

Jose Fernando de Carvalho Ferreira
Orientador: Romuere Rodrigues Veloso e Silva

**Uma Ferramenta para Predição de Peso
Corporal em Pequenos Ruminantes em
Rebanhos Baseada em Medidas Lineares e
Algoritmos de Regressão**

Picos - PI
22 de Janeiro de 2024

Jose Fernando de Carvalho Ferreira
Orientador: Romuere Rodrigues Veloso e Silva

Uma Ferramenta para Predição de Peso Corporal em Pequenos Ruminantes em Rebanhos Baseada em Medidas Lineares e Algoritmos de Regressão

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação. Orientador: Prof. Dr. Romuere Rodrigues Veloso e Silva

Universidade Federal do Piauí
Campus Senador Heuvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
22 de Janeiro de 2024

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

F383f Ferreira, José Fernando de Carvalho.

Uma ferramenta para predição de peso corporal em pequenos ruminantes em rebanhos baseada em medidas lineares e algoritmos de regressão./ José Fernando de Carvalho Ferreira. – 2024.

39 f.

1 Arquivo em PDF

Indexado no catálogo *online* da biblioteca José Albano de Macêdo-CSHNB
Aberto a pesquisadores, com restrições da Biblioteca

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Piauí, Curso de Bacharelado em Sistemas de Informação, Picos, 2024.

“Orientação: Prof. Dr. Romuere Rodrigues Veloso e Silva”

1. Ruminantes-peso. 2. Medidas lineares. 3. Previsão corporal.
- I. Ferreira, José Fernando de Carvalho. II. Silva, Romuere Rodrigues Veloso e. III. Título.

CDD 614.362

UMA FERRAMENTA PARA PREDIÇÃO DE PESO CORPORAL EM PEQUENOS
RUMINANTES EM REBANHOS BASEADA EM MEDIDAS LINEARES E
ALGORITMOS DE REGRESSÃO

JOSE FERNANDO DE CARVALHO FERREIRA

Monografia aprovada como exigência parcial para obtenção do grau de Bacharel em Sistemas
de Informação.

Data de Aprovação

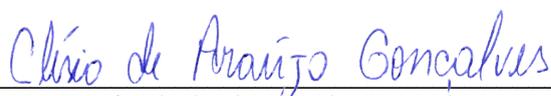
Picos – PI, 01 de fevereiro de 2024



Prof. Romuere Rodrigues Veloso e Silva



Prof. Júlio Vitor Monteiro Marques



Prof. Clésio de Araújo Gonçalves

Agradecimentos

Desejo expressar minha gratidão começando pelos alicerces da minha jornada, meus pais, Edlene e Antônio. Sem o amor, apoio e dedicação deles, eu não teria chegado até aqui. São a luz que guiou meus passos desde o início, e por isso, meu coração transborda de agradecimento.

À Daniela, um presente valioso que a UFPI me concedeu. Além dos conhecimentos compartilhados, ela se tornou um pilar essencial em minha vida acadêmica. Sempre presente nos momentos de necessidade, sei que posso contar com ela para o que der e vier. É um privilégio ter alguém tão especial ao meu lado.

Ao David, agradeço por sua generosidade e ajuda inestimável nos momentos em que mais precisei. Sua colaboração foi crucial e, por isso, serei eternamente grato. E também ao Ericksulino, Miqueias e Jansen pelas conversas e risadas pelos corredores.

Não posso deixar de mencionar o Professor Dr. Romuere Rodrigues Veloso e Silva, que proporcionou a oportunidade de ampliar meus horizontes na IC/IT. Sua orientação e incentivo foram fundamentais para o meu crescimento acadêmico. Agradeço por sua dedicação em compartilhar conhecimento.

Às incansáveis tias da limpeza, que sempre recebem todos com um sorriso no rosto. Seu trabalho muitas vezes passa despercebido, mas é fundamental para manter o ambiente acolhedor e propício ao aprendizado. Obrigado por tudo que fazem, pois vocês tornam nosso dia mais leve.

A todos aqueles que, diariamente, se esforçam para manter esse lugar especial, expresso minha sincera gratidão. Seja através do trabalho nos bastidores ou do compromisso constante, cada um de vocês contribui para a construção e preservação deste espaço que se tornou tão significativo em minha vida.

Olhando para trás, percebo que minha jornada foi de apoio, amizade e oportunidades únicas. Gostaria de agradecer a todos que se juntaram a mim nesta jornada e fizeram dela uma experiência inesquecível.

A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.

Arthur Schopenhauer

Resumo

Esta pesquisa é dedicada à previsão de peso em pequenos ruminantes e ao desenvolvimento de uma ferramenta para ajudar os produtores a monitorar o crescimento animal. Neste estudo, utilizamos técnicas avançadas para selecionar cuidadosamente recursos e ajustar parâmetros para atingir com precisão o objetivo de previsão de peso.

Dentre os modelos de regressão implementados, destacam-se o Linear, Ridge, MLP (Multilayer Perceptron) e Random Forest. A pesquisa revelou que os modelos linear e Ridge se destacaram, alcançando um coeficiente de determinação (R^2) de 85%. Esses resultados indicam a eficácia desses modelos na predição precisa do peso dos pequenos ruminantes.

A abordagem sistemática não só forneceu informações valiosas que definem as características da previsão de peso, mas também levou à criação de uma ferramenta prática e funcional. A ferramenta, desenvolvida com base no modelo identificado, oferece aos produtores uma solução eficaz para otimizar a gestão do rebanho e permite a tomada de decisões estratégicas informadas. Este estudo não só melhorou significativamente a compreensão da previsão do peso corporal em pequenos ruminantes, mas também traduziu este conhecimento numa ferramenta aplicável, promovendo assim a modernização e eficiência da produção pecuária.

Palavras-chaves: Previsão de peso, Modelos de regressão, Medidas lineares.

Abstract

This research is dedicated to predicting the weight of small ruminants and developing a tool to assist producers in monitoring animal growth. In this study, we employed advanced techniques to carefully select features and adjust parameters to accurately achieve the weight prediction goal.

Among the implemented regression models, the Linear, Ridge, MLP (Multilayer Perceptron), and Random Forest models stand out. The research revealed that the Linear and Ridge models excelled, achieving a coefficient of determination (R^2) of 85%. These results indicate the effectiveness of these models in precisely predicting the weight of small ruminants.

The systematic approach not only provided valuable insights defining the characteristics of weight prediction but also led to the creation of a practical and functional tool. The tool, developed based on the identified model, offers producers an effective solution to optimize herd management and enables informed strategic decision-making. This study not only significantly improved the understanding of body weight prediction in small ruminants but also translated this knowledge into an applicable tool, thereby promoting the modernization and efficiency of livestock production.

Lista de ilustrações

Figura 1 – Demonstração de tipos de Aprendizado de Máquina. Fonte: (MAIA, 2020)	15
Figura 2 – Demonstração da Tela Inicial.	32
Figura 3 – Demonstração da Tela de Predição.	33
Figura 4 – Demonstração da Tela de Instruções.	34
Figura 5 – Demonstração da Tela Sobre	34

Lista de tabelas

Tabela 1 – Trabalhos relacionados	24
Tabela 2 – Resultados obtidos com os modelos de regressão em diferentes conjuntos de características utilizando a métrica R^2	30
Tabela 3 – As medidas com melhores resultados.	31
Tabela 4 – Parâmetros com melhores resultados.	31

Lista de abreviaturas e siglas

MAE	Erro Médio Absoluto
MSE	Erro Quadrático Médio
RMSE	Raiz do Erro Quadrático Médio
MAPE	Erro Percentual Absoluto Médio
R ²	Coefficiente de Determinação
MLP	Multilayer perceptron
IBGE	Instituto Brasileiro de Geografia e Estatística
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
PT	Perímetro Torácico
PB	Perímetro de Barril
AG	Altura de Garupa
LG	Largura de Garupa
CG	Comprimento da Garupa
CMP	Compacidade de Carcaça
AC	Altura da Cernelha
CC	Comprimento Corporal
PV	Peso Vivo
VC	Volume corporal
CT	Circunferência do Tórax
P	Peso
CC	Comprimento Corporal
LT	Largura do tórax
LL	Largura lombar
D_Cer-5VT	Dist. da Cernelha a 5 ^a vertebra

D_5VT-13VT	Dist. da 5 ^a vertebra torácica a 13 ^a vertebra torácica
D_13VT-1S	Dist. da 13 ^a vertebra torácica a 1 ^a vertebra sacral
D_1S-IC	Dist. 1 ^a vertebra sacral a inserção da cauda
D_EE	Dist. entre as espinhas da escapula
CP	Comprimento da perna
LP	Largura da perna
CCox	Comprimento da coxa
CAb	Comprimento do antebraço
LAB	Largura do antebraço
CB	Comprimento do braço (úmero)
CEsc	Comprimento da escápula
D_Esc-IL	Dist. entre a espinha da escápula- íleo
D_Esc-IS	Dist. entre a articulação escápula-umeral e ísquio
D_Esc-FT	Dist. entre art. escápula-umeral e art. Fêmur-tíbia
AT	Altura do Tórax
AL	Altura Lombar
PCF	Peso da carcaça fria
RCF	Rendimento da carcaça fria
CIC	Comprimento interno da carcaça
CEC	Comprimento externo da carcaça
PG	Perímetro da garupa
ECCPi	Escore de condição corporal
CCFi	Comprimento corporal final
PTFi	Perímetro torácico final
PCV	Peso de corpo vazio
PVA	Peso vivo ao abate

