




Attribution ShareAlike (CC BY-SA) 

Glosario

1. **Accuracy:** Es la proporción del número de predicciones correctas sobre las predicciones totales. Al multiplicar tales fracciones por 100, podemos obtener un puntaje de precisión porcentual. (Skiena, 2017)
2. **Agent based modeling:** Un sistema se modela como una colección de entidades autónomas de toma de decisiones llamadas agentes. Cada agente evalúa individualmente su situación y toma decisiones sobre la base de un conjunto de reglas. Los agentes pueden ejecutar varios comportamientos apropiados para el sistema que representan, por ejemplo, producir, consumir o vender. Las interacciones competitivas repetitivas entre agentes son una característica del modelado basado en agentes, que se basa en el poder de las computadoras para explorar dinámicas fuera del alcance de los métodos matemáticos puros. En simple, consiste en un sistema de agentes y la relación entre ellos. (Bonabeau, 2002)
3. **Algoritmo:** Conjunto de instrucciones sobre cómo una computadora debe realizar una tarea en particular. Son utilizados por muchas organizaciones para tomar decisiones y asignar recursos basados en grandes conjuntos de datos. Se comparan con las recetas, que toman un conjunto específico de ingredientes y los transforman a través de una serie de pasos explicables en un resultado predecible. (Caplan et al., 2018)
4. **Análisis bayesiano:** La estadística bayesiana se basa en la probabilidad subjetiva, trabaja con la actualización de la evidencia considerando los conocimientos adquiridos previos a una investigación, más la evidencia obtenida con esta. La interpretación de los resultados requiere la

especificación de las hipótesis por contrastar y su probabilidad a priori antes del estudio. La evidencia del estudio se mide con el factor Bayes (razón de la compatibilidad de los datos bajo las hipótesis propuestas). La conjunción de las probabilidades a priori de las hipótesis con el factor Bayes permite calcular la probabilidad a posteriori de cada una. La hipótesis con mayor grado de certidumbre en su actualización es la aceptada para la toma de la decisión. En esta revisión se muestran tres ejemplos de hipótesis por contrastar: diferencia de promedios, correlación y asociación. (Rendón-Macías et al. 2018)

5. **Analítica predictiva:** El uso de estadísticas para hacer predicciones sobre ciertos eventos o resultados. El aprendizaje automático es una forma de análisis predictivo, pero el análisis predictivo también puede incluir técnicas analíticas que no se basan en predicciones de aprendizaje de datos de entrenamiento. La "analítica predictiva" y el "aprendizaje automático" a menudo se combinan y se usan indistintamente. (Peña-López, 2018)
6. **API (application programming interface):** Es la interfaz de usuario de una biblioteca de funcionalidades para el programador que la usa (Stylos et al., 2009). Un ejemplo de una API, es la que disponibiliza Google Maps para que los desarrolladores puedan calcular distancias, tiempos de viaje o rutas entre 2 puntos, entre otras cosas.
7. **Árbol de Decisión:** Es una de las técnicas más simples e intuitivas en el aprendizaje automático, basada en el paradigma de divide and conquer. La idea básica detrás de los árboles de decisión es dividir el espacio en parches y ajustar un modelo a un parche. (Igual y Seguí, 2017)
8. **Base de datos:** Una pieza de software que almacena datos de un tipo particular en un formato que admite el acceso de baja latencia. Es un marco para almacenar y acceder de manera eficiente a los datos. Una base de datos prototípica es un servidor fornido, que contiene más datos de los que cabría en una computadora normal, los almacena de manera que se pueda

acceder rápidamente (esto generalmente implica una tonelada de optimizaciones ocultas para que el usuario la pueda usar más fácilmente), y está preparado para recibir solicitudes de otras computadoras para acceder o modificar los datos. (Cady, 2017)

9. **Big Data:** Se refiere a conjuntos de datos gigantescos (volumen); más diversificados, incluyendo datos estructurados, semiestructurados y no estructurados (variedad), y llegando de manera inmediata (velocidad). Big Data Analytics se refiere al proceso de recopilación, organización y análisis de grandes conjuntos de datos para descubrir diferentes patrones y otra información útil. (Russom, 2011)
10. **Ciencia de Datos:** Se define comúnmente como una metodología mediante la cual se pueden inferir ideas procesables a partir de los datos. La realización de la ciencia de datos es una tarea con un objetivo ambicioso: la producción de creencias informadas por datos y para ser utilizadas como base para la toma de decisiones. (Iguar y Seguí, 2017)
11. **Clustering:** Agrupación de objetos (por ejemplo, instancias en un conjunto de datos) que son similares entre sí y (relativamente) diferentes a los objetos que pertenecen a otros grupos. (Bramer, 2016)
12. **Conocimiento:** Respuesta apropiada y persistente a una entrada dada. La idoneidad está determinada por las interacciones externas, que generalmente mantienen la viabilidad del respondedor. Para un sistema basado en modelos: resultados acordado/aceptados de aportes dados a un conjunto de creencias generalizado, a largo plazo y consistente. La base del modelo es estrictamente un conjunto de creencias, pero se aproxima a la verdad/conocimiento si el uso/aceptación es, por ejemplo, exitoso y generalizado. (Sanders, 2016)
13. **Correlación:** Decimos que x e y están correlacionados cuando el valor de x tiene algún poder predictivo sobre el valor de y . El coeficiente de

correlación $r(X, Y)$ es una estadística que mide el grado en que Y es una función de X, y viceversa. El valor del coeficiente de correlación varía de -1 a 1, donde 1 significa totalmente correlacionado y 0 no implica ninguna relación o variables independientes. Las correlaciones negativas implican que las variables están anti-correlacionadas, lo que significa que cuando X sube, Y baja. (Skiena, 2017)

14. **CSV (comma separated value):** Es un formato simple y popular para intercambiar datos entre programas. Cada línea representa un solo registro, con campos separados por comas. La mejor prueba de si su archivo csv está formateado correctamente es si Microsoft Excel o algún otro programa de hoja de cálculo puede leerlo sin problemas. (Skiena, 2017)
15. **Data Leakage:** Se puede definir como un evento en el que la información clasificada como datos sensibles, protegidos o confidenciales han sido vistos, robados o utilizados por alguien que no está autorizado para hacerlo. La fuga de datos causa problemas serios y costosos a las empresas y organizaciones, porque el número de eventos continúa aumentando. (Skrop, 2015)
16. **Data Warehouse:** Es una base de datos para el procesamiento analítico en línea (OLAP) para ayudar a la toma de decisiones. Está diseñado de acuerdo con un modelo dimensional que tiene como objetivo observar hechos a través de medidas, también llamados indicadores, de acuerdo con las dimensiones que representan los ejes de análisis. (Dehdouh et. al, 2015)
17. **Datos:** Una representación espacial organizada a partir de un flujo de información. En general, la organización proviene de la mente y la representación se construye simbólicamente. Cualquier propósito se define mediante procesamiento o decodificación. (Sanders, 2016)

- 18. Datos Personales:** Cualquier información vinculada o referida a una persona natural identificada o identificable. Se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, mediante información combinada con otros datos, en particular mediante un identificador, tales como el número de cédula de identidad, el análisis de elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona, excluyendo aquellos casos en que el esfuerzo de identificación sea desproporcionado. (Ley 10.628)
- 19. Deep Learning:** Una clase de algoritmos de aprendizaje automático que realizan sus tareas construyendo abstracciones de los datos de entrada que analizan. Los métodos de deep learning más populares utilizan "redes neuronales", arquitecturas de algoritmos de aprendizaje automático que aprenden potencialmente cientos de millones de parámetros. Los modelos creados a través del deep learning son notoriamente difíciles de entender por un humano. (Peña-López, 2018)
- 20. Detección de Anomalías:** Es el proceso de encontrar los patrones en un conjunto de datos cuyo comportamiento no es normal dentro de lo esperado. Estos comportamientos inesperados también se denominan anomalías o valores atípicos (outliers). Las anomalías no siempre se pueden clasificar como un ataque, pero puede ser un comportamiento sorprendente que anteriormente no se conocía. Puede o no ser dañino. La detección de anomalías proporciona información muy significativa y crítica en diversas aplicaciones, por ejemplo, robos de tarjetas de crédito o robos de identidad. (Agrawal y Agrawal, 2015)
- 21. Economía del comportamiento (Behavioral economics):** Explora cómo afectan las decisiones económicas de las personas y las consecuencias de esas decisiones para los precios de mercado, los rendimientos y la asignación de recursos. La investigación económica tradicional supone que las decisiones económicas de las personas se basan en la regla de maximizar la utilidad. La economía del comportamiento, por el contrario,

no supone que las personas sean buenas en la maximización de la utilidad ni que sea el único objetivo de las personas. Utilizando herramientas empíricas, los economistas del comportamiento han demostrado más bien que las personas tienen prejuicios psicológicos, recursos cognitivos limitados y se preocupan por otros valores como la equidad, todo lo cual podría socavar su comportamiento de maximización de la utilidad. (Miller, Amit y Posten, 2015)

