

Para los tres modelos he aplicado la siguiente configuración de opciones para los datos:

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

Modelo 1

Logistic Regression

Time taken to build model: 3.27 seconds

Correctly Classified Instances	1441	96.0667 %
Incorrectly Classified Instances	59	3.9333 %
Kappa statistic	0.9541	
Mean absolute error	0.016	
Root mean squared error	0.0989	
Relative absolute error	6.5376 %	
Root relative squared error	28.2815 %	
Total Number of Instances	1500	

Modelo 2

RandomForest

Bagging with 100 iterations and base learner

Time taken to build model: 0.28 seconds

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	

Modelo 3

Naive Bayes Classifier

Time taken to build model: 0.01 seconds

Correctly Classified Instances	1216	81.0667 %
Incorrectly Classified Instances	284	18.9333 %
Kappa statistic	0.7791	

Mean absolute error	0.0554
Root mean squared error	0.2258
Relative absolute error	22.6144 %
Root relative squared error	64.5548 %
Total Number of Instances	1500

De los tres modelos de clasificación aplicados, el modelo que funcionaría mejor en este caso sería el de Random Forest. Como podemos observar en los datos adquiridos de su aplicación y planteados en este documento pese a que la aplicación del modelo 3 (Naive) supondría una mejora en cuanto a recursos consumidos (Time taken to build model: 0.01 second) vemos como los mejores porcentajes en clasificación (97.8667 %) corresponden a este modelo 2.

Los valores de detalle de la Accuracy también nos indican que la correlación es alta:

=== Detailed Accuracy By Class ===

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

