

# Clasificación de Clientes - Full Data

Muestren la tabla de métricas del Test and Score. ¿Qué modelo logró el mejor AUC y cuál el mejor F1-score?

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.906	0.908	0.896	0.892	0.908	0.398
Naive Bayes	0.825	0.784	0.820	0.891	0.784	0.353
Logistic Regression	0.925	0.915	0.910	0.908	0.915	0.498

El mejor AUC (0.925) lo obtuvo Logistic Regression, lo que indica que es el modelo con mejor capacidad de discriminar entre las clases.

También fue el mejor F1-Score (0.910), mostrando un equilibrio sólido entre precisión y recall.

**A partir de la Matriz de Confusión de la Regresión Logística, identifiquen y reporten los valores de TP, TN, FP y FN.**

		Predicted		
		no	yes	$\Sigma$
Actual	no	712	26	738
	yes	44	41	85
$\Sigma$		756	67	823

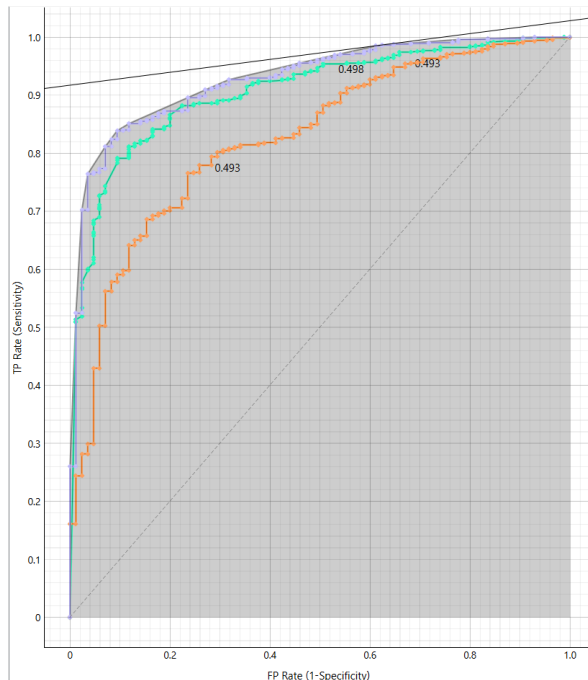
TP (Verdaderos Positivos): 41

TN (Verdaderos Negativos): 712

FP (Falsos Positivos): 26

FN (Falsos Negativos): 44

**Analicen la curva ROC: ¿Cuál de los tres modelos exhibe el mejor trade-off entre Tasa de Verdaderos Positivos (TPR) y Tasa de Falsos Positivos (FPR)? Justifiquen su respuesta basándose en la gráfica.**



El mejor trade-off lo logra Logistic Regression, ya que su curva está más cerca del vértice superior izquierdo (TPR alto, FPR bajo). Esto coincide con su mayor AUC (0.925), confirmando que es el modelo con mejor desempeño global.

## Ejercicio Teórico: Análisis de Métricas de Clasificación (Churn)

### A. Cálculo de Métricas

1. **Precisión:**  $VP / (VP + FP)$ 
  - a.  $350 / (350 + 100) = 0,7777$
2. **Exhaustividad:**  $VP / (VP + FN)$ 
  - a.  $350 / (350 + 150) = 0,7$
3. **Exactitud:**  $(VP + VN) / (VP + VN + FP + FN)$ 
  - a.  $(350 + 100) / (350 + 100 + 150 + 4400) = 0,09$
4. **Puntuación F1:**  $2 \times (Precision \cdot Recall) / (Precision + Recall)$ 
  - a.  $2 \times (0.7777 \cdot 0.7) / (0.7777 + 0.7) = 0.7368$

### B. Análisis de Impacto Estratégico (1 Punto)

#### 1. Implicaciones de Errores:

Falso Positivo: Terminamos ofreciendo descuentos/bonos a clientes que no se iban a ir. Eso encarece la operación (beneficios innecesarios) y además desenfoca el presupuesto de retención que podríamos usar en casos realmente riesgosos.

Falso Negativo: Perdemos al cliente sin haber hecho nada. Hay pérdida directa de ingresos y de LTV, posible impacto en churn rate y quizá costos extras posteriores para intentar re-adquirir a ese cliente (mucho más caro que retenerlo).

#### 2. Prioridad de Métrica:

Si la dirección pide evitar a toda costa regalar beneficios a quien no iba a abandonar, la métrica más relevante es la Precisión: mide qué porcentaje de los predichos como “abandona” realmente lo son.

En la práctica, se puede subir el umbral de decisión para privilegiar precisión y monitorear la curva Precision–Recall para elegir el punto que mejor balancee costo/beneficio.