

## UM SISTEMA DE RECOMENDAÇÃO BASEADO EM MACHINE LEARNING PARA SELEÇÃO DE VARIEDADES DE CAFÉ

EDUARDO RODRIGUES PINTO<sup>1</sup>, LANDIR SAVINIEC<sup>2</sup>

<sup>1</sup> Graduando em Engenharia Agrícola, Universidade Federal do Paraná, eduardorodrg.ea@gmail.com

<sup>2</sup> Doutor em Matemática Computacional, Universidade Federal do Paraná, landir.saviniiec@ufpr.br

Apresentado no  
LII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2023  
18 a 21 de outubro de 2023 – Ribeirão Preto - SP, Brasil

**RESUMO:** A necessidade de produzir alimentos em grande escala e de forma sustentável tem sido um motor de impulsionamento para o rápido avanço da agricultura de precisão. Sendo assim, técnicas de Machine Learning para extração de conhecimento em dados agrícolas têm ganhado cada vez mais aplicações na agricultura. Algoritmos supervisionados e não-supervisionados identificam padrões (agrupamento/classificação) ou preveem acontecimentos (previsão) baseados em dados. O algoritmo K-Means pode ser utilizado para agrupar dados e os algoritmos Random Forest e KNN para previsão e classificação. Este trabalho desenvolveu um sistema de recomendação baseado nesses algoritmos de Machine Learning para ajudar produtores de café a escolher as variedades adequadas para o plantio. Foi utilizada uma base de dados com 49 variedades de café, mapeando 17 características das plantas. O algoritmo K-Means foi utilizado para agrupar as variedades por similaridade e detectou 8 grupos diferentes. Através de uma interface simples, o produtor informa as características desejadas e invoca os algoritmos KNN ou Random Forest, fazendo a recomendação. Como resposta, o programa retorna o grupo de variedades de café mais similar as características desejadas

**PALAVRAS-CHAVE:** cafeicultura, k-means, machine learning.

### A RECOMMENDATION SYSTEM BASED ON MACHINE LEARNING FOR SELECTION OF COFFEE VARIETIES

**ABSTRACT:** The need to produce food on large scales and in a sustainable manner has led to rapid advances in the field of precision agriculture. Therefore, Machine Learning techniques for extracting knowledge from agricultural data have been gaining more and more applications in agriculture. Supervised and unsupervised algorithms identify patterns (grouping or classification) or predict events (prediction) based on data. The K-Means algorithm can be used for data clustering and the Random Forest and KNN algorithms for prediction and classification. This research developed a recommendation system based on these Machine Learning algorithms to help coffee farmers in choosing suitable varieties for planting. A database with 49 coffee varieties was used, mapping 17 plant characteristics. The K-Means algorithm was used to group the varieties by similarity and it detected 8 different groups. Through a simple interface, the farmer inputs the desired characteristics and invokes the KNN or Random Forest algorithms to make the recommendation. In response, the program returns the group of coffee varieties most similar to the desired characteristics.

**KEYWORDS:** coffee farming, k-means, machine learning.

**INTRODUÇÃO:** O uso de dados vem sendo crescente para solucionar problemas nas mais diversas áreas, principalmente utilizando ferramentas de Machine Learning (SOUZA, 2019). Segundo Fernandes (2021), algoritmos supervisionados aprendem com dados rotulados e geram resultados esperados, podendo ser usados para classificação ou regressão. Algoritmos não-supervisionados, por outro lado, não têm variável resposta e buscam agrupar dados similares ou reduzir sua dimensionalidade. Na cafeicultura, a escolha de qual variedade/cultivar plantar tem grande peso no futuro da lavoura, devido cada uma ter suas próprias particularidades de produção, necessidade de nutrientes, temperaturas, relevos, resistências a pragas e doenças, e entre outros aspectos que podem variar de local para local. Neste trabalho foi investigado a possibilidade de desenvolver um sistema de recomendação, baseado em algoritmos de Machine Learning, para sugerir ao produtor o cultivar de café que melhor se ajusta as características de sua propriedade. Isto é, através da uma interface o produtor informa as características desejadas e o programa retorna o grupo de variedades de café mais similar as características desejadas. Foram utilizados os algoritmos K-Means (supervisionado) para realizar o agrupamento dos dados e os algoritmos K-Nearest Neighbors (KNN) e Random Forest (RF) para realizar a classificação/recomendação.

