

Prueba final Weka

Curso: Algoritmos de Big Data para ingenieros.

Estudiante: Juan Pablo Delzo Melendez

PRIMER EJERCICIO:

1. J48

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: segment

Instances: 1500

Attributes: 20

- region-centroid-col
- region-centroid-row
- region-pixel-count
- short-line-density-5
- short-line-density-2
- vedge-mean
- vedge-sd
- hedge-mean
- hedge-sd
- intensity-mean
- rawred-mean
- rawblue-mean
- rawgreen-mean
- exred-mean
- exblue-mean
- exgreen-mean
- value-mean
- saturation-mean
- hue-mean
- class

Test mode: 10-fold cross-validation

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

J48 pruned tree

```
region-centroid-row <= 155
| value-mean <= 91.4444
| | rawred-mean <= 24.6667
```

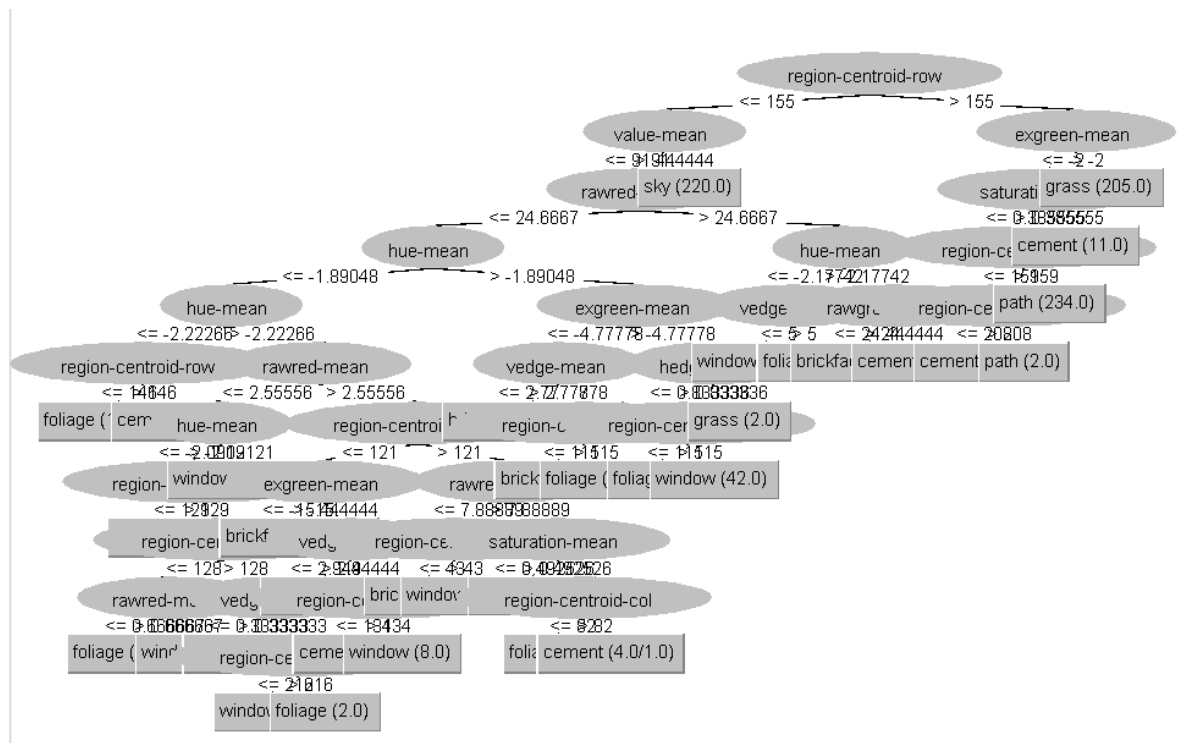
```

| | | hue-mean <= -1.89048
| | | | hue-mean <= -2.22266
| | | | | region-centroid-row <= 146: foliage (102.0/1.0)
| | | | | region-centroid-row > 146: cement (3.0)
| | | | hue-mean > -2.22266
| | | | | rawred-mean <= 2.55556
| | | | | | hue-mean <= -2.09121
| | | | | | | region-centroid-row <= 129: foliage (50.0)
| | | | | | | region-centroid-row > 129
| | | | | | | | region-centroid-col <= 128
| | | | | | | | | rawred-mean <= 0.666667: foliage (30.0/4.0)
| | | | | | | | | rawred-mean > 0.666667: window (5.0)
| | | | | | | | | region-centroid-col > 128
| | | | | | | | | vedge-mean <= 0.333333: window (11.0)
| | | | | | | | | vedge-mean > 0.333333
| | | | | | | | | | region-centroid-col <= 216: window (3.0)
| | | | | | | | | | region-centroid-col > 216: foliage (2.0)
| | | | | | | | | | hue-mean > -2.09121: window (38.0/1.0)
| | | | | | | | | rawred-mean > 2.55556
| | | | | | | | | region-centroid-row <= 121
| | | | | | | | | | exgreen-mean <= -15.4444: brickface (2.0/1.0)
| | | | | | | | | | exgreen-mean > -15.4444
| | | | | | | | | | vedge-mean <= 2.94444: window (75.0)
| | | | | | | | | | vedge-mean > 2.94444
| | | | | | | | | | region-centroid-col <= 134: cement (2.0)
