v7.n7.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

DESENVOLVIMENTO DE APLICATIVO COM APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DO IMPACTO DE RIZOBACTÉRIAS NO CRESCIMENTO DE PLANTAÇÕES DE ARROZ

DEVELOPMENT OF A MACHINE LEARNING APPLICATION FOR PREDICTING THE IMPACT OF RHIZOBACTERIA ON RICE CROP GROWTH

DESARROLLO DE UNA APLICACIÓN CON APRENDIZAJE AUTOMÁTICO PARA PREDECIR EL IMPACTO DE LAS RIZOBACTERIAS EN EL CRECIMIENTO DE CULTIVOS DE ARROZ

João Paulo dos Santos Brito[1], Pedro Leonan Gonzaga de Freitas[1], Antônia Wiviane de Oliveira Mendes[2], Alessandra Jackeline Guedes de Moraes[3], Marta de Oliveira Barreiros[4]

e778407
https://doi.org/10.47820/recima21.v7i7.8407

PUBLICADO: 07/2026

RESUMO

O uso de rizobactérias promotoras de crescimento de plantas (RPCPs) apresenta-se como alternativa sustentável para a agricultura, porém a predição de seus efeitos envolve múltiplas variáveis. Este trabalho teve como objetivo desenvolver um aplicativo móvel apoiado por aprendizado de máquina para análise preditiva do impacto de RPCPs no crescimento de arroz. A metodologia abrangeu quatro fases: levantamento de requisitos com especialista, análise exploratória de uma base de dados com 6.038 registros experimentais, desenvolvimento e avaliação de modelos de classificação e implementação do sistema. Foram comparados os algoritmos *KNN*, *Random Forest* e *XGBoost*, sendo este último selecionado por apresentar maior acurácia (0,945) e menor desvio padrão (0,010) na validação cruzada. A arquitetura Cliente-Servidor integrou um aplicativo *Android* em *Kotlin* com *Jetpack Compose* a uma *API RESTful* em *FastAPI*, operando em duas modalidades: não destrutiva, baseada em medições de campo, e destrutiva, com métricas de biomassa seca. Os resultados indicam que a ferramenta pode auxiliar a tomada de decisão ao reduzir a necessidade de coletas destrutivas em determinadas situações, contribuindo para práticas agrícolas mais sustentáveis.

PALAVRAS-CHAVE: Aprendizado de máquina. Rizobactérias. Cultivo de arroz.

ABSTRACT

The use of plant growth-promoting rhizobacteria (PGPR) presents itself as a sustainable alternative for agriculture; however, predicting their effects involves multiple variables. This study aimed to develop a mobile application supported by machine learning for the predictive analysis of the impact of PGPR on rice growth. The methodology comprised four phases: requirements gathering with a specialist, exploratory analysis of a dataset with 6,038 experimental records, development and evaluation of classification models, and system implementation.

[1] Graduando em Engenharia de Software pela Universidade do Estado do Pará – Campus XX Castanhal.
[2] Graduanda em Engenharia Florestal pela Universidade do Estado do Pará – Campus XX Castanhal.
[3] Doutora em Fitopatologia pela Universidade Federal Rural do Pernambuco – UFPE, docente da Universidade do Estado do Pará – Campus XX Castanhal.
[4] Doutora em Engenharia Elétrica pela Universidade Federal do Maranhão – UFMA, docente na Universidade do Estado do Pará – Campus XX Castanhal.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

The KNN, Random Forest, and XGBoost algorithms were compared, with the latter selected for presenting the highest accuracy (0.945) and the lowest standard deviation (0.010) in cross-validation. The Client-Server architecture integrated an Android application developed in Kotlin with Jetpack Compose to a RESTful API implemented in FastAPI, operating in two modes: non-destructive, based on field measurements, and destructive, using dry biomass metrics. The results indicate that the tool can support decision-making by reducing the need for destructive sampling in certain situations, contributing to more sustainable agricultural practices.

KEYWORDS: *Machine Learning. Rhizobacteria. Rice Cultivation.*

RESUMEN

El uso de rizobacterias promotoras del crecimiento de plantas (RPCP) se presenta como una alternativa sostenible para la agricultura; sin embargo, la predicción de sus efectos involucra múltiples variables. Este trabajo tuvo como objetivo desarrollar una aplicación móvil apoyada por aprendizaje automático para el análisis predictivo del impacto de las RPCP en el crecimiento del arroz. La metodología abarcó cuatro fases: levantamiento de requisitos con un especialista, análisis exploratorio de una base de datos con 6.038 registros experimentales, desarrollo y evaluación de modelos de clasificación e implementación del sistema. Se compararon los algoritmos KNN, Random Forest y XGBoost, siendo este último seleccionado por presentar la mayor exactitud (0,945) y la menor desviación estándar (0,010) en la validación cruzada. La arquitectura Cliente-Servidor integró una aplicación Android desarrollada en Kotlin con Jetpack Compose a una API RESTful implementada en FastAPI, operando en dos modalidades: no destructiva, basada en mediciones de campo, y destructiva, con métricas de biomasa seca. Los resultados indican que la herramienta puede apoyar la toma de decisiones al reducir la necesidad de muestreos destructivos en determinadas situaciones, contribuyendo a prácticas agrícolas más sostenibles.

PALABRAS CLAVE*: Aprendizaje automático. Rizobacterias. Cultivo de arroz.*

1. INTRODUÇÃO

A agricultura global enfrenta o desafio contínuo de aumentar a produtividade e contribuir para a segurança alimentar, especialmente em culturas de base como o arroz (*Oryza sativa* L.), o trigo e o milho, reconhecidas como importantes para a alimentação mundial (Farooq *et al.*, 2023). No caso do arroz, esse desafio torna-se ainda mais relevante diante de fatores ambientais que podem afetar sua produção, como a escassez hídrica associada às mudanças climáticas (Farooq *et al.*, 2023). Ao mesmo tempo, a busca por maior produtividade precisa considerar os impactos associados ao uso contínuo ou excessivo de fertilizantes químicos, uma vez que tais práticas podem alterar propriedades físico-químicas e biológicas do solo, afetar sua fertilidade e comprometer comunidades microbianas benéficas (Pahalvi *et al.,* 2021).

Nesse contexto, a utilização de microrganismos do solo, particularmente as rizobactérias promotoras de crescimento de plantas (RPCPs), apresenta-se como uma alternativa relevante para o desenvolvimento de práticas agrícolas mais sustentáveis.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Essas rizobactérias podem promover o crescimento vegetal por mecanismos diretos e indiretos, incluindo o aumento da disponibilidade de nutrientes, a produção de fitormônios, o desenvolvimento da parte aérea e das raízes, a proteção contra fitopatógenos e a redução de doenças (Andrade *et al.*, 2023). Além disso, o estudo de Andrade *et al.* (2023) também indica que as RPCPs podem contribuir para a redução do uso de fertilizantes e pesticidas sintéticos, favorecendo a saúde do solo e a sustentabilidade da produção agrícola.

No cultivo de arroz, as RPCPs têm sido investigadas por sua relação com a tolerância a estresses abióticos, como seca, salinidade, toxicidade por metais pesados e altas temperaturas. Esses fatores podem afetar negativamente a produtividade e a qualidade da cultura, tornando necessário o estudo de alternativas que auxiliem o desenvolvimento das plantas em condições adversas (Sonowal *et al.*, 2024). As RPCPs podem atuar por meio de mecanismos como absorção de nutrientes, síntese de fitormônios, modificação de vias de sinalização relacionadas ao estresse e estímulo a genes responsivos, contribuindo para a adaptação das plantas a diferentes condições ambientais (Sonowal *et al.*, 2024).

No entanto, a eficiência das RPCPs como bioinoculantes depende, entre outros fatores, da capacidade de colonizar a rizosfera, competir com outros microrganismos e sobreviver sob diferentes condições ambientais (Santoyo *et al.*, 2021). No caso do arroz, a literatura também aponta dificuldades relacionadas à variabilidade entre cepas, às diferenças de solo, clima, pH, disponibilidade de nutrientes, práticas agronômicas, ausência de protocolos padronizados e desafios de aplicação em escala de campo (Sonowal *et al.*, 2024). Diante desse cenário de múltiplas variáveis biológicas e ambientais, este trabalho propõe o uso de técnicas de aprendizado de máquina como ferramenta de apoio à organização e à análise de dados experimentais relacionados ao efeito de rizobactérias no desenvolvimento de plantas de arroz.

Técnicas de aprendizado de máquina têm sido exploradas em estudos agrícolas como apoio à análise de dados e à construção de modelos preditivos. Botero-Valencia *et al.* (2025) indicam que essas técnicas podem auxiliar na integração de múltiplas fontes de dados, na análise de padrões complexos, no monitoramento da saúde do solo e na tomada de decisão em sistemas agrícolas. Entretanto, a adoção dessas abordagens também apresenta desafios, como a necessidade de infraestrutura adequada, integração de dados e adaptação dos modelos às condições locais (Botero-Valencia *et al.*, 2025). Portanto, neste estudo, o uso de aprendizado de máquina é compreendido como estratégia de apoio à interpretação de dados e à tomada de decisão, e não como substituição automática de análises agronômicas ou experimentais.

A aplicação de sistemas computacionais de apoio à decisão em contextos agrícolas também tem sido investigada em estudos recentes, incluindo soluções que

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

utilizam algoritmos de aprendizado de máquina para recomendar culturas com base em características do solo, clima e produção histórica (Senapaty; Ray; Padhy, 2024). Considerando esse cenário, este trabalho propõe o desenvolvimento de um aplicativo baseado em aprendizado de máquina como ferramenta de apoio a profissionais e pesquisadores da área agrícola e ambiental. A proposta busca auxiliar a análise preditiva do impacto de rizobactérias no desenvolvimento de plantas de arroz, a partir de dados experimentais, contribuindo para a organização das informações, a interpretação dos resultados e o apoio à tomada de decisão em pesquisas e aplicações relacionadas ao uso de RPCPs.

Dessa forma, busca-se articular conhecimentos da Engenharia de Software ao desenvolvimento de ferramentas computacionais voltadas à análise de dados agrícolas. O objetivo geral deste trabalho é desenvolver um aplicativo móvel, apoiado por técnicas de aprendizado de máquina, para auxiliar a análise do impacto de rizobactérias promotoras de crescimento no desenvolvimento de plantas de arroz.

Para alcançar esse objetivo, foram definidos os seguintes objetivos específicos:

- Coletar e organizar dados experimentais sobre rizobactérias promotoras de crescimento e seus impactos diretos e indiretos no crescimento de plantas de arroz;
- Realizar o levantamento de requisitos para a criação do aplicativo;
- Desenvolver o aplicativo e integrá-lo a um modelo de aprendizado de máquina adequado ao objetivo preditivo do sistema;
- Implementar o sistema para uso e geração de resultados a partir dos dados informados;
- Avaliar o desempenho técnico e a eficácia preditiva dos modelos de aprendizado de máquina integrados ao aplicativo, verificando sua viabilidade e adequação computacional para o contexto de predição agronômica.

Este trabalho está organizado em seções. A segunda seção apresenta o referencial teórico e os trabalhos relacionados. A terceira seção detalha a metodologia, as ferramentas e as técnicas usadas para o desenvolvimento. A quarta seção apresenta os resultados e a discussão. Por fim, a última seção traz as considerações finais.

2. REFERENCIAL TEÓRICO

2.1. Rizobactérias promotoras de crescimento de plantas

As rizobactérias promotoras de crescimento de plantas são bactérias associadas à rizosfera capazes de colonizar raízes e favorecer o desenvolvimento vegetal por diferentes mecanismos. De acordo com Andrade *et al.* (2023), esses microrganismos podem atuar de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

forma direta ou indireta, contribuindo para o aumento da disponibilidade de nutrientes, produção de fitormônios, desenvolvimento da parte aérea e das raízes, proteção contra fitopatógenos e redução de doenças. Além disso, as RPCPs podem auxiliar plantas submetidas a estresses abióticos, como seca, salinidade e presença de metais pesados.

A atuação dessas bactérias também está relacionada à possibilidade de reduzir a dependência de fertilizantes e pesticidas sintéticos. Andrade *et al.* (2023) destacam que as RPCPs têm sido estudadas como estratégia para uma produção agrícola mais sustentável, pois podem melhorar a absorção de nutrientes, favorecer a saúde do solo e contribuir para práticas agrícolas menos dependentes de insumos químicos. Assim, o uso de RPCPs se insere em um contexto de busca por alternativas biológicas voltadas ao crescimento vegetal e à sustentabilidade da produção.

2.2. Colonização da rizosfera e eficiência dos bioinoculantes

A rizosfera corresponde à região do solo diretamente influenciada pelas raízes das plantas, onde ocorre intensa interação entre compostos liberados pelas raízes e comunidades microbianas. Santoyo *et al.* (2021) afirmam que a eficiência das RPCPs como bioinoculantes depende da capacidade dessas bactérias de colonizar a rizosfera, competir com outros microrganismos e sobreviver sob diferentes condições ambientais. Dessa forma, a simples presença de microrganismos promotores de crescimento não garante sua eficácia, pois sua atuação depende da capacidade de estabelecimento no ambiente radicular.

Entre os mecanismos associados à colonização da rizosfera, Santoyo *et al.* (2021) citam o reconhecimento de sinais químicos e nutrientes provenientes dos exsudatos radiculares, a produção de biofilme, a motilidade bacteriana, atividades antioxidantes e a produção de compostos antimicrobianos, como antibióticos, compostos orgânicos voláteis e enzimas líticas. Esses fatores contribuem para que determinadas cepas consigam competir no ambiente da rizosfera e exercer efeitos benéficos sobre as plantas. Por isso, a seleção de cepas competitivas é um aspecto importante para o desenvolvimento e aplicação de bioinoculantes.

2.3. RPCPs no cultivo de arroz e tolerância a estresses abióticos

No cultivo de arroz, as RPCPs têm sido investigadas em razão de sua relação com a tolerância a estresses abióticos. Sonowal *et al.* (2024) destacam que o arroz, por ser uma cultura de importância alimentar, é vulnerável a fatores como seca, salinidade, toxicidade por

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

metais pesados e altas temperaturas. Esses fatores podem afetar negativamente a produtividade e a qualidade da cultura, tornando relevante o estudo de microrganismos capazes de auxiliar as plantas em condições adversas.

Segundo Sonowal *et al.* (2024), as RPCPs podem favorecer o desenvolvimento do arroz por meio de mecanismos como absorção de nutrientes, síntese de fitormônios, modificação de vias de sinalização relacionadas ao estresse e estímulo a genes responsivos. O estudo também indica que esses microrganismos podem contribuir para reduzir a dependência de pesticidas e fertilizantes químicos, ao mesmo tempo em que favorecem a produtividade vegetal, a saúde do solo e práticas agrícolas ambientalmente mais sustentáveis.

Apesar desse potencial, a aplicação de RPCPs em arroz apresenta desafios. Sonowal *et al.* (2024) apontam que diferentes cepas podem apresentar efeitos variados, e que fatores como tipo de solo, clima, pH, disponibilidade de nutrientes, práticas agronômicas, interação com outros microrganismos, ausência de protocolos padronizados e aplicação em escala de campo podem interferir nos resultados. Esses elementos reforçam a complexidade da análise dos efeitos das RPCPs sobre o desenvolvimento vegetal.

2.4. Aprendizado de máquina e apoio à decisão na agricultura

O uso de técnicas de aprendizado de máquina tem sido explorado em estudos agrícolas para análise de dados, construção de modelos preditivos e apoio à tomada de decisão. Botero-Valencia *et al.* (2025) afirmam que essas técnicas permitem analisar grandes volumes de dados, identificar padrões complexos e integrar informações provenientes de diferentes fontes. No contexto da agricultura sustentável, o aprendizado de máquina tem sido associado a aplicações como monitoramento da saúde do solo, integração de dados agrícolas, apoio à tomada de decisão e agricultura de precisão.

Entretanto, a adoção dessas técnicas também envolve desafios. Botero-Valencia *et al.* (2025) destacam limitações relacionadas à infraestrutura, ao gerenciamento de grandes volumes de dados, à qualidade e representatividade das bases utilizadas, à integração entre diferentes fontes de informação e à adaptação dos modelos às condições locais. Assim, o uso de aprendizado de máquina na agricultura exige que os modelos sejam avaliados de forma criteriosa, considerando o contexto experimental e a finalidade da aplicação.

