

Análise comparativa de métodos de boosting aplicados à recomendação de cultivos agrícolas

Comparative analysis of boosting methods applied to agricultural crop recommendations

José Airton Azevedo dos Santos¹
Dalila Natalia Costa da Luz²

RESUMO

O emprego de tecnologias preditivas, como algoritmos de *machine learning*, surge como uma alternativa viável e eficiente para orientar escolhas e recomendações de cultivo. Estes algoritmos aprimoram a precisão das decisões agronômicas e mitigam riscos ambientais e econômicos. Neste contexto, este trabalho teve como objetivo comparar o desempenho, na recomendação de culturas agrícolas, dos modelos XGBoost (*eXtreme Gradient Boosting*) e CATBoost (Categorical Boosting) e ADABOOST (*Adaptive Boosting*). Utiliza-se, para realizar esta comparação, uma base de dados com 6 diferentes culturas, 2 parâmetros físicos e 9 nutrientes. Os modelos XGBoost, CATBoost e ADABOOST tiveram seus hiperparâmetros otimizados com o auxílio da biblioteca Optuna e foram avaliados por meio das métricas *Accuracy*, *Kappa* e *f1-score*. Os resultados obtidos indicam que os três modelos apresentaram, na tarefa de recomendação de culturas agrícolas, muito bom desempenho, com valores elevados de acurácia, *Kappa* e *f1-score*.

Palavras-chave: Aprendizado de máquina; XGBoost; CATBoost; ADABOOST; Optuna.

ABSTRACT

The use of predictive technologies, such as machine learning algorithms, emerges as a viable and efficient alternative to guide crop choices and recommendations. These algorithms improve the accuracy of agronomic decisions and mitigate environmental and economic risks. In this context, this work aimed to compare the performance of the XGBoost (*eXtreme Gradient Boosting*) and CATBoost (Categorical Boosting) and ADABOOST (*Adaptive Boosting*) models in crop recommendation. A database with 6 different crops, 2 physical parameters, and 9 nutrients was used for this comparison. The XGBoost, CATBoost, and ADABOOST models had their hyperparameters optimized using the Optuna library and were evaluated using the *Accuracy*, *Kappa*, and *f1-score* metrics. The results indicate that all three models performed very well in the task of crop recommendation, with high accuracy, *Kappa*, and *f1-score* values.

Keywords: Machine learning; XGBoost; CATBoost; ADABOOST; Optuna.

¹ Doutor em Engenharia Elétrica. Docente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: airton@utfpr.edu.br ORCID: <https://orcid.org/0000-0002-2568-5734>

² Discente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: dalilanataliacostadaluz@gmail.com ORCID: <https://orcid.org/0009-0009-4701-0875>

1. INTRODUÇÃO

A agricultura desempenha um papel fundamental no cenário global, sendo essencial tanto para a segurança alimentar quanto para o desenvolvimento econômico de diversas nações. A escolha da cultura a ser desenvolvida está no cerne da gestão de risco e do planejamento de longo prazo na agricultura. Uma decisão acertada resulta em maior produtividade e rentabilidade. Uma decisão equivocada pode levar à perda total do investimento. Neste contexto, técnicas de Aprendizado de Máquina (*Machine Learning* - ML) podem transformar a tomada de decisão agrícola ao fornecer análises preditivas altamente precisas (KWAGHTYO; EKE, 2023; KWAGHTYO et al., 2024; IORZUA et al., 2025).

A utilização destas técnicas, de Aprendizado de Máquina, amplia significativamente a capacidade de análise e de planejamento dos produtores, indo além da simples automação de tarefas. Essas técnicas permitem integrar dados de solo, clima, histórico de produtividade, insumos e manejo, gerando modelos preditivos que apoiam decisões estratégicas, como a seleção das culturas mais adequadas ao contexto produtivo, econômico e ambiental de cada propriedade (MORAIS et al., 2021; AKHTER; SOFI, 2022; IORZUA et al., 2025).

Apesar do avanço recente no uso de algoritmos de *Machine Learning* na agricultura, a literatura ainda é mais consolidada em temas como previsão de produtividade ou detecção de doenças. Enquanto a tarefa de recomendação de culturas agrícolas, a partir de atributos físico-químicos do solo, permanece relativamente pouco explorada. Embora existam estudos aplicando modelos de *boosting* em problemas agrônômicos correlatos, a utilização conjunta e comparativa de XGBoost, AdaBoost e CatBoost especificamente para apoiar a escolha de culturas, com base em nutrientes e parâmetros físicos, ainda é limitada. Assim, o presente estudo, ao comparar o desempenho desses três algoritmos de *boosting*, na recomendação de culturas agrícolas, busca preencher essa lacuna, oferecendo evidências que podem subsidiar tanto a literatura acadêmica quanto aplicações práticas em recomendação de cultivo (MAHESH; SOUNDRAPANDIYAN, 2024; SILVA et al., 2021; SUNDIANTO; CAHYADI, 2025).

Neste contexto, este trabalho teve como objetivo comparar o desempenho, na recomendação de culturas agrícolas, dos modelos XGBoost (*eXtreme Gradient Boosting*) e CATBoost (*Categorical Boosting*) e ADABOOST (*Adaptive Boosting*).

Esta proposta de pesquisa está organizada da seguinte forma: Na seção 2 são apresentados a base e o tratamento dos dados, os algoritmos XGBoost, CATBoost e ADABOOST e as métricas utilizadas na avaliação de desempenho dos modelos. Na Seção 3 são apresentados e analisados os resultados obtidos da aplicação dos três modelos. Comentários finais, na Seção 4, finalizam o artigo.

2. MATERIAIS E MÉTODOS

Nesta seção, apresenta-se uma breve descrição dos algoritmos XGBoost, CATBoost e ADABOOST, a base de dados, bem como as métricas utilizadas.

