

Unidad Temática 4 - Algoritmos no Lineales

Trabajo de Aplicación 3 Árboles de Decisión

Equipo 5

Ejercicio 1

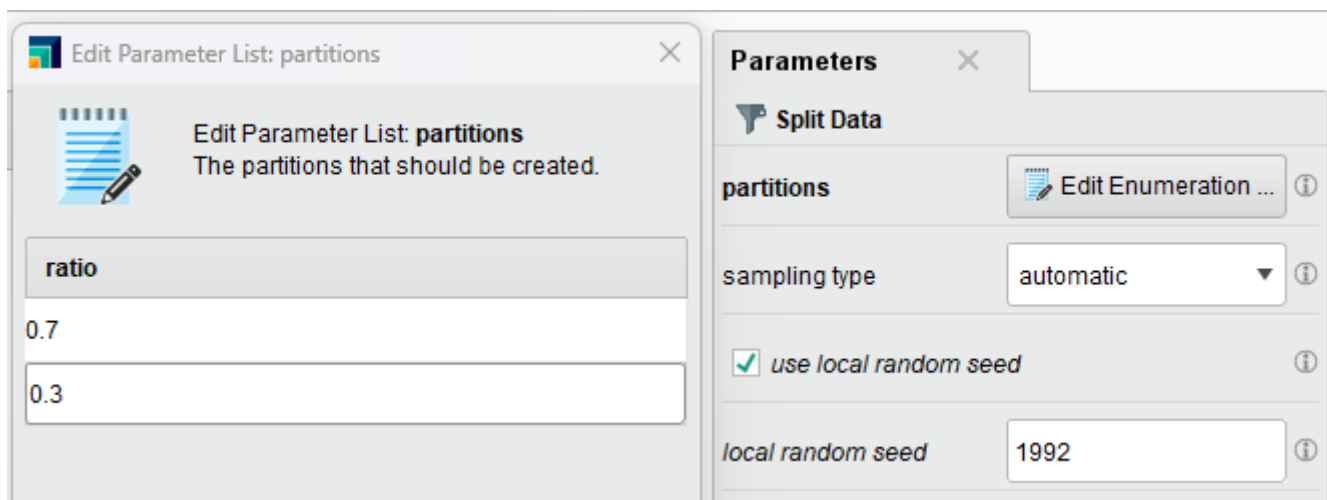
En este caso trabajaremos con el dataset de "eReader_training" que contiene información sobre los usuarios de un servicio de lectura de libros electrónicos. El objetivo es predecir si un usuario es o no un usuario Premium.

El dataset contiene 661 registros y 11 atributos.

1. Carga de datos

Se introdujeron los datos en el RapidMiner, se observó que el dataset no contiene valores perdidos. Además que no se aprecian valores atípicos o nulos.

2. Split en Training y Prueba



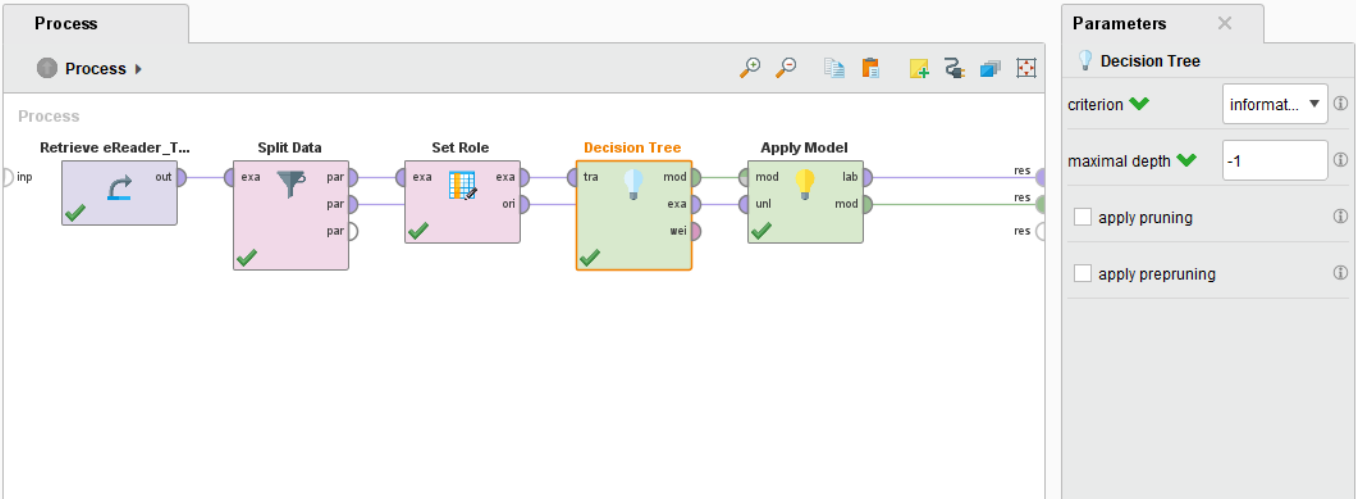
Se dividió el dataset en dos conjuntos, uno de entrenamiento y otro de prueba, con 70% y 30% respectivamente.

