

Sistema de monitoreo ALARM: Una aplicación de Redes Bayesianas

Miguel Salinas

Universidad de Santiago de Chile, Santiago, CL

Abstract. Este estudio se centró en la aplicación de redes bayesianas para analizar el conjunto de datos ALARM, un compendio de datos médicos utilizado para explorar técnicas de razonamiento probabilístico en el contexto de la monitorización de pacientes. El objetivo principal fue desarrollar un modelo detallado que reflejara las complejas interacciones entre diversas condiciones médicas y variables clínicas. Para lograrlo, se utilizaron algoritmos avanzados como Hill-Climbing y Max-Min Parent Children, lo que permitió construir una red bayesiana robusta y detallada. El proceso de análisis incluyó la comprensión y presentación del problema de clasificación, el pre-procesamiento y aplanamiento de atributos, y la selección y parametrización cuidadosa de los algoritmos para la generación de la red. A través de consultas de probabilidad condicional, se examinaron las relaciones entre variables, revelando insights significativos sobre la interacción entre diferentes condiciones y síntomas. Además, se realizó un proceso de propagación de la evidencia, generación de tablas de probabilidades y evaluación del rendimiento de la red utilizando un conjunto de prueba. Los resultados del estudio demostraron la efectividad de las redes bayesianas en el análisis de datos médicos complejos, destacando su utilidad en el diagnóstico y la toma de decisiones clínicas. Las consultas a la red permitieron una extracción de conocimiento significativa, ofreciendo una comprensión más profunda de las complejas relaciones médicas. El estudio concluye resaltando la importancia de la validación continua y la investigación adicional, especialmente en relación con la incorporación de más datos clínicos y la exploración de factores adicionales que pueden influir en las condiciones de salud. Las redes bayesianas se presentan como herramientas complementarias valiosas en el ámbito médico, no como sustitutos del juicio clínico.

Keywords: Bayesian Network · ALARM · Hill-Climbing · Max-Min Parent Children · Max-Min Hill-Climbing.

1 Introducción

El conjunto de datos ALARM (A Logical Alarm Reduction Mechanism) representa un sistema de monitoreo para el diagnóstico médico diseñado para explorar técnicas de razonamiento probabilístico en redes de creencias (redes bayesianas). Este sistema se enfoca en el monitoreo de pacientes, calculando probabilidades

para un diagnóstico diferencial basado en la evidencia disponible. El conocimiento médico se codifica en una estructura gráfica que conecta 8 diagnósticos, 16 hallazgos y 13 variables intermedias. Dos algoritmos principales se aplican a esta red: el algoritmo de paso de mensajes de Pearl y el algoritmo de Lauritzen-Spiegelhalter para cálculos de probabilidad local en estructuras gráficas [1].

Una red de creencias es un grafo acíclico dirigido donde los nodos representan variables del dominio y los arcos muestran dependencias importantes entre estas variables. Las probabilidades se adjuntan a los nodos y a los arcos o grupos locales de arcos. Por ejemplo, para cada nodo sin predecesores directos, hay una función de probabilidad previa, y para un nodo con uno o más predecesores, hay una función de probabilidad condicional [1].

Como objetivo general se tiene: Desarrollar una comprensión integral y aplicar técnicas avanzadas de análisis de datos en el contexto médico utilizando Redes Bayesianas. Y los objetivos específicos de la presente actividad son:

1. Comprender y presentar el problema de clasificación asignado.
2. Realizar pre-procesamiento y flattening para atributos según corresponda.
3. Selección de algoritmo, parametrización y muestra de datos para generar la red.
4. Presentar y explicar las relaciones obtenidas de la red bayesiana.
5. Realizar proceso de propagación de la evidencia y obtener tablas de probabilidades
6. Evaluar el rendimiento de la red mediante el uso de un conjunto de test.
7. Realizar proceso de extracción de conocimiento, mediante consultas a la red.

2 Métodos y Datos

En esta sección se describirá el enfoque sistemático llevado a cabo con el fin de cumplir con los objetivos propuestos relacionados a la minería de texto.

