

SUPORTE A DECISÃO DE MANEJO DE PASTAGENS: APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA DISPONIBILIDADE DE FORRAGEM

Wilian da Silva Ricce^{1*}, Tiago Celso Baldissera², Cassiano Eduardo Pinto², Fabio Cervo Garagorry³

¹Epagri - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina, CIRAM, Florianópolis, Santa Catarina, Brasil.

²Epagri - Estação Experimental de Lages, Lages, Santa Catarina, Brasil.

³Empresa Brasileira de Pesquisa Agropecuária, Bagé, RS, Brasil.

*Autor para correspondência: Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina, Rodovia Admar Gonzaga, 1347 – Itacorubi, Florianópolis, SC – Brasil – CEP 88034-901, fone: (48) 3665-5006, wilianricce@epagri.sc.gov.br.

RESUMO

Algoritmos de aprendizado de máquina são utilizados para a predição de valores a partir de resultados obtidos em base de dados gerando informações robustas com baixo custo. O objetivo deste trabalho foi avaliar alguns algoritmos de aprendizado de máquina na predição da disponibilidade de pastagem em sistemas integrados de produção agropecuária. O trabalho foi desenvolvido em Ponta Grossa – PR, de janeiro de 2011 a junho de 2013, em sistema integrado de produção agropecuária consorciado com as seguintes pastagens: *Axonopus catharinensis*, *Brachiaria brizantha* cv. Marandu, *Megathyrus maximus* cv. Aruana, *Hemarthria altissima* cv. Flórida, *Cynodon spp.* híbrido Tifton 85 e *Paspalum notatum* cv. Pensacola. Os tratamentos utilizados foram doses de N (0 e 300 kg/ha/ano), condição de cultivo (sol ou sombra) e severidade de desfolha (50 e 70%). Foram realizados medidos a altura das pastagens, a interceptação da radiação solar e a produção de matéria seca em g/m². Para a estimativa da produção de pastagem, foram avaliados os modelos de regressão linear múltipla, árvore de decisão, random forest e k-vizinhos, avaliando MSE, RMSE, MAE e R². A árvore de decisão foi o modelo com melhor desempenho. Conclui-se que algoritmos de aprendizado de máquina são ferramentas que podem ser utilizadas para a predição da produção de pastagens.

Palavras-chave: Algoritmos de predição, Produção de matéria seca, Sistemas Integrados de Produção Agropecuária.

SUPPORT THE PASTURE MANAGEMENT DECISION: MACHINE LEARNING TO PREDICT FORAGE AVAILABILITY

ABSTRACT

Machine learning algorithms have been gaining importance in several branches of the economy, including agriculture. The objective of this work was to evaluate machine learning algorithms in the prediction of pasture availability. The work was carried out in Ponta Grossa - PR, from January 2011 to June 2013, in an integrated crop-livestock system combined with the following pastures: *Axonopus catharinensis*, *Brachiaria brizantha* cv. Marandu, *Megathyrus maximus* cv. Aruana, *Hemarthria altissima* cv. Flórida, *Cynodon spp.* hybrid Tifton 85 e *Paspalum notatum* cv. Pensacola. The treatments used were doses of N (0 and 300 kg/ha/ano), growing condition (sun or shade) and defoliation severity (50 and 70%). Several cuts were made in the period and then the height of the pastures, interception of solar radiation and the production of dry matter in g / m² were measured. Multiple linear regression, decision tree, random forest and k-neighbors were studied for the estimation of pasture production, evaluating MSE, RMSE, MAE and R². Decision tree was the model with the best performance. It is concluded that machine learning algorithms are interesting tools to be used in the prediction of pasture production.

Keywords: Prediction algorithms, Dry matter production, integrated crop-livestock systems.

INTRODUÇÃO

A produção animal no Brasil e em Santa Catarina é altamente baseada em sistemas pastoris, em que as pastagens estão implantadas dentro de uma gama de fatores bióticos e abióticos para expressar o potencial produtivo, visando atender as exigências do rebanho. Em sistemas integrados de produção agropecuária, a complexidade é ainda maior, pois integra diversos componentes dentro de um mesmo sistema, como por exemplo a introdução de árvores (MORAES et al., 2014). Isto reflete em maiores desafios para o sistema produtivo e para o uso de ferramentas de predição.

A estimativa da produção de pastagens é de grande interesse para os produtores, pois auxilia em diversas tomadas de decisão, como determinação de lotação animal, adubação, dentre outros, refletindo diretamente no desempenho animal e consequentemente na rentabilidade das propriedades.

