

Sprawozdanie z badań nad algorytmem genetycznym LAB 1

Kurs: Sztuczna Inteligencja i Inżynieria Wiedzy – Laboratorium

Prowadzący: mgr inż. Jan Jakubik

Autor: Konrad Liuras

Numer indeksu: 246752

Spis treści

1.	Wstęp – opis problemu	3
2.	Metodyka.....	3
3.	Wyniki badań	4
3.1	Badanie wpływu parametrów ilościowych.....	4
3.1.1	Badanie rozmiaru populacji.....	4
3.1.2	Badanie liczby pokoleń	7
3.1.3	Badanie rozmiaru turnieju	8
3.2	Porównanie operatorów selekcji.....	11
3.3	Zbadanie wpływu parametrów prawdopodobieństwa	13
3.3.1	Badanie wpływu prawdopodobieństwa krzyżowania.....	13
3.3.2	Badanie wpływu prawdopodobieństwa mutacji	14
3.4	Porównanie algorytmu genetycznego z metodą losową.....	16
4.	Podsumowanie.....	18
5.	Możliwe usprawnienia	18

1. Wstęp – opis problemu

Problem, który był rozwiązywany przez algorytm genetyczny polegał na zaprojektowaniu sieci fizycznych połączeń w taki sposób, aby dowolne dwa punkty lutownicze na płytce PCB zostały fizycznie połączone wtedy i tylko wtedy, jeżeli występowało między nimi planowane połączenie strukturalne. Rozwiązaniem problemu, jest zestaw połączeń między punktami lutowniczymi na danej płytce PCB. Jakość rozwiązania jest wyznaczana na podstawie funkcji oceny, na którą składa się suma iloczynów cech rozwiązania i ich wag. Im niższa wartość funkcji oceny, tym lepsze jest rozwiązanie. Na potrzeby realizowanego badania, zostały przyjęte następujące wagi:

Łączna długość wszystkich ścieżek (a): 15

Łączna liczba segmentów (b): 5

Liczba ścieżek, które wychodzą poza obszar płytki (c): 200

Suma długości segmentów, które znajdują się poza obszarem płytki (d): 100

Liczba kolizji między ścieżkami (e): 100

$$\text{Ocena} = 15*a + 5*b + 200*c + 100*d + 100*e$$

2. Metodyka

Badania zostały przeprowadzone w dużej części automatycznie, dzięki wcześniejszemu przygotowaniu odpowiedniego programu, który dla zadanego zakresu parametrów algorytmu, 10-krotnie generował rozwiązanie, a następnie wyliczał wszystkie potrzebne wartości, czyli:

średnią z najlepszych wyników, najlepszy z najlepszych wyników, najgorszy z najlepszych wyników oraz odchylenie standardowe przy średniej.

Po wyliczeniu, ww. wartości były zapisywane w pliku tekstowym (wraz z wartościami parametrów, dla których wykonywane było badanie), a następnie konwertowane do rozszerzenia .csv, aby ułatwić import danych do arkusza kalkulacyjnego. Następnie w arkuszu dane były manualnie przygotowywane do analizy i na ich podstawie powstały zestawienia i wykresy.

Dla wszystkich badań przyjąłem zestaw domyślnych wartości parametrów, których nie modyfikowałem, jeżeli nie sprawdzałem ich zależności w danym badaniu. Wybór ten został podjęty głównie przez pryzmat czasu, który byłby potrzebny do przeprowadzenia badań.

Wartości domyślne parametrów:

Prawdopodobieństwo krzyżowania (Px): 0.5

Prawdopodobieństwo krzyżowania (Pm): 0.5

Rozmiar populacji: 200

Liczba pokoleń: 20

Operator selekcji: Turniej

Rozmiar turnieju: 10% rozmiaru populacji

3. Wyniki badań

3.1 Badanie wpływu parametrów ilościowych

3.1.1 Badanie rozmiaru populacji

Cel badania: Wykrycie zależności między rozmiarem populacji, a jakością rozwiązania

Parametry:

Px: 0.5

Pm: 0.5

Rozmiar populacji: [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]

