
**APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NA
CLASSIFICAÇÃO DE TIPOS DE SOLO COM BASE EM SEUS
PERFIS DE NUTRIENTES E ANÁLISES FITOPATOLÓGICAS**

**APPLICATION OF MACHINE LEARNING TECHNIQUES IN
CLASSIFICATION OF SOIL TYPES BASED ON THEIR
NUTRIENT PROFILES AND PHYTOPATHOLOGICAL ANALYSIS**

Eron Ponce Pereira¹
Sergio Kenji Sawasaki Tanaka²
Bruno Henrique Coletto³
Sergio Akio Tanaka⁴

RESUMO

A agricultura de precisão é uma área que cresce exponencialmente, visto que este setor é o maior contribuinte para o PIB brasileiro. O problema a ser resolvido é a morosidade na identificação de padrões em análise de solo, sendo necessário uma demanda de diversos atores no processo da agricultura de precisão. Este trabalho tem uma aplicação prática utilizando a Inteligência Artificial com foco em aprendizagem de máquina com a visão computacional e análise de dados. Um estudo de caso prático, na área do agronegócio, está sendo implementado para resolver problemas na classificação de tipos de solos, com base em seus perfis de nutrientes e a identificação de fatores fitopatológicos em plantas. Foi elaborado um *workflow* para o entendimento do processo da análise de solo. Os resultados iniciais gerados com a implementação do projeto são promissores. Ao utilizar uma base de dados de plantio de solo contendo 2200 linhas foram gerados oito modelos de aprendizado supervisionado, variando em uma confiança de resultados de 93% a 99%. Portanto, verificou-se que é possível identificar a melhor cultura a ser plantada com os dados de uma análise de solo existente.

Palavras-chave: inteligência artificial; aprendizado de máquina; agricultura de precisão; gestão de solo.

ABSTRACT

Precision agriculture is an area that is growing exponentially, as this sector is the largest contributor to Brazilian GDP. The problem to be resolved is the slowness in identifying patterns in soil analysis, requiring demand from different actors in the precision agriculture process.

¹ Graduando do Curso de Ciência da Computação do Centro Universitário Filadélfia - UniFil. eronponcepereira@edu.unifil.br

² Graduando do Curso de Ciência da Computação do Centro Universitário Filadélfia - UniFil. serginho.k.s.tanaka@edu.unifil.br

³ Co-orientador: Professor do Centro Universitário Filadélfia - UniFil. bruno.coletto@unifil.br

⁴ Orientador: Professor e Orientador do Projeto do Centro Universitário Filadélfia - UniFil. sergio.tanaka@unifil.br

This work has a practical application using Artificial Intelligence with a focus on machine learning with computer vision and data analysis. A practical case study, in the area of agribusiness, is being implemented to solve problems in the classification of soil types, based on their nutrient profiles and the identification of phytopathological factors in plants. A workflow was developed to understand the soil analysis process. The initial results generated with the implementation of the project are promising. When using a soil planning database containing 2,200 lines, eight supervised learning models were generated, varying in confidence in results from 93% to 99%. Therefore, it is possible to identify the best crop to be planted with data from an existing soil analysis.

Keywords: artificial intelligence; machine learning; computer vision; precision agriculture; soil management.

1 INTRODUÇÃO

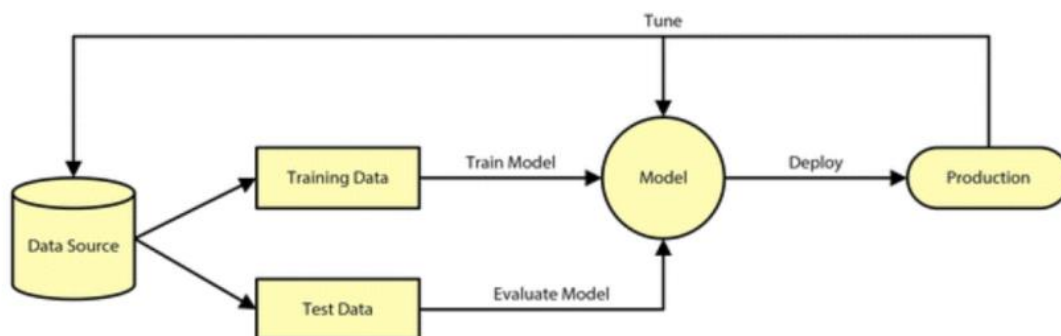
A função da Inteligência Artificial (IA) é concentrar-se na capacidade de receber informações do ambiente e tomar decisões com base nelas. Esses agentes utilizam uma variedade de técnicas e abordagens para interpretar e tomar decisões de acordo com o ambiente, sendo eles, o aprendizado de máquina (*machine learning*), tanto supervisionado ou não supervisionado (Dwivedi *et al.*, 2021).

