

1. Título: Datos de segmentación de imágenes
2. Información relevante: Las instancias se extrajeron aleatoriamente de una base de datos de 7 imágenes exteriores.
Las imágenes se segmentaron manualmente para crear uno de clasificación para cada píxel. Cada instancia es una región 3x3.
3. Número de instancias:
 - a. Datos de entrenamiento: 210
 - b. Datos de prueba: 2100
4. Número de atributos: 19 atributos continuos
5. Valores de atributos que faltan: ninguna
6. Distribución de clases:
 - a. Clases: ladrillo, cielo, follaje, cemento, ventana, camino, hierba.
 - b. 30 instancias por clase para datos de formación.
 - c. 300 instancias por clase para datos de prueba.

1º prueba → Classifier → Trees → model J48

- Training set

- Número de capes(hojas): 34
- Tamaño del árbol: 67

```
Classifier output

=== Summary ===

Correctly Classified Instances      1485           99      %
Incorrectly Classified Instances    15            1      %
Kappa statistic                    0.9883
Mean absolute error                 0.0048
Root mean squared error             0.0488
Relative absolute error             1.9473 %
Root relative squared error         13.9545 %
Total Number of Instances          1500
```

Parece que el modelo nos muestra que ha clasificado de manera correcta un total de 1485 instancias con un 99% en la fase de training set y de manera incorrecta ha clasificado 15 instancias representado el 1%.

Aunque los datos son bastante buenos, quizás es demasiado bueno, sospecho que esto es debido a que no están balanceadas las muestras.

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	AUC Area	Cla
1,000	0.003	0.981	1,000	0.990	0.990	0.989	0.999	0.990	bri1
1,000	0.000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	sky
0.986	0.005	0.972	0.986	0.979	0.979	0.975	0.999	0.992	fol
0.986	0.001	0.995	0.986	0.991	0.989	0.989	1,000	0.999	cem
0.956	0.003	0.980	0.956	0.968	0.968	0.963	0.997	0.986	wln
1,000	0.000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	pat
1,000	0.000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	gra
Weighted Avg.	0.990	0.002	0.990	0.990	0.990	0.988	0.999	0.996	

```

=== Confusion Matrix ===

```

a	b	c	d	e	f	g	<-- classified as
205	0	0	0	0	0	0	a = brickface
0	220	0	0	0	0	0	b = sky
1	0	205	0	2	0	0	c = foliage
1	0	0	217	2	0	0	d = cement
2	0	6	1	195	0	0	e = window
0	0	0	0	0	236	0	f = path

De manera más visual, weka nos indica cuál es la matriz de correlación que ha obtenido y nos desengrana las instancias clasificadas de manera casi perfecta.

Se observa que aquellas instancias que no se han clasificado de manera correcta se encuentran en la columna “a” y con un total de 4 instancias mal clasificadas, en la columna “c” con un total de 6 instancias erróneas, en la columna “d” con un total de 1 instancia mal clasificada y en la columna “e” un total de 4

2º prueba → Classifier → Trees → **model J48**

- **Cross-validation** → Folds: 10
- Número de capes(hojas): 34
- Tamaño del árbol: 67

En esta ocasión la validación cruzada 10 veces proporciona una precisión media del clasificador y en esta ocasión los resultados parecen que han empobrecido, con una tasa de clasificación correcta de un 96% aproximadamente frente a un 4.3% mal clasificadas.

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      1436          95.7333 %
Incorrectly Classified Instances     84          4.2667 %
Kappa statistic                    0.9502
Mean absolute error                 0.0138
Root mean squared error             0.1057
Relative absolute error              5.6471 %
Root relative squared error         30.2115 %
Total Number of Instances          1500

=== Detailed Accuracy By Class ===

```

	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	

```

=== Confusion Matrix ===
 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

```

3a prueba → Classifier → Trees → **Randomforest**

- **Cross-validation** → Folds: 10

```

=== 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              5.1639 %
Root relative squared error         22.4623 %
Total Number of Instances          1500

=== 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	0,968	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	

```

=== 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  5 185  0  0  0 | e = window
  0  0  0  1  0 235  0 | f = path
  0  0  0  0  0  0 207 | g = grass

```

Los resultados obtenidos indican que el modelo random forest es capaz con un 98% aproximadamente de clasificar bien 1468 instancias frente al 2% que clasifica mal (32 instancias)

Comparado con el modelo j48 en modo cross-validation, éste modelo, es mucho más preciso a la hora de hacer la clasificación. Vamos mejorando

4º prueba → Classifier → functions → **Logistic**

- **Cross-validation** → Folds: 10

```
=== Summary ===

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
```

El modelo logístico revela un porcentaje de resultados similar a la que expresa el modelo de árbol j48, ambos utilizando la misma opción de cross-validation, por lo que descarto la función logística.

5º prueba → Classifier → functions → **multilayer Perceptron**

- **Cross-validation** → Folds: 10

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

Correctly Classified Instances      1456           97.0667 %
Incorrectly Classified Instances     44           2.9333 %
Kappa statistic                    0.9658
Mean absolute error                 0.0135
Root mean squared error             0.0841
Relative absolute error             5.5049 %
Root relative squared error         24.04 %
Total Number of Instances          1500

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,980	0,002	0,990	0,980	0,985	0,983	0,999	0,996	brickface	
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	sky	
0,928	0,010	0,937	0,928	0,932	0,922	0,989	0,959	foliage	
0,977	0,011	0,939	0,977	0,958	0,951	0,994	0,980	cement	
0,902	0,011	0,929	0,902	0,915	0,902	0,991	0,955	window	
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	path	
1,000	0,001	0,995	1,000	0,998	0,997	1,000	1,000	grass	
Weighted Avg.	0,971	0,005	0,971	0,971	0,971	0,966	0,996	0,985	

```

=== Confusion Matrix ===
 a  b  c  d  e  f  g  <-- classified as
201  0  0  3  1  0  0 | a = brickface
  0 220  0  0  0  0  0 | b = sky
  0  0 193  6  9  0  0 | c = foliage
  0  0  1 215  4  0  0 | d = cement
  2  0 12  5 184  0  1 | e = window
  0  0  0  0  0 236  0 | f = path
  0  0  0  0  0  0 207 | g = grass

```

he observado que el modelo multilayer Perceptron devuelve unos porcentajes de clasificación bastante buenos, en cambio, si nos fijamos en la matriz de correlación las instancias mal clasificadas son más abundantes en columnas en concreto, lo cual no es un buen modelo para llevar a cabo el análisis de los datos, ya que por ejemplo los otros modelos que he visto anteriormente, a pesar de contar con un mayor grado de porcentaje de error, comparado con el MP, la cantidad de error acumulados por columna es menor.

Conclusión: El modelo que mejor se ajusta a la hora de clasificar las instancias correctamente, parece que es el Random Forest con la opción de cross-validation. Eso es posible ya que las muestras no están balanceadas. Gracias a que el modo de cross-validation añade aleatoriedad en la comparación de las muestras permite hacer una predicción más ajustada.