PCQ	Peso de carcaça quente
RCQ	Rendimento de carcaça quente
ICC	Índice de compacidade da perna
PER	Pernil
LOM	Lombo
COS	Costela
PES	Pescoço
PAL	Paleta
RPER	Rendimento do pernil
RLOM	Rendimento do lombo
RCOS	Rendimento da costela
RPES	Rendimento do pescoço
RPAL	Rendimento da paleta
EC	Escore corpora
LARG	Largura do músculo Longissimus dorsi
PROF	Profundidade do músculo Longissimus dorsi
EG	Espessura de gordura subcutânea
AA	Altura de anterior
MUSC	Medida encontrada pela razão da largura e profundidade do músculo Longissimus dorsi

Sumário

1	Introdução	13
1.1	Objetivos	13
1.1.1	Objetivos Específicos	14
1.2	Contribuições	14
1.3	Estrutura do trabalho	14
2	Referencial Teórico	15
2.1	Aprendizado de Máquina	15
2.2	Algoritmo de Regressão	16
2.3	Métricas de regressão	18
2.3.1	Erro Médio Absoluto (MAE)	18
2.3.2	Erro Quadrático Médio (MSE)	18
2.3.3	Raiz do Erro Quadrático Médio (RMSE)	18
2.3.4	Erro Percentual Absoluto Médio (MAPE)	19
2.3.5	Coefficiente de Determinação (R^2)	19
3	Trabalhos Relacionados	20
4	Metodologia Proposta	25
4.1	Conjunto de Dados	25
4.2	Pré-Processamento	25
4.3	Seleção de Características	26
4.4	Modelos implementados	26
4.5	Treinamento e Métricas	27
4.6	Desenvolvimento da Ferramenta	29
4.6.1	Escolha da Linguagem de Programação	29
4.6.2	Kivy e KivyMD para Interfaces Gráficas	29
4.6.3	Prototipação Visual	29
5	Resultados	30
5.1	Resultado dos modelos	30
5.2	Resultados do desenvolvimento da ferramenta	31
5.2.1	Interação com algoritmos de regressão	31
5.2.2	Estilo visual e capacidade de resposta	31
5.2.3	Tela Inicial	31
5.2.4	Tela Predição de Peso	31

5.2.5	Tela Instruções	32
5.2.6	Tela Sobre	32
6	Conclusão	35
	Referências	37

1 Introdução

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil possuía em 2022 um rebanho de aproximadamente 12,3 milhões de caprinos (*Capra hircus*) e 21,5 milhões de ovinos (*Ovis aries*). Essas duas espécies compõem uma categoria de produção conhecida como pequenos ruminantes e são exploradas comercialmente na ovino-caprinocultura. Segundo a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), a comercialização de ovinos e caprinos entre 2017 e 2022 teve um acréscimo de 56.21% e 49.82%, respectivamente. Apesar dos crescentes investimentos nos últimos anos, a ovino-caprinocultura no Brasil ainda não chegou a patamares comparáveis aos da bovinocultura, que representa uma das atividades pecuárias responsáveis pelo destaque do país no cenário mundial, com um rebanho de mais de 200 milhões de animais (MARTINS et al., 2020). Utilizando-se como base a bovinocultura brasileira, é possível estimar que a ovino-caprinocultura ainda tem espaço para crescer substancialmente no setor de produção de proteína animal (SONG et al., 2018).

Na literatura encontra-se o trabalho de (NASCIMENTO, 2019) o estudo aborda a importância do acompanhamento individual de animais de produção, especialmente o peso corporal diário. Para superar limitações das balanças estáticas, investimentos têm sido feitos em alternativas, como a tecnologia *Walk-over-Weighing* (WoW). É analisado o desempenho de um protótipo de balança automática WoW voltada para pequenos ruminantes. Duas fases foram conduzidas, envolvendo ovelhas Lacaune e Romane. A primeira fase calibrou o equipamento, enquanto a segunda expôs 237 ovelhas ao WoW durante 29 dias. Métodos estatísticos, incluindo a análise de concordância de Bland e Altman e o coeficiente de correlação de concordância de Lin, foram utilizados para validar o protótipo. Após período de adaptação, os animais mostraram comportamento desejado, e a correlação entre o peso corporal obtido pela balança WoW e o peso de referência foi alta (entre 0,95 e 0,99), indicando precisão e confiabilidade do método.

1.1 Objetivos

O objetivo geral deste trabalho é desenvolver uma ferramenta de aprendizagem de máquina destinada a prever o peso corporal de pequenos ruminantes, com o propósito de simplificar a identificação do peso desses animais, resultando na redução de custos e economia de tempo nos processos de monitoramento em fazendas.

1.1.1 Objetivos Específicos

- Aplicar algoritmos de aprendizagem de máquina para aprimorar a predição do peso dos pequenos ruminantes.
- Desenvolver uma metodologia de execução para a análise do peso no contexto proposto.
- Implementar uma ferramenta que permita a execução do processo de predição de peso, tornando-o eficiente e acessível.

1.2 Contribuições

A ferramenta desenvolvida para predição de peso apresenta as seguintes contribuições:

- Melhoria na Gestão Pecuária com a redução de custos e perdas.
- Aumento da precisão na pecuária.
- Facilidade de uso e acesso além da modernização do setor pecuário.

1.3 Estrutura do trabalho

Além deste capítulo introdutório, o restante da monografia está organizada da seguinte forma: o Capítulo 2 apresenta os fundamentos teóricos da metodologia abordada. No Capítulo 3 são apresentados os trabalhos relacionados. No Capítulo 4 é apresentada a metodologia proposta. No Capítulo 5 são discutidos os resultados obtidos com a metodologia proposta. Por fim, no Capítulo 6 é apresentada a conclusão.

2 Referencial Teórico

Essa seção irá abranger alguns conceitos necessários para o entendimento do deste projeto. Os conceitos que serão apresentados são: Aprendizado de Máquina e Algoritmo de Regressão.

2.1 Aprendizado de Máquina

Segundo Müller e Guido (2016) o aprendizado de máquina é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos capazes de aprender a partir de dados de entrada. O objetivo do aprendizado de máquina é construir modelos que possam fazer previsões ou tomar decisões com base em padrões identificados em conjuntos de dados. Existem três principais tipos de aprendizado de máquina: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

Em Mahesh (2020) explica que no aprendizado supervisionado, o modelo é treinado usando dados rotulados, ou seja, dados que já possuem a resposta correta. O modelo tenta aprender a relação entre as características de entrada e a saída desejada (rótulo), para que possa fazer previsões precisas em dados nunca antes vistos. No aprendizado não supervisionado, o modelo é treinado em dados não rotulados, e o objetivo é encontrar padrões ou estrutura oculta nos dados. Os dados são agrupados em clusters com base em sua similaridade. No aprendizado por reforço, o modelo aprende a tomar decisões em um ambiente dinâmico, recebendo feedback em forma de recompensa ou penalidade, com o objetivo de maximizar a recompensa ao longo do tempo.

O aprendizado de máquina é amplamente utilizado em aplicações como reconhecimento de fala, processamento de linguagem natural, visão computacional, detecção de fraudes, sistemas de recomendação, entre outros.



Figura 1 – Demonstração de tipos de Aprendizado de Máquina. Fonte: (MAIA, 2020)

2.2 Algoritmo de Regressão

Regressão é um tipo de algoritmo de aprendizado de máquina utilizado para prever um valor numérico baseado em dados de entrada. Existem vários tipos de algoritmos de regressão, incluindo regressão linear, regressão logística e regressão de árvores de decisão.

A Regressão Linear é um algoritmo de aprendizado de máquina que tenta prever um valor numérico a partir de dados de entrada. Ele assume que existe uma relação linear entre as variáveis de entrada (ou características) e a variável de saída (ou alvo). A regressão linear tenta encontrar os melhores coeficientes (ou pesos) para cada característica, de modo que a soma ponderada dos valores das características seja o mais próximo possível do valor alvo (MATOS, 1995).

A regressão linear é uma técnica simples, mas pode ser muito eficaz em muitas aplicações. No entanto, ela só é válida quando existe uma relação linear entre as variáveis de entrada e de saída.

Em (MOREIRA et al., 2020) explica o conceito e a utilidade da análise de regressão como uma ferramenta estatística para prever um resultado (variável dependente) com base em uma ou mais variáveis explicativas (variáveis independentes). Ele destaca que a análise de regressão permite avaliar o impacto de cada variável independente sobre o resultado desejado, indo além da simples correlação entre variáveis.

