



**EDUCACIÓN**

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

Instituto Tecnológico de Orizaba

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

OPCIÓN I.- TESIS

TRABAJO PROFESIONAL

“SISTEMA PARA LA PREDICCIÓN DE CONCENTRACIÓN  
DE COLESTEROL EN LA SANGRE UTILIZANDO BIG DATA  
Y TÉCNICAS DE SIMULACIÓN MESOSCÓPICA CON  
DINÁMICA DE PARTÍCULAS DISIPATIVAS (DPD)”

QUE PARA OBTENER EL GRADO DE:  
MAESTRA EN SISTEMAS  
COMPUTACIONALES

PRESENTA:

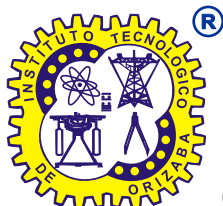
*I.S.C. Reyna Nohemy Soriano Machorro*

DIRECTOR DE TESIS:

*Dr. José Luis Sánchez Cervantes*

CODIRECTOR DE TESIS:

*Dr. Luis Rolando Guarneros Nolasco*



ORIZABA, VERACRUZ, MÉXICO.

NOVIEMBRE 2022



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

Instituto Tecnológico de Orizaba  
División de Estudios de Posgrado e Investigación

Orizaba, Veracruz, **11/noviembre/2022**  
Dependencia: **División de Estudios de  
Posgrado e Investigación**  
Asunto: **Autorización de Impresión**  
OPCION: I

**C. REYNA NOHEMY SORIANO MACHORRO**  
**CANDIDATO A GRADO DE MAESTRO EN:**  
**SISTEMAS COMPUTACIONALES**  
**P R E S E N T E.-**

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

**" Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD)"**

comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica®*  
CIENCIA - TÉCNICA - CULTURA®

**DR. MARIO LEONCIO ARRIJOJA RODRÍGUEZ**  
**JEFE DE LA DIVISIÓN DE ESTUDIOS**

**DE POSGRADO E INVESTIGACIÓN**

SECRETARÍA DE EDUCACIÓN PÚBLICA NACIONAL DE MÉXICO

**INSTITUTO TECNOLÓGICO DE  
ORIZABA  
DIV. DE EST. DE  
POSGRADO E INVEST.**

OG-13-F06



Av. Oriente 9 Núm.852, Colonia Emiliano Zapata. C.P. 94320 Orizaba, Veracruz. Tel. 01  
(272)1105360 e-mail: cyd\_orizaba@tecnm.mx tecnm.mx | orizaba.tecnm.mx



**2022 Flores**  
Año de Magón  
PRELUDIO DE LA REVOLUCIÓN MEXICANA



Orizaba, Veracruz, 31/octubre/2022  
Asunto: **Revisión de trabajo escrito**

**C. MARIO LEONCIO ARRIJOJA RODRÍGUEZ  
JEFE DE LA DIVISIÓN DE ESTUDIOS  
DE POSGRADO E INVESTIGACIÓN  
P R E S E N T E.-**

Los que suscriben, miembros del jurado, han realizado la revisión de la Tesis del (la) C.

**REYNA NOHEMY SORIANO MACHORRO**

La cual lleva el título de:

**Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data  
y técnicas de simulación mesoscópica con Dinámica de Partículas Disipativas (DPD)**

Y concluyen que se acepta.

**ATENTAMENTE**  
Excelencia en Educación Tecnológica®  
CIENCIA - TÉCNICA - CULTURA®

**PRESIDENTE: DR. JOSÉ LUIS SÁNCHEZ CERVANTES**

FIRMA

**SECRETARIO: M.C. MA. ANTONIETA ABUD FIGUEROA**

FIRMA

**VOCAL: M.C.E BEATRIZ ALEJANDRA OLIVARES  
ZEPAHUA**

FIRMA

**VOCAL SUP.: DR. LUIS ROLANDO GUARNEROS  
NOLASCO**


FIRMA

TA-09-18



## *CARTA DE ORIGINALIDAD*


En la ciudad de Orizaba, Veracruz, el día 25 del mes de noviembre del año 2022, el (la) que suscribe Soriano Machorro Reyna Nohemy, alumno (a) del programa de Maestría en Sistemas Computacionales con número de control M20011509, manifiesta que es autor (a) del trabajo de tesis titulado “Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD).” y declaro que el trabajo es original, ya que sus contenidos son producto de mi directa contribución intelectual. Todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal y, en los casos que así lo requieran, cuento con las debidas autorizaciones de quienes poseen los derechos patrimoniales. Por lo tanto, me hago responsable de cualquier litigio o reclamación relacionada con derechos de propiedad intelectual, exonerando de toda responsabilidad al Tecnológico Nacional de México / Instituto Tecnológico de Orizaba.

  
Reyna Nohemy Soriano Machorro  
Nombre y firma

## *CARTA DE CESIÓN DE DERECHOS*

En la ciudad de Orizaba, Veracruz, el día 25 del mes de noviembre del año 2022, el (la) que suscribe Soriano Machorro Reyna Nohemy, alumno (a) del programa de Maestría en Sistemas Computacionales con número de control M20011509, manifiesta que es autor (a) del trabajo de tesis bajo la dirección de Dr. José Luis Sánchez Cervantes y cede los derechos del trabajo de tesis titulado “Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD).” al TecNM/Instituto Tecnológico de Orizaba para su difusión y divulgación, con fines académicos y de investigación.

Queda estrictamente prohibido reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del Tecnológico Nacional de México/Instituto Tecnológico de Orizaba. Este puede obtenerse escribiendo a la siguiente dirección: [msc@orizaba.tecnm.mx](mailto:msc@orizaba.tecnm.mx) . Si el permiso se otorga, cualquier usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

  
Reyna Nohemy Soriano Machorro  
Nombre y firma

## **Agradecimientos**

Al Tecnológico Nacional de México por la oportunidad de ingresar a la maestría en sistemas computacionales.

Al Consejo Nacional de Ciencia y Tecnología, CONACyT, por el apoyo económico otorgado para la realización de mis estudios de maestría.

A mis profesores de la maestría en sistemas computacionales, gracias por compartir sus conocimientos, por su dedicación en cada clase impartida, especialmente al Dr. José Luis Sánchez Cervantes y al Dr. Luis Rolando Guarneros Nolasco por su apoyo en la realización de este proyecto de tesis.

A mis compañeros y amigos de la maestría gracias por los buenos momentos que hemos compartido durante los trabajos en equipo.

A mi familia gracias por siempre impulsarme a cumplir mis sueños, alentarme en la búsqueda del conocimiento, enseñarme a mantener la esperanza en los momentos más oscuros de la vida y a disfrutar de cada instante de felicidad.

A mis abuelitos por compartir su sabiduría y ser un ejemplo de perseverancia.

# Índice general

Índice de figuras.....	iv
Índice de tablas.....	v
Índice de lista .....	vi
Resumen.....	vii
Introducción .....	ix
Capítulo 1 Antecedentes.....	1
1.1 Marco teórico .....	1
1.1.1 Colesterol .....	1
1.1.2 Colesterol de tipo HDL .....	1
1.1.3 Colesterol de tipo LDL.....	1
1.1.4 Enfermedades cardiovasculares .....	1
1.1.5 Algoritmo .....	2
1.1.6 Simulación.....	2
1.1.7 Simulación Mesoscópica.....	2
1.1.8 Dinámica Molecular.....	2
1.1.9 Dinámica de Fluidos.....	2
1.1.10 Dinámica de partículas disipativas .....	2
1.1.11 Ensemble gran canónico.....	3
1.1.12 Ensemble canónico (Número de partículas, volumen y temperatura constantes).....	3
1.1.13 Método Monte Carlo .....	3
1.1.14 Fluido confinado .....	3
1.1.15 Computación acelerada por GPU .....	3
1.1.16 Arquitectura CUDA .....	3
1.1.17 Apache Cassandra .....	4
1.1.18 MongoDB.....	4
1.1.19 Minería de datos .....	4
1.1.20 5 Vs del Big Data .....	4
1.1.21 Regresión Lineal .....	5
1.1.22 Árbol de decisión CART.....	5
1.1.23 Regresión no lineal.....	5

1.1.24 Regresión lineal multivariante.....	5
1.1.25 Regresión Logística.....	5
1.2 Situación tecnológica, económica y operativa de la empresa .....	6
1.3 Planteamiento del problema .....	6
1.4 Objetivo general y específico .....	7
1.4.1 Objetivo General .....	7
1.4.2 Objetivos específicos.....	7
1.5 Justificación.....	8
Capítulo 2 Estado de la práctica.....	10
2.1 Trabajos relacionados.....	10
2.1.1 Modelos para simulación de fluidos.....	10
2.1.2 Aplicación de Minería de Datos para la prevención de enfermedades cardiovasculares .....	15
2.2 Análisis comparativo.....	18
Capítulo 3 Aplicación de la Metodología.....	25
3.1 Planificación.....	26
3.2 Diseño del sistema.....	29
3.2.1 Arquitectura del sistema.....	30
3.2.1.1 Capas de la arquitectura .....	31
3.2.1.2 Componentes de la arquitectura .....	31
3.2.1.3 Flujo de Trabajo .....	32
3.2.2 Análisis de requerimientos .....	32
3.2.3 Modelo de Datos .....	33
En el modelo de datos se observan ocho documentos, en la siguiente tabla se muestran los documentos del modelo con su descripción. ....	34
3.2.4 Modelado de funcionamiento.....	38
3.2.5 Diseño de la aplicación .....	39
3.2.6 Mapa de navegación.....	48
3.2.7 Modelo de contenido.....	50
3.3 Desarrollo.....	51
3.3.1 Módulo de Simulación .....	51
3.3.2 Módulo de Predicción .....	57
3.3.3 Capa de Presentación .....	59



Capítulo 4 Resultados .....	65
4.1 Caso de estudio .....	65
Capítulo 5 Conclusiones y recomendaciones .....	94
5.1 Conclusiones .....	94
5.2 Recomendaciones.....	95
Bibliografía .....	96

## Índice de figuras

Figura 3.1 Fases de la metodología XP.....	25
Figura 3.2 Arquitectura para el sistema .....	30
Figura 3.3 Diagrama de Caso de usos .....	32
Figura 3.4 Modelo de datos.....	34
Figura 3.5 Diagrama de actividades de la predicción de concentración de colesterol en la sangre .....	38
Figura 3.6 Diagrama de actividades actividad para acceder a predicciones ya registradas .....	39
Figura 3.7 <i>Mockup</i> para autenticación del usuario. ....	40
Figura 3.8 <i>Mockup</i> Registro de usuarios.....	41
Figura 3.9 <i>Mockup</i> de datos generales de usuario.....	42
Figura 3.10 <i>Mockup</i> Análisis .....	43
Figura 3.11 <i>Mockup</i> Predicción de niveles de Colesterol .....	44
Figura 3.12 <i>Mockup</i> de simulación de colesterol total en la sangre .....	45
Figura 3.13 <i>Mockup</i> Menú .....	46
Figura 3.14 <i>Mockup</i> Predicciones .....	47
Figura 3.15 <i>Mockup</i> Perfil de Paciente .....	48
Figura 3.16 Mapa navegacional .....	49
Figura 3.17 Diagrama de paquetes.....	50
Figura 3.18 210 mg/dl de colesterol en la sangre.....	52
Figura 3.19 290 mg/dl de colesterol en la sangre.....	53
Figura 3.20 Gráfica colesterol por edad. ....	57
Figura 3.21 Gráfica obtenida de la predicción del rango de edad de 20 a 80 años. ....	59
Figura 3.22 Interfaz para el inicio de sesión. ....	60
Figura 3.23 Interfaz para el registro de usuario.....	60
Figura 3.24 Interfaz menú de usuario.....	61
Figura 3.25 Interfaz para registro de Análisis. ....	61
Figura 3.26 Interfaz de predicción de los niveles de colesterol. ....	62
Figura 3.27 Interfaz del modelo de mesoscópico.....	64

## Índice de tablas

Tabla 2.1 Comparación de trabajos relacionados con simulación .....	18
Tabla 2.2 Análisis comparativo de artículos relacionados con Minería de Datos.....	22
Tabla 3.1 Historia de usuario 1 .....	26
Tabla 3.2 Historia de usuario 2 .....	26
Tabla 3.3 Historia de usuario 3 .....	27
Tabla 3.4 Historia de usuario 4 .....	27
Tabla 3.5 Historia de usuario 5 .....	27
Tabla 3.6 Descripción del diagrama de caso de uso de la figura 3.2 .....	33
Tabla 3.7 Descripción del modelo de datos de la figura 3.4. ....	35
Tabla 3.8 Descripción de los datos que contiene el documento Paciente de tipo Local del modelo de datos de la figura 3.4. ....	35
Tabla 4.1 Ejemplo de datos de análisis proporcionados por el sanatorio Escudero.....	66
Tabla 4.2 Llenado de formulario para la predicción de colesterol de los pacientes.....	67
Tabla 4.3 Resultado de la predicción de cada paciente de este caso de estudio.....	73
Tabla 4.4 Resultado de la predicción de cada paciente de este caso de estudio.....	79
Tabla 4.5 Reporte de la predicción de cada paciente. ....	85
Tabla 4.6 Reporte de la predicción de cada paciente. ....	91

## Índice de lista

Lista 1 fragmento de código del módulo de simulación. ....	51
Lista 2 fragmento de posiciones de moléculas. ....	54
Lista 3 fragmento de bibliotecas utilizadas para la construcción de DPD. ....	55
Lista 4 Fragmento para la construcción de un modelo DPD. ....	55
Lista 5 fragmento para la construcción de un modelo DPD con datos de un archivo con formato .xyz. ....	56
Lista 6 Fragmento para la colorear el modelo. ....	57
Lista 7 Fragmento de código del modelo para la regresión logística. ....	58
Lista 8 Fragmento de código para la gráfica de predicción. ....	63

## **Resumen**

Cuatro de cada diez adultos mexicanos tienen colesterol alto, según el Instituto Nacional de Cardiología. El colesterol es esencial para la producción de sustancias en el organismo por ejemplo hormonas, metabolismo de la vitamina D, es imprescindible para la absorción de calcio y ácidos biliares, el exceso de colesterol genera su endurecimiento y estrechamiento en las paredes de las arterias y forman un coágulo que causa un ataque cardíaco o un accidente cerebrovascular.

Teniendo en cuenta esta problemática, en este proyecto de tesis se desarrolló la aplicación “PREDIROL” (Programa para la predicción de niveles de saturación de colesterol) en la cual se presenta un enfoque con Big Data y técnicas de simulación mesoscópica con un método de Dinámica de Partículas DPD (Dinámica de partículas disipativas). Para la construcción del modelo se utilizó cómputo en paralelo empleando CUDA.

Este proyecto busca la concientización sobre el cuidado de los niveles de concentración del colesterol, ya que el tener los niveles altos es perjudicial para la salud, pero al tener los niveles de concentración bajos, el organismo no produce células en el organismo. Esta es una herramienta para la medicina preventiva y mejorar el estilo de vida de los usuarios antes de desarrollen padecimientos más graves e incluso infartos o accidentes cerebrovasculares.

## **Abstract**

Four out of ten Mexican adults have high cholesterol, according to the National Institute of Cardiology. Cholesterol is essential for the production of substances in our body such as hormones, vitamin D metabolism, it is essential for the absorption of calcium and bile acids, excess cholesterol causes hardening and narrowing in the walls of the arteries and can form a clot that causes a heart attack or stroke.

Taking into account this problem, this thesis project develops the application "PREDIROL" (Program for the prediction of cholesterol saturation levels) which presents an approach with Big Data and mesoscopic simulation techniques with a method of Particle Dynamics DPD (Dissipative Particle Dynamics). Parallel computing using CUDA was used to build the model.

PREDIROL seeks to raise awareness about the care of cholesterol concentration levels, since having high levels is detrimental to health, but having low concentration levels, the body does not produce cells in the body. This is a tool for preventive medicine and to improve the lifestyle of users before they develop more serious ailments and even heart attacks or strokes.

## **Introducción**

El colesterol es una sustancia cerosa y parecida a la grasa que se encuentra en todas las células del cuerpo humano. El colesterol es importante para la buena salud y se necesita para fabricar las paredes de las células, los tejidos, las hormonas, la vitamina D y el ácido de la bilis. El colesterol también proviene del consumo de alimentos de origen animal, como las yemas de huevo, la carne y los productos lácteos elaborados con leche entera. Existen dos tipos de colesterol en el cuerpo humano, la HDL (Lipoproteína de alta densidad compuesta por lípidos y proteínas), conocido comúnmente como colesterol bueno y la LDL (Lipoproteína de baja densidad) conocido como el colesterol malo, el cual cuando hay demasiado colesterol en la sangre, se acumulan en las paredes de los vasos sanguíneos, bloquear el flujo de sangre a los tejidos y los órganos, y aumentar el riesgo de contraer enfermedades del corazón y sufrir accidentes cerebrovasculares y de acuerdo al Instituto Nacional de Cardiología, cuatro de cada diez adultos mexicanos tienen colesterol alto.

Por este motivo, en proyecto de tesis presente se propone la creación de un sistema para la predicción de concentración de colesterol en la sangre, para concientizar a la población sobre el cuidado y prevención de enfermedades cardiovasculares con un enfoque con Big Data y técnicas de simulación mesoscópica con un método de Dinámica de Partículas DPD.

Este documento se organiza en cinco capítulos principales: en el capítulo uno abarca el marco teórico, el planteamiento del problema, el objetivo general y los objetivos específicos, además de la justificación; el capítulo dos abarca el estado de arte y un análisis comparativo; el tercer capítulo presenta la metodología seleccionada para el desarrollo de la solución, abarca cada una de las fases y las actividades realizadas; en el cuarto capítulo se describen los resultados obtenidos y el caso de estudio en la cual se aplicó la solución; en el capítulo cinco se realizan una conclusión general del trabajo y se anexan las recomendaciones.

## Capítulo 1 Antecedentes

En este capítulo se presentan los conceptos más relevantes relacionados con el trabajo a realizar, también se presenta el problema, objetivos y justificación del proyecto a desarrollar.

### 1.1 Marco teórico

A continuación, se presentan los conceptos relacionados

#### 1.1.1 Colesterol

El colesterol es un esteroide que constituye el componente esencial de la membrana celular y es precursor de las hormonas esteroideas, varias vitaminas y la bilis. Una fracción del colesterol se obtiene de los alimentos, pero la mayor proporción se sintetiza en el hígado y de allí pasa a la circulación general, siendo transportado por las lipoproteínas, las cuales son de diferente peso molecular [1].

#### 1.1.2 Colesterol de tipo HDL

La HDL (Lipoproteína de alta densidad compuesta por lípidos y proteínas) tiene un diámetro entre 25 y 10 nm y los fosfolípidos son su principal lípido. Son producidos por el hígado (30%) y el intestino (70%) y su función principal es extraer el colesterol sobrante de las células y transportarlo al hígado para su eliminación en forma de ácidos biliares y colesterol en las heces [2].

#### 1.1.3 Colesterol de tipo LDL

Lipoproteína de baja densidad son partículas ricas en colesterol con un diámetro de 20-25 nm que son captadas por las células del cuerpo y de ese modo se proveen del colesterol que requieran [2].

#### 1.1.4 Enfermedades cardiovasculares

Las enfermedades cardiovasculares abarcan una amplia gama de trastornos, incluidas las enfermedades del músculo cardíaco y del sistema vascular que irriga el corazón, el cerebro y otros órganos vitales. Es la principal causa de muerte en todo el mundo [3].



### **1.1.5 Algoritmo**

El algoritmo es un modelo lógico-matemático que declara reglas definidas, no ambiguas, ordenadas para solucionar un problema o procesar datos, por mencionar algunas, se expresa en lenguaje natural, diagramas de flujo, lenguajes de programación y pseudocódigo, entre otros [4].

### **1.1.6 Simulación**

La simulación es el proceso para diseñar un modelo de un sistema real, para lo cual se utiliza la técnica numérica para conducir experimentos en una computadora, con la finalidad de entender el comportamiento del sistema, así como los riesgos dentro de límites impuestos, también se utiliza para calcular la probabilidad con la que ocurren ciertos eventos complejos en experimentos de simulación [5].

### **1.1.7 Simulación Mesoscópica**

La simulación mesoscópica es un método alternativo, limpio y económico para estudiar medios en los cuales no habrá exposición para los experimentos, permite superar las limitaciones para la simulación molecular, porque se dirige al comportamiento de los átomos [6].

### **1.1.8 Dinámica Molecular**

La dinámica molecular (MD) es un tipo de simulación computacional para analizar la evolución de un sistema (físico, químico o biológico) en determinado tiempo y su comportamiento, su cálculo está conformado por las ecuaciones de movimiento de Newton en donde se aplica a las fuerzas entre los átomos que conforma la molécula [7].

### **1.1.9 Dinámica de Fluidos.**

La dinámica de fluidos estudia los fluidos en movimiento con un conjunto de leyes de conservación para describir su comportamiento. Este se representa por modelos, ya que son muy complejos [8].

### **1.1.10 Dinámica de partículas disipativas**

La dinámica de partículas disipativas es un método para la simulación mesoscópica que fue desarrollada para estudiar el comportamiento hidrodinámico de fluidos complejos, también es útil para el cálculo de propiedades dinámicas [9].

### **1.1.11 Ensamble gran canónico**

El ensamble gran canónico describe los estados de un sistema con un número de partículas arbitrario siendo especificado su promedio, volumen, temperatura y su potencial químico de cada constituyente. Este ensamble es natural para el estudio de sistemas físicos, ya que el sistema intercambia partículas o energía con el entorno. [10]

### **1.1.12 Ensamble canónico (Número de partículas, volumen y temperatura constantes)**

El ensamble (también llamado colectivo canónico o colectividad canónica) describe los estados de un sistema con número de partículas constantes, volumen constante y temperatura constante, en contacto entre sí, este intercambia energías, pero no partículas. Con estas condiciones la energía no es constante en cada sistema y se da a conocer la energía media del conjunto [10].

### **1.1.13 Método Monte Carlo**

Es un método que originalmente se desarrolló para evaluar ecuaciones integrales multidimensionales u obtener la solución de ecuaciones integrales, pero también se utiliza para simular moléculas en líquidos [11].

### **1.1.14 Fluido confinado**

Se refiere a cualquier fluido dentro de un conducto lleno completamente, que en ocasiones tiene un flujo que es potencial o laminado entre fronteras sólidas en reposo, como ocurre en un tubo, las fronteras o paredes son lisas o porosas [12].

### **1.1.15 Computación acelerada por GPU**

Este tipo de computación acelerada por GPU (Unidad de Procesamiento Gráfico), permite trabajar en paralelo junto a una CPU (Unidad Central de Procesamiento) para acelerar el funcionamiento de las aplicaciones [13].

### **1.1.16 Arquitectura CUDA**

CUDA (*Compute Unified Device Architecture*, Arquitectura Unificada de Dispositivos de Cómputo) es una interfaz de programación de programación que permite acelerar los *Kernels* de cómputo en la GPU, trabaja con código escrito en C o C++, funciona en las tarjetas de la marca nVidia a partir de la serie G8X [14].

### 1.1.17 Apache Cassandra

Apache Cassandra es un sistema de almacenamiento de datos de código abierto, se diferencia con otros gestores de base de datos ya que es NoSQL (*No Structured Query Language*, No es un lenguaje de consulta estructurada), esta tiene un soporte robusto para múltiples centros de datos. Sus tablas son creadas, eliminadas y alteradas en tiempo de ejecución, sin que las actualizaciones o consultas se bloqueen [15].

### 1.1.18 MongoDB

La base de datos MongoDB está basada en documentos, ya que almacena datos en forma de documentos tipo JSON (*JavaScript Object Notation*, notación de objeto de JavaScript) llamados BSON (similar a JSON) para ser más expresiva y potente, es de código abierto y es NoSQL [16].

### 1.1.19 Minería de datos

La minería de datos es la extracción del conocimiento de grandes cantidades de datos, el cual busca un conjunto de datos valiosos o patrones interesantes a partir de un conjunto de datos grande. Para extraer o minar datos existen técnicas para generar distintos modelos, estos provienen de la inteligencia artificial y la estadística [17].

### 1.1.20 5 Vs del Big Data

Las 5 Vs del *Big Data* (BD) son las características de un conjunto de datos enorme y complejo, a continuación se describen brevemente sus características [18].

- Volumen: se refiere al tamaño de datos que se generan hoy en día.
- Velocidad: los datos se almacenan y se generan a una gran velocidad.
- Variedad: el origen de los datos es heterogéneo, ya que proviene de muchas herramientas o soportes, estos datos no se encuentran en una forma estructurada.
- Veracidad: por la cantidad de datos almacenados en la BD no es posible que el 100% de los datos sean correctos.
- Valor: este es generado por la cantidad de datos una vez convertidos en información.

### **1.1.21 Regresión Lineal**

La regresión lineal es una técnica de la minería de datos la cual se utiliza para modelar la predicción con variables de entradas y salidas continuas, en este modelo la media de una variable Y es dependiente de una o más variables X. La variable Y es llamada variable de salida o dependiente, mientras que la variable X es llamada variable de entrada o explicativa (en ocasiones las variables X no son dependientes entre sí) [19].

