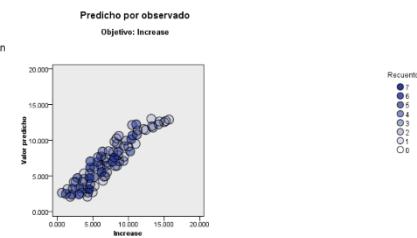
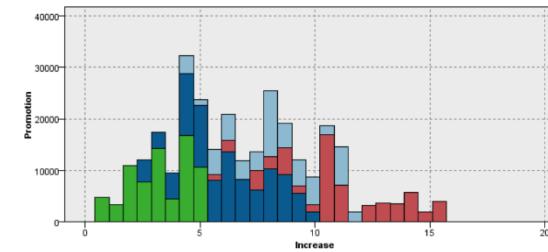
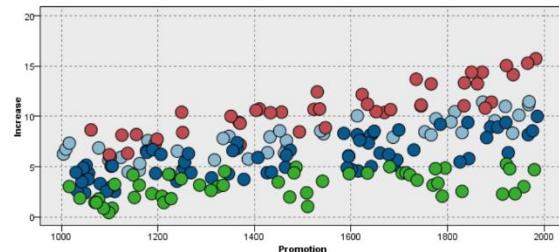


UEA Datos a Gran Escala

Caso de Estudio en el Mercado de Bienes de Consumo



Dr. Pedro Pablo González Pérez

e-mail: pgonzalez@correo.cua.uam.mx

<http://dcni.cua.uam.mx/division/usuario?p=31#>

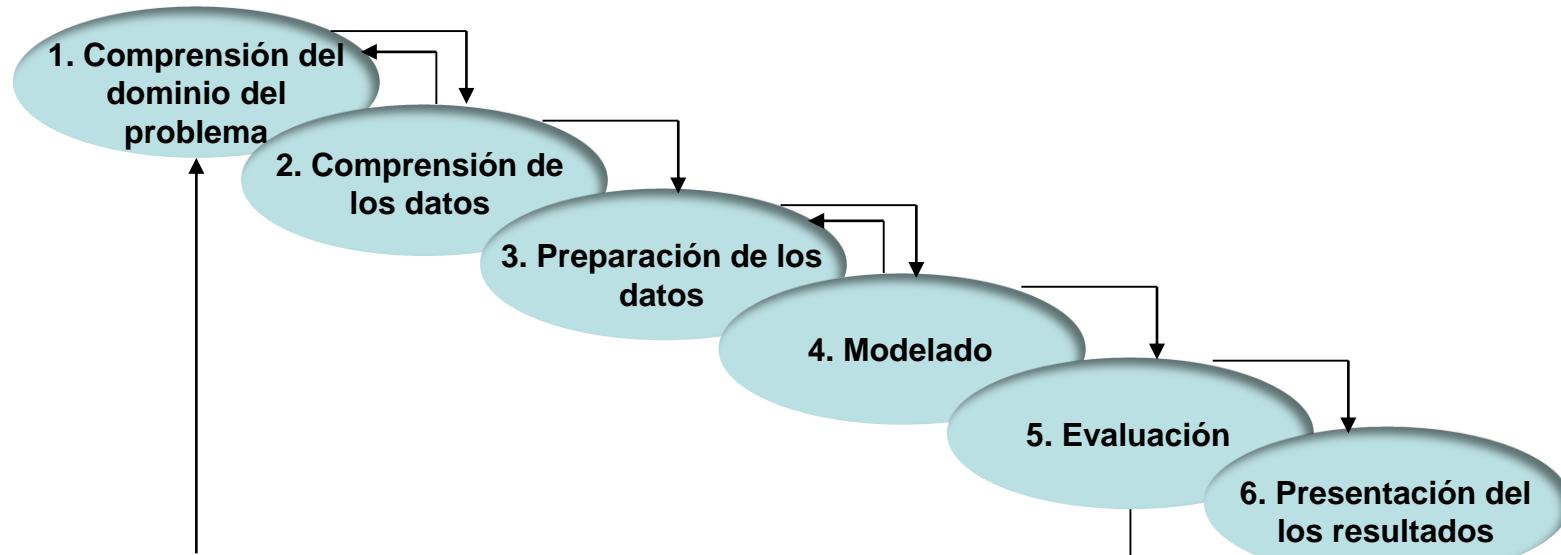
Departamento de Matemáticas Aplicadas y Sistemas

UEA Datos a Gran Escala

Enfoque metodológico basado en la minería de datos

Preprocesamiento y Análisis de los Datos: Enfoque metodológico basado en la minería de datos

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

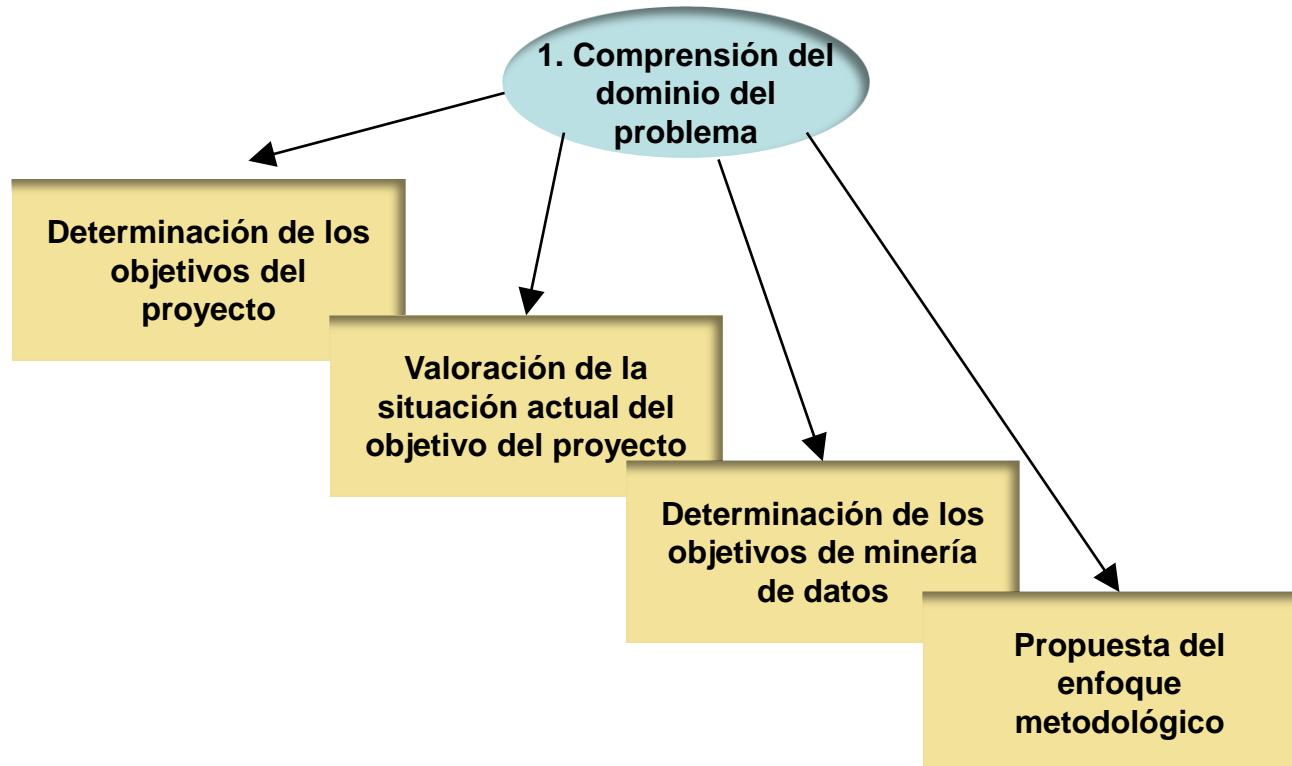


UEA Datos a Gran Escala

Enfoque metodológico basado en la minería de datos

CRISP-DM (Cross Industry Standard Process for Data Mining)
[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.



UEA Datos a Gran Escala

Enfoque metodológico basado en la minería de datos

CRISP-DM (Cross Industry Standard Process for Data Mining)
[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

- Determinación de los objetivos del proyecto.

Caso de estudio en el mercado de bienes de consumo

- Comprender los factores que influyen positivamente en el incremento de las ventas en cuatro categorías de bienes de consumo: “luxury”, “drink”, “meat” y “confection”.
- Comparar el comportamiento del incremento de las ventas entre las cuatro categorías de bienes de consumo, en base a la promoción aplicada.
- Predecir el incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

- Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Se comprende de forma clara el problema que se intenta abordar?

- Un bien de consumo es el producto final de un proceso de producción, el cual va dirigido a satisfacer de forma directa las necesidades de las personas. El mercado de bienes de consumo se refiere a aquel tipo de mercado donde los productos comercializados están destinados a satisfacer las necesidades del cliente final.
- Los bienes de consumo pueden ser de consumo inmediato (no perdurables), por ejemplo, alimentos y bebidas; o de consumo duradero, por ejemplo, vivienda, muebles, equipos electrodomésticos, electrónica, etc.
- A diferencia de otros bienes, por ejemplo, los bienes de capital, los bienes de consumo no crean otros bienes o productos, sino que éstos benefician directamente al cliente final.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Se comprende de forma clara el problema que se intenta abordar?
 - Aunque existen diferentes categorías para clasificar los bienes de consumo, una de las principales es aquella que los clasifica según su tiempo de uso o duración:
 - ✓ Bienes de consumo duraderos.
 - ✓ Bienes de consumo no duraderos.
 - De forma paralela, existen varias formas de clasificar los bienes de consumo según el tipo de productos que representa, por ejemplo:
 - ✓ Electrodomésticos.
 - ✓ Electrónica.
 - ✓ Bebidas y licores.
 - ✓ Alimentos confeccionados.
 - ✓ Carnes.
 - ✓ Perfumería.
 - ✓ ...

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Se comprende de forma clara el problema que se intenta abordar?
 - Dos características claves de los bienes de consumo son su clasificación y precio. Por ejemplo:
 - ✓ Bienes de consumo duradero:
 - Electrodomésticos:
 - Refrigerador marca AX200, precio: \$18,600.00.
 - Lavadora marca BZ1000, precio: \$14,500.00.
 - Muebles:
 - Sala de cuatro piezas modelo FT500, precio: \$24,000.00.
 - Recámara modelo ITA300, precio: \$18,400.00.
 - Electrónica:
 - TV 60 pulgadas 4K marca IO2000, precio: \$18,900.00.
 - Laptop marca Spin600, precio: \$14,800.00.
 - Automóviles:
 - Auto sedán marca AT150, modelo MS200, precio: \$360,000.00.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

❖ ¿Se comprende de forma clara el problema que se intenta abordar?

- ✓ Bienes de consumo no duradero:
 - Carnes:
 - Milanesa de res, precio: \$180.00/kg.
 - Pechuga de pollo, precio: \$120.00/kg.
 - Alimentos procesados:
 - Galletas surtidas, precio: \$80.00 la caja de 400 g.
 - Atún en lata, marca AT30, precio: \$24.00 la lata de 380 g.
 - Pastel de chocolate, precio: \$120.00/kg.
 - Bebidas y licores:
 - Vino tinto, marca VT10, precio: \$265.00.
 - Whisky, marca Wh80, precio: \$320.00.
 - Cerveza, marca CR40, precio: \$45.00

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

❖ ¿Se comprende de forma clara el problema que se intenta abordar?

- La gran mayoría de los bienes de consumo se encuentran sujetos a ofertas y promociones sobre su precio de mercado, en determinados períodos del año, con la finalidad de incentivar el consumo.
- Para lograr que las ofertas y promociones logren realmente incentivar el consumo, y por lo tanto reflejen un incremento real en las ventas, es necesario considerar varios factores, entre ellos:
 - 1) período del año en que se lanza la oferta o promoción,
 - 2) tipo de bien de consumo (duradero o no duradero),
 - 3) categoría a la que pertenece el bien de consumo (electrodoméstico, electrónica, automóviles, alimentos elaborados, bebidas y licores, perfumería, etc.),
 - 4) precio comercial del bien de consumo,
 - 5) porcentaje de descuento respecto al precio comercial, y
 - 6) características de la población de las áreas o regiones consideradas para la oferta o promoción

CRISP-DM (Cross Industry Standard Process for Data Mining)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Existen datos disponibles para efectuar el análisis?
 - Sí. Se cuenta con suficientes datos que describen los volúmenes de ventas registrados antes y después de aplicar promociones, para cuatro clases de productos de consumo:
 - Bebidas y licores.
 - Carnes.
 - Alimentos elaborados.
 - Artículos de lujo.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

1) Comprensión del dominio del problema o negocio.

1. Comprensión del dominio del problema

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Existen datos disponibles para efectuar el análisis? De contar con datos disponibles, ¿cuál es la fuente de estos datos y de qué tipo son?
 - La fuente de estos datos son los registros de ventas antes y después de la aplicación de promociones en cuatro clases de productos de consumo.

1) Comprensión del dominio del problema o negocio.

➤ Valoración de la situación actual del objetivo del proyecto.

Caso de estudio en el mercado de bienes de consumo

- ❖ ¿Existen datos disponibles para efectuar el análisis? De contar con datos disponibles, ¿cuál es la fuente de estos datos y de qué tipo son?

1	Class	Cost	Promotion	Before	After
2	Confection	23.99	1467	114957	122762
3	Drink	79.29	1745	123378	137097
4	Luxury	81.99	1426	135246	141172
5	Confection	74.18	1098	231389	244456
6	Confection	90.09	1968	235648	261940
7	Meat	69.85	1486	148885	156232
8	Meat	100.15	1248	123760	128441
9	Luxury	21.01	1364	251072	268134
10	Luxury	87.32	1585	287043	310857
11	Drink	26.58	1835	240805	272863
12	Drink	65.23	1194	212406	227836
13	Meat	79.82	1596	174022	181489
14	Confection	41.39	1161	270631	283189
15	Meat	36.82	1151	231281	235722
16	Meat	44.05	1482	178138	185934
17	Drink	84.62	1623	247885	278031
18	Confection	51.82	1969	148597	165598
19	Confection	90.08	1462	215102	228696
20	Luxury	57.3	1842	246885	270082
21	Drink	11.02	1370	164984	176802
22	Confection	95.86	1815	257882	284835

1) Comprensión del dominio del problema o negocio.

➤ Determinación de los objetivos de minería de datos.

Caso de estudio en el mercado de bienes de consumo

- ❖ **Problema de predicción:** Construir varios modelos predictivos supervisados, que permitan predecir el incremento en las ventas para cada clase de bienes de consumo. Seleccionar el modelo predictivo con mayor precisión (menor tasa de error).

1) Comprensión del dominio del problema o negocio.

- Propuesta del enfoque metodológico (plan de proyecto de minería de datos) en forma de tabla.

1. Comprensión del dominio del problema

Caso de estudio en el mercado de bienes de consumo

Fase	Tiempo a dedicar	Recursos humanos y tecnológicos	Riesgos atribuibles
1. Comprensión del dominio del problema	1 semana	Experto en mercadotecnia, experto en minería de datos.	No se han identificado riesgos.
2. Comprensión de los datos	2 semanas	Experto en el dominio del problema, experto en minería de datos. Tablas, gráficos y resúmenes estadísticos que faciliten la comprensión de los datos.	No se trata de un gran volumen de datos, lo cual podría influir en la precisión de los resultados, y por lo tanto en el poder de la herramienta predictiva.
3. Preparación de los datos	2 semanas	Experto en mercadotecnia, experto en minería de datos. Herramientas para el análisis exploratorio de datos. Paquete <i>IBM SPSS Modeler</i> .	No se han identificado riesgos.

1) Comprensión del dominio del problema o negocio.

- Propuesta del enfoque metodológico (plan de proyecto de minería de datos) en forma de tabla.

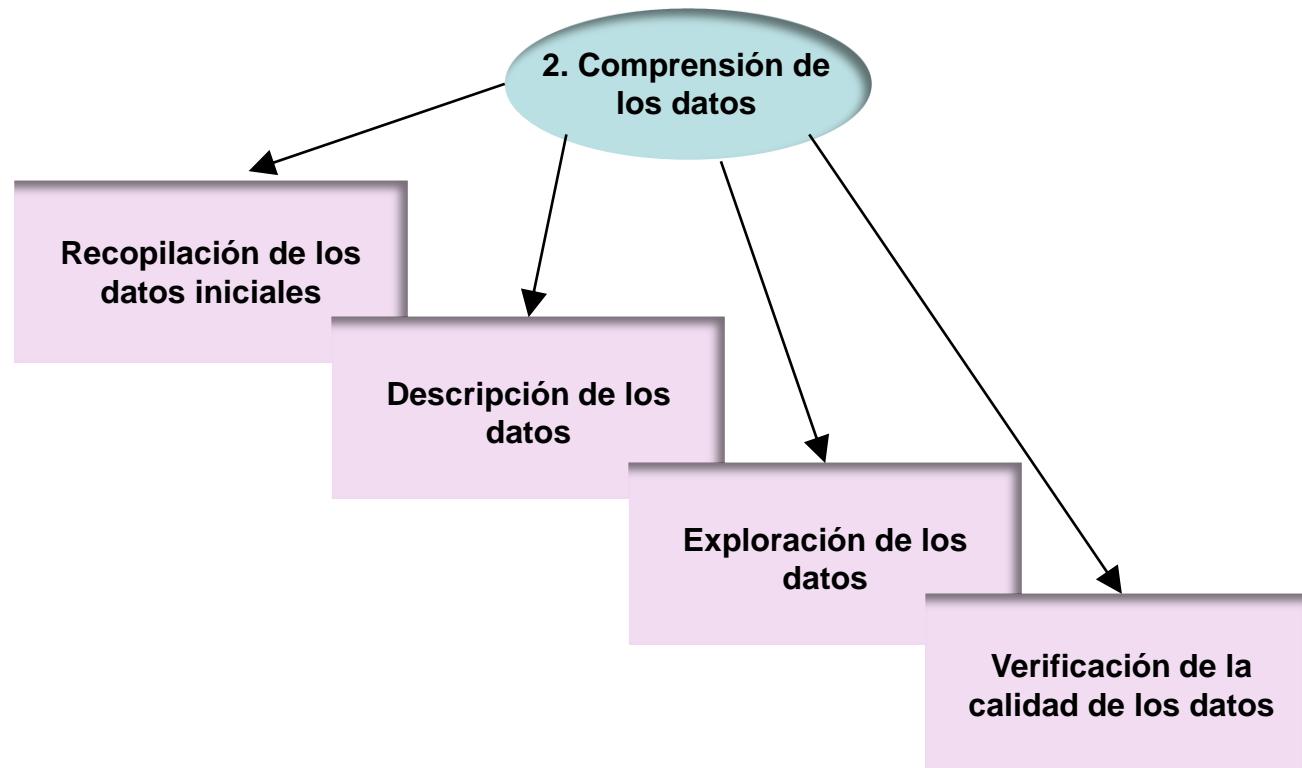
1. Comprensión del dominio del problema

Caso de estudio en el mercado de bienes de consumo

Fase	Tiempo a dedicar	Recursos humanos y tecnológicos	Riesgos atribuibles
4. Modelado	2 semanas	Experto en minería de datos y experto en técnicas de <i>machine learning</i> . Herramientas para la implementación de modelos de <i>machine learning</i> . Paquete <i>IBM SPSS Modeler</i> .	No se han identificado riesgos.
5. Evaluación	1 semana	Experto en mercadotecnia, experto en minería de datos. Paquete <i>IBM SPSS Modeler</i> .	No se han identificado riesgos.
6. Presentación	1 semana	Experto en mercadotecnia, experto en minería de datos, directiva de la organización. Paquete <i>IBM SPSS Modeler</i> .	No se han identificado riesgos.

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.



CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.

2. Comprensión de los datos

➤ Recopilación de los datos iniciales.

Caso de estudio en el mercado de bienes de consumo

	Class	Cost	Promotion	Before	After
1	Confection	23.990	1467.000	114957....	122762....
2	Drink	79.290	1745.000	123378....	137097....
3	Luxury	81.990	1426.000	135246....	141172....
4	Confection	74.180	1098.000	231389....	244456....
5	Confection	90.090	1968.000	235648....	261940....
6	Meat	69.850	1486.000	148885....	156232....
7	Meat	100.1...	1248.000	123760....	128441....
8	Luxury	21.010	1364.000	251072....	268134....
9	Luxury	87.320	1585.000	287043....	310857....
10	Drink	26.580	1835.000	240805....	272863....
11	Drink	65.230	1194.000	212406....	227836....
12	Meat	79.820	1596.000	174022....	181489....
13	Confection	41.390	1161.000	270631....	283189....
14	Meat	36.820	1151.000	231281....	235722....
15	Meat	44.050	1482.000	178138....	185934....
16	Drink	84.620	1623.000	247885....	278031....
17	Confection	51.820	1969.000	148597....	165598....
18	Confection	90.080	1462.000	215102....	228696....
19	Luxury	57.300	1842.000	246885....	270082....
20	Drink	11.020	1370.000	164984....	176802....

Tabla compuesta por 6 atributos y 200 registros

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprendión de los datos.

➤ Descripción de los datos.

