

Fakulteta za elektrotehniko, računalništvo in informatiko
Univerza v Mariboru

Analiza vrednosti nepremičnin

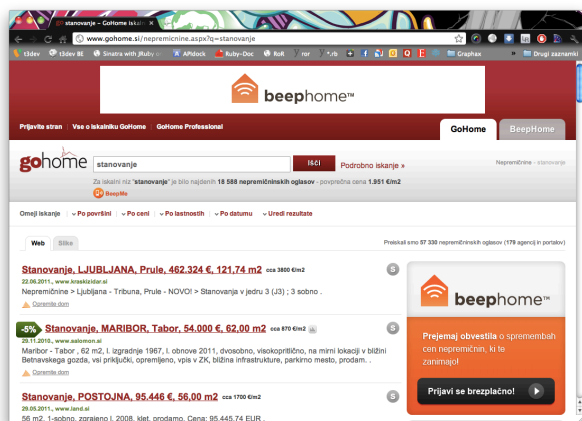
Poročilo 2. naloge pri predmetu sistemi znanja

Predstavitev primera	3
Analiza strojnega učenja z uporabo M5P	4
Analiza strojnega učenja z uporabo J48	6
Primerjava rezultatov strojnega učenja	8

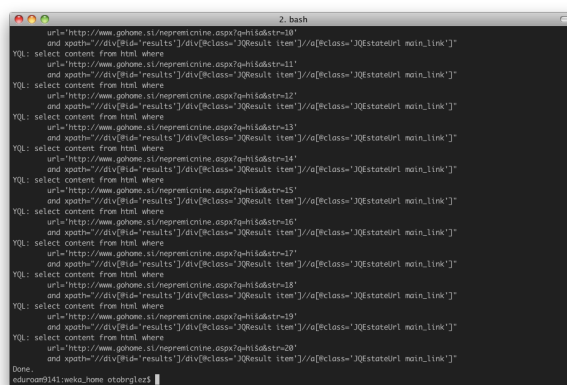
Predstavitev primera

Z uporabo metode "scrape" sem iz spletnega portala GoHome.si pridobil podatke o nepremičninah v večjih slovenskih mestih. Pridobil sem podatke: lokacija, velikost, cena in tip nepremičnine (hiša ali stanovanje). Podatke sem klasificiral glede na lokacijo.

Namen naloge je bil izdelati model, ki lahko predvidi ceno nepremičnine glede na lokacijo, velikost in tip.



