

## ANEXO 1

---

En este anexo se muestra para cada dataset sus atributos y para cada uno de ellos el valor máximo y mínimo, la media y la desviación estándar. Se presentan también los diagramas de dispersión en los que se relacionan cada par de grupos de cada dataset y que permiten determinar visualmente qué tan separados están las clases. Estos diagramas se realizaron con el programa Rapidminer [59] y cada dataset fue preprocesado en primer lugar para normalizar sus valores entre cero y uno (Ecuación **¡Error! No se encuentra el origen de la referencia.**) y luego para eliminar los atributos menos representativos, aquellos cuyas desviaciones estándar (calculadas con los atributos normalizados) están por debajo de un valor seleccionado manualmente, es decir, se probaron diferentes valores hasta que se eliminaron los atributos suficientes como para realizar los diagramas.

### DATASET IRIS

En la Figura 1 se muestran los atributos del dataset Iris y para cada uno el valor máximo, el valor mínimo, la media y la desviación estándar.

|              | Max | Min | Media | Desviación |
|--------------|-----|-----|-------|------------|
| Sepal length | 7.9 | 4.3 | 5.843 | 0.828      |
| Sepal width  | 4.4 | 2   | 3.054 | 0.434      |
| Petal length | 6.9 | 1   | 3.759 | 1.764      |
| Petal width  | 2.5 | 0.1 | 1.199 | 0.763      |

Figura 1: Atributos dataset iris

En Figura 2 se muestra el diagrama de dispersión del dataset Iris. Los atributos seleccionados son aquellos cuyas desviaciones estándar son superiores a 0.22.

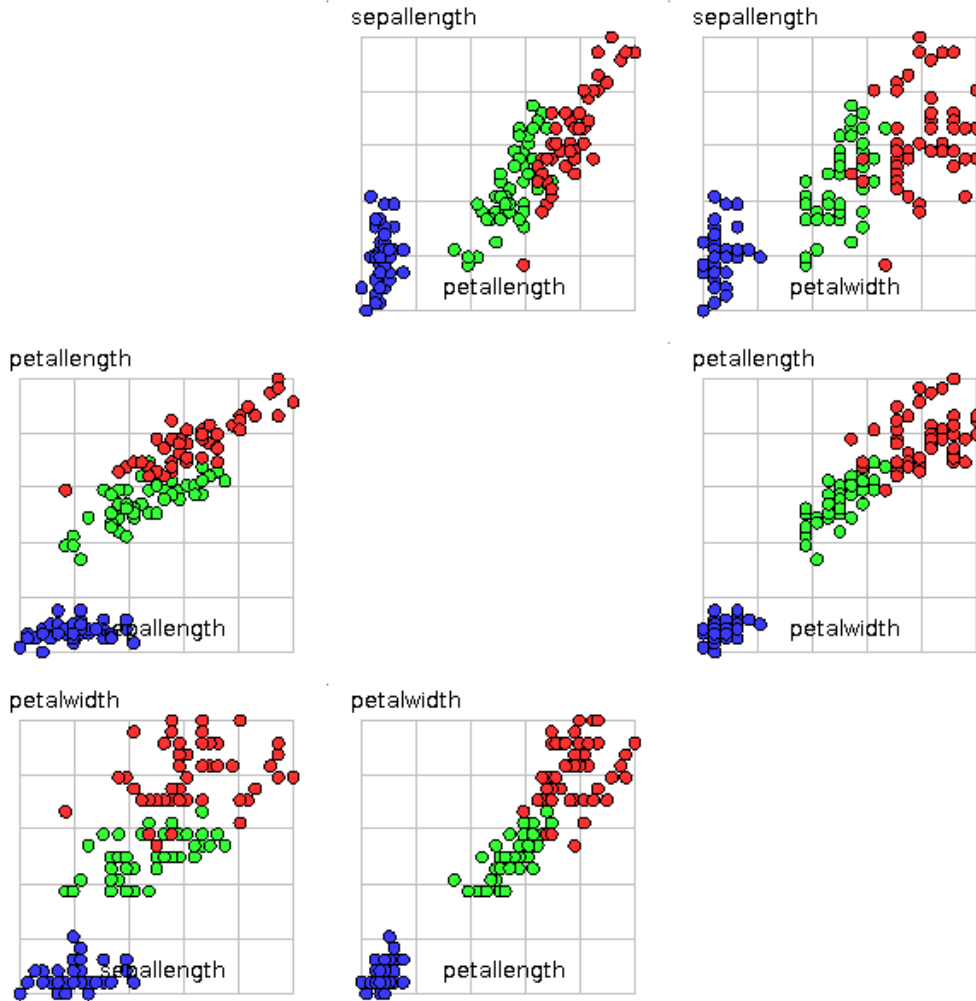


Figura 2: Diagrama de dispersión dataset iris

Se puede notar que el grupo correspondiente a iris-setosa (azul) no está solapado con ningún otro grupo mientras que los grupos iris-versicolor (verde) e iris-virginica (rojo) lo están ligeramente. Por otro lado se puede notar que los grupos son compactos, aunque iris-virginica lo es en menor medida que los demás.

## DATASET GLASS

En la Figura 3 se muestran los atributos del dataset Glass y para cada uno el valor máximo, el valor mínimo, la media y la desviación estándar

|    | Max   | Min   | Media  | Desviación |
|----|-------|-------|--------|------------|
| Ri | 15.34 | 15.11 | 15.18  | 0.003      |
| Na | 17.38 | 10.73 | 13.408 | 0.817      |
| Mg | 4.49  | 0     | 2.685  | 1.442      |
| Al | 3.5   | 0.29  | 1.445  | 0.499      |
| Si | 75.41 | 69.81 | 72.651 | 0.775      |
| K  | 6.21  | 0     | 0.497  | 0.652      |
| Ca | 16.19 | 5.43  | 8.957  | 1.423      |
| Ba | 3.15  | 0     | 0.175  | 0.497      |
| Fe | 0.51  | 0     | 0.057  | 0.097      |

Figura 3: Atributos dataset Glass

En Figura 4 se muestra el diagrama de dispersión del dataset Glass. Los atributos seleccionados son aquellos cuyas desviaciones estándar son superiores a 0.15.

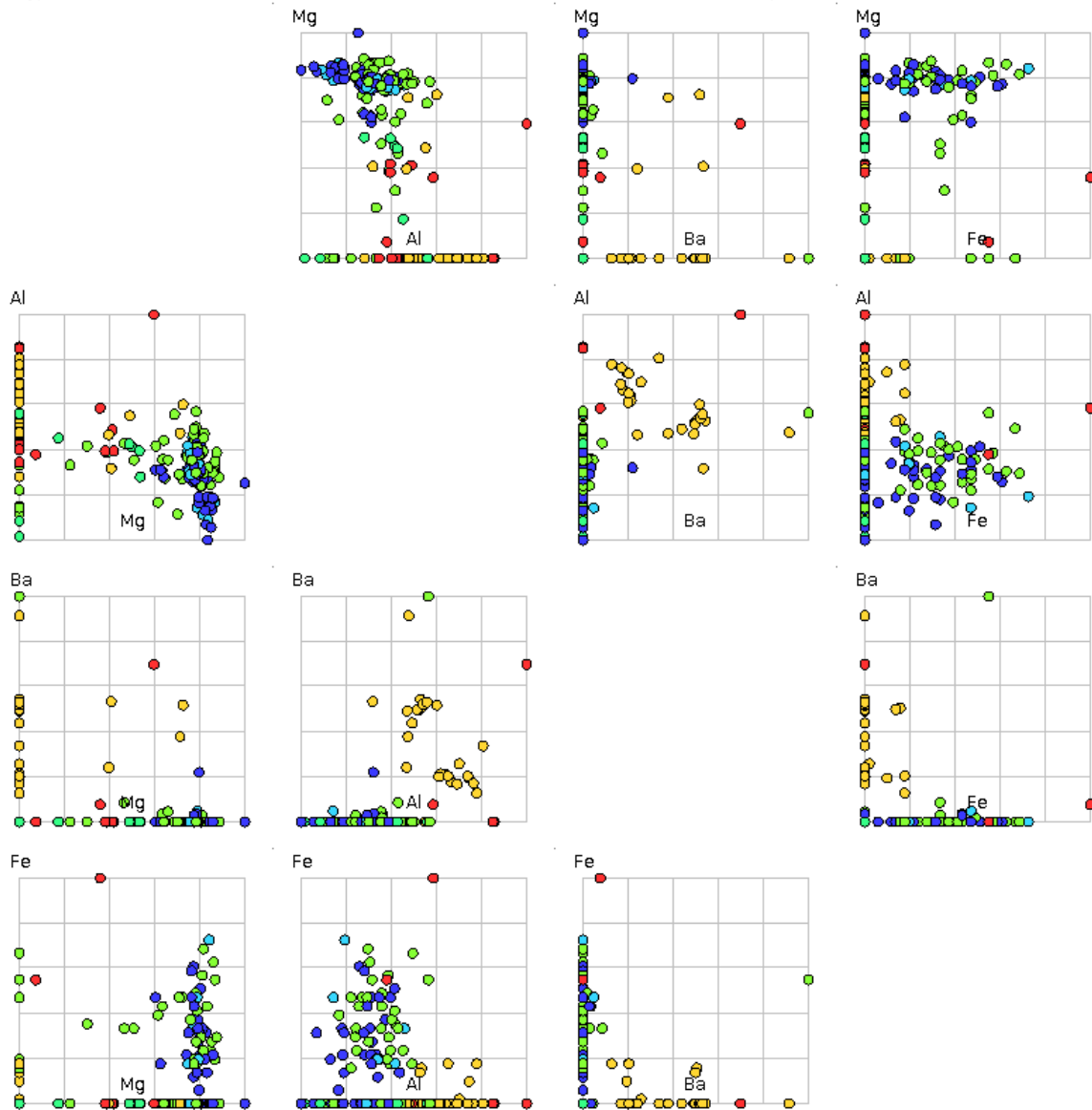


Figura 4: Diagrama de dispersión dataset glass

Se puede notar que los grupos build wind float (azul oscuro), vehic wind float (azul claro), tableware (verde azulado), build wind non-float (verde), headlamps (amarillo) y containers (rojo) no están bien delimitados, más bien están muy solapados y no son compactos.