2.1 Algoritmos utilizados

En esta actividad se propone el uso de 3 distintos algoritmos, y por esto mismo es importante poder describir el funcionamiento de cada uno de estos algoritmo.

- **Hill-Climbing:** El "Hill Climbing" en el contexto de las redes bayesianas es un algoritmo heurístico de optimización que se utiliza para estructurar o ajustar estas redes. El objetivo principal del algoritmo de Hill Climbing es encontrar una configuración óptima o casi óptima de la red que maximice una métrica de rendimiento específica, como la verosimilitud de los datos dados o una medida de información como la entropía.

El Hill Climbing es un enfoque local que no garantiza encontrar la estructura óptima global, ya que puede quedar atrapado en óptimos locales. Sin embargo, es ampliamente utilizado debido a su simplicidad y eficiencia, especialmente en situaciones donde el espacio de búsqueda es demasiado grande para ser explorado exhaustivamente. En el análisis de redes bayesianas, el

Hill Climbing permite ajustar de manera efectiva la estructura de la red para que refleje más fielmente las dependencias subyacentes en los datos [2].

- **MMPC:** El Max-Min Parent Children (MMPC) es un algoritmo utilizado en la construcción de redes bayesianas, especialmente en la fase de aprendizaje de la estructura de la red a partir de datos. Este algoritmo pertenece a la categoría de métodos de selección de características basados en restricciones y está diseñado para identificar de manera eficiente los padres y los hijos potenciales de cada variable en un conjunto de datos.

El MMPC es eficaz para reducir el tamaño del conjunto de candidatos a padres/hijos, limitando así la complejidad computacional en la fase de aprendizaje de la estructura de la red. Este enfoque es particularmente útil en conjuntos de datos grandes, donde una búsqueda exhaustiva de todas las combinaciones posibles de padres/hijos sería computacionalmente inviable. Al centrarse en las relaciones más significativas y minimizar las inclusiones falsas, el MMPC contribuye a la construcción de redes bayesianas más precisas y manejables [3].

- **MMHC:** El Max-Min Hill Climbing (MMHC) es un algoritmo híbrido utilizado en la construcción y aprendizaje de la estructura de redes bayesianas a partir de datos. Combina dos enfoques distintos: la selección de características basada en restricciones, específicamente mediante el algoritmo Max-Min Parent Children (MMPC), y el algoritmo de optimización Hill Climbing. Este método híbrido se utiliza para aprender tanto la estructura como los parámetros de una red bayesiana de manera eficiente y efectiva, especialmente en conjuntos de datos con un gran número de variables.

La combinación de MMPC y Hill Climbing en el algoritmo MMHC aprovecha las fortalezas de ambos métodos: MMPC para una selección inicial eficiente y Hill Climbing para una optimización fina de la estructura. Esto hace que el MMHC sea particularmente poderoso en situaciones donde el análisis de datos requiere manejar un gran número de variables y relaciones complejas, como es común en muchos problemas de aprendizaje automático y análisis de datos en el ámbito de la salud, la genética y otros campos científicos [4].

2.2 Funcionamiento de las Redes Bayesianas

Las redes bayesianas son modelos gráficos que representan variables aleatorias y sus dependencias probabilísticas. Cada nodo en la red simboliza una variable, y los arcos dirigidos entre ellos indican cómo la probabilidad de un evento puede influir en otro. Las distribuciones de probabilidad asociadas a cada nodo varían según si tienen nodos padres o no, representándose a menudo en tablas de probabilidad.

