

**Análisis de métodos de razonamiento e incertidumbre** (Grupo. 301)

Selección oportuna de alimentos en la dieta para afrontar la diabetes mediante Redes de Decisión

**Docente:**

Daniel Otero Fadul

**Alumnos:**

Ricardo Kaleb Flores Alfonso | A01198716

Iván Alejandro López Valenzuela | A01284875

David Alejandro Acuña Orozco | A00571187

Raúl Correa Ocañas | A01722401

Carol Jatziry Rendon Guerrero | A01425341

# **1. Problematización**

La diabetes es una enfermedad crónica caracterizada por mantener elevados los niveles de azúcar en sangre y afecta al menos a 112 millones de adultos en la región de las Américas. Según la Organización Panamericana de la Salud, el punto de partida para vivir con una calidad aceptable teniendo diabetes es un diagnóstico temprano. Sin embargo, la identificación en individuos sigue siendo un desafío, por todas las variables que se ven envueltas como hábitos de ejercicio, alimentación, predisposición genética y antecedentes médicos.

Normalmente los problemas asociados con la diabetes y la salud presentan soluciones que suelen ser costosas, complejas o poco accesibles. Es por ello que existe la necesidad de crear modelos predictivos y probabilísticos que favorezcan la toma de decisiones y que aparte de ser capaces de integrar y relacionar todas las múltiples variables y factores de riesgo de la diabetes también funcionen como herramientas intuitivas e interpretables. Un modelo de este tipo no solo puede mejorar la calidad de vida de los pacientes al permitir la detección temprana sino que también reduce los costos de instituciones y sistemas de salud. Asimismo, su implementación puede ayudar a identificar patrones en poblaciones específicas y diseñar estrategias y recomendaciones pertinentes para reducir la incidencia de la enfermedad.

# **2. Enfoque**

Para la implementación se utiliza un enfoque metodológico de red de decisión basado en redes bayesianas, por ser un modelo que se puede representar tanto probabilísticamente como gráficamente sus relaciones de dependencia entre variables y seleccionar la acción más conveniente conectando estos factores. Estas redes son especialmente útiles para problemas de predicción y diagnóstico como el que se presenta, ya que permiten realizar inferencias basadas en datos incompletos o inciertos. El enfoque del proyecto se resume en 5 puntos fundamentales:

1. Preprocesamiento de datos: se utiliza un dataset con factores de riesgo de diabetes, como índice de masa corporal, presión arterial, nivel de glucosa, entre otras, las cuales deben ser discretizadas para facilitar la construcción de la red bayesiana.
2. Estructuración de la red: definición de un grafo causal en el que los nodos representan variables (por ejemplo, BMI y glucosa) y los arcos representan relaciones causales.
3. Cálculo de probabilidades: estimación de distribuciones de probabilidad condicional para cada nodo, considerando los datos disponibles.
4. Validación: el modelo se evalúa mediante experimentos utilizando evidencias para comprobar su capacidad de calcular probabilidades y generar predicciones precisas dados ciertos hechos ya determinados.
5. Análisis de utilidad: se implementan nodos de decisión y utilidad para complementar la red de decisión y así evaluar diferentes acciones preventivas que determinen la estrategia más adecuada según el estado del paciente.

**3. Propósito**

El propósito de este proyecto fue desarrollar un modelo de red de decisión basado en redes bayesianas que permita primeramente predecir la probabilidad de riesgo de padecer diabetes con el objetivo de contribuir al manejo preventivo y de tratamiento de la misma mediante la toma de decisiones bajo el criterio del máximo beneficio que se vea influenciado por la probabilidad ya mencionada. Este propósito incluye no solo la implementación de un modelo adaptable y capaz, sino también la validación de este mediante experimentos controlados, garantizando resultados precisos y que mediante la interpretación proporcione información valiosa y confiable.

**4. Información**

La información utilizada en principio para obtener mediante una red bayesiana ya estructurada las distribuciones de probabilidad de los posibles diagnósticos que sirvan como nodos aleatorios y facilitador de las funciones principales que tiene el proyecto proviene del conjunto de datos “diabetes-dataset” que se encuentra disponible en la plataforma Kaggle.

Este contiene los registros recabados de 2000 pacientes en cuanto a mediciones de distintos factores de riesgo relacionados con la diabetes y finalmente el diagnóstico definitivo de la enfermedad. De todos los pacientes, se encuentran 684 que padecen diabetes mientras que este no es el caso para 1316 de ellos. Además, se discretizan el resto de variables no binarias de manera que se vuelvan columnas categóricas para adaptar mejor su manejo en la construcción de evidencias para los experimentos. Se puede acceder al conjunto de datos original en [este enlace](https://www.kaggle.com/datasets/vikasukani/diabetes-data-set).

En cuanto a la información referente a los nodos de decisión y de utilidad que se construyen para completar la red de decisión, las acciones entre las que se debe decidir en el primero de estos incluyen la recomendación de alimentos bajos, medios o altos en índice glucémico e influye en esta decisión el beneficio que se genera con la recomendación dependiendo del estado de diagnóstico de diabetes del paciente, convirtiéndolos a estos pares en el nodo de utilidad y resultando en que el respectivo desglose de estas se vea representado de la siguiente manera.

| **Diabetes** | **Food** | **Utilidad** |
| --- | --- | --- |
| 0 | Low GI | 60 |
| 0 | Medium GI | 100 |
| 0 | High GI | 80 |
| 1 | Low GI | 100 |
| 1 | Medium GI | 50 |
| 1 | High GI | 0 |

Tabla 1. Beneficios de la recomendación de dietas según el estado de diagnóstico del paciente

# **5. Razonamiento**

En el contexto del problema y el enfoque que se le está dando, la implementación en código consiste prácticamente en la integración de los componentes que conforman una red de decisión para seleccionar una acción, que en este caso sería la recomendación de la dieta adecuada basada en alimentos con cierto nivel de índice glucémico que se mide con un valor del 1 al 100, siendo los de nivel bajo los que cuentan con un GI menor a 55, los de nivel medio con GI entre 55 y 69 y los de nivel alto con GI mayor a 70.

En primer lugar, la inicialización de la red bayesiana con la que ya se estaba familiarizado y que modela la dependencia entre factores de riesgo relacionados con la diabetes y su reconfiguración dados posibles escenarios mediante la función *evidence()* constituye lo que son los nodos aleatorios, de los cuales el más relevante para este caso resulta el de Outcome que sería el diagnóstico.

En adición, se genera el nodo de decisión añadiendo una lista con las 3 acciones entre las que se decide y otra con los estados posibles de Outcome. Asimismo, para comenzar con la generación del nodo de utilidad se crea primero un diccionario en el que a cada par de estos posibles estados y acciones se le asigna un valor beneficio acorde a como se observa en la tabla de la sección anterior. De esta manera, la función *utility\_function()* interactúa con dicho diccionario según los parámetros que se le ingresen para retornar el beneficio que le corresponda.

Finalmente, *maximize\_utility()* se encarga de modelar la relación e interacción de toda la red de decisión en sus nodos aleatorios, de decisión y de utilidad al calcular y comparar los beneficios de las acciones para retornar aquella que represente la máxima utilidad dados los parámetros que se le ingresen, donde para esta función se trata de la distribución de probabilidad del nodo aleatorio. Es así como con el seguimiento de esta secuencia de pasos y funciones, se pueden ajustar escenarios donde las decisiones seleccionadas indiquen posibles tendencias en el tipo de alimentación recomendada para prevenir y afrontar la diabetes.

# **6. Conclusiones**

Luego de una serie de 7 experimentos con nuestra red de decisión en los que las evidencias ingresadas fueron arrojando distintas distribuciones de probabilidad que fueron variando en proporción tanto para diagnósticos negativos como positivos en el nodo aleatorio, se llegaron a conclusiones muy pertinentes mediante la observación de los resultados sobre las decisiones tomadas.

En primer lugar, se observa algo que en la práctica guarda un tanto de lógica y es que ninguno de los experimentos recomendó una dieta basada en alimentos altos en GI. Por más que estuviera cargada la probabilidad hacia un diagnóstico positivo o negativo, esta acción no parece ser viable primero desde la utilidades que esto generaría, lo cual se interpreta como una acción que puede resultar perjudicial ya sea cuando se está previniendo y peor aún combatiendo la diabetes.

Además, se concluye que mientras sea más probable que exista padecimiento de diabetes para el paciente, conviene más seguir una dieta basada en alimentos bajos en GI. En el caso contrario, se recomienda una con alimentos que cuenten con un nivel medio de este índice, con lo cual se procura en ambos situaciones que se regule el impacto que se tiene en el aumento de la glucosa en la sangre, que representa distinta peligrosidad según el caso, pero al mismo tiempo que se cumplan con los requerimientos alimenticios. Los resultados que permiten dar esta conclusión se piensa que fueron principalmente favorecidos por las utilidades, sobre todo las de 100, que se habían establecido para dichas acciones en los posibles estados de diagnóstico.

**Referencias**

OPS/OMS. (2024). Diabetes. *Organización Panamericana De La Salud.* Recuperado de: <https://www.paho.org/es/temas/diabetes>

S.A. (2022). Índice glucémico y diabetes. *MedlinePlus.* Recuperado de: <https://medlineplus.gov/spanish/ency/patientinstructions/000941.htm>

**A. Anexos**

Acceso al código ejecutado en [este enlace](https://drive.google.com/file/d/1kmWgB_1Jf9kpKBj6A_H1d6UozeUvjNDM/view?usp=sharing)