O aprendizado de máquinas para a agricultura usa algoritmos para analisar dados e, a partir de métodos estatísticos, aprender com o conjunto de dados para trazer uma resposta ou para classificar grupos. No caso de disponibilidade de forragem em pastagens, onde a variável alvo é contínua, algoritmos supervisionados são de interesse para gerar as previsões, pois são fornecidos os dados com a resposta que está sendo estudada para o treinamento do algoritmo. A partir desse treinamento, o algoritmo vai aplicar o que foi aprendido em outro conjunto de dados e trazer os resultados.

O objetivo deste trabalho foi avaliar modelos de aprendizado de máquina para a previsão da produção de pastagem em sistema integrado de produção agropecuária com a presença de árvores.

METODOLOGIA

O trabalho foi desenvolvido na Estação Experimental do Instituto Agrônomo do Paraná (IAPAR) em Ponta Grossa - PR (latitude 25°07'22"S, longitude 50°03'01"O e altitude 880m). o clima segundo Koppen é Cfb (ALVARES et al., 2013), o solo é classificado com Latossolo Bruno distroférico, com textura em torno de 30% de argila.

Seis espécies forrageiras perenes C4, de amplo uso em sistemas pastoris do Brasil foram utilizadas e avaliadas em sistema sombreado por Eucalipto e em pleno sol. As espécies utilizadas foram *Axonopus catharinensis* (Ac), *Brachiaria brizantha* cv. Marandu (Mb), *Megathirsus maximus* cv. Aruana (Mm), *Hemarthria altissima* cv. Flórida (Ha), *Cynodon* spp. hybrid Tifton 85 (Cc) e *Paspalum notatum* cv. Pensacola (Pn). O sistema sombreado era composto por árvores de *Eucalyptus dunnii* e durante o experimento, o nível de sombreamento se manteve entre ~40 a ~59%.

As forrageiras foram implantadas no ano de 2010 em parcelas de 4,5m² (1,5 x 3m) no pleno sol vs. 100m² (5 x 20m) no sistema sombreado. Em cada sistema, foram avaliadas duas doses de nitrogênio (0 e 300 kg ha⁻¹ ano⁻¹) em três repetições. No pleno sol também foram avaliadas as severidades de desfolha de 50 e 70%. O nitrogênio foi aplicado, em forma de uréia, em dose única no início da estação de crescimento de cada ano.

Nos anos de 2011, 2012 e 2013 foram coletados os dados, onde foram avaliadas as variáveis de massa de forragem, altura do dossel e interceptação luminosa. As medições das parcelas eram realizadas de forma aleatória, já no ambiente sombreado, eram realizadas medições em posições equidistante dos renques de árvores (2, 4, 10, 16 e 18m de um dos renques), com o objetivo melhor avaliar o efeito da sombra sobre as variáveis. A determinação

da massa de forragem foi realizada apenas no estrato superior (50% da altura da pastagem), utilizando quadros de 0,25m². No pleno sol, em cada unidade experimental foi realizado um corte, já na sombra, foram realizados cinco cortes. Após o corte, as amostras foram secas em estufa a 60°C até atingirem peso constante. A interceptação luminosa (IL) foi medida com o uso de ceptômetro (Decagon - AccuPAR LP-80). A altura do dossel foi medida com uso de um bastão graduado (*sward stick*, Barthram, 1985), em 10 pontos das parcelas a pleno sol e 20 pontos no sistema sombreado. As pastagens foram manejadas com cortes mecânicos ao resíduo de 30 ou 50% da altura correspondente aos 95% de IL.

Foi utilizado o software Orange (versão 3.26, DEMSAR J. et al., 2013) para a estruturação dos modelos e análise dos resultados. Os modelos utilizados foram: regressão linear sem regularização (Linear Regression); árvores de decisão com número mínimo de 5 (Tree), instâncias por folha, com subconjuntos de até 20 amostras e profundidade máxima de 5; random forest com 5 árvores e k-vizinhos (KNN) com 5 vizinhos. Os dados foram divididos na proporção de 60% para treino (794 dados) e 40% para teste (528 amostras) de um total de 1322 dados. Para a comparação dos modelos, foram utilizadas as métricas Erro médio quadrado (MSE), raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e coeficiente de determinação (R²), disponibilizadas pelo software.

