

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL
CURSO DE DOUTORADO

**MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA
PROTEÍNA BRUTA DE PASTOS DE CAPIM-TAMANI: DESENVOLVIMENTO,
AVALIAÇÃO E INTERPRETABILIDADE COM A TÉCNICA SHAP**

GABRIELA OLIVEIRA DE AQUINO MONTEIRO

CAMPO GRANDE, MS

2025

MONTEIRO, G.O.A.	MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA PROTEÍNA DE PASTOS DE CAPIM- TAMANI: DESENVOLVIMENTO, AVALIAÇÃO E INTERPRETABILIDADE COM SHAP	2025
---------------------	---	------

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL
CURSO DE DOUTORADO

**MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA
PROTEÍNA BRUTA DE PASTOS DE CAPIM-TAMANI: DESENVOLVIMENTO,
AVALIAÇÃO E INTERPRETABILIDADE COM SHAP**

MACHINE LEARNING MODELS FOR PREDICTING CRUDE PROTEIN FROM
TAMANI GRASS PASTURES: DEVELOPMENT, EVALUATION AND
INTERPRETABILITY WITH SHAP

Gabriela Oliveira de Aquino Monteiro

Orientador: Prof. Dr. Gelson dos Santos Difante

Coorientadores: Prof. Dr. Edson Takashi Matsubara

Dra. Jéssica Gomes Rodrigues

Tese apresentada à Universidade
Federal de Mato Grosso do Sul,
como requisito à obtenção do título
de Doutora em Ciência Animal.

Área de concentração: Produção
Animal

CAMPO GRANDE, MS

2025



Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



Certificado de aprovação

GABRIELA OLIVEIRA DE AQUINO MONTEIRO

**MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA PROTEÍNA BRUTA DE
PASTOS DE CAPIM-TAMANI: DESENVOLVIMENTO, AVALIAÇÃO E INTERPRETABILIDADE COM
A TÉCNICA SHAP**

**MACHINE LEARNING MODELS FOR PREDICTING CRUDE PROTEIN FROM TAMANI GRASS
PASTURES: DEVELOPMENT, EVALUATION AND INTERPRETABILITY WITH SHAP**

Tese apresentada à Universidade Federal de Mato Grosso do Sul, como requisito para obtenção do título de Doutora em Ciência Animal. Área de concentração: Produção Animal.

Aprovado em: 30-05-2025

BANCA EXAMINADORA:

Dr. Gelson dos Santos Difante
(UFMS) – Presidente

Dra. Ana Cláudia Ruggieri
(UNESP)

Dr Edson Takashi Matsubara
(UFMS)

Dr. Luis Carlos Vinhas Itavo
(UFMS)

Dra. Valéria Pacheco Batista Euclides
(EMBRAPA)

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Ana Cláudia Ruggieri, Usuário Externo**, em 30/05/2025, às 14:24, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Gelson dos Santos Difante, Professor do Magisterio Superior**, em 30/05/2025, às 14:34, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Luis Carlos Vinhas Itavo, Professor do Magisterio Superior**, em 30/05/2025, às 14:34, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Valeria Pacheco Batista Euclides, Usuário Externo**, em 30/05/2025, às 15:21, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Edson Takashi Matsubara, Professor do Magisterio Superior**, em 30/05/2025, às 15:46, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufms.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5658655** e o código CRC **316DE6EF**.

COLEGIADO DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

Av Costa e Silva, s/nº -

Cidade Universitária Fone:

CEP 79070-900 - Campo Grande - MS

Referência: Processo nº 23104.001236/2021-33
nº 5658655

SEI

*Dedico ao meu marido Felipe Soares Santos
por ser meu esteio e fonte de amor.*

AGRADECIMENTOS

Agradeço profundamente a Deus por ter me amparado até aqui com saúde e sua constante proteção.

Aos meus pais Flavia O. Aquino e Antonio J. Pereira por não medirem esforços para que eu pudesse ter acesso a uma educação de qualidade. Vocês são meus exemplos de integridade.

Ao meu marido Felipe Soares Santos, por ser tão companheiro, atencioso e por me apoiar e alegrar nos momentos difíceis. Amo você.

À minha irmã Ana Clara O. de Aquino Pereira, que mesmo distante sempre se fez presente com palavras de carinho e incentivo. Aos meus irmãos Mariene G. Pereira e Antonio M. G. Pereira pelo apoio e momentos de alegria.

Ao meu avô Izolino Alves Aquino (*in memoriam*) por ter sido tão amável e me ensinar a contemplar o belo e o simples da vida.

Ao meu orientador, Prof. Dr. Gelson dos Santos Difante, minha profunda gratidão por todas as oportunidades, pela amizade e pelo constante incentivo. Seu apoio e orientação foram fundamentais na minha formação profissional.

Ao Prof. Dr. Edson Takashi Matsubara, obrigada pela oportunidade em trabalhar com sua equipe, por compartilhar seus conhecimentos. Agradeço por cada orientação na construção deste trabalho, pelas reuniões e ideias enriquecedoras.

À Dra. Jéssica Gomes Rodrigues, minha sincera gratidão pelos direcionamentos precisos, pelo apoio constante e pelos momentos de alegria e carinho ao longo desta jornada.

Ao Prof. Dr. Anderson Bessa da Costa, minha gratidão por sua contribuição na construção desta tese. Os insights foram fundamentais para o desenvolvimento deste trabalho.

À Dra. Denise Baptaglin Montagner, por ser uma profissional que tanto admiro e me espelho.

À Universidade Federal de Mato Grosso do Sul (UFMS), em especial ao Programa de Pós-graduação em Ciência Animal (PPGCA) e a todos os professores que ensinaram e contribuíram com minha formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pela concessão de bolsa.

À EMBRAPA Gado de Corte e UNIPASTO, por toda infraestrutura cedida (de laboratórios e campo) e financiamento da pesquisa.

Em especial à Marislaine Gusmão Pereira, Juliana Caroline Santos Santana, Hitalo Rodrigues da Silva, Jecelen Campos e Rafael Torres pela amizade, momentos de alegria, paciência e disponibilidade em compartilhar seus conhecimentos. A todos que contribuíram e contribuem para meu desenvolvimento pessoal e profissional, muito obrigada.

“O correr da vida embrulha tudo,
a vida é assim: esquenta e esfria,
aperta e daí afrouxa, sossega e depois desinquieta.
O que ela quer da gente é coragem.”

Guimarães Rosa

RESUMO

MONTEIRO, G.O.A. Modelos de aprendizado de máquina para predição da proteína bruta de pastos de capim-tamani: desenvolvimento, avaliação e interpretabilidade com Shap. 2025. Tese (Doutorado) - Faculdade de Medicina Veterinária e Zootecnia, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, 2025.

Conhecer a qualidade da forragem é essencial para atender às exigências dos animais e otimizar a produção. Os métodos tradicionais para aferir o valor nutritivo do pasto são destrutivos, onerosos e demorados. Modelos de aprendizado de máquina (AM) tem sido utilizado em tarefas de predição na agropecuária. No entanto, compreender como o modelo faz uma predição é tão importante quanto predizer. O estudo teve como objetivos: i) testar a aplicabilidade de modelos de aprendizado de máquina com dados tabulares como variáveis climáticas, interceptação de luz (IL), doses de nitrogênio (N), intervalos entre pastejo (IEP) e alturas de pré (ALTpre) e pós-pastejo (ALTpos) para predição do teor de proteína bruta das folhas (PB) de pastos de capim-tamani; ii) utilizar a técnica SHAP para interpretar como essas variáveis atuam no modelo XGBoost. Foi utilizado um banco de dados robusto coletado de área de pastagem formada com capim-tamani manejado intensivamente no bioma Cerrado brasileiro de outubro de 2020 a abril de 2022. O conjunto de dados foi composto por 90 instâncias com 80% para treino e validação e 20% para teste com validação cruzada com 5 folds e 100 iterações. Os hiperparâmetros foram ajustados com *grid-search* no conjunto de treino. Foram testados Regressão Linear (RL), Multilayer Perceptron (MLP), Decision Trees (DT), Random Forest (RF) e XGBoost. A técnica Shapley Additive Explanations (SHAP) foi utilizada para entender a força e direção de cada variável na predição do modelo. As métricas de desempenho obtidas foram coeficiente de correlação de Pearson (r), coeficiente de determinação (R^2) e o erro absoluto médio (MAE). Os modelos MLP ($r=0.75$, $R^2=43.94\%$, $MAE=1.56$), RF ($r=0.78$, $R^2=49.23\%$, $MAE=1.60$) e XGBoost ($r=0.78$, $R^2=56.90\%$, $MAE=1.45$) apresentaram os melhores resultados de predição ($p<0.001$). As variáveis com maior importância preditora do teor de PB são IEP, seguido de doses de N, ALTpre e ALTpos. XGBoost se destacou entre os modelos testados ($p<0.001$). O impacto de IEP depende da interação com outras variáveis. A maior dose de N resulta em impacto positivo sobre o teor de PB das folhas de capim-tamani. Observou-se que precipitações entre 100 e 180 mm possuem impacto positivo no modelo, elevando a PB e valores fora desse intervalo possuem impacto negativo sobre teor de PB das folhas. O uso de 240 kg/ha/ano de N em pastos de capim-tamani reduz a influência das condições ambientais sobre o teor de PB das folhas. Quando se utiliza 80 kg/ha/ano de N, o teor de PB das folhas do capim-tamani sofre maior influência da precipitação. O uso de dados tabulares como doses de N, IEP, ALTpre, ALTpos, IL e variáveis climáticas é uma alternativa viável a predição de PB em AM. Os resultados reforçam a importância do manejo integrado das alturas, metas de manejo do pastejo, adubação nitrogenada, intervalo entre pastejos como estratégias-chave para elevar a proteína bruta das folhas de capim-tamani.

Palavras-chave: algoritmos, machine learning, manejo de pastagens, *Panicum maximum*, precision livestock farming, Shapley Additive Explanations.

Sumário

Introdução Geral	9
Capítulo 1: Revisão de Literatura	11
Evolução das tecnologias de manejo de pastagens	11
Valor nutritivo e composição química do pasto	14
Aprendizado de máquina	16
SHAP (SHapley Additive Explanations)	18
Referências	21
Capítulo 2. Modelos de aprendizado de máquina para predição de proteína bruta de pastos de capim-tamani	26
Introdução	27
Material e Métodos	29
Aquisição de dados	29
Machine Learning	31
Análise estatística	34
Resultados	35
Respondendo RQ1: Como os atributos se relacionam linearmente?	35
Respondendo RQ2 - Entre os algoritmos clássicos de AM qual é aquele mais adequado para predição de proteína bruta?	37
Respondendo RQ 3 – Quais são as variáveis mais relevantes para os melhores modelos?	38
Conclusão	44
Referências	45
Capítulo 3. Interpretabilidade do modelo de aprendizado de máquina usando valores SHAP: aplicação a uma tarefa de predição de proteína bruta de pastos de capim-tamani	50
Introdução	51
Material e Métodos	52
Descrição dos dados	52

SHapley Additive exPlanations (SHAP)	53
Resultados.....	54
Respondendo P.P. 1: Qual é a contribuição individual da precipitação, radiação solar e temperatura na predição do teor de PB da folha de capim-tamani, desconsiderando o efeito da estação do ano?	55
Respondendo P.P. 2: Qual faixa de precipitação exerce impacto positivo na predição do teor de PB de capim-tamani?	58
Respondendo P.P. 3: As doses de nitrogênio influenciam a hierarquia de importância das variáveis de manejo e ambientais na predição do teor de proteína bruta da folha de capim-tamani?	61
Discussão	63
Conclusão	67
Referências	67
Considerações Finais	71

1 **Introdução Geral**

2 As áreas de pastagens correspondem a aproximadamente 70% das terras agrícolas
3 do mundo (Squires et al., 2018). Em escala mundial, a pecuária é predominantemente
4 baseada no uso de forragens como principal fonte de alimentação para o rebanho (Lussem
5 et al., 2019). No Brasil, aproximadamente 154 milhões de hectares de pastagens são
6 destinados a produção de ruminantes. Assim, as pastagens exercem papel importante no
7 abastecimento de proteína de origem animal no mundo (Reinermann et al., 2020) e
8 contribuem para a segurança alimentar (Tripathi et al., 2018). O manejo desse ecossistema
9 permite otimizar a eficiência produtiva e sustentabilidade da produção animal além de
10 sequestrar carbono, conservar a biodiversidade e promover a mitigação das mudanças
11 climáticas (O'Mara, 2012; Zhao et al., 2020).

12 Com a intensificação da pecuária, o uso de plantas forrageiras de alta
13 produtividade e elevado valor nutritivo é fundamental para atender à demanda alimentar
14 dos ruminantes (Mwendia et al., 2022). Nesse cenário, as cultivares do gênero *Panicum*
15 se destacam (Caldas et al., 2019), quando bem manejadas. A cultivar BRS Tamani foi
16 lançada em 2015 como opção de forrageira de porte baixo, abundância de folhas, elevados
17 teores de proteína bruta e digestibilidade in vitro da matéria orgânica (Embrapa, 2015). O
18 sucesso na utilização de gramíneas tropicais depende das características da planta,
19 suprimento de nutrientes, manejo e interação com o meio. A produtividade dos sistemas
20 depende do potencial produtivo da planta forrageira e de sua qualidade, sendo os manejos
21 da pastagem e do pastejo fundamentais para otimizar tanto a produção forrageira quanto
22 o desempenho animal (Delevatti et al., 2019a).

23 Sistemas de produção animal em pastagens eficientes e sustentáveis exigem o
24 monitoramento regular da quantidade e valor nutritivo da forragem pois a essência do
25 manejo do pastejo consiste em equilibrar o crescimento da planta, seu consumo e o
26 desempenho animal (Hodgson, 1990). Portanto, conhecer a composição química do pasto
27 permite ajustar estratégias de manejo, além de otimizar o desempenho animal (Rouquette,
28 2016). No entanto, as metodologias disponíveis para avaliar a composição da forragem
29 são demoradas, caras e requerem grande quantidade de mão de obra (Horrocks et al.,
30 2020; t'Mannetje et al., 2000), o que as tornam impraticáveis para fornecer informações
31 em tempo real sobre a dieta animal e dar suporte as decisões de manejo (Safari et al.,
32 2016).

Nesse cenário, as novas tecnologias desenvolvidas com o advento da pecuária de precisão podem fornecer informações precisas e rápidas para apoiar tomadas de decisão no manejo de pastagens. A modelagem preditiva baseada em aprendizado de máquina (AM) tem sido utilizada para sistemas de produção em pastagens (Bretas et al., 2021; Dos Reis et al., 2020; Murphy et al., 2022). Porém, a maioria dos estudos utilizam imagens provenientes de satélites, veículos aéreos não tripulados (VANT) e sensores. Segundo Stump et al. (2023), a compreensão sobre a eficácia das técnicas de AM ainda é limitada e há poucas evidências sobre a relação entre fatores bióticos e abióticos na modelagem preditiva. A inclusão de variáveis não espectrais, como dados meteorológicos, possibilita aumentar a precisão dos modelos (Jaffrés, 2019; Farmers Guide, 2022). Além disso, são escassos os estudos que consideram componentes nutritivos da forragem, como o teor de proteína bruta (PB) (Bretas et al., 2023) e que considerem a modelagem explicativa (Shmuel, 2010).

Dessa forma, dados tabulares que são rotineiramente obtidos nos protocolos experimentais e envolvem variáveis climáticas, de manejo, estações do ano e o teor de proteína bruta da folha do capim-tamani aliado ao aprendizado de máquina podem fornecer informações rápidas e precisas que auxiliem tomadas de decisão em sistemas de produção animal em pastagens.

Diante disso, as hipóteses testadas são: (1) é possível prever a PB de capim-tamani com dados tabulares e (2) é possível explicar como o modelo prevê esta variável. O objetivo foi avaliar a aplicabilidade de modelos de aprendizado de máquina com dados tabulares para predição do teor de proteína bruta da folha (PB) de pastos de capim-tamani e compreender como o modelo prevê esta variável.

Capítulo 1: Revisão de Literatura

Evolução das tecnologias de manejo de pastagens

O Brasil possui aproximadamente 154 milhões de hectares de pastagens (MapBiomass, 2024) e a produção animal baseada no recurso forrageiro representa um componente essencial da agropecuária nacional, considerando a vasta extensão territorial do país e a importância do agronegócio para sua economia (Martuscello et al., 2019). Nesse contexto, o manejo de pastagens constitui um dos pilares da pecuária brasileira, influenciando diretamente indicadores zootécnicos, ambientais e econômicos e sendo determinante para a sustentabilidade dos sistemas de produção.

Historicamente, a substituição de pastagens nativas, geralmente de baixo valor nutritivo, por forrageiras cultivadas e geneticamente melhoradas aumentou significativamente a produtividade da pecuária no país (Dias Filho, 2015). Entre os principais avanços obtidos por meio da pesquisa científica, destaca-se o melhoramento genético de gramíneas tropicais, especialmente do gênero *Panicum*. O lançamento das cultivares Tanzânia e Mombaça, a partir da década de 1990, representou um marco na intensificação do uso das pastagens, impulsionado pela alta produção de forragem e pelo bom valor nutritivo desses materiais. Nos anos seguintes, novas cultivares como Aruana, Massai, BRS Zuri, BRS Tamani e BRS Quênia foram incorporadas ao portfólio forrageiro nacional, ampliando as opções disponíveis os diversos sistemas produtivos.

Além dos avanços no melhoramento genético, houve uma mudança na forma como o manejo de pastagens passou a ser estudado. O foco na ecofisiologia de gramíneas tropicais e na ecologia do pastejo passaram a orientar os protocolos experimentais, contribuindo para uma compreensão mais aprofundada sobre os processos que regulam o crescimento das plantas forrageiras e sua interação com os animais (Da Silva & Nascimento Jr., 2007). Nesse contexto, um dos principais marcos foi a adoção do conceito de índice de área foliar crítico (IAF), que estabelece como condição ideal de pré-pastejo o momento em que o dossel forrageiro intercepta 95% da luz incidente. Nessa fase, observa-se elevada produção de forragem, com alta proporção de folhas, reduzida presença de colmos e baixo acúmulo de material morto (Carnevalli et al., 2006; Barbosa et al., 2007; Pedreira et al., 2007). Diversos estudos com pastagens tropicais,

especialmente sob regime de lotação intermitente, reforçam que esse nível de interceptação de luz (IL) representa o estágio mais adequado para interromper a rebrota (Euclides et al., 2022).

Com o objetivo de tornar o conceito de interceptação de IL mais prático, Carnevalli (2006) e Barbosa (2007) demonstraram que a altura do dossel no momento do pré-pastejo está diretamente relacionada ao nível de IL, podendo, assim, ser utilizada como um critério eficaz para o controle da rebrotação e do manejo do pastejo. Foi evidenciado que o nível de 95% de IL, considerado ideal para a interrupção da rebrota varia conforme a espécie, cultivar, condições sazonais e adubação, influenciando diretamente os intervalos entre pastejos. As alturas correspondentes a esse nível foram de 90 cm para *Panicum maximum* cv. Mombaça (Carnevalli, 2006), 70 cm para cv. Tanzânia (Barbosa et al., 2007) e 30 cm para cv. Xaraés (Pedreira et al., 2009). Essas variações reforçam a importância do conhecimento da ecofisiologia de plantas forrageiras, uma vez que práticas de manejo afetam de maneira distinta o crescimento, o perfilhamento e a estrutura do dossel, impactando diretamente a produtividade das pastagens (Sbrissia et al., 2007).

Além disso, a intensidade de pastejo deve ser ajustada de modo que, ao final da desfolha, o resíduo remanescente contenha quantidade suficiente de tecido fotossinteticamente ativo. Essa condição é essencial para garantir a sobrevivência da planta, promover uma rebrota rápida e favorecer o acúmulo de forragem para o próximo ciclo de pastejo (Costa et al., 2017). Dessa forma, o manejo do pastejo passou a ser considerado um componente central para a sustentabilidade dos sistemas pastoris, por equilibrar o crescimento do pasto com o consumo animal e a produção zootécnica (Euclides et al., 2021).

Pesquisas mais recentes sugerem ainda que a adoção de 90% de IL como critério para o pastejo animal permitindo maior flexibilidade ao manejo e melhor aproveitamento dos pastos, principalmente nos períodos de transição. Alvarenga et al. (2020) demonstraram que pastagens de capim-mombaça podem ser manejadas com IL entre 90% e 95%, correspondendo a alturas de 80 a 90 cm, com estrutura do dossel e valor nutritivo semelhante. Gurgel et al. (2017) avaliaram o capim-massai sob diferentes metas de IL (90% e 95%) e duas alturas de pós-pastejo (15 e 25 cm). Os autores concluíram que a altura de resíduo de 25 cm, durante o período chuvoso, garantiu melhor estrutura do pasto para o período seco, independentemente da meta de pré-pastejo adotada. Zanine et al.

(2020) destacaram que, no capim-piatã (*Brachiaria brizantha*), a meta de 90% de IL aliada à adubação nitrogenada favorece o controle do acúmulo de colmos e material morto, aumentando a eficiência de colheita e o número de ciclos de pastejo.

Além das práticas de manejo baseadas em IL, a adubação nitrogenada tem se destacado como estratégia eficaz para elevar a produtividade das pastagens. Diversos estudos relatam aumentos na produção de forragem com o uso de nitrogênio (Yasuoko et al., 2018; Galindo et al., 2018). Gomes et al. (2020) avaliaram as cultivares BRS Tamani e BRS Quênia submetidas em duas doses de nitrogênio observaram que 300 kg/ha/ano de N resultou em maior número de perfilhos por metro quadrado quando comparado a dose de 200 kg/ha/ano.

Nos últimos anos, a introdução de tecnologias digitais por meio da pecuária de precisão tem ganhado destaque nas pesquisas. Essa abordagem visa otimizar o desempenho animal e a eficiência dos sistemas produtivos por meio do monitoramento em tempo real de variáveis relacionadas ao solo, planta, animal e clima. Ferramentas como sensores de proximidade, estações meteorológicas automáticas, drones e softwares de apoio à tomada de decisão vêm sendo cada vez mais empregadas. A utilização de imagens obtidas por satélites, drones ou sensores permite a coleta de dados em escala espacial e temporal elevada. Esses dados podem ser integrados a modelos preditivos e sistemas de apoio à decisão, possibilitando ajustes mais precisos nas práticas de manejo.

