso em: 20 de dez. 2021.

MEDERSKI, P. S.; BORZ, S. A.; ĐUKA, A.; LAZDIŅŠ, A. challenges in forestry and forest engineering - case studies from four countries in east europe. **Croatian Journal of Forest Engineering**, Zagreb, v. 42, n. 1, p. 117-134, 2021. Disponível em: https://doi.org/10.5552/crojfe.2021.838. Acesso em: 07 de abr. 2022.

MEENACHI, L.; RAMAKRISHNAN, S. Differential evolution and ACO based global optimal feature selection with fuzzy rough set for cancer data classification. **Soft Computing**, Berlin, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s00500-020-05070-9. Acesso em: 18 de jul. 2020.

MEHARI, T.; STRODTHOFF, N. Self-supervised representation learning from 12-lead ECG data. **Computers in Biology and Medicine**, Amsterdam, v. 141, n. 105114, 2022. Disponível em: https://doi.org/10.1016/j.compbiomed.2021.105114. Acesso em: 10 de fev. 2022.

MELANDER, L.; EINOLA, K.; RITALA, R. Fusion of open forest data and machine fieldbus data for performance analysis of forest machines. **European Journal of Forest Research**, Berlin, v. 139, p. 213-227, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s10342-019-01237-8. Acesso em: 14 de jul. 2020.

MELANDER, L.; RITALA, R.; STRANDSTRÖM, M. Classifying soil stoniness based on the excavator boom vibration data in mounding operations. **Silva Fennica**, Helsinki, v. 53, n. 2, p. 1-21. Disponível em: https://doi.org/10.14214/sf.10068. Acesso em: 14 de jul. 2020.

MENDOZA, G. A.; PRABHU, R. Multiple criteria decision making approaches to assessing forest sustainability using criteria and indicators: a case study. **Forest Ecology and Management**, Amsterdam, v. 131, n. 1-3, p. 107-126, 2000. Disponível em: https://doi.org/10.1016/S0378-1127(99)00204-2. Acesso em: 21 de dez. 2021.

MHAISEN, N.; FETAIS, N.; ERBAD, A.; MOHAMED, A.; GUIZANI, M. To chain or not to chain: A reinforcement learning approach for blockchain-enabled IoT monitoring applications. **Future Generation Computer Systems**, Amsterdam, v. 111, p. 39-51, 2020. Disponível em: https://doi.org/10.1016/j.future.2020.04.035. Acesso em: 10 de jul. 2020.

MIRKOUEI, A.; MIRZAIE, P.; HAAPALA, K. R.; SESSIONS, J.; MURTHY, G. S. Reducing the cost and environmental impact of integrated fixed and mobile bio-oil refinery supply chains. **Journal of Cleaner Production**, Amsterdam, v. 113, p. 495-507, 2016. Disponível em: https://doi.org/10.1016/j.jclepro.2015.11.023. Acesso em: 03 de ago. 2020.

MITROSHIN, P.; SHITOVA, Y.; SHITOV, Y.; VLASOV, D.; MITROSHIN, A. Big data and data mining technologies application at road transport logistics. **Transportation Research Procedia**, Amsterdam, v. 61, p. 462-466, 2022. Disponível em: https://doi.org/10.1016/j.trpro.2022.01.075. Acesso em: 13 de fev. 2022.

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**, Zagreb, v. 41, n. 2, p. 219-229, 2020. Disponível em: http://www.crojfe.com/site/assets/files/4462/miyajima-1.pdf. Acesso em: 18 de jul. 2020.

MIYAJIMA, R. H.; PASSOS, J. R. S.; FENNER, P. T.; SIMÕES, D. Extração de eucalipto com grapple skidder: abordagem de produtividade operacional e custos de produção. **Scientia Forestalis**, Piracicaba, v. 48, n. 128, p. e3298, 2020. Disponível em: https://doi.org/10.18671/scifor.v48n128.14. Acesso em: 18 de jul. 2020.

MOGUERZA, J. M.; MUÑOZ, A. Support vector machines with applications. **Statistical Science**, Beachwood, v. 21, n. 3, p. 322-336. Disponível em: https://doi.org/10.1214/088342306000000493. Acesso em: 03 de jan. 2021.

MONDAL, S.; BARMAN, A. D. Speech activity detection using time-frequency auditory spectral pattern. **Applied Acoustics**, Amsterdam, v. 167, n. 107403, 2020. Disponível em: https://doi.org/10.1016/j.apacoust.2020.107403. Acesso em: 10 de jul. 2020.

MONTEIRO, S. T.; RIBEIRO, C. H. C. Desempenho de algoritmos de aprendizagem por reforço sob condições de ambiguidade sensorial em robótica móvel. **Sba Controle & Automação**, Campinas, v. 15 n. 3, p. 320-338, 2004. Disponível em: https://doi.org/10.1590/S0103-17592004000300008. Acesso em: 10 de jul. 2020.

MORETTE, N.; HEREDIA, L. C. C.; DITCHI, T.; MOR, A. R.; OUSSAR, Y. Partial discharges and noise classification under HVDC using unsupervised and semi-supervised learning. **International Journal of Electrical Power & Energy Systems**,

London, v. 121, n. 106129, 2020. Disponível em: https://doi.org/10.1016/j.ijepes.2020.106129. Acesso em: 10 de jul. 2020.

MUMALI, F. Artificial neural network-based decision support systems in manufacturing processes: A systematic literature review. **Computers & Industrial Engineering**, Amsterdam, v. 165, n. 107964, 2022. Disponível em: https://doi.org/10.1016/j.cie.2022.107964. Acesso em: 28 de fev. 2022.

