

SISTEMAS INTELIGENTES APLICADOS NA PREDIÇÃO DE PREENHEZ DE NOVILHAS DE DIFERENTES GRUPOS GENÉTICOS ACASALADAS AOS 14-16 MESES DE IDADE

VAZ, Ricardo Zambarda¹
RAMOS, Guyomar Ubyrathan²
PACHECO, Rangel Fernandes³
ESPIGOLAN, Rafael⁴
RESTLE, João

RESUMO

A intensificação sustentável da pecuária exige alta eficiência reprodutiva. Este estudo objetivou prever a prenhez em novilhas de corte, identificando as variáveis de maior impacto. Foram utilizados dados de 98 novilhas (Charolês, Nelore e cruzas), com 14 meses de idade, avaliadas quanto ao peso corporal inicial (IBW) e final (FBW), escore de condição corporal inicial (IBS) e final (FBS), idade (AGE) e composição genética. Uma Análise de Componentes Principais (PCA) foi conduzida para exploração dos dados, onde os dois primeiros componentes (PC1 e PC2) explicaram 60,7% da variância total. O PC1 (44,6%) foi fortemente associado ao desenvolvimento corporal (peso e idade), enquanto o PC2 (16,1%) relacionou-se ao escore de condição corporal e à genética zebuína. O biplot da PCA revelou uma tendência de separação, com novilhas prenhas agrupadas com maiores pesos e idades. Subsequentemente, um modelo *Random Forest* (RF) foi treinado para classificar o status de prenhez, alcançando uma Área Sob a Curva ROC (AUC) de 0,80, indicando bom poder discriminatório. As variáveis de maior importância preditiva foram IBW (30,1%), FBW (19,9%), IBS (14,4%) e FBS (12,4%). Conclui-se que as variáveis fenotípicas ligadas ao desenvolvimento e estado nutricional foram mais determinantes para a prenhez do que a genética. A combinação da PCA e do RF demonstrou ser uma abordagem robusta para identificar fatores críticos e prever o sucesso reprodutivo em novilhas de corte.

PALAVRAS-CHAVE: análise de componentes principais, aprendizado de máquina, árvore de decisão, escore de condição corporal, peso corporal.

INTELLIGENT SYSTEMS APPLIED TO THE PREDICTION OF PREGNANCY IN HEIFERS FROM DIFFERENT GENETIC GROUPS BRED AT 14–16 MONTHS OF AGE

ABSTRACT

Sustainable intensification of livestock production requires high reproductive efficiency. This study aimed to predict pregnancy in beef heifers by identifying the variables with the greatest impact. Data from 98 heifers (Charolais, Nelore, and crossbred), aged 14 months, were used and evaluated for initial body weight (IBW) and final body weight (FBW), initial body condition score (IBS) and final body condition score (FBS), age (AGE), and genetic composition. A Principal Component Analysis (PCA) was conducted for data exploration, in which the first two components (PC1 and PC2) explained 60.7% of the total variance. PC1 (44.6%) was strongly associated with body development (weight and age), whereas PC2 (16.1%) was related to body condition score and Zebu genetics. The PCA biplot revealed a tendency for separation, with pregnant heifers clustered with higher weights and ages. Subsequently, a Random Forest (RF) model was trained to classify pregnancy status, achieving an Area Under the ROC Curve (AUC) of 0.80, indicating good discriminative power. The variables with the highest predictive importance were IBW (30.1%), FBW (19.9%), IBS (14.4%), and FBS (12.4%). It is concluded that phenotypic variables related to development and nutritional status were more determinant for pregnancy than genetics. The combination of PCA and RF proved to be a robust approach for identifying critical factors and predicting reproductive success in beef heifers.

KEYWORDS: principal component analysis, machine learning, decision tree, body condition score, body weight.

¹ Doutor em Zootecnia pela Universidade Federal do Rio Grande do Sul – UFRGS. Professor do Departamento de Zootecnia e Ciências Biológicas da Universidade Federal de Santa Maria - UFSM, Campus Palmeira das Missões (RS), Brasil.

² Acadêmico de Zootecnia pela Universidade Federal de Santa Maria - UFSM, Campus de Palmeira das Missões, Palmeira das Missões (RS), Brasil.

³ Doutor em Zootecnia Professor do Departamento de Zootecnia o Instituto Federal Farroupilha – Campus de Frederico Westphalen - IFFAR, RS, Brasil

⁴ PhD Universidade |Federal de Goiás – Goiânia – GO - Brasil

1. INTRODUÇÃO

Em um cenário de instabilidade geopolítica, mudanças climáticas e crescimento populacional, a oferta global de proteína animal tornou-se eixo estratégico para a segurança alimentar e para a economia brasileira, principalmente no Sul do Brasil, reconhecido pela produção de carne de alta qualidade (Farias *et al.*, 2023). No entanto, no extremo Sul do Brasil, é recorrente a sazonalidade forrageira, além da pressão por área, com a entrada expressiva de produção de grãos (Jaurena *et al.*, 2021; Kuplich; Capoane; Costa, 2018).

Para que a pecuária de corte evolua de forma competitiva e ambientalmente sustentável, é indispensável intensificar a produção sem ampliar a área, aumentando a taxa de prenhez e reduzindo a idade do primeiro parto das novilhas. Esses indicadores impactam diretamente a produção de bezerros e o fluxo de caixa dos sistemas produtivos (Pravia *et al.*, 2014) e a pegada de carbono da atividade (Cullen *et al.*, 2016). A otimização desses parâmetros permite maximizar a produção de bezerros, reduzir custos com a manutenção de animais improdutivos, acelerar o retorno do investimento em genética e manejo, além de uma maior produção de quilogramas de proteína por unidade de emissão de carbono (Cullen *et al.*, 2016).

A prenhez de novilhas é um processo complexo, influenciado por fatores genéticos e ambientais (Kgari *et al.*, 2020; Rovira, 1974). Dentre os fatores genéticos, a escolha do correto indivíduo é fundamental, podendo ser determinada por uma raça mais adaptada ao sistema de produção. No entanto, não são encontradas raças completas que se adaptem a todos os ambientes de produção, sendo o cruzamento de raças para o aproveitamento da heterose uma alternativa de aumento de produção e eficiência para os sistemas de produção (Restle *et al.*, 1999). Nos parâmetros ambientais, o desenvolvimento adequado, resultado de boa nutrição e bons manejos, são fatores determinantes de melhores desempenho reprodutivos (Eloy *et al.*, 2022). Na busca por eficiência nos sistemas intensivos de produção, sendo a prenhez uma característica multifatorial, e por isso complexa, cada vez mais é importante novas formas de prever variáveis-chave para os sistemas de produção.

