**Curso: 23/CIFOFSE/670/0182279/006**

**Evaluación final**

**Agustín Paternoster**

Para el análisis del dataset segment-challenge.arff propuesto en la evaluación final he elegido 3 modelos de análisis de clasificación:

1. Modelo de regresión logística
2. Modelo de regresión de Red neuronal
3. Modelo de random Forest.

El objetivo es entrenar a un modelo que realice una clasificación de una imagen exterior de acuerdo con 7 categorías (pared de ladrillo, cielo, vegetación, cemento, ventana, camino, césped) a partir de 19 atributos continuos.

**Modelo de regresión logística**

Resultados

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

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0,980 0,005 0,971 0,980 0,976 0,972 0,997 0,974 brickface

1,000 0,000 1,000 1,000 1,000 1,000 1,000 1,000 sky

0,923 0,014 0,914 0,923 0,919 0,906 0,994 0,953 foliage

0,918 0,005 0,967 0,918 0,942 0,932 0,985 0,962 cement

0,897 0,022 0,867 0,897 0,882 0,863 0,987 0,921 window

1,000 0,000 1,000 1,000 1,000 1,000 1,000 1,000 path

1,000 0,000 1,000 1,000 1,000 1,000 1,000 1,000 grass

Weighted Avg. 0,961 0,006 0,961 0,961 0,961 0,955 0,995 0,974

=== Confusion Matrix ===

a b c d e f g <-- classified as

201 0 0 2 2 0 0 | a = brickface

0 220 0 0 0 0 0 | b = sky

0 0 192 1 15 0 0 | c = foliage

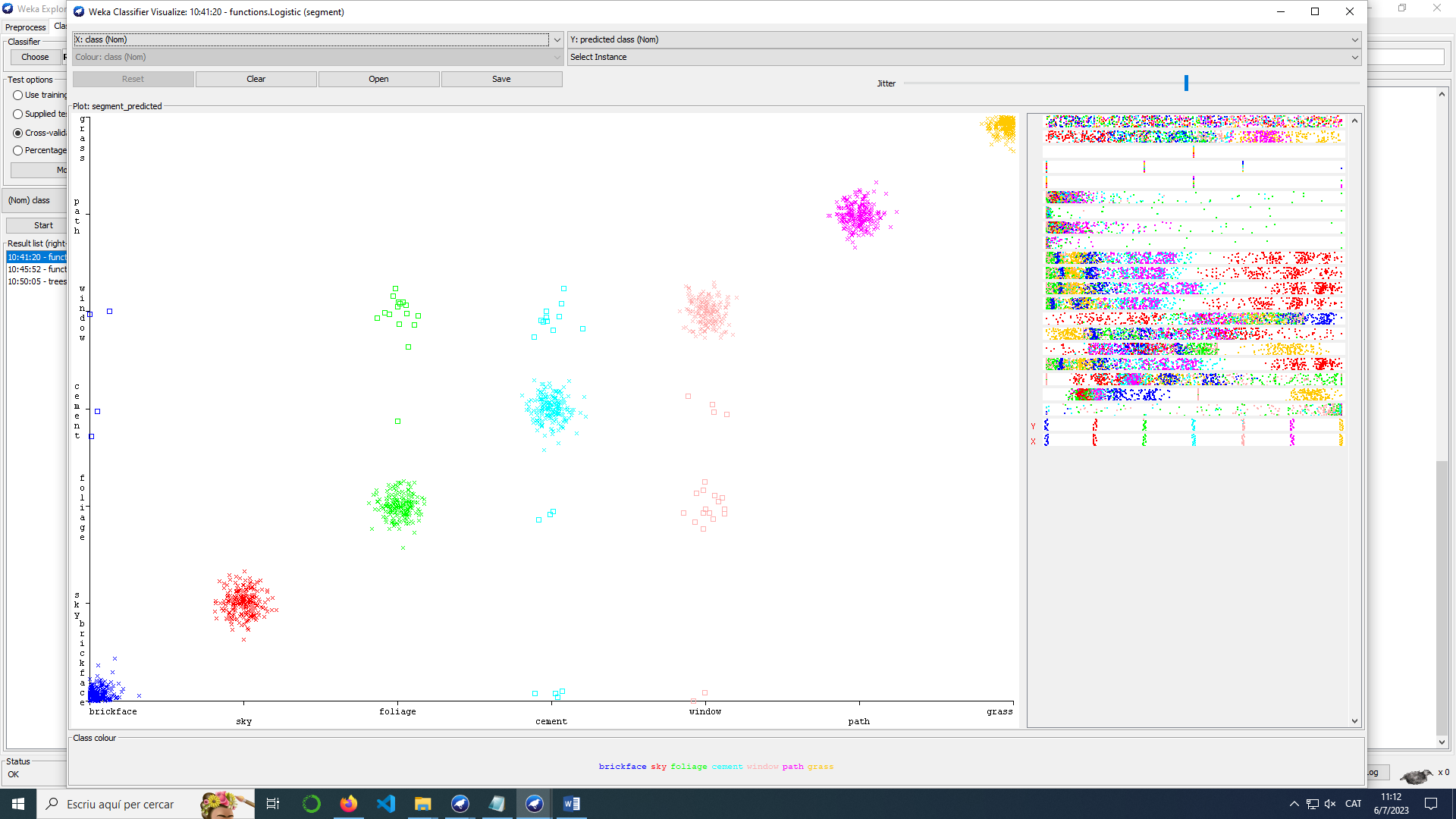
4 0 3 202 11 0 0 | d = cement

2 0 15 4 183 0 0 | e = window

0 0 0 0 0 236 0 | f = path

0 0 0 0 0 0 207 | g = grass

Grafico Clasificacion errores



**Modelo de regresión de Red neuronal**

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

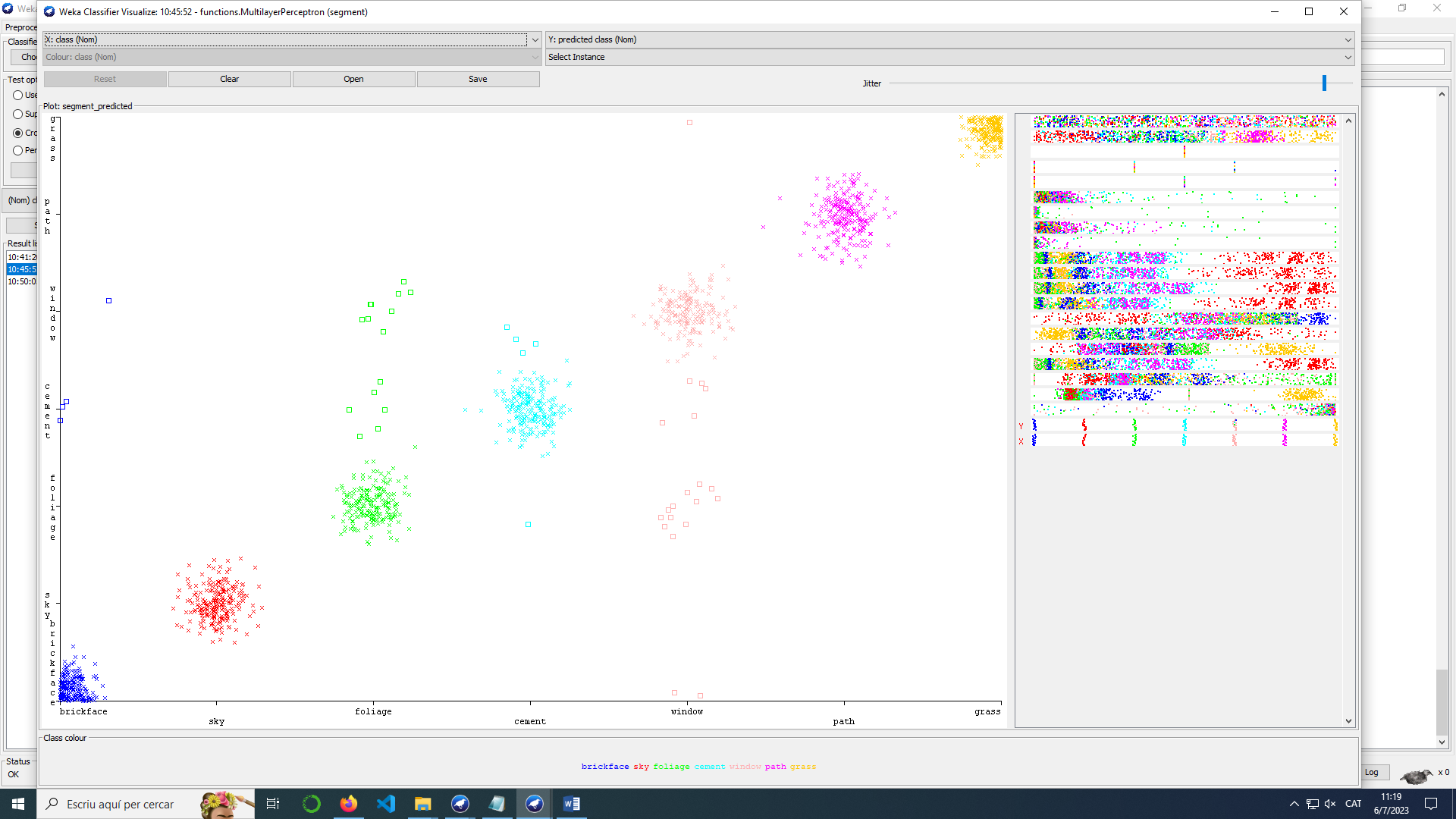
0 0 1 215 4 0 0 | d = cement

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

Grafico Clasificacion errores



**Modelo de random Forest.**

Resultados

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 ===

