



VšĮ „Informatikos mokslų centras“

DUOMENŲ ANALIZĖS ĮRANKIO

**DAMIS**

INSTRUKCIJA NAUDOTOJUI

(atnaujinta 2015-04-15)

Vilnius, 2015

## Turinys

Įvadas.....	4
1. DAMIS ir lygiagrečiųjų bei paskirstytųjų skaičiavimų resursų sąveika ir galimybės .....	5
2. Testinės duomenų aibės.....	6
3. Duomenų analizės įrankio naudotojo instrukcija .....	8
3.1 Duomenų analizės įrankio naudojimas.....	8
3.2 Langas „Prisijungimas“ .....	8
3.3 Langas „Pamiršote slaptažodį“ .....	9
3.4 Langas „Registracija“ .....	9
3.5 Pagrindinis puslapio meniu „Pagrindinis puslapis“ .....	10
3.5.1 Meniu skiltis „Profilis“ .....	11
3.5.2 Meniu skiltis „Nustatymai“ .....	11
3.5.3 Meniu skiltis „Pagalba“ .....	11
3.5.4 Kalbų pasirinkimo skiltis.....	12
3.6 Pagrindinio puslapio meniu skiltis „Kurti eksperimentą“ .....	12
3.6.1 Kuriamo eksperimento vykdymo seka .....	13
3.6.2 Kairiojo meniu skiltis „Duomenų įkėlimas“ .....	13
3.6.3 Kairiojo meniu skiltis „Pirminis apdorojimas“ .....	20
3.6.4 Kairiojo meniu skiltis „Statistiniai primityvai“ .....	44
3.6.5 Kairiojo meniu skiltis „Dimensijos mažinimas“ .....	48
3.6.6 Kairiojo meniu skiltis „Klasifikavimas ir grupavimas“ .....	76
3.6.7 Kairiojo meniu skiltis „Rezultatų peržiūra“ .....	88
3.7 Pagrindinio puslapio meniu skiltis „Eksperimentai“ .....	94
3.8 Pagrindinis puslapio meniu „Failų sąrašas“ .....	95
3.8.1 Skiltis „Įkelti failą iš kompiuterio“ .....	96
3.8.2 Veiksmo „Redaguoti failą“ langas.....	97
3.9 Naujo algoritmo įkėlimas .....	97
3.10 Eksperimento vykdymas .....	98
3.10.1 Mygtukas „Naujas eksperimentas“ .....	98
3.10.2 Mygtukas „Išsaugoti“ .....	98
3.10.3 Mygtukas „Vykdyti“ .....	99
3.10.4 Atskiros komponentės vykdymas .....	100

3.11	Pavyzdiniai eksperimentų scenarijai .....	100
3.11.1	Stuburo duomenų statistinė analizė .....	101
3.11.2	Stuburo duomenų normavimas .....	104
3.11.3	Stuburo duomenų grupavimosi tendencijos.....	107
3.11.4	Stuburo duomenų klasifikavimas.....	110
3.12	Langai „Pagalba“ ir „D.U.K.“ .....	114

## Įvadas

Šis dokumentas – tai duomenų analizės įrankio (toliau DAMIS) instrukcija naudotojui. Įrankis sukurtas pagal 2013 m. vasario 12 d. duomenų analizės metodų algoritmizavimo ir pilotinio įrankio programinės įrangos sukūrimo paslaugų sutartį Nr. APS-580000-243 ir atnaujintas pagal 2015 m. kovo 12 d. pilotinio duomenų analizės įrankio (DAMIS) programinės įrangos atnaujinimo paslaugų sutartį Nr. APS-13300-447 tarp VŠĮ „Informatikos mokslų centras“ ir Vilniaus universiteto. Paslaugos vykdytos Vilniaus universitetui įgyvendinant Europos regioninės plėtros fondo remiamą projektą „Nacionalinės atviros prieigos mokslo informacijos duomenų archyvo“ pagal „Informacinė visuomenė visiems“ prioriteto įgyvendinimo priemonę Nr. VP2-3.1-IVPK-13-V „Mokslo duomenų archyvas“.

DAMIS (duomenų analizės įrankis) – tai atvira mokslo infrastruktūra, skirta duomenų analizei atlikti. Įrankio paskirtis – sudaryti galimybę specializuotoje aplinkoje bendradarbiaujantiems mokslininkams ar jų grupėms atlikti pagrindinius duomenų analizės tyrimus (grupavimą, klasifikavimą ir kt.) skaičiavimo poreikius atitinkančioje aplinkoje; vizualios analizės priemonėmis tirti daugiamačių duomenų projekcijas į plokštumą, duomenų grupavimąsi, duomenų panašumus, atskirų daugiamačių duomenų požymių įtaką ir tarpusavio priklausomybes; stebėti bei apdoroti vizualizacijos ar našųjų skaičiavimų aplinkoje gautus tyrimų rezultatus.

DAMIS ir jo dokumentacijos autoriai – VŠĮ „Informatikos mokslų centras“ kolektyvas.

Veikiantis DAMIS prototipas patalpintas adresu <http://damis.lt>. Įrankio programinis kodas prieinamas adresu <https://github.com/InScience/DAMIS>, kodo dokumentacija pateikta adresu [https://github.com/InScience/DAMIS\\_docs](https://github.com/InScience/DAMIS_docs).

## 1. DAMIS ir lygiagrečių bei paskirstytųjų skaičiavimų resursų sąveika ir galimybės

Šiame skyriuje pateiktas programinės sąsajos aprašas, kuriame paaiškinama duomenų analizės įrankio ir lygiagrečių bei paskirstytųjų skaičiavimų resursų web servisų pagrindu.

DAMIS aplinkoje yra galimybė pasirinkti vieną iš dviejų lygiagrečių bei paskirstytųjų skaičiavimų resursų:

- VU Matematikos ir informatikos instituto (MII) kompiuterių klasterį,
- VU Matematikos ir informatikos fakulteto (MIF) superkompiuterį.

VU Matematikos ir informatikos instituto klasteryje (*hpc.mii.vu.lt*) yra 32 skaičiavimo mazgai po 4 Intel x86 architektūros branduolius:

- 16 su Intel I5-760 2,8 GHz 44,8 GFLOP procesoriumi po 4 branduolius;
- 14 su Intel Quad Core Q9400 2,66 GHz 42,656 GFLOP procesoriumi po 4 branduolius;
- 2 su Intel Quad Core Q6600 2,4 GHz 38,4 GFLOP procesoriumi po 4 branduolius;
- kiekvienas skaičiavimo mazgas turi po 4 GB operatyviosios atminties, 320 GB diskinės atminties.

Viso 128 Intel x86 architektūros branduoliai, maksimalus teorinis našumas apie 1,4 TFLOP, 0,5 TB operatyviosios atminties.

VU informacinių technologijų atviros prieigos centras (ITAPC) buvo sukurtas vykdant 2010 m. rugšėjo 30 d. VU Senato nutarimą Nr. S-2010-06-30 ir ES struktūrinės paramos projektą VP2-1.1-ŠMM-04-V-01-002 „Informacinių technologijų atviros prieigos centro sukūrimas“. ITAPC veiklos uždaviniai yra pateikti naudojimui kompiuterinius išteklius, vystyti grid ir debesijos skaičiavimų technologijas, kurti patrauklų skaičiavimų ir paslaugų centrą. ITAPC naudojama techninė įranga yra įsigyta Švietimo ir mokslo ministerijos administruojamo „Santaros“ slėnio projekto, finansuoto Europos Sąjungos, lėšomis.

VU MIF superkompiuteris – vienas pajėgiausių superkompiuterių Lietuvoje, turintis:

- 1920 Intel Xeon branduolius: 224 procesoriai (po 6 branduolius) X5650 2,66 GHz 63,984 GFLOP, 48 procesoriai (po 8 branduolius) X7550 2 GHz 64 GFLOP, 48 procesoriai (po 4 branduolius) E5520 2,26 GHz 36,256 GFLOP;
- 3,6 TB operatyviosios atminties;
- 620 TB duomenų saugyklą;
- maksimalus teorinis našumas apie 19 TFLOP.

Skaičiavimų resursų pasirinkimui duomenų analizės įrankyje yra sukurtos dviejų tipų komponentės, kurios žymimos VU MII ar VU MIF logotipais. Pasirinkus vieno ar kito tipo komponentę, jos skaičiavimai bus vykdomi pasirinktame lygiagrečių ir paskirstytųjų skaičiavimų resurse. Šių resursų integravimas į DAMIS web servisų pagalba suteikia galimybę tyrėjams greičiau atlikti daug skaičiavimų reikalaujančią duomenų analizę.

Naudotojas, vykdydamas duomenų analizę, gali pasirinkti skaičiavimams naudojamų procesorių skaičių. Jei nurodomas daugiau nei vienas procesorius, skaičiavimai vykdomi lygiagrečiai. Sistema automatiškai parenka vykdymui lygiagrečiąją duomenų analizės algoritmų versiją. Naudotojui nebūtina turėti specifinių žinių apie algoritmų lygiagretinimą bei užduočių paldidimą lygiagrečių ir paskirstytųjų skaičiavimų telkiniuose.

## 2. Testinės duomenų aibės

Duomenų analizės įrankiui išbandyti galima naudoti kelias įrankyje pateiktas testines duomenų aibes. Analizuojamų duomenų objektų, juos charakterizuojančių požymių bei klasių skaičius pateiktas 1 lentelėje. Pirmosios šešios duomenų aibės paimtos iš duomenų archyvo „UCI Machine Learning Repository“ (<http://archive.ics.uci.edu/ml/>), septintoji duomenų aibė – tai dirbtinai sugeneruota duomenų aibė. Šios duomenų aibės dažnai yra naudojamos duomenų analizės algoritmams testuoti, kadangi žinomos duomenų savybės.

1 lentelė. Testavimui naudojamos duomenų aibės

Nr.	Duomenų aibė	Objektų skaičius	Požymių skaičius	Klasių skaičius
1.	Irisai	150	4	3
2.	Vynai	178	13	3
3.	Krūties vėžys	683	9	2
4.	Abalone	4177	7	15
5.	Magic	19020	10	2
6.	Diabetas	768	8	2
7.	Elipsoidai	1115	50	20
8.	Stuburo ligų duomenų aibė	310	6	2

*Fišerio irisų* duomenys (iris.arff) – tai klasikiniai testiniai duomenys, naudojami daugiamačių duomenų analizėje, dažnai vadinami tiesiog irisais arba irisų duomenimis. Yra išmatuoti trijų veislių gėlių (Iris Setosa (I klasė), Iris Versicolor (II klasė) ir Iris Virginica (III klasė)) šie požymiai:

- vainiklapių pločiai (angl. *petal weight*),
- vainiklapių ilgiai (angl. *petal height*),
- taurėlapių pločiai (angl. *sepal weight*),
- taurėlapių ilgiai (angl. *sepal height*).

Iš viso matuota 150 gėlių žiedų. Sudaryti 4-mačiai taškai. Įvairiais duomenų analizės metodais yra nustatyta, kad I klasės irisai „atsiskiria“ nuo kitų dviejų klasių (II ir III). II ir III klasės dalinai persidengia.

*Vynų* duomenų aibė (wine.arff) sudaryta iš tame pačiame Italijos regione gaminamų trijų skirtingų vynų rūšių cheminės analizės rezultatų. Visos trys vynų rūšys vertinamos pagal 13 skirtingų cheminės sudėties kriterijų, todėl sudaryti 13-mačiai taškai. I klasei priklauso 59 objektai, II klasei – 71, III klasei – 48. Iš viso vertinti 178 objektai.

*Krūties vėžio* duomenų aibė (breast\_cancer.arff) sudaryta fiksuojant 683 susirgimo atvejus: nepiktybinio naviko 444 atvejus (I klasė) ir piktybinio – 239 atvejus (II klasė). Vertinti 9 požymiai, parametrų reikšmės kategorinės, kiekvieno parametro reikšmė yra tarp 1 ir 10.

*Abalone* duomenų aibės (*abalone.arff*) kiekvienas objektas yra charakterizuojamas 7 moliuskų požymiais:

- ilgis (ilgiausia kiauto dalis),
- skersmuo (statmenas ilgiui),
- kiauto aukštis,
- moliusko svoris kartu su kiautu,
- moliusko svoris be kiauto,
- vidaus organų svoris,
- kiauto svoris be moliusko.

Moliusko žiedų skaičius nusako klasę, viso yra iki 29 žiedų. Kadangi moliuskų skaičius tam tikrose klasėse nedidelis, tai iš 29 klasių sudaryta 15 klasių, apjungiant 1–4, 18–20, 21–29 klases.

*Magic* teleskopo duomenų aibę (*magic.arff*) sudaro 19020 įrašų: I klasė - gama signalo (g) 12332 įrašai, II klasė - hadron arba fono tipo (h) 6688 įrašai. Vertinta 10 požymių.

*Diabeto* duomenų aibė (*diabetes.arff*) sudaryta iš 768 objektų. Čia objektai yra pacientai tirti siekiant nustatyti diabetą. Visos pacientės (moterys) virš 21 metų amžiaus iš Phoenix, Arizonos regiono, JAV. Vertinti 8 požymiai. 500 tiriamųjų diabetas nėra nustatytas, 268 tiriamųjų diabeto testas buvo teigiamas.

*Elipsoidų* duomenų aibę (*ellipsoid.arff*) sudaro 1115 taškų, kurių dimensija lygi 50. Duomenų aibės taškai suformuoja 20 persidengiančių elipsoidinio tipo klasterių. Duomenų aibė sugeneruota naudojant elipsoidinių klasterių generatorių kuris aprašytas darbe: „Handl, J.; Knowles, J. 2005. Cluster generators for large high-dimensional data sets with large numbers of clusters“. Šis generatorius sukuria elipsoidinius klasterius. Klasterių ribos apibrėžiamos keturiais požymiais:

- centras,
- tarpžidininis atstumas, kurio reikšmės tolygiai pasiskirstę intervale [1,0; 3,0],
- pagrindinės ašies kryptis tolygiai keičiama generuojant kiekvieną atskirą klasterį,
- maksimali atstumų nuo sugeneruoto taško iki dviejų židinių sumos reikšmė, priklausanti intervalui [1,05; 1,15].

Generuojami taškai kiekvienam klasteriui atskirai, tikrinama ar neperžengtos elipsoidui apibrėžtos ribos ir netinkami taškai atmetami.

***Stuburo ligų duomenų aibė*** (angl. Vertebral Column Database), failas *stuburo\_su\_klasem.txt*. Duomenų rinkinį galima klasifikuoti į 3 klases – sveiki, stuburo disko išvarža, spondilolistezė (angl. normal, disk hernia, spondilolysthesis) – arba į 2 klases – sveiki, sergantys (angl. normal, abnormal). Visą duomenų rinkinį sudaro 310 pacientų. Kiekvieną pacientą apibūdina šeši biomechaniniai požymiai: dubens dažnis (angl. pelvic incidence), dubens tentas (angl. pelvic tilt), juosmens kampas (angl. lumbar lordosis angle), sakraliniai nuolydžiai (angl. sacral slope), dubens spindulys (angl. pelvic radius) ir spondilolistezės klasė (angl. the grade of spondylolisthesis). Eilutės atitinka vertintą paciento atvejį, stulpeliai matuotų požymių reikšmės, paskutinis stulpelis klasės numeris.

Duomenų aibė paimta iš duomenų archyvo „UCI Machine Learning Repository” (<http://archive.ics.uci.edu/ml/datasets/Vertebral+Column>), galima atsisiųsti originalią duomenų aibę. Rocha Neto, A., R. Sousa, G. Barreto, and J. Cardoso (2011). Diagnostic of pathology on the vertebral column with embedded reject option. In Pattern Recognition and Image Analysis, pp. 588–595. Springer.

### 3. Duomenų analizės įrankio naudotojo instrukcija

#### 3.1 Duomenų analizės įrankio naudojimas

**Paskirtis:** atlikti duomenų analizę naudojantis duomenų analizės įrankiu (DAMIS).

**Pradžia:** procesas pradedamas, kai registruotas paslaugų gavėjas vidinėje MIDAS portalo dalyje pasirenka meniu punktą „Duomenų analizės įrankis“.

**Rezultatas:** iš paslaugų gavėjo asmeninės arba tyrimo erdvės į duomenų analizės įrankį perduotas duomenų failas, įvykdyta analizė ir gautas analizės rezultatas išsaugotas asmeninėje erdvėje, tyrimo erdvėje arba laikinoje saugykloje.

Veikiantis duomenų analizės įrankio grafinės naudotojo sąsajos prototipas patalpintas adresu <https://damis.midas.lt>. Duomenų analizės įrankio svetainės pradinis langas pateiktas 1 pav. Iš šio lango naudotojas jungiasi prie savo darbo aplinkos arba registruojasi kaip naujas naudotojas.



1 pav. Duomenų analizės įrankio svetainės pradinis langas

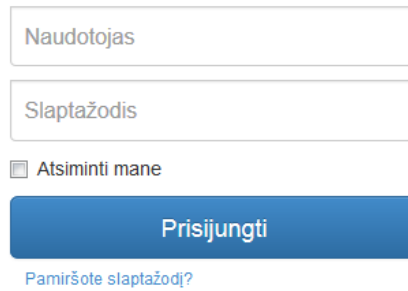
#### 3.2 Langas „Prisijungimas“

Paspaudus nuorodą [Prisijungti](#), atidaromas prisijungimo langas, pateiktas 2 pav. Šiame lange naudotojas įveda naudotojo vardą ir slaptažodį. Paspaudus prisijungimo mygtuką, vyksta naudotojo tikrinimas. Jei naudotojo registracija patvirtinta, ir jo įvesti prisijungimo duomenys teisingi, naudotojas prijungiamas prie sistemos ir jam suteikiamas vaidmuo pagal jo statusą. Jei toks naudotojas sistemoje nėra registruotas, sistema rodo pranešimą pradiniam prisijungimo lange, jog naudotojas nėra registruotas svetainėje. Jei naudotojas įvedė neteisingus prisijungimo duomenis, jam rodomas klaidos pranešimas apie neteisingai įvestus duomenis ir leidžia naudotojui dar kartą įvesti savo prisijungimo duomenis.

Jei į DAMIS patenkama iš MIDAS aplinkos, jokio papildomo autentifikavimo nereikia.



## Prisijungti



Naudotojas

Slaptažodis

Atsiminti mane

Prisijungti

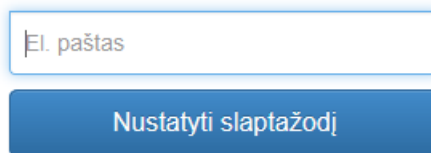
[Pamiršote slaptažodį?](#)

2 pav. Svetainės prisijungimo langas

### 3.3 Langas „Pamiršote slaptažodį“

Naudotojas pamiršęs savo slaptažodį, turi prisijungimo puslapyje paspausti nuorodą „Pamiršote slaptažodį?“ (2 pav.). Tuomet bus atidaromas slaptažodžio nustatymo langas (3 pav.), kuriame naudotojas turi įvesti savo el. pašto adresą ir jam nurodytu adresu išsiunčiamas laiškas su nuoroda, kurią paspaudus galima pakeisti savo slaptažodį.

## Nustatyti slaptažodį



El. paštas

Nustatyti slaptažodį

3 pav. Slaptažodžio nustatymo langas

### 3.4 Langas „Registracija“

Naudotojas norėdamas registruotis sistemoje turi paspausti meniu skiltį „Registruotis“. Tuomet atidaromas registracijos langas (4 pav.). Naudotojas turi įvesti reikiamus registracijai duomenis:

- Sugalvoti naudotojo vardą;
- Įvesti savo el. pašto adresą;
- Įvesti savo vardą;
- Įvesti savo pavardę;
- Įvesti savo organizaciją;
- Įvesti slaptažodį;
- Pakartoti slaptažodį.

Įvedęs reikiamus duomenis, naudotojas turi spausti mygtuką „Registruotis“. Tada bus

vykdomas įvestų duomenų tikrinimas. Jei bus neteisingai įvestas laukas ar jis bus tuščias, naudotojui bus rodomas klaidos pranešimas. Jei naudotojas visus laukus užpildė teisingai, jo registracijos duomenys bus išsiųsti. Naudotojui reiks sulaukti, kol administratorius patvirtins jo registraciją, apie tai jis bus informuotas gavęs laišką į nurodytą el. pašto adresą.

## Registruotis

Naudotojas
El. paštas
Vardas
Pavardė
Organizacija
Slaptažodis
Patvirtinti
<b>Registruotis</b>

4 pav. Registracijos langas

### 3.5 Pagrindinis puslapio meniu „Pagrindinis puslapis“

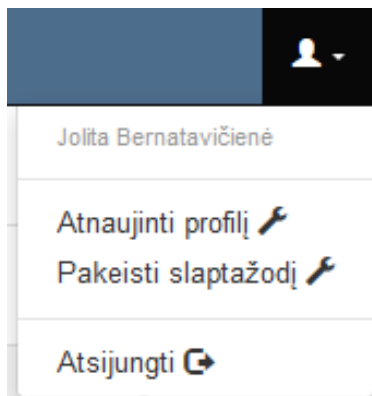
Šis langas apima svetainės meniu (5 pav.), kuris susideda iš:

- Kurti eksperimentą;
- Eksperimentai;
- Failų sąrašas;
- Profilis;
- Nustatymai (meniu skiltis matoma tik administratoriaus teises turinčiam naudotojui);
- Pagalba;
- Kalba.

### 5 pav. Pagrindinio puslapio administratoriaus meniu

#### 3.5.1 Meniu skiltis „Profilis“

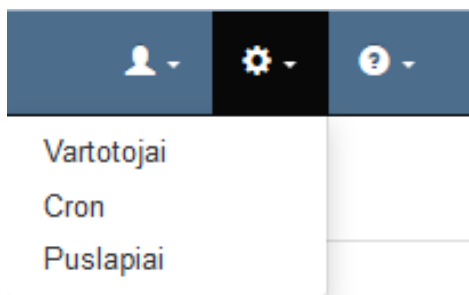
Naudotojas prisijungęs prie svetainės, mato savo profilį, gali jį keisti, jei to reikia. Taip pat gali pakeisti esamą slaptažodį bei atsijungti. Lango vaizdas pateiktas 6 pav.



6 pav. Profilio skiltis

#### 3.5.2 Meniu skiltis „Nustatymai“

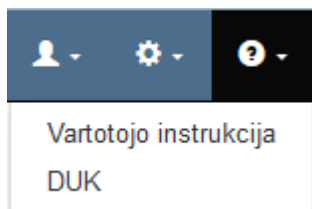
Administratoriaus pagrindinio puslapio meniu turi papildomą skiltį „Nustatymai“. Naudotojas turintis administratoriaus teises gali keisti registruotų naudotojų prieigas, t. y. aktyvuoti arba išjungti naudotojus. Taip pat gali keisti naudotojų asmeninius duomenis. Gali peržiūrėti „Cron“ ir keisti puslapių informaciją, pakeisti jų pavadinimus, pozicijas, kalbą, tekstą ir kt. Lango vaizdas pateiktas 7 pav.



7 pav. Nustatymų skiltis

#### 3.5.3 Meniu skiltis „Pagalba“

Naudotojas gali pasinaudoti pagalba ir peržiūrėti DUK. Lango vaizdas pateiktas 8 pav.



8 pav. Pagalbos skiltis

### 3.5.4 Kalbų pasirinkimo skiltis

Naudotojas gali pakeisti naudojimo kalbą, jei nustatyta naudojimo kalba lietuvių, tai galima pasirinkti anglų ir atvirkščiai ( 9 pav.)



9 pav. Kalbos nustatymas

### 3.6 Pagrindinio puslapio meniu skiltis „Kurti eksperimentą“

Paspaudus meniu skiltį „Kurti eksperimentą“ (5 pav.), naudotojas pirmiausiai mato eksperimento kūrimo darbalaukį.

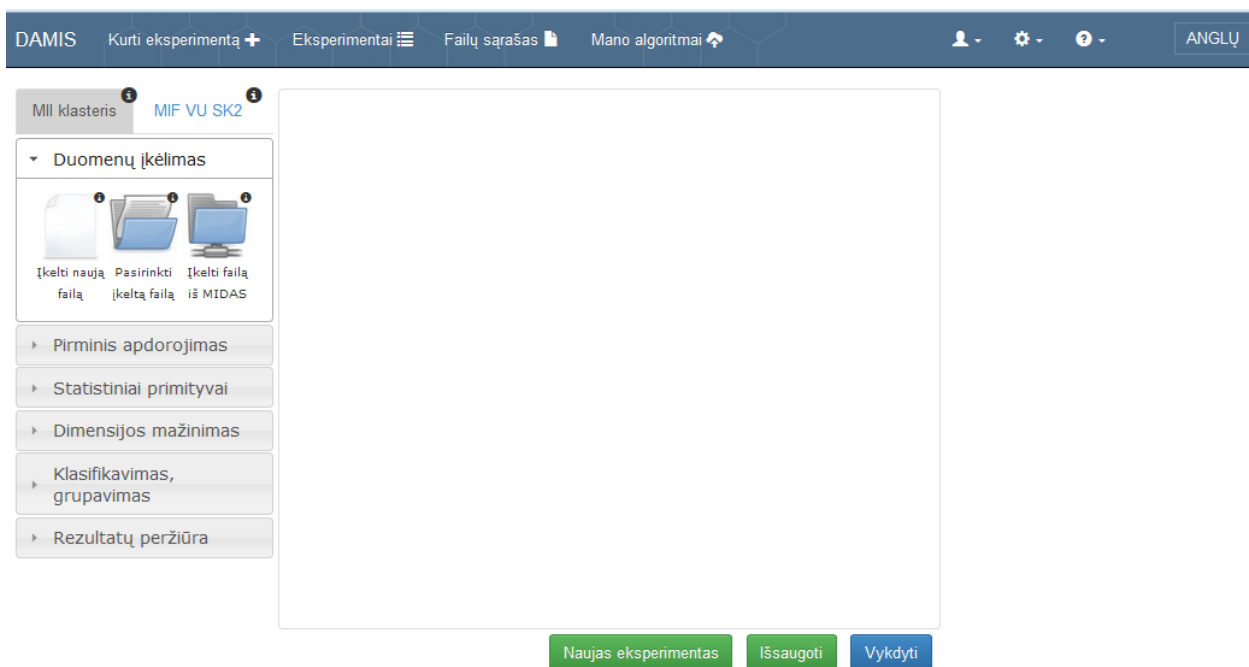
Eksperimentų kūrimo langas pavaizduotas 10 pav. Planuojant eksperimentą, naudotojui pirmiausia reikia peržiūrėti skaičiavimų išteklių apkrovą ir pasirinkti tą išteklių, kurio apkrova tuo metu yra mažesnė. Tam, kad peržiūrėti apkrovą, naudotojui reikia paspausti informacinį ženkluką prie skaičiavimo išteklių pavadinimo. Tam, kad naudotojas pasirinktų išteklių, reikia paspausti ant skilties „MII klasteris“ ar „MIF VU SK2“.

Pasirinkus skaičiavimų išteklių, naudotojas gali rinktis iš meniu laukų komponentes ir tempti jas į darbalaukį. Suplanavus eksperimento seką ar sekas, naudotojas turi spausti mygtuką „Vykdyti“, tuomet įvyks suplanuotų sekų patikrinimas. Jei nebus rastos klaidos, naudotojas gaus patvirtinimo pranešimą. Tuomet naudotojui reikės sulaukti apskaičiuotų eksperimento rezultatų.

Kai naudotojui reikia išsaugoti savo eksperimentą, jis turi spausti mygtuką „Išsaugoti“. Jei naudotojui reikia išvalyti darbalaukį nuo įkeltų komponentių, jis turi spausti mygtuką „Naujas eksperimentas“.

Realizuotos komponentės svetainėje surūšiuotos pagal naudojimo kategoriją:

- Duomenų įkėlimas;
- Pirminis apdorojimas;
- Statistiniai primityvai;
- Dimensijos mažinimas;
- Klasifikavimas, grupavimas;
- Rezultatų peržiūra.



10 pav. Eksperimentų kūrimo langas

### 3.6.1 Kuriamo eksperimento vykdymo seka

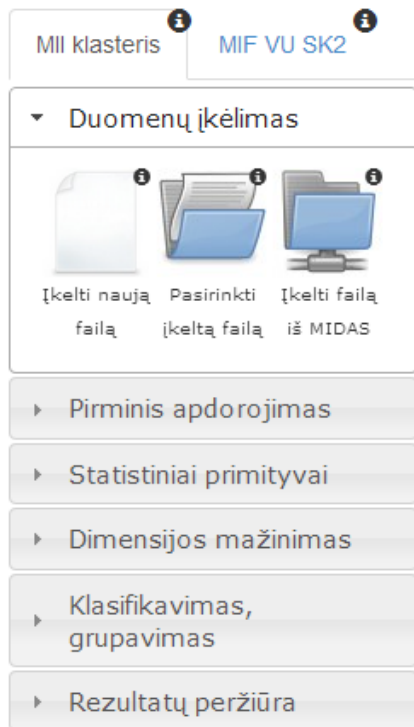
Duomenų analizės eksperimentai vykdomi pagal tokį scenarijų:

1. Analizuojami duomenys paruošiami *.arff* formatu.
2. Pasirenkamas skaičiavimų telkinys, kuriame bus atliekami eksperimentai: Vilniaus universiteto Matematikos ir informatikos instituto paskirstytųjų skaičiavimų klasteris arba Vilniaus universiteto Matematikos ir informatikos fakulteto paskirstytųjų skaičiavimų superkompiuteris.
3. DAMIS aplinkoje analizuojamas duomenų failas įkeliamas, naudojant duomenų įkėlimo komponentę.
4. Pasirenkamas duomenų analizės algoritmas.
5. Prijungiamos rezultatų peržiūros komponentės (techninė informacija, matricinis atvaizdavimas, grafinis vaizdavimas).
6. Vykdomas eksperimentas.

Analizuojamai duomenų aibei vykdomi duomenų analizės algoritmai, esant įvairioms jų parametrų reikšmėms, fiksuojamos paklaidos, skaičiavimo laikai, rezultatų failai. Sprendžiant dimensijos mažinimo uždavinį, kai projekcijos dimensija lygi 2, gauti dvimačiai taškai atvaizduojami Dekarto koordinatų sistemoje ir stebima, ar dimensijos mažinimas leidžia išlaikyti žinomas testinių duomenų savybes.

### 3.6.2 Kairiojo meniu skiltis „Duomenų įkėlimas“

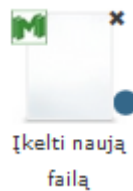
Meniu skilties „Duomenų įkėlimas“ laukas skirtas duomenų failo įkėlimui. Šioje skiltyje yra trys skirtingos komponentės skirtos failų įkėlimui (11 pav.). Pasirinkus komponentę „Įkelti naują failą“, naudotojas turės įkelti duomenų failą iš savo kompiuterio. Kai pasirenka komponentę „Pasirinkti įkeltą failą“, naudotojas turės pasirinkti failą iš failų sąrašo, kuriuos pats yra anksčiau įkėlęs.



11 pav. Meniu skiltis „Duomenų įkėlimas“

### 3.6.2.1 Komponentė „Įkelti naują failą“

Įkėlus į darbalaukį (eksperimentų kūrimo langą) ir paspaudus du kartus komponentę „Įkelti naują failą“



atidaromas langas (12 pav.).

12 pav. Langas „Įkelti naują failą“

Paspaudus mygtuką „Pasirinkti“ atidaromas katalogų langas, kuriame naudotojas turi pasirinkti failą, kurį nori įkelti ir spausiti mygtuką „Open“. Esant reikalui naudotojas gali pakeisti failo pavadinimą ir pateikti įkelto failo aprašymą. Paspaudus mygtuką „Patvirtinti“, vyksta failo patikrinimas, ar jis atitinka formato reikalavimus (yra vienas iš šių formatų: tab, txt, csv, xls, xlsx, zip, xml, arff arba archyvas zip). Jei jis tinkamo formato, tuomet langas uždaromas ir failas yra išsaugomas failų sąrašė. Jei tikrinimo metu buvo rastos klaidos, ties neteisingai užpildytais laukais rodomas klaidos pranešimas (13 pav.). Jei klaidų nėra, rodomas langas, pavaizduotas 14 pav. Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas. Papildomai atidarius komponentę „Įkelti naują failą“ naudotojui rodomi užpildyti laukai (jei naudotojas juos užpildė anksčiau).

13 pav. Langas, kai nenurodytos visų laukų reikšmės įkeliant naują failą

**14 pav. Langas, kai įkeliant failą nurodomos tinkamos laukų reikšmės**

Jei naudotojas nori pakeisti failą, kurį įkėlė, jam reikia vėl du kartus paspausti komponentę „Įkelti naują failą“. Tuomet atsidaręs langas atrodo taip, kaip pavaizduota 15 pav.

**15 pav. Įkelto failo keitimo langas**

Ties užrašu „Eksperimentui naudojamas failas“ rašomas anksčiau įkelto failo pavadinimas. Paspaudus pabraukto failo pavadinimą, atskirame lange parodomas failo turinys. Jei naudotojas nori įkelti failą iš naujo, jis turi spausti mygtuką „Naujas failas“. Paspaudęs mygtuką, naudotojui atsidarys naujo failo įkėlimo forma, viršuje lieka eksperimentui naudojamą failo pavadinimas, o žemiau reikia įkelti naują failą (16 pav.).



