

Klasifikasi Transaksi Kartu Kredit Menggunakan Algoritma Random Forest dan XGBoost

ANGGITA GHOZALI

Fraud & Machine Learning

Fraud



Kecurangan

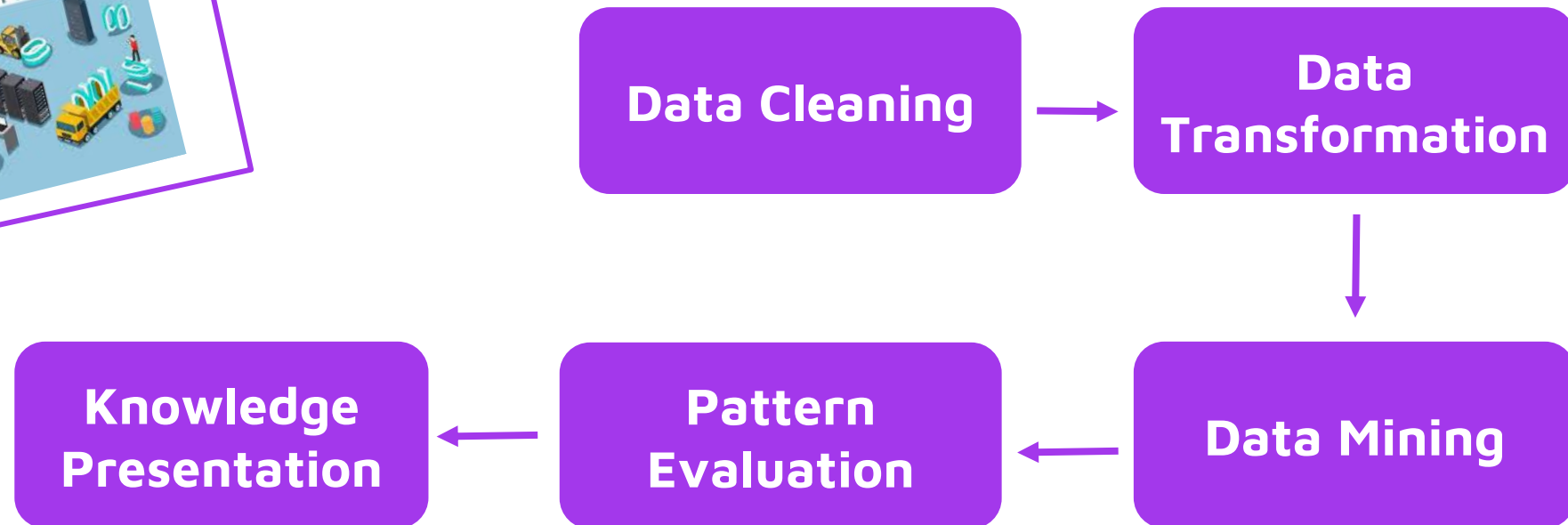
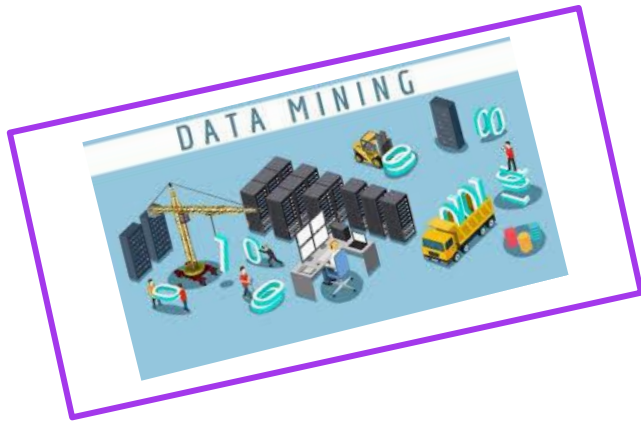