XGBoost: O XGBoost é um algoritmo de *gradient boosting* de ponta, amplamente aplicado em tarefas de regressão e classificação. Ele integra múltiplas árvores de decisão (modelos base simples) para formar um *ensemble* mais robusto e preciso. Constrói, adotando a estratégia de *boosting*, essas árvores de forma sequencial, com cada nova iteração focada em corrigir os erros (resíduos) da anterior. Além disso, incorpora mecanismos de regularização para mitigar o *overfitting* e elevar a eficiência geral do modelo. O XGBoost, como ferramenta versátil e potente, destaca-se pela capacidade de processar eficientemente grandes volumes de dados, tornando-o ideal para cenários de *Machine Learning* em escala (SOLANO et al., 2022).

AdaBoost: O ADABOOST é um algoritmo de *ensemble learning* que combina múltiplos classificadores fracos (geralmente árvores de decisão simples chamadas "*stumps*") para criar um classificador forte. Ele funciona através de um processo iterativo onde cada novo classificador se concentra nos erros do anterior. Inicialmente todos os dados recebem pesos iguais, mas após cada iteração os exemplos classificados incorretamente têm seus pesos aumentados enquanto os acertados têm pesos reduzidos. O classificador final é uma combinação ponderada de todos os classificadores fracos treinados sequencialmente. Onde cada um contribui, para a decisão final, proporcionalmente à sua acurácia. Isto resulta num modelo robusto que adaptivamente melhora seu desempenho, focando nas instâncias mais difíceis de classificar. Este algoritmo, com relação a outros métodos de *boosting* mais complexos, possui muita boa capacidade preditiva (DING et al., 2022).

CATBoost: O CatBoost é um poderoso algoritmo de *gradient boosting* de código aberto, desenvolvido pelo Yandex, que se baseia na construção sequencial de árvores de decisão.

Ele visa otimizar uma função de perda (*loss function*) adicionando preditores que corrigem os erros das árvores anteriores, de forma similar a outros algoritmos de *boosting*. A principal inovação do CatBoost, mesmo ao focar apenas em variáveis numéricas, reside na técnica de *Gradient Boosting* com *Ordered Boosting* (Ordenado). Neste esquema, o algoritmo utiliza um método de estimação de gradiente sem viés (*unbiased*), onde a estimativa do gradiente, para uma determinada amostra de treinamento, é calculada usando um subconjunto de dados de treinamento diferente do conjunto usado para treinar a árvore atual. Este método mitiga o viés introduzido na estimativa do gradiente e reduz significativamente o risco de superajuste (*overfitting*), que é uma preocupação comum nos algoritmos de *boosting* tradicionais. Além disso, o CatBoost é altamente otimizado para o uso de *hardware* moderno, incluindo suporte para treinamento em GPUs, o que permite o processamento eficiente de grandes conjuntos de dados, e mantém um conjunto de parâmetros padrão robustos que proporcionam um bom desempenho *out-of-the-box* (DOROGUSH et al., 2018).

Base de dados:

A base de dados *Crop recommender dataset with soil nutrition*, utilizada nesta pesquisa, é um conjunto de dados público (*license CCO: Public Domain*) fornecido pelo repositório Kaggle do Google LLC (SANJAY; SRI HARI, 2021). A base de dados é composta por 620 linhas e 12 colunas (9 nutrientes, 2 parâmetros físicos e 6 culturas). Os 9 nutrientes, juntamente com suas unidades de medição, são apresentados na Tabela 1.

Tabela 1. Nutrientes

| Símbolo | Nome do Nutriente | Unidade de Medição |
|---------|--------------------------------|--------------------|
| N | Nitrogênio (<i>Nitrogen</i>) | kg/ha |
| P | Fósforo (<i>Phosphorus</i>) | kg/ha |
| K | Potássio (<i>Potassium</i>) | kg/ha |
| S | Enxofre (<i>Sulfur</i>) | % |
| Cu | Cobre (<i>Copper</i>) | ppm |
| Mn | Manganês (<i>Manganese</i>) | ppm |
| Zn | Zinco (<i>Zinc</i>) | ppm |
| B | Boro (<i>Boron</i>) | ppm |
| Fe | Ferro (<i>Iron</i>) | ppm |

Fonte: Sanjay; Sri Hari (2020).

Os dois parâmetros físicos são: pH (Medida da acidez ou alcalinidade do solo -

na faixa de 0 a 14) e CE (Condutividade Elétrica - dS/m). Os 6 tipos de culturas alvo são: *Pomegranate* (romã), *Mango* (manga), *Grapes* (uva), *Mulberry* (amora), *Ragi* (milheto) e *Potato* (batata). Os nutrientes e os parâmetros físicos são as variáveis de entrada e as culturas a variável alvo (*label*).

Na Figura 1 apresentam-se as cinco primeiras linhas da base de dados *Crop recommender dataset with soil nutrition*.

