



BCI CHALLENGE

Presentado por:



William Berrios

Data Scientist (3 años)

Researcher Computer Vision (1 año)

Ganador competencias DS - LATAM





Agenda

Parte 1: Problemática



Parte 2: Oportunidad del Reto



Parte 3: Solución Analítica



Parte 4: Impacto en el Negocio



Parte 5: Conclusiones Y Trabajo Futuro

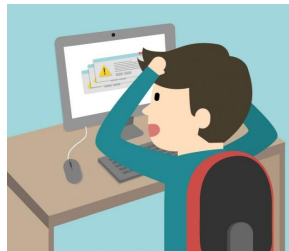


“Actualmente el banco BCI presenta problemas para la obtención de información de ventas de los clientes del segmento empresa”

Causas:



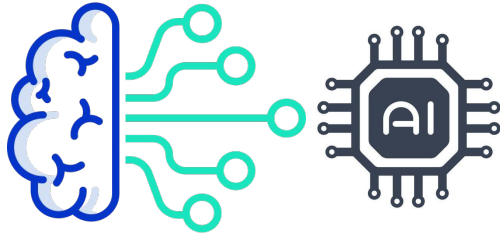
Proceso costoso y manual



Información parcial o discontinuada



Entrega de información de ventas no recurrente



Generación de un modelo(s) de machine learning utilizando variables anonimizadas:

- Data transaccional
- Características del cliente



Estimar el récord de Ventas
(Ventas actuales y pasadas)



Toma de Decisiones Crediticias



Oportunidades de Negocio

NUESTRA SOLUCIÓN





A. Tratamiento
de los Datos

B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Limpieza de Datos

- Gran cantidad de nulos (>98%) !!!

Transformación de los Datos

- Imputación de valores nulos
Imputación con 0 (mejores resultados)



A. Tratamiento de los Datos

B. Esquema de Validación

C. Generación de Variables

D. Modelado

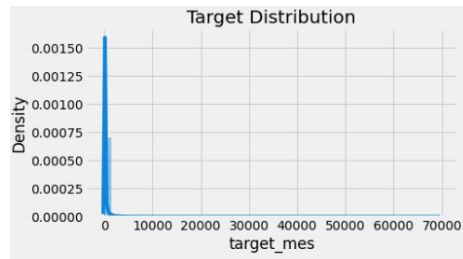
Limpieza de Datos

Análisis Exploratorio

- Gran cantidad de nulos (>98%) !!!

Transformación de los Datos

- Imputación de valores nulos
Imputación con 0 (mejores resultados)



“Transformación a la Variable Objetivo”



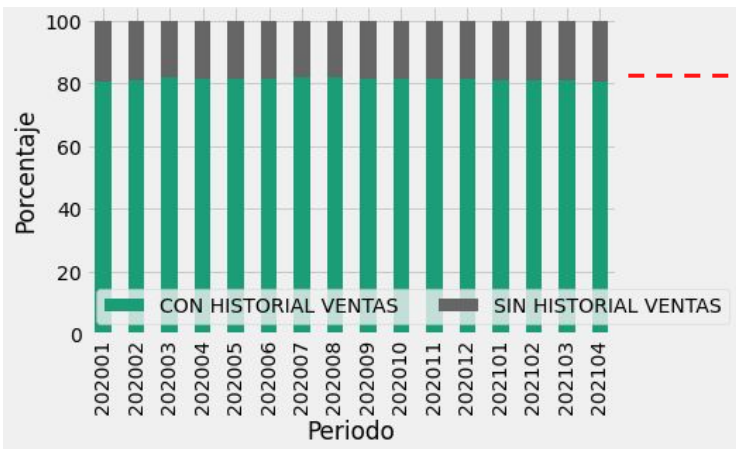
A. