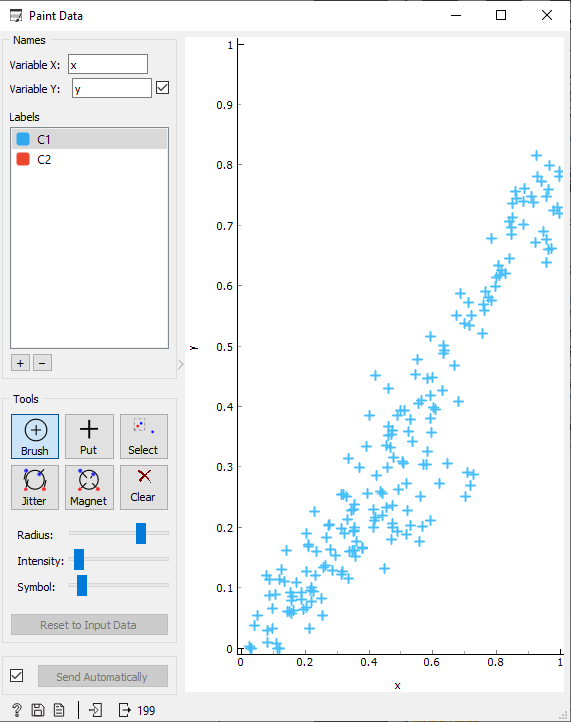
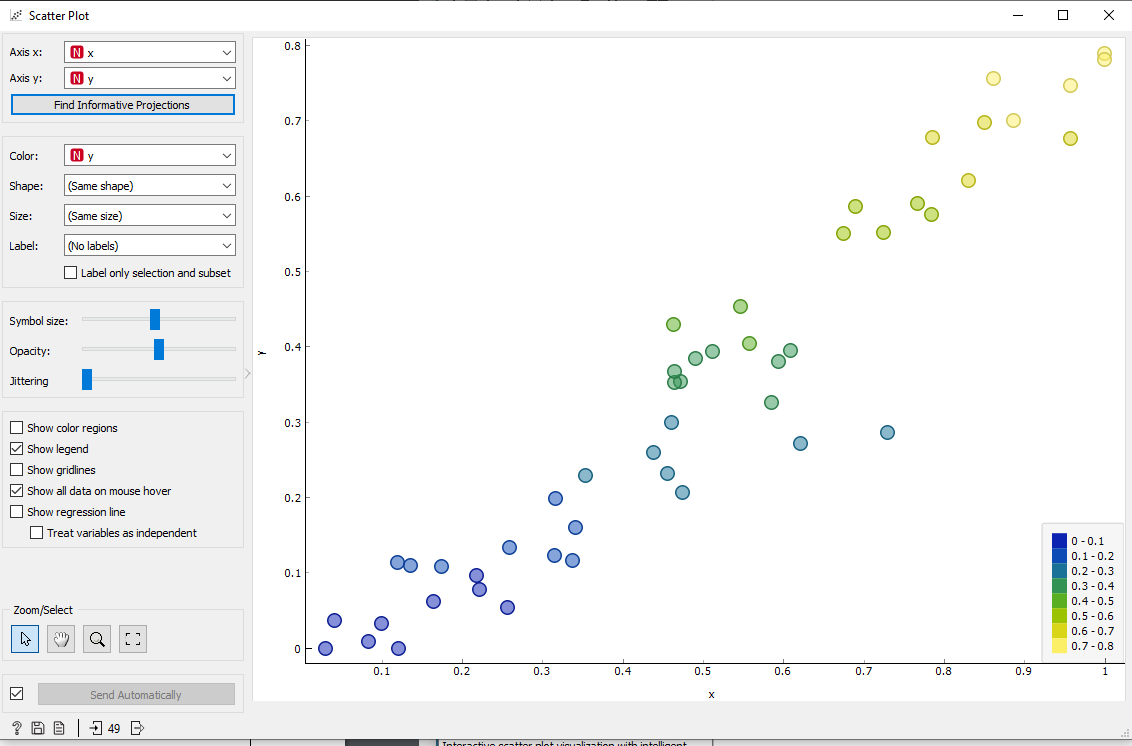
**RUAP LV5**

Zadatak 1.

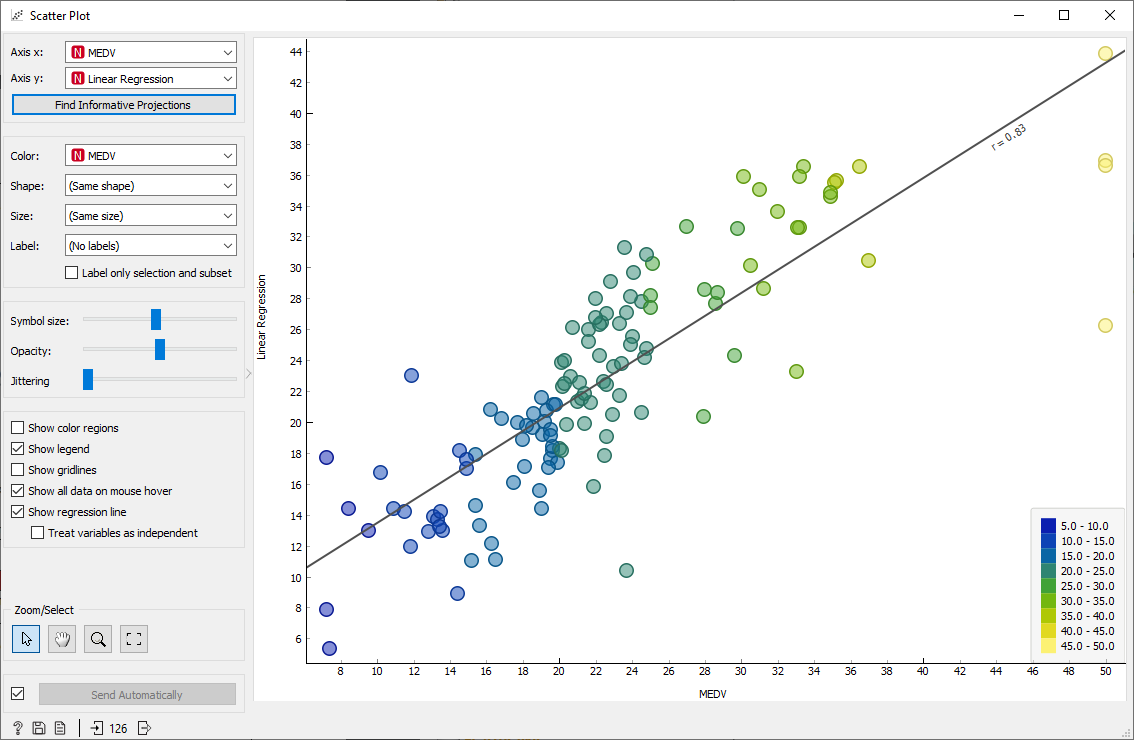
Paint Data



Scatter Plot

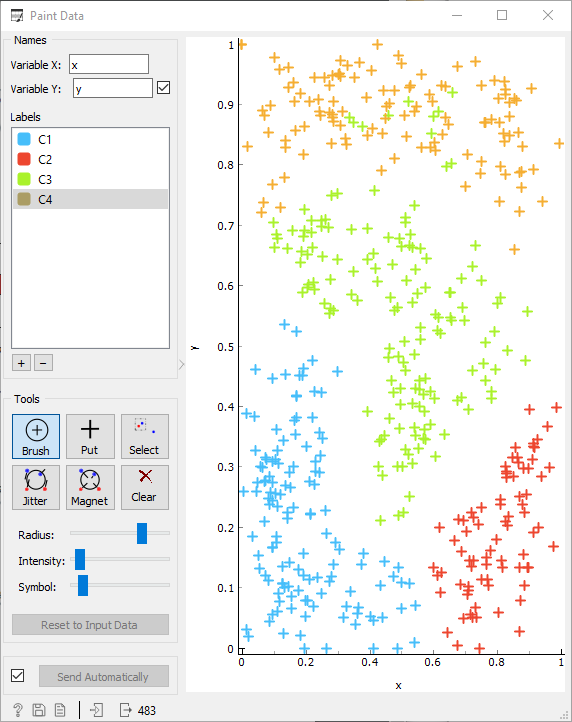


Na slici iznad može se primjetiti da su slike slične, što govori da je su se predviđeni podaci slični onima na orignalu (Paint data).  
Zadatak 2.