MURRAY, B.; PERERA, L. P. A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data. **Ocean Engineering**, Amsterdam, v. 209, n. 107478, 2020. Disponível em: https://doi.org/10.1016/j.oceaneng.2020.107478. Acesso em: 10 de jul. 2020.

NABUURS, G.; SCHELHAAS, M.; ORAZIO, C. *et al.* European perspective on the development of planted forests, including projections to 2065. **New Zealand Journal of Forestry Science**, Rotorua, v. 44, p. 1-7, 2014. Disponível em: http://www.nzjforestryscience.com/content/44/S1/S8. Acesso em: 10 de jul. 2020.

NORIHIRO, J.; ACKERMAN, P.; SPONG, B. D.; LÄNGIN, D. Productivity model for *cut-to-length harvester* operation in south african *eucalyptus* pulpwood plantations. **Croatian Journal of Forest Engineering**, Zagreb, v. 39, n. 1, p. 1-13, 2018. Disponível em: http://www.crojfe.com/site/assets/files/4117/norihiro.pdf. Acesso em: 18 de jul. 2020.

OBI, O. F.; VISSER, R. Including exogenous factors in the evaluation of harvesting crew technical efficiency using a multi-step data envelopment analysis procedure. **Croatian Journal of Forest Engineering**, Zagreb, v. 39, n. 2, p. 153-162, 2018. Disponível em: http://www.crojfe.com/site/assets/files/4216/obi.pdf. Acesso em: 18 de jul. 2020.

OKARO, I. A.; JAYASINGHE, S.; SUTCLIFFE, C.; BLACK, K.; PAOLETTI, P.; GREEN, P. L. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. **Additive Manufacturing**, Amsterdam, v. 27, p. 42-53, 2019. Disponível em: https://doi.org/10.1016/j.addma.2019.01.006. Acesso em: 11 de jul. 2020.

OKORI, W.; OBUA, J. Supervised learning algorithms for famine prediction. **Applied Artificial Intelligence**, London, v. 25, p. 822-835, 2011. Disponível em: https://doi.org/10.1080/08839514.2011.611930. Acesso em: 26 de fev. 2022.

OLIVE, D.J. Multiple linear regression. *In*: **Linear Regression**. Cham, Springer, 2017. Chapter 2, p. 17-83. Disponível em: https://doi.org/10.1007/978-3-319-55252-1_2. Acesso em: 30 de dez. 2021.

OLSSON, L. Road investment scenarios in Northern Sweden. **Forest Policy and Economics**, Amsterdam, v. 7, n. 4, p. 615-623, 2005. Disponível em: https://doi.org/10.1016/j.forpol.2003.11.002. Acesso em: 03 de ago. 2020.
OLSSON, L.; LOHMANDER, P. Optimal forest transportation with respect to road investments. **Forest Policy and Economics**, Amsterdam, v. 7, n. 3, p. 369-379,

2005. Disponível em: https://doi.org/10.1016/j.forpol.2003.07.004. Acesso em: 03 de ago. 2020.

OROZCO, A. L. S.; HUAMÁN, C. Q.; ÁLVAREZ, D. P.; JAVIER, L.; VILLALBA, G. A machine learning forensics technique to detect post-processing in digital videos. **Future Generation Computer Systems**, Amsterdam, v. 111, p.199-212, 2020. Disponível em: https://doi.org/10.1016/j.future.2020.04.041. Acesso em: 08 de jul. 2020.

OSUSZEK, L.; STANEK, S.; TWARDOWSKI, Z. Leverage big data analytics for dynamic informed decisions with advanced case management. **Journal of Decision Systems**, London, v. 25, p. 426-449, 2016. Disponível em: https://doi.org/10.1080/12460125.2016.1187401. Acesso em: 03 de mar. 2022.

PALANDER, T.; ERONEN, J.; KÄRHÄ, K.; OVASKAINEN, H. Development of a wood damage monitoring system for mechanized harvesting. **Annals of forest research**, Brașov, v. 61, n. 2, p. 243-258, 2018. Disponível em: https://doi.org/10.15287/afr.2018.1084. Acesso em: 14 de jul. 2020.

PARAJULI, R; CHIZMAR, S.; HOY, M.; *et al*. Economic contribution analysis of urban forestry in the northeastern and midwestern States of the United States in 2018. **Urban Forestry & Urban Greening**, Amsterdam, v. 69, n. 127490, p. 1-9, 2022. Disponível em: https://doi.org/10.1016/j.ufug.2022.127490. Acesso em: 14 de jul. 2020.

PARIS, F.; STRINO, F.; NADLER, B.; KLUGER, Y. Ranking and combining multiple predictors without labeled data. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 111, n. 4, p. 1253-1258, 2014. Disponível em: https://doi.org/10.1073/pnas.1219097111. Acesso em: 11 de jul. 2020.

PATIL, M. M.; HIREMATH, B. N. A systematic study of data wrangling. *International Journal of Information Technology and Computer Science*, Hong Kong, v. 10, n. 1, p. 32-39, 2018. Disponível em: https://doi.org/10.5815/ijitcs.2018.01.04. Acesso em: 04 jan. 2022.

PEI, L.; SUN, Z.; YU, T.; LI, W.; HAO, X.; HU, Y.; YANG, C. Pavement aggregate shape classification based on extreme gradient boosting. **Construction and Building Materials**, Amsterdam, v. 256, n. 119356, 2020. Disponível em: https://doi.org/10.1016/j.conbuildmat.2020.119356. Acesso em: 12 de jul. 2020.