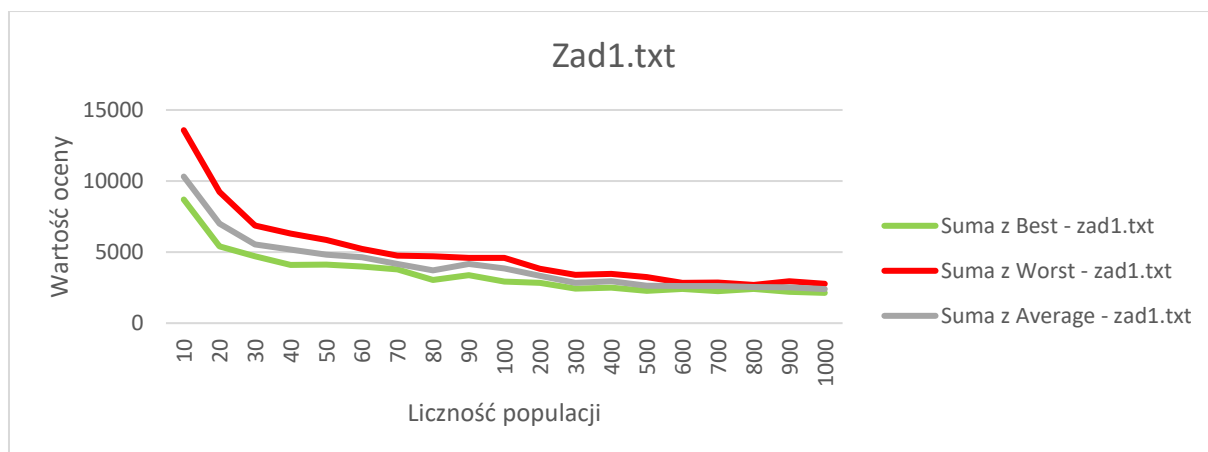
Liczba pokoleń: 20

Operator selekcji: Turniej

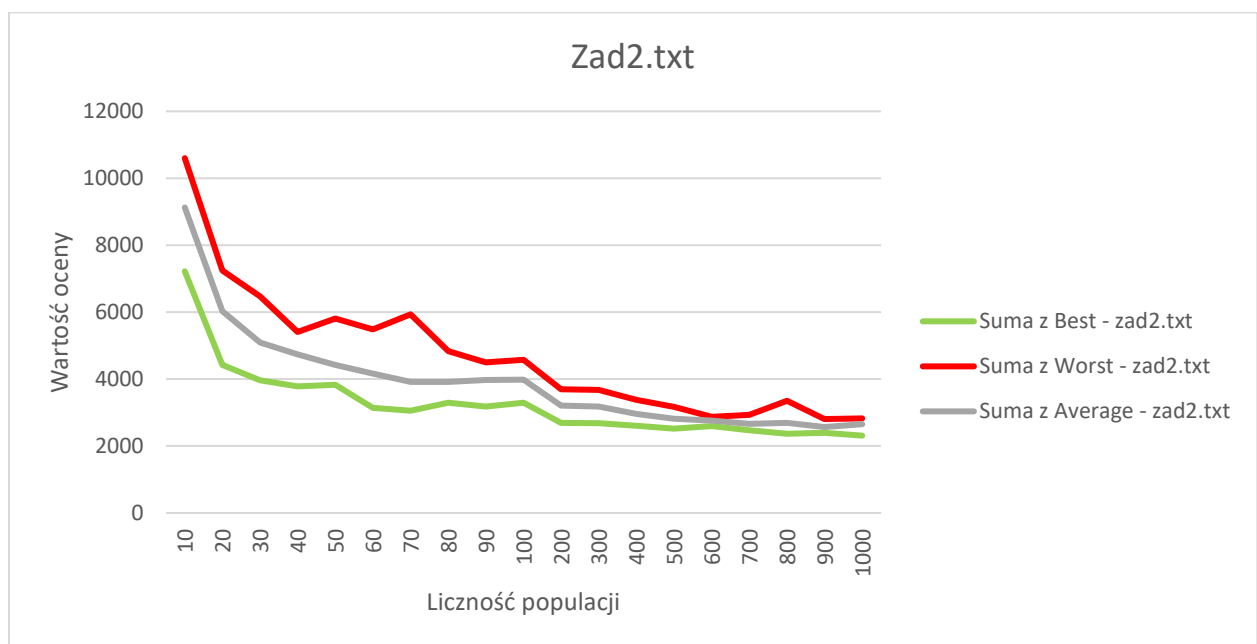
Rozmiar turnieju: 10% rozmiaru populacji

Wyniki:

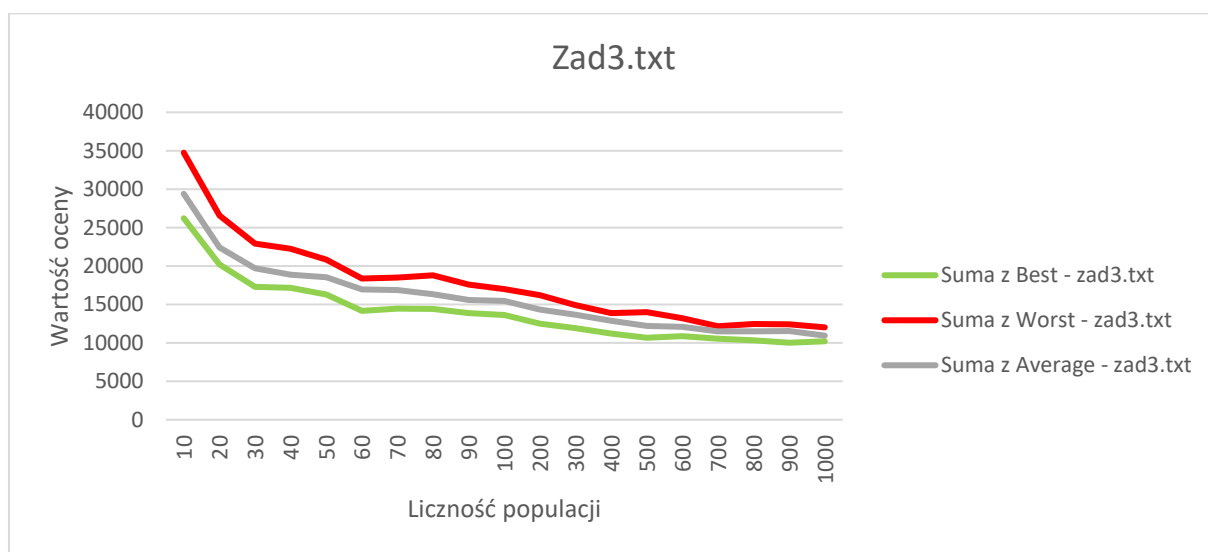
Populacja	Best	Worst	Average	Std
10	8700	13580	10314,50	1407,36
20	5395	9240	7005,50	1019,80
30	4695	6860	5529,00	729,16
40	4100	6305	5172,00	627,79
50	4115	5845	4813,50	483,67
60	3985	5215	4647,50	403,86
70	3785	4745	4161,00	299,95
80	3030	4710	3712,00	528,18
90	3385	4585	4160,50	408,36
100	2930	4600	3856,50	517,40
200	2845	3835	3321,50	273,10
300	2440	3400	2834,50	280,77
400	2505	3465	2938,00	296,43
500	2280	3245	2623,00	285,28
600	2415	2835	2617,50	137,14
700	2255	2855	2605,50	175,61
800	2400	2690	2548,00	96,47
900	2210	2940	2516,00	198,16
1000	2130	2770	2396,50	192,20



Populacja	Best	Worst	Average	Std
10	7220	10600	9128,0	1097,43
20	4425	7245	6032,0	888,55
30	3965	6470	5091,5	664,51
40	3785	5405	4736,5	521,02
50	3830	5810	4419,0	516,67
60	3145	5485	4165,0	608,01
70	3050	5935	3918,5	833,77
80	3290	4830	3916,0	411,95
90	3175	4495	3972,5	428,49
100	3290	4575	3984,5	446,89
200	2695	3695	3204,0	279,80
300	2680	3680	3179,5	321,70
400	2610	3380	2959,5	230,29
500	2515	3170	2811,0	195,25
600	2595	2870	2762,0	95,63
700	2470	2935	2666,5	141,58
800	2370	3350	2689,5	250,35
900	2395	2805	2567,5	132,10
1000	2310	2825	2657,0	142,64



Populacja	Best	Worst	Average	Std
10	26220	34745	29393,0	2494,97
20	20220	26575	22404,0	1899,72
30	17295	22905	19693,0	1586,20
40	17170	22245	18858,5	1490,87
50	16285	20850	18545,0	1433,86
60	14175	18370	16954,5	1252,36
70	14460	18495	16860,0	1098,10
80	14425	18780	16338,5	1324,49
90	13885	17585	15597,5	1157,26
100	13615	17010	15478,0	969,91
200	12485	16205	14346,5	1111,20
300	11925	14925	13670,0	1055,23
400	11215	13875	12894,0	829,92
500	10685	14015	12191,0	848,60
600	10870	13195	12085,5	678,09
700	10535	12150	11515,5	492,80
800	10325	12470	11491,0	632,66
900	10020	12420	11533,5	647,09
1000	10200	12025	10943,5	496,75



Wnioski:

Dla każdego z zadań, **wraz z rosnącą licznością populacji**, spadała wartość funkcji oceny (zarówno najlepsza, najgorsza, jak i średnia), więc **rozwiązania stawały się coraz lepsze**.

Wraz z polepszaniem się rozwiązania, malało także odchylenie standardowe, więc algorytm zwraca bardziej przewidywalne, oscylujące bliżej średniej wyniki dla większych populacji.

Wzrost jakości rozwiązania jest bardziej wyraźny przy mniejszych rozmiarach populacji, a od ok. 400-500 osobników, oscyluje w bliskim otoczeniu.

Im większa jest populacja startowa, tym większe prawdopodobieństwo wylosowania osobnika z przydatnymi dla populacji cechami, które mogą zostać wykorzystane do osiągnięcia najlepszego rozwiązania.

