

1. Describa brevemente cuál es el fundamento de las técnicas de balanceo de clases usadas en machine learning.

Los algoritmos de Machine Learning suponen que las clases están balanceadas, en caso de desbalance, pueden llegar a predecir más cierta clase que está sobrerrepresentada en los set de datos. Las técnicas de balanceo buscan, valga la redundancia, balancear las clases existentes para prevenir sobreajuste de algún modelo predictivo hacia una clase más común que otra.

2. En las sesiones de trabajo usamos 2 métodos de balanceo, sin embargo, existen otros más, mencione 2 de estos, sin importar su objetivo.

Algoritmos de detección de anomalías como One-class SVMs, and Isolation Forests.

3. A la hora de llevar a cabo nuestra sesión de laboratorio comparamos el efecto de unir métodos de balanceo versus solamente selección de variables. Resuma brevemente estos hallazgos.

Con solamente feature selection con BORUTA:

Accuracy : 0.9773

Sensitivity : 0.8996

Specificity : 0.9973

Balanced Accuracy : 0.9484

'Positive' Class : 1

	-1	1
-1	1092	63
1	5	1034

Con Balanceo y feature selection:

Accuracy : 0.9243

Sensitivity : 0.9179

Specificity : 0.9259

Balanced Accuracy : 0.9219

'Positive' Class : 1

	-1	1
-1	512	11
1	41	123

Con estos datos se puede apreciar un gran desbalance de clases en la primera matriz de confusión, más de 1000 de la clase -1 y menos de 70 en las clase 1. Después del balanceo, la matriz de confusión se ve mejor y con menos probabilidad de sobreajuste. El balanceo trae una baja en accuracy, lo que nos hace pensar que es muy probable que haya estado sobre ajustado en el caso sin balanceo.