**Įkelti naują failą**

Eksperimentui naudojamas failas: iris

**Pavadinimas\***

**Failas\***

**Aprašymas**

\* Žvaigždute pažymėti laukai yra privalomi

16 pav. Įkelto failo keitimo kitu langas

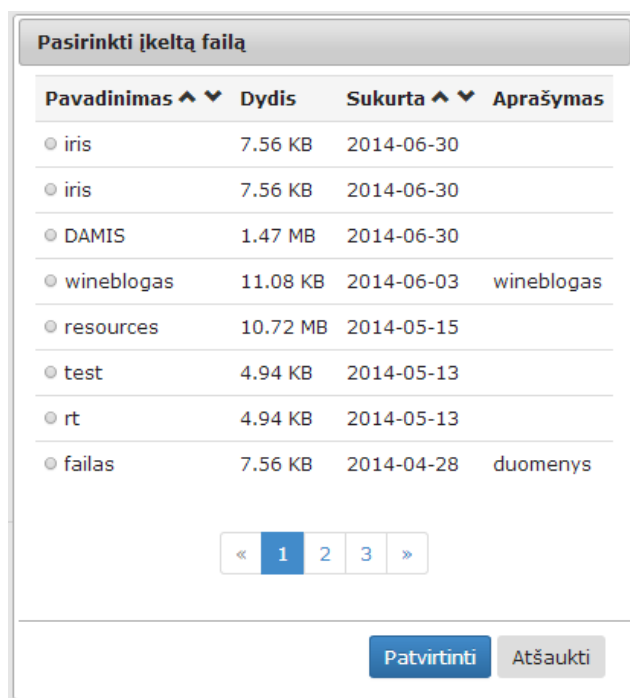
### 3.6.2.2 Komponentė „Pasirinkti įkeltą failą“

Komponentę „Pasirinkti įkeltą failą“



Pasirinkti  
įkeltą failą

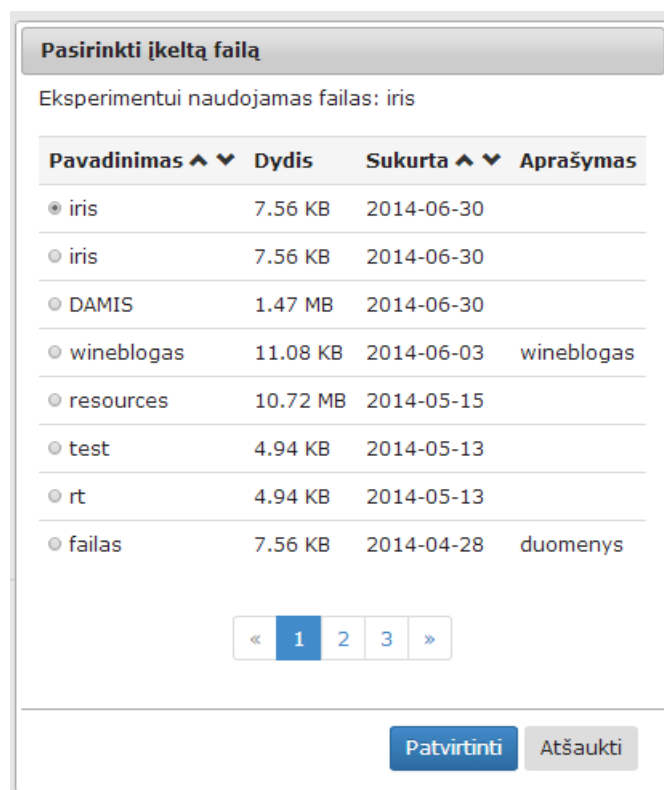
perkėlus į darbalaukį ir ją spragtelėjus du kartus atidaromas langas, kuriame matomas naudotojo anksčiau įkeltų failų sąrašas (17 pav.).



17 pav. Langas „Pasirinkti įkeltą failą“

Ilgam failų sąrašui yra užtikrintas puslapiavimas, viename puslapyje rodomi 10 įkeltų failų įrašų. Yra galimybė rūšiuoti failų sąrašą pagal pavadinimą ir sukūrimo datą. Paspaudus rodyklę į apačią ties užrašais „Pavadinimas“ arba „Sukurta“, įrašai surūšiuojami didėjimo tvarka. Paspaudus rodyklę į viršų, įrašai surūšiuojami mažėjimo tvarka. Paspaudus mygtuką „Patvirtinti“, pasirinktas failas bus naudojamas tolimesniuose skaičiavimuose. Paspaudus mygtuką „Atšaukti“, failo pasirinkimas nėra atliekamas. Grįžtama į eksperimento planavimo langą.

Kai naudotojas jau yra pasirinkęs failą, spragtelėjus du kartus komponentę „Pasirinkti įkeltą failą“, atsiranda langas (18 pav.), kuriame iš karto atvertas tas puslapis, kuriame yra pasirinktas naudotojo failas, jis yra pažymėtas. Ties užrašu „Pasirinktas failas:“ matosi pasirinkto failo pavadinimas, kurį spragtelėjus atskirame lange parodomas failo turinys. Naudotojas gali pasirinkti kitą failą. Tuomet pasikeičia pasirinkto failo pavadinimas ties įrašu „Pasirinktas failas“. Paspaudus mygtuką „Patvirtinti“, pasirinktas failas bus naudojamas tolimesniuose skaičiavimuose grįžtama į eksperimento planavimo langą. Jei paspaudžiamas mygtukas „Atšaukti“, langas uždaromas ir anksčiau pasirinktas failas lieka nepakeistas, grįžtama į eksperimento planavimo langą.



18 pav. Langas „Pasirinkti įkeltą failą“, kai jau failas yra įkeltas

### 3.6.2.3 Komponentė „Įkelti failą iš MIDAS“

Įkėlus į darbalaukį ir paspaudus du kartus komponentę „Įkelti failą iš MIDAS“



atidaromas langas, kuris pavaizduotas 19 pav. Reikia pasirinkti katalogą, kuriame saugomas norimas failas. Rodomas tik DAMIS tinkamų formatų failų sąrašas. Paspaudus mygtuką „Patvirtinti“ failas atsiunčiamas iš MIDAS į DAMIS.

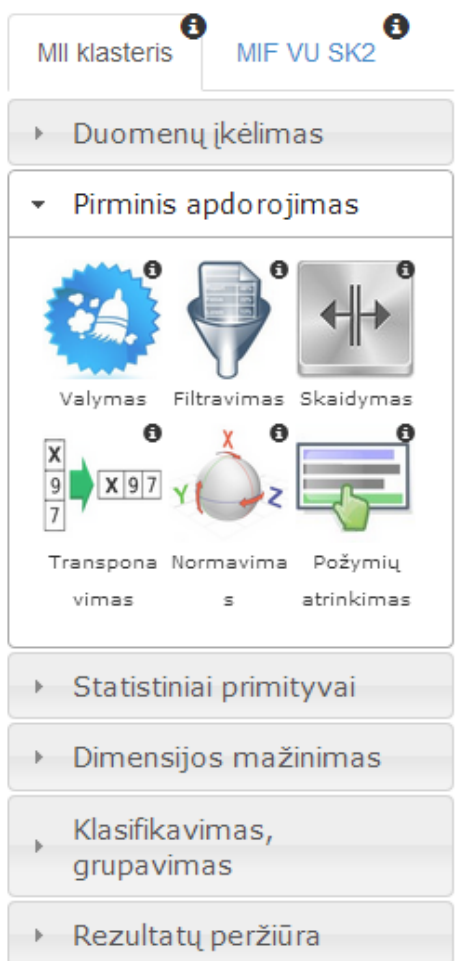


19 pav. Langas „Įkelti failą iš MIDAS“

### 3.6.3 Kairiojo meniu skiltis „Pirminis apdorojimas“

Meniu skilties „Pirminis apdorojimas“ hierarchija pavaizduota 61 pav. Šis meniu laukas skirtas pirminiam duomenų apdorojimui. Pasirinkus komponentę „Duomenų valymas“ – naudotojas pasirinks duomenų valymą. Pasirinkus „Filtravimas“ – naudotojas pasirinks duomenų filtravimo komponentę. „Skaidymas“ – tai skaidymo komponentė, dalinanti analizuojamą duomenų aibę į du atskirus poaibius. „Normavimas“ – tai duomenų normavimo komponentė, atliekanti duomenų normavimą pagal pasirinktą normavimo būdą. „Transponavimas“ – tai duomenų transponavimo komponentė. „Požymių atrinkimas“ – tai požymių atrinkimo komponentė, formuojanti naują duomenų aibę pagal pasirinktus požymius.

Nutempus bet kurią komponentę, naudotojas turi du kartus spragtelėti ją, tuomet atsidarys komponentės parametrų pildymo langas. Naudotojas turi nurodyti parametrų reikšmes ir spausti mygtuką „Patvirtinti“, tuomet bus tikrinami įvesti parametrai. Jei bus rasti neteisingai nurodyti parametrai, tuomet bus rodomas klaidos pranešimas prie neteisingai nurodytų parametrų. Jei klaidų nebus rasta, tuomet langas bus uždarytas, o parametrai išsaugoti. Pirminio apdorojimo meniu skiltis pavaizduota 20 pav.



20 pav. Meniu skiltis „Pirminis apdorojimas“

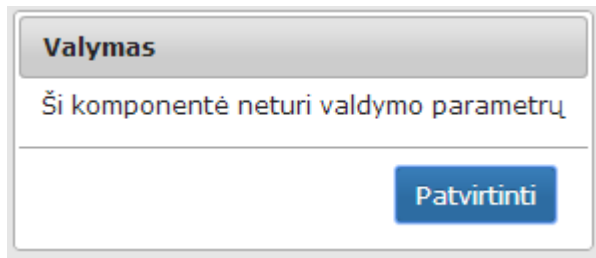
Toliau bus pateikiamas pirminio apdorojimo komponentių aprašymas.

### 3.6.3.1 Komponentė „Valymas“

Į darbalaukį įkėlus, sujungus ir paspaudus su kartus komponentę „Valymas“

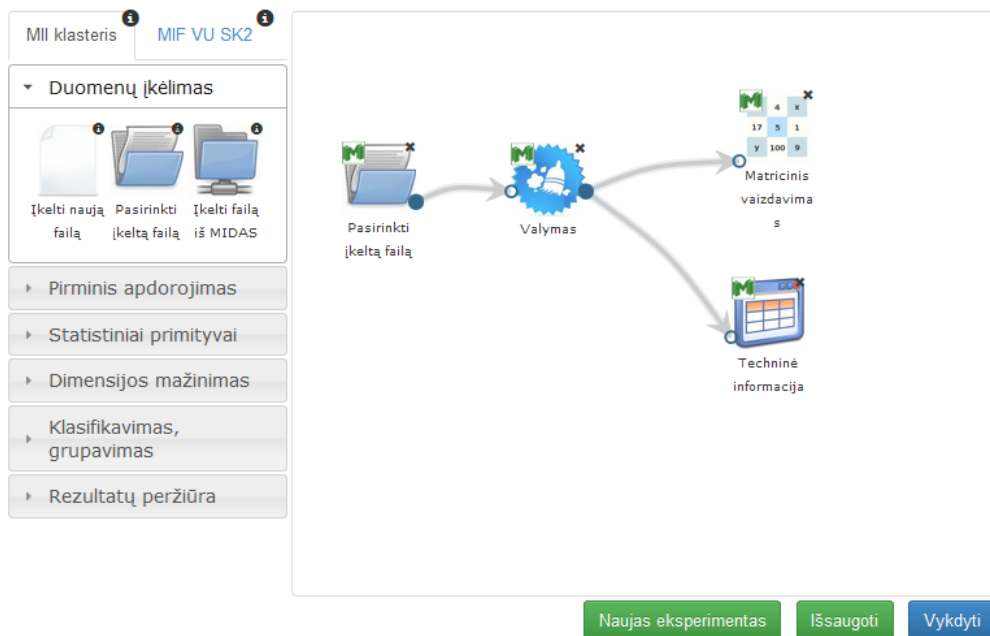


atsiranda langas. Kadangi komponentei nereikia įvesti valdymo parametru, atsidaręs langas atodo taip kaip pavaizduotas 21 pav. Jei naudotojas paspaudžia „Patvirtinti“ grįžtama į eksperimento planavimo langą.



21 pav. Langas „Valymas“

Duomenų valymas – tai veikla, kurios metu yra užtikrinama, kad duomenys turėtų vientisą struktūrą ir būtų tinkami tolimesniam apdorojimui ir analizei. Duomenų valymo eksperimentas vykdomas pagal darbų seką, pateiktą 22 pav.: įkeliami duomenų įkėlimo komponentė, pasirenkamas įkeliamas duomenų failas, įkeliami valymo komponentė ir sujungiami su duomenų įkėlimo komponente, taip pat sujungiami su įkelta su rezultatų peržiūros komponente ar keliomis komponentėmis.



22 pav. Duomenų valymo eksperimento darbų seka

Ši komponentė valdymo parametru neturi.

Atlikus eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 23 pav. ir

24 pav. testavimas MII klasteryje, 25 pav. ir 26 pav. – MIF VU superkompiuteryje.

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.01
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/24f7187e19450391e1e0c841c">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/24f7187e19450391e1e0c841c</a>

Atsisiųsti Atšaukti

23 pav. Komponentės „Valymas“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas				
Objekto nr.	SEPALLENGTH (REAL)	SEPALWIDTH (REAL)	PETALLENGTH (REAL)	PETALWIDTI (REAL)
1.	5.100000	3.500000	1.400000	0.200000
2.	4.900000	3.000000	1.400000	0.200000
3.	4.700000	3.200000	1.300000	0.200000
4.	4.600000	3.100000	1.500000	0.200000
5.	5.000000	3.600000	1.400000	0.200000
6.	5.400000	3.900000	1.700000	0.400000

Atsisiųsti Atšaukti

24 pav. Komponentės „Valymas“ testavimo rezultatas: matricinis vaizdavimas (MII klasteris)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.1
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/c59dcd0d952d1cabad0ba2e37e">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/c59dcd0d952d1cabad0ba2e37e</a>

25 pav. Komponentės „Valymas“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas				
Objekto nr.	ATTR0 (REAL)	ATTR1 (REAL)	ATTR2 (REAL)	ATTR3 (REAL)
1.	-0.29868	0.12644	-0.0775405	-0.252869
2.	0.137492	0.0835017	-0.0329761	-0.0337818
3.	0.070258	-0.0149113	0.0442346	-0.010622
4.	0.0590369	-0.225898	-0.0467255	-0.106744
5.	-0.000334845	-0.000214195	-0.065086	0.0441215
6.	0.317867	-0.108291	-0.0976692	-0.0716518
7.	0.172744	-0.0191883	0.0784447	-0.258299
8.	-0.187151	-0.141026	-0.00458187	0.025917

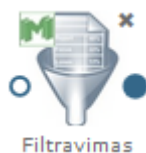
26 pav. Komponentės Valymas testavimo rezultatas: matricinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.3.2 Komponentė „Filtravimas“

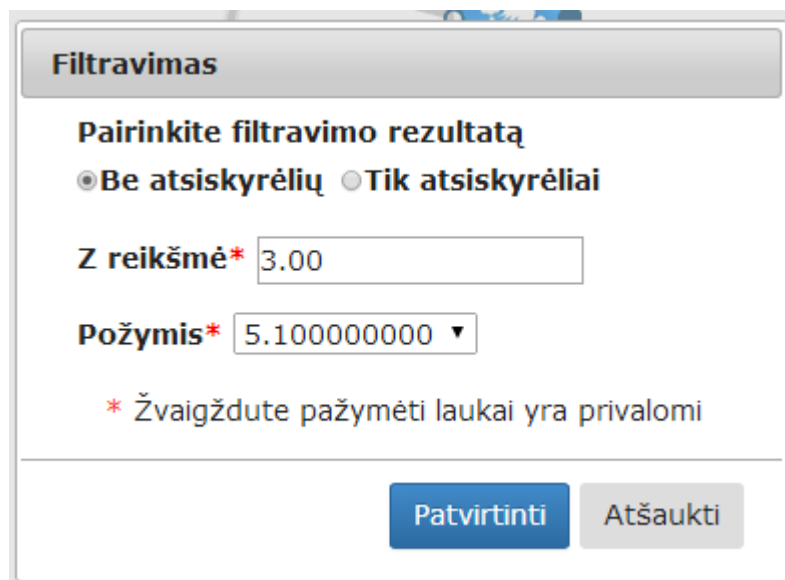
Filtravimas yra tam tikromis savybėmis pasižyminčių įrašų atmetimas iš nagrinėjamų įrašų aibės. Filtravimo rezultatu gali būti arba duomenų aibė be išsiskiriančių objektų, arba objektai atsiskyreliai. Vykdamas filtravimo komponentę galima pasirinkti pagal kokį vieną požymį bus atliekamas filtravimas ir kokia yra slenksčio reikšmė (kvantilis).

Komponentės „Filtravimas“ rezultatas: nauja duomenų matrica be išsiskiriančių objektų arba duomenų matrica, kurią sudaro tik objektai atsiskyrėliai (priklauso nuo pasirinktų filtravimo rezultatų).

Į darbalaukį įkėlus filtravimo komponentę,



ją sujungus su duomenų įkėlimo komponente ir paspaudus su kartus komponentę „Filtravimas“ parodomas jos langas. Naudotojas turi įvesti reikalingus valdymo parametus. Numatytosios reikšmės filtravimo rezultatas – „be atsiskyrėlių“, z reikšmė – 3, požymis – atr1 (sąrašas generuojamas iš arff failo, pasirenkamas pirmas požymis). Numatytosios reikšmės gali būti keičiamos, jei naudotojas pageidauja. Pakeistos reikšmės išsaugomos paspaudus mygtuką „Patvirtinti“. Langas pavaizduotas 27 pav.

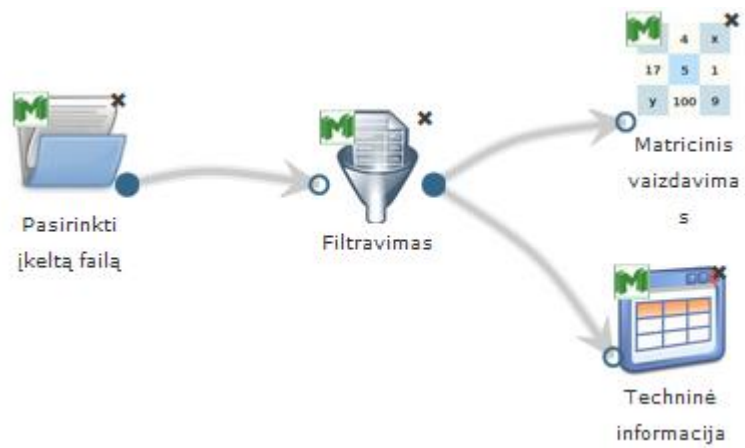


27 pav. Langas „Filtravimas“

Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinama Z reikšmė. Leistinos yra teigiamos parametro Z reikšmės (nebūtinai sveikieji skaičiai). Jei pateikta reikšmė yra negalima, klaidos pranešimas rodomas šalia įvesties lauko („Z reikšmė negali būti neigiama“, angl. „Z value can not be negative“). Jei naudotojas paspaudžia „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

28 pav. pateikiama bendra Filtravimas komponentės testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, Filtravimo komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas).





**28 pav. Duomenų filtravimo eksperimento darbų seka**

Vykdamas filtravimo komponentę buvo pasirinktas filtravimo rezultatas be atsiskyrėlių; Z reikšmė lygi 3.00; požymis – petallength.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti 29 pav. ir 30 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.01
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07f6f896453acefe5a7904332e57/experiment">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07f6f896453acefe5a7904332e57/experiment</a>

**29 pav. Komponentės „Filtravimas“ testavimo rezultatas: Techninė informacija (MII klasteris)**

**Matricinis vaizdavimas**

Objekto nr.	SEPALLENGTH (REAL)	SEPALWIDTH (REAL)	PETALLENGTH (REAL)	PETALWIDTI (REAL)
1.	5.100000	3.500000	1.400000	0.200000
2.	4.900000	3.000000	1.400000	0.200000
3.	4.700000	3.200000	1.300000	0.200000
4.	4.600000	3.100000	1.500000	0.200000
5.	5.000000	3.600000	1.400000	0.200000
6.	5.400000	3.900000	1.700000	0.400000

Atsisiųsti Atšaukti

30 pav. Komponentės „Filtravimas“ testavimo rezultatas: matricinis vaizdavimas (MII klasteris)

Įvykdžius analogišką eksperimentą naudojant *Ellipsoid* duomenų aibę MIF VU superkompiuteryje, analizuojamos darbų sekos rezultatai pateikti 31 pav. ir 32 pav.

**Techninė informacija**

Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.05
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/89722755e4d5a704b916a25a8">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/89722755e4d5a704b916a25a8</a>

Atsisiųsti Atšaukti

31 pav. Komponentės „Filtravimas“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas					
Objekto nr.	ATTR0 (REAL)	ATTR1 (REAL)	ATTR2 (REAL)	ATTR3 (REAL)	ATTR4 (REAL)
1.	0.117587	0.616645	0.318557	0.0801464	0.277455
2.	0.0656895	0.610787	0.296116	0.0961459	0.268968
3.	0.0941724	0.728045	0.304592	0.0859315	0.290502
4.	-0.0944014	0.522241	-0.188681	-0.145707	-0.0349787
5.	0.118921	0.760899	0.314854	0.127502	0.338449
6.	0.00651647	0.639546	-0.317156	-0.180186	-0.0181631
7.	0.00773307	0.612607	-0.304004	-0.178574	-0.0229073
8.	0.0961704	0.577234	0.270478	0.110039	0.260326

Atsisiųsti

Atšaukti

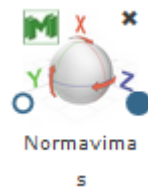
32 pav. Komponentės „Filtravimas“ testavimo rezultatas: matricinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.3.3 Komponentė „Normavimas“

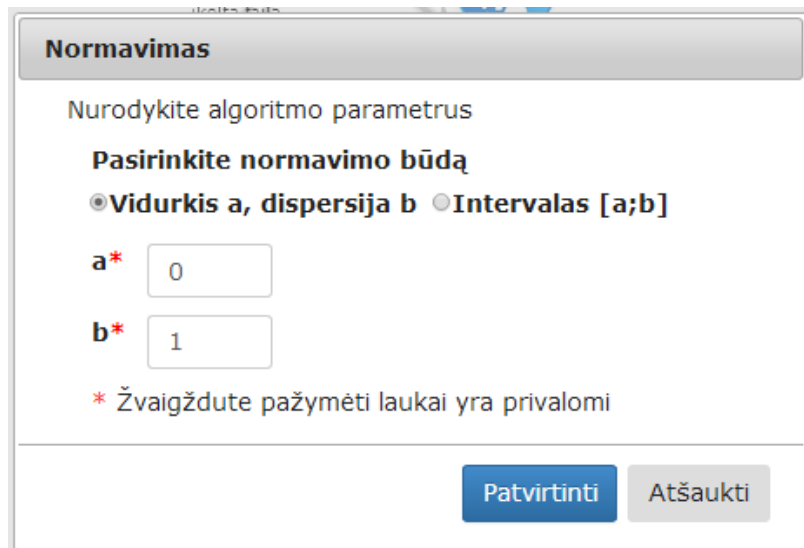
Paprastai konkretaus analizuojamo daugiamačių duomenų rinkinio parametrų reikšmės kinta skirtinguose intervaluose arba jos išreikštos skirtingais matavimo vienetais (pavyzdžiui, kilogramai, metrai, laipsniai). Todėl prieš analizuojant duomenis būtina suvienodinti šių reikšmių mastelius. Normavimas – tai duomenų reikšmių keitimas kitomis suvienodinant jų mastelius. Galimi normavimo du būdai:

- normavimas pagal vidurkį ir dispersiją, kai reikšmės pakeičiamos taip, kad kiekvieno požymio vidurkiai būtų lygūs 0, o dispersija – 1;
- normavimas į norimą intervalą – reikšmių intervalų keitimas, kai reikšmės pakeičiamos taip, kad kiekvieno požymio minimalios ir maksimalios reikšmės būtų intervale  $[a, b]$ .

Į darbalaukį įkėlus normavimo komponentę, sujungus su duomenų įkėlimo komponente ir paspaudus du kartus komponentę „Normavimas“

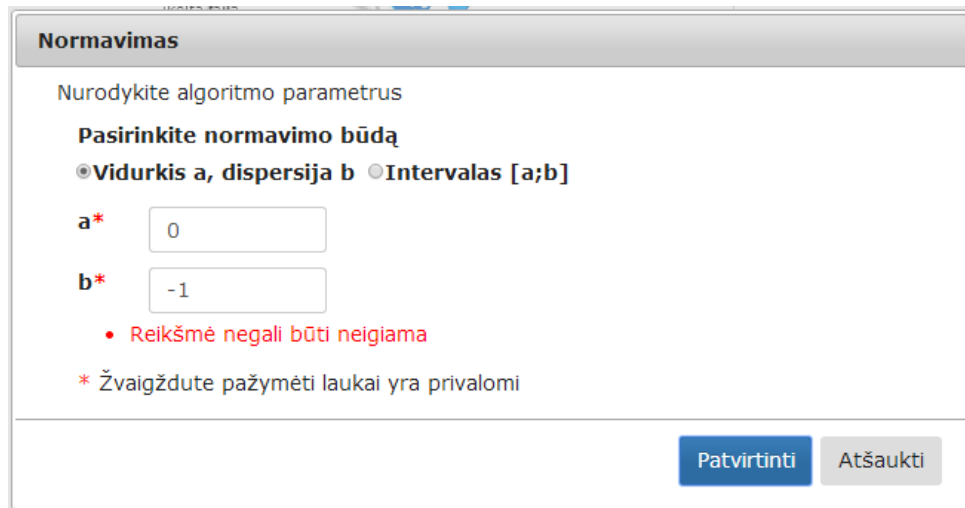


parodomas langas (33 pav.). Naudotojas turi įvesti reikalingus valdymo parametrus.



33 pav. Langas „Normavimas“

Numatytosios reikšmės normavimo būdas – „Vidurkis a, dispersija b“,  $a = 0$ ,  $b = 1$ .  $b$  parametras negali būti neigiamas („Reikšmė negali būti neigiama“, angl. „Value cannot be negative“). Pasirinkus „Intervalas [a; b]“ (numatytosios reikšmės  $a=0$ ,  $b=1$ ) turi būti tenkinama sąlyga  $a < b$  („Intervalo viršutinis rėžis turi būti didesnis nei apatinis“, angl. „Interval upper bound must be greater than lower“). Įvesties laukai tikrinami, jei įvestos negalimos reikšmės rodomas klaidos pranešimas prie kiekvieno įvesties lauko (34 pav.).

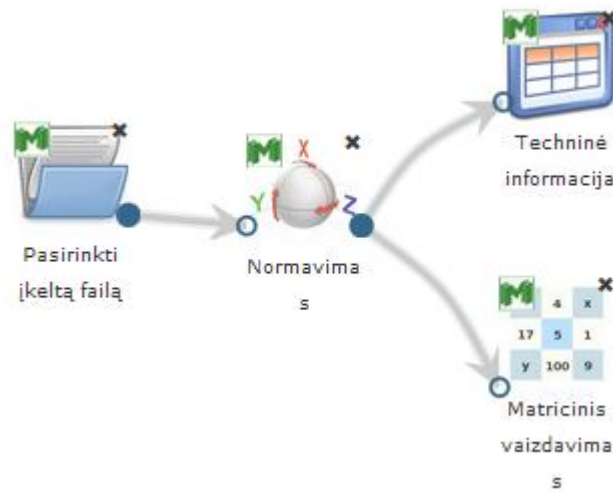


34 pav. Pranešimas apie įvedimo klaidą

Paspaudus „Patvirtinti“, jei parametrai nurodyti teisingai, parametrai išsaugomi, langas uždaromas ir grįžtama į eksperimento planavimą. Jei naudotojas paspaudžia „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „Norm data“, „Choose norm method:“, „Mean a, Standard deviation b“, „Fields marked with \* are mandatory“. Mygtukai „Ok“ ir „Cancel“.

35 pav. pateikiama bendra komponentės Normavimas testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, Normavimo komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas).



35 pav. Duomenų normavimo eksperimento darbų seka

Vykdamas normavimo komponentę, parametrų lange buvo pasirinktas „Vidurkis a, dispersija b“ normavimo būdas, kai  $a = 0$ ,  $b = 1$ . Skaičiavimai vykdyti MII klasteryje, naudota *Krūties vėžio* duomenų aibė.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 36 pav.,

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	1.06
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a/experiment/94cfa1e228e89946225fabce">http://www.damis.lt/assets/b3dc68655f2eee81930b90a/experiment/94cfa1e228e89946225fabce</a>

36 pav. Komponentės „Normavimas“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas			
Objekto nr.	CLUMP_THICKNESS (REAL)	UNIFORMITY_OF_CELL_SIZE (REAL)	UNI
1.	0.197905	-0.702212	-0.7
2.	0.197905	0.277252	0.26
3.	-0.511643	-0.702212	-0.7
4.	0.552679	1.5832	1.60
5.	-0.156869	-0.702212	-0.7
6.	1.26223	2.23618	2.27
7.	-1.22119	-0.702212	-0.7
8.	-0.866417	-0.702212	-0.4

37 pav. Komponentės „Normavimas“ testavimo rezultatas: matricinis vaizdavimas(MII klasteris)

Analogiškas eksperimentas įvykdytas MIF VU superkompiuteryje. Naudota *Elipsoidų* duomenų aibė.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose:

38 pav., 39 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	12.13
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/ffe7e1236ccb2f6acfb1978fcd1e">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/ffe7e1236ccb2f6acfb1978fcd1e</a>

38 pav. Komponentės „Normavimas“ testavimo rezultatas: Techninė informacija(MIF VU superkompiuteris)

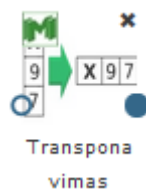
Matricinis vaizdavimas						
Objekto nr.	ATTR0 (REAL)	ATTR1 (REAL)	ATTR2 (REAL)	ATTR3 (REAL)	ATTR4 (REAL)	
1.	0.0752813	0.482318	0.411027	0.16214	0.414369	0
2.	0.575184	0.449275	0.461153	0.429657	0.453782	0
3.	0.498126	0.373541	0.547999	0.457936	0.27141	0
4.	0.485265	0.211177	0.445688	0.340566	0.272578	0
5.	0.417219	0.384851	0.425036	0.524781	0.456265	0
6.	0.781914	0.301681	0.388386	0.383415	0.43595	0
7.	0.615587	0.37025	0.586478	0.15551	0.283215	0
8.	0.203106	0.27649	0.493091	0.502552	0.455432	0

39 pav. Komponentės „Normavimas“ testavimo rezultatas: matricinis vaizdavimas(MIF VU superkompiuteris)

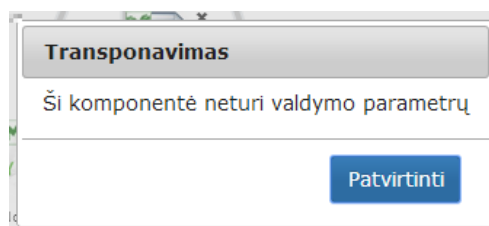
### 3.6.3.4 Komponentė „Transponavimas“

Duomenų transponavimas – procesas, kurio metu iš pradinės matricos gaunama nauja duomenų matrica, pakeičiant kiekvieną jos eilutę (stulpelį) stulpeliu (eilute), turinčiu tą pačią indeksą.

Į darbalaukį įkėlus transponavimo komponentę,



sujungus su duomenų įkėlimo komponente, paspaudus du kartus komponentę „Transponavimas“ atsiranda langas „Transponavimas“. Kadangi komponentei nereikia įvesti valdymo parametrų, atsidaręs langas atodo taip kaip pavaizduotas 40 pav. Jei naudotojas paspaudžia „Patvirtinti“ grįžtama į eksperimento planavimo langą.



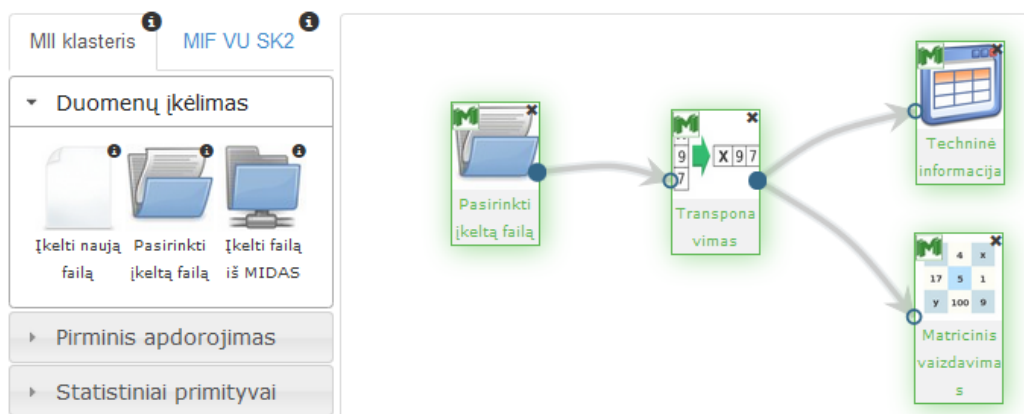
40 pav. Langas „Transponavimas“

Naudojami pavadinimai anglų kalba yra šie: Komponentės pavadinimas – „Transpose data“;

tekstas formoje – „Component does not have control parameters“, mygtukas – „Ok“.

41 pav. pateikiama bendra komponentės Transponavimas testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, transponavimo komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas). Ši komponentė neturi valdymo parametrų. Skaičiavimai vykdyti MIF VU superkompiuteryje.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 42 pav., 43 pav. Naudota *Iris* duomenų aibė.



41 pav. Duomenų pirminio apdorojimo algoritmų testavimas: Transponavimas (MII klasteris)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.01
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/d8d5a5a11edb9c2614e117201">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/d8d5a5a11edb9c2614e117201</a>

Atsisiųsti Atšaukti

42 pav. Komponentės „Transponavimas“ testavimo rezultatas: Techninė informacija (MII klasteris)

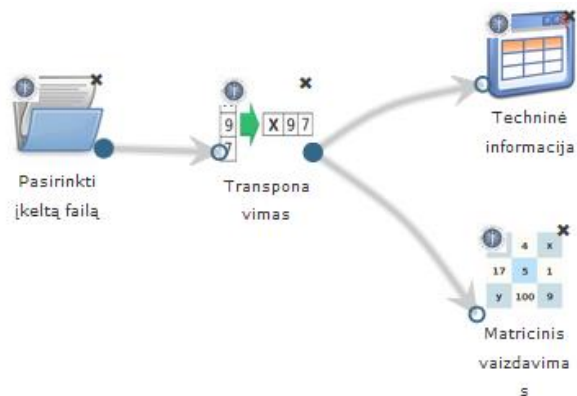


Matricinis vaizdavimas					
Objekto nr.	attr1 (NUMERIC)	attr2 (NUMERIC)	attr3 (NUMERIC)	attr4 (NUMERIC)	(NUMERIC)
1.	5.1	4.9	4.7	4.6	5
2.	3.5	3	3.2	3.1	3.6
3.	1.4	1.4	1.3	1.5	1.4
4.	0.2	0.2	0.2	0.2	0.2

43 pav. Komponentės „Transponavimas“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)

44 pav. pateikiama bendra komponentės „Transponavimas“ testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, transponavimo komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas). Skaičiavimai atlikti MIF VU superklasteryje. Atliktas analogiškas eksperimentas naudojant *Elipsoidų* duomenų aibę.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti 45 pav. ir 46 pav.



