



De mirar el pasado a
anticipar el futuro



Machine Learning con Python para la Toma de Decisiones Empresariales

Jorge Israel Frometa Moya

Calidad del Dato y Metodología



- Metodología: El motor invisible y diferencial del Machine Learning en los negocios.
- Entender "Garbage In, Garbage Out" es fundamental para decisiones rentables.
- No importa qué tan avanzado sea tu algoritmo; si los datos están mal, la decisión será errónea.

Metodologías para ML en negocios:

- CRISP-DM
- TDSP (Team Data Science Process de Microsoft)
- Agile Data Science (iteraciones rápidas)



El Ciclo de Vida de un Proyecto (CRISP-DM)



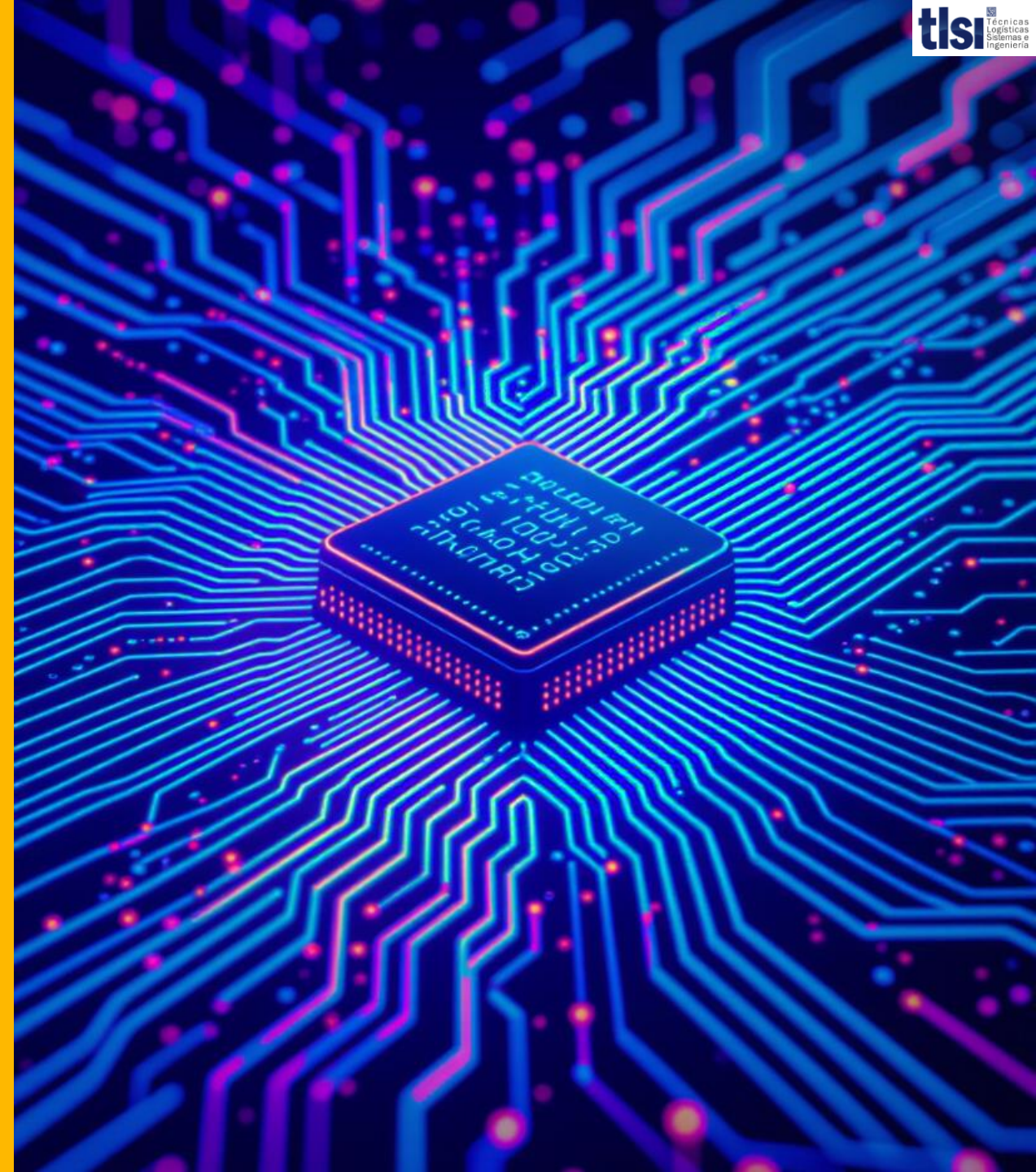
El ML no es un proceso lineal, sino iterativo y de mejora continua.

1. **Comprensión de Negocio:** ¿Qué queremos mejorar?
2. **Comprensión de Datos:** ¿Tenemos lo que necesitamos?
3. **Preparación de Datos:** Limpieza y transformación (donde pasamos el 80% del tiempo).
4. **Modelado:** Aplicar el algoritmo.
5. **Evaluación:** ¿Aporta valor económico?
6. **Despliegue:** Ponerlo en marcha.

El Problema de "Garbage In, Garbage Out" (GIGO)

Puntos clave:

- **Valores nulos:** ¿El cliente no tiene edad o no la quiso dar?
- **Outliers (Valores atípicos):** Una venta de 1.000.000€ en un negocio de tickets de 10€ ¿es un error o un cliente VIP?
- **Duplicados:** Sobreestiman la importancia de ciertos patrones.



Los 4 Pilares de la Calidad del Dato:

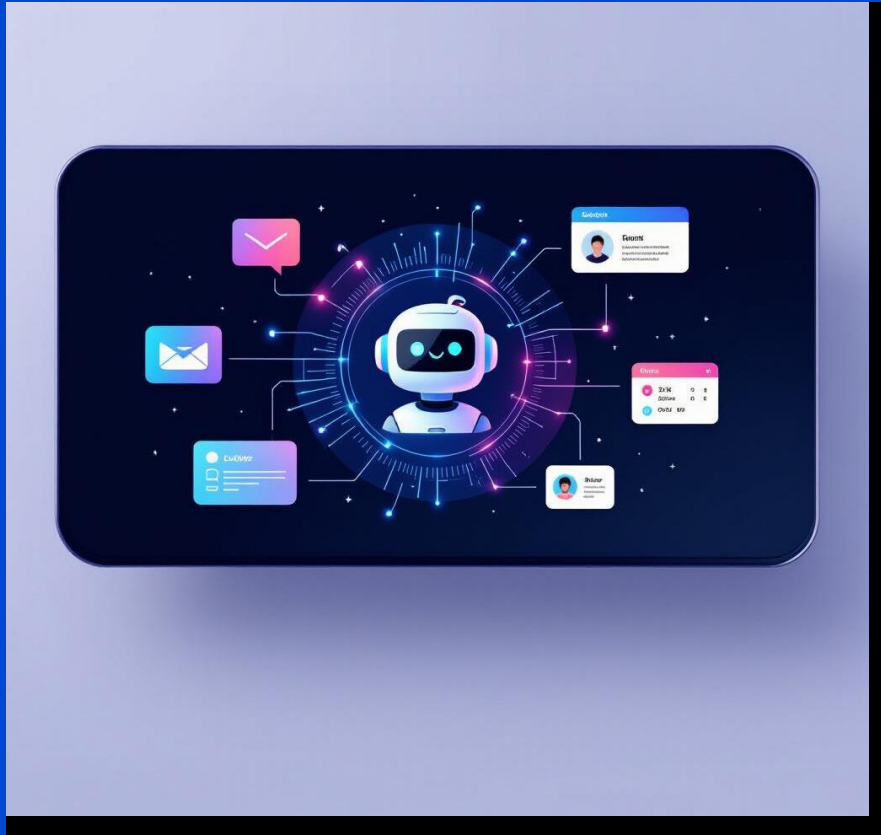
1. Integridad (¿Faltan datos?)
2. Consistencia (¿Los datos siguen reglas?)
3. Precisión (¿Los datos son correctos?)
4. Actualidad (¿Los datos están vigentes?)



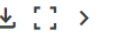
Caso práctico: OnlineRetail



Datos reales



OnlineRetail.csv (45.58 MB)




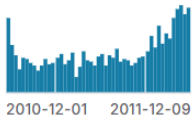

Detail Compact Column

8 of 8 columns

About this file

[Suggest Edits](#)

This file does not have a description yet.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice
25900 unique values	4070 unique values	4224 unique values			
			-80995 81.0k	2010-12-01 2011-12-09	-11.1k
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39

Conceptos claves

DEL DATO A LA DECISIÓN

Dato Crudo → Información → Conocimiento → Decisión

EL ROI DE LA CALIDAD

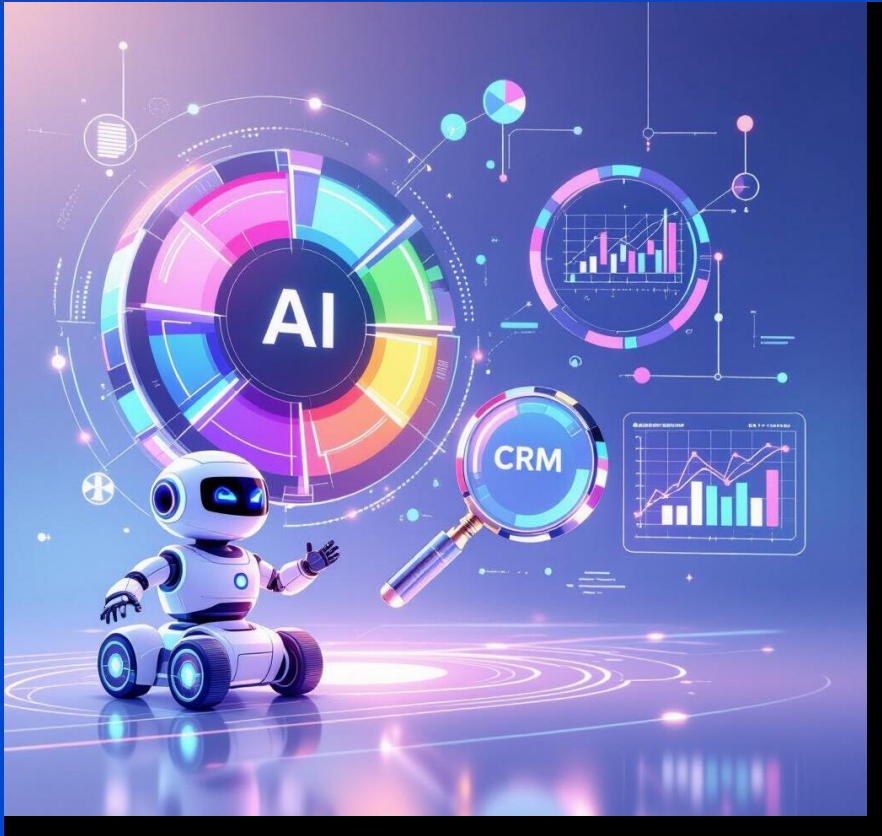
Inversión en limpieza × 10 = Retorno en mejores decisiones

PREGUNTAS CLAVES

1. ¿En qué confío? → Integridad
2. ¿Qué me estoy perdiendo? → Consistencia
3. ¿Dónde me puedo equivocar? → Precisión



Advertencia temprana



Riesgos asociados:

1. Paralización por análisis → Nunca llegar a producción.
2. Costo oculto → Limpieza manual no escalable
3. Falsa confianza → "Los datos están limpios, el modelo es perfecto"
4. Desalineación → Limpiar lo técnico vs. lo estratégico