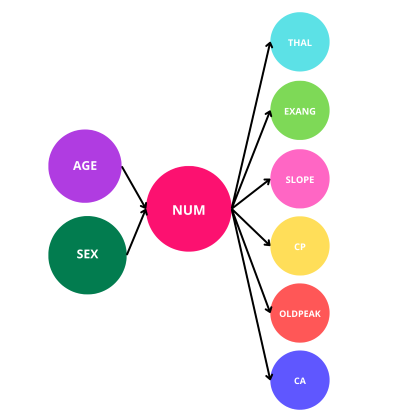
**Reporte Proyecto 2**

***Red Bayesiana:***



***Entrenamiento:***

Para la parte del entrenamiento de nuestro modelo, se dividieron los datos iniciales en 2 grupos: datos de entrenamiento y datos de testeo, para esta sección se usaron los datos de entrenamiento que corresponden al 75% y el restante corresponde a los de testeo. Usando el modelo ya creado en la anterior fase se tomaron las inferencias para analizarlas y meterlas en una lista que guardara sus valores, de forma que posteriormente se analizaran contra los datos reales que corresponden a la base de datos empleada. Con esto se realizó la matriz de confusión comparando estos 2 grupos de datos, obteniendo:

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Vemos que nuestro modelo arroja que en los casos que realmente no había enfermedad, se dijo que no había 15 veces, mientras que 24 se dijo que sí había. En caso contrario, al haber enfermedad hubo un 50% de las veces que se predijo que no había, y el otro porcentaje restante se predijo que no, esto muestra cierta imprecisión en el modelo realizado.

Para tener mayor vista de cómo fue el desempeño del modelo se calcularon 3 indicadores: precisión, exhaustividad(recall) y F1-score:

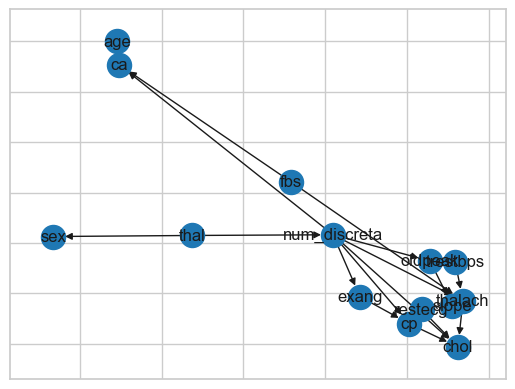
Gráfico, Gráfico de barras

Descripción generada automáticamente

Vemos que la precisión es media, este valor de 43% nos indica que de las predicciones positivas que realizó el modelo, sólo este porcentaje son realmente positivas y el modelo tiene una alta tasa de falsos positivos. Ahora, con respecto al recall del 50%, quiere decir que, de todas las instancias de la clase positiva, sólo la mitad fueron correctamente identificadas por el modelo. Finalmente, en cuanto al F1-Score, que es combinación de precisión y recall, vemos que tenemos una medida de calidad del 46% lo que muestra que el modelo no es muy bueno para la clasificación de fase positiva.

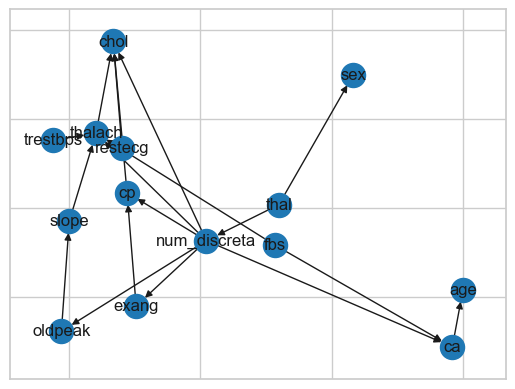
***Evaluación:***

Para la parte de la evaluación, con una parte de los datos se estimó el modelo mediante el método K2Score y HillClimb Search, obteniendo el siguiente:



Como se puede observar el estimado por estos métodos emplea la totalidad de variables de la base de datos, es por ello que se visualizan varias relaciones que permiten realizar las inferencias y demás.

***Comparación:***



Con respecto al modelo realizado por el otro grupo vemos que varía bastante tanto en comparación con el estimado justo antes y el nuestro. Este evalúa muchas más variables con relaciones diferentes, lo que implica diferentes estimaciones y por lo tanto arrojará diferentes diagnósticos.

***Visualizaciones justificación:***

Se tuvieron en cuenta para las visualizaciones mostradas en el dash las recomendaciones según los principios de Gestalt, como se observan se emplearon paletas de colores amigables y en contraste con la forma y diseño de la aplicación, además de poseer ejes claros, descripciones de leyenda y títulos que mostraran de forma clara la información que se quiere comunicar, detalles como centrar la información y demás fueron clave para que el usuario que observa pueda digerir y comprender lo que se desea mostrar.

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente con confianza media

Gráfico, Gráfico de barras

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Se evidencian en las gráficas las buenas prácticas mencionadas previamente. Se escogieron las variables más relevantes teniendo en cuenta la base de datos y el modelo estimado.