Além disso, é ressaltado que a análise de regressão é particularmente útil para avaliar a influência de um conjunto de fatores sobre doenças ou outras características de interesse. Ela permite estabelecer uma equação que descreve os relacionamentos entre a variável dependente e as variáveis independentes que se deseja investigar.

A Regressão Logística é um algoritmo de aprendizado de máquina utilizado para prever a probabilidade de um evento binário (verdadeiro ou falso) com base em dados de entrada. É um tipo específico de análise discriminante que é usado para classificação binária (FIGUEIRA, 2006). Ao contrário da regressão linear, que tenta prever um valor numérico, a regressão logística tenta prever a probabilidade de ocorrência de um evento, geralmente representado por uma função sigmóide. É comumente utilizada em problemas de classificação binária, como análise de risco, diagnóstico médico e detecção de fraude. Ele é capaz de lidar com problemas linearmente não separáveis e é robusto a *outliers*¹.

Em (GONZALEZ, 2018) destaca a importância das técnicas de mineração de dados, especialmente devido ao grande volume de dados gerados e armazenados atualmente. Ele se concentra na regressão logística como uma técnica específica para a análise de dados com resposta categórica, tanto em sua forma binária quanto múltipla. A regressão logística é apresentada como uma ferramenta poderosa para identificar padrões relevantes em grandes conjuntos de dados, permitindo a estimação da probabilidade de ocorrência de eventos e a avaliação dos fatores que contribuem para esses eventos.

¹ É uma observação que está fora do padrão geral de uma distribuição. É um ponto de dados que difere significativamente de outras observações em um conjunto de dados.

A Regressão de Árvores é um algoritmo de aprendizado de máquina utilizado para prever valores numéricos com base em dados de entrada. Ele é baseado em árvores de decisão, onde cada nó representa uma decisão a ser tomada com base em uma característica de entrada, e cada ramo representa uma possível consequência dessa decisão. A Regressão de árvore é uma técnica não paramétrica e não linear, é capaz de lidar com problemas de alta dimensionalidade, e é fácil de ser interpretado. No entanto, pode ser propenso a *overfitting*², é necessário cuidado ao lidar com o hiperparâmetro de profundidade da árvore (FERREIRA, 1999).

No trabalho do (CAPELO, 2022) é analisado o processo de construção de uma árvore de regressão difere um pouco do processo de construção de uma árvore de decisão para classificação. Em uma árvore de regressão, cada folha contém um valor numérico em vez de categorias como "Sim" ou "Não".

Para construir a árvore de regressão, primeiro é necessário determinar qual valor será utilizado como condição no primeiro nó. O algoritmo testa todos os possíveis valores de separação para os dados e seleciona aquele que minimiza a soma dos quadrados dos resíduos. Os resíduos são as diferenças entre o valor original e o valor previsto pela árvore. Ao elevar esses resíduos ao quadrado e somá-los, obtemos a soma dos quadrados dos resíduos.

A Rede Neural Artificial (RNA) representa uma abordagem computacional inspirada no funcionamento do cérebro humano, onde neurônios artificiais são organizados em camadas para realizar tarefas específicas de processamento de informações. Seu referencial teórico abrange uma variedade de conceitos fundamentais que fundamentam seu design, treinamento e aplicação em uma ampla gama de problemas de aprendizado de máquina e inteligência artificial (RAUBER, 2005).

Em seu núcleo, os neurônios artificiais, também chamados de unidades de processamento, formam os blocos de construção essenciais de uma RNA. Cada neurônio recebe entradas, que são ponderadas por pesos sinápticos, e aplica uma função de ativação para gerar uma saída. A organização desses neurônios em camadas, como entrada, oculta e saída, compõe a arquitetura da rede.

Durante o treinamento da RNA, os pesos sinápticos são ajustados iterativamente para minimizar uma função de perda, através de algoritmos de otimização como Gradiente Descendente. Esse processo de aprendizado pode ser supervisionado, onde a rede é treinada em dados rotulados, ou não supervisionado, onde ela descobre padrões nos dados por conta própria.

Diversas funções de ativação são empregadas nos neurônios para introduzir não linearidades nas saídas, permitindo que a rede modele relações complexas nos dados. Além disso,

² É um fenômeno comum em aprendizado de máquina e modelagem estatística, no qual um modelo se ajusta muito bem aos dados de treinamento, mas tem um desempenho inferior ao lidar com novos dados não vistos anteriormente.

técnicas de regularização são aplicadas para prevenir *overfitting* e melhorar a capacidade de generalização da RNA.

As aplicações das redes neurais artificiais são vastas e incluem reconhecimento de padrões, visão computacional, processamento de linguagem natural, previsão de séries temporais, diagnóstico médico, entre muitas outras. Sua capacidade de aprender representações complexas dos dados torna as RNAs uma ferramenta poderosa em uma variedade de domínios.

Em (VENDRUSCOLO et al., 2017) teve como objetivo avaliar a modelagem por regressão e por redes neurais artificiais na estimativa da altura total de árvores de teca em diferentes espaçamentos em Cáceres, MT. Os dados foram coletados a partir da medição do diâmetro à altura do peito (dap) por meio de censo florestal, os quais foram posteriormente agrupados em classes de diâmetro, com amplitude de 5 cm. A altura total (h) de 20% dos indivíduos em cada espaçamento e classe de diâmetro foi medida.

Para estimar a altura total por meio de regressão, foram utilizados modelos tanto não lineares quanto lineares. Destacou-se, entre os modelos não lineares, o modelo de Gompertz. Além disso, foram testadas redes neurais artificiais do tipo Multilayer Perceptron para essa estimativa.

Os resultados indicaram que os modelos não lineares foram superiores aos lineares na modelagem por regressão. Ambas as técnicas, regressão e redes neurais artificiais, foram eficientes na estimativa da altura total de *Tectona grandis*. No entanto, a modelagem por redes neurais artificiais apresentou erro inferior a 10

2.3 Métricas de regressão

2.3.1 Erro Médio Absoluto (MAE)

É uma medida direta da magnitude dos erros entre as previsões do modelo e os valores reais. Calcula-se a média das diferenças absolutas, oferecendo uma visão geral da precisão média do modelo.

2.3.2 Erro Quadrático Médio (MSE)

O Erro Quadrático Médio (MSE) leva em conta não apenas a magnitude, mas também a direção dos erros. Ele calcula a média dos quadrados das diferenças entre as previsões e os valores reais, penalizando erros maiores de forma mais significativa.

2.3.3 Raiz do Erro Quadrático Médio (RMSE)

É uma versão do MSE que retorna a mesma unidade dos dados originais, facilitando a interpretação. Ao extrair a raiz quadrada do MSE, essa métrica oferece uma medida do

erro médio em termos das unidades da variável alvo.

2.3.4 Erro Percentual Absoluto Médio (MAPE)

É particularmente útil quando se deseja entender o erro relativo do modelo. Ele calcula a média das porcentagens absolutas das diferenças entre as previsões e os valores reais, oferecendo uma visão da precisão relativa das previsões.

2.3.5 Coeficiente de Determinação (R^2)

O R^2 fornece uma medida da qualidade geral do ajuste do modelo aos dados. Ele indica a proporção da variância na variável dependente que é explicada pelas variáveis independentes, sendo uma medida de quão bem o modelo se ajusta aos dados.

Essas métricas fornecem uma visão abrangente do desempenho do modelo de regressão, cada uma destacando diferentes aspectos da qualidade das previsões. Ao escolher as métricas mais adequadas para um determinado problema, os praticantes podem realizar uma avaliação precisa e informada do desempenho do modelo, auxiliando na tomada de decisões e no refinamento do processo de modelagem.

3 Trabalhos Relacionados

Esta seção analisamos os estudos e descobertas relacionadas que fornecem uma base sólida para o contexto deste estudo. Estudos anteriores abordaram questões semelhantes nesta área, fornecendo informações valiosas e estabelecendo as bases para a nossa pesquisa. O objetivo da revisão da literatura é situar a nossa investigação no contexto existente, destacar lacunas no conhecimento e destacar as contribuições únicas da nossa abordagem ao campo.

No trabalho de [Koritiaki et al. \(2012\)](#) é apresentado um estudo que teve como objetivo testar correlações entre peso e medidas corporais de cordeiros Santa Inês e desenvolver regressões lineares simples e múltiplas para prever o peso por meio de medidas corporais. O estudo foi conduzido com 150 cordeiros do setor de ovinocultura da Universidade Estadual de Londrina (UEL), Brasil. Os animais foram pesados e medidos ao nascimento e ao desmame. As medidas morfométricas corporais incluíram altura de cernelha, comprimento corporal e perímetro torácico. Os resultados mostraram que todas as medidas corporais apresentaram correlação positiva com o peso, sendo a circunferência do tórax a medida com maior correlação. As regressões lineares simples foram eficazes para estimar o peso ao nascer e as regressões lineares múltiplas foram mais eficientes para estimar o peso ao desmame. O estudo conclui que as medidas corporais podem ser consideradas uma boa alternativa para prever o peso de cordeiros, facilitando o manejo animal e o controle zootécnico.

