



---

# Instituto Tecnológico y de Estudios Superiores de Monterrey

---

## MA3001B

Desarrollo de proyectos de ingeniería matemática

Ingeniería en Ciencias de Datos y Matemáticas

### Reto “Salud Digna”. Entendimiento de los datos

#### Profesores

Dr. Pedro Cervantes Lozano

Dr. Javier Edgardo Garrido Guillén

#### Integrantes de equipo:

Erik Ernesto Ocegueda Sambrano	A01639729
Grace Aviance Silva Aróstegui	A01285158
Marcos Esparza Arizpe	A01634066
Mayavel Torres Hideroa	A01661085

Zapopan, Guadalajara. 06 de junio del 2025.

# ÍNDICE

<b>1. Entendimiento del negocio</b>	<b>3</b>
1.1. Describir el contexto del negocio del reto . . . . .	3
1.2. Definir el problema a atacar . . . . .	3
1.3. Información brindada . . . . .	4
1.4. Definir objetivo y alcance del proyecto . . . . .	5
<b>2. Proceso de nuestra propuesta de solución</b>	<b>6</b>
2.1. Limpieza y Preparación de datos. . . . .	6
2.2. Preprocesamiento de datos para trabajar los Parámetros de la Simulación. . . . .	6
2.3. Parámetros de la Simulación. . . . .	6
2.4. Simulación. . . . .	7
2.5. Extra. Visualización . . . . .	8
<b>3. Estrategia Algorítmica</b>	<b>9</b>
3.1. Algoritmo Exhaustivo. . . . .	9
3.1.1. Definición. . . . .	9
3.1.2. Viabilidad . . . . .	9
3.2. Algoritmos Genéticos . . . . .	10
3.2.1. Definición. . . . .	10
3.2.2. Viabilidad . . . . .	10
3.3. Aspectos clave para la comprensión del algoritmo . . . . .	10
<b>4. Resultados e interpretaciones</b>	<b>12</b>
<b>5. Discusión</b>	<b>14</b>

# **1. Entendimiento del negocio**

## **1.1. Describir el contexto del negocio del reto**

Salud Digna es una institución de salud de origen mexicano enfocada principalmente en ofrecer servicios de análisis y estudios de alta calidad a costos accesibles. Fundada en 2003, Salud Digna ha crecido a lo largo de los años mediante la apertura de múltiples clínicas en distintos estados de la República Mexicana.

Entre sus servicios se incluyen análisis clínicos, estudios de laboratorio (sangre, orina, pruebas de COVID-19, etc.), diagnósticos de imagen (rayos X, ultrasonidos, tomografías, entre otros) y exámenes de la vista. La principal característica de Salud Digna es su enfoque en los análisis y detección oportuna de características de enfermedades, con el objetivo de brindar a la población servicios médicos confiables a precios competitivos.

Además, su modelo de atención suele ser más ágil que el de otros establecimientos, buscando atender a un gran volumen de pacientes de manera eficiente y con tecnología actualizada. En resumen, Salud Digna se ha posicionado en México como una alternativa accesible para la realización de estudios clínicos y de laboratorio, con un énfasis en la prevención y la detección temprana para mejorar el bienestar de la comunidad.

En 2024, Salud Digna hizo historia al ser reconocida como la institución con más unidades acreditadas en todo el mundo por Joint Commission International, en la modalidad de Atención Ambulatoria.

Joint Commission International es una organización estadounidense, distinguida como la máxima autoridad en materia de calidad y seguridad en la atención de los pacientes.

De esta manera, nos posicionamos en el panorama internacional como un modelo mexicano de alta calidad y confiabilidad en la atención de nuestros pacientes, que está comprometido con brindar servicios de primer nivel a las familias mexicanas.

## **1.2. Definir el problema a atacar**

Al ofrecer servicios de análisis clínicos accesibles, lo anterior implica grandes aglomeraciones de personas a ciertas horas del día (principalmente en las mañanas). Esto lleva a que muchos de los que asisten tengan tiempos de espera prolongados dentro de la sucursal.

