

Ciência da Computação / Sistemas de Informação

## **Efeito do tamanho do banco de dados na predição do comportamento ingestivo de bovinos a pasto**

Caio Matheus Leite da Silva - 5º módulo de Ciência da Computação, UFLA, bolsista PIBIC/UFLA.

Lázaro H. Silva - Doutorando Zootecnia, UFLA.

Victor A. Oliveira - 6º módulo de Zootecnia, UFLA, bolsista PIBIC/CNPq.

Erick G. Maziero - Coorientador DCA, UFLA.

Daniel R. Casagrande - Coorientador DZO, UFLA.

Marina. A. C. Danes - Orientadora DZO, UFLA. - Orientador(a)

### **Resumo**

O uso de modelos preditivos e sensores em sistemas de produção a pasto possibilita aumentos de produtividade com sustentabilidade. Contudo, ainda há questões práticas a serem respondidas, como quantos dados são necessários para criar um modelo de classificação acurado. Portanto, objetivou-se verificar a eficiência de um modelo preditivo utilizando diferentes tamanhos de banco de dados (BD). Foram realizadas 96 observações visuais de 12 horas cada, em dias não consecutivos, de novilhas da raça Tabapuã em pasto consorciado de Marandu e Amendoim forrageiro. Os comportamentos avaliados foram o pastejo e o não pastejo. Os dados observados foram unidos aos dados obtidos por sensores do tipo acelerômetro de 3 eixos, que foram vestidos na nuca das novilhas. Do BD obtido (607 mil dados), 20% foi destinado para a validação Holdout (121 mil dados), o restante (486 mil dados) foi aleatoriamente separado em 14 BDs com frações de 0.1, 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 e 100% do total de treinamento. A fração do BD foi obtida com base na fração anterior, garantindo que os mesmos dados fossem usados. O modelo preditivo selecionado foi o Random Forest, os dados passaram por uma limpeza e pela busca do melhor hiperparâmetro. Os resultados foram acessados através da acurácia para cada fração do banco de dados. Como esperado, a fração com a menor acurácia (68.8) foi a fração 0.1% e a fração com maior acurácia (79.8) foi a fração 100%. Contudo, a queda na acurácia não foi uniforme, a média da acurácia entre as frações 20, 30, 40, 50, 60, 70, 80, 90 e 100% foi de 79.2 ( $\pm 0.5$ ) enquanto a média entre as frações 0.1, 0.5, 1, 5 e 10% foi de 74.6 ( $\pm 3.4$ ). A distância entre as frações 20 e 10% foi de 1.6% enquanto a distância entre 20 e 30% foi de 0.1. Saber a quantidade mínima de dados é importante para que recursos, como baterias, não sejam utilizados de forma desnecessária. Mais importante que BDs extensos é a qualidade dos dados coletados e a correta análise dos mesmos. Com nossos resultados podemos concluir que para o nosso BD, 20% dos dados (97.200 dados) foram o suficiente para prever com acurácia próxima a totalidade do BD (100%).

Palavras-Chave: Acelerômetro, Ciência de dados, Zootecnia de precisão.

Instituição de Fomento: CNPq, FAPEMIG e CAPES

Link do pitch: <https://youtu.be/xYgxcbFhS2g>