Fraud pada Kartu Kredit



Mencegah Fraud

Data Mining



Machine Learning

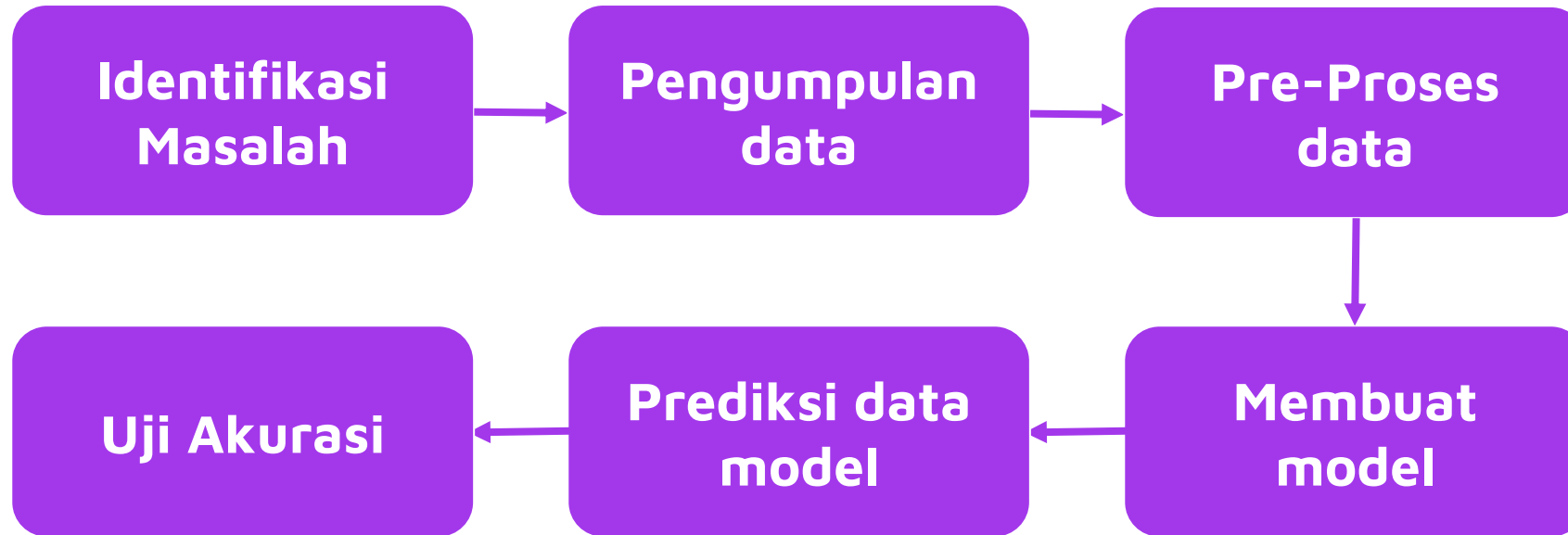
Random Forest

Merupakan pengembangan dari metode CART dengan menerapkan metode bootstrap aggregating (bagging)

XGBoost

Salah satu teknik pembelajaran mesin untuk mengatasi permasalahan regresi dan klasifikasi berdasarkan Gradient Boosting Decision Tree.

Deteksi Fraud dengan Machine Learning



Random Forest

TAHAP-TAHAP PEMROSESAN

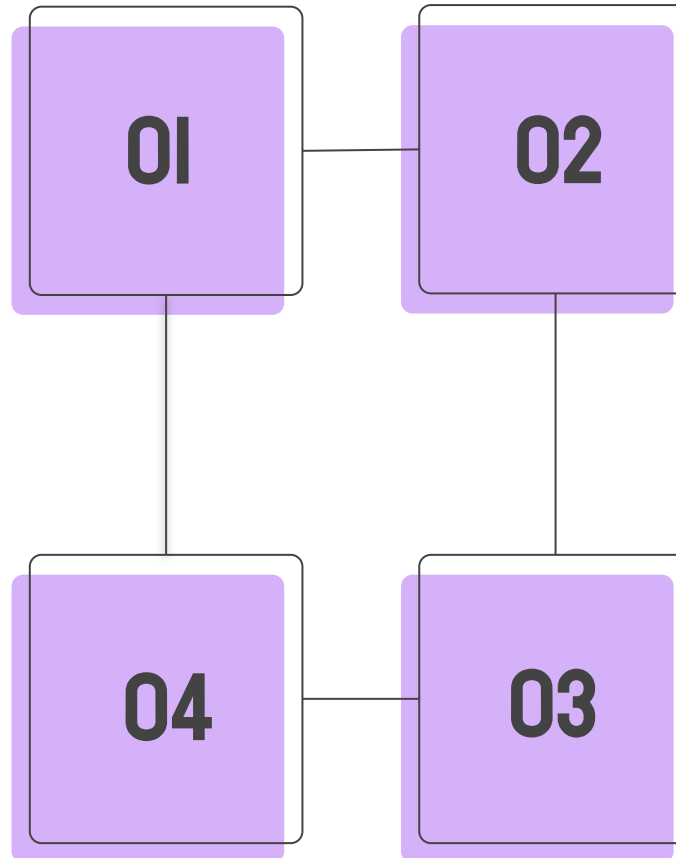
DATA CLEANING (PRE-PROCESSING)

Pada bagian ini kita membuat beberapa bentuk data:

1. Data dengan Drop Missing Values > 50%, Imputasi < 50 %, dan Outlier.
2. Data dengan Drop Missing Values > 50%, Imputasi < 50%, dan Tanpa Outlier
3. Data dengan Drop All Missing Values, dan Tanpa Outlier

PEMILIHAN MODEL TERBAIK

Pada bagian ini kita akan membandingkan model terbaik dari hasil akurasi, Confusion Matrix dan Feature Importance



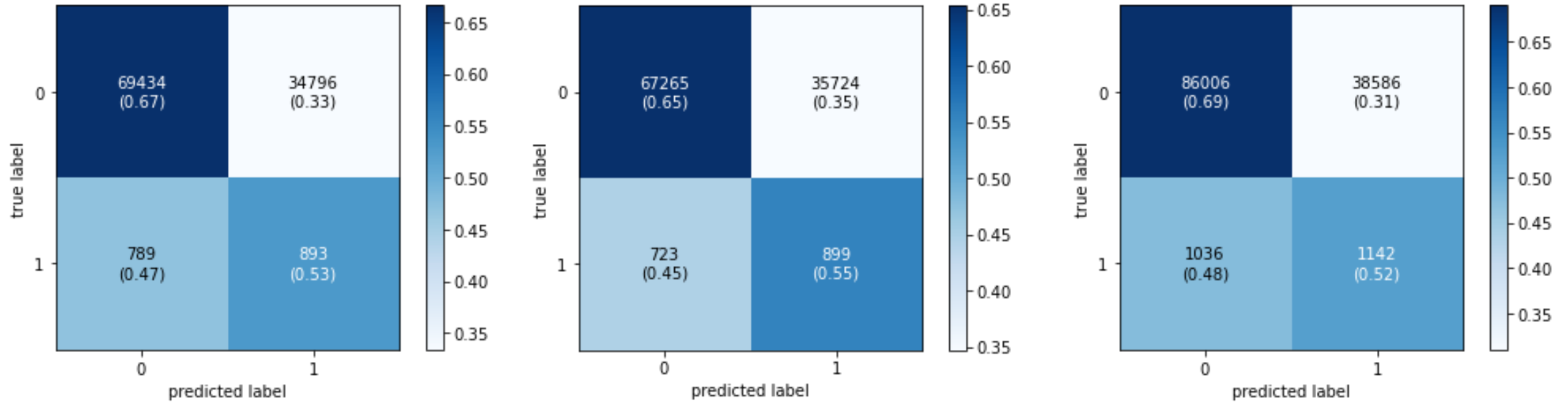
SPLIT DATA (TRAIN DAN TEST)

Pada bagian ini kita akan membuat dua model data yakni X dan y. Sebelum membuatnya kita perlu menyesuaikan tipe data agar saat diproses lebih mudah. Disini juga setelah Split Data kita akan menangani Imbalance Data dengan menggunakan Random Under Sampling dan Random Over Sampling.