Já em [Santos et al. \(2014\)](#), foi investigado a relação entre diferentes medidas morfométricas (Perímetro torácico (PT), Perímetro de barril (PB), Comprimento Corporal (CC), Altura de Cernelha (AC), Altura de Garupa (AG), Largura de garupa (LG), Comprimento de garupa (CG), estimada a Compacidade de Carcaça ($CMP = PV/CC$)) e o peso corporal em ovelhas F1 criadas a pasto. O objetivo foi determinar qual medida é a melhor para prever o peso corporal nesse tipo de criação, com potencial para auxiliar na avaliação do desempenho dos animais e na tomada de decisões relacionadas à gestão da produção. Os resultados do estudo indicam que o perímetro torácico (PT) apresentou o melhor desempenho na predição do peso corporal (PC) de ovinos F1 (Dorper x Santa Inês) criados no semiárido mineiro. O PT demonstrou um alto coeficiente de determinação ($R^2 = 0,77$), baixo coeficiente de variação ($CV = 6,89\%$), e foi altamente significativo ($P < 0,0001$). Em comparação, outras medidas como a altura de garupa (AG) e o comprimento corporal (CC) mostraram valores de R^2 similares, mas com CVs um pouco mais elevados. A compacidade de carcaça (CMP), embora também tenha apresentado um alto R^2 , não foi considerada viável, pois depende do peso vivo do animal.

Portanto, conclui-se que o PT é a medida mais prática e confiável para estimar o peso corporal desses ovinos.

O estudo de [Conrado, Arandas e Ribeiro \(2015\)](#) teve como objetivo utilizar equações de predição para estimar o peso vivo de caprinos Canindé com base em medidas morfométricas quantitativas. Foram avaliadas 347 cabras Canindé adultas de diferentes estados da região Nordeste do Brasil. Os animais foram pesados e medidas de altura da cernelha (AC), comprimento corporal (CC) e perímetro do torácico (PT) foram feitas com uma fita métrica. Correlações significativas foram observadas entre o peso e todas as variáveis avaliadas, com PT apresentando a correlação mais forte (0,69). Os modelos polinomiais forneceram o melhor ajuste, sendo o modelo de três variáveis o mais preciso. Esta metodologia mostrou-se adequada para prever o peso de cabras Canindé adultas nas condições em que o estudo foi conduzido.

Em [TRINDADE \(2021\)](#) procurou prever o peso vivo, características da carcaça e pesos de cortes comerciais de cordeiros usando medidas morfométricas realizadas em animais in vivo. As informações foram coletadas de 92 cordeiros Santa Inês e mestiços Dorper x Santa Inês abatidos entre 7 e 12 meses de idade. Os dados resultantes foram analisados por meio do algoritmo Elastic Net, o que resultou em 4 modelos de previsão - 2 para a vista dorsal e 2 para a vista lateral - usando o peso vivo como variável preditora. A análise mostrou que a utilização de medidas morfométricas utilizando o peso vivo como variável preditora é eficaz para prever características do animal vivo, características da carcaça e dos cortes comerciais, tornando-se uma ferramenta útil para o meio rural.

Em [Salazar-Cuytun et al. \(2022\)](#) para avaliar a relação entre o PV e o BV em cordeiros em crescimento, foram ajustados três modelos: linear, quadrático e alométrico. O modelo quadrático apresentou o maior coeficiente de determinação ($R^2=0,93$) e o menor erro de previsão, mostrando melhor desempenho na predição do PV. Isso está de acordo com outros estudos, que também observaram que o modelo quadrático é o mais adequado para relacionar essas variáveis em cordeiros.

A compreensão da relação entre a condição corporal e a fertilidade é de extrema importância para alcançar uma produção e reprodução eficientes dos animais. Isso fornece técnicas simples que podem guiar decisões nutricionais e reprodutivas no manejo do rebanho.

O estudo ([MACIEL et al., 2022](#)) teve como objetivo utilizar técnicas de análise multivariada para avaliar a composição das carcaças de ovinos da raça Santa Inês, bem como distinguir grupos com base nos escores de conformação e acabamento de carcaça. Foram utilizadas informações de 122 ovinos da raça Santa Inês, com 24 características de carcaça. As técnicas aplicadas incluíram análise fatorial para redução da dimensionalidade dos dados, regressão múltipla stepwise para prever a composição dos tecidos (músculo, osso e gordura), e análise discriminante multivariada para distinguir grupos com base nos escores de conformação e acabamento de carcaça.

O teste de esfericidade de Bartlett e o KMO indicaram que as variáveis de carcaça eram adequadas para a análise fatorial. A rotação ortogonal Varimax foi escolhida para simpli-

ficar a interpretação dos fatores. A utilização das variáveis latentes da análise fatorial em modelos de regressão múltipla mostrou-se útil para melhorar a acurácia da interpretação dos resultados, eliminando o problema de multicolinearidade. No entanto, os coeficientes de determinação (R^2) foram moderados para proporção muscular e gordura total e baixos para proporção óssea, indicando a necessidade de variáveis independentes mais adequadas para melhor prever a proporção dos tecidos. Na análise discriminante multivariada, as variáveis com maior poder discriminatório para os escores de conformação de carcaça foram peso de carcaça fria, comprimento externo da carcaça e peso do pescoço, enquanto para acabamento de carcaça foram peso vivo ao abate, comprimento externo da carcaça e perímetro do tórax. A análise discriminante mostrou-se eficiente para alocar os animais em seus grupos de origem (escores de carcaça).

As técnicas de análise multivariada foram úteis para compreender a composição das carcaças de ovinos da raça Santa Inês e distinguir grupos com base nos escores de conformação e acabamento de carcaça. No entanto, foram identificadas áreas de melhoria, como a necessidade de variáveis independentes mais adequadas para prever a composição dos tecidos e aprimorar a acurácia dos modelos.

O estudo do (SILVA, 2023a) teve como objetivo avaliar o desempenho e correlacionar medidas morfométricas in vivo com características da carcaça de cordeiros Texel em confinamento, submetidos a dietas de alto e baixo ganho de peso.

Os cordeiros Texel foram divididos em dois tratamentos: alto ganho (T1) e baixo ganho (T2), com diferentes dietas e períodos de confinamento. Foram realizadas avaliações de desempenho e medidas morfométricas in vivo e na carcaça. Os cordeiros do grupo de alto ganho apresentaram melhores resultados em ganho de peso, escore de condição corporal, perímetro torácico, altura de anterior e altura de posterior, em comparação com os do grupo de baixo ganho. As medidas físicas da carcaça, como peso de carcaça fria, comprimento interno de carcaça e perímetro de garupa, foram maiores no grupo de alto ganho.

Variáveis morfométricas in vivo, como escore de condição corporal, comprimento corporal e perímetro torácico, mostraram alta correlação com características da carcaça dos cordeiros Texel, independentemente da taxa de crescimento. Medidas morfométricas in vivo, como perímetro torácico, comprimento corporal e escore de condição corporal, podem ser úteis como indicadores correlacionados com características da carcaça de cordeiros Texel, independentemente da taxa de crescimento.

Em (PAULA et al., 2023) o estudo teve como objetivo correlacionar características morfométricas e de carcaça in vivo em ovelhas Dorper puras de origem em diferentes faixas etárias. Foram avaliados 39 ovelhas, estimando-se o escore de condição corporal (CE) e medidas corporais, além de analisar as carcaças por ultrassonografia. Os dados foram divididos em dois grupos com base na idade (Grupo G1: $1,4 \leq 2$ anos e Grupo G2: $2 < \text{anos} \leq 5,5$) e submetidos à análise estatística. O grupo G2 apresentou valores maiores

em várias características, incluindo escore de condição corporal, altura na cernelha, altura na garupa, perímetro torácico, comprimento do corpo, área de olho de lombo, largura e profundidade do músculo L. dorsi, e espessura de gordura subcutânea, em comparação com o grupo G1. Algumas características mostraram diferenças significativas entre os grupos, enquanto outras não. Foram encontradas correlações moderadas entre idade e escore de condição corporal, e correlações significativas entre várias características morfométricas e de carcaça, como altura na cernelha, altura na garupa, perímetro torácico, área de olho de lombo, largura e profundidade do músculo L. dorsi, e espessura de gordura subcutânea. O perímetro torácico mostrou correlação significativa com diversas medidas de carcaça.

Os resultados sugerem que a medida da razão entre largura e profundidade do músculo L. dorsi (MUSC) pode ser utilizada para selecionar ovinos com maior musculatura, enquanto o perímetro torácico (PT) pode ser utilizado como indicador de boa conformação de carcaça, permitindo avaliação *in vivo* sem a necessidade de ultrassom.

Em (SILVA, 2023b) teve como objetivo caracterizar o desenvolvimento ponderal de cordeiros e cordeiras da raça Santa Inês, do nascimento ao desmame, manejados em pastagem e recebendo suplementação concentrada via creep-feeding.

