



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

**Trečia užduotis (Daugiasluoksnis perceptronas naudojant sistemą
KNIME)**

Laboratorinio darbo ataskaita

Atliko: Klaidas Kubilius

VU el. p.: klaidas.kubilius@mif.stud.vu.lt

Vilnius
2025

1. DARBO EIGA

1.1. Duomenys

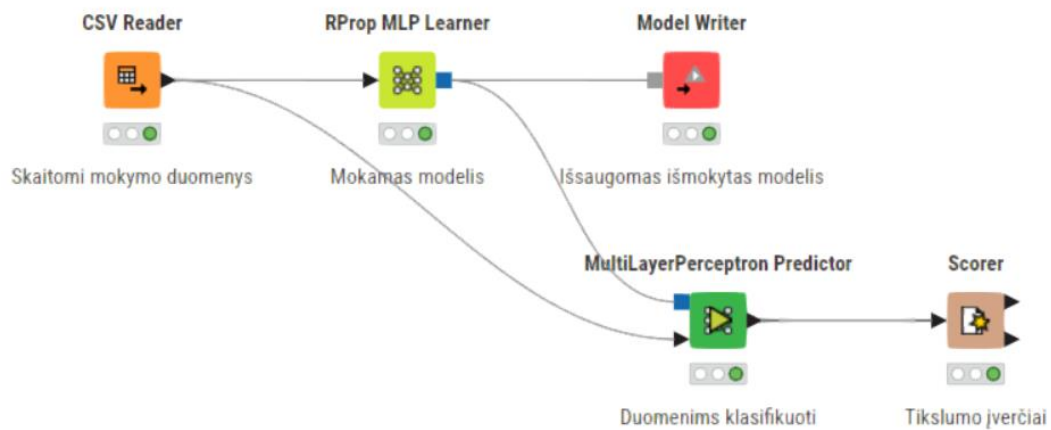
Naudojama Iris duomenų rinkinys, kuriame yra 150 įrašų ir 3 klasės. Taip pat priklauso 4 požymiai (neįskaitant klasės). Taurėlapio ilgis, plotis ir žiedlapio ilgis, plotis. Duomenys buvo padalinami naudojant „absolute 120“ ir „stratified sampling“ nustatymus. Buvo padalinta į mokymo duomenis, po 40 įrašų kiekvienai klasei (iš viso 120) ir testavimo po 10 įrašus kiekvienai klasei.



1.1 pav. Užduočių seka, skirta duomenų dalijimui

1.2. KNIME užduočių seka daugiasluoksniui perceptronui apmokyti

Pagal mokymo duomenis buvo apmokytas modelis klasifikavimui, naudojama buvo „RProp MLP Learner“.



1.2 pav. Užduočių seka daugiasluoksniui perceptronui mokyti ir modeliui išsaugoti

Klasifikavimo lentelė(confusion matrix) ir klasifikavimo tikslumo metrikos.