44 pav. Transponavimas (MIF VU superkompiuteris)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.08
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/b8dc4b1261f4b2cb78a946a914">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/b8dc4b1261f4b2cb78a946a914</a>

45 pav. Komponentės „Transponavimas“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas				
attr1111 (NUMERIC)	attr1112 (NUMERIC)	attr1113 (NUMERIC)	attr1114 (NUMERIC)	attr1115 (NUMERIC)
0.0970608	0.131066	0.0139969	-0.0418208	-0.00105799
0.253555	0.0643991	-0.0840271	0.127534	-0.0408508
-0.147135	-0.0217466	-0.248103	0.0269047	-0.0346616
-0.129593	0.132108	0.0986024	0.0447228	0.0131652
-0.0218506	-0.0156425	0.322221	0.0117744	-0.0333159
0.12152	-0.122559	-0.121421	-0.0169887	-0.0264433
0.113588	0.0269825	-0.508258	0.032414	0.00705855
0.195721	0.0234859	0.173674	-0.0460131	-0.0132984

46 pav. Komponentės „Transponavimas“ testavimo rezultatas: Matricinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.3.5 Komponentė „Skaidymas“

Duomenų skaidymas – pradinės duomenų aibės skaidymas į smulkesnius poaibius. Gautuosius poaibius galima apdoroti lygiagrečiai arba nagrinėti vieną iš jų, pavyzdžiui, imtis gali atspindėti populiacijos dėsningumus.

Duomenų aibės skaidymo būdai:

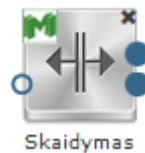
1. Nekeičiam analizuojamos duomenų aibės objektų tvarka ir aibė dalinama į du poaibius pasirinktu santykiu.
2. Pradinės duomenų aibės objektai perrikiuojami atsitiktine tvarka ir tuomet dalinama į du poaibius pasirinktu santykiu.

Komponentės „Skaidymas“ **rezultatas**: dvi duomenų matricos, gautos po pradinės duomenų aibės skaidymo į du poaibius, kurių dydžiai nustatomi komponentės parametru lange.

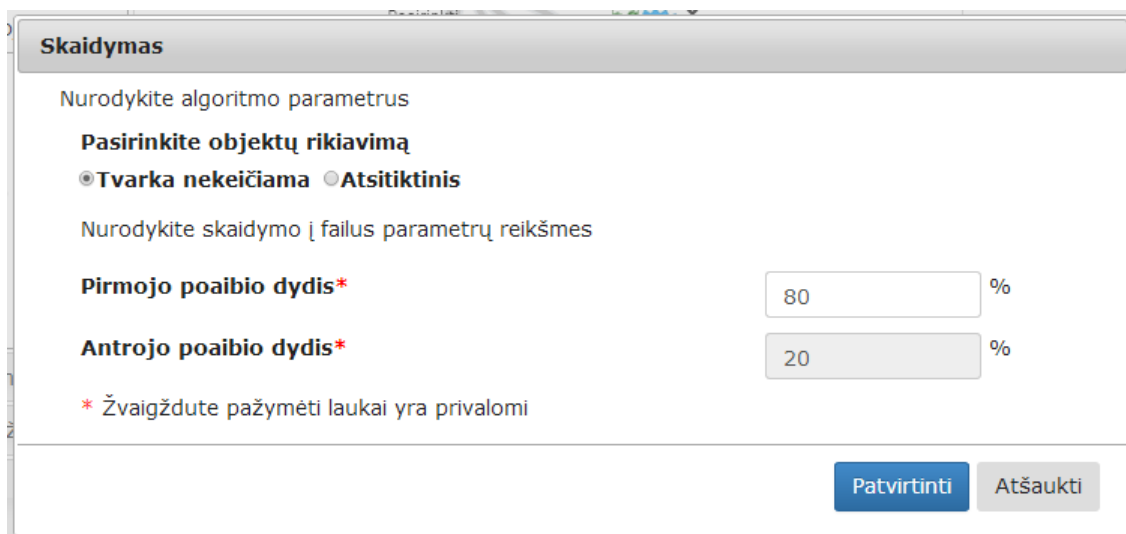
Parametrai:

- **Objektų rikiavimo tipas – Tvarka nekeičiama** (objektų išdėstymo tvarka nebus keičiama); **Atsitiktinis** (objektų išdėstymo tvarka bus keičiama atsitiktine tvarka).
- **Pirmojo poaibio dydis** – reikia pasirinkti pirmojo poaibio santykinį dydį.
- **Antrojo poaibio dydis** – antrasis poaibio dydis skaičiuojamas automatiškai taip, kad abiejų poaibių dydžių suma būtų lygi 100 %.

Į darbalaukį įkėlus skaidymo komponentę,



sujungus su duomenų įkėlimo komponente ir paspaudus du kartus komponentę „Skaidymas“ atsidaro langas (47 pav.). Naudotojas turi įvesti ir pasirinkti reikiamus valdymo parametrus.



47 pav. Langas „Skaidymas“

Reikia pasirinkti ar bus atliekamas objektų perrikiavimas, taip pat turi įvesti kokiu santykiu bus skaidomi duomenys į failus, t. y. naudotojas turi nurodyti kiek procentų viso objektų skaičiaus skiria vienam ir kitam failui. Bendras abiejų failų objektų kiekis sudaro 100 %. Kai naudotojas įveda pirmojo poaibio dydį, į antrame lauke likusi procentinė dalis apskaičiuojama ir įrašoma automatiškai. Antrojo poaibio dydžio reikšmės naudotojas keisti negali. Kai naudotojas įvedė ir pasirinko valdymo parametrus turi spausti mygtuką „Patvirtinti“, tuomet bus išsaugomi įvesti parametrai, modalinis langas uždaromas ir grįžtama į eksperimento planavimą. Jei naudotojas paspaudžia „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „Split data“, „Choose object sort type:“, „Order left intact“, „Random“, „Set splitting to files parameter values“, „First subset size“, „Second subset

size“ „Fields marked with \* are mandatory“. Mygtukai „Ok“ ir „Cancel“.

Pateikiama bendra komponentės Skaidymas testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, Skaidymo komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas).



48 pav. Duomenų pirminis apdorojimas: Skaidymas (MII klasteris)

Vykdamas filtravimo komponentę parametrų lange buvo nurodyta, kad atliekant objektų rikiavimą tvarka nekeičiama; pirmojo poaibio dydis – 70 %; antrojo poaibio dydis – 30 %. Naudota *Iris* duomenų aibė.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 49 pav., 50 pav., 51 pav. Kadangi Skaidymo komponentės rezultatas yra dvi duomenų aibės, tai ir paveiksluose pateiktos dvi gautos duomenų aibės (pirmas poaibis sudarytas iš 105 elementų, o antras – iš 45 elementų).

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.01
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/967bff95ec6abb2b525b3e4319">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/967bff95ec6abb2b525b3e4319</a>

49 pav. Komponentės „Skaidymas“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas				
98.	6.200000	2.900000	4.300000	1.300000
99.	5.100000	2.500000	3.000000	1.100000
100.	5.700000	2.800000	4.100000	1.300000
101.	6.300000	3.300000	6.000000	2.500000
102.	5.800000	2.700000	5.100000	1.900000
103.	7.100000	3.000000	5.900000	2.100000
104.	6.300000	2.900000	5.600000	1.800000
105.	6.500000	3.000000	5.800000	2.200000

50 pav. Komponentės „Skaidymas“ testavimo rezultatas: Matricinis vaizdavimas (1 poaibis) (MII klasteris)

Matricinis vaizdavimas				
36.	6.700000	3.100000	5.600000	2.400000
37.	6.900000	3.100000	5.100000	2.300000
38.	5.800000	2.700000	5.100000	1.900000
39.	6.800000	3.200000	5.900000	2.300000
40.	6.700000	3.300000	5.700000	2.500000
41.	6.700000	3.000000	5.200000	2.300000
42.	6.300000	2.500000	5.000000	1.900000
43.	6.500000	3.000000	5.200000	2.000000
44.	6.200000	3.400000	5.400000	2.300000
45.	5.900000	3.000000	5.100000	1.800000

51 pav. Komponentės „Skaidymas“ testavimo rezultatas: Matricinis vaizdavimas(2 poaibis) (MII klasteris)

Analogiškas eksperimentas atliktas MIF VU superkompiuteryje, čia naudojamos komponentės skirtos MIF VU superkompiuterio skaičiavimams. Vykdam skaidymo komponentę parametrų lange buvo nurodyta, kad atliekant objektų rikiavimą tvarka nekeičiama; pirmojo poaibio dydis – 80 %; antrojo poaibio dydis – 20 %. Gauti rezultatai iliustruojami naudojant *Elipsoidų* duomenų aibę.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 52 pav., 53 pav., 54 pav. Kadangi Skaidymo komponentės rezultatas yra dvi duomenų aibės, tai ir paveiksluose pateiktos dvi gautos duomenų aibės (pirmas poaibis sudarytas iš 223 elementų, o antras – iš 892 elementų)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.09
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/f81eb16889e969411e4caa61c2">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/f81eb16889e969411e4caa61c2</a>

52 pav. Komponentės „Skaidymas“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas					
215.	0.0950573	-0.00202503	0.0519107	-0.000283513	0.
216.	-0.123456	-0.313518	-0.0776662	-0.135374	0.
217.	0.409747	0.222106	-0.00924002	0.178613	0.
218.	0.133513	0.150896	-0.0243066	-0.0192356	0.
219.	0.0970608	0.253555	-0.147135	-0.129593	-0.
220.	0.131066	0.0643991	-0.0217466	0.132108	-0.
221.	0.0139969	-0.0840271	-0.248103	0.0986024	0.
222.	-0.0418208	0.127534	0.0269047	0.0447228	0.
223.	-0.00105799	-0.0408508	-0.0346616	0.0131652	-0.

53 pav. Komponentės „Skaidymas“ testavimo rezultatas: Matricinis vaizdavimas (1 poaibis) (MIF VU superkompiuteris)

Matricinis vaizdavimas				
884.	0.069528	0.256968	0.125556	0.051457
885.	-0.172453	-0.118947	-0.0163282	0.027075
886.	-0.0560274	0.143491	-0.0601443	0.0709194
887.	-0.0774225	-0.0859724	0.29451	0.0252526
888.	-0.0716035	0.11023	-0.0267516	0.0503304
889.	0.0409206	0.425296	0.165159	0.10934
890.	0.251333	0.224631	0.0715843	0.32218
891.	-0.0446464	-0.133388	-0.0589776	0.104867
892.	0.128524	-0.00709554	0.0250547	0.0234893

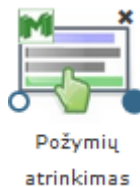
Atsisųsti

Atšaukti

54 pav. Komponentės „Skaidymas“ testavimo rezultatas: Matricinis vaizdavimas(2 poaibis)  
(MIF VU superkompiuteris)

### 3.6.3.6 Komponentė „Požymių atrinkimas“

Požymių atrinkimo komponentė



suformuoja naują duomenų aibę, sudarytą iš pasirinktų analizuojamos aibės požymių (atributų) reikšmių. Taip pat numatyta naujajame faile klase paskelbti bet kokią pasirinktą požymį.

Komponentės „Požymių atrinkimas“ **rezultatas**: nauja duomenų matrica, sudarytą iš pasirinktų analizuojamos aibės požymių (atributų) reikšmių.

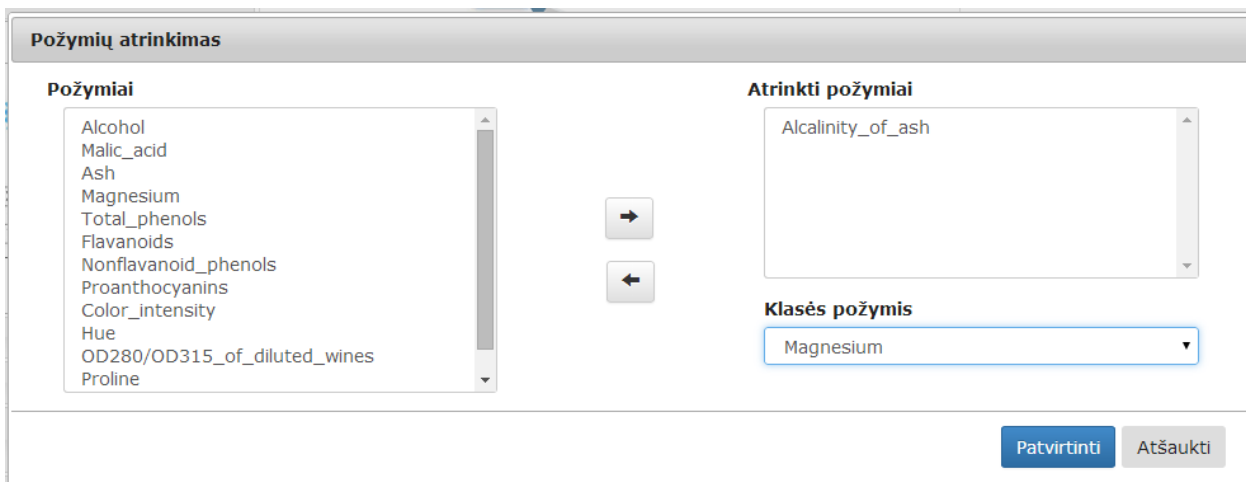
Parametrai:

- **Atrinkti požymiai** - pasirenkama požymių aibė iš kurios bus formuojama nauja duomenų aibė. Numatyta reikšmė - tuščias sąrašas.
- **Klasės požymis** - naujajame faile klase paskelbiamas bet koks pasirinktas požymis. Numatyta reikšmė - nepasirinktas joks požymis.

Į darbalaukį įkėlus požymių atrinkimo komponentę, ją sujungus su duomenų įkėlimo komponente ir paspaudus du kartus komponentę „Požymių atrinkimas“ atsidaro modalinis langas. Atsidariusiame lange naudotojas gali pasirinkti reikiamus požymius ir nurodyti, kuris iš požymių



bus klasė naujai suformuotame faile (55 pav.).



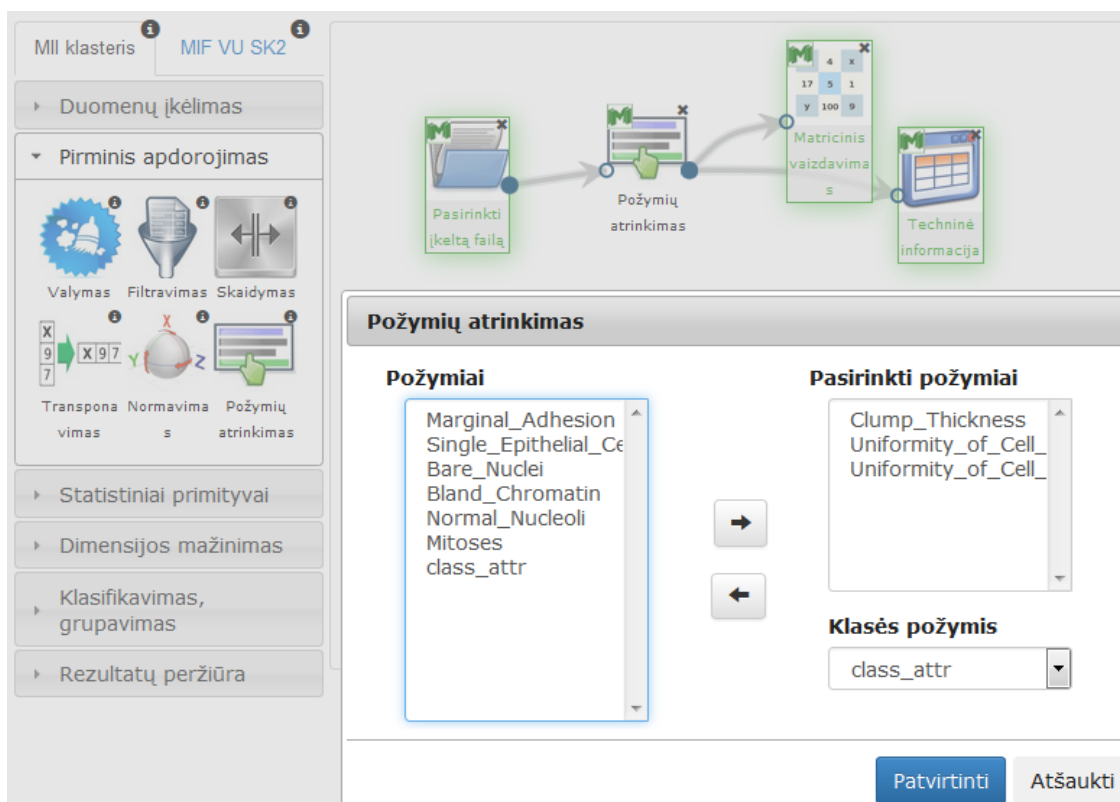
55 pav. Langas „Požymių atrinkimas“

Komponentės numatytosios reikšmės sąrašas „Atrinkti požymiai“ – tuščias, „Klasės požymis“ – nepasirinkta. Atsidarius modaliniam langui, pirmame sąrašo stulpelyje pateikiamas visas atributų sąrašas nuskaitytas iš arff failo įskaitant ir klasės atributą, tačiau prie šio automatiškai pridedama santrumpa „\_attr“. Naudotojas gali pasirinkti pažymėdamas pele vieną ar kelis požymius ir paspausti rodyklę „į dešinę“ (tooltip’as „ pridėti“ angl. „Add“). Pažymėti požymiai perkeliama į antrą sąrašo stulpelį pavadinimu „Atrinkti požymiai“. Iš pirmojo stulpelio pasirinktas atributas dingsta. Jei naudotojas nori pašalinti iš antrojo stulpelio atributą ar jų grupę, jis turi pažymėti juos ir spausti rodyklę „į kairę“ (tooltip’as „pašalinti“ angl. „Remove“), tuomet atributas turi dingti iš antro stulpelio ir turi atsirasti pirmame. Naujai formuojamo arff failo klasės požymių naudotojas gali pasirinkti iš iškrentančio meniu pavadinimu „Klasės požymis“. Iškrentančių sąrašą sudaro visi atributai esantys pradiniam arff faile (klasės atributas formuojant sąrašą nėra pervadinamas). Naudotojas naujajame faile klase paskelbti gali bet kokį atributą.

Kai naudotojas paspaudžia mygtuką „Patvirtinti“, išsaugomi naudotojo pasirinkimai, ir langas uždaromas. Jei naudotojas paspaudžia „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „Feature selection“, „Attributes“, „Selected attributes“, „Class attribute“. Mygtukai „Ok“ ir „Cancel“.

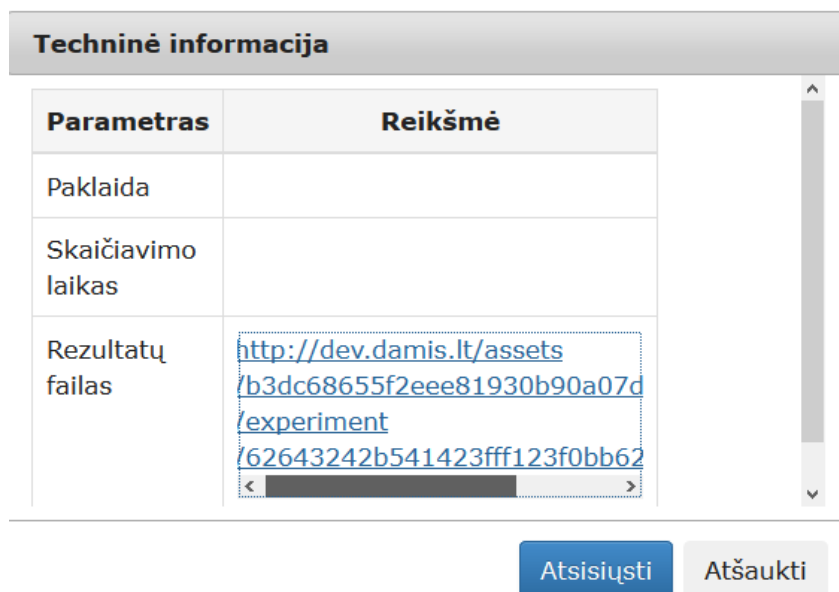
56 pav. pateikiama bendra komponentės Požymių atrinkimas testavimo darbų seka (kartu su Transponavimu ir Normavimu), sudaryta iš failo įkėlimo komponentės, Požymių atrinkimo komponentės ir rezultatų peržiūros komponentių (Techninė informacija ir Matricinis vaizdavimas).



56 pav. Duomenų pirminis apdorojimas: Požymių atrinkimas (MII klasteris)

Vykdamas požymių atrinkimo komponentę valdymo parametruose buvo atrinkti pirmieji trys požymiai. Eksperimento rezultatų iliustravimui naudota *krūties vėžio* duomenų aibė. Skaičiavimai atlikti MII klasteryje.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 57 pav. ir 58 pav.



57 pav. Komponentės Požymių atrinkimas testavimo rezultatas: Techninė informacija(MII klasteris)

Matricinis vaizdavimas			
Objekto nr.	Clump_Thickness (REAL)	Uniformity_of_Cell_Size (REAL)	Uniformi
1.	5	1	1
2.	5	4	4
3.	3	1	1
4.	6	8	8
5.	4	1	1
6.	8	10	10

58 pav. Komponentės Požymių atrinkimas testavimo rezultatas: matricinis vaizdavimas (MII klasteris)

Analogiškas eksperimentas požymių atrinkimo darbų sekos įvykdytas MIF VU superkompiuteryje.

Įvykdžius eksperimentą (pasirinkti pirmieji trys požymiai), analizuojamos darbų sekos rezultatai pateikti paveiksluose: 59 pav., 60 pav. Naudota *Elipsoidų* duomenų aibė.

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/99c2b5299ff53b8feb6ff98b18a">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/99c2b5299ff53b8feb6ff98b18a</a>

59 pav. Komponentės „Požymių atrinkimas“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Objekto nr.	attr1 (real)	attr2 (real)	attr3 (real)	CLASS ( )
1.	0.12644	-0.0775405	-0.252869	0.0112095
2.	0.0835017	-0.0329761	-0.0337818	-0.106012
3.	-0.0149113	0.0442346	-0.010622	0.0316691
4.	-0.225898	-0.0467255	-0.106744	-0.134266
5.	-0.000214195	-0.065086	0.0441215	0.0733144
6.	-0.108291	-0.0976692	-0.0716518	0.0136823
7.	-0.0191883	0.0784447	-0.258299	-0.162126
8.	-0.141026	-0.00458187	0.025917	-0.0620802
9.	-0.0478093	0.0479412	-0.0698611	-0.27243

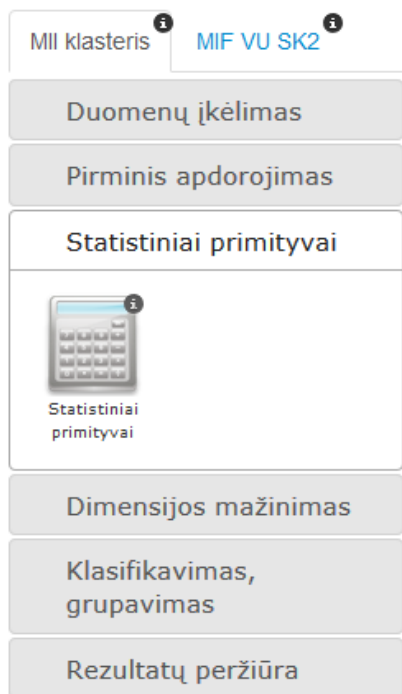
Atsisiųsti

Atšaukti

60 pav. Komponentės „Požymių atrinkimas“ testavimo rezultatas: matricinis vaizdavimas (MIF VU superkompiuteris)

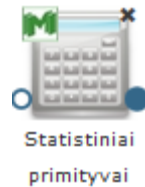
### 3.6.4 Kairiojo meniu skiltis „Statistiniai primityvai“

Pasirinkus statistinių primityvų skiltį, galima pasirinkti tik vieną komponentę „Statistiniai primityvai“ (61 pav.) – tai statistinių primityvų komponentė, kurią nutempus ir sujungus su kita komponente, bus apskaičiuojami duomenų failo statistiniai primityvai (min, max, vidurkis, mediana, standartinis nuokrypis).

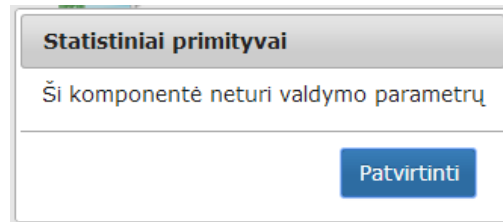


61 pav. Meniu skiltis „Statistiniai primityvai“

Į darbalaukį įkėlus statistinių primityvų komponentę,



sujungus su duomenų įkėlimo komponente ir paspaudus du kartus komponentę atsiranda langas „Statistiniai primityvai“. Kadangi komponentei nereikia įvesti valdymo parametru, atsiradęs modalinis langas atrodo taip kaip pavaizduotas 62 pav.paveiksle. Jei naudotojas paspaudžia „Patvirtinti“ grįžtama į eksperimento planavimo langą.

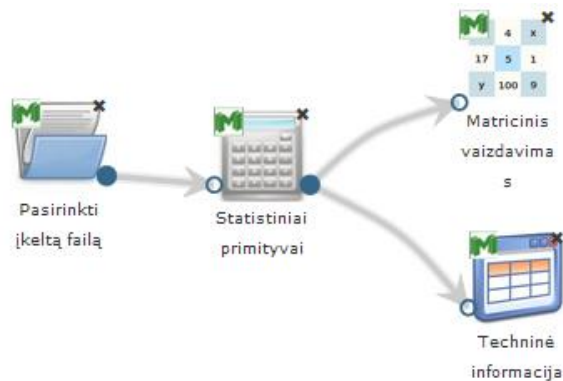


62 pav. Langas „Statistiniai primityvai“

Naudojami pavadinimai anglų kalba yra šie: Komponentės pavadinimas – „Statistical data“; tekstas formoje – „Component does not have control parameters“, mygtukas – „Ok“.

63 pav. pateikiama bendra komponentės „Statistiniai primityvai“ testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, Statistinių primityvų komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas).

Įvykdžius eksperimentą MII klasteryje, naudojant *Krūties vėžio* testinę duomenų aibę, analizuojamos darbų sekos rezultatai pateikti paveikluose: 64 pav., 65 pav.



63 pav. Statistinių primityvų algoritmo testavimas (MII klasteris)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.01
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/199484189525cb745684df02a">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/199484189525cb745684df02a</a>

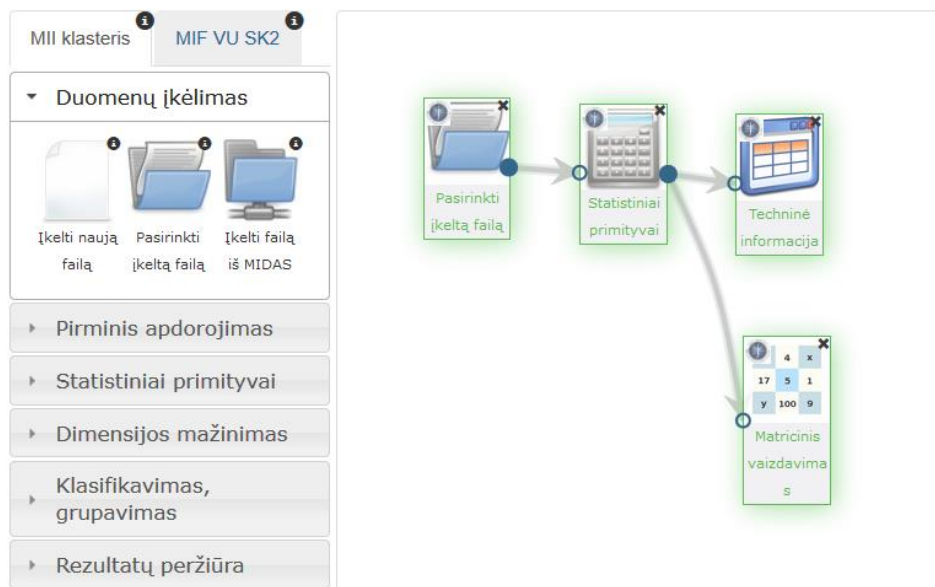
64 pav. Komponentės „Statistiniai primityvai“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas					
Objekto nr.	min (NUMERIC)	max (NUMERIC)	mean (NUMERIC)	std (NUMERIC)	r
1.	1	10	4.44217	2.8187	4
2.	1	10	3.15081	3.0629	1
3.	1	10	3.21523	2.98639	1
4.	1	10	2.83016	2.86246	1
5.	1	10	3.23426	2.22146	2
6.	1	10	3.54466	3.64119	1
7.	1	10	3.4451	2.4479	3
8.	1	10	2.86969	3.05043	1

65 pav. Komponentės „Statistiniai primityvai“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)

66 pav. pateikiama bendra komponentės Statistiniai primityvai testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, Statistinių primityvų komponentės ir rezultatų peržiūros komponentių (techninė informacija ir matricinis vaizdavimas), kuri vykdyta MIF VU superkompiuteryje.

Įvykdžius eksperimentą naudojant *Elipsoidų* testinę duomenų aibę, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 67 pav., 68 pav.



66 pav. Statistinių primityvų algoritmo testavimas (MIF VU superkompiuteris)

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.05
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/c0aca7ba833939c45b70775eed">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/c0aca7ba833939c45b70775eed</a>

67 pav. Komponentės „Statistiniai primityvai“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas					
Objekto nr.	min (NUMERIC)	max (NUMERIC)	mean (NUMERIC)	std (NUMERIC)	r (NUMERIC)
1.	-0.364364	0.50815	0.0360827	0.136063	0.00
2.	-0.500315	0.79915	0.0761967	0.17304	0.00
3.	-0.442964	0.446086	-0.00418633	0.123508	-0.00
4.	-0.385656	0.43331	-0.000369769	0.118872	0.00
5.	-0.434506	0.581354	0.00355531	0.152369	0.00
6.	-0.374911	0.30778	-0.00557058	0.110325	0.00
7.	-0.600216	0.755356	-0.0319318	0.163482	-0.00
8.	-0.30714	0.474132	0.00808711	0.120342	-0.00

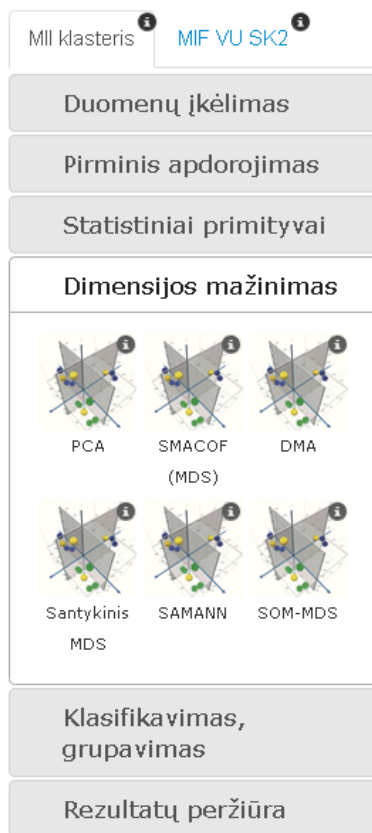
Atsisiųsti Atšaukti

68 pav. Komponentės „Statistiniai primityvai“ testavimo rezultatas: matricinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.5 Kairiojo meniu skiltis „Dimensijos mažinimas“

Pasirinkus dimensijos mažinimo skiltį, atsidaro realizuotų dimensijos mažinimo algoritmų komponentės. „PCA“ – tai pagrindinių komponentių analizės metodas, „SMACOF (MDS)“ – daugiamačių skalių algoritmas, „DMA“ – tai diagonalinio mažoravimo algoritmą realizuojanti komponentė, „Relative MDS“ – tai santykinų daugiamačių skalių algoritmą realizuojanti komponentė, „SOM-MDS“ – tai saviorganizuojančio neuroninio tinklo (SOM) ir daugiamačių skalių metodo (MDS) junginio algoritmą realizuojanti komponentė, „SAMANN“ – dirbtinio neuroninių tinklo algoritmą realizuojanti komponentė. Dimensijos mažinimo meniu skiltis pateikta 69 pav.





69 pav. Meniu skiltis „Dimensijos mažinimas“

Tolimesniuose skyriuose detaliai apžvelgiamos visos „Dimensijos mažinimas“ skilties komponentės ir pateikiamos jų naudojimosi instrukcijos.

### 3.6.5.1 Komponentė „PCA“

Pagrindinių komponentių analizė (angl. principal component analysis, PCA) yra klasikinis statistikos metodas. Tai tiesinė duomenų transformacija, plačiai naudojama duomenų analizei kaip daugiamačių duomenų dimensijos mažinimo metodas.

Pagrindinė pagrindinių komponentių analizės idėja yra sumažinti duomenų dimensiją atliekant tiesinę transformaciją ir atsisakant dalies po transformacijos gautų naujų komponentių, kurių dispersijos yra mažiausios. Iš pradžių ieškoma krypties, kuria dispersija yra didžiausia. Didžiausią dispersiją turinti kryptis vadinama pirmąja pagrindine komponente. Ji eina per duomenų centrinį tašką. Tai taškas, kurio komponentės yra analizuojamą duomenų aibę sudarančių taškų atskirų komponentių vidurkiai. Visų taškų vidutinis atstumas iki šios tiesės yra minimalus, t. y., ši tiesė yra kiek galima arčiau visų duomenų taškų. Antrosios pagrindinės komponentės ašis taip pat turi eiti per duomenų centrinį tašką ir ji turi būti statmena pirmosios pagrindinės komponentės ašiai.

Komponentės „PCA“ **rezultatas**: nauja duomenų matrica, apskaičiuota PCA dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2.

