

DIAGNOSTICO AUTOMATIZADO DE DIABETES





Contenido:

- Introducción
- Descripción del problema
- Solución propuesta
- Innovación y diferenciación competitiva
- Viabilidad técnica
- Plan de ejecución
- Próximos pasos
- Contacto

INTRODUCCIÓN

• En la actualidad, la atención médica en Mexico enfrenta retos significativos en términos de eficiencia y calidad. La carga en los sistemas de salud y la necesidad de respuestas rápidas y precisas en el diagnóstico han generado un interés creciente en la automatización de este proceso.

La integración de tecnologías emergentes, como la inteligencia artificial y el análisis de Big data, ofrece oportunidades para revolucionar el ámbito médico, permitiendo diagnósticos más rápidos y precisos.

• En este contexto, la investigación se centra en explorar y desarrollar soluciones que automatizan el diagnóstico de Diabetes, con el objetivo de mejorar la eficiencia en la atención médica y, por ende, elevar la calidad de vida de los pacientes.



DESCRIPCIÓN DEL PROBLEMA:

- La cantidad de pacientes, complejidad y la diversidad de las enfermedades presentan desafíos significativos para los profesionales de la salud.
- El tiempo requerido para realizar diagnósticos precisos puede ser prolongado, y en ocasiones, la disponibilidad limitada de expertos médicos puede retrasar el inicio del tratamiento adecuado.
- Además, la carga en los sistemas de salud ha aumentado, evidenciando la necesidad de soluciones que optimicen los recursos y mejoren la rapidez en la identificación de enfermedades.
- La automatización del diagnóstico busca abordar estos desafíos, aprovechando el potencial de la inteligencia artificial para analizar grandes conjuntos de datos clínicos relevantes de manera eficiente y precisa.



ANTECEDENTES:



- Investigaciones previas han demostrado el éxito de la inteligencia artificial en diversas aplicaciones médicas, desde la interpretación de imágenes radiológicas hasta el análisis de patrones en datos genómicos. Métodos basados en aprendizaje profundo y algoritmos de machine learning han mostrado resultados prometedores en la identificación temprana de enfermedades, permitiendo un tratamiento más oportuno.
- Sin embargo, es esencial continuar desarrollando y perfeccionando estas técnicas para garantizar su aplicabilidad generalizada y su integración efectiva en entornos clínicos. Esta investigación busca ampliar la comprensión de las soluciones automatizadas de diagnóstico, evaluando su viabilidad, precisión y contribución a la mejora del sistema de salud.

METODOLOGIA:

Selección del Modelo:

- Se utilizará una regresión logística, que es adecuada para problemas de clasificación binaria como el diagnóstico de diabetes (presencia o ausencia de la enfermedad).

Variables Predictoras:

- Índice de Masa Corporal (IMC) y Nivel de glucosa en sangre.

Preprocesamiento de datos:

- Limpiar y preparar los datos para el análisis, manejando valores perdidos y normalizando las variables.

División de datos:

- Separar los datos en conjuntos de entrenamiento y prueba.

Entrenamiento del modelo:

- Ajustar el modelo de regresión logística utilizando el conjunto de entrenamiento.

Evaluación del modelo:

- Validar el modelo con el conjunto de prueba y evaluar su desempeño utilizando métricas como la precisión, sensibilidad, especificidad y la curva ROC.