Sistemas computacionais de apoio à decisão também têm sido aplicados em contextos agrícolas. Senapaty, Ray e Padhy (2024) desenvolveram um sistema de recomendação de culturas baseado em algoritmos de aprendizado de máquina, utilizando dados

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

agrícolas para auxiliar escolhas relacionadas à seleção de culturas. O estudo comparou diferentes classificadores, incluindo *KNN*, *Random Forest* e *XGBoost*, demonstrando a aplicação de modelos de classificação em problemas agrícolas. Nesse sentido, o presente trabalho se aproxima dessa abordagem ao utilizar técnicas de aprendizado de máquina em um aplicativo móvel voltado à análise preditiva de dados relacionados ao desenvolvimento de plantas de arroz submetidas à inoculação de rizobactérias.

3. METODOLOGIA

3.1. Classificação da pesquisa

Esta pesquisa, quanto à sua natureza, classifica-se como aplicada, pois visa gerar um artefato de software para apoio à tomada de decisão no contexto agronômico. Em relação aos objetivos, assume um caráter exploratório e descritivo, ao investigar os padrões de interação entre as rizobactérias e plantas de arroz, e descrever o processo de construção de um sistema preditivo. A abordagem do problema é quantitativa, trazendo um método de pesquisa focado na análise estatística dos dados, realização de testes de hipótese e na avaliação de desempenho dos algoritmos de aprendizado de máquina. Em relação aos procedimentos técnicos, a investigação envolve pesquisa bibliográfica, pesquisa experimental (com base em dados de campo) e desenvolvimento tecnológico.

3.2. Levantamento bibliográfico

As fontes utilizadas na revisão bibliográfica foram levantadas por meio do Google Acadêmico, com recorte temporal entre 2021 e 2026. As buscas foram realizadas prioritariamente em inglês, considerando a maior disponibilidade de publicações internacionais na área. Foram empregadas combinações de termos relacionados a arroz, segurança alimentar, mudanças climáticas, fertilizantes químicos, rizobactérias promotoras de crescimento de plantas, estresses abióticos, colonização da rizosfera, aprendizado de máquina, agricultura sustentável e sistemas de apoio à decisão. Priorizaram-se revisões recentes, artigos publicados em periódicos científicos e estudos diretamente relacionados ao cultivo de arroz, às RPCPs e à aplicação de aprendizado de máquina em contextos agrícolas.

3.3. Etapas de pesquisa e desenvolvimento do sistema

A execução do trabalho seguiu um fluxo estruturado em quatro fases principais.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A Fase 1 consistiu na definição do problema, na revisão da literatura e no levantamento de requisitos para o sistema. As necessidades funcionais e os dados de entrada foram discutidos com uma especialista da área de engenharia florestal, parceira do projeto, permitindo delimitar as variáveis agronômicas relevantes e os dois cenários de análise preditiva: não destrutiva (plantas vivas) e destrutiva (análise laboratorial).

A Fase 2 abrangeu a coleta, a limpeza e a análise exploratória dos dados fornecidos. Esses dados, provenientes de experimentos de inoculação de rizobactérias em cultivos de arroz, foram submetidos a procedimentos de tratamento de valores ausentes e identificação de *outliers*, descritos na seção 2.4.

A Fase 3 envolveu o desenvolvimento do modelo de aprendizado de máquina e do aplicativo móvel. O sistema foi concebido em um ciclo evolutivo que permitiu a refatoração de ambos os componentes à medida que novas variáveis contínuas se mostraram necessárias para aprimorar a capacidade preditiva dos modelos de *machine learning*. Essa evolução levou à arquitetura multimodal adotada.

A Fase 4, por fim, dedicou-se à avaliação. O modelo com melhor desempenho nas métricas de classificação foi selecionado, integrado ao sistema e submetido a testes automatizados de *API*.

3.4. Origem e tratamento de dados

A base de dados utilizada para treinamento e avaliação dos modelos é proveniente de experimentos de campo conduzidos por profissionais de engenharia florestal e cedida ao projeto. Ela reúne 6040 registros de mensurações biométricas de plantas de arroz de diferentes tipos, submetidas à inoculação de diferentes cepas de rizobactérias promotoras de crescimento, cultivadas em diferentes cidades, mas após o pré-processamento, foram utilizados 6038 registros.

O pré-processamento inicial foi realizado com o auxílio das bibliotecas *Pandas* e *NumPy*. Inicialmente, as variáveis categóricas foram padronizadas para eliminar inconsistências de formatação de texto. Em seguida, um total de três registros que apresentavam valores ausentes nas variáveis de comprimento da parte aérea (PA) e o comprimento da raiz, foram substituídos pela mediana dos valores das suas respectivas colunas. Para a análise de *outliers*, empregou-se a visualização por meio de gráficos de caixa (*boxplot*), mas adotou-se uma abordagem conservadora, removendo-se apenas dois valores extremos na variável do comprimento da parte aérea (acima de 100 cm), pois variações biométricas acentuadas podem representar respostas fisiológicas legítimas aos tratamentos biológicos e sua exclusão em massa comprometeria a

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

capacidade de generalização dos modelos. Após essa etapa, a base consolidou-se em 6038 observações, que serviriam de entrada para as análises posteriores.

3.5. Desenvolvimento do modelo de aprendizado de máquina

Para a construção do modelo preditivo, a variável alvo foi definida a partir do Índice de Qualidade de Dickson (IQD), calculado previamente em análises laboratoriais. A mediana de IQD (0,6467) foi utilizada como limiar para dividir os dados em duas classes de impacto: “Padrão” (IQD < mediana) e “Alto” (IQD ≥ mediana). Como não foram identificados registros com IQD inferior a 0,2 — limiar agronômico associado a mudas inadequadas para plantio —, optou-se por uma classificação binária. Esse procedimento resultou em um balanceamento natural, com exatamente 3019 registros em cada classe, eliminando a necessidade de técnicas artificiais de balanceamento.

O pré-processamento dos dados de entrada incluiu a separação das variáveis preditivas (X) e da variável alvo (y). As variáveis categóricas preditivas (organismo, cidade e tipo de cultivo) foram transformadas em formato numérico utilizando o método *One-Hot Encoding* (configurado para ignorar categorias desconhecidas), enquanto a variável alvo foi codificada por meio de um *Label Encoder*, porém as variáveis numéricas contínuas foram mantidas em sua escala original, dada a proximidade de seus valores. O conjunto de dados foi então dividido aleatoriamente em subconjuntos de treinamento (80%) e teste (20%), utilizando estratificação (*stratify*) baseada na variável alvo para garantir a manutenção do balanceamento das classes e assegurar que a avaliação final ocorresse sobre dados não vistos durante o aprendizado.

Foram treinados e comparados três algoritmos de classificação — *K-Nearest Neighbors* (*KNN*), *Random Forest* e *XGBoost* —, escolhidos por suas diferentes estratégias de aprendizado e aplicabilidade a dados não lineares. Para otimizar o desempenho e combater o *overfitting*, os hiperparâmetros de todos os modelos foram ajustados automaticamente por meio da técnica de busca em grade (*GridSearchCV*). No caso do modelo *XGBoost*, os hiperparâmetros finais estabelecidos pelo otimizador foram: *n_estimators* = 200, *learning_rate* = 0,2, *max_depth* = 7, *subsample* = 0,8 e *colsample_bytree* = 0,8.

A seleção do melhor modelo foi orientada pela avaliação de métricas extraídas de relatórios de classificação (*classification report*) e matrizes de confusão, incluindo a acurácia, precisão, *recall* e *F1-Score* para cada classe. A etapa de validação foi suprida pelo uso da técnica de Validação Cruzada com 10 partições (*10-Fold Cross-Validation*) aplicada sobre o conjunto de treinamento, avaliando-se a acurácia média e o desvio padrão. Adicionalmente, para atestar a significância estatística na comparação dos desempenhos entre os modelos, foi calculado o

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Intervalo de Confiança (IC) de 95% da acurácia média, utilizando a distribuição T de *Student*. O algoritmo com melhor acurácia, precisão e equilíbrio entre viés e variância (menor desvio padrão na validação cruzada) seria selecionado para integração ao *back-end*.

