Test lại với 20% dữ liệu

+ Grid\_search->random forest: 85,9%

+ Random\_search->random forest: 85,85%

+ Decision Tree-> Gini:77,6%, Entropy:79%, Log\_loss:79,45%

+ Logistic Regression: 83%

+ Support Vector Machine: with probability:84%, with hyperameter:79%

+ Kneighbor:n=7: 82,05%

+ AdaBoost: learning\_rate=0,2 : 86%

+ Gradient Boosting: learning\_rate=0,2 : 86%

=> Dựa theo bảng thống kê theo tiêu chí Precision, Recall và F1-Score, ta nhận thấy với dữ liệu Test-Train 20% thì thuật toán GradientBoostingClassifier là model học tốt nhất.

Đoán 88%(precision) người có khả năng NOT Churn và 95%(recall) trong 88%(precesion) là thực sự NOT Churn. F1-Score là 0.92 gần tiến về 1 chứng tỏ khả năng học của mô hình cực kỳ tốt.

Đoán 85%(precision) người có khả năng Churn và 49%(recall) trong 85%(precesion) là thực sự Churn. F1-Score là 0.62 vẫn cao hơn các thuật toán khác chứng tỏ GradientBoostingClassifier phân tích hiệu quả nhất.