Raport

Parametry:

1) train_test_split
   a) test_size=0.3
   b) random_state=50
   c) shuffle = True
2) DecisionTreeClassifier
   a) criterion='gini'
3) RandomForestClassifier
   a) n_estimators = 10
   b) criterion = 'log_loss'

	Głębokość drzewa
Niezbalosowane dane	5
Under-sumpling	4
Over-sumpling	4
SMOTE	2

Wyniki pomiarów:

	Dokładność	ROC	F1
DT z niezbalosowanymi danymi	83%	66%	48%
RF z niezbalosowanymi danymi	82%	64%	44%
DT z under-sumplingiem	79%	70%	52%
RF z under-sumplingiem	78%	71%	53%
DT z over-sumplingiem	78%	71%	53%
RF z over-sumplingiem	79%	71%	54%
DT z SMOTE	82%	68%	51%
RF z SMOTE	80%	69%	51%

Wnioski:

1. Przy stosowaniu metod balansujących dane wzrasta wartość ROC i F1 kosztem dokładności.
2. Przy stosowaniu metod balansujących dane dla optymalnego wyniku należy zmniejszyś maksymalną głębokość drzewa, szczególnie dla SMOTE.
3. Drzewo Decyzyjne wydaje się nieznacznie skuteczniejsze od Random forest.
4. Biorąc pod uwagę wszystkie 3 współczynnki można stwierdzić że najskuteczniejsza jest metoda SMOTE.