### **1.1.22 Árbol de decisión CART**

El árbol de decisión CART (*Classification And Regression Trees*, Árboles de Clasificación y de Regresión) genera decisiones binarias, lo cual hace que cada nodo se divida en dos ramas para que sea lo más homogéneo posible a la variable dependiente [20].

### **1.1.23 Regresión no lineal**

La regresión no lineal es un algoritmo que genera una ecuación para describir la relación que una variable de respuesta continua y una variable productora, su solución es encontrada a través de un proceso interactivo que converge en una solución de límites [21].

### **1.1.24 Regresión lineal multivariante**

La regresión lineal multivariante o múltiple es una extensión de la regresión lineal simple, permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta se determina a partir de un conjunto de variables independientes llamadas predictores, este modelo se utiliza ampliamente para predecir el comportamiento de las variables de respuesta asociadas a cambios en las variables predictoras, una vez que se establece un grado de relación deseado [22].

### **1.1.25 Regresión Logística**

La Regresión Logística es de los algoritmos de *Machine Learning* más simples, esta es un método de regresión y es común resolver problemas de clasificación binaria con este tipo de algoritmo. Este tipo de Regresión se utiliza para calcular la probabilidad de que ocurra un evento, ya que su resultado o variable objetivo es de naturaleza dicotómica (solo hay dos clases posibles)[23].

Existen tres tipos de Regresión Logística, los cuales se describen a continuación.

- Regresión Logística Binaria: la variable objetivo tiene solo dos resultados posibles 0 o 1 (ausencia o presencia, respectivamente).
- Regresión Logística Multinomial: este tipo tiene de 3 en adelante categorías llamadas covariables, estas son cualitativas o cuantitativas, en este algoritmo se selecciona una variable dependiente cualitativa. Las covariables cualitativas deben ser dicotómicas (valor de 0 o 1). Si la covariable tuviera más de dos categorías se debe realizar una transformación de la misma en varias covariables cualitativas dicotómicas ficticias (variables *dummy*). Al hacer esta transformación cada categoría de la variable entraría en el modelo de forma individual.
- Regresión Logística Ordinal: este algoritmo da una forma a la dependencia de una respuesta ordinal politómica sobre un conjunto de predictores, que son factores o covariables. la variable objetivo tiene tres o más categorías ordinales, dar forma a la dependencia de una respuesta ordinal politómica sobre un conjunto de predictores, que son factores o covariables. Los covariables deben ser numéricos[24].

## 1.2 Situación tecnológica, económica y operativa de la empresa

El Instituto Tecnológico de Orizaba es una institución pública que depende de la Secretaría de Educación Pública SEP y pertenece al Tecnológico Nacional de México. Se encuentra ubicado en Oriente 9, Colonia Emiliano Zapata, en la ciudad de Orizaba, Veracruz. Esta institución ofrece carreras de licenciatura, maestría y doctorado. En el área de maestrías ofrece las carreras de Maestría en Ingeniería Electrónica, Maestría en Ingeniería Industrial, Maestría en Ciencias en Ingeniería Química, Maestría en Ingeniería Administrativa y la Maestría en Sistemas Computacionales. Asimismo, ofrece un Doctorado en Ciencias de la Ingeniería.

## 1.3 Planteamiento del problema

De acuerdo al Instituto Nacional de Cardiología, cuatro de cada diez adultos mexicanos tienen colesterol alto. El colesterol es una sustancia cerosa que se transporta en la sangre, adherido a las proteínas. Esta combinación de proteínas y colesterol se llama lipoproteína. Existen

diferentes tipos de colesterol, que dependen de qué transporta la lipoproteína. Ellos son: 1) Lipoproteína de baja densidad (LDL). El colesterol LDL, o “malo”, transporta las partículas de colesterol por todo el cuerpo. El colesterol LDL se acumula en las paredes de las arterias, lo que genera su endurecimiento y estrechamiento y, 2) Lipoproteínas de alta densidad (HDL). El colesterol HDL, o “bueno”, recoge el exceso de colesterol y lo lleva de nuevo al hígado. El cuerpo humano necesita colesterol para formar células sanas, pero tener altos niveles de colesterol de tipo LDL aumenta el riesgo de sufrir una enfermedad cardíaca. Con niveles altos de este tipo de colesterol, es posible que se formen depósitos grasos en los vasos sanguíneos. Con el tiempo, estos depósitos crecen y hacen que sea más difícil que fluya suficiente sangre a través de las arterias. A veces, esos depósitos consiguen romperse de repente y formar un coágulo que causa un ataque cardíaco o un accidente cerebrovascular.

Para este problema el trabajo presenta un enfoque con *Big Data* y técnicas de simulación mesoscópica con un método de Dinámica de Partículas DPD para un sistema para la predicción de concentración de colesterol en la sangre, para concientizar a la población sobre el cuidado y prevención de enfermedades cardiovasculares.

## **1.4 Objetivo general y específico**

### **1.4.1 Objetivo General**

Desarrollar un sistema de software para obtener la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica a través del modelo de Dinámica de Partículas Disipativas (DPD).

### **1.4.2 Objetivos específicos**

1. Analizar trabajos relacionados para identificar diferencias entre iniciativas existentes y el tema de tesis propuesto, y establecer la(s) principal(es) aportaciones que tendrá el proyecto.
2. Definir el perfil de usuario para identificar los niveles de saturación de colesterol en la sangre.

3. Diseñar el modelo Big Data para llevar cabo la persistencia y analítica de datos.
4. Implementar el modelo Big Data para el almacenamiento y analítica de datos.
5. Implementar el modelo de Dinámica de Partículas Disipativas para la simulación de los niveles de saturación de colesterol en la sangre.
6. Desarrollar el módulo de predicción y visualización de niveles de saturación de colesterol en la sangre.
7. Realizar las pruebas de concepto para validar el sistema desarrollado a través de al menos un caso de estudio que incluya la predicción y simulación de niveles de saturación en la sangre.

### 1.5 Justificación

El colesterol alto alcanza a heredarse, aunque suele ser el resultado de la elección de un estilo de vida poco o nada saludable, por lo que es posible prevenir y tratar. Llevar una dieta saludable, hacer ejercicio con regularidad y, a veces, tomar medicamentos ayudan a reducir el colesterol alto.

El *Big Data* ofrece alternativas para llevar a cabo el análisis de grandes volúmenes de datos, por ejemplo: 1) Cómputo de alto rendimiento (HPC) para ejecutar aplicaciones avanzadas de manera rápida, eficiente y confiable; 2) Empleo de súper-cómputo que emplea coprocesadores y aceleradores; 3) Los recursos de HPC en la nube que son cada vez más accesibles siendo para los consumidores un servicio y, 4) El análisis de datos se realiza por un analista específico o un grupo de especialistas. En el dominio médico, Big Data ayuda a los especialistas a mantener una interrelación comunicativa, inteligente y bien orientada a cumplir objetivos concretos eliminando información irrelevante para el tratamiento de enfermedades.

Por otra parte, la simulación mesoscópica, que se encuentra situada entre la escala atomística y macroscópica, es la Dinámica de Partículas Disipativas (DPD), la cual es capaz de capturar procesos que se realizan en los microsegundos, por lo que permite analizar las corrientes de moléculas y sus movimientos basados en capacidades y funciones de velocidad-densidad predefinidas. Las moléculas individuales consiguen ser rastreadas para identificar aglomeraciones o cúmulos de las mismas que ocasionen algún efecto dependiendo del dominio de aplicación.

Esta propuesta propone un sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación mesoscópica con la finalidad de identificar posibles afectaciones a la salud.

Entre los beneficios esperados se encuentran:

- Mostrar mediante simulación computacional los niveles de saturación de colesterol en la sangre
- Apoyar a las personas con la prevención de ataques cardíacos o accidentes cerebrovasculares



## Capítulo 2 Estado de la práctica

En la actualidad existen modelos para simular fluidos y estudiar su comportamiento cuando se someten a ciertos criterios. En este capítulo la primera sección muestra un análisis de trabajos relacionados con la simulación de fluidos, con diferentes métodos, diferentes arquitecturas o con diferentes tecnologías, algunos trabajos se dirigieron al campo médico. En la segunda sección del capítulo se encuentran análisis de trabajos que utilizan las técnicas de la minería de datos para la prevención de enfermedades.

### 2.1 Trabajos relacionados

En la presente sección se describe brevemente el análisis de un conjunto iniciativas literarias relacionadas con el anteproyecto de tesis. Las iniciativas analizadas están divididas o clasificadas en dos grupos: 1) Modelo para simulación de fluidos y, 2) Aplicación de Minería de Datos para la prevención de enfermedades cardiovasculares.

#### 2.1.1 Modelos para simulación de fluidos

A continuación, se presenta el análisis de trabajos relacionados con la simulación de fluidos.

En [25] para la búsqueda de nuevos métodos para eliminar el exceso de moléculas de colesterol, se utilizó el método de simulación de dinámica molecular DM con el cual se estudió el sistema compuesto por moléculas  $\beta$ -ciclodextrina y grupos de colesterol, tanto independientes como en un entorno acuático. Las simulaciones de DM fueron realizadas con el software NAMD (*Not (just) Another Molecular Dynamics program*, No es otro programa de dinámica molecular), el integrador utilizado fue el estándar (algoritmo *Brünger-Brooks-Karplus*) y visualizado con el programa de modelado molecular VMD para configurar la simulación. Para las simulaciones se utilizó el modelo TIP3P CHARMM (*Chemistry at Harvard Macromolecular Mechanics*, conjuntos de campos de fuerza ampliamente utilizado para la dinámica molecular en la universidad de *Harvard*), ajustado ya que el modelo de agua aplicado no reproduce correctamente la difusión traslacional del agua. Los autores realizaron una serie de simulaciones

DM b-ciclodextrina-colesterol en presencia y ausencia de agua. La influencia del agua es bastante significativa, ya que con agua las moléculas del colesterol se agrupan. Los resultados de las simulaciones para la eliminación del exceso del colesterol en el cuerpo se dejaron a las consideraciones futuras sobre el uso potencial de b-ciclodextrina en el contexto de la lucha contra enfermedad de la aterosclerosis.

Se estudiaron las lipoproteínas de alta densidad HDL (*High density lipoprotein*, lipoproteínas de alta densidad comúnmente llamado colesterol bueno) en [26], ya que es un poderoso factor de riesgo de enfermedad cardiovascular.

Los autores analizaron 47 sujetos con diferentes niveles de HDL (bajos y altos) y estudiaron sus análisis con el lenguaje R para la computación estadística, para realizar las simulaciones utilizaron el paquete GROMACS (v. 4.0). Se utilizó el campo de fuerza de lípidos estándar MARTINI (modelo para simulaciones biomoleculares) para composición de los componentes de lípidos y apolipoproteínas. El resultado obtenido fue que el perfil molecular de las partículas de HDL se combinó con el modelado estructural dinámico, lo cual reveló marcadas diferencias en los perfiles lipídicos de HDL, así como características clínicas y bioquímicas relacionadas entre sujetos de HDL-C bajo y alto. El enfoque permitió demostrar que los cambios en la composición de los lípidos también inducen distribuciones espaciales específicas de lípidos dentro de las partículas de HDL.

En [27] se propuso un algoritmo de simulación eficiente basado en la dinámica de partículas disipativas (DPD), este método se utilizó para estudiar fenómenos electrohidrodinámicos en fluidos electrolíticos. El flujo de fluido es imitado con partículas DPD mientras que la evolución de la concentración de la especie iónica es descrita utilizando pseudopartículas brownianas. El algoritmo se desarrolló a partir del conjunto de ecuaciones electrocinéticas para líquidos electrolíticos, lo que proponen los autores está diseñado para optimizar la eficiencia computacional de simulaciones de electrolitos a altas concentraciones fisiológicas de sal en fluidos.

Los autores observaron que el enfoque de su algoritmo llamado Condiff-DPD no es restringido a soluciones de electrolitos, ya que se aplica la misma idea a otras simulaciones de flujo de fluidos a mesoescala donde la difusión de un componente minoritario es importante, por ejemplo, en microrreactores.

Terrón-Mejía et al. [28] estudió las propiedades interfaciales y estructurales de los fluidos confinados por paredes diferentes a escala mesoscópica utilizando simulaciones de DPD en el gran conjunto canónico. Se implementó toda la metodología en código de simulación llamado SIMES que está diseñado para estudiar sistemas complejos a escala mesoscópica utilizando una unidad de procesamiento gráfico (GPU), para fluidos confinados se utilizó DPD ya que en el modelo DPD, las partículas no representan moléculas sino más bien grupos de moléculas con límites suaves la cual realizaron un híbrido con la metodología Grand Canonical Monte Carlo (MC) para simular fluidos confinados por paredes, en la escala mesoscópica y las propiedades interfaciales para fluidos confinados. Para la simulación se analizaron cuatro modelos diferentes de fluidos confinados, a saber los confinados entre paredes lisas, paredes rugosas simétricas, paredes rugosas no simétricas, y una combinación entre una pared lisa y una pared rugosa simétrica, en la conclusión de los autores se describen los diferentes comportamientos de los fluidos confinados dependiendo de las paredes con las que se llevaron a cabo las simulaciones y se define que para determinar el valor de la tensión interfacial no es fundamental la geometría de las paredes, sino la intensidad de la interacción pared sólida-fluido.

En [29] se presentó un paquete de simulación de flujo de glóbulos rojos (RBC, *red blood cells*) con propiedad de transporte químico acelerado por GPU basado en una adaptación del método de dinámica de partículas disipativas de transporte (tDPD), utilizaron el tDPD porque permite capturar la cinética de reacción a nivel mesoscópico. Para la programación se utilizaron los lenguajes C / C ++, CUDA (*Compute Unified Device Architecture*, Arquitectura Unificada de Dispositivos de Cómputo) C / C ++ y MPI (*Message Passing Interface*, interfaz de paso de mensajes). El paquete de simulación procesa todas las cargas de trabajo computacionales en paralelo por GPU e incorpora programación de múltiples flujos y comunicaciones MPI sin bloqueo para mejorar la escalabilidad entre nodos. Los autores mencionaron que utilizaron GPU por la velocidad en el procesamiento a comparación de su contraparte la Unidad Central de

Procesamiento (CPU). El modelo de glóbulos rojos presentado por los autores recupera correctamente la viscosidad de la membrana celular, la elasticidad, la rigidez a la flexión y el transporte químico a través de la membrana.

La falta de modelos de simulaciones de flujo de fluidos de estructuras complejas RCB a través de vasos capilares complejos, llevó a Phan-Thien et al. [30] a la elaboración de un método para la simulación de los flujos de glóbulos rojos en diferentes tubos, los métodos utilizados fueron la dinámica de partículas disipativas suavizadas (SDPD) y la dinámica de partículas disipativas (DPD), con parámetros que tienen un significado físico específico, combinados con fluctuaciones térmicas en una simulación de mesoescala y el método de límite sumergido (IBM), un método preferido para manejar problemas de interacción fluido-estructura, este último se ha utilizado ampliamente para manejar la interacción fluido-RBC en simulaciones. El objetivo de los autores fue acoplar SDPD e IBM para llevar a cabo las simulaciones de RBC en problemas de flujo complejos. Los autores primero desarrollaron el modelo SDPD-IBM y demostraron con simulaciones CFD (*Computational Fluid Dynamics*, fluidodinámica computacional) la capacidad del método SDPD-IBM simulando los flujos de glóbulos rojos en tubos rectangulares, cilíndricos, curvos, bifurcados y constreñidos.

Pourmousa and Pastor [31] proporcionaron una descripción general de las simulaciones DM de nanodiscos lipídicos nHDL, para la simulación hubo una solvatación de todos los nanodiscos con moléculas de agua y NaCl 0,15 M en cajas cúbicas de 15-16 nm de longitud lateral utilizando la interfaz web CHARMM-GUI (*Effective Simulation Input Generator and More*).

Los autores concluyen después de comparar simulaciones nHDL que comprender la maduración de nHDL es desafiante para estudios futuros, porque no solo requiere una gran cantidad de recursos computacionales, sino que también depende de estructuras de alta resolución de otras proteínas involucradas, la apolipoproteína APOA1 primero interactúa con el transportador ATP *Binding Cassette* A1 (ABCA1) para adquirir lípidos. Poco se sabe sobre los detalles de esta primera etapa, ya que las imágenes de una sola molécula revelaron que una forma dimérica (proteína formada por dos cadenas de aminoácidos) la interacción de ABCA1 con APOA1.

En [32] presentaron un paquete acelerado por GPU para la simulación de flujo en redes de nano a microporos con un modelo de mesoescala de dinámica de partículas disipativas de muchos cuerpos (mDPD), ya que las simulaciones mesoscópicas del flujo de hidrocarburos en las lutitas son un desafío debido a los poros heterogéneos de las lutitas con tamaños que van desde unos pocos nanómetros a unos pocos micrómetros. Los autores utilizaron el método de dinámica de partículas disipativas de muchos cuerpos (mDPD), este es un modelo mesoscópico para fluidos y moléculas sólidas, la pared se modela con una condición de límite de no deslizamiento que evita que las partículas de fluido penetren en las paredes, también usaron el algoritmo de Velocity-Verlet, los lenguajes de programación utilizados fueron CUDA C / C ++ con MPI y OpenMP. Los autores mencionaron que una simulación en CPU toma 15 veces más tiempo que utilizando la programación en paralelismo con GPU, el uso de GPU reduce efectivamente la sobrecarga en la comunicación entre *rank* / nodos

Gracias a la angioscopía no obstructiva con el dispositivo NOGA, Kojima et al. [33] tuvieron como objetivo estudiar la placa rota aórtica RP derivada de NOGA y la distribución estereográfica el incremento regional en la tensión del corte (WSS, *Wall shear stress*, esfuerzo cortante en la pared ) mediante el modelado de dinámica de fluidos computacional (CFD), para el cual se utilizó la aplicación PHOENICS-CFD Works, este modelo se utilizó para calcular la distribución de la velocidad del flujo sanguíneo y revelar la distribución tridimensional de WSS dentro del lumen del arco aórtico. Los autores estudiaron a 45 pacientes a los que se les realizó tomografía computarizada tridimensional 3D-CT antes de la coronariografía. El WSS en el arco aórtico lo midieron mediante análisis CFD basado en el método de elementos finitos usando condiciones uniformes de flujo de entrada y salida. La RP aórtica fue detectada por NOGA. Estos llegaron a la conclusión que la RP aórtica detectada por NOGA (Inyección transendocárdica con un sistema de mapeo electromecánico) se asoció fuertemente con un WSS máximo más alto en el arco aórtico derivado por CFD usando 3D-CT. El valor máximo de WSS tiene un papel importante en el mecanismo no solo de la aterosclerosis aórtica, sino también de la RP aórtica.

### **2.1.2 Aplicación de Minería de Datos para la prevención de enfermedades cardiovasculares**

Se prosigue con los trabajos que utilizan las técnicas de la minería de datos para la prevención de enfermedades.

Liu et al. [34] investigaron la asociación que existe entre la capacidad de salida de colesterol y el riesgo de mortalidad que presenta el paciente que tiene un problema cardiovascular (enfermedad arterial coronaria). Estudiaron a 1737 pacientes con enfermedad de las arterias coronarias, los autores encontraron que la capacidad de salida del colesterol no se explica simplemente por los niveles de colesterol HDL circulante o apolipoproteína AI, también se relaciona independientemente con la presencia y la extensión de la aterosclerosis, ellos concluyen sugiriendo que la capacidad de salida de colesterol es un predictor de mortalidad por todos problemas cardiovasculares tanto en el síndrome coronario agudo (SCA) como en cetoacidosis diabética (CAD).

En [35] realizaron un metanálisis de datos de pacientes sin antecedentes en enfermedades cardiovasculares, utilizaron dos métodos de complejidad creciente para modelar mediciones repetidas: valores medios acumulativos e intersecciones y pendientes específicas de cada individuo a partir de modelos de regresión lineal de efectos mixtos. El objetivo fue cuantificar el cambio en la discriminación y estratificación del riesgo de las personas de acuerdo con su riesgo de ECV previsto a 5 años cuando la información de las mediciones repetidas de los predictores de riesgo se agregó a la evaluación de las mediciones únicas de los niveles de factores de riesgo utilizados en las puntuaciones de riesgo estándar. Utilizaron 38 estudios con factores de riesgo emergentes. Los eventos identificados del riesgo de enfermedades cardiovasculares fueron infarto de miocardio no fatal o cualquier accidente cerebrovascular, en los modelos se utilizaron diferentes medidas de HDL, presión arterial, entre otras. Los autores concluyen que si se incorpora una media de una medición repetida post-basal de la presión arterial sistólica, es posible que el colesterol total y el colesterol HDL en los modelos de predicción del riesgo de enfermedades cardiovasculares den como resultado ligeras mejoras en la discriminación y reclasificación del riesgo.

En la industria de la salud, se recopilan grandes cantidades de datos que contienen información no observadas a simples vistas que son útiles para una toma de decisiones para la detección temprana de enfermedades, es por eso que Singh et al. [36] utilizaron técnicas avanzadas de minería de datos para desarrollar un sistema eficaz de predicción de enfermedades cardíacas utilizando una red neuronal. El sistema de los autores utilizó quince parámetros médicos para la predicción (edad, sexo, presión arterial, colesterol y obesidad). La herramienta que utilizaron fue Weka 3.6.11 con un conjunto de datos de 303 registros en donde el 40 % sirvió como conjunto de entrenamiento y el 60% como conjunto de prueba. Los autores demostraron que su modelo tiene los mejores resultados y ayuda a los expertos del campo médico a planificar un mejor diagnóstico a tiempo para el paciente, ya que el sistema funcionó de manera realista y en los resultados mostró que predice enfermedades cardíacas con una precisión del 100% mediante el uso de redes neuronales.

La enfermedad cardíaca según es un tipo de dolencia que está directamente relacionada con el corazón humano y con la circulación sanguínea por ese motivo el objetivo de [37] fue recopilar información de los seis parámetros que son la edad, el dolor de pecho, el electrocardiograma, la presión arterial sistólica, el azúcar en sangre en ayunas y el colesterol sérico para detectar enfermedades cardiovasculares. El modelo que propusieron para el diagnóstico de enfermedades cardiovasculares (DCD) mediante el sistema de inferencia difusa Mamdani (DCD-MFIS) mostro una precisión del 87,05%, también propusieron el modelo para el diagnóstico de enfermedades cardiovasculares incluye lógica difusa utilizando el sistema de inferencia difusa Mamdani (DCD-MFIS), la red neuronal artificial (DCD-ANN) y el enfoque de aprendizaje automático extremo profundo (DCD-DEML), estos lograron una mayor precisión y exactitud. El aprendizaje automático extremo profundo de DCD que propusieron alcanza una mayor precisión con las soluciones propuestas anteriormente que son del 92,45%. Para verificar el desempeño calcularon muchos indicadores que determinan el desempeño preciso. El entrenamiento de las redes neuronales se realiza utilizando las capas de 10 a 20 neuronas que denotan la capa oculta. DEML revela e indica una capa oculta que contiene 10 neuronas, que muestra el mejor resultado.

En [38] se presentó un sistema de eficiente de predicción de enfermedades cardíacas, utilizaron bases de datos con la información clínica de pacientes cardiacos y utilizaron la minería de datos para procesar el conjunto de datos obtenido con valores de entrada como la edad, sexo, dolor de pecho, presión arterial en reposo, colesterol sérico, azúcar en sangre en ayuno entre otros. Los autores usaron la herramienta Weka y KEEL (Extracción de conocimientos basada en el aprendizaje evolutivo), ya que es un conjunto de herramientas con lenguaje Java de código abierto (GPLv3) para implementar procesos de desarrollo para problemas de minería de datos, y el algoritmo C4.5. Al probar el sistema obtuvo 86.3% para la predicción en la fase de prueba, y en la fase de formación obtuvo una predicción del 87.3%, esta herramienta predice la tempranamente las enfermedades cardiacas.

La causa principal de enfermedad cardíaca es la presión arterial alta y colesterol alto. En [39] los autores utilizan el ojo humano para detectar niveles altos de colesterol en la sangre, ya que cuando se forman placas gruesas en las arterias, se observa una oclusión en el flujo sanguíneo de la vena retiniana. Los autores desarrollaron una aplicación móvil la cual utiliza un algoritmo, emplea *Machine Learning* y datos como altura, peso, edad, sexo, hábitos alimenticios, captura de foto del iris, entre otros. El resultado que presenta la aplicación para el usuario si es que tiene colesterol en la sangre, es un “No” o un “SI”. Como servidor los autores utilizan Heroku y para el procesamiento de imágenes utiliza módulos de Python. Los desafíos que encontraron los autores, es que las imágenes varían de personas a personas, las condiciones de la cámara de los *Smartphone* son diferentes y la situación de iluminación ya que la influencia del reflejo de luz afecta al diagnóstico.



## 2.2 Análisis comparativo

Las tablas 2.1 y 2.2 muestran un análisis comparativo, para observar diferencias y similitudes entre ellos, además de las metodologías utilizadas por los autores para resolver las diferentes problemáticas.

**Tabla 2.1 Comparación de trabajos relacionados con simulación**

Autor	Enfoque / Problema	Aportación / Solución	Modelos	Metodología	Arquitectura / Software para simulación	Estatus
Makieła <i>et al.</i> [25]	En la búsqueda de nuevos métodos para eliminar el exceso de moléculas de colesterol.	Se estudió el sistema compuesto por moléculas b –ciclodextrina y grupos de colesterol.	Simulación de moléculas b – ciclodextrina y grupos de colesterol.	TIP3P CHARMM.	Software NAMD.	Finalizado.
Yetukuri <i>et al.</i> [26]	Desconocimiento de la composición molecular de lípidos de las partículas de HDL en sujetos con niveles altos y bajos.	Simulación de HDL con estudios de 47 sujetos con colesterol alto y bajo.	Simulación de Dinámica Molecular a gran escala.	El campo de fuerza de lípidos estándar MARTINI.	Paquete de simulación GROMACS (v. 4.0).	Finalizado.

Autor	Enfoque / Problema	Aportación / Solución	Modelos	Metodología	Arquitectura / Software para simulación	Estatus
Medina <i>et al.</i> [27]	Propuso un algoritmo para el estudio de fenómenos electrohidrodinámicos en electrolitofluidos.	Método de simulación diseñado especialmente para sistemas con altas concentraciones de sal.	Algoritmo de DPD para electrolitos.	DPD.	No utiliza.	Finalizado, con configuraciones adicionales.
Terrón-Mejía <i>et al.</i> [28]	Propone una metodología DPD.	Método para la simulación de fluidos confinados con paredes rugosas.	Simulación de fluidos confinados.	DPD y MC.	GPU.	Finalizado.
Blumers <i>et al.</i> [29].	Simulación basada en partículas de una suspensión de glóbulos rojos con propiedad de transporte químico.	Paquete de simulación de RBC (recuento de glóbulos rojos) acelerado por GPU.	Simulación de transporte con glóbulos rojos.	Método de dinámica de partículas disipativas de transporte (tDPD).	GPU.	Finalizado.

Autor	Enfoque / Problema	Aportación / Solución	Modelos	Metodología	Arquitectura / Software para simulación	Estatus
Phan-Thien <i>et al.</i> [30]	Propone un método numérico híbrido con el método de límite sumergido y método de partícula dinámica disipativa suavizada, para simular el RBC en geometrías de flujo complejas.	Método numérico híbrido con el método de límite sumergido y método de partícula dinámica disipativa suavizada.	Modelo numérico.	Dinámica de partículas disipativas suavizadas (SDPD).	CFD ( <i>Computational Fluid Dynamics</i> , fluidodinámica).	Finalizado.
Pourmousa and Pastor [31]	El artículo revisa estudios previos de simulación de dinámica molecular (DM) de nanodiscos e ilustra problemas de convergencia y precisión utilizando resultados de nuevas simulaciones MD atomísticas de varios microsegundos.	Comparación de modelos para la simulación de nanodiscos HDL.	Simulación de dinámica molecular.	Dinámica Molecular con el intercambio de réplicas y Monte Carlo.	CHARMM-GUI.	Finalizado.
Xia <i>et al.</i> [32]	Simulación basada en partículas de flujo multifásico e interacción fluido-sólido en redes de poros de nano a microescala de geometrías de poro arbitrarias.	Paquete de simulación de flujo de poros mesoscópico acelerado por GPU basado en un modelo de dinámica de partículas disipativas de muchos cuerpos (mDPD).	Modelo mesoscópico para moléculas sólidas y fluidos.	Modelo de dinámica de partículas disipativas de muchos cuerpos (mDPD).	GPU.	Finalizado.

Autor	Enfoque / Problema	Aportación / Solución	Modelos	Metodología	Arquitectura / Software para simulación	Estatus
Kojima <i>et al.</i> [33]	Estudiaron RP aórtica derivada de NOGA y WSS mediante el modelo CFD.	Estudio en donde el valor máximo de WSS tiene un papel importante en el mecanismo no solo de la aterosclerosis aórtica, sino también de la RP aórtica.	Modelado de dinámica de fluidos computacional.	Dinámica de fluidos.	Aplicación PHOENICS-CFD Works.	Finalizado.

**Tabla 2.2 Análisis comparativo de artículos relacionados con Minería de Datos**

Autor	Enfoque / Problema	Aportación / Solución	Métodos / herramientas utilizadas	Estado
Liu <i>et al.</i> [34]	Investiga la asociación entre la capacidad de salida de colesterol y la mortalidad por cualquier causa y cardiovascular en pacientes con enfermedad arterial coronaria.	La capacidad de salida de colesterol sirve como una medida independiente para predecir la mortalidad cardiovascular.	Estadística.	Finalizado.
Paige <i>et al.</i> [35]	Falta de evaluación de mediciones repetidas de la presión arterial y el colesterol para predecir el riesgo de enfermedades cardiovasculares.	Estudio estadístico.	Regresión Lineal.	Finalizado.
Singh <i>et al.</i> [36]	Existe recopilación de datos que contienen información útil para predecir el nivel de riesgo de enfermedad cardíaca.	Desarrollo de algoritmo para un sistema eficaz de predicción de enfermedades cardíacas (EHDPS) utilizando una red neuronal.	Técnicas de minería de datos (Redes neuronales).	Finalizado.
Siddiqui <i>et al.</i> [37]	Recopila información de seis parámetros que son la edad, el dolor de pecho, el electrocardiograma, la presión arterial sistólica, el azúcar en sangre en ayunas y el colesterol sérico que utiliza Mamdani Fuzzy Expert para detectar enfermedades cardiovasculares.	Redes neuronales artificiales y enfoques de aprendizaje automático extremo profundo para el diagnóstico automatizado de enfermedades cardiovasculares.	Técnicas de minería de datos (Redes neuronales).	Finalizado.

Autor	Enfoque / Problema	Aportación / Solución	Métodos / herramientas utilizadas	Estado
Purushottam <i>et al.</i> [38]	Propone un sistema eficiente para la predicción de enfermedades cardíacas con un conjunto de datos proporcionados por la información clínica de pacientes cardíacos.	Sistema de eficiente de predicción de enfermedades cardíacas.	Técnicas de minería de datos (Ábol de decisión C4.5) con la herramienta Weka y Keel.	Finalizado.
Alhasawi <i>et al.</i> [39]	Propone una aplicación móvil que detecte el colesterol en la sangre con fotografía de iris y con otros datos sobre el usuario.	Aplicación móvil para la detección del colesterol en la sangre.	<i>Machine Learning</i> y para procesar imágenes con servidor Heroku con módulos de Python.	En desarrollo.

Por consiguiente, después de efectuar un análisis de los artículos abordados en este capítulo, se encontró que los trabajos [28] [29] y [32] utilizan la arquitectura GPU, la cual al utilizarla tiene ventajas como por ejemplo el tiempo, ya que es 15 veces más rápido que utilizar una arquitectura CPU, es por ese motivo para este trabajo se eligió la simulación aceleradas por GPU, en donde permite el procesamiento paralelo entre una o más GPU y CPU. En [27] y [28] utilizan el método DPD, también son utilizados métodos variantes de DPD en [29], [30] y [32] para la simulación de fluidos simples y complejos, estos trabajos permiten la justificación del uso de la DPD en este trabajo en el cual se simulará diferentes niveles de colesterol en la sangre, porque es el método que mejor describe a los fluidos complejos, ya que se tiene un mejor control de las propiedades de transporte. En [25] y [26] se simuló el colesterol con diferentes métodos y softwares de simulación, se encontraron trabajos sobre la simulación de glóbulos rojos como en [29] y [30], a diferencia de los trabajos anteriores se simulara diferentes niveles de colesterol en la sangre.

Para el campo médico también se encontró que en [33] se simula el flujo de la aorta, para una angioscopia, que utiliza datos de estudios médicos.

Los trabajos [35] [36] [37] y [38] utilizaron las técnicas de minería de datos para detectar y predecir enfermedades cardiovasculares, además en [34] [35] [36] [37] y [38] utilizaron el colesterol como predictor de riesgo para padecimientos cardiacos, mientras que en [39] propone una aplicación para detectar el colesterol en la sangre utilizando Aprendizaje automático e imágenes de la retina, entre otros datos, a diferencia de este trabajo solo se predecirá el colesterol con datos proporcionados por el usuario. En este trabajo también se aplicaran técnicas de minería de datos para la predicción de los niveles de colesterol y el riesgo de padecer enfermedades cardiacas.

## Capítulo 3 Aplicación de la Metodología

Este capítulo tiene como objetivo presentar la metodología XP la cual fue utilizada para el desarrollo de este proyecto de tesis, de acuerdo con la problemática estudiada y al objetivo trazado para el desarrollo un Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD).

La metodología utilizada en este proyecto comienza con la fase de planificación en la cual se organizan las ideas, mediante pláticas entre las partes involucradas para detallar las funcionalidades específicas del software a desarrollar, en la fase de diseño se hace énfasis en los bocetos simples y claros que servirán de guía para la siguiente fase, el desarrollo o codificación se comienza a desarrollar el proyecto, mientras que en la última fase lo desarrollado se somete a una serie de pruebas unitarias continuas, con el objetivo de corregir fallas periódicamente.

A continuación, se presentan en la siguiente imagen las fases de la metodología utilizada en forma de esquema.

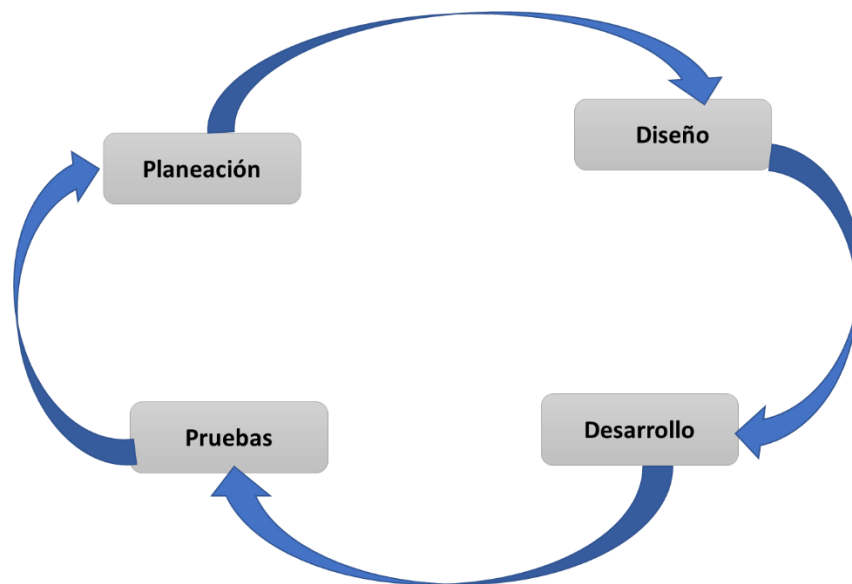


Figura 3.1 Fases de la metodología XP.



### 3.1 Planificación

En esta fase de la metodología se encuentran las historias de usuario, que permitieron el diseño de la aplicación. Las historias de usuario se recopilaron con reuniones con el director y codirector de este proyecto de Tesis. A continuación, se presenta en forma de tablas las historias de usuario.

**Tabla 3.1 Historia de usuario 1**

**Historia de usuario**

Numero: <b>1</b>	
Nombre Historia: <b>Registro de usuarios</b>	
Prioridad: <b>Alta</b> (Alta, Media, Baja)	Riesgo en Desarrollo: <b>Media</b> (Alta, Media, Baja)
Puntos Estimados: <b>1</b>	Iteración Asignada: <b>1</b>
Descripción: <b>Los usuarios podrán registrarse, llenando el formulario con sus datos personales y podrá crear su contraseña para su ingreso al sistema.</b>	
Observaciones:	

**Tabla 3.2 Historia de usuario 2**

**Historia de usuario**

Numero: <b>2</b>	
Nombre Historia: <b>Inicio de sesión</b>	
Prioridad: <b>Alta</b> (Alta, Media, Baja)	Riesgo en Desarrollo: <b>Media</b> (Alta, Media, Baja)
Puntos Estimados: <b>1</b>	Iteración Asignada: <b>1</b>
Descripción: <b>Los usuarios registrados podrán acceder a la aplicación web, con su nombre de usuario y contraseña</b>	
Observaciones: <b>Solo podrán iniciar sesión los usuarios registrados</b>	

**Tabla 3.3 Historia de usuario 3**

**Historia de usuario**

Numero: <b>3</b>	
Nombre Historia: <b>Perfil de usuario</b>	
Prioridad: <b>Alta</b> (Alta, Media, Baja)	Riesgo en Desarrollo: <b>Alta</b> (Alta, Media, Baja)
Puntos Estimados: <b>1</b>	Iteración Asignada: <b>1</b>
Descripción: <b>Una vez que inicie sesión el usuario podrá agregar sus análisis clínicos sobre sus niveles de colesterol</b>	
Observaciones:	

**Tabla 3.4 Historia de usuario 4**

**Historia de usuario**

Numero: <b>4</b>	
Nombre Historia: <b>Consulta de niveles de colesterol</b>	
Prioridad: <b>Alta</b> (Alta, Media, Baja)	Riesgo en Desarrollo: <b>Alta</b> (Alta, Media, Baja)
Puntos Estimados: <b>1</b>	Iteración Asignada: <b>1</b>
Descripción: <b>El usuario podrá consultar su nivel de colesterol actual y los niveles que podrá tener en el futuro en mg/dl con modelos en tercera dimensión.</b>	
Observaciones: <b>El usuario tiene que iniciar sesión primero.</b>	

**Tabla 3.5 Historia de usuario 5**

**Historia de usuario**

Numero: <b>5</b>	
Nombre Historia: <b>Obtención de Datos</b>	

Prioridad: <b>Alta</b> (Alta, Media, Baja)	Riesgo en Desarrollo: <b>Alta</b> (Alta, Media, Baja)
Puntos Estimados: <b>2</b>	Iteración Asignada: <b>1</b>
Descripción: <b>La obtención de datos será por medio de datos de análisis clínicos proporcionados por el perfil de usuario. Los datos serán almacenados en un repositorio.</b>	

Observaciones:

La obtención de datos será por medio de análisis clínicos que se encuentran en repositorios básicos, posteriormente los datos se utilizarán para la predicción del colesterol en diferentes lapsos de tiempo utilizando una regresión logística.

Para la regresión Logística se entrenará el modelo con los datos obtenidos de un *Dataset*, se elegirán las variables de colesterol contra edad, para obtener el modelo de predicción en diferentes lapsos de tiempo del colesterol de un usuario.

Para la representación del colesterol en la sangre, se tomó en cuenta la medida de colesterol que se utiliza en la mayoría de laboratorios, la cual es miligramos sobre un decilitro (mg/dl). Para simular la cantidad de sangre y colesterol de un usuario se presentará en un decilitro, en la cual en 55% de sangre pertenece al plasma y el 45% pertenece a la hemoglobina y otros componentes, en el plasma el 90% contienen agua y el 5% las demás proteínas, en donde se encuentra el colesterol. Debido a la capacidad de cómputo, no se podrá simular todas las moléculas que contiene un decilitro de sangre, ya que no es posible simular un millón de moléculas y en este caso la cantidad de moléculas superan los seis millones o bien se contempla simular las moléculas totales que contiene una muestra tomada a los pacientes para análisis de

colesterol, la cual es de 5 ml de sangre, muestra que posteriormente es sometida a una centrifugadora para separar el plasma de la hemoglobina para la realización de análisis clínico del colesterol; por lo tanto se utilizará una escala que representará los gramos en un decilitro, en donde una molécula representara las moléculas totales que contiene un gramo.

Para la simulación se generará un módulo el cual calculará posiciones de las moléculas y el tamaño de la caja que almacena dichas moléculas utilizando la Big Data para el almacenamiento utilizando archivos JSON.

El usuario podrá acceder a su perfil y predicción de colesterol a través de una aplicación web progresiva.

### **3.2 Diseño del sistema**

En esta sección se presenta la arquitectura del sistema que cumple con las necesidades que se requiere. En la Figura 3.2 se muestra la arquitectura a utilizar para el desarrollo de la aplicación.

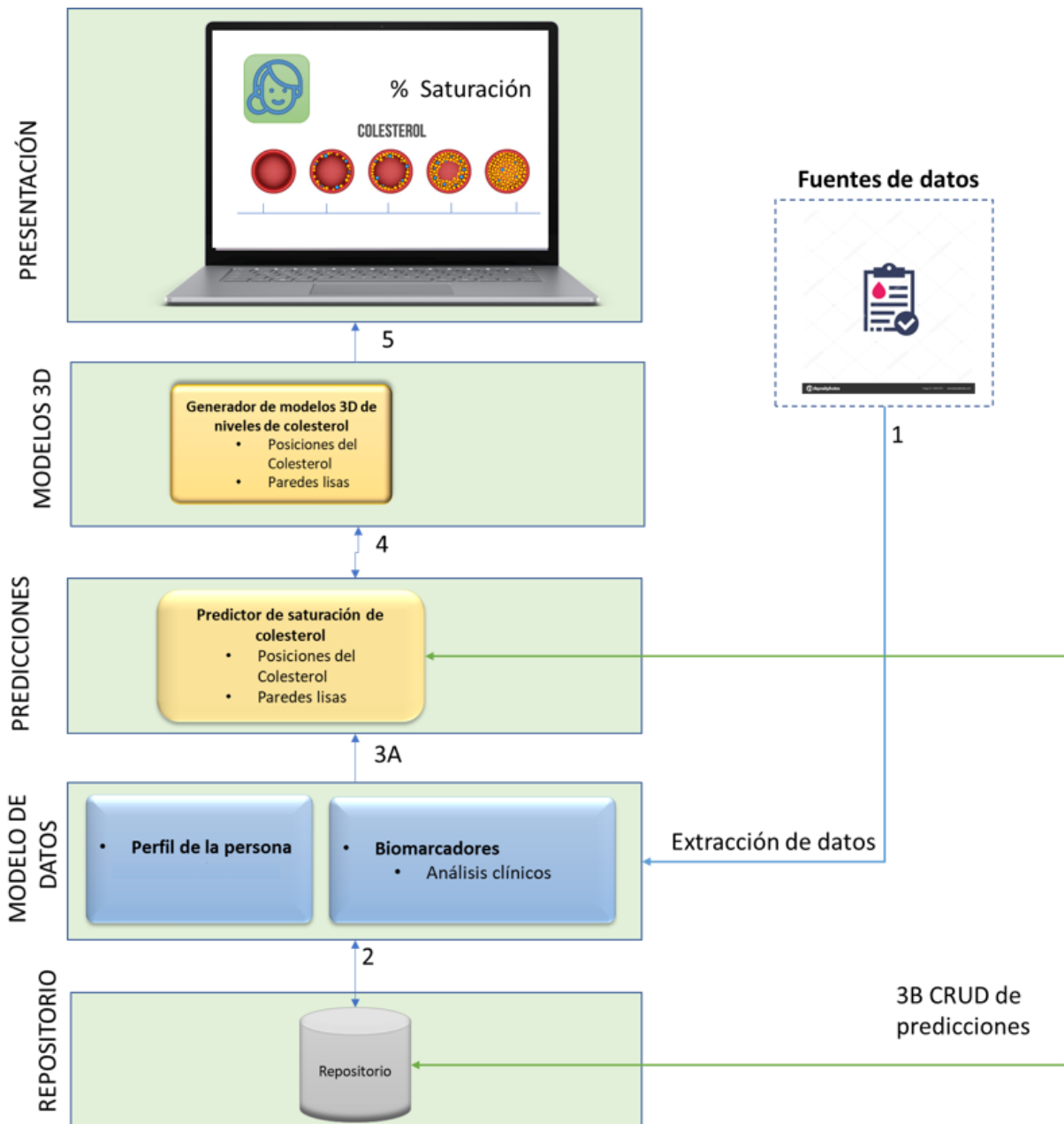


Figura 3.2 Arquitectura para el sistema.

### 3.2.1 Arquitectura del sistema

A continuación, se describen las capas, módulos y flujos que conforman la arquitectura del sistema.

### 3.2.1.1 Capas de la arquitectura

- **Presentación:** En esta capa se presenta la interfaz entre el usuario y el sistema, en donde se despliega la información del usuario y las predicciones de los niveles de colesterol.
- **Modelos 3D:** Esta capa contiene el componente generador de modelos en 3D de niveles de colesterol.
- **Predicciones:** Esta capa presenta un componente que calcula la predicción de colesterol.
- **Modelo de datos:** Esta contiene el componente del perfil de usuarios.
- **Repositorio:** El repositorio está basado en Mongo DB que contendrá todos los datos del perfil de usuarios.

### 3.2.1.2 Componentes de la arquitectura

**Fuentes de datos:** Este módulo se encarga de almacenar los análisis de los pacientes, en donde contiene los niveles de colesterol en la sangre.

**Repositorio:** El repositorio se encarga de guardar los datos del paciente, en un gestor de base de datos.

**Predicciones:** Este módulo recibe los datos del repositorio del perfil de cada usuario para predecir las posiciones de colesterol que tiene en la actualidad y que tendrá en lapsos de tiempo el usuario, también el módulo predecirá el grosor de las paredes en los diferentes lapsos de tiempo.

**Modelos 3D:** En este módulo se generan los modelos en tercera dimensión de los niveles de colesterol en los diferentes lapsos de tiempo con los datos del módulo predictor de saturación de colesterol.

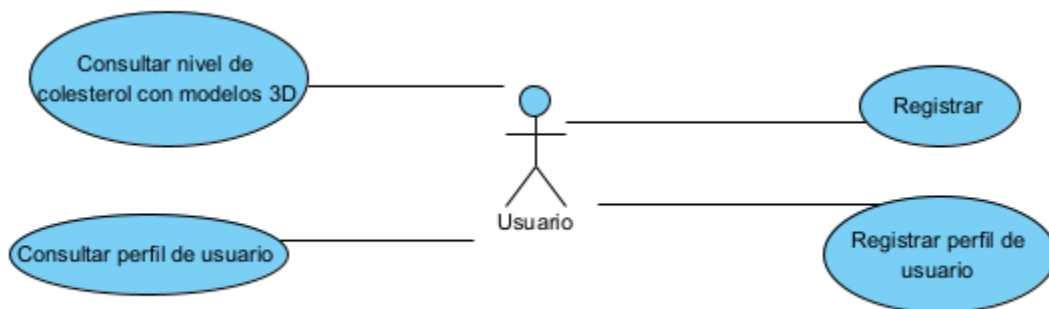
**Presentación:** En este módulo se muestran al usuario los modelos en 3D de la saturación de colesterol en los diferentes lapsos de tiempo simulados en el módulo generador de modelos 3D de niveles de colesterol, así como el perfil de usuario.

### 3.2.1.3 Flujo de Trabajo

1. A partir de las fuentes de datos (análisis clínicos), se extraerá la información y se filtrará para almacenar los datos en el repositorio.
2. A partir de la información almacenada en el repositorio se realizarán las predicciones de saturación de colesterol en la sangre en diferentes lapsos de tiempo y serán almacenados en el repositorio.
3. Las predicciones de los niveles de colesterol (3A) pasarán al componente que genera los las posiciones de las moléculas para los modelos en 3D de los diferentes lapsos de tiempo, estos modelos serán almacenados en el repositorio (3B).
4. A partir de las posiciones de las moléculas el componente encargado generará los modelos en tercera dimensión de la saturación del colesterol en los lapsos de tiempo calculados por el predictor.
5. El usuario podrá visualizar a través de la aplicación los diferentes niveles de saturación de colesterol en diferentes lapsos de tiempo en 3D.

### 3.2.2 Análisis de requerimientos

Los requerimientos identificados solo corresponden al actor usuario, el cual las tareas que realizará están representadas en el siguiente diagrama de caso.



**Figura 3.3 Diagrama de Caso de usos**

En la tabla 3.1 se encuentra la descripción del diagrama de la figura 3.1.