O aprendizado de máquina é uma técnica que permite que os agentes aprendam a partir dos dados e melhorem seu desempenho ao longo do tempo. Ele envolve o uso de algoritmos para identificar padrões nos dados e fazer previsões ou tomar decisões com base nessas descobertas (Dzulkalnine *et al.*, 2023).

O aprendizado de máquina supervisionado é o mais utilizado e comum na literatura e acontece quando o sistema recebe uma informação como entrada e produz uma saída na forma de pontuações para cada categoria reconhecida. O objetivo é que a categoria desejada tenha a pontuação mais alta entre todas as categorias. O sistema ajusta seus parâmetros internos, chamados de pesos, com base entre as pontuações de saída e o padrão desejado. Os dados de treinamento são rotulados e o agente é ensinado a fazer previsões com base nesses rótulos (HUA *et al.*, 2023).

A Figura 1 apresenta um *workflow* de um algoritmo de aprendizado supervisionado, podendo verificar que o algoritmo consiste em treinar o modelo com as determinadas categorias encontradas, testá-lo e criar um modelo a ser utilizado, podendo ser refinado posteriormente com outros dados.

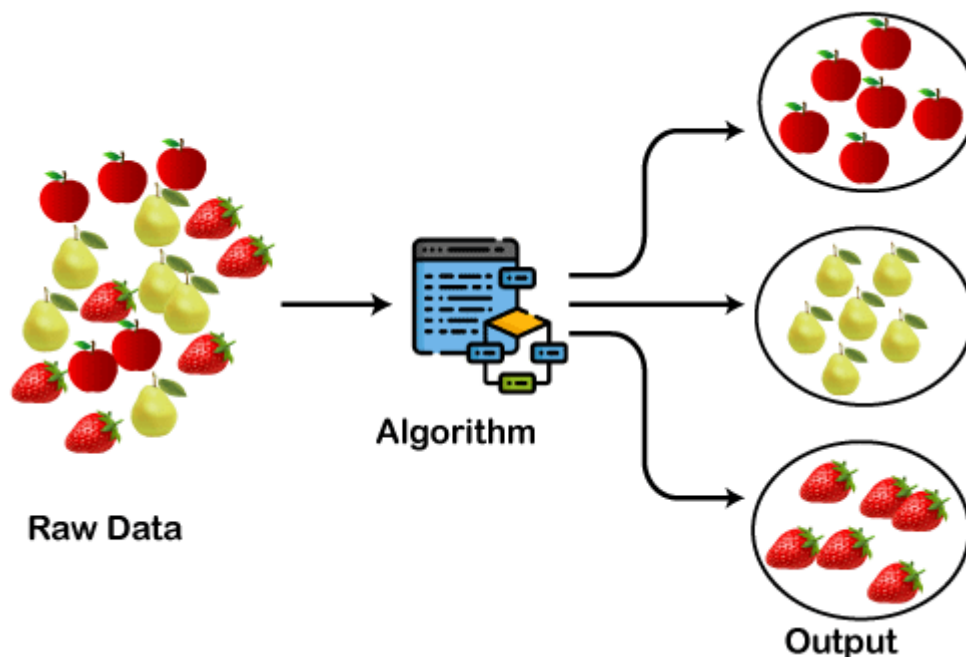
Figura 1 - Aprendizado supervisionado



Fonte - Mahesh (2018)

A Figura 2 apresenta o aprendizado de máquina não supervisionado e envolve o treinamento de sistemas em conjuntos de dados não rotulados, ou seja, dados que não possuem categorias conhecidas. Nesse tipo de aprendizado, o objetivo é encontrar estruturas, padrões ou grupos nos dados sem a orientação de rótulos predefinidos (HUA *et al.* 2023).