Caso de estudio en el mercado de bienes de consumo

Tabla compuesta por 6 atributos y 200 registros. Los atributos y tipos son los siguientes:

- Class**: categórico.
- Cost**: continuo.
- Promotion**: continuo.
- Before**: continuo.
- After**: continuo.

2. Comprensión de los datos

	Class	Cost	Promotion	Before	After
1	Confection	23.990	1467.000	114957....	122762....
2	Drink	79.290	1745.000	123378....	137097....
3	Luxury	81.990	1426.000	135246....	141172....
4	Confection	74.180	1098.000	231389....	244456....
5	Confection	90.090	1968.000	235648....	261940....
6	Meat	69.850	1486.000	148885....	156232....
7	Meat	100.1...	1248.000	123760....	128441....
8	Luxury	21.010	1364.000	251072....	268134....
9	Luxury	87.320	1585.000	287043....	310857....
10	Drink	26.580	1835.000	240805....	272863....
11	Drink	65.230	1194.000	212406....	227836....
12	Meat	79.820	1596.000	174022....	181489....
13	Confection	41.390	1161.000	270631....	283189....
14	Meat	36.820	1151.000	231281....	235722....
15	Meat	44.050	1482.000	178138....	185934....
16	Drink	84.620	1623.000	247885....	278031....
17	Confection	51.820	1969.000	148597....	165598....
18	Confection	90.080	1462.000	215102....	228696....
19	Luxury	57.300	1842.000	246885....	270082....
20	Drink	11.020	1370.000	164984....	176802....

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.

2. Comprensión de los datos

➤ Descripción de los datos.

Caso de estudio en el mercado de bienes de consumo

Descripción de los atributos

- Class.** Clase o categoría a la que pertenece el bien de consumo. En el *dataset* se identifican 4 clases de bienes de consumo: *Drink*, *Meat*, *Confection* y *Luxury*.
- Cost.** Costo unitario del bien de consumo.
- Promotion.** Monto de la promoción aplicada.
- Before.** Monto de las ventas antes de la promoción.
- After.** Monto de las ventas después de la promoción.

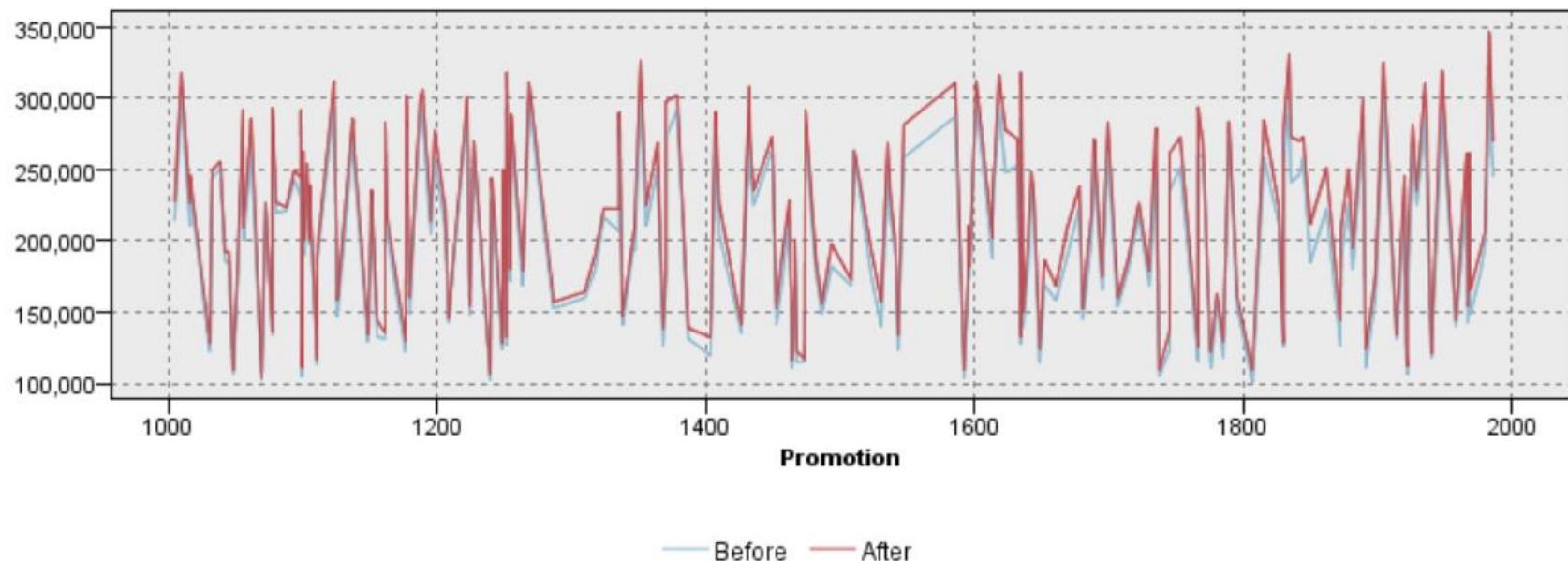
CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprendión de los datos.

2. Comprensión de los datos

➤ Exploración de los datos.

Caso de estudio en el mercado de bienes de consumo



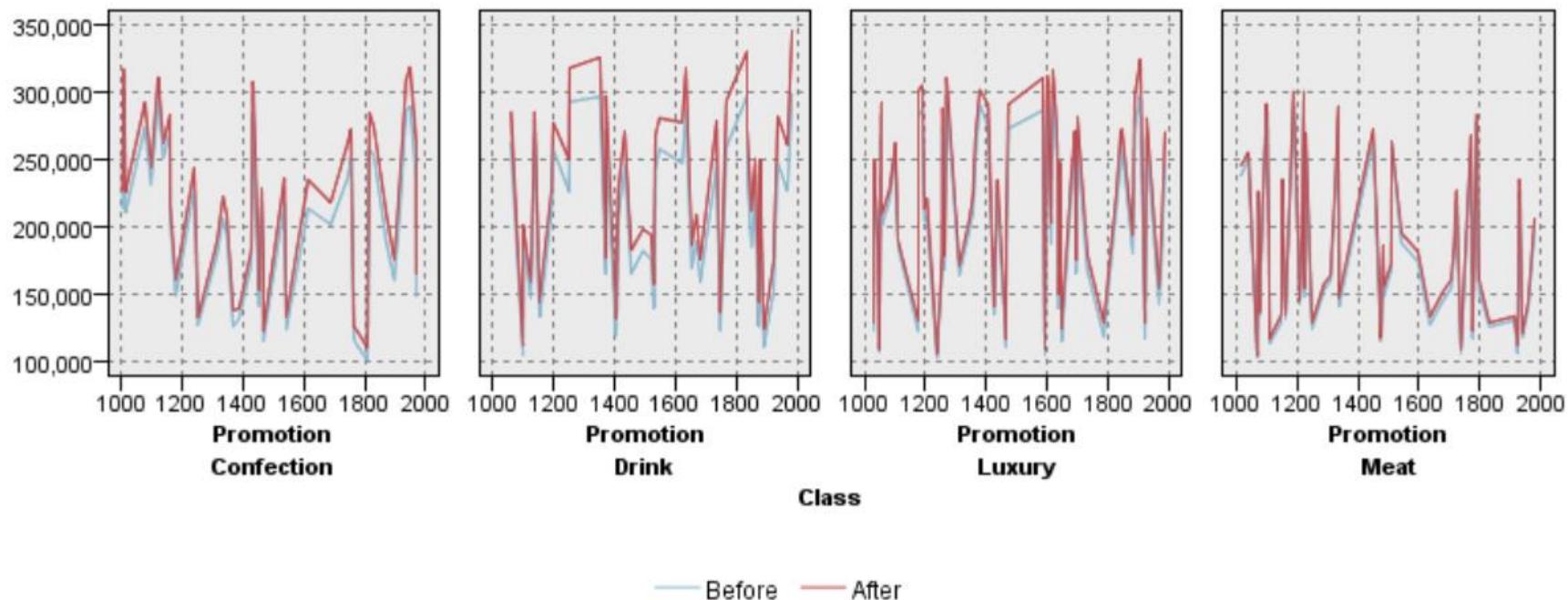
CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprendión de los datos.

➤ Exploración de los datos.

Caso de estudio en el mercado de bienes de consumo

2. Comprensión de los datos



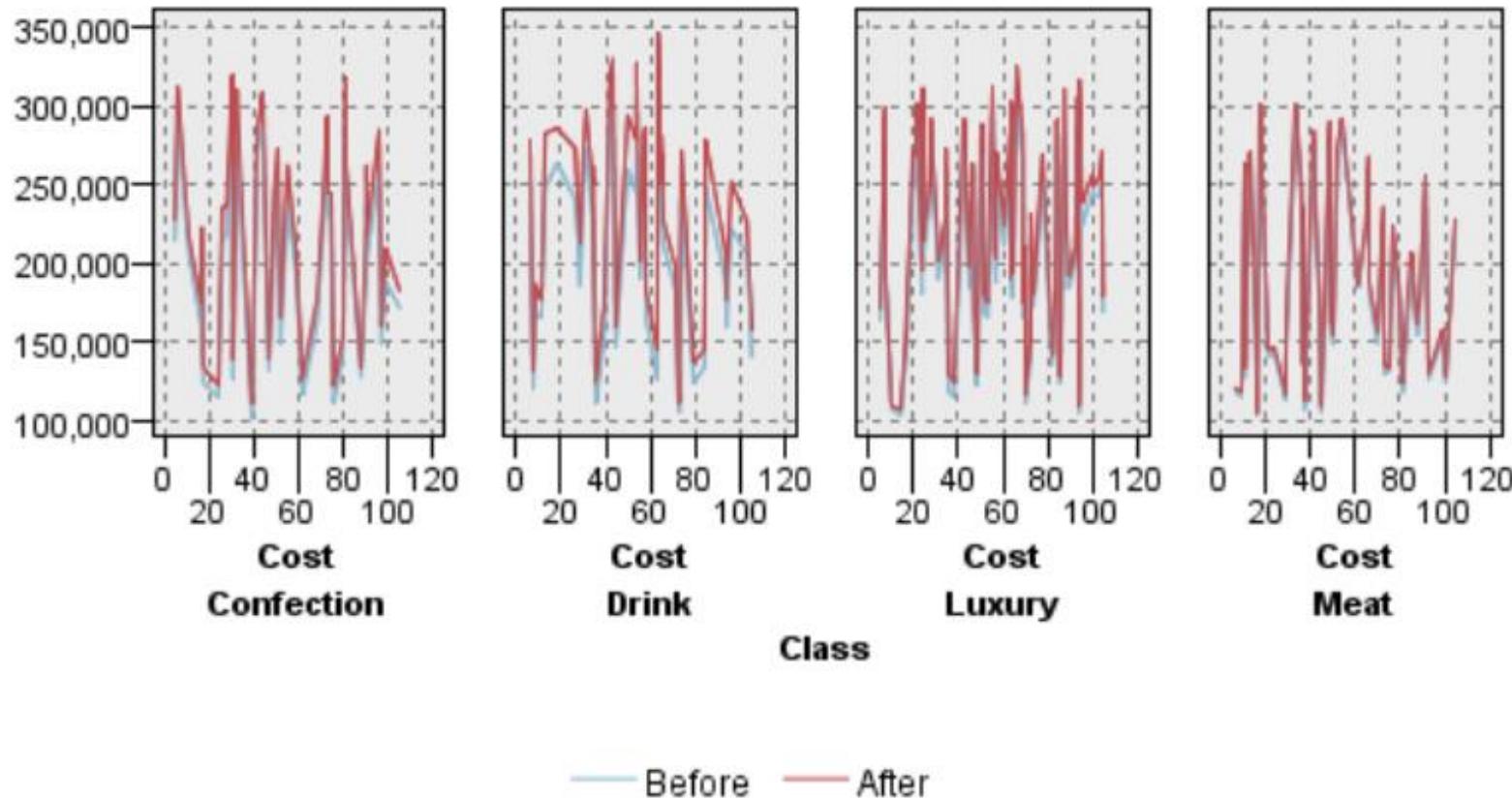
CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprendión de los datos.

2. Comprensión de los datos

➤ Exploración de los datos.

Caso de estudio en el mercado de bienes de consumo



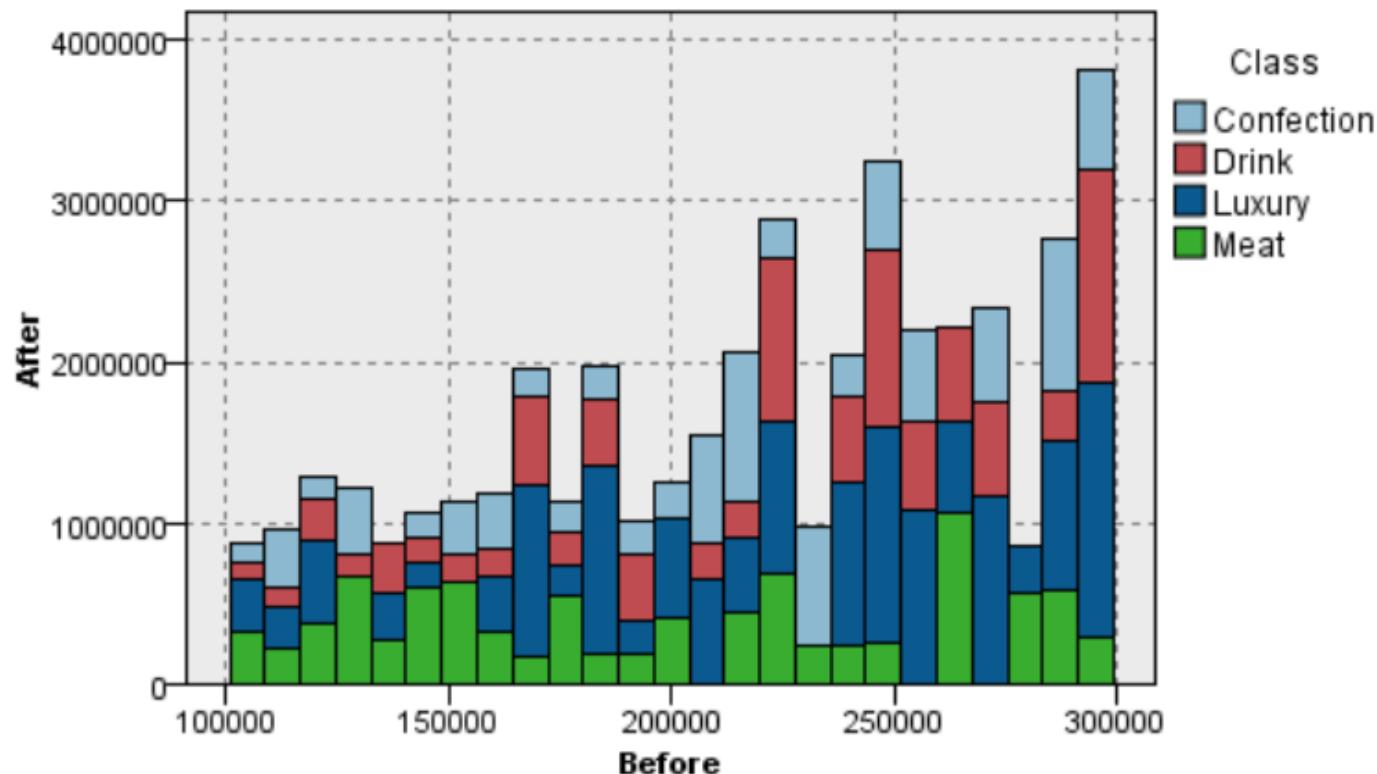
CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.

2. Comprensión de los datos

➤ Exploración de los datos.

Caso de estudio en el mercado de bienes de consumo



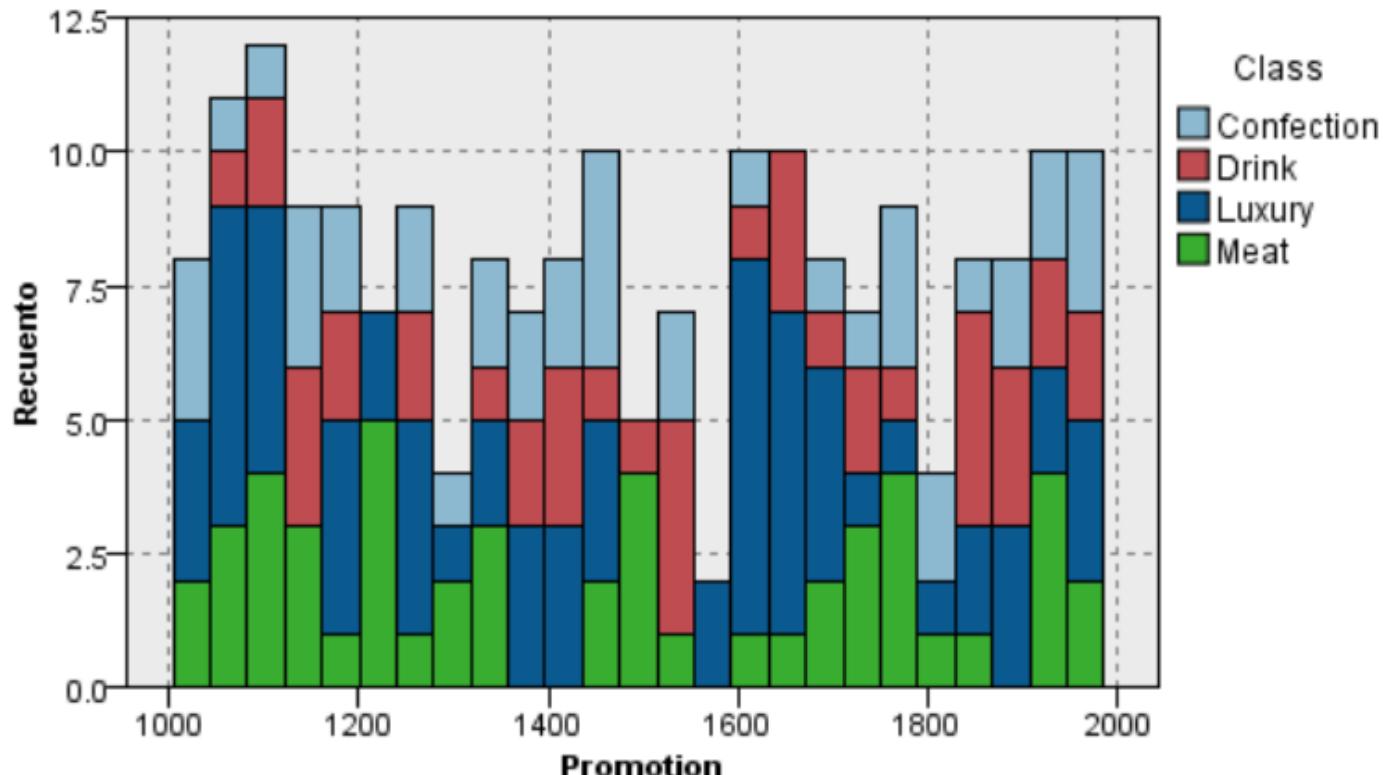
CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.

2. Comprensión de los datos

➤ Exploración de los datos.

Caso de estudio en el mercado de bienes de consumo



CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

2) Comprensión de los datos.

2. Comprensión de los datos

➤ Verificación de la calidad de los datos.

