

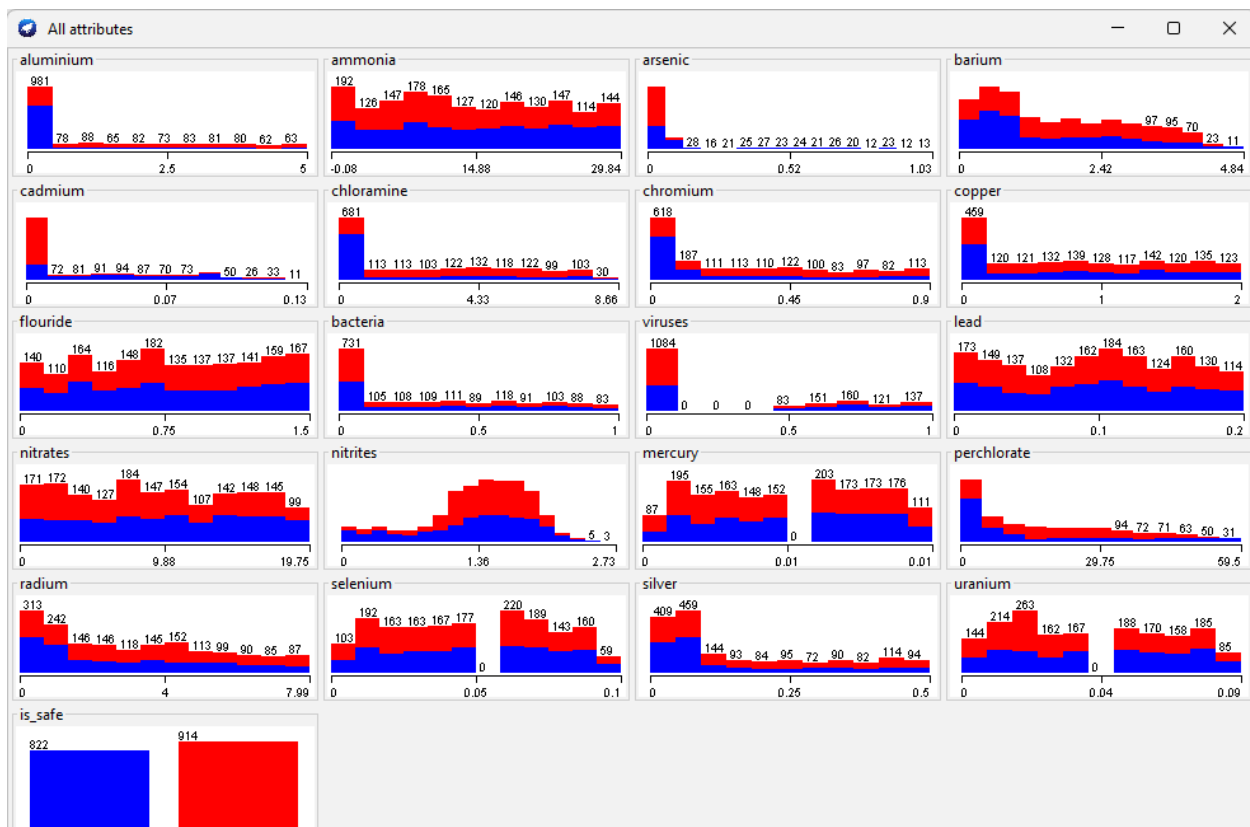
Skup podataka

Dataset: <https://www.kaggle.com/datasets/mssmartypants/water-quality>

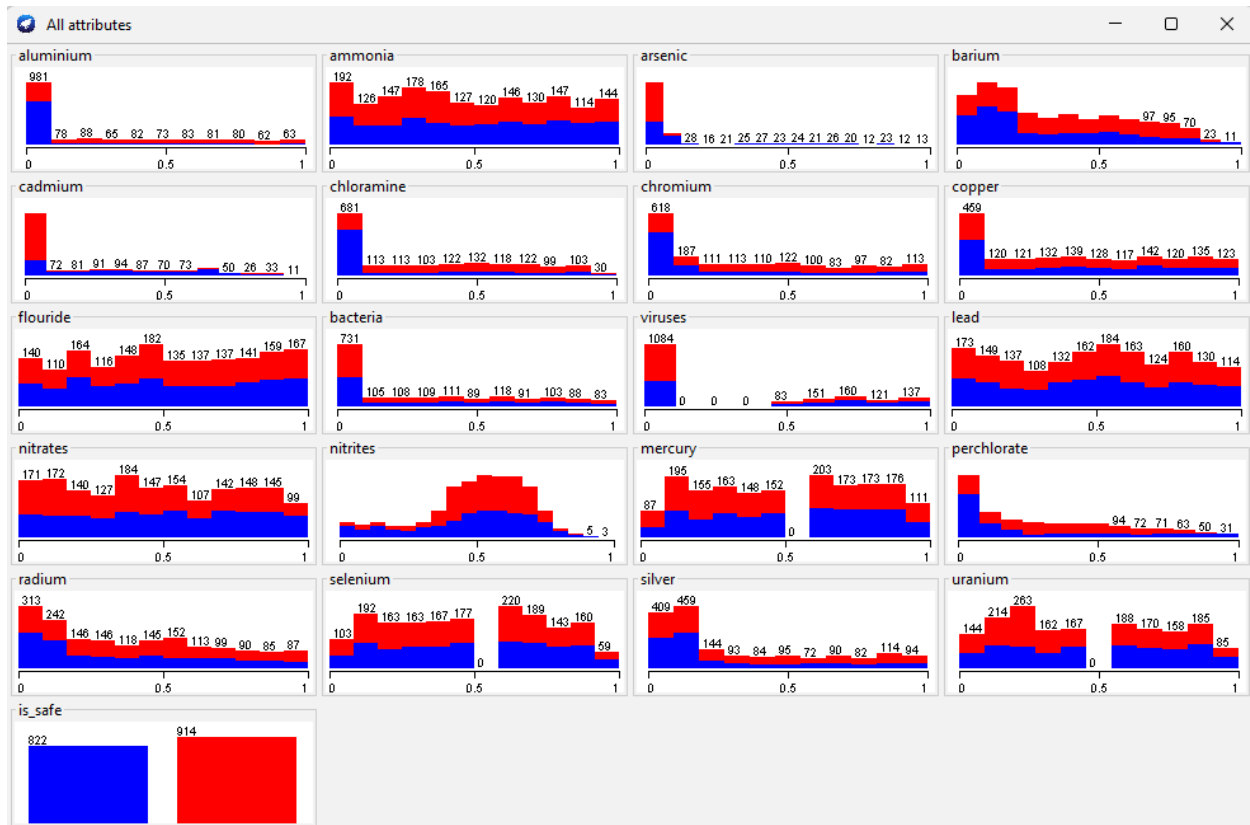
Sadržaj dataseta su supstance koje se sadrže u vodi i klasa koja označava da li je voda bezbedna za pice.

Cilj: Predvideti da li je voda bezbedna za pice.

Pocetni podaci:



Normalizovani podaci:



Priprema skupa podataka za dalju analizu

Ciscenje podataka od pogresnih vrednosti: U datasetu nemamo nevalidnih vrednosti ili vrednosti koje odskacu od opsega.

Skaliranje i normalizacija podataka: sve kolone imaju numericke vrednosti koje je moguće normalizovati osim klase koja ima vrednosti 0 i 1. Normalizujemo podatke iz razloga da ako neke kolone imaju razlicite vrednosti da se ne da posebna vaznost visokom vrednostima, normalizujemo ih tako da imaju jednaku vaznost.

Primena algoritma

1. NaiveBayes sa cross validation

The screenshot shows the Orange3 software interface with the NaiveBayes classifier selected. The 'Test options' section on the left is configured for cross-validation with 10 folds. The 'Classifier output' section on the right displays the results of the cross-validation, including a summary of performance metrics and a detailed accuracy by class table.

Test options:

- ☐ Use training set
- ☐ Supplied test set (Set...)
- ☒ Cross-validation (Folds: 10)
- ☐ Percentage split (%: 66)
- More options...

Classifier output:

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	1360	78.341 %
Incorrectly Classified Instances	376	21.659 %
Kappa statistic	0.5664	
Mean absolute error	0.2473	
Root mean squared error	0.4317	
Relative absolute error	49.6053 %	
Root relative squared error	86.469 %	
Total Number of Instances	1736	

