

ACTIVITAT 1

Analitza el conjunt de dades segment-challenge.arff amb Weka, i després de provar 3 models troba el millor per poder fer la predicció.

Genera un informe.docx mostrant comentant els resultats dels 3 models i justificant quin et sembla millor. Ha d'incloure algun gràfic que et sembli significatiu.

Els 3 models predictius seleccionats per fer la comparació han sigut el Naive Bayes, Random Tree i Multilayer Perceptron.

- Naive Bayes resultats

Test mode: 100-fold cross-validation

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

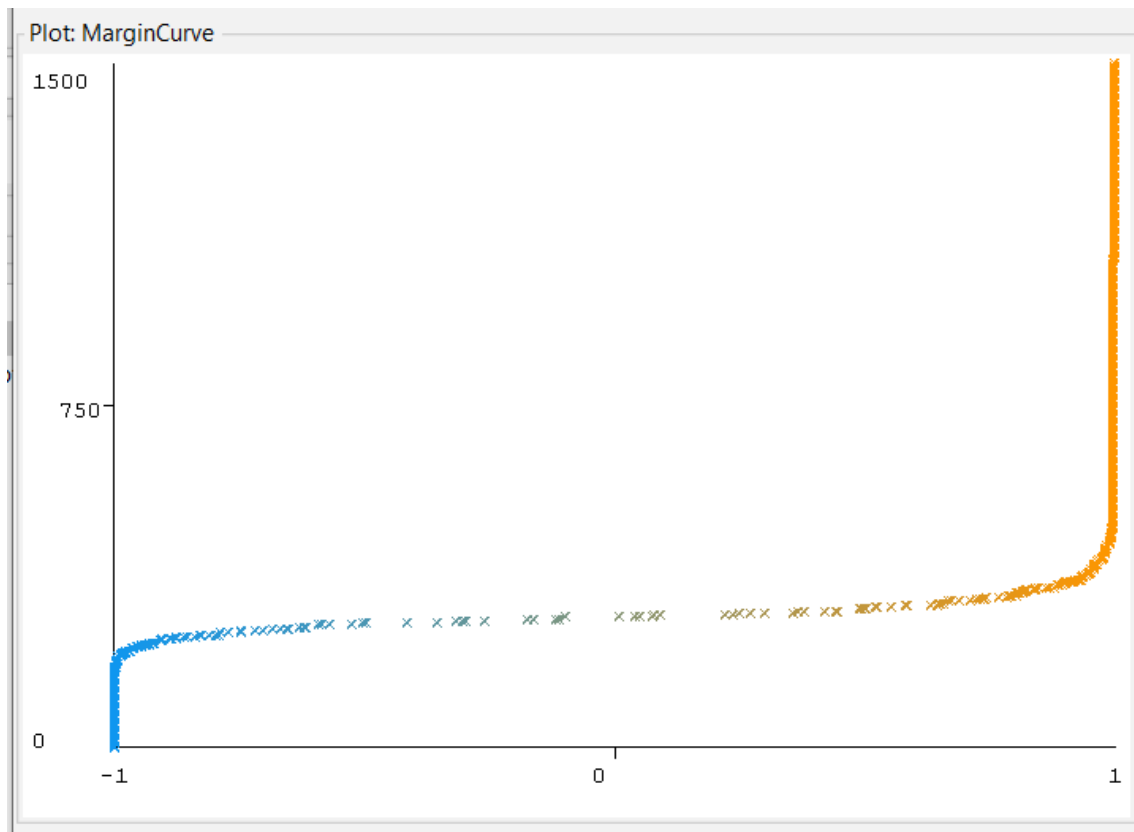
Correctly Classified Instances      1218           81.2 %
Incorrectly Classified Instances    282           18.8 %
Kappa statistic                    0.7807
Mean absolute error                 0.0554
Root mean squared error            0.2261
Relative absolute error            22.6111 %
Root relative squared error        64.6273 %
Total Number of Instances         1500

