



CHOROBY SERCA W MODELU XGBOOST

Michał Jurek

Jakub Dudajek

CHOROBY SERCA

- Główna przyczyna zgonów na świecie.
- Obejmują schorzenia naczyń, mięśnia sercowego i układu przewodzącego.
- Mogą prowadzić do bólu w klatce piersiowej, duszności, zmęczenia lub nawet nagłego zgonu.
- Za najczęstsze czynniki ryzyka podaje się: nadciśnienie, otyłość, cukrzyca, palenie tytoniu, brak aktywności fizycznej.

ZMIENNE

Źródłem jest kaggle. W zbiorze jest 299 obserwacji.

Wiek, anemia, kinaza kreatynowa, cukrzyca, frakcja wyrzutowa, wysokie ciśnienie krwi, ilość płytek krwi, kreatynina w surowicy, sód w surowicy, płeć, palenie, okres obserwacji pacjenta, wystąpienie śmierci.

XGBOOST

- Nowe rozgałęzienia poprawiają ocenę modelu
- Kolejne drzewa dodawane sekwencyjnie korygują błędy poprzednich
- Współczynnik uczenia reguluje wpływ każdego drzewa na końcową predykcję

XGBOOST

- Obserwacje przechodzą przez drzewa otrzymując oceny na podstawie wartości zmiennych
- Oceny przekształcane są w wynik, który umożliwia uporządkowanie i wybór najlepszych obserwacji.

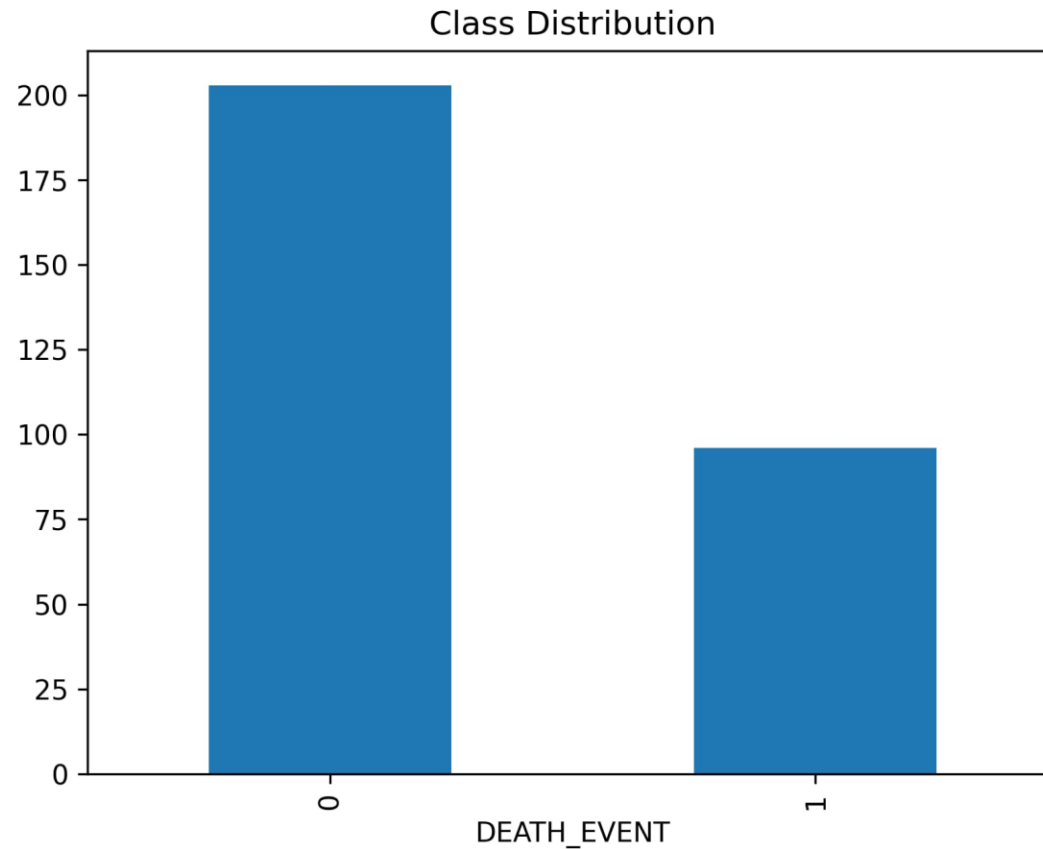
RANDOMIZEDSEARCHCV

RandomizedSearchCV to metoda optymalizacji hiperparametrów, która losowo próbuje określone liczby kombinacji hiperparametrów z podanego zakresu. Każda kombinacja jest oceniana przy użyciu walidacji krzyżowej, a najlepszy zestaw hiperparametrów jest wybierany na podstawie wyników modelu.