Figura 2 – Aprendizado não supervisionado



Fonte: Mahesh (2018)

A utilização de *machine learning* em diferentes tipos de aprendizagem é possibilitado dado a forma de dados padrão, sendo os mais básicos sendo numéricos, porém com o avanço da tecnologia e a agricultura de precisão, foi necessário fazer análises visuais de diversos setores do campo, assim dando uma ênfase maior para o estudo da visão computacional (Pandit et al., 2023).

O ramo da visão computacional, é algo que tem suas raízes em 1960 com muitas documentações de estudos realizados, como reconhecimento de caracteres, diagnóstico médico, automação (visão robótica), inteligência militar, compressão de dados e reconhecimento de fala. Tais pesquisas são apontadas e sintetizadas por Fu e Rosenfeld (1976). Através destas tecnologias também foram criados modelos de identificação de padrões das imagens geradas.

O modelo de dados chamado “*Sparse signal representation*” é adotado fortemente na literatura como exemplifica Yang e Liu (2019), com tal modelo é possível classificar imagens e modificá-las conforme o objetivo, o autor exemplifica utilizações como “*face recognition, Image super-resolution, motion and data segmentation, supervised denoising and inpainting, background modeling and image classification*”.

É possível utilizar *machine learning* em diversos domínios por meio da visão computacional, porém, não se resume somente a este tipo de dado. O aprendizado de máquina pode ser utilizado com dados sobre o solo, tais como, a temperatura do solo, precipitação acumulada entre outras, com o objetivo de fazer previsões sobre a qualidade do mesmo (Schwalbert et al., 2020).

Uma das utilizações do *machine learning* para a previsão de qualidade de solo é visto em Wadoux, Minasny e Mcbratney (2020), os autores exemplificam os dados a serem utilizados para o aprendizado da máquina, entre eles são, imagens de satélite, dados de sensores remotos, informações topográficas, dados climáticos, entre outros especificados no próprio artigo. Os autores apresentam também uma lista de autores, com base nas variáveis utilizadas e diferentes tipos de algoritmos, a fim de quantificar a qualidade do solo.

Outra utilização do *machine learning* na área agrícola, é a realização da previsão de rendimento de grãos com o auxílio de satélites, mostrado por Pinto et al. (2022). Com o foco em milho (grão produzido por 29% da área cultivável do Brasil), os modelos de aprendizados de máquina abordados são as *artificial neural networks (ANN)*, *k-nearest neighbors (KNN)*, *Random forest (RF)* e *Support vector machines (SVM)*.

O objetivo geral é desenvolver um projeto integrado na linguagem Python, que utilize algoritmos de aprendizado supervisionado, para classificar o melhor plantio de acordo com o solo, tendo como base seus nutrientes. Utilizando também técnicas de visão computacional para identificar doenças em plantas. Através da análise de dados do solo, busca-se otimizar a utilização de recursos e maximizar o rendimento das colheitas. Simultaneamente, o uso de visão computacional permite diagnósticos reduzindo a dependência de métodos manuais, permitindo intervenções mais eficazes. Este projeto integrado contribui para uma agricultura mais

sustentável e produtiva.

