

---

# PROBABILISTYCZNE MODELE GRAFOWE

## Predykcja stosowanej metody antykoncepcji

---

**Marcel Cielinski**

Wydział Informatyki i Zarządzania  
Politechnika Wrocławska  
236747@student.pwr.edu.pl

**Weronika Pawlak**

Wydział Informatyki i Zarządzania  
Politechnika Wrocławska  
236740@student.pwr.edu.pl

### 1 Wstęp

Problem polega na predykcji obecnie stosowanej metody antykoncepcji (brak, metody długoterminowe lub krótkoterminowe) na podstawie cech demograficznych i społeczno-ekonomicznych. Podjęto również dodatkowe zadanie dotyczące predykcji liczby dzieci. Rozwiązanie oparto na probabilistycznych modelach grafowych, takich jak sieci bayesowskie i model naiwnego Bayesa.

### 2 Zbiór danych

W celu realizacji zadania wykorzystano zbiór danych [1], stanowiący podzbiór krajowego badania na temat antykoncepcji, przeprowadzonego w Indonezji w 1987 r.

Atrybuty:

1. Wiek żony
  - (liczbowy, ciągły)
2. Poziom edukacji żony
  - (kategoryczny) 1=niski, 2, 3, 4=wysoki
3. Poziom edukacji męża
  - (kategoryczny) 1=niski, 2, 3, 4=wysoki
4. Liczba posiadanych dzieci
  - (liczbowy, ciągły)
5. Religia żony
  - (binarny) 0=Nie-Islam, 1=Islam
6. Czy żona obecnie pracuje?
  - (binarny) 0=Tak, 1=Nie
7. Zajęcie męża
  - (kategoryczny) 1, 2, 3, 4
8. Wskaźnik poziomu życia
  - (kategoryczny) 1=niski, 2, 3, 4=wysoki

### 9. Ekspozycja medialna

- (binarny) 0=Dobra, 1=Słaba

### 10. Stosowana metoda antykoncepcji

- (klasa) 1=brak, 2=długoterminowa, 3=krótkoterminowa

## 3 Eksploracyjna analiza danych

Zakres, rozkłady i zależności zmiennych zostały zgłębione w ramach eksploracyjnej analizy danych. Dogłębne rozumienie zbioru ułatwia planowanie i konfiguracje eksperymentów.

### 3.1 Opis zmiennych

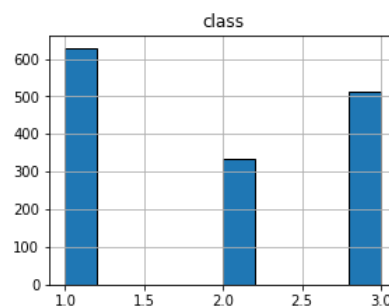
	wife age	n children
count	1473.000000	1473.000000
mean	32.538357	3.261371
std	8.227245	2.358549
min	16.000000	0.000000
25%	26.000000	1.000000
50%	32.000000	3.000000
75%	39.000000	4.000000
max	49.000000	16.000000

Tabela 1: Charakterystyka zmiennych ciągłych

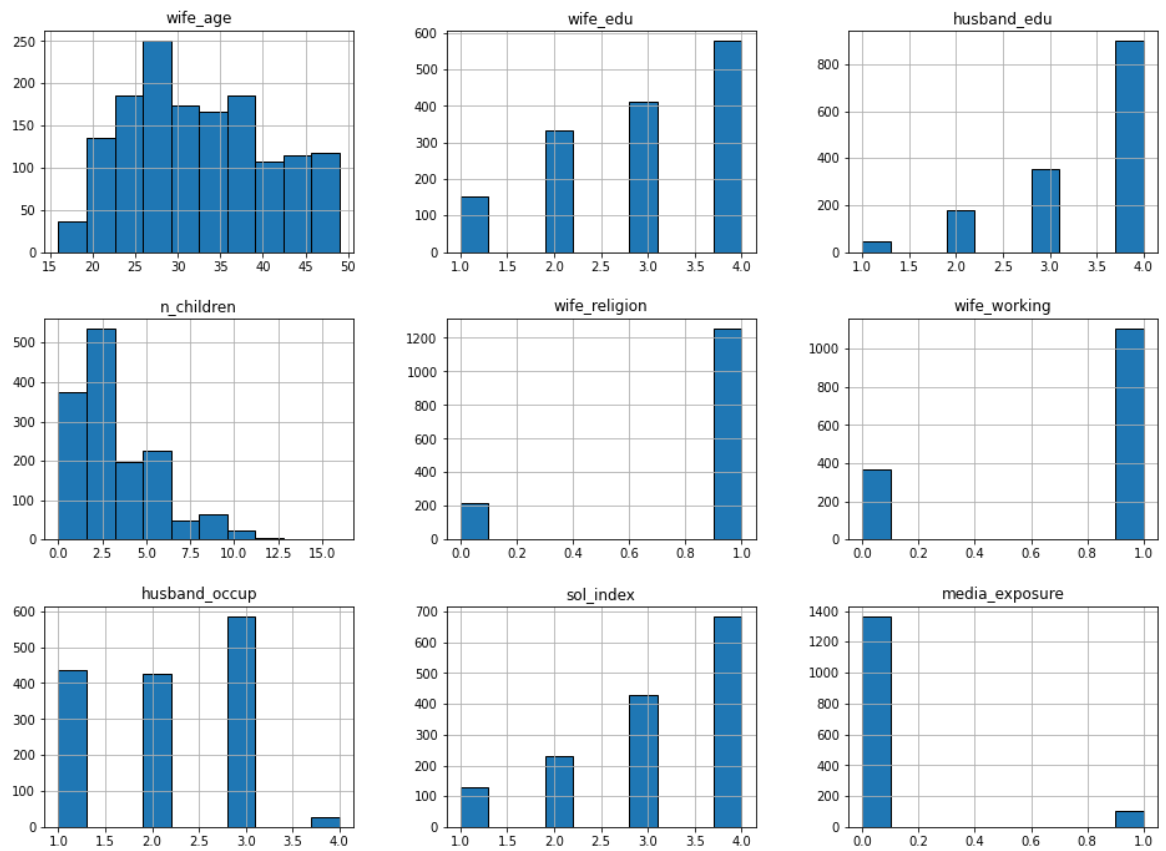
	wife edu	husband edu	wife religion	wife working	husband occup	sol index	media exposure	class
count	1473	1473	1473	1473	1473	1473	1473	1473
unique	4	4	2	2	4	4	2	3
top	high	high	islam	no	upper middle	high	good	no use
freq	577	899	1253	1104	585	684	1364	629

