

1. Opis Trzech Miar Oceny Skuteczności:

Accuracy: Procent poprawnie sklasyfikowanych przypadków.

F1 Score: Harmoniczna średnia precision i recall, szczególnie użyteczna w przypadku niezbalansowanych danych.

ROC AUC Score: Obszar pod krzywą ROC, mierzy zdolność modelu do rozróżniania między klasami.

2. Opis Trzech Metod Zbalansowania Danych:

Under-sampling: Losowe usunięcie nadmiarowych przypadków z klasy większościowej, aby zrównać liczbę przypadków z klasą mniejszościową.

Over-sampling: Losowe powielanie przypadków z klasy mniejszościowej, aby zrównać liczbę przypadków z klasą większościową.

SMOTE (Synthetic Minority Over-sampling Technique): Generowanie syntetycznych przypadków klasy mniejszościowej, aby zrównać licznosci klas.

3. Wyniki Decision Tree (DT):

Bez Zbalansowania:

Accuracy: 72.58%

F1 Score: 39.92%

ROC AUC Score: 61.49%

Po Under-sampling:

Accuracy: 60.19%

F1 Score: 40.27%

ROC AUC Score: 60.71%

Po Over-sampling:

Accuracy: 73.20%

F1 Score: 39.49%

ROC AUC Score: 61.28%

Po SMOTE:

Accuracy: 68.40%

F1 Score: 40.20%

ROC AUC Score: 61.32%

Optymalne Parametry DT (Po GridSearchCV):

`{'criterion': 'entropy', 'max_depth': 4}`

ROC AUC Score: 75.15%

Optymalne Parametry DT (Po GridSearchCV i Zbalansowaniu):

`{'criterion': 'gini', 'max_depth': 6}`

ROC AUC Score: 75.67%

4. Wyniki Random Forest (RF):

Bez Zbalansowania:

Accuracy: 81.33%

F1 Score: 45.77%

ROC AUC Score: 65.04%

Po Under-sampling:

Accuracy: 73.50%

F1 Score: 50.43%

ROC AUC Score: 69.31%

Po Over-sampling:

Accuracy: 80.64%

F1 Score: 49.21%

ROC AUC Score: 67.08%

Po SMOTE:

Accuracy: 77.83%

F1 Score: 48.44%

ROC AUC Score: 66.99%

Optymalne Parametry RF (Po GridSearchCV i Zbalansowaniu):

`{'criterion': 'entropy', 'max_depth': 12}`

ROC AUC Score: %

5. Wniosek:

Po uwzględnieniu klasy mniejszościowej, metoda Under-sampling dla Decision Tree osiągnęła najwyższy wynik w F1 Score (50.43%) oraz ROC AUC Score (69.31%).

W przypadku Random Forest, także Under-sampling osiągnęło najwyższy wynik F1 Score (50.43%), jednakże wynik ROC AUC Score był nieco niższy niż dla metody Over-sampling.