Parametrai:

- **Projekcijos dimensija** – nurodoma sumažinta pradinių duomenų dimensija (**dimensija**). Dimensija negali būti neigiamas skaičius, taip pat, jei pasirinkimas yra „Dimensija“ įvestos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius.
- **Požymių santykinė suminė dispersija** – dispersijos dalis, kurią norima išlaikyti (**dispersija**). Santykinės suminės dispersijos galimos reikšmės yra intervale (0; 100] %.

Į darbalaukį įkėlus PCA komponentę,



ją sujungus su failo įkėlimo komponente ir paspaudus du kartus komponentę „PCA“ atidaromas modalinis langas. Naudotojas turi pasirinkti ir įvesti reikalingus paramentrus. Numatytosios reikšmės – PCA projekcijos dimensija ir dispersija (gali būti realus skaičius) reikšmės yra 2 ir 80 % atitinkamai. Modalinis PCA langas ir PCA algoritmo vykdymo eksperimento darbų seka pateikiama 70 pav.

70 pav. PCA langas ir PCA algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Naudotojui paspaudus „Patvirtinti“, tikrinama įveta dimensijos/dispersijos reikšmė. Dimensija negali būti neigiamas skaičius, taip pat, jei pasirinkimas yra „Dimensija“ įvestos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius. Dimensijos parametras yra sveikas skaičius. Jei pasirinkimas yra „Požymių santykinė suminė dispersija“ – santykinės dispersijos galimos reikšmės yra intervale (0; 100] %. Jei reikšmės netinkamos,

rodomas klaidos pranešimas šalia įvesties lauko.

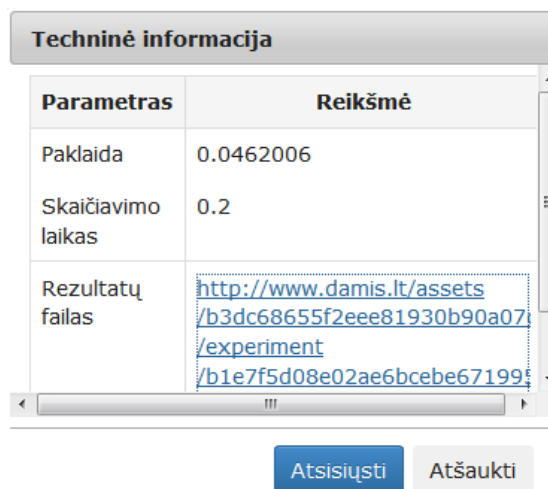
Kai naudotojas spaudžia „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „PCA“, „Specify the parameters of the algorithm.“, „Choose PCA projection:“ „Space“, „Attribute relative cumulative variance“, „Space/Variance“, „Fields marked with \* are mandatory“. Klaidos pranešimai: „Value cannot be negative“, „Relative cumulative variance must be in interval (0; 100]“, „Space cannot be real value“, „Space dimension cannot be greater than quantity of attributes in arff file“. Mygtukai: „Ok“ ir „Cancel“.

70 pav. pateikiama bendra komponentės PCA vykdymo darbų seka, sudaryta iš failo įkėlimo komponentės, PCA komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Eksperimentė naudota krūties vėžio testinė duomenų aibė.

PCA algoritmo valdymo parametrai: sumažintos dimensijos skaičius  $d$  arba dispersijos dalis, kurią siekiama išlaikyti transformuojant duomenis naudojant pagrindines komponentes. Vykdam eksperimentą PCA komponentės parametrų lange buvo nurodyta, kad PCA projekcijos dimensija yra 2.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 71 pav., 72 pav., 73 pav.



Parametras	Reikšmė
Paklaida	0.0462006
Skaičiavimo laikas	0.2
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/b1e7f5d08e02ae6bcebe67199">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/b1e7f5d08e02ae6bcebe67199</a>

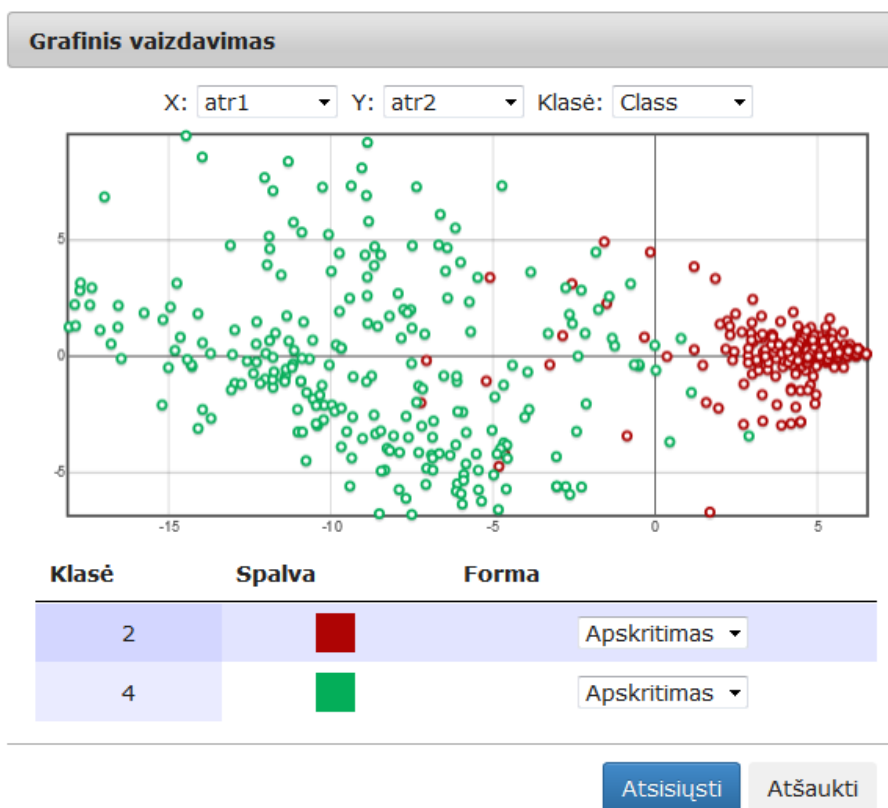
71 pav. Komponentės „PCA“ testavimo rezultatas: Techninė informacija (MII klasteris)

**Matricinis vaizdavimas**

Objekto nr.	atr1 (REAL)	atr2 (REAL)	CLASS (2, 4)
1.	4.4821	0.0240233	2
2.	-4.81965	-4.73973	2
3.	4.63296	-0.609533	2
4.	-5.09653	3.3866	2
5.	4.11643	-0.0988651	2
6.	-15.009	-0.490217	4
7.	1.68552	-6.70708	2
8.	4.97955	0.409244	2
9.	5.45816	0.979859	2

Atsisiųsti Atšaukti

72 pav. Komponentės „PCA“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)



73 pav. Komponentės „PCA“ testavimo rezultatas: Grafinis vaizdavimas (MII klasteris)

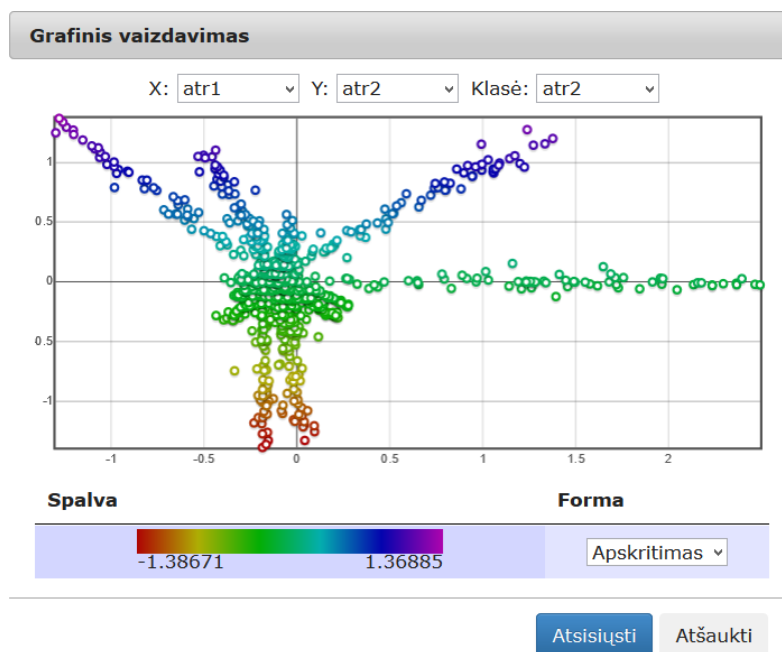
Komponentės testavimas atliktas ir MIF VU superkompiuteryje. Darbų schema analogiška pateiktai 70 pav., tik naudotos MIF VU superkompiuteriui skirtos komponentės. Testavimo rezultatai, naudojant *Elipsoidų* testinę duomenų aibę, pateikiami 74 pav., 75 pav. ir 76 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.274041
Skaičiavimo laikas	0.1
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/85b993153939c7ee5fff0fef429/">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/85b993153939c7ee5fff0fef429/</a>

74 pav. Komponentės „PCA“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas		
1.	-0.368997	-0.251897
2.	-0.266072	-0.0796039
3.	-0.234181	0.2685
4.	-0.981542	0.790298
5.	-0.061039	-0.184627
6.	-0.011376	-0.205118
7.	-0.372893	0.812845
8.	-0.315691	0.0203876
9.	-1.02205	0.983513
10.	-0.191299	0.0350656

75 pav. Komponentės „PCA“ testavimo rezultatas: Matricinis vaizdavimas (MIF VU superkompiuteris)



76 pav. Komponentės „PCA“ testavimo rezultatas: Grafinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.5.2 Komponentė „SMACOF (MDS)“

Daugiamatės skalės (angl. MDS) – tai grupė metodų, plačiai naudojamų daugiamačių duomenų analizei įvairiose šakose, ypač ekonomikoje, socialiniuose moksluose, medicinoje ir kt. Gausu šio metodo realizacijų, kurios viena nuo kitos skiriasi naudojamais vizualizavimo kokybės kriterijais, optimizavimo algoritmais ar prielaidomis apie duomenis. Naudojantis MDS, ieškoma daugiamačių duomenų projekcijų mažesnio skaičiaus matmenų erdvėje, siekiant išlaikyti analizuojamos aibės objektų artimumus – panašumus arba skirtingumus. Gautuose vaizduose panašūs objektai išdėstomi arčiau vieni kitų, o skirtingi – toliau vieni nuo kitų.

„SMACOF (MDS)“ – tai daugiamačių skalių metodas, kuriam pritaikytas vienas geriausių SMACOF optimizavimo algoritmų, taikomų daugiamačių skalių paklaidos minimizavimui. Algoritmas yra paprastas, bet efektyvus, kadangi garantuoja paklaidos funkcijos konvergavimą į lokalų minimumą su tiesiniu konvergavimo greičiu.

SMACOF algoritmui buvo pritaikytas Gauso Zeidelio metodas. Modifikacijos esmė – nauji projekcijos taškai iteracinio proceso eigoje apskaičiuojami, remiantis jau prieš tai toje pačioje iteracijoje apskaičiuotais taškais.

Komponentės „SMACOF (MDS)“ **rezultatas**: nauja duomenų matrica, apskaičiuota SMACOF dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2.

Parametrai:

- **Projekcijos dimensija**– nurodoma sumažinta pradinių duomenų dimensija (projekcijos dimensija).
- **Maksimalus iteracijų skaičius** – maksimalus iteracijų skaičius. Maksimalus iteracijų skaičius yra 1000.

- **Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų** – skirtumas tarp paklaidų, gautų tarp gretimų iteracijų. Skirtumas tarp paklaidų intervale nustatomas iš intervalo  $[10e-8; \infty)$ .
- **Ar taikyti Zeidelio modifikaciją (TAIP | NE)?** – parametras, kuris nurodo taikyti arba netaikyti Zeidelio modifikaciją. Reikšmė „True“, jei taikoma Zeidel modifikacija; jei nenurodyta kitaip, Zeidel modifikacija nebus taikoma.

Į darbalaukį įkėlus komponentę „SMACOF (MDS)“, sujungus ją su duomenų failų įkėlimo komponente ir paspaudus du kartus komponentę „SMACOF (MDS)“ atidaromas langas. Naudotojas turi pasirinkti ir įvesti reikalingus paramentrus. Numatytosios reikšmės: projekcijos dimensija – 2, maksimalus iteracijų skaičius – 100, skirtumas tarp paklaidų, gautų tarp gretimų iteracijų – 0,0001, Ar taikyti Zeidelio modifikaciją – ne. Modalinis SMACOF (MDS) langas ir SMACOF (MDS) algoritmo vykdymo eksperimento darbų seka (MII klasteris) pavaizduoti 77 pav.



77 pav. SMACOF (MDS) langas ir SMACOF (MDS) algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametų reikšmės, parametų reikšmės negali būti neigiamos, jei įvesta neigiama reikšmė rodomas klaidos pranešimas ties neteisingai įvestu lauku. Įvestos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius. Dimensijos parametras yra sveikas skaičius. Maksimalus iteracijų skaičius  $[1; 1000]$  (sveikas skaičius). Skirtumas tarp paklaidų intervale  $[10^{-8}; \infty)$ .

Naudotojui paspaudus mygtuką „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „SMACOF (MDS)“, “Specify the parameters of the

algorithm.”, “Projection space”, “Maximum number of iteration”, “Minimal stress change”, “Does apply Seidel modification? – Yes, No”, „Fields marked with \* are mandatory“. Klaidos pranešimai: „Value cannot be negative“, „Relative cumulative variance must be in interval (0; 100]”, “Projection space cannot be real value”, “Projection space cannot be greater than quantity of attributes in arff file”, “Minimal stress change must be in interval  $[10^{-8}; \infty)$ “, “Maximum number of iteration must be in interval [1; 1000]”. Mygtukai: „Ok“ ir „Cancel“.

77 pav. pateikiama bendra komponentės SMACOF (MDS) testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, SMACOF (MDS) komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Naudota *Krūties vėžio* testinė duomenų aibė.

SMACOF (MDS) valdymo parametrai: maksimalus iteracijų skaičius; minimalus skirtumas tarp apskaičiuotų MDS paklaidų, gautų gretimose iteracijose; projekcijos dimensija. Vykdamas eksperimentą, SMACOF (MDS) komponentės parametru lange buvo nurodyta, kad SMACOF (MDS) projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; skirtumas tarp paklaidų – 0,0001; Zeidelio modifikacija nebuvo taikoma.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 78 pav., 79 pav., 80 pav. Rezultatai pateikiami atlikus eksperimentą su vynu duomenimis.

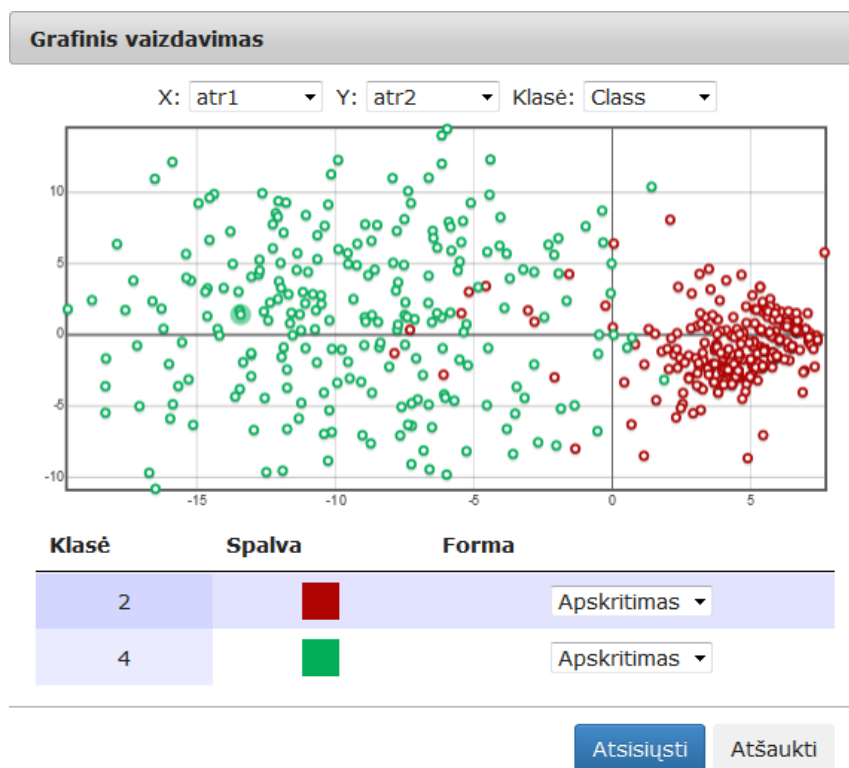
Techninė informacija	
Parametras	Reikšmė
Paklaida	0.0229664
Skaičiavimo laikas	0.05
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/2a86ce890e9dea00d4e12e1a6">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/2a86ce890e9dea00d4e12e1a6</a>

78 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Techninė informacija (MII klasteris)



Matricinis vaizdavimas			
Objekto nr.	atr1 (REAL)	atr2 (REAL)	CLASS (2, 4)
1.	4.09444	-2.28658	2
2.	-5.44588	1.52409	2
3.	4.65416	-0.215961	2
4.	-6.09098	-2.79599	2
5.	3.70353	-1.08835	2
6.	-15.2112	3.82599	4
7.	1.14713	-8.49527	2
8.	5.30677	0.443736	2
9.	6.88024	-4.03278	2

79 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)



80 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Grafinis vaizdavimas (MII klasteris)

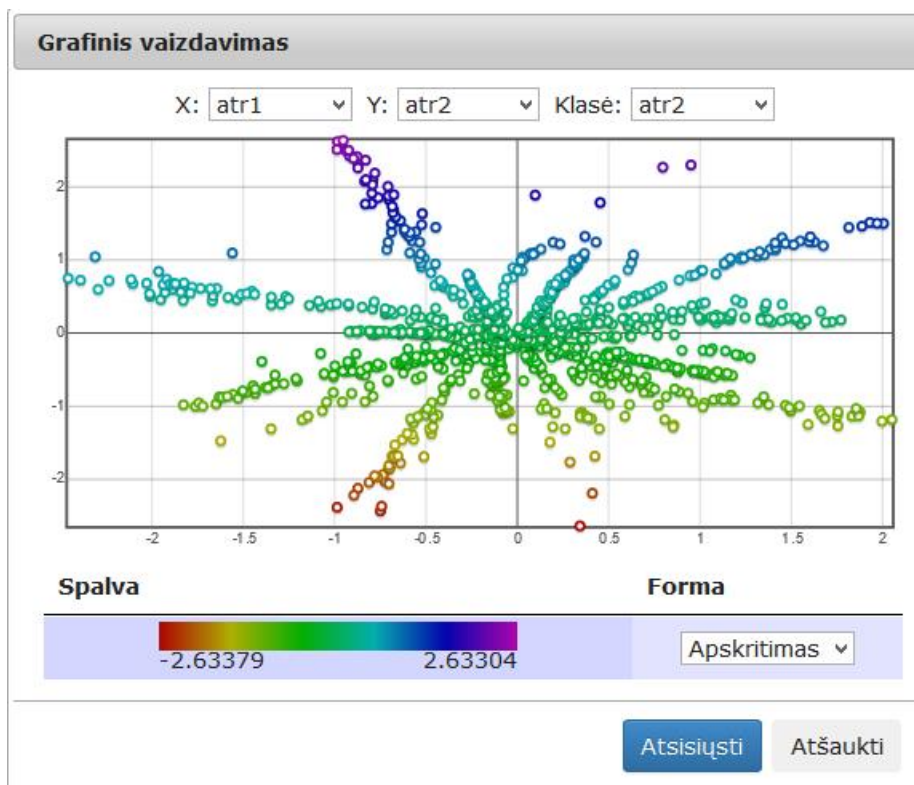
Analogiški komponenčių testavimai atlikti MIF VU superkompiuteryje. Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 81 pav., 82 pav. ir 83 pav. Rezultatai pateikiami atlikus eksperimentą su *Elipsoidų* duomenimis.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.11118
Skaičiavimo laikas	0.12
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/fa248b4c1609ce7272199b9798">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/fa248b4c1609ce7272199b9798</a>

81 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas		
Objekto nr.	atr1 (REAL)	atr2 (REAL)
1.	-0.182223	1.09719
2.	0.0678368	-0.629694
3.	0.191122	0.414405
4.	-0.0431635	1.69247
5.	0.155268	0.0305378
6.	-0.1638	-1.0684
7.	-1.46198	0.270953
8.	-1.05678	0.646354
9.	-0.0586208	-1.94607

82 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Matricinis vaizdavimas (MIF VU superkompiuteris)



83 pav. Komponentės „SMACOF (MDS)“ testavimo rezultatas: Grafinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.5.3 Komponentė „DMA“

Diagonalinio mažoravimo algoritmas (DMA) yra SMACOF algoritmo modifikacija, kurioje naudojama paprastesnė mažoravimo funkcija. SMACOF – tai vienas geriausių optimizavimo algoritmų, taikomų daugiamačių skalių paklaidos minimizavimui. Modifikacija DMA skirta atvaizduoti didesnėms duomenų aibėms.

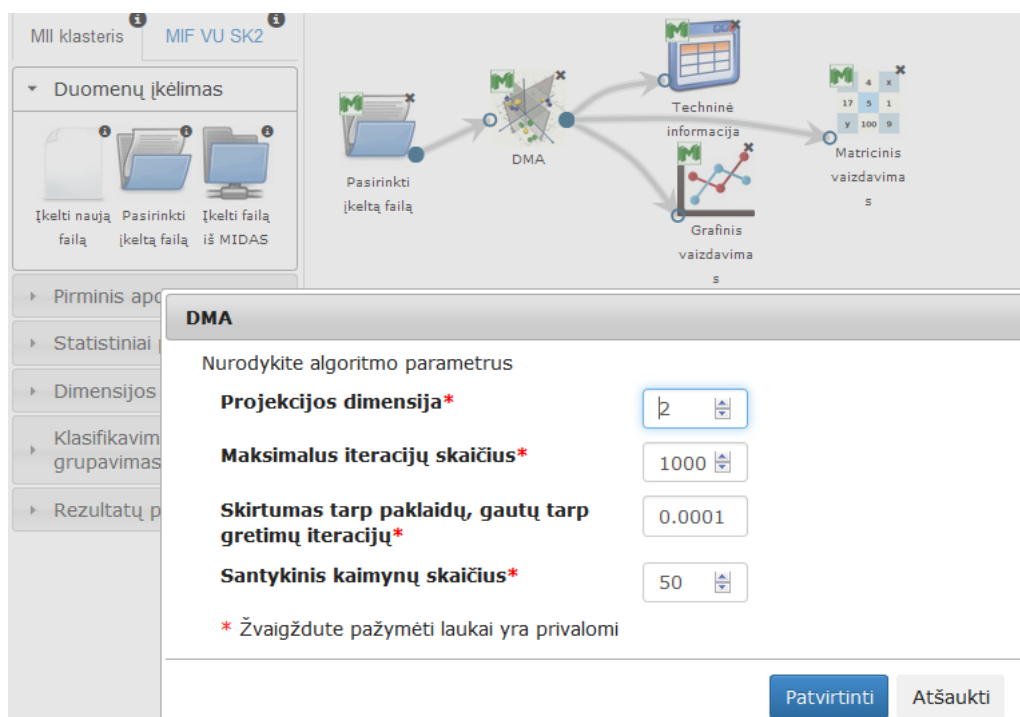
Komponentės „DMA“ **rezultatas**: nauja duomenų matrica, apskaičiuota diagonalinio mažoravimo dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2.

Parametrai:

- **Projekcijos dimensija** – nurodoma sumažinta pradinių duomenų dimensija (projekcijos dimensija).
- **Maksimalus iteracijų skaičius** – maksimalus iteracijų skaičius. Maksimalus iteracijų skaičius yra 1000.
- **Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų** – skirtumas tarp paklaidų, gautų tarp gretimų iteracijų. Skirtumas tarp paklaidų intervale nustatomas iš intervalo  $[10e-8; \infty)$ .
- **Santykinis kaimynų skaičius** – nurodomas kaimynų skaičius.

Į darbalaukį įkėlus komponentę „DMA“, sujungus ją su duomenų failų įkėlimo komponente ir paspaudus du kartus komponentę „DMA“ atidaromas langas. Naudotojas turi pasirinkti ir įvesti reikalingus paramentrus. Numatytosios reikšmės: projekcijos dimensija – 2, maksimalus iteracijų skaičius – 100, skirtumas tarp paklaidų, gautų tarp gretimų iteracijų – 0,0001, santykinis kaimynų

skaičius – 1%. DMA langas ir algoritmo vykdymo eksperimento darbų seka (MII klasteris) pavaizduoti 84 pav.



84 pav. DMA langas ir DMA algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės, parametrų reikšmės negali būti neigiamos, jei įvesta neigiama reikšmė rodomas klaidos pranešimas ties neteisingai įvestu lauku. Įvestos projekcijos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius. Dimensijos parametras yra sveikas skaičius. Maksimalus iteracijų skaičius [1; 1000] (sveikas skaičius). Skirtumas tarp paklaidų intervale  $[10^{-8}; \infty)$ , Santykinis kaimynų skaičius intervale (0; 100] %.

Naudotojui paspaudus mygtuką „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „DMA“, „Specify the parameters of the algorithm.“, „Projection space“, „Maximum number of iteration“, „Minimal stress change“, „Relative number of neighbours“, „Fields marked with \* are mandatory“. Klaidos pranešimai: „Value cannot be negative“, „Relative neighbor quantity must be in interval (0; 100] %“, „Projection space cannot be real value“, „Projection space cannot be greater than quantity of attributes in arff file“, „Minimal stress change must be in interval  $[10^{-8}; \infty)$ “. Mygtukai: „Ok“ ir „Cancel“.

84 pav. pateikiama bendra komponentės DMA testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, DMA komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Skaičiavimai MII klasteryje. Naudota testinė *Iris* duomenų aibė.

Vykdam eksperimentą, DMA komponentės parametrų lange buvo nurodyta, kad DMA projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; skirtumas tarp paklaidų – 0,0001; Santykinis kaimynų skaičius – 10.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 85 pav., 86 pav., 87 pav.

Techninė informacija	
Parametras	Reikšmė
Klaidos	0.857104
Skaičiavimo laikas	0.03
Rezultatų failas	http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/6bafb127a19d3457b80f0221e6

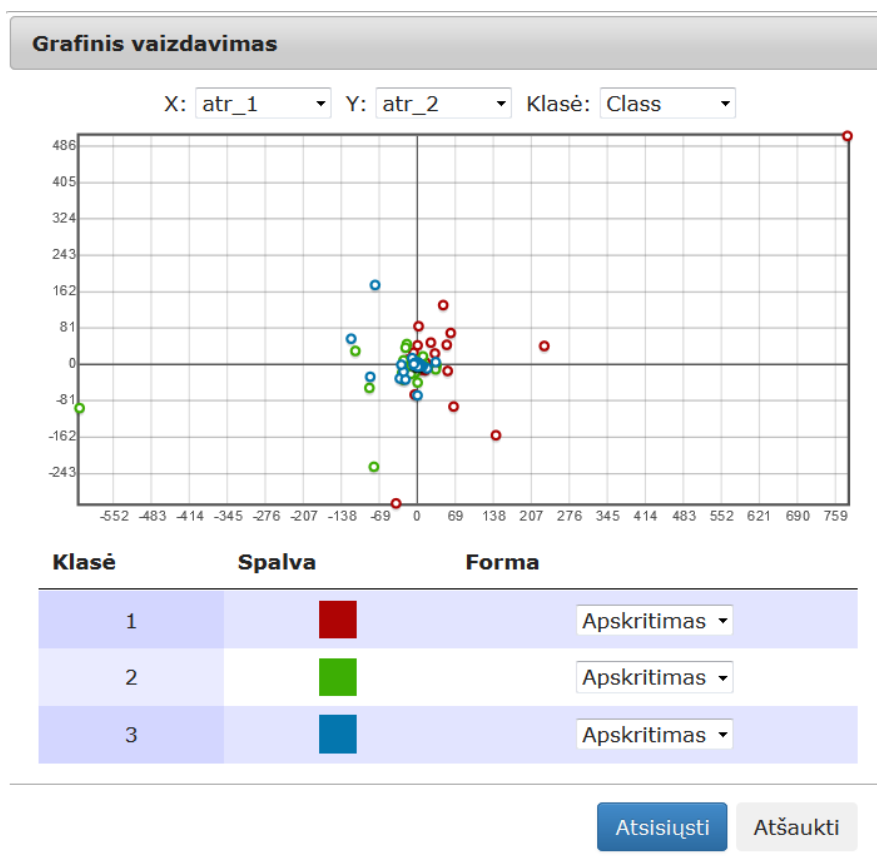
Atsisiųsti Atšaukti

85 pav. Komponentės „DMA“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas			
Objekto nr.	atr_1 (REAL)	atr_2 (REAL)	CLASS (1, 2, 3)
1.	-38.5451	-309.029	1
2.	3.20198	0.632819	1
3.	-17.243	10.618	1
4.	4.5626	-7.11674	1
5.	0.85297	4.29358	1
6.	780.921	508.789	1
7.	-6.62544	26.5496	1
8.	-3.2043	-2.24668	1
9.	-2.81713	-0.587786	1

Atsisiųsti Atšaukti

86 pav. Komponentės „DMA“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)



87 pav. Komponentės „DMA“ testavimo rezultatas: Grafinis vaizdavimas (MII klasteris)

### 3.6.5.4 Komponentė „Relative MDS“

Santykinų daugiamačių skalių algoritmas (angl. Relative MDS) skirtas didelių aibių bei naujų taškų priklausančių daugiamatei erdvei vizualizavimui, naudojant prieš tai apskaičiuotą bazinių taškų projekciją.

Naudojant klasikinį daugiamačių skalių metodą, negalima atidėti naujo taško neperskaičiuojant visos turimos duomenų aibės projekcijos. Todėl naujų taškų atvaizdavimui gali būti naudojamas santykinų daugiamačių skalių algoritmas (SDS). Nors šis metodas nėra toks tikslus kaip SMACOF, tačiau jis gali atvaizduoti dideles aibes, tam pareikalaudamas mažai kompiuterio skaičiavimo resursų.

Komponentės „Relative MDS“ **rezultatas**: nauja duomenų matrica, apskaičiuota santykinų daugiamačių skaičių dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2.

Parametrai:

- **Projekcijos dimensija** – nurodoma sumažinta pradinė duomenų dimensija (projekcijos dimensija).
- **Maksimalus iteracijų skaičius** – maksimalus iteracijų skaičius. Maksimalus iteracijų skaičius yra 1000.

- **Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų** – Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų. Skirtumas tarp paklaidų intervale nustatomas iš intervalo  $[10e-8; \infty)$ .
- **Bazinių objektų skaičius** – bazinių objektų skaičiaus nustatymas.
- **Bazinių objektų parinkimas (Atsitiktinis | Pagal PCA (dimensija 1) | Pagal didžiausių požymių dispersiją)** – bazinių objektų parinkimo strategija.

Į darbalaukį įkėlus komponentę „Relative (MDS)“, sujungus ją su duomenų failų įkėlimo komponente ir paspaudus du kartus komponentę „Relative (MDS)“ atidaromas langas. Naudotojas turi pasirinkti ir įvesti reikalingus paramentrus. Numatytosios reikšmės: projekcijos dimensija – 2, maksimalus iteracijų skaičius – 100, skirtumas tarp paklaidų, gautų tarp gretimų iteracijų – 0,0001, bazinis objektų skaičius – 1 %, bazinių objektų parinkimas – atsitiktinis. Modalinis Relative (MDS) langas ir algoritmo vykdymo eksperimento darbų seka (MII klasteris) pavaizduoti 88 pav.

The screenshot displays the MII klasteris interface. At the top, it shows 'MII klasteris' and 'MIF VU SK2'. The main area is divided into 'Duomenų įkėlimas' (Data Loading) and 'Pirminis apdorojimas' (Initial Processing). The workflow consists of several steps: 'Pasirinkti įkeltą failą' (Select loaded file), 'Relative MDS', 'Techninė informacija' (Technical information), 'Matricinis vaizdavimas' (Matrix visualization), and 'Grafinis vaizdavimas' (Graphical visualization). A configuration window for 'Relative MDS' is open, showing the following parameters:

- Projekcijos dimensija\***: 2
- Maksimalus iteracijų skaičius\***: 100
- Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų\***: 0.0001
- Bazinis objektų skaičius\***: 50 %
- Bazinių objektų parinkimas\***: Pagal didžiausių požymių disp

A note at the bottom of the window states: '\* Žvaigždute pažymėti laukai yra privalomi' (Fields marked with an asterisk are mandatory). Buttons for 'Patvirtinti' (Confirm) and 'Atšaukti' (Cancel) are visible at the bottom right.

88 pav. Relative MDS langas ir Relative MDS algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Yra trys bazinių objektų parinkimo strategijos būdai: atsitiktinis, pagal PCA (dimensija 1), pagal didžiausių požymių dispersiją. Strategijų parinkimo langas pavaizduotas 89 pav.

**Relative MDS**

Nurodykite algoritmo parametrus

**Projekcijos dimensija\***

**Maksimalus iteracijų skaičius\***

**Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų\***

**Bazinis objektų skaičius\***  %

**Bazinių objektų parinkimas\*** Pagal didžiausią požymių disp

Atsitiktinis  
 Pagal PCA(dimensija 1)  
Pagal didžiausią požymių dispersiją

\* Žvaigždute pažymėti laukai yra privalomi

**89 pav. Bazinių objektų parinkimo strategijos**

Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės, parametrų reikšmės negali būti neigiamos, jei įvesta neigiama reikšmė rodomas klaidos pranešimas ties neteisingai įvestu lauku. Įvestos projekcijos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius. Dimensijos parametras yra sveikas skaičius. Maksimalus iteracijų skaičius intervale [1; 1000] (sveikas skaičius). Skirtumas tarp paklaidų intervale  $[10^{-8}; \infty)$ , santykinis bazių objektų skaičius intervale (0; 100] %.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „Relative MDS“, “Specify the parameters of the algorithm.” “Projection space”, “Maximum number of iteration”, “Minimal stress change”, “Relative number of basis objects”, “Select Basis objects strategy – Random, By line based on PCA, By line based on max variable”, „Fields marked with \* are mandatory“. Klaidos pranešimai: „Value cannot be negative“, „Relative basis object quantity must be in interval (0; 100] %“, “Projection space cannot be real value”, “Projection space cannot be greater than quantity of attributes in arff file”, “Minimal stress change must be in interval  $[10^{-8}; \infty)$ “. Mygtukai: „Ok“ ir „Cancel“.

Vykdamas eksperimentą, komponentės „Relative MDS“ parametrų lange buvo nurodyta, kad Relative MDS projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; skirtumas tarp paklaidų – 0,0001; bazinis objektų skaičius – 50; bazinių objektų parinkimas – Pagal didžiausią požymių dispersiją. Ekasperimente naudota *Iris* duomenų aibė.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 90 pav., 91 pav., 92 pav.

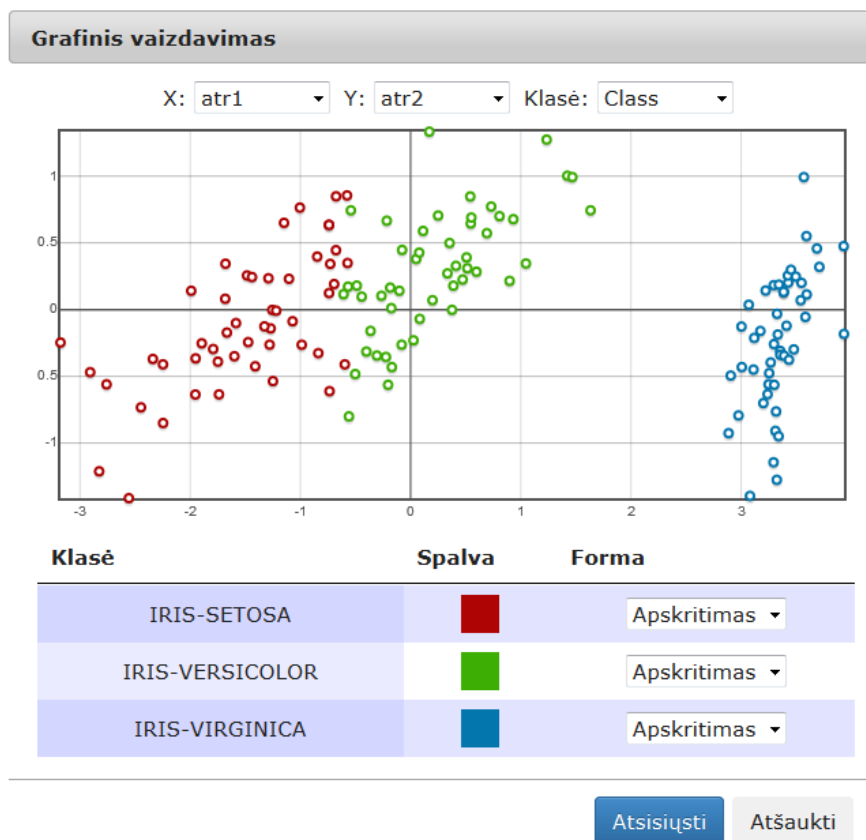


Techninė informacija	
Parametras	Reikšmė
Paklaida	0.207707
Skaičiavimo laikas	0.03
Rezultatų failas	http://damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/9d43dd6c896d5fb07eb35ff42f7

90 pav. Komponentės „Relative MDS“ testavimo rezultatas: Techninė informacija (MII klasteris)

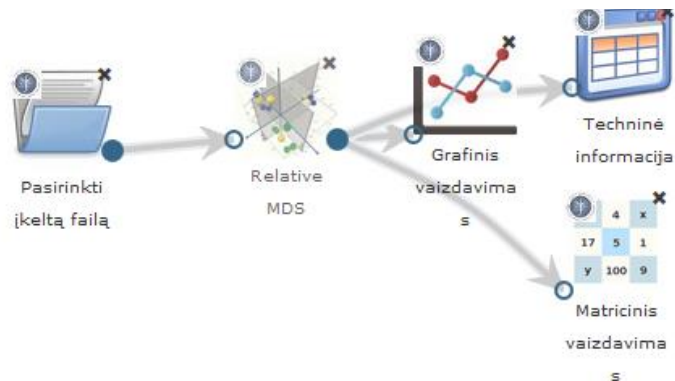
Matricinis vaizdavimas			
Objekto nr.	atr1 (REAL)	atr2 (REAL)	CLASS (IRIS-SETOSA, IRIS-VERSICOLOR, IRIS-VIRGINICA)
1.	-3.17626	-0.245214	IRIS-SETOSA
2.	-2.90723	-0.467489	IRIS-SETOSA
3.	-2.82545	-1.2106	IRIS-SETOSA
4.	-2.7588	-0.558484	IRIS-SETOSA
5.	-2.55595	-1.41177	IRIS-SETOSA
6.	-2.44631	-0.731039	IRIS-SETOSA
7.	-2.33842	-0.368442	IRIS-SETOSA
8.	-2.24537	-0.840713	IRIS-SETOSA

91 pav. Komponentės „Relative MDS“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)



92 pav. Komponentės „Relative MDS“ testavimo rezultatas: Grafinis vaizdavimas (MII klasteris)

93 pav. pateikiama bendra komponentės „Relative MDS“ testavimo darbų seka (kartu su kitais dimensijos mažinimo algoritmais), sudaryta iš failo įkėlimo komponentės, komponentės „Relative MDS“ ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Skaičiavimai vykdyti MIF VU superkompiuteryje, pateikiami testavimo rezultatai, gauti naudojant *Elipsoidų* duomenų aibę.



93 pav. Dimensijos mažinimo algoritmų testavimas: Relative MDS (MIF VU superkompiuteris)

Vykdamas eksperimentą, komponentės „Relative MDS“ parametrų lange buvo nurodyta, kad Relative MDS projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; skirtumas tarp paklaidų – 0,0001; bazinis objektų skaičius – 20; bazinių objektų parinkimas – atsitiktinis.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 94 pav.,

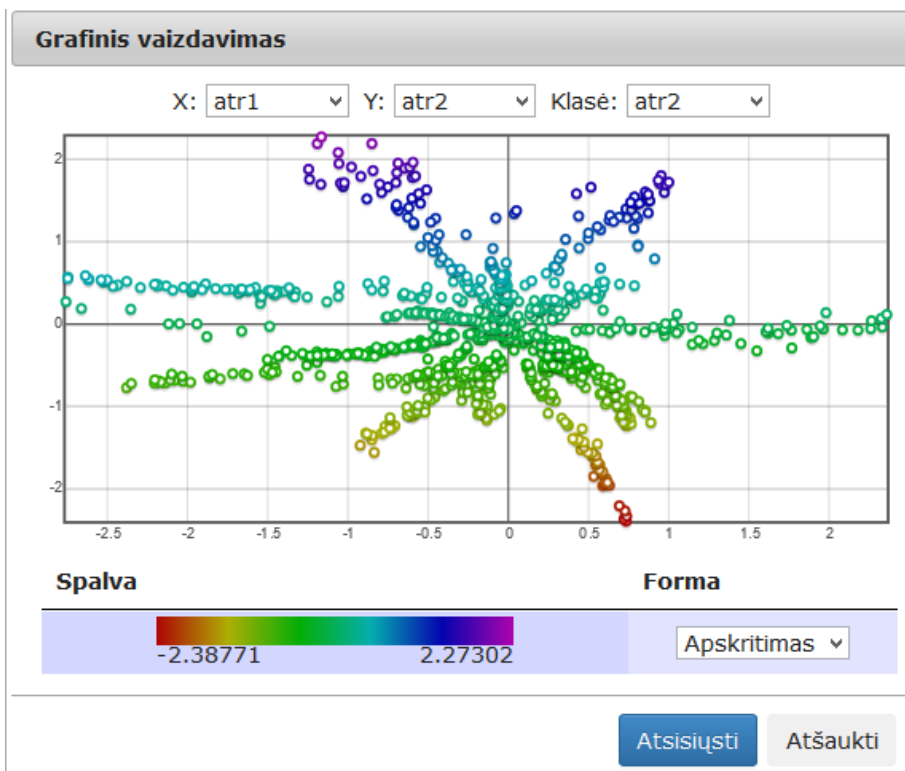
95 pav. ir 96 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.409327
Skaičiavimo laikas	0.12
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/0eca5469167fdbcad0521c39a">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/0eca5469167fdbcad0521c39a</a>

94 pav. Komponentės „Relative MDS“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas		
Objekto nr.	atr1 (REAL)	atr2 (REAL)
1.	-0.835913	-1.5548
2.	1.98041	0.141075
3.	0.035297	-0.315323
4.	-0.304367	0.0919247
5.	-1.0559	0.458805
6.	0.248146	-1.08023
7.	0.153984	0.0883811
8.	-0.258937	-0.267779
9.	-0.919659	1.79352

95 pav. Komponentės „Relative MDS“ testavimo rezultatas: Matricinis vaizdavimas (MIF VU superkompiuteris)



96 pav. Komponentės „Relative MDS“ testavimo rezultatas: Grafinis vaizdavimas (MIF VU superkompiuteris)

### 3.6.5.5 Komponentė „SAMANN“

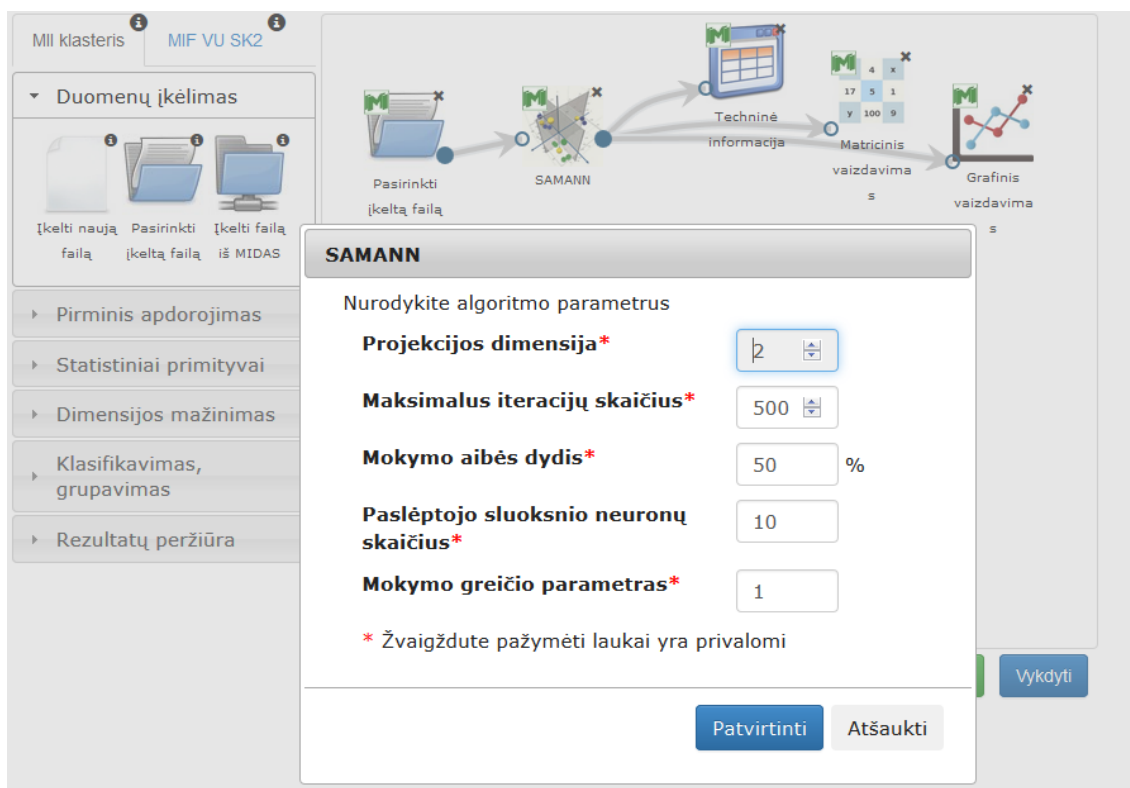
Pastaruoju metu, buvo pasiūlyta dirbtinių neuroninių tinklų algoritmų daugiamačių duomenų vizualizavimui. Pavyzdžiui, pasiūlyta specifinė „klaidos skleidimo atgal“ mokymo taisyklė, pavadinta SAMANN, kuri leidžia įprastam tiesioginio skleidimo neuroniniam tinklui realizuoti Sammono projekciją mokymo be mokytojo būdu. Sammono projekcija yra netiesinis daugelio kintamųjų objektų atvaizdavimo žemesnio matavimo erdvėje metodas. Jo idėja – atvaizduoti daugiamačius vektorius mažesnio matavimo erdvėje išlaikant santykinai panašius atstumus tarp vektorių.

Komponentės „SAMANN“ **rezultatas**: nauja duomenų matrica, apskaičiuota SAMANN dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2.

Parametrai:

- **Projekcijos dimensija** – nurodoma sumažinta pradinė duomenų dimensija (projekcijos dimensija).
- **Maksimalus iteracijų skaičius** – maksimalus iteracijų skaičius. Maksimalus iteracijų skaičius yra 1000.
- **Apmokymo aibės dydis** – SAMANN neuroninio tinklo apmokymui naudojamų pradinės aibės elementų skaičius.
- **Paslėptojo sluoksnio neuronų skaičius** – SAMANN neuroninio tinklo paslėptojo sluoksnio neuronų skaičius.
- **Mokymo greičio parametras** – mokymosi greičio parametro reikšmė (iš intervalo 0,1–10).

Į darbalaukį įkėlus komponentę „SAMANN“ sujungus ją su duomenų failų įkėlimo komponente ir paspaudus du kartus komponentę „SAMANN“ atidaromas langas. Naudotojas turi pasirinkti ir įvesti reikalingus parametrus. Numatytosios reikšmės: projekcijos dimensija – 2 (naudotojas jos keisti negali), maksimalus iteracijų skaičius – 100, apmokymų aibės dydis – 10 % (gali būti realus skaičius), paslėptojo sluoksnio neuronų skaičius – 10, mokymų greičio parametras – 1 (gali būti realus skaičius). SAMANN langas ir algoritmo vykdymo eksperimento darbų seka (MII klasteris) pavaizduoti 97 pav.



97 pav. SAMANN langas ir SAMANN algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės, parametrų reikšmės negali būti neigiamos, jei įvesta neigiama reikšmė rodomas klaidos pranešimas ties neteisingai įvestu lauku. Įvestos projekcijos dimensijos reikšmė negali būti didesnė nei arff faile esančių požymių skaičius. Dimensijos parametras yra sveikas skaičius. Maksimalus iteracijų skaičius [1; 1000] (sveikas skaičius). Apmokymo aibės dydis intervale (0; 100] %. Paslėpto sluoksnio neuronų skaičius ir mokymo greičio parametras – teigiamas reikšmės.

Naudotojui paspaudus mygtuką „Atšaukti“ langas uždaromas (lieka galioti paskutinės išsaugotos reikšmės jei tokios buvo, jei ne – numatytosios), grįžtama į eksperimento planavimą.

Naudojami pavadinimai anglų kalba: „SAMANN“, “Specify the parameters of the algorithm.”. “Projection space”, “Maximum number of iteration”, “Relative size of the training data”, “Number of neurons in the hidden layer”, “Value of the learning rate”, „Fields marked with \* are mandatory“. Klaidos pranešimai: „Value cannot be negative“, „Relative size of the training data must be in interval (0; 100] %”, “Projection space cannot be real value”, “Projection space cannot be greater than quantity of attributes in arff file”, “Minimal stress change must be in interval  $[10^{-8}; \infty)$ “. Mygtukai: „Ok“ ir „Cancel“.

97 pav. pateikiama bendra komponentės SAMANN testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, SAMANN komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Skaičiavimai atlikti MII klasteryje.

Vykdamas eksperimentą, SAMANN komponentės parametrų lange buvo nurodyta, kad SAMANN projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; apmokymo aibės dydis – 50 %; paslėptojo sluoksnio neuronų skaičius – 10; mokymo greičio parametras – 1. Skaičiavimams naudota *Iris* duomenų aibė.

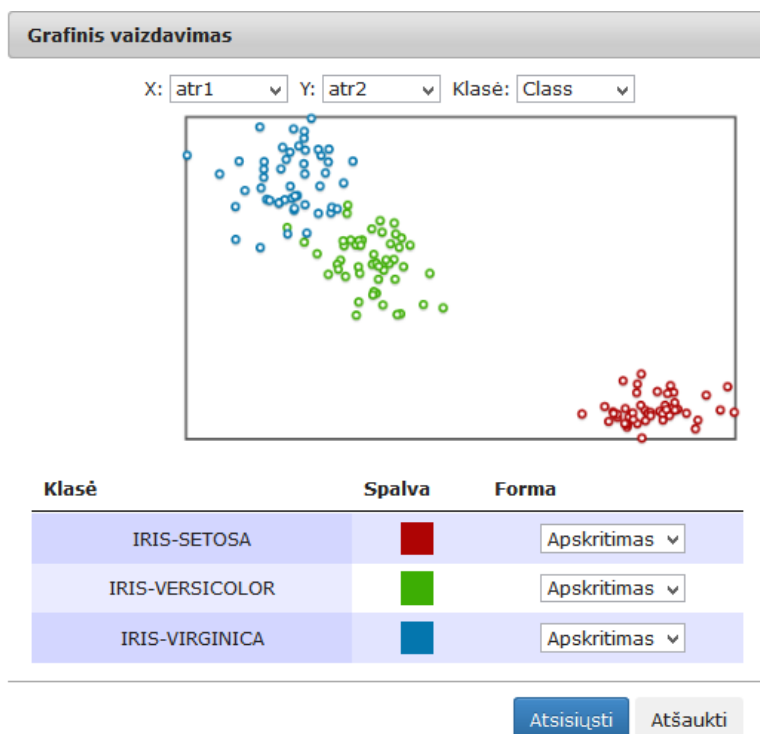
Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 98 pav., 99 pav., 100 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.907719
Skaičiavimo laikas	0.03
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/3fb35145f0131762320fc311d3">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/3fb35145f0131762320fc311d3</a>

98 pav. Komponentės „SAMANN“ testavimo rezultatas: Techninė informacija (MII klasteris)

Matricinis vaizdavimas			
Objekto nr.	atr1 (REAL)	atr2 (REAL)	CLASS (IRIS-SETOSA, IRIS-VERSICOLOR, IRIS-VIRGINICA)
1.	0.84958	0.292879	IRIS-SETOSA
2.	0.846943	0.292646	IRIS-SETOSA
3.	0.848112	0.292275	IRIS-SETOSA
4.	0.846121	0.292598	IRIS-SETOSA
5.	0.849827	0.292792	IRIS-SETOSA
6.	0.8499	0.295714	IRIS-SETOSA
7.	0.848086	0.293315	IRIS-SETOSA
8.	0.848275	0.292996	IRIS-SETOSA

99 pav. Komponentės „SAMANN“ testavimo rezultatas: Matricinis vaizdavimas (MII klasteris)



100 pav. Komponentės „SAMANN“ testavimo rezultatas: Grafinis vaizdavimas (MII klasteris)

Vykdamas eksperimentą MIF VU superkompiuteryje (eksperimento darbų seka analogiška pateiktai 97 pav., tik naudojamos MIF VU superkompiuteriui skirtos komponentės), SAMANN komponentės parametrų lange buvo nurodyta, kad SAMANN projekcijos dimensija yra 2; maksimalus iteracijų skaičius yra 500; apmokymo aibės dydis – 10 %; paslėptojo sluoksnio neuronų skaičius – 10; mokymo greičio parametras – 1. Pateikiami testavimo rezultatai, gauti naudojant *Elipsoidų* duomenų aibę. Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 101 pav., 102 pav. ir 103 pav.

**Techninė informacija**

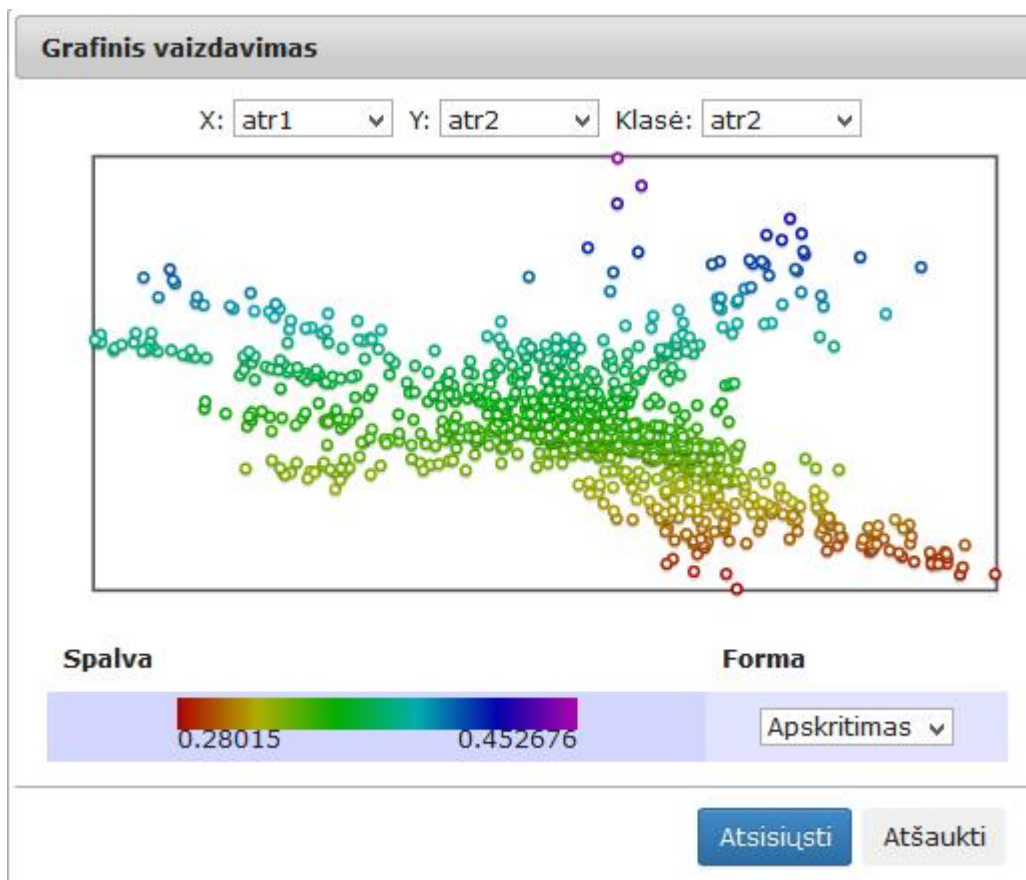
Parametras	Reikšmė
Paklaida	0.808393
Skaičiavimo laikas	0.1
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/bcc937f55c788b8804fee4c129">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07/experiment/bcc937f55c788b8804fee4c129</a>

Atsisiųsti Atšaukti

101 pav. Komponentės „SAMANN“ testavimo rezultatas: Techninė informacija (MIF VU superkompiuteris)

Matricinis vaizdavimas		
Objekto nr.	atr1 (REAL)	atr2 (REAL)
1.	0.654843	0.375694
2.	0.696664	0.347031
3.	0.708825	0.328652
4.	0.707912	0.326516
5.	0.705654	0.367254
6.	0.698196	0.359086
7.	0.72266	0.302502
8.	0.67187	0.326882
9.	0.704647	0.307895

102 pav. Komponentės „SAMANN“ testavimo rezultatas: Matricinis vaizdavimas (MIF VU superkompiuteris)



103 pav. Komponentės „SAMANN“ testavimo rezultatas: Grafinis vaizdavimas (MIF VU superkompiuteris)

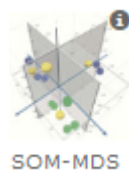


### 3.6.5.6 Komponentė „SOM–MDS“

SOM žemėlapiai naudojami ir daugiamačiams duomenims klasterizuoti ir juos vizualizuoti, t. y. rasti projekcijas mažesnės dimensijos erdvėje, įprastai plokštumoje. SOM tinklo tikslas – išlaikyti duomenų kaimyniškumus, t. y. taškai, esantys arti įėjimo vektorių erdvėje, turi būti atvaizduojami arti vieni kitų ir SOM žemėlapyje. Kartais gautus rezultatus sudėtinga interpretuoti, todėl jie papildomai analizuojami vienu iš daugiamačių duomenų projekcijos metodu. Tuo tikslu gali būti naudojamas daugiamačių skalių metodas (MDS). Vienas iš nuosekliojo junginio tikslų – pagerinti duomenų vizualizavimą, panaudojant saviorganizuojančius neuroninius tinklus. Tačiau pagrindinis nuosekliojo junginio tikslas – sumažinti skaičiavimo laiką, neprarandant vizualizavimo kokybės, atvaizduojant neuronus–nugalėtojus atitinkančius vektorius, gautus taikant SOM ir juos vizualizuojant MDS metodu, lyginant su visos duomenų aibės vizualizavimo laiku, taikant tik MDS metodą.

Komponentės „SOM-MDS“ **rezultatas**: nauja duomenų matrica, apskaičiuota SOM-MDS dimensijos mažinimo metodu iš pradinės duomenų aibės, sumažinus požymių skaičių iki 2. Vizualizuojami SOM tinklu gauti neuronai nugalėtojai.

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenų dimensiją norima sumažinti SOM-MDS algoritmu, ir paspaudus du kartus komponentę „SOM-MDS“



atidaromas langas (104 pav.).

Numatytosios reikšmės: SOM eilučių skaičius = 10, SOM stulpelių skaičius = 10, SOM mokymų epochų skaičius = 100, MDS iteracijų skaičius = 100, skirtumas tarp paklaidų, gautų tarp gretimų iteracijų = 0,0001, MDS projekcija = 2 (naudotojas keisti negali).

Naudotojas gali šias reikšmes pakeisti. Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės. Šios reikšmės negali būti neigiamos. SOM eilučių ir stulpelių skaičius turi būti intervale [3; 100] (sveikas skaičius). Epochų skaičius turi būti intervale [1; 1000] (sveikas skaičius). MDS iteracijų skaičius turi būti intervale [1; 1000] (sveikas skaičius). Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų, turi būti intervale [ $10^{-8}$ ;  $\infty$ ). Jei įvestos netinkamos reikšmės, rodomas klaidos pranešimas ties neteisingai įvestu lauku.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas ir lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios, ir grįžtama į eksperimento planavimą.

**SOM-MDS**

Nurodykite algoritmo parametrus

**SOM eilučių skaičius\***

**SOM stulpelių skaičius\***

**SOM mokymų epochų skaičius\***

**MDS projekcijos dimensija\***

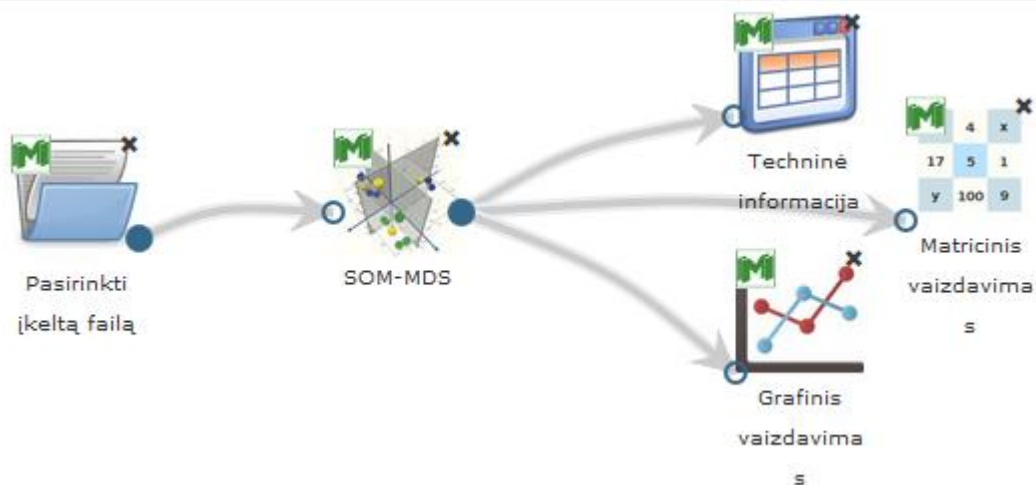
**MDS iteracijų skaičius\***

**Skirtumas tarp paklaidų, gautų tarp gretimų iteracijų\***

\* Žvaigždute pažymėti laukai yra privalomi

104 pav. Langas „SOM-MDS“

105 pav. pateikiama bendra komponentės SOM-MDS testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, SOM-MDS komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas). Naudojama *iris* testinė duomenų aibė.



105 pav. SOM-MDS algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Vykdamas eksperimentą, SOM-MDS komponentės parametrų lange buvo nurodyta, kad SOM eilučių skaičius – 10; SOM stulpelių skaičius – 10; SOM mokymo epochų skaičius – 100; MDS projekcijos dimensija – 2; MDS iteracijų skaičius – 100; Skirtumas tarp paklaidų – 0,0001.

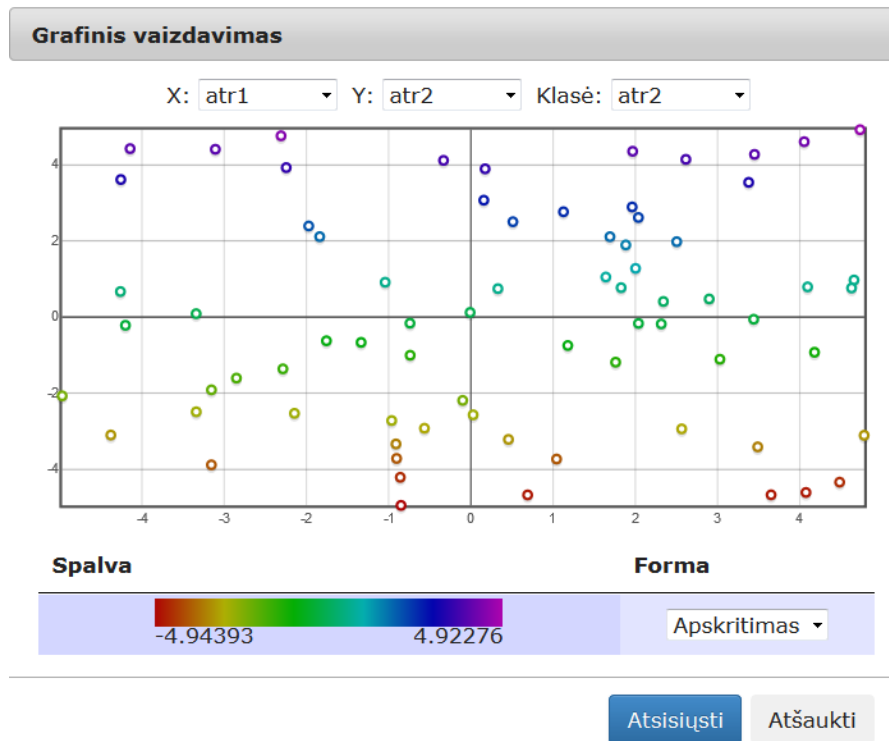
Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 106 pav., 107 pav., 108 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	2.13767
Skaičiavimo laikas	0.34
Rezultatų failas	<a href="http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/88209db2882f24aaa2b21341a6">http://dev.damis.lt/assets/b3dc68655f2eee81930b90a07c/experiment/88209db2882f24aaa2b21341a6</a>

106 pav. Komponentės „SOM-MDS“ testavimo rezultatas: Techninė informacija

Matricinis vaizdavimas		
Objekto nr.	atr1 (REAL)	atr2 (REAL)
1.	4.48485	-4.33355
2.	3.43952	-0.0550341
3.	0.0279574	-2.56564
4.	2.5036	1.9818
5.	-3.15508	-3.87859
6.	-0.00907091	0.123804
7.	4.07511	-4.60473
8.	1.82737	0.774482
9.	3.02878	-1.10559

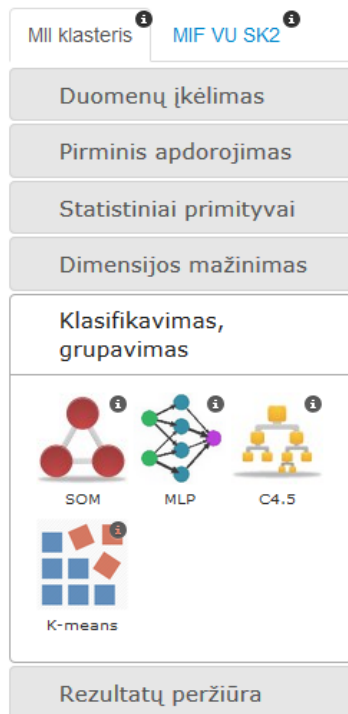
107 pav. Komponentės „SOM-MDS“ testavimo rezultatas: Matricinis vaizdavimas



108 pav. Komponentės „SOM-MDS“ testavimo rezultatas: Grafinis vaizdavimas

### 3.6.6 Kairiojo meniu skiltis „Klasifikavimas ir grupavimas“

Pasirinkus klasifikavimo ir grupavimo skiltį, atsidaro realizuotų klasifikavimo ir grupavimo algoritmų komponentės. „SOM“ – tai saviorganizuojantį neuroninį tinklą realizuojanti komponentė, „MLP“ – tai daugiasluksnį perceptroną, pritaikytą klasifikavimo uždaviniams spręsti, realizuojanti komponentė, „RDF“ – tai RDF klasifikavimo algoritmą realizuojanti komponentė, „K-means“ – tai grupavimo algoritmą k-means (k-vidurkiai) realizuojanti komponentė. Meniu skiltis pateikta 109 pav.



109 pav. Meniu skiltis „Klasifikavimas, grupavimas“

Toliau detalai bus aprašytos visos „Klasifikavimo, grupavimo“ skilties komponentės ir pateikiamos jų naudojimosi instrukcijos.

### 3.6.6.1 Komponentė „SOM“

SOM žemėlapiai naudojami daugiamačiams duomenims vizualizuoti (t. y. rasti projekcijas mažesnės dimensijos erdvėje, įprastai plokštumoje) ir klasterizuoti. SOM tinklo tikslas – išlaikyti duomenų kaimyniškumus, t. y. taškai, esantys arti įėjimo vektorių erdvėje, turi būti atvaizduojami arti vieni kitų ir SOM žemėlapyje.