Foram utilizados dados de 212 ovinos da raça Santa Inês. Os animais foram mantidos em pastagem de capim massai, com suplementação mineral fornecida em cochos. Os cordeiros receberam suplementação concentrada desde o primeiro dia de vida. O ganho de peso total (GPT) e o ganho médio diário (GMD) foram avaliados a cada 7 dias, enquanto medidas biométricas foram tomadas a cada 28 dias. Foram testados modelos de regressão linear, simples e múltipla, para estimar o peso vivo utilizando medidas biométricas como variáveis independentes. O Modelo não linear de Gompertz foi utilizado para descrever o desenvolvimento ponderal ao longo do tempo.

Houve interação entre tipo de parto e sexo para o peso ao nascer (PN), e efeito do tipo de parto no GMD, GPT e peso ao desmame. As curvas de crescimento de acordo com o modelo de Gompertz mostraram alta acurácia. As correlações entre medidas biométricas e peso foram positivas e significativas, permitindo a criação de equações simples e múltiplas com alta acurácia para prever o peso ao desmame.

Cordeiros de parto duplo, recebendo suplementação concentrada via creep-feeding e manejados a pasto, desenvolvem-se mais lentamente que animais de parto simples, mas apresentam ganhos compensatórios ao desmame. O modelo de Gompertz pode ser útil para prever o desenvolvimento dos animais e auxiliar na tomada de decisões de manejo. As equações lineares, simples e múltiplas, são eficientes para prever o peso ao desmame por meio de medidas biométricas, com destaque para a equação utilizando comprimento corporal e perímetro de barril.

A Tabela 1 apresenta uma síntese dos métodos e medidas utilizados nos estudos relacionados pelos autores e fornece informações valiosas sobre as abordagens utilizadas para analisar características específicas.

Autor	Metodos	Medidas
TRINDADE (2021)	Algoritmo Elastic Net	CT,CC, P, LT, LL, LG, D_Cer-5VT,D_5VT-13VT, D_13VT-1S,D_1S-IC,D_EE. CP, LP, CCox, CAb, LAb, CEsc, D_Esc-IL, D_Esc-IS, D_Esc-FT, AC, AT, AL, AG. VC, PT, CC.
Salazar-Cuytun et al. (2022)	Regressão linear, quadrático e alométrico	AO, CC, CT.
Koritiaki et al. (2012)	Regressões lineares simples e múltiplas	PT, PB, CC, AC, AG,LG,CG, P, CMP.
Santos et al. (2014)	Regressão linear	AC, CC, PT.
Conrado, Arandas e Ribeiro (2015)	Regressão lineares, quadrática e polinomial	Peso final, PCF,RCF, pH, Temperatura, CG, Grau de conformação, CP, CIC, CEC, LG, PP, PG, CP, ECCPi, CCFi, PTFi.
(SILVA, 2023a)	A análise multivariada	PCV, PVA, PCQ, PCFRB, RCQ, RCF, CEC, CIC, CP, LG, LT, PG, PP, PT, ICC, PER, LOM, COS, PES, PAL, RPER, RLOM, RCOS, RPES, RPAL.
(MACIEL et al., 2022)	Análise Fatorial e Regressão Múltipla Stepwise	EC, AC, PT, CC, AOL, LARG, PROF, EG.
(PAULA et al., 2023)	Análise estatística	CC, AA, AP, LP, LG, PT.
(SILVA, 2023b)	Regressão linear simples e múltipla	

Tabela 1 – Trabalhos relacionados

Esta análise resumida concentra-se nos diferentes métodos estatísticos utilizados pelos autores para examinar as relações entre variáveis morfológicas específicas em seus respectivos estudos. Cada abordagem metodológica reflete uma decisão estratégica para compreender e modelar as características do objeto de estudo.

Os trabalhos relacionados forneceram informações valiosas sobre a relação entre medidas morfométricas e peso corporal em pequenos ruminantes, mas revelaram a falta significativa de praticidade e acessibilidade na predição de peso em pequenos ruminantes, especialmente no contexto de produtores rurais. Este trabalho aborda essa falta ao desenvolver uma aplicação móvel que traduz os conhecimentos adquiridos em um instrumento prático e eficiente para os produtores.

4 Metodologia Proposta

Neste capítulo descreve os aspectos fundamentais relacionados à manipulação e análise do conjunto de dados estudado e à implementação e avaliação do modelo proposto.

4.1 Conjunto de Dados

A base de dados privada utilizada consiste em 1186 amostras, cada amostra apresenta diversas características, oferecendo uma visão ampla dos animais examinados.

Os dados incluem informações como o número do animal, identificação dos reprodutores (pai e mãe), padrão racial, sexo do animal, data de nascimento e a idade do animal (em dias). Além disso, são fornecidas informações sobre o proprietário do animal, a fazenda e o município local.

A coleta de dados também inclui a data da mensuração, permitindo a análise temporal das características físicas dos animais. O estado fisiológico do animal é exibido e fornece informações sobre seu estado de saúde e estágio reprodutivo.

Dentre as características físicas medidas, destacam-se as medidas obtidas pela ultrasonografia como a área de olho de lombo, comprimento de olho de lombo, profundidade de olho de lombo e a relação entre comprimento e profundidade de olho de lombo. Além disso, são coletadas informações sobre espessura da gordura subcutânea, espessura da gordura esternal e marmoreio do olho de lombo.

O conjunto de dados também inclui parâmetros como escore corporal, peso corporal, altura da cernelha, altura da garupa, comprimento corporal, altura do peito, perímetro da canela, circunferência torácica, comprimento da orelha, comprimento da cabeça, largura da cabeça, largura do íleo, largura do ísquio, comprimento da garupa, comprimento da perna, perímetro da perna, comprimento da cauda, diâmetro da cauda e circunferência escrotal.

Esses dados fornecem uma base rica para análise e permitem a compreensão das características físicas e genéticas dos animais. Estes são valiosos para pesquisas relacionadas à pecuária e ao melhoramento genético. A diversidade e o número de variáveis presentes no conjunto de dados permitem um estudo detalhado e abrangente das características dos animais estudados.

4.2 Pré-Processamento

O pré-processamento dos dados desempenha um papel importante na preparação adequada para análises subsequentes. Neste contexto, as informações qualitativas e que não

podiam ser obtidas por meio de fita métrica foram criteriosamente eliminadas, visando otimizar recursos e simplificar a análise dentro do escopo do estudo. Essa abordagem permite concentrar os esforços nos elementos de maior importância, contribuindo para uma análise mais eficiente e focada.

Adicionalmente, foi realizado o descarte das amostras que apresentavam dados ausentes em algum dos atributos para a investigação em curso. Esta ação foi tomada com o intuito de assegurar a integridade e consistência dos dados, prevenindo potenciais distorções ou lacunas que poderiam afetar a qualidade das análises posteriores.

Dessa forma, a estratégia de pré-processamento adotada não apenas simplifica o conjunto de dados, tornando-o mais gerenciável, mas também garante que apenas informações relevantes e completas sejam consideradas nas etapas subsequentes da pesquisa. Essas práticas não só beneficiam a confiabilidade e validade dos resultados obtidos durante o estudo, como também fortalecem sua integridade metodológica.

4.3 Seleção de Características

Um passo importante na metodologia adotada para o desenvolvimento da ferramenta foi a seleção de características. Foi dada prioridade à inclusão de características que pudessem ser medidas diretamente durante a de coleta de dados.

A escolha por características mensuráveis foi motivada pela busca de simplicidade e praticidade na implementação da ferramenta, garantindo, ao mesmo tempo, uma relação direta com as variáveis de interesse, ou seja, os dados de peso dos animais. Dessa forma, características como perímetro da perna, peso corporal, altura da cernelha, altura da garupa, comprimento corporal, altura do peito, perímetro da canela, circunferência torácica, comprimento da garupa e comprimento da perna, que podem ser medidas com facilidade e precisão, foram incorporadas à análise.

Esta abordagem não apenas simplifica o processo de coleta de dados, mas também facilita a interpretação dos resultados e fornece uma compreensão intuitiva da relação entre características mensuráveis e peso corporal do animal. A seleção proposital de características diretamente mensuráveis visa otimizar a eficácia das ferramentas preditivas e fornecer uma abordagem prática e eficiente para a estimativa do peso corporal em pequenos ruminantes.

4.4 Modelos implementados

Neste projeto, foram implementados quatro modelos de regressão utilizando a linguagem de programação Python, a Regressão Linear, Ridge Regression, Random Forest foram implementadas da biblioteca scikit-learn e a MLP foi implementada utilizando a biblioteca TensorFlow/Keras . A importância de escolher esses modelos reside no fato

de que cada um deles possui características específicas que podem se adequar melhor à tarefa de predição de peso corporal em pequenos ruminantes. Este projeto busca explorar as nuances de cada modelo e adaptá-los de forma a maximizar sua eficácia na predição de peso corporal. Essa abordagem permite uma avaliação comparativa dos modelos em termos de desempenho e precisão, auxiliando na seleção do mais adequado para a análise dos dados de pequenos ruminantes.