Diversos estudos têm demonstrado o potencial dessas tecnologias para estimar variáveis-chave do manejo, como altura do dossel e biomassa. Batistoti et al. (2019) encontraram $R^2=0.81$ para predição de altura e $R^2=0.80$ para biomassa de capim-tamani com o uso de veículo aéreo não tripulado (VANT). Dos Reis et al. (2020) demonstraram que imagens PlanetScope, combinadas com medidas de textura e o algoritmo XGBoost, permitem estimar a biomassa ($R^2=0.65$) e a altura do dossel ($R^2=0.68$) em pastagens manejadas intensivamente. Em contrapartida, Rueda-Ayala et al. (2019) relataram correlações fracas ($R < 0,6$) entre dados espectrais nas bandas vermelho, verde e azul (RGB) e a altura do pasto em azevém-perene, atribuindo a baixa precisão a interferências ambientais como vento, variações de luminosidade solar e cobertura de nuvens.

Resultados promissores também foram verificados por Bretas et al. (2023), que utilizaram imagens do satélite Sentinel-2 para classificar a altura do dossel e estimar a biomassa de capim-mombaça, com desempenho preditivo moderado para biomassa fresca

($R^2=0,61$) e concentração de matéria seca ($R^2=0,69$). Bretas et al. (2020) ao avaliar imagens dos satélites Landsat-8 e Sentinel-2 para prever a biomassa de pastagens de *Brachiaria* encontraram bom desempenho para biomassa fresca ($R^2=0.85$). Defalque et al. (2024) encontraram desempenho de predição de biomassa ($R^2=0.76$) e o teor de matéria seca ($R^2=0.77$) de pastagens de *Brachiaria brizantha* cv. Marandu com imagens do satélite Sentinel 2. Apesar do potencial evidenciado, ainda existem lacunas a serem preenchidas e desafios para facilitar a adoção da tecnologia de forma prática e acessível ao produtor.

Existem limitações quanto ao uso eficaz do sensoriamento remoto, principalmente devido à alta heterogeneidade espacial e temporal desses sistemas, com folhas verdes e senescentes que afeta significativamente a resposta espectral das pastagens e o desafio para alcançar resolução espacial adequada para detectar variações relevantes ao manejo, pois sensores com alta precisão são caros (Murphy et al., 2020). Para que os métodos de estimativa de produtividade sejam amplamente adotados, é fundamental que apresentem baixo custo e sejam de fácil aplicação prática pelos produtores (Matos et al., 2022). Nesse cenário, observa-se uma subexploração do potencial de dados tabulares na predição de variáveis de manejo em sistemas pastoris. Além disso, ainda há escassez de tecnologias consolidadas que permitam medir, em tempo real, a qualidade da forragem (Murphy et al., 2020). O conhecimento do valor nutritivo em tempo real possibilita otimizar o processo de pastejo e o manejo nutricional do rebanho (Martelli et al., 2025).

Valor nutritivo e composição química do pasto

O valor nutritivo da forragem engloba a composição química, que abrange as concentrações de matéria seca (MS), proteína bruta (PB), fibra em detergente neutro (FDN), fibra em detergente ácido (FDA), lignina, extrato etéreo, carboidratos não estruturais e minerais e a digestibilidade desses componentes (Balseca et al., 2015).

Existem métodos destrutivos e não destrutivos para estimar a qualidade da forragem. Geralmente os produtores utilizam critérios visuais como a cor e estágio de maturidade da planta (Wijeshingha et al., 2020). O método tradicional é feito de forma destrutiva e após a coleta de amostras no campo, são realizadas análises químicas via úmida ou por espectroscopia de refletância no infravermelho próximo (NIRS) em laboratórios (Horrocks et al., 1999). No caso da tecnologia NIRS, é preciso fazer calibrações específica para cada cultivar utilizada (Schellberg et al., 2008). Determinar a

composição química e a digestibilidade de gramíneas tropicais com análises químicas e NIRS, requer tempo, grande quantidade de mão de obra e são dispendiosos (t'Mannetje et al., 2000), dificultando seu uso na prática.

No entanto, o manejo eficiente das pastagens também depende das estratégias de suplementação alimentar, que depende do conhecimento preciso sobre o valor nutritivo da forragem consumida pelos animais (Tedeschi et al., 2019). Dentre os componentes do valor nutritivo, o teor de proteína da forragem é um parâmetro fundamental na nutrição de ruminantes, pois fornece os aminoácidos necessários à síntese de tecidos e enzimas, ao metabolismo animal, ganho de peso e produção de carne e leite (Van Soest, 1994). Além disso, a proteína da forragem exerce papel crucial no funcionamento do rúmen, pois os microrganismos utilizam a fração proteica para a síntese de proteína microbiana que constitui a principal fonte de aminoácidos para os ruminantes (NRC, 2001). Quando a quantidade de proteína na dieta é insuficiente, menor que 7%, a degradação ruminal é prejudicada, diminuindo a digestibilidade da fibra e, conseqüentemente, o consumo de matéria seca (Van Soest, 1994).

Em pastagens tropicais, a composição química da forragem pode variar em função da cultivar, maturidade da planta e fatores ambientais (Lopes et al., 2017). O capim-tamani é uma planta forrageira que apresenta elevado valor nutritivo, em sua composição química, os teores de PB variam de 10 a 22% e 8 a 14% da matéria seca, respectivamente, na estação chuvosa e seca (Pasto Certo, 2021).

A maturidade da planta também influencia sua composição química pois a concentração dos principais constituintes orgânicos varia em função do estágio de desenvolvimento (Hodgson, 1990). Plantas jovens apresentam elevada qualidade, mas com o amadurecimento, sua qualidade tende a diminuir (Lacerda et al., 2009). Isso ocorre porque conforme a maturidade da planta aumenta, há aumento dos componentes fibrosos e diluição dos teores de PB (Zubieta et al., 2021).

Dentre os fatores abióticos, a radiação solar influencia o crescimento, produtividade e qualidade das cultivares forrageiras, pois afeta diretamente a fotossíntese e a composição química. A quantidade de luz interceptada pelo dossel forrageiro possui relação direta com a fotossíntese e assimilação de nitrogênio, incluindo proteínas principalmente nos estágios iniciais de crescimento (Da Silva & Nascimento Júnior, 2007). Em condições de alta radiação solar, o crescimento acelerado da gramínea pode

diluir a concentração de PB devido ao aumento na deposição de carboidratos estruturais e lignina (Van Soest, 1994). Além disso, o alongamento de colmos sob alta luminosidade pode reduzir a proporção de folhas, que possuem o maior teor proteico quando comparado ao colmo.

Outro fator importante é a temperatura que, de maneira geral, com seu aumento ocorre maior lignificação da parede celular reduzindo a qualidade das folhas e colmos (Van Soest, 1994). Além disso, a composição dos constituintes das plantas está diretamente relacionada à fertilidade do solo e à disponibilidade de nutrientes fornecidos pela adubação que, geralmente, contribui para a melhoria da qualidade da forragem, desde que seja associada a práticas de manejo que garantam a colheita no momento ideal (Hodgson, 1990). Nesse cenário, a adubação nitrogenada possibilita melhoria no valor nutritivo em gramíneas tropicais (Campos et al., 2016), que é corroborado por Paciullo et al. (2017), que observaram aumento dos teores de PB com o aumento das doses de nitrogênio.

A disponibilidade hídrica também é um fator essencial para o crescimento e qualidade das gramíneas e influencia o teor de PB. A água é fundamental para absorção e transporte de nutrientes do solo, incluindo o nitrogênio, elemento importante na síntese de proteínas. Em condições de baixa precipitação, a absorção de nitrogênio pelas raízes pode ficar comprometida devido à baixa umidade no solo, o que resulta em menor conteúdo de PB na planta (Silva et al., 2020). Por outro lado, disponibilidade hídrica adequada pode favorecer a absorção de nutrientes favorecendo a síntese de aminoácidos e proteínas. No entanto, essa condição também pode estimular o crescimento da cultivar forrageira, com alongamento de colmo e maior lignificação da parede celular.

Diante disso, as variações nos processos de crescimento das plantas forrageiras em resposta a alterações na radiação solar, temperatura, disponibilidade hídrica, fornecimento de nutrientes e a interação entre esses fatores influenciam a qualidade da forragem. Identificar como esses fatores interagem entre si possibilita estabelecer estratégias de manejo para atender as exigências nutricionais de animais em pastejo.

Aprendizado de máquina

O aprendizado de máquina (AM) é uma subárea da pesquisa da inteligência artificial (IA), esta por sua vez, tem o objetivo de criar máquinas que resolvem problemas que, quando solucionados por humanos, exigem inteligência (Grosan e Abrahan, 2011).

O AM desenvolve e utiliza algoritmos com capacidade de aprender e reconhecer padrões automaticamente a partir de dados e usa o modelo aprendido para generalizar decisões em novos casos (Provost e Kohavi, 1998). Segundo Mitchell e Mitchell (1997), um programa de computador é dito aprender a partir de uma experiência (E) com respeito a uma classe de tarefas (T) e medida de desempenho (P), se seu desempenho em tarefas de (T) medido por (P), melhora com a experiência (E). A classe de tarefas pode ser classificação ou regressão.

Os algoritmos são conjuntos de regras e procedimentos matemáticos usados para resolver problemas computacionais (Papadimitriou, 1994). O método mais utilizado é o aprendizado de máquina supervisionado (Ludermir, 2021), que utiliza um conjunto de dados para o treinamento, onde cada dado já possui a resposta esperada, envolvendo um modelo matemático que relaciona as entradas aos resultados desejados (Russel e Norvig, 2016).

Dentre o aprendizado supervisionado, a regressão linear (RL) é um dos métodos estatísticos mais básicos, usado para prever um valor numérico com base em variáveis independentes. Quando há uma relação linear entre as variáveis independentes e a variável dependente, é possível alcançar um bom desempenho mesmo com um número reduzido de amostras. No entanto, em casos de relacionamentos não lineares, o desempenho tende a ser insatisfatório (Štepanovský et al., 2017). O modelo busca encontrar a melhor reta que minimiza a soma dos erros quadráticos entre os valores previstos e reais (Montgomery et al., 2012).

As redes neurais artificiais têm solucionado diversos problemas, elas são inspiradas na estrutura de neurônios biológicos (Ludermir, 2021). O Multilayer Perceptron (MLP) é um tipo de rede neural artificial composta por múltiplas camadas de neurônios, sendo capaz de modelar relações complexas entre entrada e saída. Utiliza a retropropagação do erro (*backpropagation*) para ajustar os pesos das conexões entre os neurônios (Haykin, 2009).

O algoritmo Decision Tree (DT) constrói um modelo em uma estrutura de árvore, onde cada nó representa uma condição sobre um atributo e cada ramo representa um possível resultado. Esse método é amplamente utilizado para tarefas de classificação e regressão (Quinlan, 1996). O Random Forest (RF) é um conjunto de múltiplas árvores de decisão, em que cada árvore é treinada com um subconjunto aleatório dos dados

288 (*bagging*). A predição final é obtida pela média para regressão ou votação para
289 classificação (Breiman, 2001).

290 O XGBoost é um método baseado em árvores de decisão que utiliza *boosting*, uma
291 técnica onde modelos fracos são treinados sequencialmente, corrigindo os erros dos
292 anteriores (Chen e Guestrin, 2016). O modelo emprega uma árvore de decisão para tarefas
293 de regressão, segmentando os dados em categorias menores de acordo com diferentes
294 limites das variáveis de entrada. Essa segmentação forma uma estrutura hierárquica
295 semelhante a uma árvore, onde a raiz representa o ponto inicial da divisão, os nós internos
296 correspondem às subdivisões intermediárias e as folhas indicam a saída final do modelo
297 (Bisong et al., 2019). Sua eficiência e precisão o tornam um dos algoritmos mais
298 utilizados em competições de aprendizado de máquina.

299 *SHAP (SHapley Additive Explanations)*

300 Modelos de aprendizado de máquina podem apresentar alta complexidade,
301 tornando desafiadora a interpretação do processo que conduz a uma determinada
302 predição. Esses modelos são frequentemente chamados de "caixa preta", pois oferecem
303 uma compreensão limitada sobre seu funcionamento interno (Lubo et al., 2020).
304 Compreender as razões por trás das predições de um modelo pode ser tão essencial quanto
305 sua precisão, pois isso contribui para gerar confiança e fornecer insights valiosos aos
306 usuários (Lundberg et al., 2017).

307 Diante disso, a técnica SHAP (SHapley Additive Explanations) é um dos métodos
308 mais avançados para explicar o impacto das variáveis de entrada nas predições de
309 modelos de AM. Baseada na teoria dos valores de Shapley, originada da teoria dos jogos
310 cooperativos, essa abordagem distribui a importância de cada característica de forma
311 justa, considerando todas as combinações possíveis das variáveis preditoras. Os valores
312 SHAP atribuem a cada característica a mudança na predição esperada do modelo ao
313 condicionar essa característica. Eles explicam como obter o valor base $E[f(z)]$ que seria
314 previsto se não soubéssemos nenhuma característica para a saída atual $f(x)$ (Lundberg e
315 Lee, 2017).

316 Na analogia com jogos cooperativos, o modelo de AM é tratado como um jogo no
317 qual os "jogadores" são os atributos de entrada, e a função característica define o valor
318 (recompensa) atribuído a cada coalizão de variáveis (Figura 1). Essa função pode
319 representar medidas como ganho de predição, desempenho, ou qualquer métrica de

interesse, como acurácia ou recompensa em aprendizado por reforço (Rozemberczki et al., 2022).

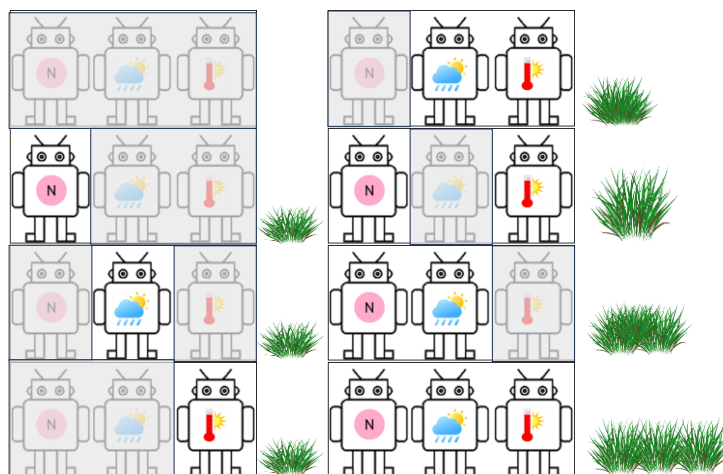


Figura 1. Valor SHAP em jogos cooperativos. Cada robô representa um atributo de entrada do modelo: nitrogênio (N), precipitação e luz solar e temperatura. As combinações (coalizões) de robôs, dispostas em retângulos, formam subconjuntos de variáveis preditoras. A quantidade de pasto ao lado de cada coalizão simboliza a predição do modelo (valor da função característica), ou seja, o valor associado àquela combinação de atributos. Adaptado de Rozemberczki et al. (2022).

Nessa estrutura, o valor SHAP permite quantificar a contribuição individual de cada atributo para uma predição específica. A predição de um dado ponto é decomposta em contribuições atribuídas a cada variável, de modo que a soma dessas partes seja exatamente igual à predição realizada, princípio conhecido como eficiência. Essa decomposição local é fiel à predição do modelo, pois considera todas as combinações possíveis de atributos e calcula a contribuição marginal média de cada um. Diferentemente de métodos como *feature importance* do Random Forest ou coeficientes da Regressão Linear, o SHAP captura interações complexas entre variáveis, tornando-se uma abordagem robusta para modelos não lineares, como XGBoost, Redes Neurais e Random Forest (Lundberg et al., 2018).

Além disso, apresenta vantagens como consistência, equidade e interpretação local e global (Lundberg, 2017). Na consistência, a soma dos valores SHAP corresponde à diferença entre a predição do modelo e o valor médio das predições. Na equidade, todas

as variáveis são avaliadas de maneira justa, considerando seu impacto marginal. Com a interpretação local e global, o SHAP pode explicar previsões individuais (análise local) e fornece uma visão geral do modelo (análise global). A visualização dos resultados de SHAP gera representações gráficas mais intuitivas, facilitando a interpretação dos resultados pelos usuários (Cesaro, 2021).

Assim, os valores de SHAP se consolidam como uma ferramenta poderosa na interpretação de modelos de AM, permitindo avaliar com precisão a importância de cada característica individual na formação de uma determinada previsão. Diante disso, a combinação de modelos AM com o valor SHAP aplicado a dados tabulares do capim-tamani em tarefa de regressão poderia facilitar o entendimento das interações entre os fatores ambientais e de manejo, gerando insights para otimizar o manejo de pastagens.

371

372 **Referências**

373

374 Alvarenga, CA, Euclides, VP, Montagner, DB, Sbrissia, AF, Barbosa, RA, & De Araujo,
 375 AR (2020). Desempenho animal e características do pasto de capim-guiné Mombaça
 376 submetidos a duas frequências de pastejo. *Pastagens Tropicais-Forrajes*
 377 *Tropicales* , 8 (1), 1-10. [10.17138/TGFT\(8\)1-10](https://doi.org/10.17138/TGFT(8)1-10).

378 Balseca, D. G., Rivas, E. C., López, H. B., Guevara, H. P., & Martínez, J. C. (2015).
 379 Nutritional value of Brachiarias and forage legumes in the humid tropics of Ecuador.
 380 *Ciencia e Investigación Agraria: Revista Latinoamericana de Ciencias de la Agricultura*,
 381 42(1), 57–63.

382

383 Barbosa, R. A., Nascimento Júnior, D. D., Euclides, V. P. B., Silva, S. C. D., Zimmer, A.
 384 H., & Torres Júnior, R. A. D. A. (2007). Capim-tanzânia submetido a combinações entre
 385 intensidade e frequência de pastejo. *Pesquisa Agropecuária Brasileira*, 42, 329-340.

386

387 Barrios, S. C. L., Carromeu, C., Crivellaro, L. L., Verzignassi, J. R., Zimmer, A. H.,
 388 Santos, M. F., Jank, L., Valle, C. B., José, M. R., Gomes, O. C. O., Matsubara, E. T., &
 389 Silva, M. A. I. (2021). *Pasto Certo – versão 3.0®: Aplicativo para dispositivos móveis e*
 390 *desktop sobre forrageiras tropicais*. Embrapa Gado de Corte.

391

392 Batistoti, J., Marcato Junior, J., Ítavo, L., Matsubara, E., Gomes, E., Oliveira, B., ... &
 393 Dias, A. (2019). Estimativa da biomassa de pastagens e da altura do dossel no Cerrado
 394 brasileiro utilizando fotogrametria de VANT. *Sensoriamento Remoto* , 11 (20), 2447.
 395 <https://doi.org/10.3390/rs11202447>

396

397 Bisong, E. (2019). *Building machine learning and deep learning models on Google Cloud*
 398 *Platform*. Apress.

399

400 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

401

402 Bretas, I. L., Valente, D. S. M., De Oliveira, T. F., Montagner, D. B., Euclides, V. P. B., &
 403 Chizzotti, F. H. M. (2023). Canopy height and biomass prediction in Mombaça guinea
 404 grass pastures using satellite imagery and machine learning. *Precision Agriculture*, 1-25.
 405 <https://doi.org/10.1007/s11119-023-10017-6>

406 Bretas, I.L., Valente, D.S., Silva, F.F., Chizzotti, M.L., Paulino, M.F., D'Áurea, A.P., ... &
 407 Chizzotti, F.H. (2021). Predição da biomassa aérea e do teor de matéria seca em pastagens
 408 de braquiária pela combinação de dados meteorológicos e imagens de satélite. *Grass and*
 409 *Forage Science* , 76 (3), 340-352. <https://doi.org/10.1111/gfs.12517>

410

411 Caldas, M. B., Diniz, J. P., da Silva, A. J., Baio, S. P. da S., Borges, M. C. R. Z., Nogueira,
 412 K. B., Roque, C. G., & Teodoro, P. E. (2020). Forms of nitrogen fertilizer application in
 413 *Panicum maximum*. *Bioscience Journal*, 36(1), 23–29. [https://doi.org/10.14393/BJ-](https://doi.org/10.14393/BJ-V36N1A2020-39714)
 414 [V36N1A2020-39714](https://doi.org/10.14393/BJ-V36N1A2020-39714)

415

416 Campos, F. P., Nicácio, D. R. O., Sarmiento, P., Cruz, M. C. P., Santos, T. M., Faria, A. F.

- G., Ferreira, M. E., Conceição, M. R. G., & Lima, C. G. (2016). Chemical composition and in vitro ruminal digestibility of hand-plucked samples of Xaraes palisade grass fertilized with incremental levels of nitrogen. *Animal Feed Science and Technology*, 215, 1–12. <https://doi.org/10.1016/j.anifeedsci.2015.12.013>
- Carnevalli, R. A., Da Silva, S. C., Bueno, A. D. O., Uebele, M. C., Bueno, F. O., Hodgson, J., ... & Moraes, J. P. G. (2006). Herbage production and grazing losses in *Panicum maximum* cv. Mombaça under four grazing managements.
- Cesaro, J. (2021). *Avaliação de discriminação em aprendizagem de máquina usando técnicas de interpretabilidade* [Dissertação de mestrado, Escola Politécnica da Universidade de São Paulo].
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Da Silva, S.C.; Nascimento Jr., D. Avanços na pesquisa com plantas forrageiras tropicais em pastagens: características morfofisiológicas e manejo do pastejo. *Revista Brasileira de Zootecnia*, v.36, p.121-138, 2007 (supl. especial).
- Defalque, G., Santos, R., Bungenstab, D., Echeverria, D., Dias, A., & Defalque, C. (2024). Machine learning models for dry matter and biomass estimates on cattle grazing systems. *Computers and Electronics in Agriculture*, 216, 108520. <https://doi.org/10.1016/j.compag.2023.108520>
- Delevatti, L. M., Romanzini, P., Fabiano, J., Koscheck, W., Luis Da Ross De Araujo, T., Renesto, D. M., Ferrari, A. C., Pavezzi Barbero, R., Mulliniks, J. T., & Reis, R. A. (2019). Forage management intensification and supplementation strategy: Intake and metabolic parameters on beef cattle production. *Animal Feed Science and Technology*, 247, 74–82. <https://doi.org/10.1016/j.anifeedsci.2018.11.004>
- Dias Filho, M. B., & Dias Filho, M. B. (2015). Estratégias de recuperação de pastagens degradadas na Amazônia brasileira.
- Dos Reis, A.A., Werner, J.P., Silva, B.C., Figueiredo, G.K., Antunes, J.F., Esquerdo, J.C., ... & Magalhães, P.S. (2020). Monitoramento da biomassa aérea e da altura do dossel de pastagens em um sistema integrado de lavoura e pecuária usando informações texturais de imagens do PlanetScope. *Sensoriamento Remoto*, 12 (16), 2534. 0.3390/rs12162534
- Embrapa. (2015). *BRS Tamani, forrageira híbrida de Panicum maximum*. Embrapa. Disponível em: <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1011507>
- Euclides, V. P. B., Montagner, D. B., de Araújo, A. R., de Aragão Pereira, M., dos Santos Difante, G., de Araújo, I. M. M., Barbosa, L. F., Barbosa, R. A., & Gurgel, A. L. C. (2022). Biological and economic responses to increasing nitrogen rates in Mombaça guinea grass pastures. *Scientific Reports* 2022 12:1, 12(1), 1–14. <https://doi.org/10.1038/s41598-022-05796-6>
- Foster, P., & Kohavi, R. (1998). On applied research in machine learning. *Machine Learning-Boston*, 30, 127–132.

- Galindo, F. S., Beloni, T., Buzetti, S., Carvalho, M., Filho, M. T., Dupas, E., Gaioto, M., & Ludkiewicz, Z. (2018). Technical and economic viability and nutritional quality of mombasa guinea grass silage production. *Acta Scientiarum. Agronomy*, 40(1), e36395. <https://doi.org/10.4025/ACTASCIAGRON.V40I1.36395>
- Gomes, E. S., Bittar, D. Y., & Sà, A. C. O. (2020). Produção de forrageiras da espécie panicum maximum submetidas a doses de nitrogênio. *Ipê Agronomic Journal*, 4(2), 1-8. Doi: <https://doi.org/10.37951/2595-6906.2020v4i2.6242>
- Grosan, C., & Abraham, A. (2011). *Intelligent systems: A modern approach* (17^a ed.). Springer-Verlag. <https://doi.org/10.1007/978-3-642-21004-4>
- Gurgel, A. L. C., Difante, G. S., Emerenciano Neto, J. V., Souza, J. S., Veras, E. L. L., Costa, A. B. G., & Roberto, F. F. S. (2017). Estrutura do pasto e desempenho de ovinos em capim-massai na época seca em resposta ao manejo do período das águas. *Boletim de Indústria Animal*, 74(2), 86-95. 10.17523/bia.v74n2p86
- Haykin, S. (2009). *Neural networks and learning machines* (3^a ed.). Pearson.
- Hodgson, J. (1990). *Grazing management—Science into practice*. Longman Scientific & Technical.
- Hoover, W. H., & Stokes, S. R. (1991). Balanceamento de carboidratos e proteínas para rendimento microbiano ruminal ótimo. *Journal of Dairy Science*, 74(10), 3630–3644.
- Horrocks, R. D., & Vallentine, J. F. (1999). Forage quality—the basics. In *Harvested Forage* (pp. 17–47). Academic Press. ISBN 9780977304561
- Lacerda, M. D. S. B., Alves, A. A., de Oliveira, M. E., Rogério, M. C. P., Carvalho, T. B., & Veras, V. S. (2009). Composição bromatológica e produtividade do capim-andropogon em diferentes idades de rebrota em sistema silvipastoril. *Acta Scientiarum. Animal Sciences*, 31(2), 123-129. <https://doi.org/10.4025/actascianimsci.v31i2.4549>
- Leng, R. A. (1984). Factors affecting the utilization of poor-quality forages by ruminants particularly under tropical conditions. *Nutrition Research and Review*, 3, 277–303.
- Lopes, C. M., Paciullo, D. S. C., Araújo, S. A. C., Gomide, C. D. M., Morenz, M. J. F., & Villela, S. D. J. (2017). Massa de forragem, composição morfológica e valor nutritivo de capim-braquiária submetido a níveis de sombreamento e fertilização. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 69, 225-233. <https://doi.org/10.1590/1678-4162-9201>
- Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K. J., & Pranter, M. J. (2020, October). Machine learning model interpretability using SHAP values: Application to a seismic facies classification task. In *SEG international exposition and annual meeting* (p. D021S008R006). SEG. <https://doi.org/10.1190/segam2020-3428275.1>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://arxiv.org/abs/1705.07874>

- Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.
- MapaBiomias. (2022). *Coleção 7 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil*. Projeto MapBiomias. <https://mapbiomas.org>
- Martelli, R., Ali, A., Rondelli, V., & Barbanti, L. (2025). Are We Up to the Best Practises in Forage and Grassland Precision Harvest? A Review. *Grass and Forage Science*.<https://doi.org/10.1111/gfs.12701>
- Martuscello, J. A., Rios, J. F., Ferreira, M. R., de Assis, J. A., & dos Santos Braz, T. G. (2019). Produção e morfogênese de capim BRS Tamani sob diferentes doses de nitrogênio e intensidades de desfolhação. *Boletim de Indústria Animal*. <https://doi.org/10.17523/bia.2019.v76.e1441>
- Morais, L. F., Cavalcante, A. C. R., Aquino, D. N., & Cândido, M. J. D. (2022). Remote Sensing Applied to Grassland Ecosystems in Regions with Climatic Vulnerability. In *OPEN SCIENCE RESEARCH VII* (Vol. 7, pp. 150-165). Editora Científica Digital.
- Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1, No. 9). New York: McGraw-hill.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5^a ed.). Wiley.
- Murphy, D. J., O'Brien, B., O'Donovan, M., Condon, T., & Murphy, M. D. (2022). A near infrared spectroscopy calibration for the prediction of fresh grass quality on Irish pastures. *Information Processing in Agriculture*, 9(2), 243-253. <https://doi.org/10.1016/j.inpa.2021.04.012>
- Murphy, D. J., Murphy, M. D., O'Brien, B., & O'Donovan, M. (2021). A review of precision technologies for optimising pasture measurement on Irish grassland. *Agriculture*, 11(7), 600.<https://doi.org/10.3390/agriculture11070600>
- NRC - National Research Council. (2001). *Requisitos de nutrientes para gado leiteiro* (7^a ed.). National Academies Press.
- O'Mara, F. P. (2012). The role of grasslands in food security and climate change. *Annals of Botany*, 110(6), 1263–1270. <https://doi.org/10.1093/aob/mcs209>
- Pedreira, B. C., Pedreira, C. G. S., & Silva, S. C. D. (2009). Acúmulo de forragem durante a rebrotação de capim-xaraés submetido a três estratégias de desfolhação. *Revista Brasileira de Zootecnia*, 38, 618-625.
- Projeto MapBiomias, 2024, "Coleção 9 dos Mapas Anuais de Cobertura e Uso do Solo do Brasil (1985-2023)", <https://doi.org/10.58053/MapBiomias/XXUKA8> , Dados MapBiomias, V1.
- Rozemberczki, B., Watson, L., Bayer, P., Yang, H. T., Kiss, O., Nilsson, S., & Sarkar, R. (2022, July). The shapley value in machine learning. In *The 31st International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial*

Intelligence (pp. 5572-5579). International Joint Conferences on Artificial Intelligence Organization.

Rueda-Ayala, V., Peña, J., Höglind, M., Bengochea-Guevara, J., Andújar, D. (2019). Comparing UAV-Based Technologies and RGB-D Reconstruction Methods for Plant Height and Biomass Monitoring on Grass Ley. *Sensors*, 19, 535. <https://doi.org/10.3390/s19030535>

Sbrissia, A. F., da Silva, S. C., & Nascimento Júnior, D. (2007). Ecofisiologia de plantas forrageiras e o manejo do pastejo. *Simpósio sobre Manejo da pastagem*, 24, 1-27.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3). <https://doi.org/10.1214/10-STS330>

Squires, V. R., Dengler, J., Feng, H., & Hua, L. (Eds.). (2018). *Grasslands of the world: Diversity, management*.

Stuart, J. R., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited.

t'Mannetje, L., & Jones, R. (2000). *Field and laboratory methods for grassland and animal production research*. CABI Publishing. ISBN 0-85199-351-6

Van Soest, P. J. (1994). *Nutritional ecology of the ruminant* (2^a ed.). Cornell University Press, New York.

Zanine, A., Farias, L., Ferreira, D., Farias, L., Ribeiro, M., Souza, A., ... & Nascimento, T. (2020). Effect of season and nitrogen fertilization on the agronomic traits and efficiency of Piatã grass in Brazilian savanna. *Agriculture*, 10(8), 337. <https://doi.org/10.3390/agriculture10080337>

Zhao, Y., Liu, Z., & Wu, J. (2020). Grassland ecosystem services: A systematic review of research advances and future directions. *Landscape Ecology*, 35(4), 793–814. <https://doi.org/10.1007/s10980-020-00980-3>

Capítulo 2. Modelos de aprendizado de máquina para predição de proteína bruta de pastos de capim-tamani

Resumo: Conhecer a qualidade da forragem é essencial para atender às exigências dos animais e otimizar a produção. Os métodos tradicionais para aferir o valor nutritivo do pasto são destrutivos, onerosos e demorados, destacando a necessidade de técnicas mais rápidas e acessíveis para a tomada de decisão. O estudo teve como objetivos: i) testar a aplicabilidade de modelos de aprendizado de máquina com dados tabulares como variáveis climáticas, interceptação de luz (IL), doses de nitrogênio (N), intervalos entre pastejo (IEP) e alturas de pré (ALTpre) e pós-pastejo (ALTpos) para predição do teor de proteína bruta das folhas (PB) de pastos de capim-tamani; ii) identificar as variáveis que mais contribuem para predição de PB. Para isso foi utilizado um conjunto de 90 instâncias com 80% para treino e validação e 20% para teste com validação cruzada com 5 folds e 100 iterações. Os hiperparâmetros foram ajustados com *grid-search* no conjunto de treino. Testou-se Regressão Linear (RL), Multilayer Perceptron (MLP), Decision Trees (DT), Random Forest (RF) e XGBoost. Os modelos MLP ($r=0.75$, $R^2=43.94\%$, $MAE=1.56$), RF ($r=0.78$, $R^2=49.23\%$, $MAE=1.60$) e XGBoost ($r=0.78$, $R^2=56.90\%$, $MAE=1.45$) apresentaram os melhores resultados de predição ($p<0.001$). As variáveis com maior importância preditora do teor de PB são IEP, seguido por doses de N, ALTpre e ALTpos. XGBoost se destacou entre os modelos testados ($p<0.001$). O uso de dados tabulares como dose de N, IEP, ALTpre, ALTpos, IL e variáveis climáticas é uma alternativa viável a predição de PB. Existem evidências fortes de que o manejo influencia mais a composição química das folhas do capim-tamani do que as condições ambientais.

Palavras-chave: machine learning, precision livestock farming, manejo de pastagens, *Panicum maximum*.

Introdução

As pastagens ocupam aproximadamente um terço da superfície terrestre e constituem grande parte das terras agrícolas do planeta, além de exercer serviços ecossistêmicos como base de alimentação para pecuária, substrato de biomassa para produção de energia e sequestro de carbono (Renemann et al., 2020; Wijeshingha, 2020). A utilização de gramíneas tropicais, aliada ao manejo adequado da pastagem e do pastejo, é fundamental para explorar o recurso forrageiro de maneira eficiente. Para atingir bons resultados, o manejo do pastejo deve ser feito visando produtividade animal e perenidade das pastagens, baseando-se na relação entre valor nutritivo, exigência animal e massa de forragem (Bretas et al., 2022).

A massa de forragem é a quantidade de forragem em uma área, expressa em quilograma de matéria seca por hectare. Já o valor nutritivo da forragem, é determinado pela sua digestibilidade e composição química, que inclui os níveis de proteína bruta, carboidratos estruturais e não estruturais (Balseca et al., 2015). A composição química das plantas forrageiras pode ser influenciada pela espécie ou cultivar, condições edafoclimáticas e manejo adotado (Lopes et al., 2017).

A partir da massa de forragem o animal seleciona sua dieta e o ganho de peso diário sofre influência da quantidade e qualidade da forragem consumida (Rouquette, 2016). Pastos com baixo valor nutritivo estão frequentemente associados a altos níveis de fibra e teores reduzidos de proteínas, características comuns em plantas forrageiras em estágio de maturidade avançado (Zubieta et al., 2021) ou que não foram manejadas adequadamente. Isto resulta em alterações no consumo, comprometendo a capacidade de atender às exigências nutricionais dos animais. Portanto, conhecer a qualidade da forragem é essencial para atender às exigências dos animais e aumentar a eficiência do sistema de produção.

Os métodos para aferir valor nutricional do pasto são feitos de forma destrutiva, por meio de análises químicas em laboratório e espectroscopia de refletância no infravermelho próximo (NIRS), requerem tempo, mão de obra e investimentos (Horrocks et al., 2020; t'Mannetje et al., 2000). Diante disso, desenvolver métodos e técnicas capazes de fornecer informações precisas, com menor exigência de mão de obra, de baixo custo e curto prazo é fundamental para tomadas de decisão em sistemas de produção em pasto.

Modelos estatísticos são amplamente utilizados para prever a produtividade agrícola (Gornott e Wechsung, 2016). No entanto, esses modelos geralmente assumem que os dados são independentes e identicamente distribuídos (iid) e exigem um conhecimento prévio sobre sua distribuição para uma modelagem adequada. Em contraste, modelos de aprendizado de máquina vêm sendo cada vez mais empregados na estimativa do rendimento de culturas, pois não dependem dessas suposições rígidas e conseguem identificar padrões diretamente a partir dos dados disponíveis (Jeong et al., 2016). Essa flexibilidade reduz a necessidade de conhecimento prévio sobre a estrutura dos dados, tornando esses modelos mais adaptáveis a diferentes contextos. Ademais, estudos indicam que o aprendizado de máquina frequentemente apresenta maior precisão nas estimativas quando comparado a métodos estatísticos tradicionais (Le Maire et al., 2011; Freitas et al., 2020; Vahedi et al., 2019).

Diante dessas vantagens, modelos de aprendizado de máquina vêm sendo amplamente aplicados na agricultura, incluindo culturas como eucalipto, cana-de-açúcar, trigo e milho (Vani et al., 2015; Jeong et al., 2016; da Silva et al., 2021; Ramos et al., 2020). Entretanto, sua aplicação ainda é limitada no manejo de pastagens, especialmente para cultivares da espécie *Panicum maximum* (Sin. *Megathyrsus maximus*), amplamente utilizadas no Brasil em sistemas intensificados devido à alta produtividade e valor nutritivo (Mwendia et al., 2022).

Estudos recentes sobre predição de massa de forragem e teor de proteína bruta utilizam aprendizado de máquina com dados provenientes de imagens de satélites, drones e análises laboratoriais, o que pode tornar essas abordagens caras e de difícil aplicação em larga escala (Reis et al., 2020; Rosa et al., 2021; Lugasi et al., 2021; Fernandes et al., 2024). Até o presente momento, não foram encontrados estudos que explorem a predição da proteína bruta da folha sem a utilização de imagens. Além disso, não há registros de trabalhos que utilizem variáveis como dados climáticos, doses de nitrogênio, intervalo entre pastejo e alturas de pré e pós-pastejo para estimar o teor de proteína bruta (PB) em pastos de capim-tamani. O presente trabalho segue uma abordagem distinta, propondo uma solução simples e acessível, mas também eficaz, para essa predição do teor de proteína bruta no capim-tamani. Utiliza-se um conjunto de dados de fácil obtenção, dispensando o uso de sensores remotos e análises laboratoriais, além de algoritmos que não requerem alto poder computacional para treinamento e avaliação, tornando a proposta viável para aplicação em sistemas produtivos.

As contribuições deste trabalho são:

- A proposta de predição de PB utilizando informações de baixo custo e fácil acesso potencializa a sua aplicabilidade e consequentemente o impacto desta pesquisa. A PB, importante elemento da composição química do pasto, é resultado de dados climáticos, adubação nitrogenada, intervalo entre pastejo e altura de pré e pós-pastejo. Nenhuma destas características requerem voo de drone ou análises laboratoriais. As informações utilizadas podem ser obtidas pelo registro e coleta das informações que normalmente já fazem parte da maioria dos testes dos campos experimentais.
- A análise de correlação que indica que dose de nitrogênio (N), estação de águas, altura de pré-pastejo (ALT_{pre}), altura de pós-pastejo (ALT_{pos}), intervalo entre pastejo (IEP) e precipitação (Precip) possuem correlações significativas (teste Tukey com $p < 0.05$), podendo estes serem utilizados como características relevantes para predição de PB.
- Avaliação experimental robusta utilizando algoritmos tradicionais de aprendizado de máquina, demonstrando que é possível obter resultados promissores com a abordagem proposta. O melhor resultado foi obtido com o algoritmo XGBoost apresentando uma correlação de Pearson (r) de 0.78 e coeficiente de determinação (R^2) de 56.65%.

Material e Métodos

Aquisição de dados

Os dados coletados foram de experimento realizado na Embrapa Gado de Corte, em Campo Grande, MS (20°27'S e 54°37'W, a 530 m de altitude), de outubro de 2020 a abril de 2022. O padrão climático da região é descrito segundo a classificação de Köppen, como tropical chuvoso de savana, subtipo Aw, caracterizado pela distribuição sazonal de chuvas. Os dados de temperatura, precipitação e radiação solar do período experimental foram coletados pela estação meteorológica da Embrapa Gado de Corte, distante aproximadamente 2,4 km da área experimental. O solo da área experimental é classificado como Nitossolo Vermelho Distrófico Latossólico (Embrapa, 2018), caracterizado por textura argilosa.

Os dados foram coletados de uma área de 0,96 ha, formada com capim-tamani dividida em quatro blocos, cada bloco foi subdividido em quatro piquetes (0,06 ha). Os

pastos foram manejados com os tratamentos constituídos por duas frequências de pastejo representadas pela interceptação de luz (IL) 90% de IL e 95% de IL e duas doses de nitrogênio (N) 80 e 240 kg/ha/ano de N. A intensidade de pastejo foi mantida fixa, 50% da altura de entrada. Os dados foram obtidos nas seguintes estações chuvosas e secas entre os anos de 2020 e 2022 visando gerar uma grande variabilidade e, a partir disso, construir modelos robustos que possam ser utilizados em qualquer época do ano. As variáveis selecionadas como entrada nos modelos testados tiveram como base sua facilidade de obtenção e que fossem adquiridas rotineiramente ao manejar o pasto.

O conjunto de dados construído foi composto com um total de 14 variáveis, sendo uma variável dependente e 13 variáveis independentes, e um total de 90 instâncias. No resumo do conjunto de dados (Tabela 1) foi apresentado o seu significado, a unidade utilizada, a média juntamente com o desvio padrão (somente para variáveis numéricas). A variável categórica nominal estação foi transformada em um conjunto de variáveis binárias através da técnica *One Hot Encoding*. Calculou-se a média das variáveis binárias com o intuito de caracterizar sua distribuição (Tabela 1).

Tabela 1. Descrição das variáveis que compõe o conjunto de dados coletado. No total são 14 variáveis, sendo 5 variáveis categóricas e 9 variáveis numéricas. A variável que se deseja prever é a proteína bruta (PB).

Variável (símbolo)	Significado	Unidade	Média experimental
Águas 20/21	estação do ano	variável bin. (1/0)	0.28 (± 0.45)
Águas Seca 21	estação do ano	variável bin. (1/0)	0.13 (± 0.33)
Seca Águas 21	estação do ano	variável bin. (1/0)	0.15 (± 0.36)
Águas 21/22	estação do ano	variável bin. (1/0)	0.38 (± 0.49)
Águas Seca 22	estação do ano	variável bin. (1/0)	0.06 (± 0.23)
TEMP	temperatura média	graus °C	27.8 (± 1.6)
RAD	radiação solar	KJ/m ²	1589 (± 378.5)
PREC	precipitação	mm	229.8 (± 195.3)
DOSE DE N	dose de nitrogênio	kg/ha/ano	181.3 (± 77.5)
IL	interceptação de luz	%	91.7 (± 2.4)
IEP	intervalo entre pastejo	dias	31.9 (± 13.5)
ALTpré	altura de pré-pastejo	cm	34.6 (± 6.4)
ALTpós	altura de pós-pastejo	cm	17.2 (± 3.3)
PB	proteína bruta	% da MS	12.52 (± 3.0)

Bin.: binária, MS: matéria seca

Machine Learning

A Figura 1 apresenta o fluxograma de coleta, processamento de dados, treinamento e validação de modelos de aprendizado de máquina para predição de proteína bruta da folha de pastos de capim-tamani.

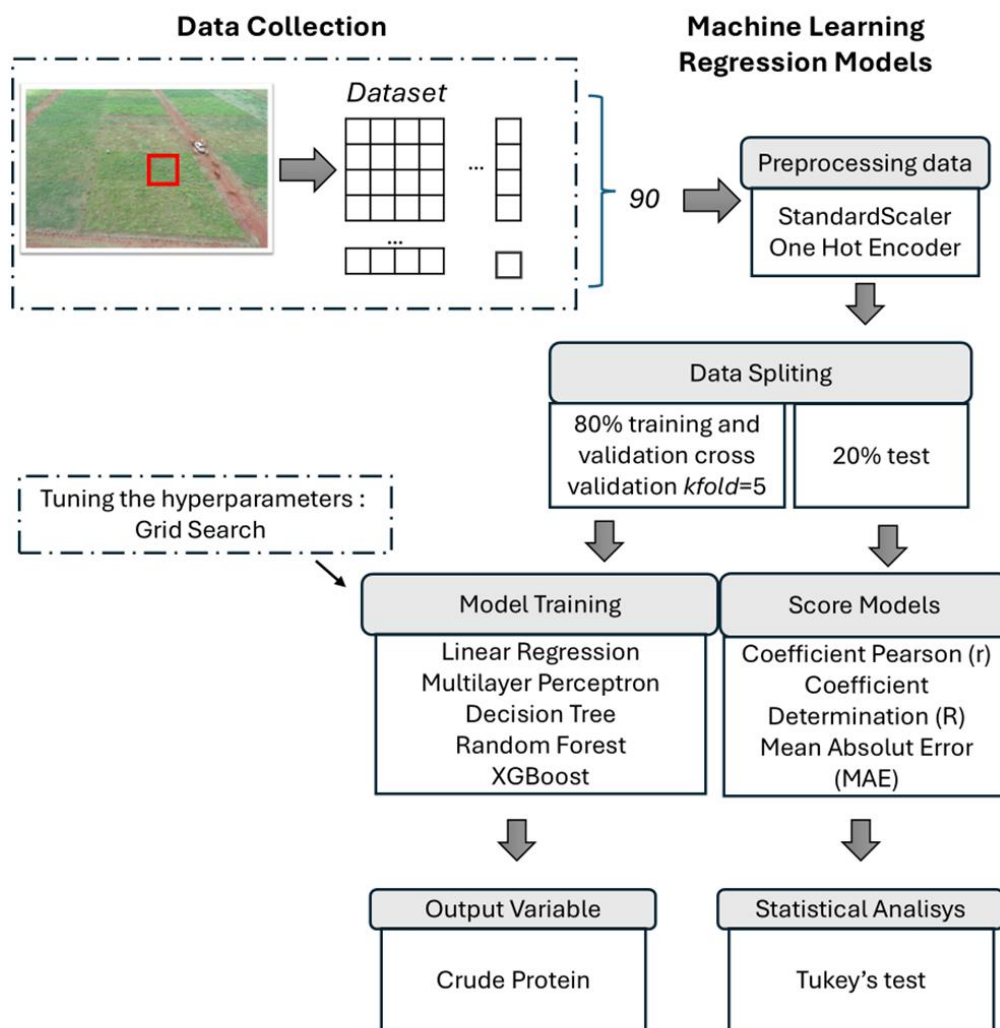


Figura 1. Fluxograma de coleta, processamento de dados, treinamento, validação e teste de modelos aprendizado de máquina para predição de proteína bruta da folha.

Os modelos de aprendizado de máquina tiveram como objetivo prever a proteína bruta (PB) das folhas do capim-tamani. As variáveis selecionadas como entrada foram: estação do ano, dados climáticos (temperatura média, radiação solar e precipitação), dose de nitrogênio (N) (80 e 240 kg/ha/ano de N), interceptação de luz (IL), intervalo entre pastejo (IEP) e altura de pré (ALTpré) e pós-pastejo (ALTpós) foram incluídas como

773 dados de treinamento dos modelos. O teor de proteína bruta da folha (PB) é o atributo
774 alvo (Tabela 1).

775 A implementação e treinamento dos modelos foram realizados utilizando Pandas
776 e Scikit-learn (Scikit-learn, 2011). De um conjunto de 90 amostras, foram separadas 72
777 amostras (80%) para treino e validação 18 (20%) para teste. A estimativa de avaliação
778 dos modelos foi feita com validação cruzada (cross-validation) com $k=5$ partições, 100
779 iterações e os ajustes dos hiperparâmetros foram realizados com *grid-search* no conjunto
780 de treino.

781 Foram testados cinco modelos: (i) Regressão Linear (RL); (ii) Multilayer
782 Perceptron (MLP); (iii) Decision Trees (DT); (iv) Random Forest (RF) e (v) XGBoost. A
783 Tabela 2 apresenta informações sobre os modelos avaliados, com destaque para os
784 hiperparâmetros utilizados e os intervalos de variação dos hiperparâmetros.

785 A regressão linear (RL) é uma técnica tradicional para previsão usada como
786 controle contra os modelos de aprendizado de máquina. Se existe relação linear entre as
787 variáveis independentes e dependente, é possível obter um bom desempenho mesmo com
788 poucas amostras. Entretanto, quando o relacionamento não é linear, o desempenho não é
789 satisfatório. Utiliza o critério de Akaike para seleção de modelos e é capaz de lidar com
790 instâncias ponderadas (Tabela 2) (Štepanovský et al. 2017).

791 Para redes neurais foi utilizado o modelo Multilayer Perceptron (MLP), um
792 regressor que usa retropropagação para aprender um perceptron multicamadas para
793 prever instâncias. Para MLP foi utilizada a configuração (selecionada de maneira ad
794 hoc) com uma única camada oculta com 100 neurônios, com a função de ativação *tanh*,
795 taxa de aprendizado $\alpha=0.0001$ adaptável, treinado usando *Stochastic Gradient Descent*
796 (*SGD*) com o limite de 5000 iterações. (Tabela 2).

797 Decision Trees (DT) é um aprendizado rápido da árvore de decisão. Constrói uma
798 árvore de decisão/regressão usando ganho/variância de informação poda de erro reduzido
799 (com *backfitting*). Classifica valores para atributos numéricos apenas uma vez (Raj et al.,
800 2020). São modelos simples e interpretáveis, seu desempenho pode ser limitado devido à
801 alta variância e tendência ao sobreajuste. Foi utilizado profundidade máxima de 40, com
802 3 variáveis por divisão, 8 amostras por folha e divisão de um nó quando houve pelo menos
803 4 amostras nele (Tabela 2).

Tabela 2. Machine learning models and experimental setups used for prediction.

Experimento	Modelo	Hiperparâmetros	Intervalo	Referência
#1	RL	Copy X: true, fit intercept: true, n jobs: 1, positive: false	Copy X: [True, False], fit_intercept: [True, False], n_jobs: [1], positive: [True, False]	Štepanovský et al. (2017)
#2	MLP	Activation: tanh, Alpha: 0.0001, hidden layer sizes: (100,), learning rate: adaptative, max iter: 5000, solver: sgd	Alpha: [0.001, 0.01], hidden layer sizes: [(100,), (100, 50,)], (100, 100,)], learning rate: [constant, adaptive], max iter: [10000], solver: [adam]	Raj et al. (2020)
#3	RF	Criterion: Absolute error, Max depth: 1000, max features: 2, min samples leaf: 2, min samples split: 2, n estimators: 200	Max depth: np.arange(2, 100, 40), max features: np.arange(2, 10, 2), min samples leaf: np.arange(2, 10, 2), min samples split: np.arange(2, 8, 2), n estimators: [200, 600, 1000]	Belgiu et al. (2016)
#4	DT	Criterion: Absolute error, Max depth: 40, max features: 3, min samples leaf: 8, min samples split: 4	Max depth: [10, 40, 80, None], max features: np.arange(1, 14, 2), min samples leaf: [2, 4, 8, 16], min samples split: [2, 4, 8]	Saha et al. (2020)
#5	XGBoost	Colsample bytree: 0.6, learning rate: 0.1, max depth: 2, n estimators: 200, reg alpha: 0.6, reg lambda: 2, subsample: 0.6	Colsample bytree: [0.6, 0.8, 1.0], learning rate: [0.001, 0.01, 0.1], max depth: np.arange(2, 8, 2), n estimators: [200, 600, 1000], reg alpha: [0, 0.2, 0.6], reg lambda: [1, 1.5, 2]	Bisong et al. (2019)

RL: regressão linear, MLP: multilayer perceptron, RF: random forest, DT: decision tree.

Random Forest (Breiman, 2001) é um método ensemble baseado em *bagging* (*Bootstrap Aggregating*), no qual múltiplas árvores de decisão são treinadas com diferentes amostras dos dados e subconjuntos de variáveis. Essa abordagem reduz a variância e melhora a generalização do modelo. A predição final é obtida por meio da média das respostas das árvores no caso de regressão ou pelo voto majoritário na classificação. RF foi configurado com 200 árvores, profundidade máxima de 1000. Cada nó utilizou no máximo duas variáveis, dividido com, pelo menos, duas amostras com folhas contendo no mínimo duas amostras. Foi calculada a importância do atributo por meio da diminuição média da impureza que mede o impacto de cada variável na divisão dos nós ao longo das árvores da floresta. Esse cálculo é baseado no conceito de impureza dos nós, quantificada pelo erro quadrático médio em casos de regressão.

Apesar de sua robustez, métodos mais avançados, como o XGBoost, aprimoram ainda mais o desempenho ao incorporar técnicas de *boosting* para otimizar a aprendizagem do modelo. O modelo utiliza uma árvore de decisão para tarefas de regressão com divisão dos dados em categorias menores conforme os diferentes limites de recursos de entrada. Essas divisões criam uma estrutura semelhante a uma árvore, em que a raiz representa um limite inicial para a divisão de dados, os nós internos indicam divisões internas e as folhas representam a saída final do modelo (Bisong et al., 2019). O ajuste de hiperparâmetros de XGBoost foi configurado com árvores rasas de profundidade máxima de 2, com 200 iterações e taxa de aprendizado de 0.1 com 60% dos dados e 60% das features por árvore (Tabela 2). Foi calculada a importância dos atributos usando XGBoost com base no *total gain* que mede a importância de uma variável com base no ganho total de informação que ela proporciona ao longo de todas as árvores do modelo.

Análise estatística

Foi realizada análise de correlação de Pearson entre as variáveis preditoras e a variável a ser predita (PB). O desempenho dos modelos de aprendizado de máquina (AM) foi avaliado por meio das métricas de coeficiente de Pearson (r), coeficiente de determinação (R^2) e erro médio absoluto (MAE) calculados entre os valores preditos e observados. As médias das 100 repetições utilizadas (folds) foram realizadas com *Random State=None* e *Shuffle=True* para cada métrica de desempenho e submetidas ao teste de Tukey a 5% de probabilidade.

803 Os experimentos foram executados em uma CPU Intel® Core™ i5 com 8 Gb de
804 RAM e todos hiperparâmetros foram definidos na biblioteca scikit-learn.

805 **Resultados**

806 Os experimentos foram estruturados em quatro avaliações buscando responder às
807 seguintes perguntas de pesquisa (Research Questions - RQ):

808 -RQ 1: Como os atributos se relacionam linearmente?

809 -RQ 2: Entre os algoritmos clássicos de AM qual é aquele mais adequado para predição
810 de proteína bruta?

811 -RQ 3: Quais são as variáveis mais relevantes para os melhores modelos?

812 As seções seguintes representam o respectivo experimento/avaliação executado
813 com o objetivo de responder cada uma das perguntas acima.

814 *Respondendo RQ1: Como os atributos se relacionam linearmente?*

815 Para mensurar a correlação linear, utilizamos o coeficiente de correlação de Pearson (r).
816 A correlação mensura a direção e o grau da relação linear entre duas variáveis (Moore,
817 2007). O coeficiente de correlação de Pearson possui intervalo de -1 a 1 e o sinal indica
818 a direção da relação, positiva ou negativa, enquanto o valor representa a intensidade da
819 relação entre as variáveis. Sua fórmula é a seguinte:

820

$$821 \quad r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

822

823 Onde X_i são os valores da variável X em uma amostra, $\bar{X}$ é a média dos valores da
824 variável X, Y_i são os valores da variável Y em uma amostra e $\bar{Y}$ é a média dos valores da
825 variável Y.

826 A Figura 2 exibe os coeficientes de Pearson entre as variáveis do conjunto de dados.
827 A PB apresentou correlação moderada e negativa com ALTpre ($r=-0.49$), ALTpos ($r=-$
828 0.46) e Águas 20/21 ($r=-0.43$), fraca negativa com IEP ($r=-0.23$) e PREC ($r=-0.33$) e
829 moderada positiva com dose de N ($r=0.48$) e Águas 21/22 ($r=0.40$) (Figura 2).

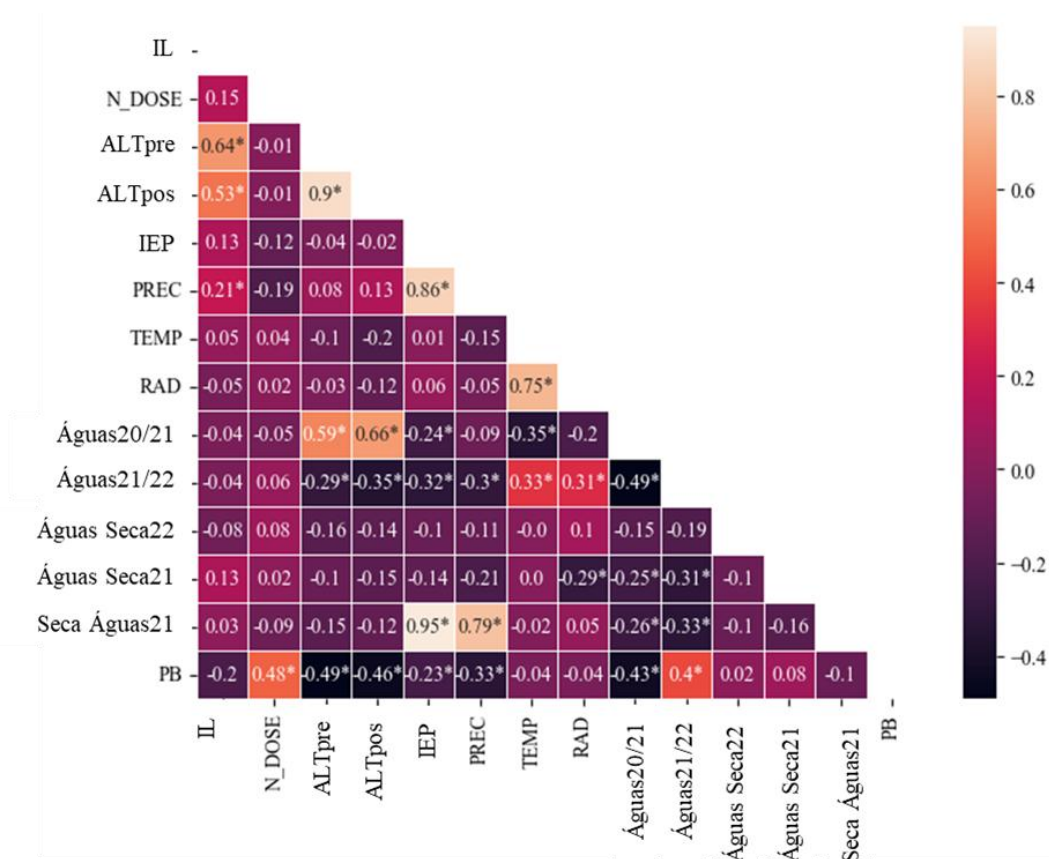


Figura 2. Coeficientes de correlação da interceptação de luz (IL), dose de N, estações, altura de pré-pastejo (ALTpre), altura de pós-pastejo (ALTpos), intervalo entre pastejo (IEP), precipitação (PREC), temperatura (TEMP), radiação solar (RAD) e proteína bruta da folha (CP). Correlações com asterisco são significativas pelo teste Tukey ($p < 0,05$).

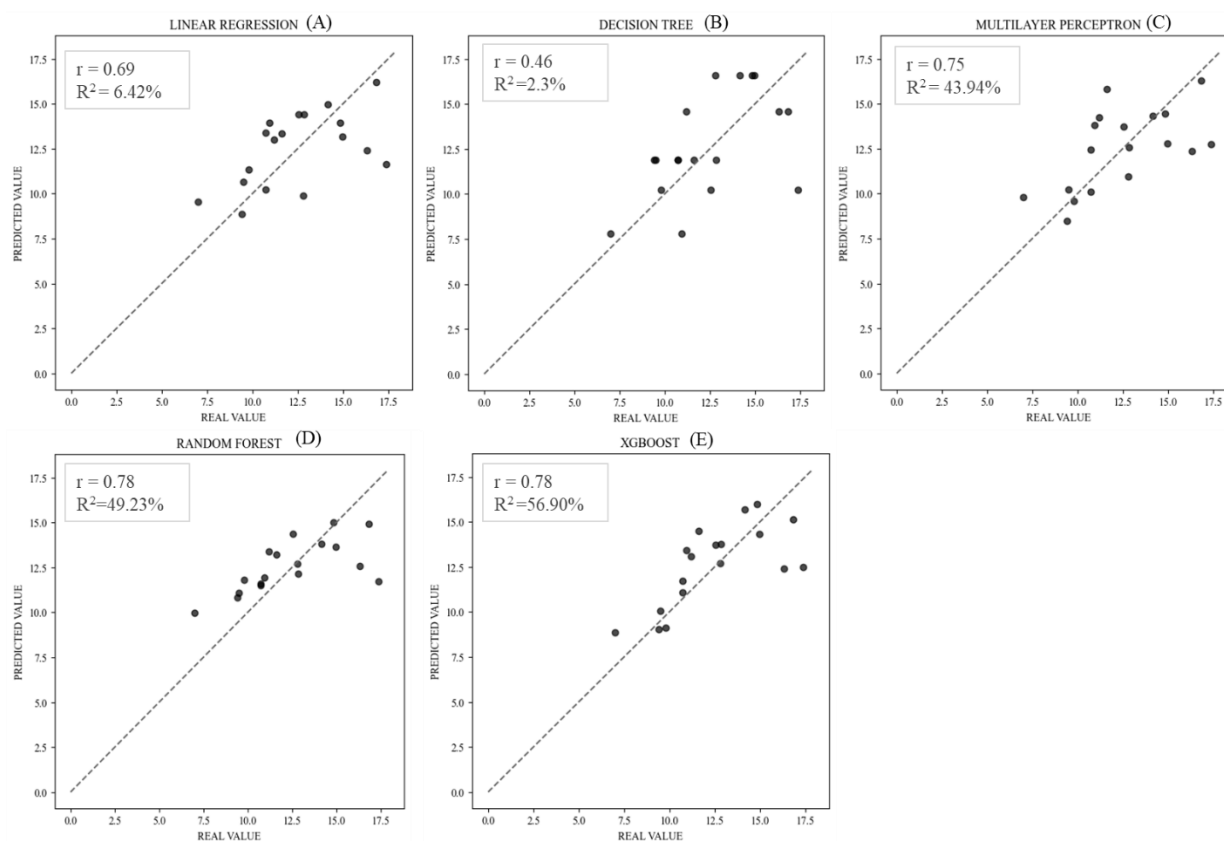
Observou-se correlação forte e positiva de Seca Águas 21 com IEP ($r=0.95$) e PREC ($r=0.79$) e negativa e fraca com Águas 20/21 ($r=-0.26$) e Águas 21/22 ($r=-0.33$). A estação Águas Seca 21 apresentou correlação fraca e negativa com RAD ($r=-0.29$), Águas 20/21 ($r=-0.25$) e Águas 21/22 ($r=-0.31$). Águas 21/22 apresentou correlação fraca e negativa com ALTpre ($r=-0.29$), ALTpos ($r=-0.35$), IEP ($r=-0.32$), PREC ($r=-0.30$), moderada negativa com Águas 20/21 ($r=-0.49$), fraca positiva com TEMP ($r=0.33$) e RAD ($r=0.31$). Foi observada correlação moderada e positiva de Águas 20/21 com ALTpre ($r=0.59$) e ALTpos ($r=0.66$) e fraca negativa com IEP ($r=-0.24$) e TEMP ($r=-0.35$) (Figura 2).

A RAD apresentou correlação forte e positiva com TEMP ($r=0.75$). Observou-se correlação forte e positiva entre PREC e IEP ($r=0.86$). ALTpos apresentou correlação

moderada e positiva com IL ($r=0.53$) e forte e positiva com ALTpre ($r=0.9$). HRE apresentou correlação moderada e positiva com IL ($r=0.64$) (Figura 2).

Respondendo RQ2 - Entre os algoritmos clássicos de AM qual é aquele mais adequado para predição de proteína bruta?

Na Figura 3, são apresentados os gráficos de dispersão entre os valores observados e valores preditos em cada modelo testado. O modelo RL apresentou alguns dados com bom ajuste de distribuição. No entanto, foram observados valores superestimados e subestimados com grande amplitude de erro e dados outliers (Figura 3A). Em DT, os dados se apresentaram muito dispersos do valor real (Figura 3B), corroborando com os dados de avaliação dos modelos (Figura 4). Ao testar o modelo MLP, observou-se melhoria na distribuição dos dados em relação a LR e DT (Figura 3C). A comparação entre o observado e predito demonstra que as previsões do modelo RF e XGBoost correspondem fortemente às observações (Figura 3D e 3E).



860

Figura 3. Gráficos de dispersão entre valores observados e preditos de cada modelo testado: Linear Regression (LR), Decision Tree (DT), Multilayer Perceptron (MLP), Random Forest (RF), e XGBoost para predição de proteína bruta da folha de capim-tamani.

864

A Tabela 3 exibe os coeficientes de correlação (r), erro médio absoluto (MAE) e o coeficiente de determinação (R^2) dos modelos de aprendizado máquina, testados por meio de Linear Regression (LR), Decision Tree (DT), Multilayer Perceptron (MLP), Random Forest (RF), e XGBoost com $k\text{-fold}=5$ e 100 repetições com as seguintes variáveis preditoras: IL, dose de N, dados climáticos, altura pré-pastejo, altura pós-pastejo e IEP para predição de PB de capim-tamani.

Tabela 3. Coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e coeficiente de determinação (R^2) de cada modelo de Machine Learning (ML): Linear Regression (LR), Decision Tree (DT), Multilayer Perceptron (MLP), Random Forest (RF), e XGBoost para predição de proteína bruta da folha de capim-tamani.

Models	r	MAE	R^2
LR	0.69b	1.67b	6.42b
DT	0.46c	2.23a	2.30b
MLP	0.75a	1.56bc	43.94a
RF	0.78a	1.60b	49.23a
XGBoost	0.78a	1.45c	56.90a
p-value	<0.001	<0.001	<0.001
SEM	0.014	0.034	5.53

SEM: standard error of the mean. Médias seguidas por letras minúsculas diferentes na coluna diferem entre si, pelo teste de Tukey a 5% de probabilidade.

O modelo DT apresentou menor valor de r de Pearson, ao testar RL, o coeficiente melhorou em relação a DT ($p<0.05$). Os modelos MLP ($r=0.75$), RF ($r=0.78$) e XGBoost ($r=0.78$) apresentaram os maiores valores de r de Pearson ($p<0.05$), não diferindo entre si ($p>0.05$). Para a métrica MAE, observou-se menor valor para XGBoost (MAE=1.45) ($p<0.05$). Os modelos RF, MLP e RL apresentaram valores intermediários de MAE e não diferiram entre si ($p<0.05$). Observou-se maior MAE para o modelo DT (MAE=2.22) ($p<0.05$). Para a métrica R^2 , foi observado destaque para XGBoost com o maior valor ($R^2=56.90\%$), seguido de RF ($R^2=49.23\%$) e MLP ($R^2=43.94\%$) ($p<0,05$). Assim, respondendo a RQ2 podemos afirmar que o melhor modelo de aprendizado de máquina avaliado para prever PB de pastos de capim-tamani foi XGBoost com um desempenho de $r=0.78$, MAE=1.45 e $R^2=56.90$ (Tabela 3).

Respondendo RQ 3 – Quais são as variáveis mais relevantes para os melhores modelos?

A contribuição dos atributos individuais para estimar a PB das folhas do capim-tamani (Figura 4) foi estimada com RF considerando todos os dados de entrada para

confrontar a importância dos atributos com base na diminuição média de impurezas (e no número de nós que usam esse atributo) em RF.

