

“Big data” y economía en la era de las “fake news” y la verdades imprecisas

José García Montalvo
Catedrático de Economía de la UPF

Fundación Ramón Areces
10 de Diciembre de 2019



Universitat
Pompeu Fabra
Barcelona



Esquema

1. Redes sociales y el apogeo de las “fake news”
2. “Data science” y soluciones de mercado
3. “Data science” y retos sociales
4. “Data science” y economía: como los datos están cambiando la enseñanza de la economía y la investigación económica
5. Ventajas y limitaciones del “data science”
6. Conclusiones



Introducción

- “Without data you are just one more person with an opinion” – Edwards Deming (Data Scientist)
- “Nos ahogamos en información y, a la vez, estamos hambrientos de conocimiento” - John Naisbitt
- “It is easy to lie with statistics. It is hard to tell the truth without it – Andrejs Dunkels (Mathematician)

“Fake news”

- Análisis de 126.000 historias de Twitter entre 2006 y 2017 (4,5 millones de tuits)
- La falsedad se difunde más lejos, más rápido, más profundamente y más ampliamente que la verdad y los efectos son mayores para las noticias falsas de políticos que para noticias falsas de terrorismo, desastres naturales, ciencia, leyendas urbanas o información financiera
- Las personas que difunden noticias falsas tienen menos seguidores pero sin embargo las noticias falsas son más virales -> la transmisión se produce por “bots” y humanos en la misma proporción
- El valor novedoso de las noticias falsas alienta su difusión
- *Aral, Vosoughi y Roy (2018), "The spread of true and false news online," Science ; D. Lazer et al. (2018), "The science of fake news," Science*

Verdades “imprecisas”

La desigualdad en España entre 2008 y 2017:

- A. Se ha duplicado
- B. Prácticamente no ha aumentado
- C. Ha aumentado un 50%

Solo un tercio de los niños de cero a tres años están escolarizados

ANA TORRES, **Madrid**
Solo un 37,9% de los niños de cero a tres años están escolarizados en España, a pesar de que los expertos recomiendan la educación temprana. El dato resulta de un reparto muy desigual por territorios: mientras en el País Vasco van a la escuela un 52,4% de niños de esa edad, en Canarias apenas son el 16,8%. Las cifras contrastan con una escolarización del 97% entre los menores de tres a seis años, etapa que es obligatoria y gratuita desde 2002. En un intento de promover la educación de los menores de tres años, el Gobierno y Unidos Podemos han pactado destinar 330 millones de euros en los Presupuestos de 2019 a convertir también esta primera etapa de la enseñanza en universal y gratuita.

zación del 97% entre los menores de tres a seis años, etapa que es obligatoria y gratuita desde 2002. En un intento de promover la educación de los menores de tres años, el Gobierno y Unidos Podemos han pactado destinar 330 millones de euros en los Presupuestos de 2019 a convertir también esta primera etapa de la enseñanza en universal y gratuita.

PÁGINA 25

La escolarización en España de 0 a 3 años:

- A. Está un 25% por debajo de la media de la UE y la OCDE
- B. Está un 50% por debajo de la media de la UE y la OCDE
- C. Está en la media de la OCDE y por encima de la media de la UE



Verdades “imprecisas”

La desigualdad en España entre 2008 y 2017:

- A. Se ha duplicado
- B. Prácticamente no ha aumentado
- C. Ha aumentado un 50%

“Además, la sociedad española continúa siendo más desigual que antes de la crisis. ... Así lo refleja la encuesta en el ratio S80/S20. Este indicador compara los ingresos del 20% más rico de la población con los del 20% más pobre. Aunque mejoró seis décimas el año pasado (del 6,6 al 6,0) todavía es mayor que en 2008 (5,9), lo que quiere decir que la diferencia entre lo que ganan los más ricos y los más pobres es todavía más grande que hace una década. Lo mismo sucede con el índice de Gini, un indicador que ... aún está 0,3 puntos por encima de 2008.”

Verdades “imprecisas”

Solo un tercio de los niños de cero a tres años están escolarizados

ANA TORRES, **Madrid**
Solo un 37,9% de los niños de cero a tres años están escolarizados en España, a pesar de que los expertos recomiendan la educación temprana. El dato resulta de un reparto muy desigual por territorios: mientras en el País Vasco van a la escuela un 52,4% de niños de esa edad, en Canarias apenas son el 16,8%. Las cifras contrastan con una escolarización del 97% entre los menores de tres a seis años, etapa que es obligatoria y gratuita desde 2002. En un intento de promover la educación de los menores de tres años, el Gobierno y Unidos Podemos han pactado destinar 330 millones de euros en los Presupuestos de 2019 a convertir también esta primera etapa de la enseñanza en universal y gratuita.

zación del 97% entre los menores de tres a seis años, etapa que es obligatoria y gratuita desde 2002. En un intento de promover la educación de los menores de tres años, el Gobierno y Unidos Podemos han pactado destinar 330 millones de euros en los Presupuestos de 2019 a convertir también esta primera etapa de la enseñanza en universal y gratuita.

PÁGINA 25

La escolarización es España de 0 a 3 años:

- A. Está un 25% por debajo de la media de la UE y la OCDE
- B. Está un 50% por debajo de la media de la UE y la OCDE
- C. Está en la media de la OCDE y por encima de la media de la UE



Verdades “imprecisas”

Programa de Estabilidad 2019-2022 del gobierno español.

En el diagnóstico de los desequilibrios que persisten en la economía española señala “... y aumento de la desigualdad (salarios, renta, consumo y riqueza, distribución intergeneracional, territorial y de género)”.

“Fake news” y polarización política

- La polarización política y la segregación ideológica de los votantes por medio de los canales digitales (Twitter, Facebook, etc.) tienen un enorme potencial de socavar la democracia aunque no está claro si polariza más estar rodeado de gente que piensa como tú o lo contrario
- Las “fake news” tienen un caldo de cultivo natural en comunidades con una gran segregación ideológica y de influir en elecciones como se ha demostrado en varias contiendas electorales
- Pero el “big data” combinado con ML podría ayudar a librarnos de las “fake news” -> “fact checkers”



Introducción: soluciones de mercado





Pero, ¿qué es “big data”?

- Ingente cantidad de información: “la muestra es la población”
- Enorme heterogeneidad de formatos: sensores, GPS, clicks, logs de servidores, correos electrónicos, imágenes, voz, etc
- Bajo nivel de señal sobre ruido -> análisis asociado normalmente a técnicas de “machine learning” computacionalmente muy intensivas
- Reutilización y posibilidad de cruzar datos
- No pretende explicaciones causales sino meramente predictivas -> causalidad no es lo más relevante -> solo importa la correlación. Mayer y Cukier (2013): “creciente respecto por la correlación en lugar de la búsqueda de la elusiva causalidad”



Pero, ¿qué es “big data”?

- Algunos hechos básicos sobre big data
 - La mayor cantidad de datos en la actualidad no la genera Internet sino los sensores y las grandes infraestructuras científicas: el Large Hadron Collider del CERN produce 600 TB/sec con sus 15 millones de sensores y, después de filtrado, necesita almacenar 25 PB/año
 - El problema fundamental en la actualidad no es la capacidad de computación sino la creación de información a un ritmo más rápido que la capacidad de almacenarla y la energía necesaria para mover la información entre el procesador y el dispositivo de almacenaje de la información



Big data y mercado

- Target
- Amazon empleó hasta 2001 a docenas de críticos y editores para sugerir títulos a sus clientes-> el algoritmo de recomendación “item-by-item” los sustituyó
- Netflix: mismo algoritmo de recomendación
- Google: subastas de anuncios
- Seguros: sensores
- Control de fraude en tarjetas y extracción de efectivo en cajeros
- Servicios financieros: scoring y big data
 - Contactos en LinkedIn: probabilidad de perder empleo y encontrar otro
 - Amigos en Facebook en listas de morosos
 - Número de amigos en Facebook y sus características: educación, empleo, etc
 - ¿Capitalizar?

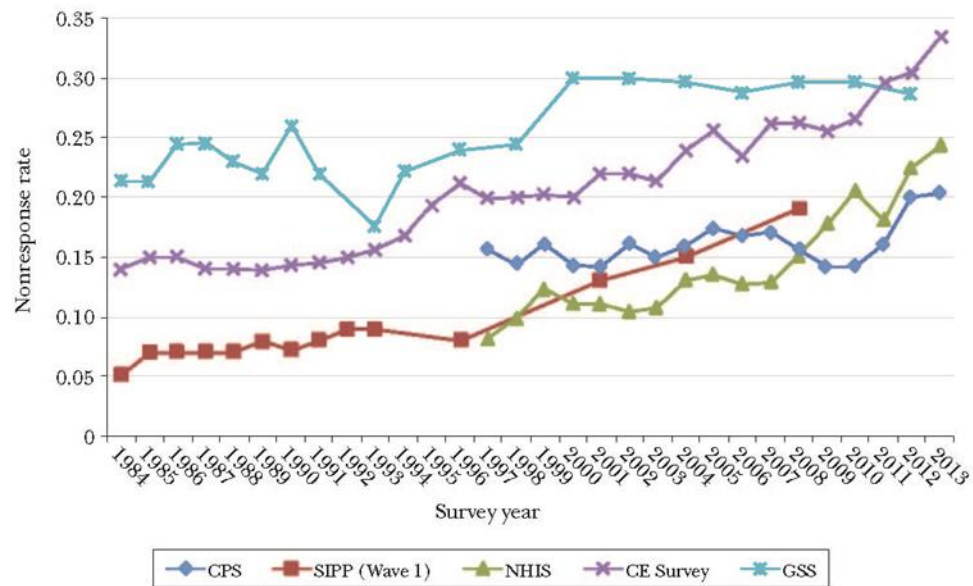


Big data: máquinas contra humanos

- Ajedrez: Deep Blue vs Kasparov (1997)
- Go: AlphaGo vs Fan Hui (2 dan) (2015): 5-0
- Go: AlphaGo vs Lee Sedol (9 dan, héroe nacional de Corea) Marzo 2016 (Google DeepMind Challenge Match): 4-1
- AlphaGo muy diferente: Inteligencia Artificial basada en redes neuronales que aprende jugando millones de partidas consigo mismo para conseguir experiencia (“reinforcement learning”). En cierta forma es original pues no tiene integrados los vicios y servidumbres del pasado de los humanos y los programas anteriores
- Humanos: ¿podemos ser sustituidos o somos complementarios?

“Big data” y avance social

- Problema para avanzar en los medios tradicionales: cada vez las encuestas presentan mayor nivel de no respuesta y son menos informativas (los encuestados parece que mienten cada vez más..)



Source: Meyer, Mok and Sullivan (2015)



Big data y avance social

- El efecto Hawthorne en experimentos presenciales/encuestas. Stephens Davidowitz “Todo el mundo miente”... menos en las búsquedas en Internet (“suero de la verdad”):
 - “es mi marido”...



Big data y avance social

- El efecto Hawthorne en experimentos presenciales/encuestas. Stephens Davidowitz “Todo el mundo miente”... menos en las búsquedas en Internet (“suero de la verdad”):
 - “es mi marido”... buscador rellena “homosexual” no “infidel”
 - Correlación con desempleo en Estados Unidos: ¿se busca “oficina de empleo” o “vacante”?



Big data y avance social

- El efecto Hawthorne en experimentos presenciales/encuestas. Stephens Davidowitz “Todo el mundo miente”... menos en las búsquedas en Internet (“suero de la verdad”):
 - “es mi marido”... buscador rellena “homosexual” no “infidel”
 - Correlación con desempleo en Estados Unidos: ¿se busca “oficina de empleo” o “vacante”? No. Se busca “porno” y el juego del solitario

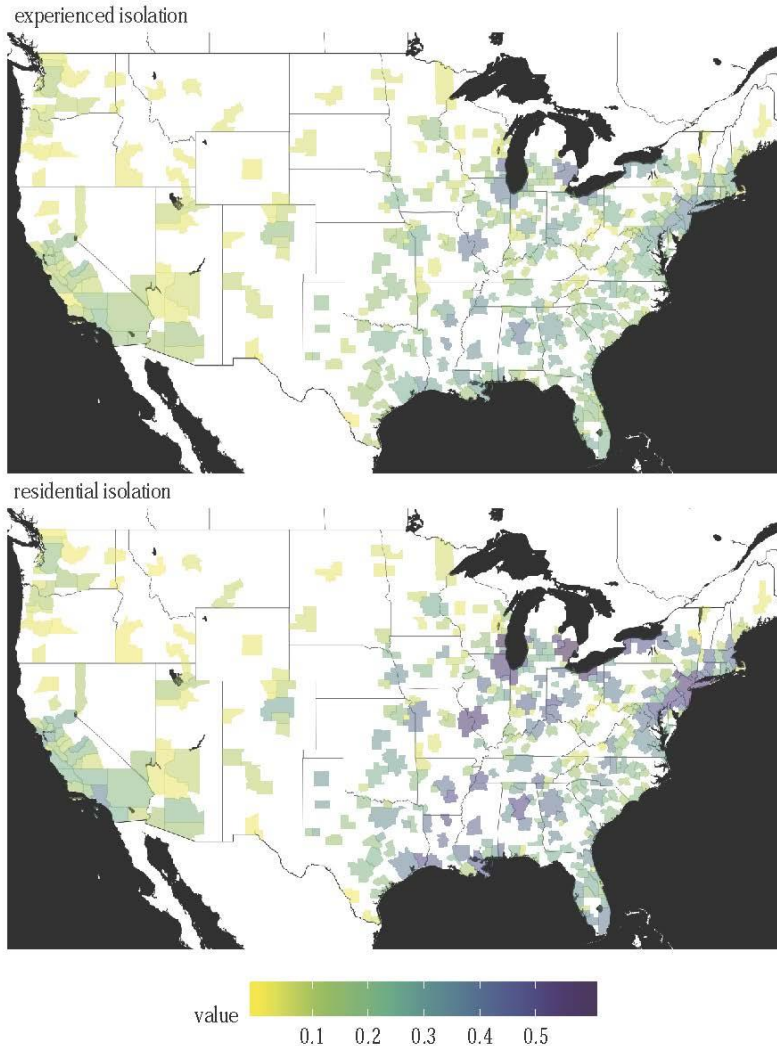
Big data y avance social

- El efecto Hawthorne en experimentos presenciales/encuestas. Stephens Davidowitz “Todo el mundo miente”... menos en las búsquedas en Internet (“suero de la verdad”):
 - “es mi marido”... buscador rellena “homosexual” no “infidel”
 - Correlación con desempleo en Estados Unidos: ¿se busca “oficina de empleo” o “vacante”? No. Se busca “porno” y el juego del solitario
 - Se busca más “tamaño del pene” o “alargamiento de pene” que como cambiar una rueda. Por su parte las mujeres muestran preocupación por “olor en las partes íntimas”
 - Se busca más “mi hijo es un superdotado” que “mi hija es una superdotada” y, sin embargo, se busca más “mi hija es obesa/guapa/fea” que “mi hijo” cuando los niños obesos son una proporción superior que las niñas
 - No busquen palabras que correlacionan con el precio de las acciones o del oro... ya lo intentó Larry Summers, antiguo secretario del Tesoro de Estados Unidos, y fracasó

“Big data” y avance social

- La disponibilidad creciente de enormes bases de datos, en muchos casos geocodificadas, que fusionan información de procedencia diversa hace de la economía una disciplina cada vez más científica
 - The Billion Prices Project: estimación en tiempo real de la evolución de los precios utilizando millones de precios de tiendas on-line. Con este proyecto se muestra que la evolución oficial de la inflación y la de precios on-line sigue patrones parecidos en Brasil, Chile, Venezuela o Colombia pero no en Argentina (diferencia acumulada entre 2007 y 2011 un 65%)
 - Central de Información de Riesgos del Banco de España: aprender sobre la relación entre pol. monetaria y riesgo de crédito: 24 millones de créditos
 - Seguimiento de los participantes en el proyecto STAR
 - Choi y Varian (2014): complementar la información de pasado de una serie con la búsqueda en algunas categorías -> ejemplo: modelo AR(1) de subsidio de desempleo completado con Google Trends para palabras como jobs, welfare o unemployment

"Big data" y avance social



Segregación residencial frente a segregación de experiencias

Uso de la localización de teléfonos móviles

Alta correlación entre segregación residencial y segregación de experiencias

El aislamiento en la experiencia no cambia el ranking de las ciudades en términos de mezcla racial

El mayor aislamiento racial se da en los domicilios, colegios e iglesias. La mayor exposición a otras razas se produce en universidades y restaurantes

Hay menos aislamiento racial a medio día

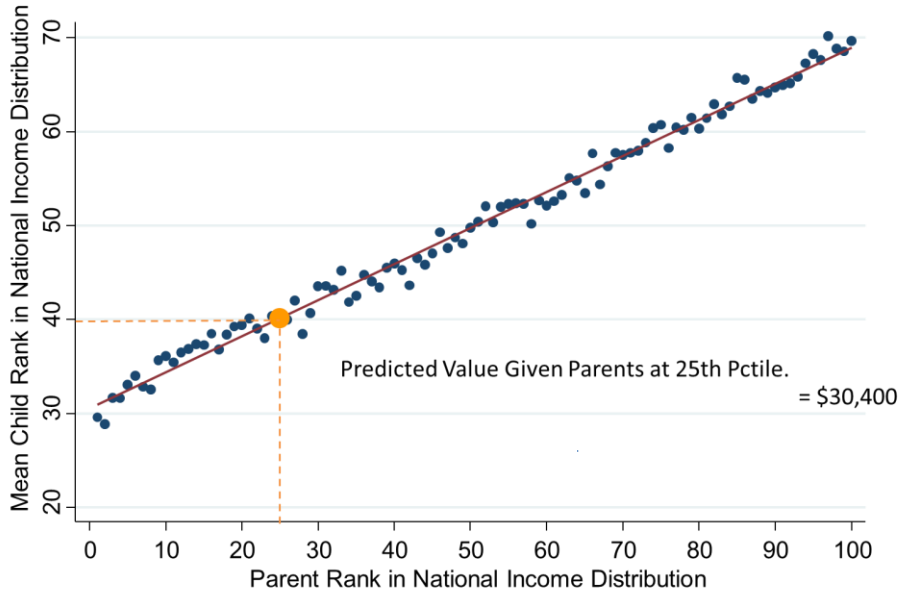


Big data y avance social: datos administrativos

- Los datos administrativos no están diseñados para la investigación (reutilización) y, por lo general, son más complejos que los datos de encuestas -> se requiere bastante tiempo para organizarlos y “limpiarlos”
- El tamaño de las bases de datos es ingente aunque normalmente los datos recogidos en redes sociales, transacciones comerciales o seguimiento de GPS suelen tener tamaños superiores
- Utilización de datos administrativos en economía: propuesta de Card, Chetty, Fledstein y Saez (2010)
- Infraestructura de uso de datos administrativos para la investigación muy avanzada en Noruega, Finlandia, Suecia y, especialmente, Dinamarca
- En Estados Unidos y Reino Unido más problemas para cruzar información por falta de número único de identificación

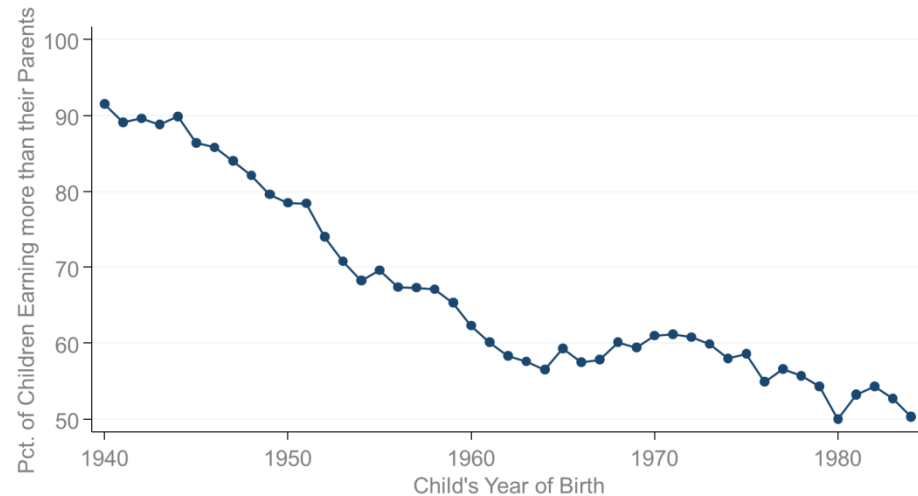


Big data y avance social: datos administrativos



Movilidad social e igualdad de Oportunidades: Raj Chetty y coautores

Fusionar datos de la agencia tributaria para muchos años y los censos

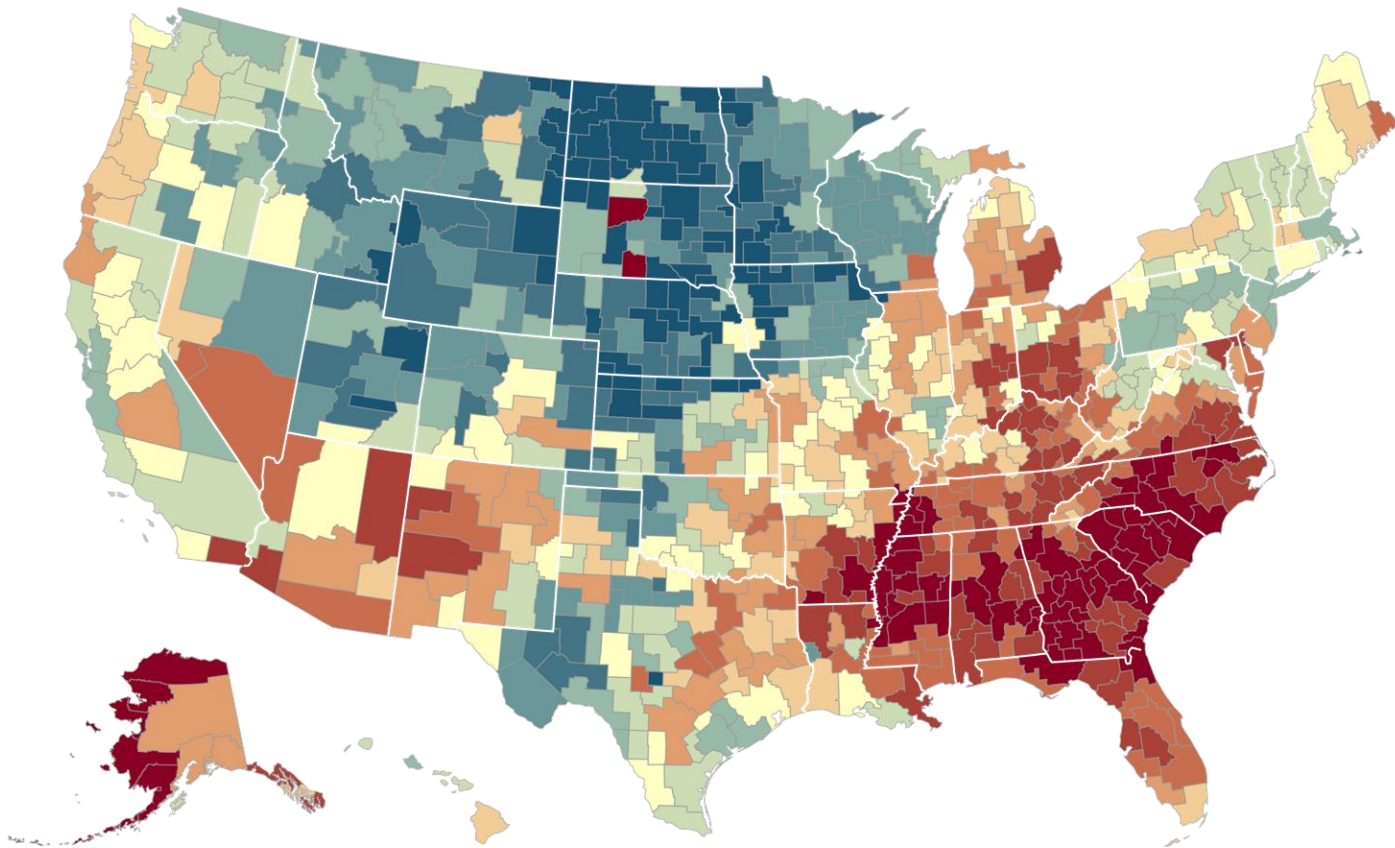




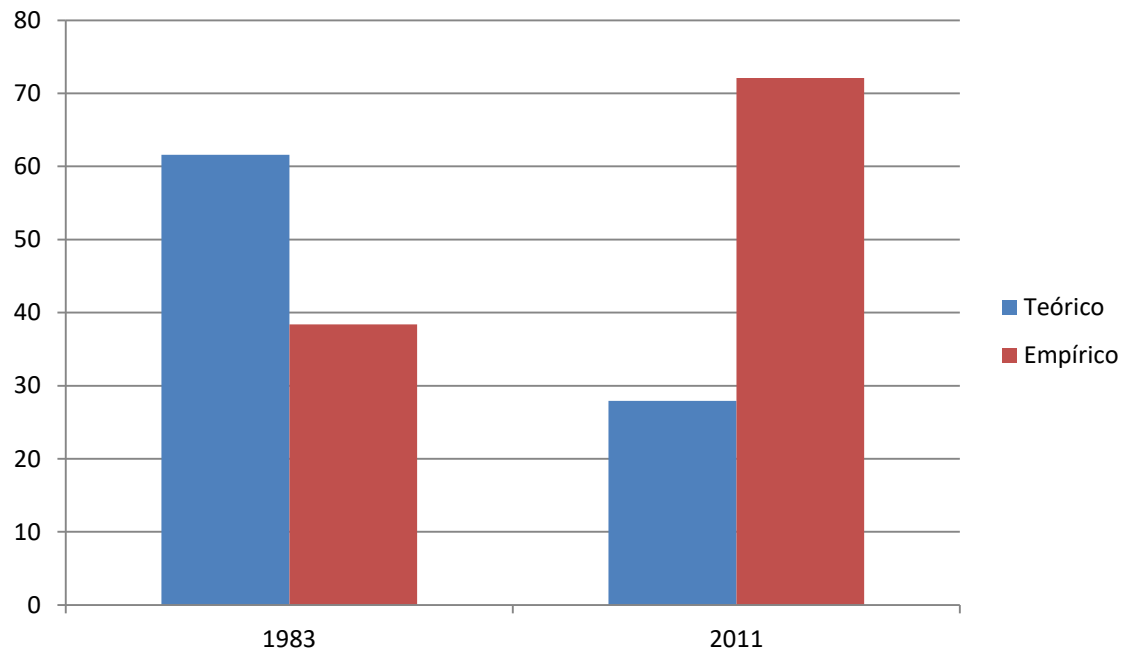
Big data y avance social: datos administrativos

The Geography of upward mobility in the US

Renta media de una familia con un hijo cuyos padres ganan 27.000 dólares



Datos y economía académica





Big data y la enseñanza de la economía

EC10 antes de marzo de 2019
(Greg Mankiw)...

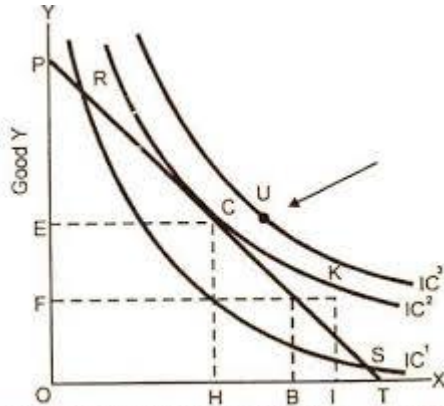
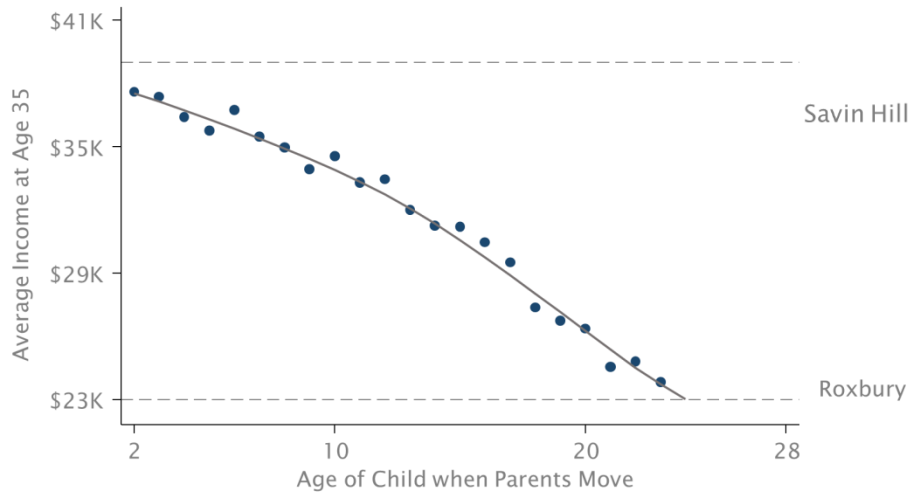


Fig. 3.11: Consumer's equilibrium through indifference Curves



Y ahora (Raj Chetty)

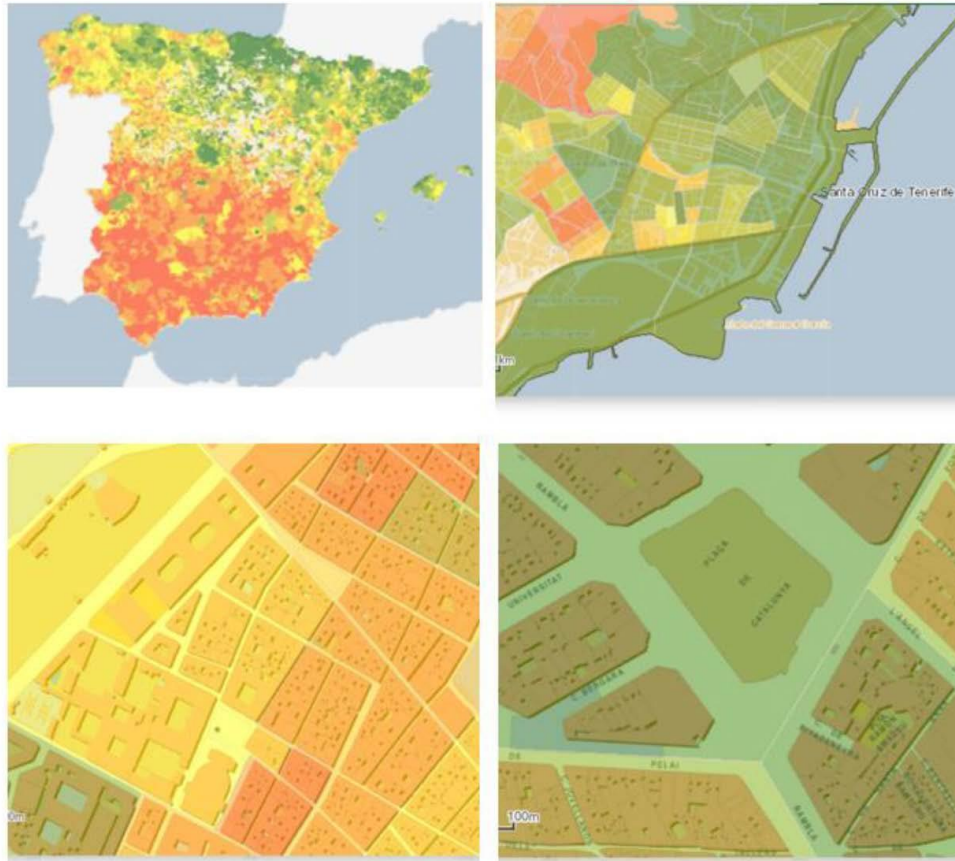


Datos administrativos: el caso español

- Retraso frente a otros países a pesar de tener unos datos administrativos muy ricos
- Ejemplos de bases de datos generales disponibles:
 - Muestra continua de vidas laborales (MCVL): 4% de inscritos en la Seguridad Social. Enlazados con IRPF (aunque no en todos los años)
 - Base de datos del Instituto de Estudios Fiscales: 14% de los declarantes del IRPF y de las personas en la EPA y la ECV
 - PET: panel de datos empresa-trabajadores de la SS. Misma información que MCVL pero para todas las personas que hayan trabajado en una muestra de empresas en algún momento durante un período de cuatro
 - DataResSS (BarcelongaGSE e IDESCAT)
- Mi experiencia
 - Subasta de viviendas públicas y mercado laboral
 - Muchos años para mover cualquier cosa
 - Depende de las personas y no hay estrategia de favorecer la investigación...
 - ...pero con las empresas privadas se es mucho más “permisivo”

Datos administrativos: el caso español

Atlas de distribución de la renta de los hogares
INE – Septiembre 2019

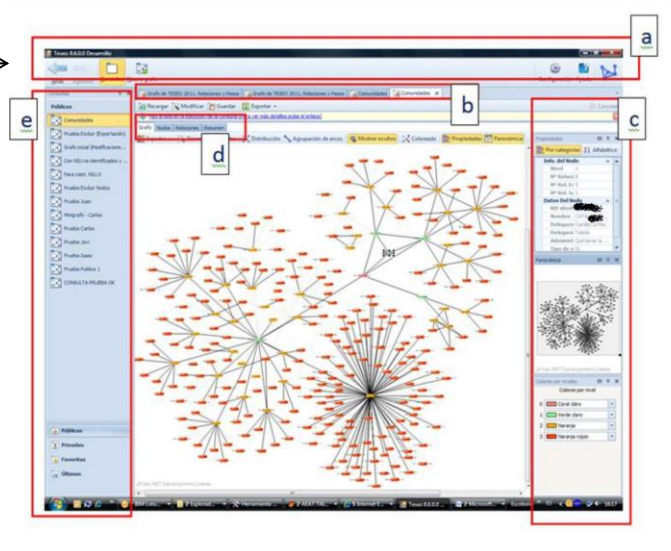
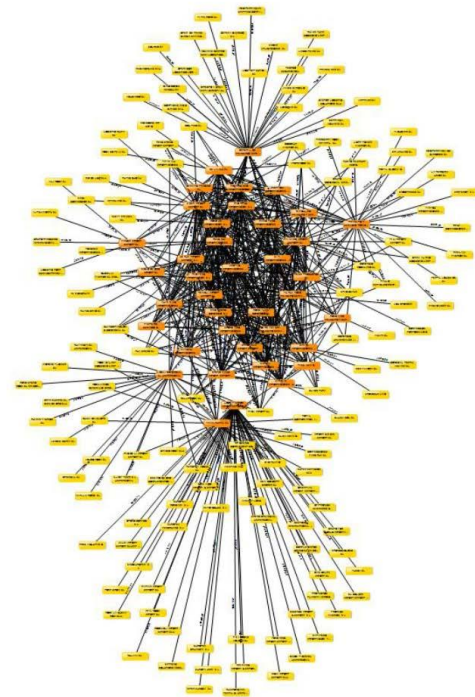


Datos administrativos: el caso español

La agencia tributaria

Sistemas de “big data” e inteligencia artificial:

- Zujar: unos 20 Billones de datos
- Prometeo
- Infonor/genio
- Midas – ROI
- Web scrapping
- Electra
- Teseo



Algunos grupos de contribuyentes tienen una distribución anómala de relaciones comerciales Internas versus externas

Datos administrativos: el caso español

- Sin embargo para la evaluación del gasto público no hay tanta tecnología...
- De hecho hay políticas para las cuales no sabemos ni el monto total del gasto destinado
- Lo único importante parece ser solo anunciar la política de turno y gastar lo presupuestado
- Parece que es mejor no aprender de lo que ha funcionado, o no, que exponerse a que una evaluación determine que la política no ha sido ineficaz, y no se pueda “vender” en la próxima cita electoral
- Al menos la AIREF lo intenta con sus “Spending reviews”

Las operadoras cobrarán medio millón de euros por ceder los datos al INE

A. MAQUEDA / R. MUÑOZ. Madrid. Las operadoras telefónicas cobrarán al Instituto Nacional de Estadística (INE) casi medio millón de euros por cederle los datos para que elabore un estudio sobre la movilidad de los españoles. Estas empresas explican que pueden vender sin problemas la información porque ha sido anonimizada. Es decir, no hay forma de conocer qué clientes están detrás de esas cifras y, por tanto, no se infringe de ningún modo la legislación de protección de datos.

Así va a rastrear el INE tu móvil durante ocho días: a qué operadoras afecta y cómo evitarlo

SARA RÍOS 02.11.2019 - 12:01H

- [El INE estudiará los desplazamientos de los teléfonos móviles desde este noviembre.](#)
- [¿Es legal que el INE obtenga la posición de tu teléfono móvil sin preguntarte primero?](#)



Protección de Datos pide más información

El proyecto del INE de usar datos de móviles ha desencadenado una tormenta. ¿Es legal? ¿Vulnera nuestra privacidad? La Agencia Española de Protección de Datos aseguró ayer que ha pedido más información al organismo, mientras que las asociaciones de consumidores piden prudencia. "El uso de datos con fines estadísticos está regulado en la ley y en el reglamento comunitario, así que en principio no es ilegal", explica Ileana Izvernieceanu, portavoz de la OCU. Rubén Sánchez, de Facua, entiende que la noticia "pueda escandalizar", pero recuerda que empresas como Google muestran publicidad en función de dónde estemos con un fin comercial y no público. El abogado experto en tecnología Borja Adsuara tampoco duda sobre la legalidad del proyecto, ya que se basa en datos agregados que son anónimos, pero lamenta la poca transparencia del organismo. "Esto perjudica a la economía del conocimiento y crea miedo al big data, que es la estadística del siglo XXI". / L. DELLE FEMME

El INE va a espiar tu móvil 8 días: ¿Qué puedes hacer?

Las compañías telefónicas se van a embolsar medio millón de euros por esta iniciativa

AGENCIAS(AUTOR/-)/META(AGENCIAS)/ 30 OCT 2019 / 17:22 H.



Big data y la demanda de graduados

ANTES

Tasa de inserción de los universitarios graduados en Matemáticas y Estadística mediocres y tasas altas de desempleo. Graduados de informática tenían mejores indicadores pero tampoco estelares

Notas de corte bajas

DESPUÉS

Tasas de inserción de las más altas de todas las titulaciones y muy baja tasa de desempleo tanto matemáticas como estadística, ciencia de datos e informática

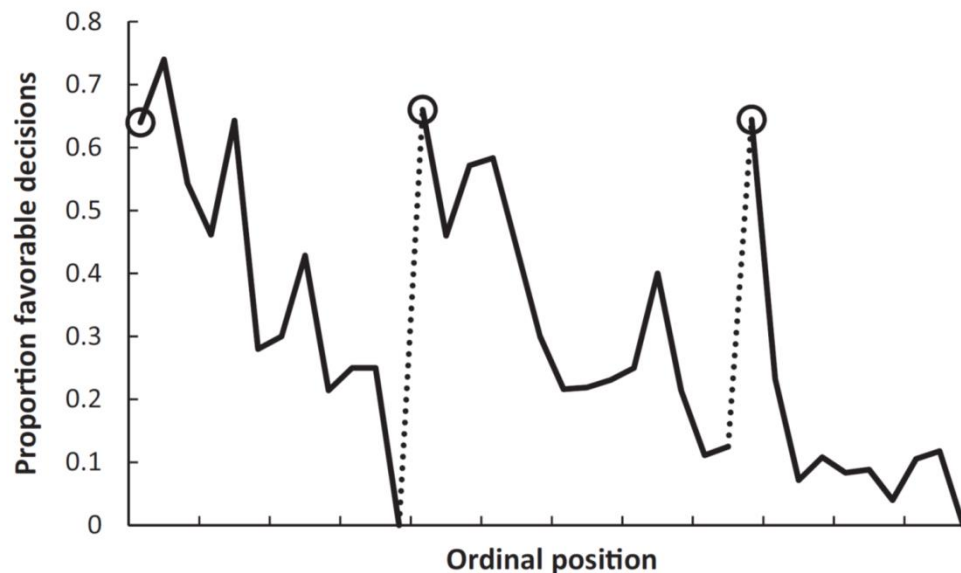
Notas de corte muy altas (especialmente en combinación con física)

Ventajas y limitaciones del "big data" y el ML

- Ventajas
 - Puede superar algunos sesgos presentes en los humanos
 - Permite analizar en tiempo real decisiones muy complejas
 - Permite realizar análisis a un nivel geográfico minúsculo

- Limitaciones
 - Peligros para la privacidad
 - Los algoritmos replican la discriminación existente en la realidad
 - Los resultados pueden no ser estables en el tiempo
 - Sin conocimiento se pueden llegar a decir muchas tonterías

- Sesgos en las decisiones humanas
 - Bertand and Mullainathan (2004): estudio sobre curriculum ficticios con nombres típicos de blancos (ej. Emily Walsh or Greg Backer) y afroamericanos (ej. Lakisha Washington or Jamal Jones)
 - Edelman, Luca y Sversky (2017) lo mismo pero para Airbnb
 - Danziger et al. (2011): fatiga de los jueces produce una reducción de la concesión de libertad bajo fianza



- Sesgos en las decisiones humanas

Behavioral Tendency	Description	Sources
Loss Aversion	Tendency to avoid options that result in a loss relative to one's current reference point, and to perceive losses as more impactful than gains of equal value	Kahneman & Tversky (1979); Tversky & Kahneman (1991); McNeil, Pauker, Sox & Tversky (1982); Tversky & Kahneman (1986); Wertebroch & Dhar (2000)
Endowment Effect	Tendency to attribute increased value to an owned item or entity	Thaler (1980); Kahneman, Knetsch & Thaler (1990);
Status Quo Bias	Tendency to select a default option when one is present	Samuelson & Zeckhauser (1988); Madrian & Shea (2001)
Affective Forecasting Error	Tendency to inaccurately predict future emotional states	Loewenstein & Schkade (1999); Simonson (1990); Gilbert et al. (1998); Loewenstein (1996)
Context-Dependent Preferences	Tendency to change one's preferences based on context, including how many options are being compared and the nature of their comparison (joint or separate)	Simonson & Tversky (1992); Tversky & Simonson (1993); Hsee & LeClerc (1998)
Affective-Cognitive Decision Making	Tendency to be more influenced by affective reactions than cognitive reactions when cognitive resources are limited	Shiv & Fedorikhin (1999)
Introspection and Consideration Override	Tendency to alter one's preferences when prompted to analyze them	Wilson & Schooler (1991); Amir & Ariely (2007)

Table 1. Summary of relevant behavioral tendencies



Las ventajas del “big data” y el ML

- Decisiones afectadas por factores que no influyen en los sistemas de inteligencia artificial y “machine learning”: decisiones de jueces sobre fianzas
- Kleinberg et al. (2017) comparan la fiabilidad de las decisiones de jueces frente a un sistema de “machine learning” en el sistema judicial de EE.UU. (10 millones de arrestos anuales)
- Tras el arresto decisión de fianza o no para minimizar el riesgo de fuga
- Datos: 750.000 detenciones en Nueva York entre 2008-13
- Técnica: árboles de decisión



Las ventajas del “big data” y el ML

- Los jueces tienen la ventaja de disponer de todos los datos del sistema de ML pero además ven el comportamiento del acusado...
- Lo que es un inconveniente dado que los indicios y pistas en la sala de juicios no son buenos predictores (como muestran los experimentos de la contratación de músicos “a ciegas” y de los CVs falsos)
- ML vence con una gran ventaja:
 - El crimen se puede reducir un 25% sin cambiar tasa de encarcelamiento
 - La población encarcelada puede reducirse un 42% sin cambio en las tasas de criminalidad



Las limitaciones del “big data” y el ML

- Si bien es cierto que “big data” proporciona herramientas muy útiles, no es menos cierto que la transformación de un proyecto de “big data” en un programa de éxito no está garantizada
 - Rendimientos decrecientes de la acumulación de información
 - Datos no proporcionan ventaja si no se analizan correctamente
 - Coste-Beneficio: ROI
- Grave peligro de injerencia en la privacidad de las personas con el objetivo de mejorar la capacidad predictiva de los modelos o su utilización para objetivos poco éticos (el escándalo de Cambridge Analytics) o discriminatorios



Las limitaciones del “big data” y el ML

- Privacidad de los datos y reutilización. Legislación de protección de datos: conferencias en la FRA organizadas por M. E. Casas
- La posibilidad de que los errores en la captura, fusión o limpieza de los datos generen consecuencias negativas para los ciudadanos a partir de la aplicación de técnicas de “big data” a problemas concretos. Un ejemplo es la industria de generación de “credit scores” a partir de “big data” captado en Internet-> NCLC (2014) (National Consumer Law Center)
- La posibilidad de sesgos en las evaluaciones crediticias:
 - Tarjeta de crédito de Apple: New York State Department of Financial Services investiga por qué el algoritmo carga un tipo de interés más alto a las mujeres -> “cualquier algoritmo que intencionadamente o no resulte en un tratamiento discriminatorio de la mujer viola las leyes de Nueva York”
 - Lo mismo sucede con los afroamericanos pero, en ese caso, es una ley federal

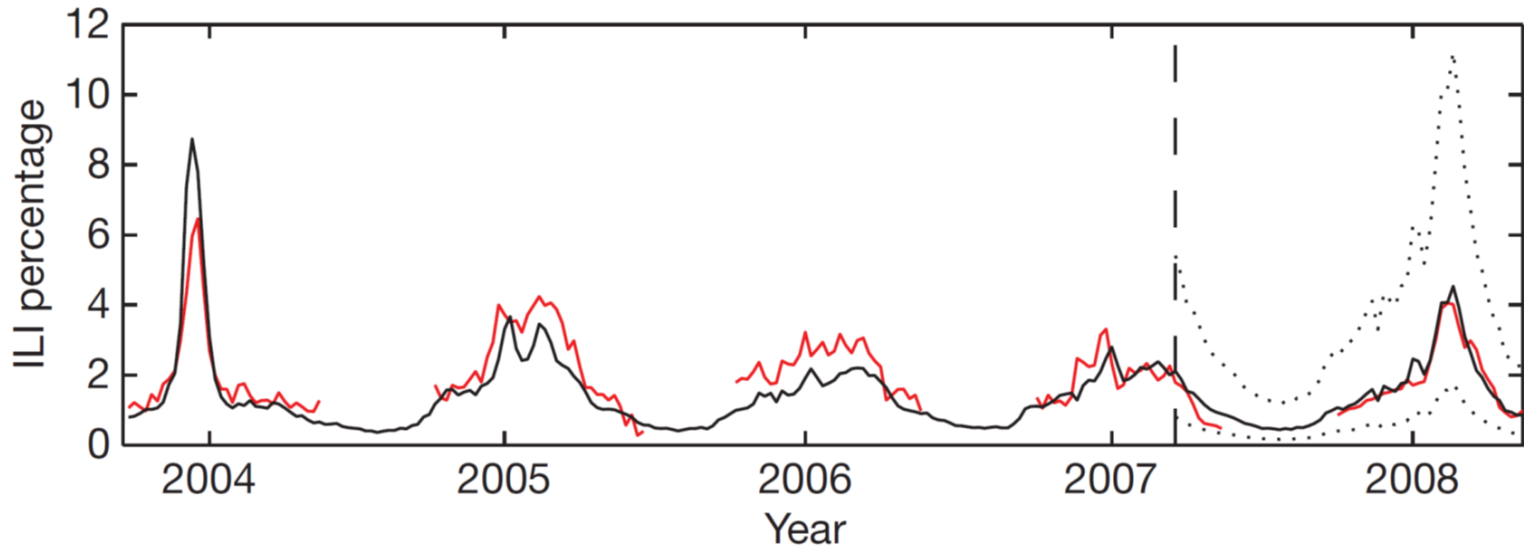


Las limitaciones del “big data” y el ML

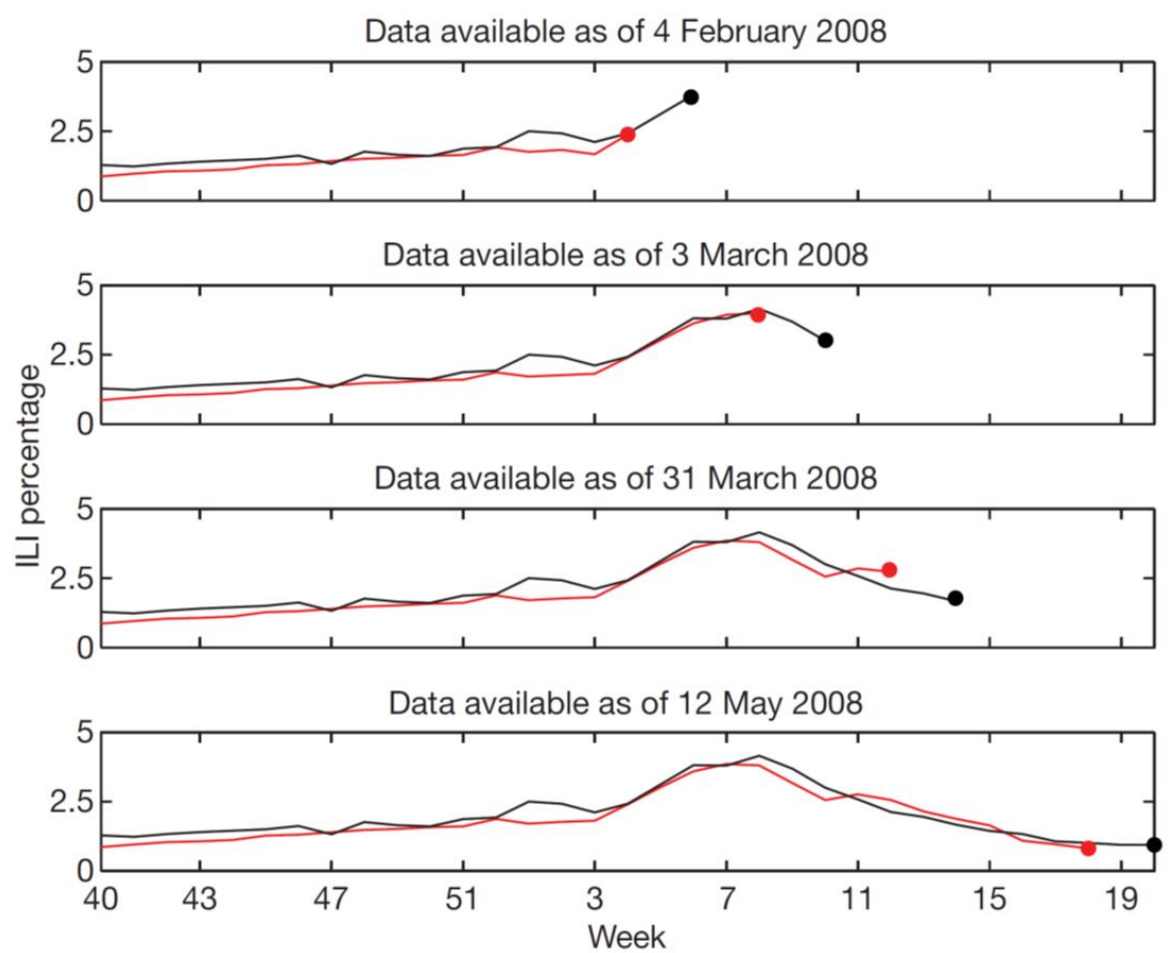
- La existencia de grandes cantidades de datos no puede hacer olvidar los fundamentos de la ciencia estadística, la influencia de los errores de medida o la precaución contra la utilización de correlaciones espurias
- Además del conocimiento técnico hace falta estar dispuestos a analizar constantemente la capacidad predictiva de los modelos y hacer ajustes a medida que el sistema pierde potencia explicativa. La experiencia de Google Trends (Google Flu Trends)
- Ginsberg et al. (2009) proponen una nueva fuente de datos para hacer seguimiento de la gripe: búsquedas en Google. Ventaja:
 - Seguimiento más rápido que usando los datos del CDC
 - Gran precisión geográfica de las predicciones

Las limitaciones del "big data" y el ML

La predicción de la gripe usando las búsquedas de Google

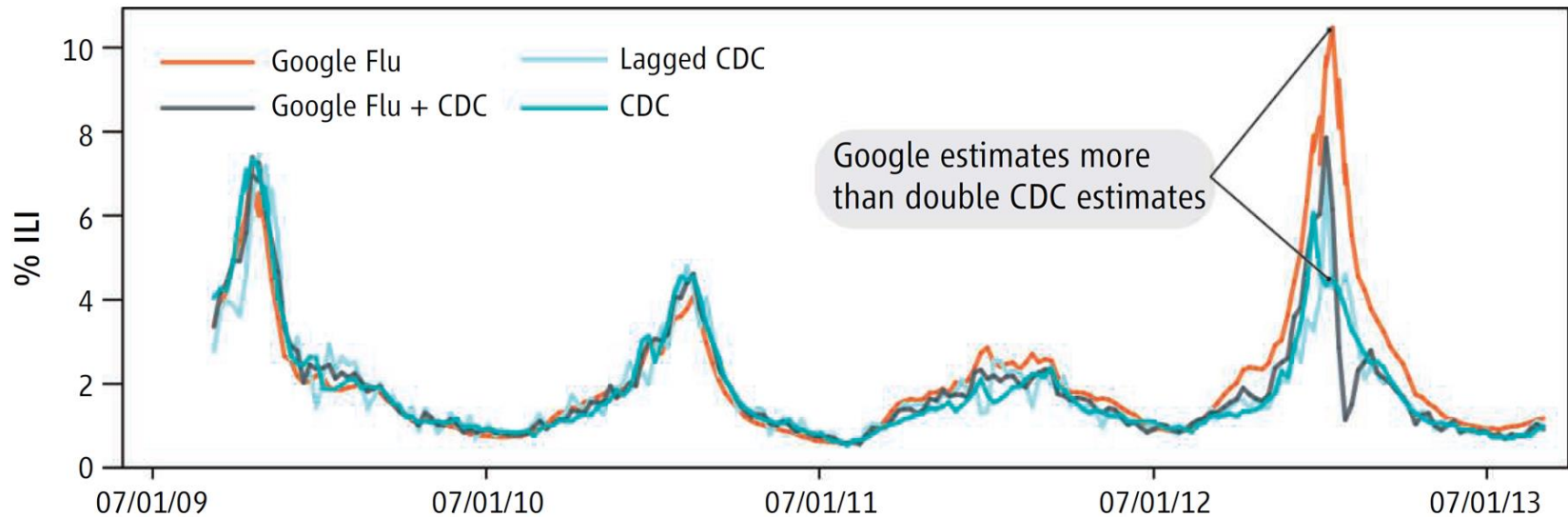


Las limitaciones del "big data" y el ML



Las limitaciones del "big data" y el ML

La capacidad predictiva del modelo empieza a fallar en 2012 y genera predicciones muy alejadas de las realizaciones. Lazer et al (2014) documentan en Science ese fallo



Las limitaciones del “big data” y el ML

- La existencia de grandes cantidades de datos no puede hacer olvidar los fundamentos de la ciencia estadística, la influencia de los errores de medida o la precaución contra la utilización de correlaciones espurias o metodologías no apropiadas
- Princeton versus Facebook
 - Cannarella y Spechler (investigadores de Princeton) concluyeron en 2014 que Facebook habría perdido el 80% de sus usuarios en 2017 usando un modelo epidémico sobre el número de veces que la palabra Facebook se busca en Google y usando MySpace como comprobación (“Epidemiological modelling of online social network dynamics”)
 - Usando los mismos principios Facebook contra ataca: “Utilizando el principio correlación implica causalidad y los mismos criterios de Princeton la universidad desaparecerá próximamente”



Conclusiones

- El ML y la enorme cantidad de datos disponibles abren una cantidad inmensa de posibilidades de mejorar los procesos productivos, empresariales y de conocimiento de los problemas sociales y evaluación de políticas públicas
- Ha avanzado mucho más la utilización en el mercado que el uso para resolver problemas sociales
- En España hay una utilización intensiva de estas técnicas en la recaudación impositiva pero están bastante ausentes en la evaluación del gasto público con excepción de los esfuerzos del AIREF
- Las nuevas oportunidades vienen asociadas a nuevos peligros
- El “big data” y las técnicas de ML no son una panacea



**Universitat
Pompeu Fabra**
Barcelona