| | | | | | | | | | region-centroid-col > 134: window (8.0)
| | | | | | | | | region-centroid-row > 121
| | | | | | | | | | rawred-mean <= 7.88889
| | | | | | | | | | region-centroid-col <= 43: brickface (2.0)
| | | | | | | | | | region-centroid-col > 43: window (13.0/2.0)
| | | | | | | | | rawred-mean > 7.88889
| | | | | | | | | | saturation-mean <= 0.492526: cement (15.0)
| | | | | | | | | | saturation-mean > 0.492526
| | | | | | | | | | region-centroid-col <= 82: foliage (2.0)
| | | | | | | | | | region-centroid-col > 82: cement (4.0/1.0)
| | | | hue-mean > -1.89048
| | | | | exgreen-mean <= -4.77778
| | | | | | vedge-mean <= 2.77778: brickface (198.0/2.0)
| | | | | | vedge-mean > 2.77778
| | | | | | | region-centroid-row <= 115: brickface (4.0)
| | | | | | | region-centroid-row > 115: foliage (3.0/1.0)
| | | | | | | exgreen-mean > -4.77778
| | | | | | | | hedge-mean <= 0.833336
| | | | | | | | region-centroid-col <= 115: foliage (4.0)
| | | | | | | | region-centroid-col > 115: window (42.0)
| | | | | | | | hedge-mean > 0.833336: grass (2.0)
| | | rawred-mean > 24.6667
| | | | hue-mean <= -2.17742

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	
Class									
	0,956	0,004	0,975	0,956	0,966	0,960	0,981	0,954	brickface
	1,000	0,001	0,995	1,000	0,998	0,997	1,000	0,995	sky
	0,942	0,018	0,895	0,942	0,918	0,905	0,975	0,889	foliage
	0,941	0,009	0,945	0,941	0,943	0,933	0,978	0,946	cement
	0,877	0,017	0,891	0,877	0,884	0,866	0,961	0,881	window
	0,987	0,001	0,996	0,987	0,991	0,990	0,997	0,992	path
	0,990	0,000	1,000	0,990	0,995	0,994	1,000	1,000	grass
Weighted Avg.	0,957	0,007	0,958	0,957	0,957	0,951	0,985	0,952	

```
a b c d e f g <-- classified as
196 0 3 1 5 0 0 | a = brickface
0 220 0 0 0 0 0 | b = sky
0 1 196 2 9 0 0 | c = foliage
2 0 4 207 6 1 0 | d = cement
3 0 16 6 179 0 0 | e = window
0 0 0 3 0 233 0 | f = path
0 0 0 0 2 0 205 | g = grass
```



2. Random Forest

=== Run information ===

