



LÁZARO HENRIQUE DA SILVA

**PREDIÇÃO DO COMPORTAMENTO INGESTIVO DE
BOVINOS EM PASTEJO AO LONGO DO REBAIXAMENTO
DO DOSSEL A PARTIR DE DADOS DE ACELERÔMETROS.**

LAVRAS – MG

2021

LÁZARO HENRIQUE DA SILVA

PREDIÇÃO DO COMPORTAMENTO INGESTIVO DE BOVINOS EM PASTEJO AO LONGO DO REBAIXAMENTO DO DOSSEL A PARTIR DE DADOS DE ACELERÔMETROS.

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia, área de concentração em Produção e Nutrição de Ruminantes, para a obtenção do título de Mestre.

Orientador: Marina de Arruda Camargo Danes, Dra. UFLA

Coorientador: Daniel Rume Casagrande, Dr. UFLA

LAVRAS – MG

2021

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Silva, Lázaro Henrique da.

Predição do comportamento ingestivo de bovinos em pastejo ao longo do rebolamento do dossel a partir de dados de acelerômetros.

/ Lázaro Henrique da Silva. - 2021.

99 p. : il.

Orientador(a): Marina de Arruda Camargo Danes.

Coorientador(a): Daniel Rume Casagrande.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Pecuária de precisão. 2. Uso de sensores para prever o comportamento ingestivo. 3. Estratégias de validação. I. Danes, Marina de Arruda Camargo. II. Casagrande, Daniel Rume. III. Título.

LÁZARO HENRIQUE DA SILVA

**PREDIÇÃO DO COMPORTAMENTO INGESTIVO DE BOVINOS EM PASTEJO
AO LONGO DO REBAIXAMENTO DO DOSSEL A PARTIR DE DADOS DE
ACELERÔMETROS.**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia, área de concentração em Produção e Nutrição de Ruminantes, para a obtenção do título de Mestre.

Aprovado em 29 de julho de 2021

Dra Marina de Arruda Camargo Danés, UFLA

Dr Adenilson José Paiva, UFRRJ

Dr Tiago Bresolin, UW-Madison (EUA)

Orientadora: Prof^ª. Dra. Marina de Arruda Camargo Danes, UFLA.

Coorientador: Prof. Dr. Daniel Rume Casagrande, Dr. UFLA.

LAVRAS – MG

2021

Para todos que acreditam e defendem a ciência brasileira, em especial aos meus pais, minha namorada e meu irmão que sempre me apoiaram nesse caminho.

Dedico

AGRADECIMENTOS

À minha namorada Kássia Fernanda, minha principal confidente, pelo amor, companheirismo e paciência nos momentos difíceis. Começamos a namorar no ensino médio e hoje lhe agradeço ao finalizar o mestrado. Também agradeço todas as consultorias técnicas durante esse período.

Aos meus pais, Eliane e Arlindo, por todo suporte fornecido até este momento e pelo suporte nos meus próximos passos. Por todos os ensinamentos, conselhos e valores transmitidos. Ao meu irmão Lucas por toda companhia, torcida e conselhos ao longo dos anos, é uma honra ter você como irmão.

À minha orientadora Marina de Arruda Camargo Danes por todos os ensinamentos, conselhos e “puxões de orelha” ao longo dos últimos anos, saiba que hoje escrevo muito mais evoluído como pessoa, zootecnista e pesquisador graças a você. Agradeço também ao meu coorientador Daniel Rume Casagrande por todo apoio durante esses dois anos.

Ao grupo INPPAR - Inovação, Pesquisa e Pessoas para Alimentação de Ruminantes, por todo suporte na condução do experimento, sem vocês seria impossível realizar esse trabalho. Aprendi muito com vocês e espero que também tenha sido capaz de ensinar algo.

Ao NEFOR, em especial os orientados do professor Daniel Rume Casagrande pelo trabalho em conjunto durante o trabalho a campo.

Aos doutores Tiago Bresolin e João Ricardo Rebouças Dórea pela parceria na condução das análises.

Aos doutores Marina Danes, Tiago Bresolin, Adenilson Paiva, Daniel Casagrande e Thiago Bernardes, por aceitarem o pedido para compor minha banca.

Aos professores e servidores da UFLA, em especial aos professores e servidores do Departamento de Zootecnia, por todo ensinamento e cuidado com os estudantes.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“A vida é muito curta e ansiosa para aqueles que esquecem o passado, negligenciam o presente e temem o futuro.”

Sêneca

RESUMO

O uso de novas tecnologia tem contribuído para o aumento na eficiência da produção bovina, porém pouco dessa tecnologia tem sido usada na produção a pasto, base da pecuária no Brasil. Pensando nisso, o objetivo desse trabalho foi avaliar diferentes modelos preditivos, estratégias de validação e composições do conjunto de dados para a predição do comportamento ingestivo de bovinos a pasto, com base em dados gerados pelo sensor do tipo acelerômetro. O experimento foi conduzido em pasto consorciado de braquiária brizantha e amendoim forrageiro. Para testar a mudança do comportamento ingestivo ao longo do rebaixamento, a área foi manejada no sistema rotativo de pastejo, com altura de entrada de 25 cm e três alturas de saída (20, 15 e 10 cm). As observações ocorreram ao longo de nove meses, em dias não consecutivos, durante 12 horas por dia. Os modelos preditivos utilizados foram: *generalizer linear regression* (GLR), *random forest* (RF), *gradiente boosting* (GB) e *artificial neural network* (ANN). As validações utilizadas foram: *holdout*, *leave-animals-out* (LAO), *leave-days-out* (LDO), *leave-height-out* de 10 cm (LHO10) e *leave-height-out* de 25 cm (LHO25). Foram utilizados dois conjuntos de dados, o primeiro foi o conjunto de dados PRO, com as observações de pastejo, ruminação e ócio, o segundo foi o conjunto de dados PNP, com as observações de pastejo e não pastejo, e um banco de dados externo. Por fim, os parâmetros usados para a avaliação dos modelos preditivos foram: acurácia, taxa de erro (para o conjunto de dados PNP), sensibilidade, especificidade, valor predito positivo e valor predito negativo. Para o conjunto de dados PRO, o melhor modelo preditivo foi o ANN, principalmente ao predizer o comportamento de pastejo, com acurácia de 60,5% (LAO), 65,3% (LDO), 71,8% (*holdout*) 60,5% (LHO10) e 63,2% (LHO25). Já no conjunto de dados PNP os melhores modelos preditivos foram o ANN, com acurácia de 63,6% (LAO), 65,8% (LDO), 73,0% (*holdout*), 60,8% (LHO10) e 64,9% (LHO25), e o RF, com acurácia de 62,5% (LAO), 64,4% (LDO), 73,3% (*holdout*), 59,9% (LHO10) e 61,7% (LHO25). De maneira geral, os modelos foram mais eficientes ao predizer apenas dois comportamentos do que ao predizer três comportamentos, principalmente pela dificuldade em predizer o comportamento de ócio, com sensibilidade abaixo de 26% em quase todas as estratégias de validação utilizadas. Outro ponto importante a se considerar é que a estratégia de validação adotada pode interferir nos resultados dos parâmetros de avaliação, como observado com a *holdout*, que teve acurácia superior as demais estratégias de validação ao inflar os modelos preditivos. O conjunto de dados externo expôs os modelos a uma nova situação, onde a estratégia *holdout* não foi superior as demais, com acurácia de 57,5% (LAO), 59,4% (LDO) e 59,4% (*holdout*), mostrando a necessidade de expor os modelos preditivos a situações novas, como a entrada de novos animais no piquete e a diferentes estruturas do pasto.

Palavras chave: comportamento animal. pecuária de precisão. pastejo. validação cruzada. modelos preditivos.

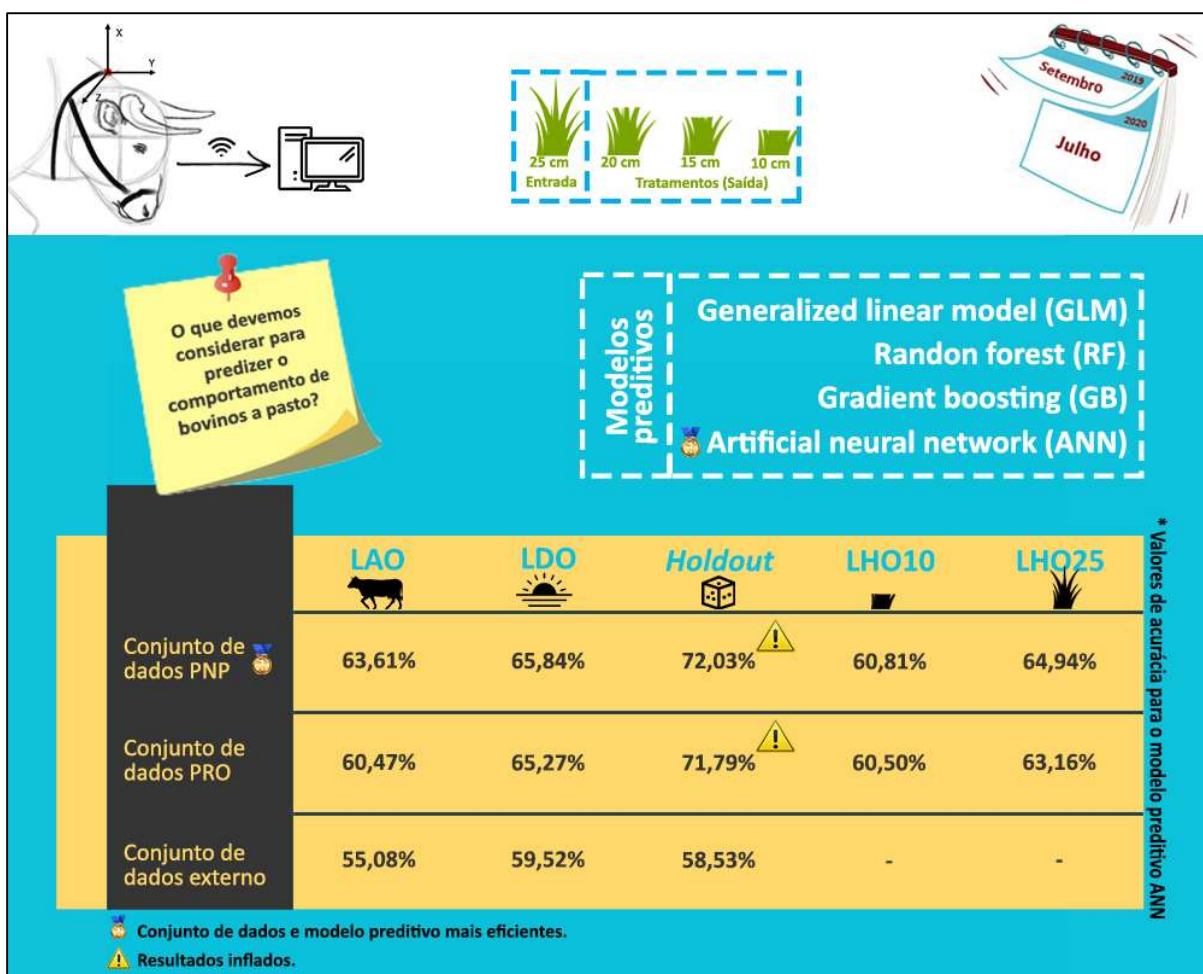
ABSTRACT

The use of new technology has contributed to the increase in the efficiency of cattle production. Still, little of this technology has been used in pasture production, the basis of livestock in Brazil. With this in mind, the objective of this work was to evaluate different predictive models, validation strategies, and dataset compositions for the prediction of ingestive behavior of grazing cattle based on data generated by an accelerometer-type sensor. The experiment was carried out in intercropped pasture of *Urochloa brizantha* cv Marandu and *Arachis pintoi*. To test the change in ingestive behavior during lowering, the area was managed in a rotational grazing system, with an entry height of 25 cm and three exit heights (20, 15, and 10 cm). Observations took place over nine months, on non-consecutive days, for 12 hours a day. The predictive models used were: generalized linear regression (GLM), random forest (RF), gradient boosting (GB), and artificial neural network (ANN). The validations used were: holdout, leave-animals-out (LAO), leave-days-out (LDO), 10 cm leave-height-out (LHO10) and 25 cm leave-height-out (LHO25). Two datasets were used. The first was the PRO dataset, with grazing, rumination, and idle observations; the second was the PNP dataset, with grazing and non-grazing observations and an external database. Finally, the parameters used to evaluate the predictive models were: accuracy, error rate (for the PNP dataset), sensitivity, specificity, positive predicted value, and negative predicted value. For the PRO dataset, the best predictive model was the ANN, mainly in predicting grazing behavior, with an accuracy of 60.5% (LAO), 65.3% (LDO), 71.8% (holdout), 60.5% (LHO10), and 63.2% (LHO25). In the PNP dataset, the best predictive models were the ANN, with an accuracy of 63.6% (LAO), 65.8% (LDO), 73.0% (holdout), 60.8% (LHO10) and 64.9% (LHO25), and the RF, with an accuracy of 62.5% (LAO), 64.4% (LDO), 73.3% (holdout), 59.9% (LHO10) and 61.7% (LHO25). In general, the models were more efficient in predicting only two behaviors than in predicting three behaviors, mainly due to the difficulty in predicting idle behavior, with sensitivity below 26% in almost all validation strategies used. Another important point to consider is that the adopted validation strategy can interfere with the results of the evaluation parameters, as observed with the holdout, which had greater accuracy than other validation strategies when inflating the predictive models. The external dataset exposed the models to a new situation, where the holdout strategy was not superior to the others, with an accuracy of 57.5% (LAO), 59.4% (LDO), and 59.4% (holdout), showing the need to expose predictive models to new situations, such as the entry of new animals into the paddock and to different pasture structures.

Keywords: animal behavior. precision-livestock. grazing. validation. predictive models.

RESUMO INTERPRETATIVO E INFOGRÁFICO

O uso de novas tecnologia contribui para a eficiência na produção bovina, porém pouco dessa tecnologia é utilizada na pecuária a pasto, base da produção brasileira. Por isso, o objetivo foi desenvolver modelos capazes de prever o comportamento ingestivo de bovinos em pastejo a partir de dados de acelerômetros. Os modelos preditivos testados foram *generalizer linear regression* (GLR), *random forest* (RF), *gradiente boosting* (GB) e *artificial neural network* (ANN). As validações foram: *holdout*, que seleciona o subconjunto de teste aleatoriamente, *leave-animals-out* (LAO), simulando a utilização dos modelos em novos animais, *leave-days-out* (LDO), expondo o modelo a diferentes estruturas da pastagem, *leave-height-out* de 10 cm (LHO10) e *leave-height-out* de 25 cm (LHO25), expondo os modelos a diferentes alturas da pastagem. Foram utilizados dois conjuntos de dados: o conjunto de dados PRO, com as observações de pastejo, ruminação e ócio, o conjunto de dados PNP, com as observações de pastejo e não pastejo, e um banco de dados externo. Para a maioria das avaliações, o modelo ANN foi o mais eficiente. Os modelos foram mais acurados no conjunto de dados PNP. A estratégia de *holdout* infla os resultados ao ignorar as inter-relações entre animal, planta e ambiente. Novas estratégias são necessárias para refletir a realidade da pecuária.



Legenda: *Leave-animals-out* (LAO), *leave-days-out* (LDO), *Holdout* (20% dos dados selecionados para testagem), *Leave-height-out* 10 cm (LHO10), *Leave-height-out* 25 cm (LHO25), conjunto de dados PNP, com as categorias pastejando e não pastejando, conjunto de dados PRO, com as categorias pastejando, ruminando e ócio.

LISTA DE ILUSTRAÇÕES

<i>Figura 1: Artigos publicados sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto ao longo dos últimos 29 anos.</i>	20
<i>Figura 2: Mapa dos países que mais publicaram artigos sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto.</i>	21
<i>Figura 3: Boxplot da quantidade de animais (n) utilizados nos artigos sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto.</i>	22
<i>Figura 4: Sensores mais comumente utilizados para prever o comportamento ingestivo de animais a pasto com base nos artigos presentes nessa revisão de literatura.</i>	23
<i>Figura 5: Locais de fixação dos sensores mais utilizados para prever o comportamento ingestivo de animais a pasto com base nos artigos presentes nessa revisão de literatura.</i>	24
<i>Figura 6: Imagem esquemática de um sistema de pastejo rotacionado composto por oito piquetes, onde o lote principal inicia o rebaixamento e um lote secundário finaliza o rebaixamento.</i>	25
<i>Figura 7: Tempo, em minutos, destinados as atividades de pastejo, ruminação e ócio, de bovinos da raça nelore, à pasto de capim Xaraés durante 12 horas de observações diárias.</i>	28
<i>Figura 8: Cabresto com o switch (à esquerda); bovino utilizando o switch no cabresto e o equipamento para registro no dorso (à direita).</i>	30
<i>Figura 9: Switch para a detecção de movimentos mandibulares. A movimentação plataforma superior (A), ocasionada pelo movimento da mandíbula, aciona o contato entre as peças B e C.</i>	30
<i>Figura 10: Quantidade de artigos publicados sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto ao longo de 30 anos.</i>	31
<i>Figura 11: Mapa com a localização dos principais centros de ensino e pesquisa em tecnologia agropecuária do Brasil, e as regiões com um ecossistema mais desenvolvido.</i>	33
<i>Figura 12: Imagem esquemática sobre o funcionamento de um sistema automatizado. Com um sensor capaz de captar um estímulo ambiental, um microcontrolador que processa o sinal e um display que retorna uma resposta.</i>	34
<i>Figura 13: Três tipos de aprendizado para um modelo de machine learning e suas respectivas aplicações mais comuns.</i>	36
<i>Figura 14: Representação da classificação de um modelo preditivo, o primeiro com o erro underfitting, outro com a classificação adequada e o último com o erro overfitting.</i>	37
<i>Figura 15: Separação de um banco de dados em dados para treinamento (training set), dados para testagem do modelo (test set) e dados para validação do modelo (validation set), com uma estratégia de validação denominada cross-validation.</i>	38
<i>Figura 16: Esquema de uma matriz de confusão 2x2. A categoria positivo e negativo é determinada pelo cientista de dados.</i>	39
<i>Figura 17: Etapas para a aplicação de modelos preditivos.</i>	40
<i>Figura 18: A) Esquema de uma rede neural biológica, com os neurônios e suas sinapses. B) Esquema de uma rede neural artificial com suas unidades de processamento e conexões.</i>	41

<i>Figura 19: Esquema de um modelo Ensemble Learning, com o banco de dados, a utilização de vários modelos e a geração da predição.</i>	42
<i>Figura 20: Esquema de um modelo bagging (esquerda) com o banco de dados sendo dividido aleatoriamente em vários conjuntos de dados para treinamento e os resultados sendo agregados. Esquema de um boosting (direita) com os treinamentos sequenciais e resultados sendo agregados.</i>	43
<i>Figura 21: Imagem esquemática de um modelo preditivo de Random Forest com “n” decision trees.</i>	44
<i>Figura 22: Modelo esquemático de um gradient boosting com 50 predições. Mostrando os modelos preditivos (Esquerda) e erros residuais (Direita) das predições 1, 2 e 50.</i>	45
<i>Figura 23: Esquema de uma validação k-fold Cross-validation, com a separação do banco de dados (all data) em banco de dados para treinamento (training data) e para teste (test data) e a separação dos dados de treinamento em cinco subconjuntos (Folds).</i>	46
<i>Figura 24: Duas estratégias de validação leave-one-out. LOAO, ao selecionar um animal para a validação e LODO, ao selecionar um dia do rebaixamento para compor a validação.</i>	47
<i>Figura 25: Distribuição dos três tratamentos em 12 unidades experimentais (de 1 a 12) e sua distribuição pela área experimental, cada UE foi dividida em três piquetes.</i>	49
<i>Figura 26: Representação dos sensores utilizados, suas dimensões e seu funcionamento. Orientação dos eixos.</i>	50
<i>Figura 27: Representação da fixação dos sensores em bovinos e os eixos capturados pelo mesmo, eixo X (vertical), Y e Z (horizontais).</i>	51
<i>Figura 28: Sensores antes da utilização. Já protegidos pelo balão e presos com amarrilho, barbante e fita. Todo o conjunto era fixado no cabresto com auxílio de barbante e fita.</i>	51
<i>Figura 29: Imagem esquemática do funcionamento dos sensores no pasto. As linhas representam as conexões, tanto wi-fi quanto via cabo.</i>	52
<i>Figura 30: Boxplot comparando a altura real com a esperada para todos os tratamentos, desfolhação leve, média e severa, e altura de entrada (25 cm).</i>	61
<i>Figura 31: Porcentagem do tempo gasto nas atividades de pastejando, ruminando, ócio e não pastejando (ruminando e ócio), para a alturas de 25 cm, 20 cm, 15 cm e 10 cm.</i>	62
<i>Figura 32: Divisão do conjunto de dados PNP em subconjunto de treinamento e subconjunto de validação cruzada (teste) para as diferentes estratégias de validação.</i>	63
<i>Figura 33: Distribuição dos dados brutos ao longo de uma hora, escolhido aleatoriamente, de observações comportamentais de pastejando (superior) e não pastejando (inferior).</i>	64
<i>Figura 34: Divisão do conjunto de dados PRO em subconjunto de treinamento e subconjunto de validação cruzada (teste) para as diferentes estratégias de validação.</i>	68
<i>Figura 35: Distribuição dos dados brutos, ao longo de uma hora, de pastejando (superior), ruminando (centro) e ócio (inferior) para o conjunto de dados PRO.</i>	69

LISTA DE TABELAS

<i>Tabela 1: Combinações entre os modelos preditivos de classificação e estratégias de validação utilizados. ____</i>	<i>56</i>
<i>Tabela 2: Média de precipitação, temperatura ao longo do dia e umidade relativa nos dias de observações comportamentais, separados por estação do ano. _____</i>	<i>58</i>
<i>Tabela 3: Precipitação, temperatura máxima e mínima para os sete dias utilizados na validação leave-days-out (LDO). _____</i>	<i>59</i>
<i>Tabela 4: Porcentagem de falhas no recebimento de dados por intensidade de desfolhação, na entrada e saída dos animais dos piquetes, e a porcentagem total. _____</i>	<i>59</i>
<i>Tabela 5: Comparação entre a porcentagem de falha no recebimento de dados por sensor. _____</i>	<i>60</i>
<i>Tabela 6: Descrição dos melhores hiper parâmetros ou arquitetura identificada na pesquisa de grade aleatória para modelos preditivos testados no banco de dados PNP. _____</i>	<i>64</i>
<i>Tabela 7: Desempenho do treinamento dos modelos para prever os comportamentos de pastejando ou não pastejando usando diferentes estratégias de validação. _____</i>	<i>66</i>
<i>Tabela 8: Desempenho da validação dos modelos preditivos para prever categorias de comportamento de pastejo ou não pastejo visualmente observadas em novilhas tabapuã usando diferentes estratégias de validação. _____</i>	<i>67</i>
<i>Tabela 9: Descrição dos melhores hiper parâmetros ou arquitetura identificada na pesquisa de grade aleatória para modelos preditivos capazes de prever os comportamentos de pastejando, ruminando e ócio. _____</i>	<i>70</i>
<i>Tabela 10: Desempenho do treinamento de modelos preditivos para prever categorias de comportamento de pastejo, ruminação e ócio visualmente observados em novilhas Tabapuã usando diferentes estratégias de validação. _____</i>	<i>71</i>
<i>Tabela 11: Validação dos modelos preditivos para prever categorias de comportamento de pastejo, ruminação e ócio visualmente observados em novilhas tabapuã usando diferentes estratégias de validação. _____</i>	<i>73</i>
<i>Tabela 12: Validação externa dos modelos preditivos para prever categorias de comportamento de pastejo ou não pastejo visualmente observadas em novilhas e bovinos Tabapuã usando diferentes estratégias de validação. _____</i>	<i>76</i>

LISTA DE ABREVIATURAS

ANN – *Artificial neural network*

DT – *Decision Trees*

EL - *Ensemble Learning*

FN – Falso negativo

FP – Falso positivo

GB – *Gradient boosting*

GLM - *Generalized linear model*

IA – Inteligência Artificial

IPO - Oferta pública inicial

KNN - *K-Nearest Neighbor*

LAO – *Leave-animals-out*

LDO - *leave-days-out*

LHO10 - *Leave-height-out* 10 cm

LHO25 - *Leave-height-out* 25 cm

NB - *Naive Bayes*

RF – *Random Forest*

TN – Verdadeiro negativo

TP – Verdadeiro positivo

VPN – Valor preditivo negativo

VPP – Valor preditivo positivo

SUMÁRIO

1 - INTRODUÇÃO	16
2 – REFERENCIAL TEÓRICO	19
2.1 – Uso de sensores para predizer o comportamento a pasto	19
2.2 – Manejo Rotacionado	24
2.3 - Comportamento ingestivo	27
2.4 – Automação	29
2.4.1 - Pecuária de precisão	29
2.4.2 – Instrumentação na pecuária	33
2.5 - Machine Learning	35
2.5.1 – Pré-processamento e Treinamento	37
2.5.2 – Modelos preditivos	40
2.6 – Validação	45
3 - MATERIAL E MÉTODOS	48
3.1 Animais e período experimental	48
3.2 Sensor e software	50
3.3 – Análise exploratória dos dados	53
3.4 - Modelos preditivos	53
3.4 – Estratégias de validação	55
4 – RESULTADOS	58
4.1 – Conjunto de dados PNP: Pastejando x Não Pastejando	62
4.2 – Conjunto de dados PRO: Pastejando x Ruminando x Ócio	68
4.3 – Conjunto de dados externo	76
5 – DISCUSSÃO	77
6 – CONCLUSÃO	85
REFERÊNCIAS	86
APÊNDICE	95

1 - INTRODUÇÃO

No ano de 2019 a pecuária foi responsável por 8,5% do PIB brasileiro, o equivalente a 618,5 bilhões de reais, sendo um importante setor da economia. A maior parte do plantel de bovinos brasileiros é mantido em sistemas a pasto. Segundo a ABIEC (2020) aproximadamente 86% dos bovinos abatidos no Brasil foram terminados em sistemas a pasto.

A produtividade dos bovinos a pasto merece um destaque. No ano de 1990 a produtividade era de 1,6 arrobas por hectare ao ano, já no ano de 2019 a produtividade foi de 4,3 arrobas por hectare ao ano, que representa um aumento de 169% nos últimos trinta anos. Em contrapartida houve uma redução no uso de áreas de pastagem, diminuindo de 192,2 milhões de hectares em 1990 para 162,5 milhões de hectares em 2019 (ABIEC, 2020). Dois fatores explicam como o Brasil foi capaz de ampliar sua produção de bovinos diminuindo a quantidade de pastagens: os avanços nos estudos sobre planta, animal e na correlação entre ambos, e o aumento no uso de tecnologias (ABIEC, 2020).

A utilização do método de pastejo rotacionado, que busca aumentar a quantidade de animais na propriedade, sem prejudicar a pastagem é um desses avanços. Esse sistema consiste na divisão da área de pastagem em vários piquetes, que são submetidos a uma rotação no seu uso. Dessa maneira, parte da pastagem está em uso, enquanto o restante está em um período de descanso, para recuperar do pastejo anterior.

Diversos procedimentos podem ser adotados para auxiliar na tomada de decisão quanto ao manejo do pasto, principalmente para determinar o momento ideal de retirada dos animais dos piquetes.

A altura do dossel é o método referência para determinar a entrada e saída dos animais dos piquetes, porém é um método que apresenta algumas limitações. É um método desgastante para os funcionários, além de exigir um treinamento para que a medição seja confiável. Algumas propriedades separam os animais entre um lote de maior exigência e um de menor, o primeiro lote recebe o pasto em melhores condições e quando o pasto rebaixa o segundo lote entra no piquete para finalizar o rebaixamento. Saber o momento exato da troca de lotes é uma tarefa complicada, pois exige medições

constantes. Piquetes muito grandes também são um desafio, pois são necessárias várias amostras e algumas vezes é até inviável.

Outra estratégia utilizada para determinar o momento ideal de saída dos animais do pasto é a visualização do comportamento ingestivo. O comportamento dos animais se altera de acordo com a altura do pasto, o tempo gasto pelos animais para o pastejo aumenta à medida que o pasto é rebaixado, em contrapartida o tempo em ócio diminui com o rebaixamento, o tempo ruminando se mantém. Porém quando o pasto está muito rebaixado o tempo em ócio aumenta, enquanto o tempo pastejando e ruminando caem, demonstrando que os animais não estão conseguindo apreender o alimento. Identificar essa mudança de comportamento pode se tornar uma alternativa para auxiliar na tomada de decisão sobre o manejo da pastagem (DE FREITAS BARBOSA *et al.*, 2013). Comercialmente, observação visual do comportamento ingestivo dos animais é uma alternativa inviável, uma vez que é necessário um funcionário treinado, capaz de permanecer horas observando os animais sem interferir no seu comportamento natural. Uma possibilidade de viabilizar esse método é a automação na coleta de dados.

A automação visa utilizar técnicas computacionais, elétricas ou mecânicas com o objetivo de otimizar tarefas cotidianas, e auxiliar as necessidades humanas, principalmente em atividades repetitivas, desgastantes ou que ofereça riscos. Nos últimos anos pesquisas relacionadas a adoção de tecnologias para o monitoramento do comportamento animal vêm crescendo fortemente (NEETHIRAJAN; KEMP, 2021).