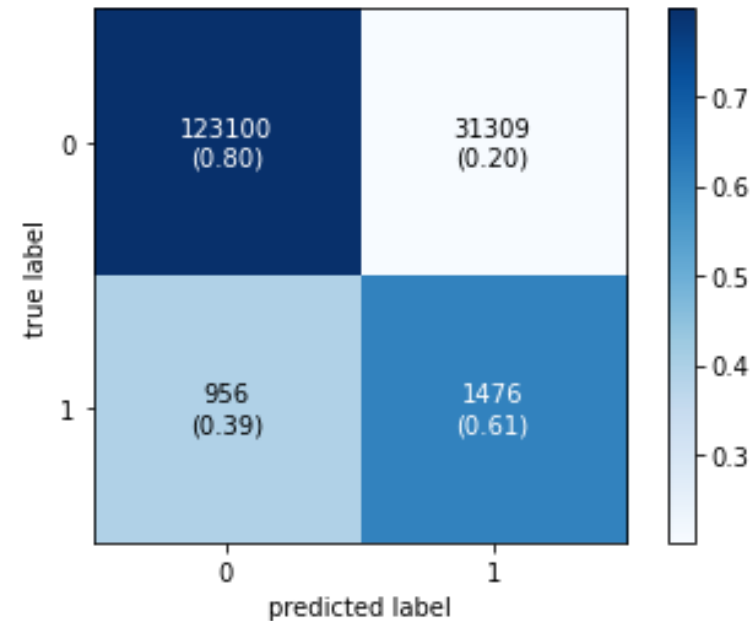
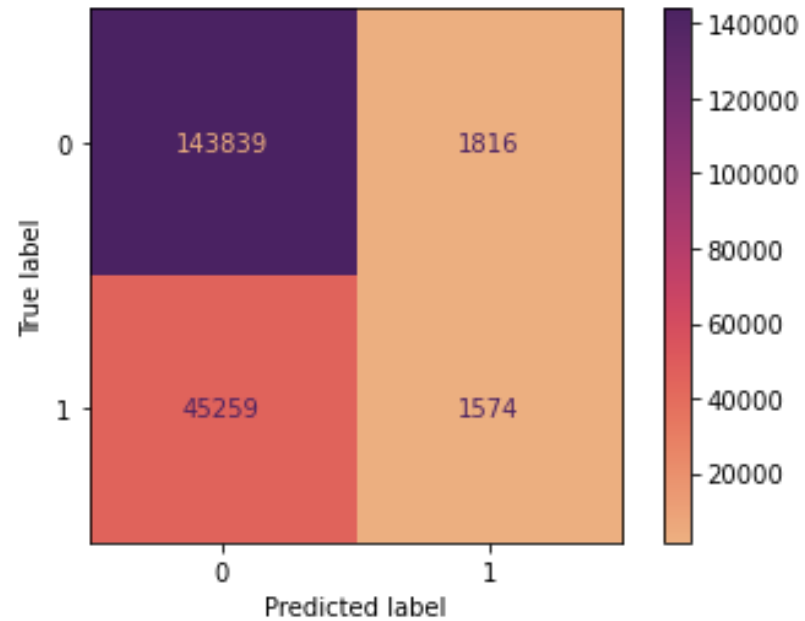
MODELING DENGAN ALGORITMA XGBOOST

Pada bagian ini kita akan membuat model dengan algoritma XGBoost & Random Forest dalam menentukan akurasi pada transaksi fraud atau legitimate.

