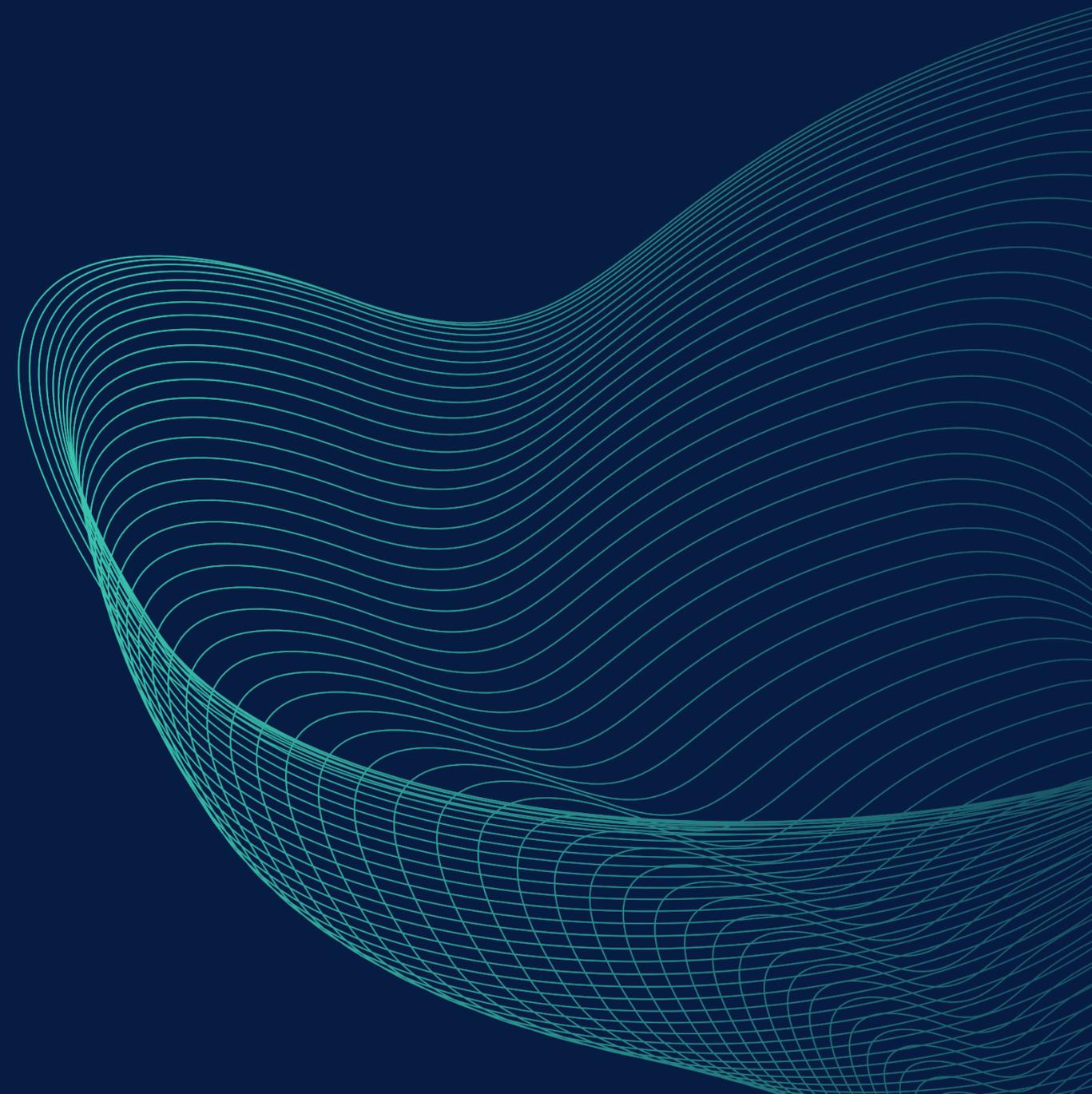


# Statistical Modeling: The Two Cultures

Leo Breiman (2001)

**Integrantes:**

Matías Bustamante, Manuel Mora, Isidora Anabalón,  
Francesca Simonetti, Yetsy Carvallo