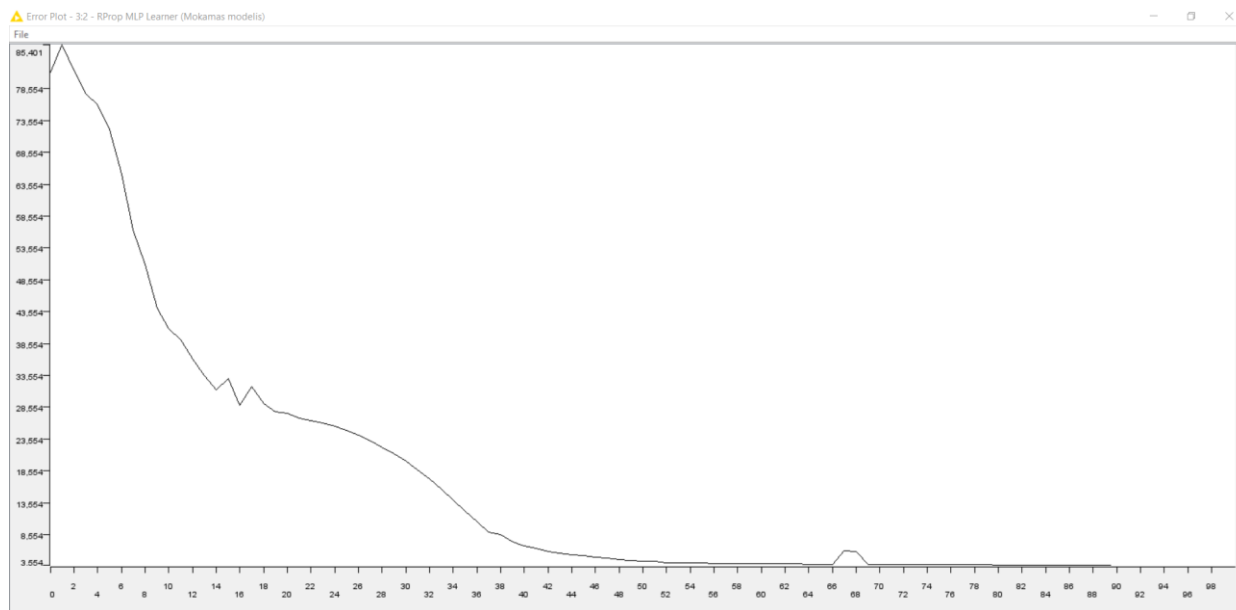
1.1 lentelė. Klasifikavimo lentelė pagal mokymo duomenis

Tikroji / Prognozė	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	40	0	0
Iris-versicolor	0	38	2
Iris-virginica	0	1	39

1.2 lentelė. Klasifikavimo tikslumo reikšmės pagal mokymo duomenis

Metrika	Reikšmė
Tikslumas (Accuracy)	0,975
Precision (macro)	0,975
Sensitivity	0,975
Recall	0,975
F-measure	0,975

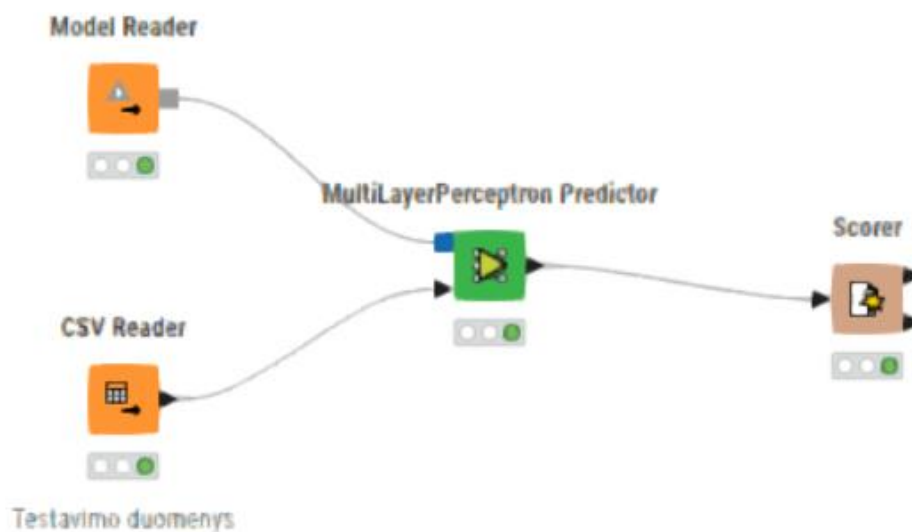
Mokymo metu paklaidos kitimo grafikas.



1.3 pav. Mokymo metu paklaidos kitimo grafikas

1.3. KNIME užduočių seka daugiasluoksniam perceptronui testuoti

Testuoti modelio tikslumą naudojamas testavimo rinkinys.