Tabela 2: Charakterystyka zmiennych kategoriycznych

### 3.2 Rozkłady zmiennych

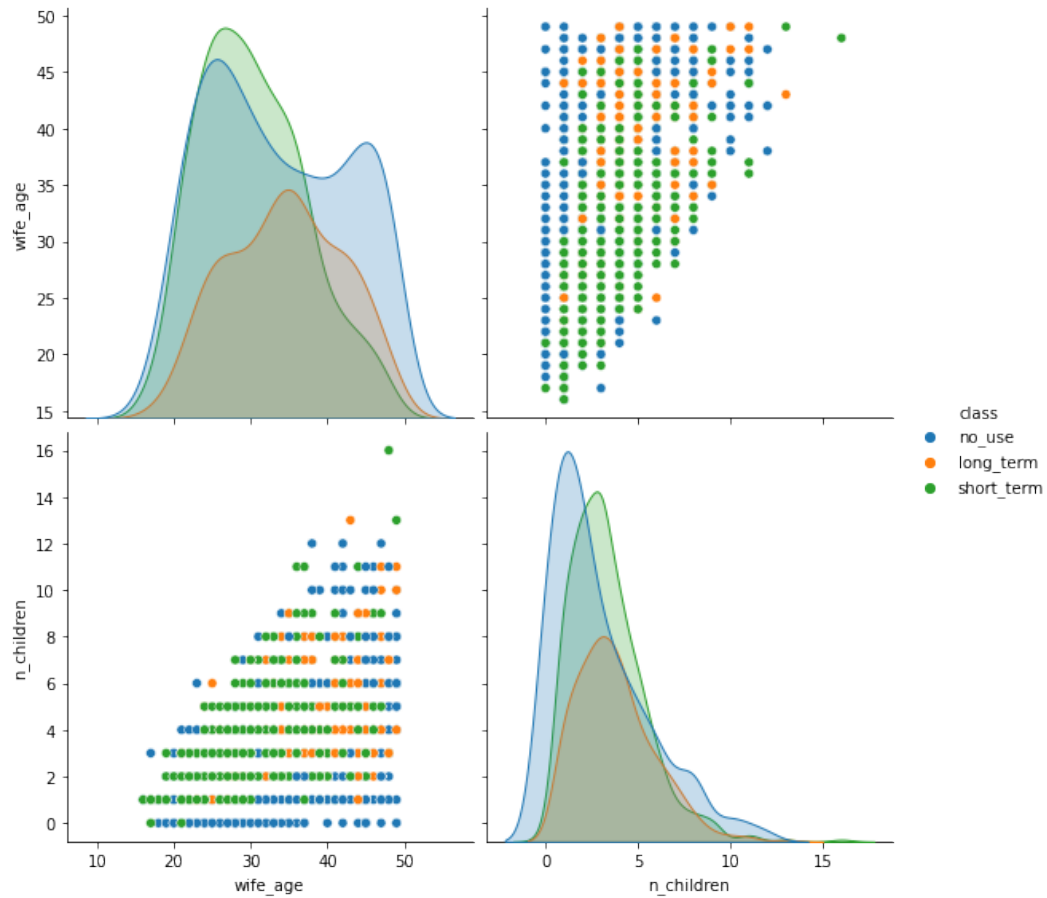


Rysunek 1: Rozkład klas

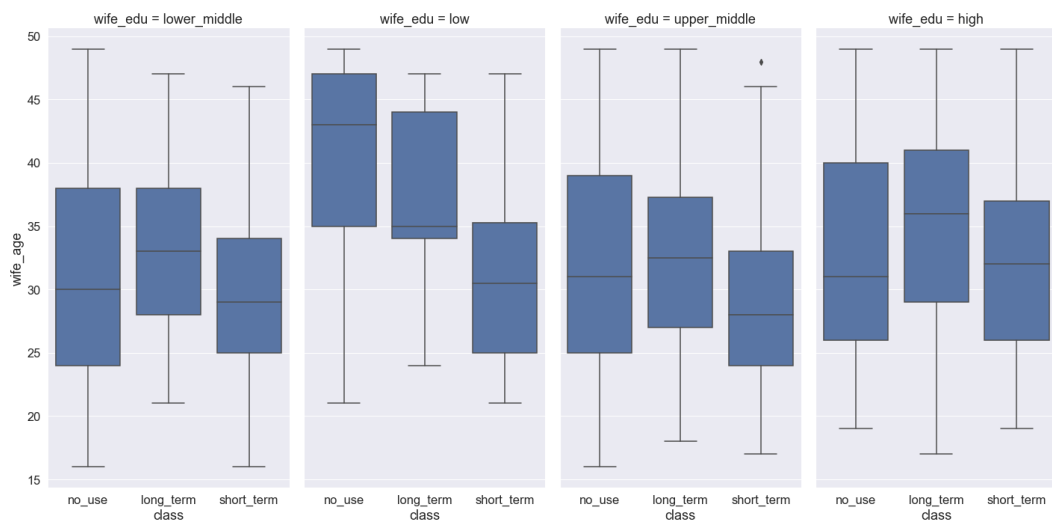


Rysunek 2: Rozkłady zmiennych

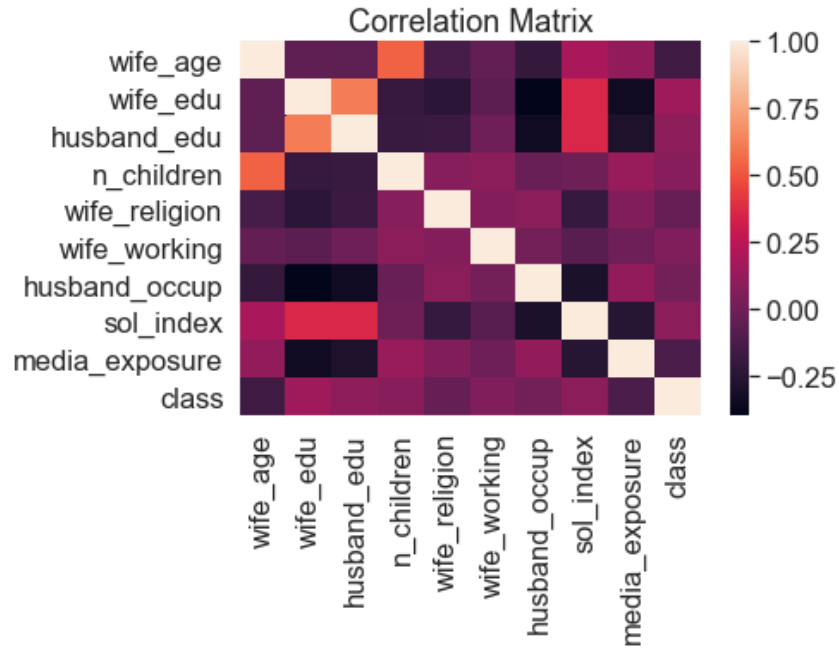
### 3.3 Zależności parami



Rysunek 3: Zależność zmiennych ciągłych



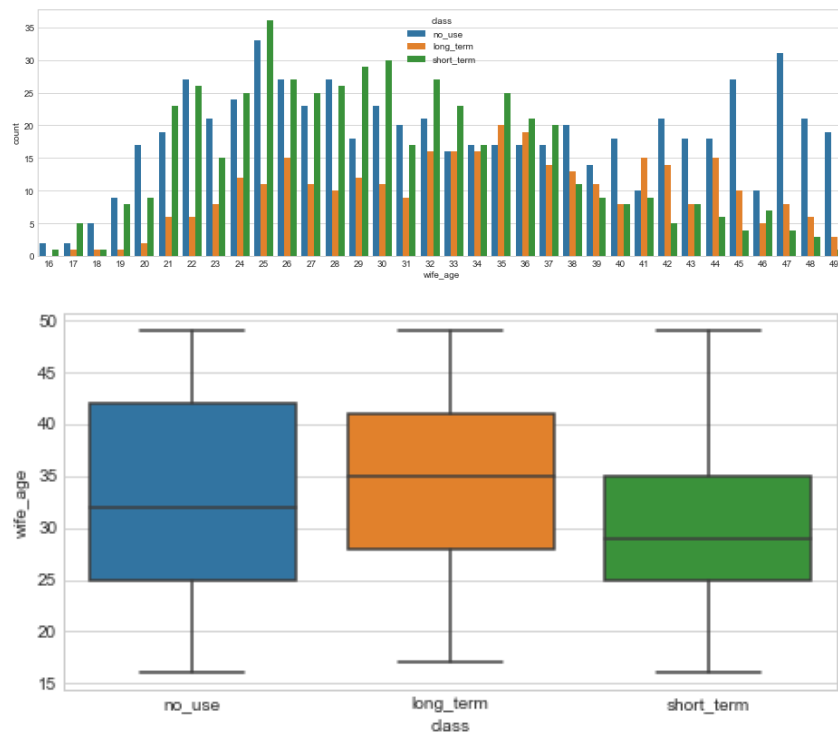
Rysunek 4: Zależność klasy od wieku i poziomu wykształcenia respondentki



Rysunek 5: Macierz korelacji

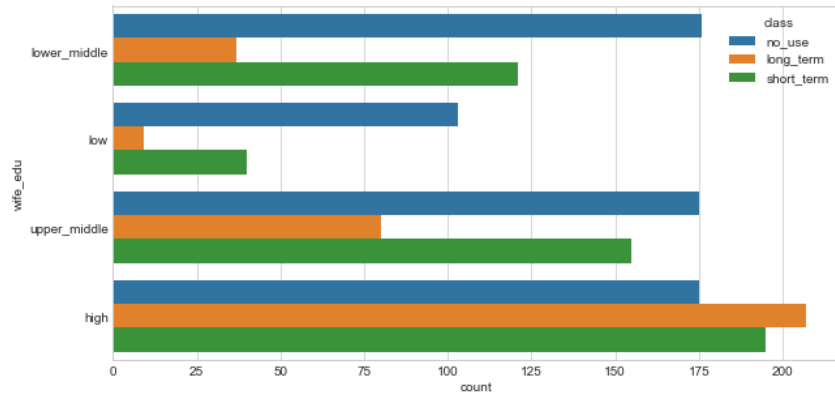
### 3.4 Charakterystyka wybranych zmiennych

#### 3.4.1 Wiek respondentki



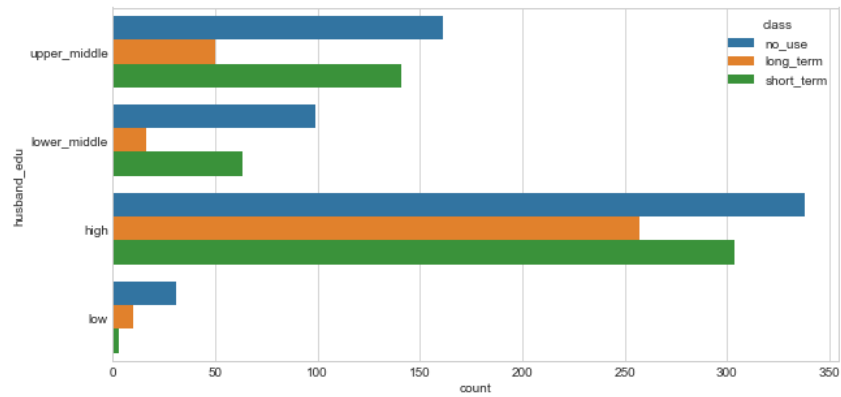
Rysunek 6: Zależność klasy od wieku respondentki

### 3.4.2 Poziom edukacji respondentki



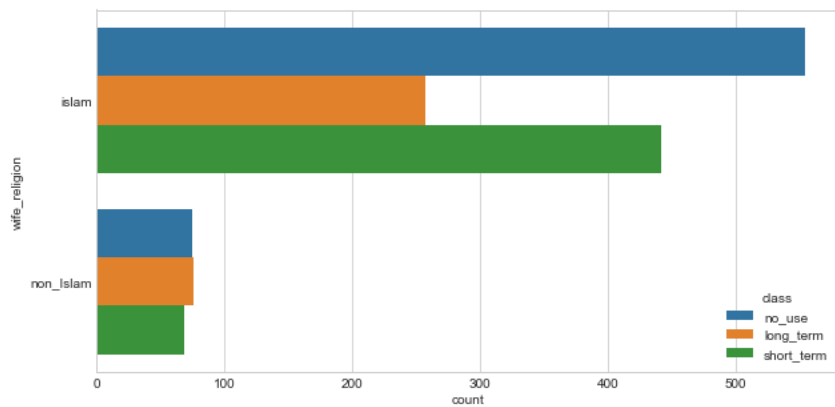
Rysunek 7: Zależność klasy od poziomu edukacji respondentki

### 3.4.3 Poziom edukacji męża respondentki



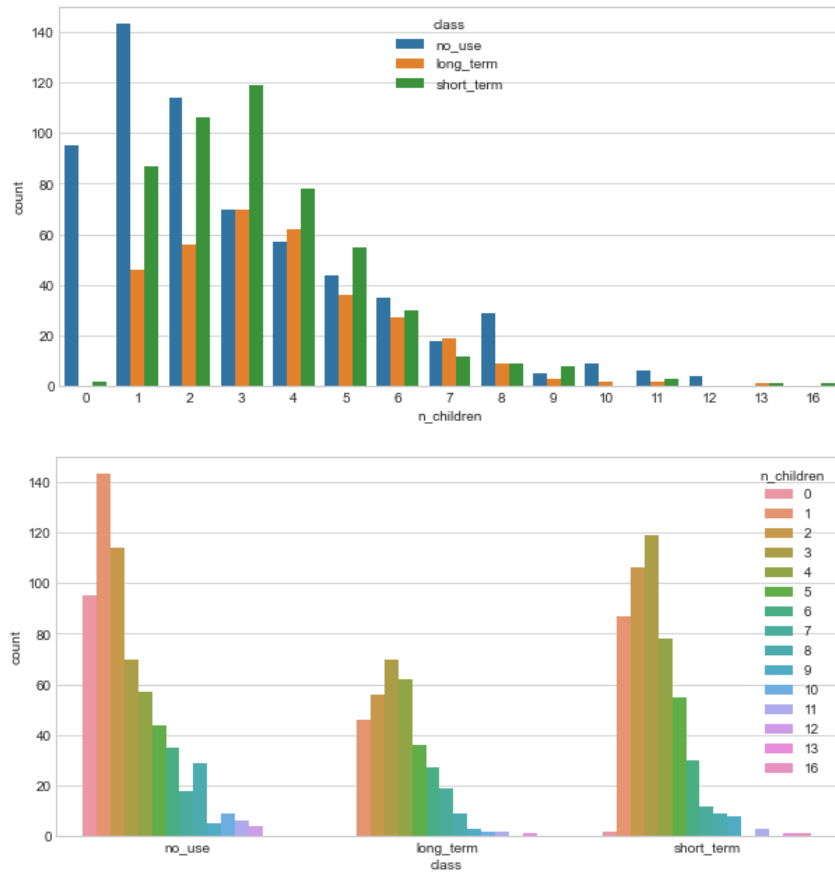
Rysunek 8: Zależność klasy od poziomu edukacji męża respondentki

### 3.4.4 Religia respondentki



Rysunek 9: Zależność klasy od religii respondentki

### 3.4.5 Liczba dzieci



Rysunek 10: Zależność klasy od liczby posiadanych dzieci

### 3.5 Wnioski z analizy

Na podstawie eksploracyjnej analizy danych, możemy stwierdzić wysoką korelację zmiennych:

- liczba dzieci + wiek respondentki
- edukacja respondentki + edukacja męża
- edukacja respondentki + wskaźnik poziomu życia
- edukacja męża + wskaźnik poziomu życia
- edukacja męża + zajęcie męża
- edukacja respondentki + zajęcie męża

Z dodatkowych analiz ukierunkowanych na badanie zależności klasy od jednej zmiennej widać, że liczba dotychczas posiadanych dzieci i wiek respondentki istotnie wpływają na wybraną metodę antykoncepcji. Wyniki są zgodne z intuicją - kobiety starsze, jak i kobiety z dużą liczbą dzieci częściej wybierają metody długoterminowe. Odnotowano też wpływ poziomu edukacji na zmienną określającą klasę.

## 4 Modele

### 4.1 Sieć Bayesowska

Sieć Bayesowska jest probabilistycznym modelem grafowym, który reprezentuje zbiór zmiennych i ich warunkowe zależności za pomocą skierowanego grafu acyklicznego (DAG). Sieci bayesowskie idealnie nadają się do analizowania zaistniałych zdarzeń i przewidywania prawdopodobieństwa, że czynnikiem sprawczym była jedna z kilku możliwych przyczyn.

Formalnie sieci Bayesowskie są skierowanymi grafami acyklicznymi, których węzły reprezentują zmienne, a krawędzie zależności warunkowe. Węzły, które nie są połączone (żadna ścieżka nie łączy jednego węzła z drugim) reprezentują zmienne, które są warunkowo niezależne. Każdy węzeł jest powiązany z funkcją prawdopodobieństwa, która przyjmuje na wejściu określony zestaw wartości zmiennych nadrzędnych węzła i daje na wyjściu prawdopodobieństwo (lub rozkład prawdopodobieństwa) zmiennej reprezentowanej przez węzeł.

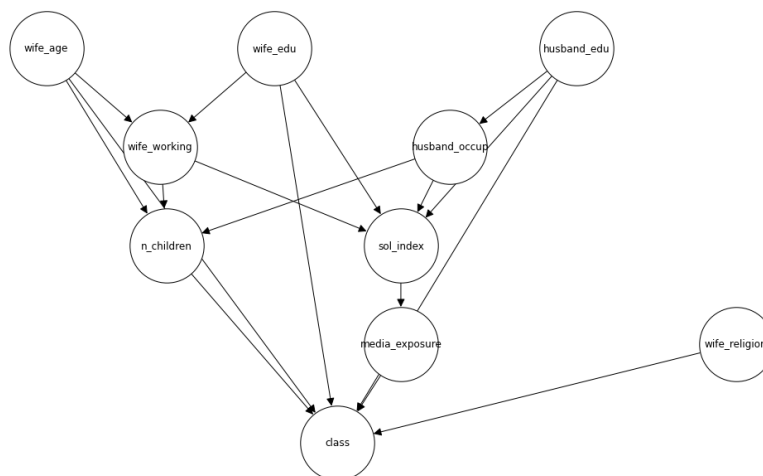
### 4.2 Struktura sieci

Ważnym elementem w sieciach Bayesowskich jest zatem struktura. Jeśli problem jest prosty lub posiadamy odpowiednią wiedzę dziedzinową istnieje możliwość stworzenia struktury wedle własnych intuicji. W przeciwnym przypadku można wykorzystać odpowiednie algorytmy, które wspomogą generowanie struktury sieci z danych.

Do problemu generowania grafu zostało wykorzystanych kilka podejść, które zostaną opisane w kolejnych podrozdziałach.

#### 4.2.1 Budowanie grafu na podstawie własnej wiedzy

Najprostszym podejściem jest zbudowanie zależności zmiennych na podstawie własnych intuicji. W przypadku podjętego zagadnienia, zależności te nie są oczywiste, jednak ich określenie nie wymaga specjalistycznej wiedzy dziedzinowej. Podjęto próbę zbudowania sieci empirycznie, zostało zaproponowanych kilka struktur.

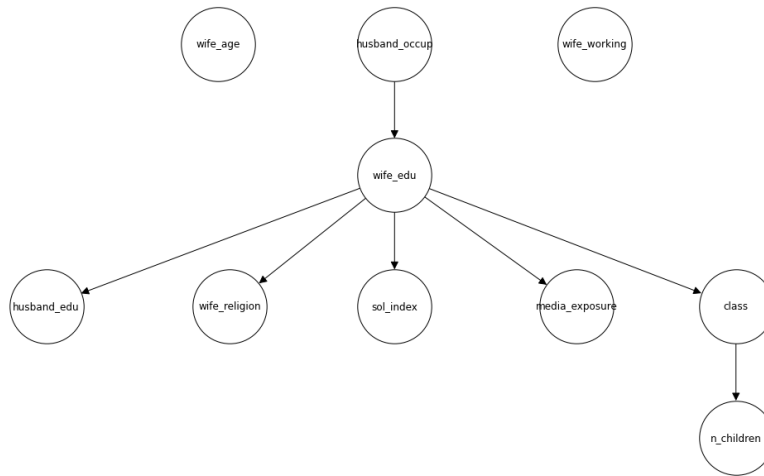


Rysunek 11: Przykładowa struktura sieci zbudowana empirycznie

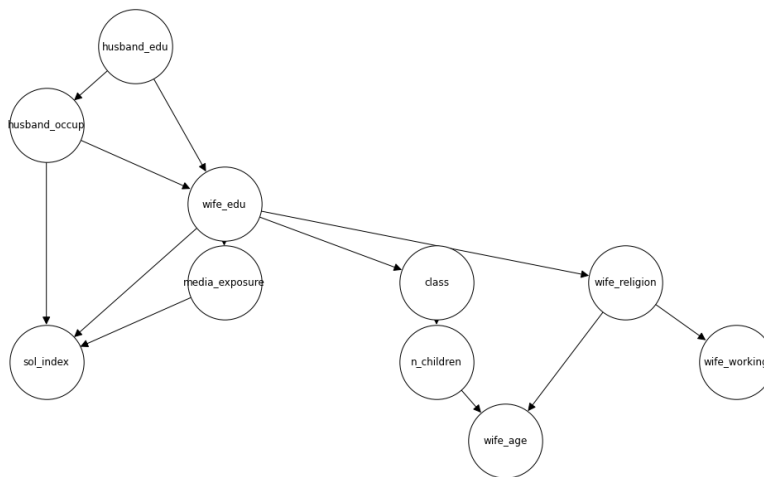


#### 4.2.2 Algorytm Tabu Search

Za pomocą algorytmu Tabu Search, który bazuje na scoringu można z danych wygenerować optymalną strukturę grafu. Algorytm modyfikuje sieć, aż do momentu osiągnięcia lokalnego maksimum (zgodnie z podaną metodą scoringu). Można podać wejściowy graf, w przypadku gdy parametr z nim związany jest pusty, algorytm rozpoczyna działanie na sieci niepołączonej. Przetestowano trzy metody scoringu.



