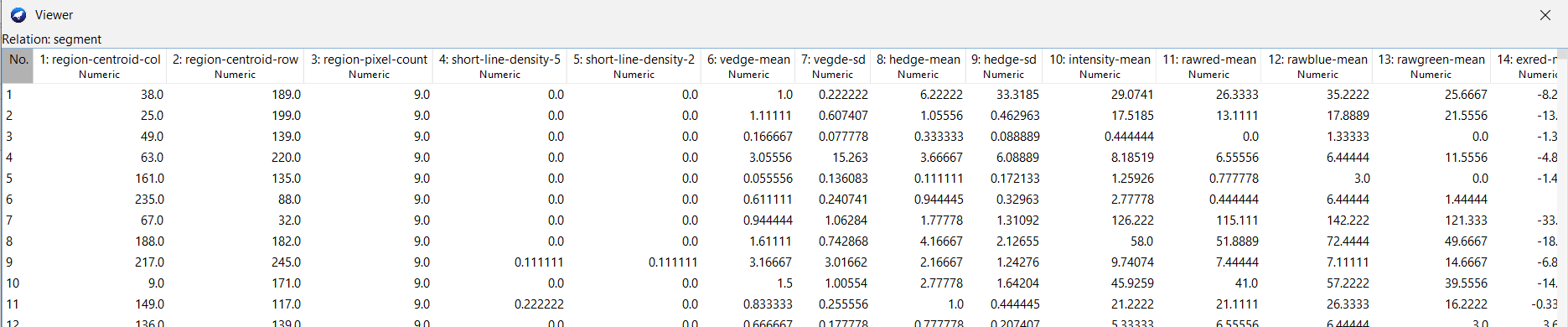
**segment-challenge**

Al observar los datos podemos ver que todos son numéricos.



