

# Ella es callaita..

## El Potencial del Machine Learning para la Evaluación de Políticas Públicas

Gonzalo Villa-Cox  
José Luis Castillo B.  
Mayo 31 de 2022

Breakthrough in  
Machine Learning  
will be worth ten  
Microsofts.

Bill Gates

Machine Learning



“Machine learning is  
today’s discontinuity”  
Jerry Yang, CEO, Yahoo



“Machine learning  
is the next Internet”  
(Tony Tether,  
Director, DARPA)



# Picantería Delicias del Norte

🇨🇴 🇨🇷 🇵🇷 🇪🇨 🇨🇺 🇨🇦 🇨🇱 🇨🇰 🇨🇯 🇨🇰



Regalos para Papá  
Descúbrelo todo para Papá

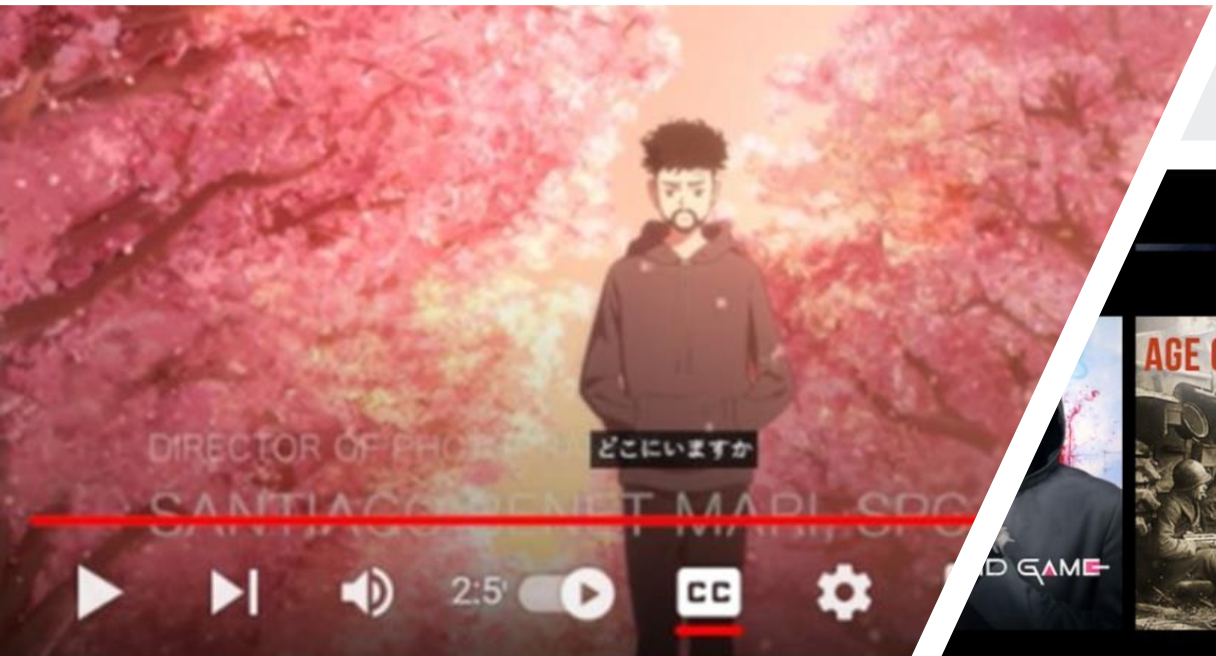
### Seguir comprando



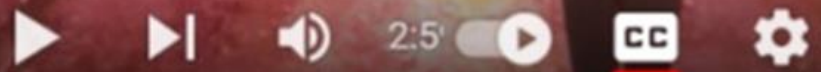
### Compra por Categoría



### Ofertas y Promociones

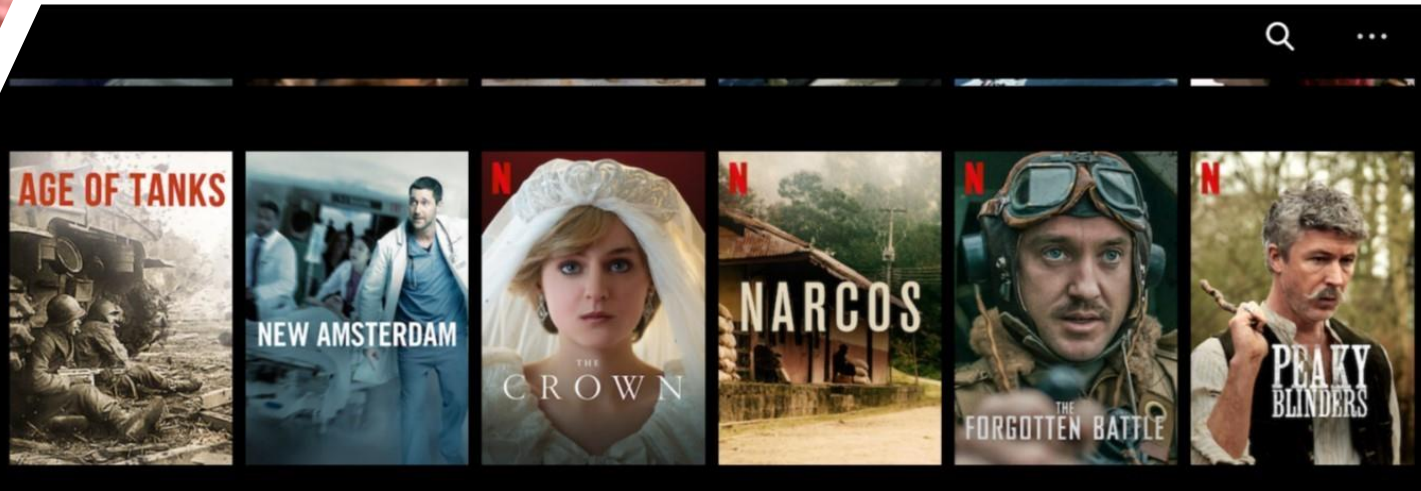


どこにいますか



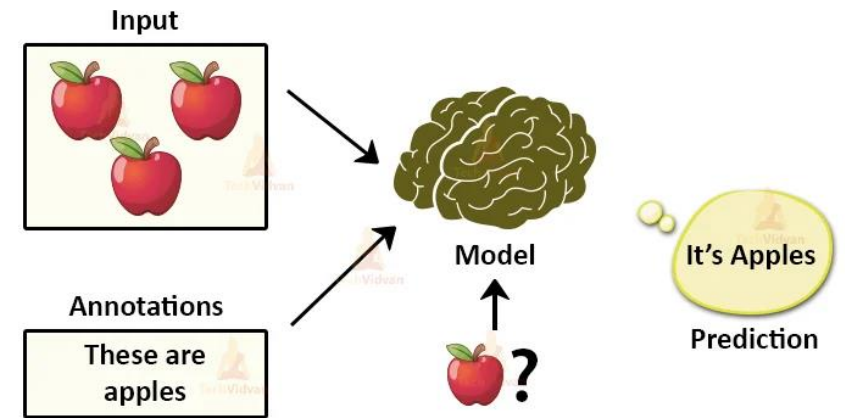
#41 TOP MUSIC VIDEO

Red Bunny Yoneguni (Official Video)

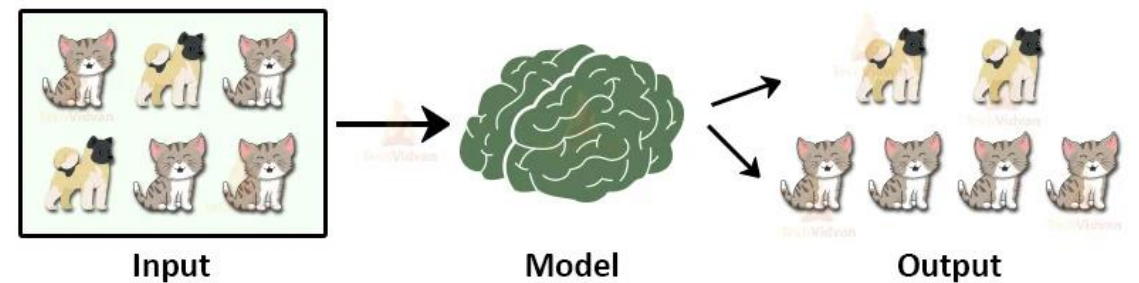


¿Cómo aprende la maquina?

