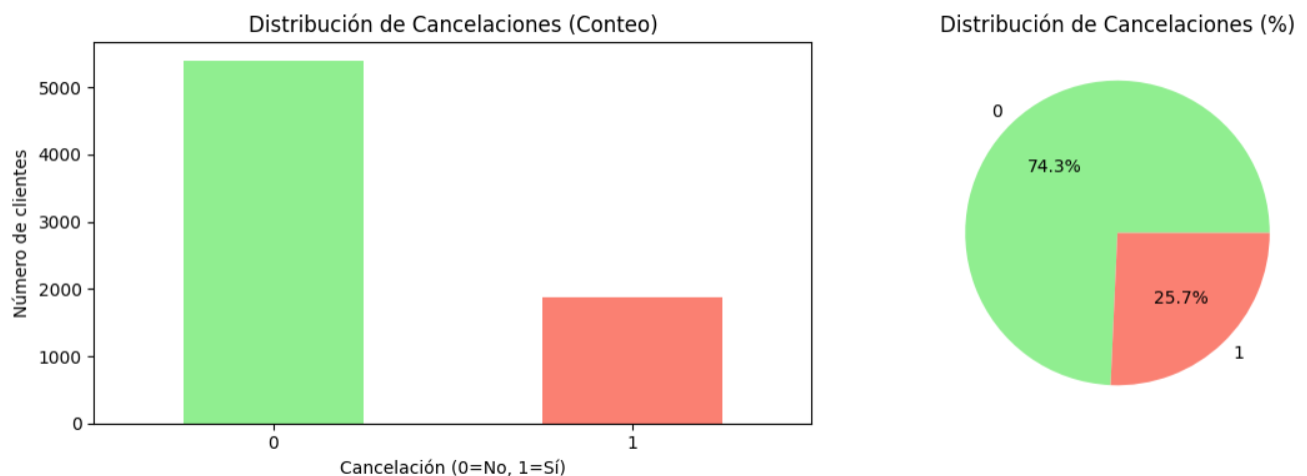


Análisis de Cancelación de Clientes en Telecomunicaciones

Introducción

La cancelación de clientes (churn) es uno de los principales desafíos en el sector de telecomunicaciones, ya que impacta directamente en la rentabilidad y sostenibilidad de las compañías. Con el objetivo de anticipar este comportamiento, se trabajó con un dataset de **7,267 clientes y 21 características**, que describe información demográfica, tipo de contrato, servicios contratados y variables de facturación.



Tras un proceso de **limpieza y preparación de los datos**, se desarrolló un análisis integral que incluyó:

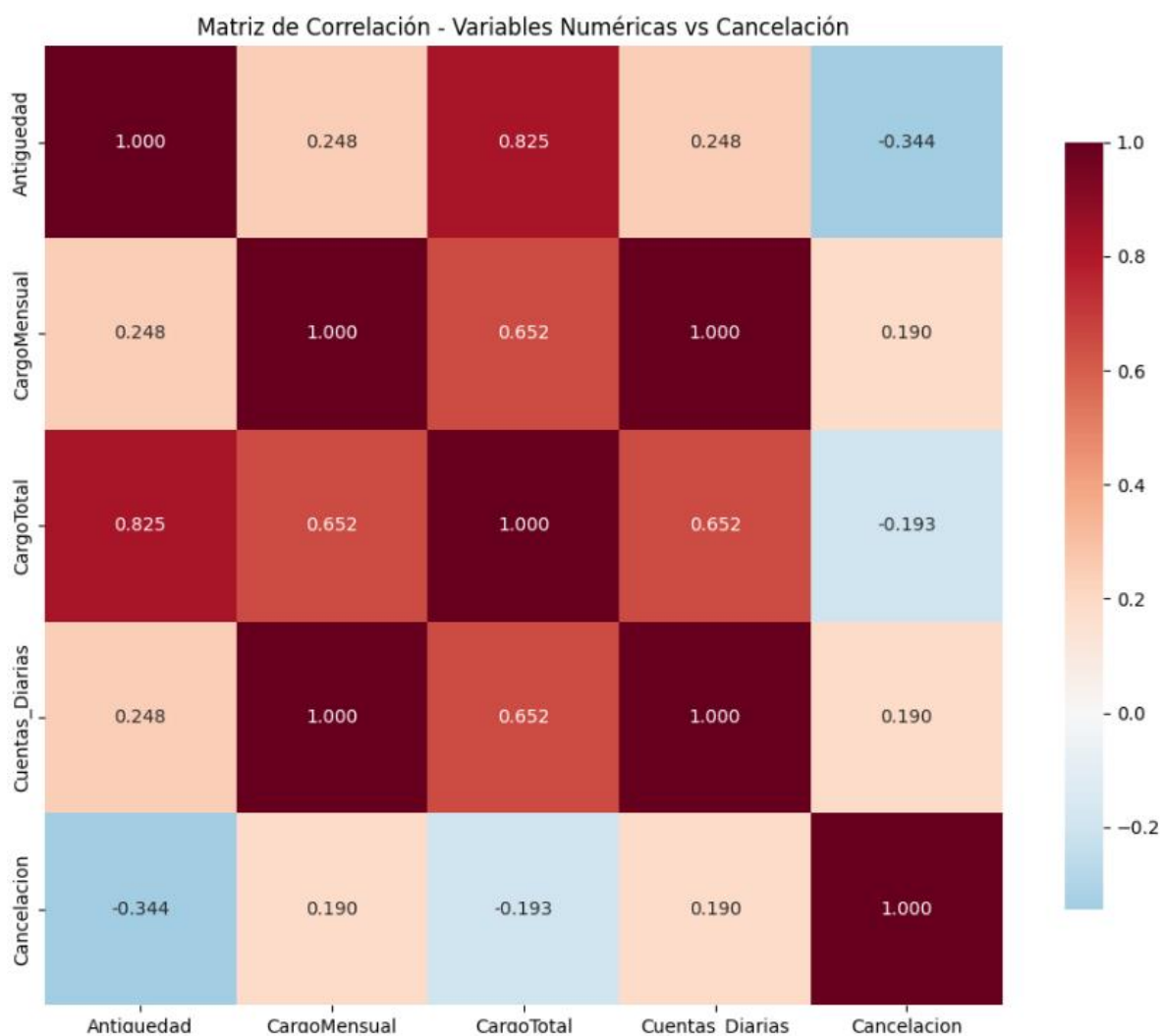
1. **Exploración de datos y análisis descriptivo**, para identificar patrones en variables categóricas y numéricas.
2. **Modelado predictivo** mediante diferentes algoritmos de machine learning.
3. **Comparación de desempeño** en métricas clave (Exactitud, Precisión, Recall, F1-Score, AUC-ROC).
4. **Interpretación de factores críticos de cancelación**, a través de modelos explicativos.

Este enfoque permitió comprender no solo **qué tan bien predicen los modelos la cancelación**, sino también **qué características influyen más en la decisión de los clientes**.

