



# UNIVERSIDAD NACIONAL DEL ALTIPLANO

## ESCUELA DE POSGRADO

### DOCTORADO EN ESTADÍSTICA APLICADA



TESIS

#### COMPARACIÓN Y CLASIFICACIÓN DE LAS TÉCNICAS DE APRENDIZAJE ESTRUCTURAL DE LAS REDES BAYESIANAS

PRESENTADA POR:

ANGEL JAVIER QUISPE CARITA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTORIS SCIENTIAE EN ESTADÍSTICA APLICADA

PUNO, PERÚ

2023

Reporte de similitud

NOMBRE DEL TRABAJO

**COMPARACION Y CLASIFICACION DE LAS TECNICAS DE APRENDIZAJE ESTRUCTURAL DE LAS REDES BAYESIANAS**

AUTOR

**Angel Javier Quispe Carita**

RECUENTO DE PALABRAS

**19901 Words**

RECUENTO DE CARACTERES

**106841 Characters**

RECUENTO DE PÁGINAS

**93 Pages**

TAMAÑO DEL ARCHIVO

**882.9KB**

FECHA DE ENTREGA

**Dec 29, 2023 2:12 PM GMT-5**

FECHA DEL INFORME

**Dec 29, 2023 2:13 PM GMT-5**

● **3% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos es:

- 3% Base de datos de Internet
- Base de datos de Crossref
- 2% Base de datos de trabajos entregados
- 0% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 10 palabras)



  
Dr. VLADIMIRO IBANEZ QUISPE  
DOCENTE PRINCIPAL - UNA - P  
CIP. 47043



# UNIVERSIDAD NACIONAL DEL ALTIPLANO

## ESCUELA DE POSGRADO

### DOCTORADO EN ESTADÍSTICA APLICADA

TESIS

COMPARACIÓN Y CLASIFICACIÓN DE LAS TÉCNICAS DE APRENDIZAJE  
ESTRUCTURAL DE LAS REDES BAYESIANAS



PRESENTADA POR:

ANGEL JAVIER QUISPE CARITA

PARA OBTAR EL GRADO ACADÉMICO DE:

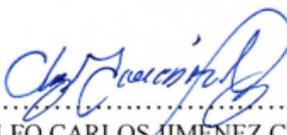
DOCTORIS SCIENTIAE EN ESTADÍSTICA APLICADA

APROBADA POR EL JURADO SIGUIENTE:

PRESIDENTE

  
.....  
D. Sc. LEONEL COYLA IDME

PRIMER MIEMBRO

  
.....  
D.Sc. ADOLFO CARLOS JIMENEZ CHURA

SEGUNDO MIEMBRO

  
.....  
D.Sc. JUAN CARLOS JUAREZ VARGAS

ASESOR DE TESIS

  
.....  
Dr. VLADIMIRO IBÁÑEZ QUISPE

Puno, 19 de diciembre de 2023

ÁREA: Inteligencia Artificial

TEMA: Comparación y clasificación de las técnicas de aprendizaje estructural de las redes bayesianas

LINEA: Técnicas Multivariadas Supervisadas



## DEDICATORIA

*A mi amada esposa, cuyo apoyo  
incondicional y amor me han sostenido en los momentos más desafiantes  
y a mi hija, luz y motivación constante en mi vida.*

*Angel.*



## AGRADECIMIENTOS

En primer lugar deseo agradecer a Dios fuente inagotable de fortaleza y guía en cada paso de este viaje.

A mi asesor Dr. Vladimiro Ibañez Quispe, por la orientación constante, en el desarrollo personal y académico y la confianza puesto en mí para la conclusión de este trabajo.

A mis jurados, Dr. Leonel Coyla Idme, Dr. Adolfo Carlos Jimenez Chura, Dr. Juan Carlos Juarez Vargas que contribuyeron con sus valiosos aportes para el refinamiento y consolidación de este trabajo de investigación.

Finalmente, deseo expresar mi gratitud a la Escuela de Posgrado de la Universidad Nacional del Altiplano por brindarme la oportunidad de crecer tanto académica como personalmente en sus instalaciones.



## ÍNDICE GENERAL

	<b>Pág.</b>
DEDICATORIA	i
AGRADECIMIENTOS	ii
ÍNDICE GENERAL	iii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS	vii
ÍNDICE DE ANEXOS	viii
ÍNDICE DE ACRÓNIMOS	ix
RESUMEN	x
ABSTRACT	xi
INTRODUCCIÓN	1

### CAPÍTULO I

#### REVISIÓN DE LITERATURA

1.1	Marco teórico	3
1.1.1	Redes Bayesianas	3
1.1.2	Inferencia bayesiana	10
1.1.3	Aprendizaje bayesiano	12
1.1.4	Aprendizaje Paramétrico	12
1.1.5	Aprendizaje Estructural	14
1.1.6	Métodos basados en restricción	15
1.1.7	Métodos basados en puntuación	18
1.1.8	Métodos híbridos	21
1.2	Antecedentes	23
1.2.1	Trabajos sobre Aprendizaje Estructural	23

### CAPÍTULO II

#### PLANTEAMIENTO DEL PROBLEMA

2.1	Identificación del problema	28
-----	-----------------------------	----

...



2.2	Enunciados del problema	29
2.3	Justificación	30
2.4	Objetivos	31
2.4.1	Objetivo general	31
2.4.2	Objetivos específicos	31
2.5	Hipótesis	31

### **CAPÍTULO III**

#### **MATERIALES Y MÉTODOS**

3.1	Ámbito o Lugar de estudio	32
3.2	Población	34
3.3	Muestra	35
3.4	Método de Investigación	35
3.5	Descripción de las técnicas de aprendizajes estructural aplicados	36
3.5.1	Grow-Shrink (GS)	36
3.5.2	Incremental Association Markov Blanket (IAMB)	36
3.5.3	Hill Climbing (HC)	37
3.5.4	Tabu Search (TABU)	38
3.5.5	Max-Min Hill Climbing (MMHC)	38
3.5.6	Two-phase Restricted Maximization (RSMAX2)	39
3.6	Métricas	40
3.6.1	Validación Cruzada	40
3.6.2	Criterio de Información Bayesiano (Bayesian Information Criterion, BIC)	41
3.6.3	Pérdida Esperada (Expected Loss)	43
3.6.4	Hardware y Software	44

### **CAPÍTULO IV**

#### **RESULTADOS Y DISCUSIÓN**

4.1	Aplicación de los algoritmos de aprendizaje Estructural	47
4.1.1	Aplicación del algoritmo Grow-Shrink (GS)	47

...



4.1.2	Aplicación del algoritmo Incremental Association Markov Blanket (IAMB)	49
4.1.3	Aplicación del algoritmo Hill Climbing (HC)	51
4.1.4	Aplicación del algoritmo Tabu Search (TABU)	53
4.1.5	Aplicación del algoritmo Max-Min Hill Climbing (MMHC)	55
4.1.6	Aplicación del algoritmo Two-Phase Restricted Maximization (RSMAX2)	57
4.2	Comparación de las técnicas de aprendizaje de estructuras de una BN	59
4.3	Clasificación de las técnicas de aprendizaje	62
	CONCLUSIONES	67
	RECOMENDACIONES	68
	BIBLIOGRAFÍA	69
	ANEXOS	75



## ÍNDICE DE TABLAS

	<b>Pág.</b>
1. Descripción de las variables en el dataset Asia	35
2. Resumen de la Red Bayesiana aprendida aplicando el algoritmo Grow-Shrink (GS)	48
3. Resumen de la Red Bayesiana aprendida aplicando el Incremental Association Markov Blanket (IAMB)	49
4. Resultado del Aprendizaje de Redes Bayesianas con Método Basado en Hill Climbing (HC)	52
5. Resultado del Aprendizaje de Redes Bayesianas con Métodos Basados Tabu Search (TABU)	54
6. Resultado del Aprendizaje de Redes Bayesianas mediante el Método Max-Min Hill Climbing (MMHC)	56
7. Resultado del Aprendizaje de Redes Bayesianas mediante el Método Two-Phase Restricted Maximization (RSMAX2)	58
9. Resultados del valor BIC en cada fold de la validación cruzada	60
8. Comparación de diferentes técnicas de aprendizaje estructural.	61
10. Análisis de varianza para las técnicas de aprendizaje estructural	62
11. Clasificación de técnicas de aprendizaje estructural basados en BIC aplicando Tukey	63
12. Perdida esperada de la función de perdida log-verosimilitud disc. en la validación cruzada de los métodos de aprendizaje estructural de una BN	65
13. Valores BIC de k-folds para diferentes técnicas de aprendizaje (tratamientos)	80

## ÍNDICE DE FIGURAS

	<b>Pág.</b>
1. Estructuras básicas existentes en la teoría de grafos.	6
2. Ejemplo de una Red Bayesiana con seis variables	8
3. Estimación de estructura: Ilustración del algoritmo PC	17
4. Algoritmo ascenso de colina (Hill Climbing)	21
5. Estimación de estructura: abordaje basada en puntuaciones	21
6. Red Objetivo del problema ASIA	33
7. Validación Cruzada de 10-Folds.	41
8. Representación de DAG aplicando el algoritmo Grow-Shrink (GS)	47
9. Representación de DAG aplicando el algoritmo Incremental Association Markov Blanket (IAMB)	49
10. Representación de DAG aplicando el algoritmo Hill Climbing (HC)	51
11. Representación de DAG aplicando el algoritmo Tabu Search (TABU)	53
12. Representación de DAG aplicando el algoritmo Aplicación del algoritmo Max-Min Hill Climbing (MMHC)	55
13. Representación de DAG aplicando el algoritmo Aplicación del algoritmo Two-Phase Restricted Maximization (RSMAX2)	57
14. Diagrama de Cajas de los valores BIC generadas en cada fold de la Validación Cruzada	62
15. Clasificación de las técnicas de aprendizaje estructural aplicando Tukey	64
16. Aplicación Overleaf para la redacción del documento científico	79



## ÍNDICE DE ANEXOS

	<b>Pág.</b>
1. Dataset ASIA	76
2. codificación ejecutada en R para el aprendizaje estructural	77
3. Generación de documento con Overleaf	79
4. Conjunto de datos BIC para la realización de comparación de medias	80



## ÍNDICE DE ACRÓNIMOS

**AI** Inteligencia Artificial (Artificial intelligence )

**BN** Red Bayesiana (Bayesian Network)

**DAG** Gráfo Acíclico Dirigido (Directed Acyclic Graph)

**BIC** Criterio de Información Bayesiano (Bayesian Information Criterion)

**GS** Grow-Shrink (Grow-Shrink)

**IAMB** Manta de Markov con Asociación Incremental ( Incremental Association Markov  
Blanket)

**HC** Hill-Climbing

**TABU** Tabu Search

**MMHC** Max-Min Hill-Climbing

**RSMAX2** Two-Phase Restricted Maximization



## RESUMEN

Actualmente, la Inteligencia Artificial (AI) es una de las áreas más estudiadas, en la que se aplican técnicas y métodos en escenarios de incertidumbre y adquisición de conocimiento para la toma de decisiones. Entre las técnicas de AI, destacan las Redes Bayesianas (BNs), que se implementan en dos fases: aprendizaje paramétrico y estructural. La fase de aprendizaje estructural es especialmente compleja y requiere la intervención de un experto, que en muchas situaciones no está disponible, por lo que surgen innovadoras técnicas de aprendizaje estructural, que es el punto de partida de esta investigación cuyo objetivo fue comparar y clasificar las técnicas de aprendizaje estructural de la Red Bayesiana según su desempeño. La investigación fue de tipo comparativo y cuantitativo, seleccionándose técnicas de aprendizaje estructural utilizadas ampliamente en el campo de las redes bayesianas, aplicado a una base de datos que es un benchmark estándar bastante difundido en la literatura bayesianas, para cada técnica se aplicó la validación cruzada, Criterio de Información Bayesiano (Bayesian Information Criterion, BIC), Pérdida esperada y ANOVA encontrándose que las técnicas Hill Climbing (HC) y Tabu Search (TABU) destacaron del resto, obteniendo un valor BIC medio (-1168.277) que las ubicó por encima de sus competidoras. Las técnicas se clasificaron en dos grupos a y b, la clasificación y comparación proporciona una guía útil para los investigadores y profesionales que buscan seleccionar la técnica más adecuada para sus propias investigaciones o aplicaciones.

**Palabras clave:** Redes Bayesianas, aprendizaje estructural, Hill Climbing, Tabu Search, BIC



## ABSTRACT

Currently, Artificial Intelligence (AI) is one of the most studied areas, in which techniques and methods are applied in scenarios of uncertainty and knowledge acquisition for decision making. Among AI techniques, Bayesian Networks (BNs) stand out, which are an extension of classical probabilistic, decision theory models. The task of the structural and parametric learning of a Bayesian Network is especially complex and requires the intervention of an expert, often not available, hence the emergence of innovative techniques for structural learning, which is the starting point of this research whose objective is to compare and classify the techniques of structural learning of the Bayesian Network according to their performance. The research was comparative and quantitative and have been selected the structural learning techniques widely used in the field of bayesian networks, applied to a data sets with benchmarking standards from the bayesian literature, for each technique was applied cross-validation, Bayesian Information Criterion (BIC), Expected Loss, and ANOVA, finding that the Hill Climbing (HC) and Tabu Search (TABU) techniques stood out from the rest, obtaining a BIC average value (-1168.277) that placed it above its competitors. The techniques were classified into two groups a and b, and the classification and comparison provide a useful guide for researchers and professionals seeking to select the most appropriate technique for their own investigations or applications.

**Keywords:** Bayesian Networks, structural learning, Hill Climbing, Tabu Search, BIC

## INTRODUCCIÓN

En los últimos años la Inteligencia Artificial (Artificial intelligence, AI) ha tenido un gran impacto en la manera como interactuamos con la tecnología y la forma como abordamos problemas complejos con aplicaciones en múltiples dominios, desde atención médica hasta logística y marketing. La AI está tornando radicalmente la manera en que resolvemos problemas, analizamos datos y realizamos descubrimientos (Russell & Norvig, 2016).

En la AI, se encuentra dos principales categorías: el aprendizaje simbólico y el aprendizaje sub simbólico, siendo el simbólico dependiente de lógicas y algoritmos explícitos, en cuanto al sub simbólico, incluye técnicas como las redes neuronales, basado en el aprendizaje a partir de los datos, reproduciendo la manera en que funciona el cerebro humano (Goodfellow *et al.*, 2016).

Una técnica destacada en el aprendizaje simbólico son las Red Bayesiana (Bayesian Network, BN), que son modelos probabilísticos, representados gráficamente, que muestran dependencias condicionales entre un conjunto de variables aleatorias mediante grafos dirigidos acíclicos (Pearl, 1988) tales modelos han sido aplicados en distintos áreas del conocimiento y han demostrado ser extremadamente eficientes en el manejo de la incertidumbre y la causalidad.

El aprendizaje en las Redes Bayesianas es dividido en dos principales aspectos: el aprendizaje paramétrico y estructural, el aprendizaje paramétrico se centra en la estimación de los parámetros de la red bayesiana una vez conocida la estructura de la red, mientras que el aprendizaje estructural se enfoca en encontrar la mejor estructura de la red que se ajusta a la relación causal de los datos (Koller & Friedman, 2009).

El aprendizaje estructural presenta desafíos significativos, puesto que es una tarea NP-Hard, reto computacional, empero, una variedad de técnicas y abordajes han sido desarrollados que en el presente trabajo de investigación se analizará.

El trabajo de investigación está organizado de la siguiente manera: en el capítulo I, se detallan los conceptos que sustentan el trabajo, y se presentan los antecedentes relacionados al estado del arte; en el capítulo II, se desarrollan el problema, objetivos e hipótesis; en el capítulo III, se describen el diseño de la investigación, materiales y métricas utilizadas, así como detalles sobre el *dataset* compuesto; en el capítulo IV, se presentan los resul-



tados organizados en dos grupos: cuantitativos y cualitativos representados en tablas y figuras. Finalmente, se hace mención a las conclusiones y recomendaciones a considerar para futuras investigaciones.

## CAPÍTULO I

### REVISIÓN DE LITERATURA

El capítulo se realizó la revisión literario respecto al ambito de las Redes Bayesianas y las técnicas de aprendizaje que estas redes presentan conforme se han ido desarrollando, tanto el aprendizaje paramétrico y específicamente el aprendizaje estructural y los algoritmos que intervienen. El capítulo pretende ser una revisión exhaustiva que no solo cubra la teoría básica sino que también profundiza en los métodos y algoritmos más avanzados, proporcionando un panorama completo sobre las Redes Bayesianas y su aprendizaje

#### 1.1 Marco teórico

##### 1.1.1 Redes Bayesianas

Los modelos gráficos probabilísticos son representaciones gráficas de la dependencia condicional entre variables aleatorias. Esta relación probabilística se expresa a través de gráficos. Tales modelos han sido algunos de los principales intereses del área de aprendizaje de máquina en las últimas dos décadas (Ghahramani, 2015). Entre estos modelos, las redes bayesianas - también conocidas como redes de creencia, redes causales y gráficos de dependencia probabilística - son representadas por un tipo específico de grafo denominado Gráfo Acíclico Dirigido (Directed Acyclic Graph, DAG). Así, son modelos gráficos que representan, de forma visual, las dependencias probabilísticas entre un conjunto de variables aleatorias (Nagarajan *et al.*, 2013)

Las Redes Bayesianas son una de las familias clásicas de modelos gráficos dirigidos, junto con los Campos Aleatorios de Markov, que son una familia de modelos gráficos no dirigidos. Ambas familias de modelos gráficos son fundamentales en la teoría de

los Modelos Gráficos Probabilísticos y son aplicadas en varios dominios, incluyendo la inferencia causal y los métodos generativos modernos (Risteski, 2022).

Las Redes Bayesianas tienen una amplia gama de aplicaciones en diversos campos como la bioinformática, la medicina, la ingeniería, la clasificación, la fusión de datos y los sistemas de apoyo a la toma de decisiones debido a su interpretación natural y a su capacidad para modelar relaciones probabilísticas complejas entre variables. Además, los métodos probabilísticos asociados con las redes Bayesianas, como los métodos de muestreo, el método de maximización de la expectativa, el método variacional Bayesiano, y el método de mezcla, son cruciales para la inferencia y el aprendizaje de los modelos. Tales modelos han sido alguno de los principales intereses del área de aprendizaje de máquina en las dos últimas décadas (Ghahramani, 2015).

Las Redes Bayesianas pueden ser utilizadas para construir modelos a partir de datos y/o opiniones de expertos, y son especialmente útiles en tareas que incluyen diagnósticos, razonamiento, modelado causal, toma de decisiones bajo incertidumbre, detección de anomalías, *insight* automatizado y predicción (Bayes Server, 2023). Algunas aplicaciones actuales se extienden a las áreas como diagnóstico de fallas (Cai *et al.*, 2017; Quispe *et al.*, 2013), agricultura (Drury *et al.*, 2017; Ianishi *et al.*, 2020), previsión de ataques cibernéticos (Okutan *et al.*, 2017), desastres y evaluación de impactos ambientales y cambios climáticos (Sperotto *et al.*, 2019; Yin *et al.*, 2021), problemas de tránsito (Afrin & Yodo, 2021), así como en la asistencia y diagnóstico médicos (Arora *et al.*, 2019; McLachlan *et al.*, 2020; Zagorecki *et al.*, 2013).

La teoría de las redes bayesianas se basa exhaustivamente en el Teorema de Bayes, expuesto en la próxima sección.

### **Teorema de Bayes**

El Teorema de Bayes es un principio fundamental en la teoría de la probabilidad y la estadística que describe cómo actualizar las probabilidades de hipótesis en función de nuevas evidencias. Formalmente, el Teorema de Bayes se puede expresar como sigue:

Supongamos que tenemos una partición del espacio muestral  $\{B_1, B_2, \dots, B_n\}$ ,

donde cada  $B_i$  es mutuamente excluyente y colectivamente exhaustivo. En este contexto, estamos interesados en encontrar  $P(B_j|A)$ , la probabilidad de que el evento  $B_j$  ocurra dado que ha ocurrido el evento  $A$ .