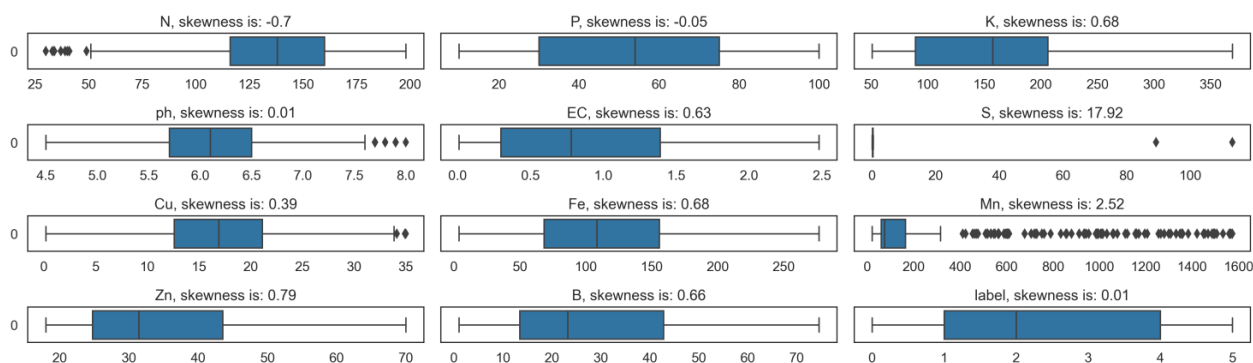
Figura 1. Cinco primeiras linhas da base de dados.

| | N | P | K | ph | EC | S | Cu | Fe | Mn | Zn | B | label |
|---|-----|----|-----|-----|------|------|-------|--------|-------|-------|-------|-------------|
| 0 | 143 | 69 | 217 | 5.9 | 0.58 | 0.23 | 10.20 | 116.35 | 59.96 | 54.85 | 21.29 | pomegranate |
| 1 | 170 | 36 | 216 | 5.9 | 0.15 | 0.28 | 15.69 | 114.20 | 56.87 | 31.28 | 28.62 | pomegranate |
| 2 | 158 | 66 | 219 | 6.8 | 0.34 | 0.20 | 15.29 | 65.87 | 51.81 | 57.12 | 27.59 | pomegranate |
| 3 | 133 | 45 | 207 | 6.4 | 0.94 | 0.21 | 8.48 | 103.10 | 43.81 | 68.50 | 47.29 | pomegranate |
| 4 | 132 | 48 | 218 | 6.7 | 0.54 | 0.19 | 5.59 | 63.40 | 56.40 | 46.71 | 31.04 | pomegranate |

Inicialmente, na etapa de pré-processamento, observou-se que a base de dados não continha valores ausentes, mas possuía *outliers* (Figura 2). Os *outliers*, neste estudo, foram substituídos pelos Limites Inferior e Superior do *boxplot* de cada uma das variáveis. *Outliers* com valores menores que o Limite Inferior foram substituído por este Limite e valores maiores que o limite superior foram ajustados para o valor do limite superior.

Figura 2. Base de dados com *outliers*.

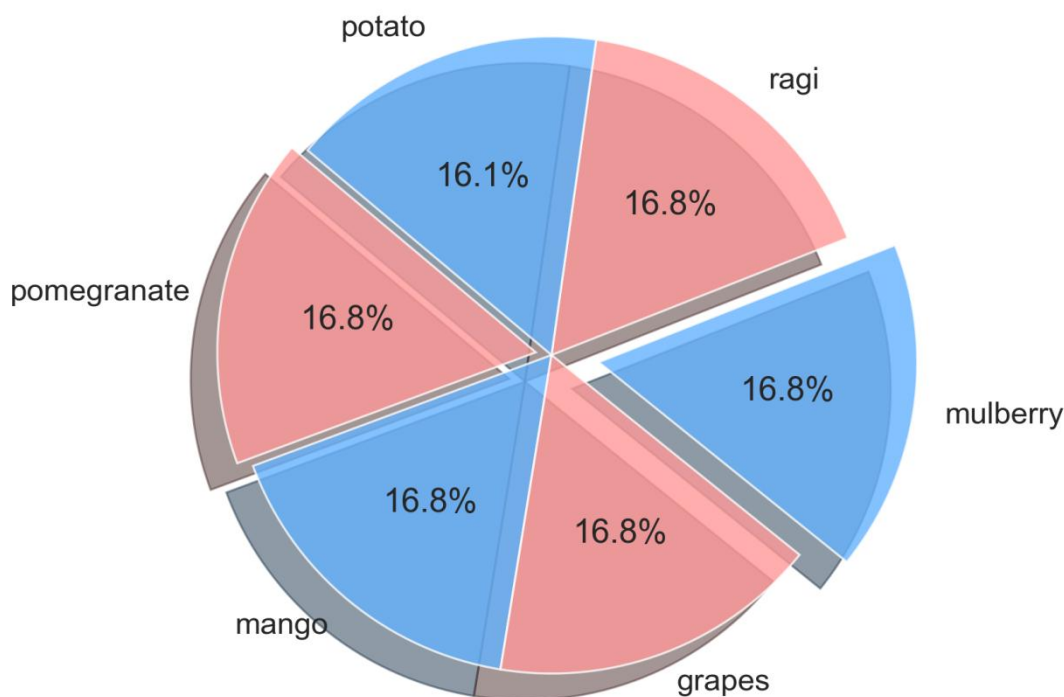
Boxplots for each variable



Na sequência, observou-se que a distribuição por classe é apenas levemente desbalanceada (104 (16,8%) amostras para cinco culturas *Pomegranate*, *Mango*, *Grapes*,

Mulberry e *Ragi*) e 100 (16,1%) para *Potato*) (Figura 3). Portanto, não utilizaremos, neste trabalho, técnicas de balanceamento de classes.

Figura 3. Desbalanceamento entre as classes.



O conjunto de dados, para mensurar a acurácia da classificação, foi estratificado em 80% para treinamento e 20% para teste. Além disso, adotou-se, na etapa de treinamento do modelo, a validação cruzada *K-fold* com $k=5$, segmentando os dados em cinco subconjuntos (*folds*) equivalentes. O modelo, nessa abordagem, é treinado, a cada rodada, em quatro *folds* e avaliado no *fold* remanescente. Isto gera uma métrica de desempenho média que minimiza vieses e maximiza a generalização.

