



Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey

Modelación del aprendizaje con inteligencia artificial

Jose Eduardo Ferrer Cruz

Aprendizaje no supervisado

Leonardo Laureles Olmedo A01659241

Mariana Rincón Flores A01654973

Carlos Mateos Pérez A01654085

Daniel Núñez López A01654137

Situación Problema: Identificar enfermedades y diagnóstico

Explicación del problema

El diagnóstico precoz engloba una serie de medidas para determinar la presencia de una enfermedad potencialmente grave. Las ventajas son muchas, disminuyen la tasa de mortalidad de estas enfermedades, la probabilidad de éxito de los tratamientos aumenta, las complicaciones se reducen, así como las secuelas de la enfermedad y del mismo tratamiento, de igual manera se podrían llegar a reducir los costos en caso de que la enfermedad se llegue a agravar.

Las pruebas de detección son exámenes que los doctores usan para detectar enfermedades y problemas de salud antes de que existan signos o causen síntomas. Estas pruebas sirven para detectar los problemas a tiempo, cuando pueden ser más fáciles de tratar. La prueba de detección dependerá del problema que se quiera detectar.

La actual pandemia SARS-CoV-2 plantea retos sanitarios, sociales y económicos de enorme magnitud. El diagnóstico correcto y rápido de la infección por SARS-CoV-2 es crítico, tanto desde el punto de vista epidemiológico, dado que muchas personas infectadas están asintomáticas, como desde el punto de vista clínico, para identificar y tratar a los pacientes cuanto antes.

Las técnicas de aprendizaje supervisado pueden ser de ayuda para predecir qué tan probable es que un paciente haya sido infectado con COVID-19 con base en sus comorbilidades habiendo estado en contacto con algún contacto positivo y con una gran cantidad de datos.

El objetivo del agente será recabar información sobre pacientes que fueron contagiados previamente, junto con la lista de comorbilidades que presente, para entrenarse y así utilizar un modelo de clasificación para dar la probabilidad de estar contagiado de COVID-19.

Descripción del Ambiente

La información se obtendrá de los sensores que se utilizan para detectar las enfermedades, para detectar la neumonía, se utiliza una radiografía de tórax; para la diabetes son los análisis de sangre que muestran si el nivel de glucosa en la sangre es demasiado alto; para detectar el asma se realizan espirometrías, las cuales son pruebas que miden cuánto aire se inhala y se exhala de los pulmones; el sensor que se utilizan para detectar la hipertensión es el esfigmomanómetro, que es el instrumento que se coloca en el brazo y se infla; para

detectar las enfermedades cardiovasculares, se utilizan electrocardiogramas como sensores; para detectar la obesidad se utiliza la báscula para medir la altura y el peso, de ahí se procede a realizar el cálculo del índice de masa corporal; para la insuficiencia renal crónica se realizan exámenes de sangre y orina; y finalmente, para detectar el tabaquismo se realiza un análisis de sangre y pruebas de nicotina.

Luego de esto se creará una base de datos y se realizará un cálculo de probabilidades y correlaciones entre las variables para determinar la forma de la red bayesiana para el uso del modelo.

El ambiente es observable ya que se trata de un programa que hará un test donde el agente tiene total conocimiento de las preguntas y respuestas. Es estratégico, ya que el agente busca algo en específico, únicamente basado en las variables de entrada y no las de salida, a diferencia del otro tipo de aprendizaje supervisado.

El ambiente es completamente accesible ya que los sensores detectan todos los aspectos relevantes a la elección de una acción, ya que si los sensores de un agente le permiten tener acceso al estado total del ambiente se dice que es accesible. También es considerado semi-dinámico ya que el ambiente no cambia con el paso del tiempo pero si va modificando la calificación asignada al desempeño del agente, el ambiente es discreto porque existe una cantidad limitada de percepciones y acciones y, finalmente es episódico porque la respuesta que se da a una pregunta no afectará a la siguiente, es decir cada vez que se pregunte algo la estrategia se va a replantear.

