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 6 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 | 0.72 | 0.728 | 0.7276 | 5.3s | 0.3s |
| 05 | LG | DB02 | max\_iter=1000,  random\_state=42 | 0.739 | 0.747 | 0.7468 | 0.5s | 0s |
| 06 | Stacking | DB02 | RF,LG và model cuối là LG | 0.739 | 0.748 | 0.748 | 35.2s | 0.4s |