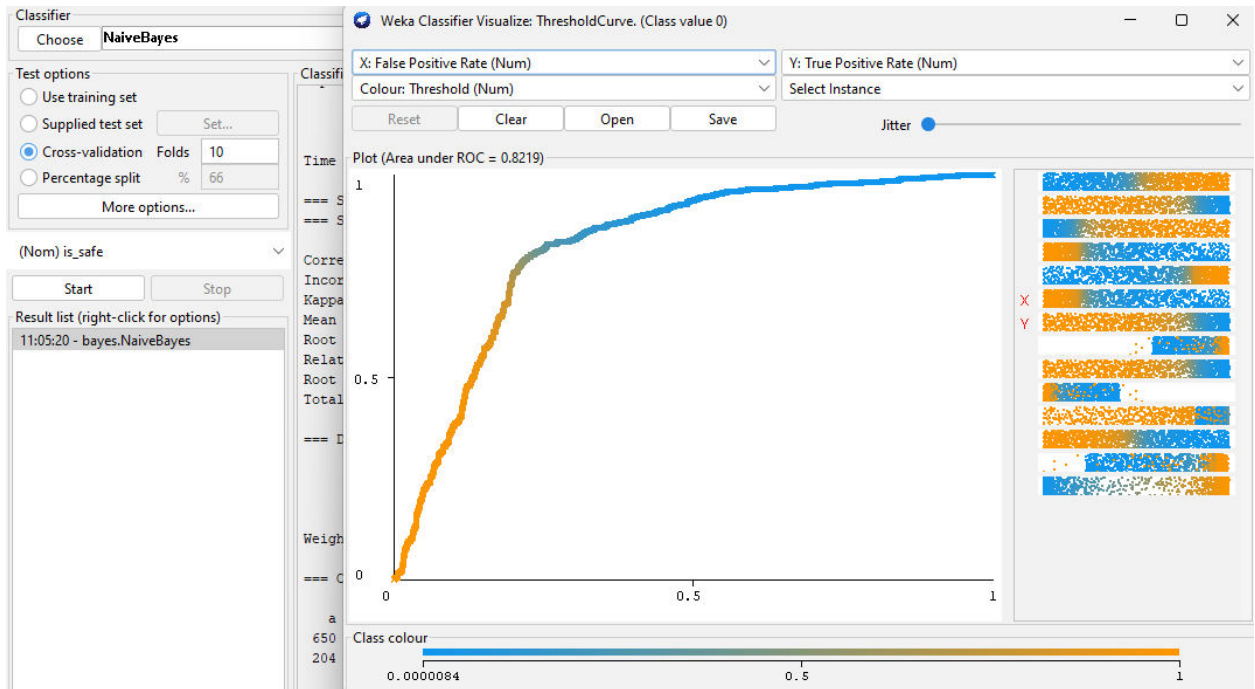
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.791	0.223	0.761	0.791	0.776	0.567	0.822	0.741	0
	0.777	0.209	0.805	0.777	0.791	0.567	0.822	0.843	1
Weighted Avg.	0.783	0.216	0.784	0.783	0.784	0.567	0.822	0.795	

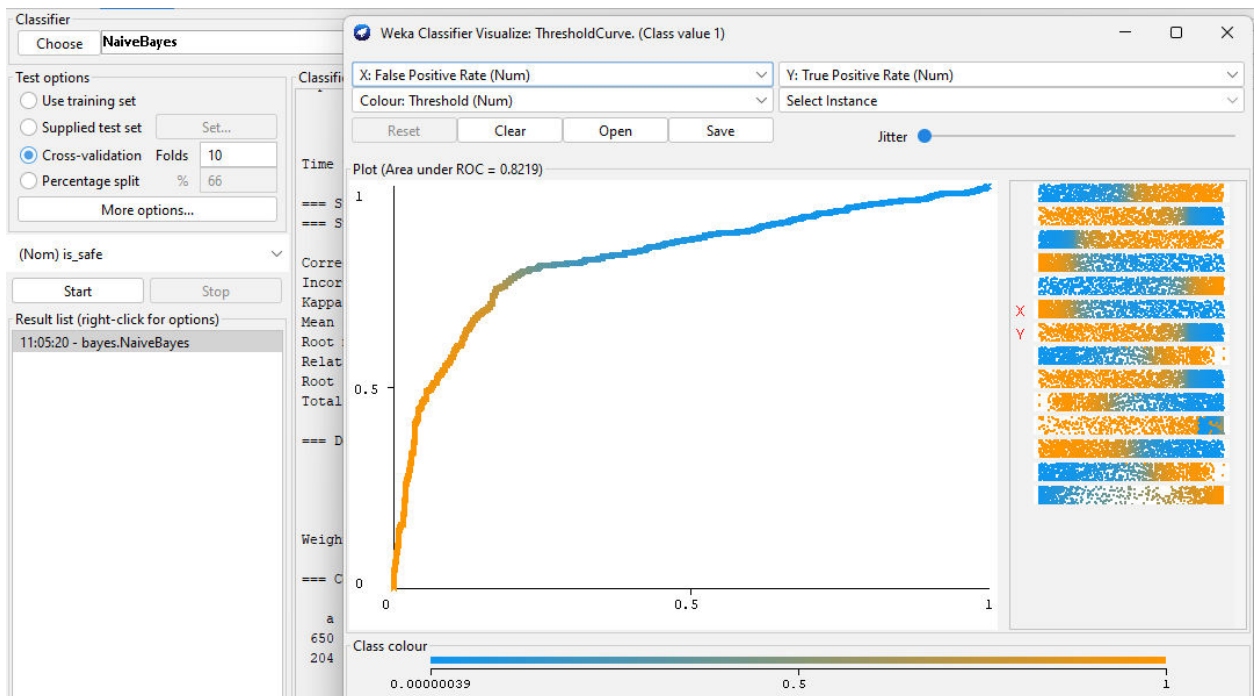
=== Confusion Matrix ===

```
a  b  <-- classified as
650 172 | a = 0
204 710 | b = 1
```

