



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Machine Learning for Social Science: An Agnostic Approach.

Curso: Aproximación a las políticas públicas desde los datos

Fecha: 9 de junio de 2023

Profesor: José Daniel Conejeros Pavéz

Integrantes: Josefa Anselmo – Sofía Fraile – Fabián Gil – Enzo Loiza – Gariel Sotomayor

Agenda

I. Machine Learning for Social Science: An agnostic approach

II. Desafíos para la generación de conocimiento científico

III. Desafíos para la práctica investigativa

IV. Discusión

I. Machine Learning for Social Science: An agnostic approach

- 1. Machine Learning para ciencias sociales**
- 2. Tesis del artículo**
- 3. Herramientas**
- 4. Objetivos en CS con aplicaciones de ML**
- 5. Conclusiones**

1. Machine Learning para Ciencias Sociales

Paper...

Es un Review a la inclusión de la cultura del Machine Learning en la ciencias sociales: no sólo lleva a repensar aplicaciones, sino también metodologías y prácticas.

Machine learning (ML) ...

Es un campo que reúne clases de algoritmos flexibles y técnicas estadísticas para predicción y reducción de dimensionalidad.

Efectos...

Los efectos del ML en los enfoques metodológicos y en las maneras de aproximarse a conceptualizaciones empíricas.

2.Tesis del artículo

ANTES, LA ESCASEZ DE DATOS DEFINÍA LAS CIENCIAS SOCIALES



METODOLOGÍA DEDUCTIVA



APROXIMACIÓN AGNÓSTICA A LAS HERRAMIENTAS DE ML:

LA METODOLOGÍA DUDA DE LOS MODELOS (EN SU CAPACIDAD DE REPRESENTAR LA REALIDAD) PERO CONFIA EN LAS VALIDACIONES (QUÉ HACEMOS COMO INVESTIGADORES)

AHORA, LAS CIENCIAS SOCIALES TIENEN ABUNDANCIA DE DATOS



METODOLOGÍA INDUCTIVA



3.Herramientas

✓ Evaluación de desempeño

Herramienta 1: Sample Splitting

Herramienta 2: V-fold Crossing Validation

✓ Ajustes de modelo

Herramienta 3: Regularización

Herramienta 4: Hiperparámetros

✓ Ingeniería de características automáticas

Herramienta 5: AFE

✓ Reducción de dimensionalidad

Herramienta 6: Reducción de dimensionalidad no supervisada

✓ El progreso como campo

Herramienta 7: Common task framework

4. Objetivos en Ciencias Sociales con aplicaciones de Machine Learning



Descubrimiento

- Clustering y Admixtures
- Factor Models and Embedding
- Fictitious Prediction Problems



Medición

- Democratization de la medición
- ML y estrategias de medición
- Dawid Skene y las ciencias sociales
- Regresión y aprendizaje supervisado
- Extrapolación MrP / Regresión Multinivel
- Evaluación de modelos de medición



Predicción e Inferencia Causal

Predicciones y ciencias sociales

Inferencia Causal

¿En qué se diferencian?

5. Conclusiones del paper

+

Ciencias sociales y un nuevo enfoque inductivo de la investigación con una mayor cantidad de datos.

+

El ML funcionará bien si se aplica adecuadamente a los problemas de investigación.

+

La teoría es importante para guiar el uso de ML en CS.

+

La abundancia no garantiza el progreso, pero si puede generar nuevos conocimientos.