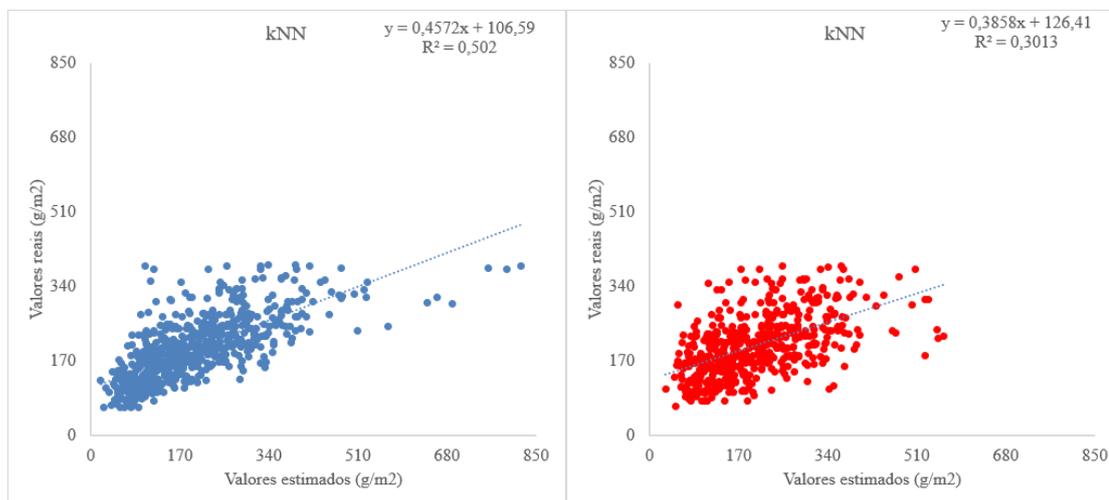
RESULTADOS

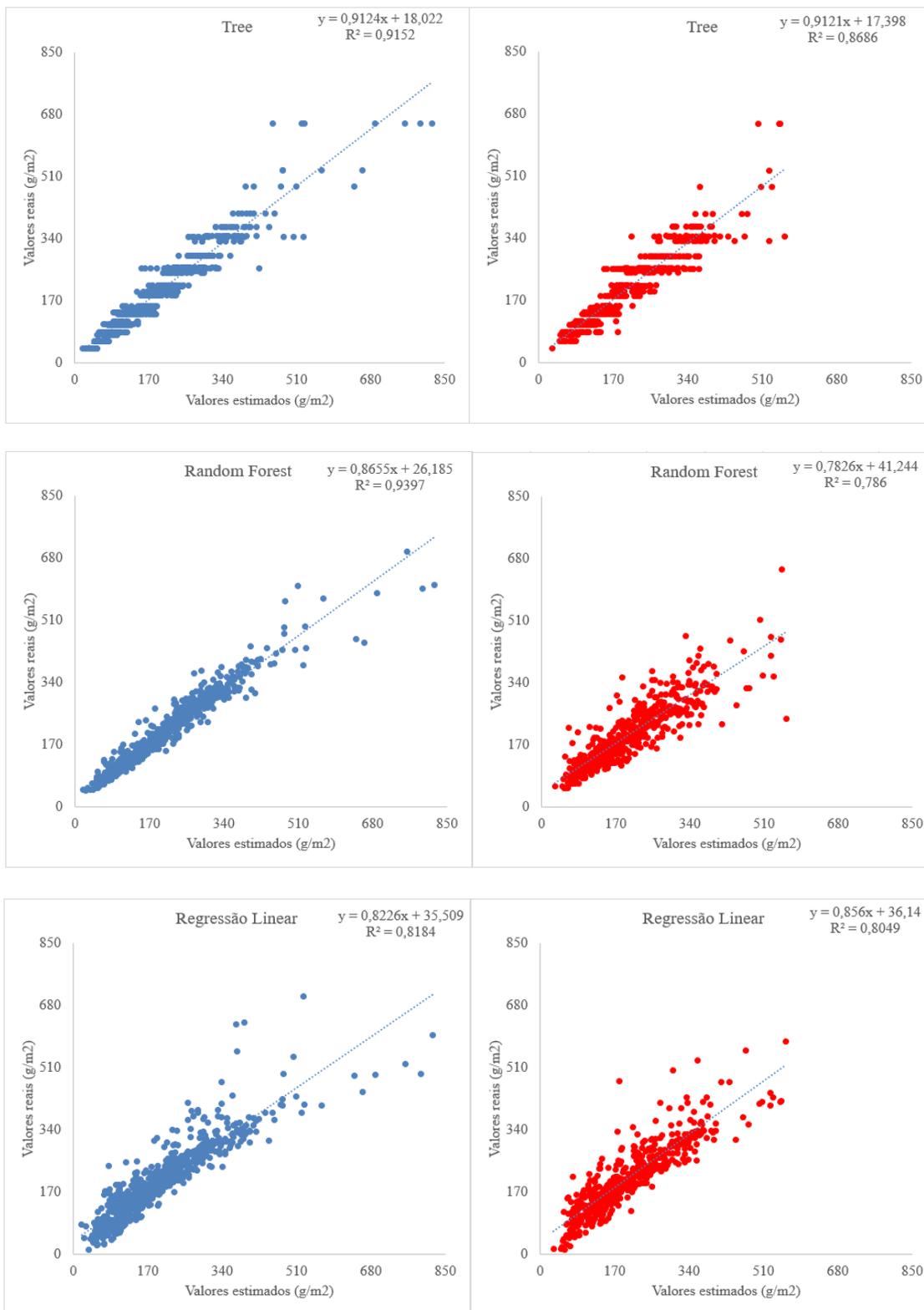
As métricas utilizadas para avaliação dos modelos são apresentadas na Tabela 1, onde os modelos o modelo Tree apresentou melhor desempenho tanto nos dados de treinamento quanto no teste, seguido pelo modelo Random Forest (Tabela 1). Para melhor visualizar os resultados preditos em relação aos dados originais, foram construídos gráficos de dispersão apresentados na Figura 1.

Tabela 1. Erro médio quadrado (MSE), raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e coeficiente de determinação (R^2) utilizados para a avaliação das predições obtidas pelos modelos de aprendizado de máquina estudados, resultados apresentados com os dados de treino e dados de teste.

Modelos	MSE	RMSE	MAE	R^2
Dados de treino (60%)				
kNN	6202,8	78,76	55,51	0,50
Tree	1104,0	33,23	21,93	0,91
Random Forest	819,7	28,63	17,59	0,93
Regressão Linear	2231,2	47,24	31,48	0,82
Dados de teste (40%)				
kNN	7135,4	84,47	63,86	0,28
Tree	1317,9	36,30	25,60	0,87
Random Forest	2117,5	46,02	31,38	0,79
Regressão Linear	2007,8	44,81	31,42	0,80

Figura 1. Dispersão entre os valores reais de produção das pastagens (g/m^2) e os valores preditos pelos modelos estudados. Pontos em azul representando os dados de treinamento e os pontos vermelhos representando os dados de teste.





DISCUSSÃO

Existe uma grande complexidade nos dados utilizados para a predição dos modelos, com diferentes níveis de severidade de desfolha (70 e 50%), doses de nitrogênio (0 e 300 kg ha⁻¹

ano⁻¹) e cortes de massa de forragem em diferentes épocas do ano, contudo, dois modelos testados apresentaram (Tree e Random Forest) bons ajustes e predição dos dados, com valores similares ou superiores a regressão linear. Já o modelo KNN teve resultados insatisfatórios de ajuste. Na Figura 2, vale observar que os dados preditos pelo modelo Tree vem em classes, não apresentando distribuição contínua nos resultados.

CAMARGO et al. (2017), avaliando a previsão da produtividade de pastagens a partir de indicadores de solo, obtiveram resultados de até 92% de acerto com o uso do algoritmo Random Forest, porém o conjunto de dados utilizados foi bem pequeno, o que pode levar ao overfitting dos modelos. Comparando com os resultados aqui apresentados e em função da complexidade dos dados, fica claro que devem ser consideradas outras variáveis para que as predições sejam mais precisas.

CONCLUSÃO

Algoritmos de aprendizado de máquina são ferramentas viáveis para serem utilizadas na predição de produção de pastagens.

REFERÊNCIAS

ALVARES, C. A.; et al. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, p. 711–728, 2013.

CAMARGO, L. da S.; et al. Previsão da Produtividade de Pastagens a partir de Indicadores de Solo. Em: V Simpósio da Ciência do Agronegócio, v. 1, p. 40-48, 2017.

DEMSAR, J.; et al. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, v. 14, p. 2349–2353, 2013.

MORAES, A.; et al. Integrated crop–livestock systems in the Brazilian subtropics. *Europ. J. Agronomy*, v. 57, p. 4-9, 2014.