A busca por ferramentas mais eficientes de predição de variáveis produtivas, como a prenhez, tem ganhado força nas pesquisas (Brand *et al.*, 2021; Rabagliolo *et al.*, 2023). Nesse contexto, o aprendizado de máquina (*Machine Learning*) emerge como uma tecnologia promissora, oferecendo a capacidade de analisar grandes e complexos volumes de dados e identificar padrões sutis que podem escapar à análise estatística convencional (Schleder; Fazzio, 2021).

A aplicação de algoritmos de *Machine Learning* permite a construção de modelos preditivos que podem auxiliar na tomada de decisões estratégicas, desde a seleção de animais com maior potencial produtivo até a otimização de programas de manejo (Zhang *et al.*, 2021). Em particular, a

predição da prenhez em novilhas de corte por meio de *Machine Learning* permite a identificação precoce de animais com maior ou menor probabilidade de conceber, possibilitando intervenções de manejo direcionadas e mais eficientes (Marques *et al.*, 2024). Entre os diversos algoritmos de aprendizado de máquina, as árvores de decisão e, mais especificamente, o Random Forest, destacam-se por sua robustez e capacidade de trabalhar e analisar dados complexos, realizando médias de um conjunto de árvores de decisão e evitando fenômenos de sobreajuste (*overfitting*; (Breiman, 2001).

O Random Forest, introduzido por Breiman (2001), é um método de conjunto (*ensemble*) que constrói múltiplas árvores de decisão durante o treinamento e produz a classe que é a moda das classes (classificação) ou a previsão média (regressão) das árvores individuais. A capacidade do Random Forest de identificar a importância das variáveis preditoras o torna uma ferramenta valiosa para desvendar os fatores mais influentes na prenhez de novilhas. Portanto, o estudo objetiva prever a prenhez em novilhas de corte, a partir da identificação e quantificação de variáveis fenotípicas e genéticas, utilizando de uma abordagem de aprendizado de máquina baseada em Random Forest.

2. FUNDAMENTAÇÃO TEÓRICA

Quando se planeja um sistema de produção de bovinos de corte, o tema que se ressalta e é caracterizado como a base da cadeia de produção é a fase cria, e dentro dela, a eficiência reprodutiva é um pilar fundamental na produção de bovinos, influenciando diretamente a produção de carne e leite e, conseqüentemente, na rentabilidade das propriedades rurais (Pohler *et al.*, 2020). Mas para entender sobre a eficiência produtiva das vacas de corte, é necessário compreender quais são os fatores que contribuem e formam a fase reprodutiva de uma fêmeas bovinas.

Em termos biológicos, portanto, o sucesso reprodutivo é fruto de uma conjuntura de adequada programação fetal da fêmea, do desenvolvimento puberal, da ciclicidade estral regular com ovulações férteis, e ainda, formar ambiente uterino propício para concepção e manutenção da gestação da sua prole.

Nos últimos anos, diversos estudos vêm sendo publicados, explanando ser a eficiência produtiva da fêmea bovina definida desde seu desenvolvimento em ambiente uterino, seja pela oferta adequada de pastagens, ou ainda, com corretivos de manejo como a suplementação proteica ao final da gestação da matriz. Essa correta nutrição fará com que a puberdade de sua prole, seja mais precocemente atingida, ou pelo melhor desenvolvimento celular da mesma e maior eficiência no ganho de peso corporal.

Após o parto, uma vez que a bezerra está “ao pé da vaca”, os seus desafios começam a ser mais

complicados, pois está em um ambiente hostil, onde diversos fatores bióticos e abióticos vão acontecer durante seu desenvolvimento até atingir sua puberdade, e consequentemente, o início à vida reprodutiva nos rodeios de cria.

Sendo a puberdade o marco inicial da vida reprodutiva da fêmea, ou seja, é o momento em que a bezerra, agora novilha, está apta à reprodução podendo ser inserida dentro de um sistema de acasalamento, sendo isso de extrema importância, uma vez que, quanto mais cedo este momento ocorrer, irá reduzir os custos de produção e aumentar a eficiência do sistema de produção.

No entanto, para atingir a puberdade são necessários alguns fatores essenciais. Assim, a reprodução dos bovinos é regulada por complexas interações endócrinas e metabólicas. Os hormônios metabólicos que são responsáveis pela vida reprodutiva das fêmeas, sendo eles, a leptina, IGF-1 e grelina, que sinalizam ao eixo hipotálamo-hipófise-gônadas, influenciando a liberação de GnRH e a função ovariana. Entretanto, para a novilha atingir as concentrações ideais destes hormônios, são necessárias condições mínimas de peso, escore de condição corporal e adequado nível nutricional, onde a resposta do sistema endócrino reprodutivo em condição nutricional e metabólica adequada, níveis elevados de leptina estimulam os neurônios Kiss1 no hipotálamo, levando a um aumento na liberação de GnRH pela hipófise e consequente elevação das gonadotrofinas LH e FSH, promovendo a função gonadal. Já em situação de déficit nutricional, a queda nos níveis de leptina suprime a atividade de Kiss1, reduzindo a liberação de GnRH e, por sua vez, a secreção de LH e FSH, resultando em supressão da função gonadal.

Condições de peso e escore de condição corporal, reflexo do adequado nível nutricional, através de diversos estudos, a equipe do National Research Council (NRC, 2016), categorizou a necessidade de que as fêmeas de origem europeia (*Bos taurus taurus*) atinjam aproximadamente 60% do seu peso na maturidade. Já novilhas de origem zebuína (*Bos taurus indicus*) necessitam de um aporte nutricional mais elevado, pois sua puberdade se dá por volta dos 65% do peso à maturidade.

Novilhas com peso e ganho adequados atingem a puberdade mais cedo e apresentam maior fertilidade ao longo da vida. Uma vez que novilhas apresentem precocidade à puberdade, conseguem engravidar mais cedo dentro da primeira estação de monta, consequentemente, parindo mais cedo e tendo mais tempo de recuperação de condição corporal e fatores fisiológicos para conceber no segundo período reprodutivo (Reis *et al.*, 2023).

Esta busca pela precocidade à puberdade essa pode ser influenciada por diversos fatores como raça, peso da vaca (Eloy *et al.*, 2022; NRC, 2016), velocidade de ganho e níveis nutricionais no pré e pós-desmame (Vaz *et al.*, 2012).

Uma vez que a puberdade foi atingida, é preciso entender que a fertilidade em bovinos é um fenômeno multifatorial, influenciado por atributos individuais e biológicos como a manutenção da

ciclicidade da fêmea, que possui em média 21 dias e envolve fases foliculares e luteais. Entre os principais fatores já verificados, tanto por produtores como por pesquisadores, estão os ganhos de pesos corporais, o tipo de pastagem, a carga animal, os pesos e os escores de condição corporal das fêmeas no início e ao fim do período de acasalamento, a suplementação nutricional e mineral, a raça, ou grupo genético originário de sistemas de cruzamento e as condições climáticas, todos afetando as chances de uma fêmea bovina ficar prenha.