### 1.3. Información brindada

Se nos brindaron 4 bases de datos. Realizamos el Análisis Exploratorio de ellas. Observamos que MPSF y Turnos Pacientes, abarcan datos respecto a la espera en Caja. Mientras que Data-Tiempos y TurnosPacientes abordan datos de la espera de los pacientes en las modalidades de sus diagnósticos. Optamos por analizar exhaustivamente TurnosPacientes y TiemposPacientes, por ser datos mucho más completos y con más variables para analizar.

	MPSF	DataTiempos	TurnosPacientes	TiemposPacientes
No. Observaciones	58,451	109,167	85,801	109,164
Variables	10	11	16	16
Numericas	FechaID	ReservaciónSP	FechaID	EstudioModalidad
	TurnoHorainicio	TEPFechaEspera	TurnoHorainicio	TEPHorainicio
	TurnoMinutosEspera	TEPHorainicio	TurnoHoraFin	TEPHoraFin
	TAPRecepcionMinutos	TEPHoraFin	TurnoMinutosEspera	TEPMinutos
	TAPRecepcionCaja	TEPMinutos	TAPRecepcionMinutos	TAPFechaAtención
		TAPHorainicio	TAPRecepcionCaja	TAPHorainicio
		TAPHoraFin	PacienteID	TAPHoraFin
		TAPMinutos	PacienteFechaNacimiento	TAPMinutos
			PacienteID	PacienteID
				PacienteFechaNacimiento
Categoricas	Sucursal	Sucursal	Sucursal	Sucursal
	Turno	EstudioModalidad	PacienteGenero	PacienteGenero
	TurnoTipo		Turno	
			TurnoTipo	
Eliminadas	TurnoID		TurnoID	TEPFechaEspera
	Orden		HoraLlegada	ReservaciónSP
			PacienteSP	PacienteSP
			PacienteCodigoPostal	PacienteCodigoPostal

Después de examinar las bases, tuvimos un hallazgo significativo e importante para el enfoque de cómo abordaríamos el problema: de los 85411 datos después de remover duplicados, 1645 tienen tiempos de espera mayores a 20 minutos, lo cual conforma aproximadamente el 2 % de la muestra. Esto nos dice que no es necesario optimizar los tiempos de espera en recepción. Y el optimizarlos más incrementaría el cuello de botella.

Dato	Descripción
Sucursal	Sucursal SD
EstudioModalidad	Servicio realizado
ReservacionSP	Clave servicio realizado
TEPFechaEspera	Fecha tiempo de espera
TEPHorainicio	Hora inicio de espera
TEPHoraFin	Hora fin espera
TEPMinutos	Minutos de espera
TAPFechaAtencion	Fecha tiempo de atencion
TAPHorainicio	Hora inicio de atención
TAPHoraFin	Hora fin atención
TAPMinutos	Minutos de atención

Figura 1: Diccionario de datos de TiemposPacientes

La Figura 1 muestra las principales variables que componen la base de datos de Salud Digna. A continuación se describen con mayor detalle:

- Sucursal: Coyoacán, Culiacán, Culiacán Cañadas, Culiacán La Conquista, Culiacán Las Torres, Culiacán Nakayama, Culiacán Colegio Militar y Culiacán Colegio Militar.

- Estudio Modalidad: Se ofrecen diversas modalidades de servicios, tales como:
  - Densitometría
  - Electrocardiograma
  - Laboratorio
  - Mastografía
  - Nutrición
  - Optometría
  - Papanicolaou
  - Rayos X
  - Resonancia Magnética
  - Tomografía
  - Ultrasonido
  
- Reservación SP: Clave que identifica la reservación realizada por el paciente para un servicio específico, dentro de una sucursal y fecha determinada.
  
- TEP Fecha Espera: Fecha en la que el paciente llegó a la unidad de atención e inició su periodo de espera.
  
- TEP Hora Inicio: Hora exacta en que comenzó el tiempo de espera del paciente.
  
- TEP Hora Fin: Hora en que finalizó el tiempo de espera, justo antes de iniciar la atención médica.
  
- TEP Minutos: Duración total del tiempo de espera, expresada en minutos a partir de la hora de inicio y fin.
  
- TAP Fecha Atención: Fecha en la que el paciente fue atendido por el personal médico.
  
- TAP Hora Inicio: Hora exacta en la que comenzó la atención al paciente.
  
- TAP Hora Fin: Hora en la que finalizó la atención médica.
  
- TAP Minutos: Duración total de la atención, expresada en minutos, calculada a partir de la hora de inicio y fin.

#### 1.4. Definir objetivo y alcance del proyecto

**Objetivo:** *Minimizar el tiempo de espera de los pacientes en las filas de las modalidades, en una ventana máxima de 20 minutos*

**Alcances:** El algoritmo debe de ser lo suficientemente robusto, escalable y flexible para poder aplicarse en cualquier sucursal a nivel nacional.

## 2. Proceso de nuestra propuesta de solución

### 2.1. Limpieza y Preparación de datos.

- Eliminación de columnas no necesarias para el análisis.
- Conversión de columnas de fecha y hora al formato `datetime`.
- Eliminación de duplicados.
- Corrección de registros con fechas erróneas (por ejemplo, año 1900).
- Eliminación de registros con valores inconsistentes, como duración negativa o secuencia temporal incorrecta (eventos fuera de orden lógico).

### 2.2. Preprocesamiento de datos para trabajar los Parámetros de la Simulación.

- Reordenar los eventos de cada estudio por fecha y hora.
- Validar el orden lógico de los eventos (es decir, primero la espera, luego la atención).
- Calcular el tiempo de espera (`TEPMinutos`) y tiempo de atención (`TAPMinutos`).
- Extraer la duración total de atención por estudio.

### 2.3. Parámetros de la Simulación.

Los 3 parámetros clave para nuestra simulación fueron generar para cada sucursal:

- La distribución de llegadas por día, a lo largo del tiempo en cada día de la semana
- Distribución de la cantidad de estudios por la cantidad máxima de cada sucursal
- Probabilidad de que un paciente se realice cada estudio

Para apreciar un caso, podemos observar en las siguientes imágenes estos parámetros para la sucursal de Coyoacán.

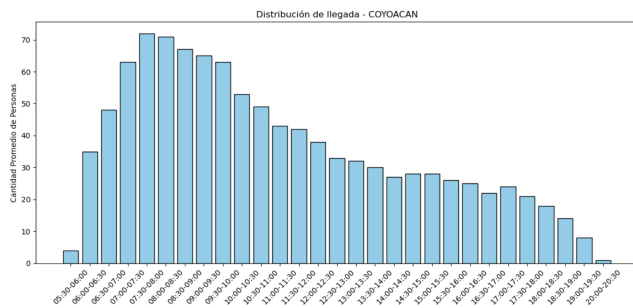


Figura 2: Parámetro 1: Distribución de llegadas

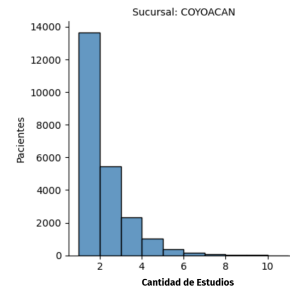


Figura 3: Parámetro 2: Distribución de cantidad de estudios

Sucursal	EstudioModalidad	Probabilidad
COYOACAN	LABORATORIO	0.366373
COYOACAN	RAYOS X	0.13232
COYOACAN	ULTRASONIDO	0.13118
COYOACAN	OPTOMETRIA	0.112957
COYOACAN	DENSITOMETRIA	0.066428
COYOACAN	ELECTROCARDIOGR	0.061017
COYOACAN	MASTOGRAFIA	0.031705
COYOACAN	PAPANICOLAOU	0.028284
COYOACAN	NUTRICION	0.028125
COYOACAN	RESONANCIA MAGN	0.021263
COYOACAN	TOMOGRAFIA	0.020347

Figura 4: Parámetro 3: Probabilidad de realizarse un estudio

## 2.4. Simulación.

Primeramente generamos los datos para la simulación y con ellos fue que ejecutamos nuestra estrategia algorítmica, la cual se detalla en la siguiente sección de este reporte.

