



# Instituto Tecnológico y de Estudios Superiores de Monterrey

Análisis de métodos de razonamiento e incertidumbre  
MA2014.101

## PBL2. Muertes por diagnósticos médicos erróneos

Karla Andrea Palma Villanueva / A01754270  
Daniela Márquez Campos / A00833345  
Julio Eugenio Guevara Galván / A01704733  
Adrian Pineda Sánchez / A00834710  
David Fernando Armendáriz Torres / A01570813  
Kevin Antonio González Díaz / A01338316

Docente: Daniel Otero Fadul

Monterrey, Nuevo León, México. 28 agosto 2023

# Índice

<b>1. Problematización</b>	<b>2</b>
<b>2. Enfoque</b>	<b>2</b>
<b>3. Propósito</b>	<b>3</b>
<b>4. Información</b>	<b>3</b>
4.1. Redes Bayesianas . . . . .	3
4.2. Diabetes . . . . .	4
<b>5. Razonamiento</b>	<b>4</b>
<b>6. Conclusiones</b>	<b>6</b>
<b>Referencias</b>	<b>7</b>

# 1. Problematización

La transformación cultural en relación con el manejo de los errores en el ámbito médico representa uno de los desafíos más significativos no solo para la medicina, sino también para otras disciplinas implicadas en la atención de la salud. Aunque este proceso de cambio se inició hace más de dos décadas, su implementación sigue siendo limitada, lo que nos plantea un reto considerable en el horizonte. No obstante, se han logrado avances notables, aunque gradualmente, en cuanto a la seguridad del paciente, lo que ha contribuido a la reducción de errores y la minimización de riesgos en la prestación de atención médica [1].

A través de una investigación que examinó un gran número de autopsias realizadas entre 1966 y 2002. Se determinó que, en promedio, alrededor del 23,5 % de los casos presentaban errores en los diagnósticos médicos como causa de muerte [2].

Un análisis de datos de tasas de mortalidad médica realizado por expertos en seguridad del paciente de la Universidad Johns Hopkins revela que más de 250,000 muertes al año en Estados Unidos son resultado de errores médicos [3]. Esta cifra supera a la tercera causa de muerte en el país, la enfermedad respiratoria.

Estos estudios resaltan la necesidad de una evaluación más profunda y de medidas efectivas para mejorar la precisión diagnóstica y, por consiguiente, elevar la calidad de la atención médica.

# 2. Enfoque

La aplicación del concepto de redes bayesianas a esta problemática permite el uso de la información que se tiene (lo cual se detallará a continuación) para el cálculo de la probabilidad de diabetes, ajustándose a las circunstancias específicas de cada paciente. Esta capacidad es inherente a la estadística bayesiana, que describe la actualización de probabilidades dada la aparición de nueva evidencia, que en este caso es constituida por los valores que toman las variables.

Estas variables vienen de un dataset llamado "diabetes-dataset", el cual contiene 2,000 registros, cada uno de los cuales corresponde a una persona de quien se obtuvo información -numérica- como cantidad de embarazos, nivel de glucosa, presión arterial, grosor de la piel, insulina, índice de masa corporal (BMI, por sus siglas en inglés), edad, y finalmente, si padecen diabetes o no, lo cual se utiliza para etiquetar los datos ya que esta es la variable que se busca predecir con base en las demás.

Posterior a un tratamiento de los datos, y otros procesos que incluyen pero no se limitan a una estructuración de las variables basada en sus relaciones, es posible implementar un modelo que utilice datos usualmente disponible para revelar información importante: en este caso, la presencia de una enfermedad de gravedad como lo es la diabetes.

