

Tiesioginio sklidimo DNT Wekoje

Pijus Petkevičius

2023 m. lapkričio 27 d.

Turinys

1	Įvadas	3
1.1	Tikslas	3
1.2	Uždaviniai	3
2	Eksperimentų vykdymas	4
2.1	Duomenys	4
2.1.1	Duomenų paruošimas	4
2.1.2	Duomenų savybės	4
2.2	Užduočių sekos	6
2.2.1	Neuroninio tinklo hiperparametrų parinkimo užduočių seka	6
2.2.2	Naujų duomenų klasifikavimo užduočių seka	7
2.2.3	Klasifikavimo ir testavimo užduočių seka	8
2.3	Excel skaičiavimai	8
3	Rezultatai	9
3.1	Neuroninio tinklo hiperparametrų parinkimas	9
3.2	Naujų duomenų klasifikavimas	10
3.3	Klasifikavimas ir testavimas	11
3.4	Excel skaičiavimų ir WEKA palyginimas	11
4	Išvados	13

1 Įvadas

1.1 Tikslas

Išmokyti neuroninį tinklą teisingai klasifikuoti duomenis naudojant sistemą WEKA.

1.2 Uždaviniai

- Paruošti duomenis išskaidant juos į mokymo-testavimo ir naujus, nepriširtus jokiai klasei, duomenis.
- Sukonstruoti WEKA sistemoje tokias užduočių sekas:
 - bandymų atlikimui, skirto parinkti tinkamus dirbtinio neuroninio tinklo parametrus
 - naujų duomenų klasifikavimo
 - klasifikavimo ir testavimo
- Atlikti bandymus su pirmąja užduočių seka, nustatyti tinkamus dirbtinio neuroninio tinklo parametrus.
- Suklasifikuoti naujus duomenis su gautu DNT
- Atlikti paskutinę užduočių seką ir nustatyti DNT svorius, kurie bus naudojami Excel skaičiavimuose
- Paskaičiuoti neuronų išėjimų reikšmes su gautais DNT svoriais Excel programoje
- Palyginti gautas neuronų išėjimų reikšmes WEKA programoje ir Excel.

2 Eksperimentų vykdymas

2.1 Duomenys

2.1.1 Duomenų paruošimas

Tinklo mokymui ir testavimui panaudota 120 duomenų eilučių (atitinkamai po 40 kiekvienos klasės duomenų eilučių).

Panaudota 30 duomenų eilučių su nežinomomis klasėmis (atitinkamai parinkta po 10 kiekvienos klasės duomenų eilučių).

Pirmoje [Neuroninio tinklo hiperparametrų parinkimo užduočių sekoje](#) panaudota kryžminė patikra. Duomenys buvo 10 kartų skirtingai padalinti į mokymo ir testavimo aibes.

