

Proyecto 1 Modelos Predictivos de Fuga

Integrantes:

Antonio Jara
Juan Pablo Ureta
Nicolas Yañez
Rodrigo Espina



Objetivos del proyecto

Objetivo Principal: Aplicar un algoritmo de árboles de decisión, clasificador u otros para construir un modelo predictivo de fuga de clientes en la industria de las telecomunicaciones.

Objetivos específicos:

- Utilizar la base de datos entregada para obtener los registros de comportamiento de clientes de telefonía celular.
- Predecir la fuga utilizando una ventana de tiempo de 1 mes como marco temporal.
- Realización de análisis.
- Proponer acciones para retener clientes basadas en el análisis de predicción de fugas realizado.



Contexto del Proyecto

Antecedentes del Proyecto:

- Conjunto de datos consta de 6590 registros y 28 columnas.
- La variable dependiente clave es “Fugado”. Es la que se quiere predecir.

Marco Temporal: Se basa de una ventana de tiempo de 1 mes respecto a la predicción de fuga de clientes.

Descripción del contexto:

- Se enfoca en la industria de las telecomunicaciones.
- Se busca ayudar a la empresa a predecir la fuga de clientes y tomar medidas proactivas para retenerlos.



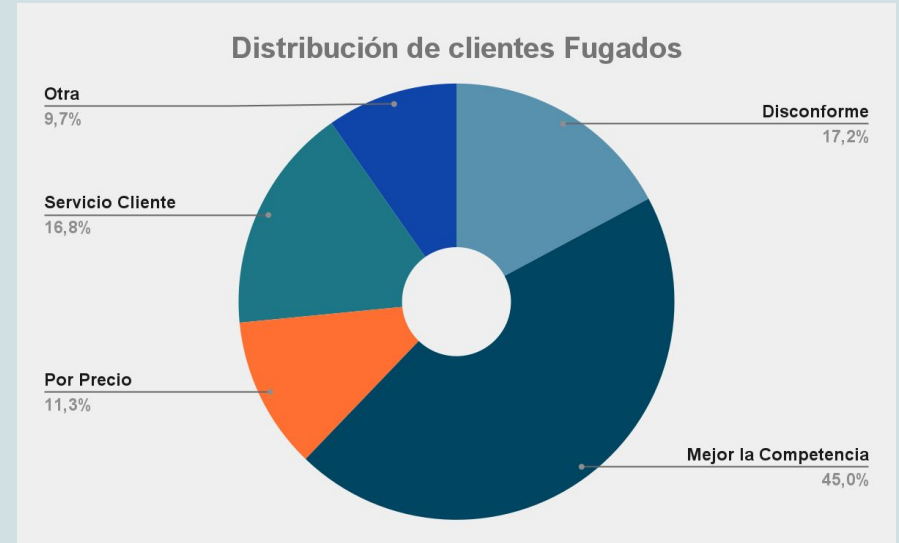
Características base de datos

- 12 columnas con Valores NA.
- Variable “Sexo” en igual proporcion para variable “Fugado” (Si y No).
- Variable “Causa Fuga” un valor NA para el 100% de los casos no fugados
- Variables relacionadas a servicios digitales (GB, Streaming, Nubes...etc) presentan valor NA cuando variable “Servicio Internet” es No. (9 columnas con valores NA)
- Variable “Histórico de Devoluciones” presenta un 92% de sus valores como 0.
- Variable “Histórico Cargos extra datos” presenta un 90% de sus valores como 0.
- Solo el 6% de los clientes con Plan Anual o más largo se fugó.
- Los clientes “Casados” recomiendan la empresa. Los “Soltero” no hacen recomendaciones

Análisis preliminar de causa de fuga

Se puede observar que las 4 razones más probable de fuga son:

- 45% se fuga debido a planes poco competitivos.
- 17% ser fuga por problemas técnicos.
- 16.8% por mala atención o soporte técnico.
- 11.3 % Costos de servicios.



Limpieza de base y balanceo de dataset



Al explorar dataset, se aprecian variables que presentan campos vacíos y que en primera instancia no aportarían al modelo a aplicar. Por lo mismo en la limpieza de la base se realizan las siguientes acciones:

- Se elimina columnas de **Id** y **Causa de Fuga**, esta última pues la variable que se busca predecir es si cliente abandona o no la compañía, sin considerar la causa, esto para la predicción.
- Luego para variable de **GB Mensuales Consumidos**, donde existen algunos casos vacíos, se determina dejar valores en 0, pues estos casos se da para clientes sin **Servicio de Internet**.
- Dada la misma causa anterior, las variables asociadas a **Servicio de Internet** con campos vacíos, se ingresa el valor de “No”.
- La variable **Cargo Mensual por Llamadas** se completa con valores 0, pues al visualizar columna de **Historico Cargo de Llamadas** esta presenta valores en 0 para los mismos casos.
- Variable **Múltiples Líneas**, determinamos una relación con las variables anteriores (**Cargo Mensual Llamadas** e **Historico Cargo Llamadas**) y al tener valores en 0, se asume que los clientes no debieran tener múltiples líneas, por lo que se completa con valor “No”.

Finalmente se realiza un balanceo de la base, pues variable a predecir Fugado, estaba sobrerrepresentada para los que “No” se fugan, pudiendo afectar la predicción final. Dado esto se genera un factor de balanceo, resultando un dataset para trabajar más equilibrado (45-55) en las clases de variable de interés.

