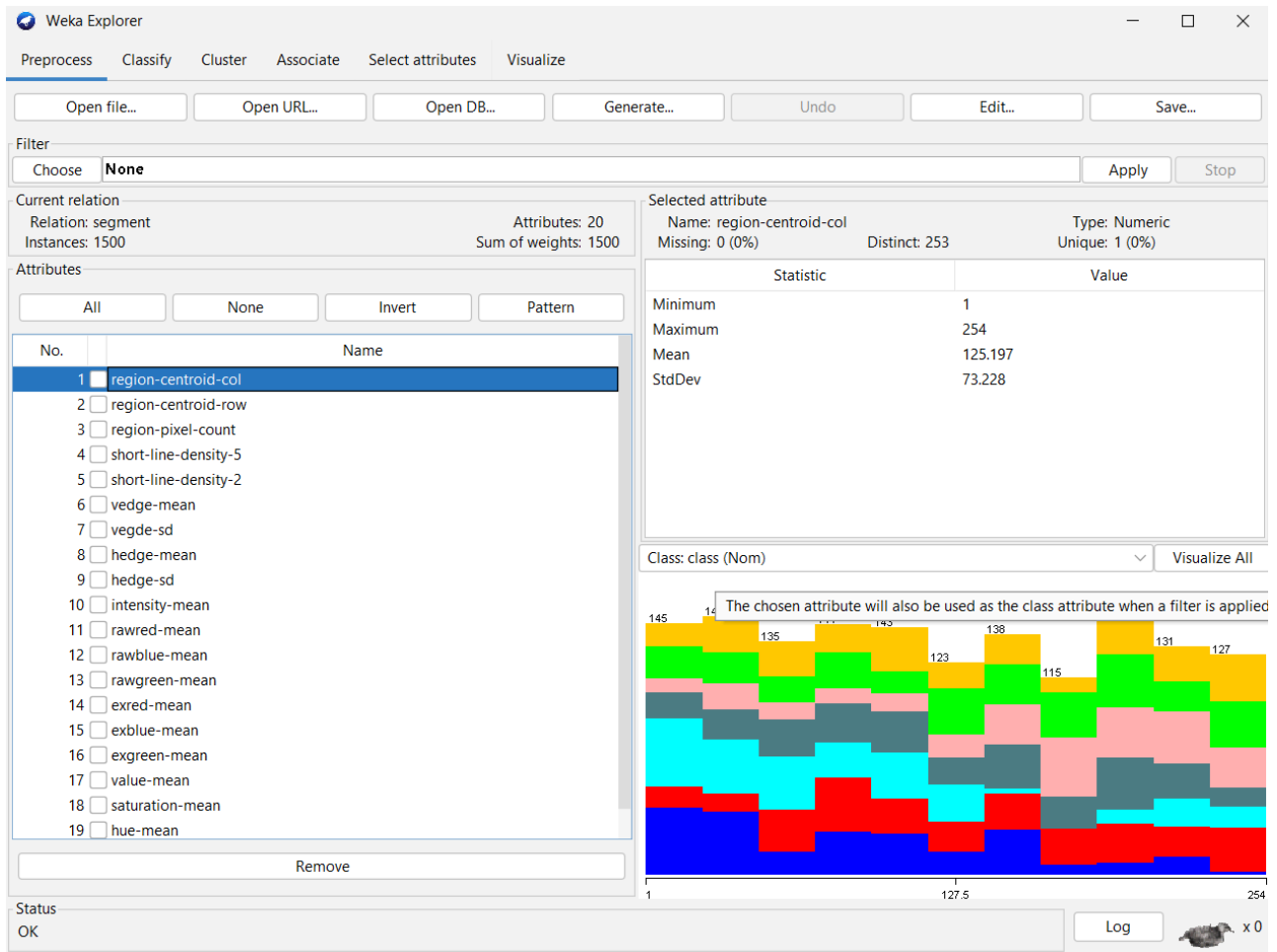