La formulación del Teorema de Bayes es:

$$P(B_j|A) = \frac{P(A|B_j) \times P(B_j)}{\sum_{i=1}^n P(A|B_i) \times P(B_i)} \quad (1)$$

Comenzamos con la definición de probabilidad condicional:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (3)$$

A partir de la ecuación (2), despejamos  $P(A \cap B)$ :

$$P(A \cap B) = P(A|B) \times P(B) \quad (4)$$

Usamos la Ley de la Probabilidad Total para  $P(A)$ :

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i) \times P(B_i) \quad (5)$$

Finalmente, sustituimos en la ecuación para  $P(B|A)$ :

$$P(B|A) = \frac{P(A|B) \times P(B)}{\sum_{i=1}^n P(A|B_i) \times P(B_i)} \quad (6)$$

También se puede presentar como:

$$P(B_i|A) = \frac{P(A|B_i)}{\sum_{i=1}^n P(A|B_i) \times P(B_i)} \times P(B_i) \propto P(A|B_i) \times P(B_i) \quad (7)$$

Siendo  $P(B_i)$  probabilidad *a priori* del evento  $B_i$ ,  $P(B_i|A)$  probabilidad *a posteriori* del evento  $B_i$ ,  $P(A|B_i)$  y  $\sum_{i=1}^n P(A|B_i) \times P(B_i)$  conocidas como verosimilitud y constante normalizadora, respectivamente. En otras palabras podemos decir que la

probabilidad *a posteriori* es proporcional a la probabilidad *a priori* multiplicada por la verosimilitud.

### Grafo acíclico no direccionado

En una red bayesiana, todas las variables se asumen como variables aleatorias, consideradas como los nodos de un grafo. Así, la dependencia condicional entre ellas es representada por arcos dirigidos y la ausencia de estos arcos implica independencia condicional (Abellán *et al.*, 2006). Estas relaciones se descodifican visualmente a través de un grafo (conectado) acíclico dirigido (DAG - *Directed Acyclic Graph*). El término “dirigido” hace referencia a la presencia de dirección en los arcos, el término “conectado” se utiliza para designar que todos los nodos están conectados en la red y, finalmente, el término “acíclico” se refiere a la propiedad de no regresar a un nodo después de seguir la dirección de los arcos. La estructura de conexión, sin orientación de arcos entre los nodos, se llama esqueleto.

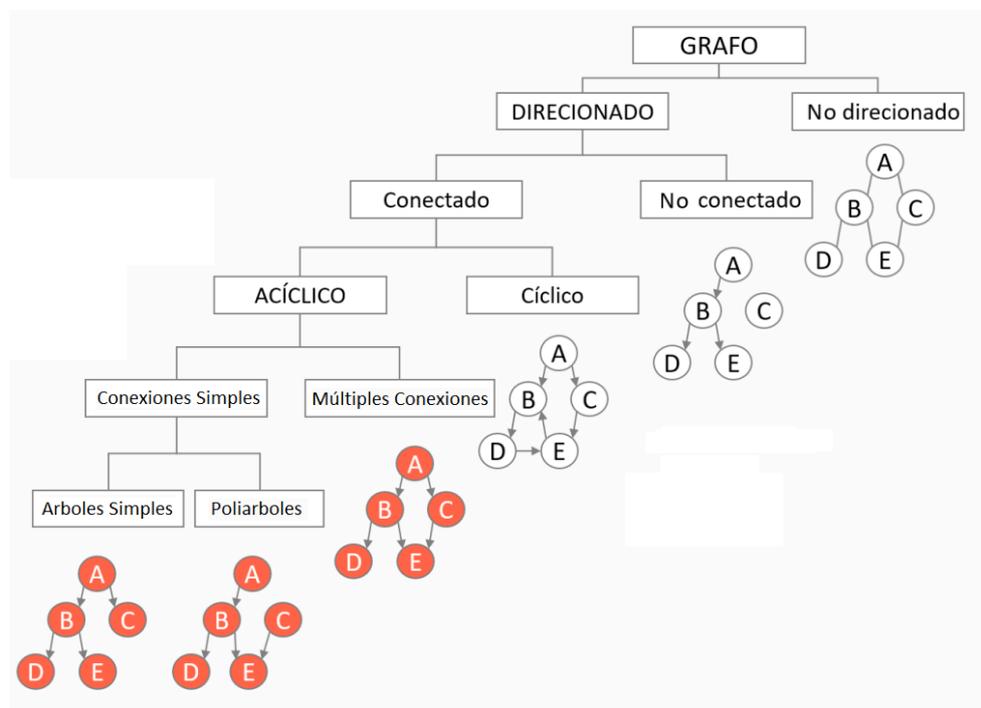


Figura 1. Estructuras básicas existentes en la teoría de grafos.

Adaptado de (Ara, 2023)

De esta forma, a través de la figura 1 se observa que las redes bayesianas involu-

cran algunos tipos de estructuras básicas: la estructura de conexiones simples, que engloba las estructuras de árboles simples y poliárboles y la estructura de múltiples conexiones, Para las estructuras de conexiones simples esta dada la regla general de que existe solo un camino que conecta una variable con otra, independiente de la dirección de los arcos. Análogamente, para las estructura de múltiples conexiones hay mas de una posibilidad que conecta una variable a otra, independiente de las direcciones de los arcos(Ara, 2023).

En este contexto, la terminología de parentesco se utiliza para entender la estructura de una red bayesiana. Para el caso de la Figura 1, en la estructura poliárbol, tenemos que la variable  $A$  es llamada padre de la variable  $B$ , siendo la variable  $B$  hija de la variable  $A$  y, al mismo tiempo,  $A$  es cónyuge de la variable  $C$  y padre de  $E$ . Los ancestros de la variable  $E$  son dados por el conjunto  $\{A, B, C\}$ . Además, las variables  $A$  y  $C$  son llamadas raíces de la red y las variables  $D$  y  $E$  son denominadas hojas de la red.

El estudio de los grafos se lleva a cabo en la rama de matemáticas llamada teoría de grafos y se refiere al estudio de las relaciones de sus elementos, que comúnmente se llaman nodos y arcos. Los nodos son los principales elementos que representan las variables aleatorias en el problema y generalmente se representan con círculos. Los arcos son flechas que representan la relación directa de dependencia entre un nodo y otro, es decir, representan una dependencia probabilística directa entre dos variables. Los arcos sin dirección se llaman aristas, elementos que no forman parte de un DAG.

Sea  $X = (X_1, \dots, X_p)$  el vector aleatorio con  $p$  componentes, la estructura del DAG, es decir, el patrón de conexiones de nodos y arcos, puede ser denotada por  $G$  y codifica las relaciones de independencia entre esas variables. Es decir, la estructura gráfica  $G$  puede ser representada por  $X$  y un conjunto de arcos (o aristas) que podemos denotar por  $E$ , siendo  $G = \{X, E\}$ . De esta manera, una vez que la separación gráfica en  $G$  implica en independencia condicional en probabilidad,  $G$  induce la factorización

$$P(X|G, \Theta) = \prod_{i=1}^p P(X_i | \Pi_{X_i}, \Theta_{X_i})$$

donde la distribución conjunta de  $X$  (con parámetros  $\Theta$ ) se descompone en una factorización local para cada  $X_i$  (con parámetros  $\Theta_{X_i}$ ), siendo  $\bigcup_{i=1}^p \Theta_{X_i} = \Theta$  y condicional al conjunto de sus padres, denotado por  $\Pi_{X_i}$ . Tal descomposición es posible debido a la suposición que una red bayesiana cumple con la condición de Markov (Neapolitan *et al.*, 2004), de que cada nodo es condicionalmente independiente de los nodos no descendientes dados sus padres.

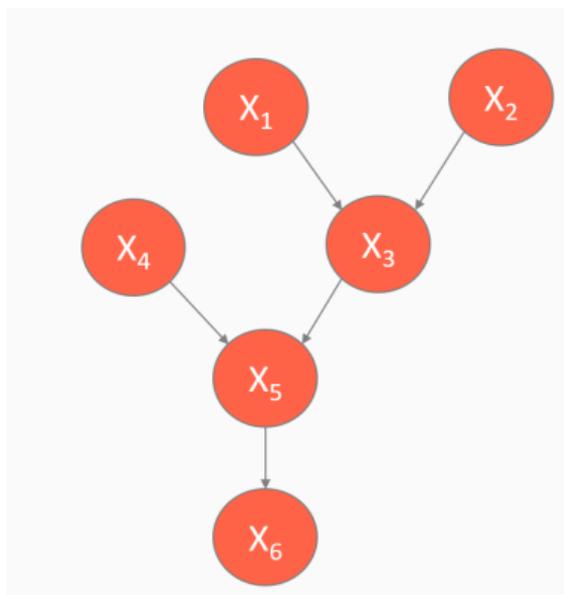


Figura 2. Ejemplo de una Red Bayesiana con seis variables

Adaptado de (Ara, 2023)

Para el caso de la red bayesiana presentada en la Figura 2, la descomposición de la distribución conjunta de  $(X_1, \dots, X_6)$  es dada por:

$$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1)P(X_2)P(X_3|X_1, X_2)P(X_4)P(X_5|X_3, X_4)P(X_6|X_5)$$

### Causalidad

La importancia de las redes bayesianas para la causalidad está relacionada con el trabajo de 1956 de Hans Reichenbach sobre causalidad (Reichenbach, 1991), que propuso el siguiente principio (Korb & Nicholson, 2010):

**Principio de la causa común:** Si dos variables son marginalmente dependientes probabilísticamente, entonces una causa a la otra (directa o indirectamente) o ellas poseen un ancestro común.

El principio, en esencia, afirma que detrás de cada dependencia probabilística hay una dependencia causal explicativa. Tales relaciones de dependencia están vinculadas a las estructuras de conexiones.

El contexto, la suposición de suficiencia causal es importante, la cual afirma que no existen variables no observadas que sean padres de una o más variables observadas. Es decir, no existen variables no observadas que puedan explicar las dependencias o independencia entre las demás variables. Esta suposición es inválida en la mayoría de los casos, principalmente debido a la facilidad de imaginar más variables que podrían ser incluidas como hijos o padres de una de las variables involucradas (Margaritis *et al.*, 2003).

Aún más, las características de clases de equivalencia actúan en el sentido de que, para algunas estructuras, existe una indistinguibilidad probabilística, lo que puede impedir la interpretación de causa-efecto del direccionamiento de arcos (Nagaranjan *et al.*, 2013).

Así, la influencia de arcos en una estructura de una red bayesiana debe ser investigada de forma práctica por un equipo de especialistas en el área, ya que debido a tales situaciones, sólo la orientación de arcos no puede ser entendida o garantizada como causalidad (Uusitalo, 2007).

Una red Bayesiana debe satisfacer las siguientes propiedades:

1. **Representación de un Grafo Acíclico Dirigido (DAG):** La red debe ser un DAG, donde los nodos representan variables aleatorias y los arcos representan dependencias directas entre estas variables. No puede haber ciclos, lo que significa que no se puede volver a un nodo siguiendo la dirección de los arcos.
2. **Factorización del Producto:** La distribución conjunta debe poder factorizarse como el producto de las distribuciones condicionales de cada variable dado sus padres en el DAG, Matemáticamente, si  $X_1, X_2, \dots, X_n$  son variables aleatorias y  $Pa(X_i)$  denota el conjunto de los padres de  $X_i$  en la red entonces:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (8)$$

3. **Independencias Condicionales:** La red debe reflejar las independencias condicionales entre las variables. Una red bayesiana encodes independencias condicionales a través de su estructura. Si un nodo  $X$  es condicionalmente independiente de  $Y$  dado un conjunto de nodos  $Z$ , entonces esta independencia se debe reflejar en la estructura de la red.
4. **Localidad:** Las distribuciones condicionales de una variable están determinadas únicamente por sus padres inmediatos en la red, y no por otras variables del sistema. Esto permite que la inferencia se realice de manera local.
5. **Consistencia de Probabilidades Marginales y Condiciones:** Las probabilidades marginales y condicionales derivadas de la red deben ser consistentes entre sí, lo que significa que la sumatoria o integración sobre las variables no observadas produce la distribución correcta sobre las variables restantes.

### 1.1.2 Inferencia bayesiana

La inferencia bayesiana es un enfoque de la estadística basado en el teorema de Bayes, que permite actualizar probabilidades a medida que se obtiene más información. A diferencia de la estadística clásica, que busca parámetros fijos basados en una muestra, la estadística bayesiana trata los parámetros como distribuciones de probabilidad.

#### Fundamentos Teóricos

El teorema de Bayes formula cómo actualizar probabilidades basadas en nueva evidencia:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Donde:

- $P(H|E)$  es la probabilidad posterior de la hipótesis  $H$  después de observar la evidencia  $E$ .

- $P(H)$  es la probabilidad priori, nuestra creencia inicial en  $H$  antes de observar cualquier evidencia.
- $P(E|H)$  es la probabilidad de observar la evidencia  $E$  suponiendo que  $H$  es verdadera.
- $P(E)$  es la probabilidad total de observar  $E$ , independientemente de la veracidad de  $H$ .

### **Distribuciones Priori y Posteriori**

Una parte esencial de la inferencia bayesiana es la elección de una distribución priori. Esta distribución refleja nuestro conocimiento previo sobre el parámetro antes de observar los datos. Después de observar los datos, actualizamos esta distribución para obtener la distribución posterior, que combina nuestra creencia previa y la nueva evidencia.

### **Estimación Bayesiana**

La inferencia bayesiana permite estimar parámetros y hacer predicciones. Para estimar un parámetro, se calcula el valor esperado de la distribución posterior. También se pueden obtener intervalos de credibilidad, que son análogos a los intervalos de confianza en la estadística clásica, pero tienen una interpretación probabilística directa.

### **Modelos Jerárquicos y Modelos Gráficos**

Los modelos jerárquicos permiten incluir estructura adicional en los modelos bayesianos, como efectos aleatorios. Los modelos gráficos, por otro lado, representan las relaciones condicionales entre variables en un gráfico, donde los nodos representan variables y los bordes representan dependencias.

### **Ventajas y Desafíos**

La principal ventaja de la inferencia bayesiana es su capacidad para incorporar información previa y tratar los parámetros como distribuciones en lugar de va-

lores fijos. Sin embargo, la inferencia bayesiana puede ser computacionalmente intensiva, especialmente en modelos complejos.

### 1.1.3 Aprendizaje bayesiano

Quizás la tarea más desafiante al tratar con redes bayesianas sea aprender su estructura. Sin embargo; investigar en esta dirección es esencial debido a su enorme utilidad, tanto para aplicaciones destinadas al usuario final como para el aprendizaje de redes causales en Biología, Medicina, Química, Física y en las ciencias en general.

Se presenta una visión general de las técnicas predominantes que se utilizan para aprender la Red Bayesiana (Bayesian Network, BN). primero describimos cómo se pueden aprender los parámetros de las BNs, dada la estructura. En las secciones siguientes centramos en aprender la estructura en sí. Existen dos amplias clases de algoritmos para aprender la estructura de las BNs. Una clase "puntúa una BN según lo bien que se ajuste a los datos e intenta producir una que optimice esa puntuación, inmediatamente después, se presenta el enfoque alternativo, que utiliza restricciones como las relaciones de independencia que se podría conocer en los datos, para reconstruir la estructura. Esta última clase es más relevante para la investigación, aunque aplicamos el aprendizaje basado en puntuación a una aplicación donde ajustar bien los datos es una prioridad.

### 1.1.4 Aprendizaje Paramétrico

Aprender los parámetros de una estructura de red fija (modelo estadístico) es un problema ampliamente reconocido en estadísticas. En la literatura sobre Red Bayesiana (Bayesian Network, BN), donde las metodologías bayesianas parecen predominar, el problema se plantea de la siguiente manera: se asume una distribución previa sobre los parámetros de las funciones de densidad de probabilidad (pdf) locales antes de usar los datos. Es deseable que esta distribución previa sea conjugada; una familia de distribución es denominada *prior conjugado* cuando el posterior de los parámetros pertenece a la misma familia. Aunque con diferentes *hiperparámetros* (los parámetros de una distribución sobre otros parámetros, a veces llamados así). En el presente estudio, solo usamos multinomiales para las pdfs locales, así que

detallaremos solo este caso. En otros contextos, como la regresión lineal con ruido gaussiano, ver Buntine (1993); Heckerman y Geiger (1995), o en enfoques más complejos representados por redes neuronales artificiales, consultar Monti y Cooper (1998a).

Para distribuciones multinomiales, el prior conjugado proviene de la familia Dirichlet. Si denotamos la probabilidad de cada bin  $p_{ijk,k=1,\dots,r}$  en la pdf local de la variable  $X_i$  para la configuración de padre  $pa_j$ , la distribución de Dirichlet sobre estos parámetros se expresa así.

Dado el conjunto de datos  $\mathcal{D}$ , la probabilidad posterior sobre los parámetros también pertenece a la familia Dirichlet, ya que es un prior conjugado al multinomial. Por lo tanto, tenemos:

$$Pr(p|\mathcal{G}) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \frac{p_{ijk}^{\alpha_{ijk}-1}}{\Gamma(\alpha_{ijk})} \quad (9)$$

La probabilidad puede expresarse como:

$$Pr(p_{i1}, p_{i2}, \dots, p_{ir_i} | \mathcal{G}, \mathcal{D}) = Dir(N_{i1} + \alpha_{i1}, N_{i2} + \alpha_{i2}, \dots, N_{ir_i} + \alpha_{ir_i}) \quad (10)$$

y

$$Pr(p|\mathcal{G}, \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \frac{(\alpha_{ijk} + N_{ijk})^{N_{ijk} + \alpha_{ijk} - 1}}{\Gamma(N_{ijk} + \alpha_{ijk})} \quad (11)$$

Donde  $N_{ijk}$  es el número de muestras en el bin  $k$  para la pdf de la variable  $X_i$  con una configuración de padre  $pa_j$  (y  $N_{ijk}$  son las estadísticas suficientes de esa pdf). Al utilizar esta distribución para predecir el valor de cualquier cantidad  $Q(X_1, X_2, \dots, X_n)$  basada en las variables del dominio, se promedia sobre todos los posibles valores de los parámetros (que son desconocidos).

La probabilidad de  $Q(X_1, X_2, \dots, X_n)$  ajustada por la probabilidad posterior de cada valor es:

$$Pr(Q(X_1, X_2, \dots, X_n) | \mathcal{G}, \mathcal{D}) = \int Q(X_1, X_2, \dots, X_n) Pr(p | \mathcal{G}, \mathcal{D}) dp \quad (12)$$

Con frecuencia, y por conveniencia, se utilizan los parámetros de máxima verosimilitud (ML) en lugar de la distribución completa. El ML para  $p_{ijk}$  se define como:

$$p_{ijk} = \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + N_{ij}} \quad (13)$$

Por comodidad, especialmente en casos donde los datos son abundantes, se suelen ignorar los hiperparámetros (asumiendo que  $\alpha_{ijk} \ll N_{ijk}$ ), y se utiliza la fracción  $N_{ijk}/N_{ij}$  en su lugar. Esto ocurre, por ejemplo, en métodos basados en puntuación, donde se emplea la puntuación Criterio de Información Bayesiano (Bayesian Information Criterion, BIC). Esta puntuación es una aproximación basada en muestras grandes del posterior y ya asume que los efectos de un prior son despreciables.

### 1.1.5 Aprendizaje Estructural

El aprendizaje de redes bayesianas busca modelos que representen de la mejor manera cómo las variables se conectan entre sí, de modo que tales conexiones caractericen las dependencias marginales y condicionales entre sus nodos (Koller & Friedman, 2009). Dado que las redes bayesianas están compuestas por dos elementos, su estimación se divide en dos tareas principales: la estimación de la estructura  $G$  y de los parámetros de  $P(\mathbf{X})$ .

La estructura de una red bayesiana puede ser descrita por especialistas y, en este caso, el proceso de estimación se restringirá a la estimación de los parámetros (Uusitalo, 2007). Sin embargo, el proceso de estimación puede ser realizado a través de los datos y de la utilización de algoritmos de estimación.

Encontrar la estructura óptima de una red bayesiana a partir de datos es un proceso difícil. Chickering (2002) destaca que la estimación de estructura es una tarea de alta complejidad computacional (NP-difícil), incluso cuando cada nodo está restringido a tener un máximo de dos padres.

En este contexto, los métodos de estimación de estructura son moderadamente re-

cientes y aún están en evolución, pero han sido eficaces para la solución de algunos problemas de modelado. Además, son capaces de ajustar modelos de alta dimensión que no son afectados por el problema conocido como maldición de la dimensionalidad (Scutari, 2010).

Se tiene dos clases de métodos para estimar la estructura de redes bayesianas a partir de datos son comunes en la literatura: métodos basados en restricciones y métodos basados en puntuación.