**MATERIAIS E MÉTODOS:** O sistema de recomendação foi desenvolvido em R, uma linguagem dinâmica e multifuncional para análise de dados. O RStudio (2022) é um software gratuito que permite criar algoritmos em R, manipular dados e gerar gráficos. Como base de dados, foi usado um projeto de Souza et al. (2004), que divulgaram características de diferentes cultivares de café, como cores de folhas e frutos, vigor, tamanho de folhas, copa, diâmetro, semente, a produtividade, resistência a pragas, entre outras características. Porém, os dados estavam em formato qualitativo. Para serem utilizados pelos algoritmos, os dados passaram por um pré-processamento, trocando os valores qualitativos por quantitativos. A Tabela 1 apresenta uma amostra do resultado dessa conversão.

Tabela 1. Dados qualitativos convertidos para quantitativos.

Cultivar	Porte (altura da planta)	Copa	Diâmetro da copa	Comprimento do internódio
Acaía	4	3	2	3
Acaia Cerrado Mg 1474	4	3	2	3
Acauã	2	2	2	1
Acauã Novo	2	2	2	1
Araponga Mg1	1	1	2	1
Arara	1	2	2	1
Asa Branca	2	2	2	1
Catiguá Mg1	1	1	2	1
Legenda	1 - Baixo	1 - Cônico	1 - Pequeno	1 - Curto
	2 - Baixo a Médio	2 - Cilíndrica	2 - Médio	2 - Médio
	3 - Médio	3 - Cônica ou cilíndrica	3 - Grande	3 - Longo
	4 - Alto	4 - Cônica e compacta	4 - Médio a grande	

O K-Means foi o primeiro algoritmo a ser utilizado. Ele agrupou as cultivares por similaridade de características. Sendo assim, quando um novo vetor de características de café é informado ao programa (características desejadas pelo produtor), a interface invoca os

algoritmos classificadores KNN e RF para determinar qual grupo de cultivares é mais similar as características desejadas, usando como rótulos, os grupos indicados pelo K-Means. Com isso, é feita uma filtragem no conjunto de dados e são retornados as cultivares pertencentes ao grupo indicado pelos algoritmos classificadores. Para criar uma interface que facilite o contato entre o usuário e o algoritmo de recomendação de cultivares de café, usou-se a biblioteca Shiny, presente na Linguagem R, que permite adicionar *inputs* e *outputs* como em um aplicativo. Assim, gerou-se uma página web onde o usuário insere as características do café que deseja, formando o vetor que será usado no algoritmo de classificação. O usuário pode escolher entre os dois algoritmos de classificação e após a seleção, o algoritmo retorna o conjunto de cultivares mais parecido com os dados informados pelo usuário. Em outra aba, o aplicativo mostra as características de cada cultivar do grupo retornado, para que o usuário possa decidir qual implantar em sua propriedade.

**RESULTADOS E DISCUSSÃO:** O algoritmo K-Means requer a definição da quantidade ideal de grupos para separar os dados, para isso, empregou-se o método do cotovelo, que testa o algoritmo K-Means com várias quantidades de clusters e mostra qual delas tem o melhor equilíbrio entre as distâncias euclidianas. Observou-se que a quantidade de 8 clusters é a que apresenta o menor erro, sendo escolhida para a próxima etapa, onde se executou novamente o K-Means para agrupar a base de dados em 8 grupos. Após o agrupamento, a informação de grupo foi adicionada a cada cultivar na base de dados, que será usada como variável “rótulo” para os algoritmos de classificação.