Rysunek 12: Struktura sieci zbudowana za pomocą algorytmu Tabu Search z użyciem BIC score



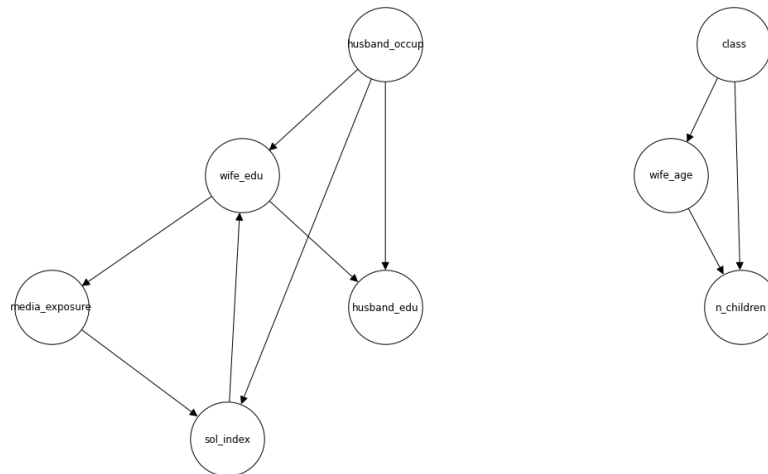
Rysunek 13: Struktura sieci zbudowana za pomocą algorytmu Tabu Search z użyciem K2 score



Rysunek 14: Struktura sieci zbudowana za pomocą algorytmu Tabu Search z użyciem BDeu score

#### 4.2.3 Algorytm PC

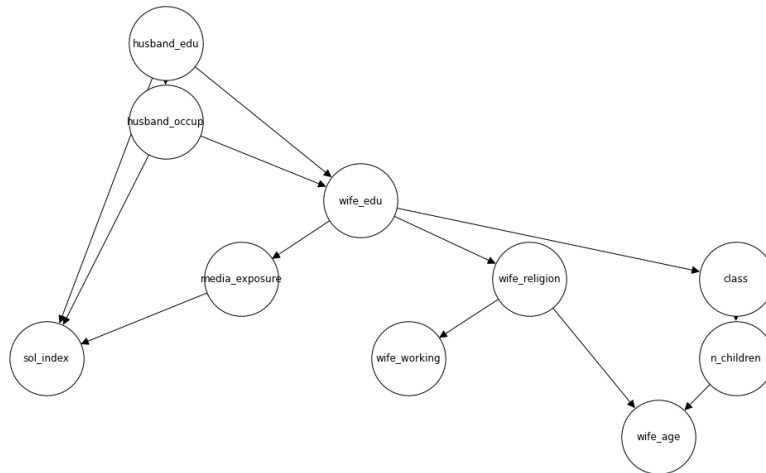
Algorytm dokonuje estymacji optymalnej struktury grafu poprzez identyfikację warunkowych zależności zmiennych, która odbywa się na podstawie testów statystycznych.



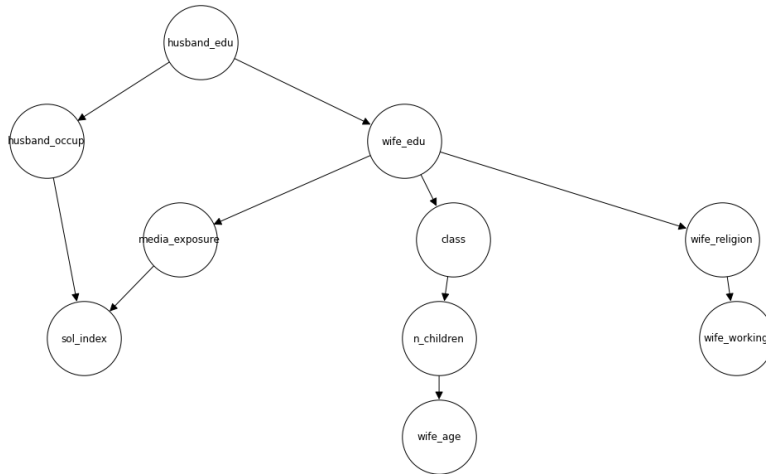
Rysunek 15: Struktura sieci zbudowana za pomocą algorytmu PC

#### 4.2.4 Podejścia hybrydowe

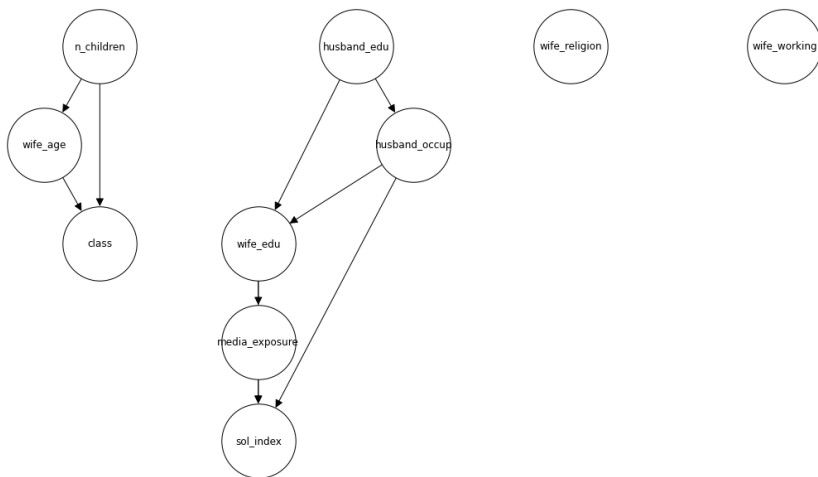
Przebadano również podejście hybrydowe w dwóch wersjach. Połączenie własnej struktury z metodą opartą na scoringu oraz budowa szkieletu za pomocą PC i Tabu Search do wyznaczenia kierunku krawędzi. Przykładowe grafy wygenerowane za pomocą podejść hybrydowych.



Rysunek 16: Struktura sieci zbudowana za pomocą własnej struktury i Tabu Search (K2 score)



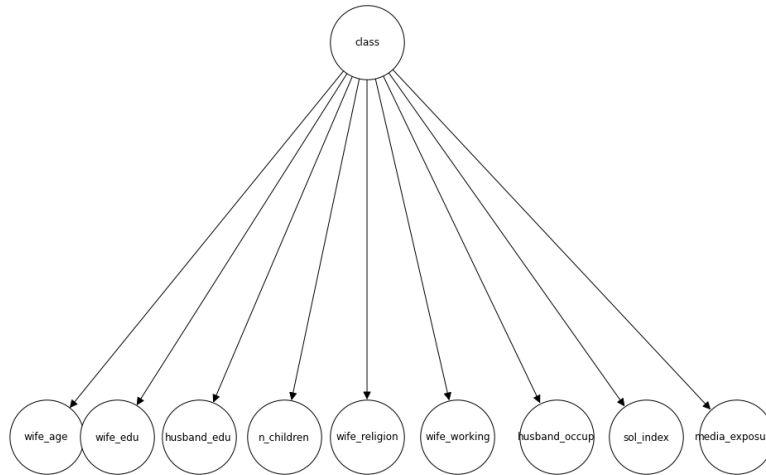
Rysunek 17: Struktura sieci zbudowana za pomocą własnej struktury i Tabu Search (BDeu score)



Rysunek 18: Struktura sieci zbudowana za pomocą szkieletu PC i Tabu Search (K2, BDeu score)

### 4.3 Naiwny Bayes

Model Naiwnego Bayesa to uproszczenie (specjalny przypadek) sieci bayesowskiej. Wykorzystuje się założenie, że wszystkie zmienne są warunkowo niezależne. Sieć naiwnego Bayesa dla podjętego zagadnienia wygląda następująco.



