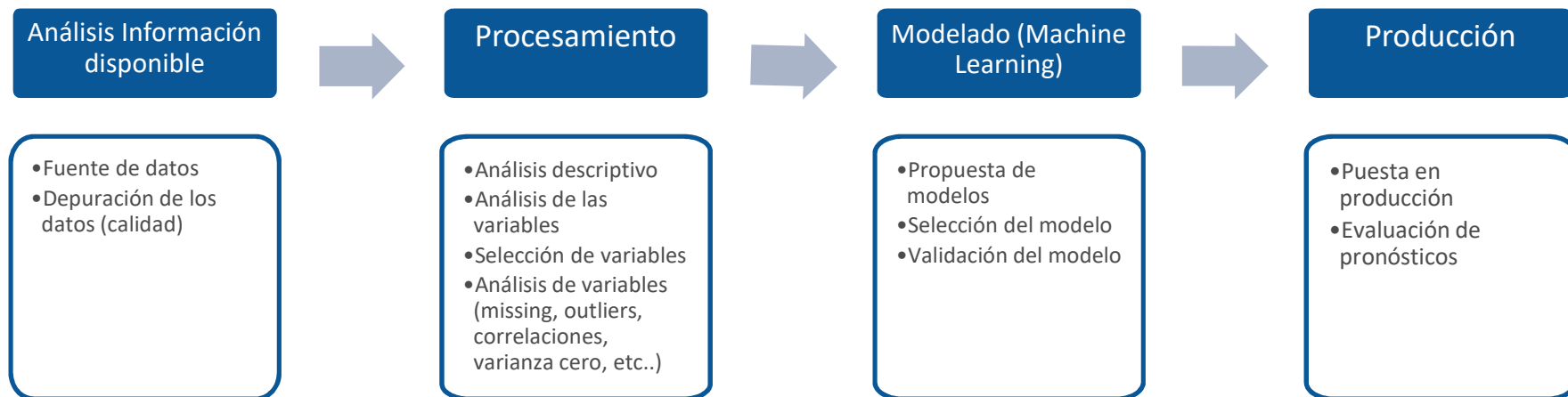


# Modelo de predicción para captación de clientes

Advanced Analytics  
Noviembre 2018

## Construcción de un modelo predictivo



Retail Financiero

## Antecedentes: Información disponible

### Demográficos

- Rut
- Sexo
- Edad
- Región
- Comuna
- Ciudad

### Comportamiento captación

- Historia en la base de Captación
  - Nuevo
  - 12 meses
  - 18 meses
- Historia de asignación a la campaña de captación últimos 12 meses

### Comportamiento Compra / Deuda

- SBIF
  - Deuda vigente
  - Deuda morosa
  - Deuda vencida
  - Deuda c.consumo
  - Deuda c.hipotecario
  - Monto líneas de c. disponible
  - Número de instituciones
- Fidelidad
  - Compras por retail

Estas variables cuentan con información de los últimos 13 meses

## Antecedentes: Creación nuevas variables

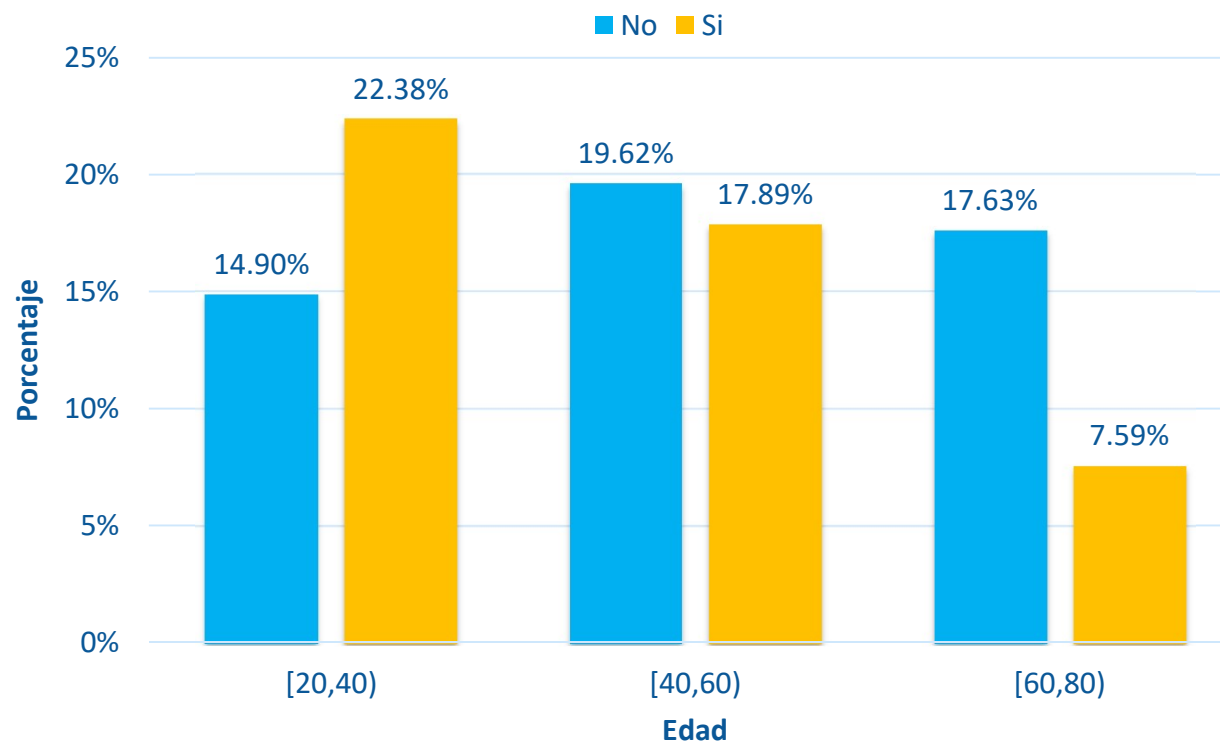
### Comportamiento SBIF

- Deuda directa vigente
- Deuda directa morosa
- Deuda directa vencida
- Deuda directa c.consumo
- Deuda directa c.hipotecario
- Monto líneas de c. disponible
- Compra retail últimos 13 m.
- Número de instituciones

### Nuevas variables

- Promedio deuda directa vigente
- Tiene deuda (SI/NO)
- Aumento deuda últimos 3 meses (SI/NO)
- Aumento deuda últimos 6 meses (SI/NO)
- Índice tend\_1 (deuda directa vigente / máx deuda directa vigente)
- Índice tend\_1 (deuda directa vigente / máx deuda directa vigente)
- Frecuencia compra retail últimos 6 meses
- Frecuencia compra retail últimos 13 meses
- Aumento N° instituciones últimos 6 meses

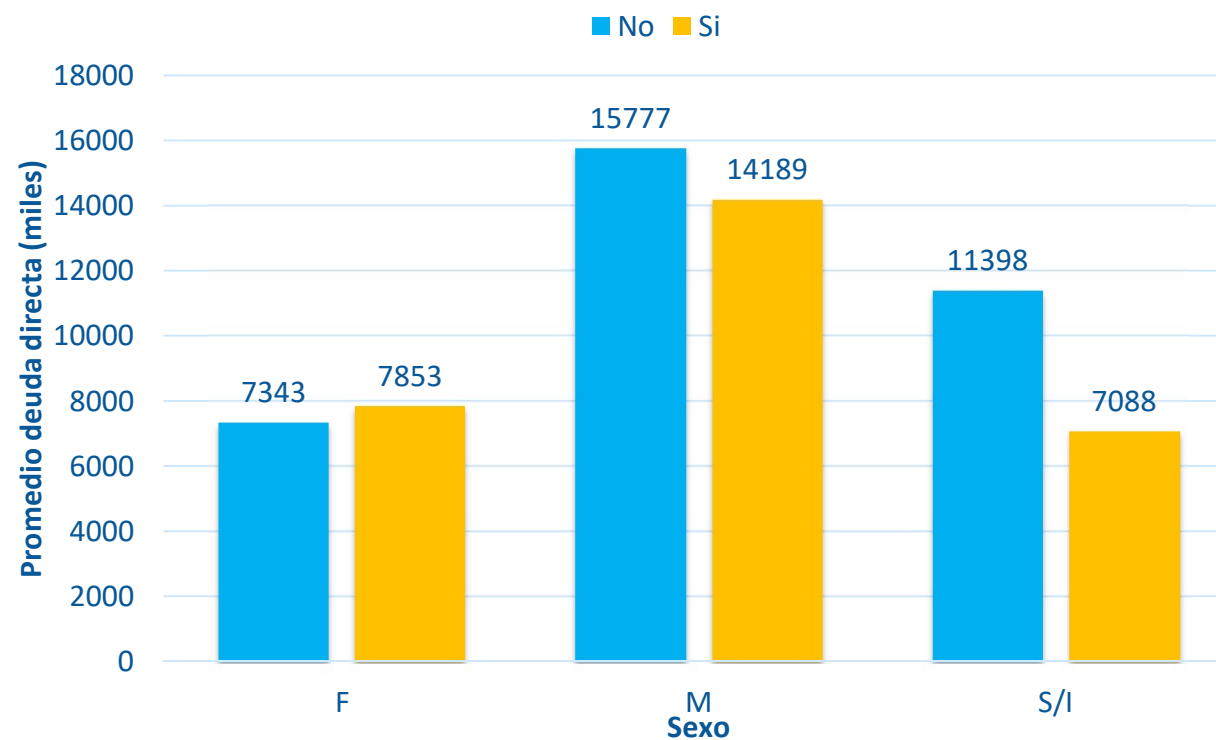
## Edad vs tiene/no tiene tarjeta



Retail Financiero



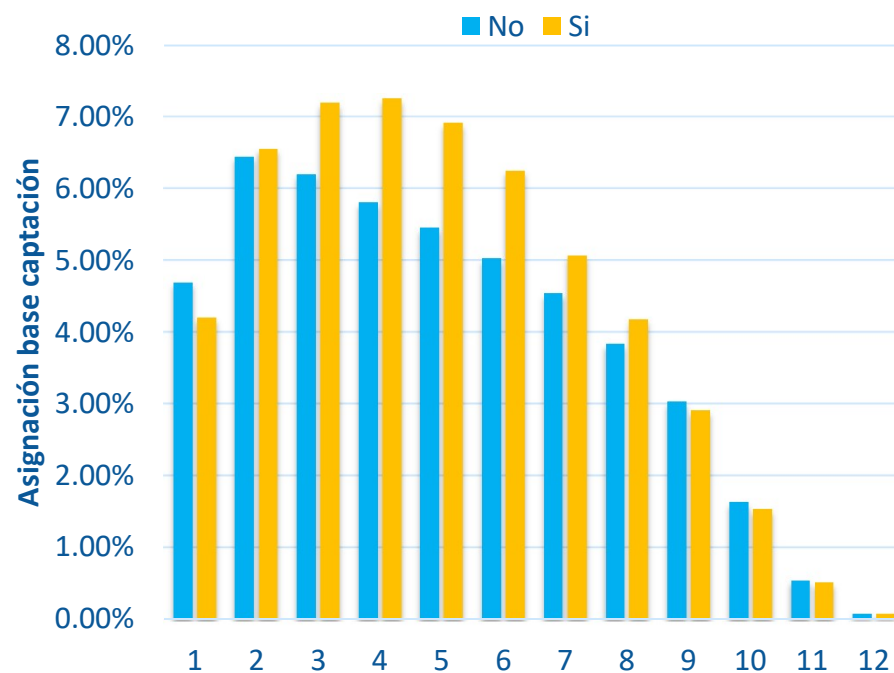
## Promedio deuda directa vigente vs tiene/no tiene tarjeta (SEXO)



Retail Financiero



## Asignación 12 meses vs tiene/no tiene tarjeta

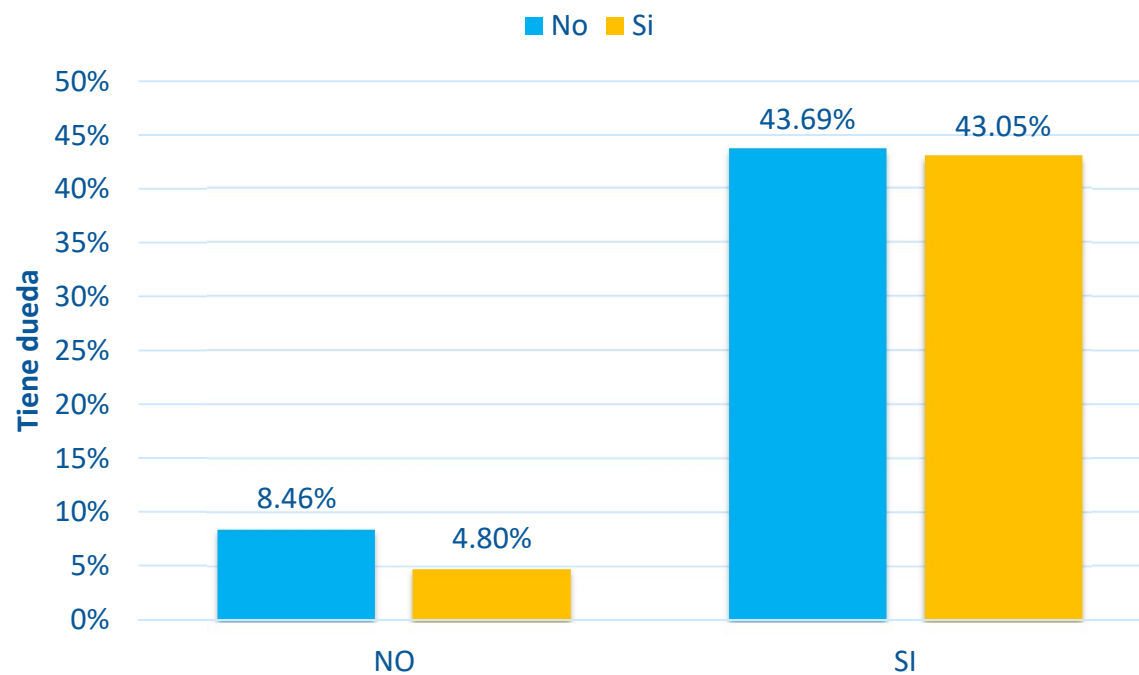


N veces	0	1	Total general
0	0.00%	0.00%	0.00%
1	4.70%	4.21%	8.90%
2	6.45%	6.56%	13.01%
3	6.20%	7.20%	13.40%
4	5.81%	7.26%	13.07%
5	5.46%	6.92%	12.37%
6	5.03%	6.26%	11.29%
7	4.55%	5.07%	9.62%
8	3.83%	4.19%	8.02%
9	3.03%	2.91%	5.94%
10	1.63%	1.54%	3.17%
11	0.54%	0.51%	1.05%
12	0.07%	0.07%	0.15%
	<b>47.31%</b>	<b>52.69%</b>	<b>100%</b>

Retail Financiero



## Tiene deuda últimos 13 meses vs tiene/ no tiene tarjeta



Retail Financiero





## Modelado de datos

- División conjunto de datos inicial en Datos entrenamiento (70%) y Datos de prueba (30%). El propósito de esto es:
  - Método para evaluación rápida del modelo predictivo
  - Comparar las predicciones en los datos de prueba permite calcular una medida de rendimiento para el modelo propuesto
- Propuesta de 8 modelos para predecir la probabilidad de que un cliente pueda ser captado y obtener tarjeta
- Metodologías evaluadas:
  - Regresión Logística
  - Análisis discriminante lineal
  - Árbol de Decisión
  - Random Forest
  - KNN (K-Nearest Neighbors)
  - Máquina vector soporte
  - GBM (Gradient Boosting Model)
  - Red Neuronal

Retail Financiero



## Métricas de evaluación

### Curva ROC

Una curva ROC es un gráfico que muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Esta curva representa dos parámetros:

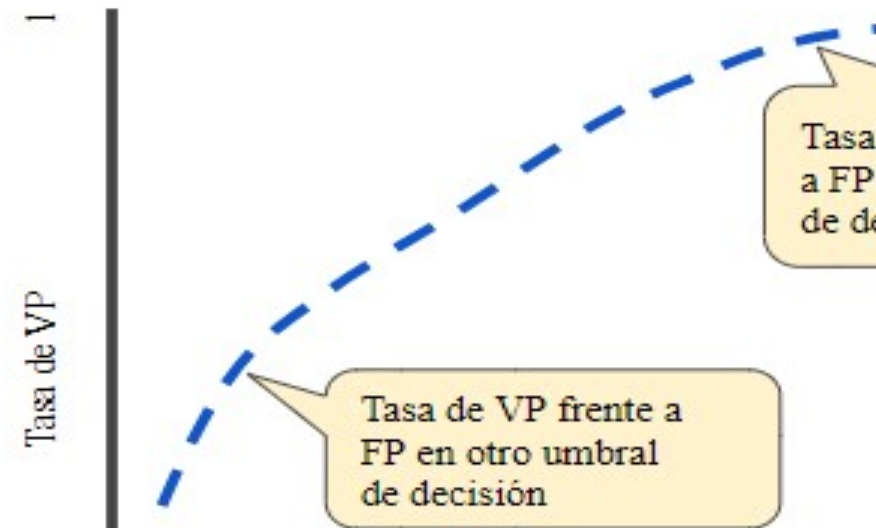
- Tasa de verdaderos positivos

$$TPR = \frac{VP}{VP+FN}$$

- Tasa de falsos positivos

$$FPR = \frac{FP}{FP+VN}$$

Una curva ROC representa TPR frente a FPR en diferentes umbrales de clasificación.



## Métricas de evaluación

- Para evaluar la predicción de los modelos, se construye una **matriz de confusión**, la que compara la predicción del modelo con la realidad

		Predicción	
		No Toma TSC	Toma TSC
Real	No Toma TSC	<b>VN</b> : Valor Negativo	<b>FP</b> : Falso Positivo
	Toma TSC	<b>FN</b> : Falso Negativo	<b>VP</b> : Valor Positivo

- A partir de esto, se construyen 4 ratios:

$$\text{Accuracy} = \frac{VP+VN}{VP+FP+FN+VN}$$

$$\text{Precision} = \frac{VP}{VP+FP}$$

$$\text{Recall} = \frac{VP}{VP+FN}$$

$$\text{fScore} = \frac{2(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

## RESULTADOS

M_propuestos	Accuracy	Recall	Precision	fScore	AUC
modelo1_LO	0.667	0.610	0.666	0.637	0.664
modelo2_LO	0.665	0.611	0.663	0.636	0.663
modelo3_LO	0.625	0.581	0.615	0.598	0.624
modelo4_LO	0.629	0.589	0.617	0.603	0.627
modelo5_LO	0.626	0.558	0.622	0.588	0.624
modelo6_LO	0.631	0.571	0.626	0.597	0.629
modelo7_LO	0.629	0.559	0.626	0.590	0.626
modelo8_LO	0.604	0.521	0.599	0.557	0.601
modelo9_LO	0.604	0.522	0.599	0.558	0.601
modelo10_LO	0.604	0.514	0.601	0.554	0.600
modelo1_AD	0.677	0.671	0.660	0.666	0.677
modelo2_AD	0.676	0.674	0.657	0.666	0.676
modelo3_AD	0.633	0.713	0.598	0.650	0.636
modelo4_AD	0.638	0.674	0.610	0.640	0.639
modelo5_AD	0.637	0.653	0.613	0.632	0.637
modelo6_AD	0.643	0.665	0.618	0.641	0.644
modelo7_AD	0.642	0.673	0.615	0.643	0.643
modelo8_AD	0.604	0.508	0.602	0.551	0.600
modelo9_AD	0.602	0.530	0.594	0.560	0.599
modelo10_AD	0.606	0.643	0.580	0.610	0.608
modelo1_RF	0.679	0.644	0.672	0.658	0.678
modelo2_RF	0.681	0.654	0.671	0.662	0.680
modelo3_RF	0.634	0.688	0.603	0.643	0.636
modelo4_RF	0.643	0.662	0.618	0.639	0.643
modelo5_RF	0.640	0.679	0.612	0.644	0.642
modelo6_RF	0.651	0.633	0.637	0.635	0.645
modelo7_RF	0.650	0.630	0.625	0.640	0.650
modelo8_RF	0.651	0.631	0.630	0.630	0.642
modelo9_RF	0.620	0.625	0.628	0.632	0.620
modelo10_RF	0.625	0.644	0.620	0.633	0.610