Do ponto de vista de manejo, a adoção de estratégias nutricionais que garantam boa condição das matrizes ao parto e durante a estação de monta é crucial. Práticas como suplementação alimentar, pastagens de alto valor nutricional, são fatos esses, já verificado há décadas, como cruciais na reprodução de fêmeas bovinas. A nutrição é importante na função reprodutiva de fêmeas bovinas, sendo o peso vivo e o escore de condição corporal (ECC), parâmetros há décadas reconhecidos como base da fertilidade em novilhas e vacas no pós-parto.

Indicadores zootécnicos como taxa de prenhez, intervalo de partos e taxa de desmame são utilizados para mensurar a eficiência reprodutiva e embasar decisões de manejo e seleção genética, impactando diretamente na rentabilidade geral. Nos últimos anos, ganhou destaque também a investigação de parâmetros para predição de prenhez, ou seja, variáveis mensuradas antes ou durante a estação de monta que permitam antecipar, quais fêmeas têm maior probabilidade de emprenhar. Dentre essas variáveis, destacam-se pesos corporais, ganhos de pesos, ECC, idade, dias amamentação, histórico de doenças periparto, entre outras, além de indicadores fisiológicos (por exemplo, concentrações hormonais) e até marcadores moleculares (Marques *et al.*, 2024). Esses avanços da bioinformática têm revolucionado o setor da pecuária e gerando melhores insights para a produção de bezerros.

Marques *et al.* (2024) desenvolveram um modelo preditivo de prenhez em vacas leiteiras integrando dados de monitoramento automatizado de atividade (AAM) com fatores individuais da vaca e do ambiente. Nesse estudo, foram combinados registros de ruminação, atividade e intensidade de estro coletados por colares eletrônicos com informações de cada vaca (idade, ordem de lactação, ECC, escore de locomoção, secreção vulvar, doenças no periparto), dados do protocolo reprodutivo (momento da inseminação, touro utilizado) e variáveis ambientais (estação do ano, umidade relativa, THI no dia da inseminação). Após essa gama de dados serem analisados, constatou-se que alguns modelos preditivos conseguiram atingir 85% de acurácia na predição da prenhez de vacas em sistemas confinados.

Em outro exemplo, técnicas de aprendizado de máquina (*machine learning*) foram aplicadas a dados ômicos: Rabaglino *et al.* (2023) empregaram ML para integrar perfis de transcriptoma de embriões bovinos (blastocistos) e identificar assinaturas gênicas capazes de prever quais embriões

possuem maior competência para se desenvolver e gerar prenhez.

3. METODOLOGIA

O experimento foi desenvolvido nas dependências do Departamento de Zootecnia da Universidade Federal de Santa Maria, localizado na região fisiográfica denominada de Depressão Central do Estado do Rio Grande do Sul. A altitude é de 95 m, latitude 29° 43' sul e longitude de 53° 42' oeste. De acordo com a classificação de Köppen, o clima é subtropical (Alvares *et al.*, 2013).

Foram utilizadas 98 novilhas, as quais foram avaliadas quanto a diferentes características corporais e genéticas no período de acasalamento. A Tabela 1 apresenta a distribuição dos animais conforme o status reprodutivo (prenhas e não prenas), categoria, predominância genética, grupo genético e escores de condição corporal no início e ao final do acasalamento.

Tabela 1 – Dados descritivos os animais do experimento.

Variáveis	Categoria	Prenhas		Não Prenhas	
		N	(%)	N	(%)
Categoria	Novilha	58	59,18	40	40,82
Predominância	Taurino	44	74,58	15	25,42
	Zebuino	9	30,00	21	70,00
Grupos Genéticos	Charolês	14	56,00	11	44,00
	3343	26	89,66	3	10,34
	3344	3	60,00	2	40,00
	4433	2	50,00	2	50,00
	4434	3	30,00	7	70,00
	Nelore	1	10,00	9	90,00
	5833	4	80,00	1	20,00
	5844	5	50,00	5	50,00
ECCI_EM ¹	2,5	1	11,11	8	88,89
	3	35	61,40	22	38,60
	3,5	20	68,97	9	31,03
	4	2	66,67	1	33,33
	2,5	0	0,00	1	100

ECCF_EM ²	3	21	50,00	21	50,00
	3,5	27	64,29	15	35,71
	4	10	76,92	3	23,08

3344 - ½ CH ½ NE; 4433 - ½ NE ½ CH; 3343 - ¾ CH ¼ NE; 4434 - ¾ NE ¼ CH; 5833 - 5/8 CH 3/8 NE e 5844 - 5/8 NE 3/8 CH. ¹Escore de condição corporal no início do acasalamento; ²Escore de condição corporal ao final do acasalamento.

A Tabela 2 apresenta os valores descritivos das variáveis contínuas analisadas, incluindo idade no início do acasalamento, peso no início e ao final do acasalamento.

Tabela 2 – Análise descritiva das variáveis analisadas.

Prenhas					Não prenhas			
Variáveis	n	Mín.	Média	Max.	n	Mín.	Média	Max.
ID_EM ¹	58	353,0	417,2	459,0	40	352,0	402,1	453,0
PI_EM ³	58	202,0	279,6	382,0	40	158,0	238,1	346,0
PF_EM ⁴	58	248,0	329,2	423,0	40	192,0	286,9	390,0

¹ Idade no início do acasalamento (dias); ² Peso no início do acasalamento (kg); ³ Peso ao final do acasalamento (kg).

Os touros pais das novilhas Charolês e Nelores definidas foram os mesmos pais das novilhas mestiças. Previamente ao experimento, as fêmeas foram mantidas em pastagem natural com as suas mães até a época do desmame realizado com idade média de sete meses. Dos sete aos treze meses permaneceram em pastagem cultivada de inverno composta de Triticale (*Tritico secale*) e Azevém (*Lolium multiflorum*), com uma lotação média de 5,7 fêmeas/ha ou 1200 kg de peso vivo/ha.

Durante o período de acasalamento (dezembro a fevereiro), as novilhas foram mantidas em pastagem natural (lotação média de 3,0 novilhas/ha ou 810 kg de peso corporal/ha), recebendo suplementação concentrada na proporção de 0,42% do peso corporal.

O potreiro de pastagens naturais utilizado, anteriormente, foi adubado durante o período de inverno com introdução da cultura de aveia via semeadura direta. Após o término da cultura da aveia, a pastagem remanescente foi roçada e diferida por 45 dias (com o objetivo de aumento de matéria seca -MS), previamente ao período acasalamento.