1.4 pav. Užduočių seka daugiasluoksniam perceptronui mokyti ir modeliui išsaugoti

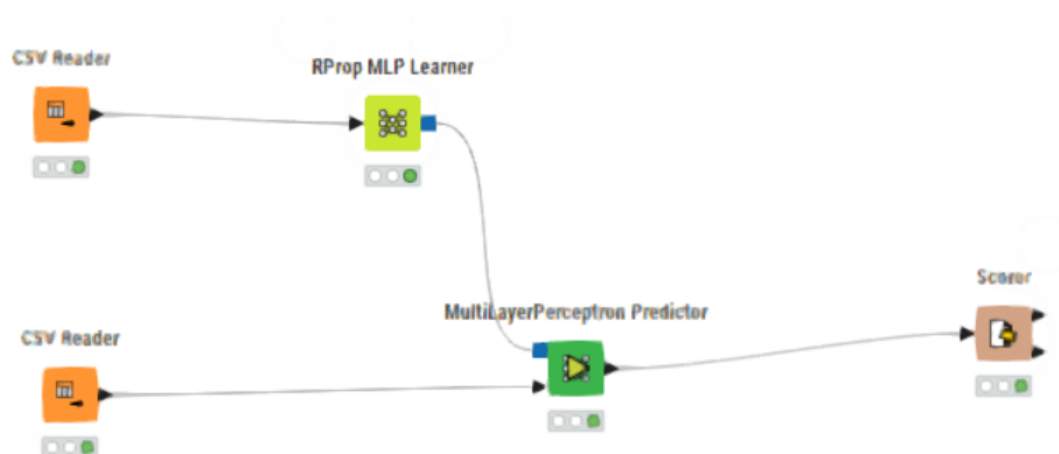
1.3 lentelė. Klasifikavimo lentelė pagal testavimo duomenis

Tikroji / Prognozė	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	10	0	0
Iris-versicolor	0	9	1
Iris-virginica	0	0	10

1.4 lentelė. Klasifikavimo tikslumo reikšmės pagal testavimo duomenis

Metrika	Reikšmė
Tikslumas (Accuracy)	0,967
Precision (macro)	0,97
Sensitivity	0,967
Recall	0,967
F-measure	0,967

1.4. Bendra KNIME užduočių seka daugiasluoksniui perceptronui mokyti ir testuoti



1.5 pav. Bendra KNIME užduočių seka daugiasluoksniui perceptronui mokyti ir testuoti

Buvo eksperimentuota su šiais hiperparametrų rinkiniais:

- 1 rinkinys – 100 iteracijų; 1 paslėptasis sluoksnis; 10 paslėptieji neuronai per sluoksnį.
- 2 rinkinys – 100 iteracijų; 10 paslėptasis sluoksnis; 50 paslėptieji neuronai per sluoksnį.
- 3 rinkinys – 300 iteracijų; 10 paslėptasis sluoksnis; 100 paslėptieji neuronai per sluoksnį.

1.5 lentelė. Klasifikavimo tikslumo reikšmės su skirtingais hiperparametrų rinkiniais

Metrika / Hiperparametrų rinkinys	1 rinkinys	2 rinkinys	3 rinkinys
Tikslumas (Accuracy)	0,967	0,967	0,933
Precision	0,97	0,97	0,933

Sensitivity	0,967	0,967	0,933
Recall	0,967	0,967	0,933
F-measure	0,967	0,967	0,933

Skirtomas tarp tikslumų buvo neryškus, tik labai didėjant sluoksnių skaičiui tikslumas pradėjo kristi. Optimaliausias pasirinktas 1 rinkinys, nes kai kuriais atvejais tikslumas pasiekia 100 %.

