

UNIVERSIDADE FEDERAL DA FRONTEIRA SUL

CAMPUS REALEZA

**PROGRAMA DE PÓS GRADUAÇÃO EM SAÚDE, BEM-ESTAR E
PRODUÇÃO ANIMAL SUSTENTÁVEL NA FRONTEIRA SUL**

ISAAC DE JESUS DE OLIVEIRA

**APLICAÇÃO DO APRENDIZADO DE MÁQUINA SOBRE A QUALIDADE DA
PRODUÇÃO FAMILIAR DO LEITE**

REALEZA

2024

ISAAC DE JESUS DE OLIVEIRA

**APLICAÇÃO DO APRENDIZADO DE MÁQUINA SOBRE A QUALIDADE DA
PRODUÇÃO FAMILIAR DO LEITE**

Dissertação apresentada ao Programa de Pós-graduação em Saúde, Bem-estar e Produção Animal Sustentável na Fronteira Sul da Universidade Federal da Fronteira Sul, como requisito para obtenção do título de Mestre.

Área:

Medicina Veterinária

Orientadora:

Prof.^a Dr.^a Maiara Garcia Blagitz Azevedo

REALEZA

2024

Bibliotecas da Universidade Federal da Fronteira Sul - UFFS

Oliveira, Isaac de Jesus de
APLICAÇÃO DO APRENDIZADO DE MÁQUINA SOBRE A QUALIDADE
DA PRODUÇÃO FAMILIAR DO LEITE / Isaac de Jesus de
Oliveira. -- 2024.
28 f.

Orientadora: Doutora Maiara Garcia Blagitz Azevedo
Co-orientadora: Doutora Dalila Moter Benvegnu
Dissertação (Mestrado) - Universidade Federal da
Fronteira Sul, Programa de Pós-Graduação em Saúde,
Bem-Estar e Produção Animal Sustentável Na Fronteira
Sul, Realeza, PR, 2024.

1. Produção de leite, Qualidade do leite, Aprendizado
de Máquina, mineração de dados. I. Azevedo, Maiara
Garcia Blagitz, orient. II. Benvegnu, Dalila Moter,
co-orient. III. Universidade Federal da Fronteira Sul.
IV. Título.

Elaborada pelo sistema de Geração Automática de Ficha de Identificação da Obra pela UFFS
com os dados fornecidos pelo(a) autor(a).

ISAAC DE JESUS DE OLIVEIRA

**APLICAÇÃO DO APRENDIZADO DE MÁQUINA SOBRE A QUALIDADE DA
PRODUÇÃO AGROECOLÓGICA DO LEITE**

Dissertação apresentada ao Programa de Pós-graduação em Saúde, Bem-estar e Produção Animal Sustentável na Fronteira Sul da Universidade Federal da Fronteira Sul, como requisito para obtenção do título de Mestre.

Área: Medicina Veterinária

Orientadora: Prof.^a Dr.^a Maiara Garcia Blagitz Azevedo

Este trabalho foi defendido e aprovado pela banca em 20/09/2024.

BANCA EXAMINADORA



Prof.^a Dr.^a Maiara Garcia Blagitz Azevedo – UFFS
Orientadora



Prof. Dr. Edmilson Rodigo Daneze – Unoesc
Avaliador

Documento assinado digitalmente

gov.br

KARINA RAMIREZ STARIKOFF

Data: 05/11/2024 16:10:58-0300

Verifique em <https://validar.iti.gov.br>

Prof.^a Dr.^a Karina Ramirez Starikoff – UFFS
Avaliador

Dedico este trabalho aos meus pais, que não
pouparam esforços para que eu
pudesse concluir meus estudos.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me conservado para que pudesse chegar até este momento com força e determinação.

Aos meus pais e minha irmã que sempre me apoiaram, motivaram e deram condições de poder ingressar na vida acadêmica a qual sempre foi meu sonho.

Aos meus amigos, em especial, colegas do mestrado Ana Letícia Rodrigues Marques, Vitor Mamguê, Rafael Perin e Fernando Gnoatto, por todo trabalho árduo e parceria que estabelecemos nesse processo.

Ao meu colega de apartamento Luiz Gustavo que foi um apoio tanto emocional quanto do cotidiano, conservando e tornando nosso local de moradia um lar real.

E por último, mas não menos importante a Beatriz Costa, que não só me deu apoio incondicional quando o fardo era mais pesado, como seu amor e carinho comigo me deram forças pra seguir em frente.

RESUMO

O Paraná se destaca na produção leiteira, com um crescimento significativo na atividade. Entretanto, há grande heterogeneidade no perfil dos produtores. A região Sudoeste do Paraná é caracterizada pela produção de leite oriunda principalmente da Agricultura Familiar, que é uma das mais produtivas do estado. O objetivo desta pesquisa foi aplicar a técnica de mineração de dados (aprendizado de máquina) através de modelos de árvores de decisão, compostas por regras, para identificar padrões que permitam prever a qualidade do leite pelos indicadores de contagem de células somáticas (CCS) e contagem bacteriana total (CBT) em dados coletados (questionários e análises laboratoriais) de produtores de leite da região Sudoeste do Paraná. Foram criados modelos para as classes finais CCS e CBT, resultando em duas árvores de decisão. Utilizou-se uma base de dados para treinamento (70%) e uma base de dados para testes (30%), totalizando 221 registros. Na categorização dos dados, a classificação do leite para CCS foi: A (até 400.000 células/mL), B (até 500.000 células/mL) ou C (acima de 500.000 células/mL), e para CBT: A (até 100.000 UFC/mL), B (até 300.000 UFC/mL) e C (acima de 300.000 UFC/mL). Após essas etapas, realizou-se a mineração de dados e extração do conhecimento. Em relação aos resultados para as árvores de decisão, os atributos relevantes para a geração do modelo de CCS foram teor de lactose e tipo de ordenha. Para o modelo de CBT, os atributos principais foram o número de animais e o uso de água sanitária na higienização do resfriador. Para ambos os modelos (CCS e CBT), a produção diária de leite em litros e a origem da água utilizada nos procedimentos de ordenha foram fatores importantes. A eficiência dos modelos gerados foi de 63,64% para CCS e 74,24% para CBT.

Palavras chave: Agricultura familiar. Bovinocultura leiteira. Data mining.

ABSTRACT

Paraná stands out in dairy production with significant growth in this activity. However, there is great heterogeneity in the profile of the producers. The Southwest region of Paraná is characterized by milk production mainly from family farming and is one of the most productive in the state. The objective of this research was to apply the data mining technique (machine learning) through decision tree models, composed of rules, to identify patterns to predict milk quality using the indicators of somatic cell count (SCC) and total bacterial count (TBC) in data collected (questionnaires and laboratory analyses) from milk producers in the Southwest region of Paraná. Models were created for the final classes SCC and TBC, resulting in two decision trees. A database was used for training (70%) and a database for testing (30%), totaling 221 records. In data categorization, the classification of milk for SCC was: A (up to 400,000 cells/mL), B (up to 500,000 cells/mL), or C (above 500,000 cells/mL), and for TBC: A (up to 100,000 CFU/mL), B (up to 300,000 CFU/mL), and C (above 300,000 CFU/mL). After these steps, data mining and knowledge extraction were performed. Regarding the decision tree results, the relevant attributes for generating the SCC model were lactose content and type of milking. For the TBC model, the main attributes were the number of animals and the use of bleach in cooling tank sanitation. For both models (SCC and TBC), the daily milk production in liters and the origin of the water used in the milking procedures were important factors. The efficiency of the generated models was 63.64% for SCC and 74.24% for TBC.

Key words: Family farming. Dairy cattle farming. Data mining.

SUMÁRIO

1. INTRODUÇÃO	10
2. MATERIAL E MÉTODOS	13
3. RESULTADOS E DISCUSSÃO	16
3.1 Modelo de árvore de decisão para contagem células somáticas (CCS).....	16
3.2 Modelo de árvore de decisão para contagem bacteriana total (CBT).....	19
4. CONCLUSÃO	23
REFERÊNCIAS	24

1. INTRODUÇÃO

O setor dos laticínios é um ramo de destaque mundial, e como todo segmento de importância econômica, deve superar desafios como controle de doenças, bem-estar animal e ineficiências da cadeia produtiva (Neethirajan; Kemp, 2021). Com o avanço da tecnologia e novas ferramentas como a Inteligência Artificial (IA) incorporada aos sensores, os paradigmas da indústria leiteira têm sido modificados (Bari *et al.*, 2022).

O Brasil ocupa uma posição de destaque no cenário global da produção de leite, sendo o terceiro maior produtor mundial, com um volume anual superior a 34 bilhões de litros. Em 2023, a produção de leite no país apresentou crescimento em relação ao ano anterior, impulsionada principalmente pelos altos preços do produto e pela modernização de processos. A maior parte dessa produção vem de pequenas e médias propriedades, presente em 98% dos municípios brasileiros, e o setor emprega cerca de 4 milhões de pessoas. Mato Grosso, por exemplo, registrou um aumento de 2,86% em sua captação de leite, refletindo uma tendência positiva após uma fase de queda no setor. Esse desenvolvimento demonstra o impacto das políticas públicas e do investimento em tecnologia para sustentar o setor lácteo, que enfrenta desafios, mas tem grande potencial de crescimento (Embrapa, 2023; Ministério da Agricultura e Pecuária, 2024; Agrolink, 2024)

O Paraná se destaca na produção de leite, desde de 2015, quando o Estado passou a se tornar o maior de leite produtor da região sul do Brasil. O Paraná é constituído por 39 microrregiões, das quais, 10 são consideradas especializadas na produção de leite. Além de ser o maior produtor da região sul, o Estado é considerado também com a maior heterogeneidade no perfil produtivo. Essa característica se deve ao nível de tecnologia empregado na atividade pelos produtores (Telles *et al.*, 2018).

A Região Sudoeste do Paraná é considerada a maior produtora do Estado em volume de leite. Porém os municípios com a maior produção no Estado, Castro e Carambeí, são pertencentes a região Centro oriental. Contudo a região ainda, na maioria das propriedades, é carente de tecnificação e a atividade é pautada principalmente na Agricultura Familiar, apesar de todos os avanços da região (Pereira; Nunes, 2021).

Para se aferir a qualidade do leite produzido, ele deve estar em conformidade com alguns parâmetros bem estabelecidos pela legislação. Dentre os parâmetros de qualidade do leite, a contagem de células somáticas (CCS) e a contagem padrão em

placas (CPP) (outrora conhecida como contagem bacteriana total, CBT) são preconizadas para avaliar a celularidade e conseqüentemente as condições físico-químicas desse leite (MAPA, 2018).

O leite de vaca é extremamente reconhecido pelo seu perfil nutricional balanceado, sendo uma fonte rica de proteínas, lipídios, vitaminas e minerais essenciais. Em média, 100 mL de leite integral fornece cerca de 64 kcal, 3,3 g de proteína e 3,6 g de gordura, incluindo uma quantidade significativa de ácidos graxos saturados (cerca de 2,4 g) e 14 mg de colesterol. Ele é especialmente valorizado pela presença de cálcio (120 mg por 100 mL), fundamental para a saúde óssea, e fornece outros nutrientes como fósforo e potássio, essenciais para funções celulares e musculares. Além dos macronutrientes, o leite é rico em vitaminas A, D e B12, que desempenham papéis importantes na saúde dos ossos, sistema e produção imunológica de energia (Cidesp, 2023; Embrapa, 2023; Nutritotal, 2023).

Devido ao seu alto teor de proteínas, composto majoritariamente por caseína (80%) e proteínas do soro (20%), o leite é particularmente benéfico para a saúde muscular, auxiliando na recuperação e no crescimento dos músculos, especialmente em pessoas fisicamente ativas. Já o teor de gordura do leite varia conforme a raça do animal e a alimentação, afetando também seu conteúdo de ácidos graxos e características sensoriais, como sabor e textura (Embrapa, 2023; Nutritotal, 2023).

No Brasil os valores máximos estabelecidos pelas INs 76 e 77 para CCS e CBT são, de 500.000 CCS/mL e 300.000UFC/mL para a CBT, a contar do dia 1 de julho de 2014 até 30 de junho de 2016, prazo limite alterado pelo IN N° 7, de 3 de maio de 2016 e posteriormente pela IN N° 31, de 29 de junho de 2018, até 30 de junho de 2019 (Brasil, 2011).

Altas taxas de celularidade no leite podem indicar falha no controle da mastite nas propriedades, que trata-se de uma afecção multifatorial que irá alterar as propriedades do leite devido ao alto número de células inflamatórias para o local, visando combater infecções por agentes patogênicos ou mesmo por reações de hipersensibilidade (Botton *et al.*, 2019, Costa *et al.*, 2019).

Infelizmente, muitos dos produtores não sabem interpretar os dados de CCS e CBT fornecidos pelos laticínios, assim, não têm consciência sobre os impactos gerados por esses parâmetros tanto na qualidade quanto no retorno econômico da atividade leiteira, visto que muitas empresas de lácteos pagam pela qualidade do leite (Lordão *et al.*, 2013).

Para seleção e processamento eficiente de informações, técnicas para mineração de dados podem ser aplicadas. A técnica de árvores de decisão, podem ser utilizadas para a interpretação de dados colhidos em diversas áreas, como é o caso da área da saúde ou mesmo de áreas produtivas. Os modelos criados, especificamente as árvores de decisão, podem ser utilizados para desenvolver alguma ferramenta computacional que auxilie no diagnóstico, prevenção e classificação de doenças (Freddo *et al.*, 2019).