Observaciones a Recolectar

Las observaciones que serán importantes para el agente son el sexo de la persona, si ha padecido alguna enfermedad antes como neumonía, diabetes, asma, hipertensión, enfermedades cardiovasculares, obesidad, insuficiencia renal crónica, tabaquismo. A diferencia del proyecto con agente supervisado, no utilizaremos la edad ya que es el único valor que no podemos tener en binario y esto complicaría mucho el cálculo de probabilidades para la red bayesiana.

Se deben recolectar tantas observaciones como preguntas que haya en el cuestionario, Luego de esto, se hará el cálculo de probabilidades y se asignan a cada observación para el correcto funcionamiento de la red bayesiana.

Se deben recolectar tantas observaciones como preguntas que haya en el cuestionario, pero al momento que se realice cada pregunta el agente deberá recolectar una sola observación a la vez, para la creación de la base de datos.

Toma de Decisiones

El agente decidirá al final con base en las probabilidades los resultados obtenidos de la base de datos creada con las observaciones previas, hasta no tener toda la información no podrá dar una predicción.

No tendrá rangos de tiempo, será con base en los valores de probabilidad de cada observación de las variables predictoras, es decir si la persona tiene neumonía, enfermedades cardiovasculares, enfermedades renales, diabetes, asma, obesidad e hipertensión es más probable que sea infectado por COVID-19 que un adolescente sin ninguna de estas comorbilidades.

Una vez que se tengan las probabilidades para cada observación se dará una predicción sobre si la persona o no tenga COVID-19 en términos de probabilidades esto será con base en la cantidad de comorbilidades y sexo que tenga el paciente.

Acciones Disponibles

Una vez que se tome la decisión se dirá si la persona tiene más probabilidades que sea infectado al haber estado en contacto con algún contagiado, esto quiere decir que la acción será discreta, ya que será un sí o no únicamente como respuesta.

Representación Gráfica del Modelo

Para realizar las redes bayesianas utilizamos la aplicación basada en Java llamada Belief Networks de AI Space, en esta aplicación creamos los nodos de cada observación y luego formulamos una estructura con base en un árbol de decisión como en el primer proyecto, solamente que a diferencia de este, no consideramos como un nodo a la edad, así mismo, lo realizamos con una menor cantidad de registros para simular una red con una base de datos no tan extensa.

Una vez realizada la red, llenamos las probabilidades de cada nodo, intentando buscar en internet las probabilidades de cada una de ellas, sin embargo, no encontramos todas y las que faltaban las intentamos relacionar y deducir por nuestra cuenta.

Finalmente, comprobamos nuestra red con cinco queries, en las que definimos si la observación era verdadera o falsa y daba como resultado la probabilidad de salir positivo o negativo a una prueba de COVID-19.

Cabe mencionar que en sexo, consideramos el valor de positivo como hombre y el valor falso como mujer.

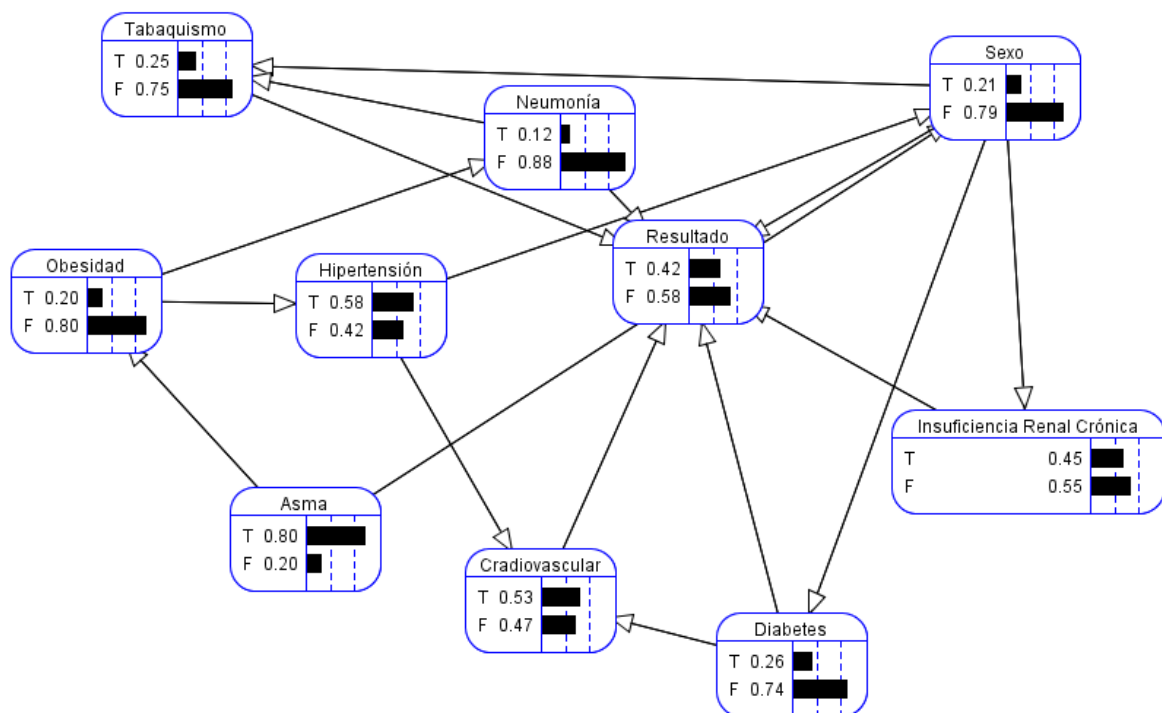


Figura 1. Probabilidades de salir positivo a COVID-19 sin saber si se tiene una comorbilidad.

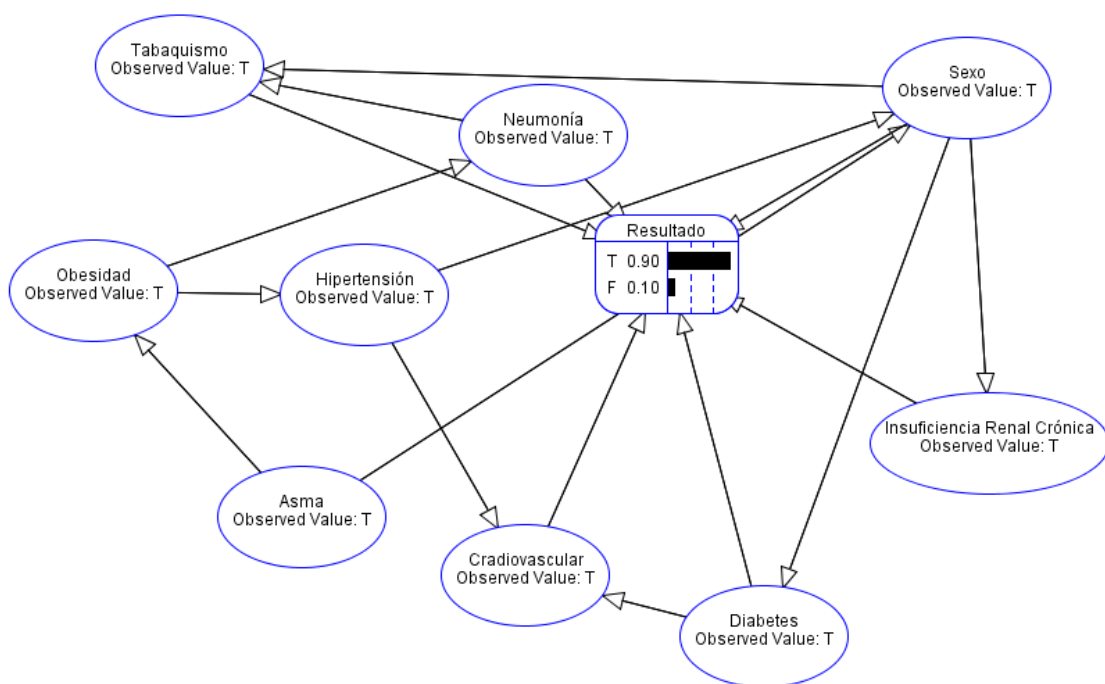


Figura 2. Probabilidad de salir positivo a COVID-19 siendo hombre y teniendo todas las comorbilidades.

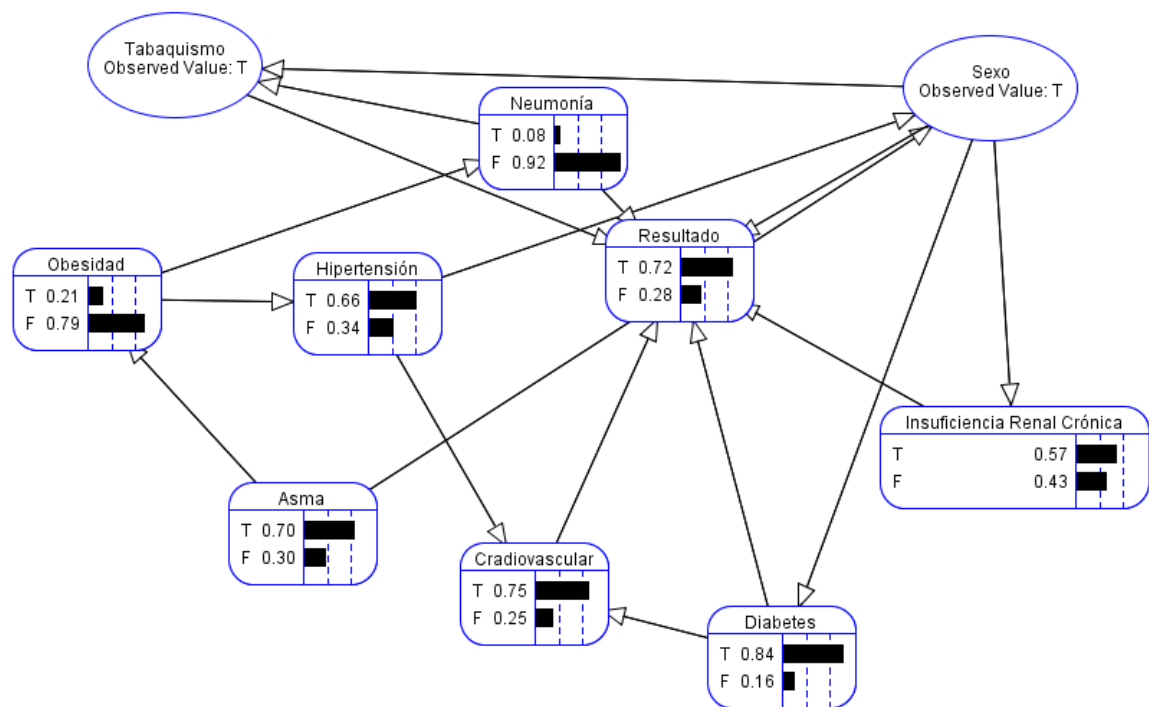


Figura 3. Probabilidad de salir positivo a COVID-19 teniendo tabaquismo y siendo hombre.

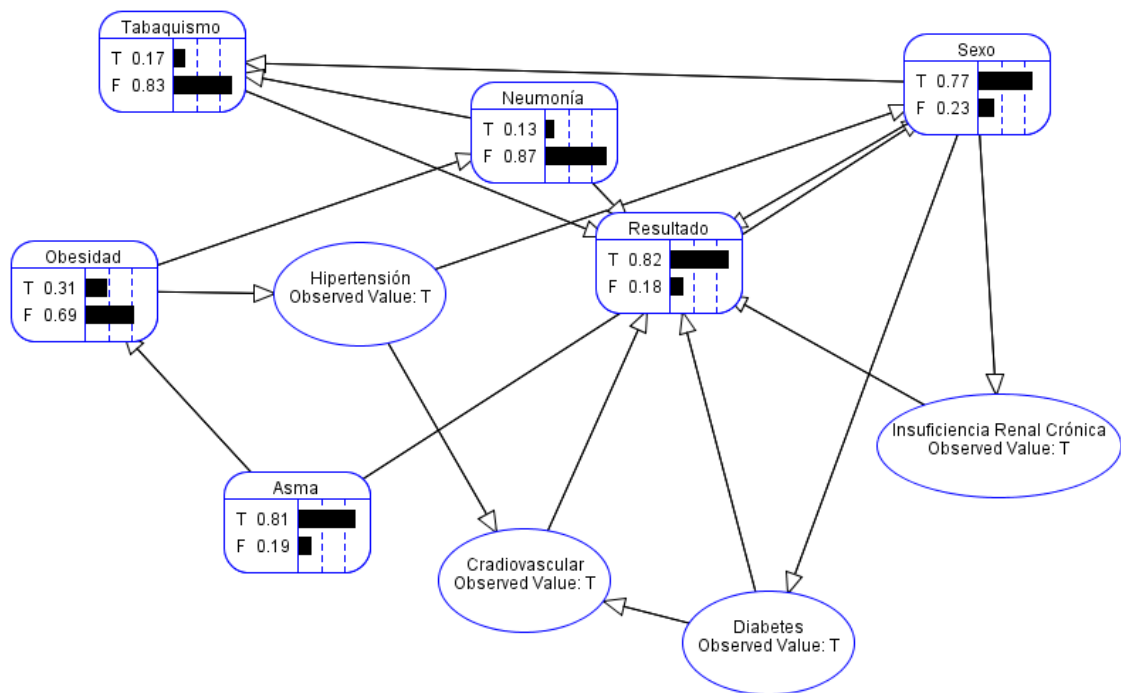


Figura 4. Probabilidad de salir positivo a COVID-19 teniendo problemas cardiovasculares, diabetes e insuficiencia renal siendo hombre.

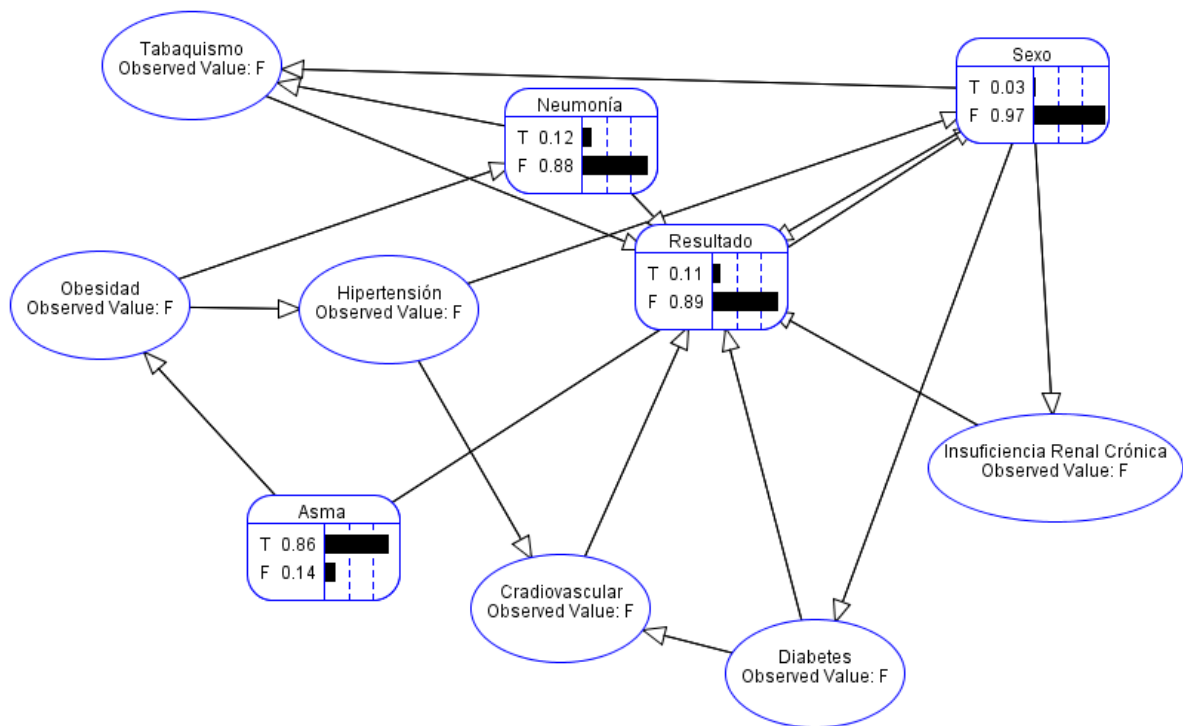


Figura 5. Probabilidad de tener COVID-19 siendo mujer y teniendo tabaquismo, obesidad, hipertensión, problemas cardiovasculares, diabetes e insuficiencia renal crónica.

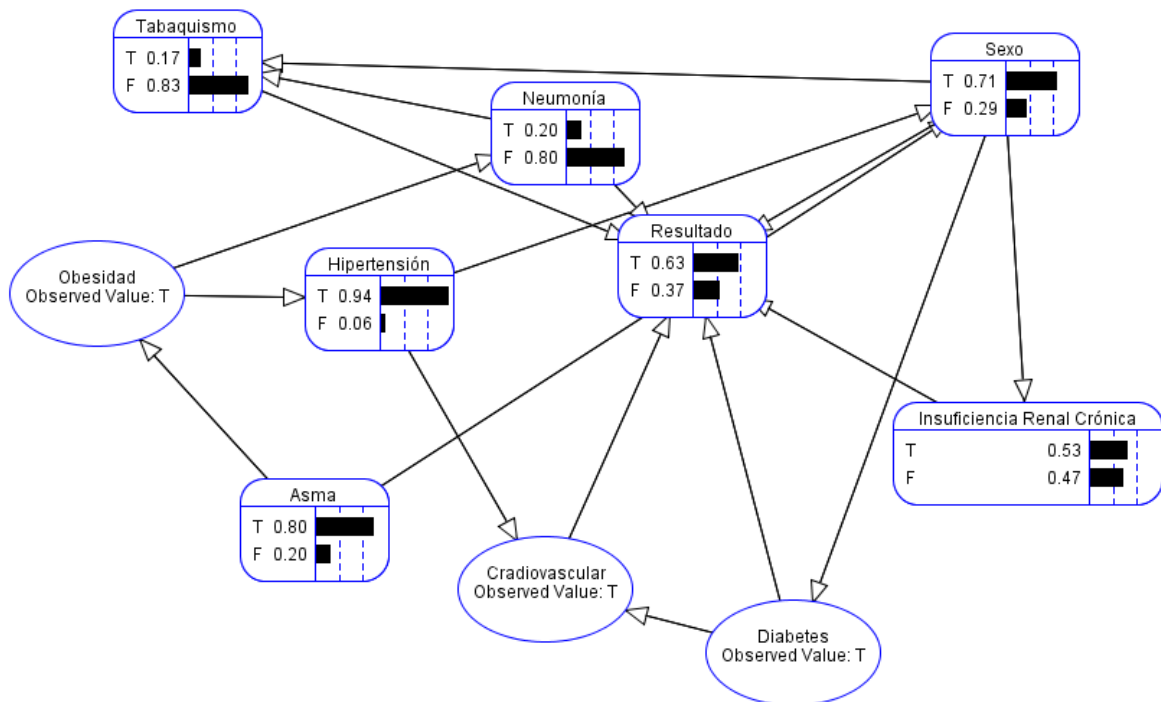


Figura 6. Probabilidad de tener COVID-19, siendo hombre y teniendo obesidad, problemas cardiovasculares, diabetes.

Referencias

Kaggle. (2020, 5 junio). *México COVID-19 clinical data*. Recuperado 14 de junio de 2022, de

<https://www.kaggle.com/datasets/marianarfranklin/mexico-covid19-clinical-data/meta-data>

NIH. (2021, 20 junio). *Características clínicas y comorbilidades asociadas a mortalidad en pacientes con COVID-19 en México*. National Library of Medicine. Recuperado 14 de junio de 2022, de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7816557/>