

Universidad de Guadalajara



Licenciatura en *Inteligencia Artificial y Ciencia de Datos*



Proyecto Final Programación

Programación

Impartida por

Mtra. Guadalupe Jeannete Gonzalez Dias

P r e s e n t a

Iris Alina Perez Rivera

Fernanda Garcia Rodriguez

Karol Paola Rosales Miranda

Diana Garcia Trujillo

Zapopan, Jalisco; 6 de Noviembre de 2025

Propuesta de Proyecto

Título y Descripción del Problema:

Sistema de detección (Pre-diagnóstico) de SOP (Síndrome de Ovario Poliquístico) para médicos.

El trastorno endocrino y metabólico que se conoce como síndrome de ovario poliquístico (SOP) es común en mujeres en edad reproductiva; se distingue por una gran diversidad en su manifestación bioquímica y clínica. Su diagnóstico requiere la integración de una variedad de criterios hormonales, ecográficos y clínicos, lo que puede hacer que el proceso sea lento y propenso a errores, particularmente en contextos con recursos escasos o sin acceso inmediato a pruebas especializadas.

El objetivo de este proyecto es crear un simulador de diagnóstico probabilístico en Python, con el propósito de respaldar la actividad médica. Para esto, el sistema incorporará elementos hormonales, clínicos y de estilo de vida para calcular la posibilidad de que una paciente tenga SOP.

Además de ello, se busca optimizar la trazabilidad clínica, agilizar el diagnóstico y propiciar el seguimiento a largo plazo de las pacientes, impulsando así una herramienta que sea reproducible, clara y ampliable para la comunidad médica.

Justificación de la POO: ¿Por qué este problema es una buena opción para la POO? ¿Qué "cosas" (objetos) interactúan?

Dado que el problema es modular y estructurado, y que involucra a diferentes entidades que colaboran de manera coordinada en el proceso de diagnóstico, este proyecto se beneficia del uso de la programación orientada a objetos (POO), pues ofrece varios beneficios para la creación del sistema de detección de SOP.

La abstracción hace posible esconder la complejidad de la lógica diagnóstica tras métodos abstractos y clases base, por ejemplo, ReglaDiagnóstico. Esto facilita la comprensión y el uso del sistema.

Por medio del encapsulamiento, se asegura la integridad y la consistencia de los datos almacenados al protegerlos con propiedades y validaciones. Esto incluye las pruebas clínicas y hormonales de cada paciente.

Mediante jerarquías de clases, la herencia permite que se reutilice el código común, lo cual propicia la inclusión de nuevas estrategias o reglas sin tener que cambiar la estructura general y evita la redundancia.

Finalmente, el polimorfismo facilita que diversas clases de estrategias o reglas diagnósticas sean manejadas equitativamente a través de métodos compartidos, como la evaluación del paciente, lo cual aumenta la escalabilidad y flexibilidad del sistema.

Objetos principales que interactúan:

- **Persona:** Clase base con atributos comunes (nombre, edad, sexo).
 - **Paciente:** Hereda de Persona, y agrega datos clínicos, síntomas y resultados de laboratorio.
 - **Médico:** También hereda de Persona, con credenciales y acciones clínicas (dar de alta, solicitar estudios, emitir diagnósticos).
- **ReglaDiagnóstico:** Clase abstracta para modelar criterios de diagnóstico.
- **SistemaDiagnóstico:** Coordina la aplicación de reglas y la obtención del puntaje final.

Identificación de Clases (Propuesta inicial):

Superclase Base:

La regla que se utilizará como modelo para todas las reglas de diagnóstico es la clase principal ReglaDiagnóstico. Esta clase establece lo que debe incluir toda

regla: un nombre, una probabilidad empírica y un peso (qué tan relevante es). Es el fundamento para cualquier criterio que un doctor aplique para examinar el SOP.

ReglaDiagnóstico

Atributos: nombre, prob_empirica, peso.

Método abstracto: `evaluar(paciente)`.

Subclases (Herencia):

1. **ReglaSimple** = Evalúa una sola condición (por ejemplo, “IMC > 25”).
2. **ReglaCompuesta** = Combina reglas con operadores lógicos (AND, OR).
3. **ReglaHormonal** = Especializa la evaluación de valores hormonales según rangos esperados.

Otras clases relevantes:

Persona = Clase base de Paciente y Médico.

Paciente = Contiene síntomas, datos clínicos y resultados hormonales; valida entradas.

Médico = Gestiona diagnósticos y generación de reportes.

DatasetManager = Carga y limpia los datos del dataset PCOS_data_1.xlsx y calcula probabilidades empíricas.

SistemaDiagnóstico = Aplica las reglas a los pacientes, consolida los resultados y entrega un score probabilístico.

Propuesta de Polimorfismo

El método `evaluar(paciente)` se comporta de forma diferente según la subclase:

En **ReglaSimple**: evalúa una condición directa (ej. testosterona > valor_umbral).

En **ReglaCompuesta**: combina resultados de varias reglas con operadores lógicos.

En **ReglaHormonal**: analiza intervalos hormonales y asigna pesos probabilísticos según su grado de anormalidad.

Además, el **SistemaDiagnostico** puede integrar polimorfismo en la estrategia de combinación de los resultados:

CombinerBase = Clase abstracta base

- PromedioPonderado
- BayesSimple

Cada una calcula la probabilidad total de manera diferente, pero todas comparten la misma interfaz, garantizando flexibilidad y escalabilidad del sistema.

| Área / Módulo | Horas estimadas | Jr (30 USD/h) MXN | Mid (50 USD/h) MXN | Sr (80 USD/h) MXN |
|--|-----------------|-------------------|--------------------|-------------------|
| 1. Análisis y diseño | 8–12 h | 4,320–6,480 | 7,200–10,800 | 11,520–17,280 |
| 2. Estructura del proyecto | 6–8 h | 3,240–4,320 | 5,400–7,200 | 8,640–11,520 |
| 3. Modelado OOP y reglas de diagnóstico | 12–16 h | 6,480–8,640 | 10,800–14,400 | 17,280–23,040 |
| 4. Interfaz de consola | 10–14 h | 5,400–7,560 | 9,000–12,600 | 14,400–20,160 |
| 5. Modo Dataset Real | 8–14 h | 4,320–7,560 | 7,200–12,600 | 11,520–20,160 |
| 6. Modo Simulación de Datos | 6–10 h | 3,240–5,400 | 5,400–9,000 | 8,640–14,400 |
| 7. Pruebas y validación | 8–12 h | 4,320–6,480 | 7,200–10,800 | 11,520–17,280 |
| 8. Documentación y empaquetado | 8–12 h | 4,320–6,480 | 7,200–10,800 | 11,520–17,280 |
| Totales aproximados | 58–89 h | 35,640–52,920 | 59,400–88,200 | 95,040–142,920 |