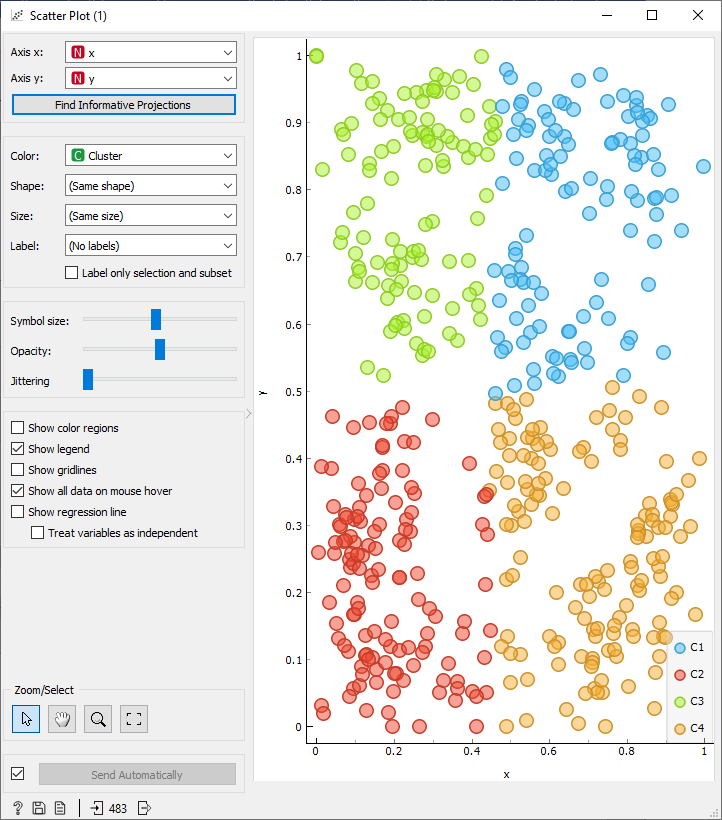
   
S povećanjem podataka povećava se težina predviđanja u ovom primjeru.

Zadatak 3.