Métricas de desempenho:

Os desempenhos dos modelos foram avaliados pelas seguintes métricas (CHANG et al., 2023; MAULANA et al., 2023; BUDHOLIYA et al., 2022; IMANI et al., 2025):

Estatística de Cohen's Kappa (κ): É uma ferramenta estatística essencial para avaliar o grau de concordância entre dois conjuntos de dados. Esta estatística, desenvolvida por Jacob

Cohen em 1960, é amplamente usada em campos como psicologia, medicina, ciências sociais e *Machine Learning*.

O coeficiente Kappa de Cohen é interpretado através de uma escala de concordância que classifica os valores da seguinte forma: valores abaixo de 0 indicam ausência de concordância; entre 0 e 0,20 a concordância é considerada ruim; de 0,21 a 0,40 é razoável; de 0,41 a 0,60 é moderada; de 0,61 a 0,80 é boa e valores acima de 0,81 indicam concordância muito boa (BASTIANI et al., 2018).

f1-score: O *f1-score* é uma métrica de avaliação de aprendizado de máquina que combina duas outras métricas fundamentais: a Precisão (*Precision*) e a Revocação (*Recall*) (Equação 1). A métrica *Precision* mede a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo. Já *Recall* mede a proporção de casos positivos reais que o modelo conseguiu identificar corretamente. A pontuação *f1* varia de 0 a 1.

$$f1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad 1$$

Accuracy: é uma das métricas mais simples e intuitivas para avaliar o desempenho geral de um modelo. Ela mede a proporção de previsões corretas em relação ao total de previsões feitas (Equação 2).

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \quad 2$$

Onde: TP: Verdadeiro Positivo; TN: Verdadeiro negativo; FP: Falso Positivo e FN: Falso Negativo.

3. RESULTADOS E DISCUSSÃO

Treinamento:

É essencial, para garantir a robustez e o alto desempenho de modelos de previsão, maximizar sua capacidade de generalização, processo que ocorre através da otimização de hiperparâmetros. Neste trabalho, adotou-se o Optuna, um *framework open-source* reconhecido, na área de aprendizado de máquina, pela eficácia na otimização de hiperparâmetros. Utilizando estratégias de amostragem adaptativa, como a Otimização Bayesiana, a ferramenta navega inteligentemente pelo espaço de busca. O Optuna, ao aprender com o histórico das iterações, assegura uma convergência acelerada para a configuração ideal dos parâmetros (LAI et al., 2023).

Na Tabela 2 apresentam-se os parâmetros dos três modelos, os intervalos de busca e os melhores parâmetros encontrados pelo Optuna.

Tabela 2. Hiperparâmetros dos modelos

| Algoritmo | Hiperparâmetro | Intervalo de Busca | Hiperp. Ótimo |
|-----------|-------------------------|---------------------------------------|---------------|
| XGBoost | <i>n_estimators</i> | [50,1000] | 70 |
| | <i>learning_rate</i> | [1e ⁻⁵ ,1e ⁻¹] | 0.031 |
| | <i>max_depth</i> | [1,15] | 2 |
| | <i>min_child_weight</i> | [1,10] | 2 |
| | <i>subsample</i> | [0.0,1.0] | 0.643 |
| | <i>colsample_bytree</i> | [0.0,1.0] | 0.738 |
| | <i>reg_alpha</i> | [0.0,1.0] | 0.088 |
| CATBoost | <i>reg_lambda</i> | [0.0,1.0] | 0.684 |
| | <i>iterations</i> | [0,20] | 16 |
| | <i>depth</i> | [4,10] | 10 |
| | <i>learning_rate</i> | [0.01, 0.2] | 0.026 |
| ADABOOST | <i>l2 leaf reg</i> | [1,10] | 4.92 |
| | <i>n_estimators</i> | [50,500] | 269 |
| | <i>learning_rate</i> | [0.01,2] | 0.126 |
| | <i>algorithm</i> | ['SAMME', 'SAMME.R'] | SAMME |

