

# Projekt - Drzewa Decyzyjne II

Sebastian Michoń 136770, Marcin Zatorski 136834  
grupe L5

## 1 Wybrane najważniejsze atrybuty

1. sex - nawet jeśli płeć sama w sobie może nie być dobrze skorelowana z tym, czy student zdał, wpływa ona na to, jak dany atrybut wpływa na studenta - np. dla mężczyzn wyższą korelację z wynikiem końcowym może mieć chęć podjęcia edukacji wyższej raczej niż zdrowie, dla kobiet - wręcz przeciwnie.
2. reason - to, dlaczego student wybrał daną szkołę.
3. failures
4. higher
5. Dalc - dzienne spożycie alkoholu w dni robocze
6. health
7. absences

## 2 Wybrane metryki

1. Accuracy - jako, że dataset jest dosyć zrównoważony (115/85 dla zestawu treningowego i 118/77 dla testowego) wykorzystanie procentu trafień jest zasadne.
2. F1 measure - jeśli jakaś inna miara poza celnością jest zasadna, to najprędzej F1 - pozwala ona bowiem zagregować informację o precyzji i czułości
3. Ostatnią miarą, która zostanie użyta jest Balanced accuracy -  $\frac{TPR+TNR}{2}$ , czyli suma czułości i selektywności podzielona przez 2 - nieco lepsza niż Accuracy, jako że proporcja w tym datasetcie to 3:2 (gdzie więcej osób nie zdało niż zdało)

Table 1: Wyniki algorytmu J48 dla wybranych atrybutów - Math dataset

Binary Split	Confidence factor	Minimum objects	TP	FP	FN	TN	Accuracy	F1	Balanced Accuracy
0	0.1000	2	31	37	46	81	0.5744	0.4276	0.5445
0	0.2000	2	31	37	46	81	0.5744	0.4276	0.5445
0	0.3000	2	31	37	46	81	0.5744	0.4276	0.5445
0	0.4000	2	31	37	46	81	0.5744	0.4276	0.5445
0	0.5000	2	27	28	50	90	0.6000	0.4091	0.5567
0	0.6000	2	28	35	49	83	0.5692	0.4000	0.5335
0	0.7000	2	28	35	49	83	0.5692	0.4000	0.5335
0	0.8000	2	28	35	49	83	0.5692	0.4000	0.5335
0	0.9000	2	28	35	49	83	0.5692	0.4000	0.5335
1	0.1000	2	31	37	46	81	0.5744	0.4276	0.5445
1	0.2000	2	31	37	46	81	0.5744	0.4276	0.5445
1	0.3000	2	27	29	50	89	0.5949	0.4060	0.5524
1	0.4000	2	26	23	51	95	0.6205	0.4127	0.5714
1	0.5000	2	26	23	51	95	0.6205	0.4127	0.5714
1	0.6000	2	30	30	47	88	0.6051	0.4380	0.5677
1	0.7000	2	30	30	47	88	0.6051	0.4380	0.5677
1	0.8000	2	30	30	47	88	0.6051	0.4380	0.5677
1	0.9000	2	30	30	47	88	0.6051	0.4380	0.5677
0	0.1000	3	31	37	46	81	0.5744	0.4276	0.5445
0	0.2000	3	31	37	46	81	0.5744	0.4276	0.5445
0	0.3000	3	31	37	46	81	0.5744	0.4276	0.5445
0	0.4000	3	31	37	46	81	0.5744	0.4276	0.5445
0	0.5000	3	23	27	54	91	0.5846	0.3622	0.5349
0	0.6000	3	24	34	53	84	0.5538	0.3556	0.5118
0	0.7000	3	24	34	53	84	0.5538	0.3556	0.5118
0	0.8000	3	24	34	53	84	0.5538	0.3556	0.5118
0	0.9000	3	24	34	53	84	0.5538	0.3556	0.5118
1	0.1000	3	31	37	46	81	0.5744	0.4276	0.5445
1	0.2000	3	31	37	46	81	0.5744	0.4276	0.5445
1	0.3000	3	31	37	46	81	0.5744	0.4276	0.5445
1	0.4000	3	25	29	52	89	0.5846	0.3817	0.5395
1	0.5000	3	25	25	52	93	0.6051	0.3937	0.5564
1	0.6000	3	28	28	49	90	0.6051	0.4211	0.5632
1	0.7000	3	28	28	49	90	0.6051	0.4211	0.5632
1	0.8000	3	28	28	49	90	0.6051	0.4211	0.5632
1	0.9000	3	28	28	49	90	0.6051	0.4211	0.5632
0	0.1000	4	31	37	46	81	0.5744	0.4276	0.5445
0	0.2000	4	31	37	46	81	0.5744	0.4276	0.5445
0	0.3000	4	31	37	46	81	0.5744	0.4276	0.5445
0	0.4000	4	31	37	46	81	0.5744	0.4276	0.5445
0	0.5000	4	31	37	46	81	0.5744	0.4276	0.5445
0	0.6000	4	25	27	52	91	0.5949	0.3876	0.5479
0	0.7000	4	25	27	52	91	0.5949	0.3876	0.5479
0	0.8000	4	25	27	52	91	0.5949	0.3876	0.5479
0	0.9000	4	25	27	52	91	0.5949	0.3876	0.5479
1	0.1000	4	31	37	46	81	0.5744	0.4276	0.5445
1	0.2000	4	31	37	46	81	0.5744	0.4276	0.5445
1	0.3000	4	25	28	52	90	0.5897	0.3846	0.5437
1	0.4000	4	25	28	52	90	0.5897	0.3846	0.5437
1	0.5000	4	25	24	52	94	0.6103	0.3968	0.5606
1	0.6000	4	25	24	52	94	0.6103	0.3968	0.5606
1	0.7000	4	25	24	52	94	0.6103	0.3968	0.5606
1	0.8000	4	25	24	52	94	0.6103	0.3968	0.5606
1	0.9000	4	25	24	52	94	0.6103	0.3968	0.5606
0	0.1000	5	40	53	37	65	0.5385	0.4706	0.5352
0	0.2000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.3000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.4000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.5000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.6000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.7000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.8000	5	19	16	58	102	0.6205	0.3393	0.5556
0	0.9000	5	19	16	58	102	0.6205	0.3393	0.5556
1	0.1000	5	36	49	41	69	0.5385	0.4444	0.5261
1	0.2000	5	22	18	55	100	0.6256	0.3761	0.5666
1	0.3000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.4000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.5000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.6000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.7000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.8000	5	24	27	53	91	0.5897	0.3750	0.5414
1	0.9000	5	24	27	53	91	0.5897	0.3750	0.5414

### 3 Działanie algorytmu - Generowane drzewa

#### 3.1 Dyskretny Age

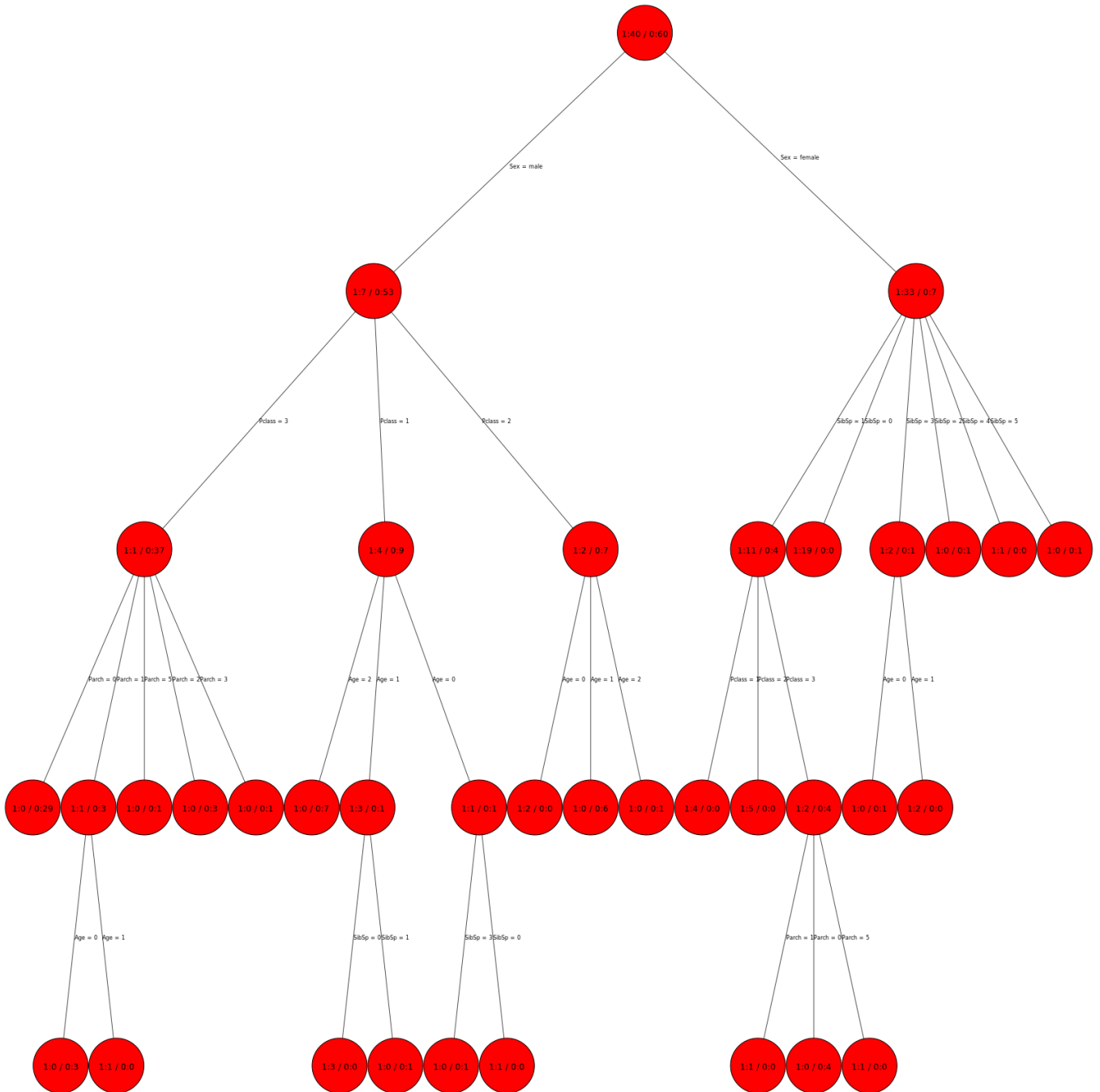


