

Apertura científica a través del uso de datos abiertos



Héctor Guacapiña, Yasmina Vizuite-Salazar
Fundación Ecuatoriana para el Desarrollo de la Investigación Aplicada
y Transferencia Tecnológica – FEDIAT
(Colegio de Científicos de Datos de Ecuador),
Quito-Ecuador,
hec.tom1994@hotmail.com ,
vizuetey@gmail.com

ORGANIZADORES

data-lat.org



AUSPICIANTES



CONTENIDO:

- Objetivo
- Metodología
- Pre-procesamiento de datos
- Análisis de datos

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Objetivo

- Los investigadores dan mayor importancia de la ciencia abierta, ya que comprenden la necesidad de compartir datos, software e infraestructura. Cada paso en el proceso de investigación sería público con datos accesibles a la ciudadanía, que permitiría una mayor participación de actores en la resolución de una situación en específico.
- En el caso del sector público, se requiere iniciar con el conocer los datos que se tiene a disposición en portales abiertos, lo que permite a los investigadores (o ciudadanos científicos) procesar estos datos e intercambiar los resultados para contribuir a la formulación de política pública.
- El objetivo de este taller es la promoción de la apertura científica a través del impulso a los datos abiertos y que permita al sector público enriquecerse de estos datos que lleven a contribuir a desarrollar política pública.

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

Chan
Zuckerberg
Initiative

GORDON AND BETTY
MOORE
FOUNDATION

Metodología

- Preprocesamiento de datos
 - Usaremos el portal de datos abiertos, de donde nos descargaremos una base de datos y su Diccionario. Utilizaremos la base de datos de educación del examen 'ser bachiller', y analizaremos los datos juntos con su diccionario para identificar errores o inconsistencias. Finalmente, nos enfocaremos en valores perdidos y su imputación para obtener una base de datos depurada.
- Análisis de datos
 - Con la base de datos depurada, realizaremos estadística descriptiva y detección de patrones. Formularemos alternativas que permitan definir la aplicación de política pública con base en los datos y patrones identificados.

Preprocesamiento de datos



Héctor Guacapiña

Fundación Ecuatoriana para el Desarrollo de la Investigación

Aplicada y Transferencia Tecnológica – FEDIAT

(Colegio de Científicos de Datos de Ecuador),

Quito-Ecuador,

hec.tom1994@hotmail.com

ORGANIZADORES

data-lat.org



AUSPICIANTE



SUMARIO/CONTENIDO:

- Portal de datos abiertos del Ecuador
- Importancia del preprocesamiento de los datos
- Colección de datos
- Limpieza e imputación de datos

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Portal de Datos Abiertos Ecuador

Busca en nuestro catálogo de Datos Abiertos

Explora más de **5778 recursos**
más de **776 conjuntos** de datos abiertos
de **95 instituciones**.

Busca por palabras claves



Fuente: datosabiertos.gob.ec

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES



Importancia

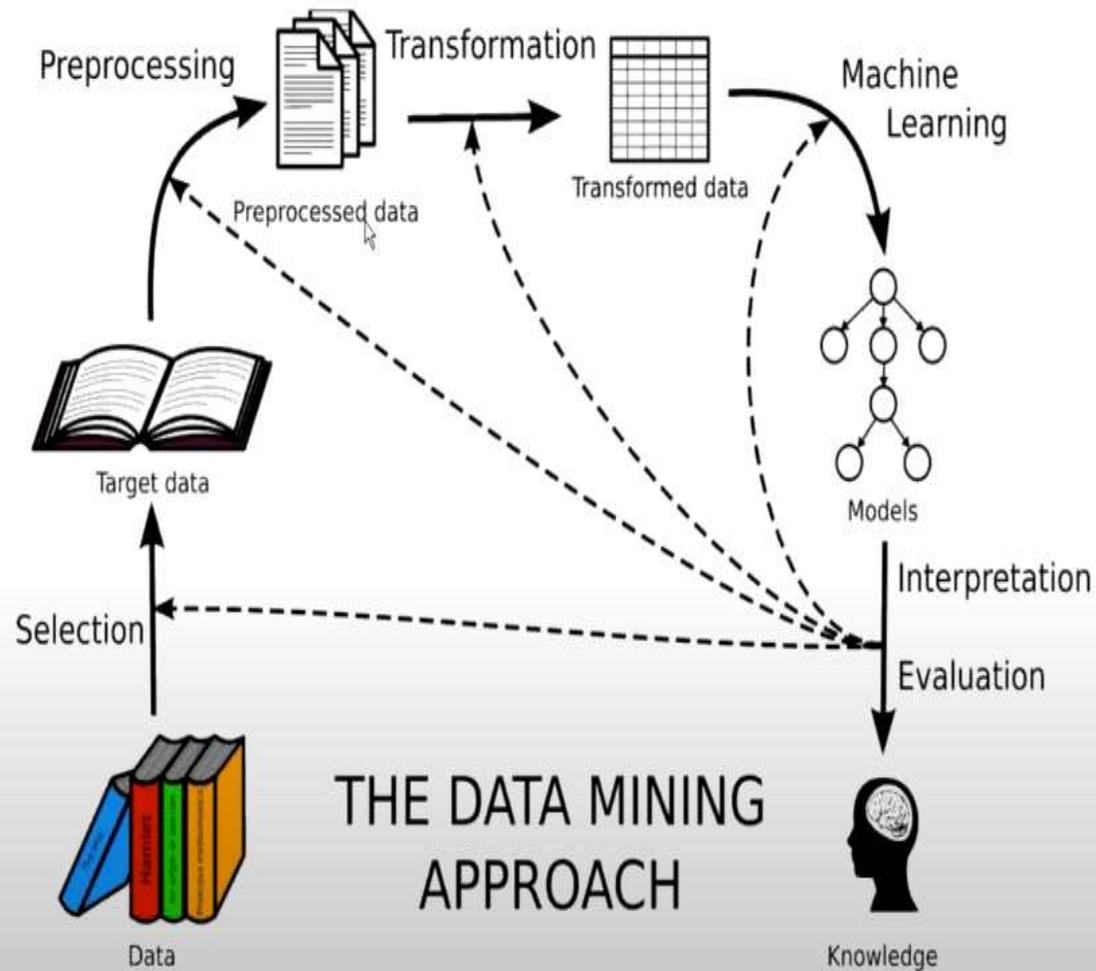
“El propósito fundamental de la preparación de datos es manipular y transformar los datos sin procesar para que el contenido de la información incluida en el conjunto de datos pueda exponerse o hacerse más fácilmente accesible”.