2 METODOLOGIA

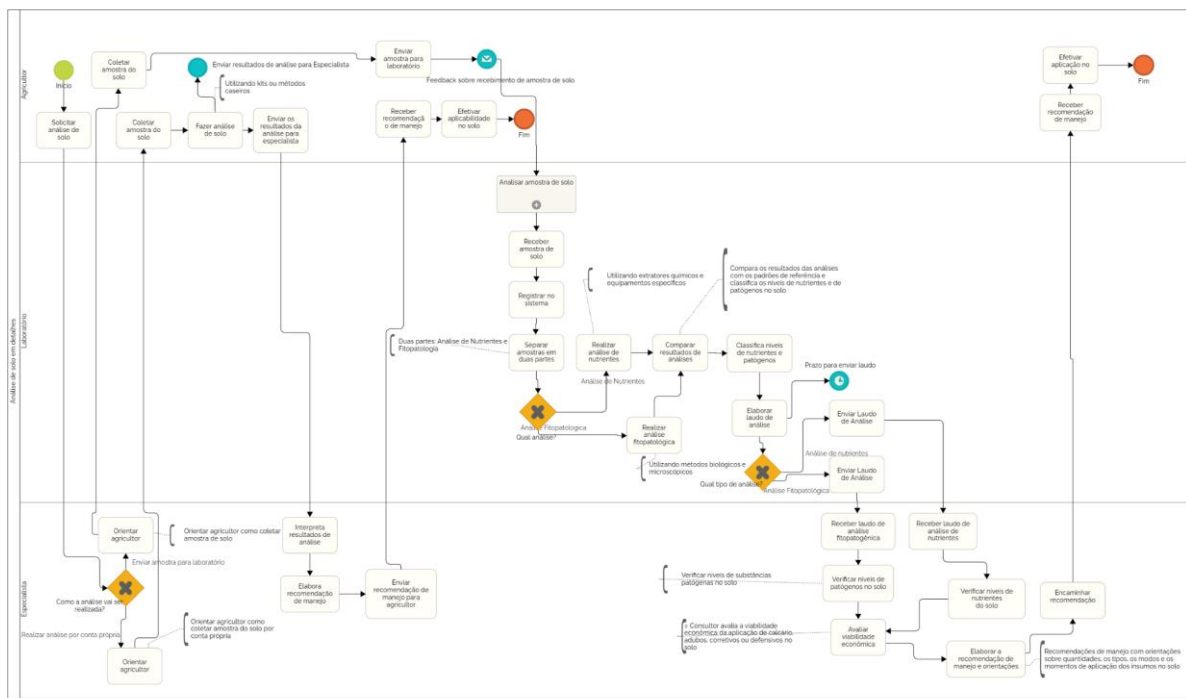
A Figura 3 apresenta um *workflow* utilizando o modelo *business process management & notation (BPMN)*, utilizado para documentar o processo da análise de solo, desde a coleta de amostras, e seus diversos fluxos de análise e síntese de resultados. O objetivo do estudo de caso é desenvolver uma solução para minimizar o processo árduo e custoso da análise de solo.

Para a coleta dos dados de solo, três principais atores participam do processo, o agricultor (responsável pela solicitação da análise), o especialista (responsável por orientar o agricultor de como realizar a extração dos dados e o laboratório (local onde as análises são feitas).

As análises de nutrientes são utilizadas para a verificar o manejo de plantio, podendo ser utilizado uma Inteligência Artificial para catalogação automatizada. A análise fitopatológica é usada para verificação dos patógenos no solo com o objetivo de identificar manualmente um ator externo, por meio da visão computacional e inteligência artificial é possível fazer o reconhecimento deste processo automaticamente.

5

Figura 3 - Workflow detalhado análise de solo



3 ESTUDO DE CASO

O estudo de caso permite a realização de análises que forneçam informações para correção ou manutenção do solo. Isso pode abrir potenciais oportunidades para a venda de novos produtos, como defensivos agrícolas, por exemplo.

A exploração de análises mais aprofundadas dos nutrientes presentes no solo, busca identificar com precisão a presença de determinados nutrientes, como o fósforo, potássio, nitrogênio, entre outros. Isso pode contribuir para estratégias de manejo mais eficientes e produtivas.

Uma das utilizações é a verificação da melhor cultura a ser plantada com determinado padrão de nutrientes de um solo, sendo menos custoso ao agricultor para direcionar qual cultura combina mais com o respectivo solo, e quais modificações seriam necessárias para o regulamento do mesmo.

A realização de análises detalhadas do solo, inclui, além de nutrientes, os fungos e bactérias presentes. Esse tipo de análise permite um melhor entendimento das condições do solo e contribui para realizar estratégias de manejo direcionadas ao controle de doenças e otimização do crescimento das plantas.