Existem relatos sobre o uso de mineração de dados na pecuária leiteira para detectar o ciclo estral de vacas leiteiras, utilizando a técnica de árvore de decisão (NÄÄS *et al.*, 2008). Além disso, Perissinoto e De Moura (2007) analisaram o estresse térmico em vacas holandesas confinadas, utilizando o programa Weka para a classificação dos dados e o algoritmo J48 para validação cruzada e geração de árvores de decisão. Steensels *et al.* (2016) desenvolveram um modelo de detecção de saúde com base em comportamento e desempenho de vacas pós-parto, em uma fazenda com sistema de ordenha robotizada, utilizando árvores de decisão. O objetivo era identificar vacas doentes com base em sensores comerciais disponíveis, alcançando uma precisão de 78%.

Bakir *et al.* (2010) examinaram os efeitos do período seco, paridade da lactação, fazenda, época de parto e idade de vacas holandesas na produção de leite em 305 dias, usando o método de árvores de regressão. Essa mesma técnica foi aplicada para avaliar os impactos da gestão e do ambiente na produção de leite durante a primeira lactação (Mikail, Bakir, 2019). Pinzón-Sánchez, Cabrera e Ruegg (2011) usaram árvores de decisão para analisar o impacto econômico de diferentes durações de tratamento intramamário para o primeiro caso de mastite clínica leve ou moderada que ocorre no início da lactação. Além disso, Dorshorst, Collins e Lombard (2006) criaram um modelo de árvore de decisão para o controle da paratuberculose em rebanhos leiteiros comerciais.

Cak, Keskin e Yilmaz (2013) analisaram os fatores que influenciam a produção de leite em cada fase de lactação de vacas da raça *Brown Swiss*, utilizando o método de árvores de decisão. Brum *et al.* (2017) também aplicaram essa técnica para avaliar o potencial produtivo de propriedades leiteiras no Noroeste do Rio Grande do Sul. Sitkowska *et al.* (2017) e Piwczyński *et al.* (2020) usaram árvores de decisão para prever o aumento da contagem de células somáticas (CCS) no leite de vacas em sistemas de ordenha robotizada.

Utilizando sistema de árvore de decisões na pecuária leiteira Aerts *et al.* (2022), indicaram a melhor combinação de fatores e seus níveis que contribuíram para demonstrar maior eficiência na ordenha em rebanhos ordenhados em mecanicamente e com auxílio de robôs. Outro estudo também utilizou a mineração de dados e o sistema de decisão para traçar relação entre fatores da qualidade do leite como CCS, CBT e componentes como gordura, proteína e lactose (Botton *et al.*, 2019).

As árvores de decisão foram utilizadas para avaliar e prever o peso ao nascer de cordeiros (Eyduvan *et al.*, 2008) e bovinos (Topal *et al.*, 2010; Yilmaz; Eyduvan; Kaygisiz, 2013). Também foram usadas para identificar as relações entre o peso corporal e as características morfométricas de ovinos (Mohammad *et al.*, 2012; Yakubu, 2012; Tariq *et al.*, 2012; Khan *et al.*, 2014) e patos (Oguntunji, 2017). Celik *et al.* (2017) compararam diversos algoritmos de mineração de dados para prever o peso corporal de cães com base em várias medidas corporais e morfométricas.

Além disso, redes neurais, outra técnica de mineração de dados, foram usadas para avaliar a produção de leite de vacas holandesas (Takma *et al.*, 2012). Karadas *et al.* (2017) previram a produção de leite e a lactação em ovelhas, enquanto Ali *et al.* (2015) usaram técnicas de mineração de dados para prever o peso corporal de ovinos de 6 a 9 meses com base em características biométricas. Esses autores compararam o desempenho de métodos como CART, modelos lineares gerais, qui-quadrado e redes neurais artificiais.

Diante do exposto, o objetivo deste estudo consistiu em empregar técnicas de aprendizado de máquina, mais especificamente, árvores de decisão para prever a qualidade do leite no que se refere à CCS e CBT de leites oriundos de sistemas de bovinocultura familiar.

2. MATERIAL E MÉTODOS

A pesquisa foi realizada após aprovação pelo Comitê de Ética em Pesquisa - CEP da Universidade Federal da Fronteira Sul - UFFS (sob número 88506018.7.0000.5564), e o período de aplicação dos questionários aos produtores de leite foi de julho de 2018 a junho 2019.

O método utilizado para extrair conhecimento a partir dos dados (Knowledge Discovery in Databases - KDD) possui as seguintes etapas: (1) seleção, (2) pré-processamento, (3) transformação ou preparação dos dados, (4) mineração de dados e

(5) extração e avaliação do conhecimento (Mitchell, 1997). Na sequência, as etapas foram descritas detalhadamente.

Etapa 1: Seleção de dados

Os dados utilizados foram obtidos de questionários aplicados aos produtores de leite da região Sudoeste do Paraná e dos resultados das análises laboratoriais do leite da mesma propriedade fornecidos por um laticínio responsável. As propriedades foram selecionadas de acordo com a rota que o automóvel do laticínio executava para buscar o leite, além do consentimento dos produtores.

Foram selecionadas 27 questões do questionário formulado por Da Costa; Rajala-Schultz; Schunemann (2016) sobre características das propriedades, como: número e raça dos animais, tipo de alimentação fornecida aos animais, origem da água, manejo e limpeza nos procedimentos de ordenha, instalações e tipo de resfriador, bem como, práticas de biossegurança aplicadas na propriedade.

Com relação às análises do leite foram utilizados os parâmetros de contagem de células somáticas (CCS), contagem bacteriana total (CBT), proteína total, sólidos totais, lactose e gordura, de um período de quatro meses: dezembro de 2018 a março de 2019, utilizando-se a média desse período.

Etapa 2: Pré-processamento dos dados

Nesta etapa foi realizada a identificação de dados duplicados, incorretos, faltantes e atípicos. Dos 235 questionários respondidos, 14 foram retirados da base de dados por apresentarem dados inconsistentes como: questões não preenchidas, dados faltantes. A base de dados foi composta, portanto, por 221 questionários.

Etapa 3: Transformação e Preparação dos dados

A preparação dos dados consistiu na categorização de valores numéricos em intervalos de valores para a criação de uma classificação.

Para CCS o leite foi classificado em A (até 400.000 células/mL), B (até 500.000 células/mL) ou C (acima de 500.000 células/mL), para CBT em: A (até 10.000 UFC/mL), B (até 300.000 UFC/mL) e C (acima de 300.000 UFC/ mL), e em função da composição nutricional (lactose, proteínas totais, gordura e sólidos totais) em ideal ou não ideal (parâmetros estabelecidos pela legislação vigente: Instrução Normativa 76/2018) (BRASIL, 2018).