3.1.2 Badanie liczby pokoleń

Cel badania: Wykrycie zależności między liczbą pokoleń, a jakością rozwiązania

Parametry:

Px: 0.5

Pm: 0.5

Rozmiar populacji: 200

Liczba pokoleń: [10, 15, 20, 25, 30, 35, 40]

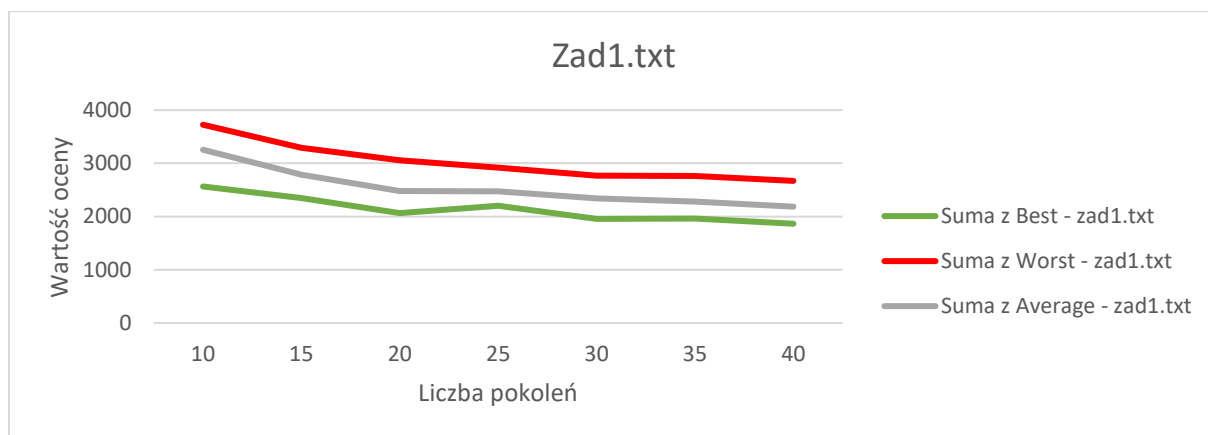
Operator selekcji: Turniej

Rozmiar turnieju: 10% rozmiaru populacji

Wyniki:

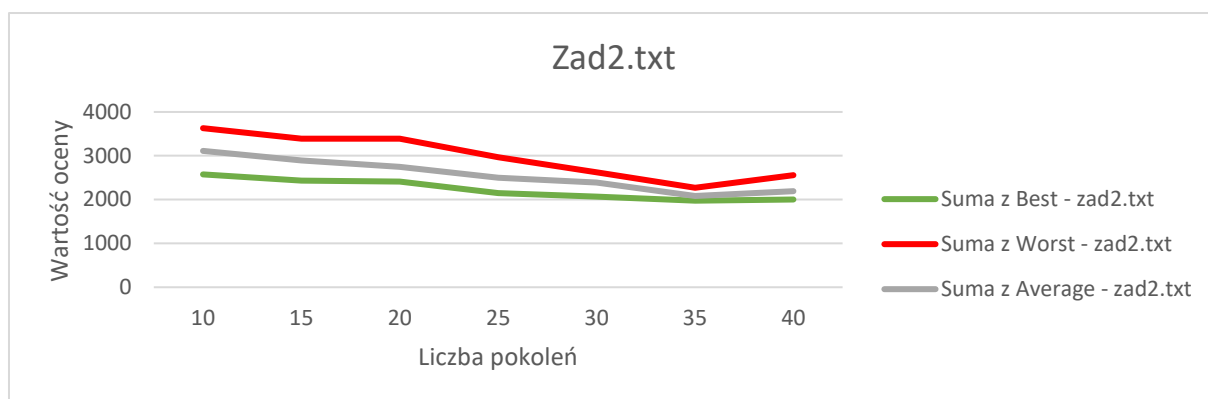
Zadanie 1.

Pokolenia	Best	Worst	Average	Std
10	2565	3725	3255,0	407,93
15	2350	3290	2783,5	307,81
20	2065	3060	2480,5	361,97
25	2200	2920	2474,0	227,12
30	1955	2770	2343,0	255,20
35	1965	2765	2281,5	231,74
40	1865	2670	2186,0	232,88



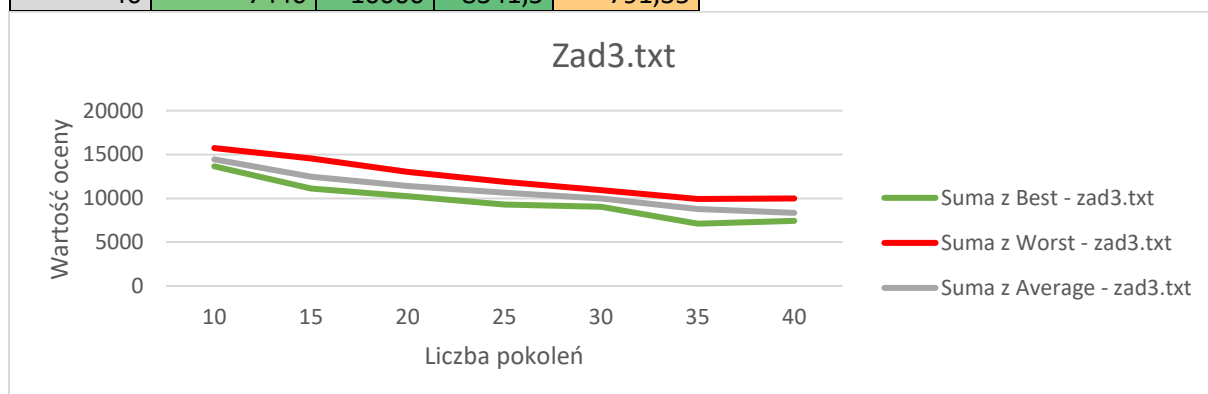
Zadanie 2.

Pokolenia	Best	Worst	Average	Std
10	2575	3630	3111,5	332,03
15	2430	3390	2896,5	269,94
20	2410	3390	2745,0	327,93
25	2145	2965	2501,0	251,84
30	2070	2620	2387,5	201,26
35	1975	2270	2082,5	106,07
40	2000	2555	2195,0	168,33



Zadanie 3.

Pokolenia	Best	Worst	Average	Std
10	13650	15735	14446,0	588,58
15	11100	14535	12460,0	1024,68
20	10240	13025	11413,0	716,26
25	9285	11890	10645,5	695,01
30	9035	10945	9971,5	599,71
35	7105	9925	8764,0	808,32
40	7440	10000	8341,5	791,35



Wnioski:

Podobnie jak w przypadku liczebności populacji, **wraz z rosnącą liczbą pokoleń, poprawia się jakość rozwiązania**, jednak w tym przypadku, dla zbadanych wartości, wykres przypomina bardziej zależność liniową. Dla większej liczby pokoleń wydłużamy proces poszukiwania najlepszego rozwiązania, stąd zwiększa się szansa na jego znalezienie.

3.1.3 Badanie rozmiaru turnieju

Cel badania: Wykrycie zależności między rozmiarem turnieju, a jakością rozwiązania

Parametry:

Px: 0.5

Pm: 0.5

Rozmiar populacji: 200

Liczba pokoleń: 20

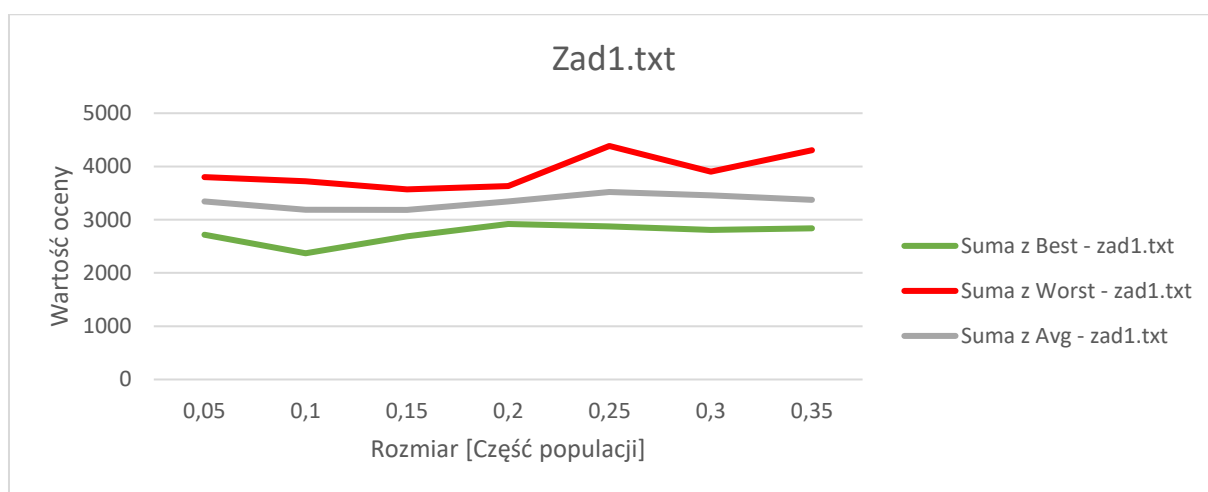
Operator selekcji: Turniej

Rozmiar turnieju: [0.1, 0.15, 0.20, 0.25, 0.30, 0.35, 0.55]

Wyniki:

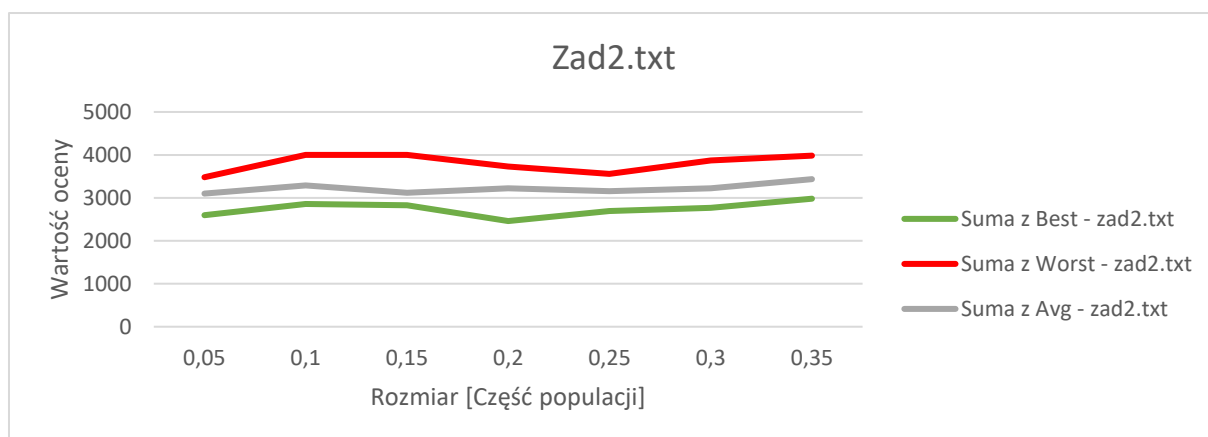
Zadanie 1.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	2720	3800	3343,0	328,53
Turniej	0,1	2370	3725	3186,0	364,39
Turniej	0,15	2690	3570	3184,5	285,64
Turniej	0,2	2920	3635	3341,0	243,16
Turniej	0,25	2875	4385	3521,0	481,07
Turniej	0,3	2810	3900	3457,0	318,16
Turniej	0,35	2840	4305	3376,5	436,50
Turniej	0,55	3150	4465	3835,5	423,16



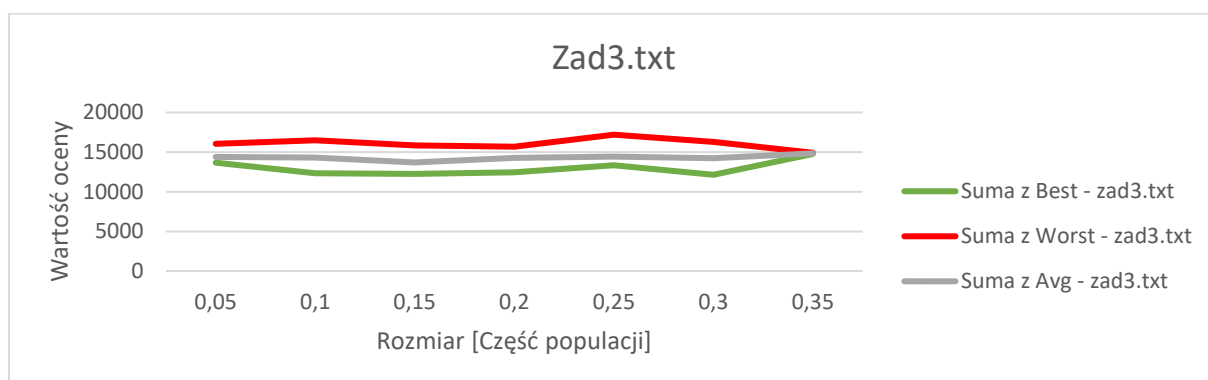
Zadanie 2.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	2595	3480	3099,5	239,34
Turniej	0,1	2860	4000	3292,0	353,69
Turniej	0,15	2830	4000	3115,0	333,83
Turniej	0,2	2460	3730	3221,5	330,37
Turniej	0,25	2695	3560	3155,0	289,31
Turniej	0,3	2770	3870	3220,0	280,86
Turniej	0,35	2980	3985	3435,5	347,94
	0,55	2935	4145	4338,5	418,75



Zadanie 3.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	13655	16060	14393,5	671,33
Turniej	0,1	12335	16495	14325,0	1389,58
Turniej	0,15	12255	15835	13695,0	1079,77
Turniej	0,2	12440	15670	14261,0	1005,31
Turniej	0,25	13330	17200	14418,0	1174,23
Turniej	0,3	12140	16305	14248,0	1175,87
Turniej	0,35	14765	14920	14842,5	77,50
Turniej	0,55	13820	18925	15854,5	1387,83



Wnioski:

Nie widać wyraźnej zależności między rozmiarem turnieju, a jakością rozwiązania. Dla każdej z płytek najlepsze rozwiązanie pojawiło się przy innej wartości tego współczynnika. Jednakże **najlepsze wartości średniego rozwiązania** we wszystkich przypadkach koncentrują się w okolicach rozmiaru turnieju **ok. 10-20%**, więc z moich obserwacji wynika, iż to właśnie te wartości są najlepsze dla sprawnego działania algorytmu. Najśłabsze rozwiązania wychodzą dla **rozmiarów 35% i 50%** i pogarszają się one wraz ze wzrostem tego parametru od pewnego momentu, ponieważ rośnie wtedy ciśnienie selekcyjne i wiodący osobnik szybko dominuje populację, czego następstwem jest zatrzymanie ewolucji.