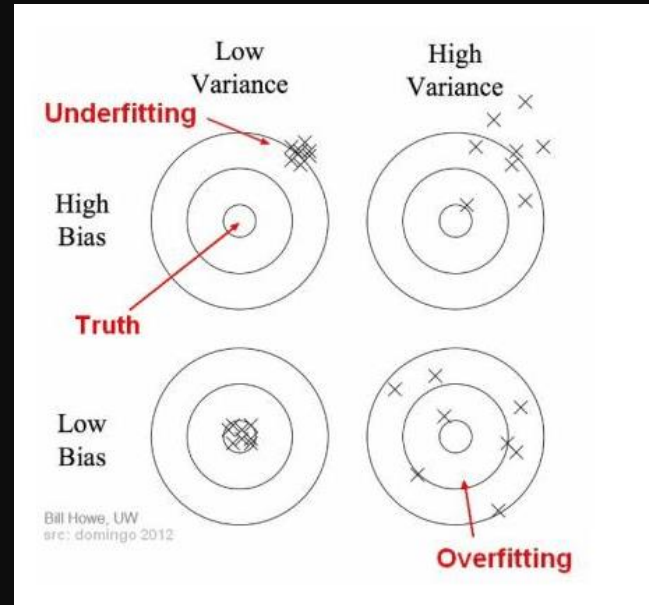
## Supervised Learning in ML



## Unsupervised Learning in ML



¿Qué características debería tener un modelo ideal?



¿Cómo se lo descubre?



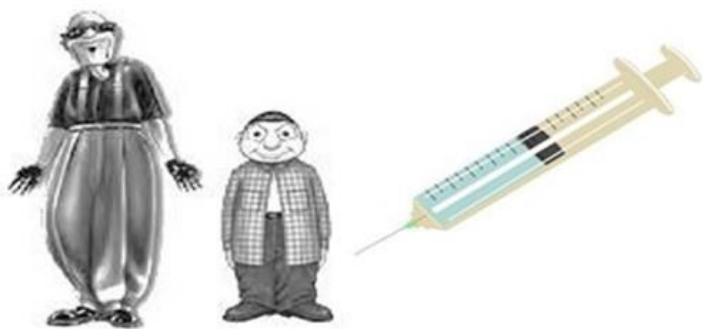
TRAINING  
+  
ACCURACY



Predecir no es lo mismo que evaluar...

# Evaluar (identificar) el impacto de un programa no es nada trivial

El impacto de un programa es la diferencia entre:



Que ocurrió con el programa

V/S



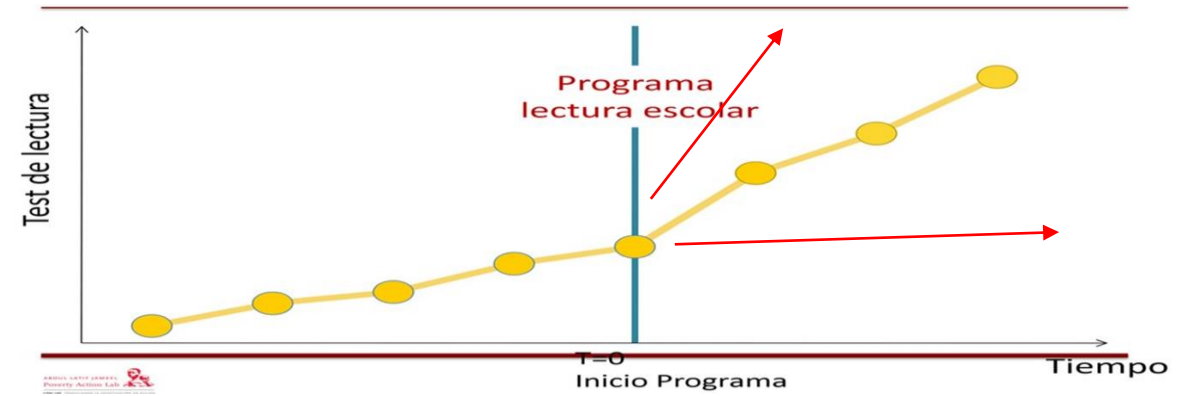
Que hubiera ocurrido sin el programa

*Contrafactual*



# 2 errores típicos

## (I) Antes vs Después



## (II) Tratados vs No Tratados

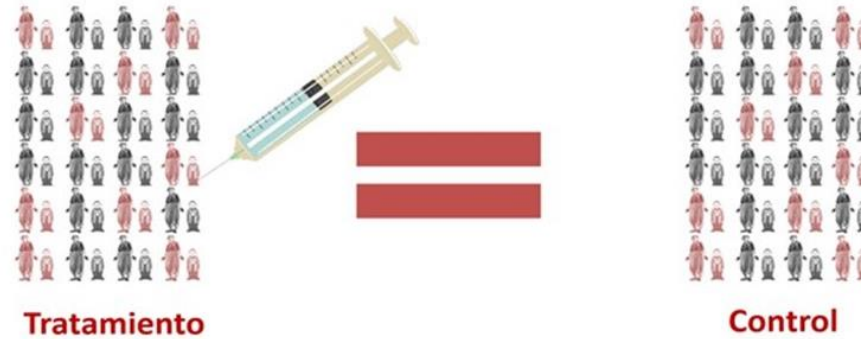


# ¿Cuál es la solución ideal?

---

## Evaluación Aleatoria:

Tome una muestra de postulantes a un programa. Asígnelos de forma aleatoria (una lotería) a un grupo de tratamiento y a un grupo de control.

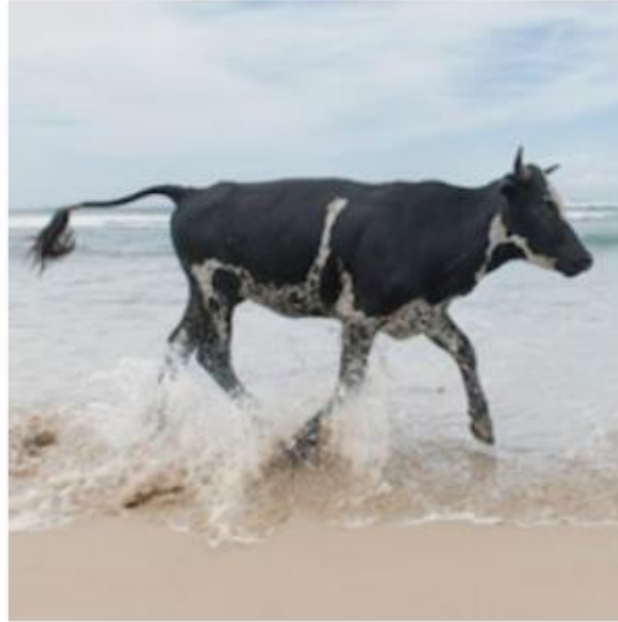


# ¿Y si no es factible implementar un RCT?

---

- Métodos No Experimentales:
  - Variables Instrumentales
  - Regresión Discontinua
  - Diferencias en Diferencias y Efectos Fijos
  - Matching, Controles Sintéticos
- Supuestos más fuertes que con RCTs
- Validez Interna vs Validez Externa





¿Puede ML  
resolver estos  
problemas?

- *Garbage in, garbage out*
- *Contextos le causan problemas a ML*



# Machine learning y Causalidad:

¿Cómo está el “estado de las artes”?

# Apuestas seguras: Causal Learning

- Métodos que dependan fuertemente de predecir o asociar unidades de estudio con alta precisión. Gonzalo Villa-Cox  
José Luis
- Métodos Intensivos en data y variables
- Por ejemplo:
  - Matching y Controles Sintéticos
  - Discontinuidades
  - 1era etapa del 2SLS IV (?)



# Servicios Ecosistémicos (SE) en la agricultura y el manejo del uso del suelo: Aplicación de *Structural Topic Modelling (STM)*.