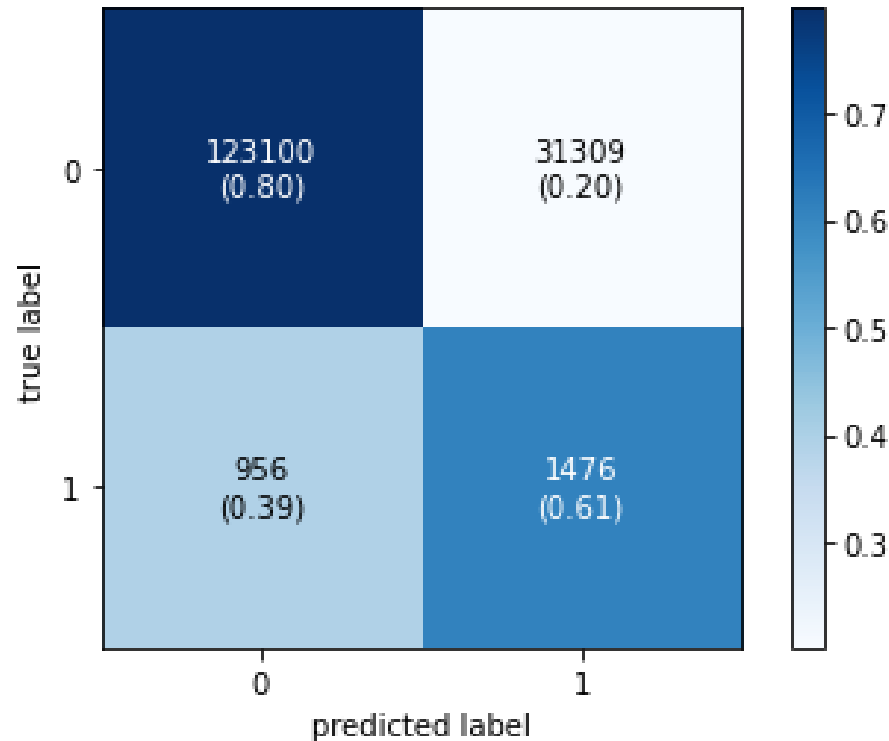
Model Random Forest



Model Random Forest



Model Terbaik Random Forest

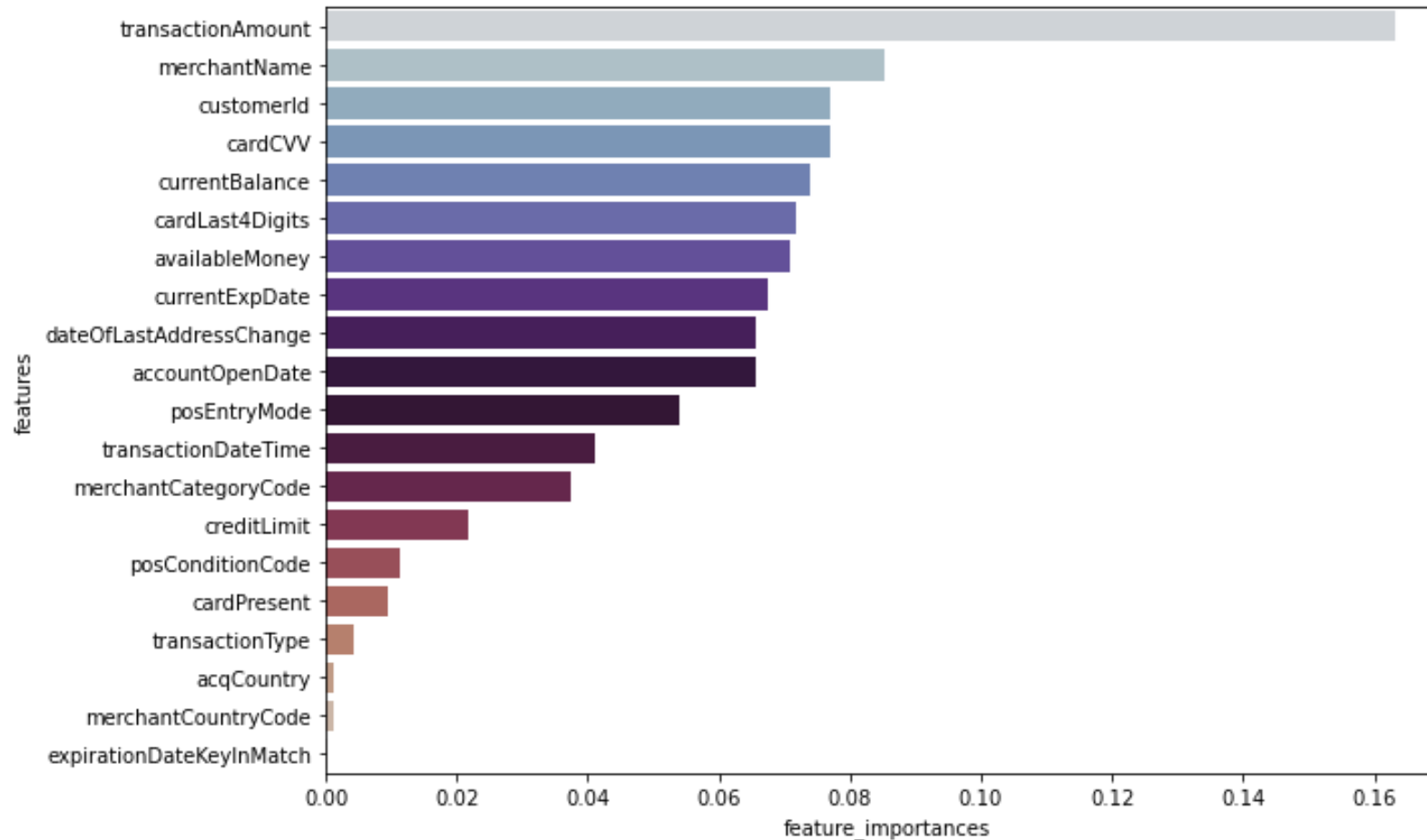


bootstrap = 'True', n_estimators = 250,
max_features = 10, max_depth = 20)

