

COMPARAÇÃO DE ALGORITMOS PARA CLASSIFICAÇÃO AUTOMÁTICA DE SEMENTES DE TRIGO

PIAZZA, C. G. ¹

¹ Universidade Federal Do Pampa (Unipampa) – Bagé – RS – Brasil –

giovanapiazza.aluno@unipampa.edu.br

RESUMO

A classificação de sementes é, na maioria das vezes, um processo baseado em trabalho humano. Este estudo apresenta uma abordagem de classificação automática de três variedades de sementes de trigo: Rosa, Kama e Canadense, utilizando técnicas de inteligência artificial. A base de dados consistiu em 210 amostras, cada uma descrita por 7 características, que foram pré-processadas e normalizadas para o intervalo de 0 a 1. As amostras foram divididas em 80% para treinamento e 20% para teste do modelo. Foram comparados dois algoritmos de classificação: árvores de decisão e redes neurais multicamadas (MLP). Com parâmetros padrão, a avaliação dos modelos nos dados de teste mostrou que as árvores de decisão alcançaram uma acurácia de 88%, enquanto as redes neurais obtiveram 95%. Os resultados indicam que as técnicas de inteligência artificial são promissoras para a classificação de sementes, com potencial para otimizar processos na agricultura.

Palavras-chave: Inteligência Artificial, Redes Neurais Artificiais, Agronegócio.

1 INTRODUÇÃO

A produtividade agrícola é fortemente influenciada pela qualidade das sementes, pois esta impacta diretamente o desempenho das colheitas. Nos últimos anos, técnicas de classificação automática, como árvores de decisão e redes neurais artificiais (RNAs), vêm ganhando destaque no setor agrícola. As RNAs, inspiradas no funcionamento do cérebro humano, são modelos computacionais com capacidade de aprender com dados e realizar previsões complexas. Uma das principais vantagens das RNAs é a habilidade de detectar padrões não lineares em grandes volumes de dados, o que permite análises mais profundas e precisas.

O objetivo deste estudo é comparar a performance dos algoritmos de árvores de decisão e redes neurais artificiais (RNAs) na tarefa de classificação de sementes de trigo. Trabalhos anteriores, como o de Sousa et al. (2020) no artigo "Python e predição de dados usando redes neurais multicamadas", demonstram o uso de

MLPs para predição de dados, ressaltando o potencial das RNAs para aprendizado supervisionado em problemas complexos. Da mesma forma, Rauber (2005) apresenta um trabalho fundamental sobre RNAs, abordando seu funcionamento, métodos de aprendizado e diversas aplicações, destacando as redes neurais como uma ferramenta essencial na inteligência artificial, com um grande potencial em diferentes campos.

2 METODOLOGIA (MATERIAL E MÉTODOS)

O estudo utilizou o conjunto de dados "Seeds" do UCI Machine Learning Repository, contendo 210 amostras de sementes das variedades de trigo Kama, Rosa e Canadense, descritas por sete atributos geométricos, incluindo área, perímetro e coeficiente de compactação. Todas as variáveis foram normalizadas para o intervalo de 0 a 1.

O desenvolvimento do modelo de rede neural multicamadas (MLP) foi realizado no Google Colab, com suporte de bibliotecas de aprendizado de máquina em Python. As amostras foram divididas aleatoriamente, com 80% dos dados destinados ao treinamento e 20% para testes. A análise preliminar incluiu métodos estatísticos descritivos e testes de variância para identificar diferenças entre as variedades de trigo. Em seguida, os modelos de árvore de decisão e MLP foram treinados e avaliados.

3 RESULTADOS E DISCUSSÃO

Os resultados evidenciaram que o modelo de rede neural multicamadas (MLP) apresentou uma acurácia de 95% nos dados de teste, enquanto as árvores de decisão alcançaram 88%. Esse desempenho reflete a capacidade das RNAs em capturar padrões complexos, sendo mais eficazes que as árvores de decisão, especialmente em dados de alta dimensionalidade.

As árvores de decisão, conforme ilustrado na Figura 1 retirada do artigo de Piazza e Camargo (2024), são modelos lineares que funcionam bem para problemas mais simples e oferecem uma interpretação clara e intuitiva. Sua estrutura, baseada em divisões sucessivas, facilita a compreensão das regras geradas pelo modelo, o que as torna adequadas para a análise de problemas em que a explicabilidade é importante. No entanto, quando confrontadas com dados mais complexos, seu desempenho tende a ser limitado.

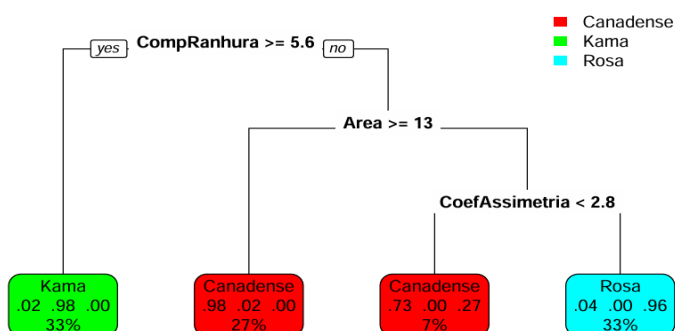


Figura 1: Modelo de árvores de decisão criado para classificar as variedades.

