

EJERCICIO

ALUMNO: Pacsi Ramos Jose Luis Armando

Predicción de Compra de Cliente(10,000)

Supongamos que estamos trabajando para una tienda en línea y queremos predecir si un cliente realizará una compra en base a ciertas características. El dataset contiene las siguientes columnas:

- `age`: Edad del cliente.
- `income`: Ingreso mensual del cliente.
- `gender`: Género del cliente.
- `web_visits`: Número de visitas al sitio web en el último mes.
- `purchased`: (Variable objetivo) indica si el cliente compró (1) o no (0).

Crearemos un dataset sintético con estas variables y luego aplicaremos tres métodos para compararlos

1.- Curvas ROC y AUC

2.- XGBoost

3.- Máquina de Soporte Vectorial (SVM)

4.- Regresión Logística

5.- Random Forest

4.- Árbol de Decisión

Finalmente Realizar la comparación de los modelos y su interpretación de los resultados.

INTERPRETACIÓN

1. Ejes y Elementos Básicos del Gráfico:

- **Eje X:** En el eje horizontal, vemos tres modelos: **Árbol de Decisión**, **Random Forest** y **XGBoost**. Cada grupo de barras representa un modelo de predicción.
- **Eje Y:** En el eje vertical, está el valor de cada métrica, que va de 0 a 1. Un valor más alto indica un mejor desempeño en esa métrica.
- **Colores de las Barras:** Cada color representa una métrica específica:
 - **Rojo:** AUC (Área Bajo la Curva ROC).
 - **Verde:** Precisión.
 - **Azul:** Sensibilidad.
 - **Morado:** Especificidad.

2. Significado de las Métricas:

- **AUC:** Esta métrica nos ayuda a entender qué tan bien el modelo distingue entre los clientes que compran y los que no. Un AUC cercano a 1 significa que el modelo es muy bueno para hacer esta distinción.
- **Precisión:** Indica cuántas de las predicciones de compra del modelo fueron correctas. Si este valor es alto, el modelo está acertando en la mayoría de sus predicciones positivas.
- **Sensibilidad:** Mide la capacidad del modelo para identificar correctamente a los clientes que compraron. Si es alta, el modelo está captando bien a los compradores.
- **Especificidad:** Evalúa cuántos de los que realmente no compraron fueron identificados correctamente por el modelo.

3. Comparación entre Modelos:

- **Árbol de Decisión:** En general, tiene un AUC y precisión más bajos en comparación con los otros dos modelos. Esto sugiere que este modelo no es tan efectivo para distinguir entre compradores y no compradores. La sensibilidad también es baja, lo que indica que pierde algunos casos de clientes que realmente compran.
- **Random Forest:** Este modelo tiene métricas bastante equilibradas y mejor rendimiento en comparación con el Árbol de Decisión. Su AUC, precisión, sensibilidad y especificidad son todas moderadamente altas, lo que lo convierte en una opción sólida y confiable para predecir la compra.
- **XGBoost:** Destaca con valores más altos en la mayoría de las métricas, especialmente en AUC y precisión, lo que indica que es el modelo que mejor predice quién comprará. XGBoost también tiene buena sensibilidad y especificidad, lo que significa que es excelente tanto para captar compradores como para evitar falsos positivos.

4. Conclusión:

- **Mejor Modelo:** De acuerdo con el gráfico, **XGBoost** es el modelo que ofrece el mejor rendimiento general, siendo más preciso y efectivo en la predicción de compra.
- **Modelo de Respaldo:** **Random Forest** también muestra buen desempeño, por lo que puede ser una alternativa si queremos un modelo menos complejo.
- **Modelo Menos Eficiente:** El **Árbol de Decisión** es el que tiene un rendimiento menor en la mayoría de las métricas, así que podría no ser la mejor opción para este problema.

En resumen, si nuestra tienda en línea quiere optimizar las predicciones de compra de clientes, **XGBoost** sería la mejor elección, seguido por **Random Forest** como una opción de respaldo.

