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 kroswalidacji(k = 5) . Otrzymano poniższe wyniki:

ACC w przypadku powtornego podstawienia

1.0

ACC w przypdaku kroswalidacji

0.8709523809523809

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

ACC w przypadku powtornego podstawienia - optymalne drzewo

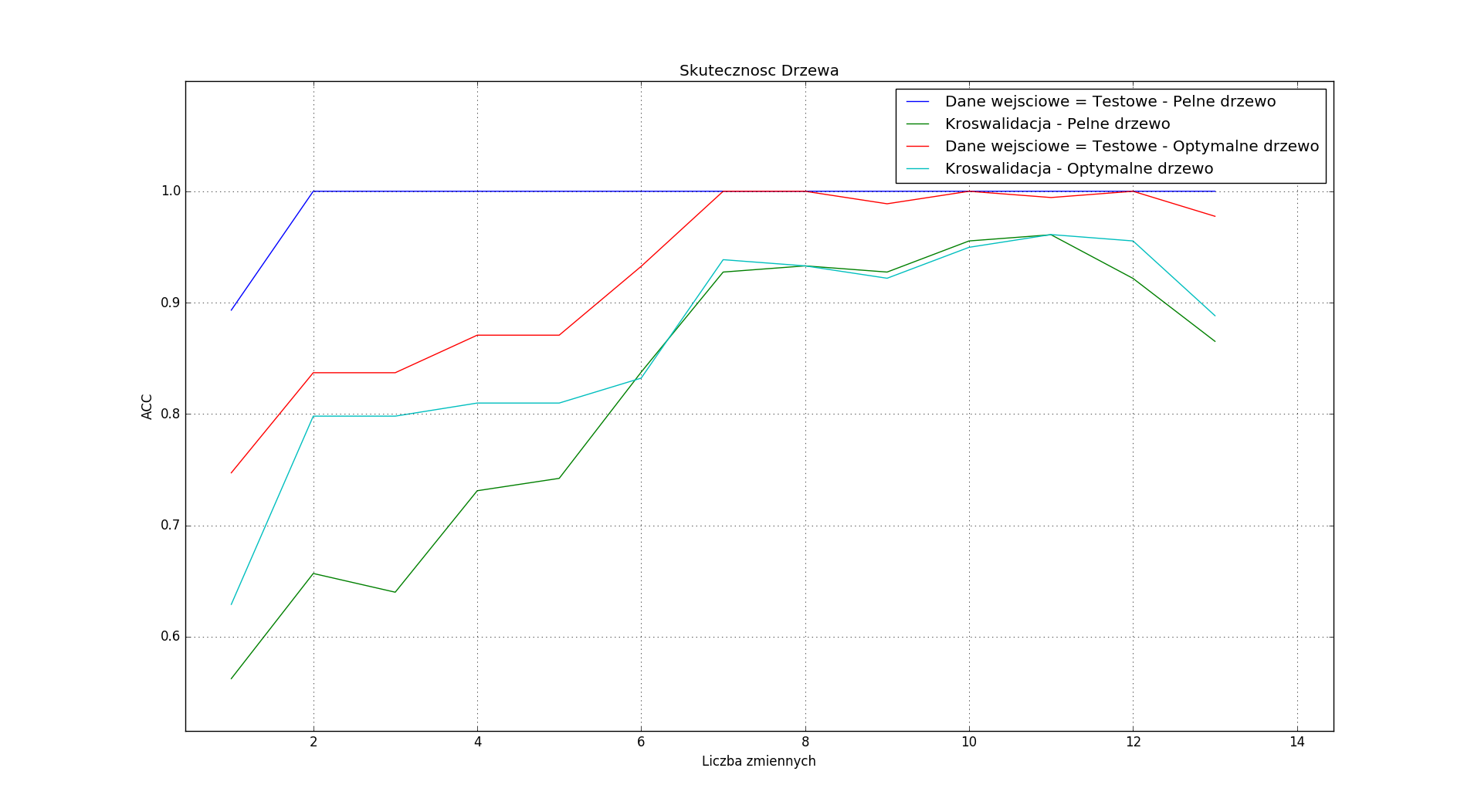
0.9550561797752809

ACC w przypdaku kroswalidacji - optymalne drzewo

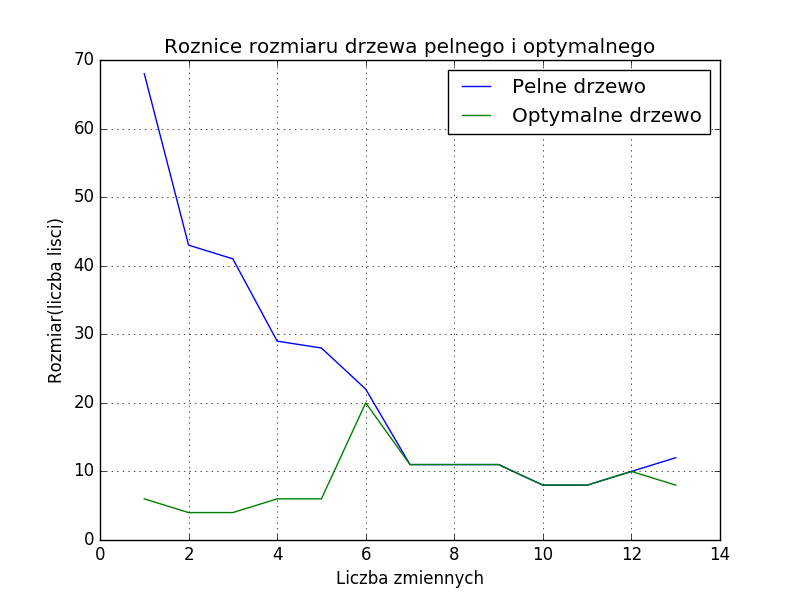
0.882063492063492

Można zauważyć, że w przypadku powtórnego 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 powtórnego podstawienia oraz kroswalidacji. 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.



Pełne Drzewo

Optymalne Drzewo

