Thử nghiệm model:

-Data : diabetes\_binary\_5050split\_health\_indicators\_BRFSS2015.csv

-Kí hiệu RF: random forest

LR:Logistic regression

DB01: là từ Data xóa các sample trùng

DB02 là từ DB01 xóa 8 feature k quan trọng

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Thuật toán | Xử lí data | Tham số được thay đổi | Cross  Validation  mean | Độ chính xác | F1 score | Time  train | Time predict |
| 01 | RF | DB01 | n\_estimators=100, random\_state=42 | 0.729 | 0.7386 | 0.738 | 7s | 0.3s |
| 02 | LG | DB01 | max\_iter=1000,  random\_state=42 | 0.743 | 0.7513 | 0.7512 | 1.3s | 0s |
| 03 | Stacking | DB01 | RF,LG và model cuối là LG | 0.742 | 0.752 | 0.752 | 42s | 0.5s |
| 04 | RF | DB02 | n\_estimators=100, random\_state=42 | 0.717 | 0.7235 | 0.723 | 5.3s | 0.3s |
| 05 | LG | DB02 | max\_iter=1000,  random\_state=42 | 0.7386 | 0.7475 | 0.7473 | 0.5s | 0s |
| 06 | Stacking | DB02 | RF,LG và model cuối là LG | 0.738 | 0.747 | 0.747 | 35.2s | 0.4s |
| 07 | RF | DB02  Scale sử dụng StandardScale | 'n\_estimators': [50, 100, 200, 300],      'max\_depth': [None, 10, 20, 30],      'min\_samples\_split': [2, 5, 10],      'min\_samples\_leaf': [1, 2, 4],      'max\_features': ['sqrt', 'log2'],      'bootstrap': [True, False] | 0.742 | 0.7516 | 0.751 | 8ph 22.9s | 0.3s |
| 08 | LG | DB02  Scale sử dụng StandardScale | param\_grid = {      'penalty': ['l1', 'l2',],      'C': [0.001, 0.01, 0.1, 1, 10, 100],      'solver': ['liblinear'],      'class\_weight': [None, 'balanced'],      'max\_iter': [1000, 1500, 2000]  } | 0.738 | 0.747 | 0.747 | 31.3s | 0s |
| 09 | Stacking | DB02  Scale sử dụng StandardScale | RF ở bước 7,LG ở bước 08 và LG(max\_iter=1000,random\_state=42)  ) là final estimate | 0.742 | 0.7516 | 0.751 | 39ph 38.4s | 0.3s |
| 10 | Stacking | DB02  Scale sử dụng StandardScale | RF ở bước 7,LG ở bước 08 và LG ở bước 08 là final estimate | 0.74 | 0.751 | 0.75 | 39ph 21.8s | 0.3s |
| 11 | Stacking | DB02  Scale sử dụng StandardScale | RF ở bước 7,LG ở bước 08 và RF ở bước 07 là final estimate | 0.7407 | 0.7513 | 0.75 | 56ph 10.6ss | 0.4s |