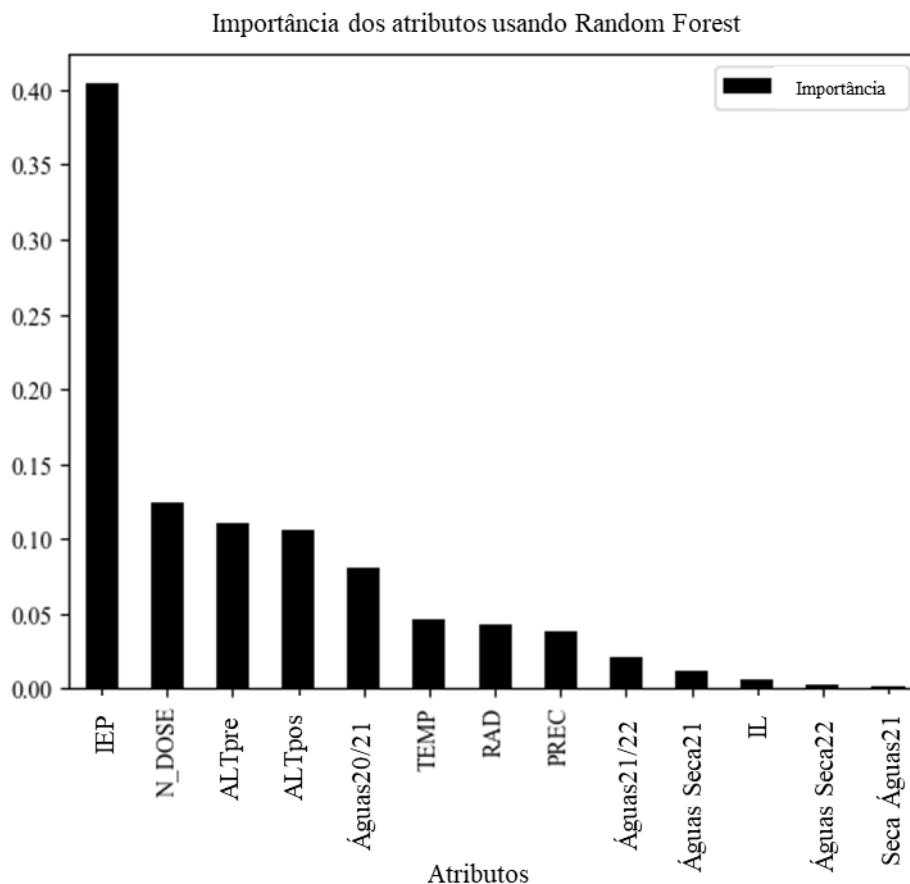
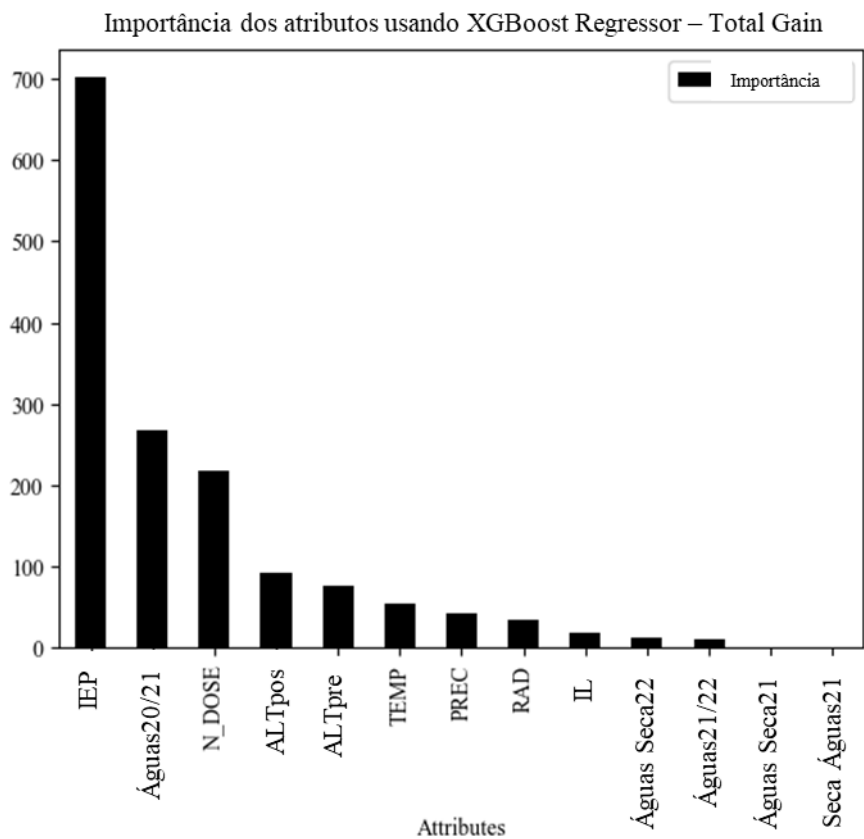


Figura 4. Importância dos atributos com base na diminuição média de impurezas (e no número de nós que usam esse atributo) em Random Forest. IEP: intervalo entre pastejo, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, PREC: precipitação, IL: interceptação de luz.

Respondendo a RQ3, as variáveis mais relevantes para o modelo RF são IEP seguido de dose de nitrogênio, altura pré-pastejo, altura pós-pastejo, estação de águas 20/21, precipitação, temperatura, radiação, com maior importância na predição e por último as estações de águas 21/22, transição águas-seca 21, transição águas-seca 22, IL e transição seca-águas 21 com menor importância (Figura 4).

A contribuição dos atributos individuais para estimar a PB de capim-tamani (Figura 5) foi estimada com XGBoost considerando todos os dados de entrada para confrontar a importância dos atributos em XGBoost com base em total gain.

910



911

912 **Figura 5.** Importância dos atributos usando XGBoost com base no total gain. IEP:
913 intervalo entre pastejo, ALTpos: altura pós-pastejo, ALTpre: altura pré-pastejo, TEMP:
914 temperatura, RAD: radiação solar, IL: interceptação de luz.

915

916

917 A ordem de importância das variáveis muda entre os modelos testados (Figuras 4
918 e 5). Respondendo a RQ3, as variáveis mais importantes no modelo XGBoost para
919 prever a PB são IEP seguido de águas 20/21, dose de nitrogênio, altura pós-pastejo,
920 altura pré-pastejo e temperatura (Figura 5). Precipitação, radiação, IL, e as estações
921 águas-seca 22 e águas 21/22 são os atributos com menor importância em contribuir na
922 predição (Figura 5).

923

924 **Discussão**

925 As correlações significativas entre PB e as doses de N, estação, ALTpre, ALTpos, IEP
926 e Precip, demonstram a importância desses fatores em influenciar a composição química

da forragem e consequentemente sua qualidade. Sabe-se que a planta forrageira responde aos fatores abióticos luz, água, temperatura e nutrientes (Nabinger, 1997). Ainda que não tenha sido observado correlações entre PB e temperatura e radiação ($p>0,05$), houve correlação com as estações de águas que reúne os elementos climáticos como temperatura, radiação, umidade e pressão.

Da mesma forma, PB não se correlacionou com IL, mas foi observado correlação com ALTpre e ALTpos. Já é bem elucidado na literatura que a altura do dossel forrageiro possui relação com IL (Carnevali et al., 2006; Pedreira et al., 2009), sendo utilizada como forma de controle da estrutura do dossel forrageiro. Além disso, observou-se correlação forte entre IL e ALTpre e entre ALTpre e ALTpos. Em sistemas de produção que utilizam a lotação intermitente como método de pastejo, a frequência de pastejo irá determinar o período de descanso da planta forrageira o que pode influenciar a produtividade, a composição química e a capacidade de rebrota (Costa et al., 2019).

A correlação negativa observada entre PB e IEP, demonstra sua influência na composição química do pasto ao passo que quanto maior o IEP mais velha será a planta forrageira o que indica maior lignificação dos tecidos e aumento de estruturas de sustentação e componentes fibrosos (Geremia et al., 2018). A concentração de compostos de carbono e nitrogênio na planta varia de acordo com a maturidade do tecido vegetal (Hodgson 1990), havendo mudanças na composição química da parede celular.

A correlação positiva observada entre os teores de PB com o nitrogênio (N) pode ser explicada pelo seu papel fundamental na fisiologia vegetal. Esse macronutriente constitui componentes celulares como a clorofila, aminoácidos e ácidos nucleicos (Taiz et al., 2017). O uso correto da adubação nitrogenada pode resultar em maior produção de forragem e aumentos dos teores de PB (Delevatti et al., 2019a). Nesse sentido, Paciullo et al. (2017) relataram incrementos de PB em cultivares de Massai e Tanzânia em resposta ao aumento das doses de N.

Foi avaliado o potencial de predição de proteína bruta da folha de pastos de capim-tamani com modelos de aprendizado de máquina. Utilizou-se um banco de dados robusto coletado durante 18 meses de pastos de capim-tamani com duas doses de nitrogênio (80 e 240 kg/ha/ano de N) e duas frequências de pastejo (90 e 95%IL). A escolha das variáveis usadas como preditoras ocorreu baseada em sua facilidade de obtenção e baixo custo, uma vez que, determinar o teor de PB da forragem requer coletas destrutivas e análises laboratoriais de alto custo. O que torna necessário o estudo e aplicabilidade de métodos

indiretos para obter informações importantes nos sistemas de produção animal em pastagens com menor custo (Bretas et al., 2021).

O aprendizado de máquina apresentou potencial para prever o teor de PB da folha, com base no desempenho observado, $r=0.78$ e $R^2=49.07\%$ para o modelo RF e $r=0.78$ e $R^2= 56.65\%$ para o modelo XGBoost. Esses resultados foram obtidos a partir de informações de baixo custo e fácil acesso, o que potencializa sua aplicabilidade. Nenhuma das variáveis utilizadas exige uso de drones, imagens de satélites ou análises laboratoriais, podendo ser obtidas por meio de registros que normalmente já fazem parte da maioria dos testes dos campos experimentais. Esses resultados demonstram a importância das variáveis escolhidas e destaque para o desempenho superior do modelo XGBoost em relação ao RF.

Ramos et al. (2020) verificaram que o modelo RF apresenta bom desempenho ($r=0,72$ a $0,78$) para prever variáveis de finalidade agrícola. Bretas et al. (2023) encontraram desempenho preditivo para biomassa fresca acima do solo e concentração de matéria seca de ($R^2= 61\%$ e 69% , respectivamente) utilizando RF. Dos Reis et al. (2020) observaram desempenho superior para XGBoost ($R^2= 65\%$ e 89%) para biomassa fresca acima do solo e altura, respectivamente quando comparado com RF ($R^2= 60\%$ e 88%) para biomassa fresca acima do solo e altura, respectivamente. Vale ressaltar que esses autores utilizaram imagens que requerem o uso de equipamentos específicos como drones, câmeras multiespectrais ou satélites, além de condições ambientais favoráveis para capturas consistentes (Kamilaris e Prenafeta, 2018). Além disso, processar as imagens demanda maior capacidade computacional e consequentemente custos mais elevados.

Outra limitação importante refere-se à complexidade do pré-processamento e da anotação dos dados. O uso de imagens geralmente exige etapas como correção geométrica, segmentação e rotulagem manual, o que pode ser demorado e sujeito a erros humanos (Lu e Weng, 2007). Ademais, os modelos baseados em visão computacional tendem a ser menos interpretáveis em comparação aos modelos baseados em dados tabulares, o que pode dificultar a adoção por usuários que demandam explicações claras para a tomada de decisão (Doshi Velez e Kim, 2017).

O IEP se destaca como variável mais influente no teor de PB, no entanto é uma variável resultante do manejo e condições ambientais adequadas que quando somadas a adubação nitrogenada permitem encurtar o período de descanso das pastagens de clima

tropical (Cunha et al., 2022). Isso resulta em pastos mais jovens e com maior teor de PB ao passo que, com o aumento da maturidade ocorre efeito do decréscimo da proteína e aumento dos componentes fibrosos (Zubieta et al., 2021).

O nitrogênio está entre as variáveis que mais influencia o teor de PB, pois quando utilizado adequadamente potencializa os processos de fotossíntese para fixação de carbono em biomassa, ocasionando aumento dos constituintes celulares, além de atuar em zonas de divisão e alongamento foliar e em zonas meristemáticas de crescimento de tecidos vegetais (Taiz et al., 2017; Galindo et al., 2018; Li et al., 2019). Assim, o nitrogênio possui efeito direto sobre o fluxo de tecidos permitindo redução no intervalo entre pastejo (Zanine et al., 2020), o que influencia a época de colheita e qualidade da forragem.

Além dos fatores IEP e doses de N serem os atributos de maior importância em modelos de predição de PB da folha, as alturas de pré e pós-pastejo e Águas20/21 também se destacam. A definição de metas de pastejo exerce influência na qualidade da forragem disponível ao pastejo animal (Alvarenga et al., 2020). Quando estabelecidas adequadamente, permitem aumentar significativamente a produtividade do pasto pois o acúmulo de matéria seca está fortemente correlacionado com a frequência de pastejo, que interage com a intensidade da desfolhação. Isso afeta diretamente a produtividade da forragem, influenciando seus componentes e, conseqüentemente, sua composição química ao longo do ano (Da Silva et al., 2015).

A precipitação, radiação solar e temperatura e estações do ano possuem menor influência sobre o teor de PB quando comparada as variáveis de manejo. Vale destacar que as estações do ano são constituídas pelo conjunto de características ambientais, sendo elas radiação solar, temperatura e disponibilidade hídrica. A luz solar aumenta a relação entre conteúdo celular e parede celular, pois aumenta a formação de carboidratos não estruturais e aminoácidos, o que pode resultar em melhora da qualidade de plantas forrageiras (Van Soest, 1994). A temperatura atua diretamente nos processos bioquímicos, físicos e morfogênicos, influenciando a qualidade da forragem (Gillet, 1984; Reis et al., 2013). Altas temperaturas resultam em maior lignificação da parede celular o que diminui a relação entre conteúdo celular e parede celular (Van Soest, 1994). A disponibilidade hídrica tem papel fundamental no metabolismo vegetal por ser responsável pelo

transporte de nutrientes, via de absorção e meio de dissipar energia solar e excesso de temperatura (Lemaire, 2001) atuando de forma indireta no valor nutritivo da forragem.

O que explica o fato de Águas 20/21, diferente das demais estações, ter grande influência no teor de PB pois foi o período em houve maior precipitação que propiciou maior número de ciclos de pastejos durante o período experimental. O ecossistema pastagem é dinâmico e sensível a mudanças de manejo e alterações climáticas. Chen et al. (2021) ressaltaram a importância de dados meteorológicos como temperatura e precipitação como fatores influenciadores no desenvolvimento do pasto. O estágio de maturidade da planta forrageira é considerado fator primário na alteração dos componentes de valor nutritivo, enquanto variações ambientais como umidade e fertilidade do solo, são fatores secundários (Tesk et al., 2018). O estágio de maturidade da planta pode ser controlado pelas estratégias de manejo exercendo maior influência sobre a composição química da planta do que os fatores abióticos. Os resultados confirmam isso, uma vez que as variáveis IEP, dose de N, altura de pré e pós-pastejo estão entre os atributos que mais contribuíram na predição do modelo e as variáveis que apresentaram maiores correlações com o teor de PB do capim-tamani. Estratégias de manejo adequadas devem ser adotadas quando o objetivo é produzir pastos com boa composição química independente dos fatores abióticos.

Conclusão

O uso de informações como dados climáticos, adubação nitrogenada, intervalo entre pastejo, altura de pré e pós-pastejo e interceptação de luz são importantes para estimar a proteína bruta da folha de pastos de capim-tamani.

As variáveis com maior importância preditora são intervalo entre pastejo, doses de nitrogênio, alturas de pré e pós-pastejo reforçando que o manejo interfere mais na composição química do que as condições ambientais.

Os modelos de aprendizado de máquina Random Forest e, principalmente o XGBoost apresentaram resultados promissores mesmo utilizando variáveis de fácil obtenção e baixo custo. Esse estudo constitui uma abordagem promissora para fornecer informações sobre o teor de proteína bruta da folha, parâmetro importante no ajuste das exigências nutricionais em sistemas de produção animal em pastagens. No entanto, o tamanho

1053 reduzido do banco de dados pode limitar a generalização dos resultados, indicando a
 1054 necessidade de estudos com amostras maiores.

1055 Referências

- 1056
 1057 Alvarenga, C. A. F., Euclides, V. B. P., & Montagner, D. B. et al. (2020). Animal
 1058 performance and sward characteristics of Mombaça guineagrass pastures subjected to two
 1059 grazing frequencies. *Tropical Grasslands-Forrajes Tropicales*, 8(1), 1-10.
 1060 [https://doi.org/10.17138/TGFT\(8\)1-10](https://doi.org/10.17138/TGFT(8)1-10)
- 1061 Balseca, D. G., Rivas, E. C., López, H. B., Guevara, H. P., & Martínez, J. C. (2015).
 1062 Nutritional value of Brachiarias and forage legumes in the humid tropics of Ecuador.
 1063 *Ciencia e investigación agraria: Revista latinoamericana de ciencias de la agricultura*,
 1064 42(1), 57-63.
- 1065 Belgiu, M., & Dragu, L. (2016). Random forest in remote sensing: A review of
 1066 applications and future directions. *ISPRS Journal of Photogrammetry and Remote*
 1067 *Sensing*, 114, 24–31.
- 1068 Bisong, E. (2019). *Building machine learning and deep learning models on Google cloud*
 1069 *platform* (pp. 59-64). Berkeley, CA: Apress.
- 1070 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
 1071 <https://doi.org/10.1023/A:1010933404324>
- 1072
 1073 Bretas, I. L., Valente, D. S., Silva, F. F., Chizzotti, M. L., Paulino, M. F., D'Áurea, A. P.,
 1074 ... & Chizzotti, F. H. (2021). Prediction of aboveground biomass and dry-matter content
 1075 in Brachiaria pastures by combining meteorological data and satellite imagery. *Grass and*
 1076 *Forage Science*, 76(3), 340-352. <https://doi.org/10.1111/gfs.12517>
- 1077
 1078 Bretas, I. L., Valente, D. S. M., De Oliveira, T. F., Montagner, D. B., Euclides, V. P. B., &
 1079 Chizzotti, F. H. M. (2023). Canopy height and biomass prediction in Mombaça guinea
 1080 grass pastures using satellite imagery and machine learning. *Precision Agriculture*, 1-25.
 1081 <https://doi.org/10.1007/s11119-023-10017-6>
- 1082
 1083 Carnevalli, R. A., Da Silva, S. C., BUENO, A. D. O., Uebele, M. C., Bueno, F. O.,
 1084 Hodgson, J., ... & Morais, J. P. G. (2006). Herbage production and grazing losses in
 1085 Panicum maximum cv. Mombaça under four grazing managements.
- 1086
 1087 Chen, Y., Guerschman, J., Shendryk, Y., Henry, D., & Harrison, M. T. (2021). Estimating
 1088 pasture biomass using Sentinel-2 imagery and machine learning. *Remote Sensing*, 13(4),
 1089 603. <https://doi.org/10.3390/rs13040603>
- 1090
 1091 Costa, N. L., Jank, L., Magalhães, J. A., Rodrigues, B. H. N., & Santos, F. J. S. (2019).
 1092 Resposta de pastagens de Megathyrsus maximus cv. Zuri à frequência de desfolhação.
 1093 *Research Society and Development*, 8(8), e1185. <https://doi.org/10.33448/rsd-v8i8.1185>
- 1094
 1095 Cunha, A. M. Q., Macedo, V. H. M., De Oliveira, J. K. S., Melo, D. De M., Domingues,
 1096 F. N., Cândido, E. P., Faturi, C., & Do Rêgo, A. C. (2022). Nitrogen fertilisation as a
 1097 strategy for intensifying production and improving the quality of Massai grass grown in

- a humid tropical climate. *Journal of Plant Nutrition*, 45(14), 2213-2227.
<https://doi.org/10.1080/01904167.2022.2046078>
- Da Silva, A. K. V., Borges, M. V. V., Batista, T. S., Da Silva Junior, C. A., Furuya, D. E. G., Prado Osco, L., Pistori, H. (2021). Predicting eucalyptus diameter at breast height and total height with UAV-based spectral indices and machine learning. *Forests*, 12(5), 582.
<https://doi.org/10.3390/f12050582>
- Da Silva, S. C., Sbrissia, A. F., & Pereira, L. E. T. (2015). Ecophysiology of C4 forage grasses-understanding plant growth for optimizing their use and management. *Agriculture*, 5(3), 598-625. <https://doi.org/10.3390/agriculture5030598>
- Delevatti, L. M., Cardoso, A. S., Barbero, R. P., Leite, R. G., Romanzini, E. P., Ruggieri, A. C., & Reis, R. A. (2019). Effect of nitrogen application rate on yield, forage quality, and animal performance in a tropical pasture. *Scientific Reports*, 9(1), 7596.
<https://doi.org/10.1038/s41598-019-44138-x>
- Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608.
- Freitas, E. C. S., de Paiva, H. N., Neves, J. C. L., Marcatti, G. E., & Leite, H. G. (2020). Modeling of eucalyptus productivity with artificial neural networks. *Industrial Crops and Products*, 146, 112149. <https://doi.org/10.1016/j.indcrop.2020.112149>
- Galindo, F. S., Beloni, T., Buzetti, S., Teixeira Filho, M. C. M., Dupas, E., & Ludkiewicz, M. G. Z. (2018). Technical and economic viability and nutritional quality of mombasa guinea grass silage production. *Acta Scientiarum. Agronomy*, 40, e36395.
<https://doi.org/10.4025/actasciagron.v40i1.36395>
- Geremia, E. V., Crestani, S., Mascheroni, J. D. C., Carnevalli, R. A., Mourão, G. B., & Gillet, M. (1984). *Las gramíneas forrajeras: descripción, funcionamiento, aplicaciones al cultivo de la hierba*. Zaragoza: Acribia, 355p.
- Gornott, C., & Wechsung, F. (2016). Statistical regression models for assessing climate impacts on crop yields: a validation study for winter wheat and silage maize in Germany. *Agricultural and Forest Meteorology*, 217, 89–100.
<https://doi.org/10.1016/j.agrformet.2015.10.005>
- Hodgson, J. (1990). *Grazing management. Science into practice*. Longman Group UK Ltd.
- Horrocks, R. D., & Vallentine, J. F. (1999). Forage quality—the basics. In *Harvested Forage* (pp. 17–47). Academic Press. ISBN 9780977304561
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Kim, S.-H. (2016). Random Forests for global and regional crop yield predictions. *PLoS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). *Deep learning in agriculture: A survey*. *Computers and Electronics in Agriculture*, 147, 70-90.
<https://doi.org/10.1016/j.compag.2018.02.016>

- 1148 Le Maire, G., Marsden, C., Nouvellon, Y., Grinand, C., Hakamada, R., Stape, J. L.,
1149 Laclau, J. P. (2011). MODIS NDVI time-series allow the monitoring of Eucalyptus
1150 plantation biomass. *Remote Sensing of Environment*, 115, 2613–2625.
1151 <https://doi.org/10.1016/j.rse.2011.05.017>
- 1152 Lemaire, G. (2001). Ecophysiology of grasslands: dynamic aspects of forage plant
1153 populations in grazed swards. *International Grassland Congress*, 19, 29–37.
1154
- 1155 Li, P., et al. (2019). Elevated carbon dioxide and nitrogen supply affect photosynthesis
1156 and nitrogen partitioning of two wheat varieties. *Journal of Plant Nutrition*, 42(11–12),
1157 1290–1300. <https://doi.org/10.1080/01904167.2019.1616758>
1158
- 1159 Lopes, C. M., Paciullo, D. S. C., Araújo, S. A. C., Gomide, C. A. M., Morenz, M. J. F., &
1160 Villela, S. D. J. (2017). Massa de forragem, composição morfológica e valor nutritivo de
1161 capim-braquiária submetido a níveis de sombreamento e fertilização. *Arquivo Brasileiro*
1162 *de Medicina Veterinária e Zootecnia*, 69(1), 225–233. [https://doi.org/10.1590/1678-](https://doi.org/10.1590/1678-4162-9201)
1163 [4162-9201](https://doi.org/10.1590/1678-4162-9201)
1164
- 1165 Lu, D., & Weng, Q. (2007). *A survey of image classification methods and techniques for*
1166 *improving classification performance*. *International Journal of Remote Sensing*, 28(5),
1167 823–870. <https://doi.org/10.1080/01431160600746456>
- 1168 Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W.,
1169 Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando
1170 Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., et al. (2020). A random forest ranking
1171 approach to predict yield in maize with UAV-based vegetation spectral indices.
1172 *Computers and Electronics in Agriculture*, 178, 105791.
1173 <https://doi.org/10.1016/j.compag.2020.105791>.
1174
- 1175 McKinney, W., & Team, P. D. (2015). Pandas—Powerful Python Data Analysis Toolkit.
1176 *Pandas—Powerful Python Data Analysis Toolkit*, 1625.
1177
- 1178 Moore, David S. & McCabe, George. (2004). *Introduction to the practice of statistics*.
1179 New York, Freeman.
1180
- 1181 Mwendia, S. W., Ohmstedt, U., Nyakundi, F., Notenbaert, A., & Peters, M. (2022). Does
1182 harvesting Urochloa and Megathyrsus forages at short intervals confer an advantage on
1183 cumulative dry matter yields and quality? *Journal of the Science of Food and Agriculture*,
1184 102(2), 750–756. <https://doi.org/10.1002/JSEFA.11407>.
1185
- 1186 Nabinger, C. (1997). Eficiência do uso de pastagens: disponibilidade e perdas de
1187 forragem. In *Simpósio sobre Manejo da Pastagem*, 14, 213–251.
1188
- 1189 Paciullo, D. S. C., Gomide, C. A. M., Castro, C. R. T., Maurício, R. M., Fernandes, P. B.,
1190 & Morenz, M. J. F. (2017). Morphogenesis, biomass and nutritive value of Panicum
1191 maximum under different shade levels and fertilizer nitrogen rates. *Grass and Forage*
1192 *Science*, 72(3), 590–600. <https://doi.org/10.1111/gfs.12264>
1193
- 1194 Pedreira, B. C., Pedreira, C. G. S., & da Silva, S. C. (2009). Acúmulo de forragem durante
1195 a rebrotação de capim-xaraés submetido a três estratégias de desfolhação. *Revista*
1196 *Brasileira de Zootecnia*, 38(4), 618–625.

- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org/>.
- Raj, P., & Evangeline, P. (2020). The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases. Academic Press, USA.
- Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., ... & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178, 105791.
- Dos Reis, A. A., Werner, J. P., Silva, B. C., Figueiredo, G. K., Antunes, J. F., Esquerdo, J. C., ... & Magalhães, P. S. (2020). Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery. *Remote Sensing*, 12(16), 2534. <https://doi.org/10.3390/rs12162534>
- Reis, R. A., Bernardes, T. F., & Siqueira, G. R. (2013). Forragicultura: ciência, tecnologia e gestão dos recursos forrageiros. Jaboticabal: Gráfica Multipress.
- Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P. R., & Rowlings, D. W. (2021). Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, 180, 105880. <https://doi.org/10.1016/j.compag.2020.105880>
- Rouquette, F. M. (2016). Invited review: The roles of forage management, forage quality, and forage allowance in grazing. *The Professional Animal Scientist*, 32(1), 10–18. <https://doi.org/10.15232/pas.2015-01408>
- Saha, S., Saha, M., Mukherjee, K., Arabameri, A., Ngo, P. T. T., & Paul, G. C. (2020). Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. *Science of the Total Environment*, 730, 139197. <https://doi.org/10.1016/j.scitotenv.2020.139197>
- Scikit-learn Documentation. (2024). One Hot Encoding. Recuperado em 20 de agosto de 2024, de <https://scikit-learn.org/stable/modules/preprocessing.html#one-hot-encoding>.
- Scott, A.J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507–512.
- Štepanovský, M., Ibrová, A., Buk, Z., & Velemínská, J. (2017). Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic science international*, 279, 72-82. <https://doi.org/10.1016/j.forsciint.2017.08.005>
- t’Mannetje, L., & Jones, R. (2000). *Field and Laboratory Methods for Grassland and Animal Production Research*. CABI Publishing: Oxon, UK.
- Taiz, L., Zeiger, E., Moller, I. M., & Murphy, A. (2017). *Fisiologia e desenvolvimento vegetal Diversidade vegetal* (6ª ed.). Artmed, Porto Alegre.

- Team, R.C. (2013). R: A Language and Environment for Statistical Computing.
- Tesk, C. R. M., Pedreira, B. C., Pereira, D. H., Pina, D. D. S., Ramos, T. A., & Mombach, M. A. (2018). Impact of grazing management on forage qualitative characteristics: a review. *Scientific Electronic Archives*, 11(5), 188-197. <https://doi.org/10.36560/1152018667>.
- Vahedi, A. A. (2017). Monitoring soil carbon pool in the Hyrcanian coastal plain forest of Iran: Artificial neural network application in comparison with developing traditional models. *CATENA*, 152, 182–189. <https://doi.org/10.1016/j.catena.2017.01.022>.
- Van Soest, P. J. (1994). *Nutritional ecology of the ruminant* (2^a ed.). Cornell University Press, New York.
- Vani, S., Sukumaran, R. K., & Savithri, S. (2015). Prediction of sugar yields during hydrolysis of lignocellulosic biomass using artificial neural network modeling. *Bioresource Technology*, 188, 128–135. <https://doi.org/10.1016/j.biortech.2015.01.083>
- Zubieta, A. S., Savian, J. V., Souza Filho, G., et al. (2021). Does grazing management provide opportunities to mitigate methane emissions by ruminants in pastoral ecosystems? *Science of the Total Environment*, 754. <https://doi.org/10.1016/j.scitotenv.2020.142029>

Capítulo 3. Interpretabilidade do modelo de aprendizado de máquina usando valores SHAP: aplicação a uma tarefa de predição de proteína bruta de pastos de capim-tamani

Resumo: Modelos de aprendizado de máquina tem sido utilizados em tarefas de predição na agropecuária. No entanto, compreender por que um modelo faz uma predição é tão importante quanto predizer. O objetivo foi interpretar o impacto das variáveis de altura de pré-pastejo (ALTpre) e pós-pastejo (ALTpos), intervalo entre pastejo (IEP), dose de nitrogênio (N), interceptação de luz (IL), precipitação, temperatura (TEMP) e radiação solar (RAD) atuam no modelo XGBoost na tarefa de predição de proteína bruta da folha (PB) de pastos de capim-tamani. A técnica Shapley Additive Explanations (SHAP) foi utilizada para entender a força e direção de cada variável na predição do modelo. O impacto de IEP depende da interação com outras variáveis, baixos valores de IEP podem ter impacto positivo e negativo na predição. A dose 240 kg/ha/ano de N possui impacto positivo no modelo e 80 kg/ha/ano de N, impacto negativo. Observou-se que precipitações entre 100 e 180 mm elevam e valores fora desse intervalo diminuem o teor de PB das folhas de capim-tamani. O uso de 240 kg/ha/ano de N em pastos de capim-tamani reduz a influência das condições ambientais sobre o teor de PB. Quando se utiliza 80 kg/ha/ano de N, o teor de PB da folha do capim-tamani sofre maior influência da precipitação. Os resultados reforçam a importância do manejo integrado das alturas, metas de manejo do pastejo, adubação nitrogenada, intervalo entre pastejos como estratégias-chave para elevar o teor de proteína bruta da folha de capim-tamani.

Palavras-chave: algoritmos, shapley additive explanations, manejo de pastagens, precision livestock farming, *Panicum maximum*.

Introdução

A produtividade dos sistemas de produção animal que utilizam o pasto como principal fonte de alimentação para ruminantes pode ser otimizada com a adoção de tecnologias e práticas de manejo aprimoradas (Wani e Habibi, 2022). Nesse cenário, monitorar a qualidade do pasto é essencial para determinar o momento ideal para o pastejo, equilibrar a produção e qualidade da forragem e permitir aumento do desempenho animal (Silva et al., 2020c).

Assim, informações precisas e confiáveis sobre as pastagens possuem importância para a pesquisa que busca novas tecnologias capazes de otimizar os índices produtivos, economizando recursos, mão de obra e terras (Manyika et al. 2017; Rööös et al., 2017). Isso exige o desenvolvimento e adoção de ferramentas que possibilitem obter informações mais rápidas e menos onerosas.

Com os avanços contínuos das tecnologias baseadas em inteligência artificial, as técnicas de aprendizado de máquina assumem destaque, pois conseguem analisar e obter insights com bom desempenho o que possibilita a tomada de decisão mais rápida (Morota et al., 2018; Wolfert et al., 2017). No entanto, modelos de aprendizado de máquina podem ser altamente complexos, o que dificulta a compreensão do processo que leva a uma determinada predição, mesmo quando se tem acesso ao que o modelo aprendeu. Estes modelos são vistos como ‘caixa preta’, e fornecem insights limitados (Lubo-Robles et al., 2020). A necessidade de entender por que um modelo faz determinada predição pode ser tão importante quanto à precisão da predição, permitindo gerar confiança e insights aos usuários (Lundberg et al., 2017).

Diante disso, a técnica SHapley Additive exPlanations (SHAP) possibilita a explicação de modelos, demonstrando como a contribuição de cada variável afeta o modelo (Lundberg et al., 2017). O objetivo da técnica é explicar a predição de uma amostra x através do cálculo da contribuição de cada variável (Molnar et al., 2020). Assim, o SHAP atribui a cada variável um valor de importância baseado no seu impacto na predição, quando a variável está presente ou não durante a estimativa (Lundberg et al., 2018; Lundberg et al., 2020), permitindo entender a força e a direção de cada variável o que o possibilita a interpretação do modelo tanto de forma global como local.

O objetivo do trabalho foi utilizar a técnica SHAP para interpretar o impacto das variáveis de alturas de manejo, intervalo entre pastejo, doses de nitrogênio, interceptação

de luz, precipitação, temperatura e radiação solar no modelo XGBoost na tarefa de predição do teor de proteína bruta da folha (PB) de pastos de capim-tamani.

Material e Métodos

Descrição dos dados

Os dados foram coletados de uma área de pastagem da Embrapa Gado de Corte, formada com 0.96 ha de capim-tamani, manejado intensivamente no período de outubro de 2020 a abril de 2022. Os pastos foram manejados com os tratamentos constituídos por duas frequências de pastejo representadas pela interceptação de luz (IL) 90% de IL e 95% de IL e duas doses de nitrogênio (N) 80 e 240 kg/ha/ano de N com delineamento em blocos casualizados, em esquema fatorial 2x2. A intensidade de pastejo foi mantida fixa em 50% da altura do pasto no pré-pastejo para os quatro tratamentos. O método de pastejo foi *mob stocking* (Allen et al., 2011).



Figura 1. Área experimental de pastos de capim-tamani delimitado em vermelho (Fonte: Silveira, 2021).

As variáveis temperatura, precipitação e radiação solar do período experimental foram coletados pela estação meteorológica da Embrapa Gado de Corte, distante

aproximadamente 2,4 km da área experimental. Os dados foram organizados nas seguintes estações Águas 20/21 (dezembro de 2020 até 28 de fevereiro de 2021), transição Águas Seca 21 (março e abril de 2022), Seca Águas 21 (primeiro ciclo de pastejo de cada tratamento após o período seco), Águas 21/22 (dezembro a 28 de fevereiro de 2022) e Águas Seca 22 (março e abril de 2022).

As estações, temperatura, radiação solar, precipitação, dose de nitrogênio (N) (80 e 240 kg/ha/ano de N), interceptação de luz (IL) (90% e 95% IL), intervalo entre pastejo (IEP) e altura de pré-pastejo (ALTpré) e altura pós-pastejo (ALTpós) foram utilizadas como variáveis de entrada no modelo de aprendizado de máquina para prever o teor de proteína bruta da folha (PB) de pastos de capim-tamani.

O conjunto de dados foi composto por 14 variáveis, sendo uma dependente e 13 variáveis independentes, com o total de 90 instâncias. O conjunto de dados foi dividido em 80% para treino e validação e 20% para teste com validação cruzada com 5 folds e 100 iterações. Foi realizado ajuste dos hiperparâmetros com *grid-search* no conjunto de treino.

O modelo de aprendizado de máquina utilizado foi o XGBoost que se baseia no *gradiente boosting* para otimizar o aprendizado. Sua otimização em árvores de decisão divide os dados em categorias menores baseados nos diferentes limites das variáveis de entrada. Essa abordagem cria uma estrutura hierárquica em formato de árvore, onde a raiz define o primeiro critério de divisão, os nós internos representam subdivisões intermediárias e as folhas correspondem à saída final do modelo (Bisong et al., 2019). A configuração dos hiperparâmetros utilizada foi árvores rasas de profundidade máxima de 2, 200 iterações e taxa de aprendizado de 0.1 com 60% dos dados e 60% das features por árvore.

SHapley Additive exPlanations (SHAP)

A técnica SHAP foi utilizada para interpretar e explicar os resultados de predição, além de fornecer a medida da importância de cada variável no modelo XGBoost para predição de PB. Para calcular os valores SHAP do modelo, foi utilizada a implementação TreeExplainer, desenvolvida por Lundberg et al. (2018), que possui um método eficiente e preciso em modelos de aprendizado de máquina baseado em árvores. Foram gerados gráficos de importância global das variáveis, gráficos de resumo SHAP para investigar o comportamento global do modelo e gráficos de força SHAP para examinar a predição do

modelo em amostras específicas (Lundberg et al., 2018; Lundberg et al., 2020; Molnar, 2020). Para implementação da técnica SHAP, as estações, temperatura, radiação solar, precipitação, dose de nitrogênio (N) (80 e 240 kg/ha/ano de N), interceptação de luz (IL) (90% e 95% IL), intervalo entre pastejo (IEP) e altura de pré-pastejo (ALTpré) e altura pós-pastejo (ALTpós) foram utilizadas como variáveis de entrada no modelo XGBoost para prever o teor de proteína bruta da folha (PB) de pastos de capim-tamani.

Resultados

Os experimentos foram estruturados em três avaliações buscando responder às seguintes perguntas de pesquisa (P.P.):

-P.P. 1: Qual é a contribuição individual da precipitação, radiação solar e temperatura na predição do teor de PB da folha de capim-tamani, desconsiderando o efeito da estação do ano?

-P.P. 2: Qual faixa de precipitação exerce impacto positivo na predição do teor de PB de capim-tamani?

-P.P. 3: As doses de nitrogênio influenciam a hierarquia de importância das variáveis de manejo e ambientais na predição do teor de proteína bruta da folha de capim-tamani?

As seções seguintes representam o respectivo experimento/avaliação executado com o objetivo de responder cada uma das perguntas acima.

A classificação da contribuição dos atributos/características para estimar o teor de PB das folhas do capim-tamani (Figura 2) foi feita utilizando a técnica SHAP. Todos os dados de entrada foram considerados para confrontar a importância dos atributos.

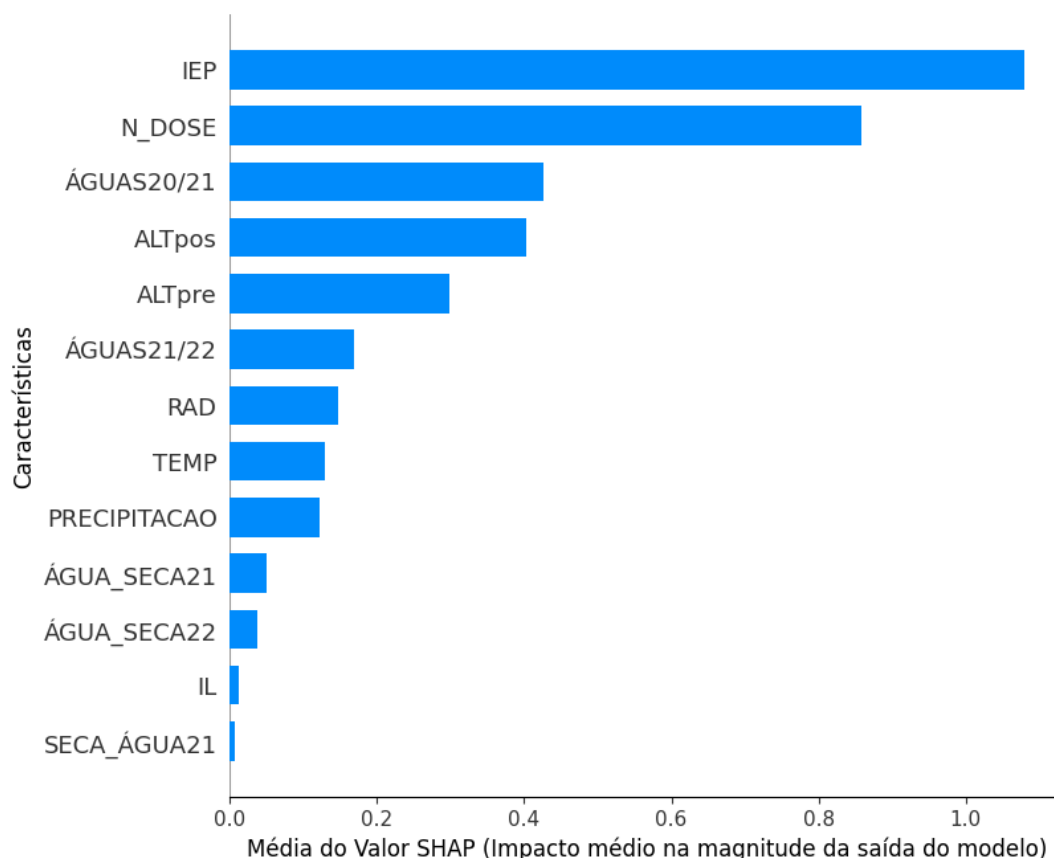


Figura 2. Interpretação global do modelo XGBoost usando valores SHAP para tarefa de predição de PB: gráfico de importância de característica global SHAP. IEP: intervalo entre pastejo, N: nitrogênio, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, IL: interceptação de luz.

O atributo que possui maior impacto na magnitude da saída do modelo é IEP, seguido das doses de N, estação Águas 20/21, altura pós-pastejo e altura pré-pastejo. As variáveis radiação, precipitação, temperatura e IL possuem menor influência, mas ainda afetaram os resultados do modelo dependendo do contexto (Figura 2).

Respondendo P.P. 1: Qual é a contribuição individual da precipitação, radiação solar e temperatura na predição do teor de PB da folha de capim-tamani, desconsiderando o efeito da estação do ano?

Como a estação é caracterizada pelo conjunto dos fatores precipitação, radiação e temperatura, ela foi excluída a fim de compreender como cada fator contribui de maneira isolada (Figura 3).

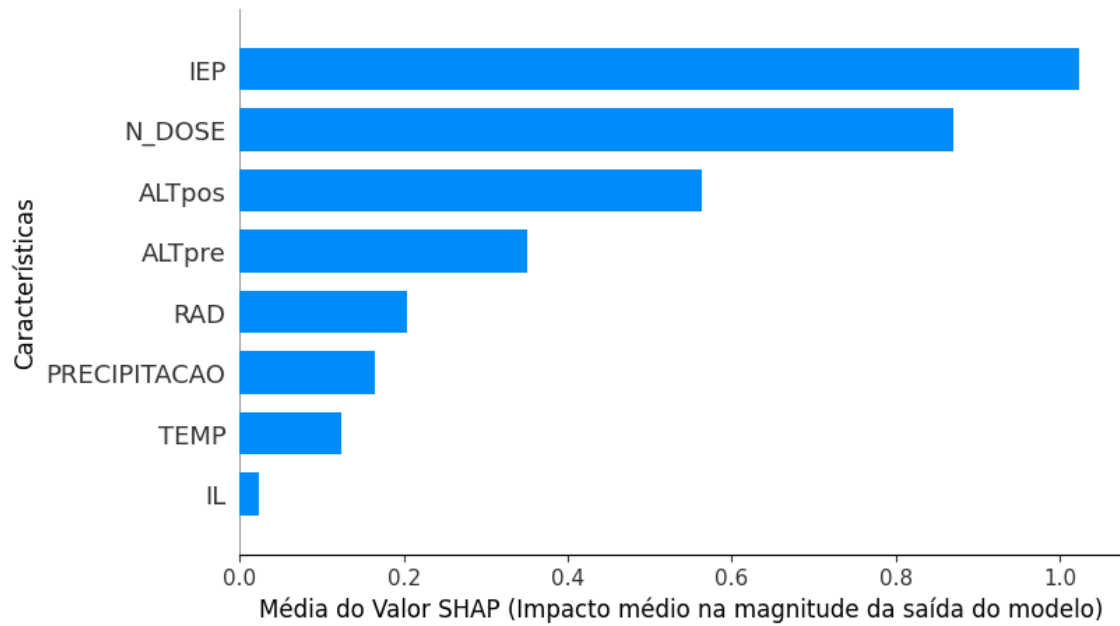


Figura 3. Interpretação global do modelo XGBoost usando valores SHAP para tarefa de predição de PB: gráfico de importância de característica global SHAP. IEP: intervalo entre pastejo, N: nitrogênio, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, IL: interceptação de luz.

Houve alteração na ordem de importância dos atributos, embora IEP e dose de N permanecerem com maior impacto, as alturas de manejo vieram em seguida e depois os fatores climáticos com menor influência (Figura 3). Entre os fatores climáticos, a ordem de importância também foi modificada, com a radiação solar ocupando a primeira posição, seguida pela precipitação e, por último, a temperatura. Respondendo a P.P. 1, pode-se afirmar que a contribuição individual dos fatores climáticos é menor em relação às variáveis de manejo IEP, doses de N e alturas de pré e pós-pastejo.

O gráfico de resumo SHAP foi utilizado para explorar a importância do atributo com sua direção e magnitude no modelo (Figura 4). Na interpretação global do modelo usando os valores SHAP, as variáveis que tem maior influência na predição estão no topo do eixo y (Figura 4a). Cada ponto na figura representa dados individuais. Os valores SHAP representam a contribuição de cada característica para o resultado do modelo em que, valores positivos sugerem que a variável tem impacto positivo na predição do modelo. As cores representam os valores da característica, a cor rosa diz que a variável tem alto valor e a cor azul refere-se a baixo valor. As variáveis IEP e dose de N foram identificadas como as de maior importância relativa, apresentando a maior dispersão de valores SHAP, o que indica forte impacto sobre a saída do modelo. O impacto do IEP na

predição depende de sua interação com outras variáveis, uma vez que, baixos valores de IEP (cor azul) podem ter impacto positivo e negativo no modelo. Altos valores de IEP (cor rosa) possuem impacto negativo na predição (Figura 4a).

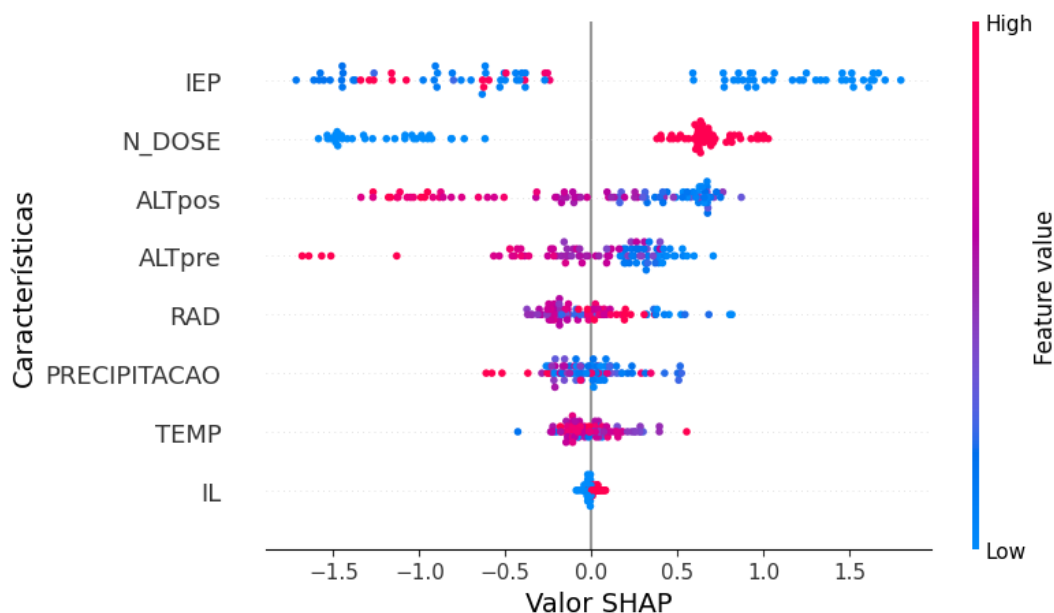


Figura 4. Interpretação global do modelo XGBoost usando valores SHAP para tarefa de predição de PB: gráfico de resumo SHAP. Cada ponto representa uma amostra do conjunto de dados, com a posição no eixo horizontal indicando o impacto da variável no valor predito pelo modelo. As variáveis estão ordenadas por importância decrescente. A coloração dos pontos representa o valor da variável (azul para valores baixos e vermelho para valores altos). IEP intervalo entre pastejo, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, IL: interceptação de luz.

Na variável dose de N, foi observado que 240 kg/ha/ano de N (cor rosa) possui impacto positivo e 80 kg/ha/ano de N (cor azul) possui impacto negativo na predição do modelo (Figura 4).

Altos valores de ALTpos, ALTpre e RAD (cor rosa) possuem impacto negativo no modelo, quanto maior o valor de cada uma delas direciona o modelo a reduzir a predição de PB. De maneira geral, baixos valores de ALTpos, ALTpre e RAD (cor azul) possuem impacto positivo no modelo, contribuindo para elevar a predição de PB na saída do modelo (Figura 4).

Respondendo P.P. 2: Qual faixa de precipitação exerce impacto positivo na predição do teor de PB de capim-tamani?

Discretizar a precipitação em quatro classes representadas pelos limites 60, 100, 180 e 1200 mm (Figura 4b) permitiu uma melhor interpretação de sua influência no modelo.

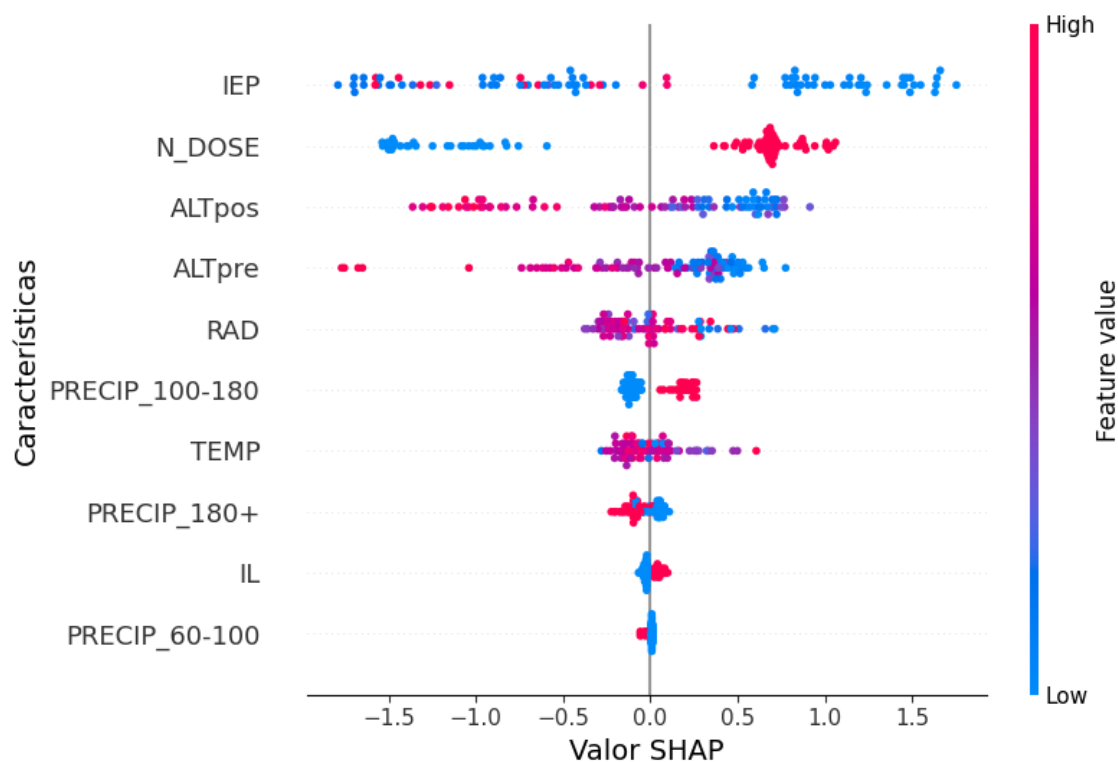


Figura 5. Interpretação global do modelo XGBoost usando valores SHAP para tarefa de predição de PB: gráfico de resumo SHAP com precipitação discretizada. Cada ponto representa uma amostra do conjunto de dados, com a posição no eixo horizontal indicando o impacto da variável no valor predito pelo modelo. As variáveis estão ordenadas por importância decrescente. A coloração dos pontos representa o valor da variável (azul para valores baixos e vermelho para valores altos). IEP: intervalo entre pastejo, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, PRECIP.100-180: precipitação discretizada 100 a 180 mm, PRECIP.180+: precipitação discretizada acima de 180 mm, PRECIP.60-90: precipitação discretizada 60 a 90 mm, IL: interceptação de luz.

Foi observado que quando a precipitação discretizada está na faixa de 100 a 180 mm (Precip 100-180 = 1), o impacto é positivo (cor rosa) no modelo e quando a precipitação está fora dessa faixa (Precip 100-180 = 0), o impacto é negativo (cor azul). Precipitação acima de 180 mm (Precip 180+ = 1) possui impacto negativo (cor rosa) no modelo. A precipitação na faixa de 60 a 100mm (Precip 60-100 = 1) e IL não tem impacto

no modelo (valor SHAP=0) (Figura 5). Respondendo a P.P. 2, a precipitação na faixa de 100 a 180 mm possui impacto positivo no modelo, elevando a predição de PB.

As Figuras 6, 7, 8, 9 e 10 exibem os gráficos de força que apresentam a interpretabilidade local usando valores SHAP de cinco amostras. Os valores SHAP representam diferentes forças que direcionam o aumento ou a diminuição das previsões do modelo Molnar (2020). Cada predição parte de um valor base, correspondente à média de todas as probabilidades para cada valor de PB no conjunto de dados, considerando a ausência de informações sobre os atributos de entrada (Lundberg e Lee, 2017). No presente estudo, o teor de PB tem uma probabilidade de valor base = 12.5% (Figuras 6, 7, 8, 9 e 10). As variáveis que contribuíram positivamente para essa elevar a predição são representadas em vermelho, enquanto as que contribuíram negativamente aparecem em azul.

Quando a dose usada foi 80 kg/ha/ano de N e a RAD de 1,761 direcionou o modelo a diminuir a predição do teor de PB de 14.5% para 13.57% (Figura 6). Quando o IEP foi de 26 dias, ALTpos de 15.5 cm, classe de precipitação no intervalo 100 a 180 mm, ALTpre de 29.3 cm direcionou o modelo a aumentar a predição. Esse conjunto de fatores resultou na predição do teor de PB da folha de capim-tamani de 13.57% (Figura 6).

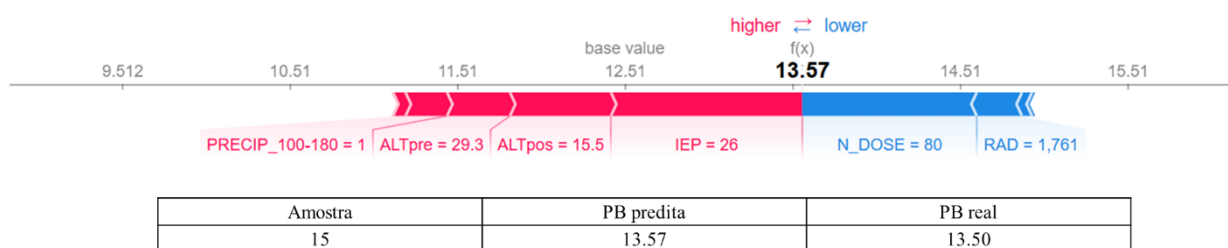


Figura 6. Gráfico de força SHAP para interpretabilidade local amostra 15.

A variável RAD com o valor de 1,421 direcionou o modelo a diminuir a predição do teor de PB das folhas do capim-tamani (Figura 7). As demais variáveis direcionaram o modelo a aumentar a predição de PB. Partindo do valor base 12.5% de PB, quando a precipitação foi no intervalo de 100 a 180 mm, 27.22°C de temperatura, ALTpre de 28.4

1535 cm, ALTpos de 13.45 cm combinado com a dose de 240 kg/ha/ano de N e o IEP de 13
 1536 dias elevaram a predição para o teor de PB para 16.19% (Figura 7).

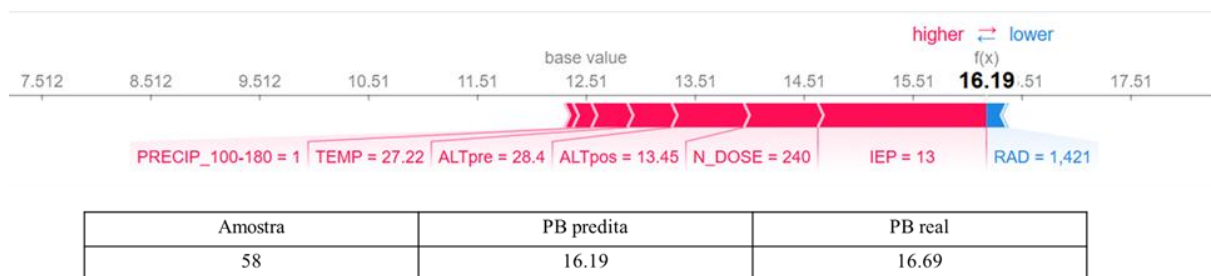


Figura 7. Gráfico de força SHAP para interpretabilidade local amostra 58.

1539 O IEP de 27 dias e a precipitação na faixa de 100 a 180 mm atuaram elevando a
 1540 predição do teor de PB das folhas do capim-tamani (Figura 8). Enquanto a temperatura
 1541 de 28.7 °C, RAD de 1,653, ALTpre de 38.9 cm, ALTpos de 19.5 cm e a dose de 80
 1542 kg/ha/ano de N diminuíram a predição do modelo, resultando em 10.73% de PB (Figura
 1543 8).

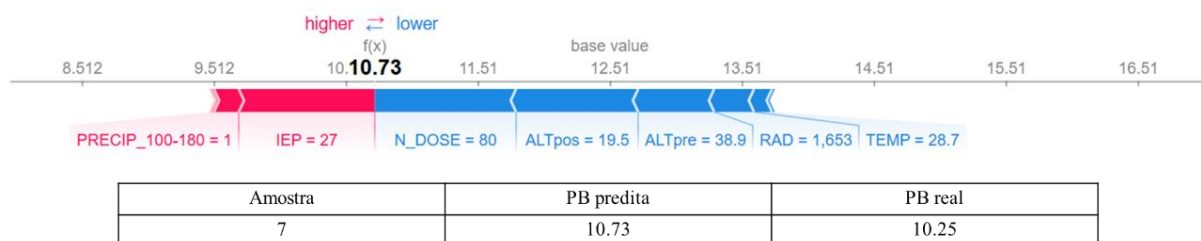


Figura 8. Gráfico de força SHAP para interpretabilidade local amostra 7.

1545 Apenas a variável dose (240 kg/ha/ano) de N contribuiu para aumentar a predição
 1546 do teor de PB das folhas do capim-tamani enquanto as demais variáveis atuaram
 1547 diminuindo a predição (Figura 9). A precipitação fora da faixa 100 a 180 mm, temperatura
 1548 de 27.42 °C e RAD de 1,475 direcionaram o modelo a predizer 12.5% (valor base) de
 1549 PB. Somado a isso, com ALTpre de 39 cm, IEP de 32 dias e ALTpos de 21 cm fez com
 1550 que o modelo diminuísse a predição de 12.5% para 10.26% de PB (Figura 9).

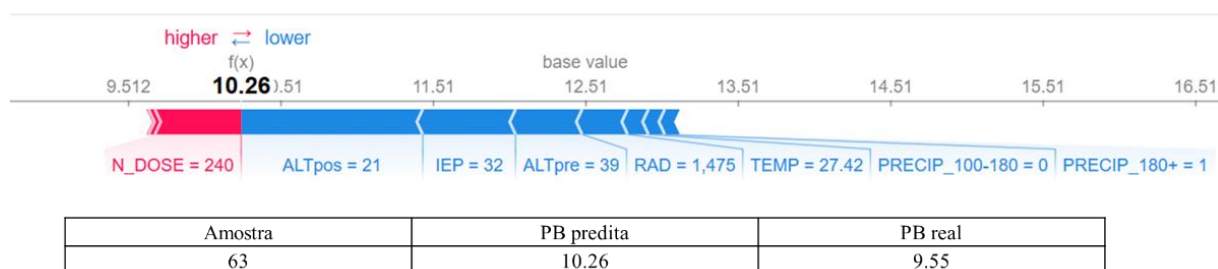
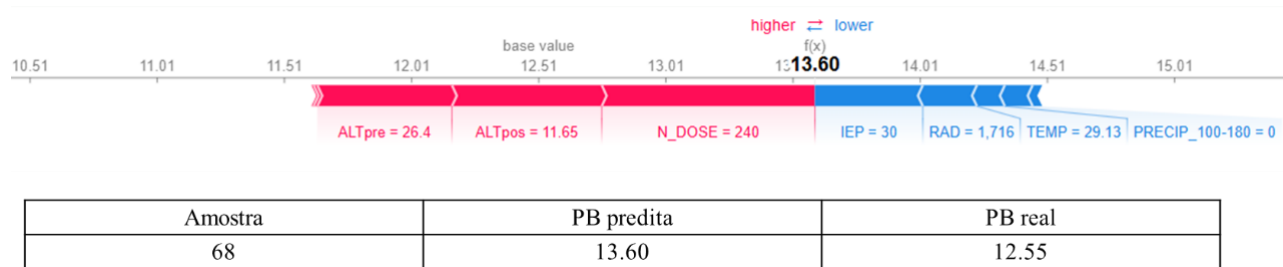


Figura 9. Gráfico de força SHAP para interpretabilidade local amostra 63.

1553 A ALTpre de 26.4 cm, ALTpos de 11.65 cm associadas a dose de 240 kg/ha/ano
 1554 de N atuaram elevando a predição do teor de PB da folha do capim-tamani. Em
 1555 contrapartida, o IEP de 30 dias, RAD de 1,716, temperatura de 29.13°C e a precipitação
 1556 fora do intervalo de 100 a 180 mm atuaram reduzindo a predição. Todos esses fatores em
 1557 conjunto resultaram na predição de 13.6% de PB (Figura 10).



1558 **Figura 10.** Gráfico de força SHAP para interpretabilidade local amostra 68.

1559

1560 Temperaturas com valores muito próximos atuaram de maneira antagônica no
 1561 modelo em que, 27.22 °C elevou (Figura 7) e 27.42 °C (Figura 9) diminuiu o teor de PB
 1562 das folhas do capim-tamani.

1563 *Respondendo P.P. 3: As doses de nitrogênio influenciam a hierarquia de importância das*
 1564 *variáveis de manejo e ambientais na predição do teor de proteína bruta da folha de*
 1565 *capim-tamani?*

1566 Com o intuito de entender se a ordem de importância das variáveis altera de acordo
 1567 com a dose de N, o banco de dados foi separado nas doses de N. A Figura 11 mostra o
 1568 impacto das variáveis na saída do modelo na dose de 80 kg/ha/ano de N (Figura 11a) e na
 1569 dose de 240 kg/ha/ano de N (Figura 11b).

1570 A ordem de importância das características foi alterada (Figura 11). Quando se
 1571 utiliza 80 kg/ha/ano de N, o atributo que possui maior impacto na magnitude da saída do
 1572 modelo é a precipitação, seguido de ALTpos e ALTpre (Figura 11a). Apenas a variável
 1573 precipitação na faixa 100 a 180mm, precipitação acima de 180 mm e IL demonstraram
 1574 impacto de forma distinta. Precipitações na classe de 100 a 180 mm (precip 100-180 = 1)
 1575 (cor rosa) e precipitações fora da classe 180+ (precip 180+ = 0) (cor azul) possuem
 1576 impacto positivo. O manejo com 90% de IL (cor azul) possui impacto positivo no modelo
 1577 enquanto o manejo com 95% de IL (cor rosa) possui impacto negativo no modelo (Figura
 1578 11a).

1579

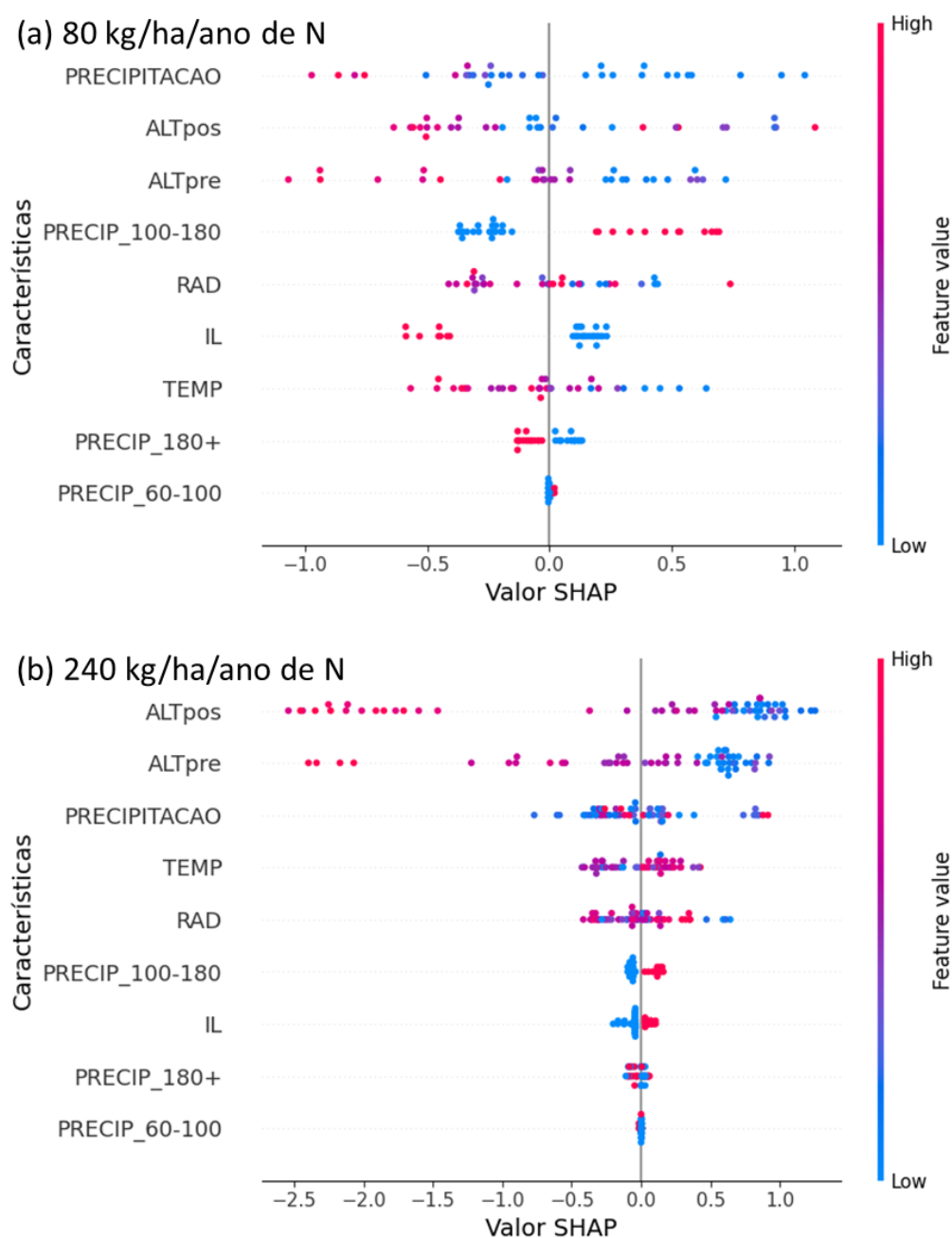


Figura 11. Gráfico de resumo SHAP para predição de PB dividido em dois bancos de dados (a) 80 kg/ha/ano de N e (b) 240 kg/ha/ano de N. Cada ponto representa uma amostra do conjunto de dados, com a posição no eixo horizontal indicando o impacto da variável no valor predito pelo modelo. As variáveis estão ordenadas por importância decrescente. A coloração dos pontos representa o valor da variável (azul para valores baixos e vermelho para valores altos). IEP: intervalo entre pastejo, ALTpre: altura pré-pastejo, ALTpos: altura pós-pastejo, TEMP: temperatura, RAD: radiação solar, PRECIP.100-180: precipitação discretizada 100 a 180 mm, PRECIP.180+: precipitação discretizada acima de 180 mm, PRECIP.60-90: precipitação discretizada 60 a 90 mm, IL: interceptação de luz.

1590

Na dose de 240 kg/ha/ano de N, a ordem de importância global fica com as variáveis de manejo: ALTpos seguido de ALTpre, no topo e depois as variáveis climáticas com menor magnitude de impacto (Figura 11b). Respondendo a P.P. 3, a ordem de importância das variáveis é alterada conforme a dose de N, em pastos que receberam 240 kg/ha/ano de N as alturas de manejo possuem maior influência. Em pastos que receberam 80 kg/ha/ano de N, a precipitação possui a maior influência sobre o teor de PB da folha do capim-tamani.

Discussão

O ecossistema pastagem é dinâmico e influenciado tanto pelo manejo quanto por fatores abióticos, como disponibilidade de água, luz, nutrientes e temperatura (Santos et al., 2012). No entanto, foi possível observar que as variáveis de manejo são mais importantes na proteína bruta da folha de pastos de capim-tamani do que os fatores ambientais (Figura 3). O IEP é determinante para o modelo, pois influencia o estágio de maturidade que afeta o valor nutritivo da planta forrageira no momento da desfolha pelo animal (Figura 4). Isso ocorre porque, após as primeiras semanas de rebrota, quando os níveis de proteína e digestibilidade são mais elevados, há a necessidade de aumento das estruturas de suporte da planta. Com isso, ocorre o crescimento do colmo, havendo a redução do conteúdo celular, altamente digestível, e a intensificação da deposição de fibra e lignina, que torna a digestão da forragem mais lenta e menos eficiente (Tesk et al., 2018). O que consequentemente diminui a taxa de ingestão do animal reduzindo o seu desempenho (Benvenuti et al., 2008).

Os valores SHAP evidenciaram um comportamento contrastante entre as doses de N avaliadas em que, a dose de 80 kg/ha/ano de N apresentou impacto negativo na predição do modelo, enquanto 240 kg/ha/ano de N exerceu impacto positivo (Figura 4). Esse resultado indica que a dose mais elevada de nitrogênio foi determinante para o aumento dos teores de PB. O nitrogênio é um dos principais fatores que limitam o crescimento da planta e, quando manejado adequadamente, potencializa a produção e o valor nutritivo forragem (Delevatti et al., 2019a, Galindo et al., 2019).

As precipitações dentro da faixa 100 a 180 mm elevaram a predição do teor de PB do capim-tamani (Figuras 5, 6 e 7). Durante a estação das águas, em pastos bem manejados, a PB e a digestibilidade in vitro da matéria orgânica da forrageira tendem a ser mais elevadas, enquanto a fibra em detergente neutro e a lignina aumentam nos períodos de

escassez de chuva (Hughes et al., 2014). Dessa forma, os efeitos positivos da quantidade de chuva no intervalo 100 a 180 mm podem ser explicados pela disponibilidade adequada de água no solo que favorece o crescimento vegetativo e a renovação foliar resultando em pastos mais jovens e de maior qualidade.

As menores alturas de pré-pastejo (ALT_{pre} menor que 30 cm) e pós-pastejo (ALT_{pos} menor que 19 cm), precipitação na faixa 100 a 180 mm e intervalo entre pastejo 26 dias elevaram o teor de PB enquanto a dose de 80 kg/ha/ano de N e a radiação atuaram reduzindo a predição na amostra 15, manejada com 90% de IL nas águas 21/22 (Figura 6). Esse resultado reforça que o manejo da altura do dossel do capim-tamani possibilita um controle mais eficaz da estrutura do pasto, influenciando diretamente sua composição química (Hodgson 1990). No entanto, na amostra 58, quando essas mesmas condições (mesmo período e IL) foram associadas à dose de 240 kg/ha/ano de N (Figura 7), o intervalo entre pastejos foi de 13 dias, metade do observado em 80 kg/ha/ano de N (Figura 6). Esse contraste evidencia o papel fundamental da adubação nitrogenada na dinâmica de rebrota, visto que a dose mais elevada de N em condições adequadas de precipitação acelerou o crescimento do capim-tamani e reduziu o tempo necessário para o dossel atingir a altura alvo para pastejo (Figura 7). Essas condições permitem otimizar o acúmulo de forragem em menor intervalo de tempo, favorecendo a eficiência do pastejo.

A predição de 10.73% inferior à média (12.5%) observada na amostra 7 (Figura 8) foi resultado de condições que limitaram o teor de PB da folha do capim-tamani. O efeito positivo da precipitação na classe 100 a 180 mm e o intervalo entre pastejo de 27 dias não foram suficientes para superar os efeitos negativos das demais variáveis. Os dados da amostra 7 correspondem ao início do manejo da pastagem no período das águas 20/21 que associado a baixa dose de N pode ter resultado em menor densidade populacional de perfilhos que ocasiona menor cobertura do solo e menor expansão foliar por unidade de área, reduzindo a eficiência do dossel na interceptação de luz (Sbrissia e Silva, 2008). Como estratégia compensatória, a planta aumentou a altura do dossel em detrimento da menor densidade para atingir a meta alvo (IL) para pastejo.

A alta dose de N atuou elevando os teores de PB do capim-tamani (Figuras 9 e 10), porém a predição variou de 10.26% a 13.60% de PB. Essa variação pode ser atribuída às diferentes combinações dos fatores ambientais e de manejo e envolvidos. A predição mais elevada (Figura 10) indica que, com a menor altura pré-pastejo, o dossel mais jovem pode

ter favorecido a renovação de tecido foliar e maior concentração de compostos nitrogenados, resultando em maior teor de PB. Além disso, a amostra 68 (Figura 10) pertence à estratégia de manejo com 90% de IL e segundo ano de manejo da pastagem. Em contrapartida, a amostra 63 (Figura 9) pertence à estratégia de manejo com 95% de IL e primeiro ano de manejo da pastagem. Nessa estratégia, o dossel atinge 95% de IL em altura maior quando comparado a 90% de IL, o que explica as maiores alturas de pré e pós-pastejo e IEP observados (Figura 9). Esses resultados estão de acordo com a literatura, que utiliza a altura do pasto como ferramenta prática para representar o nível de IL como meio de controle da estrutura do dossel forrageiro influenciando sua produtividade e composição química (Carnevali et al., 2006; Pedreira et al., 2009). O modelo reconhece que o efeito isolado da adubação não é suficiente para elevar os teores de PB quando as condições estruturais não são favoráveis.

O efeito da radiação solar em reduzir a predição da PB da folha do capim-tamani (Figuras 6,7,8,9,10) pode estar relacionado ao aumento da taxa fotossintética promovida por elevados níveis de radiação, que favorece a produção de carboidratos estruturais em detrimento da síntese de compostos nitrogenados. Sob altas intensidades de radiação, há tendência de aumento da lignificação dos tecidos vegetais, processo que contribui para a diluição do teor de proteína bruta (Van Soest, 1994).

A temperatura em torno de 27°C apresentou efeitos antagônicos no modelo (Figuras 7 e 9), reforçando a complexa interação entre planta e ambiente. Quando combinada com menores alturas de pastejo (ALT_{pre} de 28,4 cm e ALT_{pos} de 13,45 cm), IEP reduzido (13 dias), dose de 240 kg/ha/ano de N e precipitação dentro da faixa 100 a 180 mm, a temperatura em 27,22°C favoreceu o aumento da PB (Figura 7). No entanto, quando a temperatura foi 27,42°C associada à mesma dose de N, com precipitação acima de 180 mm, alturas de manejo mais elevadas (ALT_{pre} de 39 cm e ALT_{pos} de 21 cm) e IEP de 32 dias, resultou em desequilíbrio entre os fatores que influenciam a produção vegetal e nesta situação, atuaram reduzindo a PB (Figura 9). Portanto, a combinação de metas adequadas de manejo possibilita controlar o crescimento da planta em resposta à disponibilidade de fatores abióticos permitindo que a cultivar expresse seu potencial, refletindo na estrutura do pasto e, conseqüentemente, no desempenho animal (Tesk et al., 2018).

A ordem de importância global das variáveis (Figura 11) foi modificada quando o banco de dados foi dividido entre as duas doses de nitrogênio. O uso de 80 kg/ha/ano de N não foi suficiente para o capim-tamani, gramínea exigente em fertilidade, com recomendação de 100-150 e 200-300 kg/ha/ano de N para sistemas de médio a alto nível tecnológico, respectivamente (Pereira et al., 2018).

Em pastos com baixa dose de N (Figura 11a), a precipitação passa a exercer papel predominante na determinação do teor de PB da folha do capim-tamani. Nesse contexto, a precipitação entre 100 e 180 mm apresentou impacto positivo na predição da PB, sugerindo condições hídricas favoráveis ao acúmulo de compostos nitrogenados. O manejo com 95% de IL apresentou impacto negativo no teor de PB (Figura 11a) provavelmente em razão da menor densidade populacional de perfilhos devido à baixa disponibilidade de N. Essa condição pode retardar o alcance da meta de IL para o pastejo, refletindo em maiores alturas do dossel e consequente alteração na composição química da forragem.

Em pastagens que alcançaram a maturidade ou são manejadas de forma inadequada, geralmente há redução no teor de PB e aumento no teor de fibra (Pedreira et al., 2013). Diante disso, quando não há suprimento adequado de nutrientes mesmo que as metas de manejo sejam respeitadas, se não houver precipitação, o teor de PB do capim-tamani pode ficar comprometido.

Em contrapartida, quando há suprimento adequado de N (240 kg/ha/ano) (Figura 11b), as variáveis de altura de manejo ALT_{pos} e ALT_{pre} são determinantes na composição química do capim-tamani. Nessa condição, o controle da altura deve ser conduzido com rigor, pois, sob disponibilidade hídrica, o dossel atinge rapidamente a meta de pastejo. Essa dinâmica exige maior atenção ao manejo, a fim de evitar o avanço da maturidade das plantas, o que pode comprometer o teor de PB. O fornecimento adequado de N reduz o intervalo entre pastejos em pastagens tropicais (Cunha et al., 2022), ao acelerar o metabolismo das plantas e favorecer a rápida reconstituição dos tecidos foliares após o desfolhamento (Sousa et al., 2021). Sabe-se que plantas mais jovens apresentam maior teor de proteína em comparação com plantas mais maduras (Lima et al., 2019; Van Cleef et al., 2019), o que contribui para um uso mais eficiente da forragem oferecida aos animais.

A análise interpretável do modelo XGBoost com a técnica SHAP mostra-se uma ferramenta valiosa para entender como variáveis ambientais e o manejo interagem na construção da predição do modelo, permitindo identificar padrões e oferecer suporte para tomada de decisões mais assertivas.

Conclusão

De acordo com a técnica SHAP, valores de precipitação dentro da faixa 100 a 180 mm aumentam o teor de proteína bruta da folha de pastos de capim-tamani.

O uso de 240 kg/ha/ano de N em pastagens de capim-tamani reduz a influência das condições ambientais, principalmente da precipitação, sobre o teor de PB da folha.

Quando se utiliza 80 kg/ha/ano de N, o teor de proteína bruta da folha do capim-tamani sofre maior influência do ambiente, tornando a precipitação fator determinante na composição química da forragem.

Os resultados reforçam a importância do manejo integrado das alturas metas de manejo do pastejo, adubação nitrogenada com as condições climáticas como estratégias-chave para elevar a proteína bruta da folha de capim-tamani.

Referências

- Allen, V. G., Batello, C., Berretta, E. J., Hodgson, J., Kothmann, M., Li, X., McIvor, J., Milne, J., Morris, C., Peeters, A., & Sanderson, M. (2011). An international terminology for grazing lands and grazing animals. *Grass and Forage Science*, 66(1), 2–28. <https://doi.org/10.1111/j.1365-2494.2010.00780.x>
- Benvenuti, M.A., Gordon, I.J., Poppi, D.P., Crowther, R., Spinks, W. (2008). Foraging mechanics and their outcomes for cattle grazing reproductive tropical swards. *Applied Animal Behaviour Science* 113: 15-31.
- Bisong, E. (2019). *Building machine learning and deep learning models on Google cloud platform* (pp. 59-64). Berkeley, CA: Apress.
- Carnevali, R. A., Da Silva, S. C., Bueno, A. D. O., Uebele, M. C., Bueno, F. O., Hodgson, J., & Morais, J. P. G. (2006). Produção de forragem e perdas de pastejo em *Panicum maximum* cv. Mombaça sob quatro manejos de pastejo. *Revista Brasileira de Zootecnia*, 35(3), 809–818. <https://doi.org/10.1590/S1516-35982006000300004>
- Cunha, A. M. Q., Macedo, V. H. M., de Oliveira, J. K. S., Melo, D. de M., Domingues, F. N., Cândido, E. P., Faturi, C., & do Rêgo, A. C. (2022). Nitrogen fertilisation as a strategy for intensifying production and improving the quality of Massai grass grown in a humid

- 1750 tropical climate. *Journal of Plant Nutrition*, 45(14), 2213–2227.
 1751 <https://doi.org/10.1080/01904167.2022.2046078>
- 1752 Delevatti, L. M., Romanzini, P., Fabiano, J., Koscheck, W., Luis Da Ross De Araujo, T.,
 1753 Renesto, D. M., Ferrari, A. C., Pavezzi Barbero, R., Mulliniks, J. T., & Reis, R. A.
 1754 (2019a). Forage management intensification and supplementation strategy: Intake and
 1755 metabolic parameters on beef cattle production. *Animal Feed Science and Technology*,
 1756 247, 74–82. <https://doi.org/10.1016/j.anifeedsci.2018.11.004>
- 1757 Galindo, F. S., Beloni, T., Buzetti, S., Carvalho, M., Filho, M. T., Dupas, E., Gaioto, M.,
 1758 & Ludkiewicz, Z. (2018). Technical and economic viability and nutritional quality of
 1759 mombasa guinea grass silage production. *Acta Scientiarum. Agronomy*, 40(1), e36395.
 1760 <https://doi.org/10.4025/ACTASCIAGRON.V40I1.36395>
- 1761 Geremia, E. V., Crestani, S., Mascheroni, J. D. C., et al. (2018). Sward structure and
 1762 herbage intake of *Brachiaria brizantha* cv. Piatã in a crop-livestock-forestry integration
 1763 area. *Livestock Science*, 212, 83–92. <https://doi.org/10.1016/j.livsci.2018.03.020>
- 1764 Hodgson, J. (1990). Grazing management. Science into practice. Logman Group UK Ltd.
 1765 p.203.
- 1766 Hughes, M., Mlambo, V., Jennings, P. G. A., & Lallo, C. H. O. (2014). A precisão da
 1767 previsão da digestibilidade *in vitro* da matéria orgânica ruminal a partir de componentes
 1768 químicos de pastagens tropicais varia com a estação e o método de colheita. *Tropical*
 1769 *Agriculture*, 91, 131–146.
- 1770 Kunrath, T. R., Lemaire, G., Sadras, V. O., & Gastal, F. (2018). Water use efficiency in
 1771 perennial forage species: Interactions between nitrogen nutrition and water deficit. *Field*
 1772 *Crops Research*, 222, 1–11. <https://doi.org/10.1016/j.fcr.2018.02.031>
- 1773 Lima, MA; Paciullo, DS; Silva, FF; Morenz, MJ; Gomide, CA; Rodrigues, RA; Bretas,
 1774 IL; Chizzotti, FH Avaliação de um sistema silvipastoril de *Brachiaria decumbens* há
 1775 muito estabelecido: Características da planta e valor alimentar para o gado. *Crop Pasture*
 1776 *Sci.* 2019 , 70 , 814–825.
- 1777 Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K. J., & Pranter, M.
 1778 J. (2020, October). Machine learning model interpretability using SHAP values:
 1779 Application to a seismic facies classification task. In *SEG international exposition and*
 1780 *annual meeting* (p. D021S008R006). SEG. DOI:10.1190/segam2020-3428275.1
- 1781 Lundberg, S. M., B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston,
 1782 D. K.-W. Low, S. Newman, J. Kim, and S. Lee, 2018, Explainable machine-learning
 1783 predictions for the prevention of hypoxaemia during surgery: *Nature Biomedical*
 1784 *Engineering*, 2, 749–760.
- 1785 Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J.
 1786 Himmelfarb, N. Bansal, and S. Lee, 2020, From local explanations to global
 1787 understanding with explainable AI for trees: *Nature Machine Intelligence*, 2, 56–67.
- 1788 Lundberg, S. M., G. G. Erion, and S. Lee, 2018, Consistent individualized feature
 1789 attribution for tree ensembles: arXiv preprint arXiv:1802.03888.

- 1790 Lundberg, SM, & Lee, SI (2017). Uma abordagem unificada para interpretar previsões de
1791 modelos. *Avanços em sistemas de processamento de informações neurais*, 30. DOI:
1792 arXiv:1705.07874v2 [cs.AI] 25 Nov 2017
- 1793 Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst,
1794 M. (2017). *A future that works: Automation, employment, and productivity*. McKinsey
1795 Global Institute. Retrieved from [https://www.mckinsey.com/global-themes/digital-](https://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works)
1796 [disruption/harnessing-automation-for-a-future-that-works](https://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works)
- 1797 Molnar, C., 2020, Interpretable Machine Learning: A Guide for Making Black Box
1798 Models Explainable: <https://christophm.github.io/interpretableml-book/shap.html>
- 1799 Murphy, D. J., Murphy, M. D., O'Brien, B., & O'Donovan, M. (2021). A review of
1800 precision technologies for optimising pasture measurement on Irish
1801 grassland. *Agriculture*, 11(7), 600. <https://doi.org/10.3390/agriculture11070600>
- 1802 Pedreira, B. C., Pereira, L. E. T., & Paiva, A. J. (2013). Eficiência produtiva e econômica
1803 na utilização de pastagens adubadas. *SIMPÓSIO MATOGROSSENSE DE*
1804 *BOVINOCULTURA DE CORTE*, 2, 2013-17.
- 1805 Pedreira, B. C., Pedreira, C. G. S., & Silva, S. C. D. (2009). Acúmulo de forragem durante
1806 a rebrotação de capim-xaraés submetido a três estratégias de desfolhação. *Revista*
1807 *Brasileira de Zootecnia*, 38, 618-625.
- 1808 Pereira, L. E. T., Nishida, N. T., Carvalho, L. D. R., & Herling, V. R. (2018).
1809 Recomendações para correção e adubação de pastagens tropicais. *Pirassununga:*
1810 *Faculdade de Zootecnia e Engenharia de Alimentos da USP*.
- 1811 Rööß, E., Bajželj, B., Smith, P., Patel, M., Little, D., & Garnett, T. (2017). Greedy or
1812 needy? Land use and climate impacts of food in 2050 under different livestock futures.
1813 *Global Environmental Change*, 47, 1–12.
1814 <https://doi.org/10.1016/J.GLOENVCHA.2017.09.001>
- 1815 Santos, M. R., Da Fonseca, D. M., Gomes, V. M., Da Silva, S. P., Silva, G. P., & Reis, M.
1816 (2012). Correlações entre características morfogênicas e estruturais em pastos de capim-
1817 braquiária. *Ciência Animal Brasileira/Brazilian Animal Science*, 13(1), 49-56.
1818 <https://10.5216/cab.v13i1.13041>
- 1819 Silva, RO, Miotto, FRC, Neiva, JNM, Silva, LFFM, Freitas, IB, Araújo, VL, Restle, J.
1820 (2020c). Efeitos do aumento dos níveis de nitrogênio no capim Mombaça em pastagem
1821 características, composição química e desempenho da pecuária de corte nos trópicos
1822 úmidos da Amazônia. *Tropical Animal Health and Production*, 52, 3293-3300.
1823 <https://doi.org/10.1007/s11250-020-02360-0>
- 1824 Sousa, C. C. C., Montagner, D. B., Araujo, A. R., Euclides, V. P. B., Difante, G. S.,
1825 Gurgel, A. L. C., & Souza, D. L. (2021). Soil-plant interface in a Mombasa Guineagrass
1826 pasture subjected to different nitrogen doses in rotational stocking. *Revista Mexicana de*
1827 *Ciências Pecuárias*, 12(4). <https://doi.org/10.22319/rmcp.v12i4.5904>
- 1828 Tesk, C. R. M., Pedreira, B. C., Pereira, D. H., Pina, D. D. S., Ramos, T. A., & Mombach,
1829 M. A. (2018). Impact of grazing management on forage qualitative characteristics: a
1830 review. *Scientific Electronic Archives*, 11(5), 188-197
1831 <http://dx.doi.org/10.36560/1152018667>

- 1832 Van Cleef, FOS; van Cleef, EHCB; Longhini, VZ; Nascimento, TS; Ezequiel, JMB;
1833 Ruggieri, AC Desempenho em confinamento, características de carcaça e características
1834 da carne de cordeiros criados em sistemas silvipastoris. *Can. J. Anim. Sci.* 2019 , 100 ,
1835 385–388.
- 1836 Wani, S., & Habib, R. (2022). Intensification of Forage Production and Quality
1837 Parameters: A Review. 40(12), 513–521.
1838 <https://doi.org/10.9734/AJAEES/2022/v40i121834>
- 1839 Wolfert, S., Ge, L., Verdouw, C. e Bogaardt, MJ (2017). Big data na agricultura inteligente
1840 – uma revisão. *Sistemas agrícolas*, 153, 69-80.
- 1841
- 1842
- 1843
- 1844
- 1845
- 1846
- 1847
- 1848
- 1849
- 1850
- 1851
- 1852
- 1853
- 1854
- 1855
- 1856
- 1857
- 1858
- 1859
- 1860
- 1861
- 1862
- 1863
- 1864
- 1865

Considerações Finais

O uso de dados tabulares é uma ferramenta eficiente e acessível para a predição do teor de proteína bruta. A partir de variáveis simples como altura do dossel, intervalo entre pastejos, dose de nitrogênio e condições climáticas foi possível desenvolver modelos preditivos robustos, com desempenho satisfatório.

Além disso, a utilização dessas variáveis permite maior interpretabilidade dos modelos, contribuindo para a tomada de decisão no manejo das pastagens de forma mais fundamentada e precisa.

A análise interpretável do modelo XGBoost com a técnica SHAP mostra-se uma ferramenta valiosa para entender como variáveis de manejo e clima interagem na construção da predição do modelo, permitindo identificar padrões e oferecer suporte para tomada de decisões mais assertivas de manejo.

Assim, esta abordagem se mostra promissora para apoiar a intensificação sustentável da pecuária, ao mesmo tempo em que amplia o uso de ciência de dados aplicada à forragicultura.