Fonte: Piazza e Camargo (2024).

As redes neurais artificiais (RNAs) são modelos não lineares muito mais poderosos, capazes de lidar com padrões complexos e dados de alta dimensionalidade. Sua flexibilidade permite que capturem relações mais sutis que os métodos lineares, como as árvores, não conseguem identificar. A implementação de um modelo de rede neural multicamadas (MLP) no estudo demonstrou resultados promissores na classificação das três variedades de sementes de trigo: Rosa, Kama e Canadense. Com a normalização e o embaralhamento dos dados, os conjuntos de treinamento e teste foram separados corretamente, garantindo uma avaliação justa e imparcial do modelo.

A acurácia de treinamento ao longo das 5000 épocas é apresentada, revelando um aumento constante na precisão do modelo durante o aprendizado. Apesar de pequenas oscilações, o gráfico confirma que o modelo foi capaz de melhorar gradualmente sua capacidade de classificação à medida que o treinamento avançava.

Na Figura 2 demonstra a matriz de confusão gerada para o conjunto de teste das árvores de decisão. Ela fornece uma visão abrangente do desempenho do modelo de classificação, permitindo identificar onde ele acerta e onde comete erros.

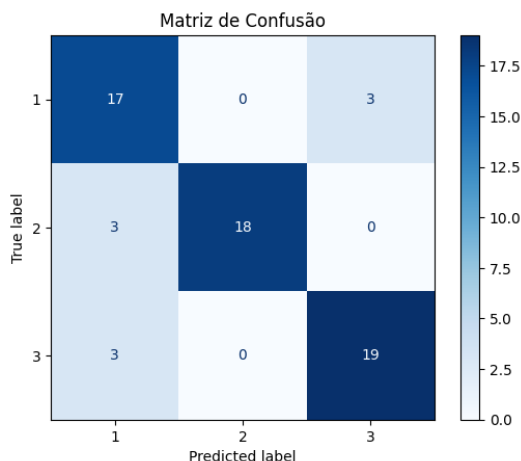


Figura 2: Matriz de Confusão da Arvore De Decisão.

Fonte: próprio autor.

A Figura 3 mostra a matriz de confusão gerada para o conjunto de teste das redes neurais artificiais, onde é possível observar o desempenho do modelo ao comparar os rótulos previstos com os rótulos reais.

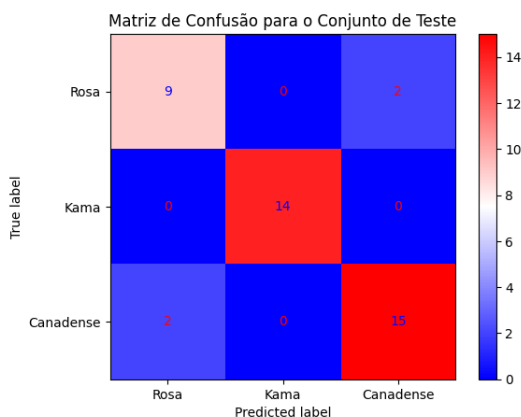


Figura 3: Matriz de Confusão para o Conjunto de Teste Das RNAs.

Fonte: próprio autor.

A matriz de confusão proporciona uma visão detalhada da eficácia do modelo, evidenciando uma taxa de acerto superior ao empregar redes neurais artificiais em comparação com árvores de decisão, mesmo diante de pequenos erros de classificação entre classes cujas características são mais semelhantes.

4 CONCLUSÕES

A comparação entre árvores de decisão e redes neurais artificiais (RNAs) no processo de classificação de sementes de trigo mostrou que as RNAs apresentam um desempenho superior, alcançando uma acurácia acima de 95%. Esse resultado demonstra o potencial das RNAs para resolver problemas de classificação complexos na agricultura, destacando sua capacidade de lidar com padrões

complexos e alta dimensionalidade dos dados. A análise das matrizes de confusão indicou que, embora existam erros em classes com características similares, as RNAs se mostraram mais eficazes do que as árvores de decisão. Esses achados sugerem que a utilização de RNAs pode otimizar processos como a seleção de sementes, podendo ser ainda aprimorada com ajustes e otimizações futuras para aumentar a precisão na distinção entre classes complexas.

REFERÊNCIAS

DE SOUSA, Jonathan Romeiro et al. Python e predição de dados usando redes neurais multicamadas. Brazilian Journal of Development, v. 6, n. 7, p. 54181-54185, 2020.

CHARYTANOWICZ, M, NIEWCZAS, J, KULCZYCKI, P; KOWALSKI, P; LUKASIK, S. (2012). Seeds. UCI Machine Learning Repository. <https://doi.org/10.24432/C5H30K>.

RAUBER, Thomas Walter. Redes neurais artificiais. Universidade Federal do Espírito Santo, v. 29, 2005.

PIAZZA, G. C.; CAMARGO, S. S.. Usando aprendizado de máquina para classificação automática de sementes de trigo. In: Salão Internacional de Ensino, Pesquisa e Extensão, 2024, São Borja. Anais do 16º SIEPE, 2024. v. 1.