

Opis Implementacji Algorytmów Klasyfikacji

Weronika Kłujso (223599), Michał Korzeniewski (223399),
Miłosz Malinowski (223391), Piotr Misiejuk (223302)

June 10, 2025

Plik `data.py`

- Moduł pomocniczy zawierający wspólne funkcje wykorzystywane w całym projekcie klasyfikacji
- Dostarcza narzędzia do:
 - Wczytywania i preprocessingu danych
 - Usuwania wartości odstających (oparte o IQR)
 - Transformacji i skalowania cech (StandardScaler, MinMaxScaler)
 - Generowania statystyk opisowych
 - Generowania macierzy korelacji i wizualizacji
 - Eksportu tabeli statystyk do formatu LaTeX

Implementacje Random Forest

`forest.py`

- Implementacja klasyfikacji przy użyciu algorytmu Random Forest z danymi po standaryzacji
- Generuje wyniki klasyfikacji oraz metryki jakości (Accuracy, F1-score, ROC AUC)

`forest_minmax.py`

- Wersja Random Forest z danymi po MinMax skalowaniu
- Umożliwia porównanie wpływu różnych metod skalowania na skuteczność klasyfikatora

Implementacje modelu hybrydowego

hybrid_model.py

- Implementacja modelu hybrydowego VotingClassifier (Random Forest + KNN + LDA + XGBoost)
- Wersja ze standaryzacją danych
- Zawiera ocenę jakości klasyfikacji

hybrid_model_minmax.py

- Wersja modelu hybrydowego VotingClassifier dla danych po MinMax skalowaniu

Implementacje regresji

LogReg.py

- Klasyfikacja z wykorzystaniem regresji logistycznej na danych po standaryzacji
- Ocena jakości klasyfikacji

LogReg_minmax.py

- Wersja regresji logistycznej na danych po MinMax skalowaniu

Implementacje Linear Discriminant Analysis (LDA)

LDA.py

- Implementacja LDA dla klasyfikacji danych po standaryzacji

LDA_minmax.py

- Wersja LDA dla danych po MinMax skalowaniu

Pozostałe pliki pomocnicze

transform.py

- Funkcje do transformacji danych:
 - Skalowanie (StandardScaler, MinMaxScaler)
 - Transformacje logarytmiczne wybranych zmiennych
 - Przygotowanie danych do modelowania

statystyki.py

- Skrypt do generowania statystyk opisowych zbioru danych
- Wykorzystuje funkcje z data.py
- Eksport statystyk do formatu LaTeX

visualization.py

- Biblioteka funkcji do wizualizacji danych:
 - Histogramy i wykresy pudełkowe
 - Macierze korelacji
 - Wizualizacja rozkładu cech

model_visualisation.py

- Funkcje wizualizujące wyniki modeli:
 - Wykresy SHAP dla interpretacji modeli
 - Dodatkowe wizualizacje wspomagające analizę wyników klasyfikacji