

UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

Introducción a los sistemas inteligentes (75.50)

Profesor Adjunto: Ochoa María

Ayudante: Vazquez Diego



Tutorial Elvira

Introducción a Elvira	pág. 2
Creación y edición de redes	pág. 5
Definición de la red	pág. 7
Inferencia en Elvira	pág. 9
Cargando datos externos	pág. 16

Introducción a Elvira

El programa Elvira es fruto de un proyecto de investigación financiado por la CICYT y el Ministerio de Ciencia y Tecnología, en el que participan investigadores de varias universidades españolas y de otros centros.

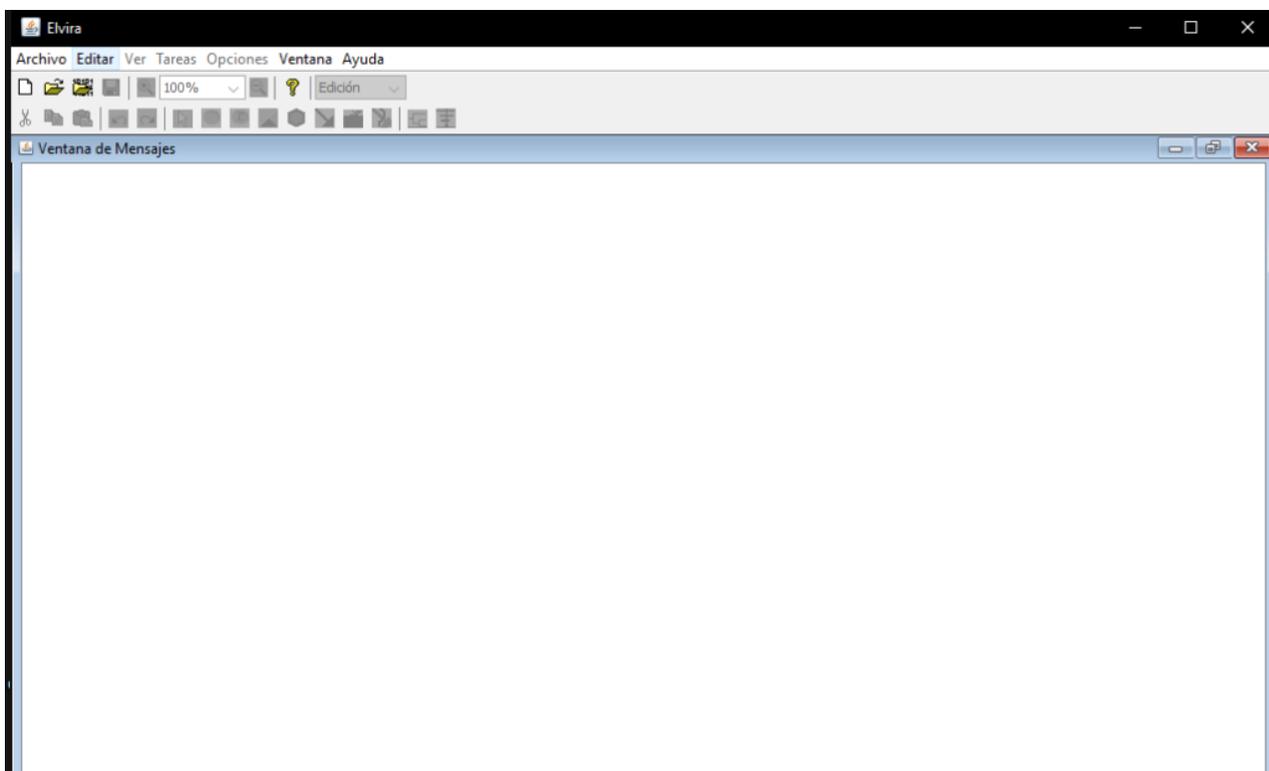
El programa Elvira está destinado a la edición y evaluación de **modelos gráficos probabilistas**, concretamente redes bayesianas y diagramas de influencia. Elvira cuenta con un formato propio para la codificación de los modelos, un lector-intérprete para los modelos codificados, una interfaz gráfica para la construcción de redes, con opciones específicas para modelos canónicos (puertas OR, AND, MAX, etc.), algoritmos exactos y aproximados (estocásticos) de razonamiento tanto para variables discretas como continuas, métodos de explicación del razonamiento, algoritmos de toma de decisiones, aprendizaje de modelos a partir de bases de datos, fusión de redes, etc.

Elvira está escrito y compilado en Java, lo cual permite que funcione en diferentes plataformas y sistemas operativos (Windows, linux, Solaris, etc.).

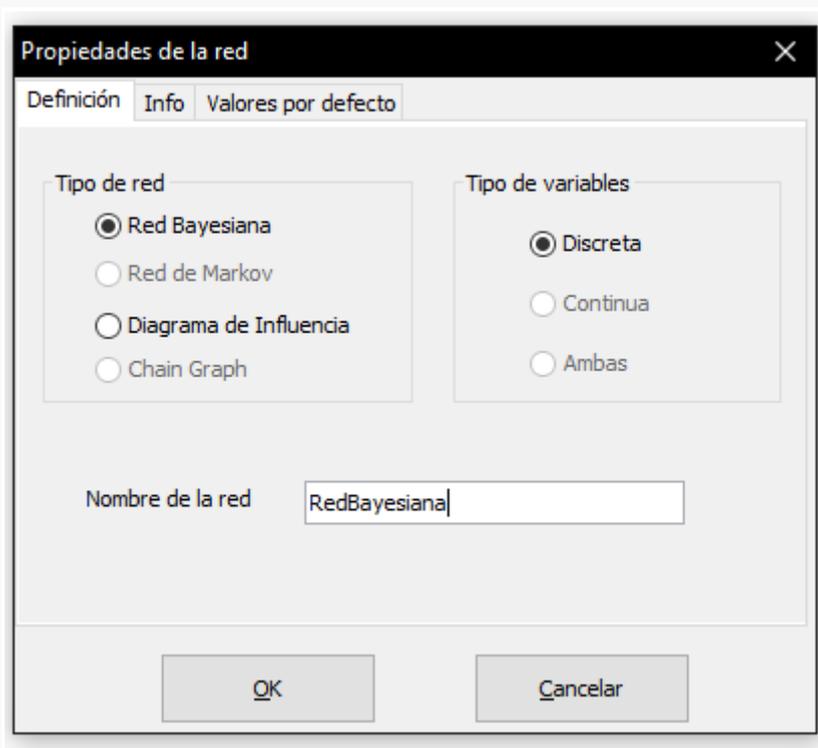
Para descargar el programa Elvira podemos hacerlo desde el siguiente enlace:

<http://www.ia.uned.es/~elvira/instalar/Elvira.zip>

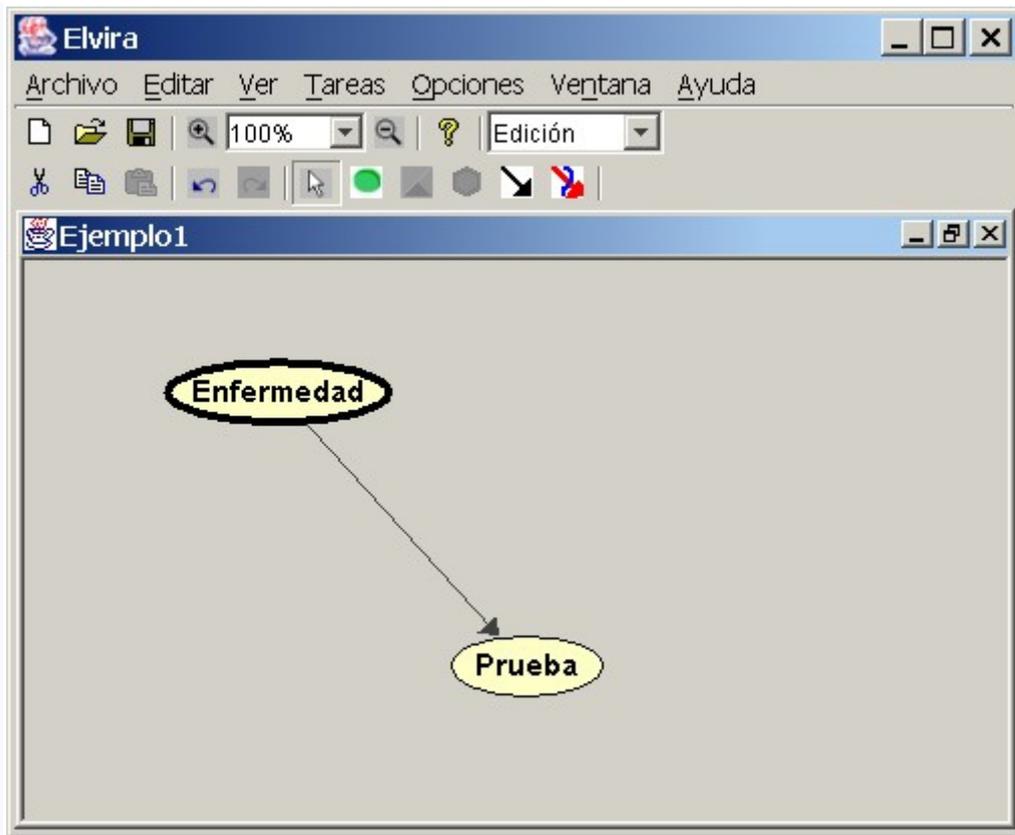
La instalación del programa es muy sencilla y cuando terminen de instalar les aparecerá la siguiente ventana:



Para crear la red bayesiana desde cero hacemos clic en **Archivo -> Nuevo**



1. En la barra de menú ofrece siete opciones: *Archivo, Editar, Ver, Tareas, Opciones, Ventana y Ayuda*, aunque, lamentablemente, la versión actual de Elvira no ofrece ayuda en línea.
2. En la primera barra de herramientas hay botones para crear, abrir y guardar redes, hacer zoom, obtener información sobre Elvira y seleccionar el modo básico: *Edición, Inferencia o Aprendizaje*.
3. La segunda barra de herramientas varía dependiendo del modo básico en que nos encontremos: *Edición, Inferencia o Aprendizaje*. La barra de edición, que es la que aparece en la figura anterior, tiene botones para cortar, copiar, pegar, deshacer/rehacer cambios y cinco botones para escoger la herramienta de edición, que puede ser una de las cinco siguientes:
 - seleccionar
 - añadir nodo aleatorio
 - añadir nodo decisión (sólo está activo para diagramas de influencia)
 - añadir nodo utilidad (sólo activo para diagramas de influencia)
 - añadir enlace
4. Debajo de la barra de herramientas hay un espacio destinado a mostrar la ventana de mensajes y las redes abiertas



Creación y edición de redes

Supongamos que queremos resolver el siguiente ejemplo mediante Elvira:

Ejemplo 1. Para una enfermedad que tiene una prevalencia del 8% existe una prueba que tiene una sensibilidad del 75% y una especificidad del 96% respecto de dicha enfermedad. ¿Cuáles son los valores predictivos de la prueba?

Vemos que en este problema intervienen dos variables: Enfermedad y Prueba. La enfermedad puede tener dos valores, presente y ausente, y la prueba puede dar dos resultados, positivo y negativo.

Dado que la primera variable influye causalmente sobre la segunda, vamos a trazar un enlace

Enfermedad->Prueba.

Definición de la red

Para definir una red en Elvira, pulsamos pulsamos el icono Crear nueva red (), que se encuentra en la primera barra de herramientas. Se abrirá la ventana Propiedades de la red. El tipo de red que Elvira escoge por defecto es Red bayesiana, así que no necesitamos modificar este valor. En el campo Nombre de red, podemos escribir Ejemplo1. Luego pulsamos el botón Aceptar.

Estructura de la red

Para **crear los nodos**, vamos a la segunda barra de herramientas, pulsamos sobre el icono *Añadir nodo aleatorio* ().

Luego hacemos clic en el lugar donde queremos que aparezca el nodo Enfermedad. Haciendo doble-clic sobre este nodo, se abre la ventana *Propiedades de nodo*. En el campo *Título* de esta ventana escribimos Enfermedad. Pulsando en la pestaña *Valores* de la misma ventana podemos comprobar que los valores que Elvira ha asignado por defecto a esta variable son "presente" y "ausente", precisamente los que nosotros deseábamos. Por eso pulsamos el botón *Aceptar*.

Del mismo modo creamos un segundo nodo, hacemos doble-clic sobre él para abrir la ventana *Propiedades de nodo*, y en el campo *Título* escribimos Prueba. En la pestaña *Valores* de esta ventana, vamos al campo *Tipo de valores* y en la lista desplegable seleccionamos *Positivo-Negativo*. Luego pulsamos en *Aceptar* para cerrar la ventana.

Para completar la estructura de la red, vamos a **añadir un enlace** desde el primer nodo hasta el segundo. Primero hemos de cambiar de la herramienta *Añadir nodo aleatorio* () a *Añadir enlace dirigido* (), que está también en la segunda barra de herramientas. Después hay que arrastrar el ratón desde el nodo origen, Enfermedad, hasta el nodo destino, Prueba. ("Arrastrar" significa pulsar con el botón izquierdo del ratón sobre el nodo Enfermedad y, sin soltar el botón, desplazamos el puntero hasta el nodo Prueba.) El resultado debe ser similar al que aparece en la figura anterior.

Si desea mover de posición un nodo, cambie a la herramienta *Selección* pulsando en el icono correspondiente de la segunda barra de herramientas () y arrástrelo con el ratón.

Nota sobre los nombres de los nodos. En Elvira cada nodo tiene un nombre y, opcionalmente, un título. La diferencia es que el nombre es más restrictivo en cuanto al tipo de caracteres que puede contener; por ejemplo, el nombre no puede contener espacios, mientras que el título sí. Además, Elvira a veces da problemas cuando el nombre de un nodo tiene eñes o caracteres acentuados. Por eso recomendamos que el nombre contenga sólo letras (excepto la "ñ") sin acentuar, números y el guión bajo, "_".

Generalmente, Elvira muestra el título de cada nodo (si lo tiene; si no, muestra el nombre). Este comportamiento puede cambiarse escogiendo la opción *Ver / Por nombre* en vez de *Ver / Por título*

Probabilidades condicionada

Una vez que tenemos las variables y enlaces, vamos a introducir la información numérica, es decir, la tabla de probabilidad condicionada (TPC) de cada nodo.