3. Construcción del modelo AD

Al ajustar el AD con los siguientes valores:

- Criterio: Information Gain
- Profundidad: -1
- No pruning
- No prepruning

Se obtuvo el siguiente árbol:



Se observa que el árbol se vuelve muy complejo, por tener tantas ramas y nodos.

4. Evaluación del modelo

Table View Plot View

accuracy: 48.48%

	true MayoriaTardia	true Innovador	true AdoptanteTempra...	true MayoriaTemprana	class precision
pred. MayoriaTardia	33	3	7	13	58.93%
pred. Innovador	6	9	7	6	32.14%
pred. AdoptanteTempr...	8	13	32	15	47.06%
pred. MayoriaTemprana	5	4	15	22	47.83%
class recall	63.46%	31.03%	52.46%	39.29%	

Una vez que se setea el prepruning se obtienen diferentes valores de performance para el modelo.

Resultados con prepruning:

ion Tree) PerformanceVector (Performance)

Table View Plot View

accuracy: 56.57%

	true MayoriaTardia	true Innovador	true AdoptanteTempra...	true MayoriaTemprana	class precision
pred. MayoriaTardia	43	5	6	13	64.18%
pred. Innovador	1	12	7	6	46.15%
pred. AdoptanteTempr...	4	11	33	13	54.10%
pred. MayoriaTemprana	4	1	15	24	54.55%
class recall	82.69%	41.38%	54.10%	42.86%	

PerformanceVector (Performance (2))

Tree (Decision Tree (2))

☒ Table View
 ☐ Plot View

accuracy: 50.50% +/- 10.90% (micro average: 50.51%)

	true MayoriaTardia	true Innovador	true AdoptanteTempra...	true MayoriaTemprana	class precision
pred. MayoriaTardia	41	3	9	16	59.42%
pred. Innovador	1	14	15	5	40.00%
pred. AdoptanteTempra...	4	11	24	14	45.28%
pred. MayoriaTemprana	6	1	13	21	51.22%
class recall	78.85%	48.28%	39.34%	37.50%	

TRAINING

#	Accuracy	Error	Method
1	52.25	+ - 4.85	Standar no pruning o prepruning
2	52.90	+ - 5.50	Prepruning
3	53.54	+ - 6.69	Prepruning y pruning
4	52.04	+ - 5.01	Pruning- confidence 0.1
5	52.47	+ - 4.57	Pruning- confidence 0.5

TEST

#	Accuracy	Error	Method
1	48.08	+ - 9.24	Standar no pruning o prepruning
2	47.05	+ - 8.38	Prepruning
3	47.05	+ - 8.38	Prepruning y pruning
4	48.08	+ - 9.24	Pruning- confidence 0.1
5	48.02	+ - 9.24	Pruning- confidence 0.5

Se observa que el mejor resultado se obtiene con el prepruning y pruning, para el caso de Training. Y en caso de Test es igual el Standar con pruning de confidencia 0.1.