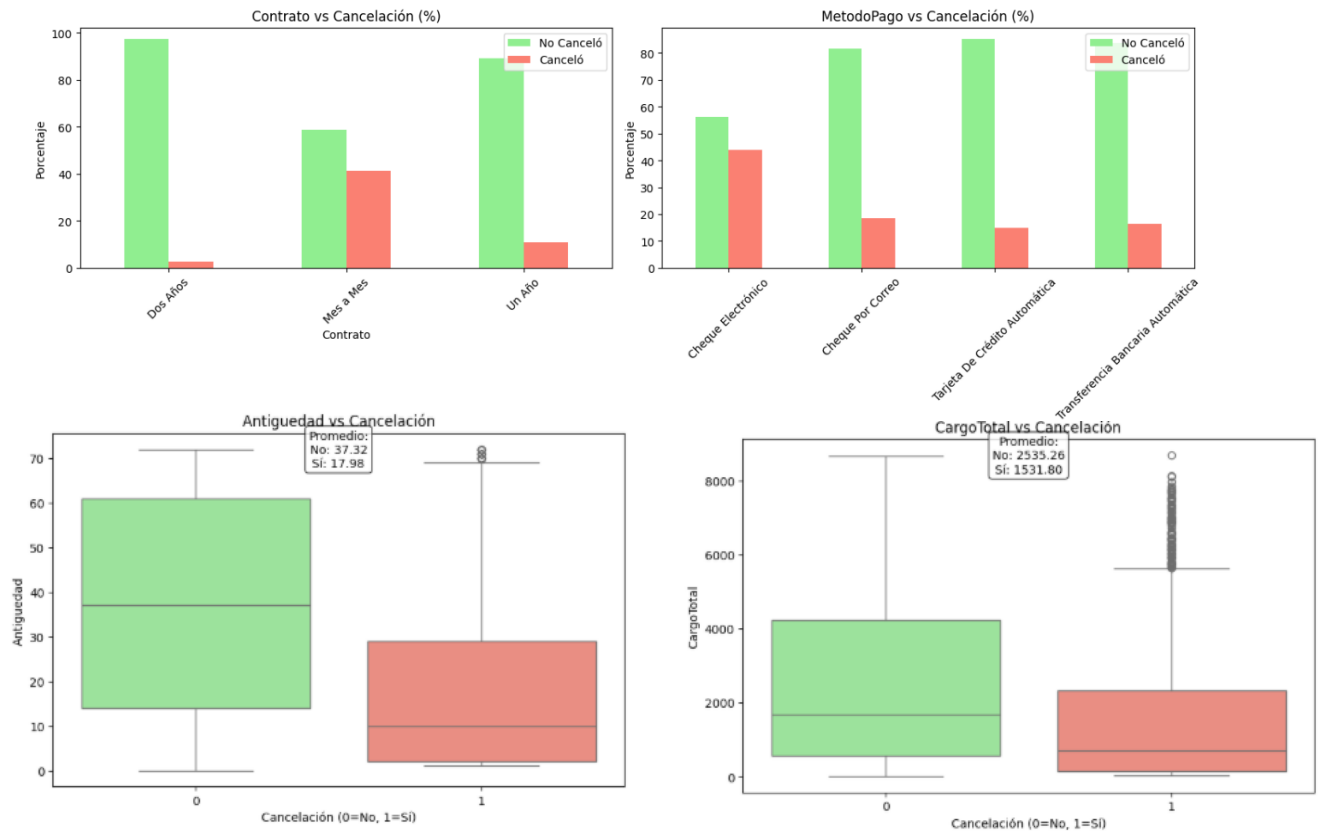
Desarrollo

1. Análisis Exploratorio

- La **tasa de cancelación fue del 25.7%**, evidenciando un dataset **desbalanceado**.



- Variables categóricas como **Contrato**, **Método de Pago** y **Servicio de Internet** mostraron fuerte relación con la cancelación.



2. Modelado Predictivo

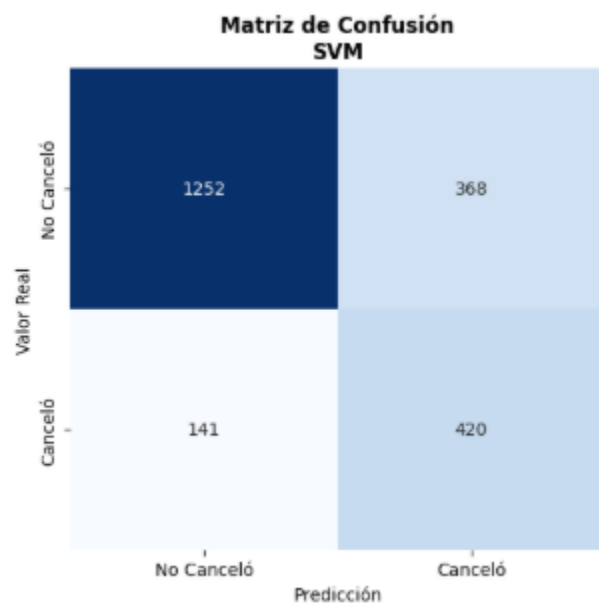
Se evaluaron **7 algoritmos de machine learning** con balanceo de clases vía **SMOTE**:

- **Regresión Logística**
- **Random Forest**
- **Gradient Boosting**
- **SVM (Support Vector Machine)**
- **K-Neighbors**
- **Decision Tree**
- **Naive Bayes**

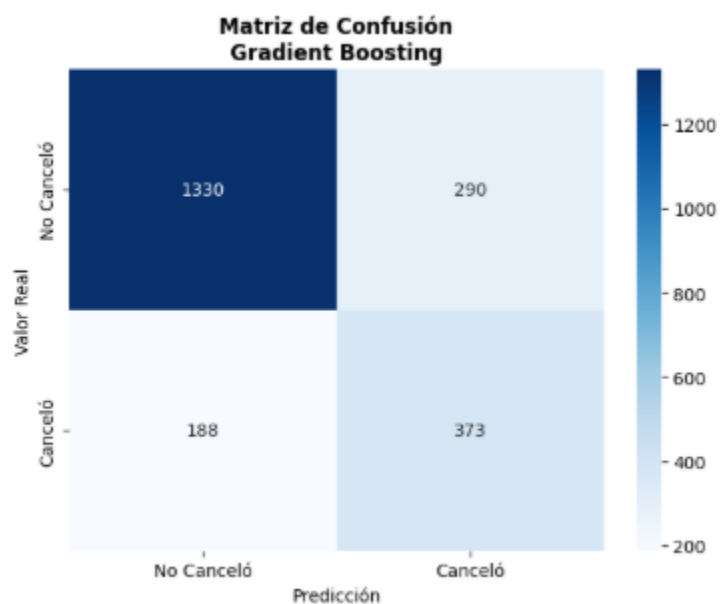
Cada modelo fue medido en **Exactitud**, **Precisión**, **Recall**, **F1-Score** y **AUC-ROC**, además de tiempos de entrenamiento y predicción.

3. Resultados Destacados

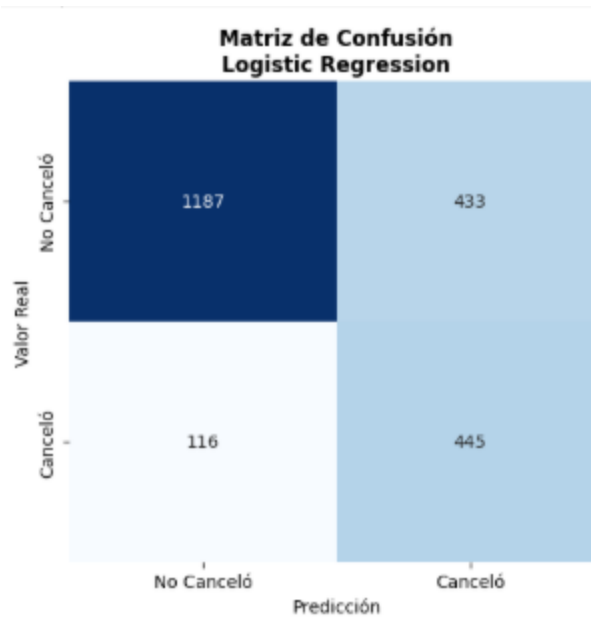
- **Mejor Modelo Global (F1-Score): SVM** con 0.6227.