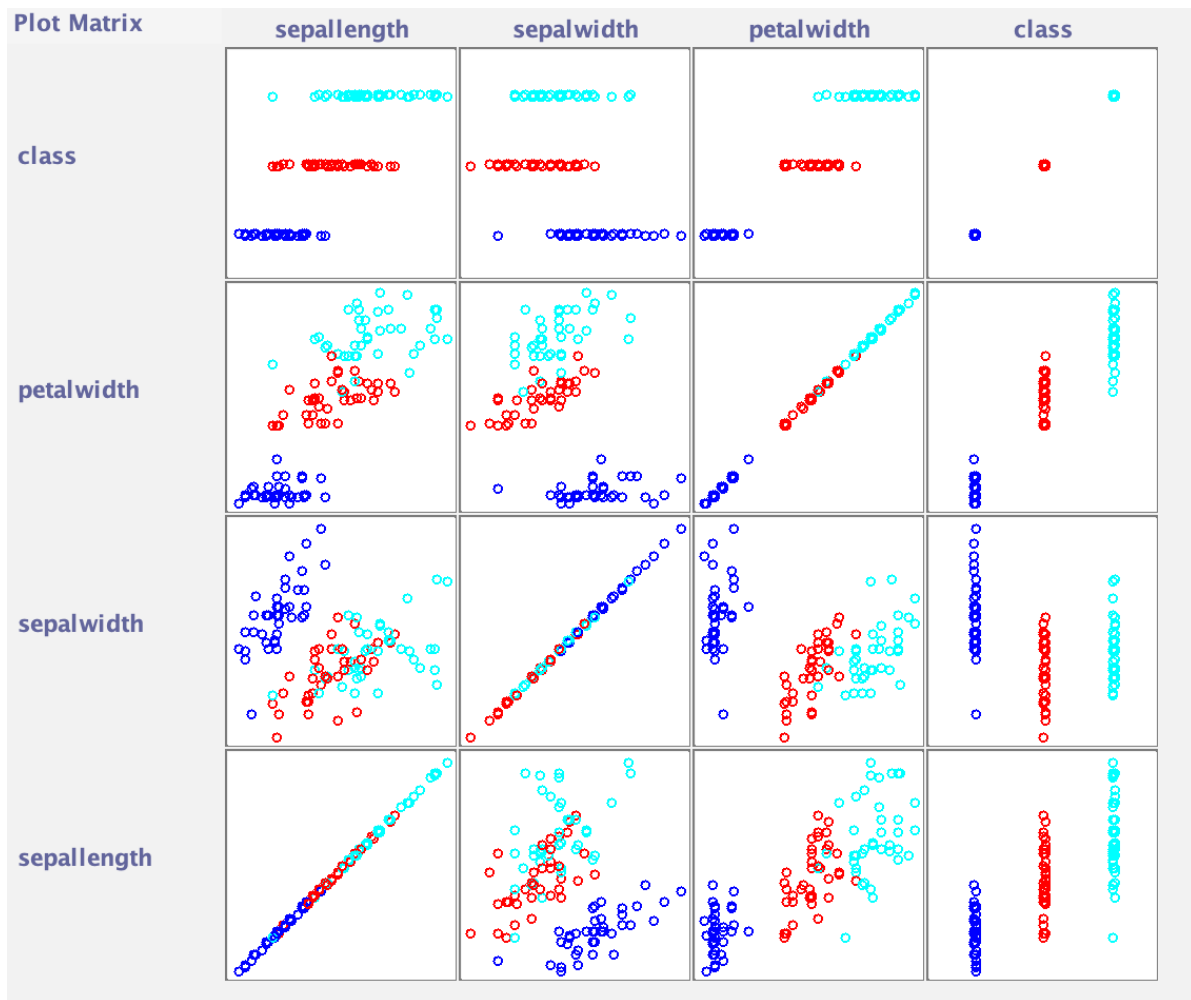
Turint 0 užduoties variantą, iš irisų duomenų aibės buvo pasirinkta tokie požymiai:

- sepallength
- sepalwidth
- petalwidth

2.1.2 Duomenų savybės

[1 pav.](#) pavaizduota mokymo ir testavimo aibės duomenų požymių porų vaizdai. Mėlyna spalva pažymėta Iris-setosa, raudona spalva Iris-versicolor ir žydra spalva Iris-Virginica.

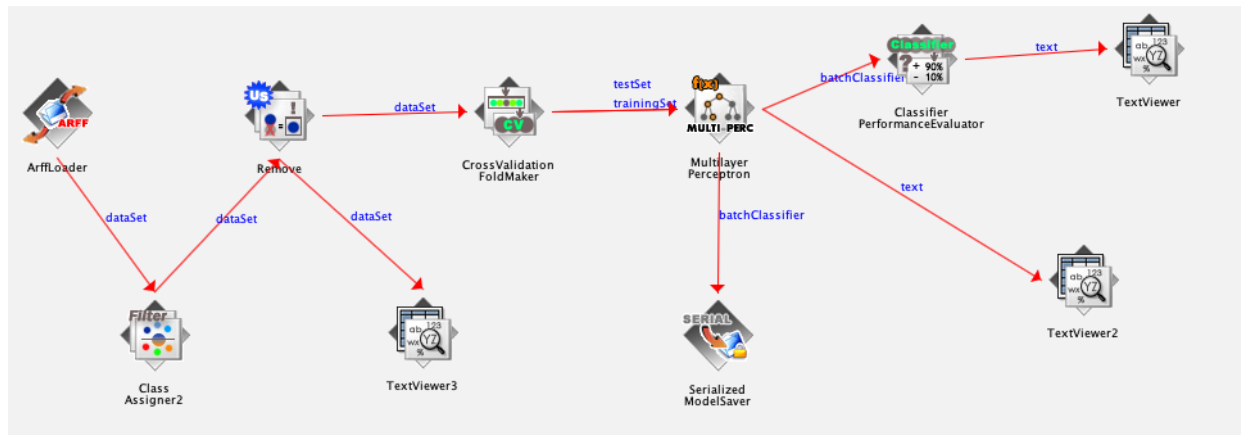
[1 pav.](#) galime pastebėti, kad Iris-versicolor ir Iris-virginica požymių porų reikšmės ganėtinai artimos viena kitai, dažnai sutampa, o Iris-setosa klasės požymiai išsiskiria iš kitų dviejų klasių.



1 pav.: Duomenų požymių porų vaizdai

2.2 Užduočių sekos

2.2.1 Neuroninio tinklo hiperparametrų parinkimo užduočių seka

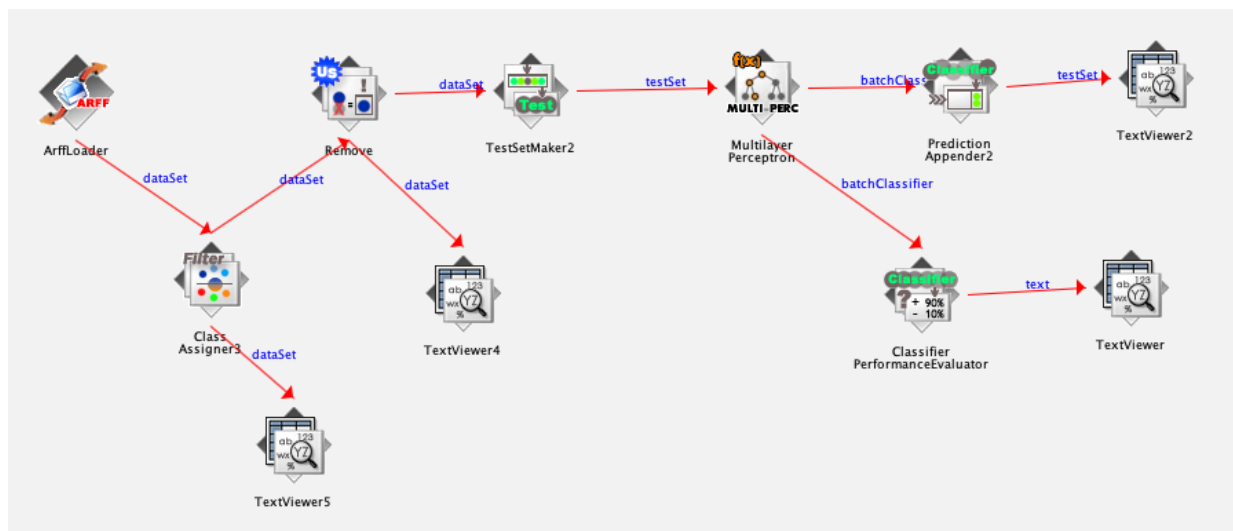


2 pav.: Neuroninio tinklo hiperparametrų parinkimo užduočių seka

2 pav. užduočių sekos komponentės:

- *ArffLoader* - skaito arff tipo įvesties duomenų failą.
- *ClassAssigner* - nurodoma, kuris duomenų požymis yra klasė.
- *Remove* - nurodoma, kurie duomenų aibės parametrai yra pašalinami.
- *CrossValidationFoldMaker* - sukuria skirtingus kryžminės patikros duomenų pjūvius. Išskaidoma į mokymo ir testavimo aibes.
- *ClassifierPerformanceEvaluator* - parodomas klasifikavimo rezultatų tikslumas.
- *Textviewer* - surenka ir atvaizduoja gautus tekstinius duomenis.

2.2.2 Naujų duomenų klasifikavimo užduočių seka

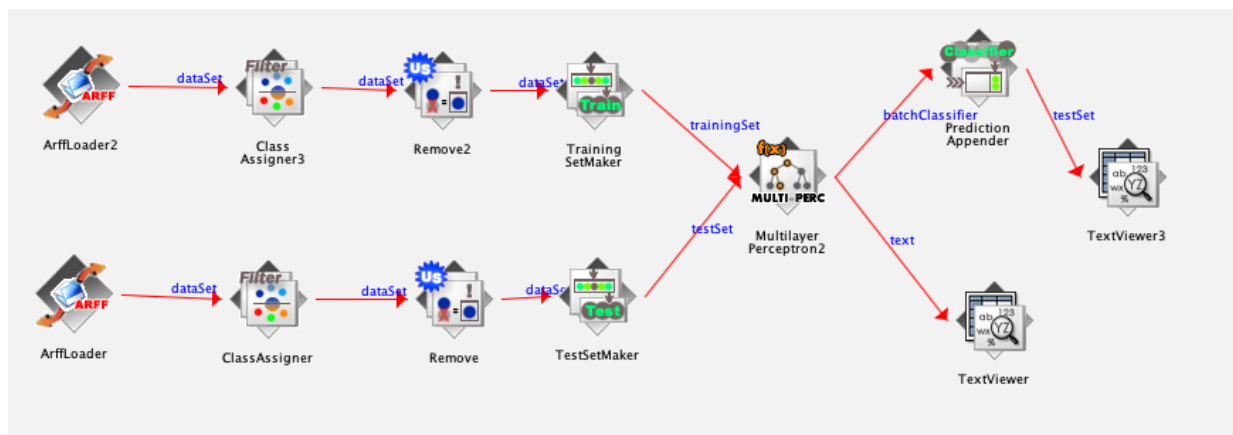


3 pav.: Naujų duomenų klasifikavimo užduočių seka

3 pav. užduočių sekos komponentės:

- *TestSetMaker* - sukuria testavimo duomenų aibes iš pateiktų duomenų.
- *PredictionAppender* - prognozuoja duomenų klasių reikšmes.
- *ArffSaver* - sukuria arff tipo duomenų failą.

2.2.3 Klasifikavimo ir testavimo užduočių seka



4 pav.: Klasifikavimo ir testavimo užduočių seka

4 pav. užduočių sekos komponentės:

- *TrainingSetMaker* - sukuria mokymo duomenų aibę iš pateiktų duomenų.

2.3 Excel skaičiavimai

1. Įsikelti naujų duomenų klasifikavimo rezultatus su tikimybėmis, kad eilutė priskirta kuriai nors iš klasių, gauti WEKA programoje.
2. Normalizuoti duomenų įėjis, kad kiekvieno iš parametrų reikšmės priklausytų intervalui $[0; 1]$.
3. Įsikelti paslėpto ir paskutinio sluoksnio neuronų svoriai, gauti WEKA.
4. Paskaičiuoti aktyvacijos funkcijos parametrus a kiekvienai iš normalizuotų duomenų eilučių, naudojantis paslėpto sluoksnio neuronų svoriais.
5. Paskaičiuoti sigmoidinės funkcijos reikšmes kiekvienam iš paskaičiuotų a parametrų.
6. Paskaičiuoti aktyvacijos funkcijos parametrų a kiekvienam iš 5 žingsnyje paskaičiuotų rezultatų, naudojantis paskutinio sluoksnio neuronų svoriais.
7. Paskaičiuoti sigmoidinės funkcijos reikšmes kiekvienam iš paskaičiuotų a parametrų. Paskaičiavus gaunamos tikimybės, kad duomenų eilutė priklauso vienai ar kitai irisų klasei.