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,961	0,046	0,767	0,961	0,853	0,834	0,987	0,919	brickface
	0,991	0,000	1,000	0,991	0,995	0,995	0,999	0,998	sky
	0,183	0,012	0,704	0,183	0,290	0,316	0,946	0,662	foliage
	0,864	0,028	0,841	0,864	0,852	0,826	0,964	0,886	cement
	0,721	0,129	0,468	0,721	0,568	0,499	0,893	0,551	window
	0,945	0,002	0,987	0,945	0,965	0,959	0,999	0,995	path
	0,990	0,000	1,000	0,990	0,995	0,994	0,999	0,997	grass
Weighted Avg.	0,812	0,030	0,829	0,812	0,794	0,781	0,971	0,863	

```
=== Confusion Matrix ===

 a  b  c  d  e  f  g  <-- classified as
197  0  0  6  2  0  0 | a = brickface
 0 218  0  2  0  0  0 | b = sky
 5  0 38  7 158  0  0 | c = foliage
17  0  5 190  5  3  0 | d = cement
38  0  6  13 147  0  0 | e = window
 0  0  5  8  0 223  0 | f = path
 0  0  0  0  2  0 205 | g = grass
```



D'aquest model podem destacar que encara que el cost computacional ha sigut mínim, fixar-se quasi instantani el marge d'error és d'un 18.8% sent aquest molt elevat si el comparem amb els altres dos models que segueixen a continuació.

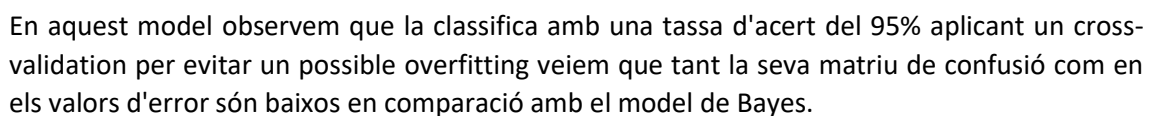
- Random Tree resultats

Test mode: 100-fold cross-validation

Correctly Classified Instances	1425	95	%
Incorrectly Classified Instances	75	5	%
Kappa statistic	0.9416		
Mean absolute error	0.0143		
Root mean squared error	0.1195		
Relative absolute error	5.8356	%	
Root relative squared error	34.1624	%	
Total Number of Instances	1500		

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,980	0,003	0,980	0,980	0,980	0,977	0,989	0,964	brickface
	1,000	0,001	0,995	1,000	0,998	0,997	1,000	0,995	sky
	0,880	0,015	0,901	0,880	0,891	0,873	0,932	0,810	foliage
	0,941	0,015	0,916	0,941	0,928	0,916	0,963	0,870	cement
	0,858	0,021	0,866	0,858	0,862	0,840	0,919	0,763	window
	0,992	0,003	0,983	0,992	0,987	0,985	0,994	0,976	path
	0,990	0,000	1,000	0,990	0,995	0,994	0,995	0,992	grass
Weighted Avg.	0,950	0,008	0,950	0,950	0,950	0,942	0,971	0,912	

	a	b	c	d	e	f	g	<-- classified as
201	0	0	2	2	0	0		a = brickface
	0	220	0	0	0	0		b = sky
	0	1	183	4	19	1		c = foliage
	2	0	3	207	6	2		d = cement
	1	0	16	11	175	1		e = window
	0	0	1	1	0	234		f = path
	1	0	0	1	0	0		g = grass



- Multilayer Perceptron resultsats

Test mode: 100-fold cross-validation

```

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

Correctly Classified Instances      1446           96.4   %
Incorrectly Classified Instances     54           3.6   %
Kappa statistic                     0.958
Mean absolute error                 0.0151
Root mean squared error             0.0937
Relative absolute error              6.1716 %
Root relative squared error         26.7766 %
Total Number of Instances          1500

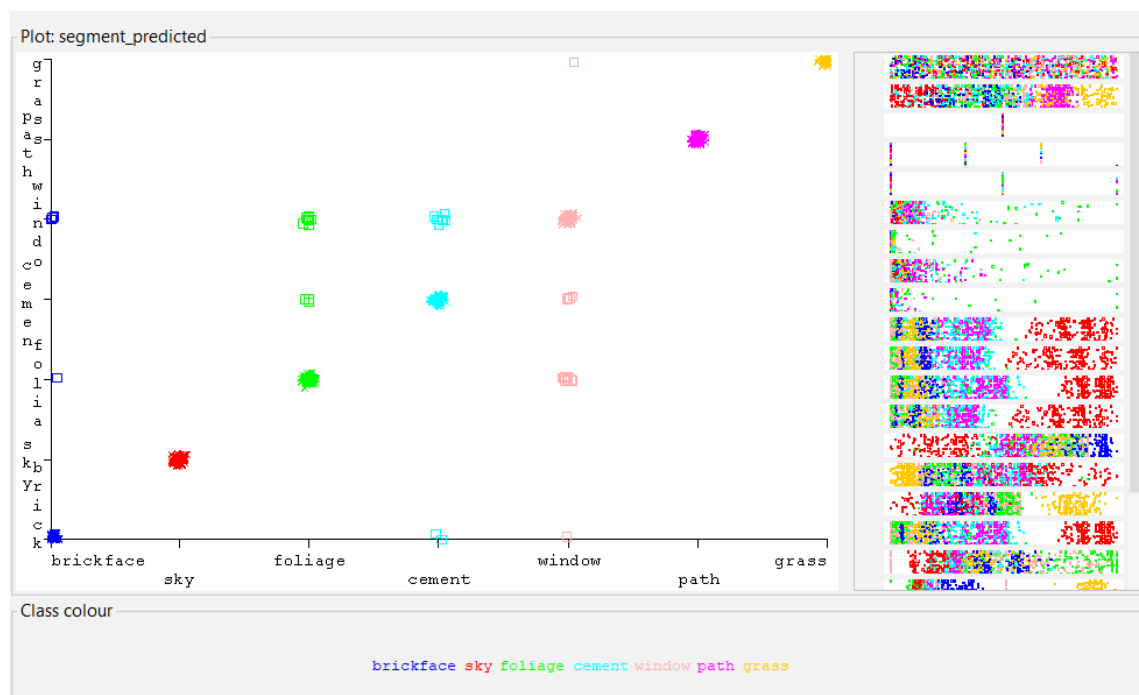
=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
0,980    0,002    0,985    0,980    0,983    0,980    0,999    0,995    brickface
1,000    0,000    1,000    1,000    1,000    1,000    1,000    1,000    sky
0,918    0,011    0,932    0,918    0,925    0,913    0,991    0,962    foliage
0,950    0,009    0,950    0,950    0,950    0,941    0,993    0,982    cement
0,892    0,019    0,879    0,892    0,886    0,868    0,983    0,929    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.   0,964    0,006    0,964    0,964    0,964    0,958    0,995    0,982

=== Confusion Matrix ===

  a  b  c  d  e  f  g  <-- classified as
201  0  1  0  3  0  0 | a = brickface
  0 220  0  0  0  0  0 | b = sky
  0  0 191  4 13  0  0 | c = foliage
  2  0  0 209  9  0  0 | d = cement
  1  0 13  7 182  0  1 | e = window
  0  0  0  0  0 236  0 | f = path
  0  0  0  0  0  0 207 | g = grass

```



Com es pot observar en aquest model s'ha millorat els resultats anteriors amb un acert del 96.4% , encara que el cost computacional en realitzar una cross-validation ha sigut molt més lenta que la dels altres dos models aplicats en aquesta prova. També s'ha millorat el valor de la root mean square error, el que indica que la dispersió dels errors és menor en comparació amb els anteriors.

Conclusions:

Dels models seleccionats si es vol decidir pel que sigui més precís escolliria el Multilayer Perceptron encara que aquesta elecció té una contra que és el cost computacional, ja que per assegurar-se d'evitar un overfitting del model tarda més temps.

En canvi, si es volgués tindre un model que tingués una ràtio d'acert elevat, però més àgil segons la prova realitzada optaria pel de Random Tree en aquest cas d'estudi.