UNIDAD DIDÁCTICA 5

EVALUACIÓN Y PROTECCIÓN DE DATOS

MÓDULO PROFESIONAL: DIGITALIZACIÓN APLICADA A LOS SECTORES PRODUCTIVOS (GS)

Índice

RESUMEN INTRODUCTORIO

En esta unidad se abordará la distinción entre datos e información, destacando que los datos son hechos objetivos y que la información se genera a partir de su interpretación. Se explorará el ciclo de vida de los datos, desde su creación hasta su eliminación, enfatizando que su gestión adecuada es esencial para garantizar la calidad de la información.

Además, se enfatizará la ciencia de datos como herramienta clave para identificar patrones y mejorar la toma de decisiones en las organizaciones, integrando técnicas como inteligencia artificial y aprendizaje automático. También se señalará la importancia de la ciberseguridad en la protección de datos sensibles y la responsabilidad de las empresas en implementar medidas adecuadas para salvaguardar la información.

Asimismo, se resaltarán los beneficios que la ciencia de datos puede ofrecer a las organizaciones, incluyendo la mitigación de riesgos, la personalización de productos, la optimización de la experiencia del cliente y la automatización de procesos.

Por último, se abordará la importancia de aplicar estos conceptos de acuerdo con la legislación vigente para que las empresas puedan aprovechar al máximo las oportunidades que presenta la era de la información.

INTRODUCCIÓN

En la era digital actual, donde la información fluye a gran velocidad y las organizaciones se ven inundadas por una creciente cantidad de datos, la gestión eficaz de esta información se ha convertido en un pilar fundamental para el éxito y la competitividad empresarial. La comprensión de los conceptos de “dato” e “información” es esencial para navegar en este vasto entorno.

Comprender las fases del ciclo de vida de los datos es vital para garantizar su calidad, accesibilidad y seguridad. La ciencia de datos se presenta como una disciplina clave, integrando técnicas avanzadas para extraer patrones y tendencias valiosas de grandes volúmenes de datos. Elementos como la inteligencia artificial, el aprendizaje automático y el aprendizaje profundo son fundamentales para transformar la operación de las empresas, ofreciendo soluciones innovadoras para problemas complejos.

La ciberseguridad también desempeña un papel crítico en la protección de datos sensibles. Las organizaciones deben implementar estrategias robustas para salvaguardar la información de sus clientes y cumplir con las normativas vigentes. Esto

asegura la integridad de la información y fomenta la confianza entre empresas y clientes. Es importante que los futuros profesionales exploren estos conceptos, analizando su interrelación y ofreciendo una visión integral sobre la importancia de la ciencia de datos y la seguridad en un entorno empresarial cada vez más complejo.

CASO INTRODUCTORIO

Trabajas en una empresa que ha decidido implementar inteligencia artificial y tecnologías de análisis de datos para mejorar su eficiencia operativa, adaptarse a las demandas del mercado y optimizar el uso de sus recursos. La compañía maneja una gran cantidad de datos que provienen de sus transacciones diarias, inventarios y plataformas digitales, pero hasta ahora no han sido explotados completamente para generar valor. La dirección está interesada en automatizar tareas clave, realizar análisis predictivos para mejorar la planificación de inventarios y personalizar la experiencia del cliente a través de herramientas digitales.

Además de estos objetivos, la empresa está comprometida con la protección de los datos personales de sus clientes y empleados. A medida que se implementan soluciones basadas en inteligencia artificial y análisis de grandes volúmenes de datos, surgen preocupaciones sobre la seguridad de la información y el cumplimiento de las normativas de protección de datos. Tu equipo ha sido asignado para liderar este proyecto, investigando las aplicaciones de la inteligencia artificial que puedan mejorar los procesos empresariales y, al mismo tiempo, asegurando que se cumplan las normativas de seguridad y privacidad de los datos.

Al finalizar esta unidad, serás capaz de identificar las aplicaciones más relevantes de la inteligencia artificial y el análisis de datos para optimizar los procesos en la empresa. Además, comprenderás la importancia del manejo adecuado de los datos a lo largo de su ciclo de vida, aplicando soluciones de IA de manera ética y conforme a las normativas de protección de datos y ciberseguridad.

1. DATO VS INFORMACIÓN

En tu empresa se está implementando un sistema de inteligencia artificial para optimizar la gestión de inventarios y mejorar la eficiencia operativa. Como primer paso, trabajas con tus compañeros para comprender cómo se transforman los datos recopilados sobre ventas y niveles de stock en información útil que facilite la toma de decisiones. Tenéis que identificar qué datos son relevantes y organizarlos de manera que puedan ser aprovechados por el sistema de IA. Este proceso te permite observar cómo la IA puede adaptarse a las necesidades de la empresa y apoyar la operación sin interrumpir el flujo de trabajo diario.

Aunque en el campo del conocimiento es frecuente el uso de los términos "dato" e "información" como si fueran sinónimos, en realidad representan ideas distintas y cumplen funciones diferentes en el análisis y la comprensión. Para un mejor entendimiento, a continuación, se explicará su significado y la importancia de su diferenciación para una óptima interpretación.

En primer lugar, el término **dato** hace referencia a un conjunto definido de factores objetivos relacionados con un hecho real. No obstante, aunque este concepto alude a la **documentación de operaciones**, por sí solo **no proporciona información** sobre las causas o el motivo detrás de los eventos, y carece de relevancia o propósito significativo. Por sí mismos, no guían la toma de decisiones ni indican qué acciones tomar. Sin embargo, los datos constan como un fundamento basal para generar información más significativa y útil.

Por otro lado, nos encontramos con el término de **información**, que se puede entender como un mensaje generalmente presentado a modo de documento o como una **comunicación** que podemos escuchar o ver. Por su parte, a diferencia del concepto anterior, este sí **cuenta con significado** dotado de relevancia para el receptor y propósito organizado.

Una vez tenemos delimitadas ambas nociones y habiendo entendido que son muy diferentes entre sí, es el momento de conocer por qué acostumbramos a relacionarlos y confundirlos. Como ya se ha comentado, **los datos constan como el principio esencial que, tras su unificación, da lugar a la información completa**. Dicha transformación es el producto de diferentes métodos que tratan de asegurar que comprendemos para qué se recopilaron los datos, cómo se organizan, cómo se han evaluado numéricamente, cómo se han corregido posibles errores y cómo se han presentado de manera más compacta para facilitar su comprensión.

Analizando la información obtenida a través de los datos ofrecidos

Fuente: https://www.freepik.es/foto-gratis/vista-superior-companeros-trabajo-hablando-grafico-barras_854382.htm

2. CICLO DE VIDA DEL DATO

Mientras trabajas en la comprensión de la automatización de la gestión de inventarios, te enfrentas al desafío de analizar el ciclo de vida de los datos para entender cómo estos se convierten en información útil. Cada fase del ciclo, desde la creación hasta el almacenamiento, tiene objetivos que debes conocer para garantizar la seguridad y accesibilidad de la información.

El ciclo de vida de los datos trata de describir las diferentes **fases por las que pasan los datos** a lo largo de su existencia dentro del proceso de organización, donde evolucionan, se transforman y almacenan. Estas etapas se determinan según distintos criterios y se completan a medida que los datos realizan diversas funciones o cumplen ciertos objetivos.

Análisis de datos

Fuente: https://www.freepik.es/foto-gratis/oficinistas-que-usan-graficos-finanzas_42621749.htm

Este ciclo **abarca todo el tiempo en que los datos están activos en una entidad**, desde su creación inicial hasta su eventual eliminación o reutilización en distintos repositorios de investigación. Dicho proceso se denomina "ciclo" porque los conocimientos adquiridos de un proyecto de datos a menudo sirven de fundamento para el siguiente. Así, la fase final del proceso alimenta de nuevo la etapa inicial.

Como se ha dicho, cada fase de este proceso tiene objetivos y propiedades particulares, las cuales son imprescindibles entender para manejar los datos de manera eficaz. Un buen ejemplo de ello son las organizaciones y empresas que, a través del **Big Data**, gestionan y protegen sus datos frente a pérdidas, eliminaciones, ciberataques u otros problemas similares, facilitando que se establezcan directrices sobre cómo manejar, usar, almacenar y compartir sus datos, asegurando el cumplimiento de las normativas

sobre privacidad y reduciendo el riesgo de filtraciones de datos y evitando el uso inapropiado de información sensible. Además, ayuda a **preservar la integridad de los datos** a lo largo de su ciclo de vida, lo que a su vez optimiza los procesos y eleva la eficiencia.

El ciclo de vida en ciencia de datos es una serie de **pasos repetitivos** que se siguen para completar un proyecto o análisis. Se debe tener en cuenta que, aunque cada proyecto y equipo puede tener su propio enfoque particular, la mayoría de ellos tiende a seguir una estructura general similar en su proceso.

VÍDEO DE INTERÉS

Atiende a esta explicación sobre la relación directa que cabe entre los datos y los humanos:

2.1 Análisis de datos

En un primer momento, el proceso comienza con la **obtención de datos** procedentes de una o varias fuentes externas, como mediante la introducción manual de datos, sensores, registros automáticos u operaciones digitales, entre otros.

Acto seguido de dicha obtención, se procede con el **análisis de datos**, que consiste en el procesamiento de identificación de los datos previamente obtenidos para definir su utilidad, así como la obtención de información de interés que se transformará a un formato más accesible, pudiendo ser comprimido o cifrado como una medida de seguridad. En este análisis pueden ser implementados métodos como la **minería de datos, la inteligencia artificial y el aprendizaje automático**, que se dedican a identificar patrones, tendencias y extraer información valiosa.

Para poder completar esta etapa, se debe dar una **interpretación y uso** a los datos analizados. Como pueden ser empleados de diversas formas, es crucial implementar las precauciones adecuadas para prevenir su uso indebido.

2.2 Almacenamiento de datos

Una vez cumplimentada la etapa anterior, es necesario **almacenar y organizar** de un modo seguro los datos recogidos. En este momento del ciclo, los datos se archivan en sistemas de almacenamiento, bases de datos u otros repositorios. Ahí es donde los datos se organizan y procesan para posteriormente ser supervisados de manera continua con el objetivo de asegurar que se mantengan accesibles y optimizados.

ENLACE DE INTERÉS

Accede para saber más sobre la computación en la nube:

Una correcta **estructuración y clasificación** permiten una recuperación eficiente de la información para cuando se requiera. Además, es destacable que, en esta etapa, las tecnologías de **almacenamiento en la nube y las bases de datos** juegan un papel fundamental, ya que, aunque se definen políticas para decidir el período de conservación de los datos según su relevancia y su posible uso futuro, también se realizan copias de seguridad que puedan asegurar la recuperación en caso de pérdida de datos debido a fallos técnicos o incidentes imprevistos.

Por otro lado, los datos pueden ser gestionados como **memoria a corto o largo plazo**. La memoria de acceso aleatorio (RAM) se encarga de procesar la memoria temporal, gestionando las solicitudes y acciones mientras la computadora realiza cálculos específicos (denominados tareas). Después de completar los cálculos, los datos se almacenan a largo plazo en diferentes medios de almacenamiento, como, por ejemplo, en la nube.

Los diferentes tipos de almacenamiento que encontramos en la nube son:

- **Almacenamiento en bloques:** Divide un volumen (como un nodo en la nube) en pequeñas unidades llamadas bloques. Es rápido y de baja latencia, ideal para cargas de trabajo que requieren un alto rendimiento.

- **Almacenamiento de objetos:** Cada unidad de datos se asocia con identificadores únicos llamados metadatos. Dado que los objetos no están comprimidos ni cifrados, se puede acceder a ellos de manera rápida y en grandes cantidades, lo que lo hace perfecto para aplicaciones nativas de la nube.

- **Almacenamiento de archivos:** Utilizado principalmente en sistemas NAS, organiza los datos en una estructura jerárquica, facilitando su exploración desde el inicio hasta el final. Sin embargo, esta estructura puede aumentar el tiempo necesario para procesar los datos.

2.3 Etapas de la ciencia de datos

La ciencia de datos está transformando todos los aspectos de nuestras vidas, abarcando áreas como las finanzas, la educación, la salud, las compras y el deporte. En este marco, los proyectos de ciencia de datos se han vuelto fundamentales para las empresas, ya que proporcionan soluciones a problemas, responden preguntas y ofrecen una visión integral del negocio.

El término como tal surgió por primera vez en la década de 1960 como una forma alternativa de referirse a la **estadística**. Pero no fue hasta finales de los años 90 cuando los expertos en computación comenzaron a establecer el concepto de manera más formal. Una posible definición la valoraba como un campo autónomo que abarca tres elementos: **el diseño, la recopilación y el análisis de datos**. Sin embargo, tuvo que pasar una década más para que el término empezara a emplearse fuera del entorno académico.

Por ello, los analistas de datos desarrollan modelos para prever resultados y descubrir patrones, utilizando una serie de pasos repetitivos que se siguen para completar un proyecto o análisis. Sin embargo, es fundamental reconocer que hay varias etapas preliminares importantes antes de llegar al modelado. Aunque en algunos proyectos se pueden omitir ciertas etapas, la mayoría sigue una secuencia de pasos ampliamente aceptada, que puede variar según las necesidades del proyecto o del equipo.

De este modo, las etapas de la ciencia de datos en forma general son las siguientes:

1. **Definición y entendimiento del problema:** Esta ciencia va más allá de los algoritmos y tecnologías, ya que se centra en entender cómo los datos impactan el negocio y sus implicaciones. Aunque esta fase puede ser rápida si ya se tiene un buen conocimiento del negocio, a menudo requiere una investigación detallada para comprender completamente la situación.

2. **Adquisición y análisis de datos:** Implica identificar dónde se encuentran los datos, comprender su significado y cómo extraerlos, ya sea de fuentes internas, públicas o proveedores externos, además de considerar aspectos como el formato y volumen de los datos.

   Una vez obtenidos, los datos deben ser explorados, limpiados y transformados para su análisis. Los científicos de datos o analistas dirigirán estas actividades para asegurar que los datos sean adecuados y útiles. Este paso es crucial, ya que, sin datos adecuados, el análisis no puede proceder.

3. **Modelización:** Se incluye el análisis exploratorio de datos, la ingeniería de características, el entrenamiento de modelos y la validación. Dado que estos procesos son cíclicos, un ajuste en la creación de una variable puede afectar tanto al entrenamiento como a la validación del modelo. Por ejemplo, si durante el análisis exploratorio se identifica que una variable no aporta valor, ésta debe eliminarse y todo el proceso (análisis, ingeniería, entrenamiento y validación) debe repetirse.

   En esta fase, se construye y evalúa el modelo de aprendizaje automático, asegurando que funcione correctamente a través de pruebas rigurosas. Aunque en las empresas se suele poner mucho énfasis en la modelización, en la práctica, los científicos de datos a menudo dedican más tiempo a las etapas previas. Esto implica probar diferentes modelos, medir su rendimiento, seleccionar el más adecuado, entrenarlo y volver a evaluarlo para garantizar su eficacia.

4. **Despliegue:** El modelo desarrollado se pone en funcionamiento automáticamente y se despliega. Después, se monitoriza su rendimiento en producción. Si surgen problemas, es necesario revisar y ajustar el modelo, a veces regresando a etapas anteriores.

   Esta fase es generalmente manejada por especialistas en ingeniería de datos, ingeniería en la nube, aprendizaje automático y control de calidad. Además, el despliegue permite evaluar el rendimiento real del modelo mediante plataformas como Model Store, servicios web (AWS, Azure, Google Cloud, IBM) y demás aplicaciones inteligentes.

Persona creando datos

Fuente: https://www.freepik.es/foto-gratis/holograma-tecnologico-interiores_330180632.htm

PARA SABER MÁS

Es importante que comprendas la importancia del ciclo de vida de los datos y sus fases:

EJEMPLO PRÁCTICO

Amanda trabaja en una pequeña empresa que maneja una gran cantidad de archivos de clientes. Después de experimentar una pérdida de datos debido a un fallo en el disco duro, propone implementar un sistema de respaldo de datos efectivo.

Amanda investiga y plantea a sus compañeros un servicio de almacenamiento en la nube que ofrece copias de seguridad automáticas y cifrado de datos. Además, sugiere establecer un cronograma de copias de seguridad diarias y expone a sus compañeros la importancia de mantener copias de seguridad y cómo acceder a ellas en caso de emergencia. Como resultado, la empresa no solo protege su información crítica, sino que también mejora la confianza de los clientes al garantizar que sus proyectos se mantendrán seguros y disponibles.

3. BIG DATA Y CLOUD COMPUTING

Tras vuestro análisis, la empresa ha decidido adoptar Cloud Computing y Big Data para mejorar la gestión y protección de los datos. Sabes cómo estas tecnologías permiten almacenar información de forma segura y analizar grandes volúmenes de datos para obtener insights valiosos, por lo que comprendes el impacto de implementar estas herramientas de manera efectiva, destacando la importancia de cumplir con las normativas de privacidad sin interrumpir las operaciones diarias de la empresa.

Hoy en día, muchas personas usan el **Cloud Computing** para almacenar datos en servidores de distintos proveedores, lo que reduce el riesgo de pérdida o robo de información. Los usuarios particulares suelen emplear esta tecnología en sus teléfonos móviles para guardar fotos, datos bancarios y documentos, mientras que las empresas confían en el Cloud Computing para guardar y proteger toda su información y sistemas.

El concepto de Cloud Computing se entiende como un modelo que proporciona acceso a recursos informáticos como **redes, servidores, almacenamiento, aplicaciones y servicios a través de la red**. Estos recursos se pueden configurar, activar y desactivar rápidamente con poco esfuerzo de gestión y sin necesidad de interactuar extensamente con el proveedor del servicio.

Por su parte, el **Big Data** es el método utilizado habitualmente por las empresas para la **producción y recolección de información masiva** que posteriormente se procesa para su mejora.

El uso combinado de Cloud Computing y Big Data ha revelado la forma en que estas tecnologías se complementan y ha permitido comprender sus características y ventajas en conjunto. Analizar cómo el Cloud Computing influye en el almacenamiento de datos y cómo se conecta con la administración de información que ofrece Big Data puede ayudar a las empresas a mejorar la eficiencia de sus recursos y procesos. Esta integración permite una gestión de datos más eficiente y efectiva, optimizando la administración y el desempeño general de la empresa.

BIBLIOGRAFÍA RECOMENDADA

Marr, B. (2016). *Big Data en la práctica: Cómo empresas de grandes éxitos utilizan Big Data para obtener resultados extraordinarios*. Anaya.

3.1 Relación entre Big Data, análisis de datos, machine/deep learning e IA

Dentro de la ciencia de datos, la cual, como se ha explicado, se entiende como la disciplina que se encarga de obtener información a partir de datos, los expertos en este ámbito han llegado a desarrollar nuevas tecnologías o técnicas estadísticas y de análisis de datos, desde una perspectiva múltiple, combinando fundamentos y métodos de diferentes disciplinas para poder evaluar y almacenar grandes conjuntos de datos.

Una de estas tecnologías es el **Big Data**, que abarca el conjunto de técnicas necesarias para estudiar datos de forma estructurada y a gran escala. Las habilidades y técnicas necesarias para trabajar con Big Data son diferentes de las que se utilizan en otras técnicas de las que hablaremos a continuación, y aunque a menudo se confunden, también actúa como la base de información para estos modelos.

Estos modelos son:

- **Inteligencia Artificial:** Se originó con el objetivo de permitir que las máquinas imiten funciones del cerebro humano para realizar tareas que requieren inteligencia, como el reconocimiento de voz o la toma de decisiones. La IA se divide en dos categorías según sus capacidades:

    - **IA general (o fuerte):** Aunque actualmente sigue en desarrollo y no se ha alcanzado totalmente, busca crear máquinas o software que posean inteligencia en un sentido amplio, capaces de entender, pensar y razonar en una variedad de situaciones, similar a lo que puede hacer cualquier ser humano. Dentro de esta categoría se incluyen el aprendizaje automático avanzado con sistemas que pueden reconocer imágenes; robots automáticos que interactúen con su entorno de manera inteligente; y sistemas expertos que utilizan sus conocimientos para resolver problemas como lo haría un experto.

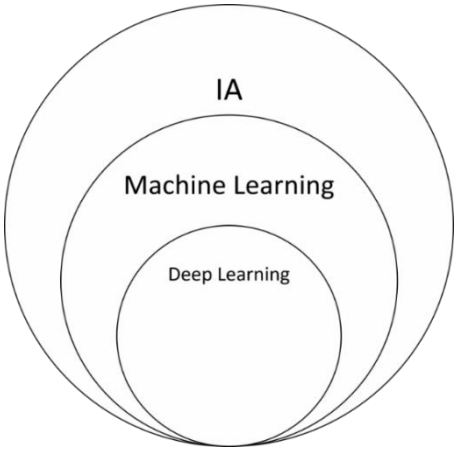
    - **IA estrecha (o débil):** Esta es la IA aplicada en campos específicos, como asistentes virtuales o algoritmos de recomendación, que realizan tareas muy concretas con gran precisión, sin necesidad de demostrar una inteligencia general comparable a la del ser humano. Ejemplos de este tipo de IA se pueden encontrar en asistentes virtuales como Siri o Alexa, sistemas de reconocimiento de voz para la transcripción automática o sistemas de detección de spam.

VÍDEO DE INTERÉS

Accede para conocer el reglamento europeo sobre Inteligencia Artificial:

- **Machine Learning:** También llamado aprendizaje automático, es un subconjunto o rama contenida de la inteligencia artificial que permite a las máquinas aprender y "pensar" de manera similar a los humanos. A través de algoritmos, estas máquinas analizan datos y extraen patrones sin necesidad de ser programadas explícitamente para cada tarea. El ML es una herramienta clave en los proyectos de ciencia de datos, ya que permite obtener información automatizada a partir de grandes volúmenes de datos, facilitando así el aprendizaje a partir de la experiencia.

- **Deep Learning:** Es un subcampo dentro del Machine Learning que utiliza redes neuronales artificiales para procesar datos de manera más sofisticada. Inspiradas en el funcionamiento del cerebro humano, estas redes tienen múltiples capas que permiten realizar análisis complejos, como reconocimiento de voz o procesamiento de imágenes. Su estructura multicapa permite que el sistema aprenda de manera más efectiva a partir de grandes volúmenes de información.



Esquema de la relación entre IA, Machine Learning y Deep Learning

Fuente: https://www.josenalda.com/blog/que-son-el-data-science-big-data-inteligencia-artificial-machine-learning-y-deep-learning-y-en-que-se-diferencian/#Relacion-entre-el-Data-Science-Big-Data-Inteligencia-Artificial-Machine-Learning-y-Deep-Learning

PARA SABER MÁS

Aquí puedes profundizar en las diferencias entre Data Science, Big Data, Inteligencia Artificial, Machine Learning y Deep Learning. Además, encontrarás esquemas explicativos:

A modo de resumen, todos los términos descritos están relacionados entre sí, aunque con unas diferencias tan mínimas que llegan incluso a confundirse. A continuación, se explica mejor la diferenciación de estos conceptos entre sí.

Deep Learning es un tipo de **Machine Learning**, que a su vez forma parte de la Inteligencia Artificial. Esto significa que el aprendizaje profundo es una técnica específica dentro de un conjunto más amplio de herramientas de IA.

La ciencia de datos y el Big Data comparten un área de intersección. Mientras que el Big Data se enfoca en el almacenamiento y procesamiento de grandes volúmenes de información, la ciencia de datos incluye una amplia gama de técnicas analíticas para interpretar esos datos y obtener conclusiones valiosas. Ambas disciplinas emplean métodos y herramientas similares, como los algoritmos de Machine Learning.

Inteligencia Artificial y Big Data también están conectados. Los modelos de IA necesitan grandes cantidades de datos para entrenar y mejorar sus predicciones, mientras que el análisis de esos datos es posible gracias a las herramientas que proporciona el Big Data.

En definitiva, cada una de estas áreas tiene sus propias aplicaciones y técnicas, pero comparten herramientas y métodos comunes. Mientras que el Big Data se centra en gestionar grandes volúmenes de información, la Inteligencia Artificial y el Machine Learning aprovechan esa información para automatizar decisiones y mejorar procesos, haciendo posible que las organizaciones extraigan valor de los datos de forma más eficiente.

3.2 Aplicación y objetivos de la ciencia de datos en empresas

El propósito de aplicar la ciencia de datos en cualquier organización es **identificar patrones y tendencias que faciliten la toma de decisiones estratégicas para el negocio**. Esta disciplina se ha vuelto fundamental para las organizaciones que han optado por adentrarse en la transformación digital, asumiendo un rol de liderazgo en el entorno empresarial actual.

Las herramientas de inteligencia de negocios y ciencia de datos pueden incrementar la productividad laboral en un 40%, permitiendo un uso más eficiente del tiempo. Esto no solo mejora la productividad económica de la empresa, sino que también la introduce en una cultura ágil, un aspecto crucial en el proceso de transformación digital.

EJEMPLO PRÁCTICO

Sergio trabaja en una empresa que quiere mejorar su estrategia comercial utilizando ciencia de datos. Junto con su equipo, debe desarrollar un plan para analizar los datos de clientes y operaciones.

Sergio comienza organizando los datos disponibles y propone implementar un sistema de Big Data para almacenarlos y procesarlos eficientemente. Piensa que, utilizando Machine Learning, se podría segmentar a los clientes y personalizar las ofertas, mejorando así la efectividad de las campañas de marketing. Junto con el equipo, sugieren crear modelos predictivos para anticipar la demanda de productos y optimizar la gestión de inventarios.

Como resultado, la empresa observa un aumento en la satisfacción del cliente y una reducción de costos operativos, logrando una mayor agilidad en su respuesta al mercado.

Visto desde otro punto de vista, los **beneficios** que la ciencia de datos proporciona a las empresas incluyen:

- **Mitigación de riesgos y fraudes:** Mediante metodologías estadísticas y modelos avanzados, es posible identificar datos con comportamientos anómalos, generando alertas que permiten reaccionar de manera oportuna ante operaciones inusuales.

- **Oferta de productos adecuados:** La ciencia de datos facilita identificar el momento y lugar óptimos para ofrecer productos y servicios, además de desarrollar innovaciones que respondan a las necesidades de los consumidores.

- **Experiencia personalizada para el cliente:** El análisis de patrones de consumo y comportamiento, junto con la segmentación según características específicas de los compradores, permite a los equipos de ventas y marketing comprender a fondo a su audiencia, logrando una visión 360° de la misma.

- **Definición de acciones basadas en tendencias:** Explorar y analizar los datos de la organización ayuda a establecer metas que mejoren el rendimiento y aumenten la rentabilidad.

- **Toma de decisiones fundamentadas en datos:** La integración y análisis de información de diversas fuentes permiten simular estrategias y cursos de acción para evaluar cuál generará los mejores resultados en áreas administrativas, comerciales, operativas o financieras.

- **Reducción de tareas manuales:** Automatizando procesos de forma rápida y eficiente, se facilita el manejo de grandes volúmenes de datos que, de otro modo, serían difíciles de analizar. Esto proporciona una comprensión más profunda de la información, lo que permite tomar decisiones de manera más efectiva.

Por otro lado, es bueno tener en cuenta también las **desventajas** pertenecientes a la aplicación del análisis de datos en las empresas, las cuales son:

- **Decisiones poco efectivas:** Si los datos no se procesan de manera adecuada, las decisiones podrían basarse en conjeturas o intuiciones, en lugar de hechos verificables. Esto puede conducir a elecciones equivocadas o ineficientes, afectando negativamente el rendimiento y la rentabilidad de la empresa.

- **Desconocimiento del cliente:** Sin un análisis adecuado de la información del cliente, es difícil obtener una comprensión profunda de estos aspectos, lo que impide personalizar estrategias de marketing, segmentar el mercado de manera precisa o responder a las necesidades específicas de los consumidores.

- **Ineficiencia en el uso de recursos y tiempo:** Si no se define un enfoque claro hacia los datos relevantes y una estrategia de análisis adecuada, las empresas pueden perder tiempo valioso procesando información que no resulta útil.

- **Falta de identificación de problemas y riesgos:** Cuando no se realiza un análisis detallado, los problemas o los riesgos en el negocio pueden pasar inadvertidos, lo que podría llevar a pérdidas financieras.

Una vez conocidos tanto los beneficios como las desventajas que nos proporciona esta herramienta, cabe preguntarse **cómo se debe implementar en una empresa** para poder disfrutar de dichos beneficios.

Aprovechar las oportunidades que brinda la ciencia de datos **requiere decisiones de inversión específicas y la creación de una cultura organizacional** que favorezca el intercambio de información y utilice un enfoque cuantitativo para resolver los retos. Para lograrlo, se debe:

- Establecer un mandato que ponga la estrategia de datos en primer plano.
- Alinear las prácticas de gestión de datos para acelerar las prioridades de la empresa, como mejorar la experiencia del cliente, desarrollar nuevos negocios y aumentar los ingresos.
- Implementar acuerdos de colaboración de datos con terceros.
- Solucionar las deficiencias en la infraestructura y potenciar las capacidades internas.
- Invertir en inteligencia artificial y aprendizaje automático.

Gestionar, procesar y organizar la información de forma ágil es un desafío importante para las empresas, ya que la ciencia de datos será cada vez más clave para monitorear, gestionar y recopilar indicadores de rendimiento que mejoren la toma de decisiones en toda la organización.

Una vez resueltos estos retos, pasamos a los **pasos** que toda empresa debe seguir para alcanzar y prosperar en su objetivo de implementación de la ciencia de datos.

PASO 1 **Definición de objetivos y estrategias**	El primer paso es establecer qué se quiere lograr, cómo mejorar la eficiencia o aumentar la satisfacción del cliente. Definir claramente estos objetivos orientará las decisiones sobre tecnología y diseño del sistema. Además, la empresa debe adoptar una cultura basada en datos, donde las decisiones se tomen en función de análisis concretos, no de intuiciones.
PASO 2 **Recolección de datos**	Con los objetivos claros, es importante identificar qué datos se necesitan y cómo se recopilarán, ya

	sean internos (transacciones) o externos (redes sociales). También es crucial asegurar la calidad de los datos y fomentar la recolección en toda la organización.
PASO 3 **Almacenamiento de datos**	Se debe elegir entre almacenamiento local o en la nube, considerando factores como seguridad, escalabilidad y costos. El almacenamiento local brinda control total, pero es costoso de mantener, mientras que la nube es más flexible y rápida de implementar, aunque puede generar dependencia del proveedor.
PASO 4 **Análisis de datos**	El análisis es el núcleo del Big Data. Es fundamental tener un equipo especializado que limpie y analice los datos usando técnicas avanzadas como machine learning. Además, la empresa debe fomentar el uso de datos en todas las áreas para una toma de decisiones más informada.
PASO 5 **Visualización e interpretación**	Para que los datos sean útiles, deben presentarse de manera comprensible. Herramientas de visualización como Tableau o PowerBI ayudan a transformar datos complejos en gráficos y dashboards, facilitando su interpretación y el uso de insights para mejorar las operaciones empresariales. Este enfoque asegura que los datos se utilicen de manera efectiva para impulsar decisiones estratégicas dentro de la organización.

ENLACE DE INTERÉS

Conoce 5 empresas que usan Big Data y han conseguido los mejores resultados:

Ordenador analizando datos

Fuente: https://www.pexels.com/es-es/foto/ordenador-portatil-negro-y-gris-546819/

4. IMPORTANCIA DE LA SEGURIDAD EN EL MANEJO DE DATOS

Ante el aumento de ciberamenazas, tu empresa ha decidido reforzar su estrategia de ciberseguridad. Gracias a tus conocimientos, puedes colaborar para salvaguardar la información sensible y asegurar el cumplimiento de las normativas de protección de datos vigentes.

La seguridad en el manejo de datos es crucial para proteger tanto la **integridad de la información** como la **privacidad de los usuarios**. Con el creciente volumen de datos que las organizaciones recopilan, procesan y almacenan, el riesgo de violaciones de seguridad y ciberataques ha aumentado de manera significativa.

Una gestión adecuada de la seguridad de los datos no solo garantiza el cumplimiento de regulaciones y normativas, sino que también preserva la confianza de los clientes y protege los activos más valiosos de la empresa. La implementación de estrategias de seguridad efectivas es esencial para mitigar riesgos, evitar pérdidas financieras y asegurar la continuidad del negocio en un entorno cada vez más interconectado.

La **protección de datos**, también conocida como seguridad informática o de la información, es un componente fundamental en las tecnologías de la información (TI) para empresas de cualquier tamaño y sector.

Su propósito principal es **salvaguardar los datos de accesos no autorizados** y evitar su alteración o pérdida a lo largo de su ciclo de vida. Este campo abarca conceptos como la **encriptación**, que es el proceso de aplicar un algoritmo para convertir datos en texto simple a un formato ilegible, conocido como texto cifrado; **la tokenización**, que es el proceso de transformar un dato valioso, como un número de cuenta, en una secuencia aleatoria de caracteres llamada token, que carece de significado y no proporciona información útil en caso de ser comprometido; **y la gestión de claves**, la cual cubre todo el ciclo de vida de las claves criptográficas, que incluye su creación, almacenamiento, protección, distribución, actualización y, por último, su eliminación. Dichos conceptos son esenciales para asegurar la información en todas las plataformas y aplicaciones que utiliza una organización.

En la actualidad, muchas empresas a nivel global están invirtiendo considerablemente en tecnologías de ciberseguridad para resguardar sus activos más valiosos, como su reputación, propiedad intelectual y los datos sensibles de sus clientes. Se debe destacar que, a la hora de implementar estrategias de seguridad de la información, hay tres factores clave que todas las organizaciones deben considerar: **las personas**

involucradas, los procesos establecidos y las herramientas tecnológicas empleadas. Estos elementos son la base de una defensa sólida contra las amenazas cibernéticas.

4.1 Ciberseguridad

La **ciberseguridad** se refiere a las **acciones destinadas a proteger dispositivos, redes, aplicaciones de software, sistemas esenciales y datos frente a amenazas digitales**. Las empresas tienen el deber de salvaguardar la información para conservar la confianza de sus clientes y cumplir con las regulaciones vigentes. Para ello, utilizan diversas herramientas y estrategias de ciberseguridad que impiden el acceso no autorizado a datos sensibles, además de minimizar posibles interrupciones en sus operaciones causadas por actividades maliciosas en la red. Las organizaciones fortalecen su seguridad digital combinando la defensa de sus recursos tecnológicos, el establecimiento de procesos eficientes y la capacitación de su personal.

En diversas industrias, como la energía, el transporte, el comercio minorista y la manufactura, se utilizan sistemas digitales y conectividad de alta velocidad para ofrecer un servicio al cliente eficiente y realizar operaciones rentables. Al igual que protegen sus activos físicos, estas organizaciones deben asegurar sus recursos digitales y sistemas contra accesos no intencionados. Un acceso no autorizado a un sistema informático, red o recursos conectados, que ocurre de manera accidental o maliciosa, se conoce como ciberataque. **Si un ciberataque tiene éxito, puede resultar en la exposición, robo, eliminación o modificación de datos sensibles**. Las estrategias de ciberseguridad están diseñadas para defenderse de estos ciberataques.

EJEMPLO PRÁCTICO

Rita trabaja en una institución que ha enfrentado varios intentos de ciberataques en los últimos meses. La dirección de la empresa le encarga a Rita y a su equipo desarrollar un plan para fortalecer la ciberseguridad y proteger la información sensible de los clientes. Rita comienza reflexionando sobre las prácticas actuales y se da cuenta de que ni ella ni sus compañeros han recibido capacitación sobre la protección de datos y las mejores prácticas de seguridad.

Para abordar esta situación, Rita sugiere a la dirección que se organice un programa de formación accesible para todos los empleados, que podría incluir talleres sobre la identificación de comportamientos sospechosos en correos electrónicos y la importancia de utilizar contraseñas seguras. Además, propone establecer políticas claras sobre el uso de dispositivos personales en el trabajo y la necesidad de reportar cualquier actividad inusual.

Tras la capacitación y la implementación de estas nuevas políticas, la empresa experimenta una notable reducción en los intentos de acceso no autorizado, lo que refuerza la confianza de sus clientes en la protección de sus datos.

Las principales ciberamenazas, o al menos las más comunes a las que se enfrentan los profesionales de ciberseguridad, son:

CIBERAMENAZA	DESCRIPCIÓN	TIPOS
Malware	Software malicioso diseñado para permitir que terceros accedan de manera no autorizada a información sensible o para interrumpir el funcionamiento normal de infraestructuras críticas.	**Troyanos:** Programas que se disfrazan como software legítimo para engañar a los usuarios y facilitar el acceso no autorizado a sus sistemas. **Spyware:** Software que recopila información sobre un usuario sin su conocimiento, a menudo registrando datos personales o financieros. **Virus:** Programas que se replican y se propagan a través de archivos, afectando el rendimiento del sistema y corrompiendo datos.
Ransomware	Modelo de extorsión que utiliza diversas tecnologías para chantajear a organizaciones con el fin de obtener dinero. Ya sea que esté comenzando o utilizando AWS, existen recursos específicos disponibles para ayudar a proteger sus sistemas críticos y datos confidenciales contra el ransomware.	**Ransomware de cifrado:** Cifra archivos y pide un rescate a cambio de la clave de descifrado. **Ransomware de bloqueo:** Impide el acceso al sistema hasta que se pague un rescate.
Ataque intermediario	Un ataque de intermediario ocurre cuando un tercero intenta acceder de forma no	**Intercepción de redes:** Colocándose en medio de una

	autorizada a una red durante un intercambio de datos. Estos ataques aumentan el riesgo para la seguridad de información confidencial, como datos financieros.	conexión Wi-Fi pública para robar información. **Suplantación de DNS:** Redirigiendo a las víctimas a sitios web maliciosos a través de un servidor DNS comprometido.
Phishing	El phishing es una amenaza cibernética que utiliza técnicas de ingeniería social para engañar a los usuarios y hacer que revelen información personal. Por ejemplo, los atacantes pueden enviar correos electrónicos que instan a los usuarios a hacer clic en enlaces y proporcionar detalles de tarjetas de crédito en un sitio web falso. Además, los ataques de phishing pueden incluir solicitudes para descargar archivos adjuntos maliciosos que instalan malware en los dispositivos de la empresa.	**Phishing por correo electrónico:** Mensajes que aparentan ser de fuentes legítimas, solicitando información personal. **Spear phishing:** Ataques dirigidos a individuos específicos, personalizando los mensajes para aumentar su credibilidad. **Vishing:** Uso de llamadas telefónicas para engañar a las personas y obtener datos confidenciales.
DDoS	Un ataque de denegación de servicio distribuido (DDoS) es un intento malicioso de interrumpir el funcionamiento normal de un servidor, servicio o red al inundarlo con un volumen abrumador de tráfico. Estos ataques se llevan a cabo a través de múltiples sistemas comprometidos, lo que dificulta su mitigación.	**Inaccesibilidad del servicio:** Los usuarios legítimos no pueden acceder a la plataforma, lo que afecta la experiencia del cliente y la reputación de la empresa. **Pérdidas financieras:** La incapacidad de operar puede resultar en pérdidas significativas de ingresos.
Amenaza interna	Una amenaza interna es un riesgo de seguridad que se origina en el interior de una	**Robo de datos:** Empleados deshonestos pueden sustraer

	organización, generalmente a partir de empleados, contratistas o socios que tienen acceso autorizado a los sistemas. Estos individuos pueden actuar de manera intencionada o accidentalmente, causando daño a la infraestructura de TI.	información confidencial para beneficio personal. **Errores humanos:** El personal puede, sin querer, comprometer la seguridad al ignorar políticas de seguridad o al hacer clic en enlaces maliciosos.

Para reducir este tipo de amenazas, las organizaciones deben establecer políticas de control de acceso rigurosas, llevar a cabo auditorías de seguridad de manera regular y fomentar una cultura de concienciación sobre la seguridad entre su personal.

PARA SABER MÁS

Conoce más tipos de ciberamenazas, además de las expuestas:

4.2 Regulación y normativa en relación a los datos

La protección de los datos personales y los derechos digitales son cuestiones fundamentales en la sociedad actual, caracterizada por un constante flujo de información y el uso generalizado de tecnologías digitales.

Nuestra ley orgánica tiene como objetivo principal ajustar la normativa española a las exigencias del Reglamento Europeo sobre protección de datos. Este reglamento busca **garantizar que los datos personales de las personas estén debidamente protegidos** y que puedan circular libremente dentro de la Unión Europea. Asimismo, la ley también se **propone reforzar los derechos digitales de los ciudadanos**, asegurando que sus derechos sean respetados y protegidos en el entorno digital, tal como establece la Constitución Española. En esencia, se trata de un marco legal que busca no solo proteger la privacidad de las personas, sino también fomentar un uso responsable de los datos en el contexto actual de digitalización.

Dentro de este mismo marco, se exponen a continuación los aspectos más importantes que se deben destacar:

1. **Protección de datos personales:** La ley regula cómo deben ser tratados los datos personales de los ciudadanos en España, siguiendo el principio de transparencia, legalidad y proporcionalidad. Los datos personales solo pueden ser recogidos y procesados para fines específicos, legítimos y explícitos, y no deben ser utilizados de manera incompatible con esos fines.

   Los responsables del tratamiento de los datos están obligados a adoptar medidas técnicas y organizativas adecuadas para garantizar la protección y seguridad de los datos, evitando el acceso no autorizado o el uso indebido. Esto incluye medidas como la pseudonimización y el cifrado de datos.

2. **Derechos de los ciudadanos:** La ley establece y refuerza una serie de derechos que las personas pueden ejercer sobre sus datos personales. Estos derechos incluyen:

   - **Derecho de acceso:** Permite al ciudadano saber qué datos están siendo tratados y obtener una copia de los mismos.

   - **Derecho de rectificación:** Posibilita que las personas corrijan datos incorrectos o incompletos.

   - **Derecho de supresión:** Conocido también como “derecho al olvido”, permite solicitar la eliminación de los datos cuando ya no sean necesarios o cuando el consentimiento para su tratamiento se haya retirado.

   - **Derecho a la limitación del tratamiento:** El ciudadano puede solicitar que el tratamiento de sus datos se restrinja, por ejemplo, mientras se comprueba la exactitud de los mismos o se resuelve una reclamación.

   - **Derecho a la portabilidad:** Permite al individuo recibir sus datos en un formato estructurado, de uso común y lectura mecánica, para transferirlos a otro responsable del tratamiento.

   - **Derecho de oposición:** La persona puede oponerse al tratamiento de sus datos en determinadas circunstancias, especialmente en casos de marketing directo o cuando se trate de decisiones automatizadas.

- **Derecho a no ser objeto de decisiones automatizadas:** Protege a los ciudadanos de ser evaluados o perfilados únicamente a través de algoritmos sin intervención humana, lo que puede afectar significativamente sus derechos o libertades.

NORMATIVA DE INTERÉS

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos):

3. **Consentimiento:** La ley enfatiza que el tratamiento de los datos personales solo puede llevarse a cabo cuando el ciudadano ha dado su consentimiento explícito y claro.

   - El consentimiento no puede ser asumido, como en el caso de las casillas marcadas por defecto. Debe ser **específico, informado, inequívoco y revocable**.

   - Si el tratamiento de los datos se basa en el consentimiento, este puede ser **retirado en cualquier momento**, y el tratamiento posterior debe cesar inmediatamente.

   - El consentimiento de los menores de edad está regulado de forma especial. Para aquellos menores de 14 años, el consentimiento debe ser otorgado por sus **tutores legales**.

4. **Delegados de protección de datos:** La ley introduce la figura del Delegado de Protección de Datos (DPO), cuya función principal es asegurar el cumplimiento de la normativa de protección de datos dentro de una organización. Este rol es obligatorio para entidades públicas y para algunas organizaciones privadas, especialmente aquellas que procesan grandes cantidades de datos sensibles.

El DPO es responsable de:

- **Supervisar** que se respeten las leyes de protección de datos.
- **Asesorar y capacitar** a la organización sobre las obligaciones en cuanto a protección de datos.
- Actuar como punto de contacto con la **Agencia Española de Protección de Datos (AEPD)** y con los ciudadanos que deseen ejercer sus derechos.
- Supervisar la **gestión de riesgos** relacionados con el tratamiento de datos.

EJEMPLO PRÁCTICO

En una agencia que recopila datos de clientes para sus campañas, se ha observado que muchos clientes son reacios a compartir su información personal. Para abordar esta situación, el equipo quiere implementar un sistema de gestión de consentimiento de datos que cumpla con las normativas de protección de datos.

El grupo propone el desarrollo de un formulario de consentimiento que explique, de manera clara, cómo se utilizarán los datos y los derechos que tienen los clientes sobre su información. Además, sugieren crear otro procedimiento sencillo para que los clientes puedan retirar su consentimiento fácilmente en cualquier momento.

Al implementar este sistema, la agencia nota un aumento en la disposición de los clientes a compartir sus datos. Esto no solo mejora la efectividad de las campañas de marketing, sino que también asegura el cumplimiento de las normativas vigentes, generando mayor confianza entre los clientes.

5. **Garantía de derechos digitales:** Se añaden a los derechos tradicionales relacionados con la protección de datos. Estos derechos buscan adaptarse a la era tecnológica actual y ofrecer protecciones en el ámbito digital. Entre estos derechos se encuentran:
   - **Derecho a la intimidad y uso de dispositivos en el ámbito laboral:** Los empleados tienen derecho a la privacidad en el uso de dispositivos digitales proporcionados por la empresa, y las organizaciones deben informar de manera clara y previa sobre el uso de sistemas de monitoreo

o videovigilancia. La vigilancia sólo puede hacerse respetando la dignidad y privacidad del trabajador.

- **Derecho a la desconexión digital:** Se garantiza el derecho de los empleados a desconectarse de sus dispositivos digitales y comunicaciones laborales fuera de su jornada de trabajo, promoviendo así un equilibrio adecuado entre la vida laboral y personal. Las empresas deben establecer políticas claras para respetar este derecho.

- **Acceso universal a internet:** Se reconoce el derecho de todos los ciudadanos a acceder a internet de manera asequible y de calidad, como parte del ejercicio de sus derechos fundamentales.

- **Protección de los menores en internet:** La ley establece medidas para garantizar la protección de los menores en el entorno digital, especialmente en relación con el uso de redes sociales y la divulgación de su información personal. Se promueve la educación digital de los menores para que hagan un uso responsable de las tecnologías.

- **Derecho a la neutralidad de la red:** Se establece que los usuarios deben poder acceder y utilizar internet sin discriminación ni limitaciones, garantizando la neutralidad de la red, es decir, que todo el tráfico de internet debe ser tratado de manera igualitaria.

- **Derecho al testamento digital:** Este derecho regula el acceso a las cuentas y datos digitales de una persona fallecida, permitiendo a los herederos gestionar o eliminar su presencia en internet, como perfiles en redes sociales, correos electrónicos, etc.

Derechos digitales

Fuente: https://www.freepik.es/foto-gratis/ojo-robot-futurista_11309679.htm

RESUMEN FINAL

En esta unidad se ha examinado el impacto transformador de la digitalización en los sectores productivos, destacando el rol crucial del manejo de datos, el Big Data y el Cloud Computing. Estos elementos permiten a las empresas optimizar sus operaciones, tomar decisiones basadas en información precisa y mejorar la personalización y satisfacción del cliente. El ciclo de vida del dato, desde su creación hasta su almacenamiento, procesamiento y eventual eliminación, ha sido abordado como base para la calidad y seguridad de la información.

Además, se ha explorado cómo la ciencia de datos y las tecnologías habilitadoras, como la inteligencia artificial y el machine learning, permiten el análisis de grandes volúmenes de datos, facilitando la detección de patrones, la previsión de tendencias y la mejora en la toma de decisiones empresariales. Las aplicaciones de estas tecnologías benefician a las organizaciones en múltiples frentes: desde la mitigación de riesgos hasta la mejora de la eficiencia y competitividad.

La unidad también ha enfatizado la importancia de la ciberseguridad en la protección de datos sensibles. Las empresas deben cumplir con normativas, como el Reglamento General de Protección de Datos (RGPD) y la Ley de Protección de Datos en España, que regulan el tratamiento y la seguridad de la información personal. Asimismo, la figura del Delegado de Protección de Datos (DPO) se presenta como clave para garantizar el cumplimiento normativo y la seguridad de los datos en un entorno digital cada vez más interconectado y vulnerable a amenazas.