### 3. Propósito

La finalidad de crear este programa es obtener las distribuciones de probabilidad de las variables dadas a través de Redes Bayesianas, para después aplicar algoritmos de inferencia que ayuden a realizar deducciones de utilidad para realizar diagnósticos centrados a dos posibilidades: tener o no tener diabetes, siendo posible identificar las secuencias de probabilidad conjuntas de acuerdo a las variables, en este caso, diversos parámetros médicos que se relacionan con esta enfermedad en particular. La aplicación eficaz de este método puede apoyar a continuar minimizando los errores y el riesgo que como se mencionó comprometen la seguridad del paciente. Aunado a ello, esta minimización de incertidumbre propiciaría la realización de diagnósticos médicos más precisos a través de evaluaciones más complejas y por consiguiente, elevaría la calidad en el servicio de prestación de atención médica. Finalmente, la mejoría en la precisión de los diagnósticos médicos garantiza la mejoría en la esperanza de vida de la comunidad.

### 4. Información

#### 4.1. Redes Bayesianas

Las redes bayesianas son instrumentos estadísticos que han surgido en el ámbito de la Inteligencia Artificial y que posibilitan abordar situaciones de investigación que presentan estas cualidades. [4] Una red bayesiana se configura como un diseño gráfico que posibilita la representación de las interconexiones entre un conjunto de variables mediante nodos. A través de esta representación, es factible proporcionar datos de relevancia sobre la naturaleza de las conexiones entre las variables del contexto, las cuales en ocasiones pueden ser comprendidas como vínculos de causa y efecto, padre e hijo. [4]

Por consiguiente, se establece que la distribución de probabilidad perteneciente a cualquier nodo  $x_i$  que tenga padres será condicional de la forma  $P(x_i|padres(x_i))$ . Mientras que si dicho nodo  $x_i$  no tiene padres, se le asociará una probabilidad marginal, tal que  $P(x_i|padres(x_i)) = P(x_i)$ . [8]

La configuración de la red posibilita la definición de la probabilidad conjunta de estas variables como el resultado de multiplicar funciones de probabilidad condicional, que suelen ser de naturaleza más simple, lo cual se puede representar con la siguiente ecuación:

$$P(x_1, x_2, \dots, x_n) = P(x_1) \prod_{i=2}^n P(x_i|x_{i-1}, \dots, x_1) \quad (1)$$

Adaptando este concepto a redes bayesianas, la probabilidad conjunta se define por:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i|padres(x_i)) \quad (2)$$

Este método se presenta como una efectiva táctica para abordar cuestiones ligadas a la incertidumbre, en situaciones donde las conclusiones no pueden formarse únicamente basándose en el conocimiento previo del problema. [4]

Una red bayesiana es conocida por ser una estructura que permite crear un modelo de

probabilidad a través de pruebas y representaciones de dependencias entre las variables aleatorias. Su función básicamente puede servir para la predicción en diferentes situaciones.

## 4.2. Diabetes

La diabetes es una enfermedad metabólica crónica caracterizada por niveles elevados de glucosa en sangre, cuyas consecuencias son problemas en el corazón, vasos sanguíneos, ojos riñones y nervios, puesto que eleva el azúcar en la sangre a un nivel más alto de lo normal [5] Se genera cuando el organismo no utiliza de manera eficaz la insulina, hormona que regula la glucosa en la sangre.

Es una de las principales causas de mortalidad por el alto riesgo que genera de padecer enfermedades cardiovasculares y tuberculosis. En México, de acuerdo con los datos de mortalidad para 2020, se reportaron 1 086 743 fallecimientos , de los cuales 14 % (151 019) correspondieron a defunciones por diabetes mellitus. De estas, 52 % (78 922) ocurrieron en hombres y 48 % (72 094) en mujeres. Del total de fallecimientos 98 % (144 513) fueron por diabetes No insulino dependiente y de otro tipo y 2 % (3 506) por diabetes Insulino dependiente. [7]

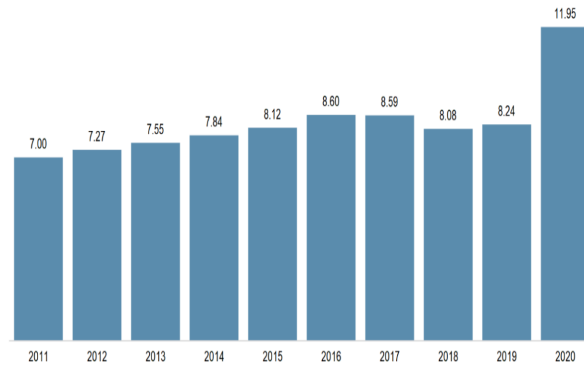


Figura 1: Tasa de mortalidad por cada 10 mil habitantes 2011-2020

## 5. Razonamiento

Primeramente, para el desarrollo del algoritmo planteado fue requerido llevar a cabo una fase de preparación de los datos, la cual consistió en generar variables compuestas y discretizar algunas otras. La variable compuesta generada 'Overweight' clasifica aquellos registros con un  $BMI \geq 25$  como 1 indicando sobrepeso, en caso de lo contrario se presenta un 0. Dicha variable fue generada para apegarse al diseño preestablecido de la red bayesiana. A su vez, el proceso de discretización se realizó con base en los cuartiles de cada variable con la excepción de 'Outcome' y 'Overweight' dado que estas dos son variables binarias. Valores menores al primer cuartil se clasificaron como 0, valores entre el 1er y 2do cuartil se clasificaron como 1, aquellos entre el 2do y 3er cuartil se clasifican como 3 y el resto como 4. La discretización se realizó para acatar las necesidades del paquete pybbn empleado que requiere de datos discretizados para poder generar una red bayesiana.

Una vez finalizada la preparación de los datos, como parte del proceso para la creación de un algoritmo de inferencia bayesiana, se diseñó la estructura de la red. Para esto se generaron dos diccionarios, 'graph' y 'values', cuyas llaves corresponden a cada elemento de la red bayesiana y sus valores, respectivamente son una lista de sus elementos padre y una lista de los posibles valores a tomar. Adicionalmente, se calcularon todas las probabilidades asociadas con algún nodo específico dado los diversos estados de sus nodos padre; para aquellos nodos hueranos, se calcula la probabilidad marginal, mientras que para los nodos con padre, se calcula la probabilidad condicional de todas las posibles combinaciones de sus padres.

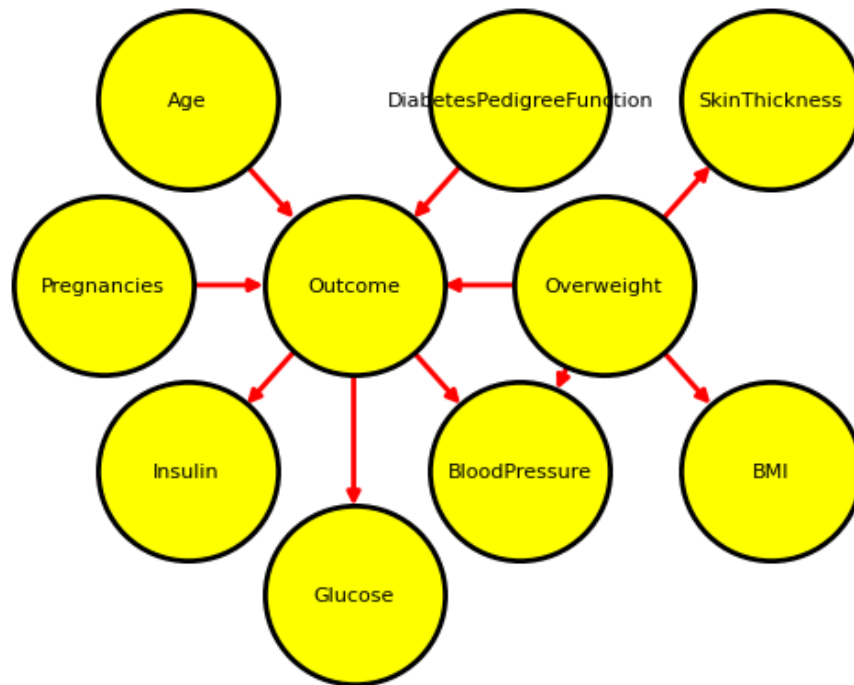


Figura 2: Red Bayesiana generada a partir de la base de datos 'diabetes-dataset.csv' [6]