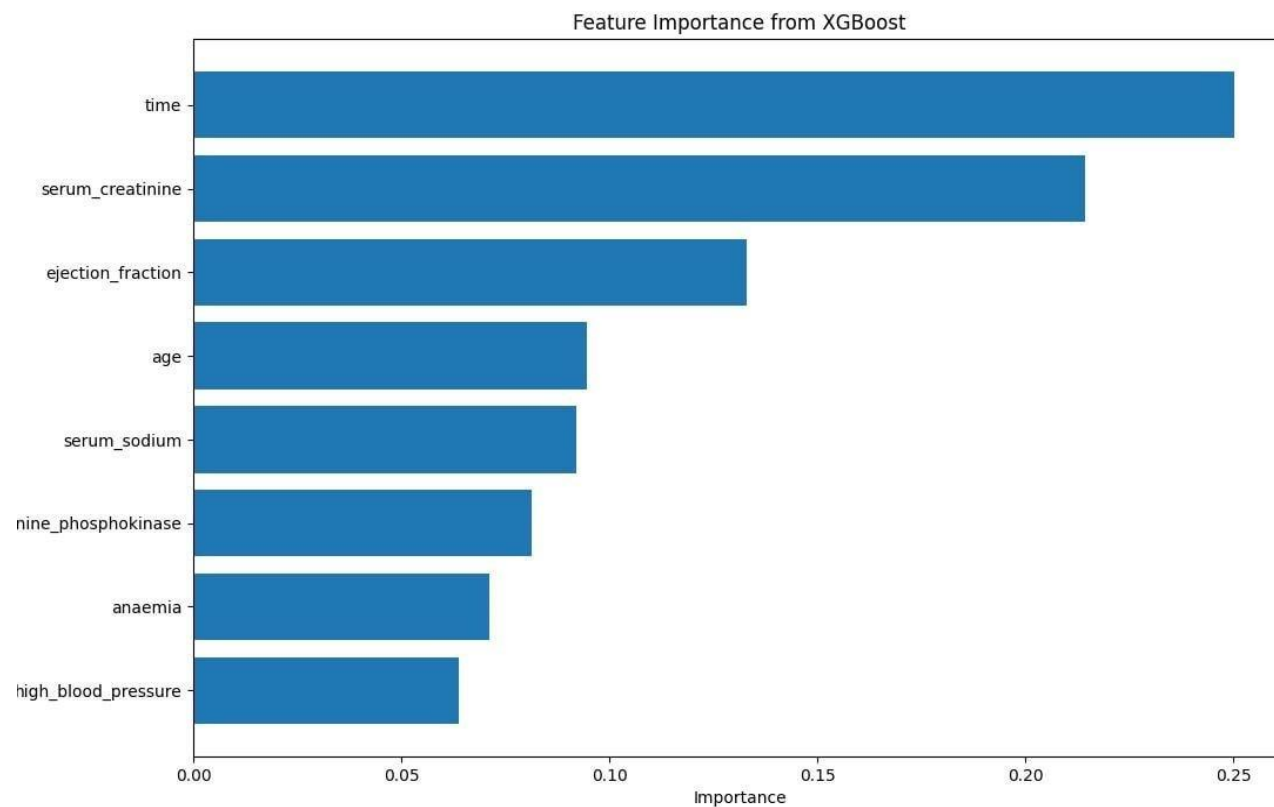
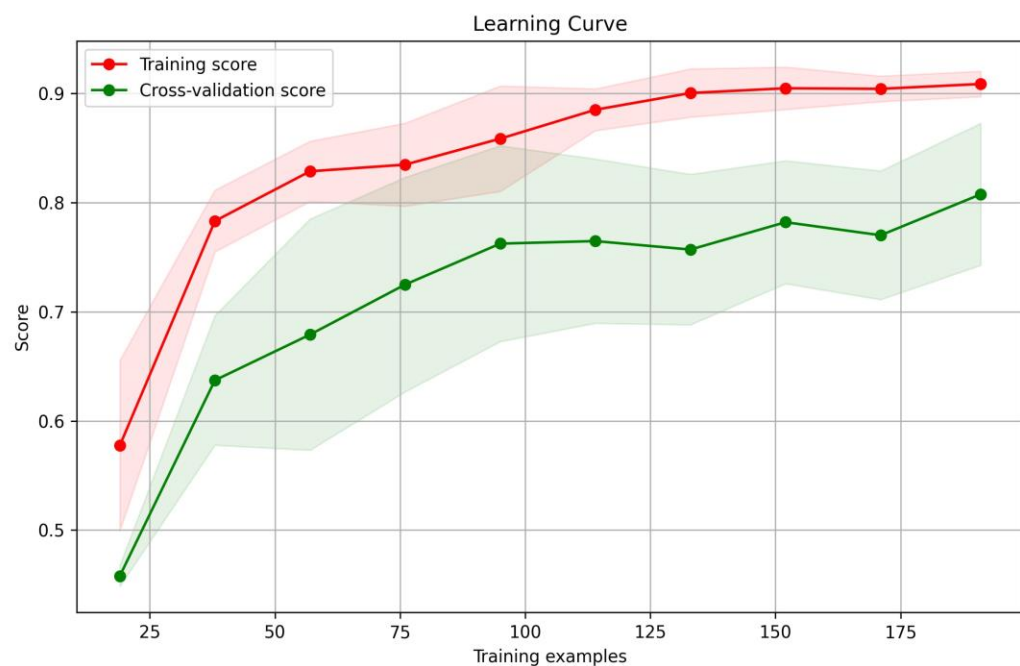
RFECV

