

Resultados e Insights Relevantes

1. Desempenho dos Modelos

a. K-Nearest Neighbors (KNN):

- **Desempenho:** O modelo apresentou boa acurácia geral, especialmente em classes com características bem distintas.
- **Pontos Fortes:**
 - Fácil de interpretar e implementado com bons resultados em dados balanceados.
 - Não faz suposições sobre a distribuição dos dados.
- **Limitações:**
 - Menor eficiência em classes que têm grande sobreposição de características.
 - Sensível ao número de vizinhos (*n_neighbors*), o que exige cuidado na otimização.

b. Support Vector Machine (SVM):

- **Desempenho:** O modelo SVM apresentou alta precisão e F1-score, especialmente ao utilizar o kernel RBF (otimizado).
- **Pontos Fortes:**
 - Excelente para dados com separação clara no espaço de características.
 - Resultados consistentes em identificar classes distintas.
- **Limitações:**
 - Tempo de treinamento maior, especialmente com grandes volumes de dados.
 - Depende fortemente da escolha do kernel e de parâmetros como C e gamma.

c. Random Forest:

- **Desempenho:** Consistentemente, foi o modelo com melhor desempenho geral, destacando-se pela robustez e capacidade de generalização.
- **Pontos Fortes:**
 - Modelo robusto a dados ruidosos.
 - Possibilidade de interpretar a importância dos atributos.
- **Limitações:**
 - Modelo mais complexo, que pode ser menos eficiente em tempo de predição.

2. Insights Extraídos

a. Importância das Características:

- A análise da importância dos atributos no Random Forest indicou que **área, perímetro e comprimento do sulco** são as características mais relevantes para distinguir variedades de grãos.
- Isso sugere que características relacionadas ao tamanho e morfologia do grão têm um impacto direto na classificação.

b. Dificuldades de Classificação:

- Os erros de classificação observados nas matrizes de confusão ocorreram principalmente entre as classes **1 e 2**.

Isso pode indicar que essas variedades têm características físicas muito próximas, exigindo maior precisão nos sensores ou atributos adicionais para melhorar a separação.

c. Aplicações no Contexto Agrícola:

- Os modelos são úteis para separar automaticamente variedades de trigo no pós-colheita, reduzindo a necessidade de intervenção humana.
- Podem ser utilizados em sistemas agrícolas automatizados para prever a variedade com base nas características do grão, ajudando em decisões como escolha de sementes e ajustes nos processos de plantio e irrigação.

3. Conclusões Significativas

- a. O **Random Forest** foi o modelo mais adequado, equilibrando precisão, robustez e interpretabilidade. Ele pode ser adotado como solução primária para o problema.
- b. Características físicas do grão, especialmente tamanho e forma, são determinantes para identificar variedades, alinhando-se à prática agrícola.
- c. Para classes com maior sobreposição, adicionar novos atributos ou melhorar a qualidade das medições (sensores) pode aumentar a precisão.
- d. O uso desses modelos automatiza a classificação de grãos, economizando tempo e recursos, e contribui para práticas agrícolas mais eficientes e sustentáveis.