**RStudio - 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.

**Comparación de Resultados**

**Conclusiones**

* **Regresión Logística** es ideal cuando es crítico evitar falsos negativos, por ejemplo, en la detección de fraudes graves donde la omisión de un caso positivo puede ser costosa.
* **Random Forest** ofrece un mejor equilibrio entre sensibilidad y especificidad. Es una buena opción cuando se puede tolerar un cierto nivel de falsos positivos, pero se busca también minimizar los falsos negativos.
* **Árbol de Decisión** es efectivo para maximizar la detección de casos positivos, pero debido a su baja especificidad, podría no ser adecuado en contextos donde sea necesario evitar falsos positivos.

Si el objetivo es maximizar la detección de fraudes sin preocuparse demasiado por los falsos positivos, Regresión Logística o Árbol de Decisión son opciones adecuadas. Sin embargo, para un enfoque más equilibrado, donde se busca mejorar tanto la precisión como la reducción de falsos positivos, Random Forest sería la elección preferida.

**Regresión Logística:**

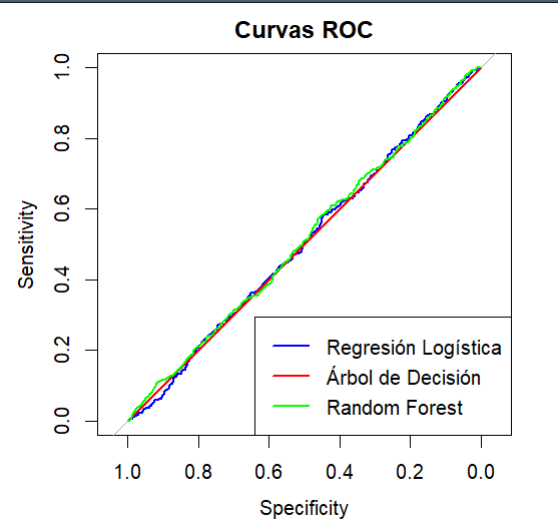
* Precisión: 71%
* Sensibilidad: 100%
* Especificidad: 3%
* F1 Score: 83%

