



**C3**

# **EPIPUMA:** **Una plataforma universitaria** **de inteligencia epidemiológica** **de SARS-CoV-2**

**T**un 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

# ¿Quién esta ayudando?

Chris Stephens C3 y ICN, UNAM

Rodolfo Rivas IMSS

Victor Villalobos CENAPRECE

Raúl Sierra Alcocer CONABIO

Juan Carlos Salazar CONABIO

Pedro Romero CONABIO

Gabriel Garcia FMZN, UNAM

Constantino Gonzalez CCA, UNAM

Julián Velasco CCA, UNAM

Antonio Barajas FM UNAM

Luís Flores FM UNAM

Carlos A. Hernández IIMAS UNAM

Marco Rosas, C3, UNAM

Adriana Robles, IIB

Estefa Espitia, C3, UNAM

Miriam Barajas, C3, UNAM

Jonathan Easton C3, UNAM

Hugo Flores C3, UNAM

Dagmara Wrzcionkowska F Psic, UNAM

Ivette Rizzo C3, UNAM

Rolando Díaz-Loving, Fac. Psic., UNAM

Manuel Suarez Igg, UNAM

Juan Pablo Gutierrez FM, UNAM

Gustavo Olaiz CIPSS, UNAM

Armando Sanchez IIE, UNAM

Jésica Tapia C3, UNAM

Lilian Castillo IIMAS, UNAM

Alejandro Ruíz IIMAS, UNAM

Alejandro Salinas IIMAS, UNAM

Tanya Arenas C3, UNAM

Adrian Ghirardi CIGA, UNAM

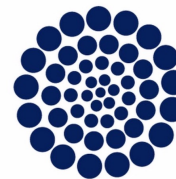
Gerardo Suzán FMVZ, UNAM

Romel Calero C3, UNAM

José Luis Gordillo C3, UNAM



CONABIO



CONACYT



dga

Dirección General de Asuntos  
del Personal Académico

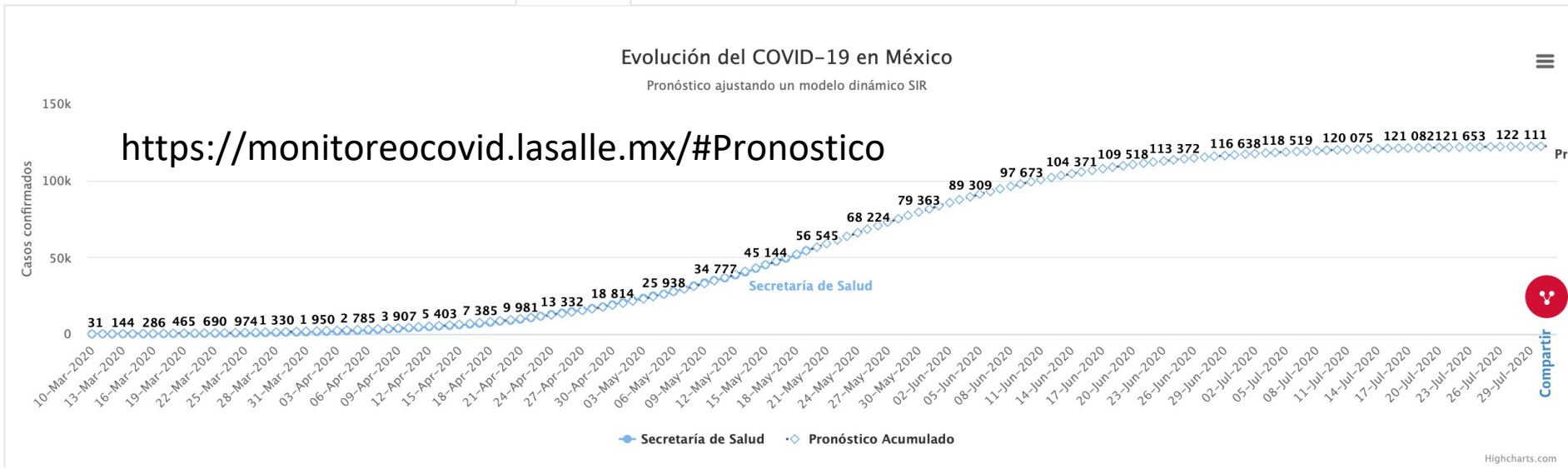


Microsoft



# Casos confirmados

- Modelo Neuronal
- Modelo Gausiano
- Modelo Logsig
- Modelo SIR**
- Modelo SIR Ponderado



## Pronósticos utilizando redes neuronales artificiales y cómputo evolutivo

# Casos confirmados

- Modelo Neuronal
- Modelo Gausiano
- Modelo Logsig
- Modelo SIR
- Modelo SIR Ponderado**

