

Um Modelo Preditivo de Incidência da Ferrugem Asiática da Soja: Uma Aplicação no Estado de Mato Grosso

Moiseis dos S. Ceconello,¹ André Krindges²
UFMT, Cuiabá, MT, FCA - UNICAMP, Campinas, SP
Nayara L. S. Zagui³
IFMT, Juína, MT

Resumo. Este estudo apresenta um modelo preditivo para a incidência da ferrugem da soja, utilizando técnicas de ciência de dados. A modelagem proposta foi aplicada para prever ocorrências da doença no município de Campo Verde, localizado no estado de Mato Grosso. Duas abordagens de ciência de dados são comparadas para avaliar a eficácia e a robustez dos resultados do modelo. **Palavras-chave.** Machine Learning, Regressão Logística, Support Vector Machines, Ferrugem Asiática

1 Introdução

Presente no Brasil desde o ano de 2001, a ferrugem asiática é uma doença infecciosa que figura entre as doenças mais devastadoras e de rápido avanço que afetam a cultura da soja (*Glycine max*). Tendo como agente causador o fungo *Phakopsora pachyrhizi*, a ferrugem asiática pode causar uma redução de até 90% na produtividade da lavoura, trazendo inúmeros prejuízos sociais e econômicos para a agricultura brasileira e mundial [1].

De acordo com o Companhia Nacional de Abastecimento, o cultivo da soja é uma das principais fontes da economia do estado de Mato Grosso, com uma produção de aproximadamente 46 milhões de toneladas na safra 2022/2023 [5]. Essa produção representa aproximadamente 30% da produção nacional, cerca de 10% do Produto Interno Bruto do estado em 2023, consolidando Mato Grosso na liderança da produção dentre as unidades federativas do país. Além do mais, a atividade econômica advinda do cultivo da soja é a principal fonte de geração de riqueza em muitos municípios do estado. Na Figura 1 apresentamos os dois municípios com os maiores números acumulados de ocorrências da ferrugem asiática da soja no período 2004-2023

Considerando a importância do cultivo da soja para a economia do estado de Mato Grosso, e também do país, tornam-se indispensáveis medidas de contenção e controle da dispersão da ferrugem asiática nas áreas de plantio da cultura, impactando os custos da produção agrícola. Dessa forma, modelos de previsão de incidência se tornam aliados indispensáveis para a redução de custos e aumento da produtividade agrícola

Nessa perspectiva, em [10] os autores propuseram um modelo matemático espaço-temporal para descrever a dinâmica de propagação da ferrugem asiática nas áreas de plantios do estado de Mato Grosso. A modelagem apresentada pelos autores consistem em considerar equações diferenciais parciais do tipo difusão-advecção para modelar a dispersão de esporos do fungo. A partir da

¹moiseis@gmail.com

²krindges@gmail.com

³nayara.sartor@jna.ifmt.edu.br

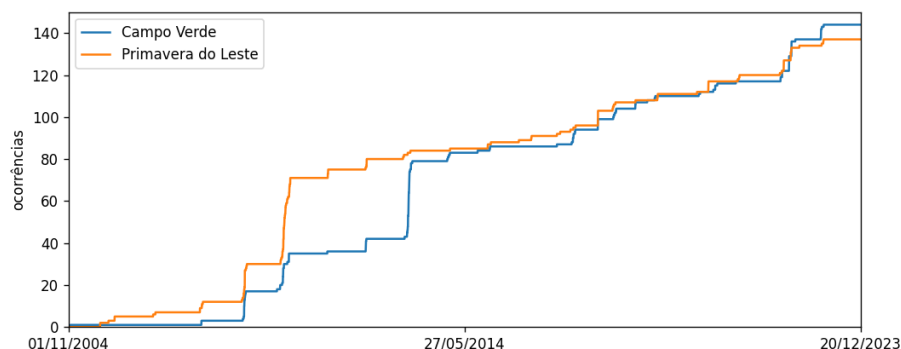


Figura 1: Número acumulado de ocorrências de ferrugem-asiática nos municípios de Campo Verde e Primavera do Leste, em Mato Grosso. Fonte: autor.

dinâmica de dispersão e considerando dados climáticos, os autores utilizam a lógica fuzzy para definir uma medida cuja magnitude indica o grau de vulnerabilidade à doença da região.

A abordagem que propomos aqui vai em outra direção: a partir dos dados de observações históricas, propomos usar métodos de ciências de dados para treinar modelos preditivos da incidência da ferrugem asiática em uma determinada localidade levando em consideração tanto a presença da doença nas proximidades quando condições climáticas da região de interesse.

Este artigo está estruturado em quatro seções principais. A Seção 2, apresenta as características e dinâmica da ferrugem asiática da soja. Em seguida, na Seção 3, apresentamos o modelo proposto para prever a incidência da ferrugem asiática com base em técnicas de ciência de dados. Na Seção 4, apresentamos uma aplicação prática do modelo preditivo considerando os dados reais de incidência da ferrugem asiática no município de Campo Verde, situado no estado de Mato Grosso. Por fim, na Seção 5, apresentamos as conclusões finais sobre a proposta apresentada.

2 Caracterização e Dinâmica da Ferrugem Asiática

A ferrugem asiática da soja é uma doença que se caracteriza por manchas de cor marrom-avermelhada nas folhas inferiores da soja, que posteriormente se espalham para as folhas superiores, hastes e vagens. Essas manchas são constituídas por esporos do fungo *Phakopsora pachyrhizi*, que são facilmente dispersos pelo vento e pela água da chuva. As condições favoráveis para o desenvolvimento da ferrugem asiática incluem temperaturas entre 20°C e 28°C e alta umidade relativa do ar [7, 8].

A transmissão da ferrugem asiática ocorre principalmente por meio de esporos do fungo que são transportados pelo vento. Os esporos podem viajar grandes distâncias e se depositar nas folhas da soja, onde encontram condições favoráveis para germinar e infectar a planta. Além disso, a ferrugem asiática pode ser transmitida por meio de sementes infectadas e equipamentos agrícolas contaminados.

No Brasil, as estratégias de manejo para o controle da doença incluem a ausência de plantio de soja e a eliminação de plantas voluntárias na entressafra por meio do vazio sanitário, a utilização de cultivares de ciclo precoce e semeaduras no início da época recomendada como estratégia de escape da doença, a utilização de cultivares com genes de resistência, o monitoramento da lavoura desde o seu início de desenvolvimento para definir o melhor momento do controle químico, a utilização de

fungicidas preventivamente ou no aparecimento dos sintomas e a definição de janelas de semeaduras para reduzir o número de aplicações de fungicidas [1, 2].

De toda forma, o uso das técnicas recomendadas de controle da ferrugem asiática aumenta os custos de produção além de causar impactos negativos sobre o meio ambiente e a saúde humana.

3 O Modelo Preditivo

Como mencionado anteriormente, há dois fatores predominantes na dinâmica de transmissão da ferrugem asiática: a presença de esporos e condições climáticas favoráveis para o desenvolvimento do fungo.

Devido as característica de dispersão, a presença de esporos do fungo tanto na localidade de interesse quanto em sua vizinhança são fatores determinantes para a proliferação da doença. Além da presença de esporos na vizinhança, as condições climáticas na localidade de interesse também são fatores preponderantes para a fixação da ferrugem da soja.

Tendo em vista que os dados de ocorrências são indicativos da presença de esporos na vizinhança da localidade de considerada, a ideia central do modelo preditivo que propomos é usar dados tanto históricos de ocorrência na vizinhança bem como dados climáticos de uma determinada localidade, até um instante t , para prever a ocorrência da doença nessa localidade em um período futuro.

De maneira mais precisa, considere uma localidade l e seja Ω_l uma vizinhança de l . Para um determinado instante t , o modelo preditivo genérico é definido pela equação

$$p_{(t,t+\tau_1]}^{(l)} = M_{\Omega_l}(o_{[t-\tau_2,t]}, c_{[t-\tau_3,t]}) \quad (1)$$

em M_{Ω_l} é um modelo de classificação, $p_{(t,t+\tau_1]}^{(l)}$ é alguma medida que indica a ocorrência/não-ocorrência da doença no intervalo $[t, t+\tau_1]$ na localidade l , $o_{[t-\tau_2,t]}$ são as ocorrências na vizinhança Ω_l no período $[t-\tau_2, t]$ e $c_{[t-\tau_3,t]}$ representa as condições climáticas da localidade l no intervalo $[t-\tau_3, t]$.

4 Aplicação

Com a finalidade de validar a proposta anterior, vamos considerar uma aplicação do modelo para predição de ocorrências no município de Campo Verde - MT. De acordo com os dados processados a partir do site do Consórcio Antiferrugem [6], o município apresenta o maior número de ocorrências da doença em MT no final da safra 2022-2023. Nessa aplicação, consideramos a vizinhança Ω_l como sendo a região definida pelos 8 municípios mais próximos de Campo Verde, sendo eles: Dom Aquino, Jaciara, Chapada dos Guimarães, Juscimeira, Nova Brasilândia, Poxoréo, Primavera do Leste e Cuiabá.

Para as condições climáticas, utilizamos os dados de temperatura, umidade relativa e número de horas do dia com umidade relativa superior à 80%. Para cada uma dessas variáveis, consideramos a média dos 15 dias anterior a data de referência. Esses dados foram processados a partir de interpolação dos dados históricos das estações de coleta de dados disponibilizados pelo Instituto Nacional de Metrologia [4].

Os parâmetros τ_1 e τ_2 foram definidos como 7 e 30 dias, respectivamente. Assim, construímos um banco de dados contendo 1753 (1423 valores 0 e 330 valores 1) entradas com 15 colunas como descrito na Tabela 1.

Para fins comparativos, empregamos este conjunto de dados para ajustar os parâmetros de dois modelos de aprendizado de máquina: Regressão Logística e *Support Vector Machine* (SVM). Seguindo as práticas convencionais em processos de aprendizado de máquina [3], as seguintes etapas foram realizadas para o ajuste de parâmetros:

Tabela 1: Descrição do banco de dados para os algoritmos de aprendizado de máquina.

Nome da coluna	Descrição da informação
Data	Dia de referência
Município da vizinhança	Número de ocorrências no município nos últimos 30 dias
Temperatura	Média da temperatura diária dos 15 dias anteriores a data de referência
Umidade relativa	Média da umidade relativa diária dos 15 dias anteriores a data de referência
Umidade $\geq 80\%$	Média dos 15 dias anteriores do número de horas do dia com umidade não inferior a 80%
Ocorrência nos próximos 7 dias	Valor 0 ou 1 indicando a não-ocorrência/ocorrência nos próximos 7 dia a partir da data de referência

1. Divisão banco de dados em dois conjunto distintos: treinamento (70%) e teste (30%);
2. Ajuste dos parâmetros do modelo de aprendizado de máquina no conjunto de treino;

A sequência acima é realizada 50 vezes para cada modelo com a finalidade de minimizar os efeitos aleatórios das divisões da etapa 1. Para realizar essas simulações, utilizamos os métodos *LogisticRegression* e *SVC* do pacote SciKit-learn [9]. Os resultados dessas simulações estão sumarizados na Tabela 2.

Tabela 2: Valores médios das métricas de 50 simulações. Para o método *LogisticRegression* foi definido $C = 80$ para o parâmetro de regularização. Já para o *SVC*, consideramos: $C = 88$, $\gamma = 1$ e kernel RBF.

Métrica	RL Treino	RL Teste	SVM Treino	SVM Teste
Área Curva ROC	0,82	0,80	0,99	0,95
Acurácia	0,73	0,72	0,98	0,92
Recall	0,82	0,78	0,99	0,73
F1 Score	0,52	0,50	0,95	0,77
Precisão	0,38	0,37	0,91	0,82

Ao analisar os resultados, observa-se que o modelo SVM demonstrou consistentemente um desempenho superior em comparação com o modelo de Regressão Logística em todas as métricas avaliadas nos conjuntos de treinamento e teste. Especificamente, o SVM alcançou uma área sob a curva ROC (ROC AUC) de 0.99 no conjunto de treinamento e 0.95 no conjunto de teste, indicando uma excelente capacidade de discriminação entre as classes em ambas as situações. Além disso, o SVM exibiu uma alta acurácia, recall, F1-score e precisão tanto no conjunto de treinamento quanto no conjunto de teste, refletindo sua habilidade de classificação robusta e precisa.

Por outro lado, embora o modelo de Regressão Logística tenha obtido valores razoáveis em algumas métricas, como a área sob a curva ROC e a acurácia, sua performance foi inferior à do SVM, especialmente no que se refere ao recall, F1-score e precisão. Isso sugere que o SVM é mais eficaz na identificação correta de exemplos positivos, minimizando tanto os falsos negativos quanto os falsos positivos, em comparação com a Regressão Logística.

5 Considerações Finais

Neste artigo, apresentamos um modelo baseado na aplicação de métodos de ciência de dados para a previsão da ocorrência de ferrugem asiática da soja em uma determinada região. O modelo

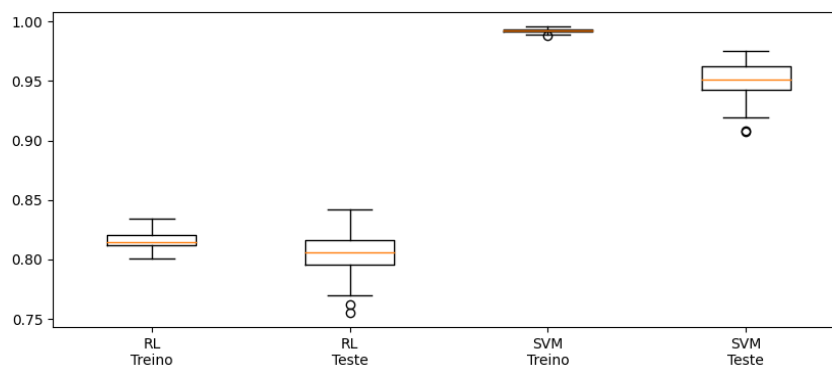


Figura 2: Boxplot mostrando o resultado de 50 simulações para cada método de predição: Regressão Logística e Support Vector Machines. Fonte: autor.

proposto utiliza como entrada o estado de contágio na área de interesse e em suas proximidades, bem como as condições climáticas em um momento específico. Como resultado, o modelo fornece uma estimativa da probabilidade de ocorrência da doença em um intervalo de tempo futuro. A relação entre as variáveis de entrada e saída são então determinadas por meio de aplicação de métodos de machine learning. Para aplicação do modelo, utilizamos dados de contágio do município de Campo Verde - MT, treinando os métodos de Regressão Logística e Support Vector Machines. Uma análise dos valores das métricas da aplicação realizada, Tabela 2 e Figura 2, nos mostra que a modelagem proposta apresenta resultados muito satisfatórios em prever a ocorrência da ferrugem asiática.

A robustez dos resultados indica que esta proposta de modelagem pode, em certa medida, representar uma alternativa viável aos modelos clássicos de dispersão espaço-temporal de doenças infectocontagiosas. A principal vantagem da abordagem proposta reside no fato de que muitos dos fatores determinantes para a propagação de doenças, frequentemente desafiadores de serem incorporados em modelagens baseadas em equações diferenciais, podem ser identificados por meio de técnicas de ciência de dados. No entanto, é importante observar que, além da falta de interpretabilidade, uma desvantagem evidente é a necessidade de volumes significativos de dados para garantir a eficácia desses modelos, volumes estes que frequentemente não estão disponíveis.

Agradecimentos

Os autores Moiseis Santos Ceconello e André Krindges agradecem a Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas pela parceria em projetos de pós-doutorado. A autora Nayara Longo Sartor Zagui agradeceu à Fundação de Amparo à Pesquisa de Mato Grosso por ter parte da pesquisa custeada pelo Edital FAPEMAT 014/2022.

Referências

- [1] C. V. Godoy et al. "Asian soybean rust in Brazil: past, present, and future". Em: **Pesquisa Agropecuária Brasileira** 51.5 (mai. de 2016), pp. 407–421. ISSN: 0100-204X. DOI: 10.1590/s0100-204x2016000500002.

- [2] C. V. Godoy et al. **Eficiência de fungicidas para o controle da ferrugem-asiática da soja, *Phakopsora pachyrhizi*, na safra 2022/2023: resultados sumarizados dos ensaios cooperativos**. Circular Técnica 195. Londrina, PR: Embrapa Soja, jun. de 2023.
- [3] T. Hastie, R. Tibshirani e J. Friedman. **The elements of statistical learning: data mining, inference and prediction**. 2^a ed. Springer, 2009. ISBN: 9787519296865.
- [4] INMET. **Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP)**. Online. Acessado em: 12/11/2023, <https://bdmep.inmet.gov.br>.
- [5] B. Junior et al. **Análise da produção, área cultivada, produtividade e estabilidade produtiva da soja nos estados brasileiros em 25 safras (1996/1997 a 2020/2021)**. Boletim de Pesquisa e Desenvolvimento 27. Londrina: Embrapa Soja, 2022, p. 20.
- [6] Laboratório de Epidemiologia de Plantas. **Consórcio Antiferrugem**. Online. Acessado em: 12/11/2023, <http://www.consorcioantiferrugem.net/>.
- [7] S. Li, W. F. Moore, B. L. Spinks, B. C. Wells, G. L. Sciumbato, S. J. Robinson e L. Libous-Bailey. “Occurrence of Asian Soybean Rust Caused by *Phakopsora pachyrhizi* in Mississippi”. Em: **Plant Health Progress** 8.1 (jan. de 2007). ISSN: 1535-1025. DOI: 10.1094/php-2007-0917-02-br.
- [8] X. Li, P. D. Esker, Z. Pan, A. P. Dias, L. Xue e X. B. Yang. “The Uniqueness of the Soybean Rust Pathosystem: An Improved Understanding of the Risk in Different Regions of the World”. Em: **Plant Disease** 94.7 (jul. de 2010), pp. 796–806. ISSN: 1943-7692. DOI: 10.1094/pdis-94-7-0796.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot e E. Duchesnay. “Scikit-learn: Machine Learning in Python”. Em: **Journal of Machine Learning Research** 12 (2011), pp. 2825–2830.
- [10] N. L. S. Zagui, A. Krindges, A. D. P. Lotufo e C. R. Minussi. “Spatio-Temporal Modeling and Simulation of Asian Soybean Rust Based on Fuzzy System”. Em: **Sensors** 22.2 (jan. de 2022), p. 668. ISSN: 1424-8220. DOI: 10.3390/s22020668.