

Explorando los desafíos lógicos en la visualización y análisis de datos en arquitecturas de *Big data*: un enfoque en falacias, sesgos y paradojas



Exploring Logical Challenges in Data Visualization and Analysis in *Big data* Architectures: A Focus on Fallacies, Biases, and Paradoxes

DOI: 10.22458/rna.v15i1.5150



Roberto Antonio Zárate-Sánchez¹

1.Universidad Nacional de Costa Rica, UNA, Heredia, Costa Rica
roberto.zarate.sanchez@est.una.ac.cr

Recepción: 16 de enero de 2024

Corrección: 14 de mayo de 2024

Aceptación: 10 de junio de 2024

RESUMEN

Esta investigación cualitativa se centra en identificar y caracterizar los errores lógicos frecuentes en el análisis y la visualización de datos dentro de las arquitecturas de *Big data*, a través de una revisión bibliográfica exhaustiva, en la que se categorizan los errores en falacias, sesgos y paradojas. Este estudio busca servir como guía para profesionales en entornos públicos y privados, así como señalar áreas de investigación relacionadas con la epistemología y la ética en el ámbito del *Big data*. El texto plantea dos preguntas: ¿Cuáles son los errores lógicos más comunes que se encuentran en el análisis y la visualización de datos dentro de las arquitecturas de *Big data*? ¿Cómo pueden abordarse estos errores para mejorar tanto la calidad de la toma de decisiones como la ética en este campo? El artículo tiene cuatro objetivos: identificar y caracterizar las falacias, sesgos y paradojas más comunes en el análisis y la visualización de datos en arquitecturas de *Big data*; proporcionar orientación y conocimiento a profesionales que trabajan en el análisis y la visualización de datos en entornos públicos y privados; destacar la importancia de la epistemología y la ética en el contexto del *Big data*; y establecer líneas de investigación adicionales relacionadas con la mejora de la calidad de los análisis de datos y la promoción de prácticas éticas.

ABSTRACT

This qualitative research focuses on identifying and characterizing common logical errors in data analysis and visualization within *Big data* architectures. This is done through a profound literature review, categorizing errors into fallacies, biases, and paradoxes. The study aims to serve as a guide for professionals in both public and private realms and to highlight areas of research related to epistemology and ethics in the realm of *Big data*. The text raises two questions: what are the most common logical errors found in data analysis and visualization within *Big data* architectures, and how can these errors be addressed to improve both decision-making quality and ethics in this field? In the other hand, the article has four objectives: to identify and characterize the most common fallacies, biases, and paradoxes in data analysis and visualization in *Big data* architectures; to provide guidance and knowledge to professionals working in data analysis and visualization in public and private realms; to emphasize the importance of epistemology and ethics in the context of *Big data*; and to establish additional lines of research related to improving the quality of data analysis and promoting ethical practices.

PALABRAS CLAVE:

BIG DATA, FALACIAS, SESGOS, PARADOJAS, ANÁLISIS DE DATOS.

KEYWORDS:

BIG DATA, FALLACIES, BIASES, PARADOXES, DATA ANALYSIS.





RESUMÉ

Cette recherche qualitative se concentre sur l'identification et la caractérisation des erreurs logiques courantes dans l'analyse et la visualisation des données au sein des architectures *Big data*. Cette analyse est réalisée à travers un examen approfondi de la littérature, classant les erreurs en fallacies, biais et paradoxes. L'étude vise à servir de guide pour les professionnels des secteurs public et privé et à mettre en lumière les domaines de recherche liés à l'épistémologie et à l'éthique dans le domaine du *Big data*. Le texte soulève deux questions : quelles sont les erreurs logiques les plus courantes dans l'analyse et la visualisation des données au sein des architectures *Big data*, et comment ces erreurs peuvent-elles être traitées pour améliorer à la fois la qualité de la prise de décision et l'éthique dans ce domaine ? D'autre part, l'article a quatre objectifs : identifier et caractériser les fallacies, biais et paradoxes les plus courants dans l'analyse et la visualisation des données dans les architectures *Big data* ; fournir des orientations et des connaissances aux professionnels travaillant dans l'analyse et la visualisation des données dans les secteurs public et privé ; mettre l'accent sur l'importance de l'épistémologie et de l'éthique dans le contexte du *Big data* ; et établir des lignes de recherche supplémentaires visant à améliorer la qualité de l'analyse des données et à promouvoir des pratiques éthiques.

MOTS-CLÉS:

BIG DATA, SOPHISMES, BIAIS, PARADOXES, ANALYSE DE DONNÉES

RESUMO

Esta pesquisa qualitativa foca na identificação e caracterização dos erros lógicos comuns na análise e visualização de dados dentro das arquiteturas de *Big data*. Isso é feito por meio de uma revisão profunda da literatura, categorizando os erros em falácias, vieses e paradoxos. O estudo tem como objetivo servir como um guia para profissionais nos setores público e privado e destacar áreas de pesquisa relacionadas à epistemologia e ética no campo do *Big data*. O texto levanta duas questões: quais são os erros lógicos mais comuns encontrados na análise e visualização de dados dentro das arquiteturas de *Big data*, e como esses erros podem ser abordados para melhorar tanto a qualidade da tomada de decisão quanto a ética nesse campo? Por outro lado, o artigo tem quatro objetivos: identificar e caracterizar as falácias, vieses e paradoxos mais comuns na análise e visualização de dados em arquiteturas de *Big data*; fornecer orientação e conhecimento para profissionais que trabalham na análise e visualização de dados nos setores público e privado; enfatizar a importância da epistemologia e ética no contexto do *Big data*; e estabelecer linhas adicionais de pesquisa relacionadas à melhoria da qualidade da análise de dados e à promoção de práticas éticas.

PALAVRAS-CHAVE:

BIG DATA, FALÁCIAS, VIESES, PARADOXOS, ANÁLISE DE DADOS.

INTRODUCCIÓN

Las sociedades actuales están caracterizadas por la presencia de datos de diversa índole en todos los ámbitos de la vida, especialmente a partir del uso de diferentes dispositivos tales como celulares, relojes inteligentes y computadoras. Lo anterior no implica que los datos no hayan existido antes, sino que la cantidad y la intensidad han incrementado en los actuales contextos digitales. Además, los datos se han convertido en un activo económico muy relevante; incluso se les ha llamado el “nuevo oro” (Servando, 2020). Es importante tener en cuenta que estos también poseen implicaciones éticas significativas, como las relacionadas con la privacidad, el control político, la libertad y el consentimiento.¹

Asimismo, cabe señalar que ha surgido, para describir a las sociedades actuales, el concepto de dataísmo que, acuñado hace una década, hace alusión al papel central de los datos para entender la realidad, en sentido filosófico, en todos sus ámbitos (Brooks, 2013). Por otra parte, según Drayson y Bashir (s.f), en una obra titulada *The Evolution of Data Management*, a nivel histórico se pueden identificar diferentes etapas o eras de la gestión de datos, siendo la actual *the digitally orchestrated age*, la cual se caracteriza por la completa digitalización de las actividades empresariales, la automatización, el machine learning y el papel de la inteligencia de negocio como eje central.

En otras palabras: vivimos en una época atravesada por los datos. Sin embargo, es fundamental tener en cuenta que un dato, entendido como una representación o construcción simbólica de una variable determinada, por sí mismo, no dice nada. Incluso, después de su procesamiento, los datos deben ser dotados de sentido. Es un error creer que, por ejemplo, son como el conjunto de símbolos verdes que se deslizan frente a Neo en *Matrix*², los cuales explicaban, por sí mismos, una supuesta realidad objetiva³. T.S Elliot, poeta estadounidense, señalaba este error, de una manera más amplia, en su poema titulado *El primer coro de la roca*⁴: “¿Dónde está la sabiduría que hemos perdido en conocimiento? / ¿Dónde el conocimiento que hemos perdido en información?⁵”. Es decir, el conocimiento o la comprensión de un fenómeno requiere de algo más que hechos aislados, proposiciones individuales o, en este caso, datos. Es importante que se enmarquen en contextos, narrativas, intereses y discursos específicos, por lo que es fundamental tener en cuenta ciertos sesgos, paradojas y falacias que, al realizar la contextualización y dotación de sentido, llevan a que los datos no se apeguen a ciertas realidades y que no sean adecuados para la toma de decisiones.

Teniendo en cuenta lo anterior, el propósito de este artículo es ofrecer una reflexión y sistematización de los principales sesgos, paradojas y falacias que afectan el proceso de comprensión, análisis y visualización de un conjunto específico de datos y sus asociaciones. Es crucial destacar que el análisis se centra en cómo se utilizan los datos para tomar decisiones, no necesariamente en los procesos de automatización algorítmica. Para ello, se realiza inicialmente un acercamiento a los conceptos de análisis y visualización de datos en el contexto de las capas de una arquitectura de *Big data*, así como a los de sesgo, paradoja y falacia. Luego, se detallan los aspectos más relevantes dentro de las capas previamente definidas; finalmente, se presentan las reflexiones conclusivas.

APUNTES METODOLÓGICOS

En el nivel metodológico, el artículo adopta un enfoque documental y bibliográfico, el cual, según García y Pérez (2022), implica la recuperación, el análisis y la interpretación crítica de información que ha sido recopilada y analizada previamente por otros investigadores. A la luz de esto, la pregunta que guía esta investigación es: ¿cuáles son las principales falacias, sesgos y paradojas que afectan los procesos de visualización y análisis de datos en las arquitecturas de *Big data*? Para responder a esta pregunta, se realizó una revisión de artículos científicos y académicos, ejemplos en blogs y periódicos, casos de estudio, así como manuales y libros relevantes en el campo de la analítica de datos. Durante la presentación de los resultados, se citan los documentos y ejemplos consultados en cada caso.

1 Un aporte relevante para comprender esto es la obra *La era del capitalismo de la vigilancia* de la escritora Shoshana Zuboff. Fue escrito en el 2020.
2 Película de 1999.
3 Supuesta en el sentido de que la realidad explicada por los datos era ficticia e ilusoria.
4 Es un poema de 1934, con diferentes interpretaciones, en el que, de acuerdo a Atkins (2013), se realiza un acercamiento a ciertas ideas de comunidad en contraposición a la atomización social.
5 Traducción de Jorge Luis Borges. Recuperado de la plataforma de poesía Triartans (2022).

En cuanto al tipo de investigación, es relevante destacar que es cualitativa, centrada en el aspecto conceptual, es decir, en los significados (Barrantes, 2016) y cómo estos se interrelacionan. Específicamente, consiste en identificar, a partir de la caracterización teórica de las capas de visualización y análisis, las falacias, sesgos y paradojas sistematizadas en diversas fuentes literarias⁶ sobre estadística, analítica de datos, epistemología y lógica formal, como también en ejemplos comunes en blogs especializados y prensa que pueden estar presentes en estos procesos de *Big data*. Este proceso se muestra en la figura 1.

Figura 1. Proceso metodológico de comparación conceptual



Fuente: Elaboración propia

En cuanto al proceso de selección de los textos y fuentes, es fundamental considerar que, al abordar sesgos, falacias y paradojas, conceptos ampliamente estudiados en la filosofía, se realizó una revisión de enciclopedias como la *Stanford Encyclopedia of Philosophy* y la *Routledge Encyclopedia of Philosophy*, así como de elaboraciones nacionales de instituciones públicas, específicamente en materiales de la materia de Filosofía del MEP.

El objetivo fue identificar qué tipos de errores lógicos, tal y como se conceptualizan teóricamente más adelante, pueden manifestarse en las arquitecturas de *Big data*. Esta revisión se complementó con la consulta de literatura sobre errores estadísticos, con base en las categorías de sesgo, paradoja y falacia. Además, se amplió la investigación a textos de otras disciplinas, como medicina y física, para determinar si estos errores podrían presentarse de manera similar en los procesos de analítica de datos. Mediante la diversidad de referencias estudiadas se planteó una comprensión más completa de la posible presencia y efectos de estos errores en el modelo de arquitectura propuesto.

Cabe señalar que lo anterior se profundizó a partir de artículos especializados en ciencias sociales y filosofía, con el fin de brindar una aproximación actual y crítica en torno a los sesgos, paradojas y falacias estudiadas. Además, se revisaron documentos y guías de actores importantes en el sector, como por ejemplo

6 Libros, manuales, diccionarios, artículos científicos, ensayos académicos y prensa.

la guía práctica de NTT⁷, los aportes de DataHeroes⁸ y la conceptualización en torno a las arquitecturas propuesta por IBM y por Solano y Leiva⁹ (2014). Asimismo, se tomaron en cuenta sistematizaciones de experiencias y estudios de caso tanto en artículos académicos como en notas periodísticas.

De igual forma, se estudiaron blogs populares, tales como el de *Litera*¹⁰, *Geckoboard*¹¹, *VisualCapitalist*¹² y *Connectif*¹³, los cuales tienen entradas referentes a *data fallacies*. Cabe señalar que, al ser blogs, suelen simplificar la información, como por ejemplo no tomar en cuenta la distinción entre falacias, sesgos y paradojas¹⁴, ni tampoco discutir los problemas filosóficos y de fundamentación epistemológica¹⁵. Es importante destacar que las publicaciones académicas con respecto a estos errores lógicos en *Big data* son escasas, por lo que se recurrió a estos blogs como fuente de información, ya que sistematizan las experiencias de empresas y usuarios.

El proceso de selección de los errores lógicos, estuvo determinado por la construcción teórica en torno a la definición de sesgo, falacia y paradoja, es decir, el análisis consistió en la identificación de la coherencia conceptual entre los errores lógicos generales y las definiciones teóricas problematizadas. En otras palabras, se analizó qué errores se ajustaban a las definiciones brindadas en el marco teórico de la investigación.

Respecto a los alcances y limitaciones, es importante tener en cuenta que la propuesta es meramente bibliográfica y conceptual, por lo que no profundiza en otros aspectos que pueden estar presentes en las dinámicas en las que se enmarcan los errores lógicos: cultura organizacional, consideraciones sociológicas, tecnociencia, teoría del poder, etc. Además de lo anterior, la investigación plantea un registro inicial de estos errores, sin embargo, no es un inventario final, ya que el artículo se plantea como una propuesta que pueda ser ampliada y discutida por trabajos futuros.

Consideraciones teóricas

Se presentan acá los conceptos relacionados a diferentes posturas teóricas que fundamentan la investigación. Así, se profundiza, especialmente, en el concepto de *Big data* y sus arquitecturas, como también en los sesgos, las falacias y las paradojas. Entonces, a nivel metodológico, se tiene en cuenta lo expuesto en el apartado anterior, es decir, el marco teórico que permite comparar los conceptos de visualización y análisis con los diferentes errores identificados, es decir, los relaciona y, por tanto, los justifica.

ANÁLISIS Y VISUALIZACIÓN DE DATOS EN *BIG DATA*

El *Big data* como concepto de análisis es reciente, debido a que fue utilizado por primera vez en la década de los noventa. De esta forma, cabe señalar que, a pesar de ser un concepto ubicuo, no existe un consenso claro en torno a la manera en la que se acuña el término: John Mashey en Silicon Valley, el artículo de Michael Cox y David Ellsworth de la NASA en 1998, Weiss e Indurkha en el ámbito de la computación en 1998 o las reformulaciones en el ámbito de la genética (Diebold, 2012; Tapia, 2022). Ahora, para el presente artículo no se problematiza este aspecto, puesto que el interés no es su historia, sino, más bien, su contenido conceptual. Además de que se debe tener en cuenta que el *Big data* aparece en un contexto determinado y a partir de algunos antecedentes e hitos relevantes a nivel histórico: la memoria virtual de Fritz-Rudolf Güntsch, la ley bibliométrica del aumento exponencial de Derek Price, la ley de Parkinson de Tjomsland, como también la automatización (Sánchez, 2019).

7 Nippon Telegraph and Telephone Corporation, una de las principales compañías asiáticas.

8 Importante plataforma para el acceso a bibliotecas datos para procesos de machine learning.

9 Uno de los pocos trabajos en el ámbito costarricense que presenta una propuesta de arquitectura, además de que fue realizada teniendo en cuenta las principales empresas del sector.

10 Se puede consultar en el siguiente link: <https://www.litera.com/>

11 Acceder en: <https://www.geckoboard.com/best-practice/statistical-fallacies/>

12 Se puede acceder al blog en el siguiente enlace: <https://www.visualcapitalist.com/>

13 Revisar en el siguiente enlace: <https://connectif.ai/>

14 Parte de la justificación de la presente investigación consiste, precisamente, en la necesidad de sistematizar de manera crítica y rigurosa información que se encuentra dispersa y presentada con poca profundidad.

15 Un documento importante a tomar en cuenta es la tesis titulada *El problema epistemológico de los Big Data* en la producción de conocimiento científico, elaborada por Sergio Martín en el 2023.

De manera similar, si bien no existe una definición totalmente compartida de lo que es *Big data* (Camargo-Vega, Camargo-Ortega y Joyanes-Aguilar, 2014), se puede entender como un conjunto de activos de información de gran variedad, alta velocidad y volumen, que requiere formas innovadoras, en el marco de los avances tecnológicos del siglo XXI, de procesamiento y análisis, es decir, no con base en los métodos tradicionales. Cabe señalar que el fin del *Big data*, dentro de los procesos de gestión pública y privada, es el procesamiento de información capaz de propiciar un conocimiento basado en la identificación y comprensión de asociaciones y patrones relevantes para la toma de decisiones, así como para la automatización de procesos (Hernández-Leal, Duque-Méndez y Moreno-Cadavid, 2017). En el caso de la presente investigación, se enfatiza en el primer aspecto, es decir, no necesariamente en la automatización. Dentro de la literatura del *Big data* se suele asumir la existencia de una serie de adjetivos conocidos como las V: Variedad, Veracidad, Valor, Volumen y Velocidad (Márquez, 2020). Estos adjetivos describen las características ideales que debería tener un conjunto de datos categorizado como *Big data*.

Además, en gestión de datos se suele hablar de arquitecturas, mismas que son la estructura de los diferentes sistemas de gestión de datos empresariales, es decir, la forma en que se organizan los componentes y la manera en la que interactúan entre sí. Aquí cabe señalar que están modeladas por los servicios que ofrecen las empresas que controlan el sector: IBM, Microsoft, Oracle y Cloudera (Leiva y Solano, 2014). Así, pues, se debe considerar, para entender una arquitectura de datos, el concepto del ciclo de vida de la data. En general, se suele asumir, tal y como señalan El Arass y El Souissi (2018), que existen cinco fases generales: adquisición, almacenamiento, procesamiento, uso y eliminación. Cabe señalar que tienen relación con las capas lógicas, las cuales, según Mysore, Shrikant y Jain (2014), son obtención, tratamiento, análisis y visualización. Ahora bien, las capas lógicas se refieren al diseño o abstracción, mientras que, por otra parte, la arquitectura es más concreta y se refiere a la materialización de ese diseño. En otras palabras, las capas lógicas son los elementos abstractos generales que le brindan coherencia al modelo teórico y a la materialización de la arquitectura.

Respecto a esto, a nivel nacional, en uno de los pocos artículos académicos elaborados en torno a arquitectura de datos, en este caso *Big data Analytics: propuesta de una arquitectura* (Solano y Leiva, 2014), las personas autoras plantean cuatro capas: origen de datos, recuperación y transformación, análisis y visualización. Para el presente trabajo, se enfatizó solo en las capas de análisis y visualización, es decir, la primera, entendida como el proceso orientado a identificar, mediante herramientas matemáticas, econométricas y estadísticas, las tendencias, asociaciones, patrones y, en general, información útil para la toma de decisiones; la visualización, por su parte, se refiere a la forma en la que se sistematiza y presenta la información (Valero, 2014). Esta última se encuentra ligada al *Storytelling*, que es la práctica de generar, a partir de los datos, un relato coherente (Nussbaumer, 2015). Es, en muchas ocasiones, a la hora de desarrollar la narrativa y el contexto discursivo de sentido, en la que se comenten la mayor cantidad de falacias, sesgos y paradojas.

Sesgos, paradojas y falacias comunes en análisis y visualización de datos

Al trabajar con datos, se debe entender que estos no son entes puros, es decir, son recolectados, interpretados, analizados y presentados por personas, en el caso de *Big data*, con ayuda de inteligencia artificial, por lo que pueden verse afectados por sesgos y falacias. Además, puede haber intereses políticos, dado que en muchos casos los nuevos datos generan resistencias, es decir, contradicen narrativas o creencias instauradas. En la literatura de analítica de datos, esto se suele ilustrar mediante el efecto Semmelweis, entendido como la acción de rechazar evidencia que contradice normas, paradigmas y creencias preestablecidas.¹⁶ (Gupta, Saini, Oberoi, Kalra y Nasir, 2020).

Con respecto a las arquitecturas de *Big data*, especialmente en la etapa de visualización de los datos, en el ámbito organizacional, institucional o empresarial, estos sesgos pueden llevar a que las decisiones y

¹⁶ Actualmente existen discusiones filosóficas en torno al efecto Semmelweis: su relación con el concepto de paradigma y la inducción en sentido de Popper (Scholl, 2012, Parusniková, 2023), así como la propuesta de inferencia causal (Scholl, 2012). Además, en el contexto de la analítica, se ha utilizado como ejemplo de la importancia del *Storytelling* (Dykes, 2016); sin embargo, este acercamiento es simplista, ya que asume que si Semmelweis hubiera presentado sus resultados de otra forma, quizás hubieran sido aceptados, pero no toma en cuenta las dinámicas sociales de los paradigmas.

sus acciones respectivas no tengan los efectos deseados o que, incluso, tengan consecuencias contrarias a lo planteado en las estrategias e instrumentos de planificación. Esto se conoce como el efecto cobra, entendido como el desarrollo de acciones que conllevan el efecto opuesto al esperado, agravando en muchas ocasiones el problema que se intenta solucionar. Su nombre proviene del libro *Der kobra effekt*, escrito por el economista Horst Siebert en 2001. Hace alusión a la decisión tomada por el gobierno de la India, ante la presencia de muchas cobras en la ciudad, de pagar por cada cobra muerta, lo que llevó a que las personas las reprodujeran en sus casas para matarlas y recibir la recompensa, provocando, al final, una mayor cantidad de cobras en la ciudad. (Ventura, 2020).

Volviendo a Semmelweis, a nivel empresarial y de *management*, especialmente en *startups*, existe un error común debido a una creencia compartida: un DAU/MAU de aproximadamente 45 %, es decir, similar al de Facebook en sus inicios, indica éxito. Sin embargo, más allá de las creencias generales en torno a esta métrica, se encuentran una serie de hechos y evidencias que se suelen ignorar porque contradicen la idea de que el DAU/MAU es un indicador concluyente por sí mismo: muchas de las personas usuarias no usan el núcleo de la aplicación, es decir, sus características centrales; así como que el DAU/MAU no describe el uso que se le da a la aplicación. Razonar de manera errónea, en este caso, sería asumir que, en una lectura superficial a partir del DAU/MAU, la *startup* es exitosa. Una situación similar se suele presentar con los KPI o indicadores claves de calidad, esto cuando se enfatiza en uno o dos sin tomar en cuenta otros o cuando no se los analiza en conjunto. Es una simplificación de los problemas, lo que se expone más adelante al hablar del efecto McNamara.

En cuanto a uno de los errores estudiados en este artículo, en este caso las falacias, estas se refieren a diversos tipos de argumentos formalmente erróneos e incorrectos, pero que, en muchas ocasiones, suelen ser, en sentido psicológico, convincentes. A nivel cognitivo, pueden aceptarse de manera intuitiva, es decir, parecen razonables dentro de determinados marcos de sentido o narrativas; sin embargo, a nivel estructural, el paso de las premisas a la conclusión es equivocado. Es debido a lo anterior que su identificación es de gran importancia, ya que, como se expone más adelante, en muchas ocasiones se realizan de forma adrede, pero, en otras, de forma inconsciente o intuitiva, ya que, por ejemplo, con respecto a la falacia *ad hominem*, muchas personas la emplean dentro de sus procesos argumentativos sin darse cuenta.

Un sesgo se refiere, en términos generales, a un defecto de interpretación en el que se asigna un valor desigual a favor o en contra de, en este caso, unas determinadas conclusiones obtenidas a partir de un *dataset* o un conjunto de datos específico. Cabe señalar que las conclusiones se utilizan como argumentos a favor o en contra de ciertas posibles decisiones o acciones, por lo que los sesgos, en muchos casos, son sistemáticos, es decir, no aleatorios o azarosos (Manterola y Otzen, 2015).

Ahora bien, una paradoja es una proposición, teoría o idea que resulta ilógica, incoherente o contradictoria. En este caso, como se expone más adelante al describir la Paradoja de Simpson, son proposiciones o conclusiones que, en apariencia, resultan contradictorias al sentido común o a un grupo de creencias preestablecidas. Asimismo, estas han sido muy populares para ilustrar ideas o teorías en ciencia y filosofía: la paradoja del abuelo, influyente en las reflexiones en torno a los viajes en el tiempo; la paradoja de Epiménides, muy relevante en la matemática del siglo XX; el gato de Schrödinger en el campo de la cuántica o la paradoja de los gemelos utilizada por Einstein para reflexionar sobre la teoría de la relatividad especial. En este caso, más que la paradoja en sí, en analítica lo que se estudia es la respuesta que se le suele dar.

Seguidamente una vez definidas las falacias, los sesgos y las paradojas, se caracterizan algunas de las más comunes y relevantes en el ámbito del análisis y visualización de datos en el contexto de las capas de análisis y visualización en las arquitecturas de *Big data*. Es decir, como se explicó, se describen los errores que se comenten al intentar dotar de sentido a los datos.

RESULTADOS

En este punto se exponen, a partir de la revisión documental y bibliográfica, los principales errores, para este efecto, sesgos, paradojas y falacias, identificados para las fases de análisis y visualización, teniendo en cuenta las definiciones brindadas con anterioridad.

Cherry Picking

Es un sesgo muy similar al dragado de datos o *data dredging*, que consiste en seleccionar, de un conjunto amplio, solo los datos que refuerzan y consolidan la narrativa elegida, es decir, la que es funcional para ciertos intereses. En el ámbito de la analítica de datos, por ejemplo, con un mismo *dataset*, dos personas pueden llegar a conclusiones diferentes. En este caso, si se realiza de forma adrede, además de un error lógico o epistemológico, se puede considerar como un problema ético.

Cabe señalar que es una acción y un error muy común, en algunos casos, en el desarrollo general de la ciencia. Por ejemplo, recientemente se dio una discusión, en el ámbito de la filosofía y la epistemología, en torno al realismo científico¹⁷, puesto que Kyle Stanford y Peter Vickers realizaron una crítica en la que, de acuerdo a Mizrahi (2015), donde solamente se tomaron en cuenta las proposiciones que estaban en consonancia con la tesis a defender y no las evidencias en contra.

Overfitting o sobreajuste

En términos generales, se refiere a un sesgo en el que se asume que un modelo estadístico y matemático debe corresponderse casi totalmente con su conjunto de entrenamiento o con su *dataset*. Este es un problema asociado al *machine learning*, pues, si, por ejemplo, el modelo estadístico se corresponde totalmente, el modelo falla al incorporar nueva información y realizar generalizaciones. Las principales soluciones son algorítmicas: *early stopping*, referida a la detención del entrenamiento de los datos antes de la aparición de ruidos¹⁸; incorporación de más datos en el modelo; así como “*network-reduction*”, que consiste en una disminución de la data (Xing, 2019).

Asimismo, se considera un error debido a que las personas analistas y encargadas de la visualización en la arquitectura de capas lógicas señalada en el presente artículo cometen el error de razonamiento de asumir que los resultados de un modelo que se adapta casi completamente a su *dataset* de entrenamiento se adecuan a la realidad, cuando, como se expuso en el párrafo anterior, implica lo contrario. Es un sesgo común, debido a que es intuitivo asumir que a mayor porcentaje de correspondencia existe mayor capacidad de aprendizaje en un proceso de *machine learning*.

El sesgo del superviviente

El sesgo del superviviente es un error de razonamiento en el contexto de la argumentación lógica, presente en la analítica de datos, que implica poner la atención únicamente en una serie de elementos que han superado una determinada selección, ignorando todos aquellos que no lo superaron. Este sesgo puede implicar que la persona que observa llegue a ignorar la existencia de todos esos elementos, los cuales, en muchas ocasiones, pueden contar con características significativas y relevantes (Elston, 2021).

En la historia existen dos eventos que, a pesar de las dudas en torno a su historicidad, resultan paradigmáticos (Felton, 2021): en la Primera Guerra Mundial, los generales estadounidenses se alarmaron porque llegaban muchos soldados con heridas en la cabeza, específicamente los que utilizaban el casco Brodie, modelo diseñado en acero en 1915, por lo que asumieron que los cascos no eran efectivos. Sin embargo, el caso era el contrario: los soldados sobrevivían y recibían solo una herida gracias a los Brodie. El otro caso, el de la Segunda Guerra Mundial, se refiere a los aviones de combate estadounidenses, los cuales volvían a la base militar con agujeros de bala en el fuselaje, las alas exteriores y la cola, por lo que reforzaron estas áreas. Sin embargo, la única información que recibían, tal y como señaló el matemático Abraham Wald, era de los sobrevivientes. Es decir, realmente las áreas que habían recibido las balas eran las más fuertes, por lo que no necesitaban ser reforzadas.

En el ámbito de la analítica es bastante común en los análisis financieros. Por ejemplo, cuando se estudia el desempeño de grupos de inversiones, usualmente fondos mutuos con diferentes personas, sociedades y empresas, se realiza tomando en cuenta solo a los sobrevivientes, excluyendo a los fondos y empresas que

17 Se considera que el realismo científico es una postura que asume que existe una realidad objetiva más allá de las interpretaciones humanas, por lo que esta postura se centra en las discusiones relacionadas en torno al grado y tipo de conocimiento que la ciencia puede brindar acerca del mundo (Borge, 2015).

18 En términos generales, se entienden como las distorsiones que aparecen en el proceso de aprendizaje de los datos. De acuerdo a la página especializada Data Heroes (2023), son fluctuaciones inesperadas que influyen en la identificación de patrones y relaciones.

ya no aportan o, incluso, existen, debido a sus rendimientos deficientes. Lo anterior implica que el estudio conlleva un sesgo, proporcionando una visión muy optimista sobre el rendimiento que no necesariamente se corresponde con la realidad.

Efecto Hawthorne

Elton Mayo, autor australiano y considerado como uno de los teóricos más relevantes de la historia del *management* (Martino, 2023), realizó una serie de experimentos en Hawthorne Works, una fábrica eléctrica en Chicago, para estudiar la influencia de los diferentes tipos de iluminación en la productividad. Los estudios señalaron que la productividad disminuyó tras el estudio, es decir, cuando las personas que trabajaban en la fábrica ya no estaban siendo observadas (Sujatha, Mayurnath y Pathak, 2019). Por tanto, el efecto Hawthorne se entiende como el comportamiento de las personas que son parte de un experimento y que muestran una modificación en algún aspecto de su conducta por el hecho de saber que están siendo estudiadas y observadas. Si bien existen algunas críticas o reformulaciones en torno a la existencia de este efecto (Witton y Elbourne, 2014), se puede asumir, siguiendo a Sedwick y Greenwood (2015), que existe algún tipo de influencia en el comportamiento de las personas que participan en un estudio.

En el caso del *Big data*, es fundamental que la persona analista tenga en cuenta este sesgo y, por tanto, una conciencia clara de dónde vienen sus datos y la manera en la que fueron recolectados, incluso si son producto de procesos de automatización. Esto tiene que ver con lo señalado previamente con respecto a las V del **Big data**, específicamente en torno a los adjetivos de valor y veracidad.

Error de McNamara

El sesgo del Error de McNamara se refiere a la tendencia a enfocarse en métricas y datos cuantitativos bajo una supuesta pretensión de objetividad, sin considerar otros factores cualitativos (Singh y Shah, 2023). Se suele priorizar los datos más sencillos de cuantificar, similar a la Navaja de Ockham, eligiendo el dato más simple ante un conjunto de opciones. Su nombre proviene de la práctica de McNamara durante la guerra de Vietnam, donde la principal métrica era el recuento de bajas enemigas, ignorando variables cruciales.

Esta falacia también fue notable durante la gestión inicial de la pandemia de SARS-CoV-2, donde se basaron decisiones en métricas fácilmente medibles como pruebas positivas y hospitalizaciones. En la visualización y análisis de datos, es crucial reconocer la falacia de McNamara para evitar conclusiones simplistas que pueden llevar al fracaso de estrategias en diversos ámbitos, un error vinculado al efecto cobra mencionado teóricamente.

Paradoja de Simpson

En términos generales, se entiende como una paradoja en la que una tendencia determinada que aparece en diferentes conjuntos de datos, desaparece cuando estos conjuntos son combinados y, usualmente, en su lugar surge la tendencia opuesta. Por otra parte, el fenómeno se presenta, también, cuando una relación estadística entre dos variables en una población desaparece al ser dividida en subpoblaciones (Stanford Encyclopedia of Philosophy, 2021) o, en términos más simples, cuando la relación entre dos variables se modifica cuando se tiene en cuenta el posible efecto de otra variable no contemplada hasta el momento. Es muy utilizada en epidemiología. Un ejemplo muy claro es el de comparar la efectividad de dos hospitales tomando en cuenta el número de personas fallecidas y considerar, por tanto, que el que tiene menores muertes es el más eficaz; sin embargo, al introducir variables relacionadas al tipo de patologías que se abordan en cada hospital, la conclusión puede cambiar (Molinero, 2001). En el caso costarricense, algunas reacciones a noticias relacionadas a la vacunación en torno al SARS-CoV-2 pudieron haber cometido, de manera intuitiva, el error de no tomar en cuenta los supuestos de la Paradoja de Simpson, sobre todo como respuesta a los titulares.¹⁹ Por ejemplo, aDiarioCR.com publicó una nota, el 31 de enero del 2022, titulada

19 La manera en la que se redactan los titulares ha sido estudiada por diferentes disciplinas, ya que determinan la manera en la que se leen los artículos y las conclusiones que se obtienen, ya que, además, es muy común que las personas, en el contexto de un mundo digital y de redes sociales, solo lean los titulares (Nkiru y Uchechukwu, 2020)

el 70% de los hospitalizados por Covid-19 están vacunados, ante lo que, en los comentarios²⁰, algunas personas empezaron a cuestionar la efectividad de las vacunas, esto a pesar de que en la nota se clarificaba que existían otros elementos a tomar en cuenta. Aspectos relacionados²¹ a la Paradoja de Simpson, en este caso, consisten en asumir una relación entre personas vacunadas y hospitalizaciones ignorando variables que podrían modificar las conclusiones: factores de riesgo, número de vacunas, etc.

A propósito, en este artículo, se expuso de manera muy general la paradoja de Simpson para tener una visión general del problema a la hora de trabajar con datos en el ámbito de arquitecturas de *Big data*. Sin embargo, su correcto análisis implica la cuantificación estadística de las relaciones entre las variables, así como un riguroso estudio de las relaciones causales en un fenómeno específico.

Falacia del jugador

La falacia del jugador, o falacia de Montecarlo, consiste en la idea de que un evento aleatorio en el pasado condiciona los hechos aleatorios en el presente. Por ejemplo, uno de los errores comunes es el de asumir que un suceso aleatorio tiene más probabilidad de ocurrir porque no ha sucedido durante un periodo de tiempo extenso. Esto se suele observar en la lotería, ya que muchas personas consideran que, si un número lleva mucho tiempo sin salir, probablemente salga. De hecho, el año anterior, *El Financiero* publicó una nota (Cerdas, 16 de diciembre de 2023) en la que indicaba, para el público en general, los números que nunca han salido premiados, esto para que fueran tenidos en cuenta, ya que el año anterior, en 2022, salió el 00 por primera vez.

En análisis y visualización de datos es una falacia muy común, pues, por ejemplo, se suele cometer el error de asumir que, debido a condiciones contextuales y azarosas que ocurrieron en el pasado, una tendencia a la baja en ciertos índices de rendimiento va a ser revertida porque, al igual que los números de la lotería que saldrían por llevar algún tiempo sin salir, es el momento de una tendencia a la alta. Esto implica pasividad organizacional y que no se tomen decisiones para revertir la situación.

Falacia de las manos calientes

Similar a la falacia anterior, ambas se suelen estudiar de forma conjunta (Stöckl, Huber, Kirchler y Lindner, 2015). Como su nombre lo dice, hace alusión al básquetbol, específicamente al momento en el que un jugador se encuentra en una racha de canastas. Un ejemplo reciente es el caso de Stephen Curry, jugador de la NBA y de los Golden State Warriors, quien tuvo, hasta el 18 de diciembre del 2023, una racha de 268 partidos anotando triples. Cuando se encontraba en la racha, se asumía como un hecho que iba a anotar al menos un triple durante el juego.

A nivel empresarial, se suele considerar la misma actitud: cuando hay rachas positivas a nivel de rendimiento, sustentadas por los datos, se asume que la situación va a ser la misma en el futuro. Al igual que en la falacia del jugador, esto lleva a la pasividad o a la asunción de riesgos innecesarios. La persona analista de datos debe ser consciente de esto al llevar a cabo sus investigaciones y posteriores visualizaciones.

Falacia de la falsa causalidad o causa falsa

Es una falacia muy común en todos los ámbitos, no solo en analítica, que consiste en la atribución de una causa a un evento a pesar de que no exista evidencia. En analítica, el error se presenta al confundir correlación con causalidad. A estos errores se les llama *relaciones espurias* y las personas que gestionan, analizan y visualizan datos las deben tener en cuenta.

Los ejemplos de este tipo de correlaciones utilizando datasets abundan: la tasa de divorcio en Maine y su correlación con el consumo de margarina; el número de personas que se ahogan en las piscinas y su relación con la cantidad de películas de Nicolas Cage; o la correlación estadística entre gastos en ciencia, espacio y tecnología y los suicidios por estrangulamiento (Vigen, s.f.).²²

20 Consultar en: <https://adiariocr.com/salud/70-de-los-hospitalizados-por-covid-19-están-vacunados/>

21 Se habla de "aspectos relacionados", ya que, como se explica más adelante en el artículo, la Paradoja de Simpson implica la cuantificación estadística.

22 Tyler Vigen, autor del libro *Spurious Correlations*, presenta una serie de ejemplos similares en su página web, la cual se puede consultar

CONCLUSIONES

Las sociedades contemporáneas se caracterizan por la integración de los datos en todas las esferas de la vida, lo cual se debe en parte a su valor como recurso político y económico significativo. Sin embargo, los datos en sí mismos no ofrecen información hasta que se les interpreta mediante análisis y visualización. Durante estos procesos es común cometer varios errores, como sesgos y falacias, incluidas tres falacias y una paradoja específicas, así como siete sesgos, destacando el efecto cobra y el efecto Semmelweis, que se analizaron en el marco teórico.

Por tanto, es fundamental destacar que los errores identificados no son meramente lógicos, sino que también presentan dimensiones éticas, ejemplificadas por el fenómeno del *Cherry Picking*. Adicionalmente, se observó que algunas falacias y la Paradoja de Simpson demandan un enfoque más profundo en estadística y lógica formal. Asimismo, otras falacias ampliamente reconocidas, como la apelación a la emoción, la generalización apresurada y la falacia ad hominem, no se abordaron en este análisis, focalizado en la dinámica de los datos.

Asimismo, en este trabajo se buscó llegar a aquellos profesionales dedicados al análisis y la visualización de datos dentro de estructuras de *Big data*; sin embargo, las falacias, sesgos y paradojas discutidas son extensibles a otros ámbitos y disciplinas, especialmente en las Ciencias Sociales.

Par terminar, este artículo propone explorar líneas de investigación sobre el rol de la ética en la interpretación de sesgos, falacias y paradojas, así como sobre la legitimidad epistemológica del *Big data*, enfocándose en su carácter científico derivado de los procesos analíticos y de visualización. Además, el análisis conceptual, documental y bibliográfico llevado a cabo en esta investigación podría ampliarse mediante estudios de caso, encuestas y entrevistas en el ámbito privado y público para entender mejor la prevalencia de estos errores en distintas organizaciones y desarrollar estrategias efectivas para mitigar su incidencia.

REFERENCIAS

- Atkins, Hazel, "Raising "The Rock": The Importance of T. S. Eliot's Pageant-Play". *Christianity and Literature*, 62, n.o 2 (2013): 261- 282. <https://doi.org/10.1177/014833311306200>.
- Barrantes, Rodrigo, "Investigación: un camino al conocimiento", San José, Costa Rica: EUNED, 2016.
- Borge, Bruno, "Realismo científico hoy: a 40 años de la formulación del Argumento del No-Milagro". *Revista Acta Scientiarum* 37, n.o 2 (julio, 2015): 221-233. URL: <http://www.redalyc.org/articulo.oa?id=307343306010>.
- Brooks, David, "The Philosophy of Data", *New York Times*, 4 de febrero del 2013, edición en español, <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>
- Camargo-Vega, Juan., Camargo-Ortega, Jonathan., y Joyanes-Aguilar, Luis. "Conociendo *Big data*". *Revista de la Facultad de Ingeniería*, 24, n.o 38(enero-junio 2015): 63-77, <http://www.redalyc.org/articulo.oa?id=413940775006>
- Cerdas, Mónica, "Lotería Navideña: estos son los números que nunca han salido con el premio mayor", *El Financiero*, 16 de diciembre del 2023, <https://www.elfinancierocr.com/finanzas/loteria-navidena-estos-son-los-numeros-que-nunca/WGMMVS5TWZEH5CLRIW5YQ5CZPU/story/>
- Chen, Andrew, "DAU/MAU is an important metric to measure engagement, but here's where it fails", @andrewchen, s.f, <https://andrewchen.com/dau-mau-is-an-important-metric-but-heres-where-it-fails/>
- Data Heroes, Noise in Machine Learning, *Data Heroes blog*, s.f, <https://dataheroes.ai/glossary/noise-in-machine-learning/>
- Diebold, Francis, "On the Origin(s) and Development of the Term \Big data", PIER Working Paper n.o 12-037, septiembre 2012, <https://dx.doi.org/10.2139/ssrn.2152421>

en el siguiente link: <https://www.tylervigen.com/spurious-correlations>

- Drayson, Matthew y Bashir, Amjad, “The Evolution of Data Management: A Practitioner Perspective”, <https://www.dimensiondata.com/-/media/ntt/global/solutions/intelligent-business/intelligent-business-landing-page/evolution-of-data-management-ebook.pdf>
- Dykes, Brent, “A History Lesson On The Dangers Of Letting Data Speak For Itself”, Forbes, 9 de febrero del 2016, <https://www.forbes.com/sites/brentdykes/2016/02/09/a-history-lesson-on-the-dangers-of-letting-data-speak-for-itself/?sh=7aed8c9d20e1>
- Elston, Dirk, “Survivorship Bias”, Journal of the American Academy of Dermatology, (2021), <https://doi.org/10.1016/j.jaad.2021.06.845>
- Felton, James, “How A Helmet And A Bullet-Riddled Plane Perfectly Demonstrates Survivor Bias”, IFL Science, 14 de septiembre del 2021, <https://www.iflscience.com/how-a-helmet-and-a-bulletriddled-plane-perfectly-demonstrates-survivor-bias-60930>
- García, Yamileth y Pérez, Yolanda, “Aplicación del discurso teórico en diversos espacios de la realidad social: un acercamiento sociológico entre actores sociales”, Revista Abra, 42, n.o. 65 (2022): 83-96, <https://doi.org/10.15359/abra.42-65.5>
- Gupta, Vipin., Saini, Chhavi., Oberoi, Meher., Kalra, Gagan., y Imran, Nasir, “Simmelweis Reflex: An Age-Old Prejudice”, World Neurosurgery, 136 (2020): e119-e125, <https://doi.org/10.1016/j.wneu.2019.12.012>.
- Hernández-Leal, Emilcy., Duque-Méndez, Néstor y Moreno-Cadavid, Julián, “*Big data*: una exploración de investigaciones, tecnologías y casos de aplicación”, TecnoLógicas 20, n.o 29, (2017): 17-24, http://www.scielo.org.co/scielo.php?pid=S0123-77992017000200002&script=sci_arttext
- IBM, “Understanding the architectural layers of a *Big data* solution”, IBM blog, 15 de octubre del 2013, <https://developer.ibm.com/articles/bd-archpatterns3/>
- Manterola, Carlos y Otzen, Tamara, “Los Sesgos en Investigación Clínica”, International Journal of Morphology 33, n.o. 3, (2015): 1156-1164, URL: https://www.scielo.cl/scielo.php?pid=S0717-95022015000300056&script=sci_arttext&tlng=pt.
- Márquez, Jairo, “Inteligencia artificial y *Big data* como soluciones frente a la COVID-19”, Revista de Bioética y derecho. No.50, (2020): 315-331, https://scielo.isciii.es/scielo.php?pid=S1886-58872020000300019&script=sci_arttext
- Martén, Sergio. “El problema epistemológico de los *Big data* en la producción de conocimiento científico”. Tesis de Maestría en Filosofía (2023). Universidad de Costa Rica.
- Martínez, Sergio, “La navaja de Ockham y la heterogeneidad de las representaciones: hacia una ontología de lo abstracto”, Azafea Revista de filosofía 12, n.o. (2011):97-118, <https://doi.org/10.14201/7999>.
- Martino, Silvia, “Repensar la empresa, la persona y el trabajo: Elton Mayo y algunos aportes antropológicos de Leonardo Polo”, Revista de Pensamiento y Cultura volumen 10, (2023): 173-195. DOI: <http://dx.doi.org/10.31207/colloquia.v10i0.154>
- McCambridge, Jim., Witton, John y Elbourne, Diana, “Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects”, Journal of Clinical Epidemiology 67, n.o. 3, (2014): 267-277, doi: 10.1016/j.jclinepi.2013.08.015.
- Molinero, Luis, “La paradoja de Simpson”, Asociación de la Sociedad Española de Hipertensión, octubre del 2021, <https://www.alceingenieria.net/bioestadistica/simpson.pdf>
- Mizrahi, Moti, “Historical Inductions: New Cherries, Same Old Cherry-picking”, International Studies in the Philosophy of Science, 29, n.o 2, (2015): 129-148, DOI:10.1080/02698595.2015.1119413
- Mysore, Divakar., Khupat, Shrikant., y Jain, Shweta, “Understanding the architectural layers of a *Big data* solution”, IBM blog, 14 de octubre del 2013, <https://developer.ibm.com/articles/bd-archpatterns3/>
- Nussbaumer, Cole, “Storytelling with data. A data visualization guide for business professionals”, New Jersey, Estados Unidos: John Wiley & Sons.

- Nwammuo, Angela & Nwafor, Gideon, “How Online Newspaper Headlines Sway Opinion: Analysis of Online Newspaper Readership Patterns among Facebook Users”, *International Journal of Informatics, Technology & Computers* 5, n.o. 1, (2020): 1 – 10.
- Pastor, Servando, “Alquimia: Cómo los datos se están transformando en oro”, *Perfiles Económicos*, n.o. 10, (2020): 173-177, <http://doi.org/10.22370/pe.2020.10.2663>.
- Parusniková, Zuzana, “Popperian methodology and the Semmelweis case”, *Medical Health Care and Philosophy* 26, n.o. 4, (2023): 529–537. <https://doi.org/10.1007/s11019-023-10167-7>
- Redacción, “70% de los hospitalizados por Covid-19 están vacunados”, *aDiarioCR*, 31 de enero del 2022, <https://adiariocr.com/salud/70-de-los-hospitalizados-por-covid-19-estan-vacunados/>
- Sánchez, Óscar, “El *Big data* y su aplicación práctica en la estrategia de la empresa: Social Media Analytics”, Trabajo Final de Grado en Administración y Dirección de Empresas, Universidad Pontificia Comillas, 2019.
- Scholl, Raphael, “Causal inference, mechanisms, and the Semmelweis case. *Studies in History and Philosophy of Science*”, 44, n.o. 1 (2013): 66–76. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0039368112000350>
- Sedwick, Philip y Greenwood, Nan, “Understanding the Hawthorne effect”, *The BMJ: Endgames*, (2015), <https://doi.org/10.1136/bmj.h4672>.
- Singh y Shah, “Competency-based medical education and the McNamara fallacy: Assessing the important or making the assessed important?” *Postgrad Med* 69, n.o. 1 (2023):35-40. Doi: 10.4103/jpgm.jpgm_337_22.
- Solano, Jonathan y Leiva, Estefany, “Big data Analytics: propuesta de una arquitectura”, 2014, San José, Costa Rica: ULACIT.
- Soussi, Nisrrine y El Arass, Mohammed, “Data Cycle: From *Big data* to Smart Data”, Quinto congreso Internacional de IEEE, Marruecos: octubre del 2018.
- Stanford Encyclopedia of Philosophy, Simpson’s Paradox, 21 de marzo del 2023, <<https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/>>.
- Stöckl, Thomas., Jürgen, Huber., Kirchler, Michael y Lindner, Florian, “Hot hand and gambler’s fallacy in teams: Evidence from investment experiments”, *Journal of Economic Behavior & Organization*, 117, (septiembre del 2015): 327-339, <https://doi.org/10.1016/j.jebo.2015.07.004>
- Sujatha., Mayurnath, Reddy y Pooja, Pathak, “Camouflage in research – the hawthorne effect”. *International Journal of Development Research*, 09, n.o. 04, (2019): p. 26996-26999. ISSN: 2230-9926.URL: <https://www.journalijdr.com/camouflage-research-%E2%80%93-hawthorne-effect>
- Tapia, Jeimy, “Macrodatos: almacenamiento y consulta de datos aplicados en las organizaciones”, *Revista Ciencia administrativa*, 2, 2022, ISSN 1870-9427.
- Trianarts, “T. S. Eliot: El primer coro de la roca”, *Trianarts blog*, 12 de mayo del 2022, <https://trianarts.com/acerca-de/#sthash.pG6DBi6R.dpbs>
- Valero, José, “La visualización de datos”, *Revista Ámbitos*, núm. 25, julio-diciembre 2014, <https://www.redalyc.org/pdf/168/16832256009.pdf>
- Ventura, Dalia, “Qué es el “efecto cobra” (y cómo demuestra que a veces es peor el remedio que la enfermedad)”, *BBC News Mundo*, 21 de noviembre del 2020, <https://www.bbc.com/mundo/noticias-54935306>
- Vigen, Tyler, “Spurious correlations”, s.f., <https://www.tylervigen.com/spurious-correlations>
- Ying, Sue, “An Overview of Overfitting and its Solutions. *Journal of Physics*”, Conference Series 1168, n.o 2, 2019, doi:10.1088/1742-6596/1168/2/022022