# Hay muchas predicciones



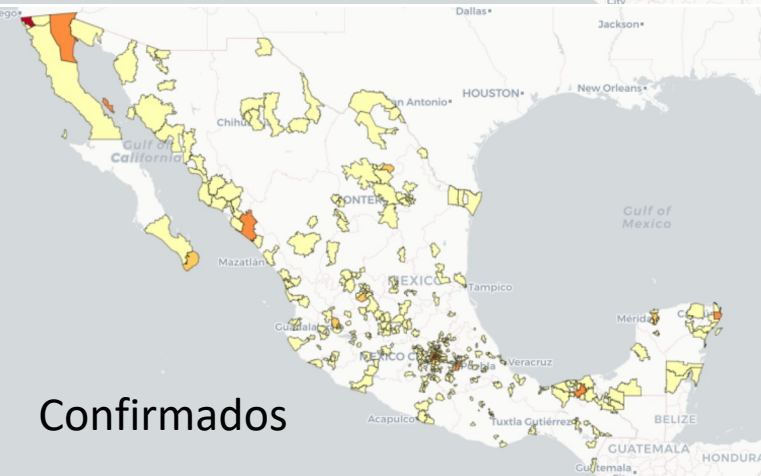


Confirmados

01/03/2020

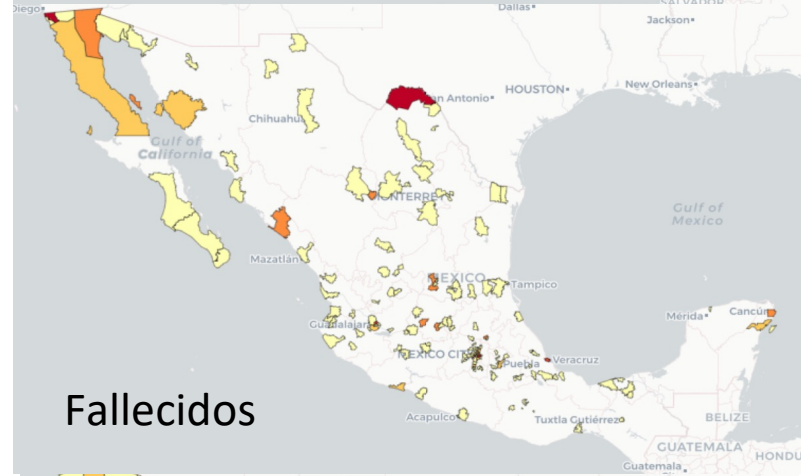


Fallecidos

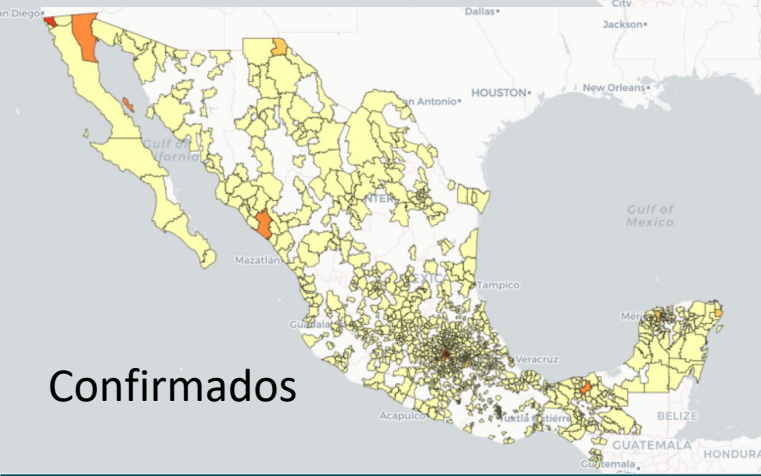


Confirmados

01/04/2020

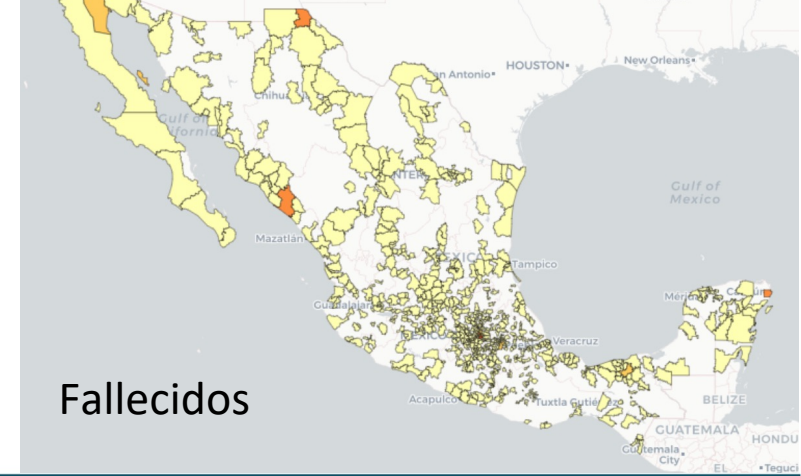


Fallecidos



Confirmados

01/05/2020



Fallecidos

### Gustavo A. Madero

Nombre	Edad	Genero	Fecha Colecta
COVID-19 CONFIRMADO	41-49	Femenino	2020-02-28
COVID-19 CONFIRMADO	50-55	Masculino	2020-02-23

### Miguel Hidalgo

Nombre	Edad	Genero	Fecha Colecta
COVID-19 CONFIRMADO	19-30	Masculino	2020-04-01
COVID-19 CONFIRMADO	50-55	Masculino	2020-04-01
COVID-19 CONFIRMADO	19-30	Femenino	2020-04-01
COVID-19 CONFIRMADO	41-49	Femenino	2020-04-01

Confirmados  
01/03/2020

Confirmados  
01/04/2020

### Iztapalapa

Nombre	Edad	Genero	Fecha Colecta
COVID-19 CONFIRMADO	50-55	Masculino	2020-05-01
COVID-19 CONFIRMADO	50-55	Masculino	2020-05-01
COVID-19 CONFIRMADO	50-55	Masculino	2020-05-01
COVID-19 CONFIRMADO	50-55	Masculino	2020-05-01

Confirmados  
01/05/2020

**¿Por qué "prefiere" ciertos lugares versus otros o ciertas personas versus otras?**

Porque tiene un nicho

Un aspecto de la pandemia/una pregunta queremos contestar:

- Municipios dónde no hay confirmados pero habrá
- Municipios dónde habrá incrementos
- Cuadro de síntomas que es mas acertado para positivos
- Cuadro de síntomas que es mas acertado para difuntos
- Hospitales dónde habrá riesgo de falta de insumos
- Municipios donde...

Preguntas sobre:

¿Dónde? ¿Cuándo?  
versus ¿Quien?

$$P(C(t) | X(t'))$$

Se necesita poder representar la pregunta como una clase – en un ensemble de lugares (países, estados, municipios, AGEs, calles,...) o de personas

$X(t) = (X_1(t), X_2(t), \dots, X_N(t))$ , el universo de factores (de riesgo) que afectan  $C(t)$

- demografía, socioeconomía, historia
- clínica, contaminación, movilidad,
- genética, insumos de salud, clima,...

Si  $P(C(t) | X(t')) > P(C(t))$  entonces  $X(t)$  es nicho para  $C$  y si  $P(C(t) | X(t')) < P(C(t))$  es anti-nicho.



# Para "Dónde" – EPIPUMA Para "Quién" – Proyecto 42

Trabajando eventualmente para tener una única Plataforma

Plataforma como un servicio (PaaS)

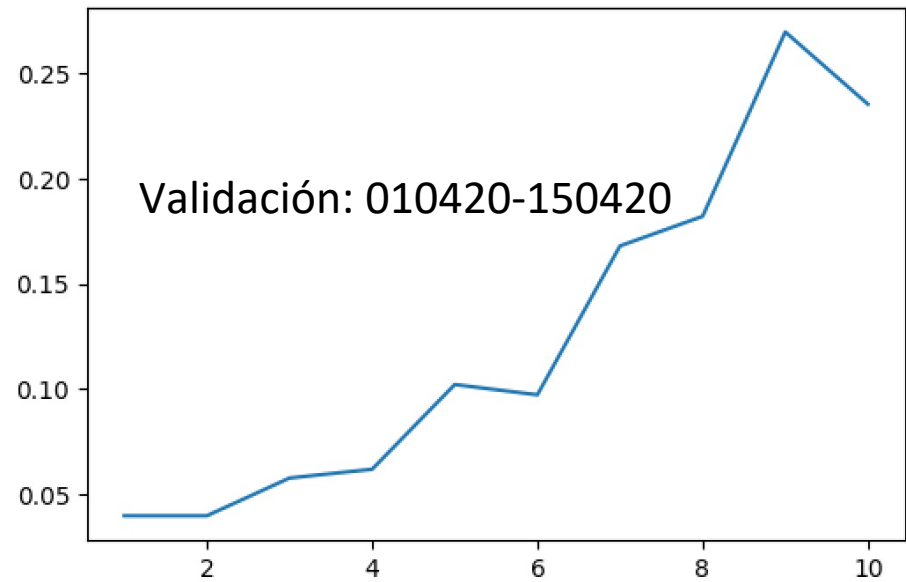
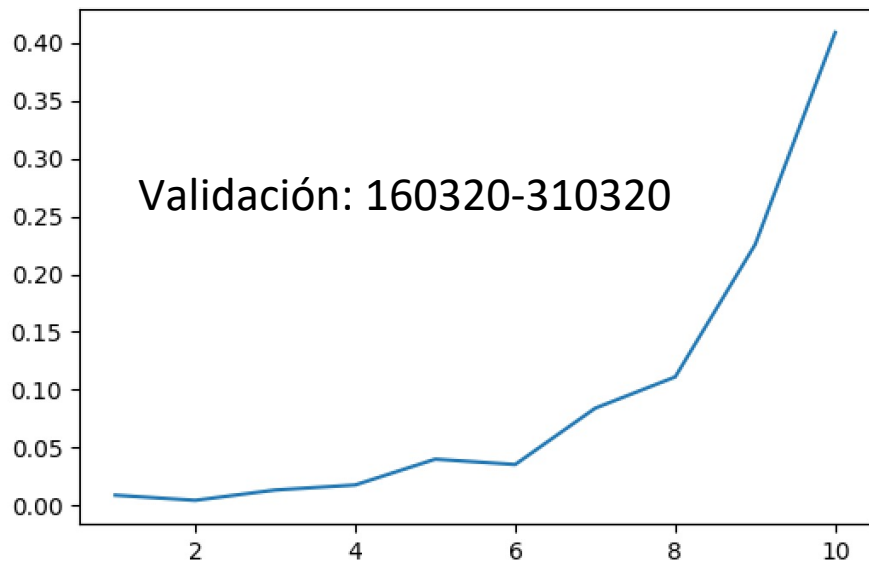
Disponible a todo público (siempre y cuando no hay datos  
"privados")

