

**INTELIGÊNCIA ARTIFICIAL NA PLATAFORMA OPENSOILS:
IMPLEMENTAÇÃO DE ALGORITMOS DE CLASSIFICAÇÃO PREDITIVA
PARA ANÁLISE DOS DADOS DO SOLO**

Caroline Silva De Oliveira (caroline@pet-si.ufrj.br)

Sergio Manuel Serra Da Cruz (sergioserra@gmail.com)

O solo é um recurso natural limitado e de suma importância para a vida, o que demanda pesquisas contínuas para otimizar seu uso e minimizar perdas. Nesse contexto, a plataforma OpenSoils, lançada em 2018, emerge como uma iniciativa crucial ao reunir dados de solos brasileiros, alinhada com princípios internacionais de Open Science, e-Science, Data Science e os Princípios FAIR. Este trabalho destaca a necessidade de aproveitar os recursos do solo, especialmente para pequenos produtores rurais que frequentemente enfrentam a limitação de dados sobre suas propriedades. O principal objetivo do projeto foi desenvolver um algoritmo de análise preditiva dos dados de solo da plataforma OpenSoils, utilizando técnicas de aprendizado de máquina. A pesquisa visa demonstrar que, mesmo com um conjunto de dados reduzido, é possível tomar decisões assertivas, otimizando a produção rural e reduzindo significativamente o desperdício. A contribuição direta é a melhoria na classificação de amostras de solo, o que, indiretamente, prolonga a vida útil da terra, garantindo que os produtos agrícolas plantados sejam os mais adequados para o tipo de solo. A metodologia, de caráter experimental, propôs uma melhoria no processo de análise de dados de solo, fornecendo análises mais assertivas em pequenas amostras de dados. O dataset utilizado foi

proveniente da Fazenda agroecológica Km47, disponível na plataforma OpenSoils sob a sigla SIPA47, contendo informações sobre as propriedades físicas e químicas do solo. Para a análise preditiva, foram aplicados dois algoritmos de aprendizado de máquina supervisionado: Árvore de Decisão e Support Vector Machine (SVM). O projeto foi desenvolvido na linguagem de programação Python e no ambiente Google Colab, com a avaliação do desempenho dos modelos baseada em métricas como acurácia, precisão, recall e F1-Score. Os experimentos buscaram identificar qual das técnicas utilizadas apresentaria o melhor resultado ao lidar com datasets limitados. O desempenho de ambos os modelos foi avaliado usando métricas como acurácia, precisão, recall e F1-Score, com o dataset dividido em diferentes proporções entre dados de treinamento e teste (80/20 a 60/40). Os resultados demonstram que, embora a utilização de modelos de aprendizado de máquina em datasets limitados seja possível, ela impõe desafios significativos de generalização. O modelo de Árvore de Decisão apresentou uma acurácia ligeiramente superior em certas configurações, mas com precisão e F1-Score consistentemente baixos, o que indica uma dificuldade em identificar corretamente a classe positiva e minimizar falsos positivos. Em contraste, o modelo SVM, apesar de uma acurácia geral um pouco inferior, mostrou-se mais promissor em termos de precisão, alcançando 0,667 em uma divisão de 40% dos dados para teste, revelando maior potencial para fazer previsões positivas confiáveis. Desta forma conclui-se que a aplicação de algoritmos de classificação preditiva em datasets de solo limitados é viável e eficaz. O estudo aponta que o desempenho dos modelos foi impactado pela escassez de dados, o que foi a principal limitação do projeto, junto à falta de variação de parâmetros, como o kernel do modelo SVM. Para trabalhos futuros, sugere-se a investigação de outras técnicas de aprendizado de máquina, como algoritmos de ensemble (ex: Random Forests ou Gradient Boosting), que podem apresentar maior robustez em cenários de dados limitados e alcançar um desempenho mais confiável e generalizável. Além disso, um estudo mais aprofundado sobre a relação entre as métricas de desempenho no modelo SVM pode revelar insights importantes.

Palavras-chave: aprendizado de máquina; classificação preditiva; opensoils:.