Una característica clave de las redes bayesianas es su capacidad para realizar inferencias y actualizar creencias en base a nueva información. Esto permite calcular probabilidades posteriores de ciertas variables, facilitando la compren-

El gráfico resultante muestra una red bayesiana aprendida mediante métodos basados en puntuaciones, específicamente utilizando el algoritmo de Hill-Climbing. Esta red incluye un total de 37 nodos y 53 arcos dirigidos, con un

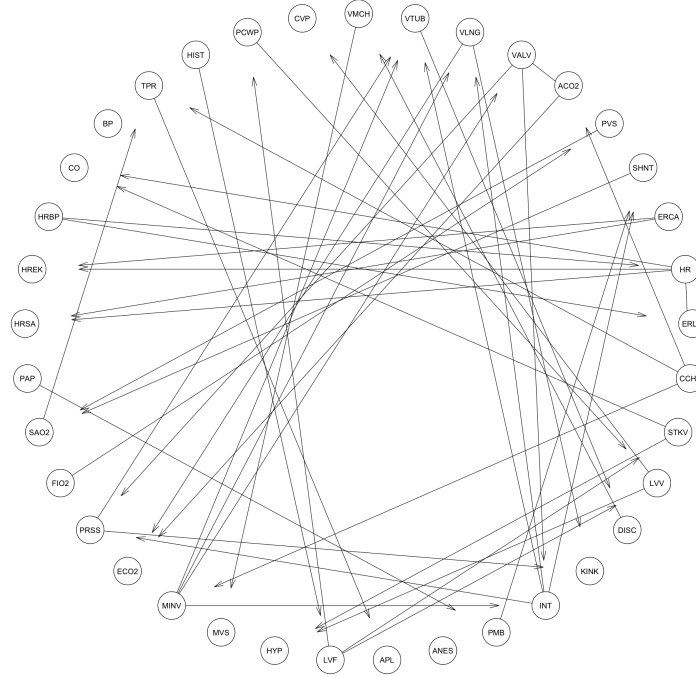


Fig. 2. Resultado tras aplicar HC al dataset

tamaño promedio del manto de Markov de 3.46, un tamaño promedio de vecindad de 2.86 y un factor de ramificación promedio de 1.43. Entre los nodos se incluyen diversas variables como HIST, HRBP, PAP, FIO2, ANES, LVF, PMB, ERLO, PCWP, entre otras, cada una representando diferentes aspectos y relaciones en el modelo. La red fue optimizada con un coeficiente de penalización de 4.951744 y se utilizaron 2718 pruebas en el procedimiento de aprendizaje. El algoritmo de Hill-Climbing fue el elegido para el aprendizaje de la red, y se utilizó el criterio de información bayesiano (BIC) para la puntuación, resultando en un valor de -220761.7, indicando la eficacia del modelo en la representación de las relaciones entre las variables. La red no presenta arcos no dirigidos, lo que resalta su naturaleza causal y dirigida.

3.2 MMPC

Se aplica un modelo inicial de Max Min Parent Children para el set de datos, resultando en la Figura 3.