Komponentės „SOM“ **rezultatas**: nauja duomenų matrica apskaičiuota SOM metodu, ir sudaryta iš duomenų matricos ir priskirtų kiekvienam duomenų įrašui neurono nugalėtojo indekso.

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenų dimensiją norima sumažinti SOM algoritmu, ir paspaudus du kartus komponentę „SOM“



atsidaro langas (110 pav.).

Numatytosios reikšmės: eilučių skaičius = 10, stulpelių skaičius = 10, mokymų epochų skaičius = 100. Naudotojas gali šias reikšmes pakeisti. Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės. Šios reikšmės negali būti neigiamos. Eilučių ir stulpelių skaičius turi būti intervale [3; 100] (sveikas skaičius). Epochų skaičius turi būti intervale [1; 1000] (sveikas skaičius). Jei įvestos netinkamos reikšmės, rodomas klaidos pranešimas ties neteisingai įvestu lauku.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas ir lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios, ir grįžtama į eksperimento planavimą.

**SOM**

Nurodykite algoritmo parametrus

**Eilučių skaičius\***

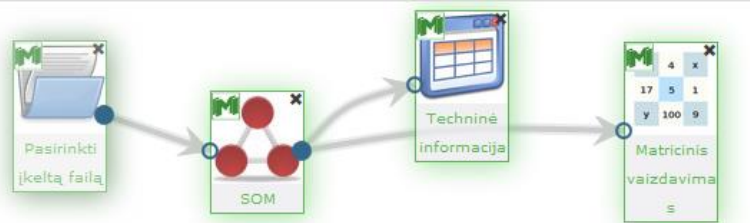
**Stulpelių skaičius\***

**Mokymų epochų skaičius\***

\* Žvaigždute pažymėti laukai yra privalomi

110 pav. Langas „SOM“

111 pav. pateikiama bendra SOM komponentės testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, SOM komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas).



111 pav. SOM algoritmo vykdymo eksperimento darbų seka (MII klasteris)

Vykdamas eksperimentą SOM komponentės parametrų lange buvo nurodyti tokie parametrai: eilučių skaičius – 10; stulpelių skaičius – 10; mokymo epochų skaičius – 100.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 112 pav., 113 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	15.5037
Skaičiavimo laikas	0.03
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/1207ad4d619eda85774d55f8c5">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/1207ad4d619eda85774d55f8c5</a>

112 pav. Komponentės „SOM“ testavimo rezultatas: Techninė informacija

Matricinis vaizdavimas				
Objekto nr.	atr1 (REAL)	atr2 (REAL)	atr3 (REAL)	atr4 (REAL)
1.	6.15753	2.94008	5.11887	1.94337
2.	5.20244	3.08348	2.41782	0.621929
3.	5.66461	2.8608	3.7643	1.17093
4.	6.48075	3.02316	5.37913	2.06596
5.	6.46917	3.02621	5.37002	2.07137
6.	5.97559	2.75347	4.93516	1.73619
7.	6.42409	3.03387	5.3356	2.08261

113 pav. Komponentės „SOM“ testavimo rezultatas: Matricinis vaizdavimas

### 3.6.6.2 Komponentė „k-means“

K-vidurkių metodas yra grupavimo algoritmas, skirtas suskirstyti duomenų aibę į kompaktiškas grupes, esančias kuo toliau viena nuo kitos. Paprastai nurodomas parametras k, pasakantis į kiek grupių reikia padalinti vektorių aibę.

Komponentės „K-Means“ **rezultatas**: nauja duomenų matrica, kurioje kiekvienam pradinės duomenų aibės vektoriui priskirta klasė, t. y. vektoriai suskirstomi į grupes, kurių skaičius

nustatomas komponentės parametru lango.

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenis norima klasterizuoti k-vidurkių (k-means) algoritmu, ir paspaudus du kartus komponentę „K-MEANS“



atsiranda langas (114 pav.).

Numatytosios reikšmės: maksimalus iteracijų skaičius = 100, klasterių skaičius = 10. Naudotojas gali šias reikšmes pakeisti. Naudotojui paspaudus mygtuką „Patvirtinti“, tikrinamos įvestos parametru reikšmės. Šios reikšmės turi būti sveikieji teigiami skaičiai. Maksimalus iteracijų skaičius turi būti intervale [1; 1000]. Klasterių skaičius turi būti intervale [1; 100]. Jei įvestos netinkamos reikšmės, rodomas klaidos pranešimas ties neteisingai įvestu lauku.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas ir lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios, ir grįžtama į eksperimento planavimą.

**K-MEANS**

Nurodykite algoritmo parametrus

**Maksimalus iteracijų skaičius\***

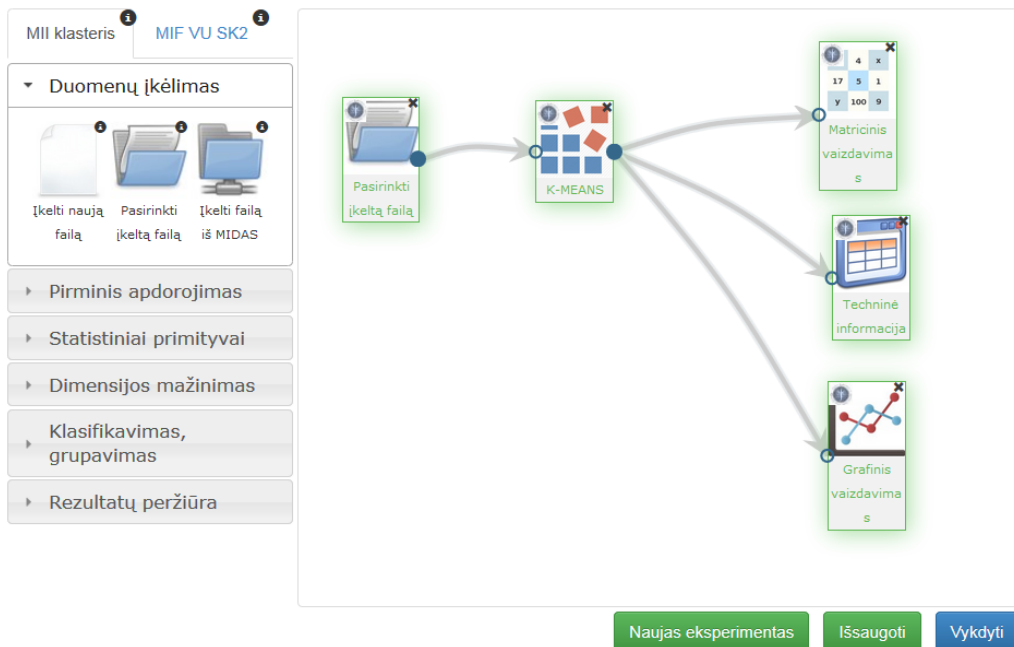
**Klasterių skaičius\***

\* Žvaigždute pažymėti laukai yra privalomi

114 pav. Langas „K-MEANS“

115 pav. pateikiama bendra K-MEANS komponentės testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, K-MEANS komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas, Grafinis vaizdavimas).





115 pav. K-MEANS algoritmo vykdymo darbų seka

Vykdamas eksperimentą K-MEANS komponentės parametrų lange buvo nurodyti tokie parametrai: maksimalus iteracijų skaičius – 100; maksimalus klasterių skaičius – 4.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 116 pav., 117 pav., 118 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	
Skaičiavimo laikas	0.03
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/28a31041aaafa7b28a4a35c277">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/28a31041aaafa7b28a4a35c277</a>

Atsisiųsti Atšaukti

116 pav. Komponentės „K-MEANS“ testavimo rezultatas: Techninė informacija

### Matricinis vaizdavimas

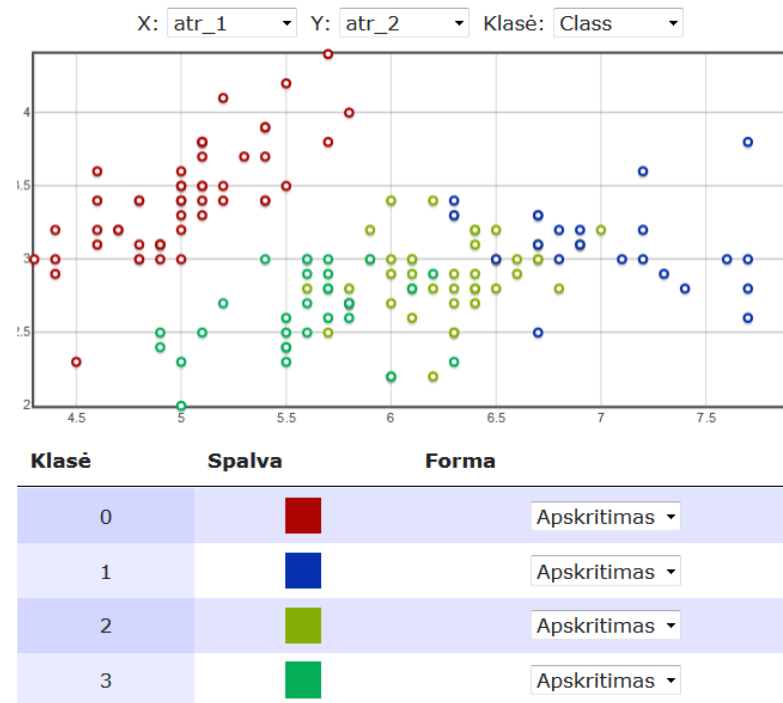
47.	5.1	3.8	1.6	0.2	0
48.	4.6	3.2	1.4	0.2	0
49.	5.3	3.7	1.5	0.2	0
50.	5	3.3	1.4	0.2	0
51.	7	3.2	4.7	1.4	2
52.	6.4	3.2	4.5	1.5	2
53.	6.9	3.1	4.9	1.5	2
54.	5.5	2.3	4	1.3	3
55.	6.5	2.8	4.6	1.5	2
56.	5.7	2.8	4.5	1.3	3

Atsisiųsti

Atšaukti

117 pav. Komponentės „K-MEANS“ testavimo rezultatas: Matricinis vaizdavimas

### Grafinis vaizdavimas



Atsisiųsti

Atšaukti

118 pav. Komponentės „K-MEANS“ testavimo rezultatas: Grafinis vaizdavimas

### 3.6.6.3 Komponentė „MLP“

Daugiasluoksnis perceptronas (DSP) (angl. multilayer perceptron) yra tiesioginio sklaidimo neuroninis tinklas, apmokomas klaidos sklaidimo atgal (angl. error back-propagation) metodu. Daugiasluoksnis perceptronas taikomas klasifikavimo uždaviniams spręsti.

Komponentės „MLP“ **rezultatas**: nauja duomenų matrica, kurioje kiekvienam pradinės duomenų aibės vektoriui nurodomos galimų klasių tikimybės ir nustatoma jo klasė.

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenis norima klasifikuoti MLP algoritmu, ir paspaudus du kartus komponentę „MLP“



atsiranda langas (119 pav.).

Numatytosios reikšmės: maksimalus iteracijų skaičius = 100, neuronų skaičius pirmame sluoksnyje = 5, antrame = 0, mokymo aibės dydis  $k = 90\%$ .

Naudotojui paspaudus „Patvirtinti“, tikrinamos įvestos reikšmės. Jos negali būti neigiamos. Jei parametro reikšmė negalima, rodomas klaidos pranešimas ties neteisingai įvestu lauku. Maksimalus iteracijų skaičius intervale [1; 1000] (sveikas skaičius). Neuronų skaičius paslėptuose sluoksniuose: pirmame bent vienas, antrame – neneigiamas sveikas skaičius. Mokymo aibės dydžio  $k$  mažiausia galima lauko reikšmė – 1% (galimos reikšmės intervale [1,100]), kryžminio validavimo parametro reikšmė – sveikas skaičius intervale [2,100].

Naudotojas gali šias reikšmes pakeisti. Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės. Šios reikšmės negali būti neigiami skaičiai. Maksimalus iteracijų skaičius turi būti intervale [1; 1000] (sveikas skaičius). Neuronų skaičius paslėptuose sluoksniuose turi būti: pirmame bent vienas, antrame – neneigiamas sveikas skaičius. Mokymo aibės dydžio  $k$  mažiausia galima lauko reikšmė – 1% (galimos reikšmės intervale [1,100]), kryžminio validavimo parametro reikšmė – sveikas skaičius intervale [2,100]. Jei įvestos netinkamos reikšmės, rodomas klaidos pranešimas ties neteisingai įvestu lauku.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas ir lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios, ir grįžtama į eksperimento planavimą.

**MLP**

Nurodykite algoritmo parametrus

**Maksimalus iteracijų skaičius\***

**1-as sluoksnis\***  **2-as sluoksnis\***

Neuronų skaičius paslėptuose sluoksniuose

**Pasirinkite mokymo parametą k reiškmę**

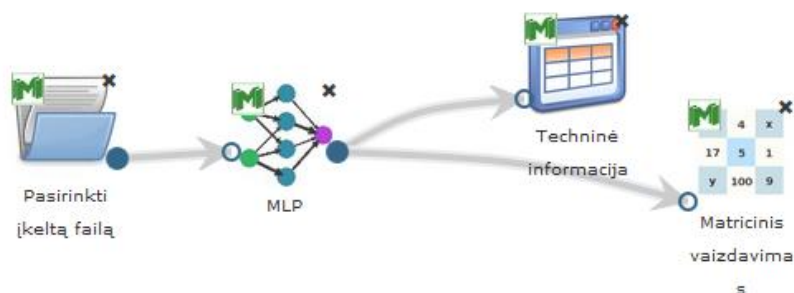
**Mokymo aibės dydis k (%)**  **Kryžminio validavimo parametras k**

**Parametras k\***

\* Žvaigždute pažymėti laukai yra privalomi

119 pav. Langas „MLP“

120 pav. pateikiama bendra MLP algoritmo vykdymo darbų seka (kartu su kitais klasifikavimo, grupavimo algoritmais), sudaryta iš failo įkėlimo komponentės, MLP komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas).



120 pav. MLP algoritmo vykdymo darbų seka

Vykdamas eksperimentą MLP komponentės parametruose buvo nurodyti tokie parametrai: maksimalus iteracijų skaičius – 100; neuronų skaičius pirmame sluoksnyje – 5; neuronų skaičius antrame sluoksnyje – 0; mokymo aibės dydžio parametras k – 90 %.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 121 pav., 122 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.0406595
Skaičiavimo laikas	0.02
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/066dcb2053919db153564d8b1">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/066dcb2053919db153564d8b1</a>

Atsisiųsti

Atšaukti

121 pav. Komponentės „MLP“ testavimo rezultatas: Techninė informacija

Matricinis vaizdavimas						
Objekto nr.	atr1 (REAL)	atr2 (REAL)	atr3 (REAL)	atr4 (REAL)	probClassIRIS-SETOSA (REAL)	probVER (REAL)
1.	5.1	3.5	1.4	0.2	0.992468	0.007
2.	4.9	3	1.4	0.2	0.991071	0.008
3.	4.7	3.2	1.3	0.2	0.992346	0.007
4.	4.6	3.1	1.5	0.2	0.992006	0.007
5.	5	3.6	1.4	0.2	0.992657	0.007
6.	5.4	3.9	1.7	0.4	0.992528	0.007

Atsisiųsti

Atšaukti

122 pav. Komponentės „MLP“ testavimo rezultatas: Matricinis vaizdavimas

**3.6.6.4 Komponentė „RDF“**

Atsitiktinis miškas (angl. *random forest*) yra populiarus ir efektyvus sprendimo medžių grupės klasifikavimo algoritmas. Pagrindinė atsitiktinio miško formavimo idėja yra tokia, kad reikia suformuoti tikslų klasifikatorių, apjungiant sprendimus daugelio binarinių sprendimų medžių, užaugintų naudojant skirtingus duomenų poaibius iš originalios duomenų aibės, ir atsitiktinai parinktus požymių poaibius iš požymių aibės. Toks binarinių medžių rinkinys pasižymi atsparumu persimokymui ir, medžių skaičiui augant, bendros klaidos konvergavimu iki stabilios reikšmės.

Komponentės „RDF“ **rezultatas**: nauja duomenų matrica, kurioje kiekvienam pradinės duomenų aibės vektoriui nurodomos galimų klasių tikimybės ir nustatoma jo klasė.

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenis norima klasifikuoti RDF algoritmu, ir paspaudus du kartus komponentę „RDF“



atsiranda langas (123 pav.).

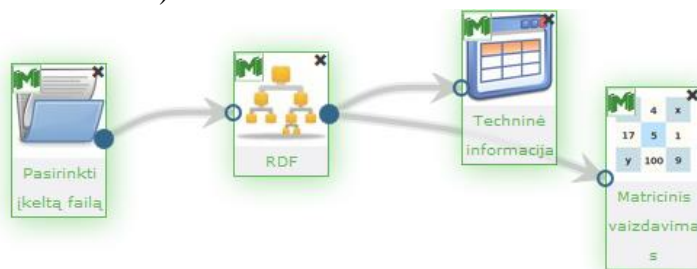
Numatytosios reikšmės: atsparumas triukšmui = 0,63, santykinis mokymo objektų kiekis = 80 %, santykinis testavimo objektų kiekis = 20 %.

Naudotojas gali šias reikšmes pakeisti. Naudotojui paspaudus mygtuką „Patvirtinti“ tikrinamos įvestos parametrų reikšmės. Šios reikšmės negali būti neigiami skaičiai. Atsparumo triukšmui reikšmė turi būti intervale (0; 1] (realusis skaičius). Mokymo, testavimo objektų suma turi sudaryti sudaro 100 %. Mokymo objektų kiekį galima keisti (mažiausia galima reikšmė – 1 %), testavimo objektų kiekio reikšmė apskaičiuojama automatiškai. Jei įvestos netinkamos reikšmės, rodomas klaidos pranešimas ties neteisingai įvestu lauku.

Naudotojui paspaudus mygtuką „Atšaukti“, langas uždaromas ir lieka galioti paskutinės išsaugotos reikšmės, jei tokios buvo, jei ne – numatytosios, ir grįžtama į eksperimento planavimą.

123 pav. Langas „RDF“

124 pav. pateikiama bendra RDF komponentės testavimo darbų seka, sudaryta iš failo įkėlimo komponentės, RDF komponentės ir rezultatų peržiūros komponentių (Techninė informacija, Matricinis vaizdavimas).



124 pav. RDF algoritmo vykdymo darbų seka

Vykdamas eksperimentą RDF komponentės parametrų lange buvo nurodyti tokie parametrai: atsparumo triukšmui parametras (0,63 pagal nutylėjimą), mokymo aibės objektų kiekio parametras (80 % pagal nutylėjimą), testavimo aibės objektų kiekio parametras (20 % pagal nutylėjimą). Nurodžius mokymo aibės dydį, testavimo aibę apskaičiuojama automatiškai.

Įvykdžius eksperimentą, analizuojamos darbų sekos rezultatai pateikti paveiksluose: 125 pav., 126 pav.

Techninė informacija	
Parametras	Reikšmė
Paklaida	0.193333
Skaičiavimo laikas	0.02
Rezultatų failas	<a href="http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/4219f4cbad2fc5c4b0494075a46">http://www.damis.lt/assets/b3dc68655f2eee81930b90a07d/experiment/4219f4cbad2fc5c4b0494075a46</a>

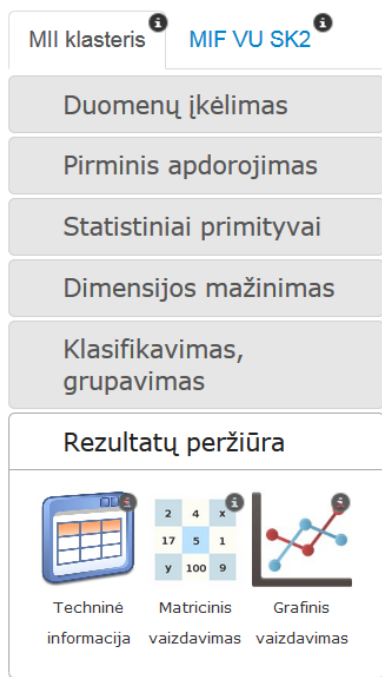
125 pav. Komponentės „RDF“ testavimo rezultatas: Techninė informacija

Matricinis vaizdavimas						
Objekto nr.	atr1 (REAL)	atr2 (REAL)	atr3 (REAL)	atr4 (REAL)	probability_IRIS-SETOSA (REAL)	p
1.	5.1	3.5	1.4	0.2	1	0
2.	4.9	3	1.4	0.2	0.986667	0
3.	4.7	3.2	1.3	0.2	1	0
4.	4.6	3.1	1.5	0.2	1	0
5.	5	3.6	1.4	0.2	1	0
6.	5.4	3.9	1.7	0.4	1	0
7.	4.6	3.4	1.4	0.3	1	0

126 pav. Komponentės „RDF“ testavimo rezultatas: Matricinis vaizdavimas

### 3.6.7 Kairiojo meniu skiltis „Rezultatų peržiūra“

Naudotojas rezultatų peržiūros komponentes galės naudoti tik tuomet, kai bus apskaičiuoti eksperimento rezultatai. „Grafinis vaizdavimas“ – tai komponentė, kurios pagalba naudotojas galės peržiūrėti rezultatų grafiką, sujungus su vykdyto algoritmo komponente ir spragtelėjus ją du kartus. Atsidariusiame grafike naudotojas galės matyti grafiką. „Matricinis vaizdavimas“ – tai komponentė, kurios pagalba naudotojas galės peržiūrėti duomenis ar algoritmų rezultatus matricos (lentelės) pavidalu, kai ši komponentė bus sujungta su failo įkėlimo ar algoritmo komponente. „Techninė informacija“ – tai komponentė, skirta techninei informacijai peržiūrėti: skaičiavimo laikui, gautos paklaidos ir kt. Komponentės „Grafinis vaizdavimas“ ir „Matricinis vaizdavimas“ turi galimybę atsiųsti rezultatus. Tam du kartus spragtelėjus komponentę, atsiradusiame rezultatų peržiūros lange, reikia paspausti mygtuką „Atsiųsti“. Naudotojas norėdamas atnaujinti rezultatų grafiką, turi spausti mygtuką „Atnaujinti“. Jei reikia uždaryti langą, reikia spausti „Patvirtinti“. Rezultatų peržiūros meniu skiltis pateikta 127 pav.



127 pav. Meniu skiltis „Rezultatų peržiūra“

Toliau detalai bus aprašytos visos „Rezultatų peržiūros“ skilties komponentės ir pateikiamos jų naudojimosi instrukcijos.

#### 3.6.7.1 Komponentė „Matricinis vaizdavimas“

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenis (rezultatus) norima pavaizduoti, ir paspaudus du kartus komponentę „Matricinis vaizdavimas“,

2	4	x
17	5	1
y	100	9

Matricinis  
vaizdavimas

atidaromas langas, kuriame matoma duomenų (rezultatų) lentelė (matrica) (128 pav.).



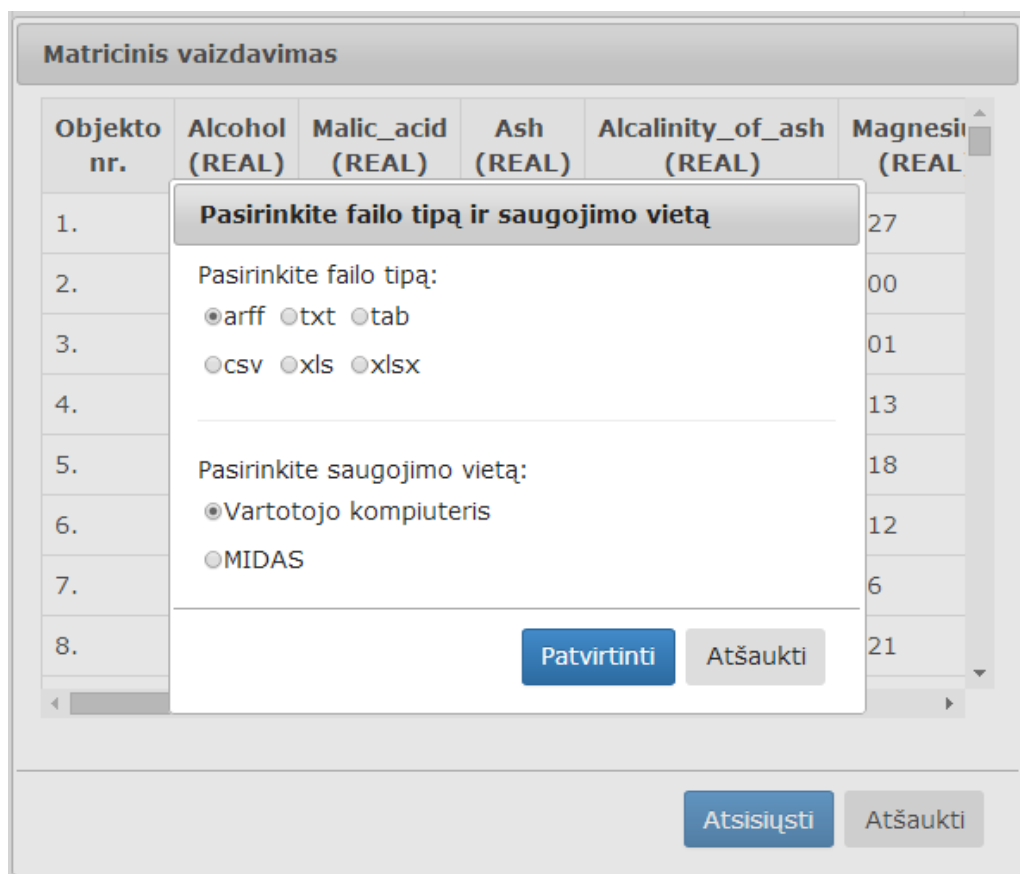
Požymių pavadinimai nuskaitomi iš duomenų failo. Atsiranda vertikalios ir horizontalios slinkties juostos, jei duomenų lentelė turi daugiau duomenų, nei jai yra skirta vietos. Traukiant vertikalią slinkties juostą, lentelės antraštės eilutė lieka matymo zonoje, t. y. jos pozicija yra fiksuota. Slenkant horizontalią slinkties juostą, lentelės antraštės eilutė pasislenka kartu su stulpeliais į dešinę arba kairę. Paspaudus mygtuką „Atšaukti“, grįžtama į eksperimento planavimo langą.

Matricinis vaizdavimas					
Objekto nr.	Alcohol (REAL)	Malic_acid (REAL)	Ash (REAL)	Alcalinity_of_ash (REAL)	Magnesium (REAL)
1.	14.23	?	?	15.6	127
2.	13.2	1.78	2.14	11.2	100
3.	13.16	2.36	2.67	18.6	101
4.	14.37	1.95	2.5	16.8	113
5.	13.24	2.59	2.87	21	118
6.	14.2	1.76	2.45	15.2	112
7.	14.39	1.87	2.45	14.6	96
8.	14.06	2.15	2.61	17.6	121

128 pav. Langas „Matricinis vaizdavimas“

Jei naudotojas nori atsisiųsti rezultatų lentelę, jis turi paspausti mygtuką „Atsisiųsti“. Tuomet atsidaro langas (129 pav.). Naudotojas gali pasirinkti pageidaujamą failo formatą ir vietą, kur failas bus išsaugotas. Atsidariusiame lange pagal nutylėjimą yra pažymėtas failo tipas – arff, saugojimo vieta – naudotojo kompiuteris. Naudotojas gali pasirinkti kitą saugojamo failo tipą bei nurodyti, kad failas būtų saugomas MIDAS archyve.

Naudotojui paspaudus mygtuką „Patvirtinti“, atsidaro failo parsisiuntimo langas, atlikus reikiamus veiksmus, matricinio vaizdavimo langas uždaromas, ir grįžtama į eksperimento planavimo langą. Paspaudus „Atšaukti“ langas uždaromas, failas nėra išsaugomas nei naudotojo kompiuteryje, nei MIDAS archyve.



129 pav. Langas „Matricinis vaizdavimas“ paspaudus mygtuką „Atsisiųsti“

### 3.6.7.2 Komponentė „Grafinis vaizdavimas“

Į darbalaukį įkėlus, sujungus su komponente, kurios duomenis norima pavaizduoti grafiškai, ir paspaudus du kartus komponentę „Grafinis vaizdavimas“,



Grafinis

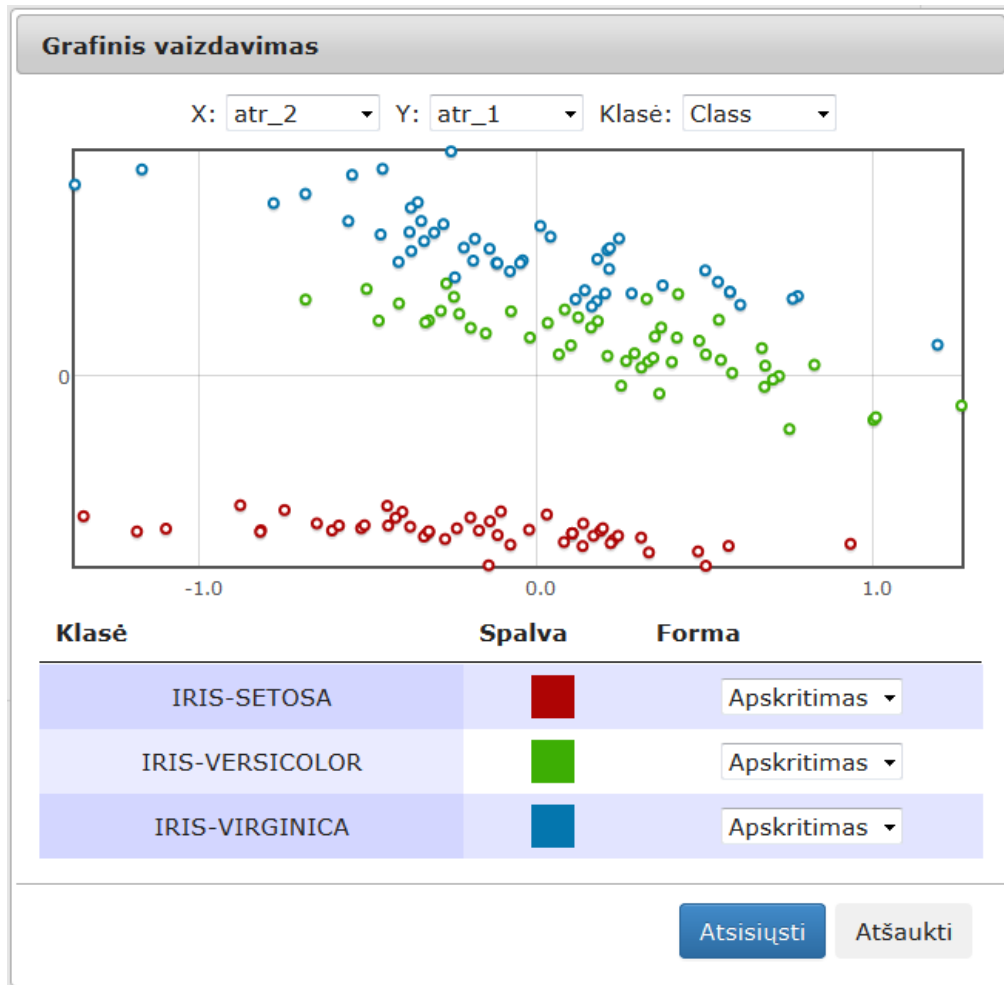
vaizdavimas

atidaromas langas (130 pav.). Lango viršuje yra pateikiami trys pasirinkimai. Naudotojas gali pasirinkti, kurio požymio reikšmės bus atvaizduotos X ir Y ašyse, kuris požymis reiškia klasę. X ir Y sąrašo reikšmės pagal nutylėjimą priskiriamos atitinkamai iš eilės iš arff failo.

Jei klasės požymis yra žinomas iš arff failo, grafike automatiškai klasės laukui priskiriamas klasės požymis. Grafiko apačioje pateikiamas klasių sąrašas, kurį sudaro šie elementai:

- Klasės pavadinimas;
- Klasę atitinkanti spalva (pradžioje priskiriama atsitiktinė spalva, tačiau naudotojas ją gali pakeisti į kitą, tam reikia klikelti ant spalvos ir parinkti norimą);
- Taško forma (pradžioje vis taškai yra apskritimai, tačiau naudotojas gali pakeisti į kitą formą (galimos formos: apskritimas, kvadratas, rombas, trikampis, kryžiuokas).

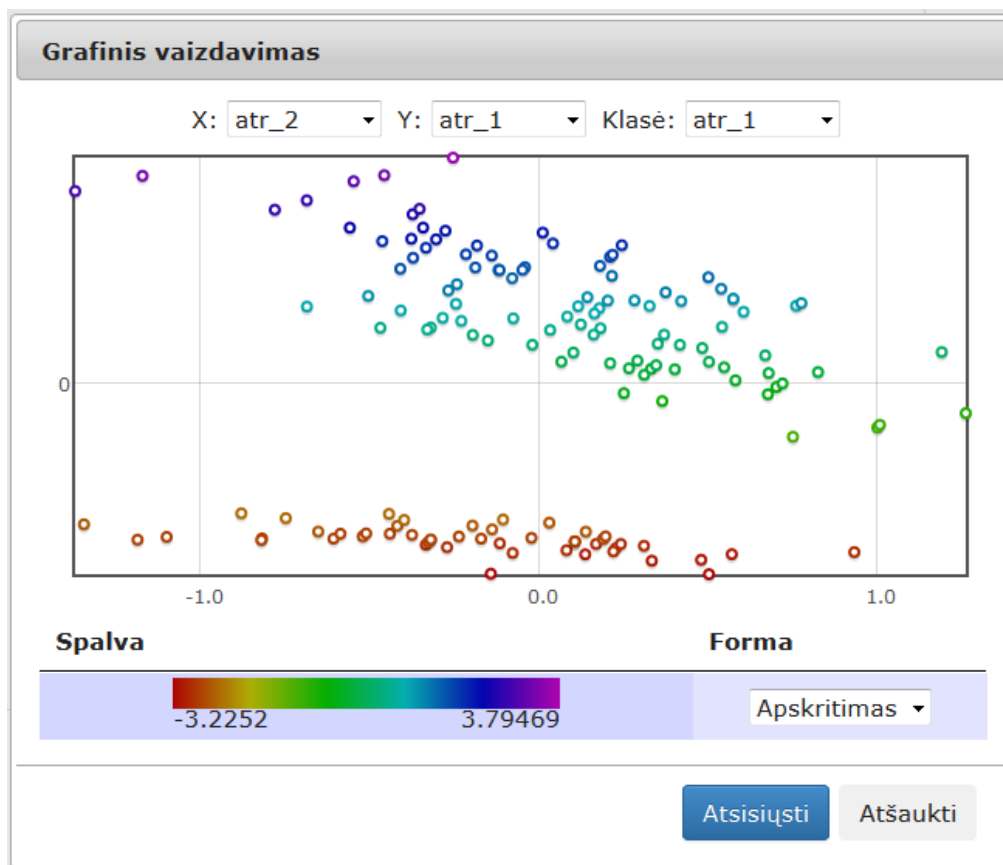
Parodomas visas klasių sąrašas, jei jo elementų kiekis yra ne didesnis nei 10. Jei sąrašė elementų yra daugiau, atsiranda vertikali slinkties juosta.



