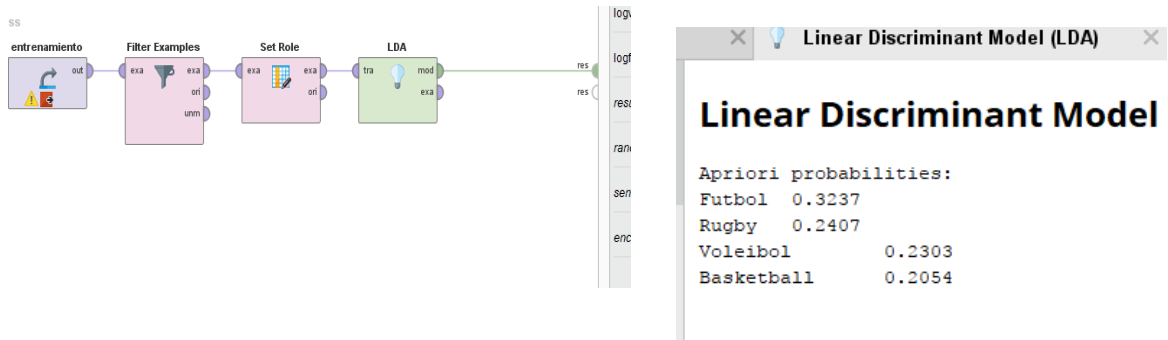


## Ejercicio 1 - Preparación de Datos

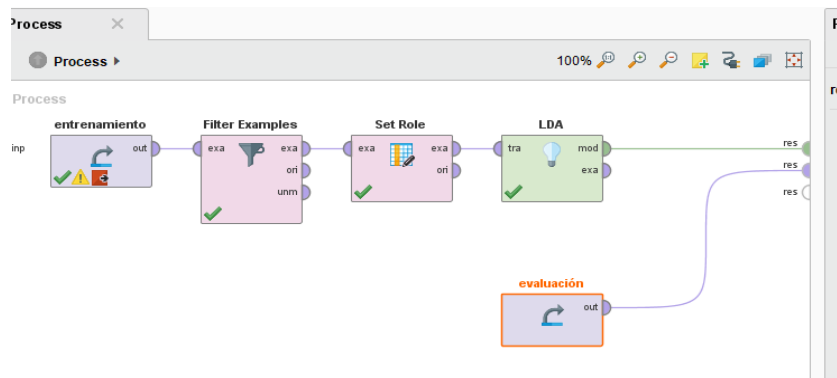
1. Importa en RM el dataset de entrenamiento ("*sport-training.csv*").
  - a) Agrega el dataset a un nuevo proceso, y renombra el operador "retrieve" del dataset a ""entrenamiento".
  - b) Sabemos, de la descripción, que habrá que hacer algo de preparación, en particular en el atributo "**CapacidadDecision**". Analizar los valores incorrectos ( $< 3$  o  $> 100$ ). En principio, se sugiere eliminar los ejemplos si son pocos.... (utilizar el operador "filter examples")
  - c) Setea este atributo como "label" utilizando un operador "set role"
  - d) Analiza las distribuciones, outliers y otras características de los demás predictores. Para cada uno, describe los hallazgos

## Ejercicio 2 - Modelado

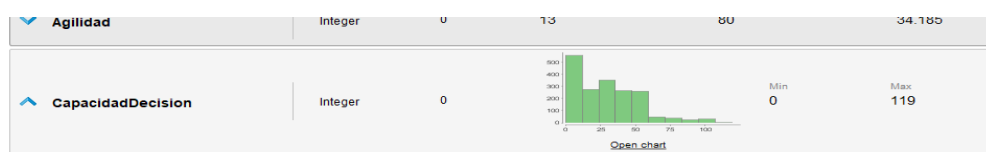
1. Agregar un operador “LDA”, conectar la salida del “set role” a su entrada “tra” y su salida “mod” a un puerto de salida
2. Ejecuta el modelo: observa ahora las probabilidades “a priori” calculadas

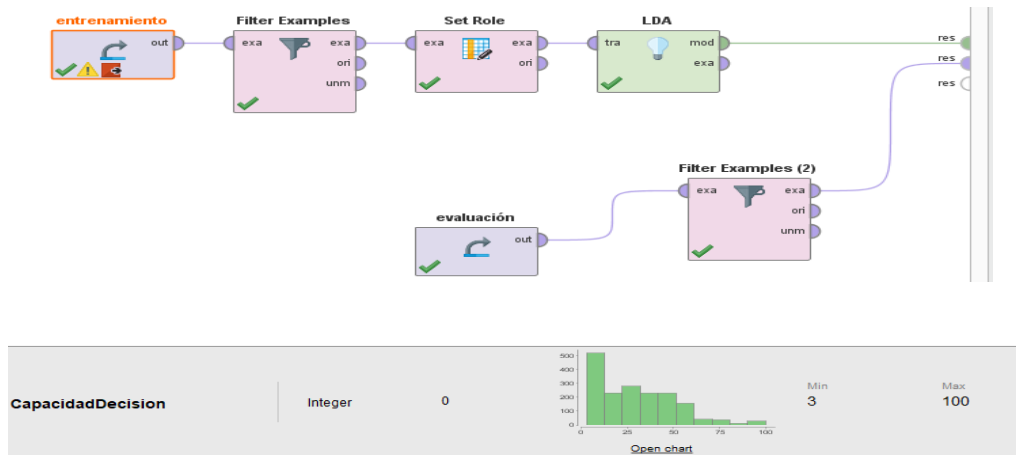


3. Estas probabilidades suman 1.
4. ¿Cómo se calculan? Demostrarlo para las 4 clases.
5. Estas probabilidades, junto con los valores para cada atributo, se utilizarán para predecir la clasificación de “DeportePrimario” para cada uno de los clientes actuales de la academia del Maestro, que están representados en el dataset “*sport-scoring.csv*”.
6. Agrega el dataset “*sport-scoring.csv*” al modelo, y conéctalo directo a la salida.



7. Ejecuta el modelo nuevamente. Ahora RM mostrará un tab adicional con los datos del dataset “*sport-scoring*”.
  - a. Verifica el tipo de datos de los atributos sea el correcto.
  - b. Analiza los rangos de los mismos en comparación con los del dataset de entrenamiento
  - c. Observa que también hay incongruencias en los valores del atributo “CapacidadDecision”. Resuelve este problema de la misma forma que se hizo para el dataset de entrenamiento.





- d. De 1841 ejemplos, nos quedan 1767 después de los filtros.
- e. Completa el proceso de importación y renombrar el operador “retrieve” a “evaluación”
8. Corre el modelo y comparar los rangos de los atributos entre los datasets de entrenamiento y evaluación.
  - a. ¿cómo son, comparativamente, estos rangos?
  - b. ¿están todos los atributos de los ejemplos de evaluación / predicción en los rangos de los atributos del dataset de entrenamiento?
    - i. ¿por qué tenemos que verificar esto?
  - c. ¿hay más tareas de preparación previa de los datos para hacer?
9. **agrega** un operador “apply model”, y conectar a sus entradas:
  - a. a “mod” conectar la salida del operador “LDA”
  - b. a “unl” el dataset “scoring”
10. **verifica** que los puertos “lab” y “mod” estén conectados a las salidas “res” del proceso

