



Politechnika
Wrocławska

Analiza Danych Empirycznych

Ocena Wydajności Modeli Klasyfikacji

Wiktor Wołkowski (290298)

Politechnika Wrocławska

15.12.2025



unite!
University Network for Innovation
Technology and Engineering



HR EXCELLENCE IN RESEARCH

Zaliczenie zadania

- Repozytorium: <https://github.com/Wilk1717/zadanie3.git>

Plan prezentacji

- 1 Charakterystyka porównywanych algorytmów
- 2 Implikacje praktyczne wdrożenia
- 3 Dane i Metodologia
- 4 Wyniki
- 5 Podsumowanie

Charakterystyka porównywanych algorytmów

Regresja Logistyczna

- ▶ Reprezentant prostych modeli liniowych[cite: 61].
- ▶ **Zaleta:** Łatwa interpretowalność zależności[cite: 62].
- ▶ **Ograniczenie:** Zakłada liniową relację między cechami a wynikiem, co może być niewystarczające[cite: 62].

Las Losowy (Random Forest)

- ▶ Algorytm zespołowy oparty na wielu drzewach decyzyjnych[cite: 65].
- ▶ Modeluje złożone, nieliniowe granice decyzyjne[cite: 66].
- ▶ Lepiej radzi sobie ze specyfiką danych finansowych[cite: 66].

Implikacje praktyczne wdrożenia

Wdrożenie modelu Lasu Losowego niesie za sobą konkretne korzyści biznesowe:

- ▶ **Automatyzacja procesów:** Zastąpienie manualnej weryfikacji automatyczną walidacją transakcji[cite: 95].
- ▶ **Redukcja błędów:** Wysoka precyzja modelu minimalizuje ryzyko błędnych decyzji (np. odrzucenia poprawnej transakcji)[cite: 95].
- ▶ **Efektywność kosztowa:** Ograniczenie strat finansowych wynikających z fraudów oraz zmniejszenie nakładu pracy ludzkiej[cite: 95].

Charakterystyka zbioru danych

- ▶ **Zbiór:** Syntetyczne dane finansowe (1000 rekordów).
- ▶ **Cel:** Wykrywanie oszustw (klasyfikacja binarna).
- ▶ **Kluczowe atrybuty:**
 - ▶ Czas i kwota transakcji.
 - ▶ Typ klienta (detaliczny/korporacyjny).
 - ▶ Zmienna docelowa (0 - fraud, 1 - legalna).

Przetwarzanie danych

1. **Czyszczenie:** Imputacja braków średnią arytmetyczną.
2. **Normalizacja:** Skalowanie Min-Max do przedziału $[0, 1]$.
3. **Inżynieria cech:** Kodowanie One-Hot dla zmiennych kategoriycznych.
4. **Podział zbioru:**
 - ▶ Treningowy: 70%
 - ▶ Walidacyjny: 15%
 - ▶ Testowy: 15%

Porównanie modeli



Algorytm	Dokładność	F1-Score
Regresja Logistyczna	89.5%	0.88
Las Losowy	94.2%	0.95

Tabela: Wyniki na zbiorze testowym

Wnioski końcowe

- ▶ **Las Losowy** okazał się skuteczniejszy niż Regresja Logistyczna.
- ▶ Modele liniowe nie radzą sobie w pełni ze złożonymi relacjami w danych finansowych.
- ▶ Wysoki F1-Score (0.95) potwierdza stabilność wybranego rozwiązania.
- ▶ Implementacja modelu może zredukować potrzebę manualnej weryfikacji transakcji.

Wykorzystane materiały

-  John Doe, *Sztuczna Inteligencja: Podstawy i Zastosowania*, Tech Press, 2023.
-  Alice Smith, *Znaczenie liniowej regresji w uczeniu maszynowym*, Journal of Computational Science, 2022.

Dziękuję za uwagę.