1.6 lentelė. Klasių prognozės ir tikimybės

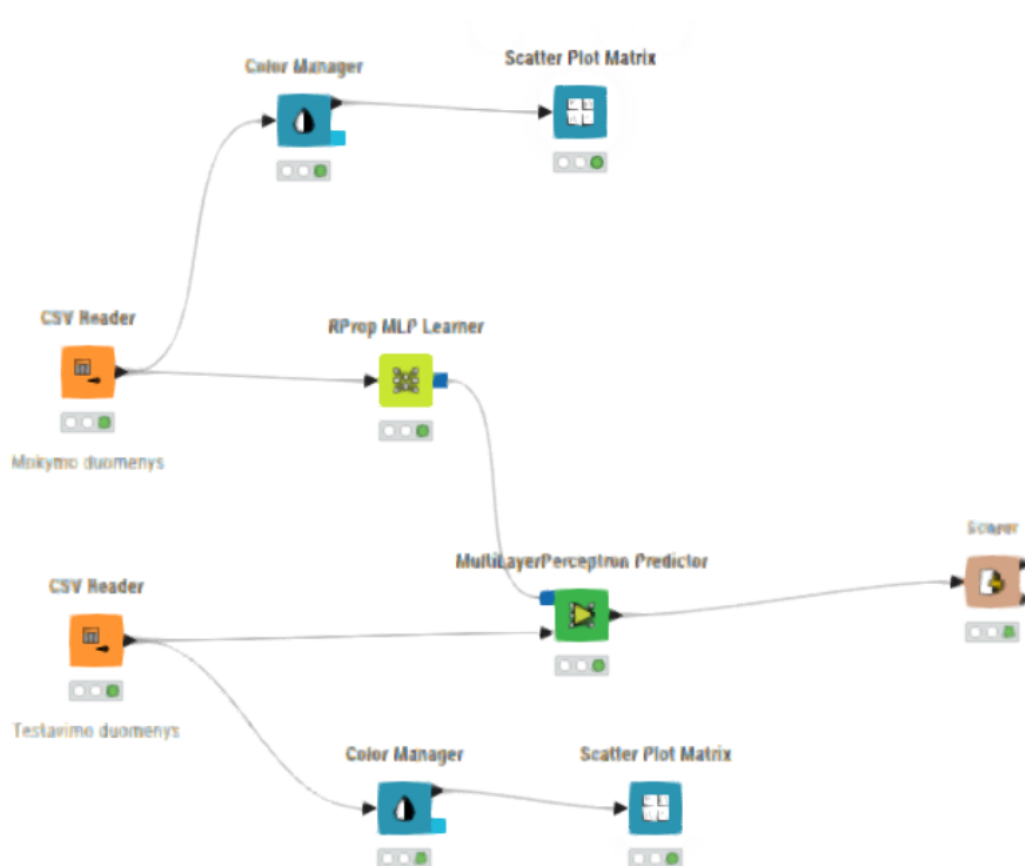
Sep len	Sep wid	Pet len	Pet wid	True class	Pred setosa	Pred versicolor	Pred virginica	Pred class
4,7	3,2	1,3	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,0	3,4	1,5	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
4,6	3,6	1,0	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,0	3,0	1,6	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,2	3,5	1,5	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
4,8	3,1	1,6	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,4	3,4	1,5	0,4	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,0	3,5	1,6	0,6	Iris-setosa	1,0	0,507	0,0	Iris-setosa
5,1	3,8	1,6	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
5,0	3,3	1,4	0,2	Iris-setosa	1,0	0,0	0,0	Iris-setosa
6,4	3,2	4,5	1,5	Iris-versicolor	0,0	0,998	0,0	Iris-versicolor
6,0	2,2	4,0	1,0	Iris-versicolor	0,0	1,0	0,0	Iris-versicolor
6,1	2,9	4,7	1,4	Iris-versicolor	0,0	0,98	0,016	Iris-versicolor

5,6	3,0	4,5	1,5	Iris- versicolo r	0,0	0,929	0,019	Iris- versicolo r
6,2	2,2	4,5	1,5	Iris- versicolo r	0,0	0,896	0,13	Iris- versicolo r
5,9	3,2	4,8	1,8	Iris- versicolo r	0,0	0,229	0,549	Iris- virginica
6,1	2,8	4,0	1,3	Iris- versicolo r	0,0	1,0	0,0	Iris- versicolo r
5,8	2,7	3,9	1,2	Iris- versicolo r	0,0	1,0	0,0	Iris- versicolo r
6,3	2,3	4,4	1,3	Iris- versicolo r	0,0	0,997	0,002	Iris- versicolo r
5,6	3,0	4,1	1,3	Iris- versicolo r	0,0	1,0	0,0	Iris- versicolo r
6,3	2,9	5,6	1,8	Iris- virginica	0,0	0,011	0,984	Iris- virginica
7,3	2,9	6,3	1,8	Iris- virginica	0,0	0,01	0,986	Iris- virginica
6,5	3,2	5,1	2,0	Iris- virginica	0,0	0,06	0,888	Iris- virginica
6,4	3,2	5,3	2,3	Iris- virginica	0,0	0,003	0,99	Iris- virginica
6,1	3,0	4,9	1,8	Iris- virginica	0,0	0,155	0,745	Iris- virginica
6,4	2,8	5,6	2,1	Iris- virginica	0,0	0,002	0,995	Iris- virginica
7,2	3,0	5,8	1,6	Iris- virginica	0,0	0,334	0,835	Iris- virginica
7,7	3,0	6,1	2,3	Iris- virginica	0,0	0,003	0,994	Iris- virginica
6,0	3,0	4,8	1,8	Iris- virginica	0,0	0,184	0,653	Iris- virginica
6,5	3,0	5,2	2,0	Iris- virginica	0,0	0,027	0,953	Iris- virginica

1.5. Duomenų pavaizdavimas XY koordinatinių sistemoje

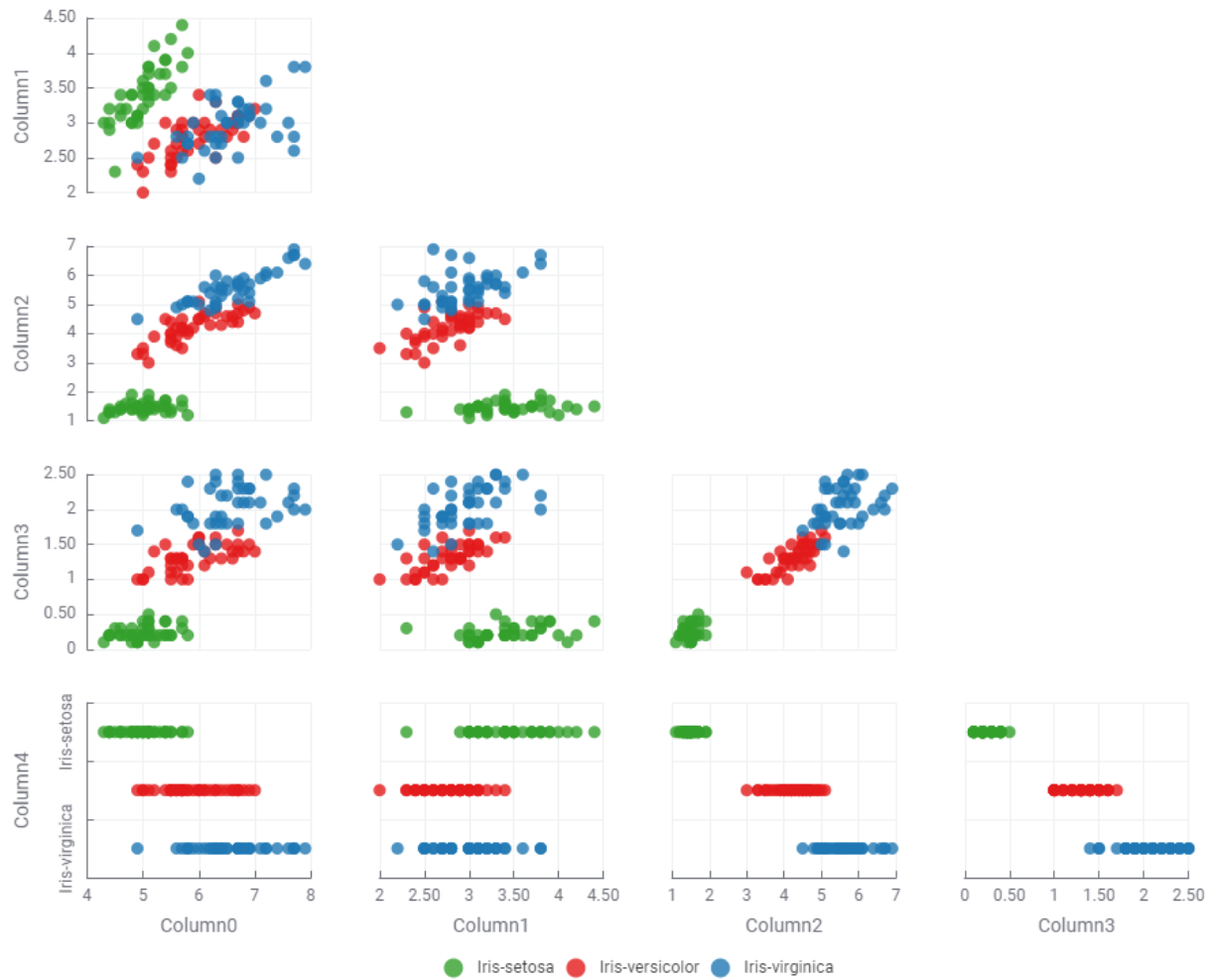
Mokymo ir testavimo duomenims pavaizduoti XY koordinatinių sistemoje buvo naudota „color manager“ ir „scatter plot matrix“. Taškų dydis 9. Spalvos naudojamos klasėms atskirti:

- Iris-setosa – žalia
- Iris-versicolor – raudona
- Iris-virginica – mėlyna



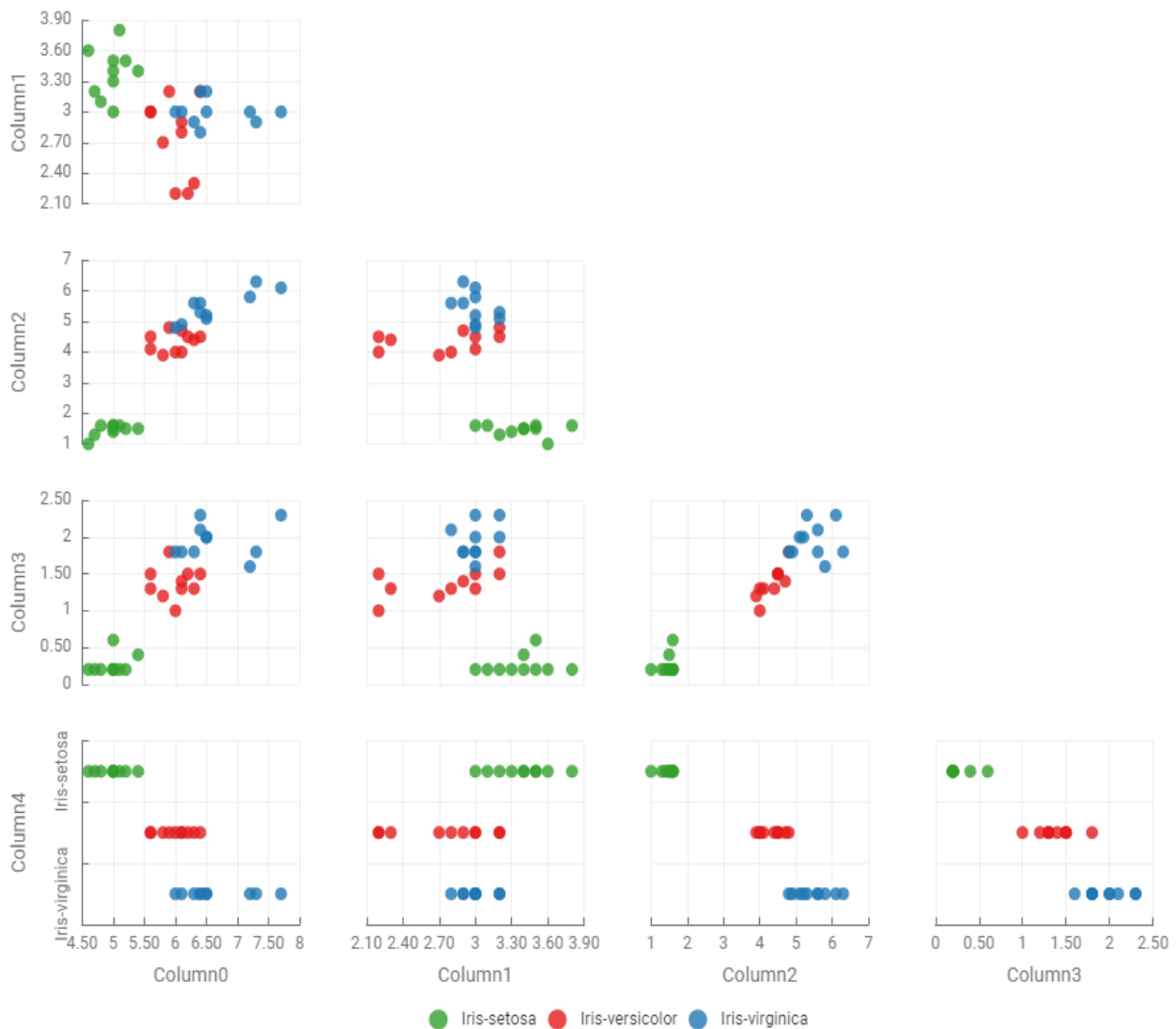
1.6 pav. Užduočių seka duomenų pavaizdavimui XY koordinatinių sistemoje

