

Ejercicio Práctico en Orange: Clasificación de Clientes (Regresión Logística)

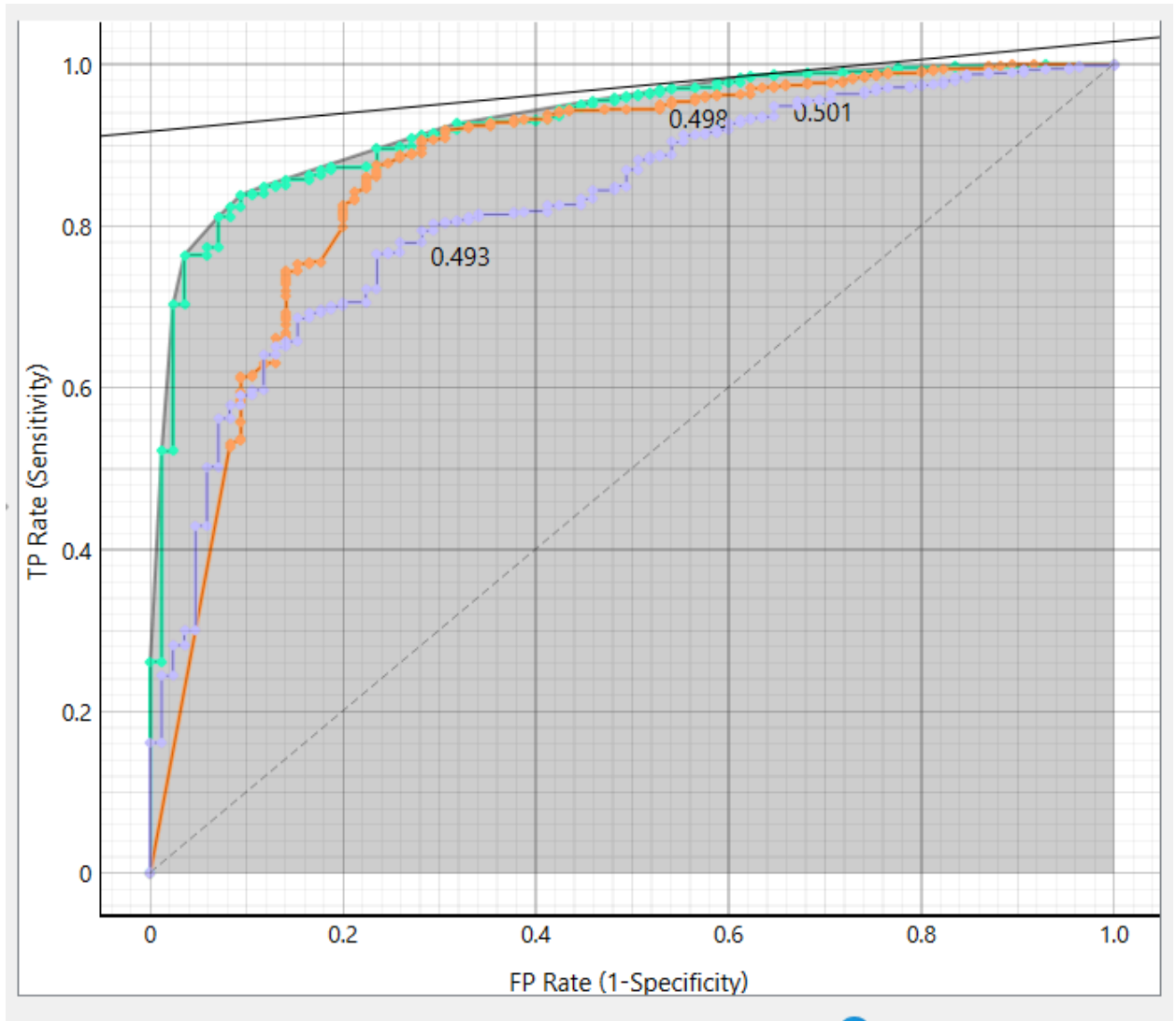
1. Muestren la tabla de métricas del Test and Score. ¿Qué modelo logró el mejor AUC y cuál el mejor F1-score?

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.925	0.915	0.910	0.908	0.915	0.498
Random Forest	0.867	0.908	0.896	0.892	0.908	0.398
Naive Bayes	0.825	0.784	0.820	0.891	0.784	0.353

- a.
 - i. El modelo que mejor logró el AUC fue el modelo de Logistic Regression y el mejor F1 Score también fue el de Logistic Regression
2. 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	Σ
Actual	no	712	26	738
	yes	44	41	85
Σ		756	67	823

- a.
 - i. TP = 41
 - ii. TN = 712
 - iii. FP = 26
 - iv. FN = 44
3. 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.



4.

EJERCICIO CLASE 9/10

1 2 3 4 Ejercicio Teórico: Análisis de Métricas de Clasificación (*Churn*)

Problema: Una empresa de telecomunicaciones desarrolló un modelo para predecir si un cliente **abandonará** (se dará de baja, *churn*) el servicio en el próximo mes. La acción de retención es contactar y ofrecer costosos incentivos.

Definición de Clases:

- **Positivo:** El cliente **abandona** (*Churn*).
- **Negativo:** El cliente **permanece** (*No Churn*).

El modelo generó la siguiente matriz de confusión en un conjunto de prueba de **5,000 clientes**:

	Predicción: Abandona (Positivo)	Predicción: Permanece (Negativo)
Realidad: Abandona	350 (TP)	150 (FN)
Realidad: Permanece	100 (FP)	4400 (TN)

Tareas

A. Cálculo de Métricas

Calculen las siguientes métricas, mostrando la fórmula y el resultado numérico exacto (o hasta 4 decimales):

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

B. Análisis de Impacto Estratégico (1 Punto)

1. Implicaciones de Errores:

- Describan la consecuencia económica o estratégica de un **Falso Positivo (FP)** en este contexto.

Una consecuencia económica de un falso positivo puede ser que al predecir que abandonaría pero en realidad permanece la empresa pudo haber presupuestado de menos y no lograr satisfacer el servicio de todos los clientes.

- Describan la consecuencia económica o estratégica de un **Falso Negativo (FN)** en este contexto.

La consecuencia económica que trae un Falso Negativo impacta de manera significativa a la compañía de telecomunicaciones, ya que, el modelo predice que el usuario permanecerá, pero termina abandonando. Esto provoca que la empresa no obtenga retorno de su inversión en adquisición y retención, además puede perder ingresos futuros. Y esto puede implicar costos adicionales para atraer nuevos usuarios o retener a los que se abandonaron.

2. **Prioridad de Métrica:** Si la dirección de la empresa establece que el costo de **gastar dinero en incentivos innecesarios** (es decir, en clientes que no iban a abandonar) debe ser minimizado a toda costa, ¿qué métrica de las calculadas en la sección A sería la más relevante para evaluar el modelo? **Justifiquen su elección.**

La más relevante para evaluar el modelo sería **precisión** ya que estaríamos evaluando los falsos positivos, es decir los que el modelo predijo que abandonan pero en realidad permanecen.

$$\text{Accuracy} = (VP + VN) / (VP + VN + FP + FN)$$

$$\text{Precision} = VP / (VP + FP)$$

$$\text{Recall} = VP / (VP + FN)$$

$$\text{F1-Score} = 2 \times ((\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}))$$