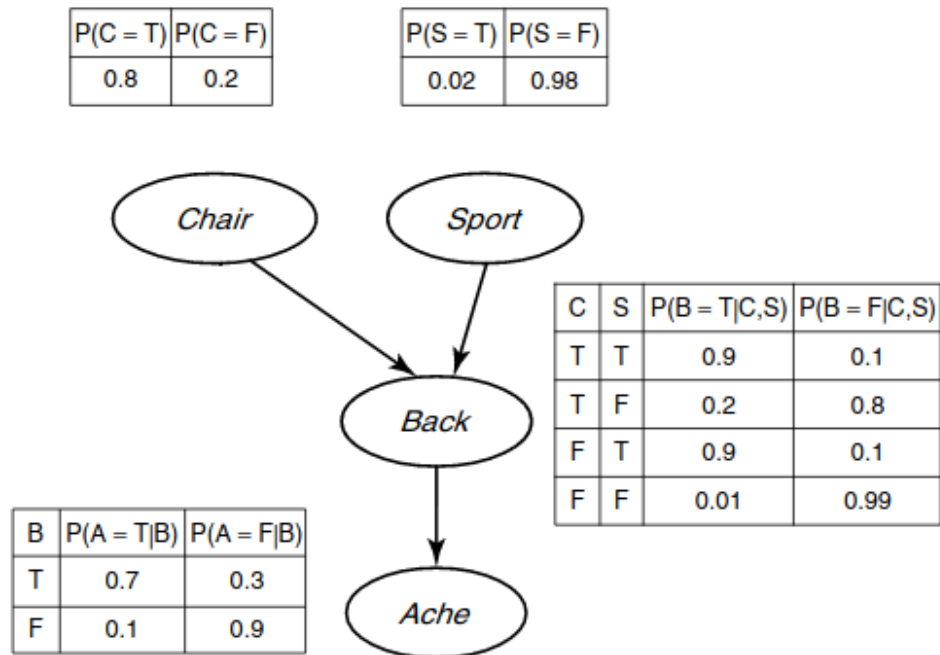


WSI lab 7

Proszę zaimplementować losowy generator danych, który działa zgodnie z rozkładem reprezentowanym przez daną sieć bayesowską. Sieć ta opisuje zależności między (zero-jedynkowymi) zmiennymi losowymi i dana jest w postaci opisu grafu połączeń oraz tabel prawdopodobieństw warunkowych. Wejście algorytmu: ile przykładów wygenerować, opis struktury prostej sieci (według własnego formatu) oraz tabele prawdopodobieństw należy wczytać z pliku tekstowego. Wyjście: plik tekstowy z przykładami. Strukturę sieci i tabele prawdopodobieństw widać na rysunku.

Klasa to „Ache” (czy bolą plecy), pozostałe węzły to atrybuty („Back” to uszkodzenie kręgosłupa (drobne, czasem nie skutkujące bólem)). Wytworzony zbiór podzielić i użyć do treningu i testowania klasyfikatora utworzonego na wcześniejszych ćwiczeniach. Jakie uzyskujemy wyniki? Wnioski?