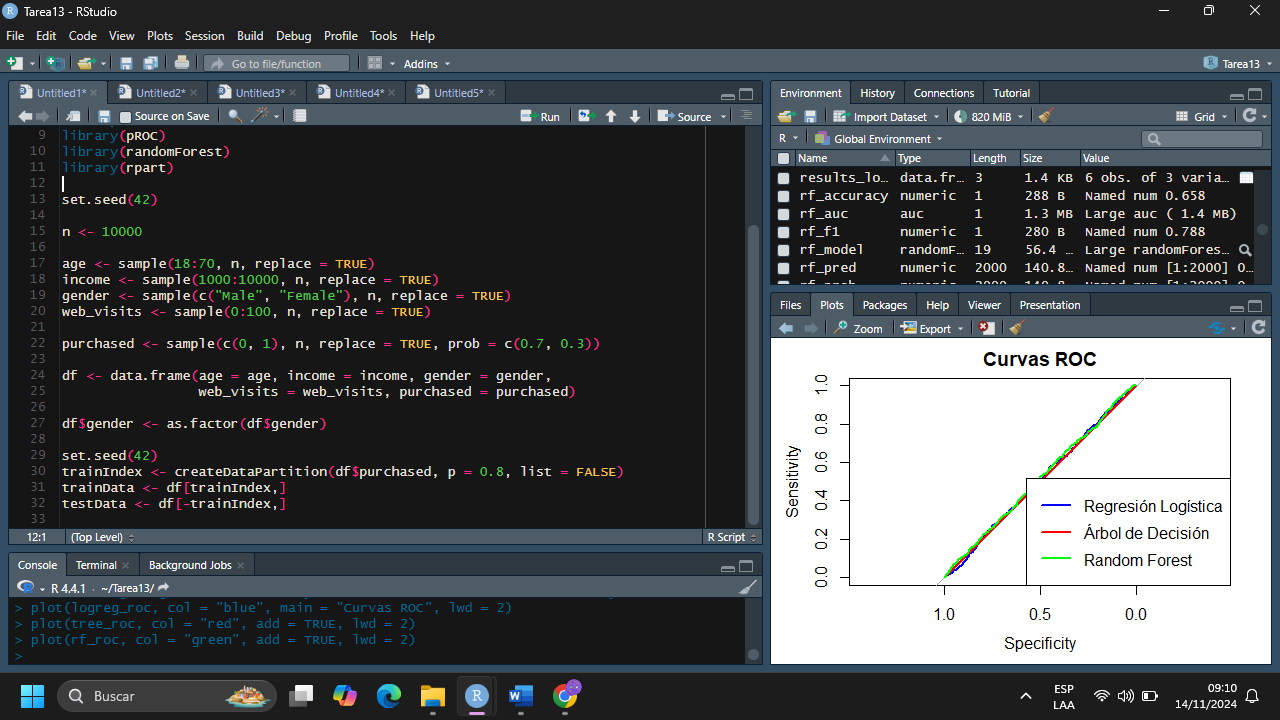
**Random Forest:**

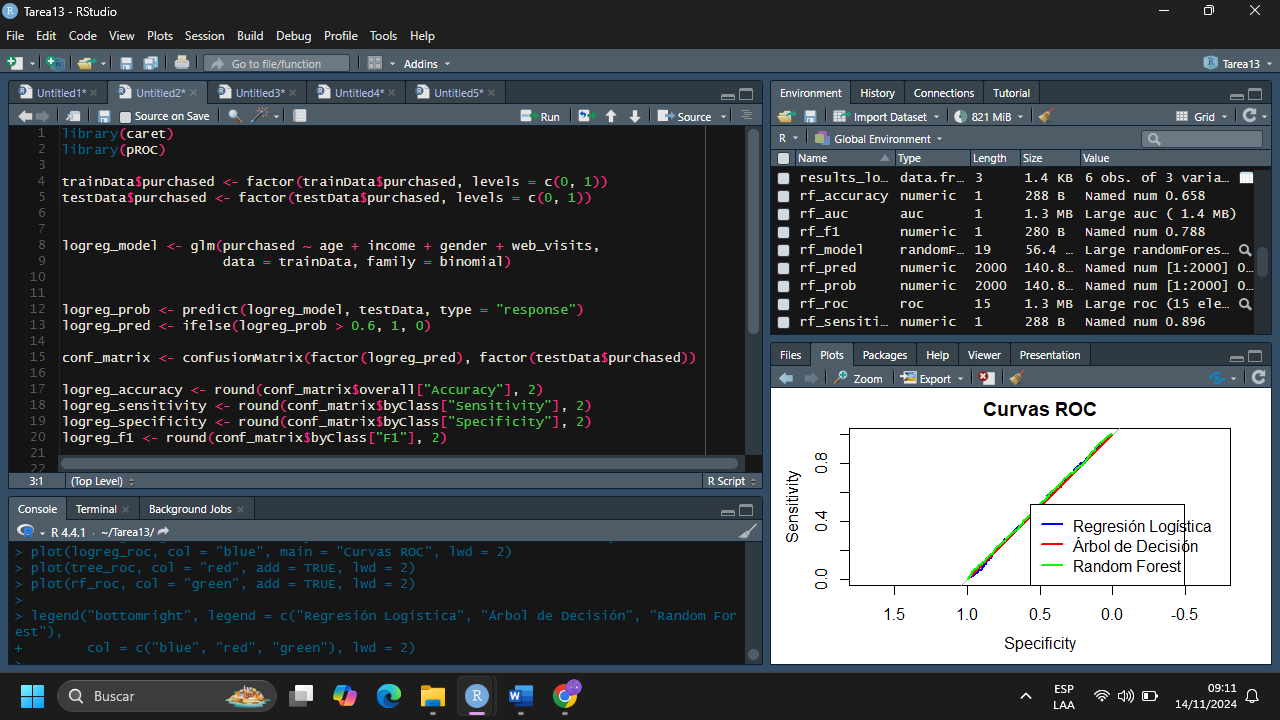
* Precisión: 66.1%
* Sensibilidad: 89.90%
* Especificidad: 8.39%
* F1 Score: 78.97%

**Árbol de Decisión:**

* Precisión: 70.8%
* Sensibilidad: 100%
* Especificidad: 0%
* F1 Score: 82.90%





Pantalla de computadora

Descripción generada automáticamente