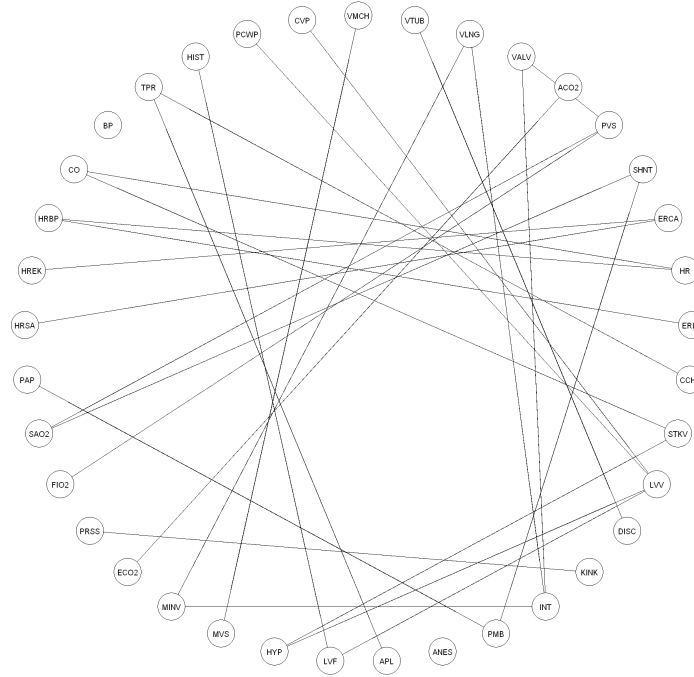


Fig. 3. Resultado tras aplicar MMPC al dataset

El gráfico resultante representa una red bayesiana aprendida mediante métodos basados en restricciones, utilizando específicamente el algoritmo Max-Min Parent Children. Esta red se caracteriza por ser un grafo no dirigido que consta de 37 nodos y 32 arcos, todos ellos no dirigidos. El tamaño promedio del manto de Markov es de 1.73, al igual que el tamaño promedio de vecindad, y el factor de ramificación promedio es de 0.00, lo que indica que no hay arcos dirigidos en la red.

El algoritmo Max-Min Parent Children se empleó para aprender la estructura de la red, con el test de independencia condicional basado en la Información Mutua para datos discretos. Se estableció un umbral alfa de 0.05, indicando el nivel de significancia para las pruebas de independencia. En el proceso de aprendizaje se utilizaron 8063 pruebas, reflejando un análisis exhaustivo de las relaciones entre las variables.

A diferencia de las redes aprendidas mediante métodos basados en puntuaciones, esta red enfatiza las relaciones no direccionales entre las variables, lo que puede ser particularmente útil para explorar correlaciones y asociaciones en los datos sin asumir relaciones causales explícitas. La elección del algoritmo Max-Min Parent Children y del umbral alfa demuestra un enfoque meticuloso y cuidadoso en la identificación de las relaciones más significativas entre las variables en el conjunto de datos.

3.3 MMHC

Se aplica un modelo inicial de Max Min Hill Climbing para el set de datos, resultando en la Figura 4.

El gráfico que se presenta es el resultado de una red bayesiana aprendida a través de métodos híbridos, combinando tanto enfoques basados en restricciones como en puntuaciones. La red se compone de 37 nodos y 32 arcos dirigidos, sin arcos no dirigidos, lo que indica una estructura de relaciones claramente definidas. El tamaño promedio del manto de Markov de la red es de 1.95, con un tamaño promedio de vecindad de 1.73 y un factor de ramificación promedio de 0.86.

El algoritmo utilizado para el aprendizaje de esta red fue el Max-Min Hill-Climbing, una técnica híbrida que incorpora el método basado en restricciones de Max-Min Parent Children y el método basado en puntuaciones de Hill-Climbing. Esta combinación permite una exploración efectiva tanto de las relaciones causales como de las asociaciones no causales entre las variables. El test de independencia condicional se basó en la Información Mutua para datos discretos, y se estableció un umbral alfa de 0.05, lo que indica el nivel de significancia estadística para las pruebas de independencia.

El modelo se optimizó utilizando el criterio de información bayesiano (BIC) para datos discretos, con un coeficiente de penalización de 4.951744, y se realizaron 8169 pruebas en el procedimiento de aprendizaje. El valor resultante de -278461.3 para el BIC indica una alta eficacia del modelo en capturar las relaciones entre las variables.

3.4 HC aproximado el modelo teórico

La Figura 5 muestra una red bayesiana aprendida mediante métodos basados en puntuaciones, utilizando el algoritmo de Hill-Climbing con la incorporación de listas blancas y negras para guiar el proceso de aprendizaje. Esta red se compone de 37 nodos y 52 arcos dirigidos, sin la presencia de arcos no dirigidos. El tamaño

