

Tarea 2
Proceso de minería de datos

1. Aplique el proceso de minería de datos para construir un modelo de clasificación con el conjunto de datos "SEISMIC-BUMPS" (repositorio UCI). Utilice la técnica de K-vecinos más cercanos (K-NN). Construya una gráfica que muestre cómo varía el error de predicción y g-medias en función del valor de K.

Justifique cada decisión tomada.

(4 ptos)

2.- Aplique el proceso de minería de datos para construir un modelo descriptivo basado en agrupación para el conjunto de datos "WHOLESALE CUSTOMERS" (repositorio UCI). El objetivo es determinar patrones de compra de clientes. Realice la interpretación de los grupos obtenidos tomando en cuenta el contexto del problema.

Justifique cada decisión tomada.

(10 ptos)

3.- Aplique el proceso de minería de datos para realizar un análisis de asociación sobre el conjunto de datos "BANK MARKETING" (repositorio UCI). El objetivo es encontrar asociaciones frecuentes que permitan caracterizar a los clientes que podrían suscribirse o no a los servicios de depósitos bancarios. Utilice un umbral para el soporte de 70% y una confianza de 80%. Tomando en cuenta el contexto del problema, realice una interpretación de las reglas generadas.

Justifique cada decisión tomada.

(6 ptos)

4.- **OPCIONAL:** Aplique el proceso de minería de datos para obtener un modelo de clasificación a partir del conjunto de datos "STUDENT ALCOHOL CONSUMPTION" (repositorio UCI), el cual está relacionado con la predicción de consumo de alcohol en estudiantes. Utilice los algoritmos C4.5, RIPPER y K-NN.

a) Con base en las medidas sensibilidad, g-medias y precisión ¿Cuál modelo considera mejor?

b) A continuación, utilice el algoritmo SMOTE para corregir la desproporción entre las clases. Construya los modelos de clasificación con las tres técnicas. Ahora, ¿Cuál modelo considera mejor?

Justifique cada decisión tomada.

(10 ptos)

Fecha de entrega: 15/04/16

MD/Hn