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

=== 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

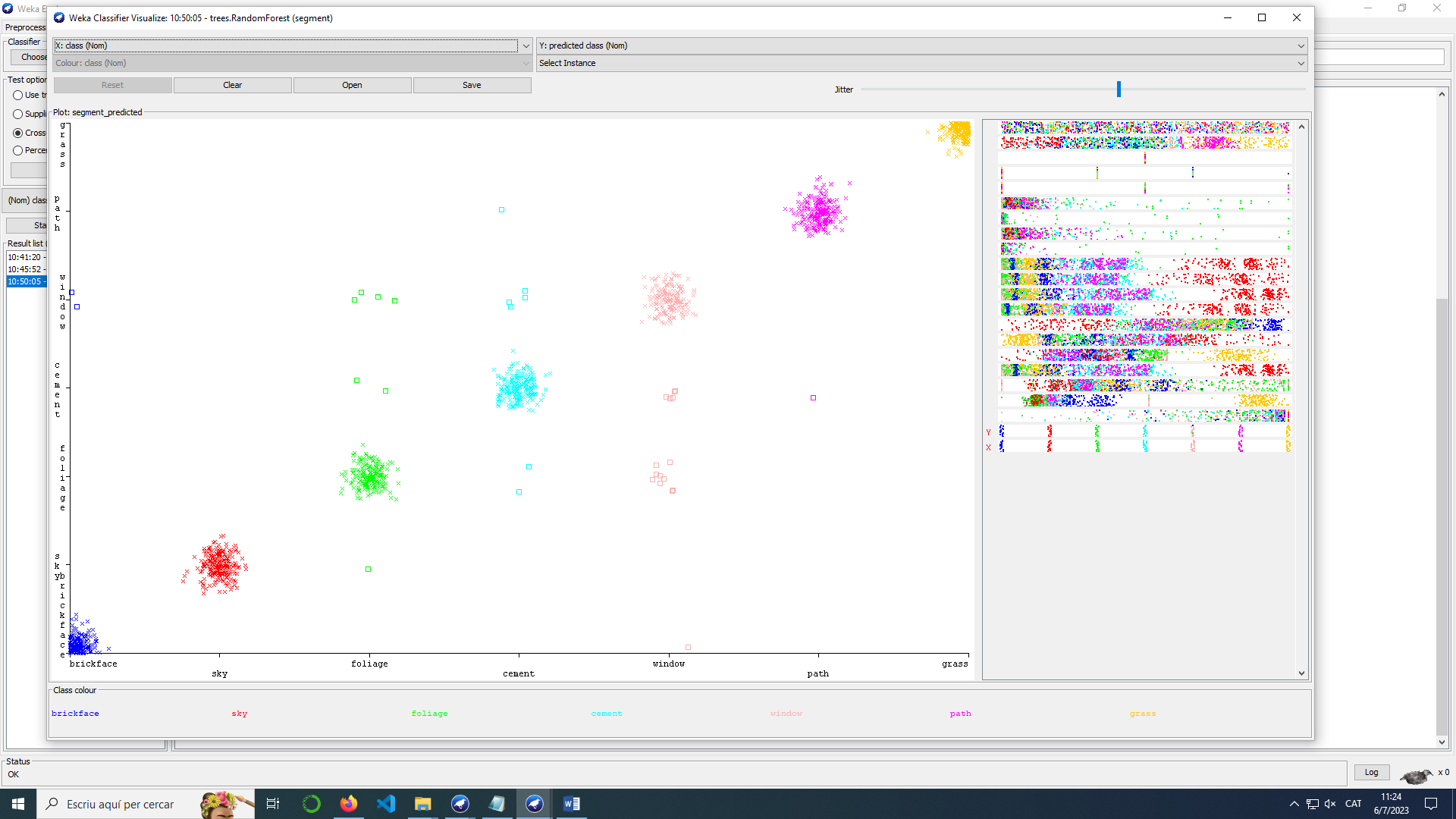
0 0 2 213 4 1 0 | d = cement

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

Grafico Clasificación errores



Conclusiones

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Clasificaciones Incorrectas** | **TR Rate** | **FP Rate** | **Precision** |
| **regresión logística** | 59 | 0,961 | 0,006 | 0,961 |
| **regresión de Red neuronal** | 44 | 0,971 | 0,005 | 0,971 |
| **random Forest** | 32 | 0,979 | 0,003 | 0,979 |

Al analizar la información comprobamos el modelo de random forest es el que menos resultados incorrectos da y por tanto seria el modelo a escoger. Al analizar las gráficas vemos que todos los modelos presentan dificultad con las categorías de ventanas, cemento y vegetación, aunque de los tres el randon forest tiene la tasa de aciertos más elevada.