**Pronóstico de ventas de Walmart**

La ciencia de datos juega un papel muy importante en la previsión de ventas y riesgos en el sector minorista. La mayoría de las tiendas minoristas líderes implementan Data Science para realizar un seguimiento de las necesidades de sus clientes y tomar mejores decisiones comerciales. Walmart es uno de esos minoristas.

**Declaración del problema**: analizar el conjunto de datos de ventas de Walmart para predecir las ventas por departamento para cada una de sus tiendas.

**Descripción del conjunto de datos**: el conjunto de datos utilizado para este proyecto contiene datos de captación de información histórica, que cubren los detalles de ventas desde 2010-02-05 hasta 2012-11-01. Para el análisis de este problema se utilizan las siguientes variables predictoras:

* Tienda – el número de la tienda
* Departamento: el número del departamento
* Fecha – la semana
* IPC: el índice de precios al consumidor
* Weekly\_Sales: ventas para el departamento dado en la tienda dada
* IsHoliday: si la semana es una semana festiva especial

Al estudiar la dependencia de estas variables predictoras en la variable de **respuesta**, puede predecir o pronosticar las ventas para los próximos meses.

**Lógica:**

**Importar el conjunto de datos**: el conjunto de datos necesario para este proyecto se puede descargar desde Kaggle.

**Limpieza de datos**: en esta etapa, debe asegurarse de deshacerse de todas las inconsistencias, como valores faltantes y variables redundantes.

**Exploración de datos**: en esta etapa, puede trazar diagramas de caja y qplots para comprender la importancia de cada variable predictora.

**Modelado de datos**: para esta declaración de problema en particular, dado que el resultado es una variable continua (Número de ventas), es razonable construir un modelo de regresión. El algoritmo de regresión lineal se puede usar para resolver tales problemas, ya que se usa específicamente para predecir variables dependientes continuas.

**Valide el modelo**: en esta etapa, debe evaluar la eficiencia del modelo de datos utilizando el conjunto de datos de prueba y, finalmente, calcular la precisión del modelo utilizando una matriz de confusión.

<https://medium.com/analytics-vidhya/walmart-sales-forecast-41c6dc1028b8>

<https://rpubs.com/erickamtzrdz/940298>