

中國醫藥大學附設醫院1JZ0-醫研人工智能智能演演開發工程師上機試題

(1) 信號測試題

預測癌症是良性還是惡性

特徵是根據乳房腫塊的細針抽吸 (FNA) 的數字化圖像計算的。它們描述了圖像中存在的細胞核的特徵。

屬性信息：

1. 身份證號碼

2. 診斷 (M = 惡性 ,B = 良性)

3. 為每個細胞核計算十個實值特徵：

· 半徑 (從中心到周邊點的距離平均值) · 紋理 (灰度值的標準偏差) · 周長

· 區域

· 光滑度 (半徑長度的局部變化) · 緊湊度 (-1.0) · 凹度

(輪廓凹入部分的嚴重程度) · 凹點 (輪廓凹入部分的分數)

形維數 (“海岸線近似” - 1)

為每張圖像計算這些特徵的平均值、標準誤差和“最差”或最大 (三個最大值的平均值) ,得到 30 個特徵。例如 ,字段 3 是平均半徑 ,字段 13 是半徑 SE ,字段 23 是最差半徑。

缺少屬性值 :無

類別分佈 :357良性 ,212惡性

問題：1. 顯示

所有特徵之間的相關性。

2. 使用其中一種特徵選擇方法來選擇 5 個最重要的特徵。

3. 使用 2. 中選擇的特徵通過任何 ML 模型 (SVM、邏輯回歸、隨機森林、xgboost、NN 等)預測風險 ,並進行以下數據拆分 :x_train、x_test、y_train、y_test = train_test_split(x, y, test_size = 0.25, random_state = 16)

4. 調整至少兩個超參數 (網格搜索、隨機搜索……)以提高模型的性能。

5. 可視化模型性能 (混淆矩陣、precision-recall 曲線、F1-score、...)

6. 請提交代碼生成測試結果。

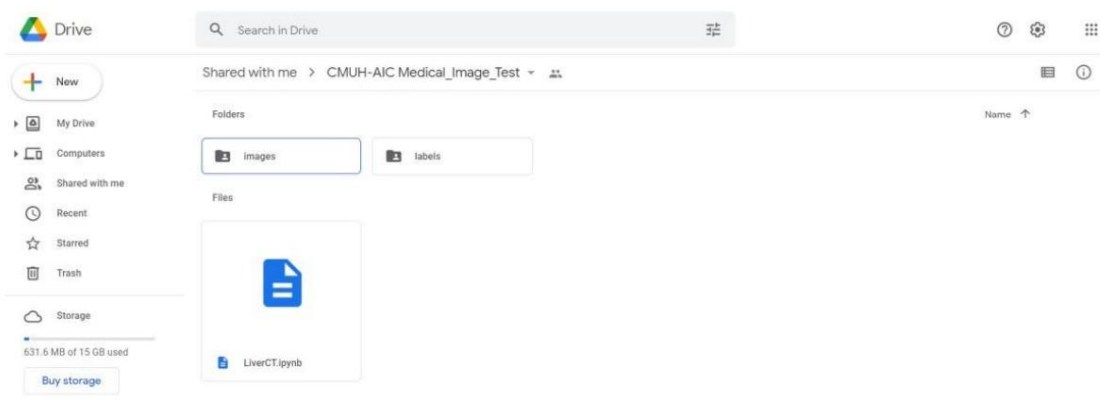
(2) 圖像測試題

中國醫藥大學附屬題設醫院 人工智慧醫療診斷中心醫學影像

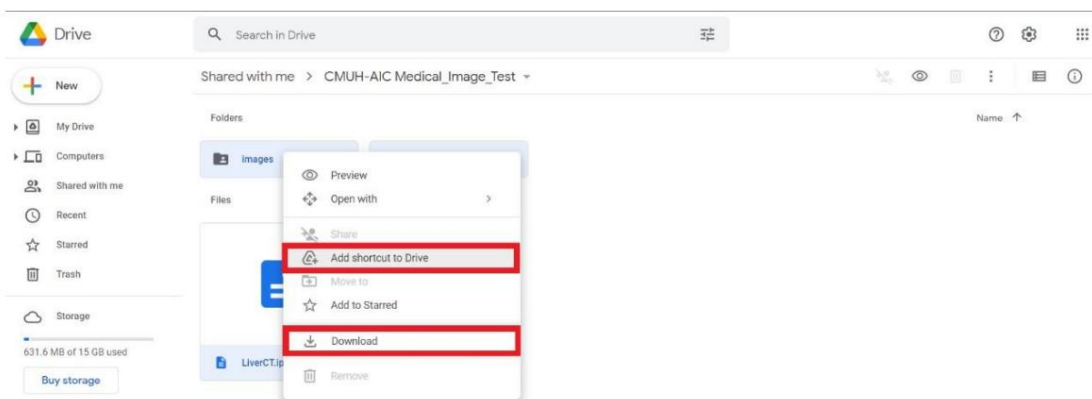
<https://drive.google.com/drive/folders/1cXzaqbw9RhU5PB5lHQWQ11hnPIZNEXQF?usp=分享>

請先至上列網址將圖像標記及標記下載或加入自己的 Google Drive 中，使用方法如下。

1. 進入下載檔案網址



2. 全選選擇下載文件夾到自己的硬盤硬盤中。



3. 請在 Test.ipynb 內作答，並回傳此檔案。若有其他補充資料可另以其他文字檔或圖檔一併

附上。

(3) NLP 測試題

IMDb 上的文本分類

數據集信息：

<http://ai.stanford.edu/~amaas/data/sentiment/>

數據集網址：

http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz

問題：

1. 基於任何機器學習方法訓練模型。
2. 可視化模型的性能（混淆矩陣、precision-recall 曲線、F1-score，...）。
3. 請簡要說明您在每個訓練步驟中所做的工作。
4. 請提交您的型號和代碼以產生測試結果。