Durante essa etapa, identificou-se a necessidade de explorar dois caminhos preditivos distintos: uma análise não destrutiva, utilizando apenas variáveis categóricas e medições de fácil obtenção em campo (comprimento da parte aérea e diâmetro do coleto); e uma análise destrutiva, que incorpora métricas adicionais destrutivas das plantas (PSR, PSA, PST e comprimento da raiz), obtidas em laboratório. O mesmo algoritmo selecionado foi então treinado separadamente para cada cenário, gerando os dois modelos que coexistem na arquitetura final do aplicativo móvel.

3.6. Desenvolvimento do aplicativo móvel

O desenvolvimento do aplicativo “Biogrow” partiu do levantamento de requisitos com a especialista de domínio, que indicou a necessidade de uma ferramenta para consulta rápida em campo e laboratório, com o objetivo de oferecer suporte à tomada de decisão e reduzir a necessidade de coletas destrutivas. Para apoiar a obtenção e a especificação dessas funcionalidades, foram elaborados protótipos de interface no ambiente Figma, permitindo organizar e refinar os módulos antes da codificação. O fluxo projetado foi organizado em três módulos lógicos — apresentação, coleta de dados e exibição de resultados — e a navegação foi posteriormente adaptada para acomodar as duas modalidades de análise (não destrutiva e destrutiva).

A arquitetura do sistema adota o modelo Cliente-Servidor. O cliente móvel foi desenvolvido em linguagem Kotlin com o *framework Jetpack Compose*, utilizando o padrão de arquitetura *MVVM* (*Model-View-ViewModel*) para separar a lógica de apresentação dos dados. A comunicação com o servidor é feita de forma assíncrona por meio da biblioteca *Retrofit*. O *back-end* foi implementado em *Python*, com o *framework FastAPI*, e disponibiliza uma *API RESTful* com dois *endpoints* de predição. O gerenciamento do ciclo de vida da aplicação, via manipulador *lifespan*, carrega os modelos *XGBoost* e os codificadores na memória uma única vez durante a inicialização, otimizando o tempo de resposta das requisições. A validação dos dados de entrada é realizada com modelos *Pydantic*, que rejeitam automaticamente requisições malformadas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

3.7. Avaliação do Sistema

A avaliação do sistema foi conduzida pela abordagem quantitativa, que consistiu na análise de desempenho dos modelos preditivos sobre o conjunto de teste, utilizando as métricas de acurácia, precisão, *recall* e *F1-Score*, além da matriz de confusão para verificar a distribuição dos acertos por classe. Adicionalmente, uma suíte de testes automatizados foi implementada com o *framework Pytest* sobre a *API*, simulando tanto cenários de sucesso quanto de falha (ausência de parâmetros e indisponibilidade dos artefatos), para verificar a robustez do *back-end*.

Por fim, realizou-se uma validação qualitativa preliminar da ferramenta com a especialista de domínio em engenharia florestal e sua equipe de pesquisa. Essa avaliação prática confirmou que o aplicativo atende de forma satisfatória aos requisitos operacionais do projeto, solucionando a necessidade de prever o impacto das rizobactérias sem a exigência de métodos destrutivos, o que viabiliza a entrega da solução e justifica a postergação de testes formais de usabilidade em larga escala para trabalhos futuros.

4. RESULTADOS E DISCUSSÃO

4.1. Caracterização da base de dados e análise exploratória

O processo de limpeza descrito na Metodologia resultou em uma base consolidada de 6038 registros, cada um representando um microrganismo inoculado em diferentes tipos de plantações de arroz, em diferentes cidades e sistema de cultivo. A Tabela 1 sumariza as estatísticas descritivas das variáveis contínuas, podendo-se observar que variáveis como o Índice de Qualidade de Dickson (IQD) e o Comprimento da Parte Aérea (PA) apresentam distribuição aproximadamente simétrica (assimetria de -0,28 e -0,05, respectivamente), enquanto Diâmetro do Coleto e Comprimento da Raiz exibem forte assimetria positiva e caudas pesadas (curtose de 12,38 e 1,04), sugerindo que a maioria das plantas se concentra em valores menores, com alguns poucos espécimes atingindo medidas muito acima da média. Essa heterogeneidade é esperada em experimentos com organismos vivos e reforça a relevância de algoritmos que não assumem normalidade.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

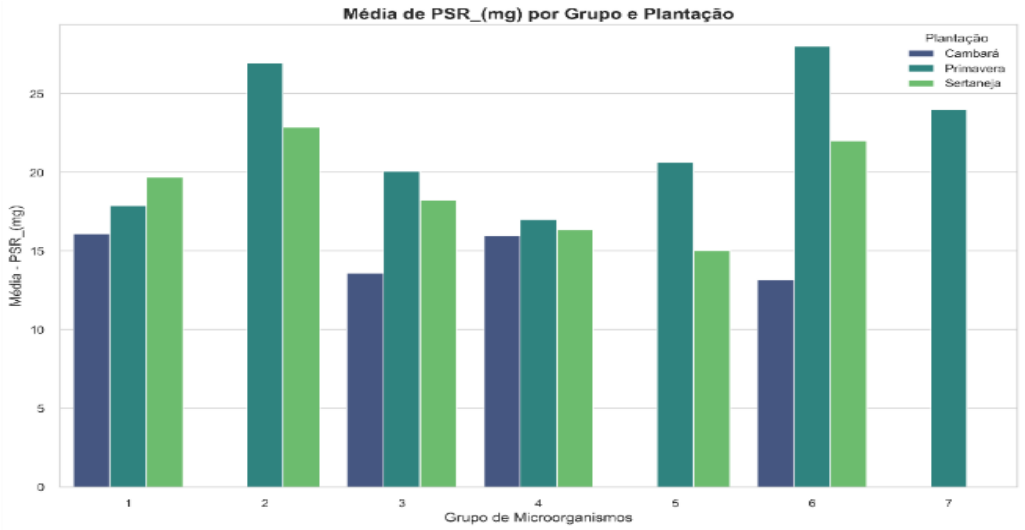
Tabela 1. Estatísticas descritivas das principais variáveis contínuas

Variável	Média	Desvio Padrão	Assimetria	Curtose	Interpretação
PA (cm)	36,35	5,73	-0,05	0,43	Aproximadamente simétrica
Diâmetro do Coleto (mm)	1,56	0,45	2,29	12,38	Forte assimetria positiva
Comprimento da Raiz (cm)	11,81	5,07	0,77	1,04	Assimetria positiva moderada
IQD	0,64	0,08	-0,28	0,85	Aproximadamente simétrica

Fonte: Autoria Própria (2026)

Os gráficos de barras agrupadas (Figura 1) evidenciam o comportamento diferencial dos sete grupos de microrganismos em relação às médias das variáveis numéricas. Destacam-se os grupos 2, 6 e 7 no cultivo “Primavera”, com médias de peso seco total (PST) até 25% superiores às do cultivo “Testemunha”. O grupo 4, por sua vez, apresentou um pico de comprimento radicular no mesmo sistema, indicando que o efeito promotor pode se manifestar de forma específica em diferentes órgãos da planta, a depender da cepa. Essas interações, visualmente perceptíveis, sustentam a decisão de incluir tanto o microrganismo quanto o tipo de cultivo como variáveis de entrada no modelo preditivo.