- **Mejor Exactitud: Gradient Boosting (0.7808).**



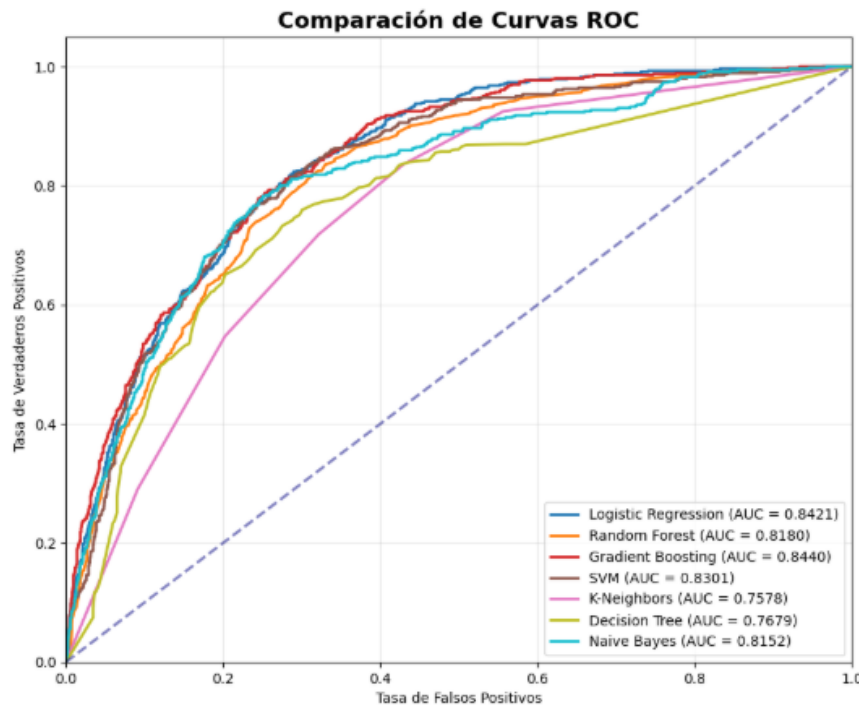
- **Mejor Precisión: Random Forest (0.5655).**



- **Mejor Recall: Naive Bayes (0.8342).**
- **Modelo más eficiente en tiempo:** Naive Bayes (entrenamiento y predicción más rápidos).

La **Regresión Logística con SMOTE** alcanzó un **AUC-ROC de 0.8421**, lo que confirma un buen poder discriminativo, además de permitir identificar **factores clave de cancelación**:

- **Antigüedad (coef. -1.457):** mayor permanencia reduce probabilidad de cancelar.
- **Contrato Mes a Mes (1.353):** clientes en este esquema presentan mayor riesgo.
- **Servicio de Internet Fibra Óptica (1.364):** asociado a mayor cancelación.
- **Cargo Total (0.826):** mayor gasto acumulado incrementa riesgo de baja.



Conclusión

El análisis realizado permitió **comprender el fenómeno de cancelación desde dos perspectivas:**

1. **Predictiva:** los modelos de machine learning, en particular **SVM y Gradient Boosting**, lograron un rendimiento sólido para anticipar cancelaciones, siendo útiles para sistemas de alerta temprana.
2. **Explicativa:** la Regresión Logística facilitó la interpretación de variables críticas, destacando la importancia de **tipo de contrato, antigüedad, cargo total y servicio de internet** como factores clave.

En términos estratégicos, se recomienda:

- Enfocar acciones de retención en clientes con **contratos mes a mes** y en los que utilizan **fibra óptica**.
- Desarrollar incentivos y programas de fidelización para **clientes con baja antigüedad**.
- Implementar **modelos híbridos (ensambles)** para mejorar precisión y recall simultáneamente.
- Mantener un **monitoreo continuo** del churn con técnicas de balanceo y validación cruzada.

Con ello, la compañía podrá no solo **anticipar cancelaciones con alta efectividad**, sino también **diseñar estrategias de retención personalizadas**, incrementando la satisfacción y reduciendo pérdidas por evasión de clientes.