Na avaliação de comportamento animal, essa automação acontece principalmente pela adoção de sensores vestíveis, como colares e brincos, contendo, principalmente, um acelerômetro que coleta dados de movimentação dos animais. Apesar de ser uma tecnologia extensamente estudada em animais confinados, o uso de sensores em animais em pastejo ainda é incipiente (PARK; PARK, 2021). Os poucos dados gerados em pastagens temperadas (ARABLOUEI *et al.*, 2021; SHAFIULLAH *et al.*, 2019) são pouco aplicáveis em locais de clima tropical, por causa da diferença na heterogeneidade da estrutura do dossel em pastos adaptados a essas regiões. Essa maior complexidade da relação animal x ecossistema de pastagens em forrageiras tropicais pode exigir modelos preditivos mais complexos, capazes de lidar com relações não lineares entre as variáveis (AIKEN *et al.*, 2019). Além disso, os métodos de validação cruzada dos modelos preditivos precisam levar em conta as interações biológicas

entre as observações, já que uma validação cruzada completamente aleatória, como é feito na maioria dos estudos, pode inflar a acurácia dos modelos (RIBEIRO *et al.*, 2021)

Os objetivos foram comparar o desempenho dos modelos preditivos utilizados em dois diferentes bancos de dados, comparar se os modelos são mais eficientes ao prever dois ou três comportamentos, e verificar as diferentes estratégias de validação.

Para isso, o presente trabalho baseia-se no uso de sensores do tipo acelerômetro triaxial para prever o comportamento ingestivo de bovinos a pasto em três diferentes manejos de pastagem: rebaixamento leve, moderado e severo. Os modelos preditivos foram testados em dois diferentes conjuntos de dados oriundos do mesmo banco de dados. O primeiro conjunto de dados difere os comportamentos entre pastejando, ruminando e ócio, enquanto o segundo distinguindo entre pastejando e não pastejando. Foram utilizadas quatro técnicas de *machine learning*: *Generalized Linear Model* (GLM), *Random Forest* (RF), *Artificial Neural Network* (ANN) e *Gradiente Boosting* (GB). As estratégias de validação utilizadas foram *holdout*, *leave-animals-out* (LAO), *leave-days-out* (LDO), *leave-height-out* de 10 cm (LHO10) e *leave-height-out* de 25 cm (LHO25). Uma validação externa foi realizada utilizando o presente banco de dados para validar modelos treinados previamente, em outro trabalho do nosso grupo (RIBEIRO *et al.*, 2021).

2 – REFERENCIAL TEÓRICO

A automação é uma realidade em vários setores da sociedade. Na pecuária não seria diferente, cada vez mais nos deparamos com a aplicação da automação, e da ciência de dados, no que chamamos de pecuária de precisão. Porém, diferente de alguns outros setores, a pecuária de precisão precisa levar em consideração os fatores animal, planta e sua interrelação. Por exemplo, a intensidade e a frequência do pastejo afetam o desenvolvimento das plantas, conseqüentemente afetam o desenvolvimento dos animais, pois interferem no tempo pastejando, no tempo em ócio, na quantidade de alimento digerida, e na facilidade em apreender o alimento (CASAGRANDE *et al.*, 2011; FERREIRA *et al.*, 2022).

2.1 – Uso de sensores para prever o comportamento a pasto

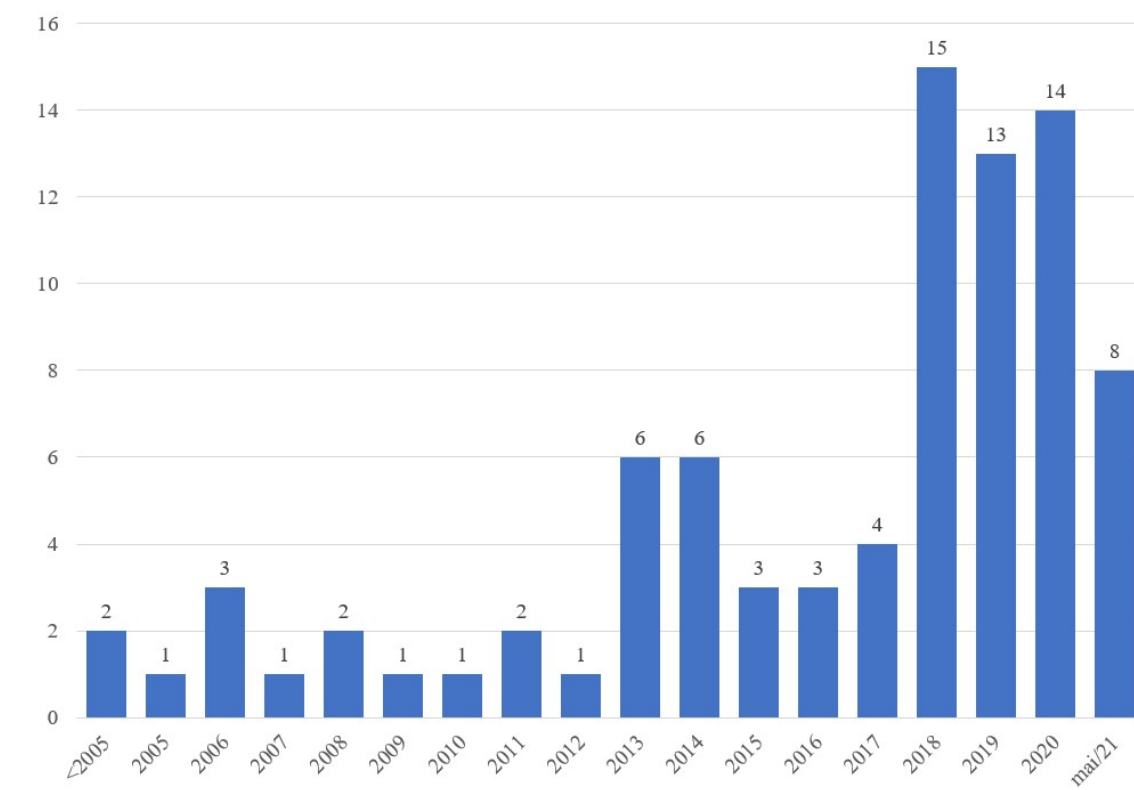
Essa revisão bibliográfica foi realizada com pesquisas na base SCOPUS de artigos publicados até o mês de maio de 2021. Foram utilizados quatro conjuntos de palavras chaves. A primeira pesquisa utilizou as palavras chaves *sensor + behavior + grazing*, retornando 182 artigos. A segunda pesquisa utilizou as palavras *sensor + behavior + animal*, retornando 116 artigos. A terceira pesquisa utilizou as palavras *behavior + machine learning*, retornando 23 artigos. Por fim a quarta pesquisa utilizou as palavras *Sensor + behavior + machine learning*, retornando 29 artigos. Dos 350 artigos encontrados 86 permaneceram após verificação de duplicidade e retirada de artigos que não estavam em conformidade com o tema proposto. As segunda, terceira e quarta pesquisas retornaram valores elevados de artigos encontrados, dessa forma foram selecionados apenas artigos das revistas *Journal of dairy Science*, *Animal Science*, *Computers and electronics in agriculture* e *IEEE sensor journal*, as revistas foram escolhidas pelo seu fator de impacto, utilizando a plataforma *Scimago Journal & Country Rank* para verificação.

As informações observadas foram: ano de publicação, país, espécie animal utilizada, número de animais, tipos de sensores utilizados, posição dos sensores e modelos preditivos utilizados.

Apesar de não ser um conceito novo, a zootecnia de precisão, principalmente a pecuária de precisão, tem um destaque recente. Dos artigos encontrados apenas 12,8% são anteriores ao ano de 2010, a maioria dos estudos se concentra nos últimos dez anos,

principalmente nos últimos quatro anos, que correspondem a mais da metade dos artigos publicados, 58% das publicações, como pode ser observado na Figura 1.

Figura 1: Artigos publicados sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto ao longo dos últimos 29 anos.

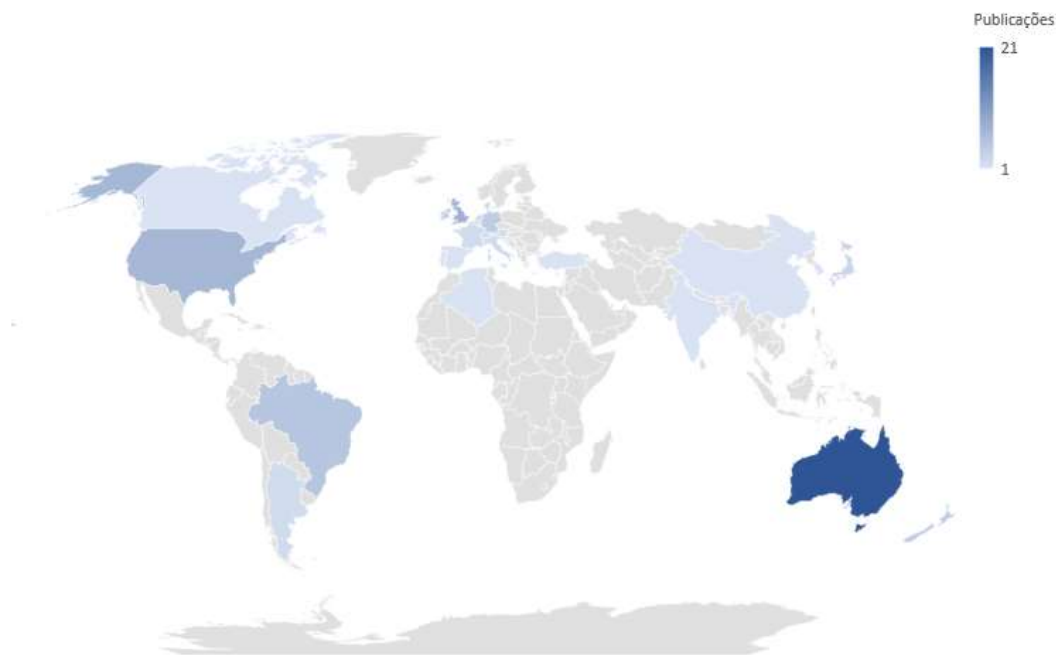


Legenda: Publicados até maio de 2021

Fonte: Do autor (2021).

As pesquisas encontradas foram desenvolvidas em 25 países diferentes. Os quatro países com mais artigos, Austrália, Reino Unido, Estados Unidos da América e Brasil, são responsáveis por 47% das publicações. Dos 21 países restantes 13 publicaram apenas um ou dois artigos (Figura 2). A Austrália é responsável por 21 publicações, enquanto o Brasil é responsável por 5 artigos. Esse mapa mostra o quanto as pesquisas sobre o uso de sensores para prever o comportamento de bovinos a pasto está concentrado em poucos países, principalmente os países ditos como primeiro mundo como Austrália, Reino Unido e Estados Unidos da América.

Figura 2: Mapa dos países que mais publicaram artigos sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto.



Legenda: Austrália (21), Reino Unido (8), Estados Unidos da América (7), Brasil (5), Alemanha (4), Israel (4), Nova Zelândia (4), Dinamarca (3), Irlanda (3), Japão (3), Países Baixos (3), Suíça (3), Argentina (2), Áustria (2), Bélgica (2), França (2), Itália (2), Argélia (1), Canadá (1), China (1), Coreia do Sul (1), Espanha (1), Índia (1), Portugal (1), Turquia (1)

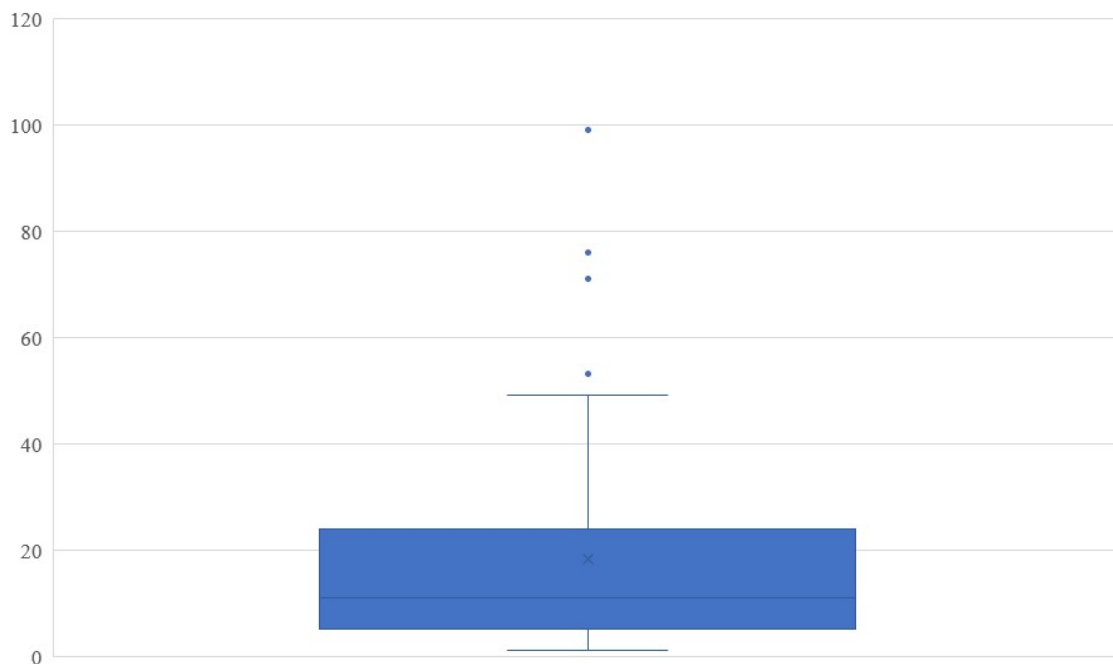
Fonte: Do autor (2021).

A espécie mais utilizada são os bovinos que aparecem em 59 artigos, 40 deles para bovinos de leite e 19 para bovinos de corte. A segunda espécie mais utilizada são os ovinos, com 22 artigos, em sequência aparecem os caprinos com seis artigos. Três artigos utilizam equinos e um artigo utiliza cervídeos. Os caprinos foram muito utilizados no Reino Unido, 6 dos 8 artigos publicados.

O número de animais utilizados varia entre os artigos e é dependente da quantidade de sensores disponíveis. Dois artigos utilizam uma quantidade de animais superior a 100 e foram retirados dessa análise, pois alteraram muito a média da quantidade de animais utilizados. O valor médio da quantidade de animais utilizados é 18 animais. Contudo, quatro artigos utilizaram mais de 50 animais, o que aumentou a média, mesmo com a mediana em 11 animais. A moda é de apenas três animais,

quantidade de animais presentes em 9 dos 86 artigos. Na Figura 3 podemos observar um gráfico em *boxplot* contendo a média da quantidade de animais usados, a mediana, os maiores e menores valores e os *outliers*.

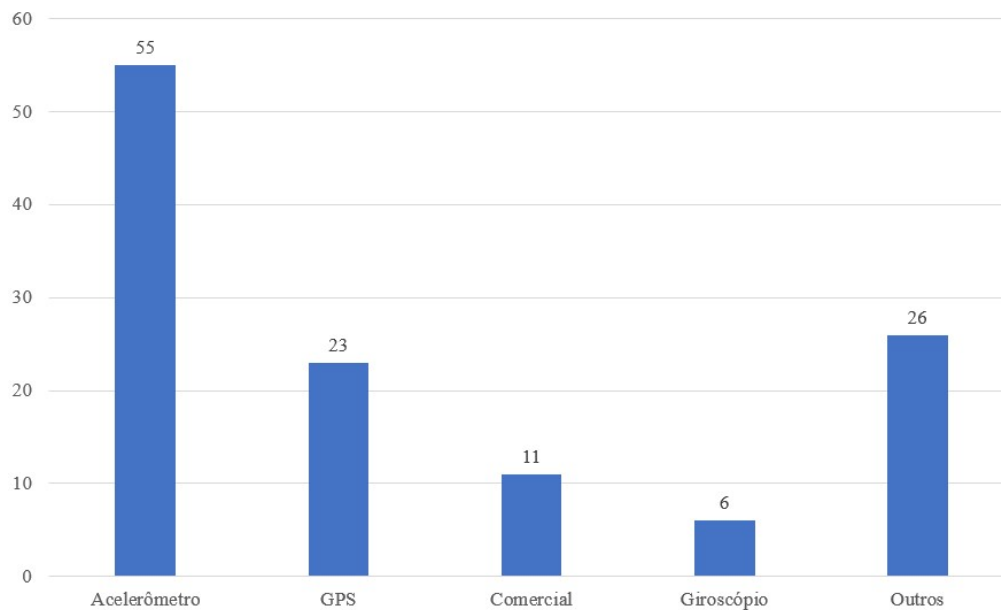
Figura 3: *Boxplot* da quantidade de animais (n) utilizados nos artigos sobre o uso de sensores para predizer o comportamento ingestivo de animais a pasto.



Fonte: Do autor (2021).

Vários sensores podem ser utilizados para o monitoramento do comportamento ingestivo dos animais a pasto, alguns aparecem com mais frequência, enquanto outros são pouco utilizados. O sensor mais utilizado é o acelerômetro, que foi utilizado em 46% dos artigos pesquisados. Em segundo lugar vem o GPS, que aparece em 19% dos artigos. Os equipamentos comerciais, que não indicam quais sensores são utilizados em seu *hardware*, aparecem em terceiro lugar, com 9% das publicações. O giroscópio apareceu em 5% dos artigos, mas sempre em conjunto com o acelerômetro, em um sistema denominado *Inertial Navigation System*. Todos os demais sensores, como microfones e eletrodos, foram classificados como “Outros” e aparecem em 21% artigos. Em 26 artigos foram utilizados mais de um sensor. Nesses casos os dados dos sensores são mesclados para a classificação do comportamento. Na Figura 4 é possível observar a quantidade de artigos que utilizam cada um dos sensores.

Figura 4: Sensores mais comumente utilizados para prever o comportamento ingestivo de animais a pasto com base nos artigos presentes nessa revisão de literatura.

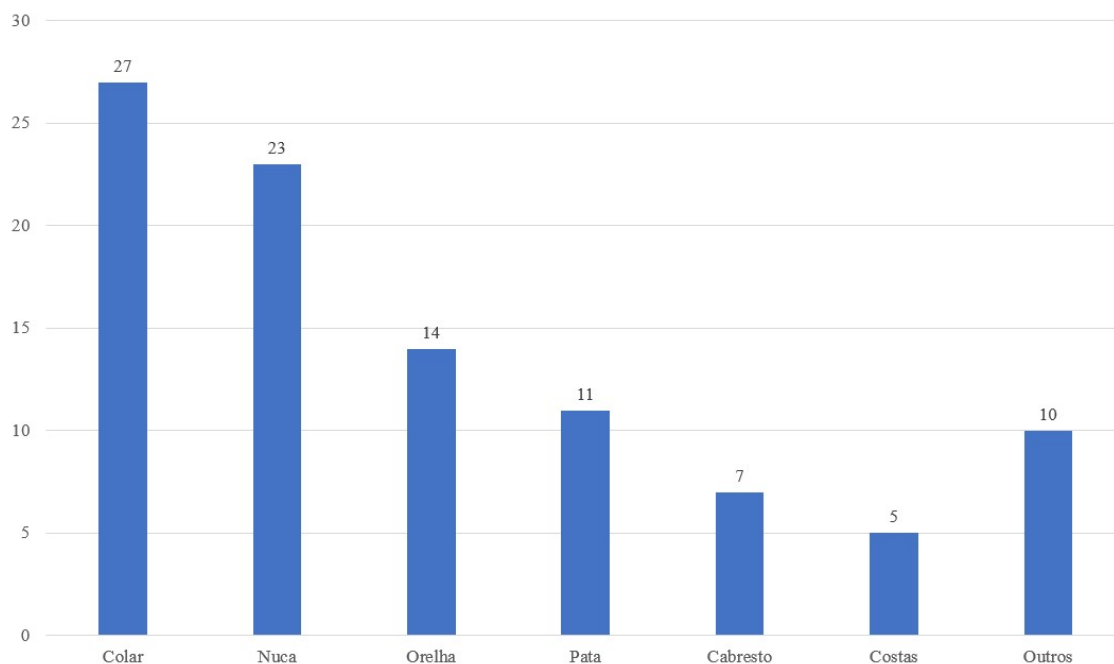


Fonte: Do autor (2021).

O local de fixação dos sensores é um fator importante que deve ser avaliado, pois a programação do modelo preditivo pode variar dependendo do local escolhido. Além disso, os componentes eletrônicos contidos nos sensores são equipamentos sensíveis e necessitam de proteção de acordo com o ambiente no qual serão inseridos, logo são diferentes para cada local de fixação por estarem expostos a diferentes intempéries. Por outro lado, a proteção pode aumentar o preço do produto, é necessário então um equilíbrio entre o custo e a proteção. Não foi encontrado um artigo que compare os diferentes locais de fixação.

O local mais comum para a fixação dos sensores foi em colares, seguido da nuca e da orelha (Figura 5). Todos esses locais utilizam a movimentação da cabeça para realizar a classificação dos comportamentos. Diferente dos sensores fixados nas patas, quarto local mais comum, que utiliza o caminhar para identificar os comportamentos. A fixação em cabrestos e nas costas aparecem em quinto e sexto lugares, respectivamente. Todos os outros lugares de fixação foram classificados como “Outros” e incluem drone, cauda, músculo masseter, maxilar e rúmen.

Figura 5: Locais de fixação dos sensores mais utilizados para prever o comportamento ingestivo de animais a pasto com base nos artigos presentes nessa revisão de literatura.



Fonte: Do autor (2021).

Mais de 10 estratégias de modelagem preditivas foram utilizadas nos artigos encontrados, demonstrando que várias técnicas podem ser aplicadas, dependendo das características do banco de dados e da decisão do cientista de dados. Algumas técnicas são mais comuns como GLM, *Decision tree* e *Random Forest*.

Uma dificuldade encontrada durante a leitura dos artigos foi identificar a metodologia utilizada, grande parte artigos utilizados na revisão tinham informações faltando, como número de animais, quais sensores utilizados, qual modelo preditivo escolhido e, principalmente, qual estratégia de validação adotada.

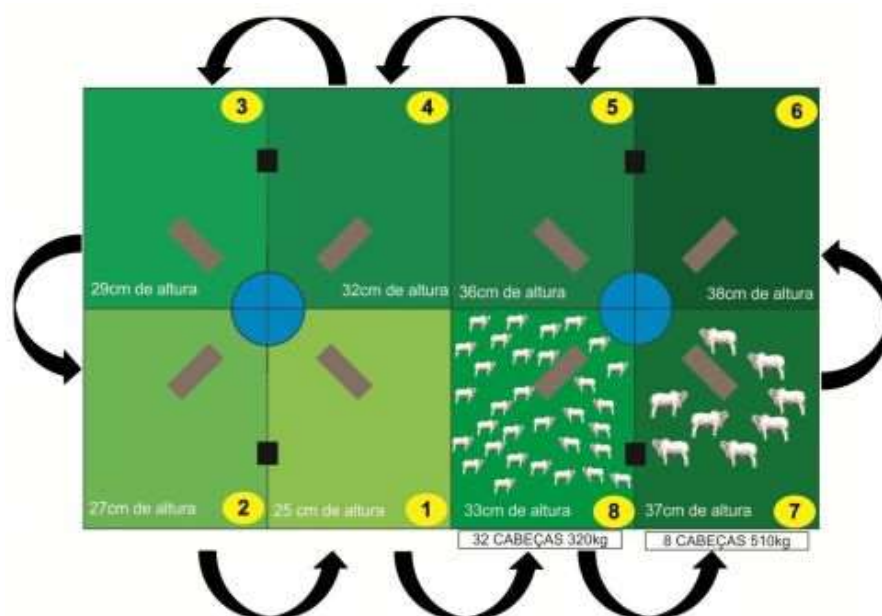
2.2 – Manejo Rotacionado

O manejo rotacionado é caracterizado pela divisão de uma área em piquetes e pela rotação periódica dos animais por esses piquetes, consequentemente rotacionando os piquetes em uso e em descanso. A Figura 6 ilustra o manejo rotacionado, os animais de alta produção então no piquete 7, assim que chegar na altura ideal eles vão para o

piquete 6, já os animais jovens estão no piquete 8, eles entraram após a saída dos animais de alta produção. Os demais piquetes estão em descanso.

Esse manejo visa favorecer a rebrota da forrageira, sem prejudicar a produção animal, os estudos sobre o manejo do pasto no Brasil permitiram ajustes no planejamento e execução dessa estratégia (DE GIMENES *et al.*, 2011; EUCLIDES *et al.*, 2014, 2018; SAVIAN *et al.*, 2021). Também é relatado uma melhora no temperamento dos bovinos mantidos no manejo rotacionado, devido a maior interação humano-animal (CEBALLOS *et al.*, 2018; ROCHA GÓIS *et al.*, 2016).

Figura 6: Imagem esquemática de um sistema de pastejo rotacionado composto por oito piquetes, onde o lote principal inicia o rebaixamento e um lote secundário finaliza o rebaixamento.



Fonte: Nepomuceno (2016).

O momento ideal de entrada e saída dos animais é extremamente importante para a produção em manejo rotacionado e influencia diretamente na intensidade do pastejo (PEREIRA, LILIAN ELGALISE TECHIO; POLIZEL, 2016; TRINDADE *et al.*, 2007). A intensidade de pastejo, caracterizado pelo resíduo pós pastejo, permite que a modifiquemos de acordo com o objetivo do produtor. Desfolhação leve diminui a taxa de lotação e o ganho por área, mas aumenta o consumo de matéria seca por animal. Já a desfolhação intensa aumenta a taxa de lotação e o ganho por área, mas diminui o

consumo de matéria seca por animal (DE MOURA ZANINE *et al.*, 2011; EUCLIDES *et al.*, 2014; LAMPERT *et al.*, 2020; SAVIAN *et al.*, 2021).

Apesar dessa possibilidade de escolha, é preciso ficar atento para o manejo adequado da intensidade de pastejo. O manejo inadequado pode ocasionar uma menor produtividade da forrageira, além da má formação da estrutura de pasto, com áreas degradadas ou acúmulo de colmo e material morto. Ambos os casos comprometem a produção animal por interferirem no comportamento ingestivo. Quando a pastagem é manejada de maneira severa, o intervalo entre pastejos aumenta, pois é necessário que a planta use suas reservas energéticas para formar a primeira folha, e só então realizar a fotossíntese. Se o pastejo é severo por longos períodos de tempo, as reservas das plantas podem se exaurir e ela não serão mais capazes de fazer o rebrotamento, tornando a pastagem uma área degradada. Já quando o pastejo é leniente por vários ciclos, ocorre um acúmulo de material morto, que por si só já é uma perda financeira, pois é um alimento que deixou de ser ofertado aos animais.

A literatura cita como altura ideal para a retirada dos animais, quando esses removam entre 40% e 60% da altura de entrada, caso o produtor não opte por uma estratégia de priorizar a pastagem ou os animais (CARNEVALLI *et al.*, 2006; SILVEIRA *et al.*, 2013; ZANINI; SANTOS; SBRISSIA, 2012).

Dentre as várias técnicas para determinação do momento de entrada e saída dos animais de um piquete, uma possibilidade é a utilização dias fixos, onde os animais permanecem nos piquetes por uma quantidade de dias pré-determinados, tendo como vantagem a facilidade desse manejo. O grande problema quando se opta por essa alternativa de tomada de decisão é que ela não leva em consideração a sazonalidade da produção. A produção da forrageira é dependente de diversos fatores, tais como temperatura e chuvas, e eles não são homogêneos ao longo do ano, ou mesmo ao longo de uma mesma estação, conseqüentemente a produção não é homogênea (OKUI *et al.*, 2020; PEDREIRA; PEDREIRA; DA SILVA, 2007). De Gimenes *et al.* (2011) verificaram em seu trabalho que o crescimento da forrageira foi próximo ao esperado apenas no verão, nas demais estações o crescimento foi inferior ao esperado para os dias pré-estabelecidos. Quando trabalhamos com dias fixos não temos controle sobre a altura de entrada ou saída dos animais, podendo ocasionar áreas com subpastejo ou superpastejo.

Outro método bastante indicado é a utilização da altura da pastagem como indicador do momento de saída dos animais, por sua alta correlação com a interceptação de luz (IL). Existem diversas ferramentas para realizar as medições, como régua específicas, marcações na cerca e outros. Apesar de ser um método preciso, é mais trabalhoso que a opção de dias fixo, esse tipo de manejo demanda que um funcionário treinado realize constantemente a verificação de altura dos piquetes, tarefa na qual despense bastante tempo além de ser uma atividade desgastante. (DE MOURA ZANINE *et al.*, 2011; SILVA, S. C.; NASCIMENTO JÚNIOR, 2007).

2.3 - Comportamento ingestivo

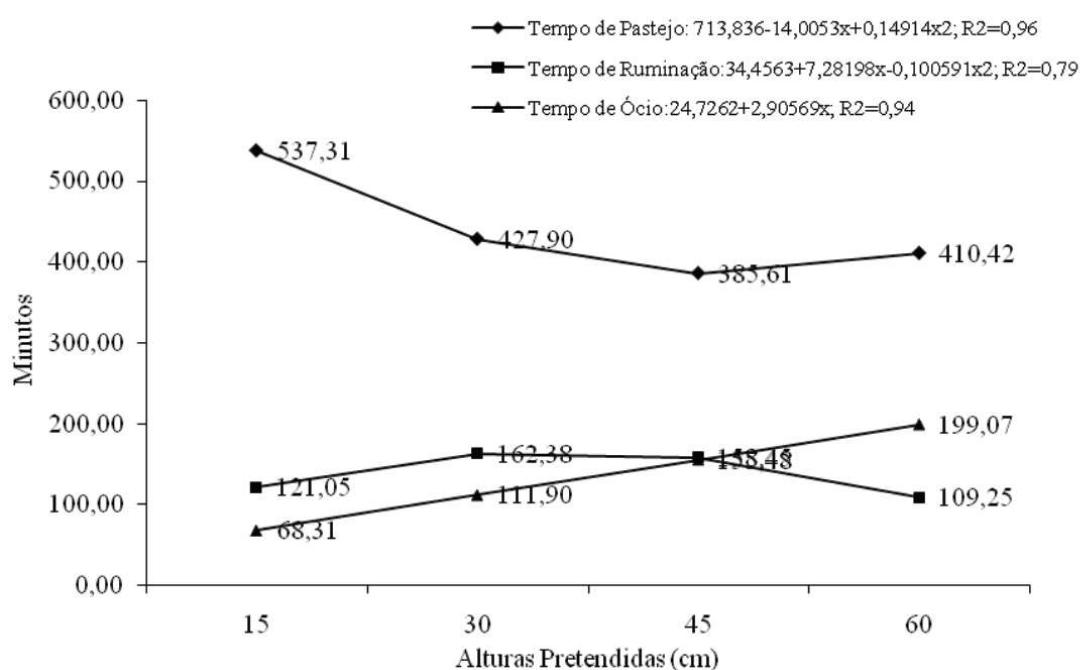
O comportamento ingestivo dos bovinos a pasto é um processo complexo, que envolve características da planta, animal e ambiente, com destaque para as características estruturais da forrageira (DA SILVA, S. C.; CARVALHO, 2005; DONALD, 1986). A estrutura da pastagem é um fator determinante no comportamento ingestivo de animais a pasto, pois os animais ajustam seu comportamento, até um certo limite, em decorrência da estrutura da pastagem, principalmente em situações desfavoráveis (EUCLIDES *et al.*, 2014; PALHANO *et al.*, 2006; TRINDADE *et al.*, 2007).

Em pastos com maior disponibilidade de forragem, ou seja, pastagens que se encontram com altura elevada e alta IL, os animais gastam menos tempo na atividade de pastejo. Nesse cenário, a colheita de forragem é otimizada, porém acima de uma determinada altura a estrutura do pasto passa a atrapalhar a ingestão de alimentos, ampliando o tempo gasto pastejando. À medida que o pasto é rebaixado os animais gastam mais tempo pastejando, até alcançarem um limite, onde passa a ser inviável permanecer procurando alimento. (BARCELLOS *et al.*, 2004; DA COSTA LIMA *et al.*, 2020; GONTIJO NETO *et al.*, 2006).

De Freitas Barbosa *et al.* (2013) avaliaram em seu trabalho o tempo despendido para as atividades de pastejo, ruminação e ócio, em quatro diferentes alturas para pastos Xaraés, com observações de 12 horas (Figura 7). Quando a pastagem estava com altura 60 cm, superior ao recomendado como entrada, os animais gastaram menos tempo ruminando que em ócio, demonstrando que, apesar do tempo pastejando ter sido próximo à altura de 30 cm, os animais não conseguiram apreender alimento com facilidade. Para a altura de 45 cm, superior a recomendada pela literatura, os animais gastaram menos tempo pastejando e tempos similares para ruminando e ócio, com

tempo ruminando bem próximo ao gasto para a altura de 30 cm, considerada a altura ideal de entrada. À medida que a pastagem foi rebaixada, os animais gastaram mais tempo pastejando e menos tempo em ócio. Na altura de 15 cm, considerada altura ideal para a saída, os animais gastaram quase duas horas a mais pastejando quando comparado à altura de 30 cm, apesar disso o tempo ruminando foi 41 minutos inferior, demonstrando que os animais estavam tendo dificuldades em localizar alimentos.

Figura 7: Tempo, em minutos, destinados as atividades de pastejo, ruminação e ócio, de bovinos da raça nelore, à pasto de capim Xaraés durante 12 horas de observações diárias.



Fonte: De Freitas Barbosa et al. (2013).

Difante *et al.* (2009) observou uma correlação negativa entre o tempo pastejando e a altura do pasto em capim Tanzânia. À medida que a altura diminuía mais tempo era despendido para o pastejo, resultado semelhante ao encontrado por Hendricksen; Minson 1980 com forrageira de clima temperado. Esses estudos demonstram também uma alteração na taxa de bocado.

A taxa de bocado, frequência de bocados por minuto, permite analisar a facilidade que os animais tem de selecionar e apreender os alimentos. A taxa de bocado varia em função da estrutura do pasto. Em pastagens mais rebaixadas, os animais tem

maior taxa de bocado, enquanto pastos mais altos, os animais tem menores taxas de bocado. Isso demonstra que pastos mais baixos dificultam a apreensão de alimento, além de indicar uma maior procura por alimento à medida que o pasto é rebaixado (FLORES *et al.*, 2008; SARMENTO, 2003).

2.4 – Automação

De maneira resumida a automação pode ser definida como o uso da potência elétrica ou mecânica para acionar algum tipo de dispositivo, ou sistema, com a mínima interferência manual, diminuindo o esforço de trabalho humano.

2.4.1 - Pecuária de precisão

A pecuária de precisão é o monitoramento de variáveis em níveis individuais, desenvolvimento de modelos preditivos que descrevam a interação animal e ambiente, e a comparação entre os modelos preditivos com o observado. Para tal faz-se uso de técnicas de automação, como sensoreamento, *big data* e *machine learning* (ANDRIAMANDROSO; HERINAINA; *et al.*, 2016; COCKBURN, 2020; ELLIS *et al.*, 2020).

Uma das tecnologias da automação muito utilizada na pecuária de precisão são os sensores, que são capazes de medir as variáveis de interesse e transformar essas variáveis em dados, captam sinais analógicos de algum parâmetro físico não elétrico e convertem para um sinal elétrico equivalente. Outra tecnologia empregada é a *big data*, que consiste na análise e interpretação de um conjunto de dados extenso, muitas vezes oriundos de um sensor. Enquanto técnicas de *machine learning* são empregadas para fazer as análises com *Big data*, algo inviável para um humano realizar (GARCIA, 2020; KABUGO *et al.*, 2020; RIBEIRO, MARCO ANTÔNIO, 1999).

Um dos trabalhos pioneiros nesse campo de estudos foi publicado por Duckworth; Shirlaw (1955) intitulado: *The Development of an apparatus to record the jaw movements of cattle*. Nesse artigo foi proposto um aparelho capaz de medir a movimentação da mandíbula dos bovinos (Figura 8) para registrar o tempo gasto pastejando, a taxa de bocado e o tempo gasto ruminando.

O aparelho consistia em um *Switch*, que era ligado ou desligado dependendo da movimentação da mandíbula (Figura 9). Quando o bovino abria a boca a peça “A”

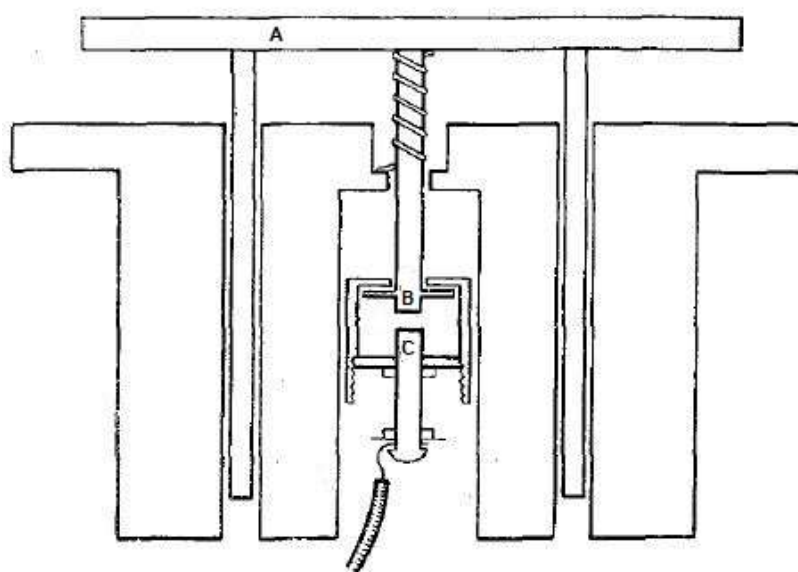
empurrava a “B”, e essa entrava em contato a “C” provocando um sinal elétrico. Esse sinal mantinha-se ligado enquanto o bovino permanecesse com a boca aberta e era cessado quando o bovino fechava a boca, assim era possível medir a movimentação mandibular.

Figura 8: Cabresto com o *switch* (à esquerda); bovino utilizando o *switch* no cabresto e o equipamento para registro no dorso (à direita).



Fonte: Adaptado de Duckworth; Shirlaw (1955).

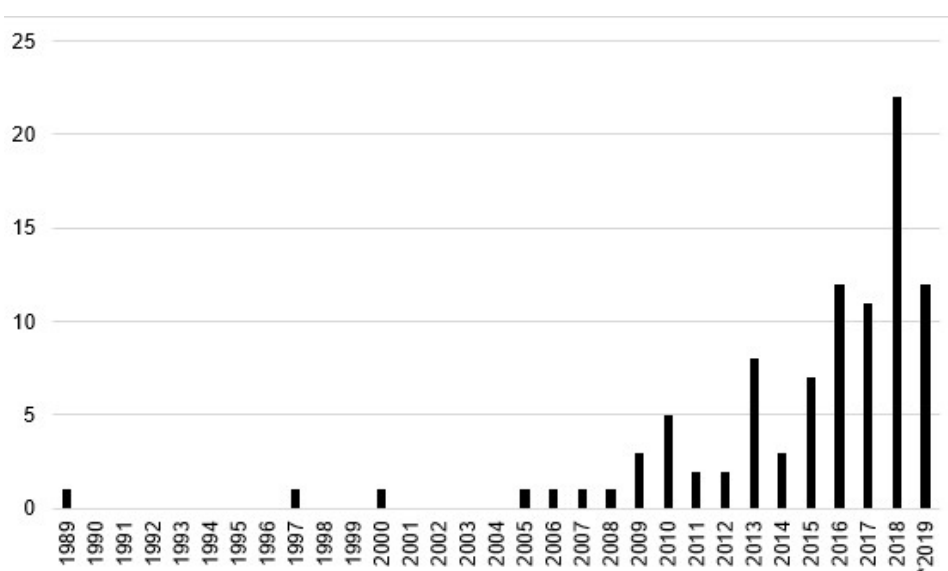
Figura 9: *Switch* para a detecção de movimentos mandibulares. A movimentação plataforma superior (A), ocasionada pelo movimento da mandíbula, aciona o contato entre as peças B e C.



Fonte: Adaptado de Duckworth; Shirlaw (1955).

A tecnologia avançou entre o ano de 1955 e os dias atuais. As técnicas de programação, os sensores e os computadores melhoraram, ficaram menores e mais acessíveis. A produção animal também acompanhou essas mudanças, mas de maneira tímida, de modo que a pecuária de precisão só recebeu a devida atenção nos últimos anos. Ribeiro, Leonardo Augusto Coelho (2020) demonstrou um crescimento no número de publicações sobre o uso de sensores para prever o comportamento ingestivo dos animais nos últimos anos (Figura 10). Dos 94 artigos estudados pelo autor, apenas 10 foram anteriores ao ano de 2010. Resultados semelhantes aos encontrados nessa revisão de literatura.

Figura 10: Quantidade de artigos publicados sobre o uso de sensores para prever o comportamento ingestivo de animais a pasto ao longo de 30 anos.



Fonte: Ribeiro, Leonardo Augusto Coelho (2020).

Semelhante ao crescimento nas pesquisas, também pode ser observado um aumento na quantidade de produtos de base tecnológica destinados a pecuária. Esses produtos desempenham diversas atividades como ferramentas de gerenciamento, técnicas de modelagem para simulações financeiras, monitoramento por imagem, uso da robótica, monitoramento dos animais através da microeletrônica e outros (AUGUSTO, 2019). Muitas dessas tecnologias são imprescindíveis para o desenvolvimento das

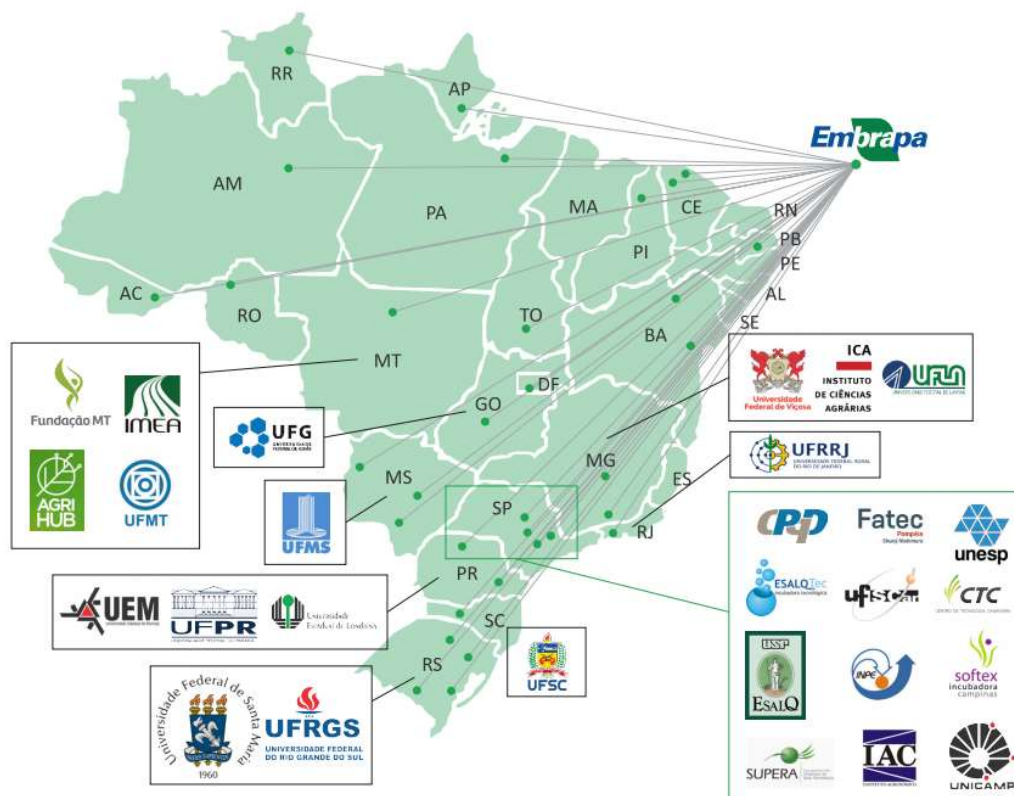
propriedades rurais como empresas, com decisões pautadas em dados e monitoramento eficiente, propriedades cada vez maiores e com escassez de mão de obra especializada.

Uma tecnologia adotada é o colar de ruminção, que através da microeletrônica possibilita ao produtor identificar o comportamento dos bovinos, acompanhar o consumo de alimento e prever doenças ou cio. Algumas empresas já contam com IA atuando como secretária, que são capazes de emitir alertas e lembretes sobre as decisões que os produtores precisam tomar em suas propriedades. Porém grande parte desses colares são destinados a bovinos de leite e bovinos de corte em sistemas de confinamento, ou são baseados em dados de regiões temperadas e não se aplicam a realidade das pastagens tropicais. Outro ponto é que as características únicas da pecuária a pasto no Brasil dificultam sua utilização, como a conectividade em propriedades extensas e a heterogeneidade das pastagens tropicais. O desenvolvimento de uma tecnologia de monitoramento a pasto é um mercado em potencial, uma vez que a pastagem é a base para a produção de bovinos no Brasil

Grande parte das tecnologias utilizadas na pecuária de precisão são desenvolvidas por empresas jovens, de base tecnológica, modelo de negócios repetível e escalável, também denominadas Startups. Hoje existem no Brasil, e no mundo, diversas startups voltadas ao agronegócio, também chamadas de *AgTech*, assim como *hubs* especializados nesse tipo de startup. A *AgTech* mais notória é a *Climate Corporation*, que se tornou o primeiro “unicórnio” do setor ao ser comprada pela Monsanto no ano de 2013. Unicórnios são empresas que atingem US \$1 bilhão em valor de mercado antes de abrir capital (IPO).

Segundo Dias; Jardim; Sakuda (2019) o Brasil conta com aproximadamente 1125 *startups* distribuídas em mais de 20 *hubs* especializados em *AgTech* e diversos outros que olham para as *AgTechs* como pilares de interesse. Grande parte dessas empresas estão localizados em regiões com um ecossistema já desenvolvido e próximo de Instituições de Ensino e Pesquisa em tecnologia agropecuária no Brasil (Figura 11).

Figura 11: Mapa com a localização dos principais centros de ensino e pesquisa em tecnologia agropecuária do Brasil, e as regiões com um ecossistema mais desenvolvido.



Fonte: Dias; Jardim; Sakuda (2019).

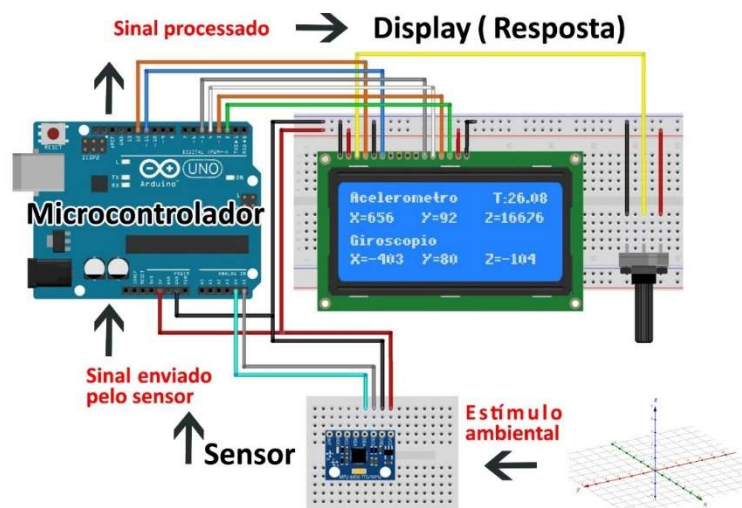
2.4.2 – Instrumentação na pecuária

Instrumentação é a ciência que estuda, desenvolve e aplica, instrumentos de medição, transmissão, indicação, registro e controle de variáveis físicas. Entre esses instrumentos podemos citar os sensores, que são dispositivos capazes de detectar e responder a estímulos físicos ou químicos, transformando esses estímulos em grandezas elétricas (LACA, 2009; RIBEIRO, MARCO ANTÔNIO, 1978).

Os sensores tem o funcionamento semelhante aos nossos órgãos sensoriais, como pele, olhos e ouvidos. Nossos órgãos sensoriais são capazes de captar estímulos ambientais e transformar em sinais, que são enviados ao sistema nervoso central, onde são processados e geram uma resposta. Os sensores captam um estímulo ambiental, transformam essa informação em sinais e os envia até algum processador, ou outro

dispositivo, que será encarregado de processar essa informação e transformar em uma resposta, como pode ser visto no exemplo da Figura 12.

Figura 12: Imagem esquemática sobre o funcionamento de um sistema automatizado. Com um sensor capaz de captar um estímulo ambiental, um microcontrolador que processa o sinal e um display que retorna uma resposta.



Fonte: Do autor (2021).

Vários tipos de sensores são utilizados na produção animal para desempenhar diferentes funções, principalmente na pecuária de precisão. Entre os sensores utilizados podemos citar o acelerômetro, capaz de medir a aceleração aplicada sobre um corpo. O giroscópio, capaz de medir a mudança na direção de um objeto em movimento. O termômetro, capaz de medir a temperatura de um objeto. As câmeras, capazes de captar a luminosidade rebatida ou emitida por um objeto. Entre outros.

O acelerômetro é muito utilizado para detectar o comportamento ingestivo de animais a pasto, algumas vezes sendo utilizado sozinho, outras sendo utilizado em conjunto com um giroscópio, ou com um GPS (ATKINSON *et al.*, 2009; CHAPA *et al.*, 2020; GUO *et al.*, 2018). A presente revisão demonstrou que o sensor mais utilizado é o acelerômetro, resultando em concordância com Ribeiro, Leonardo Augusto Coelho (2020) que também observou um uso mais recorrente desse sensor. Os acelerômetros são sensores capazes de transformar a energia mecânica do movimento em sinais elétricos mensuráveis. Dessa forma, são capazes de medir a aceleração de um objeto, a

aceleração devido à gravidade, a frequência da aceleração e a aceleração dinâmica, obtida com a movimentação dos animais, por exemplo a movimentação da cabeça (BRAUN; TSCHONER; HÄSSIG, 2014; DE ALMEIDA *et al.*, 2013; MANSBRIDGE *et al.*, 2018).

Segundo Ribeiro, Leonardo Augusto Coelho (2020) o local de fixação mais utilizado é a nuca, dos 96 artigos estudados pelo autor 40 a utilizavam como local de fixação, porém o autor não fez distinção para os colares. Em nossa revisão a nuca foi um local importante, com 23 dos 86 artigos. A nuca é muito utilizada pelo movimento que os animais fazem com a cabeça ao pastejar, e pela facilidade de fixar os sensores nesses locais.

2.5 - Machine Learning

Para entender as técnicas de *machine learning* primeiro precisamos entender a inteligência artificial (IA). Muitos conceitos são apresentados para definir uma IA, mas podemos defini-la como um ramo da ciência da computação que busca desenvolver dispositivos que simulem a capacidade humana de raciocinar, aprender, tomar decisões e resolver problemas (GOMES, 2010; JUNIOR BERTOZZO, 2019).

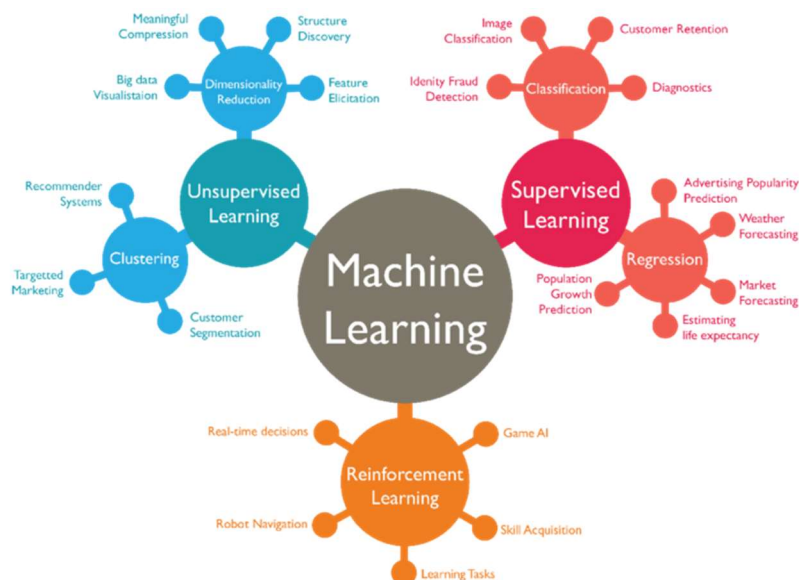
Machine learning é uma subárea da IA e é caracterizado pela criação de modelos capazes de fazer com que uma máquina aprenda a desempenhar uma tarefa, esse modelo utiliza um conjunto de dados de entrada e, com base em algum modelo, possibilita a geração de saídas. Nesse processo o modelo é melhorado, por isso quanto maior o banco de dados (*dataset*), mais preciso é o resultado (CARBONELL; MICHALSKI; MITCHELL, 1983; YANG *et al.*, 2019).

As técnicas de *machine learning* normalmente são classificadas em três grupos quanto a forma de aprendizado. O aprendizado supervisionado, onde o modelo aprende de resultados pré-estabelecidos, ou seja, o banco de dados foi rotulado, classificado ou categorizado previamente, também ocorre um *feedback* direto. O aprendizado não supervisionado, onde a ML aprende com dados de teste que não foram rotulados, classificados ou categorizados previamente, sem *feedback* direto. E o aprendizado por reforço, onde a máquina tenta aprender qual a melhor ação com base nas circunstâncias que se encontra, recebendo reforços positivos para comportamentos desejados e

reforços negativos para comportamentos indesejados (AMORIM; BARONE; MANSUR, 2008; DOUGHERTY; KOHAVI; SAHAMI, 1995; KENJI, 2019).

O aprendizado supervisionado é utilizado para testes de regressão e classificação. O aprendizado não supervisionado é utilizado para testes de *clustering* e de *dimensionality reduction*. O aprendizado por reforço é usado para atividades que exigem uma resposta em tempo real, como pode ser visto na Figura 13 (ELLIS *et al.*, 2020; SILVA, ÉMERSON, 2019; VALLETTA *et al.*, 2017). Utilizaremos técnicas de *machine learning* com aprendizado supervisionado, pois o intuito do presente trabalho é classificar o banco de dados em diferentes comportamentos.

Figura 13: Três tipos de aprendizado para um modelo de *machine learning* e suas respectivas aplicações mais comuns.

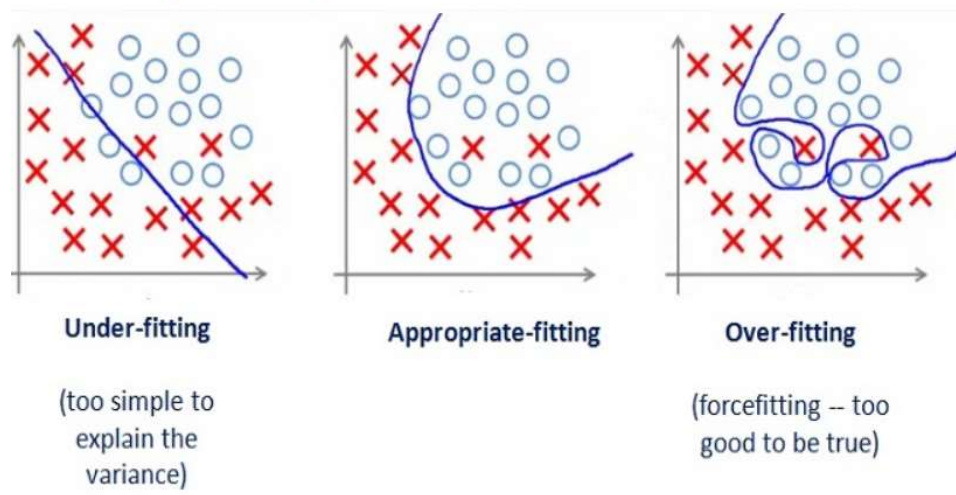


Fonte: Silva, Émerson (2019).

As técnicas de ML podem apresentar problemas caso o banco de dados utilizado seja pequeno ou com ruídos. Também é possível ocorrer erros durante o treinamento, causando *overfitting* ou *underfitting*, problemas que afetam a generalização do modelo preditivo (Figura 14). O *underfitting* é o erro causado por suposições errôneas do modelo de aprendizado, perdendo relações importantes entre os *inputs* e *outputs*. Isso faz com que o modelo não seja capaz de classificar os dados de maneira eficaz e a generalização é muito alta. O *overfitting* é o erro de sensibilidade para pequenas flutuações do banco de dados usado para o treinamento, fazendo com que o modelo

utilize ruídos aleatórios para o treinamento. Isso faz com que o modelo só seja eficiente para o banco de dados utilizado no treinamento e perde generalização (BHANDE, 2018).

Figura 14: Representação da classificação de um modelo preditivo, o primeiro com o erro *underfitting*, outro com a classificação adequada e o ultimo com o erro *overfitting*.



Fonte: Bhande (2018).

2.5.1 – Pré-processamento e Treinamento

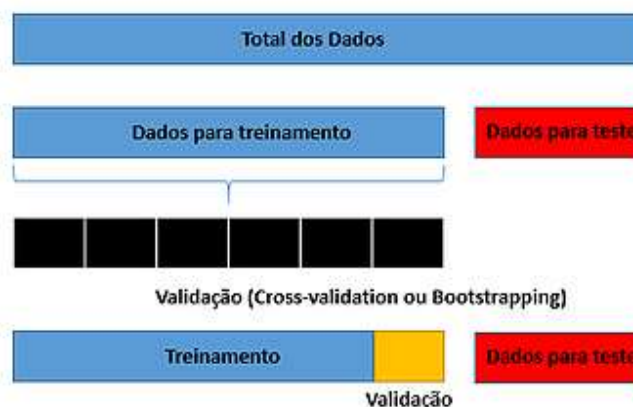
Antes de realizar o treinamento de um banco de dados é necessário fazer um pré-processamento do mesmo, onde são feitas uma limpeza e a organização dos dados. A coleta de dados pode ser realizada em diversas fontes, ou podem apresentar falhas, com isso o banco de dados pode conter alguns ruídos e valores ausentes. Utilizar um banco de dados sem tratamento para o treinamento do modelo pode produzir valores incorretos. As principais tarefas do pré-processamento são a limpeza dos dados, transformação dos dados para padronizá-los, separação de amostras para testagem, diferenciação de dados convertendo atributos contínuos para atributos categóricos, e limpeza de textos (LARRIVA-NOVO *et al.*, 2020; MICROSOFT, 2020; UYSAL; GUNAL, 2014).

O banco de dados pode conter dados estruturados e não estruturados. Dados estruturados são aqueles contidos em tabelas, com colunas e linhas bem definidas que identificam as informações, são facilmente interpretados por modelos. Os dados não estruturados são os dados mais comumente encontrados, devido as redes sociais e aos e-

mails, são dados que não estão contidos em tabelas, portanto os modelos não conseguem lê-los com tanta facilidade, fazem parte desse grupo os textos, áudios, vídeos e fotos. Na pecuária de precisão podemos ter dados estruturados, como a aceleração nos eixos X, Y e Z, mas também podemos ter dados não estruturados, como imagens ou padrões sonoros.

Depois do pré-processamento o banco de dados é dividido em três partes, o banco de dados para treinamento do modelo preditivo (*training set*), o banco de dados para validação do modelo preditivo (*validation set*) e o banco de dados para testagem do modelo (*test set*). A validação pode ser realizada com diversas técnicas. Já a testagem ocorre para medir a capacidade de generalização da rede, para isso o modelo encontrado é utilizado e então é verificado se o resultado encontrado está de acordo com o esperado. Para manter a imparcialidade na avaliação do modelo preditivo, os valores contidos no banco de dados para testagem não podem ter sido usados para o treinamento (KABUGO *et al.*, 2020; KENJI, 2019), conforme podemos observar na Figura 15.

Figura 15: Separação de um banco de dados em dados para treinamento (*training set*), dados para testagem do modelo (*test set*) e dados para validação do modelo (*validation set*), com uma estratégia de validação denominada *cross-validation*.



Fonte: site 3 Dimensões ([S.d.]).

Após a definição dos *Sets* é feito o treinamento do modelo preditivo. Como dito anteriormente, existem vários modelos que podem ser divididos em três grupos de acordo com o tipo de aprendizado, cada um com suas características. Da mesma forma diversas linguagens de programação podem ser usadas para realizar o treinamento do

modelo, como python e R, além de *softwares* como o Matlab. O cientista de dados define qual plataforma será utilizada. O resultado do treinamento é chamado modelo, uma *Application Programming Interface*.

Definidos os modelos e a linguagem utilizada é necessário utilizar um *script*, nele o banco de dados será carregado, os parâmetros serão definidos e as bibliotecas necessárias para os algoritmos serão adicionadas. O modelo será aplicado ao conjunto de dados para teste (*training set*), ao conjunto de dados para validação (*validation set*) e ao conjunto de dados para testagem (*test set*). Os resultados serão plotados como uma matriz de confusão e outros parâmetros obtidos através dessa matriz.

A matriz de confusão, ou Tabela de Confusão (Figura 16), é uma matriz de tamanho $n \times n$, onde n representa a quantidade de classes que se deseja prever. É voltada para modelos de classificação com o objetivo de calcular a quantidade de falsos positivos (FP), falsos negativos (FN), verdadeiros positivos (TP), verdadeiros negativos (TN), acurácia, taxa de erro, sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN). A acurácia é a proporção de predições feitas corretamente em relação ao total de predições. A taxa de erro é a quantidade de predições feitas de maneira errônea sobre o total de predições realizadas. A sensibilidade é a proporção de TP na população de positivos reais. A especificidade é proporção de VN na população de negativos reais. O VPP é a proporção de TP na população de indivíduos positivos preditos. Por fim o VPN é a proporção de TN na população de indivíduos negativos preditos (NOGARE, 2020; THARWAT, 2018; VASCONCELLOS, 2018).

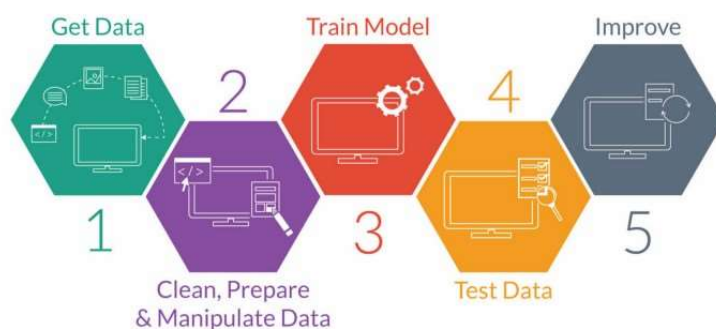
Figura 16: Esquema de uma matriz de confusão 2x2. A categoria positivo e negativo é determinada pelo cientista de dados.

		Valor real	
		Positivo (P)	Negativo (N)
Valor predito	Verdadeiro (T)	Verdadeiro positivo (TP)	Falso positivo (FP)
	Falso (F)	Falso negativo (FN)	Verdadeiro negativo (TN)

Fonte: Adaptado de Tharwat (2018)

Todos os passos comuns para a aplicação de um modelo preditivo abordados podem ser vistos na Figura 17. A coleta dos dados que serão avaliados, seguido da limpeza e preparação do banco de dados, etapa mais demorada e que exige maior atenção durante a aplicação de modelos de predição. O treinamento do modelo, através de diversas técnicas e a aplicação de uma estratégia de validação (testagem). Por fim a aplicação do modelo.

Figura 17: Etapas para a aplicação de modelos preditivos.



Fonte: Oliveira, Neto, (2018)

2.5.2 – Modelos preditivos

Vários modelos preditivos podem ser utilizados para classificação quando trabalhamos com um aprendizado supervisionado. Nessa revisão iremos apresentar apenas os modelos que iremos utilizar.

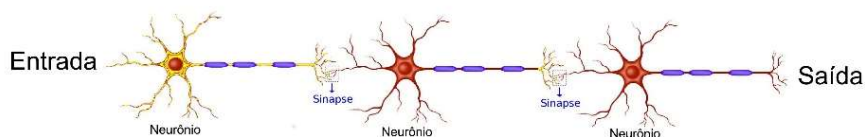
Os *Generalized Linear Model* (GLM), são uma adaptação dos modelos de regressão, com o objetivo é obter uma generalização desses modelos, portanto os GLM admitem mais suposições, deixando de analisar apenas as relações lineares entre as variáveis explicativas e a resposta. Para um GLM deve-se definir a função de ligação que irá conectar as variáveis explicativas com a resposta, diversas equações podem ser utilizadas, cada uma com suas características. Uma das equações utilizadas é a *Elastic Net*, a grande vantagem dessa equação em relação as demais é permitir uma efetiva regularização, aonde a regressão seleciona os coeficientes impondo uma restrição. Isso é uma vantagem principalmente em grupos de preditores correlacionados, como ocorre bovinos a pasto (LI; LINY, 2010; TAY; NARASIMHAN; HASTIE, 2021).

A *Artificial Neural Network* (ANN), ou Rede neural artificial, é uma das mais conhecidas técnicas de *machine learning* da atualidade. Esse método se baseia nas redes

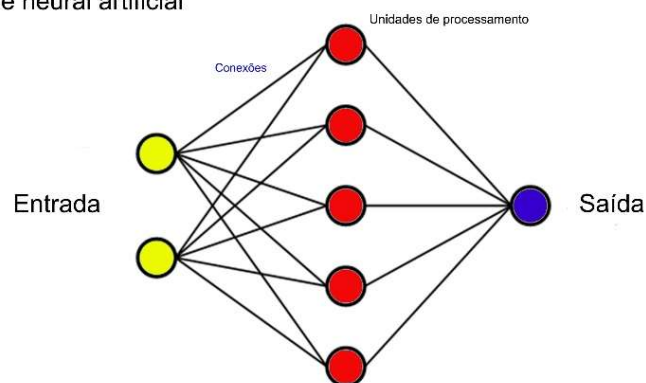
neurais dos animais, onde os neurônios recebem algum dado de entrada, ou *input*, e retorna um valor de saída, *output*. Nos animais os *inputs* são os estímulos vindos do ambiente através de algum “sensor”, como as papilas gustativas, ou outro neurônio, via dendritos, e emitem uma saída por meio dos axônios. O objetivo dessa técnica é simular o cérebro dos animais, inclusive em seu comportamento de aprendizado. As ANNs possuem unidades de processamento, que por sua vez possui ligação com outras unidades de processamento, cada unidade possui uma capacidade de armazenamento e também é atribuído um peso as decisões. Por fim somam-se as entradas e retorna uma saída, como pode ser observado na Figura 18 (DÓREA *et al.*, 2018; JAMES *et al.*, 2013; MANICA, 2013). As ANNs permitem ajustes, como no número de camadas, no peso dado as conexões, quantas conexões serão realizadas, e outros ajustes, aumento sua complexidade (JUNIOR BERTOZZO, 2019).

Figura 18: A) Esquema de uma rede neural biológica, com os neurônios e suas sinapses. B) Esquema de uma rede neural artificial com suas unidades de processamento e conexões.

A) Rede neural biológica



B) Rede neural artificial

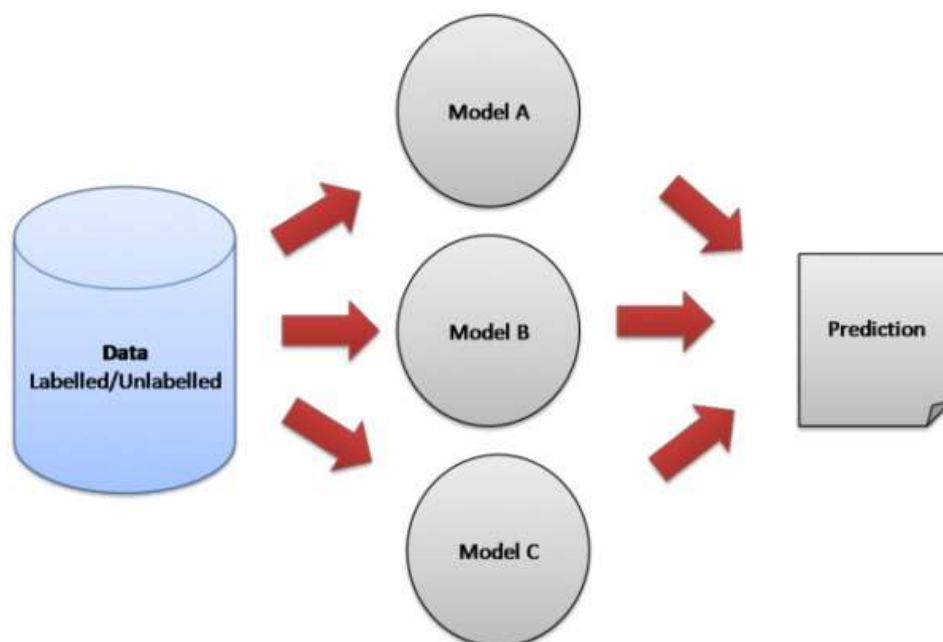


Fonte: Do autor (2021).

Decision Tree (DT) é um modelo preditivo baseado em estruturas semelhantes a árvores. Nesse modelo vários pontos de decisão são criados, como os nós de uma árvore. Em cada nó a decisão seguirá por um caminho ou por outro, simulando os ramos da árvore. Esse modelo é utilizado para prever o comportamento animal com base em uma sequência de perguntas e respostas de sim ou não. (JAMES *et al.*, 2013; KAMIŃSKI; JAKUBCZYK; SZUFEL, 2018). Esse modelo preditivo não será implementado nesse experimento, mas serve como base para outros modelos preditivos utilizados.

Antes de dar continuidade aos próximos dois modelos é necessário explicar o conceito de *Ensemble Learning* (EL), pois os próximos são modelos baseados nessa metodologia. Em uma pesquisa de opinião, por exemplo, não pegamos apenas o resultado “mais adequado” e sim a combinação de várias respostas. O princípio do EL é semelhante, consiste na junção de diversos modelos considerados mais fracos, como DT e *k-nearest neighbors*, com o objetivo de tornar o modelo mais robusto, mais confiável e minimizar as desvantagens desses modelos “mais fracos” (DUARTE, 2020; SAGI; ROKACH, 2018). O banco de dados é utilizado para a predição em diversos modelos que, juntos, retornam um resultado mais acurado (Figura 19).

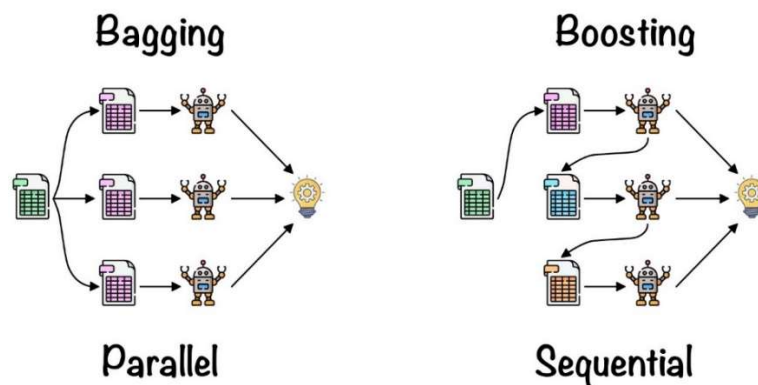
Figura 19: Esquema de um modelo *Ensemble Learning*, com o banco de dados, a utilização de vários modelos e a geração da predição.



Fonte: DUARTE (2020).

O EL pode ser dividido em várias categorias dependendo do método utilizado, mas aqui falaremos apenas dos métodos *Bagging* e *Boosting*. O *Bagging* seleciona diversos conjuntos de dados para treinamento individuais dentro do banco de dados, de maneira aleatória, e faz o treinamento, no final os resultados são agregados. Já o *Boosting* seleciona os modelos de maneira sequencial, ajustando o modelo treinado anteriormente, com maior foco nas observações que os modelos anteriores erraram mais (DUARTE, 2020; LÓPEZ, 2021). Ambas as metodologias de *ensemble learning* podem ser observadas na Figura 20.

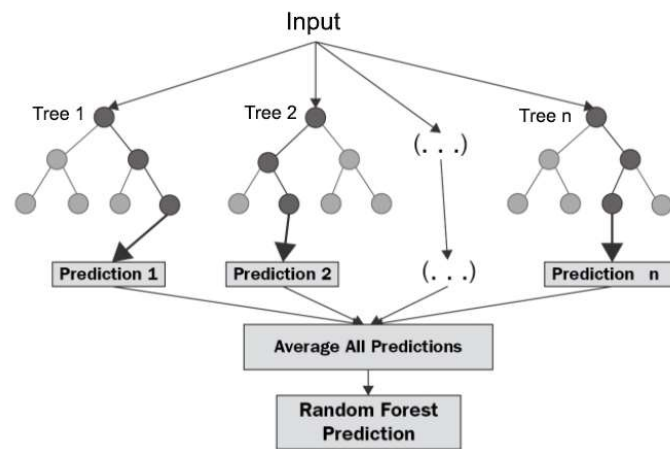
Figura 20: Esquema de um modelo *bagging* (esquerda) com o banco de dados sendo dividido aleatoriamente em vários conjuntos de dados para treinamento e os resultados sendo agregados. Esquema de um *boosting* (direita) com os treinamentos sequenciais e resultados sendo agregados.



Fonte: López (2021).

Retornando aos modelos preditivos. O *Random forest* (RF) é um modelo de *Bagging*, tem como objetivo principal reduzir o problema de *overfitting* presente no DT. Para isso várias árvores de decisões são criadas de maneira aleatória, simulando uma floresta, ou seja, uma floresta randômica. O resultado de várias árvores é agregado em um resultado final. As árvores serão diferentes entre si, pois a seleção das amostras e a seleção das variáveis acontece de maneira aleatória, como podemos observar na Figura 21 (AMIT; GEMAN, 1994; VALLETTA *et al.*, 2017).

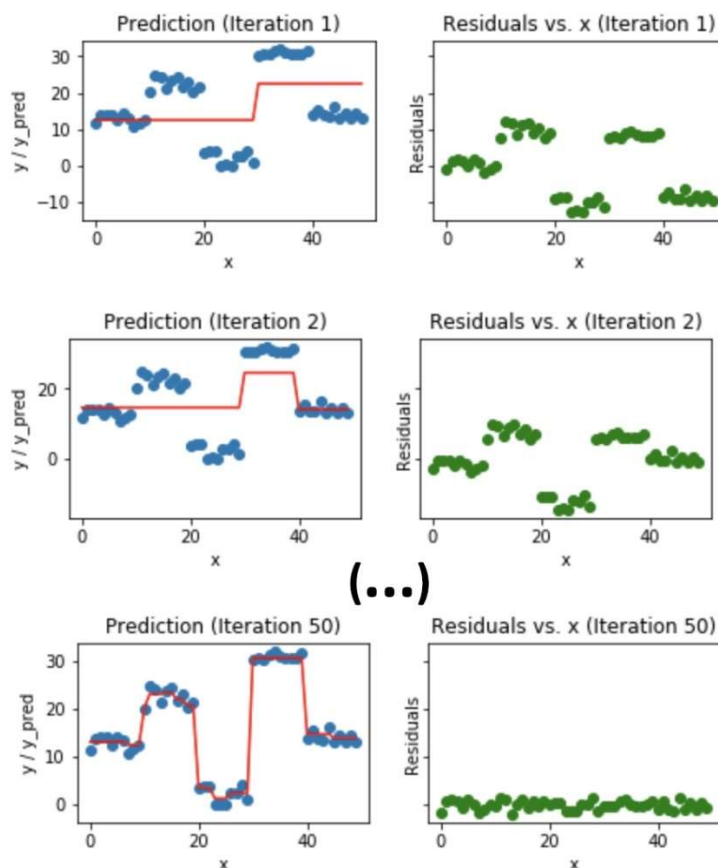
Figura 21: Imagem esquemática de um modelo preditivo de *Random Forest* com “n” *decision trees*.



Fonte: Bakshi (2020).

O *Gradient Boosting* (GB), como o nome diz, é um modelo de *boosting* que usa diversos “modelos fracos” para realizar sua classificação. É um modelo preditivo que treina novos modelos diretamente no erro dos modelos anteriores, ou seja, o primeiro modelo faz uma aproximação da predição e gera alguns erros residuais, um novo treinamento é realizado nesses erros residuais para tentar prever o erro do primeiro modelo, e assim por diante. Ao final obtemos uma “versão” mais corrigida da primeira predição (FRIEDMAN, 2001; SAGI; ROKACH, 2018). A Figura 22 exemplifica esse método, nele podemos ver os modelos preditores (Esquerda), com os valores preditos (Linha vermelha) e os valores reais (Azul), também podemos observar os erros residuais (Esquerda). Temos uma diminuição nos erros residuais à medida que mais modelos são aplicados, chegando a 50 modelos no exemplo, bem como uma melhora nos valores preditos.

Figura 22: Modelo esquemático de um *gradient boosting* com 50 predições. Mostrando os modelos preditivos (Esquerda) e erros residuais (Direita) das predições 1, 2 e 50.



Fonte: Adaptado de Grover (2017).

Os modelos RF, GB e ANN foram selecionados por serem resistentes a ruídos, comuns em bancos de dados com base na pecuária, e por sua capacidade de aprendizado rápido. Enquanto o GLM é um modelo extremamente utilizado e mais antigo, sendo bastante resistente a ruídos quando comparado a outros modelos lineares.

2.6 – Validação

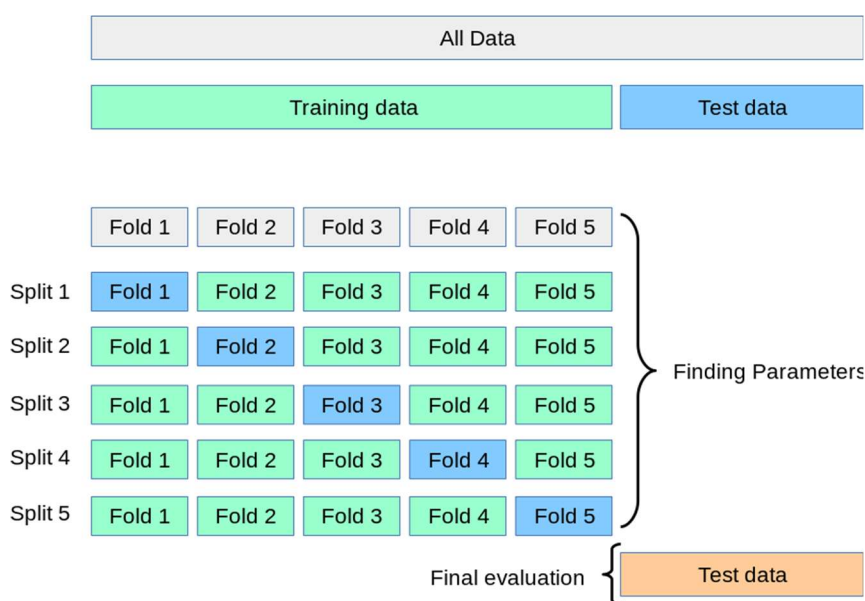
A validação cruzada é uma das técnicas mais comuns para validar modelos no comportamento animal. Essa técnica consiste no particionamento do banco de dados em subconjuntos e o uso de alguns desses subconjuntos para o treinamento do modelo, enquanto o restante é utilizado para a validação. Dentre os vários métodos de validação

cruzada podemos destacar o *Holdout*, *k-fold Cross-validation* e o *Leave-one-out* (HAMILTON *et al.*, 2019; HASTIE; TIBSHRANI; FRIEDMAN, 2008).

Na validação *Holdout* os dados de validação são obtidos aleatoriamente dentro do banco de dados. Diversas proporções podem ser utilizadas, como a 80:20, onde são selecionados 20% dos dados para a validação e 80% são destinados para compor o banco de dados de treinamento. Um problema desse método é que, apesar de se esperar uma boa representatividade do banco de dados total, pode ser que o banco de dados selecionado para validar o modelo não seja representativo do total, sendo necessários técnicas para evitar tal problema (BUNKER; THABTAH, 2019; DWORK *et al.*, 2015).

O método *k-fold Cross-validation* divide o conjunto de dados utilizado para treinamento em k vezes, gerando k subconjuntos mutuamente exclusivos e de mesmo tamanho. Então o modelo usará $k-1$ partes para o treinamento, e usará 1 parte para formar o banco de dados para validação. Esse processo é realizado k vezes, dessa forma todo os dados serão usados para o treinamento e para a validação (Figura 23). Ao final do processo o modelo treinado é submetido a uma testagem com o banco de dados de teste (KOHAVI, 1995; SCIKIT-LEARN.ORG, [S.d.]; VASCONCELLOS, 2018).

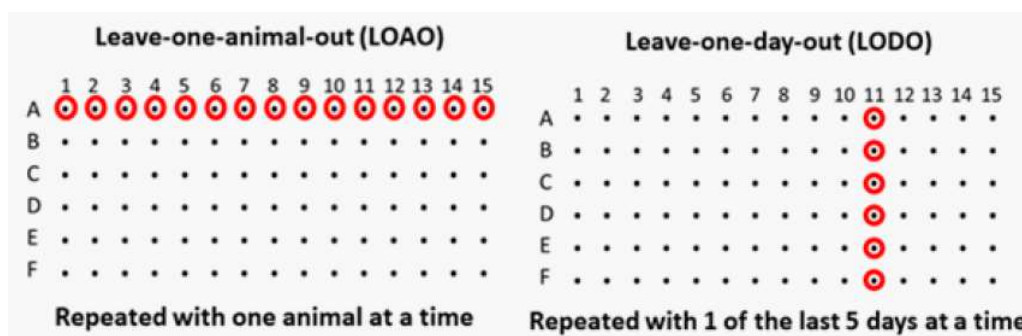
Figura 23: Esquema de uma validação *k-fold Cross-validation*, com a separação do banco de dados (*all data*) em banco de dados para treinamento (*training data*) e para teste (*test data*) e a separação dos dados de treinamento em cinco subconjuntos (Folds).



Fonte: Scikit-learn.org ([S.d.]).

A validação *Leave-one-out* é um caso específico do método de *k-fold Cross-validation*, onde k é semelhante a quantidade total de dados (REFAEILZADEH; TANG; HUAN, 2008). Essa técnica de validação pode ser aplicada de várias formas distintas. Ribeiro, Leonardo Augusto Coelho (2020) utilizaram duas metodologias dessa técnica, a *Leave-one-animal-out* (LOAO), onde os dados de um animal foram retirados para compor o banco de dados destinados a validação, enquanto o restante dos dados foi utilizado como banco de dados para treinamento. E a metodologia *Leave-one-day-out* (LODO), onde os dados de um dia, ou seja, uma altura de rebaixamento, foram retirados para compor o banco de dados para validação, e o restante foi utilizado como banco de dados para treinamento (Figura 24). Essa técnica pode ser aplicada de diversas maneiras, como separar para o *validation set* os dados referentes a estação do ano, índice de pluviosidade, altura do pasto, entre outros.

Figura 24: Duas estratégias de validação *leave-one-out*. LOAO, ao selecionar um animal para a validação e LODO, ao selecionar um dia do rebaixamento para compor a validação.



Fonte: Ribeiro, Leonardo Augusto Coelho (2020).

3 - MATERIAL E MÉTODOS

3.1 Animais e período experimental

O experimento foi conduzido em área pertencente ao Departamento de Zootecnia da Universidade Federal de Lavras, em Lavras, MG, localizada a 21°14'04.5"S de latitude, 44°58'08.6"W de longitude e 918 metros de altitude, entre os meses de novembro de 2019 e julho de 2020, totalizando nove meses de observações e três estações do ano: primavera, verão e outono. Todos os procedimentos foram aprovados pelo comitê de ética para uso de animais, protocolo 016/2018.

A área experimental foi constituída por 0,8 ha de pastagem, com consórcio de *Urochloa brizantha* (Hochst ex A. Rich) Stapf cv. Marandu e *Arachis pintoi* Krap. & Greg. cv. Mandobi, formada no ano de 2007. A área foi dividida em 12 unidades experimentais (UE), distribuídas aleatoriamente entre três tratamentos experimentais. O método de pastejo adotado foi lotação rotativa com altura de entrada dos animais quando o dossel atingisse 25 cm de altura. Os tratamentos experimentais foram três intensidades de desfolhação: leve, moderada e severa, com altura alvo de resíduo de 20 cm, 15 cm e 10 cm, respectivamente, equivalente a 20%, 40% e 60% de desfolhação em relação à altura de entrada. As medições da altura de entrada foram realizadas semanalmente, com a ajuda de uma régua, já as alturas de saída foram realizadas um dia após a saída, pela manhã.

O pastejo foi realizado por duas novilhas da raça Tabapuã com peso médio inicial de aproximadamente 250 kg. Quando necessário, foram adicionados mais animais semelhantes, da mesma raça e peso similar, para auxiliar no rebaixamento do pasto. Cada UE foi dividida em três piquetes onde os animais permaneceram por três dias em cada piquete, conseqüentemente nove dias de permanência em cada UE (Figura 25). Para garantir diferentes intensidades de desfolhação com taxas de lotação semelhantes, os piquetes tinham tamanhos distintos. Os piquetes do tratamento de desfolhação leve, moderada e severa tinham 4.320 m², 2.640 m² e 1.920 m², respectivamente. Os dois primeiros piquetes foram destinados para adaptação dos animais à estrutura do pasto, totalizando seis dias de adaptação, enquanto as avaliações foram realizadas no último piquete, ou piquete de avaliação.

Figura 25: Distribuição dos três tratamentos em 12 unidades experimentais (de 1 a 12) e sua distribuição pela área experimental, cada UE foi dividida em três piquetes.



Fonte: Do autor (2021).

As observações visuais do comportamento ingestivo foram realizadas no primeiro e no último dia do piquete de avaliação, obtendo assim o comportamento ingestivo no início do rebaixamento (quando a altura do dossel era 25 cm), e no final do rebaixamento, quando a altura do dossel era próxima das alturas residuais de 20, 15 e 10 cm, dependendo do tratamento adotado na UE. As observações visuais ocorreram por um período de 12 horas para cada dia de observação, iniciando às 06 horas e encerrando às 18 horas.

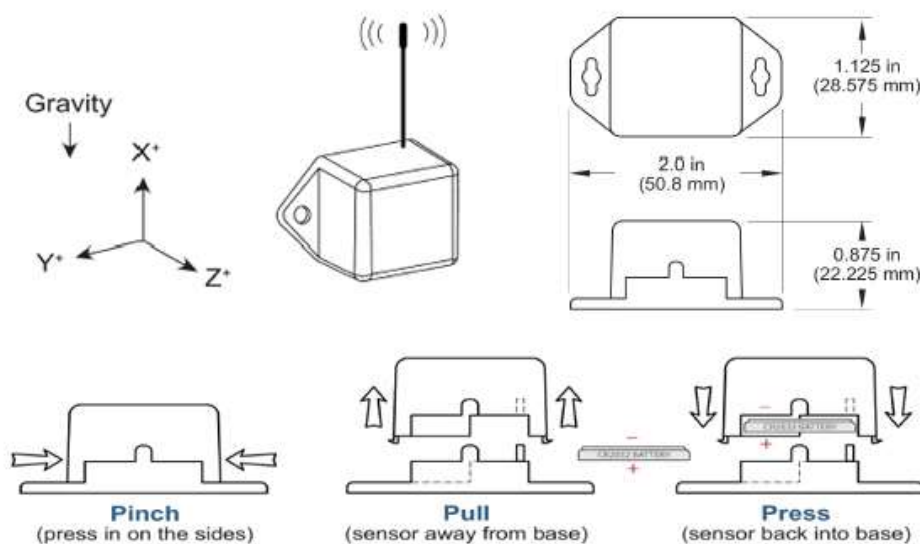
Todas as observações visuais foram anotadas em planilha manual e posteriormente transferidas para uma planilha no Excel. Foi registrado o horário de início das atividades de pastejo, ócio, ruminando, locomovendo, ingerindo água e outros. O objetivo foi adquirir observações visuais contínuas, com o momento exato em que os animais alteram seu comportamento, e o tempo que permaneceram em cada

atividade. Foram anotados hora, minutos e segundos com o auxílio de relógios de pulso sincronizados com o horário dos sensores.

3.2 Sensor e software

O sensor utilizado foi o acelerômetro de 3 eixos da empresa MONNIT (MONNIT, Salt Lake City, Utah, Estados Unidos da América) gravando a uma taxa de 800 Hz, utilizando uma bateria de lítio CR2032 (Figura 26). Os sensores possuíam a identificação 434110, 434136, 434137, 434138 e 434139. O *software* iMonnit (MONNIT, Salt Lake City, Utah, Estados Unidos da América) foi utilizado para receber e exportar os dados brutos em formato CSV.

Figura 26: Representação dos sensores utilizados, suas dimensões e seu funcionamento. Orientação dos eixos.

