

1. Bases cognitivas

1.1 Datos y la pirámide DIKW

Los datos son la materia prima del razonamiento y de la toma de decisiones y, dado su gran volumen y diversidad, son la característica fundamental de esta era. Sin embargo, y a pesar de que existe un claro consenso al respecto, no existe una definición absolutamente aceptada de lo que es un dato ni de sus productos derivados. Una excelente discusión al respecto se presenta en [1].

Una definición aceptable de datos, en el contexto que nos interesa, es:

Los datos son un conjunto de símbolos o signos que representan estímulos del medio ambiente o señales, y que son registrados a partir de la observación.

A partir de los datos se generan, sucesivamente, constructos de mayor abstracción: la información, el conocimiento y la sabiduría. La relación entre los datos y estos otros conceptos suele describirse a través de la llamada *pirámide de datos*, también conocida como *pirámide de la información*, *del conocimiento*, *de la sabiduría* o, simplemente, *pirámide DIKW* (por las siglas en inglés *data*, *information*, *knowledge*, *wisdom*). En la figura 1.1 se muestra la un diagrama de la pirámide DIKW con un ejemplo simple:

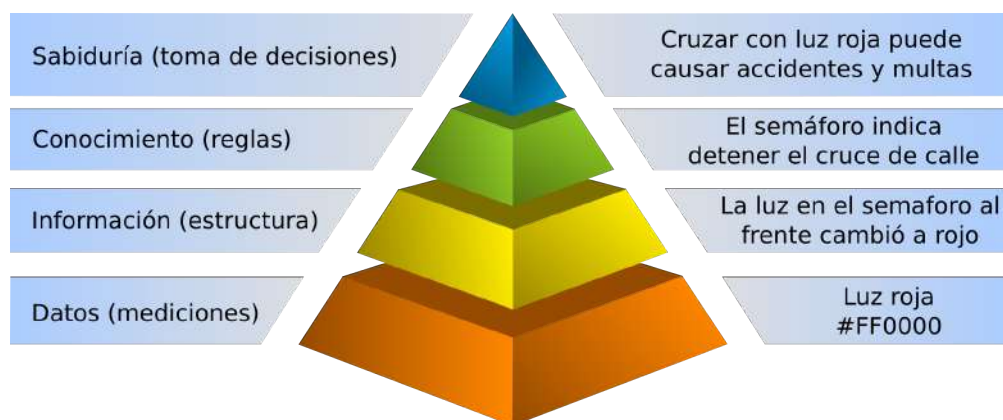


Figura 1.1: Pirámide DIKW, con un ejemplo simple.

De manera similar, podemos definir el término *información* como un conjunto de datos organizados para un propósito específico y presentados en un contexto que les da significado y relevancia para el usuario. El conocimiento es el conjunto de reglas, en un contexto dado, que resultan de la interpretación y apropiación de la información por parte del usuario. La sabiduría, por su parte, es la capacidad de tomar decisiones adecuadas utilizando el conocimiento, la experiencia, la comprensión y la percepción. El fin de la ciencia de datos es transformar datos en acciones pertinentes a través de la generación de información, conocimiento y sabiduría.

1.2 Ciencia de datos y otras disciplinas

La **ciencia de datos** es un nuevo campo interdisciplinario cuyo objetivo es analizar y desarrollar explicaciones de fenómenos complejos a partir del análisis de los datos generados por eventos particulares.

La ciencia de datos se relaciona con otras disciplinas para el tratamiento y comprensión de datos, entre las que destacan las siguientes:

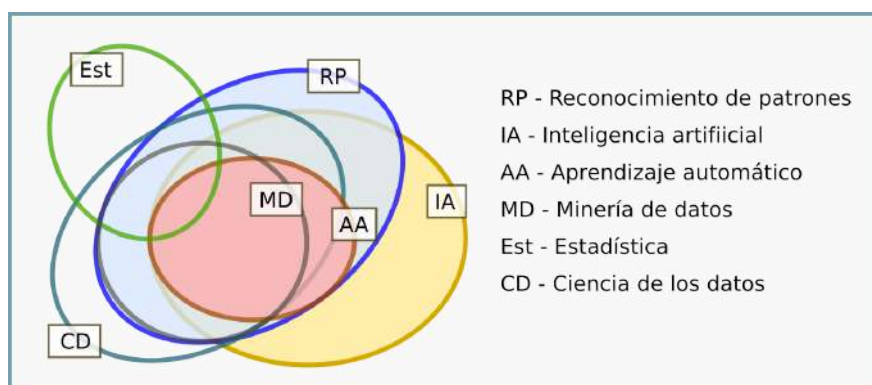


Figura 1.2: Bosquejo de la relación entre el reconocimiento de patrones y otras disciplinas relacionadas.

- La **estadística**, una rama de las matemáticas que se ocupa de la recopilación, organización, análisis, interpretación y presentación de datos para una mejor comprensión de la naturaleza. Como tal, su agenda principal consiste en inferir, a partir de ese estudio, reglas generales o conceptos que pueden aplicarse a observaciones nuevas o desconocidas.
- La **inteligencia artificial**, una disciplina que tiene el objeto de desarrollar métodos de solución de problemas que, por sus mecanismos complejos de deducción, suelen compararse con los mecanismos utilizados por los seres vivos, particularmente el ser humano.
- El **reconocimiento de patrones**, una de las disciplinas de las ciencias de la computación, relacionada con el análisis inteligente de datos. Su objetivo es la identificación de patrones o regularidades en los datos disponibles en un determinado contexto. El reconocimiento de patrones utiliza elementos de estadística, gramáticas formales y metáforas con la naturaleza para desarrollar sus métodos propios.
- El **aprendizaje automático**, una disciplina orientada a desarrollar métodos que les permitan a las máquinas construir modelos analíticos a partir de datos. Como tal, puede considerarse una rama de la inteligencia artificial, basada en el reconocimiento de patrones, o como un área del

reconocimiento de patrones basada en un enfoque de inteligencia artificial.

- La **minería de datos**, una disciplina cuyo objetivo es descubrir patrones en conjuntos de datos complejos. La complejidad puede deberse a su volumen (espacial o de velocidad de generación) o a su falta de estructura. Para ello, se vale de métodos especializados de reconocimiento de patrones, provenientes lo mismo del aprendizaje automático que de la estadística, además de técnicas especializadas para manejo de bases de datos.

Si bien, la ciencia de datos está vinculada con estas disciplinas y su interrelación con todas ellas se vuelve cada vez más estrecha y compleja, aún es importante distinguir su naturaleza como una disciplina con una agenda propia, que se nutre de otras disciplinas como la ingeniería de software y la gestión. A diferencia de otras disciplinas, la finalidad principal de la ciencia de datos es generar información útil para la toma de decisiones en problemas específicos, por lo cual se adoptan métodos provenientes de cualquier disciplina que aporte fortaleza a la solución, con un punto de vista dinámico y flexible.

De entre todas las disciplinas relacionadas con la ciencia de datos, el reconocimiento de patrones juega un papel crucial, al proporcionar los principales elementos de generación de conocimiento y englobar las aportaciones de otras disciplinas a la ciencia de datos, como la estadística y la inteligencia artificial.

1.3 Patrones y percepción

Un patrón puede definirse como una regularidad en un sistema. El **reconocimiento de patrones**, entonces, puede definirse, de una forma general, como la capacidad para identificar estas regularidades.



Figura 1.3: Paisaje agreste mexicano. Las regularidades en la imagen nos permiten reconocer los límites imprecisos entre áreas.

El reconocimiento de patrones es una habilidad natural del ser humano, y de los seres vivos en general, y es lo que nos permite reconocer objetos en ambientes complejos y tomar decisiones de forma cotidiana. En la imagen 1.3, por ejemplo, podemos observar un camino de terracería y aunque se trata

de tierra, piedras y vegetación mezclados, es muy simple distinguir los límites (imprecisos) entre el camino y los bordes, las piedras a un costado y la vegetación. Incluso podemos distinguir en el camino las zonas por donde transitan regularmente los automóviles.

Existen diferentes tipos de regularidades que pueden ser detectadas de forma natural por los seres humanos, entre las que se pueden destacar las siguientes:

- **Armonía:** La armonía puede definirse como un arreglo agradable de las partes de un sistema. Proveniente del griego *ἁρμονία* (harmonía), que significa “conjunto, acuerdo, concordia”, esta palabra se usa comúnmente para denotar formas de concordancia y acuerdo en el ámbito social y musical. Sin embargo, podemos hablar también de armonías visuales, olfativas, gustativas o táctiles.



Figura 1.4: En música, la armonía es el uso de notas simultáneas que juntas producirán una firma sonora única y que son el soporte de la melodía.

- **Ritmo:** El ritmo es cualquier movimiento regular y recurrente, marcado por la sucesión regular de elementos débiles y fuertes, o bien de condiciones opuestas o diferentes. Una onda periódica es una forma simple de comportamiento rítmico, sin embargo, otros comportamientos rítmicos pueden no ser necesariamente periódicos; un caso particularmente interesante son los sistemas cuasi periódicos, como el atractor de Lorenz.

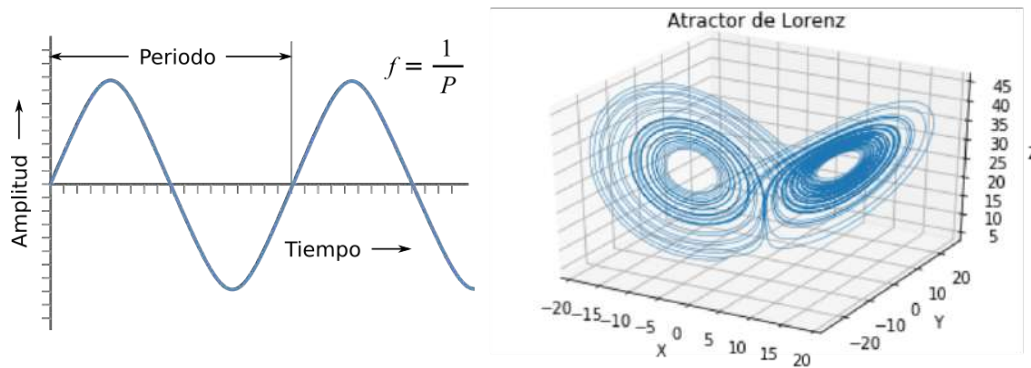


Figura 1.5: La onda periódica es un caso simple de comportamiento rítmico. El atractor de Lorenz, por otra parte, es un ejemplo de sistema rítmico no periódico.

En la definición del ritmo, el movimiento puede hacer referencia a una secuencia “apreciada” de estados: un escalón que “sigue a otro”, por ejemplo.

- **Flujo:** La palabra flujo, proveniente del latín *fluxus* hace referencia al movimiento de una substancia, un fluido, que puede tratarse de una corriente o de una emanación [2]. Aunque típicamente se asocia el flujo a un líquido, actualmente el término se utiliza para referirse a cualquier efecto en el que una cosa parece pasar o viajar, ya sea que se mueva o no. Así, podemos hablar del flujo de actividades en un proceso o del flujo de efectivo en una empresa.

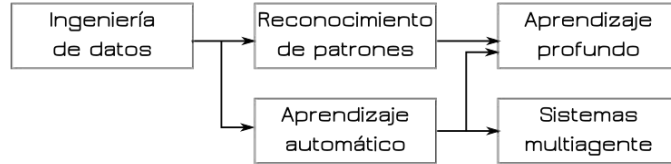


Figura 1.6: El flujo es una regularidad que refleja movimiento en una dirección bien definida, lo mismo de sustancia que de elementos abstractos.

- **Simetría:** La simetría puede definirse como la invariancia de un objeto ante determinada operación. El caso más conocido y, por ello obvio, es la la simetría geométrica.

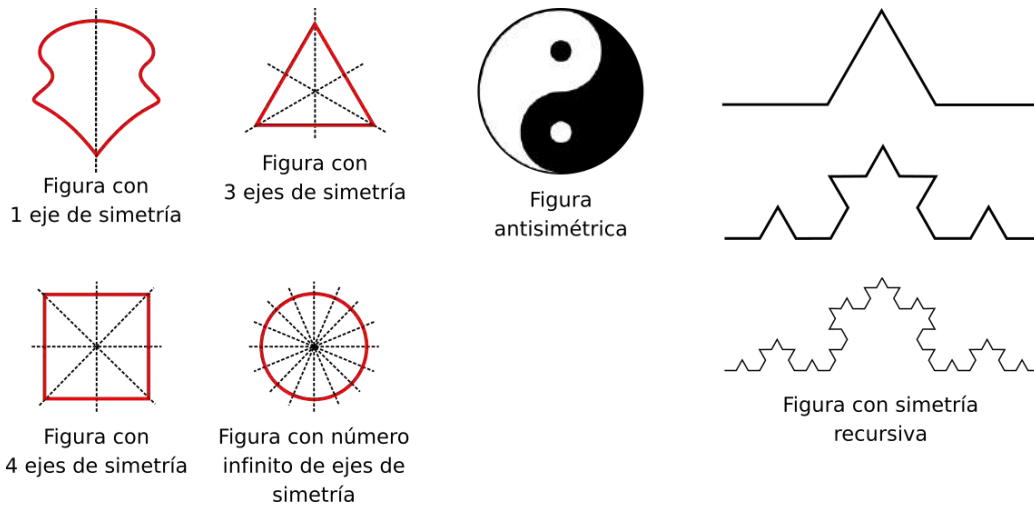


Figura 1.7: Diversas formas de simetría.

- **Proporcionalidad:** La proporcionalidad es una relación o razón constante entre magnitudes medibles.

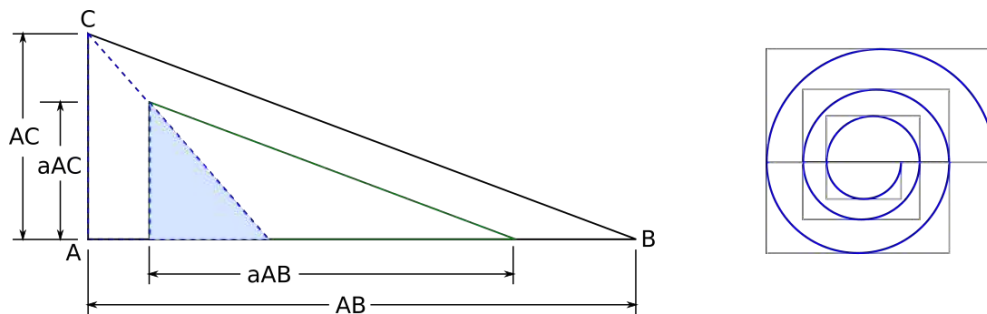


Figura 1.8: La proporcionalidad establece una relación de escalabilidad entre variables.

1.4 Reconocimiento de patrones y conocimiento

El ser humano ha tratado, desde el surgimiento de la sociedad, de identificar las regularidades en un ambiente, típicamente azaroso, para generar modelos abstractos. Estos modelos, utilizados por los seres humanos para dominar su entorno, son representaciones parciales de la realidad, limitadas por las capacidades del ser humano para sentir su entorno. Esta idea fue planteada hace cerca de 2500 años por Protágoras y sujeta, desde entonces, a controversia.

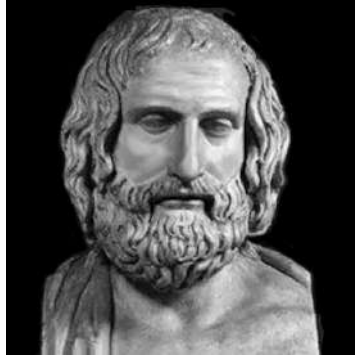


Figura 1.9: Se atribuye a Protágoras de Abdera (c. 485 a. C.- c. 411 a. C.) la frase “El hombre es la medida de todas las cosas”. Esta frase ha sido objeto de gran controversia desde su origen, como se refleja en obras de Platón y Aristóteles. Una de sus interpretaciones es que el ser humano evalúa las cosas de acuerdo a su percepción.

Un punto de vista alterno y complementario, para valorar la calidad de los modelos, se deriva de la alegoría de la caverna, presentada por Platón alrededor de 380 a. C. En esta metáfora, Platón describe a un grupo de prisioneros que han vivido encadenados en una cueva viendo hacia una pared desnuda. Los prisioneros juzgan el mundo a través de las sombras proyectadas en ella por objetos comunes, en contraposición a los filósofos que ven los objetos como son. De esta manera, Platón discute la importancia de la educación: los modelos de los prisioneros son pobres, en este caso, debido a la falta de conocimiento, pero pueden ser mejorados gracias a un estudio más profundo.



Figura 1.10: Alegoría de la caverna de Platón, grabado de Jan Saenredam (1604), Museo británico.

Al margen de esta discusión por demás, filosóficamente interesante (si los modelos utilizados por los seres humanos son incompletos o erróneos por ignorancia o por imposibilidad de aprehensión de la realidad), es importante destacar la naturaleza incompleta de cualquier modelo, ya sea producto de la

inferencia que los métodos de reconocimiento de patrones, por ejemplo, o de teorías científicas bien establecidas.

1.4.1 Pensamiento computacional

El pensamiento computacional es un concepto que describe el proceso de razonamiento que conduce a la resolución de un problema utilizando medios computacionales. El término fue introducido originalmente por Seymour Papert [3] para resaltar el potencial de las representaciones computacionales para expresar ideas complejas, particularmente mediante el uso de simulaciones para enfatizar los aspectos dinámicos de los fenómenos físicos, en la enseñanza. Posteriormente, el concepto fue abordado por Jeannette M. Wing [4] en el sentido que se utiliza actualmente, como la habilidad para descomponer un problema de manera que se pueda resolver de manera algorítmica. Wing enfatiza su importancia como una competencia básica para el aprendizaje, adicional a las de lectura, escritura y aritmética. De forma controversial, en algunos casos el pensamiento computacional es presentado como una tercera forma de generar conocimiento, además de las formas tradicionales de teoría y experimentación [5].

De acuerdo a este concepto, las actividades de resolución de problemas se pueden organizar en cuatro etapas:

- Descomposición: En esta etapa, el problema se replantea en pequeños subproblemas más manejables.
- Reconocimiento de patrones: En esta etapa, cada subproblema es analizado tratando de encontrar similitudes con otros casos resueltos u observados anteriormente.
- Abstracción: Esta etapa consiste en replantear el problema global y sus subproblemas, así como la solución, tratando de eliminar los aspectos específicos al caso y generalizar problema y solución.
- Algoritmos: Finalmente, se presenta la solución al problema en términos de una secuencia de pasos sistemáticos.

Hay que notar que, en este contexto, el término de reconocimiento de patrones suele utilizarse de manera más general que el utilizado en ciencias de la computación, es decir, no sólo se trata del reconocimiento de patrones a partir de los datos, sino de segmentos de solución de problemas en general. También es digno de resaltar la coincidencia del enfoque del pensamiento computacional con el de ciencia de datos, en el sentido de estar enfocado en casos específicos y en el uso de las herramientas necesarias para sistematizar su análisis.

Conclusiones

La ciencia de datos es un campo interdisciplinario cuyo objetivo es generar el conocimiento acerca de un problema que permita tomar decisiones, enfocándose más a la solución de casos concretos que a la creación de modelos generales. En esta tarea, la ciencia de datos echa de toda clase de herramienta que ofrezca información de mayor valor, desde la estadística hasta las ciencias sociales y administrativas. Una de las áreas centrales de la ciencia de datos es el reconocimiento de patrones.

El reconocimiento de patrones, es una disciplina de las ciencias de la computación que busca identificar regularidades a partir de observaciones. Sin embargo, su fundamento está fuertemente anclado en las habilidades de los seres vivos para adaptarse a su ambiente. Una consideración importante que conviene hacer, al abordar el estudio de esta disciplina, es la relatividad del valor de los modelos, tanto los desarrollados mediante las diversas técnicas de reconocimiento de patrones, como los modelos

existentes desarrollados a partir de teorías científicas bien establecidas. En cualquier caso, el objetivo es establecer una relación explicativa clara entre los modelos teóricos y los datos observados: el objetivo de un modelo es servir para entender, explicar y predecir el comportamiento del entorno.

Bibliografía

Articles

- [1] C. Zins. «Conceptual approaches for defining data, information, and knowledge». En: *Journal of the American Society for Information Science and Technology* 58.4 (2007) (véase página 11).
- [3] S. Papert. «An exploration in the space of mathematics educations». En: *International Journal of Computers for Mathematical Learning* 1.1 (1996) (véase página 17).
- [4] J.M. Wing. «Computational Thinking». En: *Communications of the ACM* 49.33 (2006) (véase página 17).
- [5] P.J. Denning. «Computational Thinking in Science». En: *American Scientist* 105 (2017) (véase página 17).

Books

- [2] E. Cram D. Linn; A.R.; Nowak. *History of Linguistics 1996: Volume 2: From Classical to Contemporary Linguistics*. John Benjamins Publishing, 1999 (véase página 14).
- [6] P. Chapman y col. *CRISP-DM 1.0 Step-by-step data mining guides*. Reporte técnico. SPSS Inc., 2000 (véanse páginas 19, 20).
- [7] D.O. Hebb. *The Organization of Behavior. A Neuropsychological Theory*. New York: John Wiley & Sons, Inc., 1949 (véase página 197).