Roc kriva za klasnu vrednost 0:



Roc kriva za klasnu vrednost 1:



Primecujemo da su ROC vrednosti iste 0,82. Tj. da se jednako dobro predviđaju 0 i 1 vrednosti

Kako NaiveBayes algoritam pretpostavlja da su atributi nezavisni a normalizacija neće promeniti medjusobne odnose izmedju njih normalizacija neće znacajno uticati na rezultate,takodje ovaj algoritam nije osetljiv na razlike u opsegu atributa pa su rezultati sa normalizovanim i nenormalizovanim podacima slicni.

2. Koriscenjem percentage split opcije 70%, 80%, 90% redom:

Classifier

ChooseNaiveBayes

Test options

☐ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds

10

☒ Percentage split

%

70

More options...

(Nom) is_safe

Start

Stop

Result list (right-click for options)

11:05:20 - bayes.NaiveBayes

11:08:29 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	418	80.2303 %
Incorrectly Classified Instances	103	19.7697 %
Kappa statistic	0.604	
Mean absolute error	0.2348	
Root mean squared error	0.411	
Relative absolute error	47.0587 %	
Root relative squared error	82.2311 %	
Total Number of Instances	521	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.793	0.189	0.796	0.793	0.794	0.604	0.844	0.781	0
	0.811	0.207	0.808	0.811	0.810	0.604	0.844	0.870	1
Weighted Avg.	0.802	0.198	0.802	0.802	0.802	0.604	0.844	0.827	

=== Confusion Matrix ===

a

b

<-- classified as

199

52

|

a = 0

51

219

|

b = 1

Classifier

Choose NaiveBayes

Test options

☐ Use training set
☐ Supplied test set
☐ Cross-validation Folds 10
☒ Percentage split % 80

(Nom) is_safe

Result list (right-click for options)

11:05:20 - bayes.NaiveBayes
11:08:29 - bayes.NaiveBayes
11:08:57 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	286	82.4207 %
Incorrectly Classified Instances	61	17.5793 %
Kappa statistic	0.648	
Mean absolute error	0.2145	
Root mean squared error	0.3898	
Relative absolute error	42.9439 %	
Root relative squared error	77.903 %	
Total Number of Instances	347	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.153	0.834	0.800	0.817	0.649	0.857	0.805	0
	0.847	0.200	0.815	0.847	0.831	0.649	0.857	0.868	1
Weighted Avg.	0.824	0.177	0.825	0.824	0.824	0.649	0.857	0.837	

=== Confusion Matrix ===

a b <-- classified as
136 34 | a = 0
27 150 | b = 1

Classifier

Choose NaiveBayes

Test options

☐ Use training set
☐ Supplied test set
☐ Cross-validation Folds 10
☒ Percentage split % 90

(Nom) is_safe

Result list (right-click for options)

11:05:20 - bayes.NaiveBayes
11:08:29 - bayes.NaiveBayes
11:08:57 - bayes.NaiveBayes
11:09:17 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	145	83.3333 %
Incorrectly Classified Instances	29	16.6667 %
Kappa statistic	0.6643	
Mean absolute error	0.2063	
Root mean squared error	0.3757	
Relative absolute error	41.4035 %	
Root relative squared error	75.3089 %	
Total Number of Instances	174	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.802	0.140	0.833	0.802	0.818	0.665	0.870	0.815	0
	0.860	0.198	0.833	0.860	0.847	0.665	0.870	0.880	1
Weighted Avg.	0.833	0.171	0.833	0.833	0.833	0.665	0.870	0.850	

=== Confusion Matrix ===

a b <-- classified as
65 16 | a = 0
13 80 | b = 1

3. ZeroR

Classifier

ChooseZeroR

Test options

☐ Use training set

☐ Supplied test set

Set...

☒ Cross-validation

Folds10

☐ Percentage split

%90

More options...