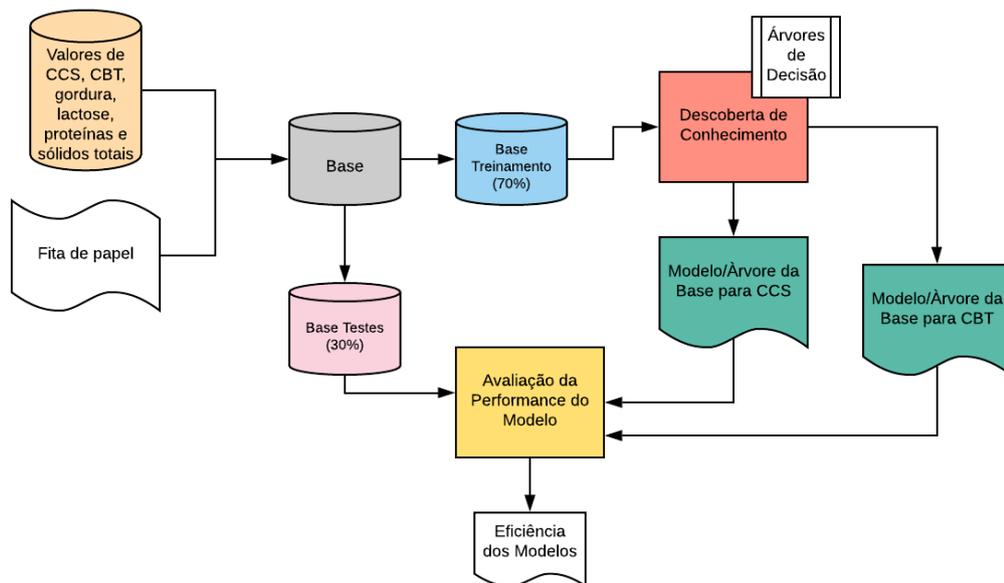
Após a categorização, os dados foram transformados para o formato CSV (Comma Separated Values), que é requerido pela interface funcional do programa utilizado para a mineração dos dados no ambiente R (ferramenta R-Studio).

Etapa 4: Mineração de Dados

Esta etapa compreendeu a aplicação das técnicas de mineração de dados denominada de árvore de decisão. Assim, foram criados os modelos para as classes finais CCS e CBT, sendo duas árvores de decisão.

Para que se possa aplicar o aprendizado de máquina, se faz necessária a existência de uma quantidade de dados de treinamento e teste. Portanto, optou-se por utilizar 70% das bases de dados para treinamento (155 registros para CCS e 152 registros para CBT) e 30% para testes (66 registros para CCS e 69 registros para CBT). Esta divisão foi gerada aleatoriamente pelas ferramentas computacionais na base de dados.

Figura 1. Construção dos modelos da árvore de decisão.



A fita de papel corresponde ao questionário aplicado aos produtores de leite. E os valores são referentes às médias de quatro meses (dezembro de 2018 a março de 2019) de contagem de células somáticas (CCS), contagem bacteriana total (CBT), gordura, lactose, proteínas e sólidos totais, dados esses que foram fornecidos pelo laticínio em que o produtor estava cadastrado. Fonte: Elaborado pelos autores.

Etapa 5: Extração do conhecimento

Ao aplicar as técnicas de mineração de dados, os resultados obtidos são os modelos (árvores de decisão) ou padrões que são interpretados nesta etapa para a geração ou extração de conhecimento. Portanto, buscou-se nesta etapa interpretar os

modelos para que se tenha conclusões acerca do problema da qualidade do leite, a fim de verificar os fatores influentes sobre os valores de CCS e CBT. Verifica-se a possibilidade de obter a qualidade do leite, ou seja, obter a categorização de CCS e CBT a partir das respostas dos questionários e com base nos valores de proteínas totais, gordura, lactose e sólidos totais.

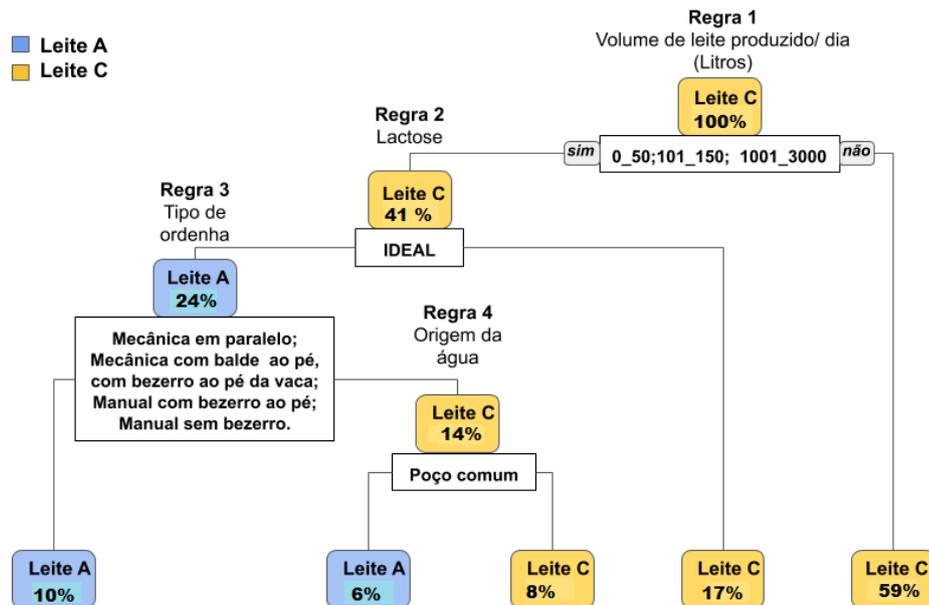
3. RESULTADOS E DISCUSSÃO

3.1 Modelo de árvore de decisão para contagem células somáticas (CCS)

Para o atributo CCS, de um total de 66 propriedades, 10 (15,15%) tiveram seu leite como classificado A, enquanto 56 (84,85%) foram classificados como C. Assim, 84,85% das propriedades apresentaram contagens superiores a 500.000 CS/mL, o que está em desacordo com as normas vigentes e indica baixa qualidade do leite cru produzido. Ribas *et al.* (2014) relataram que, no Paraná, das 1.950.034 amostras de leite avaliadas, 65,95% apresentaram altos escores de CCS, fato que indica alto nível de mastite em rebanhos paranaenses, possivelmente com número significativo de quartos mamários infectados e consideráveis perdas produtivas. Portanto, para estar em conformidade com a legislação, é crucial que práticas diárias de prevenção e controle da mastite sejam aplicadas no processo de ordem, conforme recomendado por Lordão *et al.* (2013).

A Figura 2 apresenta a árvore de decisão gerada para a classe final CCS, das quais as classes finais são descritas em A (Leite A) e C (Leite C).

Figura 2. Árvore de decisão gerada para contagem de células somáticas (CCS) em amostras de leite.



Legenda: Leite A: $CCS \leq 400.000$ células/mL; Leite C: $CCS > 500.000$ células/mL. Fonte: elaborado pelos autores.

Considerando a falta de resultados da classe B (Leite B), não houve a criação de regras na árvore para esta classe final.

Os atributos considerados mais relevantes para gerar a árvore da decisão para CCS foram: produção de leite diária em litros, tipo de ordenha, teor de lactose, e origem da água utilizada para os procedimentos de ordenha.

Neste trabalho não houve influência do volume de produção de leite diário em relação à CCS, pois, a classe final do leite predominante foi C para Regra 1, ou seja, as contagens apresentaram-se acima de 500.000 células somáticas por mL.

São diversos os fatores que influenciam a qualidade do leite cru, entre os quais destacam-se os índices zootécnicos, associados ao manejo, saúde da glândula mamária, potencial genético, suplementação e dieta fornecidas aos animais (Lopez *et al.*, 2022). Além desses, a elevação da CCS está diretamente associada à redução na produção de leite. Esta ocorre devido às alterações de permeabilidade vascular dos alvéolos secretores e lesões nas células epiteliais da glândula mamária (Fonseca; Santos, 2000).

Miller e Nesi (2012), no município de Ipuçu, localizado no Oeste de Santa Catarina, definiram três níveis de produção leiteira: até 100 litros por dia (pequena), entre 101 e 300 litros por dia (média) e acima de 301 litros por dia (grande). A partir dessa categorização, observaram que, com o aumento da produção de leite, os valores de CCS diminuíram. Além disso, Dong, Hennessy e Jensen (2012) sugerem que fazendas maiores (com um número maior de animais e, consequentemente, maior

produção de leite) possuem melhores condições para alcançar um limite mais rigoroso de CCS devido aos seus recursos de atendimento.

Na Regra 2, valores não ideais de lactose preveem a qualidade do leite tipo C em relação à CCS em 17% dos registros. Por outro lado, teores ideais de lactose indicaram que, em 24% dos registros, a classificação do leite seria A para CCS. Assim, altos níveis de CCS apontam para alterações nos constituintes do leite, especialmente na lactose, que geralmente apresenta uma concentração reduzida no leite contaminado. Isso ocorre devido à alteração da permeabilidade vascular e à diminuição da capacidade de síntese das células secretoras, resultantes de processos inflamatórios ou infecciosos que danificam a glândula mamária. Além disso, a redução na síntese de lactose contribui para a diminuição do volume de leite produzido, uma vez que esse carboidrato desempenha um papel central na regulação osmótica do leite (Ribas *et al.*, 2014).

Na regra 3, se o tipo de ordenha utilizado fosse mecânica em paralelo, mecânica com balde ao pé, com bezerro ao pé da vaca, ordenha manual com bezerro ao pé ou ordenha manual sem bezerro ao pé, então o leite seria classificado como tipo A para CCS (em 10% dos registros). Caso contrário, se o tipo de ordenha fosse mecânica em espinha de peixe, mecânica em carrossel, mecânica voluntária, robotizada ou mecânica com balde ao pé sem bezerro ao pé da vaca, a Regra 4 seria aplicada, e o leite seria classificado como tipo C para CCS (em 14% dos registros).

Neste trabalho, 10% das propriedades foram classificadas como A (CCS inferior ou igual a 400.000 células/ mL). Não é necessariamente o tipo de ordenha o causador de quadros de mastite, mas sim as falhas operacionais relacionadas às práticas de manejo e higiene que estão associadas à capacitação do ordenhador, e que influenciam diretamente na qualidade do leite cru produzido no que se refere à CCS e CBT (Bari *et al.*, 2022).

Barbosa, Benedetti e Guimarães (2009) avaliaram o índice de CCS em relação ao tipo de ordenha empregado e observaram que o sistema mecânico apresentou um valor elevado (1.167.308 células/mL), seguido pela ordenha mecânica de circuito fechado (886.396 células/mL) e, com menor CCS, a ordenha manual (352.670 células/mL). Além disso, não constataram influência do bezerro ao pé na ocorrência de mastite em nenhum tipo de ordenha.

Esses dados estão em concordância com Silva *et al.* (2018), que verificaram que os valores médios de CCS do leite ordenhado mecanicamente foram superiores aos do leite obtido por ordenha manual. O uso inadequado e a limpeza deficiente dos

equipamentos de ordenha, assim como a falta de antissepsia dos tetos antes e após a ordenha, contribuem para a ocorrência de mastite, resultando em uma CCS mais alta no leite proveniente de sistemas mecânicos de ordenha.

Causas que levem à contaminação da água podem favorecer o desenvolvimento bacteriano, que por consequência, podem carrear contaminação do leite por microrganismos patogênicos (Silva *et al.*, 2018), bem como esses também podem desencadear quadros de mastite, prejudicando o desempenho produtivo desses animais, e caracteriza-se por altos valores de CCS (Amaral *et al.*, 2004; Stepaniak *et al.*, 2018).

Quanto à origem da água utilizada nos procedimentos de ordenha (Regra 4), se a água fosse proveniente de poço comum, a classe do leite seria A (em 6% dos registros). Por outro lado, se a água fosse de outras fontes, como rede de abastecimento público (SANEPAR), poço artesiano, curso d'água, açude ou fonte, o leite seria classificado como tipo C (em 8% dos questionários).

Brum *et al.* (2017) também aplicaram técnicas de mineração de dados com o intuito de avaliar indicadores socioeconômicos, ambientais e produtivos de propriedades leiteiras do Noroeste do Rio Grande do Sul, utilizando o algoritmo RPART e obtiveram uma acurácia de 63,33%, valor semelhante ao encontrado para o modelo gerado para CCS, o qual foi de 63,64%.

3.2 Modelo de árvore de decisão para contagem bacteriana total (CBT)

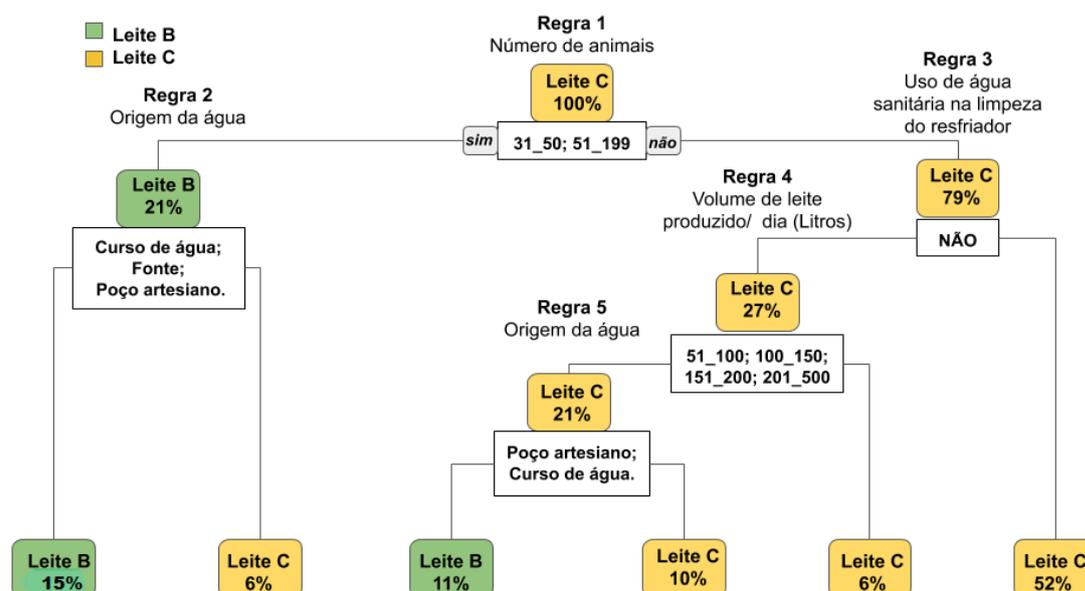
Em relação à CBT, nenhuma das propriedades foi classificada com leite A, indicando que as contagens estavam acima de 10.000 UFC/mL. Assim, as classes finais apresentadas na árvore de decisão para CBT foram: B em 19 propriedades (27,54%), com contagens menores ou iguais a 300.000 UFC/mL, e C em 50 propriedades (72,46%), com valores superiores a 300.000 UFC/mL.