II. Desafíos para la generación de conocimiento científico

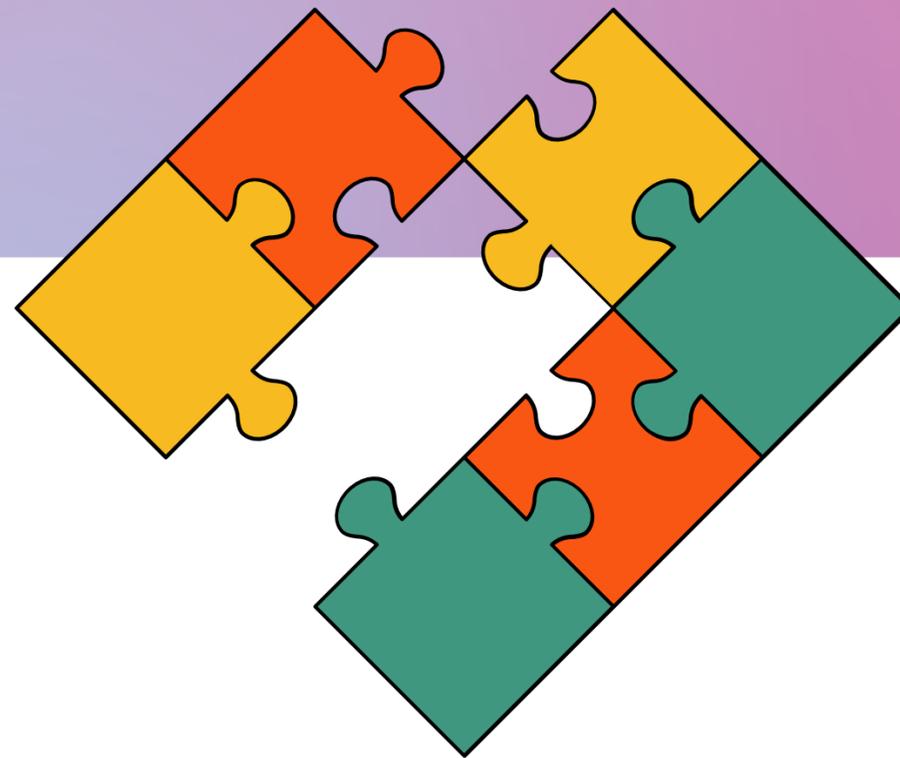
- 1. Enfoque Inductivo y agnóstico en Ciencias Sociales**
- 2. Desafíos en la producción de conocimiento**
- 3. Problemas de inducción**

1. Enfoque Inductivo y agnóstico en Ciencias Sociales



Los autores proponen un modelo inductivo, que parte del análisis exploratorio de los datos hacia la generalización del conocimiento y no asume la existencia de un modelo correcto para acercarse a los datos, ni asume que el modelo representa el data-generating process.

2.Desafíos en la producción de conocimiento



¿Cómo se genera o acumula conocimiento científico a partir de este enfoque? ¿Cómo se teoriza a partir de dicho conocimiento?

3. Problemas de inducción



La acumulación de evidencia inductiva no necesariamente establece un criterio para afirmar que algo es conocimiento científico.

III. Desafíos para la práctica investigativa

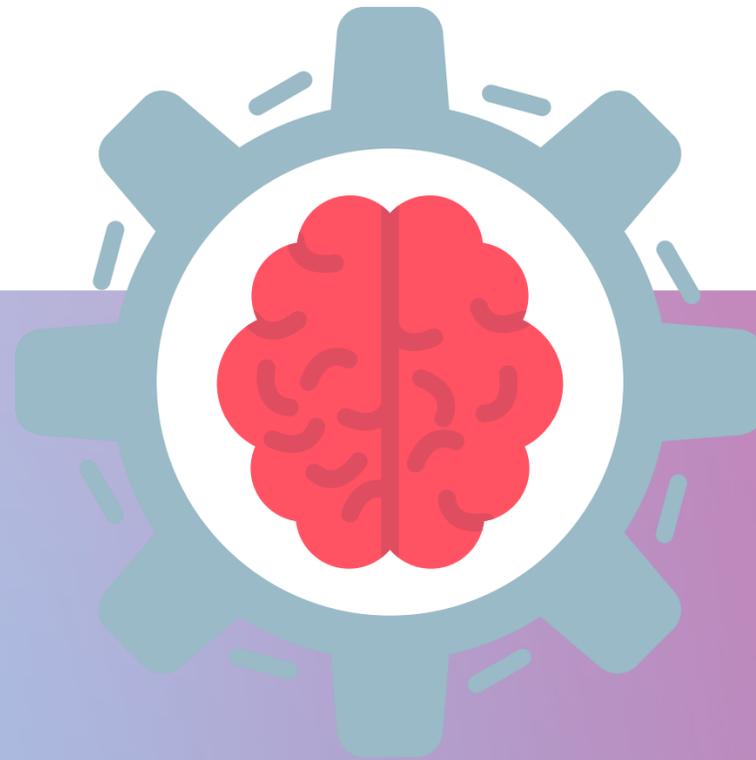
- 1. Interacción con la Práctica de Investigación Real**
- 2. Machine Learning y Problemas de Interpretabilidad**
- 3. Desafíos Éticos/Metodológicos**