Nepremičninski iskalnik GoHome.si



Scrape z uporabo YQL v Ruby

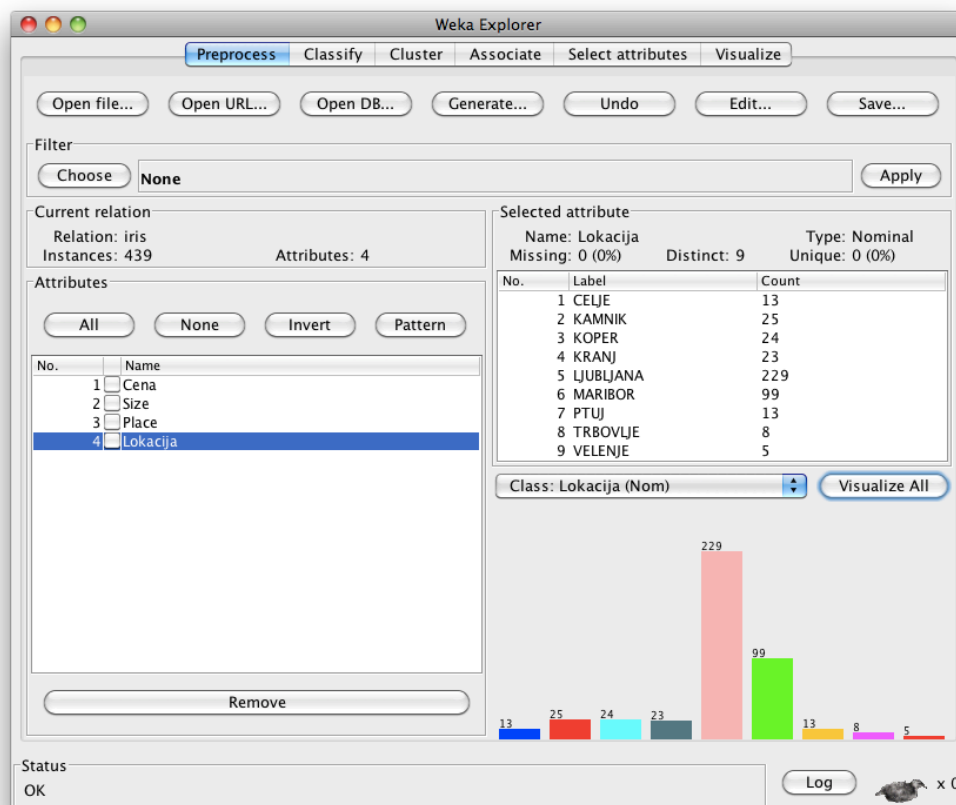
```

1 @RELATION iris
2 @ATTRIBUTE Cena REAL
3 @ATTRIBUTE Size REAL
4 @ATTRIBUTE Place {Stanovanje,Hiša}
5 @ATTRIBUTE Lokacija {CELJE,KAMNIK,KOPER,KRANJ,LJUBLJANA,MARIBOR,PTUJ,TRBOVLJE,VELENJE}
6 @DATA
7 462324,121.74,Stanovanje,LJUBLJANA
8 108000,81.42,Stanovanje,MARIBOR
9 135000,80.00,Stanovanje,KAMNIK
10 110000,40.00,Stanovanje,KOPER
11 100000,24.00,Stanovanje,LJUBLJANA
12 109990,66.44,Stanovanje,MARIBOR
13 115000,109.00,Stanovanje,CELJE
14 300000,149.00,Stanovanje,LJUBLJANA
15 95000,71.00,Stanovanje,MARIBOR
16 165000,75.00,Stanovanje,LJUBLJANA
17 125000,40.90,Stanovanje,LJUBLJANA
18 139900,69.99,Stanovanje,LJUBLJANA
19 61999,42.00,Stanovanje,MARIBOR
20 134000,55.00,Stanovanje,LJUBLJANA
21 199665,145.00,Stanovanje,MARIBOR
22 550000,165.00,Stanovanje,LJUBLJANA
23 69900,23.65,Stanovanje,LJUBLJANA
24 219000,115.00,Stanovanje,MARIBOR
25 220000,76.00,Stanovanje,LJUBLJANA
26 160000,72.00,Stanovanje,LJUBLJANA
27 155000,120.00,Stanovanje,LJUBLJANA
28 119000,35.21,Stanovanje,LJUBLJANA
29 83000,34.00,Stanovanje,KAMNIK
30 58000,44.00,Stanovanje,MARIBOR
31 320000,80.00,Stanovanje,LJUBLJANA
32 108000,80.45,Stanovanje,MARIBOR
33 320000,75.00,Stanovanje,LJUBLJANA
34 80000,66.00,Stanovanje,MARIBOR
35 110000,84.35,Stanovanje,MARIBOR
36 311905,67.59,Stanovanje,LJUBLJANA

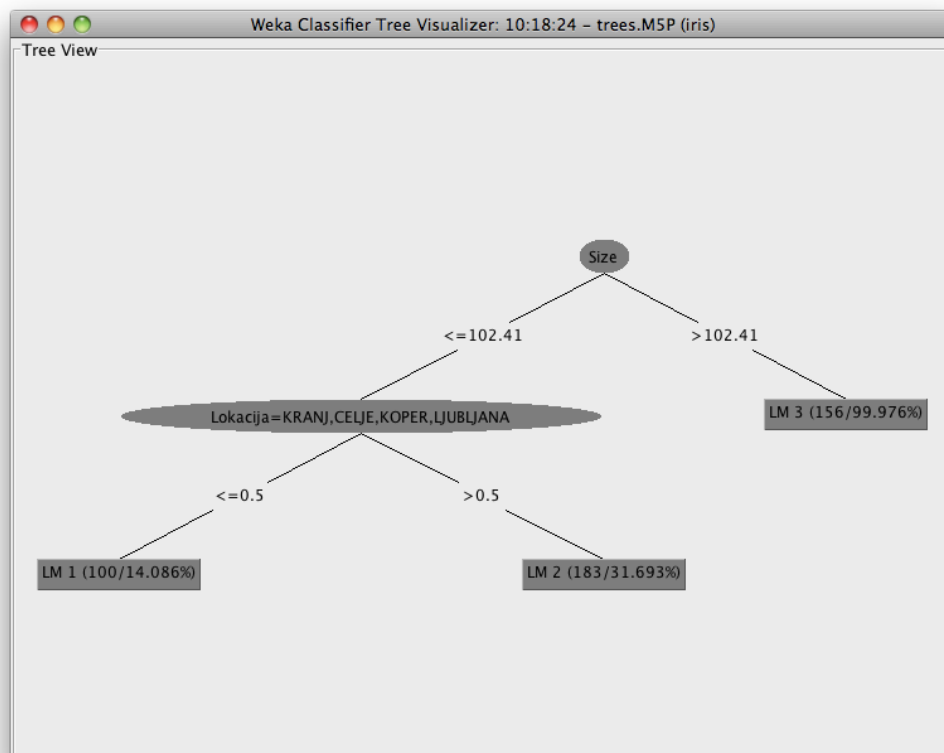
```

Izsek ARFF datoteke pripravljene za ogrodje WEKA

Analiza strojnega učenja z uporabo M5P



Pripravljeno WEKA ogrodje za analizo za 9 lokacij.



Drevo pridobljeno z M5P algoritmom

=== Run information ===

Scheme: weka.classifiers.trees.M5P -M 4.0
Relation: iris
Instances: 439
Attributes: 4

Cena
Size
Place
Lokacija

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

M5 pruned model tree:
(using smoothed linear models)

Size <= 102.41 :
| Lokacija=KRANJ,CELJE,KOPER,LJUBLJANA <= 0.5 : LM1 (100/14.086%)
| Lokacija=KRANJ,CELJE,KOPER,LJUBLJANA > 0.5 : LM2 (183/31.693%)
Size > 102.41 : LM3 (156/99.976%)

LM num: 1

Cena =
1265.0479 * Size
+ 2626.353 * Place=Hi°a
+ 24786.0285 * Lokacija=MARIBOR,KAMNIK,KRANJ,CELJE,KOPER,LJUBLJANA
+ 29575.7229 * Lokacija=KAMNIK,KRANJ,CELJE,KOPER,LJUBLJANA
+ 2086.9022 * Lokacija=KRANJ,CELJE,KOPER,LJUBLJANA
- 2814.6741 * Lokacija=CELJE,KOPER,LJUBLJANA
+ 6636.0886 * Lokacija=KOPER,LJUBLJANA
+ 6381.347 * Lokacija=LJUBLJANA
- 20622.0071

LM num: 2

Cena =
2364.006 * Size
+ 2626.353 * Place=Hi°a
+ 5391.7792 * Lokacija=MARIBOR,KAMNIK,KRANJ,CELJE,KOPER,LJUBLJANA
+ 2806.3272 * Lokacija=KAMNIK,KRANJ,CELJE,KOPER,LJUBLJANA
+ 2086.9022 * Lokacija=KRANJ,CELJE,KOPER,LJUBLJANA
- 32005.0376 * Lokacija=CELJE,KOPER,LJUBLJANA
+ 63684.7448 * Lokacija=KOPER,LJUBLJANA
+ 23430.9346 * Lokacija=LJUBLJANA
- 36647.7626

LM num: 3

Cena =
326.253 * Size
+ 4576.9192 * Place=Hi°a
+ 107781.712 * Lokacija=MARIBOR,KAMNIK,KRANJ,CELJE,KOPER,LJUBLJANA
+ 3636.8237 * Lokacija=KRANJ,CELJE,KOPER,LJUBLJANA
+ 130730.6935 * Lokacija=LJUBLJANA
+ 79909.204

