

Minería de Datos

Caso Insurance

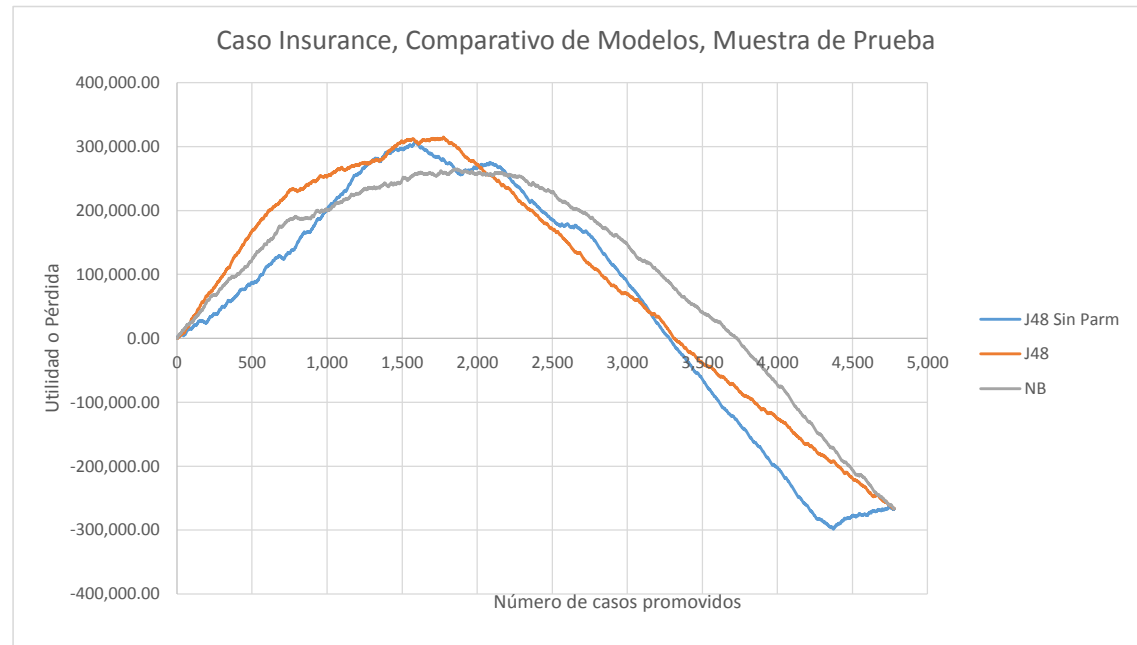
Tarea No 1.- Comparativo de modelos en términos de Curva de Utilidad

Muestra de Prueba (45% de los datos)

Clase	casos	prop	U o P	u_p
0	3,134	0.655785729	400	- 1,253,600.00
1	1,645	0.344214271	600	987,000.00
	4,779			- 266,600.00

Densidades	original	0.344214271
	umbral	0.4

Test			
	J48_SinParam	J48, conf=0.0125 20 casos por nodo	NB
UMP	987,000.00	987,000.00	987000
UC	307,600.00	314,600.00	265,000.00
UC/UMP	31.17%	31.87%	26.85%
CAP	1,586	1,776	1,850
CAP/Muestra	33.19%	37.16%	38.71%
#1's capt	942	1,025	1,005
#0's inc	644	751	845
densCapt	0.593947037	0.57713964	0.543243243
LiftCapt	1.726	1.677	1.578
uProm_CAP	193.95	177.14	143.24
CC	634,400.00	710,400.00	740,000.00
ROI	48.49%	44.28%	35.81%
Score de corte	0.2625	0.391304	0.004353



La utilidad absoluta mayor se obtiene para el modelo de árbol J48 con parámetros de poda de 0.0125 y 20 elementos mínimo en los nodos para poder aplicar la partición, sin embargo el árbol J48 con los parámetros por default (árbol granulado) presenta un mejor Retorno de la Inversión (no financiero), una mejor utilidad promedio y esto se debe a su más reducido número de casos a promover y de ahí su menor Costo e Campaña.

En la gráfica se observa que de ambos árboles el J48 "tuneado" tiene una curva más suave respecto a la curva del modelo con los parámetros sin tocar. Sin embargo la curva del modelo bayesiano ingenuo es más suave.

Otra situación interesante es que las curvas se cruzan en varias ocasiones y ello implica que en función de los casos que se promuevan va a resultar el "mejor" modelo.