

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Praktinė užduotis Nr. 3
Tiesioginio sklidimo DNT naudojant sistemą WEKA

Atliko:

Programų sistemų 4 k. 1 gr. stud. Rokas Petrauskas

VILNIUS, 2023

Turinys

Tikslas	3
Naudojami duomenys	3
Užduočių sekos	3
Geriausi rasti parametrai	4
Kiti parametrai	4
Neuroninio tinklo vaizdas	6
Naujų duomenų klasifikavimo rezultatai	6
Požymių porų vaizdai	8
Svoriai	8
MS Excel ir WEKA rezultatai	9
MS Excel programoje konstruotas neuroninis tinklas	9
Gauti rezultatai ir išvada	9

Tikslas

Išmokyti neuroninį tinklą teisingai klasifikuoti duomenis naudojant sistemą WEKA.

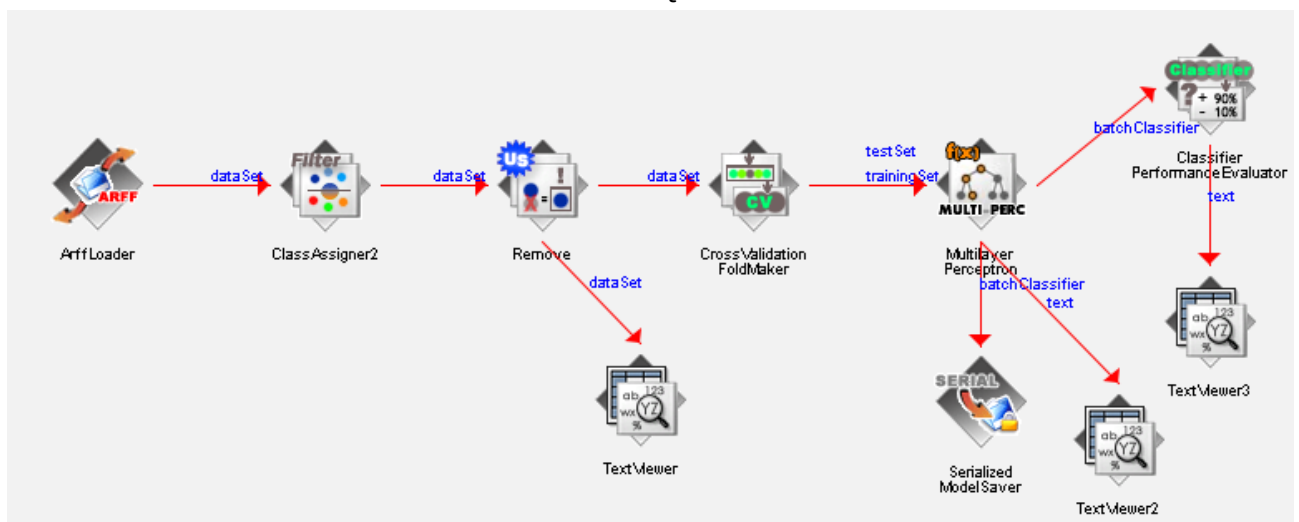
Naudojami duomenys

Dirbtinio neuroninio tinklo mokymui ir testavimui naudota Irisų duomenų aibė:

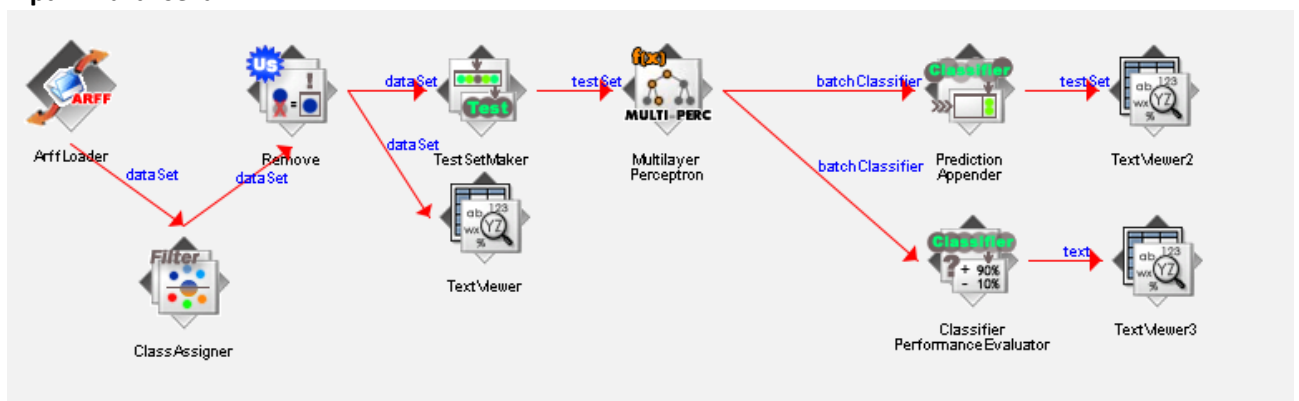
<https://archive.ics.uci.edu/dataset/53/iris>.

Duomenų aibę sudaro 150 įrašų po 4 požymius ir klasės identifikatorių. Mokymui ir testavimui naudojau 3 požymius – sepalwidth, petallength, petalwidth (sepalwidth išmetama). Yra po 50 visų 3 klasių įrašų, todėl mokymui naudojau po 40 įrašų (viso 120), o testavimui po 10 (viso 30).

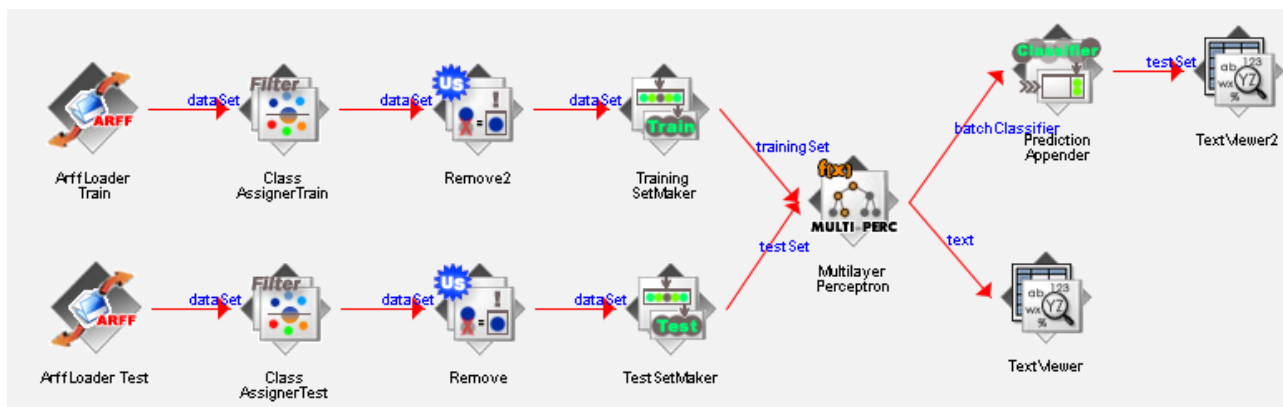
Užduočių sekos



1 pav. 1 užd. seka



2 pav. 2 užd. Seka



3 pav. 3 užd. Seka

Geriausi rasti parametrai

Geriausi gauti klasifikavimo tikslumo įverčiai su šiais parametrais:

Paketo dydis: 10

Paslėptų neuronų skaičius: 3

Mokymo greitis: 0,01

Impulsas (momentum): 0,3

Klasifikavimo tikslumo įverčiai su šiais parametrais atvaziduoti 4 pav.

```
Correctly Classified Instances      115          95.8333 %
Incorrectly Classified Instances     5           4.1667 %
Kappa statistic                    0.9375
Mean absolute error                 0.2272
Root mean squared error             0.2797
Relative absolute error             51.1264 %
Root relative squared error         59.3391 %
Total Number of Instances          120
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-setosa
	0,875	0,000	1,000	0,875	0,933	0,907	0,996	0,992	Iris-versicolor
	1,000	0,063	0,889	1,000	0,941	0,913	0,996	0,992	Iris-virginica
Weighted Avg.	0,958	0,021	0,963	0,958	0,958	0,940	0,997	0,995	

=== Confusion Matrix ===

```
a b c  <-- classified as
40 0 0 | a = Iris-setosa
0 35 5 | b = Iris-versicolor
0 0 40 | c = Iris-virginica
```

4. pav. klasifikavimo tikslumo įverčiai

Kiti parametrai

Paketo dydis: 10

Paslėptų neuronų skaičius: 5

Mokymo greitis: 0,01

Impulsas (momentum): 0,3

Klasifikavimo tikslumo įverčiai su šiais parametrais atvaziduoti 5 pav.

```

Correctly Classified Instances      114          95    %
Incorrectly Classified Instances     6           5    %
Kappa statistic                    0.925
Mean absolute error                 0.2226
Root mean squared error             0.2768
Relative absolute error             50.0807 %
Root relative squared error         58.7082 %
Total Number of Instances          120

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-setosa
	0,850	0,000	1,000	0,850	0,919	0,889	0,996	0,992	Iris-versicolor
	1,000	0,075	0,870	1,000	0,930	0,897	0,996	0,993	Iris-virginica
Weighted Avg.	0,950	0,025	0,957	0,950	0,950	0,929	0,997	0,995	

=== Confusion Matrix ===

```

a  b  c  <-- classified as
40  0  0 | a = Iris-setosa
 0 34  6 | b = Iris-versicolor
 0  0 40 | c = Iris-virginica

```

5. pav. klasifikavimo tikslumo įverčiai

Paketo dydis: 10

Paslėptų neuronų skaičius: 5

Mokymo greitis: 0,1

Impulsas (momentum): 0,3

Klasifikavimo tikslumo įverčiai su šiais parametrais atvaziduoti 6 pav.

```

Correctly Classified Instances      115          95.8333 %
Incorrectly Classified Instances     5           4.1667 %
Kappa statistic                    0.9375
Mean absolute error                 0.0539
Root mean squared error             0.1396
Relative absolute error             12.1332 %
Root relative squared error         29.6087 %
Total Number of Instances          120

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-setosa
	0,975	0,050	0,907	0,975	0,940	0,909	0,994	0,990	Iris-versicolor
	0,900	0,013	0,973	0,900	0,935	0,906	0,994	0,989	Iris-virginica
Weighted Avg.	0,958	0,021	0,960	0,958	0,958	0,938	0,996	0,993	

=== Confusion Matrix ===

```

a  b  c  <-- classified as
40  0  0 | a = Iris-setosa
 0 39  1 | b = Iris-versicolor
 0  4 36 | c = Iris-virginica

```

6. pav. klasifikavimo tikslumo įverčiai

Paketo dydis: 10

Paslėptų neuronų skaičius: 6

Mokymo greitis: 0,1

Impulsas (momentum): 0,1

Klasifikavimo tikslumo įverčiai su šiais parametrais atvaziduoti 7 pav.