Accuracy: 79.42821073571324 %

	precision	recall	f1-score	support
0	0.99	0.80	0.88	154409
1	0.05	0.61	0.08	2432
accuracy			0.79	156841
macro avg	0.52	0.70	0.48	156841
weighted avg	0.98	0.79	0.87	156841

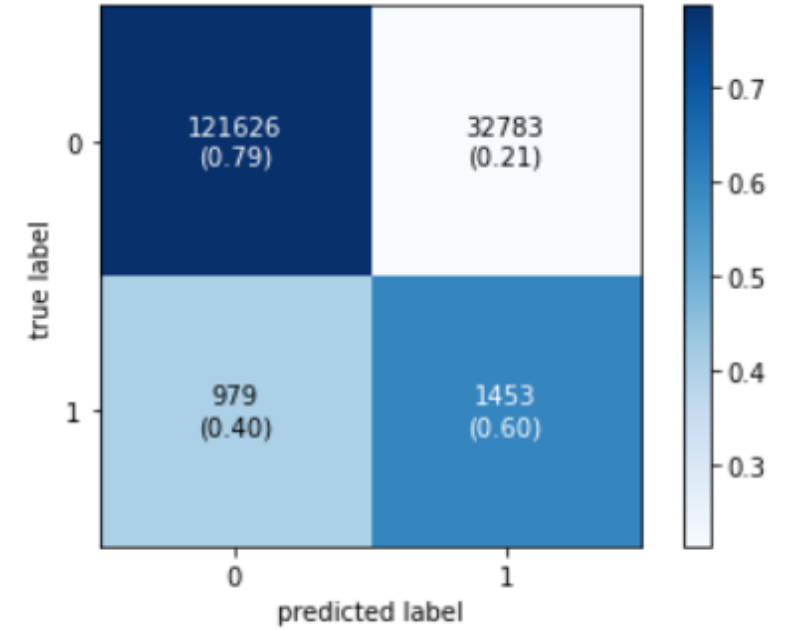
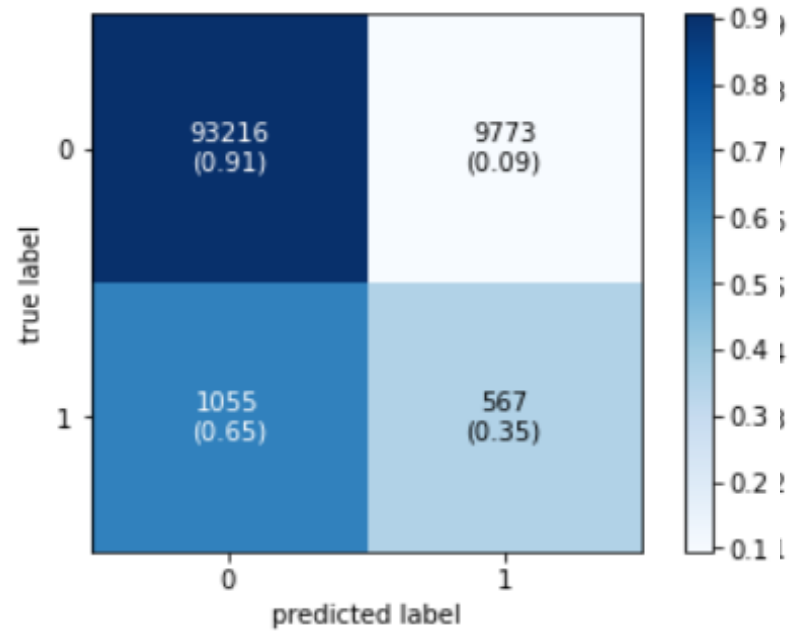
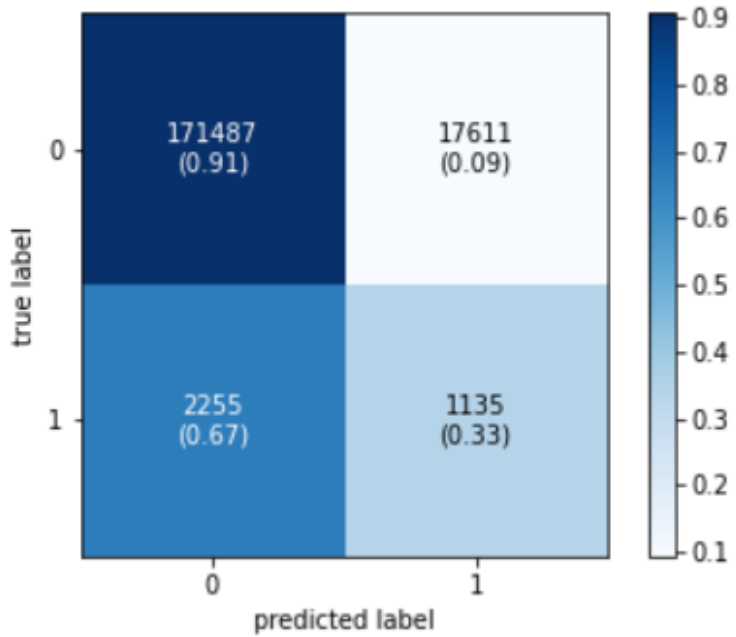
Feature Importance Random Forest



