**Laboratorio de:** Inteligencia de negocios

**Práctica No.:** 2

**Tema: Arboles de decisión con WEKA**

**Nombre:** Díaz Padilla Danny Sebastián  **Fecha:** 13/01/2020

1. **Objetivos:** 
   1. **Objetivo General**

*Crear un árbol de decisión utilizando el clasificador J48 de WEKA*

* 1. **Objetivos Específicos**
* Utilizar el archivo “breastcancer.arf” de los datos de muestra de WEKA.
* Ejemplificar las reglas de clasificación de un árbol de decisión.
* Explicar la matriz de confusión generada al crear un árbol de decisión.

1. **Marco teórico:** 
   1. **Data mining**

Es el proceso de extracción de información significativa de grandes bases de datos, información que revela inteligencia del negocio, a través de factores ocultos, tendencias y correlaciones para permitir al usuario realizar predicciones que resuelven problemas del negocio proporcionando una ventaja competitiva.

Las herramientas de Data Mining predicen las nuevas perspectivas y pronostican la situación futura de la empresa, esto ayuda a los mismos a tomar decisiones de negocios proactivamente. [1]

* 1. **WEKA**

Weka es un software programado en Java que está orientado a la extracción de conocimientos desde bases de datos con grandes cantidades de información. Existen otras herramientas similares como Oracle Data Miner o Clementine, pero el hecho de que Weka sea desarrollado bajo licencia GPL lo ha hecho una alternativa muy interesante. [2]

* 1. **Árbol de decisión**

Diagrama ordenado de secuencias de decisiones alternativas, sucesos futuros con distinta probabilidad de ocurrencia y resultados finales valorados objetiva o subjetivamente. Sirve para visualizar las decisiones a tomar, ahora y en el futuro, y para facilitar la expresión de probabilidades, valoraciones de diversos sucesos y resultados por cada uno de los miembros del equipo que toma la decisión. [3]

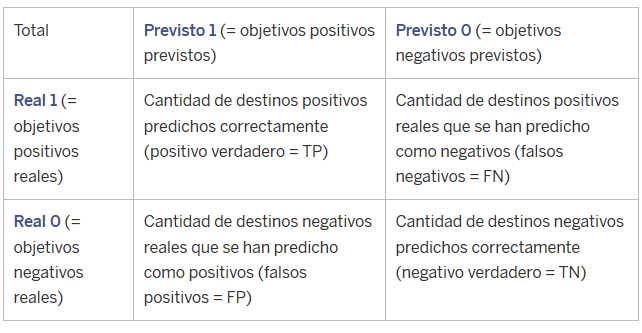
**Clasificador J48**

Es el código para definir al clasificador basado en árboles de decisión (J48) dentro de la herramienta WEKA.

* 1. **Matriz de confusión**

Es una tabla que muestra el rendimiento de un modelo predictivo de clasificación mediante la comparación del valor previsto de la variable de destino con su valor real.

Cada columna de la Matriz de confusión representa las observaciones de una categoría prevista, mientras que cada línea representa las observaciones en una clase real:



**Precisión**

Porcentaje de ítems seleccionados que son correctos.

Fórmula: No. Verdaderos positivos / (No. Verdaderos positivos + No. Falsos positivos)

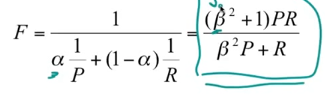
**Recall**

Porcentaje de ítems correctos que son seleccionados.

Fórmula: No. Verdaderos positivos / (No. Verdaderos positivos + No. Falsos negativos)

**F measure**

Es una medida combinada que evalúa el equilibrio P / R (media armónica ponderada):

 [4]

1. **Desarrollo de la práctica:**
2. Abrir el archivo breastcancer.arf con WEKA y crear un árbol de decisión utilizando el clasificador J48.
3. Dar dos ejemplos de reglas de clasificación con los datos del árbol creado.
4. Explicar los datos presentados en la matriz de confusión y los parámetros Precisión, Recall y F measure.
5. Preparar un nuevo archivo con datos para clasificación con el modelo generado
6. Cargar el modelo generado y realizar la clasificación
7. **Análisis de resultados:**
8. **Conclusiones y recomendaciones:**

* Se logró crear un árbol de decisión utilizando el clasificador J48 de WEKA.
* Se realizó un clasificador usando el archivo de muestra sobre “breast cáncer” dado por WEKA.
* Se definieron reglas de clasificación para el árbol de decisión generado.
* Se explicó la matriz de confusión.
* .
* .
* Se recomienda..
* Es recomendable…..

1. **Bibliografía:**

[1]G. .com, "¿Qué es Data Mining? - GestioPolis", Gestiopolis.com, 2001. [Online]. Available: https://www.gestiopolis.com/que-es-data-mining/. [Accessed: 13- Jan- 2020].

[2]"Introducción a la minería de datos con Weka", Locualo.net, 2007. [Online]. Available: http://www.locualo.net/programacion/introduccion-mineria-datos-weka/00000018.aspx. [Accessed: 13- Jan- 2020].

[3]U. S.L., "Árbol de decisión", Expansion.com. [Online]. Available: https://www.expansion.com/diccionario-economico/arbol-de-decision.html. [Accessed: 13- Jan- 2020].

[4]R. Merino García, Precision, Recall, and the F measure. 2012. [Online]. Available: https://www.youtube.com/watch?v=XOgvlpwu0hI [Accessed: 13- Jan- 2020].