Finalmente, al estructurar y generar la red bayesiana, Figura 2, se puede obtener la probabilidad de distribución conjunta utilizando lo previamente calculado e implementar nueva evidencia para generar inferencias con nuevas distribuciones. En el contexto del problema, la evidencia que se decidió agregar fue la existencia de sobrepeso con un valor de 1 y la edad con un valor correspondiente al tercer cuartil (3), por lo que el sistema calculó la distribución de probabilidad acumulada de si se padece diabetes, dado que se tiene sobrepeso y una edad correspondiente al 3er cuartil, es decir, con un valor de 40 años.

## 6. Conclusiones

La implementación de Redes Bayesianas en el ámbito abordado por el presente documento presenta varias ventajas notables. Una de estas ventajas radica en la habilidad de calcular las probabilidades asociadas a eventos específicos después de haber observado la ocurrencia de uno o varios eventos que pueden tener una relación causal con dichos sucesos. Esta característica es especialmente útil para evaluar cómo la ocurrencia de ciertos eventos afecta las probabilidades de otros eventos en el sistema.

Las funciones desarrolladas en el marco de esta implementación permiten la actualización dinámica de información a medida que se observan nuevos eventos. Esto implica que las probabilidades calculadas pueden ajustarse según las observaciones, lo que refleja la naturaleza dinámica de los datos y permite mantener el modelo actualizado. Además, estas funciones también permiten retroceder al modelo original, lo que brinda flexibilidad y control sobre la manipulación de la información.

Otra ventaja clave es la representación visual de las redes bayesianas mediante nodos y conexiones. Esta representación gráfica facilita enormemente la comprensión de las relaciones causales entre los eventos. La visualización de los nodos permite identificar las conexiones de manera más clara.

Aunque el algoritmo desarrollado no puede reemplazar la experiencia y el juicio clínico de un médico especializado en el diagnóstico de diabetes, sí puede desempeñar un papel importante. Proporciona una herramienta que, basándose en las probabilidades calculadas, puede evaluar la probabilidad de que un paciente padezca la enfermedad. Esta información no solo puede ayudar al profesional médico en la toma de decisiones, sino que también brinda una base adicional para respaldar la validez del diagnóstico médico.

## Referencias

- [1] Ceriani Cernadas, José M. (2015). Errores de diagnóstico en la práctica médica. Archivos argentinos de pediatría, 113(3), 194-195.
- [2] Shojania, K. G., Burton, E. C., McDonald, K. M., Goldman, L. (2003). Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. Jama, 289(21), 2849-2856.
- [3] Study suggests medical errors now third leading cause of death in the U.S. - 05/03/2016. (s/f). Hopkinsmedicine.org. Recuperado el 28 de agosto de 2023, de [https://www.hopkinsmedicine.org/news/media/releases/study\\_suggests\\_medical\\_errors\\_now\\_third\\_leading\\_cause\\_of\\_death\\_in\\_the\\_us](https://www.hopkinsmedicine.org/news/media/releases/study_suggests_medical_errors_now_third_leading_cause_of_death_in_the_us)
- [4] Santiesteban, C., Dianet Utria Pérez, Enrique, C., Carlos. (2012, June 21). Definición de Redes Bayesianas y sus aplicaciones. Revista Vinculando. <https://vinculando.org/articulos/redes-bayesianas.html>
- [5] Diabetes. (s. f.). OPS/OMS — Organización Panamericana de la Salud. <https://www.paho.org/es/temas/diabetes#:~:text=La%20diabetes%20es%20una%20enfermedad,los%20ri%C3%B1ones%20y%20los%20nervios>.
- [6] Ukani, V. (2020). Diabetes Data Set [Data set]. <https://www.kaggle.com/vikasukani/diabetes-data-set>
- [7] INEGI. (2021). [https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP\\_Diabetes2021.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP_Diabetes2021.pdf)
- [8] Otero Fadul, D. (2023). Redes Bayesianas. Análisis de Métodos de Razonamiento e Incertidumbre. Tecnológico de Monterrey.