```

Correctly Classified Instances      115          95.8333 %
Incorrectly Classified Instances    5           4.1667 %
Kappa statistic                    0.9375
Mean absolute error                 0.0495
Root mean squared error            0.1373
Relative absolute error             11.1287 %
Root relative squared error        29.1241 %
Total Number of Instances         120

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-setosa
	0,975	0,050	0,907	0,975	0,940	0,909	0,995	0,990	Iris-versicolor
	0,900	0,013	0,973	0,900	0,935	0,906	0,995	0,990	Iris-virginica
Weighted Avg.	0,958	0,021	0,960	0,958	0,958	0,938	0,996	0,993	

=== Confusion Matrix ===

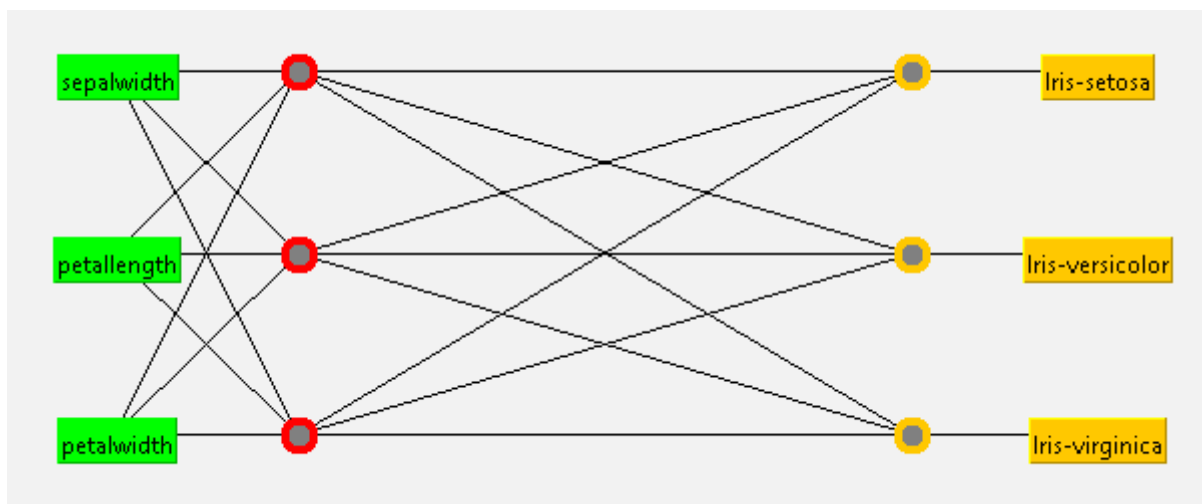
```

a b c <-- classified as
40 0 0 | a = Iris-setosa
0 39 1 | b = Iris-versicolor
0 4 36 | c = Iris-virginica