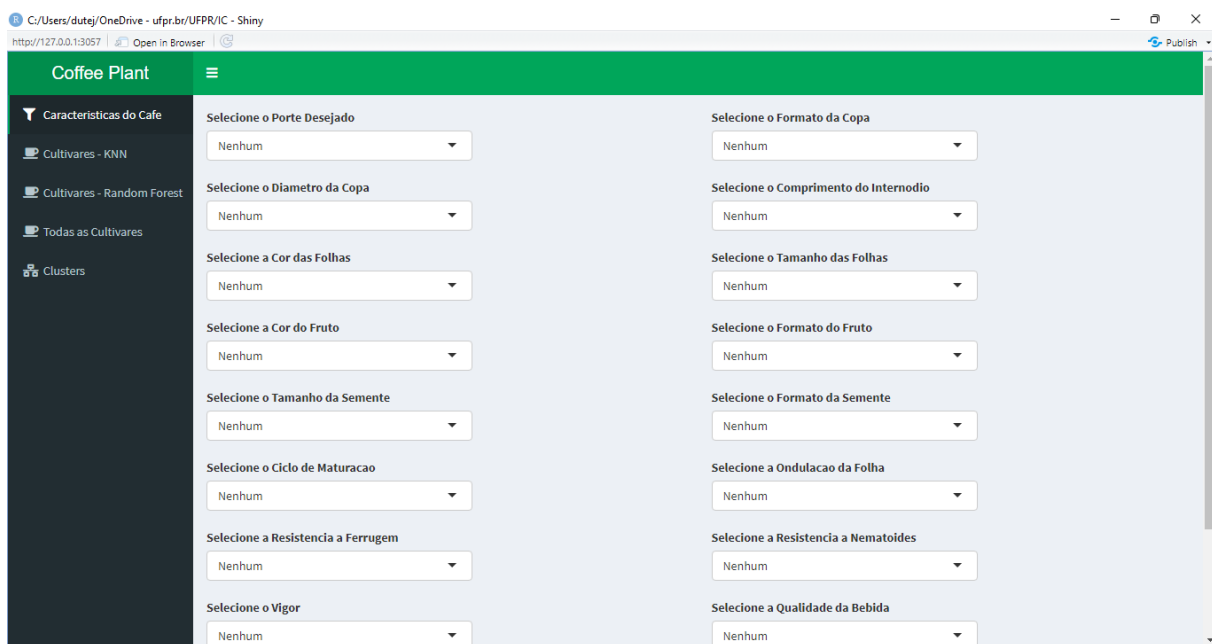


FIGURA 1. Interface do aplicativo mostrando as informações da primeira aba.

A Figura 1 apresenta a interface final do aplicativo. A interface está dividida em 5 abas, mostradas no lado esquerdo da tela, sendo elas: “Características do Café”, onde o usuário insere as características do café que ele deseja implantar na sua propriedade rural e escolhe um classificador (KNN ou RF) para fazer a recomendação; “Cultivares – KNN” e “Cultivares – Random Forest”, apresentam o resultado da classificação feita pelos algoritmos KNN e Random Forest, respectivamente, o usuário é encaminhado para estas abas logo após clicar no tipo de classificador que ele deseja executar; “Todas as Cultivares”, mostra toda a base de

dados usada no programa, contendo todas as cultivares e suas características; “Clusters”, exibe de forma gráfica os grupos criados pelo algoritmo K-Means.

**CONCLUSÕES:** O sistema de recomendação desenvolvido em Linguagem R funcionou conforme o esperado. A interface com o usuário é simples e intuitiva, podendo ser usada por produtores rurais. Um desafio para a performance do algoritmo é a escassez de dados sobre as cultivares do café. Uma base de dados mais ampla e diversificada poderia melhorar a classificação e a recomendação. Além disso, o aplicativo também pode ser adaptado para outras culturas agrícolas, bastando trocar os dados cafeeiros por outros de interesse.

#### **REFERÊNCIAS:**

FERNANDES, F. T. **Machine Learning em saúde e segurança do trabalhador: perspectivas, desafios e aplicações.** 2021.

RSTUDIO. **RStudio: ambiente integrado de desenvolvimento para R.** RStudio, PBC, 2022. Disponível em: <<https://www.rstudio.com/>>.

SOUSA, M. C. C. **Uma Análise Do Algoritmo K-Means Como Introdução Ao Aprendizado De Máquinas.** Sistema de Bibliotecas da Universidade Federal do Tocantins, v. 126, n. 1, 2019.

SOUZA, F. DE F.; SANTOS, J. C. F.; COSTA, J. N. M.; SANTOS, M. M. DOS. **Características das principais variedades de café cultivadas em Rondônia.** 2004. Porto Velho, RO: Embrapa Rondônia.