

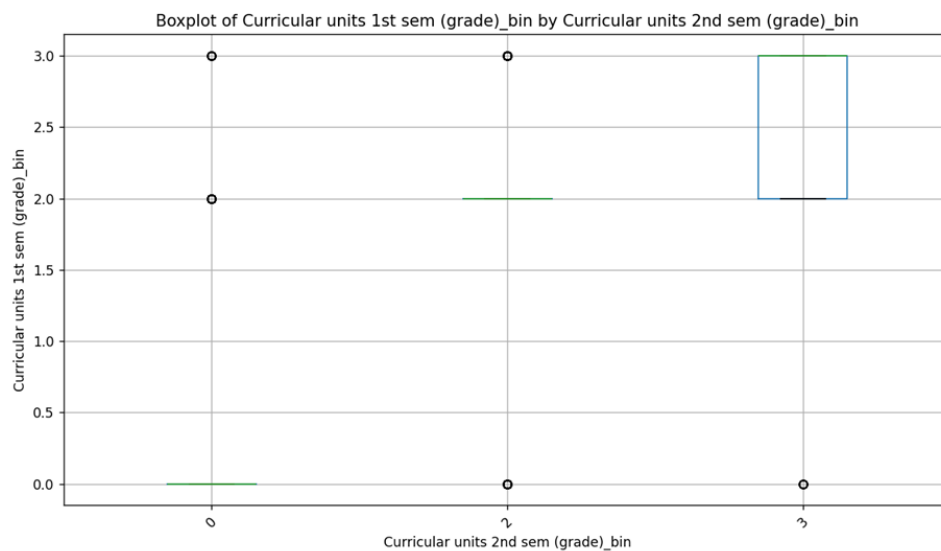
Reporte Proyecto 2

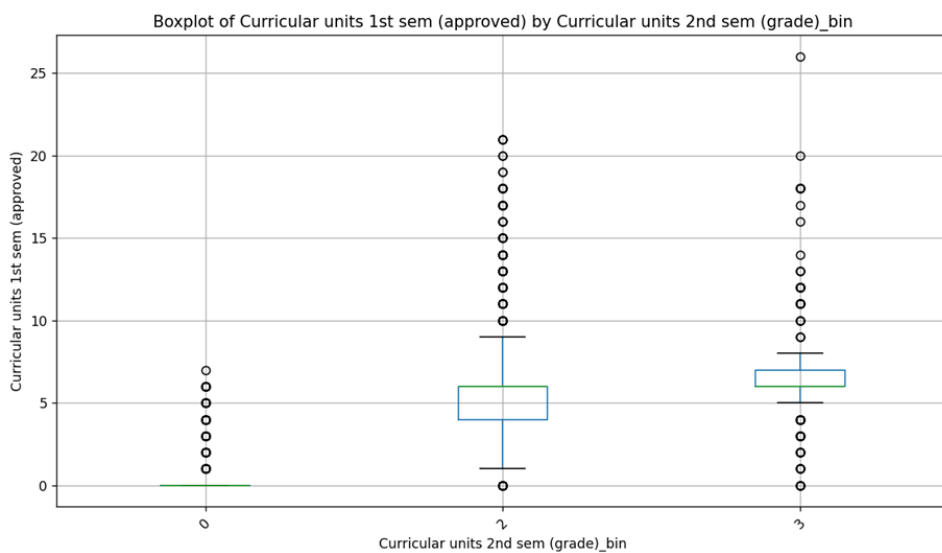
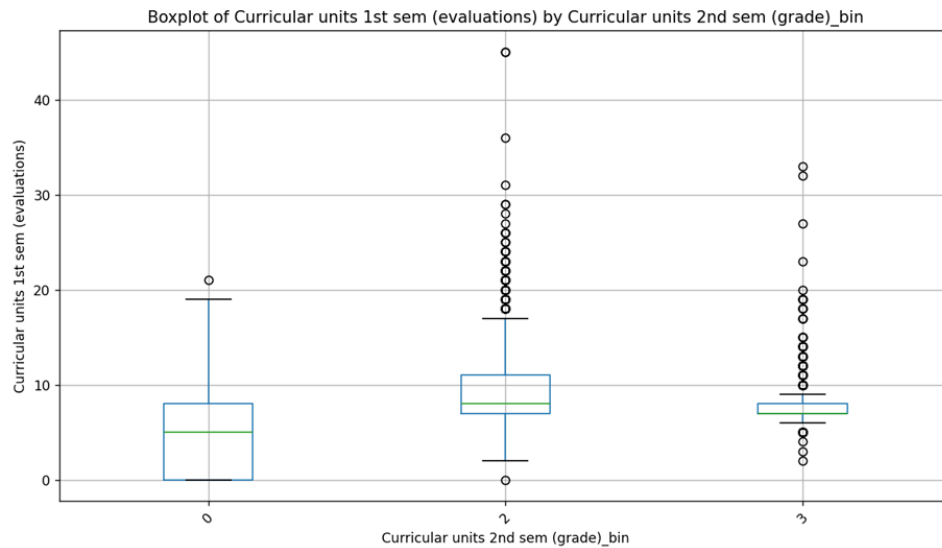
Para empezar en la construcción del modelo se discretizaron las siguientes variables:

	0	1	2	3
'Previous qualification (grade)'	(94.905, 118.75]	(118.75, 142.5]	(142.5, 166.25]	(166.25, 190.0]
'Admission grade'	(94.905, 118.75]	(118.75, 142.5]	(142.5, 166.25]	(166.25, 190.0]
'Unemployment rate'	[(7.591, 9.75]	(9.75, 11.9]	(11.9, 14.05]	(14.05, 16.2]
'Inflation rate'	(-0.804, 0.325]	(0.325, 1.45]	(1.45, 2.575]	(2.575, 3.7]
'GDP'	(-4.068, -2.167]	(-2.167, -0.275]	(-0.275, 1.617]	(1.617, 3.51]
'Curricular units 2nd sem (grade)'	(-0.0186, 4.643]	(4.643, 9.286]	(9.286, 13.929]	(13.929, 18.571]
'Curricular units 1st sem (grade)'	(-0.0189, 4.719]	(4.719, 9.438]	(9.438, 14.156]	(14.156, 18.875]
'Age at enrollment'	(16.947, 30.25]	(30.25, 43.5]	(43.5, 56.75]	(56.75, 70.0]
'Curricular units 1st sem (evaluations)'	(-0.045, 11.25]	(11.25, 22.5]	(22.5, 33.75]	(33.75, 45.0]

Modelación:

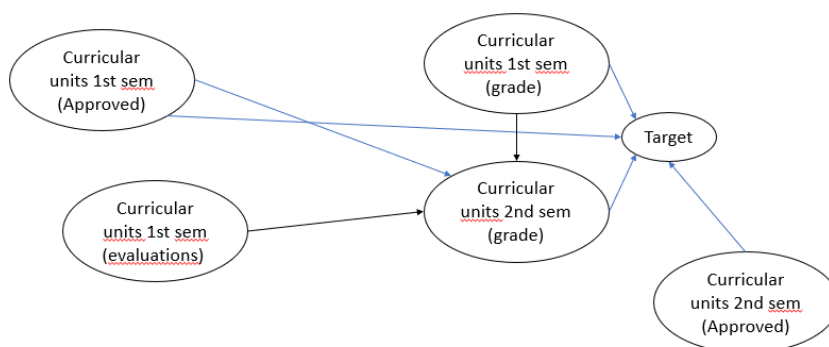
El modelo del proyecto 1 se replanteo para justificar la relación de padres entre variables dando después de un análisis de gráficos de caja adicional al del proyecto 1





Con lo anterior se obtuvo este modelo con un estimador bayesiano basado en una distribución a priori uniforme

Modelo Proyecto 1:

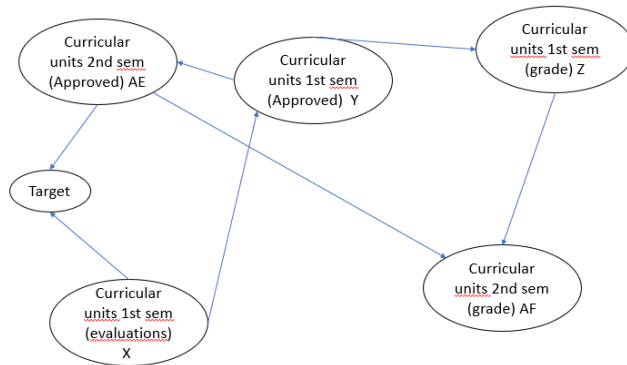


Para estimar la red del modelo 2 se usó el método de scoring teniendo como posibles nodos los mismos de la red del proyecto 1 en el cual se tuvieron los siguientes intentos

1.

[('Z', 'AF'), ('AE', 'AF'), ('AE', 'Target'), ('Y', 'AE'), ('Y', 'Z'), ('X', 'Y'), ('X', 'Target')]

K2 score -
19023.08902
5124696

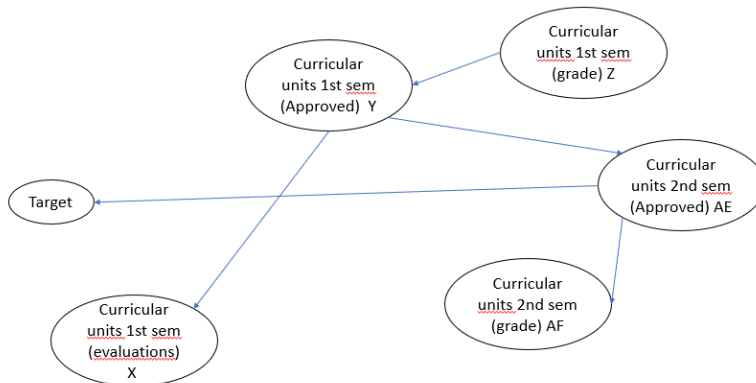


Este se estimó con K2 score, obteniendo –19023.08902 como puntaje

2.

