



Vilniaus universitetas
Matematikos ir informatikos fakultetas
Informatikos katedra

Dirbtiniai neuroniniai tinklai sistemoje WEKA

prof. dr. Olga Kurasova
Olga.Kurasova@mii.vu.lt

2018

Tiesioginio sklidimo DNT sistemoje WEKA

- Šiuo metu yra **sukurta programinių sistemų**, skirtų tik neuroniniams tinklams.
- Yra duomenų tyrybos sistemų, kuriose be kitų metodų yra **įgyvendinti ir neuroniniai tinklai**.
- **WEKA** – tai viena populiariausių duomenų tyrybos sistemų (www.cs.waikato.ac.nz/ml/weka/).
- Tai **atviro kodo nemokama sistema**, turinti plačias duomenų tyrybos funkcijas.

Sistema WEKA (1)



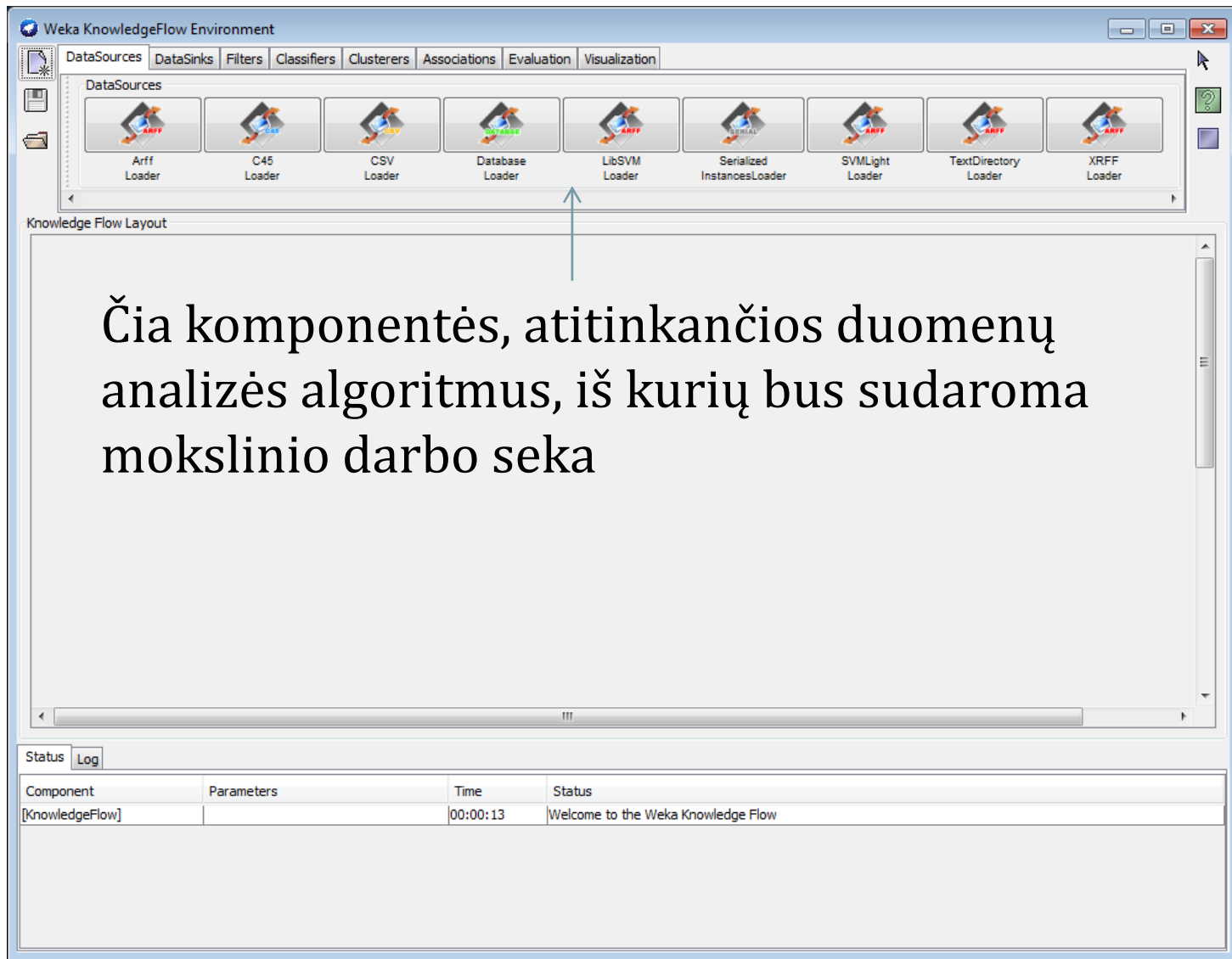
- Sistemoje duomenų tyrybos algoritmus galima taikyti **standartiniu būdu** arba sukūrus vadinamąsias **mokslinio darbo sekas**.
- Mokslinio darbo seka – tai sujungtų **komponenčių junginys**, kur kiekviena komponentė atitinka tam tikrą duomenų įkėlimo, apdorojimo ar tyrybos algoritmą, taip pat rezultatų peržiūrą.
- Tai patogus įrankis, kai norima tokiu pat būdu analizuoti **kitus duomenis**, ar tirti juos, **pakeitus** algoritmų **valdymo parametrų reikšmes**.

Sistema WEKA (2)

Norint pradėti **kurti mokslinio darbo seką**, reikia paspausti mygtuką ***KnowledgeFlow***.



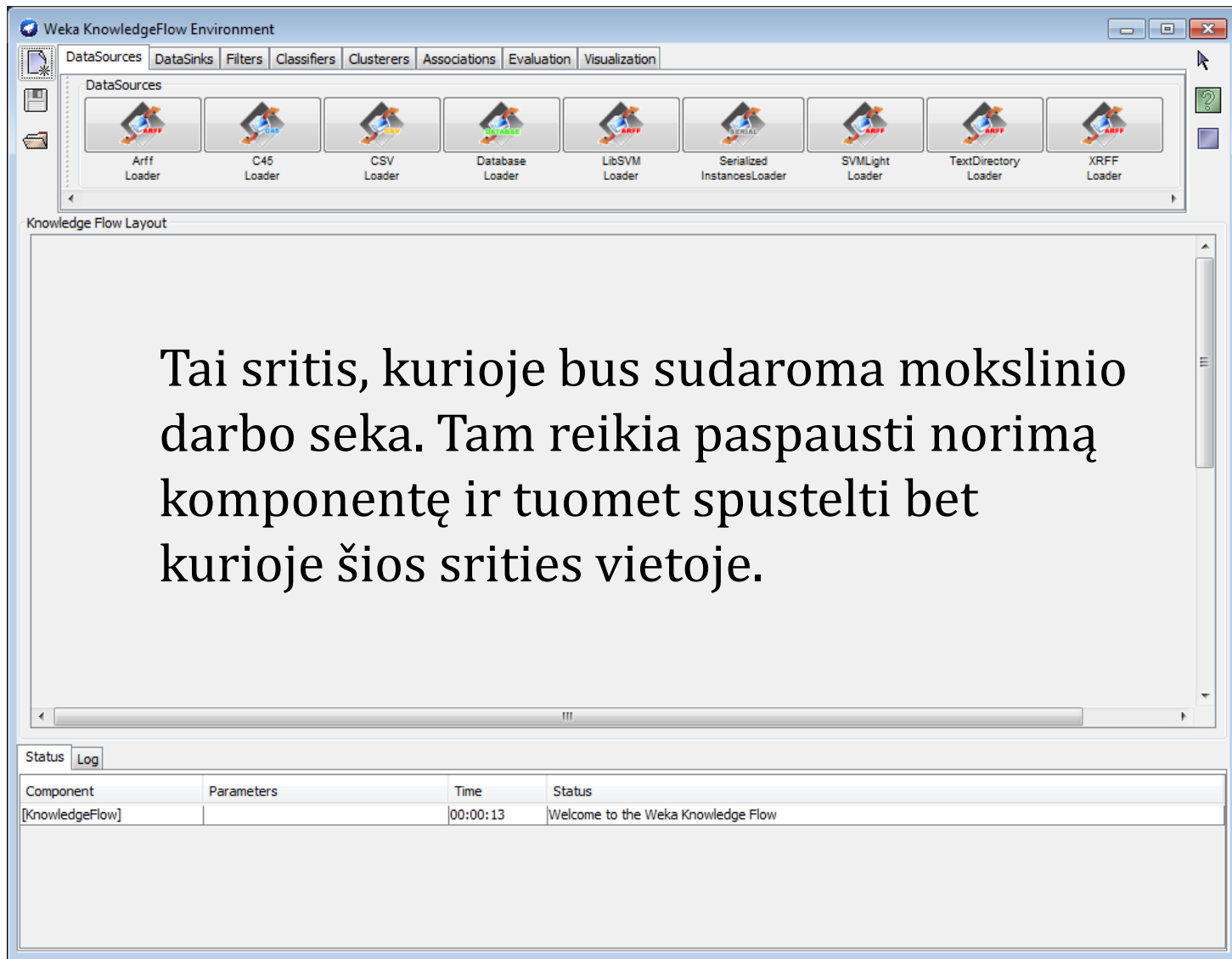
Sistema WEKA (3)