**130 pav. Langas „Grafinis vaizdavimas“, kai klasės požymis yra žinoma iš arff failo**

Jei klasės atributas nėra žinomas iš arff failo arba naudotojas pakeičia klasės atributą į bet kurį kitą, neturintį klasės požymio, naudotojas mato langą, pavaizduotą 131 pav. Pasirinkto klasės požymio reikšmės yra perskaičiuojamos į intervalą  $[0; 255]$ , kuris atitinka taško spalvą. Grafiko apačioje matoma spalvų paletė su nurodytomis klasės požymio minimalia ir maksimalia reikšmėmis. Naudotojas gali pakeisti taškų formą. Viena forma pritaikoma visiems taškams.

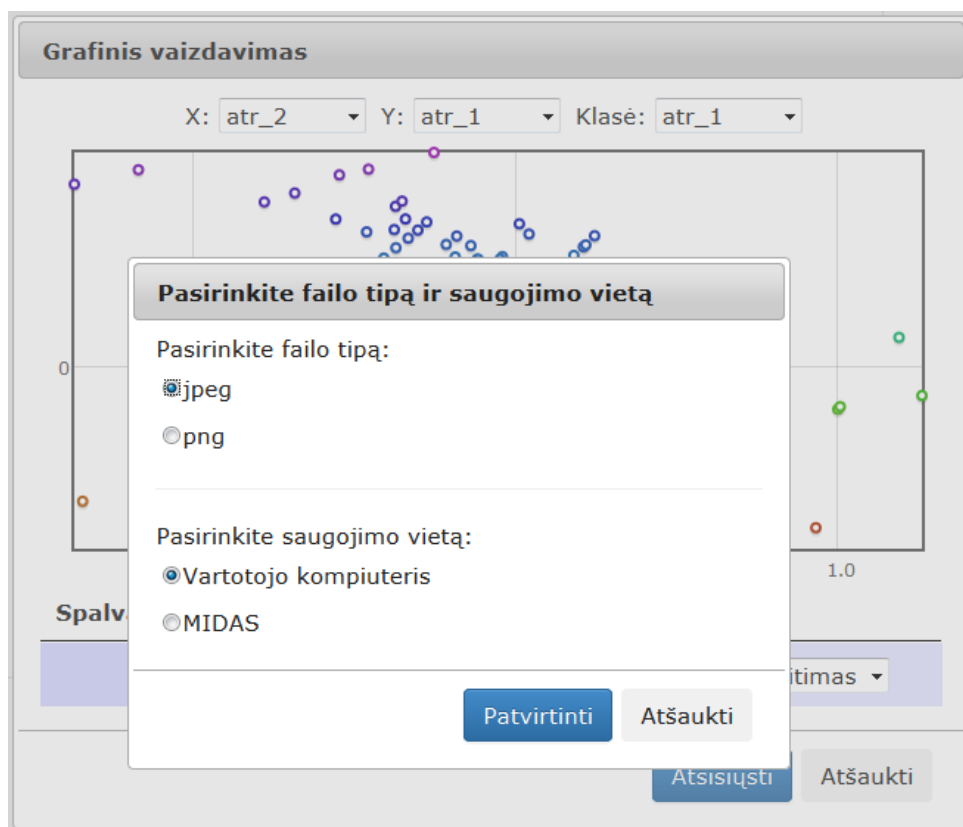
Pakeitus bet kurio objekto būseną (spalvą, formą, pasirinkus kitą požymį, atvaizduojamą X ašyje ar kt.), duomenys grafike atnaujinami automatiškai.



131 pav. Langas „Grafinis vaizdavimas“, kai klasės požymis nurodo naudotojas

Jeigu naudotojas nori atsisiųsti grafiką, jis turi paspausti mygtuką „Atsisiųsti“. Tuomet atsidaro langas, kuriame naudotojas turi pasirinkti pageidaujamą failo formatą ir vietą, kur failas bus išsaugotas (132 pav.). Naudotojas gali pasirinkti norimą failo formatą (jpeg ar png) ir vietą, kur failas bus išsaugotas (naudotojo kompiuteryje ar MIDAS archyve). Pagal nutylėjimą yra pažymėtas failo tipas – jpeg, saugojimo vieta – naudotojo kompiuteris.

Naudotojui paspaudus mygtuką „Patvirtinti“, atsidaro failo parsisiuntimo langas, atlikus reikiamus veiksmus, grafinio vaizdavimo langas uždaromas, ir grįžtama į eksperimento planavimo langą. Paspaudus „Atšaukti“ langas uždaromas, failas nėra išsaugomas nei naudotojo kompiuteryje, nei MIDAS archyve.



132 pav. Langas „Grafinis vaizdavimas“, paspaudus mygtuką „Atsisiųsti“.

### 3.6.7.3 Komponentė „Techninė informacija“

Į darbalaukį įkėlus, sujungus su komponente, kurios techninę informaciją norima pamatyti, ir paspaudus du kartus komponentę „Techninė informacija“



Techninė  
informacija

rodomas langas, kuriame pateikiama algoritmo vykdymo techninė informacija arba iš serviso gautas klaidos pranešimas (133 pav.).

Šiame lange informacija pateikiama dviejų stulpelių lentelės pavidalu. Pirmajame stulpelyje nurodyti parametrų pavadinimai, antrajame – jų reikšmės. Rodoma algoritmo gautos paklaidos reikšmė, pvz., MDS paklaida, SOM kvantavimo paklaida ir kt., taip pat skaičiavimo laikas bei nuoroda į rezultatų failą, kurią paspaudus galima rezultatų failą parsisiųsti arff formatu.

Paspaudus mygtuką „Atsisiųsti“ techninę informaciją galima parsisiųsti į savo kompiuterį ar MIDAS archyvą. Paspaudus „Atšaukti“ langas uždaromas, failas nėra išsaugomas nei naudotojo kompiuteryje, nei MIDAS archyve.

Techninė informacija	
Parametras	Reikšmė
Klaidos	0.00174947
Skaičiavimo laikas	0.41
Rezultatų failas	<a href="http://dev.damis.lt/assets/f285280dd0aea2ab56ab4580a4/experiment/3838f987da7c9816683130e78c">http://dev.damis.lt/assets/f285280dd0aea2ab56ab4580a4/experiment/3838f987da7c9816683130e78c</a>

133 pav. Lango „Techninė informacija“





























### 3.7 Pagrindinio puslapio meniu skiltis „Eksperimentai“

Eksperimentų istorijos langas pateiktas 134 pav. Lango elementai yra šie:

- Eksperimento pavadinimas;
- Statusas:
  - Išsaugotas;
  - Vykdomas;
  - Įvykdytas;
  - Klaida.
- Veiksmai:
  - Redaguoti;
  - Peržiūrėti.

Pažymėtus eksperimentus naudotojas gali trinti. Ištrinus eksperimentą pašalinami visi su juo susiję duomenys. Ištrinus atstatyti buvusio eksperimento naudotojas nebegali. Kiekvieną eksperimentą rodomą sąraše galima peržiūrėti arba redaguoti (tokiu atveju atidaromas eksperimento planavimo langas su jame atvaizduota eksperimentą nusakančių komponenčių seka). Bet koks išsaugoto eksperimento keitimas (išskyrus rezultatų peržiūros komponenčių pridėjimą) vykdymo atveju kuria naują eksperimentą.

## Eksperimentų istorija

Pavadinimas	Statusas	Veiksmai
exp108	Įvykdytas	 
<u>SOM</u> -MDS	Įvykdytas	 
MDS	Klaida	 
exp105	Klaida	 
exp104	Klaida	 
exp102	Įvykdytas	 
exp101	Klaida	 
exp100	Klaida	 
exp099	Išsaugota	
exp098	Klaida	 
exp095	Įvykdytas	 
exp094	Įvykdytas	 
Filtravimas klaida dėl <u>blogų</u> klasės atributų	Klaida	 
exp003	Išsaugota	
exp001	Klaida	 

« 1 2 3 4 5 6 »

134 pav. Meniu skiltis „Eksperimentų peržiūra“

### 3.8 Pagrindinis puslapio meniu „Failų sąrašas“

Šiame lange naudotojas galės tvarkyti savo įkeltus duomenų failus. Pirmiausiai rodomas įkeltų failų sąrašas. Naudotojas gali atsisiųsti reikiamus failus į savo kompiuterį, pasirinkus norimą failo formatą. Taip pat naudotojas gali redaguoti failo aprašą, bei įkelti naują failą. Norint redaguoti įkelto duomenų failo sąrašą, naudotojui reikia paspausti paveiksluką „Redaguoti“ ties failo pavadinimu, tuomet atsidarys forma, kurioje naudotojas galės pakeisti failo aprašą bei pakeisti patį duomenų failą nauju. Failų sąrašo langas pateiktas 135 pav. Sąrašo elementus galima rūšiuoti pagal:

- Pavadinimą;
- Dydį;
- Sukūrimo datą.

Paspaudus failo pavadinimą arba pasirinkus veiksmą „Redaguoti“, atidaromas failo duomenų redagavimo langas (3.5.2 punktas). Paspaudus mygtuką žymintį failo plėtinį, inicijuojamas failo parsisiuntimas į naudotojo kompiuterį pasirinktu formatu. Siuntimo įgyvendinimas perduodamas naršyklei.

Pažymėtus failus naudotojas gali ištrinti. **Pastaba.** Jei failas yra naudojamas eksperimente,

ištrinamas failas tik iš naudotojo failų sąrašo. Eksperimente naudojamas failas išlieka tol kol nėra trinamas eksperimentas.

Pavadinimas	Dydis	Sukurta	Aprašymas	Veiksmai	Atsisiuntimas
wineblogas	11.08 KB	2014-06-03	wineblogas	[pen]	CSV ARFF TXT TAB XLSX
resourees	10.72 MB	2014-05-15		[pen]	CSV ARFF TXT TAB XLSX
test	4.94 KB	2014-05-13		[pen]	CSV ARFF TXT TAB XLSX
rt	4.94 KB	2014-05-13		[pen]	CSV ARFF TXT TAB XLSX
failas	7.56 KB	2014-04-28	duomenys	[pen]	CSV ARFF TXT TAB XLSX
bxt (1)	0.49 KB	2014-04-23		[pen]	CSV ARFF TXT TAB XLSX
elips	1.41 MB	2014-04-09		[pen]	CSV ARFF TXT TAB XLSX
elips	0.54 MB	2014-04-09	ghfdhfxhxf	[pen]	CSV ARFF TXT TAB XLSX
ddd	1.41 MB	2014-04-09		[pen]	CSV ARFF TXT TAB XLSX
wine_su_klasem	14.19 KB	2014-04-09	vynas su klas4m	[pen]	CSV ARFF TXT TAB XLSX
smacof-2-0001b	22.67 KB	2014-04-09	hfgdhfxhxf	[pen]	CSV ARFF TXT TAB XLSX
wine	11.08 KB	2014-04-09	hfgdhfxhxf	[pen]	CSV ARFF TXT TAB XLSX
vynas	11.08 KB	2014-04-09	hfgdhfxhxf	[pen]	CSV ARFF TXT TAB XLSX
iris	2.48 KB	2014-04-08		[pen]	CSV ARFF TXT TAB XLSX
iris	5.68 KB	2014-04-04		[pen]	CSV ARFF TXT TAB XLSX

135 pav. Langas „Failų sąrašas“

### 3.8.1 Skiltis „Įkelti failą iš kompiuterio“

Norėdamas įkelti naują duomenų failą, naudotojas turi paspausti meniu skiltį „Įkelti failą iš kompiuterio“, tuomet atsidarys langas su laukais, kuriuos reikės užpildyti. Naudotojui reikia įvesti duomenų failo vardą, taip pat įkelti failą iš kompiuterio ir pateikti failo aprašymą (jis nėra būtinas). Paspaudus mygtuką „Įkelti duomenis“, vykdomas įkelto failo tikrinimas, ar įvestas pavadinimas. Jei pavadinimas įvestas ir failo formatas tinkamas, įkeltas failas išsaugomas ir naudotojui atidaromas failų sąrašo langas. Priešingu atveju, jei rastos klaidos, rodomas klaidos pranešimas prie neteisingai užpildytų laukų. Langas „Įkelti failą iš kompiuterio“ pateiktas 136 pav.



136 pav. Langas „Įkelti failą iš kompiuterio“

### 3.8.2 Veiksmo „Redaguoti failą“ langas

Norėdamas redaguoti duomenų failo aprašą, naudotojas turi paspausti paveiksliuką „Redaguoti“, tuomet atsidarys forma su laukais, užpildytais ankstesniais duomenų failo parametrais. Norėdamas pakeisti vieną ar kitą lauką, naudotojas turi ištrinti seną įrašą ir pakeisti jį nauju. Tuomet reikia spausti mygtuką „Patvirtinti“. Jei naudotojas pakeitė neteisingai, jam bus rodomas klaidos pranešimas. Jei naudotojas pakeitė teisingai, jo pakeitimai bus išsaugoti. Langas „Redaguoti failą“ pateiktas 137 pav.

137 pav. Langas „Redaguoti failą“

### 3.9 Naujo algoritmo įkėlimas

Naudotojas norėdamas įkelti savo algoritmą, pasirinkęs meniu punktą „Mano algoritmai“, turei pasirinkti kortelę „Įkelti algoritmą iš kompiuterio“. Atsiradusioje formoje reikia nurodyti algoritmo pavadinimą, įkelti norimą failą paspaudus mygtuką „pasirinkti“. Be to, galima pateikti algoritmo aprašymą, kuriame nurodyti ne tik informaciją apie algoritmą, bet ir pateikti instrukcijas apie algoritmo programinio kodo vykdymą (138 pav.).

The screenshot shows the top navigation bar of the DAMIS system with the following items: 'DAMIS', 'Kurti eksperimentą +', 'Eksperimentai ☰', 'Failų sąrašas 📄', and 'Mano algoritmai 📁'. Below the navigation bar, there are two tabs: 'Mano algoritmai' and 'Įkelti algoritmą iš kompiuterio'. The 'Įkelti algoritmą iš kompiuterio' tab is active. The form contains the following fields and buttons:

- Pavadinimas\***: A text input field.
- Failas\***: A file selection input field with a blue button labeled 'Pasirinkti'.
- Aprašymas**: A large text area for description.
- A blue button labeled 'Įkelti algoritmą' at the bottom.

138 pav. Langas „Įkelti algoritmą“

### 3.10 Eksperimento vykdymas

#### 3.10.1 Mygtukas „Naujas eksperimentas“

Eksperimento planavimo lange paspaudus mygtuką „Naujas eksperimentas“

Naujas eksperimentas

eksperimentų langas yra išvalomas ir galima pradėti planuoti naują eksperimentą.

#### 3.10.2 Mygtukas „Išsaugoti“

Eksperimento planavimo lange paspaudus mygtuką „Išsaugoti“

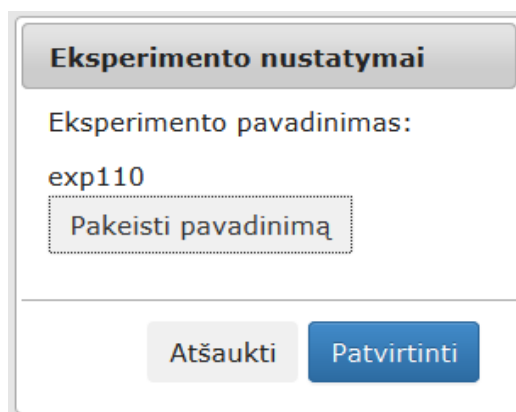
Išsaugoti

parodomas eksperimento nustatymo langas (139 pav.).

Automatiškai sukuriamas eksperimento pavadinimas, kurį naudotojas gali pakeisti, paspaudęs mygtuką „Pakeisti pavadinimą“ ir atsiradusiame lange nurodęs norimą pavadinimą. Paspaudus mygtuką „Patvirtinti“, suplanuotas eksperimentas (darbų seka) yra išsaugomas naudotojo eksperimentų sąrašė, jo būseną – „Išsaugotas“.

Paspaudus mygtuką „Atšaukti“ langas uždaromas, eksperimentas nėra išsaugomas, grįžtama į eksperimento planavimą.

Norint vėliau įvykdyti išsaugotą eksperimentą, reikia rinktis pagrindinio puslapio meniu skiltį „Eksperimentai“, atsiradusiame lange prie išsaugoto eksperimento paspausti redagavimo piktogramą. Eksperimento planavimo lange bus parodyta išsaugoto eksperimento darbų seka, norint įvykdyti eksperimentą, reikia paspausti mygtuką „Vykdyti“.



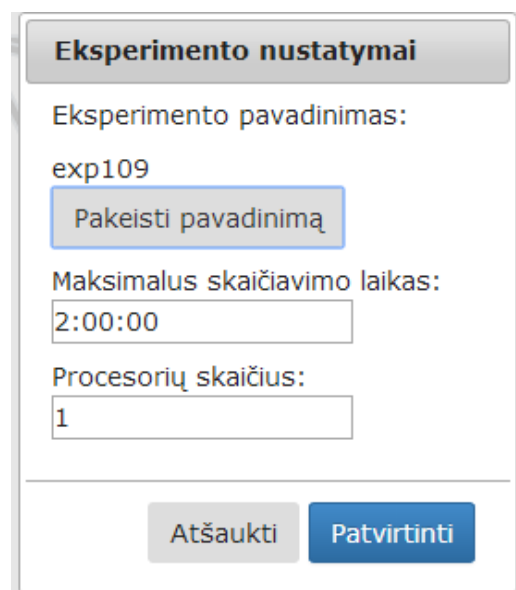
139 pav. Langas „Eksperimento nustatymai“, norint išsaugoti eksperimentą

### 3.10.3 Mygtukas „Vykdėti“

Naudotojas, kuris suplanavo eksperimento seką ir nori įvykdyti suplanuotą eksperimentą, turi paspausti mygtuką „Vykdėti“



Atidaromas langas, kuris pavaizduotas 140 pav. Automatiškai sukuriamas eksperimento pavadinimas, kurį naudotojas gali pakeisti, paspaudęs mygtuką „Pakeisti pavadinimą“ ir atsiradusiame lange nurodęs norimą pavadinimą. Taip pat galima pakeisti maksimalų skaičiavimo laiką (numatytoji reikšmė – 2 val.) bei procesorių skaičių (numatytoji reikšmė – 1). Jei procesorių skaičius daugiau nei 1, skaičiavimai vykdomi lygiagrečiai naudojant nurodytą procesorių skaičių. Sistema automatiškai parenka vykdymui lygiagrečiąją duomenų analizės algoritmų versiją. Naudotojui nebūtina turėti specifinių žinių apie algoritmų lygiagretinimą bei užduočių paleidimą lygiagrečiųjų ir paskirstytųjų skaičiavimų telkiniuose.



140 pav. Langas „Eksperimento nustatymai“, norint įvykdyti eksperimentą

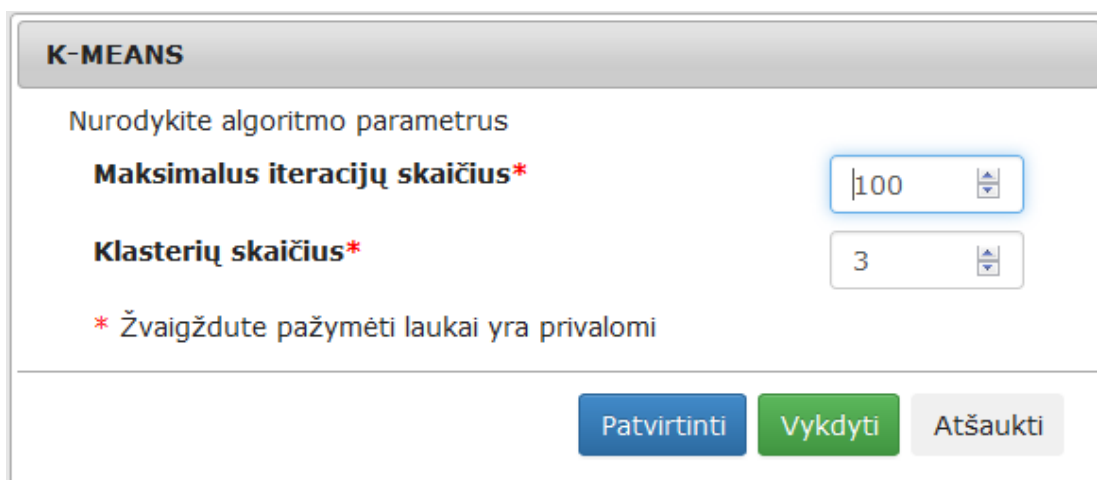
Paspaudus mygtuką „Patvirtinti“ tikrinama preliminarus skaičiavimo laiko parametro reikšmė. Laikas turi būti parašytas skaičiais, kurie atskiriami dvitaškiu

(valandos : minutės : sekundės). Eksperimento pavadinimo laukas negali būti tuščias. Eksperimentas išsaugomas naudotojo eksperimentų sąrašė, jo statusas – „Vykdomas“. Paspaudus mygtuką „Atšaukti“ langas uždaromas, eksperimentas nėra nei išsaugomas ir nei vykdomas, grįžtama į eksperimento planavimą.

Paleidus eksperimentą vykdymui, yra validuojami kiekvienos komponentės parametrai. Jei eksperimentas buvo sėkmingai įvykdytas (nebuvo komponentių parametrų validavimo ir vykdymo klaidų) eksperimento statusas pasikeičia į „Įvykdytas“ ir galima peržiūrėti rezultatus, paspaudus peržiūros piktogramą ties eksperimento pavadinimu. Jei buvo gautas klaidos pranešimas, komponentė lieka raudona, eksperimentas nutraukiamas ir eksperimento statusas pasikeičia į „Klaida“.

### 3.10.4 Atskiros komponentės vykdymas

Sukurus eksperimento seką galima vykdyti ne visą eksperimentą, o tik atskiras komponentes. Tam reikia paspausti du kartus sukurtos sekos norimą komponentę, atsidaro komponentės parametrų keitimo langas (pavyzdžiui, 141 pav.), kuriame paspaudus mygtuką „Vykdyti“, įvykdoma sudarytos sekos dalis iki šios komponentės.



141 pav. Komponentės parametrų nustatymo langas

### 3.11 Pavyzdiniai eksperimentų scenarijai

Sukurti pavyzdiniai scenarijai naudojant stuburo duomenų aibę. Stuburo ligų duomenų rinkinį galima klasifikuoti į 3 klases – sveiki, stuburo disko išvarža, spondilolistezė (angl. normal, disk hernia, spondylolisthesis) – arba į 2 klases – sveiki, sergantys (angl. normal, abnormal). Visą duomenų rinkinį sudaro 310 pacientų. Kiekvieną pacientą apibūdina šeši biomechaniniai požymiai: dubens dažnis (angl. pelvic incidence), dubens tentas (angl. pelvic tilt), juosmens kampas (angl. lumbar lordosis angle), sakraliniai nuolydžiai (angl. sacral slope), dubens spindulys (angl. pelvic radius) ir spondilolistezės klasė (angl. the grade of spondylolisthesis). Eilutės atitinka vertintą paciento atvejį, stulpeliai matuotų požymių reikšmės, paskutinis stulpelis klasės numeris.

Kiekvienas administratoriaus teisės turintis naudotojas gali savo sukurtas eksperimentų sekas padaryti pavyzdinėmis, t.y. jos bus įkeltos į visų naudotojų pavyzdinių eksperimentų sąrašą

(142 pav.). Tam reikia paspausti veiksmų mygtuką

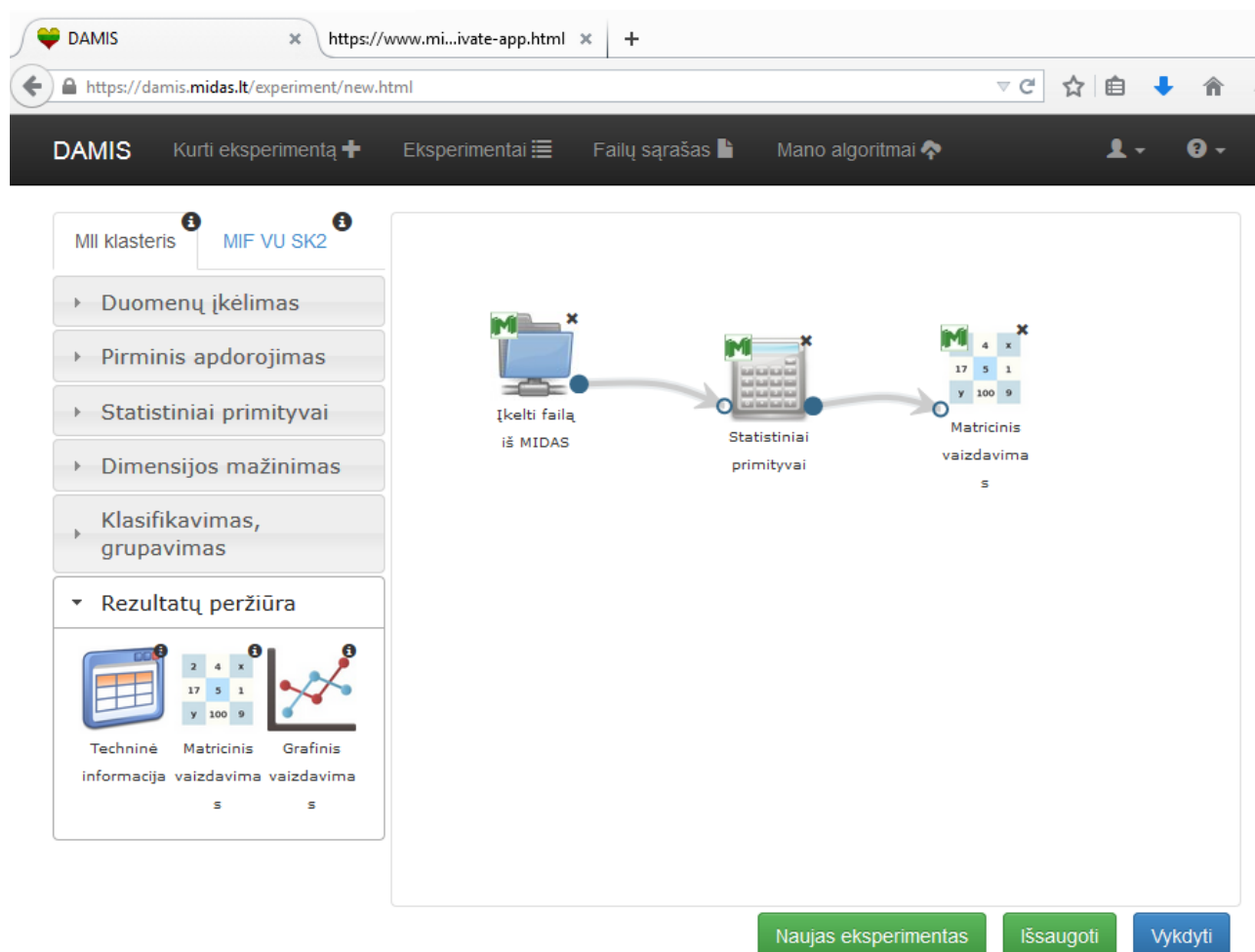


Pavadinimas	Statusas	Pradžia	Pabaiga	Veiksmai
Norm_stub_duom_analize	Įvykdytas	2015-05-23 14:34	2015-05-23 14:36	

142 pav. Eksperimento įkėlimas į pavyzdinių eksperimentų sąrašą

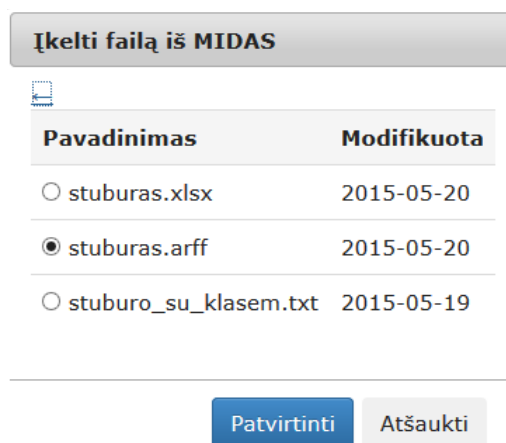
### 3.11.1 Stuburo duomenų statistinė analizė

Prieš pradėdant analizuoti nežinomą duomenų aibę, pirmiausia reikia atlikti duomenų statistinę analizę. Atlikus eksperimentą gauname pagrindinius duomenis apie tiriamų duomenų kiekvieno parametro matuotas reikšmes: minimali, maksimali kiekvieno parametro reikšmė, parametro reikšmių vidurkis, dispersija ir mediana. Statistinė analizė atliekama pagal tokį scenarijų, pateiktą 143 pav.:



143 pav. Statistinės analizės eksperimento scenarijus

1. Viršutinėje meniu juostoje spaudžiama nuoroda **Kurti eksperimentą**. Atidaromas Naujo eksperimento kūrimo langas.
2. Tempiant į darbalaukį įkeliamos eksperimentui atlikti reikalingos komponentės: Duomenų įkėlimo komponentė, statistinių primityvų komponentė ir duomenų peržiūros komponentė.
3. Pateiktame scenarijuje duomenims įkelti naudojama duomenų įkėlimo komponentė **Įkelti failą iš MIDAS**. Du kartus kliktelėjus ant komponentės, atidaromas failo įkėlimo langas (144 pav.). Pasirinkus tinkamą analizei failą, spaudžiamas mygtukas **Patvirtinti**.



**144 pav. Duomenų įkėlimo langas**

4. Visos komponentės eksperimento atlikimo tvarka sujungiamos į vieną seką, duomenų įkėlimo komponentėje paspaudžiamas pelyte pilnavidūris skrituliukas ir tempiamas iki statistinių primityvų komponentės tuščiavidūrio skrituliuko. Analogiškai sujungiamos ir statistinių primityvų bei matricinio vaizdavimo komponentės.
5. Sukūrus eksperimento seką, spaudžiamas mygtukas **Vykdyti**. Sukurtam eksperimentui visada suteikiamas vardas pagal nutylėjimą, bet tyrėjas gali eksperimentą pavadinti savaip, bei Eksperimento nustatymų lange nustatyti eksperimento atlikimo parametrus (145 pav.). Pasirinkus vykdymo parametrus, spaudžiamas mygtukas **Patvirtinti**.

Eksperimento nustatymai

Eksperimento pavadinimas:  
 Statistinė\_analizė

Maksimalus skaičiavimo laikas:

Procesorių skaičius:

Pirminė atmintis (RAM):  
 MB

Antrinė atmintis (HDD):  
 MB

Pradėti eksperimentą nuo:

---

**145 pav. Eksperimento nustatymo langas**

6. Norint peržiūrėti, koks sukurto eksperimento statusas, spaudžiama nuoroda viršutinėje meniu juostoje „Eksperimentai“
7. Kai sukurto eksperimento statusas tampa „Įvykdytas“, galima peržiūrėti ir atsisiųsti gautus rezultatus. Paspaudus akutę, atidaromas Eksperimento langas.

<input type="checkbox"/>	Statistinė_analizė	Įvykdytas	2015-05-22 10:59	2015-05-22 11:37	<input type="button" value="✎"/> <input type="button" value="📄"/>
--------------------------	--------------------	-----------	------------------	------------------	---

Įvykdyto eksperimento rezultatas pateiktas paveiksle. Matome šešių matuotų parametų visas statistinių primityvų reikšmes (146 pav.).

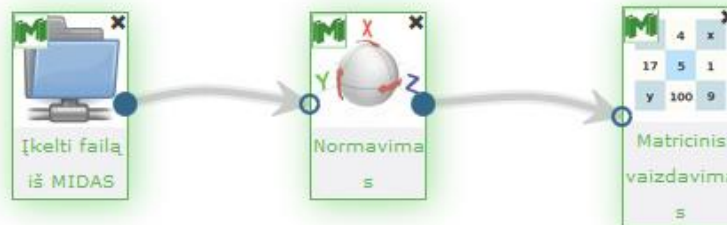
Matricinis vaizdavimas					
Objekto nr.	min (NUMERIC)	max (NUMERIC)	mean (NUMERIC)	std (NUMERIC)	med (NUMERIC)
1.	26.15	118.48	60.4599	17.0723	58.78
2.	-6.55	49.43	17.5429	9.99198	16.42
3.	14	101.02	51.851	18.2572	49.78
4.	13.37	121.43	42.9539	13.4011	42.44
5.	70.08	163.07	117.921	13.2961	118.34
6.	-11.06	148.75	25.3499	30.6652	12.07

< \_\_\_\_\_ >

146 pav. Matricinio vaizdavimo langas rezultatų pateikimui

### 3.11.2 Stuburo duomenų normavimas

Peržiūrėjus stuburo duomenų statistinius primityvus, galime įvertinti požymių reikšmių kitimo intervalus, vidurkius, duomenų išsibarstymą. Išanalizavus rezultatus matyti, kad atskirų požymių reikšmių intervalai skiriasi, todėl požymiai gali skirtingai įtakoti duomenų tyrybos rezultatus. Siekiant to išvengti, atliekamas duomenų normavimas. Stuburo duomenų normavimo seka pateikta 147 pav.