(Nom) is_safe

Start

Stop

Result list (right-click for options)

11:05:20 - bayes.NaiveBayes

11:08:29 - bayes.NaiveBayes

11:08:57 - bayes.NaiveBayes

11:09:17 - bayes.NaiveBayes

11:09:58 - rules.ZeroR

Classifier output

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

ZeroR predicts class value: 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	914	52.6498 %
Incorrectly Classified Instances	822	47.3502 %
Kappa statistic	0	
Mean absolute error	0.4986	
Root mean squared error	0.4993	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.498	0.472	0
	1.000	1.000	0.526	1.000	0.690	?	0.498	0.525	1
Weighted Avg.	0.526	0.526	?	0.526	?	?	0.498	0.500	

=== Confusion Matrix ===

a b <-- classified as

0	822		a = 0
0	914		b = 1

4. OneR

Classifier

Choose **OneR -B 6**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR

Classifier output

>= 0.0095 -> 0
(1397/1736 instances correct)

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1397	80.4724 %
Incorrectly Classified Instances	339	19.5276 %
Kappa statistic	0.6105	
Mean absolute error	0.1953	
Root mean squared error	0.4419	
Relative absolute error	39.1651 %	
Root relative squared error	88.5044 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.848	0.234	0.765	0.848	0.804	0.614	0.807	0.721	0
	0.766	0.152	0.848	0.766	0.805	0.614	0.807	0.773	1
Weighted Avg.	0.805	0.191	0.809	0.805	0.805	0.614	0.807	0.748	

=== Confusion Matrix ===

a	b	<-- classified as
697	125	a = 0
214	700	b = 1

OneR je jednostavan klasifikator koji bira jedan atribut i za njega definiše pravila na osnovu kojih se radi klasifikacija. U našem slučaju je to aluminium atribut.

5. K Nearest Neighbors (kNN)

KNN sa vrednoscu 1:

Classifier

Choose **IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk

Classifier output

IBk instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1325	76.3249 %
Incorrectly Classified Instances	411	23.6751 %
Kappa statistic	0.523	
Mean absolute error	0.2371	
Root mean squared error	0.4863	
Relative absolute error	47.5508 %	
Root relative squared error	97.3887 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.706	0.185	0.774	0.706	0.738	0.525	0.763	0.697	0
	0.815	0.294	0.755	0.815	0.784	0.525	0.763	0.723	1
Weighted Avg.	0.763	0.243	0.764	0.763	0.762	0.525	0.763	0.711	

=== Confusion Matrix ===

a	b	<-- classified as
580	242	a = 0
169	745	b = 1

KNN sa vrednoscu 6:

Classifier

Choose **IBk -K 8 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk

Classifier output

IB1 instance-based classifier
using 6 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1439	82.8917 %
Incorrectly Classified Instances	297	17.1083 %
Kappa statistic	0.6557	
Mean absolute error	0.2436	
Root mean squared error	0.3562	
Relative absolute error	48.8614 %	
Root relative squared error	71.3467 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.786	0.132	0.842	0.786	0.813	0.657	0.897	0.889	0
	0.868	0.214	0.818	0.868	0.842	0.657	0.897	0.874	1

=== Confusion Matrix ===

```

a  b  <-- classified as
646 176 |  a = 0
121 793 |  b = 1

```

KNN sa vrednoscu 9:

Classifier

Choose **IBk -K 9 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk

Classifier output

IB1 instance-based classifier
using 9 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1431	82.4309 %
Incorrectly Classified Instances	305	17.5691 %
Kappa statistic	0.644	
Mean absolute error	0.2517	
Root mean squared error	0.3523	
Relative absolute error	50.4814 %	
Root relative squared error	70.562 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.713	0.075	0.895	0.713	0.794	0.657	0.905	0.902	0
	0.925	0.287	0.782	0.925	0.847	0.657	0.905	0.890	1

=== Confusion Matrix ===

```

a  b  <-- classified as
586 236 |  a = 0
69 845 |  b = 1

```

6. J48 algoritam

Classifier

Choose

J48 -C 0.25 -M 2

Test options

☐ Use training set
☐ Supplied test set

Set...

☒ Cross-validation Folds

10

☐ Percentage split %

90

More options...

(Nom) is_safe

Start

Stop

Result list (right-click for options)

11:05:20 - bayes.NaiveBayes

11:08:29 - bayes.NaiveBayes

11:08:57 - bayes.NaiveBayes

11:09:17 - bayes.NaiveBayes

11:09:58 - rules.ZeroR

11:10:37 - rules.OneR

11:11:18 - lazy.JBk

11:11:46 - lazy.JBk

11:12:10 - lazy.JBk

11:12:16 - lazy.JBk

11:12:25 - lazy.JBk

11:12:31 - lazy.JBk

11:13:07 - lazy.JBk

11:13:15 - lazy.JBk

11:13:21 - lazy.JBk

11:14:22 - trees.J48

Classifier output

Size of the tree : 101

Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1616	93.0876 %
Incorrectly Classified Instances	120	6.9124 %
Kappa statistic	0.8613	
Mean absolute error	0.0777	
Root mean squared error	0.2521	
Relative absolute error	15.5795 %	
Root relative squared error	50.4931 %	
Total Number of Instances	1736	