Valores muito altos de CBT geralmente resultam de procedimentos inadequados de manejo, especialmente nas práticas higiênico-sanitárias durante e após a ordenha (Hoogerheide; Mattioda, 2012). Baron *et al.* (2016) relataram que a maioria dos produtores de leite da Agricultura Familiar no município de Realeza, no Sudoeste do Paraná, possuem conhecimentos insuficientes sobre manejo do rebanho e práticas de higiene na ordenha para a produção de leite de qualidade, o que faz com que as especificações estabelecidas pela legislação vigente não sejam atendidas.

Entre os atributos da base de dados, os mais importantes para gerar essa árvore de decisão, conforme mostrado na Figura 3, foram: número de animais, origem da água

utilizada nos procedimentos de ordenha na propriedade, produção diária de leite em litros e o uso de água sanitária na higienização do resfriador.

Figura 3. Árvore de decisão gerada para contagem bacteriana total (CBT) em amostras de leite.



A partir da extração de conhecimento das árvores de decisão, foram formadas as seguintes regras:

Regra 1: Caso o rebanho fosse composto por 31 a 50 animais ou 51 a 199 animais, o leite seria do tipo B para CBT (21% dos registros) e a regra 2 seria executada. Caso contrário, se o número de animais não estivesse dentro desses intervalos (79% dos registros), o leite seria classificado como C para CBT e seguir-se-ia para a regra 3. Propriedades com maior produção consideram a pecuária leiteira uma atividade rentável, possuem melhor infraestrutura e são mais bem informadas, investindo em tecnologia. Por esses motivos, obtêm uma qualidade microbiológica do leite superior (Belli *et al.*, 2017). Taffarel *et al.* (2015) observaram menor CBT em propriedades com maior produção (acima de 15.000 litros mensais) e número de animais, dados que corroboram com os deste estudo, onde produtores com rebanho acima de 31 animais apresentaram CBT conforme a legislação vigente (contagens menores ou iguais a 300.000 UFC/mL).

Regra 2: Se a origem da água usada na propriedade fosse de curso d'água de origem natural (açude, rio ou fonte) ou poço artesiano, o leite seria do tipo B para CBT (15% dos produtores). Se a água utilizada não fosse dessas fontes, mas de poço comum ou rede de abastecimento público (SANEPAR), o leite seria classificado como C para CBT (6% dos produtores). A origem da água utilizada nos procedimentos de ordenha também foi um atributo utilizado para a geração da Regra 5 na árvore de decisão para CBT. Portanto, se a água era oriunda de poço artesiano ou de curso d'água, a classe para CBT do leite seria B (11% dos produtores). Em contrapartida, se a água fosse proveniente de outra origem, o leite seria do tipo C para CBT (10% dos registros).

A qualidade microbiológica do leite é influenciada por múltiplos fatores. Apesar da origem da água influenciar a qualidade do leite, a adoção de boas práticas de ordenha, como o uso de *pré-dipping* e de detergentes específicos na limpeza dos equipamentos, tende a reduzir a contaminação do leite, controlando os valores de CBT (Silva *et al.*, 2018). Procedimentos de higiene antes, durante e após a ordenha, bem como a sanitização dos equipamentos, têm um impacto mais significativo na CBT em comparação com a CCS; contudo, eles podem levar a um novo quadro de infecção intramamária e, conseqüentemente, ao aumento da CCS quando esses requisitos não são totalmente atendidos (Costa *et al.*, 2019).

Amaral *et al.* (2004) avaliaram a qualidade da água de propriedades leiteiras e verificaram que 90% das amostras das fontes de abastecimento estavam fora dos padrões microbiológicos de potabilidade. Águas não potáveis oferecem alto risco à qualidade do leite, à saúde da glândula mamária e também à saúde pública.

Regra 3: Observa-se que o uso de água sanitária para higienização do resfriador não influenciou na qualidade do leite cru produzido, pois a classificação foi predominantemente do tipo C para CBT. Produtores de leite tendem a usar saneantes para a limpeza de equipamentos, mas não consideram concentrações e diluições adequadas desses produtos nem avaliam as propriedades químicas da água, o que interfere negativamente na higienização e na qualidade do leite armazenado (Cerqueira *et al.*, 2006; Silva *et al.*, 2018). Dessa maneira, pode haver um aumento significativo da CBT quando o leite entra em contato com equipamentos que receberam uma limpeza e sanitização inadequadas, pois os microrganismos proliferam em locais onde ocorrem acúmulos de resíduos de leite, como recipientes, junções e borrachas (Guerreiro *et al.*, 2005).

Assim como para CCS, a produção diária de leite (Regra 4) não influenciou diretamente na qualidade final do leite para CBT, pois, independentemente do volume de leite produzido, a classe final majoritária foi C (contagem acima de 300.000 UFC/mL). Resultados equivalentes foram apresentados por Ramires, Berger e Almeida (2009), onde o volume de produção não foi significativo em relação à CBT. Portanto, o volume de produção de leite não influencia diretamente a qualidade microbiológica da matéria-prima; os fatores determinantes são o manejo dos animais e as práticas de ordenha aplicadas.

Em estudo realizado no Oeste do Paraná, a CBT foi significativamente superior no leite oriundo de propriedades com menor produção de leite, aquelas com o volume de até 4.500 litros mensais (Taffarel *et al.*, 2015). Propriedades com menor nível tecnológico e produtivo tendem a apresentar CBT superiores, pois uma baixa produção resulta em menor receita, deixando o investimento na atividade em segundo plano. Assim, práticas de rotina e manejo dos animais são realizadas com menores gastos possíveis (Belli *et al.*, 2017).

A acurácia para esse modelo foi de 74,24%, superior ao modelo gerado para CCS. Ou seja, o modelo de árvore de decisão para CBT classificou corretamente 74,24% dos dados na fase de teste.