22.F-score: Una medida del rendimiento del clasificador. Es la media armónica de la precisión y el recall. (Cady, 2017)

23.Inferencia Causal: Su objetivo no es sólo inferir creencias o probabilidades bajo condiciones estáticas, sino también la dinámica de creencias bajo condiciones cambiantes, por ejemplo, cambios inducidos por tratamientos o intervenciones externas. Para establecer causalidad, se requiere cierto conocimiento del proceso de generación de datos; no pueden calcularse únicamente de los datos, ni de las distribuciones que los gobiernan. (Pearl, 2009)

24.Información: Es un flujo (generalmente dependiente del tiempo, por ejemplo, de energía), que puede modificarse por el efecto de las interacciones y los cambios que ocurren a lo largo de cada canal/ruta tomada a través del espacio local. La información puede considerarse en términos de señales. La señal codificada puede procesarse luego para producir una secuencia de datos y reprocesarse para algún propósito subyacente. Información como esta se utiliza para las comunicaciones entre transmisores y receptores. (Sanders, 2016)

25.Inteligencia Artificial (también AI, IA): Es un término general para una gama de tecnologías basadas en algoritmos que a menudo intentan imitar el pensamiento humano para resolver tareas complejas. (The Alan Turing Institute)

26. IoT (internet of things): En un mundo donde miles de millones de objetos pueden detectar, comunicarse y compartir información, todos interconectados a través de redes de protocolo de Internet (IP) públicas o privadas, los objetos interconectados tienen datos recopilados, analizados y utilizados regularmente para iniciar acciones, proporcionando una gran cantidad de inteligencia para la planificación, la gestión y la toma de decisiones. Es una red de objetos físicos. Internet no es solo una red de computadoras, sino que se ha convertido en una red de dispositivos de todo tipo y tamaño, vehículos, teléfonos inteligentes, electrodomésticos, juguetes, cámaras, instrumentos médicos y sistemas industriales, animales, personas, edificios, todo conectado, todo comunicando y compartiendo información basada en protocolos estipulados para lograr reorganizaciones inteligentes, posicionamiento, rastreo, seguridad y control e incluso monitoreo personal en línea en tiempo real, actualización en línea, control de procesos y administración. (Patel y Patel, 2016)

27. JSON (JavaScript Object Notation): Es un formato para transmitir objetos de datos entre programas. Es una forma natural de comunicar el estado de las variables / estructuras de datos de un sistema a otro. Esta representación es básicamente una lista de pares de atributos-valores correspondientes a nombres de variables / campos y los valores asociados. (Skiena, 2017)

28. K-means: Es un algoritmo de partición dura con el objetivo de asignar cada punto de datos a un solo clúster. El algoritmo K-means divide un conjunto de n muestras X en k grupos disjuntos c_i , $i = 1, \dots, k$, cada uno descrito por la media μ_i de las muestras en el grupo. Los medios se denominan comúnmente centroides de agrupamiento. El algoritmo K-means supone que todos los grupos k tienen la misma varianza. (Igual y Seguí, 2017)

29. K-Nearest Neighbors: Al representar puntos en un espacio n -dimensional podemos calcular la distancia geométrica entre cualquier par de puntos. Cuanto más cerca estén los puntos en el espacio de instancias, mayor será

su similitud mutua. El ejemplo de entrenamiento con la menor distancia de x en el espacio de instancia es, el vecino más cercano de x . En dominios ruidosos, no se puede confiar en el testimonio del vecino más cercano. Un enfoque más sólido identifica no uno, sino varios vecinos más cercanos, y luego les permite votar. Esta es la esencia del denominado clasificador k -NN, donde k es el número de vecinos votantes, generalmente un parámetro especificado por el usuario. (Kubat, 2017)