## DATASET SONAR

En la Figura 5 se muestran los atributos del dataset Sonar y para cada uno el valor máximo, el valor mínimo, la media y la desviación estándar.

|              | Max   | Min   | Media | Desviación |
|--------------|-------|-------|-------|------------|
| attribute_1  | 0.137 | 0.002 | 0.029 | 0.023      |
| attribute_2  | 0.234 | 0.001 | 0.038 | 0.033      |
| attribute_3  | 0.306 | 0.002 | 0.044 | 0.038      |
| attribute_4  | 0.426 | 0.006 | 0.054 | 0.047      |
| attribute_5  | 0.401 | 0.007 | 0.075 | 0.056      |
| attribute_6  | 0.382 | 0.01  | 0.105 | 0.059      |
| attribute_7  | 0.373 | 0.003 | 0.122 | 0.062      |
| attribute_8  | 0.459 | 0.006 | 0.135 | 0.085      |
| attribute_9  | 0.683 | 0.008 | 0.178 | 0.118      |
| attribute_10 | 0.711 | 0.011 | 0.208 | 0.134      |
| attribute_11 | 0.734 | 0.029 | 0.236 | 0.133      |
| attribute_12 | 0.706 | 0.024 | 0.25  | 0.14       |
| attribute_13 | 0.713 | 0.018 | 0.273 | 0.141      |
| attribute_14 | 0.997 | 0.027 | 0.297 | 0.164      |
| attribute_15 | 1     | 0.003 | 0.32  | 0.205      |
| attribute_16 | 0.999 | 0.016 | 0.378 | 0.233      |
| attribute_17 | 1     | 0.035 | 0.416 | 0.264      |
| attribute_18 | 1     | 0.038 | 0.452 | 0.262      |
| attribute_19 | 1     | 0.049 | 0.505 | 0.258      |
| attribute_20 | 1     | 0.066 | 0.563 | 0.263      |
| attribute_21 | 1     | 0.051 | 0.609 | 0.258      |
| attribute_22 | 1     | 0.022 | 0.624 | 0.256      |
| attribute_23 | 1     | 0.056 | 0.647 | 0.25       |
| attribute_24 | 1     | 0.024 | 0.673 | 0.239      |
| attribute_25 | 1     | 0.024 | 0.675 | 0.245      |
| attribute_26 | 1     | 0.092 | 0.7   | 0.237      |
| attribute_27 | 1     | 0.048 | 0.702 | 0.246      |
| attribute_28 | 1     | 0.028 | 0.694 | 0.237      |
| attribute_29 | 1     | 0.014 | 0.642 | 0.24       |
| attribute_30 | 1     | 0.061 | 0.581 | 0.221      |
| attribute_31 | 0.966 | 0.048 | 0.504 | 0.214      |
| attribute_32 | 0.931 | 0.04  | 0.439 | 0.213      |
| attribute_33 | 1     | 0.048 | 0.417 | 0.207      |
| attribute_34 | 0.965 | 0.021 | 0.403 | 0.231      |
| attribute_35 | 1     | 0.022 | 0.393 | 0.259      |

|              |       |       |       |       |
|--------------|-------|-------|-------|-------|
| attribute_36 | 1     | 0.008 | 0.385 | 0.264 |
| attribute_37 | 0.95  | 0.035 | 0.364 | 0.24  |
| attribute_38 | 1     | 0.038 | 0.34  | 0.213 |
| attribute_39 | 0.986 | 0.037 | 0.326 | 0.199 |
| attribute_40 | 0.93  | 0.012 | 0.311 | 0.179 |
| attribute_41 | 0.899 | 0.036 | 0.289 | 0.171 |
| attribute_42 | 0.825 | 0.006 | 0.278 | 0.169 |
| attribute_43 | 0.773 | 0     | 0.247 | 0.139 |
| attribute_44 | 0.776 | 0     | 0.214 | 0.133 |
| attribute_45 | 0.703 | 0     | 0.197 | 0.152 |
| attribute_46 | 0.729 | 0     | 0.161 | 0.134 |
| attribute_47 | 0.552 | 0     | 0.122 | 0.087 |
| attribute_48 | 0.334 | 0     | 0.091 | 0.062 |
| attribute_49 | 0.198 | 0     | 0.052 | 0.036 |
| attribute_50 | 0.083 | 0     | 0.02  | 0.014 |
| attribute_51 | 0.1   | 0     | 0.016 | 0.012 |
| attribute_52 | 0.071 | 0.001 | 0.013 | 0.01  |
| attribute_53 | 0.039 | 0.001 | 0.011 | 0.007 |
| attribute_54 | 0.035 | 0.001 | 0.011 | 0.007 |
| attribute_55 | 0.045 | 0.001 | 0.009 | 0.007 |
| attribute_56 | 0.039 | 0     | 0.008 | 0.006 |
| attribute_57 | 0.035 | 0     | 0.008 | 0.006 |
| attribute_58 | 0.044 | 0     | 0.008 | 0.006 |
| attribute_59 | 0.036 | 0     | 0.008 | 0.006 |
| attribute_60 | 0.044 | 0.001 | 0.007 | 0.005 |

Figura 5: Atributos dataset Sonar

En Figura 6 se muestra el diagrama de dispersión del dataset Sonar. Los atributos seleccionados son aquellos cuyas desviaciones estándar son superiores a 0.27. En la misma se puede notar que el grupo rock (azul) y mine (rojo) están muy solapados y dispersos.

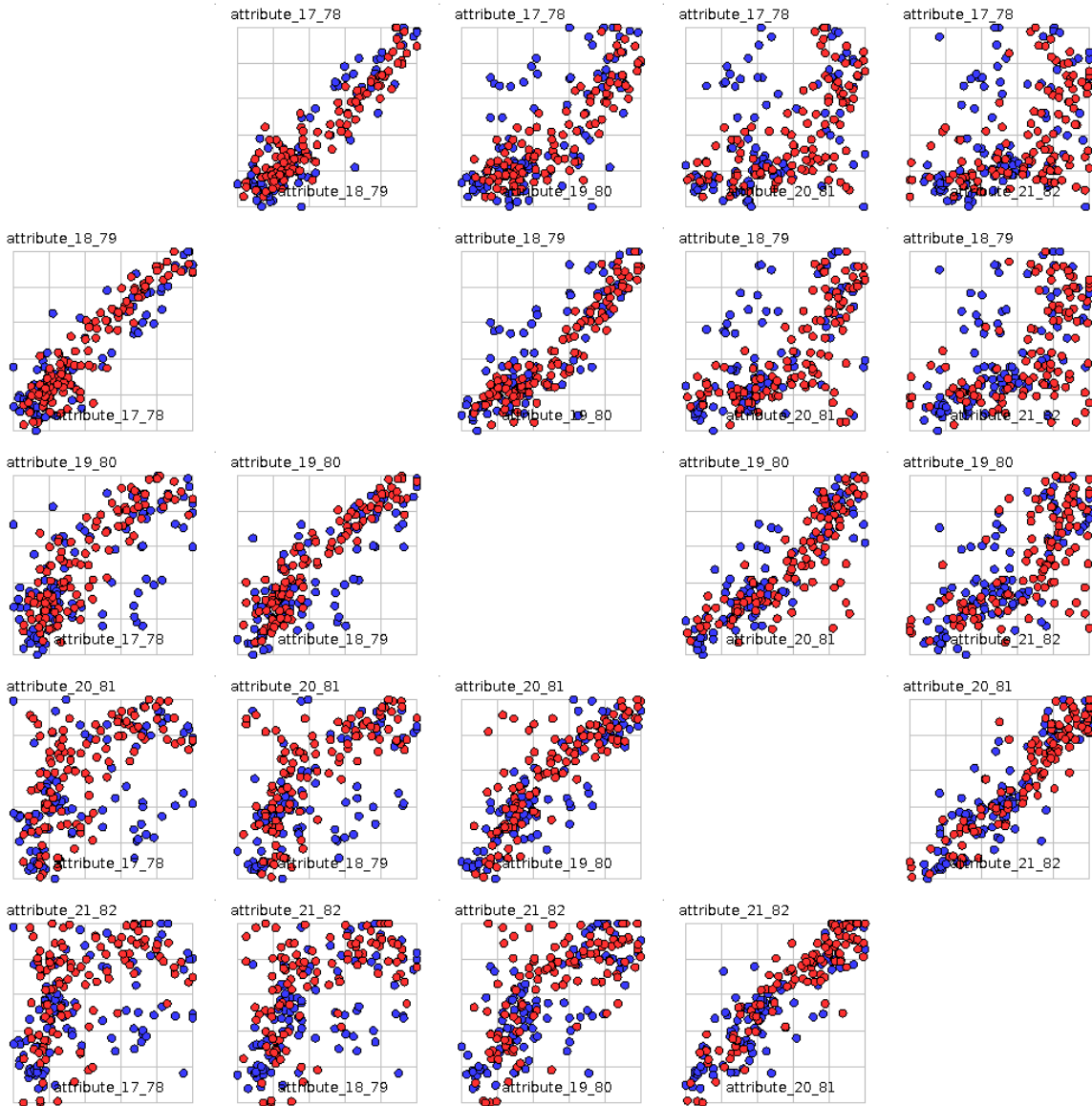


Figura 6: Diagrama de dispersión dataset sonar

## DATASET WDBC

En la Figura 7 se muestran los atributos del dataset WDBC y para cada uno el valor máximo, el valor mínimo, la media y la desviación estándar.

En Figura 8 se muestra el diagrama de dispersión del dataset WDBC. Los atributos seleccionados son aquellos cuyas desviaciones estándar son superiores a 0.17.