Para la variable Enfermedad, la TPC viene dada por su prevalencia, que nos dice que $P(\text{Enfermedad}=\text{presente}) = 0.08$. Para introducir este dato, hacemos doble-clic sobre el nodo Enfermedad y seleccionamos la pestaña *Relación*; en el menú *Tipo de relación* escogemos "General", luego escribimos el 0.09 en la casilla correspondiente y pulsamos la tecla *Intro*. Observe que el valor de la celda inferior cambia a 0.92, porque la suma de las probabilidades debe ser la unidad.

Estado	Probabilidad
presente	0.08
ausente	0.92

En esta figura aparecen ambas probabilidades porque está seleccionado el botón *Todos los parámetros*. En cambio, si seleccionamos el botón *Parámetros independientes* sólo aparece el valor 0.08; el valor 0.92 no aparece, porque no es necesario mostrarlo.

Para la variable Prueba, la TPC se construye teniendo en cuenta que la sensibilidad (75%) es la probabilidad de que la prueba dé positivo cuando la enfermedad está presente, y la especificidad

(96%) es la probabilidad de que la prueba dé negativo cuando la enfermedad está ausente. La TPC introducida en Elvira debe quedar así:

Enfermedad...	presente	ausente
positivo	0.75	0.04
negativo	0.25	0.96

En principio, el grafo de la red no ofrece información sobre las probabilidades condicionadas. Sin embargo, al pulsar el botón , Elvira ofrece una idea aproximada de la información numérica contenida. Esto hace que cada nodo se muestre con un color y una anchura propios. Hablando *grosso modo*, los enlaces coloreados en rojo indican que hay una influencia positiva, es decir que al aumentar el valor del nodo padre aumenta el valor del nodo hijo; los enlaces en azul indican influencia negativa; los enlaces negro indican que el enlace no transmite información; los enlaces en violeta indican que hay influencia ambigua, es decir, que en unos para ciertos valores del padre hay influencia positiva y para otros negativa. Cuanto mayor es la influencia de un nodo sobre otro, mayor es la anchura del enlace.

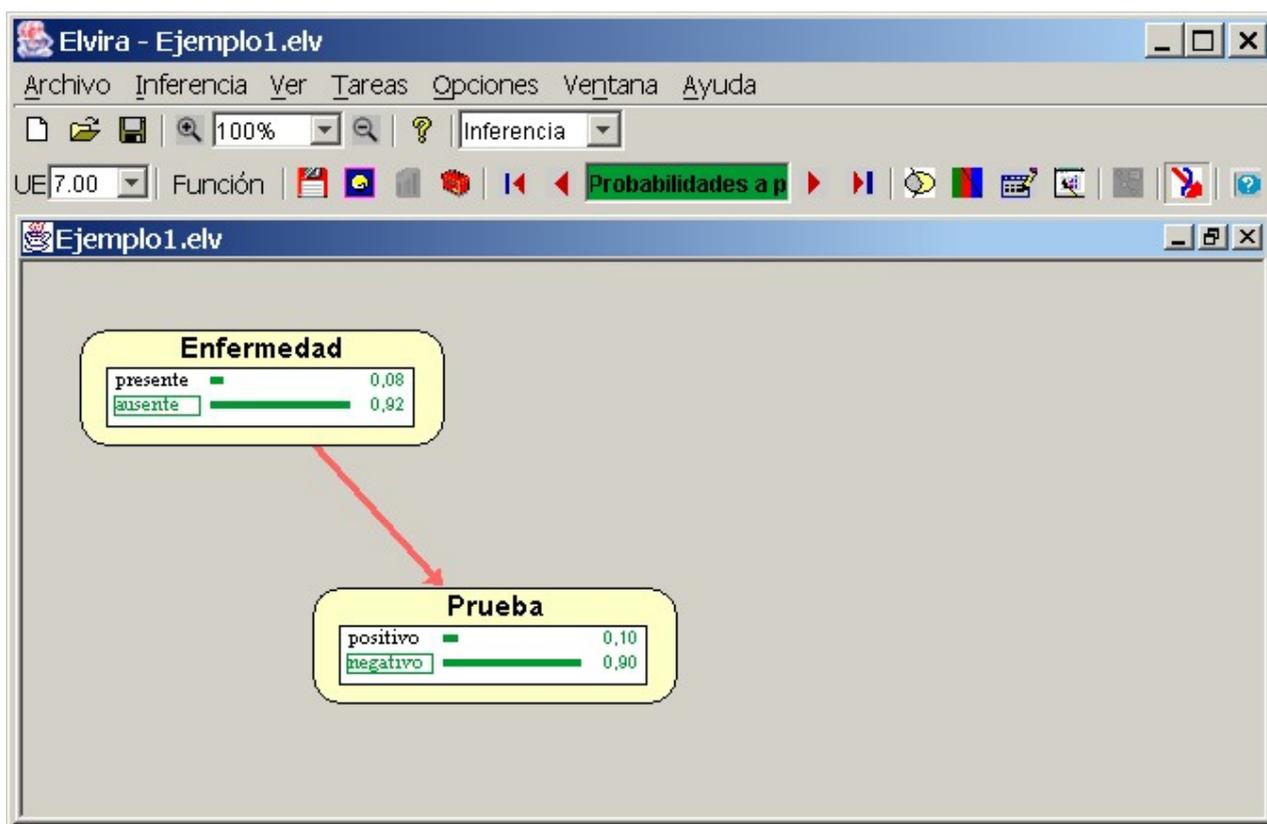
Guardar la red

Aunque no es obligatorio, antes de pasar el modo inferencia conviene guardar en disco la red que hemos creado. La forma más sencilla de hacerlo consiste en pulsar el icono *Guardar red* (). También se puede guardar a través de la opción *Archivo* de la barra de menú, seleccionando *Guardar red*, *Guardar red como*, o *Guardar todas las redes*. Por último, si decidimos salir de Elvira cuando aún hay una red modificada y no guardada, aparecerá una ventana en la que se nos preguntará si queremos guardar la red.

Inferencia en Elvira

De modo edición a modo inferencia

Una vez que tenemos creado y guardado la red, vamos a ver cómo podemos utilizarla para resolver el problema planteado en la sección anterior. Para ello, vamos a la lista desplegable de la primera barra de herramientas, y donde decía *Edición* seleccionamos *Inferencia*. La apariencia de Elvira será ésta:



Se observa que la segunda barra de herramientas ha cambiado: han desaparecido los botones de edición, como cortar, copiar y pegar, y han aparecido otros botones y campos nuevos: umbral de expansión, guardar caso en fichero, guardar caso en memoria, expandir/contraer nodo, opciones de inferencia, etc.

Hay un segundo cambio, no visible para el usuario, que consiste en que Elvira crea una estructura de datos interna, que será utilizada para calcular la probabilidad de cada variable. Este proceso, conocido como **compilación de la red**, es el que hace que para redes grandes Elvira tarde unos segundos en pasar de modo edición a modo inferencia.

También se observa que los nodos se han expandido. De hecho, en el ejemplo que estamos describiendo, comprobamos que al pasar a modo inferencia, Elvira nos muestra la probabilidad de

cada valor de dos modos: mediante una barra de longitud proporcional a la probabilidad y mediante un número. Como aún no hemos introducido ningún hallazgo, aparecen las probabilidades a priori. La probabilidad de Enfermedad=presente es 0.08, uno de los datos que habíamos introducido de acuerdo con el enunciado del problema. En cambio, la probabilidad a priori para la variable prueba ha sido calculada por Elvira. En efecto, podemos comprobar que:

$$\begin{aligned}
 P(\text{Prueba}=\text{positivo}) &= P(\text{Prueba}=\text{positivo} | \text{Enfermedad}=\text{presente}) \times P(\text{Enfermedad}=\text{presente}) \\
 &\quad + P(\text{Prueba}=\text{positivo} | \text{Enfermedad}=\text{ausente}) \times P(\text{Enfermedad}=\text{ausente}) \\
 &= 0.75 \times 0.08 + 0.04 \times 0.92 = 0.0968 \\
 P(\text{Prueba}=\text{negativo}) &= P(\text{Prueba}=\text{negativo} | \text{Enfermedad}=\text{presente}) \times P(\text{Enfermedad}=\text{presente}) \\
 &\quad + P(\text{Prueba}=\text{negativo} | \text{Enfermedad}=\text{ausente}) \times P(\text{Enfermedad}=\text{ausente}) \\
 &= 0.25 \times 0.08 + 0.96 \times 0.92 = 0.9032
 \end{aligned}$$

En principio, Elvira muestra los valores numéricos con dos decimales. Para ver más decimales, seleccione la opción *Ver* de la barra de menú y escoja la *Precisión* que desee.

Introducción de hallazgos. Probabilidades a posteriori

La pregunta que nos planteaba el problema inicial era calcular los valores predictivos. El valor predictivo positivo (*VPP*) es la certeza con que podemos diagnosticar la enfermedad cuando la prueba da positivo: $P(\text{Enfermedad}=\text{presente} | \text{Prueba}=\text{positivo})$. Para calcularlo con Elvira, introducimos el hallazgo Prueba=positivo y vemos cuál es la probabilidad a posteriori para el valor Enfermedad=presente.

La forma de introducir un hallazgo en Elvira consiste en hacer doble-clic en el valor correspondiente. En este caso, debemos ir al nodo Prueba y hacer doble-clic en "positivo". Se observa que la probabilidad de "Prueba=positivo" pasa a ser 1, porque conocemos con certeza este valor, y la probabilidad del otro valor, "Prueba=negativo", pasa a ser 0, porque los valores de una variable son mutuamente exclusivos. También se observa que la probabilidad de Enfermedad=presente ha aumentado de 0.08 a 0.62, que es precisamente el valor que estábamos buscando:

$$VPP = P(\text{Enfermedad}=\text{presente} | \text{Prueba}=\text{positivo}) = 0.62 = 62\%$$

Comprobamos así que, como era de esperar, un resultado positivo en el test tiende a confirmar la presencia de la enfermedad.

El valor predictivo negativo (*VPN*) es la certeza con que podemos descartar la enfermedad cuando la prueba da negativo: $P(\text{Enfermedad}=\text{ausente} | \text{Prueba}=\text{negativo})$.

Para calcularlo con Elvira, introducimos el hallazgo "Prueba=negativo" y vemos cuál es la probabilidad a posteriori para el valor "Enfermedad=ausente".

Si nos interesa observar cómo varía la probabilidad de Enfermedad en función del resultado de la prueba, debemos almacenar el **caso de evidencia** actual. (Un caso de evidencia se caracteriza por un conjunto de hallazgos. En este caso, el único hallazgo que tenemos es "Prueba=positivo".) Para almacenar el caso de evidencia actual hay que pulsar en el icono *Guardar el caso en memoria* () de la segunda barra de herramientas. Se observa que en esta barra de herramientas el campo de texto que decía *Caso número 1* sobre fondo rojo, ahora dice *Caso número 2* sobre fondo azul.

Ahora que estamos en el caso de evidencia nº 2, hacemos doble-clic sobre el valor "negativo" del nodo Prueba. La probabilidad de "Prueba=negativo" pasa a ser 1 y la de "Prueba=positivo" se queda en 0. Observamos también que $P(\text{Enfermedad=presente}|\text{Prueba=negativo}) = 0.02$ y $P(\text{Enfermedad=ausente}|\text{Prueba=negativo}) = 0.98$. Por tanto,

$$VPN = P(\text{Enfermedad=ausente} | \text{Prueba=negativo}) = 0.98 = 98\%$$

con lo que queda resuelto el problema que teníamos planteado.

Trabajando con varios casos de evidencia

En Elvira, cada caso de evidencia está constituido por un conjunto de hallazgos. En el ejemplo anterior hemos trabajado considerado tres casos:

- *Caso inicial* (color verde): no tenía ningún hallazgo; por eso la probabilidad de cada variable era su probabilidad a priori.
- *Caso número 1* (color rojo): tenía como hallazgo "Prueba=positivo"; por eso, la probabilidad correspondiente a este valor era 1 y la barra alcanzaba su máxima longitud, mientras que la probabilidad de "Prueba=negativo" era 0 y la barra no era visible. La probabilidad de "Enfermedad=presente" era 0.62, bastante más alta que la probabilidad a priori (0.08), como era de esperar.
- *Caso número 2* (color azul): el único hallazgo es "Prueba=negativo", que hace que la probabilidad de este valor sea 1 y la del otro valor, "Prueba=positivo", sea 0. Como era de esperar, la probabilidad de "Enfermedad=ausente" (0.02) es aún más baja que la del caso inicial (0.08), en que no teníamos ningún hallazgo, pues un resultado negativo en la prueba aporta evidencia para descartar la enfermedad.

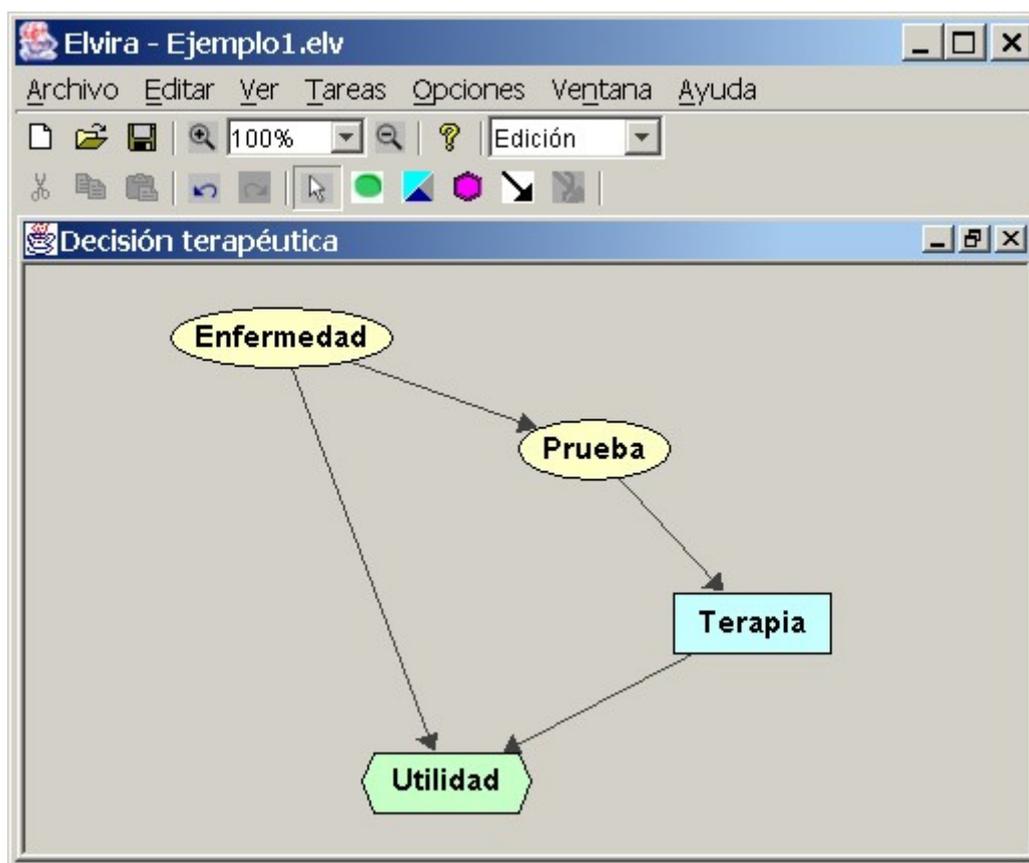
De este modo, podemos observar gráficamente el impacto que cada hallazgo ejerce sobre las probabilidades de las demás variables, lo cual ayuda a comprender el funcionamiento de la red bayesiana y, en el caso de que deseemos mejorar la fiabilidad de la red, podemos corregir la estructura o las probabilidades cuando los resultados no sean los deseados.