Scatter Plot Matrix



1.7 pav. Mokymo duomenų pavaizdavimas XY koordinatų sistemoje

Scatter Plot Matrix



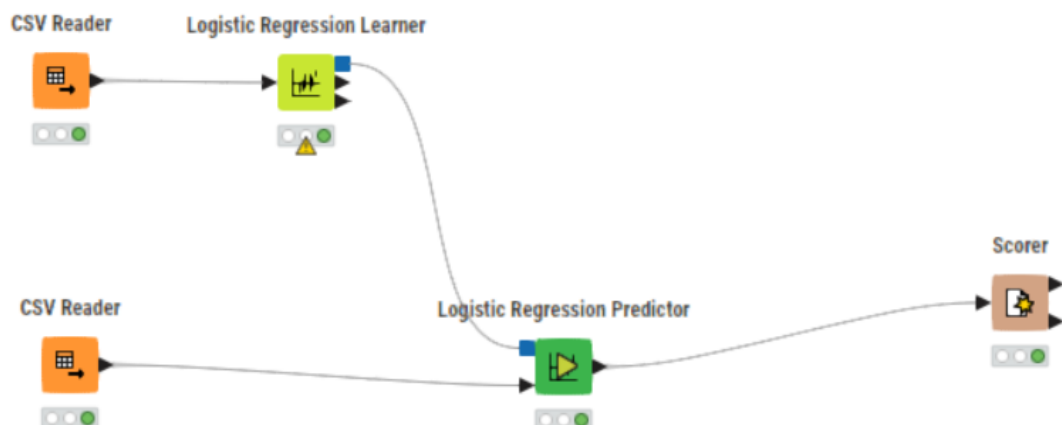
1.8 pav. Testavimo duomenų pavaizdavimas XY koordinatinių sistemoje