XGBOOST

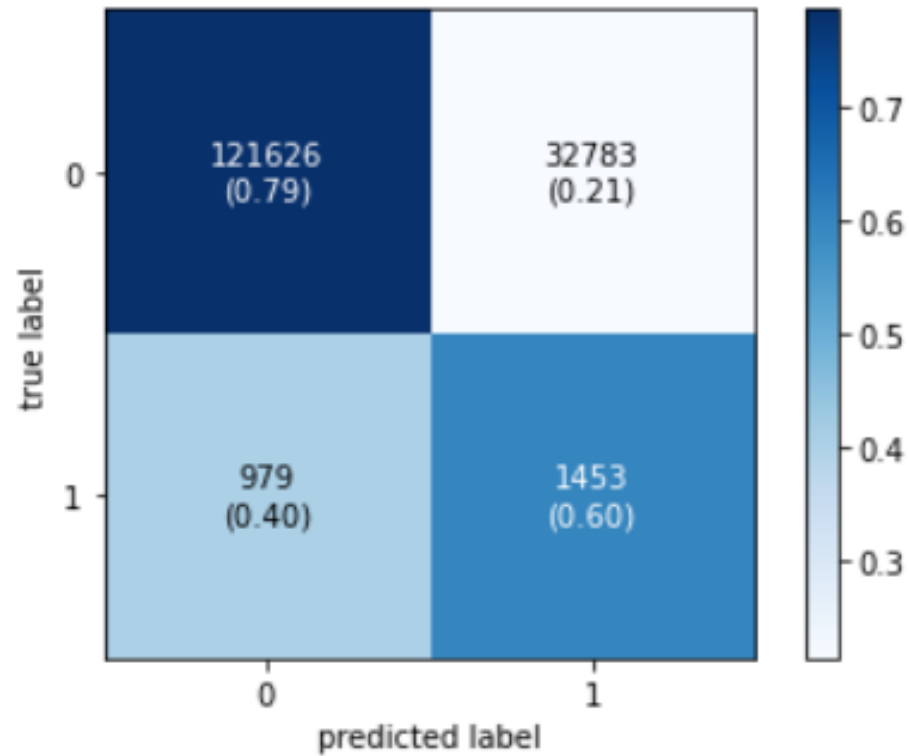


MODEL XGBOOST



MODEL TERBAIK PADA XGBOOST

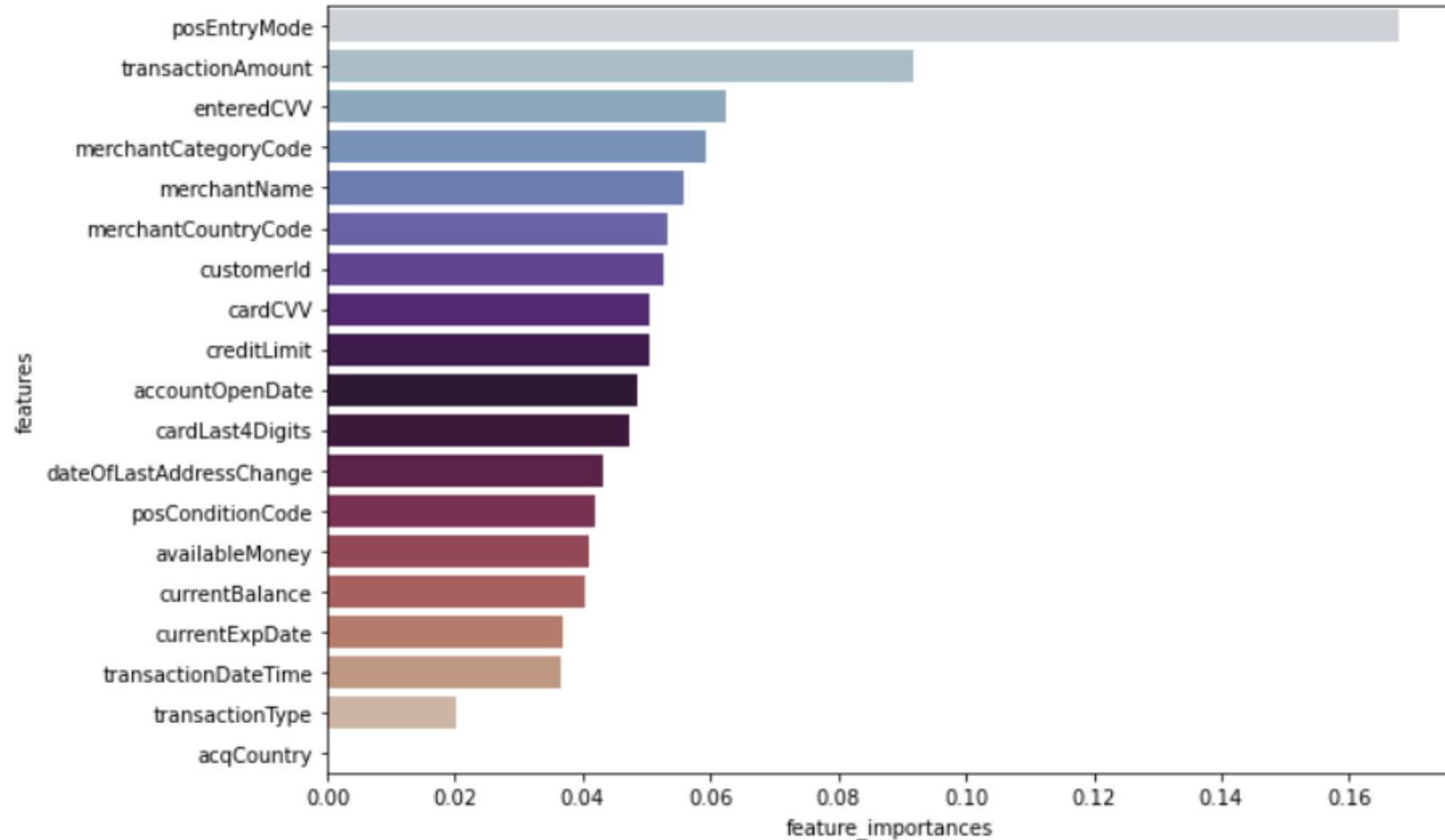
Parameter : Default



Classification Report of XGB Classifier:

	precision	recall	f1-score	support
0	0.99	0.79	0.88	154409
1	0.04	0.60	0.08	2432
accuracy			0.78	156841
macro avg	0.52	0.69	0.48	156841
weighted avg	0.98	0.78	0.87	156841

FEATURE IMPORTANCE



KESIMPULAN

Berdasarkan hasil dari confusion matrix Random Forest dan XGBoost, maka diperoleh model yang paling baik yaitu Random Forest dengan parameter (bootstrap = 'True', n_estimators = 250, max_features = 10, max_depth = 20) dan akurasi sebesar 79%, sehingga faktor yang paling berpengaruh terhadap kecurangan pada transaksi kartu kredit yaitu TransactionAmount.