[('Z', 'Y'), ('AE', 'AF'), ('AE', 'Target'), ('Y', 'AE'), ('Y', 'X')]

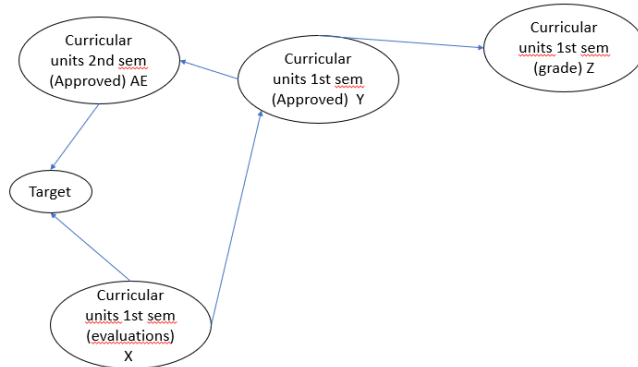
BIC score -
20782.1707
54046416



Este se estimó con Bic score, obteniendo –20782.17075 como puntaje

3.

```
[('AE', 'Target'), ('Y', 'AE'), ('Y', 'Z'), ('X', 'Y'), ('X', 'Target')]
```

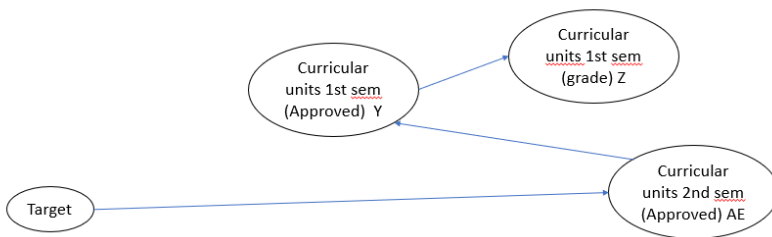


K2 v2 score -
-
17798.07206
6555695

El tercer intento se realizó modificando la red obtenida en el primer intento por k2 score, quitando nodos que no aportaban a la variable objetivo y se obtuvo un puntaje de -17798.0720666

4.

```
[('Target', 'AE'), ('AE', 'Y'), ('Y', 'Z')]
```



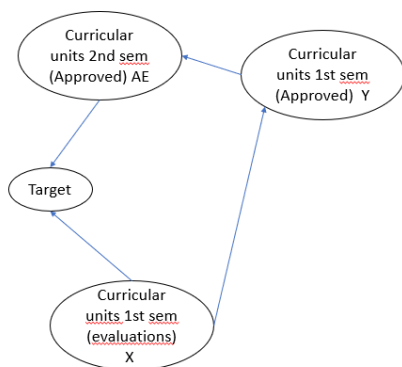
BIC v2 score

17806.6209
6690137

El cuarto intento se realizó modificando la red obtenida en el primer intento por Bic score, quitando nodos que no aportaban a la variable objetivo y se obtuvo un puntaje de -17806.620967