A identificação precoce de problemas pode significar uma redução significativa nos custos de tratamento e pode também contribuir para estratégias mais eficazes de controle de pragas. A Figura 3 apresenta os passos para a análise fitopatológica. Ao entender os tipos de feridas e os sintomas associados, os agricultores podem agir rapidamente para mitigar os problemas, melhorando assim a saúde geral das plantas e, conseqüentemente, a produtividade agrícola.

4 DESENVOLVIMENTO E RESULTADOS

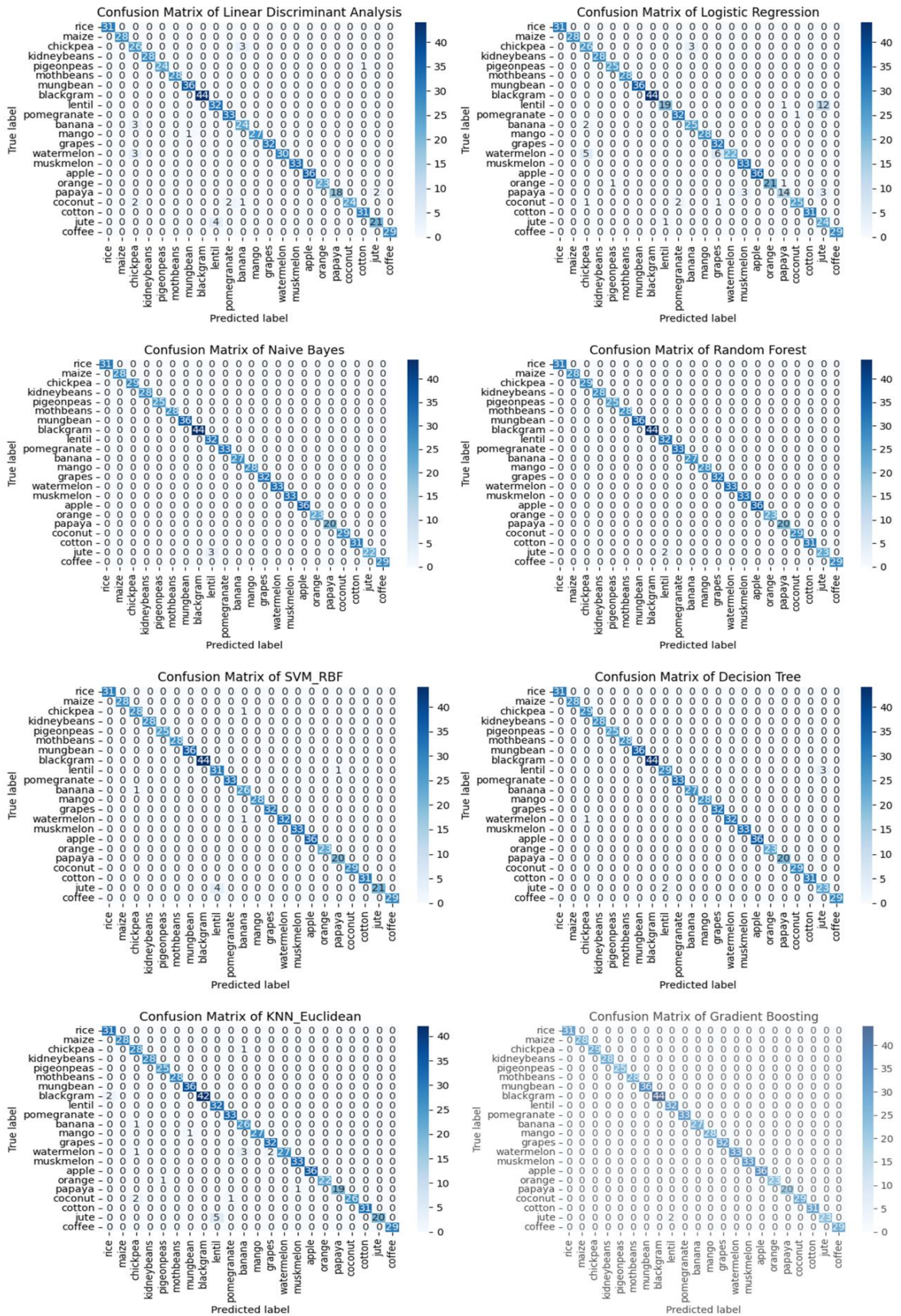
O projeto desenvolvido é um sistema de recomendação de cultivos implementado com o auxílio da biblioteca em *python Django* que permite aos usuários inserir parâmetros como temperatura, umidade e nutrientes do solo para prever o cultivo mais adequado. O sistema utiliza diversos algoritmos de aprendizado de máquina, como *Logistic Regression*, *Random Forests*, *Support Vector Machines with RBF kernel*, *Decision Trees*, *Gradient Boosting*, *K-NN with Euclidean metric*, *Gaussian Naive Bayes* e *Linear Discriminant Analysis* para fazer as previsões (Pedregosa *et al.*, 2011).