A adaptação das novilhas ao manejo e ao suplemento teve duração de 19 dias, realizada de forma gradativa. O suplemento utilizado foi um concentrado comercial peletizado composto por casquinha de soja, farelo de trigo e melaço, totalizando 12,5 % de proteína bruta (PB) e 67,3% de nutrientes digestíveis totais (NDT).

O período total de suplementação teve duração de 90 dias, sendo essa realizada às 7:00 horas. A suplementação foi realizada em cochos com disponibilidade de área de acesso de 1,0 metro

linear/animal, com acesso pelos dois lados, para proporcionar desta forma oportunidade de consumo mais uniforme do suplemento. As novilhas não eram mochas, contemporâneas, não ocorrendo dominância. O lote de novilhas, sempre foi mantido em grupo único após o desmame, nas mesmas áreas de pastagem, sempre providas com água de boa qualidade, tendo à disposição mineralização adequada para a categoria fornecida em cochos.

As pesagens foram realizadas no nascimento, ao desmame, no início e fim do período reprodutivo, com jejum prévio mínimo de oito horas. Ainda eram realizadas pesagens periódicas a cada 21 dias, sendo essas pesagens intermediárias para o monitoramento do ganho do peso, ajuste da carga animal e durante o período reprodutivo para o ajuste do consumo de suplemento. O ganho médio diário (GMD) foi obtido pela diferença entre os pesos finais e iniciais das novilhas divididos pelo número de dias do período avaliado. Por ocasião das pesagens ainda foram realizadas as avaliações subjetivas do escore de condição corporal (Rasby *et al.*, 2014), sendo avaliada visualmente, utilizando escores de 1 a 5, onde 1= Muito magro; 2= Magro; 3= Médio; 4= Gordo e 5= Muito gordo.

O período de acasalamento, teve duração de noventa dias, sendo utilizado na sua totalidade o uso de inseminação artificial como método reprodutivo, sem a utilização de rufiões, com duas recorridas diárias para a observação do cio, sendo uma pela parte da manhã e outra pela parte da tarde. As novilhas que manifestaram cio pela manhã eram inseminadas à tarde, e as que manifestavam cio pela parte da tarde eram inseminadas na manhã do dia seguinte. Utilizou-se sêmen de um único touro da raça Aberdeen Angus (variedade Red) adquirido em uma única compra, sendo todas as doses da mesma partida. A prenhez foi avaliada 30 dias após o término do período de acasalamento através de palpação retal. Por ocasião da inseminação artificial foram avaliados o peso e idade das novilhas.

O manejo sanitário composto de vacinações seguiu o calendário da Secretaria da Agricultura – RS, e o controle de endo e ectoparasitas, foram realizados com produtos específicos, de maneira estratégica, e quando necessário, de acordo com o grau de infestação.

Os dados foram submetidos à análise utilizando o procedimento Random Forest (RF), implementado com o *RandomForestClassifier* do *scikit-learn*, para identificar as variáveis mais relevantes associadas à ocorrência de prenhez. Embora a PCA tenha sido empregada para visualização da estrutura dos dados, a seleção de variáveis no modelo RF foi realizada de forma independente, com base na redução média do índice de Gini. O classificador RF opera gerando um conjunto de árvores de decisão treinadas em subconjuntos aleatórios de amostras (*bootstrap=false*) e subconjuntos aleatórios de variáveis em cada split, melhorando a robustez do modelo e reduzindo o sobreajuste.

As variáveis explicativas submetidas ao RF foram categorizadas em quatro grupos: (1) peso

corporal (IBW e FBW), (2) escore de condição corporal (IBS e FBS), e (3) predominância genética (Breed Zebu e Breed Sint), com Taurino como valor referência e (4) idade da novilha em meses (AGE). A variável resposta foi a ocorrência de prenhez.

O conjunto de dados foi dividido aleatoriamente em dois subconjuntos, com 70% das observações utilizadas para treinamento e os 30% restantes para teste, utilizando *random_state=42* para garantir reprodutibilidade. Essa abordagem de validação interna (holdout) foi adotada como alternativa à validação externa. O modelo Random Forest foi treinado com 1.000 árvores (*n_estimators=1000*) e profundidade máxima de 4 níveis (*max_depth=4*), para priorizar divisões mais robustas e globais, alinhadas ao conhecimento de campo e visando maior interpretabilidade.

Para avaliar o desempenho do modelo, foram calculadas a matriz de confusão, a acurácia, precisão, sensibilidade (recall), especificidade, a partir destes resultados, a curva ROC (Receiver Operating Characteristic) e a área sob a curva ROC (AUC).

A importância de cada variável foi determinada com base na média de sua contribuição para a redução da impureza de Gini nos nós das árvores. Além disso, foi extraída uma árvore de decisão representativa do modelo RF (composto por 1.000 árvores), selecionada como aquela que apresentou a maior acurácia individual no conjunto de teste. Essa árvore foi exportada com a função *export_graphviz* para visualização.

Como etapa complementar, foi também treinada uma árvore de decisão completa (*DecisionTreeClassifier*), sem restrição de profundidade (*max_depth=None*), utilizando os mesmos dados de treinamento. A árvore de decisão completa foi utilizada para avaliar o potencial máximo de separação das variáveis, com vistas a fornecer um panorama completo da modelagem.

As análises estatísticas para o RF foram conduzidas em ambiente Python (versão 3.11), com uso das bibliotecas *pandas*, *numpy*, *scikit-learn*, *matplotlib* e *seaborn*, em conformidade com práticas recomendadas para modelagem supervisionada em estudos biológicos (Breiman, 2001; Hastie; Tibshirani; Friedman, 2009).

Ferramentas de inteligência artificial generativa (ChatGPT Plus, com recurso Data Analyst GPT) foram empregadas exclusivamente como suporte à automação de tarefas auxiliares, geração de scripts, formatação de resultados e revisão de comandos sem participação no delineamento analítico ou na modelagem estatística dos dados.

