



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Especialização em Ciência de Dados



PREDIÇÃO DE IMA E DISTRIBUIÇÃO DE CLONES POR REGIÃO

Lílian Alves Carvalho Reis

João Monlevade, MG

2024

Lilian Alves Carvalho Reis

PREDIÇÃO DE IMA E DISTRIBUIÇÃO DE CLONES POR REGIÃO

Trabalho de conclusão de curso apresentado ao curso de Especialização em Ciência dos Dados do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto, como parte dos requisitos necessários para a obtenção do título de Especialista em Ciência dos Dados.

Orientadora: Profa. Dra. Sarah Negreiros de Carvalho Leite

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

R375p Reis, Lilian Alves Carvalho.
Predição de imã e distribuição de clones por região. [manuscrito] /
Lilian Alves Carvalho Reis. Sarah Negreiros de Carvalho Leite. - 2024.
32 f.: il.: color., gráf., tab., mapa.

Orientadora: Profa. Dra. Sarah Negreiros de Carvalho Leite.
Produção Científica (Especialização). Universidade Federal de Ouro
Preto. Departamento de Engenharia de Produção.

1. Análise de regressão. 2. Florestas - Produtividade. 3.
Levantamentos florestais. 4. Previsão (Estatística). I. Leite, Sarah
Negreiros de Carvalho. II. Leite, Sarah Negreiros de Carvalho. III.
Universidade Federal de Ouro Preto. IV. Título.

CDU 519.2

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Lilian Alves Carvalho Reis

Predição de IMA e distribuição de clones por região

Trabalho de conclusão de curso apresentado ao curso de Especialização em Ciência de Dados da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Especialista em Ciência de Dados

Aprovada em 28 de fevereiro de 2024

Membros da banca

Dra. Sarah Negreiros de Carvalho Leite - Orientadora - Instituto Tecnológico de Aeronáutica
Dr. George Henrique Godim da Fonseca - Universidade Federal de Ouro Preto
Dra. Anna Carolyne Fernandes Ferreira - Aperam South America

Sarah Negreiros de Carvalho Leite, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 01/04/2024



Documento assinado eletronicamente por **Thiago Augusto de Oliveira Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 03/04/2024, às 15:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0680702** e o código CRC **471B8788**.

Agradecimentos

Agradeço à Deus por ter me dado meu filho Luís Darwin que nasceu um mês antes do início desse curso, ele tem sido a maior força inspiradora que já tive em minha vida.

Agradeço aos colegas do curso Aperam BioEnergia LTDA, pois foram suporte sempre que precisei e mão amiga.

Agradeço a minha irmã que sempre me deu forças para continuar em nossas conversas e contribuiu muito com o presente trabalho.

Agradeço ao meu namorado Cristiano Henrique Lopes Rocha e aos seus filhos Laura e Henrique pelo carinho, apoio, dedicação e compreensão.

Agradeço a Ângela, Dona Marli e a escola crescer de Capelinha por ser apoio para cuidar do Luís.

Também, preciso agradecer o apoio dado pela Divina e suas filhas, que cuidam do Luís nos horários não comerciais sempre que preciso, com muito amor, zelo e carinho. Sem a rede de apoio nada seria possível, pois uma mãe preocupada não é produtiva.

Agradeço a Sara sua irmã Daniele e a toda família pelo apoio no trabalho e cuidado com o Luís desde o seu nascimento até dias atuais.

Agradeço ao Jesimon, por suas veneráveis contribuições e total disposição para me ajudar com as dúvidas de programação.

Agradeço a Aperam BioEnergia LTDA por ter financiado e acreditado no meu potencial para fazer o curso.

Agradeço também a UFOP pelos excelentes profissionais e em especial minha orientadora Sarah Negreiros de Carvalho Leite e a banca avaliadora.

Resumo

A produtividade florestal é o principal indicador a ser acompanhado em uma floresta. Ela pode ser aferida por metodologias de inventário florestal ou no momento da colheita. No presente trabalho foram usados dados de colheita dos últimos 5 anos das áreas produtivas da empresa Aperam BioEnergia LTDA, para prever a produtividade da floresta nos próximos anos. Com o objetivo de obter o menor erro foram testadas três técnicas de predição: regressão linear, árvore de decisão e LASSO (*Least Absolute Shrinkage and Selection Operator*). Em cada abordagem se analisou o modelo usando a variável com maior correlação ou todas as variáveis disponíveis na base de dados. Também foi testada a aplicação do melhor método com o uso da função softmax para direcionar as proporções de cada clone para as regiões. O melhor desempenho de predição foi obtido usando o modelo de regressão linear com todas as variáveis. Esta abordagem apresentou o menor erro médio absoluto percentual quando comparado com os outros métodos testados. O uso da função softmax gerou maior produtividade, quando comparada à recomendação feita manualmente, aumentando a assertividade na recomendação clonal.

Palavras chaves: Floresta, predição linear, árvore de decisão, inventário florestal.

Abstract

Forest productivity is the main indicator to be monitored in a forest. It can be measured using forest inventory methodologies or at the time of harvest. In this work, harvest data from the last 5 years of the productive areas of the company Aperam BioEnergia LTDA were used to predict forest productivity in the coming years. In order to obtain the lowest error, three prediction techniques were tested: linear regression, decision tree and LASSO (Least Absolute Shrinkage and Selection Operator). In each approach, the model was analyzed using the variable with the highest correlation or all variables available in the database. The application of the best method was also tested using the softmax function to direct the proportions of each clone to the regions. The best prediction performance was obtained using the linear regression model with all variables. This approach presented the lowest mean absolute percentage error when compared to the other methods tested. The use of the softmax function generated greater productivity, when compared to the recommendation made manually, increasing the assertiveness of the clonal recommendation.

Keywords: Forest, linear prediction, decision tree, forest inventory.

LISTA DE FIGURAS

Figura 1 - Plantio em mosaico com diferentes clones (Aperam BioEnergia, 2021).....	10
Figura 2 - Fluxograma do Planejamento florestal	12
Figura 3 - Representação das áreas de florestas nativas (verdes) e de plantios (amarelo).....	12
Figura 4 - Caracterização do ciclo florestal adotado pela empresa	13
Figura 5 - Localização e subdivisão das áreas nas unidades de ordenamento	14
Figura 6 - Gráfico de calor	22
Figura 7 - Gráfico de correlação.....	23
Figura 8 - Boxplot variável VOL m ³ /ha	24
Figura 9 - Boxplot contendo todas as variáveis	25

LISTA DE TABELAS

Tabela 1 - Uso do solo Aperam BioEnergia LTDA	10
Tabela 2 - Informações dos clones e proporções de plantio	12
Tabela 3 - Erro médio absoluto para regressão linear (%).....	12
Tabela 4 - Erro médio absoluto para árvore de decisão (%).....	13
Tabela 5 - Erro médio absoluto para LASSO (%).....	14
Tabela 6 - Informações sobre as % de clones disponíveis para plantio	22
Tabela 7 - Recomendação clonal feita pelo profissional	23
Tabela 8 - Recomendação clonal feita pelo softmax	24
Tabela 9 - IMA por UPER de acordo com o banco de dados de colheita 2019 – 2023.....	25

SUMÁRIO

1.	Introdução.....	1
2.	Contextualização Teórica	3
3.	Metodologia.....	7
3.1.	Base de dados	7
3.2.	Pré-processamento dos dados	9
3.3.	Normalização dos dados.....	9
3.4.	Análise de correlação.....	10
3.5.	Seleção de variáveis.....	10
3.6.	Modelos de predição.....	10
3.6.1	Regressão Linear	11
3.6.2	Árvore de decisão	11
3.6.3	LASSO	11
3.6.4	Função softmax.....	12
4.	Resultados e Discussão.....	14
5.	Conclusão	21
	REFERÊNCIAS	22

1. Introdução

A cultura do Eucalipto é uma *commodity* usada para a produção industrial de vários itens como celulose, aço, construção civil, plástico, cosmético, medicamento entre outros. A área plantada no Brasil em 2022 com Eucalipto foi 0,3% maior que o ano anterior, chegando a 9,94 milhões de ha plantados. Os plantios ocorreram principalmente nas regiões sudeste e centro oeste, com destaque para Minas Gerais que possui a maior área plantada, correspondente a 29% do total plantado (IBÁ, 2022). Porém a produtividade florestal pode variar de acordo com vários fatores, sendo eles agrupados em fatores abióticos e bióticos. Dentre os fatores abióticos podemos citar variações de fertilidade e física dos solos, pluviométricas, geadas e ventos. Já quanto às variações bióticas se têm ações de pragas, doença e variabilidade genética (Assis, 2015).

Com o objetivo de obter maior produtividade, a Aperam BioEnergia LTDA começou o programa de melhoramento genético a mais de 50 anos, assim como o plantio das áreas produtivas que são localizadas no Vale do Jequitinhonha. Atualmente, são plantados aproximadamente 75 mil ha e em média se colhe cerca de 12 mil ha por ano (Tabela 1).

Tabela 1 – Uso do solo Aperam BioEnergia LTDA.

