

Relatório Comparativo: KNN Hardcore vs. Sklearn

Este relatório apresenta uma comparação da métrica de acurácia entre uma implementação do zero (manual) do algoritmo K-Nearest Neighbors (KNN) e a versão otimizada disponível na Sklearn.

Metodologia

O estudo utilizou o conjunto de dados Iris, um clássico em problemas de classificação de aprendizado de máquina, que consiste em 150 amostras de flores de íris divididas em 3 classes. O conjunto foi dividido em dados de treino (80%) e teste (20%) para avaliar o desempenho de ambos os modelos.

- **Implementação Manual:** Um classificador KNN foi construído do zero, usando a distância euclidiana para determinar os vizinhos mais próximos. Para este teste, o valor de k (número de vizinhos) foi definido como 1, 3, 5 e 7.
- **Implementação Sklearn:** O classificador KNeighborClassifier da biblioteca sklearn foi utilizado com os mesmos parâmetros (k=1, 3, 5 e 7) para garantir uma comparação justa.

Ambos os modelos foram treinados com os mesmos dados e avaliados no mesmo conjunto de teste. As métricas de acurácia, precisão e revocação foram calculadas para cada modelo.

Resultados

A tabela a seguir resume os resultados de desempenho obtidos:

	k = 1	k = 3	k = 5	k = 7
Manual	100%	96,67%	96,67%	100%
Sklearn	100%	96,67%	96,67%	100%

Análise e Conclusão

Conforme mostrado na tabela, as implementações obtiveram accuracies idênticas para todos os valores de K.

Este resultado indica que, para um problema relativamente simples e limpo como o conjunto de dados Iris, uma implementação básica do KNN consegue um desempenho perfeito, idêntico ao da biblioteca. Isso demonstra que a lógica fundamental do algoritmo é robusta e eficaz.

A principal vantagem da versão do sklearn não reside na sua acurácia básica em um cenário ideal, mas em sua otimização para performance, sua capacidade de lidar com grandes volumes de dados e sua interface padronizada, que permite fácil integração com outras ferramentas de pré-processamento e avaliação. Para a maioria das aplicações práticas, a versão sklearn é a escolha recomendada devido à sua eficiência e recursos adicionais.