Rysunek 19: Struktura sieci dla modelu naiwnego Bayesa

## 5 Eksperymenty

### 5.1 Metodologia badań

1. Dyskretyzacja zmiennych ciągłych
  2. Budowa struktur z wykorzystaniem metod opisanych w punkcie 4.2
  3. Badania dla dwóch różnych zadań klasyfikacji:
    - predykcja metody antykoncepcyjnej (zadanie podstawowe)
    - predykcja liczby dzieci (zadanie dodatkowe)
- z wykorzystaniem różnych struktur i parametrów dyskretyzacji
4. Porównanie wyników modeli Bayesowskich i implementacji *scikit-learn*

## 5.2 Wyniki

	NETWORK	N BINS	ESTIMATOR	ACCURACY	PRECISION	RECALL	F1	TIME [s]
0	network_1	3_bins	BayesianEstimator	0.464	0.468	0.466	0.462	114.207
1	network_1	3_bins	MaxLLEstimator	0.464	0.468	0.466	0.462	110.766
2	network_2	3_bins	BayesianEstimator	0.431	0.380	0.372	0.352	34.065
3	network_2	3_bins	MaxLLEstimator	0.431	0.380	0.372	0.352	33.628
4	network_3	3_bins	BayesianEstimator	0.431	0.380	0.372	0.352	15.008
5	network_3	3_bins	MaxLLEstimator	0.431	0.380	0.372	0.352	14.659
6	network_4	3_bins	BayesianEstimator	0.431	0.380	0.372	0.352	11.140
7	network_4	3_bins	MaxLLEstimator	0.431	0.380	0.372	0.352	11.186
8	naive_bayes	3_bins	BayesianEstimator	0.458	0.461	0.464	0.456	3.855
9	naive_bayes	3_bins	MaxLLEstimator	0.458	0.461	0.464	0.456	4.028
10	hcs_bic	3_bins	BayesianEstimator	0.447	0.428	0.380	0.355	3.436
11	hcs_bic	3_bins	MaxLLEstimator	0.447	0.428	0.380	0.355	3.593
12	hcs_bdeu	3_bins	BayesianEstimator	0.451	0.497	0.433	0.435	4.458
13	hcs_bdeu	3_bins	MaxLLEstimator	0.451	0.497	0.433	0.435	4.695
14	hcs_k2	3_bins	BayesianEstimator	0.461	0.510	0.440	0.423	12.298
15	hcs_k2	3_bins	MaxLLEstimator	0.461	0.510	0.440	0.423	12.122
16	pc	3_bins	BayesianEstimator	0.410	0.270	0.356	0.307	8.576
17	pc	3_bins	MaxLLEstimator	0.410	0.270	0.356	0.307	8.388
18	hcs_network_1	3_bins	BayesianEstimator	0.461	0.510	0.440	0.423	10.625
19	hcs_network_1	3_bins	MaxLLEstimator	0.461	0.510	0.440	0.423	10.607
20	hcs_network_2	3_bins	BayesianEstimator	0.461	0.510	0.440	0.423	8.693
21	hcs_network_2	3_bins	MaxLLEstimator	0.461	0.510	0.440	0.423	8.126
22	hcs_network_3	3_bins	BayesianEstimator	0.461	0.510	0.440	0.423	10.174
23	hcs_network_3	3_bins	MaxLLEstimator	0.461	0.510	0.440	0.423	10.654
24	hcs_network_4	3_bins	BayesianEstimator	0.461	0.510	0.440	0.423	10.223
25	hcs_network_4	3_bins	MaxLLEstimator	0.461	0.510	0.440	0.423	10.221
26	hcs_naive_bayes	3_bins	BayesianEstimator	0.441	0.465	0.400	0.368	9.384
27	hcs_naive_bayes	3_bins	MaxLLEstimator	0.441	0.465	0.400	0.368	9.134
28	hcs_pc	3_bins	BayesianEstimator	0.410	0.270	0.356	0.307	8.476
29	hcs_pc	3_bins	MaxLLEstimator	0.410	0.270	0.356	0.307	8.973
30	network_1	6_bins	BayesianEstimator	0.539	0.519	0.494	0.494	357.917
31	network_1	6_bins	MaxLLEstimator	0.539	0.519	0.494	0.494	331.033
32	network_2	6_bins	BayesianEstimator	0.529	0.507	0.496	0.490	43.525
33	network_2	6_bins	MaxLLEstimator	0.529	0.507	0.496	0.490	42.790
34	network_3	6_bins	BayesianEstimator	0.529	0.507	0.496	0.490	25.440
35	network_3	6_bins	MaxLLEstimator	0.529	0.507	0.496	0.490	25.108
36	network_4	6_bins	BayesianEstimator	0.529	0.507	0.496	0.490	18.066
37	network_4	6_bins	MaxLLEstimator	0.529	0.507	0.496	0.490	17.823
38	naive_bayes	6_bins	BayesianEstimator	0.478	0.472	0.473	0.471	3.613
39	naive_bayes	6_bins	MaxLLEstimator	0.485	0.479	0.482	0.479	3.579
40	hcs_bic	6_bins	BayesianEstimator	0.492	0.479	0.424	0.396	3.252
41	hcs_bic	6_bins	MaxLLEstimator	0.492	0.479	0.424	0.396	3.436
42	hcs_bdeu	6_bins	BayesianEstimator	0.461	0.490	0.446	0.453	4.070
43	hcs_bdeu	6_bins	MaxLLEstimator	0.461	0.490	0.446	0.453	4.072
