

# Raport

## Skuteczność Modeli Drzewa Decyzyjnego (DT)

### Znalezione przez Grid Search Parametry DT:

- max\_depth: 30
- min\_samples\_leaf: 4
- min\_samples\_split: 2
- **DT Oryginalne:**
  - Accuracy: 75.2%
  - F1 Score: 0.411
  - ROC AUC: 0.622
- **DT po RUS:**
  - Accuracy: 64.2%
  - F1 Score: 0.426
  - ROC AUC: 0.625
- **DT po Over-Sampling:**
  - Accuracy: 72.8%
  - F1 Score: 0.415
  - ROC AUC: 0.622
- **DT po SMOTE:**
  - Accuracy: 72.0%
  - F1 Score: 0.414
  - ROC AUC: 0.621

## Skuteczność Modeli Lasu Losowego (RF)

### Znalezione przez Grid Search Parametry RF:

- max\_depth: None
- min\_samples\_leaf: 4
- min\_samples\_split: 100
- n\_estimators: 300
- **RF Oryginalne:**
  - Accuracy: 82.1%
  - F1 Score: 0.480
  - ROC AUC: 0.660
- **RF po RUS:**
  - Accuracy: 75.5%
  - F1 Score: 0.539
  - ROC AUC: 0.713
- **RF po Over-Sampling:**
  - Accuracy: 80.5%
  - F1 Score: 0.538
  - ROC AUC: 0.699

- **RF po SMOTE:**
  - Accuracy: 79.9%
  - F1 Score: 0.531
  - ROC AUC: 0.696

## **Wnioski:**

### **Las Losowy (RF) vs Drzewo Decyzyjne (DT):**

- Modele RF generalnie wykazują lepszą skuteczność niż DT, szczególnie pod względem Accuracy i F1 Score.

### **Wpływ Balansowania Danych:**

- **DT:** Zastosowanie metod zbalansowania danych przynosi różne efekty. Wzrost skuteczności nie jest znaczący, choć Over-Sampling i SMOTE oferują lepsze wyniki niż RUS pod względem Accuracy.
- **RF:** Metody balansowania danych znacznie poprawiają wyniki dla modeli RF, zwłaszcza RUS i SMOTE, które zwiększają wartości F1 Score i ROC AUC, sugerując lepsze radzenie sobie z niezbalansowanymi danymi.

### **Podsumowanie:**

- Modele RF generalnie przewyższają DT we wszystkich zastosowaniach, z wyjątkową poprawą w przypadku metod balansowania danych.
- W przypadku DT, wybór metody balansowania danych powinien być starannie rozważony, ponieważ różnice w skuteczności są mniej znaczące niż w przypadku RF.