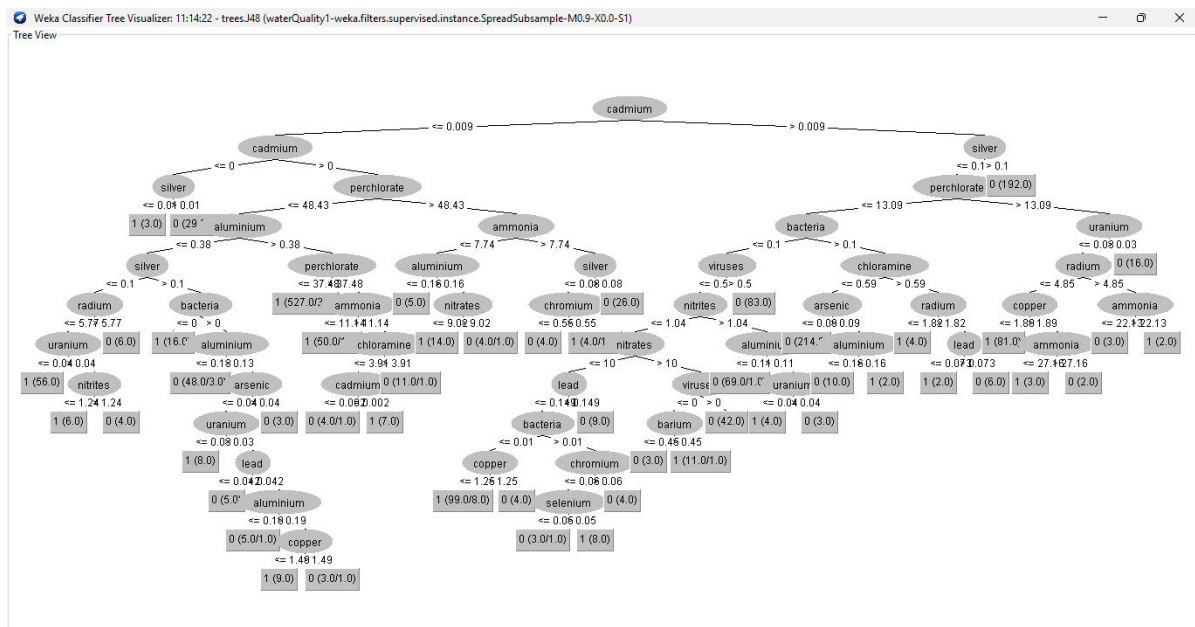
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.925	0.063	0.929	0.925	0.927	0.861	0.941	0.925	0
	0.937	0.075	0.932	0.937	0.934	0.861	0.941	0.910	1
Weighted Avg.	0.931	0.070	0.931	0.931	0.931	0.861	0.941	0.917	

=== Confusion Matrix ===

a	b	<-- classified as
760	62	a = 0
58	856	b = 1

Izgled stable:



Ukoliko izaberemo da postoji odsecanje stable dobijamo za nijansu losije rezultate (92,1083%)

7. Random forest

Classifier

Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest

Classifier output

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.99 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1635	94.182 %
Incorrectly Classified Instances	101	5.818 %
Kappa statistic	0.8831	
Mean absolute error	0.1489	
Root mean squared error	0.2307	
Relative absolute error	29.8602 %	
Root relative squared error	46.2114 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.923	0.042	0.952	0.923	0.938	0.884	0.984	0.983	0
	0.958	0.077	0.933	0.958	0.945	0.884	0.984	0.984	1
Weighted Avg.	0.942	0.060	0.942	0.942	0.942	0.884	0.984	0.983	

=== Confusion Matrix ===

a	b	-- classified as
759	63	a = 0
38	876	b = 1

8. Classification via Regression

Classifier

Choose **ClassificationViaRegression** -W weka.classifiers.trees.M5P -- -M 4.0 -num-decimal-places 4

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe

Result list (right-click for options)

- 11:05:20 - bayes.NaiveBayes
- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest
- 11:17:04 - meta.ClassificationViaRegression

Classifier output

Number of Rules : 62

Time taken to build model: 1.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1600	92.1659 %
Incorrectly Classified Instances	136	7.8341 %
Kappa statistic	0.8425	
Mean absolute error	0.1296	
Root mean squared error	0.2477	
Relative absolute error	25.9912 %	
Root relative squared error	49.6145 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.895	0.055	0.936	0.895	0.915	0.843	0.975	0.973	0
	0.945	0.105	0.909	0.945	0.927	0.843	0.975	0.978	1
Weighted Avg.	0.922	0.081	0.922	0.922	0.922	0.843	0.975	0.976	