Tratando de dar oportunidad para cada persona ser un científico  
de datos

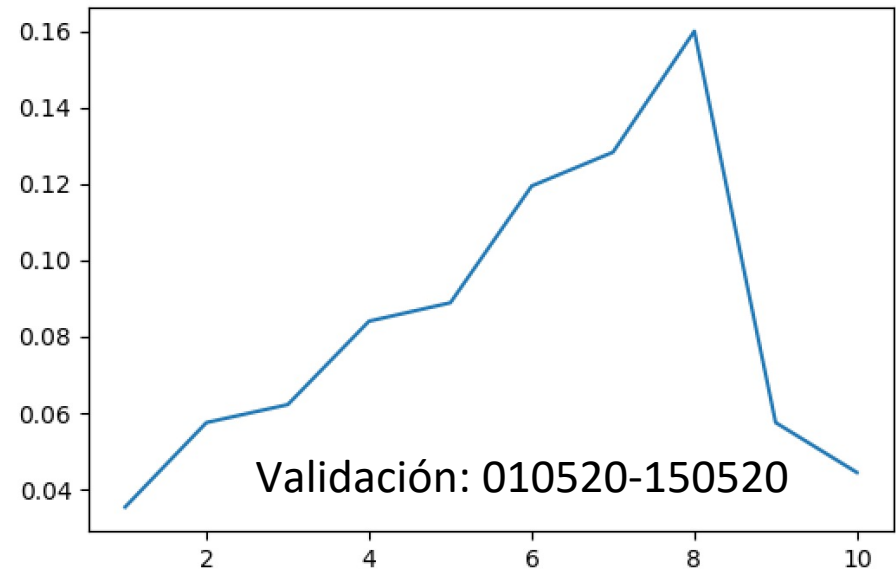
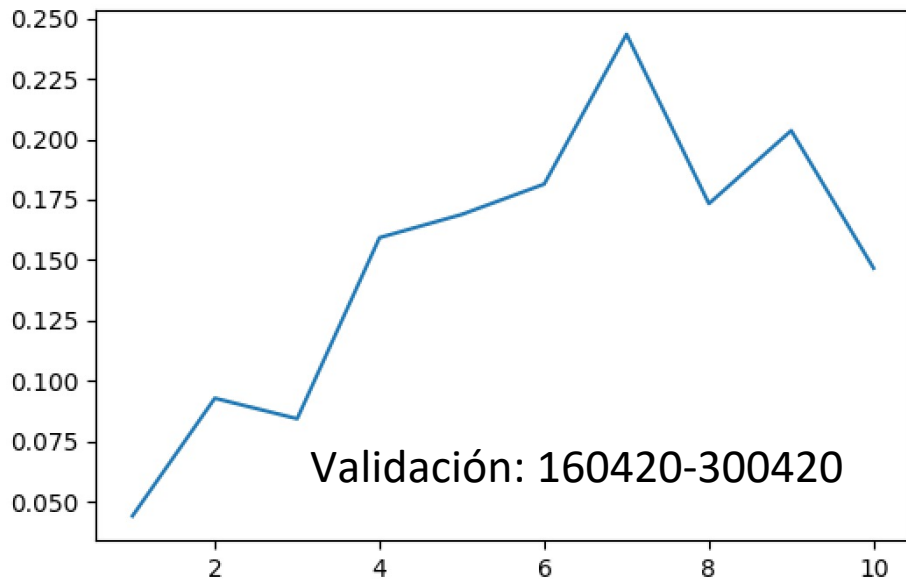
# Vamos a ver EPIPUMA

Plataforma como un servicio (PaaS)

Disponible a todo público (siempre y cuando no hay datos  
“privados”)

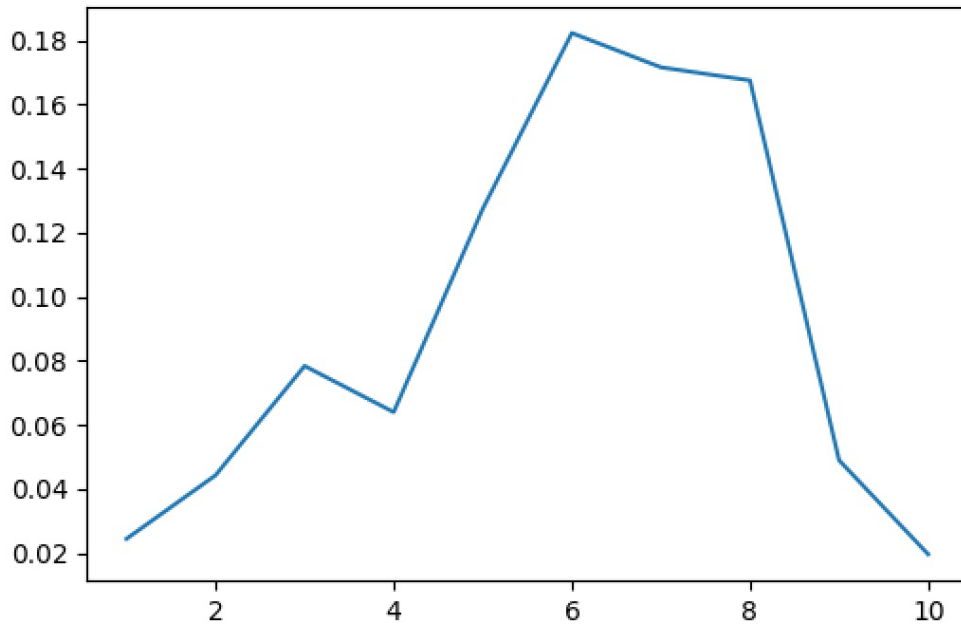


Desempeño de un modelo para predecir confirmados hacia el futuro.  
 Periodo de entrenamiento: 010120-150320

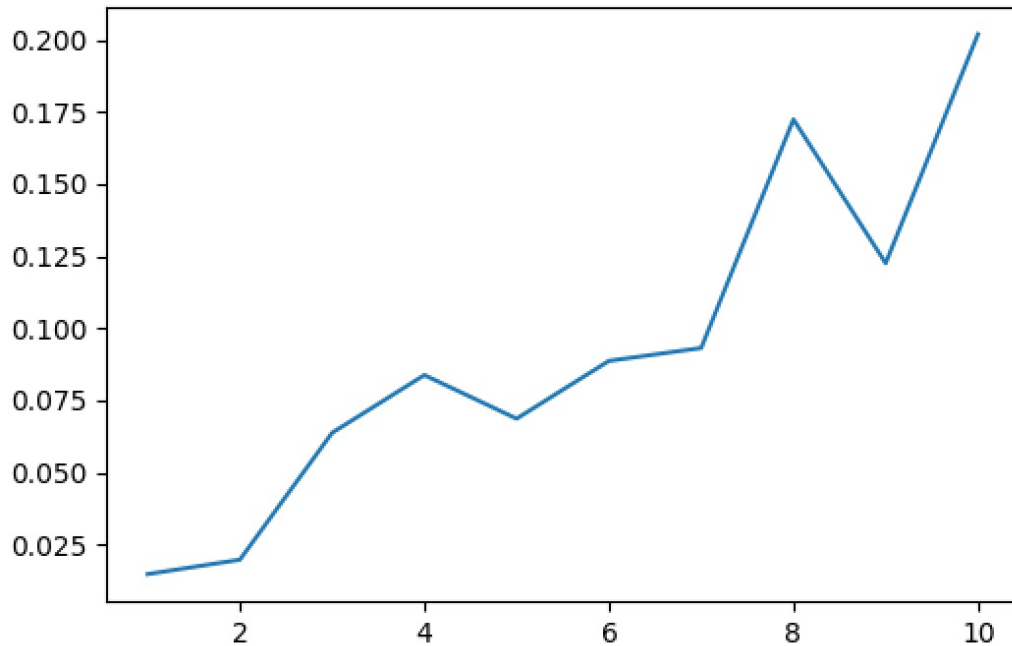


# COVID19 se adapta

Entonces se necesitan modelos que adaptan. ¿Cómo lo tomamos en cuenta? Con-sin memoria – la belleza del Bayesiano



Desempeño del modelo con  
entrenamiento del 010120-150320  
y validación 010520-150520



Desempeño del modelo con  
entrenamiento del 150420-300420  
y validación 010520-150520

# Proyecto 42

## El ensemble de “quien”



Modelo para predecir quien es positivo para la prueba de SARS-Cov-2

Tenemos un modelo de riesgo para difuntos, un modelo para distinguir entre COVID-19 y otras enfermedades respiratorias, un modelo para optimizar el diagnóstico de COVID-19 a partir de los síntomas expuestas, y...

# ¿Qué falta?

## Todo, por supuesto

Preguntas por contestar (capas por agregar)

Insumos – camas, ventiladores, medicos, tasas de cambio, magnitudes,...

Predictores para dar la respuesta (capas por agregar) –

Hay tanto – repositorio Universitario de datos de COVID-19 (no de artículos)

Modelos para entender – ¿por qué “viviendas con internet” era tan importante?

Funcionalidades – Redes Complejas Inferenciales; diferentes UX para diferentes usuarios; modelos dinámicos SIRS etc.;

# Virtudes de EPISPECIES y PROYECTO 42

PaaS – disponible a todos

Integrador de datos de múltiples fuentes, tipos y disciplinas

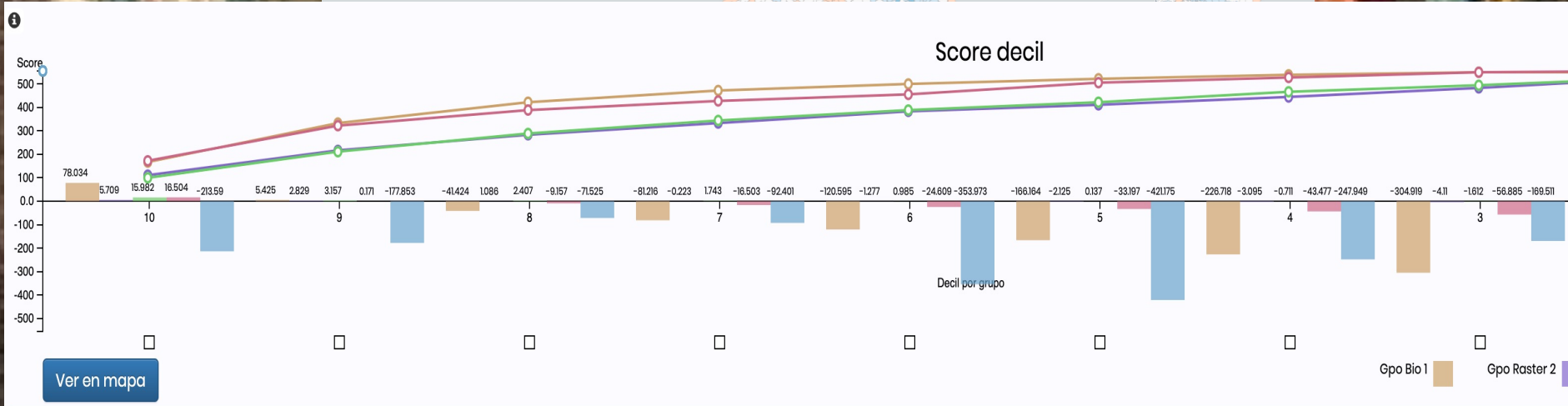
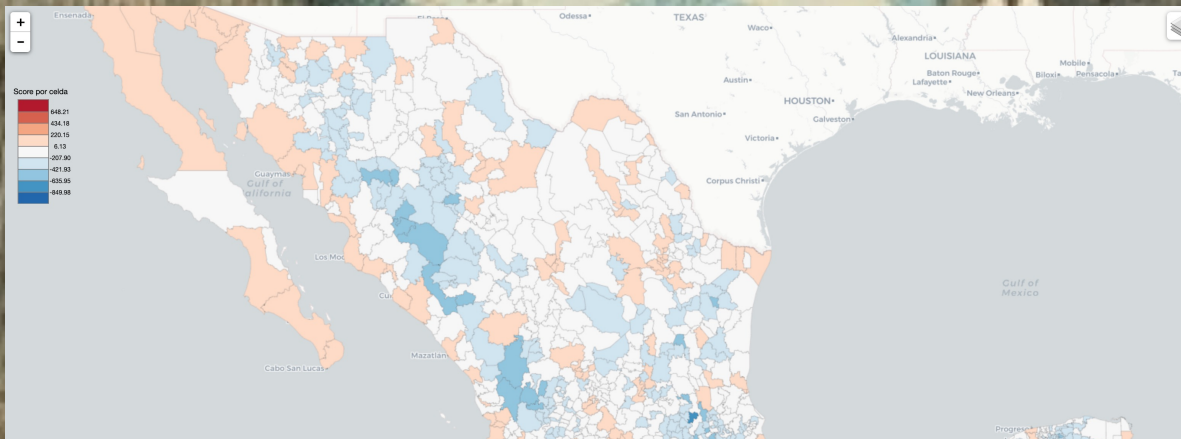
Modelos (hipotesis) con cientos o hasta miles de variables  
en segundos o minutos