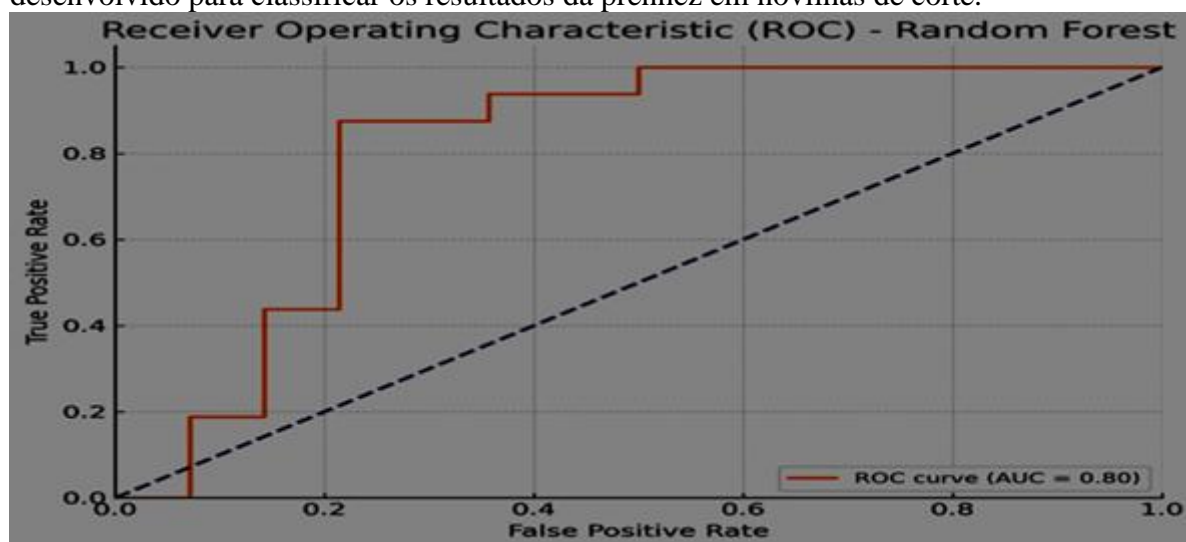
4. ANÁLISE E DISCUSSÃO DO RESULTADOS

Os modelos supervisionados como a Random Forest (RF), a partir da curva ROC (Receiver Operating Characteristic) permite verificar a capacidade preditiva do modelo RF utilizado para

classificar a prenhez em novilhas de corte. Essa curva é construída a partir da taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos ($1 - \text{especificidade}$), ao longo de diferentes limiares de classificação. O modelo apresentou uma área sob a curva (AUC) de 0,8. Esse valor indica que, ao selecionar aleatoriamente uma novilha prenha ou não prenha, o modelo tem 80% de chance de atribuir maior probabilidade de prenhez à novilha que de fato está prenha (Figura 1).

A forma da curva de característica de operação do receptor (ROC), que se distancia nitidamente da linha diagonal de referência (que representa um classificador aleatório), reforça a capacidade do modelo em distinguir entre os dois grupos de novilhas prenhas e não prenhas. A curva representa a sensibilidade do modelo (taxa de verdadeiros positivos) em relação à especificidade (taxa de falsos positivos) em uma série de limites de classificação. A área sob a curva ($\text{AUC} = 0,80$) indica um bom poder discriminatório do classificador de conjunto.

Figura 1 – Curva ROC (curva de característica de operação do receptor) do modelo Random Forest desenvolvido para classificar os resultados da prenhez em novilhas de corte.



Fonte: Dados da pesquisa.

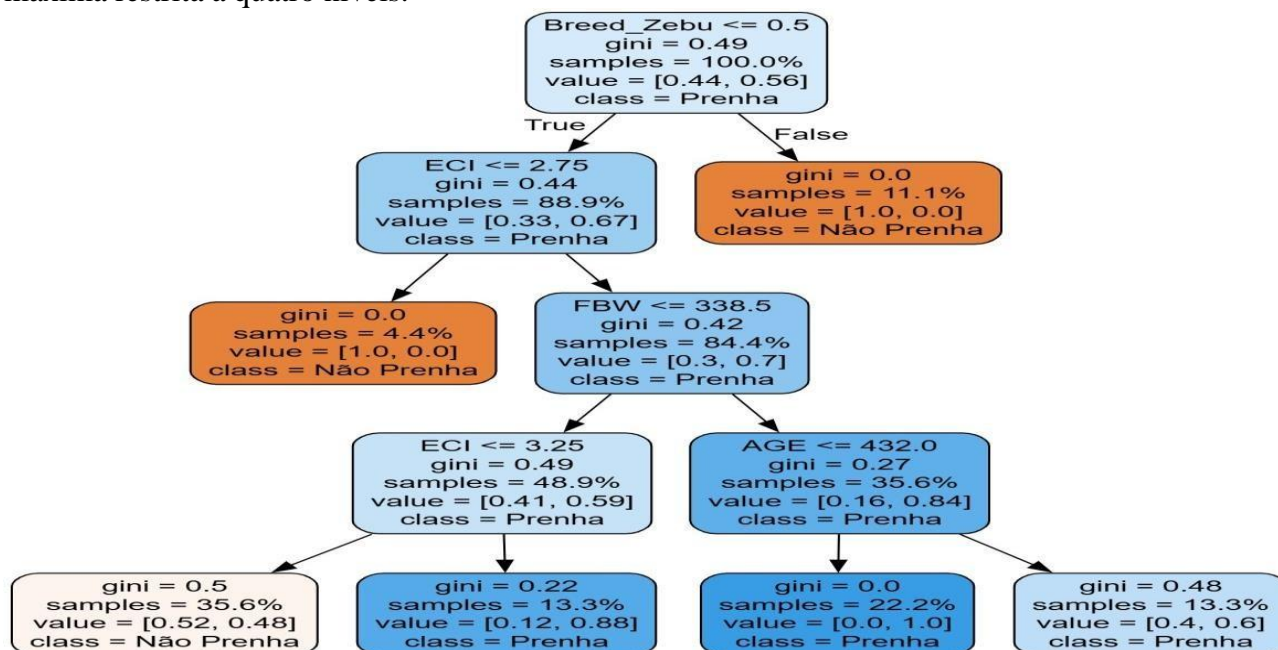
A variável de maior importância na Random Forest (RF) foi o peso corporal inicial (IBW), seguida pelo peso corporal final (FBW) e pelo escore de condição corporal final (FBS) (Tabela 3). O escore de condição corporal final (FBS), embora com menor importância relativa do que seu valor inicial (IBS), também foi identificado como uma variável preditora relevante, o que aponta para o papel cumulativo do manejo nutricional no sucesso reprodutivo. Outro ponto observado é que, logo em seguida das variáveis de cunho nutricional, destacam-se a predominância genética zebuína e a idade. Ressalta-se também que a predominância genética taurina apresentou baixa capacidade discriminatória em comparação com todas as outras variáveis.

Tabela 3 – Importância das variáveis avaliadas na Random Forest.

