



UNIVERSIDAD NACIONAL DE LOJA

FACULTAD DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES

CARRERA EN COMPUTACIÓN

PASOS PARA LA APLICACIÓN DEL OneR EN WEKA

Carrera: Ingeniería en Sistemas “A”

Fecha: 6 de febrero de 2022

Materia: Inteligencia Artificial

Integrantes:

- Francisco Agreda Sánchez
- Israel Campoverde Peñaherrera
- Eduardo Leon Castillo
- Josué Ortega Jaramillo

PASOS:

Para realizar la aplicación del algoritmo se debe tener instalado la herramienta Weka y tener descargado previamente el data set de “Harry Potter Movies Dataset” que se encuentra en este Github o descargarlo de la página oficial:

<https://www.kaggle.com/maricinnamon/harry-potter-movies-dataset>

Luego, se procede a abrir la herramienta, se mostrará el área de trabajo con una serie de opciones que se detalla a continuación:

Preprocess:

En esta sección permite abrir el data set seleccionado, para ello presionamos el botón Open file y seleccionamos el archivo con extensión .arff o csv, una vez abierto el data set se mostrará los atributos que consta nuestra base de datos como: Species, Gender, Patronus, Character name, Etc, como se muestra en la figura 1.

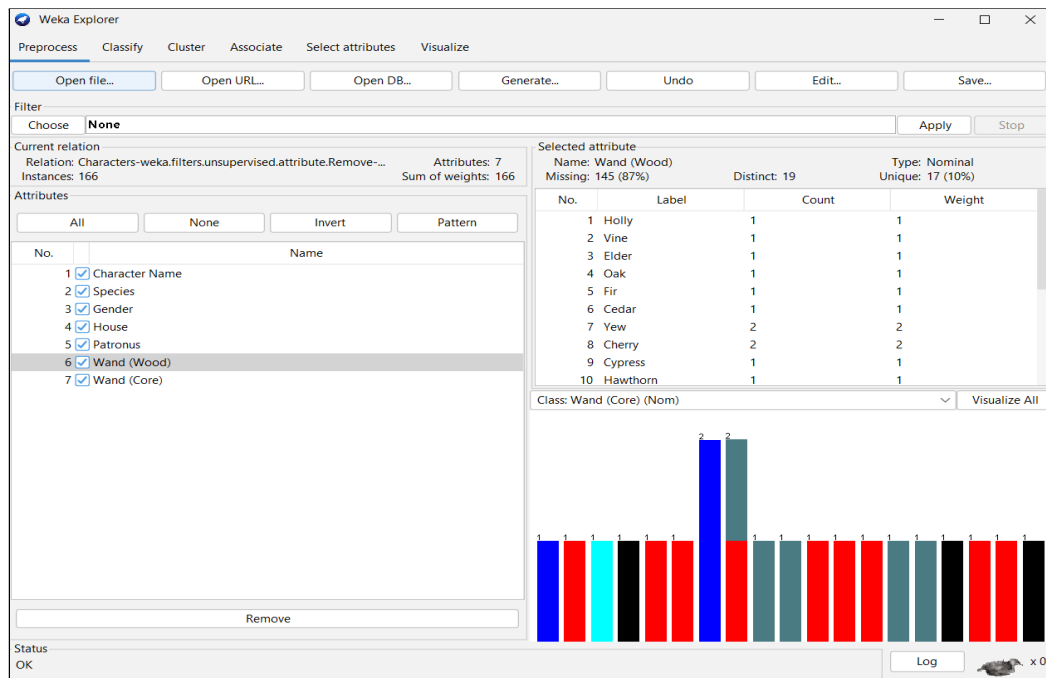


Figura 1: Área de trabajo de weka sección “Preprocess”

Classification:

Nos permite seleccionar el algoritmo, el cual hemos visto conveniente utilizar el algoritmo **OneR**. Luego en “*Test options*” hemos escogido “*Use training set*” porque nos permite utilizar todos los registros o los datos que se cargaron del data set en WEKA.

El (Nom) House nos ayuda a que el atributo según su clasificación, en este caso estoy escogiendo que haga el proceso de clasificado según las Casas que tiene la película de Harry Potter, el resultado que como se muestra en la figura 2.

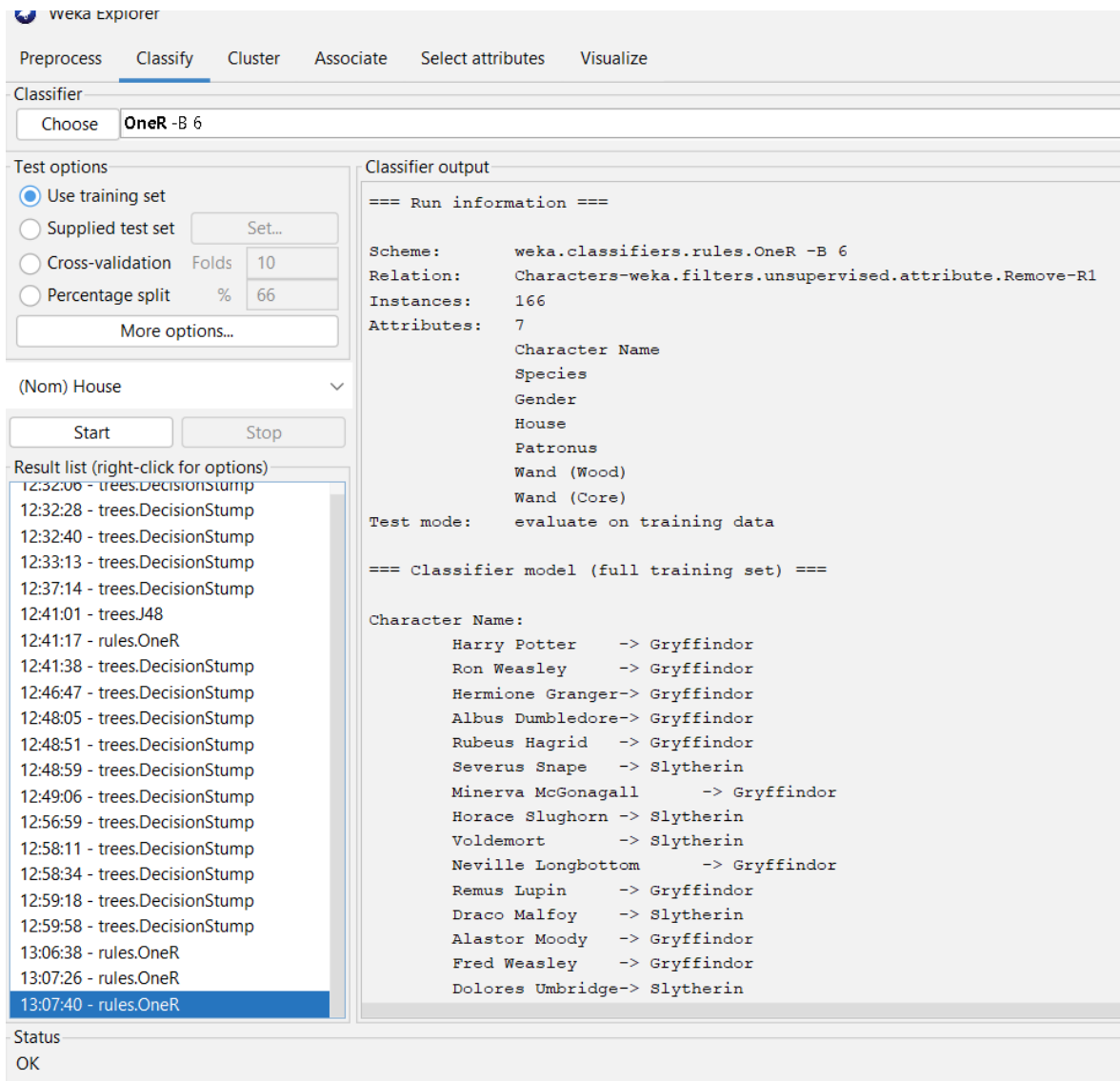


Figura 2: Área de trabajo de weka, sección "Classify"

El resultado que se mostrará una vez seleccionado las opciones anteriores es el siguiente: Como se puede observar en la figura 3, se tiene 166 instancias y 7 atributos los cuales son: Character Name, Species, Gender, House, Patronus, Wand (Wood), Wand (Core).

