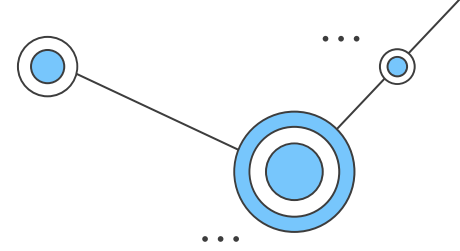




Calidad de datos

Problemas comunes

¿Qué es calidad de datos?



Es una medida del estado de los mismos basada en factores como la exactitud, completitud, consistencia, credibilidad y actualidad.



La condición en los sistemas empresariales es cada vez más importante a medida que el procesamiento de datos se convierte en una parte integral de las operaciones y que las soluciones analíticas se utilizan cada vez más como herramienta de apoyo en la toma de decisiones.



**David Loshin, President of
Knowledge Integrity, Inc:**

"Para ser capaces de relacionar los problemas de la calidad de datos con su impacto en el negocio, tenemos que ser capaces de clasificar tanto nuestras expectativas de calidad de datos, como los criterios de impacto en la empresa."





Métricas para medir la calidad de datos

...

1. **Disponibilidad:**

- Accesibilidad
- Oportunidad

2. **Usabilidad:**

- Credibilidad

3. **Confiabilidad:**

- Exactitud
- Consistencia
- Integridad
- Compleitud

4. **Pertinencia:**

- Conveniencia

5. **Calidad de presentación:**

- Legibilidad





Matriz de riesgos de datos

Problema	Edicion	Riesgo Empresarial	Cuantificador	Coste (anual)
Este es el problema de calidad de los datos que reside en su conjunto de datos.	Estos son los diversos problemas que pueden surgir debido al problema de los datos.	Este es el impacto que el problema puede tener en la empresa.	Esto cuantifica el impacto en términos de una medida empresarial.	Esto proporciona una estimación periódica de los costes incurridos debido al impacto empresarial.
Ejemplo				
Nombre del cliente e información de contacto mal escritos	Registros duplicados creados para el mismo cliente	Servicio de atención al cliente: Aumento del número de llamadas entrantes	Mayor tiempo del personal	Se necesitan 30.000 dólares más de tiempo del personal
		Servicio al cliente: Disminución de la satisfacción del cliente	Reducción de pedidos, pérdida de clientes	Unos 500 pedidos menos este año (en comparación con la estimación)
		Cuentas: Impacto en el flujo de caja	Aumento de la volatilidad del flujo de caja	No se puede confiar en el flujo de caja estimado alrededor del 20% de las veces
		Cuentas: Aumento de la demanda de auditoría	Mayor tiempo del personal	Se necesitan 20.000 dólares más de tiempo del personal

Problemas en calidad de datos

Datos Sesgados

Demasiados
Datos

Silos de Datos

Problemas de
Etiquetado

Datos Inexactos

Datos
Inconsistentes

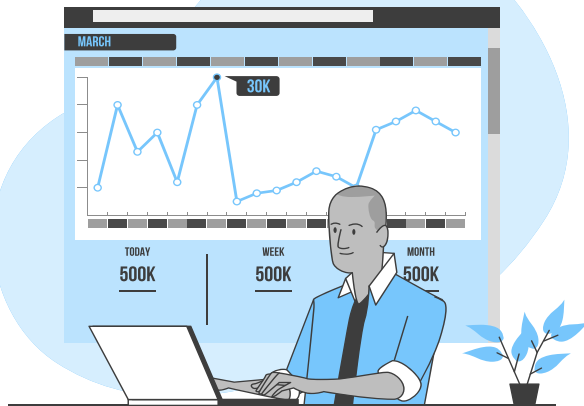
Pocos Datos



Datos Sesgados

Los datos pueden seleccionarse de conjuntos de datos más grandes de formas que no transmitan adecuadamente el mensaje del conjunto de datos más amplio.

Quizás hay algunos problemas con la forma en que se recopilan o generan los datos que dan como resultado un resultado final sesgado.



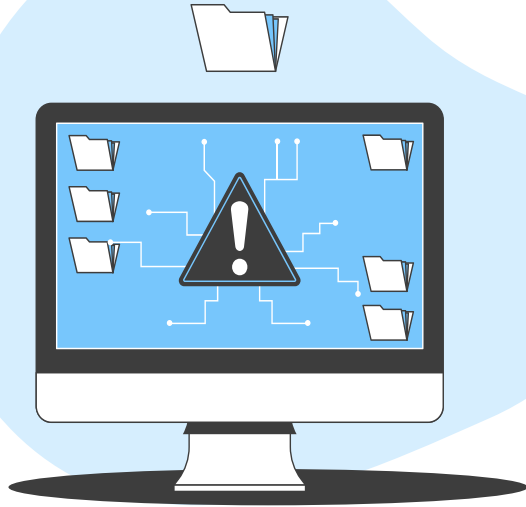
Datos Inexactos

Estos problemas de datos pueden variar desde datos incorrectos en el origen hasta datos que no se han limpiado o preparado correctamente. Es posible que los datos estén en los campos incorrectos o que se hayan aplicado etiquetas incorrectas.



Solución

Limpieza y depuración de datos inexactos. Los enfoques tradicionales simplemente no escalan, lo que ha resultado en nuevas herramientas impulsadas por IA para ayudar a detectar y limpiar problemas de datos.



En muchas circunstancias, los mismos registros pueden existir varias veces en diferentes conjuntos de datos pero con diferentes valores, lo que da como resultado inconsistencias. Los datos duplicados son uno de los mayores problemas para las empresas basadas en datos.

**Datos
Inconsistentes**

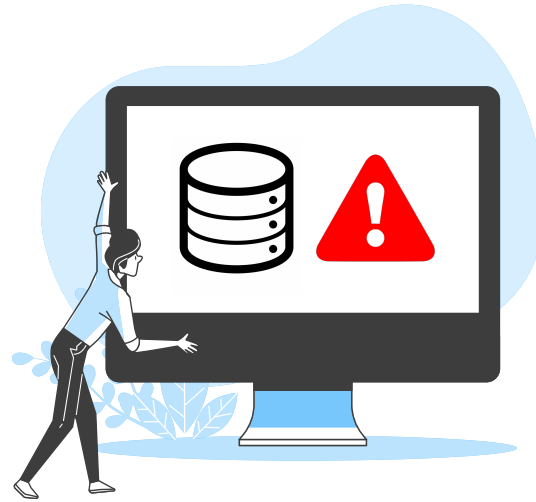
Pocos Datos

Si bien entrenar un modelo en un pequeño conjunto de datos puede producir resultados aceptables en un entorno de prueba, llevar este modelo de la prueba de concepto o la etapa piloto a la producción generalmente requiere más datos.



En general, los conjuntos de datos pequeños pueden producir resultados de baja complejidad, sesgados o demasiado ajustados y no serán precisos al trabajar con datos nuevos.

Demasiados Datos



Si bien puede parecer que demasiados datos nunca pueden ser algo malo, la mayoría de las veces, una buena parte de **los datos no es utilizable o relevante**.

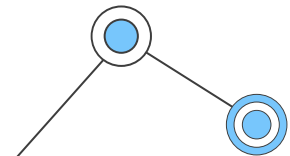
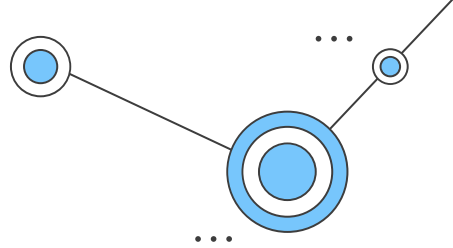
Tener que pasar para **separar los datos útiles** de este gran conjunto de datos **desperdicia** los **recursos** de la organización.

Silos de Datos

Un silo de datos es donde solo un cierto grupo o un número limitado de personas en una organización tienen acceso a un conjunto de datos.



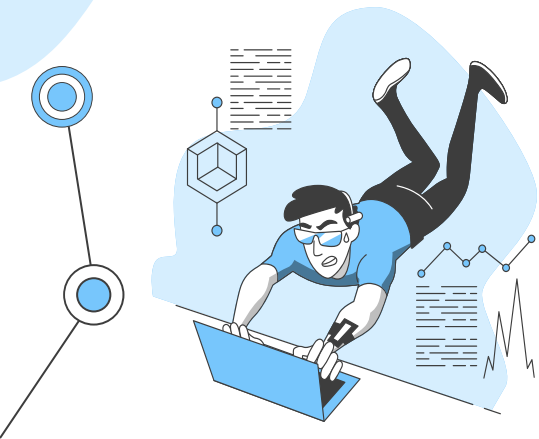
Los silos de datos pueden resultar de varios factores, incluidos desafíos técnicos o restricciones en la integración de conjuntos de datos, así como problemas con el control de acceso de seguridad o de propiedad de los datos



Problemas de Etiquetado

El etiquetado de datos es una tarea difícil, que a menudo requiere recursos humanos para colocar metadatos en una amplia gama de tipos de datos. Esto puede resultar complejo y costoso.

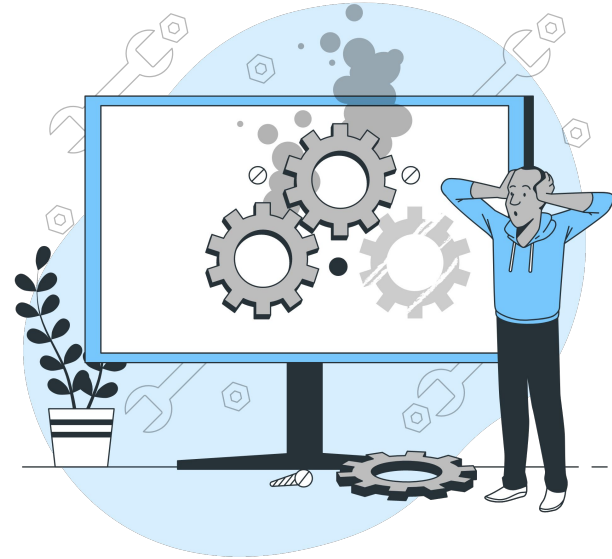
Los datos etiquetados con precisión garantizan que los sistemas de aprendizaje automático establezcan modelos confiables para el reconocimiento de patrones. Los datos etiquetados de buena calidad son primordiales para entrenar con precisión al sistema de IA sobre los datos que se están alimentando.



Riesgos empresariales

Oportunidades Perdidas

Una empresa es propensa a perder oportunidades en múltiples frentes si tiene una mala calidad de datos en conjuntos de datos dispares.



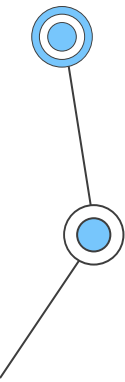
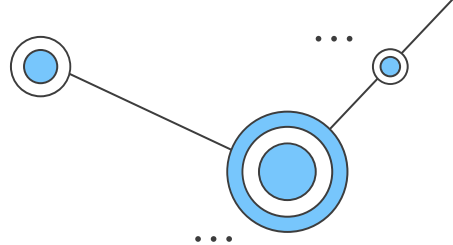
Pérdida de ingresos



Los datos incompletos o incorrectos pueden hacerle perder clientes potenciales e incurrir en pérdidas de ingresos como resultado.

Insatisfacción del cliente

Para lograrlo, las empresas utilizan una gran cantidad de datos generados por los clientes para comprender su comportamiento y sus preferencias. Si estos datos tienen graves defectos, es obvio que acabará infiriendo detalles erróneos sobre sus clientes o compradores potenciales.



Daño a la reputación

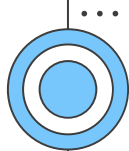
La mala información del producto es una de las principales razones por las que los clientes devuelven los productos comprados. El producto no era como se comercializaba. La mala calidad de los datos puede costarle más que los productos devueltos.



Falta de cumplimiento

Las normas de cumplimiento de datos están obligando a las empresas a revisar sus estrategias de gestión de datos.

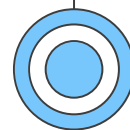




...



...



...



joinmyquiz.com



...

...

...