En la barra de inferencia hay un botón que permite borrar la evidencia del caso actual, y otro que permite borrar todos los casos de evidencia (excepto el *caso inicial*, correspondiente a la probabilidad a priori).

Edición de diagramas de influencia

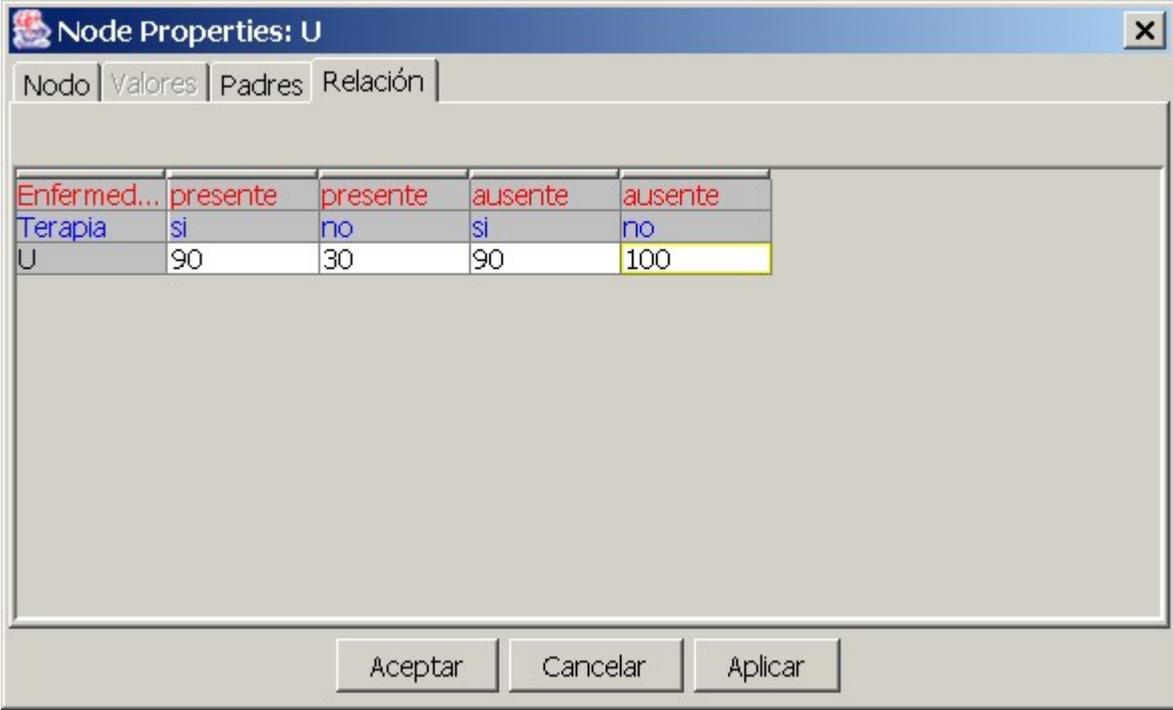
La edición de diagramas de influencia es similar a la edición de redes. Como ejemplo, para editar el grafo de un diagrama de influencia puede seguir los pasos siguientes.

- a) Pulse el icono *Crear nueva red* () , seleccione *Diagrama de influencia* y escriba como título "Decisión terapéutica". Introduzca las variables Enfermedad y Prueba y sus correspondientes tablas de probabilidad, como hemos indicado anteriormente.
 - b) Si no desea repetir los pasos anteriores, puede transformar la red bayesiana del ejemplo anterior en un diagrama de influencia. Para ello, haga clic con el botón derecho del ratón en alguna zona vacía de la ventana de la red bayesiana (es decir, donde no haya un nodo ni un enlace) y seleccione *Editar propiedades de la red* y luego seleccione *Diagrama de influencia*. Observe cómo se han activado los iconos que permiten añadir nodos de decisión y de utilidad.
2. Seleccione luego el icono *Insertar nodos de decisión* () y añada el nodo Terapia. Haciendo doble clic sobre él, asígnele los valores "sí" y "no". Trace un enlace Prueba->Terapia. Éste es un enlace de información, que indica que en el momento de tomar la decisión de aplicar terapia o no aplicarla se conoce ya el resultado de la prueba.
3. Seleccione el icono *Insertar nodos de utilidad* () y añada el nodo Utilidad. Trace un enlaces Enfermedad->Utilidad y otro Terapia->Utilidad. Estos dos enlaces significan que el estado del paciente depende de si tiene o no tiene la enfermedad y de si se le aplica la terapia o no.



Una vez editado el grafo, se pueden introducir las tablas de probabilidad condicionada para los nodos Enfermedad y Prueba. El nodo Terapia no tiene una tabla de probabilidad porque no representa una variable aleatoria sino una decisión, que se tomará en función de la información disponible, como veremos en el apartado siguiente. Sólo falta incluir la tabla de utilidad para el nodo Utilidad.

Para ello, haga doble clic en el nodo utilidad e introduzca los valores que aparecen en la siguiente figura:



The screenshot shows a dialog box titled "Node Properties: U" with a close button (X) in the top right corner. Below the title bar are four tabs: "Nodo", "Valores", "Padres", and "Relación". The "Valores" tab is selected. The main area contains a table with the following data:

Enfermedad...	presente	presente	ausente	ausente
Terapia	si	no	si	no
U	90	30	90	100

At the bottom of the dialog box are three buttons: "Aceptar", "Cancelar", and "Aplicar".

Evaluación de un diagrama de influencia

Siguiendo con el diagrama de influencia del diagrama anterior, vaya a la lista desplegable de la primera barra de herramientas, y donde decía *Edición* seleccione *Inferencia*. Haga clic con el botón derecho sobre el nodo Terapia, escoja la opción "Mostrar política óptima para esta decisión": la tabla que se abre muestra la utilidad asociada a cada decisión (en función de la información disponible al tomar la decisión, que en este ejemplo es el resultado de la prueba).

