

Nombres: _____
Códigos: _____

Nota competencia (60%): La conformación de 6 equipos de trabajo los cuales deben competir por obtener los mejores resultados en términos de capacidad predictiva específica (15%), capacidad predictiva general (20%) y selección de regresores (25%). Para esto a los equipos de trabajo se les entregarán 3 bases de datos, las cuales corresponden a variables dependientes con soportes continuo, binario y conteo. Sobre cada una de las bases de datos, se medirá el desempeño de cada uno de los equipos de trabajo en cada uno de los ítems de evaluación, tal que, el equipo con el mejor desempeño obtendrá una nota de 5, el segundo mejor 4.5, y así sucesivamente hasta 2.5. Cabe resaltar que, en caso de empate en alguna medida de desempeño, se promedian las notas respectivas a la posición superior e inferior, y se asigna esta nota a los respectivos equipos, tal que el siguiente equipo en desempeño se le asigna la nota inferior menos 0.5 unidades. El entregable debe ser un documento que establezca el proceso seguido para realizar la tarea encomendada para cada base de datos. Este archivo será utilizado para juzgar si el grupo de trabajo en cuestión merece una mejora en la nota obtenida en la competencia. Los otros archivos son los códigos de programación, tal que sean replicables sin ningún inconveniente, y que lean dos archivos de trabajo, el primero, la base de datos utilizada para escoger la metodología para realizar los pronósticos y la selección de variables, y el segundo, un archivo que contiene la base de datos sobre la cual se realiza la evaluación de la capacidad predictiva general, y específica.

La capacidad predictiva general será el error cuadrático medio para las variables continuas y conteo, está última debe tener un criterio para ser discretizada, y el accuracy (verdaderos positivos + verdaderos negativos dividido el tamaño muestral) para la variable binaria. La capacidad predictiva específica, será la correcta clasificación de los valores inferiores y superiores a -1 para la variable continua. Para la variable binaria, será el área bajo la curva ROC. En el caso de la variable de conteo, será la correcta clasificación de los valores iguales a 0, y mayores a 0. En el caso de la selección de las variables cada correcta detección dará un valor de 1, y cada incorrecta clasificación restará 1.