PENG, B.; TONG, L.; YAN, D.; HUO, W. Experimental research and artificial neural network prediction of free piston expander-linear generator. **Energy Reports**, Amsterdam, v. 8, p. 1966-1978, 2022. Disponível em: https://doi.org/10.1016/j.egyr.2022.01.021. Acesso em: 19 de fev. 2022.

PENG, Y.; LIU, H.; LI, X.; HUANG, J.; WANG, W. Machine learning method for energy consumption prediction of ships in port considering green ports. J**ournal of Cleaner Production**, Amsterdam, v. 264, n. 121564, 2020. Disponível em: https://doi.org/10.1016/j.jclepro.2020.121564. Acesso em: 13 de jul. 2020.

PEREA, F.; MENEZES, M. B. C.; MESA, J. A.; RUBIO-DEL-REY, F. Transportation infrastructure network design in the presence of modal competition: computational complexity classification and a genetic algorithm. **TOP: Revista de Investigación Operativa**, Berlin, v. 28, p. 442-474, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1007/s11750-019-00537-x. Acesso em: 03 de ago. 2020.

PEREIRA, A. L. N.; LOPES, E. S.; DIAS, N. A. Análise técnica e de custo do feller buncher e skidder na colheita de madeira em diferentes produtividades do povoamento. **Ciência Florestal**, Santa Maria, v. 25, n. 4, p. 981–989, 2015. Disponível em: https://doi.org/10.5902/1980509820659. Acesso em: 03 de ago. 2020.

PÉREZ-RODRÍGUEZ, F.; NUNES, L.; AZEVEDO, J. C. A software tool for accurate assessment of costs and CO2 emissions in wood transport using OpenStreetMap©. **Mathematical and Computational Forestry and Natural-Resource Sciences**, East Lansing, v. 8, n. 1, p. 35-53, 2016. Disponível em: https://mcfns.net/index.php/Journal/article/view/MCFNS.8.35/8.35. Acesso em: 03 de ago. 2020.

PERFIL'EV, P. N. Study of optimal bundling methods of round timber. **Lesnoy Zhurnal-Forestry Journal**, Arkhangelsk, v. 2, p. 77-87, 2018. Disponível em: 10.17238/issn0536-1036.2018.2.77. Acesso em: 03 de ago.2020.

PHAM, D. T.; PACKIANATHER, M. S.; AFIFY, A. A. Artificial neural networks. In: ANDINA, D.; PHAM, D. T. (ed.). *Computational intelligence: for engineering and manufacturing*. Boston: Springer, 2007. p. 67-92. Disponível em: https://doi.org/10.1007/0-387-37452-3_3. Acesso em: 05 jan. 2022.

PIRDASHTI, M.; CURTEANU, S.; KAMANGAR, M. H.; HASSIM, M. H.; KHATAMI, M. A. Artificial neural networks: applications in chemical engineering. **Reviews in Chemical Engineering**, Berlim, v. 29, n. 4, p. 205-239, 2013. Disponível em: https://doi.org/10.1515/revce-2013-0013. Acesso em: 05 de jan. 2022.

POLLARD, S. J. T.; BROOKES, A.; EARL, N.; LOWE, J.; KEARNEY, T.; NATHANAIL, C. P. Integrating decision tools for the sustainable management of land contamination. **Science of The Total Environment**, Amsterdam, v. 325, n. 1-5, p. 15-28. Disponível em: https://doi.org/10.1016/j.scitotenv.2003.11.017. Acesso em: 22 de dez. 2021.

POURGHASEMI, H. R.; GAYEN, A.; EDALAT, M.; ZARAFSHR, M.; TIEFENBACHER, J. P. Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management? **Geoscience Frontiers**, Amsterdam, v. 11, n. 4, p. 1203-1217, 2020. Disponível em: https://doi.org/10.1016/j.gsf.2019.10.008. Acesso em: 14 de jul. 2020.

PUKKALA, T.; KANGAS, J. A heuristic optimization method for forest planning and decision making. **Scandinavian Journal of Forest Research**, London, v. 8, p. 560-570. Disponível em: https://doi.org/10.1080/02827589309382802. Acesso em: 21 dez. 2021.

RAHMAN, A. T. M. S.; HOSONO, T.; QUILTY, J. M.; DAS, J.; BASAK, A. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. **Advances in Water Resources**, Amsterdam, v. 141, n. 103595, 2020. Disponível em: https://doi.org/10.1016/j.advwatres.2020.103595. Acesso em: 13 de jul. 2020.

RAMOS, D.; FARIA, P.; MORAIS, A.; VALE, Z. Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office Building. **Energy Reports**, Amsterdam, v. 8, p. 417-422, 2022. Disponível em: https://doi.org/10.1016/j.egyr.2022.01.046. Acesso em: 13 de fev. 2022.

RAULERSON, S.; JACKSON, C. R.; MELEAR, N. D.; YOUNGER, S. E.; DUDLEY, M.; ELLIOTT, K. J. Do southern Appalachian Mountain summer stream temperatures respond to removal of understory rhododendron thickets? **Hydrological Processes**, Malden, v. 34, p. 3045-3060, 2020. Disponível em: https://doi-org.ez87.periodicos.capes.gov.br/10.1002/hyp.13788. Acesso em: 18 de jul. 2020.

REN, J.; ZHANG, Q.; LIU, F. Analysis of factors affecting traction energy consumption of electric multiple unit trains based on data mining. **Journal of Cleaner Production**, Amsterdam, v. 262, n. 121374, 2020. Disponível em: https://doi.org/10.1016/j.jclepro.2020.121374. Acesso em: 09 de jul. 2020.

REZAGHOLIZADEH, M.; HAIDAR, M. A. Reg-gan: semi-supervised learning based on generative adversarial networks for regression. **IEEE – Institute of Electrical and Electronics Engineers**, New York, p. 2806-2810. Disponível em: https://doi.org/10.1109/ICASSP.2018.8462534. Acesso em: 11 de jul. 2020.

RICHHARIYA, B.; TANVEER, M. Least squares projection twin support vector clustering (LSPTSVC). **Information Sciences**, Amsterdam, v. 533, p. 1-23, 2020. Disponível em: https://doi.org/10.1016/j.ins.2020.05.001. Acesso em: 10 de jul. 2020.

ROBINS, A. V. incorporating supervised learning in the domains account of categorization. **Connection Science**, London, v. 4, p. 45-56, 2007. Disponível em: https://doi.org/10.1080/09540099208946603. Acesso em: 26 de fev. 2022.

RODRIGUES, C. K.; LOPES, S. E.; MÜLLER, M. M. L. Avaliação da qualidade estrutural do solo em colheita mecanizada (harvester e forwarder) de Eucalyptus grandis. **Revista Agrarian**, Dourados, v. 13, n. 47, p. 60-67, Dourados, 2020. Disponível em: https://ojs.ufgd.edu.br/agrarian/article/view/8484/5570. Acesso em: 24 jan. 2025.

ROOPA, S. N.; NAGARAJAN, N. Efficient semi-supervised learning and sparse structural learning for feature selection of leukemia dataset. **Journal of Medical Imaging and Health Informatics**, Valley Village, v. 10, n. 8, p. 1815-1824, 2020. Disponível em: https://doi.org/10.1166/jmihi.2020.3110. Acesso em: 11 de jul. 2020.

SANTOS-PEREIRA, J.; GRUENWALD, L.; BERNARDINO, J. Top data mining tools for the healthcare industry. **Journal of King Saud University - Computer and Information Sciences**, Amsterdam, In Press, 2021. Disponível em: https://doi.org/10.1016/j.jksuci.2021.06.002. Acesso em: 13 de fev. 2020.

SAMADI, E.; BADRI, A.; EBRAHIMPOUR, R. Decentralized multi-agent based energy management of microgrid using reinforcement learning. **International Journal of Electrical Power & Energy Systems**, London, v. 122, n. 106211, 2020. Disponível em: https://doi.org/10.1016/j.ijepes.2020.106211. Acesso em: 10 de jul. 2020.

SARKER, I. H. Machine learning: algorithms, real-world applications and research directions. **SN Computer Science**, Cham, v. 2, n. 160, p. 1-23, 2021. Disponível em: https://doi.org/10.1007/s42979-021-00592-x. Acesso em: 29 de dez. 2021.

SARRAZIN; LEBEL; LEHOUX. Analyzing the impact of implementing a logistics center for a complex forest network. **Canadian Journal of Forest Research**, Ottawa, v. 49, n. 2, p. 179-189, 2019. Disponível em: https://doi.org/10.1139/cjfr-2018-0346. Acesso em: 03 de ago. 2020.

SCHETTINO, S.; MINETTE, L. J.; SCHETTINO, C. F.; REBOLETO, I. D. Feasibility analysis of the use of light and medium trucks in timber transport in rural properties. **Revista Árvore**, Viçosa, v. 42, n. 6, e420608, 2018. Disponível em: http://dx.doi.org/10.1590/1806-90882018000600008. Acesso em 03 de ago. 2020.

SCHUH, G.; PROTE, J. P.; HÜNNEKES, P. Data mining methods for macro level process planning. **Procedia CIRP**, Amsterdam, v. 88, p. 48-53, 2020. Disponível em: https://doi.org/10.1016/j.procir.2020.05.009. Acesso em: 09 de jul. 2020.

SHABANI, S.; POURGHASEMI, H.R.; BLASCHKE, T. Forest stand susceptibility mapping during harvesting using logistic regression and boosted regression tree machine learning models. **Global ecology and conservation**, Amsterdam, v. 22, n. e00974, 2020. Disponível em: https://doi.org/10.1016/j.gecco.2020.e00974. Acesso em: 06 de jul. 2020.

SHABANKAREH, M. J.; SHABANKAREH, M. A.; NAZARIAN, A.; RANJBARAN, A.; SEYYEDAMIRI, N. A stacking-based data mining solution to customer churn prediction. **Journal of Relationship Marketing**, London, In Press, 2021. Disponível em: https://doi.org/10.1080/15332667.2021.1889743. Acesso em: 12 de jul. 2020.

SHAHEEN, M. ZAFAR, T.; KHAN, S. A. Decision tree classification: Ranking journals using IGIDI. **Journal of Information Science**, London, v. 46, n. 3, p. 325-339, 2020. Disponível em: https://doi.org/10.1177/0165551519837176. Acesso em: 12 de jul. 2020.

SHAIKH, A. A.; DOSS, A. N.; SUBRAMANIAN, M.; JAIN, V.; NAVED, M.; MOHIDDIN, M. K. Major applications of data mining in medical. **Materials Today: Proceedings**, Amsterdam, In Press, 2021. Disponível em: https://doi.org/10.1016/j.matpr.2021.11.642. Acesso em: 13 de fev. 2022.

SHAKORSHAHABI, R.; QARAHASANLOU, A. N.; AZIMI, S. R.; MOTTAHEDI, A. Application of data mining in Iran's Artisanal and Small-Scale mines challenges analysis. **Resources Policy**, Amsterdam, v. 74, n. 102337, In Press, 2021.

Disponível em: https://doi.org/10.1016/j.resourpol.2021.102337. Acesso em: 13 de fev. 2022.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, New York, v. 27, p. 379–423, 1948. Disponível em: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x. Acesso em: 12 jul. 2020.

SHAMSARA, J.; SCHÜÜRMANN, G. A machine learning approach to discriminate MR1 binders: The importance of the phenol and carbonyl fragments. **Journal of Molecular Structure**, Amsterdam, v. 1217, n. 128459, 2020. Disponível em: https://doi.org/10.1016/j.molstruc.2020.128459. Acesso em: 08 de jul. 2020.

SHARMA, A.; CHOPRA, A. Artificial neural networks: Applications in management. **Journal of Business and Management**, Ghaziabad, v. 12, n. 5, p. 32-40, 2013. Disponível em: https://www.iosrjournals.org/iosr-jbm.html. Acesso em: 05 de jan. 2022.

SHEIKHPOUR, R.; SARRAM, M. A.; GHARAGHANI, S.; CHAHOOKI, M. A. Z. A robust graph-based semi-supervised sparse feature selection method. **Information Sciences**, Amsterdam, v. 531, p. 13-30, 2020. Disponível em: https://doi.org/10.1016/j.ins.2020.03.094. Acesso em: 10 de jul. 2020.

SHEN, S.; GU, K.; CHEN, X.; YANG, M.; WANG, R. movements classification of multi-channel semg based on cnn and stacking ensemble learning. **IEEE Access**, Amsterdam, v.7, p. 137489-137500. Disponível em: https://doi.org/10.1109/ACCESS.2019.2941977. Acesso em: 20 de dez. 2021.

SHI, Y.; LIU, J.; WANG, B.; QI, Z.; TIAN, Y. Deep learning from label proportions with labeled samples. **Neural Networks**, Amsterdam, v. 128, p. 73-81, 2020. Disponível em: https://doi.org/10.1016/j.neunet.2020.04.026. Acesso em: 09 de jul. 2020.

SHIM, S.; KIM, J.; LEE, S. W.; CHO, G. C. Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. **Automation in Construction**, Amsterdam, v. 135, n. 104139, In Press, 2022. Disponível em: https://doi.org/10.1016/j.autcon.2022.104139. Acesso em: 10 de fev. 2022.

SHMILOVICI, A. Support vector machines. In: MAIMON, O.; ROKACH, L. (ed.). Data mining and knowledge discovery handbook. Cham: Springer, 2010. cap. 12, p. 231-247. Disponível em: https://doi.org/10.1007/978-0-387-09823-4_12. Acesso em: 03 jan. 2022.

SHUKLA, A.; BILLARD, A. Augmented-SVM for gradient observations with application to learning multiple-attractor dynamics. In: MA, Y.; GUO, G. (ed.). **Support vector machines applications**. 1. ed. Cham: Springer, 2014. cap. 1, p. 1-21. Disponível em: https://doi.org/10.1007/978-3-319-02300-7. Acesso em: 03 jan. 2022.

SILVA, A. R.; DIAS JÚNIOR, M. S.; LEITE, F. P. Avaliação da intensidade de tráfego e carga de um forwarder sobre a compactação de um Latossolo Vermelho-Amarelo.

Revista Árvore, Viçosa, v. 35, n. 3, p. 547–554, 2011. Disponível em: https://doi.org/10.1590/S0100-67622011000300017. Acesso em: 12 de jul. 2020.

SINGER, G.; ANUAR, R.; BEN-GAL, I. A weighted information-gain measure for ordinal classification trees. **Expert Systems with Applications**, Amsterdam, v. 152, n. 113375, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113375. Acesso em: 12 de jul. 2020.

SKRYPNIK, A. N.; SHCHELCHKOV, A. V.; GORTYSHOV, Y. F.; POPOV, I. A. Artificial neural networks application on friction factor and heat transfer coefficients prediction in tubes with inner helical-finning. **Applied Thermal Engineering**, Amsterdam, v. 206, n. 118049, 2022. Disponível em: https://doi.org/10.1016/j.applthermaleng.2022.118049. Acesso em: 25 de fev. 2022.

SMITH, D.; SMITH, I.; COLLETT, I.; ELMS, S. Forest health surveillance in Victoria. **Australian Forestry**, Abingdon, v. 71, n.3, p.188-195, 2008. Disponível em: https://doi.org/10.1080/00049158.2008.10675034. Acesso em: 03 de mar. 2022.

SOLEYMANI, F.; PAQUET, E. Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder-DeepBreath. **Expert Systems with Applications**, Amsterdam, v. 156, n. 113456, 2020. Disponível em: https://doi.org/10.1016/j.eswa.2020.113456. Acesso em: 10 de jul. 2020.

SUN, M.; XIAO, J.; LIM, E. G.; XIE, Y.; FENG, J. Adaptive ROI generation for video object segmentation using reinforcement learning. **Pattern Recognition**, Amsterdam, v. 106, n. 107465, 2020. Disponível em: https://doi.org/10.1016/j.patcog.2020.107465. Acesso em: 10 de jul. 2020.

SUN, P.; YANG, L. Low-rank supervised and semi-supervised multi-metric learning for classification. **Knowledge-Based Systems**, Amsterdam, v. 236, n. 107787, In Press, 2022. Disponível em: https://doi.org/10.1016/j.knosys.2021.107787. Acesso em: 10 de fev. 2022.

SUN, W.; TREVOR, B. A stacking ensemble learning framework for annual river ice breakup dates. **Journal of Hydrology**, Amsterdam, v. 561, p. 636-650, 2018. Disponível em: https://doi.org/10.1016/j.jhydrol.2018.04.008. Acesso em: 27 de dez. 2021.

SUNHARE, P.; CHOWDHARY, R. R.; CHATTOPADHYAY, M. K. Internet of Things and Data Mining: An Applications Oriented Survey. J**ournal of King Saud University - Computer and Information Sciences**, Amsterdam. In Press, 2020. Disponível em: https://doi.org/10.1016/j.jksuci.2020.07.002. Acesso em: 09 de jul. 2020.

ŠVAČ, V.; CAGÁŇOVÁ, D. Managerial skills for innovation support. **Mobile Networks and Applications**, New York, v. 25, p. 925-931, 2020. Disponível em: https://doi.org/10.1007/s11036-020-01517-3. Acesso em: 10 de jul. 2020.

STEINWART, I.; CHRISTMANN, A. **Support Vector Machines**: Information Science and Statistics. 1. ed. New York, NY: Editora Springer, 2008. Disponível em: https://doi.org/10.1007/978-0-387-77242-4. Acesso em: 03 de jan. 2021.

TA, N. B. D. FC2: cloud-based cluster provisioning for distributed machine learning. **Cluster Computing**, New York, v. 22, p. 1299-1315, 2019. Disponível em: https://doi.org/10.1007/s10586-019-02912-6. Acesso em: 10 de jul. 2020.

TAM, V. W. Y.; BUTERA, A.; LE, K. N.; SILVA, L. C. F.; EVANGELISTA, A. C. J. A prediction model for compressive strength of CO2 concrete using regression analysis and artificial neural networks. **Construction and Building Materials**, Amsterdam, v. 324, n. 126689, 2022. Disponível em: https://doi.org/10.1016/j.conbuildmat.2022.126689. Acesso em: 26 de fev. 2022.

TIAN, X.; VRIES, M. P.; VISSCHER, S. W. J.; PERMENTIER, H. P.; BISCHOFF, R. selective maleylation-directed isobaric peptide termini labeling for accurate proteome quantification. **Analytical Chemistry**, Washington, v. 92, n. 11, p. 7836–7844, 2020. Disponível em: https://dx.doi.org/10.1021/acs.analchem.0c01059. Acesso em: 11 de jul. 2020.

TRIGUERO, I.; GARCÍA-GIL, D.; MAILLO, J.; LUENGO, J.; GARCÍA, S.; HERRERA, F. Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Hoboken, p. e1289, 2018. Disponível em: https://doi.org/10.1002/widm.1289. Acesso em: 03 de jan. 2022.

TRUONG, D.; CHOI, W. Using machine learning algorithms to predict the risk of small unmanned aircraft system violations in the national airspace system. **Journal of Air Transport Management**, Amsterdam, v. 86, n. 101822, 2020. Disponível em: https://doi.org/10.1016/j.jairtraman.2020.101822. Acesso em: 13 de jul. 2020.

TSENG, K. K.; ZHANG, Y.; ZHU, Q.; YUNG, K. L.; IP, W. H. Semi-supervised image depth prediction with deep learning and binocular algorithms. **Applied Soft Computing**, Berlin, v. 92, n. 106272, 2020. Disponível em: https://doi.org/10.1016/j.asoc.2020.106272. Acesso em: 10 de jul. 2020.

ÜNVER-OKAN, S. Modelling of work efficiency in cable traction with tractor implementing the least-squares methods and robust regression. **Croatian Journal of Forest Engineering**, Zagreb, v. 41, n. 1, p. 109-117, 2020. Disponível em: https://doi.org/10.5552/crojfe.2020.677. Acesso em: 07 de abr. 2022.

UYANIK, G. K.; GÜLER, N. A Study on multiple linear regression analysis. **Procedia - Social and Behavioral Sciences**, Amsterdam, v. 106, p. 234-240, 2013. Disponível em: https://doi.org/10.1016/j.sbspro.2013.12.027. Acesso em: 29 de dez. 2021.

VATSAVAI, R. R.; SHEKHAR, S.; BURK, T. E. An efficient spatial semi-supervised learning algorithm. **International Journal of Parallel, Emergent and Distributed Systems**, London, v. 22, p. 427-437, 2007. Disponível em: https://doi.org/10.1080/08839514.2011.611930. Acesso em: 26 de fev. 2022.

VILLE, B.; NEVILLE, P. *Decision Trees for Analytics Using Sas Enterprise Miner*. Cary: Sas Institute, 2013. 268 p.

WAGNER, J. E. Misinterpreting the internal rate of return in sustainable forest management planning and economic analysis. **Journal of Sustainable Forestry**, London, v. 31, n. 3, p. 239-266, 2012. Disponível em: https://doi.org/10.1080/10549811.2011.582830. Acesso em: 03 de mar. 2022.

WANG, D.; DENG, H.; PAN, Z. MRCDRL: Multi-robot coordination with deep reinforcement learning. **Neurocomputing**, Amsterdam, v. 406, p. 68-76, 2020. Disponível em: https://doi.org/10.1016/j.neucom.2020.04.028. Acesso em: 11 de jul. 2020.