O sistema também oferece funcionalidades para salvar instâncias de plantio com seus parâmetros e previsões, permitindo que os usuários carreguem essas informações posteriormente. Além disso, o código inclui uma função para treinar os modelos de aprendizado de máquina e salvar suas acurácias.

O algoritmo foi desenvolvido contendo diversas bibliotecas de aprendizado supervisionado, utilizando como uma base de dados de plantio fornecido pela base de dados *Kaggle* pelo usuário TheEyesChico (2023). A Figura 6 apresenta as matrizes de confusão, como exemplifica Görtler *et al.* (2020) “A matriz de confusão, é uma visualização onipresente para ajudar as pessoas a avaliar modelos de aprendizado de máquina, é um layout tabular que compara rótulos de classes previstos com rótulos de classes reais em todas as instâncias de dados.” Os dados previstos são apresentados pelo eixo *Y*, e os dados corretos na base de dados o eixo *X*, essa relação representa as acurácias dos modelos treinados.

Dentre os modelos apresentados na Figura 6, o *Logistic Regression* obteve uma acurácia de 93%. O *KNeighbors Classifier* e o *Linear Discriminant Analysis* ambos alcançaram uma acurácia de 97%. O *Decision Tree Classifier* e o *SVC* alcançaram uma acurácia de 99% e os modelos *Gradient Boosting Classifier*, *Gaussian Naive Bayes* e *Random Forest Classifier* atingiram a acurácia máxima de aproximadamente 100%. Tais resultados são satisfatórios para os modelos utilizados e apontam uma ótima oportunidade para aplicação em um teste real em um solo.

Figura 6 - Matrizes de confusão dos resultados



A Figura 7 apresenta uma das interfaces gráficas do projeto que foi desenvolvido. O sistema de recomendação de solo foi implementado e está em fase de teste. Ao inserir informações sobre o solo e as condições climáticas, como níveis de nitrogênio, fósforo, potássio, temperatura e umidade, é possível que o usuário faça a consulta dos dados apresentando a cultura prevista, a confiança do modelo em sua escolha, sua acurácia, fornecida pela matriz de confusão, e seu *z-score* dado que representa qual a discrepância de valor específico em relação à média da amostras coletadas.

Figura 7 - Interface gráfica do sistema de recomendação de plantio

Nome da Instância: Autor:

Valores de Entrada Adicionar Linha

Salvar Parâmetros

N	P	K	Temperatura C°	Umidade	pH	Chuva	Ações
<input type="text" value="20,0"/>	<input type="text" value="27,0"/>	<input type="text" value="29,0"/>	<input type="text" value="31,0"/>	<input type="text" value="50,0"/>	<input type="text" value="5,7"/>	<input type="text" value="94,0"/>	Remover

Executar Algoritmo

Informação de predição

Nome do modelo	Predição	Confiança máxima	Acuracia
DecisionTreeClassifier	mango	1.0	0.99
GradientBoostingClassifier	mango	1.0	1.0
KNeighborsClassifier	mango	1.0	0.97
LinearDiscriminantAnalysis	mango	1.0	0.97
LogisticRegression	mango	0.5	0.93
GaussianNB	mango	1.0	1.0
RandomForestClassifier	mango	1.0	1.0
SVC	mango	0.95	0.99

Parâmetro	N	P	K	Temperatura C°	Umidade	pH	Chuva
Z-score	0.0	0.0	0.0	0.0	0.0	0.1	0.0

6 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho catalogou duas áreas de pesquisa, sendo elas, a visão computacional, sistema que permite o diagnóstico automático de doenças em plantas, e o sistema de recomendação de solo. Acelerando processos que atualmente exigem inspeção manual, permitindo intervenções mais rápidas e eficazes no tratamento de doenças das plantas, economizando tempo e recursos.

