

# Classificazione di frutti esotici

TropicTaste Inc., leader nella distribuzione di frutti esotici, intende migliorare l'efficienza e l'accuratezza nel processo di classificazione dei frutti. L'obiettivo è sviluppare un modello di machine learning in grado di predire il tipo di frutto basandosi su caratteristiche numeriche.

L'attuale processo di classificazione dei frutti esotici è manuale e soggetto a errori, risultando inefficiente e dispendioso in termini di risorse. La necessità di un sistema automatizzato e preciso è cruciale per ottimizzare le operazioni aziendali e mantenere alti standard qualitativi.

Implementando un modello di classificazione automatizzato, TropicTaste Inc. potrà:

- **Migliorare l'Efficienza Operativa:** Automatizzare la classificazione ridurrà il tempo e le risorse necessarie, aumentando la produttività.
- **Ridurre gli Errori Umani:** Un modello di machine learning minimizzerà gli errori di classificazione, garantendo una maggiore precisione.
- **Ottimizzare l'Inventario:** Una classificazione accurata permetterà una migliore gestione dell'inventario, assicurando condizioni ottimali di conservazione per ogni tipo di frutto.
- **Aumentare la Soddisfazione del Cliente:** Una corretta identificazione e classificazione dei frutti contribuirà a mantenere elevati standard di qualità, migliorando la soddisfazione dei clienti.

## Dettagli del Progetto:

1. **Dataset:** Utilizzo di un dataset contenente varie caratteristiche numeriche di diversi frutti esotici.
2. **Algoritmo:** Implementazione dell'algoritmo K-Nearest Neighbors (KNN) per la classificazione.
3. **Output:** Il modello deve predire correttamente il tipo di frutto basandosi sui dati forniti.

## Requisiti del Progetto:

1. **Preparazione del Dataset:**
  - Caricamento e preprocessing dei dati sui frutti esotici.
  - Gestione di eventuali valori mancanti, normalizzazione e scalatura dei dati.
2. **Implementazione del Modello KNN:**
  - Sviluppo e addestramento del modello KNN.
  - Ottimizzazione dei parametri per migliorare l'accuratezza predittiva.
3. **Valutazione delle Performance:**
  - Utilizzo di tecniche di validazione incrociata per valutare la capacità di generalizzazione del modello.
  - Calcolo delle metriche di performance, come l'accuratezza e l'errore di classificazione.
4. **Visualizzazione dei Risultati:**
  - Creazione di grafici per visualizzare e confrontare le performance del modello.
  - Analisi e interpretazione dei risultati per identificare eventuali aree di miglioramento.

## Variabili del Dataset

Il dataset è scaricabile da qui: <https://proai-datasets.s3.eu-west-3.amazonaws.com/fruits.csv>

Contiene le seguenti variabili:

1. **Frutto:** Il tipo di frutto. Questa è la variabile di destinazione (target) che vogliamo prevedere.
2. **Peso (g):** Il peso del frutto in grammi. Variabile continua.
3. **Diametro medio (mm):** Il diametro medio del frutto in millimetri. Variabile continua.
4. **Lunghezza media (mm):** La lunghezza media del frutto in millimetri. Variabile continua.
5. **Durezza buccia (1-10):** La durezza della buccia del frutto su una scala da 1 a 10. Variabile continua.
6. **Dolcezza (1-10):** La dolcezza del frutto su una scala da 1 a 10. Variabile continua.
7. **Acidità (1-10):** L'acidità del frutto su una scala da 1 a 10. Variabile continua.

Ricorda di eseguire una corretta analisi esplorativa del dataset prima di procedere con il modello. Ricordati anche di applicare il corretto preprocessing ai dati e di misurare le performance del modello secondo le metriche che ritieni più utili.

Accompagna ogni passo dell'analisi con grafici appropriati e commenta adeguatamente ogni passaggio, estraendo degli insight utili all'analisi e alla comprensione dell'informazione trasmessa dal dato.

**Modalità di consegna:** Link pubblico a notebook di Google Colab

