

4.1. Analítica, Business Intelligence y Big Data

Conceptos principales	1
¿Qué es la inteligencia artificial?	1
¿Qué es el Machine Learning?	2
¿Qué es el deep learning?	4
¿Qué es el análisis de datos y la visualización de datos?	5
¿Qué es el big data y el small data?	5
Volumen	5
Variedad	5
Velocidad	5
Cadena de valor del dato	6
Beneficios del uso de datos en una organización	6
Sinergias con otras tecnologías (cloud, inteligencia artificial, etc.)	7
Machine Learning: algoritmos y aplicaciones	10
Herramientas de visualización de datos	11
Tableau	11
PowerBI	13
Looker	14
MicroStrategy	15

Conceptos principales

¿Qué es la inteligencia artificial?

La inteligencia artificial (IA) se refiere a la capacidad de las máquinas y sistemas informáticos para realizar tareas que normalmente requieren inteligencia humana. Es un campo multidisciplinario que combina la informática, la estadística, las matemáticas y la neurociencia, entre otros.

El objetivo de la inteligencia artificial es desarrollar programas y sistemas capaces de simular y emular el pensamiento humano, así como de realizar tareas de manera autónoma y adaptativa. Estos sistemas pueden aprender, razonar,

reconocer patrones, tomar decisiones, comprender y comunicarse en lenguaje natural, y mejorar su desempeño a través de la experiencia.

Existen diferentes enfoques y técnicas en el campo de la IA, entre ellos:

Aprendizaje automático (Machine Learning): Es una rama de la IA que se centra en el desarrollo de algoritmos y modelos que pueden aprender y mejorar automáticamente a partir de datos. Los modelos de aprendizaje automático pueden reconocer patrones, hacer predicciones y tomar decisiones basadas en la información proporcionada.

Aprendizaje profundo (Deep Learning): Es una subcategoría del aprendizaje automático que se basa en redes neuronales artificiales profundas. Estas redes tienen múltiples capas y son capaces de aprender características y representaciones complejas a partir de grandes volúmenes de datos.

Procesamiento del lenguaje natural (Natural Language Processing, NLP): Se enfoca en la interacción entre las computadoras y el lenguaje humano. Los sistemas de NLP permiten a las máquinas comprender, interpretar y generar lenguaje natural, lo que se utiliza en aplicaciones como la traducción automática, los chatbots y la analítica de texto.

Visión por computadora (Computer Vision): Se ocupa del análisis, procesamiento e interpretación de imágenes y videos por parte de las máquinas. Los sistemas de visión por computadora pueden reconocer objetos, rostros, emociones, realizar seguimiento de movimiento y llevar a cabo tareas de reconocimiento visual.

La inteligencia artificial se utiliza en una amplia gama de aplicaciones y sectores, como la medicina, la robótica, los vehículos autónomos, los sistemas de recomendación, el análisis de datos, el comercio electrónico y muchas más. A medida que avanza la tecnología, la IA continúa evolucionando y transformando la forma en que interactuamos con la tecnología y el mundo que nos rodea.

¿Qué es el Machine Learning?

El Machine Learning, o aprendizaje automático en español, es una rama de la inteligencia artificial que se enfoca en desarrollar algoritmos y modelos que permiten a las computadoras aprender y mejorar automáticamente a partir de los datos, sin ser programadas explícitamente para realizar una tarea específica.

En lugar de seguir reglas o instrucciones predefinidas, los modelos de Machine Learning utilizan datos de entrada para identificar patrones, hacer predicciones y tomar decisiones basadas en la información proporcionada. El objetivo principal es permitir que las máquinas aprendan de manera autónoma, a través de la experiencia, y mejoren su desempeño a medida que se exponen a más datos.

El proceso de **Machine Learning** generalmente implica los siguientes pasos:

Recopilación de datos: Se recopilan y se preparan conjuntos de datos relevantes que contienen ejemplos y características relacionadas con el problema que se quiere abordar.

Entrenamiento del modelo: Se utiliza el conjunto de datos de entrenamiento para alimentar el modelo de Machine Learning y ajustar sus parámetros. Durante el entrenamiento, el modelo aprende a reconocer patrones y realizar predicciones a partir de los datos proporcionados.

Evaluación del modelo: Se utiliza un conjunto de datos separado, llamado conjunto de validación o conjunto de prueba, para evaluar el rendimiento del modelo entrenado. Esto permite medir su precisión, eficacia y capacidad de generalización.

Mejora y ajuste: Si el rendimiento del modelo no es satisfactorio, se realizan ajustes en su estructura, algoritmos o parámetros para mejorar su desempeño. Este proceso de mejora iterativa puede involucrar la modificación de hiperparámetros o la selección de diferentes algoritmos.

Implementación y predicción: Una vez que el modelo ha sido entrenado y evaluado de manera satisfactoria, se puede utilizar para realizar predicciones o tomar decisiones en nuevos datos de entrada.

El **Machine Learning** se aplica en una amplia gama de áreas, como la detección de fraudes, la recomendación de productos, el análisis de sentimientos, la detección de anomalías, la clasificación de imágenes y textos, entre otros. Los algoritmos de Machine Learning pueden ser supervisados, no supervisados o semi-supervisados, dependiendo de la disponibilidad y la naturaleza de los datos de entrenamiento.

¿Qué es el deep learning?

El Deep Learning, o aprendizaje profundo en español, es una rama del campo de la inteligencia artificial (IA) que se basa en redes neuronales artificiales de gran escala para aprender y realizar tareas de manera autónoma. Está inspirado en el funcionamiento del cerebro humano y se centra en el procesamiento y análisis de datos para extraer patrones y características complejas.

A diferencia de los algoritmos tradicionales de aprendizaje automático, que se basan en la extracción manual de características, el Deep Learning permite que los modelos aprendan representaciones de características de manera automática y jerárquica a partir de los datos. Esto se logra mediante el uso de redes neuronales artificiales con múltiples capas ocultas, lo que permite la extracción de características a diferentes niveles de abstracción.

El **Deep Learning** ha tenido un gran impacto en una variedad de áreas, como el reconocimiento de imágenes y de voz, el procesamiento del lenguaje natural, la visión por computadora, la traducción automática, la generación de contenido, entre otros. Al tener un mayor número de capas y parámetros en comparación con otros modelos de aprendizaje automático, los modelos de Deep Learning pueden aprender características más complejas y realizar predicciones más precisas en una amplia gama de problemas.

Una de las técnicas más utilizadas en el Deep Learning es la denominada Red Neuronal Convolucional (**CNN**), que se emplea principalmente en el procesamiento de imágenes y la visión por computadora. Otra técnica popular es la Red Neuronal Recurrente (**RNN**), utilizada en el procesamiento del lenguaje natural y otras tareas que implican secuencias de datos.

El Deep Learning ha logrado avances significativos en muchas aplicaciones y ha impulsado el desarrollo de tecnologías como los vehículos autónomos, los asistentes virtuales, el reconocimiento facial, la detección de fraudes y muchas más. Sin embargo, también requiere grandes conjuntos de datos y recursos computacionales potentes para entrenar y ejecutar los modelos, lo que puede representar un desafío en términos de tiempo y recursos para algunas aplicaciones.

¿Qué es el análisis de datos y la visualización de datos?

¿Qué es el *big data* y el *small data*?

La principal diferencia entre big data y small data radica en la escala y el alcance de los conjuntos de datos utilizados en cada caso. Aquí están las diferencias clave:

Volumen

Big Data: Se refiere a conjuntos de datos extremadamente grandes y complejos que no se pueden gestionar fácilmente con herramientas de procesamiento de datos tradicionales. Los conjuntos de datos de big data suelen tener terabytes o incluso petabytes de tamaño.

Small Data: Se refiere a conjuntos de datos más pequeños que pueden ser gestionados y analizados con herramientas de procesamiento de datos convencionales. Los conjuntos de datos de small data suelen tener tamaños de gigabytes o menos.

Variedad

Big Data: Implica una amplia variedad de tipos y fuentes de datos, incluyendo datos estructurados (como bases de datos), datos no estructurados (como texto, imágenes, audio, video) y datos semiestructurados (como archivos XML). El big data puede provenir de diversas fuentes, como redes sociales, sensores, registros de transacciones, etc.

Small Data: Por lo general, se basa en datos estructurados y más homogéneos, que siguen un formato y una organización coherentes.

Velocidad

Big Data: Se caracteriza por la velocidad a la que se generan los datos y la necesidad de analizarlos en tiempo real o cerca de tiempo real. Los sistemas de big data están diseñados para manejar grandes volúmenes de datos en tiempo real y generar resultados rápidos.

Small Data: Por lo general, se procesa a un ritmo más lento y no tiene la misma exigencia de tiempo real para el análisis.

Valor:

Big Data: El enfoque principal en el big data es extraer información valiosa y conocimiento significativo a partir de grandes volúmenes de datos, lo que puede ayudar a las organizaciones a tomar decisiones estratégicas y obtener una ventaja competitiva.

Small Data: El valor se encuentra en el análisis detallado de datos más pequeños y específicos, lo que permite comprender mejor a un grupo o individuo en particular y tomar decisiones más personalizadas.

En resumen, la principal diferencia entre big data y small data radica en la escala y el alcance de los conjuntos de datos utilizados, así como en los enfoques y herramientas requeridos para su gestión y análisis. El big data se refiere a conjuntos de datos masivos, diversos y en constante crecimiento, mientras que el small data se centra en conjuntos de datos más pequeños y específicos que se pueden analizar con herramientas convencionales.

Cadena de valor del dato

¿Cómo aporta valor el dato desde que se extrae hasta su utilización en un producto o su uso para tomar decisiones?

Beneficios del uso de datos en una organización

El uso de datos en una organización puede brindar una serie de beneficios significativos. Aquí hay algunos de ellos:

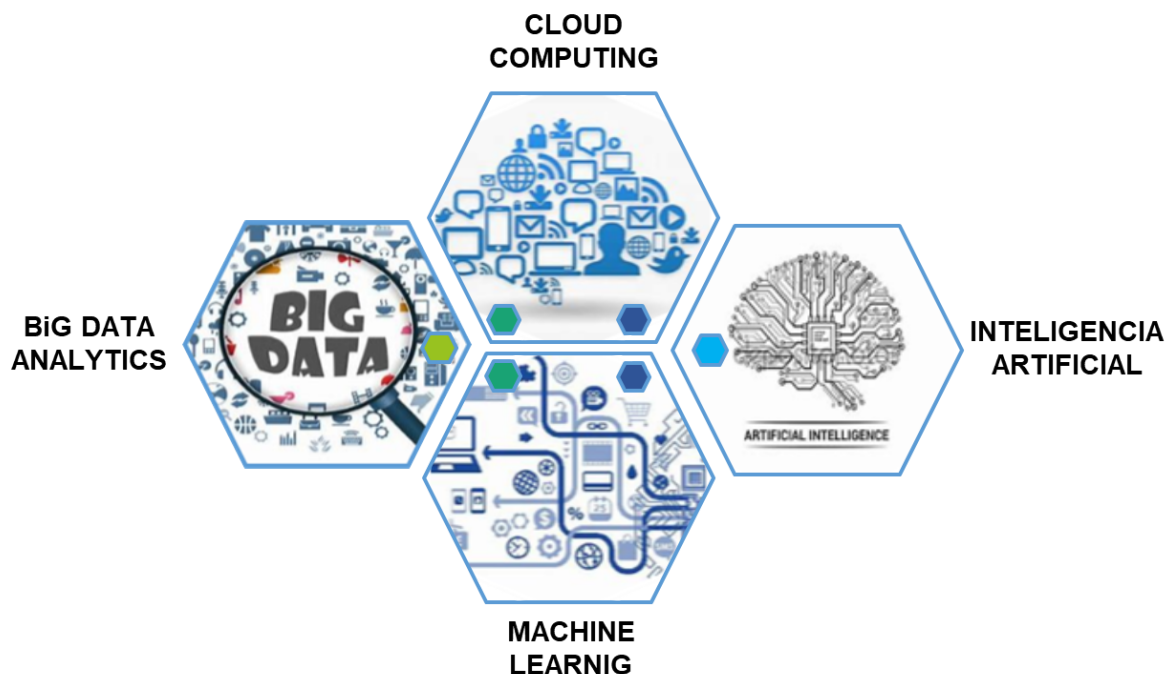
1. **Toma de decisiones basada en evidencia:** Los datos proporcionan información objetiva y basada en hechos, lo que permite tomar decisiones empresariales fundamentadas en lugar de depender únicamente de la intuición o la experiencia subjetiva.
2. **Identificación de tendencias y patrones:** Al analizar grandes volúmenes de datos, es posible identificar tendencias, patrones y correlaciones que pueden ayudar a la organización a comprender mejor el comportamiento de

sus clientes, el rendimiento de sus productos o servicios, y las oportunidades emergentes en el mercado.

3. **Mejora de la eficiencia y productividad:** El análisis de datos puede revelar procesos ineficientes o ineficaces dentro de una organización. Al identificar estas áreas problemáticas, se pueden implementar medidas correctivas para mejorar la eficiencia y la productividad, lo que puede conducir a ahorros de costos significativos.
4. **Personalización y mejora de la experiencia del cliente:** Con datos precisos sobre los clientes, sus preferencias y comportamientos, las organizaciones pueden personalizar sus productos, servicios y comunicaciones para satisfacer las necesidades individuales de los clientes. Esto puede mejorar la experiencia del cliente, fomentar la lealtad y aumentar las ventas.
5. **Detección temprana de problemas y riesgos:** El análisis de datos puede ayudar a identificar problemas o riesgos potenciales antes de que se conviertan en situaciones graves. Esto permite a la organización tomar medidas preventivas o correctivas de manera oportuna, minimizando el impacto negativo en la operación y la reputación de la empresa.
6. **Innovación y desarrollo de productos:** Al recopilar datos sobre las preferencias y necesidades de los clientes, una organización puede obtener ideas valiosas para la innovación y el desarrollo de nuevos productos o servicios. Estos datos también pueden ayudar a mejorar los productos existentes para satisfacer mejor las demandas del mercado.
7. **Ventaja competitiva:** El uso efectivo de los datos puede brindar una ventaja competitiva significativa. Al comprender mejor a los clientes, optimizar los procesos internos y aprovechar las oportunidades del mercado, una organización puede diferenciarse de sus competidores y aumentar su posición en la industria.

Estos son solo algunos ejemplos de los beneficios del uso de datos en una organización. Sin embargo, es importante destacar que la gestión y el uso adecuado de los datos también requieren consideraciones éticas y de privacidad para garantizar que se cumplan las regulaciones y proteger la información sensible de los individuos.

Sinergias con otras tecnologías (cloud, inteligencia artificial, etc.)



Sinergia entre el big data y el cloud

La computación en la nube (cloud computing) se ha convertido en una opción popular para almacenar y procesar grandes volúmenes de datos, conocidos como big data. La nube ofrece una serie de beneficios para el procesamiento de big data, que incluyen:

- **Escalabilidad:** Los proveedores de servicios en la nube ofrecen recursos escalables, lo que significa que puedes aumentar o disminuir la capacidad de almacenamiento y procesamiento según tus necesidades. Esto es especialmente útil para el big data, ya que los conjuntos de datos pueden ser enormes y requieren una infraestructura flexible.
- **Elasticidad:** La nube permite escalar horizontalmente, es decir, agregar o reducir recursos según la carga de trabajo. Esto es beneficioso para el procesamiento de big data, ya que puedes asignar recursos adicionales durante períodos de alta demanda y liberarlos cuando la carga disminuye, lo que te permite optimizar costos y rendimiento.

- **Facilidad de implementación:** Configurar y mantener una infraestructura para el procesamiento de big data puede ser complejo y costoso. Al utilizar la nube, puedes aprovechar las capacidades y servicios existentes, lo que te permite implementar rápidamente una arquitectura de big data sin invertir en infraestructura física y administración de servidores.
- **Disponibilidad y fiabilidad:** Los servicios en la nube suelen ofrecer altos niveles de disponibilidad y redundancia, asegurando que tus datos y aplicaciones de big data estén accesibles y protegidos. Los proveedores de la nube suelen tener centros de datos en múltiples ubicaciones geográficas, lo que reduce el riesgo de pérdida de datos debido a fallas o desastres.
- **Servicios de datos administrados:** Los proveedores de la nube ofrecen una amplia gama de servicios de big data administrados, como almacenamiento de datos, procesamiento distribuido, análisis y visualización. Estos servicios facilitan la gestión y el análisis de big data, ya que los proveedores se encargan de la configuración, el escalado y el mantenimiento de la infraestructura subyacente.
- **Integración con herramientas y ecosistemas:** Los proveedores de la nube suelen proporcionar integración con herramientas y ecosistemas populares de big data, como Hadoop, Spark y bases de datos NoSQL. Esto te permite aprovechar las herramientas y tecnologías existentes y utilizar los servicios de la nube como complemento para el procesamiento de big data.

Sinergia entre el big data y la inteligencia artificial

Por otro lado, la inteligencia artificial (IA) y el big data están estrechamente relacionados y se complementan entre sí. La IA se beneficia del big data al utilizar grandes conjuntos de datos para entrenar y mejorar modelos de aprendizaje automático (machine learning) y algoritmos de inteligencia artificial. A su vez, la IA también puede ayudar a extraer conocimientos y obtener valor de los datos almacenados en el big data. A continuación, se explican algunas formas en las que la IA se relaciona con el big data:

- **Análisis de texto y procesamiento del lenguaje natural (NLP):** La IA se utiliza en el análisis de texto y el procesamiento del lenguaje natural para extraer información y conocimientos de grandes volúmenes de texto no estructurado. El big data proporciona una gran cantidad de texto, como comentarios de clientes, reseñas, noticias, redes sociales, entre otros, que pueden ser analizados y procesados mediante técnicas de IA como NLP. Esto

permite a las organizaciones comprender las opiniones de los clientes, identificar tendencias y realizar análisis de sentimiento.

- **Recomendaciones personalizadas:** La IA se utiliza en sistemas de recomendación para ofrecer recomendaciones personalizadas a los usuarios en función de su comportamiento, preferencias y datos históricos. El big data, que incluye información sobre el comportamiento de los usuarios y sus interacciones, se utiliza para entrenar modelos de IA que pueden generar recomendaciones precisas y relevantes. Esto se ve comúnmente en plataformas de comercio electrónico, servicios de transmisión de contenido y redes sociales.
- **Automatización de tareas:** La IA puede automatizar tareas que tradicionalmente requerían intervención humana, como clasificación y etiquetado de datos, detección de anomalías, segmentación de clientes y procesamiento de imágenes. El big data proporciona los datos necesarios para entrenar y mejorar los algoritmos de IA que realizan estas tareas automatizadas. Esto ayuda a acelerar el procesamiento de grandes volúmenes de datos y reduce la carga de trabajo manual.

Machine Learning: algoritmos y aplicaciones

Existen varios tipos de algoritmos de Machine Learning, cada uno con características y aplicaciones específicas. Aquí están algunos de los tipos más comunes y sus aplicaciones:

Aprendizaje supervisado (Supervised Learning): Este tipo de algoritmo se basa en conjuntos de datos etiquetados, donde se conocen las respuestas correctas. El algoritmo aprende a partir de estos datos y puede hacer predicciones o clasificar nuevos ejemplos en función de las características proporcionadas. Ejemplos de aplicaciones incluyen:

- Clasificación de correo electrónico como spam o no spam.
- Predicción de precios de viviendas basado en características.
- Diagnóstico médico basado en síntomas y resultados de pruebas.

Aprendizaje no supervisado (Unsupervised Learning): Este tipo de algoritmo se utiliza cuando no hay etiquetas o respuestas predefinidas en los datos. El algoritmo busca patrones y estructuras inherentes en los datos sin una guía explícita.

Ejemplos de aplicaciones incluyen:

- Agrupamiento de clientes para segmentación de mercado.
- Detección de anomalías en datos de sensores.
- Reducción de dimensionalidad para visualización de datos.

Aprendizaje por refuerzo (Reinforcement Learning): En este enfoque, un agente de aprendizaje interactúa con un entorno y toma decisiones en función de las acciones tomadas y las recompensas recibidas. El objetivo es maximizar la recompensa acumulada a lo largo del tiempo. Ejemplos de aplicaciones incluyen:

- Control de robots y sistemas autónomos.
- Optimización de carteras de inversión.
- 1. Juegos de estrategia, como el ajedrez o Go.

Aprendizaje semisupervisado (Semi-Supervised Learning): Este enfoque combina elementos de aprendizaje supervisado y no supervisado. Se utilizan conjuntos de datos etiquetados y no etiquetados para mejorar la precisión y escalabilidad del modelo. Ejemplos de aplicaciones incluyen:

- Análisis de texto y etiquetado automático de documentos.
- Clasificación de imágenes con datos parcialmente etiquetados.
- Detección de fraudes con datos transaccionales.

Estos son solo algunos ejemplos de algoritmos y sus aplicaciones en Machine Learning. Además, existen otros enfoques y técnicas como el aprendizaje profundo (Deep Learning), el aprendizaje por transferencia (Transfer Learning), entre otros, que también tienen aplicaciones en diversas áreas. Cada tipo de algoritmo tiene fortalezas y limitaciones, por lo que la elección del algoritmo adecuado depende del problema específico y los datos disponibles.

Herramientas de visualización de datos

Tableau



Tableau es una plataforma que permite a las empresas y organizaciones analizar, visualizar y compartir datos de manera efectiva. A continuación, te explicaré cómo se aplica en una empresa.

Tableau Software se utiliza en las empresas para analizar grandes conjuntos de datos y convertirlos en información visualmente atractiva y comprensible. Algunas de las formas en que se aplica en la empresa son las siguientes:

- **Visualización de datos:** Tableau permite crear gráficos interactivos, tablas, mapas y otros tipos de visualizaciones para explorar y comprender los datos. Estas visualizaciones ayudan a identificar patrones, tendencias y relaciones ocultas en los datos, lo que puede ser útil para tomar decisiones informadas.
- **Análisis de datos:** La plataforma de Tableau ofrece herramientas de análisis avanzadas que permiten a los usuarios realizar análisis complejos de datos de manera intuitiva. Los usuarios pueden aplicar filtros, realizar cálculos,

crear grupos y segmentar los datos para obtener información más detallada y profunda.

- **Creación de paneles de control:** Tableau permite crear paneles de control personalizados que muestran información clave de manera visual. Estos paneles pueden incluir múltiples visualizaciones y se actualizan automáticamente a medida que los datos subyacentes cambian. Los paneles de control son útiles para monitorear el rendimiento empresarial en tiempo real y tomar decisiones basadas en datos actualizados.
- **Compartir información y colaboración:** Tableau facilita la colaboración en la empresa al permitir compartir visualizaciones, paneles y análisis con otros miembros del equipo. Los usuarios pueden publicar sus creaciones en la web o en un servidor local para que otros puedan acceder a ellas y realizar comentarios o aportaciones.
- **Integración con otras herramientas:** Tableau se integra con una variedad de herramientas y plataformas de datos, como bases de datos, sistemas de almacenamiento en la nube y aplicaciones empresariales. Esto permite a las empresas utilizar Tableau en conjunto con otras soluciones existentes y aprovechar al máximo sus datos.

PowerBI



Power BI es una plataforma de análisis empresarial desarrollada por Microsoft. Permite a las organizaciones recopilar, analizar y visualizar datos desde una amplia variedad de fuentes, y compartir información y conocimientos de manera fácil y efectiva. Power BI se compone de varios componentes que trabajan juntos para brindar capacidades de inteligencia empresarial.

Comparte las mismas formas de aplicación en empresa que tableau anteriormente descrito.

Looker



Looker es una plataforma de análisis y visualización de datos basada en la nube. Fue adquirida por Google en 2020 y ahora forma parte de la plataforma Google Cloud. Looker permite a las empresas explorar, analizar y compartir datos de manera colaborativa, brindando información en tiempo real y facilitando la toma de decisiones basadas en datos.

Algunos de los casos de uso más comunes de Looker son:

- **Análisis de negocios:** Looker es utilizado para el análisis de negocios, permitiendo a las organizaciones obtener una visión profunda de su rendimiento financiero, operativo y estratégico. Se pueden crear informes y paneles personalizados para monitorear métricas clave, identificar tendencias y patrones, y tomar decisiones basadas en datos.
- **Inteligencia de marketing:** Las empresas utilizan Looker para analizar y visualizar datos de marketing, como métricas de rendimiento de campañas publicitarias, análisis de segmentación de clientes y análisis de atribución. Esto ayuda a las organizaciones a optimizar sus estrategias de marketing y tomar decisiones informadas para maximizar el retorno de inversión.
- **Analítica de productos:** Looker es utilizado para analizar el rendimiento de productos y servicios, lo que incluye seguimiento de ventas, análisis de uso de productos, retroalimentación de clientes y análisis de satisfacción del cliente. Esto ayuda a las empresas a comprender mejor el comportamiento de los clientes y tomar decisiones basadas en datos para mejorar la experiencia del usuario y la calidad del producto.

- **Analítica de operaciones:** Looker se utiliza para analizar y optimizar las operaciones empresariales, como la gestión de inventario, el seguimiento del rendimiento de la cadena de suministro, el análisis de costos y la optimización de procesos. Esto permite a las organizaciones identificar áreas de mejora, reducir costos y aumentar la eficiencia operativa.
- **Analítica de recursos humanos:** Looker ayuda a las organizaciones a analizar datos de recursos humanos, como la gestión del talento, la planificación de la fuerza laboral, el análisis de la diversidad y la evaluación del rendimiento del personal. Esto permite a las empresas tomar decisiones estratégicas relacionadas con la gestión de recursos humanos y el desarrollo de talento.

MicroStrategy

MicroStrategy es una empresa de software de inteligencia empresarial con sede en Estados Unidos. Ofrece una plataforma integral de análisis y visualización de datos que permite a las organizaciones recopilar, analizar y presentar datos de manera efectiva para la toma de decisiones.

En el contexto español, MicroStrategy se encuentra en grandes empresas multinacionales que fueron pioneras en el análisis de datos empresarial.