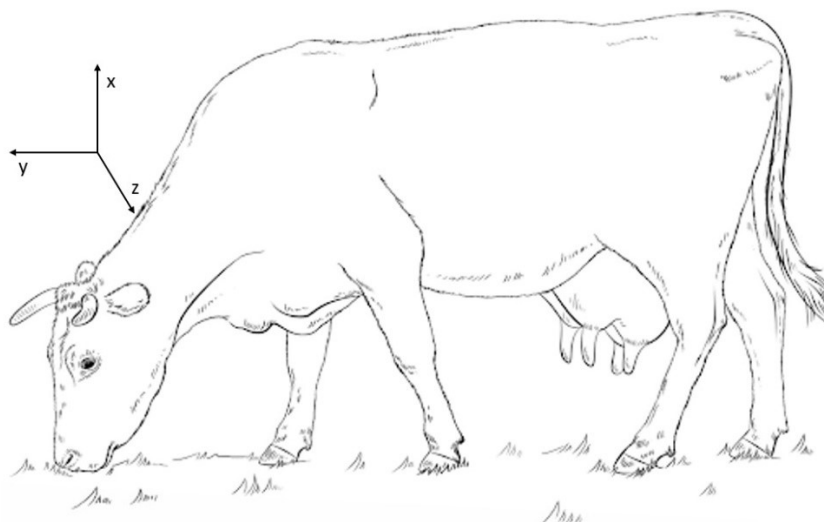


Fonte: MONNIT® (2021).

O acelerômetro utilizado nesse trabalho é capaz de registrar a aceleração em três eixos, sendo eles X, Y e Z, o primeiro é o eixo vertical, que registra a gravidade, enquanto os dois últimos são os eixos horizontais (Figura 27). O sensor foi fixado na nuca dos animais, no cabresto, possibilitando a identificação da movimentação da cabeça enquanto executa diferentes comportamentos, além da facilidade de fixação e remoção dos sensores. Os sensores eram sempre posicionados com a antena para a esquerda, mantendo a mesma angulação. Para proteger os sensores de intempéries eles

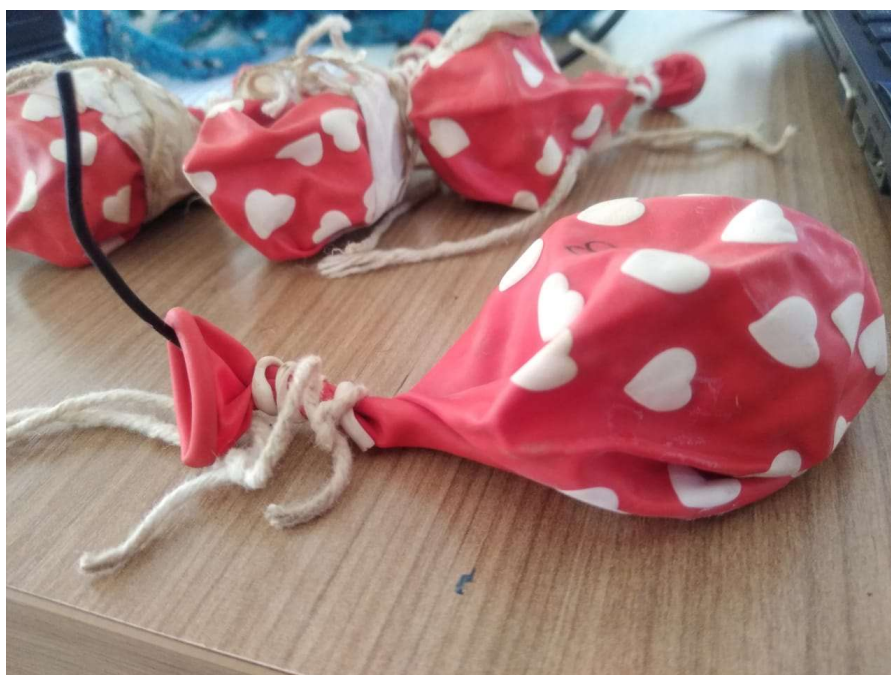
foram colocados dentro de bexigas N 10, e amarrados utilizando um fecho de arame amarelo, barbante e fita crepe (Figura 28).

Figura 27: Representação da fixação dos sensores em bovinos e os eixos capturados pelo mesmo, eixo X (vertical), Y e Z (horizontais).



Fonte: Do autor (2021).

Figura 28: Sensores antes da utilização. Já protegidos pelo balão e presos com amarelo, barbante e fita. Todo o conjunto era fixado no cabresto com auxílio de barbante e fita.

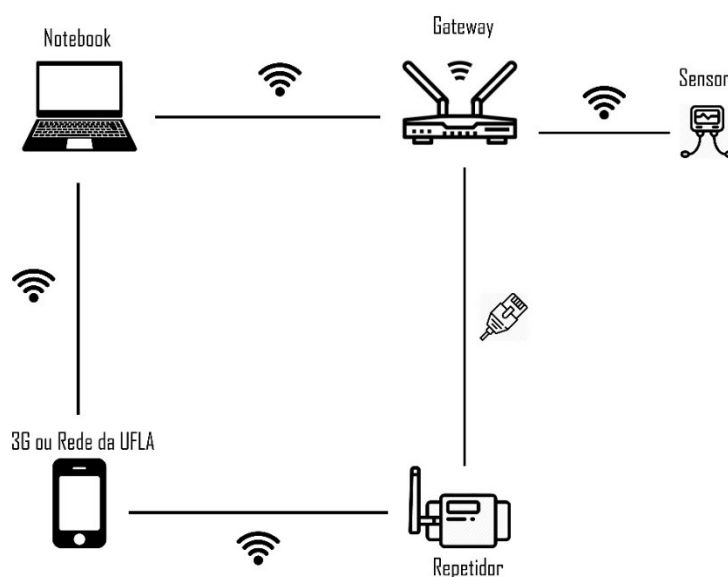


Fonte: Do autor (2021).

Os sensores e o *software* necessitam de uma conexão constante com a internet, através de uma conexão *ethernet*. Portanto foi necessário utilizar a rede móvel de um celular e um repetidor para transformar o sinal *wifi* em sinal *ethernet*.

Um notebook foi utilizado para suportar o *software* dos sensores, para seu funcionamento utilizamos uma extensão de 200 metros conectada a uma tomada presente em um poste próximo, levando todo o hardware para perto do piquete em observação. Um celular foi utilizado como roteador de rede 3G, levando sinal *wi-fi* para o computador e para o repetidor. O repetidor captava esse sinal *wi-fi* e o transmitia em sinal *ethernet* (a cabo) que, posteriormente, era conectado no *gateway* próprio da empresa MONNIT (MONNIT, Salt Lake City, Utah, Estados Unidos). O *gateway* foi responsável por captar os sinais transmitidos pelos sensores e enviar para o computador, onde era lido pelo *software* iMonnit. Essa configuração pode ser vista na Figura 29.

Figura 29: Imagem esquemática do funcionamento dos sensores no pasto. As linhas representam as conexões, tanto *wi-fi* quanto via cabo.



Fonte: Do autor (2021).

Os sensores, o computador e os relógios eram preparados antes das observações e desmontados ao final da mesma. Para auxiliar na fixação dos sensores os animais eram direcionados a um tronco de contenção presente no local.

3.3 – Análise exploratória dos dados

O objetivo da análise exploratória é resumir, compreender e explorar as características básicas de um banco de dados. Em uma planilha foram adicionadas as informações de data da observação, piquete, altura esperada, altura observada, animais, sensor utilizado, tempo da observação, recebimento de dados e estação do ano. Foram utilizadas medidas de tendência central e de dispersão que pudessem demonstrar a eficiência de envio de dados, as alturas reais mensuradas e a distribuição das observações entre os animais utilizados, as alturas esperadas, os sensores utilizados e as estações do ano.

Outra análise exploratória realizada foi a porcentagem do tempo gasto pelos animais em cada comportamento ao longo do rebaixamento do dossel. Os comportamentos foram separados em: pastejando, ruminando, ócio e não pastejando, junção dos tempos em ócio e ruminando. Foi feita uma média do tempo gasto nas atividades das repetições dos tratamentos, 25, 20, 15 e 10 cm.

3.4 - Modelos preditivos

Antes de desenvolver os modelos preditivos foi necessário preparar o banco de dados, ou seja, realizar a limpeza dos dados e unir todas as planilhas em uma única. Para tal foi utilizado a linguagem R. As planilhas contendo os dados dos sensores e as observações visuais foram unidas utilizando como parâmetro a data da observação, o horário em segundos e o sensor utilizado, todas as linhas faltantes e incompletas foram removidas. O banco de dados contém as informações da identificação do sensor, a data e a hora, os dados brutos da aceleração nos três eixos, a identificação dos animais, o comportamento observado, a identificação do piquete e a altura esperada para o piquete. Os dados referentes aos comportamentos ingerindo água, locomovendo e outros foram excluídos, pois representavam uma fração muito pequena do banco de dados e também apresentaram padrões de movimento no sensor que não se encaixavam em outras categorias.

O primeiro objetivo do trabalho foi comparar alguns modelos preditivos e diferentes técnicas de validação. Para isso foram criados dois conjuntos de dados a partir do banco de dados, modificando a coluna “comportamento”. No primeiro, denominado PRO, foram observados três comportamentos: pastejando, ruminando e

ócio. Já no segundo, denominado PNP, foram observados três comportamentos: pastejando e não pastejando, que é uma junção dos comportamentos de ruminando e ócio.

Os modelos preditivos selecionados foram: *generalized linear model* (GLM), *artificial neural network* (ANN), *random forest* (RF) e *gradient boosting* (GB). Os melhores hiper parâmetros ou arquitetura para os modelos foram definidos utilizando uma pesquisa de grade aleatória para as abordagens de aprendizado de máquina.

Os modelos GLM são uma generalização da regressão linear, assim são capazes de analisar relações não lineares entre as variáveis explicativas e a resposta. Diversas equações podem ser utilizadas nos modelos GLM, a que utilizamos é denominada *Elastic Net*, pois é uma equação que admite grupos de preditores correlacionados.

O ANN é um modelo preditivo muito conhecido atualmente, o objetivo desse modelo é simular as redes neurais dos animais. Sua forma de aprendizado garante uma enorme versatilidade, além de permitir relações complexas entre os dados de entrada (*input*) e a resposta (*output*). Outras vantagens são sua auto-organização e a tolerância a falhas e ruídos. Porém é um modelo que exige maior poder computacional, principalmente em redes mais complexas. A função de ativação (*Activation function*) traz para as redes neurais um componente de não linearidade, fazendo com que elas aprendam mais do que as relações lineares entre as variáveis dependentes e independentes, atribuindo pesos as relações. As camadas ocultas (*Hidden layers*) são as camadas que realizam cálculos intermediários, entre a camada de entrada e a camada de saída, auxiliando a rede a encontrar os *outputs*, cada camada possui um número definido de neurônios (*Neurons per layer*), que de fato executam os cálculos. A taxa de abandono (*dropout ratio*) é a taxa com que o modelo abandona, aleatoriamente, unidades na rede neural visando prevenir o sobreajuste dos dados de treinamento (*overfitting*), que é a perda da generalização do modelo por se tornar muito específico. Por fim temos a regularização em L1 e L2, que também auxilia a evitar *overfitting*. A regularização L1 (*L1 regularization*) produz modelos mais robustos reduzindo os ruídos presentes no modelo, já a regularização L2 (*L2 regularization*) estabiliza o modelo quando há correlação entre os *inputs*. Tais valores variam conforme a técnica de validação utilizada.

Já os modelos preditivos RF e GB são classificados como *Ensemble Learning* (EL), técnica que une diversos modelos considerados mais fracos, como o *decision tree* (DT), em um modelo mais robusto, que minimize as desvantagens desses modelos mais fracos. O EL pode ser dividido em várias categorias, mas utilizamos duas: *bagging* e *boosting*. Os modelos classificados como *bagging*, como o RF, selecionam diversos subconjuntos de dados para treinamento (*training set*) individuais, de maneira aleatória, e realiza diversos treinamentos, ao final os resultados de todos os treinamentos são agregados em um único resultado. Os modelos preditivos classificados como *boosting*, um exemplo é o GB, selecionam os modelos mais fracos de maneira sequencial, ajustando os modelos com base nos erros dos modelos anteriores, ao final temos um modelo ajustado para ter o menor erro ao prever a resposta.

O RF é um modelo de *bagging*, que utiliza o DT como base, seu objetivo é reduzir o *overfitting* presente no DT. Várias árvores (*number of trees*) são criadas de maneira aleatória, deixando cada DT diferente da outra, e seus resultados são agregados em um resultado final. A profundidade máxima (*maximum depth*) é uma maneira de determinar o quão profunda uma árvore pode ser, seu objetivo é evitar *overfitting* nas DTs criadas. O *minimum row* define a quantidade mínima de conexões que devem ser criadas, deixando a resposta mais precisa. O número de bins (*number of bins*) define a quantidade de bins presentes no modelo.

O GB é um modelo preditivo de *boosting*, assim como RF utiliza DT como base para realizar suas classificações. Sua principal vantagem é criar um novo modelo em cima do erro do anterior, reduzindo erros por ruídos no banco de dados e erros de classificação. Cada nova DT é treinada em cima dos erros residuais encontrados pela DT anterior. Assim como RF precisamos definir a quantidade de árvores utilizadas, a profundidade máxima das árvores, o *minimum row* e o número de bins. Também definimos a taxa de amostragem, histograma e taxa de amostragem das colunas.

3.4 – Estratégias de validação

Diferentes estratégias de validação cruzada também foram avaliadas, utilizando conjuntos de dados para validação escolhidos aleatoriamente (*holdout*) ou de acordo com algum critério específico, como descrito a seguir.

A primeira foi o *holdout*, na qual um subconjunto de dados é aleatoriamente retirado do banco de dados e utilizado na validação dos modelos. A proporção de dados usados para treinamento e de dados utilizados para validação pode variar, conforme interesse do cientista de dados. Foi decidido utilizar a porcentagem 80:20.

As outras três estratégias de validação cruzada buscaram remover algum tipo de interdependência biológica existente entre observações, evitando que observações relacionadas estejam tanto no conjunto de treinamento quanto no de validação. Sendo assim, na estratégia *leave-animals-out* (LAO), todos os dados de dois animais, selecionados aleatoriamente, foram removidos do conjunto de treinamento e utilizados para validação. Na estratégia *leave-days-out* (LDO), todas as observações de sete dias, escolhidos aleatoriamente, foram excluídas do conjunto de treinamento e utilizadas para validação. Por fim, a estratégia *leave-one-height-out* (LHO) foi utilizada nas duas alturas extremas testadas no experimento, 10 cm (LHO10) e 25 cm (LHO25), nas quais todas as observações coletadas em piquetes com essas alturas alvo foram excluídas do conjunto de treinamento e utilizados para validação.

Todos os modelos preditivos e as estratégias de validação utilizados em ambos os bancos de dados, PRO e PNR, podem ser vistas na Tabela 1.

Tabela 1: Combinações entre os modelos preditivos de classificação e estratégias de validação utilizados.

		Modelos preditivos			
		RF / <i>Holdout</i>	ANN / <i>Holdout</i>	GB / <i>Holdout</i>	GLM / <i>Holdout</i>
		RF / LAO	ANN / LAO	GB / LAO	GLM / LAO
Validação		RF / LDO	ANN / LDO	GB / LDO	GLM / LDO
		RF / LHO10	ANN / LHO10	GB / LHO10	GLM / LHO10
		RF / LHO25	ANN / LHO25	GB / LHO25	GLM / LHO25

Legenda: *Leave-animals-out* (LAO), *leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Holdout* (20% dos dados selecionados para testagem), *Leave -height-out* 10 cm (LHO10) e *Leave-height-out* 25 cm (LHO25).

Fonte: Do autor (2021).

As métricas utilizadas para avaliar os modelos preditivos e as validações são obtidas após a plotagem da matriz de confusão. Essa matriz calcula a porcentagem de vezes que o modelo preditivo conseguiu prever como verdadeiro algum valor que de

fato era verdadeiro, ou verdadeiro positivo (TP), a porcentagem de vezes que o modelo conseguiu prever como falso algum valor que de fato era falso, ou verdadeiro negativo (TN), a quantidade de vezes que o modelo predisse como verdadeiro um dado que na realidade era falso, ou falso positivo (FP), e a quantidade de vezes que o modelo predisse como falso um dado que na realidade era verdadeiro, ou falso negativo (FN). As métricas utilizadas foram: Acurácia, taxa de erro, sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN).

A acurácia é a proporção de predições feitas corretamente em relação ao total de predições, pode ser obtida pela equação 1.1. A taxa de erro é a quantidade de predições feitas de maneira errônea sobre o total de predições realizadas, pode ser obtida pela equação 1.2. A sensibilidade é a proporção de TP na população de positivos reais, obtida pela equação 1.3. A especificidade é proporção de VN na população de negativos reais, obtida pela equação 1.4. O valor preditivo positivo (VPP) é a proporção de TP na população de indivíduos positivos preditos, obtido pela equação 1.5. Por fim o valor preditivo negativo (VPN) é a proporção de TN na população de indivíduos negativos preditos, obtido pela equação 1.6.

$$1.1 \text{ Acurácia} = \frac{TP+TN}{TP+TN+FP+F}$$

$$1.2 \text{ Taxa de erro} = \frac{FP+FN}{TP+TN+FP+F}$$

$$1.3 \text{ Sensibilidade} = \frac{TP}{TP+FN}$$

$$1.4 \text{ Especificidade} = \frac{TN}{TN+FP}$$

$$1.5 \text{ VPP} = \frac{TP}{TP+F}$$

$$1.6 \text{ VPN} = \frac{TN}{TN+FN}$$

Outro objetivo do trabalho foi utilizar o presente banco de dados para uma validação externa dos modelos gerados no trabalho anterior do nosso grupo de pesquisa (RIBEIRO *et al.*, 2021). Como no referido trabalho os dados observados foram classificados em apenas duas categorias (pastejando e não pastejando), o mesmo tipo de classificação foi aplicado em nossos dados. Os comportamentos de ócio e ruminando foram unidos na categoria não pastejando.

4 – RESULTADOS

Uma observação do comportamento pode ser definida como o período de 12 horas que cada animal passou sendo observado. Ao todo foram realizadas 96 observações do comportamento ingestivo dos animais, divididos em 35 dias não contínuos. Das 96 observações 47 foram na entrada dos animais (25 cm), 16 nos piquetes de desfolhação leve (20 cm), 13 nos piquetes de desfolhação média (15 cm) e 20 nos piquetes de desfolhação severa (10 cm). As observações foram divididas ao longo de 8 meses, iniciando em novembro de 2019 e encerrando em junho de 2020, abrangendo 3 estações. As médias de precipitação (mm), temperatura a meia noite (°C), temperatura ao meio dia (°C), temperatura as seis da tarde (°C), umidade relativa (%) e velocidade do vento (m/s) dos dias de observações do comportamento ingestivo podem ser vistas na Tabela 2, separados por estação do ano. As informações foram coletadas na estação meteorológica 83687, localizada em Lavras-MG e são disponibilizadas pelo INMET: Instituto Nacional de Meteorologia.

Tabela 2: Média de precipitação, temperatura ao longo do dia e umidade relativa nos dias de observações comportamentais, separados por estação do ano.

	Primavera	Verão	Outono
Precipitação (mm)	2,0	3,6	0,10
Temperatura 00:00 (°C)	21,9	19,6	17,6
Temperatura 12:00 (°C)	22,2	22,5	16,7
Temperatura 18:00 (°C)	25,9	28,1	24,2
Umidade relativa (%)	77,6	72,6	73,5

Fonte: Adaptado do INMET (2020).

Para a validação LDO foram selecionados sete dias aleatoriamente, esses dias foram utilizados para compor o banco de dados de teste enquanto os demais foram utilizados para treinar o modelo. A temperatura registrada meio dia (6 horas do início da observação), a temperatura registrada as 18:00 (12 horas após o início da observação), precipitação, umidade relativa e velocidade do vento registrada ao longo dos sete dias do conjunto de validação podem ser vistos na Tabela 3.

Tabela 3: Precipitação, temperatura máxima e mínima para os sete dias utilizados na validação leave-days-out (LDO).

Data	Precipitação (mm)	Temperatura 18:00 (°C)	Temperatura 12:00 (°C)	Umidade relativa (%)
17/11/2019	0	27,8	20,6	63,0
02/03/2020	17,6	21,4	18,2	83,5
19/03/2020	4,2	23,8	23,2	85,5
21/03/2020	Falha	Falha	Falha	Falha
30/03/2020	0	28,4	22,0	72,5
22/05/2020	0	25,6	17,4	68,5
30/06/2020	0	23,2	16,0	68,0

Fonte: Adaptado do INMET (2020).

Os sensores foram calibrados para enviar os dados de segundo em segundo durante as 12 horas em que as observações visuais estavam sendo coletadas, porém houve uma enorme falha no recebimento dos sinais sem motivo aparente. Na Tabela 4 está demonstrada a porcentagem de falha na transmissão dos dados para cada tratamento, tanto na entrada quanto na saída, e também a média total de falha dos sensores. Já na Tabela 5, podemos observar a porcentagem de falha de cada sensor e o quanto eles foram heterogêneos entre si.

Tabela 4: Porcentagem de falhas no recebimento de dados por intensidade de desfolhação, na entrada e saída dos animais dos piquetes, e a porcentagem total.

Desfolhação severa (10 cm)			Desfolhação média (15 cm)			Desfolhação leve (20 cm)			Média
Entrada	Saída	Total	Entrada	Saída	Total	Entrada	Saída	Total	Total
60%	50%	55%	55%	54%	50%	63%	54%	54%	53%

Fonte: Do autor (2021).

Tabela 5: Comparação entre a porcentagem de falha no recebimento de dados por sensor.

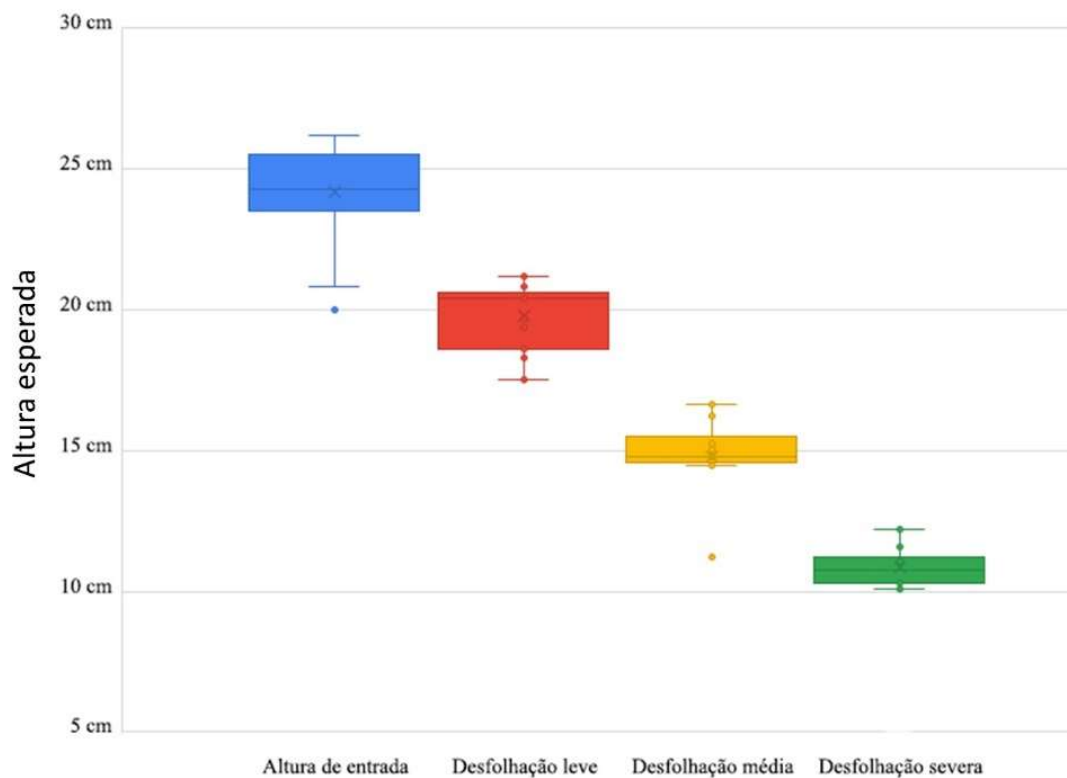
Sensor	Falha	Sensor	Falha
434110	92%	434138	80%
434136	42%	434139	41%
434137	44%		

Fonte: Do autor (2021).

Apesar das falhas observadas, o banco de dados final ainda foi robusto, com mais de 550 mil dados coletados, o que nos possibilitou realizar o treinamento dos modelos preditivos com dois e três categorias de comportamentos distintos. A porcentagem dos dados pertencentes aos comportamentos de pastejo, ruminação e ócio foi semelhante a porcentagem de tempo desprendido em cada uma dessas atividades, registrado pela observação visual, demonstrando que não houve um desequilíbrio do recebimento dos dados que pudesse comprometer as previsões.

Outro ponto analisado foi a altura real e esperada. Ao formular os modelos preditivos utilizamos a altura esperada como padrão, porém a altura real não pode ser muito discrepante da esperada, caso contrário estaremos classificando todo um dia de observações em uma categoria equivocada. Podemos observar que a altura real foi próxima para todos os tratamentos, com média de 10,9 cm para desfolhação severa (10 cm), 14,8 cm para desfolhação média (15 cm), 19,8 cm para desfolhação leve (20 cm) e 24,2 cm para as alturas de entrada (25 cm), com desvio padrão de 0,68; 1,48; 1,16 e 1,57 cm, respectivamente. A Figura 31 apresenta um gráfico de boxplot das alturas observadas em contraste com as alturas alvo. Na imagem podemos observar um *outlier* para a altura de entrada, com aproximadamente 20 cm, e outro para a desfolhação média, com aproximadamente 12 cm.

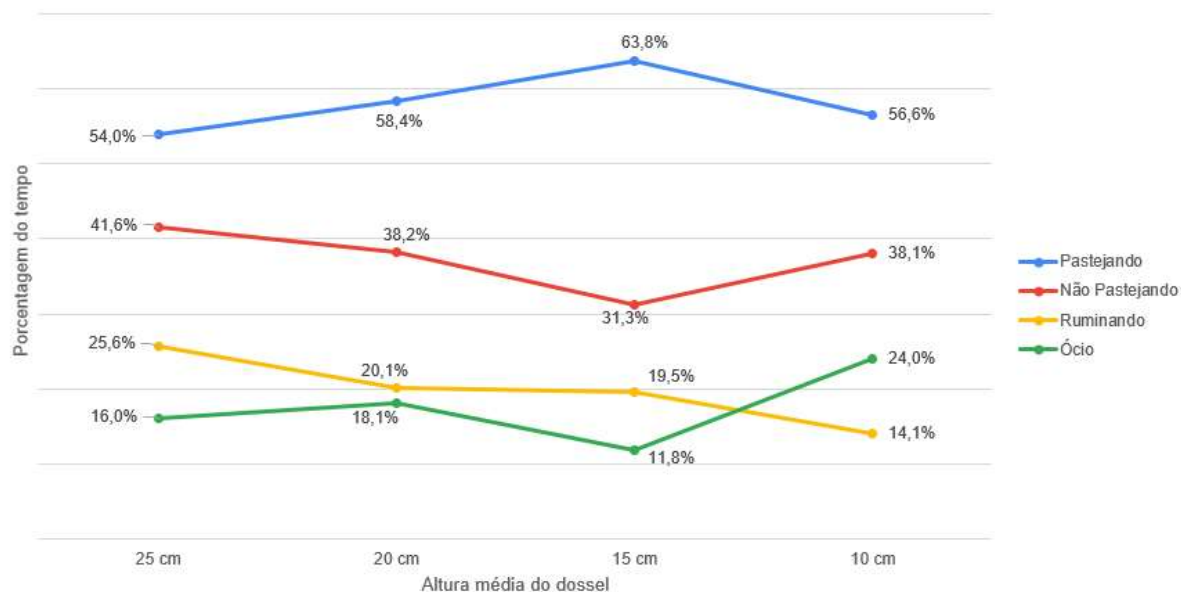
Figura 30: Boxplot comparando a altura real com a esperada para todos os tratamentos, desfolhação leve, média e severa, e altura de entrada (25 cm).



Fonte: Do autor (2021).

O comportamento ingestivo dos bovinos se alterou conforme o rebaixamento do dossel, como pode ser observado na Figura 32. Esses três comportamentos são os mais observados, ocupando 95,6% do tempo na altura de entrada, 96,6% do tempo para o tratamento de 20 cm, 95,1% do tempo para o tratamento de 15 cm e 94,7% do tempo para o tratamento de 10 cm.

Figura 31: Porcentagem do tempo gasto nas atividades de pastejando, ruminando, ócio e não pastejando (ruminando e ócio), para a alturas de 25 cm, 20 cm, 15 cm e 10 cm.