Para darle cierto realismo a la simulación, agregamos los siguientes parámetros de ruido que se presentan en la vida real

- Variabilidad en hora de llegada de pacientes.(2.0 min.)
- Variación en tiempo de atención al paciente. (30 %)
- Tiempo base para moverse entre modalidades. (1.5 min)
- Variabilidad en tiempo de desplazamiento. (0.5 min)
- Tiempo medio de preparación previa al siguiente paciente. (1.0 min)
- Probabilidad de que un estudio necesite repetirse. (3 %):

Para realizar una simulación más precisa de los datos, se optó por analizarlos de manera desagregada por sucursal. Es decir, para cada sucursal se calcularon las medias por estudio de forma independiente. Esta estrategia se aplicó a todas las sucursales con el objetivo de reflejar con mayor fidelidad las condiciones particulares de cada una. Esta decisión permite una mayor precisión en la simulación, ya que no todas las sucursales cuentan con la misma capacidad operativa ni con el mismo volumen de pacientes. Utilizar una media general habría implicado una pérdida significativa de información relevante.

2.5. Extra. Visualización

Simulación de colas en tiempo real:





### 3. Estrategia Algorítmica

Para el desarrollo de esta propuesta de solución se tuvo en consideración el uso de dos diferentes algoritmos, un algoritmo exhaustivo y uno genético (GA). Decidimos optar por la implementación de dos algoritmos debido a la naturaleza del problema, al Salud Digna necesitar soluciones rápidas y óptimas decidimos segmentar en dos casos distintos donde, en cada caso particular, utilizar el algoritmo indicado. Nuestro criterio fue el siguiente, si un paciente se realizará menos de 6 estudios se utilizará el algoritmo exhaustivo, caso contrario se utilizará el algoritmo genético, a continuación en cada sección correspondiente se encontrará más detalladamente la razón de esta decisión.

#### 3.1. Algoritmo Exhaustivo.

##### 3.1.1. Definición.

Un **algoritmo exhaustivo**, también conocido como *búsqueda por fuerza bruta*, es un procedimiento sistemático que explora todas las posibles soluciones de un problema, una por una, con el fin de encontrar la solución óptima o verificar si existe una que cumpla con ciertos criterios.

Este tipo de algoritmo garantiza encontrar la solución correcta (si existe), ya que no descarta ninguna posibilidad sin evaluarla. Sin embargo, su principal desventaja es el alto costo computacional, especialmente cuando el espacio de búsqueda es grande, lo que puede llevar a tiempos de ejecución exponenciales.

##### 3.1.2. Viabilidad

Este algoritmo, como su nombre lo indica, realiza una búsqueda exhaustiva buscando el mejor caso entre todos los posibles, sin embargo esto se puede concebir como un arma de doble filo debido a que si el espacio de búsqueda es muy grande la cantidad de combinaciones crecen en orden  $n!$ , lo que directamente implica tiempos de procesamiento largos, volviéndose lento e ineficiente. Aunque lo anterior sea un hecho, este algoritmo sigue teniendo sus ventajas dependiendo el contexto en el que se desenvuelva, en este caso es fácil explotar sus ventajas debido a que el espacio de búsqueda, en la mayoría de los casos es pequeño, entonces sería capaz de encontrar la solución óptima con pocos recursos.

Es por ello que se optará por el uso de este algoritmo cuando la cantidad de estudios a realizar son pocos ( $\leq 5$ ).

## 3.2. Algoritmos Genéticos

### 3.2.1. Definición.

Los **algoritmos genéticos** son una clase de algoritmos de búsqueda y optimización inspirados en los principios de la selección natural y la genética evolutiva, propuestos por John Holland en la década de 1970.

Estos algoritmos trabajan con una población de soluciones potenciales (llamadas individuos o cromosomas), que evolucionan a lo largo de múltiples generaciones. En cada generación, se aplican operadores genéticos como *selección*, *cruce* (crossover) y *mutación* para producir nuevas soluciones. La calidad de cada individuo se evalúa mediante una función de aptitud (*fitness function*), y los más aptos tienen mayor probabilidad de transmitir sus características a la siguiente generación.

Los algoritmos genéticos son especialmente útiles en problemas donde el espacio de búsqueda es grande, complejo o poco conocido, y donde los métodos tradicionales de optimización pueden ser ineficaces.

### 3.2.2. Viabilidad

Este algoritmo es viable porque ofrece una estrategia robusta y flexible para resolver problemas complejos de búsqueda y optimización, especialmente cuando:

- El espacio de búsqueda es grande.
- Es necesario evitar mínimos locales y buscar una solución cercana al óptimo global.
- El problema involucra múltiples variables interdependientes o restricciones difíciles de modelar matemáticamente.

Por ello, optamos por utilizar este algoritmo cuando el paciente se realizará un número mayor de estudios ( $\geq 5$ ).

## 3.3. Aspectos clave para la comprensión del algoritmo

A lo largo del desarrollo de la solución se tomaron en consideración ciertos aspectos, se describen a continuación.

- Por motivos prácticos, se realizó un cambio en el formato de los tiempos. En lugar de trabajar con horas en formato convencional (por ejemplo, 6:00a.m.), se transformaron a

minutos transcurridos desde el inicio del día. Por ejemplo, una llegada a las 6:00a.m. se representa como 360 minutos ( $360 \div 60 = 6$ ). Esta conversión permite simplificar las operaciones posteriores en el algoritmo. Cuando un paciente llega a una cabina que se encuentra ocupada, se suma al tiempo de llegada el promedio de espera correspondiente a ese estudio. Por ejemplo, si el paciente llegó en el minuto 360 y el tiempo promedio de espera es de 5 minutos, la cabina se libera en el minuto 365. Este enfoque facilita el manejo de los tiempos en el algoritmo y permite realizar los cálculos de manera más eficiente y consistente.

- Otra consideración importante es que se asumió un tiempo constante de traslado entre estudios. Es decir, si el algoritmo determina que un paciente finaliza un estudio en el minuto 365, se considera que llegará al siguiente estudio en el minuto 366.5, suponiendo un tiempo fijo de traslado de un minuto y medio. Esta simplificación no solo facilita el cálculo secuencial de los tiempos de atención y mantiene la coherencia temporal en la simulación, sino que también incorpora un factor realista al comportamiento del algoritmo.
- Finalmente, es importante destacar una situación de vital relevancia: aunque se trabaja con datos, estos provienen de acciones humanas, lo que introduce un componente impredecible y genera ruido en la información. Este factor humano puede ser mitigado, pero no eliminado por completo, especialmente en contextos donde los recursos son limitados. Por ello, resulta fundamental implementar estrategias que permitan al algoritmo operar de manera óptima, incluso ante la presencia de incertidumbre y falta de precisión inherente a los datos.

## 4. Resultados e interpretaciones

La solución se probó con datos simulados basados en los datos originales, esto con el fin de obtener un resultado aproximado del funcionamiento del algoritmo en la vida real.

Antes de pasar al análisis de los resultados obtenidos es importante destacar algunos aspectos. Aunque se introdujo ruido para simular la realidad, este no reproduce con exactitud el comportamiento humano, lo cual representa un factor relevante a considerar. En consecuencia, el buen desempeño del algoritmo puede atribuirse, en gran medida, a que opera en un entorno completamente controlado e idealizado, esto no quiere decir que el algoritmo no funcione, sino que los resultados tan ideales se deben a lo anteriormente explicado.

Otro aspecto importante a aclarar es que, en los gráficos presentados, el eje  $y$  representa la frecuencia relativa. Sin embargo, debido a la naturaleza del problema y al contexto en el que se desarrolla, esta puede interpretarse como una probabilidad. Por esta razón, en adelante se utilizará el término 'probabilidad' en lugar de 'frecuencia relativa'. A continuación se detallan individualmente los gráficos obtenidos.

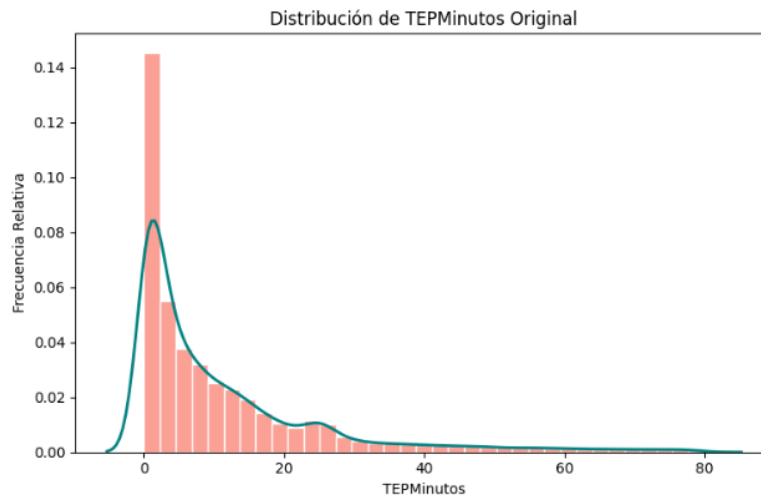


Figura 5: Distribución de los tiempos de espera sin optimizar

En la Figura 5 se muestran los datos de la variable TEPMinutos correspondientes al conjunto original. Puede observarse que la gran mayoría de los valores se concentran entre 0 y 20 minutos, con una probabilidad aproximada del 30 % de que los pacientes esperen dentro de este intervalo. No obstante, la situación se complica al notar que, más allá de los 20 minutos, aún existen numerosos casos de espera, algunos de los cuales alcanzan hasta los 80 minutos. Aunque estos representan porcentajes relativamente bajos, continúan siendo motivo de preocupación.

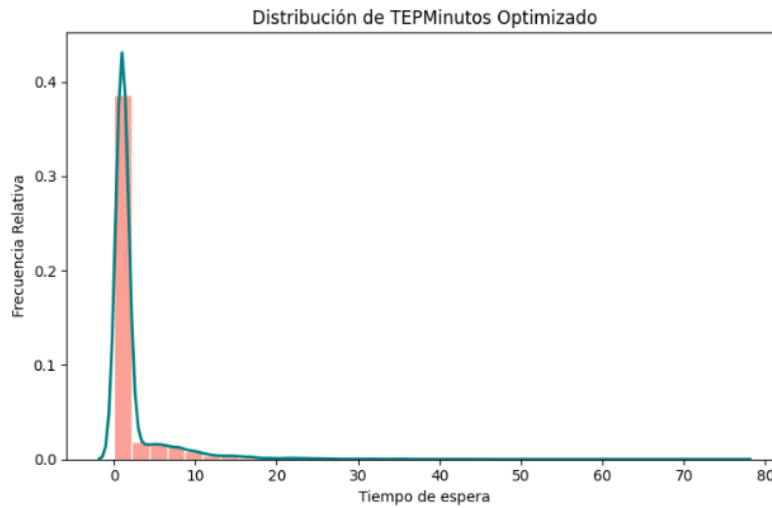


Figura 6: Distribución de los tiempos de espera optimizados

Como se puede observar en la Figura 6, los resultados muestran una mejora sustancial y tangible en los tiempos de espera de los pacientes tras la aplicación del algoritmo. La distribución de los tiempos se ha concentrado completamente en intervalos inferiores a los 20 minutos, lo que cumple con el objetivo planteado inicialmente para el proyecto. Es especialmente destacable el notable pico que se presenta en los primeros 5 minutos, reflejando que una gran proporción de pacientes experimenta ahora un tiempo de espera muy reducido, lo que sin duda impacta positivamente en su experiencia y satisfacción durante el proceso de atención.

Este avance no solo representa una optimización numérica, sino que también implica una mejora en la eficiencia operativa de las sucursales, contribuyendo a la reducción de aglomeraciones y a una mejor distribución de los flujos de pacientes a lo largo del día. Sin embargo, es importante recordar que estos resultados fueron obtenidos en un entorno controlado y simulado, con datos que, aunque basados en la realidad, no capturan por completo las variabilidades y factores impredecibles inherentes a la atención humana y a la dinámica diaria en cada sucursal.

