

## Guía metodológica para el uso de minería de datos en la Plataforma Nacional de Transparencia

*Methodological Guide for the Use of Data Mining  
in the National Transparency Platform*

Eva Grissel Castro Coria

 <https://orcid.org/0000-0001-6106-5631>

Instituto de Investigaciones Económicas y Empresariales, Universidad Michoacana de San Nicolás de Hidalgo  
Correo electrónico: [eva072001@hotmail.com](mailto:eva072001@hotmail.com)

Pamela Ruiz Flores López

 <https://orcid.org/0000-0001-5905-5370>

Instituto de Investigaciones Económicas y Empresariales, Universidad Michoacana de San Nicolás de Hidalgo  
Correo electrónico: [pamland24@gmail.com](mailto:pamland24@gmail.com)

Recepción: 17 de febrero de 2023  
Aceptación: 9 de marzo de 2023

**Resumen:** El objetivo de esta investigación es generar una guía metodológica para usar la minería de datos en la Plataforma Nacional de Transparencia (PNT), específicamente en el Sistema de Portales de Obligaciones de Transparencia. Lo anterior resulta importante en virtud de que obtener datos confiables y de naturaleza pública es esencial para comprender la realidad social. Para ello se utilizó el modelo identificado por el acrónimo SEMMA adecuado a la PNT, generando cinco pasos para la migración y análisis de datos. Por último, se desagregan las limitaciones relacionadas con que la guía está diseñada a un sistema de los cuatro sistemas que compone la PNT, y las futuras líneas de investigación.

**Palabras clave:** datos abiertos, transparencia, minería de datos, derecho a la información.

**Abstract:** The objective of this research is to generate a methodological guide to use data mining in the National Transparency Platform, specifically in the Transparency Obligations Portal System. This is important because obtaining reliable and public

data is essential to understand the social reality. For this, the model identified by the acronym SEMMA appropriate to the National Transparency Platform was used, generating five steps for data migration and analysis. Finally, the limitations related to the fact that the guide is designed for a system of the four systems that make up the National Transparency Platform, and future lines of research are disaggregated.

Keywords: open data, transparency, data mining, right to information.

Sumario: I. *Introducción*. II. *El Sistema Nacional de Transparencia en México*. III. *Los datos abiertos en la Plataforma Nacional de Transparencia*. IV. *Guía metodológica*. V. *Conclusiones*. VI. *Bibliografía*.

## I. Introducción

El objetivo de esta investigación es generar una serie de pasos, estructurados en una guía metodológica, para aplicar la minería de datos (MD), y migrar datos de la Plataforma Nacional de Transparencia (PNT), concretamente del Sistema de Portales de Obligaciones de Transparencia (SIPOT). Por esa razón, el apartado se divide en cuatro momentos.

En el primer momento, se analiza el sistema nacional de transparencia en México, el cual reconoce el derecho humano al acceso a la información pública y protección de datos personales. En ese orden de ideas, se puede acceder de manera directa a un apartado de información pública, utilizando la PNT, para obtener información en manos de las autoridades.

La PNT alberga tanto datos abiertos como documentos de carácter público, los cuales están al alcance de cualquier persona. De allí su importancia, ya que pueden ser usados, reutilizados y redistribuidos en un formato accesible. Lo anterior genera una gran base de datos, organizada y confiable, porque los sujetos obligados en todos sus niveles deben transparentar toda la información, definida como *de carácter público*, de manera trimestral, y la misma debe conservarse, es decir, es de naturaleza permanente.

Además, para generar orden y los datos estén homologados, atendiendo a la gran cantidad de sujetos obligados, y con el objetivo de eliminar o reducir cualquier apreciación subjetiva, se elaboran los *Lineamientos técnicos generales*, los cuales determinan los elementos técnicos aplicables a la obligación de transparencia.

El segundo momento teórico analizado está relacionado con los datos abiertos en la PNT, los cuales tienen características normativas que los hacen accesibles, integrales, gratuitos, no discriminatorios, oportunos, permanentes, primarios, legibles, y pueden ser interpretados por equipos electrónicos en for-

matos abiertos y de libre uso. Este apartado analiza por qué los datos abiertos de la PNT son sumamente importantes para la investigación, ya que sirven para sustentar la metodología y efectuar el análisis.

El tercer momento de análisis teórico está relacionado con la MD, que consiste en migrar datos para identificar, extraer, procesar y analizar la información existente. Aquí el primer paso es identificar la base de datos que habrá que migrar, de modo que el descubrimiento de las bases de datos es primordial para aplicar tanto la migración, como las técnicas de análisis.

El cuarto apartado introduce al lector en el modelo SEMMA, como herramienta que genera una guía para el proceso de selección, exploración y modelado de datos, adecuada a la base con la que se va a trabajar, en este caso la PNT.

Analizada la parte teórica se desagrega la guía, que consiste en:

- 1) *Sample*. Muestree: Consiste en identificar la base de datos con la que se va a trabajar, delimitar la investigación a un sujeto, tipo y periodo de información, para generar el universo de datos a analizar, limpiar los datos y obtener la muestra.
- 2) *Explore*. Explore: Permite generar escenarios posibles de investigación, considerando que los *Lineamientos técnicos generales* tienen definiciones y clasificadores específicos, para generar criterios de confiabilidad.
- 3) *Modify*. Modifique: Es el momento en que se otorga la posibilidad de que los datos y los escenarios sean modificados.
- 4) *Model*. Modele: Es la etapa de proceso en donde se comienza con la manipulación de los datos para transformarlos e identificar los datos más significativos
- 5) *Assess*. Evalúe: Se identifica la validez de los resultados y por ende de los datos, para verificar si fueron idóneos.

Por último, se presentan las conclusiones, en donde se desagregan las aportaciones teóricas y metodológicas, así como las limitantes y posibles líneas de investigación.

## II. El Sistema Nacional de Transparencia en México

«La información es poder». Esta frase es una premisa fundamental en la administración pública (Velasco Zamora, 2019), ya que tener información es trascendental para realizar análisis metódicos y racionales (Vivas, 2016), y esto es importante porque genera nuevas formas de participación ciudadana

(OCDE, 2019, p. 17), nuevas estructuras de gobierno (Castillo, 2019), y, por ende, mejores niveles de gobierno y ciudadanía (Ibujés y Crespo, 2019) en el foro de la Organización de las Naciones Unidas (ONU).

En México se reconoce el derecho humano al acceso a la información pública y protección de datos personales (CPEUM, 2019, art. 6o.). Esto implica que todo sujeto puede solicitar documentos en posesión de las autoridades, por medio de solicitudes de transparencia; o bien, de manera directa, puede acceder a un apartado de información pública, utilizando la PNT.

Materializar el derecho humano al acceso a la información pública no ha sido una labor sencilla —por lo menos seis reformas constitucionales dan cuenta de ello (DOF 2007b; 1977; 2007a; 2013; 2014; 2016)—, ya que obtener información de naturaleza pública, permite cristalizar otros derechos. Conjuntamente, se genera un entendimiento objetivo y racional de las actuaciones del Estado, concibiendo niveles de confianza en el gobierno para exigir su transformación; es decir, permite generar nuevas formas de gobernabilidad al coadyuvar con las autoridades (Castro y Rosales, 2020).

En ese orden de ideas se crea el Sistema Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (SNTA), el cual establece procedimientos, instrumentos y políticas en materia de transparencia, acceso a la información y protección de datos personales (LGTAIP, 2015, arts. 27-36).

Una parte del SNTA se ocupa de la generación de datos abiertos digitales de carácter público que pueden ser usados, reutilizados y redistribuidos por cualquier persona (LGTAIP, 2015, art. 3o.), y que están a disposición a través de medios electrónicos (CPEUM, 2019, art. 6o, frac. V) en un formato accesible y bajo criterios de calidad. Este medio electrónico es la PNT, que alberga distintos sistemas (LGTAIP, 2015, arts. 49-52):

- 1) Sistema de solicitudes de acceso a la información.
- 2) Sistema de gestión de medios de impugnación.
- 3) Sistema de portales de obligaciones de transparencia, y
- 4) Sistema de comunicación entre Organismos garantes y sujetos obligados.

Ante ello, esta investigación se centra en el análisis y establecimiento de una guía metodológica para el apartado 3, es decir, para el SIPOT.

Lo anterior implica que los sujetos obligados en todos sus niveles deben transparentar la información definida como *de carácter público* (LGTAIP, art. 70). Para ello, y atendiendo a la cantidad de información y de sujetos obligados, se generan los *Lineamientos técnicos generales* (SNT, 2020), que

establecen los parámetros técnicos deben seguir los sujetos obligados para ingresar al SIPOT, con la intención de que estén homologados y estandarizados (LGTAIP, 2015, art. 31, frac. IV).

Estos *Lineamientos técnicos generales* (SNT, 2020) establecen obligaciones específicas para los sujetos obligados en cuanto a:

- 1) Tiempos en que se debe entregar la información: trimestrales o anual.
- 2) Tipo de obligación: mayormente relacionadas con el manejo de recursos públicos, aunque también se reportan servicios, trámites, directorio, padrón de beneficiarios, entre otros.
- 3) Formato en que se reporta la obligación: hoja de cálculo estandarizada.

En consecuencia, la PNT alberga la información de cada sujeto obligado, en atención a las obligaciones que les corresponden, y se genera una importante base de datos de los tres niveles de gobierno y de personas morales y físicas que obtuvieron recursos públicos para realizar acciones específicas o realizan funciones públicas, de obtención gratuita, homogénea.

### III. Los datos abiertos en la Plataforma Nacional de Transparencia

Los datos abiertos existentes en la PNT deben cumplir con las siguientes características: inicialmente deben ser *accesibles*, es decir, estar disponibles para la mayor cantidad de usuarios; deben ser *integrales*, esto es, que describan a detalle los elementos que lo componen; deben ser *gratuitos*, ya que su obtención no debe generar ningún cobro; y, deben ser *no discriminatorios*. Ante ello, se generan medios tecnológicos para que cualquier persona pueda disponer de los mismos. Además, los datos también deben ser *oportunos*, es decir, deben estar actualizados; *permanentes*, se debe conservar en el tiempo; *de origen primario*, porque devienen de la fuente inicial; y *legibles*, es decir, que pueden ser interpretados por equipos electrónicos en formatos abiertos y de libre uso (LGTAIP, 2015, art. 3o.).

Los datos son sumamente importantes para cualquier toma de decisión; sirven para realizar investigación académica o periodística, de corte cualitativo, cuantitativo o, en su caso, mixto (Torres y Navarro, 2007). También permiten efectuar análisis estadísticos, de eficiencia, calidad o de productividad, entre otros; y son necesarios para que los *softwares* de análisis de datos cualitativos funcionen —SPSS, STATA, Python, Excel, R., etcétera—. La obtención de datos debe estar apegada a una secuencia lógica, o guía metodológica, para

que los resultados puedan ser confiables (Subiyakto, 2015) y den respuesta a la investigación o al planteamiento del problema (Alvesson y Sandberg, 2011).

Mucho se ha abordado metodológicamente sobre la importancia de los datos en las investigaciones sociales (Hernández *et al.*, 2006; 2014; Ansola-behere *et al.*, 2018), y se ha insistido en que “se requiere de datos suficientes, para que la investigación cuente con validez interna, esto es, que hay coherencia en el proceso interno y consistencia en los resultados” (Le Dain, Blanco, Summers, 2013).

Por tal motivo, los datos son importantes, ya que sirven para sustentar la metodología y efectuar el análisis (Sargent, 2011). Sin embargo, también se ha resaltado que los datos no son fáciles de obtener, por el hecho de que el investigador tiene diferentes medios para su obtención: encuestas, entrevistas, datos públicos o privados, ya sean gratuitos o de cobro.

En ese orden de ideas, su obtención y la cantidad de datos a utilizar, es el primer obstáculo que se presenta; el segundo, sería que los datos sean apropiados, precisos y suficientes (Vilalta, 2016). Aquí es donde sobresale el SIPOT, para obtener datos accesibles, integrales, gratuitos, oportunos homologados y estandarización en formatos oficiales, y de naturaleza pública (LGTAIP, 2015, arts. 61 y 65).

#### IV. La minería de datos

El objetivo de aplicar la minería de datos (MD) en cualquier investigación, consiste en aprovechar el acceso a las bases de datos que se desprenden de la información resguardada, clasificada y colocada en ambientes virtuales y que son de acceso público.

La MD permite implementar un proceso metodológico, estableciendo una serie de pasos para mejorar la toma de decisiones, tanto de organismos públicos como privados. Es importante referir que la MD ofrece varias técnicas de análisis de datos. La pretensión es identificar, extraer, procesar y analizar la información existente (Marcano y Talavera, 2007) y, en ese orden de ideas, aplicar la técnica adecuada.

La MD consiente en migrar datos, y aquí resalta su importancia, ya que sirve especialmente para aquellas bases de datos que están constituidas de un gran número de datos (Riquelme y Gilbert, 2006). Es por ello que sus antecedentes se remontan al ámbito privado. Específicamente a los negocios, los cuales requirieron construir herramientas para el análisis de los elementos que se suscitan al interior de la organización y, con ello, mejorar la toma de decisio-

nes, cimentado en elementos objetivos y racionales, es decir, implica concebir un análisis para generar negocios inteligentes (Mikut y Reischl, 2011).

Debe señalarse que la evolución de la MD ha permitido transitar del sector privado al público (Romero y Ventura, 2013; Obenshain, 2004). De allí que la MD esté asociada a la predicción. Por consiguiente, surgen a la par herramientas que complementan el análisis de datos, como los árboles de decisión, ya que el objetivo de la toma de decisiones es pronosticar en un mundo de alternativas, la que sea más benéfica para la organización, o la que genere el menor daño.

En ese orden de ideas, pareciera que el principio de la MD es la migración, como primer paso; y la toma de decisión, como el último —personificada en una representación visual—. Sin embargo, la construcción de herramientas permitió que los pasos fueran heterogéneos, desde el uso de machine learning, el reconocimiento de patrones o pattern recognition (Bouza y Santiago, 2012), o el descubrimiento de conocimiento en bases de datos o knowledge discovery in databases (KDD), que consiste en usar estadística para validar la hipótesis (Asencios, 2004, p. 86).

Ahora bien, migrar datos no es el único paso de esta metodología. Si bien ha sido señalado como el primero, esta investigación afirma que el primer paso implica identificar la base de datos que habrá que migrar (Riquelme et al., 2006, p. 12), para procesarla, clasificarla e implementar asociaciones o patrones (Martínez, 2011). Es por ello que el descubrimiento de las bases de datos es primordial para aplicar tanto la migración como las técnicas de análisis (Lee y Siau, 2001).

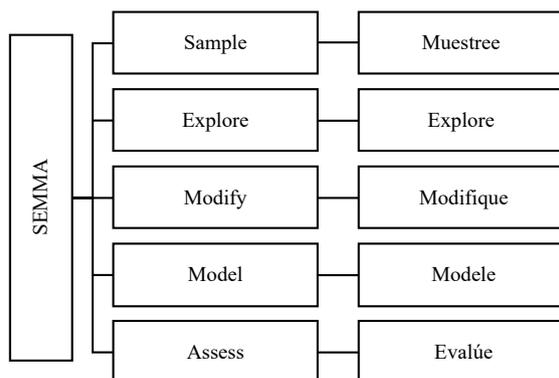
Invariablemente, ante la diversidad de herramientas en el análisis de MD, surge la necesidad de realizar una guía metodológica para minar datos, con el objetivo de dar un correcto tratamiento a los datos; que se ajusten a la necesidad de la organización, y con ello evitar errores en la migración y por ende en los resultados e interpretación (Córdova y Rosales, 2017).

## V. Guía Metodológica

Para esta investigación se considera idónea la aplicación de SEMMA, una herramienta desarrollada por SAS *Institute*. Se considera “idónea” porque establece un guía para el proceso de selección, exploración y modelado de datos para descubrir patrones desconocidos, evaluando aspectos técnicos. Por ello es que existe una etapa de *muestreo* que es sumamente importante, pues genera grados de confiabilidad en los resultados (Miguel, Gordillo y Haed, 2011).

En atención a lo anterior, y considerando los *Lineamientos técnicos generales* (SNT, 2020), y por ello, los parámetros técnicos de los datos abiertos en la PNT, es que se presenta una guía bajo el modelo SEMMA, adecuada a la base de con la que se va a trabajar la PNT. Ver ilustración 1.

Ilustración 1. Acrónimo SEMMA



FUENTE: elaboración propia. Con información de Obenshain 2004

## 1. Sample. *Muestree*

El primer paso implica identificar la base de datos con que se va a trabajar, por lo que se debe:

- 1) Ingresar a la PNT.
- 2) Identificar el apartado de información pública.
- 3) Seleccionar el ámbito de gobierno en que recae la investigación: “Estado o Federación”.
- 4) Delimitar el tipo de información detallada en la PNT como “obligaciones generales o especiales de información” correspondiente a la materia de estudios.
- 5) Delimitar el periodo de investigación, puede ser por año o por trimestre y se puede extraer información de varios años.

El objetivo de este primer paso es delimitar la investigación a un sujeto, tipo y periodo de información, para generar el universo de datos a analizar. Lo anterior implica determinar rangos de búsqueda para descargar los formatos en Excel.

Una vez que se tiene delimitada la información, el siguiente paso consiste en limpiar los datos, ya que, si bien los *Lineamientos técnicos generales* (SNT, 2020) tienen la pretensión de generar una guía para que todos los sujetos obligados carguen la información de manera ordenada, sistematizada y homogénea, lo anterior no siempre es así, y existe la posibilidad del error humano.

La limpieza de datos consiste en identificar si todos los datos son del mismo tipo. De no ser así, el proceso implica codificar los datos para que sean homogéneos; o, en su caso, eliminar datos duplicados o incompletos (Vilalta J, 2016). Una vez realizado lo anterior, se obtendrá una muestra que debe ser estadísticamente representativa para ser utilizada (López y Fachelli, 2017).

## 2. Explore. *Explore*

Explorar los datos implica identificar relaciones útiles para su aplicación, es decir, generar escenarios posibles de investigación (Moine, Gordillo y Haedo, 2011). Lo anterior obliga a identificar el tipo de dato —numérico o textual— y su posible relación con el modelo de corte cualitativo, cuantitativo o mixto. Para ello, los *Lineamientos técnicos generales* (SNT, 2020) y la Ley General de Transparencia (LGTAI, art. 70) proporcionan una gama interesante de escenarios de información, por mencionar algunos:

- 1) Manejo de recursos públicos.
- 2) La remuneración bruta y neta de todos los servidores públicos.
- 3) Gastos de representación y viáticos.
- 4) El número total de las plazas y del personal de base y confianza.
- 5) Contrataciones de servicios profesionales.
- 6) Declaraciones patrimoniales de los servidores públicos.
- 7) Sanciones administrativas definitivas.
- 8) Deuda pública.
- 9) Montos destinados a la comunicación social y publicidad oficial.
- 10) Resultados de las auditorías al ejercicio presupuestal de cada sujeto obligado.
- 11) Licitaciones públicas, adjudicaciones o procedimientos de invitación restringida.

Además, los *Lineamientos técnicos generales* (SNT, 2020) tienen definiciones y clasificadores específicos para generar criterios de confiabilidad, con los que se busca eliminar la interpretación subjetiva de lo que debe reportarse y, por lo tanto, lo que debe entenderse. Lo anterior está definido dentro de los lineamientos como *criterios sustantivos de contenido*.

### 3. Modify. *Modifique*

Identificada las características de los datos, así como el escenario de investigación, previo a manipular los datos se otorga la posibilidad de que los mismos sean modificados en atención a los parámetros previamente definidos, es decir:

- 1) Sujeto.
- 2) Tipo de investigación.
- 3) Periodo de información.
- 4) Tipo de dato, y
- 5) Tipo de modelo.

Esta fase, al igual que las anteriores, está considerada como de *preparación de datos*, y permite modificar los alcances antes de comenzar con la etapa de proceso (Moine, Gordillo y Haedo, 2011).

Además de realizar modificaciones, esta etapa es el primer momento en donde al efectuar la revisión la misma otorga la posibilidad de clasificar los datos en variables e indicadores. Un ejemplo de lo anterior, si se va a realizar un análisis relacionado con recursos públicos, las variables pueden estar relacionadas con el gasto y los indicadores con la moneda en que se realiza el gasto; es decir, implica generar variables homogéneas y significativas para el estudio.

### 4. Model. *Modele*

La etapa de proceso comienza con la manipulación de los datos para transformarlos e identificar los datos más significativos, por lo que, hasta el momento, la información debe estar:

- 1) Concentrada.
- 2) Clasificada.
  - a) Identificada por sujeto, por periodo, por escenario de investigación, por variable e indicador.
- 3) Identificadas las técnicas —estadística descriptiva o inferencial— y el *software* que se utilizará.

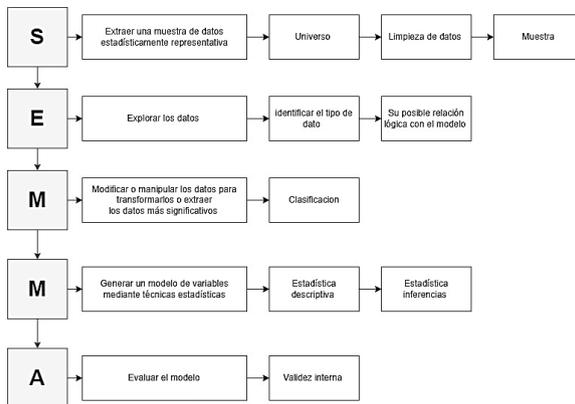
### 5. Assess. *Evalúe*

Evaluar pretende identificar la validez de los resultados y, por ende, de los datos, pues los mismos son el objeto de la investigación. Para ello, se debe resumir las características de estos:

- Tipo de información: pública.
- Medio de obtención: PNT.
- Tipo de investigación: en atención a los *Lineamientos técnicos generales* (SNT, 2020) y la Ley General de Transparencia (LGTAIP, art. 70).
- Ámbito de gobierno: federal o estatal.
- Sujeto obligado o institución: federal o estatal.
- Periodo de investigación: anual o trimestral.
- Cantidad de datos procesados: del universo y de la muestra.
- Tipo de dato: numérico o textual.
- Tipo de modelo: de corte cualitativo, cuantitativo o mixto.
- Variable e indicador: relacionado con los datos y el escenario de investigación.

De la evaluación se debe desprender que los datos fueron en homogéneos, y suficientes para generar una muestra significativa, de no haber sido homogéneos inicialmente, se sometieron a una limpieza para ser estandarizados, además de que existe un elemento congruente con el tipo de investigación, es decir, que fueron idóneos, ver ilustración 2:

Ilustración 2. Técnica de minería de datos



FUENTE: elaboración propia con información de Obenshain (2004).

Esta guía metodológica ofrece la posibilidad de analizar una gran cantidad de datos públicos y gratuitos que existen actualmente en la PNT, con el objetivo de generar investigación académica, periodística o simplemente como un referente para el conocimiento social. Sin embargo, se recalca la necesidad de que el sujeto tenga los instrumentos para saber procesar la información y, en

ese orden de ideas, esté informado desde parámetros de objetividad y racionalidad utilizando para ello la MD.

## V. Conclusiones

Esta investigación generó una guía metodológica para usar la MD en la PNT, específicamente en el SIPOT. Para ello se utilizó y adecuó el modelo identificado con el acrónimo SEMMA, para generar una serie de pasos ordenados y coherentes.

Las aportaciones teóricas que se identificaron están relacionadas con los criterios, clasificadores y definiciones de los *Lineamientos técnicos generales*, que prevén reducir o eliminar cualquier criterio subjetivo en el reporte de obligaciones de transparencia, y, con ello, homogenizar y generar criterios de confiabilidad en el uso de los datos.

Las aportaciones metodológicas están vinculadas con el modelo SEMMA, el cual se adecua a la PNT, generando cinco pasos que consisten en: 1) delimitar la investigación para generar el universo de datos, limpiar los datos y obtener la muestra; 2) generar escenarios posibles de investigación; 3) permitir que los datos y los escenarios sean modificados; 4) manipular los datos para transformarlos e identificar los datos más significativos; y 5) evaluar la validez de los resultados.

Por último, las limitaciones de esta investigación están relacionadas con que la guía solo está diseñada a un sistema de los cuatro sistemas que compone la PNT. Es decir, sólo está diseñada para el SIPOT. Por lo que, para futuras líneas de investigación, se puede considerar el análisis del sistema de solicitudes de acceso a la información; el sistema de gestión de medios de impugnación; y el sistema de comunicación entre organismos garantes y sujetos obligados.

## VI. Bibliografía

- Alvesson, M. y Sandberg, J. (2011). Generating research questions through problematization. *The Academy of Management Review*, 36(2). <https://www.jstor.org/stable/41318000>
- Ansolabehere, K., Cortés, F., Martínez, L. y Zaremborg, G. (2018). *Diseños de investigación: metodología en tesis de ciencias sociales*. Flacso.
- Asencios, V. V. (2004). Data Mining y el descubrimiento del conocimiento. *Revista de la Facultad de Ingeniería Industrial*, 7(2).

- Cámara de Diputados del H. Congreso de La Unión (2015). Ley General de Transparencia y Acceso a la Información Pública.
- Cámara de Diputados del H. Congreso de La Unión (2019). *Constitución Política de los Estados Unidos Mexicanos*. Cámara de Diputados del H. Congreso de la Unión. [http://www.diputados.gob.mx/LeyesBiblio/pdf/1\\_201219.pdf](http://www.diputados.gob.mx/LeyesBiblio/pdf/1_201219.pdf)
- Castillo, A. M. C. (2019). Gobierno cercano y moderno: una mirada desde los estándares internacionales del acceso a la información pública. *Estudios en Derecho a la Información*, 1(8). <https://doi.org/10.22201/ijj.25940082e.2019.8.13883>
- Castro Coria, E. G. y Rosales Huerta, C. (2020). El derecho a la información como llave de acceso a nuevas formas de gobierno. *Estudios en Derecho a la Información*, 1(10). <https://doi.org/10.22201/ijj.25940082e.2020.10.14660>
- Córdova Gallegos, O. J. y Rosales Gallardo, C. P. (2017). *Desarrollo de una guía metodológica sobre minería de datos*. Pontificia Universidad Católica del Ecuador. <http://repositorio.puce.edu.ec/handle/22000/13181>
- Diario Oficial de la Federación (1977). Reforma al artículo 60. *Comisión Permanente del Honorable Congreso de la Unión*. [https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM\\_ref\\_086\\_06dic77\\_ima.pdf](https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_086_06dic77_ima.pdf)
- Diario Oficial de la Federación (2007a). Reforma al artículo 60. *Comisión Permanente del Honorable Congreso de la Unión*. [https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM\\_ref\\_178\\_13nov07\\_ima.pdf](https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_178_13nov07_ima.pdf)
- Diario Oficial de la Federación (2007b). Reforma al artículo 60. *Comisión Permanente del Honorable Congreso de la Unión*. [https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM\\_ref\\_174\\_20jul07\\_ima.pdf](https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_174_20jul07_ima.pdf)
- Diario Oficial de la Federación (2013). Reforma al artículo 60. *Comisión Permanente del Honorable Congreso de la Unión*.
- Diario Oficial de la Federación (2014). Reforma que se reforman y adicionan diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, en materia de transparencia. *Comisión Permanente del Honorable Congreso de la Unión*. [https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM\\_ref\\_215\\_07feb14.pdf](https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_215_07feb14.pdf)
- Diario Oficial de la Federación (2016). Decreto por el que se declaran reformadas y derogadas diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, en materia de la reforma política de la Ciudad de México. *Comisión Permanente del Honorable Congreso de la Unión*. [https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM\\_ref\\_227\\_29ene16.pdf](https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_227_29ene16.pdf)

- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, M. del P. (2006), *Metodología de la investigación* (4a. ed.).
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, M. del P. (2014). *Metodología de la Investigación* (6a. ed. ).
- Ibujés Villacís, J. M. y Franco Crespo, A. (2019). Uso de las TIC y relación con los Objetivos de Desarrollo Sostenible en Ecuador. *Retos*, 9(17). <https://orcid.org/0000-0001-8439-3048>
- PNT. (2020). *Página de inicio*. <https://www.plataformadetransparencia.org.mx/web/guest/inicio>
- Le Dain, M. A., Blanco, E. y Summers, J. D. (2013, 19-22 de agosto). *Assessing design research quality: Investigating verification and validation criteria*. International Conference on Engineering Design, ICED13. Sungkyunkwan University. Corea.
- Lee, S. J. y Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems*, 101(1). <https://doi.org/10.1108/02635570110365989/FULL/XML>
- López-Roldán, P. y Fachelli, S. (2017). El diseño de la muestra. En P. López-Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa*. <https://ddd.uab.cat/record/185163>
- Martínez Luna, G. L. (2011). Minería de datos: Cómo hallar una aguja en un pajar. *Ciencia*, 62(3), 55-63.
- Marcano Aular, Y. J. y Talavera Pereira, R. (2007). Minería de Datos como soporte a la toma de decisiones empresariales. *Opcion*, 23(52). [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S1012-15872007000100008](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1012-15872007000100008)
- Moine, J. M., Gordillo, S. y Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. En: *XVII Congreso Argentino de Ciencias de la Computación*. <http://sedici.unlp.edu.ar/handle/10915/18749>
- Mikut, R. y Reischl, M. (2011). Data mining tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(5). <https://onlinelibrary.wiley.com/doi/full/10.1002/widm.24>
- N. Bouza, C. y Santiago, A. (2012). La minería de datos: arboles de decisión y su aplicación en estudios médicos. *Modelación Matemática de Fenómenos del Medio Ambiente y la Salud*, 2. [https://rideca.cs.buap.mx/web/files/articulo\\_itBUo0uWIAaJENf.pdf](https://rideca.cs.buap.mx/web/files/articulo_itBUo0uWIAaJENf.pdf)
- Obenshain, M. K. (2004). Application of Data Mining Techniques to Healthcare Data. *Infection Control & Hospital Epidemiology*, 25(8). <https://www.cambridge.org/core/journals/infection-control-and-hospital-epidemiology>

[gy/article/abs/application-of-data-mining-techniques-to-healthcare-data/7EE5E7B1FA8B1C535FBC7A3881EC42E0](https://www.juridicas.unam.mx/revistas/ijj/revista.asp?area=11&I=17&V=18782)

- OCDE (2019). *La Integridad Pública en América Latina y el Caribe 2018-2019*. OCDE. <https://www.oecd.org/gov/ethics/integridad-publica-america-latina-caribe-2018-2019.pdf>
- Riquelme, J. C., RUIZ, R. y Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, (29). <http://www.aepia.org>
- Romero, C. y Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1). <https://onlinelibrary.wiley.com/doi/full/10.1002/widm.1075>
- Sargent, R. G. (2011). Verification and validation of simulation models. En S. Jain, R. R. Creasey, J. Himmelspach, K. P. White y M. Fu (eds.), *Proceedings of the 2011 Winter Simulation Conference*.
- SNT (2020). *Lineamientos Técnicos Generales para la publicación, homologación y estandarización de la información de las obligaciones establecidas en el Título Quinto y en la fracción IV del artículo 31 de la Ley General de Transparencia y Acceso a la Información Pública, que deben de difundir los sujetos obligados en los portales de Internet y en la PNT*. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales. [https://www.transparencia.ipn.mx/Apoyo/SIPOT/LTG\\_DO28122020.pdf](https://www.transparencia.ipn.mx/Apoyo/SIPOT/LTG_DO28122020.pdf)
- Subiyakto, A., Ahlan, A. R., Putra, S. J. y Kartiwi, M. (2015). Validation of Information System Project Success Model: A Focus Group Study. *SAGE Open* 5(2). <https://journals.sagepub.com/doi/full/10.1177/2158244015581650>
- Torres Hernández, Z. y Navarro Chávez, J. C. L. (2007). *Conceptos y principios fundamentales de epistemología y de metodología*. Universidad Michoacana de San Nicolás de Hidalgo. [https://www.dropbox.com/s/nrsofnh0jytfys5/Conceptos\\_y\\_Principios\\_Fundamentales\\_de\\_Epistemología\\_y\\_de\\_Metodología.pdf?dl=0](https://www.dropbox.com/s/nrsofnh0jytfys5/Conceptos_y_Principios_Fundamentales_de_Epistemología_y_de_Metodología.pdf?dl=0)
- Velasco Zamora, C. H. (2019). *La desviación de poder*. Leyer Editores.
- Vilalta J., C. (2016). *Análisis de Datos*. Centro de Investigación y Docencia Económicas.
- Vivas, Á. P. (2016). Gobierno abierto, el valor social de la información pública. *Estudios en Derecho a la Información*, 1(2).