|            | Max   | Min   | Media   | Desviación |
|------------|-------|-------|---------|------------|
| Feature_1  | 28.11 | 6.981 | 12.127  | 3.523      |
| Feature_2  | 39.28 | 9.71  | 19.29   | 4.301      |
| Feature_3  | 188.5 | 43.79 | 91.969  | 24.299     |
| Feature_4  | 2501  | 143.5 | 654.889 | 351.914    |
| Feature_5  | 0.163 | 0.053 | 0.096   | 0.014      |
| Feature_6  | 0.345 | 0.019 | 0.104   | 0.053      |
| Feature_7  | 0.427 | 0     | 0.089   | 0.08       |
| Feature_8  | 0.201 | 0     | 0.049   | 0.039      |
| Feature_9  | 0.304 | 0.106 | 0.181   | 0.027      |
| Feature_10 | 0.097 | 0.05  | 0.063   | 0.007      |
| Feature_11 | 2.873 | 0.112 | 0.405   | 0.277      |
| Feature_12 | 4.885 | 0.36  | 1.217   | 0.552      |
| Feature_13 | 21.98 | 0.757 | 2.866   | 2.022      |
| Feature_14 | 542.2 | 6.802 | 40.337  | 45.491     |
| Feature_15 | 0.031 | 0.002 | 0.007   | 0.003      |
| Feature_16 | 0.135 | 0.002 | 0.025   | 0.018      |
| Feature_17 | 0.396 | 0     | 0.032   | 0.03       |
| Feature_18 | 0.053 | 0     | 0.012   | 0.006      |
| Feature_19 | 0.079 | 0.008 | 0.021   | 0.008      |
| Feature_20 | 0.03  | 0.001 | 0.004   | 0.003      |
| Feature_21 | 36.04 | 7.93  | 16.269  | 4.833      |
| Feature_22 | 49.54 | 12.02 | 25.677  | 6.146      |
| Feature_23 | 251.2 | 50.41 | 107.261 | 33.603     |
| Feature_24 | 425.4 | 185.2 | 880.583 | 569.357    |
| Feature_25 | 0.223 | 0.071 | 0.132   | 0.023      |
| Feature_26 | 1.058 | 0.027 | 0.254   | 0.157      |
| Feature_27 | 1.252 | 0     | 0.272   | 0.209      |
| Feature_28 | 0.291 | 0     | 0.115   | 0.066      |
| Feature_29 | 0.664 | 0.157 | 0.29    | 0.062      |
| Feature_30 | 0.207 | 0.055 | 0.084   | 0.018      |



Figura 7: Atributos dataset WDBC

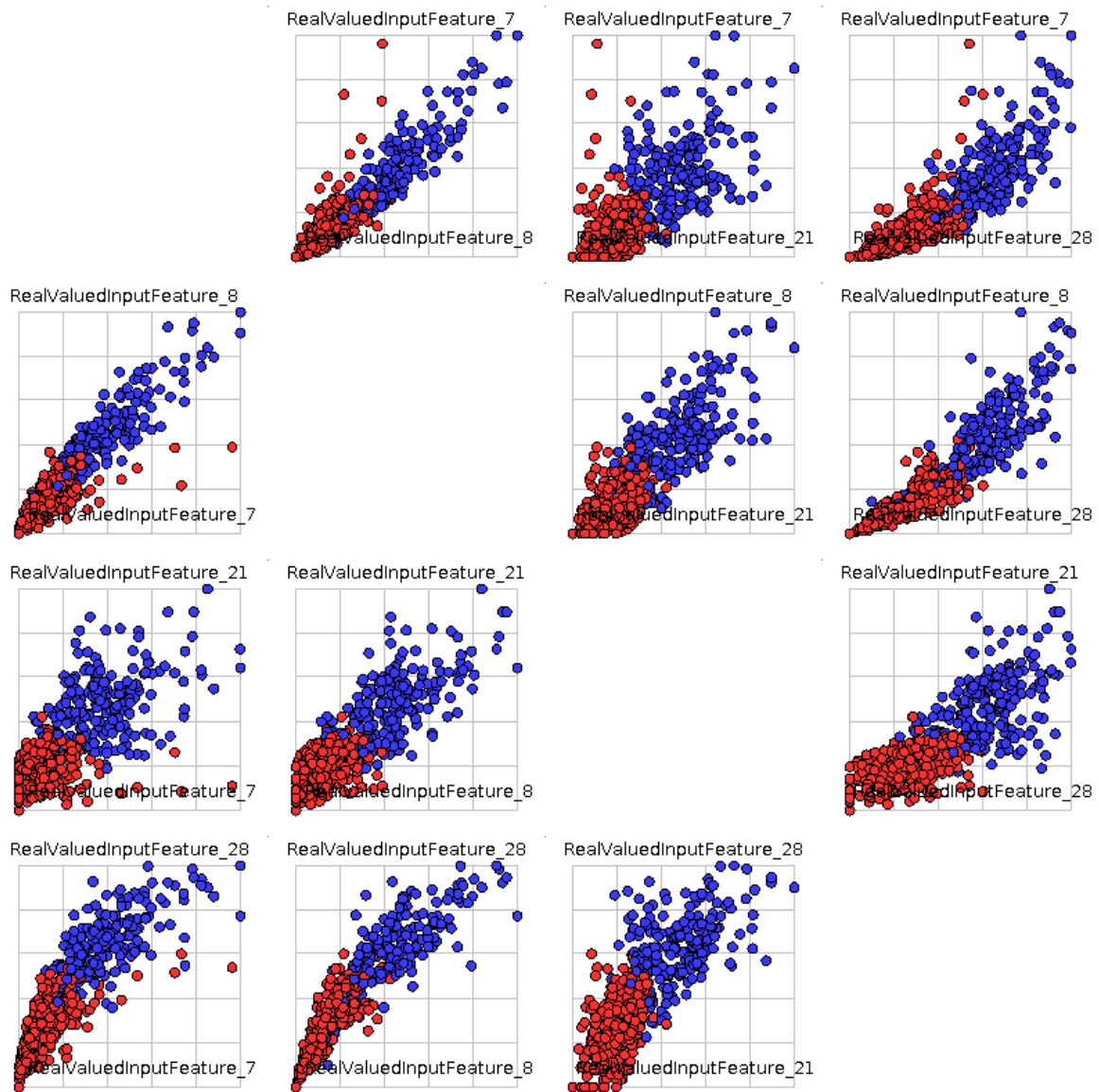


Figura 8: Diagrama de dispersión dataset wdbc

Puede verse que los dos grupos M (azul) y B (rojo) son compactos y ligeramente solapados.

## DATASET WINE

En la Figura 9 se muestran los atributos del dataset Wine y para cada uno el valor máximo, el valor mínimo, la media y la desviación estándar.

|                              | Max   | Min   | Media   | Desviación |
|------------------------------|-------|-------|---------|------------|
| Alcohol                      | 14.83 | 11.03 | 13.001  | 0.812      |
| Malic_acid                   | 5.8   | 0.74  | 2.336   | 1.117      |
| Ash                          | 3.23  | 1.36  | 2.367   | 0.274      |
| Alcalinity_of_ash            | 30    | 10.6  | 19.495  | 3.34       |
| Magnesium                    | 162   | 70    | 99.742  | 14.282     |
| Total_phenols                | 3.88  | 0.98  | 2.295   | 0.626      |
| Flavanoids                   | 5.08  | 0.34  | 2.029   | 0.999      |
| Nonflavanoids_phenols        | 0.66  | 0.13  | 0.362   | 0.124      |
| Proanthocyanins              | 3.58  | 0.41  | 1.591   | 0.572      |
| Color_intensity              | 13    | 1.28  | 5.058   | 2.318      |
| Hue                          | 1.71  | 0.48  | 0.957   | 0.229      |
| OD280/OD315_of_diluted_wines | 4     | 1.27  | 2.612   | 0.71       |
| Proline                      | 1680  | 278   | 746.893 | 314.907    |

Figura 9: Atributos dataset Wine

En la Figura 10 se muestra el diagrama de dispersión del dataset Wine. Los atributos seleccionados son aquellos cuyas desviaciones estándar son superiores a 0.22.