SOLUCIÓN PROPUESTA

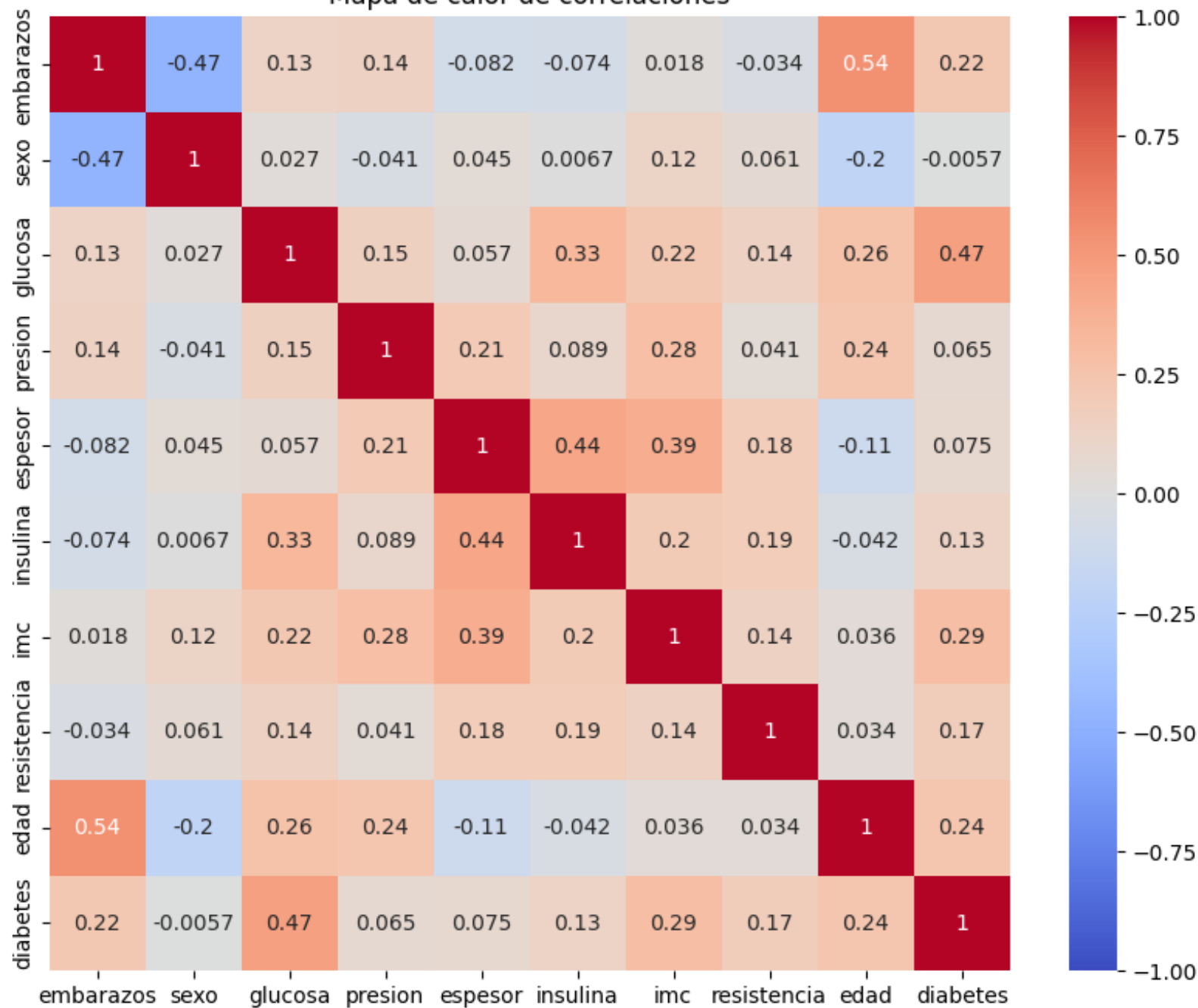
- La solución propuesta en el proyecto de investigación es la automatización del diagnóstico de diabetes mediante el uso de tecnologías emergentes como la inteligencia artificial y el análisis de Big Data. El objetivo es mejorar la eficiencia en la atención médica y elevar la calidad de vida de los pacientes al proporcionar diagnósticos más rápidos y precisos.
- Todo lo anterior con la creación de un algoritmo simple, el cual después de analizar la información proporcionada realice un pre diagnóstico, lo más acertado posible.
- Cabe resaltar que esta herramienta no busca eliminar la tarea de un médico humano, sino que ayudar a eficientar esta tarea para mejorar el nivel de vida de los pacientes.

SELECCIÓN DE DATOS

- La base de datos contiene la información de 3,000 pacientes, presentado de la siguiente manera:

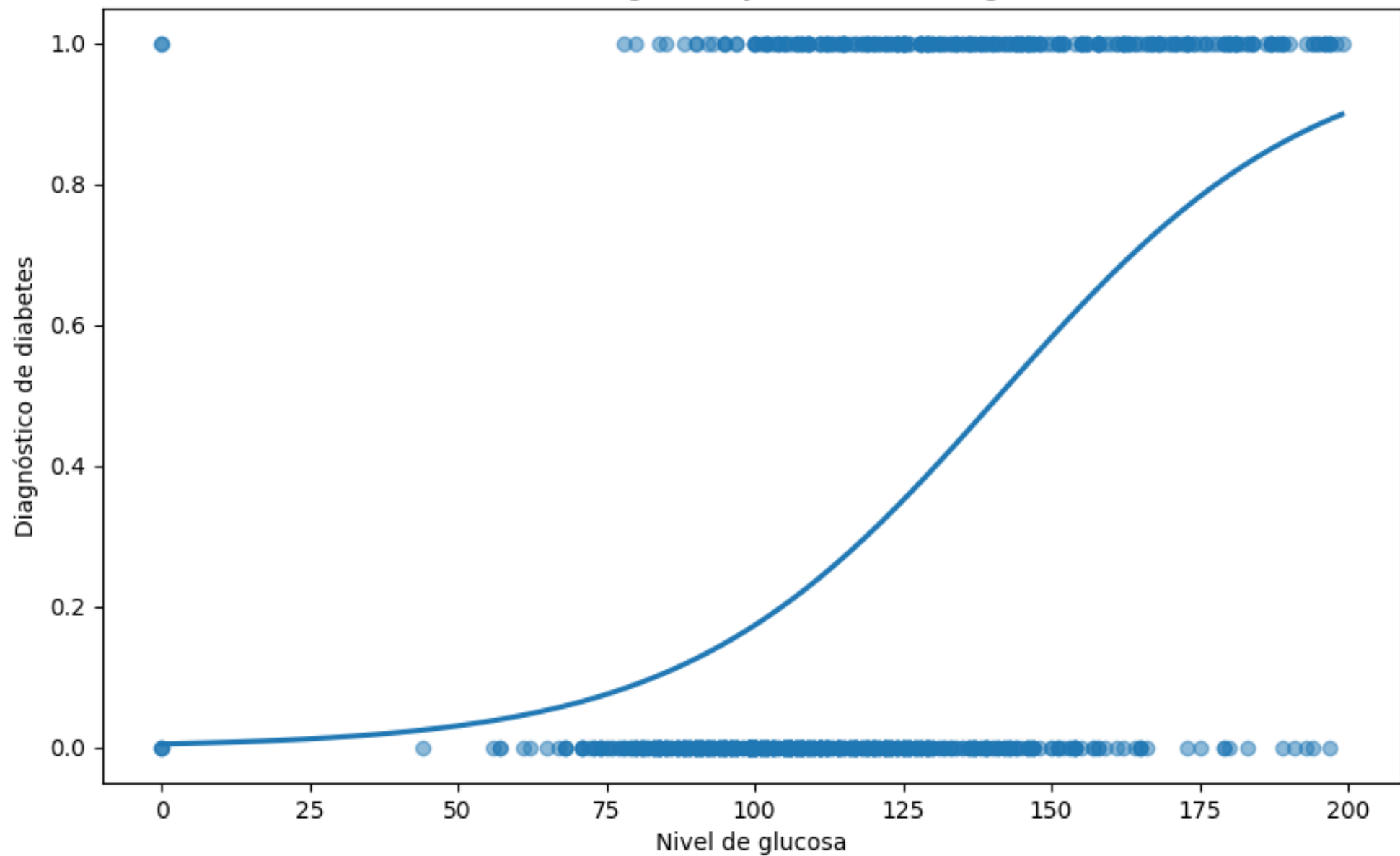
Index	embarazos	sexo	glucosa	presion	espesor	insulina	imc	resistencia	edad	diabetes
1	6	0	148	72	35	0	33.6	0.627	50	1
2	1	0	85	66	29	0	26.6	0.351	31	0
3	8	0	183	64	0	0	23.3	0.672	32	1
4	1	0	89	66	23	94	28.1	0.167	21	0
5	0	1	137	40	35	168	43.1	2.288	33	1
6	5	0	116	74	0	0	25.6	0.201	30	0
7	3	0	78	50	32	88	31	0.248	26	1
8	10	0	115	0	0	0	35.3	0.134	29	0
9	2	0	197	70	45	543	30.5	0.158	53	1
10	8	0	125	96	0	0	0	0.232	54	1
11	4	0	110	92	0	0	37.6	0.191	30	0
12	10	0	168	74	0	0	38	0.537	34	1
13	10	0	139	80	0	0	27.1	1.441	57	0
14	1	0	189	60	23	846	30.1	0.398	59	1
15	5	0	166	72	19	175	25.8	0.587	51	1
16	7	0	100	0	0	0	30	0.484	32	1
17	0	1	118	84	47	230	45.8	0.551	31	1
18	7	0	107	74	0	0	29.6	0.254	31	1
19	1	0	103	30	38	83	43.3	0.183	33	0
20	1	0	115	70	30	96	34.6	0.529	32	1
21	3	0	126	88	41	235	39.3	0.704	27	0
22	8	0	99	84	0	0	35.4	0.388	50	0
23	7	0	196	90	0	0	39.8	0.451	41	1
24	9	0	119	80	35	0	29	0.263	29	1
25	11	0	143	94	33	146	36.6	0.254	51	1

Mapa de calor de correlaciones

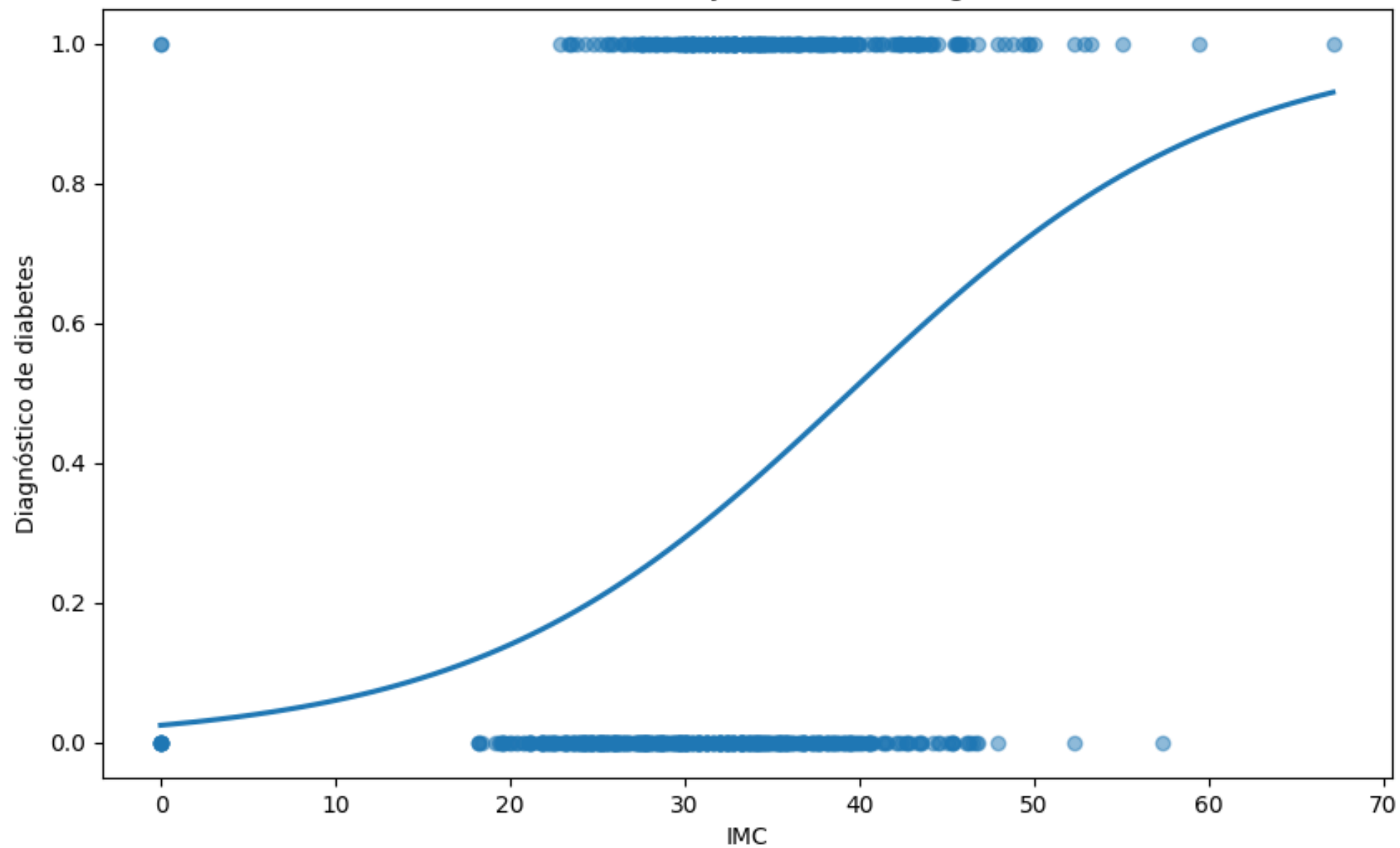


- Representa las relaciones entre variables en forma de matriz bidimensional.
- En esta matriz, cada celda representa la correlación entre dos variables específicas, y el color de la celda indica la fuerza y la dirección de la correlación.
- A partir de aquí y los resultados, se toma la decisión de usar el nivel de Glucosa en sangre y el Índice de Masa Corporal como las variables que mejor explicarían el modelo.

Relación entre glucosa y diabetes con regresión



Relación entre IMC y diabetes con regresión



REGRESIÓN LOGÍSTICA

Precisión del modelo: 0.7662337662337663

Nos indica que aproximadamente el 77 % de las predicciones realizadas por el modelo de regresión logística son correctas. Este es un buen punto de partida.

Coeficientes de la Regresión Logística:

Intercepto: [-0.82437897]

Coeficientes: [[0.6547683 1.09673749]]

Los coeficientes indican la relación logarítmica entre las variables independientes y la probabilidad de tener diabetes.

Intercepto: -0.82437897

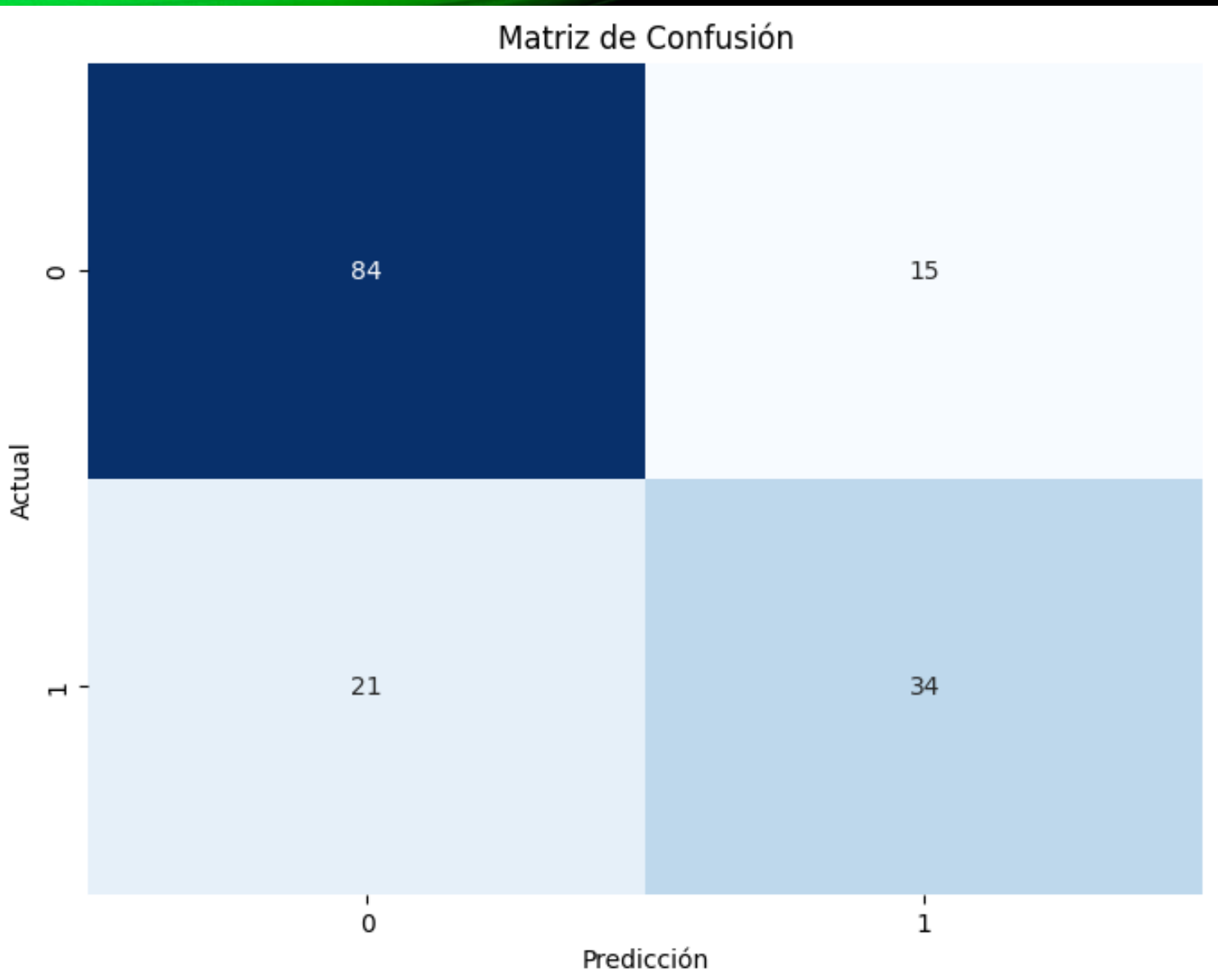
Término constante del modelo. Representa el logaritmo de las probabilidades de la variable dependiente (diabetes) cuando todas las variables independientes (IMC y glucosa) son iguales a cero. En contextos prácticos, este valor tiene un significado limitado, ya que un IMC y nivel de glucosa de cero no son posibles en el ámbito médico.

Coeficiente del IMC: 0.6547683

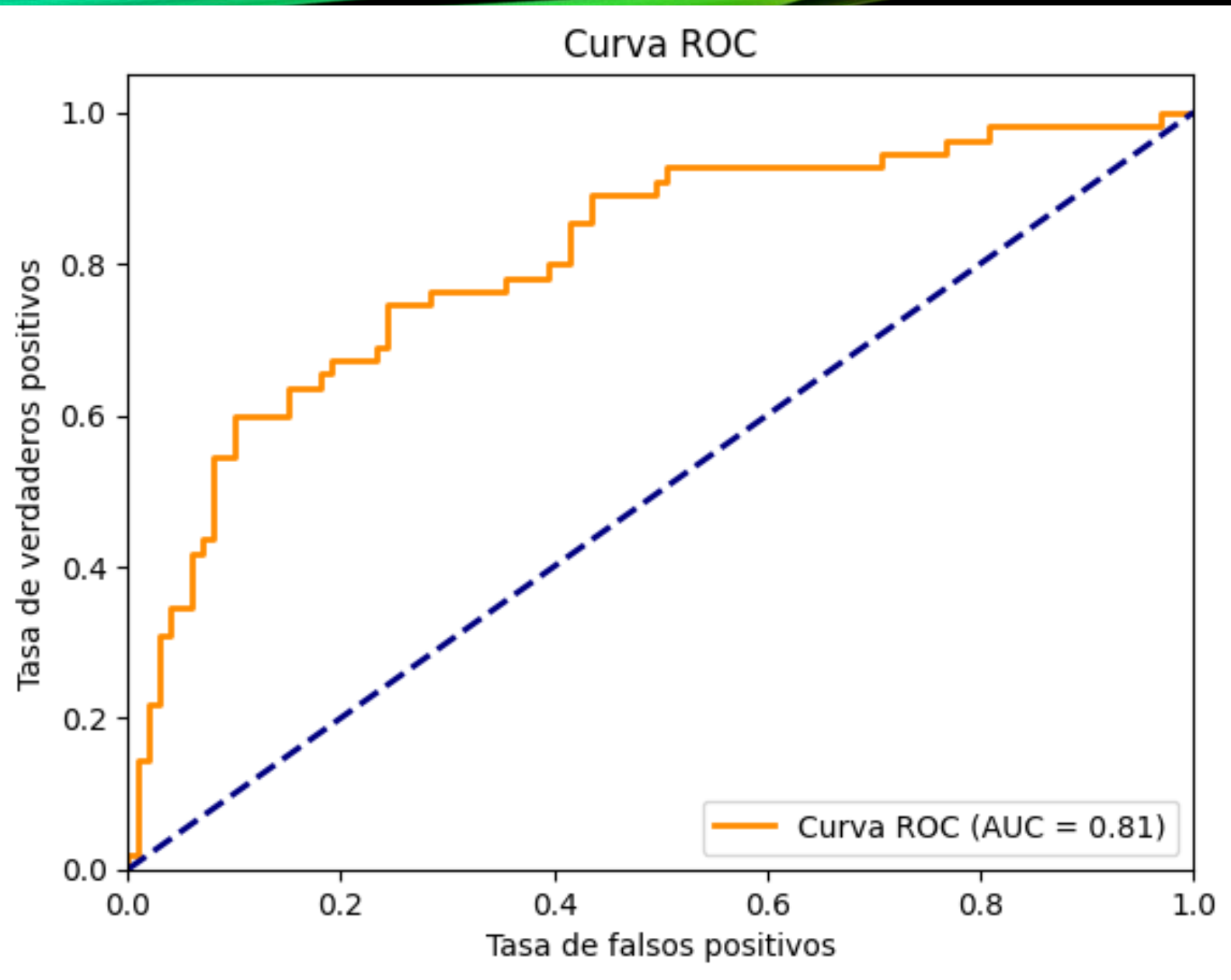
Indica que a medida que el IMC aumenta, la probabilidad de tener diabetes también aumenta. Cada unidad adicional de IMC está asociada con un incremento en el logaritmo de las probabilidades de tener diabetes, en este caso, un incremento de 0.6547683.

Coeficiente del Nivel de Glucosa: 1.09673749

Este también es positivo y tiene un valor mayor que el del IMC. Esto sugiere que el nivel de glucosa tiene una relación más fuerte con la probabilidad de tener diabetes en comparación con el IMC. Un aumento de una unidad en el nivel de glucosa está asociado con un incremento de 1.09673749 en las probabilidades de tener diabetes.



- Se creó una Matriz de Confusión para analizar y ver el desglose de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.
- Resultados:
Verdaderos Negativos (TN): 84
Falsos Positivos (FP): 15
Falsos Negativos (FN): 21
Verdaderos Positivos (TP): 34
- Nos muestra que la precisión general del modelo es aceptable, pero hay margen de mejora, especialmente en la identificación de casos positivos de diabetes



- Se ilustra la capacidad de un modelo para discriminar entre clases positivas y negativas con una curva ROC.
- El AUC (Área bajo la curva) de 0.81 indica que se tiene una buena capacidad para discriminar entre las clases positivas y negativas.
- También sugiere que hay un 81% de probabilidad de que el modelo clasifique correctamente una instancia positiva como más probable que una instancia negativa al azar.



PRE DIAGNOSTICO

- Despues del planteamiento del problema y la construcción del modelo y/o algoritmo, se puede pasar a la parte donde se genera el diagnostico ...
- Continuar la presentación en el navegador ...

RESULTADOS

- Los resultados indican que tanto el IMC como el nivel de glucosa son factores importantes y positivos en la predicción de la diabetes.
- Sin embargo, el nivel de glucosa tiene un impacto mayor en la predicción en comparación con el IMC, dado que su coeficiente es mayor.
- La precisión del modelo del 76.62% sugiere que el modelo tiene un desempeño razonablemente bueno, pero puede haber margen para mejoras.
- Es posible que la adición de más características, un ajuste más fino de los hiperparámetros del modelo, o el uso de técnicas más avanzadas de ingeniería de características puedan mejorar el rendimiento del modelo.
- Este proyecto tiene como objetivo sentar las bases para el desarrollo de un algoritmo más complejo, con el uso de más variables para hacerlo cada vez más exacto en el diagnóstico de esta enfermedad y en un futuro de otras comorbilidades.