Retail Financiero



## RESULTADOS

### modelo4\_RF

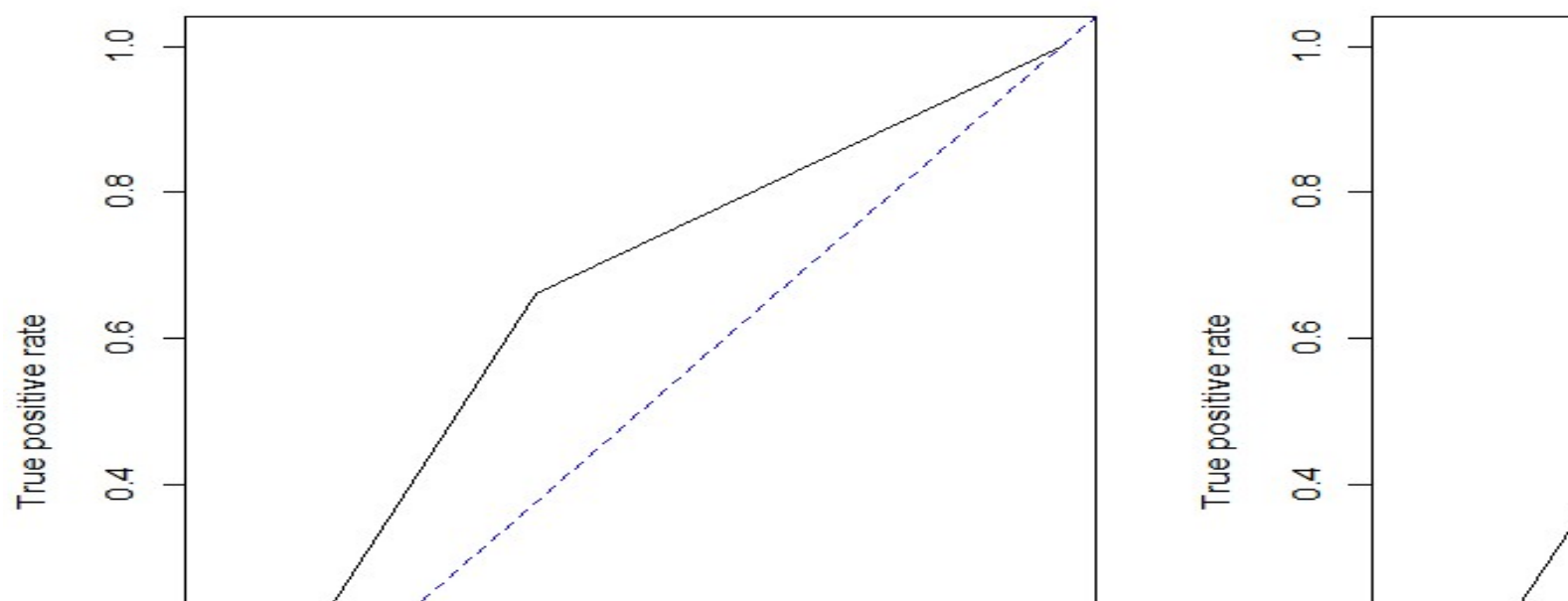
- Captación 18 meses
- Asignación 12 meses
- Frecuencia compra retail 6 meses
- Edad
- Tiene deuda
- Frecuencia compra retail 13 meses
- Aumento deuda 6 meses
- Aumento deuda 3 meses
- Sexo
- Captación nuevo
- Aumento N° instituciones 6 meses
- Promedio deuda directa vigente

### modelo5\_RF

- Captación 18 meses
- Asignación 12 meses
- Frecuencia compra retail 6 meses
- Edad
- Tiene deuda
- Frecuencia compra retail 13 meses
- Aumento deuda 6 meses
- Sexo
- Historia base captación nuevo
- Aumento N° instituciones 6 meses
- Promedio deuda directa vigente

## RESULTADOS

Curva ROC Modelo 4



M_propuestos	Accuracy	Recall	Precision	fScore	AUC
modelo4_RF	0.643	0.662	0.618	0.639	0.643
modelo5_RF	0.640	0.679	0.612	0.644	0.642

Retail Financiero

## Variables de importancia en el modelo elegido



Retail Financiero



## Resumen

- Se evaluaron 30 modelos con diferentes predictores (variables de entrada) bajo 4 metodologías distintas.
- Se determinó la probabilidad de que un cliente adquiriera tarjeta en base a la historia del comportamiento de estos en el retail.
- El modelo elegido corresponde a un Random Forest el cual presentó mejores métricas de evaluación (Accuracy, Recall, Precision, fScore y AUC)
- Las variables que explican de mejor forma si un cliente tiene o no tiene tarjeta son:
  - Edad
  - Promedio deuda directa vigente
  - Asignación en la base de captación últimos 12 meses



## Modelo de predicción para captación de clientes

Advanced Analytics  
Octubre 2018