

FACULDADE DE TECNOLOGIA DE SÃO PAULO

**CURSO SUPERIOR DE TECNOLOGIA EM DESENVOLVIMENTO DE
SOFTWARE MULTIPLATAFORMA**

ANÁLISE DE ADEQUAÇÃO DE SOLO PARA CULTIVO DE MANDIOCA:

Comparativo entre Perceptron e K-Nearest Neighbors (KNN)

Desenvolvido por:

João Pedro Dias Barreto
Leonardo Wicher Lopes Ferreira

Relatório Técnico - AT2 - Técnicas de Aprendizagem de Máquina

São Paulo
December de 2025

RESUMO

Este trabalho apresenta a implementação e análise comparativa de dois algoritmos de aprendizado de máquina, Perceptron e K-Nearest Neighbors (KNN), para a classificação da adequação de solo no cultivo de mandioca (*Manihot esculenta*). Utilizando uma base de dados com parâmetros edáficos (pH, potássio, drenagem e compactação), o algoritmo KNN foi avaliado com diferentes valores de K (3, 5 e 7). Os resultados demonstraram que o KNN obteve 100% de acurácia em todas as configurações testadas, superando ou igualando a performance do Perceptron (97,78%). O estudo confirma a eficácia de algoritmos baseados em instância para conjuntos de dados de pequeno porte e alta dimensionalidade relativa.

Palavras-chave: KNN, Perceptron, Classificação de Solo, Mandioca, Aprendizado de Máquina.

1. INTRODUÇÃO

A mandioca é uma cultura vital para a segurança alimentar global. A determinação correta da aptidão do solo é crucial para maximizar a produtividade. Enquanto modelos lineares como o Perceptron oferecem simplicidade e interpretabilidade, algoritmos não-paramétricos como o K-Nearest Neighbors (KNN) podem capturar relações mais complexas entre os dados sem assumir uma distribuição específica ou separabilidade linear.

Este relatório técnico (AT2) expande o trabalho anterior focando na implementação do algoritmo KNN, variando o hiperparâmetro K, e comparando sua eficácia com o modelo Perceptron previamente desenvolvido.

2. METODOLOGIA

2.1 Algoritmo K-Nearest Neighbors (KNN)

O KNN é um algoritmo de aprendizado supervisionado baseado em instâncias (lazy learning). A classificação de um novo exemplo é realizada através de uma votação majoritária entre os seus K vizinhos mais próximos no espaço de características.

2.2 Métrica de Distância

Foi utilizada a Distância Euclidiana para calcular a proximidade entre as amostras:

$$d(p, q) = \sqrt{\sum (p_i - q_i)^2}$$

2.3 Configuração Experimental

- **Pré-processamento:** Normalização Min-Max para atributos contínuos (pH e Potássio) e codificação binária para categóricos.
- **Valores de K testados:** 3, 5 e 7.
- **Divisão dos Dados:** 30 exemplos para treinamento e 15 para teste (mesma divisão do estudo anterior para comparabilidade).
- **Base de Dados:** 45 registros contendo pH, Potássio, Drenagem, Compactação e a Classe (Apto/Não Apto).

3. RESULTADOS E DISCUSSÃO

Tabela 1 - Performance do KNN com diferentes valores de K

Valor de K	Acurácia	Acertos (de 15)	Observações
K = 3	100,00%	15	Excelente estabilidade
K = 5	100,00%	15	Excelente estabilidade
K = 7	100,00%	15	Excelente estabilidade

Tabela 2 - Comparativo Perceptron vs KNN

Modelo	Melhor Acurácia	Características
Perceptron	97,78%	Modelo Linear, Rápido Treinamento
KNN (K=3,5,7)	100,00%	Não-Linear, Sem Treinamento (Lazy)

3.1 Análise dos Resultados

O algoritmo KNN demonstrou desempenho superior, atingindo 100% de acurácia em todas as configurações de K testadas. Isso sugere que as classes são bem separadas no espaço vetorial multidimensional, mas também que a estrutura local dos dados (vizinhança) é altamente preditiva.

Ao contrário do Perceptron, que busca um hiperplano separador global, o KNN toma decisões baseadas na similaridade local. A robustez do KNN neste dataset indica que amostras com características de solo similares (pH, potássio, etc.) tendem fortemente a ter a mesma classificação de aptidão, o que é consistente com a teoria agronômica.

4. CONCLUSÃO

A implementação do algoritmo KNN para a análise de solo de mandioca provou-se extremamente eficaz. A variação de K (3, 5, 7) não alterou o resultado final de 100% de acurácia no conjunto de teste, demonstrando a estabilidade do método para este problema específico.

Comparado ao Perceptron, o KNN mostrou-se ligeiramente superior em acurácia bruta. No entanto, ambos os modelos são válidos para a tarefa. O KNN é recomendado quando se dispõe de poder computacional para armazenar e consultar a base de dados em tempo real, enquanto o Perceptron é ideal para sistemas embarcados com memória limitada, pois armazena apenas os pesos.

Este trabalho cumpre os requisitos da AT2, demonstrando a aplicação prática de técnicas de aprendizado de máquina na agricultura de precisão.

REFERÊNCIAS

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21-27, 1967.

FIX, E.; HODGES, J. L. Discriminatory analysis: nonparametric discrimination: consistency properties. **USAF School of Aviation Medicine**, Randolph Field, TX, 1951.

COCK, J. H. **Cassava: new potential for a neglected crop**. Boulder: Westview Press, 1985.