Paint Brush



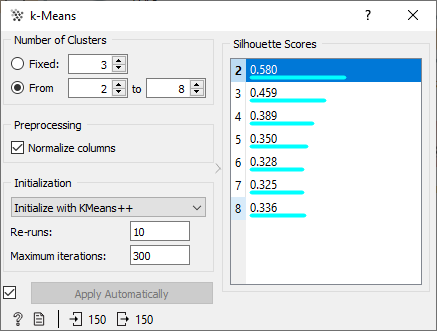
Scatter Plot



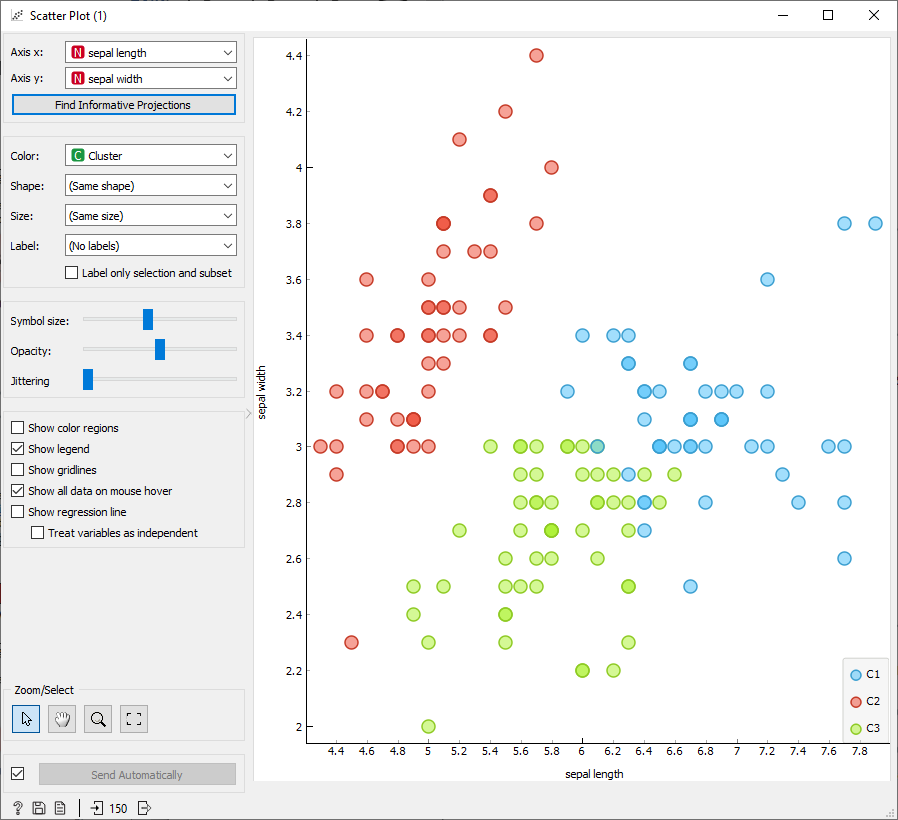
Može se vidjeti da su clusteri relativno dobro podjeljeni. Ako bi prilikom nanošenja podataka u Paint Data ostavili veći razmak između dva clustera, bolje bi bili podijeljeni.

Zadatak 4.

1. Iris

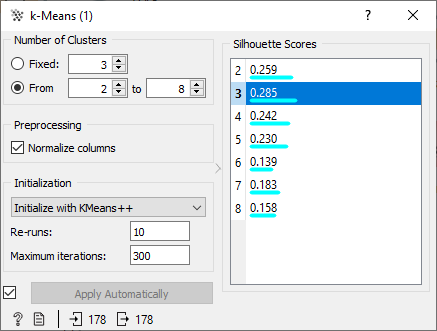


Scatter plot

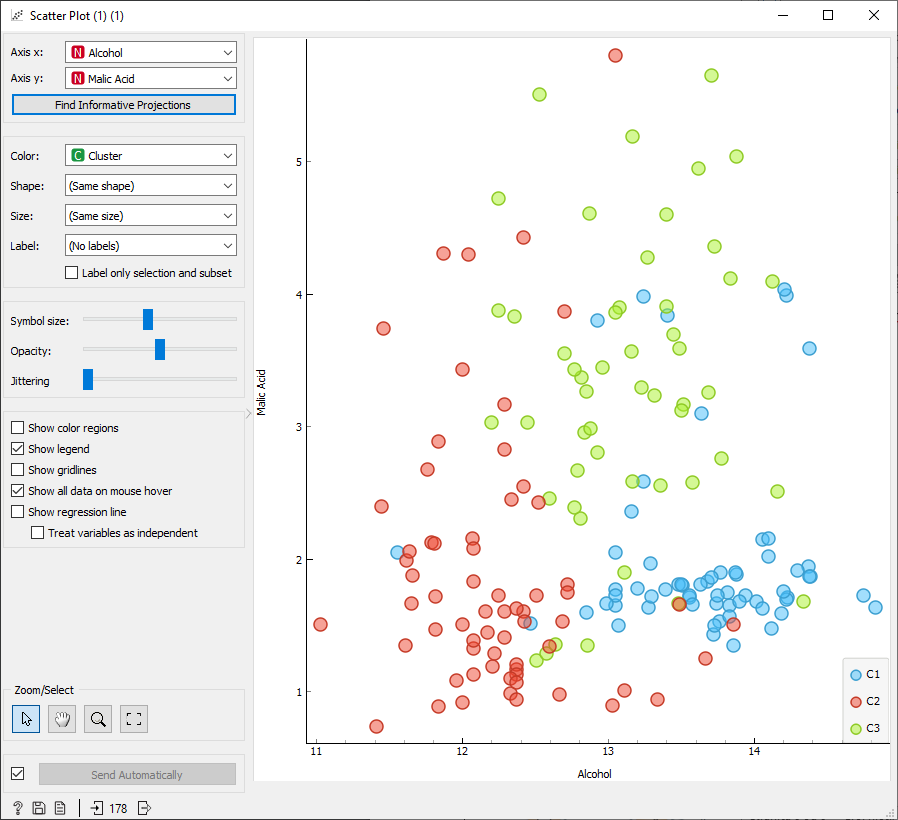


K-means nam pokazuje da je najbolja kvaliteta clustera za unesene podatke razdvajanja u dva clustera.

