

Zadanie 4

Bartosz Kozłowski

Korzystając z danych z zadania 2 stworzono pełne drzewo decyzyjne korzystając z wszystkich parametrów oraz wyznaczono jego skuteczność w przypadku ponownego podstawienia i krosvalidacji($k = 5$) . Otrzymano poniższe wyniki:

ACC w przypadku powtórzonego podstawienia

1.0

ACC w przypadku krosvalidacji

0.8709523809523809

Następnie wykorzystując algorytm kosztu-łożoności znaleziono i narysowano optymalne drzewo, oraz wyznaczono jego skuteczność w obu przedstawionych powyżej przypadkach.

ACC w przypadku powtórzonego podstawienia - optymalne drzewo

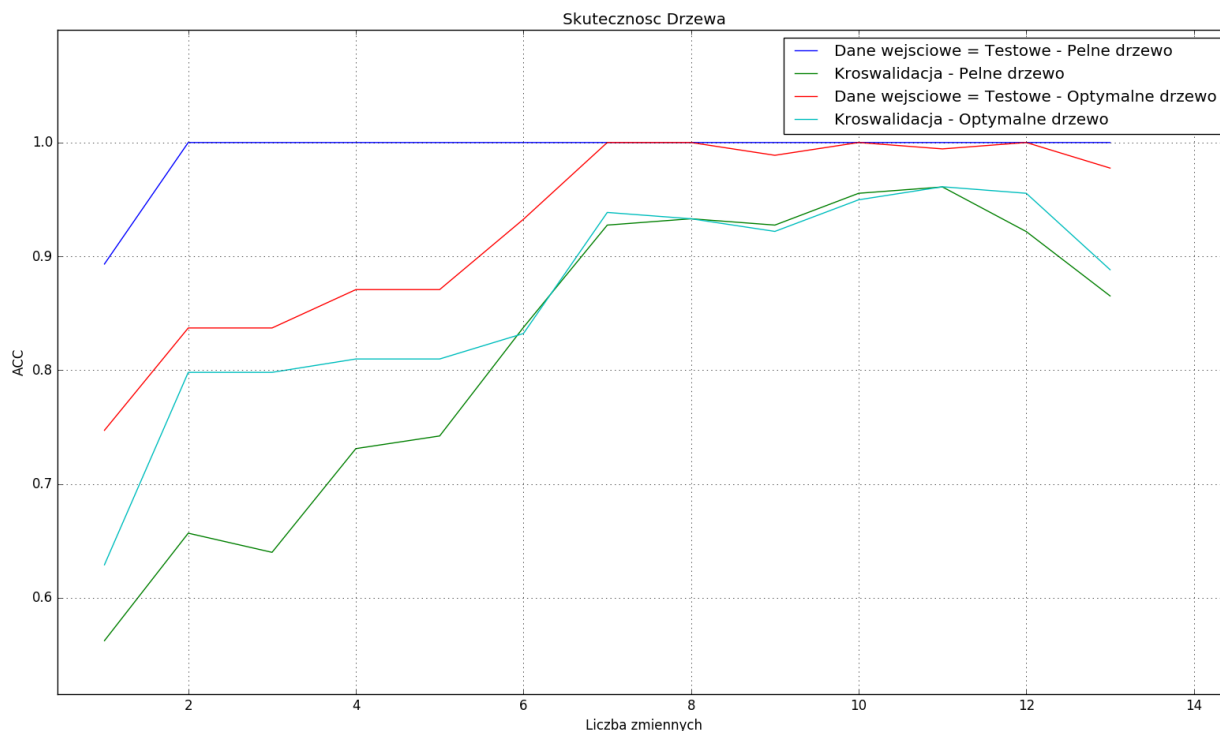
0.9550561797752809

ACC w przypadku krosvalidacji - optymalne drzewo

0.882063492063492

Można zauważyć, że w przypadku powtórzonego podstawienia wyniki są najlepsze, jest to spowodowane tym że model jest dopasowany do danych, które są na nim sprawdzane. W przypadku testowania z użyciem innych danych skuteczność byłaby niższa. Widać, że zastosowanie optymalnego drzewa pozwala na uzyskanie lepszej skuteczności. Rysunki drzewa pełnego i optymalnego są załączone wraz z tym raportem w formie pdf, odpowiednio tree.pdf i tree_best.pdf.

Stworzono wykres skuteczności w zależności od liczby wykorzystanych parametrów od 1 do 13(wszystkich). Skuteczność przedstawiono dla drzewa pełnego i optymalnego dla przypadku powtórzonego podstawienia oraz krosvalidacji. Z wykresu wynika że najlepszą skuteczność osiąga się dla dużej ilości parametrów(10 – 12).



Narysowano także wykres zależności rozmiaru drzewa pełnego i optymalnego od liczby użytych parametrów. Rozmiary optymalnego drzewa zawsze niższe lub równe od pełnego drzewa. Rozmiary obu drzew zbliżają się do siebie dla dużej ilości użytych parametrów.

