**5. Proyecto de implementación de un prototipo del sistema utilizando Weka Oracle**

Se trata de incluir en este apartado la documentación del desarrollo del proyecto de implementación, utilizando la tecnología B, del sistema cuyos requisitos funcionales se enumeraron en el apartado 2.

**5.1 Documentación de diseño**

Hay que incluir la descripción del diseño del prototipo, incluyendo diagramas, y el diseño de la interfaz de usuario.

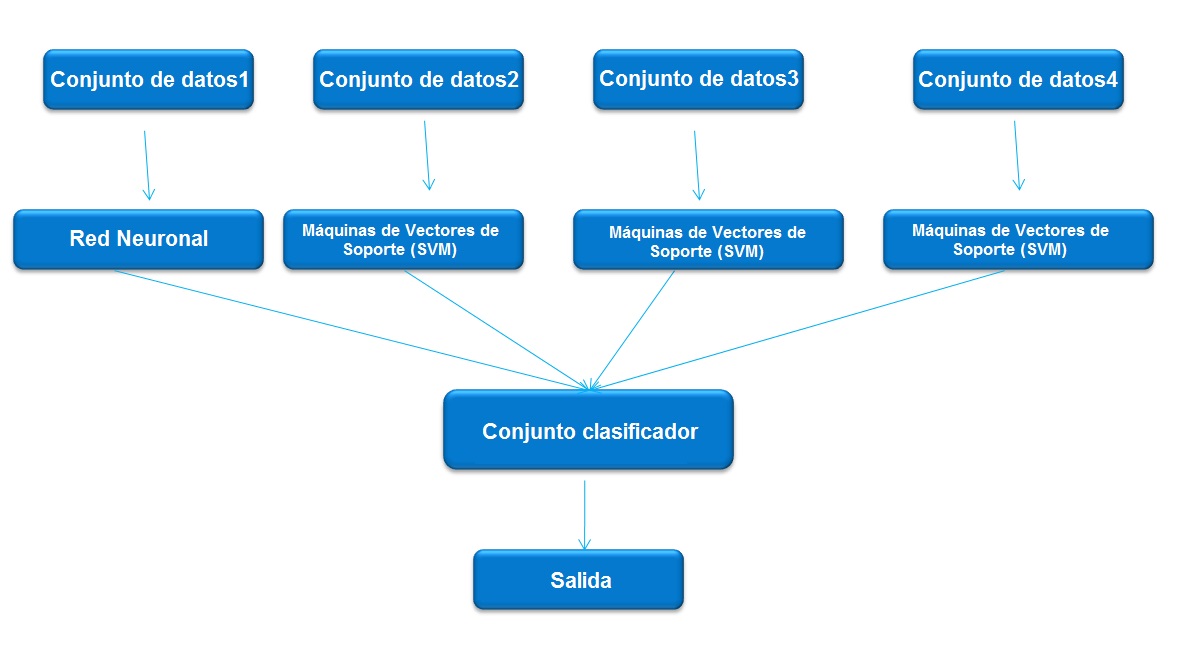
El programa de implementación cumple con todos lo requisitos propuestos anteriormente.

El programa Weka Oracle se puede instalar en diferentes sistemas operativos, ya que es multiplataforma, los pasos explicados para su instalación están en el apartado 5.4 Documentación de instalación.

Se utiliza el lenguaje Java.

Para cargar los datos, necesitamos que estos sean de tipo ARFF, DATA, CSV o JSON. Cuando cargamos estos datos, nos muestra las variables y podemos seleccionar tantas variables como tenga el archivo para luego comparar datos y poder analizarlos.

Una vez generado el árbol de decisión y no haya errores en la implementación, este nos mostrará una serie de resultados. Donde se nos permitirá analizar y ver cómo de buenos son los resultados.



**5.3 Documentación de pruebas**

**Errores**

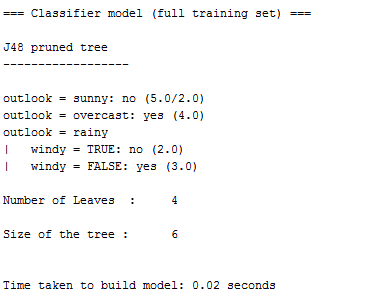


Error al abrir un fichero.

**Resultado**

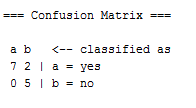
Una vez cargado el archivo.arff y tenemos las variables de “outlook”, “temperature”, “humidity”, “windy” y “play” cargadas, pasamos a clasificar los datos.

Lo que hacemos es seleccionar el algoritmo, en este caso de árbol, el J48, y seleccionamos la variable “play”. El resultado es el siguiente:



Lo que nos dice este resultado es que si el clima es soleado, no juegan 2 de 5 personas; si está nublado, juegan 4 personas; y si llueve, dependiendo si hace o no aire, no juegan 2 personas y juegan 3 respectivamente. Tambíen nos muestra el número de hojas (4) y el tamaño del árbol (6).

Otros de los datos que nos muestran es la matriz de confusión:



En este caso vemos que la variable “a” son las personas que sí juegan, mientras la “b” son las personas que no juegan.

Nos dice que de 7 registros que se analizan de personas que juegan, hay 0 errores. Y de 5 registros que se analizan de personas que no juegan, hay 2 errores.

En la matriz de confusión debemos tener en cuenta que la diagonal debe de ser mayor, este caso, 7 es mayor que 0, y 5 es mayor que 2. Por lo tanto, la matriz tiene sentido.

La forma gráfica de visualizar el árbol es la siguiente:

