



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko

Koroška cesta 46
2000 Maribor, Slovenija



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 celotno strojno učenje. Pridobljene rezultate bom primerjal s svojim programom in izpostavil razne razlike oziroma podobnosti.

Rezultati v Orange3

The screenshot displays the 'Test and Score' widget in Orange3. The left sidebar contains settings for sampling and model comparison. The main area shows two tables: 'Evaluation Results' and 'Model Comparison by AUC'.

Sampling Settings:

- ☒ Cross validation
- Number of folds: 5
- ☒ Stratified
- ☐ Cross validation by feature
- ☐ Random sampling
- Repeat train/test: 10
- Training set size: 66 %
- ☒ Stratified
- ☐ Leave one out
- ☐ Test on train data
- ☐ Test on test data

Target Class: (Average over classes)

Model Comparison: Area under ROC curve

☐ Negligible difference: 0.1

Evaluation Results Table:

Model	AUC	CA	F1	Precision	Recall	Specificity
kNN	1.000	1.000	1.000	1.000	1.000	1.000
SVM	1.000	1.000	1.000	1.000	1.000	1.000
Random Forest	0.999	0.993	0.993	0.993	0.993	0.993
Neural Network	1.000	1.000	1.000	1.000	1.000	1.000
Naive Bayes	0.949	0.902	0.902	0.902	0.902	0.899
AdaBoost	0.985	0.985	0.985	0.985	0.985	0.985

Model Comparison by AUC Table:

	kNN	SVM	Random Forest	Neural Network	Naive Bayes	AdaBoost
kNN		0.500	0.745	0.500	0.999	0.983
SVM	0.500		0.745	0.500	0.999	0.983
Random Forest	0.255	0.255		0.255	1.000	0.986
Neural Network	0.500	0.500	0.745		0.999	0.983
Naive Bayes	0.001	0.001	0.000	0.001		0.001
AdaBoost	0.017	0.017	0.014	0.017	0.999	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

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

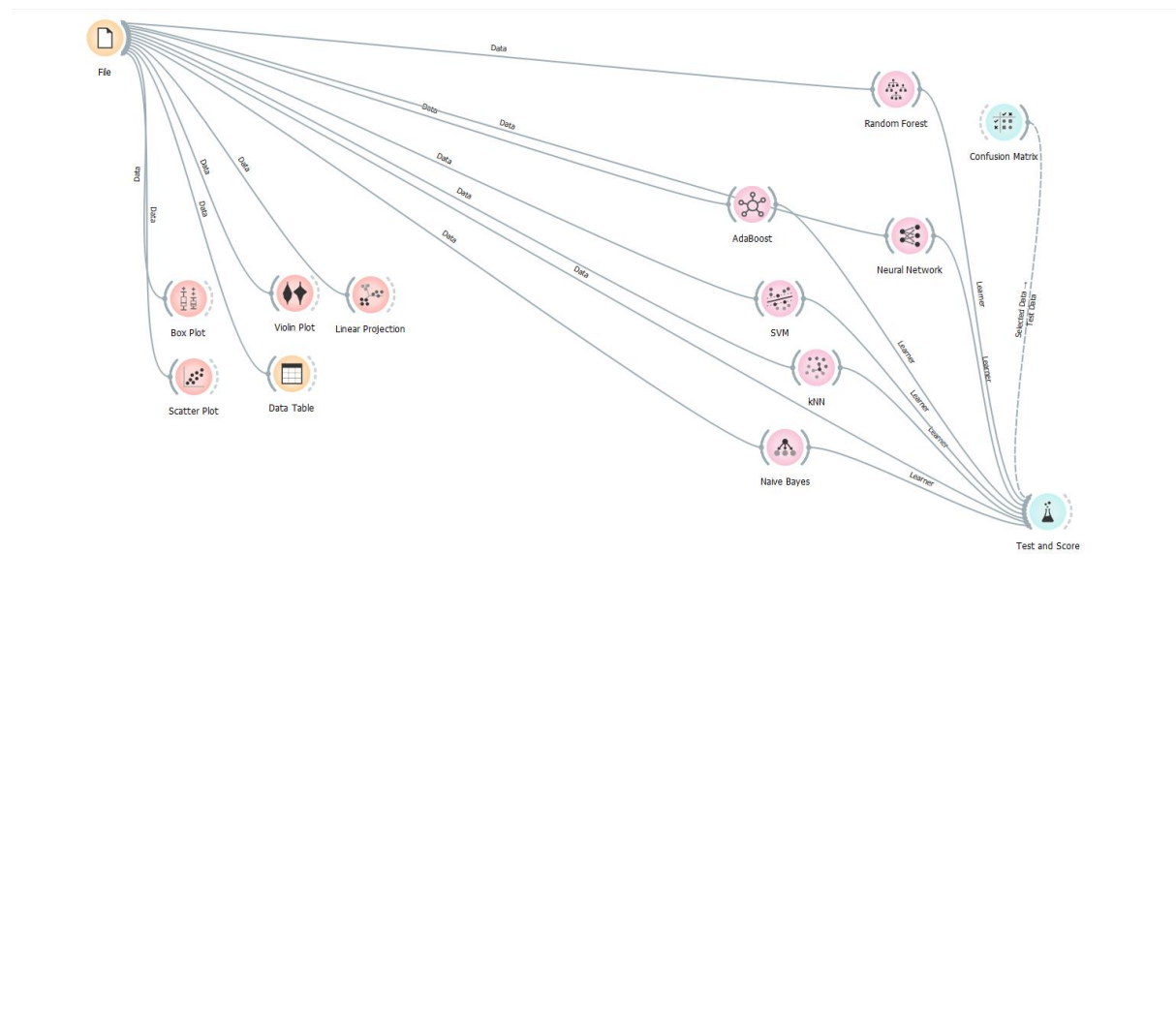
Rezultati mojega programa

```
~> REZULTATI <~  
Povrečna vrednost priklic -> 0.9993421052631579  
Povrečna vrednost priklic -> 0.9993421052631579  
Povrečna vrednost preciznost -> 0.9991869918699188  
Povrečna vrednost fmera -> 0.9992618037313935  
Povrečna vrednost senзитivnost -> 0.9993421052631579  
Povrečna vrednost specifičnost -> 1.0
```