147 pav. Duomenų normavimo scenarijus

1. Viršutinėje meniu juostoje spaudžiama nuoroda **Kurti eksperimentą**. Atidaromas Naujo eksperimento kūrimo langas.



2. Tempiant į darbalaukį įkeliamos eksperimentui atlikti reikalingos komponentės: Duomenų įkėlimo komponentė, normavimo komponentė ir duomenų peržiūros komponentė.
3. Pateiktame scenarijuje duomenims įkelti naudojama duomenų įkėlimo komponentė **Įkelti failą iš MIDAS**. Du kartus kliktelėjus ant komponentės, atidaromas failo įkėlimo langas. Pasirinkus tinkamą analizei failą, spaudžiamas mygtukas **Patvirtinti**. Jei duomenų failas jau yra DAMIS įrankyje, duomenų įkėlimui galima naudoti ir kitą komponentę **Pasirinkti įkeltą failą**.
4. Visos komponentės eksperimento atlikimo tvarka sujungiamos į vieną seką, duomenų įkėlimo komponentėje paspaudžiamas pelyte pilnavidūris skrituliukas ir tempiamas iki normavimo komponentės tuščiavidūrio skrituliuko. Analogiškai sujungiamos ir normavimo bei matricinio vaizdavimo komponentės.
5. Du kartus kliktelėjus ant normavimo komponentės atidaromas normavimo būdo pasirinkimo langas (148 pav.). Galimi du normavimo būdai: pagal vidurkį ir dispersiją, bei normavimas į nurodytą intervalą. Pasirinkus tinkamą normavimo būdą, spaudžiamas mygtukas **Patvirtinti**.

**Normavimas**

Nurodykite algoritmo parametrus

**Pasirinkite normavimo būdą**

**Vidurkis a, dispersija b**  **Intervalas [a;b]**

**a\***

**b\***

\* Žvaigždute pažymėti laukai yra privalomi

---

**148 pav. Normavimo būdo pasirinkimo langas**

6. Sukūrus eksperimento seką, spaudžiamas mygtukas **Vykdyti**. Sukurtam eksperimentui visada suteikiamas vardas pagal nutylėjimą, bet tyrėjas gali eksperimentą pavadinti savaip, bei Eksperimento nustatymų lange nustatyti eksperimento atlikimo parametrus. Pasirinkus vykdymo parametrus, spaudžiamas mygtukas **Patvirtinti**.

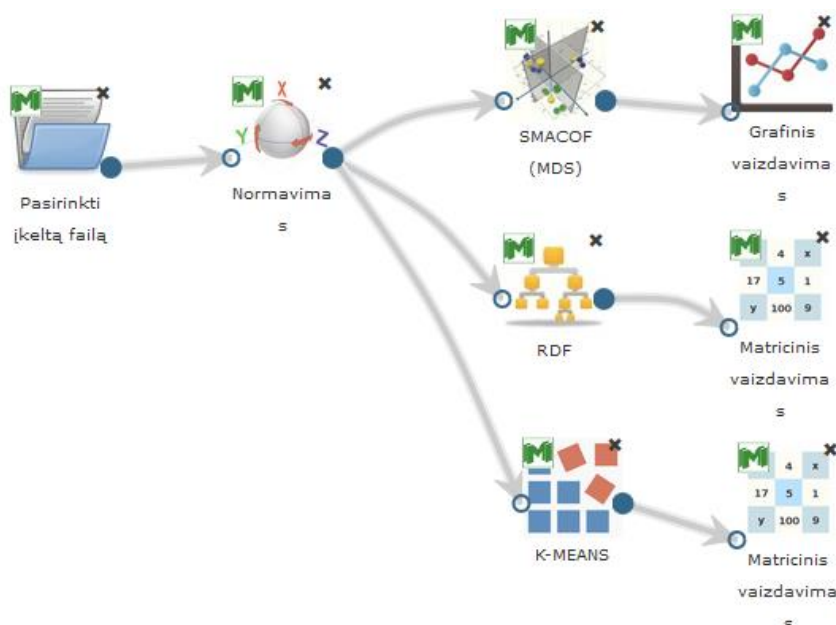
7. Norint peržiūrėti, koks sukurto eksperimento statusas, spaudžiama nuoroda viršutinėje meniu juostoje „**Eksperimentai**“
8. Kai sukurto eksperimento statusas tampa „**Įvykdytas**“, galima peržiūrėti ir atsisiųsti gautus rezultatus. Paspaudus akutę, atidaromas Eksperimento langas.

Matricinis vaizdavimas			
Objekto nr.	PELVIC_INCIDENCE (REAL)	PELVIC_TILT (REAL)	LUMBAR_LORDOSIS_ (REAL)
1.	0.150544	0.501111	-0.670474
2.	-1.25349	-0.748891	-1.46961
3.	0.490276	0.468085	-0.0964533
4.	0.517806	0.71128	-0.413041
5.	-0.629669	-0.789923	-1.28886
6.	-1.18378	-0.362581	-1.46413
7.	-0.411771	-0.168425	-0.804119
8.	-0.883882	-0.678834	-1.24942

**149 pav. Normavimo rezultatai**

Įvykdyto eksperimento rezultatas pateiktas 149 pav. Čia šešių matuotų parametų normuotos reikšmės su kuriomis toliau galime tęsti eksperimentus, gautus rezultatus galima išsisaugoti savo kompiuteryje arba MIDAS archyve.

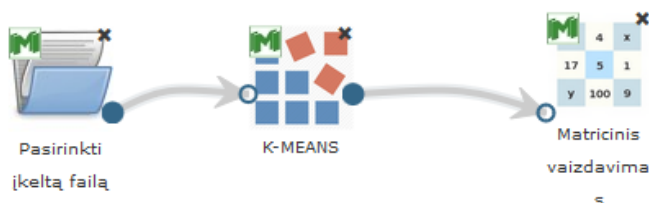
Galimos eksperimentų sekos su normuotais duomenimis pateiktos 150 pav.



150 pav. Duomenų analizė su normuotais duomenimis: vizualizavimas, klasifikavimas bei grupavimas

### 3.11.3 Stuburo duomenų grupavimosi tendencijos

Tyrinėjant nežinomus duomenis labai naudinga panaudojus grupavimo algoritmus pažiūrėti, į kokias grupes algoritmai tiriamus objektus suskirsto ir kokios tendencijos būdingos sugrupuotiems objektams. Duomenų grupavimui naudojamas eksperimento scenarijus pateikiamas 151 pav.



151 pav. Grupavimo eksperimento užduočių seka

1. Viršutinėje meniu juostoje spaudžiama nuoroda **Kurti eksperimentą**. Atidaromas naujo eksperimento kūrimo langas.
2. Tempiant į darbalaukį įkeliamos eksperimentui atlikti reikalingos komponentės: Duomenų įkėlimo komponentė, grupavimo k-means komponentė ir duomenų peržiūros komponentė.
3. Pateiktame scenarijuje duomenims įkelti naudojama duomenų įkėlimo komponentė **Pasirinkti įkeltą failą**. Du kartus kliktelėjus ant komponentės, atidaromas failo įkėlimo langas. Pasirinkus tinkamą analizei failą, spaudžiamas mygtukas **Patvirtinti**.

4. Visos komponentės eksperimento atlikimo tvarka sujungiamos į vieną seką, duomenų įkėlimo komponentėje paspaudžiamas pelyte pilnavidūris skrituliukas ir tempiamas iki grupavimo komponentės tuščiavidūrio skrituliuko. Analogiškai sujungiamos ir grupavimo bei matricinio vaizdavimo komponentės.

**K-MEANS**

Nurodykite algoritmo parametrus

**Maksimalus iteracijų skaičius\***

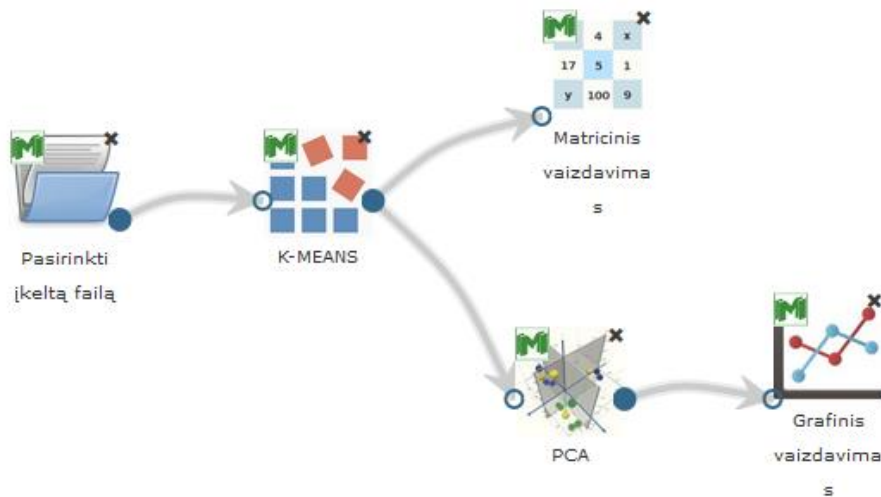
**Klasterių skaičius\***

\* Žvaigždute pažymėti laukai yra privalomi

#### 152 pav. Grupavimo algoritmo parametrų parinkimo langas

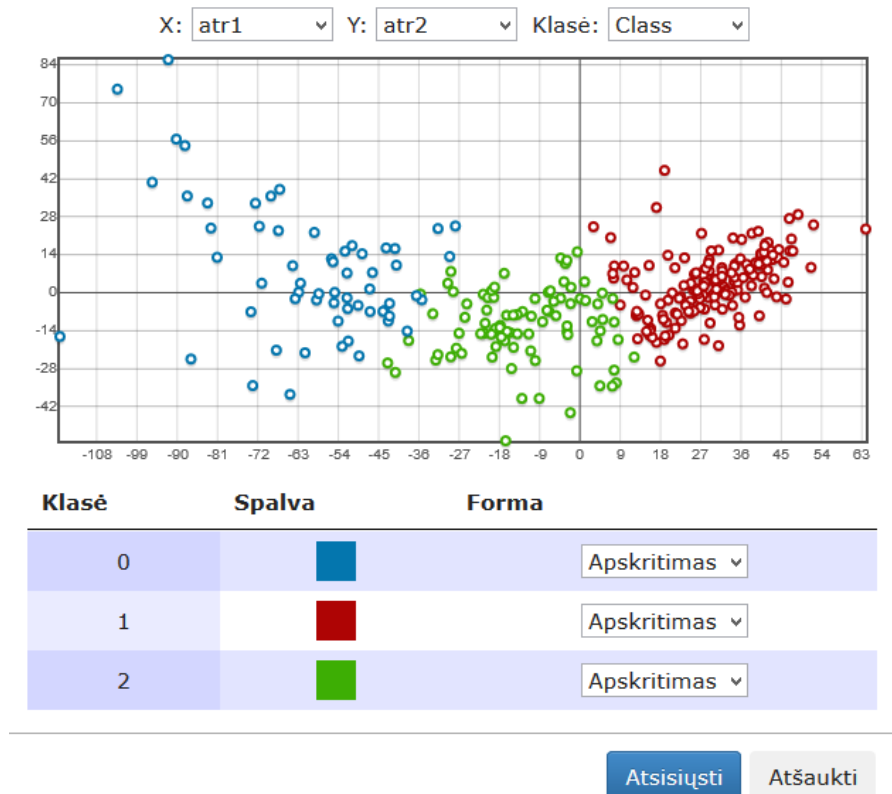
5. Du kartus kliktelėjus ant grupavimo komponentės atidaromas grupavimo būdo pasirinkimo langas (152 pav.). Galima nurodyti maksimalų iteracijų skaičių, bei į kelis klasterius duomenis sugrupuoti. Pasirinkus tinkamus grupavimo algoritmo parametrus, spaudžiamas mygtukas **Patvirtinti**.
6. Sukūrus eksperimento seką, spaudžiamas mygtukas **Vykdyti**. Sukurtam eksperimentui visada suteikiamas vardas pagal nutylėjimą, bet tyrėjas gali eksperimentą pavadinti savaip, bei Eksperimento nustatymų lange nustatyti eksperimento atlikimo parametrus. Šis scenarijus pavadintas **Stuburo\_duom\_grupavimas**. Pasirinkus vykdymo parametrus, spaudžiamas mygtukas **Patvirtinti**.
7. Norint peržiūrėti, koks sukurto eksperimento statusas, spaudžiama nuoroda viršutinėje meniu juostoje „**Eksperimentai**“
8. Kai sukurto eksperimento statusas tampa „**Įvykdytas**“, galima peržiūrėti ir atsisiųsti gautus rezultatus. Paspaudus akutę, atidaromas Eksperimento langas.

Norint iš kart pamatyti duomenų grupavimosi tendencijas eksperimento užduočių seką galima praplėsti, papildant ją dimensijos mažinimo PCA komponente, kuri sugrupuotus duomenis suspaus į dvimatę erdvę, ir grafinio vaizdavimo komponente, kuri atvaizduos gautus rezultatus. Pakeistas eksperimento užduočių scenarijus pateiktas 153 pav.



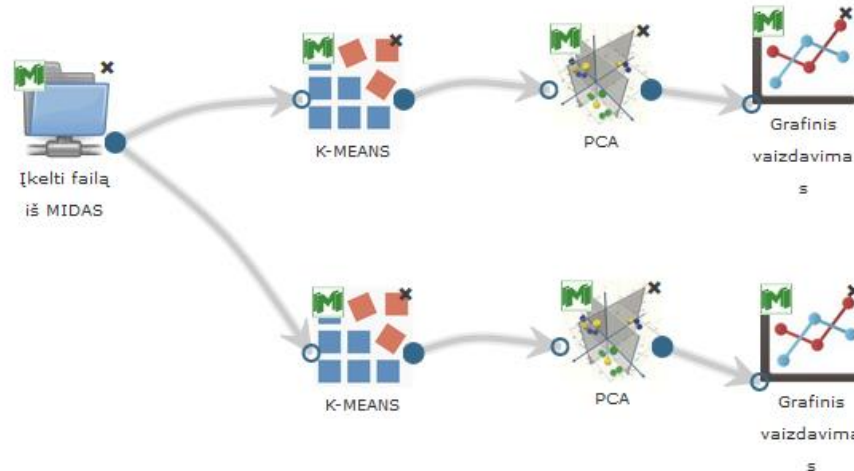
153 pav. Praplėsta grupavimo eksperimento užduočių seka

Atlikus eksperimentą pagal duotą scenarijų, gauname sugrupuotų stuburo duomenų vaizdą, atvaizduotą plokštumoje, kur vieną šešiamatį objektą atitinka vienas taškas plokštumoje. Gautas grafikas pateiktas 154 pav.



154 pav. Grupavimo rezultatai atvaizduoti plokštumoje

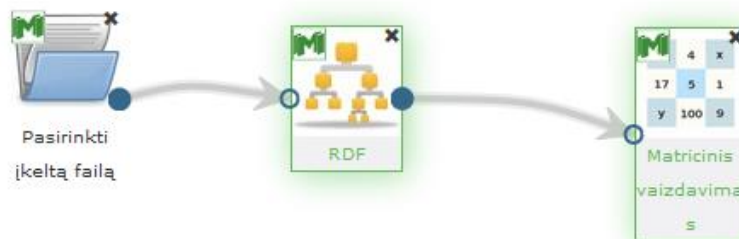
Dabar matydami algoritmo sugrupuotus objektus, galima nagrinėti kiekvieną išskirtą grupę atskirai, ieškoti kiekvienai išskirtai grupei būdingų požymių. Duomenų aibes galima nagrinėti įvairiais atpektais, sukurti scenarijų, kur iš karto aibę galima grupuoti į du, tris ar daugiau klasterių, ir palyginti gautus rezultatus. Tam galima naudoti 155 pav. pateiktą scenarijų **Grupavimosi\_tendencijos**.



155 pav. Grupavimo ir vizualizavimo rezultatai: Stuburo duomenys grupuoti į 2 ir 3 klasterius

### 3.11.4 Stuburo duomenų klasifikavimas

Stuburo duomenų aibės objektai sudaryti iš šešių matuotų parametų reikšmių ir objektui priskirtos klasės. Viso objektai suskirstyti į tris klases. 156 pav. pateikiamas scenarijus **Stuburo duomenų klasifikavimas**, kurį įvykdžius klasifikatorius apmokomas turimais duomenimis ir apmokytas klasifikatorius klasifikuoja duomenų aibėje esamus nežinomus duomenis (tuos, kuriems klasė nebuvo priskirta).



156 pav. Klasifikavimo eksperimento scenarijus

1. Viršutinėje meniu juostoje spaudžiama nuoroda **Kurti eksperimentą**. Atidaromas naujo eksperimento kūrimo langas.
2. Tempiant į darbalaukį įkeliamos eksperimentui atlikti reikalingos komponentės: Duomenų įkėlimo komponentė, Klasifikavimo RDF komponentė ir duomenų peržiūros komponentė Matricinis vaizdavimas.

3. Pateiktame scenarijuje duomenims įkelti naudojama duomenų įkėlimo komponentė **Pasirinkti įkeltą failą**. Du kartus kliktelėjus ant komponentės, atidaromas failo įkėlimo langas. Pasirinkus tinkamą analizei failą, spaudžiamas mygtukas **Patvirtinti**. Pasirenkamas failas **stuburas\_be\_klasiu.arff**, kuris jau yra įkeltas DAMIS įrankyje. Šioje duomenų aibėje yra objektai, kurių klasės iš anksto žinomos, ir objektai, kuriems klasė dar nepriskirta.
4. Visos komponentės eksperimento atlikimo tvarka sujungiamos į vieną seką, duomenų įkėlimo komponentėje paspaudžiamas pelyte pilnavidūris skrituliukas ir tempiamas iki klasifikavimo komponentės tuščiavidūrio skrituliuko. Analogiškai sujungiamos ir klasifikavimo bei matricinio vaizdavimo komponentės.
5. Du kartus kliktelėjus ant klasifikavimo komponentės atidaromas klasifikavimo algoritmo parametrų nustatymo langas (157 pav.). Galima nurodyti atsparumo triukšmui parametras, bei mokymo ir testavimo aibių dydžius procentais. Pasirinkus tinkamus klasifikavimo algoritmo parametrus, spaudžiamas mygtukas **Patvirtinti**.

**RDF**

Nurodykite algoritmo parametrus

**Atsparumas triukšmui\***

**Mokymo objektų kiekis\***  %

**Testavimo objektų kiekis\***  %

\* Žvaigždute pažymėti laukai yra privalomi

**Patvirtinti** **Atšaukti**

157 pav. Klasifikavimo algoritmo parametrų nustatymo langas

6. Sukūrus eksperimento seką, spaudžiamas mygtukas **Vykdyti**. Sukurtam eksperimentui visada suteikiamas vardas pagal nutylėjimą, bet tyrėjas gali eksperimentą pavadinti savaip, bei Eksperimento nustatymų lange nustatyti eksperimento atlikimo parametrus. Šis scenarijus pavadintas **Stuburo\_duomenų\_klasifikavimas**. Pasirinkus vykdymo parametrus, spaudžiamas mygtukas **Patvirtinti**.

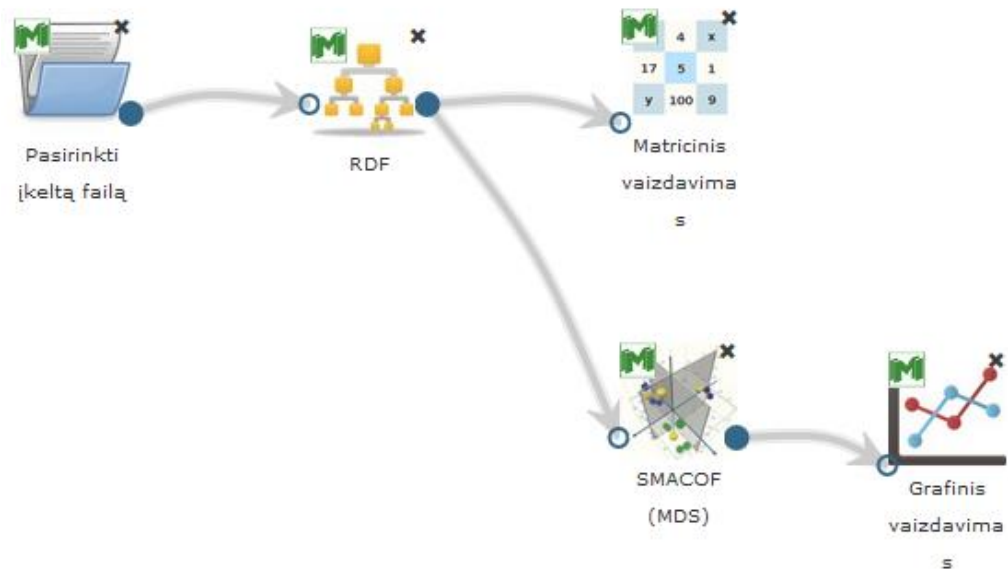
7. Norint peržiūrėti, koks sukurto eksperimento statusas, spaudžiama nuoroda viršutinėje meniu juostoje „**Eksperimentai**“
8. Kai sukurto eksperimento statusas tampa „**Ivykdytas**“, galima peržiūrėti ir atsisiųsti gautus rezultatus. Paspaudus akutę, atidaromas Eksperimento langas. Tuomet galima peržiūrėti gautus klasifikavimo rezultatus atsidarius matricinio vaizdavimo komponentę, du kartus ant jos kliktelėjus (158 pav.). Atidarytame lange pateikiami klasifikatoriaus rezultatai, kur visiems objektams nurodomos kiekvienos klasės tikimybės ir klasifikatoriaus priskirta klasė.

Matricinis vaizdavimas					
atr5 (REAL)	atr6 (REAL)	probability_1 (REAL)	probability_2 (REAL)	probability_3 (REAL)	CLASS (1, 2, 3)
.67	-0.25	0.826667	0.0133333	0.16	1
4.41	4.56	0.92	0	0.08	1
5.99	-3.53	0.84	0.0533333	0.106667	1
1.87	11.21	0.786667	0.133333	0.08	1
8.17	7.92	0.893333	0.0666667	0.04	1
0.33	2.23	0.813333	0	0.186667	1
0.57	5.99	0.746667	0.0533333	0.2	1
7.27	-10.68	0.946667	0	0.0533333	1

158 pav. Klasifikavimo rezultatai

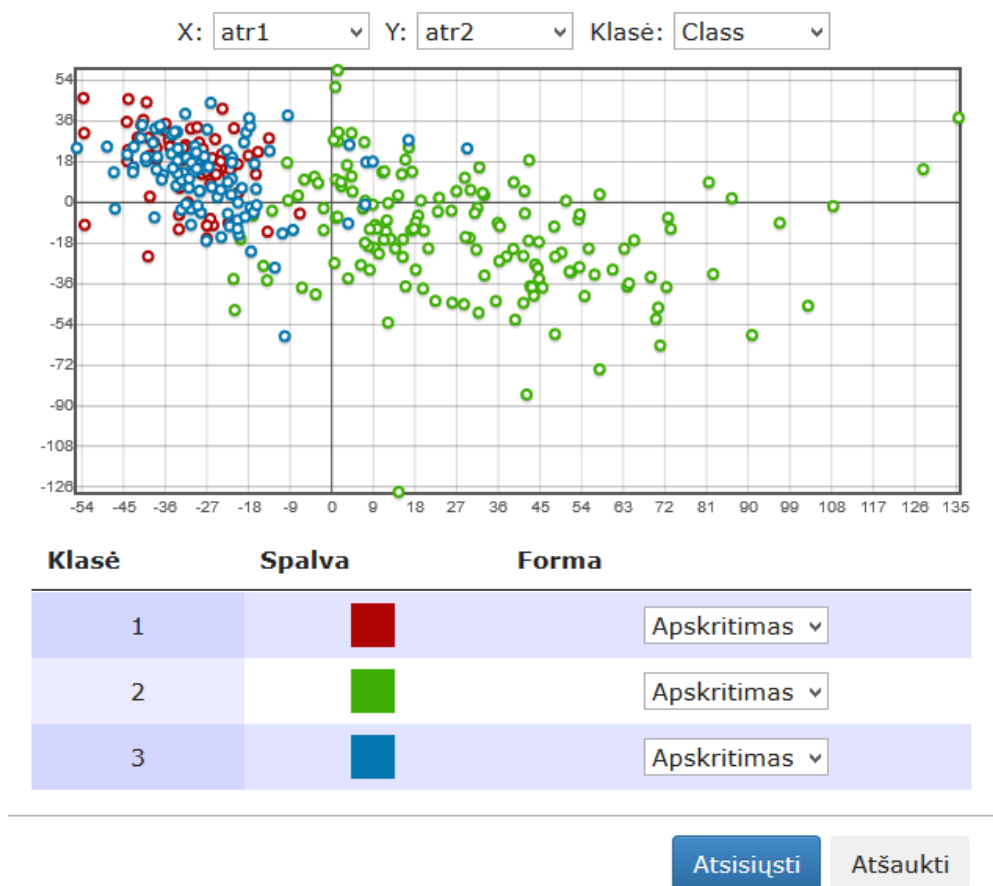
Norint iš kart pamatyti duomenų klasifikavimo rezultatus ir pažiūrėti, kaip klasės atvaizduojamos plokštumoje, eksperimento užduočių seką galima praplėsti, papildant ją dimensijos mažinimo komponente (šiam scenarijuje naudosime SMACOF komponentę), kuri klasifikuojamus duomenis suspaus į dvimatę erdvę, ir grafinio vaizdavimo komponente, kuri atvaizduos gautus rezultatus. Pakeistas eksperimento užduočių scenarijus pateiktas 159 pav.





159 pav. Klasifikavimo ir vizualizavimo scenarijus

Klasifikavimo ir vizualizavimo rezultatai pateikti 160 pav. Kaip matyti pateiktame paveiksle, pirma ir trečia klasės persidengia, o antra klasė atsiskiria. Galima daryti išvadą, kad pirmos ir trečios klasės objektai pagal matuotus parametrus labai panašūs.



160 pav. Klasifikavimo ir vizualizavimo rezultatai

### 3.12 Langai „Pagalba“ ir „D.U.K.“

Norėdamas peržiūrėti eksperimentų naudojimo instrukciją, naudotojas turi paspausti meniu skiltį „Pagalba“. Tuomet atsidarys svetainės langas, kuriame aprašyta kiekviena komponentė ir jos paskirtis; be to, aprašytas elementarus eksperimento planavimo scenarijus bei kaip įvykdyti eksperimentą.

Taip pat naudotojas gali perskaityti dažniausiai užduodamus klausimus. Paspaudus „D.U.K.“ meniu skiltį atidaromas langas, kuriame pateikti dažniausiai kylantys klausimai planuojant savo eksperimentus arba ruošiantis peržiūrėti apskaičiuotus eksperimento rezultatus.

Toliau pateikti dažniausiai užduodami klausimai ir atsakymai.

#### ***Kas yra DAMIS?***

DAMIS (duomenų analizės įrankis) – tai atvira mokslo infrastruktūra, skirta duomenų analizei atlikti. Įrankio paskirtis – sudaryti galimybę specializuotoje aplinkoje bendradarbiaujantiems mokslininkams ar jų grupėms atlikti pagrindinius duomenų analizės tyrimus (grupavimą, klasifikavimą ir kt.) skaičiavimo poreikius atitinkančioje aplinkoje; vizualios analizės priemonėmis tirti daugiamačių duomenų projekcijas į plokštumą, duomenų grupavimąsi, duomenų panašumus, atskirų daugiamačių duomenų požymių įtaką ir tarpusavio priklausomybes; stebėti bei apdoroti vizualizacijos ar našiųjų skaičiavimų aplinkoje gautus tyrimų rezultatus.

### ***Kokie duomenų analizės metodai yra prieinami DAMIS įrankyje?***

DAMIS įrankyje šiuo metu yra prieinami šie duomenų analizės metodai:

- Pagrindinių komponentių analizės (PCA) algoritmas.
- Daugiamatčių skalių grupei (MDS) priklausantis klasikinis SMACOF algoritmas.
- SMACOF algoritmo Zeidelio modifikacija.
- Diagonalinis mažoravimo algoritmas (DMA).
- Santykinės daugiamatės skalės.
- Dirbtiniais neuroniniais tinklais ir daugiamatėmis skalėmis grindžiamas SAMANN algoritmas.
- Saviorganizuojančiais neuroniniais tinklais (SOM) grindžiamas algoritmas.
- MDS ir SOM junginys.
- Daugiasluoksnis perceptronas grindžiamas klaidos sklidimo atgal taisykle.
- RDF (random decision forest) algoritmas.
- K-vidurkių (k-means) grupavimo algoritmas.

### ***Kaip eksperimente panaudoti prieš tai įkeltą duomenų failą?***

Tam skirta komponentė „Pasirinkti įkeltą failą“. Pirmiausia šią komponentę nutempkite į eksperimento planavimo langą, du kartus paspaudę šią komponentę, pasirinkite norimą failą. Tuomet šią komponentę galima naudoti darbų sekai sudaryti.

### ***Kokiais formatais duomenys turi būti pateikiami sistemai?***

DAMIS įrankiui duomenys gali būti pateikiami šiais formatais: arff, tab, txt, csv, xml, xls, zip. Taip pat naudotojas turi galimybę redaguoti arba ištrinti jau įkeltus failus. Pasirinkus failų valdymo skiltį yra rodomas jau įkeltų duomenų failų sąrašas, kuriame pateiktas kiekvieno saugomo failo pavadinimas ir dydis. Failų sąrašą galima rūšiuoti pagal duomenų failų pavadinimus, įkėlimo datą. Įkeltus ir gautus rezultatų failus po darbų sekų įvykdymo naudotojas gali atsiųsti į savo kompiuterį arba patalpinti MIDAS archyve šiais formatais: arff, zip, tab, csv, xls, xlsx.

### ***Kodėl darbui su DAMIS geriau pasirinkti ARFF formato failą ir kaip tinkamai aprašyti duomenis?***

Norint išvengti klaidų ir kitų nepatogumų, patartina pasirinkti ARFF formato failą. ARFF (Attribute-Relation File Format) failas – ASCII tekstinis failas, kuriame objektai, sudarantys konkrečią analizuojamų objektų aibę, yra apibūdinami bendrais požymiais. ARFF failas sudarytas iš dviejų dalių: antraštė ir duomenys.

ARFF failo antraštę sudaro: informacija apie duomenų aibės pavadinimą (@RELATION), požymių sąrašas (@ATTRIBUTE) ir jų tipai (NUMERIC (real arba integer), STRING, DATA, <galimų\_reikšmių\_aibė>):

```
@RELATION iris
@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

ARFF failo duomenys aprašomi, kaip pavaizduota pavyzdyje:

```
@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

Trūkstamos reikmės keičiamos klaustuko ženklų „?“:

@DATA  
?,3.5,?,0.2,Iris-setosa

### ***Kaip peržiūrėti eksperimentų rezultatus?***

Norint peržiūrėti gautus eksperimentų rezultatus, prie eksperimento darbų sekos komponentės, kurios rezultatus norima pamatyti, reikia prijungti vieną iš rezultatų peržiūros komponentių:

- „Techninė informacija“ – tai komponentė, skirta techninei informacijai peržiūrėti: skaičiavimo laikui, gautos paklaidos ir kt.
- „Matricinis vaizdavimas“ – tai komponentė, kurios pagalba naudotojas galės peržiūrėti duomenis ar algoritmų rezultatus matricos (lentelės) pavidalu, kai ši komponentė bus sujungta su failo ar algoritmo komponente.
- „Grafinis vaizdavimas“ – tai komponentė, kurios pagalba naudotojas galės peržiūrėti duomenų (rezultatų) grafiką, kai ši komponentė bus sujungta su failo ar algoritmo komponente.

Šios komponentės turi galimybę atsiųsti rezultatus.

### ***Kaip atlikti eksperimentą?***

Norint atlikti eksperimentą su norima duomenų aibe, reikia sukurti norimą darbų seką ir ją įvykdyti. Tam reikia nutempti į eksperimento planavimo darbalaukį norimas komponentes, užpildyti reikiamų valdymo parametrų reikšmes, sujungti reikiamas komponentes ir paspausti mygtuką „vykdyti“. Tuomet atidaromas eksperimento vykdymo parametrų užpildymo langas. Užpildžius reikiamus laukus ir paspaudus mygtuką „Patvirtinti“, vyksta sekos patikrinimas: tikrinami kiekvienos komponentės įvesti valdymo parametrai ir kiekvienos komponentės sujungimas. Jei sekoje rastos klaidos, rodomas klaidos pranešimas prie nesujungtos komponentės arba pažymima raudonai ta komponentė, kurioje neteisingai įvesti valdymo parametrai. Priešingu atveju, jei komponentės sujungtos teisingai ir valdymo parametrai užpildyti tinkamai, tuomet eksperimentų istorijoje atsiranda naujas eksperimentas, kurio statusas „Vykdomas“.

### ***Kaip sužinoti ar jau suformuoti eksperimento rezultatai?***

Eksperimento planavimo lange sukūrus norimą darbų seką ir paspaudus mygtuką „Vykdyti“, eksperimentų istorijoje atsiranda naujas eksperimentas su nauju pavadinimu ir statusu „Vykdomas“. Kai eksperimentas įgauna statusą „Įvykdytas“, eksperimentų rezultatai bus prieinami peržiūrai.

### ***Ar naudojantis DAMIS galima analizuoti duomenų failą, kuriame yra praleistų reikšmių?***

Jeigu duomenų failas turi praleistų reikšmių, tai norint jį analizuoti dimensijos mažinimo, klasifikavimo arba grupavimo metodais, duomenys turi būti išvalyti, prie duomenų įkėlimo komponentės prijungus komponentę „Valymas“.

***Pamiršau prisijungimo slaptažodį. Ką man daryti?***

Pamiršus slaptažodį jį galima atstatyti. Prisijungimo lange paspauskite nuorodą „Pamiršau slaptažodį“, įveskite el. pašto adresą, kuriuo registravotės, ir sekite tolimesnius nurodymus.

***Sėkmingai įvykdžius eksperimentą, noriu sužinoti algoritmo veikimo laiką. Kaip tai padaryti?***

DAMIS yra realizuota galimybė peržiūrėti duomenų analizės algoritmo veikimo technines charakteristikas: veikimo laiką, gautą paklaidą ir kitas charakteristikas. Norint pamatyti šią informaciją prie duomenų analizės komponentės reikia prijungti komponentę „Techninė informacija“, esančią rezultatų peržiūros dalyje.