b)vino



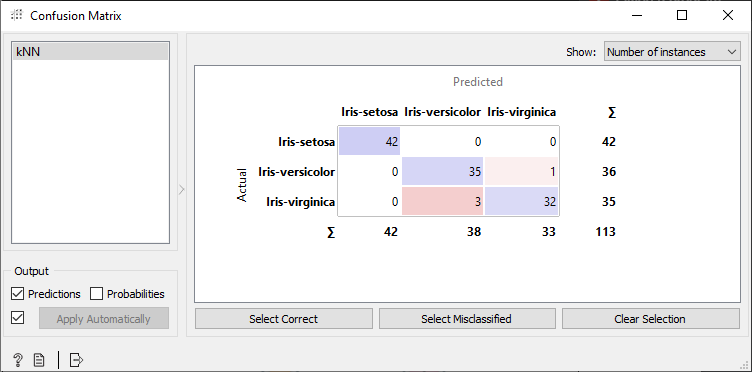
Scatter plot

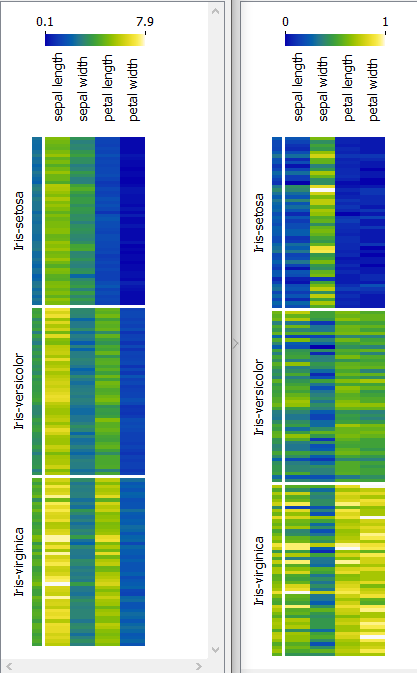


Problem se događa kada ima previše preklapanja, tada može doći do progreše u k-means. Kao što se dogodilo na primjeru Iris gdje nam se preklapaju točke kada imamo tri clustera.

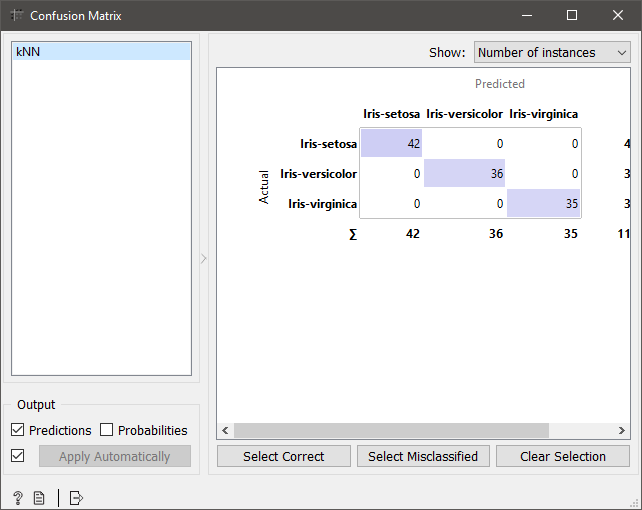
Zadatak 5.