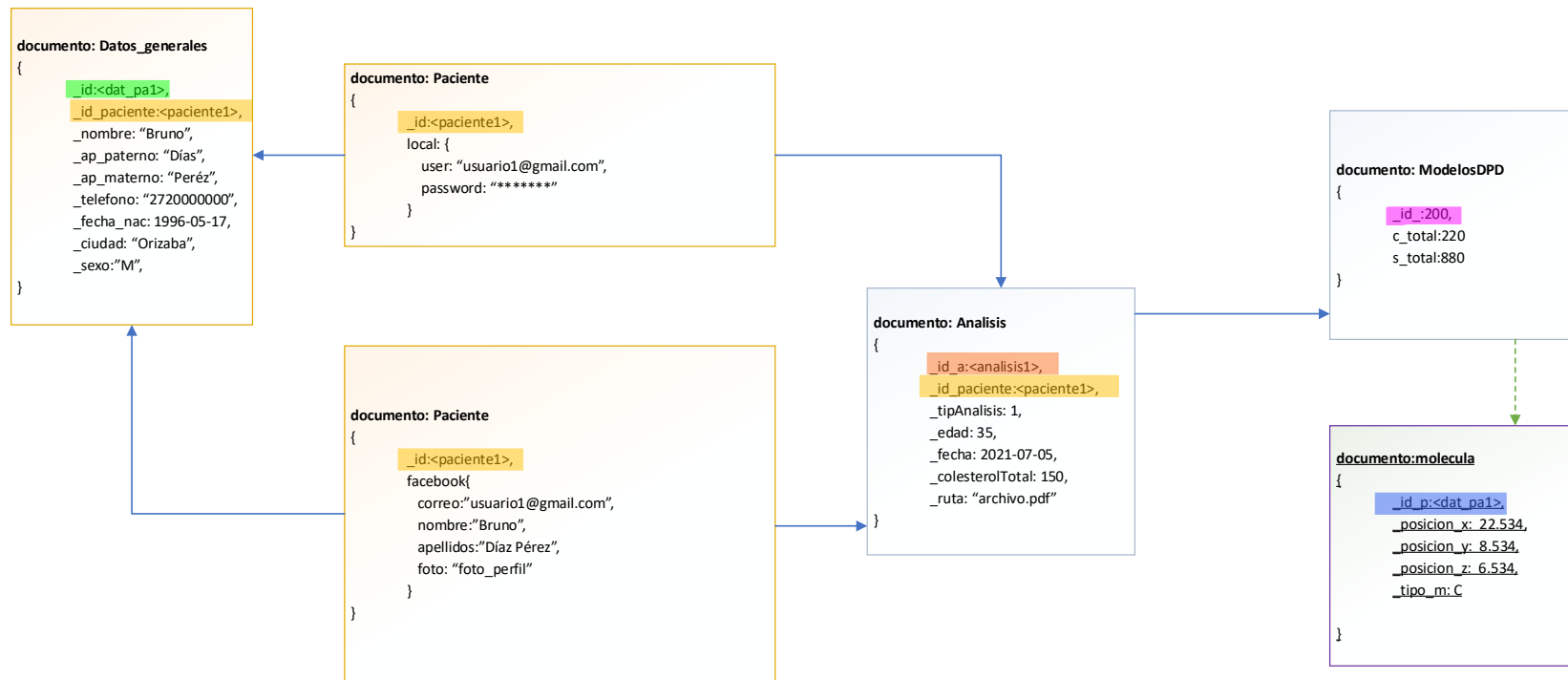
**Tabla 3.6 Descripción del diagrama de caso de uso de la figura 3.2.**

<i>Clave</i>	<i>Caso de Uso</i>	<i>Descripción</i>
CU01	Registrar	El usuario se registrará en el sistema.
CU02	Registrar perfil de usuario	El usuario puede registrar su perfil en el cual registrará su nivel de colesterol actual.
CU03	Consultar perfil de usuario	El usuario consultará su perfil en donde se encontrarán sus datos personales.
CU04	Consultar niveles de colesterol con modelos 3D	El usuario consultará los niveles de colesterol actual y que posiblemente tendrá en un futuro, si es alto el nivel de colesterol malo se mostrará las consecuencias que tendrá si no cambia el estilo de vida que lleva el usuario.

### 3.2.3 Modelo de Datos

En la figura 3.4 se muestra el modelo de datos, implementado en MongoDB para el almacenamiento de los datos del usuario y para el de la predicción.





**Figura 3.4 Modelo de datos.**

En el modelo de datos se observan ocho documentos, en la siguiente tabla se muestran los documentos del modelo con su descripción.

**Tabla 3.7 Descripción del modelo de datos de la figura 3.4.**

Documento	Descripción
<b>Paciente</b>	Este documento almacenará los datos del paciente para iniciar sesión.
<b>Datos_generales</b>	Este almacena los datos del paciente para complementar el registro como nombre completo, edad, entre otros.
<b>Analisis</b>	El documento almacena la cantidad de colesterol, así como la fecha del análisis, tipo del análisis que se realizó el paciente.
<b>ModeloDPD</b>	En este documento tiene almacenados los datos para la construcción de los modelos de DPD de la concentración de colesterol en la sangre.
<b>molecula</b>	En este documento se almacena las posiciones de las moléculas y su tipo al que pertenece.

A continuación, en las siguientes tablas se describen los datos y sus tipos que se almacenan en los documentos.

**Tabla 3.8 Descripción de los datos que contiene el documento Paciente de tipo Local del modelo de datos de la figura 3.4.**

Documento: Paciente			
Datos	Tipo	Descripción	Relación
_id	ObjectID	Este es el identificador del paciente.	No tiene.
_nom_usuario	String	Este dato almacena el nombre de usuario.	No tiene.
_password	String	En este dato se almacena la contraseña de usuario para iniciar sesión.	No tiene.

**Tabla 3.9 Descripción de los datos que contiene el documento Paciente tipo Facebook del modelo de datos de la figura 3.4.**

Documento: Paciente			
Datos	Tipo	Descripción	Relación
_id	ObjectID	Este es el identificador del paciente.	No tiene.
_correo	String	Este dato almacena el nombre de correo.	No tiene.
_nom_usuario	String	Este dato almacena el nombre de usuario.	No tiene.
_apellidos	String	En este dato almacena los apellidos del usuario.	No tiene.
_foto	String	En este dato almacena el enlace de la foto de perfil del usuario de la plataforma Facebook.	No tiene.

**Tabla 3.10 Descripción de los datos que contiene el documento Datos\_generales del modelo de datos de la figura 3.4.**

Documento: Datos_generales			
Datos	Tipo	Descripción	Relación
_id	ObjectID	Este dato es el identificador de Datos generales.	No tiene.
_id_paciente	Object	Este dato hace referencia al Paciente al que pertenecen esos datos.	Con identificador del documento Paciente.
_nombre	String	El dato guarda el nombre o los nombres (en caso de tener más de uno) del paciente.	No tiene.
_ap_paterno	String	Este dato guarda el apellido paterno del paciente.	No tiene.
_ap_materno	String	Este dato guarda el apellido materno del paciente.	No tiene.
_telefono	String	Este dato guarda el teléfono del paciente.	No tiene.
_fecha_nac	Date	El dato guarda la fecha de nacimiento del paciente.	No tiene.
_ciudad	String	Este guarda la ciudad de origen del paciente.	No tiene.
_sexo	String	Este dato contiene el sexo del paciente.	No tiene.
_edad	Integer	El contiene la edad del paciente.	No tiene.

**Tabla 3.11 Descripción de los datos que contiene el Análisis Paciente del modelo de datos de la figura 3.4.**

Documento: Análisis			
Datos	Tipo	Descripción	Relación
_id_a	ObjectID	Este dato es el identificador del documento Análisis.	No tiene.
_id_paciente	Object	Este dato contiene el id del paciente al que le corresponde el análisis.	Con identificador del documento Paciente.
_edad	Integer	Este dato indica la edad actual del paciente.	No tiene.
_tipAnálisis	Boolean	Este dato indica que tipo de análisis se realizó el paciente, ya que para medir los niveles de colesterol existen dos (perfil lipídico y bioquímica sanguínea).	No tiene.
_fecha	Date	Este indica la fecha en que se realizó el análisis el paciente.	No tiene.
_colesterolTotal	Integer	Este dato indica la cantidad de colesterol total del paciente.	No tiene.

**Tabla 3.12 Descripción de los datos que contiene el documento Modelo DPD de datos de la figura 3.4.**

Documento: Predicción			
Datos	Tipo	Descripción	Relación
_id	ObjectID	Este dato es el identificador del documento Predicción.	No tiene.
_c_total	Integer	Este dato contiene los niveles de colesterol a simular.	No tiene.
_s_total	Integer	Este dato contiene la cantidad total de moléculas a simular.	No tiene.

**Tabla 3.13 Descripción de los datos que contiene el documento molécula del modelo de datos de la figura 3.4.**

Documento: globulosSanguineos			
Datos	Tipo	Descripción	Relación
_id_p	ObjectID	Este dato es el identificador de la partícula.	No tiene.
_posicion_x	Double	El dato almacena la posición del eje X.	No tiene.
_posicion_y	Double	El dato almacena la posición del eje Y.	No tiene.
_posicion_z	Double	El dato almacena la posición del eje Z.	No tiene.
_tipo_m	String	Este dato almacena el tipo de molécula al que pertenece.	No tiene.

### 3.2.4 Modelado de funcionamiento

Para representar la secuencia de comunicación entre el usuario y la aplicación Web, se diseñaron los siguientes diagramas de actividades presentados a continuación.

En la figura 3.5 se presenta el diagrama de actividades de la predicción de concentración de colesterol en la sangre, para esta interacción el usuario debe estar registrado e iniciado sesión en el sistema.

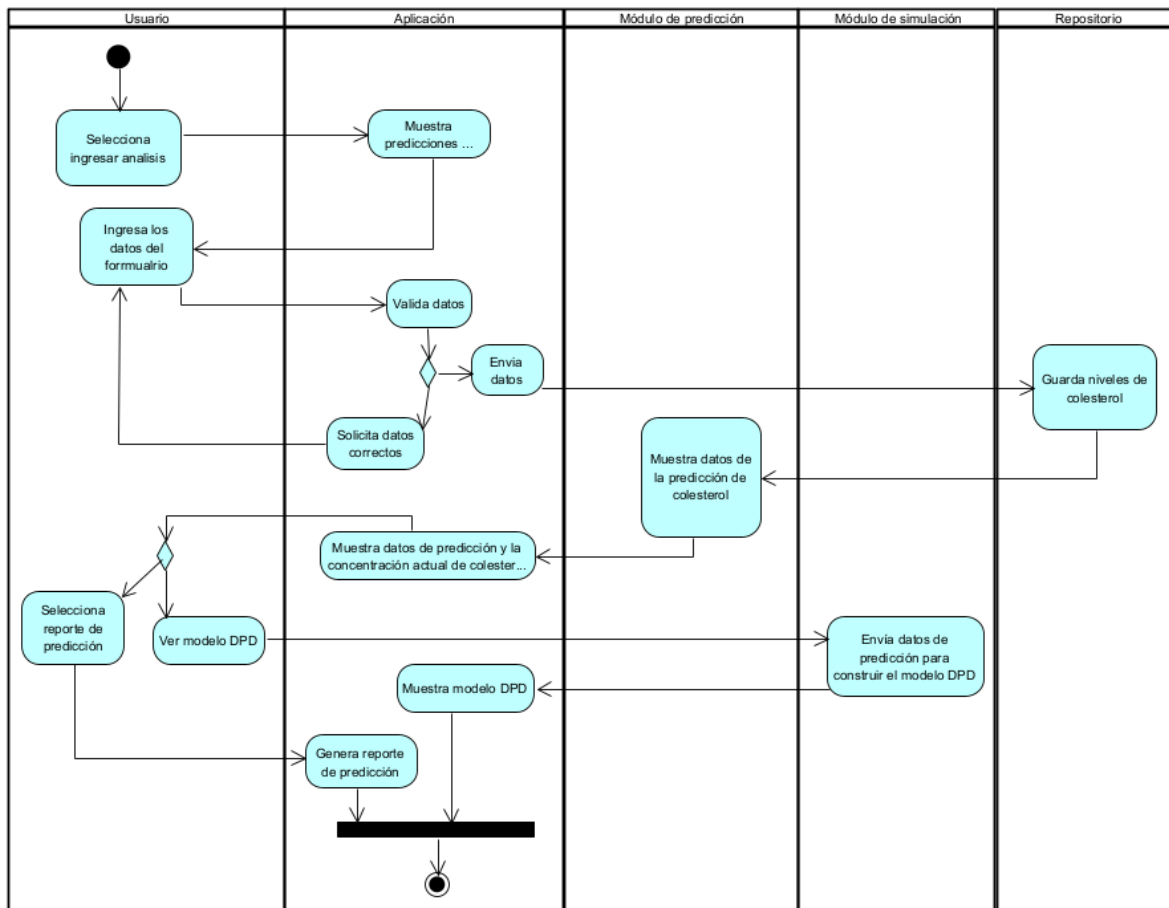
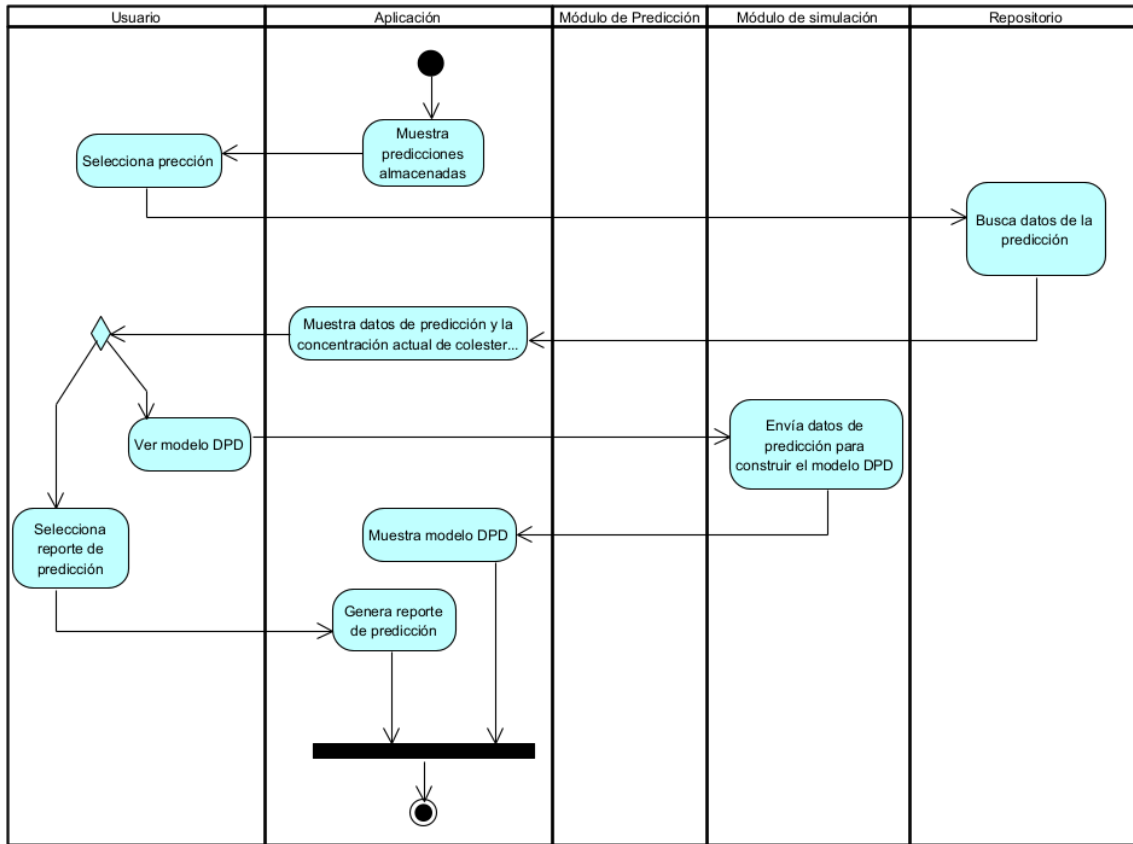


Figura 3.5 Diagrama de actividades de la predicción de concentración de colesterol en la sangre.

En la figura 3.6 se muestra el diagrama de actividad para acceder a predicciones ya registradas, de igual manera, para esta interacción el usuario debe estar registrado e iniciado sesión en el sistema.

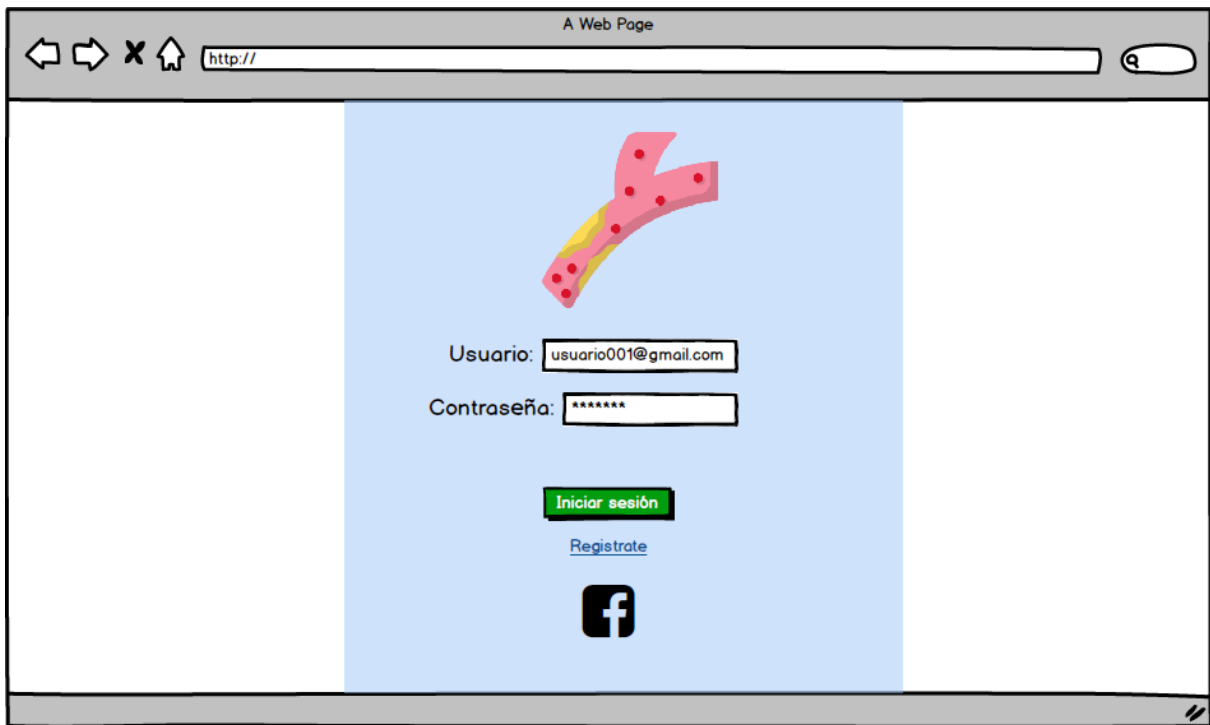


**Figura 3.6 Diagrama de actividades actividad para acceder a predicciones ya registradas.**

### 3.2.5 Diseño de la aplicación

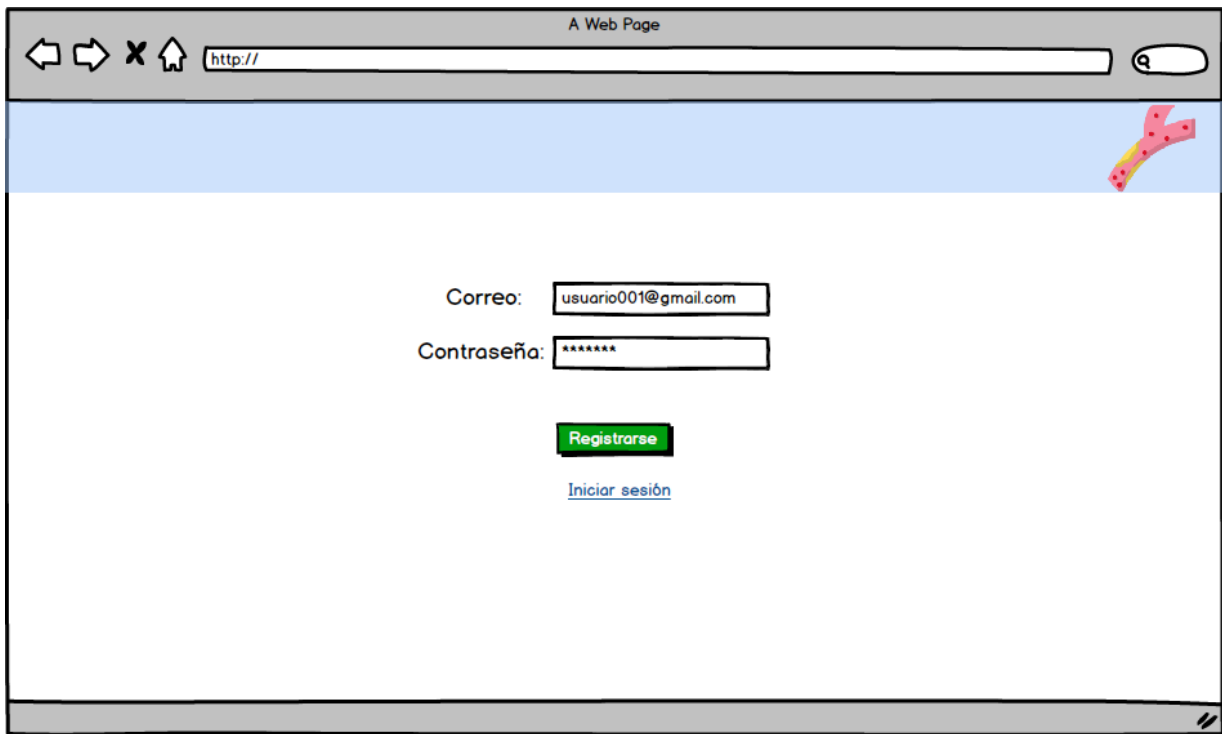
En esta sección se presentan los *Mockups* para las interfaces gráficas adquisición y visualización de resultados.

En la siguiente figura se muestra el *Mockup* para que inicie sesión el usuario, el cual podrá iniciar sesión por Facebook, en caso de no estar registrado y no tener cuenta en Facebook también se encuentra una opción para registrarse.



**Figura 3.7** *Mockup* para autenticación del usuario.

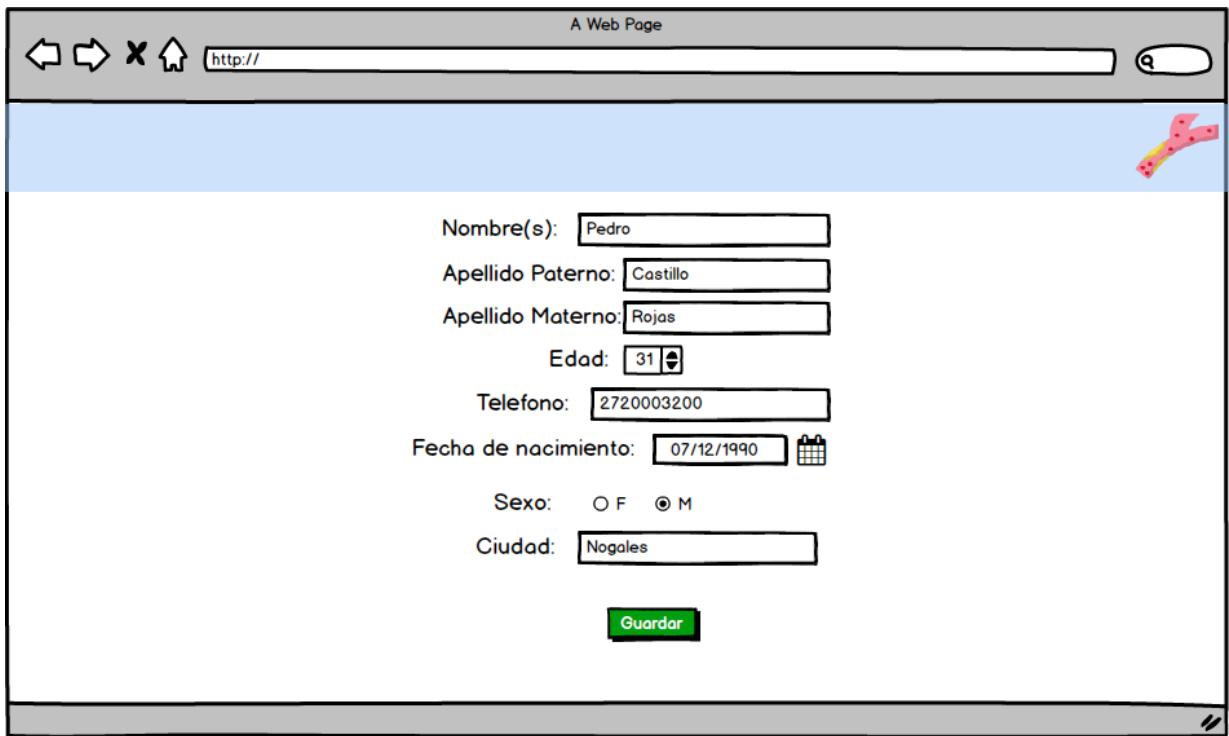
El siguiente *Mockup* es para el registro de los usuarios que no tengan cuenta en Facebook, en el cual deberán proporcionar su usuario y contraseña.



**Figura 3.8 Mockup Registro de usuarios.**

Después de que el registro del usuario sea satisfactorio, se mostrará un formulario para el llenado de datos generales del usuario, como se muestra en la figura 3.9.





A Web Page

http://


Nombre(s):

Apellido Paterno:

Apellido Materno:

Edad:

Telefono:

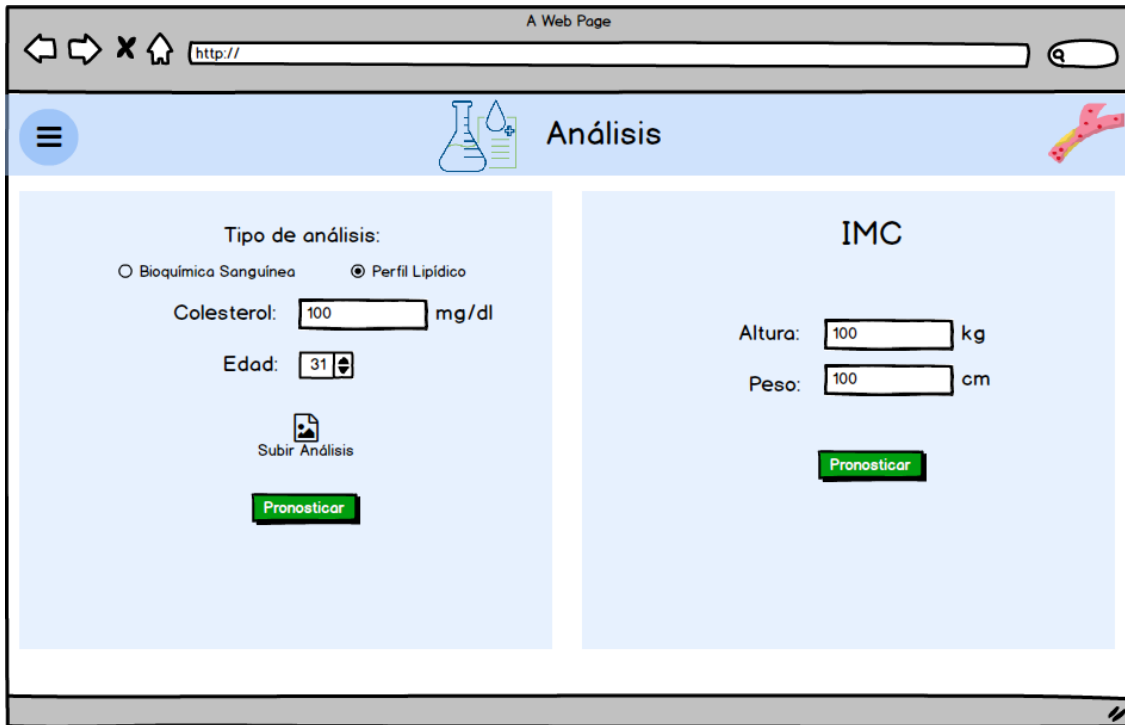
Fecha de nacimiento:  

Sexo:  F  M

Ciudad:

**Figura 3.9** *Mockup* de datos generales de usuario

Al iniciar sesión por primera vez el paciente subirá los datos de sus análisis como se muestra en el siguiente *Mockup* (Figura 3.10), en el cual se selecciona el tipo de análisis que se realizó, y después se eligen los niveles de colesterol que marca el análisis en mg/dl.



**Figura 3.10 Mockup Análisis.**

Para mostrar el resultado de la predicción al paciente se mostrará el colesterol total en el lapso de tiempo de 20 a 80 años como se muestra en la figura 3.11, la cual será visualizada con una gráfica y una tabla, el usuario podrá descargar un reporte de los niveles de colesterol en formato .pdf seleccionando el botón de descargar reporte, posteriormente podrá visualizar los diferentes niveles de colesterol correspondientes a la predicción, en un modelo 3D que representará la concentración de colesterol en la sangre en un decilitro, como se muestra en la figura 3.12.

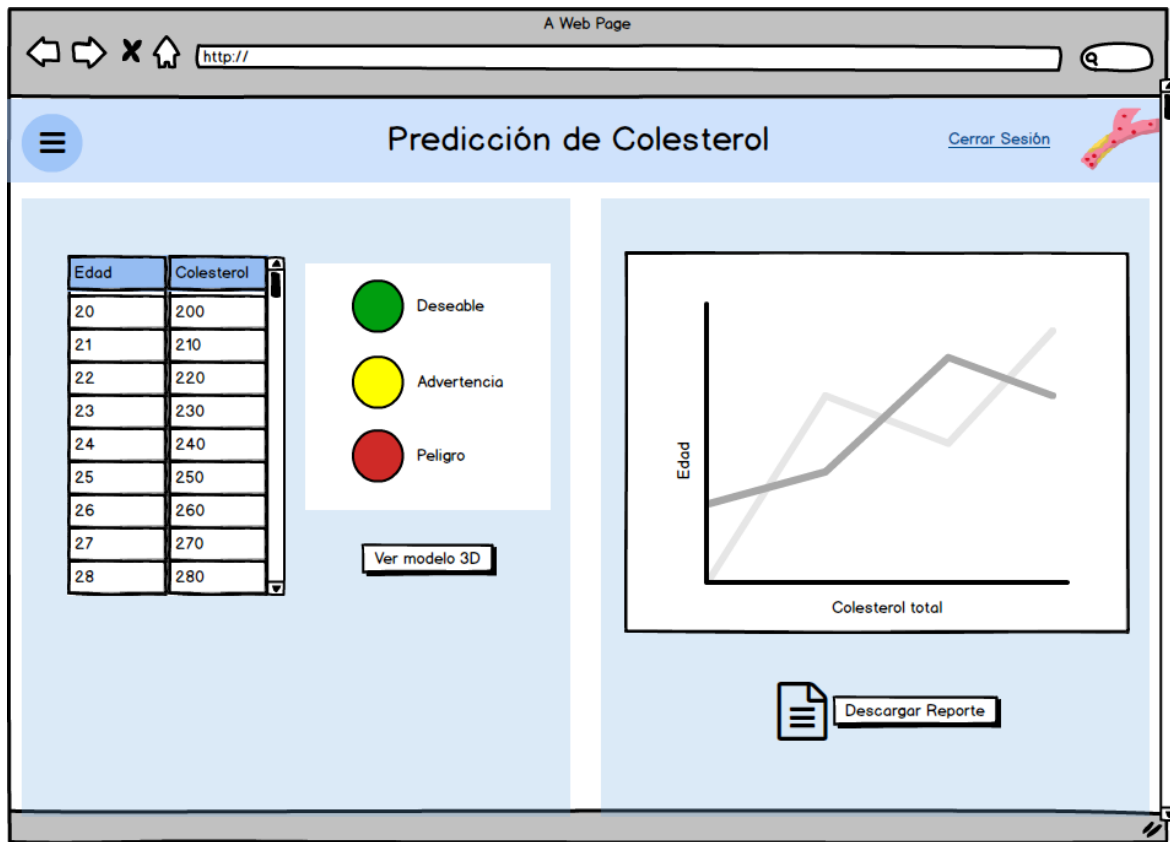
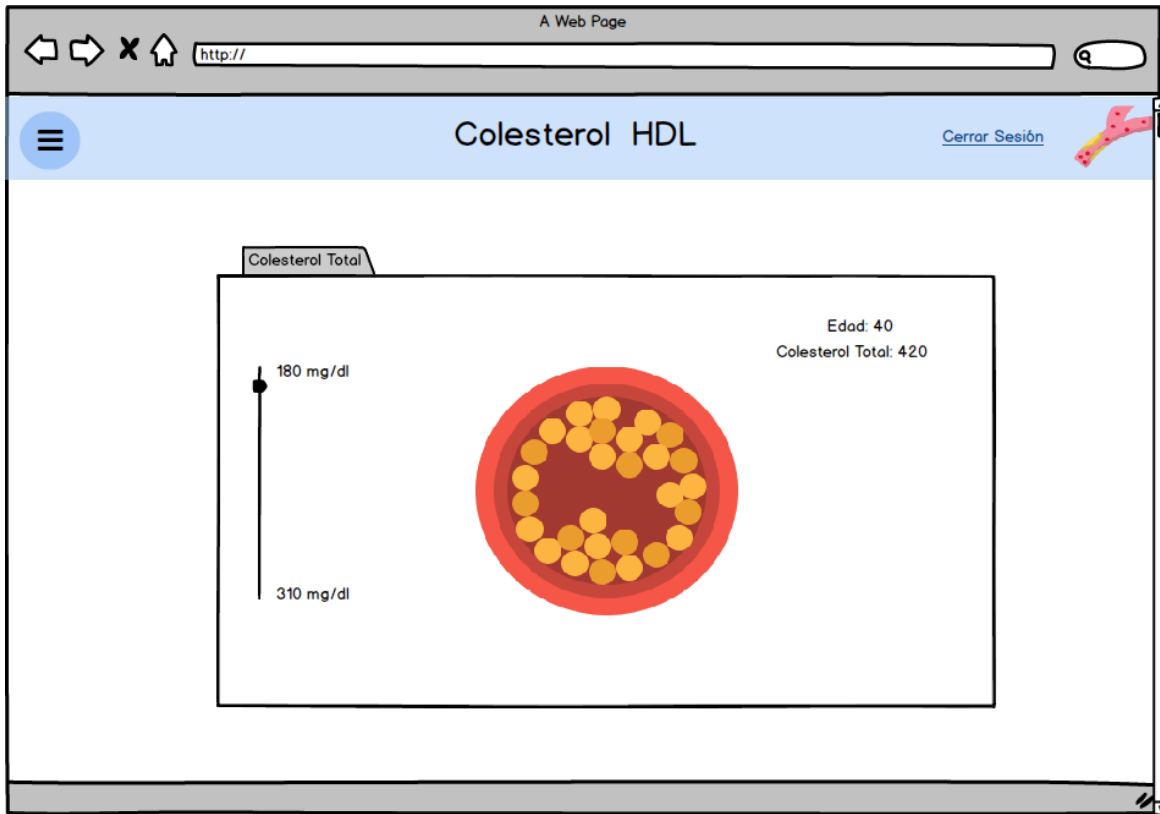


Figura 3.11 *Mockup* Predicción de niveles de Colesterol.



**Figura 3.12** *Mockup* de simulación de colesterol total en la sangre.

En la siguiente figura muestra el menú, en el cual se observan las opciones de ver perfil del paciente, Registro de análisis, el cual se muestra en la figura 3.13 y la opción de predicciones que contendrá los análisis registrados del paciente. Este se visualizará cuando el usuario tenga iniciada la sesión, ya que el diseño es despegable.

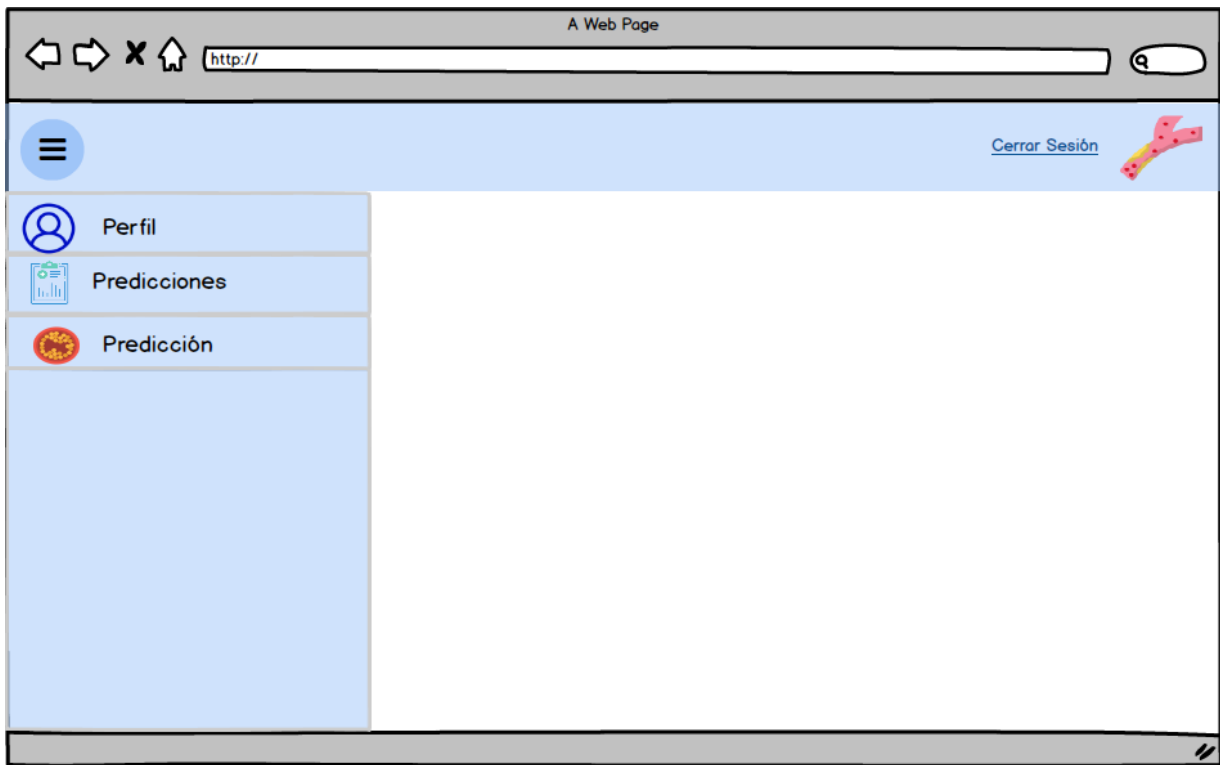


Figura 3.13 *Mockup* Menú.

A continuación, se presenta el *Mockup* de la vista de las predicciones registradas por un usuario, el cual podrá visualizar de nuevo la predicción de colesterol seleccionada.

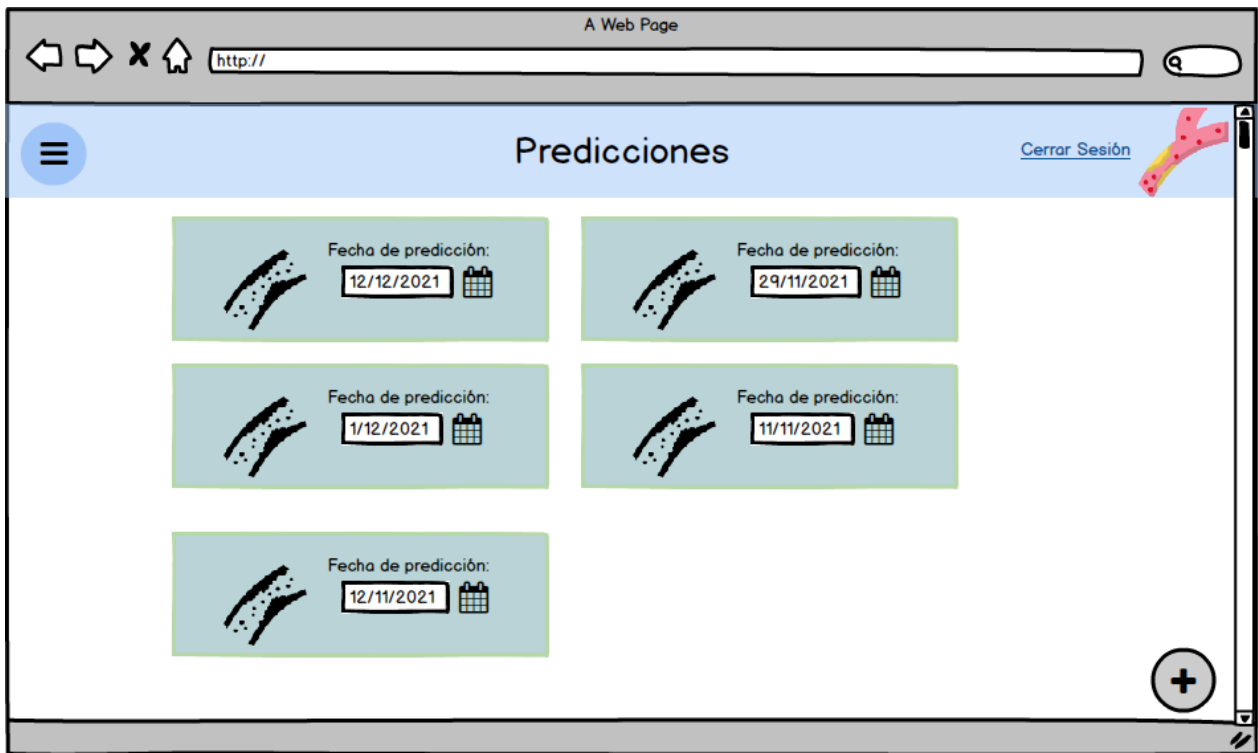


Figura 3.14 *Mockup* Predicciones.

En el *Mockup* Perfil (Figura 3.15) muestra los datos del paciente y tiene una opción de editar la cual conduce a un formulario con el objetivo de actualizar los datos generales del usuario.

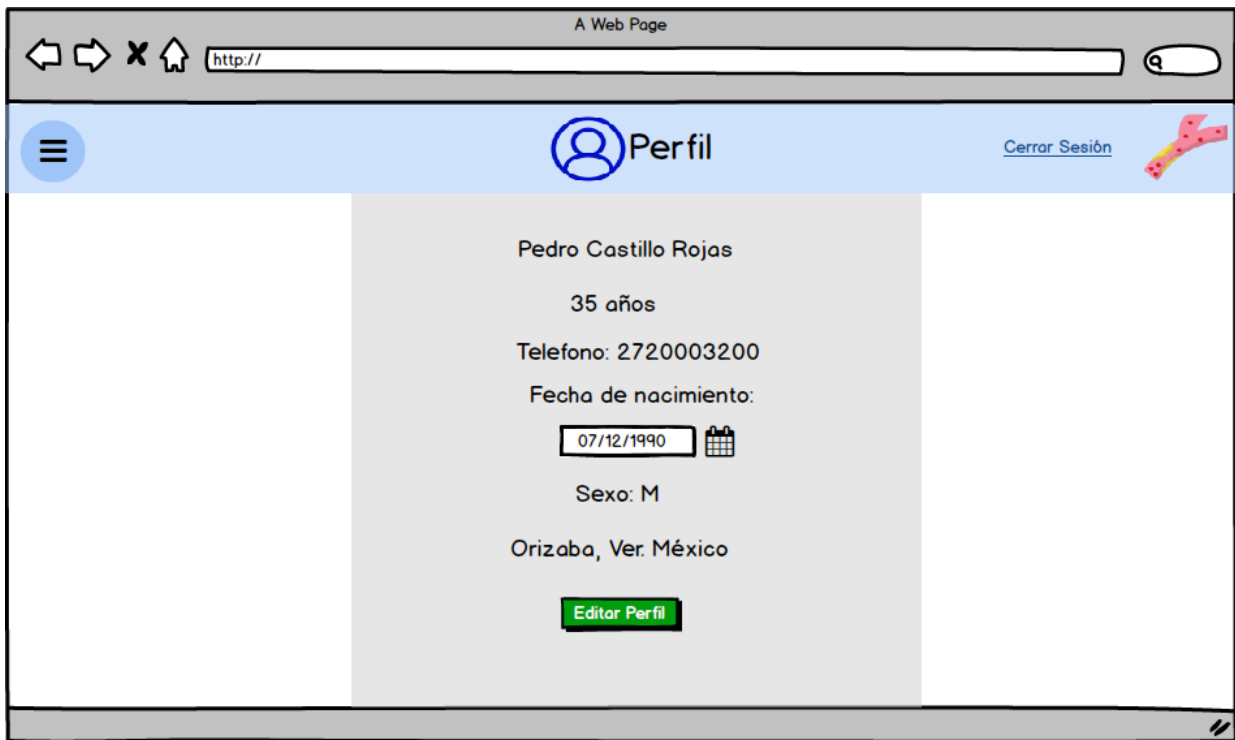


Figura 3.15 *Mockup* Perfil de Paciente.

### 3.2.6 Mapa de navegación

En esta sección se presenta el mapa navegacional de la aplicación en la figura 3.16, el usuario iniciará sesión para tener acceso a las simulaciones del colesterol, también los usuarios tienen la opción de registrarse por Facebook, por Gmail o al ingresar sus datos.

Una vez iniciada la sesión el usuario navegar en las opciones descritas anteriormente en los diseños de los *mockups*.

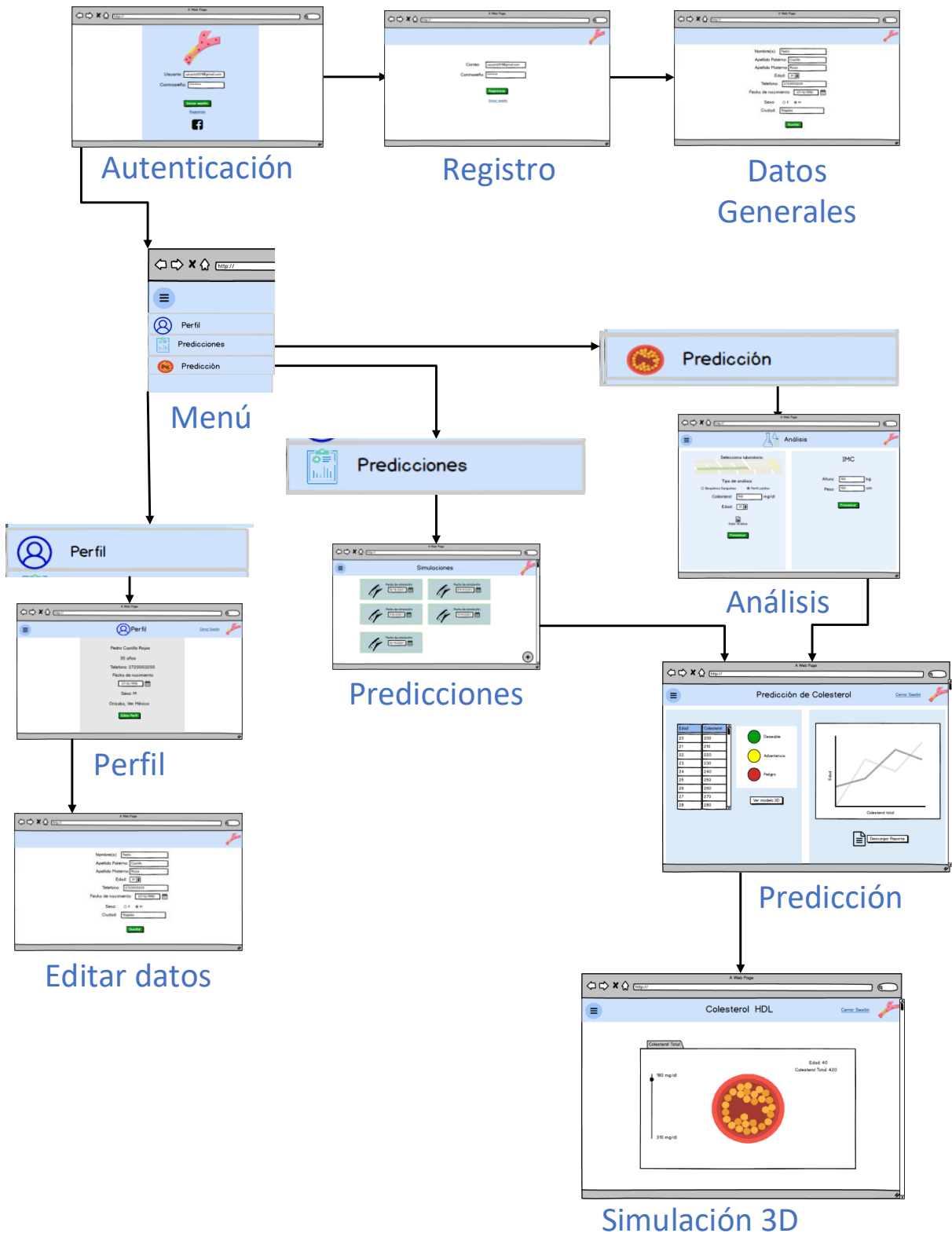


Figura 3.16 Mapa navegacional.



### 3.2.7 Modelo de contenido

Para este proyecto se presenta una composición de la aplicación web, representada por el siguiente diagrama de paquetes, como se muestra en la Figura 3.17, en la cual “routes” funciona como un controlador y “passport” controla los registros y las sesiones de los usuarios. El “colesterol” representa los métodos utilizados para la predicción que genera el sistema, mientras que en el paquete “views”, se encuentran todas las interfaces de usuario. La clase “dpd” representa los métodos para la visualización de los modelos DPD que se encuentran en “3Dmol”.

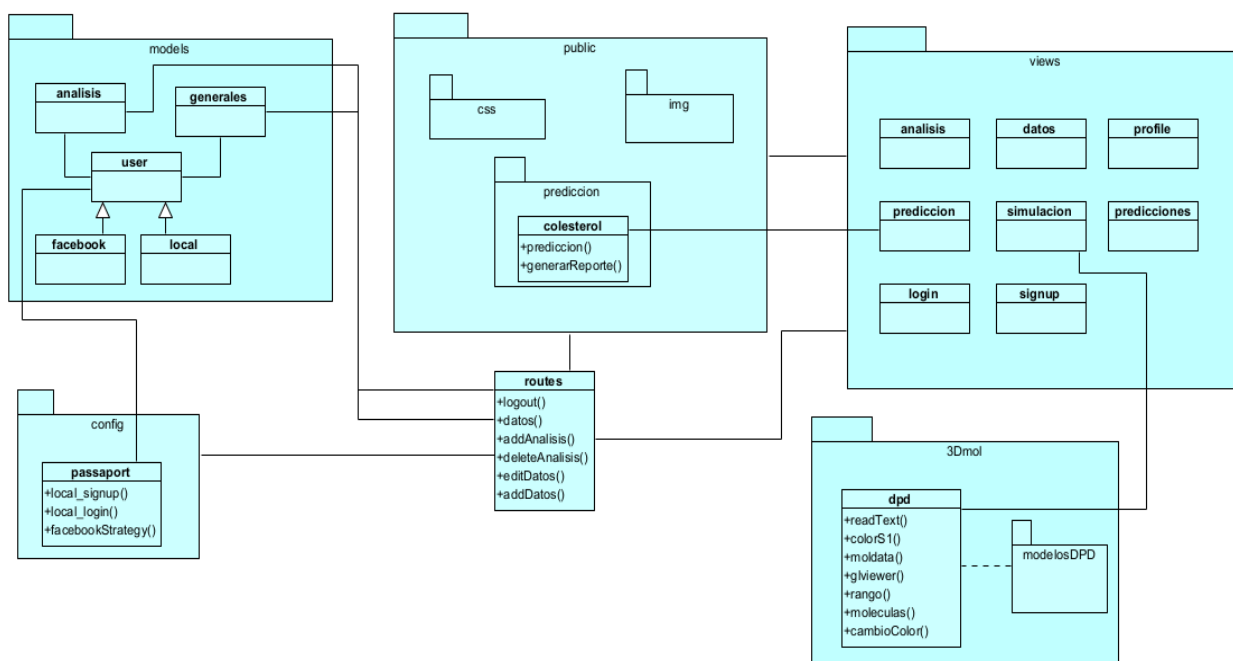


Figura 3.17 Diagrama de paquetes.

En la sección pasada se presentó el diseño del modelo de datos, que incluye la descripción de las tablas, el tipo de datos y sus relaciones, esto permite establecer la estructura de la información con la que trabajará el sistema, así como se definió el diseño de las interfaces de usuario del mismo, posteriormente en el mapa navegacional se explica la forma en que se navegará en la aplicación web.

Una vez presentado el diseño se continuó con el desarrollo del sistema que se describe en la sección 3.3.

### 3.3 Desarrollo

En esta sección se presenta el desarrollo del proyecto.

#### 3.3.1 Módulo de Simulación

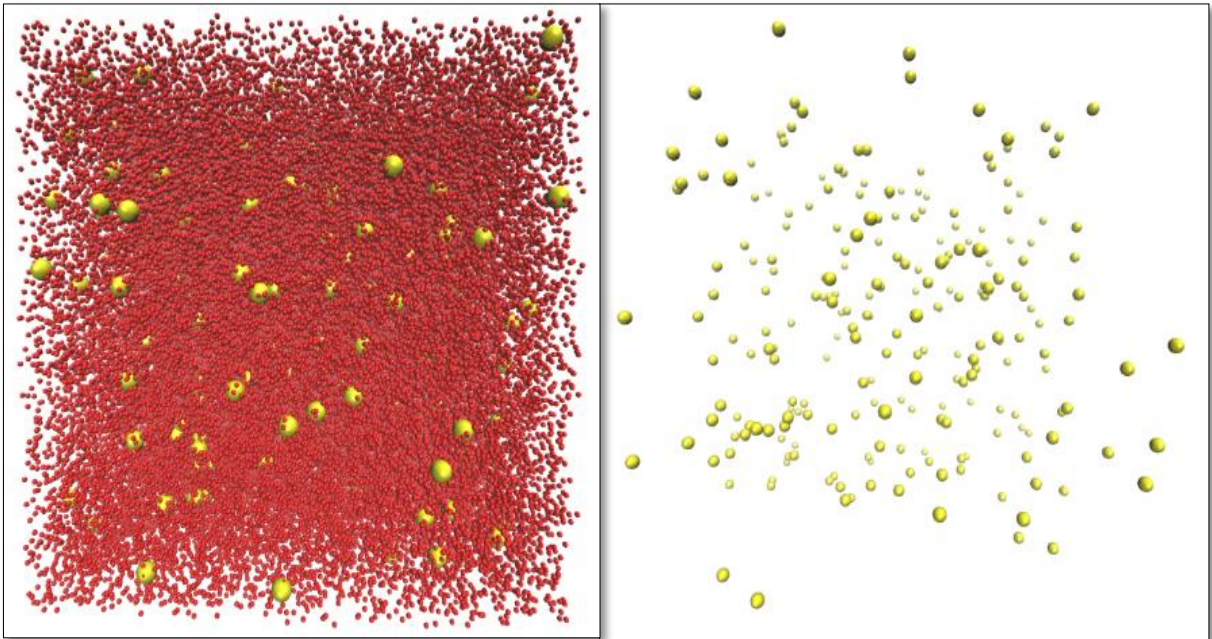
El módulo de simulación calcula posiciones al azar dentro de un cubo representando dos tipos de moléculas, la cantidad de sangre que se representa es un decilitro y cada molécula simulada representa un miligramo. A continuación, se presenta lista con el fragmento de código del módulo de simulación, escrito en lenguaje Python.

```

1 print("\n Creando configuracion inicial \n\n")
2  Natoms1 = Natoms.natoms
3  Natomsf = Natoms1
4  num = Natoms1/Rho
5  Lx= num**(1/3)
6  Ly = Lx
7  Lz = Lx
8  print(" Longitud de la caja de simulacion\n\n", Lx)
9  Rx[1]= (frand()*Lx
10  Ry[1]= (frand()*Ly
11  Rz[1]= (frand()*Lz
12  i=2
13  while i <= Natoms1:
14      rxp = (frand()*Lx
15      ryp = (frand()*Ly
16      rzp = (frand()*Lz
17      #print(rxp,ryp,rzp)
18      addparticle = 0
19      j=1
20      i2=i-1
21      while j <= i2 :
22          rxj = Rx[j]
23          ryj = Ry[j]
24          rzj=Rz[j]
25          #print(rxj,ryj,rzj)
26          rxij = rxj - rxp
27          ryij = ryj - ryp
28          rzij = rzj - rzp
29          n3=((rxij*rxij) + (ryij*ryij) + (rzij*rzij))
30          rij = math.sqrt(n3)
31          if rij <= 0.25:
32              addparticle = 1
33              j=i
34          else:
35              j+=1
36      if addparticle == 0 :
37          Rx[i] = rxp
38          Ry[i] = ryp
39          Rz[i] = rzp
40          i+= 1
41          if i%100000 == 0 :
```

**Lista 1** fragmento de código del módulo de simulación.

En esta lista de código se calcula el tamaño de la caja que contiene los dos tipos de moléculas a simular, ya que el tamaño depende de la cantidad de moléculas totales a representar, posteriormente el módulo calcula las posiciones de los dos tipos de moléculas al azar en forma de coordenadas (eje X, eje Y , eje Z) dentro de la caja, en la figura 3.18 se muestra un ejemplo de colesterol a 210 mg/dl, en donde de color rojo se representa la sangre y con color amarillo se representa el colesterol.



**Figura 3.18 210 mg/dl de colesterol en la sangre.**

La escala elegida para la representación en mg/dl final es de 1 a 1000, esto quiere decir que una molécula representará a las moléculas que contiene 1000 mg.

En la figura 3.19 se muestra un ejemplo con la escala final, en donde se representa 290 mg/dl de colesterol en la sangre, en donde nuevamente de color rojo se representa la sangre y con color amarillo se representa el colesterol.

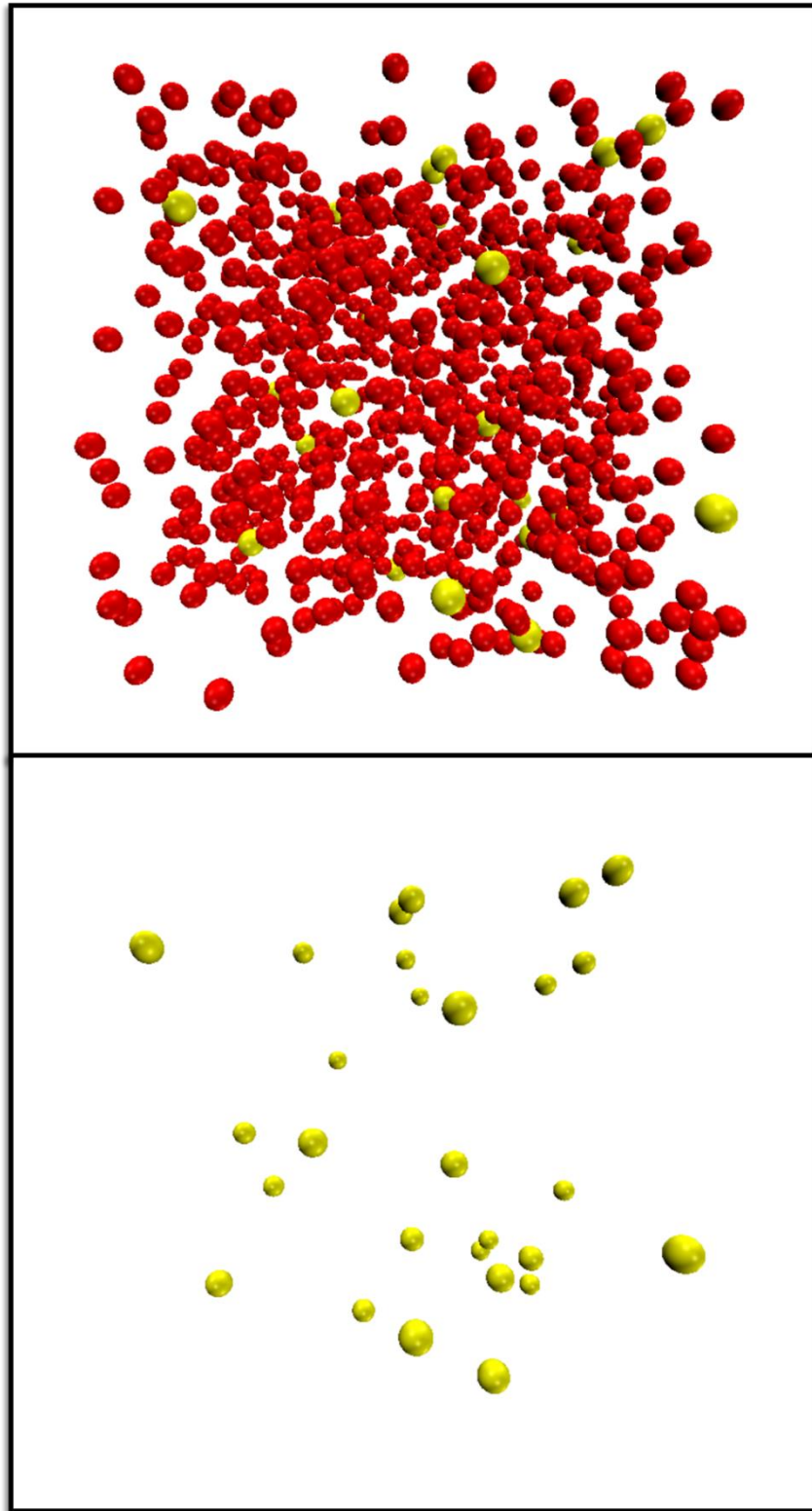


Figura 3.19 290 mg/dl de colesterol en la sangre.

A continuación, se presenta un fragmento de las posiciones de las moléculas generadas, en donde se observa el tipo, la posición en el eje X, la posición en Y y la posición en Z. Para los modelos que genera este módulo se eligió el tipo de molécula “A” para las moléculas de la sangre y el tipo de molécula “B” para representar a las del colesterol.

```
B 18.639 22.066 16.456
B 10.544 28.008 16.304
B 24.199 19.746 29.014
B 16.353 28.385 21.291
B 27.467 27.837 25.648
B 18.808 17.134 20.774
B 26.950 14.334 25.333
B 0.148 0.588 27.977
B 11.048 23.807 24.476
B 14.479 13.260 13.681
B 3.387 14.672 28.142
B 23.440 10.976 15.303
B 17.549 21.922 17.102
A 8.466 7.032 5.814
A 19.395 25.303 15.561
A 1.656 10.181 27.154
A 29.152 16.913 21.316
A 2.073 0.893 5.756
A 2.165 23.534 25.209
A 3.490 23.227 19.778
A 9.097 14.386 5.744
A 17.680 5.007 10.102
A 12.525 28.392 27.282
A 17.396 11.576 19.349
A 8.404 6.052 27.753
A 28.008 21.149 19.499
A 24.920 1.767 27.414
A 26.888 17.713 29.747
A 2.685 4.077 27.578
A 19.288 24.885 21.042
A 2.012 1.051 3.912
```

**Lista 2 fragmento de posiciones de moléculas.**

Para la visualización del modelo se utilizó la biblioteca de JavaScript basada en WebGL llamada “3Dmol.js”.

En la siguiente lista se muestra el fragmento se muestra las bibliotecas que se utilizaron para construir el modelo DPD.

```

1 <script src="http://3Dmol.csb.pitt.edu/build/3Dmol-min.js"></script>
2 <link type="text/css" rel="stylesheet" href="./3Dmol.js/sunlight.default.css">
3 <link type="text/css" rel="stylesheet" href="./3Dmol.js/site.cerulean.css">

```

### Lista 3 fragmento de bibliotecas utilizadas para la construcción de DPD.

Para construir un modelo se muestra un ejemplo en el siguiente fragmento (este es modelo es el logo de la aplicación), en el cual se observa la construcción de cada molécula involucrada y sus propiedades, como el radio que medirá cada molécula, el color, entre otras. La propiedad responsable de construir el modelo, es “render()”, mientras que “zoom (3,1000)” permite ejecutar la escala de tamaño a observar.

```

1 function() {
2   let element = $('#container-05');
3   let config = { backgroundColor: '#b55eb5' };
4   let viewer = $3Dmol.createViewer( element, config );
5   viewer.addSphere({ center: {x:0, y:0, z:0}, radius: .8, color: 'yellow' });
6   viewer.addSphere({ center: {x:1, y:0, z:1}, radius: .8, color: 'red' });
7   viewer.addSphere({ center: {x:0, y:1, z:0}, radius: .8, color: 'red' });
8   viewer.addSphere({ center: {x:0, y:0, z:1}, radius: .8, color: 'yellow' });
9   viewer.addSphere({ center: {x:1, y:0, z:0}, radius: .8, color: 'red' });
10  viewer.addSphere({ center: {x:1, y:1, z:1}, radius: .8, color: 'yellow' });
11  viewer.zoomTo();
12  viewer.render();
13  viewer.zoom(3, 1000);
14 }

```

### Lista 4 Fragmento para la construcción de un modelo DPD.

Para la construcción del modelo DPD de concentración de colesterol, se elaboró un método para leer los datos de un documento, en donde se encuentra las posiciones de las moléculas y su concentración de colesterol elegida. El archivo a leer se encuentra en la variable “moldata”, con el método “glviewer.addModel” se especifica el tipo de archivo, para estas simulaciones se utiliza el tipo “.xyz”. En la lista 6, se muestra el fragmento de código para construir modelos de un archivo con formato .xyz.

```

1  function modelo(){
2      moldata = data = $("#180").val();
3      glviewer = $3Dmol.createViewer("gldiv", {
4          defaultcolors : $3Dmol.rasmolElementColors
5      });
6      glviewer.setBackgroundColor(0xffffffff);
7
8      receptorModel = m = glviewer.addModel(data, "xyz");
9
10     atoms = m.selectedAtoms({});
11
12     for ( var i in atoms) {
13         var atom = atoms[i];
14         atom.clickable = true;
15         atom.callback = atomcallback;
16     }
17
18     glviewer.mapAtomProperties($3Dmol.applyPartialCharges);
19     glviewer.setStyle({}, {sphere:{radius:.15}});
20     colorS1(glviewer, 'red', 'yellow');
21     glviewer.zoomTo();
22     glviewer.render();
23
24 }

```

**Lista 5 fragmento para la construcción de un modelo DPD con datos de un archivo con formato .xyz.**

De igual manera, se le agregan propiedades al modelo, la propiedad “*sphere*” indica el tipo de representación de la molécula, seguido con el tamaño del radio que tendrá este modelo, con el método “colorS1 ()” se le indica que las moléculas tipo “A” se representarán de color rojo, mientras que las moléculas del tipo “B” se representarán de color amarillo, si existe otro tipo de molécula en el archivo, esta se representará de color blanco. En la lista 6 se muestra un fragmento de este método.

```

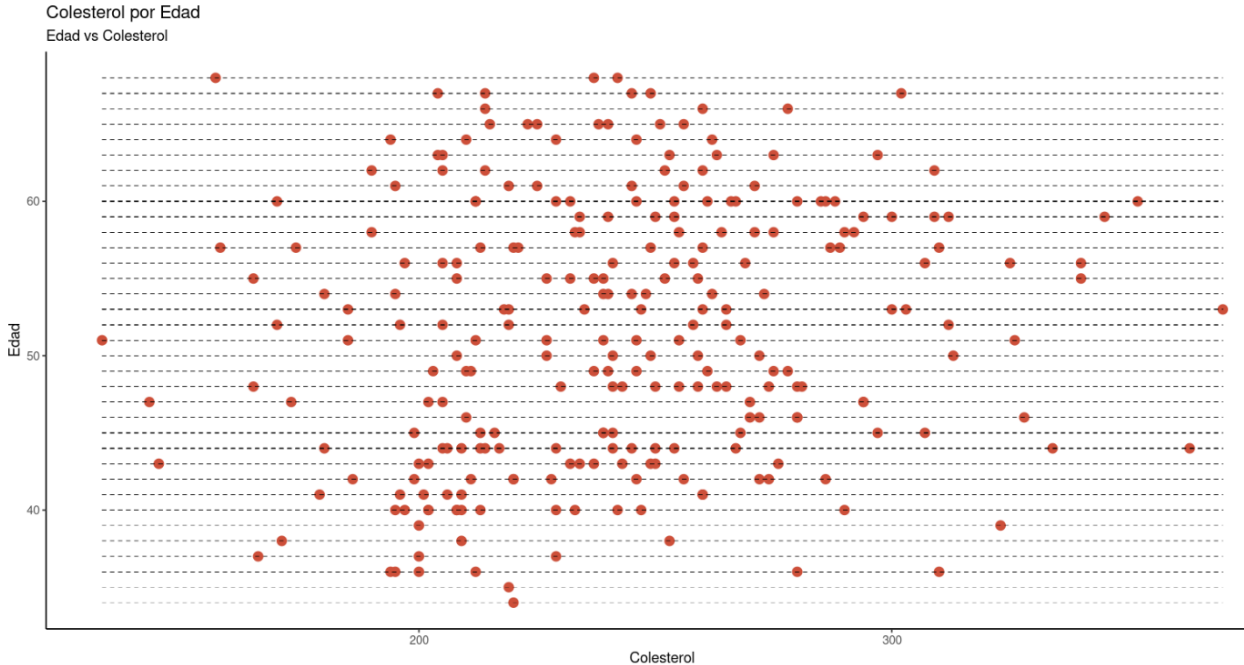
1   var colorS1 = function(viewer,c1,c2) {
2       var m = viewer.getModel();
3       m.setColorByFunction({}, function(atom) {
4           console.log(atom.elem);
5           if(atom.elem == 'A') return c1;
6           else if(atom.elem == 'B') return c2;
7           else return "white";
8       });
9       viewer.render();
10  }

```

**Lista 6 Fragmento para la colorear el modelo**

### 3.3.2 Módulo de Predicción

El módulo de predicción se utilizó la regresión logística utilizando las variables edad frente al colesterol total en mg/dl, en la figura 3.20 se muestra una gráfica con los datos para el entrenamiento del modelo.



**Figura 3.20 Gráfica colesterol por edad.**



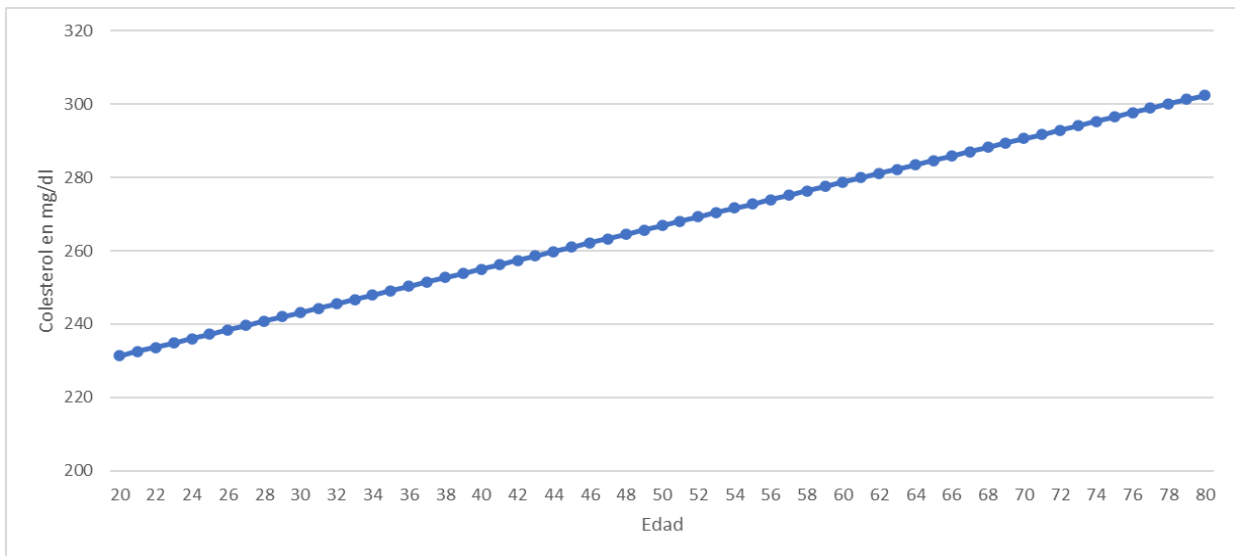
El modelo se creó en lenguaje R, en el siguiente listado se muestra un fragmento de código para la creación del modelo, el cual posteriormente se entrenó.

```
1 #Creating model
2 set.seed(1234)
3 modeloLR <- lm(Colesterol ~ Edad, data = nltrain)
4 summary(modeloLR)
5 #Prediction for dataset training
6 y_predict <- predict(modeloLR, nltrain)
7 ggplot() + geom_point(data = nltrain, aes(x = Edad, y = Colesterol), size = 0.9) +
8 geom_line(aes( x = nltrain$Edad, y = y_predict),
9           color = "red") +
10 xlab("Edad") +
11 ylab("Colesterol") +
12 ggtitle("Curva de Ajuste sobre Conjunto de
13 Entrenamiento (nltrain)")
14 ##Predicting cholesterol between two ages
15 rango.edades <- data.frame(Edad = seq(20, 60))
16 predict_value <- predict(modeloLR, rango.edades)
17 glimpse(predict_value)
18 df_predic=as.data.frame(predict_value)
19 write.csv(df_predic, 'prediccion20_60.csv')
```

**Lista 7 Fragmento de código del modelo para la regresión logística.**

El módulo de predicción se consulta el rango de edad de 20 a 80 años, ya que el trabajo se enfoca en la población adulta y la prevención de enfermedades que posiblemente provoque el tener altos niveles de colesterol.

En la siguiente figura 3.21 se muestra la predicción de los niveles de colesterol obtenida en el rango de edad de 20 a 80 años.



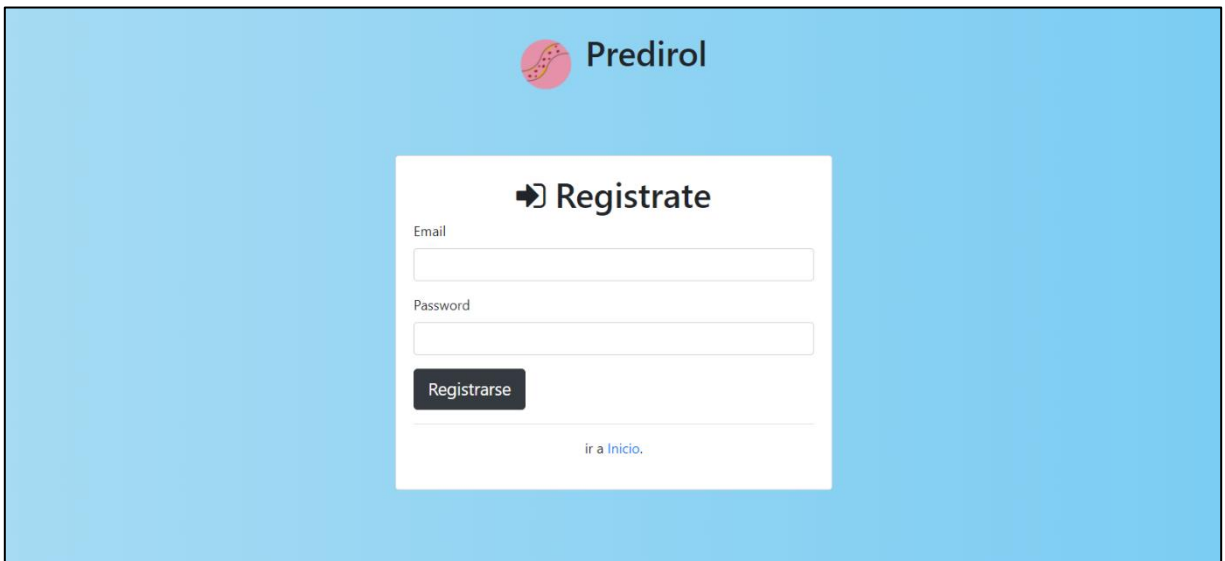
**Figura 3.21** Gráfica obtenida de la predicción del rango de edad de 20 a 80 años.

### 3.3.3 Capa de Presentación

Para el desarrollo de la aplicación se utilizó el entorno NodeJs, a continuación, se presenta las interfaces de usuario para el sistema comenzando por la de “Inicio de sesión” presentada en la figura 3.22. Para esta aplicación el usuario tiene dos opciones para iniciar sesión; la primera consiste en que después de registrarse en el sistema, el usuario proporcionará su correo electrónico y una contraseña para acceder como se muestra en la figura 3.23, y la segunda opción es con una cuenta de la plataforma Facebook, la cual le solicitará permiso al usuario para acceder con sus datos de la misma plataforma.



**Figura 3.22** Interfaz para el inicio de sesión.



**Figura 3.23** Interfaz para el registro de usuario.

Una vez iniciada la sesión el usuario tiene acceso al menú de opciones, se observa en la figura 3.24, en donde aparece la opción de predicción, la cual lo dirigirá a una interfaz en donde tendrá dos opciones para ver las predicciones, una con los resultados de sus análisis y la segunda con la opción de obtener su IMC (Índice de Masa Corporal) como se muestra en la figura 3.25.



Figura 3.24 Interfaz menú de usuario.



Figura 3.25 Interfaz para registro de Análisis.

Posteriormente la predicción aparecerá en una interfaz (figura 3.26), la cual cuenta con la tabla de predicción de los niveles de colesterol de acuerdo con la edad y una gráfica con los datos

obtenidos de la predicción, así también se tiene la opción de ver en 3D la cantidad de colesterol en un decilitro de sangre a escala seleccionando el botón “Modelo 3D”.

Para la elaboración de la gráfica se utilizó la biblioteca “Morris.js”, en la siguiente lista se muestra un fragmento del código que construye la gráfica de predicción (lista 8), en donde se obtiene de la regresión logística de la predicción la concentración de colesterol y con “grafica.push” se especifica el eje X como la edad y el eje Y la concentración de colesterol.



Figura 3.26 Interfaz de predicción de los niveles de colesterol.

```

1 var url = 'http://162.214.187.236:8010/cholesterol_by_range?Edad_In=20&Edad_Fin=80';
2 fetch(url)
3   .then(res => res.json())
4   .then(data => {
5     let l = 20-1;
6     data.forEach((data) => {
7       l +=1;
8       grafica.push({x:l, y: ch2});
9       new Morris.Line({
10        element: 'grafica',
11        data: grafica,
12        xkey: 'Edad',
13        ykeys: ['Colesterol_P', 'Colesterol'],
14        labels: ['Colesterol Actual', 'Colesterol'],
15        lineColors: [color, '#0070c8'],
16        ymin:150,
17        ymax:350,
18        parseTime: false,
19        xLabelMargin: 10,
20        integerYLabels: true,
21        resize: true,
22        pointRadius: 5
23      });
24    })
  }

```

**Lista 8 Fragmento de código para la gráfica de predicción.**

En la figura 3.27 se muestra la interfaz en donde se ve el modelo mesoscópico de la concentración de colesterol en la sangre, el usuario tiene la opción de elegir los colores de las moléculas representadas en el modelo, para distinguir unas de otras, así como manipular el tamaño del modelo, elegir desde qué eje observar, escoger la concentración de colesterol a visualizar, entre otras opciones.



Figura 3.27 Interfaz del modelo de mesoscópico.

## Capítulo 4 Resultados

Este capítulo tiene como objetivo presentar los resultados de este proyecto de tesis y el caso de estudio, de acuerdo con la problemática estudiada y al objetivo trazado para el Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD), llamado “PREDIROL”.

Esta aplicación brindará una predicción de niveles aproximados de colesterol y su relación con la edad, en la cual se podrá observar el aumento de colesterol si la persona no lleva una vida saludable, el colesterol alto es un indicativo de enfermedades asociadas al sistema cardiovascular.

A continuación, se describe el caso de estudio de este proyecto de Tesis.

### 4.1 Caso de estudio

Para este proyecto se obtuvieron los datos y realizaron pruebas en el “Sanatorio Escudero”, ubicado en Poniente 7 No. 905, Centro, Orizaba, Veracruz. El sanatorio ofrece los siguientes servicios:

- Clínica estética
- Clínica oncológica
- Clínica de hemodiálisis
- Clínica de la mujer (ginecológica)
- Obesidad y bariatría
- Tratamientos de varices Heridas Urgencias
- Compañías aseguradoras
- Análisis clínicos
- Exámenes médicos



Los datos de análisis clínicos de pacientes fueron recolectados desde el mes de febrero del año 2022, la información recolectada fue de 150 pacientes durante un periodo de 3 meses. La información proporcionada por el Sanatorio fue edad, colesterol en la sangre, triglicéridos y diagnóstico de los pacientes. El rango de edad fue de 18 años a 88 años.

En la siguiente tabla se muestra algunos datos proporcionados por el sanatorio, omitiendo aquellos datos considerados como sensibles o privados, estos datos se sustituyeron de manera sintética (nombres aleatorios).

**Tabla 4.1 Ejemplo de datos de análisis proporcionados por el sanatorio Escudero.**


NOMBRE	EDAD	DIAGNOSTICO	COLESTEROL	TRIGLICERIDOS
Paciente 1	73	DM	201.5	221.1
Paciente 2	84	PB. DM	244	174.6
Paciente 3	40	DM/IRC	402	660.1
Paciente 4	43	PB. DM	105.1	152.5
Paciente 5	73	DM	201.5	221.1
Paciente 6	24	S/DX	201.5	207.6
Paciente 7	54	PB/ DM	344.6	238.1
Paciente 8	63	S/DX	286.2	222
Paciente 9	31	S/DX	208.2	184.1
Paciente 10	31	S/DX	199.5	128.8
Paciente 11	36	DMII	211.2	222.9
Paciente 12	43	S/DX	269.3	273
Paciente 13	22	ANEMIA/HIV	209.2	216.1
Paciente 14	62	PROSTATA	235.9	94.1
Paciente 15	43	DMII/GEPI	350.7	503.6
Paciente 16	28	S/DX	209.2	216.1
<b>Diagnosticos</b> DM= Diabetes Mellitus 2 IRC = Insuficiencia Renal Cronica S/DX = Sin Diagnostico PROSTATA = Hiperplasia Prostatica Beninga GEPI = Gastroenteritis Probablemente Infecciosa HIV = Infección por el virus de inmunodeficiencia humana				

De esta muestra, seis pacientes realizaron la interacción con el sistema, en la siguiente tabla se muestra los datos de los se muestra el registro de análisis de dichos pacientes de este caso de estudio.


**Tabla 4.2 Llenado de formulario para la predicción de colesterol de los pacientes.**

Paciente 6	<b>Edad</b>	<b>24</b>
	<b>Colesterol</b>	201.5 mg/dl
	<b>Registro de resultado de análisis</b>	

Predirol


## Registro de Análisis


Análisis

Edad Actual

Colesterol

Tipo de Análisis

Archivo de análisis  oariza-9-35-1-ce.pdf

<b>Edad</b>	54
<b>Colesterol</b>	344.6 mg/dl

**Registro de resultado de análisis**

## Predirol

### Registro de Análisis

#### Análisis

Edad Actual

Colesterol

Tipo de Análisis


Archivo de análisis  oariza-9-35-1-ce.pdf

Paciente 7

<b>Edad</b>	63
<b>Colesterol</b>	286.2 mg/dl

**Registro de resultado de análisis**

**Paciente 8**

**Prediol** 

**Registro de Análisis**

**Análisis**

Edad Actual

Colesterol


Tipo de Análisis

Archivo de análisis  Reporte (8).pdf

**Obtener Predicción**

<b>Edad</b>	31
<b>Colesterol</b>	208.2 mg/dl

**Registro de resultado de análisis**

Predirol

## Registro de Análisis

### Análisis

Edad Actual

Colesterol

Tipo de Análisis


Archivo de análisis  Reporte (9).pdf

Paciente 9

<b>Edad</b>	43
<b>Colesterol</b>	269.3 mg/dl

**Registro de resultado de análisis**

**Paciente 12**

**Predirol** 

**Registro de Análisis**

**Análisis**

Edad Actual

Colesterol

Tipo de Análisis


Archivo de análisis  Reporte.pdf

**Obtener Predicción**

Paciente 14

<b>Edad</b>	62
<b>Colesterol</b>	235.9 mg/dl

**Registro de resultado de análisis**

**Predirol**


### Registro de Análisis

**Análisis**

Edad Actual

Colesterol

Tipo de Análisis

Archivo de análisis  Reporte.pdf

Luego de llenar los datos del formulario para el registro de resultados de los análisis, los pacientes eligieron la opción de “Obtener Predicción”, en la siguiente tabla muestra se muestra los resultados obtenidos de la aplicación.

Tabla 4.3 Resultado de la predicción de cada paciente de este caso de estudio.





Paciente 7

<b>Edad</b>	54
<b>Colesterol</b>	344.6 mg/dl

Registro de resultado de análisis

# Predirol

## Niveles de Colesterol

### Predicción

24	236.06
25	237.24
26	238.43
27	239.61
28	240.80
29	241.98

Nivel de Colesterol elevado: 344.6 a la edad de 54 años se recomienda consultar a un Medico

- Nivel de Colesterol Deseable
- Nivel de Colesterol superior al rango normal
- Nivel de Colesterol elevado

Modelo 3D

### Gráfica de predicción de colesterol de acuerdo con la edad

Descargar Reporte

Paciente 8

Edad	63
Colesterol	286.2 mg/dl

Registro de resultado de análisis

# Predirol

## 🔹 Niveles de Colesterol

### 🔗 Predicción

Edad	Colesterol mg/dl
20	231.32
21	232.50
22	233.69
23	234.87
24	236.06

Nivel de Colesterol elevado: 286.2 a la edad de 63 años se recomienda consultar a un Medico

Nivel de Colesterol Deseable

Nivel de Colesterol superior al rango normal

Nivel de Colesterol elevado

Modelo 3D

### 📄 Gráfica de predicción de colesterol de acuerdo con la edad

Descargar Reporte

Paciente 9

Edad	31
Colesterol	208.2 mg/dl

Registro de resultado de análisis

# Predirol

## 🔹 Niveles de Colesterol

### 🔗 Predicción

Edad	Colesterol mg/dl
20	231.32
21	232.50
22	233.69
23	234.87
24	236.06

Nivel de Colesterol Deseable: 208.2 a la edad de 31 años

Nivel de Colesterol Deseable

Nivel de Colesterol superior al rango normal

Nivel de Colesterol elevado


Modelo 3D

### 📄 Gráfica de predicción de colesterol de acuerdo con la edad

Descargar Reporte

Edad	43
Colesterol	269.3 mg/dl

Registro de resultado de análisis

**Prediorol** 

## Niveles de Colesterol

### Predicción

Edad	Colesterol mg/dl
20	231.32
21	232.50
22	233.69
23	234.87
24	236.06

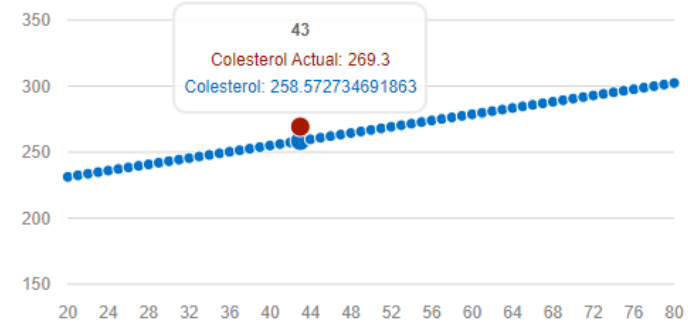
Nivel de Colesterol elevado: 269.3 a la edad de 43 años se recomienda consultar a un Medico



- Nivel de Colesterol Deseable
- Nivel de Colesterol superior al rango normal
- Nivel de Colesterol elevado

**Modelo 3D**

### Gráfica de predicción de colesterol de acuerdo con la edad




**Descargar Reporte**

Paciente 14

Edad	62
Colesterol	235.9 mg/dl

Registro de resultado de análisis

# Predirol




## Niveles de Colesterol

### Predicción

Edad	Colesterol mg/dl
20	231.32
21	232.50
22	233.69
23	234.87
24	236.06

Nivel de Colesterol Deseable: 235.9 a la edad de 62 años



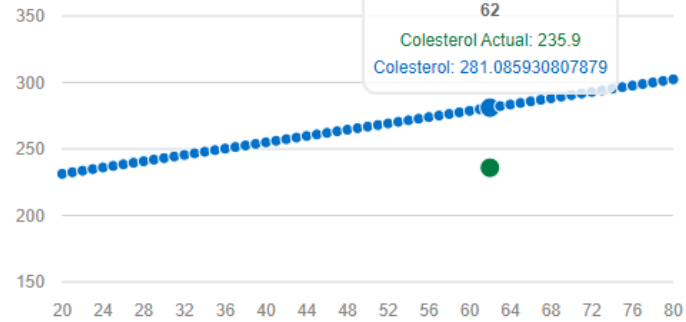
**Nivel de Colesterol Deseable**

**Nivel de Colesterol superior al rango normal**

**Nivel de Colesterol elevado**

Modelo 3D

### Gráfica de predicción de colesterol de acuerdo con la edad



Descargar Reporte

En la tabla 4.4 se muestran los modelos DPD obtenido de cada paciente, con su colesterol total actual, en la misma pantalla el usuario observa otros niveles de colesterol, ya sea en aumento o viceversa.

**Tabla 4.4 Resultado de la predicción de cada paciente de este caso de estudio.**

Paciente 6	Edad	24
	Colesterol	201.5 mg/dl
	Resultado de predicción	

Predirol

### 👁 Simulación Mesoscópica de colesterol en la Sangre



180 mg/dl 


 310 mg/dl

Colesterol= 201 'nivel optimo de colesterol'

Colesterol actual = 201 Edad actual = 24

**● Representación de Moléculas**

Colesterol

Sangre

**⚙ Ajustes de Color**

Moléculas de Colesterol

 Moléculas de Sangre

✎ Cambio de Color

✎ Color por defecto

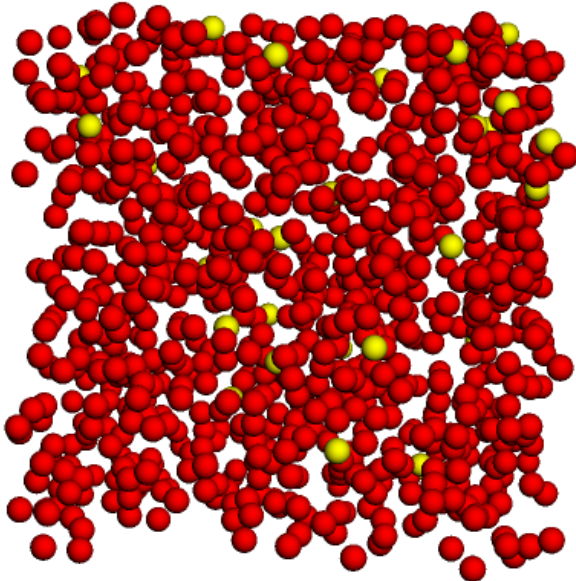
Edad	54
Colesterol	344.6 mg/dl

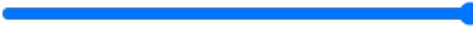
Registro de resultado de análisis

**Predirol** 

## 👁 Simulación Mesoscópica de colesterol en la Sangre

Paciente 7





180 mg/dl  310 mg/dl



Colesterol= 310 Rango de Edad = 78 a 80 años



Colesterol actual = 344 Edad actual = 54

### ● Representación de Moléculas

Colesterol   
Sangre 

### ⚙ Ajustes de Color

 Moléculas de Colesterol  
 Moléculas de Sangre

 Cambio de Color  
 Color por defecto

	Edad	63
	Colesterol	286.2 mg/dl

Registro de resultado de análisis

## Predirol

### 👁 Simulación Mesoscópica de colesterol en la Sangre



180 mg/dl  310 mg/dl  
Colesterol actual = 286 Edad actual = 63

#### ● Representación de Moléculas

Colesterol	
Sangre	

#### ⚙ Ajustes de Color

<input type="checkbox"/>	Moléculas de Colesterol
<input type="checkbox"/>	Moléculas de Sangre

✎ Cambio de Color  
✎ Color por defecto

Paciente 8




Paciente 9

Edad	31
Colesterol	208.2 mg/dl



Registro de resultado de análisis

## Predirol



### Simulación Mesoscópica de colesterol en la Sangre



180 mg/dl  310 mg/dl  
Colesterol actual = 208 Edad actual = 31

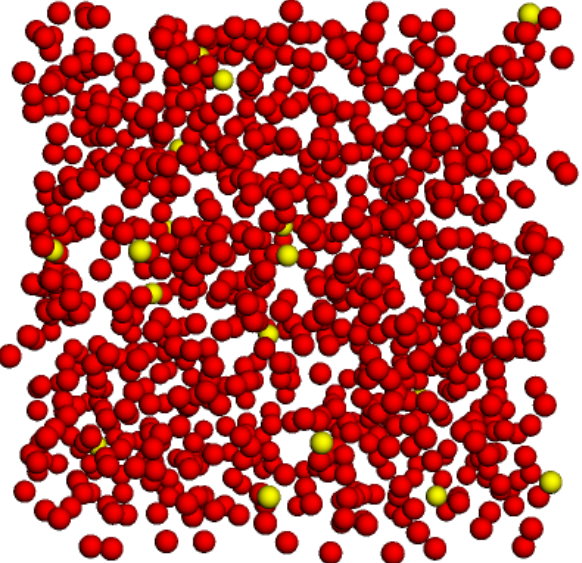
#### Representación de Moléculas

Colesterol   
Sangre 

#### Ajustes de Color

 Moléculas de Colesterol  
 Moléculas de Sangre

 Cambio de Color  
 Color por defecto



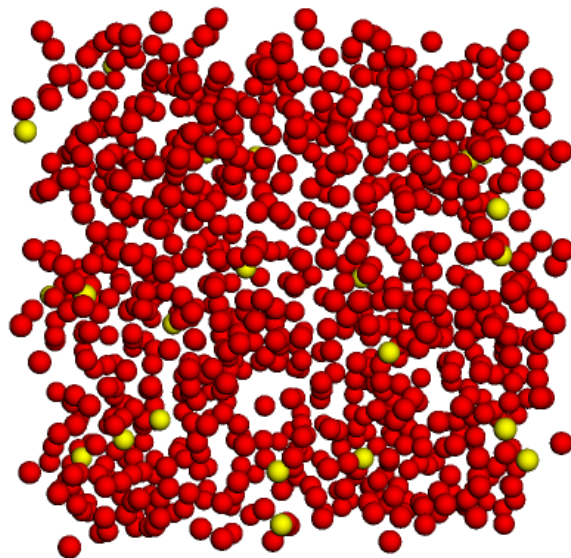
Paciente 12

Edad	43
Colesterol	269.3 mg/dl

Registro de resultado de análisis

**Predirol** 

## 👁 Simulación Mesoscópica de colesterol en la Sangre





180 mg/dl  310 mg/dl



Colesterol = 269 Rango de Edad = 45 a 53 años



Colesterol actual = 269 Edad actual = 43

### ● Representación de Moléculas

Colesterol   
Sangre 

### ⚙ Ajustes de Color

-  Moléculas de Colesterol
-  Moléculas de Sangre

-  Cambio de Color
-  Color por defecto


Paciente 14

Edad	62
Colesterol	235.9 mg/dl

Registro de resultado de análisis

## Predirol


### Simulación Mesoscópica de colesterol en la Sangre


180 mg/dl  310 mg/dl

Colesterol = 235 Rango de Edad = 20 a 27 años


Colesterol actual = 235 Edad actual = 62


#### ● Representación de Moléculas


Colesterol 


Sangre 

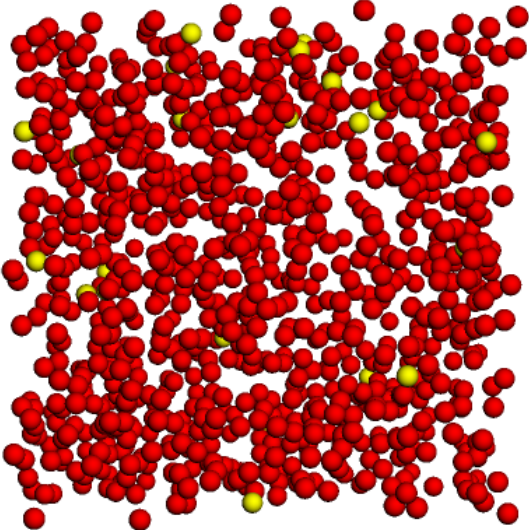
#### ⚙️ Ajustes de Color

 Moléculas de Colesterol

 Moléculas de Sangre

 Cambio de Color

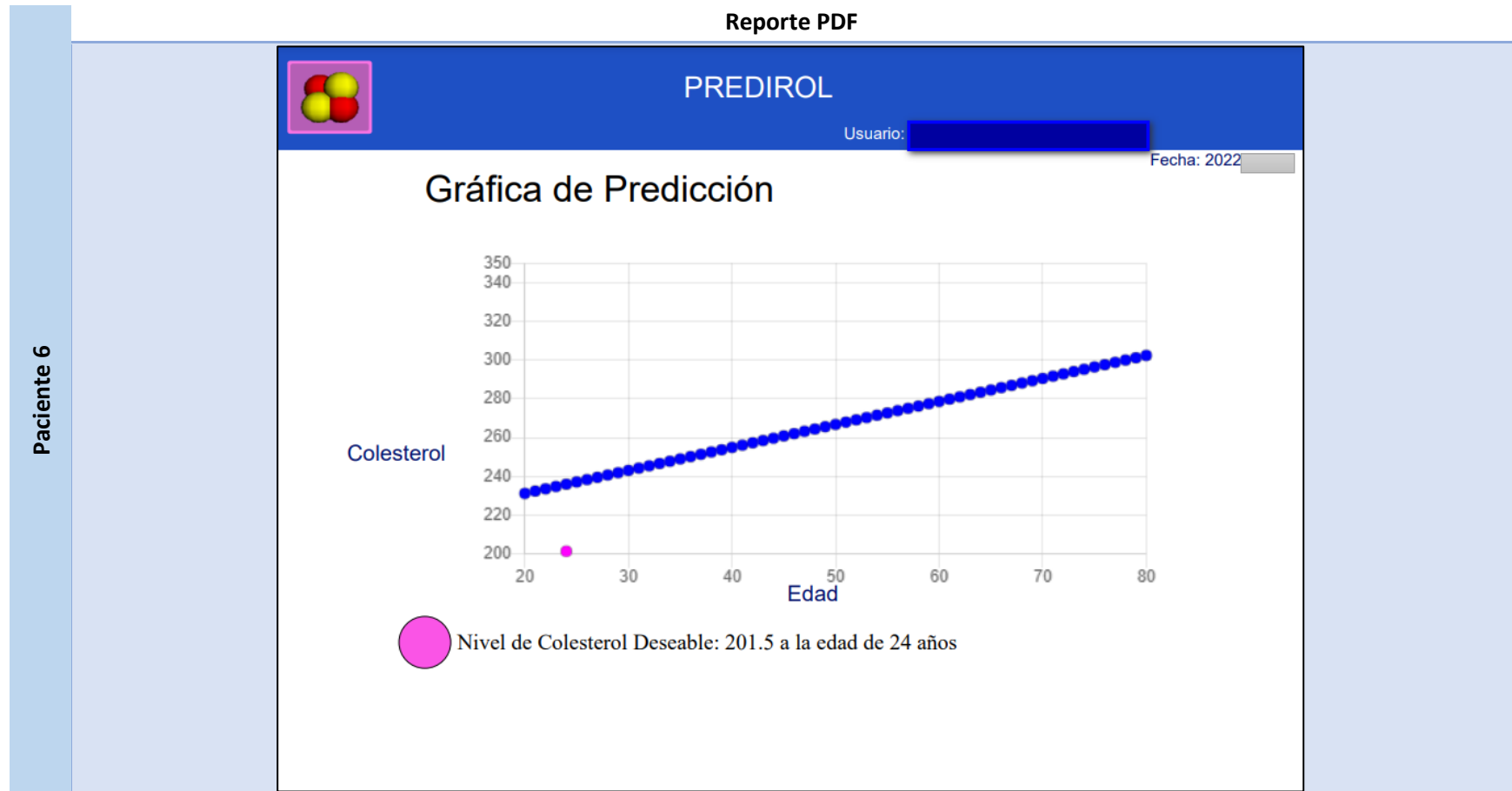
 Color por defecto



The simulation shows a dense cluster of red spheres representing blood molecules and a few scattered yellow-green spheres representing cholesterol molecules.