4. CONCLUSÃO

Para a árvore de decisão no caso da CCS, as classes finais de qualidade do leite foram A e C, enquanto que no modelo para CBT, as classes finais foram B e C. Esses resultados revelam a dificuldade dos produtores em cumprir com os requisitos das normas vigentes, resultando em leite de baixa qualidade tanto para CCS quanto para aspectos microbiológicos. Isso indica falhas no processo produtivo, principalmente em relação às boas práticas de ordem e à higienização dos equipamentos, que não estão sendo executadas de maneira adequada.

Os atributos que influenciaram a geração da árvore para CCS incluíram: a produção diária de leite em litros, o tipo de ordem, o teor de lactose e a origem da água usada nos procedimentos de ordem. Para o CBT, os atributos mais relevantes foram: o número de animais, a origem da água utilizada na ordem, a produção diária de leite em litros e o uso de água sanitária para a higienização do resfriador.

A árvore de decisão para CBT mostrou-se mais eficiente na classificação dos registros, resultando em um índice de acurácia superior ao do modelo para CCS. Essa diferença nos índices de eficiência pode ser atribuída ao pequeno conjunto de dados de treinamento ou à inadequação dos atributos na criação de um modelo preciso e assertivo.

Portanto, para melhoramento das análises por parte da inteligência artificial, deve-se aumentar o número de dados (questionários), e avaliar a combinação e melhor seleção dos atributos em relação às variáveis analisadas. Ainda, sugere-se a aplicação de outros algoritmos para geração das árvores de decisão, bem como, considerar a aplicação de outras técnicas de aprendizado de máquina.

5. REFERÊNCIAS

AERTS, J. et al. Forecasting milking efficiency of dairy cows milked in an automatic milking system using the decision tree technique. **Animals: An Open Access Journal from MDPI**, v. 12, n. 8, 2022.

AGROLINK. Produção de leite registra crescimento em 2023. Disponível em: https://www.agrolink.com.br/noticias/producao-de-leite-registra-crescimento-em-2023_473108.

ALI, M. et al. Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. **Pakistan Journal of Zoology**, v. 47, n. 6, p. 1579-1585, 2015.

AMARAL, L. A. do et al. Qualidade da água em propriedades leiteiras como fator de risco à qualidade do leite e à saúde da glândula mamária. **Arquivo do Instituto Biológico**, v. 71, n. 4, p. 417-421, 2004.

BAKIR, G. et al. Determination of the effective factors for 305 days milk yield by regression tree (RT) method. **Journal of Animal and Veterinary Advances**, v. 9, n. 1, p. 55-59, 2010.

BARBOSA, C. P.; BENEDETTI, E.; GUIMARÃES, E. C. Incidência de mastite em vacas submetidas a diferentes tipos de ordenha em fazendas leiteiras na região do Triângulo Mineiro. **Bioscience Journal**, v. 25, n. 6, 16 dez. 2009.

BARON, C. P. et al. Caracterização das condições de higiene de ordenha na produção leiteira da agricultura familiar no município de Realeza-Sudoeste Paranaense. **Revista Brasileira de Higiene e Sanidade Animal: RBHSA**, v. 10, n. 4, p. 693-707, 2016.

BARI, M. S. et al. Subclinical mastitis in dairy cows in south-Asian countries: A review of risk factors and etiology to prioritize control measures. **Veterinary Research Communications**, [s. l.], v. 46, n. 3, p. 621-640, 2022.

BELLI, C. Z. P. et al. Qualidade do leite cru refrigerado obtido em unidades produtivas no Sudoeste do Paraná. **Revista de Ciências Agroveterinárias**, v. 16, n. 2, p. 109-120, 2017.

BOTTON, F. S. et al. Relationship of total bacterial and somatic cell counts with milk production and composition - Multivariate analysis. **Acta Scientiarum. Animal Sciences**, v. 41, n. 1, p. e42568, 2019.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa n. 76, de 28 de novembro de 2018. Regulamentos técnicos que fixam a identidade e as características de qualidade que devem apresentar o leite cru refrigerado, o leite pasteurizado e o leite pasteurizado tipo A. **Diário Oficial da República Federativa do Brasil**, Brasília, Seção 1, p. 10, 2018.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa n. 77, de 28 de novembro de 2018. Regulamento técnico de produção, identidade e

qualidade do leite tipo A; o regulamento de identidade e qualidade de leite cru refrigerado; o regulamento técnico de identidade e qualidade de leite pasteurizado; e o regulamento técnico da coleta de leite cru refrigerado e seu transporte a granel. **Diário Oficial da República Federativa do Brasil**, Brasília, Seção 1, p. 9, 2018b.

BRUM, L. M. da L. et al. Aplicação de técnicas de mineração em dados de propriedades leiteiras do município de Derrubadas-RS. In: **Simpósio da Ciência do Agronegócio**, 5, 2017, Porto Alegre. Anais... Porto Alegre: UFRGS, 2017.

CAK, B.; KESKIN, S.; YILMAZ, O. Regression tree analysis for determining of affecting factors to lactation milk yield in Brown Swiss cattle. **Asian Journal of Animal and Veterinary Advances**, v. 8, n. 4, p. 677-682, 2013.

CELIK, S. et al. Comparison of different data mining algorithms for prediction of body weight from several morphological measurements in dogs. **Journal of Animal and Plant Sciences**, v. 27, p. 57-64, 2017.

COSTA, G. M. et al. Risk factors for high bulk milk somatic cell counts in dairy herds from Campos das Vertentes region, Minas Gerais State, Brazil: A case-control study. **Pesquisa Veterinária Brasileira**, v. 39, n. 8, p. 606-613, 2019.

DA COSTA, L. B.; RAJALA-SCHULTZ, P. J.; SCHUENEMANN, G. M. Management practices associated with presence of *Staphylococcus aureus* in bulk tank milk from Ohio dairy herds. **Journal of Dairy Science**, v. 99, n. 2, p. 1364-1373, 2016.

DONG, F.; HENNESSY, D. A.; JENSEN, H. H. Factors determining milk quality and implications for production structure under somatic cell count standard modification. **Journal of Dairy Science**, v. 95, n. 11, p. 6421-6435, 2012.

DORSHORST, N. C.; COLLINS, M. T.; LOMBARD, J. E. Decision analysis model for paratuberculosis control in commercial dairy herds. **Preventive Veterinary Medicine**, v. 75, n. 1-2, p. 92-122, 2006.

EMBRAPA. Desempenho recente do setor lácteo e perspectivas para 2023. Disponível em: <https://www.embrapa.br/noticia/23405619/desempenho-recente-do-setor-lacteo-e-perspectivas-para-2023>.

EYDURAN, E. et al. Determination of factors influencing birth weight using regression tree (RT) method. **Journal of Applied Animal Research**, v. 34, n. 2, p. 109-112, 2008.