A **Regressão Linear**¹ é um modelo simples e básico. Foi utilizado para determinar uma linha de melhor ajuste entre as variáveis de entrada (medidas) e a variável de saída (peso corporal). O modelo assume relações lineares entre as variáveis.

A **Ridge Regression**² é uma extensão da Regressão Linear que inclui um termo de regularização para lidar com a multicolinearidade. Essa técnica ajuda a evitar overfitting, principalmente quando há uma alta correlação alta entre as variáveis independentes.

O **Random Forest Regressor**³ é um algoritmo de aprendizado de máquina usado para prever valores contínuos em tarefas de regressão. Com base na árvore de decisão, uma “floresta” de árvores é construída usando amostras aleatórias de dados e características. Durante a previsão, a árvore gera previsões individuais e a média dessas previsões é considerada a estimativa final. Essa abordagem de conjunto fornece robustez, capacidade de lidar com dados complexos e risco reduzido de overfitting.

O **MLP**⁴ é um tipo de rede neural artificial com múltiplas camadas. Foi escolhido devido à sua capacidade de capturar relações não lineares complexas nos dados. Foi criada e treinada uma rede neural, ajustando a arquitetura da MLP conforme necessário para a tarefa específica.

Cada modelo foi selecionado com base em suas características únicas e adaptado para a previsão do peso corporal em pequenos ruminantes. A implementação foi realizada cuidadosamente, utilizando bibliotecas poderosas em Python, e os modelos foram treinados e avaliados para garantir a eficácia na tarefa proposta. Essa diversidade de abordagens permite uma compreensão mais profunda das relações presentes nos dados e contribui para a robustez das previsões.

4.5 Treinamento e Métricas

Os dados foram pré-processados para lidar com valores ausentes, e em seguida, divididos em conjuntos de treinamento e teste. Uma proporção de 70% foi utilizada para treinamento e 30% para testes. O objetivo desta escolha foi equilibrar a quantidade de

¹ https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

² https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html

³ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

⁴ https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html

dados disponíveis para treinar o modelo e, ao mesmo tempo, avaliar de forma confiável a capacidade de generalização do modelo.

Além disso, para cada conjunto de dados de treinamento e teste, foram utilizadas todas as combinações de 3, 4, 5 e 6 medidas, incluindo ac, ag, pc, ct, cc, cperna, cga, ap e perperna. Essas combinações foram geradas utilizando a função `list(combinations(características, quantidade))`, permitindo explorar diferentes conjuntos de características durante o treinamento e teste do modelo.

Os modelos de regressão foram treinados com os conjuntos de treinamento. Cada modelo foi ajustado aos dados usando parâmetros específicos obtidos através de uma busca em grade. Durante o treinamento, os dados foram normalizados utilizando `StandardScaler`⁵ para garantir estabilidade e convergência do algoritmo.

Durante a busca em grade⁶, os modelos foram ajustados com o objetivo de otimizar sua capacidade de generalização, utilizando métricas de desempenho. A normalização dos dados e a seleção de hiperparâmetros adequados foram aspectos essenciais para aprimorar a qualidade das previsões.

Os hiperparâmetros específicos utilizados para o ajuste de cada modelo foram os seguintes:

- Regressão Linear: Não foram especificados hiperparâmetros adicionais.
- Ridge: Utilizou-se o hiperparâmetro alpha, variando entre valores em uma escala logarítmica de -6 a 6 com 13 pontos.
- MLP Regressor: Os hiperparâmetros foram definidos como `hidden_layer_sizes`, com opções de arquitetura de rede neural de uma camada oculta com 50 neurônios, uma camada oculta com 100 neurônios, duas camadas ocultas com 50 neurônios cada ou uma camada oculta com 100 neurônios seguida por uma camada com 50 neurônios, e alpha, variando entre 0.0001, 0.001 e 0.01.
- Random Forest Regressor: Os hiperparâmetros foram definidos como `n_estimators`, com opções de 50, 100 ou 200 árvores na floresta, e `max_depth`, com opções de profundidade máxima de cada árvore sendo None, 10 ou 20.

A validação cruzada foi aplicada à pesquisa em grade usando k-fold com 5 compartimentos. Esta técnica contribuiu para uma avaliação mais confiável do desempenho do modelo considerando diferentes partições dos dados de treinamento.

Para avaliar o desempenho dos modelos, foram empregadas métricas comuns para problemas de regressão. Utilizando uma única métrica, o R^2 , possibilita uma comparação

⁵ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

⁶ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

direta e descomplicada entre os modelos. Isso simplifica a identificação do modelo que melhor se adapta aos dados, sem a necessidade de levar em conta várias métricas com interpretações distintas.

Os valores de R^2 representam a proporção da variabilidade nos dados que os modelos conseguiram explicar. Um R^2 próximo de 1 indica uma boa capacidade de predição, enquanto valores mais baixos sugerem que o modelo pode não estar capturando efetivamente os padrões nos dados.

Os resultados obtidos foram analisados considerando as expectativas iniciais do estudo. Destacaram-se pontos fortes e limitações de cada modelo, proporcionando *insights* valiosos sobre sua adequação ao problema específico de predição de peso de pequenos ruminantes.

4.6 Desenvolvimento da Ferramenta

A criação da ferramenta foi conduzida utilizando a linguagem de programação Python, aproveitando os recursos das bibliotecas Kivy e KivyMD para elaborar uma interface gráfica interativa. O foco foi direcionado para garantir uma excelente usabilidade, proporcionando uma experiência fluida e intuitiva para os usuários.

4.6.1 Escolha da Linguagem de Programação

A linguagem Python foi escolhida com base em sua versatilidade, ampla variedade de bibliotecas e comunidade ativa de desenvolvedores. Essa decisão permitiu implementar com eficiência lógica preditiva, manipulação de dados e integração da interface do usuário.

4.6.2 Kivy e KivyMD para Interfaces Gráficas

Utilizado como framework principal para o desenvolvimento de interfaces gráficas multi-touch. A simplicidade e flexibilidade do Kivy foram fatores chave que nos permitiram criar uma interface intuitiva e responsiva.

Ao introduzir elementos e padrões de design de Material Design, KivyMD garantiu uma estética moderna e consistente para interfaces de ferramentas. Esta biblioteca estende o Kivy com componentes pré-construídos que facilitam a criação de interfaces visualmente atraentes.

4.6.3 Prototipação Visual

Antes da implementação completa, a primeira fase de desenvolvimento incluía a prototipagem visual da interface gráfica do usuário. Ao trabalhar com Kivy e KivyMD, conseguimos criar projetos preliminares que refletiam com precisão o posicionamento dos elementos da interface, acelerando o processo de desenvolvimento.

5 Resultados

Este capítulo apresenta os resultados obtidos para os quatro modelos descritos no Capítulo 4 e o desenvolvimento da ferramenta.

5.1 Resultado dos modelos

Os experimentos foram realizados com os modelos e com as combinações geradas, a Tabela 2 mostra os melhores resultados obtidos pelos modelos de regressão ao serem aplicados em quatro conjuntos de características variados. A métrica utilizada para avaliar o desempenho dos modelos é o coeficiente de determinação (R^2), que indica a proporção da variabilidade na variável dependente que é explicada pelo modelo.

Os modelos de Regressão Linear e Ridge apresentam um desempenho consistente à medida que o número de características aumenta. Os valores de R^2 para ambos os modelos variam de 0.8407 com três características até 0.8500 com seis características como é demonstrado na Tabela 2, Tabela 3 e Tabela 4. Esses resultados sugerem que a capacidade preditiva do modelo não muda substancialmente quando características adicionais são incluídas.

O modelo MLP, que representa uma Rede Neural de Perceptron Multicamadas, apresenta uma performance ligeiramente inferior em comparação com os modelos lineares. À medida que a complexidade do conjunto de recursos aumenta, o valor R^2 varia entre 0,8248 e 0,8181, como é demonstrado na Tabela 2. Isto pode indicar que o MLP tem certa sensibilidade à inclusão de recursos adicionais.

O modelo Random Forest, que é um *ensemble* de árvores de decisão, mostra uma tendência interessante nos resultados. O R^2 inicial para três características é 0,8152, mas um aumento significativo para 0,8337 é observado para seis características.

A Tabela 3 mostra as características que apresentou o melhor desempenho em cada modelo de regressão considerando diferentes conjuntos de características. Esses resultados enfatizam a importância de escolher o modelo adequado para um determinado conjunto de características. A Regressão Linear e Ridge mantêm uma estabilidade consistente, enquanto o MLP e o Random Forest são mais sensíveis à complexidade do conjunto de características.

Modelo de Regressão	Três Características	Quatro Características	Cinco Características	Seis Características
Regressão Linear	0,8407	0,8443	0,8480	0,8500
Ridge	0,8407	0,8443	0,8480	0,8500
MLP	0,8248	0,8256	0,8193	0,8181
Random Forest	0,8152	0,8256	0,8314	0,8337

Tabela 2 – Resultados obtidos com os modelos de regressão em diferentes conjuntos de características utilizando a métrica R^2 .