S normalizacijom u intervalu od [0,1], k=5, weight: uniform

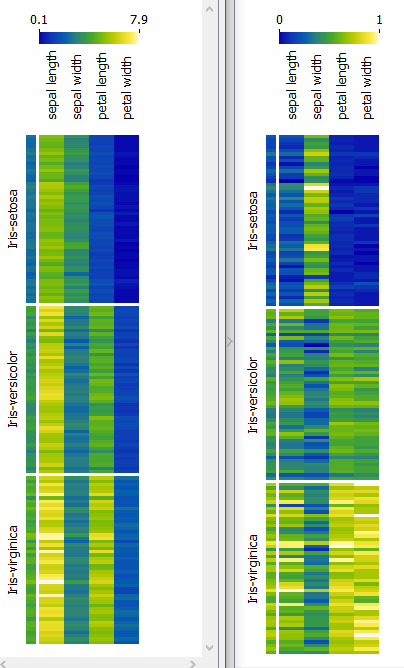




S normalizacijom u intervalu od [0,1], k=15, weight: distance

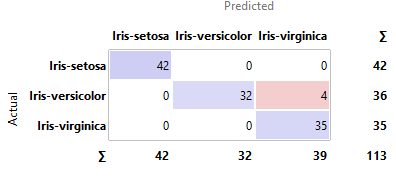


S povećanjem susjeda možemo vidjeti da se rezulat poboljšao, tako što se svi podaci nalazen na glavnoj dijagonali.

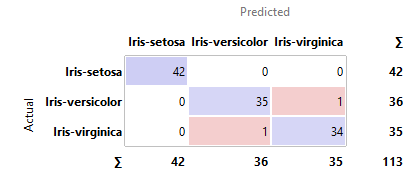


Zadatak 6.

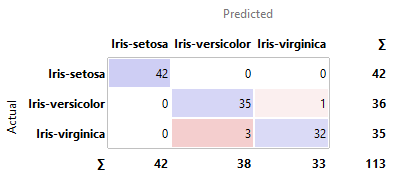
Dana: Iris  
Random Forest



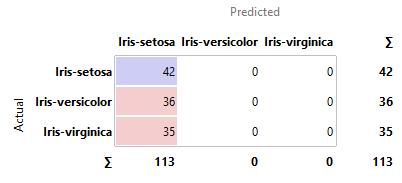
SVM



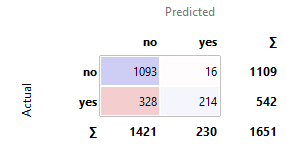
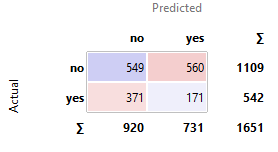
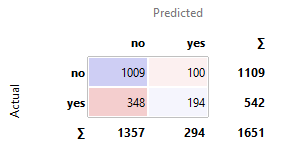
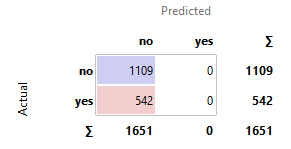
kNN



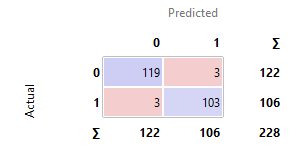
Constant(Majority)

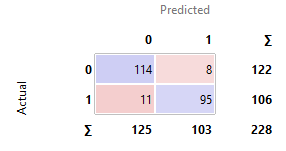


Za ovaj skup podataka najbolje nam odgovara SVM jer ima nema elemenata nigdje osim na glavnoj dijagonali. A najlošiji je majority.

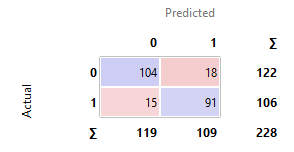
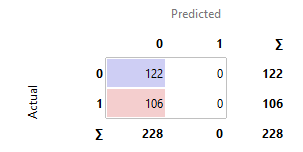
Data: Titanic  
Random Forest:  
  
SVM  
  
kNN  
  
Constant  


U ovom slucaju nam je također Random Forest najbolji odabir a SVM najgori.

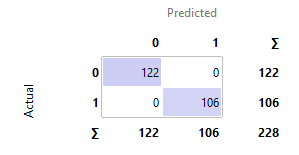
Data: Heart\_disease  
Random Forest  
  
SVM



kNN

  
Constant  


U ovom slučaju najbolji je Random Forest a najgori je Constant(Majority)

U svakom od ispitivanja postoji ovisnost o podešenosti parametara svakog model. Npr Ukoliko je kNN Weight promjenjen na distance Confusion matrix će izgledati ovako:  
kNN – Weight: distance

U odnosu na ova tri testa Random forest se predstavlja kao najbolji model a constant(majority) kao najlošiji.