30.Licencia Social: Describe cómo las expectativas de la sociedad con respecto a algunas actividades pueden ir más allá del cumplimiento de los requisitos de la regulación formal; aquellos que no cumplan las condiciones para la licencia social (incluso si cumplen formalmente) pueden experimentar desafíos y cuestionamientos continuos. Son las expectativas de la sociedad con respecto al conducta y actividades de corporaciones que van más allá de los requisitos de la regulación formal. (Carter, Laurie y Dixon-Woods, 2015)

31.Machine Learning: El aprendizaje automático o machine learning es una forma de IA que permite que un sistema aprenda de datos en lugar de a través de programación explícita. Usa una variedad de algoritmos que iterativamente aprenden de los datos para mejorar, describir datos y predecir resultados. (Hurwitz y Kirsrh, 2018)

32.Margen de error (estadística): El "reverso" de la precisión predictiva (accuracy) de un clasificador. Una precisión predictiva de 0.8 (es decir, 80%) implica un margen de error de 0.2 (es decir, 20%) (Bramer, 2016)

33.Minería de Datos (o data mining): Es el proceso de explorar y analizar grandes cantidades de datos para descubrir patrones en esos datos. Los algoritmos se usan para encontrar relaciones y patrones en los datos, y luego esta información sobre los patrones se usa para hacer pronósticos y predicciones. La minería de datos se utiliza para resolver una variedad de problemas comerciales, como la detección de fraudes, el análisis de la cesta

de la compra y el análisis de abandono de clientes. (Hurwitz y Kishrsh, 2018, p.12)

34. **Minería de texto (Text Mining):** Debido a que los datos de texto a menudo se representan en el modelo de bolsa de palabras, la extracción frecuente de patrones puede ayudar a identificar términos y palabras clave concurrentes. Tales términos concurrentes tienen numerosas aplicaciones de minería de texto. (Aggarwal, 2015)
35. **MySQL:** La familia SQL es una gran clase de bases de datos relacionales, que tienen interfaces casi idénticas. De estos, MySQL es la versión de código abierto más popular y la que es más probable que encuentre en la ciencia de datos. (Cady, 2017)
36. **Naive Bayes:** Imagine que dos eventos A y B son independientes si $p(A \text{ y } B) = p(A) \cdot p(B)$. Si A es el evento de que "mi equipo deportivo favorito gana hoy" y B es "el mercado de valores sube hoy", entonces presumiblemente A y B son independientes. Pero esto no es cierto en general. Considere el caso si A es el evento de que "obtengo una A en Data Science este semestre" y B es "obtengo una A en un curso diferente este semestre". Hay dependencias entre estos eventos: el entusiasmo renovado por estudiar o beber afectará el rendimiento del curso de manera correlacionada. Si todo fuera independiente, el mundo de la probabilidad sería un lugar mucho más simple. El algoritmo de clasificación de Naive Bayes asume la independencia, para evitar la necesidad de calcular estas probabilidades condicionales desordenadas. (Skiena, 2017)
37. **Opacidad algorítmica:** Es la falta de visibilidad de los procesos computacionales tal que los humanos no son capaces de inspeccionar el funcionamiento interno para cerciorarse como los resultados y conclusiones fueron computadas. (Paudyal, 2018)

- 38.**Outlier:** Los valores atípicos son muestras de datos con un valor que está lejos de la tendencia central. (Igal y Seguí, 2017)
- 39.**Parámetros:** Son constantes que caracterizan una característica o una distribución probabilística o función de regresión de una población o valor desconocido que describe una relación poblacional. (Wooldridge, 2010 y Stock y Watson 2012)
- 40.**PostgreSQL:** Es un potente sistema administrador de base de datos relacional de objetos de código abierto con más de 30 años de desarrollo activo que le ha valido una sólida reputación de fiabilidad, solidez de características y rendimiento. (postgresql.org)
- 41.**Precisión y Recall:** La precisión es el porcentaje de verdaderos positivos, entre todos los ejemplos que el clasificador ha etiquetado como positivos, es la probabilidad de que el clasificador sea correcto al etiquetar un ejemplo como positivo. Recall es la probabilidad de que un ejemplo positivo sea reconocido correctamente como tal (por el clasificador). Por lo tanto, el valor se obtiene dividiendo el número de positivos verdaderos, por el número de positivos en el conjunto dado. Mientras que la precisión es la frecuencia de los verdaderos positivos entre todos los ejemplos considerados positivos por el clasificador, el recall es la frecuencia de los mismos verdaderos positivos entre todos los ejemplos positivos del conjunto. (Kubat, 2017)
- 42.**Privacidad:** Es un derecho reconocido en muchas constituciones nacionales y consiste en la protección de los datos personales. Muchos de los principios mundialmente reconocidos están establecidos en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea que entró en vigor en el año 2018. (Buenadicha et al., 2019)
- 43.**Procesamiento de Lenguaje Natural:** Es el análisis del texto y la lingüística computacional. (Bramer, 2016)