### **Métodos basados en restricción**

Estos algoritmos se basan en estimar la estructura de la red analizando las relaciones probabilísticas que surgen de la propiedad de Markov y de pruebas de independencia condicional y, a continuación, construyendo un DAG (gráfico acíclico dirigido) que satisfaga las declaraciones de d-separación. Los modelos resultantes son frecuentemente interpretados como modelos causales, incluso cuando son estimados a partir de datos observacionales (Pearl, 1988).

### **Métodos basados en puntuación**

Estos algoritmos asignan una puntuación a cada estructura de red bayesiana candidata e intentan maximizarla a través de metaheurísticas. Algoritmos de búsqueda ávida (como el ascenso de colina o búsqueda tabú) son una elección común, pero muchos otros tipos de procedimientos de búsqueda pueden ser utilizados.

Una tercera clase de métodos de estimación está compuesta por algoritmos híbridos, que combinan ambas aproximaciones mencionadas anteriormente.

#### **1.1.6 Métodos basados en restricción**

Esta clase de algoritmos se basa en pruebas de independencia condicional para verificar las conexiones de arcos. De forma general, cuando hay independencia (marginal o condicional) entre dos variables, el arco entre dichas variables es retirado (Zhang *et al.*, 2012). De esta manera, las pruebas de independencia condicional deben ser

definidas previamente y están vinculadas al tipo de dato en análisis; para el caso discreto, las pruebas comunes son vía información mutua condicional y  $\chi^2$  de Pearson, para el caso continuo, bajo la suposición de normalidad, una elección natural es la prueba vía correlaciones parciales de Pearson (Nagarajan *et al.*, 2013).

### Algoritmo PC

Entre los algoritmos más comunes de los métodos basados en restricción, se tiene el algoritmo PC (nombrado en homenaje a sus autores, Peter y Clark), siendo considerado una implementación del algoritmo IC (Inductive Causation) propuesto por (Pearl & Verma, 1995), un algoritmo conceptual relativo a la estimación basada en restricción. El algoritmo PC fue modernizado por Colombo, Maathuis *et al.* (2014), en una versión conocida como PC-estable.

Pseudocódigo del algoritmo IC:

1. Para cada par de variables  $A$  y  $B$ , busque un conjunto  $S_{AB}$  tal que  $A \perp B | S_{AB}$ . Construya un grafo no dirigido, tal que los vértices  $A$  y  $B$  estén conectados con una arista si y solo si ningún  $S_{AB}$  es encontrado.
2. Para cada par de variables no adyacentes  $A$  y  $B$  con un vecino común  $C$ , verifique si  $C \in S_{AB}$ . Si está contenido, continúe; de lo contrario, oriente  $A \rightarrow C \leftarrow B$ .
3. En el grafo parcialmente dirigido resultante, oriente el máximo posible de aristas no dirigidas, sujeto a dos condiciones: (i) la orientación no debe crear una nueva estructura *head-to-head*; y (ii) la orientación no debe crear un ciclo dirigido.

Así, el algoritmo PC comienza con un grafo completo no dirigido y excluye las aristas recursivamente con base en decisiones basadas en independencia condicional. La Figura 3 ilustra cómo funciona el algoritmo PC para cuatro variables, donde:

- (i) presenta el verdadero grafo,
- (ii) el inicio del algoritmo a través de un grafo no dirigido y totalmente conectado,

- (iii) la arista  $X_1 - X_2$  es removida porque  $X_1 \perp X_2$ ,
- (iv) las aristas  $X_1 - X_4$  y  $X_2 - X_4$  son removidas debido a  $X_1 \perp X_4 | X_3$  y  $X_2 \perp X_4 | X_3$ ,
- (v) presenta el grafo parcialmente conectado después de encontrar las estructuras *head-to-head*.
- (vi) presenta el grafo final después de la orientación de los demás arcos.

Básicamente, el algoritmo PC encuentra el esqueleto de la red y, después de eso, identifica las estructuras *head-to-head* basándose en las propiedades de *d*-separación. Si las decisiones de independencia condicional son correctas para una muestra suficientemente grande, el algoritmo PC tiende a converger a la verdadera Clase de Equivalencia de Markov (Glymour *et al.*, 2019).

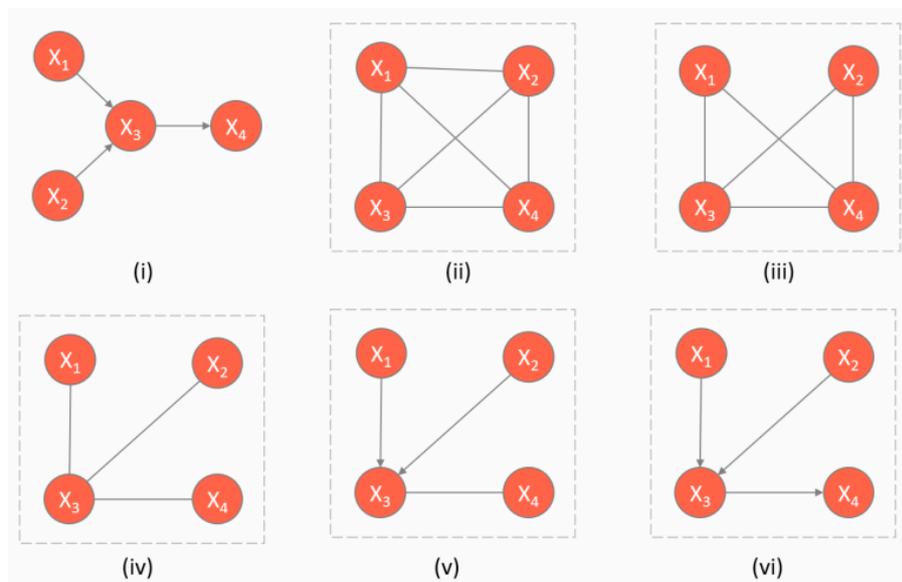


Figura 3. Estimación de estructura: Ilustración del algoritmo PC

Adaptado de (Glymour *et al.*, 2019)

### Algoritmo PC-estable

Este algoritmo, propuesto por Colombo, Maathuis *et al.* (2014), aborda la dependencia del orden de las variables en el algoritmo de PC. Los mismos autores demostraron que el orden de las variables para datos de baja dimensión no es un problema significativo, pero para datos de alta dimensión, puede producir resultados bastante variados para la estructura estimada. Además, fueron propuestos

algunas modificaciones que podrían eliminar parcial o completamente la dependencia del orden de las variables en el algoritmo de PC. Las modificaciones del algoritmo se realizaron a lo largo de la fase de estimación del esqueleto de la red.

### 1.1.7 Métodos basados en puntuación

Uno de los métodos más comunes para generar Redes Bayesianas (BNs) a partir de datos, en particular para estimar funciones de densidad de probabilidad (pdf), es el enfoque basado en puntuaciones. El método otorga una puntuación a cada BN candidata, que mide típicamente cuán adecuadamente la BN describe el conjunto de datos  $D$ . Dada una estructura  $G$ , su puntuación se da por:

$$\text{Puntuación}(G, D) = Pr(G|D), \quad (14)$$

es decir, la probabilidad posterior de  $G$  dada la muestra de datos. Un algoritmo basado en puntuaciones busca maximizar esta medida. Esta computación puede simplificarse usando el teorema de Bayes:

$$\text{Puntuación}(G, D) = \frac{Pr(D|G)Pr(G)}{Pr(D)}. \quad (15)$$

Para optimizar esta puntuación, basta con maximizar el numerador, ya que el denominador es independiente de  $G$ . Hay distintas maneras de determinar  $Pr(G)$  a partir de información previa. Por simplicidad en esta discusión, consideraremos que  $Pr(G)$  sigue una distribución uniforme sobre las posibles estructuras.

Para calcular  $Pr(D|G)$ , la metodología bayesiana considera todos los posibles parámetros, ponderando cada uno según su probabilidad posterior:

$$Pr(D|G) = \int Pr(D|p, G)Pr(p|G) dp. \quad (16)$$

Sin embargo, todas las soluciones disponibles tienden a basarse en un principio común conocido como parsimonia o navaja de Occam (Russell & Norvig, 2016), que considera que debemos preferir modelos más simples en lugar de modelos más com-

plejos.

Utilizando la intuición de complejidad del modelo, podemos definir una clase común de medidas de puntuación,

$$s(G|D) = l(G|D) - \phi(n) \times \|G\|,$$

donde el primer componente  $l(G|D)$  es la función de log-verosimilitud. El segundo componente,  $\phi(n) \times \|G\|$ , es un término de penalización que favorece modelos más simples, es decir, modelos con menos parámetros, siendo  $\|G\|$  el número de parámetros en  $G$ . Además,  $\phi(n) \geq 0$  es una función del tamaño del conjunto de datos  $n$  que controla el peso general de la penalización.

De esta manera, existen diferentes tipos de medidas bayesianas y no bayesianas de puntuación que pueden ser utilizadas y la elección se realiza previamente.

El procedimiento de estimación de estructura para los algoritmos basados en puntuación consiste en una búsqueda recursiva por conexiones, seguida del cálculo de una función objetivo que mide el grado de ajuste de la red a los datos (Beretta *et al.*, 2018). La repetición recursiva de estos pasos busca encontrar la mejor estructura  $g^*$  en  $G$  que maximice la función objetivo  $s(G|D)$ .

Para esto, el algoritmo realiza una búsqueda vía *greedy search* en el espacio de posibles grafos  $G$  con el fin de encontrar una estructura que no cause pérdida en  $s(G|D)$  en relación con la opción anterior. Los algoritmos ampliamente utilizados son el ascenso por la colina (*hill-climbing*) y la búsqueda tabú (*tabu search*):

- Ascenso por la colina: algoritmo construido vía *greedy search* que explora el espacio de los grafos acíclicos dirigidos por medio de las operaciones de adición, remoción y reversión de un único arco. Este método es llamado algoritmo codicioso porque se mueve, en cada paso, hacia un buen estado vecino de manera oportunista (Russell & Norvig, 2016). Comúnmente es implementado con reinicios aleatorios para evitar óptimos locales.
- Búsqueda TABU: es el algoritmo de ascenso por la colina modificado para minimizar el estancamiento en óptimos locales, permitiendo que durante su

búsqueda sea seleccionada una red que disminuya la puntuación mínimamente. Así, este algoritmo limita los lugares que ya fueron testados anteriormente en la lista tabú, los cuales no pueden ser revisados por iteraciones predefinidas. Además, cuando encuentra un máximo, realiza más algunas iteraciones para garantizar que haya alcanzado un máximo global (Scutari *et al.*, 2018).

Los algoritmos *greedy search* siguen una heurística dirigida a hacer la elección ideal localmente, en cada paso, con la esperanza de encontrar un óptimo global ver figura 4. En muchos problemas, esta estrategia no produce en general una solución óptima, pero, aún así, puede producir soluciones localmente óptimas que se acercan a una solución óptima global en un tiempo razonable (Gámez *et al.*, 2011).

Para el caso de estimación de estructura de redes bayesianas, el paso principal del algoritmo consiste en realizar operaciones de adición, remoción o reversión de arco, tomando como nueva candidata la estructura que genera un aumento en la puntuación. El proceso termina cuando no hay ninguna alteración de un único arco que aumente la puntuación. El pseudocódigo del algoritmo de ascenso de colina se muestra a continuación, así como ilustrado genéricamente a través de la Figura 5 (Margaritis *et al.*, 2003).

1.  $E \leftarrow \emptyset$
2.  $g \leftarrow TPC(E, D)$
3.  $s(g|D) = -\infty$
4. Repetir:
  - a)  $maxscore \leftarrow score;$
  - b) para cada par de variables  $(X, Y)$ , hacer:
    - 1) para cada  $E' \in \{g \cup \{X \rightarrow Y\}, g - \{X \rightarrow Y\}, g - \{X \rightarrow Y\} \cup \{X \leftarrow Y\}\}$ 
      - $a' \theta \leftarrow TPC(E', D)$
      - $b' g' \leftarrow (X, E', \theta')$
      - $c' s_{nuevo} \leftarrow s(g'|D)$
      - $d' \text{ si } s_{nuevo} > s(g|D) \text{ entonces } g \leftarrow g'$
5. Mientras  $s_{nuevo} > s_{max}$

## 6. Retornar $g$

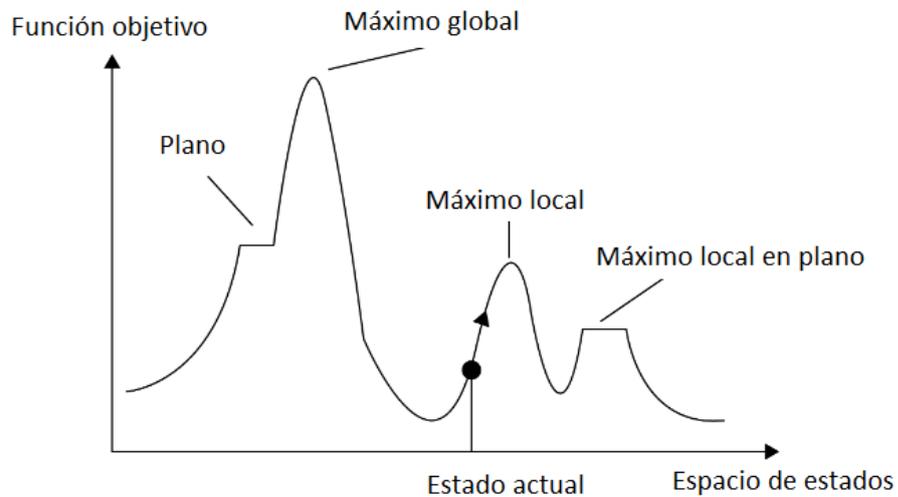


Figura 4. Algoritmo ascenso de colina (Hill Climbing)

Adaptado de (Russell & Norvig, 2016)

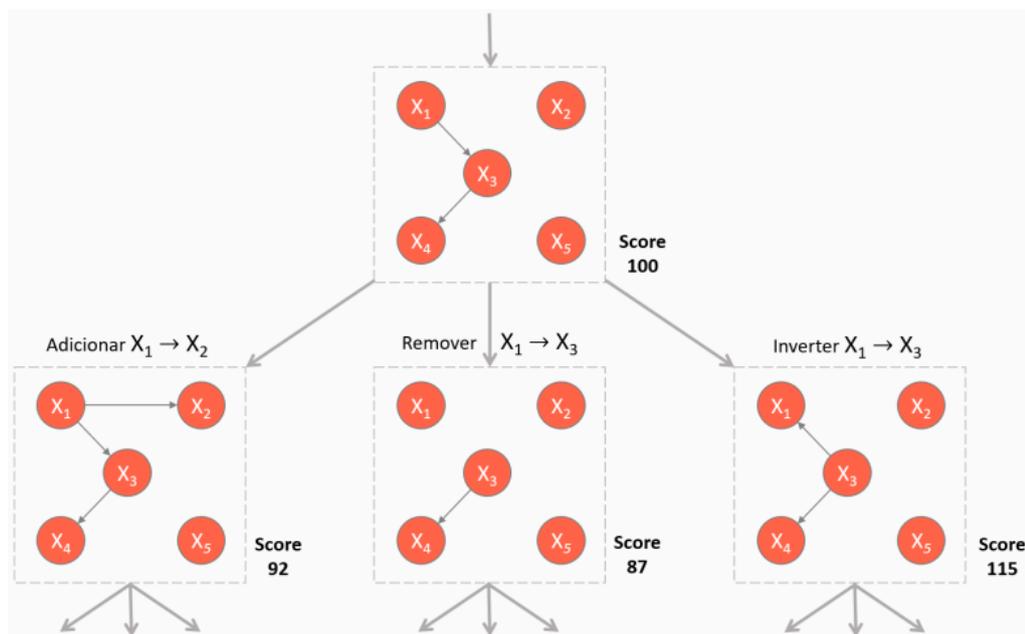


Figura 5. Estimación de estructura: abordaje basada en puntuaciones

Adaptado de (Margaritis *et al.*, 2003)

### 1.1.8 Métodos híbridos

El aprendizaje estructural en redes Bayesianas híbridas integra enfoques basados en restricciones y basados en puntuaciones para superar sus respectivas limitaciones

individuales. Estos métodos buscan combinar las fortalezas de cada enfoque para mejorar la eficiencia computacional y la precisión de la estructura aprendida. Estos métodos consisten en explorar el espacio de posibles estructuras de la red y evaluarlas en función de una función de puntuación que mide cuán bien la estructura se ajusta a los datos.

### **Algoritmos MCMC para el Aprendizaje de Estructuras**

Los métodos MCMC son ampliamente empleados para explorar el espacio de DAGs, teniendo en cuenta la incertidumbre estructural y realizando inferencias mediante el muestreo de DAGs de la distribución posterior a partir de los datos. Los avances en MCMC, como el MCMC de orden y el MCMC de partición, han mejorado la eficiencia computacional y la precisión al reducir el espacio de muestreo a espacios de particiones ordenadas de nodos.

### **El Algoritmo Partitioned PC**

El algoritmo Partitioned PC (pPC) mejora la eficiencia del algoritmo PC y conserva sus propiedades teóricas deseables y precisión empírica en el aprendizaje de estructuras. El pPC se basa en la estrategia de partición, estimación y fusión (PEF) y utiliza un enfoque de partición para aprender la estructura de redes Bayesianas discretas.

Las funciones de puntuación comúnmente utilizadas incluyen el Criterio de Información Bayesiana (BIC), el Criterio de Información de Akaike (AIC) o puntuaciones basadas en la verosimilitud que pueden incorporar creencias previas sobre la estructura de la red. El objetivo es encontrar la estructura que maximice la puntuación, lo que indica el mejor equilibrio entre la complejidad del modelo y el ajuste a los datos.

El procedimiento de búsqueda puede realizarse mediante diversos algoritmos, como el algoritmo de ascenso por la colina (hill-climbing), búsqueda tabú o algoritmos genéticos. El ascenso por la colina, por ejemplo, comienza con una estructura aleatoria y realiza cambios iterativos, como agregar o eliminar aristas, para mejorar la puntuación. La búsqueda tabú mejora el ascenso por la colina al mantener una lista

de estructuras recientemente visitadas para evitar ciclos y potencialmente escapar de máximos locales.

## 1.2 Antecedentes

La primera Subsección 1.2.1 presenta los antecedentes orientados a los trabajos de investigación sobre diferentes investigaciones aprendizaje estructural de una Red Bayesiana

### 1.2.1 Trabajos sobre Aprendizaje Estructural

El trabajo de Cooper y Herskovits (1992) es uno de los primeros que utiliza un método bayesiano para aprender tanto la estructura como los parámetros de una red Bayesiana a partir de datos. Este trabajo presenta un método Bayesiano para construir redes probabilísticas a partir de bases de datos, con un enfoque particular en la construcción de redes de creencias Bayesianas. El principal objetivo fue aplicar estas redes en áreas como la hipótesis asistida por computadora, el descubrimiento científico automatizado y la construcción automatizada de sistemas expertos probabilísticos. La metodología propuesta extiende el método básico para manejar datos faltantes y variables ocultas (latentes), y muestra cómo realizar inferencias probabilísticas promediando las inferencias de múltiples redes de creencias. Se presentan resultados de una evaluación preliminar de un algoritmo para construir una red de creencias a partir de una base de datos de casos. Como conclusión, se relacionan los métodos propuestos con trabajos anteriores y se discuten problemas abiertos.

Margaritis *et al.* (2003) En su tesis presentada, el objetivo principal fue abordar la determinación de la estructura de modelos estadísticos dirigidos, específicamente utilizando redes bayesianas para descubrir estructuras causales subyacentes y realizar predicciones sobre eventos difíciles de medir. La metodología implementada incluyó el desarrollo de un algoritmo para discernir la estructura de la red bayesiana basada en declaraciones de independencia estadística, la aplicación de un test de independencia para variables continuas y la demostración de una aplicación práctica del aprendizaje estructural en toma de decisiones. Las contribuciones significativas de la tesis abarcan la introducción de este algoritmo, el test de independencia y la

validación de la utilidad del modelo aprendido en escenarios prácticos, superando en algunos aspectos a enfoques contemporáneos.

Pifer y Guedes (2007) indican que Las Redes Bayesianas sirven para describir distribuciones de probabilidad a través de una representación gráfica y manejar incertidumbres de sistemas reales. Recientemente, se ha intensificado el interés en aprender las estructuras de estas redes desde conjuntos de datos; sin embargo, este aprendizaje es un problema NP-Difícil. El objetivo de este trabajo fue abordar el aprendizaje de la estructura de la red ALARM, un benchmark estándar. La metodología emplea el algoritmo de búsqueda K-2, pero con una modificación en la métrica MDL para puntuar el modelo. La principal contribución es esa adaptación de la métrica MDL. Finalmente, la conclusión fue que los modelos puntuados con parámetros menos restrictivos (que favorecen estructuras de red más simples) superan a los más restrictivos, y que la métrica MDL modificada ofrece mejores resultados que la métrica MDL original.