2400+ papers publicados

$M$  Documents



Collection of Text Documents

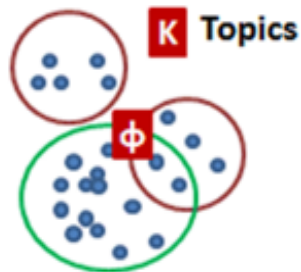
$\alpha$   
 $\beta$  Concentration Parameter

STM



$X$   
Metadata

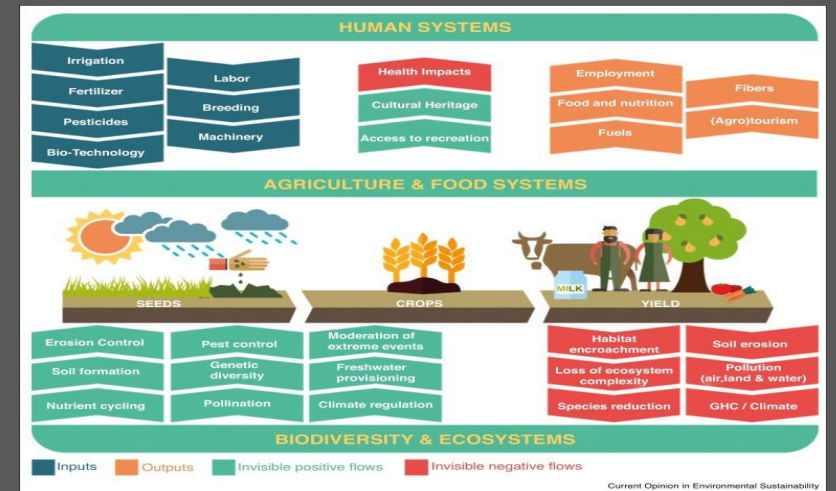
Cluster of word by topic



Cluster of document by topic

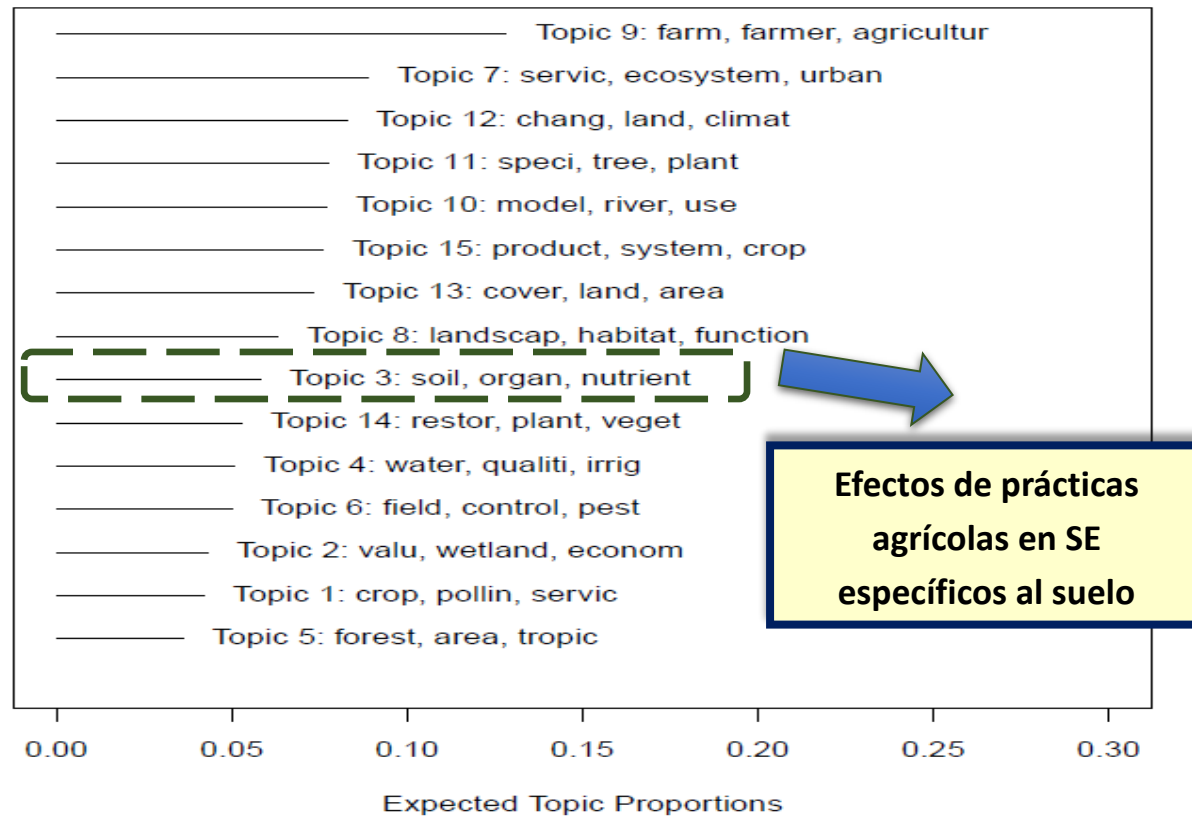
Correlated topic prevalence

Fuente de imagen: DeCleck, et. al. (2016).