Modelo de Regressão	Três Características	Quatro Características	Cinco Características	Seis Características
Regressão Linear	pc, ct, cperna	ag, pc, ct, cperna	ag, ct, cc, cga, perperna	ag, ct, cc, cga, ap, perperna
Ridge	pc, ct, cperna	ag, pc, ct, cperna	ag, ct, cc, cga, perperna	ag, ct, cc, cga, ap, perperna
MLP	ct, cc, perperna	ag, ct, cc, perperna	ac, ag, ct, cc, ap	ac, ag, ct, cc, ap, perperna
Random Forest	ct, cc, perperna	ag, ct, cc, perperna	ag, pc, ct, cperna, cga	ag, pc, ct, cc, cperna, cga

Tabela 3 – As medidas com melhores resultados.

Modelo de Regressão	Três Características	Quatro Características	Cinco Características	Seis Características
Regressão Linear	-	-	-	-
Ridge	'alpha': 1e-06	'alpha': 1e-06	'alpha': 1e-06	'alpha': 1e-06
MLP	'alpha': 0.01, 'hidden_layer_sizes': (100, 50)	'alpha': 0.001, 'hidden_layer_sizes': (100, 50)	'alpha': 0.01, 'hidden_layer_sizes': (100, 50)	'alpha': 0.001, 'hidden_layer_sizes': (100, 50)
Random Forest	'max_depth': 10, 'n_estimators': 100	'max_depth': 10, 'n_estimators': 200	'max_depth': 20, 'n_estimators': 50	'max_depth': 20, 'n_estimators': 200

Tabela 4 – Parâmetros com melhores resultados.

5.2 Resultados do desenvolvimento da ferramenta

5.2.1 Interação com algoritmos de regressão

A lógica por trás das previsões de peso suportadas pelos algoritmos de regressão é integrada de forma transparente na interface gráfica. A ferramenta Kivy simplificou o tratamento de eventos de interface para garantir uma interação suave entre usuários e recursos de previsão.

5.2.2 Estilo visual e capacidade de resposta

KivyMD forneceu um estilo visual consistente com os princípios do Material Design para uma aparência moderna e intuitiva. Além disso, a interface do usuário tornou-se responsiva e pode ser adaptada com eficiência a diferentes tamanhos de tela e dispositivos.

5.2.3 Tela Inicial

Na Figura 2 mostra a tela inicial da ferramenta é o ponto central a partir do qual o usuário pode explorar e utilizar os recursos fornecidos pelo aplicativo.

5.2.4 Tela Predição de Peso

A tela Predição de Peso tem como objetivo fornecer ao usuário uma interface intuitiva e eficiente para inserir medidas específicas para previsão de peso de pequenos ruminantes. A disposição dos componentes visa facilitar o processo de entrada de dados de forma organizada.

Na Figura 3a, cada medida relevante é associada a um campo de entrada na tela, indicando qual medida deve ser inserida. Estes campos possibilitam ao usuário introduzir medidas específicas para o animal em questão. Após o preenchimento de todos os campos, ao clicar no botão "Calcular", inicia-se o processo de predição de peso, o algoritmo de regressão selecionado foi o Ridge. O resultado da predição é então apresentado, proporci-

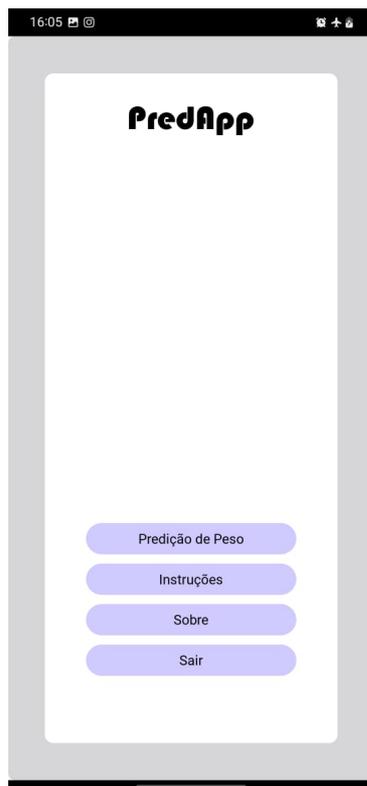


Figura 2 – Demonstração da Tela Inicial.

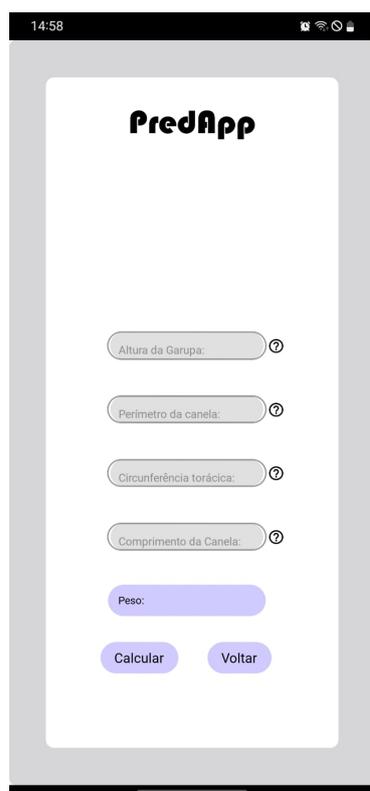
onando aos usuários uma estimativa do peso do animal, conforme ilustrado na Figura 3b. Cada campo de medição é acompanhado por um botão de ajuda. Ao clicar nesse botão, os usuários têm acesso a informações contextuais e dicas sobre como obter ou medir as dimensões específicas necessárias, como demonstrado na Figura 3c. O objetivo é oferecer orientação adicional, especialmente para usuários não familiarizados com a terminologia técnica. A Figura 3d exemplifica que, caso o usuário insira informações incorretas, uma mensagem de erro será exibida.

5.2.5 Tela Instruções

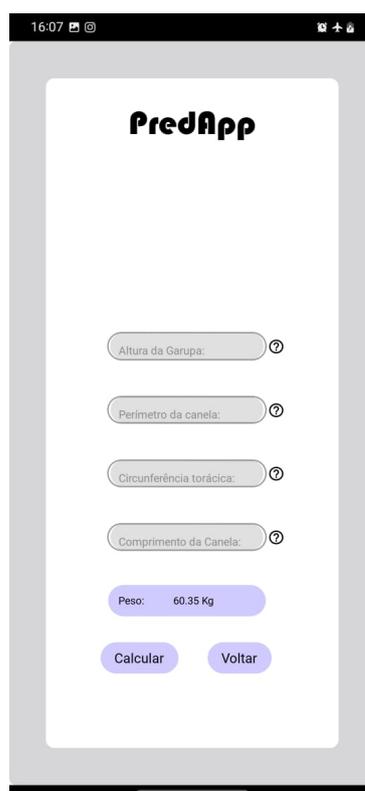
A tela instruções fornece ao usuário instruções detalhadas sobre como usar efetivamente a ferramenta e mais detalhes para realizar cada medida. O objetivo deste recurso é responder às suas perguntas e fornecer uma compreensão abrangente de como o aplicativo funciona, como mostra a Figura 4.

5.2.6 Tela Sobre

A Figura 5 apresenta a tela Sobre que tem como objetivo fornecer informações contextuais sobre a ferramenta de previsão de peso de pequenos ruminantes e fornecer ao usuário uma visão geral de seu desenvolvimento, finalidade e tecnologia utilizada.



(a) Tela para inserir as medidas.



(b) Tela informando o peso.



(c) Tela ao clicar no botão de ajuda.



(d) Tela com mensagem de erro.

Figura 3 – Demonstração da Tela de Predição.

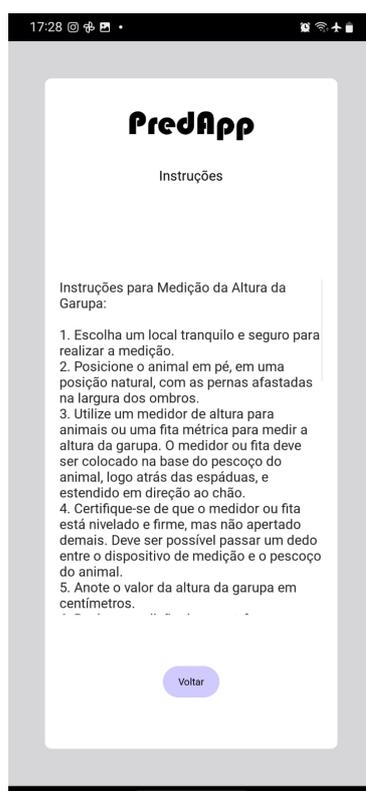


Figura 4 – Demonstração da Tela de Instruções.