1. Interacción con la práctica de investigación real



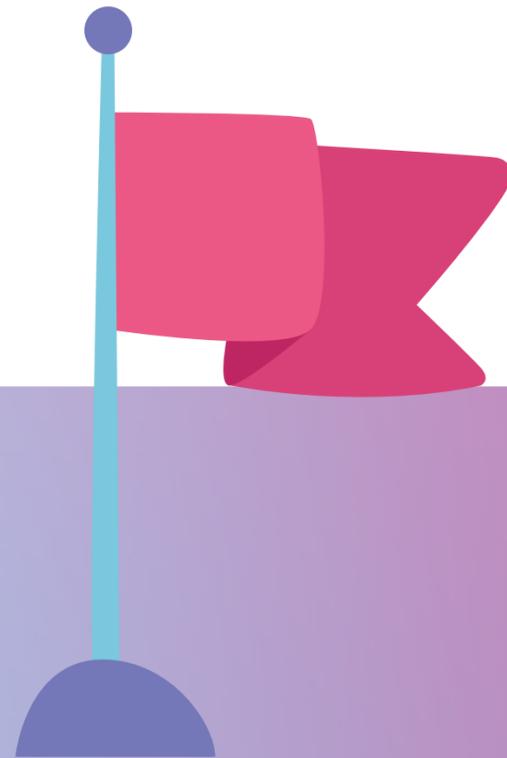
La acumulación de evidencia empírica inductiva no siempre se alinea con la investigación real que suele basarse en supuestos, intereses y conceptos, lo que se asemeja más a un modelo deductivo. Un modelo abductivo (mencionado por los autores) similar al de la lógica de investigación cualitativa de la teoría fundamentada, puede resolver parcialmente los desafíos anteriores.

2. Machine Learning y problemas de interpretabilidad



Los modelos complejos de machine learning pueden limitar la generación de conocimiento científico debido a su falta de interpretabilidad. ¿Cómo se pueden teorizar los resultados inductivos con modelos estadísticos que pueden volverse cada vez más "cajas negras"?

3.Desafíos éticos/metodológicos



La lógica de machine learning puede chocar con la ciencia reproducible. Mientras que las herramientas computacionales pueden mejorar la apertura de la ciencia, también pueden promover malas prácticas como el p-hacking y las hipótesis ad-hoc, particularmente en investigaciones con un diseño exploratorio.

IV. Discusión

¿Cuál creen que es el mayor beneficio que el ML puede aportar a las Ciencias Sociales?
¿Pueden identificar desventajas?

¿Que desafíos metodológicos implica el uso de ML en el ámbito de la investigación social y de políticas públicas (en contraste con su origen en ciencias de la computación)?

¿Qué desafíos éticos plantea el uso de machine learning en CS desde el punto de vista de la investigación y su relación con la sociedad?



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Machine Learning for Social Science: An Agnostic Approach.

Curso: Aproximación a las políticas públicas desde los datos

Fecha: 9 de junio de 2023

Profesor: José Daniel Conejeros Pavéz

Integrantes: Josefa Anselmo – Sofía Fraile – Fabián Gil – Enzo Loiza – Gariel Sotomayor