

<b>Systemy inteligentne 2</b>		
<b>Temat: Analiza danych z wykorzystaniem drzew decyzyjnych</b>		
<b>Numer komputera: 3</b>		<b>Numer projektu: 4</b>
<b>Data: 11.12.2017 r.</b>	<b>Autor: Bartłomiej Osak</b>	<b>Grupa: 3ID13B</b>

## 1. Cel projektu.

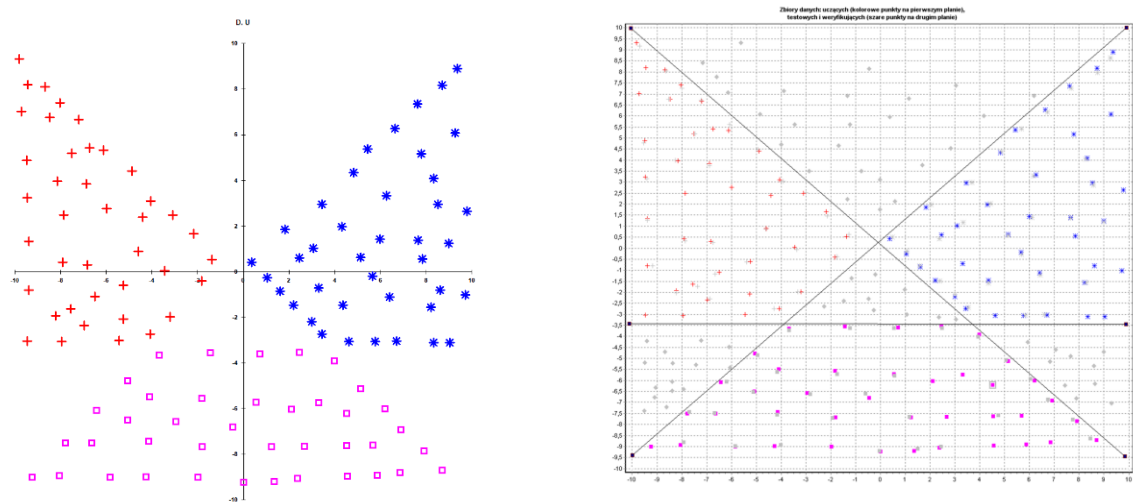
Celem projektu było na podstawie przygotowanych podczas poprzednich projektów danych uczących i testowych skonstruować odpowiednie drzewa decyzyjne oraz ich warianty w różnych stopniach przycięcia drzewa maksymalnego.

2. Dane liniowo separowalne.

2.1 Dane uczące:

Lp.	we	we	wy	wy	wy
	x1	x2	klasa1	klasa2	klasa3
1.	-9,822	9,322	1	0	0
2.	-8,678	8,094	1	0	0
3.	-7,2	6,669	1	0	0
4.	-6,118	5,343	1	0	0
5.	-4,89	4,409	1	0	0
6.	-4,058	3,107	1	0	0
7.	-3,101	2,493	1	0	0
8.	-2,185	1,658	1	0	0
9.	-1,373	0,528	1	0	0
10.	-1,831	-0,405	1	0	0
11.	-3,205	-1,977	1	0	0
12.	-4,079	-2,739	1	0	0
13.	-5,452	-3,009	1	0	0
14.	-7,949	-3,058	1	0	0
15.	-9,469	-3,034	1	0	0
16.	-9,385	-0,798	1	0	0
17.	-9,385	1,339	1	0	0
18.	-9,469	3,23	1	0	0
19.	-9,489	4,876	1	0	0
20.	-9,718	7,013	1	0	0
21.	-7,512	5,195	1	0	0
22.	-5,993	2,763	1	0	0
23.	-4,599	0,897	1	0	0
24.	-5,244	-0,602	1	0	0
25.	-6,493	-1,093	1	0	0
26.	-7,908	0,43	1	0	0
27.	-7,866	2,493	1	0	0
28.	-8,157	3,967	1	0	0
29.	-6,888	3,844	1	0	0
30.	-6,826	0,307	1	0	0
31.	-3,454	0,037	1	0	0
32.	-5,265	-2,076	1	0	0
33.	-6,971	-2,346	1	0	0
34.	-7,554	-1,634	1	0	0
35.	-8,22	-1,928	1	0	0
36.	-8,47	6,767	1	0	0
37.	-9,448	8,192	1	0	0
38.	-8,033	7,406	1	0	0
39.	-6,742	5,416	1	0	0
40.	-4,412	2,395	1	0	0
41.	0,375	0,43	0	1	0
42.	1,832	1,855	0	1	0
43.	3,455	2,96	0	1	0
44.	4,828	4,336	0	1	0
45.	5,432	5,367	0	1	0
46.	6,639	6,276	0	1	0
47.	7,617	7,357	0	1	0
48.	8,72	8,168	0	1	0
49.	9,365	8,905	0	1	0
50.	1,041	-0,258	0	1	0
51.	1,603	-0,847	0	1	0
52.	2,206	-1,462	0	1	0
53.	2,997	-2,199	0	1	0
54.	3,434	-2,739	0	1	0
55.	4,62	-3,058	0	1	0
56.	5,765	-3,058	0	1	0
57.	6,701	-3,034	0	1	0
58.	8,345	-3,107	0	1	0
59.	9,032	-3,107	0	1	0
60.	2,456	0,602	0	1	0
61.	3,309	-0,7	0	1	0
62.	4,35	-1,462	0	1	0
63.	4,308	1,977	0	1	0
64.	5,14	0,626	0	1	0
65.	5,661	-0,184	0	1	0
66.	6,41	-1,118	0	1	0
67.	6,264	3,328	0	1	0
68.	7,658	1,388	0	1	0
69.	7,846	0,553	0	1	0
70.	8,616	-0,798	0	1	0
71.	7,783	5,171	0	1	0
72.	8,324	4,09	0	1	0
73.	8,532	2,96	0	1	0
74.	8,99	1,241	0	1	0
75.	9,282	6,08	0	1	0
76.	9,781	2,641	0	1	0
77.	9,719	-1,019	0	1	0
78.	8,22	-1,56	0	1	0
79.	5,994	1,437	0	1	0
80.	3,08	1,019	0	1	0
81.	-9,24	-9,003	0	0	1
82.	-7,783	-7,504	0	0	1
83.	-6,43	-6,08	0	0	1
84.	-5,057	-4,778	0	0	1
85.	-3,683	-3,648	0	0	1
86.	-1,436	-3,55	0	0	1
87.	0,708	-3,599	0	0	1
88.	2,456	-3,525	0	0	1
89.	3,975	-3,918	0	0	1
90.	5,14	-5,122	0	0	1
91.	6,202	-6,006	0	0	1
92.	6,909	-6,915	0	0	1
93.	7,908	-7,848	0	0	1
94.	8,699	-8,708	0	0	1
95.	-4,099	-5,49	0	0	1
96.	-1,831	-5,564	0	0	1
97.	0,541	-5,711	0	0	1
98.	3,309	-5,736	0	0	1
99.	-6,638	-7,504	0	0	1
100.	-4,141	-7,431	0	0	1
101.	-1,81	-7,676	0	0	1
102.	1,228	-7,676	0	0	1
103.	2,664	-7,652	0	0	1
104.	4,537	-7,627	0	0	1
105.	5,681	-7,603	0	0	1
106.	-8,053	-8,929	0	0	1
107.	-5,848	-9,003	0	0	1
108.	-4,266	-8,978	0	0	1
109.	-1,977	-9,003	0	0	1
110.	0	-9,224	0	0	1
111.	1,353	-9,199	0	0	1
112.	2,373	-9,052	0	0	1
113.	4,558	-8,954	0	0	1
114.	5,869	-8,905	0	0	1
115.	6,847	-8,806	0	0	1
116.	-0,458	-6,792	0	0	1
117.	-2,955	-6,571	0	0	1
118.	-5,078	-6,497	0	0	1
119.	2,102	-6,031	0	0	1
120.	4,516	-6,203	0	0	1

