

Recolección, tratamiento y visualización de datos para inteligencia artificial

Breve descripción:

Este componente formativo aborda la recolección, tratamiento y visualización de datos para inteligencia artificial. Incluye tipos de datos, fuentes según origen y calidad, principios éticos como el consentimiento informado, y herramientas actuales para el análisis y visualización, facilitando la interpretación y la toma de decisiones basadas en datos.

Tabla de contenido

Introducción	4
1. Recolección y caracterización de datos	7
1.1. Tipos de datos	9
1.2. Técnicas de definición de datos	12
1.3. Fuentes de información.....	16
1.4. Principios éticos en el tratamiento de datos	19
2. Procesamiento y preparación de datos	21
2.1. Técnicas de limpieza de datos.....	21
2.2. Preparación de datos para inteligencia artificial	25
2.3. Estadística descriptiva	26
3. Inteligencia Artificial (IA) aplicada al tratamiento de datos	29
3.1. Concepto, origen y evolución de la inteligencia artificial	29
3.2. Funciones y aplicaciones de la IA en el procesamiento de datos	31
3.3. Tipologías de IA: descriptiva, predictiva y generativa.....	35
3.4. IA generativa: características, usos y retos	36
3.5. Interacción con modelos generativos: el papel del prompt.....	39
4. Análisis y visualización de datos	43
4.1. Herramientas de visualización de datos	43

4.2. Herramientas de analítica de datos	44
4.3. Introducción al aprendizaje automático (machine learning)	45
Síntesis	47
Material Complementario	48
Glosario	50
Referencias bibliográficas	52
Créditos	54

Introducción

Este componente formativo aborda los fundamentos esenciales para la recolección, procesamiento, análisis y visualización de datos, integrando herramientas de inteligencia artificial (IA) que potencian su uso en distintos contextos. A lo largo del contenido, se exploran tipos y fuentes de datos, principios éticos para su tratamiento, técnicas de preparación y limpieza, así como nociones clave de estadística descriptiva. Además, se profundiza en el papel de la IA en el procesamiento de datos, diferenciando entre IA generativa y descriptiva, y se presentan herramientas tecnológicas actuales para su análisis y visualización. Esta formación busca fortalecer competencias técnicas necesarias para integrar datos de forma eficiente y responsable en procesos apoyados por inteligencia artificial. Para comprender la importancia del contenido y los temas abordados, se recomienda acceder al siguiente video:

Video 1. Recolección, tratamiento y visualización de datos para inteligencia artificial



Enlace de reproducción del video

Síntesis del video: Recolección, tratamiento y visualización de datos para inteligencia artificial

Los datos están en todas partes, desde interacciones cotidianas en redes sociales hasta complejos sistemas de inteligencia artificial. Para que esta información sea útil, primero debe ser recolectada, caracterizada y comprendida. Es por eso que resulta esencial reconocer los tipos de datos, definirlos correctamente, identificar sus fuentes y garantizar un tratamiento ético.

Una vez obtenidos, los datos pasan por procesos de limpieza y preparación, donde la calidad y la organización son clave. Aquí entra en juego la estadística descriptiva, que permite interpretar la información y extraer patrones significativos.

Con datos bien preparados, la inteligencia artificial se convierte en una poderosa aliada. Comprender su evolución, sus funciones y sus múltiples aplicaciones en el procesamiento de datos permite aprovechar su potencial al máximo. Desde modelos que describen lo ocurrido, hasta sistemas que predicen o generan contenido nuevo, la IA transforma datos en conocimiento accionable. La interacción con modelos generativos, como los que crean texto o imágenes, depende en gran parte de saber formular correctamente un prompt.

Por último, la visualización y el análisis de datos ofrecen herramientas para tomar decisiones basadas en evidencia. A través de gráficos, paneles y modelos analíticos, es posible entender lo que los datos realmente nos dicen. Y gracias al

aprendizaje automático, las máquinas pueden incluso aprender de ellos y anticipar comportamientos.

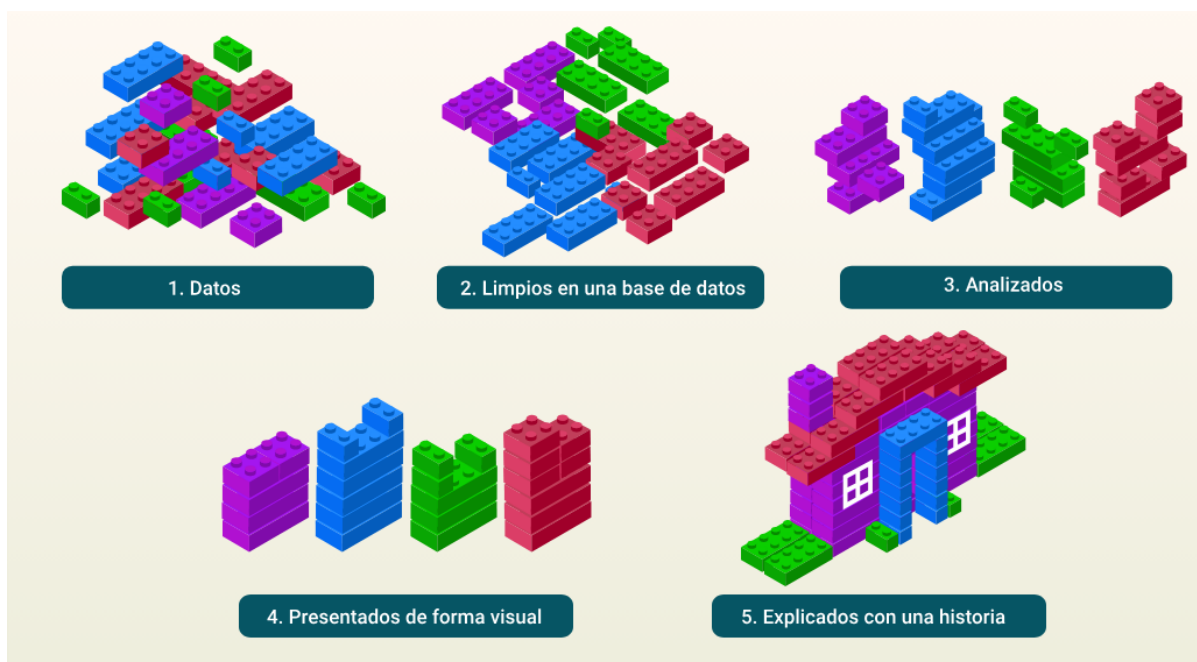
Un viaje completo que une datos, inteligencia y tecnología en una sola experiencia de aprendizaje.

1. Recolección y caracterización de datos

El análisis de datos es un proceso que inicia mucho antes del uso de herramientas estadísticas o informáticas. Parte de una comprensión profunda sobre qué son los datos, cómo se obtienen, de dónde provienen y bajo qué principios deben ser tratados. Por esta razón, en este apartado se abordan los aspectos fundamentales relacionados con la recolección y caracterización de datos, entendiendo que estos constituyen la base sobre la cual se construyen procesos analíticos, diagnósticos y soluciones efectivas a problemas reales.

Para facilitar esta comprensión, se propone la siguiente figura del ciclo de transformación de los datos, presentado a través de bloques de construcción:

Figura 1. Analogía de los datos



Nota. Tomado de Manuel Rodríguez (2025).

Esta figura permite representar de forma didáctica el proceso que siguen los datos desde su recolección hasta su uso significativo:

- ✓ **Datos:** en su estado inicial, los datos aparecen como elementos dispersos, sin organización ni estructura. Representan información cruda recolectada de distintas fuentes, que aún no ha sido procesada.
- ✓ **Limpios en una base de datos:** en esta etapa, los datos se organizan y depuran para eliminar inconsistencias, errores o duplicaciones. Esta limpieza permite clasificarlos y almacenarlos adecuadamente en una base de datos estructurada.
- ✓ **Analizados:** una vez organizados, los datos se someten a análisis para identificar patrones, tendencias o relaciones. Esta fase corresponde al uso de técnicas estadísticas o computacionales para extraer conocimiento útil.
- ✓ **Presentados de forma visual:** los resultados del análisis se traducen en representaciones gráficas, visualizaciones o cuadros de datos que permiten una comprensión más clara de la información.
- ✓ **Explicados con una historia:** finalmente, los datos analizados y visualizados se contextualizan dentro de una narrativa que permite tomar decisiones informadas, comunicar hallazgos y construir conocimiento. Este paso también implica una reflexión ética sobre cómo se interpretan y comunican los resultados.

Esta secuencia demuestra que el trabajo con datos requiere no solo habilidades técnicas, sino también pensamiento crítico, criterio ético y capacidad de comunicación. En ese sentido, los subtemas de este apartado tipos de datos, técnicas de definición,

fuentes de información y principios éticos, permiten al aprendiz comprender de forma integral las primeras etapas de un proceso analítico riguroso y confiable.

1.1. Tipos de datos

Los datos pueden definirse como hechos u observaciones concretas, recolectados a partir de diversas fuentes, que por sí mismos pueden carecer de significado hasta que son organizados, procesados y analizados. Es en este proceso cuando se transforman en información útil para la toma de decisiones. En este punto es importante definir cuál es la diferencia entre un dato y la información:

Los datos son el insumo bruto, como cifras, palabras, imágenes, sonidos, entre otros. En cambio, la información es el resultado procesado que tiene valor contextual y utilidad. Esta distinción es clave: así como el carbón se convierte en diamante tras someterse a un proceso riguroso, los datos se transforman en información mediante técnicas de análisis, modelado y presentación.

La naturaleza de los datos hace referencia a su forma de representación, su nivel de organización y la manera en que pueden ser interpretados o manipulados por sistemas informáticos. Esta clasificación resulta fundamental al momento de diseñar bases de datos, seleccionar herramientas de análisis y determinar los procesos requeridos para transformar los datos en información útil.

En términos generales, los datos pueden clasificarse en tres grandes grupos:

a) Datos estructurados

Son aquellos que se organizan de manera ordenada en filas y columnas, como en las bases datos relacionales. Esta organización facilita su búsqueda, análisis y

recuperación. Son comunes en sistemas administrativos, registros de usuarios y encuestas cerradas.

b) Datos no estructurados

Corresponden a aquellos que no siguen un esquema fijo ni una organización tabular. Incluyen archivos de texto libre, imágenes, audios, videos y publicaciones en redes sociales. Aunque son ricos en contenido, requieren herramientas específicas para su procesamiento, como algoritmos de procesamiento de lenguaje natural (PLN) o análisis de imágenes.

c) Datos semiestructurados

Tienen una estructura flexible, generalmente basada en etiquetas o jerarquías, como ocurre con los archivos XML, JSON, o YAML. Aunque no están organizados en una base relacional, si contienen metadatos que permiten cierta organización lógica y facilitan su interpretación por programas especializados.

Además, es importante considerar una clasificación más detallada según el tipo específico de dato, que puede incluir representaciones numéricas (enteros o decimales), cadenas de texto, fechas y horas, archivos multimedia o valores booleanos, como se presenta en la siguiente tabla:

Tabla 1. Clasificación según el tipo

Tipo de dato	Característica principal	Ejemplo
Bits y bytes	Representación básica binaria de todos los datos digitales	0 y 1
Números enteros	Sin decimales	145
Números decimales	Con cifras decimales	45.78
Texto	Codificado con estándares como ASCII o UTF	“SENA”, “Bogotá”
Fecha y hora	Referencia temporal, expresada en distintos formatos	2025-05-30, 14:30:00
Imagen	Representación visual en pixeles	JPG, PNG
Video	Secuencia de imágenes con audio y frecuencia de reproducción (FPS)	MP4, AVI, MOV
Booleano	Solo dos valores posibles.	Verdadero (1) o Falso (0)

Esta clasificación no solo permite identificar la forma en que se almacenan y presentan los datos, sino que también ayuda a establecer los requerimientos técnicos para su tratamiento posterior, como el tipo de almacenamiento, los métodos de análisis y las herramientas computacionales requeridas. Comprender la naturaleza y el tipo de datos es el primer paso para caracterizarlos adecuadamente, evaluar su calidad y garantizar su utilidad en entornos de análisis de información, modelamiento predictivo y toma de decisiones basadas en datos.

Además de su tipo y estructura, los datos poseen una serie de características complementarias que influyen directamente en su gestión y análisis:

- ✓ **Formato:** modo de almacenamiento o presentación (CSV, Excel, JSON).
- ✓ **Estructura:** grado de organización interna (estructurado o no).
- ✓ **Origen:** fuente que los produce (sistema, encuesta, sensor).
- ✓ **Volumen:** cantidad de datos disponibles.
- ✓ **Velocidad:** rapidez con la que se generan.
- ✓ **Veracidad:** grado de confiabilidad.
- ✓ **Valor:** utilidad que aportan para resolver un problema.

1.2. Técnicas de definición de datos

Los datos están en constante transformación: se trasladan, se procesan y evolucionan. Su objetivo principal es convertirse rápidamente en información útil que genere el valor que las organizaciones necesitan. Por esta razón, resulta fundamental establecer normas claras de gestión que permitan administrar y controlar eficazmente su ciclo de vida, ya que los datos se han convertido en el motor clave de la transformación digital.

La gestión de los datos no depende exclusivamente de los sistemas informáticos. Su eficacia está directamente relacionada con la alineación entre las personas, los procesos y la tecnología. Para aplicar técnicas de definición de datos, se pueden utilizar herramientas y enfoques organizados en tres categorías principales:

- a) **Herramientas de gestión de datos:** describen las funciones y responsabilidades de los perfiles clave involucrados en el tratamiento y análisis de los datos:
 - ✓ **Data Scientist:** especialistas en datos, tradicionalmente con formación en matemáticas o estadística. Su labor se enfoca en analizar grandes volúmenes de datos ya preparados, con el fin de identificar patrones y generar modelos predictivos que respalden la toma de decisiones.
 - ✓ **Data Analyst:** se encarga de transformar los datos en información útil y comprensible mediante herramientas analíticas. Su función principal es generar informes e interpretaciones que permitan entender el comportamiento de los datos.
 - ✓ **Data Engineer:** colabora estrechamente con los analistas y científicos de datos, asegurando que la infraestructura técnica esté preparada para recibir, procesar y organizar los datos de manera óptima.
- b) **Metadatos:** permiten definir, describir y contextualizar la información. Actúan como etiquetas que eliminan ambigüedades y aseguran una interpretación uniforme. Por ejemplo, un campo como “fecha” puede tener múltiples formatos y significados; el metadato define si se refiere a

una fecha de creación, modificación o vencimiento, así como su formato estándar. En este sentido, los metadatos son para los datos lo que una etiqueta es para un producto.

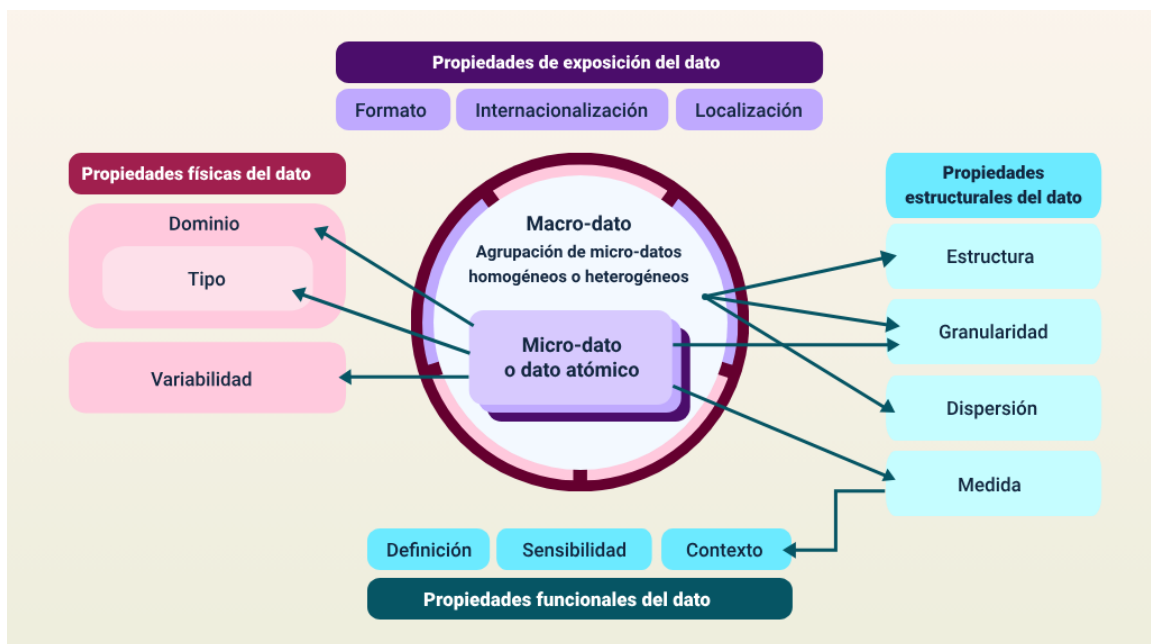
c) **Seguridad de los datos:** la protección de la información almacenada es un aspecto esencial. Se deben aplicar mecanismos que garanticen su seguridad sin comprometer su accesibilidad. Entre las técnicas más comunes se encuentran:

- ✓ **Anonimización y seudonimización:** modifican los datos para eliminar o sustituir cualquier elemento que permita identificar a una persona, impidiendo así la identificación directa del individuo. La anonimización elimina por completo los identificadores, mientras que la seudonimización los reemplaza por códigos.
- ✓ **Cifrado de información:** convierte los datos legibles en información ininteligible mediante algoritmos, tanto cuando los datos están almacenados como cuando están en tránsito. Esto impide que terceros no autorizados puedan acceder a su contenido. Actualmente, servicios como Google Drive, AWS S3 y Microsoft OneDrive protegen los datos mediante cifrado. Asimismo, las páginas web que utilizan HTTPS cifran la información entre el navegador y el servidor mediante protocolos como SSL/TLS.

Para complementar las técnicas de definición, es importante comprender las propiedades fundamentales de los datos. Estas propiedades se agrupan en cuatro grandes dimensiones tales como físicas, estructurales, funcionales y de exposición.

Comprenderlas permite caracterizar mejor los datos y asegurar su uso eficiente y contextualizado.

Figura 2. Propiedades físicas, estructurales, funcionales y de exposición del dato



Nota. Tomado de Benoit Cayla.

Propiedades físicas, estructurales, funcionales y de exposición del dato

✓ **Propiedades de exposición del dato**

- ✓ Formato.
- ✓ Internacionalización.
- ✓ Localización.

a) **Macro-dato:** Agrupación de micro-datos/homogéneos o heterogéneos.

✓ **Propiedades físicas del dato**

- Dominio (Tipo)=Micro dato o dato atómico.

- Variabilidad= Micro dato o dato atómico.

✓ **Propiedades estructurales del dato**

- Estructura.
- Granularidad.
- Dispersión.
- Medida: Definición, sensibilidad y contexto.

En la actualidad, no es posible hablar de datos sin mencionar el papel creciente de la Inteligencia Artificial (IA). Más que una tendencia tecnológica, la IA se ha convertido en un elemento estratégico para empresas, científicos de datos y desarrolladores de software. Sin embargo, su uso excesivo y difuso ha contribuido a que el término pierda claridad, generando confusión sobre su verdadero alcance. Aun así, sus aplicaciones prácticas en la automatización del análisis, la predicción y la optimización de procesos hacen que la IA sea una herramienta imprescindible en cualquier estrategia moderna de datos.

1.3. Fuentes de información

Las fuentes de información constituyen un pilar fundamental en los procesos de investigación, análisis y toma de decisiones. Permiten acceder al conocimiento, sustentar argumentaciones y validar resultados. La elección adecuada de las fuentes garantiza la calidad, la objetividad y la veracidad de la información utilizada.

Las fuentes se clasifican, principalmente, en función de su origen, nivel de elaboración del contenido y propósito comunicativo. A continuación, se presentan los principales tipos de fuentes de información:

- a) **Fuentes primarias:** corresponden a documentos originales que contienen información inédita o datos directos sobre un tema específico. Incluyen investigaciones científicas, informes técnicos, estadísticas oficiales, documentos legales, entrevistas, testimonios y registros históricos. Su valor reside en ofrecer datos de primera mano, sin intermediarios ni interpretaciones previas.
- b) **Fuentes secundarias:** se basan en el análisis, interpretación o síntesis de fuentes primarias. Ejemplos comunes son los artículos de revisión, libros de texto, ensayos críticos y enciclopedias. Estas fuentes son útiles para contextualizar la información, facilitar su comprensión y proporcionar una visión más amplia del tema tratado.
- c) **Fuentes terciarias:** recopilan, organizan y resumen tanto fuentes primarias como secundarias. Manuales, catálogos, índices bibliográficos, guías temáticas y bases de datos pertenecen a esta categoría. Su función principal es facilitar el acceso y la localización de otras fuentes más detalladas.

Las fuentes de información también pueden clasificarse según su procedencia. Estas pueden ser académicas, institucionales, gubernamentales, comerciales o populares. Las fuentes académicas, como artículos publicados en revistas científicas y libros especializados, se distinguen por su rigurosidad metodológica y su proceso de revisión por pares. En contraste, las fuentes comerciales o populares, como blogs, redes sociales o sitios web sin revisión, pueden carecer de objetividad, contener sesgos o carecer de fundamentos científicos.

El contenido de las fuentes puede ser de tipo cuantitativo (datos numéricos y medibles) o cualitativo (descripciones, opiniones o experiencias), y presentarse de forma estructurada (tablas, gráficos o esquemas) o no estructurada (textos libres, imágenes o audios). La elección del tipo de contenido debe responder al propósito de la consulta o investigación.

Para garantizar la validez y confiabilidad de la información consultada, es necesario aplicar criterios que permitan evaluar la calidad de las fuentes:

- 1) **Autoridad:** identifica quién es el autor o la institución responsable, y valora su experiencia, formación y reconocimiento en el tema.
- 2) **Exactitud:** verifica que los datos sean precisos, contrastables y estén libres de errores.
- 3) **Actualidad:** considera la fecha de publicación para asegurar que la información sea vigente y relevante.
- 4) **Objetividad:** evalúa la neutralidad del contenido, identificado posibles sesgos, intereses comerciales o ideológicos.
- 5) **Cobertura:** analiza si la fuente trata el tema en profundidad, de forma completa y con enfoque integral.
- 6) **Verificabilidad:** comprueba si la fuente proporciona referencias, citas o evidencia que respalde lo expuesto.

En síntesis, reconocer y seleccionar fuentes confiables es una competencia crítica en el tratamiento de la información. La rigurosidad en este proceso fortalece la solidez de cualquier trabajo investigativo o proyecto de análisis de datos.

1.4. Principios éticos en el tratamiento de datos

En un entorno cada vez más digitalizado, el uso responsable de los datos personales se ha convertido en una prioridad ética y legal. El tratamiento adecuado de la información no solo implica el cumplimiento de normativas, sino también el respeto por los derechos fundamentales de las personas, garantizando transparencia, equidad y protección.

Para asegurar una gestión responsable de los datos, se deben considerar los siguientes principios éticos de datos:

- ✓ **Legalidad:** todo tratamiento de datos debe estar respaldado por una base jurídica clara, en cumplimiento con la normativa vigente. En el contexto colombiano, esto implica adherirse a la Ley 1581 de 2012 y sus decretos reglamentarios.
- ✓ **Consentimiento informado:** los titulares de los datos deben conocer de forma clara y previa cómo, por qué y para qué se recolectan y usan sus datos. Este consentimiento debe ser libre, específico, informado e inequívoco.
- ✓ **Finalidad:** los datos recolectados deben utilizarse exclusivamente para los fines definidos al momento de su obtención. Usos distintos requieren una nueva autorización.
- ✓ **Seguridad:** es necesario implementar medidas técnicas, organizativas y administrativas que garanticen la protección de los datos contra pérdida, manipulación, acceso no autorizado o uso indebido.

- ✓ **Transparencia:** las organizaciones deben proporcionar información clara y accesible sobre el tratamiento de los datos, permitiendo a los titulares ejercer sus derechos de acceso, rectificación, cancelación y oposición.
- ✓ **Minimización:** se deben recolectar únicamente los datos estrictamente necesarios para cumplir con el propósito definido, evitando el exceso de información innecesaria.

Con el crecimiento exponencial de la Inteligencia Artificial (IA), los principios éticos en el tratamiento de datos adquieren una dimensión más compleja. La IA se nutre de grandes volúmenes de datos para entrenar modelos, tomar decisiones y generar predicciones. No obstante, su uso sin un marco ético puede derivar en sesgos algorítmicos, violaciones a la privacidad, discriminación o exclusión de ciertos grupos poblacionales.

En este contexto, se hace urgente establecer marcos regulatorios específicos que rijan el uso de tecnologías emergentes, así como promover una cultura organizacional centrada en la ética del dato. Esta debe basarse en la responsabilidad, la inclusión, la transparencia y el compromiso con la equidad, garantizando que la innovación tecnológica esté al servicio de las personas y no en su contra.

En conclusión, el tratamiento ético de los datos no es solo una obligación legal, sino un imperativo moral en la sociedad contemporánea. Su cumplimiento fortalece la confianza en las instituciones, protege los derechos ciudadanos y promueve un desarrollo tecnológico justo y sostenible.

2. Procesamiento y preparación de datos

La etapa de procesamiento y preparación de datos constituye una de las fases más críticas en cualquier proyecto de análisis de datos, ciencia de datos o desarrollo de sistemas de inteligencia artificial. Esta fase determina en gran medida la calidad y la eficacia de los modelos y análisis que se construyan posteriormente. Se encarga de transformar datos crudos, frecuentemente desordenados, incompletos o con errores, en un conjunto de información precisa, coherente y organizada que pueda ser utilizada de forma eficiente por herramientas analíticas o algoritmos automatizados.

En este proceso, se desarrollan una serie de actividades sistemáticas que van desde la limpieza básica de datos hasta la transformación y reestructuración profunda de los mismos, con el objetivo de adaptarlos al contexto específico del problema que se desea resolver. Esto incluye tareas como la detección de valores atípicos, el tratamiento de datos faltantes, la normalización de escalas, la codificación de variables categóricas y la integración de diferentes fuentes de información.

Una adecuada preparación de datos no solo garantiza la integridad de los análisis, sino que también mejora el rendimiento de los modelos de inteligencia artificial, minimizando errores, reduciendo sesgos y aumentando la capacidad predictiva. Por ello, esta etapa se considera un requisito indispensable para alcanzar resultados confiables y significativos en cualquier proceso basado en datos.

2.1. Técnicas de limpieza de datos

La limpieza de datos es una fase fundamental dentro del procesamiento de datos, ya que garantiza la calidad, consistencia y precisión de la información que será utilizada en los análisis posteriores. A menudo, los datos recolectados presentan errores, valores

incompletos, duplicaciones o formatos inconsistentes que pueden afectar la validez de los resultados. Por esta razón, es indispensable aplicar un conjunto de técnicas sistemáticas que permitan corregir, transformar o eliminar estos problemas antes de alimentar modelos analíticos o de inteligencia artificial.

Entre las principales técnicas de limpieza de datos se encuentran:

- ✓ **Eliminación o imputación de valores faltantes:** los valores ausentes pueden tratarse eliminando filas o columnas incompletas, o bien imputando dichos valores utilizando métodos estadísticos simples (como media, mediana o moda) o técnicas más avanzadas como modelos de regresión, árboles de decisión o algoritmos de imputación múltiple.
- ✓ **Corrección de valores erróneos o inconsistentes:** los datos pueden presentar errores tipográficos, unidades mal definidas o registros ilógicos (como fechas futuras en historiales pasados). Estas inconsistencias se detectan mediante reglas de validación, análisis de dominios de valores o comparación con fuentes confiables.
- ✓ **Eliminación de duplicados:** la presencia de registros duplicados genera un sesgo en los análisis. Es común aplicar funciones de deduplicación que identifican y eliminan entradas repetidas basadas en claves únicas o atributos combinados.
- ✓ **Normalización y estandarización:** estas técnicas ajustan la escala de los datos numéricos. La normalización transforma los valores para que estén en un rango específico (por ejemplo, entre 0 y 1), mientras que la estandarización los convierte a una distribución con media 0 y desviación

estándar 1. Esto es especialmente importante para modelos sensibles a la escala, como los basados en distancia euclidiana.

- ✓ **Conversión de tipos de datos:** asegurar que cada variable tenga el tipo de dato adecuado (números, fechas, booleanos y texto) es fundamental para realizar operaciones precisas y evitar errores en cálculos o visualizaciones.
- ✓ **Tratamiento de valores atípicos (outliers):** los valores extremos pueden distorsionar los resultados estadísticos o los modelos predictivos. Estos valores se detectan mediante análisis exploratorio y pueden eliminarse, transformarse o mantenerse si son significativos para el fenómeno que se analiza.

Una correcta ejecución de estas técnicas no solo mejora la calidad de los datos, sino que también optimiza el desempeño de los modelos de inteligencia artificial y aumenta la confiabilidad de los procesos de toma de decisiones basados en datos.

Además, con el auge de la inteligencia artificial generativa, es fundamental comprender los distintos tipos de IA y cómo estos dependen de la calidad de los datos procesados. Una correcta limpieza y preparación de los datos impacta directamente en el rendimiento de los sistemas inteligentes, ya sean descriptivos o generativos. A continuación, se presenta una comparación entre dos enfoques clave de la IA:

Tabla 2. Características de la IA Descriptiva y la IA Generativa

Característica	IA Descriptiva	IA Generativa
Objetivo	Analizar y describir datos existentes	Crear contenido nuevo a partir de datos

Característica	IA Descriptiva	IA Generativa
Resultado	Informes, visualizaciones y resúmenes	Textos, imágenes, música y modelos 3D
Técnica base	Estadística, minería de datos y ML supervisado	Redes neuronales profundas, GANs y transformers
Ejemplos	Tableau, Power BI y dashboards	ChatGPT, DALL·E, MidJourney y GitHub Copilot

La diferencia entre estos tipos de inteligencia artificial también se manifiesta en sus aplicaciones prácticas, donde el procesamiento previo de datos es un paso determinante para garantizar la utilidad y eficiencia de los sistemas. Algunos ejemplos destacados incluyen:

a) Sector salud

La IA facilita la integración de registros clínicos dispersos (Historias medicas, resultados de laboratorio, prescripciones, etc.) en un único expediente por paciente . Esto se logra mediante algoritmos de aprendizaje automático y técnicas de duplicación inteligente, lo cual mejora la calidad de la atención y reduce errores administrativos.

b) Comercio electrónico

En plataformas como Amazon o Mercado Libre, la IA permite consolidar información proveniente de múltiples proveedores (como precios, inventarios o

descripciones de productos) a través del reconocimiento de patrones y el procesamiento del lenguaje natural (PLN). Esta integración optimiza la experiencia del usuario, mejora los motores de recomendación y contribuye a una gestión más eficiente del catálogo.

Este contexto refuerza la importancia del procesamiento y preparación de datos, ya que la capacidad de los sistemas de IA para ofrecer resultados útiles y precisos depende en gran medida de la calidad, integridad y estructura de los datos que reciben.

2.2. Preparación de datos para inteligencia artificial

La preparación de datos constituye una fase crítica dentro del ciclo de vida de los proyectos de Inteligencia Artificial (IA) y aprendizaje automático (machine learning). Este proceso abarca una serie de actividades destinadas a transformar los datos crudos en insumos útiles y adecuados para alimentar modelos analíticos. La calidad, integridad y estructura de los datos impactan directamente en el desempeño, precisión y validez de los modelos construidos.

Entre las tareas esenciales en esta etapa se encuentran:

- ✓ Integración de fuentes de datos.
- ✓ Selección de variables relevantes.
- ✓ Tratamiento de valores faltantes.
- ✓ Conversión de formatos.
- ✓ Transformación de variables categóricas.
- ✓ Conversión de formatos.

Todo esto se realiza con el objetivo de garantizar que los datos sean consistentes, completos, estructurados y libres de errores o valores extremos que puedan afectar negativamente el análisis posterior.

En el contexto de la IA, esta preparación también implica alinear los datos con los requerimientos específicos del algoritmo que se va a emplear. Por ejemplo, los modelos de clasificación requieren variables codificadas, mientras que los modelos de redes neuronales suelen necesitar normalización de valores y estructuras matriciales específicas.

Además, se deben tener en cuenta aspectos como el sesgo en los datos, la representatividad de las muestras y la ética en el manejo de información sensible, especialmente en aplicaciones como salud, finanzas o justicia. Una adecuada preparación de datos no solo mejora el rendimiento de los modelos, sino que también contribuye a la generación de resultados más confiables, interpretables y útiles para la toma de decisiones.

2.3. Estadística descriptiva

La estadística descriptiva es una herramienta fundamental en el análisis de datos, ya que permite resumir, organizar y visualizar la información recopilada de forma clara y comprensible. Este conjunto de técnicas constituye el punto de partida para explorar los datos antes de aplicar modelos más complejos de análisis predictivo o aprendizaje automático. Su función es identificar patrones generales, distribuciones, tendencias y posibles valores atípicos que podrían afectar la interpretación posterior.

Para aplicar la estadística descriptiva correctamente, es importante reconocer los tipos de datos con los que se trabaja:

- a) **Cualitativos (categóricos):** representan cualidades o atributos no numéricos, como el color de un producto o el género de una persona.
- b) **Cuantitativos (numéricos):** expresan cantidades y se subdividen en:
 - ✓ **Discretos:** toman valores contables, como el número de hijos o de productos vendidos.
 - ✓ **Continuos:** pueden asumir infinitos valores dentro de un rango, como la altura o el peso.

Además, los datos se pueden clasificar según su nivel de medición, lo cual influye en los tipos de análisis estadístico que se pueden realizar:

- ✓ **Nominal:** variables categóricas sin un orden definido, como los nombres de ciudades.
- ✓ **Ordinal:** categorías con jerarquía, como los niveles de satisfacción (bajo, medio, alto).
- ✓ **Intervalo:** variables numéricas sin un cero absoluto, como la temperatura en grados Celsius.
- ✓ **Razón:** variables numéricas con cero absoluto, como los ingresos o el tiempo.

En función de su naturaleza, las variables también se distinguen en:

- ✓ **Categóricas:** corresponden atributos o cualidades (de nivel nominal u ordinal).
- ✓ **Numéricas:** representan cantidades medibles (de nivel intervalo o razón).

Entre las herramientas más utilizadas para describir los datos se encuentran:

- 1) **Histogramas:** permiten visualizar la distribución de frecuencias de variables continuas, facilitando el análisis de la forma, la simetría y la dispersión de los datos.
- 2) **Tablas de contingencia:** también conocidas como tablas cruzadas, se utilizan para observar la relación entre dos variables categóricas, identificando patrones conjuntos, dependencias o asociaciones.
- 3) **Diagramas de barras:** representan frecuencias de variables categóricas. Son útiles para comparar categorías de forma clara y rápida.
- 4) **Diagramas de caja (boxplots):** muestran la mediana, los cuartiles y los valores atípicos de una variable numérica. Son especialmente útiles para comparar distribuciones entre grupos.
- 5) **Medidas de tendencia central y dispersión:** como la media, mediana, moda, rango, desviación estándar y varianza. Estas permiten resumir cuantitativamente el comportamiento de una variable.
- 6) **Diagramas de dispersión (scatter plots):** permiten explorar la relación entre dos variables numéricas, ayudando a identificar correlaciones o tendencias.

El uso adecuado de estas herramientas permite obtener una primera comprensión del comportamiento de los datos y facilita la toma de decisiones informadas antes de implementar modelos analíticos o predictivos más avanzados.

3. Inteligencia Artificial (IA) aplicada al tratamiento de datos

La inteligencia artificial ha transformado el tratamiento de datos al facilitar el procesamiento de grandes volúmenes de información de manera eficiente, precisa y automatizada. Sus aplicaciones abarcan desde tareas básicas como la limpieza, organización y clasificación de datos, hasta procesos avanzados como la detección de patrones, el análisis predictivo y la generación automática de contenido. Este apartado aborda los principales enfoques, características y tipos de inteligencia artificial que fortalecen la gestión y el análisis de datos, con especial atención a las capacidades y desafíos actuales de la inteligencia artificial generativa.

3.1. Concepto, origen y evolución de la inteligencia artificial

La inteligencia artificial (IA) es una rama de la informática que estudia y desarrolla sistemas capaces de realizar tareas que, tradicionalmente, requieren inteligencia humana, como el razonamiento lógico, la resolución de problemas, el reconocimiento de patrones, el aprendizaje y la toma de decisiones. Su evolución ha sido continua y se puede dividir en etapas que reflejan avances científicos, tecnológicos y filosóficos significativos.

a) Década de 1950 Los orígenes

El término “inteligencia artificial” fue acuñado en 1956 durante la Conferencia de Dartmouth, organizada por John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester, considerada el punto de partida oficial del campo. Previamente, Alan Turing había propuesto en 1950 el célebre “Test de Turing” como criterio para determinar si una máquina puede exhibir un comportamiento inteligente indistinguible del humano.

En esta etapa se desarrollan los primeros programas simbólicos como Logic Theorist (1955) y General Problem Solver (1957).

b) Décadas de 1960 y 1970 Primeros logros y limitaciones

Se crean los primeros lenguajes de programación orientados a IA, como LISP (1958), y surgen programas capaces de demostrar teoremas y jugar ajedrez. Sin embargo, las expectativas iniciales no se cumplieron del todo, lo que provocó una disminución del financiamiento y la entrada en el llamado “invierno de la IA”.

c) Décadas de 1980 Auge de los sistemas expertos

La IA resurge con fuerza gracias al desarrollo de los sistemas expertos, programas capaces de emular el razonamiento humano en contextos específicos, como el diagnóstico médico (ej. MYCIN) o la ingeniería (ej. XCON de Digital Equipment Corporation). Se refuerzan conceptos como las redes semánticas, los motores de inferencia y las bases de conocimiento.

d) Décadas de 1990 y 2000 Expansión del aprendizaje automático

El avance del machine learning y el aumento del poder computacional propician una nueva era. Se perfeccionan algoritmos como las máquinas de soporte vectorial (SVM), los árboles de decisión y las primeras redes neuronales profundas. Un hito destacado fue en 1997, cuando Deep Blue (IBM) derrotó al campeón mundial de ajedrez Garry Kasparov.

e) Desde 2010 hasta la actualidad El auge del (deep learning) y la IA generativa

El desarrollo de modelos basados en aprendizaje profundo (deep learning), impulsado por redes neuronales convolucionales y grandes volúmenes de datos (big data), lleva la IA a nuevos niveles de rendimiento. Destacan logros como AlphaGo (DeepMind, 2016), capaz de vencer a jugadores profesionales del complejo juego Go, y el surgimiento de modelos generativos como GPT (OpenAI), BERT (Google) o DALL·E, capaces de producir texto, imágenes y sonido con altos niveles de coherencia y creatividad. Estas herramientas redefinen la relación entre humanos y máquinas, con aplicaciones en educación, salud, arte y ciencia.

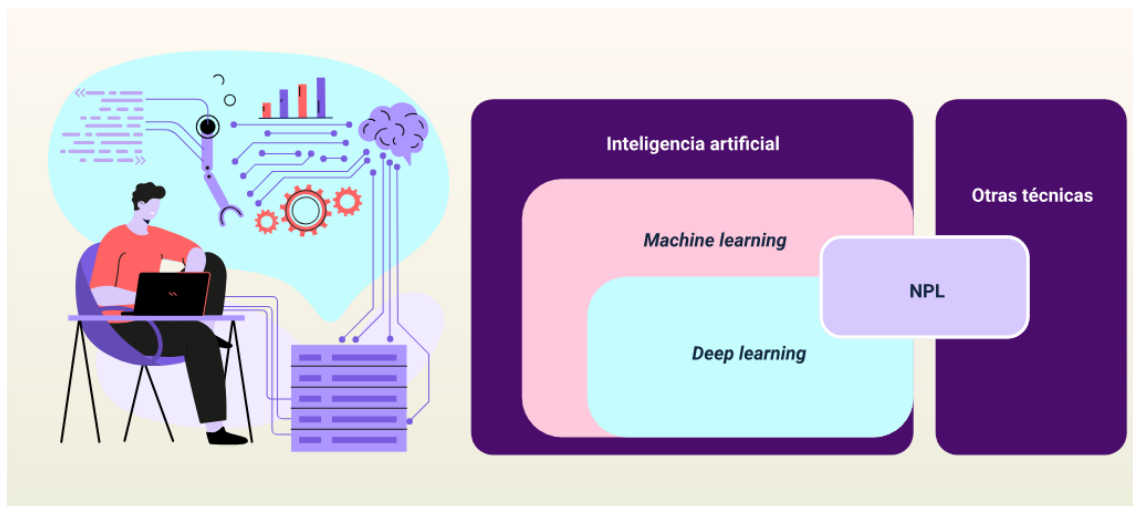
3.2. Funciones y aplicaciones de la IA en el procesamiento de datos

La Inteligencia Artificial (IA) es una rama de la informática que busca desarrollar sistemas capaces de simular el pensamiento humano y aprender a partir de datos. En el contexto del análisis de datos, la IA permite procesar grandes volúmenes de información, identificar patrones, hacer predicciones y tomar decisiones automatizadas, contribuyendo a la eficiencia y precisión en diversos procesos.

Dentro de la IA se encuentran subcampos como el machine learning (aprendizaje automático) y el deep learning (aprendizaje profundo), los cuales permiten entrenar algoritmos con datos para mejorar su desempeño sin necesidad de programación explícita. Además, técnicas como el procesamiento de lenguaje natural (NLP, por sus siglas en inglés) posibilitan la interpretación y generación de lenguaje humano, ampliando las aplicaciones de la IA al análisis de textos, conversaciones y documentos.

A continuación, se presenta la relación jerárquica entre algunos componentes clave de la inteligencia artificial:

Figura 3. Áreas de la inteligencia artificial



Nota. Adaptado de Benoit Cayla (2024).

Esta figura describe cómo el machine learning y el deep learning forman parte del campo más amplio de la inteligencia artificial, junto con otras técnicas especializadas como el NLP (Natural Language Processing).

La IA desempeña un papel fundamental en el ciclo de vida de los datos, desde su recolección y limpieza hasta su análisis e interpretación. Sus capacidades automatizadas permiten optimizar tareas complejas, reducir errores y extraer información valiosa de fuentes estructuradas y no estructuradas. Entre sus funciones más destacadas se encuentran:

- ✓ **Automatización del análisis:** los algoritmos de IA procesan rápidamente millones de registros, detectando patrones, anomalías y relaciones que serían difíciles de identificar de forma manual.
- ✓ **Limpieza y transformación de datos:** la IA identifica errores, elimina duplicados, completa datos faltantes y normaliza variables, mejorando la calidad del conjunto de datos.
- ✓ **Clasificación y segmentación:** mediante modelos supervisados y no supervisados como árboles de decisión o técnicas de clustering, la IA agrupa elementos y clasifica información según características comunes.
- ✓ **Predicción y prescripción:** con base en datos históricos, la IA puede prever comportamientos futuros y recomendar acciones específicas para mejorar procesos y resultados.
- ✓ **Procesamiento de datos no estructurados:** gracias a tecnologías como NLP y visión por computador, la IA puede analizar texto, audio, imágenes y videos, facilitando la gestión de datos complejos.

La inteligencia artificial ha demostrado ser una herramienta versátil, aplicable a diversos sectores económicos y sociales. Sus soluciones permiten automatizar tareas, optimizar recursos y apoyar la toma de decisiones basada en datos. Algunas de sus aplicaciones más representativas incluyen:

- ✓ **Empresas:** optimización de procesos internos, segmentación de clientes, análisis predictivo y toma de decisiones basada en datos.
- ✓ **Gobierno:** monitoreo de servicios públicos, análisis de políticas, prevención del delito y mejora en la atención ciudadana.

- ✓ **Salud:** diagnóstico asistido por IA, análisis de historias clínicas, predicción de brotes epidemiológicos y personalización del tratamiento médico.

Actualmente, existen áreas específicas en las que la IA ha alcanzado un desarrollo significativo, impactando de manera directa distintos campos del conocimiento y la industria. Estos ámbitos concentran importantes avances tecnológicos e investigaciones:

- a) **Apoyo a la toma de decisiones:** más allá de la automatización tradicional, la IA permite tomar decisiones en entornos inciertos y dinámicos. Por ejemplo, en el sector salud, sistemas inteligentes analizan bases de datos clínicas estructuradas y no estructuradas para apoyar diagnósticos médicos con mayor precisión. En el comercio electrónico, los motores de recomendación personalizan ofertas según el comportamiento del usuario, optimizando la experiencia del cliente.
- b) **Visión por Computadora (Computer Vision):** gracias al deep learning y al avance de la capacidad computacional, las máquinas pueden interpretar imágenes y videos con una precisión similar a la humana. Algunas de sus aplicaciones más relevantes son:
 - ✓ Clasificación de objetos.
 - ✓ Detección y localización en imágenes.
 - ✓ Reconocimiento facial.
 - ✓ Reconocimiento óptico de caracteres (OCR).
 - ✓ Procesamiento inteligente de documentos (IDP).
 - ✓ Segmentación y reconstrucción de imágenes.

- ✓ Análisis médico por imágenes y gestión urbana (monitoreo del tráfico y lectura de matrículas).

c) **Procesamiento del Lenguaje Natural (NLP):** permite a las máquinas comprender, interpretar y generar lenguaje humano. Esta disciplina combina análisis fonético, léxico, semántico, sintáctico y prosódico para lograr una interacción fluida con los usuarios. Entre sus aplicaciones más comunes se encuentran:

- ✓ Asistentes virtuales.
- ✓ Traducción automática.
- ✓ Clasificación de textos.
- ✓ Análisis de sentimientos.
- ✓ Generación de contenido automatizado.

3.3. Tipologías de IA: descriptiva, predictiva y generativa

La inteligencia artificial puede clasificarse en función del tipo de análisis que realiza y del valor que aporta a la gestión de los datos. Las tres principales tipologías son la inteligencia artificial descriptiva, predictiva y generativa. Cada una de ellas cumple funciones específicas y tiene aplicaciones en distintos sectores productivos.

- a) **IA descriptiva:** analiza datos históricos y actuales para identificar patrones, tendencias y relaciones. Su propósito es responder a la pregunta “¿qué ha pasado?”.
- ✓ **Ejemplo:** los tableros de control en herramientas como Power BI o Tableau que muestran métricas de rendimiento de una empresa en tiempo real.

- b) **IA predictiva:** utiliza modelos estadísticos y de aprendizaje automático para anticipar lo que podría suceder, con base en datos previos. Responde a la pregunta “¿qué podría pasar?”.
 - ✓ **Ejemplo:** un sistema bancario que predice el riesgo de impago de un cliente o una app de transporte que estima el tiempo de llegada según el tráfico y la ruta.
- c) **IA generativa:** es capaz de crear contenido nuevo a partir de grandes volúmenes de datos. No se limita a analizar o predecir, sino que genera textos, imágenes, música o código. Responde a “¿qué puedo crear?”.
 - ✓ **Ejemplo:** herramientas como ChatGPT que redactan textos o responden preguntas, y plataformas como DALL-E que crean imágenes originales a partir de descripciones escritas.

En muchos entornos digitales, estas tipologías se complementan. Por ejemplo, una plataforma de comercio electrónico puede describir el comportamiento de los usuarios (descriptiva), predecir qué productos comprarán (predictiva) y generar automáticamente anuncios personalizados (generativa).

3.4. IA generativa: características, usos y retos

La inteligencia artificial generativa se refiere a un tipo de IA capaz de crear contenido nuevo y original a partir de datos existentes. A diferencia de los modelos tradicionales que se centran en clasificar o predecir, los modelos generativos imitan patrones complejos para producir textos, imágenes, sonidos, videos, código y otros formatos.

a) Características principales de la IA generativa

- ✓ **Creatividad algorítmica:** utiliza modelos entrenados con grandes volúmenes de datos para generar contenido que simula la creatividad humana.
- ✓ **Aprendizaje profundo:** se basa en redes neuronales profundas, especialmente en arquitecturas como los transformers, que permiten entender contextos complejos.
- ✓ **Interacción con lenguaje natural:** acepta instrucciones o prompts en lenguaje humano, facilitando su uso por personas sin conocimientos técnicos.
- ✓ **Generalización:** puede producir resultados que no están literalmente en los datos de entrenamiento, sino que combinan elementos para formar algo nuevo.

b) Usos actuales de la IA generativa

- ✓ **Comunicación:** redacción automática de textos, generación de resúmenes, traducción de idiomas, chatbots conversacionales.
- ✓ **Diseño e imagen:** creación de ilustraciones, logotipos, renders o imágenes a partir de descripciones.
- ✓ **Educación:** generación de actividades, retroalimentación automática o explicaciones personalizadas para estudiantes.
- ✓ **Entretenimiento:** producción de guiones, música, personajes o mundos virtuales.
- ✓ **Programación:** asistentes que generan código a partir de instrucciones en lenguaje natural.

c) Retos y desafíos

- ✓ **Ética y desinformación:** el contenido generado puede ser indistinguible del creado por humanos, lo que plantea riesgos relacionados con noticias falsas, plagio o suplantación.
- ✓ **Propiedad intelectual:** la autoría y los derechos sobre lo generado por IA son aún objeto de debate legal.
- ✓ **Sesgos algorítmicos:** los modelos pueden reproducir estereotipos o prejuicios presentes en los datos con los que fueron entrenados.
- ✓ **Uso responsable:** se requiere desarrollar criterios y marcos regulatorios que garanticen un uso justo, inclusivo y transparente.

A medida que la inteligencia artificial generativa evoluciona, surgen múltiples herramientas que permiten materializar su potencial en distintos campos. Estas soluciones, desarrolladas por empresas tecnológicas líderes, ya se utilizan ampliamente en entornos creativos, educativos, profesionales e industriales. A continuación, se presentan algunos ejemplos representativos de estas herramientas, organizados según el tipo de contenido que generan:

a) Generación de texto

- ✓ Chat GPT (OpenAI).
- ✓ Claude (Anthropic).
- ✓ Gemini (Google).
- ✓ Copilot (Microsoft).

b) Generación de imágenes

- ✓ DALL- E (OpenAI).
- ✓ Midjourney.

- ✓ Stable Disffusion.

c) Generación de música y audio

- ✓ Suno AI.
- ✓ AIVA.
- ✓ Voicemod.

d) Generación de video

- ✓ Sora (OpenAI).
- ✓ Runway Gén-2.
- ✓ Pika.

e) Asistentes de Código

- ✓ GitHub Copilot.
- ✓ CodeWhisperer (Amazon).
- ✓ Codey (Google).

Estos ejemplos presentan cómo la IA generativa ya está integrada en múltiples industrias, aportando eficiencia, creatividad y nuevas formas de interacción entre humanos y máquinas.

3.5. Interacción con modelos generativos: el papel del prompt

En el campo de la inteligencia artificial generativa, un prompt es la entrada o instrucción que se le proporciona a un modelo para obtener una respuesta específica. Puede ser una pregunta, una frase, una descripción o incluso una serie de directrices detalladas. Es el punto de partida desde el cual la IA genera contenido textual, visual, sonoro o de otro tipo.

Cuanto más clara y específica sea la instrucción, mayor será la calidad y relevancia de la respuesta. Por ejemplo, un prompt genérico como “Escribe sobre el reciclaje” generará una salida básica, mientras que una solicitud más estructurada como “Redacta un artículo de 300 palabras sobre los beneficios del reciclaje en zonas urbanas de América Latina, dirigido a estudiantes de secundaria”, producirá una respuesta más precisa y contextualizada.

Para comprender mejor cómo influye un prompt bien formulado en la calidad de las respuestas generadas por la IA, se presenta a continuación una situación cotidiana. A través de este ejemplo, se evidencia cómo la claridad, el contexto y los detalles proporcionados pueden transformar una solicitud simple en una solución personalizada y útil.

Era una tarde de marzo y Laura, madre de dos niños pequeños y apasionada por los viajes, se sentó frente a su computador con un reto: organizar unas vacaciones memorables para junio de 2024. Cansada de intentos fallidos, decidió probar algo nuevo. Abrió ChatGPT y escribió:

“Actúa como un experto en viajes”. Luego especificó: “¿Me puedes preparar un itinerario personalizado para nuestras vacaciones?”.

Pero sabía que la clave estaba en los detalles, así que agregó:

“Queremos viajar a un país asiático durante 21 días. No más de cuatro días por ciudad. Buscamos playas y metrópolis, evitar aglomeraciones, buen precio y al menos un parque acuático. Nuestro presupuesto es de 6.000 euros para dos adultos y dos niños. Queremos una agenda día por día”.

En segundos, la IA propuso un itinerario adaptado a su ritmo, presupuesto e intereses. Laura comprendió entonces que un buen prompt convierte a la inteligencia artificial en una verdadera asistente personal.

Para optimizar los resultados con herramientas generativas, se recomienda incluir los siguientes elementos en la redacción del prompt:

- 1) Definir el tono: académico, informal, técnico, etc.
- 2) Establecer restricciones claras: extensión, formato y estilo.
- 3) Especificar idioma y público objetivo.
- 4) Indicar el propósito de la respuesta.
- 5) Asignar un rol a la IA (experto, asesor, crítico, etc.).
- 6) Proporcionar contexto, ejemplos o antecedentes.
- 7) Pedir balance de opiniones, referencias, datos estadísticos o elementos visuales, si es necesario.

Los prompts pueden adaptarse a múltiples propósitos según las necesidades del usuario. A continuación, se presentan ejemplos orientados a diversos contextos, como la generación de textos, la solución de problemas, la planificación de tareas o la creación de contenidos visuales, mostrando cómo la redacción adecuada influye en la calidad de las respuestas.

Tabla 3. Ejemplos de prompt

Tipo de contenido	Prompt ejemplo	Objetivo del usuario
Texto (ChatGPT)	Actúa como sociólogo. Explica en 200 palabras	Obtener un texto argumentativo académico.

Tipo de contenido	Prompt ejemplo	Objetivo del usuario
	cómo influye la tecnología en la interacción social.	
Imagen (DALL·E, Midjourney)	Ilustra una biblioteca futurista en medio del desierto, con arquitectura orgánica y cielo estrellado.	Generar una imagen conceptual creativa.
Código (ChatGPT, Copilot)	Crea una función en Python que convierta grados Celsius a Fahrenheit con validación de entrada.	Resolver una tarea de programación.
Música (Soundraw, Amper)	Compón una pista ambiental con sonidos naturales y ritmo suave para usar en sesiones de meditación.	Generar música personalizada.
Video (Sora, Synthesia)	Crea un video animado con voz en off explicando qué es la IA generativa, en menos de 90 segundos.	Crear un recurso educativo visual.

Estos ejemplos ilustran cómo un prompt bien formulado permite aprovechar al máximo las capacidades de los modelos generativos, transformando simples solicitudes en soluciones complejas, creativas y útiles.

4. Análisis y visualización de datos

La visualización y el análisis de datos son procesos fundamentales en la toma de decisiones basada en evidencia. Permiten identificar patrones, relaciones y tendencias ocultas dentro de grandes volúmenes de información. A través del uso de herramientas digitales, es posible representar los datos de forma gráfica e interactiva, facilitando su comprensión y comunicación. Al combinar la visualización con técnicas de análisis, se obtiene una visión más profunda que puede aplicarse en múltiples campos, desde la gestión empresarial hasta la investigación científica. Este apartado introduce herramientas clave de visualización y analítica, así como una primera aproximación al aprendizaje automático como extensión del análisis avanzado

4.1. Herramientas de visualización de datos

La visualización de datos consiste en representar gráficamente información compleja para facilitar su interpretación. Este proceso convierte cifras y registros en gráficos, mapas, tablas dinámicas o dashboards interactivos, que permiten detectar patrones, comparaciones y comportamientos de manera intuitiva. Las herramientas de visualización no solo ayudan a comunicar hallazgos de forma clara, sino que también son clave para la exploración inicial de los datos y la toma de decisiones estratégicas.

Entre las herramientas más utilizadas se encuentran:

- ✓ **Microsoft Power BI:** permite crear informes interactivos, integrarse con múltiples fuentes de datos y compartir visualizaciones en la nube.

- ✓ **Tableau**: se destaca por su capacidad de crear dashboards dinámicos, con una interfaz intuitiva que facilita el análisis visual.
- ✓ **Google Data Studio (ahora Looker Studio)**: una solución gratuita y accesible que permite generar informes personalizados a partir de datos de hojas de cálculo, Google Analytics, entre otros.
- ✓ **Excel**: aunque tradicional, sigue siendo útil para representaciones básicas como gráficos de barras, líneas o sectores, especialmente en entornos no especializados.
- ✓ **Plotly, D3.js y Matplotlib**: bibliotecas de programación que permiten desarrollar visualizaciones más avanzadas y personalizadas, especialmente en entornos de análisis con Python o JavaScript.

Estas herramientas ayudan a convertir datos en conocimiento, promoviendo una cultura de toma de decisiones informada y basada en evidencias.

4.2. Herramientas de analítica de datos

La analítica de datos se refiere al conjunto de técnicas y herramientas utilizadas para examinar, procesar e interpretar grandes volúmenes de información con el fin de obtener conclusiones útiles. Este proceso abarca desde el análisis descriptivo (que resume lo que ocurrió), hasta el análisis predictivo (que anticipa tendencias futuras) y el análisis prescriptivo (que sugiere acciones óptimas basadas en los datos).

Entre las herramientas más representativas se encuentran:

- 1) **Excel con Power Query y Power Pivot**: utilizadas para análisis exploratorios, combinación de fuentes de datos y creación de modelos simples.

- 2) **Python y sus bibliotecas (Pandas, NumPy y Scikit-learn)**: ampliamente empleadas en la analítica avanzada y el desarrollo de modelos predictivos y algoritmos de machine learning.
- 3) **R**: un lenguaje especializado en estadística, ideal para análisis profundos y visualización avanzada de datos.
- 4) **Google BigQuery**: una solución en la nube que permite el análisis de grandes volúmenes de datos con alto rendimiento.
- 5) **Apache Hadoop y Apache Spark**: diseñadas para procesar datos masivos distribuidos, son comunes en entornos de big data.
- 6) **RapidMiner y KNIME**: plataformas que permiten realizar analítica predictiva sin necesidad de programación avanzada, a través de interfaces visuales.

Estas herramientas permiten a las organizaciones y profesionales comprender mejor sus datos, mejorar procesos, identificar oportunidades y optimizar recursos mediante decisiones basadas en evidencia cuantitativa.

4.3. Introducción al aprendizaje automático (machine learning)

El aprendizaje automático, o machine learning, es una disciplina dentro de la inteligencia artificial que se enfoca en desarrollar sistemas capaces de aprender de los datos y mejorar su desempeño con el tiempo sin necesidad de programación explícita. A diferencia de los algoritmos tradicionales que siguen instrucciones fijas, los modelos de machine learning detectan patrones, hacen inferencias y toman decisiones a partir de grandes volúmenes de datos, lo que los hace altamente adaptables a tareas complejas y cambiantes.

Este enfoque se clasifica principalmente en tres tipos:

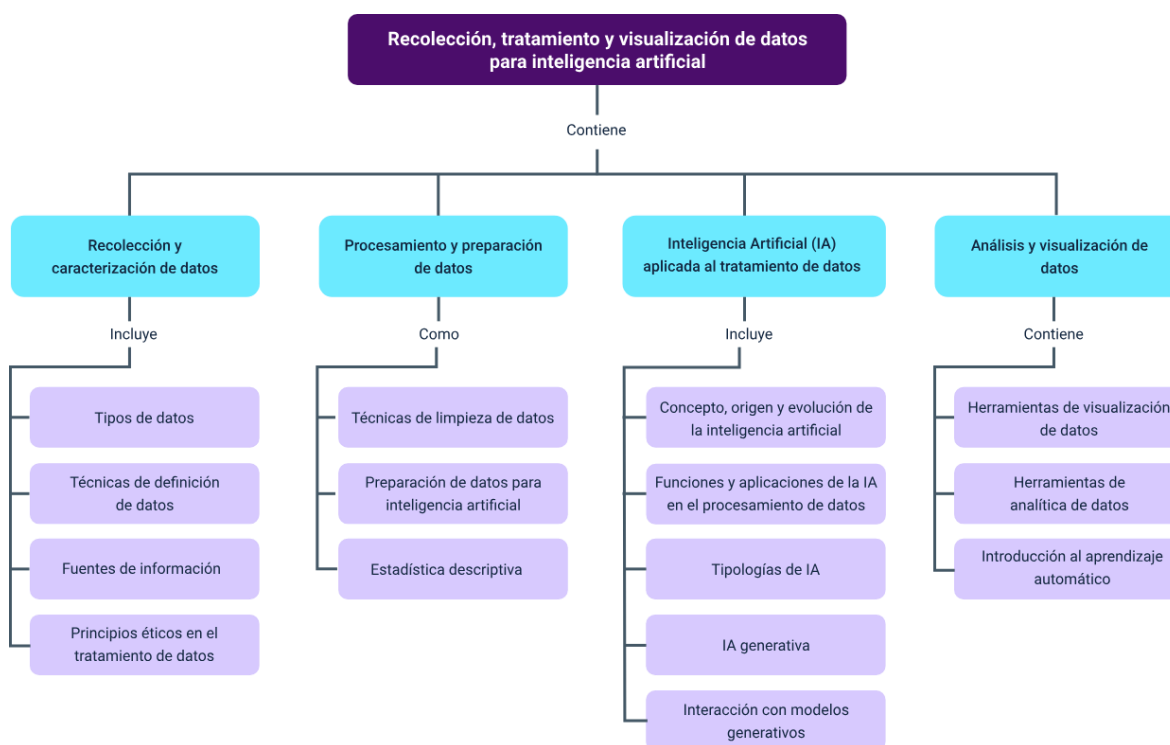
- a) Aprendizaje supervisado, en el que se entrena un modelo con datos etiquetados, es decir, datos de entrada con sus respectivas salidas esperadas, para que pueda predecir resultados similares con nuevos datos.
- b) Aprendizaje no supervisado, que se utiliza cuando los datos no tienen etiquetas y el sistema busca estructuras o agrupaciones dentro de ellos, como patrones ocultos o segmentos similares.
- c) Aprendizaje por refuerzo, donde el modelo aprende a través de la interacción con su entorno, recibiendo recompensas o penalizaciones en función de sus decisiones, lo que refuerza comportamientos óptimos con el tiempo.

El machine learning es una tecnología clave en muchas aplicaciones actuales, como los motores de recomendación de plataformas de streaming, la detección de fraudes financieros, el diagnóstico médico asistido por IA, los asistentes virtuales, y los sistemas de predicción de demanda. Además, es la base de desarrollos más avanzados como el reconocimiento facial, los vehículos autónomos y los modelos generativos actuales.

A medida que el volumen de datos crece, el aprendizaje automático se convierte en una herramienta esencial para convertir esa información en conocimiento útil y decisiones automatizadas que optimicen procesos en prácticamente todos los sectores económicos y científicos.

Síntesis

Este componente formativo aborda los fundamentos clave de la inteligencia artificial, su evolución y aplicaciones actuales, con énfasis en la IA generativa y su papel en la transformación digital. Se exploran conceptos como el aprendizaje automático, los tipos de inteligencia artificial, los modelos generativos y la importancia de los prompts para una interacción eficaz con estas herramientas. Asimismo, se introducen estrategias de análisis, visualización de datos y el uso de herramientas tecnológicas que permiten interpretar grandes volúmenes de información, preparando a los aprendices para comprender, aplicar y adaptarse a los avances tecnológicos en entornos personales, académicos y laborales.



Material Complementario

Tema	Referencia	Tipo de material	Enlace del recurso
1. Recolección y caracterización de datos	Ecosistema de Recursos Educativos Digitales SENA. (2022). Datos e información [Video]. YouTube.	Video	https://www.youtube.com/watch?v=2GkGn3Mei6o
2.3. Estadística descriptiva	Ecosistema de Recursos Educativos Digitales SENA. (2023). Estadística descriptiva, gráficas e informes estadísticos [Video]. YouTube.	Video	https://www.youtube.com/watch?v=v5UMIXHe2nM
3. Inteligencia Artificial (IA) aplicada al tratamiento de datos	Ecosistema de Recursos Educativos Digitales SENA. (2023). Inteligencia artificial en los datos [Video]. YouTube.	Video	https://www.youtube.com/watch?v=-hYXrGAUYAE
4.3. Introducción al aprendizaje automático	Ecosistema de Recursos Educativos Digitales SENA. (2023). Machine Learning con Python [Video]. YouTube.	Video	https://www.youtube.com/watch?v=noMy4-zjR9Q

Tema	Referencia	Tipo de material	Enlace del recurso
(machine learning)			

Glosario

Algoritmo: conjunto de reglas o instrucciones definidas para realizar una tarea o resolver un problema de manera sistemática, especialmente en programación y análisis de datos.

Analítica de datos: conjunto de técnicas que permiten examinar y extraer conclusiones significativas a partir de datos. Puede ser descriptiva, predictiva o prescriptiva.

Datos estructurados: información organizada en formatos predefinidos como tablas, filas y columnas, que facilita su almacenamiento y análisis.

Datos no estructurados: información sin un formato predefinido, como imágenes, videos, audios o texto libre, que requiere técnicas especializadas para su procesamiento.

IA generativa: tecnología basada en inteligencia artificial capaz de crear contenido nuevo (texto, audio, imágenes, etc.) a partir de datos y patrones aprendidos.

Limpieza de datos: proceso de detección y corrección de errores, valores faltantes o inconsistencias en un conjunto de datos para asegurar su calidad.

Machine learning (aprendizaje automático): rama de la inteligencia artificial que permite a las máquinas aprender a partir de datos y mejorar su desempeño sin intervención humana directa.

Prompt: instrucción o entrada que se da a un modelo de inteligencia artificial generativa para que produzca una salida específica (texto, imagen, código, etc.).

Visualización de datos: representación gráfica de datos mediante gráficos, mapas o diagramas que permite identificar patrones, tendencias y anomalías.

Referencias bibliográficas

Cayla, B. (2024). El dato: Guía de supervivencia en el mundo de los datos. Editorial ENI.

Chollet, F. (2021). Deep learning con Python (2.ª ed.). Ediciones Anaya Multimedia.

Chollet, F. (2022). Deep learning con Python (2.ª ed., J. Barceló, Trad.). Anaya Multimedia. (Obra original publicada en 2021)

Cuatrecasas. (2024). Guía para el uso responsable de IA y datos personales en Colombia. <https://www.cuatrecasas.com/es/spain/tecnologia-medios-digitales/art/guia-uso-responsable-ia-datos-personales-colombia>

De Prado, M. (2023). Inteligencia artificial y análisis de datos: Fundamentos y aplicaciones. Ediciones Pirámide.

Fernández-Macías, E., & Sánchez, Á. (2022). El impacto de la inteligencia artificial en el empleo y el trabajo. Fundación Cotec.

García, E., & López, F. (2023). Inteligencia artificial: Fundamentos, ética y aplicaciones. Alfaomega.

Goodfellow, I., Bengio, Y., & Courville, A. (2022). Deep learning: Introducción práctica con Python (2.ª ed., J. Cordero, Trad.). Marcombo.

González, M. A. (2023). Inteligencia artificial generativa: Fundamentos, herramientas y aplicaciones prácticas. Editorial Alfaomega.

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2022). Estadística aplicada a los negocios y la economía (17.^a ed.). Cengage Learning.

Maldonado, G. (2021). Estadística aplicada a las ciencias sociales y administrativas. Ediciones de la U.

Provost, F., & Fawcett, T. (2020). Data science para los negocios: Lo que necesitas saber sobre minería de datos y análisis predictivo. O'Reilly Media.

Rodríguez, M. (2025). Los datos son como piezas de LEGO [Publicación en LinkedIn]. https://www.linkedin.com/posts/manuel-rodriguez-1a1703242_los-datos-son-como-piezas-de-lego-activity-7274364694881054720-fJnz/?originalSubdomain=es

Suárez, E. (2024). Guía práctica sobre los distintos tipos de fuentes de información. Experto Universitario. <https://expertouniversitario.es/blog/fuentes-de-informacion/>

Créditos

Nombre	Cargo	Centro de Formación y Regional
Milady Tatiana Villamil Castellanos	Responsable Ecosistema de Recursos Educativos Digitales (RED)	Dirección General
Diana Rocío Possos Beltrán	Responsable de línea de producción	Centro de Comercio y Servicios - Regional Tolima
Javier Eduardo Díaz Machuca	Experto temático	Centro de Comercio y Servicios - Regional Tolima
Viviana Esperanza Herrera Quiñonez	Evaluadora instruccional	Centro de Comercio y Servicios - Regional Tolima
Oscar Ivan Uribe Ortiz	Diseñador web	Centro de Comercio y Servicios - Regional Tolima
Jose Yobani Penagos Mora	Diseñador web	Centro de Comercio y Servicios - Regional Tolima
Sebastian Trujillo Afanador	Desarrollador full stack	Centro de Comercio y Servicios - Regional Tolima
Gilberto Junior Rodríguez Rodríguez	Animador y productor audiovisual	Centro de Comercio y Servicios - Regional Tolima
Jorge Eduardo Rueda Peña	Evaluador de contenidos inclusivos y accesibles	Centro de Comercio y Servicios - Regional Tolima

Nombre	Cargo	Centro de Formación y Regional
Jorge Bustos Gómez	Validador y vinculator de recursos educativos digitales	Centro de Comercio y Servicios - Regional Tolima