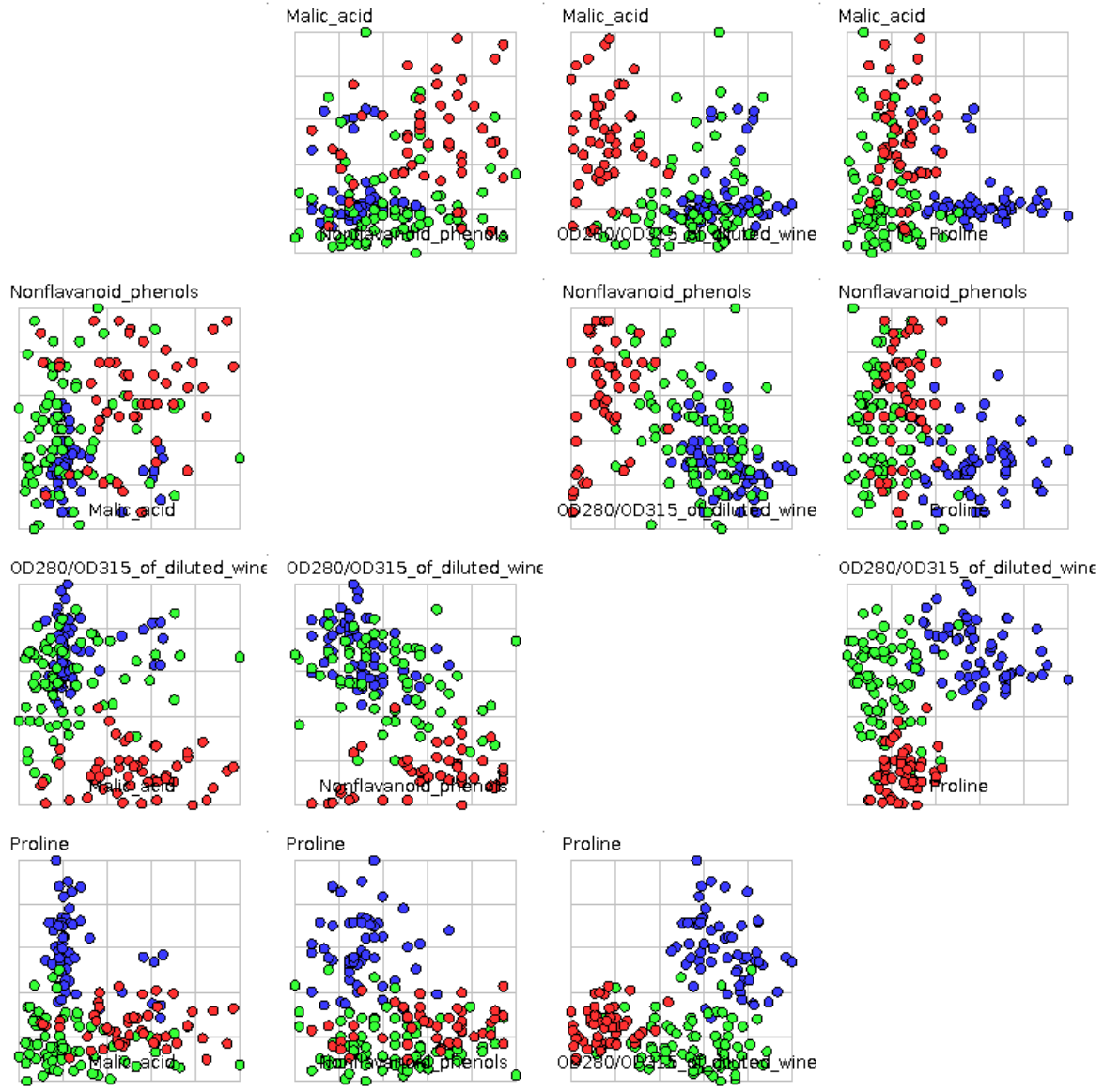


Figura 10: Diagrama de dispersión dataset Wine

Se puede observar que los grupos 1 (azul), 2 (verde) y 3 (rojo) no son compactos, además están muy solapados, donde el grupo 1 parece estar un poco más separado de los otros dos.

### VALORES DE AFINACIÓN

El proceso de afinación de los parámetros de los algoritmos consistió en generar de manera aleatoria los valores de minPar, maxPar, hmcr, hms, po, nlt, maxK y maxKMeans; generar las tareas como se vio en la sección **¡Error! No se encuentra el origen de la referencia.** con los otros parámetros definidos en el archivo json y la ejecución de las mismas. Finalmente se calcula el promedio de los errores obtenidos en cada ejecución y se almacena en un archivo csv los valores de los parámetros junto al error promedio. Con el parámetro tuneUplt se determina cuantas veces se ejecuta el proceso de afinación. Aquí se muestran los valores de parámetros obtenidos tras 57 ejecuciones del proceso de afinación.

| minPar | maxPar | hmcr | hms   | Po   | nlt     | maxK  | maxKMeans | ER    |
|--------|--------|------|-------|------|---------|-------|-----------|-------|
| 0,43   | 0,48   | 0,43 | 18,00 | 0,16 | 938,00  | 2,00  | 43,00     | 34,47 |
| 0,95   | 1,00   | 0,99 | 28,00 | 0,81 | 164,00  | 2,00  | 109,00    | 34,49 |
| 0,27   | 0,85   | 0,92 | 9,00  | 0,94 | 698,00  | 2,00  | 49,00     | 34,52 |
| 0,37   | 0,79   | 0,61 | 14,00 | 0,77 | 528,00  | 2,00  | 24,00     | 34,53 |
| 0,74   | 0,88   | 0,68 | 27,00 | 0,50 | 572,00  | 2,00  | 13,00     | 34,67 |
| 0,47   | 0,86   | 0,34 | 47,00 | 0,09 | 736,00  | 3,00  | 46,00     | 34,99 |
| 0,39   | 0,45   | 0,66 | 35,00 | 0,21 | 865,00  | 4,00  | 104,00    | 36,44 |
| 0,22   | 0,99   | 0,66 | 18,00 | 0,40 | 766,00  | 4,00  | 41,00     | 36,69 |
| 0,16   | 0,75   | 0,34 | 7,00  | 0,55 | 1054,00 | 7,00  | 32,00     | 37,21 |
| 0,45   | 0,55   | 0,87 | 29,00 | 0,89 | 122,00  | 6,00  | 96,00     | 37,26 |
| 0,93   | 1,00   | 0,40 | 10,00 | 0,85 | 601,00  | 5,00  | 71,00     | 37,35 |
| 0,94   | 0,99   | 0,49 | 13,00 | 0,11 | 668,00  | 5,00  | 34,00     | 37,43 |
| 0,94   | 0,97   | 0,06 | 24,00 | 0,74 | 296,00  | 7,00  | 74,00     | 37,54 |
| 0,04   | 0,61   | 0,67 | 37,00 | 0,87 | 429,00  | 7,00  | 19,00     | 37,74 |
| 0,32   | 0,45   | 0,15 | 14,00 | 0,79 | 914,00  | 14,00 | 61,00     | 37,76 |
| 0,18   | 0,81   | 0,15 | 8,00  | 0,25 | 506,00  | 8,00  | 56,00     | 37,87 |
| 0,73   | 0,75   | 0,01 | 34,00 | 0,74 | 1087,00 | 8,00  | 36,00     | 37,92 |
| 0,72   | 0,88   | 0,95 | 13,00 | 0,72 | 537,00  | 10,00 | 49,00     | 37,93 |
| 0,13   | 0,70   | 0,89 | 29,00 | 0,70 | 196,00  | 14,00 | 42,00     | 37,97 |
| 0,97   | 1,00   | 0,38 | 3,00  | 0,61 | 851,00  | 15,00 | 91,00     | 38,00 |
| 0,10   | 0,60   | 0,27 | 40,00 | 0,74 | 220,00  | 8,00  | 38,00     | 38,04 |
| 0,67   | 0,80   | 0,04 | 42,00 | 0,22 | 322,00  | 8,00  | 45,00     | 38,15 |
| 0,61   | 0,69   | 0,26 | 20,00 | 0,50 | 1052,00 | 21,00 | 35,00     | 38,15 |
| 0,37   | 0,98   | 0,82 | 38,00 | 0,77 | 863,00  | 17,00 | 70,00     | 38,16 |
| 0,95   | 0,97   | 0,77 | 39,00 | 0,41 | 1011,00 | 8,00  | 88,00     | 38,19 |
| 0,32   | 0,91   | 0,20 | 18,00 | 0,38 | 464,00  | 21,00 | 100,00    | 38,33 |
| 0,33   | 0,60   | 0,32 | 7,00  | 0,75 | 473,00  | 9,00  | 64,00     | 38,42 |
| 0,93   | 0,96   | 0,07 | 47,00 | 0,75 | 1046,00 | 18,00 | 76,00     | 38,44 |
| 0,30   | 0,37   | 0,20 | 16,00 | 0,99 | 520,00  | 21,00 | 79,00     | 38,45 |
| 0,98   | 1,00   | 0,72 | 17,00 | 0,48 | 109,00  | 11,00 | 26,00     | 38,53 |
| 0,72   | 0,82   | 0,50 | 17,00 | 0,31 | 687,00  | 12,00 | 57,00     | 38,55 |
| 0,71   | 0,95   | 0,05 | 49,00 | 0,67 | 208,00  | 18,00 | 106,00    | 38,58 |