```
Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0
-V 0.001 -S 1
```

Relation: segment

Instances: 1500

Attributes: 20

region-centroid-col
region-centroid-row
region-pixel-count
short-line-density-5

short-line-density-2
 vedge-mean
 vegde-sd
 hedge-mean
 hedge-sd
 intensity-mean
 rawred-mean
 rawblue-mean
 rawgreen-mean
 exred-mean
 exblue-mean
 exgreen-mean
 value-mean
 saturation-mean
 hue-mean
 class

Test mode: 10-fold cross-validation

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

RandomForest

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: 0.27 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1468	97.8667 %
Incorrectly Classified Instances	32	2.1333 %
Kappa statistic	0.9751	
Mean absolute error	0.02	
Root mean squared error	0.0786	
Relative absolute error	8.1639 %	
Root relative squared error	22.4623 %	
Total Number of Instances	1500	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
brickface	0,990	0,001	0,995	0,990	0,993	0,992	1,000	1,000
sky	1,000	0,001	0,995	1,000	0,998	0,997	1,000	1,000
foliage	0,966	0,009	0,948	0,966	0,957	0,950	0,998	0,988
cement	0,968	0,006	0,964	0,968	0,966	0,960	0,998	0,993

	0,926	0,008	0,950	0,926	0,938	0,928	0,997	0,983	window
	0,996	0,001	0,996	0,996	0,996	0,995	1,000	1,000	path
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	grass
Weighted Avg.	0,979	0,003	0,979	0,979	0,979	0,975	0,999	0,995	

=== Confusion Matrix ===

```

a b c d e f g <-- classified as
203 0 0 0 2 0 0 | a = brickface
0 220 0 0 0 0 0 | b = sky
0 1 201 2 4 0 0 | c = foliage
0 0 2 213 4 1 0 | d = cement
1 0 9 5 189 0 0 | e = window
0 0 0 1 0 235 0 | f = path
0 0 0 0 0 0 207 | g = grass

```

3. IBK (KNN)

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A
 "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R
 first-last\""

Relation: segment

Instances: 1500

Attributes: 20

```

region-centroid-col
region-centroid-row
region-pixel-count
short-line-density-5
short-line-density-2
vedge-mean
vedge-sd
hedge-mean
hedge-sd
intensity-mean
rawred-mean
rawblue-mean
rawgreen-mean
exred-mean
exblue-mean
exgreen-mean
value-mean
saturation-mean
hue-mean
class

```

Test mode: 10-fold cross-validation

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

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1443	96.2 %
Incorrectly Classified Instances	57	3.8 %
Kappa statistic	0.9556	
Mean absolute error	0.012	
Root mean squared error	0.1039	
Relative absolute error	4.9184 %	
Root relative squared error	29.7112 %	
Total Number of Instances	1500	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	
Class									
	0,995	0,003	0,981	0,995	0,988	0,986	0,996	0,978	brickface
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	sky
	0,942	0,013	0,920	0,942	0,931	0,920	0,964	0,877	foliage
	0,914	0,006	0,962	0,914	0,937	0,927	0,956	0,892	cement
	0,877	0,019	0,882	0,877	0,880	0,861	0,932	0,790	window
	1,000	0,003	0,983	1,000	0,992	0,990	0,998	0,984	path
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	grass
Weighted Avg.	0,962	0,006	0,962	0,962	0,962	0,962	0,956	0,978	0,933

=== Confusion Matrix ===

```
a b c d e f g <-- classified as
204 0 0 0 1 0 0 | a = brickface
0 220 0 0 0 0 0 | b = sky
0 0 196 0 12 0 0 | c = foliage
2 0 2 201 11 4 0 | d = cement
2 0 15 8 179 0 0 | e = window
0 0 0 0 0 236 0 | f = path
0 0 0 0 0 0 207 | g = grass
```

4. Conclusión:

El mejor método es Random Forest (97% de precisión), seguido de KNN (96%) y por último de J48(Decision Tree con 96%).