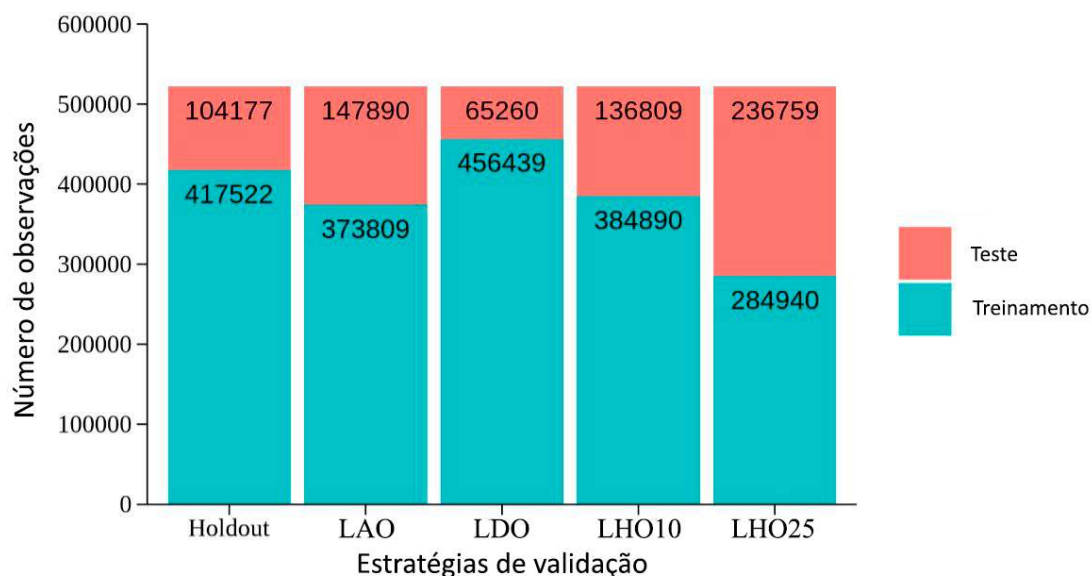


Fonte: Do Autor (2021).

4.1 – Conjunto de dados PNP: Pastejando x Não Pastejando

O total de dados coletados foi de 604.841 dados, após a limpeza do banco de dados restaram 567.940 dados e esses foram usados para compor os bancos de dados. O conjunto de dados PNP continha 309.763 (54,5%) dados do comportamento pastejando e 258.177 (45,5%) do comportamento não pastejando, obtido após a junção dos dados dos comportamentos de ócio e ruminando. A quantidade de dados utilizado para a treinamento e para o teste variou conforme a estratégia de validação, pois diferentes subconjuntos foram retirados para compor o banco de dados para os testes, enquanto o restante ficou no banco de dados para o treinamento (Figura 33).

Figura 32: Divisão do conjunto de dados PNP em subconjunto de treinamento e subconjunto de validação cruzada (teste) para as diferentes estratégias de validação.

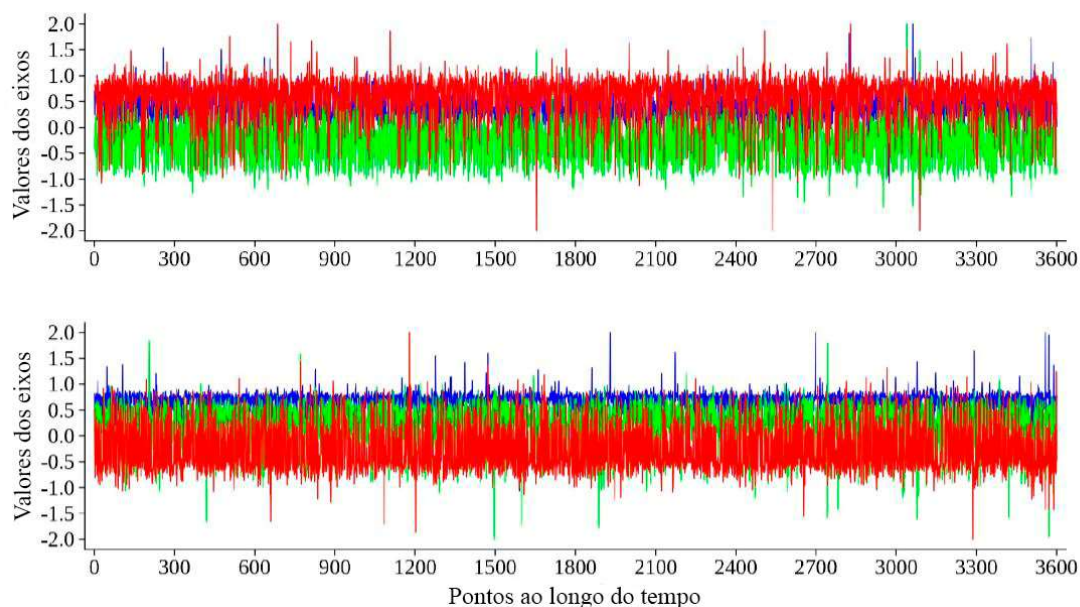


Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do Autor (2021).

Na Figura 34 podemos observar uma amostra da movimentação dos sensores em dois comportamentos, pastejando e não pastejando, durante o período de uma hora. O primeiro ponto a observar é a concentração de dados do eixo X (azul) no valor de 0,25 para o comportamento de pastejo, enquanto o mesmo eixo se concentra no valor de 0,75 para o comportamento de não pastejando. Isso ocorre porque para pastejar os bovinos mantem a cabeça mais baixa, próxima do solo. Podemos observar que os valores de Z (vermelho) oscilam em ambos os comportamentos, enquanto os valores do eixo Y (verde), lado a lado, oscilaram mais quando pastejando. Por fim podemos observar a concentração dos valores. Para a atividade de pastejo temos os valores de Z mais altos, se concentrando em 0,75, seguido pelos valores de Y concentrados em -0,5. Enquanto quando não pastejando temos o eixo Y concentrado em 0,5, seguido pelo eixo Z, com valores concentrados em -0,75. A diferença entre os comportamentos é claramente visível.

Figura 33: Distribuição dos dados brutos ao longo de uma hora, escolhido aleatoriamente, de observações comportamentais de pastejando (superior) e não pastejando (inferior).



Legenda: Os valores dos eixos do acelerômetro X (vertical), Y e Z (horizontais) podem ser vistos nas colorações azul, verde e vermelho, respectivamente.

Fonte: Do autor (2021).

A arquitetura dos modelos preditivos e seus hiper parâmetros influenciam na sua predição. Pensando nisso, foi realizada uma pesquisa de grade aleatória para identificar quais as melhores arquiteturas e hiper parâmetros em todos os modelos utilizados. O resultado dessa pesquisa pode ser visualizado na Tabela 6.

Tabela 6: Descrição dos melhores hiper parâmetros ou arquitetura identificada na pesquisa de grade aleatória para modelos preditivos testados no banco de dados PNP.

	LAO	LDO	Holdout	LHO10	LHO25
<i>Elastic Net Generalized Linear Regression</i>					
Alpha	0,1	0,3	0,9	0,9	0,9
Lambda	0,1	0,1	0,0	0,0	0,0
<i>Random Forest</i>					
Número de árvores	200	40	40	40	200
Profundidade máxima	20	20	20	20	20
Mínimo de observações	1	10	10	10	1
Número de bins	2	6	6	6	2
<i>Artificial Neural Network</i>					

Função de ativação	Maxout	Maxout	Maxout	Maxout	Maxout
Camadas ocultas	3	3	3	3	3
Neurônios por camada	50, 50, 50	50, 50, 50	20, 20, 20	20, 20, 20	20, 20, 20
Taxa de abandono	0,2	0,2	0,2	0,2	0,2
Regularização L1	0,000001	0,000001	0,000039	0,000039	0,000039
Regularização L2	0,000089	0,000089	0,000026	0,000026	0,000026
<i>Gradient Boosting</i>					
Número de árvores	80	80	80	80	80
Profundidade máxima	80	80	80	80	80
Mínimo de observações	10	10	10	10	10
Número de bins	3	3	3	3	3
Tx de amostragem	0,26	0,26	0,26	0,26	0,26
Histograma	Uniform Adaptive				
Tx de AmCol	0,62	0,62	0,62	0,62	0,62
Tx de AmCol - level	0,94	0,94	0,94	0,94	0,94
Tx de AmCol - árvore	0,93	0,93	0,93	0,93	0,93

Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados, taxa (Tx) e amostragem colunas (AmCol).

Fonte: Do autor (2021).

As métricas de desempenho nos auxiliam a avaliar os modelos desenvolvidos. Para gerar a matriz de confusão foi atribuído valor positivo para o comportamento de pastejo e negativo para o comportamento de não pastejo. Na Tabela 7 é possível observar o desempenho dos modelos durante o treinamento para prever os comportamentos de pastejo e não pastejo. O desempenho dos modelos no treinamento pode nos mostrar como a estratégia de validação afeta na avaliação dos modelos. O modelo preditivo GLM foi inferior aos outros modelos. Sua capacidade prever corretamente o comportamento observado (acurácia), não ultrapassando os 51%, logo a porcentagem de vezes que o modelo errou (taxa de erro) foi superior a 49% para todas as estratégias de validação. A principal dificuldade que esse modelo encontrou foi classificar o comportamento de pastejo, é possível observar isso por sua baixa sensibilidade, quantidade de vezes que o modelo acertou ao prever o comportamento de pastejando e de fato o animal estava pastejando, e alta especificidade, quantidade de vezes que o modelo acertou ao prever o comportamento de não pastejo e de fato o animal não estava pastejando, ou seja, o modelo classificou a maioria dos dados como não pastejando. Os parâmetros de valor preditivo positivo (VPP) e valor preditivo negativo (VPN) corroboram com essa observação, pois são a quantidade de acertos em cima das predições realizadas, e em ambos os casos sua porcentagem foi inferior aos

outros modelos preditivos. O modelo GB vem logo depois, com parâmetros ligeiramente mais baixos, seguido pelo ANN e do RF, esses últimos com resultados bem próximos.

Tabela 7: Desempenho do treinamento dos modelos para predizer os comportamentos de pastejando ou não pastejando usando diferentes estratégias de validação.

Modelo preditivo	Acurácia (%)	Taxa de erro (%)	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)
<i>Holdout (20%)</i>						
GLM	50,1	49,9	2,9	99,0	74,1	49,7
RF	73,3	26,7	65,2	81,7	78,6	69,4
GB	67,6	32,4	55,5	80,2	74,3	63,5
ANN	71,7	28,3	60,7	83,1	78,7	67,2
<i>Leave-animals-out</i>						
GLM	50,7	49,6	7,7	99,2	69,4	50,12
RF	73,7	26,3	65,3	82,1	78,6	70,2
GB	67,8	32,2	54,0	81,7	74,8	63,8
ANN	72,2	27,8	62,8	82,9	78,0	68,4
<i>Leave-days-out random</i>						
GLM	50,4	49,6	2,5	99,1	73,7	50,0
RF	73,9	26,2	65,25	82,6	79,2	70,1
GB	68,4	31,7	55,8	81,1	75,0	64,4
ANN	72,5	27,5	62,3	82,9	78,8	68,4
<i>Leave-height-out 10 cm</i>						
GLM	50,9	49,1	3,4	98,6	70,9	50,4
RF	75,0	25,0	68,4	81,6	78,9	72,0
GB	70,2	29,8	61,3	79,1	74,6	67,1
ANN	73,5	26,5	64,9	82,2	78,5	70,0
<i>Leave-height-out 25 cm</i>						
GLM	47,7	52,3	1,7	99,2	69,3	47,4
RF	74,1	25,9	68,9	79,9	79,3	69,6
GB	68,8	31,3	60,9	77,6	75,3	63,9
ANN	71,1	28,9	61,9	81,4	78,8	65,6

Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out 10 cm* (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out 25 cm* (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do autor (2021).

Já na Tabela 8 podemos visualizar o desempenho da validação desses modelos preditivos, quando expostos aos dados selecionados para a testagem. As mesmas métricas utilizadas para avaliar o desempenho foram utilizadas para a validação.

Novamente o modelo preditivo GLM foi inferior aos demais, mantendo a tendência do treinamento do modelo. O RF e o ANN foram semelhantes nos parâmetros de acurácia, taxa de erro VPP e VPN, porém o RF foi ligeiramente melhor quando olhamos a sensibilidade e a especificidade, o que indica que esse modelo conseguiu prever com mais exatidão o real comportamento registrado.

Tabela 8: Desempenho da validação dos modelos preditivos para prever categorias de comportamento de pastejo ou não pastejo visualmente observadas em novilhas tabapuã usando diferentes estratégias de validação.

Modelo preditivo	Acurácia (%)	Taxa de erro (%)	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)
<i>Holdout (20%)</i>						
GLM	50,4	49,7	3,1	98,9	74,0	49,8
RF	73,3	26,7	64,2	82,6	79,1	69,2
GB	68,3	31,8	57,2	79,6	74,2	64,4
ANN	72,0	28,0	62,5	81,8	77,9	68,0
<i>Leave-animals-out</i>						
GLM	48,3	51,7	1,8	99,5	81,2	47,9
RF	62,5	37,5	59,3	66,0	65,8	59,5
GB	60,6	39,4	56,2	65,3	64,1	57,5
ANN	63,6	36,4	46,0	83,0	74,9	58,2
<i>Leave-days-out random</i>						
GLM	46,8	53,2	1,1	99,4	68,7	46,7
RF	64,4	35,6	57,2	72,7	70,6	59,6
GB	59,0	41,0	48,2	71,3	65,9	54,5
ANN	65,8	34,2	51,4	82,4	77,1	59,6
<i>Leave-height-out 10 cm</i>						
GLM	48,5	51,5	3,8	98,3	71,3	47,8
RF	59,9	40,1	55,7	64,6	63,7	56,6
GB	56,0	44,0	53,2	59,1	59,2	53,1
ANN	60,8	39,2	57,5	64,5	64,4	57,6
<i>Leave-height-out 25 cm</i>						
GLM	52,4	47,6	2,0	99,6	80,9	52,1
RF	61,7	38,3	58,3	65,0	60,9	62,5
GB	58,0	42,0	49,3	66,2	57,7	58,3
ANN	64,9	35,1	51,0	78,0	68,4	63,0

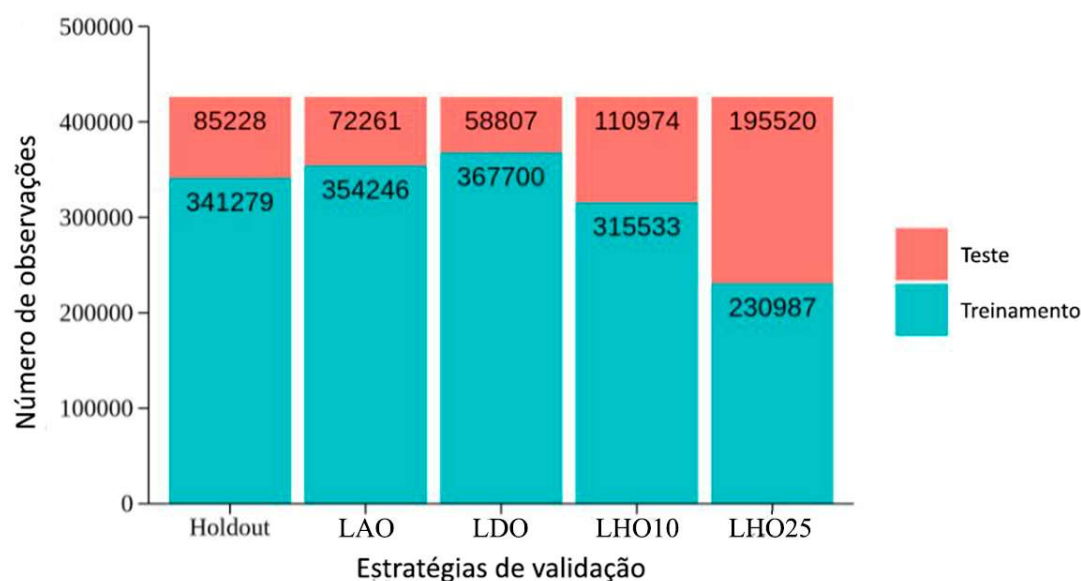
Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out 10 cm* (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out 25 cm* (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do autor (2021).

4.2 – Conjunto de dados PRO: Pastejando x Ruminando x Ócio

Ao todo, 604.841 dados foram coletados, porém, após a limpeza do banco de dados, restaram 567.940 observações. O conjunto de dados PRO possuía 309.763 (54,5%) dados de pastejando, 141.787 (25%) de ruminando e 116.390 (20,5%) de ócio. Assim como no conjunto de dados PRP, a quantidade de dados utilizadas para a treinamento e para validação variou conforme a estratégia de validação (Figura 35).

Figura 34: Divisão do conjunto de dados PRO em subconjunto de treinamento e subconjunto de validação cruzada (teste) para as diferentes estratégias de validação.



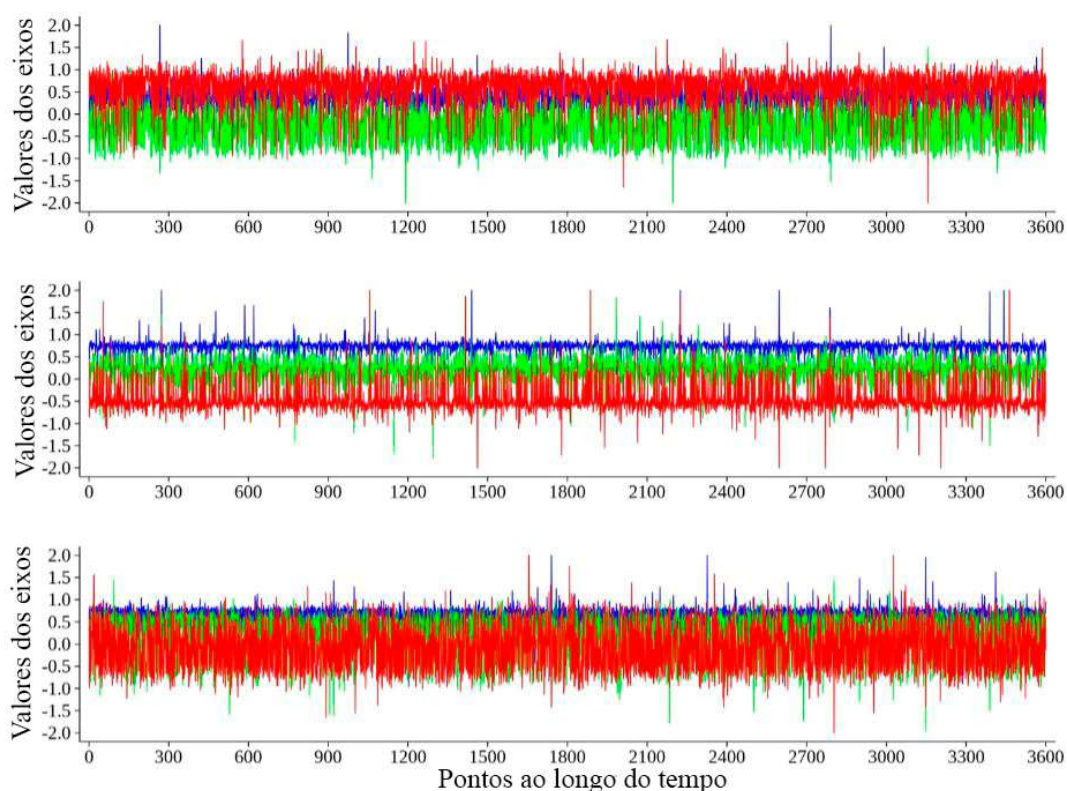
Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do Autor (2021).

A Figura 36 é uma representação gráfica dos dados brutos, de uma amostra, para os comportamentos de pastejando, ruminando e ócio durante uma hora de observações. Semelhante a Figura 34 podemos observar uma concentração do eixo X em 0,25 para a atividade de pastejo, enquanto para as atividades de ócio e ruminando o valor ficou concentrado em 0,75. Também podemos observar uma semelhança entre os valores de Y e Z para os comportamentos de ruminando e ócio, diferindo na oscilação desses eixos, principalmente o eixo Z que é muito mais oscilante no ócio. Os valores dos eixos

Y e Z também foram diferentes quando comparamos o comportamento de ruminção e ócio. Quando a novilha estava ruminando o eixo Y se concentrou no valor de 0,25, já em ócio a concentração do mesmo eixo foi no valor de 0,5. O eixo Z foi mais concentrado no valor de -0,5 quando ruminando e mais concentrado em -0,25 quando em ócio.

Figura 35: Distribuição dos dados brutos, ao longo de uma hora, de pastejando (superior), ruminando (centro) e ócio (inferior) para o conjunto de dados PRO.



Legenda: Os valores dos eixos do acelerômetro X (vertical), Y e Z (horizontais) podem ser vistos nas colorações azul, verde e vermelho, respectivamente.

Fonte: Do autor (2021).

Os resultados da pesquisa de grade aleatória para identificar as melhores arquiteturas e hiper parâmetros em todos os modelos com o conjunto de dados PRO estão na Tabela 9.

Tabela 9: Descrição dos melhores hiper parâmetros ou arquitetura identificada na pesquisa de grade aleatória para modelos preditivos capazes de prever os comportamentos de pastejando, ruminando e ócio.

	LAO	LDO	Holdout	LHO10	LHO25
<i>Elastic Net Generalized Linear Regression</i>					
Alpha	0,0	0,9	0,90	0,0	0,0
Lambda	0,0	0,0	0,0	0,0	0,0
<i>Random Forest</i>					
Número de árvores	40	40	40	40	40
Profundidade máxima	20	20	20	20	20
Mínimo de observações	10	10	10	10	10
Número de bins	6	6	6	6	6
<i>Artificial Neural Network</i>					
Função de ativação	Maxout	Maxout	Maxout	Maxout	Maxout
Camadas ocultas	3	3	3	3	3
Neurônios por camada	20, 20, 20	20, 20, 20	20, 20, 20	20, 20, 20	20, 20, 20
Taxa de abandono	0,2	0,2	0,2	0,2	0,2
Regularização L1	0,000010	0,000039	0,000039	0,000039	0,000039
Regularização L2	0,000089	0,000026	0,000026	0,000026	0,000026
<i>Gradient Boosting</i>					
Número de árvores	80	53	80	80	80
Profundidade máxima	80	80	80	80	80
Mínimo de observações	10	10	10	10	10
Número de bins	3	3	3	3	3
Tx de amostragem	0,26	0,26	0,26	0,26	0,26
Histograma	Uniform Adaptive				
Tx de AmCol	0,62	0,62	0,62	0,62	0,62
Tx de AmCol - level	0,94	0,94	0,94	0,94	0,94
Tx de AmCol - árvore	0,93	0,93	0,93	0,93	0,93

Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados, taxa (Tx) e amostragem colunas (AmCol).

Fonte: Do autor (2021).

Na Tabela 10 é possível observar o desempenho dos modelos durante o treinamento para prever os comportamentos de pastejo, ruminação e ócio. Por utilizarmos um banco de dados com três categorias, não é possível calcular a taxa de erro, uma vez que para avaliar os modelos preditivos são geradas três matrizes de confusão. A primeira aonde é atribuído valor positivo para o comportamento de pastejo e negativo para ruminando e ócio, outra aonde é atribuído como positivo o

comportamento de ruminção e negativo para pastejo e ócio, e a última aonde é atribuído valor positivo para ócio e negativo para pastejo e ruminção.

De maneira geral os modelos tiveram dificuldades para classificar o comportamento de ócio. Todos os modelos apresentaram baixa sensibilidade, apesar da alta acurácia e alta especificidade. O GLM foi o pior modelo em termos de desempenho geral, mostrando uma dificuldade ainda maior em classificar o comportamento de ócio, com uma sensibilidade inferior a 1% e VPP abaixo de 30% em três das cinco estratégias de validação. Os demais modelos se mostraram equilibrados entre si para classificar as categorias de pastejando e ruminando.

Tabela 10: Desempenho do treinamento de modelos preditivos para predizer categorias de comportamento de pastejo, ruminção e ócio visualmente observados em novilhas Tabapuã usando diferentes estratégias de validação.

Modelo preditivo	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)
<i>Holdout (20%)</i>					
<i>GLM</i>					
Pastejando	58,2	76,6	46,0	48,4	74,9
Ruminando	63,8	39,2	76,0	44,7	71,6
Ócio	70,8	10,8	93,1	36,7	73,8
<i>RF</i>					
Pastejando	73,4	79,5	69,4	63,2	83,6
Ruminando	76,3	64,8	81,9	64,0	82,5
Ócio	75,7	35,5	90,7	58,6	79,1
<i>GB</i>					
Pastejando	69,5	76,0	65,2	59,1	80,4
Ruminando	73,5	59,7	80,4	60,1	80,1
Ócio	73,9	31,3	89,8	53,2	77,9
<i>ANN</i>					
Pastejando	72,3	73,4	71,5	63,0	80,2
Ruminando	71,1	66,6	73,4	55,3	81,6
Ócio	72,4	24,5	90,2	48,1	76,3
<i>Leave-animals-out</i>					
<i>GLM</i>					
Pastejando	61,4	75,7	51,8	51,2	76,1
Ruminando	63,4	55,6	67,7	48,5	73,6
Ócio	75,2	0,2	99,6	16,8	75,4
<i>RF</i>					
Pastejando	75,9	82,2	71,7	66,0	85,8
Ruminando	76,7	69,2	80,8	66,3	82,7
Ócio	78,7	33,5	93,4	62,4	81,1

<i>GB</i>					
Pastejando	71,8	79,0	67,0	61,6	82,7
Ruminando	73,9	64,6	79,0	62,8	80,3
Ócio	77,1	28,4	93,0	57,0	79,9
<i>ANN</i>					
Pastejando	75,3	75,1	75,4	67,1	81,9
Ruminando	71,1	71,9	70,6	57,2	82,1
Ócio	75,6	22,3	93,0	51,0	78,6
<i>Leave-days-out random</i>					
<i>GLM</i>					
Pastejando	55,5	79,5	39,1	47,2	73,6
Ruminando	63,0	40,6	74,1	43,8	71,5
Ócio	73,5	0,6	99,2	21,8	73,9
<i>RF</i>					
Pastejando	74,0	81,2	69,0	64,2	84,3
Ruminando	77,5	66,4	83,0	66,0	83,2
Ócio	76,9	34,8	91,8	60,0	79,9
<i>GB</i>					
Pastejando	68,8	80,6	60,7	58,4	82,1
Ruminando	74,8	59,6	82,4	62,8	80,4
Ócio	75,5	26,6	92,8	56,5	78,2
<i>ANN</i>					
Pastejando	72,4	74,6	70,9	63,7	80,3
Ruminando	72,6	67,1	75,4	57,6	82,1
Ócio	73,3	25,1	90,4	48,0	77,3
<i>Leave-height-out 10 cm</i>					
<i>GLM</i>					
Pastejando	62,7	70,2	58,0	51,7	75,2
Ruminando	61,2	61,0	61,3	47,7	73,0
Ócio	75,8	0,1	99,9	28,0	75,8
<i>RF</i>					
Pastejando	75,2	80,3	71,9	64,8	85,1
Ruminando	76,3	72,4	78,6	66,3	83,1
Ócio	78,4	28,8	94,2	61,2	80,6
<i>GB</i>					
Pastejando	71,9	77,0	68,5	61,1	82,3
Ruminando	73,6	68,6	76,4	62,8	80,8
Ócio	77,2	24,9	94,0	56,7	79,7
<i>ANN</i>					
Pastejando	74,0	76,7	72,3	64,0	82,8
Ruminando	71,8	74,5	70,2	59,2	82,6
Ócio	76,1	14,9	95,6	51,9	77,9
<i>Leave-height-out 25 cm</i>					
<i>GLM</i>					
Pastejando	51,9	83,6	29,2	45,9	71,2
Ruminando	69,9	11,9	92,2	37,0	73,1
Ócio	66,4	19,1	87,0	39,2	71,1

<i>RF</i>					
Pastejando	73,5	82,1	67,3	64,3	84,0
Ruminando	80,2	58,3	88,6	66,3	84,7
Ócio	74,1	43,9	87,3	60,2	78,1
<i>GB</i>					
Pastejando	70,0	79,5	63,1	60,8	81,1
Ruminando	78,4	54,1	87,7	62,9	83,3
Ócio	72,1	39,3	86,4	55,8	76,5
<i>ANN</i>					
Pastejando	72,1	74,5	70,3	64,3	79,3
Ruminando	75,6	57,8	82,4	55,8	83,5
Ócio	69,7	37,6	83,7	50,2	75,4

Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do autor (2021).

Na Tabela 11 podemos observar o desempenho dos modelos preditivos quando expostos a validação utilizando os dados selecionados para a testagem. Assim como no treinamento dos modelos podemos observar que todos tiveram dificuldades para classificar o comportamento de ócio, principalmente o modelo GLM com as menores porcentagens de sensibilidade. Os modelos tiveram maior acurácia e especificidade ao predizerem o comportamento de ruminação, porém tiveram menor sensibilidade e VPP quando comparado ao predizer o comportamento de pastejo. Todos os modelos tiveram valores satisfatórios de VPN. Quando comparamos as estratégias de validação podemos perceber que os modelos tiveram melhores resultados quando foram expostos a validação *Holdout* e as piores quando expostos as validações LAO e LDO.

Tabela 11: Validação dos modelos preditivos para prever categorias de comportamento de pastejo, ruminação e ócio visualmente observados em novilhas tabapuã usando diferentes estratégias de validação.