Marcot (2012) presenta un conjunto seleccionado de métricas existentes y nuevas para evaluar el rendimiento y la incertidumbre de los modelos de redes Bayesianas. El objetivo fue proporcionar herramientas para analizar la sensibilidad del modelo, evaluar escenarios, representar la complejidad del modelo, evaluar el rendimiento de la predicción y medir la incertidumbre de las probabilidades posteriores del modelo. La metodología involucra el uso de diversas métricas, como la reducción de la varianza, análisis de influencia, coeficiente de Gini, entre otros. Además, se presentan ejemplos de la aplicación de estas métricas en modelos reales relacionados con análisis y gestión de poblaciones de fauna silvestre. La conclusión subraya que el uso de tales métricas puede fortalecer significativamente la credibilidad, aceptación y aplicación adecuada de los modelos, especialmente en decisiones de gestión.

Costa *et al.* (2013) en su disertación propone la *aplicación de métodos de Monte Carlo vía Cadenas de Markov* en la construcción de estructuras de redes Bayesianas. El objetivo fue abordar problemas para los cuales no se puede obtener una solución exacta, creando estructuras de redes Bayesianas a partir de datos. La metodología consiste en fijar nodos de la estructura inicial, construir una estructura de red a partir de los nodos fijados y utilizar algoritmos de búsqueda heurística y Monte Carlo para

iterar y mejorar la estructura. La siguiente fase altera aleatoriamente el modelo para encontrar el que mejor se adapte a los datos, seleccionando operaciones en los nodos (incluir, excluir, invertir) al azar. Finalmente, se evalúan los modelos generados y se selecciona el mejor utilizando métricas de score. La conclusión indica que los resultados son satisfactorios, con una alta tasa de representación y solo el 7% de los datos clasificados incorrectamente en una de las pruebas, subrayando la efectividad del método propuesto para identificar modelos óptimos.

Kim y Jun (2013) propone un modelo gráfico probabilístico nuevo que incluye una variable latente no observable que influye en todas las demás variables observables. El objetivo principal es aplicar este modelo a la evaluación de clasificación de instituciones mediante un conjunto de indicadores de rendimiento. La metodología utiliza modelos lineales Gaussianos para expresar las relaciones causales entre variables y un método iterativo que combina algoritmos de descubrimiento causal basados en puntuación y en restricciones para encontrar la estructura de la red; para la estimación de parámetros, se emplean el muestreo de Gibbs y análisis de regresión. La variable latente, que representa las puntuaciones de clasificación de las instituciones, es estimada y las clasificaciones se determinan mediante la comparación de estas puntuaciones. Concluyendo que la estimación del intervalo de clasificación de una institución se obtiene mediante un procedimiento repetitivo, y que el procedimiento propuesto ha sido aplicado tanto a datos reales como artificiales.

Li y Chen (2014) Concluye que el aprendizaje de redes Bayesianas comprende el aprendizaje de estructuras y de parámetros. Abordar la estructura en grandes bases de datos es computacionalmente desafiante, clasificado como un problema NP-difícil. Este estudio tiene el objetivo de proponer un algoritmo híbrido que mejore este proceso integrando los algoritmos MMPC (max-min padres e hijos), PSO (optimización de enjambre de partículas) y GA (algoritmo genético). Primero, se construyó el marco de la red no dirigida con MMPC y luego se fueron aplicados PSO y GA en la búsqueda por puntuación. Gracias a la optimización global de PSO y la capacidad de cálculo paralelo de GA, se repara eficientemente el espacio de búsqueda y se determina la dirección de los bordes en la red. En comparación con los algoritmos PSO y GA convencionales, los resultados experimentales indican que el algoritmo propuesto es superior en velocidad de convergencia. Por lo tanto, la principal contribución

es un algoritmo más eficiente para aprender estructuras de redes Bayesianas.

O’Gorman *et al.* (2015) introduce un método para aprender la estructura de una Red Bayesiana utilizando el algoritmo cuántico adiabático. El objetivo fue lograr una reformulación eficiente de una función estándar de puntuación de probabilidad a posteriori en grafos como una función pseudo-booleana, que es equivalente a un sistema de spins Ising de 2-cuerpos. La metodología implicó introducir términos de penalización adecuados para hacer cumplir las restricciones de la reformulación y el método propuesto necesita  $\mathcal{O}(n^2)$  qubits para  $n$  variables de Red Bayesiana. Además, se demuestran cotas inferiores para la ponderación de estos términos de penalización. La principal contribución es esta reformulación eficiente y la estructura lógica resultante tiene la atractiva propiedad de ser independiente de la instancia para un número dado de variables de Red Bayesiana y es independiente del número de casos de datos.

Zou, Yue *et al.* (2017) En su estudio, consideraron diversos factores que afectan la seguridad vial para desarrollar un modelo de red Bayesiana basado en el análisis probabilístico orientado a la causación de accidentes de tráfico. Tomando como ejemplo el Distrito Central de Negocios (CBD) de Adelaida en Australia, el objetivo fue establecer la estructura de la red Bayesiana integrando el algoritmo K2 con el conocimiento de expertos. La metodología empleada incorporó el algoritmo de Expectation-Maximization para gestionar datos faltantes en el aprendizaje de parámetros en Netica, que luego se utilizó para realizar razonamientos de probabilidad posterior, explicaciones probables y análisis inferencial. La conclusión reveló que el modelo Bayesiano puede explorar eficazmente las relaciones lógicas complejas en accidentes de tráfico, predecir la probabilidad de un accidente en condiciones de tráfico específicas y identificar las principales causas y combinaciones de estados desfavorables. Esto brinda apoyo teórico a las autoridades de gestión vial para mejorar la seguridad del sistema de tráfico urbano.

Sun *et al.* (2019) en su investigación menciona que desde el inicio del siglo XXI, la investigación en inteligencia artificial ha progresado significativamente, con las redes Bayesianas emergiendo como un logro central. El principal objetivo de este estudio fue abordar el desafío en el aprendizaje de estructuras de redes Bayesianas donde

el método tradicional que utiliza el conocimiento experto está siendo reemplazado por un método basado en datos. Debido a la vastedad del espacio de búsqueda, este método enfrenta problemas de baja precisión o alta complejidad. Como metodología, se propone una técnica híbrida de optimización que discretiza y aplica el algoritmo de colonia de abejas artificiales para el aprendizaje de estructuras. Los resultados experimentales demuestran que el algoritmo propuesto presenta una mejor estructura y convergencia. En conclusión, este trabajo aporta una nueva técnica de optimización híbrida para mejorar la eficiencia y precisión en el aprendizaje de estructuras de redes Bayesianas.

Kitson *et al.* (2023) realizó el análisis de los algoritmos de aprendizaje de las Redes Bayesianas encontrando que son herramientas populares para razonar bajo incertidumbre en múltiples campos como medicina, biología y economía. Indica que a pesar de su utilidad, determinar la estructura gráfica de una BN, especialmente bajo supuestos causales, es un reto significativo. Este trabajo tuvo el objetivo de ofrecer una revisión exhaustiva de algoritmos combinatorios propuestos para aprender la estructura de BN a partir de datos. La metodología abarca la descripción de 74 algoritmos, que incluyen enfoques prototípicos, consolidados y de vanguardia, destacando sus similitudes y diferencias. Se discuten métodos para evaluar estos algoritmos y su rendimiento comparativo, así como la coherencia de las afirmaciones en la literatura. Además, se tratan enfoques para gestionar el ruido en los datos y la incorporación de conocimientos expertos en el proceso de aprendizaje. En consecuencia, la principal contribución del trabajo fue una revisión profunda y comparativa de los algoritmos para aprender estructuras de BN.

## CAPÍTULO II

### PLANTEAMIENTO DEL PROBLEMA

El capítulo aborda al problema que dio origen al presente trabajo de investigación. Para ello, se describe sobre la identificación y enunciado del problema, justificación, los objetivos de la investigación y la hipótesis de estudio.

#### 2.1 Identificación del problema

Actualmente se tiene muchos estudios realizados sobre el campo de Inteligencia Artificial (Artificial intelligence , AI) mostrando el enorme número de aplicaciones en las diferentes áreas de conocimiento, cuyos estudios son cada vez más específicos que requieren mayor especialización en cada una de las técnicas que ofrece para generar nuevos conocimientos.

El aprendizaje de estructuras en redes Bayesianas es un tema de gran relevancia en el ámbito de la ciencia de datos, el aprendizaje automático y la inteligencia artificial (Koller & Friedman, 2009; Murphy, 2012). Aunque diversas técnicas han sido propuestas para abordar este desafío, todavía existe una falta de consenso sobre cuál es el método más efectivo para diferentes tipos de datos y contextos (D. H. et al., 1995; Scutari, 2010). Los métodos actuales de aprendizaje estructural incluyen, pero no se limitan a, algoritmos basados en puntuación, búsqueda y restricción, y técnicas de optimización (Chickering, 2002; I. T. et al., 2003).

Dado que las redes Bayesianas son ampliamente utilizadas en diversas aplicaciones, desde la medicina y la biología hasta la ingeniería y la economía (Neapolitan *et al.*, 2004; Pearl, 1988), la eficacia de los algoritmos de aprendizaje estructural es crucial para la inferencia y la toma de decisiones en estos campos (N. Z. et al., 2018; P. S. et al., 2000). Sin embargo,

pocas investigaciones han abordado una comparación y clasificación exhaustivas de estos métodos en términos de eficiencia computacional, precisión y aplicabilidad en diferentes dominios (Jensen & Nielsen, 2007; Korb & Nicholson, 2010).

Aunque se han realizado varios estudios comparativos, estos a menudo están limitados en su alcance y no abarcan la gama completa de técnicas disponibles (C. A. et al., 2010; R. N. et al., 2013). Además, las métricas utilizadas para evaluar la efectividad de los métodos de aprendizaje estructural a menudo varían entre estudios, lo que hace difícil una comparación directa (N. F. et al., 1999; Margaritis *et al.*, 2003). Esto representa una brecha significativa en la literatura que necesita ser abordada para permitir un uso más eficiente de las redes Bayesianas en aplicaciones prácticas.

La falta de una taxonomía bien definida y una comparación rigurosa de las técnicas de aprendizaje estructural impide que los investigadores y profesionales seleccionen el algoritmo más adecuado para sus necesidades específicas (Cussens, 2011; Koivisto & Sood, 2004).

Finalmente, se resalta que el modelo clásico muchas veces puede tornarse complejo y costoso en función al problema abordado dado a la gran cantidad de variables y datos incompletos que dificultan el aprendizaje paramétrico y de forma especial el aprendizaje estructural siendo este aprendizaje la parte más compleja dentro de la red bayesiana, así métodos automatizados pueden ser potencialmente más interesante, porque optimiza el tiempo de construcción de la red bayesiana, además de proveer una visualización adecuada de la información estructural de la red bayesiana, que en muchos casos están basado en conocimientos de especialista con un alto grado de incerteza.

## 2.2 Enunciados del problema

Por lo mencionado, se realizó la siguiente formulación del problema: ¿Cual de las técnicas de aprendizaje estructural de la Redes Bayesianas tiene mayor desempeño y cual es su clasificación?

## 2.3 Justificación

Las redes Bayesianas son una herramienta poderosa para modelar incertidumbre, dependencias causales y probabilidades condicionales en una amplia variedad de campos, incluidas las ciencias de la vida, la ingeniería y las ciencias sociales (Neapolitan *et al.*, 2004; Pearl, 1988). Sin embargo, la eficacia de una red Bayesiana depende en gran medida de su estructura, la cual es a menudo desconocida y debe ser aprendida a partir de datos (Koller & Friedman, 2009; Scutari, 2010). La importancia de los métodos de aprendizaje estructural radica en su capacidad para descubrir estas relaciones de dependencia, lo que a su vez permite hacer inferencias más precisas y tomar decisiones mejor informadas (Chickering, 2002; D. H. et al., 1995).

A pesar de la existencia de múltiples algoritmos y técnicas para el aprendizaje estructural de redes Bayesianas, hay una falta de estudios que proporcionen una comparación rigurosa y exhaustiva de estos métodos (C. A. et al., 2010; R. N. et al., 2013). Esta falta de claridad crea un obstáculo para los investigadores que buscan aplicar redes Bayesianas en diferentes dominios, ya que no hay pautas claras sobre qué método es el más adecuado para un problema específico (N. Z. et al., 2018; Korb & Nicholson, 2010).

Además, las diferencias en eficiencia computacional entre los métodos pueden tener un impacto significativo en aplicaciones a gran escala o en tiempo real (Cussens, 2011; Koi-visto & Sood, 2004). También existe una brecha en la literatura en cuanto a la adaptabilidad de estos métodos a diferentes tipos de datos y problemas, como los que implican datos de alta dimensionalidad o estructuras de dependencia complejas (N. F. et al., 1999; Jensen & Nielsen, 2007).

Por lo tanto, esta investigación busca llenar este vacío en la literatura al proporcionar una comparación y clasificación sistemáticas de las técnicas de aprendizaje estructural. No solo servirá como una guía para los investigadores y practicantes, sino que también podría llevar a la identificación de nuevos métodos o híbridos que superen las limitaciones de las técnicas actuales (I. T. et al., 2003; Margaritis *et al.*, 2003).

## 2.4 Objetivos

### 2.4.1 Objetivo general

Evaluar las técnicas de aprendizaje estructural de la Red Bayesiana según su desempeño.

### 2.4.2 Objetivos específicos

- Comparar las técnicas de aprendizaje estructural de la Red Bayesiana.
- Clasificar las técnicas de aprendizaje estructural de la Red Bayesiana según su desempeño
- Conocer técnicas estructurales de las redes bayesianas en la literatura.

## 2.5 Hipótesis

Existe diferencia entre las técnicas de aprendizaje de las redes bayesianas el cual conlleva a su posterior clasificación

## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1 Ámbito o Lugar de estudio

La presente investigación, se realizó en primera instancia de forma exploratoria para construir la teoría y estuvo basado enteramente en la revisión bibliográfica de métodos de aprendizaje estructural de las Redes Bayesianas en las diferentes bases de investigación.

Se utilizó el modelo ASIA que aborda el diagnóstico del cáncer de pulmón (Lauritzen & Spiegelhalter, 1988). La disnea, o falta de aire, puede originarse debido a enfermedades como la tuberculosis, el cáncer de pulmón, la bronquitis, entre otras causas. Se sabe que fumar incrementa el riesgo de desarrollar cáncer de pulmón y bronquitis, mientras que viajar a Asia puede elevar las posibilidades de adquirir tuberculosis. Sin embargo, una simple radiografía de tórax no es conclusiva para determinar la presencia de cáncer pulmonar o tuberculosis, al igual que la disnea no es un indicador definitivo por sí sola. El modelo ASIA se compone de 8 variables y 16 estados posibles en total. La red correspondiente tiene 8 conexiones y se ilustra en la Figura 6

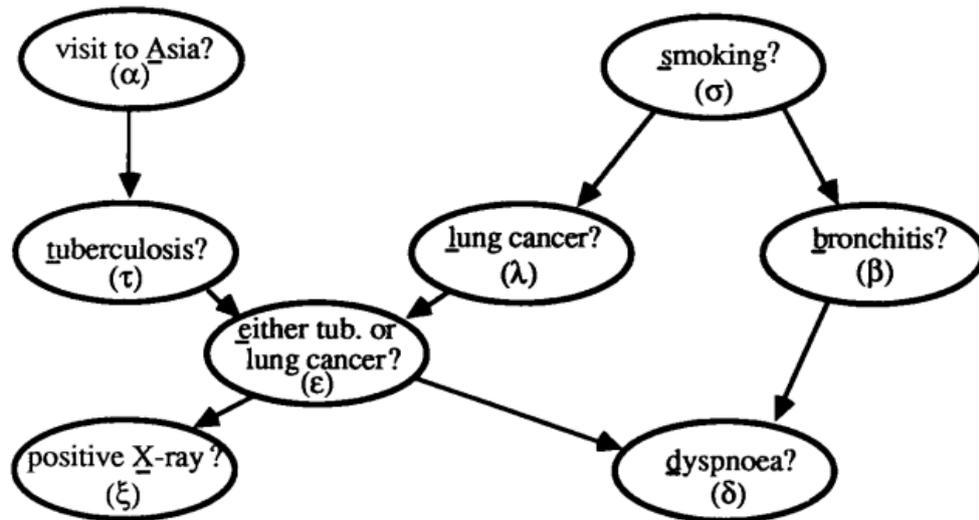


Figura 6. Red Objetivo del problema ASIA

fuelle: (Lauritzen & Spiegelhalter, 1988)

El modelo ASIA, desarrollado por Lauritzen y Spiegelhalter en 1988, es uno de los ejemplos más conocidos en la literatura sobre redes bayesianas. También se le conoce como “El modelo de la red del cáncer de pulmón”. Fue diseñado principalmente como una herramienta didáctica para ilustrar los conceptos básicos de las redes bayesianas y no necesariamente para ser un modelo clínico detallado.

El modelo se desarrolló con base en el conocimiento experto y en algunos estudios existentes sobre las relaciones entre ciertas enfermedades y condiciones. Su estructura refleja las dependencias causales entre diferentes factores y enfermedades. Las principales variables en el modelo son:

1. **Visita a Asia:** Una variable binaria que indica si una persona ha visitado Asia recientemente.
2. **Tabaquismo:** Indica si una persona fuma o no.
3. **Tuberculosis:** Una enfermedad que se puede contraer, especialmente después de visitar Asia.
4. **Cáncer de Pulmón:** Una enfermedad que tiene una mayor probabilidad de ocurrir en fumadores.
5. **Bronquitis:** Otra enfermedad que es más común entre los fumadores.

6. **Disnea (falta de aire):** Un síntoma que puede ser causado tanto por la tuberculosis como por el cáncer de pulmón.
7. **Radiografía del tórax:** Un examen que puede mostrar manchas u otras anomalías asociadas con el cáncer de pulmón o la tuberculosis.
8. **Resultado Positivo en la prueba de Tuberculosis:** Indica si una prueba específica para la tuberculosis es positiva.

Las relaciones entre las variables están establecidas por probabilidades condicionales, lo que permite a la red hacer inferencias basadas en la evidencia proporcionada. Por ejemplo, si una persona ha visitado Asia recientemente y tiene una radiografía de tórax anormal, la probabilidad de que tenga tuberculosis podría aumentar en el modelo.

Lauritzen y Spiegelhalter usaron el modelo ASIA como un caso de estudio en su influyente artículo de 1988 sobre redes bayesianas, estableciendo una base para muchos desarrollos futuros en el campo.

### 3.2 Población

Se consideró como elemento de comparación todos los datos del modelo ASIA (Lauritzen & Spiegelhalter, 1988), que es un *benchmark* bastante difundido en la literatura de Redes Bayesianas (Pifer & Guedes, 2007), que consta de un conjunto de datos (*asia10K.csv*), que contiene 10.000 casos generados aleatoriamente a partir de la red bayesiana, en la tabla 1 se muestra la descripción de las variables del dataset ASIA, con su respectiva abreviatura

Tabla 1  
*Descripción de las variables en el dataset Asia*

Variable	Abreviación	Descripción	Niveles
Disnea	D	Dificultad para respirar	Sí, No
Tuberculosis	T	Presencia de tuberculosis	Sí, No
Cáncer de Pulmón	L	Presencia de cáncer de pulmón	Sí, No
Bronquitis	B	Presencia de bronquitis	Sí, No
Visita a Asia	A	Si el individuo ha visitado Asia recientemente	Sí, No
Fumador	S	Si el individuo es fumador	Sí, No
Rayos X en el tórax	X	Resultado de la radiografía de tórax	Normal, Anormal
Prueba de Tuberculosis	E	Resultado de la prueba de tuberculosis	Positivo, Negativo

### 3.3 Muestra

En la investigación propuesta, se adoptó un enfoque de muestreo estratificado para dividir el dataset en conjuntos de entrenamiento y validación, siguiendo el método sugerido por Kohavi *et al.* (1995) en su artículo *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Específicamente, el 90% de los datos se asignó para entrenamiento, mientras que el 10% restante se utilizó para validación. Esta división permitió una evaluación robusta y confiable del modelo desarrollado, asegurando que el modelo tenga suficiente capacidad para generalizar bien en datos no vistos.

### 3.4 Método de Investigación

La presente investigación es de tipo comparativo y cuantitativo. Esta investigación busca evaluar y contrastar diversas técnicas de aprendizaje estructural basadas en criterios específicos, como la precisión, eficiencia computacional y la robustez ante diferentes tipos de datos. Se adoptó un enfoque experimental, donde cada técnica fue aplicada al conjunto de

datos del modelo ASIA, descritos en la sección 3.1. Este método permitió no sólo identificar las técnicas más eficaces en distintos escenarios, sino también proporcionar una clasificación basada en su desempeño práctico.

### 3.5 Descripción de las técnicas de aprendizajes estructural aplicados

#### 3.5.1 Grow-Shrink (GS)

El algoritmo Grow-Shrink (Grow-Shrink, GS) es una técnica basada en tests de independencia condicional. Inicia con una red vacía y, a partir de un nodo seleccionado, incrementa gradualmente el conjunto de nodos adyacentes al evaluar la dependencia. Una vez que se encuentra una variable independiente condicional, se detiene el crecimiento y comienza la fase de reducción (Margaritis, 2000).

**procedimiento** GS(datos)

```
    para cada variable X en datos hacer
        MB = conjunto vacio // Manta de Markov
        mientras X sea dependiente de alguna variable Y no en MB hacer
            aumentar Y a MB
        fin mientras
        para cada variable Y en MB hacer
            si X es independiente de Y entonces
                eliminar Y de MB
            fin si
        fin para
    fin para
fin procedimiento
```

#### 3.5.2 Incremental Association Markov Blanket (IAMB)

El algoritmo Manta de Markov con Asociación Incremental ( Incremental Association Markov Blanket, IAMB) mejora el algoritmo GS añadiendo una fase posterior donde se revisan los nodos previamente descartados para reconsiderar su adición. Esta fase adicional ayuda a evitar errores prematuros al eliminar nodos (I. T. et al., 2003).

**procedimiento** IAMB(datos)

```
para cada variable X en datos hacer
    MB = conjunto vacio
mientras X sea dependiente de alguna variable Y no en MB hacer
    aumentar Y a MB
fin mientras
para cada variable Y que fue eliminada de MB hacer
    si X es dependiente de Y entonces
        aumentar Y a MB
    fin si
fin para
fin para
fin procedimiento
```

### 3.5.3 Hill Climbing (HC)

El algoritmo Hill-Climbing (HC) es un método de búsqueda local que comienza con una estructura inicial y realiza cambios iterativos en la misma, como añadir, eliminar o invertir arcos. Cada cambio se evalúa basándose en una función de puntuación (Cooper & Herskovits, 1992).

```
procedimiento HC(datos, puntuacion)
    G = grafo inicial
    mejorPuntuacion = calcularPuntuacion(G)
hacer
    puntuacionActual = mejorPuntuacion
para cada posible operac. (aumentar,eliminar,invertir arco) hacer
    G' = aplicar operacion a G
    si calcularPuntuacion(G') > mejorPuntuacion entonces
        mejorPuntuacion = calcularPuntuacion(G')
        G = G'
    fin si
fin para
mientras puntuacionActual < mejorPuntuacion
fin procedimiento
```

### 3.5.4 Tabu Search (TABU)

Tabu Search (TABU) es similar al HC, pero mantiene un registro de los últimos movimientos realizados para evitar repetirlos, evitando así quedar atrapado en óptimos locales (Glover, 1989).

**procedimiento** TS(datos, puntuacion)

G = grafo inicial

mejorPuntuacion = calcularPuntuacion(G)

ListaTabu = lista vacia

**hacer**

puntuacionActual = mejorPuntuacion

**para** cada operacion no en ListaTabu **hacer**

G' = aplicar operacion a G

**si** calcularPuntuacion(G') > mejorPuntuacion **entonces**

mejorPuntuacion = calcularPuntuacion(G')

G = G'

aumentar operacion a ListaTabu

**fin si**

**fin para**

eliminar operaciones antiguas de ListaTabu

**mientras** puntuacionActual < mejorPuntuacion

**fin procedimiento**

### 3.5.5 Max-Min Hill Climbing (MMHC)

Max-Min Hill-Climbing (MMHC) combina características del GS e IAMB con el HC. En la primera fase, se identifica el Markov Blanket de cada nodo. En la segunda fase, se realiza una búsqueda local como en el HC, pero se restringe a los arcos identificados en la primera fase (Tsamardinos *et al.*, 2006).

**procedimiento** MMHC(datos, puntuacion)

G = ejecutar GS(datos) // Paso GS

mejorPuntuacion = calcularPuntuacion(G)

**hacer**

puntuacionActual = mejorPuntuacion

**para** cada posible operacion **hacer**

G' = aplicar operacion a G

**si** calcularPuntuacion(G') > mejorPuntuacion **entonces**

```
mejorPuntuacion = calcularPuntuacion(G')
G = G'
fin si
fin para
mientras puntuacionActual < mejorPuntuacion
fin procedimiento
```

### 3.5.6 Two-phase Restricted Maximization (RSMAX2)

El algoritmo de Two-Phase Restricted Maximization (RSMAX2) es una técnica de aprendizaje de estructuras de redes bayesianas híbridas que combina métodos basados en restricciones y puntuación para estructurar la red. RSMAX2 utiliza una fase de restricción, donde se emplean algoritmos como `si.hiton.pc` para limitar el espacio de búsqueda de estructuras potenciales, seguido de una fase de maximización que utiliza métodos como HC para encontrar la estructura que maximiza una función de puntuación dada. Este enfoque híbrido está diseñado para equilibrar la eficiencia de la búsqueda con la calidad de la estructura de la red aprendida, permitiendo una exploración más dirigida y eficaz del espacio de modelos posibles (Scutari, 2021).

**procedimiento** RSMAX2 (Datos)

Entrada: Datos, un conjunto de datos **para** aprender la red bayesiana

Salida: RedBayesiana, la estructura de la red bayesiana aprendida

```
// Fase de restriccion
```

```
Restricciones := si.hiton.pc(Datos)
```

```
EstructuraRestringida := aplicar restricciones a la estructura
```

```
// Fase de maximizacion
```

```
RedBayesiana := hc(EstructuraRestringida, Datos)
```

```
// Bucle de aprendizaje
```

```
repetir
```

```
// Generar vecindario con operaciones de arco
```

```
Vecindario := generarVecindario(RedBayesiana)
```

```
para cada estructura en Vecindario hacer
```

```
si estructura no esta en Restric. y mejora la punt. entonces
```

```
RedBayesiana := estructura
  fin si
fin para

// Actualizar restricciones si es necesario
Restricciones := actualizarRestric.(RedBayesiana, Restric., Datos)
hasta que no haya mejora en la puntuac. o se alcance criterio de parada

devolver RedBayesiana
fin procedimiento
```

## 3.6 Métricas

### 3.6.1 Validación Cruzada

La validación cruzada es una técnica estadística utilizada para evaluar la habilidad de modelos predictivos. Es especialmente útil en situaciones donde la muestra de datos es limitada. En esencia, este método divide el conjunto de datos en múltiples partes, o *folds*, para luego entrenar el modelo en algunos de estos y validarlos en el resto. Este proceso se repite varias veces, cambiando el conjunto de datos de validación en cada iteración. La validación cruzada más común es la *k-fold cross-validation*, donde el conjunto de datos se divide en  $k$  partes. El promedio de las  $k$  mediciones de rendimiento en los conjuntos de validación es el rendimiento validado cruzadamente. La Figura 7 ilustra este proceso para  $k = 10$ , es decir, validación cruzada de 10 veces. En la primera iteración, el primer subconjunto sirve como conjunto de validación  $D_{val,1}$  y los nueve subconjuntos restantes sirven como conjunto de entrenamiento  $D_{train,1}$ . En la segunda iteración, el segundo subconjunto es el conjunto de validación y los subconjuntos restantes son el conjunto de entrenamiento, y así sucesivamente (Berrar *et al.*, 2019).

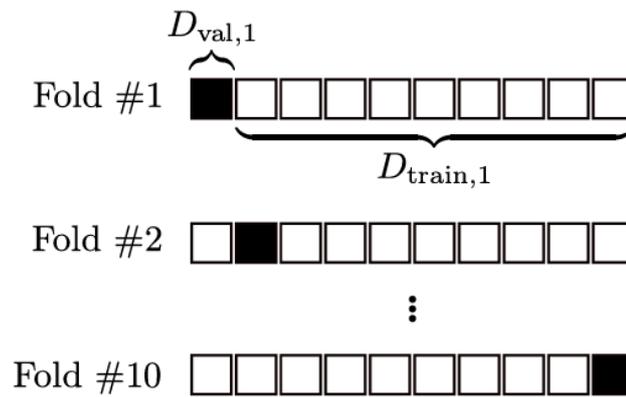


Figura 7. Validación Cruzada de 10-Folds.

fuelle: (Berrar *et al.*, 2019)

Por ejemplo, la precisión de la validación cruzada es el promedio de todas las precisiones obtenidas en los conjuntos de validación. Más generalmente, dejemos que  $\hat{f}_{-k}$  denote el modelo que fue entrenado en todos excepto en el  $k$ -ésimo subconjunto del conjunto de aprendizaje. El valor  $\hat{y}_i = \hat{f}_{-k}(x_i)$  es el valor predicho o estimado para la clase real,  $y_i$ , del caso  $x_i$ , que es un elemento del  $k$ -ésimo subconjunto. La validada cruzada estimada del error de predicción,  $\epsilon_{cv}$ , se da entonces como

$$\epsilon_{cv} = \frac{1}{n} \sum_{i=1}^n L(y_i, \hat{f}_{-k}(x_i)) \quad (1)$$

donde  $L$  es una función de pérdida que mide la diferencia entre el valor predicho y el valor real.

### 3.6.2 Criterio de Información Bayesiano (Bayesian Information Criterion, BIC)

El *Criterio de Información Bayesiana* (BIC), también conocido como el criterio de Schwarz, es una medida de selección de modelos en estadística y aprendizaje automático. Desarrollado por Gideon Schwarz en 1978, el BIC es particularmente útil cuando se comparan modelos con diferentes números de parámetros libres, o sea, modelos de diferente complejidad.

Para introducir formalmente el BIC, se considera el siguiente marco de selección de modelos. Supongamos que nos esforzamos por encontrar un modelo adecuado para describir una colección de  $n$  mediciones de respuesta  $y$ . Asumiremos que  $y$  ha sido

generado de acuerdo a una densidad desconocida  $g(y)$ . Nos referimos a  $g(y)$  como el modelo verdadero o generador. Un modelo formulado por el investigador para describir los datos y se llama un modelo candidato o modelo aproximado. Asumiremos que cualquier modelo candidato corresponde estructuralmente a una clase paramétrica de distribuciones. Específicamente, para un modelo candidato particular  $M_k$ , asumimos que existe una clase paramétrica  $k$ -dimensional de funciones de densidad (Neath & Cavanaugh, 2012).

$$\mathcal{F}(k) = \{f(y|\theta_k) : \theta_k \in \Theta(k)\},$$

una clase en la cual el espacio paramétrico  $\Theta(k)$  consiste en vectores  $k$ -dimensiones cuyos componentes son funcionalmente independientes. Permita que  $L(\theta_k|y)$  denote la verosimilitud correspondiente a la densidad  $f(y|\theta_k)$ , es decir,  $L(\theta_k|y) = f(y|\theta_k)$ . Deje que  $\hat{\theta}_k$  denote un vector de estimaciones obtenido de la maximización  $L(\theta_k|y)$  sobre  $\Theta(k)$ .

Suponiendo que se formuló una colección de modelos candidatos  $M_{k1}, M_{k2}, \dots, M_{kl}$  basados en diferentes subconjuntos de variables explicativas, diferentes estructuras de media y varianza/covarianza, e incluso diferentes especificaciones para el tipo de distribución de la variable de respuesta. El objetivo es buscar entre esta colección el modelo que 'mejor' la mejor aproximación  $g(y)$ . El BIC para el modelo candidato  $M_k$  se define como:

$$\text{BIC} = -2 \ln L(\hat{\theta}_k|y) + k \ln(n). \quad (2)$$

En la práctica, el BIC se calcula para cada uno de los modelos  $M_{k1}, M_{k2}, \dots, M_{kl}$ , y se selecciona el modelo que corresponda al valor mínimo de BIC.

El objetivo al utilizar BIC es seleccionar el modelo con el valor más bajo de BIC entre un conjunto de modelos candidatos. Un valor más bajo indica un mejor equilibrio entre la complejidad del modelo y su ajuste a los datos, asumiendo ciertas condiciones asintóticas.

Es importante señalar que el BIC se deriva bajo un marco bayesiano y asume que

los datos son generados por un modelo dentro del conjunto considerado. Además, asume que los errores en los datos son independientes y distribuidos normalmente.

En la práctica, el BIC se utiliza comúnmente para la selección de modelos en contextos como la regresión lineal múltiple, análisis de series temporales, modelos de aprendizaje y clasificación. A pesar de ser una herramienta poderosa, el BIC tiene limitaciones y no siempre puede garantizar la selección del "mejor" modelo en todas las situaciones prácticas.

### 3.6.3 Pérdida Esperada (Expected Loss)

La métrica de expected loss en la validación cruzada para el aprendizaje estructural de redes bayesianas se enfoca en calcular el promedio de una función de pérdida como el BIC o la verosimilitud negativa a través de los pliegues de validación. Cada pliegue proporciona una estimación de la pérdida en una parte del conjunto de datos. La pérdida esperada se calcula luego como el promedio de estas estimaciones.

$$E[L] = \frac{1}{k} \sum_{i=1}^k L(D_{\text{val}}^i, M^i) \quad (3)$$

donde  $E[L]$  es la pérdida esperada,  $k$  es el número de pliegues en la validación cruzada,  $L$  es la función de pérdida,  $D_{\text{val}}^i$  es el conjunto de datos de validación para el  $i$ -ésimo pliegue, y  $M^i$  es el modelo aprendido del  $i$ -ésimo pliegue del conjunto de entrenamiento.

Como criterio de selección de un modelo implica elegir el modelo que minimiza la pérdida promedio esperada sobre un conjunto de datos. En el aprendizaje de modelos, incluyendo redes bayesianas, esto podría significar seleccionar el modelo que, en promedio, tiene el menor error o la menor discrepancia entre las predicciones del modelo y los valores reales. Este enfoque es particularmente útil cuando las consecuencias de diferentes tipos de errores en la predicción varían, permitiendo que la función de pérdida refleje el costo relativo de estos errores.

En resumen indicar que la validación cruzada es una técnica crucial para evaluar la calidad y robustez de las estructuras aprendidas en las redes bayesianas. La metodología implica dividir el conjunto de datos en varios subconjuntos o *folders*. Para cada

división, un subconjunto se utiliza como conjunto de prueba mientras que los demás se usan para el entrenamiento. La estructura de la red se aprende repetidamente con los distintos subconjuntos de entrenamiento y se evalúa en el conjunto de prueba correspondiente. Una métrica común utilizada en este contexto es el Criterio de Información Bayesiano (Bayesian Information Criterion, BIC), que combina la verosimilitud del modelo con un término de penalización basado en la complejidad del mismo. El valor medio BIC obtenido de todas las iteraciones de la validación cruzada proporciona una estimación robusta de la calidad general de la estructura aprendida, equilibrando el ajuste del modelo a los datos y su simplicidad. Además como criterio adicional de selección la pérdida esperada (Expected Loss) de los k-folds.

### 3.6.4 Hardware y Software

#### Hardware

La comparación y clasificación de técnicas de aprendizaje estructural de redes bayesianas requiere de un hardware adecuado dada la naturaleza computacionalmente intensiva de estos procesos. Fue esencial contar con un procesador de múltiples núcleos o, preferentemente, múltiples procesadores para paralelizar los cálculos y reducir los tiempos de ejecución. Una cantidad sustancial de memoria RAM (al menos 16 GB, aunque 32 GB o más sería ideal) es necesaria para manejar grandes conjuntos de datos y realizar múltiples simulaciones simultáneamente. Además, si se planea implementar algoritmos que se benefician de la computación en GPU, como algunas redes neuronales bayesianas, una tarjeta gráfica potente con una memoria dedicada adecuada es imprescindible. Finalmente, un sistema de almacenamiento rápido, como un SSD, permitiría un acceso ágil a los datos y guardar resultados intermedios con eficiencia.

#### Software

El software utilizado durante el desarrollo para la investigación se dividió según su necesidad, para el análisis de datos y modelado, se destacaron R, Netica y Hugin como principales herramientas de análisis. Python es la principal herramienta de codificación, conocida por su versatilidad y amplia gama de aplicaciones. En

términos de librerías específicas, Scikit-learn en Python y Bnlearn en R fueron esenciales para la ciencia de datos y el aprendizaje estructural de las redes Bayesianas, y finalmente Overleaf que es una herramienta fundamental tanto por su necesidad para aquellos que requieren un entorno de edición LaTeX en línea como para la generación de la documentación de este trabajo de investigación, cuyos software se detallan a continuación:

- **Python:** Python es un lenguaje de programación de alto nivel, interpretado y con una sintaxis clara y legible. Ha ganado popularidad en diversos campos, especialmente en el desarrollo web, automatización y, sobre todo, en el análisis de datos y aprendizaje automático. Su extensa biblioteca estándar, junto con una amplia gama de paquetes adicionales, hace que Python sea una herramienta versátil para todo tipo de aplicaciones.
- **Scikit-learn:** Scikit-learn es una librería de Python especializada en herramientas para el aprendizaje automático. Proporciona algoritmos de clasificación, regresión y clustering simples y eficientes, así como herramientas para la selección y evaluación de modelos. Su interoperabilidad con otras librerías de Python, como NumPy y Pandas, la convierte en una opción predilecta para el análisis predictivo.
- **R:** R es un lenguaje de programación y ambiente de software libre orientado a la estadística y la visualización de datos. Es ampliamente utilizado por estadísticos, bioinformáticos y científicos de datos para el análisis y modelado de datos complejos. Ofrece una variedad de paquetes especializados para diferentes áreas de investigación y aplicaciones.
- **Bnlearn:** Bnlearn es un paquete de R diseñado para el aprendizaje y la inferencia con redes bayesianas. Ofrece herramientas para la estructuración, parametrización y utilización de estos modelos, permitiendo a los usuarios implementar algoritmos avanzados de aprendizaje estructural y realizar inferencias basadas en los datos.
- **Netica:** Netica es una aplicación de software para trabajar con redes bayesianas. Permite construir, entrenar, y utilizar estas redes para realizar inferencias. Con una interfaz gráfica intuitiva, Netica facilita la visualización y

manipulación de estructuras complejas, siendo una herramienta valiosa para expertos y novatos en el campo.

- **Hugin:** Hugin es otro software destacado en el ámbito de las redes bayesianas. Proporciona herramientas avanzadas para la construcción, entrenamiento e inferencia con estos modelos. Su capacidad para manejar redes grandes y complejas, junto con una serie de características avanzadas, lo hace adecuado para aplicaciones industriales y de investigación.
- **Overleaf:** Overleaf es un editor de LaTeX en línea, que permite a los usuarios crear, editar y compartir documentos en este formato de manera colaborativa. Su interfaz intuitiva, la previsualización en tiempo real y las capacidades de colaboración lo han convertido en una opción popular para la redacción de documentos académicos, artículos científicos y otros trabajos que requieran una presentación de alta calidad.

## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

Con base en el objetivo general y los objetivos específicos planteados en la investigación, se describe los resultados de la aplicación de la metodología descrita en el capítulo III, los cuales se muestran a continuación:

#### 4.1 Aplicación de los algoritmos de aprendizaje Estructural

Se evaluaron y entrenaron los algoritmos indicados en la sección 3.5 que se describen las siguientes secciones:

##### 4.1.1 Aplicación del algoritmo Grow-Shrink (GS)

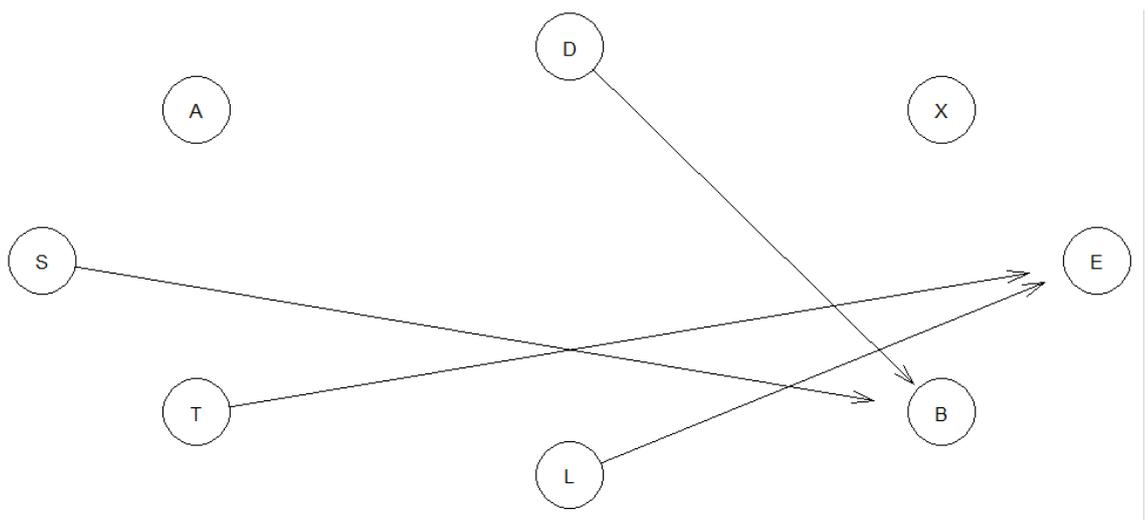


Figura 8. Representación de DAG aplicando el algoritmo Grow-Shrink (GS)

Tabla 2  
*Resumen de la Red Bayesiana aprendida aplicando el algoritmo Grow-Shrink (GS)*

Characteristic	Value
Model	$[A][S][T][L][X][D][B S : D][E T : L]$
Nodes	8
Arcs	4
Undirected arcs	0
Directed arcs	4
Average Markov blanket size	1.50
Average neighbourhood size	1.00
Average branching factor	0.50
Learning algorithm	Grow-Shrink
Conditional independence test	Mutual Information (disc.)
Alpha threshold	0.05
Tests used in the learning procedure	135

En la tabla 2 La Red Bayesiana descrita ha sido aprendida mediante métodos basados en restricciones. La estructura de la red consta de 8 nodos interconectados por 4 arcos dirigidos, lo que implica que todas las relaciones son causales. El tamaño promedio de la manta de Markov, que refleja el conjunto de nodos adyacentes a un nodo específico, es de 1.50. Esta red fue aprendida utilizando el algoritmo Grow-Shrink y emplea la métrica de Información Mutua discreta para evaluar la independencia condicional entre nodos con un umbral de significancia de 0.05. Durante el proceso de aprendizaje, se realizaron 135 pruebas de independencia.

#### 4.1.2 Aplicación del algoritmo Incremental Association Markov Blanket (IAMB)

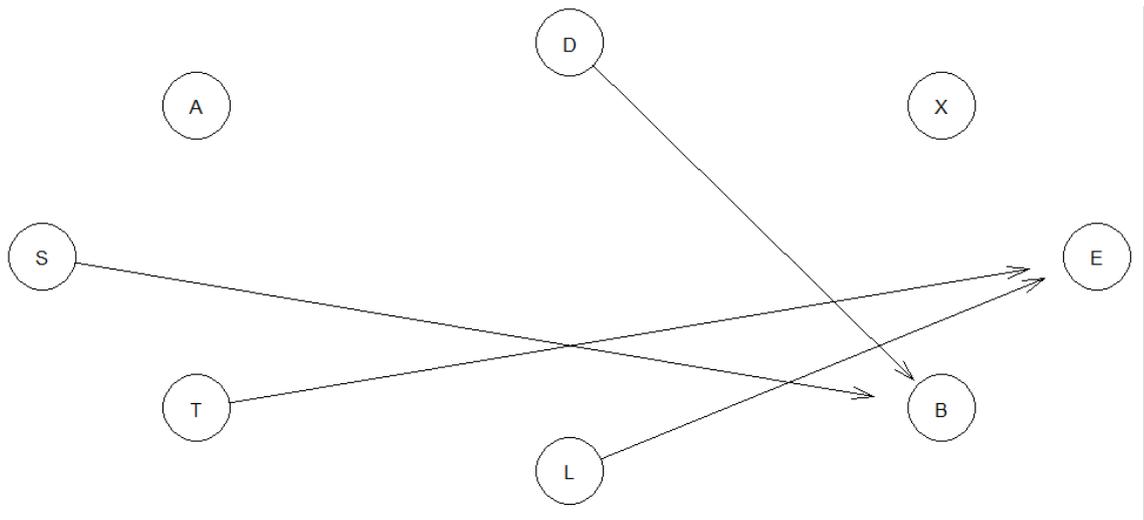


Figura 9. Representación de DAG aplicando el algoritmo Incremental Association Markov Blanket (IAMB)

Tabla 3

Resumen de la Red Bayesiana aprendida aplicando el Incremental Association Markov Blanket (IAMB)

Characteristic	Value
Model	$[A][S][T][L][X][D][B S : D][E T : L]$
Nodes	8
Arcs	4
Undirected arcs	0
Directed arcs	4
Average Markov blanket size	1.50
Average neighbourhood size	1.00
Average branching factor	0.50
Learning algorithm	IAMB
Conditional independence test	Mutual Information (Disc.)
Alpha threshold	0.05
Tests used in the learning procedure	173

En la tabla 3 se observa que la Red Bayesiana descrita ha sido aprendida mediante métodos basados en restricciones. La estructura de la red consta de 8 nodos interconectados por 4 arcos dirigidos, lo que implica que todas las relaciones son causales. El tamaño promedio de la manta de Markov, que refleja el conjunto de nodos adyacentes a un nodo específico, es de 1.50. Esta red fue aprendida utilizando el algoritmo Incremental Association Markov Blanket (IAMB) y emplea la métrica de Información Mutua discreta para evaluar la independencia condicional entre nodos con un umbral de significancia de 0.05. Durante el proceso de aprendizaje, se realizaron 173 pruebas de independencia.

En los resultados mostrados en la figura 8 y la figura 9, así como en las tablas 2 y 3, la estructura aprendida de las redes bayesianas son similares, sobre esto Scutari *et al.* (2018) indica que el algoritmo GS es una implementación práctica del trabajo seminal de Pearl sobre modelos gráficos causales, y junto con el IAMB, utiliza heurísticas más rápidas para implementar las primeras etapas de aprendizaje de la estructura, lo que puede llevar a diferencias en las estructuras aprendidas. Esto implica que aunque hay situaciones en las que los resultados pueden ser similares, es posible que los dos métodos produzcan diferentes estructuras debido a la manera en que abordan el aprendizaje y a los datos específicos con los que trabajan.

### 4.1.3 Aplicación del algoritmo Hill Climbing (HC)

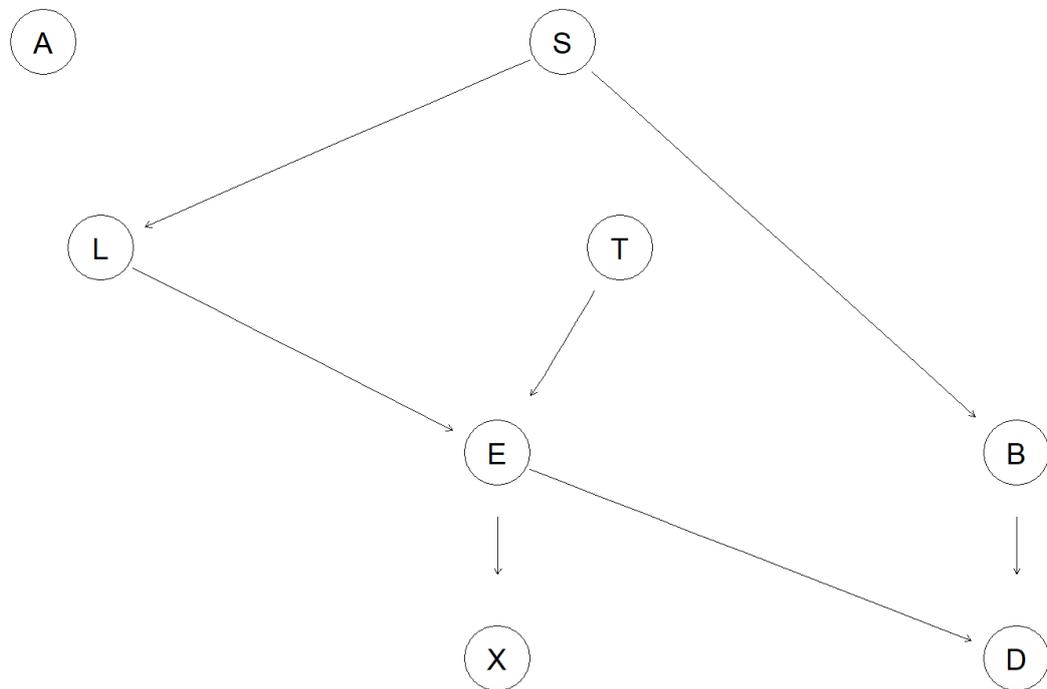


Figura 10. Representación de DAG aplicando el algoritmo Hill Climbing (HC)

Tabla 4  
*Resultado del Aprendizaje de Redes Bayesianas con Método Basado en Hill Climbing (HC)*

Descripción	Valor
Model	[A][S][T][L][S][B][S][E][T:L][X][E][D][B:E]
Nodes	8
Arcs	7
Undirected arcs	0
Directed arcs	7
Average markov blanket size	2.25
Average neighbourhood size	1.75
Average branching factor	0.88
Learning algorithm	Hill-Climbing
Score	BIC (disc.)
Penalization coefficient	4.205916
Tests used in the learning procedure	77
Optimized	TRUE

El resultado presentado en la tabla 4 es una red bayesiana aprendida mediante métodos basados en puntuación. La red consta de 8 nodos y 7 arcos dirigidos, sin arcos no dirigidos. La manta de Markov promedio para los nodos es de 2.25, indicando la cantidad promedio de nodos que, una vez conocidos, hacen que otros nodos sean independientes del nodo de interés. El tamaño medio del vecindario es de 1.75, reflejando la cantidad promedio de nodos conectados directamente a cada nodo. El factor medio de ramificación, 0.88, señala el promedio de hijos que tiene un nodo. Se empleó el algoritmo *Hill-Climbing* con una puntuación BIC discreta y un coeficiente de penalización de 4.205916. Durante el procedimiento de aprendizaje, se realizaron 77 pruebas y el proceso fue optimizado.

#### 4.1.4 Aplicación del algoritmo Tabu Search (TABU)

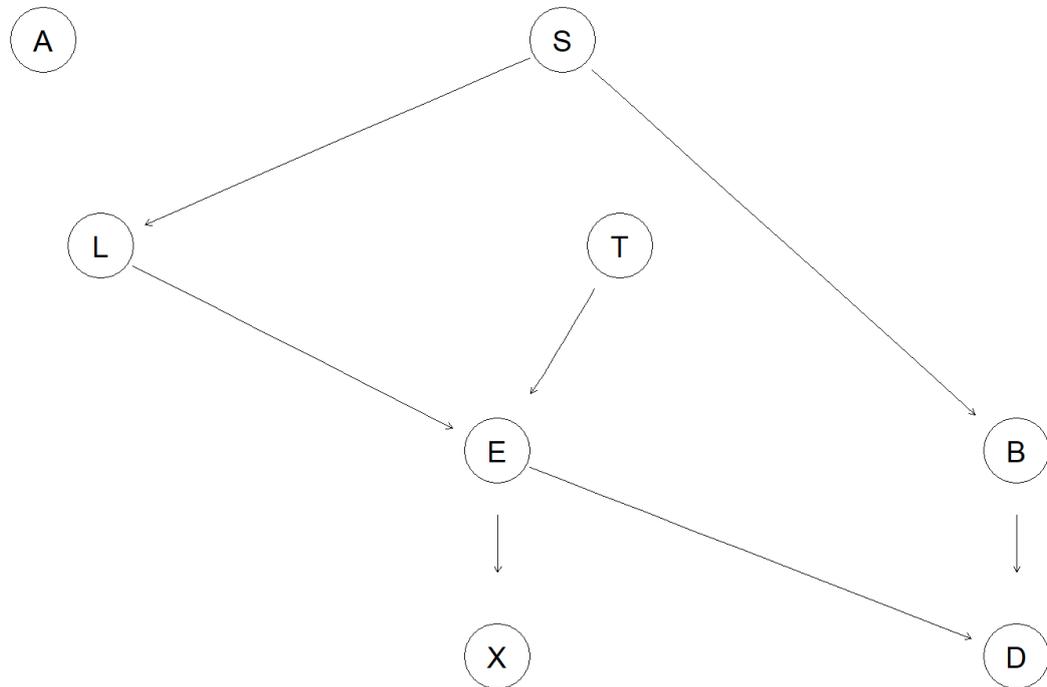


Figura 11. Representación de DAG aplicando el algoritmo Tabu Search (TABU)

Tabla 5  
*Resultado del Aprendizaje de Redes Bayesianas con Métodos Basados Tabu Search (TABU)*

Descripción	Valor
Model	[A][S][T][L][S][B][S][E][T:L][X][E][D][B:E]
Nodes	8
Arcs	7
Undirected arcs	0
Directed arcs	7
Average markov blanket size	2.25
Average neighbourhood size	1.75
Average branching factor	0.88
Learning algorithm	Tabu search
Score	BIC (disc.)
Penalization coefficient	4.258597
Tests used in the learning procedure	204
Optimized	TRUE

En la tabla 5 se muestra el resultado de una red bayesiana aprendida mediante métodos basados en puntuación. La red consta de 8 nodos y 7 arcos dirigidos, sin arcos no dirigidos. La manta de Markov promedio para los nodos es de 2.25, indicando la cantidad promedio de nodos que, una vez conocidos, hacen que otros nodos sean independientes del nodo de interés. El tamaño medio del vecindario es de 1.75, reflejando la cantidad promedio de nodos conectados directamente a cada nodo. El factor medio de ramificación, 0.88, señala el promedio de hijos que tiene un nodo. Se empleó el algoritmo Tabu Search (TABU) con una puntuación BIC discreta y un coeficiente de penalización de 4.258597. Durante el procedimiento de aprendizaje, se realizaron 204 pruebas y el proceso fue optimizado.

Ji *et al.* (2011) sugiere que Hill-Climbing (HC) y Tabu Search (TABU) ambos son métodos de búsqueda y puntuación y que han sido utilizados en el aprendizaje de estructuras de redes bayesianas. Dado por la naturaleza de dos métodos, esto explicaría

por que las estructura resultaron iguales con la diferencia que se realizaron diferentes números de pruebas en el proceso de aprendizaje 77 y 204 respectivamente y coeficiente de penalización 4.205916 y 4.258597.

#### 4.1.5 Aplicación del algoritmo Max-Min Hill Climbing (MMHC)

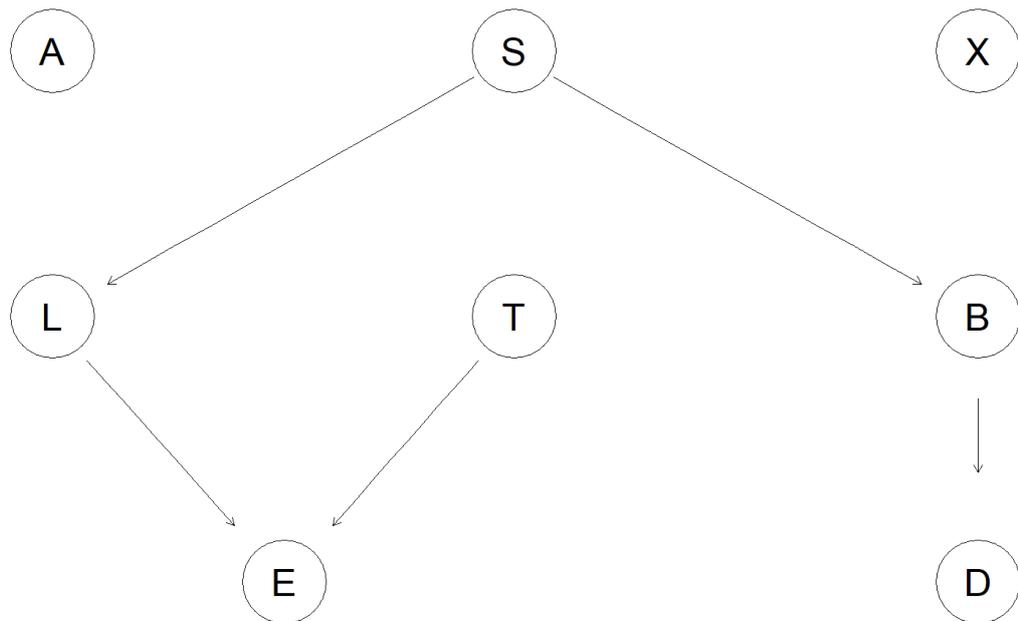


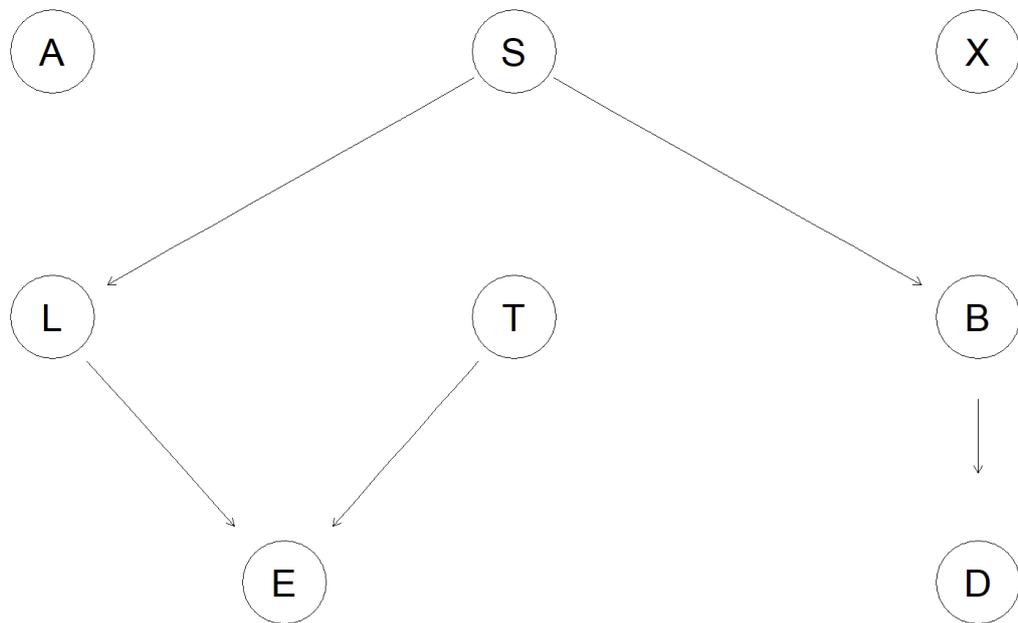
Figura 12. Representación de DAG aplicando el algoritmo Aplicación del algoritmo Max-Min Hill Climbing (MMHC)

Tabla 6  
*Resultado del Aprendizaje de Redes Bayesianas mediante el Método Max-Min Hill Climbing (MMHC)*

Descripción	Valor
Model	[A][S][T][X][L S][B S][E T:L][D B]
Nodes	8
Arcs	5
Undirected arcs	0
Directed arcs	5
Average markov blanket size	1.50
Average neighbourhood size	1.25
Average branching factor	0.62
Learning algorithm	Max-Min Hill-Climbing
Constraint-based method	Max-Min Parent Children
Conditional independence test	Mutual Information (disc.)
Score-based method	Hill-Climbing
Score	BIC (disc.)
Alpha threshold	0.05
Penalization coefficient	4.258597
Tests used in the learning procedure	254
Optimized	TRUE

La tabla 6 muestra una red bayesiana aprendida a través de métodos híbridos. Dicha red consta de 8 nodos con 5 arcos dirigidos. Se observa un tamaño medio de la manta de Markov de 1.50, un tamaño promedio del vecindario de 1.25 y un factor medio de ramificación de 0.62. Estos valores reflejan la estructura y relaciones dentro de la red. El algoritmo principal de aprendizaje es el Max-Min Hill-Climbing (MMHC), combinando el método *Max-Min Parent Children* para las restricciones y el método MMHC para la puntuación. Se utiliza la puntuación BIC discreta, con un umbral alfa de 0.05 y un coeficiente de penalización de 4.258597. Durante el proceso de aprendizaje, se llevaron a cabo 254 pruebas, y el método está optimizado.