- 44.**Python:** Es un lenguaje de programación maduro pero también tiene excelentes propiedades para programadores novatos, por lo que es ideal para personas que nunca antes han programado. Python es un lenguaje interpretado, por lo que el código se ejecuta inmediatamente en la consola de Python sin necesidad del paso de compilación al lenguaje máquina. (Bramer, 2016)
- 45.**R:** Es un sistema para análisis estadísticos y gráficos creado por Ross Ihaka y Robert Gentleman. R es tanto un software como un lenguaje considerado como un dialecto del lenguaje S creado por los Laboratorios AT&T Bell. R se distribuye libremente bajo los términos de la Licencia Pública General de GNU. Tiene muchas funciones para análisis estadísticos y gráficos; estos últimos son visualizado inmediatamente en su propia ventana y se puede guardar en varios formatos. Los resultados de un análisis estadístico son mostrados en la pantalla, algunos resultados intermedios (valores P, coeficientes de regresión, residuos, ...) pueden guardarse, escribirse en un archivo o usarse en posteriores análisis. El lenguaje R permite al usuario, por ejemplo, programar bucles para analizar sucesivamente varios conjuntos de datos. También es posible combinar en un solo programa diferentes funciones estadísticas para realizar análisis más complejos. (Paradis, 2005)
- 46.**Random Forest:** Pueden verse como una generalización del método básico de bagging, como se aplica a los árboles de decisión. Los random forest se definen como un conjunto de árboles de decisión, en los que la aleatoriedad se ha insertado explícitamente en el proceso de construcción del modelo de cada árbol de decisión. (Aggarwal, 2015)
- 47.**Redes Neuronales:** Están diseñados para emular cómo funciona el cerebro humano para que las computadoras puedan ser entrenadas para lidiar con abstracciones y problemas que están mal definidos. Constan de tres o más capas: una capa de entrada, una o varias capas ocultas y una capa de salida.

Los datos se ingieren a través de la capa de entrada. Luego, los datos se modifican en la capa oculta y las capas de salida en función de los pesos aplicados a estos nodos. La red neuronal típica puede consistir en miles o incluso millones de nodos de procesamiento simples que están densamente interconectados. El término deep learning se usa cuando hay múltiples capas ocultas dentro de una red neuronal. Usando un enfoque iterativo, una red neuronal se ajusta continuamente y hace inferencias hasta que se alcanza un punto de detención específico. Las redes neuronales se utilizan a menudo para el reconocimiento de imágenes y aplicaciones de visión por computadora. (Hurwitz y Kirsch, 2018)

- 48.**Regresión Lineal:** Modelo lineal en sus parámetros, donde la variable dependiente es una función de las variables independientes más un término de error. Es útil para un análisis ceteris paribus debido a que permite controlar de manera explícita muchos factores que afectan en forma simultánea a la variable dependiente. (Wooldridge, 2010)
- 49.**Regresión Logística:** Modelo de regresión no lineal para una variable dependiente binaria en la que la función de regresión poblacional se modela utilizando la función de distribución logística acumulada. (Stock y Watson, 2012)
- 50.**Responsabilidad Algorítmica:** Se refiere a la asignación de responsabilidad de cómo se crea un algoritmo y su impacto en la sociedad; si se produce un daño. Los sistemas responsables incluyen un mecanismo de reparación. (Caplan, 2018)
- 51.**Sesgo (estadística):** Posible consecuencia de negar a determinados miembros de la población la oportunidad de ser seleccionados para la muestra. Como resultado, la muestra puede no ser representativa de la población. (Lind, Marchall y Wathen, 2005)

- 52.**SQL (Structured Query Language):** Es la familia de bases de datos más grande que admite el modelo de base de datos relacional (RDB). Por lo general, son extremadamente rápidas y admiten un procesamiento muy extenso de los datos, pero a cambio de este poder, son extremadamente rígidos en cuanto al tipo y los formatos de los datos que puede incluir en ellos. (Cady, 2017)
- 53.**Support Vector Machine:** Es una técnica de aprendizaje diseñada inicialmente para ajustarse a un límite lineal entre las muestras de un problema binario, asegurando la máxima robustez en términos de tolerancia a la incertidumbre isotrópica. (Igal y Seguí, 2017)
- 54.**Transformación Digital:** La Estrategia de Transformación Digital es parte fundamental de la de Modernización del Estado, ya que busca materializar una serie de principios asociados al cambio tecnológico que está sucediendo a nivel global y que los expertos han llamado la cuarta revolución industrial. Actualmente, el Estado no está preparado estructuralmente para estos cambios, dado que aún existen problemas de integración y homologación, falta de procesos estandarizados y la persistente necesidad de coordinación entre una gran cantidad de instituciones con diversidad de realidades internas. Por eso, lo que buscamos con esta estrategia es romper este paradigma de compartimentos estancos y generar un Estado Digital abierto, eficiente y amigable. (<https://digital.gob.cl/plan/estrategia>)
- 55.**XML (Extensible Markup Language):** Lenguaje que permite crear etiquetas para organizar e intercambiar contenidos más eficientemente. Corresponde a una versión simplificada de SGML (Standard Generalized Markup Language). (Decreto 14 de 2014, Ministerio de Economía, Fomento y Turismo)

Referencias

1. Skiena, S. S. (2017). The data science design manual. Springer.
2. Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl 3), 7280-7287.
3. Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (2018). Algorithmic accountability: A primer. *Data & Society*, 18.
4. Rendón-Macías, M. E., Riojas-Garza, A., Contreras-Estrada, D., & Martínez-Ezquerro, J. D. (2018). Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso. *Revista Alergia México*, 65(3), 285-298.
5. Peña-López, I. (2018). Algorithmic Accountability Policy Toolkit.
6. Stylos, J., Faulring, A., Yang, Z., & Myers, B. A. (2009, September). Improving API documentation using API usage information. In *2009 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 119-126). IEEE.
7. Igual, L., & Seguí, S. (2017). Introduction to Data Science. In *Introduction to Data Science* (pp. 1-4). Springer, Cham.
8. Cady, F. (2017). The data science handbook. John Wiley & Sons.
9. Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, 19(4), 1-34.
10. Igual, L., & Seguí, S. (2017). Introduction to Data Science. In *Introduction to Data Science* (pp. 1-4). Springer, Cham.
11. Bramer, M. (2016). Data for data mining. In *Principles of data mining* (pp. 9-19). Springer, London.
12. Sanders, J. (2016, July). Defining terms: Data, information and knowledge. In *2016 SAI Computing Conference (SAI)* (pp. 223-228). IEEE.
13. Skiena, S. S. (2017). The data science design manual. Springer.

14. Skiena, S. S. (2017). *The data science design manual*. Springer.
15. Skrop, A. (2015). DATALEAK: Data Leakage Detection System. *MACRo* 2015, 1(1), 113-124.
16. Dehdouh, K., Bentayeb, F., Boussaid, O., & Kabachi, N. (2015). Using the column oriented NoSQL model for implementing big data warehouses. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)* (p. 469). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)
17. Sanders, J. (2016, July). Defining terms: Data, information and knowledge. In *2016 SAI Computing Conference (SAI)* (pp. 223-228). IEEE.
18. PL de reforma a la Ley 19.628 sobre protección a la vida privada.
19. Peña-López, I. (2018). *Algorithmic Accountability Policy Toolkit*.
20. Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques.
21. *Procedia Computer Science*, 60, 708-713.
22. Miller, J. E., Amit, E., & Posten, A. C. (2015). Behavioral economics. *Encyclopedia of global bioethics*, 1-6.
23. Cady, F. (2017). *The data science handbook*. John Wiley & Sons.
24. Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, 3, 96-146.
25. Sanders, J. (2016, July). Defining terms: Data, information and knowledge. In *2016 SAI Computing Conference (SAI)* (pp. 223-228). IEEE.
26. The Alan Turing Institute. (2020) "Explaining decisions made with AI, part I: The basics of explaining AI".
27. Patel, K. K., & Patel, S. M. (2016). Internet of things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges. *International journal of engineering science and computing*, 6(5).
28. Skiena, S. S. (2017). *The data science design manual*. Springer.
29. Igual, L., & Seguí, S. (2017). Introduction to Data Science. In *Introduction to Data Science* (pp. 1-4). Springer, Cham.

30. Kubat, M. (2015). Similarities: Nearest-Neighbor Classifiers. In *An Introduction to Machine Learning* (pp. 43-64). Springer, Cham.
31. Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care. data ran into trouble. *Journal of medical ethics*, 41(5), 404-409.
32. Hurwitz, J., & Kirsch, D. (2018). *Machine learning for dummies*. IBM Limited Edition, 75.
33. Bramer, M. (2016). Data for data mining. In *Principles of data mining* (pp. 9-19). Springer, London.
34. Hurwitz, J., & Kirsch, D. (2018). *Machine learning for dummies*. IBM Limited Edition, 75.
35. Aggarwal, C. C. (2015). *Data mining: the textbook*. Springer.
36. Cady, F. (2017). *The data science handbook*. John Wiley & Sons.
37. Skiena, S. S. (2017). *The data science design manual*. Springer.
38. Paudyal, P., & William Wong, B. L. (2018, September). Algorithmic opacity: Making algorithmic processes transparent through abstraction hierarchy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 192-196). Sage CA: Los Angeles, CA: SAGE Publications.
39. Igual, L., & Seguí, S. (2017). Introduction to Data Science. In *Introduction to Data Science* (pp. 1-4). Springer, Cham.
40. Wooldridge, J. M. (2010). *Introducción a la Econometría*. 4ta. Edición. Cengage Learning. / Stock,
41. J. H., Watson, M. W., & Larrión, R. S. (2012). *Introducción a la Econometría* (No. 330.1543 S8). Pearson.
42. PostgreSQL.org: <https://www.postgresql.org/>
43. Kubat, M. (2015). Similarities: Nearest-Neighbor Classifiers. In *An Introduction to Machine Learning* (pp. 43-64). Springer, Cham.
44. Buenadicha, C., Galdon, G., Hermosilla, M. P., Loewe, D., & Pombo, C. (2019). *La Gestión Ética de los Datos. Por qué importa y cómo hacer un uso justo de los datos en un mundo digital*. BID, editor.
45. Bramer, M. (2016). Data for data mining. In *Principles of data mining* (pp. 9-19). Springer, London.

46. Bramer, M. (2016). Data for data mining. In Principles of data mining (pp. 9-19). Springer, London.
47. Paradis, E. (2002). R for Beginners.
48. Aggarwal, C. C. (2015). Data mining: the textbook. Springer.
49. Hurwitz, J., & Kirsch, D. (2018). Machine learning for dummies. IBM Limited Edition, 75.
50. Wooldridge, J. M. (2010). Introducción a la Econometría. 4ta. Edición. Cengage Learning.
51. Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). Introducción a la Econometría (No. 330.1543 S8). Pearson.
52. Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (2018). Algorithmic accountability: A primer. Data & Society, 18.
53. Lind, D. A., Marchal, W. G., & Wathen, S. A. (2005). Estadística aplicada a los negocios y la economía. McGraw-Hill,.
54. Cady, F. (2017). The data science handbook. John Wiley & Sons.
55. Igual, L., & Seguí, S. (2017). Introduction to Data Science. In Introduction to Data Science (pp. 1-4). Springer, Cham.
56. Página web Gobierno Digital: Estrategia de Transformación Digital (<https://digital.gob.cl/plan/estrategia>)
57. Decreto 14 (27/02/2014), Modifica decreto N° 181, de 2002, que aprueba reglamento de la ley 19.799 sobre documentos electrónicos, firma electrónica y la certificación de dicha firma, y deroga los decretos que indica. Ministerio de Economía, Fomento y Turismo; Subsecretaría de Economía y Empresas de Menor Tamaño.