

ESTIMATIVA DE DIAS TRABALHÁVEIS EM ATIVIDADES AGRÍCOLAS UTILIZANDO RANDOM FOREST

RAFAEL DOS SANTOS ESTECHE¹; DANILO DOS SANTOS LEITE²; JULIANA PEREIRA PINO³; FABRÍCIO ARDAIS MEDEIROS⁴; MAURO FERNANDO FERREIRA⁵

¹ Universidade Federal de Pelotas – PPGSPAF – rafael.esteche@ufpel.edu.br

² Universidade Federal de Pelotas – PPGSPAF – daniloengagricola@gmail.com

³ Universidade Federal de Pelotas – PPGSPAF – moviciclo@gmail.com

⁴ Universidade Federal de Pelotas – PPGSPAF – medeiros.ardais@gmail.com

⁵ Universidade Federal de Pelotas – PPGSPAF – mauof@ufpel.edu.br

1. INTRODUÇÃO

Para o desenvolvimento e aumento da produtividade, o setor agrícola depende de recursos naturais, como água, solo e das condições climatológicas do local. Desse modo, a agrometeorologia, atua como um campo da ciência que auxilia a atividade agrícola a aumentar a eficiência através da redução dos custos envolvidos nas tarefas e maior conhecimento sobre os riscos climáticos (RADIN & MATZENAUER, 2016).

Para atingir mais eficiência nas operações agrícolas, leva-se em consideração as variações climáticas para analisar a real disponibilidade de tempo para as atividades (ESTRADA et al., 2015). A análise das condições climatológicas propicia a adequação das atividades mecanizadas, tornando-se essencial para determinar de forma precisa as dimensões, capacidade e tipos de implementos a serem usados. Deve-se levar em conta as capacidades operacionais apropriadas dos equipamentos para o manejo da cultura ao longo do seu desenvolvimento, fator que pode influenciar na diminuição de custos e desta forma em maior rentabilidade da propriedade (MELLO et al., 2019).

Quando há umidade excessiva no solo, ocorre dificuldade de locomoção dos tratores e máquinas para a colheita, assim torna-se relevante otimizar o número de dias favoráveis para a utilização das máquinas agrícolas (ATAÍDE et al., 2012).

Estudos na literatura buscam estimar os dias favoráveis para se trabalhar com máquinas agrícolas (ATAÍDE et al., 2012; ESTRADA et al., 2015), ambos estudos empregaram o uso dos métodos de cadeias de Markov para determinar os dias favoráveis para o trabalho de campo, no qual a técnica estabelece o dado evento ocorrer em relação ao evento que ocorreu no dia anterior.

Com o avanço tecnológico foi obtido dados e informações úteis geradas pela agricultura digital. A implementação de novas tecnologias, como inteligência artificial, identifica padrões e conhecimentos importantes para a tomada de decisão do produtor (VILLAFUERTE et al., 2018), sendo possível por exemplo, a predição das quantidades de nutrientes nitrogênio, fósforo e potássio (NPK) do solo usando *random forest* (MOHAPATRA et al., 2017), verificar os dias favoráveis com rede neural artificial (MOGHADDAM et al., 2019), estimar a produtividade das culturas usando *random forest* (BASHA et al., 2020). Todos os estudos empregam técnicas de inteligência artificial. A Random Forest é um conjunto de diversas árvores de decisão, em que cada árvore é construída com um subconjunto aleatório de variáveis, e todas suas previsões são combinadas para gerar o resultado. Essa diversidade entre as árvores, auxilia a reduzir o risco de sobreajuste, que acontece

quando um modelo “decora” demais os dados de treino e acaba aprendendo ruídos ou detalhes específicos, em vez de entender o padrão geral.

Analisando os estudos presentes na literatura e tendo o conhecimento das vantagens da implementação das novas tecnologias no campo, o objetivo deste trabalho é a criação de uma metodologia, baseada em *random forest*, para estimar os dias trabalháveis com máquinas agrícolas.

2. METODOLOGIA

Para a criação da proposta da metodologia foram utilizados dados meteorológicos diários, referentes ao período de 1976 a 2025, obtidos do banco de dados do Instituto Nacional de Meteorologia (INMET), na área de abrangência da Estação Meteorológica Convencional de Pelotas/RS (código da Estação 83985). A área para a aferição do modelo teve as coordenadas de Latitude -31.78333333° , Longitude de -52.41666666° e Altitude de 13m.

O solo incluído no modelo possuiu a classificação conforme a Sistema Brasileiro de Classificação de solos (SiBCS) como Planossolo Háplico Eutrófico Solódico (DA CUNHA & SILVEIRA, 1996).

Adotou-se o valor de 100mm para a capacidade de água disponível no solo (CAD). Segundo ATAÍDE et al. (2012) as condições para se considerar o dia como favorável ao trabalho de máquinas agrícolas são as precipitações menores que 5mm e o armazenamento de água no solo (ARM) entre 40% e 90% da CAD.

O banco de dados é composto por 10 colunas e 17.899 linhas, que através dos critérios determinados anteriormente foi inserido uma nova coluna contendo a informação sobre favorabilidade de trabalho com implementos agrícolas.

A plataforma Google Colab é uma ferramenta gratuita baseada na nuvem que permite executar código da linguagem de programação Python para criação e validação de modelos de aprendizado de máquina, como exemplo a *Random Forest*. Nesta, é possível importar as bibliotecas necessárias para a manipulação dos dados e criação do modelo de aprendizado de máquina, subdividindo o conjunto de dados em 80% para o treinamento do modelo e o restante para o teste. Implementando a separação das variáveis, em variável-alvo (coluna de dias favoráveis) e características (o restante das colunas).