#### 4.1.6 Aplicación del algoritmo Two-Phase Restricted Maximization (RSMAX2)



*Figura 13.* Representación de DAG aplicando el algoritmo Aplicación del algoritmo Two-Phase Restricted Maximization (RSMAX2)

Tabla 7  
*Resultado del Aprendizaje de Redes Bayesianas mediante el Método Two-Phase Restricted Maximization (RSMAX2)*

Descripción	Valor
Model	[A][S][T][X][L S][B S][E T:L][D B]
Nodes	8
Arcs	5
Undirected arcs	0
Directed arcs	5
Average markov blanket size	1.50
Average neighbourhood size	1.25
Average branching factor	0.62
Learning algorithm	Two-Phase Restricted Maximization
Constraint-based method	Semi-Interleaved HITON-PC
Conditional independence test	Mutual Information (disc.)
Score-based method	Hill-Climbing
Score	BIC (disc.)
Alpha threshold	0.05
Penalization coefficient	4.258597
Tests used in the learning procedure	157
Optimized	TRUE

La tabla 7 muestra la red bayesiana resultante, aprendida a través de métodos híbridos, exhibe una estructura con ocho nodos y cinco arcos dirigidos, lo que indica una simplicidad en su configuración al no presentar arcos no dirigidos. El tamaño promedio del manto de Markov es de 1.50 y el tamaño promedio de vecindario es de 1.25, con un factor de ramificación promedio de 0.62, reflejando una conectividad moderada entre los nodos de la red. El proceso de aprendizaje estructural se efectuó utilizando el algoritmo RSMAX2 (Two-Phase Restricted Maximization), que es un método híbrido que combina un enfoque basado en restricciones, implementado aquí a través del Semi-Interleaved HITON-PC, y un enfoque basado en puntuación,

ejecutado mediante el algoritmo Hill-Climbing. Para las pruebas de independencia condicional, se empleó la Información Mutua en su forma discreta, y la evaluación de las estructuras de la red se llevó a cabo utilizando el criterio BIC (Bayesian Information Criterion), también en su modalidad discreta. Se estableció un umbral alfa de 0.05 y se aplicó un coeficiente de penalización de 4.258597, indicativo de una fuerte penalización para la complejidad de la red, lo que posiblemente influyó en la simplicidad de la estructura final de la red. A lo largo del procedimiento de aprendizaje, se realizaron un total de 157 pruebas, y el atributo 'optimizado' se ha marcado como 'TRUE', sugiriendo que el algoritmo logró encontrar una estructura que considera óptima de acuerdo con los parámetros establecidos.

La estructura de la red bayesiana generada por Two-Phase Restricted Maximization (RSMAX2) resultó idéntica a la generada por el método Max-Min Hill-Climbing (MMHC), como afirma Scutari (2010) con la configuración adecuada, podrían generar estructuras de red idénticas. El método de aprendizaje estructural de redes bayesianas Two-Phase Restricted Maximization (RSMAX2) y el algoritmo Max-Min Hill Climbing (MMHC) son técnicas híbridas que, bajo ciertas condiciones, pueden producir estructuras de red equivalentes. Aunque cada uno tiene su propia metodología, la función `mmhc()` puede considerarse un caso específico de `rsmax2()` con ciertos parámetros predefinidos. Si la fase de restricción en RSMAX2 se configura para usar `mmhc` y la fase de maximización para usar Hill-Climbing (`hc`), los dos procesos pueden teóricamente converger hacia la misma estructura de red bayesiana.

## 4.2 Comparación de las técnicas de aprendizaje de estructuras de una BN

La comparación de técnicas de aprendizaje estructural en redes bayesianas es esencial para determinar cuál de ellas es más eficaz en diferentes escenarios de datos. Una técnica ampliamente aceptada para evaluar y comparar estos métodos es la *validación cruzada*. Esta técnica divide el conjunto de datos en  $k$  partes; se entrena el modelo en  $k - 1$  partes y se valida en la parte restante. Este proceso se repite  $k$  veces, utilizando una parte diferente como conjunto de validación en cada iteración. El error de validación cruzada se calcula promediando los errores obtenidos en las  $k$  iteraciones. La ventaja de este enfoque es que utiliza todo el conjunto de datos tanto para el entrenamiento como para la validación,

proporcionando una estimación robusta del rendimiento del modelo en datos no vistos (Hastie *et al.*, 2009).

Basándonos en el criterio del valor medio BIC (Bayesian Information Criterion) como se muestra en la tabla 8 y figura 14, podemos observar que las técnicas Grow-Shrink (GS) e IAMB tienen el mismo BIC, que es el más bajo entre todos los métodos. Un valor BIC más bajo indica un mejor ajuste del modelo a los datos, teniendo en cuenta la complejidad del modelo. Esto sugiere que, para este conjunto de datos particular, GS e IAMB podrían ser las técnicas preferidas. Sin embargo, Hill Climbing (HC) y Tabu Search (TABU) tienen valores BIC ligeramente más altos, pero podrían ser considerados dependiendo de otras métricas o requerimientos específicos. MMHC y RSMAX2, por otro lado, tiene un valor intermedio entre GS/IAMB y HC/TABU, lo que sugiere un equilibrio entre la simplicidad del modelo y el ajuste a los datos.

Tabla 9

*Resultados del valor BIC en cada fold de la validación cruzada*

GS	IAMB	HC	TABU	MMHC	RSMAX2
-1307.994	-1307.994	-1185.44	-1185.44	-1290.542	-1290.542
-1252.374	-1252.374	-1134.686	-1134.686	-1233.615	-1233.615
-1260.371	-1260.371	-1150.08	-1150.08	-1239.614	-1239.614
-1318.351	-1318.351	-1232.199	-1232.199	-1302.994	-1302.994
-1243.711	-1243.711	-1152.098	-1152.098	-1235.312	-1235.312
-1258.163	-1258.163	-1155.834	-1155.834	-1244.201	-1244.201
-1302.29	-1302.29	-1179.364	-1179.364	-1275.719	-1275.719
-1294.339	-1294.339	-1135.235	-1135.235	-1254.749	-1254.749
-1345.857	-1345.857	-1192.337	-1192.337	-1312.459	-1312.459
-1269.32	-1269.32	-1165.495	-1165.495	-1255.987	-1255.987

Tabla 8  
*Comparación de diferentes técnicas de aprendizaje estructural.*

Técnicas de Aprendizaje	Grow-Shrink (GS)	IAMB	Hill Climbing (HC)	Tabu Search (TABU)	MMHC	RSM
Nodos	8.000	8.000	8.000	8.000	8.000	8.0
Arcos	4.000	4.000	7.000	7.000	5.000	5.0
Arcos no dirigidos	0.000	0.000	0.000	0.000	0.000	0.0
Arcos dirigidos	4.000	4.000	7.000	7.000	5.000	5.0
Tamaño medio de Markov Blanket	1.500	1.500	2.250	2.250	1.500	1.5
Tamaño medio del vecindario	1.000	1.000	1.750	1.750	1.250	1.2
Factor de ramificación promedio	0.500	0.500	0.880	0.880	0.620	0.6
Umbral alfa	0.050	0.050	-	-	0.050	0.0
Pruebas usadas en el aprendizaje	135.000	169.000	77.000	204.000	256.000	157.000
BIC medio	-1285.277019	-1285.277019	-1168.276828	-1168.276828	-1264.519291	-1264.519291

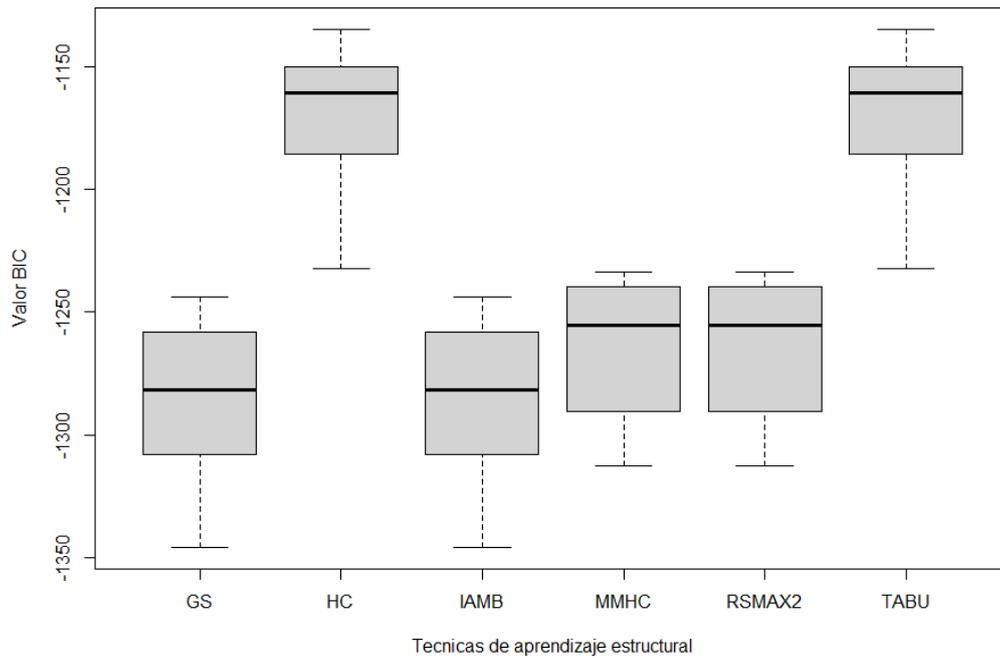


Figura 14. Diagrama de Cajas de los valores BIC generadas en cada fold de la Validación Cruzada

La prueba ANOVA resultó ser significativa  $p - valor = 3,39 \times 10^{-15}$ , lo que indica diferencias significativas entre los grupos, con un nivel de significancia del 0.05 ( $\alpha = 0,05$ ) y confianza de 0.95 ( $1 - \alpha = 0,95$ ), es decir las técnicas de aprendizaje estructural de BN son diferentes, concordando con la figura 14 generada a partir de la tabla 9. Como se muestra en la tabla de análisis de varianza, tabla 10.

Tabla 10

Análisis de varianza para las técnicas de aprendizaje estructural

Fuente de Variación	GL	SC	CM	$F_c$	P-valor
Técnicas de aprendizaje	5	155884	31177	32.73	3.39E-15
Total	59	207325			

### 4.3 Clasificación de las técnicas de aprendizaje

La clasificación de técnicas de aprendizaje automático frecuentemente involucra el uso de criterios estadísticos para evaluar y comparar modelos. Un enfoque común es el uso del Criterio de Información Bayesiano (BIC), que equilibra la complejidad del modelo contra

el ajuste de los datos, penalizando los modelos más complejos para evitar el sobreajuste. En un procedimiento de validación cruzada, donde el conjunto de datos se divide en  $k$  subconjuntos ( $k = 10$  en este caso), el BIC medio de estos pliegues provee una base sólida para la evaluación del rendimiento del modelo.

### **Análisis Post Hoc con Prueba de Tukey**

Después de la evaluación inicial con el BIC, se realizó un análisis post hoc para hacer comparaciones adicionales entre las técnicas. La prueba post hoc de Tukey, o la Prueba Honestamente Significativa de Diferencias (HSD) de Tukey, es un método estadístico utilizado comúnmente para comparar pares de medias y determinar si las diferencias entre ellas son estadísticamente significativas. Este análisis es útil cuando se han realizado múltiples comparaciones y se desea controlar la tasa de error de Tipo I.

Este método de análisis post hoc es crucial para desglosar las comparaciones entre las técnicas de aprendizaje de una manera más granular, proporcionando una comprensión detallada de las diferencias en el rendimiento. Además, discute las implicaciones prácticas de estos métodos para la selección de modelos en aplicaciones del mundo real y cómo pueden influir en la toma de decisiones en la implementación de sistemas de aprendizaje automático.

Tabla 11  
*Clasificación de técnicas de aprendizaje estructural basados en BIC aplicando Tukey*

Técnica	valor medio BIC	Grupo
Hill-Climbing (HC)	-1168.277	a
Tabu Search (TABU)	-1168.277	a
Max-Min Hill-Climbing (MMHC)	-1264.519	b
Two-Phase Restricted Maximization (RSMAX2)	-1264.519	b
Grow-Shrink (GS)	-1285.277	b
Incremental Association Markov Blanket (IAMB)	-1285.277	b

La evaluación comparativa de las técnicas de aprendizaje estructural, tal como se refleja en la tabla 11 y figura 15, proporciona una visión detallada del rendimiento de cada método aplicado al conjunto de datos ASIA. Los resultados han sido categorizados en dos grupos

distintos, denominados 'a' y 'b'. El grupo 'a' demuestra un ajuste superior, lo que sugiere una mayor precisión en la modelización de las dependencias entre los datos. En contraste, el grupo 'b' exhibe un ajuste intermedio, que podría ser preferible en situaciones donde se requiere un balance entre la simplicidad del modelo y la exactitud de las predicciones. Estas diferencias subrayan la importancia de una elección meticulosa de la técnica de aprendizaje estructural para alinearse con las particularidades inherentes al conjunto de datos y con los objetivos específicos de modelado.

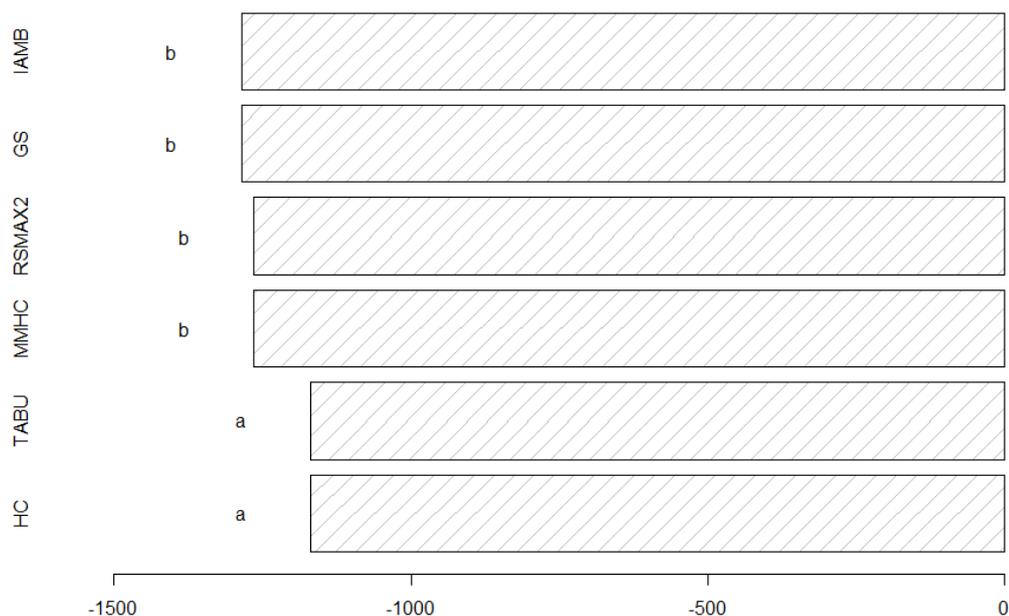


Figura 15. Clasificación de las técnicas de aprendizaje estructural aplicando Tukey

### 1. Técnicas con el Mejor Ajuste (BIC más bajo):

- Hill Climbing (HC)
- Tabu Search (TABU)

### 2. Técnica con Ajuste Intermedio:

- MMHC (Max-Min Hill Climbing)
- Two-Phase Restricted Maximization (RSMAX2)
- Grow-Shrink (GS)
- IAMB (Incremental Association Markov Blanket)

Tabla 12

*Perdida esperada de la función de perdida log-verosimilitud disc. en la validación cruzada de los métodos de aprendizaje estructural de una BN*

Algoritmos	Expected Loss
Grow-Shrink (GS)	2.495107
Incremental Association MarkovBlanket (IAMB)	2.495477
Hill Climbing (HC)	2.209267
Tabu Search (TABU)	2.209924
Max-Min Hill Climbin (MMHC)	2.423207
Two-Phase Restricted Maximization (RSMAX)	2.42319

En la tabla 12 La evaluación de los métodos de aprendizaje estructural para redes bayesianas basada en la pérdida esperada revela que los algoritmos Hill Climbing (HC) y Tabu Search (TABU) presentan las menores pérdidas, indicando una posible mayor precisión en la reconstrucción de la estructura de la red. En contraste, los métodos Grow-Shrink (GS) y Incremental Association Markov Blanket (IAMB) exhiben una pérdida esperada más alta. La literatura sugiere que los algoritmos de restricción y puntuación pueden seleccionar estructuras idénticas bajo ciertas condiciones, y que los algoritmos híbridos, como el MMHC y el RSMAX2, podrían ofrecer ventajas en términos de velocidad y precisión en diferentes contextos y tamaños de conjuntos de datos. Scutari *et al.* (2018), en una comparación general entre diversos algoritmos, muestran que los algoritmos híbridos no son más rápidos que los algoritmos basados en restricciones o puntuación, así como no hay diferencia sistemática entre la precisión de los algoritmos híbridos y los basados en restricciones. Estos hallazgos enfatizan la importancia de considerar la eficiencia computacional y la escalabilidad al seleccionar un algoritmo para el aprendizaje estructural, particularmente en datos de alta dimensión o conjuntos de datos complejos (Scutari *et al.*, 2018).

Los métodos de aprendizaje estructural de redes bayesianas se pueden clasificar según la pérdida esperada de la función de pérdida log-verosimilitud. Los algoritmos Hill Climbing (HC) y Tabu Search (TABU) presentan las pérdidas esperadas más bajas, indicando que son posiblemente los más precisos para la reconstrucción estructural de la red bayesiana. Los métodos Grow-Shrink (GS) e Incremental Association Markov Blanket (IAMB)



muestran pérdidas ligeramente mayores, mientras que Max-Min Hill Climbing (MMHC) y Two-Phase Restricted Maximization (RSMAX2) tienen las mayores pérdidas esperadas, sugiriendo una precisión menor en sus estructuras aprendidas.

El aprendizaje de redes bayesianas es un problema NP-hard y, a medida que aumenta el número de nodos, los algoritmos clásicos se vuelven ineficientes (Kaminsky & Deeva, 2022). Los métodos que limitan el espacio de búsqueda, como los algoritmos de estructuras locales que utilizan la información mutua (MI) para conectar nodos, pueden ser más efectivos en la gestión de la complejidad computacional y el tamaño de los datos (Liu *et al.*, 2016). Además, se ha desarrollado el algoritmo BigBraveBN, que limita el espacio de búsqueda y se compara favorablemente con los algoritmos existentes (Kaminsky & Deeva, 2022).

Sin embargo, la elección final de la técnica dependerá de otros factores, como la capacidad de interpretación, el tiempo de ejecución y las especificidades del problema en cuestión.

## CONCLUSIONES

Se realizó la comparación y clasificación de técnicas de aprendizaje estructural de la Red Bayesiana, utilizando un dataset ASIA, considerando un bentmach en la literatura bayesiana e investigaciones de inteligencia artificial alcanzando las siguientes conclusiones:

1. La investigación profundizó en el complejo mundo del aprendizaje automático y las redes bayesianas. Usando las técnicas populares como Grow-Shrink (Grow-Shrink, GS), Manta de Markov con Asociación Incremental ( Incremental Association Markov Blanket, IAMB), Hill-Climbing (HC), Tabu Search (TABU), Max-Min Hill-Climbing (MMHC) y Two-Phase Restricted Maximization (RSMAX2). Cada técnica fue evaluada en términos de su capacidad para identificar y modelar las relaciones entre diferentes variables en redes bayesianas. Encontrándose diferencia significativa con un  $p - valor = 3,39 \times 10^{-15}$ , con un nivel de significancia del 0.05 ( $\alpha = 0,05$ ).
2. El proceso de validación cruzada se llevó a cabo con especial cuidado y atención al detalle. Los resultados de esta validación cruzada señalaron que las técnicas Hill Climbing (HC) y Tabu Search (TABU) destacaron del resto, obteniendo un valor BIC medio (-1168.277) que las colocó por encima de sus competidoras. El resultado sugiere que, para el conjunto de datos utilizado, Hill Climbing (HC) y Tabu Search (TABU) son técnicas más eficaz y precisas para identificar estructuras óptimas en redes bayesianas, el cual fue corroborado por el valor de la pérdida esperada (Expected Loss, 2.209267).
3. La clasificación de las técnicas de aprendizaje fue otro hallazgo del trabajo. agrupando en dos grupos a y b siendo parte del grupo a las técnicas: Hill Climbing (HC) y Tabu Search (TABU) detacando un mejor ajuste (BIC más bajo) y el grupo b las técnicas de Max-Min Hill Climbing (MMHC), Grow-Shrink (GS), Incremental Association Markov Blanket (IAMB) y Two-Phase Restricted Maximization (RSMAX2). La clasificación proporciona una guía útil para aquellos investigadores y profesionales que buscan seleccionar la técnica más adecuada para sus propias investigaciones o aplicaciones.