3.2 Porównanie operatorów selekcji

Cel badania: Porównanie jakości rozwiązań otrzymanych przy użyciu Turnieju oraz Ruletki jako operatora selekcji

Parametry:

Px: 0.5

Pm: 0.5

Rozmiar populacji: 200

Liczba pokoleń: 20

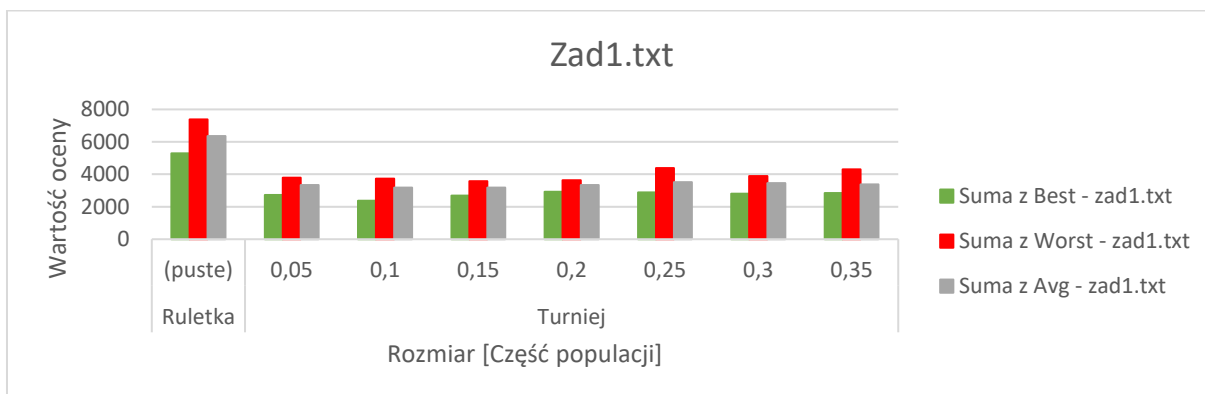
Operator selekcji: [Turniej, Ruletka]

Rozmiar turnieju: [0.1, 0.15, 0.20, 0.25, 0.30, 0.35]

Wyniki:

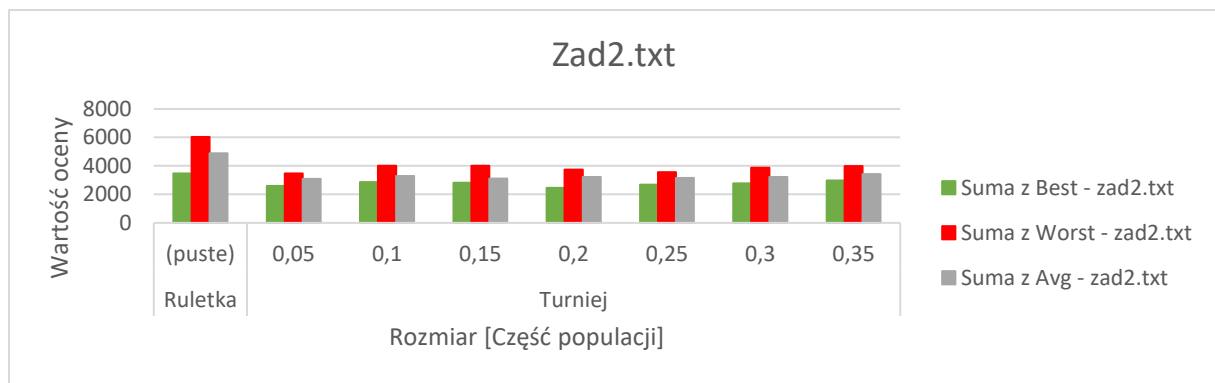
Zadanie 1.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	2720	3800	3343,0	328,53
Turniej	0,1	2370	3725	3186,0	364,39
Turniej	0,15	2690	3570	3184,5	285,64
Turniej	0,2	2920	3635	3341,0	243,16
Turniej	0,25	2875	4385	3521,0	481,07
Turniej	0,3	2810	3900	3457,0	318,16
Turniej	0,35	2840	4305	3376,5	436,50
Ruletka		5295	7375	6364,0	598,01



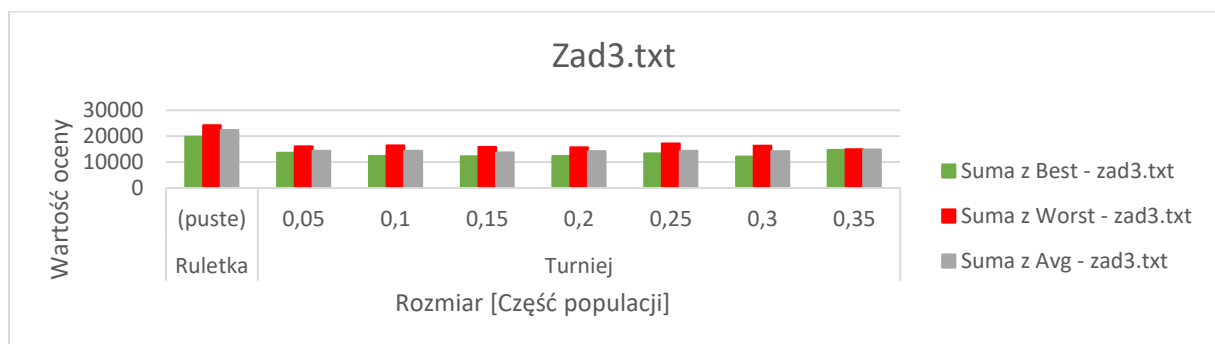
Zadanie 2.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	2595	3480	3099,5	239,34
Turniej	0,1	2860	4000	3292,0	353,69
Turniej	0,15	2830	4000	3115,0	333,83
Turniej	0,2	2460	3730	3221,5	330,37
Turniej	0,25	2695	3560	3155,0	289,31
Turniej	0,3	2770	3870	3220,0	280,86
Turniej	0,35	2980	3985	3435,5	347,94
Ruletka		3475	6040	4888,5	704,05



Zadanie 3.

Operator selekcji	Rozmiar turnieju	Best	Worst	Avg	Std
Turniej	0,05	13655	16060	14393,5	671,33
Turniej	0,1	12335	16495	14325,0	1389,58
Turniej	0,15	12255	15835	13695,0	1079,77
Turniej	0,2	12440	15670	14261,0	1005,31
Turniej	0,25	13330	17200	14418,0	1174,23
Turniej	0,3	12140	16305	14248,0	1175,87
Turniej	0,35	14765	14920	14842,5	77,50
Ruletka		19855	24280	22372,0	1146,29



Badanie zostało powtórzone dla **30 pokoleń** i nadal nie było widać żadnej poprawy Ruletki względem Turnieju:

Operator selekcji	Zadanie	Best	Worst	Avg	Std
Turniej	Zad1.txt	2065	3060	2480,5	361,97
Ruletka	Zad1.txt	4170	5405	4763,0	360,92
Turniej	Zad2.txt	2410	3390	2745,0	327,93
Ruletka	Zad2.txt	3880	5045	4332,5	319,60
Turniej	Zad3.txt	10240	13025	11413,0	716,26
Ruletka	Zad3.txt	18375	22355	20250,0	1251,32

Wnioski: Turniej w każdym przypadku, tj. dla każdego zadania i z każdym badanym rozmiarem dał lepszy rezultat niż ruletka. **Turniej jest dla badanego algorytmu dużo lepszą formą selekcji rodzica niż ruletka.** (Przynajmniej dla przyjętych współczynników domyślnych).

