

# Estudo de Algoritmos de Classificação aplicados ao FAMACHA em Caprinos e Ovinos

Mauricio Benjamin<sup>1</sup>, Lazaro Claubert<sup>1</sup>, Pedro Vital<sup>1</sup>

1

R. Cícero Duarte, 905 - Junco, Picos - PI, 64607-670

2

Universidade Federal do Piauí (UFPI)

{mauriciobenjamin}@ufpi.edu.br

## 1. Introdução

A criação de cabras e ovelhas exerce um papel significativo na economia agrícola brasileira, proporcionando diversos benefícios ao setor agropecuário. Tais animais são importantes fontes de proteína, fornecendo carne, leite e derivados para a população e contribuindo para a segurança alimentar do país. A criação é uma atividade que impacta positivamente as áreas rurais, gerando empregos e promovendo o desenvolvimento econômico local [Nascimento 2022]. A pecuária desses animais é reconhecida por sua sustentabilidade, já que são adaptados ao clima tropical brasileiro e têm menor impacto ambiental comparado a outros ruminantes. Ademais, a criação de cabras e ovelhas desempenha um papel essencial na diversificação da produção agrícola do Brasil, permitindo a participação de pequenos e médios produtores [Magalhães et al. 2021].

Cuidados adequados com a saúde de cabras e ovelhas são fundamentais para o bem-estar, produtividade e rentabilidade da criação. Isso inclui medidas preventivas como vacinação, nutrição equilibrada e higiene, além da observação constante para detectar sinais de doença ou parasitismo. Um grande desafio é combater o parasita *Haemonchus contortus*, prevalente em regiões tropicais e subtropicais, causando anemia grave, perda de peso e baixo desempenho reprodutivo [Martins et al. 2016]. A resistência do parasita a medicamentos torna o controle complexo, tornando essencial investir em cuidados de saúde e estratégias eficazes para garantir um rebanho saudável e produtivo [Arshad et al. 2024]. Isso contribuirá para a sustentabilidade da criação de cabras e ovelhas no Brasil, fortalecendo o setor agropecuário e fornecendo produtos de qualidade.

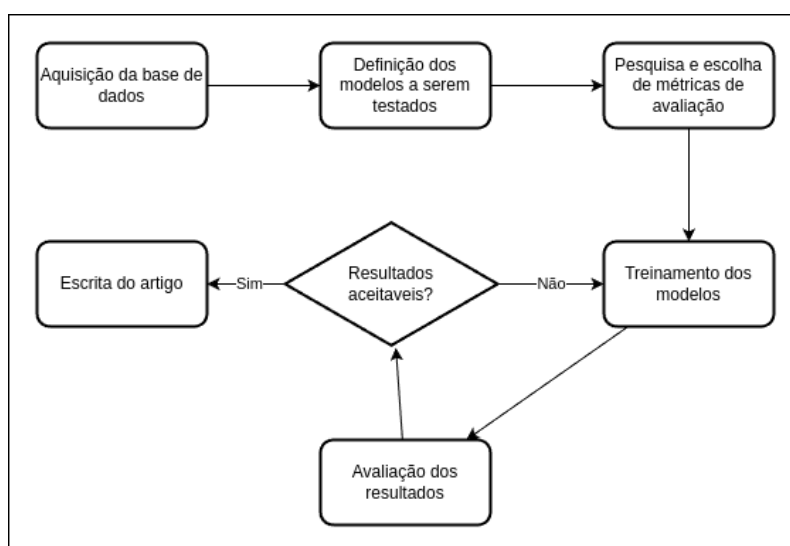
O método FAMACHA é usado na seleção de animais para vermifugação, baseada na observação da coloração da mucosa ocular do animal em comparação com um cartão que associa as cores a seus respectivos graus anêmicos [de Oliveira Diniz et al. 2022]. Este teste consiste em categorizar a coloração em cinco níveis, onde as cores vermelhas indicam boa saúde e as cores esbranquiçadas indicam possíveis traços de anemia, sinalizando a necessidade de tratamento [Teixeira et al. 2023]. Embora seja uma ferramenta útil, sua eficácia pode ser afetada pela subjetividade do examinador, exigindo profissionais treinados para realizar a avaliação consistentemente em todo o rebanho. É importante destacar que variações ou erros podem ocorrer devido à fadiga durante o procedimento, tornando a avaliação do rebanho por um único profissional altamente recomendável.

Com base no exposto, este trabalho consiste em realizar um estudo sobre modelos de aprendizado de máquina, para identificar a necessidade de vermifugação em cabras e

ovelhas. A identificação será efetuada por meio de dados extraídos a partir de imagens da mucosa ocular dos animais, possibilitando uma análise rápida e precisa. A referida metodologia poderá ser implementada em ferramentas fornecendo auxílio para os profissionais do ramo, viabilizando a realização do teste em quantidades substanciais de animais ao longo do tempo, sem que haja uma degradação na coerência dos dados. Com esta solução, mais indivíduos poderão efetuar o teste no mesmo rebanho, resultando em uma maior eficiência e precisão do processo de detecção da necessidade de medicação dos animais.

## 2. Metodologia

A metodologia adotada neste trabalho está contida na Figura 1, onde foi realizada uma busca por uma base de dados adequada para classificação, modelos de classificação e métricas foram definidas com base na base de dados. Ao final foi realizado o treinamento e avaliação dos resultados até que os resultados estivessem satisfazendo nas necessidades do trabalho, para que assim fosse possível realizar a escrita do artigo.



**Figura 1. Metodologia adotada neste trabalho**

### 2.1. Aquisição da base de dados

Para este trabalho foi feita a aquisição de uma base privada de dados com características de cores RGB de imagens de caprinos e ovinos para a classificação do grau FAMACHA. A Tabela 1 apresenta como os dados estão estruturados e apresenta que as cores foram descritas através de média, mediana e desvio padrão de cada canal de cor. A base possui um total de 548 linhas, onde cada linha representa as características de uma imagem referente ao seu respectivo animal.

A coluna FAMACHA contém um grupo das classes 3, 4 e 5 do método FAMACHA, representando animais suscetíveis a anemia e precisam ser direcionados a vermifugação e das classes 1 e 2 referentes a animais saudáveis. Os valores da coluna estão binariados de forma que o valor zero representa ausência anêmica representando pelas classes 1 e 2 e o valor um refere-se as animais anêmicos.

| FAMACHA | Mean_R | Mean_G | Mean_B | Median_R | Median_G | Median_B | Std_R | Std_G | Std_B |
|---------|--------|--------|--------|----------|----------|----------|-------|-------|-------|
| 1       | 39.08  | 43.85  | 48.15  | 41.0     | 43.5     | 48.0     | 35.14 | 26.39 | 25.03 |
| 0       | 51.66  | 39.12  | 41.83  | 61.0     | 39.0     | 41.0     | 36.79 | 21.99 | 21.26 |
| 1       | 152.45 | 128.64 | 125.36 | 174.0    | 130.0    | 129.0    | 62.61 | 43.2  | 39.56 |
| 1       | 51.65  | 29.81  | 33.4   | 59.0     | 33.0     | 37.0     | 28.6  | 11.62 | 13.75 |
| 1       | 141.32 | 126.26 | 140.11 | 164.0    | 131.0    | 149.0    | 65.99 | 54.99 | 61.08 |

**Tabela 1. Tabela de dados FAMACHA com média, mediana e desvio padrão para R, G, B.**

## 2.2. Definição dos modelos a serem testados

Para este trabalho foi feita a escolha dos algoritmos *K-Nearest Neighbors* (KNN), *Support Vector Classifier* (SVC) e *Random Forest Classifier* (RF) para o experimento de classificação, pois possuem abordagens diversas, garantindo uma análise comparativa em cenários diferentes. O KNN é baseado na similaridade entre os dados, sendo particularmente útil para conjuntos de dados onde a proximidade entre pontos é relevante. O SVC, por sua vez, utiliza a margem de separação máxima entre classes, oferecendo potencial em problemas de classificação binária e multiclases, especialmente quando as classes são bem definidas. Por fim, o RF é um modelo de *ensemble* que combina múltiplas árvores de decisão, oferecendo alta precisão a sobreajustes em conjuntos de dados de maior dimensão e com variabilidade significativa. Assim, a escolha desses algoritmos permite explorar abordagens baseadas em distância (KNN), maximização de margens (SVC) e métodos de ensemble (RF), possibilitando uma análise mais ampla sobre o desempenho dos modelos no conjunto de dados.

## 2.3. Pesquisa e escolha de métricas de avaliação

As métricas escolhidas para avaliar o desempenho dos modelos foram acurácia, precisão e F1-score. Tais métricas foram selecionadas para fornecer uma visão da capacidade dos modelos de classificação. A acurácia mede a proporção de previsões corretas sobre o total de previsões, sendo uma métrica direta que mostra o desempenho geral do modelo. Entretanto, a acurácia pode ser insuficiente em casos de classes desbalanceadas, onde é importante analisar outras métricas.

A precisão avalia a proporção de verdadeiros positivos sobre o total de previsões positivas feitas pelo modelo. Essa métrica é importante em cenários onde o custo de falsos positivos é alto, pois indica a confiabilidade das previsões positivas. Por outro lado, o F1-score fornece uma média harmônica entre precisão e sensibilidade (recall), sendo uma métrica útil para equilibrar a análise entre falsos positivos e falsos negativos. O F1-score é especialmente relevante em cenários onde tanto precisão quanto sensibilidade são importantes e quando há desbalanceamento entre classes, fornecendo uma medida única e ponderada do desempenho do modelo.

## 2.4. Treinamento dos modelos

Os modelos foram treinados todos juntos, usando a técnica *K-Fold* com valor de  $k = 5$ , de forma que toda a base foi usada para treino e toda a base foi usada para teste com a divisão de 80% para treino e 20% para teste. O valor de  $k$  foi definido levando em consideração a mais quantidade de dados da classe 1 comparada a classe 0, assim aumentando a quantidade de amostras nos testes para prevenir algum viés negativo. Ao treinar todos os modelos

em conjunto podemos garantir que cada modelo recebe os mesmos dados, de forma que a qualidade de seu resultado vai depender da técnica do modelo e do problema abordado pela base de dados.

## 2.5. Avaliação dos resultados

O K-fold cross-validation é uma técnica de validação cruzada amplamente utilizada para avaliar o desempenho de modelos de *machine learning*, permitindo uma estimativa mais confiável da generalização do modelo em dados não vistos. Nesse método, o conjunto de dados é dividido em  $K$  subconjuntos (ou "folds") de tamanhos aproximadamente iguais. O modelo é treinado  $K$  vezes, a cada iteração utilizando um fold diferente como conjunto de teste, enquanto os  $K - 1$  folds restantes são usados para o treinamento. Após cada iteração, os resultados das métricas de desempenho são registrados, e a média dessas métricas é calculada ao final do processo.

O uso do K-fold é abordagem que ajuda a reduzir o viés e a variância nas estimativas de desempenho, proporcionando uma avaliação mais equilibrada, especialmente em conjuntos de dados limitados. Além disso, o *K-fold cross-validation* é útil para prevenir o problema de *overfitting*, pois garante que o modelo seja testado em diferentes partes dos dados, promovendo uma análise mais completa do seu comportamento e desempenho geral. A Figura 2 apresenta de forma visual o que foi abordado.

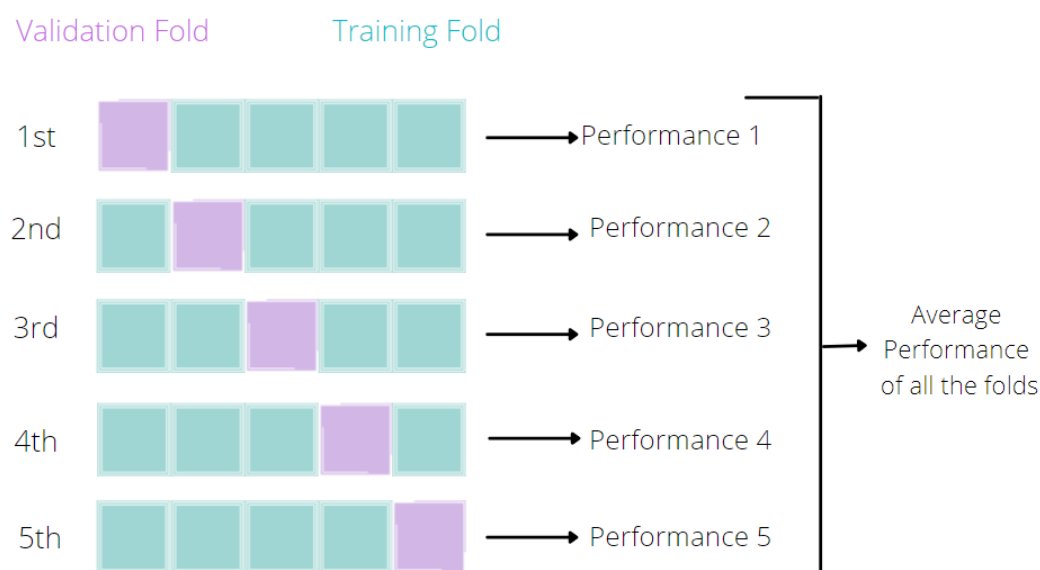
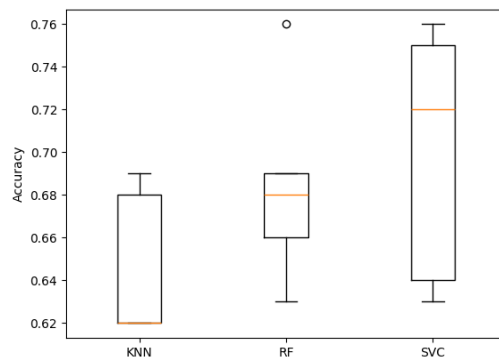