Figura 1. Barras agrupadas – médias por grupo e cultivo

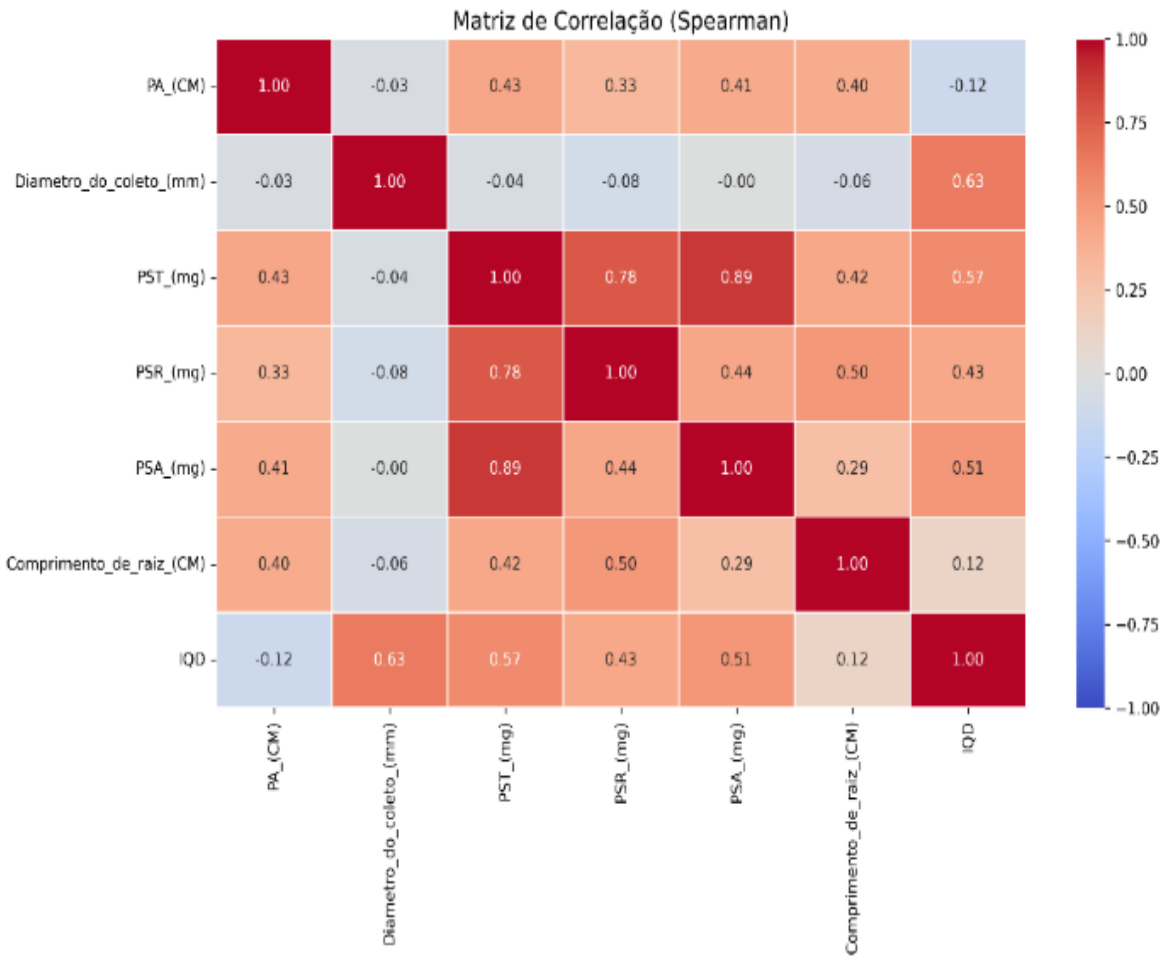


Fonte: Autoria Própria (2026)

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A matriz de correlação de *Spearman* (Figura 2) revelou associações relevantes entre as variáveis contínuas. O PST e o Peso Seco da Parte Aérea (PSA) apresentaram correlação de 0,89, o que indica redundância informacional na análise destrutiva, embora biologicamente esperada. Para a modalidade não destrutiva, o resultado mais relevante foi a correlação positiva entre Diâmetro do Coleto e IQD (0,63), porém o Comprimento da Parte Aérea mostrou correlação praticamente nula com o IQD (-0,12). Esse resultado oferece suporte empírico para a viabilidade da predição em campo: prever a qualidade do crescimento vegetal utilizando o diâmetro do coleto, uma medição simples e não invasiva, dispensando a coleta de dados destrutivos das plantas.

Figura 2. Heatmap de correlação de Spearman



Fonte: Autoria Própria (2026)

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A aplicação do teste de *Shapiro-Wilk* às 49 subamostras (7 variáveis x 7 grupos de microrganismos) indicou que 45 delas (91,8%) rejeitaram a hipótese nula de normalidade, com o valor de p inferior a 0,05. A *ANOVA* de uma via (*One-Way*), entretanto, avaliou simultaneamente a existência de diferenças significativas entre as médias dos grupos experimentais ($p < 0,001$).

Para identificar quais grupos diferiam entre si, foi realizado o teste *post-hoc* de *Tukey*. As comparações múltiplas permitiram identificar as cepas de microrganismos, como os Grupos 2, 6 e 7, que apresentaram médias de IQD significativamente superiores ao grupo testemunha ($p < 0,05$), caracterizando os tratamentos com maior potencial de promoção do crescimento vegetal.

4.2. Desenvolvimento do Sistema "Biogrow" – Prototipação, Aplicativo e API

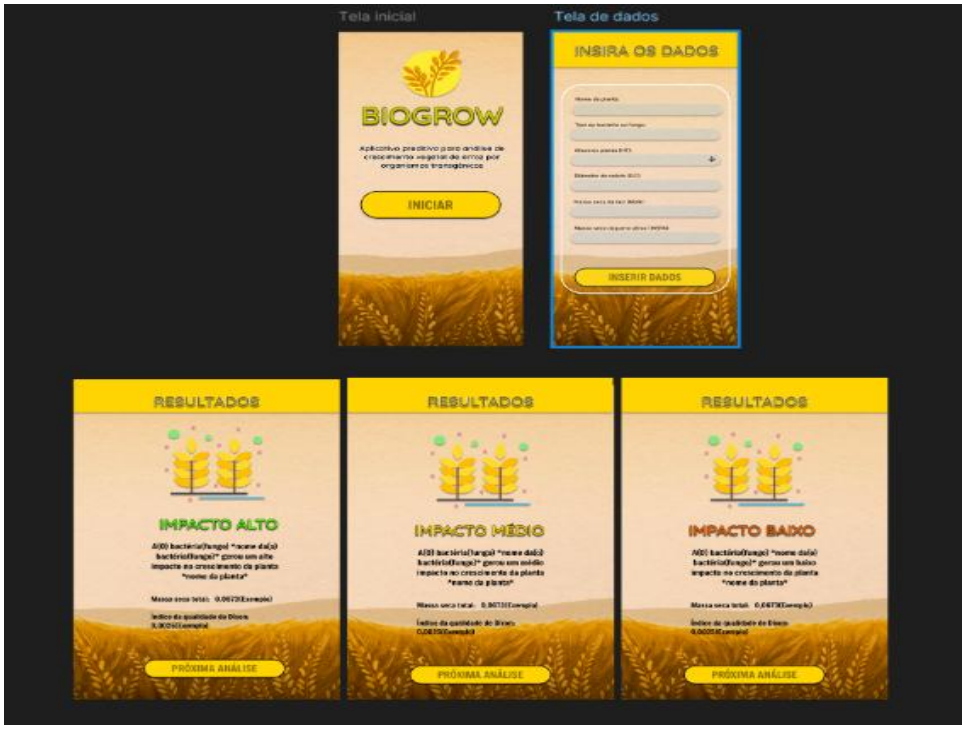
Nesta seção, são apresentados os artefatos gerados nas etapas de design e implementação, compreendendo o resultado da prototipação, o aplicativo móvel final e o *back-end* integrado, como entregas de um mesmo sistema.

4.2.1. Prototipação no Figma

A etapa de prototipação resultou em um projeto visual completo na plataforma Figma, abrangendo o fluxo de telas, a paleta de cores e a disposição dos componentes da interface. A Figura 3 mostra a prancha geral do projeto, na qual é possível visualizar o encadeamento entre a tela de abertura, o formulário para inserção de dados e a tela de resultados, e permite visualizar previamente a sequência de ações que conduz ao resultado da análise preditiva, funcionando como artefato de apoio à transição entre requisitos e codificação.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 3. Prancha geral do projeto no Figma com as telas



Fonte: Autoria Própria (2026)

4.2.2. Aplicativo Móvel e Integração com o Back-end

A implementação do aplicativo em *Kotlin* utilizando o *framework Jetpack Compose* resultou em um sistema funcional para dispositivos Android. A interface foi projetada para oferecer uma navegação simples e intuitiva, iniciando com uma tela de apresentação do sistema e, em seguida, disponibilizando ao usuário a seleção entre as modalidades de análise não destrutiva e destrutiva. A partir dessa escolha, o aplicativo direciona o usuário para os formulários específicos de entrada de dados, que são processados pelo modelo de aprendizado de máquina hospedado no *back-end*.