Graficzna postać danych uczących z programu Neuronix 4.0 oraz z edytora danych 2D:



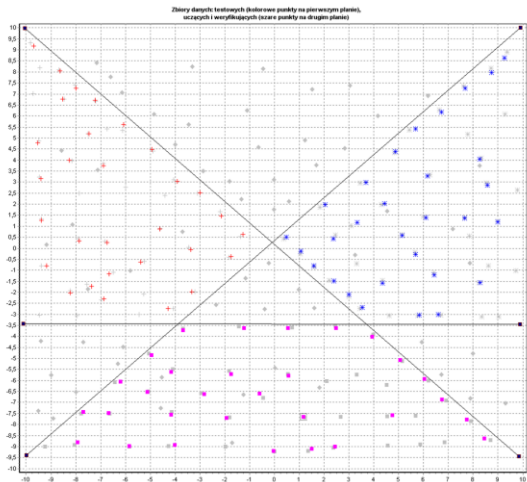
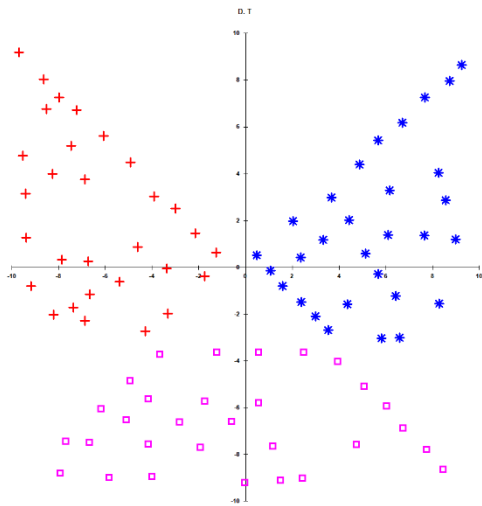
2.2 Dane testowe:

Lp.	we	we	wy	wy	wy
	x1	x2	klasa1	klasa2	klasa3
1.	-9,697	9,175	1	0	0
2.	-8,636	8,045	1	0	0
3.	-7,991	7,259	1	0	0
4.	-7,221	6,718	1	0	0
5.	-6,076	5,613	1	0	0
6.	-4,932	4,483	1	0	0
7.	-3,912	3,034	1	0	0
8.	-2,996	2,518	1	0	0
9.	-2,143	1,462	1	0	0
10.	-1,269	0,626	1	0	0
11.	-1,748	-0,381	1	0	0
12.	-3,371	-0,061	1	0	0
13.	-4,62	0,872	1	0	0
14.	-6,888	3,746	1	0	0
15.	-8,262	3,992	1	0	0
16.	-9,406	3,157	1	0	0
17.	-9,385	1,265	1	0	0
18.	-6,742	0,258	1	0	0
19.	-7,866	0,332	1	0	0
20.	-8,22	-2,027	1	0	0
21.	-7,367	-1,732	1	0	0
22.	-6,659	-1,167	1	0	0
23.	-5,39	-0,626	1	0	0
24.	-4,287	-2,739	1	0	0
25.	-3,329	-1,977	1	0	0
26.	-6,867	-2,297	1	0	0
27.	-9,177	-0,798	1	0	0
28.	-7,471	5,195	1	0	0
29.	-8,532	6,767	1	0	0
30.	-9,531	4,778	1	0	0
31.	6,722	6,178	0	1	0
32.	7,679	7,259	0	1	0
33.	8,741	7,971	0	1	0
34.	9,261	8,634	0	1	0
35.	5,681	5,416	0	1	0
36.	4,87	4,385	0	1	0
37.	3,684	2,985	0	1	0
38.	2,04	1,977	0	1	0
39.	0,479	0,504	0	1	0

40.	1,062	-0,135	0	1	0
41.	1,582	-0,798	0	1	0
42.	2,393	-1,486	0	1	0
43.	2,997	-2,1	0	1	0
44.	3,538	-2,69	0	1	0
45.	4,35	-1,584	0	1	0
46.	5,681	-0,282	0	1	0
47.	5,14	0,577	0	1	0
48.	6,098	1,388	0	1	0
49.	7,658	1,363	0	1	0
50.	8,99	1,191	0	1	0
51.	8,574	2,862	0	1	0
52.	8,262	4,041	0	1	0
53.	6,16	3,279	0	1	0
54.	4,433	2,027	0	1	0
55.	3,33	1,167	0	1	0
56.	2,373	0,43	0	1	0
57.	5,827	-3,034	0	1	0
58.	6,597	-3,009	0	1	0
59.	6,431	-1,216	0	1	0
60.	8,283	-1,56	0	1	0
61.	-4,953	-4,851	0	0	1
62.	-3,683	-3,722	0	0	1
63.	-1,228	-3,623	0	0	1
64.	0,541	-3,623	0	0	1
65.	2,477	-3,623	0	0	1
66.	3,933	-4,016	0	0	1
67.	5,057	-5,073	0	0	1
68.	6,035	-5,932	0	0	1
69.	6,743	-6,866	0	0	1
70.	7,742	-7,775	0	0	1
71.	8,449	-8,634	0	0	1
72.	0,562	-5,785	0	0	1
73.	-1,748	-5,711	0	0	1
74.	-0,603	-6,596	0	0	1
75.	-2,83	-6,62	0	0	1
76.	-4,162	-5,613	0	0	1
77.	-5,119	-6,522	0	0	1
78.	-4,162	-7,554	0	0	1
79.	-6,68	-7,48	0	0	1
80.	-7,7	-7,431	0	0	1

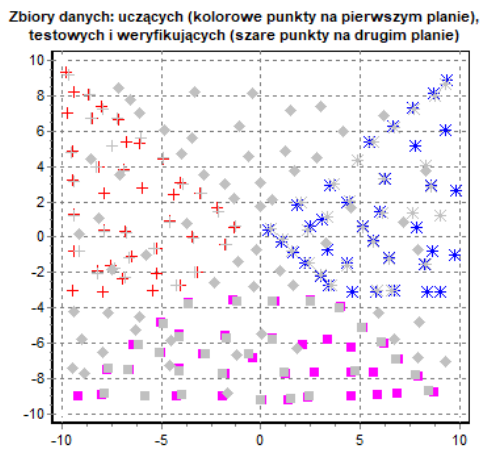
81.	-0,041	-9,199	0	0	1
82.	1,499	-9,101	0	0	1
83.	2,435	-9,003	0	0	1
84.	4,745	-7,578	0	0	1
85.	-7,929	-8,806	0	0	1
86.	-5,848	-8,978	0	0	1
87.	-4,016	-8,929	0	0	1
88.	-1,935	-7,701	0	0	1
89.	-6,201	-6,055	0	0	1
90.	1,166	-7,652	0	0	1

Graficzna postać danych testowych z programu Neuronix 4.0 oraz z edytora danych 2D:

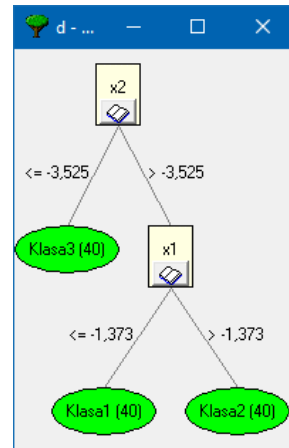


## 2.3 Ilustracja graficzna uzyskanych reguł decyzyjnych:

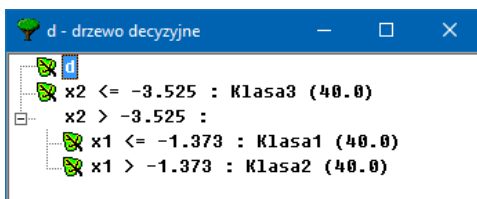
a) Zbiór danych:



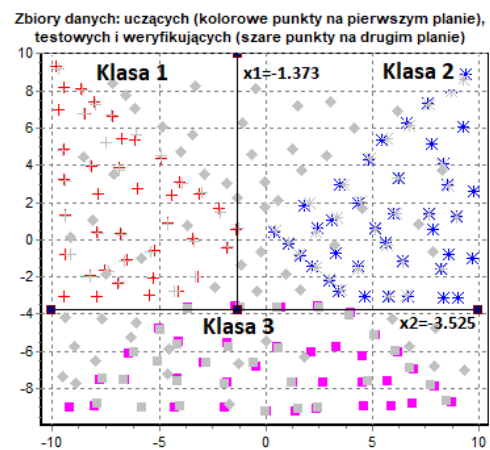
b) Maksymalne drzewo decyzyjne:



c) Zbiór reguł dla drzewa decyzyjnego z rys. b):



d) Ilustracja graficzna uzyskanych reguł decyzyjnych:



Parametr 1: Min. Liczba przykładów tworzących liść drzewa	Parametr 2: Przycinanie drzewa decyzyjnego	PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
1	• Nie przycinaj drzewa	0	1.11	3	100	98.89

Legenda:

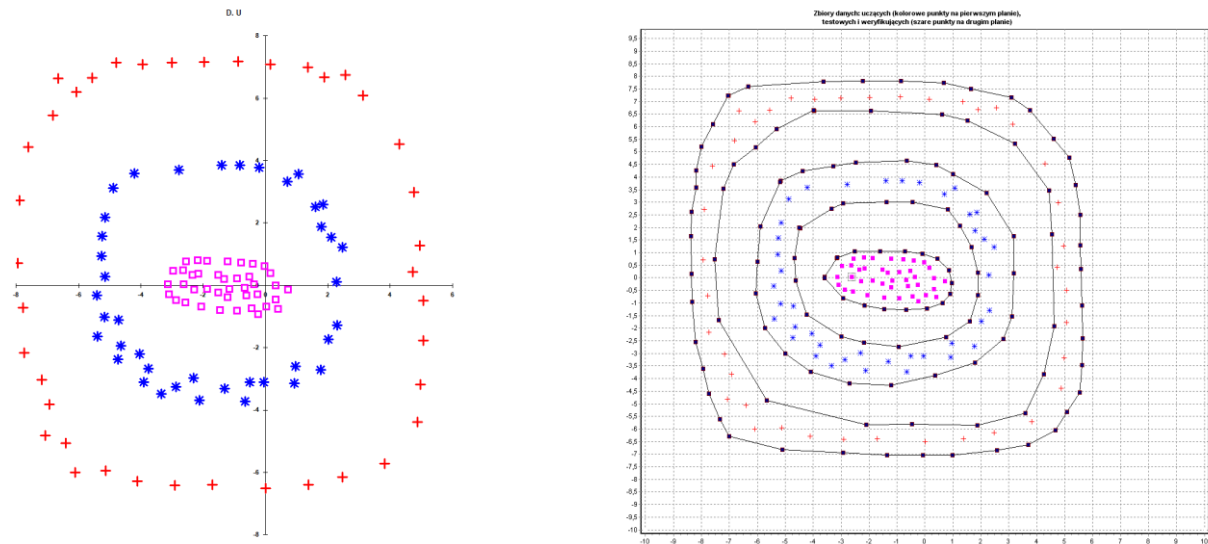
- RD – Rozmiar Drzewa (liczba liści (reguł)),
- PPBS-DT[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Testowe),
- PPBS-DU[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Uczące),
- PPDDT[%] – Procent Poprawnych Decyzji dla Danych Testowych,
- PPDDU[%] – Procent Poprawnych Decyzji dla Danych Uczących.

3. Dane liniowo nieseparowalne.

3.1 Dane uczące.

Lp.	we	we	wy	wy	wy
	x1	x2	klasa1	klasa2	klasa3
1.	-6,656	6,629	1	0	0
2.	-3,944	7,087	1	0	0
3.	-3,012	7,136	1	0	0
4.	-0,88	7,184	1	0	0
5.	0,155	7,087	1	0	0
6.	2,557	6,749	1	0	0
7.	3,137	6,097	1	0	0
8.	4,296	4,528	1	0	0
9.	4,772	2,982	1	0	0
10.	4,958	1,267	1	0	0
11.	5,062	-0,495	1	0	0
12.	5,062	-1,775	1	0	0
13.	4,979	-3,176	1	0	0
14.	4,876	-4,384	1	0	0
15.	3,82	-5,712	1	0	0
16.	2,474	-6,146	1	0	0
17.	1,377	-6,388	1	0	0
18.	0,01	-6,509	1	0	0
19.	-1,708	-6,388	1	0	0
20.	-2,909	-6,412	1	0	0
21.	-4,11	-6,291	1	0	0
22.	-5,124	-5,953	1	0	0
23.	-6,097	-6,002	1	0	0
24.	-6,408	-5,06	1	0	0
25.	-7,07	-4,818	1	0	0
26.	-6,925	-3,828	1	0	0
27.	-7,174	-3,031	1	0	0
28.	-7,733	-2,162	1	0	0
29.	-7,774	-0,713	1	0	0
30.	-7,94	0,712	1	0	0
31.	-7,898	2,716	1	0	0
32.	-7,608	4,431	1	0	0
33.	-6,822	5,445	1	0	0
34.	-6,076	6,194	1	0	0
35.	-5,559	6,653	1	0	0
36.	-4,793	7,136	1	0	0
37.	-1,977	7,16	1	0	0
38.	1,356	6,991	1	0	0
39.	1,894	6,677	1	0	0
40.	4,731	0,422	1	0	0
41.	-4,876	3,127	0	1	0
42.	-4,213	3,586	0	1	0
43.	-1,397	3,851	0	1	0
44.	-0,197	3,779	0	1	0
45.	0,694	3,32	0	1	0
46.	1,604	2,523	0	1	0
47.	2,122	1,533	0	1	0
48.	2,288	0,108	0	1	0
49.	2,308	-1,292	0	1	0
50.	1,77	-2,717	0	1	0
51.	-0,052	-3,104	0	1	0
52.	-1,315	-3,321	0	1	0
53.	-2,867	-3,249	0	1	0
54.	-3,758	-2,669	0	1	0
55.	-4,648	-1,944	0	1	0
56.	-5,165	-1,027	0	1	0
57.	-5,269	0,929	0	1	0
58.	-5,145	2,185	0	1	0
59.	-2,785	3,707	0	1	0
60.	-0,818	3,851	0	1	0
61.	1,066	3,562	0	1	0
62.	1,853	2,596	0	1	0
63.	2,474	1,219	0	1	0
64.	2,019	-1,727	0	1	0
65.	0,963	-2,596	0	1	0
66.	-0,507	-3,104	0	1	0
67.	-2,308	-2,983	0	1	0
68.	-4,047	-2,21	0	1	0
69.	-4,731	-2,379	0	1	0
70.	-5,393	-1,631	0	1	0
71.	-5,414	-0,326	0	1	0
72.	-5,248	1,581	0	1	0
73.	-5,145	0,277	0	1	0
74.	-4,71	-1,123	0	1	0
75.	-3,344	-3,49	0	1	0
76.	-3,903	-3,104	0	1	0
77.	-0,652	-3,732	0	1	0
78.	-2,122	-3,683	0	1	0
79.	0,921	-3,152	0	1	0
80.	1,791	1,871	0	1	0
81.	-2,557	0,76	0	0	1
82.	-2,184	0,809	0	0	1
83.	-1,874	0,784	0	0	1
84.	-1,211	0,76	0	0	1
85.	-0,776	0,736	0	0	1
86.	-0,404	0,688	0	0	1
87.	-0,031	0,615	0	0	1
88.	0,155	0,398	0	0	1
89.	0,321	-0,012	0	0	1
90.	0,321	-0,447	0	0	1
91.	0,031	-0,689	0	0	1
92.	-0,445	-0,737	0	0	1
93.	-0,921	-0,809	0	0	1
94.	-1,46	-0,785	0	0	1
95.	-2,06	-0,689	0	0	1
96.	-2,578	-0,544	0	0	1
97.	-2,867	-0,471	0	0	1
98.	-3,095	-0,278	0	0	1
99.	-3,137	0,036	0	0	1
100.	-2,971	0,471	0	0	1
101.	-2,64	0,495	0	0	1
102.	-2,163	0,374	0	0	1
103.	-1,335	0,205	0	0	1
104.	-0,921	0,229	0	0	1
105.	-0,528	0,277	0	0	1
106.	-0,362	0,036	0	0	1
107.	-0,673	-0,085	0	0	1
108.	-1,17	-0,109	0	0	1
109.	-1,874	-0,109	0	0	1
110.	-2,288	-0,133	0	0	1
111.	-1,211	-0,375	0	0	1
112.	-0,611	-0,326	0	0	1
113.	-0,259	-0,278	0	0	1
114.	-1,522	0,326	0	0	1
115.	-2,35	0,326	0	0	1
116.	-1,522	-0,23	0	0	1
117.	-0,238	-0,93	0	0	1
118.	0,404	-0,761	0	0	1
119.	0,714	-0,133	0	0	1
120.	-2,619	0,036	0	0	1

Graficzna postać danych uczących z programu Neuronix 4.0 oraz z edytora danych 2D:



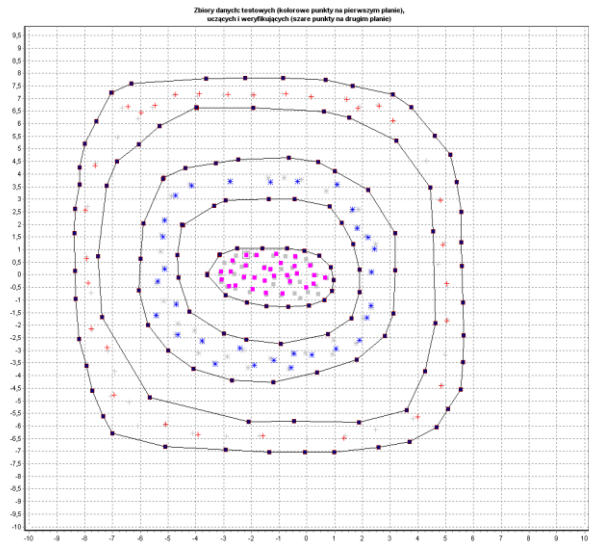
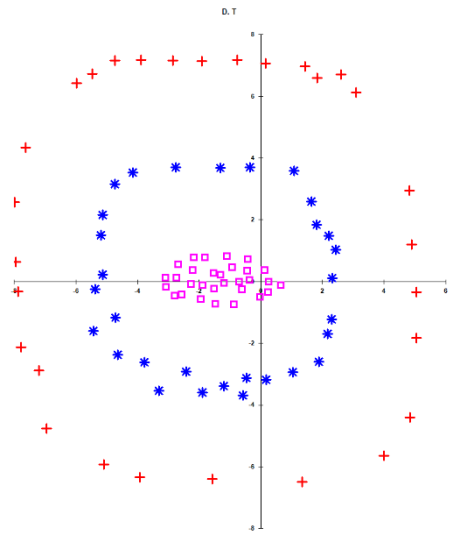
3.2. Dane testowe.

Lp.	we	we	wy	wy	wy
	x1	x2	klasa1	klasa2	klasa3
1.	-6,449	6,677	1	0	0
2.	-5,973	6,435	1	0	0
3.	-5,476	6,725	1	0	0
4.	-4,731	7,16	1	0	0
5.	-3,882	7,184	1	0	0
6.	-2,847	7,16	1	0	0
7.	-1,915	7,136	1	0	0
8.	-0,756	7,184	1	0	0
9.	0,155	7,063	1	0	0
10.	1,439	6,967	1	0	0
11.	1,832	6,604	1	0	0
12.	2,598	6,701	1	0	0
13.	3,095	6,122	1	0	0
14.	4,813	2,958	1	0	0
15.	4,896	1,195	1	0	0
16.	5,041	-0,351	1	0	0
17.	5,041	-1,824	1	0	0
18.	3,985	-5,639	1	0	0
19.	-1,584	-6,388	1	0	0
20.	-3,923	-6,34	1	0	0
21.	-6,946	-4,77	1	0	0
22.	-7,194	-2,886	1	0	0
23.	-7,774	-2,138	1	0	0
24.	-7,878	-0,326	1	0	0
25.	-7,94	0,64	1	0	0
26.	-7,981	2,572	1	0	0
27.	-7,629	4,334	1	0	0
28.	-5,083	-5,929	1	0	0
29.	1,335	-6,485	1	0	0
30.	4,834	-4,408	1	0	0
31.	-4,731	3,151	0	1	0
32.	-4,151	3,537	0	1	0
33.	-1,315	3,682	0	1	0
34.	-0,342	3,707	0	1	0
35.	1,087	3,586	0	1	0
36.	1,646	2,596	0	1	0
37.	2,205	1,485	0	1	0
38.	2,433	1,026	0	1	0
39.	2,329	0,108	0	1	0

40.	2,308	-1,22	0	1	0
41.	2,163	-1,703	0	1	0
42.	1,894	-2,596	0	1	0
43.	1,046	-2,935	0	1	0
44.	0,176	-3,176	0	1	0
45.	-0,466	-3,128	0	1	0
46.	-0,569	-3,683	0	1	0
47.	-1,19	-3,393	0	1	0
48.	-1,894	-3,587	0	1	0
49.	-3,778	-2,621	0	1	0
50.	-4,648	-2,379	0	1	0
51.	-4,71	-1,172	0	1	0
52.	-5,372	-0,254	0	1	0
53.	-5,124	0,229	0	1	0
54.	-5,186	1,509	0	1	0
55.	-2,764	3,707	0	1	0
56.	-3,302	-3,538	0	1	0
57.	-2,412	-2,91	0	1	0
58.	1,812	1,847	0	1	0
59.	-5,435	-1,606	0	1	0
60.	-5,124	2,161	0	1	0
61.	-1,812	0,784	0	0	1
62.	-1,108	0,833	0	0	1
63.	-0,942	0,471	0	0	1
64.	-1,315	0,229	0	0	1
65.	-1,19	-0,037	0	0	1
66.	-1,48	-0,713	0	0	1
67.	-1,956	-0,568	0	0	1
68.	-1,894	-0,109	0	0	1
69.	-2,743	0,132	0	0	1
70.	-3,095	0,132	0	0	1
71.	-3,074	-0,182	0	0	1
72.	-2,805	-0,447	0	0	1
73.	-0,031	-0,495	0	0	1
74.	0,238	-0,351	0	0	1
75.	0,259	-0,012	0	0	1
76.	0,114	0,374	0	0	1
77.	-0,362	0,06	0	0	1
78.	0,652	-0,109	0	0	1
79.	-0,88	-0,737	0	0	1
80.	-0,611	-0,254	0	0	1

81.	-0,714	-0,012	0	0	1
82.	-0,445	0,35	0	0	1
83.	-0,424	0,736	0	0	1
84.	-2,681	0,567	0	0	1
85.	-2,205	0,374	0	0	1
86.	-2,578	-0,423	0	0	1
87.	-2,267	-0,085	0	0	1
88.	-1,522	-0,23	0	0	1
89.	-1,542	0,277	0	0	1
90.	-2,184	0,784	0	0	1

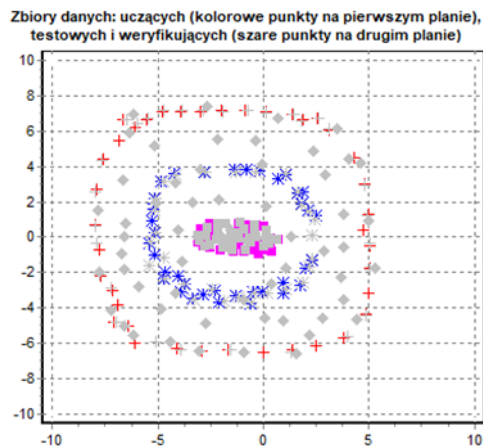
Graficzna postać danych testowych z programu Neuronix 4.0 oraz z edytora danych 2D:



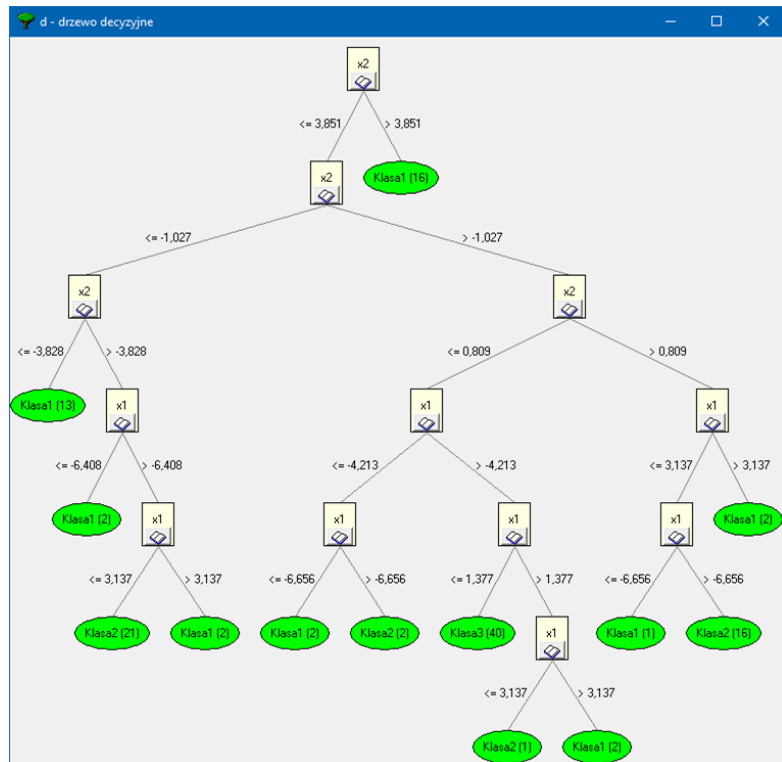
### 3.3 Ilustracja graficzna uzyskanych reguł decyzyjnych:

#### 3.3.1 Drzewo decyzyjne bez przycinania (minimalna liczba przykładów tworzących liść drzewa równa 1):

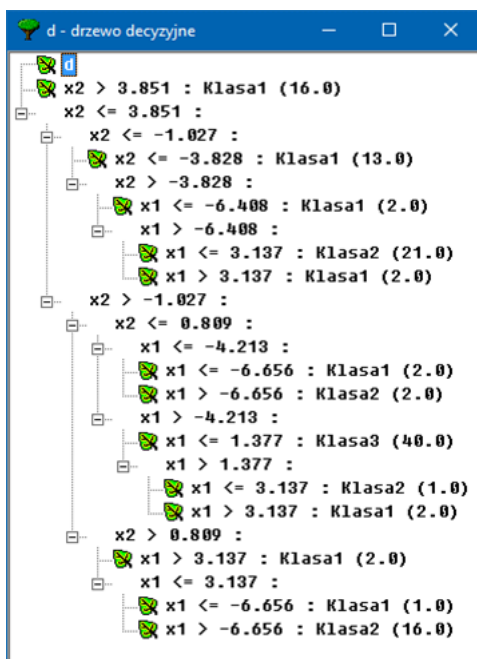
a) Zbiór danych:



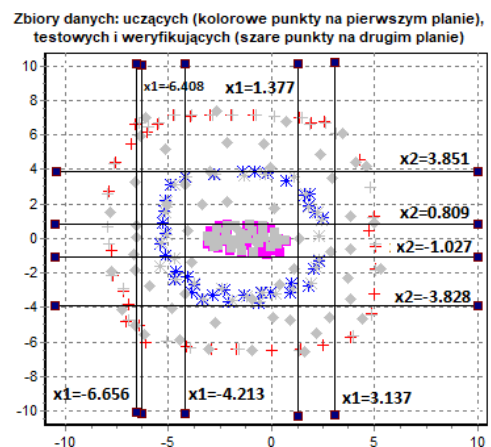
b) Uzyskane drzewo decyzyjne:



c) Zbiór reguł dla drzewa decyzyjnego z rys. b):



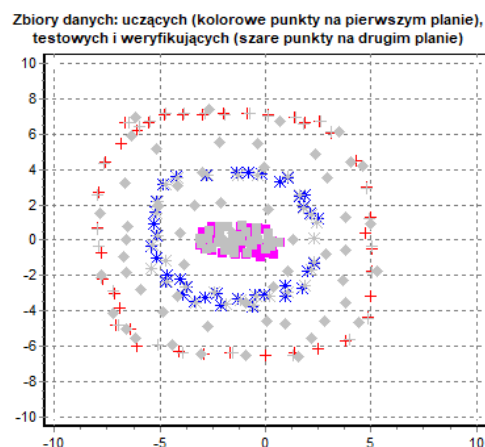
d) Ilustracja graficzna uzyskanych reguł decyzyjnych:



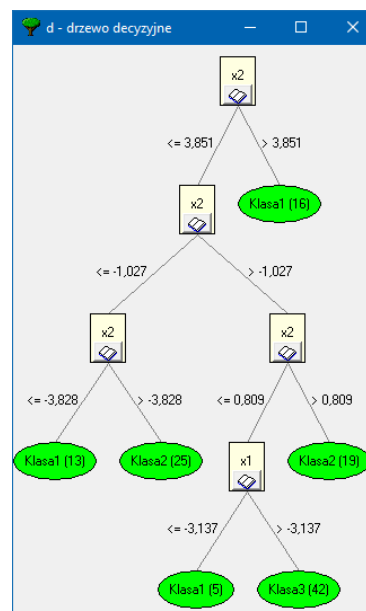


### 3.3.2 Drzewo decyzyjne przycięte w połowie (minimalna liczba przykładów tworzących liść drzewa równa 5):

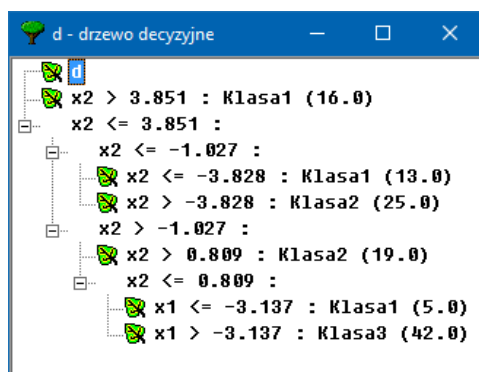
a) Zbiór danych:



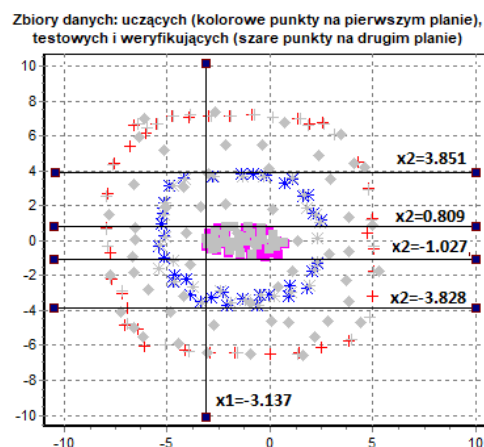
b) Uzyskane drzewo decyzyjne:



c) Zbiór reguł dla drzewa decyzyjnego z rys. b):



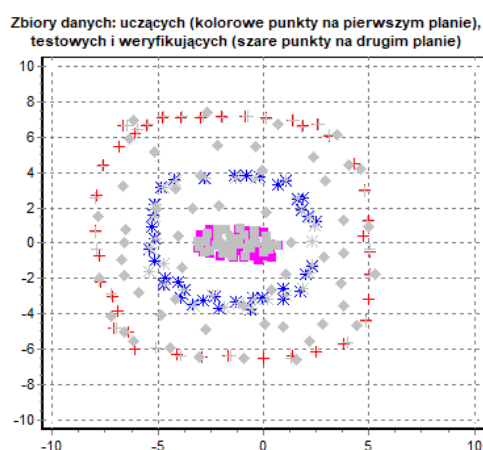
d) Ilustracja graficzna uzyskanych reguł decyzyjnych:



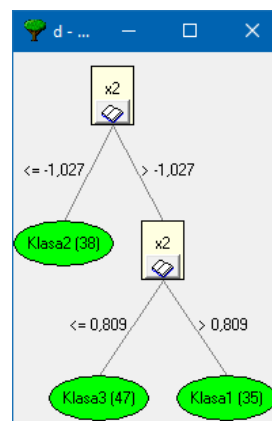


### 3.3.3 Drzewo decyzyjne przycięte maksymalnie (minimalna liczba przykładów tworzących liść drzewa równa 20):

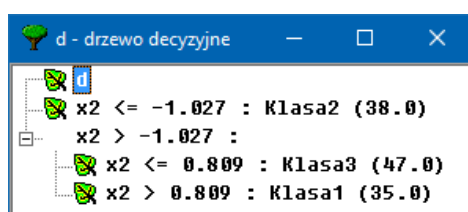
a) Zbiór danych:



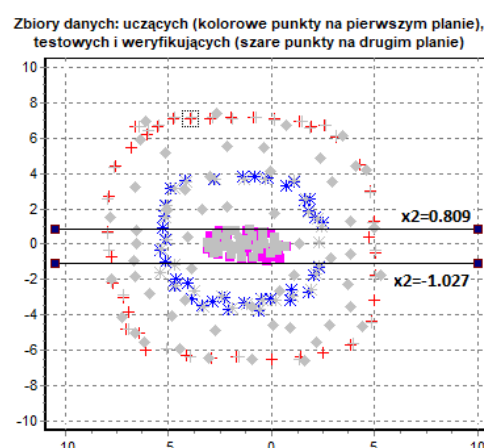
b) Uzyskane drzewo decyzyjne:



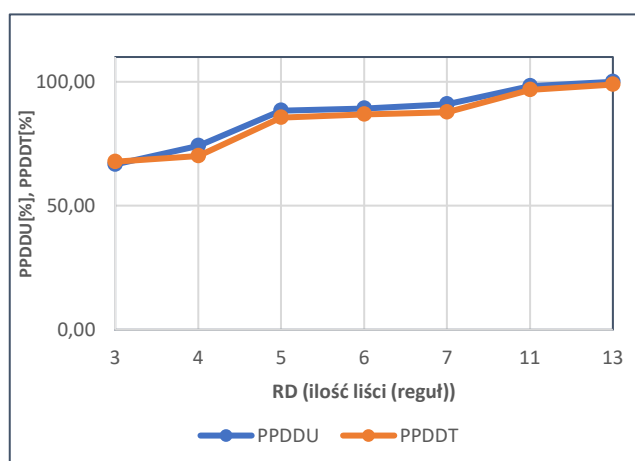
c) Zbiór reguł dla drzewa decyzyjnego z rys. b):



d) Ilustracja graficzna uzyskanych reguł decyzyjnych:



Parametr 1: Min. Liczba przykładów tworzących liść drzewa	Parametr 2: Przycinanie drzewa decyzyjnego	PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
1	• Nie przycinaj drzewa	0	1.11	13	100	98.89
2	• Nie przycinaj drzewa	1.67	3.33	11	98.33	96.67
4	• Nie przycinaj drzewa	9.17	12.22	7	90.83	87.78
5	• Nie przycinaj drzewa	10.83	13.11	6	89.27	86.89
8	• Nie przycinaj drzewa	11.67	14.44	5	88.33	85.56
18	• Nie przycinaj drzewa	25.83	30	4	74.17	70
20	• Nie przycinaj drzewa	33.33	32.22	3	66.67	67.78



#### Legenda:

- RD – Rozmiar Drzewa (liczba liści (reguł)),
- PPBS-DT[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Testowe),
- PPBS-DU[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Uczące),
- PPDDT[%] – Procent Poprawnych Decyzji dla Danych Testowych,
- PPDDU[%] – Procent Poprawnych Decyzji dla Danych Uczących.

#### 4. Dane w postaci graficznej.

##### 4.1 Dane uczące 8x10 pikseli (analogicznie dla obrazów 16x20 pikseli).

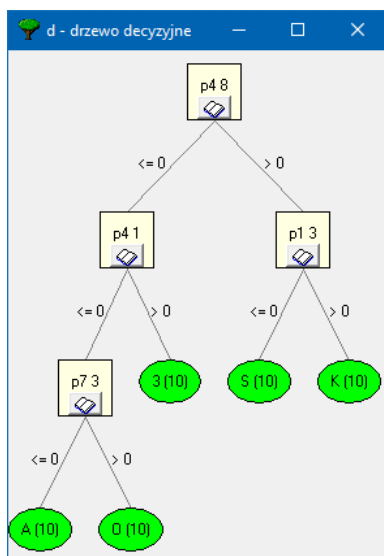


##### 4.2 Dane testowe 8x10 pikseli (analogicznie dla obrazów 16x20 pikseli).

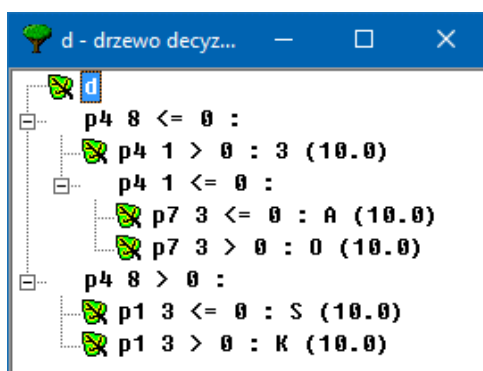


### 4.3 Ilustracja graficzna uzyskanych reguł decyzyjnych – matryca 8x10 pikseli:

a) Maksymalne drzewo decyzyjne:



b) Zbiór reguł dla drzewa decyzyjnego z rys. a):