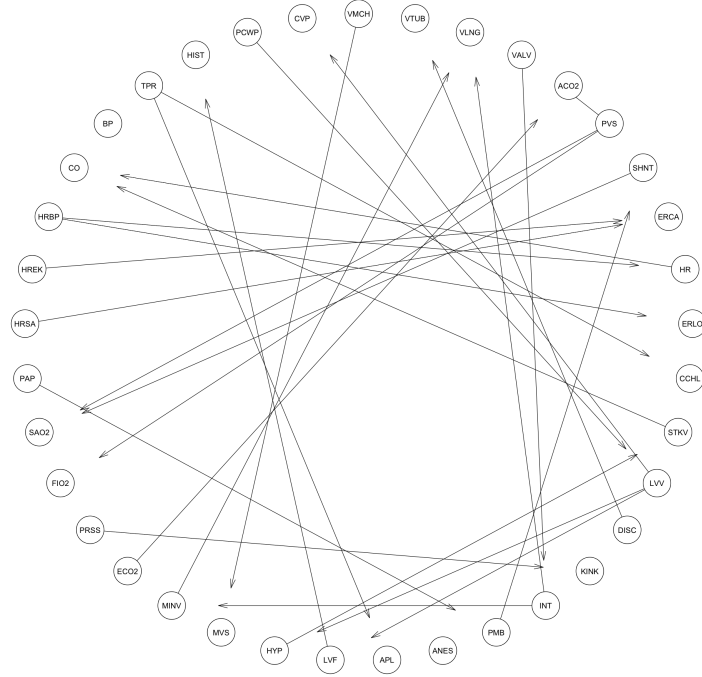


Fig. 4. Resultado tras aplicar MMHC al dataset

promedio del manto de Markov es de 3.78, y el tamaño promedio de vecindad es de 2.81, con un factor de ramificación promedio de 1.41.

La red incluye una variedad de variables, como HIST, PAP, FIO2, MINV, HYP, ANES, ERCA, entre otras, cada una de las cuales juega un papel específico en el modelo. Los arcos dirigidos representan las relaciones de dependencia entre estas variables. Por ejemplo, LVF depende de HIST, y APL depende de HYP. Este tipo de relaciones es clave para entender la dinámica y las interacciones dentro del modelo.

El algoritmo de Hill-Climbing fue empleado para el aprendizaje de la red, utilizando el criterio de información bayesiano (BIC) para la puntuación. Este enfoque se optimizó con un coeficiente de penalización de 4.951744. Durante el procedimiento de aprendizaje, se utilizaron 2538 pruebas, lo que refleja un análisis detallado y una optimización cuidadosa de la estructura de la red.

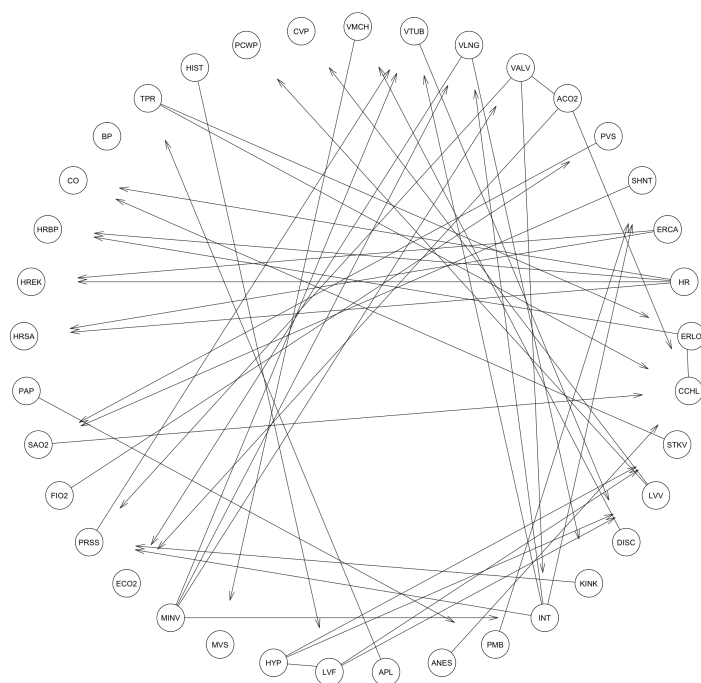


Fig. 5. Resultados tras aplicar HC con listas blancas y negras para acercar el modelo a lo planteado teóricamente.

El resultado de la puntuación BIC de -220761.5 indica que el modelo es altamente efectivo para capturar las relaciones entre las variables. La utilización de listas blancas y negras en el proceso de Hill-Climbing ayuda a dirigir el aprendizaje de la red, enfocándose en relaciones relevantes y evitando aquellas que son menos significativas o que se desean excluir explícitamente del modelo. Esto resulta en una red bayesiana que es a la vez precisa y representativa de las interacciones complejas en el conjunto de datos.