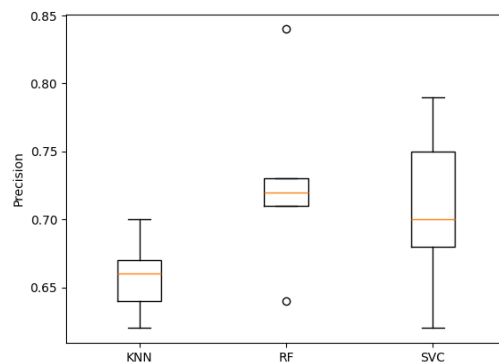
Figura 2. Representação visual do k-fold com k=5 [AI 2023]

## 3. Resultados

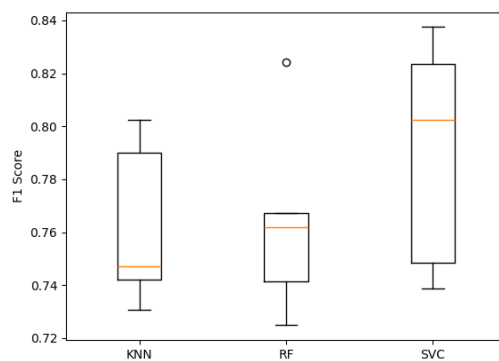
Após a realização dos experimentos, foram coletados os dados e convertidos em figuras para auxiliar na visualização. As Figuras 3, 4 e 5 apresentam os resultados obtidos por cada modelo dado a métrica aplicada. Ao observar as figuras podemos concluir que tanto em acurácia quanto em f1-score, o modelo SVC apresentou vantagens significativas comparadas aos demais modelos, perdendo apenas em precisão para o RF.



**Figura 3. Comparativo de acurácia entre os modelos**



**Figura 4. Comparativo de precisão entre os modelos**



**Figura 5. Comparativo de f1-score entre os modelos**

#### 4. Conclusão

Ao final dos experimentos podemos concluir que o SVC se destacou, tornando-se uma excelente opção para a problemática atual levando em conta a sua acurácia e f1-score. Entretanto, outros modelos que não foram selecionados para experimentação podem desempenhar bons resultados, tornando-se interessante, novos testes e experimentos. Vale

salientar que refinamentos na base de dados também podem ser possíveis melhorias para com os resultados do trabalho, auxiliando na descrição do alvo a ser classificado.

## Referências

- AI, T. (2023). K-fold cross-validation for machine learning models. Accessed: 2024-11-14.
- Arshad, M. F., Burrai, G. P., Varcasia, A., Sini, M. F., Ahmed, F., Lai, G., Polinas, M., Antuofermo, E., Tamponi, C., Cocco, R., et al. (2024). The groundbreaking impact of digitalization and artificial intelligence in sheep farming. *Research in Veterinary Science*, page 105197.
- de Oliveira Diniz, V. A., Teixeira, L. F., de Almeida, G. D., Meirelles, A. C. F., Junior, E. G., and Figueira, P. T. (2022). Utilização do método famacha como auxílio no controle de verminose em ovinos. *Pubvet*, 16:195.
- Magalhães, K. A., Holanda Filho, Z. F., and Martins, E. C. (2021). Pesquisa pecuária municipal 2020: rebanhos de caprinos e ovinos. *Boletim do Centro de Inteligência e Mercado de Caprinos e Ovinos*, 1(16).
- Martins, E. C., Magalhães, K., Souza, J., Guimarães, V., Barbosa, C., and Holanda Filho, Z. (2016). Cenários mundial e nacional da caprinocultura e da ovinocultura. *Boletim ativos de ovinos e caprinos*, 3(2):1–6.
- Nascimento, M. I. d. S. S. (2022). *Produção, comercialização e consumo dos produtos e subprodutos caprinos e ovinos no Nordeste do Brasil*. PhD thesis, Universidade Federal Rural de Pernambuco.
- Teixeira, M., Matos, A. F. I. M. d., Oliveira, T. N. d., and Souza, C. M. (2023). Cartilha controle de verminoses em caprinos e ovinos: recomendações para o controle integrado de verminose em caprinos e ovinos no sertão do são francisco. *PE*.