Modelo preditivo	Acurácia (%)	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)
<i>Holdout (20%)</i>					
<i>GLM</i>					
Pastejando	58,1	76,6	46,0	48,3	74,9
Ruminando	63,7	39,2	75,8	44,5	71,6
Ócio	70,8	10,6	93,2	36,9	73,6
<i>RF</i>					
Pastejando	73,7	79,8	69,7	63,5	84,0
Ruminando	76,3	64,8	82,0	64,0	82,5

Ócio	75,9	36,1	90,8	59,3	79,2
<i>GB</i>					
Pastejando	69,7	75,9	65,7	59,3	80,5
Ruminando	73,6	60,1	80,2	60,1	80,3
Ócio	74,0	31,6	89,8	53,5	77,9
<i>ANN</i>					
Pastejando	71,8	76,0	69,0	61,8	81,4
Ruminando	71,5	65,8	74,4	56,0	81,4
Ócio	73,2	23,4	91,9	51,7	76,3
<i>Leave-animals-out</i>					
<i>GLM</i>					
Pastejando	59,8	50,1	65,8	47,8	67,9
Ruminando	48,9	68,8	43,2	25,6	83,0
Ócio	60,5	0,2	99,8	42,9	60,6
<i>RF</i>					
Pastejando	57,9	64,0	54,2	46,6	70,7
Ruminando	68,3	38,7	76,7	32,0	81,5
Ócio	58,4	23,2	81,4	44,8	62,0
<i>GB</i>					
Pastejando	56,7	62,8	52,8	45,4	69,4
Ruminando	66,7	39,7	74,3	30,4	81,3
Ócio	58,1	19,8	83,1	43,3	61,4
<i>ANN</i>					
Pastejando	60,5	56,2	63,1	48,8	69,7
Ruminando	62,8	46,6	67,4	28,8	81,6
Ócio	58,3	22,5	81,7	44,4	61,8
<i>Leave-days-out random</i>					
<i>GLM</i>					
Pastejando	65,1	73,8	60,4	49,5	81,5
Ruminando	58,6	58,2	58,9	40,3	74,7
Ócio	66,2	2,1	98,1	35,5	66,8
<i>RF</i>					
Pastejando	61,9	73,3	56,0	46,7	79,9
Ruminando	62,1	35,7	74,6	40,1	70,9
Ócio	64,5	22,6	85,5	43,6	68,9
<i>GB</i>					
Pastejando	60,8	73,7	54,1	45,7	79,6
Ruminando	60,3	36,6	71,5	38,0	70,3
Ócio	64,4	16,6	88,2	41,3	68,0
<i>ANN</i>					
Pastejando	65,3	61,9	67,0	49,7	77,0
Ruminando	57,4	45,9	62,9	37,1	70,9
Ócio	63,3	20,6	84,5	39,9	68,1
<i>Leave-height-out 10 cm</i>					
<i>GLM</i>					
Pastejando	51,5	74,6	34,9	45,1	65,7
Ruminando	60,9	32,0	69,5	23,7	77,5

Ócio	64,6	0,0	99,9	28,3	64,6
<i>RF</i>					
Pastejando	56,4	70,6	46,3	48,5	68,8
Ruminando	64,2	28,2	74,8	25,0	77,8
Ócio	63,9	17,8	89,1	47,1	66,4
<i>GB</i>					
Pastejando	54,4	69,0	43,9	46,8	66,4
Ruminando	64,5	29,2	75,0	25,8	78,1
Ócio	63,4	16,0	89,4	45,2	66,0
<i>ANN</i>					
Pastejando	60,5	66,3	56,3	52,1	70,0
Ruminando	63,8	39,9	70,9	28,9	79,9
Ócio	64,4	21,3	88,0	49,3	67,1
<i>Leave-height-out 25 cm</i>					
<i>GLM</i>					
Pastejando	54,5	81,3	38,4	44,1	77,4
Ruminando	60,1	10,8	92,2	47,2	61,4
Ócio	68,7	30,1	80,4	31,7	79,2
<i>RF</i>					
Pastejando	58,9	75,6	48,9	47,0	77,1
Ruminando	61,9	27,6	84,3	53,3	64,2
Ócio	69,3	25,4	82,5	30,4	78,6
<i>GB</i>					
Pastejando	55,4	72,2	45,4	44,2	73,2
Ruminando	60,9	26,5	83,3	50,8	63,6
Ócio	69,0	22,5	83,0	28,6	78,0
<i>ANN</i>					
Pastejando	63,2	69,6	59,3	50,6	76,5
Ruminando	59,8	35,7	75,4	48,6	64,3
Ócio	69,1	25,4	82,3	30,2	78,5

Legenda: *Holdout* com 20% dos dados selecionados aleatoriamente para testagem, *Leave-animals-out* (LAO) com dois animais selecionados aleatoriamente, *Leave-days-out* (LDO) com sete dias selecionados aleatoriamente, *Leave-height-out* 10 cm (LHO10) com todos os dados de 10 cm selecionados e *Leave-height-out* 25 cm (LHO25) com todos os dados de 25 cm selecionados.

Fonte: Do autor (2021).

4.3 – Conjunto de dados externo

O banco de dados PNP foi utilizado também para validação externa dos modelos preditivos desenvolvidos em trabalho anterior do nosso grupo (RIBEIRO et al., 2021), com GLM, RF e ANN. Os resultados dessa validação externa podem ser observados na Tabela 12. Todos os modelos tiveram dificuldades em realizar as predições, com acurácia entre 48 e 59,4%.

Tabela 12: Validação externa dos modelos preditivos para prever categorias de comportamento de pastejo ou não pastejo visualmente observadas em novilhas e bovinos Tabapuã usando diferentes estratégias de validação.

Modelo preditivo	Acurácia (%)	Taxa de erro (%)	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)
<i>Leave-animals-out</i>						
GLM	48,0	51,9	25,5	71,4	47,9	48,2
RF	57,5	42,5	59,5	55,5	58,0	57,0
ANN	55,1	44,9	53,2	57,1	56,1	54,2
<i>Leave-days-out</i>						
GLM	49,2	50,8	37,0	61,7	49,9	48,7
RF	59,4	40,6	75,2	43,0	57,7	62,7
ANN	59,5	40,5	73,3	45,3	58,0	62,2
<i>Holdout (20%)</i>						
GLM	49,2	50,8	27,4	71,7	49,9	48,9
RF	59,4	40,6	70,3	48,2	58,3	61,1
ANN	58,5	41,5	59,1	57,9	59,2	57,9

Legenda: *Generalized linear model* (GLM), *random forest* (RF), *artificial neural network* (ANN), valor preditivo positivo (VPP) e valor preditivo negativo (VPN).

Fonte: Do autor (2021).

5 – DISCUSSÃO

O comportamento ingestivo de animais em pastejo se altera ao longo do rebaixamento do dossel (CASAGRANDE *et al.*, 2011; FERREIRA *et al.*, 2022). O tempo despendido na atividade de pastejo pode refletir o aumento na dificuldade de colheita conforme a altura do pasto diminui, até se tornar tão inviável que os animais diminuem sua procura por alimento. Além disso, o tempo de ruminação pode ser um indicativo do consumo de forragem pelos animais. Eficiência de colheita e consumo de pasto são variáveis chave na definição da rentabilidade e sustentabilidade de um sistema de produção de leite/carne em pastagens.

Dessa forma, a automação da observação do comportamento ingestivo e a consequente geração de informações e alertas que auxiliem nas tomadas de decisão no manejo do pastejo se destacam como potenciais tecnologias de alto impacto nesse sistema produtivo (ANDRIAMANDROSO; MERCATORIS; *et al.*, 2016; HOMBURGER *et al.*, 2014). Por isso, este trabalho teve como objetivo desenvolver modelos capazes de prever adequadamente variáveis de comportamento animal de bovinos em pastejo a partir de dados de acelerômetros vestíveis. A utilização de vários animais, ao longo de três estações do ano e sob 3 estratégias de manejo possibilitou grande amplitude nas variáveis de comportamento, potencialmente aumentando a representatividade e robustez do banco de dados.

Os três comportamentos animais de interesse inicial foram pastejo, ruminação e ócio (conjunto de dados PRO). Todos os modelos tiveram dificuldades ao classificar o comportamento de ócio, provavelmente por esse possuir mais ruídos que os demais comportamentos, como os animais se coçando ou percorrendo curtas distâncias. Dunne *et al.* (2017) relataram problema semelhante com ruídos no banco de dados ao utilizarem colares GPS para prever o comportamento de bovinos. Ruídos são erros aleatórios no banco de dados que podem ocorrer por diversos motivos, dificilmente são eliminados, mas podem ser minimizados (KALAPANIDAS *et al.*, 2003; WATSON-HAIGH; O'NEILL; KADARMIDEEN, 2012). Além disso a quantidade de dados desse comportamento foi inferior a quantidade presente para pastejando e ruminando, assim os ruídos podem ter sido ainda mais significativos, dificultando a realização das previsões (BHAVSAR; GANATRA, 2012; DUTTA *et al.*, 2015).

Quase todos os modelos preditivos tiveram menor acurácia, ainda que pequena a diferença, ao predizer o comportamento de pastejo em comparação com a ruminação. Porém os modelos apresentaram maiores valores de sensibilidade (porcentagem de acertos quando o modelo predisse o comportamento de pastejo), VPP (porcentagem de acertos quando o modelo tentou predizer o comportamento de pastejo) e VPN (porcentagem de acertos quando o modelo tentou predizer o que não era pastejo) ao predizerem o comportamento de pastejo. Resultados semelhantes aos encontrados na literatura, que descrevem uma melhor performance dos modelos preditivos ao categorizar o comportamento de pastejo (BARWICK *et al.*, 2018; DECANDIA *et al.*, 2018; SMITH *et al.*, 2016).

A baixa sensibilidade na classificação do comportamento de ócio, principalmente no modelo GLM, podem indicar uma situação de *underfitting*, erro causado por suposições errôneas do modelo durante o treinamento, fazendo com que a generalização seja muito alta (BHANDE, 2018). Valletta *et al.*, (2017) descreveu a possibilidade de ocorrência desse erro, e do erro *overfitting*, ao utilizar técnicas de ML para predizer o comportamento animal. Os valores altos de especificidade para a predição desses comportamentos e os valores de VPN ainda nos indicam uma possível causa desse erro ter ocorrido. Quando os modelos tentaram predizer o comportamento de ócio eles classificaram boa parte dos dados como não ócio (negativo), assim a especificidade foi alta, pois ela mede a quantidade de acertos sobre os valores reais, como quase todos os dados foram classificados como não ócio o acerto foi grande. Já o valor de VPN mede a quantidade de acertos ao classificar como não ócio (negativo) sobre os valores preditos, então toda vez que os modelos classificaram um comportamento de ócio como não ócio o VPN foi afetado negativamente, os valores de VPN ficam próximos de 65%, bem inferiores aos valores de especificidade. Situação parecida ocorreu quando os modelos tentaram classificar o comportamento de ruminação.

De maneira geral o melhor modelo para realizar as predições foi o ANN, principalmente ao predizer o comportamento de pastejo, semelhante ao relatado na literatura (ALVARENGA *et al.*, 2016; NADIMI *et al.*, 2012; RAHMAN *et al.*, 2018). O modelo RF e o modelo GB foram bem semelhantes em desempenho, ficando pouco atrás do ANN na sensibilidade e VPP ao classificar a ruminação. O modelo RF, é muito utilizado em diversas áreas da ciência aonde é necessário realizar alguma predição, isso

inclui a predição do comportamento ingestivo (MANSBRIDGE *et al.*, 2018; RIBEIRO *et al.*, 2021). O modelo GLM foi o modelo que apresentou pior desempenho, com baixos valores de acurácia e sensibilidade, principalmente ao predizer o comportamento de ócio.

O segundo interesse foi verificar se o desempenho dos modelos preditivos melhoraria com uma simplificação na classificação dos dados para categorias binárias (pastejando e não pastejando). Ao juntar os comportamentos de ruminando e ócio, a quantidade de dados entre os comportamentos foi mais próxima, com 309 mil dados para pastejando e 258 mil para não pastejando.

O modelo preditivo GLM teve um desempenho baixo quando comparado aos outros modelos, com valores ínfimos de sensibilidade, acurácia e VPN. Assim como ocorreu no conjunto de dados PRO, o modelo GLM apresentou um erro *underfitting* em todas as estratégias de validação com sensibilidade abaixo de 4% e especificidade acima de 98%, esse erro é ainda mais visível quando olhamos para os valores de VPP e VPN com valores próximos de 72% e 50% respectivamente. Esse modelo provavelmente classificou quase todos os dados como não pastejando. Os modelos RF e ANN tiveram valores de acurácia, VPP e VPN bem semelhantes em todas as estratégias. A ANN teve maior valor de especificidade enquanto a RF teve maior valor de sensibilidade. Resultados semelhantes aos encontrados por Ribeiro *et al.*, 2021 onde o modelo GLR também apresentou o pior desempenho quando comparado aos modelos RF e ANN.

Os parâmetros utilizados, acurácia, sensibilidade, especificidade, VPP e VPN, indicam que os modelos preditivos foram melhores ao classificar apenas dois comportamentos em comparação a classificação de três comportamentos. A diferença entre os conjuntos de dados PRO e PNP é a junção dos comportamentos de ócio e ruminação em não pastejo, então essa melhora ao predizer os comportamentos pode estar relacionada a maior quantidade de dados.

Ainda mais importante que o tamanho do banco de dados é a qualidade dos dados coletados. Diferenciar melhor os comportamentos, separando ócio e ruminando em ócio deitado, ócio em pé, ruminando em pé e ruminando deitado pode auxiliar na diminuição dos ruídos ao classificar esses comportamentos. Outra opção é combinar dois, ou mais, diferentes sensores com a finalidade de diminuir os ruídos.

Melhorar a qualidade dos dados, diminuindo os ruídos, permite um intervalo maior na recepção dos dados, conseqüentemente aumenta a vida útil das baterias utilizadas pelos sensores. Em uma propriedade rural, principalmente com animais a pasto, é difícil realizar a troca de baterias com frequência, sendo necessárias estratégias que preservem o equipamento em funcionamento por mais tempo. Decandia *et al.* (2018) demonstraram que intervalos de até um minuto não afetam negativamente na acurácia, precisão, sensibilidade e especificidade dos modelos utilizados para prever os comportamentos de pastejo, ruminação e outros em ovinos a pasto. Não foi encontrado na literatura artigos que comparem diferentes intervalos de recepção para dados para bovinos a pasto, informação essencial para o desenvolvimento de um produto comercial.

Grande parte das pesquisas envolvendo pecuária de precisão serve de base para o desenvolvimento de produtos que sanem dores dos pecuaristas, portanto as pesquisas devem ser condizentes com a realidade da pecuária, além de antecipar possíveis contratempos para a utilização das tecnologias. O modelo preditivo utilizado deve ser generalista, ou seja, deve ser aplicável em qualquer situação, independentemente de como foram realizadas as coletas dos dados para o treinamento. Uma maneira de verificar a generalização dos modelos preditivos é expô-los a testes com dados que não foram utilizados para o treinamento, também chamado de validação (BRESOLIN; DÓREA, 2020; KOHAVI, 1995).

A validação pode ser realizada seguindo diversas estratégias. Em nosso trabalho utilizamos cinco diferentes estratégias. A primeira estratégia é a *holdout*, mais utilizada na pecuária de precisão (BRESOLIN; DÓREA, 2020; RIBEIRO *et al.*, 2021). A segunda estratégia LAO tem o objetivo de simular a entrada de novos animais no rebanho. A estratégia LDO avalia dias diferentes ao longo do ano, conseqüentemente avalia diferentes condições do pasto. As estratégias de LHO10 e LHO25 simulam diferentes alturas da pastagem, 10 e 25 cm respectivamente, que por sua vez interferem na estrutura da mesma (BARCELLOS *et al.*, 2004; DA COSTA LIMA *et al.*, 2020).

Para ambos os conjuntos de dados a estratégia de validação *holdout* apresentou maiores valores de acurácia, sensibilidade, especificidade, VPP e VPN quando comparado as demais estratégias de validação. Estratégias que utilizam um conjunto de dados para teste que, de alguma forma, são correlacionados com o banco de dados, como a *holdout*, não são representativas da realidade, pois inflam os resultados das

métricas de avaliação de um modelo preditivo ao reduzir a importância de situações comuns do cotidiano, como a interrelação biológica dos bovinos, o ambiente e a forragem (BRESOLIN; DÓREA, 2020; DÓREA *et al.*, 2018). A aleatoriedade na seleção dos dados que são utilizados para a testagem dos modelos mascara essas relações e infla os resultados obtidos ao predizer um comportamento, já que dados com interrelações biológicas (por exemplo dados do mesmo animal) estão no conjunto de treinamento e no de validação (RIBEIRO *et al.*, 2021). Diferentes condições ambientais, de manejo, estação do ano e a espécie da forrageira interferem na estrutura do pasto, como a massa da forragem e a relação folha: colmo que, por sua vez, interferem no comportamento ingestivo dos bovinos (DA SILVA, S. C.; CARVALHO, 2005; DONALD, 1986). Ao utilizarmos a estratégia de *holdout* mantemos tais interrelações nos dados e ignoramos a heterogeneidade da pecuária a pasto no Brasil, não só entre as áreas plantadas como ao longo do ano. Pensando nessa heterogeneidade faz-se necessário utilizar outras estratégias de validação, que de fato simulem diferentes condições de pasto e não sejam apenas um subconjunto idêntico ao banco de dados.

Na estratégia LAO os dados de dois animais foram retirados do treinamento e utilizados para os testes, o objetivo é simular a entrada de novos animais no rebanho. A aquisição de novos animais, seja pela compra ou pelo nascimento, é uma situação comum na pecuária, então produtos comerciais necessariamente serão treinados em animais diferentes dos utilizados para realizar as predições, por isso os modelos devem ser capazes de predizer o comportamento em indivíduos de grupos genéticos diferentes. O comportamento dos bovinos é definido por características genéticas e ambientais, dentre as características ambientais podemos citar a hierarquia no rebanho, que pode afetar o consumo de forragem pelos bovinos (FERNANDES *et al.*, 2017; JENSEN, 2017). Apesar de haver diferença no comportamento individual dos animais pela personalidade ou por hierarquia no rebanho, os animais pertenciam a mesma categoria e peso, então as mudanças individuais no comportamento não foram críticas nos modelos ANN, RF e GB, que apresentaram desempenho semelhantes. Em situação comercial os animais são mais diferentes entre si e uma validação externa pode indicar essa mudança.

Na estratégia LDO os dados de sete dias foram retirados aleatoriamente do treinamento e utilizados para a validação, o objetivo é simular diferentes estruturas do pasto, uma vez que as mudanças climáticas alteram a estrutura ao longo do ano. Apesar disso a estrutura do pasto não foi um fator crítico, com valores de acurácia,

sensibilidade e VPN semelhantes às encontradas na estratégia LAO e valores de especificidade e VPP superiores. Dos sete dias selecionados quatro pertenciam a uma mesma estação, com temperatura e precipitação próximas, além disso, o pasto utilizado é muito bem manejado ao longo do ano, esses fatores podem ter impedido uma diferença grande na estrutura do pasto.

Nas estratégias LHO10 e LHO25 o objetivo foi verificar se a altura interfere na predição do comportamento, para isso foram retirados do treinamento os dados referentes às alturas de 10 cm e 25 cm respectivamente. Os modelos submetidos a estratégia LHO10 apresentaram os menores valores de acurácia, sensibilidade, especificidade, VPP e VPN, enquanto os modelos apresentaram melhor desempenho na estratégia LHO25 quando comparados a LHO10, mas ainda inferior aos apresentados nas outras estratégias (*Holdout*, LAO e LDO). Isso já era esperado, pois ao pastejar os bovinos fixam as patas no chão e apreendem todo o alimento presente no semicírculo a sua frente, posição denominada estação de pastejo, em pastos mais rebaixados a oferta de forragem nessas estações de pastejo é reduzido e rapidamente atingem o ponto de abandono, os animais estão abandonam essa estação e procuram uma próxima, conseqüentemente se deslocam mais (CARVALHO; MORAES, 2005; DA SILVA; CARVALHO, 2005). A estratégia LHO10 seleciona os dados coletados em uma altura inferior a recomendada para a forrageira utilizada, logo os animais se locomovem mais a procura de alimento. Submeter os modelos a diferentes alturas é importante para a predição do comportamento ingestivo a pasto, pois além da diferença de altura ao longo do rebaixamento temos a diferença de altura entre as forrageiras.

O comportamento dos bovinos se altera à medida que o pasto é rebaixado (DIFANTE *et al.*, 2009), assim mudanças no padrão de comportamento podem ajudar a indicar o momento de retirar os animais do piquete. De Freitas Barbosa *et al.* (2013) avaliaram em seu trabalho o tempo despendido para as atividades de pastejo, ruminação e ócio, em quatro diferentes alturas. Quando o pasto estava acima da altura recomendada os animais gastaram menos tempo ruminando que em ócio, apesar do tempo pastejando ser semelhante ao encontrado na altura ideal. À medida que o pasto foi rebaixado os animais gastaram mais tempo pastejando, menos tempo em ócio e o mesmo tempo ruminando. Na altura de saída recomendada os animais gastaram 18% a mais do tempo pastejando em comparação à altura de entrada, apesar disso o tempo ruminando e ócio foi menor aos observados na altura de entrada.

O presente estudo encontrou resultados semelhantes. O tempo pastejando aumenta à medida que o pasto é rebaixado, até alcançar um ponto máximo, quando a altura da pastagem está próxima de 60% da altura de entrada, a partir desse ponto os animais diminuíram a procura por alimento, devido as dificuldades encontradas. O tempo que os animais gastaram ruminando se mantem estável até esse mesmo ponto, mas despencaram após esse momento, demonstrando que os animais não conseguiram consumir forragem o suficiente. Já o tempo gasto em ócio é menor quando o pasto alcance 60% da altura de entrada, pois os animais gastam mais tempo pastejando, e volta a subir após esse ponto. É importante detectar a mudança no comportamento antes de alcançar esse momento, pois a partir dele os animais já estão em um pasto abaixo do recomendado.

O último interesse foi expor os modelos preditivos treinados anteriormente, em outro trabalho do nosso grupo (RIBEIRO *et al.*, 2021), ao nosso banco de dados. A validação externa é muito utilizada por empresas que querem testar seus produtos em um conjunto de dados completamente novo (GRINTER; CAMPLER; COSTA, 2019; PEREIRA, G. M. *et al.*, 2020; WEINERT; WERNER; WILLIAMS, 2020). Nosso objetivo foi observar como os modelos preditivos se comportariam ao serem expostos a um banco de dados completamente diferente do utilizado para o treinamento. A observação mais relevante é que os modelos tiveram desempenho semelhantes em todas as estratégias de validação, ou seja, as métricas de avaliação não foram infladas pela estratégia *holdout*, como aconteceu quando utilizamos dados oriundos do banco de dados original para validar.

Uma limitação encontrada durante a execução desse trabalho foi o hardware utilizado. Além de impossibilitar melhorias e adição de novos componentes, os sensores falharam no envio dos dados, mais de 50% dos dados não foram coletados e a quantidade de dados presente no banco de dados pode auxiliar no treinamento do modelo preditivo. Devido as limitações na coleta dos dados um protótipo está sendo desenvolvido em parceria com o Departamento de Automática (DAT) da Universidade Federal de Lavras. O objetivo é criar um produto básico que possa ser utilizado nas circunstâncias exigidas pela pecuária a pasto e que não prejudique a coleta de dados. O protótipo também nos dará controle total sobre o hardware e sobre os dados obtidos, possibilitando melhorias como a adição de novos sensores e componentes, que colem

diferentes informações. Essas novas informações podem revelar mais sobre as interconexões entre animal, pastagem e ambiente.

6 – CONCLUSÃO

Os modelos apresentaram maior acuraria, sensibilidade, especificidade, VPP e VPN ao predizerem dois comportamentos, pastejando e não pastejando, do que predizendo três comportamentos, pastejando, ruminando e ócio. Para classificar três comportamentos o melhor modelo preditivo foi o ANN, principalmente ao predizer o comportamento de pastejo. Já para classificar dois comportamentos os modelos RF e ANN foram similares. O modelo GLM teve o pior desempenho em todos os cenários.

A estratégia de validação adotada interferiu no desempenho preditivo dos modelos avaliados. Estratégias que utilizam um conjunto de dados para teste que, de alguma forma, são correlacionados com o banco de dados, como a *holdout*, não são representativas da realidade, pois inflam os resultados das métricas de avaliação de um modelo preditivo ao reduzir a importância de situações comuns do cotidiano, como estação do ano e entrada de novos animais no rebanho. Essas estratégias de validação limitam a utilização dos modelos preditivos em condições diferentes das quais foram treinados. As estratégias de validação adotadas devem expor os modelos preditivos a situações novas.

REFERÊNCIAS

3 DIMENSÕES, Website. *Overfitting e Underfitting*. Disponível em: <<https://www.3dimensoes.com.br/post/overfitting-e-underfitting#:~:text=Overfitting e Underfitting são as,é a relação bias-variância.>>. Acesso em: 20 maio 2021.

ABIEC. Beef Report: Perfil da pecuária no Brasil. *BeefREPORT*, v. 1, p. 49, 2020. Disponível em: <<http://www.abiec.com.br/control/uploads/arquivos/sumario2019portugues.pdf>>.

AIKEN, Vera Cardoso Ferreira *et al.* Record linkage for farm-level data analytics: Comparison of deterministic, stochastic and machine learning methods. *Computers and Electronics in Agriculture*, v. 163, n. June, p. 104857, 2019. Disponível em: <<https://doi.org/10.1016/j.compag.2019.104857>>.

ALVARENGA, F. A.P. *et al.* Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. *Applied Animal Behaviour Science*, v. 181, p. 91–99, 2016.

AMIT, Yali; GEMAN, Donald. Randomized Inquiries about Shape; an Application to Handwritten Digit Recognition. *Technical report 401*, n. 401, p. 49, 1994. Disponível em: <<http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA290812>>.

AMORIM, Maurício J V; BARONE, Dante; MANSUR, André Uebe. Técnicas de Aprendizado de Máquina Aplicadas na Previsão de Evasão Acadêmica. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 1, n. 1, p. 666–674, 2008. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/article/view/755%0Ahttp://www.br-ie.org/pub/index.php/sbie/article/view/755>>.

ANDRIAMANDROSO; HERINAINA, Andriamasinoro Lalaina; *et al.* A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. *Biotechnology, Agronomy and Society and Environment*, v. 20, n. S1, p. 273–286, 2016.

ANDRIAMANDROSO, Andriamasinoro Lalaina Herinaina; MERCATORIS, Benoît; *et al.* A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. *Biotechnology, Agronomy and Society and Environment*, v. 20, n. S1, p. 273–286, 2016.

ARABLOUEI, Reza *et al.* In-situ classification of cattle behavior using accelerometry data. *Computers and Electronics in Agriculture*, v. 183, p. 106045, 2021. Disponível em: <<https://doi.org/10.1016/j.compag.2021.106045>>.

ATKINSON, Patrick *et al.* Comparison of ATD and driver knee positions. *SAE Technical Papers*, v. 13, n. 3, p. 140–149, 2009.

AUGUSTO, Karito. *Pecuária de Precisão: A Revolução da Produtividade na Pecuária*. Disponível em: <<https://prodap.com.br/pt/blog/o-que-e-pecuaria-de-precisao>>. Acesso em: 28 abr. 2021.

BAKSHI, Chaya. *Random Forest Regression*. Disponível em: <<https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>>. Acesso em: 17 nov. 2020.

BARCELLOS, Naíme De *et al.* Comportamento ingestivo de novilhos de corte

em pastagem de aveia preta e azevém com níveis distintos de folhas verdes. *Ciência Rural*, v. 34, n. 5, p. 1543–1548, 2004.

BARWICK, Jamie *et al.* Categorising sheep activity using a tri-axial accelerometer. *Computers and Electronics in Agriculture*, v. 145, n. January, p. 289–297, 2018. Disponível em: <<https://doi.org/10.1016/j.compag.2018.01.007>>.

BHANDE, Anup. *What is underfitting and overfitting in machine learning and how to deal with it.* Disponível em: <<https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>>. Acesso em: 5 maio 2021.

BHAVSAR, Hetal; GANATRA, Amit. A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering*, v. 2, n. 4, p. 74–81, 2012.

BRAUN, Ueli; TSCHONER, Theresa; HÄSSIG, Michael. Evaluation of eating and rumination behaviour using a noseband pressure sensor in cows during the peripartum period. *BMC Veterinary Research*, v. 10, n. 1, p. 1–8, 2014.

BRESOLIN, Tiago; DÓREA, João R.R. Infrared Spectrometry as a High-Throughput Phenotyping Technology to Predict Complex Traits in Livestock Systems. *Frontiers in Genetics*, v. 11, n. August, p. 1–20, 2020.

BUNKER, Rory P.; THABTAH, Fadi. A machine learning framework for sport result prediction. *Applied Computing and Informatics*, v. 15, n. 1, p. 27–33, 2019. Disponível em: <<https://doi.org/10.1016/j.aci.2017.09.005>>.

CARBONELL, Jaime G.; MICHALSKI, Ryszard S.; MITCHELL, Tom M. *An Overview of Machine Learning.* 1983. 3–23 f. Morgan Kaufmann., 1983. Disponível em: <<http://dx.doi.org/10.1016/B978-0-08-051054-5.50005-4>>.

CARNEVALLI, R. A. *et al.* Herbage production and grazing losses in Panicum maximum cv. Mombaça under four grazing managements. *Tropical Grasslands*, v. 40, n. 3, p. 165–176, 2006.

CARVALHO, Paulo César de Faccio; MORAES, Anibal De. Comportamento ingestivo de ruminantes: bases para o manejo sustentável do pasto. *Manejo Sustentável em Pastagem*, v. 1, p. 1–20, 2005.

CASAGRANDE, Daniel Rume *et al.* Canopy characteristics and behavior of nellore heifers in brachiaria brizantha pastures under different grazing heights at a continuous stocking rate. *Revista Brasileira de Zootecnia*, v. 40, n. 11, p. 2294–2301, 2011.

CEBALLOS, Maria C. *et al.* Frequent handling of grazing beef cattle maintained under the rotational stocking method improves temperament over time. *Animal Production Science*, v. 58, n. 2, p. 307–313, 2018.

CHAPA, Jose M. *et al.* Accelerometer systems as tools for health and welfare assessment in cattle and pigs – a review. *Behavioural Processes*, v. 181, n. October, p. 104262, 2020. Disponível em: <<https://doi.org/10.1016/j.beproc.2020.104262>>.

COCKBURN, Marianne. Review: Application and prospective discussion of machine learning for the management of dairy farms. *Animals*, v. 10, n. 9, p. 1–22, 2020.

