

UT02 – TA 01

Ejercicio 1

PROBLEMA DE PREDICCIÓN: Predice la supervivencia a la tragedia del Titanic a partir de una serie de atributos observables

analizar atributos y comprender / documentar su significado en el problema:

los atributos recabados son:

clase - Toma los valores 1, 2 o 3 dependiendo la clase en la que viajaba el pasajero

nombre

sexo

edad

Número de hermanos/esposo a bordo

Número de padres/niño a bordo

Número de ticket

Tarifa

Cabina

Puerto en el cual embarcó

Bote salvavidas (si sobrevivió)

Número de cuerpo (si no sobrevivió y fue encontrado)

Destino

Probablemente los atributos de sexo estén vinculados con si sobrevivió o no dado que mujeres subían primero a los botes salvavidas.

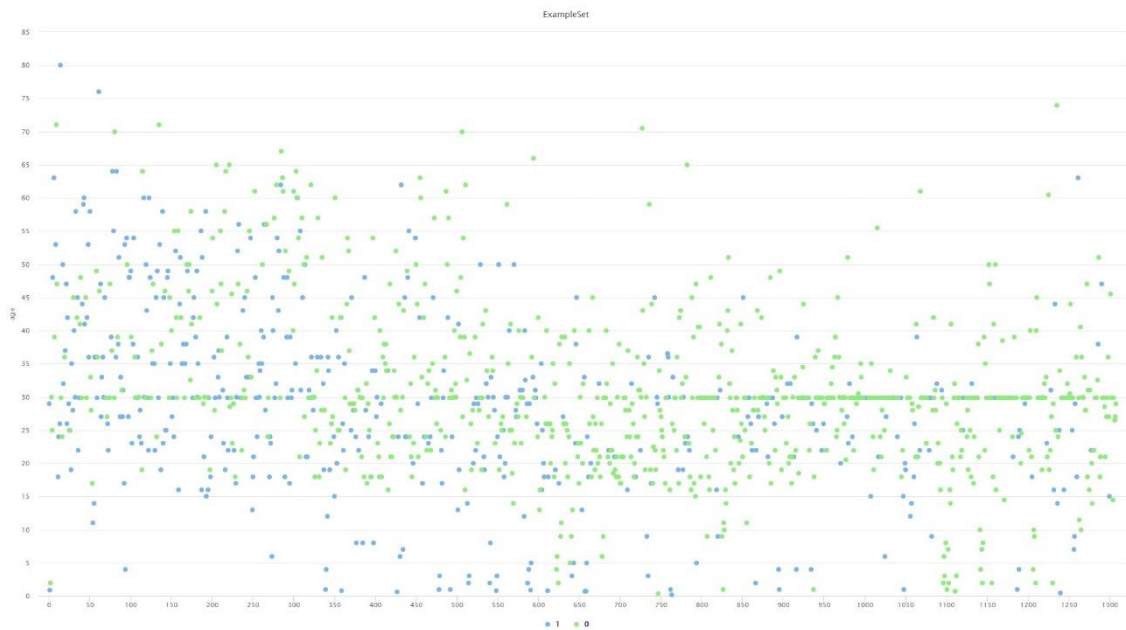
Por otro lado el número de bote y número de cuerpo están ambas muy correlacionadas con el hecho de haber sobrevivido ya que una asume un valor

en el caso de que haya sobrevivido la persona y la otra en sentido contrario.

revisar atributos faltantes y estadísticas de los atributos:

cabina, número de bote y número de cuerpo tienen alto número de missing (1014, 823 y 1188)

Gráfico de dispersión de puntos



Ejercicio 2.

Se seleccionaron atributos considerados relevantes para el problema en cuestión, dejando fuera aquellos con múltiples valores perdidos y aquellos

que obviamente no eran relevantes para el problema como el nombre o el puerto en el que subió, así como el puerto de destino.

Por otro lado, hay variables obvias como número de bote o número de cuerpo que no tiene sentido incorporarlas por desprenderse directamente de aquello

que se quiere evaluar.

Se seleccionan los atributos:

edad

sexo

Número de hermanos/esposo a bordo

Número de padres/niño a bordo

clase

número de ticket

Luego se procedió a reemplazar los valores missing para edad sustituyendo con la media por considerar la mejor solución para reparar dicha falta.

Se analizó una matriz de correlación para observar posibles vínculos entre variables, pero no se observaron correlaciones fuerte.

Ejercicio 3.

Generar matriz de correlación usando el operador “Correlation Matrix”, estudiar los parámetros que acepta el operador:

invertir selección

incluir atributos especiales

normalizar los pesos

correlación cuadrada

Se observa la vinculación de número de hermanos/esposo con número de padres/ hijos a bordo.

Así como una correlación negativa menor de edad con número de hermanos a bordo, y sexo con padre/esposo a bordo.

Por último una correlación positiva de tarifa con Número de hermanos/esposo a bordo

[illegible]