Filtro PrincipalComponents. Entre otras cosas nos permite obtener
la matriz de correlacios entre los atributos.

```
=== Run information ===

Evaluator:    weka.attributeSelection.PrincipalComponents -R 0.95 -A 5
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     iris
Instances:    150
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
Evaluation mode:    evaluate on all training data



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

Search Method:
    Attribute ranking.

Attribute Evaluator (unsupervised):
    Principal Components Attribute Transformer

Correlation matrix
  1     -0.11   0.87   0.82
 -0.11   1     -0.42  -0.36
  0.87  -0.42   1      0.96
  0.82  -0.36   0.96   1


eigenvalue  proportion  cumulative
  2.91082     0.7277      0.7277    -0.581petallength-0.566petalwidth-0.522sepallength+0.263sepalwidth
  0.92122     0.23031     0.95801   0.926sepalwidth+0.372sepallength+0.065petalwidth+0.021petallength

Eigenvectors
 V1  V2
-0.5224  0.3723 sepallength
 0.2634  0.9256 sepalwidth
-0.5813  0.0211 petallength
-0.5656  0.0654 petalwidth

Ranked attributes:
 0.2723  1 -0.581petallength-0.566petalwidth-0.522sepallength+0.263sepalwidth
 0.042   2 0.926sepalwidth+0.372sepallength+0.065petalwidth+0.021petallength

Selected attributes: 1,2 : 2
```

Comentarios a la salida de Weka:
==========
El objetivo del filtro PrincipalComponents es reducir la dimensionalidad de un problema.

Si la dimensionalidad es el número de atributos (variables independientes o características de entrada de el problema), al reducirla hacemos que los algoritmos de aprendizaje automático puedan trabajar con menos datos, de forma que podrán obtener modelos de la realidad de una manera más rápida ante una gran cantidad de información. Por otro lado, menor cantidad de información podría ayudarnos a hacer una interpretación más sencilla de un modelo.

En el caso de Iris, nos hemos quedado con solamente dos atributos de entrada en vez de 4. Cada uno de esos 2 nuevos atributos se llama componente principal, y es un combinación de los que había en origen.

Se puede aprovechar el filtro PrincipalComponents para ver como están correladas las variables de entrada o atributos de nuestro problema. Correlación significa la relación entre dos o más variables. Podemos decir que la correlación indica la fuerza y la dirección de una relación lineal y la proporcionalidad entre dos variables estadísticas. Se considera que dos variables están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra.

El valor de correlación entre variables es un valor que está en el intervalo [-1,1].
Se puede tomar de manera orientativa, que valores menores a -0.6 y mayores a 0.6 indican que un par de atributos de la matriz están correlados.

Por ejemplo, si tenemos dos variables (A y B) existe correlación entre ellas si al disminuir los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad (vea el ejemplo entre la relación de consumir helados y usar gafas de sol).

Volviendo al ejemplo de Iris, podemos ver que las variables que más relación tiene entre si son (Correlation matrix):

	sepallength	sepalwidth	petallength	petalwidth
sepallength	1	-0.11	0.87	0.82
sepalwidth	-0.11	1	-0.42	-0.36
petallength	0.87	-0.42	1	0.96
petalwidth	0.82	-0.36	0.96	1

sepallength y petallength con 0.87
petallength y petalwidth con 0.96

Podemos decir que cuando aumenta una de las variables lo hace también la otra y viceversa. Esto nos puede servir para comprender mejor el problema, pero el que la relación entre las variables sea muy fuerte (esto es, que sea casi 1 o casi -1) no significa que una de ellas sea la causa de la otra.

Si quisieramos eliminar variables, normalmente se puede probar con auquellas que tengan un valor mayor a 0.6 en su coeficiente de correlación.

Como ejemplo, petallength vs petalwidth = 0.96

¿Cúal eliminamos de las dos? La que menos correlación tenga con respecto a la salida, pero eso no lo muestra este filtro.

Para ello podríamos usar el filtro Select attributes -> CorrelationAttributeEval,Ranker. Este filtro nos ordena la importancia de los atributos con respecto a la salida, y de los atributos petallength y petalwidth eliminariamos el que esté dispuesto más tarde en el orden, ya que explicaría menos la salida.