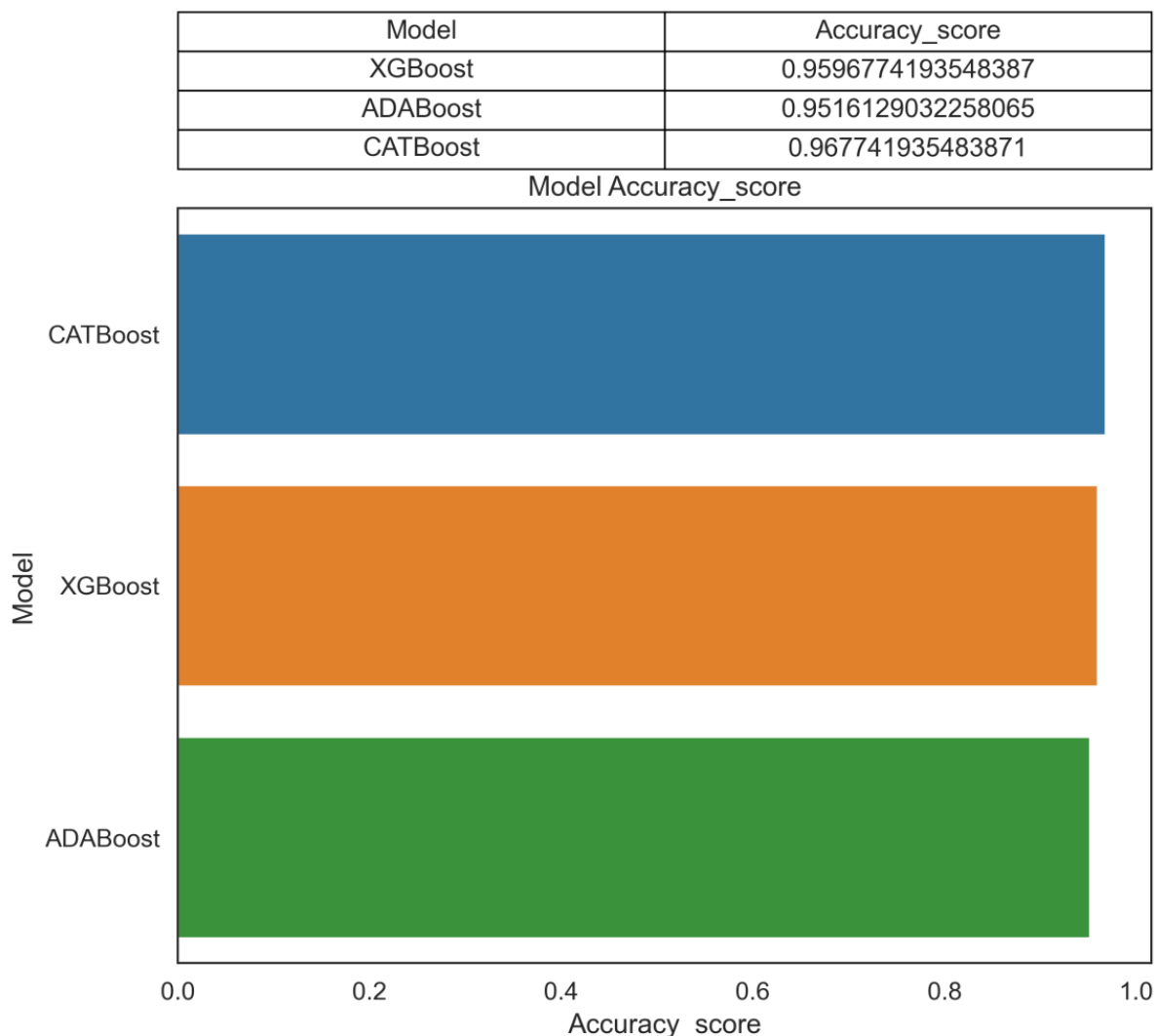
Teste:

Na sequência, detalhamos os resultados de desempenho obtidos, durante a avaliação no conjunto de teste, pelos três modelos (CATBoost, ADABOOST e XGBoost).

Accuracy: Os resultados de acurácia, apresentados na Figura 4, indicam que todos os modelos de *ensemble* testados são muito eficazes para a classificação de culturas

agrícolas. Apresentando desempenhos superior a 95%, destacando-se o CatBoost (96,77%) seguido de XGBoost (95,97%) e AdaBoost (95,16%). Isto sugere que, para a recomendação de culturas agrícolas, o uso de algoritmos de *gradient boosting*, especialmente o CatBoost, tende a fornecer previsões mais precisas e robustas. Bentéjac et al. (2020) também realizaram uma análise comparativa de algoritmos de *gradient boosting*. Concluíram, em seus estudos, que o CatBoost apresentou a melhor acurácia e generalização, embora a diferença para o XGBoost fosse pequena.