Primerjava Orange3 ter mojega programa

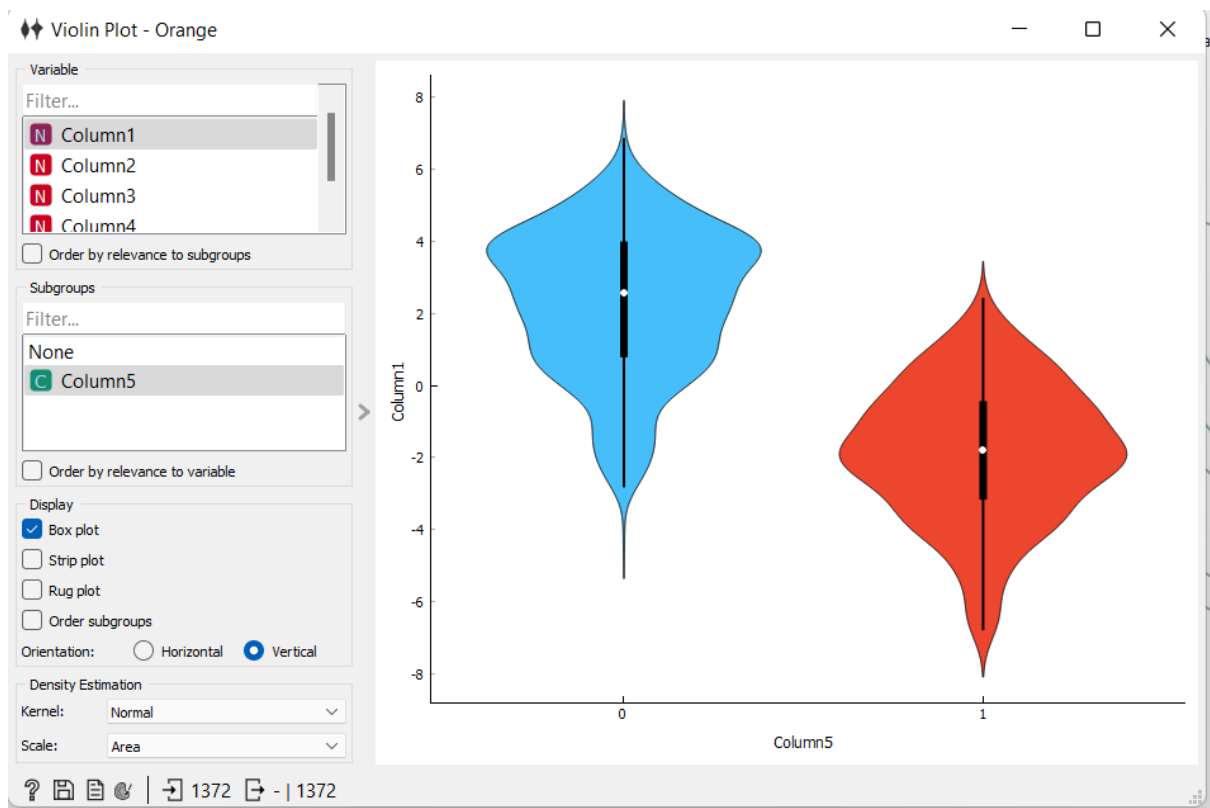
Obadva algoritma sta izračunala skoraj stoprocentno natančnost pri napovedovanju vrednosti, vendar je bil moj model kanček manj natančen.

Orange3 shema



Grafi

-> Violin plot:



-> Boxplot:

