



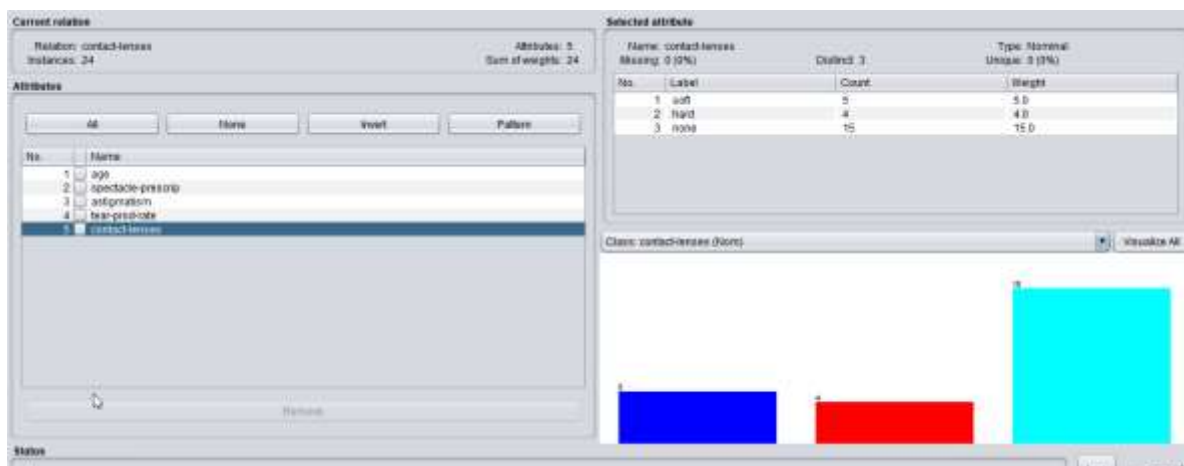
REGLAS DE ASOCIACIÓN

Las reglas de asociación permiten identificar tendencias o relaciones entre los datos almacenados en una base de datos y a partir de las coincidencias encontradas se generan reglas, de acuerdo a lo anterior es importante determinar la calidad de las reglas de tal manera que permitan garantizar una solución óptima para el análisis del comportamiento de los datos. También se hace necesario establecer una manera de reducir el número de reglas generadas por el algoritmo de inducción. Las medidas de interés o métodos de asociación son utilizadas justamente para solventar este inconveniente, pues las mismas definen criterios para evaluar la calidad de las reglas de asociación y desechar aquellas que no cumplen dichos criterios.

La actividad a realizar nos indica que a partir del archivo 'contact_lenses' debemos obtener sus valores para la confianza, lift, leverage y conviction.

Analizando la relación de que nos proporcionan encontramos que contact_lenses contiene 5 atributos, cada uno con sus respectivas etiquetas:

- Age (edad)
- Spectacle-prescrip (prescripción de gafas)
- Astigmatism (astigmatismo)
- Tear-prod-rate (tasa de punción lagrimal)
- Contact-lenses (lentes de contacto)





Minería de Datos: Actividad 04

Alumno: Bañuelos López Marco Antonio

Matrícula: 201726707



Adicionalmente el archivo cuenta con datos de tipo nominal, es decir no son cuantitativos.

Para obtener las reglas de asociación en este ejercicio, usamos el algoritmo Filtered-Associator con filtro de AttributeSelection, recordemos algunas características que usamos para calcular las mejores reglas a partir del soporte (soporte), confidence (confianza), lift, conviction y leverage

El soporte expresa el porcentaje o la fracción de registros de T que satisfacen la unión de los elementos del antecedente y consecuente de la regla: $s(X \rightarrow Y) = s(X \cup Y)$.

La confianza es la medida de efectividad de la regla, representando el porcentaje de casos en los que dado el antecedente se verifica la implicación de $c(X \rightarrow Y) = s(X \rightarrow Y) / s(X)$

El lift tiene una fuerte importancia en los resultados ya que a partir de este dato cuantificamos la relación existente entre X e Y. Definido de la siguiente manera: $lift(X \rightarrow Y) = s(X \rightarrow Y) / s(Y)$. y de acuerdo al valor que se obtenga, podemos concluir que

- Lift < 1 entonces X e Y están negativamente correlacionados.
- Lift = 1 entonces X e Y son independientes
- Lift > 1 entonces X e Y están positivamente correlacionados

Conviction como medida, tiene como propósito medir el grado que el antecedente influye en la ocurrencia del consecuente de una regla de asociación diferenciándose del lift porque convicción es una medida unidireccional, representada de la siguiente manera.

$$conv(X \rightarrow Y) = \frac{soporte(X)Soporte(\neg Y)}{soporte(X \cup \neg Y)}$$

Sus resultados se encuentran dentro del rango $(0, \infty)$. Donde 1 significa independencia, si conv es menor a uno representa dependencia negativa y su valor de conv es mayor a uno representa asociación.



Minería de Datos: Actividad 04

Alumno: Bañuelos López Marco Antonio

Matrícula: 201726707



Leverage permite medir cómo se obtiene el recuento de la co-ocurrencia del antecedente y el consecuente por medio de la siguiente notación..

$$\text{leverage}(X \rightarrow Y) = \text{soporte}(X \cup Y) - \text{soporte}(X) * \text{soporte}(Y)$$

Luego por medio de la herramienta de Weka calculando los valores con el algoritmo antes mencionados obtenemos los siguientes valores:

```
Minimum support: 0.1 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 4

Best rules found:

1. tear-prod-rate=reduced 12 ==> contact-lenses=none 12    <conf:(1)> lift:(1.6) lev:(0.19) [4] conv:(4.5)
2. contact-lenses=soft 5 ==> tear-prod-rate=normal 5      <conf:(1)> lift:(2) lev:(0.1) [2] conv:(2.5)
3. contact-lenses=hard 4 ==> tear-prod-rate=normal 4      <conf:(1)> lift:(2) lev:(0.08) [2] conv:(2)
```

Analizando las reglas de manera general encontramos que todas tienen un valor de confianza de 1, su valor de lift se encuentra fuertemente asociado respectivamente, al igual que su valor de conviction el cual es mayor que 1.

De manera detallada, y analizando cada una de las reglas, la primera nos indica la relación entre la tasa de punción lagrimal se reduce cuando el paciente no usa lentes de contacto, si bien el valor de su lift no es mayor de todos, su valor de conviction es el mayor de todos, lo que demuestra una fuerte relación unidireccional.

Considerando la segunda y tercera regla, podríamos suponer que se encuentran altamente relacionadas con respecto a la dureza de los lentes de contacto. En la segunda al ser estos suaves, aumenta la tasa de punción lagrimal, la cual en sus valores de lift y conviction son relativamente similares, mientras que la tercera regla se refiere a que si los lentes de contacto son duros, la tasa de punción lagrimal se reduce un poco. Con valores de lift y conviction iguales.