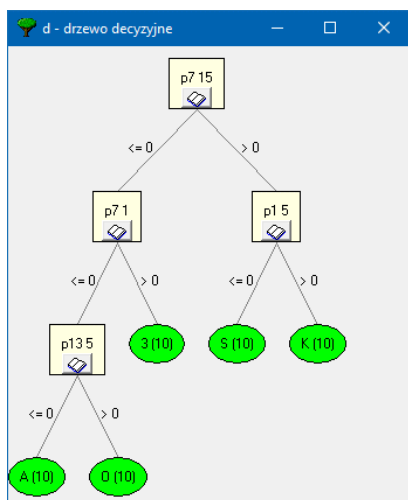
Parametr 1: Min. Liczba przykładów tworzących liść drzewa	Parametr 2: Przycinanie drzewa decyzyjnego	PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
1	• Nie przycinaj drzewa	0	52	5	100	48

#### Legenda:

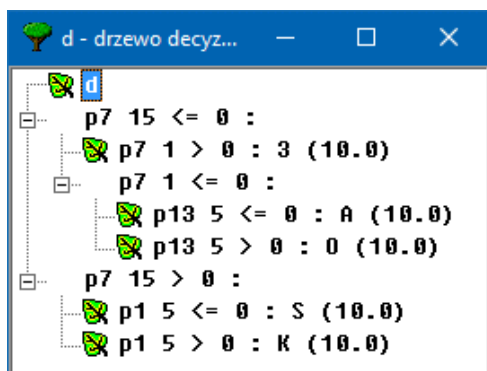
- RD – Rozmiar Drzewa (liczba liści (reguł)),
- PPBS-DT[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Testowe),
- PPBS-DU[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Uczące),
- PPDDT[%] – Procent Poprawnych Decyzji dla Danych Testowych,
- PPDDU[%] – Procent Poprawnych Decyzji dla Danych Uczących.

#### 4.4 Ilustracja graficzna uzyskanych reguł decyzyjnych – matryca 16x15 pikseli:

c) Maksymalne drzewo decyzyjne:



d) Zbiór reguł dla drzewa decyzyjnego z rys. a):



Parametr 1: Min. Liczba przykładów tworzących liść drzewa	Parametr 2: Przycinanie drzewa decyzyjnego	PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
1	• Nie przycinaj drzewa	0	52	5	100	48

Legenda:

- RD – Rozmiar Drzewa (liczba liści (reguł)),
- PPBS-DT[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Testowe),
- PPBS-DU[%] – Procent Przykładów Błędnie Sklasyfikowanych (Dane Uczące),
- PPDDT[%] – Procent Poprawnych Decyzji dla Danych Testowych,
- PPDDU[%] – Procent Poprawnych Decyzji dla Danych Uczących.

## 5. Wnioski.

Zgodnie z celem projektu, którym było skonstruowanie odpowiednich drzew decyzyjnych na podstawie danych liniowo separowalnych, liniowo nieseparowalnych oraz danych graficznych przygotowano odpowiednie drzewa decyzyjne oraz ich różne warianty w zależności od stopnia przycięcia drzewa maksymalnego. Poniżej zostanie przeprowadzona analiza uzyskanych wyników.

Pierwszym typem danych, dla których skonstruowano drzewo decyzyjne były dane uczące liniowo separowalne w postaci 120 rekordów. Uzyskano drzewo maksymalne o minimalnej liczbie przykładów tworzących liść drzewa równej 1 oraz o liczbie liści – reguł równej 3. Testowanie drzewa przeprowadzono dla danych testowych – 90 rekordów oraz danych uczących – 120 rekordów, które były podane systemowi w postaci danych testowych. Uzyskane wyniki były satysfakcjonujące – procent przykładów błędnie sklasyfikowanych dla danych testowych wynosił jedynie 1.11%, a dla danych uczących wyniósł on 0%. Na podstawie drzewa rozpisanego w postaci zbioru reguł wyznaczono nowe linie separujące dane trzech klas od siebie. Nanosząc linie o podanych współrzędnych na rysunek przedstawiający ułożenie rekordów można zauważyć, że dane zostały prawidłowo od siebie odseparowane. Można dokładnie zauważyć obszary dla klasy 1, 2 oraz 3.

PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
0	1.11	3	100	98.89

Tab. Uzyskane wyniki testowania drzewa maksymalnego dla danych testowych oraz uczących

Analizując uzyskane wyniki poprzez obserwację wskaźników RD (liczba reguł), PPBS (Procent Przykładów Błędnie Sklasyfikowanych), PPDDT (Procent Poprawnych Decyzji dla Danych Testowych) oraz PPDDU (Procent Poprawnych Decyzji dla Danych Uczących) można wywnioskować, iż dane, dla których uruchomiono drzewo były dla systemu łatwe do klasyfikacji. Ponadto proces przycinania drzewa był w tym przypadku niepotrzebny, ponieważ system zbudował drzewo z ilością liści równą liczbie klas poprawnie klasyfikując dane wejściowe.

Drugim typem danych, dla których skonstruowano drzewo decyzyjne były dane uczące liniowo nieseparowalne w postaci 120 rekordów. W przypadku tego typu danych uzyskano na początku drzewo maksymalne o minimalnej liczbie przykładów tworzących liść drzewa równej 1 oraz o liczbie reguł równej 13. Liczba reguł była znacząco większa od ilości klas – 3 klasy. Testując uzyskane drzewo maksymalne dla danych testowych – 90 rekordów oraz danych uczących – 120 rekordów uzyskano poprawne wyniki. Analogicznie do wyników dla testowania drzewa dla danych liniowo separowalnych, także tu wskaźnik PPBS – DT (Procent Przykładów Błędnie sklasyfikowanych dla Danych Testowych) wynosił jedynie 1.11%, a dla danych uczących był równy 0%. Ponadto na podstawie drzewa rozpisanego w formie zbioru reguł wyznaczono nowe linie separujące dane trzech klas od siebie. Analizując uzyskany rysunek można zauważyć, że dane wszystkich klas zostały od siebie prawidłowo odseparowane.

PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
0	1.11	13	100	98.89

Tab. Uzyskane wyniki testowania drzewa maksymalnego dla danych testowych oraz uczących

Analizując uzyskane wyniki poprzez obserwację wskaźników RD (liczba reguł), PPBS (Procent Przykładów Błędnie Sklasyfikowanych), PPDDT (Procent Poprawnych Decyzji dla Danych Testowych) oraz PPDDU (Procent Poprawnych Decyzji dla Danych Uczących) można wywnioskować, iż dane, dla których uruchomiono drzewo były dla systemu łatwe do klasyfikacji. Następnie dokonano procesu przycinania drzewa poprzez zwiększanie parametru – minimalnej liczby przykładów tworzących liść drzewa. Zmniejszając odpowiednio ten parametr uzyskano drzewo przycięte w połowie. Za drzewo przycięte w połowie uważamy drzewo decyzyjne o liczbie reguł równej 6, czyli posiada ono o połowę mniej liści względem drzewa maksymalnego. Testując uzyskane drzewo dla danych testowych oraz uczących uzyskano znacząco gorsze wyniki względem testowania dla drzewa maksymalnego. Wskaźnik PPBS – DT (Procent Przykładów Błędnie sklasyfikowanych dla Danych Testowych) wynosił 13.11%, czyli o 12% więcej rekordów zostało źle sklasyfikowanych. Wskaźnik PPBS – DU (Procent Przykładów Błędnie sklasyfikowanych dla Danych Uczących) wynosił 10.83%, więc jest znacząco większy, ponieważ dla drzewa maksymalnego wyniósł on 0%.

PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
10.83	13.11	6	89.27	86.89

Tab. Uzyskane wyniki testowania drzewa w połowie przyciętego o dla danych testowych oraz uczących

Ponadto wyznaczono nowe linie separujące dane trzech klas od siebie. Nanosząc odpowiednie współrzędne linii na rysunek przedstawiający dane uczące można zauważyć, że są one niedokładnie od siebie odseparowane. Występują obszary, w których równocześnie istnieją obiekty klasy 1 oraz klasy 2.

Proces przycinania kontynuowano do momentu uzyskania drzewa decyzyjnego maksymalnie przyciętego, czyli takiego o liczbie reguł równej liczbie klas. W tym przypadku liczb reguł wynosi 3. Drzewo maksymalnie przycięte uzyskano dla minimalnej liczby przykładów tworzących liści drzewa równej 20. Testując uzyskane drzewo dla danych testowych oraz uczących uzyskano znacząco gorsze wyniki względem testowania dla drzewa w połowie przyciętego oraz minimalnego. Wskaźnik PPBS – DT (Procent Przykładów Błędnie sklasyfikowanych dla Danych Testowych) wynosił 32.22%, czyli o 22% więcej rekordów zostało źle sklasyfikowanych względem drzewa przyciętego w połowie. Wskaźnik PPBS – DU (Procent Przykładów Błędnie sklasyfikowanych dla Danych Uczących) wynosił 33.33%, czyli co trzeci rekord danych uczących został przez system źle sklasyfikowany.

PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
33.33	32.22	3	66.67	67.78

Tab. Uzyskane wyniki testowania drzewa maksymalnie przyciętego dla danych testowych oraz uczących

Ponadto wyznaczono nowe linie separujące dane trzech klas od siebie. Nanosząc odpowiednie współrzędne linii na rysunek przedstawiający dane uczące można zauważyć, że są one bardzo niedokładnie od siebie odseparowane. W uzyskanych obszarach mieszają się rekordy wszystkich trzech klas, np.: obszary środkowym pomiędzy dwoma liniami znalazły się w wszystkie obiekty klasy 3, lecz również występują pojedyncze rekordy klas 1 oraz 2.

Analizując wszystkie wykonane procesy przycinania drzewa decyzyjnego można wywnioskować, iż system umiał prawidłowo sklasyfikować dane tylko bazując na drzewie maksymalnym, które zawierało o wiele większą ilość liści względem ilości klas. Przycinając drzewo zwiększając parametr - minimalna liczba przykładów tworzących liści drzewa uzyskiwano drzewo decyzyjne o mniejszej ilości reguł doprowadzając do drzewa przyciętego w połowie oraz maksymalnie przyciętego. Analizując kolejno wyniki wskaźników PPBS – DT oraz PPBS – DU można zauważyć, iż zwiększały się one z kolejnym etapem przycinania drzewa, czyli zmniejszając ilość reguł zwiększa się procent przykładów błędnie sklasyfikowanych przez system zarówno dla danych testowych jak i danych uczących liniowo nieseparowalnych.

Ostatnim typem danych, dla których skonstruowano drzewa decyzyjne były dane uczące graficzne o matrycach 8x10 oraz 16x15 w postaci 50 obrazów oraz dane testowe graficzne o matrycach 8x10 i 16x15 pikseli – 25 obrazów. Dane testowe stanowiły zaszumione obrazy wzorcowe o różnym stopniu zaszumienia. W przypadku tego typu danych zarówno dla obrazów 8x10 pikseli jak i 16x15 pikseli uzyskano drzewo maksymalne o minimalnej liczbie przykładów tworzących liści drzewa równej 1 oraz o liczbie reguł równej 5. Liczba liści – reguł była równa liczbie klas. Testując uzyskane drzewo maksymalne dla danych testowych – 25 rekordów oraz danych uczących – 50 rekordów uzyskano rozbieżne wyniki. Dla danych uczących, czyli obrazów wzorcowych zniekształconych maksymalnie o 10% względem wzorca system potrafił poprawnie sklasyfikować wszystkie obrazy wejściowe – zatem wskaźnik PPBS – DT (Procent Przykładów Błędnie sklasyfikowanych dla Danych Testowych) oraz PPBS – DU (Procent Przykładów Błędnie sklasyfikowanych dla Danych Uczących) wynosił 0%. Z kolei dla danych testowych system potrafił sklasyfikować poprawnie mniej niż połowę danych wejściowych. Wskaźnik PPBS-DT wyniósł 52%, zatem procent poprawnych decyzji dla danych testowych (PPDDT) wyniósł jedynie 48%.

PPBS - DU [%]	PPBS - DT [%]	RD (liczba liści (reguł))	PPDDU [%]	PPDDT [%]
0	52	5	100	48

Tab. Uzyskane wyniki testowania drzewa maksymalnego dla danych uczących i testowych – matryce 8x10 px oraz 16x15 px.

Analizując uzyskane drzewa decyzyjne oraz wyniki dla danych graficznych można stwierdzić, iż system potrafił zbudować drzewo o małej ilości reguł równej ilości klas, lecz zbiór otrzymanych reguł nie potrafił poprawnie sklasyfikować danych testowych. Ponadto zmiana matrycy obrazów z 8x10 na 16x15 pikseli nie spowodowała dużych zmian w strukturze drzew decyzyjnych. Ich wygląd był bardzo podobny do siebie. Jedyną różnicą była zmiana atrybutów, dla których występowały rozgałęzienia. Wyniki klasyfikacji danych testowych jak i wejściowych pozostawała bez zmian.

Reasumując można stwierdzić, iż dla każdego rodzaju danych uzyskiwane drzewa decyzyjne różniły się. Znaczące różnice w budowie drzewa można było zauważyć w przypadku drzewa decyzyjnego maksymalnego dla danych liniowo nieseparowalnych – uzyskane drzewo posiadało znacząco większą ilość reguł względem ilości klas. Dzięki temu można było przyciąć drzewo maksymalne i zauważyć, iż przycinanie uzyskiwanie zwiększaniem minimalnej liczby przykładów tworzących liści drzewa powodowało wzrost procentu przykładów błędnie sklasyfikowanych. Dla danych liniowo separowalnych jak i graficznych system budował drzewa o ilości reguł równej ilości klas. W przypadku danych liniowo separowalnych system bezproblemowo umiał prawidłowo sklasyfikować rekordy do domyślnej klasy. Analogiczna sytuacja odnosi się do danych graficznych – tutaj również system sklasyfikował obrazy prawidłowo. Zmiana matrycy obrazów nie miała żadnego wpływu. Uzyskane wyniki klasyfikacji pozostawały takie same. Podsumowując, dla każdego typu danych, testując uzyskiwane drzewo maksymalne system potrafił prawidłowo sklasyfikować dane wejściowe – wyjątkiem są tu zaszumione obrazy graficzne.