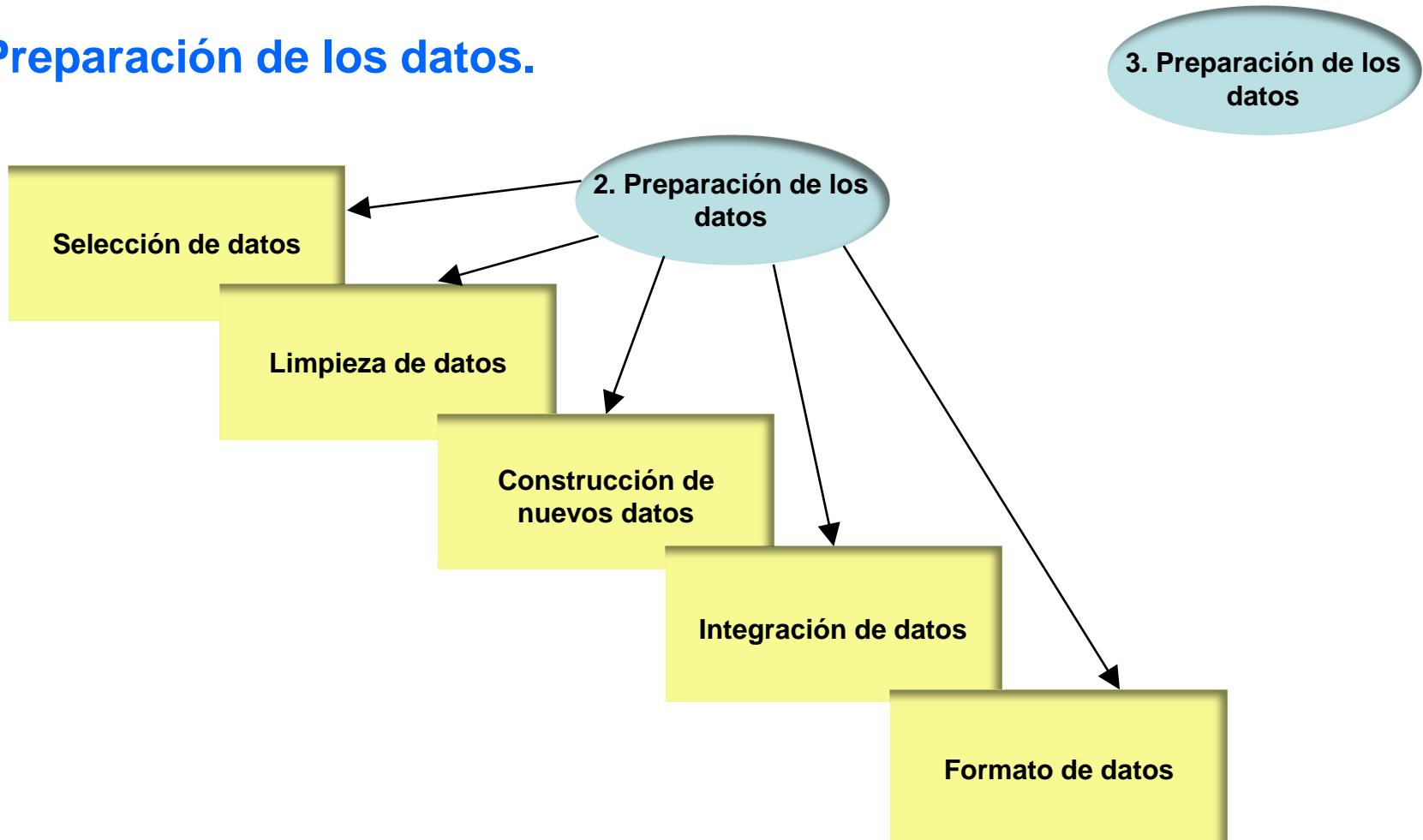
Caso de estudio en el mercado de bienes de consumo

- ❖ **Datos perdidos o vacíos:** No existen.
- ❖ **Errores en los datos:** Aparentemente no existen.
- ❖ **Errores de medición:** Aparentemente no existen.
- ❖ **Errores de codificación:** Aparentemente no existen.
- ❖ **Atributos redundantes o de escasa utilidad:** No se han identificado.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.



CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

3. Preparación de los datos

➤ Selección de datos.

Caso de estudio en el mercado de bienes de consumo

- Inicialmente, se mantendrán los 6 atributos y los 200 registros que componen el *dataset*. Como veremos más adelante, se considerará la derivación de un nuevo atributo, lo que podría ocasionar la exclusión de uno de los atributos existentes, al dejar de ser relevante.
- Como ya se indicó en la fase “**Comprensión del dominio del problema o negocio**”, el número de registros podría resultar insuficiente para garantizar una gran precisión en los objetivos de minería de datos.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

➤ Limpieza de los datos.

Caso de estudio en el mercado de bienes de consumo

3. Preparación de los datos

Como ya se determinó en la fase “**Comprensión de los datos**”, no será necesario corregir problemas relacionados con datos perdidos o vacíos, errores en los datos o errores de codificación.

- ❖ **Datos perdidos o vacíos:** No existen.
- ❖ **Errores en los datos:** Aparentemente no existen.
- ❖ **Errores de medición:** Aparentemente no existen.
- ❖ **Errores de codificación:** Aparentemente no existen.
- ❖ **Atributos redundantes o de escasa utilidad:** No se han identificado.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

3. Preparación de los datos

➤ Construcción de nuevos datos.

Caso de estudio en el mercado de bienes de consumo

- Como ya analizamos, el *dataset* define los atributos:
 - ❖ “Class”
 - ❖ “Cost”
 - ❖ “Promotion”
 - ❖ “Before”
 - ❖ “After”
- Sin embargo, se requiere derivar el atributo “*Increase*”, a partir de los atributos iniciales “*Before*” y “*After*”. De esta forma, el nuevo atributo “*Increase*” será el atributo “OBJETIVO”, es decir, el atributo cuyo valor se desea predecir.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

3. Preparación de los datos

➤ Construcción de nuevos datos.

Caso de estudio en el mercado de bienes de consumo

Derivar campo:

Increase

Derivar como: Fórmula

Tipo de campo:  <Predeterminado>

Fórmula:

1 **(After - Before) / Before * 100.0**

2

3

4

⋮

9



CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

➤ Integración de datos.

Caso de estudio en el mercado de bienes de consumo

3. Preparación de los datos

- No se cuenta con otras fuentes de datos a integrar al conjunto inicial de datos. Por lo tanto, el conjunto inicial de datos ya preparado será el conjunto de datos a analizar con los modelos que se propongan en la fase “Modelado”.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

➤ Formato de datos.

Caso de estudio en el mercado de bienes de consumo

3. Preparación de los datos

- Los datos se encuentran en el formato requerido para la aplicación de los modelos predictivos.
- Con relación al “**tipo de dato**”, sólo será necesario verificar que los atributos “Cost”, “Before”, “Promotion” y “After” sean del tipo “**Continuo**”, mientras que el atributo “Class” debería ser de tipo “**Nominal**”.
- Con relación al “**rol**”, todos los atributos iniciales deberán ser del tipo “**Entrada**”, mientras que el nuevo atributo derivado “*Increase*” deberá ser del tipo “**Destino**”.
- Dado que ahora se cuenta con el atributo derivado “*Increase*”, el atributo inicial “After” deja de ser relevante, por lo que no será considerado como atributo de “**Entrada**” a los modelos predictivos.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3) Preparación de los datos.

➤ Formato de datos.

Caso de estudio en el mercado de bienes de consumo

3. Preparación de los datos

Campo	Medida	Valores	No se enc...	Comprobar	Rol
A Class	Nominal	Confection,...		Ninguno	Entrada
Cost	Continuo	[5.08,104.98]		Ninguno	Entrada
Promotion	Continuo	[1004.0,19...]		Ninguno	Entrada
Before	Continuo	[100751.0,...]		Ninguno	Entrada
After	Continuo	[104393.0,...]		Ninguno	Ninguna
Increase	Continuo	[0.4365444...]		Ninguno	Destino

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

3. Preparación de los datos

3) Preparación de los datos – Nueva exploración de los datos.

Caso de estudio en el mercado de bienes de consumo

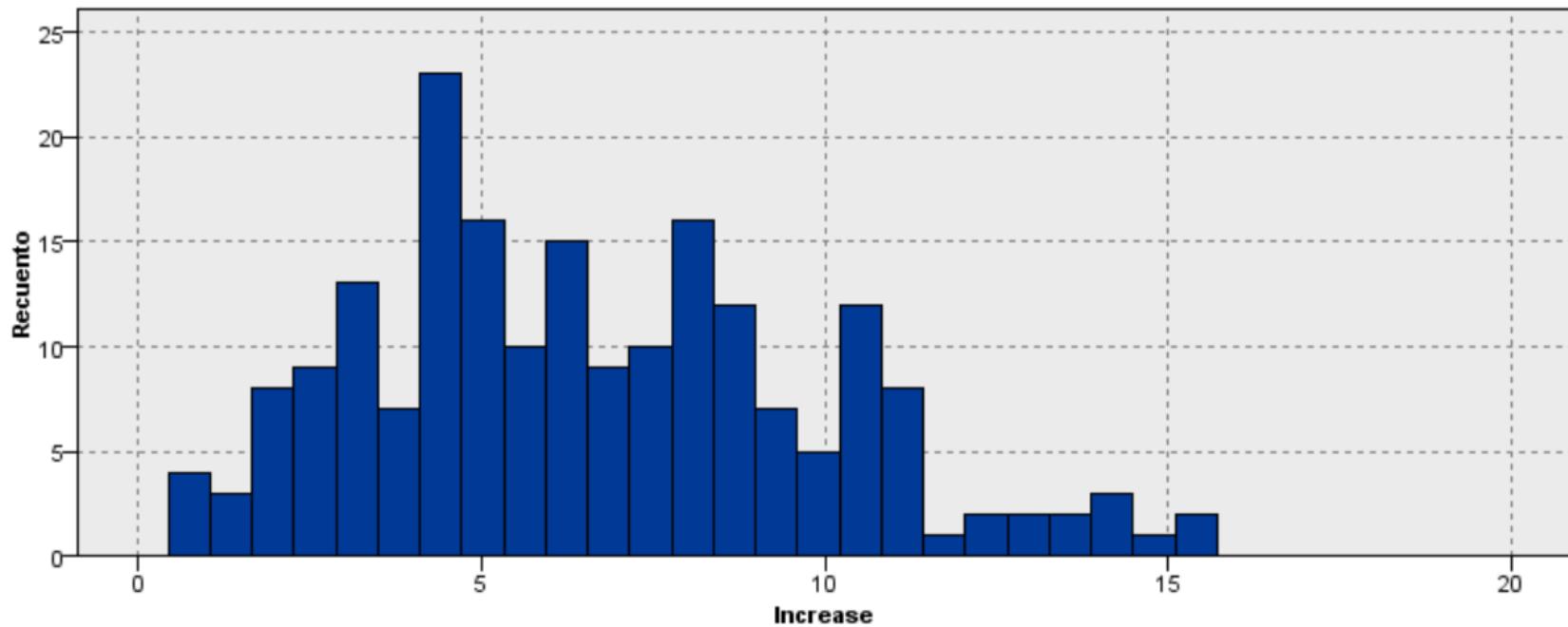
- ❖ Una vez preparados los datos, será necesario volver a explorarlos a partir de su nueva representación, tanto para comprenderlos mejor como para formular nuevas hipótesis acerca de éstos, de ser el caso.

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

3. Preparación de los datos

3) Preparación de los datos – Nueva exploración de los datos.

Caso de estudio en el mercado de bienes de consumo

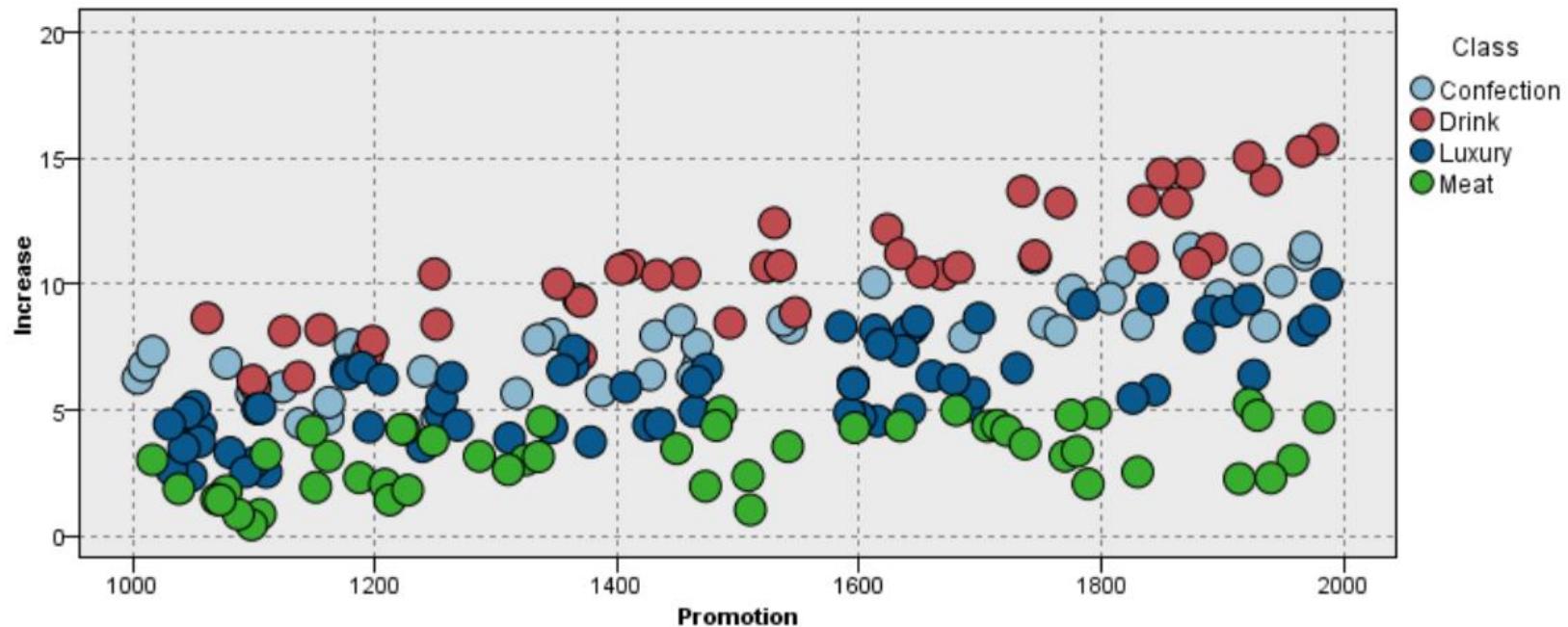


CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

3. Preparación de los datos

3) Preparación de los datos – Nueva exploración de los datos.

Caso de estudio en el mercado de bienes de consumo

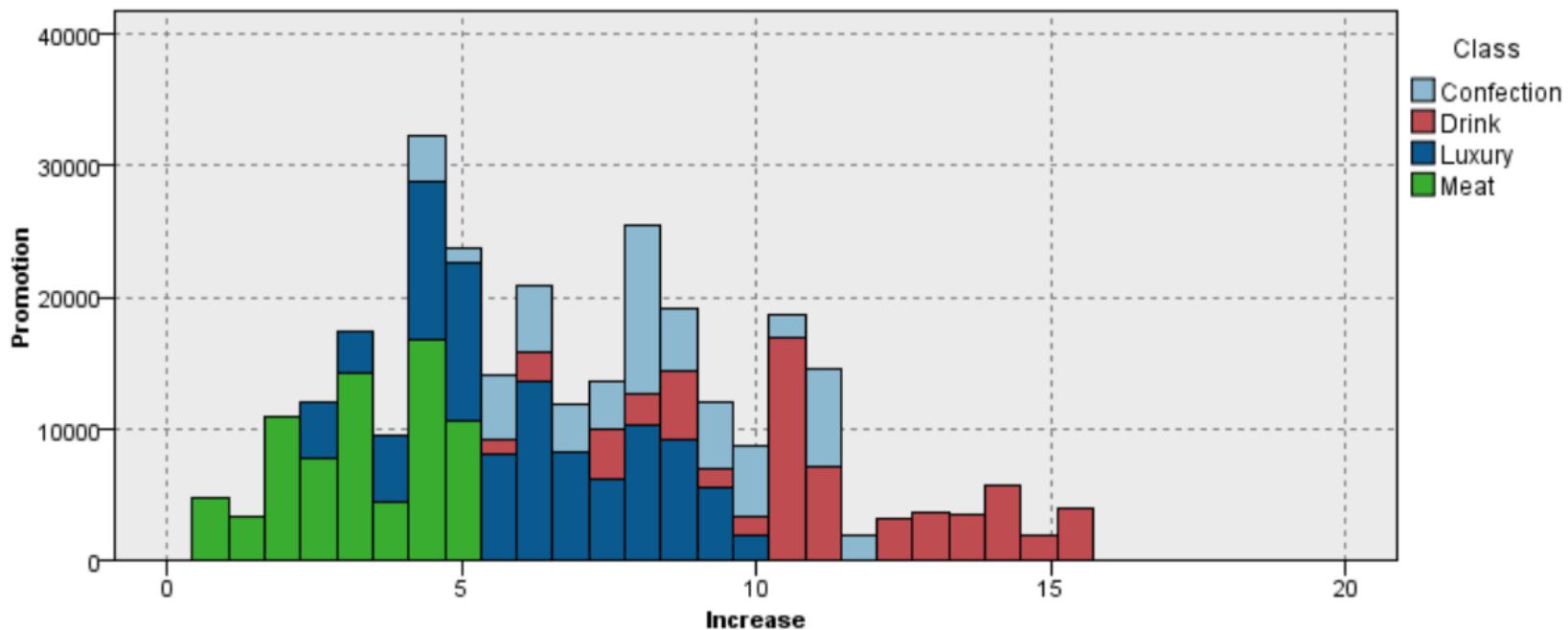


CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

3. Preparación de los datos

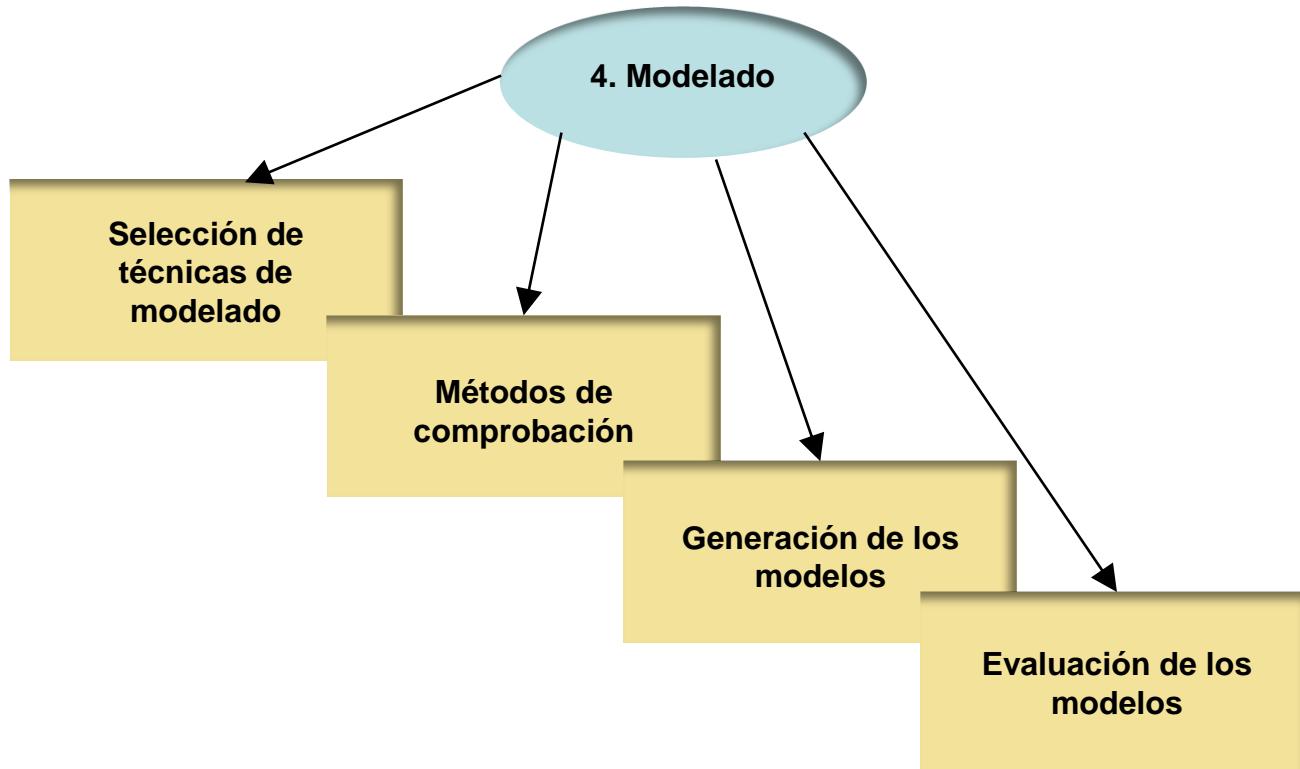
3) Preparación de los datos – Nueva exploración de los datos.

Caso de estudio en el mercado de bienes de consumo



CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

4) Modelado.



CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

4. Modelado

4) Modelado.

➤ Selección de técnicas de modelado.

Caso de estudio en el mercado de bienes de consumo

- Se escogieron 4 modelos que responden al objetivo de minería de datos que nos ocupa: predicción del incremento en las ventas, según la clase de bien de consumo. Los modelos seleccionados fueron los siguientes:
 - ❖ Árbol de decisión C&R.
 - ❖ Máquina de Soporte Vectorial (SVM).
 - ❖ Algoritmo KNN.
 - ❖ Red Neuronal Perceptron Backpropagation.

CRISP-DM (*Cross Industry Standard Process for Data Mining*) [Shearer, 2000; IBM Corporation, 2012]

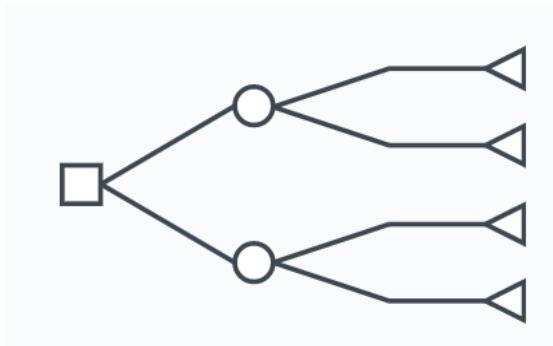
4. Modelado

4) Modelado.