W przypadku ruletki występuje większe ryzyko zbyt niskiego ciśnienia selekcyjnego, ponieważ przypadku niewiele różniących się osobników – nawet jeżeli są daleko od siebie w „rankingu” – jest

bardzo niewielka różnica szans na wylosowanie, czego następstwem jest utrudniony proces ewolucji – przestaje on być zbieżny.

3.3 Zbadanie wpływu parametrów prawdopodobieństwa

3.3.1 Badanie wpływu prawdopodobieństwa krzyżowania

Cel badania: Wykrycie zależności między prawdopodobieństwem krzyżowania, a jakością rozwiązania.

Parametry:

Px: [0.3, 0.4, 0.5, 0.6, 0.7]

Pm: 0.5

Rozmiar populacji: 200

Liczba pokoleń: 10

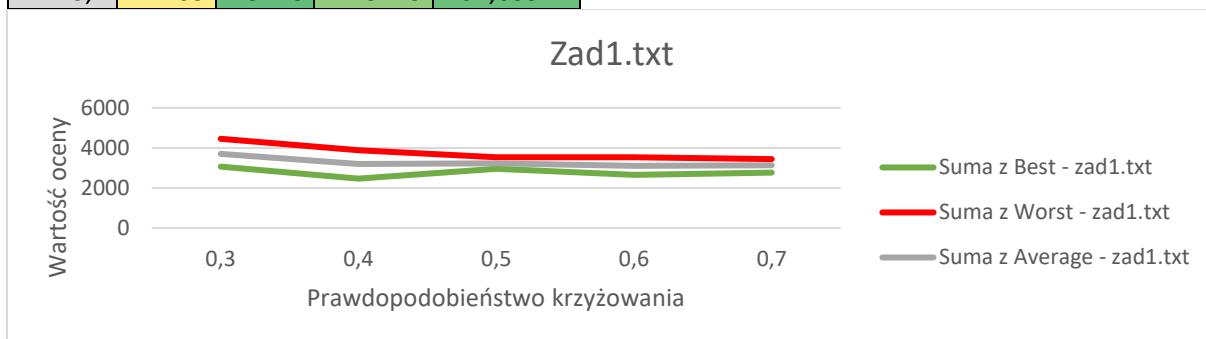
Operator selekcji: Turniej

Rozmiar turnieju: 10% rozmiaru populacji

Wyniki:

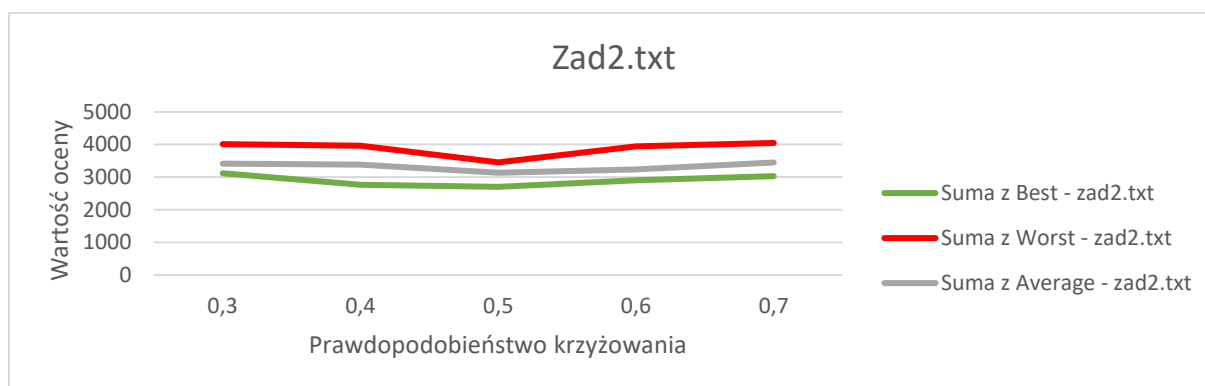
Zadanie 1.

Px	Best	Worst	Average	Std
0,3	3065	4455	3702	419,08352
0,4	2465	3885	3189,5	397,70875
0,5	2950	3540	3232	203,48464
0,6	2655	3540	3103,5	292,20755
0,7	2765	3440	3129	207,63911



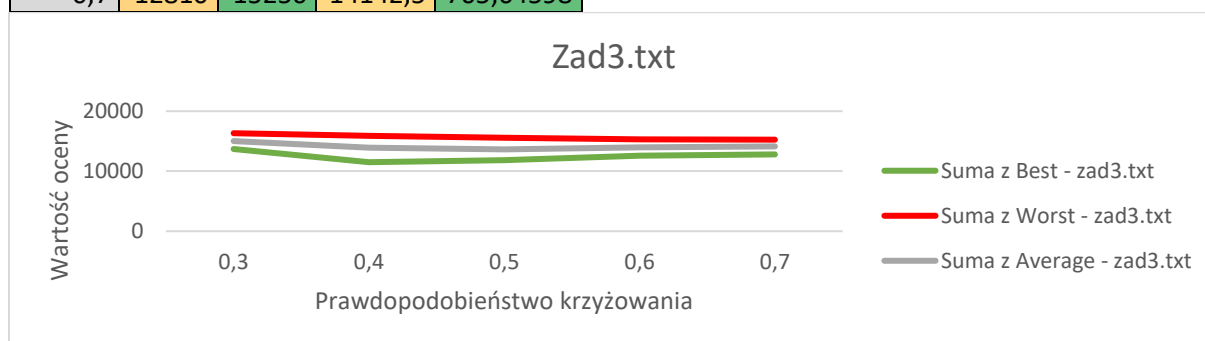
Zadanie 2.

Px	Best	Worst	Average	Std
0,3	3120	4010	3413,5	281,45204
0,4	2770	3965	3389	376,2034
0,5	2705	3455	3137	201,04975
0,6	2905	3945	3242,5	313,11539
0,7	3030	4050	3450	273,72431



Zadanie 3.

Px	Best	Worst	Average	Std
0,3	13695	16325	15017	701,91595
0,4	11500	15900	13938	1274,818
0,5	11825	15565	13633	1115,6572
0,6	12610	15325	13993	895,96652
0,7	12810	15250	14142,5	705,64598



Wnioski:

Średnia jakość rozwiązania jest najlepsza dla prawdopodobieństwa krzyżowania na poziomie ok. 50%. Najgorsze rozwiązania zostały zwrócone przy prawdopodobieństwie 30%, ponieważ algorytm działa wtedy dużo wolniej z uwagi na to, że rzadko dokonywana jest wymiana genów między osobnikami z poprzednich populacji.

3.3.2 Badanie wpływu prawdopodobieństwa mutacji

Cel badania: Wykrycie zależności między prawdopodobieństwem krzyżowania, a jakością rozwiązania.

