



## Sử dụng phương pháp SMOTE

### Dữ liệu cân bằng

### Dữ liệu gốc

Model
Logistic Regression
Gradient Boosting
Neural Network

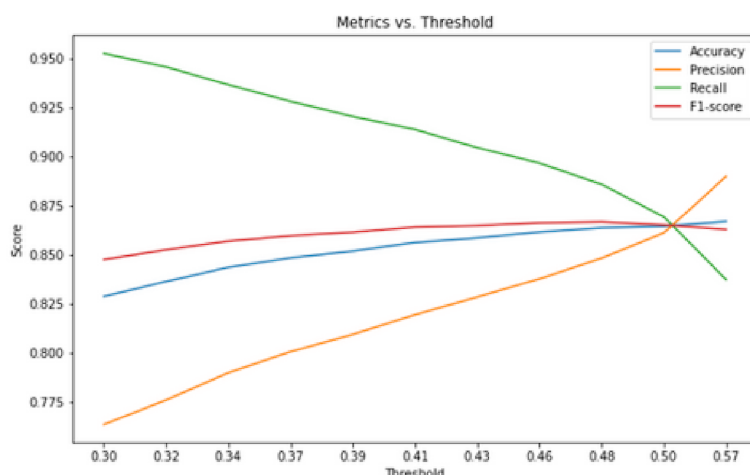
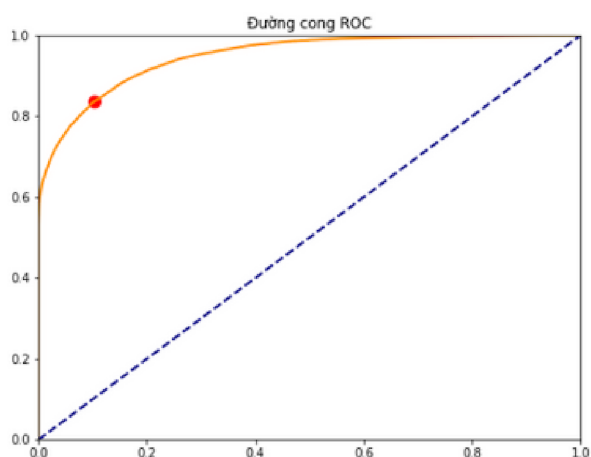
Threshold = 0.5

Accuracy	Precision	Recall	F1-Score
0.77	0.76	0.8	0.78
0.86	0.86	0.87	0.87
0.77	0.77	0.79	0.78

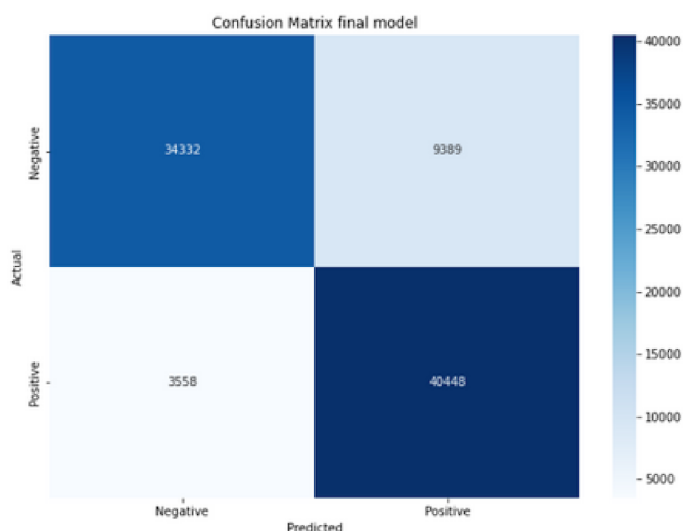
Accuracy	Precision	Recall	F1-Score
0.91	0.52	0.10	0.16
0.92	0.55	0.08	0.14
0.92	0.53	0.09	0.15

Mô hình cải thiện hiệu quả, tối ưu cả Recall và Precision. Tuy Accuracy giảm nhưng không đáng kể. Đặc biệt là mô hình sử dụng Gradient Boosting

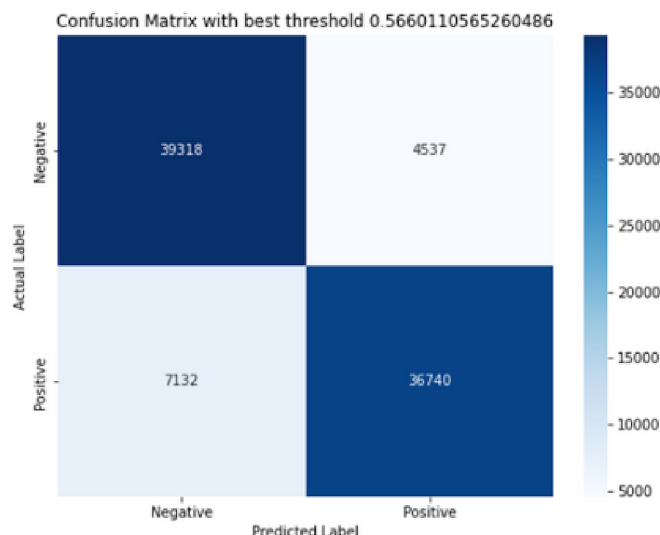
## Áp dụng đường cong ROC



Sự khác biệt giữa các threshold không quá lớn, có thể chọn threshold từ đường con ROC. Tuy nhiên cần xem xét mục đích của mô hình.



Mô hình lựa chọn



Mô hình dùng best threshold