



C3

PROYECTO 42:

Una plataforma universitaria de inteligencia epidemiológica de Enfermedades Metabólicas

Tun centro
TRANSVERSAL
para la **unam**

DR. CHRISTOPHER R. STEPHENS

Coordinador de Ciencia de los Datos del C3

Coordinador del Programa de Complejidad y Salud

Investigador Instituto de Ciencias Nucleares

Objetivos



1) Construir la base de datos más “profunda” en el mundo para el estudio de la obesidad y las enfermedades crónico-degenerativas

- ¿Por qué? A pesar de una inversión enorme en capital financiera y humana estas enfermedades siguen aumentando; no se ha obtenido datos que representen la gran multifactorialidad del problema
- Actualmente contamos con datos de más de 3000 universitarios (residentes de la CDMX) que abarca más de 5000 variables desde lo genético, fisiológico, psicológico, socio-económico, epidemiológico,...

2) Construir plataformas (Platform-as-a-Service) de Ciencia de datos que usan herramientas de Aprendizaje de Máquinas (modelos predictivos Bayesianos y Redes Complejas Inferenciales)

- ¿Por qué? Para abordar el análisis de los datos se requiere herramientas nuevas; las personas que más necesitan aprovechar de estos datos no están capacitadas en ellas
- Actualmente contamos con la plataforma de Proyecto 42 (<https://proyecto42.c3.unam.mx>) donde se puede hacer investigación usando los datos de 1076 residentes de la CDMX (trabajadores, académicos y estudiantes de la UNAM)

Conocimiento Disruptivo a través de las Tecnologías Disruptivas: Una Aplicación a la Prevención de la Obesidad y las Enfermedades Crónico-Degenerativas



3) Hacer disponibles tanto los datos como las plataformas para analizarlos a médicos, investigadores, estudiantes, tomadores de decisión y el público

- Los datos y la plataforma deberían existir para generar investigación y políticas públicas

Importancia de apoyo económico para el Desarrollo de arquitecturas de datos y el Desarrollo de software



MODELOS

Permite la creación de modelos predictivos Bayesianos de tipo

$$P(C|X_1, X_2, \dots, X_N)$$

La probabilidad de pertenecer al grupo C (diabéticos, hipertensos, sedentarios,...) acondicionado en los factores de “riesgo” X_i (edad, antecedentes familiares, estilo de vida etc.)

REDES

Permite la creación de Redes Complejas Inferenciales que exhiben la relación (aristas que representan el grado de correlación estadística entre múltiples causas y efectos). Permite identificar efectos (p.ej. altos triglicéridos) que tiene múltiples causas (nutrición, ejercicio etc.) Y causas (obesidad) que tiene múltiples efectos (hipertensión, hipertriglicideremia etc.)



Modelo NAIVE-BAYES

Selecciona los datos

POBLACIÓN
salud_medicina_1076

INSTRUMENTOS
Laboratorio

SECCIONES
Estilo de Vida

Buscar variable

- ☒ ejer1
- ☐ nr_ejer1
- ☒ ejer5

¿Cuántas horas por semana dedicabas a hacer ejercicio hace 5 años?

Variables Seleccionadas

CARGAR SELECCIÓN PREVIA

- ☒ glu_com
- ☒ id_sexo
- ☒ id_edociv
- ☒ id_gestud
- ☒ id_puesto
- ☒ edad
- ☒ talla
- ☒ peso
- ☒ calc_imc
- ☒ ejer_act
- ☒ ejer1
- ☒ ejer5

AÑADIR FILTROS



Seleccionado un modelo:

C = glucose > 110

X = datos personales, ejercicio, antropometria

Aprox. 70 variables

Modelo P(C | X) generado en 3 seg

Hipótesis: Prediabetes/diabetes

tiene estos factores de riesgo

Output del Modelo

Tabla Resultados				Significado estadístico	Peso en el modelo Bayesiano					
id	Subcategoría	Valor_Variable	Descripción	Respuesta	Epsilon	Score	nx_c	nx	pc	px_c
8	Antropometría	≥ 32.9614	Valor del Cálculo de índice de masa corporal	≥ 32.9614	3.79	0.67	26	107	0.12	0.24
4	Datos Personales	149.0	¿Qué puesto ocupas?	VIGILANTE B	3.64	0.76	10	29	0.12	0.34
5	Datos Personales	≥ 59.5	Edad del Paciente	≥ 59.5	3.61	0.63	27	116	0.12	0.23
3	Datos Personales	S	¿cuál es tu máximo grado de estudios?	SECUNDARIA	3.51	0.58	31	141	0.12	0.22
5	Datos Personales	54.5 - 59.5	Edad del Paciente	54.5 - 59.5	3.48	0.61	26	113	0.12	0.23
4	Datos Personales	131.0	¿Qué puesto ocupas?	OFICIAL DE TRANSPORTE B	3.35	0.6	5	11	0.12	0.45
3	Datos Personales	P	¿cuál es tu máximo grado de estudios?	PRIMARIA	2.85	0.59	12	46	0.12	0.26
9	Estilo de Vida	0.0	¿Cuántas horas semanales dedicas al ejercicio?	0.0	2.84	0.33	73	436	0.12	0.17
7	Antropometría	≥ 88.5	¿Cuál es tu peso actualmente?	≥ 88.5	2.81	0.51	22	103	0.12	0.21
6	Antropometría	< 1.495	Talla	< 1.495	2.76	0.5	22	104	0.12	0.21

¿Cuáles factores son “Actionable”?

Curva ROC

Incidencia de prediabetes y diabetes en la población general

Incidencia de prediabetes y diabetes en la población con X

Modelo Predictivo para ellos quienes no hacen ejercicio

Tabla Resultados

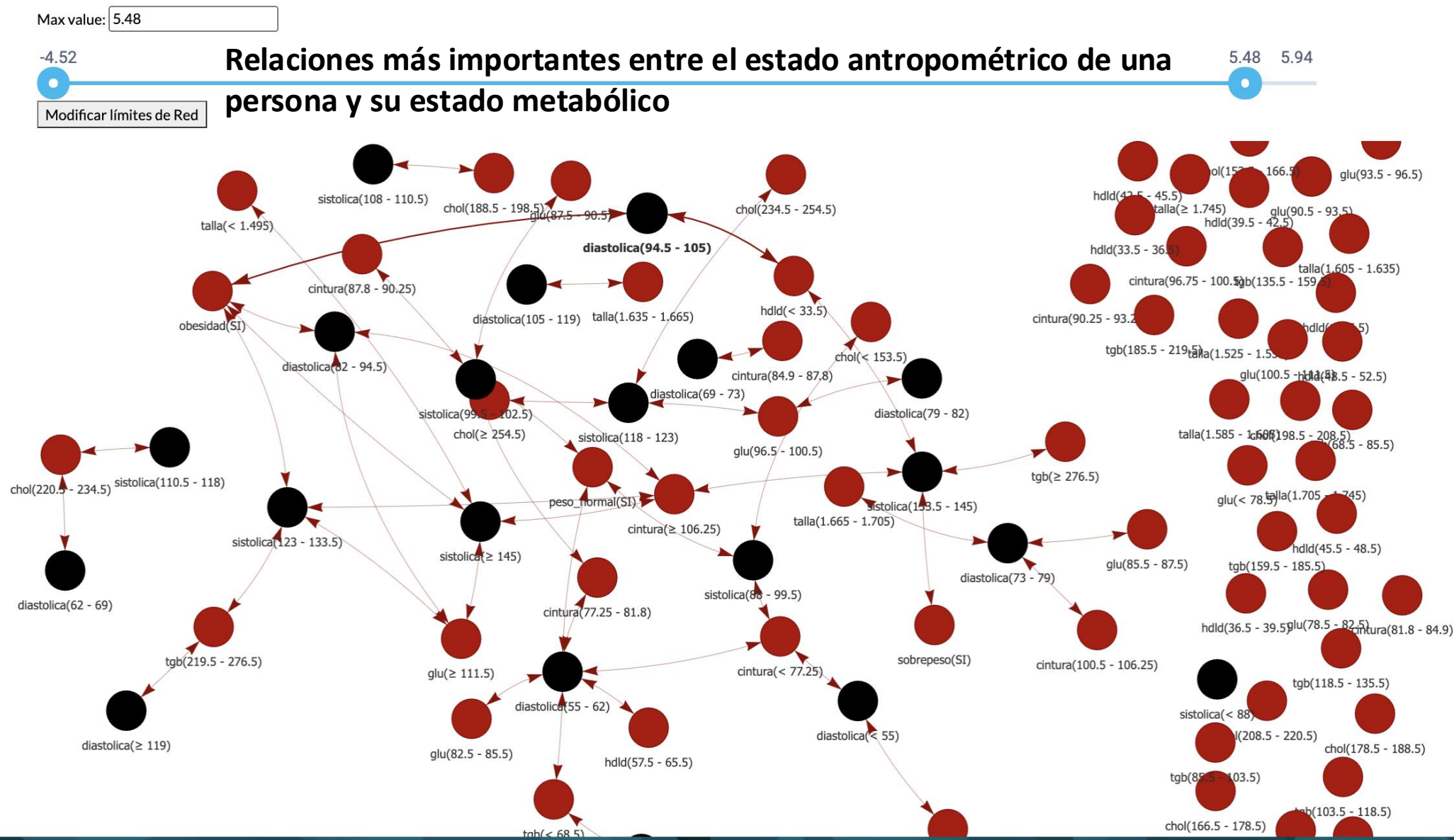
¡Ejercicio es un hábito!