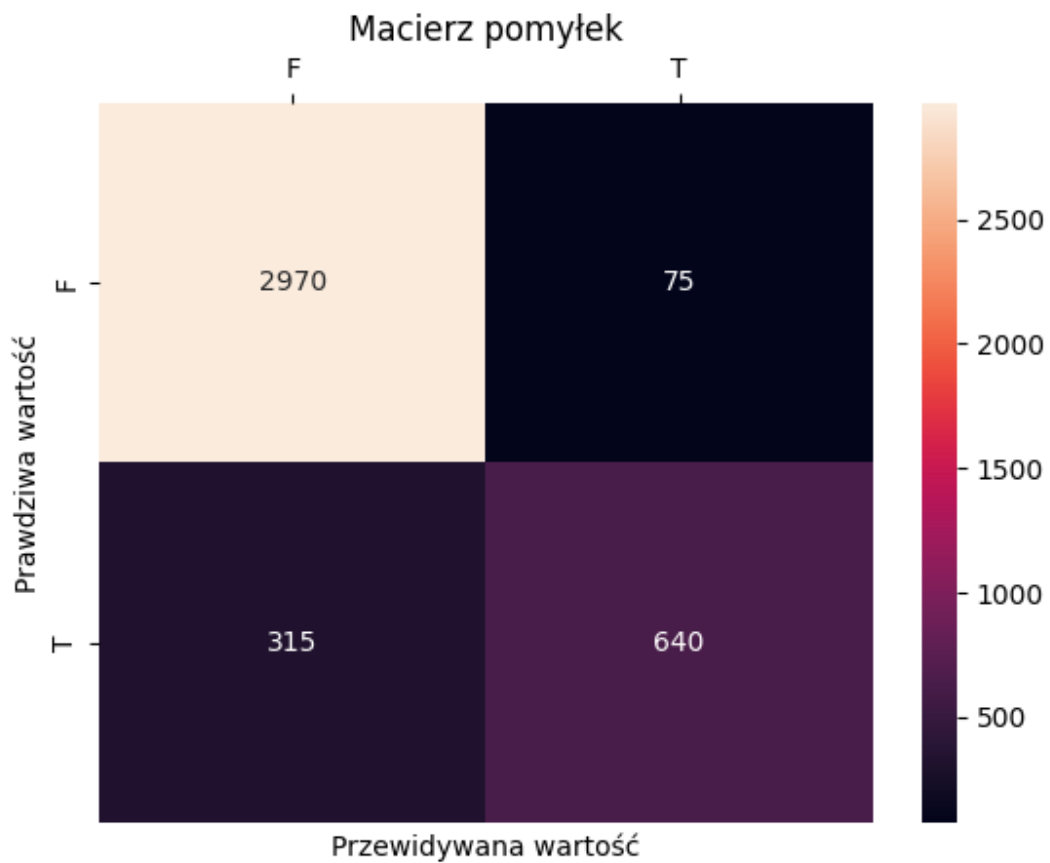


Po zaimplementowaniu generatora danych zostało wygenerowane 20 000 przykładów. Zostały one podzielone na 2 równe grupy względem sposobu losowania.

1. losowa liczba z przedziału $[0, 1]$ jest mniejsza od $P(X=F)$ dla każdego z wierzchołków
2. losowa liczba z przedziału $[0, 1]$ jest mniejsza od $P(X=T)$ dla każdego z wierzchołków

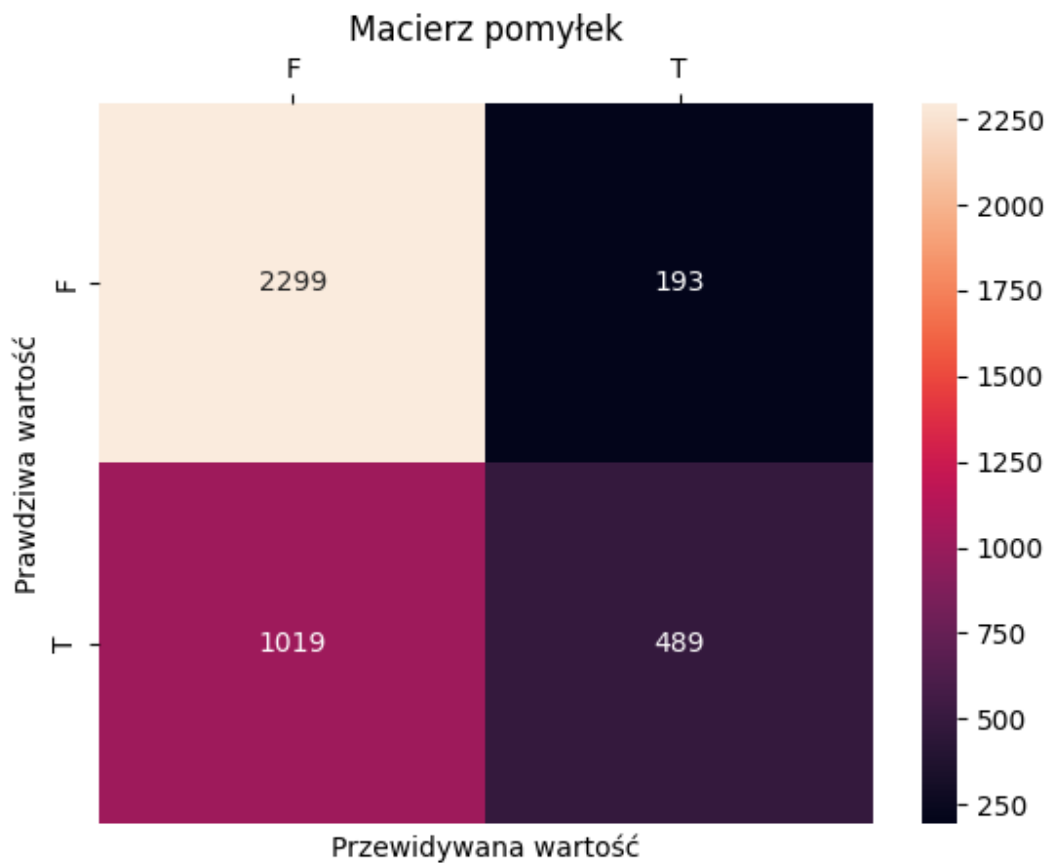
Następnie oba zbiory, zostały podzielone w stosunku 3 do 2 na zbiór treningowy i testowy. Jako klasyfikator wykorzystałem ID3, zaimplementowany na wcześniejszych ćwiczeniach.

Zauważyłem, że w przypadku dla danych wygenerowanych zgodnie z opisem w pkt. 1 klasyfikator osiągał bardzo wysoką dokładność rzędu 90%.



90,25% skuteczności dla 4000 przykładów w zbiorze testowym z pkt 1

Natomiast dla zbioru przykładów wygenerowanych w punkcie drugim, dokładność była niższa i waha się w przedziale 60-70%.



67,7% skuteczności dla 4000 przykładów w zbiorze testowym z pkt 2

Dla zadanego przykładu, klasyfikator osiągnął dużą dokładność w obu przypadkach. W tym problemie pomimo niewielkiej liczby atrybutów, klasyfikator nie ma większych problemów z odpowiednim pogrupowaniem przykładów. Wynikające różnice mogą wynikać z wartości prawdopodobieństw warunkowych.