

Malla GEN5: Ciencia de Datos en Salud (Hazla con Datos)

Este programa está diseñado para profesionales de la salud que buscan transformar su toma de decisiones mediante el análisis de datos, priorizando el desarrollo del pensamiento lógico y la estrategia por sobre la simple ejecución de código. Centrado en **R y SQL**, el curso te guiará desde los fundamentos de la programación hasta el diseño de proyectos factibles, entregándote las herramientas críticas para interrogar bases de datos clínicas, automatizar análisis y comprender el alcance real de la Inteligencia Artificial en el ecosistema sanitario.

Módulo 1: Lógica Computacional y Diseño de Proyectos

Objetivo: Desarrollar la estructura mental para programar y evaluar la viabilidad de proyectos de datos en salud.

Fundamentos del Pensamiento Computacional:

- Lógica computacional y pensamiento abstracto: Cómo "pensar" como una computadora.
- Algoritmia básica y diagramas de flujo (independiente del lenguaje).
- Fundamentos de las ciencias de la programación.

Diseño y Estrategia de Datos:

- Elementos de diseño de proyectos de ciencia de datos.
- Evaluación de factibilidad: ¿Es mi problema de salud resoluble con datos?
- Definición de preguntas de negocio/clínicas y KPIs.
- Ética y gobernanza de datos en el entorno sanitario.

Módulo 2: Programación y Gestión de Datos (R y SQL)

Objetivo: Implementar la lógica aprendida utilizando lenguajes específicos para la extracción, manipulación y análisis de datos.

Gestión de Datos con SQL:

- Fundamentos de Bases de Datos Relacionales y Consultas (Queries).

Programación Analítica con R:

- Sintaxis y estructuras fundamentales en R.

- Manipulación, limpieza y transformación de datos (Data Wrangling con tidyverse).
- Manejo de calidad de datos (Missing values, outliers).
- Automatización de reportes (RMarkdown/Quarto).

Módulo 3: Análisis Estadístico y Aplicaciones Prácticas de NLP/LLMs

Objetivo: Aplicar estadística inferencial mediante código y dominar las técnicas de NLP e integración de LLMs, manteniendo el Machine Learning tradicional solo a nivel conceptual.

Estadística Computacional en R:

- Análisis exploratorio de datos (EDA).
- Estadística descriptiva e inferencial aplicada a casos clínicos.

Procesamiento de Lenguaje Natural (NLP) Práctico:

- **Técnicas tradicionales de NLP en código** (Tokenización, *Stopwords*, Frecuencia, N-Grams, etc.) aplicadas a textos clínicos.
- **Uso de LLMs en la Nube:** Integración y uso de APIs en la nube para tareas avanzadas (resumen, clasificación).
- **LLMs en Local:** Introducción a modelos más pequeños y herramientas de código abierto.

Visualización:

- Principios de diseño y *storytelling* para comunicar hallazgos (Teoría y gráficos base en R).

Machine Learning (ML) Tradicional:

- **Revisión Conceptual:** Conceptos generales de modelos (supervisados vs. no supervisados), métricas de evaluación y Explicabilidad en salud (sin implementación con código).