



CURSO DE INTELIGENCIA ARTIFICIAL MÁLAGA, 30 mayo a 20 julio 2021

**Modelar sobre plataforma de ML
lunes 28 de junio de 2021**

Andrés González



Índice de hoy

- _01 Conexión con negocio.
- _02 Automatizaciones con Python.
- _03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.
- _04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Índice de hoy

_01 Conexión con negocio.

_02 Automatizaciones con Python.

_03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.

_04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Conexión con negocio

- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

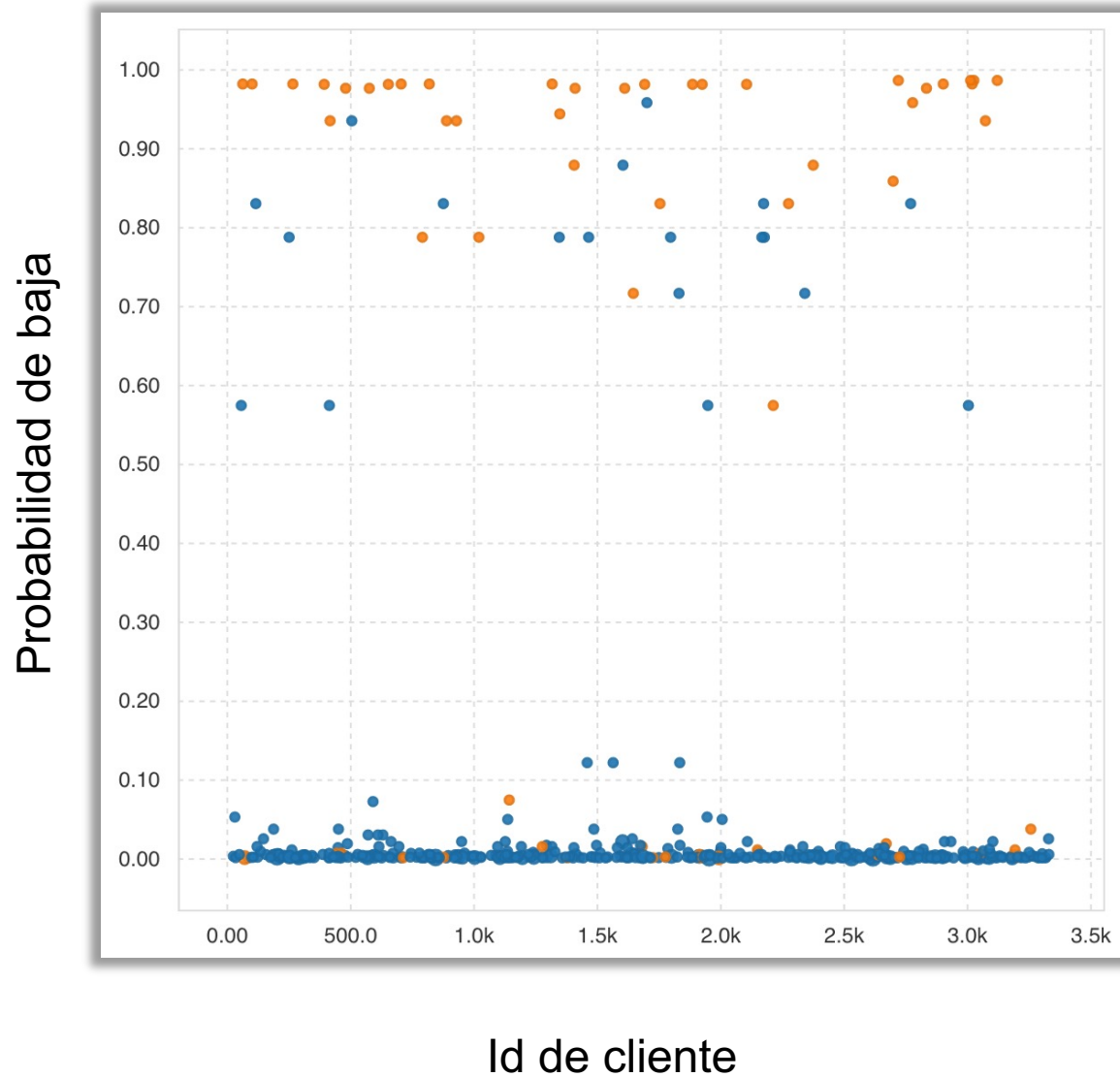
Conexión con negocio

- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

Conexión con negocio

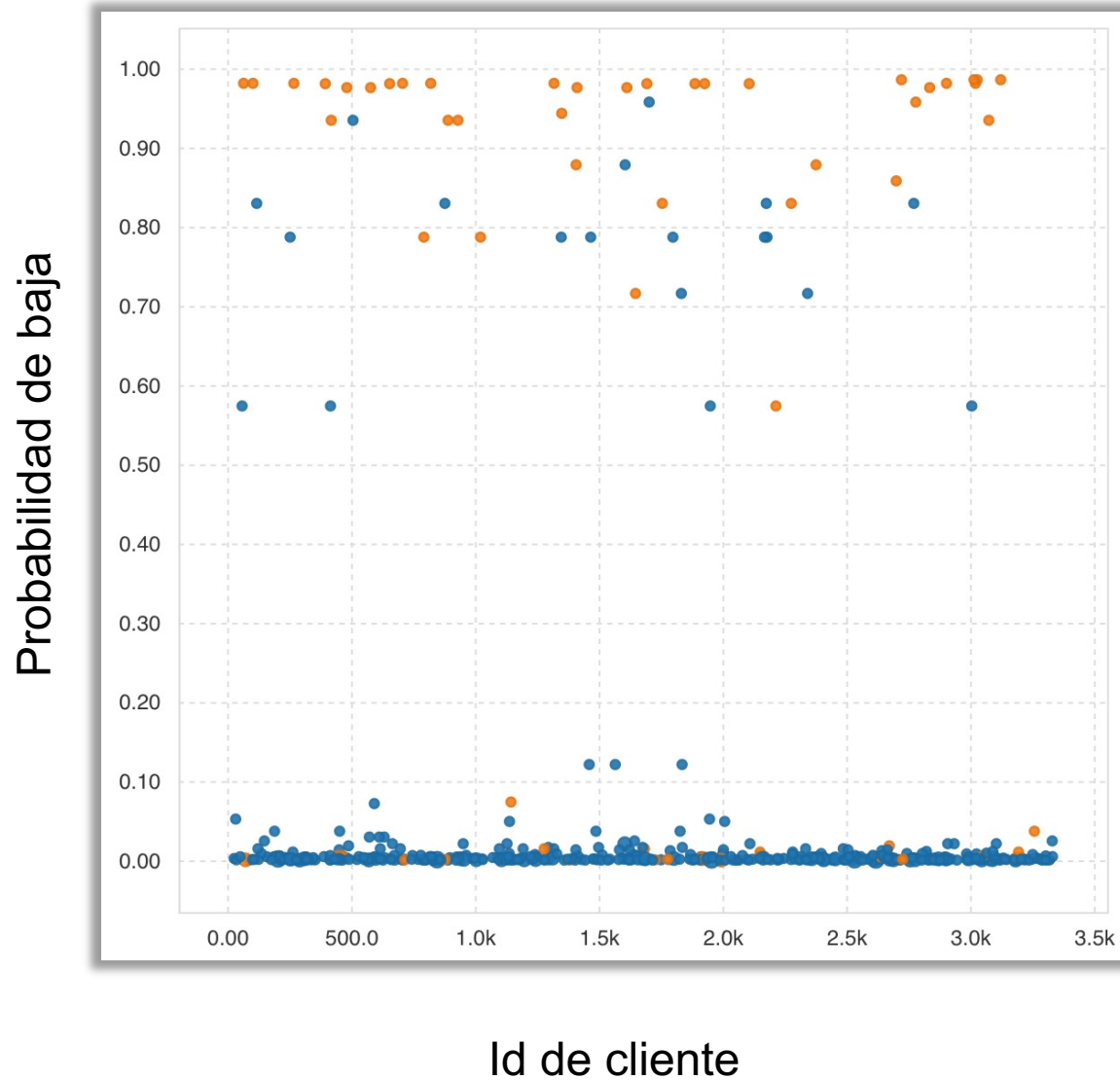
- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

Conexión con negocio



Conexión con negocio

- No baja
- Sí baja

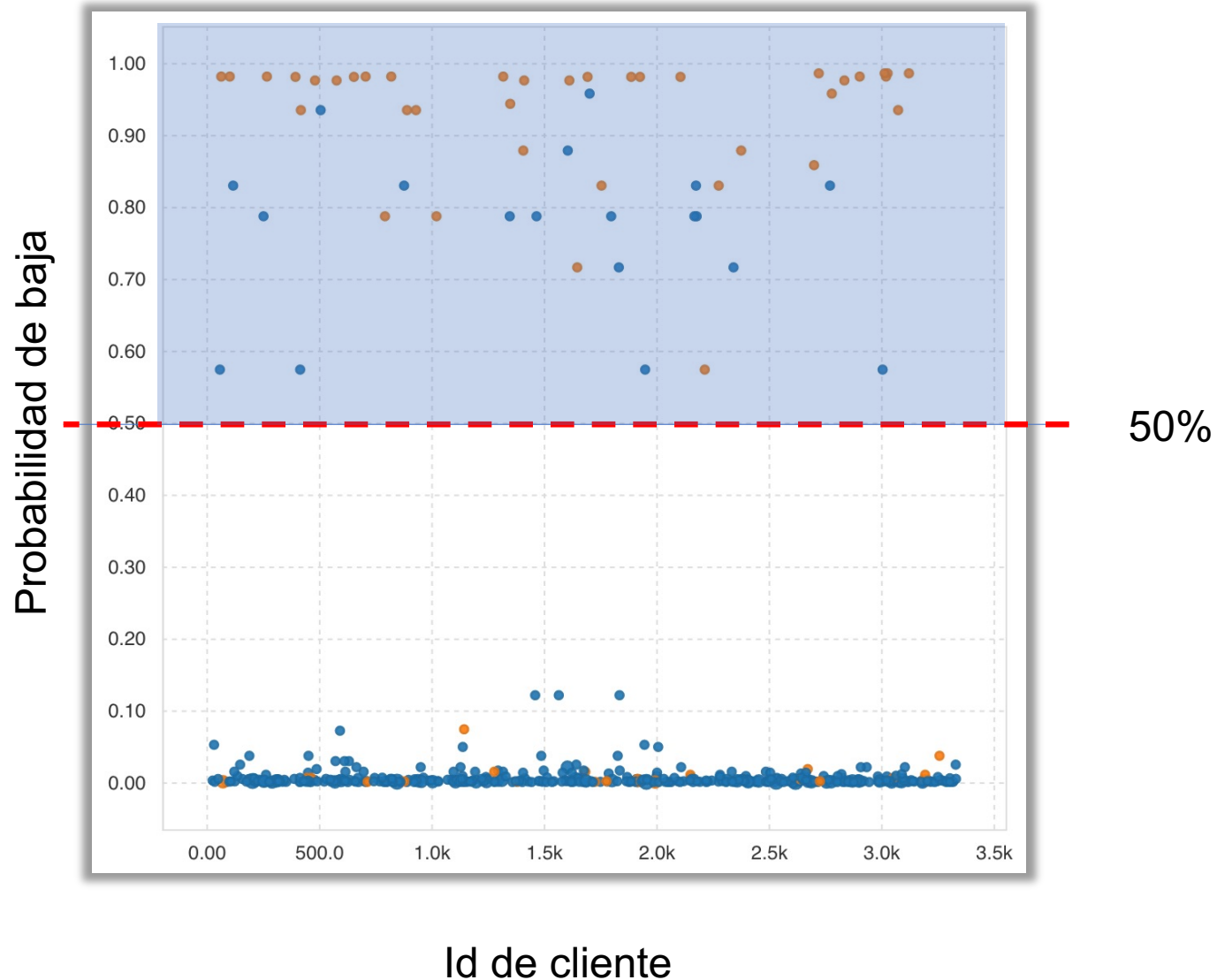


Conexión con negocio

- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

Conexión con negocio

- No baja
- Sí baja



Conexión con negocio

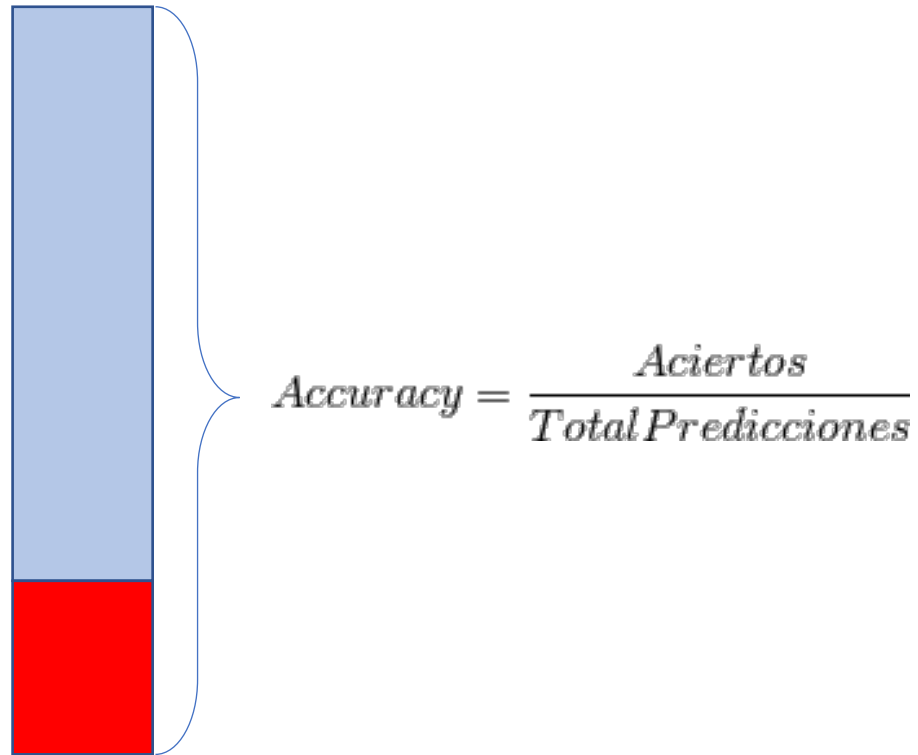
- No baja
- Sí baja



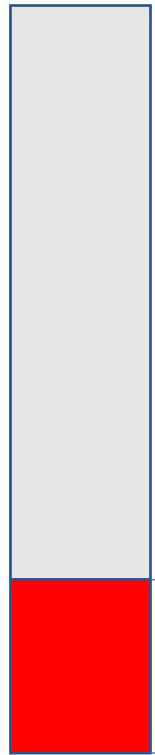
Conexión con negocio

- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

Accuracy. Indicador sobre todas las predicciones



Precisión:
Zoom a solo unas
predicciones



$$Precision = \frac{Aciertos}{Total\ Predicciones\ Rojas}$$

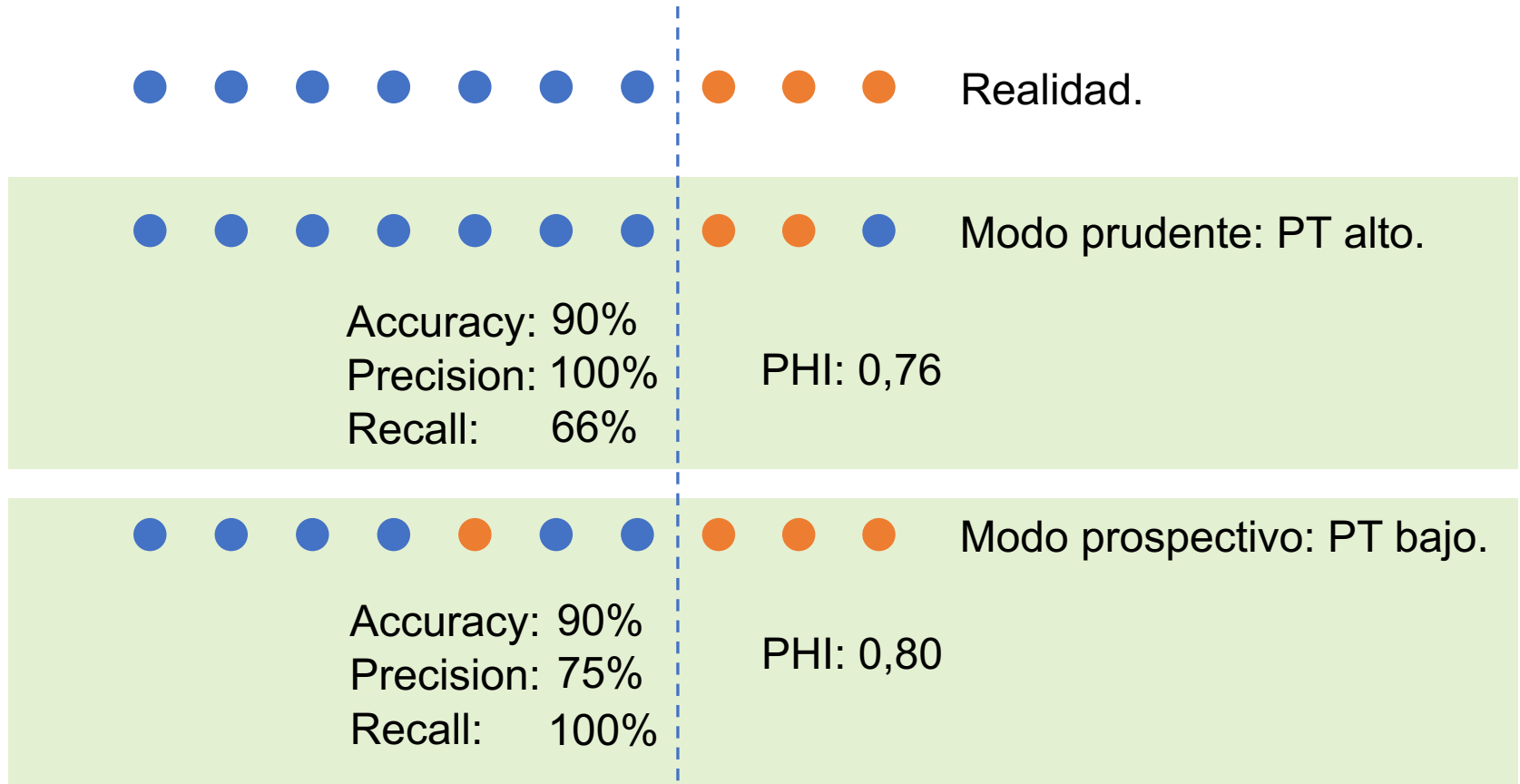
Recall:
Zoom sobre la realidad



$$Recall = \frac{Aciertos}{Total\ Realidad\ Rojas}$$

Conexión con negocio

- No baja
- Sí baja



Accuracy: sobre total predicciones

Precision: sobre las que he dicho que sí

Recall: sobre las que en realidad han sido sí

Conexión con negocio

- No baja
- Sí baja

● ● ● ● ● ● ● ● ● Realidad.

● ● ● ● ● ● ● ● ● PT = 100%

Accuracy: 70%

Precision: ?

Recall: 0%

PHI: DIV!0

● ● ● ● ● ● ● ● ● PT = 0%

Accuracy: 30%

Precision: 30%

Recall: 100%

PHI: DIV!0

Conexión con negocio

- Nuevo modelo “abandono” de clientes.
- Límite de probabilidad.
- Accuracy, precision y recall.
- ROI de la acción comercial.

Conexión con negocio

¿Cómo elegir el PT?

Matriz de costes:

Cuánto vale un TP, TN, FP, FN.

Indicador	Datos
Gasto mensual del cliente	100€
Coste de la acción	10€
Verdadero Positivo	90€
Falso Positivo	-10€
Verdadero Negativo	0€
Falso Negativo	-100€

Conexión con negocio

Costes		PT 50	
TP	80 €	TP	59 4.720 €
FP	-20 €	FP	27 -540 €
TN	0 €	TN	550 0 €
FN	-100 €	FN	31 -3.100 €
			1.080 €
https://bigml.com/shared/evaluation/0o2CloW3CunaEheMCIGZG1vReDk			
		PT 80	
		55	4.400 €
		11	-220 €
		566	0 €
		35	-3.500 €
			680 €

Índice de hoy

- _01 Conexión con negocio.**
- _02 Automatizaciones con Python.
- _03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.
- _04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Índice de hoy

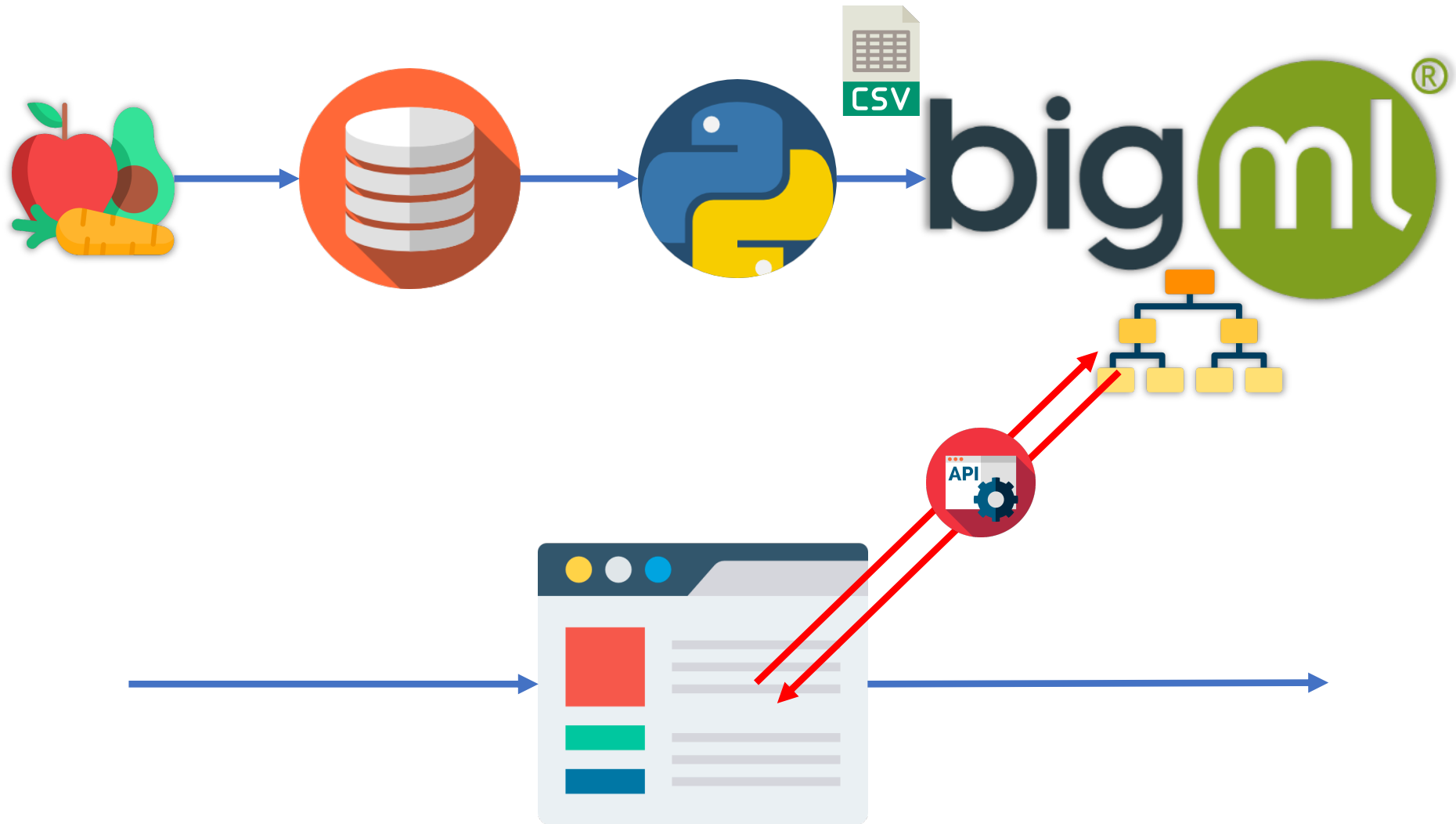
__01 Conexión con negocio.

__02 **Automatizaciones con Python.**

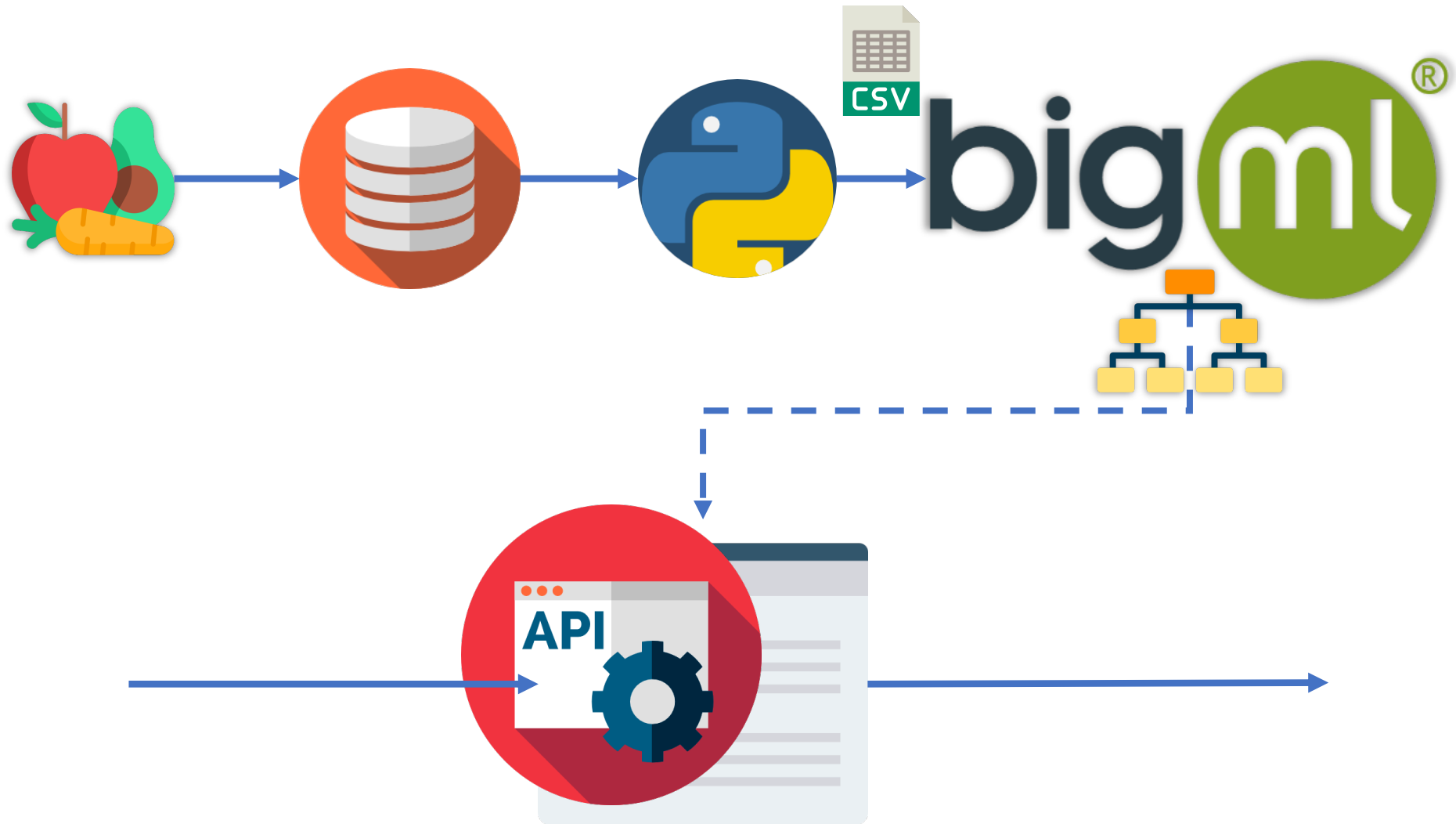
__03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.

__04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Producción modelo Direct-API



Producción modelo API local



Índice de hoy

__01 Conexión con negocio.

__02 **Automatizaciones con Python.**

__03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.

__04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Índice de hoy

- _01 Conexión con negocio.
- _02 Automatizaciones con Python.
- _03 **Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.**
- _04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Algoritmos en BigML

Árboles de decisión, Ensembles, Regresión Logística Redes Neuronales, Series Temporales

Aprendizaje supervisado

Clasificación

- Riesgo de cancelación
- Análisis de crédito
- Análisis de riesgo
- Análisis de sentimiento
- Análisis de campañas
- Mantenimiento predictivo
- Next Best Offer (“NBO”)
- Priorización de contenido
- Diagnóstico de pacientes
- Análisis de retención
- Reclutamiento por objetivos
- Análisis de fraude

Regresión

- Lifetime Value
- Publicidad predictiva
- Optimización de precios
- Estimación de ventas

Sólo para problemas de clasificación

Aprendizaje no supervisado

Clusters

- Análisis de fraude
- Segmentación de mercado
- Segmentación de clientes
- Gestión de portafolios

Detector de anomalías

- Análisis de fraude
- Limpieza de datos
- Detección de intrusos
- Autenticación

Asociaciones

- Market-Basket Analysis
- Patrones de UX
- Bioinformática
- Detección de incidentes

Modelado de tópicos

- Análisis de datos de texto.
- Evaluación de la similitud de documentos de texto

Descubriendo Asociaciones

- Qué es
- Casos de uso:
 - Market Basket Analysis
 - Recomendaciones
 - Patrones de uso web
 - Detectores de intrusos
 - Detección de fraude
 - Bioinformática
 - Factores de riesgo médicos

Entender los indicadores

Antecedente	Consecuente	Coverage (penetración)	Support	Confidence	Leverage	Lift
Atún	Aceitunas	10,4%	4,23%	40,42%	1,56%	1,58

Coverage: porcentaje de tickets con “**Atún**”.

Support: porcentaje de tickets con “**Atún**” y “**Aceitunas**”.

Confidence: de las veces que se compra “**Atún**”, qué porcentaje se compra también “**Aceitunas**”.

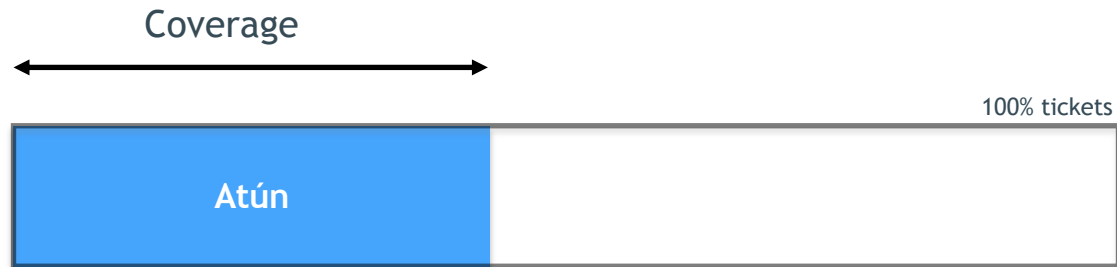
Leverage: diferencia entre la veces que “**Atún**” y “**Aceitunas**” se compran juntos y las veces que se comprarían si fueran sucesos no relacionados. Un “cero” indica que la compra simultánea de los 2 productos es aleatoria. Valores > 0 indican asociación positiva.

Lift: número de veces que es más probable que se compre “**Aceitunas**” cuando se compra “**Atún**”. Un “uno” indica que no hay asociación. Cuanto mayor es el valor, mayor es la fuerza de la asociación. “**Premia**” las asociaciones que suceden menos, pero tienen “fuerza”.

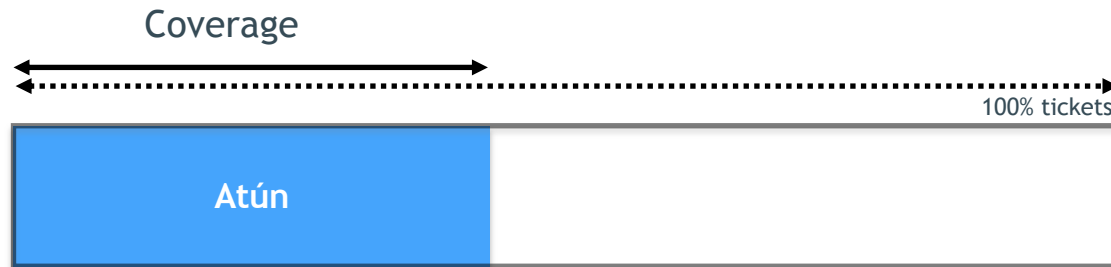
Entender los indicadores



Entender los indicadores



Entender los indicadores



Antecedente	Consecuente	Coverage (penetración)	Support	Confidence	Leverage	Lift
Atún	Aceitunas	10,4%	4,23%	40,42%	1,56%	1,58

Entender los indicadores



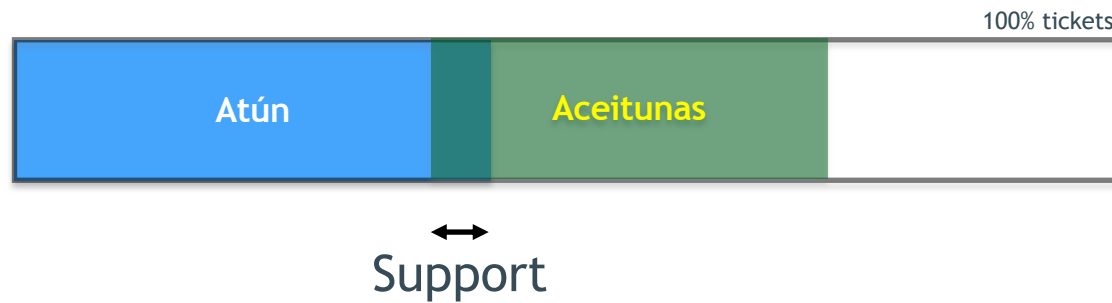
Entender los indicadores



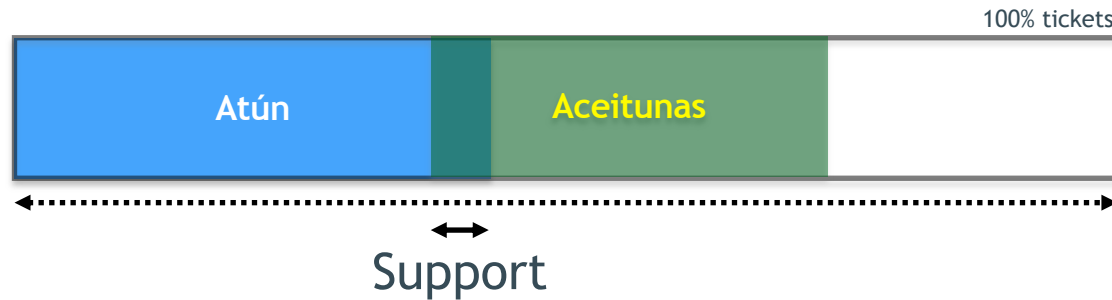
Entender los indicadores



Entender los indicadores

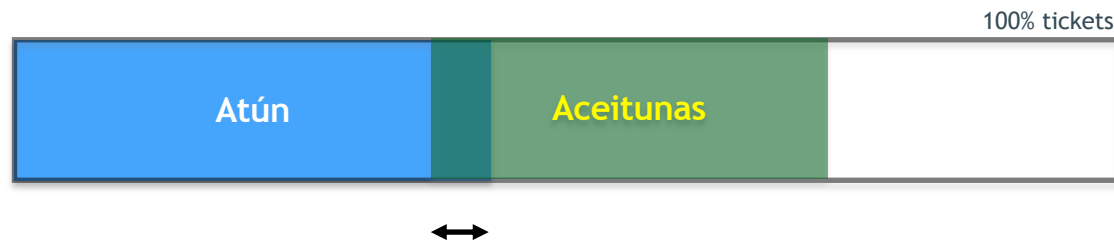


Entender los indicadores

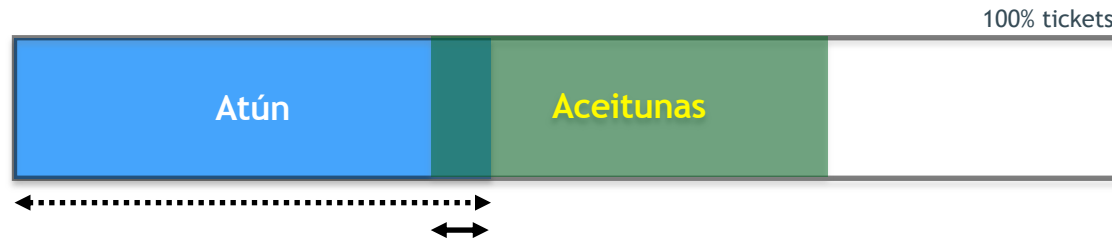


Antecedente	Consecuente	Coverage (penetración)	Support	Confidence	Leverage	Lift
Atún	Aceitunas	10,4%	4,23%	40,42%	1,56%	1,58

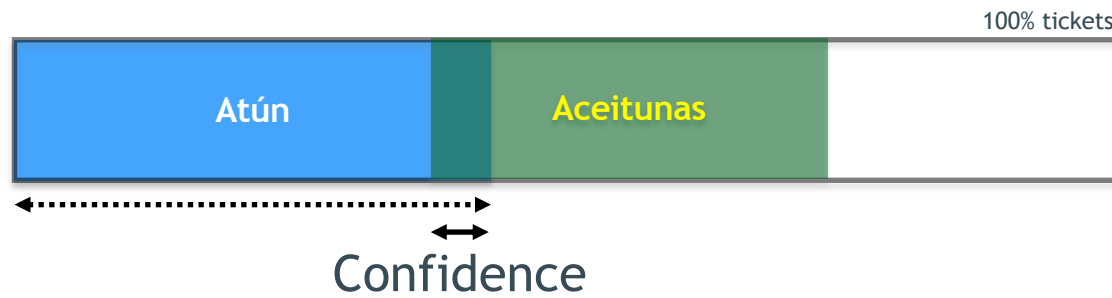
Entender los indicadores



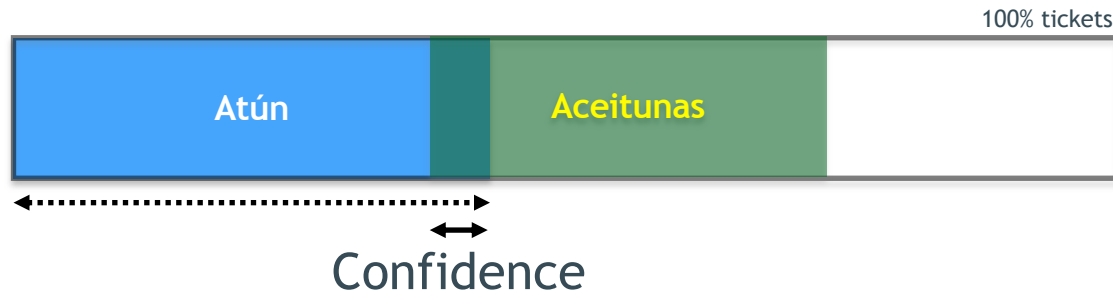
Entender los indicadores



Entender los indicadores

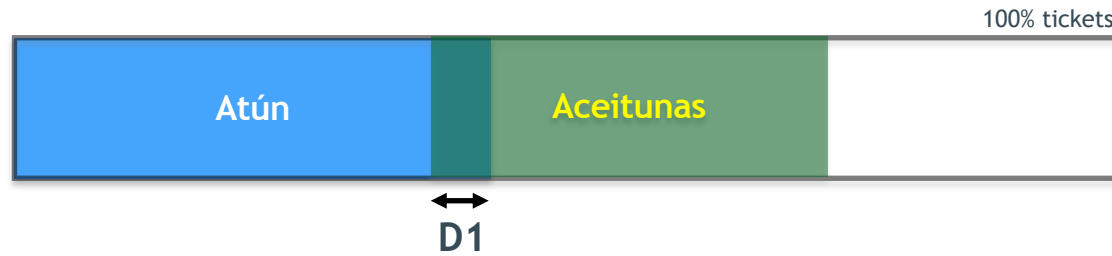


Entender los indicadores



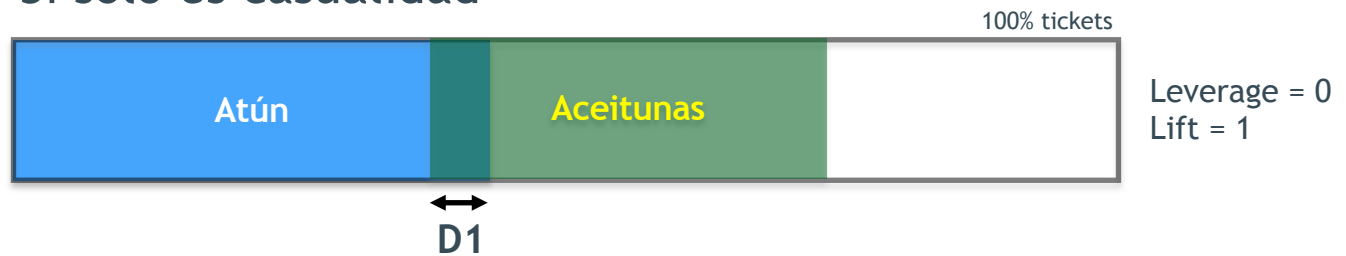
Antecedente	Consecuente	Coverage (penetración)	Support	Confidence	Leverage	Lift
Atún	Aceitunas	10,4%	4,23%	40,42%	1,56%	1,58

Entender los indicadores



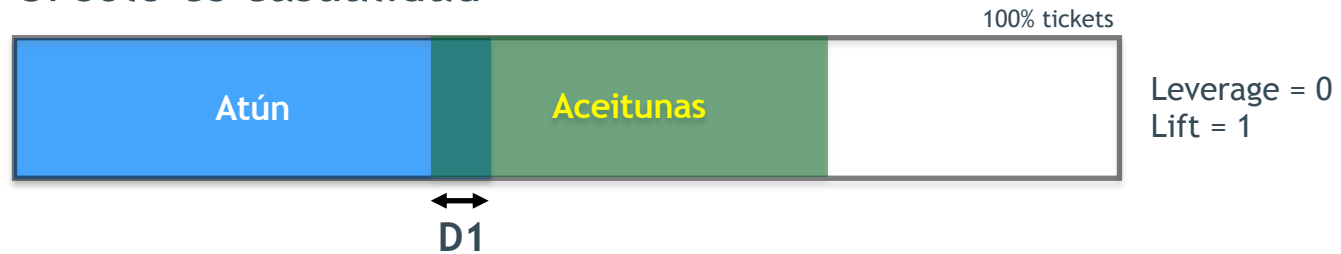
Entender los indicadores

Si solo es casualidad

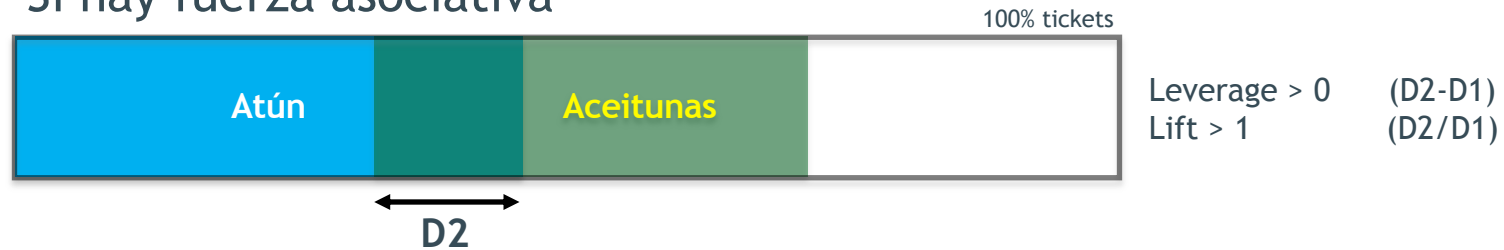


Entender los indicadores

Si solo es casualidad



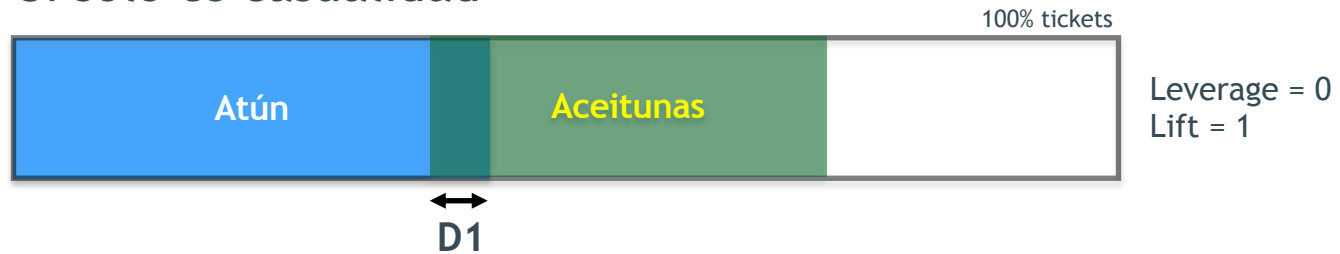
Si hay fuerza asociativa



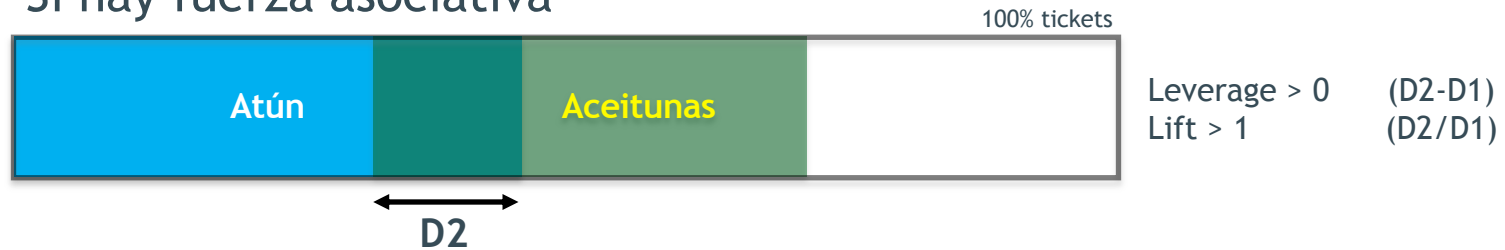
Antecedente	Consecuente	Coverage (penetración)	Support	Confidence	Leverage	Lift
Atún	Aceitunas	10,4%	4,23%	40,42%	1,56%	1,58

Entender los indicadores

Si solo es casualidad

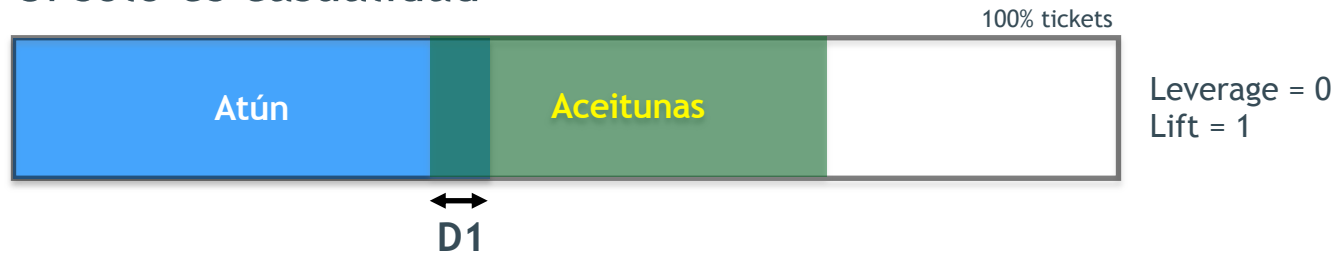


Si hay fuerza asociativa

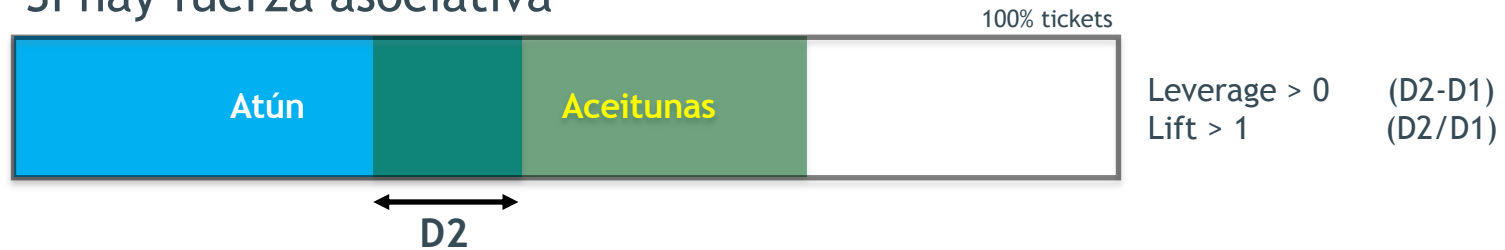


Entender los indicadores

Si solo es casualidad



Si hay fuerza asociativa



La estrategia LIFT “premia” las asociaciones que suceden menos, pero tienen “fuerza”.

Índice de hoy

- _01 Conexión con negocio.
- _02 Automatizaciones con Python.
- _03 **Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.**
- _04 Cierre del curso. Por qué una plataforma. 10 lecciones de ML.

Índice de hoy

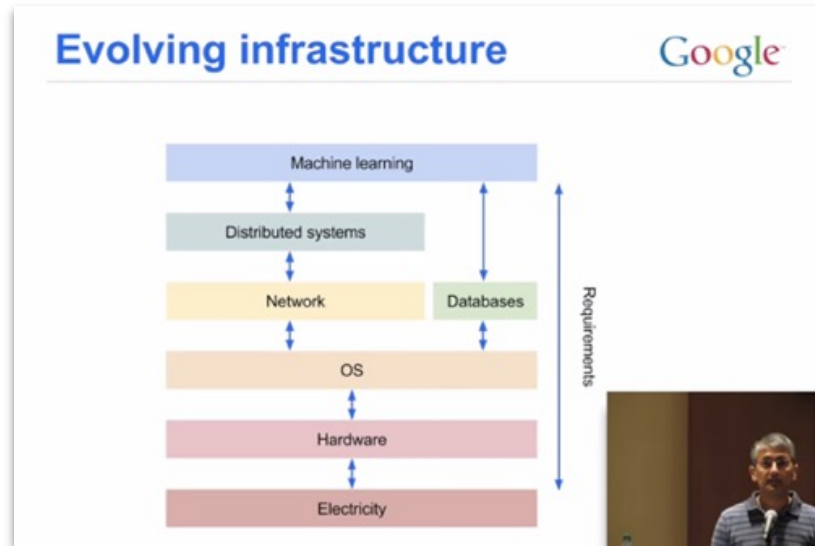
__01 Conexión con negocio.

__02 Automatizaciones con Python.

__03 Ejercicio con modelo no supervisado. Análisis de la cesta de la compra.

__04 **Cierre del curso. Por qué una plataforma. 10 lecciones de ML.**

Por qué una plataforma



- Machine Learning se está convirtiendo en una nueva capa en la infraestructura tecnológica de las empresas.
- Un desarrollador de aplicaciones de empresa tendrá acceso a una plataforma de Machine Learning.

Por qué una plataforma

La visión de una plataforma según Facebook (FBLearner)

- Cada algoritmo de ML se implementa una vez y será **reutilizable**.
- Los ingenieros deben ser capaces de escribir un código **paralelizable** que se ejecute en diferentes máquinas.
- El entrenamiento de un modelo debe ser **fácil**, independientemente de los conocimientos de ML.
- Cada paso debe ser **automatizable**.
- Cualquier ingeniero puede **buscar** experimentos pasados, ver los resultados, compartirlos e iniciar nuevas variantes basadas en un experimento anterior.
- AG: la **documentación** del modelo debe generarse automáticamente.
- AG: puedes **centrarte en los datos**, en el **negocio** y no en los modelos.
- AG: la **curva de aprendizaje** es muy rápida.

<https://engineering.fb.com/core-data/introducing-fblearner-flow-facebook-s-ai-backbone/>

10 lecciones de ML

- __01 Primero el objetivo de negocio, después los datos. Profundiza tu conocimiento del sector.
- __02 Haz un mínimo plan de negocio. ROI.
- __03 Ten en cuenta el coste de la matriz de confusión desde el punto de vista de negocio.
- __04 Piensa en cómo integras el modelo en producción y qué acciones de negocio vas a tomar.
- __05 Empieza por algo acotado que conozcas bien. No te lées con grandes proyectos. Trabajar con cientos de miles de datos es muy complejo.

10 lecciones de ML

- __06 Lo más importante son los datos, son los ingredientes de la receta. Es mejor tener pocos datos y de calidad que muchísimos y de poca calidad.
- __07 Monitoriza los modelos en producción. Se degradan con el tiempo.
- __08 No todo es Machine Learning. Se pueden “pisar callos”.
- __09 Los modelos tienen sesgos que se deben solucionar.
- __10 Hay que saber trabajar con la incertidumbre. ML es una “ciencia” experimental.

Vamos.