Sería conveniente tenerlos como nominales

En los variables podemos observar

@atributo región-centroide-col numérico

@atributo región-centroide-fila numérico

@atributo región-píxel-recuento numérico

@attribute short-line-density-5 numérico

@attribute short-line-density-2 numérico

@attribute valor medio numérico

@attribute vegde-sd numérico

@attribute hedge-mean numérico

@attribute hedge-sd numérico

@attribute intensidad-media numérica

@attribute rawred-mean numérico

@attribute rawblue-mean numérico

@attribute rawgreen-mean numérico

@attribute exred-media numérica

@attribute exblue-mean numérico

@attribute exgreen-mean numérico

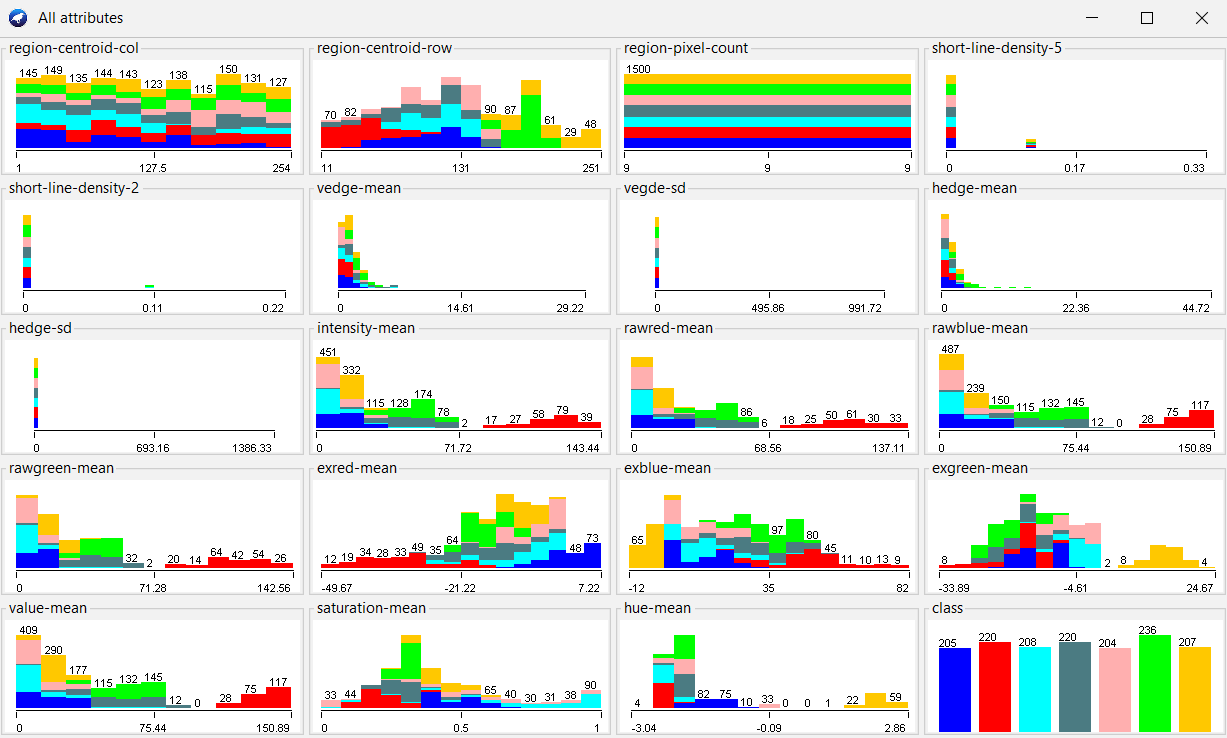
@attribute valor medio numérico

@attribute saturación media numérica

@attribute tonalidad media numérica

@attribute class {ladrillo,cielo,follaje,cemento,ventana,camino,hierba}

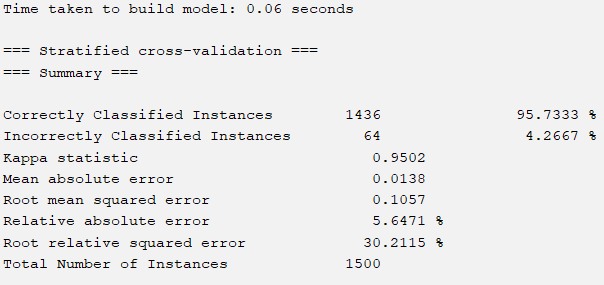
Graficas de ayuda para la clasificación



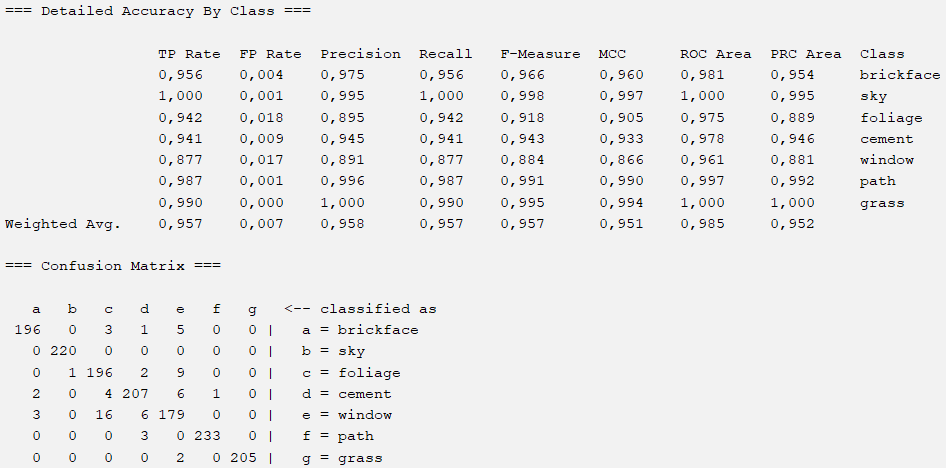
**Modelo 1: J48**

En este modelo podemos observar que tardo 0.06 segundo en clasificar.

Podemos observar que el modelo es bueno con un 95,73% instancias correctas, que clasifico como buenas 1436.



Podemos observar las predicciones son muy buenas tienen tendencia a ser 1, podemos colaborar con la matriz de confusión que la diagonal está bien marcada, también su distribución a ambos lados de la diagonal son 27 que son las 64 que están menciona como clasificaciones incorrectas.