A análise dos modelos de aprendizado de máquina mostrou um alto grau de acurácia,

variando entre 93% e 100%, esses resultados indicam um desempenho robusto e confiável dessas abordagens em tarefas de classificação, apontando para seu potencial significativo em aplicações futuras.

Portanto, o sistema implementado é satisfatório, com a possibilidade de um escalonamento para outras áreas abordadas no *workflow*, como por exemplo, a análise fitopatológica apresentada neste trabalho. Exemplificando como a aplicação dessas tecnologias são de suma importância na agricultura e como o aprendizado de máquina pode ser usado para contribuir para práticas agrícolas mais sustentáveis e eficientes.

REFERÊNCIAS

DWIVEDI, Yogesh K. et al. Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy.

International Journal Of Information Management, v. 57, p. 101994, abr. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.ijinfomgt.2019.08.002>.

DZULKALNINE, Mohamad *et al.* **Optimizing Ethanol Production in Escherichia Coli Using a Hybrid of Particle Swarm Optimization and Artificial Bee Colony**. 2022 The 6Th International Conference On Advances In Artificial Intelligence, [S.L.], p. 1-7, 21 out. 2022. ACM. <http://dx.doi.org/10.1145/3571560.3571581>.

FU, King-Sun; ROSENFELD. Pattern Recognition and Image Processing. **Ieee Transactions On Computers**, v. -25, n. 12, p. 1336-1346, dez. 1976. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tc.1976.1674602>.

GÖRTLER, Jochen *et al.* **generalizing confusion matrix visualization to hierarchical and multi-output labels**. Chi Conference On Human Factors In Computing Systems, 29 abr. 2022. ACM. <http://dx.doi.org/10.1145/3491102.3501823>.

MAHESH, Batta. **Machine learning algorithms - a review**. International Journal of Science and Research (IJSR), v. 9, p. 381-386, 2020. ISSN 2319-7064.

PANDIT, Deepanjali et al. **Computer Vision-based Automated Cashew Kernel Grading**. Proceedings Of The 2023 5Th International Conference On Image, Video And Signal Processing, [S.L.], p. 84-98, 24, mar. 2023. ACM. <http://dx.doi.org/10.1145/3591156.3591169>.

PEDREGOSA, Fabian. *et al.* **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011.

PINTO, Antonio Alves *et al.* Corn grain yield forecasting by satellite remote sensing and machine-learning models. **Agronomy Journal**, v. 114, n. 5, p. 2956-2968, ago. 2022. Wiley. <http://dx.doi.org/10.1002/agj2.21141>.

SCHWALBERT, Raí A. *et al.* Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern brazil.

Agricultural And Forest Meteorology, v. 284, p. 107886, abr. 2020. Elsevier BV.
<http://dx.doi.org/10.1016/j.agrformet.2019.107886>.

SURI, Jasjit s. Computer Vision, Pattern Recognition and Image Processing in Left Ventricle Segmentation: the last 50 years. *Pattern Analysis & Applications*, v. 3, n. 3, p. 209-242, 28 set. 2000. Springer Science and Business Media LLC.
<http://dx.doi.org/10.1007/s100440070008>.

TheEyesChico. **Crop Analysis and Prediction**. Disponível em:
<https://www.kaggle.com/code/theeyeschico/crop-analysis-and-prediction>. Acesso em:
13/09/2023

WADOUX, Alexandre M.J.-C.; MINASNY, Budiman; MCBRATNEY, Alex B.. Machine learning for digital soil mapping: applications, challenges and suggested solutions. **Earth-Science Reviews**, v. 210, p. 103359, nov. 2020. Elsevier BV.
<http://dx.doi.org/10.1016/j.earscirev.2020.103359>.