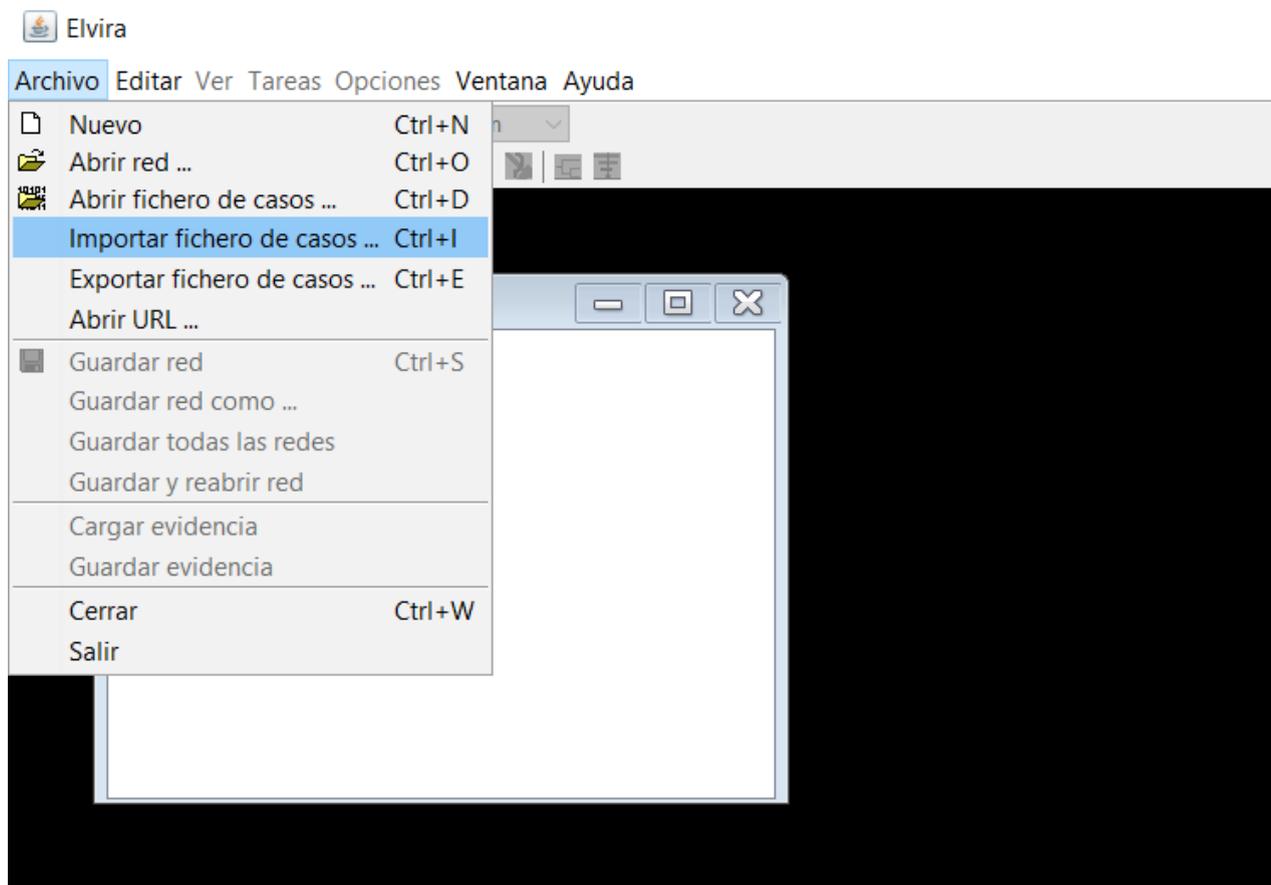
Prueba	positivo	negativo
si	83.80165...	89.77856...
no	56.61157...	98.44995...

Aceptar

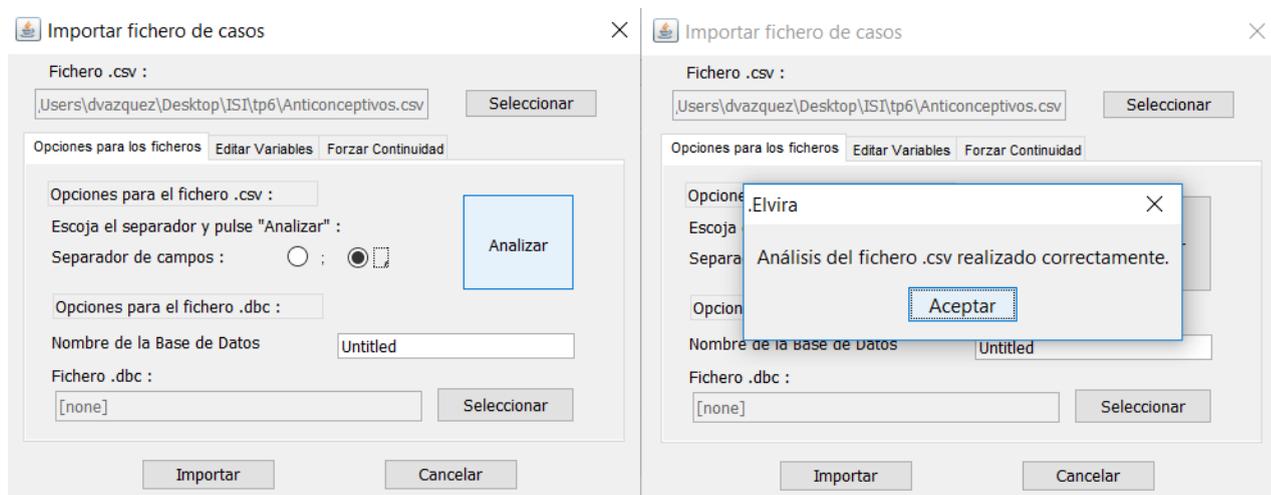
De esta tabla se deduce que cuando la prueba da positivo la mejor decisión es aplicar la terapia (porque conlleva una mayor utilidad esperada: 8.38 frente a 5.66), mientras que cuando da negativo es mejor no aplicar la terapia (la utilidad es 9.845, frente a 8.978).

Cargando datos externos

para el caso que queramos utilizar datos externos de alguna base de datos, en nuestro caso un .csv, iremos a **archivo->importar ficheros de casos**



Al hacer click nos aparecerá una ventana que nos permitirá seleccionar el fichero donde estan los casos. También nos permitirá analizarlo en el caso que este con ';', ' o ', ' :



Dentro de la importación tenemos la pestaña editar variables que nos permite modificar los nombres de las variables para que sus nombres sean más comprensibles.