Análisis de modelos y distribución de base

Una vez hecho la limpieza y balanceo de la base, aplicamos los diferentes modelos supervisados, con diferente distribución entre base de Entrenamiento y de Testeo, consiguiendo los siguientes resultados:

Utilizando todas las variables de la base

Proporción	Desempeño	RandomForest	NaiveBayes	Red Neuronal
70/30	Accuracy	0,8549	0,7561	0,7058
	Sensitivity	0,8283	0,8157	0,6741
	Specificity	0,877	0,7067	0,7899
80/20	Accuracy	0,8445	0,7655	0,70352
	Sensitivity	0,8077	0,8026	0,70769
	Specificity	0,8776	0,7321	0,69837
75/25	Accuracy	0,8463	0,7578	0,6712
	Sensitivity	0,8086	0,8266	0,644
	Specificity	0,875	0,7055	0,8045

Análisis de modelos y distribución de base



Utilizando variables significativas, según observación de modelo y árbol de decisión:

- A. TipoContrato
- B. RecomendacionesRealizadas
- C. MesesComoCliente
- D. CobroMensual
- E. NumeroDependientes
- F. Edad
- G. Casado

Proporción	Desempeño	RandomForest	NaiveBayes
70/30	Accuracy	0,8493	0,7577
	Sensitivity	0,8444	0,8587
	Specificity	0,8533	0,6741
80/20	Accuracy	0,8323	0,7473
	Sensitivity	0,8287	0,8624
	Specificity	0,8351	0,6595
75/25	Accuracy	0,856	0,7685
	Sensitivity	0,8491	0,8626
	Specificity	0,8613	0,6969



Acciones de retención

- Ofertas y Descuentos
- Mejoras en el Servicio al Cliente
- Programas de Lealtad
- Mejoras en la Calidad del Producto



Acciones de retención

Ofrecer a todos los clientes planes de internet:

- Filtrar filas con ServicioInternet igual a 'No'
- Reemplazar GBMensualesConsumidos por el promedio de los datos en casos sin uso
- Estrategia exitosa, disminución de fuga en un 13.5%

Ofrecer a todos los usuarios un servicio ilimitado:

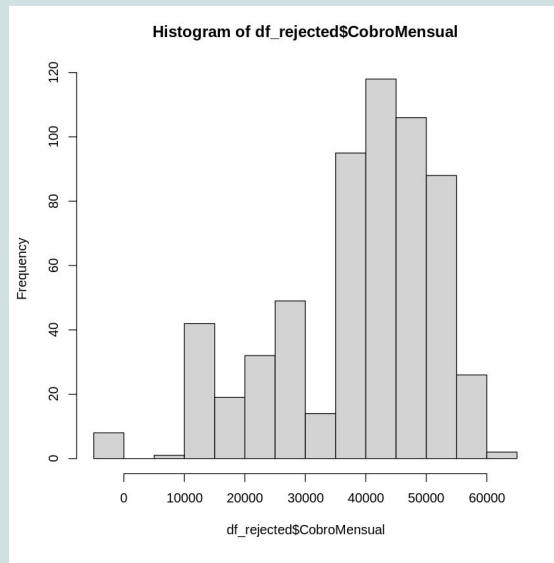
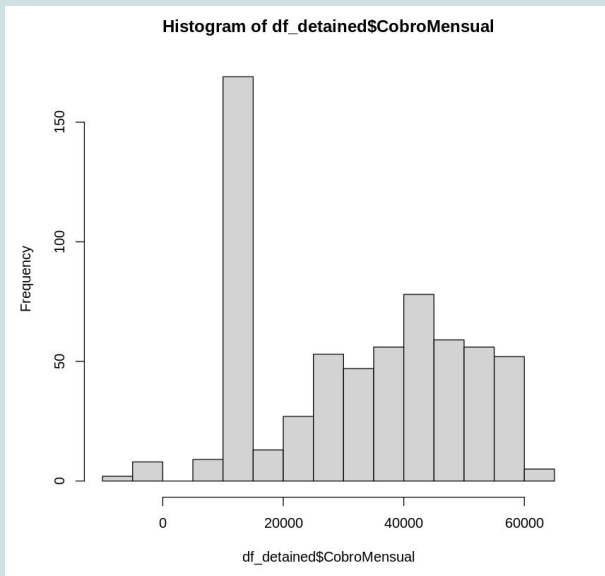
- PlanIlimitadoDatos se reemplaza por 'Si'
- Estrategia poco exitosa, fuga se mantiene en valores similares



Acciones de retención

Disminución de precios:

- Clientes fugados poseen precios altos en comparación con el resto





Acciones de retención

Disminución de precios:

- Clientes fugados poseen precios altos en comparación con el resto
- Disminuir precio para clientes con precios sobre \$30000
- Estrategia exitosa, disminución de fuga en 17%



Conclusiones

El modelo de clasificación que nos entregó mejores resultados es el RandomForest, considerando que presenta mejores valores de Accuracy, Sensitivity y Specificity.

Las acciones de retención que generan una disminución en la fuga de clientes son:

- Ofrecer internet a todos los clientes.
- Disminuir el cobro mensual.