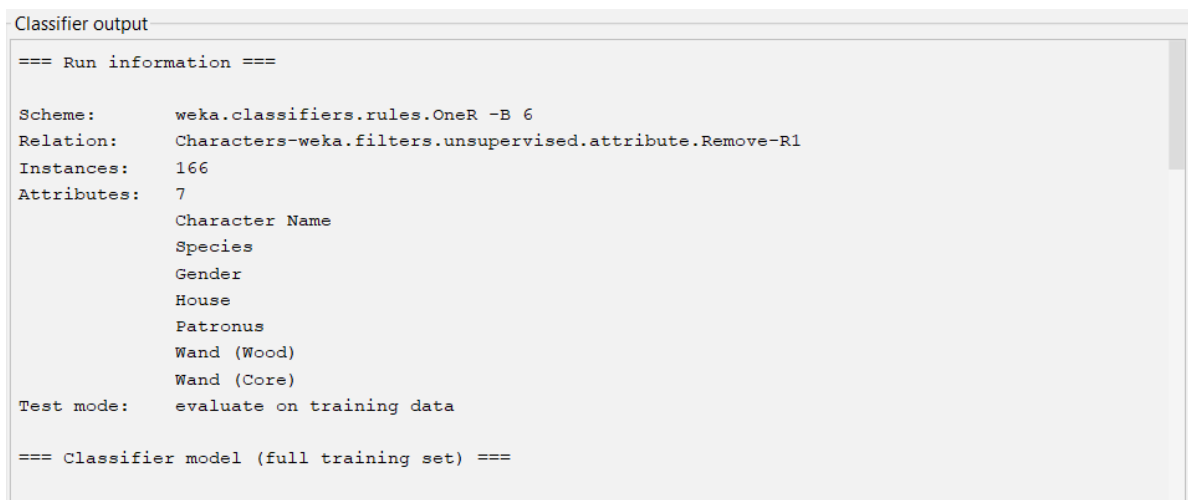


Figura 3: Sección "Classify", resultados obtenidos.

A Continuación en la figura 4, se visualiza los datos del dataset, en donde están los nombres de los personajes de Harry Potter y su Casa a la que pertenecen.

```
=== Classifier model (full training set) ===

Character Name:
  Harry Potter      -> Gryffindor
  Ron Weasley       -> Gryffindor
  Hermione Granger -> Gryffindor
  Albus Dumbledore -> Gryffindor
  Rubeus Hagrid     -> Gryffindor
  Severus Snape      -> Slytherin
  Minerva McGonagall -> Gryffindor
  Horace Slughorn    -> Slytherin
  Voldemort          -> Slytherin
  Neville Longbottom -> Gryffindor
  Remus Lupin        -> Gryffindor
  Draco Malfoy       -> Slytherin
  Alastor Moody      -> Gryffindor
  Fred Weasley       -> Gryffindor
  Dolores Umbridge -> Slytherin
  Arthur Weasley     -> Gryffindor
  Cornelius Fudge    -> Gryffindor
  Sirius Black       -> Gryffindor
  George Weasley     -> Gryffindor
  Ginny Weasley      -> Gryffindor
  Vernon Dursley     -> Gryffindor
  Bellatrix Lestrange -> Slytherin
  Lucius Malfoy      -> Slytherin
  Tom Riddle         -> Slytherin
  Luna Lovegood      -> Ravenclaw
  Molly Weasley      -> Gryffindor
  Dobby              -> Gryffindor
  Gilderoy Lockhart   -> Ravenclaw
  Seamus Finnigan    -> Gryffindor
```

Figura 4: Sección “Classify”, datos obtenidos del dataset.