Modelos adaptativos

Modelos que se puede validar

No se necesita experiencia en modelado/IA/AM





10	El grado de centralidad total del municipio, considerando el flujo laboral hacia adentro y hacia afuera. 0.053:0.416	23.58	2.47	63.49%	60.47%
10	El número de municipios totales conectados con ese municipio. 130:1018	23.58	2.47	63.49%	60.47%
10	Viviendas particulares habitadas que disponen de computadora 26.4%:75.4%	23.15	2.44	79.01%	75.89%
10	Población de 18 años y más con educación pos-básica 22.7%:62.1%	21.64	2.35	74.49%	71.54%
10	Población masculina de 18 años y más con educación pos-básica 22.9%:63.6%	21.58	2.34	72.95%	70.36%

# Sitio de EPISPECIES

[http://covid19.c3.unam.mx/geoportal\\_v0.1.html](http://covid19.c3.unam.mx/geoportal_v0.1.html)

Advertencia: Es version beta – no garantizamos ni la funcionalidad ni los datos

## ¿Quieres colaborar?

Contactar:

Chris Stephens

[stephens@nucleares.unam.mx](mailto:stephens@nucleares.unam.mx)

WhatsApp +52 5538151775

¡Tenga paciencia!


1. Desarrollar la Base de Datos más profunda en el planeta para la obesidad y las ECNT
  - i. Datos de más de 3000 universitarios
  - ii. Más de 2500 variables - desde la genética, lo fisiológico, antropométrico, psicológico, neuro-psicológico, clínico, sociológico, epidemiológico,...
  - iii. Seguir aumentando los datos con nuevas variables y seguimiento de variables ya colectadas
2. Modelación de “nicho”
  - i. Tanto variables de la persona como su ambiente
3. Hacer disponible estos datos a la comunidad de médicos y investigadores
  - i. Convertir cualquier médico o investigador a un científico de datos
  - ii. Proporcionar una Plataforma de análisis de datos usando herramientas de Inteligencia Artificial


# PROYECTO 42


## ¿QUÉ ES PROYECTO 42?


Es un Proyecto de investigación que nació de la necesidad de conocer más a fondo las causas del Síndrome Metabólico (hipertensión, dislipidemia, diabetes/prediabetes, obesidad) y el sobrepeso en México; así como formar modelos matemáticos que nos permitan determinar personalizar una escala de riesgo para desarrollar estos padecimientos tempranamente.

REGISTRARSE


 DASHBOARD


 EXPLORE


 NAIVE-BAYES


 NETWORKS

 POBLACIONES\*

 ESPACIO DE TRABAJO\*

 MODELOS\*

 TUTORIAL\*

 PERFIL\*

## ELIGE UNA POBLACIÓN



### Naive Bayes

Select you dependent and independent variables to build your model

#### Select Dependent Variables

glu (glucosa) x

#### Select Independent Variables

edad (edad\_paciente) x idsexo (sexo) x idedociv (estado\_civil) x idgestud (grado\_estudios) x talla (talla\_real) x  
peso\_normal (tiene\_peso\_normal) x sobrepeso (tiene\_sobrepeso) x obesidad (tiene\_obesidad) x  
cintura (circunferencia\_cintura) x mad\_diab (tiene\_diab\_madre) x pad\_diab (tiene\_diab\_padre) x  
mad\_sobr (tiene\_sobrepeso\_u\_obes\_madre) x pad\_sobr (tiene\_sobrepeso\_u\_obes\_padre) x

Run Model

#### Filtros

glu > 110

Add Filter










Se selecciona un trastorno/enfermedad - en este caso prediabetes - o otro grupo que quieres predecir y su selección de factores de riesgo correspondiente

Table

Model outputs

ID	SUBCATEGORÍA	VALOR VARIABLE	DESCRIPCIÓN	RESPUESTA	EPSILON	SCORE	NX C	NX	PC	PX C
1	Datos Personales	≥ 59.5	Edad del Paciente	≥ 59.5	3.89	0.69	26	116	0.11	0.22
1	Datos Personales	42.5 - 46.5	Edad del Paciente	42.5 - 46.5	0.25	-0.02	13	110	0.11	0.12
1	Datos Personales	38.5 - 42.5	Edad del Paciente	38.5 - 42.5	1.25	0.22	15	100	0.11	0.15
1	Datos Personales	54.5 - 59.5	Edad del Paciente	54.5 - 59.5	4.04	0.72	26	113	0.11	0.23
1	Datos Personales	46.5 - 50.5	Edad del Paciente	46.5 - 50.5	-0.13	-0.13	11	103	0.11	0.11
1	Datos Personales	50.5 - 54.5	Edad del Paciente	50.5 - 54.5	2.77	0.53	19	95	0.11	0.2
1	Datos Personales	34.5 - 38.5	Edad del Paciente	34.5 - 38.5	-2.16	-0.91	5	109	0.11	0.05
1	Datos Personales	30.5 - 34.5	Edad del Paciente	30.5 - 34.5	-3.59	-2.07	1	121	0.11	0.01
1	Datos Personales	26.5 - 30.5	Edad del Paciente	26.5 - 30.5	-2.75	-1.3	3	108	0.11	0.03
1	Datos Personales	< 26.5	Edad del Paciente	< 26.5	-3.53	-2.52	0	100	0.11	0
2	Datos Personales	M	Sexo	MASCULINO	0.22	0.01	44	385	0.11	0.11
2	Datos Personales	F	Sexo	FEMENINO	-0.15	-0.03	75	689	0.11	0.11
2	Datos Personales	O	Sexo	OTRO	-0.35	-0.74	0	1	0.11	0
3	Datos Personales	D	Estado civil	DIVORCIADO	-0.04	-0.16	6	55	0.11	0.11
3	Datos Personales	U	Estado civil	UNIÓN LIBRE	-0.94	-0.53	4	56	0.11	0.07
3	Datos Personales	S	Estado civil	SEPARADO	-1.85	-0.31	41	486	0.11	0.08
3	Datos Personales	C	Estado civil	CASADO	2.02	0.24	65	464	0.11	0.14
3	Datos Personales	O	Estado civil	OTRO	1.26	0.02	2	8	0.11	0.25
3	Datos Personales	P	Estado civil	P	0.44	-0.32	1	6	0.11	0.17

El sistema da el perfil de riesgo correspondiente

-  DASHBOARD
-  EXPLORE
-  NAIVE-BAYES
-  NETWORKS
-  POBLACIONES\*
-  ESPACIO DE TRABAJO\*
-  MODELOS\*
-  TUTORIAL\*
-  PERFIL\*

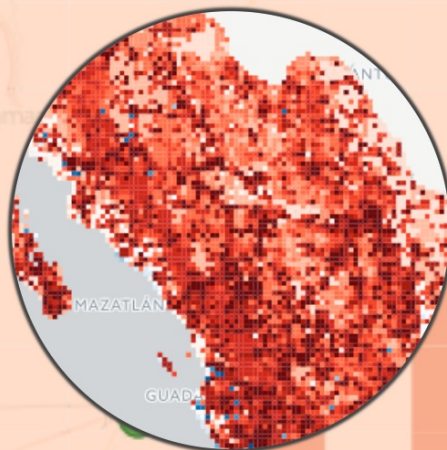
5	Antropometría	≥ 1.745	Talla	≥ 1.745	-0.24	-0.17	10	97	0.11	0.1
5	Antropometría	< 1.495	Talla	< 1.495	3.28	0.61	22	104	0.11	0.21
5	Antropometría	1.705 - 1.745	Talla	1.705 - 1.745	-0.53	-0.26	9	96	0.11	0.09
5	Antropometría	1.605 - 1.635	Talla	1.605 - 1.635	-0.05	-0.1	12	110	0.11	0.11
5	Antropometría	N/A	Talla	?	1.75	-0.13	1	2	0.11	0.5
6	Antropometría	0.0	Tiene un valor de IMC entre 18.5 y 24.9 (Peso Normal)	NO	2.97	0.3	97	660	0.11	0.15
6	Antropometría	1.0	Tiene un valor de IMC entre 18.5 y 24.9 (Peso Normal)	SI	-3.88	-0.84	21	413	0.11	0.05
6	Antropometría	N/A	Tiene un valor de IMC entre 18.5 y 24.9 (Peso Normal)	nan	1.75	-0.13	1	2	0.11	0.5
7	Antropometría	1.0	Tiene un valor de IMC mayor o igual a 25 (Sobrepeso)	SI	3.31	0.34	95	624	0.11	0.15
7	Antropometría	0.0	Tiene un valor de IMC mayor o igual a 25 (Sobrepeso)	NO	-4.02	-0.83	23	449	0.11	0.05
7	Antropometría	N/A	Tiene un valor de IMC mayor o igual a 25 (Sobrepeso)	nan	1.75	-0.13	1	2	0.11	0.5
8	Antropometría	1.0	Tiene un valor de IMC mayor o igual a 30 (Obesidad)	SI	3.81	0.56	43	226	0.11	0.19
8	Antropometría	0.0	Tiene un valor de IMC mayor o igual a 30 (Obesidad)	NO	-2.05	-0.26	75	847	0.11	0.09
8	Antropometría	N/A	Tiene un valor de IMC mayor o igual a 30 (Obesidad)	nan	1.75	-0.13	1	2	0.11	0.5
9	Antropometría	100.5 - 106.25	Cintura	100.5 - 106.25	2.08	0.4	16	89	0.11	0.18
9	Antropometría	N/A	Cintura	?	2	0.3	38	253	0.11	0.15
9	Antropometría	81.8 - 84.9	Cintura	81.8 - 84.9	-0.63	-0.35	6	69	0.11	0.09
9	Antropometría	90.25 - 93.25	Cintura	90.25 - 93.25	0.35	0	11	90	0.11	0.12
9	Antropometría	77.25 - 81.8	Cintura	77.25 - 81.8	-2.09	-1.05	3	80	0.11	0.04
9	Antropometría	≥ 106.25	Cintura	≥ 106.25	2.28	0.45	16	85	0.11	0.19
9	Antropometría	93.25 - 96.75	Cintura	93.25 - 96.75	0.62	0.07	10	75	0.11	0.13



# Enfermedades transmisibles

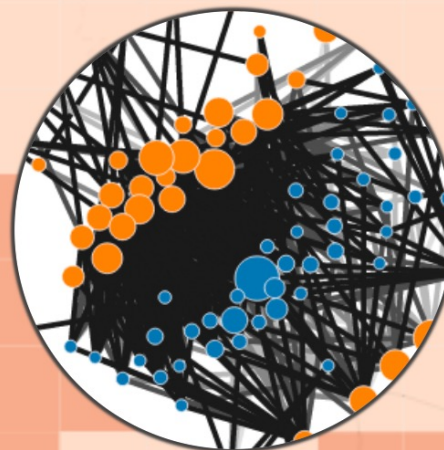
Bienvenido a la Plataforma de exploración de datos ecológicos del  
C3 y la CONABIO

Modelar Nicho  
ecológico



Mapa de presencia

Modelar Comunidad  
ecológica



Redes

**Taxones**

Región: MEXICO Resolución: 16 km

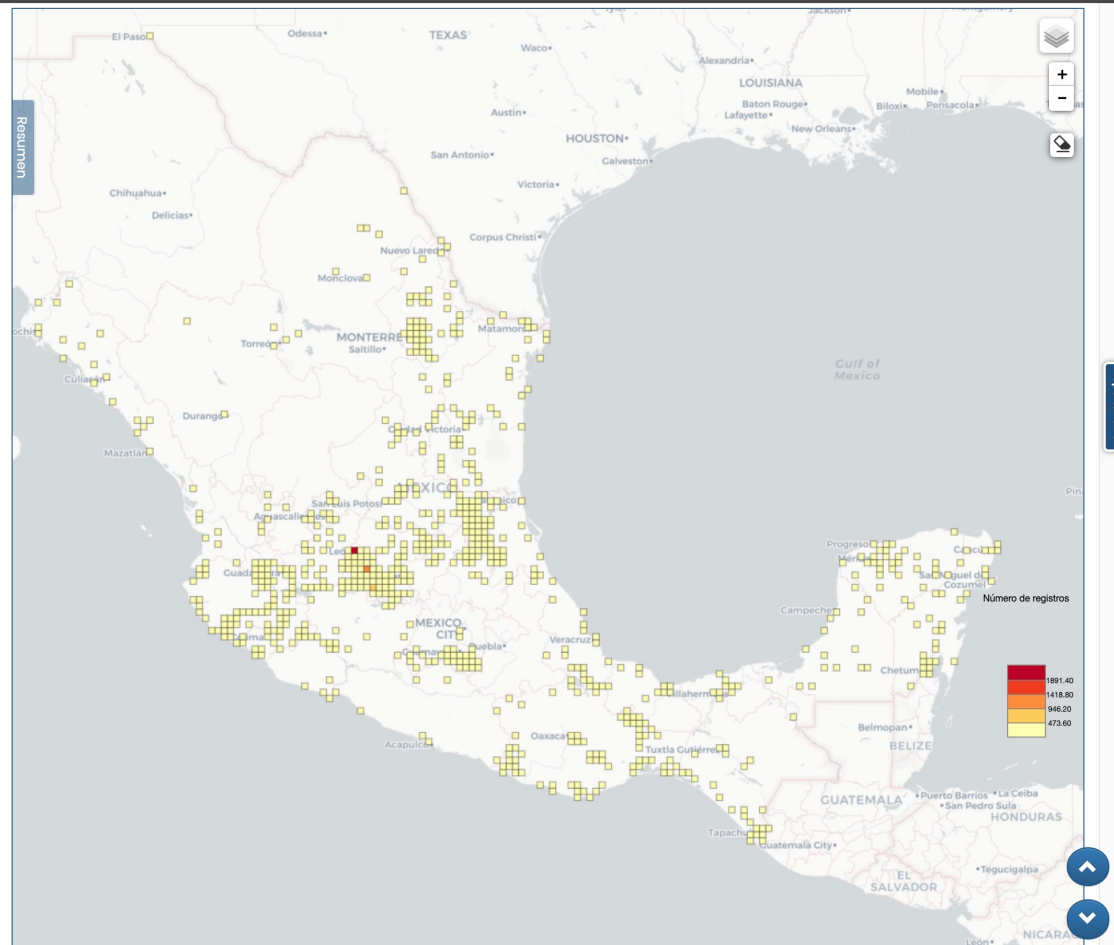
Taxonómicas  
Género ▾

Gpo Bio 1

Especie >> *Aedes aegypti*

Filtrar por fecha: 1500 - Actual  
 Registros sin fecha: No  
 Fósiles: No

**Aedes aegyptii puede ser casos de dengue, o coronavirus, o diabetes, o...**



### Grupo de variables

Taxonomías **Raster**

- Raster
  - worldclim
    - Temperatura media anual
    - Rango medio diario (Promedio mensual (temp max - temp min))
    - Forma isotérmica (BIO2/BIO7) (\* 100)
    - Temperatura estacional (desviación estandar \*100)
    - Temperatura máxima del mes mas caliente
    - Temperatura mínima de mes mas frio
    - Rango anual de temperatura (BIO5-BIO6)
    - Temperatura media del trimestre mas húmedo
    - Temperatura media del trimestre mas seco
    - Temperatura media del trimestre mas caliente
    - Temperatura media del trimestre mas frio
    - Precipitación anual
    - Precipitación del mes mas humedo
    - Precipitación del mes mas seco
    - Precipitación estacional (Coeficiente de variación)
    - Precipitación del trimestre mas humedo
    - Precipitación del trimestre mas seco
    - Precipitación del trimestre mas caliente
    - Precipitación del trimestre mas frío

+

Gpo Bio 1 ✕

Clase >> Mammalia

---

Gpo Raster 2 ✕

Raster >> worldclim

### Parámetros

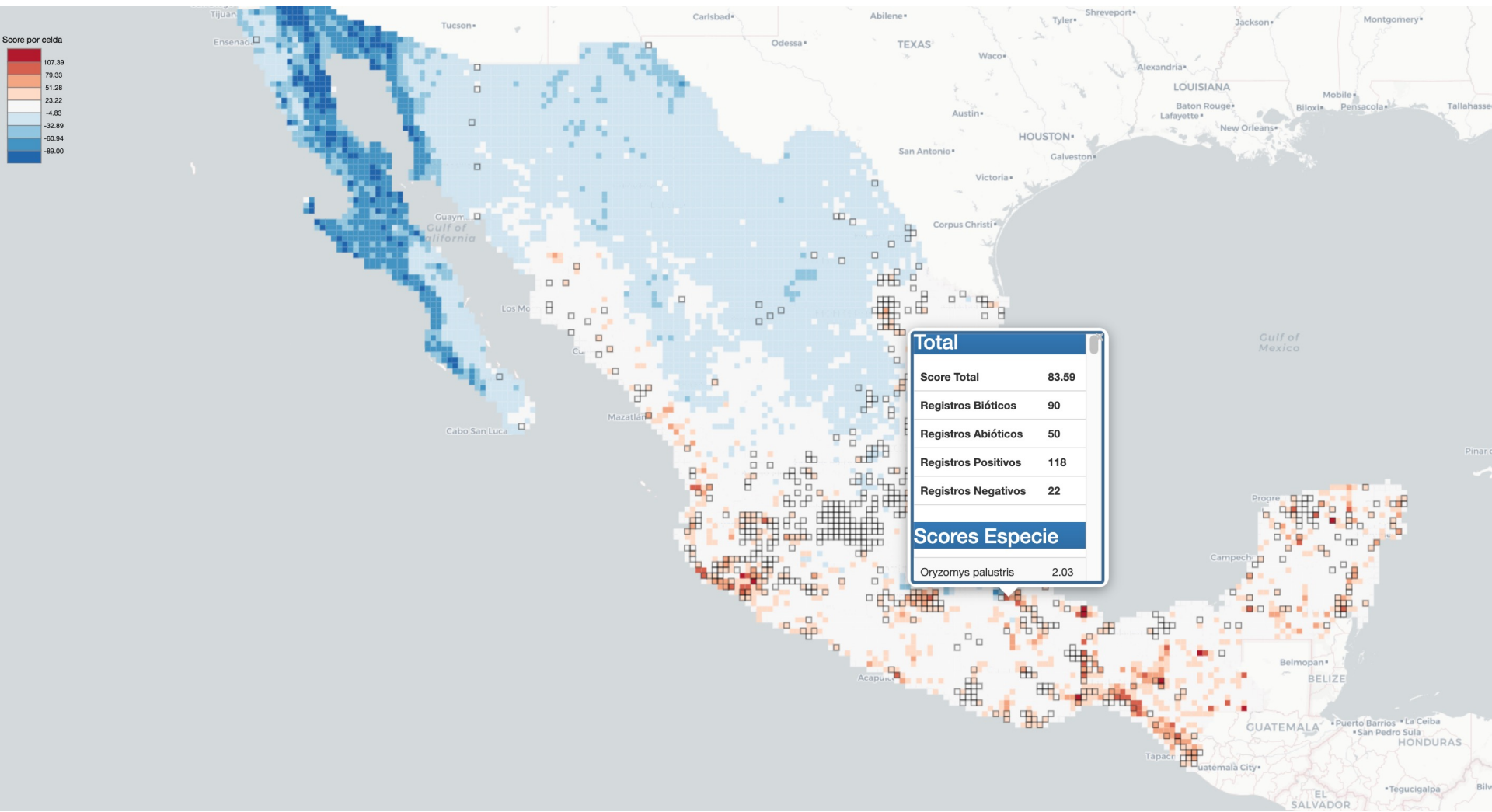
- 1 Validación: No
- 2 Min. Celdas con ocurrencia (n):  5
- 3 Añadir a priori: No
- 4 Mapa de probabilidad: No

Reportar error



Ejecutar análisis   Guarda análisis   Resultados

Se puede seleccionar como factores de riesgo hasta 60,000 factores bioticos y abioticos. Se puede incluir factores socioeconómicos, sociodemograficos, ubicaciones de clínicas, ...