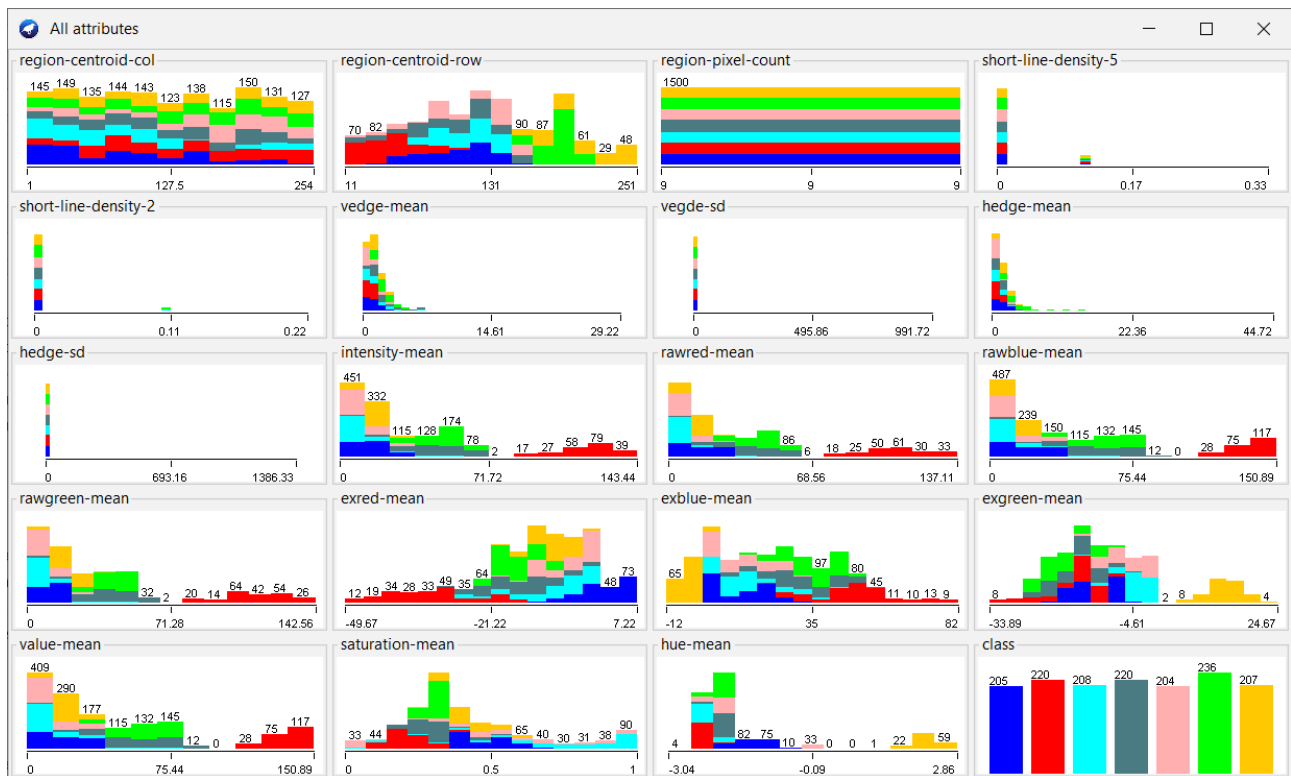