Parametry:

Px: 0.5

Pm: [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]

Rozmiar populacji: 200

Liczba pokoleń: 20

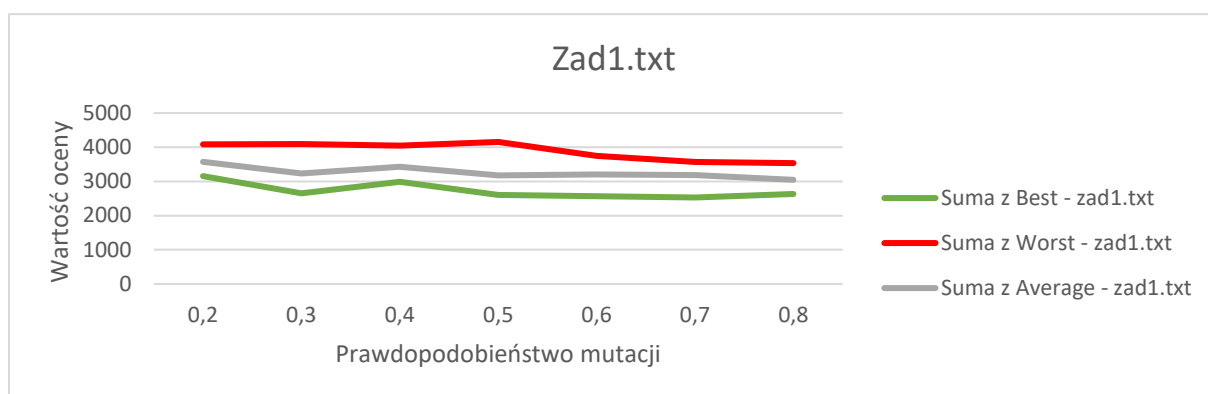
Operator selekcji: Turniej

Rozmiar turnieju: 10% rozmiaru populacji

Wyniki:

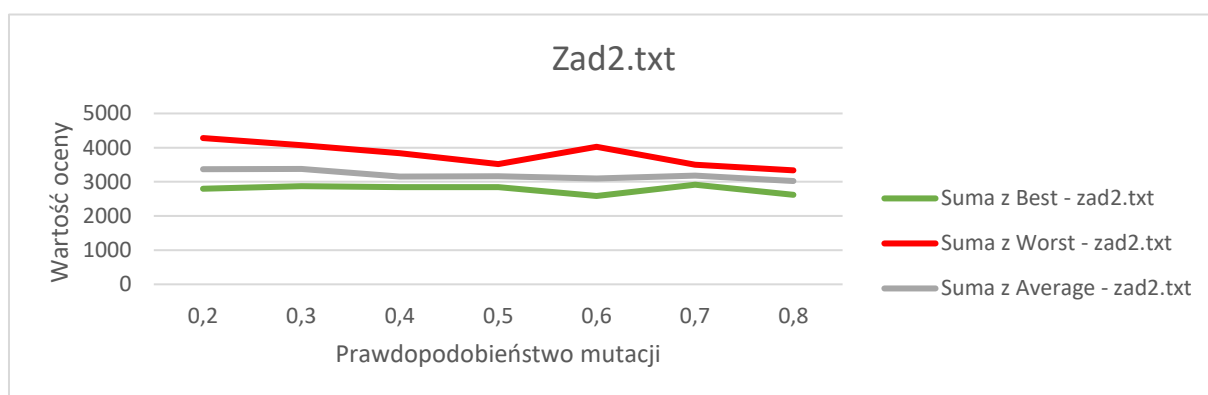
Zadanie 1.

Pm	Best	Worst	Average	Std
0,2	3155	4085	3572,5	282,65
0,3	2655	4095	3230,0	361,37
0,4	2990	4050	3433,5	338,03
0,5	2605	4155	3179,5	427,06
0,6	2565	3745	3205,0	313,69
0,7	2530	3570	3182,5	280,52
0,8	2635	3535	3046,5	244,81



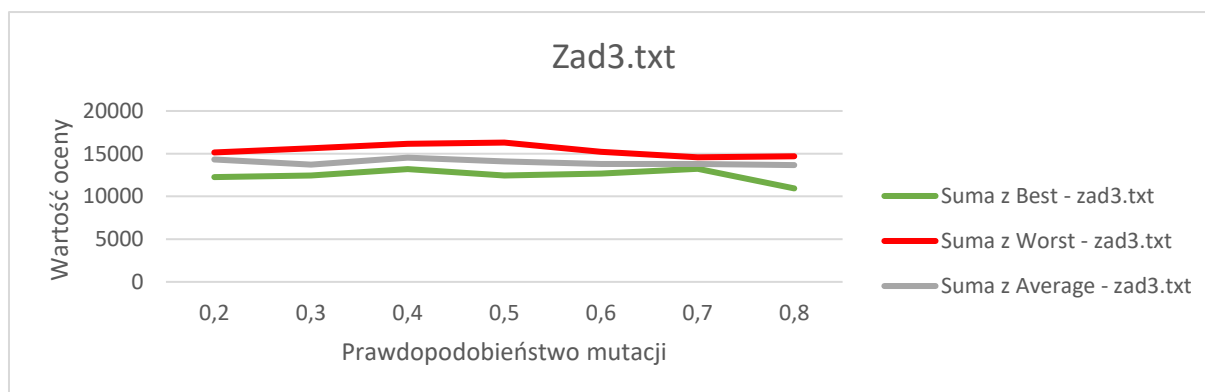
Zadanie 2.

Pm	Best	Worst	Average	Std
0,2	2800	4280	3368,0	418,56
0,3	2870	4070	3376,5	317,20
0,4	2845	3840	3150,5	318,43
0,5	2840	3520	3162,0	255,12
0,6	2585	4020	3100,5	412,36
0,7	2915	3495	3183,5	214,35
0,8	2615	3335	3023,0	214,55



Zadanie 3.

Pm	Best	Worst	Average	Std
0,2	12260	15130	14334,5	795,61
0,3	12465	15635	13738,5	850,59
0,4	13190	16150	14538,5	929,91
0,5	12455	16305	14105,0	1044,03
0,6	12660	15230	13783,0	745,25
0,7	13235	14580	13809,0	421,06
0,8	10940	14700	13669,5	1095,75



Wnioski:

Nie wykryłem liniowej zależności między prawdopodobieństwem mutacji, a jakością otrzymanego rozwiązania. Pod kątem poszukiwania najlepszego rozwiązania najlepsze okazało się prawdopodobieństwo na poziomie 80%. Algorytm przy użyciu takiej wartości parametru, dla każdego problemu zwrócił rozwiązania o najniższej średniej funkcji oceny.