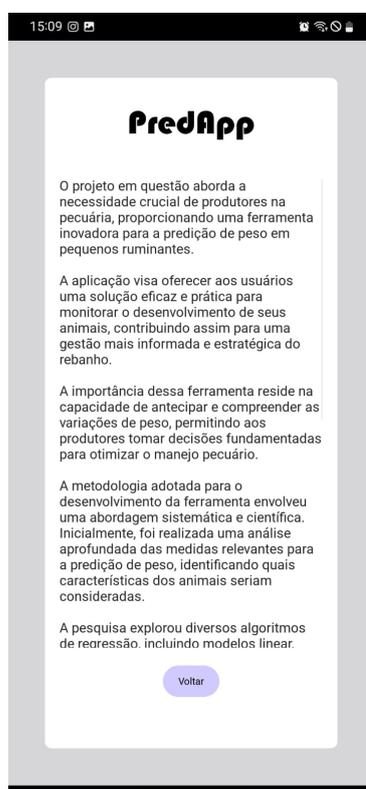


Figura 5 – Demonstração da Tela Sobre

6 Conclusão

Este estudo teve como objetivo abordar uma necessidade dos produtores, apresentando uma ferramenta para a previsão de peso em pequenos ruminantes. A relevância desta ferramenta está em sua capacidade de proporcionar aos produtores meios eficazes de monitorar o desenvolvimento animal, permitindo uma gestão mais precisa e estratégica do rebanho.

Foi explorado quatro algoritmos de regressão na previsão de peso, destacando a importância de identificar o número apropriado de medições e recursos para otimizar o desempenho do modelo. A abordagem sistemática adotada possibilitou uma avaliação criteriosa dos algoritmos, concluindo que tanto o modelo de regressão linear quanto o modelo ridge demonstraram uma robusta capacidade de lidar com os dados coletados ambos alcançando coeficiente de determinação de 0.85 com as características ag, ct, cc, cga, ap, perperna. A seleção das características diretamente mensuráveis contribuiu para a eficácia da ferramenta, simplificando o processo de coleta de dados e garantindo uma relação direta com o objetivo final de prever o peso dos animais. A utilização das bibliotecas Kivy e KivyMD no desenvolvimento da interface gráfica proporcionou uma experiência de usuário moderna e intuitiva, promovendo a acessibilidade da ferramenta.

Em conclusão, neste estudo não apenas identificamos o algoritmo mais adequado para a previsão de peso, mas também transformamos esse conhecimento em uma aplicação móvel funcional. Essa aplicação coloca nas mãos dos produtores uma ferramenta prática e eficiente para a tomada de decisões baseada em dados, contribuindo para o avanço e a sustentabilidade de suas operações pecuárias.

Para futuros trabalhos, há diversas possibilidades de aprofundamento e aprimoramento a partir dos resultados e abordagens deste estudo. Primeiramente, é importante explorar a aplicação de uma gama mais ampla de algoritmos de regressão para a previsão de peso em pequenos ruminantes, a fim de identificar opções ainda mais eficientes e precisas. Além disso, investigar a influência do número e variedade de medições e recursos na previsão de peso pode otimizar ainda mais o desempenho do modelo e sua capacidade de adaptação a diferentes contextos.

Outro aspecto relevante para futuras pesquisas é a validação dos modelos propostos em diferentes raças de pequenos ruminantes e em ambientes de criação diversos, garantindo sua aplicabilidade e precisão em diferentes cenários. Além disso, a inclusão de outras características mensuráveis no modelo de previsão de peso pode expandir sua utilidade e acurácia.

A análise da experiência do usuário na aplicação móvel desenvolvida também merece atenção, buscando identificar possíveis melhorias na interface e na usabilidade para garantir uma experiência mais intuitiva e satisfatória para os produtores. Realizar estudos

de campo para avaliar a eficácia prática da aplicação em diferentes operações pecuárias e coletar feedback dos usuários pode orientar futuras melhorias e refinamentos.

Por fim, explorar integrações com outras tecnologias ou sistemas de gestão agrícola pode ampliar o alcance e a funcionalidade da aplicação móvel, oferecendo aos produtores uma solução mais abrangente e integrada para monitorar o desenvolvimento animal e gerenciar suas operações pecuárias.

Referências

- CAPELO, P. R. A. *Florestas aleatórias e árvores de classificação e de regressão na análise de dados censurados*. Tese (Doutorado), 2022. Citado na página 17.
- CONRADO, V.; ARANDAS, J.; RIBEIRO, M. Modelos de regressão para predição do peso da raça caniné através de medidas morfométricas. *Archivos de zootecnia*, Universidad de Córdoba, v. 64, n. 247, p. 277–280, 2015. Citado 2 vezes nas páginas 21 e 24.
- FERREIRA, M. d. F. M. *Árvores de regressão e generalizações: Aplicações*. Universidade do Porto. Reitoria, 1999. Citado na página 17.
- FIGUEIRA, C. V. Modelos de regressão logística. 2006. Citado na página 16.
- GONZALEZ, L. d. A. *Regressão logística e suas aplicações*. Universidade Federal do Maranhão, 2018. Citado na página 16.
- KORITIAKI, N. A. et al. Predição do peso vivo a partir de mensurações corporais em cordeiros santa inês. *Synergismus scyentifica UTFPR*, v. 7, n. 1, 2012. Citado 2 vezes nas páginas 20 e 24.
- MACIEL, M. d. S. et al. Avaliação de carcaças ovinas da raça santa inês: uma abordagem multivariada. Universidade Federal Rural de Pernambuco, 2022. Citado 2 vezes nas páginas 21 e 24.
- MAHESH, B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], v. 9, n. 1, p. 381–386, 2020. Citado na página 15.
- MAIA, B. *Tipos de Aprendizado de Máquina 3*. 2020. Url<https://dev.to/beatrizmaiads/tipos-de-aprendizado-de-maquina-3-5d66>. Citado 2 vezes nas páginas 6 e 15.
- MARTINS, B. et al. Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements. *Livestock Science*, Elsevier, v. 236, p. 104054, 2020. Citado na página 13.
- MATOS, M. A. Manual operacional para a regressão linear. *Faculdade de Engenharia da Universidade do Porto*, v. 63, 1995. Citado na página 16.
- MOREIRA, M. S. et al. Regressão linear simples e múltipla. *Bioestatística quantitativa aplicada*. Porto Alegre: UFRGS, 2020. Cap. 10, p. 197-216, 2020. Citado na página 16.
- MÜLLER, A. C.; GUIDO, S. *Introduction to machine learning with Python: a guide for data scientists*. [S.l.]: "O'Reilly Media, Inc.", 2016. Citado na página 15.
- NASCIMENTO, H. F. A. d. *Validação de um protótipo de balança automática utilizando o sistema walk-over-weighning em ovinos*. Dissertação (B.S. thesis) — Brasil, 2019. Citado na página 13.
- PAULA, C. F. Rosa e et al. In vivo carcass evaluation and body measurements of dorper ewe in different age groups. 2023. Citado 2 vezes nas páginas 22 e 24.

- RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, v. 29, 2005. Citado na página 17.
- SALAZAR-CUYTUN, R. et al. Prediction of live weight in growing hair sheep using the body volume formula. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, SciELO Brasil, v. 74, p. 483–489, 2022. Citado 2 vezes nas páginas 21 e 24.
- SANTOS, H. P. et al. Peso predito a partir de equações de regressões sobre as medidas biométricas de ovinos mestiços. 2014. Citado 2 vezes nas páginas 20 e 24.
- SILVA, I. M. d. Avaliação morfométrica in vivo e da carcaça de cordeiros texel em sistema de confinamento. 2023. Citado 2 vezes nas páginas 22 e 24.
- SILVA, R. F. d. *Desenvolvimento ponderal de ovinos Santa Inês na fase de cria, manejados em pastagem e suplementados*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2023. Citado 2 vezes nas páginas 23 e 24.
- SONG, X. et al. Automated body weight prediction of dairy cows using 3-dimensional vision. *Journal of Dairy Science*, Elsevier, v. 101, n. 5, p. 4448–4459, 2018. Citado na página 13.
- TRINDADE, A. C. C. Uso de medidas morfométricas realizadas em animais in vivo para predição de peso vivo e características da carcaça de cordeiros terminados em confinamento. Ufra/Campus Belém, 2021. Citado 2 vezes nas páginas 21 e 24.
- VENDRUSCOLO, D. G. S. et al. Estimativa da altura de árvores de *tectona grandis* lf utilizando regressão e redes neurais artificiais. *Nativa*, v. 5, n. 1, p. 52–58, 2017. Citado na página 18.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA
“JOSÉ ALBANO DE MACEDO”**

Identificação do Tipo de Documento

- () Tese
() Dissertação
(X) Monografia
() Artigo

Eu, **Jose Fernando de Carvalho Ferreira**, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação **“Uma Ferramenta para Predição de Peso Corporal em Pequenos Ruminantes em Rebanhos Baseada em Medidas Lineares e Algoritmos de Regressão”** de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI, 02 de Fevereiro de 2024.

Jose Fernando de Carvalho Ferreira

Assinatura