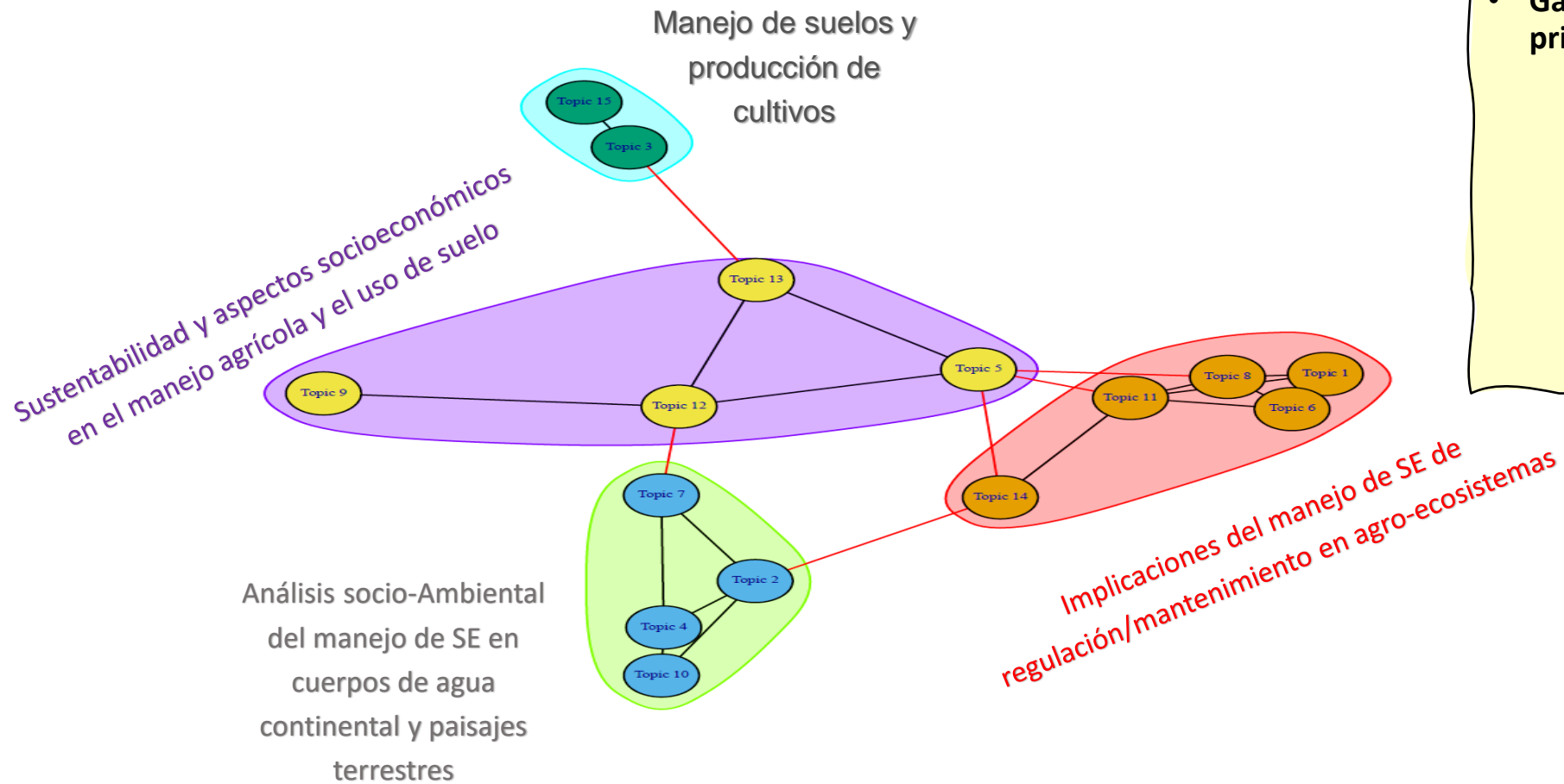


- *Structural Topic Modelling (STM)*:
  - Extracción automática de tópicos a partir de literatura científica en la temática a nivel global.
  - Estandarización de nomenclatura de SE mediante protocolo CICES v 5.1 de la Unión Europea.
  - Identificación de la estructura del discurso científico en la temática (detección de clusters de tópicos) y *gaps* en la literatura.