FONSECA, L. F. L.; SANTOS, M. V. Qualidade do leite e controle de mastite. São Paulo: **Lemos Editorial**, 2000.

FREDDO, A. R. et al. Árvores de decisão como método de mineração de dados: Análise de prontuários de uma clínica escola de nutrição. **Revista da Associação Brasileira de Nutrição - RASBRAN**, v. 10, n. 2, p. 31-37, 2019.

GUERREIRO, P. K. et al. Qualidade microbiológica de leite em função de técnicas profiláticas no manejo de produção. **Ciência e Agrotecnologia**, Lavras, v. 29, n. 1, p. 216-222, fev. 2005.

- HOOGERHEIDE, S. L.; MATTIODA, F. Qualidade bacteriológica do leite cru refrigerado em propriedades rurais do estado do Paraná. **Revista do Instituto de Laticínios Cândido Tostes**, v. 67, n. 385, p. 58-63, 2012.
- KARADAS, K. et al. Measuring predictive performance of data mining and artificial neural network algorithms for predicting lactation milk yield in indigenous Akkaraman sheep. **Pakistan Journal of Zoology**, v. 49, n. 1, 2017.
- KHAN, M. A. et al. Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. **Journal of Animal & Plant Sciences**, v. 24, n. 1, 2014.
- LORDÃO, A. C. et al. Implantação de medidas de higiene na ordenha para melhoria da qualidade do leite no município de Paty do Alferes/RJ, Brasil. **Archives of Veterinary Science**, v. 18, n. 4, 2013.
- LOPEZ, A. et al. Characterization of fat quality in cow milk from Alpine farms as influenced by seasonal variations of diets. **Animals: An Open Access Journal from MDPI**, v. 12, n. 4, p. 515, 2022.
- MIKAIL, N.; BAKIR, G. Regression tree analysis of factors affecting first lactation milk yield of dairy cattle. **Applied Ecology and Environmental Research**, Budapest, Hungary, v. 17, n. 2, p. 5293-5303, 2019.
- MILLER, E. A.; NESI, C. N. Prevalência de agentes causadores de mastite, qualidade do leite e conformidade com a IN n. 51. **Unoesc & Ciência – ACET**, v. 3, n. 2, p. 195-204, 2012.
- MITCHELL, T. M. Machine Learning. McGraw–Hill **Science/Engineering/Math**, 432 páginas, ISBN 0070428077, 1997.
- MOHAMMAD, M. T. et al. Prediction of body weight from body measurements using regression tree (RT) method for indigenous sheep breeds in Balochistan, Pakistan. **The Journal of Animal and Plant Sciences**, v. 22, p. 20-24, 2012.
- MINISTÉRIO DA AGRICULTURA E PECUÁRIA. Mapa do Leite — Políticas Públicas e Privadas para a Cadeia do Leite. Disponível em : <https://www.gov.br/agricultura> .
- NÄÄS, I. A. et al. Dairy cows estrus estimation using predictive and quantitative methods. **Ciência Rural**, v. 38, n. 8, p. 2383-2387, 2008.
- NEETHIRAJAN, S.; KEMP, B. Digital livestock farming. **Sensing and Bio-Sensing Research**, [s. l.], v. 32, p. 100408, 2021.
- OGUNTUNJI, A. O. Regression tree analysis for predicting body weight of Nigerian Muscovy duck (*Cairina moschata*). **Genetika**, v. 49, n. 2, p. 743-753, 2017.

PERISSINOTTO, M.; DE MOURA, D. J. Evaluation of thermal comfort in dairy cattle using data mining. **Revista Brasileira de Engenharia de Biosistemas**, v. 1, n. 2, p. 117-126, 2007.

PINZÓN-SÁNCHEZ, C.; CABRERA, V. E.; RUEGG, P. L. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. **Journal of Dairy Science**, v. 94, n. 4, p. 1873-1892, 2011.

RIBAS, N. P. et al. Escore de células somáticas e sua relação com os componentes do leite em amostras de tanque no estado do Paraná. **Archives of Veterinary Science**, v. 19, n. 3, 2014.

SILVA, C. G. et al. Influência da sanificação da água e das práticas de ordenha na qualidade do leite. **Arquivos Brasileiros de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 70, n. 2, p. 615-622, mar. 2018.

SITKOWSKA, B. et al. Detection of high levels of somatic cells in milk on farms equipped with an automatic milking system by decision trees technique. **Turkish Journal of Veterinary and Animal Sciences**, v. 41, n. 4, p. 532-540, 2017.

STEENSELS, M. et al. A decision-tree model to detect post-calving diseases based on rumination, activity, milk yield, BW and voluntary visits to the milking robot. **Animal**, v. 10, n. 9, p. 1493-1500, 2016.

STEPANIAK, L. et al. Avaliação microbiológica de *Escherichia coli* em leite e água de propriedades da agricultura familiar de Realeza, Paraná. In: **Seminário de Ensino, Pesquisa e Extensão da UFFS**, 8. 2018, Realeza. Anais do SEPE. UFFS, 2018.

TAFFAREL, L. E. et al. Variação da composição e qualidade do leite em função do volume de produção, período do ano e sistemas de ordenha e de resfriamento. **Semina: Ciências Agrárias**, v. 36, n. 1, p. 2287-2299, 2015.

TAKMA, C. et al. Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. **Kafkas Üniversitesi Veteriner Fakültesi Dergisi**, v. 18, n. 6, p. 941-944, 2012.

TELLES, T. S. et al. Microrregiões especializadas na produção de leite no sul do Brasil. In: **Anais do 56º Congresso da SOBER**. Anais... Campinas-SP, 2018.

TARIQ, M. M. et al. Prediction of body weight from body measurements using regression tree (RT) method for indigenous sheep breeds in Balochistan, Pakistan. **Journal of Animal & Plant Sciences**, v. 22, n. 1, p. 20-24, 2012.

TOPAL, M. et al. An analysis of the factors affecting birth weight and actual milk yield in Swedish red cattle using regression tree analysis. **Journal of Animal & Plant Sciences**, v. 20, p. 63-69, 2010.

VENÂNCIO PEREIRA, C.; NUNES, P. A estrutura da cadeia produtiva do leite no sudoeste do Paraná. [s. l.], 2021. Disponível em:

https://www.researchgate.net/publication/354184819_A_ESTRUTURA_DA_CADEIA_PRODUTIVA_DO_LEITE_NO_SUDOESTE_DO_PARANA.

YAKUBU, A. Application of regression tree methodology in predicting the body weight of Uda sheep. **Scientific Papers: Animal Science and Biotechnologies**, v. 45, n. 2, p. 484-490, 2012.

YILMAZ, I.; EYDURAN, E.; KAYGISIZ, A. Determination of non-genetic factors influencing birth weight using regression tree method in Brown-Swiss cattle. **Canadian Journal of Applied Sciences**, v. 1, n. 3, p. 382-387, 2013.