```

7. pav. klasifikavimo tikslumo įverčiai

Neuroninio tinklo vaizdas



8. pav. geriausio neuroninio tinklo vaizdas

Naujų duomenų klasifikavimo rezultatai

9 pav. pateikti naujų duomenų klasifikavimo rezultatai. Pirmi trys skaičiai yra požymių vertės, po to nurodoma tikroji klasė, likę trys skaičiai nurodo klasių tikimybes – pirmasis iš trijų paskutinių skaičių nurodo tikimybę, kad klasė yra Iris-setosa, antroji – Iris-versicolor, trečioji – Iris-virginica. Nesunku pastebėti, kad šiuo atveju tikslumas yra 100%, tai patvirtina 10 pav. matomos klasifikavimo tikslumo metrikos.

```

3.5,1.3,0.3,Iris-setosa,0.802459,0.196177,0.001365
2.3,1.3,0.3,Iris-setosa,0.774628,0.223155,0.002217
3.2,1.3,0.2,Iris-setosa,0.800542,0.198058,0.001399
3.5,1.6,0.6,Iris-setosa,0.791042,0.207149,0.001809
3.8,1.9,0.4,Iris-setosa,0.799135,0.199335,0.00153
3,1.4,0.3,Iris-setosa,0.793241,0.205131,0.001628
3.8,1.6,0.2,Iris-setosa,0.805214,0.193491,0.001295
3.2,1.4,0.2,Iris-setosa,0.799584,0.198986,0.001431
3.7,1.5,0.2,Iris-setosa,0.804863,0.193839,0.001298
3.3,1.4,0.2,Iris-setosa,0.800997,0.19761,0.001393
2.6,4.4,1.2,Iris-versicolor,0.092737,0.616888,0.290375
3,4.6,1.4,Iris-versicolor,0.071574,0.559906,0.36852
2.6,4,1.2,Iris-versicolor,0.139198,0.643356,0.217446
2.3,3.3,1,Iris-versicolor,0.289135,0.621124,0.089741
2.7,4.2,1.3,Iris-versicolor,0.099597,0.614632,0.285771
3,4.2,1.2,Iris-versicolor,0.185249,0.626686,0.188065
2.9,4.2,1.3,Iris-versicolor,0.126952,0.619335,0.253713
2.9,4.3,1.3,Iris-versicolor,0.114068,0.612315,0.273618
2.5,3,1.1,Iris-versicolor,0.367808,0.568118,0.064074
2.8,4.1,1.3,Iris-versicolor,0.124761,0.624548,0.250691
3.1,5.6,2.4,Iris-virginica,0.008887,0.342061,0.649052
3.1,5.1,2.3,Iris-virginica,0.010992,0.355864,0.633144
2.7,5.1,1.9,Iris-virginica,0.014574,0.391327,0.594099
3.2,5.9,2.3,Iris-virginica,0.008965,0.341783,0.649252
3.3,5.7,2.5,Iris-virginica,0.008553,0.337369,0.654078
3.5,2,2.3,Iris-virginica,0.010341,0.353573,0.636086
2.5,5,1.9,Iris-virginica,0.014185,0.394474,0.591341
3.5,2,2,Iris-virginica,0.013877,0.3776,0.608523
3.4,5.4,2.3,Iris-virginica,0.010884,0.349791,0.639325
3.5,1,1.8,Iris-virginica,0.019395,0.409205,0.5714

```

9. pav. klasifikavimo rezultatai

Correctly Classified Instances	30	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.2154		
Root mean squared error	0.2593		
Total Number of Instances	30		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-setosa
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-versicolor
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Iris-virginica
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

=== Confusion Matrix ===

```

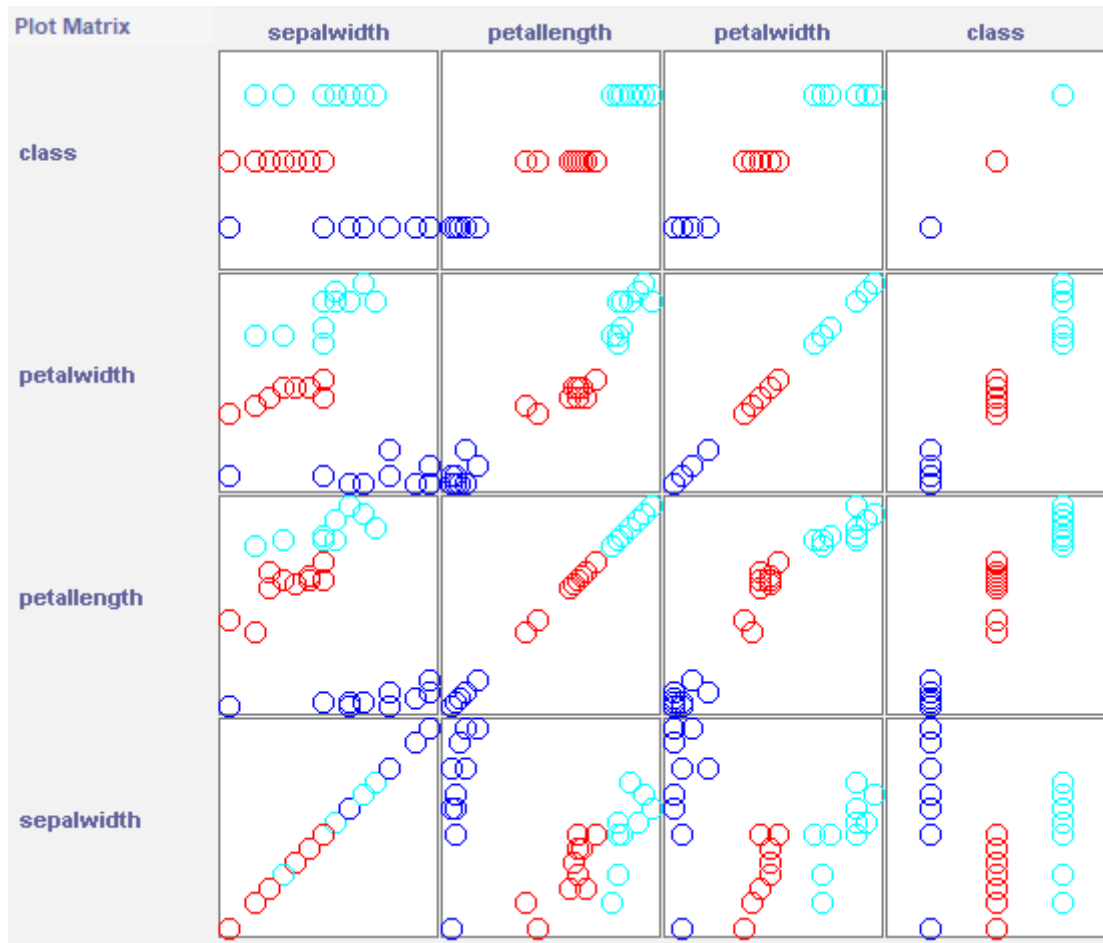
a b c <-- classified as
10 0 0 | a = Iris-setosa
0 10 0 | b = Iris-versicolor
0 0 10 | c = Iris-virginica

```

10 pav. klasifikavimo tikslumo metrikos

Poųymų porų vaizdai

11. pav. atvaizduoti poųymų porų vaizdai. Iris-setosa klasė atvaizduota tamsiai mėlyna spalva, Iris-versicolor raudona, Iris-virginica žydra.



11. pav. duomenų poųymų porų vaizdai Dekarto koordinatų sistemoje

Svoriai

1 lentelė. Paslėpto sluoksnio svoriai

Neuronas	Threshold	sepalwidth	petallength	petalwidth
Neuronas 3	-1,53487919	-0,096630716	2,155648789	2,671404487
Neuronas 4	1,628363279	0,245883283	-2,332492794	-2,733616209
Neuronas 5	-1,75621884	1,984576971	-2,497266836	-2,33357315
Neuronas 6	0,077236908	-0,128078653	-0,344881663	-0,429020623
Neuronas 7	0,505346593	-1,120856851	1,657331573	1,695276651

2 lentelė. Klasifikavimo sluoksnio svoriai

Neuronas	Threshold	w1	w2	w3	w4	w5
Neuronas 0	-1,152031473	-2,114139097	1,257703396	3,199969197	-0,299802894	-2,893621459
Neuronas 1	-0,449553784	-2,375773506	2,028520179	-3,622433001	0,041872097	0,738868935
Neuronas 2	-0,812623948	2,484520237	-3,509211474	-3,322503677	-0,816273548	2,12432979

1, 2 lentelėse aprašyti paslėpto ir klasifikavimo sluoksnio svoriai. Klasifikavimo sluoksnyje esantys svoriai atitinka neuronus taip: w1 – 3 neuronas, w2 – 4 neuronas, w3 – 5 neuronas, w4 – 6 neuronas, w5 – 7 neuronas.

MS Excel ir WEKA rezultatai

MS Excel programoje konstruotas neuroninis tinklas

MS Excel programoje naudota tas pats duomenų rinkinys kaip ir programoje WEKA. Suskaičiuojamos ir išrašomos normalizuotos požymių vertės bei WEKA programoje gauti svoriai. Kiekvienas požymis (įvestis) dauginamas su svoriu, šių sandaugų ir poslinkio suma užrašoma ir tai kartojama visiems duomenų įrašams kiekvienam neuronui. Užrašoma vertė paduodama sigmoidinei aktyvacijos funkcijai. Kadangi turime penkis neuronus, gauname po penkis skaičius kiekvienam duomenų įrašui, kuris kiekvienas dauginamas su atitinkamu klasifikavimo sluoksnio svoriu, o jų ir poslinkio suma užrašoma. Kadangi turime tris klases, šis procesas kartojamas tris kartus, kad gautume kiekvienos klasės vertes, kurioms pritaikius sigmoidinę aktyvacijos funkciją virsta klasių tikimybėmis.

Gauti rezultatai ir išvada

3 lentelė. MS Excel programa gauti rezultatai

Iris-setosa	Iris-versicolor	Iris-virginica
0,9501084	0,125980279	0,000315857
0,8394694	0,301511485	0,001020598
0,9462156	0,132519455	0,000332205
0,941756	0,137151616	0,000382056
0,9495848	0,125635031	0,000334314
0,9350548	0,150962763	0,00040144
0,9528664	0,121880431	0,000303949
0,9452215	0,134096282	0,000339337
0,9523446	0,122681625	0,000304602
0,9474423	0,130408097	0,000327397
0,0388775	0,723927494	0,176584741
0,0293731	0,584697071	0,295376333
0,0570447	0,763707655	0,108224148
0,1020244	0,828498591	0,04466754
0,0420514	0,710830138	0,168629544
0,0805712	0,698810279	0,086512939
0,0538102	0,691647951	0,137309047
0,0480111	0,680609929	0,15813366
0,1475847	0,795598553	0,031007081
0,0526309	0,713285976	0,133831152
0,0025402	0,135027151	0,956986522
0,0032189	0,161288062	0,937772803
0,0046424	0,24226935	0,881172763
0,0025508	0,133671631	0,957335736
0,0024629	0,127677221	0,960442969
0,0029997	0,156608344	0,943022962
0,0046262	0,252487369	0,87739352
0,0042223	0,20611505	0,903592182
0,0031706	0,146860381	0,943245736
0,0063904	0,279381857	0,825895472

4 lentelė. WEKA programa gauti rezultatai

Iris-setosa	Iris-versicolor	Iris-virginica
0,868171	0,131476	0,000353
0,74205	0,257059	0,000891
0,860597	0,13903	0,000373
0,841307	0,158199	0,000494
0,863977	0,135632	0,000391
0,836823	0,162707	0,00047
0,875813	0,123861	0,000326
0,85811	0,141506	0,000384
0,87458	0,125092	0,000328
0,862854	0,136779	0,000367
0,063504	0,816875	0,119621
0,048281	0,739417	0,212302
0,086071	0,833392	0,080538
0,146377	0,820665	0,032957
0,065901	0,811565	0,122534
0,112012	0,815716	0,072272
0,080382	0,812159	0,107458
0,073709	0,806283	0,120008
0,187164	0,787301	0,025534
0,078848	0,81775	0,103402
0,002785	0,144366	0,852849
0,003748	0,180851	0,8154
0,006443	0,294815	0,698741
0,002826	0,144521	0,852653
0,002603	0,13318	0,864217
0,003486	0,174611	0,821902
0,006428	0,305307	0,688265
0,005547	0,248195	0,746259
0,003581	0,164751	0,831668
0,009337	0,350985	0,639678

3 ir 4 lentelėse matomi gauti klasifikacijos rezultatai su MS Excel ir WEKA programomis. Žalias langelis nurodo tikrąją klasę, paryškinta tikimybės skaitinė vertė nurodo prognozuotą klasę. Matoma, kad tiek su MS Excel programa, tiek su WEKA programa gaunami puikūs (100% tikslumo) rezultatai. Jie gana panašūs, kadangi naudojami tie patys svoriai, tačiau nėra visiškai vienodi.