Informe WEKA

Ens presenten un seguit de dades i ens demanen que les classifiquem.
La primera visió que en tenim és la següent:



Si cliquem al botó **Visualize** visualitzarem de les dades generals com es mostren a continuació:



Les dades: region-pixel-count realment no ens aporten informació més que per indicar el nombre de pixels que hi ha bé ens els mapes amb els van extreure les dades, bé per una altra cosa.

Com que weka pot ponderar les dades que realment són importants, aquí no l'elimino ja que el propi weka dirà que no és rellevant.

Per a classificar hem d'escollir tres models i dir quin és el millor:

Utilitzarem per al model 1 MultiLayerPerceptron, per al model 2 Tree_J48 i per al model 3 NaiveBayes

Model 1 (MultiLayerPerceptron)

Time taken to build model: 2.63 seconds

=== 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	0,980	0,002	0,990	0,980	0,985	0,983
PRC Area	0,999	0,996				
	1,000	0,000	1,000	1,000	1,000	1,000
	1,000	1,000				

		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

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

Model 2 (Trees J48)

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

=== Detailed Accuracy By Class ===

=== 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

```

Model 3 (NaivesBayes)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

```

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

```

=== Detailed Accuracy By Class ===

		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
ROC Area	PRC Area	Class					
		0,946	0,048	0,758	0,946	0,842	0,820
0,987	0,917	brickface					
		0,991	0,000	1,000	0,991	0,995	0,995
0,999	0,998	sky					
		0,188	0,011	0,736	0,188	0,299	0,331
0,945	0,658	foliage					
		0,864	0,027	0,844	0,864	0,854	0,829
0,964	0,886	cement					
		0,721	0,130	0,465	0,721	0,565	0,496
0,893	0,547	window					
		0,945	0,003	0,982	0,945	0,963	0,957
0,999	0,994	path					
		0,990	0,000	1,000	0,990	0,995	0,994
0,999	0,997	grass					
Weighted Avg.		0,811	0,030	0,832	0,811	0,793	0,781
0,970	0,861						

=== Confusion Matrix ===

```

a   b   c   d   e   f   g   <-- classified as
194  0   0   6   5   0   0 | a = brickface
  0 218  0   2   0   0   0 | b = sky
  6   0  39  6 157  0   0 | c = foliage
 17   0   4 190  5   4   0 | d = cement
 39   0   6 12 147  0   0 | e = window
  0   0   4   9   0 223  0 | f = path
  0   0   0   0   2   0 205 | g = grass

```

Quan li diem quines dades són més importants per a la classificació ens dona la següent informació:

```
=== Attribute Selection on all input data ===
```

```
Search Method:
```

```
Attribute ranking.
```

```
Attribute Evaluator (supervised, Class (nominal): 20 class):
```

```
Correlation Ranking Filter
```

```
Ranked attributes:
```

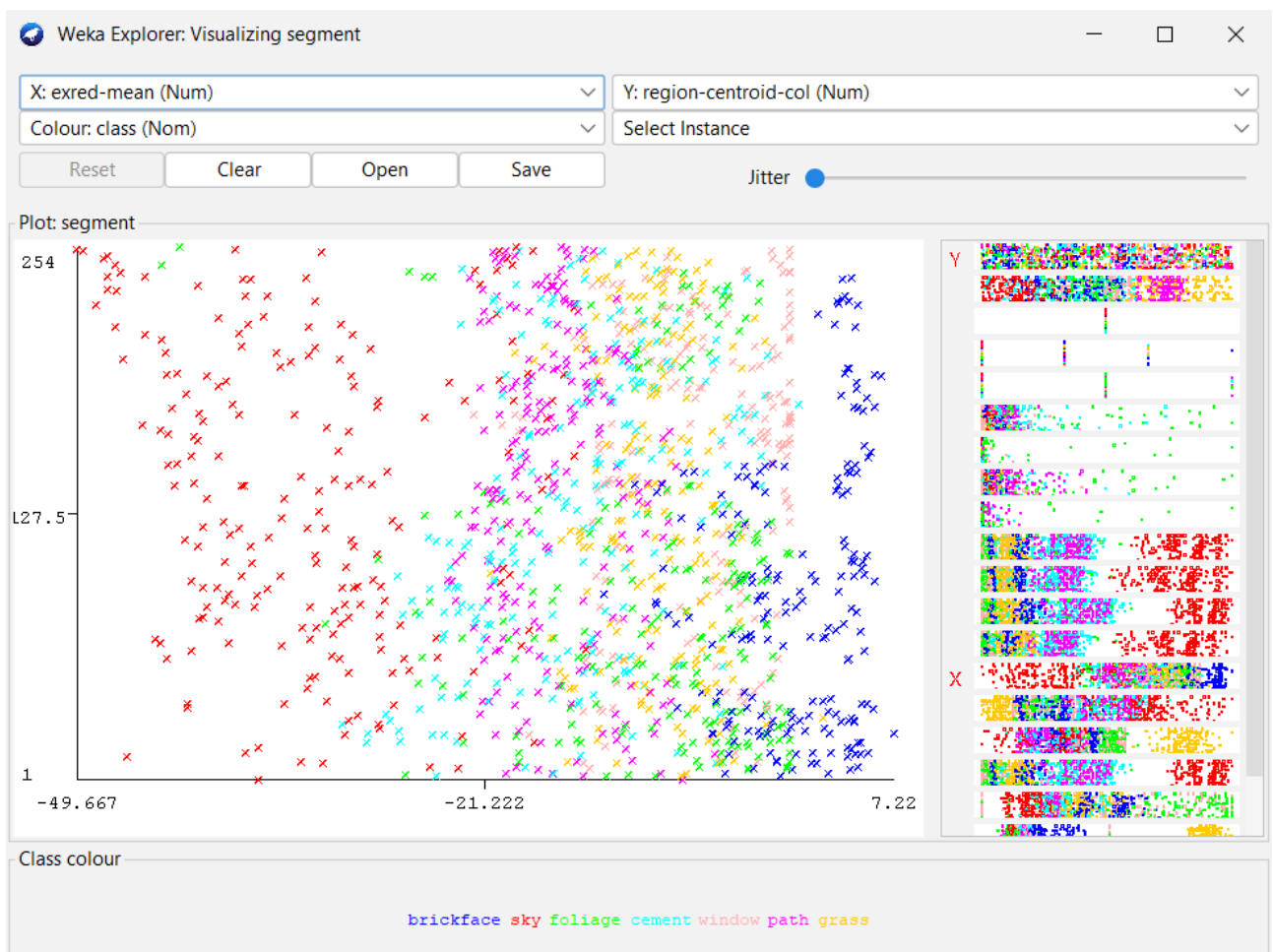
```
0.3156 15 exblue-mean
0.3155 12 rawblue-mean
0.3121 17 value-mean
0.3056 10 intensity-mean
0.3033 11 rawred-mean
0.3006 16 exgreen-mean
0.3005 2 region-centroid-row
0.2926 13 rawgreen-mean
0.2706 19 hue-mean
0.2688 14 exred-mean
0.2406 18 saturation-mean
0.1265 6 vedge-mean
0.1244 8 hedge-mean
0.107 1 region-centroid-col
0.0718 5 short-line-density-2
0.0622 7 vegde-sd
0.0613 4 short-line-density-5
0.0596 9 hedge-sd
0 3 region-pixel-count
```

```
Selected attributes: 15,12,17,10,11,16,2,13,19,14,18,6,8,1,5,7,4,9,3 : 19
```



Un exemple de la visualització de les dades. La majoria de dades eren pràcticament linealment dependents mentre que les region-centroid- respecte les intensity-mean rawwd-mean exred exblu exgreen mostra més dispersió.

Si ampliem un dels quadrats mostrats veurem la següent finestra:



Dels tres models podem resumir-ho en la següent taula:

Nom	Instàncies correctament classificades	Error mitjà absolut	Error relatiu absolut
Model 1	97.0667 %	0.0135	5.5049 %
Model 2	95.7333 %	0.0138	5.6471 %
Model 3	81.0667 %	0.0554	22.6144 %

Veiem com el model 1 té un percentatge d'instàncies correctament classificades i un error menor si ho comparem amb el model 3. En tots tres models s'ha optat per deixar les característiques (els Settings) de fàbrica. És a dir, no optimitzar cap característica. Així doncs podem dir que el model 1 (MultiLayerPerceptron) és molt més fiable que els altres dos.