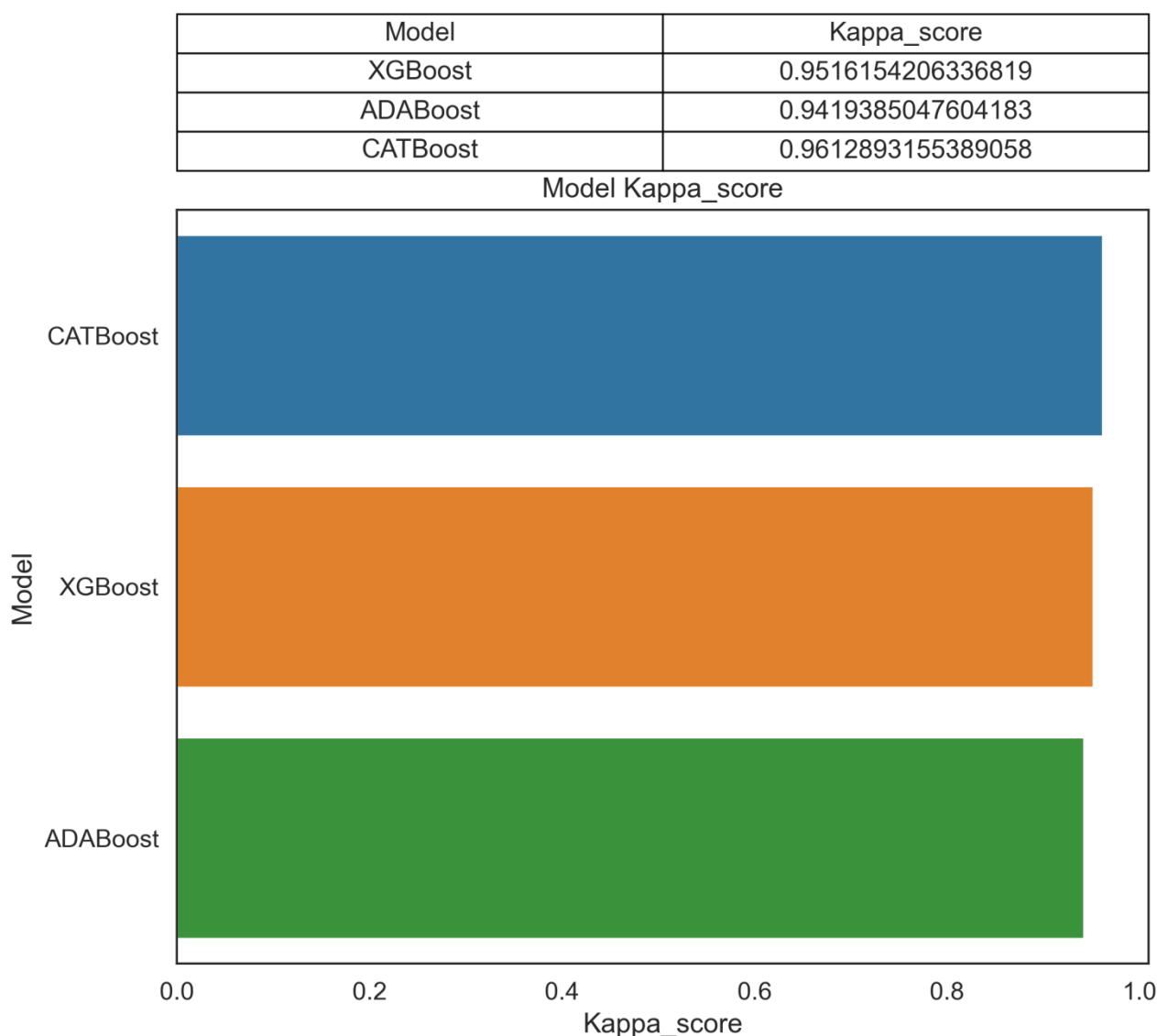
Figura 4. Métrica *accuracy*



Kappa: Na Figura 5 apresentam-se os índices *Kappa* obtidos pelos três modelos. Observe-se, deste gráfico, que os três modelos de *Machine Learning* (XGBoost, AdaBoost e CATBoost) demonstraram, na recomendação de culturas agrícolas, muito bom

desempenho. Visto que todos os seus *kappa_scores* estão acima de 0,94. Isto indica muito boa concordância entre as classificações do modelo e os rótulos verdadeiros. Especificamente, o modelo CATBoost apresentou o melhor desempenho com um *kappa_score* de aproximadamente 0,9613, seguido de perto pelo XGBoost com 0,9516 e por último, o AdaBoost com 0,9419. Isto sugere que todos são adequados para a tarefa, mas o CATBoost e o XGBoost fornecem as classificações mais robustas e confiáveis.

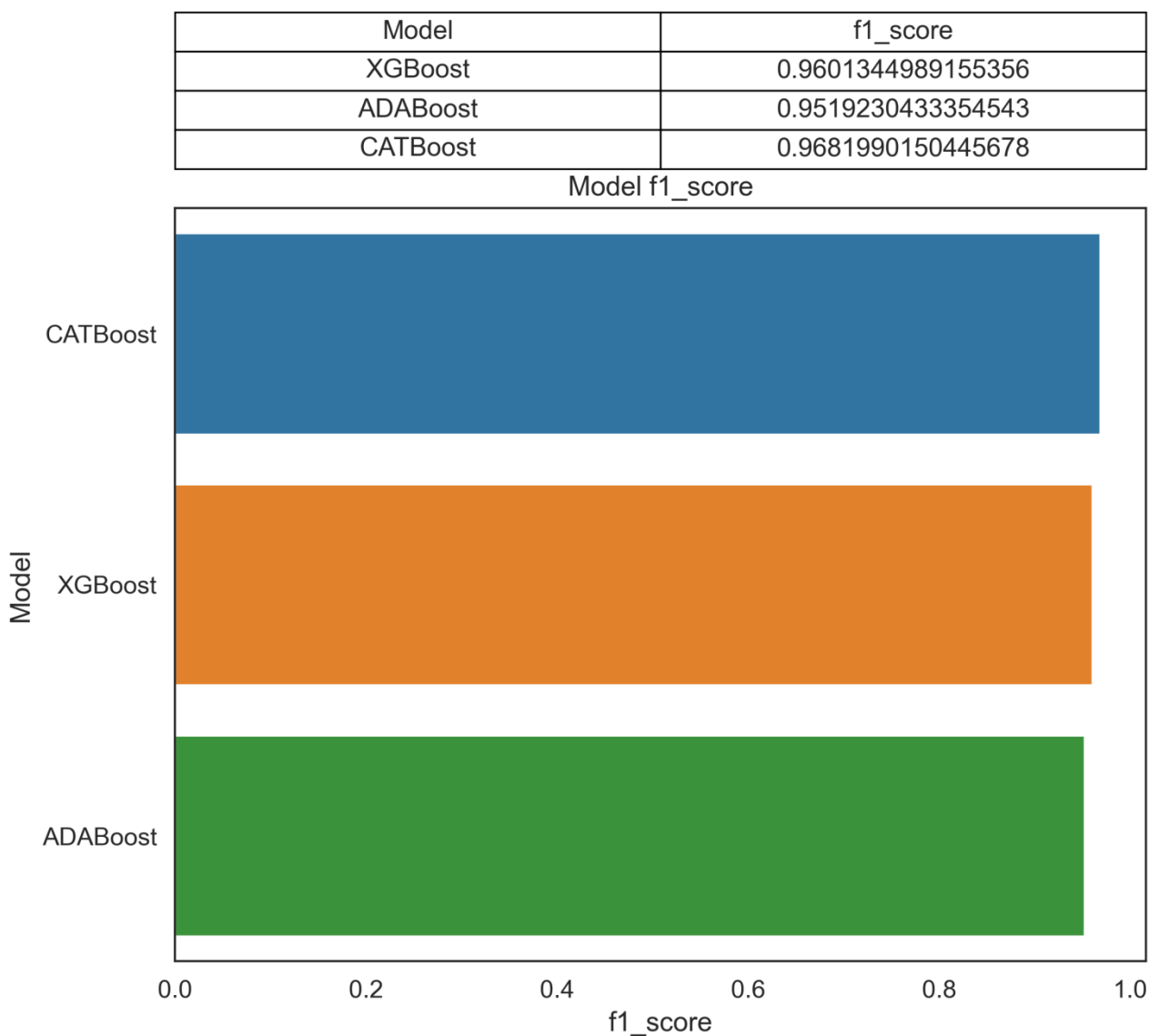
Figura 5. Métrica Kappa



f1-score: Na Figura 6 apresentam-se as pontuações *f1_score* para os três modelos de previsão. Observa-se, por meio dos dados apresentados na figura, que os três modelos de *Machine Learning* (XGBoost, AdaBoost, e CATBoost) exibiram, na classificação das culturas agrícolas, muito bom desempenho, uma vez que todos os seus *f1-scores* estão