## RECOMENDACIONES

En el ámbito de las recomendaciones, es imperativo que las futuras investigaciones consideren la incorporación de un espectro aún más amplio de técnicas de aprendizaje. Explorar diferentes parámetros y configuraciones para cada técnica podría arrojar nuevos insights y mejorar la precisión de los modelos. Además, es fundamental tener presente las limitaciones computacionales. Algunas técnicas, aunque potencialmente más precisas, podrían ser computacionalmente intensivas, lo que dificultaría su aplicación en grandes conjuntos de datos.

Comparar algoritmos de aprendizaje siempre es un desafío. Existe una amplia variedad de métricas y criterios de evaluación, y no siempre es sencillo determinar cuál es el más adecuado para un conjunto de datos o problema específico. Por lo tanto, es crucial adoptar un enfoque sistemático y basado en evidencia para estas comparaciones.

Como propuesta para trabajos futuros, se ve con entusiasmo la posibilidad de integrar técnicas emergentes basadas en inteligencia artificial y aprendizaje profundo. La aplicación de estas técnicas en diferentes conjuntos de datos no solo validaría los resultados actuales sino que también podría abrir nuevas vías de investigación, llevando el campo del aprendizaje estructural de redes bayesianas a nuevos horizontes.

## BIBLIOGRAFÍA

- Abellán, J., Gómez-Olmedo, M., Moral, S., *et al.* (2006). Some Variations on the PC Algorithm. *Probabilistic graphical models*, 1-8.
- Afrin, T., & Yodo, N. (2021). A probabilistic estimation of traffic congestion using Bayesian network. *Measurement*, 174, 109051.
- Ara, A. (2023). *Redes Bayesianas e Aplicações* [Accessed: 2023-10-24]. <http://leg.ufpr.br/~ara/teach/>
- Arora, P., Boyne, D., Slater, J. J., Gupta, A., Brenner, D. R., & Druzdzal, M. J. (2019). Bayesian networks for risk prediction using real-world data: a tool for precision medicine. *Value in Health*, 22(4), 439-445.
- Bayes Server. (2023). Introduction to Bayesian networks [Accessed: 2023-12-18]. <https://www.bayesserver.com/docs/introduction/bayesian-networks/>
- Beretta, S., Castelli, M., Gonçalves, I., Henriques, R., Ramazzotti, D., *et al.* (2018). Learning the structure of bayesian networks: A quantitative assessment of the effect of different algorithmic schemes. *Complexity*, 2018.
- Berrar, D., *et al.* (2019). Cross-Validation.
- Cai, B., Huang, L., & Xie, M. (2017). Bayesian Networks in Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, 13(5), 2227-2240. <https://doi.org/10.1109/TII.2017.2695583>
- Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3, 507-554.
- Colombo, D., Maathuis, M. H., *et al.* (2014). Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*, 15(1), 3741-3782.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9, 309-347.
- Costa, F. S., *et al.* (2013). Aprendizagem estrutural de redes Bayesianas pelo método de Monte Carlo e Cadeias de Markov.
- Cussens, J. (2011). Bayesian Network Learning with Cutting Planes. *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence*, 153-160.

- Drury, B., Valverde-Rebaza, J., Moura, M.-F., & de Andrade Lopes, A. (2017). A survey of the applications of Bayesian networks in agriculture. *Engineering Applications of Artificial Intelligence*, 65, 29-42.
- et al., C. A. (2010). Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation. *Journal of Machine Learning Research*, 11, 171-234.
- et al., D. H. (1995). A Bayesian approach to learning causal networks. *Uncertain Reasoning*, 285-318.
- et al., I. T. (2003). Algorithms for Large Scale Markov Blanket Discovery. *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, 376-381.
- et al., N. Z. (2018). Survey on Bayesian Networks and their applications. *Computational Statistics*, 3(2), 379-398.
- et al., N. F. (1999). Data Analysis with Bayesian Networks: A Bootstrap Approach. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 196-205.
- et al., P. S. (2000). *Causation, Prediction, and Search*. MIT Press.
- et al., R. N. (2013). Bayesian Networks in R with Applications in Systems Biology. *Springer*.
- Gámez, J. A., Mateo, J. L., & Puerta, J. M. (2011). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Mining and Knowledge Discovery*, 22, 106-148.
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence [On Probabilistic models]. *Nature*, 521(7553), 452-459. <https://doi.org/10.1038/nature14541>
- Glover, F. (1989). Tabu search—part I. *ORSA Journal on computing*.
- Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10, 524.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

- Ianishi, P., Gonzatto Junior, O. A., Henriques, M. J., Nascimento, D. C. d., Mattar, G. K., Ramos, P. L., Ara, A., & Louzada, F. (2020). Probability on graphical structure: a knowledge-based agricultural case. *Annals of Data Science*, 1-19.
- Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. Springer.
- Ji, J., Zhang, H.-X., Hu, R.-B., & Liu, C.-N. (2011). A Tabu-search based Bayesian network structure learning algorithm. *Beijing Gongye Daxue Xuebao/Journal of Beijing University of Technology*, 37(8), 1274-1280.
- Kaminsky, Y., & Deeva, I. (2022). BigBraveBN: algorithm of structural learning for bayesian networks with a large number of nodes. *Procedia Computer Science*, 212, 191-200.
- Kim, J.-S., & Jun, C.-H. (2013). Ranking evaluation of institutions based on a Bayesian network having a latent variable. *Knowledge-Based Systems*, 50, 87-99. <https://doi.org/https://doi.org/10.1016/j.knosys.2013.05.010>
- Kitson, N. K., Constantinou, A. C., Guo, Z., Liu, Y., & Chobtham, K. (2023). A survey of Bayesian Network structure learning. *Artificial Intelligence Review*, 1-94.
- Kohavi, R., *et al.* (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 14(2), 1137-1145.
- Koivisto, M., & Sood, K. (2004). Exact Bayesian Structure Discovery in Bayesian Networks. *Journal of Machine Learning Research*, 5, 549-573.
- Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT press.
- Korb, K. B., & Nicholson, A. E. (2010). *Bayesian artificial intelligence*. CRC press.
- Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2), 157-194.
- Li, J., & Chen, J. (2014). A Hybrid Optimization Algorithm for Bayesian Network Structure Learning Based on Database. *J. Comput.*, 9(12), 2787-2791.
- Liu, F., Zhang, S.-W., Guo, W.-F., Wei, Z.-G., & Chen, L. (2016). Inference of gene regulatory network based on local Bayesian networks. *PLoS computational biology*, 12(8), e1005024.

- Marcot, B. G. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling*, 230, 50-62. <https://doi.org/10.1016/j.ecolmodel.2012.01.013>
- Margaritis, D. (2000). Bayesian network induction via local neighborhoods. *Advances in Neural Information Processing Systems*.
- Margaritis, D., et al. (2003). *Learning Bayesian network model structure from data* (Tesis doctoral). School of Computer Science, Carnegie Mellon University Pittsburgh, PA, USA.
- McLachlan, S., Dube, K., Hitman, G. A., Fenton, N. E., & Kyrimi, E. (2020). Bayesian networks in healthcare: Distribution by medical condition. *Artificial intelligence in medicine*, 107, 101912.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in R. *Springer*, 122, 125-127.
- Neapolitan, R. E., et al. (2004). *Learning bayesian networks* (Vol. 38). Pearson Prentice Hall Upper Saddle River.
- Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 199-203.
- O’Gorman, B., Babbush, R., Perdomo-Ortiz, A., Aspuru-Guzik, A., & Smelyanskiy, V. (2015). Bayesian network structure learning using quantum annealing. *The European Physical Journal Special Topics*, 224, 163-188.
- Okutan, A., Yang, S. J., & McConky, K. (2017). Predicting cyber attacks with bayesian networks using unconventional signals. *Proceedings of the 12th Annual Conference on Cyber and Information Security Research*, 1-4.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers.
- Pearl, J., & Verma, T. S. (1995). A theory of inferred causation. En *Studies in Logic and the Foundations of Mathematics* (pp. 789-811). Elsevier.
- Pifer, A., & Guedes, L. (2007). Aprendizagem estrutural de redes bayesianas utilizando métrica MDL modificada. *IEEE Latin America Transactions*, 5(8), 644-651.

- Quispe, A. J., Cambraia Leite, L., Pedro Pires Medeiros, A., Barros, R., & Sauer, I. (2013). Bayesian Networks applied to Failure Diagnosis in Power Transformer. *IEEE Latin America Transactions*, 11(4), 1075-1082. <https://doi.org/10.1109/TLA.2013.6601752>
- Reichenbach, H. (1991). *The direction of time* (Vol. 65). Univ of California Press.
- Risteski, A. (2022). *10708 Spring 2022 : Probablistic Graphical Models - GitHub Pages* [Accessed: 2023-10-23]. <https://andrejristeski.github.io/10708-Spring2022/>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,
- Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(6), 1-22.
- Scutari, M. (2021). Bnlearn—An R Package for Bayesian Network Learning and Inference. 2021. URL <https://www.bnlearn.com/bnrepository>, 105.
- Scutari, M., Graafland, C. E., & Gutiérrez, J. M. (2018). Who Learns Better Bayesian Network Structures: Accuracy and Speed of Structure Learning Algorithms. *arXiv preprint arXiv:1805.11908*.
- Sperotto, A., Molina, J., Torresan, S., Critto, A., Pulido-Velazquez, M., & Marcomini, A. (2019). A Bayesian Networks approach for the assessment of climate change impacts on nutrients loading. *Environmental Science & Policy*, 100, 21-36.
- Sun, X., Chen, C., Wang, L., Kang, H., Shen, Y., & Chen, Q. (2019). Hybrid optimization algorithm for Bayesian network structure learning. *Information*, 10(10), 294.
- Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). Max-min hill-climbing Bayesian network structure learning based on mutual information. *Journal of Machine Learning Research*.
- Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling*, 203(3-4), 312-318.
- Yin, B., Li, B., Liu, G., Wang, Z., & Sun, B. (2021). Quantitative risk analysis of offshore well blowout using bayesian network. *Safety science*, 135, 105080.
- Zagorecki, A., Orzechowski, P., & Hołownia, K. (2013). Online diagnostic system based on Bayesian networks. *Artificial Intelligence in Medicine: 14th Conference on Artificial Intelligence in Medicine, AIME 2013, Murcia, Spain, May 29–June 1, 2013. Proceedings 14*, 145-149.



- Zhang, X., Zhao, X.-M., He, K., Lu, L., Cao, Y., Liu, J., Hao, J.-K., Liu, Z.-P., & Chen, L. (2012). Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics*, 28(1), 98-104.
- Zou, X., Yue, W. L., *et al.* (2017). A bayesian network approach to causation analysis of road accidents using netica. *Journal of advanced transportation*, 2017.



## ANEXOS

## Anexo 1. Dataset ASIA

En la tabla se muestra los 30 primeros datos del dataset ASIA.

Smoker	LungCancer	VisitToAsia	Tuberculosis	TuberculosisOrCancer	X-ray	Bronchitis	Dyspnea
yes	yes	no	no	yes	yes	yes	yes
no	no	no	no	no	yes	yes	yes
no	no	no	no	no	no	yes	yes
no	no	no	no	no	no	yes	yes
yes	no	no	no	no	no	yes	no
yes	no	no	no	no	no	yes	yes
no	no	no	no	no	no	no	no
yes	no	no	no	no	no	yes	yes
no	no	no	no	no	no	no	yes
no	no	no	no	no	no	yes	yes
yes	no	no	no	no	no	yes	yes
yes	no	no	no	no	no	no	no
yes	no	no	no	no	no	yes	yes
no	no	no	no	no	no	no	no
yes	no	no	no	no	yes	yes	yes
yes	no	no	no	no	no	yes	yes
yes	no	no	no	no	no	yes	yes
no	no	no	no	no	no	yes	yes
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
no	no	no	no	no	no	no	no
yes	no	no	no	no	no	yes	yes
yes	no	no	no	no	no	no	no
no	no	no	no	no	no	yes	no
no	no	no	no	no	no	no	no
yes	no	no	no	no	no	yes	yes
...	...	...	...	...	...	...	...

## Anexo 2. codificación ejecutada en R para el aprendizaje estructural

```
#Instalación y carga de bibliotecas
install.packages("bnlearn")
library(bnlearn)

data <- read.csv("ruta/a/tu/archivo.csv")

summary(data)

# Establecer una red inicial (opcional)
initial.bn <- model2network("[Smoker][LungCancer|Smoker][VisitToAsia]

# Ajustar parámetros del algoritmo Hill-Climbing
bn.hc <- hc(data,
            start = initial.bn,
            score = "bic",
            maxp = 5,
            iss = 10,
            restart = 5,
            perturb = 3)

graphviz.plot(bn.hc)

fitted.bn <- bn.fit(bn.hc, data=data)

result_single <- cpquery(fitted.bn, event = (LungCancer == "yes"), ev
evidence_multiple <- data.frame(Smoker = c("yes", "no"), Tuberculosis
result_multiple <- cpquery(fitted.bn, event = (LungCancer == "yes"),

cv.results <- bn.cv(data, algorithm = "hc", k = 5)

k <- 10
```



```
# Tamaño de cada pliegue
fold_size <- floor(nrow(asia) / k)

# Inicializar una lista para guardar las puntuaciones BIC
bic_scores <- numeric(k)

# Bucle para la validación cruzada de k-pliegues
for(i in 1:k) {

  test_indices <- ((i - 1) * fold_size + 1):(i * fold_size)

  # Crear conjuntos de entrenamiento y prueba
  test_set <- asia[test_indices, ]
  train_set <- asia[-test_indices, ]

  gs_model_train <- gs(train_set)

  bic_scores[i] <- score(gs_model_train, data = test_set, type = "bic")
}

mean_bic <- mean(bic_scores)

# Mostrar el BIC medio
print(paste("BIC medio a través de todos los pliegues: ", mean_bic))
print(paste("Resultado de inferencia única: ", result_single))
print("Resultados de inferencia múltiple: ")
print(result_multiple)
print("Resultados de validación cruzada: ")
print(cv.results)
```

### Anexo 3. Generación de documento con Overleaf

Captura de pantalla de la aplicación Overleaf, que permitió la redacción del presente trabajo de investigación

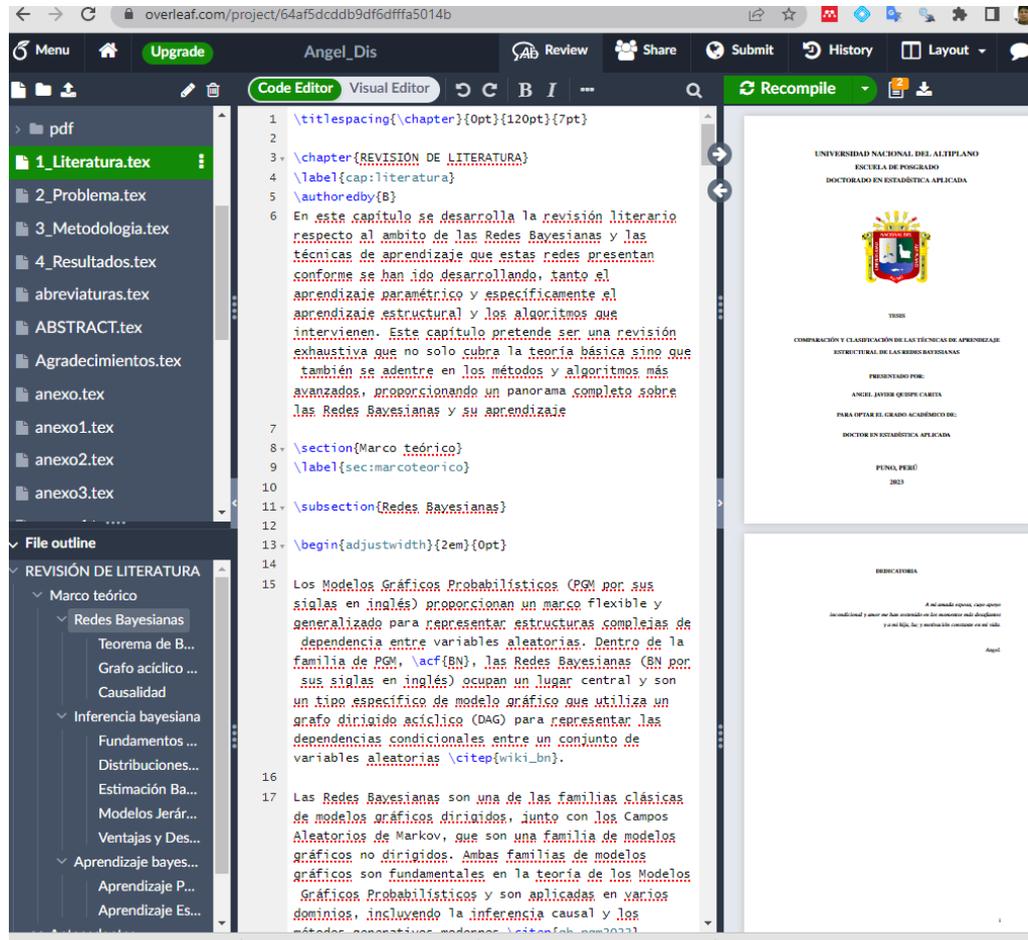


Figura 16. Aplicación Overleaf para la redacción del documento científico

#### Anexo 4. Conjunto de datos BIC para la realización de comparación de medias

Tabla 13

*Valores BIC de k-folds para diferentes técnicas de aprendizaje (tratamientos)*

Tratamiento	BIC	Tratamiento	BIC
GS	-1307.994	TABU	-1185.44
GS	-1252.374	TABU	-1134.686
GS	-1260.371	TABU	-1150.08
GS	-1318.351	TABU	-1232.199
GS	-1243.711	TABU	-1152.098
GS	-1258.163	TABU	-1155.834
GS	-1302.29	TABU	-1179.364
GS	-1294.339	TABU	-1135.235
GS	-1345.857	TABU	-1192.337
GS	-1269.32	TABU	-1165.495
IAMBB	-1307.994	MMHC	-1290.542
IAMBB	-1252.374	MMHC	-1233.615
IAMBB	-1260.371	MMHC	-1239.614
IAMBB	-1318.351	MMHC	-1302.994
IAMBB	-1243.711	MMHC	-1235.312
IAMBB	-1258.163	MMHC	-1244.201
IAMBB	-1302.29	MMHC	-1275.719
IAMBB	-1294.339	MMHC	-1254.749
IAMBB	-1345.857	MMHC	-1312.459
IAMBB	-1269.32	MMHC	-1255.987
HC	-1185.44	RSAMAX2	-1290.542
HC	-1134.686	RSAMAX2	-1233.615
HC	-1150.08	RSAMAX2	-1239.614
HC	-1232.199	RSAMAX2	-1302.994
HC	-1152.098	RSAMAX2	-1235.312
HC	-1155.834	RSAMAX2	-1244.201
HC	-1179.364	RSAMAX2	-1275.749
HC	-1135.235	RSAMAX2	-1254.749
HC	-1192.337	RSAMAX2	-1312.459
HC	-1165.495	RSAMAX2	-1255.987



Universidad Nacional  
del Altiplano Puno



Vicerrectorado  
de Investigación



Repositorio  
Institucional

## DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo ANGEL JAVIER QUISPE CORITA,  
identificado con DNI 42266179 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado

DOCTORADO EN ESTADISTICA APLICADA

informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:

“  
COMPARACION Y CLASIFICACION DE LOS TECNICOS DE APRENDIZAJE  
ESTRUCTURAL DE LAS REDES BAYESIANAS  
”

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 28 de DICIEMBRE del 20 23

FIRMA (obligatoria)



Huella



Universidad Nacional  
del Altiplano Puno



Vicerrectorado  
de Investigación



Repositorio  
Institucional

## AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo ANGEL JAVIER QUISPE CALTA,  
identificado con DNI 42266179 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado

DOCTORADO EN ESTADÍSTICA APLICADA  
informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:

“ COMPARACION Y CLASIFICACION DE LOS TECNICAS DE APRENDIZAJE  
ESTRUCTURAL DE LAS REDES BAYESIANAS ”

para la obtención de  Grado,  Título Profesional o  Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexas, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 28 de DICIEMBRE del 20 23

FIRMA (obligatoria)



Huella