5.

```
[('AE', 'Target'), ('Y', 'AE'), ('X', 'Y'), ('X', 'Target')]
```

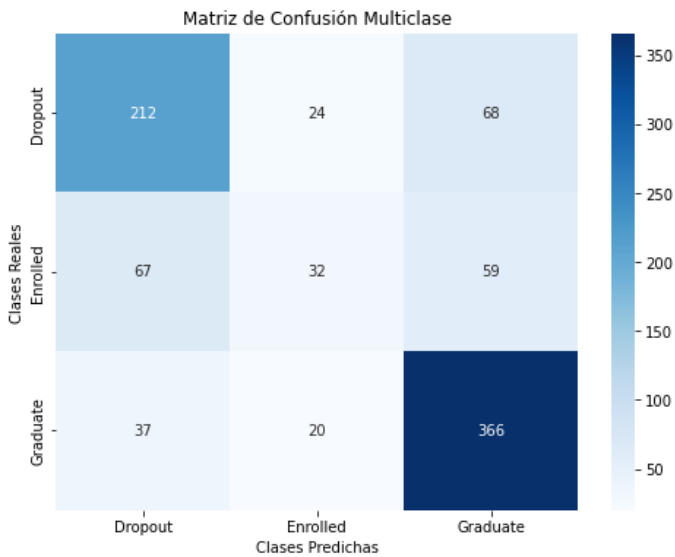


K2 v3 score -
16525.82140
8495816

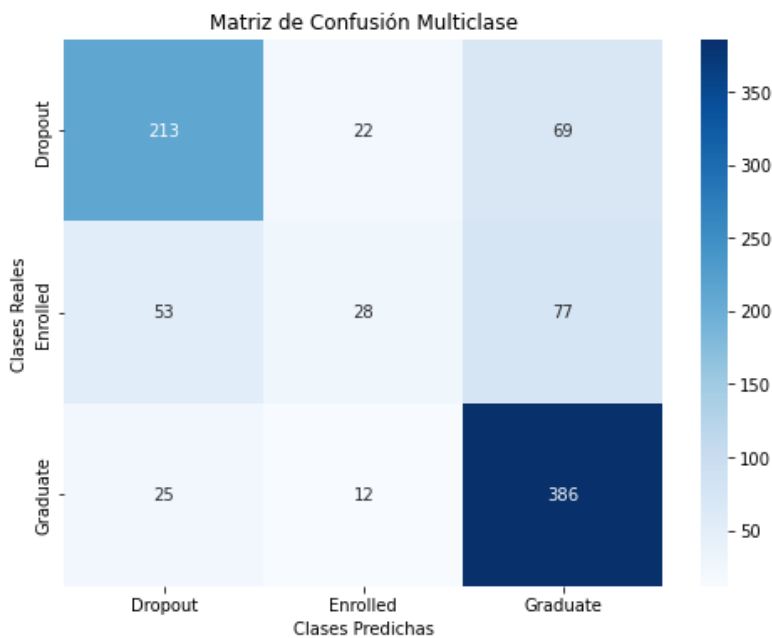
La red definitiva se hizo modificando el tercer intento quitando las variables que no aportaban información dando una red de 4 nodos y 4 arcos con un k2 score de -16525.8214085 la cual se convirtió en red y se sacó sus probabilidades por máxima verosimilitud.

Comparación de redes:

Matriz de confusión modelo proyecto 1:



Matriz de confusión modelo proyecto 2:



	Modelo proyecto 1	Modelo proyecto 2
Exactitud global	68,92%	70,85%
F1-Score Macro	58,55%	59,30%
F1-Score Micro	68,92%	70,85%

Reporte de Clasificación Modelo proyecto 1:

	precision	recall	f1-score	support
Dropout	0.67	0.7	0.68	304
Enrolled	0.42	0.2	0.27	158
Graduate	0.74	0.87	0.8	423
accuracy			0.69	885
macro avg	0.61	0.59	0.59	885
weighted avg	0.66	0.69	0.67	885

Reporte de Clasificación Modelo proyecto 2:

	precision	recall	f1-score	support
Dropout	0.73	0.7	0.72	304
Enrolled	0.45	0.18	0.25	158
Graduate	0.73	0.91	0.81	423
accuracy	0.71	885		
macro avg	0.64	0.6	0.59	885
weighted avg	0.68	0.71	0.68	885

El modelo elegido en el proyecto 2(Modelo 2) tiene una Exactitud Global (Accuracy) del 70.85%, mientras que el modelo elegido en el proyecto 1(modelo 1) tiene una Exactitud Global del 68.92%. Esto significa que el "modelo 2" tiene una tasa de predicciones correctas más alta en comparación con el "modelo 1". Por lo tanto, en términos de Exactitud Global, el "modelo 2" es mejor.

El "modelo 2" también tiene un F1-Score Macro del 59.30%, mientras que el "modelo 1" tiene un F1-Score Macro del 58.55%. El F1-Score Macro es una medida que tiene en cuenta tanto la precisión como el recall en todas las clases. El hecho de que el "modelo 2" tenga un F1-Score Macro más alto indica que tiene un mejor equilibrio entre precisión y recall en todas las clases en comparación con el "modelo 1".

En cuanto al F1-Score Micro, el "modelo 2" tiene un valor del 70.85%, mientras que el "modelo 1" tiene un valor del 68.92%. El F1-Score Micro se calcula considerando todas las predicciones y sus respectivas etiquetas reales en todas las clases. El "modelo 2" tiene un F1-Score Micro más alto, lo que sugiere que tiene un mejor rendimiento en general.

En cuanto a la matriz de confusión, se evidencia que el modelo 2 tienen mayor asertividad en los verdaderos positivos, de las clases "graduado" y "Matriculado", y también en verdaderos negativos. Esto también se evidencia en el reporte de clasificación sobre todo para las clases "Graduado" y "Dropout"

En resumen, el "modelo 2" supera al "modelo 1" en términos de Exactitud Global, F1-Score Macro y F1-Score Micro. Matriz de confusión, lo que indica un mejor rendimiento y mayor confiabilidad en la clasificación a pesar de tener menor cantidad de arcos y nodos.