44	hcs_k2	6_bins	BayesianEstimator	0.556	0.559	0.519	0.520	12.954
45	hcs_k2	6_bins	MaxLLEstimator	0.556	0.559	0.519	0.520	12.732
46	pc	6_bins	BayesianEstimator	0.559	0.544	0.521	0.524	9.078
47	pc	6_bins	MaxLLEstimator	0.559	0.544	0.521	0.524	9.265
48	hcs_network_1	6_bins	BayesianEstimator	0.556	0.559	0.519	0.520	12.942
49	hcs_network_1	6_bins	MaxLLEstimator	0.556	0.559	0.519	0.520	12.754
50	hcs_network_2	6_bins	BayesianEstimator	0.556	0.559	0.519	0.520	10.948
51	hcs_network_2	6_bins	MaxLLEstimator	0.556	0.559	0.519	0.520	10.818
52	hcs_network_3	6_bins	BayesianEstimator	0.556	0.559	0.519	0.520	12.945
53	hcs_network_3	6_bins	MaxLLEstimator	0.556	0.559	0.519	0.520	12.750
54	hcs_network_4	6_bins	BayesianEstimator	0.556	0.559	0.519	0.520	12.923
55	hcs_network_4	6_bins	MaxLLEstimator	0.556	0.559	0.519	0.520	12.751
56	hcs_naive_bayes	6_bins	BayesianEstimator	0.580	0.573	0.560	0.565	9.718
57	hcs_naive_bayes	6_bins	MaxLLEstimator	0.580	0.573	0.560	0.565	9.636
58	hcs_pc	6_bins	BayesianEstimator	0.559	0.544	0.521	0.524	9.113
59	hcs_pc	6_bins	MaxLLEstimator	0.559	0.544	0.521	0.524	9.015
60	network_1	9_bins	BayesianEstimator	0.597	0.584	0.566	0.571	339.309
61	network_1	9_bins	MaxLLEstimator	0.597	0.584	0.566	0.571	323.134
62	network_2	9_bins	BayesianEstimator	0.508	0.501	0.479	0.481	58.233
63	network_2	9_bins	MaxLLEstimator	0.508	0.501	0.479	0.481	56.598
64	network_3	9_bins	BayesianEstimator	0.508	0.501	0.479	0.481	40.163
65	network_3	9_bins	MaxLLEstimator	0.508	0.501	0.479	0.481	39.130
66	network_4	9_bins	BayesianEstimator	0.508	0.501	0.479	0.481	26.455
67	network_4	9_bins	MaxLLEstimator	0.508	0.501	0.479	0.481	25.962
68	naive_bayes	9_bins	BayesianEstimator	0.498	0.487	0.488	0.486	3.707
69	naive_bayes	9_bins	MaxLLEstimator	0.498	0.487	0.488	0.486	3.702
70	hcs_bic	9_bins	BayesianEstimator	0.525	0.540	0.495	0.498	3.592
71	hcs_bic	9_bins	MaxLLEstimator	0.525	0.540	0.495	0.498	3.576
72	hcs_bdeu	9_bins	BayesianEstimator	0.485	0.502	0.475	0.483	4.403
73	hcs_bdeu	9_bins	MaxLLEstimator	0.485	0.502	0.475	0.483	4.381
74	hcs_k2	9_bins	BayesianEstimator	0.569	0.566	0.549	0.549	18.472
75	hcs_k2	9_bins	MaxLLEstimator	0.569	0.564	0.549	0.549	17.796
76	pc	9_bins	BayesianEstimator	0.556	0.529	0.516	0.503	11.668
77	pc	9_bins	MaxLLEstimator	0.556	0.529	0.516	0.503	11.554
78	hcs_network_1	9_bins	BayesianEstimator	0.569	0.566	0.549	0.549	18.538
79	hcs_network_1	9_bins	MaxLLEstimator	0.569	0.564	0.549	0.549	17.823
80	hcs_network_2	9_bins	BayesianEstimator	0.569	0.566	0.549	0.549	16.150
81	hcs_network_2	9_bins	MaxLLEstimator	0.569	0.564	0.549	0.549	16.233
82	hcs_network_3	9_bins	BayesianEstimator	0.569	0.566	0.549	0.549	18.124
83	hcs_network_3	9_bins	MaxLLEstimator	0.569	0.564	0.549	0.549	18.153
84	hcs_network_4	9_bins	BayesianEstimator	0.569	0.566	0.549	0.549	18.164
85	hcs_network_4	9_bins	MaxLLEstimator	0.569	0.564	0.549	0.549	18.237
86	hcs_naive_bayes	9_bins	BayesianEstimator	0.597	0.590	0.567	0.568	11.614
87	hcs_naive_bayes	9_bins	MaxLLEstimator	0.600	0.592	0.569	0.570	11.486
88	hcs_pc	9_bins	BayesianEstimator	0.556	0.529	0.516	0.503	11.689
89	hcs_pc	9_bins	MaxLLEstimator	0.556	0.529	0.516	0.503	11.596
	clf	n_bins	-	accuracy	precision	recall	f1	time
0	GaussianNB	3_bins		0.424	0.438	0.445	0.427	0.014
1	DecisionTree	3_bins		0.410	0.388	0.389	0.388	0.016
2	GaussianNB	6_bins		0.441	0.458	0.456	0.442	0.008
3	DecisionTree	6_bins		0.458	0.436	0.432	0.433	0.012
4	GaussianNB	9_bins		0.447	0.465	0.461	0.447	0.009
5	DecisionTree	9_bins		0.447	0.414	0.413	0.412	0.011