acima de 0.95. Este alto valor indica que os modelos são muito bons em identificar corretamente as classes positivas (*precision*) e em encontrar todas as instâncias positivas relevantes (*recall*). O modelo CATBoost se destaca com o melhor desempenho, atingindo um *f1-score* de aproximadamente 0,9682, ligeiramente superior ao XGBoost que obteve 0,9601 e ao AdaBoost com 0.9519. Isto confirma que o CATBoost e o XGBoost não apenas identificam as culturas com alta concordância, mas também mantêm um excelente balanço entre os tipos de erros.

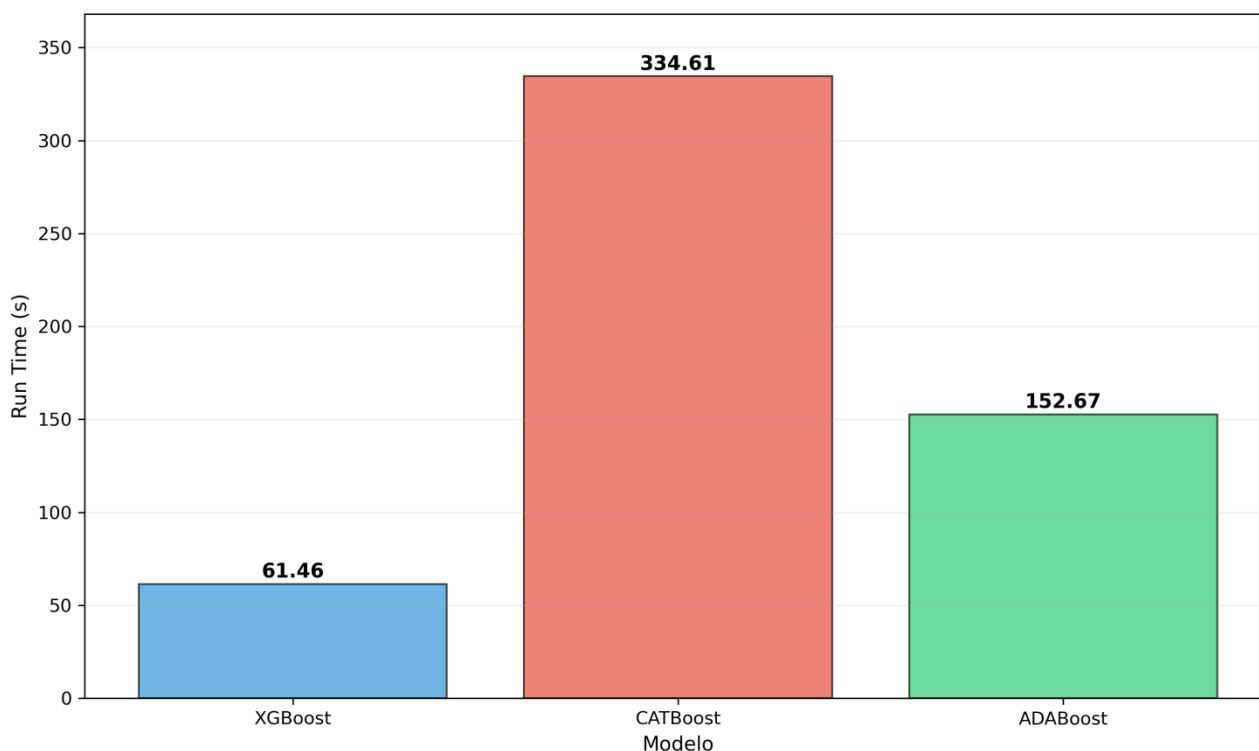
Figura 6. Métrica *f1-score*



Run Time: Os tempos de processamento, apresentados na Figura 7, indicam que o CATBoost apresenta o maior custo computacional (334,611 s), seguido pelo ADABOOST

(152,671 s) e pelo XGBoost (61,462 s), o que revela um *trade-off* claro entre desempenho preditivo e custo de execução. Embora o CATBoost tenha mostrado ligeira vantagem nas métricas *accuracy*, *Kappa* e *f1-score*, esse ganho vem acompanhado de um aumento no tempo de processamento.

Figura 7. Tempos de processamento – XGBoost, CATBoost e ADABOOST



4. CONSIDERAÇÕES FINAIS

Este trabalho cumpriu o objetivo de comparar, na tarefa de recomendação de culturas agrícolas, o desempenho dos modelos de *boosting* XGBoost, CATBoost e ADABOOST. Utilizou-se, nesta comparação de desempenho, uma base de dados composta por nove nutrientes (N, P, K, S, Cu, Mn, Zn, B e Fe), dois parâmetros físicos (pH e CE) e seis tipos de culturas: *Pomegranate* (romã), *Mango* (manga), *Grapes* (uva), *Mulberry* (amora), *Ragi* (milheto) e *Potato* (batata).

Os resultados obtidos demonstram que o emprego de tecnologias preditivas, baseadas em algoritmos de *Machine Learning*, constitui uma alternativa viável e eficiente

para orientar escolhas de cultivo, melhorando a precisão das decisões agrônômicas e mitigando riscos ambientais e econômicos.