Por lo tanto, aunque el desempeño del algoritmo es alentador y ofrece una solución prometedora para mitigar los tiempos de espera, es necesario ser cautelosos al extrapolar estos resultados a escenarios reales, donde pueden presentarse circunstancias no previstas que afecten la eficacia del sistema. En resumen, este análisis demuestra el potencial del modelo, pero también subraya la importancia de seguir validando y ajustando la solución mediante pruebas en campo y un monitoreo continuo para garantizar su efectividad a largo plazo.

## 5. Discusión

El presente proyecto abordó el problema de los tiempos de espera prolongados en las modalidades de atención de Salud Digna, particularmente en horas pico. A partir de una caracterización profunda del negocio y del análisis exploratorio de los datos, se diseñó una solución que combina simulaciones realistas y algoritmos de optimización, con el fin de reducir significativamente el tiempo de espera de los pacientes.

Uno de los aspectos más relevantes descubiertos durante el análisis fue que solamente el 2 % de los registros superaban los 20 minutos de espera en recepción. Este hallazgo resultó fundamental para redirigir el enfoque del problema: no era necesario mejorar los tiempos de espera en la recepción, sino optimizar el flujo de atención en las modalidades clínicas, donde se presentaban los verdaderos cuellos de botella.

La simulación por sucursal y el cálculo de parámetros específicos como la distribución de llegadas, los tipos de estudios solicitados y la capacidad instalada por unidad médica permitieron un enfoque más localizado y realista. Esta desagregación contribuyó significativamente a la precisión del modelo y facilitó su escalabilidad a nivel nacional.

En cuanto a la propuesta algorítmica, la estrategia de utilizar un algoritmo exhaustivo para pacientes con pocos estudios y un algoritmo genético para casos más complejos se mostró adecuada. Esta segmentación permitió balancear la necesidad de obtener soluciones óptimas sin comprometer el tiempo de respuesta, algo esencial en un contexto operativo donde se atiende a miles de pacientes diariamente. La decisión de utilizar minutos como unidad base de tiempo y asumir un tiempo constante de traslado entre estudios también ayudó a simplificar el modelo sin perder demasiado realismo.

Sin embargo, es importante reconocer ciertas limitaciones. En primer lugar, aunque se introdujo ruido a los datos para simular un comportamiento más realista, los resultados siguen estando condicionados por un entorno controlado. En situaciones reales, existen múltiples factores que pueden alterar el flujo de atención: inasistencias, retrasos del personal, fallos técnicos, emergencias médicas, entre otros. Estos elementos, al ser inherentemente impredecibles, pueden afectar la efectividad del algoritmo.

Por otro lado, el modelo parte de promedios y distribuciones históricas para estimar la demanda, lo cual puede ser insuficiente en temporadas de alta variabilidad o durante campañas especiales de atención gratuita o con descuento. También hay que considerar la resistencia al cambio por parte del personal operativo o las dificultades técnicas para integrar el algoritmo en los sistemas actuales de gestión clínica.

A pesar de estas limitaciones, los resultados obtenidos muestran una clara tendencia a la mejora. Las simulaciones permitieron observar una disminución considerable en la proporción de pacientes con tiempos de espera superiores a los 20 minutos. Esto valida la factibilidad técnica de la propuesta y sugiere que, con una implementación adecuada, podría mejorar de forma tangible la experiencia del paciente y la eficiencia operativa de las sucursales.

En futuras etapas del proyecto, sería valioso considerar una validación empírica en sucursales reales, así como una retroalimentación directa del personal médico y administrativo. Además, se podrían incorporar modelos predictivos para anticipar la demanda y adaptar dinámicamente la asignación de recursos. Finalmente, integrar un componente de aprendizaje automático que mejore con el tiempo podría convertir la solución en un sistema inteligente adaptable a contextos cambiantes.