Figure 1: Age jest atrybutem dyskretnym

## 3.2 Ciągły Age

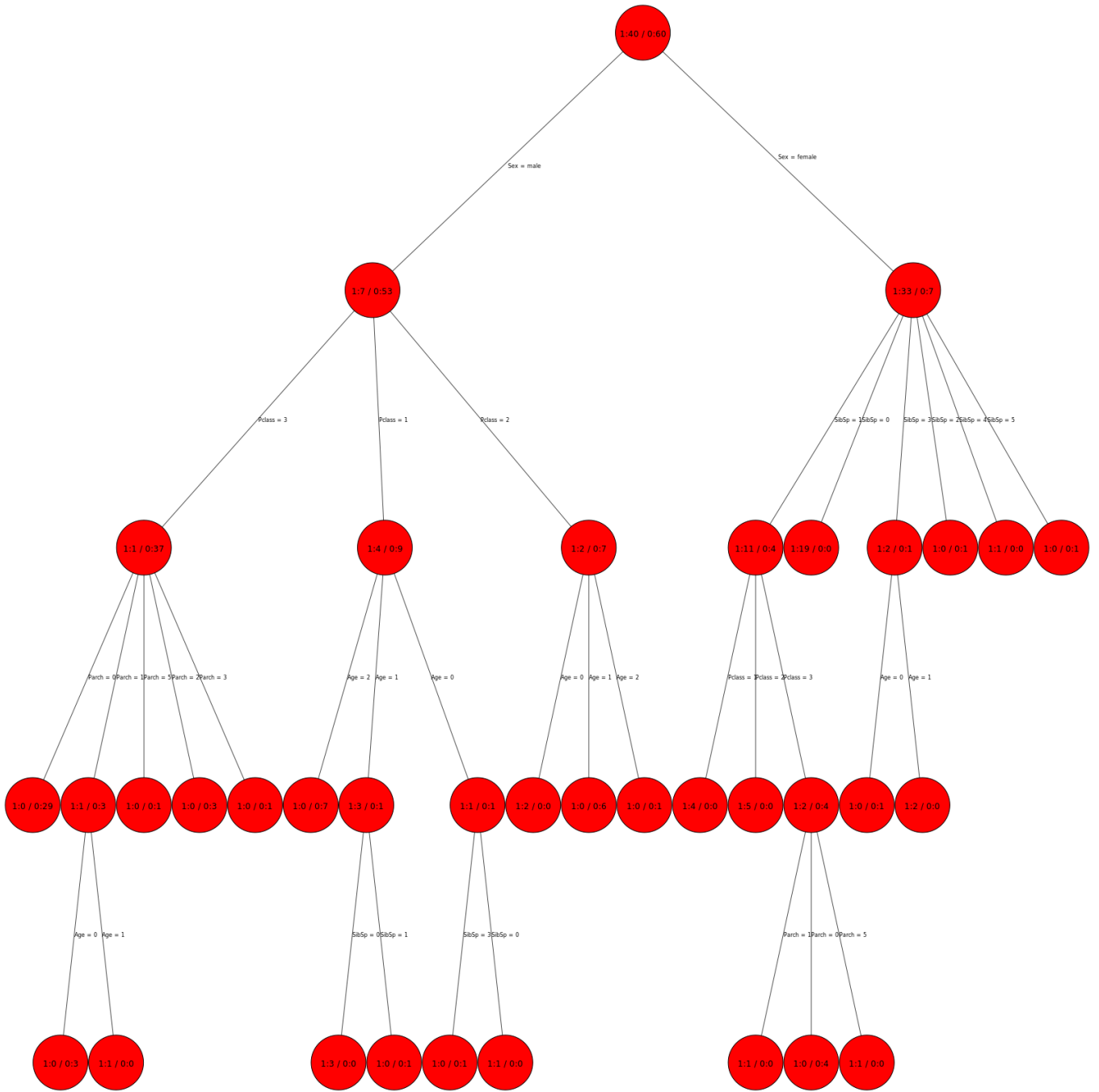


Figure 2: Age jest atrybutem ciągłym