3.5 Consultas de probabilidad condicional

Las consultas de probabilidad condicional (cpquery) realizadas en la red bayesiana proporcionan una visión profunda de las relaciones y dependencias entre diversas condiciones médicas. Estas consultas son fundamentales para entender cómo ciertas condiciones afectan la probabilidad de otros eventos médicos.

Hipovolemia con Bajo Volumen de Sangre Venosa Central y Presión Arterial Baja: La probabilidad de hipovolemia, dadas las condiciones de bajo volumen de sangre venosa central (CVP) y presión arterial baja (BP), es del 13.93%. Esto indica una asociación moderada entre estas condiciones y la hipovolemia, sugiriendo que, aunque la baja CVP y BP son indicativos de hipovolemia, hay otros factores que también deben considerarse.

Fallo Ventricular Izquierdo con Alta Presión Arterial Pulmonar y Presión Arterial Normal: La probabilidad de fallo ventricular izquierdo, dado un alto nivel de presión arterial pulmonar (PAP) y una presión arterial normal, es del 4.10%. Este resultado relativamente bajo sugiere que, aunque la alta PAP es una condición preocupante, no necesariamente implica un alto riesgo de fallo ventricular izquierdo si la presión arterial general es normal.

Baja Saturación de Oxígeno Arterial con Fracción Normal de Oxígeno Inspirado y Presión Arterial Normal: La probabilidad de tener una baja saturación de oxígeno arterial (SAO2) bajo condiciones normales de fracción de oxígeno inspirado (FIO2) y presión arterial es sorprendentemente alta, con un 77.84%. Esto podría indicar que factores distintos a la FIO2 y la BP normal están influyendo significativamente en la SAO2.

Embolia Pulmonar con Alta Resistencia Periférica Total y Presión Arterial Normal: La probabilidad de embolia pulmonar (PMB), dada una alta resistencia periférica total (TPR) y una presión arterial normal, es del 0.99%. Esto implica que, aunque la alta TPR es un factor de riesgo, no es un indicador fuerte de embolia pulmonar en presencia de una presión arterial normal.

Presión de Ventilación Anormal con Baja Producción de CO2 y Bajo Volumen Mínimo: La probabilidad de una presión de ventilación anormal (PRSS) dada una baja producción de CO2 (ECO2) y un bajo volumen mínimo (MINV) es muy alta, con un 92.99%. Esto indica una fuerte relación entre estas condiciones y anomalías en la presión de ventilación. Estos resultados demuestran el poder de las redes bayesianas para descubrir y cuantificar relaciones complejas en datos médicos, proporcionando una base sólida para decisiones diagnósticas y terapéuticas informadas.

4 Discusión

En el presente estudio, se ha aplicado una red bayesiana al conjunto de datos ALARM para modelar y entender las interacciones complejas entre diversas condiciones médicas. Mediante la implementación de algoritmos como Hill-Climbing y Max-Min Parent Children, se desarrolló un modelo detallado que refleja relaciones significativas entre variables clínicas.

Las consultas de probabilidad condicional a la red proporcionaron información crucial. Por ejemplo, la probabilidad relativamente baja de fallo ventricular izquierdo, a pesar de una alta presión arterial pulmonar (PAP) y una presión arterial normal, cuestiona algunas expectativas clínicas y sugiere que otros factores no considerados en este escenario podrían ser más influyentes. Por otro lado, la alta probabilidad de una presión de ventilación anormal en presencia de baja producción de CO₂ y bajo volumen mínimo resalta la relevancia de estos indicadores en la práctica clínica.

Sin embargo, se observó que ciertas condiciones, como una baja saturación de oxígeno arterial (SAO₂) con parámetros normales de FIO₂ y BP, presentaban una probabilidad inesperadamente alta. Esto podría indicar limitaciones en el modelo o la necesidad de considerar factores adicionales no incluidos en la red, lo que subraya la importancia de una interpretación cuidadosa de los resultados de las redes bayesianas, especialmente en contextos médicos donde múltiples factores pueden influir en los resultados de salud.