DA COSTA LIMA, Josilaine Aparecida *et al.* Ingestive diurnal behaviour of grazing beef cattle. *Semina: Ciências Agrárias*, v. 41, n. 4, p. 1335–1344, 2020.

DA SILVA, S. C.; CARVALHO, P. C. de F. Foraging behaviour and herbage intake in the favourable tropics/sub-tropics. *Grassland: A Global Resource*, n. January, p. 81–95, 2005.

DE ALMEIDA, Pedro Raposo *et al.* Testing a 3-axis accelerometer acoustic transmitter (AccelTag) on the Lusitanian toadfish. *Journal of Experimental Marine Biology and Ecology*, v. 449, p. 230–238, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.jembe.2013.09.015>>.

DE FREITAS BARBOSA, Marco Aurélio Alves *et al.* Comportamento ingestivo de bovinos mantidos em pastos de *Brachiaria brizantha* cv. Xaraés manejado em diferentes alturas de pastejo. *Semina: Ciências Agrárias*, v. 34, n. 6 SUPPL. 2, p. 4113–4120, 2013.

DE GIMENES, Flávia Maria Andrade *et al.* Ganho de peso e produtividade animal em capim-marandu sob pastejo rotativo e adubação nitrogenada. *Pesquisa Agropecuária Brasileira*, v. 46, n. 7, p. 751–759, 2011.

DE MOURA ZANINE, A. *et al.* Structural characteristics and herbage accumulation of tanzania guineagrass subjected to rotational stocking | Características estruturais e acúmulo de forragem em capim-tanzânia sob pastejo rotativo. *Revista Brasileira de Zootecnia*, v. 40, n. 11, p. 2364–2373, 2011.

DECANDIA, M. *et al.* The effect of different time epoch settings on the classification of sheep behaviour using tri-axial accelerometry. *Computers and Electronics in Agriculture*, v. 154, n. February, p. 112–119, 2018.

DIAS, Cleidson Nogueira; JARDIM, Francisco; SAKUDA, Liuz Ojima. *Radar AgTech Brasil 2019: Mapeamento das Startups do Setor Agro Brasileiro*. São Paulo: [s.n.], 2019. Disponível em: <www.radaragtech.com.br>.

DIFANTE, Gelson dos Santos *et al.* Ingestive behaviour, herbage intake and grazing efficiency of beef cattle steers on Tanzania guineagrass subjected to rotational stocking managements. *Revista Brasileira de Zootecnia*, v. 38, n. 6, p. 1001–1008, 2009.

DONALD, Jameson. What Shall We Do about Grazing Item. *Society for Range Management*, v. 8, n. 4, p. 178–179, 1986.

DÓREA, J. R.R. *et al.* Mining data from milk infrared spectroscopy to improve feed intake predictions in lactating dairy cows. *Journal of Dairy Science*, v. 101, n. 7, p. 5878–5889, 2018. Disponível em: <<http://dx.doi.org/10.3168/jds.2017-13997>>.

DOUGHERTY, James; KOHAVI, Ron; SAHAMI, Mehran. *Supervised and Unsupervised Discretization of Continuous Features*. [S.l.]: Morgan Kaufmann Publishers, Inc., 1995. Disponível em: <<http://dx.doi.org/10.1016/B978-1-55860-377-6.50032-3>>.

DUARTE, Guilherme. *Gradient Boostings Parte 1: Métodos de Ensemble Learning*. Disponível em: <<https://datarisk.io/gradient-boostings-parte-1-metodos-de-ensemble-learning/?cn-reloaded=1>>. Acesso em: 21 maio 2021.

DUCKWORTH, J.E.; SHIRLAW, D.W. The development of an apparatus to

record the jaw movements of cattle. *The British Journal of Animal Behaviour*, v. 3, n. 2, p. 56-IN2, abr. 1955. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950560155800135>>.

DUNNE, Robert *et al.* Behavior classification of dairy cows fitted with GPS collars. In: KANG, U AND LIM, EP AND YU, JX AND MOON, YS (Org.). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Lecture Notes in Artificial Intelligence. Australia: Springer Verlag, 2017. v. 10526 LNAI. p. 15–25. Disponível em: <http://link.springer.com/10.1007/978-3-319-67274-8_2>. Acesso em: 30 ago. 2019.

DUTTA, Ritaban *et al.* Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture*, v. 111, p. 18–28, 1 fev. 2015.

DWORK, Cynthia *et al.* Preserving statistical validity in adaptive data analysis. *Proceedings of the Annual ACM Symposium on Theory of Computing*, v. 14-17-June, n. 6248, p. 117–126, 2015.

ELLIS, J. L. *et al.* Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data. *Animal*, 2020.

EUCLIDES, Valéria Pacheco Batista *et al.* Beef cattle performance in response to Ipyporã and Marandu brachiariagrass cultivars under rotational stocking management. *Revista Brasileira de Zootecnia*, v. 47, n. 2000, 2018.

EUCLIDES, Valéria Pacheco Batista *et al.* Manejo do pastejo de cultivares de *Brachiaria brizantha* (Hochst) Stapf e de *Panicum maximum* Jacq. *Revista Ceres*, v. 61, p. 808–818, 2014.

FERNANDES, Tiago Albandes *et al.* Características comportamentais dos bovinos: Aspectos básicos, processo de aprendizagem e fatores que as afetam. *Revista Electronica de Veterinaria*, v. 18, n. 9, p. 1–16, 2017.

FERREIRA, Igor Machado *et al.* Twenty-five-centimeter pre-grazing canopy height in palisade grass and forage peanut. *Scientia Agricola*, v. 79, n. 2, 2022. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162022000200401&tlng=en>.

FLORES, Renata Santos *et al.* Animal performance, forage yield and structural characteristics in the palisadegrass cvs. marandu and xaraés submitted to grazing intensities. *Revista Brasileira de Zootecnia*, v. 37, n. 8, p. 1355–1365, 2008.

FRIEDMAN, Jerome H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001.

GARCIA, Marco. *Big Data: O que é, conceito e definição*. Disponível em: <<https://www.cetax.com.br/blog/big-data/>>. Acesso em: 28 abr. 2021.

GOMES, Dennis dos Santos. Inteligência Artificial: Conceitos e Aplicações. *Revista Olhar Científico*, v. 1, p. 234–246, 2010.

GONTIJO NETO, Miguel Marques *et al.* Consumo e tempo diário de pastejo por novilhos Nelore em pastagem de capim-tanzânia sob diferentes ofertas de forragem. *Revista Brasileira de Zootecnia*, v. 35, n. 1, p. 60–66, 2006.

GRINTER, L. N.; CAMPLER, M. R.; COSTA, J. H.C. Technical note: Validation of a behavior-monitoring collar's precision and accuracy to measure rumination, feeding, and resting time of lactating dairy cows. *Journal of Dairy Science*, v. 102, n. 4, p. 3487–3494, 2019. Disponível em: <<http://dx.doi.org/10.3168/jds.2018-15563>>.

GROVER, Prince. *Gradient Boosting from scratch*. Disponível em: <<https://blog.mlreview.com/gradient-boosting-from-scratch-1e317ae4587d>>. Acesso em: 21 maio 2021.

GUO, Leifeng *et al.* Comparison of grazing behaviour of sheep on pasture with different sward surface heights using an inertial measurement unit sensor. *Computers and Electronics in Agriculture*, v. 150, n. September 2017, p. 394–401, 2018. Disponível em: <<https://doi.org/10.1016/j.compag.2018.05.004>>.

HAMILTON, Andrew W. *et al.* Identification of the rumination in cattle using support vector machines with motion-sensitive bolus sensors. *Sensors (Switzerland)*, v. 19, n. 5, 2019.

HASTIE, Trevor; TIBSHRANI, Robert; FRIEDMAN, Jerome. *The Elements of Statistical Learning*. 2. ed. Palo alto - CA: Springer, 2008.

HENDRICKSEN, R; MINSON, D. J. The Feed Intake and Grazing Behaviour of Cattle Grazing a Crop of Lablab Purpureus CV. Rongai. *The Journal of Agricultural Science*, v. 95, n. 3, p. 547–554, 1980.

HOMBURGER, Hermel *et al.* Inferring behavioral states of grazing livestock from high-frequency position data alone. *PLoS ONE*, v. 9, n. 12, p. 1–22, 2014.

JAMES, Gareth *et al.* *An Introduction to Statistical Learning*. New York, NY: Springer New York, 2013. v. 103. Disponível em: <<http://link.springer.com/10.1007/978-1-4614-7138-7>>. (Springer Texts in Statistics).

JENSEN, P. *The ethology of domestic animals: an introductory text*. Wallingford: CABI, 2017. Disponível em: <<http://www.cabi.org/cabebooks/ebook/20173282665>>.

JUNIOR BERTOZZO, Richard. *Aplicação de machine learning em dataset de consultas médicas do sus*. 2019. 100 f. Universidade Federal de Santa Catarina, 2019.

KABUGO, James Clovis *et al.* Industry 4.0 based process data analytics platform: A waste-to-energy plant case study. *International Journal of Electrical Power & Energy Systems*, v. 115, n. November 2018, p. 105508, 2020. Disponível em: <<https://doi.org/10.1016/j.ijepes.2019.105508>>.

KALAPANIDAS, E. *et al.* Machine Learning Algorithms: A study on noise sensitivity. *Proc. 1st Balcan Conference in Informatics*, n. OCTOBER, p. 356–365, 2003. Disponível em: <https://perun.pmf.uns.ac.rs/old/radovanovic/dmsem/cd/install/Weka/doc/citeseer-papers/C76_Kalapanidas_BCI2003.pdf>.

KAMIŃSKI, Bogumił; JAKUBCZYK, Michał; SZUFEL, Przemysław. A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, v. 26, n. 1, p. 135–159, 2018.

KENJI, Bruno. *Machine Learning para Leigos*. Disponível em:

<<https://www.venturus.org.br/machine-learning-para-leigos/>>. Acesso em: 5 maio 2021.

KOHAVI, Ron. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference of Artificial Intelligence*, n. March 2001, 1995.

LACA, Emilio A. Precision livestock production: tools and concepts. *Revista Brasileira de Zootecnia*, v. 38, n. spe, p. 123–132, jul. 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-35982009001300014&lng=en&tlng=en>.

LAMPERT, Vinícius Do Nascimento *et al.* Modelling beef cattle production systems from the pampas in brazil to assess intensification options. *Scientia Agricola*, v. 77, n. 4, 2020.

LARRIVA-NOVO, Xavier *et al.* Efficient Distributed Preprocessing Model for Machine Learning-Based Anomaly Detection over Large-Scale Cybersecurity Datasets. *Applied Sciences*, v. 10, n. 10, p. 3430, 15 maio 2020. Disponível em: <<https://www.mdpi.com/2076-3417/10/10/3430>>.

LI, Qing; LINY, Nan. The Bayesian elastic net. *Bayesian Analysis*, v. 5, n. 1, p. 151–170, 2010.

LÓPEZ, Fernando. *Ensemble Learning: Bagging & Boosting*. Disponível em: <<https://towardsdatascience.com/ensemble-learning-bagging-boosting-3098079e5422>>. Acesso em: 21 maio 2021.

MANICA, Rafael. Aplicação de uma rede neural artificial simplificada para a identificação de degradação de depósitos turbidítico. *Geociências*, v. 32, n. 3, p. 429–440, 2013.

MANSBRIDGE, Nicola *et al.* Feature selection and comparison of machine learning algorithms in classification of grazing and rumination behaviour in sheep. *Sensors (Switzerland)*, v. 18, n. 10, p. 1–16, 2018.

MICROSOFT. *Tasks to prepare data for enhanced machine learning*. Disponível em: <<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/prepare-data>>. Acesso em: 5 maio 2021.

MONNIT®. *Remote Monitoring for Business Wireless Accelerometer - G-Force Snapshot Sensor Features of Monnit ALTA Sensors*. . Salt Lake City: [s.n.], 2021.

NADIMI, E. S. *et al.* Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Computers and Electronics in Agriculture*, v. 82, p. 44–54, mar. 2012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0168169911003206>>. Acesso em: 30 ago. 2019.

NEETHIRAJAN, Suresh; KEMP, Bas. Digital twins in livestock farming. *Animals*, v. 11, n. 4, 2021.

NEPOMUCENO, Giane Lima. *Como funciona o pastejo rotacional*. Disponível em: <http://boiapasto.com.br/noticias/como-funciona-pastejo-rotacional/8688/1#.YKcM5_zPyUI>. Acesso em: 20 maio 2021.

NOGARE, Diego. *Performance de Machine Learning – Matriz de Confusão*.

Disponível em: <<http://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/>>. Acesso em: 17 nov. 2020.

OKUI, Tatsuya *et al.* Nutrient dynamics under different regimes of stocking and cattle type of temperate pastures in Hokkaido, Japan. *Grassland Science*, n. September 2019, p. 1–16, 2020.

OLIVEIRA, Neto. *Passo a passo Machine Learning*. Disponível em: <<https://medium.com/ensina-ai/passos-a-passos-machine-learning-teoria-47f032e0db35>>. Acesso em: 21 maio 2021.

PALHANO, Ana Luisa *et al.* Padrões de deslocamento e procura por forragem de novilhas leiteiras em pastagem de capim-mombaça. *Revista Brasileira de Zootecnia*, v. 35, n. 6, p. 2253–2259, 2006.

PARK, Jung Kyu; PARK, Eun Young. Monitoring method of movement of grazing cows using cloud-based system. *ECTI Transactions on Computer and Information Technology*, v. 15, n. 1, p. 24–33, 2021.

PEDREIRA, Bruno Carneiro; PEDREIRA, Carlos Guilherme Silveira; DA SILVA, Sila Carneiro. Sward structure and herbage accumulation in *Brachiaria brizantha* cultivar Xaraés in response to strategies of grazing. *Pesquisa Agropecuária Brasileira*, v. 42, n. 2, p. 281–287, 2007.

PEREIRA, G. M. *et al.* Validation of an ear tag-based accelerometer system for detecting grazing behavior of dairy cows. *Journal of Dairy Science*, v. 103, n. 4, p. 3529–3544, 2020.

PEREIRA, Lilian Elgalise Techio; POLIZEL, Guilherme Henrique Gebim. *Princípios e recomendações para o manejo de pastagens*. [S.l: s.n.], 2016.

RAHMAN, A. *et al.* Cattle behaviour classification from collar, halter, and ear tag sensors. *Information Processing in Agriculture*, v. 5, n. 1, p. 124–133, 1 mar. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2214317317301099>>. Acesso em: 28 ago. 2019.

REFAEILZADEH, Payam; TANG, Lei; HUAN, Liu. *Cross-Validation*. Disponível em: <<https://web.archive.org/web/20110905044421/http://www.public.asu.edu/~ltang9/papers/ency-cross-validation.pdf>>. Acesso em: 15 nov. 2020.

RIBEIRO, Leonardo Augusto Coelho. *Approaches of machine learning and validation strategies to predict grazing behavior in beef cattle using sensors*. 2020. 64 f. Universidade Federal de Lavras, 2020.

RIBEIRO, Leonardo Augusto Coelho *et al.* Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data. *Journal of Animal Science*, 5 jul. 2021. Disponível em: <<https://doi.org/10.1093/jas/skab206>>.

RIBEIRO, Marco Antônio. *Automação Industrial*. 4. ed. Salvador: Tek Treinamento & Consultoria Ltda, 1999. v. 4.

RIBEIRO, Marco Antônio. *Instrumentação 8 a edição*. 8. ed. Salvador: [s.n.], 1978.

ROCHA GÓIS, Karen Camille *et al.* Using an observer rating method to assess the effects of rotational stocking method on beef cattle temperament over time. *Revista Brasileira de Zootecnia*, v. 45, n. 9, p. 501–508, 2016.

SAGI, Omer; ROKACH, Lior. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 4, p. 1–18, 2018.

SARMENTO, Daniel Oliveira de Lucena. Comportamento Ingestivo De Bovinos Em Pastos De Capim Marandu Submetidos a Regimes De Lotação Contínua. p. 76, 2003.

SAVIAN, Jean Víctor *et al.* ‘Rotatinuous’ stocking as a climate-smart grazing management strategy for sheep production. *Science of the Total Environment*, v. 753, p. 141790, 2021. Disponível em: <<https://doi.org/10.1016/j.scitotenv.2020.141790>>.

SCIKIT-LEARN.ORG. 3.1. *Cross-validation: evaluating estimator performance.* Disponível em: <https://scikit-learn.org/stable/modules/cross_validation.html>. Acesso em: 17 nov. 2020.

SHAFIULLAH, Abu Zar *et al.* Machine learning based prediction of insufficient herbage allowance with automated feeding behaviour and activity data. *Sensors (Switzerland)*, v. 19, n. 20, p. 1–19, 2019.

SILVA, Émerson. *Introdução ao Aprendizado de Máquina.* Disponível em: <<https://medium.com/lica-ufal/introdução-ao-aprendizado-de-máquina-2440b31915e3>>. Acesso em: 5 maio 2021.

SILVA, S. C.; NASCIMENTO JÚNIOR, Domicio Do. Avanços na pesquisa com plantas forrageiras tropicais em pastagens: características morfofisiológicas e manejo do pastejo. *Revista Brasileira de Zootecnia*, v. 36, n. suppl, p. 122–138, 2007.

SILVEIRA, Márcia Cristina Teixeira Da *et al.* Herbage accumulation and grazing losses on Mulato grass subjected to strategies of rotational stocking management. *Scientia Agricola*, v. 70, n. 4, p. 242–249, 2013.

SMITH, Daniel *et al.* Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. *Computers and Electronics in Agriculture*, v. 131, p. 40–50, dez. 2016.

TAY, J. Kenneth; NARASIMHAN, Balasubramanian; HASTIE, Trevor. Elastic Net Regularization Paths for All Generalized Linear Models. 5 mar. 2021. Disponível em: <<http://arxiv.org/abs/2103.03475>>.

THARWAT, Alaa. Classification assessment methods. *Applied Computing and Informatics*, 2018.

TRINDADE, Júlio Kuhn Da *et al.* Composição morfológica da forragem consumida por bovinos de corte durante o rebaixamento do capim-marandu submetido a estratégias de pastejo rotativo. *Pesquisa Agropecuária Brasileira*, v. 42, n. 6, p. 883–890, 2007.

UYVAL, Alper Kursat; GUNAL, Serkan. The impact of preprocessing on text classification. *Information Processing and Management*, v. 50, n. 1, p. 104–112, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.ipm.2013.08.006>>.

VALLETTA, John Joseph *et al.* Applications of machine learning in animal behaviour studies. *Animal Behaviour*, v. 124, p. 203–220, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.anbehav.2016.12.005>>.

VASCONCELLOS, Paulo. *Como saber se seu modelo de Machine Learning está funcionando mesmo*. Disponível em: <<https://paulovasconcellos.com.br/como-saber-se-seu-modelo-de-machine-learning-está-funcionando-mesmo-a5892f6468b>>. Acesso em: 17 nov. 2020.

WATSON-HAIGH, Nathan S.; O'NEILL, Christopher J.; KADARMIDEEN, Haja N. Proximity loggers: Data handling and classification for quality control. *IEEE Sensors Journal*, v. 12, n. 5, p. 1611–1617, 2012. Disponível em: <<http://ieeexplore.ieee.org/document/6072229/>>.

WEINERT, Jennifer R.; WERNER, Jessica; WILLIAMS, Carey A. *Validation and Implementation of an Automated Chew Sensor–Based Remote Monitoring Device as Tool for Equine Grazing Research*. *Journal of Equine Veterinary Science*. [S.l: s.n.], 2020

YANG, Xin *et al.* Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chemical Reviews*, v. 119, n. 18, p. 10520–10594, 2019.

ZANINI, Guilherme Doneda; SANTOS, Gabriela Trevisan; SBRISSIA, André Fischer. Frequencies and intensities of defoliation in aruana guineagrass swards: Accumulation and morphological composition of forage. *Revista Brasileira de Zootecnia*, v. 41, n. 4, p. 905–913, 2012.

APÊNDICE

Tabela A.1 – Sumário dos artigos revisados.

Reference	Country	Species	n	Sensor position	Sensor	Algorithm
Dohi et al., 1992	Japan	Beef cattle	5	Low back	Others	Sensor default prediction
Champion et al., 1996	UK	Sheep	8	Leg	Others	Sensor default prediction
Ungar et al., 2005	Israel	Beef cattle	6	Collar	GPS	Classification tress
Scheibe et al., 2006	Germany	Dairy cattle	4	Collar / Low back	Accelerometer	Sensor default prediction
Brosh et al., 2006	Israel	Dairy cattle	21	Collar	GPS	Stepwise regression
Guo et al., 2006	Australia	Dairy cattle	6	Collar	Accelerometer / GPS	Sensor default prediction
Schwager et al., 2007	USA	Beef cattle	3	Satellite	GPS	K-means
Ipema et al., 2008	Netherlands	Dairy cattle	Não especificado	Rumen	Bolus	Sensor default prediction
Umstatter et al., 2008	UK	Sheep	10	Collar	GPS	LDA / GLM / Decision tree
Moreau et al., 2009	Germany	Goat	2	Neck	Accelerometer	Sensor default prediction
Tolkamp et al., 2010	UK	Beef cattle	4	Leg	IceTag	Sensor default prediction
Gregorini et al., 2011	NZ	Dairy cattle	36	Ear / Neck	Accelerometer / GPS / Others	Sensor default prediction
Nadimi et al., 2011	Denmark	Sheep	11	Neck	Accelerometer	Sensor default prediction
Nadimi et al., 2011	Denmark	Sheep	11	Neck	Accelerometer	Artificial Neural Network
Mason et al., 2013	UK	Sheep	4	Neck	Accelerometer	Sensor default prediction
Navon et al., 2013	Israel	Goat / Sheep / Dairy cattle	30	Neck	Others	Sensor default prediction
Valente et al., 2013	Brazil	Beef cattle	12	Neck	GPS	GLM
Nielsen, P.P. 2013	Switzerland	Dairy Cattle	20	Halter	Accelerometer / Gyroscope	Sensor default prediction
Tani et al., 2013	Japan	Beef cattle	4	Halter	Accelerometer / Others	Sensor default prediction

Augustine et al., 2013	USA	Beef cattle	12	Collar	GPS	Decision tree
Ensing et al., 2014	Netherlands	Wild animals	Não especificado	Satellite	GPS	Pearson Correlation
Greenwood et al., 2014	Australia	Beef cattle	Não especificado	Ear / Neck	Accelerometer / Others / GPS	Sensor default prediction
Dutta et al., 2014	Australia	Dairy cattle	24	Collar	Accelerometer / Others / GPS	KNN / Decision Tree / Naïve Bayes / LDA
Dutta et al., 2014	Australia	Dairy cattle	24	Collar	Accelerometer / Others / GPS	KNN / Decision Tree / Naïve Bayes / LDA
González et al., 2014	Australia	Beef cattle	20	Collar	Accelerometer / Others / GPS	Linear regression
Homburger et al., 2014	Switzerland	Dairy cattle	3	Collar	GPS	Random Forest / SVM / LDA
Pegorini et al., 2015	Brazil	Dairy cattle	1	Halter	Others	Decision tree
Draganova et al., 2015	NZ	Dairy cattle	24	Collar	GPS / Others	Kernel smoothing
González et al., 2015	Australia	Beef cattle	Não especificado	Collar	Accelerometer / Others / GPS	Decision tree
Williams et al., 2016	UK	Dairy cattle	40	Satellite	GPS	Naïve Bayes / Random Forest
Alvarenga et al., 2016	Brazil	Sheep	4	Halter	Accelerometer	Random forest
Smith et al., 2016	Australia	Dairy cattle	24	Collar	Accelerometer / GPS	SVM / KNN / Naïve Bayes / Linear regression / RFE
Andriamasinoro et al., 2017	Belgium	Dairy cattle	19	Neck	Accelerometer / Gyroscope	Signal processing
Molfin et al., 2017	Australia	Dairy cattle	8	Neck	Accelerometer	GLM / Pearson correlation
Borchers et al., 2017	EUA	Dairy cattle	53	Coolar / Leg	Accelerometer / Others	Random forest / Linear discriminant / Artificial Neural network
Greenwood et al., 2017	Australia	Beef cattle	10	Collar	Accelerometer / Others	Linear regression

Virgilio et al., 2018	Argentina	Sheep	3	Neck	Accelerometer / Others / GPS	Sensor default prediction
Mulvenna et al., 2018	UK	Beef cattle / Sheep / Goat	6	Halter	Accelerometer / Others	GLM
Kiani, F. 2018	Turkey	Dairy cattle	5	Leg	Accelerometer	Artificial Neural Network
Nielsen et al., 2018	Denmark	Dairy cattle	10	Leg	CowScout Leg / IceTag	Sensor default prediction
Ungar et al., 2018	Israel	Beef cattle	800	Leg	IceTag	Sensor default prediction
Mansbridge et al., 2018	UK	Sheep	6	Ear / Neck	Accelerometer / Gyroscope	Random Forest / SVM / KNN / Adaptive boost
Rahman et al., 2018	Australia	Dairy cattle	Não especificado	Ear / Neck	Accelerometer / Others	Random Forest
Pereira et al., 2018	EUA	Dairy cattle	24	Ear	CowManager	Sensor default prediction
Roland et al., 2018	Austria	Dairy cattle	3	Ear	Accelerometer	Pearson correlation
Roland et al., 2018	Austria	Dairy cattle	3	Ear	Accelerometer	Pearson correlation
Decandia et al., 2018	Italy	Sheep	48	Halter	Accelerometer	
Lush et al., 2018	UK	Sheep	30	Low back	Accelerometer	Random Forest
Guo et al., 2018	China	Sheep	3	Collar	Accelerometer / Gyroscope / Others	LDA
Jemila et al., 2018	India	Dairy cattle	30	Collar	Accelerometer	Sensor default prediction
Werner et al., 2018	Ireland	Dairy cattle	12	Collar	Accelerometer	Sensor default prediction
Shafiulla et al., 2019	Germany	Dairy cattle	40	Leg / Collar	Accelerometer / Others	GLM / Others
Clothier et al., 2019	Australia	Horse	8	Leg	Accelerometer	Linear regression
Maroto-Molina et al., 2019	Spain	Beef cattle / Sheep	18	Ear / Neck	GPS / BLE tags	Sensor default prediction
Jaeger et al., 2019	Germany	Dairy cattle	49	Ear	Accelerometer	Pearson correlation

Zambelis et al., 2019	Canada	Dairy cattle	10	Ear	Accelerometer	Sensor default prediction
Achour et al., 2019	Algeria	Dairy cattle	8	Low back	Accelerometer / Gyroscope	Decision tree
Sakai et al., 2019	Japan	Goat	3	Low back	Accelerometer	KNN / Decision tree
Grinter et al., 2019	USA	Dairy cattle	24	Collar	Accelerometer	Pearson correlation
Poulopoulou et al., 2019	Italy	Beef cattle	6	Collar	RumiWatch / Accelerometer	GLM
Werner et al., 2019	Ireland	Dairy cattle	12	Collar	Accelerometer	Sensor default prediction
Werner et al., 2019	Ireland	Dairy cattle	12	Collar	Accelerometer	Sensor default prediction
Zehner et al., 2019	Switzerland	Dairy cattle	35	Collar	RumiWatch	Sensor default prediction
Campos et al., 2019	Brazil	Goat	3	Masseter muscle	Others	Linear regression
Cairo et al., 2020	Brazil	Dairy cattle	99	Ear	Others	GLM / Artificial Neural Network / Random forest
Fogarty et al., 2020	Australia	Sheep	39	Ear	Accelerometer	SVM / CART / LDA / QDA
Fogarty et al., 2020	Australia	Sheep	39	Ear	Accelerometer	SVM / CART / LDA / QDA
Dela Rue et al., 2020	NZ	Dairy cattle	37	Neck	Nedap Livestick	Sensor default prediction
Hu et al., 2020	Australia	Sheep	20	Neck	Accelerometer	Random Forest / SVM / LDA
Hut et al., 2020	Netherlands	Dairy cattle	1326	Neck	Accelerometer / Others	GLM
Ikurior et al., 2020	NZ	Sheep	24	Neck	Accelerometer	GLM
Riaboff et al., 2020	France	Dairy cattle	71	Neck	Accelerometer / GPS	Gradient Boosting
Smith et al., 2020	Australia	Sheep	76	Neck	Accelerometer	Sensor default prediction
Wagner et al., 2020	France	Dairy cattle	28	Neck	Others	KNN / Decision tree / MLP /

						Artificial Neural network
Williams et al., 2020	Australia	Beef cattle	8	Neck	Accelerometer	Sensor default prediction
Nóbrega et al., 2020	Portugal	Sheep	1	Collar	iFarmtec	Decision trees
Suparwito et al., 2020	Australia	Sheep	3	Collar	GPS	DNB
Weinert et al., 2020	EUA	Horse	8	Collar	Accelerometer	Sensor default prediction
Park et al., 2021	South Korea	Beef cattle	Não especificado	Many	ZebraNet	Sensor default prediction
Eerdeken et al., 2021	Belgium	Horse	6	Leg	Accelerometer	Random forest / Artificial Neural network / CNN
Verdon et al., 2021	Australia	Dairy cattle	8	Leg	Others	LSD
Fogarty et al., 2021	Australia	Sheep	39	Ear	Accelerometer	SVM / CART / LDA / QDA
Arablouei et al., 2021	Australia	Beef cattle	10	Neck	Accelerometer	Decision tree / Nearest Neighbors
Lama et al., 2021	Argentina	Goat	1	Drone	Accelerometer / Gyroscope	Random Forest
Pereira et al., 2021	EUA	Dairy cattle	12	Halter / Leg	RumiWatch	Sensor default prediction