3.4 Porównanie algorytmu genetycznego z metodą losową

Cel badania: Porównanie różnicy w jakości rozwiązań algorytmu genetycznego i metody losowej

Parametry:

Px: 0.5

Pm: 0.5

Rozmiar populacji: 200

Liczba pokoleń: 20

Operator selekcji: Turniej

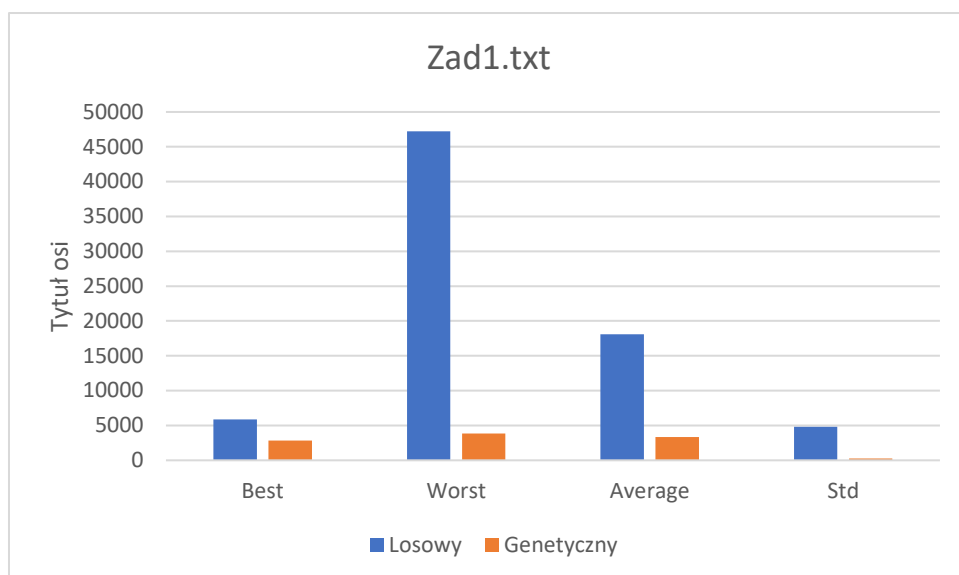
Rozmiar turnieju: 10% rozmiaru populacji

Liczba wygenerowanych osobników w metodzie losowej: $200 \cdot 20 \cdot 10 = 40\,000$

Wyniki:

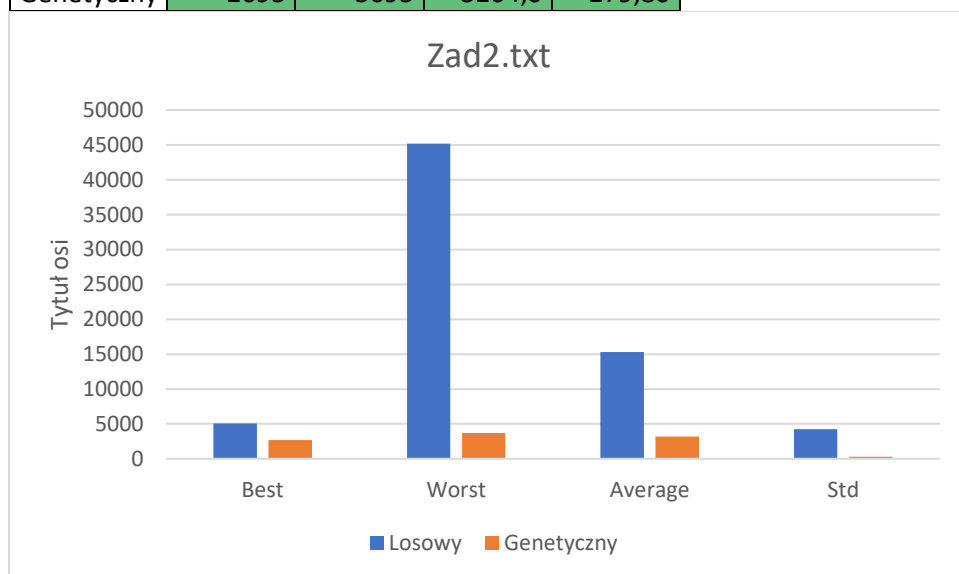
Zadanie 1.

Algorytm	Best	Worst	Average	Std
Losowy	5890	47209	18070,8	4815,46
Genetyczny	2845	3835	3321,5	273,10



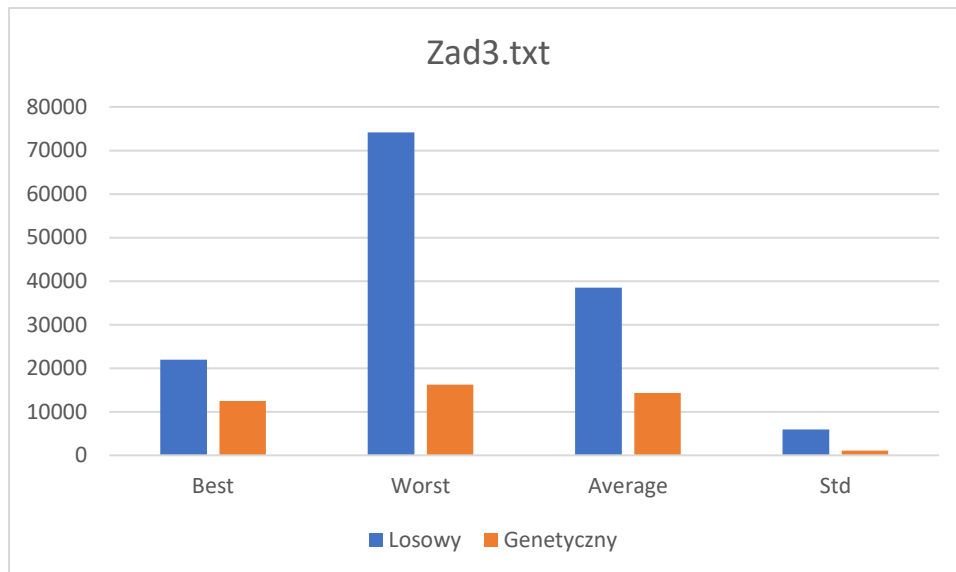
Zadanie 2.

Algorytm	Best	Worst	Average	Std
Losowy	5085	45165	15295,6	4256,75
Genetyczny	2695	3695	3204,0	279,80



Zadanie 3.

Algorytm	Best	Worst	Average	Std
Losowy	21940	74170	38519,9	5903,20
Genetyczny	12485	16205	14346,5	1111,20



Wnioski:

Algorytm genetyczny jest nieporównywalnie lepszy od metody losowej. Dla każdego zadania znajdował niemal dwukrotnie lepsze rozwiązania, a wynik jego działania jest niemożliwy do przewidzenia, ponieważ wartość odchylenia standardowego jest na bardzo wysokim poziomie

4. Podsumowanie

Przeprowadzone badania prowadzą do kilku wyraźnych wniosków:

- Algorytm Genetyczny jest dużo lepszy od algorytmu losowego
- Im większy rozmiar populacji – tym lepsze rozwiązania
- Im większa liczba pokoleń – tym lepsze rozwiązania
- Selekcja Turniejowa sprawdza się lepiej od Koła Ruletki
- Selekcja Turniejowa działa najlepiej dla rozmiaru turnieju w okolicach 10-20%
- Wartość prawdopodobieństwa mutacji na poziomie 80% daje najlepsze rozwiązania

5. Możliwe usprawnienia

Aby wyniki były bardziej rzetelne oraz możliwe było wykrycie wszystkich obecnych zależności, należałoby przeprowadzać bardziej różnorodne badania, tzn. każdy z badanych czynników zbadać pod kątem wpływu innego czynnika na badany rezultat. W obecnym przypadku było to trudne, ponieważ wymagałoby lepszego sprzętu komputerowego, który pozwoliłby na przeprowadzenie tak obszernych badań w krótszym czasie.