WANG, Z.; MARTIN, R. Model-free posterior inference on the area under the receiver operating characteristic curve. **Journal of Statistical Planning and Inference**, Amsterdam, v. 209, p. 174-186, 2020. Disponível em: https://doi.org/10.1016/j.jspi.2020.03.008. Acesso em: 08 de jul. 2020.

WANG, S.; PEDRYCZ, W.; ZHU, Q.; ZHU, W. Subspace learning for unsupervised feature selection via matrix factorization. **Pattern Recognition**, Amsterdam, v. 48, n. 1, p. 10-19, 2015. Disponível em: https://doi.org/10.1016/j.patcog.2014.08.004. Acesso em 10 de jul. 2020.

WEI, X.; CAI, L.; LIAO, B.; LU, T. Local-view-assisted discriminative band selection with hypergraph autolearning for hyperspectral image classification. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 58, n. 3, p. 2042-2055, 2020. Disponível em: https://doi.org/10.1109/TGRS.2019.2952383. Acesso em: 11 de jul. 2020.

WEI, S.; KOSOROK, M. R. Latent Supervised Learning. **Journal of the American Statistical Association**, London, v. 108, p. 957-970, 2013. Disponível em: https://doi.org/10.1080/01621459.2013.789695. Acesso em: 26 de fev. 2022.

WEI, L.; YANG, Y.; SUN, D. Rapid detection of carmine in black tea with spectrophotometry coupled predictive modelling. **Food Chemistry**, Amsterdam, v. 329, n. 127177, 2020. Disponível em: https://doi.org/10.1016/j.foodchem.2020.127177. Acesso em: 08 de jul. 2020.

WELCH, H.; BRODIE, S.; JACOX, M. G.; BOGRAD, S. J.; HAZEN, E. L. Decision-support tools for dynamic management. **Conservation Biology**, Hoboken, v. 34, n. 3, p. 589-599, 2019. Disponível em: https://doi.org/10.1111/cobi.13417. Acesso em: 22 de dez. 2021.

WEN, G.; CHEN, C. L. P.; LI, B. Optimized formation control using simplified reinforcement learning for a class of multiagent systems with unknown dynamics. **IEEE Transactions on Industrial Electronics**, Piscataway, v. 67, n. 9, p. 7879-7888, 2020. Disponível em: https://doi.org/10.1109/TIE.2019.2946545. Acesso em: 11 de jul. 2020.

WICKHAM, H.; AVERICK, M.; BRYAN, J.; CHANG, W.; MCGOWAN, L. D. 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**, Hershey, v. 4, n. 43, 1686. Disponível em: https://doi.org/10.21105/joss.01686. Acesso em: 10 de ago. 2020.

WOLFF, S.; O'DONNCHA, F.; CHEN, B. Statistical and machine learning ensemble modelling to forecast sea surface temperature. **Journal of Marine Systems**, Amsterdam, v. 208, n. 103347, 2020. Disponível em: https://doi.org/10.1016/j.jmarsys.2020.103347. Acesso em: 12 de jul. 2020.

WON, D.; MANZOUR, H.; CHAOVALITWONGSE, W. Convex optimization for group feature selection in networked data. **Informs Journal on Computing**, Catonsville, v. 32, n. 1, p. 182-198, 2020. Disponível em: https://doi.org/10.1287/ijoc.2018.0868. Acesso em: 08 de jul. 2020.

WU, T.; ZHANG, W.; JIAO, X.; GUO, W.; HAMOUD, Y. A. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. **Computers and Electronics in Agriculture**, Amsterdam, v. 184, In Press, 2021. Disponível em: https://doi.org/10.1016/j.compag.2021.106039. Acesso em: 23 de dez. 2021.

XIA, S.; XIONG, Z.; LUO, Y.; DONG, L.; ZHANG, G. Location difference of multiple distances based k-nearest neighbors algorithm. **Knowledge-Based Systems**, Amsterdam, v. 90, p. 99-110, 2015. Disponível em: https://doi.org/10.1016/j.knosys.2015.09.028. Acesso em: 03 de jan. 2021.

XU, Z.; CHENG, X.; WANG, K.; YANG, S. Analysis of the environmental trend of network finance and its influence on traditional commercial banks. **Journal of Computational and Applied Mathematics**, Amsterdam, v. 379, n. 112907, 2020. Disponível em: https://doi.org/10.1016/j.cam.2020.112907. Acesso em: 08 de jul. 2020.

XU, N.; CAI, D.; ZHAO, X. Constructing a eucalyptus identification model based on NDVI time-series remote sensing images. **Catena**, Amsterdam, v. 238, n. 107846, In Press, 2024. Disponível em: https://doi.org/10.1016/j.catena.2024.107846. Acesso em: 08 de jul. 2020.

YAKIMOVICH, A.; BEAUGNON, A.; HUANG, Y.; OZKIRIMLI, E. Labels in a haystack: Approaches beyond supervised learning in biomedical applications. **Patterns**, Amsterdam, v. 2, n. 100383, In Press, 2021. Disponível em: https://doi.org/10.1016/j.patter.2021.100383. Acesso em: 10 de fev. 2022.

YANG, J.; LI, R.; CHEN, L.; HU, Y.; DOU, Z. Research on equipment corrosion diagnosis method and prediction model driven by data. **Process Safety and Environmental Protection**, Amsterdam, v. 158, p. 418-431, In Press, 2022. Disponível em: https://doi.org/10.1016/j.psep.2021.12.006. Acesso em: 13 de fev. 2022.