# Sobre el Autor



**Leo Breiman**  
**1928 - 2005**

Estadístico de la Universidad de California, Berkeley

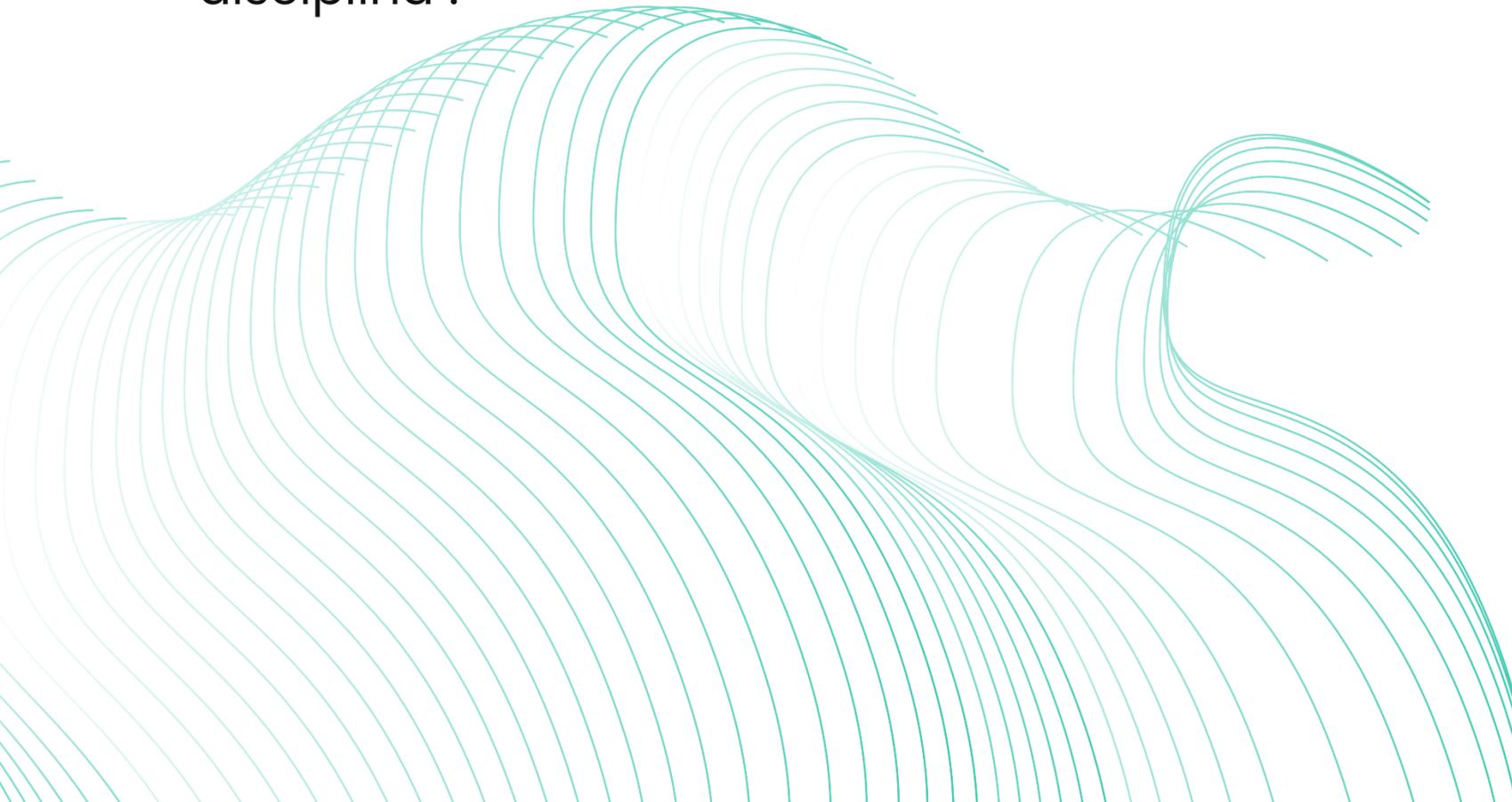
Muy influyente en el campo de Machine Learning conectando la estadística y la informática.

Escribió este artículo cuando ya estaba en retiro y reflexiona sobre cómo cambió sus técnicas de análisis en industria vs. en academia.

