**Tarea Final Optativa VI Sistemas Dinámicos**

**Conjunto de datos sobre la intención de compra de los compradores en línea.**

**Subject área: Business**

**Integrantes:**

**Ander Morales Montero**

**Kevin Caballero Olazabal**

**Jose Javier Curbelo Toledo**

**2024**

<https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset>

De las 12.330 sesiones del conjunto de datos, el 84,5% (10.422) eran muestras de clase negativa que no terminaban en compra, y el resto (1.908) eran muestras de clase positiva que terminaban en compra.

1. entrenar la red neuronal para predecir si una persona efectuará una compra o no
2. identificar qué variables o características específicas de los clientes están más relacionadas con la compra efectiva
3. La predicción de si un cliente efectuará una compra o no se basa en el análisis de un conjunto de características seleccionadas mediante un proceso de selección de características llamado Recursive Feature Elimination (RFE). Este proceso utiliza un modelo de regresión logística como base para identificar las características más influyentes en la predicción de la variable objetivo, que en este caso es la decisión de compra. La regresión logística es una técnica de análisis predictivo que modela la probabilidad de eventos binarios, permitiendo predecir la probabilidad de que un cliente efectúe una compra basándose en un conjunto de características independientes.
4. Las características más relevantes identificadas a través de RFE son aquellas que el modelo de regresión logística considera las más importantes para predecir la decisión de compra. Estas características, como 'ProductRelated', 'ExitRates', 'PageValues', 'SpecialDay', y 'Month', tienen un impacto significativo en la capacidad del modelo para predecir si un cliente efectuará una compra o no. La importancia de cada característica se determina midiendo cómo afecta el rendimiento del modelo cuando se elimina, y las características que tienen el mayor impacto en el rendimiento del modelo cuando se eliminan son consideradas las más importantes.

**Resultados de los incisos en los archivos CSV:**

<predicciones.csv>  
  
<caracteristicas_relevantes.csv>