YU, J.; LI, F.; WANG, Y.; LIN, Y.; PENG, Z.; CHENG, K. Spatiotemporal evolution of tropical forest degradation and its impact on ecological sensitivity: A case study in Jinghong, Xishuangbanna, China. **Science of The Total Environment**, Amsterdam, v. 727, n. 138678, 2020. Disponível em: https://doi.org/10.1016/j.scitotenv.2020.138678. Acesso em: 14 de jul. 2020.

ZANUNCIO, A. J. V.; CARVALHO, A. G.; SILVA, M. G.; LIMA, J. T. Importance of wood drying to the forest transport and pulp mill supply. **Revista Cerne**, Lavras, v. 23, n. 2, p. 147-152, 2017. Disponível em: https://doi.org/10.1590/01047760201723022223. Acesso em: 03 de ago. 2020.

ZHANG, J. Investment risk model based on intelligent fuzzy neural network and VaR. **Journal of Computational and Applied Mathematics**, Amsterdam, v. 371, n. 112707, 2020. Disponível em: https://doi.org/10.1016/j.cam.2019.112707. Acesso em: 10 de jul. 2020.

ZHANG, X. D.; Support Vector Machines. *In*: **A Matrix Algebra Approach to Artificial Intelligence**. Singapore, Springer, 2020. p 617-679. Disponível em: https://doi.org/10.1007/978-981-15-2770-8_8. Acesso em: 03 de jan. 2021.

ZHANG, Q.; YU, X. GrowingNet: An end-to-end growing network for semi-supervised learning. **Computer Communications**, Amsterdam, v. 151, p. 208-215, 2020. Disponível em: https://doi.org/10.1016/j.comcom.2020.01.003. Acesso em: 09 de jul. 2020.

ZHANG, Z.; LI, S.; LU, J.; WANG, J.; JIANG, X. A novel intelligent fault diagnosis method based on fast intrinsic component filtering and pseudo-normalization. **Mechanical Systems and Signal Processing**, London, v. 145, In Press, 2020a. Disponível em: https://doi.org/10.1016/j.ymssp.2020.106923. Acesso em: 10 de jul. 2020.

ZHANG, W.; WANG, R.; ZHANG, Q.; FANG, S. A joint pitch estimation and voicing detection method for melody extraction. **Applied Acoustics**, Amsterdam, v. 166, In Press, 2020b. Disponível em: https://doi.org/10.1016/j.apacoust.2020.107338. Acesso em: 10 de jul. 2020.

ZHANG, Y.; BARTEN, P. K. Watershed forest management information system (WFMIS). **Environmental Modelling & Software**, Amsterdam, v. 24, n. 4, p. 569-575, 2009. Disponível em: https://doi.org/10.1016/j.envsoft.2008.10.006. Acesso em: 18 de jul. 2020.

ZHAO, Y.; ZHANG, C.; ZHANG, Y.; WANG, Z.; LI, J. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. **Energy and Built Environment**, Amsterdam, v. 1, n. 2, p. 149-164, 2020. Disponível em: https://doi.org/10.1016/j.enbenv.2019.11.003. Acesso em: 09 de jul. 2020.

ZHI, Y.; YANG, T.; FU, D. An improved deep forest model for forecast the outdoor atmospheric corrosion rate of low-alloy steels. **Journal of Materials Science &**

Technology, Amsterdam, v. 49, p. 202-210, 2020. Disponível em: https://doi.org/10.1016/j.jmst.2020.01.044. Acesso em: 14 de jul. 2020.

ZHOU, X.; LU, P.; ZHENG, Z.; TOLLIVER, D.; KERAMATI, A. accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. **Reliability Engineering & System Safety**, London, v. 200, n. 106931, 2020a. Disponível em: https://doi.org/10.1016/j.ress.2020.106931. Acesso em: 03 de ago. 2020.

ZHOU, J.; ZHOU, Z.; ZHAO, Q.; HAN, Z.; WANG, P.; XU, J.; DIAN, Y. Evaluation of different algorithms for estimating the growing stock volume of pinus massoniana plantations using spectral and spatial information from a spot6 image. **Forests**, Basel, v. 11, n. 5, p. 540-556, 2020b. Disponível em: https://doi.org/10.3390/f11050540. Acesso em: 06 de jul. 2020.

ZHOU, Q.; LI, Z. A Comparative study of various supervised learning approaches to selective omission in a road network. **The Cartographic Journal**, London, v. 54, p. 254-264, 2016. Disponível em: https://doi.org/10.1179/1743277414Y.0000000083. Acesso em: 26 de fev. 2022.

ZONOUZI, M. N.; KARGARI, M. Modeling uncertainties based on data mining approach in emergency service resource allocation. **Computers & Industrial Engineering**, Amsterdam, v. 145, n. 106485, 2020. Disponível em: https://doi.org/10.1016/j.cie.2020.106485. Acesso em: 09 de jul. 2020.

ZOU, J.; HAN, Y.; SO, S. S. Overview of artificial neural networks. *In*: LIVINGSTONE, D. J. **Artificial Neural Networks**: Methods in Molecular Biology. 1. ed. New Jersey: Humana Press, 2008. cap. 2, v. 458, p. 14-22. Disponível em: https://doi.org/10.1007/978-1-60327-101-1_2. Acesso em: 05 de jan. 2022.