Tabela 3: Wyniki dla predykcji metody antykoncepcji

	NETWORK	N BINS	ESTIMATOR	ACCURACY	PRECISION	RECALL	F1	TIME [s]
0	network_1	3_bins	BayesianEstimator	0.834	0.323	0.327	0.317	191.747
1	network_1	3_bins	MaxLLEstimator	0.844	0.327	0.331	0.320	174.126
2	network_2	3_bins	BayesianEstimator	0.834	0.350	0.342	0.340	49.979
3	network_2	3_bins	MaxLLEstimator	0.831	0.347	0.341	0.338	47.546
4	network_3	3_bins	BayesianEstimator	0.834	0.350	0.342	0.340	21.111
5	network_3	3_bins	MaxLLEstimator	0.831	0.347	0.341	0.338	21.220
6	network_4	3_bins	BayesianEstimator	0.831	0.369	0.356	0.358	15.687
7	network_4	3_bins	MaxLLEstimator	0.827	0.365	0.355	0.357	14.923
8	naive_bayes	3_bins	BayesianEstimator	0.841	0.400	0.383	0.389	5.558
9	naive_bayes	3_bins	MaxLLEstimator	0.841	0.400	0.383	0.389	5.573
10	hcs_bic	3_bins	BayesianEstimator	0.871	0.290	0.333	0.310	4.997
11	hcs_bic	3_bins	MaxLLEstimator	0.871	0.290	0.333	0.310	4.833
12	hcs_bdeu	3_bins	BayesianEstimator	0.871	0.290	0.333	0.310	5.933
13	hcs_bdeu	3_bins	MaxLLEstimator	0.871	0.290	0.333	0.310	5.996
14	hcs_k2	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	15.558
15	hcs_k2	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	15.425
16	pc	3_bins	BayesianEstimator	0.871	0.290	0.333	0.310	13.467
17	pc	3_bins	MaxLLEstimator	0.871	0.290	0.333	0.310	13.465
18	hcs_network_1	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	15.930
19	hcs_network_1	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	15.413
20	hcs_network_2	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	12.403
21	hcs_network_2	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	11.979
22	hcs_network_3	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	15.303
23	hcs_network_3	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	15.245
24	hcs_network_4	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	15.484
25	hcs_network_4	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	15.212
26	hcs_naive_bayes	3_bins	BayesianEstimator	0.868	0.428	0.363	0.367	14.503
27	hcs_naive_bayes	3_bins	MaxLLEstimator	0.868	0.428	0.363	0.367	14.962
28	hcs_pc	3_bins	BayesianEstimator	0.871	0.290	0.333	0.310	11.642
29	hcs_pc	3_bins	MaxLLEstimator	0.871	0.290	0.333	0.310	11.797
30	network_1	6_bins	BayesianEstimator	0.569	0.264	0.269	0.264	447.184
31	network_1	6_bins	MaxLLEstimator	0.580	0.268	0.273	0.267	425.664
32	network_2	6_bins	BayesianEstimator	0.566	0.275	0.269	0.266	65.262
33	network_2	6_bins	MaxLLEstimator	0.549	0.271	0.262	0.260	63.248
34	network_3	6_bins	BayesianEstimator	0.566	0.275	0.269	0.266	37.478
35	network_3	6_bins	MaxLLEstimator	0.549	0.271	0.262	0.260	36.568
36	network_4	6_bins	BayesianEstimator	0.617	0.249	0.281	0.264	26.774
37	network_4	6_bins	MaxLLEstimator	0.624	0.252	0.284	0.267	26.064
38	naive_bayes	6_bins	BayesianEstimator	0.590	0.268	0.300	0.279	6.815
39	naive_bayes	6_bins	MaxLLEstimator	0.590	0.268	0.300	0.279	6.684
40	hcs_bic	6_bins	BayesianEstimator	0.647	0.262	0.298	0.277	5.206
41	hcs_bic	6_bins	MaxLLEstimator	0.647	0.262	0.298	0.277	5.261
42	hcs_bdeu	6_bins	BayesianEstimator	0.586	0.256	0.275	0.252	6.320
43	hcs_bdeu	6_bins	MaxLLEstimator	0.586	0.256	0.275	0.252	6.299
44	hcs_k2	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	19.510
45	hcs_k2	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	18.971
46	pc	6_bins	BayesianEstimator	0.634	0.305	0.295	0.287	16.676
47	pc	6_bins	MaxLLEstimator	0.634	0.305	0.295	0.287	16.256
48	hcs_network_1	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	19.396
49	hcs_network_1	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	19.209
50	hcs_network_2	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	16.239
51	hcs_network_2	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	16.058
52	hcs_network_3	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	19.424
53	hcs_network_3	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	19.000
54	hcs_network_4	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	19.269
55	hcs_network_4	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	18.906
56	hcs_naive_bayes	6_bins	BayesianEstimator	0.624	0.283	0.312	0.296	19.448
57	hcs_naive_bayes	6_bins	MaxLLEstimator	0.624	0.283	0.312	0.296	18.965
58	hcs_pc	6_bins	BayesianEstimator	0.627	0.249	0.280	0.261	13.952
59	hcs_pc	6_bins	MaxLLEstimator	0.620	0.250	0.278	0.261	13.720
60	network_1	9_bins	BayesianEstimator	0.495	0.258	0.253	0.253	463.321
61	network_1	9_bins	MaxLLEstimator	0.498	0.258	0.254	0.254	444.959
62	network_2	9_bins	BayesianEstimator	0.427	0.189	0.180	0.181	88.579
63	network_2	9_bins	MaxLLEstimator	0.420	0.212	0.203	0.204	85.153
64	network_3	9_bins	BayesianEstimator	0.427	0.189	0.180	0.181	59.079
65	network_3	9_bins	MaxLLEstimator	0.420	0.212	0.203	0.204	57.315
66	network_4	9_bins	BayesianEstimator	0.444	0.226	0.216	0.213	39.677
67	network_4	9_bins	MaxLLEstimator	0.444	0.229	0.216	0.213	38.208
68	naive_bayes	9_bins	BayesianEstimator	0.441	0.228	0.232	0.218	8.149
69	naive_bayes	9_bins	MaxLLEstimator	0.441	0.227	0.232	0.218	8.290
70	hcs_bic	9_bins	BayesianEstimator	0.451	0.191	0.211	0.197	6.059
71	hcs_bic	9_bins	MaxLLEstimator	0.451	0.191	0.211	0.197	6.095
72	hcs_bdeu	9_bins	BayesianEstimator	0.407	0.180	0.175	0.152	7.228
73	hcs_bdeu	9_bins	MaxLLEstimator	0.407	0.180	0.175	0.152	7.232
74	hcs_k2	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	28.956
75	hcs_k2	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	28.235
76	pc	9_bins	BayesianEstimator	0.492	0.267	0.246	0.238	26.557
77	pc	9_bins	MaxLLEstimator	0.492	0.267	0.246	0.238	24.740
78	hcs_network_1	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	28.658
79	hcs_network_1	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	28.033
80	hcs_network_2	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	25.451
81	hcs_network_2	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	25.038
82	hcs_network_3	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	28.569
83	hcs_network_3	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	28.232
84	hcs_network_4	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	28.894
85	hcs_network_4	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	27.790
86	hcs_naive_bayes	9_bins	BayesianEstimator	0.481	0.293	0.270	0.275	29.183
87	hcs_naive_bayes	9_bins	MaxLLEstimator	0.485	0.294	0.271	0.276	27.687
88	hcs_pc	9_bins	BayesianEstimator	0.464	0.198	0.225	0.210	18.670
89	hcs_pc	9_bins	MaxLLEstimator	0.461	0.197	0.223	0.209	18.595
	clf	n_bins	-	accuracy	precision	recall	f1	time
0	GaussianNB	3_bins		0.817	0.377	0.374	0.375	0.010
1	DecisionTree	3_bins		0.803	0.369	0.361	0.365	0.012
2	GaussianNB	6_bins		0.580	0.361	0.328	0.286	0.009
3	DecisionTree	6_bins		0.525	0.308	0.299	0.298	0.011
4	GaussianNB	9_bins		0.339	0.181	0.255	0.144	0.011
5	DecisionTree	9_bins		0.386	0.236	0.254	0.241	0.011

Tabela 4: Wyniki dla predykcji ilości posiadanych dzieci

W tabelach powyżej, przez network\_1 - network\_4 oznaczono struktury sieci zbudowane na podstawie własnej wiedzy. Przez hcs oznaczono sieci zbudowane za pomocą algorytmu hill climb search z użyciem różnych metod scoringu .Pc oznacza grafy zbudowane przy pomocy algorytmu pc. Przetestowano również modele baseline'owe - DecisionTree i GaussianNB z biblioteki *scikit-learn*. Pozostałe wpisy w tabeli dotyczą różnych konfiguracji hybrydowych.

## 6 Podsumowanie

Prosty model naiwnego Bayesa, przyjmujący warunkową niezależność wszystkich zmiennych, dobrze sprawdził się w obu zadaniach predykcji. Dla jednej z własnych struktur osiągnięto najlepszy wynik dla zadania predykcji metody antykoncepcyjnej. Inna w połączeniu z algorytmem tabu search dobrze sprawdziła się w dodatkowym zadaniu predykcyjnym. Niektóre ze stworzonych struktur powstały przez przycinanie, ich wyniki są takie same jak w przypadku oryginału. W pierwszym zadaniu słabo sprawdził się algorytm PC, w drugim tabu search z metodą oceny BDeu. Większa liczba binów dyskretyzacji pozytywnie wpływa na jakość predykcji, w przypadku, gdy dyskretyzowana wartość nie jest klasą. Wybór estymatora nie ma znaczącego wpływu na wyniki. W podstawowym zadaniu predykcyjnym modele baseline'owe osiągają słabsze wyniki niż modele bayesowskie. W drugim zadaniu wyniki są porównywalne.

## Literatura

- [1] T.-S. Lim. Contraceptive method choice data set.  
<https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>.