



Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Estado de México

“Análisis y Reporte sobre el desempeño del modelo”

Eduardo Acosta Hernández

A01375206

Ingeniería en Tecnologías Computacionales

Inteligencia artificial avanzada para la ciencia de datos

I

Grupo 101

Jorge Adolfo Ramírez Uresti

13 de agosto de 2022

La entrega elegida para este módulo es la entrega 2 que consta de la implementación de un algoritmo con el uso de un framework o una librería, el algoritmo implementado es el Decision Tree Classifier de la librería sklearn.

El algoritmo es capaz de clasificar múltiples clases en un dataset, requiere de un input bidimensional de los parámetros independientes y una variable target la cual contiene las diferentes clases. Este algoritmo subdivide el dataset en arboles de decisión, tomando en cuenta las características proporcionadas. Cabe mencionar que en este tipo de algoritmos es muy importante lograr identificar el número optimo de árboles para generar, debido a que es común que cuando se generan muchos subárboles se presente overfitting.

En cuanto al dataset este cuenta con una variable Species, la cual contiene las clases de pescado, de igual manera cuenta con variables que nos hablan de las dimensiones físicas del pescado como, alto, ancho, longitud, peso, etc. Con el algoritmo se buscará llegar a un modelo que nos permita predecir la especie de un pescado basándonos en sus dimensiones.

Para realizar lo anterior dividimos el dataset en una parte de entrenamiento y otra de pruebas, donde se utilizaron el 80% de los datos para entrenar y 20% para probar. Dicha división del dataset se realizó con la dependencia model selection de la librería sklearn.

```
PS C:\Users\laloa\Documents\0ITC\7.SeptimoSemestre\Reto\entrega
Documents\0ITC\7.SeptimoSemestre\Reto\entregasmodulo2\entrega2f
Species Weight Length1 Length2 Length3 Height Width
0 Bream 242.0 23.2 25.4 30.0 11.5200 4.0200
1 Bream 290.0 24.0 26.3 31.2 12.4800 4.3056
2 Bream 340.0 23.9 26.5 31.1 12.3778 4.6961
3 Bream 363.0 26.3 29.0 33.5 12.7300 4.4555
4 Bream 430.0 26.5 29.0 34.0 12.4440 5.1340
Exactitud del modelo 0.71875
Exactitud del modelo(train) 1.0
```

Siguiendo los hiperparametros en default y sin realizar tratamiento de los datos se obtuvo un accuracy del 0.71875, Lo cual nos dice que nuestro modelo es aceptable pero que aún tiene aspectos por mejorar.

Para calcular el bias y varianza de nuestro modelo usaremos [bias_variance_decomp](#) de la librería [mlxtend.evaluate](#)

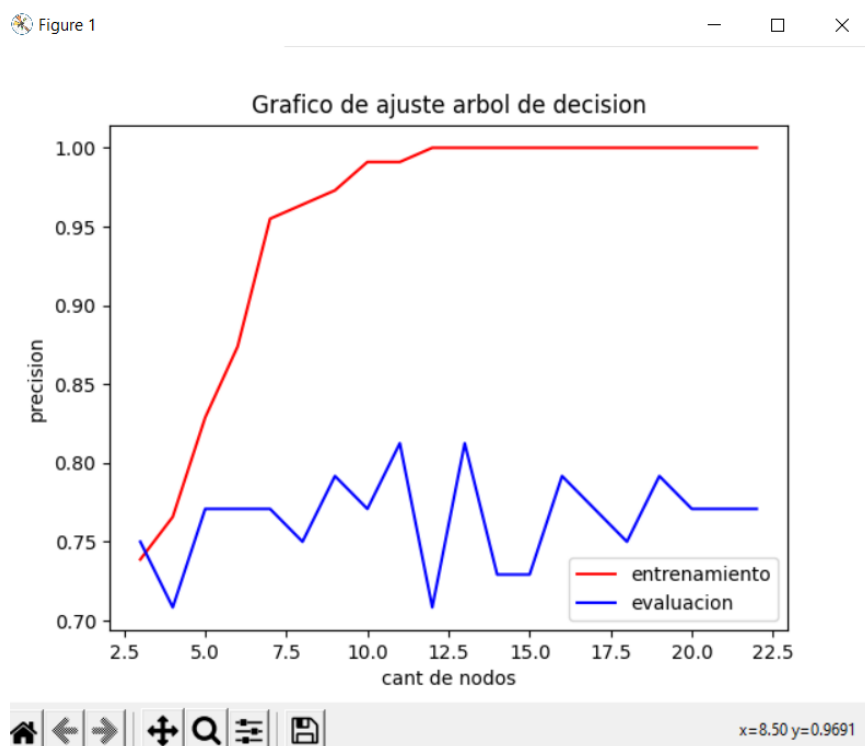
Para realizar una mejora en el modelo se ajustó la distribución de los datos para test y train siendo ahora 30% para test y 70% para train.

Con la modificación anterior se realizó una mejora considerable en la exactitud del modelo, siendo ahora de 0.7916, esto nos puede estar hablando de que no necesitamos darle tantos

datos al modelo debido a que la generación de árboles puede llegar a ser redundante y caer en overfitting.

```
PS C:\Users\laloa\Documents\0ITC\7.SextoSemestre\Reto\entregasm
Documents\0ITC\7.SextoSemestre\Reto\entregasmmodulo2\entrega2fr
Species Weight Length1 Length2 Length3 Height Width
0 Bream 242.0 23.2 25.4 30.0 11.5200 4.0200
1 Bream 290.0 24.0 26.3 31.2 12.4800 4.3056
2 Bream 340.0 23.9 26.5 31.1 12.3778 4.6961
3 Bream 363.0 26.3 29.0 33.5 12.7300 4.4555
4 Bream 430.0 26.5 29.0 34.0 12.4440 5.1340
Exactitud del modelo 0.7916666666666666
Exactitud del modelo(train) 1.0
```

A continuación creamos un grafico donde nos dice la precisión del modelo basándonos en la cantidad de nodos que generamos, podemos apreciar que el modelo en cuanto a la relación de evaluación con entrenamiento, cuando el modelo tiene 10 nodos se obtiene una menor distancia entre las líneas lo cual nos habla de que con esta cantidad el modelo puede llegar a tener un punto de optimización



Al realizar el cambio tomando 10 nodos como máximo podemos observar que la exactitud de nuestro modelo aumenta a 0.81 y al determinar un numero máximo de nodos, estamos economizando recursos y evitamos overfitting por el exceso de información que se genera con una cantidad mayor de nodos.

Documents/0ITC/7.SeptimoSemestre/Reto/entregasmodulo2/entrega3.py

	Species	Weight	Length1	Length2	Length3	Height	Width
0	Bream	242.0	23.2	25.4	30.0	11.5200	4.0200
1	Bream	290.0	24.0	26.3	31.2	12.4800	4.3056
2	Bream	340.0	23.9	26.5	31.1	12.3778	4.6961
3	Bream	363.0	26.3	29.0	33.5	12.7300	4.4555
4	Bream	430.0	26.5	29.0	34.0	12.4440	5.1340

Exactitud del modelo 0.8125

Exactitud del modelo(train) 0.9819819819819819