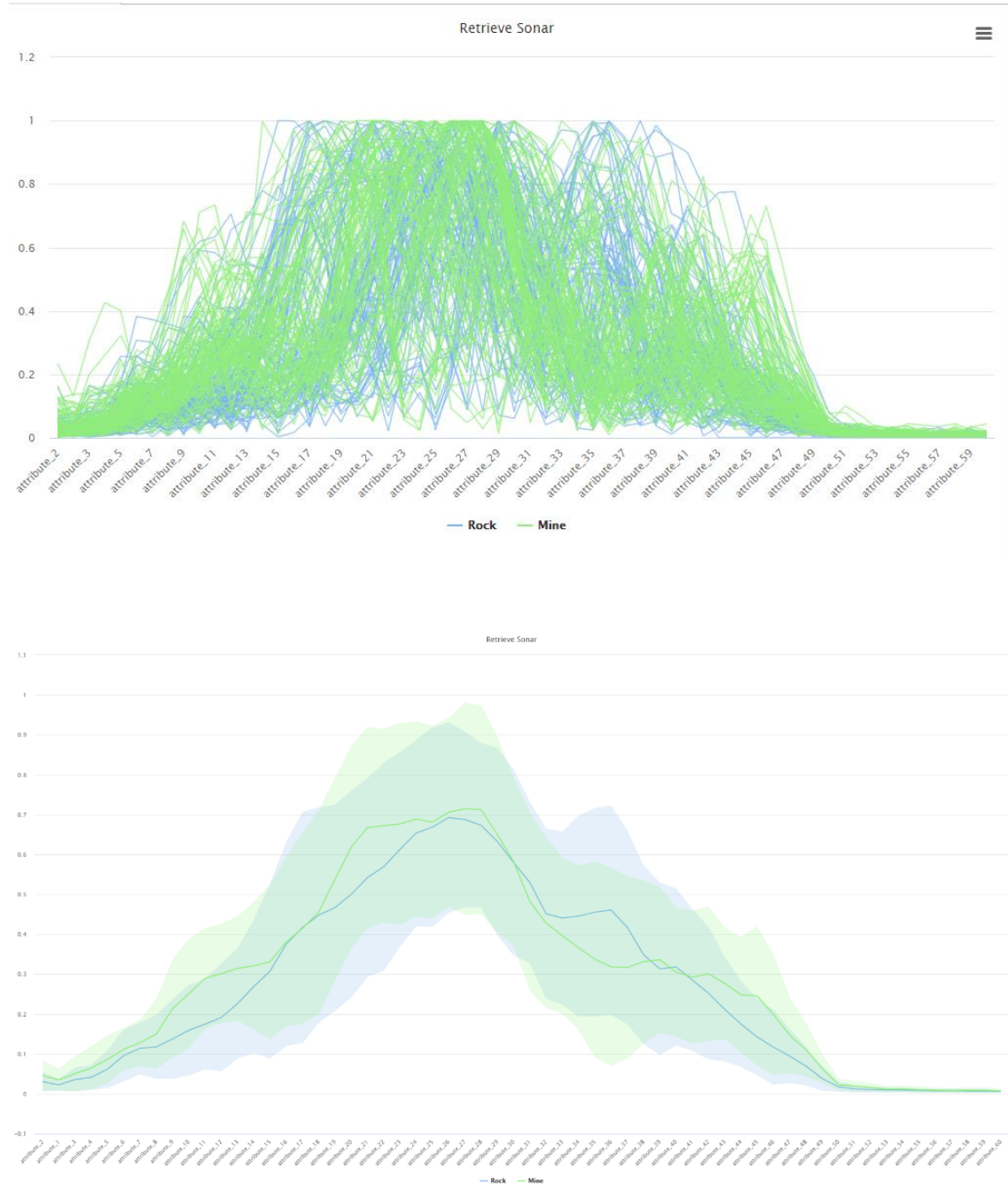


## UNIDAD TEMÁTICA 4: Selección de Atributos Trabajo de Aplicación 10

Imagen de gráfico



Los cuatro atributos en los cuales las clases se diferencian más son el atributo 11, atributo 20, atributo 35 y atributo 45

# Ejercicio 1

Se obtiene la salida del modelo:

## SimpleDistribution

Distribution model for label attribute class

Class Rock (0.466)  
60 distributions

Class Mine (0.534)  
60 distributions

Y para cada atributo numérico la media y desviación estándar calculadas por el modelo:

Row No.	class	attribute_1	attribute_2	attribute_3	attribute_4	attribute_5	attribute_6	attribute_7	attribute_8	attribute_9	attribute_10
1	Rock	0.020	0.037	0.043	0.021	0.095	0.099	0.154	0.160	0.311	0.211
2	Rock	0.045	0.052	0.084	0.069	0.118	0.258	0.216	0.348	0.334	0.287
3	Rock	0.026	0.058	0.110	0.108	0.097	0.228	0.243	0.377	0.560	0.619
4	Rock	0.010	0.017	0.062	0.021	0.021	0.037	0.110	0.128	0.060	0.126
5	Rock	0.076	0.067	0.048	0.039	0.059	0.065	0.121	0.247	0.356	0.446
6	Rock	0.029	0.045	0.028	0.017	0.038	0.099	0.120	0.183	0.210	0.304
7	Rock	0.032	0.096	0.132	0.141	0.167	0.171	0.073	0.140	0.208	0.351
8	Rock	0.052	0.055	0.084	0.032	0.116	0.092	0.103	0.061	0.146	0.284
9	Rock	0.022	0.037	0.048	0.048	0.065	0.059	0.075	0.010	0.068	0.149
10	Rock	0.016	0.017	0.035	0.007	0.019	0.067	0.106	0.070	0.096	0.025
11	Rock	0.004	0.006	0.015	0.034	0.031	0.028	0.040	0.027	0.032	0.045
12	Rock	0.012	0.031	0.017	0.031	0.036	0.010	0.018	0.058	0.112	0.084
13	Rock	0.008	0.009	0.005	0.025	0.034	0.055	0.053	0.096	0.101	0.124
14	Rock	0.009	0.006	0.025	0.049	0.120	0.159	0.139	0.099	0.096	0.190

Y a su vez las predicciones realizadas en comparación al dato observado:

Row No.	class	prediction(class)	confidence(class)	confidence(class)	attribute_1	attribute_2	attribute_3	attribute_4	attribute_5	attribute_6	attribute_7
1	Rock	Mine	1.000	0.000	0.026	0.058	0.110	0.108	0.097	0.228	0.243
2	Rock	Rock	0.001	0.999	0.022	0.037	0.048	0.048	0.065	0.059	0.075
3	Rock	Rock	0.000	1.000	0.012	0.031	0.017	0.031	0.036	0.010	0.018
4	Rock	Rock	0.004	0.996	0.012	0.043	0.060	0.045	0.060	0.035	0.053
5	Rock	Mine	0.997	0.003	0.066	0.058	0.084	0.037	0.046	0.077	0.077
6	Rock	Rock	0.000	1.000	0.020	0.003	0.014	0.006	0.013	0.015	0.054
7	Rock	Rock	0.002	0.998	0.019	0.031	0.020	0.062	0.008	0.079	0.144
8	Rock	Rock	0.014	0.986	0.031	0.049	0.069	0.083	0.008	0.020	0.098
9	Rock	Rock	0.000	1.000	0.012	0.002	0.020	0.021	0.018	0.049	0.003
10	Rock	Rock	0.000	1.000	0.009	0.027	0.022	0.034	0.030	0.117	0.145
11	Rock	Mine	1	0	0.026	0.045	0.039	0.024	0.132	0.132	0.161
12	Rock	Mine	0.721	0.279	0.037	0.028	0.023	0.022	0.018	0.073	0.084
13	Rock	Rock	0.000	1.000	0.019	0.004	0.064	0.045	0.033	0.069	0.090
14	Rock	Rock	0.004	0.996	0.012	0.058	0.062	0.060	0.140	0.188	0.142

Se observa que en algunos casos se clasifica incorrectamente.

Se observa una precisión de 67,83%

accuracy: 67.83% +/- 6.92% (micro average: 67.79%)

	true Rock	true Mine	class precision
pred. Rock	76	46	62.30%
pred. Mine	21	65	75.58%
class recall	78.35%	58.56%	

## Ejercicio 2 - Forward Selection

accuracy: 77.43% +/- 2.95% (micro average: 77.40%)

	true Rock	true Mine	class precision
pred. Rock	67	17	79.76%
pred. Mine	30	94	75.81%
class recall	69.07%	84.68%	

Al ejecutar el modelo utilizando Forward Selection la precisión del mismo aumenta a 77,43%.

Se seleccionan los atributos 12,17 15 y 18.

El atributo 12 está cerca del identificado como diferente, pero los demás no parecieran tener diferencias significativas en la gráfica. De hecho sus valores se pegan en el gráfico.

## Ejercicio 3 – Backward Elimination

El modelo presenta una mejora con respecto al inicial pero continúa siendo inferior al observado en Forward Selection:

accuracy: 73.12% +/- 6.22% (micro average: 73.08%)

	true Rock	true Mine	class precision
pred. Rock	82	41	66.67%
pred. Mine	15	70	82.35%
class recall	84.54%	63.06%	

Para un parámetro de detención “with no decrease” obtiene una precisión de 73,12% dejando fuera únicamente los atributos 3, 14, 20, 36, 47, 48, 52,59. Por tanto demuestra tener una peor performance dado que con una mayor cantidad de atributos y por ende un modelo más complejo para lograr niveles menores de precisión.

Ajustándole los parámetros a que el modelo pare con decremento significativo de la performance se obtiene el mismo porcentaje de precisión sin embargo solicitando un descenso >0,05 se obtiene un incremento en la performance a 75,04%

accuracy: 75.04% +/- 6.77% (micro average: 75.00%)

	true Rock	true Mine	class precision
pred. Rock	71	26	73.20%
pred. Mine	26	85	76.58%
class recall	73.20%	76.58%	

En este último modelo se seleccionan únicamente los atributos 12, 21 y 49 pareciéndose más al ejemplo de Backward Selection. Igualmente obtiene una performance menor al primer modelo.

## Ejercicio 4 - Selección de atributos por algoritmo evolutivo

Parámetros del nodo Optimize Selection:

**use exact number of attributes:** como su nombre lo indica si se selecciona sólo se testean las combinaciones utilizando un número exacto de atributos, los que pueden ser asignados,

**restrict maximum:** se puede restringir el máximo y mínimo de atributos a considerar

**population size:** determina el número de individuos (atributos) por generación

**maximum number of generations:** determina el máximo número de generaciones luego de utilizado el algoritmo.

**use early stopping:** si no se habilita siempre se alcance el máximo número de generaciones, de lo contrario habilita parar antes. Si se selecciona debe determinarse el criterio, es decir después de cuántas generaciones (n) sin mejoras parar.

**normalize weight:** indica si los pesos finales deben ser normalizados.

**use local random seed:** indica si se utiliza una semilla fija para la aleatorización y especifica el valor de dicha semilla en local random seed.

Se puede seleccionar a su vez obtener un gráfico de la precisión del modelo en relación a la cantidad de atributos.

A su vez permite seleccionar el esquema de selección (tournament entre ellos y asignar valores a los parámetros según modelo seleccionado).

accuracy: 77.44% +/- 5.82% (micro average: 77.40%)

	true Rock	true Mine	class precision
pred. Rock	84	34	71.19%
pred. Mine	13	77	85.56%
class recall	86.60%	69.37%	

Se obtiene una precisión similar al primer modelo de feature selection, de 77,44%, 0,1pp superior.

Selecciona un mayor número de atributos que el primer modelo, los atributos seleccionados son 1, 4, 6, 11, 12, 15, 16, 17, 21, 24, 26, 27, 30, 31, 33, 34, 36, 38, 39, 40, 42, 43, 45, 48, 49, 54, 55, 57.

El único que no selecciona en comparación al primer modelo del ejercicio es el atributo 18.

Se asemeja a los atributos seleccionados en el ejercicio 3 pero con algunas diferencias y mejor performance. Selecciona alguno (11, 45) de los atributos identificados como diferentes en la gráfica inicial aunque otros no (20, 35) pero sí selecciona aquellos que están pegados que también presentan valores parecidos (21, 24) y (34, 36).