3 Rezultatai

3.1 Neuroninio tinklo hiperparametrų parinkimas

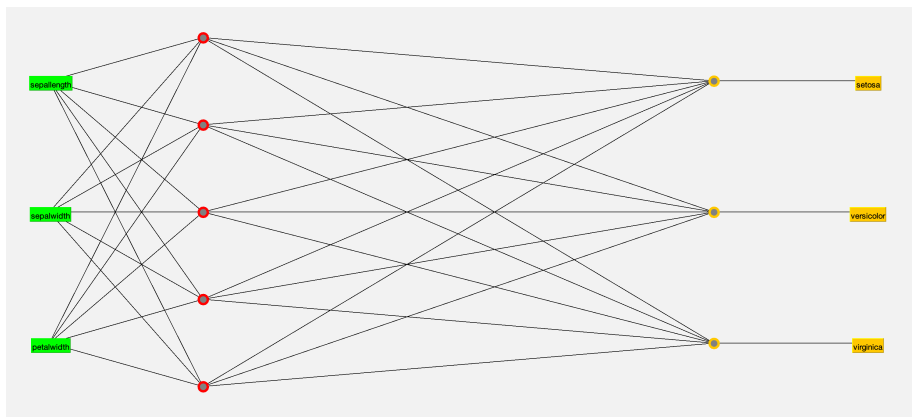
Atlikti bandymai su 27 skirtingomis paslėpto sluoksnio neurono skaičiaus, mokymo greičio ir *momentum* reikšmių kombinacijomis, kur paslėptų neuronų skaičius yra reikšmė iš aibės $\{3; 5; 7\}$, $\eta \in \{0,3; 0,6; 0,9\}$, *momentum* $\in \{0,3; 0,6; 0,9\}$. Epochų skaičius buvo 500. 1 lentelėje pavaizduoti visų bandymų klasifikavimo tikslumo rezultatai.

1 lentelė: Klasifikavimo tikslumas su skirtingais hiperparametrais.

Nr.	Paslėptų neuronų skaičius	Mokymo greitis	Momentum	Klasifikavimo tikslumas
1	3	0,3	0,3	0,9583
2	3	0,3	0,6	0,9583
3	3	0,3	0,9	0,95
4	5	0,3	0,3	0,9583
5	5	0,3	0,6	0,9583
6	5	0,3	0,9	0,9583
7	7	0,3	0,3	0,9583
8	7	0,3	0,6	0,9583
9	7	0,3	0,9	0,95
10	3	0,6	0,3	0,9667
11	3	0,6	0,6	0,9667
12	3	0,6	0,9	0,9583
13	5	0,6	0,3	0,9667
14	5	0,6	0,6	0,9667
15	5	0,6	0,9	0,95
16	7	0,6	0,3	0,9667
17	7	0,6	0,6	0,9667
18	7	0,6	0,9	0,9583
19	3	0,9	0,3	0,9667
20	3	0,9	0,6	0,9667
21	3	0,9	0,9	0,95
22	5	0,9	0,3	0,9667
23	5	0,9	0,6	0,9667
24	5	0,9	0,9	0,975
25	7	0,9	0,3	0,9667
26	7	0,9	0,6	0,95
27	7	0,9	0,9	0,9583

Gauti geriausi klasifikavimo rezultatai paryškinti 1 lentelėje. Vienas iš geriausių klasifikavimo tikslumų buvo pasiektas su 5 paslėptais neuronais, $\eta = 0,9$ ir *mo-*

momentum reikšmė 0,9, šios parametų reikšmės bus ir toliau naudojamos šiame darbe. Gautas neuroninis tinklas su 5 neuronais paslėptame sluoksnyje atvaizduotas 5 pav.