Seguidamente, en la figura 5, obtenemos el resultado de Correcta Clasificación de Instancias, que en este caso es de un 100 %.

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	75	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	75		
Ignored Class Unknown Instances	91		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	0,663	0,254	Gryffind
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Slytheri
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Ravenclo
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Hufflepu
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Beauxbat
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Durmstra
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	0,861	0,692	

Figura 4: Sección “Classify”, porcentaje obtenido de Correcta Clasificación de Instancias.

Finalmente en la figura 5, en los resultados de la Matriz de Confusión, en donde dice que “a” son los personajes pertenecientes a Gryffindor obteniendo 31 registros, “b” son de Slytherin obteniendo 20 registros, “c” son de Ravenclaw obteniendo 12 registros, “d” son de Hufflepuff obteniendo 8 registros, “e” son de Beauxbatons obteniendo 2 registros y “f” son de Durmstrang obteniendo 2 registros, debemos tener en cuenta que la diagonal principal siempre debe ser mayor a los demás registros de cada fila para saber que es totalmente confiable.

```

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
31  0  0  0  0  0 | a = Gryffindor
 0 20  0  0  0  0 | b = Slytherin
 0  0 12  0  0  0 | c = Ravenclaw
 0  0  0  8  0  0 | d = Hufflepuff
 0  0  0  0  2  0 | e = Beauxbatons Academy of Magic
 0  0  0  0  0  2 | f = Durmstrang Institute

```

Figura 5: Sección “Classify”, matriz de confusión.

Cluster:

En la figura 6 nos muestra la sección de Cluster, aquí permite seleccionar uno de los Cluster que tenemos en las opciones y seleccionar la opción “Start” la cual permite mostrar el número de iteraciones que realiza.

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'SimpleKMeans' algorithm is chosen. The 'Cluster mode' section has 'Use training set' selected. The 'Clusterer output' section shows the initial starting points, missing values, and final cluster centroids. The 'Result list' shows the execution time for the SimpleKMeans algorithm.

Cluster mode

- ☒ Use training set
- ☐ Supplied test set
- ☐ Percentage split
- ☐ Classes to clusters evaluation
- ☒ Store clusters for visualization

Clusterer output

Initial starting points (random):

Cluster 0: Pixie, Human, Male, Gryffindor, Stag, Yew, 'Dragon Heartstring'

Cluster 1: All, Human, Male, Gryffindor, Stag, Yew, 'Dragon Heartstring'

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	
	(166.0)	0	(1.0)
Character Name	Harry Potter	Harry Potter	Al:
Species	Human	Human	Humai
Gender	Male	Male	Mal
House	Gryffindor	Gryffindor	Gryffindo:
Patronus	Stag	Stag	Stag
Wand (Wood)	Yew	Yew	Yew
Wand (Core)	Dragon Heartstring	Dragon Heartstring	Dragon Heartstring

Time taken to build model (full training data) : 0.01 seconds

==== Model and evaluation on training set ====

Clustered Instances

0	165 (99%)
1	1 (1%)

Figura 6: Área de trabajo de weka, sección “Cluster”

En la figura 7, se muestran los resultados de aplicar el cluster, los resultados obtenidos son 2 iteraciones.

```
Clusterer output

=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning
Relation:    Characters-weka.filters.unsupervised.attribute.Remove-R1
Instances:   166
Attributes:  7
              Character Name
              Species
              Gender
              House
              Patronus
              Wand (Wood)
              Wand (Core)
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 2
Within cluster sum of squared errors: 318.0
```

Figura 7: Sección "Cluster", número de iteraciones.

Como se puede apreciar en la figura 8, obtenemos que el primer cluster que realizó la prueba con todos los registros, el segundo cluster realizó la prueba con 165 registros y el último cluster realizó la prueba con 1 registro.

En el primer cluster, da a conocer que el nombre predominante es Harry Potter, especie humana, género masculino, casa de Gryffindor, etc.

```

Clusterer output
Initial starting points (random):

Cluster 0: Pixie,Human,Male,Gryffindor,Stag,Yew,'Dragon Heartstring'
Cluster 1: All,Human,Male,Gryffindor,Stag,Yew,'Dragon Heartstring'

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute                                Full Data                                Cluster#
                                           (166.0)                                (165.0)                                (1.0)
=====
Character Name                          Harry Potter                          Harry Potter                          All
Species                                Human                                Human                                Human
Gender                                  Male                                  Male                                  Male
House                                  Gryffindor                          Gryffindor                          Gryffindor
Patronus                               Stag                                  Stag                                  Stag
Wand (Wood)                            Yew                                  Yew                                  Yew
Wand (Core)                            Dragon Heartstring Dragon Heartstring Dragon Heartstring

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      165 ( 99%)
1       1 ( 1%)

```

Figura 8: Sección "Cluster", resultados de los Cluster.

Associate:

En la figura 9, nos muestra la información oculta del dataset, que no se visualiza a simple vista, también nos muestra los patrones que hay en el dataset.

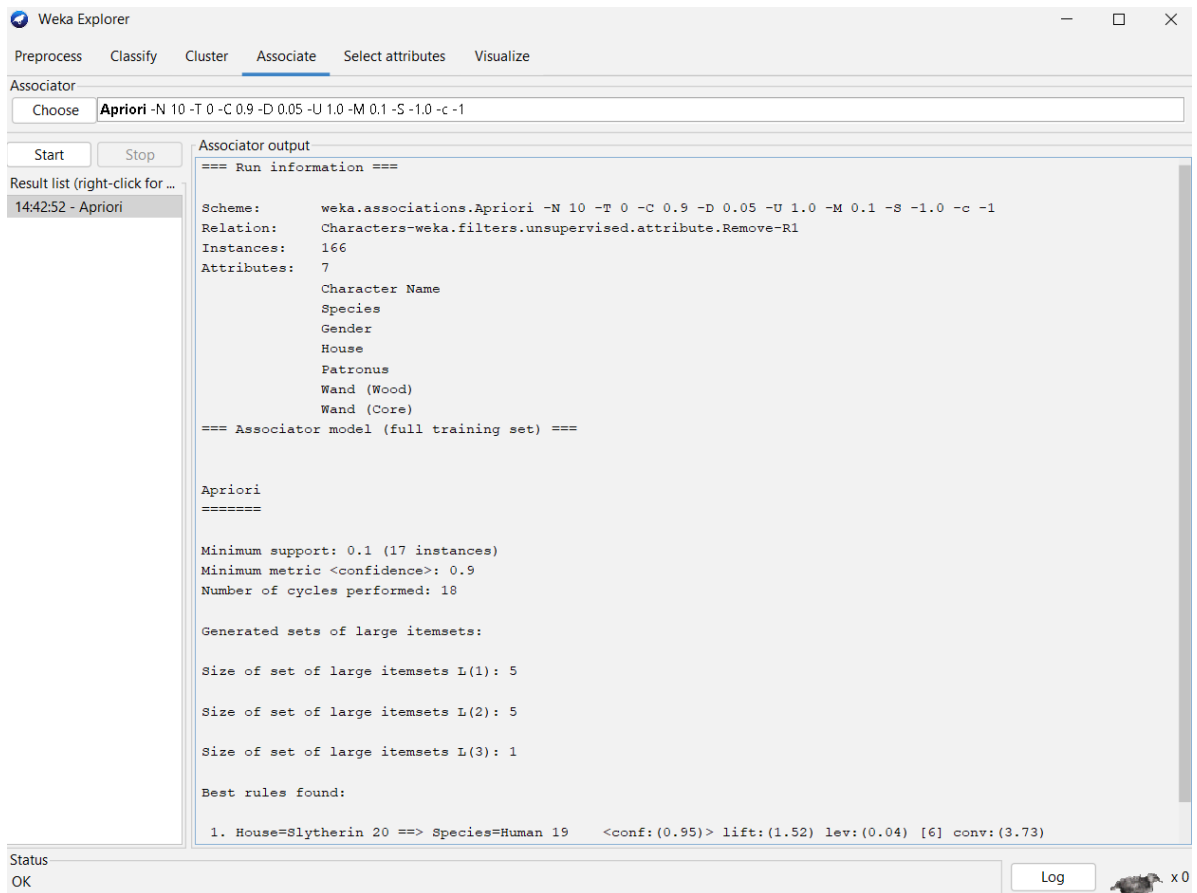


Figura 9: Área de trabajo de weka, sección “Associate”

Select Attributes:

En la figura 10, esta sección nos muestra, de qué manera realizamos la selección de atributos para ver las variables que tiene importancia en la minería de datos, en nuestro caso se muestra que las variables de importancia en nuestro dataset son Species, gender, house.

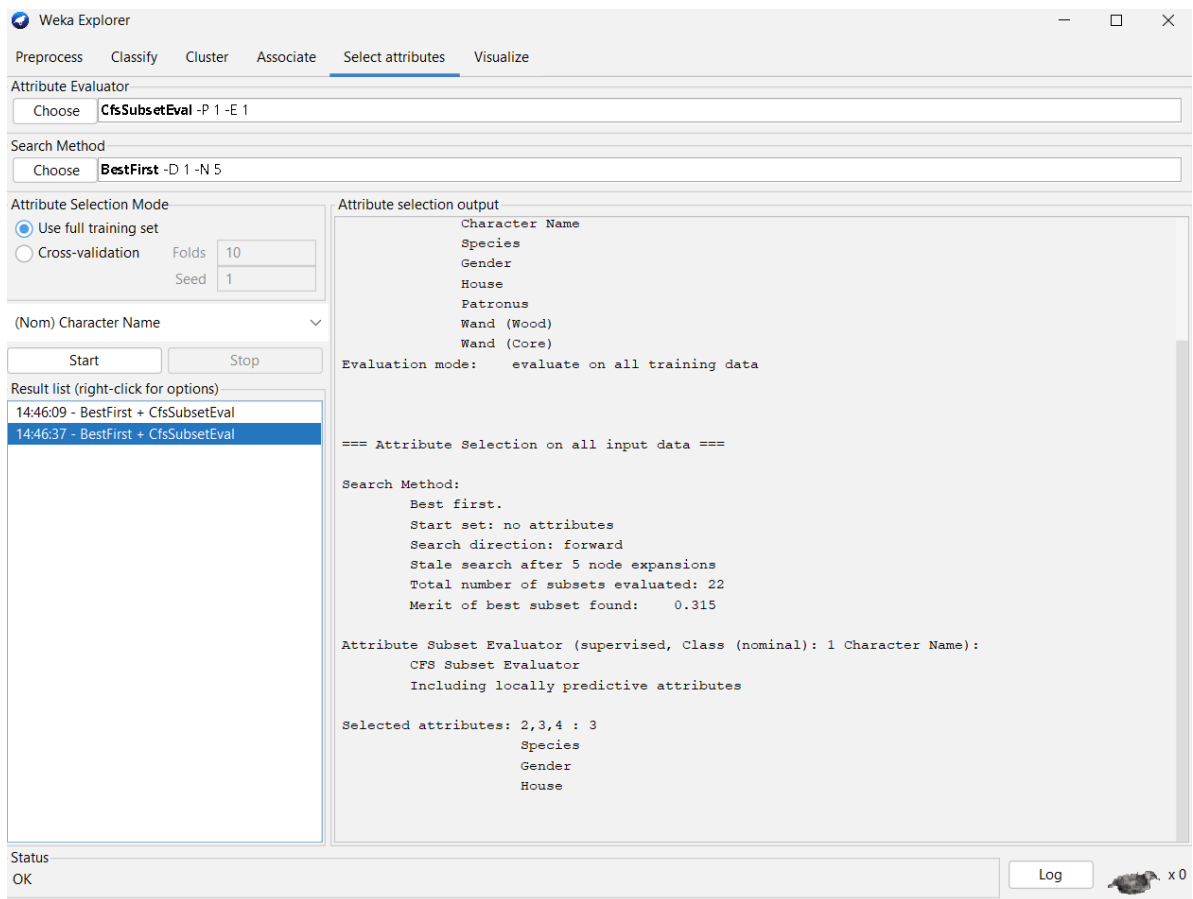


Figura 10: Área de trabajo de weka, sección “Select attributes”

