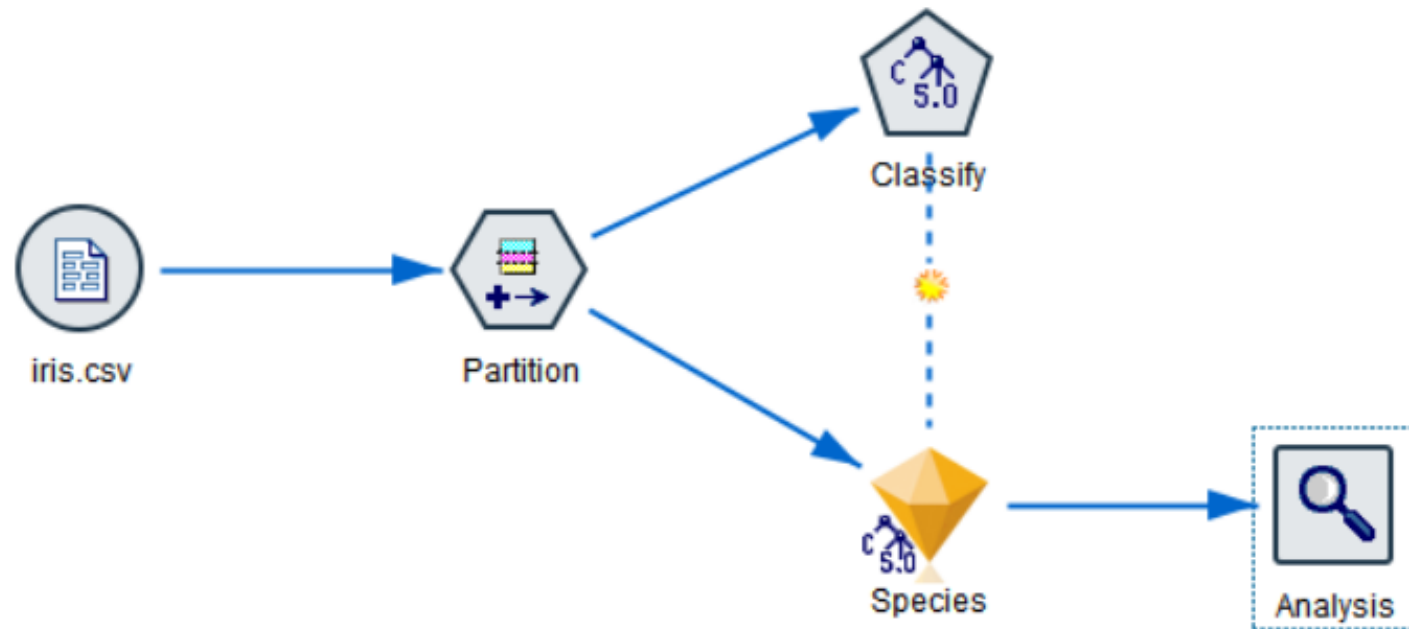


# Klasifikacija

# Primer



iris.csv

Preview

Refresh

C:\Users\Pavle\Documents\Faks\IP\Podaci\iris.csv

File

Data

Filter

Types

Annotations

Read Values

Clear Values

Clear All Values

Field	Measurement	Values	Missing	Check	Role
<div></div> Sepal_Length	<div></div> Continuous	[4.3,7.9]		None	<div></div> Input
<div></div> Sepal_Width	<div></div> Continuous	[2.0,4.4]		None	<div></div> Input
<div></div> Petal_Length	<div></div> Continuous	[1.0,6.9]		None	<div></div> Input
<div></div> Petal_Width	<div></div> Continuous	[0.1,2.5]		None	<div></div> Input
<div>A</div> Species	<div></div> Nominal	setosa,ver...		None	<div></div> Target

View current fields

View unused field settings

OK

Cancel







Apply

Reset

Za početak, u izvornom čvoru treba da se podsi atribut nad kojim hoćemo da vršimo klasifikaciju.

Rola **target** označava polje koje će se koristiti kao izlaz algoritma.