# Resultados principales: Identificación y prevalencia de tópicos.



# Resultados principales: Mapa de la estructura del discurso científico.



- Gaps identificados de 3 formas principales:
  - Ausencia de conexión entre tópicos dentro de la misma comunidad.
  - Tópicos conectados pertenecientes a distintas comunidades.
  - Efectos parciales de los metadatos y métricas de la red de tópicos.

# Polarización en redes sociales: *Casos de estudio en América Latina*



Que irresponsabilidad lanzar una noticia falsa en medio de esta crisis.

Usted verificó?



## Media Polarization and Election: A Nation Divided

*As divisions exist in the news sources  
people rely on*

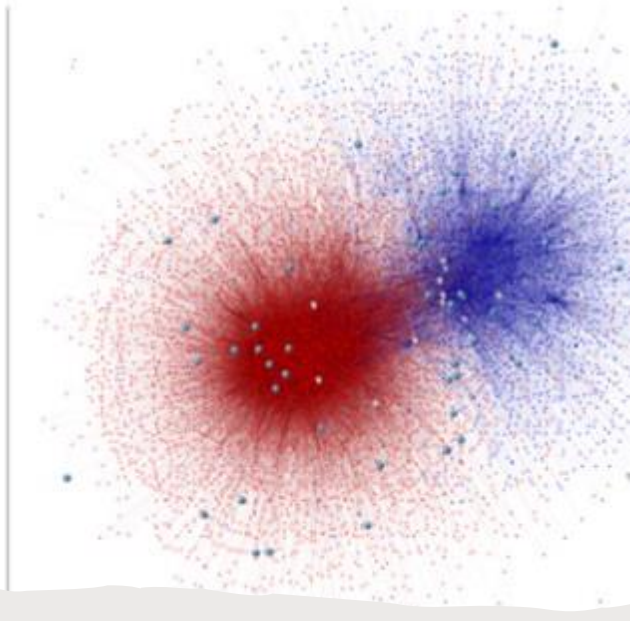
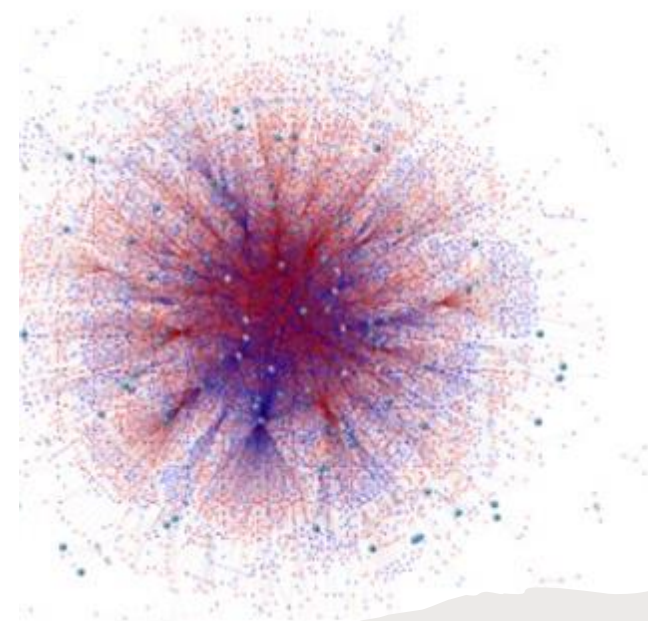
- Extracción automática de la posturas de usuarios en redes sociales durante eventos polarizados.
- *Redes Neuronales (Deep learning)* : Se entrena a nivel de texto, a nivel de usuario y a nivel de su red social.
- Predicciones son más robustas a diferentes contextos culturales o cambios de corto plazo en el evento.
- Por ejemplo, se utilizó la postura revelada por usuarios en Twitter durante las protestas en Chile en 2019 para predecir el resultado del Plebiscito en el 2020.



# Causal inference and counterfactual prediction in machine learning for actionable healthcare

Mattia Proserpio<sup>1</sup>, Yi Guo<sup>2,3</sup>, Matt Sperrin<sup>4</sup>, James S. Koopman<sup>5</sup>, Jae S. Min<sup>1</sup>, Xing He<sup>2</sup>, Shannan Rich<sup>1</sup>, Mo Wang<sup>6</sup>, Iain E. Buchan<sup>7</sup> and Jiang Bian<sup>2,3</sup>

Big data, high-performance computing, and (deep) machine learning are increasingly becoming key to precision medicine—from identifying disease risks and taking preventive measures, to making diagnoses and personalizing treatment for individuals. Predicting risks and outcomes, but also about weighing interventions. The identification of cause and effect, and the calculation of so-called counterfactuals. Observational studies are commonly affected by confounding. Domain knowledge, causal inference is not possible. But neither their parameters nor their models lead to



## ¿Cómo aterrizarlo a evaluación de programas e identificación causal?

- Laboratorios Virtuales:
  - Mediante simulaciones, se puede explorar como cambios en la arquitectura de la red social pueden afectar los niveles de polarización observados.
  - También podemos evaluar el efecto de esos niveles de polarización en la difusión de rumores.
  - En salud se ha hecho algo similar
- Se podría usar esta lógica para diseñar intervenciones de política pública en redes.

# Caminar, antes de correr..

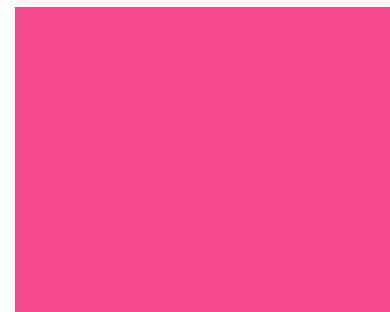
- Sistema nacional de datos administrativos
  - Liderado x esta secretaria+MEF
  - Experiencia Chilena
- Masificar el impacto de este sistema
  - Inteligencia Colectiva
    - Concursos/desafíos
    - Tercerizar
  - Data disponible para tesis
    - Con manuales



# José Luis Castillo B.



Es socio fundador de MetaData Consulting Group y presidente del Observatorio de Políticas Públicas de Guayaquil. En esas actividades mezcla su pasión y casi dos décadas de experiencia en modelamiento estadístico, problem solving y policy research. Ha sido consultor del BID, CAF, Banco Mundial y UNICEF, en temas relacionados con diseño y evaluación de programas públicos. Buena parte de su carrera la desarrolló en Chile, donde fue coautor de dos estudios actuariales donde se cuantificó el esfuerzo fiscal para financiar la reforma al sistema de pensiones. A nivel local, estuvo a cargo de desarrollar - para DINARDAP- un modelo nacional de calificación crediticia y la consultoría para el nuevo modelo de gestión del Registro Social. Ha sido profesor de métodos econométricos, business intelligence, desarrollo y políticas públicas en una docena de programas de posgrado. José Luis es candidato a doctor en policy analysis por RAND Graduate School, master en economía del envejecimiento por Tilburg University, y master en economía y políticas públicas por la Universidad de Chile, además de economista de la ESPOL.





- Econ. Gonzalo Villa-Cox, Ph. D.
  - Profesor/Investigador, Facultad de Ciencias Sociales y Humanísticas (FSCH), Escuela Superior Politécnica del Litoral (ESPOL).
  - **Correo electrónico:** [gvilla@espol.edu.ec](mailto:gvilla@espol.edu.ec)
  - **ORCID:** 0000-0002-2160-7448
- Líneas principales de Investigación:
  - Econometría aplicada a la economía agrícola y el manejo de recursos naturales.
  - Statistical Learning y Machine Learning aplicados a la economía agrícola y el manejo de recursos naturales.
  - Diseño de sistemas de Soporte a la decisión para Gestión Integral de Recursos Hídricos.
  - Societal computing y simulación de agentes heterogéneos.