Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko
Koroška cesta 46
2000 Maribor, Slovenija

FERI

Strojno učenje: algoritem K najbližjih sosedov

Projektna naloga pri predmetu Modeli in odločitveni sistemi

Avtor: Urban Vižintin

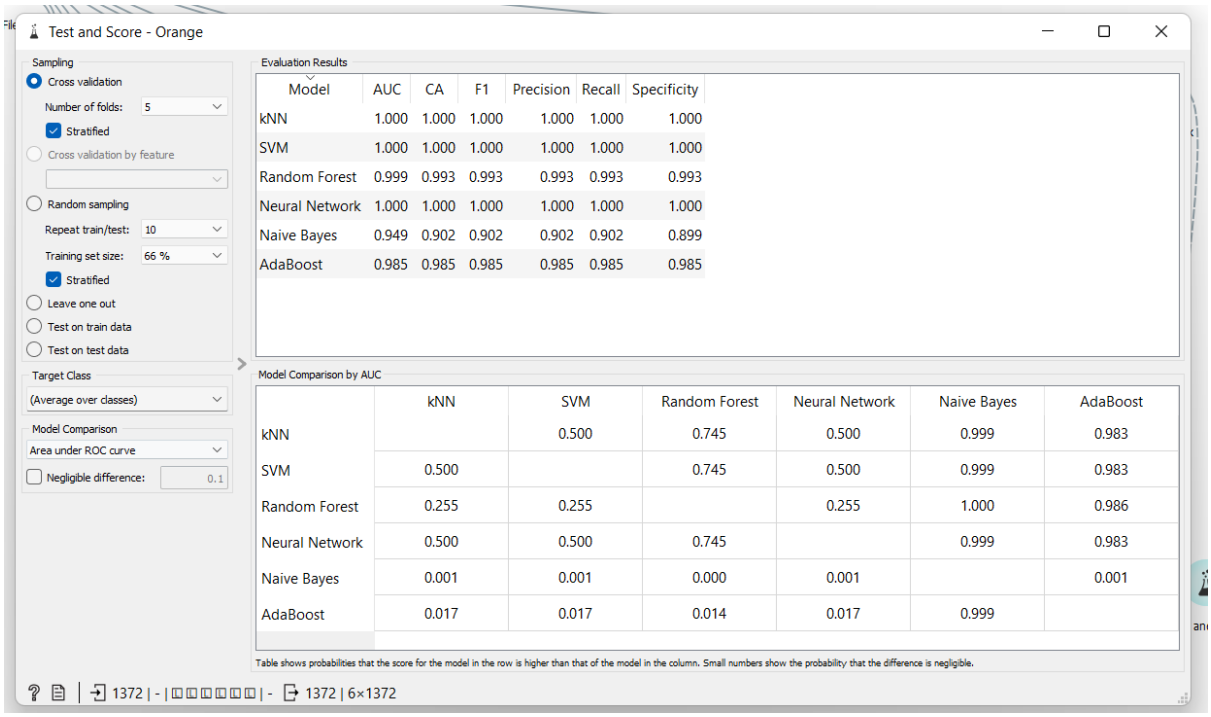
Smer študija: ITK (UN)

Študijsko leto: 2021/22

Opis projektne naloge

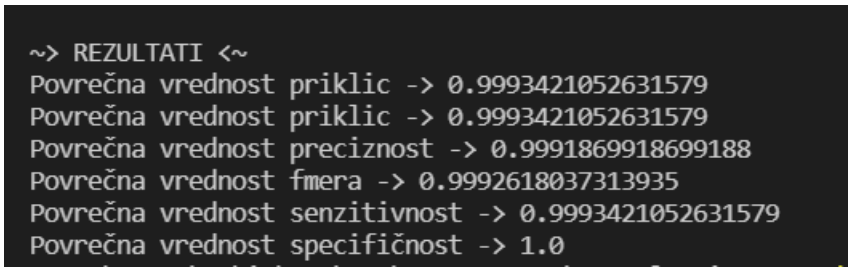
Pri sledeči projektni nalogi sem se iz ponujenih algoritmov odločil za algoritem KNN (K najbližjih sosedov), saj je meni osebno najbolj razumljiv. Za programski jezik sem si izbral Python, s pomočjo katerega bom razdelil podatkovno zbirko bankovcev na učno in testno množico (k-fold : k = 5). Pri računanju matrik ter ostalih atributov teh foldov si bom pomagal z različnimi Pythonovimi knjižnicami (Numpy ...). Celoten implementiran primer bo nato izrisan v programskem orodju Orange3, v katerega bom tudi vnesel vhodne podatke in simuliral celkotno strojno učenje. Pridobljene rezultate bom primerjal s svojim programom in izpostavil razne razlike oziroma podobnosti.

Rezultati v Orange3



Kot se vidi, je algoritem bil skorajda popolnoma točen pri napovedovanju novih primerkov.

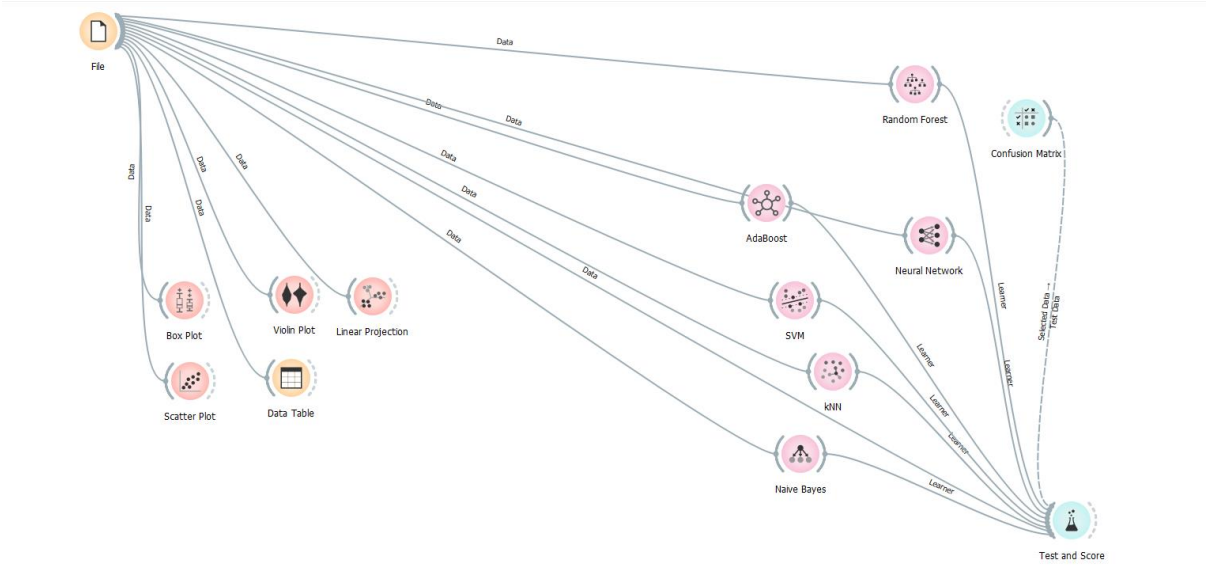
Rezultati mojega programa



Primerjava Orange3 ter mojega programa

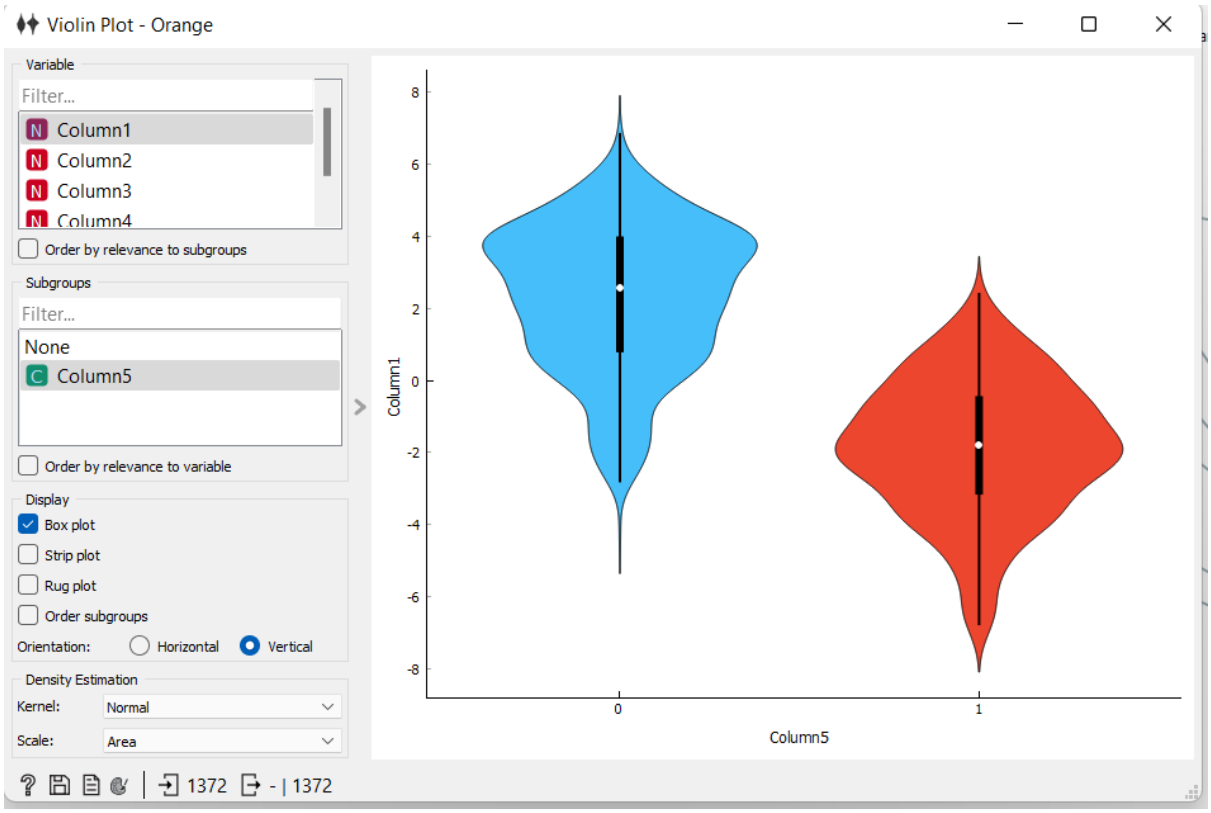
Obadva algoritma sta izračunala skoraj stoprocentno natančnost pri napovedovanju vrednosti, vendar je bil moj model kanček manj natančen.

Orange3 shema

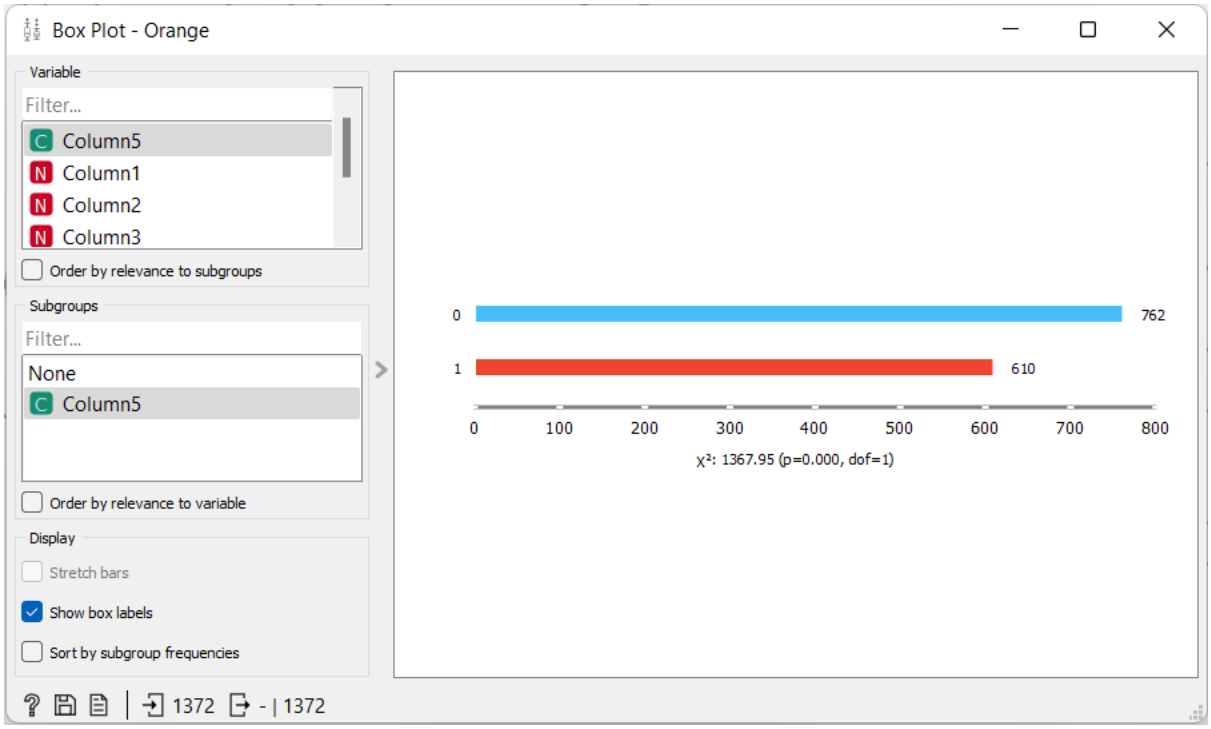


Grafi

-> Violin plot:



-> Boxplot:


Box Plot - Orange
Variable
Filter...
Column5
Column1
Column2
Column3
Order by relevance to subgroups
Subgroups
Filter...
None
Column5
Order by relevance to variable
Display
Stretch bars
Show box labels
Sort by subgroup frequencies
1372
- | 1372
0
762
1
610
0
100
200
300
400
500
600
700
800
χ²: 1367.95 (p=0.000, dof=1)