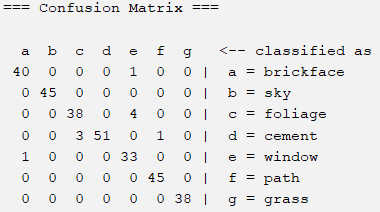


Mejora del modelo **J48**

Si trabajamos con un porcentaje del 80% el modelo mejora mucho mas

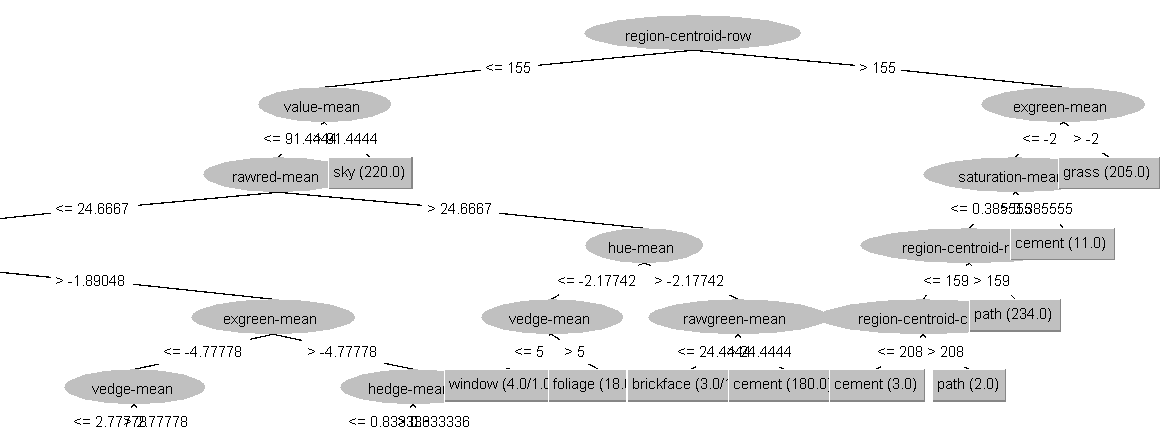


En la matriz de confusión, solo quedan 10 incorrectas



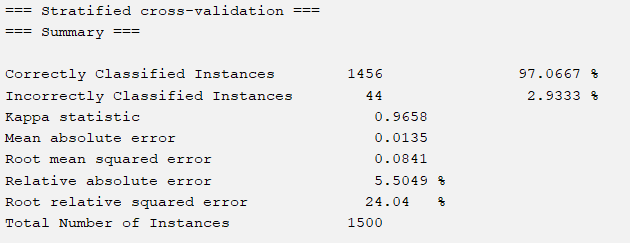
En el árbol de decisiones, las variables que dan más peso al modelo

Podemos decir si tenemos un valor en la región centroide col (region-centroid-col) y es menor o igual 155, se considera como valor medio (value\_mean) y si es mayor a 155 se considera que esta en el medio exverde (exgreen-mean),

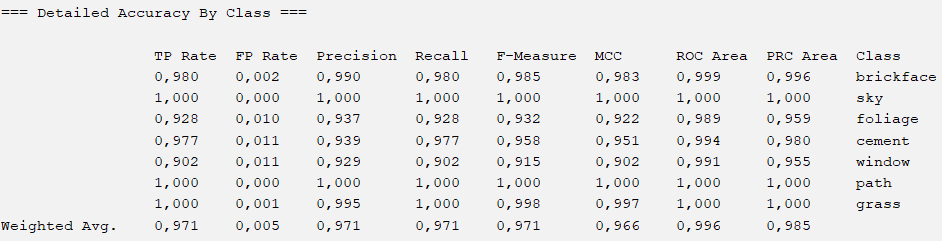


**Modelo 2: MultilayerPerceptron**

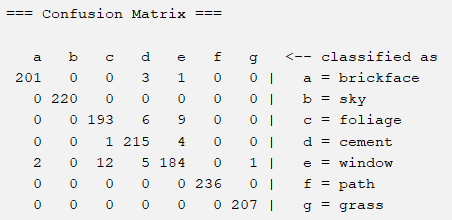
Este modelo a tardado 2.63 segundos en clasificar, es más lento al anterior (J48), las observaciones de este modelo son mejores a la anterior con un 97,067% instancias correctas, clasifico como buenas 1436.

****

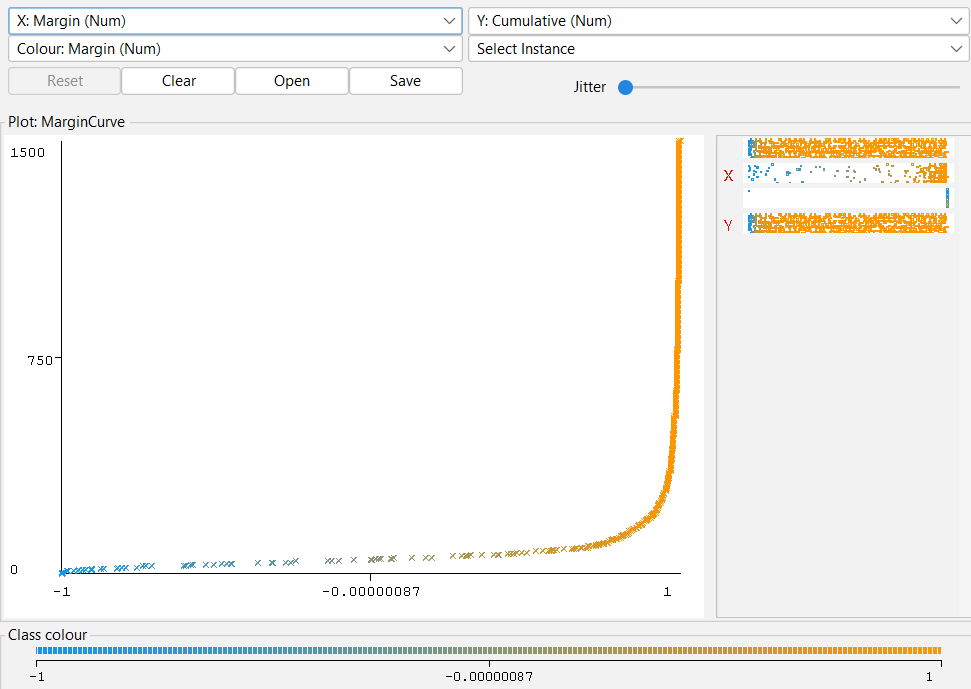
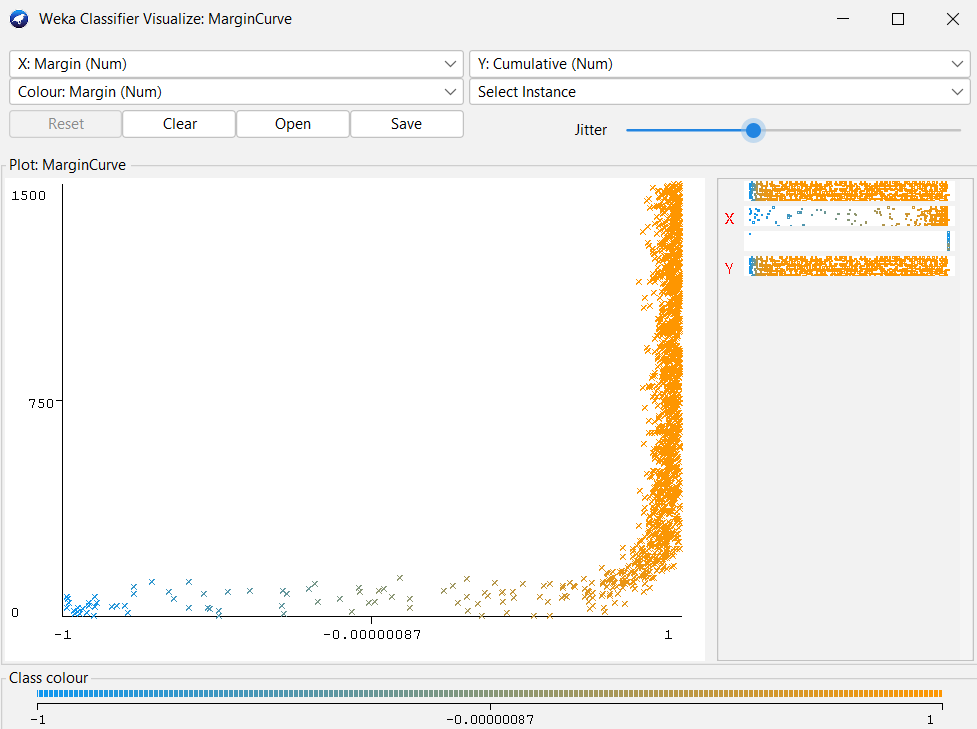
En el Recall podemos observar la cantidad de clases que podemos identificar y son muy buenas, si vemos la media de F1(F-measure) no hay discrepancias notorias.



La matriz de confusión, vemos que la diagonal está bien marcada mucho mejor que la J48, En la matriz de confusión, solo quedan 44 incorrectas



En la curva de márgenes, podemos visualizar que afecta poco el ruido aumentar, al aumentar los valores al modelo.

>>> 

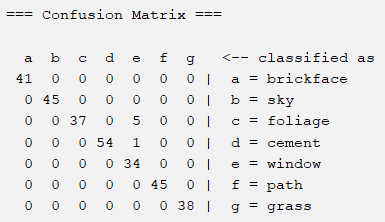
Mejora del modelo **MultilayerPerceptron**

Si trabajamos con un porcentaje del 80% el modelo mejora mucho mas

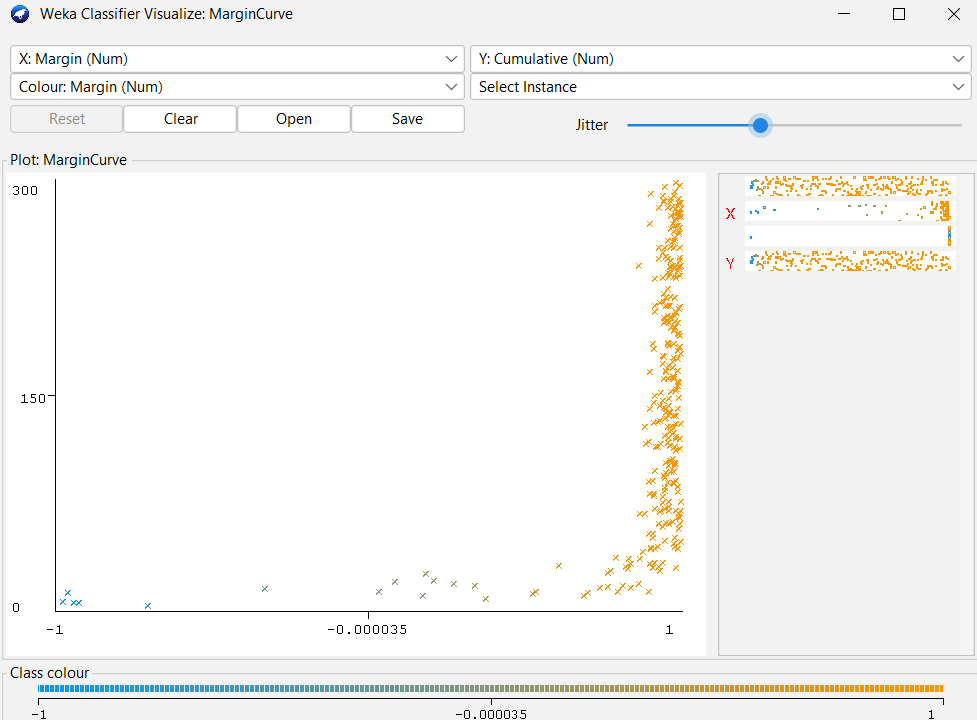


Ahora ya estamos trabajando con una precisión del 98%, obteniendo 294 clasificados como buenos.

En la matriz de confusión, solo quedan 6 incorrectas

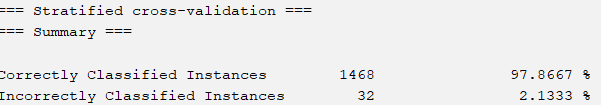


Pero aun al aumentar valores al modelo afecta el ruido.

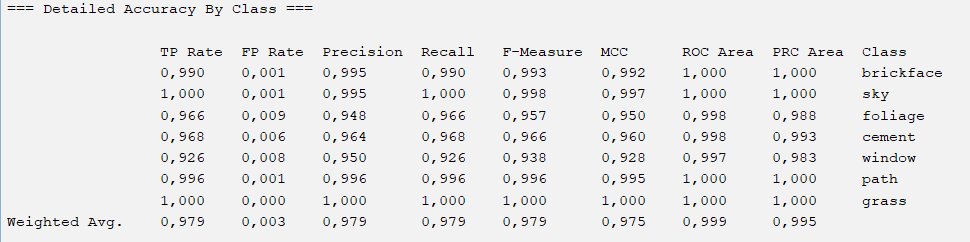


**Modelo 3: RandomForest**

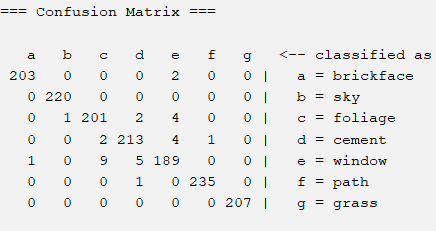
El modelo es más rápido que el modelo MultilayerPerceptron porque tardo 0.29 segundos, pero no más rápido que J48, en cuanto a las predicciones inicialmente es el mejor con un 97.86 % y 1468 que clasificó como correctas,



Tanto predicision y recall, están demasiado cerca de la media F1(F-measure) las discrepancias son mínimas.

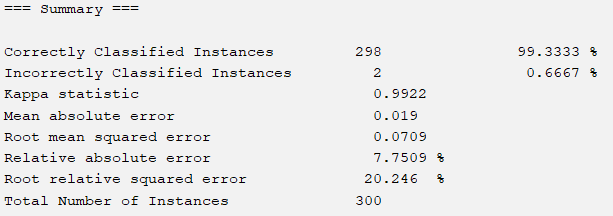


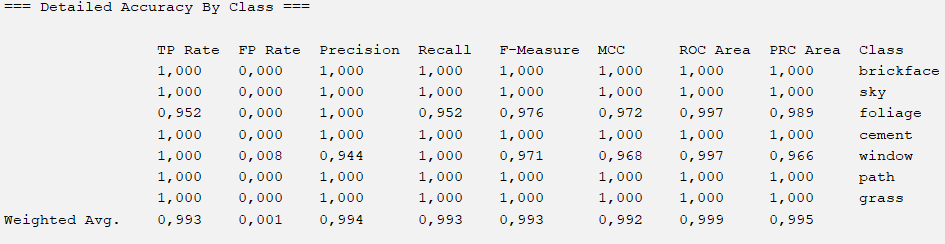
En matriz de confusión, vemos que la diagonal que está bien marcada mucho mejor que la J48 y MultilayerPerceptron, solo quedan 32 incorrectas



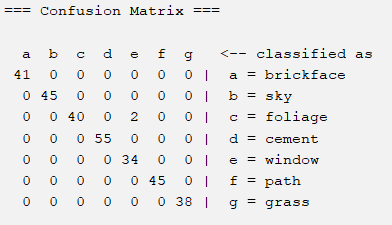
**Mejora de modelo: RandomForest**

Se puede decir que en un muestreo de un 80%, tenemos una clasificación correcta del 99.33% es decir 298 correctamente clasificados.

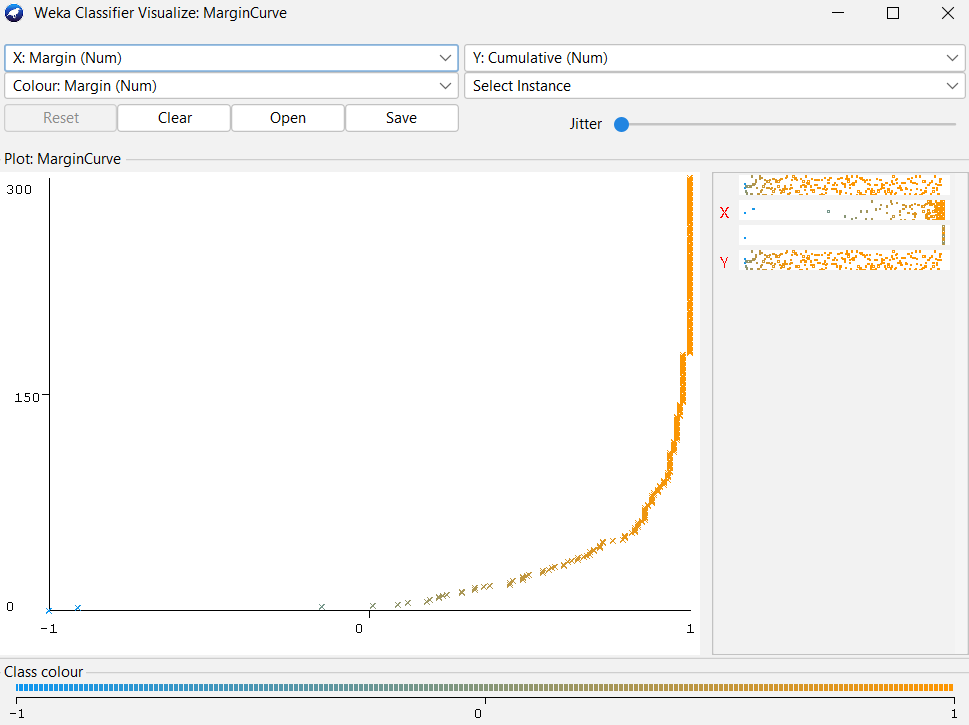
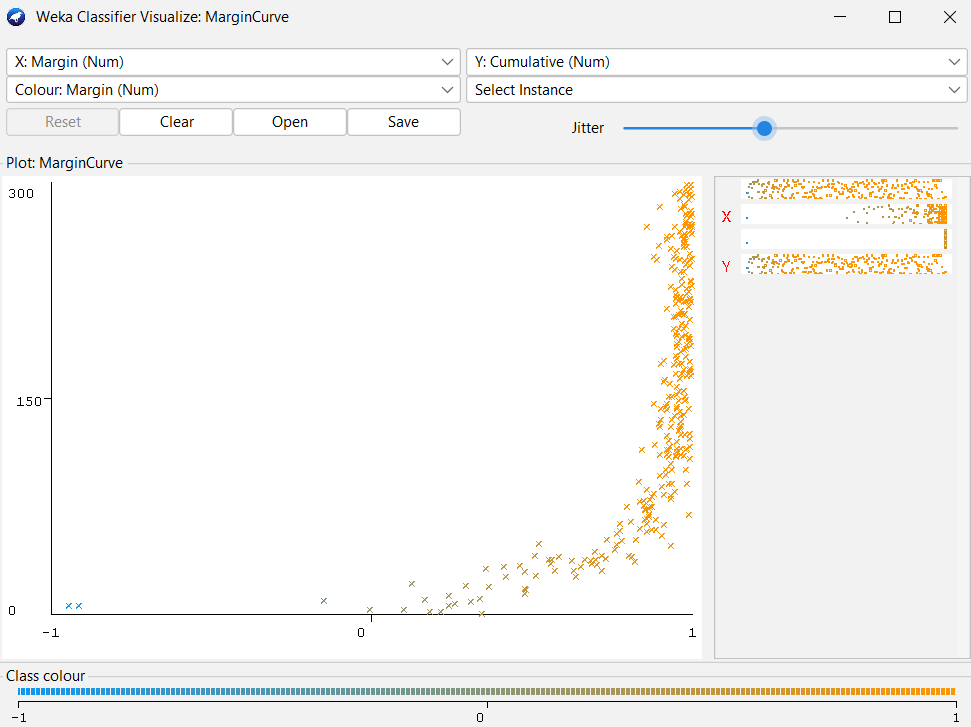




En matriz de confusión, tenemos una diagonal mejor que la J48 y MultilayerPerceptron y solo quedan 2 incorrectas



Al aumentar datos al modelo aun afecta el ruido que afecta, todos los modelos quedan afectados por el ruido.

>>

Conclusiones

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Clasificaciones correctas andes de la mejor | Clasificaciones correctas después de la mejor | Mejor tiempo, al clasificación |
| J48 | 95,73% | 96,67% | 0.06 segundo |
| Multilayer Perceptron | 97,07% | 98,00% | 2.63 segundos |
| Random Forest | 97,867% | 99,33% | 0.29 segundos |

El modelo que se acomoda a nuestros datos es el Random Forest, porque en la mejora obtuvo la mayor clasificación de datos correctos y los datos de la precisión y recall esta demasiado cerca de la media F1(F-measure).

En todos los modelos al aumentar dato son afectado por el ruido.