Variável	Importância (%)
Peso ao início do acasalamento (IBW)	30,1
Peso ao Final do acasalamento (FBW)	19,9
Escore de condição corporal ao início do acasalamento (IBS)	14,4
Escore de condição corporal ao final do acasalamento (FBS)	12,4
Cruzamento com predominância de origem genética zebuína	11,0
Idade ao acasalamento (AGE)	9,6
Cruzamento com predominância de origem genética taurina	2,7

A árvore de decisão representativa do modelo Random Forest, construída com base nas variáveis de desempenho corporal (IBW, FBW, IBS, FBS), idade e predominância de origem genética (Taurina, Zebuína), considerou a origem taurina como referência implícita no modelo (Figura 2). A árvore foi extraída do modelo Random Forest (composto por 1.000 árvores) para fins interpretativos, com profundidade máxima de quatro níveis, sendo selecionada para facilitar a visualização das regras de classificação fenotípica e genética.

Figura 2 – Árvore de decisão representativa para prever resultados de prenhez em novilhas de corte extraída de um modelo Random Forest composto por 1.000 árvores, treinado com uma profundidade máxima restrita a quatro níveis.



Breed_Zebu – Percentual zebuino (0 = taurino, 1 = zebu); **FBW** – Peso ao final do acasalamento (kg); **IBW** – Escore de condição corporal ao início da estação de monta; **AGE** – Idade ao início da estação de monta; **Gini Index** – Índice de Gini (medida de impureza da divisão dos dados); **Samples** – Percentual de amostras ou dados analisados do banco de dados total em cada nó da árvore; **Value** – Percentual de amostras para cada classe (prenha ou Não prenha) em cada nó; **Class** – Classe final do animal.

A composição genética (Breed_Zebu) é a principal variável de segmentação no nó raiz, indicando sua relevância como primeiro critério de divisão determinante da prenhez do grupo de novilhas estudadas. As novilhas são inicialmente separadas com base na sua ascendência zebuina, sendo direcionadas a caminhos distintos.

Novilhas com maior proporção de genes zebuínos (Breed_Zebu > 0,5), correspondendo a 11,1% do total de amostras, são direcionadas diretamente para um caminho que as classifica como "Não Prenha". O resultado conclusivo neste ramo, com um nó puro (gini = 0.0), sugere o forte efeito negativo da predominância da genética zebuina sobre a fertilidade neste conjunto de dados.

Por outro lado, novilhas com ausência ou baixa presença de genes zebuínos (Breed_Zebu <= 0.5), que compõem 88,9% da população estudada, seguem para uma análise mais detalhada, onde são segmentadas pelo escore de condição corporal inicial (ECI). Dentro deste grupo de base genética favorável, novilhas que apresentaram um escore de condição corporal inicial baixo ($\leq 2,75$ pontos) também foram classificadas diretamente como "Não Prenha", indicando que um escore de condição corporal inadequado no início da estação reprodutiva é um fator limitante para a concepção.

Já o grupo com escore de condição corporal inicial superior a 2,75 é novamente subdividido, desta vez com base no peso corporal ao final do acasalamento (FBW), utilizando um ponto de corte

de 338,5 kg. Para as novilhas com peso corporal ao final do acasalamento igual ou inferior a 338,5 kg, o ECI é novamente o critério de determinação, agora com um limiar mais alto de 3,25. Aquelas com escore entre 2,75 e 3,25 (35,6% das amostras) tiveram como resultante majoritária a classificação "Não Prenha", embora com alta incerteza no modelo ($\text{gini} = 0.5$). Em contraste, novilhas com escore de condição corporal inicial superior a 3,25 (13,3% das amostras) foram classificadas como "Prenha".

No subgrupo de novilhas com maior peso ao final do período reprodutivo ($\text{FBW} > 338,5 \text{ kg}$), a idade (AGE) passa a ser o fator decisivo. Animais com até 432 dias foram classificados com 100% de certeza como "Prenha", correspondendo a 22,2% do total de amostras e formando um nó puro. As novilhas mais velhas (> 432 dias) também foram majoritariamente classificadas como "Prenha", porém com menor grau de confiança no modelo.

A origem genética zebuína estar associada a menores índices de prenhez em novilhas acasaladas aos 14–16 meses de idade, também indica que a composição racial exerce papel relevante no desempenho reprodutivo de novilhas de corte. O baixo desempenho verificado nas novilhas Nelore puras, bem como naquelas com maior proporção de genes zebuínos, pode ser atribuído à sua menor adaptação ao ambiente em que o estudo foi conduzido. A estrutura dos sistemas de produção, aliada às características genéticas desses animais, não favorece condições ideais para o adequado desenvolvimento pré-acasalamento, passando essas previamente por um período de ambiente adverso, com baixas temperaturas, as quais essa origem genética não está adaptada. Nas mesmas condições ambientais, a puberdade de novilhas de corte (Restle *et al.*, 1999) e a repetição de prenhez (Vaz *et al.*, 2016) de animais Nelore é inferior aos de animais Charolês devido a adaptação das vacas Charolês às condições ambientais do presente estudo (Restle *et al.*, 1999; 2001).

A AUC (Área Sob a Curva) atingida pelo classificador Random Forest de 0,80, demonstra boa capacidade de segregação entre novilhas gestantes e não-gestantes, tendo como preditores determinantes da prenhez, a partir do percentual da média Gini, os pesos ao início e final do acasalamento com percentuais de 30,1% e 19,9% e os escores de condição corporal inicial e final do período de acasalamento com percentuais de 14,4% e 12,4%. A contribuição da variável genética zebuína (11,0%) foi relevante, enquanto a taurina (2,7%) foi secundária. O Random Forest mostra robustez na análise devido a agregação de múltiplas árvores, que captura interações não lineares e mitiga o sobreajuste, superando com frequência modelos lineares como a regressão logística (Breiman, 2001; Hastie; Tibshirani; Friedman, 2009). A região da curva mais próxima do canto superior esquerdo (Figura 2) indica que o modelo consegue manter alta sensibilidade mesmo com níveis moderados de falsos positivos, o que é desejável em situações em que a prenhez é prioritária para o manejo reprodutivo eficiente nos sistemas de produção. Esses resultados, validam o uso da Random Forest como ferramenta de apoio à tomada de decisão em programas de reprodução,

especialmente na identificação de novilhas com maiores condições de conceber com base em características fenotípicas e genéticas.

Em relação às variáveis preditoras, o Random Forest destacou o peso corporal inicial e final do acasalamento, a participação zebuína, além dos escores de condição corporal ao início e fim do período reprodutivo, como os fatores de maior influência na prenhez. Esses resultados confirmam que novilhas ao atingirem a puberdade e o acasalamento com maior peso vivo e bom escore corporal tendem a apresentar índices superiores de fertilidade (Dickinson *et al.*, 2019; Landarin *et al.*, 2016; Vaz *et al.*, 2012).