Os formulários de entrada (Figura 4) incorporam validação de tipos e formatos: campos numéricos aceitam exclusivamente dígitos, enquanto os campos categóricos utilizam menus suspensos com as opções de preenchimento de texto. Todos os dados inseridos representam as variáveis utilizadas no treinamento dos modelos de *machine learning*.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 4. Formulário (Destrutivo e Não Destrutivo) preenchido

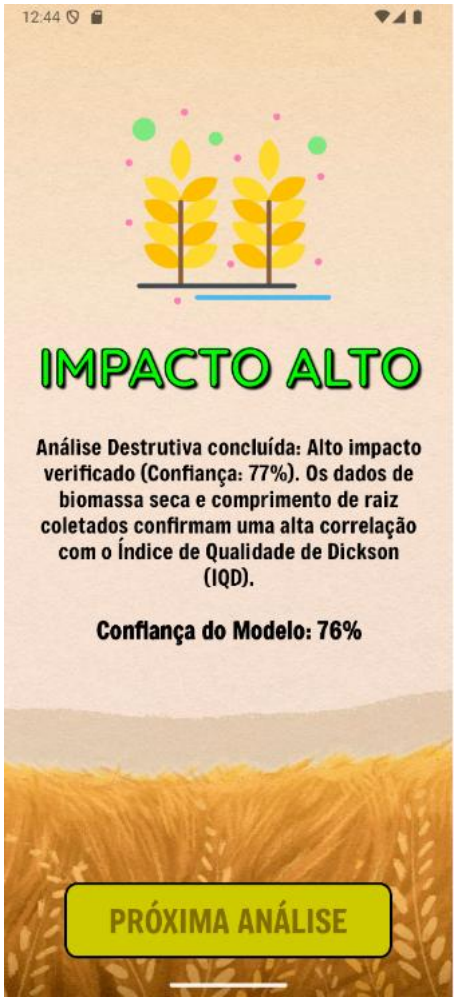


Fonte: Autoria Própria (2026)

A tela de resultados (Figura 5) apresenta a classe de impacto prevista, a probabilidade de confiança associada e uma recomendação textual para o usuário. A mudança de cor do texto — tons de verde para impacto “Alto”, tons de laranja para impacto “Padrão” — funciona como realce visual para a tomada de decisão. Esse comportamento reativo é gerenciado pelo *ViewModel*, que consome a resposta *JSON* do servidor e atualiza o estado da interface.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 5. Tela de resultado após predição



Fonte: Autoria Própria (2026)

O *back-end*, desenvolvido em *FastAPI*, responde aos *endpoints*"/predict/nao-destrutivo" e "/predict/destrutivo". A documentação automática (*Swagger*) gerada pelo *framework*, acessível via navegador, lista os esquemas *Pydantic* de entrada e saída, facilitando a inspeção dos contratos de dados. Os testes automatizados com *Pytest* cobriram o fluxo de predição correta (HTTP 200 com classe e probabilidade), a rejeição de requisições com campos faltantes (HTTP 422) e a indisponibilidade dos arquivos de modelo (HTTP 503). Todos os cenários retornaram os códigos esperados, indicando a integração entre cliente, servidor e modelos de *machine learning*, operou conforme esperado nos cenários testados.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4.3. Desempenho dos Modelos Preditivos

4.3.1. Seleção e desempenho dos algoritmos

A Tabela 2 apresenta os resultados da validação cruzada (*10-Fold*) para os três algoritmos testados. O *XGBoost* obteve a maior acurácia média (0,941) com o menor desvio padrão (0,010), indicando estabilidade superior entre os modelos, com um Intervalo de Confiança (IC 95%) entre 0,932 e 0,948. O *Random Forest* alcançou acurácia de 0,934 e um desvio padrão de 0,011 (IC 95%: 0,925 - 0,942), representando um resultado altamente competitivo. Já o modelo *KNN* ficou com uma acurácia média de 0,813, com desvio de 0,019 (IC 95%: 0,798 - 0,827), refletindo sua maior sensibilidade à distribuição assimétrica das variáveis. As métricas de *F1-Score* por classe seguiram o mesmo ordenamento, com o *XGBoost* atingindo 0,94 para ambas as classes “Padrão” e “Alto”, respectivamente.

A proximidade entre a acurácia média obtida na validação cruzada (0,941) e a acurácia no conjunto de teste isolado (0,945) para o *XGBoost* evidencia uma excelente capacidade de generalização, minimizando expressivamente os riscos de *overfitting*. Além disso, os testes de significância revelam que, embora haja um empate técnico estatístico entre o *XGBoost* e o *Random Forest* devido à sobreposição de seus intervalos, ambos são significativamente superiores ao *KNN*. O *XGBoost* consolida-se como escolha para a implementação devido a apresentar a maior média absoluta de acertos atrelada à menor variância.

Tabela 2. Resultados métricos do desempenho dos modelos não destrutivos

Algoritmo	Acurácia (Conjunto Teste)	Acurácia Média (10-Fold CV)	Desvio Padrão	F1-Score (Padrão)	F1-Score (Alto)	IC 95%
KNN	0,830	0,813	0,019	0,83	0,83	0,798 - 0,827
Random Forest	0,935	0,934	0,011	0,93	0,93	0,925 - 0,942
XGBoost	0,945	0,941	0,010	0,94	0,94	0,932 - 0,948

Fonte: Autoria Própria (2026)

A Tabela 3 apresenta os resultados da validação cruzada (*10-Fold*) para os três algoritmos testados. O *XGBoost* obteve a maior acurácia média (0,975) com o menor desvio padrão (0,006) e um Intervalo de Confiança (IC 95%) entre 0,970 e 0,980, novamente obtendo um resultado superior. O *Random Forest* alcançou acurácia de 0,940 e um desvio padrão de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

0,014 (IC 95%: 0,930 - 0,950), melhorando ligeiramente a acurácia e aumentando o desvio padrão. Já o modelo *KNN* ficou com um valor de 0,791 em acurácia média, com desvio de 0,017 (IC 95%: 0,777 - 0,803), confirmando novamente a sua sensibilidade em relação à assimetria. As métricas de *F1-Score* por classe seguiram o mesmo ordenamento, com o *XGBoost* atingindo um valor de *F1-Score* de 0,97 para ambas as classes “Padrão” e “Alto”, respectivamente.

Neste cenário destrutivo, a inclusão das variáveis biométricas de laboratório rompeu o empate técnico anterior: não há sobreposição entre o IC do *XGBoost* e os demais algoritmos, atestando sua superioridade com significância estatística absoluta. Além disso, o modelo manteve o baixo risco de *overfitting*, uma vez que o seu desempenho no conjunto de dados de teste (0,973) permaneceu rigorosamente alinhado à média obtida durante a validação cruzada (0,975).

Tabela 3. Resultados métricos do desempenho dos modelos destrutivos

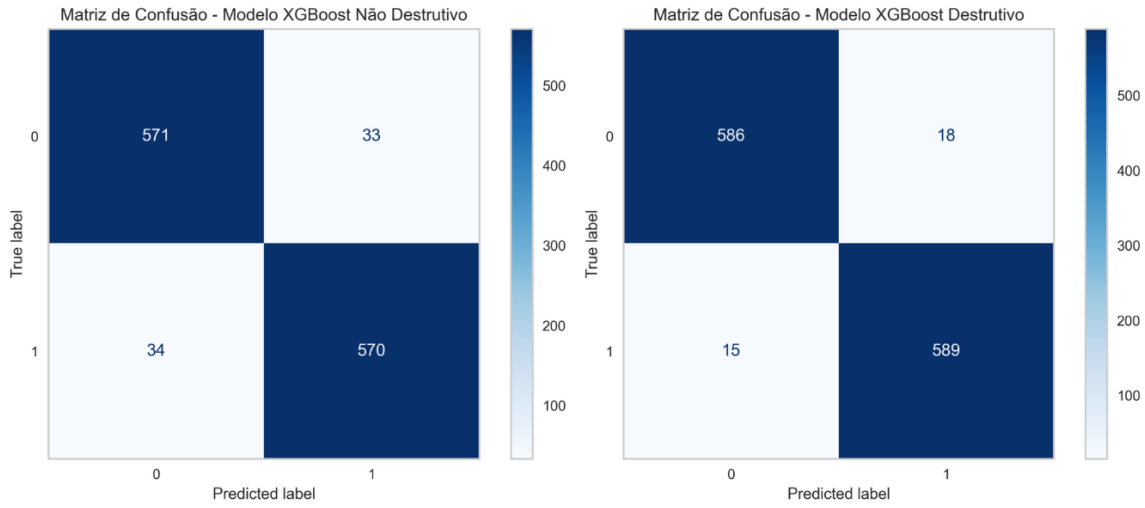
Algoritmo	Acurácia (Conjunto Teste)	Acurácia Média (10-Fold CV)	Desvio Padrão	F1-Score (Padrão)	F1-Score (Alto)	IC 95%
KNN	0,806	0,791	0,017	0,81	0,81	0,777 - 0,803
Random Forest	0,942	0,94	0,014	0,94	0,94	0,930 - 0,950
XGBoost	0,973	0,975	0,006	0,97	0,97	0,970 - 0,980

Fonte: Autoria Própria (2026)

Com base nesses resultados, o *XGBoost* foi selecionado para compor o sistema, tanto na análise destrutiva quanto na não destrutiva. As matrizes de confusão dos modelos selecionados (Figura 6) demonstram elevada capacidade de classificação para ambas as classes. No modelo não destrutivo, foram observadas apenas 67 classificações incorretas, distribuídas de forma praticamente simétrica entre falsos positivos (33) e falsos negativos (34). De maneira semelhante, o modelo destrutivo apresentou apenas 33 erros, correspondentes a 18 falsos positivos e 15 falsos negativos. Esses resultados indicam ausência de viés relevante em favor de qualquer classe, além de evidenciar a consistência do modelo na identificação dos padrões avaliados.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 6. Matriz de Confusão XGBoost (Destrutivo e Não Destrutivo)



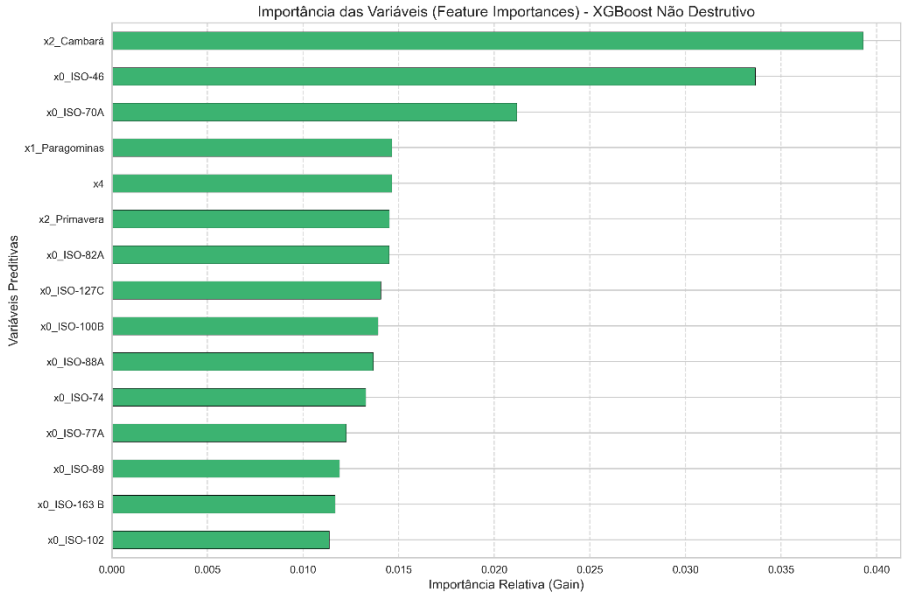
Fonte: Autoria Própria (2026)

4.3.2. Análise da Importância das Features

A análise de importância das variáveis (Figura 7) obtida pelo modelo *XGBoost* não destrutivo posiciona o Diâmetro do Coleto como a variável de maior relevância com base no ganho de informação (*gain*), seguida pelo tipo de cultivo e pela cidade, o que corrobora a correlação com o IQD identificada na análise exploratória. No modelo destrutivo, por sua vez, o Peso Seco da Raiz ganha destaque, um resultado perfeitamente coerente com a fisiologia da planta, na qual o desenvolvimento do sistema radicular reflete diretamente a eficiência na absorção de nutrientes e o acúmulo de biomassa.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 7. Importância das features – XGBoost não destrutivo

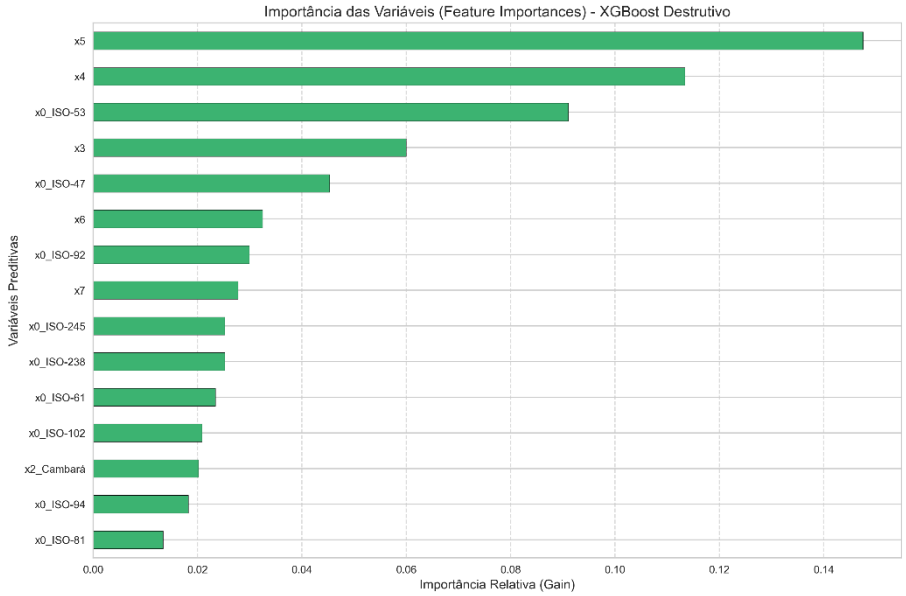


Fonte: Autoria Própria (2026)

Ao observar o comportamento do modelo destrutivo (Figura 8), o gráfico de importância gerado pelo *XGBoost* evidencia uma clara mudança no protagonismo das variáveis preditivas em relação à abordagem anterior. Com base no ganho de informação (*gain*), o Peso Seco da Raiz assume isoladamente o topo da relevância para as decisões do algoritmo. Esse destaque é fisiologicamente fundamentado, uma vez que a biomassa do sistema radicular atua como um reflexo direto da capacidade e eficiência da planta na absorção de água e nutrientes essenciais. Integrando essas métricas diretas, nota-se que a quantificação real da biomassa fornece ao classificador, indicadores estruturais muito mais precisos e definitivos sobre a qualidade final da muda, consolidando o excelente desempenho preditivo desse modelo.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 8. Importância das features – XGBoost Destrutivo



Fonte: Autoria Própria (2026)

4.4. Discussão e limitação de estudos

A comparação dos resultados obtidos com a literatura recente evidencia a consistência da abordagem analítica adotada nesse projeto. Estudos como o de Senapaty, Ray e Padhy (2024), que investigam a aplicação de algoritmos de *machine learning* para a predição no cultivo de arroz, também constataram a superioridade de modelos baseados em *Gradient Boosting* e agrupamentos de árvores (*ensembles*) para lidar com a não linearidade característica dos dados agrícolas. No presente estudo, a consolidação do XGBoost como algoritmo mais eficiente e estável, tanto na análise destrutiva quanto não destrutiva, corrobora essas observações literárias, reforçando a capacidade dessa arquitetura matemática para a modelagem de variáveis agronômicas e biológicas complexas.

Apesar do elevado desempenho preditivo e da mitigação dos riscos de *overfitting* — atestada matematicamente pela grande proximidade entre as acurácias de validação cruzada (10-Fold) e as do conjunto de dados de teste não vistos —, o modelo apresenta limitações próprias que necessitam ser destacadas. A capacidade de predição do sistema possui forte dependência da qualidade, da padronização e do escopo estrito dos dados experimentais utilizados em sua fundamentação, e como a base de dados reflete o uso de cepas de rizobactérias avaliadas em condições específicas de cultivo experimental, há potenciais

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

limitações de generalização. Consequentemente, a aplicação direta da ferramenta em diferentes regiões agrícolas, que apresentem variações de clima, microbioma do solo ou práticas de manejo, pode sofrer restrições de precisão preditiva, recomendando-se a realização de calibrações ou o retreinamento do modelo com dados locais antes de sua adoção em larga escala comercial.

Por fim, em relação à aplicação tecnológica, a arquitetura de software desenvolvida apresenta alto potencial de escalabilidade para mitigar essas limitações ao longo do tempo. A separação estrutural entre a interface móvel e a inteligência analítica isolada em uma *API* (*FastAPI*) torna o aplicativo "Biogrow" uma ferramenta completamente adaptável em relação à cultura analisada. Em relação à interação humano-computador, a restrição da avaliação de usabilidade a uma validação qualitativa preliminar com especialistas do domínio também configura uma limitação deste estudo. Embora essa validação inicial tenha sido suficiente para confirmar a eficácia da interface e o atendimento aos requisitos operacionais do projeto, a ausência de testes com uma amostragem mais ampla de usuários finais aponta para a necessidade de avaliações formais futuras.

Embora a delimitação atual do estudo seja voltada exclusivamente para as plantações de arroz (*Oryza sativa*), a lógica preditiva e a estrutura de mensuração podem ser perfeitamente expandidas para outras culturas de alta relevância econômica e segurança alimentar, como soja, trigo ou milho. Para tal adaptação, seria necessária apenas a inserção de novos conjuntos de dados experimentais correspondentes a essas espécies no processo de treinamento, reajustando-se as variáveis de biomassa esperada e os limiares de classificação do Índice de Qualidade de Dickson (IQD) de acordo com os padrões morfológicos de cada cultura.

5. CONSIDERAÇÕES FINAIS

O presente trabalho desenvolveu o aplicativo “Biogrow”, uma ferramenta móvel apoiada por aprendizado de máquina para análise preditiva do impacto de rizobactérias promotoras de crescimento no desenvolvimento de plantas de arroz. O sistema integrou dois modelos de classificação baseados no algoritmo *XGBoost*, correspondentes às modalidades de análise não destrutiva e destrutiva, disponibilizados por meio de uma *API RESTful* desenvolvida em *FastAPI* e consumidos por um cliente *Android* implementado em *Kotlin* com *Jetpack Compose*.

Os resultados da análise exploratória indicaram forte assimetria positiva em variáveis como Diâmetro do Coleto e Comprimento da Raiz, além de correlação positiva relevante entre

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Diâmetro do Coleto e IQD (0,63), o que forneceu suporte empírico à viabilidade da predição não destrutiva em campo. O modelo não destrutivo alcançou acurácia de 0,945 no conjunto de teste, enquanto o modelo destrutivo elevou esse valor para 0,973, sugerindo que a inclusão de métricas de biomassa seca agrega capacidade preditiva ao sistema. Em ambos os casos, as matrizes de confusão evidenciaram distribuição equilibrada dos erros, sem viés relevante para qualquer classe de impacto.

A arquitetura Cliente-Servidor, estruturada com o padrão *MVVM*, permitiu a comunicação assíncrona entre o aplicativo e os modelos preditivos, assegurando que o processamento das previsões não bloqueasse a interface do usuário. Os testes automatizados com *Pytest* confirmaram o comportamento esperado do *back-end,* tanto nos fluxos de sucesso quanto nos cenários de falha, indicando robustez na validação de entradas e na indisponibilidade dos artefatos.

Como contribuições, o trabalho oferece um sistema funcional que pode auxiliar profissionais e pesquisadores da área agrícola na interpretação de dados experimentais relacionados ao uso de RPCPs, apoiando a tomada de decisão sem a necessidade de coletas destrutivas em todas as situações. Ademais, a documentação dos procedimentos de pré-processamento, treinamento e avaliação dos modelos serve como referência metodológica para iniciativas semelhantes.

Para trabalhos futuros, sugere-se a ampliação da base de dados com experimentos em diferentes regiões e condições de solo e clima, de modo a fortalecer a generalização dos modelos. A incorporação de variáveis ambientais adicionais, como temperatura e umidade do solo, poderia enriquecer a análise preditiva. Por fim, recomenda-se a realização de testes formais de usabilidade com o público-alvo final — incluindo pesquisadores, agricultores e extensionistas rurais — para avaliar a experiência do usuário (*UX*) e a ergonomia da interface do aplicativo em condições reais de campo.

REFERÊNCIAS

ANDRADE, Luana Alves de; SANTOS, Carlos Henrique Barbosa; FREZARIN, Edvan Teciano; SALES, Luziane Ramos; RIGOBELO, Everlon Cid. Plant Growth-Promoting Rhizobacteria for Sustainable Agricultural Production. **Microorganisms**, v. 11, n. 4, art. 1088, p. 1-16, 2023. DOI: 10.3390/microorganisms11041088. Disponível em: https://www.mdpi.com/2076-2607/11/4/1088. Acesso em: 31 maio 2026.

BOTERO-VALENCIA, Juan *et al.* Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives. **Agriculture**, v. 15, n. 4, art. 377, 2025. DOI: 10.3390/agriculture15040377. Disponível em: https://www.mdpi.com/2077-0472/15/4/377. Acesso em: 31 maio 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

FAROOQ, Ahsan; FAROOQ, Nageen; AKBAR, Haseeb; HASSAN, Zia Ul; GHEEWALA, Shabbir H. A Critical Review of Climate Change Impact at a Global Scale on Cereal Crop Production. **Agronomy**, v. 13, n. 1, art. 162, p. 1-19, 2023. DOI: 10.3390/agronomy13010162. Disponível em: https://www.mdpi.com/2073-4395/13/1/162. Acesso em: 31 maio 2026.

PAHALVI, Heena Nisar; RAFIYA, Lone; RASHID, Sumaira; NISAR, Bisma; KAMILI, Azra N. Chemical Fertilizers and Their Impact on Soil Health. *In:* DAR, Gowhar Hamid; BHAT, Rouf Ahmad; MEHMOOD, Mohammad Aneesul; HAKEEM, Khalid Rehman (ed.). **Microbiota and Biofertilizers, Vol 2**: Ecofriendly Tools for Reclamation of Degraded Soil Environs. Cham: Springer, 2021. cap. 1, p. 1-44. DOI: 10.1007/978-3-030-61010-4_1. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-61010-4_1. Acesso em: 31 maio 2026.

SANTOYO, Gustavo; URTIS-FLORES, Carlos Alberto; LOEZA-LARA, Pedro Damián; OROZCO-MOSQUEDA, Ma. del Carmen; GLICK, Bernard R. Rhizosphere Colonization Determinants by Plant Growth-Promoting Rhizobacteria (PGPR). **Biology**, v. 10, n. 6, art. 475, p. 1-18, 2021. DOI: 10.3390/biology10060475. Disponível em: https://www.mdpi.com/2079-7737/10/6/475. Acesso em: 31 maio 2026.

SENAPATY, Murali Krishna; RAY, Abhishek; PADHY, Neelamadhab. A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms. **Agriculture**, v. 14, art. 1256, 2024. DOI: 10.3390/agriculture14081256. Disponível em: https://www.mdpi.com/2077-0472/14/8/1256. Acesso em: 31 maio 2026.

SONOWAL, Trinayana *et al.* Plant growth-promoting rhizobacteria: influence to abiotic stress tolerance in rice (Oryza sativa L.). **Journal of Applied Biology & Biotechnology**, v. 12, n. 5, p. 41-47, 2024. DOI: 10.7324/JABB.2024.167590. Disponível em: https://jabonline.in. Acesso em: 31 maio 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.