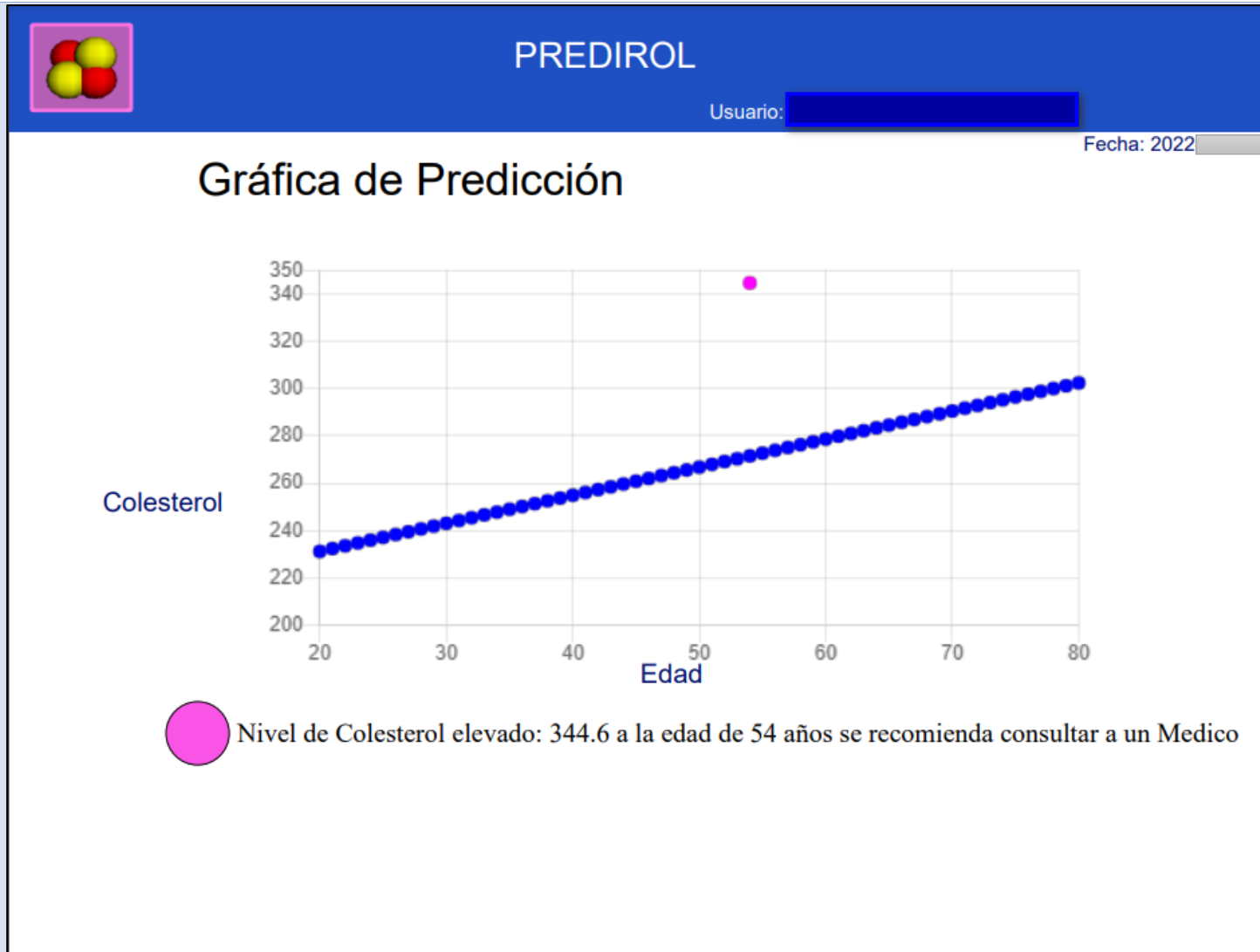
En la tabla 4.5 se muestran los reportes generados en formato “pdf” de cada paciente, con su colesterol total actual y su edad, además de la predicción.

Tabla 4.5 Reporte de la predicción de cada paciente.



Reporte PDF

Paciente 7



Reporte PDF

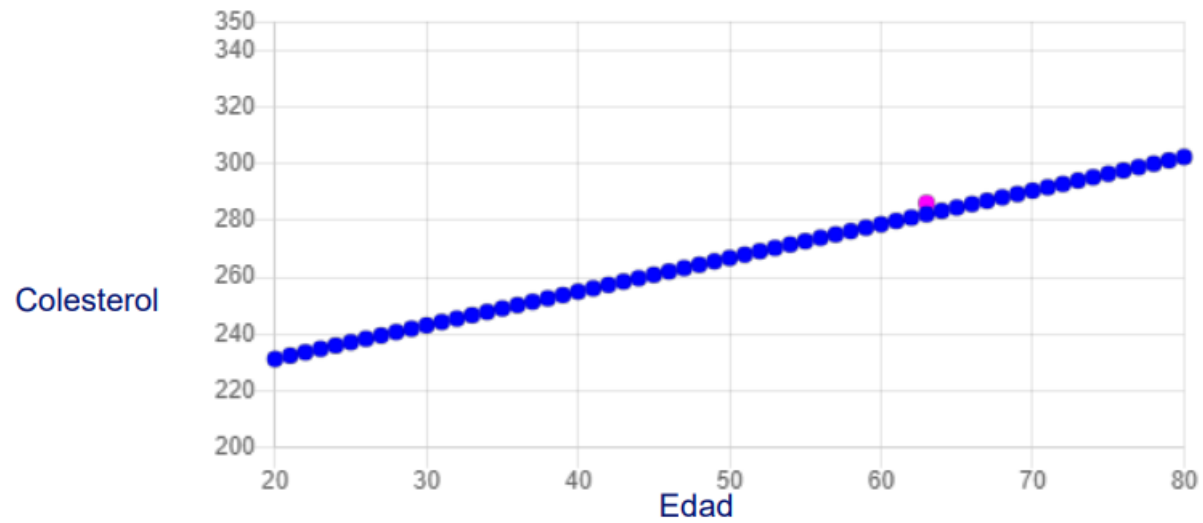


# PREDIROL

Usuario: [Redacted]

Fecha: 2022-

## Gráfica de Predicción



Nivel de Colesterol elevado: 286.2 a la edad de 63 años se recomienda consultar a un Medico

Paciente 8

Reporte PDF

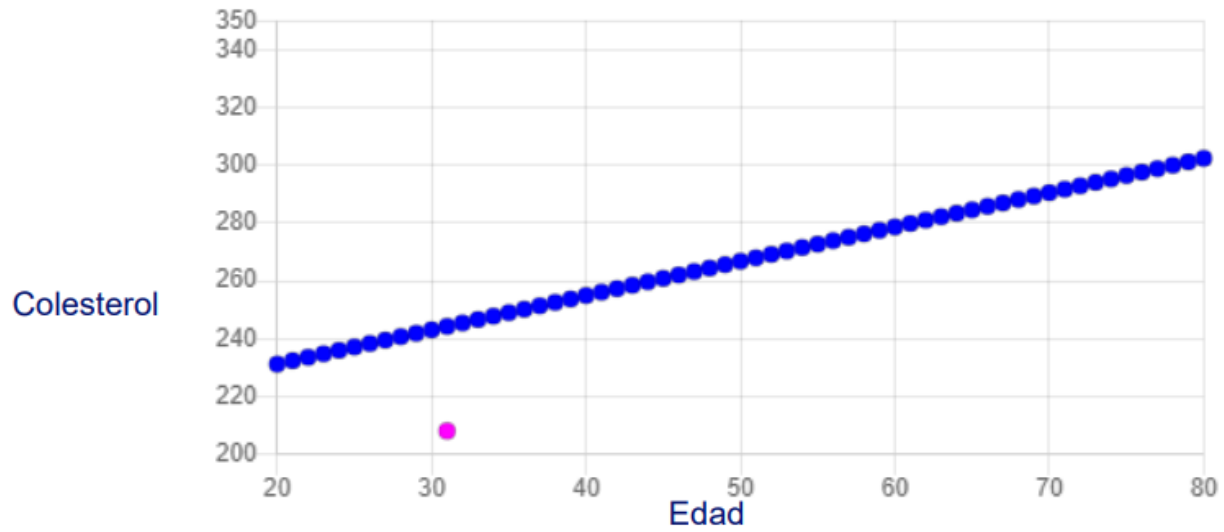



# PREDIROL

Usuario: [Redacted]

Fecha: 2022- [Redacted]

## Gráfica de Predicción



 Nivel de Colesterol Deseable: 208.2 a la edad de 31 años

Paciente 9

Reporte PDF

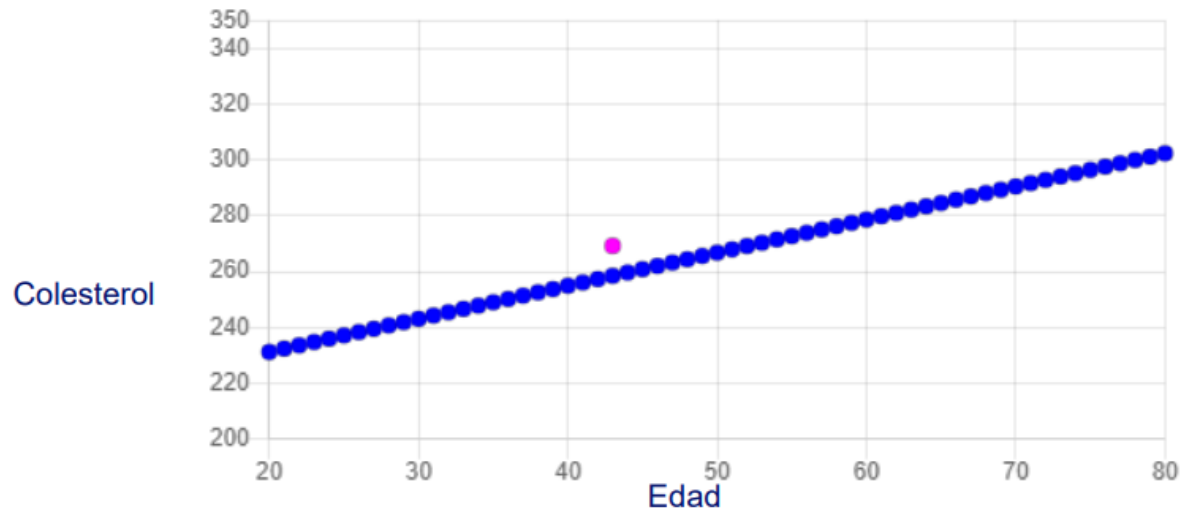



# PREDIROL

Usuario: [Redacted]

Fecha: 2022- [Redacted]

## Gráfica de Predicción



 Nivel de Colesterol elevado: 269.3 a la edad de 43 años se recomienda consultar a un Medico

Paciente 12



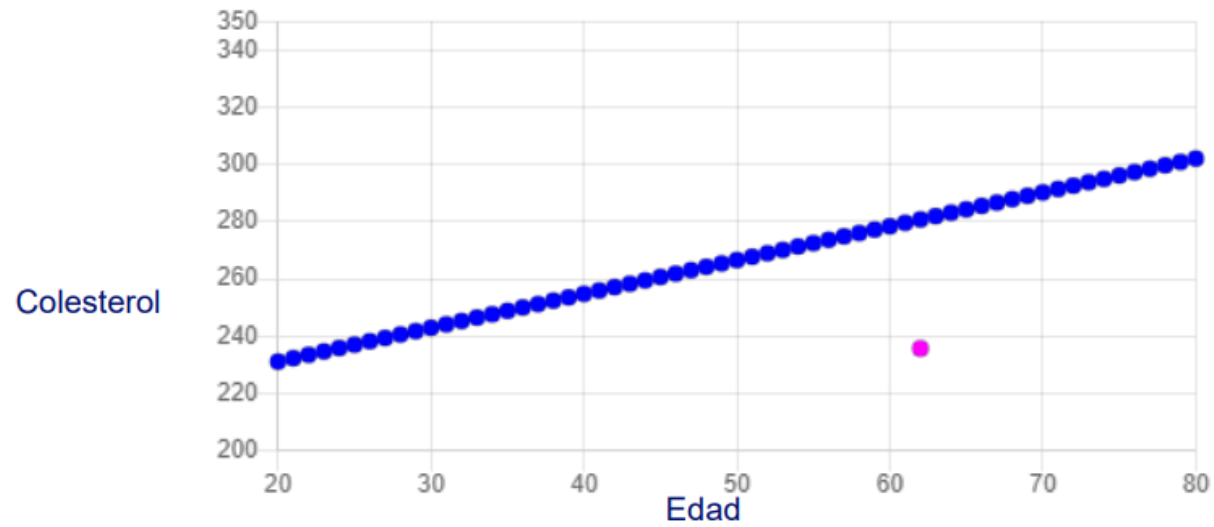
Reporte PDF




# PREDIROL

Fecha: 2022-

## Gráfica de Predicción



 Nivel de Colesterol Deseable: 235.9 a la edad de 62 años

Paciente 14

En la tabla 4.6 se muestra la predicción de acuerdo con la edad obtenida, esta tabla se descarga en el reporte del usuario, si se encuentran dentro de los valores óptimos será de color verde, amarillo indica que se necesita una valoración médica y el color rojo que los niveles de colesterol se encuentran muy elevados.

**Tabla 4.6 Reporte de la predicción de cada paciente.**

Edad	Colesterol
20	231.32
21	232.50
22	233.69
23	234.87
24	236.06
25	237.24
26	238.43
27	239.61
28	240.80
29	241.98
30	243.17
31	244.35
32	245.54
33	246.72
34	247.91
35	249.09
36	250.28
37	251.46
38	252.65
39	253.83
40	255.02
41	256.20
42	257.39

Edad	Colesterol
45	260.94
46	262.13
47	263.31
48	264.50
49	265.68
50	266.87
51	268.05
52	269.24
53	270.42
54	271.61
55	272.79
56	273.98
57	275.16
58	276.35
59	277.53
60	278.72
61	279.90
62	281.09
63	282.27
64	283.46
65	284.64
66	285.83
67	287.01
68	288.20
69	289.38
70	290.57
71	291.75
72	292.93
73	294.12
74	295.30

Edad	Colesterol
75	296.49
76	297.67
77	298.86
78	300.04
79	301.23
80	302.41

Los pacientes que participaron en este caso de estudio fueron de un rango de edad de 20 a 35 años, en este grupo se observa que sus niveles de colesterol se encuentran dentro de los parámetros normales, mientras que los pacientes de 40 a 65 años tienen un colesterol elevado, y los pacientes con valores de colesterol más altos cursan con diagnósticos de enfermedades crónico degenerativa, en estos casos el colesterol alto es asociado al desarrollo de enfermedades cardiovasculares, mientras que en casos como el del Paciente 7 y el Paciente 12 reflejan niveles elevados de colesterol que predisponen a desarrollar ataques cardíacos o accidentes cerebrovasculares en un futuro. En los casos del Paciente 8 y el Paciente 12 su edad y sus niveles de colesterol, coinciden con la predicción, a los cuales tampoco se les indicó un diagnóstico, pero de igual manera la advertencia de los niveles de colesterol, son superiores al rango normal.

## Capítulo 5 Conclusiones y recomendaciones

En este capítulo se presentan las conclusiones y las recomendaciones de la solución de este proyecto llamado “Sistema para la predicción de concentración de colesterol en la sangre utilizando Big Data y técnicas de simulación Mesoscópica con Dinámica de Partículas Disipativas (DPD)”.

### 5.1 Conclusiones

Para este proyecto de tesis se cumplieron todos los objetivos planteados al inicio, con el desarrollo de la aplicación “PREDIROL: Programa para la predicción de niveles de saturación de colesterol”. Al inicio de este proyecto, se realizaron análisis de trabajos relacionados con los temas de simulación Mesoscópica con DPD, Colesterol y la prevención de enfermedades.

Para la predicción se utilizó una regresión logística, para predecir los niveles de colesterol de acuerdo con la edad, el modelo DPD de la concentración de colesterol en la sangre de los pacientes se construyó a una escala de 1 a 1000 mg/dl, esto quiere decir que una molécula representará todas las moléculas que contiene 1000 mg en un decilitro de sangre.

La aplicación busca la concientización sobre el cuidado de los niveles de colesterol, ya que el tener los niveles altos de este provoca acumulación de colesterol y otros depósitos en las paredes de las arterias (ateroesclerosis). Estos depósitos (placas) reducen el flujo sanguíneo a través de las arterias y en un lapso de tiempo, provoca ataques cardiacos o accidente cerebrovasculares. El organismo también necesita colesterol para producción de sustancias por ejemplo hormonas, metabolismo de la vitamina D, es imprescindible para la absorción de calcio y ácidos biliares; en los datos recabados del laboratorio del “Sanatorio Escudero” se observó que los pacientes que tenían una concentración de colesterol total igual o menor a los 80 mg/dl, cursaban con diagnóstico de anemia, siendo susceptibles a otras enfermedades. Esta aplicación concientiza a los pacientes sobre el cuidado de sus niveles de colesterol total, ya que este es un indicativo de que el paciente tenga una buena salud o que este tenga algún padecimiento.

## 5.2 Recomendaciones

En la actualidad la medicina busca prevenir enfermedades detectándolas en el primer nivel de atención para ofrecer una buena calidad de vida a las personas, y así evitar que desarrollen padecimientos más graves que alteren la salud, lo cual afectaría su estilo de vida, economía, entre otros rubros. Las estrategias para lograr esto es concientizando a la población sobre sus hábitos actuales, el objetivo de este proyecto es que las personas perciban que si cursan con niveles de colesterol elevados tienden desarrollar padecimientos que requieran estudios, estancias hospitalarias y medidas terapéuticas con costos elevados.

A continuación, se enumeran recomendaciones para futuros trabajos.

1. Considerar el IMC para predicción de colesterol.
2. Agregar la simulación DPD de triglicéridos, ya que este es útil para el diagnóstico de enfermedades.

Las recomendaciones anteriores surgieron a las observaciones durante los casos de estudio.

## Bibliografía

- [1] A. Zárate, L. Manuel-Apolinar, L. Basurto, E. la Chesnaye, and I. Saldívar, "Colesterol y aterosclerosis. Consideraciones históricas y tratamiento," *Arch. Cardiol. México*, vol. 86, no. 2, pp. 163–169, 2016.
- [2] C. Carvajal, "Lipoproteínas: metabolismo y lipoproteínas aterogénicas," *Med. Leg. Costa Rica*, vol. 31, no. 2, pp. 88–94, 2014.
- [3] T. Gaziano, K. S. Reddy, F. Paccaud, S. Horton, and V. Chaturvedi, "Cardiovascular disease," *Dis. Control Priorities Dev. Countries. 2nd Ed.*, 2006.
- [4] G. Brassard, P. Bratley, and R. Garcia-Bermejo, *Fundamentos de algoritmia*, vol. 86. Prentice Hall ^ eMadrid Madrid, 1997.
- [5] R. Shannon and J. D. Johannes, "Systems Simulation: The Art and Science," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-6, no. 10, pp. 723–724, Oct. 1976, doi: 10.1109/TSMC.1976.4309432.
- [6] E. M., E. D., L. C. L., and E. N., "Estudio del proceso de adsorción-desorción de contaminantes en medios confinados mediante simulaciones computacionales," *Rev. Int. Contam. Ambient.*, vol. 27, pp. 253–261, 2011, [Online]. Available: <https://www.redalyc.org/articulo.oa?id=37020226010>.
- [7] J. Lozano-Aponte and T. Scior, "¿Qué sabe ud. acerca de... dinámica molecular?," *Rev. Mex. Ciencias Farm.*, vol. 45, no. 1, pp. 86–88, 2014.
- [8] J. Xamán, *Dinámica de fluidos computacional para ingenieros*. Palibrio, 2016.
- [9] A. Satoh, *Introduction to practice of molecular simulation: molecular dynamics, Monte Carlo, Brownian dynamics, Lattice Boltzmann and dissipative particle dynamics*. Elsevier, 2010.
- [10] J. M. Costa, *Diccionario de química física*. Ediciones Díaz de Santos, 2005.
- [11] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford university press, 2017.
- [12] G. S. Avila, *Hidráulica general*. Limusa c1993, 1999.
- [13] C. © 2020 N. Corporation, "COMPUTACIÓN ACELERADA," 2021. [96](https://www.nvidia.com/es-</a></li></ol></div><div data-bbox=)

la/drivers/what-is-gpu-computing/.

- [14] J. Cheng, M. Grossman, and T. McKercher, *Professional CUDA c programming*. John Wiley & Sons, 2014.
- [15] E. Dede, B. Sendir, P. Kuzlu, J. Hartog, and M. Govindaraju, "An evaluation of cassandra for hadoop," in *2013 IEEE Sixth International Conference on Cloud Computing*, 2013, pp. 494–501.
- [16] I. © 2021 MongoDB, "La base de datos líder para aplicaciones modernas."  
<https://www.mongodb.com/es>.
- [17] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann Ser. Data Manag. Syst.*, vol. 5, no. 4, pp. 83–124, 2011.
- [18] Ishwarappa and J. Anuradha, "A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology," *Procedia Comput. Sci.*, vol. 48, pp. 319–324, 2015, doi: <https://doi.org/10.1016/j.procs.2015.04.188>.
- [19] J. Hernández, C. Ferri, and M. J. Ramírez, "Introducción a la Minería de Datos," 2004.
- [20] C. Pérez López and D. Santín González, *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Editorial Paraninfo, 2007.
- [21] J. D. V. Henao, "Pronóstico de la Serie de Mackey-Glass usando modelos de Regresión no Lineal," *Dyna*, vol. 71, no. 142, pp. 85–95, 2004.
- [22] D. A. Yaussy, *Multivariate regression model for predicting lumber grade volumes of northern red oak sawlogs*, vol. 536. US Department of Agriculture, Forest Service, Northeastern Forest Experiment..., 1983.
- [23] J. M. Hilbe, *Logistic Regression Models*. CRC Press, 2009.
- [24] IBM, "IBM Documentación," 2017. <https://www.ibm.com/mx-es>.
- [25] D. Makiela, I. Janus-Zygmunt, K. Górny, and Z. Gburski, "Investigation of the influence of  $\beta$ -cyclodextrin on cholesterol lodgement — A molecular dynamics simulation study," *J. Mol. Liq.*, vol. 262, pp. 451–459, 2018, doi: <https://doi.org/10.1016/j.molliq.2018.04.098>.
- [26] L. Yetukuri *et al.*, "Composition and lipid spatial distribution of HDL particles in subjects with



- low and high HDL-cholesterol," *J. Lipid Res.*, vol. 51, no. 8, pp. 2341–2351, 2010.
- [27] S. Medina, J. Zhou, Z.-G. Wang, and F. Schmid, "An efficient dissipative particle dynamics-based algorithm for simulating electrolyte solutions," *J. Chem. Phys.*, vol. 142, no. 2, p. 24103, 2015.
- [28] K. A. Terrón-Mejía, R. López-Rendón, and A. G. Goicochea, "Mesoscopic modeling of structural and thermodynamic properties of fluids confined by rough surfaces," *Phys. Chem. Chem. Phys.*, vol. 17, no. 39, pp. 26403–26416, 2015.
- [29] A. L. Blumers, Y.-H. Tang, Z. Li, X. Li, and G. E. Karniadakis, "GPU-accelerated red blood cells simulations with transport dissipative particle dynamics," *Comput. Phys. Commun.*, vol. 217, pp. 171–179, 2017, doi: <https://doi.org/10.1016/j.cpc.2017.03.016>.
- [30] T. Ye, N. Phan-Thien, C. T. Lim, L. Peng, and H. Shi, "Hybrid smoothed dissipative particle dynamics and immersed boundary method for simulation of red blood cells in flows," *Phys. Rev. E*, vol. 95, no. 6, p. 63314, Jun. 2017, doi: [10.1103/PhysRevE.95.063314](https://doi.org/10.1103/PhysRevE.95.063314).
- [31] M. Pourmousa and R. W. Pastor, "Molecular dynamics simulations of lipid nanodiscs," *Biochim. Biophys. Acta - Biomembr.*, vol. 1860, no. 10, pp. 2094–2107, 2018, doi: <https://doi.org/10.1016/j.bbamem.2018.04.015>.
- [32] Y. Xia *et al.*, "A GPU-accelerated package for simulation of flow in nanoporous source rocks with many-body dissipative particle dynamics," *Comput. Phys. Commun.*, vol. 247, p. 106874, 2020, doi: <https://doi.org/10.1016/j.cpc.2019.106874>.
- [33] K. Kojima *et al.*, "High Wall Shear Stress Is Related to Atherosclerotic Plaque Rupture in the Aortic Arch of Patients with Cardiovascular Disease: A Study with Computational Fluid Dynamics Model and Non-Obstructive General Angioscopy," *J. Atheroscler. Thromb.*, p. 56598, 2020.
- [34] C. Liu *et al.*, "Cholesterol efflux capacity is an independent predictor of all-cause and cardiovascular mortality in patients with coronary artery disease: A prospective cohort study," *Atherosclerosis*, vol. 249, pp. 116–124, 2016, doi: <https://doi.org/10.1016/j.atherosclerosis.2015.10.111>.

- [35] E. Paige *et al.*, "Use of Repeated Blood Pressure and Cholesterol Measurements to Improve Cardiovascular Disease Risk Prediction: An Individual-Participant-Data Meta-Analysis," *Am. J. Epidemiol.*, vol. 186, no. 8, pp. 899–907, 2017, doi: 10.1093/aje/kwx149.
- [36] P. Singh, S. Singh, and G. S. Pandi-Jain, "Effective heart disease prediction system using data mining techniques," *Int. J. Nanomedicine*, vol. Volume 13, no. T-NANO 2014 Abstracts, pp. 121–124, Mar. 2018, doi: 10.2147/IJN.S124998.
- [37] S. Y. Siddiqui *et al.*, "Modelling, simulation and optimization of diagnosis cardiovascular disease using computational intelligence approaches," *J. Med. Imaging Heal. Informatics*, vol. 10, no. 5, pp. 1005–1022, 2020.
- [38] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," *Procedia Comput. Sci.*, vol. 85, pp. 962–969, 2016, doi: <https://doi.org/10.1016/j.procs.2016.05.288>.
- [39] Y. Alhasawi, B. Mullachery, and S. Chatterjee, "Design of a mobile-app for non-invasively detecting high blood cholesterol using eye images," 2018.