Uso do solo	Área (hectares)
Preservação Permanente	2.443,83
Reserva Legal	27.713,83
Faixas Ecológicas	10.948,30
Plantios florestais (eucalipto)	75.753,86
Outras áreas (*)	7.398,93
Total	124.258,75

** Aceiros, Estradas, Carreadores, locais de extração minerais, fragmentos e Unidades de Produção de Energia Renovável.*

Fonte: Aperam BioEnergia

Os clones plantados em maior escala na Aperam BioEnergia atualmente, são classificados como clones comerciais. Sendo eles AEC 1528, AEC 2034, AEC 2475, AEC 2129, AEC 0056 e AEC 0043. No entanto, a Aperam possui 38 clones registrados e 23 protegidos no Ministério da Agricultura, Pecuária e Abastecimento (MAPA), além de 17 clones recém-selecionados. Esses materiais são plantados em mosaico com objetivo de aumento da biodiversidade e resistência a pragas. Essa estratégia faz parte do manejo integrado de pragas e reduz o uso de herbicidas (Costa, 2008; Silva, 2020). Na Figura 1 se pode observar um plantio em mosaico.



Figura 1 – Plantio em mosaico com diferentes clones (Aperam Bio Energia , 2021).

As produtividades florestais podem ser avaliadas e feitas às previsões de produtividade com as metodologias de inventários florestal e várias outras formas. No entanto, os dados de colheita são altamente confiáveis para se analisar a produtividade da área. Pois, esses dados são a produtividade real do talhão sem previsão ou cálculos e são compostos por grande quantidade de informações de manejo como: espaçamento, volume colhido, incremento médio anual (IMA) e clone plantado.

No entanto, os dados da colheita precisam ser transformados em informações para suportar tomadas de decisões futuras, pois o que é colhido atualmente foi plantado há sete anos. Como prever de forma assertiva o que será produzido no futuro como base nos dados de colheita dos últimos anos? Alguns trabalhos apontam algumas formas de previsões como sendo assertivas, sendo elas regressão (Cunha, 2021), árvore de decisão (Diniz , 2023) e *Least Absolute Shrinkage and Selection Operator* (LASSO) (Soares, 2007; Brosfske, 2014; Wang, 2019).

Neste contexto, o objetivo principal deste estudo é analisar a base de dados disponível na Aperam BioEnergia LTDA e gerar modelos preditivos capazes de estimar a produtividade dos clones e direcionar o plantio deles para regiões em que sejam mais produtivos, considerando as proporções produzidas de mudas pelo viveiro, visando aumentar a eficiência e rentabilidade do cultivo florestal.

Para cumprimento do objetivo geral é necessário atender aos seguintes objetivos específicos:

- Os dados precisam ser confiáveis, sem *outliers*, sem dados nulos e serem normalizados.
- Os dados precisam ser em quantidade e qualidade suficiente para permitir a previsão.
- É necessário que as proporções de clones a partir da previsão proporcionem produtividade em IMA $m^3/há/ano$ maiores do que a produtividade do último ano de colheita.

2. Contextualização Teórica

O setor agrícola do Brasil é um dos setores mais forte da econômica brasileira (CEPEA, 2023), sendo uma parte importante desse setor a parte florestal, dado que essa *commodity* ficou em quarto lugar em exportação no ano de 2022. E, atualmente, possui números muito expressivos como 2,6 bilhões de empregos diretos e indiretos, receita bruta de 260 milhões de reais (IBÁ, 2022).

A principal forma de medir a produtividade do setor florestal é a avaliação do Incremento Médio Anual. No ano de 2022, os valores de IMA para o estado Mineiro variaram entre 22,0 e 42,4 m³/ha/ano, já para o país os resultados foram de 32,7 m³/ha/ano em média (IBÁ, 2022). Os dados de IMA podem ser mensurados e calculados através das variáveis de altura comercial da árvore, diâmetro a altura do peito e número total de árvores por ha, ou podem ser apontados por máquinas usadas para a colheita (Vera, 2022). Sendo os dados de colheita, os dados mais coerentes, pois são esses que de fato vão chegar à porta da fábrica, para gerar os subprodutos da floresta como: celulose, papel, carvão, aço, compensados, roupas e itens para o setor alimentício e de medicamentos.

A produtividade de um clone é afetada por questões bióticas e abióticas, entre as abióticas se podem destacar: tipo de solo, quantidade de chuva, insolação e atividades de manejo. Quando o material é plantado e está exposto a diferentes ambientes, acontece a interação entre o genótipo e o ambiente, gerando fenótipos diferentes que podem ter produtividades diferentes entre UPERs (Guo, 2023; Fernandes, 2024).

O planejamento empresarial é de fundamental importância, pois permite otimizar a capacidade atual e prever a necessidade futura para o pleno atendimento das demandas da empresa, seja qual for o horizonte de tempo. Ao se falar em planejamento florestal, esta característica é ainda mais relevante, pois a matéria-prima das empresas florestais apresenta extenso período de maturação e está sujeita às mais diferentes variáveis, tais como intempéries, qualidade do material utilizado e possibilidades de uso, avanços tecnológicos, surgimento de produtos concorrentes ou substitutos, exigências ambientais legais, disponibilidade de terras, entre outras.

Algumas definições são fundamentais neste trabalho tal como o uso da madeira e as fontes de abastecimento. O uso principal da madeira produzida pela Aperam BioEnergia destina-se à produção de carvão. Entretanto, a madeira in natura para serraria e outros usos pode ser vendida pela empresa ao mercado. O abastecimento das unidades de produção é feito, exclusivamente, a partir das florestas plantadas e mantidas pela empresa.

Em virtude das considerações e preocupações diferenciadas em cada horizonte de tempo, o planejamento florestal é dividido em categorias e está inserido em “GERIR A ESTRATÉGIA”, conforme apresentado no fluxograma dos macros processos da Aperam BioEnergia (Figura 2).

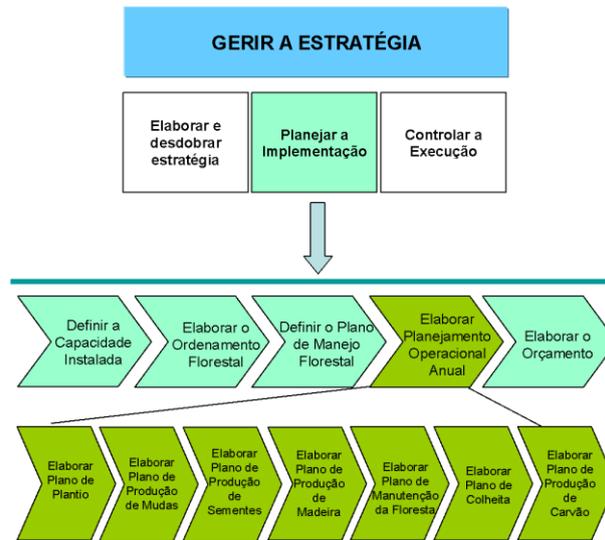


Figura 2 - Fluxograma do Planejamento florestal.

Na Figura 3, é apresentada a distribuição das florestas nativas da empresa, que representam as áreas de preservação e de reservas.

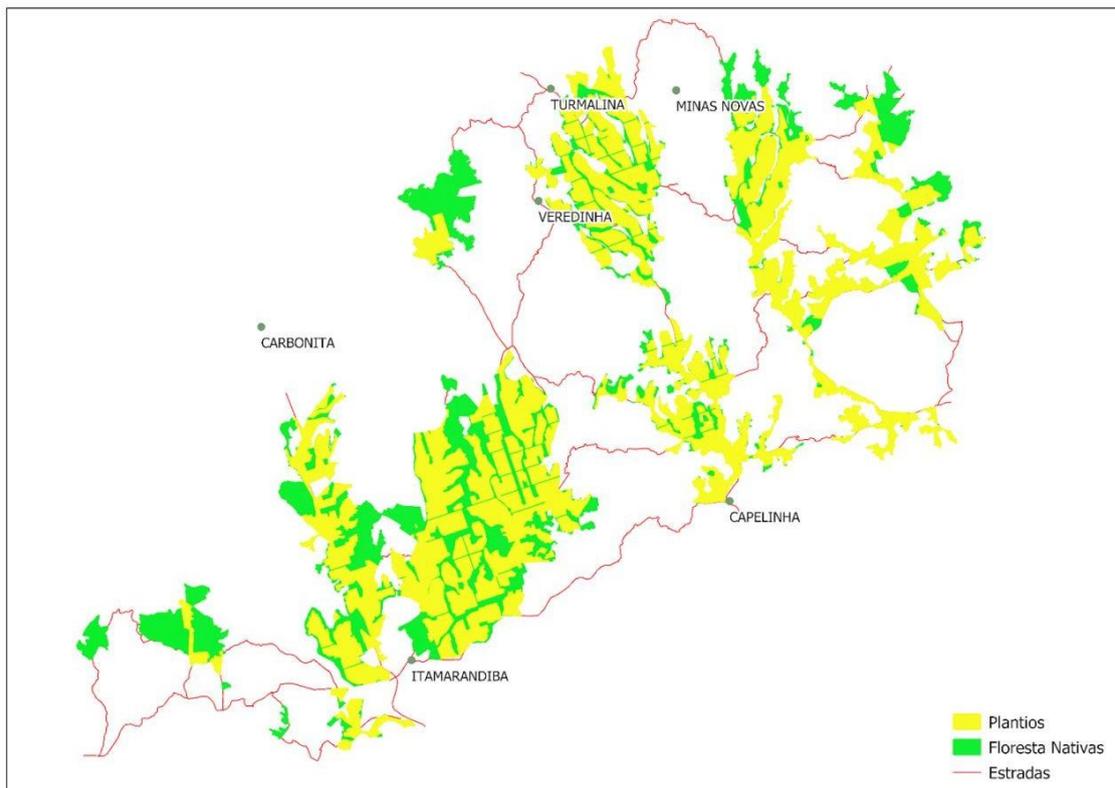


Figura 3 - Representação das áreas de florestas nativas (verdes) e de plantios (amarelo).

O planejamento de longo prazo reflete as grandes metas da empresa, traduzidas em planos de produção, conseqüentemente, na formação de florestas, manutenção e colheita florestal.

O ordenamento florestal é responsável pela avaliação das florestas e suas potencialidades produtivas, visando garantir o suprimento da demanda da Usina Siderúrgica e em carvão sólido renovável e outras demandas, visando o atendimento e o cumprimento das metas de uso múltiplo da floresta.

As florestas da empresa, a partir de 2002, estão sendo conduzidas considerando-se um ciclo de 14 anos, ou seja, duas rotações de 7 anos (Figura 4). No entanto, a empresa, em alguns casos, vem conduzindo plantios para 3ª rotação. Dentro dos princípios do planejamento florestal, é permitido o corte da floresta com idade variando entre 5,5 a 9 anos, dependendo da demanda e necessidade operacional, visando o cumprimento do abastecimento da Usina e/ou mercado de madeira.

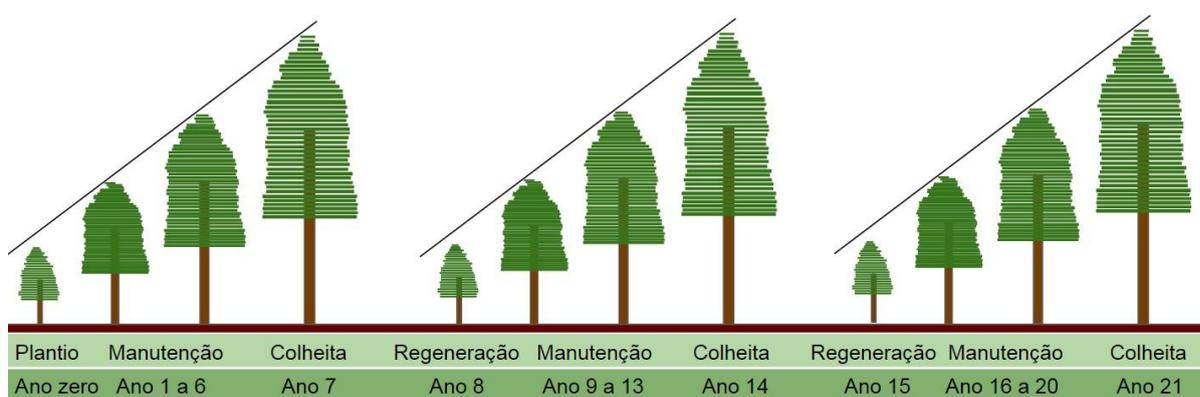


Figura 4 - Caracterização do ciclo florestal adotado pela empresa.

Os projetos florestais estão distribuídos nos diferentes municípios de atuação da Aperam BioEnergia. Portanto, a produção de madeira e carvão é distribuída em blocos, denominados de Unidades de Produção de Energia Renovável (UPER). Na Figura 5 estão identificadas as 6 UPER: São Bento, Cruz Grande, Chácara, Pontal, Palmeiras e Lagoa. Esta subdivisão é necessária para permitir uma melhor organização da produção e do planejamento operacional.

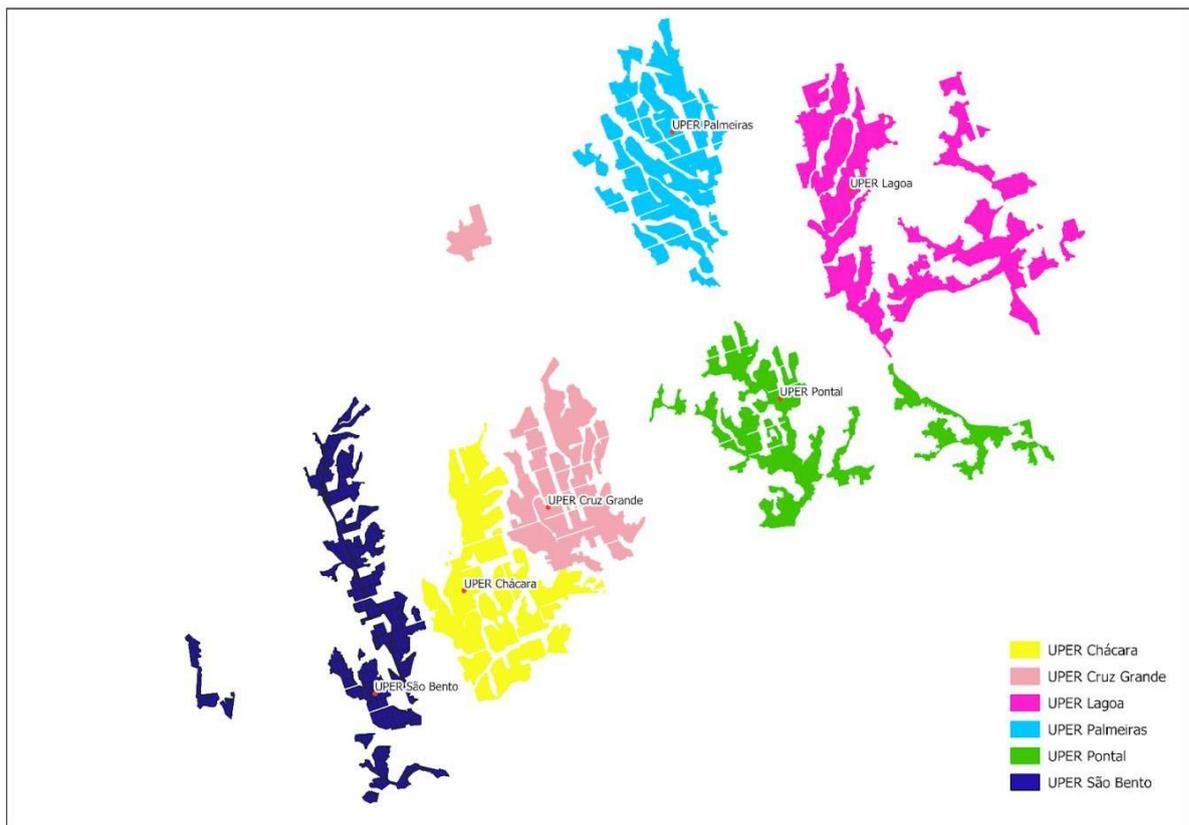


Figura 5 - Localização e subdivisão das áreas nas unidades de ordenamento.

O Inventário Florestal é a base para o planejamento florestal de longo prazo. Ele fornece informações fundamentais para a determinação dos níveis de produção, dando indicações da capacidade de suporte da floresta contida no plano de manejo.

O Inventário Florestal consiste no uso de fundamentos de amostragem para a determinação de características quantitativas ou qualitativas da floresta. Entre as características quantitativas, pode-se citar: volume, altura média das árvores (fustes), diâmetro médio, e outras. Entre as características qualitativas, pode-se citar: vitalidade das árvores, qualidade do fuste, sanidade, etc.

Como as florestas possuem uma grande população de árvores, apenas uma parte das mesmas é avaliada, por meio de parcelas. Os resultados obtidos nessas parcelas permitem fazer estimativas para toda a floresta.

Os tipos de inventários praticados na Aperam BioEnergia são:

- Inventário Florestal Contínuo (IFC);
- Inventário Pré-Corte (IPC).

A base para a realização do Inventário Florestal são os mapas dos projetos florestais, nos quais se tem a identificação dos talhões e informações de cadastro. Os procedimentos operacionais

praticados no inventário estão descritos conforme procedimentos documentados (Procedimentos Operacionais e Procedimentos Sistêmicos).

O inventário é feito a partir de amostras com repetições dentro de cada talhão, e a partir da mensuração dessas árvores o volume médio delas é usado para o cálculo de volume de todo o talhão. As amostras são feitas com 9 árvores e três repetições totalizando 27 árvores mas em um ha temos aproximadamente 1333 árvores e cada talhão tem em média 50 ha. A extrapolação da média pode gerar erro quanto à previsão do volume. Além disso, o momento da aferição pode ser meses antes da colheita e a floresta continua a crescer e a produzir. Já o volume mensurado pelas máquinas de colheita são os mais próximos possíveis do que a floresta vai realmente produzir.

No entanto, o banco de dados de colheita possui informações de vários anos e com algumas variáveis, transformar esses dados em informações que permitam tomadas de decisão com segurança é um desafio. Porém, esse desafio pode ser superado com a aplicação de análises estatísticas para verificar presença de *outlier* e normalização dos dados em grande escala. Além de ser possível usar a predição de *machine learning* para prever produções futuras e aumentar a produtividade por região.

Neste sentido é importante prever qual o IMA que as florestas que ainda não foram plantadas podem dar para que a empresa possa se adiantar com segurança no viveiro para a produção de mudas e até para a tomada de decisão de compra de terras.