Čia komponentės, atitinkančios duomenų analizės algoritmus, iš kurių bus sudaroma mokslinio darbo seka

Component	Parameters	Time	Status
[KnowledgeFlow]		00:00:13	Welcome to the Weka Knowledge Flow

Sistema WEKA (4)



Tai sritis, kurioje bus sudaroma mokslinio darbo seka. Tam reikia paspausti norimą komponentę ir tuomet spustelti bet kurioje šios srities vietoje.

Component	Parameters	Time	Status
[KnowledgeFlow]		00:00:13	Welcome to the Weka Knowledge Flow

Duomenų failų formatai sistemoje WEKA

- Sistemoje WEKA galima **įkelti kelių formatų** duomenis, tačiau rekomenduojama naudoti pačios sistemos **arff formatą**.
- Įdiegiant sistemą, į kompiuterį **įrašomi kelių duomenų failų**, pavyzdžiai. Pvz., C:\Program Files\Weka-3-6\data
- Failus arff formatu galima atverti ir redaguoti **paprastu tekstiniu redaktoriumi**.

Duomenys arff formatu

```
@RELATION iris
```

```
@ATTRIBUTE sepallength REAL
```

```
@ATTRIBUTE sepalwidth REAL
```

```
@ATTRIBUTE petallength REAL
```

```
@ATTRIBUTE petalwidth REAL
```

```
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

```
@DATA
```

```
5.1,3.5,1.4,0.2,Iris-setosa
```

```
4.9,3.0,1.4,0.2,Iris-setosa
```

```
4.7,3.2,1.3,0.2,Iris-setosa
```

```
4.6,3.1,1.5,0.2,Iris-setosa
```

```
5.0,3.6,1.4,0.2,Iris-setosa
```

```
5.4,3.9,1.7,0.4,Iris-setosa
```

```
4.6,3.4,1.4,0.3,Iris-setosa
```

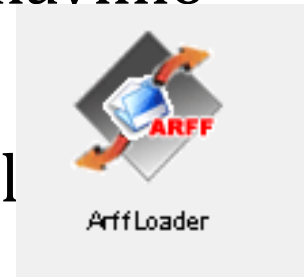
```
5.0,3.4,1.5,0.2,Iris-setosa
```