1.6. Daugiasluoksnio perceptrono palyginimas su kitais mašininio mokymosi metodais

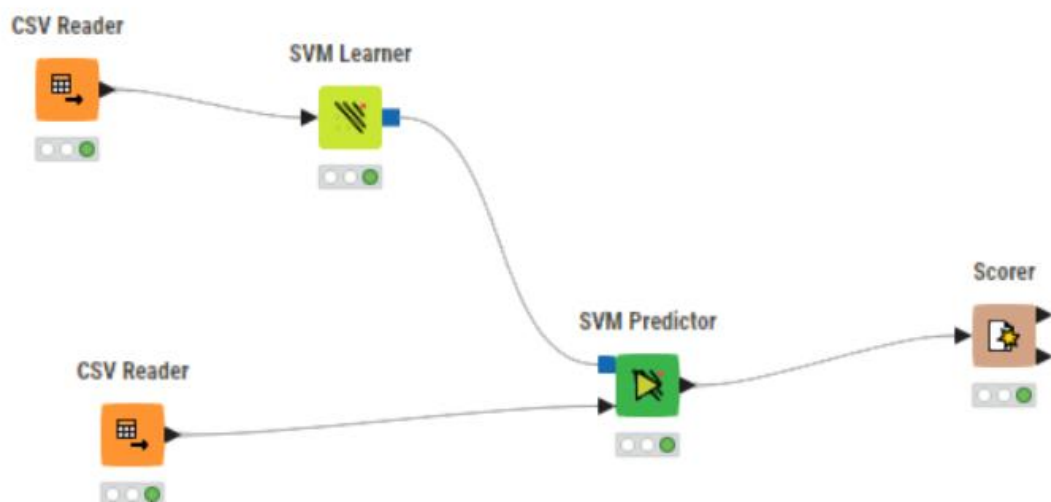
Daugiasluoksnio perceptrono palyginimas su logistinės regresijos ir SVM metodais. Palyginimui kiekvieno metodo hiperparametrai:

- Daugiasluoksnis perceptronas,
 - 100 iteracijų;
 - 1 paslėptasis sluoksnis;
 - 10 paslėptieji neuronai per sluoksnį,
- Logistinė regresija
 - 100 epochų
 - 0,1 žingsnio dydis

- SVM
 - Daugianario (polynomial)
 - 1,0 poslinkis
 - 1,0 galia
 - 1,0 gama



1.9 pav. Užduočių sekos logistinės regresijai apmokyti ir testuoti



1.10 pav. Užduočių sekos SVM metodui apmokyti ir testuoti

1.7 lentelė. Klasifikavimo tikslumo reikšmės su skirtingais hiperparametrų rinkiniais

Metrika / Metodas	Daugiasluoksniu perceptronu	Logistine regresija	SVM
Tikslumas (Accuracy)	0,967	0,967	0,933
Precision	0,97	0,97	0,933
Sensitivity	0,967	0,967	0,933
Recall	0,967	0,967	0,933
F-measure	0,967	0,966	0,933

Naudojant parinktus hiperparametrus, daugiasluoksniu perceptrono ir logistinės regresijos metodai pasiekė tokį patį tikslumo rezultatą, o SVM metodas buvo 0,034 mažiau tikslus, palyginti su kitais metodais.

2. IŠVADOS

- Keičiant daugiasluoksnio perceptrono hiperparametrus, tikslumas nesikeitė arba net pablogėjo, kai hiperparametrų reikšmės buvo didelės. Padidinus iteracijų skaičių iki 300 ir paslėptųjų neuronų skaičių iki 100, modelio tikslumas sumažėjo nuo 0,967 iki 0,933. Tai gali todėl, kad per didelis tinklo sudėtingumas lemia persimokymą (overfitting) – modelis pernelyg gerai prisitaiko prie mokymosi duomenų, tačiau blogiau apibendrina naujus, nematytus duomenis.
- Lyginant su kitais metodais, daugiasluoksnio perceptrono tikslumo rezultatai nesiskyrė arba buvo šiek tiek geresni. Tokį pat tikslumą (0,967) pasiekė ir logistinės regresijos metodas, o SVM tikslumas buvo mažesnis – 0,933. Galima teigti, kad MLP yra optimalus pasirinkimas šiam uždaviniui spręsti.
- Beveik visais atvejais, naudojant skirtingus daugiasluoksnio perceptrono parametrų rinkinius ar kitus metodus, neteisingai klasifikuojamas tik vienas įrašas. Retais atvejais klasifikacija būna tobula – visi įrašai klasifikuojami teisingai. Tai gali būti, dėl uždaviniu paprastumo ir testavimo duomenų mažumo.

3. PAPILDOMA

- Panaudota KNIME programinė įranga.
- Panaudoto ChatGPT gramatinėm klaidom taisyti.