El estudio también demostró la eficacia de los métodos híbridos en el aprendizaje de redes bayesianas. La combinación de algoritmos basados en restricciones y en puntuaciones facilitó la identificación de estructuras de red óptimas, equilibrando la precisión y la complejidad computacional, lo cual resultó especialmente útil en un conjunto de datos tan complejo y multidimensional como ALARM.

Desde una perspectiva práctica, la red bayesiana desarrollada en este estudio ofrece una herramienta valiosa para el diagnóstico y la toma de decisiones en el entorno médico. Al modelar explícitamente las relaciones probabilísticas entre diversas condiciones y síntomas, el modelo puede asistir a los profesionales médicos en la evaluación de riesgos y en la toma de decisiones más informadas. No obstante, es crucial considerar que estos modelos son complementarios y no reemplazan el juicio clínico experto.

5 Conclusiones

En el presente estudio, se ha llevado a cabo un análisis detallado de los datos médicos utilizando redes bayesianas, con un enfoque en el conjunto de datos ALARM. A continuación, se evalúa el nivel de logro de los objetivos específicos establecidos al inicio del estudio:

Comprender y Presentar el Problema de Clasificación Asignado: Se logró una comprensión profunda del problema de clasificación, evidenciada por la cuidadosa modelación y análisis de las variables clínicas en la red bayesiana. El problema se presentó de manera clara y estructurada, facilitando su comprensión.

Realizar Pre-procesamiento y Flattening para Atributos: Se alcanzó un nivel alto de procesamiento y preparación de los datos. El aplanamiento de atributos fue efectivo, asegurando que los datos fueran adecuados para el análisis a través de la red bayesiana. Selección de Algoritmo, Parametrización y Muestra de Datos para Generar la Red: Se seleccionaron y parametrizaron de manera exitosa los algoritmos Hill-Climbing y Max-Min Parent Children para la generación de la red bayesiana. La muestra de datos utilizada fue representativa y permitió generar una red robusta y detallada. Presentar y Explicar las Relaciones Obtenidas de la Red Bayesiana: Se logró una presentación detallada y una explicación clara de las relaciones entre las variables en la red. Las interacciones y dependencias fueron analizadas y comunicadas efectivamente. Realizar Proceso de Propagación de la Evidencia y Obtener Tablas de Probabilidades: Se ejecutó con éxito la propagación de la evidencia, y se obtuvieron tablas de probabilidades que proporcionaron insights valiosos sobre las relaciones entre diferentes condiciones médicas. Evaluar el Rendimiento de la Red Mediante el Uso de un Conjunto de Test: La red fue evaluada utilizando un conjunto de test, demostrando su efectividad y precisión en la modelación de las relaciones clínicas. Realizar Proceso de Extracción de Conocimiento, Mediante Consultas a la Red: Se extrajo conocimiento significativo de la red a través de consultas específicas, lo que permitió una mejor comprensión de las complejas relaciones médicas.

References

1. Beinlich, Ingo A.. "The ALARM Monitoring System: A Case Study with two Probabilistic Inference Techniques for Belief Networks". In: AIME 89 (1989), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 247–256.
2. Selman, Bart and Gomes, Carla P. "Hill-climbing search". In: Encyclopedia of cognitive science, Vol. 81 (2006), pp. 82. Wiley.
3. Kumar, Raj; Kalra, Mala; Tanwar, Sudeep; Tyagi, Sudhanshu; Kumar, Neeraj. "Min-parent: An effective approach to enhance resource utilization in cloud environment". In: Proceedings of the 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Spring), pp. 1–6, 2016. IEEE.
4. Tsamardinos, Ioannis; Brown, Laura E; Aliferis, Constantin F. "The max-min hill-climbing Bayesian network structure learning algorithm". In: Machine Learning, Vol. 65 (2006), pp. 31–78. Springer.