Segundo o NRC (2016), a novilha taurina precisa atingir 60% do seu peso adulto para ter condições de ser exposta à reprodução. Para raças de origem europeia o peso das novilhas, se aproxima aos preditos como ponto de corte no Random Forest. Já as raças zebuínas são mais tardias, sendo recomendado em torno de 65% do peso adulto para submeter à reprodução. No presente estudo, esses fatos se comprovaram, pois o índice de prenhez de fêmeas que tinham maior grau de sangue zebuino foi inferior. Além desse fato, cabe salientar, que ambas as origens genéticas estavam em situação de clima subtropical, onde naturalmente, raças de origem europeia tendem a produzir melhor que raças de origem zebuína (Restle *et al.*, 1999; 2001). Gregianini *et al.* (2021) demonstravam a importância do peso mínimo durante o período de acasalamento para novilhas Nelore acasaladas aos 11-14 meses, pois com média de 280,17 kg obtiveram um índice reprodutivo de 42%, tendo 12,43 kg de diferença e superioridade para o lote de novilhas prenhes em relação as não prenhes.

A inclusão de ambos os pesos como preditores relevantes sugere que não apenas o estado ponderal ao início do manejo, mas também a capacidade de ganho de peso e suporte ao desenvolvimento reprodutivo ao longo do período, contribuem para a taxa de prenhez. Contudo, o peso inicial isolado não garante o sucesso reprodutivo de novilhas, caso ocorra perda de peso durante a estação de monta, pois é evidenciado a importância dos pesos e dos escores de condição corporal na PCA (Vaz *et al.*, 2012).

A idade, em muitos dos casos, acaba sendo um dos principais fatores que impactam na maturidade do trato reprodutivo, descrito como preditor do estado puberal da fêmea bovina (Holm *et al.*, 2015). Todavia, é importante ressaltar que novilhas com idade inferior a 368 dias no início do primeiro acasalamento, dificilmente conseguem alcançar o sucesso reprodutivo como descrito por Dickinson *et al.* (2019), fato este, estando de acordo com o presente estudo, onde o modelo realizado apresenta que fêmeas superiores a esta idade, estão classificadas como prenhe.

Desenvolvimentos desde o nascimento fim do primeiro acasalamento quando retroativos, mostram ser determinantes para as novilhas cíclicas, maiores ganhos diários de peso, pesos e escores de condição corporal e maior idade (Da Rocha e Lobato, 2002; Azambuja *et al.*, 2008; Vaz & Lobato,

2010; Vaz *et al.*, 2012; Landarin *et al.*, 2016; Silva *et al.*, 2018).

A capacidade do modelo em distinguir entre animais gestantes e não gestantes, com uma AUC de 0,80, corrobora achados recentes que destacam o potencial do aprendizado de máquina na otimização de programas de manejo reprodutivo em bovinos. Outros estudos obtiveram AUC de 0,75 ao aplicarem algoritmos de Marching Laming para prever o sucesso reprodutivo de inseminações em vacas Holandesas (Shahinfar *et al.*, 2014).

A identificação precoce de fêmeas com menor probabilidade de prenhez permite intervenções de manejo direcionadas, como suplementação nutricional, ou ainda descarte seletivo de animais improdutivos, otimizando o retorno dos investimentos nos sistemas de produção. Isso permite que recursos genéticos mais caros possam ser alocados para aquelas novilhas que possuem uma maior probabilidade de prenhez em comparação com aquelas que não atendem aos requisitos de seleção do rebanho. Ainda, pode auxiliar na amortização do custo de produção da matriz, uma vez que essa, em situações adversas e em sistemas não muito intensivos, necessita produzir seis bezerros para diluir seus custos (Boyer; Griffith; Delong, 2020). Ocorrendo esse acerto na determinação de animais aptos para processos reprodutivos mais precoces, ocorrerá a maximização da produção e o mais rápido retorno sobre o investimento.

No entanto, esses comparativos indicam que o desempenho do nosso modelo Random Forest, embora moderado em termos de discriminação, está alinhado com a literatura atual, podendo em outros estudos, ser aprimorado com a incorporação de informações adicionais como indicadores comportamentais, fisiológicos ou ainda dados de sensores (Marques *et al.*, 2024; Rabaglino *et al.*, 2023).

5. CONSIDERAÇÕES FINAIS

No contexto específico deste conjunto de dados e condições de manejo, os fatores genéticos e nutricionais mostraram-se mais determinantes para a prenhez de novilhas de corte.

O modelo de árvore de decisão previu eficientemente a prenhez de novilhas de corte, sendo uma ferramenta acurada principalmente para tomada de decisão de sistemas de acasalamento.

REFERÊNCIAS

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES; G. J. L.; SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, n. 6, p. 711-728, 2013.

AZAMBUJA, P. S.; PILAU, A.; LOBATO, J. F. P. Suplementação alimentar de novilhas no pós-desmame: efeitos no crescimento e desempenho reprodutivo. **Revista Brasileira de Zootecnia**, v. 37, n. 6, p. 1042–1049, 2008.

BOYER, C. N.; GRIFFITH, A. P.; DELONG, K. L. Reproductive Failure and Long-Term Profitability of Spring-and Fall-Calving Beef Cows. **Journal of Agricultural and Resource Economics**, v. 45, n. 1, p. 78–91, 2020.

BRAND, W.; WELLS, A. T.; SMITH, S. L.; DENHOLM, S. J.; WALL, E.; COFFEY, M. P. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. **Journal of Dairy Science**, v. 104, n. 4, p. 4980–4990, 2021.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

CULLEN, B. R. Eckard, R. J.; Timms, M.; Phelps, D. G. The effect of earlier mating and improving fertility on greenhouse gas emissions intensity of beef production in northern Australian herds. **The Rangeland Journal**, v. 38, n. 3, p. 283–290, 2016.

Da ROCHA, M. G. LOBATO, J. F. P. Avaliação do Desempenho Reprodutivo de Novilhas de Corte Primíparas aos Dois Anos de Idade. **Revista Brasileira de Zootecnia**, v. 31, n. 3, p. 1388–1395, 2002.

DICKINSON, S. E.; ELMORE, M. F.; KRIESE-ANDERSON, L.; ELMORE, J. B.; WALKER, B. N.; DYCE, P. W.; RODNING, S. P.; BIASE, F. H. Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. **Journal of Animal Science and Biotechnology**, v. 10, n. 1, 2019.

ELOY, L. R.; BREMM, C.; LOBATO, J. F. P.; PÖTTER, L.; LACA, E. A. Direct and indirect nutritional factors that determine reproductive performance of heifer and primiparous cows. **Plos One**, v. 17, n. 10, p. e0275426, 2022.

FARIAS, P. P.; FERREIRA, O. G. L.; KRÖNING, A. B.; De AZEVEDO, G. M.; COSTA, P. T.; FLUCK, A. C.; VAZ, R. Z.; RESTLE, J. Consumers' perception of beef from cattle raised in native grasslands in the Pampa Biome. **Observatório de La Economía Latinoamericana**, v. 21, n. 4, p. 1848–1867, 2023.

GREGIANINI, H. A. G.; CARNEIRO JUNIOR, J. M.; PINTO NETO, A.; da COSTA FILHO, L. C. C.; GREGIANINI, J. T. F.; PINHEIRO, A. K.; TRENKEL, C. K.G. Sexual precocity of Nellore heifers in a herd under selection in the State of Acre, Brazil. **Research, Society and Development**, v. 10, n. 4, p. e16310413945, 2021.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. New York, NY, Springer Series in Statistics, 2009.

HOLM, D. E.; NIELEN, M.; JORRITSMA, R.; IRONS, P. C.; THOMPSON, P. N. Evaluation of pre-breeding reproductive tract scoring as a predictor of long term reproductive performance in beef heifers. **Preventive Veterinary Medicine**, v. 118, n. 1, p. 56–63, 2015.

JAURENA, M.; DURANTE, M.; DEVINCENZI, T.; SAVIAN, J. V.; BENDERSKY, D.; MOOJEN, F. G.; PEREIRA, M.; SOCA, P.; QUADROS, F. L. F.; PIZZIO, R.; NABINGER, C.; CARVALHO,

- P. C. F.; LATTANZI, F. A. Native Grasslands at the Core: A New Paradigm of Intensification for the Campos of Southern South America to Increase Economic and Environmental Sustainability. **Frontiers in Sustainable Food Systems**, v. 5, p. 547834, 2021.
- KGARI, R. D.; Muller, C. J. C.; Dzama, K.; Makgahlela, M. L. Evaluation of female fertility in dairy cattle enterprises – A review. **South African Journal of Animal Science**, v. 50, n. 6, p. 819–829, 2020.
- KUPLICH, T. M.; CAPOANE, V.; COSTA, L. F. F. O avanço da soja no bioma pampa. **Boletim Geográfico do Rio Grande do Sul**, v. 10, n. 31, p. 83–100, 2018.
- LANDARIN, C. M.; LOBATO, J. F. P.; TAROUÇO, A. K.; TAROUÇO, J. U.; ELOY, L. R.; PÖTTER, L.; ROSA, A. A. G. Growth and reproductive performance of 14- to 15-month-old Hereford heifers. **Revista Brasileira de Zootecnia**, v. 45, n. 11, p. 667–676, 2016.
- MARQUES, T. C.; MARQUES, L. R.; FERNANDES, P. B.; de LIMA, F. S.; do PRADO PAIM, T.; LEÃO, K. M. Machine Learning to predict pregnancy in dairy cows: An approach integrating automated activity monitoring and on-farm data. **Animals**, v. 14, n. 11, p. 1567, 2024.
- NATIONAL RESEARCH COUNCIL (NRC). **Nutrient Requirements of Beef Cattle**. Washington, DC: National Academy Press, 2016.
- POHLER, K. G. *et al.* New approaches to diagnose and target reproductive failure in cattle. **Animal Reproduction**, v. 17, n. 3, p. e20200057, 2020.
- PRAVIA, M. I.; RAVAGNOLO, O.; URIOSTE, J. I.; GARRIC, D.J. Identification of breeding objectives using a bioeconomic model for a beef cattle production system in Uruguay. **Livestock Science**, v. 160, n. 1, p. 21–28, 2014.
- RABAGLINO, M. B.; SÁNCHEZ, J. M.; MCDONALD, M.; O'CALLAGHAN, E.; LONERGAN, P. Maternal blood transcriptome as a sensor of fetal organ maturation at the end of organogenesis in cattle. **Biology of Reproduction**, v. 109, n. 5, p. 749–758, 2023.
- RASBY, R. J.; STALKER, A.; FUNSTON, R. N. **Body condition scoring beef cows: a tool for managing the nutrition program for beef herds**, University of Nebraska Extension EC281 (revised June 2014), 2014.
- RESTLE, J.; POLLI, V. A.; SENNA, D. B. Efeito de grupo genético e heterose efeito de grupo genético e heterose sobre a idade e peso à puberdade e sobre o desempenho reprodutivo de novilhas de corte. **Pesquisa Agropecuária Brasileira**, p. 701–707, 1999.
- RESTLE, J.; VAZ, R. Z.; ALVES FILHO, D. C.; BERNARDES, R. A. L. C.; PASCOAL, L. L.; SENNA, D. B.; POLLI, V. A. Desempenho de vacas Charolês e Nelore desterneiradas aos três ou sete meses. **Revista Brasileira de Zootecnia**, v. 30, n. 2, p. 499–507, 2001.
- REIS, N. P.; LOBATO, J. F. P.; RESTLE, J.; PACHECO, R. F.; NUÑEZ, A. J. C.; SARTORI, D. B. S.; VAZ, R.Z. Effect of the performance, calving date and lactation period on the Probability of pregnancy in beef cows. **Scientia Agrícola**, v.80, p.e20220088, 2023.
- ROVIRA, J. **Reproducción y manejo de los rodeos de cría**, Montevideo: Hemisferio Sur, 1974.

SCHLEDER, G. R.; FAZZIO, A. Machine Learning na Física, Química, e Ciência de Materiais: Descoberta e Design de Materiais. **Revista Brasileira de Ensino de Física**, v. 43, p. e20200407, 2021.

SHAHINFAR, S.; PAGE, D.; GUENTHER, J.; CABRERA, V.; FRICKE, P.; WEIGEL, K. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. **Journal of Dairy Science**, v. 97, n. 2, p. 731–742, 2014.

SILVA, M. D.; LOBATO, J. F. P.; VAZ, R. Z. Development and reproductive performance of Hereford heifers of different frame sizes up to mating at 14-15 months of age. **Revista Brasileira de Zootecnia**, v. 47, n. 1, p. e20170031, 2018.

VAZ, R. Z.; RESTLE, J.; PACHECO, P. S.; VAZ, F. N.; PASCOAL, L. L.; VAZ, M. B. Ganho de peso pré e pós-desmame no desempenho reprodutivo de novilhas de corte aos quatorze meses de idade. **Ciência Animal Brasileira**, v. 13, n. 3, p. 272–281, 2012.

VAZ, R. Z.; LOBATO, J. F. P. Efeito da idade de desmame no desempenho reprodutivo de novilhas de corte expostas à reprodução aos 13/15 meses de idade. **Revista Brasileira de Zootecnia**, v. 39, n. 1, p. 142–150, 2010.

ZHANG, M.; WANG, X.; FENG, H.; HUANG, Q.; XIAO, X.; ZHANG, X. Wearable Internet of Things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring. **Journal of Cleaner Production**, v. 312, n. 1, 2021.