Thử nghiệm model:

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

-Kí hiệu RF: random forest

LR:Logistic regression

DB01: Xóa các sample bị trùng

DB02: từ DB01 xóa các features không quan trọng

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Thuật toán | Xử lí data | Tham số được thay đổi | Cross  Validition  (Mean) | Độ chính xác | Time  train | Time predict |
| 01 | RF | DB01 | n\_estimators=100, random\_state=42 | 0.7287 | 0.7386 | 7s | 0.4s |
| 02 | LG | Db01 | max\_iter=1000,  random\_state=42 | 0.742 | 0.7513 | 1.7s | 0s |
| 03 | RF | DB02 | n\_estimators=100, random\_state=42 | 0.72 | 0.7279 | 5.4 | 0.4s |
| 04 | LG | DB02 | max\_iter=1000,  random\_state=42 | 0.7389 | 0.747 | 0.6s | 0s |