=== Confusion Matrix ===

a	b	-- classified as
736	86	a = 0
50	864	b = 1

9. JRip

Sa odsecanjem

Classifier

Choose **JRip -F 3 -N 2.0 -O 2 -S 1**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe ☒

Result list (right-click for options)

- 11:08:29 - bayes.NaiveBayes
- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest
- 11:17:04 - meta.ClassificationViaRegression
- 11:17:44 - rules.JRip

Classifier output

Number of Rules : 12

Time taken to build model: 0.53 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1582	91.129 %
Incorrectly Classified Instances	154	8.871 %
Kappa statistic	0.8218	
Mean absolute error	0.1196	
Root mean squared error	0.2792	
Relative absolute error	23.9834 %	
Root relative squared error	55.9193 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.891	0.070	0.920	0.891	0.905	0.822	0.931	0.939	0
	0.930	0.109	0.904	0.930	0.917	0.822	0.931	0.896	1
Weighted Avg.	0.911	0.091	0.912	0.911	0.911	0.822	0.931	0.916	

=== Confusion Matrix ===

a	b	<-- classified as
732	90	a = 0
64	850	b = 1

Bez odsecanja

Classifier

Choose **JRip -F 3 -N 2.0 -O 2 -S 1 -P**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) is_safe ☒

Result list (right-click for options)

- 11:08:57 - bayes.NaiveBayes
- 11:09:17 - bayes.NaiveBayes
- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest
- 11:17:04 - meta.ClassificationViaRegression
- 11:17:44 - rules.JRip
- 11:18:26 - rules.JRip

Classifier output

Number of Rules : 24

Time taken to build model: 0.16 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1605	92.4539 %
Incorrectly Classified Instances	131	7.5461 %
Kappa statistic	0.8482	
Mean absolute error	0.0818	
Root mean squared error	0.2724	
Relative absolute error	16.4062 %	
Root relative squared error	54.5558 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.892	0.046	0.946	0.892	0.918	0.849	0.920	0.909	0
	0.954	0.108	0.907	0.954	0.930	0.849	0.920	0.884	1
Weighted Avg.	0.925	0.079	0.926	0.925	0.924	0.849	0.920	0.896	

=== Confusion Matrix ===

a	b	<-- classified as
733	89	a = 0
42	872	b = 1

Selekcija atributa I primena algoritma

Weka je dosla do zakljucka da su samo ovi atributi bitni za odredjivanje klase:

No.	
1	<input type="checkbox"/> aluminium
2	<input type="checkbox"/> cadmium
3	<input type="checkbox"/> nitrates
4	<input type="checkbox"/> uranium
5	<input type="checkbox"/> is_safe

NaiveBayes sa cross validacijom:

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

Set...

☐ Cross-validation Folds 10

☒ Percentage split % 90

More options...

(Nom) is_safe

Start

Stop

Result list (right-click for options)

11:09:17 - bayes.NaiveBayes

11:09:58 - rules.ZeroR

11:10:37 - rules.OneR

11:11:18 - lazy.IBk

11:11:46 - lazy.IBk

11:12:10 - lazy.IBk

11:12:16 - lazy.IBk

11:12:25 - lazy.IBk

11:12:31 - lazy.IBk

11:13:07 - lazy.IBk

11:13:15 - lazy.IBk

11:13:21 - lazy.IBk

11:14:22 - trees.J48

11:15:29 - trees.J48

11:16:11 - trees.RandomForest

11:17:04 - meta.ClassificationViaRegression

11:17:44 - rules.JRip

11:18:26 - rules.JRip

11:21:19 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1339	77.1313 %
Incorrectly Classified Instances	397	22.8687 %
Kappa statistic	0.5419	
Mean absolute error	0.3168	
Root mean squared error	0.4216	
Relative absolute error	63.5432 %	
Root relative squared error	84.4413 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.771	0.229	0.752	0.771	0.762	0.542	0.819	0.750	0
	0.771	0.229	0.789	0.771	0.780	0.542	0.819	0.833	1
Weighted Avg.	0.771	0.229	0.772	0.771	0.771	0.542	0.819	0.794	

=== Confusion Matrix ===

a b <-- classified as

634	188		a = 0
209	705		b = 1

J48:

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % **90**

More options...