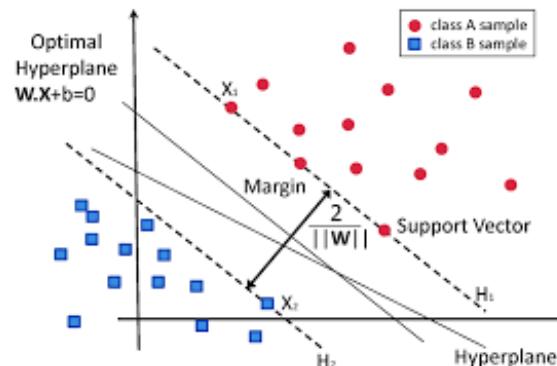
➤ Selección de técnicas de modelado.

Caso de estudio en el mercado de bienes de consumo

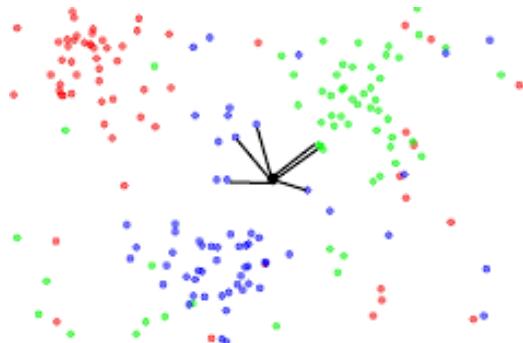
Árbol de decisión



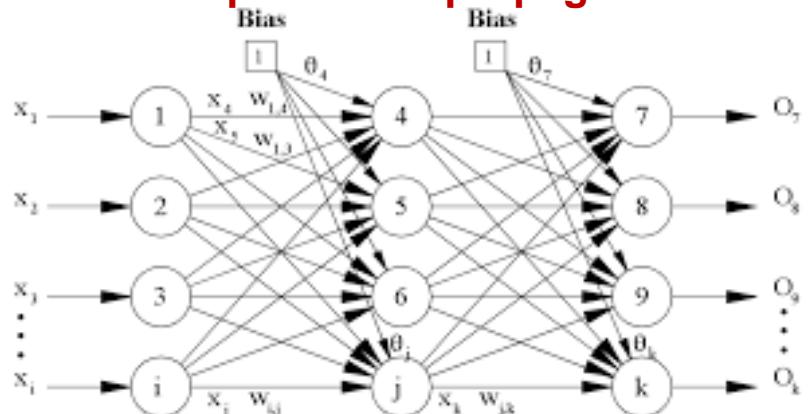
Máquina de soporte vectorial (SVM)



Algoritmo KNN



Perceptron backpropagation



CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

4. Modelado

4) Modelado.

➤ **Métodos de comprobación.**

Caso de estudio en el mercado de bienes de consumo

- Considerando que los modelos propuestos son técnicas de aprendizaje supervisado, entonces el **método de comprobación** seleccionado viene dado por el **criterio de bondad del modelo**, siendo en este caso la **tasa de error del modelo**.
- Con relación a los datos para comprobar el criterio de bondad del modelo, los 4 modelos propuestos incluyen entre sus prestaciones la partición de los datos en dos conjuntos, uno para el entrenamiento del modelo y el otro para la prueba o verificación del mismo.

CRISP-DM (*Cross Industry Standard Process for Data Mining*)

[Shearer, 2000; IBM Corporation, 2012]

4. Modelado

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

- Utilizando el paquete **IBM SPSS Modeler**, los siguientes cuatro modelos fueron generados:
 - ❖ Árbol de decisión C&R.
 - ❖ Máquina de Soporte Vectorial (SVM).
 - ❖ Algoritmo KNN.
 - ❖ Red Neuronal **Perceptron Backpropagation**.
- Inicialmente fueron utilizados los parámetros por defecto propuestos por cada modelo.
- Como ya se indicó, los cuatro modelos utilizados pertenecen a la categoría de aprendizaje supervisado, por lo que como criterio de bondad del modelo se consideró la tasa de error producida por el mismo.

CRISP-DM (Cross Industry Standard Process for Data Mining)

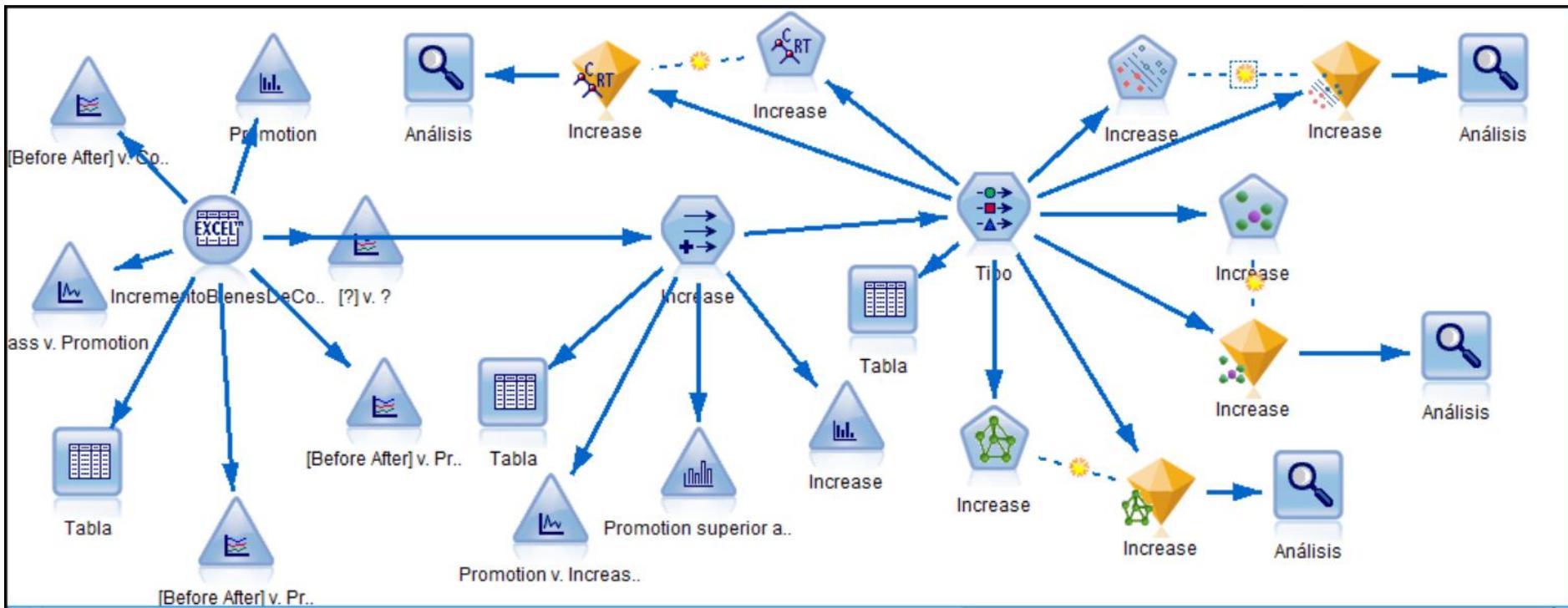
[Shearer, 2000; IBM Corporation, 2012]

4. Modelado

4) Modelado.

➤ Generación de los modelos (IBM SPSS Modeler).

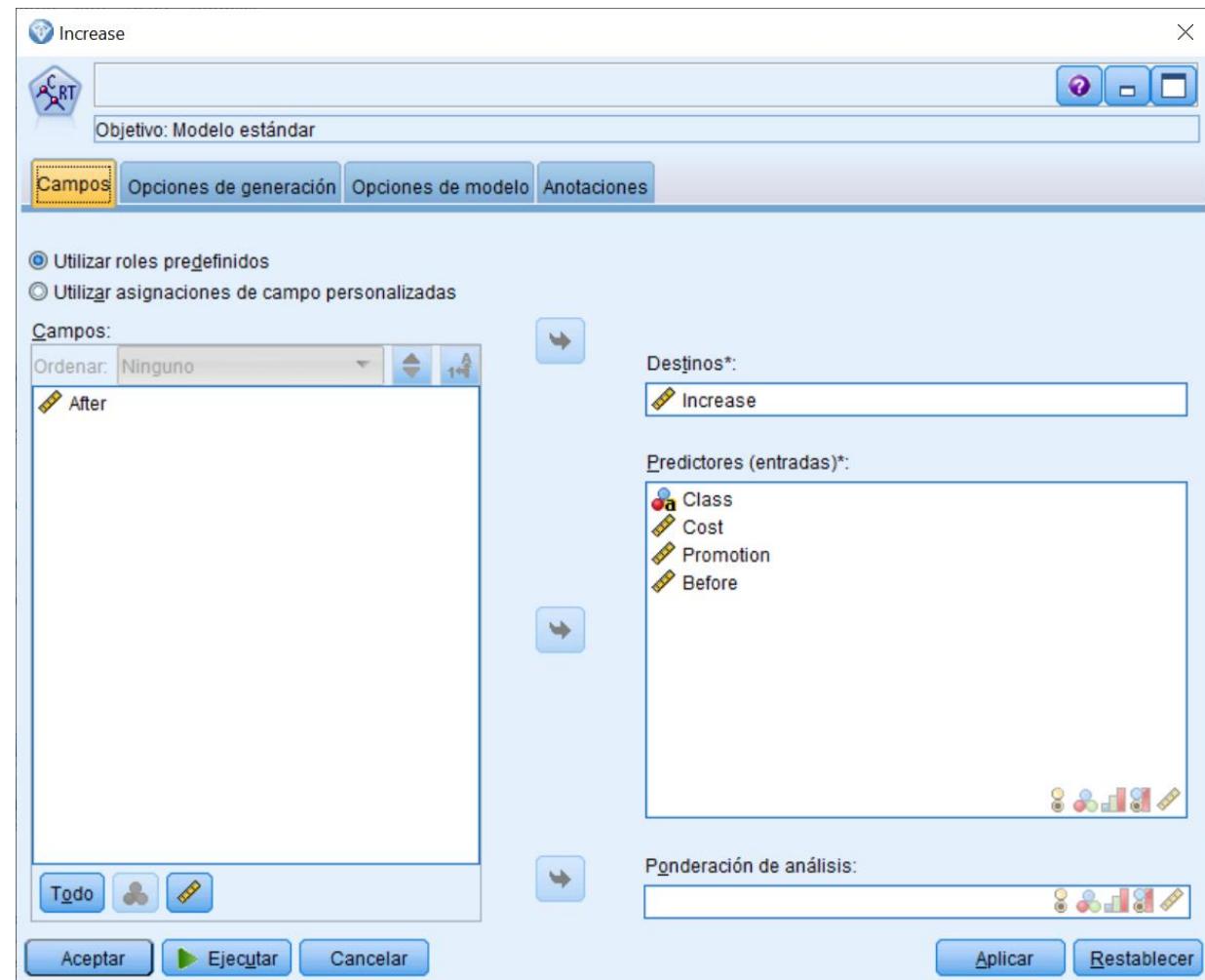
Caso de estudio en el mercado de bienes de consumo



4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

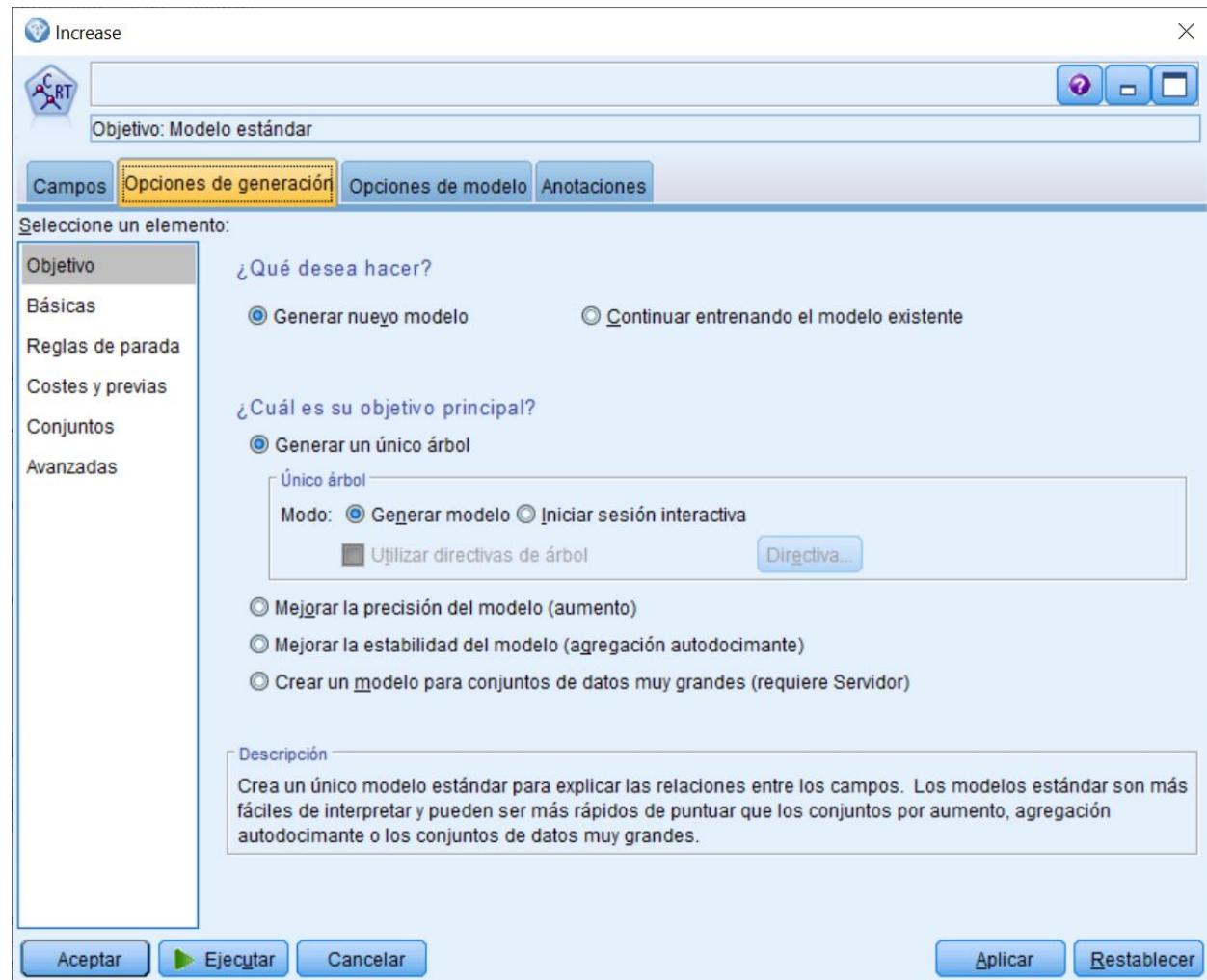


Generación del modelo
árbol de decisión
C&R

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

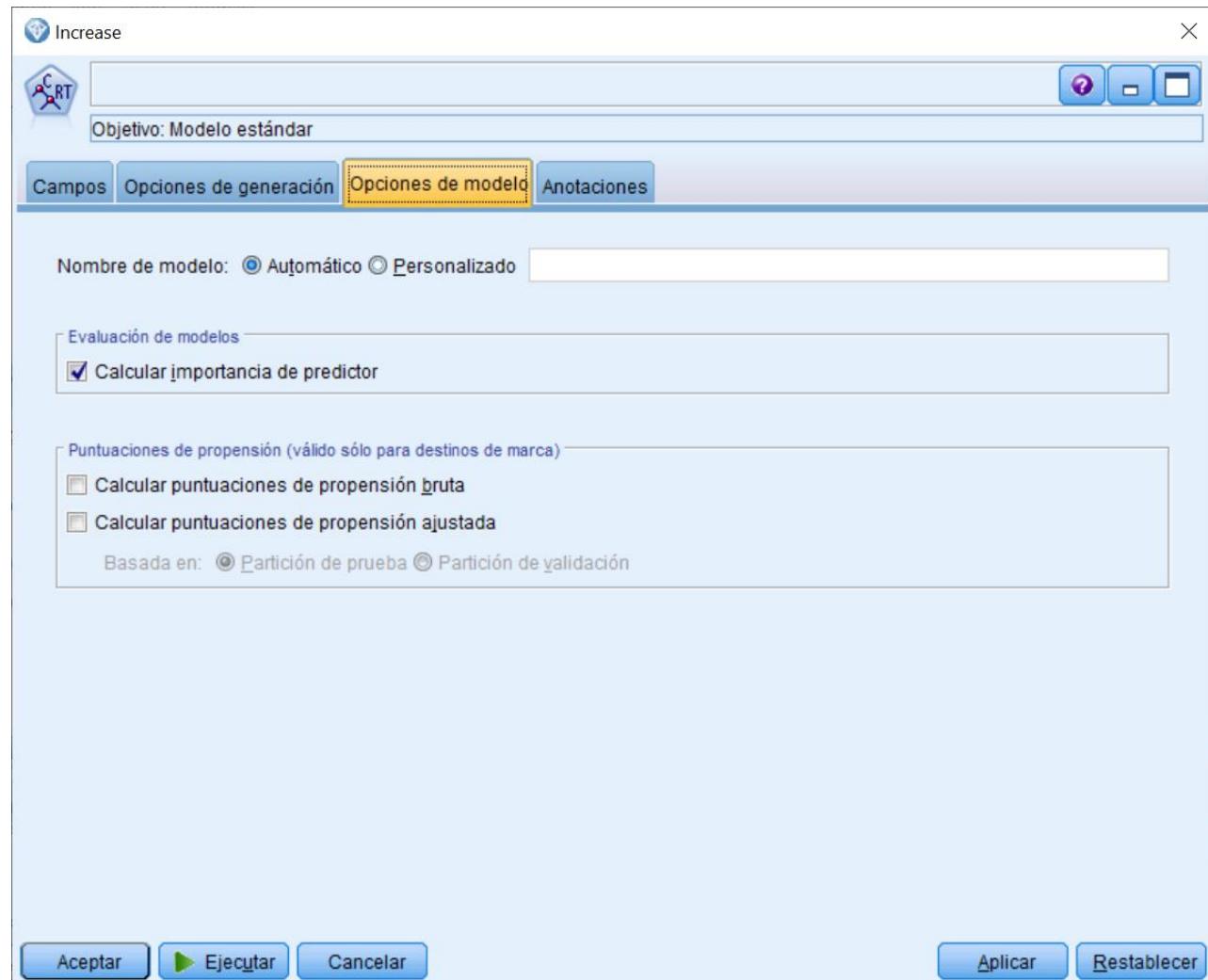


Generación del modelo
árbol de decisión
C&R

4) Modelado.

➤ Generación de los modelos.

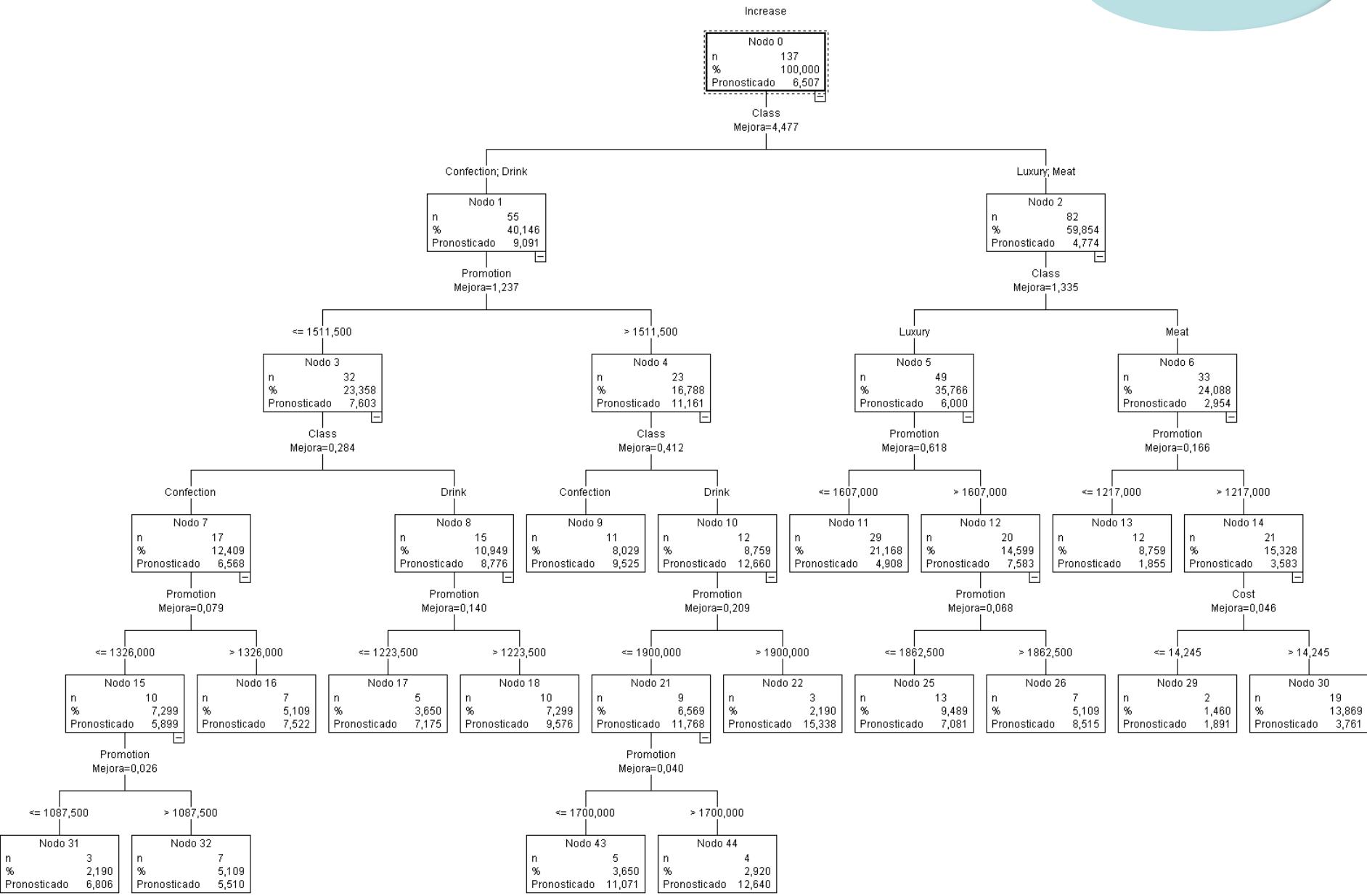
Caso de estudio en el mercado de bienes de consumo



Generación del modelo
árbol de decisión
C&R

Ejecución del modelo árbol de decisión C&R

4. Modelado

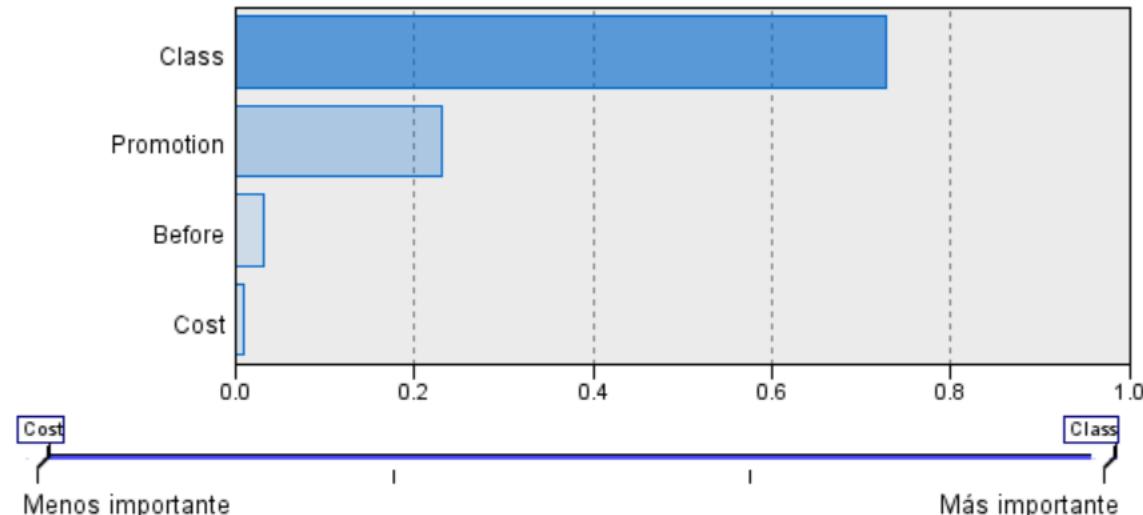


4) Modelado.

➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

Ejecución del modelo
árbol de decisión
C&R

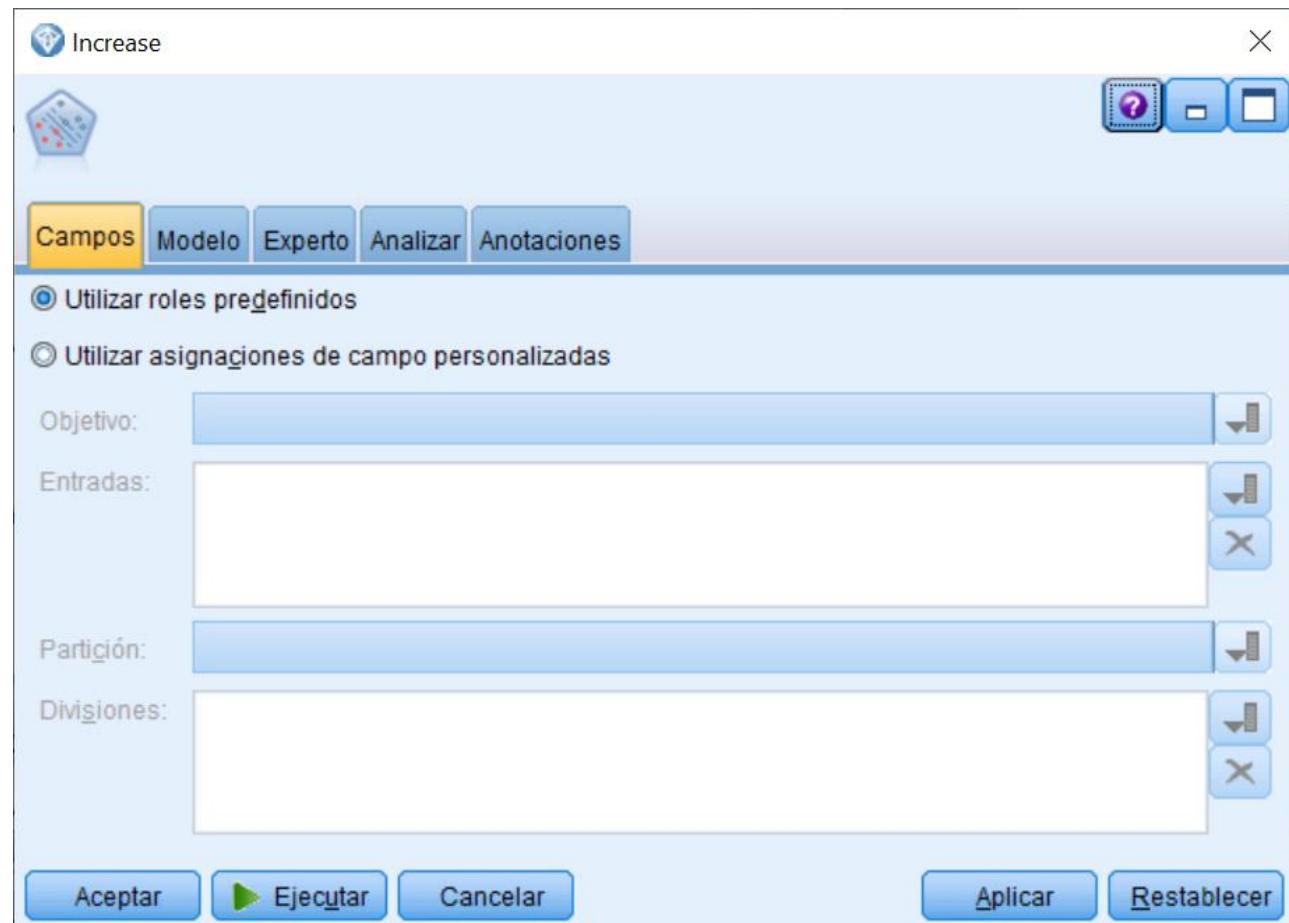


4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Generación del modelo
**máquina de soporte
vectorial (SVM)**

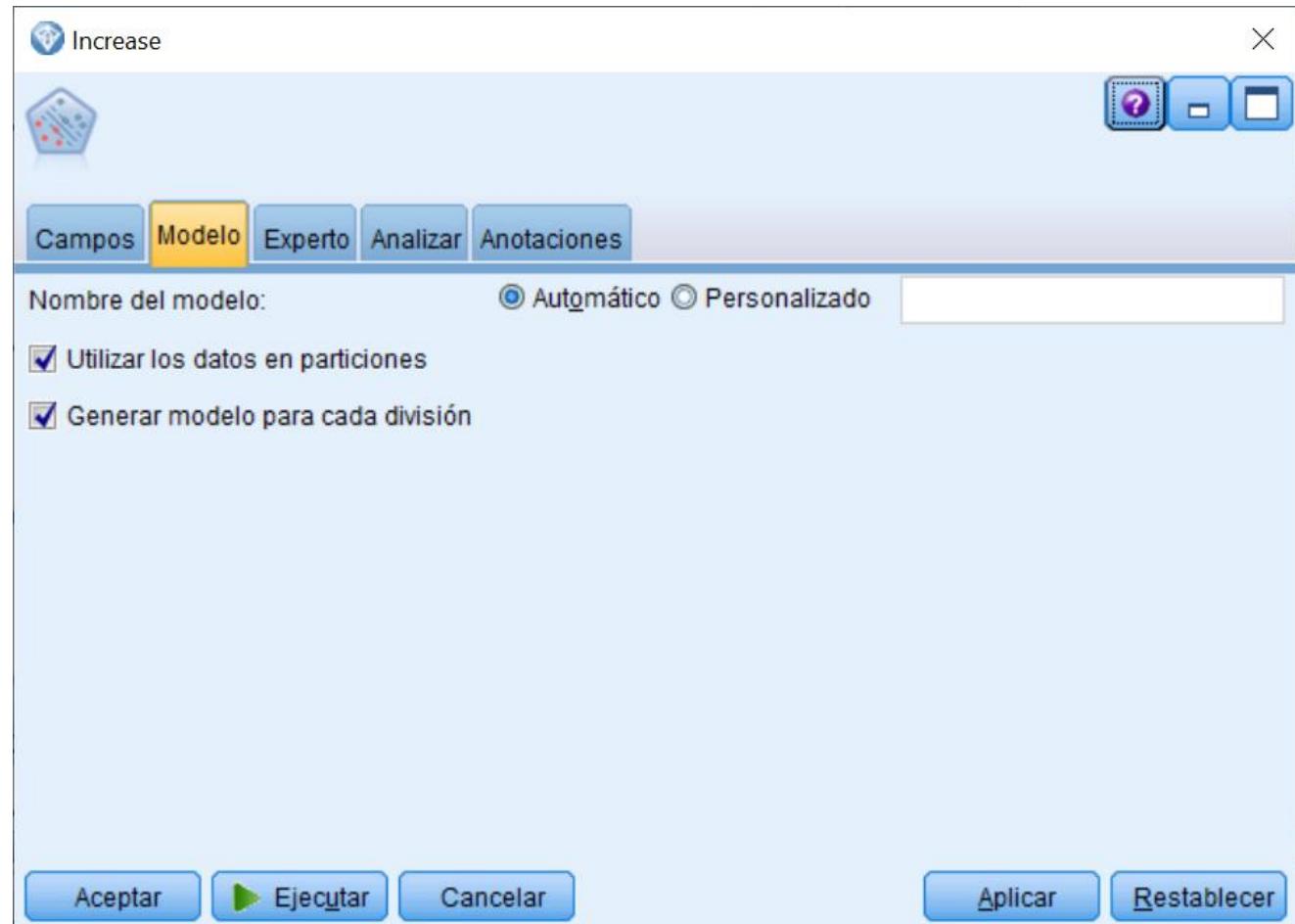


4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

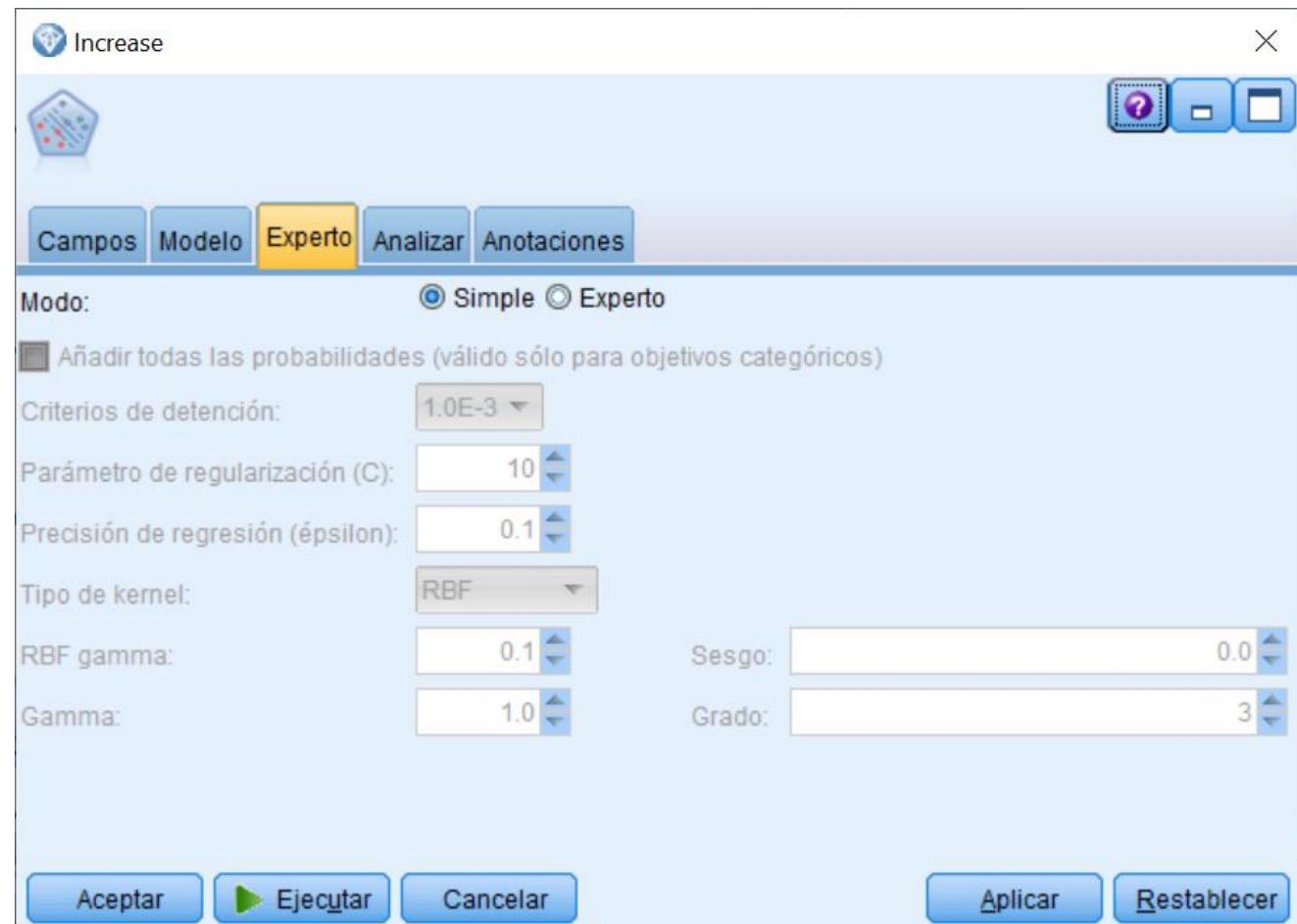
Generación del modelo
**máquina de soporte
vectorial (SVM)**



4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

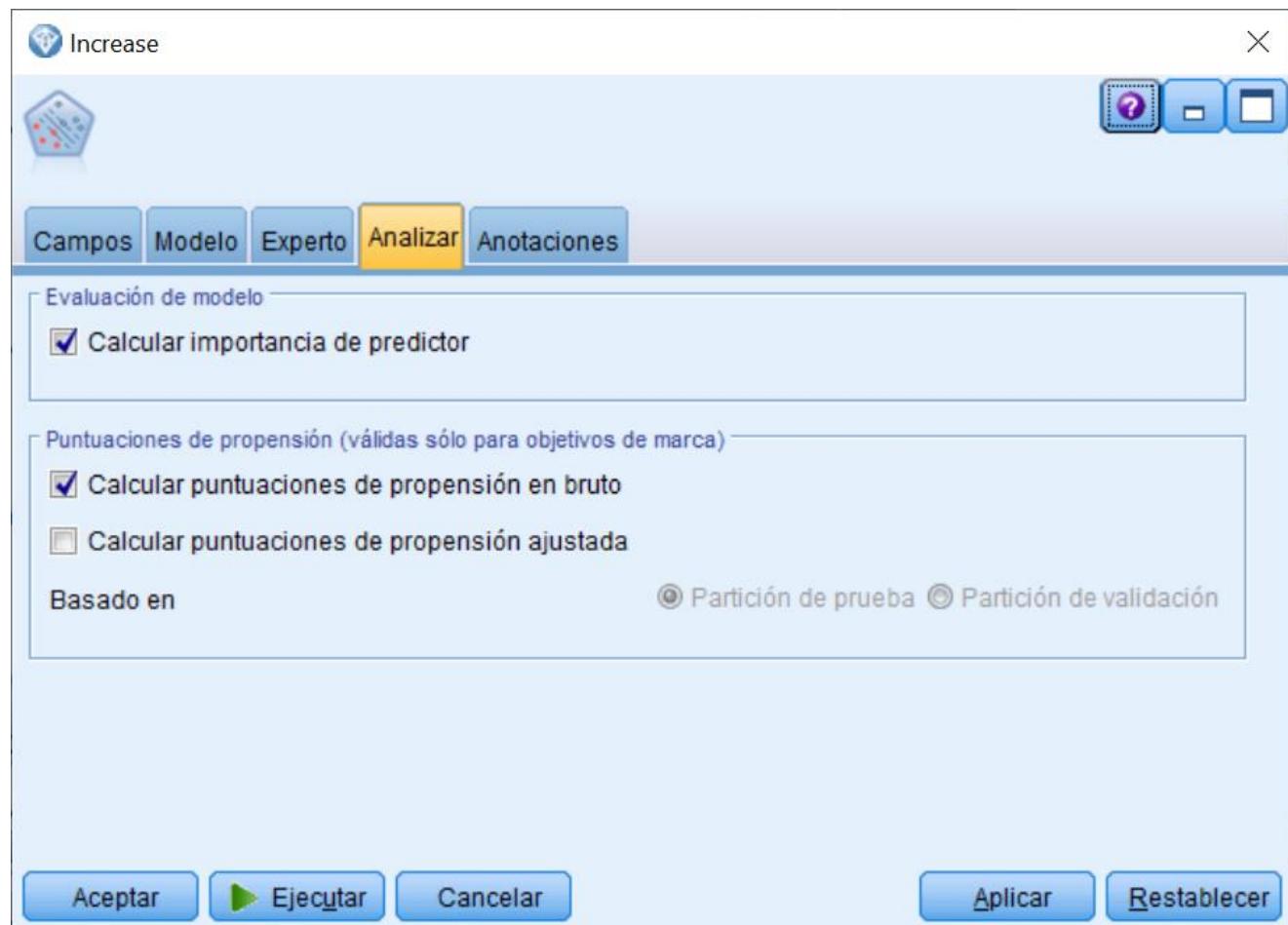


Generación del modelo
**máquina de soporte
vectorial (SVM)**

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo



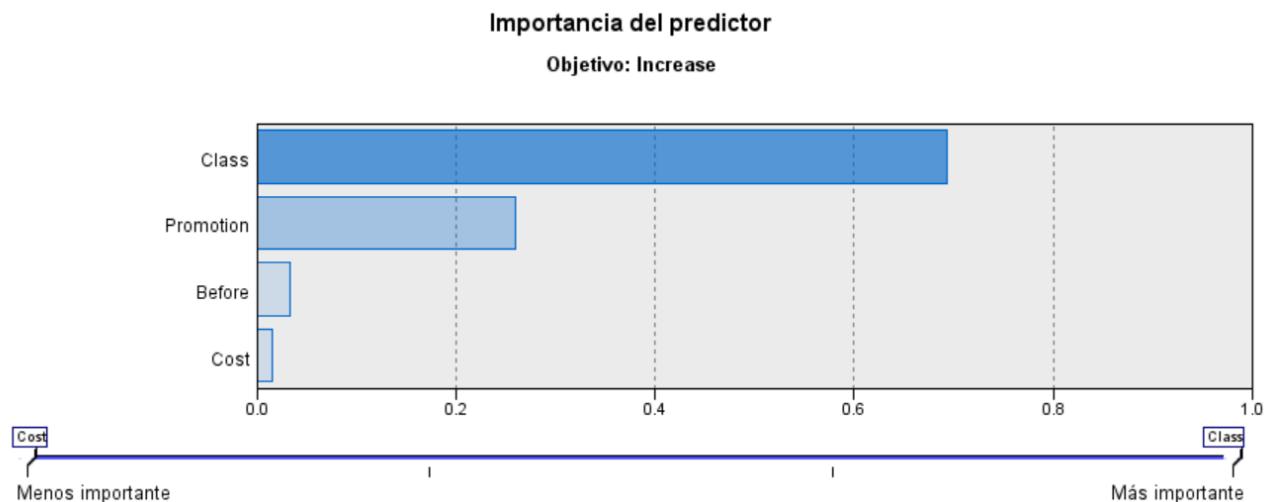
Generación del modelo
**máquina de soporte
vectorial (SVM)**

4) Modelado.

➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

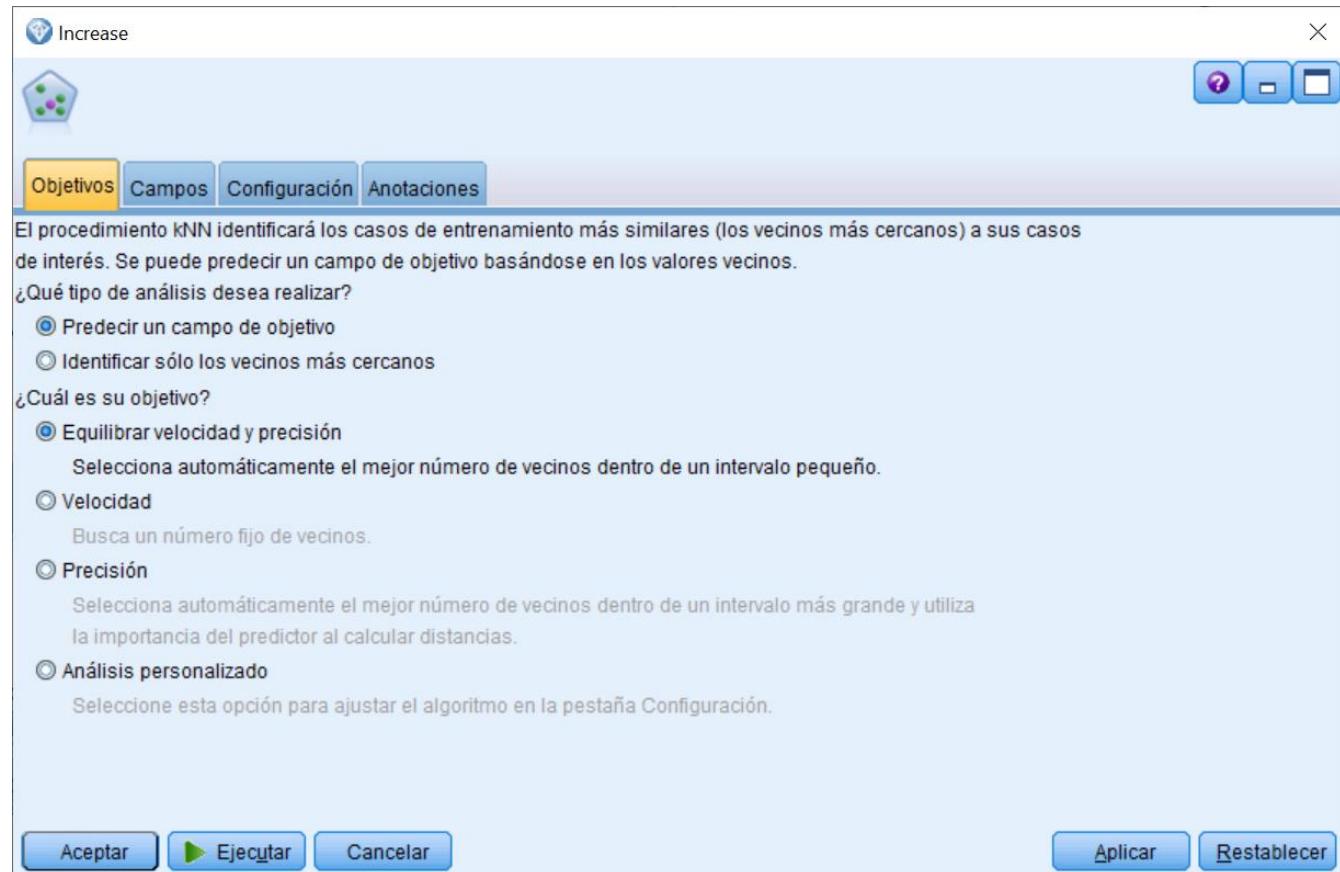
Ejecución del modelo
**máquina de soporte
vectorial (SVM)**



4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo



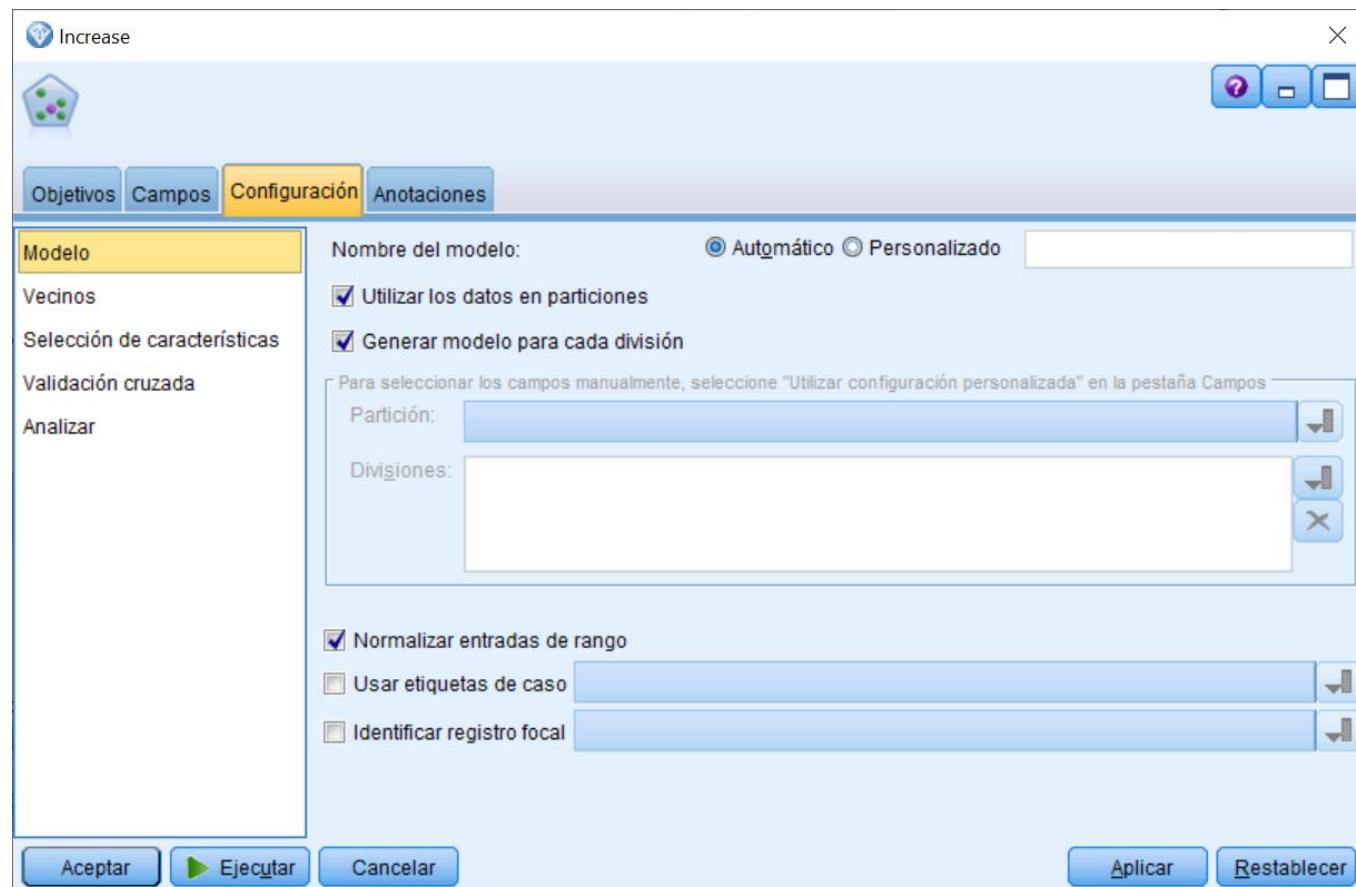
Generación del modelo algoritmo KNN

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Generación del modelo
algoritmo KNN

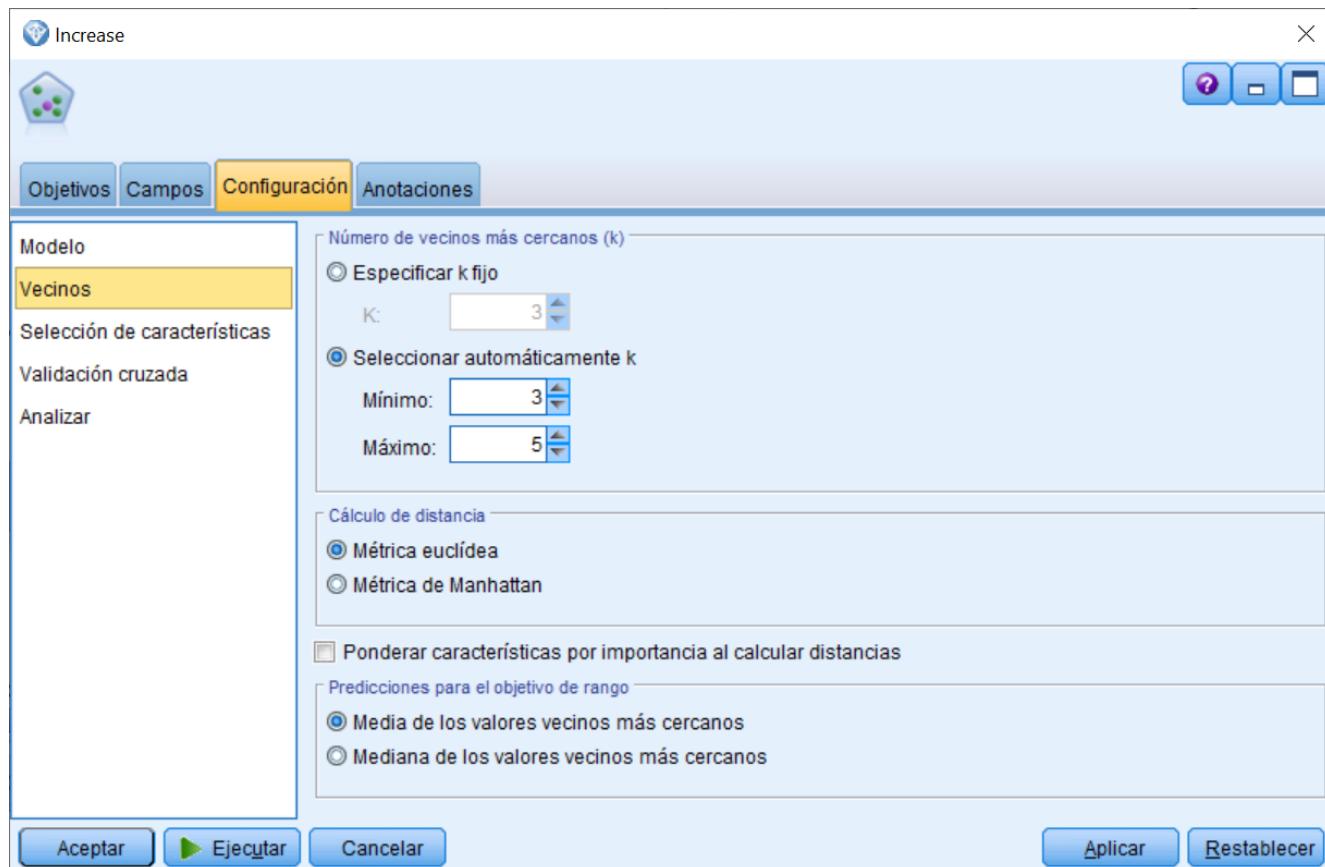


4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Generación del modelo
algoritmo KNN

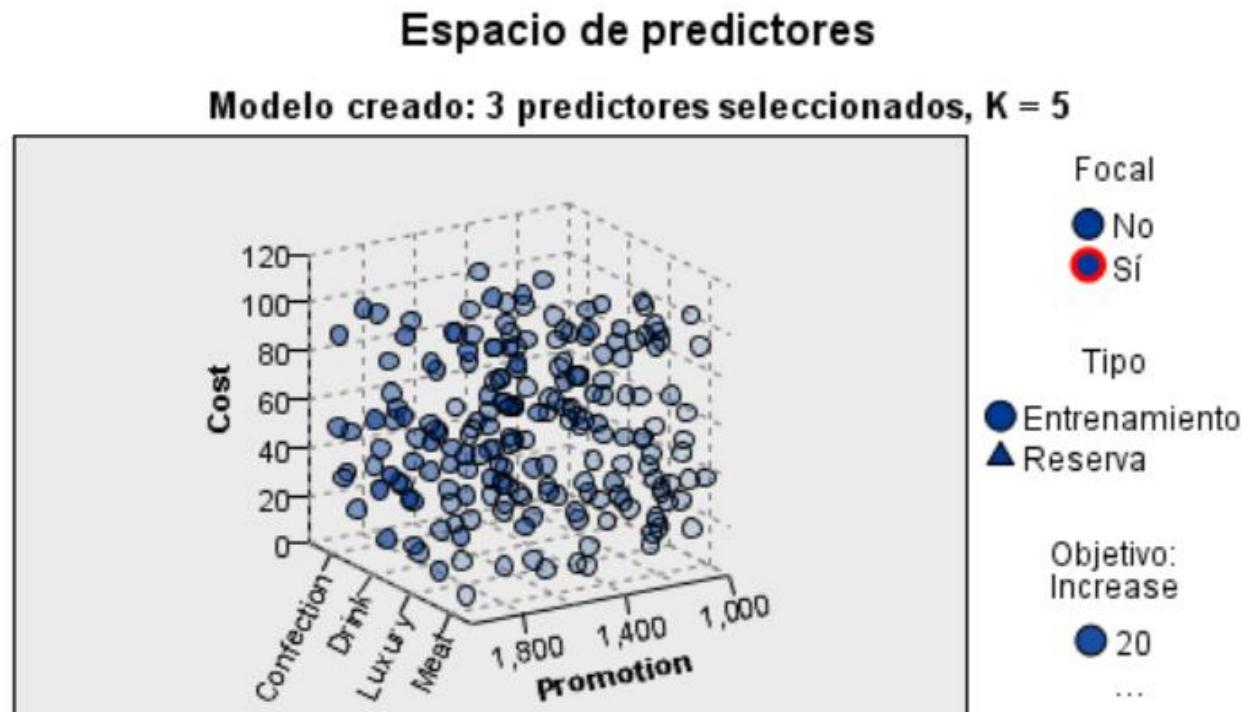


4) Modelado.

➤ Generación de los modelos: Ejecución del modelo

Caso de estudio en el mercado de bienes de consumo

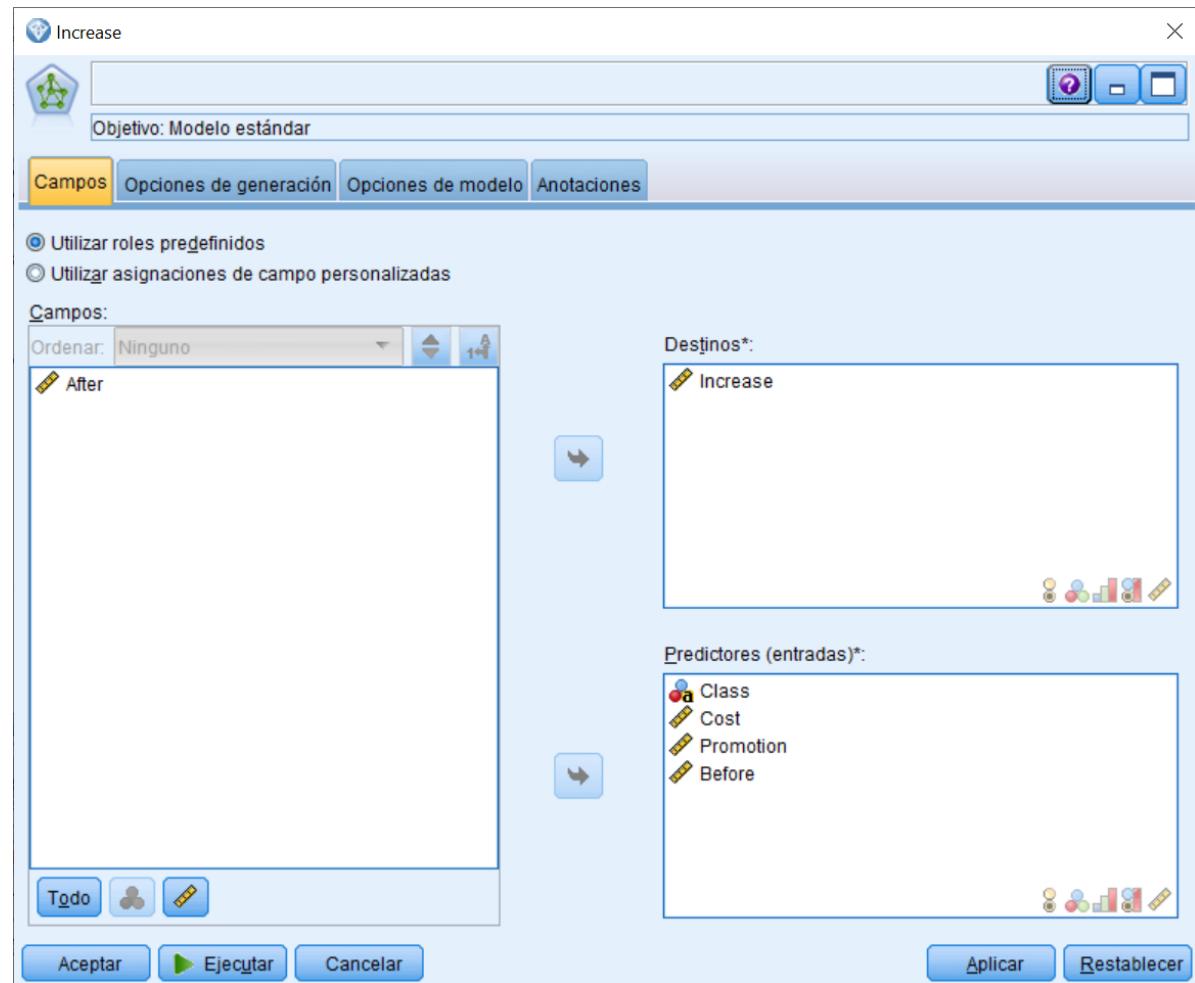
Ejecución del modelo
algoritmo KNN



4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

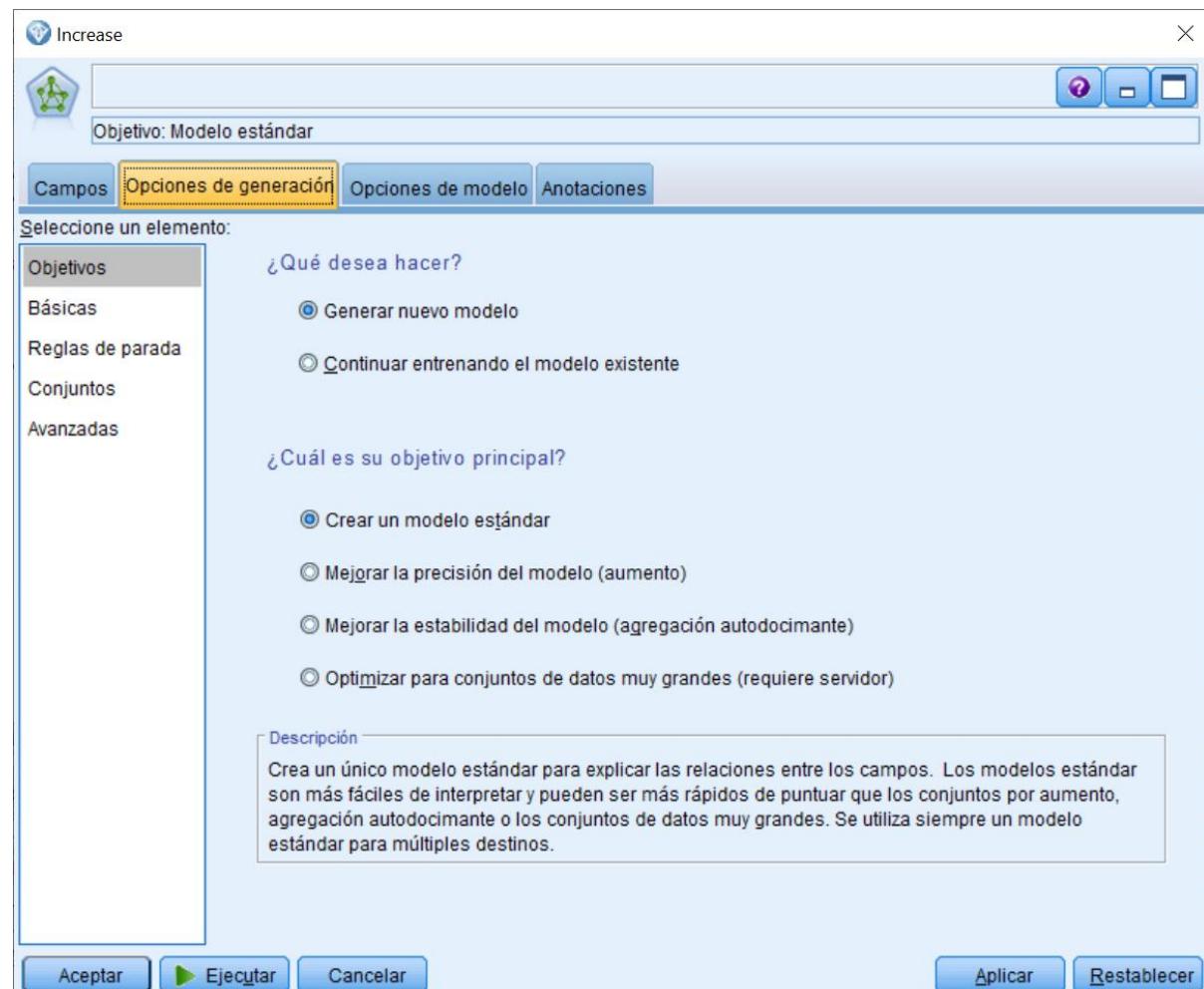


Generación del modelo
red neuronal
Perceptron
Backpropagation

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo

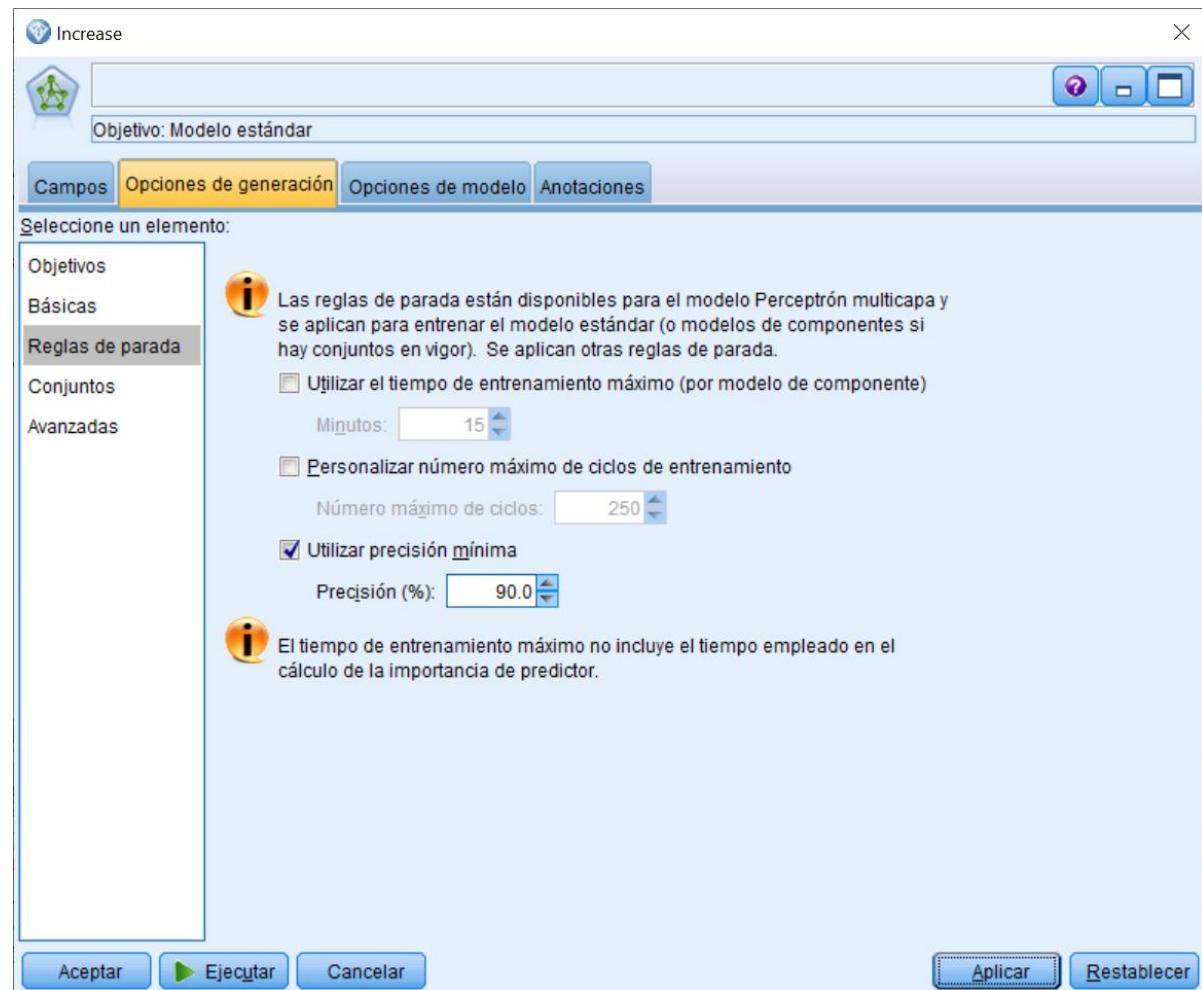


Generación del modelo
red neuronal
Perceptron
Backpropagation

4) Modelado.

➤ Generación de los modelos.

Caso de estudio en el mercado de bienes de consumo



Generación del modelo
red neuronal
Perceptron
Backpropagation

4) Modelado.

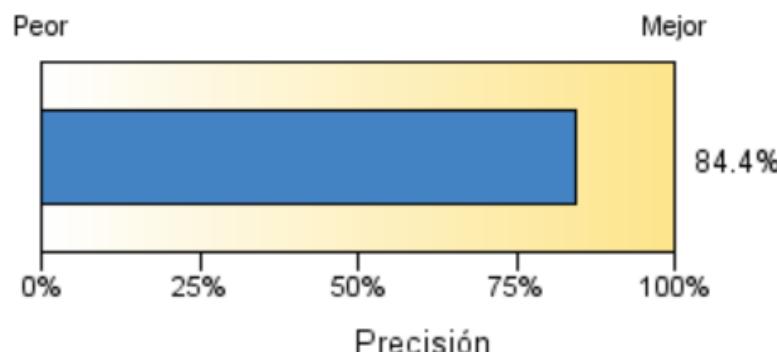
➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

Resumen del modelo

Destino	Increase
Modelo	Perceptrón multicapa
Regla de parada utilizada	El error no se puede disminuir más
Neuronas de la capa oculta 1	5

Ejecución del modelo
red neuronal
Perceptron
Backpropagation

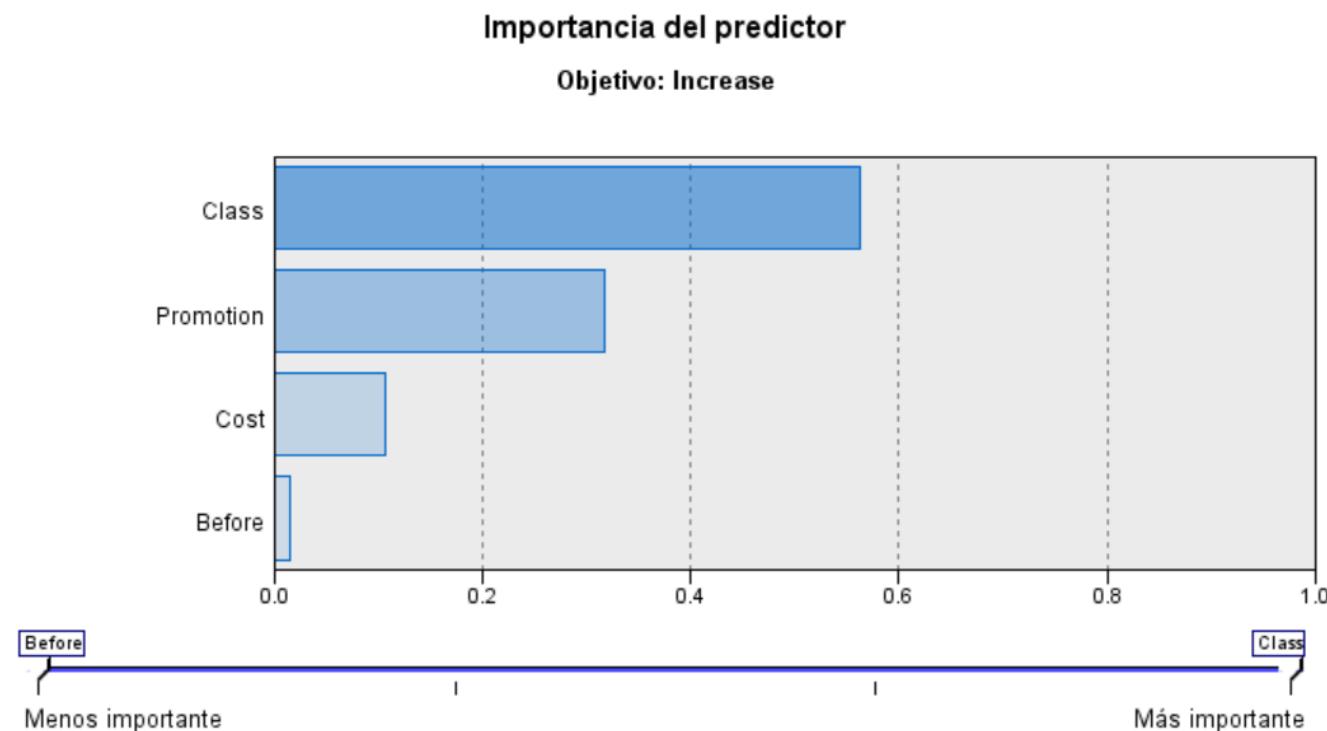


4) Modelado.

➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

Ejecución del modelo
red neuronal
Perceptron
Backpropagation

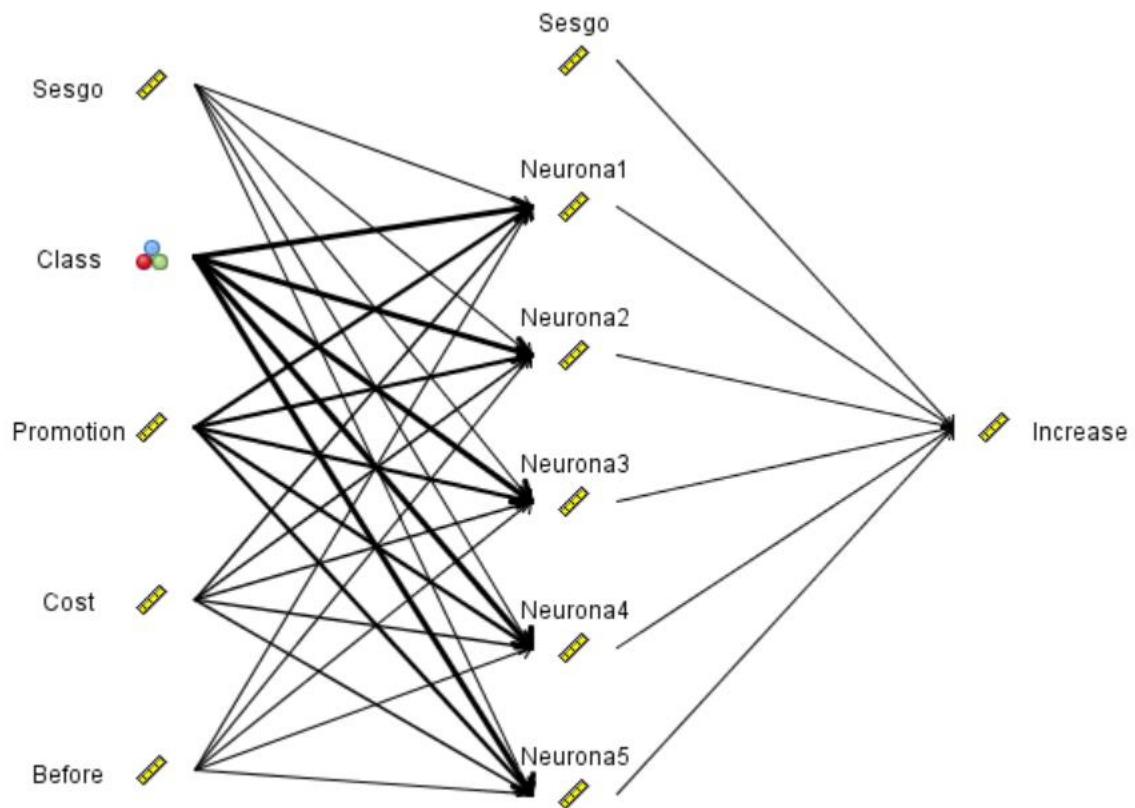


4) Modelado.

➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

Ejecución del modelo
red neuronal
Perceptron
Backpropagation

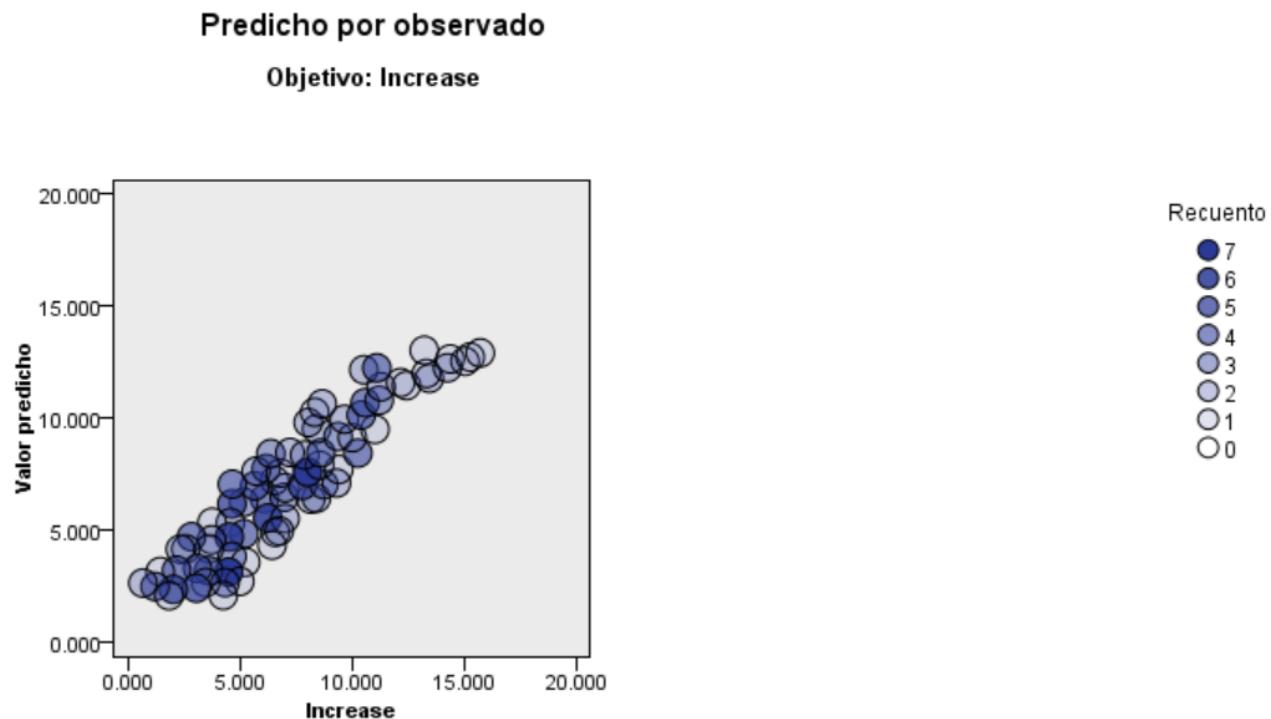


4) Modelado.

➤ Generación de los modelos: Ejecución del modelo.

Caso de estudio en el mercado de bienes de consumo

Ejecución del modelo
red neuronal
Perceptron
Backpropagation

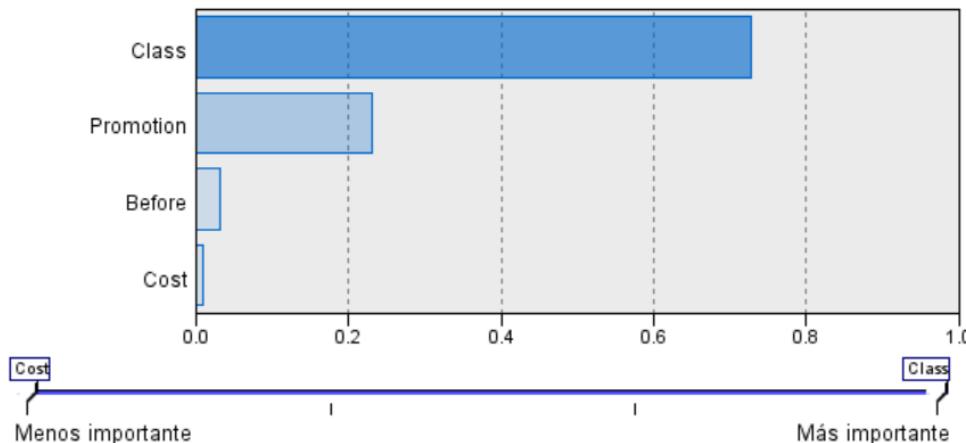


4) Modelado.

➤ Evaluación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Evaluación del modelo Árbol de Decisión C&R



- Resultados para el campo de resultado Increase
- Comparando \$R-Increase con Increase

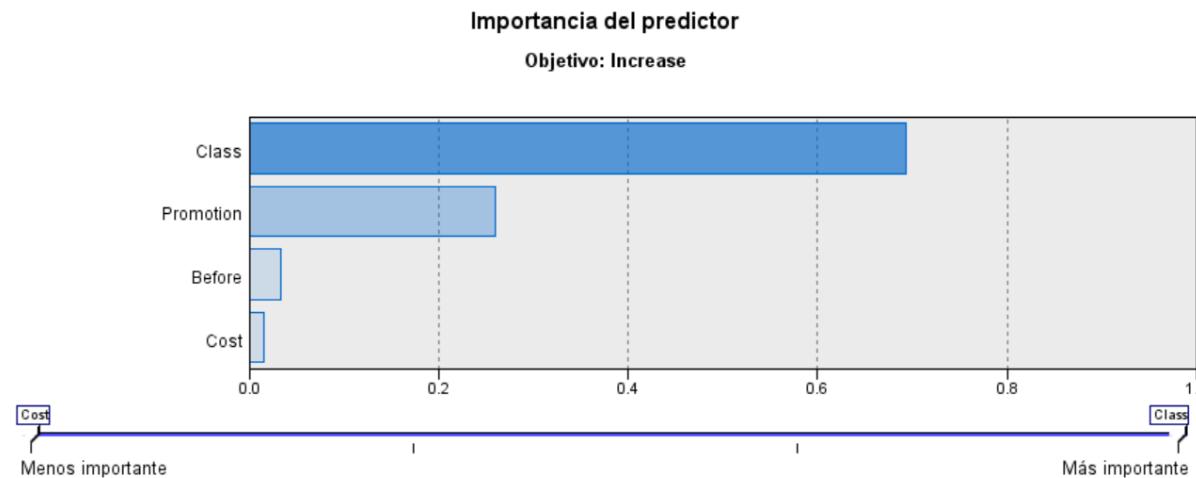
Error mínimo	-2.535
Error máximo	3.388
Error promedio	0.024
Error absoluto promedio	0.967
Desviación estándar	1.176
Correlación lineal	0.934
Ocurrencias	200

4) Modelado.

➤ Evaluación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Evaluación del modelo **Máquina de Soporte Vectorial (SVM)**



Resultados para el campo de resultado Increase

Comparando \$S-Increase con Increase

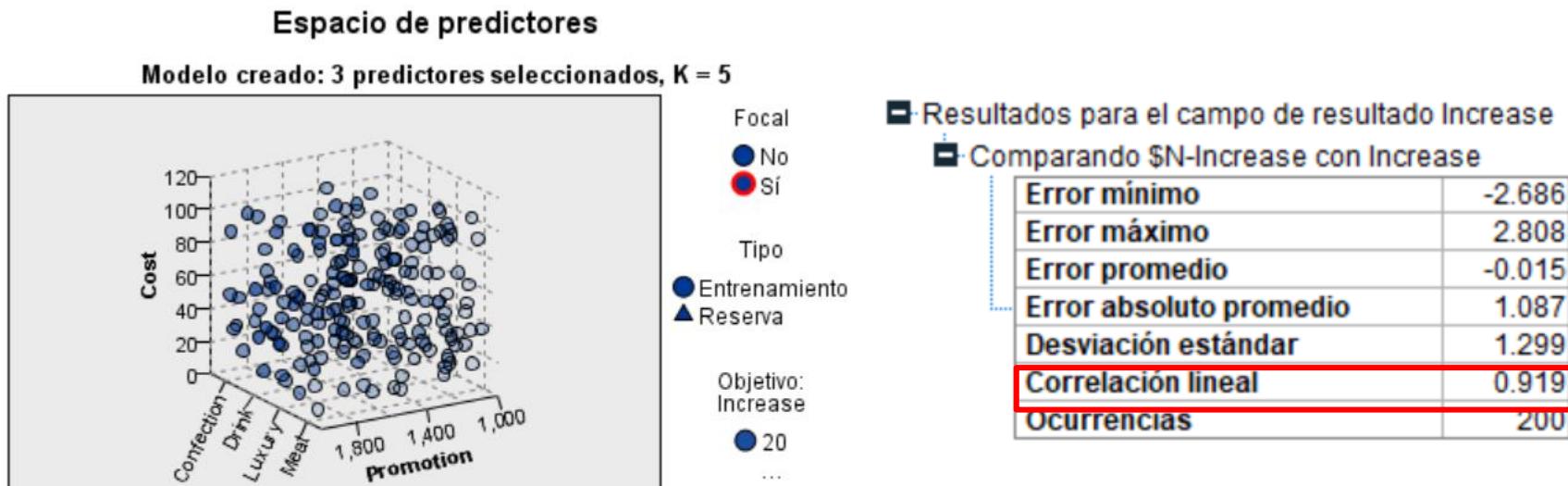
Error mínimo	-2.875
Error máximo	2.625
Error promedio	-0.114
Error absoluto promedio	1.04
Desviación estándar	1.273
Correlación lineal	0.923
Ocurrencias	200

4) Modelado.

➤ Evaluación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Evaluación del modelo Algoritmo KNN

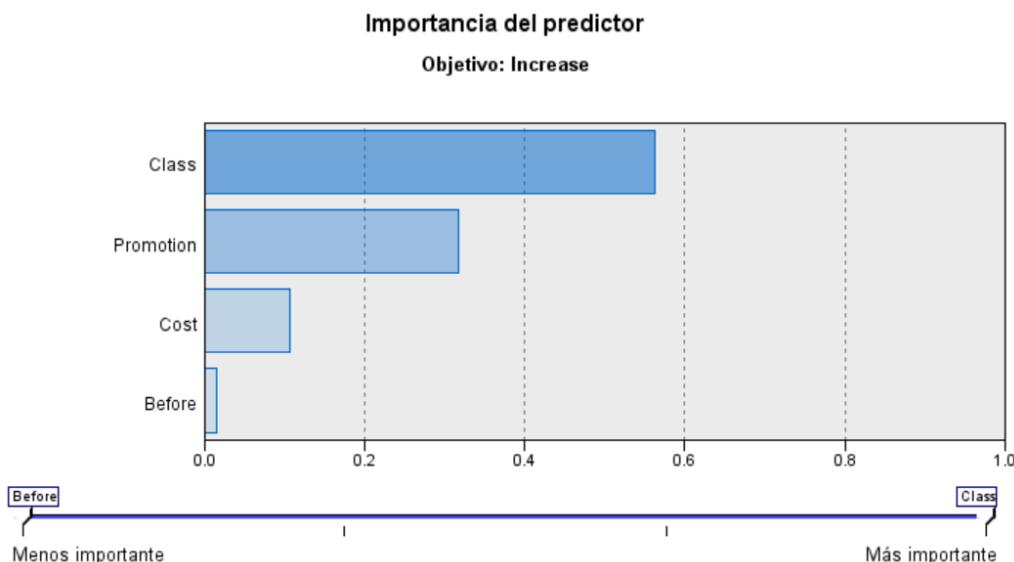


4) Modelado.

➤ Evaluación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Evaluación del modelo **Red Neuronal Perceptron Backpropagation**



Resultados para el campo de resultado Increase

Comparando \$N-Increase con Increase

Error mínimo	-2.686
Error máximo	2.808
Error promedio	-0.015
Error absoluto promedio	1.087
Desviación estándar	1.299
Correlación lineal	0.919
Ocurrencias	200

4) Modelado.

➤ Evaluación de los modelos.

Caso de estudio en el mercado de bienes de consumo

Árbol de Decisión C&R

- Resultados para el campo de resultado Increase

- Comparando \$R-Increase con Increase

Error mínimo	-2.535
Error máximo	3.388
Error promedio	0.024
Error absoluto promedio	0.967
Desviación estándar	1.176
Correlación lineal	0.934
Ocurrencias	200

Algoritmo KNN

- Resultados para el campo de resultado Increase

- Comparando \$N-Increase con Increase

Error mínimo	-2.686
Error máximo	2.808
Error promedio	-0.015
Error absoluto promedio	1.087
Desviación estándar	1.299
Correlación lineal	0.919
Ocurrencias	200

Máquina de Soporte Vectorial (SVM)

- Resultados para el campo de resultado Increase

- Comparando \$S-Increase con Increase

Error mínimo	-2.875
Error máximo	2.625
Error promedio	-0.114
Error absoluto promedio	1.04
Desviación estándar	1.273
Correlación lineal	0.923
Ocurrencias	200

Perceptron Backpropagation

- Resultados para el campo de resultado Increase

- Comparando \$N-Increase con Increase

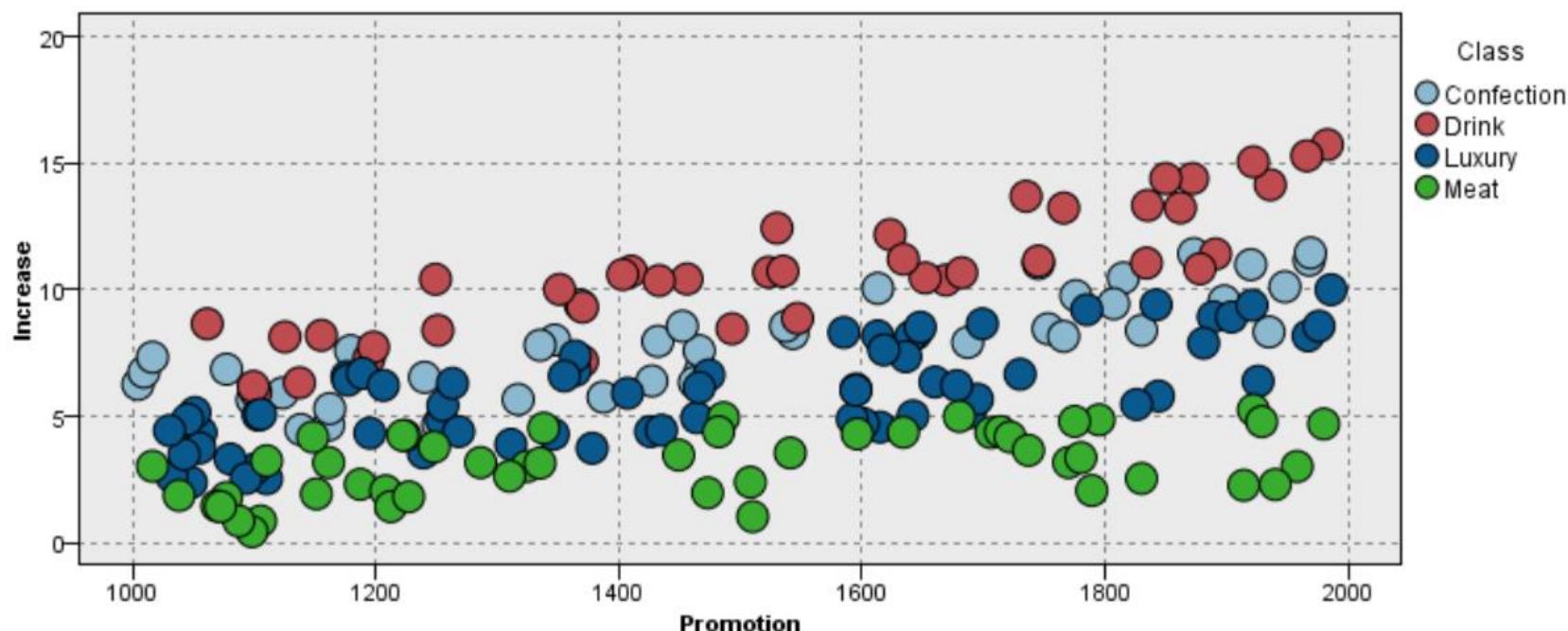
Error mínimo	-2.686
Error máximo	2.808
Error promedio	-0.015
Error absoluto promedio	1.087
Desviación estándar	1.299
Correlación lineal	0.919
Ocurrencias	200

5) Evaluación.

5. Evaluación

Caso de estudio en el mercado de bienes de consumo

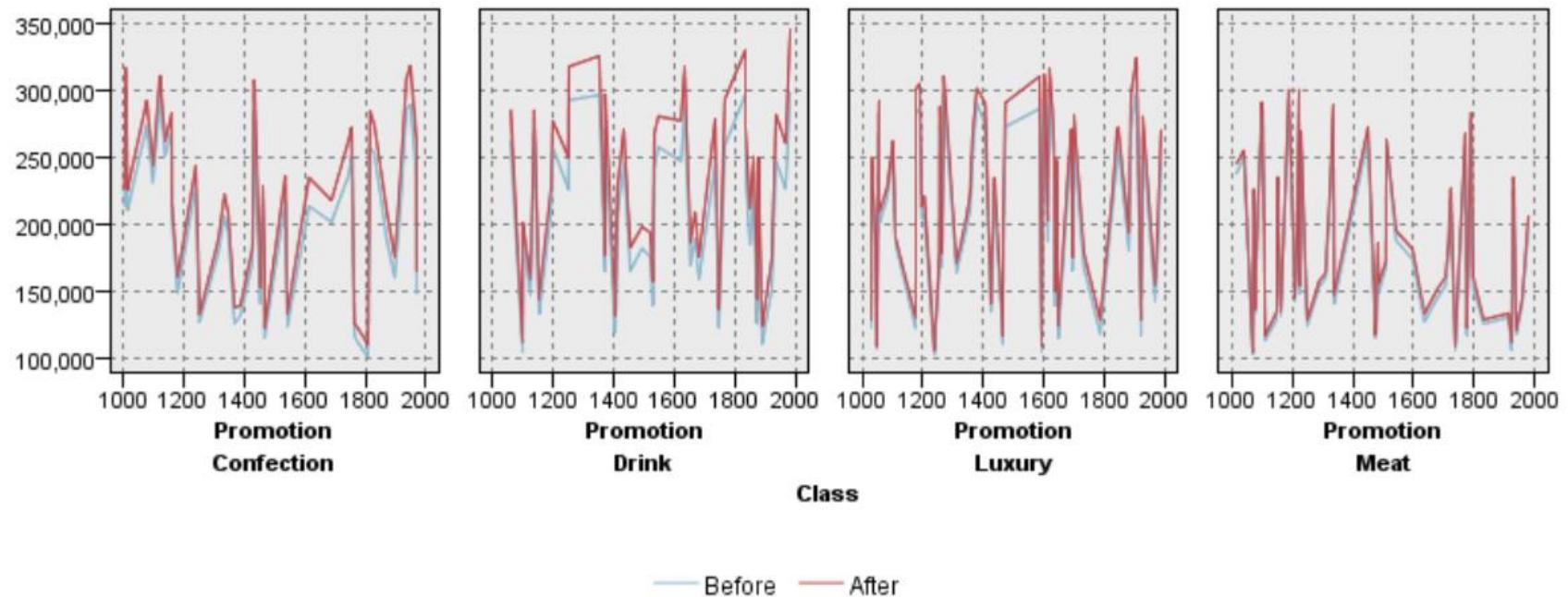
- Evaluación de los resultados producidos por los gráficos y modelos, con el objetivo de identificar mejoras para que la promoción resulte mucho más atractiva, en aquellos tipos de bienes de consumo donde no se obtuvieron los resultados previstos.



5) Evaluación.

5. Evaluación

Caso de estudio en el mercado de bienes de consumo



5) Evaluación.

5. Evaluación

Caso de estudio en el mercado de bienes de consumo

Árbol de Decisión C&R

- Resultados para el campo de resultado Increase
 - Comparando \$R-Increase con Increase

Error mínimo	-2.535
Error máximo	3.388
Error promedio	0.024
Error absoluto promedio	0.967
Desviación estándar	1.176
Correlación lineal	0.934
Ocurrencias	200

Perceptron Backpropagation

- Resultados para el campo de resultado Increase
 - Comparando \$N-Increase con Increase

Error mínimo	-2.686
Error máximo	2.808
Error promedio	-0.015
Error absoluto promedio	1.087
Desviación estándar	1.299
Correlación lineal	0.919
Ocurrencias	200

5) Evaluación.

5. Evaluación

Caso de estudio en el mercado de bienes de consumo

- A partir de los resultados producidos por los gráficos y modelos seleccionados – Árbol de Decisión C&R y Red Neuronal *Perceptron Backpropagation* – se efectuaron las siguientes recomendaciones, las cuales fueron discutidas y aceptadas por la directiva de la organización:
 - ❖ Efectuar mejoras para que la promoción resulte mucho más atractiva, en aquellos tipos de bienes de consumo donde no se obtuvieron los resultados previstos.
 - ❖ Utilizar los modelos seleccionados para predecir el incremento que se observará en las ventas – *Increase* – de un tipo particular de bien de consumo a partir de los datos de *Class, Cost, Promotion* y *Before*.

6) Despliegue.

6. Despliegue

- Considerando los objetivos comerciales del proyecto y los resultados proporcionados por el proyecto de minería de datos, la directiva de la compañía tomó las siguientes decisiones:
 - ❖ Efectuar mejoras para que la promoción resulte mucho más atractiva, en aquellos tipos de bienes de consumo donde no se obtuvieron los resultados previstos.
 - ❖ Efectuar variaciones en el monto de la promoción, en dependencia de los resultados obtenidos para cada tipo de bien de consumo.
 - ❖ Utilizar los modelos seleccionados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.

6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

6. Despliegue

- Uso de los modelos generados para la predicción del incremento en las ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Incrementos a predecir:

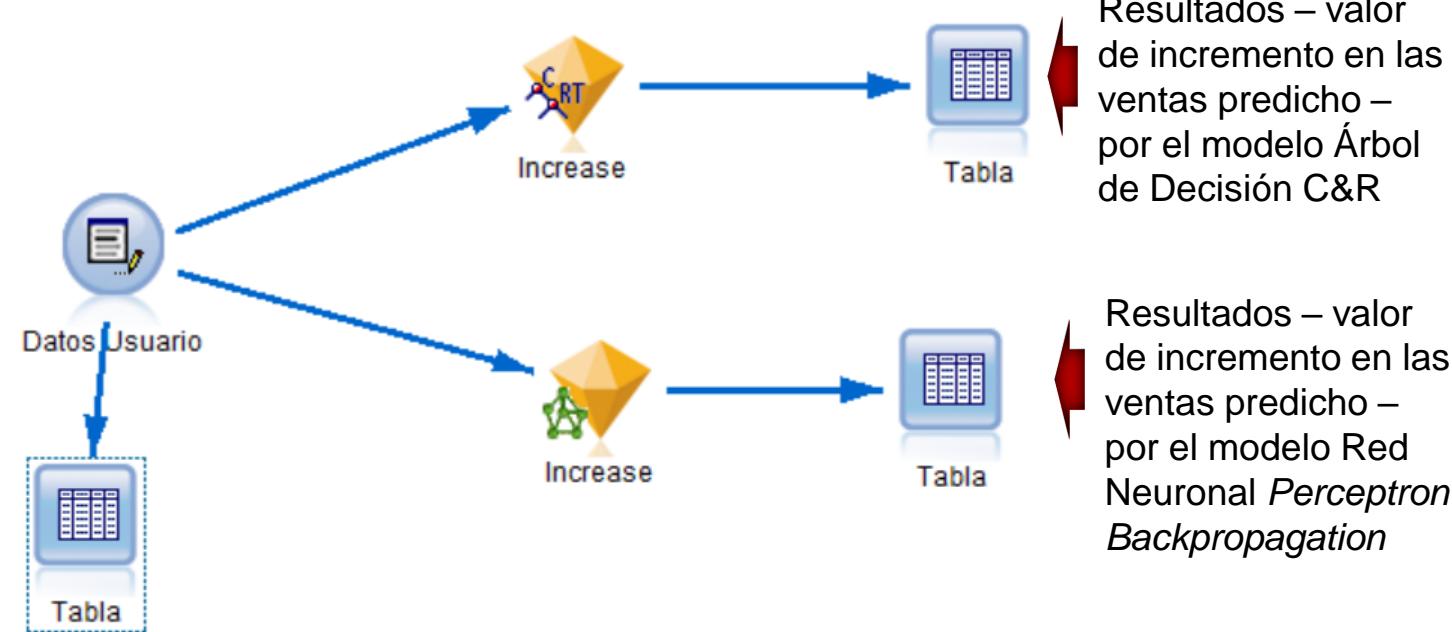
Class	Cost	Promotion	Before
Confection	24.5	1467	117400
Drink	85	1745	132262
Luxury	90	1426	148458
Meat	75	1486	159862

6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

6. Despliegue

- Uso de los modelos generados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Modelos a utilizar:



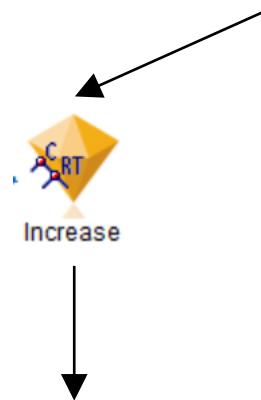
6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

6. Despliegue

- Uso de los modelos generados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Valor predicho del incremento en las ventas (**Increase**) para el bien de consumo de tipo **Confection**:

Class	Cost	Promotion	Before
Confection	24.5	1467	117400



	Class	Cost	Promotion	Before	\$R-Increase
1	Confection	24....	1467.000	117400....	7.522

	Class	Cost	Promotion	Before	\$N-Increase
1	Confection	24....	1467.000	117400....	7.559

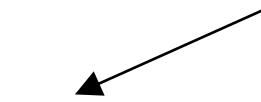
6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

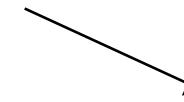
6. Despliegue

- Uso de los modelos generados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Valor predicho del incremento en las ventas (**Increase**) para el bien de consumo de tipo **Drink**:

Class	Cost	Promotion	Before
Drink	85	1745	132262



	Class	Cost	Promotion	Before	\$R-Increase
1	Drink	85....	1745.000	132262....	12.640



	Class	Cost	Promotion	Before	\$N-Increase
1	Drink	85....	1745.000	132262....	12.415

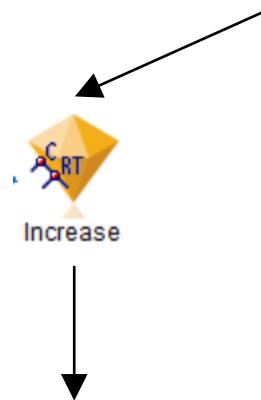
6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

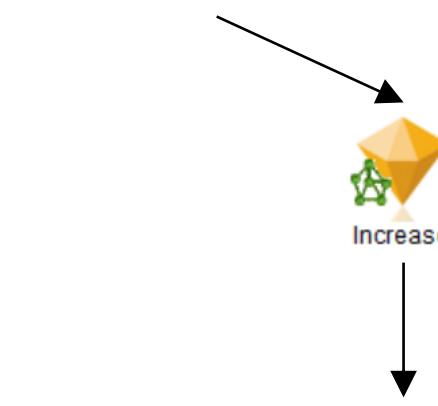
6. Despliegue

- Uso de los modelos generados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Valor predicho del incremento en las ventas (*Increase*) para el bien de consumo de tipo *Luxury*:

Class	Cost	Promotion	Before
Luxury	90	1426	148458



	Class	Cost	Promotion	Before	\$R-Increase
1	Luxury	90....	1426.000	148458....	4.908



	Class	Cost	Promotion	Before	\$N-Increase
1	Luxury	90....	1426.000	148458....	6.089

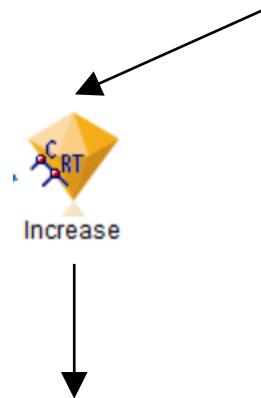
6) Despliegue.

Caso de estudio en el mercado de bienes de consumo

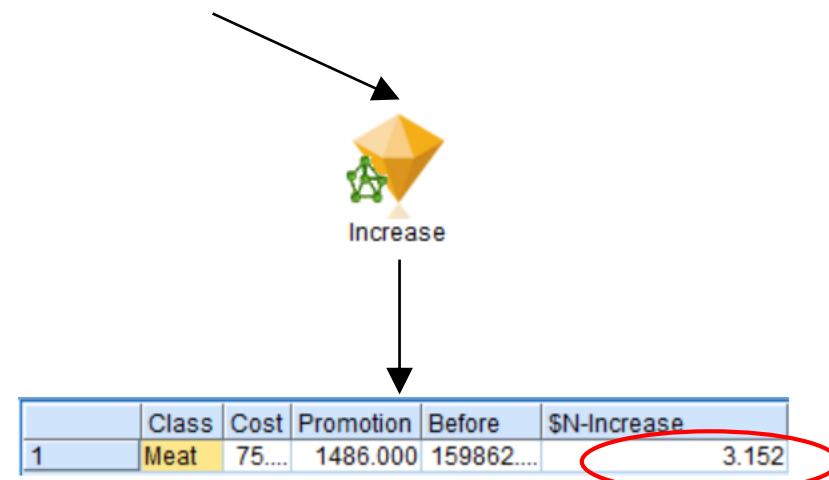
6. Despliegue

- Uso de los modelos generados para la predicción del incremento en la ventas de bienes de consumo, en base a la categoría de bien de consumo y la promoción aplicada.
- Valor predicho del incremento en las ventas (*Increase*) para el bien de consumo de tipo *Meat*.

Class	Cost	Promotion	Before
Meat	75	1486	159862



	Class	Cost	Promotion	Before	\$R-Increase
1	Meat	75....	1486.000	159862....	3.761



	Class	Cost	Promotion	Before	\$N-Increase
1	Meat	75....	1486.000	159862....	3.152