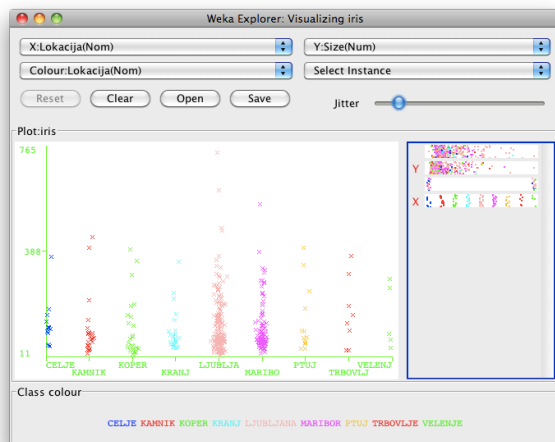
Number of Rules : 3

Time taken to build model: 0.17seconds

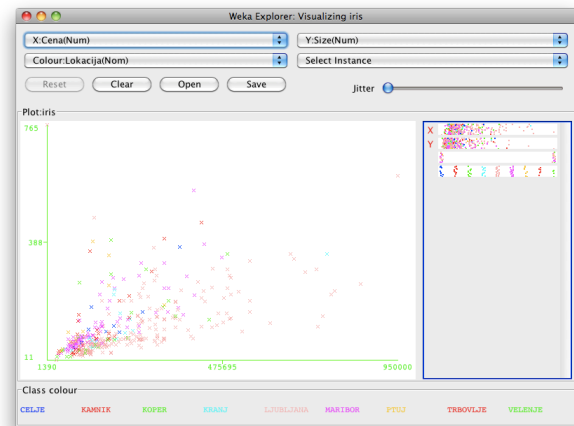
=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.7293
Mean absolute error	60038.1024
Root mean squared error	101408.3821
Relative absolute error	53.9342 %
Root relative squared error	68.3333 %
Total Number of Instances	439

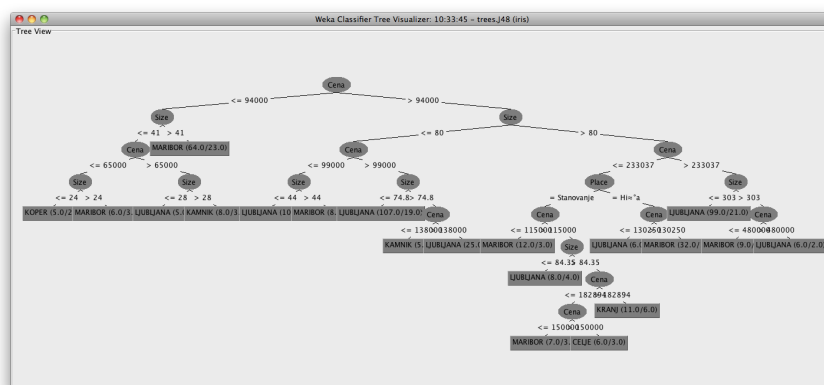


X: Lokacija / Y: Velikost

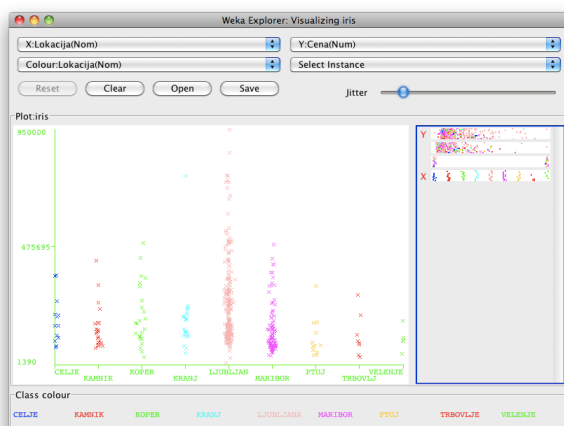


X: Cena / Y: Velikost

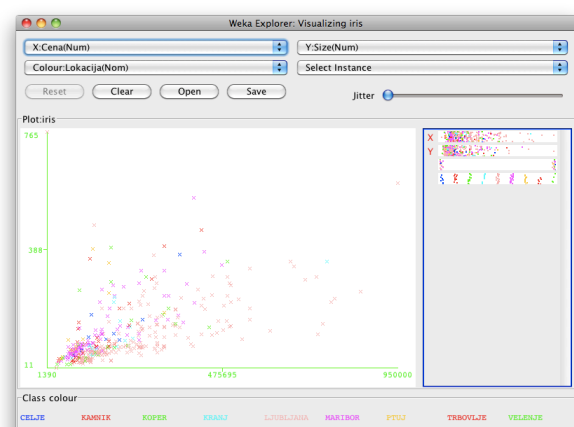
Analiza strojnega učenja z uporabo J48



Odločitveno drevo pridobljeno z J48



X: Lokacija / Y: Cena



X: Cena / Y: Velikost

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 6
 Relation: iris
 Instances:439
 Attributes:4

