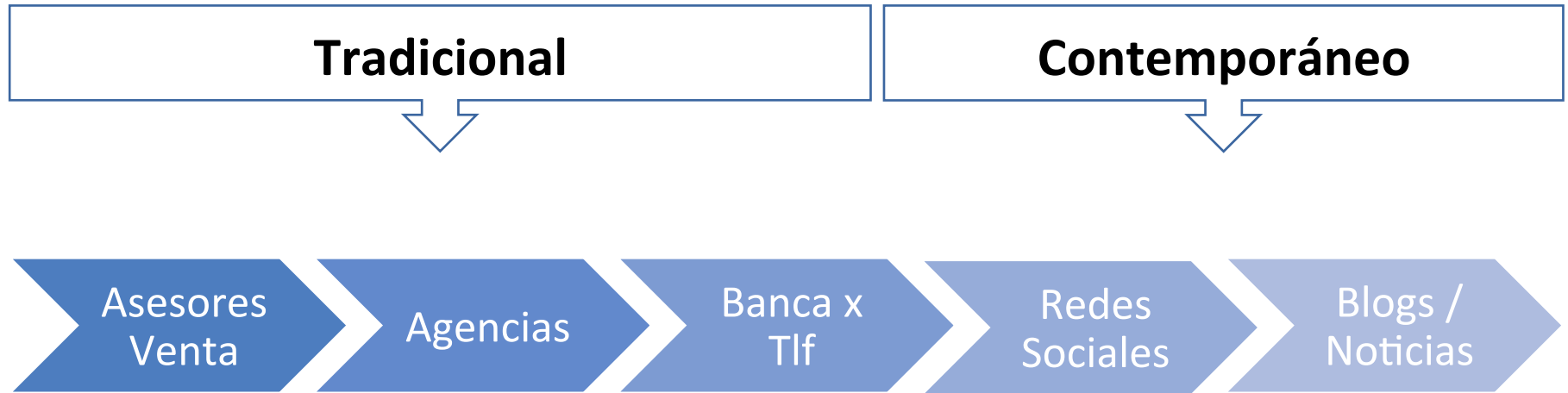


# **“Análisis de redes sociales para mejorar el rendimiento de tarjetas de crédito”**



**MSc. Leda Basombrio**

# Tarjetas de Crédito – Puntos de Contacto Cliente



***Los bancos saben utilizar esta información.***

***Los bancos todavía no aprovechan el máximo potencial de esta información.***

# Los clientes ya comparten sus ideas en redes sociales...



@zonkabe (991+) ★ 🇨🇱

Joseantonio Quiñones

@BCPComunica buenos días, hace 20 minutos que intento comunicarme via telefónica para cancelar mi tarjeta de crédito, a quien tengo q matar?



@sebastianosses (1,331+) ★ 🇨🇱 🇪🇨

Sebastián Ossés

@BCPComunica No entiendo como es posible que cobren, para permitir ver el estado de cuenta de la tarjeta de crédito en su app móvil.



@Neocriosc (14+) ★ 🇨🇱

CÉSAR RIOS

@interbank señores todo el día de ayer ha sido imposible banca por internet hoy no puedo acceder a Mi tarjeta de crédito.hasta cuándo?

# Pero, ¿los bancos están escuchando?



# Objetivos del Proyecto

Desarrollar modelos (POC) para evaluar el valor de la información de Twitter para la toma de decisiones del negocio de tarjetas de crédito.

- Fuga de clientes BCP
- Pérdida de saldos de clientes BCP
- Incremento de saldos de clientes BCP
- Adquisición de clientes BCP
- Fuga de clientes de la competencia



Amenaza

Oportunidad

# Fuentes de información

## Fuentes



Proveedor Quantico Trends

Palabras clave:



- Principales bancos (4)
- Productos, servicios y canales bancarios
- Estudios, vacaciones, viajes



SBS

Información de deudores publicada por SBS



- Saldos de productos
- Comportamiento de riesgo



BCP

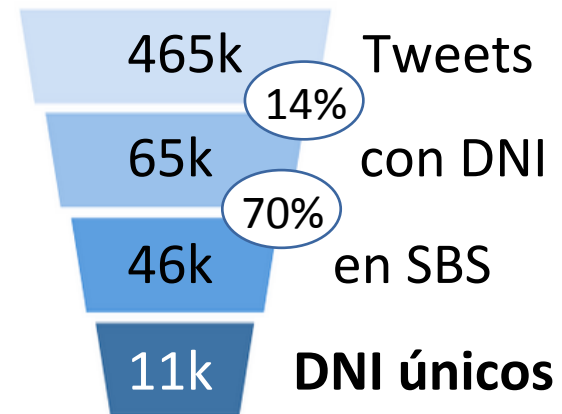
Muestra variables internas



- Del modelo actual de fuga

## Observaciones

Periodo: Ago13 - Oct16

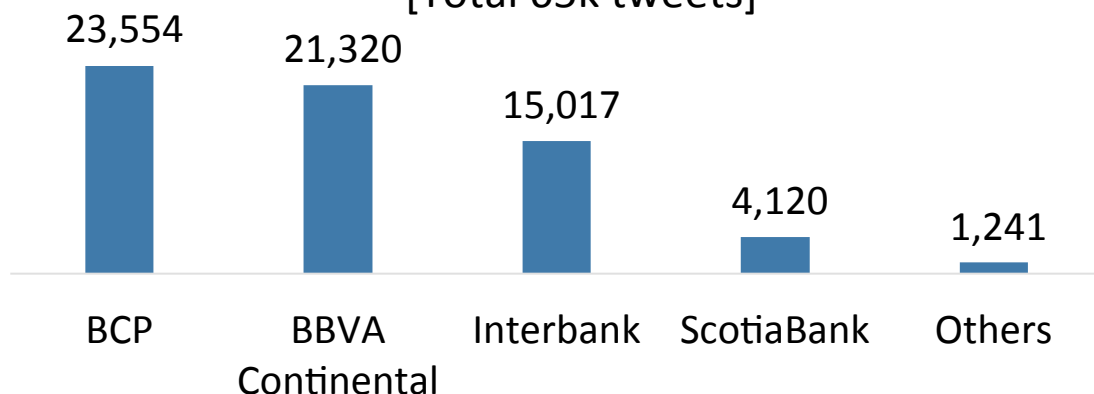


DNI's únicos por mes se agrupan para el análisis

# La mayoría de tweets corresponden a la mención de bancos

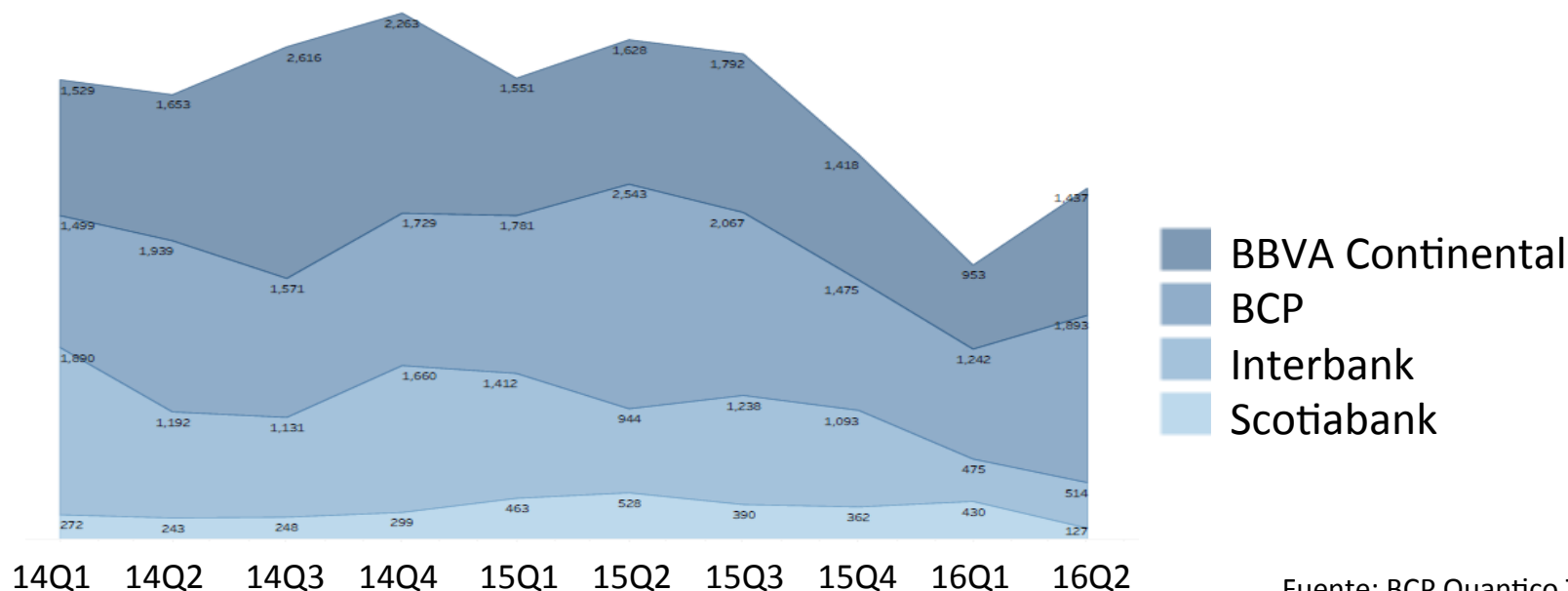
## Tweets por banco

[Total 65k tweets]



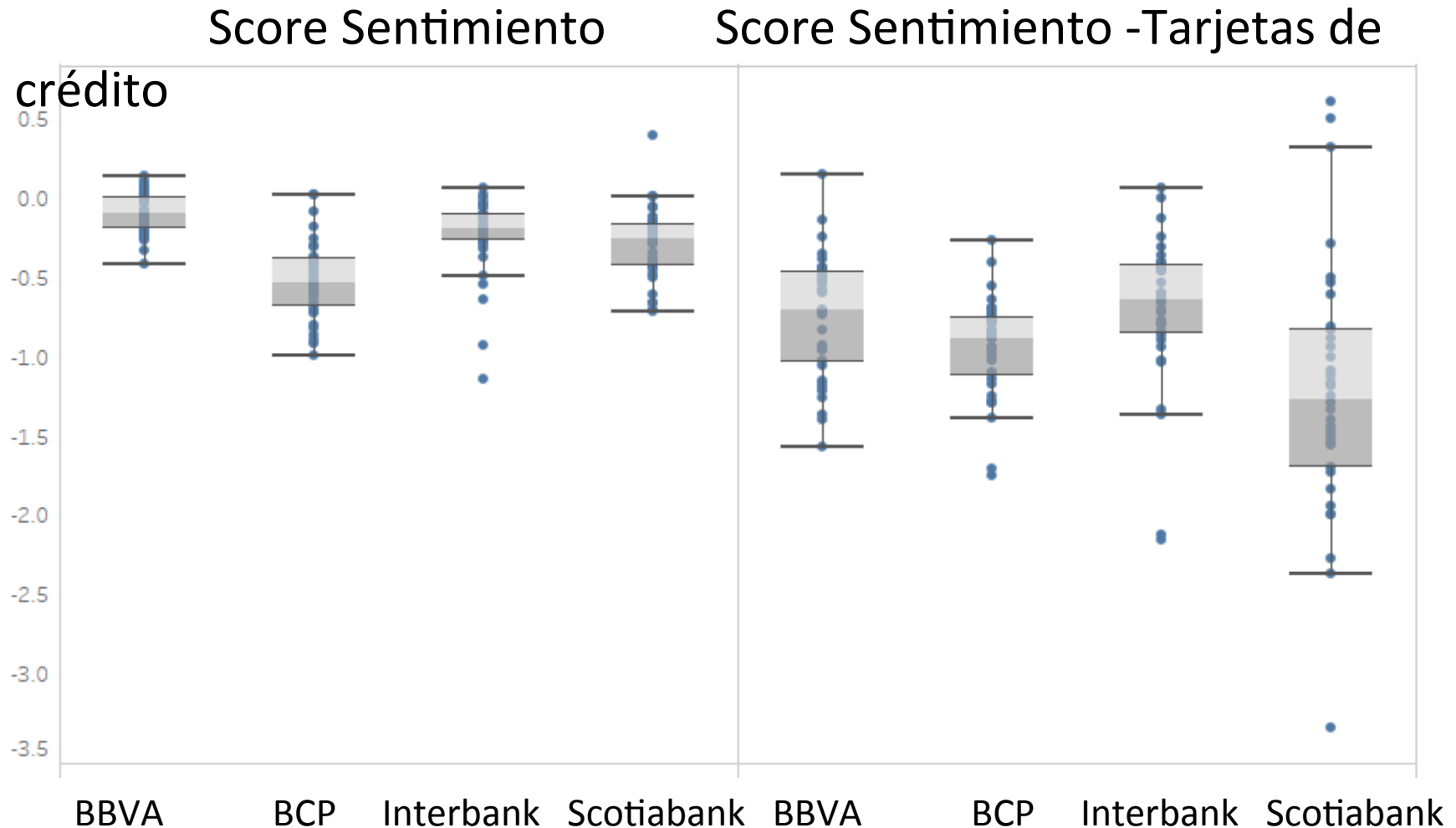
98% de los tweets mencionan a un banco y 36% al BCP.

## Tweets por banco y trimestre



# BCP tiene el puntaje de sentimiento general más bajo

Los tweets de tarjetas de créditos disminuyen el sentimiento.





# Matriz de Variables - Twitter

Agregación Tweets: { Periodo: 1, 3, 6 meses  
Cadena Texto: Solo Usuario / Historia Retweet / Tweet

## Generales

- Ctd de posts, caracteres, palabras, mayusc, emoticons, menciones por banco

422

## Términos

- Frecuencia Términos (TF)
- TF – Frec. Doc. Inversa (TF-IDF)
- Uni/Bi grams

240-1,200

## Modelo Sentimiento

- Datos de tweets con marca de sentimiento (TASS<sup>1</sup>)
- Target: Sent. Positivo
- Variables: TF / TF-IDF
- Método: SVM
- AUC (k-folds Prom): 94.5%
- Score modelo (flags)

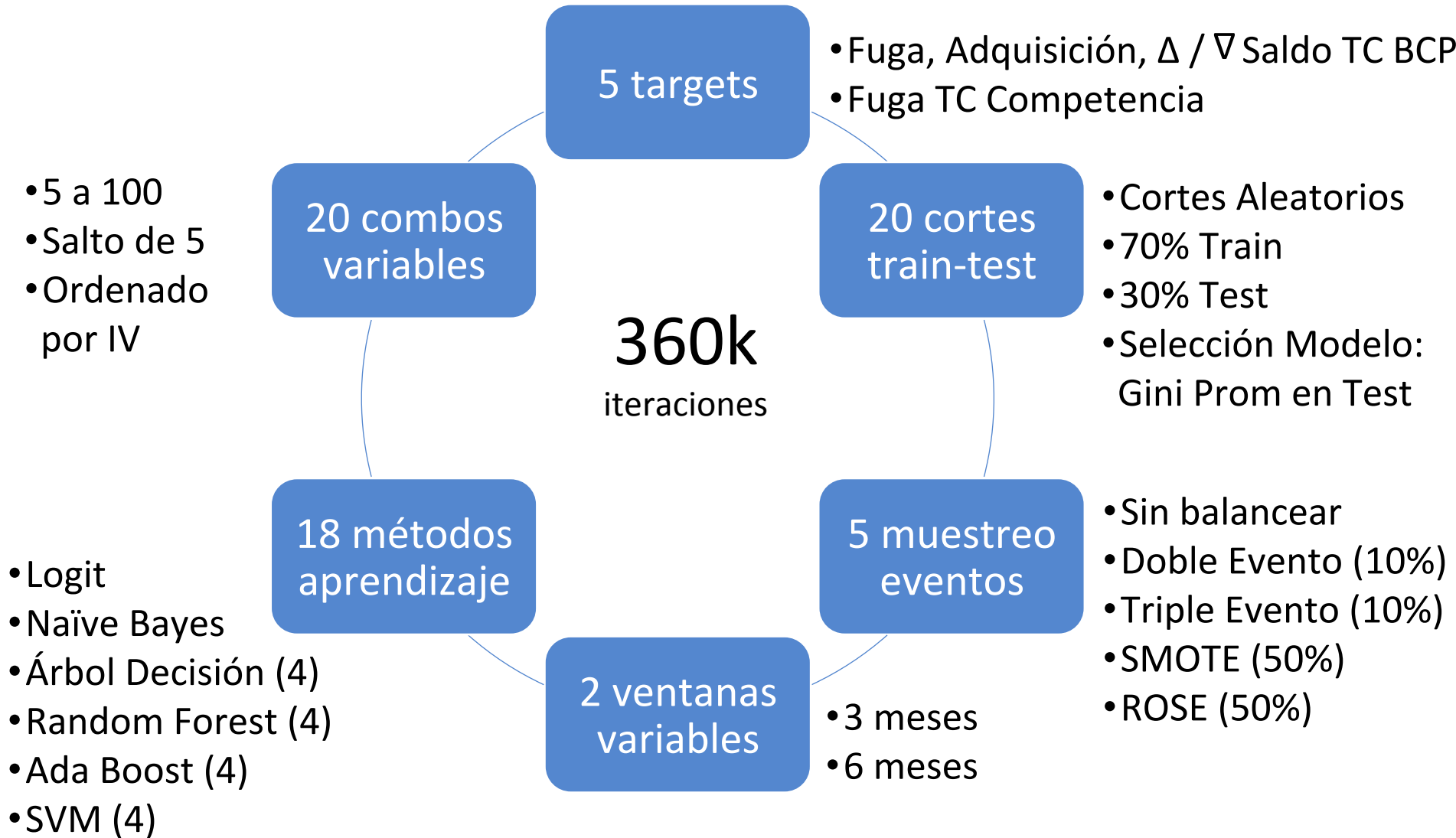
18

Limpieza Variables: Outliers capped al P99, No-informativas excluidas, Variable con mayor IV seleccionada si Corr Pearson > 0.8

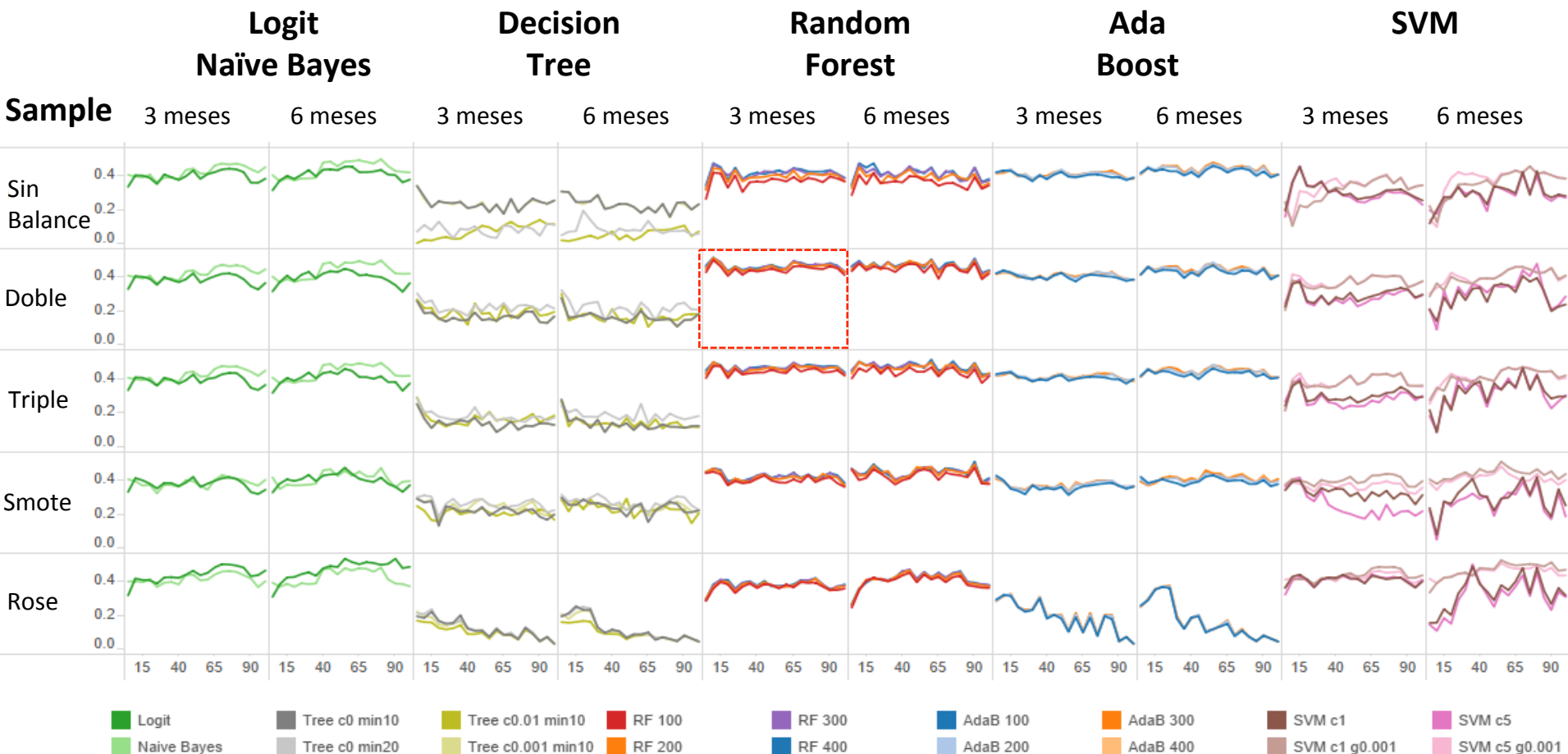
1. TASS es un Workshop de Análisis de Sentimiento en Español auspiciado anualmente por la Sociedad Española de Procesamiento Natural del Lenguaje (SEPLN).



# Marco Metodológico










# Modelo Fuga TC BCP – Coeficiente de Gini Prom

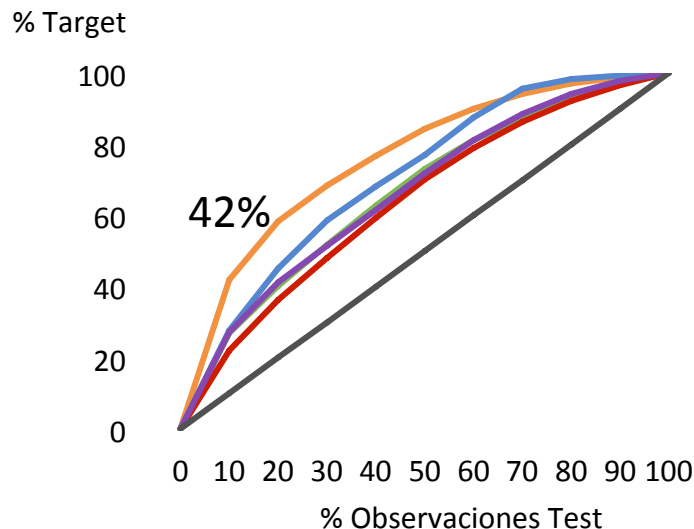


Random Forests obtiene los mejores resultados.

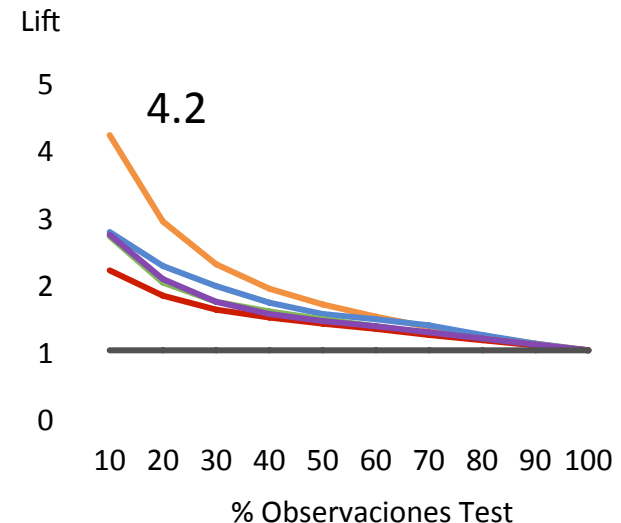
# La información de Twitter tiene mejores resultados para predecir Fuga de TC BCP

Modelo	Método	Muestreo	#Var-mes	Gini Test
 Fuga TC BCP	Random Forest	Doble	10-3m	 54.7%
 Adquisición TC BCP	Naive Bayes	Doble	52-6m	 43.0%
 Fuga TC Competen.	Random Forest	Triple	5-6m	 36.1%
 $\Delta$ Saldo TC BCP	SVM	Triple	61-6m	 35.7%
 $\nabla$ Saldo TC BCP	SVM	Doble	58-6m	 30.5%

Curva Respuesta Acumulada



Curva Lift

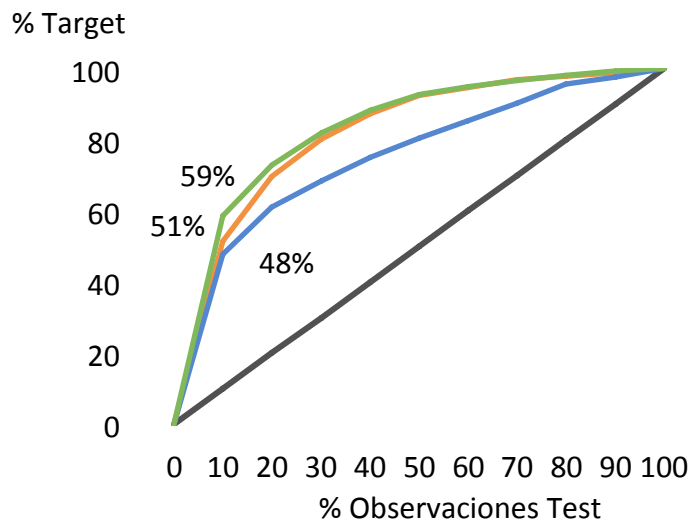


# Variables Twitter mejoran el modelo interno de Fuga TC BCP

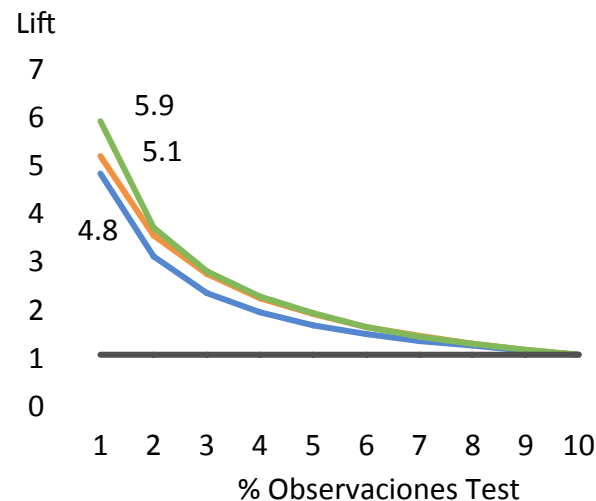
Modelo	#Vars	Gini Test
Variables internas	8	66.8%
Twitter	10	53.5%
Internas+Twitter	16	71.4%

El Gini del modelo interno aumenta en 4.6pp con variables Twitter.

















Curva Respuesta Acumulada



Curva Lift



Nota: 80% cruce entre bases (5.7k obs). Se usó Random Forest y muestreo de eventos Doble.

Variable		Internas+Twitter
	Score Behavior TC Max6	12%
	# Seguidores Prom3	11%
	# Amigos Prom3	10%
	Ingreso	10%
	% Trxs ATM 1m	8%
	# Trxs TC Prom6	7%
	# Trxs POS Tiendas Dpto Prom6	8%
	Palabras por Tweet Prom3	7%
	Total Palabras 3m	7%
	% Tweets réplica banco 3m	6%
	% Tweets horario oficina 3m	3%
	Flag Programa Fidelidad LATAM 1m	3%
	Meses desde último crédito	2%
	Flag Lima	2%
	Trxs Servicios Financieros Avg6	2%
	Flag Sentimiento Pos. Predicho 1m	3%

Twitter  
Internas

# Estrategia de Negocio

## Experimento de Retención

- Piloto de oferta de retención con un grupo de control, para medir el valor agregado de la acción e identificar el mejor corte para el modelo.
- Los beneficios de los clientes retenidos deben cubrir los costos de la acción.

Los resultados esperados son limitados por la baja identificación de DNI's en Twitter.

## Incrementar Cobertura Twitter



- Twitter ID y DNI: Puntos de contacto, onboarding, campañas



- Extender Tweet Scraping: Evaluar temas y usuarios

## Redes Sociales Adicionales

- Incluir otras redes sociales en el monitoreo (Facebook, Foros en línea, etc.)





# Conclusión

- La información de Twitter es valiosa para predecir el comportamiento de tarjetas de crédito, especialmente la fuga.



¡Los bancos  
deberían  
escuchar!

**Gracias...**

# **Estructura del Código**

# Text Mining

## Data Twitter

- Tweets colapsados por usuario



## Limpieza Texto

- Remover puntuación, números, pág. web
- Homogeneizar palabras
- Eliminar stopwords
- Stemming (truncar palabras a su raíz)



## Extracción Variables

- Creación de matriz de términos
- Uni/bi gram
- Frecuencia Términos (absoluta/ponderada)



**Otras variables  
continuas + Target**



# Limpieza de variables



## Limpieza univariada

- Outliers
- Missing values
- Vars no informativas (0s, constantes)



## Limpieza Multivariada

- Matriz de correlaciones
- Selección de mejor entre pares correlacionados según Information Value



**Tablón Final**



# Simulación de modelos