Visualize:

Finalmente en la figura 11, en la última sección se visualiza una matriz de relación con los distintos datos del data set

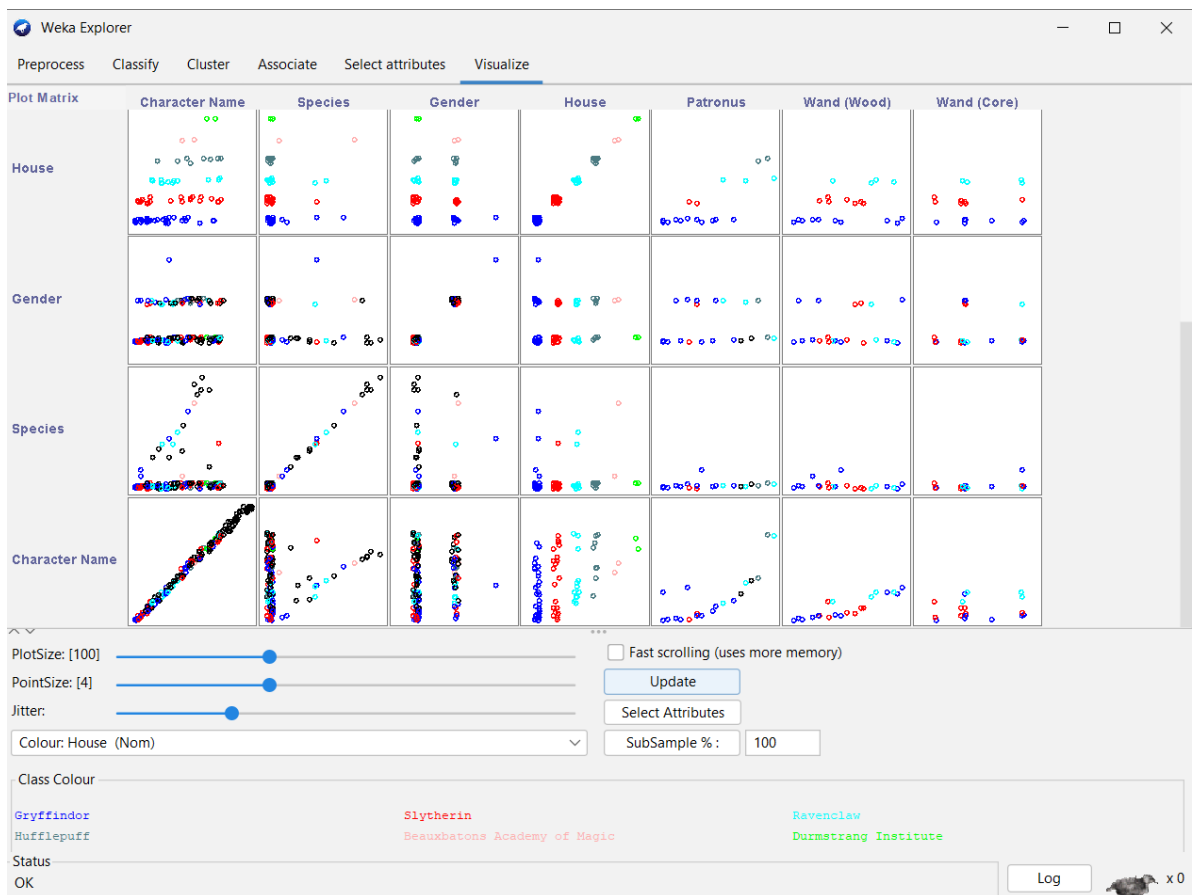


Figura 11: Área de trabajo de weka, sección "Visualize"