```
4.4,2.9,1.4,0.2,Iris-setosa
```

```
4.9,3.1,1.5,0.1,Iris-setosa
```

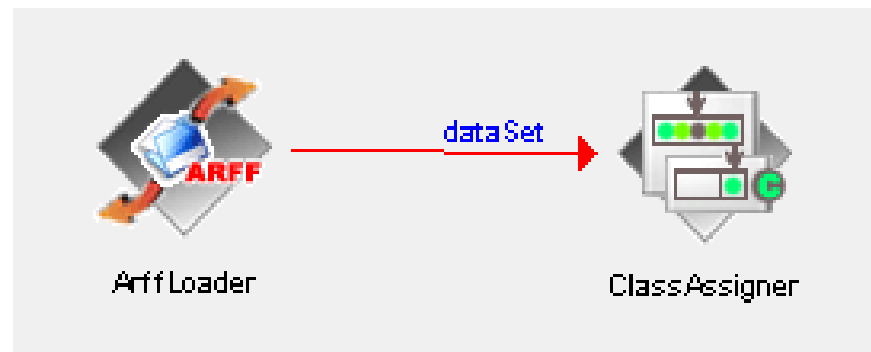

Duomenų įkėlimas sistemoje WEKA

- Norint įkelti duomenis arff formatu, reikia iš kortelės **DataSources** rinktis komponentę **Arff Loader** (nepamiršti spustelti sekos formavimo srityje).
- Du kartus ją spustelėti ir atsiradusiame l **nurodyti norimą failą**.
- Be to, jei bus sprendžiamas klasifikavimo uždavinys, reikia nurodyti, kuris duomenų požymis yra klasė. Tam skirta **Evaluation** kortelėje esanti komponentė **Class Assigner**. Ją du kartus paspaudus, reikia nurodyti tą požymį.



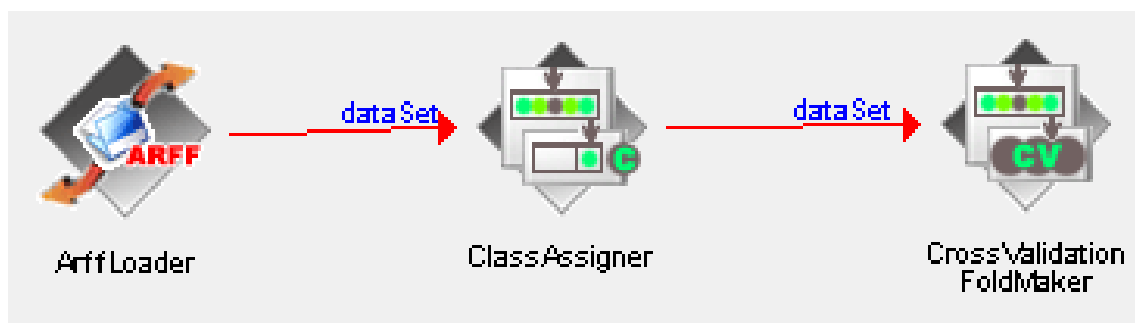
Duomenų įkėlimas sistemoje WEKA

- Dabar belieka **sujungti** šias dvi komponentes.
- Tai atliekama **tokiu būdu**:
 - ant pirmosios komponentės paspaudžiamas **dešinysis pelės mygtukas**,
 - išsirenkama **dataSet**
 - ir **paspaudžiama** ant kitos komponentės.



Kryžminė patikra

- WEKA sistemoje kryžminei patikrai atlikti skirta kortelėje **Evaluation** esanti komponentė **CrossValidation FoldMaker**.
- Du kartus ją spustelėjus, atsiradusiame lange galima nurodyti norimą **blokų skaičių** (*folds*), pagal kurį **bus suskaidoma** duomenų aibė į **mokymo** ir **testavimo** duomenis.



Daugiasluoksnio perceptrono komponentė

- **Daugiasluoksnio perceptrono** komponentės, skirta klasifikavimo uždaviniams spręsti, yra kortelėje **Classifiers** ir vadinasi **Multilayer Perceptron**.
- Ši komponentė turi būti sujungta su komponente **CrossValidation FoldMaker**.
- Įkėlus šią komponentę, ją du kartus spustelėjus pelyte, atsiranda **parametrų nustatymo langas**.

MLP pagrindiniai parametrai

- **GUI: True** – bus rodoma tinklo struktūra.
- **hiddenLayers** – kableliais atskirti neuronų skaičiai paslėptuose sluoksniuose.
- **learningRate** – mokymo greičio parametras η .
- **Momentum** – tinklo mokymo taisyklėje naudojama reikšmė.
- **Seed** – atsitiktinių skaičių generavimo „sėkla“.

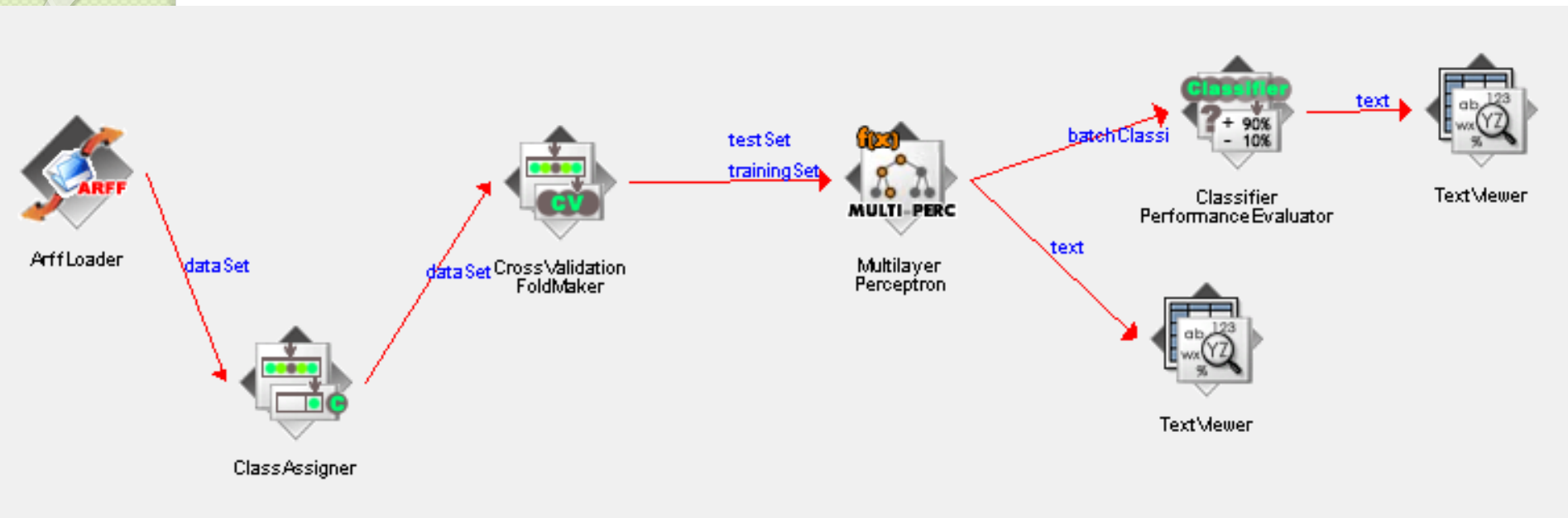
Klasifikavimo tikslumo nustatymas

- Norint pamatyti klasifikavimo rezultatų tikslumą, reikia prie komponentės **Multilayer Perceptron** prijungti kortelėje **Evaluation** esančią komponentę **Classifier PerformanceEvaluator**.
- O prie šios komponentės prijungti kortelėje **Visualization** esančią komponentę **TextViewer**.
- Tuomet reikia ant komponentės paspausti dešinę pelės mygtuką ir pasirinkti **Show Results**.

Perceptrono svoriai

- Norint pamatyti, kokie gauti išmokyto perceptrono svoriai, reikia prie komponentės **Multilayer Perceptron** prijungti komponentę **TextViewer**.
- Tuomet reikia ant komponentės paspausti dešinę pelės mygtuką ir pasirinkti **Show Results**.

Gauta mokslinio darbo seka



Norint **įvykdyti** sukurta seka reikia ant **ArffLoader** komponentės paspaudus dešinią pelės mygtuką, pasirinkti **Start Loading**.

Klasifikavimo rezultatai

=== Evaluation result ===

Scheme: MultilayerPerceptron

Options: -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H "2, 3" -G -R

Relation: iris

Correctly Classified Instances	148	98.6667 %
Incorrectly Classified Instances	2	1.3333 %
Kappa statistic	0.98	
Mean absolute error	0.0273	
Root mean squared error	0.0894	
Relative absolute error	6.1343 %	
Root relative squared error	18.9712 %	
Total Number of Instances	150	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	Iris-setosa
	0.98	0.01	0.98	0.98	0.98	0.999	Iris-versicolor
	0.98	0.01	0.98	0.98	0.98	0.999	Iris-virginica
Weighted Avg.	0.987	0.007	0.987	0.987	0.987	0.999	