(Nom) is_safe

Start Stop

Result list (right-click for options)

- 11:09:58 - rules.ZeroR
- 11:10:37 - rules.OneR
- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest
- 11:17:04 - meta.ClassificationViaRegression
- 11:17:44 - rules.JRip
- 11:18:26 - rules.JRip
- 11:21:19 - bayes.NaiveBayes
- 11:22:33 - trees.J48

Classifier output

Size of the tree : 53

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1422	81.9124 %
Incorrectly Classified Instances	314	18.0876 %
Kappa statistic	0.6375	
Mean absolute error	0.2457	
Root mean squared error	0.3735	
Relative absolute error	49.2875 %	
Root relative squared error	74.8023 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.818	0.179	0.804	0.818	0.811	0.638	0.851	0.767	0
	0.821	0.182	0.833	0.821	0.827	0.638	0.851	0.837	1
Weighted Avg.	0.819	0.181	0.819	0.819	0.819	0.638	0.851	0.804	

=== Confusion Matrix ===

a	b	<-- classified as	
672	150		a = 0
164	750		b = 1

RandomForest:

Classifier

Choose **RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % **90**

More options...

(Nom) is_safe

Start Stop

Result list (right-click for options)

- 11:11:18 - lazy.IBk
- 11:11:46 - lazy.IBk
- 11:12:10 - lazy.IBk
- 11:12:16 - lazy.IBk
- 11:12:25 - lazy.IBk
- 11:12:31 - lazy.IBk
- 11:13:07 - lazy.IBk
- 11:13:15 - lazy.IBk
- 11:13:21 - lazy.IBk
- 11:14:22 - trees.J48
- 11:15:29 - trees.J48
- 11:16:11 - trees.RandomForest
- 11:17:04 - meta.ClassificationViaRegression
- 11:17:44 - rules.JRip
- 11:18:26 - rules.JRip
- 11:21:19 - bayes.NaiveBayes
- 11:22:33 - trees.J48
- 11:22:58 - trees.J48
- 11:23:06 - trees.RandomForest

Classifier output

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.47 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1420	81.7972 %
Incorrectly Classified Instances	316	18.2028 %
Kappa statistic	0.6358	
Mean absolute error	0.2391	
Root mean squared error	0.3619	
Relative absolute error	47.9582 %	
Root relative squared error	72.4871 %	
Total Number of Instances	1736	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.831	0.194	0.794	0.831	0.812	0.636	0.891	0.870	0
	0.806	0.169	0.841	0.806	0.823	0.636	0.891	0.894	1
Weighted Avg.	0.818	0.181	0.819	0.818	0.818	0.636	0.891	0.883	

=== Confusion Matrix ===

a	b	<-- classified as	
683	139		a = 0
177	737		b = 1

Svaki algoritam koji radi nad selektovanim atributima daje losije rezultate

Najbolji rezultat:

RandomForest

Poredjenje:

Actions

Perform test

Save output

Open Explorer...

Configure test

Testing with

Paired T-Tester (corrected)

Select rows and cols

Rows

Cols

Swap

Comparison field

Percent_correct

Significance

0.05

Sorting (asc.) by

<default>

Test base

Select

Displayed Columns

Select

Show std. deviations

Output Format

Select

Test output

Tester: weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-matrix "weka.experiment.ResultMatrixPlainText -mean-prec

Analysing: Percent_correct

Datasets: 2

Resultsets: 8

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 6/6/23, 11:31 AM

Dataset

(1) bayes.Na (2) lazy. (3) meta. (4) rules (5) rules (6) rules (7) trees (8) trees

waterQuality1-weka.filter(100)

78.36 |

82.70 v

91.90 v

91.36 v

80.45 v

52.65 *

93.15 v

94.42 v

waterQuality1-weka.filter(100)

78.36 |

82.70 v

91.87 v

91.88 v

80.45 v

52.65 *

93.15 v

94.60 v

(v/ /*) |

(2/0/0)

(2/0/0)

(2/0/0)

(2/0/0)

(0/0/2)

(2/0/0)

(2/0/0)

Result list

11:31:01 - Available resultsets

11:31:13 - Number_correct - bayes.NaiveBayes " 5995231201781

11:31:24 - Percent_correct - bayes.NaiveBayes " 5995231201785

Key:

(1) bayes.NaiveBayes '' 5995231201785697655

(2) lazy.IBk '-K 6 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\\\" -308018609877706

(3) meta.ClassificationViaRegression '-W trees.M5P -- -M 4.0 -num-decimal-places 4' 4500023123618669859

(4) rules.JRip '-F 3 -N 2.0 -O 2 -S 1' -6589312996832147161

(5) rules.OneR '-B 6' -3459427003147861443

(6) rules.ZeroR '' 48055541465867954

(7) trees.J48 '-C 0.25 -M 2' -217733168393644444

(8) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698