id	Subcategoría	Valor_Variable	Descripción	Respuesta	Epsilon	Score	nx_c	nx	pc	px_c
1	Estilo de Vida	< 0.5	¿Cuántas horas por semana dedicabas a hacer ejercicio hace un año?	< 0.5	12.84	1.21	299	419	0.41	0.71
2	Estilo de Vida	< 0.5	¿Cuántas horas por semana dedicabas a hacer ejercicio hace 5 año?	< 0.5	5.41	0.53	185	336	0.41	0.55
3	Estilo de Vida	< 0.5	¿Cuántas horas por semana dedicabas a hacer ejercicio hace 10 años?	< 0.5	4.12	0.39	178	346	0.41	0.51
7	Datos Personales	149.0	¿Qué puesto ocupas?	VIGILANTE B	3.87	0.67	22	29	0.41	0.76
9	Datos Personales	S	¿cuál es tu máximo grado de estudios?	SECUNDARIA	3.57	0.47	78	141	0.41	0.55
9	Datos Personales	T	¿cuál es tu máximo grado de estudios?	CARRERA TÉCNICA	3.26	0.46	59	105	0.41	0.56
7	Datos Personales	124.0	¿Qué puesto ocupas?	AUXILIAR DE INTENDENCIA A	3.2	0.48	45	77	0.41	0.58
9	Datos Personales	P	¿cuál es tu máximo grado de estudios?	PRIMARIA	2.81	0.43	28	46	0.41	0.61
5	Datos Personales	U	Estado civil	UNIÓN LIBRE	2.53	0.37	32	56	0.41	0.57

Escolaridad y ocupación son muy importantes

Curva ROC

Output de la Funcionalidad de Redes



Usos de Proyecto 42 para la Secretaria de Salud y el programa de Salud en tu Vida

Para investigación

- Creación de modelos predictivos
- Identificación de factores de riesgo
- Cambio de conducta

Para entrenamiento/capacitación de profesionales de la salud

Para toma de decisión

- Políticas públicas que reconozcan y respetan la complejidad (multi-factorialidad)