Partition

  **Generate**  **Preview**   

**Settings** Annotations

Partition field:

Partitions: ☒ Train and test ☐ Train, test and validation

Training partition size:  Label:  Value =

Testing partition size:  Label:  Value =


Validation partition size:  Label:  Value =

Total size: 100%

Values: ☒ Use system-defined values ("1", "2" and "3")  
☐ Append labels to system-defined values  
☐ Use labels as values

☒ Repeatable partition assignment

Seed:  **Generate**

☐ Use unique field to assign partitions:  

**OK** **Cancel** **Apply** **Reset**

U čvoru za particionisanje definišemo kako želimo da se originalni skup podataka particioniše na **trening** i **test** skupove.

Biramo procenat slogova iz originalnog skupa koji će biti korišćeni za treniranje, procenat koji će biti korišćen za testiranje i obeležja koja će se koristiti za označavanje podele.

Možemo da podesimo *seed* na određenu vrednost kako bi svaki put particionisanje bilo izvršeno na isti način.

Kao rezultat se dobija početni skup podataka u kome slogovi imaju dodatni atribut koji označava da li se određeni slog nalazi u trening skupu ili u test skupu. Ovaj atribut ima rolu **partition**.

Classify

Fields Model Costs Analyze Annotations

☒ Use predefined roles  
☐ Use custom field assignments

Target: [Field Selection]

Inputs: [Field Selection]

Partition: [Field Selection]

Splits: [Field Selection]

☐ Use weight field [Field Selection]

OK Run Cancel Apply Reset

U ovom primeru se koristi C5 algoritam za klasifikaciju.

Za klasifikaciju možemo da koristimo role koje su već definisane nad ulaznim podacima čvora, ili možemo da odredimo neke druge role.

Classify

5.0

Fields **Model** Costs Analyze Annotations

Model name: ☒ Auto ☐ Custom

☒ Use partitioned data

☒ Build model for each split

Output type: ☒ Decision tree ☐ Rule set

☐ Group symbolics

☐ Use boosting Number of trials:

☐ Cross-validate Number of folds:

Mode: ☒ Simple ☐ Expert

Favor: ☒ Accuracy ☐ Generality

Expected noise (%):

OK Run Cancel Apply Reset

Potrebno je štiklirati *use partitioned data* kako bi model mogao da pročita atribut za particionisanje is ulaznog skupa podataka.

Classify

5.0

?

Fields

Model

Costs

Analyze

Annotations

☒ Use misclassification costs

Predicted

	setosa	versicolor	virginica
setosa	0.0	1.0	1.0
versicolor	1.0	0.0	1.0
virginica	1.0	1.0	0.0

Actual

OK

▶

Run

Cancel

Apply

Reset

Mogu se dodeliti cena za lošu klasifikaciju, kako bi se model trenirao u skladu sa cenama.

Povećanjem cene određene greške će smanjiti učestalost te greške ali kao posledicu ima povećanje učestalosti neke druge greške.

Analysis

Analyze \$C-Species

Analysis Output Annotations

☒ Coincidence matrices (for symbolic targets)

☐ Performance evaluation

☐ Evaluation metric (AUC & Gini, binary classifiers only)

☐ Confidence figures (if available)

Threshold for: 90 % correct

Improve accuracy: 2.0 fold

Find predicted/predictor fields using:

☒ Model output field metadata

☐ Field name format (for example, '\$<x>--<target field>')

☒ Separate by partition

☐ User defined analysis Define User Measure...

Break down analysis by fields:

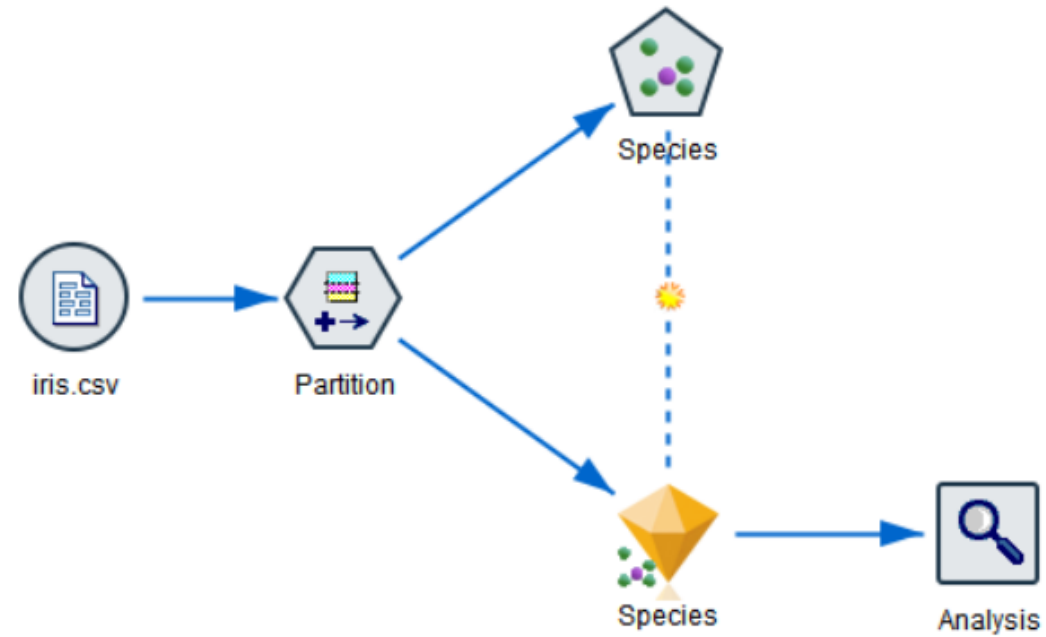
OK Run Cancel Apply Reset

Na čvor modela može da se poveže čvor za analizu pomoću koda možemo da dobijemo karakteristike dobijenog modela.


Može da se generiše matrica konfuzije, kao i neke druge metrike za evaluaciju binarnih klasifikatora.



# KNN primer



Species



?

Objectives

Fields

Settings

Annotations

The kNN procedure will identify the most similar training cases (the nearest neighbors) to your cases of interest. A target field can be predicted based on the neighboring values.

What type of analysis do you want to perform?

☒ Predict a target field

☐ Only identify the nearest neighbors

What is your objective?


☐ Balance speed and accuracy  
Automatically selects the best number of neighbors within a small range.

☐ Speed  
Finds a fixed number of neighbors.

☐ Accuracy  
Automatically selects the best number of neighbors within a larger range and uses predictor importance when calculating distances.

☒ Custom analysis  
Choose this option to fine tune the algorithm on the Settings tab.

OK

 Run

Cancel

Apply

Reset

Biramo da li želimo da model vrši predikciju ili samo da određuje  $k$  najbližih suseda.

Takođe možemo da izaberemo neka predefinisana podešavanja u zavisnosti od naših potreba.

Species

?

?

?

Objectives **Fields** Settings Annotations

☒ Use predefined roles

☐ Use custom field assignments

Target:

Inputs:

OK Run Cancel Apply Reset

Kao i kod stabla odlučivanja, možemo da izaberemo role atributa u slučaju da one već nisu definisane.

Species

Objectives Fields **Settings** Annotations

Model  
Neighbors  
Feature Selection  
Cross-Validation  
Analyze

Number of Nearest Neighbors (k)

☐ Specify fixed K

K:

☒ Automatically select k

Minimum:

Maximum:

Distance Computation

☒ Euclidean metric

☐ City Block metric

☒ Weight features by importance when computing distances

Predictions for Range Target

☒ Mean of nearest neighbor values

☐ Median of nearest neighbor values

OK Run Cancel Apply Reset

Biramo broj suseda koji se posmatra. Možemo da podesimo raspon, a da se optimalni broj suseda automatski odredi.

Biramo koju metriku koristimo i da li želimo da računamo težinsko rastojanje.

Species

Objectives Fields **Settings** Annotations

Model  
Neighbors  
Feature Selection  
**Cross-Validation**  
Analyze

**Cross-Validation Folds**  
V-fold cross-validation is performed if you choose automatic k selection but do not choose feature selection.

☒ Randomly assign cases to folds

Number of folds:

☐ Set random seed Seed:

☐ Use field to assign cases

OK Run Cancel Apply Reset

Biramo parametre za unakrsnu validaciju.