# 📘 機器學習期末模擬考題（功能擴充）

一. 題目說明：  
  
請根據下方提供的初始程式碼，加入第七至第十一章所學的功能，完成下列任務：  
  
🔹 初始程式功能：  
- 讀取數據後以 KMeans 進行分群，並將分群結果視覺化  
  
🔧 請依序完成下列任務：  
1. 在分群結果上建立一個 SVM 分類器，並使用 5-fold 交叉驗證進行模型評估，印出平均準確率。  
2. 使用 Decision Tree 與 Random Forest 分別對分群結果進行分類，並比較其效能差異。  
3. 模擬 Transfer Learning 應用情境，使用 VGG16 模型特徵提取後分類（資料可模擬為 2D 圖像資料格式）。  
4. 額外 bonus：使用 Silhouette Score 評估最佳分群數，並於圖上標示群心。  
  
請以 Python 檔案完成擴充功能，命名為 `ML\_Final\_Function\_Extension.py`

二. 題目說明：  
  
本題提供一組關於行銷媒體與銷售成效的資料集（marketing\_dataset.csv），共有五個欄位：  
- SocialMedia：社群媒體投放費用（單位：千元）  
- TV：電視廣告投放費用（單位：千元）  
- Billboard：看板廣告投放費用（單位：千元）  
- Sales：實際銷售成果（單位：千元）  
- Class：銷售成效分級（Low / Medium）  
  
請完成以下任務：  
  
1. \*\*使用 K-Means 對前三個廣告特徵進行分群，並視覺化結果與群中心。\*\*   
2. \*\*使用 SVM、Decision Tree 與 Random Forest 模型對「Class」進行分類，列出準確率並比較其優劣。\*\*  
3. \*\*假設此資料為圖像特徵轉換結果，請示意如何使用 Transfer Learning 中的 VGG16 進行特徵提取後建模。\*\*  
4. \*\*使用 Silhouette Score 評估最佳群數（k=2~5），並標示最佳 k 值。\*\*  
  
請以 Python 撰寫完整程式，檔案命名為：`marketing\_analysis.py`

三. 請根據以下提供的初始程式碼，完成下列任務：

* 任務說明：

1. 1. 模型調整：在原本的 CNN 架構中，請加入一層 Dropout(p=0.3)，以減少 overfitting。
2. 2. 訓練與測試：對修改後的模型使用 80% 訓練資料與 20% 測試資料進行訓練與測試。
3. 3. 準確率視覺化：紀錄每個 epoch 的訓練與測試準確率，並畫出準確率折線圖（train vs test）。
4. 4. Dropout 比較分析：

* - 無 Dropout 模型的最終測試準確率
* - 有 Dropout 模型的最終測試準確率
* - 說明是否因 Dropout 而造成測試準確率提升或下降，並推論可能原因。
* 附註：
* - 請使用 MNIST 作為訓練資料集（題目程式碼中已內建）。
* - 本題可在相同架構下，訓練兩次模型（一次加 Dropout，一次不加）來進行比較。

起始程式碼：請下載 CNN\_Classification.py，完成補充功能與作答。