=== Confusion Matrix ===

a	b	c	<-- classified as
50	0	0	a = Iris-setosa
0	49	1	b = Iris-versicolor
0	1	49	c = Iris-virginica

Gauti perceptrono svorjai

=== Classifier model ===

Scheme: MultilayerPerceptron

Relation: iris

Sigmoid Node 0

Inputs	Weights
Threshold	4.86297276124753
Node 5	-3.8660908414801503
Node 6	-4.705134898782587
Node 7	-9.11213313626472

Sigmoid Node 1

Inputs	Weights
Threshold	-4.667974263875816
Node 5	-7.907836952620307
Node 6	-4.191198923452797
Node 7	9.397064982396241

Sigmoid Node 2

Inputs	Weights
Threshold	-7.139047124741141
Node 5	6.407314719993383
Node 6	4.953708835603616
Node 7	2.148788678109934

Sigmoid Node 3

Inputs	Weights
Threshold	5.97182308582669
Attrib sepallength	1.2928052580000988
Attrib sepalwidth	2.6852736928332734
Attrib petallength	-7.953115255023941
Attrib petalwidth	-9.316709649525436

Sigmoid Node 4

Inputs	Weights
Threshold	0.7370366149205487

Sukurto klasifikatoriaus išsaugojimas

- Neuroninis tinklas mokomas tam, kad vėliau galėtų **nežinomų klasių duomenis priskirti** vienai iš klasių.
- Tam reikia **išsaugoti** sukurto klasifikatoriaus rezultatus, DNT atveju – **svorių reikšmes**.
- Paspaudus ant **Multilayer Perceptron** komponentės, reikia pasirinkti **Save Model** ir nurodyti saugojimo vietą bei failo vardą.

Duomenų išskaidymas į mokymo ir testavimo

- Mokant neuroninį tinklą, galima naudoti ne tik kryžminės patikros būdą, bet patiems pradžioje **išskaidyti** į mokymo ir testavimo aibes.
- Tam sistemoje WEKA yra įgyvendinti keli būdai:
 - Naudoti vieną duomenų failą, tačiau komponentės **TrainTest SplitMaker** pagalba duomenis išskaidyti į mokymo ir testavimo.
 - Naudoti du duomenų failus, vieną komponentės **Training SetMaker** pagalba priskirti mokymo duomenis, kitą komponentės **TestSet Maker** pagalba – testavimo.

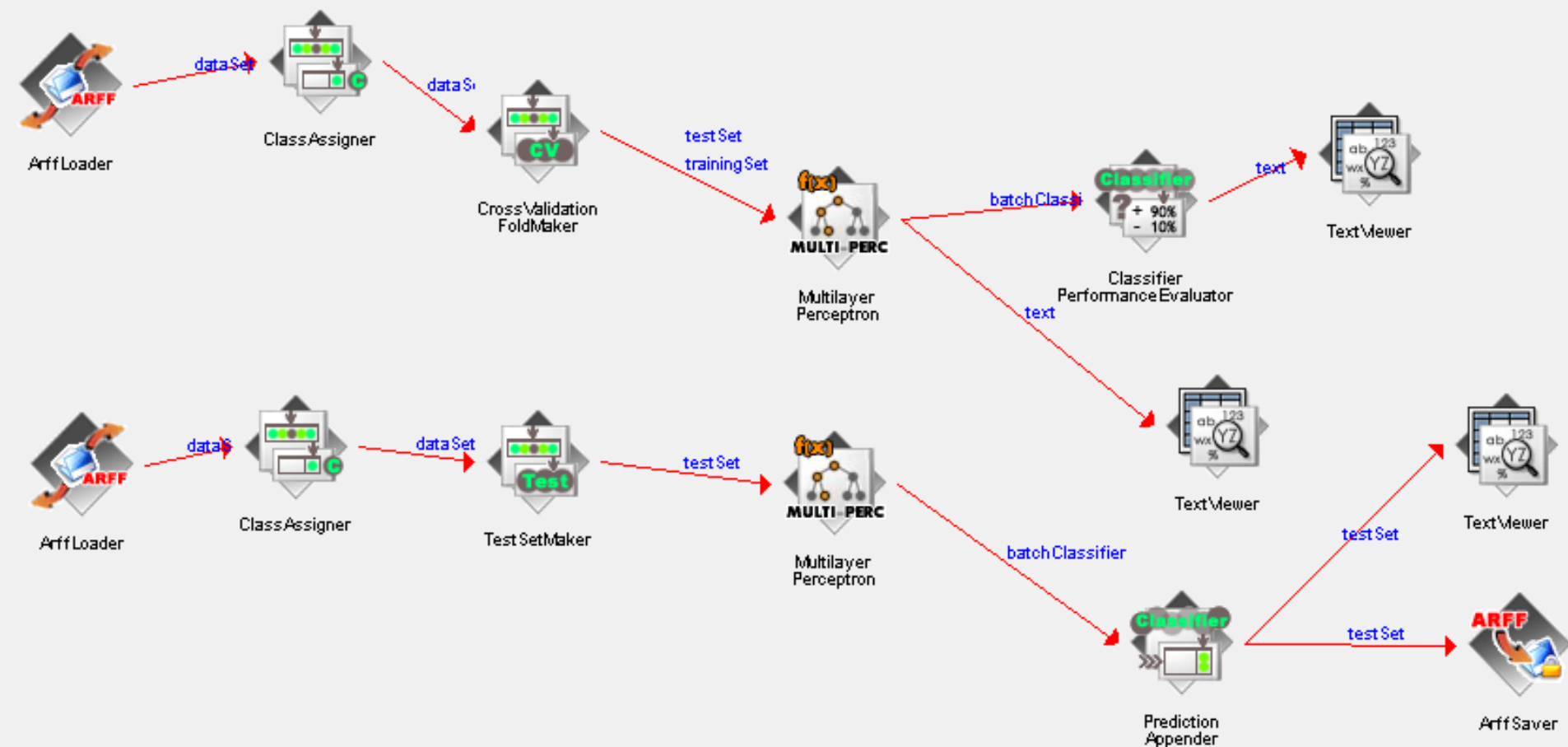
Naujų duomenų priskyrimas klasėms (1)

- Neuroninis tinklas mokomas tam, kad vėliau galėtų **nežinomų klasių duomenis priskirti** vienai iš klasių.
- Jei duomenų **klasė nėra žinoma**, tai arff faile prie vietoj klasių **įrašomas klaustukas**.
- Tuomet reikia įkelti dar vieną **ArffLoader** komponentę, nurodyti tą arff failą, kuriame yra **nauji duomenys** (su nežinomomis klasėmis).
- Taip pat įkelti ir sujungti komponentes **ClassAssigner, TestSet Maker, Multilayer Perceptron**.

Naujų duomenų priskyrimas klasėms (2)

- Paspaudus ant komponentės **Multilayer Perceptron** dešinę pelės mygtuką, pasirinkti **Load Model** ir nurodyti anksčiau išsaugotą klasifikatoriaus rezultatus (modelį).
- Tuomet prie šios komponentės reikia prijungti **Evaluation** kortelėje esančią komponentę **Prediction Appender**, o prie jos – **Text Viewer**.
- Paspaudus du kartus **Prediction Appender** galima nurodyti, ar bus rodomos klasių priskyrimo tikimybės (TRUE) ar ne (FALSE).
- Paspaudus dešinę pelės mygtuką ant komponentės **Text Viewer** ir pasirinkus **Show results**, matomi naujų duomenų klasifikavimo rezultatai.
- Prijungus kortelėje **DataSinks** esančią komponentę **Arff Saver**, klasifikavimo rezultatai bus **išsaugoti** nurodytame faile.

Gautos mokslinio darbo sekos



Duomenų vizualizavimas

- Sistemoje WEKA yra **duomenų vizualizavimo komponentės**, leidžiančios pamatyti analizuojamų duomenų struktūrą.
- Kortelėje **Visualization** pasirinkus komponentę **Scatter PlotMatrix** ir ją prijungus prie komponentės **Class Assigner**, galima pamatyti duomenų išsidėstymą Dekarto koordinatų sistemoje (imant požymių poras).