Em termos de desempenho, todos os três modelos apresentaram, na recomendação das culturas agrícolas, muito bom desempenho, atingindo valores elevados de *accuracy*, *Kappa* e *f1-score*. Sendo que o modelo CATBoost obteve os melhores resultados: *accuracy* (96,77%), *Kappa* (0,9613) e *f1-score* (0,9682), indicando uma recomendação robusta e confiável.

Contudo, a análise revelou um importante *trade-off* entre desempenho preditivo e custo computacional. Embora o CATBoost tenha tido uma ligeira vantagem nas métricas, ele também apresentou o maior custo computacional (334,611 s), seguido pelo ADABOOST (152,671 s) e, por último, pelo XGBoost (61,462 s).

Em resumo, este estudo, ao comparar de forma conjunta e específica esses algoritmos de *boosting*, preenche uma lacuna na literatura sobre recomendação de culturas. Os resultados subsidiam tanto a pesquisa acadêmica quanto as aplicações práticas, sugerindo que, embora o CATBoost ofereça o desempenho mais preciso para a recomendação, o XGBoost pode representar um balanço ideal entre robustez e eficiência computacional. Recomenda-se, como sugestão para trabalhos futuros, a comparação de desempenho dos três algoritmos para contextos agrícolas mais amplos ou para bases de dados com maior variedade de culturas ou características de solo.

REFERÊNCIAS

- AKHTER, R.; SOFI S. A. Precision agriculture using IoT data analytics and machine learning. **Journal of King Saud University - Computer and Information Sciences**, v. 34, 2022.
- BASTIANI, M.; SANTOS, J. A. A.; SCHMIDT, C. A. P.; SEPULVEDA, G. P. L. Aplicação de algoritmos de mineração de dados na gestão da produção de frangos de corte. **Revista Geintec**. v. 8, 2018.
- BENTÉJAC, C.; CSORGO, A.; MATINEZ-MUNÓZ A comparative analysis of gradient boosting algorithms. **Artificial Intelligence Review**, v. 54, 2020.
- BUDHOLIYA, K.; KUMAR, S.; VIVEK, S. An optimized XGBoost based diagnostic system for effective prediction of heart disease. **Journal of King Saud University - Computer and Information Sciences** Sharma. v. 34, 2022.
- CHANG, V.; BAILEY, J.; XU, Q. A.; SUN, Z. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. **Neural Computing and Applications**, 2023.

DING, Y.; ZHU, H.; CHEN, R.; LI, R. An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification. **Applied Science**, v. 12, 2022.

DOROGUSH, A. V.; ERSHOV, V.; GULIN, A. CatBoost: gradient boosting with categorical features support. **arXiv**, 2018.

IMANI, M.; BEIKMOHAMMADI, A.; ARABNIA, H. R. Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. **Technologies**, v. 13, 2025.

IORZUA, J. T.; KWAGHTYO, D. K. HULE, T. P. IBRAHIM, A. T.; NONGU, A. D. AI-Driven Approach to Crop Recommendation: Tackling Class Imbalance and Feature Selection in Precision Agriculture. **Journal of Future Artificial Intelligence and Technologies**. v. 2, 2025.

KWAGHTYO, D. K.; EKE, C. I. Smart farming prediction models for precision agriculture: a comprehensive survey. **Artificial Intelligence Review**, v. 56, 2023.

KWAGHTYO, D. K.; EKE, C. I.; ABAH, J.; MOSES, T. **Soil and Climate Parameters Based Crop Recommendation Model for Yandev, Gboko Local Government Area of Benue State**. In: 18th International Conference on Ubiquitous Information Management and Communication (IMCOM), Kuala Lumpur, Malasia, 2024.

LAI, J. P.; LIN, Y. L.; LIN, H. C.; SHIH, C. Y., WANG, Y. P. PAI, P. F. Tree-Based Machine Learning Models with Optuna in Predicting Impedance Values for Circuit Analysis. **Micromachines**, v. 14, n. 2, 2023.

MAHESH, P.; SOUNDRAPANDIYAN, R. Yield prediction for crops by gradient-based Algorithms. **Plos ONE**, v. 19, 2024.

MAULANA, A.; FAISAL, F. F. NOVIANDY, T. R.; RIZKIA, T.; IDROES, G. M.; IDROES, G. M.; TALLEI, T. E. Machine learning approach for diabetes detection using finite tuned XGBoost algorithm. **Infolitika Journal of Data Science**, 2023.

MORAIS, R.; MENDES, J.; SILVA, R.; SILVA, N.; SOUZA, J. J. PERES, E. A Versatile, Low-Power and Low-Cost IoT Device for Field Data Gathering in Precision Agriculture Practices. **Agriculture**, v. 11, 2021.

SANJAY, M. V.; SRI HARI K. R. (2021) **Crop Recommender Dataset with Soil Nutrients**. Kaggle. 2021. Disponível em: <<https://www.kaggle.com/dsv/2397200>>. Acesso em: 16 fev. 2025.

SILVA, V. C.; ROCHA, M. S.; FARIA, G. A.; ALVES, S. F.; JUNIOR, A. X.; OLIVEIRA, T. A.; BASTOS, A. P. Boosting algorithms for prediction in agriculture: an application of feature importance and feature selection boosting algorithms for prediction crop damage. **AgriRxiv**, 2021.

SOLANO, E. S.; DEGHANIAN, P.; AFFONSO, C. M. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. **Energies**, v. 15, 2022.

SUNDIANTO, S.; CAHYADI, E. F. A Soil Nutrient Assessment for Crop Recommendation Using Ensemble Learning and Remote Sensing. **I. J. Intelligent Systems and Applications**, v. 3, 2025.