Cena
 Size
 Place
 Lokacija

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```

Cena <= 94000
|   Size <= 41
|   |   Cena <= 65000: KOPER (11.0/8.0)
|   |   Cena > 65000
|   |   |   Size <= 31.5: LJUBLJANA (7.0/1.0)
|   |   |   Size > 31.5: KAMNIK (6.0/2.0)
|   |   Size > 41: MARIBOR (64.0/23.0)
Cena > 94000
|   Size <= 80
|   |   Cena <= 99000
|   |   |   Size <= 44: LJUBLJANA (10.0)
|   |   |   Size > 44: MARIBOR (8.0/3.0)
|   |   Cena > 99000
|   |   |   Size <= 74.8: LJUBLJANA (107.0/19.0)
|   |   |   Size > 74.8
|   |   |   |   Cena <= 185000
|   |   |   |   |   Cena <= 145000: KAMNIK (6.0/1.0)
|   |   |   |   |   Cena > 145000: LJUBLJANA (9.0/4.0)
|   |   |   |   Cena > 185000: LJUBLJANA (15.0/1.0)
|   |   Size > 80
|   |   |   Cena <= 233037
|   |   |   |   Cena <= 172000
|   |   |   |   |   Place = Stanovanje
|   |   |   |   |   |   Size <= 102: MARIBOR (20.0/8.0)
|   |   |   |   |   |   Size > 102: CELJE (6.0/3.0)
|   |   |   |   |   Place = Hiša
|   |   |   |   |   |   Cena <= 130250: LJUBLJANA (6.0/4.0)
|   |   |   |   |   |   Cena > 130250: MARIBOR (14.0/4.0)
|   |   |   |   Cena > 172000
|   |   |   |   |   Size <= 126
|   |   |   |   |   |   Size <= 98.18: LJUBLJANA (11.0/5.0)
|   |   |   |   |   |   Size > 98.18: CELJE (7.0/4.0)
|   |   |   |   |   Size > 126: MARIBOR (18.0/12.0)
|   |   |   Cena > 233037
|   |   |   |   Size <= 303: LJUBLJANA (99.0/21.0)
|   |   |   |   Size > 303
|   |   |   |   |   Cena <= 480000: MARIBOR (9.0/4.0)
|   |   |   |   |   Cena > 480000: LJUBLJANA (6.0/2.0)

```

Number of Leaves : 20
 Size of the tree : 39
 Time taken to build model: 0seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	261	59.4533 %
Incorrectly Classified Instances	178	40.5467 %
Kappa statistic	0.3249	
Mean absolute error	0.1177	
Root mean squared error	0.2582	
Relative absolute error	78.9595 %	
Root relative squared error	94.8984 %	
Total Number of Instances	439	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.154	0.009	0.333	0.154	0.211	0.673	CELJE
	0.08	0.019	0.2	0.08	0.114	0.681	KAMNIK
	0.042	0.014	0.143	0.042	0.065	0.45	KOPER
	0	0.026	0	0	0	0.581	KRANJ
	0.869	0.395	0.706	0.869	0.779	0.783	LJUBLJANA
	0.566	0.179	0.479	0.566	0.519	0.708	MARIBOR
	0	0.007	0	0	0	0.702	PTUJ
	0.125	0.005	0.333	0.125	0.182	0.63	TRBOVLJE
	0	0	0	0	0	0.494	VELENJE
Weighted Avg.	0.595	0.25	0.511	0.595	0.543	0.72	

```
=== Confusion Matrix ===
```

a	b	c	d	e	f	g	h	i	<-- classified as
2	0	0	2	4	5	0	0	0	a = CELJE
0	2	0	1	12	10	0	0	0	b = KAMNIK
0	0	1	0	18	5	0	0	0	c = KOPER
0	1	0	0	13	9	0	0	0	d = KRANJ
3	4	4	4	199	15	0	0	0	e = LJUBLJANA
1	3	0	4	32	56	2	1	0	f = MARIBOR
0	0	0	0	2	11	0	0	0	g = PTUJ
0	0	2	0	2	2	1	1	0	h = TRBOVLJE
0	0	0	0	0	4	0	1	0	i = VELENJE

Primerjava rezultatov strojnega učenja

Algoritem	Usp. klasifikacije	Kappa statistika	Pov. preklic	Pov. preciznost	F-merilo	Pov. pov. pod ROC
J48	59,4533	0,3249	0,595	0,511	0,543	0,72
BFTree	62,6424	3,429	0,626	0,525	0,555	0,713