5 pav.: Neuroninio tinklo su 5 neuronais paslėptame sluoksnyje vaizdas

3.2 Naujų duomenų klasifikavimas

Žemiau pateikta, kaip tinklas su 5 paslėptais neuronais, $\eta = 0,9$, *momentum* reikšmė 0,9 ir epochų skaičiumi 500 suklasifikavo naujus duomenis, kuriems nebuvo priskirta jokia klasė. Naujų duomenų rinkinyje buvo 30 ir 28 iš jų buvo teisingai suklasifikuoti, buvo pasiektas klasifikavimo tikslumas lygus 0,93. Testavimui skirta duomenų aibė, su jos parametrais, iš anksto priskirta reikšmė ir neuroninio tinklo priskirta reikšmė matome žemiau:

```

1 6.8,3.2,2.3,virginica,virginica
2 4.8,3,0.1,setosa,setosa
3 4.9,3.6,0.1,setosa,setosa
4 5.7,3.8,0.3,setosa,setosa
5 4.9,3,0.2,setosa,setosa
6 6.7,3.1,2.4,virginica,virginica
7 5.9,3,1.8,virginica,virginica
8 5.8,2.6,1.2,versicolor,versicolor
9 6.5,3,1.8,virginica,virginica
10 6.8,3,2.1,virginica,virginica
11 6,2.2,1.5,virginica,virginica
12 6.3,3.3,2.5,virginica,virginica
13 5.1,3.8,0.3,setosa,setosa
14 5.1,3.8,0.2,setosa,setosa
15 6.4,3.1,1.8,virginica,virginica
16 5.5,2.6,1.2,versicolor,versicolor
17 6,2.9,1.5,versicolor,versicolor

```

```

18 5.8,2.8,2.4, virginica, virginica
19 6.1,2.8,1.2, versicolor, versicolor
20 6.7,3,1.7, versicolor, virginica
21 6.2,2.2,1.5, versicolor, virginica
22 7.2,3.2,1.8, virginica, virginica
23 4.6,3.1,0.2, setosa, setosa
24 4.7,3.2,0.2, setosa, setosa
25 5.1,3.3,0.5, setosa, setosa
26 6.8,2.8,1.4, versicolor, versicolor
27 5.7,2.9,1.3, versicolor, versicolor
28 5.1,3.4,0.2, setosa, setosa
29 5.4,3,1.5, versicolor, versicolor
30 6.3,2.5,1.5, versicolor, versicolor

```

3.3 Klasifikavimas ir testavimas

Nustatyta, kad paslėptame sluoksnyje būtų 5 neuronai, $\eta = 0,9$, *momentum* reikšmė 0,9 ir epochų skaičius 500. Gauti paslėpto sluoksnio svoriai pavaizduoti [2 lentelėje](#), o paskutinio sluoksnio [3 lentelėje](#).

2 lentelė: Paslėpto sluoksnio svoriai.

Node	Threshold	sepal length	sepal width	petal width
3	-14.8646	10.6758	-10.2068	32.0955
4	-12.5104	8.1166	-8.7992	25.3611
5	-11.4965	-0.5536	-8.2919	18.3798
6	-14.9061	-0.8164	-10.5284	28.0188
7	6.2073	4.3683	-5.1927	9.6986

3 lentelė: Paskutinio sluoksnio svoriai.

