

Tarea 2

En esta tarea trabajará buscando reglas de asociación en el dataset *supermarket.arff* y con el software Weka versión 3.7.12, disponible en <http://www.cs.waikato.ac.nz/ml/weka/>.

Los algoritmos que debe probar son los siguientes:

1. Apriori
2. FP-growth

Lo que debe hacer es lo siguiente:

1. Para el algoritmo Apriori debe seleccionar las top-5000 reglas más interesantes (numRules). Debe encontrar las reglas para valores de confianza (minMetric) 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94 y 0.95. Para cada valor de confianza, contabilice el número de reglas encontradas. Luego, plotee el número de reglas (eje y) en función de la confianza (eje x). Comente sus resultados. Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas.
2. Usando la misma configuración anterior sobre el algoritmo Apriori, cambie el valor del soporte mínimo (lowerBoundMinSupport) a 0.15. Repita el experimento anterior (encontrar las reglas para valores de confianza) y plotee el número de reglas (eje y) en función de la confianza (eje x). Comente sus resultados. Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas. Justifique su elección.
3. Para el algoritmo FP-growth, suba el soporte mínimo a 0.16 y busque reglas con confianza mínima en el rango 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94 y 0.95. Para cada valor de confianza, contabilice el número de reglas encontradas. Repita el experimento con FP-growth, subiendo el soporte mínimo en 1% cada vez, hasta llegar a 0.25. Para cada valor de confianza, contabilice el número de reglas encontradas. Luego, plotee el número de reglas (eje y) en función de la confianza (eje x). Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas. Justifique su elección.

Importante: Puede hacer la tarea en grupos de hasta 3 personas. Empiece a hacerla de inmediato, puede encontrarse con dificultades en algunos pasos.

Entregables:

1. Informe: subir un informe con los resultados obtenidos en su tarea. Muestre todos los plots encontrados. Indique la configuración de parámetros usada. Comente las reglas seleccionadas como inesperadas. Indique los integrantes del grupo.

Evaluación: Las tareas serán evaluadas de acuerdo a eficacia.

Entrega: Viernes 10 de Julio, en moodle. A las 24:00 Hrs. el link de *upload* se cerrará.