A avaliação do modelo é feita utilizando a acurácia, que verifica a precisão do modelo, servindo para avaliar se o previsto como positivo foi realmente positivo, sensibilidade, verifica se o que o modelo colocou como verdadeiro e é verdadeiro e por fim, realiza-se uma tabela mostrando quantos dias foram favoráveis para se trabalhar com máquinas agrícolas, comparando o que foi observado com o previsto pelo modelo.

3. RESULTADOS E DISCUSSÃO

O Quadro 1 apresenta o resultado do desempenho do modelo, com as quatro métricas: Acurácia, Precisão, Sensibilidade e F1-score macro. Que verifica o quão preciso o modelo é para determinar o dia como trabalhável (1) e não trabalhável (0). Pode-se observar que a acurácia do modelo foi de 0.86, que significa, em geral, ele faz classificação correta 86% das vezes, apesar de que a sensibilidade não foi muito boa para classificar os dias não trabalháveis, que pode ser por conta de termos poucos dados de não trabalháveis no banco de dados disponível quando foi realizado a separação em treino e teste. Entretanto, por ter mais dias trabalháveis no banco de dados, o modelo tem uma alta sensibilidade para classificar, com uma precisão de 86 por cento.

Quadro 1- Parâmetros de desempenho do Modelo.

	Acurácia	Precisão	Sensibilidade	F1-score
0	0.84	0.72	0.17	0.28
1		0.86	0.99	0.92

Fonte: Autores, 2025.

No Quadro 2, pode-se verificar a média de dias previstos pelo modelo comparados ao conjunto de dados que foi determinado para classificar se o dia seria favorável. Como destacado anteriormente, o modelo está prevendo mais dias favoráveis do que deveria, pois o banco de dados do teste, tem mais dias favoráveis, se comparado a não favoráveis. Uma solução para ajustar, é utilizando um banco de dados maior e realizando a técnica de validação cruzada k-fold, que serve para aumentar o número de amostras muito pequeno, auxiliando a ter mais informações de número de dias não trabalháveis para o modelo.

Quadro 2 – Medias de dias observados e previstos pelo modelo

Mês	Dias Observados	Dias Previstos
Janeiro	21	25
Fevereiro	23	25
Março	26	28
Abril	26	27
Maio	27	28
Junho	26	28
Julho	26	29
Agosto	26	29
Setembro	23	25
Outubro	23	25
Novembro	26	28
Dezembro	26	29

Fonte: Autores, 2025.

O objetivo futuro deste trabalho é ajustar o conjunto de dados para ter mais informações sobre as questões climatológicas, acrescentando dados de satélites, para melhorar o desempenho do algoritmo.

4. CONCLUSÕES

Com esses resultados podemos concluir que é possível conseguir estimar os dias favoráveis a ser operável através da técnica de *machine learning random forest*, buscando uma análise mais precisa e rápida para o planejamento das execuções de máquinas agrícolas. Entretanto, para melhor eficácia da metodologia, deverá ser implementado mais dados e acrescentar mais critérios para a favorabilidade de trabalho em máquinas agrícolas, visando adicionar ao modelo amostras de não trabalháveis, aprimoramento assim, a comparação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ATAÍDE, L.T.; CARAMORI, P.; RICCE, W.S.; SILVA, D.A.B.; SOUZA, J.R.P. Probabilidade de dias potencialmente úteis de trabalho durante o ano em Londrina, **Semina: Ciências Agrárias**, Londrina, v. 33, n. 6, p. 2215-2226, 2012.
- BASHA, S.M.; RAJPUT, D.S.; JANET, J.; SOMULA, R.S.; ARUMUGAM, S.R. Principles and practices of making agriculture sustainable: crop yield prediction using Random Forest. **Scalable Computing: Practice and Experience**, Polônia, v.21, n.4, p.591-599, 2020.
- DA CUNHA, N. G.; SILVEIRA, R. J. da C. **ESTUDO DOS SOLOS DO MUNICÍPIO DE PELOTAS**. EMBRAPA/CPACT, Ed. UFPEl, 1996. 50 p.
- ESTRADA, J. S.; SCHLOSSER, J. F.; FARIAS, M. S.; SANTOS, G. O.; RÜDELL, I. Y. P. Metodologia para estimar o número de dias trabalháveis com máquinas agrícolas. **Revista Ceres**, Viçosa, v. 62, n.4, p. 410-414, 2015.
- MOGHADDAM, A.K.; ROHANI, A.; MOGHADDAM, L.K.; TROUJENI, M.E. Developing a Radial Basis Function Neural Networks to Predict the Working Days for Tillage Operation in Crop Production, **International Journal Of Agricultural Management and Development**, Irã, v.9, n.2, p. 119-133, 2019.
- MOHAPATRA, A.G.; KESWANI, B.; LENKA, S.K. Soil npk prediction using location and crop specific random forest classification technique in precision agriculture. **International Journal of Advanced Research in Computer Science**, India, v. 8, n.7, p. 1045-1050, 2017.
- RADIN, B., MATZENAUER, R. Uso das informações meteorológicas na agricultura do Rio Grande do Sul. **Agrometeoros**, Passo Fundo, v.24, n.1, p.41-54, 2016.
- VILAFUERTE, A.M.; VALADERES, F.G.; CAMPOLINA, G.F.; SILVA, M.G.P. Agricultura 4.0 - estudo de inovação disruptiva no agronegócio brasileiro. In: **9TH INTERNATIONAL SYMPOSIUM ON TECHNOLOGICAL INNOVATION**, Aracaju, 2018, v.9, p. 150-162.