---

# Tesis

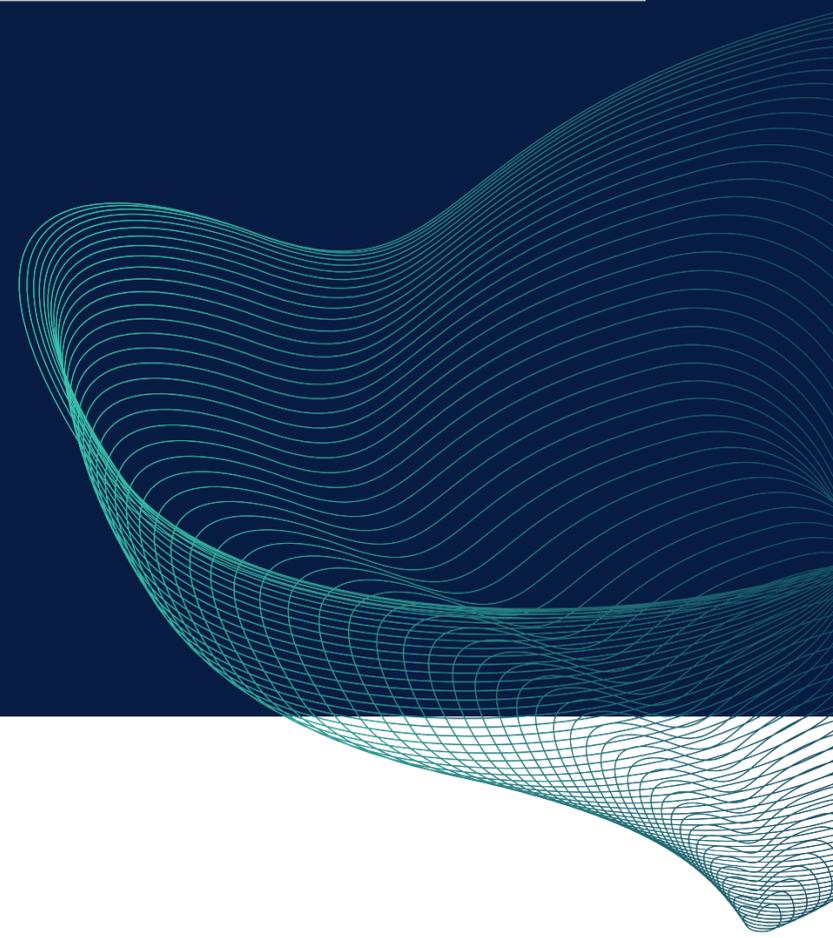
"Los estadísticos realizan análisis en base a modelos de datos, requiriendo en muchas ocasiones adaptar los datos para encajar dentro del funcionamiento del modelo. Esto ha perjudicado a la disciplina".



*"No estoy en contra de los modelos de datos per se. En algunas situaciones son la forma más adecuada de resolver el problema. Pero hay que hacer hincapié en el problema y en los datos"*

**Leo Breiman**  
pg.214

# Objetivos del análisis de datos

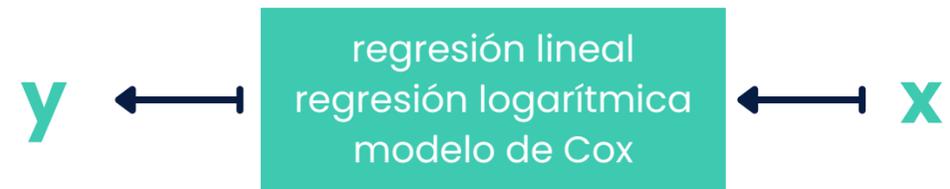


## Predicción e Información



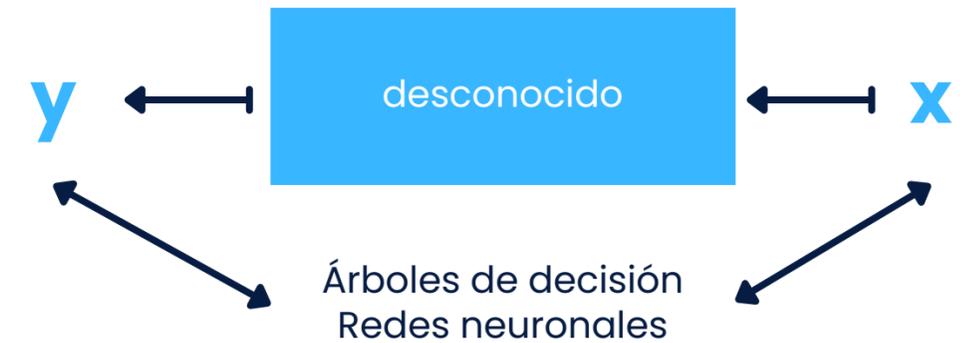
Se puede visualizar la generación de datos como la relación entre las variables independientes ( $x$ ) con la variable dependiente ( $y$ ) que están asociadas por funciones de 'naturaleza' (caja negra).

## Modelamiento Estocástico



Asume modelo de datos estocásticos  
Usado por 98% de estadísticos

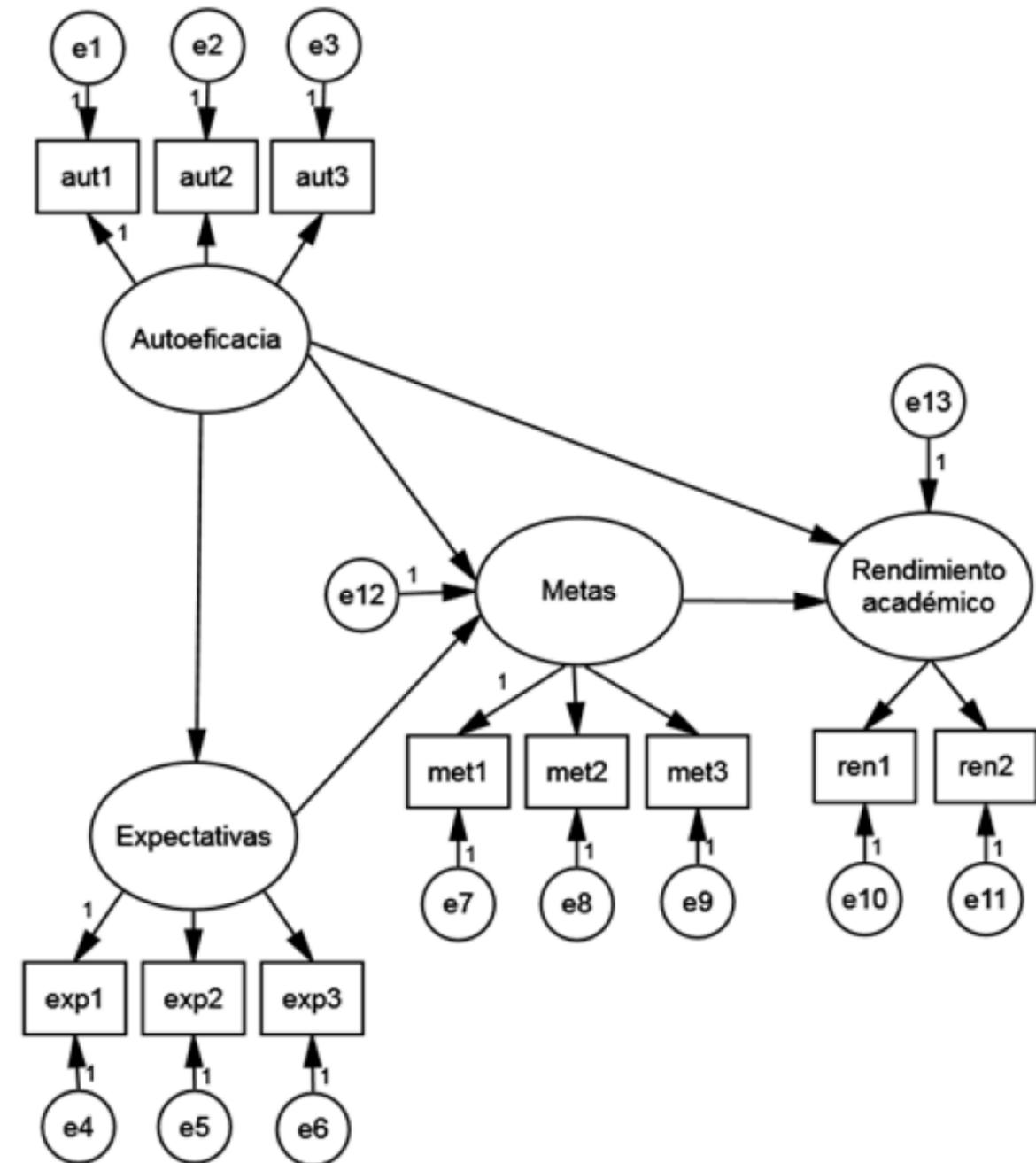
## Modelamiento Algorítmico



Considera que la caja es compleja y desconocida  
Usado por 2% de estadísticos pero más común en otros campos

# Ecuaciones estructurales

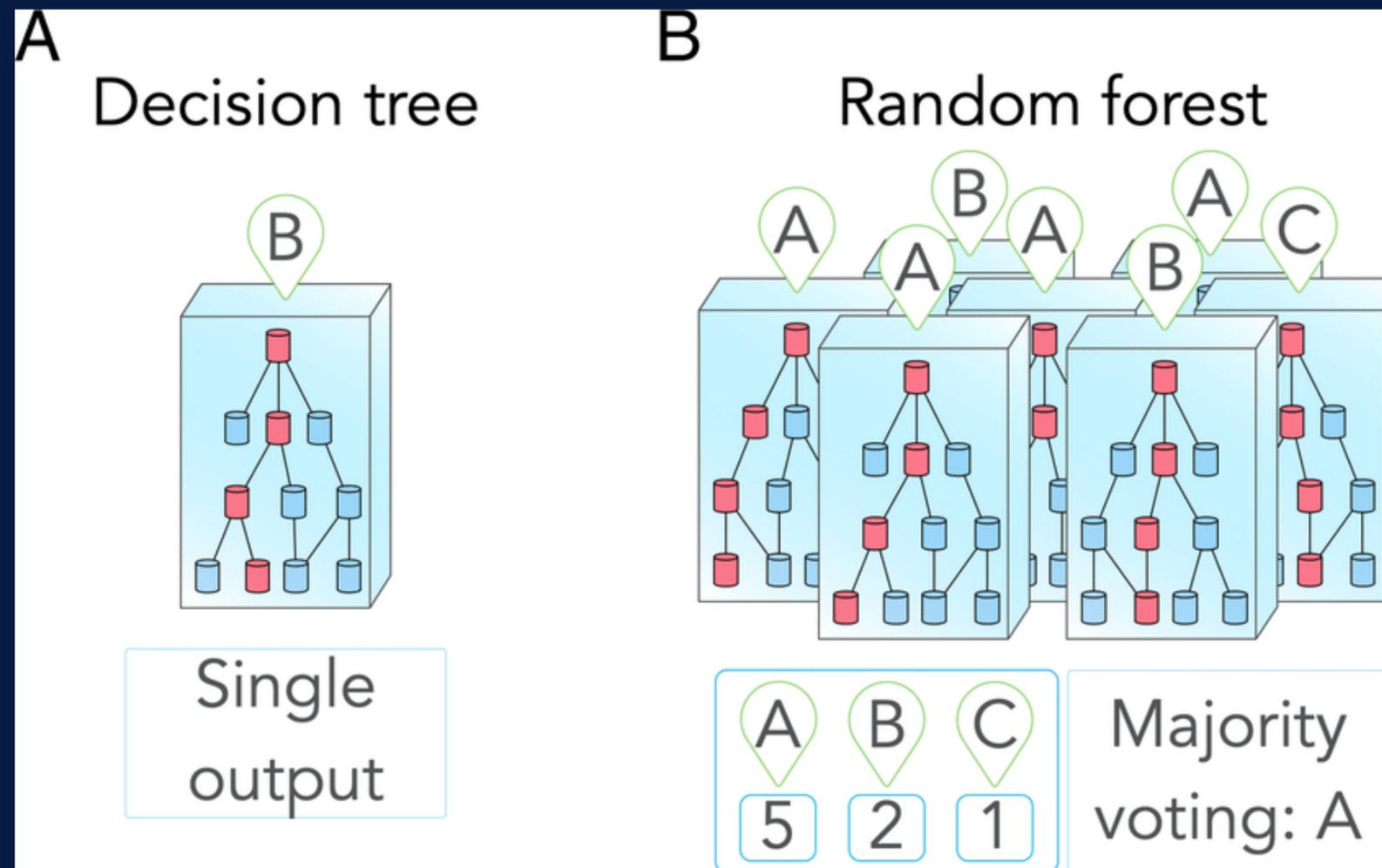
Los modelos de ecuaciones estructurales establecen la relación de dependencia entre las variables. Trata de integrar una serie de ecuaciones lineales y establecer cuáles de ellas son dependientes o independientes de otras, ya que dentro del mismo modelo las variables que pueden ser independientes en una relación pueden ser dependientes en otras.



Fuente

# Árbol de decisión / Bosque Aleatorio

Un árbol de decisión es una estructura jerárquica que utiliza una secuencia de **condiciones if-else** para tomar decisiones o realizar predicciones basadas en características de entrada.



Un bosque aleatorio es un método de **aprendizaje por conjuntos** que combina varios árboles de decisión para realizar predicciones promediando o votando los resultados de los árboles individuales.

# Pros y Contras

## Modelamiento de Datos

Ofrece una simple y comprensible descripción de la relación entre las variables independientes y las de respuestas.



Las conclusiones son sobre el mecanismo del modelo, y no sobre la naturaleza.

Las deficiencias de análisis ocurren porque se centra la atención en el modelo y no en el problema.

Hay una creencia en la infalibilidad de los modelos de datos casi religiosa.



Los datos y el problema guían la solución.

Se aprovechan un conjunto de herramientas y variables más amplias, para resolver un conjunto amplio de problemas.

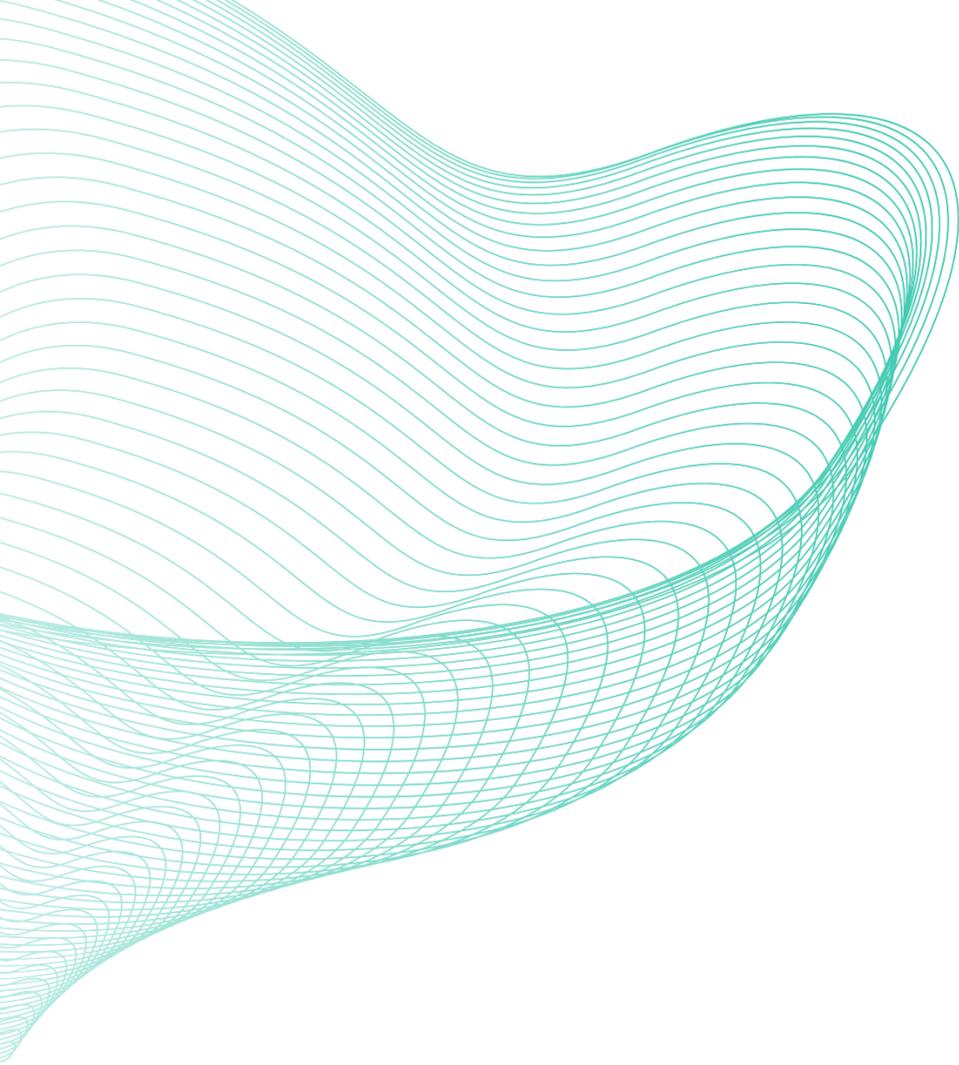
Los modelos de algoritmos tienen buena capacidad de predicción.



## Modelamiento Algorítmico

Sacrifica la interpretabilidad por la precisión en la predicción.





# Consideraciones al elegir un modelo

## **Efecto Rashomon y la multiplicidad de buenos modelos**

¿Cómo elegir el modelo adecuado para los datos que se tienen?

## **Occam y la simplicidad v/s la precisión**

¿Cómo ser simple y preciso al mismo tiempo?

## **Bellman y la maldición de la multidimensionalidad**

¿Cómo analizar una realidad compleja a partir de la reducción de sus dimensiones?

---

# Conclusiones del Paper

- La estadística destinada a la resolución de problemas, debe centrarse en la predicción.
- Los modelos algorítmicos pueden brindar una mayor precisión predictiva, por lo que la mayoría de los casos deben preferirse.
- El énfasis a la hora de investigar debe estar siempre, primero en el problema y los datos.
- Los estadísticos no se percatan de las consecuencias de su confianza en suposiciones, hipótesis Y modelos simples limitan sus resultados.

---

# Nuestras Reflexiones

- Los modelos algorítmicos son en muchos aspectos superiores a los de datos, pero ¿podemos sacrificar la interpretabilidad?
- Hay desafíos de carácter ético más fuertes en la toma de decisiones, basados en modelos algorítmicos. (Ej: privacidad de datos)
- El uso de un modelo u otro depende del problema que se quiere solucionar, para que así aporten valor práctico y conocimientos aplicables.

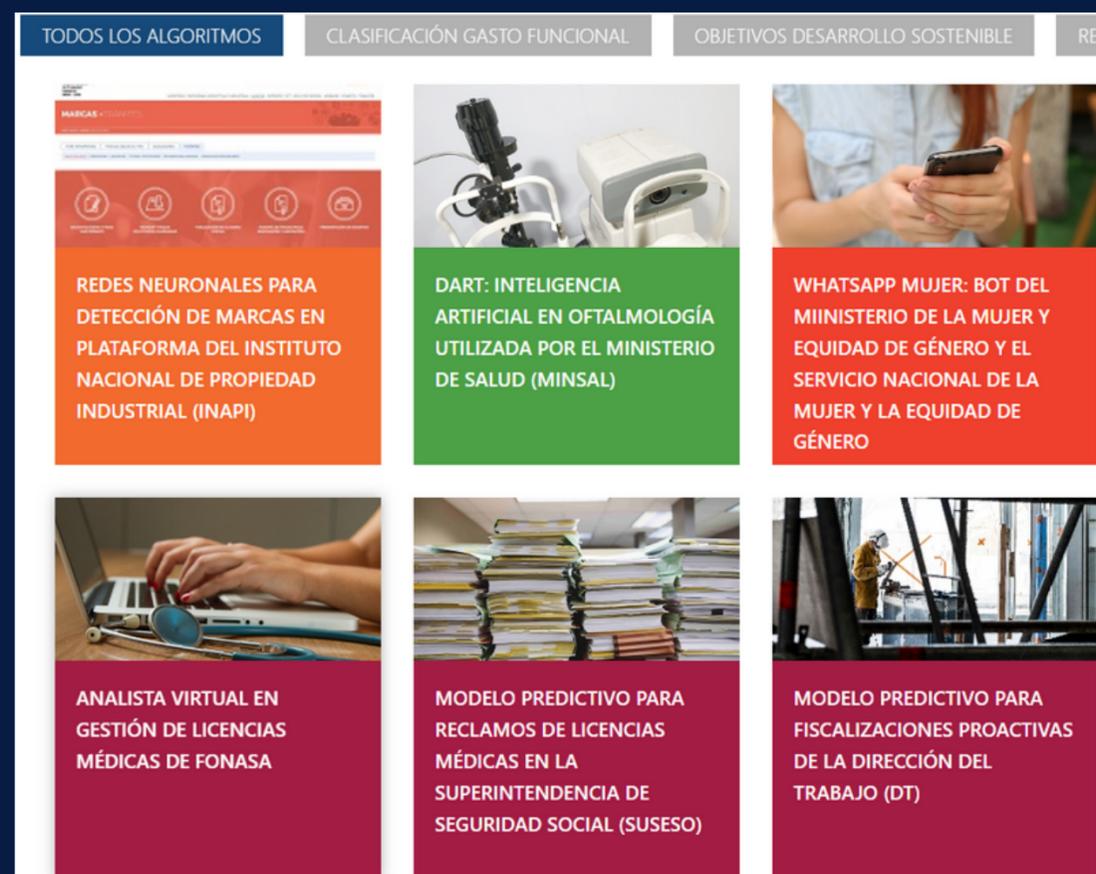
---

# Temas de Discusión





*¿Qué podría explicar el aumento en el uso de modelos algorítmicos desde la época de publicación del paper?*



[Link](#)

*¿Qué ventajas y/o desventajas tendría el uso de modelos algorítmicos para temas de políticas públicas?*

28 DE OCTUBRE DE 2021

## Chile presenta la primera Política Nacional de Inteligencia Artificial

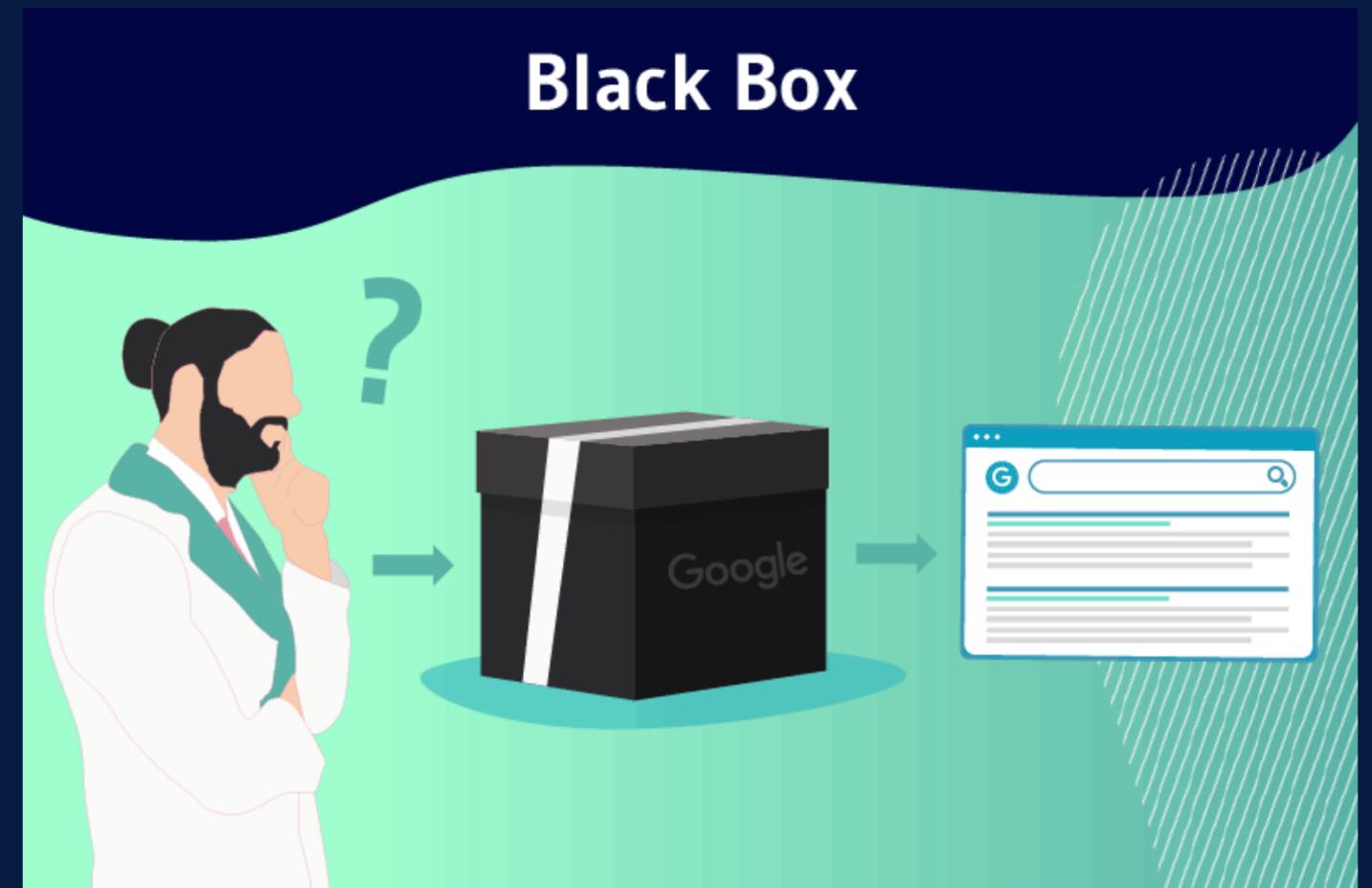
El trabajo pionero del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación contempla el desarrollo de factores habilitantes, el uso y desarrollo de esta tecnología, y aspectos de ética y seguridad. Junto a esta estrategia nacional, el ministro Andrés Couve presentó un plan de acción que reúne 70 acciones prioritarias y 185 iniciativas desde distintos servicios públicos centradas en aspectos sociales, económicos, y en la formación de talentos con un horizonte de 10 años.

Compartir esta página: [Twitter](#) [Facebook](#) [WhatsApp](#)



[Link](#)

*¿Qué opinan de priorizar la precisión de una predicción, a costa de la interpretación de cómo se llegó al resultado?*



Fuente

---

## TEMAS DE DISCUSIÓN:

- 1. ¿Qué podría explicar el aumento en el uso de modelos algorítmicos desde la época publicación del paper?*
- 2. ¿Qué ventajas y/o desventajas tendría el uso de modelos algorítmicos para temas de políticas públicas?*
- 3. ¿Qué opinan de priorizar la precisión de una predicción, a costa de la interpretación de cómo se llegó al resultado?*