Node	Threshold	Node 3	Node 4	Node 5	Node 6	Node 7
0	6.0227	-4.8094	-4.8904	-4.6948	-4.2520	-11.9854
1	-6.8969	-8.4217	-6.0713	-4.2192	-7.1382	10.5172
2	-7.6455	8.2959	6.2669	4.6664	6.7282	4.0430

3.4 Excel skaičiavimų ir WEKA palyginimas

[4 lentelėje](#) atvaizduotos tikimybės, kuriai įrisų duomenų klasei priklauso duomenų eilutė. Joje pavaizduojamos reikšmės gautos Excel skaičiavimuose ir WEKA. Lentelėje paryškintos tikimybės, kurios yra didžiausios duotai eilutei.

Didžiausios tikimybės, paskaičiuotos Excel ir WEKA, yra tos pačios kiekvienai duomenų eilutei. Gautos tikimybės skirtingais skaičiavimo būdais skiriasi ne

daugiau nei 0,25 kiekvienai duomenų eilutei, galimas atvejis, kad neuronų svoriai, naudoti Excel, turėjo per mažai skaičių po kablelio, ir paskaičiavus reikšmes neuroniniame tinkle, nukrypo(akumuliuosisi paklaidos).

4 lentelė: Excel skaičiavimuose ir WEKA gautos tikimybės.

Nr.	Tikimybės, gautos Excel			Tikimybės, gautos WEKA		
	setosa	versicolor	virginica	setosa	versicolor	virginica
1	0	0	1	0	0	1
2	0,9976	0,0010	0,0005	0,998471	0,001044	0,000484
3	0,9976	0,0010	0,0005	0,998507	0,001014	0,000479
4	0,9976	0,0010	0,0005	0,998455	0,001058	0,000487
5	0,9975	0,0010	0,0005	0,998396	0,001108	0,000496
6	0	0	1	0	0	1
7	0,0004	0,7215	0,2810	0,000075	0,156729	0,843196
8	0,0026	0,9738	0,0265	0,002578	0,971013	0,026409
9	0	0	1	0	0,000029	0,999971
10	0	0	1	0	0	1
11	0	0	1	0,00008	0,237517	0,762403
12	0	0	1	0	0	1
13	0,9976	0,0010	0,0004	0,998498	0,001022	0,00048
14	0,0026	0,9739	0,0265	0,998505	0,001016	0,000479
15	0	0,0019	0,9982	0,000003	0,00121	0,998787
16	0,0026	0,9736	0,0264	0,002602	0,971059	0,026339
17	0,0025	0,9734	0,0270	0,002512	0,97039	0,027099
18	0	0	1	0	0	1
19	0,0026	0,9734	0,0265	0,002583	0,971033	0,026384
20	0	0,0002	0,9998	0,000005	0,00228	0,997715
21	0	0	1	0,000012	0,017143	0,982845
22	0	0	1	0	0,000021	0,999979
23	0,9976	0,0010	0,0005	0,998476	0,001041	0,000484
24	0,9976	0,0010	0,0005	0,998482	0,001035	0,000483
25	0,9974	0,0011	0,0005	0,997853	0,001579	0,000568
26	0,0011	0,9157	0,0858	0,002284	0,966623	0,031093
27	0,0026	0,9734	0,0263	0,002603	0,971068	0,026329
28	0,9976	0,0010	0,0005	0,998479	0,001038	0,000483
29	0,0026	0,9733	0,0264	0,002571	0,970875	0,026554
30	0	0,0223	0,9786	0,000889	0,888518	0,110593

4 Išvados

- Su WEKA DNT atrinktais svoriais galima atkartoti klasifikavimo procesą Excel. Visgi šiame darbe gautos klasifikavimo tikimybės nėra lygios abejais atvejais ir skiriasi iki 0,25, galimais dėl slankaus kabelio skaičių panaudojimo.
- Jei pasirinktas paslėptų neuronų skaičius yra 5, rekomenduojama pasirinkti $\eta = 0,9$, *momentum* reikšmę 0,9, norint pasiekti 0,975 klasifikavimo tikslumą su irisų duomenimis.