RFECV (Recursive feature elimination with cross-validation)
to funkcja dobiera optymalną kombinację zmiennych do modelu
kierując się maksymalizacją oceny walidacji krzyżowej.

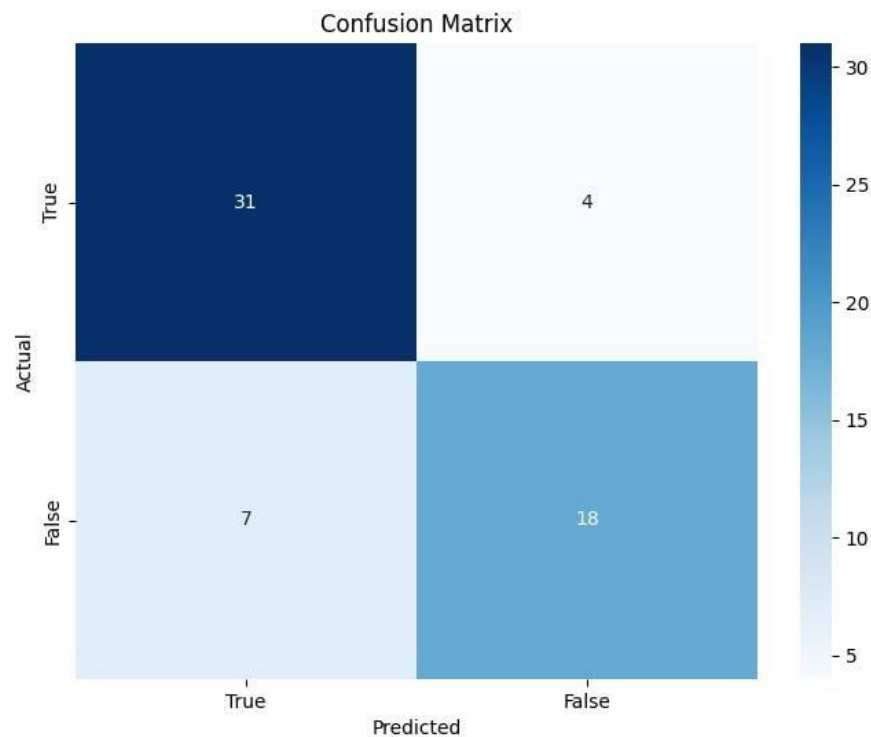
RÓWNOWAŻENIE DANYCH



OCENA MODELU



OCENA MODELU



False Positive Rate: 0,11

False Negative Rate: 0,24

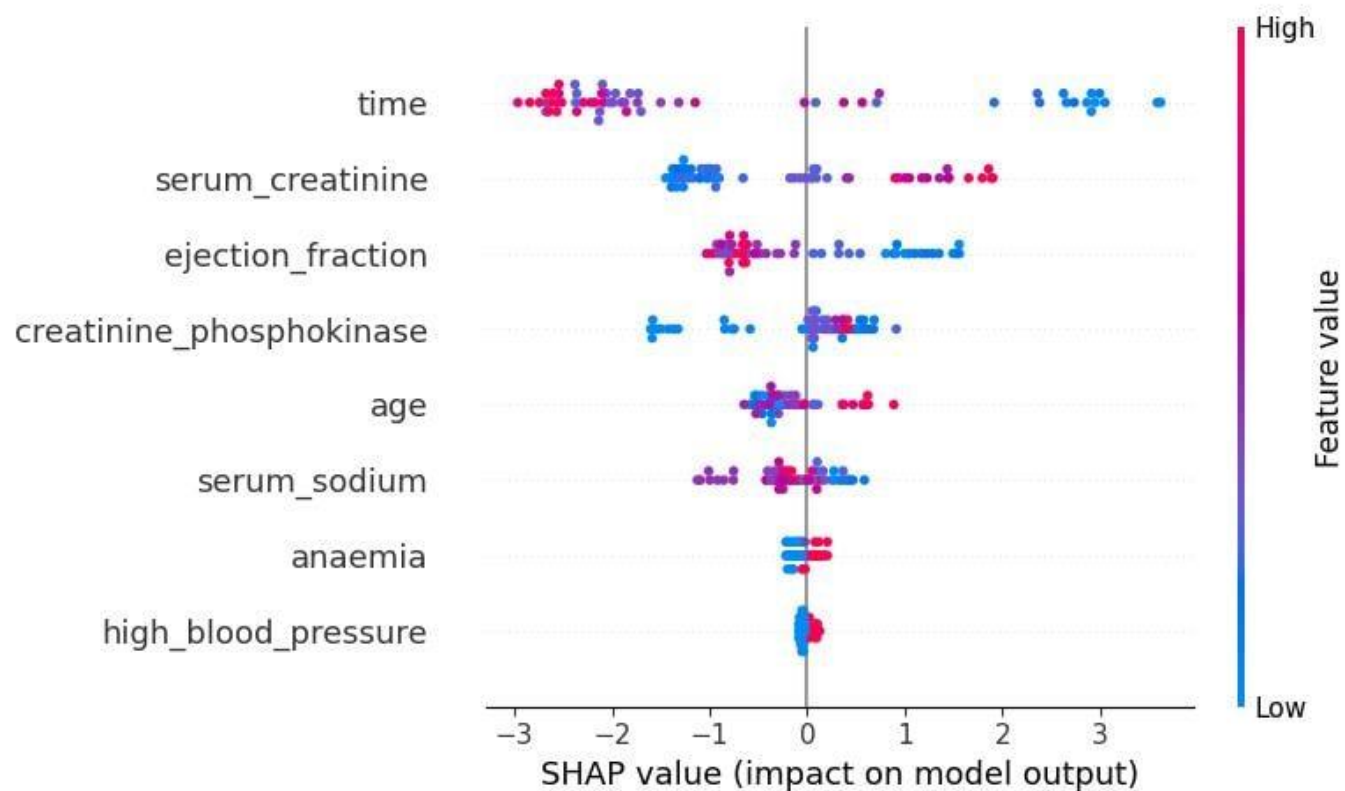
Accuracy: 0,82

Precision: 0,82

Recall: 0,72

F1-score: 0,77

SHAP



JAK DZIAŁA PROGRAM

- Interfejs CLI
- Użytkownik podaje wartości zmiennych
- W odpowiedzi zwracany jest poziom prawdopodobieństwa