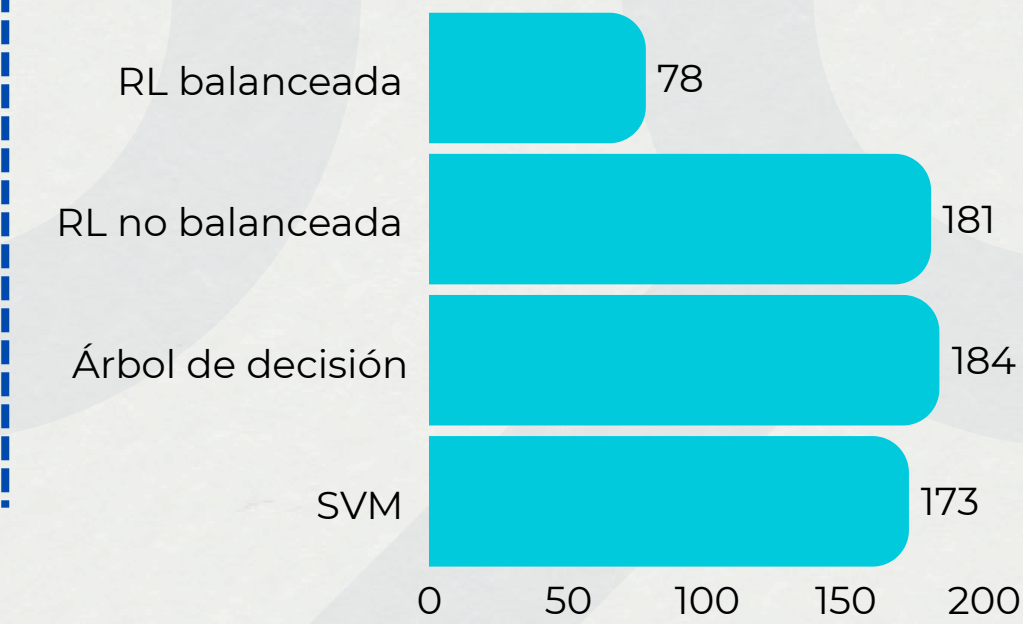


PREDICCIÓN DE DESERCIÓN EN CLIENTES TELCO



La predicción de deserción de clientes es un desafío crucial para las empresas, especialmente en sectores competitivos como el telecomunicaciones. Este estudio explora diferentes enfoques para predecir la deserción utilizando modelos de machine learning, con énfasis en la mejora del rendimiento mediante el balanceo de datos. Al enfrentar un conjunto de datos desbalanceado, se comparan tres modelos: regresión logística, árbol de decisión y SVM, con el objetivo de identificar la técnica más efectiva para predecir clientes en riesgo de abandono y optimizar las estrategias de retención.

DESEMPEÑO DE LOS MODELOS EN LA PREDICCIÓN DE DESERCIÓN DE CLIENTES



- **Regresión Logística:** Modelo estadístico que utiliza la función logística para clasificar probabilidades, eficaz con datos balanceados.
- **Árbol de Decisión:** Algoritmo de aprendizaje supervisado que se utiliza para tareas de clasificación. Tiene una estructura jerárquica de árbol
- **SVM:** Funciona correlacionando datos a un espacio de características de grandes dimensiones de forma que los puntos de datos se puedan categorizar para clasificar datos complejos.

REGRESIÓN LOGÍSTICA BALANCEADA:

La Mejor Solución para Identificar Clientes en Riesgo

El balanceo de datos fue crucial para la regresión logística. Sin balancear, el modelo dejó pasar 181 falsos negativos. Al balancear, se redujeron a 78, priorizando la identificación de clientes en riesgo.

Al balancear las clases, el modelo redujo los falsos negativos a 78, mejorando la identificación de clientes en riesgo. Aunque el accuracy bajó a 73%, esta mejora es clave para estrategias de retención efectivas.

Metricas	Regresión logística (con balanceo de datos)	Regresión logística (sin balanceo de datos)	Árbol de decisión	SVM (Support Vector Machine)
Accuracy	73%	79%	72%	80%
Recall de deserción	79%	52%	51%	54%
Recall de no deserción	71%	89%	79%	89%
Falsos negativos	78	181	184	173

Tabla comparativa de distintos métodos aplicados.

Implicaciones de los hallazgos:

Balancear los datos mejora la detección de clientes en riesgo de deserción, reduciendo falsos negativos y permitiendo una identificación más precisa. Las empresas deben enfocarse en detectar estos casos, más allá del accuracy.



Motivación detrás del estudio:

La deserción de clientes es un desafío clave para las empresas, especialmente en telecomunicaciones, donde la competencia es fuerte. Identificar con precisión a los clientes en riesgo permite tomar acciones proactivas de retención, mejorando la fidelidad y reduciendo los costos de adquisición.