

# Relatório do Trabalho da Disciplina de IA 1

## Reconhecimento de Feijões Pretos Bons e Ruins

Mateus Bazan Bespalhok

Matheus Christo Wegner

### Introdução

A seleção de feijões pretos é um trabalho demorado de ser realizado manualmente e que aumenta o valor agregado do material final. A automatização desse processo é de grande interesse para produtores e vendedores de feijão para que possam aumentar sua lucratividade de forma simples e rápida. A utilização de modelos de tomada de decisão (comumente chamadas de IA) possuem a possibilidade de automatizar e acelerar esse processo de seleção de feijões para aumentar o lucro de vendedores de feijão. Por esse motivo, iremos apresentar o trabalho de separação, seleção de características, treinamento e teste de um modelo de IA para a seleção de feijões.

### Metodologia

A forma escolhida para a seleção dos feijões foi a de análise por imagens, onde as imagens são selecionadas, separadas e 21 características são extraídas para que os feijões sejam classificados como bons ou ruins.

#### - Características

##### 1. Características de cor

Foram utilizadas 2 medidas relacionadas à coloração dos feijões: médias e desvios padrão de cada um dos canais de cor RGB (Red, Green e Blue), totalizando 6 características.

Essas características foram utilizadas pois a coloração pode indicar feijões com defeitos, manchas, fungos ou envelhecimento, que se diferenciam dos feijões saudáveis pela variação nas tonalidades de cor e uniformidade.

##### 2. Características morfológicas

Foram utilizadas 6 características morfológicas, sendo:

- **Área:** quantidade de pixels do feijão.
- **Perímetro:** contorno do grão.
- **Circularidade:** mede o quão próximo da forma circular está o feijão, calculada como  $(4 * \pi * \text{área}) / (\text{perímetro}^2)$ .
- **Aspect Ratio:** relação entre largura e altura do feijão.
- **Solidez:** mede o quão cheio é o feijão, sendo a razão entre a área e a área do casco convexo.
- **Extent:** relação entre a área do feijão e a área do retângulo delimitador (bounding box).

Estas características ajudam a diferenciar feijões quebrados, deformados ou rachados dos feijões inteiros e com formato ideal.

### 3. Características de textura

Foram utilizadas 2 características de textura:

- **Desvio padrão em tons de cinza:** mede a variação dos tons dentro do feijão.
- **Entropia:** mede a complexidade da textura, onde feijões mais danificados tendem a ter texturas menos uniformes e maior entropia.

Essas características são importantes, pois feijões defeituosos tendem a ter padrões de textura diferentes dos feijões bons.

### 4. Momentos invariantes de Hu

Foram extraídas 7 características chamadas Momentos de Hu, que representam propriedades geométricas da forma do feijão. São invariantes a rotação, escala e translação, ou seja, não importa o tamanho, posição ou orientação do feijão na imagem. Esses momentos são úteis para identificar formatos padrões ou deformações.

#### - Obtenção dos Dados

As imagens originais foram segmentadas individualmente, ou seja, cada feijão foi isolado em uma imagem própria. Posteriormente, foram extraídas todas as características descritas e organizadas em um arquivo CSV, onde:

- Cada linha representa um feijão.
- Cada coluna representa uma característica.
- A última coluna é a classe, sendo:
  - **1:** feijão bom.
  - **0:** feijão ruim.

Esse processo garantiu a criação de um conjunto de dados supervisionado, adequado para o treinamento dos modelos de inteligência artificial.

#### - Treinamento

O treinamento foi realizado utilizando o algoritmo de Random Forest Classifier, conhecido por sua robustez e boa capacidade de generalização. Esse algoritmo cria várias árvores de decisão e combina seus resultados para obter uma previsão mais precisa.

Foram utilizados três cenários de divisão dos dados, com diferentes proporções entre treino e teste, para avaliar como a quantidade de dados influencia no desempenho dos modelos:

- **70% para treino e 30% para teste**
- **80% para treino e 20% para teste**
- **90% para treino e 10% para teste**

A escolha dessas divisões permitiu avaliar a sensibilidade do modelo em relação à quantidade de dados disponíveis para treino e teste, além de verificar qual cenário produziria o modelo mais robusto.

Para cada cenário, foi aplicada uma busca em grade (Grid Search) para encontrar os melhores hiperparâmetros do modelo, incluindo:

- **Número de árvores** (`n_estimators`).
- **Profundidade máxima das árvores** (`max_depth`).
- **Número mínimo de amostras em uma folha** (`min_samples_leaf`).

Além disso, foi utilizada a estratégia de validação cruzada (K-Fold com 5 folds) para garantir que os modelos não estivessem ajustados especificamente para uma única divisão dos dados.

## Resultados

A tabela abaixo resume os resultados obtidos para cada proporção de treino/teste:

Split	Accuracy	Precision	Recall	F1	F1 Cross Validation (CV)
70/30	0.8971	0.8975	0.8971	0.8971	0.9075 ± 0.0213
80/20	0.8846	0.8854	0.8846	0.8847	0.9065 ± 0.0173
90/10	0.8846	0.8905	0.8846	0.8846	0.9036 ± 0.0193

O modelo com divisão **70/30** apresentou o melhor desempenho geral, tanto nas métricas de teste quanto na validação cruzada, demonstrando um equilíbrio ideal entre quantidade de dados de treino e teste.

## Discussão

Os resultados mostraram que a metodologia utilizada é eficiente para a classificação dos feijões.

- A abordagem de análise de imagem foi suficiente para extrair características que diferenciam feijões bons e ruins.
- As características morfológicas e de textura se mostraram altamente relevantes na separação das classes.
- A técnica de Random Forest foi adequada para este tipo de problema, por lidar bem com datasets tabulares e características diversas.
- Curiosamente, o modelo com menos dados de treino (**70/30**) teve melhor desempenho, indicando que o conjunto de teste mais robusto ajudou a validar melhor o modelo e evitar overfitting.

## Conclusão

O modelo final apresenta **acurácia superior a 89%**, mostrando-se adequado para ser aplicado em ambientes produtivos, contribuindo para a automação da classificação de feijões pretos. Mesmo que os resultados sejam próximos uns dos outros, o treinamento/teste 70/30 obteve os melhores

resultados. O resultado de acurácia determina que o modelo proposto pode ser utilizado em ambientes produtivos.