3. Metodologia

3.1. Base de dados

A base de dados usada no presente trabalho é composta por informações de 5 anos de colheita florestal contendo 1913 amostras. Destas, há somente 20 amostras com algum dado nulo e 15 variáveis que podem descrever e ou se correlacionar com a variável Incremento médio anual (IMA), sendo elas:

1. **Área:** essa variável se refere à quantidade de hectares plantados;
2. **UPER:** Unidade Produtiva de Energia Renovável;
3. **Rotação:** variável ligada ao ciclo da planta, podendo ser classificada como plantio que é a primeira rotação, depois que a área é colhida segunda rotação, e depois que a segunda rotação é colhida como terceira rotação;
4. **ID CORTE:** informa a idade que a rotação atual possui;
5. **ESPAÇAMENTO:** forma pela qual as mudas foram distribuídas na área plantada;
6. **VOL m³/ha:** volume de madeira produzido em um hectare;
7. **VOLUME DE CORTE:** quantidade apontada como colhida na área;
8. **VMI:** Volume Médio Individual das árvores de uma determinada área;

9. **Material genético:** forma pela qual o material genético ou clone é nomeado;
10. **VMIXAREA:** multiplicação das variáveis VMI e Área para conhecer o volume médio por hectare;
11. **FUSTESXÁREA:** multiplicação das variáveis FUSTE e área sendo que a variável FUSTE se refere à quantidade de árvore em uma determinada área;
12. **IMAXAREA:** multiplicação das variáveis IMA e área;
13. **IDADEXAREA:** multiplicação das variáveis de IDADE e área;
14. **ESPAÇAMENTO A:** espaço entre ruas de plantio;
15. **ESPAÇAMENTO B:** espaço entre plantas dentro da linha de plantio. O **ESPAÇAMENTO** é obtido multiplicando o **ESPAÇAMENTO A** com o **ESPAÇAMENTO B** com unidade de medida de m².

Destas, as variáveis de entrada são: Área, UPER, Rotação, ID CORTE, ESPAÇAMENTO, VOL m³/ha, VOLUME DE CORTE, VMI, Material genético. As demais variáveis são calculadas, sendo: VMIXAREA, FUSTESXÁREA, IMAXAREA, IDADEXAREA.

A área é uma variável de grande importância, pois ela se refere à área plantada com um determinado clone em um talhão, sendo que o talhão é a menor parte de terra em que se divide a área total. O tamanho do talhão na Aperam BioEnergia LTDA pode variar de 0,1 ha até 73,3 ha, durante o preparo dos dados os volumes dos talhões pequenos foram conferidos um a um. Talhões pequenos podem apresentar grandes volumes em função do efeito de borda. Árvores que ficam localizadas nas bordas dos talhões apresentam maior crescimento pela redução da competição por água e luz com outras árvores.

A variável UPER está ligada a aglomerações de talhões que ficam próximos a Unidade Produtiva de Energia Renovável (UPER), local onde a madeira é transformada em carvão, bio óleo entre outros produtos. A UPER é localizada próxima às cidades de atuação da empresa e por esse motivo podemos observar produtividades diferentes entre as UPER, principalmente pela diferença entre o volume de chuva anual que cada cidade possui.

O termo Rotação é usado para identificar se o talhão foi recentemente plantado, ou se está sendo conduzida a brotação da árvore. Se for primeira rotação, espera-se maior produtividade e maior sobrevivência por ser o período de produtividade do plantio, na segunda rotação a produtividade reduz, assim como na terceira a produtividade é menor que na segunda.

A idade de corte (ID CORTE) impacta muito o volume produzido, pois está ligada ao período que a árvore teve para fazer fotossíntese e com isso acumular carbono para o seu crescimento em volume. As variáveis VOL m³/há, VOLUME DE CORTE e VMI permitem avaliar o quanto foi produzido em volume, sendo eles (VOL m³/ha) relacionada a produtividade, VOLUME DE CORTE o quanto foi produzido no talhão total e VMI o quanto cada árvore produziu. Essa última variável pode

refletir as variações do espaçamento de plantio, pois o volume colhido é dividido pelo número de plantas plantadas por hectare para saber em média qual foi o VMI do talhão.

Para calcular o IMA m³/ha/ano são usadas as variáveis ID CORTE e VOL m³/ha, dados Aparam BioEnergia LTDA apresentam IMA médio de 28,41 m³/ha/ano, sendo respectivamente de 30,57; 26,89 e 23,47 m³/ha/ano para a primeira, segunda e terceira rotação, contendo 1893 dados com a informação de materiais genético, sendo 37 clones diferentes colhidos nos 4 anos, e o clone mais colhido foi o AEC 1528 com frequência de 697 vezes.

3.2. Pré-processamento dos dados

Os dados foram carregados para o sistema, feito upload e para isso é importante que eles estejam salvos em Excel modelo *csv*. Posteriormente as amostras que possuíam dados perdidos em qualquer variável foram eliminadas. A variável espaçamento estava composta pelos números de espaçamento entre linhas e entre ruas de plantio. Exemplo: espaçamento 3 x 1,5; o que significa que entre as ruas de plantio o espaçamento é de 3 metros e entre as plantas de 1,5 metros. Para permitir a análise ele foi transformado em espaçamento A, que trás a informação da distância entre ruas e de espaçamento B, que trás a informação entre plantas. Posteriormente foram multiplicados para obtermos o espaçamento por plantas em metros quadrados.

As variáveis categóricas, como UPERS, clones, data de corte foram transformadas em números variando entre 0 e 1 e as variáveis numéricas foram convertidas para *float* porque quando o número vem como texto não é possível processar os dados.

O número total de amostras antes da transformação eram 1913 amostras e após a retirada das amostras nulas passaram a serem 1893 amostras. Partes dos dados foram usados para construir o modelo e parte foi usada para validar o modelo, esse teste foi rodado por 5 vezes.

3.3. Normalização dos dados

A normalização dos dados possui grande importância para chegar a um resultado satisfatório com dados coerentes e de fácil compreensão. Além disso, facilita a convergência dos modelos de predição, uma vez que as variáveis de entrada ficam escaladas numa mesma ordem de grandeza.

Neste trabalho, empregou-se a normalização de máximos e mínimos no range [-1,1], que dimensiona e traduz cada característica individualmente de modo que esteja no intervalo dado no conjunto de treinamento, por exemplo, entre menos um e um.

A transformação é dada por:

$$X_{escalado} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

sendo X a variável, $\min(X)$ o menor valor da variável e $\max(X)$ o maior valor assumido pela variável.

Essa normalização não elimina o efeito de valores atípicos, mas os ajusta linearmente para um intervalo fixo, em que o maior ponto de dados corresponde ao valor máximo e o menor ponto corresponde ao valor mínimo (Braga, 2020).

3.4. Análise de correlação

O objetivo do estudo da correlação é determinar (mensurar) o grau de relacionamento entre duas variáveis. Caso os pontos das variáveis, representados num plano cartesiano (X , Y) ou gráfico de dispersão, apresentem uma dispersão ao longo de uma reta imaginária, pode-se dizer que os dados apresentam uma correlação linear. A regressão linear é uma técnica de análise de dados que prevê o valor de dados desconhecidos usando outro valor de dados relacionado, conhecido como o emprego de uma equação. Ele modela matematicamente a variável desconhecida ou dependente e a variável conhecida ou independente como uma equação linear (Martins, 2019), podendo ser aplicado em uma variável (regressão linear simples) ou em várias (regressão linear multivariada).

Foi construído um gráfico de *heatmap*, figura 6, para análise das correlações.

3.5. Seleção de variáveis

O critério usado para selecionar as variáveis que seriam usadas nos três modelos de predição foi o valor de correlação com a variável de interesse o IMA $\text{m}^3/\text{ha}/\text{ano}$.

Todos os três modelos de predição, regressão linear, árvore de decisão e LASSO, foram testados considerando dois casos:

- Empregando todas as variáveis, sendo elas: TALHAO, UPER, REGIONAL, ÁREA, DATA CORTE, ANO CORTE, ROTAÇÃO, DATA ROTAÇÃO, ID CORTE, MATGEN, ESPAÇAMENTO, VOL m^3/ha , VOLUME CORTE (m^3), FUSTES N/ha, VMI, VMIXAREA, FUSTESXAREA, IMAXAREA, IDADEXAREA, espacamentoA, espacamentoB.
- Considerando somente a variável Volume, que foi a que apresentou maior correlação com a variável de saída IMA.

3.6. Modelos de predição

Neste trabalho, os métodos usados para a predição de dados foram regressão linear, árvore de decisão e LASSO.

3.6.1 Regressão Linear

A regressão linear em resumo é o quanto uma equação linear explica a relação entre uma determinada variável x em função de uma variável y com coeficientes $w = (w_1, \dots, w_p)$ que minimizam a soma residual dos quadrados. Matematicamente resolve um problema da forma:

$$\min_w \|Xw - y\|_2^2$$

As estimativas dos coeficientes para Mínimos Quadrados Ordinários baseiam-se na independência dos recursos. O gera um coeficiente de correlação mais próximo de 1 para variáveis que são “consequências” uma das outras em um determinado experimento (Rifkin, 2007).

3.6.2 Árvore de decisão

O método da árvore de decisão é um algoritmo de aprendizado de máquina utilizado para classificação e para regressão. Isto é, pode ser usado para prever categorias discretas (sim ou não, por exemplo) e para prever valores numéricos (o valor do lucro em reais, por exemplo). Essa técnica cria modelos de fácil interpretação, mas com ela corre-se o risco de *overfitting* (sobre ajuste) dos dados de treinamento (Santana, 2023; Junior, 2019). A técnica pode ser aplicada em uma variável ou em várias.

O termo *overfitting* define o que acontece quando um modelo se ajusta muito a uma determinada base de dados, o que leva a erros futuros quando novos dados são inseridos a base.

As variáveis usadas para a construção do modelo, para o treinamento e para o teste foram: TALHAO, UPER, REGIONAL, ÁREA, DATA CORTE, ANO CORTE, ROTAÇÃO, DATA ROTAÇÃO, ID CORTE, MATGEN, ESPAÇAMENTO, VOL m³/ha, VOLUME CORTE (m³), FUSTES N/ha, VMI, VMIXAREA, FUSTESXAREA, IMAXAREA, IDADEXAREA, espacamentoA, espacamentoB.

3.6.3 LASSO

O LASSO é um modelo linear que estima coeficientes esparsos e visa identificar as variáveis e os coeficientes de regressão correspondentes que levam a um modelo que minimiza o erro de previsão. Isto é conseguido através da imposição de uma restrição nos parâmetros do modelo, que reduz os coeficientes de regressão para zero, isto é, forçando a soma do valor absoluto dos coeficientes de regressão a ser inferior a um valor fixo (λ). Variáveis com coeficiente de regressão zero após contração são excluídas do modelo. A escolha de λ é frequentemente feita usando uma abordagem automatizada de validação cruzada *k-fold*. Para esta abordagem, o conjunto de dados é particionado aleatoriamente em k subamostras de igual tamanho. Enquanto as subamostras $k-1$ são usadas para desenvolver um modelo de previsão, o restante das subamostras são usadas para validar este modelo. Este procedimento é realizado k vezes, com cada uma das k subamostras, por sua vez, sendo usadas para validação e as outras para treinamento do modelo. Neste trabalho, adotou-se $k=5$.

Um resultado geral é produzido combinando os k resultados de validação, separados em um intervalo de valores de λ , desta forma é determinado o λ preferido para o modelo final. Uma vantagem particular com esta técnica é que ela reduz o *overfitting* sem restringir um subconjunto do conjunto de dados para uso exclusivo para validação interna (Ranstan, 2018).

O *overfitting* tende a levar a um modelo otimista e enviesado, que compromete a capacidade de generalização na predição para dados não vistos durante o treinamento. O modelo tende a funcionar particularmente mal com a previsão de observações mais ‘extremos’ (muito risco alto ou muito baixo). Diversas técnicas de regressão podem ser usadas para resolver esses problemas (penalizados ou regularizados) (Ranstan, 2018).

Matematicamente, consiste em um modelo linear com um termo de regularização adicionado. A função objetivo a minimizar é:

$$\min_w \frac{1}{2n_{\text{samples}}} \|Xw - y\|_2^2 + \alpha \|w\|_1$$

A estimativa do LASSO resolve assim a minimização da penalidade dos mínimos quadrados com $\alpha \|w\|_1$ adicionado, onde α é uma constante e $\|w\|_1$ é a ℓ_1 norma do vetor de coeficientes (Kim, 2007).

As variáveis usadas para a construção do modelo, para o treinamento e para o teste foram: TALHAO, UPER, REGIONAL, ÁREA, DATA CORTE, ANO CORTE, ROTAÇÃO, DATA ROTAÇÃO, ID CORTE, MATGEN, ESPAÇAMENTO, VOL m³/ha, VOLUME CORTE (m³), FUSTES N/ha, VMI, VMIXAREA, FUSTESXAREA, IMAXAREA, IDADEXAREA, espacamentoA, espacamentoB.

Os resultados de Scolforo (2023) apontam a técnica do LASSO como sendo promissora para a predição do comportamento biológico necessário para melhor estimar o estoque de carbono no solo de florestas.

3.6.4 Função softmax

A recomendação clonal por talhão é feita com base na porcentagem clonal indicada pelo especialista em melhoramento genético. Esses dados foram usados no algoritmo de aprendizagem por reforço softmax e estão apresentados na Tabela 2:

Tabela 2 – Informações dos clones e proporções de plantio.

Clone	Área	%	Densidade	IMA	IMA Carv	Novos clones	Clone
AEC 1528	1272,35	17,5	490	36	12		AEC 1528
AEC	1756,49	24,1	588	41	15		AEC

2034							2034
AEC	1242,14	17,1	582	37	14		AEC
2233							2233
AEC	722,21	9,9	550	42	16		AEC
2475							2475
AEC	167,46	2,3	549	37	13		-
0056							
AEC	584,67	8,0	490	32	11		-
0144							
AEC	100,77	1,4	598	39	12	1,4	AEC
3704							3704
AEC	142,28	2,0	634	60	20	2,0	AEC
0043							0043
AEC	13,57	0,2	553	44	16	0,2	AEC
3166							3709
AEC	51,09	0,7	515	33	12	0,7	AEC
3709							0224

Na base de dados há vários clones que ocupam uma porcentagem da área total. A função softmax transforma as saídas, para cada classe, em valores entre 0 e 1, dando a probabilidade de a entrada estar em uma determinada classe (Silva, 2021).

O método estatístico usado, algoritmo de aprendizado por reforço softmax (Dayan, 1993), é adaptável e simples, pois necessita apenas manter as informações das recompensas coletadas e as possíveis ações que podem ser tomadas. No presente trabalho as informações necessárias foram as porcentagem possíveis de serem plantada de cada clone. O softmax atribui às ações uma probabilidade proporcional às suas estimativas de valor. O que significa que se a estimativa de valor foi maior para uma determinada UPER para um determinado clone, esse clone foi mais produtivo nessa região. O melhor clone continua tendo a maior probabilidade enquanto todas as outras são distribuídas em um ranking de acordo com suas estimativas de valor. O método mais comum para calcular a probabilidade é a distribuição de Gibbs ou Boltzmann que escolhe a ação no tempo com a probabilidade (Blanchard, 2021).

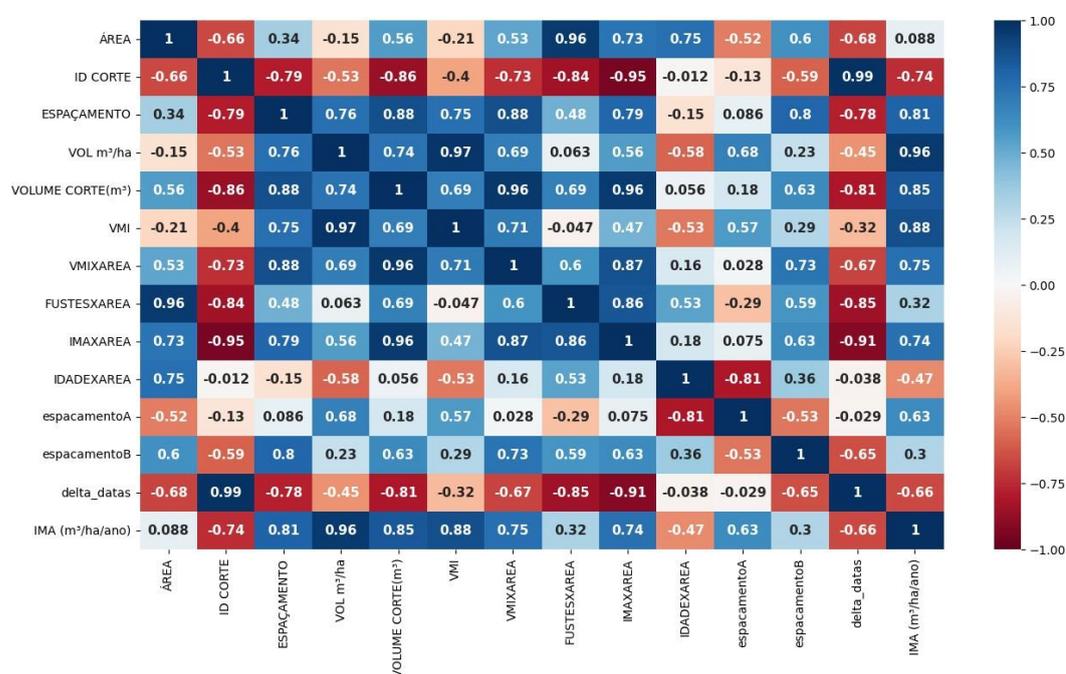
A partir dos dados preditos, foi indicada a melhor proporção de clones por região. Esta função auxiliou na recomendação clonal por talhão que é atualmente realizada de maneira manual, considerando a expertise dos funcionários. Usar o poder de predição junto com a melhor proporção do clone permite conferir qual foi o melhor resultado, o predito ou o executado a mão.

4. Resultados e Discussão

4.1 Análises dos dados

A variável IMA é a variável de saída, ou seja, a qual se deseja prever nesse estudo. Desta forma, foi feita uma análise de correlação dela com as outras variáveis, conforme apresentada na Figura 6. Nessa figura foi construído um mapa de calor. A escala do mapa indica que quanto mais escura for a cor azul, maior a correlação positiva e quanto mais escura for a cor vermelha, maior a correlação negativa, ou seja, as variáveis são inversamente proporcionais. As saídas próximas do branco/amarelo indicam baixa correlação.

Figura 6 – Correlação das variáveis de entrada e saída.

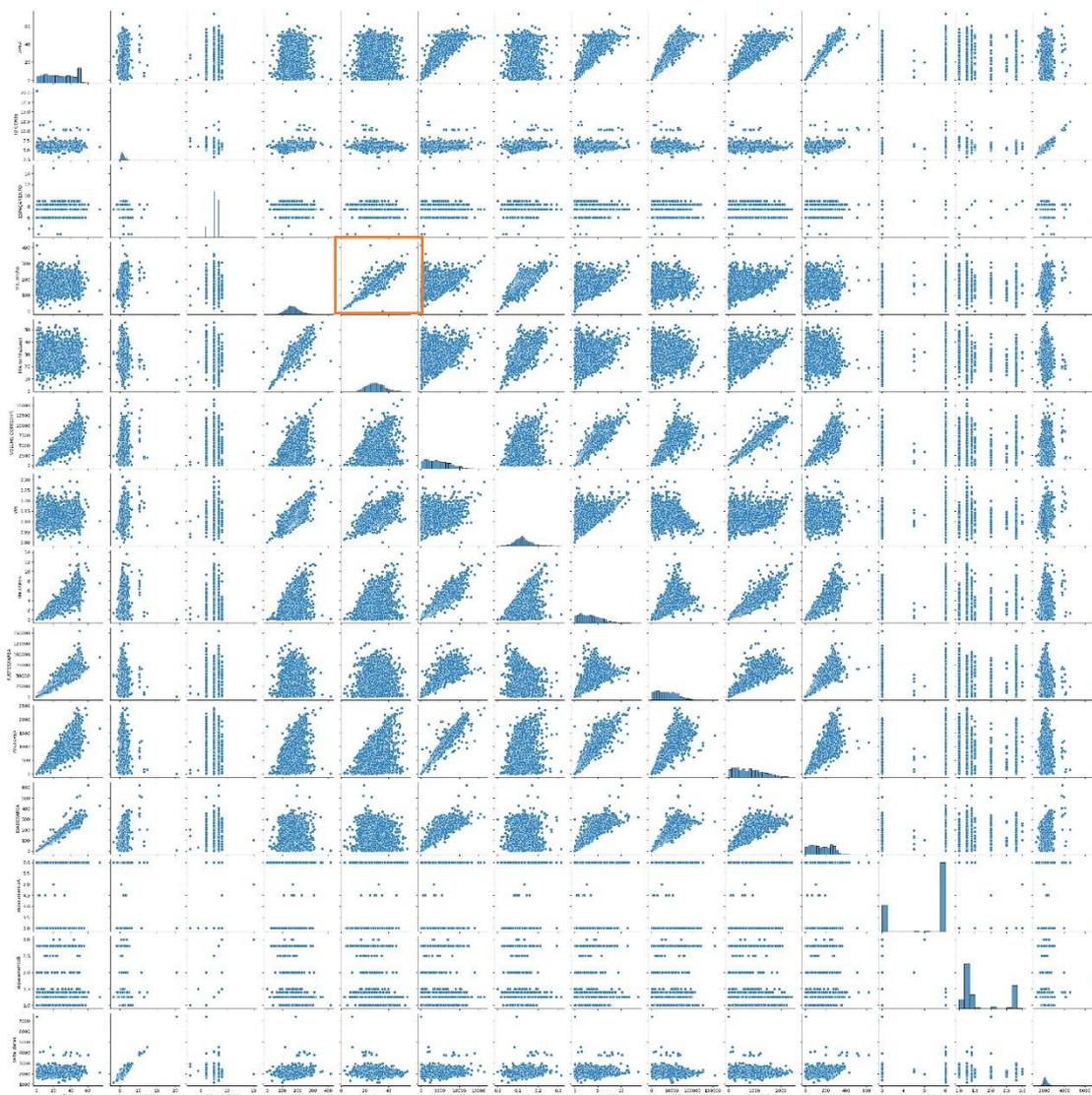


A correlação entre as variáveis foi usada na seleção da variável Volume apresentada no mapa de calor como (VOL m³/ha) para testar quanto a sua capacidade de prever a produtividade futura dos clones visto que ela obteve índice de correção com o IMA (m³/ha/ano) de 0,96 e é possível aferir a mesma ao longo do ciclo produtivo do talhão.

O volume teve correlação negativa com relação à idade de corte, ou se, quanto maior a idade de corte menor o volume. Sabe-se que ocorre redução da velocidade do aumento do volume quando ocorre aumento de idade e que provavelmente essa é uma relação de segundo grau.

A Figura 7 auxilia na análise da correlação e na apresentação das distribuições estatísticas de cada variável da base.

Figura 7 – Gráfico de correlação.

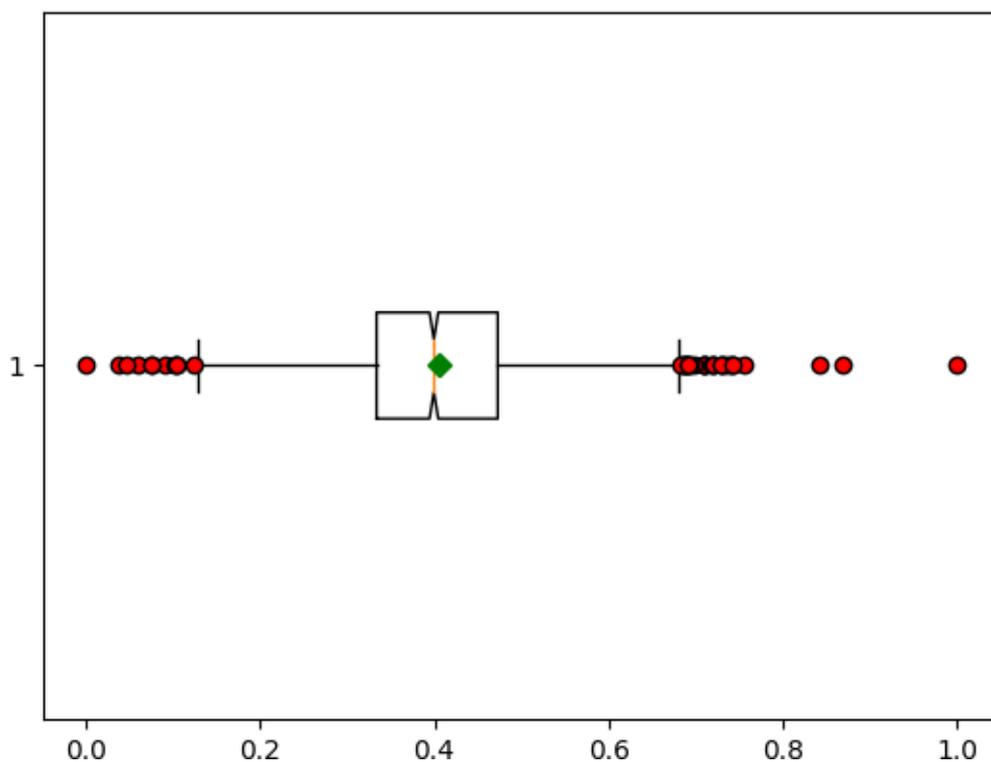


Importante destacar que o IMA e o volume tiveram correlações positivas e altas como está destacado com um quadrado alaranjado na Figura 7. As variáveis que tiveram maiores valores de correlação com IMA, foram respectivamente Volume m³/ha 0,96; Volume de corte m³ 0,85; Espaçamento 0,81; VMIXAREA 0,75; IMAXAREA 0,74; ID corte -0,74 e EspaçamentoA 0,63. No entanto, é importante salientar que Volume de corte e área são variáveis usadas para calcular volume m³/ha. Dentro dos espaçamentos presentes da Aperam BioEnergia LTDA quanto menor o espaçamento maior é a quantidade de árvores e, conseqüentemente, maior o volume colhido na área. E quanto maior a ID corte menor o crescimento da árvore por ano. Dessa forma, a variável volume m³/ha representa bem a maioria das variáveis, por estar correlacionada com as demais, conforme indica a Figura 6.

O boxplot é uma forma gráfica e padronizada de exibir a distribuição de dados usando cinco números importantes, sendo eles a médias, o percentil 25%, o percentil 75% e os valores de máximo e mínimo, sendo que os pontos acima dessas margens são representados como *outliers*.

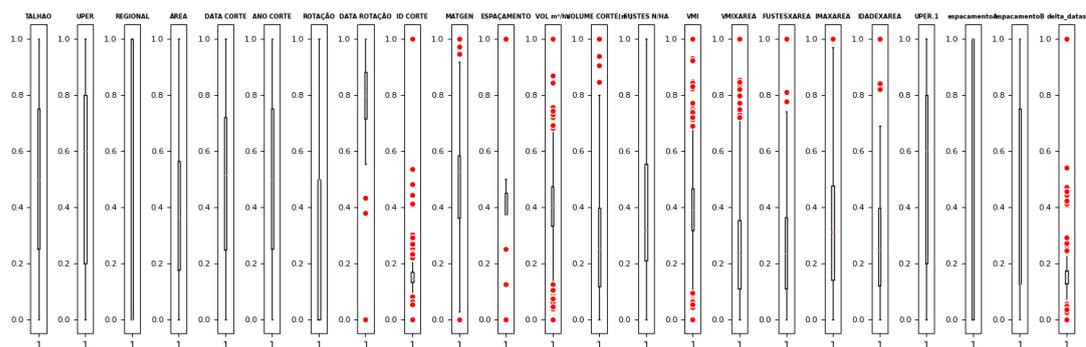
Na Figura 8, o boxplot apresentado demonstra a distribuição dos dados da variável VOL m³/ha, os *outliers* foram coloridos de vermelho, a mediana sendo uma linha amarela e a média foi destacada no formato de diamante colorida de verde.

Figura 8 – Boxplot variável VOL m³/ha.



Ao se observar todas as variáveis na Figura 9, plotadas ao mesmo tempo nota-se que existem *outliers* para as variáveis ID CORTE, ESPAÇAMENTO, VOL E VOLUME DE CORTE, FUSTES, VMI e depois calculados com suas multiplicações.

Figura 9 – Boxplots de todas as variáveis.



4.2 Modelos de predição

A Tabela 3 apresenta o erro absoluto médio percentual (MAPE) fornecido pelo modelo de regressão linear. O MAPE variou em torno de 7 a 10% quando se empregou somente a variável VOL m³/há na entrada do modelo. Quando se usou todas as variáveis de entrada, o erro caiu para o *range* de 3 a 4 % na predição do IMA, dependendo do número de execuções.

Para o método de predição usando a estrutura de árvore de decisão, o MAPE variou entre 10 e 12%, usando somente a variável VOL m³/há. Empregando todas as características, o MAPE variou entre 3 e 8%, conforme apresentado na Tabela 4. Os resultados com a técnica de predição LASSO são apresentados na Tabela 5. Observa-se que o MAPE variou entre 8 e 12%, usando somente a variável VOL m³/há, e entre 7 e 8 % empregando todas as variáveis de entrada. Os dados preditos na presente trabalho foram obtidos com erros baixos do que os apresentados no trabalho de (Azevedo, 2020).

Dentre as seis configurações testadas, o método da regressão linear usando todas as variáveis foi capaz de fornecer a melhor estima de predição dos IMAs, com um MAPE médio de 3,93 % (Tabela 3).

Valente (2021) e Teixeira (2023) também encontraram resultados semelhantes, com menores erros para regressão com múltiplas variáveis ao comparar a regressão linear.

Tabela 3 – Erro médio absoluto para regressão linear (%).

Número de execuções	1	2	3	4	5	Média
Variável volume	7,32	10,61	7,28	10,13	10,02	9,08
Todas as variáveis	3,59	3,60	2,75	4,79	4,95	3,93

Tabela 4 – Erro médio absoluto para árvore de decisão (%).

Número de execuções	1	2	3	4	5	Média
Variável volume	10,95	12,54	10,36	12,05	12,42	11,66
Todas as variáveis	3,30	3,91	3,47	5,42	8,33	4,89

Tabela 5 – Erro médio absoluto para LASSO (%).

Número de execuções	1	2	3	4	5	Média
Variável volume	9,08	12,52	8,55	11,12	11,46	10,73
Todas as variáveis	7,82	11,21	8,07	10,13	10,95	9,31

Para verificar a adequabilidade prática dos dados preditos, foi feita uma simulação considerando uma área total de plantio de aproximadamente 6.000 ha e as porcentagens de áreas disponíveis para plantio por clone e por regional (UPERs). Dessa forma foi possível comparar o resultado de IMA do ano de 2021, feito pela predição e ajuda do softmax com o valor manual indicado pelo especialista em recomendação clonal por região da empresa.

Tabela 6 – Informações sobre as % de clones disponíveis para plantio.

Clone	Área	%	Densidade	IMA	IMA Carv	Novos clones	Clone
AEC 1528	1272,35	17,5	490	36	12		AEC 1528
AEC 2034	1756,49	24,1	588	41	15		AEC 2034
AEC 2233	1242,14	17,1	582	37	14		AEC 2233
AEC 2475	722,21	9,9	550	42	16		AEC 2475
AEC 0056	167,46	2,3	549	37	13		-
AEC 0144	584,67	8,0	490	32	11		-
AEC 3704	100,77	1,4	598	39	12	1,4	AEC 3704
AEC 0043	142,28	2,0	634	60	20	2,0	AEC 0043
AEC 3166	13,57	0,2	553	44	16	0,2	AEC 3709
AEC 3709	51,09	0,7	515	33	12	0,7	AEC 0224

A coluna (%) da Tabela 6 indica os valores de acordo com o prognóstico de clones disponíveis e possíveis de serem produzidos pelo viveiro no próximo ano. Usando essa informação, a função softmax retornou as melhores (%) de cada clone por UPER.

Anualmente, o especialista em melhoramento genético faz a recomendação clonal por talhão, com o objetivo de obter a maior produtividade possível mantendo as porcentagens clonais produzidas pelo viveiro e o plantio em mosaico clonal. O mosaico clonal trata-se de uma forma de plantar que impede que sejam formados grandes blocos contendo apenas um clone. Dessa forma o plantio fica menos susceptível a pragas e a doenças. Na Tabela 7 é apresentada a recomendação feita no ano de

2021 com a informação de área recomendada por clone em cada UPER. Já na Tabela 8 é apresentada a recomendação clonal feita pela função softmax, com os dados preditos pelo modelo de regressão linear usando todas as variáveis de entrada, de acordo com o banco de dados de colheita de 2019-2023.

Tabela 7 – Recomendação clonal feita pelo profissional.

Clone	Chácara	Cruz Grande	Lagoa	Palmeiras	Pontal	São Bento	Total geral
AEC 0043	19	86	28	35	61	46	276
AEC 0144	39	43	34	40	36	39	230
AEC 1528	286	44	439	321	201	260	1550
AEC 2034	242		227	392	132	209	1202
AEC 2233	335	221	265	145	175	118	1259
AEC 2475	57	91	241	333	113	50	885
AEC 3151				7			7
AEC 3166					7		7
AEC 3481			4				4
AEC 3704	96			44	47	40	227
AEC 3709	5						5
AEC 4287				8			8
MIX Clonal	50						50
Novas espécies ou Novos clones	37	44	22	60	101	12	276
Total geral	1167	529	1259	1384	872	774	5985

Tabela 8 – Recomendação clonal feita pelo softmax.

Clone	Chácara	Cruz Grande	Lagoa	Palmeiras	Pontal	São Bento	Total geral
AEC 0043	24	24	24	24	24	24	142
AEC 0056	20	23	28	28	42	26	167
AEC 0144	71	85	68	125	136	101	585
AEC 1528	266	167	252	180	267	140	1272
AEC 2034	266	251	260	291	462	227	1756
AEC 2233	167	189	178	281	222	206	1242
AEC 2475	82	134	88	105	138	175	722
AEC 3166	2	2	2	2	2	2	14
AEC 3704	17	17	17	17	17	17	101
AEC 3709	9	9	9	9	9	9	51
Total geral	923	901	924	1060	1318	927	6053

Na Tabela 9, as produtividades dos clones Aperam BioEnergia LTDA são apresentadas de acordo com os dados do banco de dados da colheita para as diferentes UPERs. Sendo que, para os materiais genéticos que não constavam o valor da produtividade, foi atribuída a produtividade 30 m³/ha/ano. Esses dados foram usados para permitir uma comparação entre a assertividade das

recomendações manual, feita pelo especialista e a realizada pela função softmax, baseada no valor predito pelo modelo de regressão linear.

Tabela 9 – IMA por UPER de acordo com o banco de dados de colheita 2019 - 2023.

UPER	AEC 0043	AEC 0056	AEC 0144	AEC 1528	AEC 2034	AEC 2129	AEC 2233	AEC 2475	Outros	Total geral
UPER Chácara		31	34	29	33		35	42	30	30
UPER Cruz Grande	35	29	37	31	34	31	32	40	30	33
UPER Lagoa	21	28	30	27	30	30	29	32	30	28
UPER Palmeiras		24	26	23	27	25	25	32	30	25
UPER Pontal	35	31	35	28	31	27	33	37	30	30
UPER São Bento	32	28	26	31	31	29	28	38	30	30
Total geral	32	28	30	28	31	27	29	37	30	29

O IMA médio ponderado pela área obtido através da recomendação do profissional deu 30,3 m³/ha/ano já o obtido pelo softmax deu 30,7 m³/ha/ano. Só é possível verificar qual foi o mais correto, quando essa floresta for colhida em sete anos, mas optar pela versão que prevê maior produtividade em função da região é importante pela interação genótipo ambiente. As recomendações foram equivalentes em ganhos financeiros para a empresa. Porém o gasto de tempo para o especialista fazer a recomendação é muito maior, quando comparado com o tempo para processar os dados. Pois vários gráficos são construídos e uma Tabela que permita olhar a proporcionalização ao passo que os clones são alocados. Levando em média três dias de trabalho para fazer o que o sistema fez em segundos.

Khosa (2023) fizeram um estudo sobre comunicação e também usaram a predição e aplicaram o softmax para distribuição das informações, o uso dessa metodologia foi melhor que os resultados obtidos anteriormente.

Ambas as recomendações tiveram em comum a divisão da máxima quantidade de área possível por clone novo. Para que, em futuras recomendações, saiba-se em qual local ele é mais produtivo.

5. Conclusão

No presente trabalho, uma base de dados de colheita florestal foi analisada quanto à possibilidade de ser usada para prever produtividades futuras a partir de três diferentes metodologias: regressão linear, LASSO e árvore de decisão, para que a recomendação clonal fosse feita de forma a gerar mais valor para a empresa. Os objetivos foram alcançados ao obtermos valores de produtividades medias previstas pelo sistema similares aos valores obtidos pelo especialista em melhoramento florestal respectivamente, 30,3 e 30,7 m³/ha/ano, em menor tempo do que o tempo levado pelo especialista de melhoramento genético para executar a tarefa.

O trabalho trouxe a análise de seis modelos preditivos, considerando uma base de dados real que envolve 16 variáveis e 1903 amostras correspondentes ao período de cinco anos. Dentre as configurações testadas, os resultados mostram que o melhor modelo preditivo foi obtido usando a técnica de regressão linear e todas as 15 variáveis do banco de dados para prever a viável IMA m³/ha/ano. A função softmax também foi capaz de proporcionalizar as áreas produtivas por clone de forma a fornecer um resultado de IMA m³/ha/ano muito próximo ao resultado obtido pelo especialista em melhoramento genético em menor tempo, gerando um ganho de três dias de trabalho.

É importante enfatizar que a inclusão de dados, de clones novos, plantados em cada regional permitira predições mais assertivas, pois a recomendação deve ser baseada nos diferentes fenótipos possíveis. No dia a dia, a metodologia pode ser usada para fazer a primeira distribuição. Ou ser usada em discussões para ganho de produtividade conseguindo fazer a distribuição em menor tempo. Próximos trabalhos podem ser feitos com metodologias de predição e de maximização de produção, fazer análises de predição não lineares com os dados.

REFERÊNCIAS

- Azevedo, G. B., Tomiazzi, H. V., Azevedo, G. T. O. S., Teodoro, L. P. R., Teodoro, P. E., Souza, M. T. P., Batista, T. S., Eufraide-Junior, H. J., Guerra, S. P. S., Multi-volume modeling of Eucalyptus trees using regression and artificial neural networks, *PloS one*, v. 15, n. 9, p. 0238703, (2020).
- Blanchard, P. , Higham, . D.J. ,Higham, . N.J. “Accurately computing the log-sum-exp and softmax functions”, *IMA Journal of Numerical Analysis*, V. 41, Issue 4, October 2021, Pages 2311–2330.
- Braga, E. P. C., & Prates, M. H. D. A.. Desenvolvimento de um framework de redes neurais com interface gráfica. 2020. Trabalho de Conclusão de Curso. Universidade Tecnológica Federal do Paraná.
- Brosofske, K. D., Froese, R. E., Falkowski, M. J., & Banskota, A., A review of methods for mapping and prediction of inventory attributes for operational forest management. *Forest Science*, v. 60, n. 4, p. 733-756, (2014).
- Centro de Estudos Avançados Em Economia Aplicada (CEPEA) e Confederação Nacional Da Agricultura e Pecuária (CNA). PIB do agronegócio brasileiro de 1996 a 2023.
- Costa, E.C; D’Avila, M; Cantarelli, E. B; Murari, A. B; Manzoni, C.G. *Entomologia Florestal*. Santa Maria: Editora UFSM. 2008. 240 p.
- Dayan, P. and Hinton, G. (1993). Feudal reinforcement learning, *Advances in Neural Information Processing*.
- Assis, Teotônio Francisco. Melhoramento genético de Eucalyptus: desafios e perspectivas. 3º Encontro Brasileiro de Silvicultura, v. 3, p. 127-148, 2015.
- Diniz, J. M. F. S., Gama, F. F., Reis, A. A, Oliveira, C. G., Marques, E. R. G., Estimating stem volume of *Eucalyptus sp.* and *Pinus sp.* plantations in Brazil, using Sentinel-1B and ALOS-2/PALSAR-2 data; *Journal of Applied Remote Sensing*, Vol. 17, Issue 1, 014513 (February 2023).
- Fernandes, I. K., Vieira, C. C., Dias, K. O. G., Fernandes, S. B., Using machine learning to integrate genetic and environmental data to model genotype-by-environment interactions, (2024).
- Gomes da Silva, D. C. Predição de Estresse em Ovelhas Prenhas e Lactantes com o Uso de Redes Neurais Artificiais. *Ensaio e Ciência C Biológicas Agrárias e da Saúde*, [S. l.], v. 25, n. 2, p. 160–165, (2021).
- Guo, T., LI, X., Machine learning for predicting phenotype from genotype and environment. *Current Opinion in Biotechnology*, v. 79, p. 102853, 2023.
- IBÁ (2022). Relatório Anual IBÁ - Indústria Brasileira de Árvores, 2022.
- Júnior, I. D. S. T., Rocha, J. E. C. D., Ebling, Â. A., Chaves, A. D. S., Zanuncio, J. C., Farias, A. A., Leite, H. G., Artificial neural networks and linear regression reduce sample intensity to predict the commercial volume of Eucalyptus clones. *Forests*, v. 10, n. 3, p. 268, (2019).
- Khosa, S., R. Furqan., Mehmood, A., Choi, G. S., Ashraf, I., Incorporating Word Embedding and Hybrid Model Random Forest Softmax Regression for Predicting News Categories, *Springer Mayo*, p.1-17, (2023).
- Kontschieder, P., Fiterau, M., Criminisi, A., & Buló, S. R., Deep neural decision forests. In: *Proceedings of the IEEE international conference on computer vision*. p. 1467-1475, (2015).

- Li, H., Ai, Z., Yang, L., Zhang, W., Yang, Z., Peng, H., & Leng, L. Machine learning assisted predicting and engineering specific surface area and total pore volume of biochar. *Bioresource Technology*, v. 369, p. 128417, (2023).
- Martins, M. T., Marangon, G., Costa, E. A., Pfeifer, M. I., Moreira, Y. V. M., . Estimaco da altura de plantios florestais de eucalipto dor regresso e redes neurais artificiais. *BIOFIX Scientific Journal*, v. 5, n. 1, p. 141, 3 dez. (2019).
- Neto, E. M. C., Moura, M. M., Araujo, E. C. G., Santana, G. M., Corte, A. P. D., Sanquetta, C. R., Aprendizado de Mquina e Regresso Linear na Estimativa do Volume de Eucalyptus na Amaznia Oriental; *Biofix Scientific Journal* v. 6 n. 1 p. 08-14 (2021).
- Ranstam, J., Cook, J. A., LASSO regression, *British Journal of Surgery*, V. 105, n. 10, Page 1348, September (2018).
- Rifkin, Ryan M.; Lippert, Ross A. Notes on regularized least squares. 2007.
- S. -J. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, "An Interior-Point Method for Large-Scale ℓ_1 -Regularized Least Squares," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606-617, Dec. 2007.
- Santana, D. C., Santos, R. G. d., Silva, P. H. N., Pistori, H., Teodoro, L. P. R., Poersch, N. L., Azevedo, G. B., Azevedo, G. T. O. S., Junior, C. A. S., Teodoro, P.E., Machine Learning Methods for Woody Volume Prediction in Eucalyptus. *Sustainability*, v. 15, n. 14, p. 10968, (2023).
- Scolforo, P. V., Scolforo, H. F., Scolforo, J. R., McTague, J. P., Gomide, L. R., Silva, S. J., Alto, T.F., Pscoa, K. J. V., Coelho, P. A., A New Approach for Modeling the Aboveground Carbon Stock in the Brazilian Savanna, Available at.
- Silva, B. C., Costa, E. C., Saldanha, M. A., Procknow, D., Souza, P. D. de, Croda, J. P., Capitani, L. C. . Mtodos de controle e preveno de insetos-praga em povoamentos florestais / Methods of pest insect control and prevention in forest stans. *Brazilian Journal of Development*, 6(7), 48477–48496, (2020).
- Soares, T. S., Leite, H. G., Vale, A. B. do., Soares, C. P. B., Silva, G. F., Avaliaco de um modelo de passo invariante na predico da estrutura de um povoamento de *Eucalyptus sp.*. *Revista rvore*, 31(2), 275–283 (2007).
- Teixeira, M., Berra, E. F., Mallmann, A. A., Corte, A. P. D., Kruchelski, S., Imagens Planetscope na Estimativa da Biomassa acima do Solo em Sistemas ILPF. In: *Anais do XX Simpsio Brasileiro de Sensoriamento Remoto, 2023, Florianpolis*.
- Valente, E. de O.; Valente, G. de F. S. Simulao de redes neurais artificiais para estimativa de volume de madeira florestal a partir do DAP / Simulation of artificial neural networks for estimation of forest wood volume from DAP. *Brazilian Journal of Animal and Environmental Research*, [S. l.], v. 4, n. 3, p. 3748–3757, (2021).
- Vera, D. E., Lauraq, V. A., Ferreira, A. D., Couto, A. M., Crescimento e forma do eucalipto em funo da densidade de plantio. *Cincia Florestal*, Vol 32, Issue 1, p.0103-9954 (2022).
- Wang, J., Xu, J., Peng, Y., Wang, H., Shen, J., Prediction of forest unit volume based on hybrid feature selection and ensemble learning, v. 13, p. 21-32, (2020).