Dorian Pyle

Preparación de datos para minería de datos Editorial Morgan Kaufmann, 1999



Fuente: blog.powerdata.es



FLUJO ANALISIS DE DATOS

Fuente: Intelligent Systems Group.

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES



Colección e Integración de los Datos

Adquisición de los datos

Captura desde entornos de bases de datos: SQL

Mayoría de softwares, trabajan en formato plano de datos:

- Delimitado por: columnas, tab, comas, etc.
- Formato CSV: “comma separated values”

Meta-datos

Descripción de las variables

Tipos de variables:

Nominales – categóricas

Ordinales – numéricas

Roles en el fichero de datos:

input, predictor, variables para entrenamiento

Output, variable a predecir

Un identificador auxiliar

Limpieza de Datos

Relation: iris

No.	sepalength Numeric	sepalwidth Numeric	petallength Numeric	petalwidth Numeric	class Nominal
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	6.3	2.5	4.9	1.5	Iris-versicolor
5	6.1	?	4.7	1.2	Iris-versicolor
6	6.4	2.9	4.3	1.3	Iris-versicolor
7	6.5	3.0	5.2	2.0	Iris-virginica
8	6.2	3.4	5.4	2.3	Iris-virginica
9	5.9	3.0	5.1	1.8	Iris-virginica

Fuente: Intelligent Systems Group.

Valores Perdidos

Los datos pueden aparecer en distintas formas:
<empty field>, ., 999. NA, ?, ...

Necesario estandarizar

¿Qué podemos hacer con los valores perdidos?

Ignorar casos con valores perdidos

Tratar el valor perdido como un valor aparte

Imputación: rellenar el valor con la media (variables numéricas) o moda (nominales)

Posibilidad de hacer esta imputación condicionando al valor de la clase del caso concreto

Técnicas mas avanzadas: K-NN neighbours, EM algorithm

Análisis de Datos



Yasmina Vizquete-Salazar

Fundación Ecuatoriana para el Desarrollo de la Investigación

Aplicada y Transferencia Tecnológica – FEDIAT

(Colegio de Científicos de Datos de Ecuador),

Quito-Ecuador,

vizquetey@gmail.com

ORGANIZADORES

data-lat.org



REDAM
Red Ecuatoriana De
Datos Abiertos Y Metadatos

AUSPICIANTE

CS&S Code for
Science &
Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

CONTENIDO:

- Políticas Públicas
- Estadística descriptiva
- Detección de patrones
- Selección de alternativas

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Políticas Públicas

- Son conjuntos de objetivos, decisiones y acciones que lleva a cabo un gobierno para solucionar los problemas que en un momento dado los ciudadanos y el propio gobierno consideran prioritarios [1].
- Por ejemplo, el desempleo, la inseguridad ciudadana, la escasez de vivienda, la inmigración, el medioambiente, etc.
- A través de éstas se materializa la intervención del estado en la sociedad [2].

Políticas Públicas

- Toda intervención pública tiene el objetivo en alterar el estado natural de una situación social para lograr determinados impactos en la sociedad.
- Los impactos que se pretenden generar con cada política pública suelen estar contenidos en los objetivos de la misma y constituyen la razón de su existencia. Los objetivos son los resultados que se persiguen con las acciones concretas en el contexto de una política [3].
- De ahí que uno de los criterios más importantes para valorar las políticas públicas es la eficacia: si la actuación pública no ha ocasionado los impactos pretendidos o ha generado otros no deseados, entonces es considerada ineficaz [4].

