

**數據分析流程圖**

請參考上圖，協助繪製'數據分析流程圖'，本專案主要流程資料，包含以下項目：

'Data','資料清理 & EDA\資料剖析', '特徵工程','資料分割','訓練資料','演算法','模型訓練','測試資料','模型評分','模型評估','部署','預測'。

好的，這張圖是典型的機器學習（ML）專案流程。根據您提供的項目，我為您繪製了相對應的「數據分析流程圖」。

**數據分析流程圖**

這張流程圖展示了從取得原始資料到最終部署模型進行預測的完整步驟。

程式碼片段

graph TD

A["<font size=3><b>1. Data (資料)</b></font>"] --> B["<font size=3><b>2. 資料清理 & EDA/資料剖析</b></font>"];

B --> C["<font size=3><b>3. 特徵工程</b></font>"];

C --> D{"<font size=3><b>4. 資料分割</b></font>"};

D -- 訓練資料 --> F["<font size=3><b>6. 模型訓練</b></font>"];

D -- 測試資料 --> G["<font size=3><b>8. 模型評分</b></font>"];

E["<font size=3><b>5. 演算法</b></font>"] --> F;

subgraph "模型建置"

E

F

end

F -- 訓練好的模型 --> G;

G --> H["<font size=3><b>9. 模型評估</b></font>"];

H -- 最終模型 --> I["<font size=3><b>10. 部署</b></font>"];

I --> J["<font size=3><b>11. 預測</b></font>"];

%% Styling

style A fill:#f9f,stroke:#333,stroke-width:2px

style J fill:#ccf,stroke:#333,stroke-width:2px

**流程步驟說明**

1. **Data (資料)**: 專案的起點，收集原始資料。
2. **資料清理 & EDA/資料剖析**: 對資料進行清洗、處理缺失值、異常值，並透過探索性資料分析（EDA）來理解資料特性。
3. **特徵工程 (Feature Engineering)**: 根據對資料的理解，轉換、組合或創造出對模型訓練更有幫助的特徵。
4. **資料分割 (Data Splitting)**: 將處理好的資料集切分為「訓練資料」和「測試資料」。
5. **演算法 (Algorithm)**: 選擇適合解決當前問題的機器學習演算法（如：線性回歸、決策樹、神經網路等）。
6. **模型訓練 (Model Training)**: 使用「訓練資料」和所選的「演算法」來建立與訓練模型。
7. **訓練資料/測試資料**: 資料分割後的兩部分，分別用於訓練和驗證模型。
8. **模型評分 (Model Scoring)**: 使用「測試資料」來讓訓練好的模型進行預測，得出初步的預測結果（分數）。
9. **模型評估 (Model Evaluation)**: 比較模型的預測結果與「測試資料」的真實答案，使用準確率、F1-Score 等指標來評估模型的效能好壞。
10. **部署 (Deployment)**: 當模型效能達到標準後，將其部署到實際的生產環境中，使其能夠接收新資料並提供服務。
11. **預測 (Prediction)**: 已部署的模型接收真實世界的新資料，並做出預測，這也是專案的最終目的。