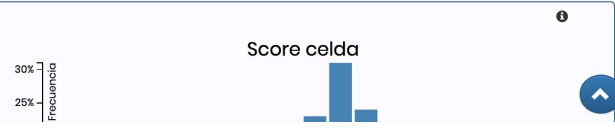
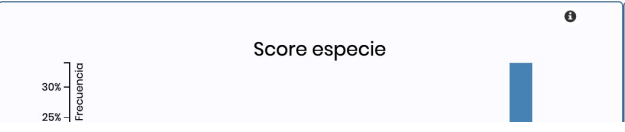
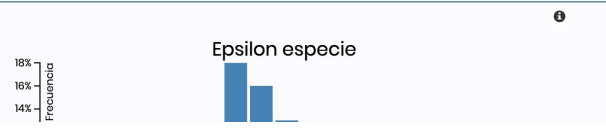


Copy CSV Excel Print

Buscar:

Decil	Variable	Epsilon	Score	Porcentaje especie	Porcentaje decil
10	Glossophaga soricina	19.96	1.49	75.00%	58.24%
10	Didelphis virginiana	19.78	1.56	63.30%	41.86%
10	Artibeus jamaicensis	18.86	1.38	72.71%	63.30%
10	Desmodus rotundus	18.25	1.43	73.08%	53.79%
10	Sigmodon hispidus	17.91	1.53	50.61%	29.42%
10	Molossus rufus	16.79	1.87	83.03%	22.75%
10	Artibeus lituratus	15.99	1.32	76.92%	55.61%
10	Heteromys irroratus	15.68	1.23	32.78%	28.21%
10	Baiomys taylori	15.43	1.47	35.79%	17.19%
10	Sciurus aureogaster	15.36	1.42	60.34%	32.46%
10	Dipodomys deserti	14.60	1.47	67.74%	30.72%

Mostrando 1 a 544 de 544 entradas



### Nodos fuente

Taxonómicas Raster

Género - Aedes

- Aedes
  - Aedes aegypti
  - Aedes albopictus
  - Aedes allotecnion
  - Aedes amabilis
  - Aedes angustivittatus
  - Aedes atactavittatus
  - Aedes atlanticus
  - Aedes atropalpus
  - Aedes bejaranoi
  - Aedes bimaculatus
  - Aedes burgeri
  - Aedes chionotum
  - Aedes condolecens
  - Aedes cozumelensis
  - Aedes diazi
  - Aedes dorsalis



Gpo Bio 1

Especie >> Aedes aegypti

Especie >> Aedes albopictus

### Nodos destino

Taxonómicas Raster

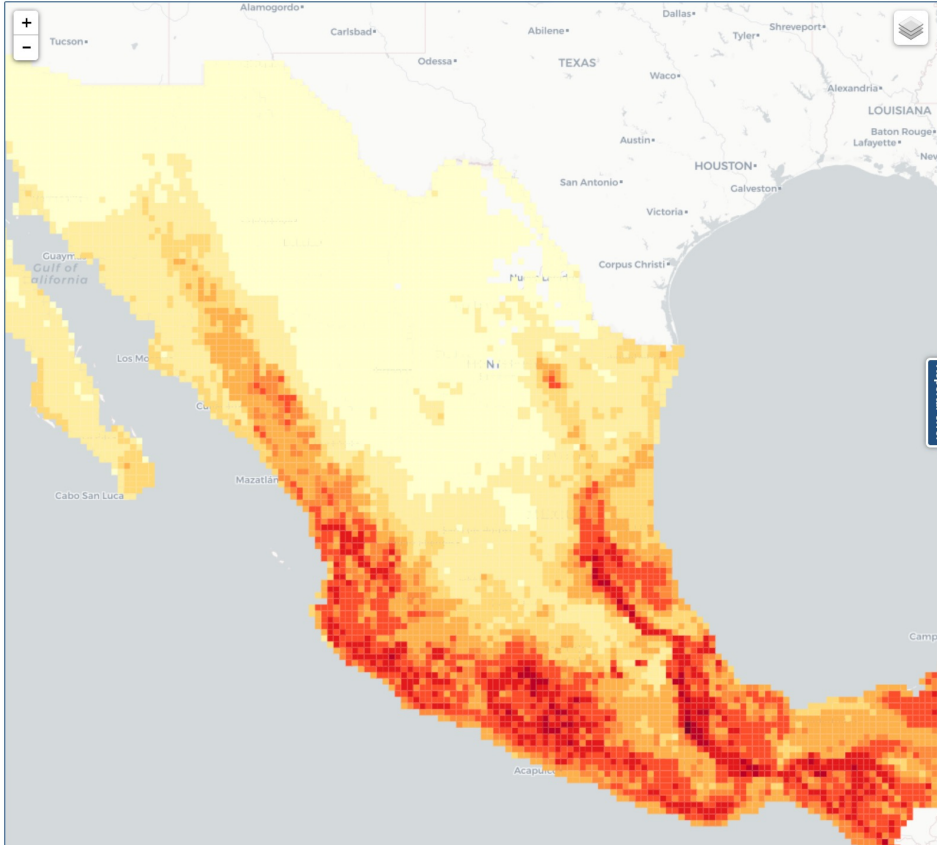
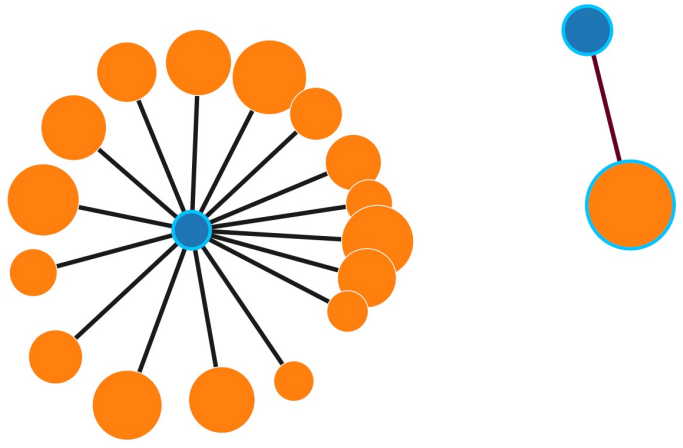
- Raster
  - worldclim
  - ENVIREM
  - USGS
  - worldclim 20 bins



Gpo Raster 1

Raster >> worldclim

Buscar especie



Copy CSV Excel Print

Buscar:

Nodo fuente	Nodo destino	nij	nj	ni	n	Epsilon
Aedes aegypti	Forma isotérmica (BIO2/BIO7) (* 100) (68.05 % : 100.00 %)	478	3245	814	9839	13.35
Aedes aegypti	Rango anual de temperatura (BIO5-BIO6) (17.30 °C : 24.50 °C)	584	4341	814	9839	12.39
Aedes aegypti	Temperatura media anual (25.66 °C : 34.13 °C)	244	1529	814	9839	10.91
Aedes aegypti	Precipitación del trimestre mas seco (78.00 mm : 137.00 mm)	184	1069	814	9839	10.61
Aedes aegypti	Temperatura media del trimestre mas frio (22.37 °C : 32.67 °C)	286	1943	814	9839	10.31
Aedes aegypti	Temperatura mínima de mes mas frio (15.20 °C : 26.50 °C)	291	2058	814	9839	9.66
Aedes aegypti	Rango medio diario (Promedio mensual (temp max - temp min)) (11.68 °C : 13.12 °C)	425	3295	814	9839	9.64



- **El IMSS tiene un sin fin de datos**
- **Hay demasiado datos para ser analizados por expertos científicos de datos**
- **Para convertir datos a inteligencia accionable requiere Plataformas de Análisis - interfaces humano-computadora - inteligentes y amigables**
- **Proyecto 42 y SPECIES-Enfermedades son prototipos de tales Plataformas de Análisis que pueden ser usados para crear y implementar un sin fin de modelos predictivos de riesgo tanto clínicos como epidemiológicos**