Ciclo de las Políticas Públicas

Las principales fases del ciclo de las políticas públicas son las siguientes [4]:

1. la identificación y definición de problemas;
2. la formulación de políticas;
3. la adopción de la decisión;
4. la implantación;
5. la evaluación.

Ciclo de las Políticas Públicas

La identificación y definición de problemas

El gobierno advierte la existencia de problemas u oportunidades en la sociedad y se plantea si debe actuar o no.

Esta fase incluye actividades de detección y selección de cuestiones, y, actividades de definición de problemas [4].

Ciclo de las Políticas Públicas

La identificación y definición de problemas

- La definición de los problemas públicos significa que el problema es percibido por los distintos actores interesados.
- La elaboración de una situación no deseada como problema público depende en gran medida de su conexión con los valores dominantes en ese momento en la sociedad de que se trate.
- La definición de los problemas públicos no tienen una formulación definitiva [4]

Ciclo de las Políticas Públicas

La formulación de las políticas

1. Establecimiento de metas y objetivos a alcanzar
2. Detección y generación de alternativas que permitan alcanzar los objetivos
3. Valoración y comparación de las alternativas
4. Selección de una opción o combinación de ellas [4]

Ciclo de las Políticas Públicas

La adopción de la decisión

Esta fase está exclusivamente en manos de uno o varios decisores públicos: para que una política sea considerada pública ésta debe haber sido generada por medios gubernamentales, debe emanar de una autoridad pública [4].

Ciclo de las Políticas Públicas

La implantación de las políticas públicas

Una vez adoptada la decisión, comprende todas las actividades y procesos desarrollados hasta que aparecen los efectos asociados con la intervención pública en cuestión [4].

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Ciclo de las Políticas Públicas

La evaluación de las políticas

- Determina en qué medida se han logrado los objetivos de la política pública en cuestión.
- La evaluación cierra el ciclo de las políticas, y puede retroalimentar el proceso en cualquiera de sus fases.
- En la práctica, la evaluación no está tan extendida como sería deseable [4]

Estadística descriptiva

Es una disciplina que se encarga de recoger, almacenar, ordenar, realizar tablas o gráficos y calcular parámetros básicos sobre el conjunto de datos [5].

Estadística descriptiva

Distribución de frecuencias

- Utilizada tanto para datos cualitativos y cuantitativos.
- Se presenta normalmente en una tabla o un gráfico.
- Es un resumen de datos agrupados en clases mutuamente excluyentes y del número de ocurrencias en cada clase respectiva.
- Se utiliza gráficos de barras, histogramas, gráficos circulares y gráficos de líneas [6].

Estadística descriptiva

Tendencia central

- La tendencia central se refiere al resumen descriptivo de un conjunto de datos utilizando un único valor que refleja el centro de la distribución de los datos.
- La media, la mediana y la moda son medidas de tendencia central [6].

Referencias

- [1] Vargas Arévalo, Cidar. Análisis de las Políticas Públicas Perspectivas, núm. 19, enero-junio, 2007, pp. 127-136 Universidad Católica Boliviana San Pablo Cochabamba, Bolivia <https://www.redalyc.org/articulo.oa?id=425942453011>
- [2] Izaguirre, G.A., Roncancio, A.F. y Restrepo Marulanda, M. (2021). La intervención estatal y el Estado social de derecho: análisis de como las ramas del poder público materializan los derechos fundamentales. Nuevo Derecho; 17(28): 1-14. <https://doi.org/>
- [3] Olavarría Gambi, Mauricio. Conceptos Básicos en el Análisis de Políticas Públicas, DOCUMENTOS DE TRABAJO N° 11 – Diciembre – 2007. INAP Instituto de Asuntos Públicos, Departamento de Gobierno y Gestión Pública, Universidad de Chile. https://www.flacsoandes.edu.ec/sites/default/files/agora/files/1276704245.mauricio_olavarria.pdf
- [4] Delgado Godoy, Leticia. Documentación sobre gerencia pública, del Subgrupo A2, Cuerpo Técnico, especialidad de Gestión Administrativa, de la Administración de la Junta de Comunidades de Castilla-La Mancha. Tema 3. Las políticas públicas. El ciclo de las políticas públicas. Clases de políticas públicas. Eficacia, legalidad y control. Indicadores de gestión. ISBN: 978-84-7788-549-8. <http://pagina.jccm.es/ear/download/A2T3.pdf>
- [5] López, José Francisco. 15 de noviembre, 2019. Estadística descriptiva. Economipedia.com. <https://economipedia.com/definiciones/estadistica-descriptiva.html#:~:text=La%20estad%C3%ADstica%20descriptiva%20es%20una,grandes%20ramas%20de%20la%20estad%C3%ADstica.>
- [6] QuestionPro. Estadística descriptiva. Qué es y su importancia. <https://www.questionpro.com/blog/es/estadistica-descriptiva/#:~:text=La%20distribuci%C3%B3n%20de%20frecuencias%20es,presentar%20los%20datos%20en%20bruto.>

CO-ORGANIZADORES



ALIADOS DE DIFUSIÓN



ORGANIZADORES



AUSPICIANTES



Octubre, 2022