|      |      |      |       |      |         |       |        |       |
|------|------|------|-------|------|---------|-------|--------|-------|
| 0,83 | 0,94 | 0,48 | 36,00 | 0,53 | 517,00  | 21,00 | 35,00  | 38,61 |
| 0,36 | 0,43 | 0,64 | 28,00 | 0,82 | 898,00  | 10,00 | 94,00  | 38,63 |
| 0,13 | 0,91 | 0,13 | 44,00 | 0,97 | 1052,00 | 15,00 | 58,00  | 38,66 |
| 0,84 | 0,98 | 0,98 | 29,00 | 0,58 | 637,00  | 18,00 | 29,00  | 38,67 |
| 0,59 | 0,88 | 0,09 | 43,00 | 0,98 | 400,00  | 18,00 | 10,00  | 38,76 |
| 0,43 | 0,74 | 0,17 | 12,00 | 0,73 | 850,00  | 8,00  | 16,00  | 38,80 |
| 0,61 | 0,75 | 0,97 | 12,00 | 0,66 | 303,00  | 9,00  | 22,00  | 38,88 |
| 0,31 | 0,89 | 0,46 | 26,00 | 0,11 | 691,00  | 10,00 | 37,00  | 38,91 |
| 0,81 | 0,96 | 0,43 | 41,00 | 0,56 | 799,00  | 19,00 | 74,00  | 38,93 |
| 0,25 | 0,37 | 0,91 | 16,00 | 0,47 | 707,00  | 10,00 | 89,00  | 38,95 |
| 0,92 | 0,99 | 0,13 | 15,00 | 0,68 | 1014,00 | 10,00 | 79,00  | 39,03 |
| 0,42 | 0,46 | 0,18 | 21,00 | 0,24 | 724,00  | 15,00 | 32,00  | 39,04 |
| 0,04 | 0,65 | 0,48 | 21,00 | 0,72 | 1028,00 | 21,00 | 31,00  | 39,08 |
| 0,74 | 0,79 | 0,69 | 15,00 | 0,11 | 546,00  | 15,00 | 50,00  | 39,11 |
| 0,96 | 0,99 | 0,16 | 14,00 | 0,88 | 460,00  | 14,00 | 17,00  | 39,20 |
| 0,85 | 0,95 | 0,76 | 15,00 | 0,34 | 842,00  | 19,00 | 107,00 | 39,95 |
| 0,24 | 0,45 | 0,29 | 4,00  | 0,87 | 620,00  | 17,00 | 30,00  | 40,74 |
| 0,24 | 0,24 | 0,75 | 37,00 | 0,07 | 233,00  | 12,00 | 16,00  | 40,82 |
| 0,77 | 0,91 | 0,34 | 23,00 | 0,19 | 221,00  | 21,00 | 108,00 | 40,96 |
| 0,17 | 0,29 | 0,94 | 1,00  | 0,04 | 237,00  | 8,00  | 108,00 | 42,66 |
| 0,43 | 0,93 | 0,88 | 8,00  | 0,11 | 920,00  | 19,00 | 77,00  | 43,01 |
| 0,34 | 0,49 | 0,54 | 12,00 | 0,07 | 433,00  | 12,00 | 31,00  | 43,19 |
| 0,41 | 0,70 | 0,24 | 16,00 | 0,01 | 347,00  | 13,00 | 103,00 | 44,15 |
| 0,23 | 0,29 | 0,90 | 2,00  | 0,03 | 731,00  | 11,00 | 37,00  | 45,56 |
| 0,26 | 0,42 | 0,83 | 8,00  | 0,04 | 246,00  | 14,00 | 26,00  | 45,97 |

Tabla 1: Valores de ejecución del proceso de afinación