Fichero .csv :

.Users\dvazquez\Desktop\ISI\tp6\Anticonceptivos.csv Seleccionar

Opciones para los ficheros **Editar Variables** Forzar Continuidad

Elija la variable a editar : Nivel_educativo... ▾
(Debe guardar cambios para cada variable)

Nombre : ativo de esposa Título : Variable 1

Comentario

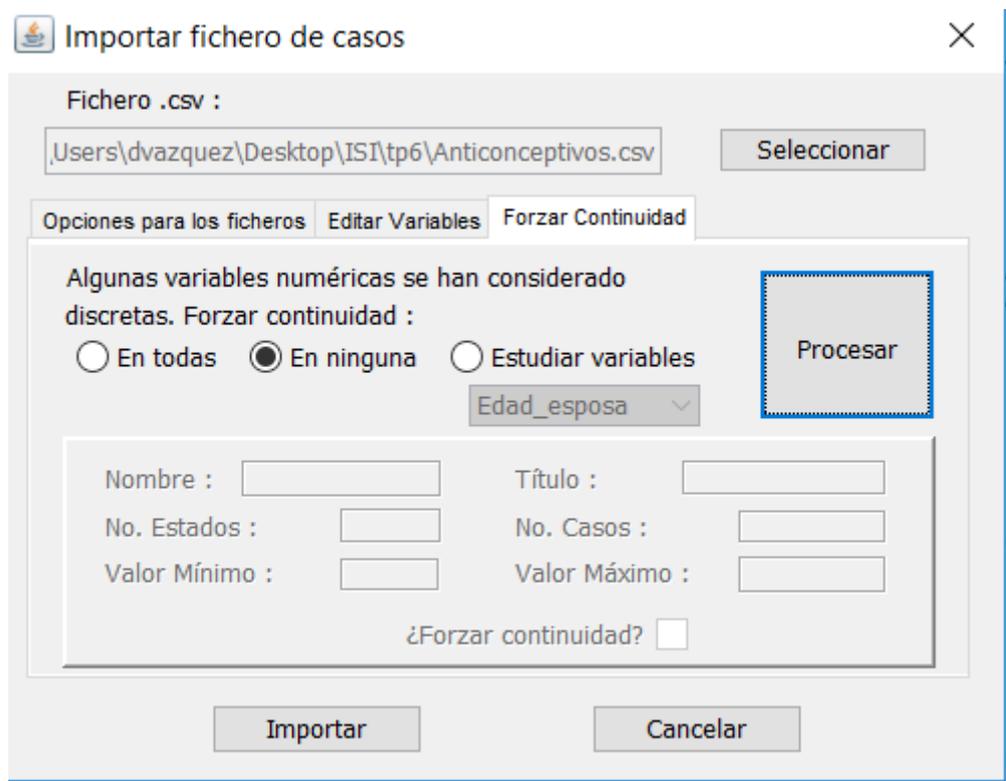
Editar nombres de estados : 2 ▾

Nombre nuevo :

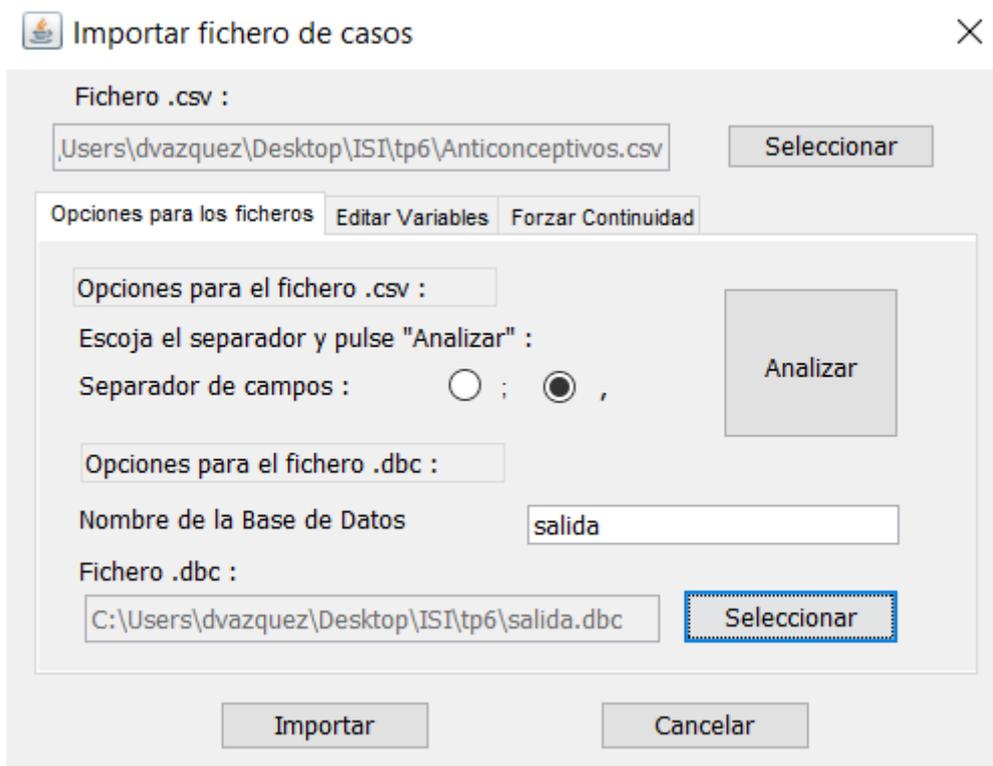
Guardar cambios

Importar Cancelar

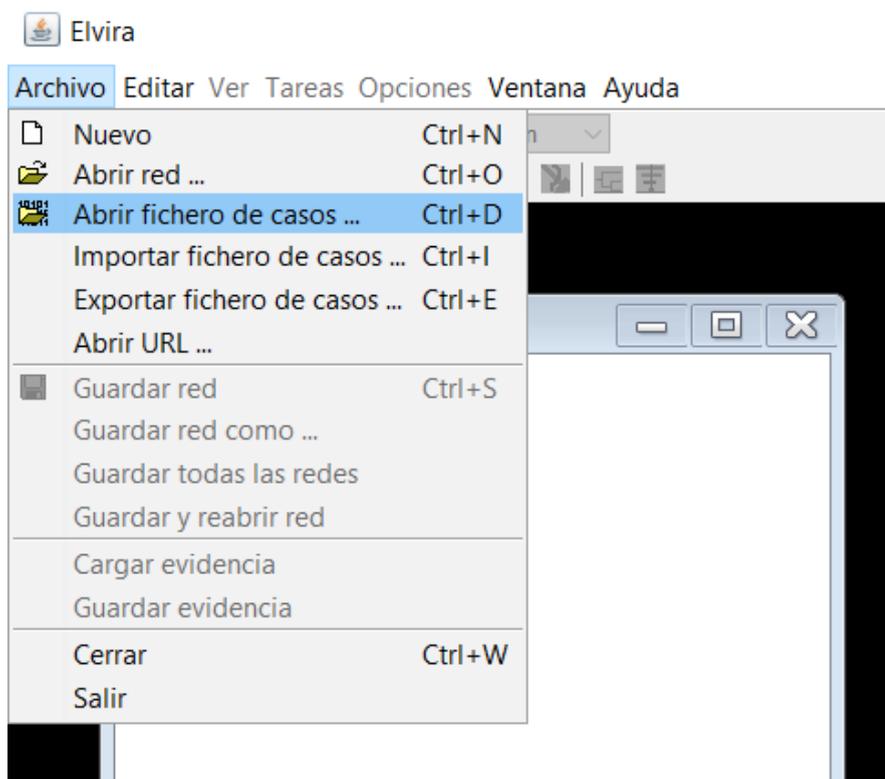
También está la posibilidad de forzar continuidad en el caso de que alguna de las variables tenga algún problema:



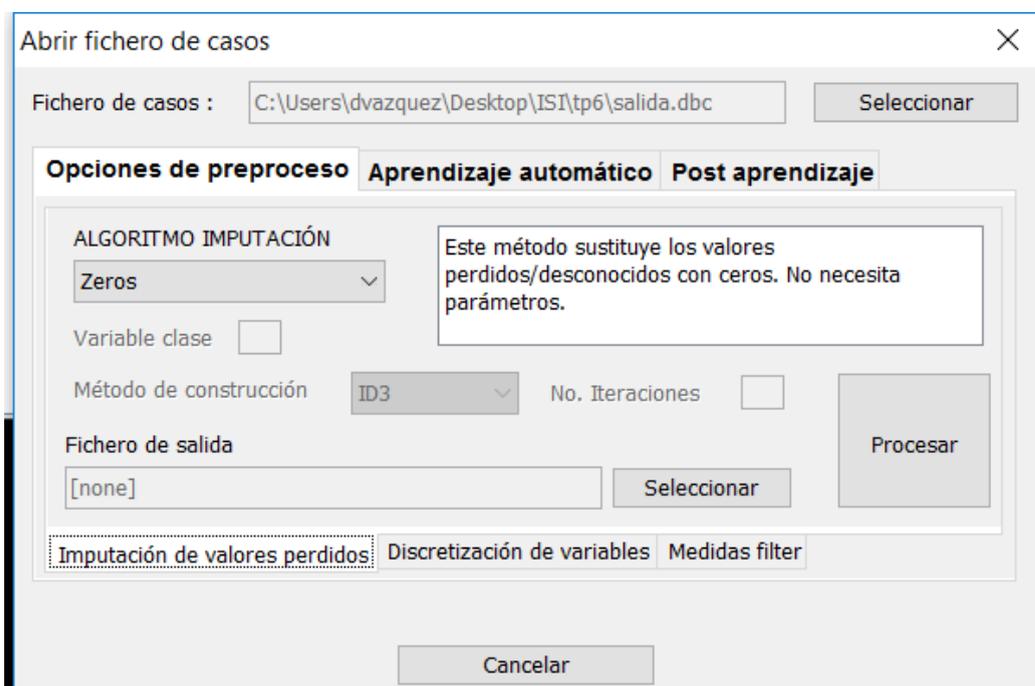
Después de estas definiciones, le damos nombre y ubicación a nuestra base de datos de salida (.dbc) e importamos:



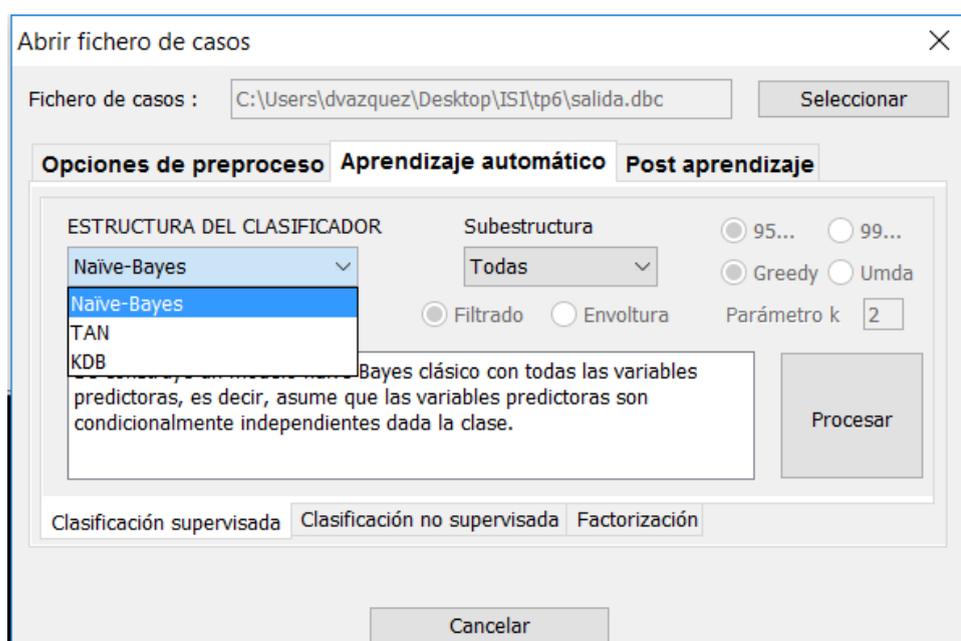
Ya cuando importamos en formato, ahora vamos a abrir la base de datos generada por Elvira, para eso vamos a **archivo->Abrir ficheros de datos**



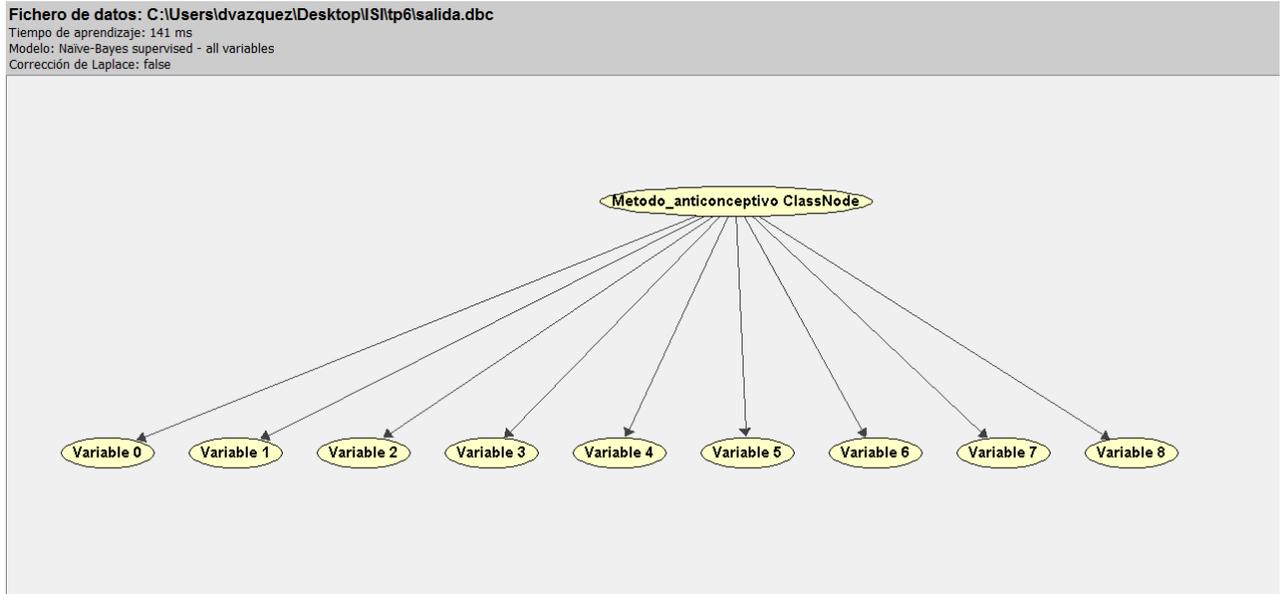
Allí se nos abrirá una ventana donde indicaremos la base de datos anteriormente importada en (.dbc) y de aplicar algoritmos, ya sea para preprocesar, aprendizaje automático y post aprendizaje



Para nuestro caso que vayamos a analizar un aprendizaje automático, nos permitirá utilizar la estructura del clasificador: Naive-Bayes, TAN y KDB. Después de seleccionar el clasificador presionaremos procesar



Y nos aparecera la red en modo edicion para el clasificador que hayamos seleccionado



Después podremos proceder de la misma forma que el apartado **Estructura de la red**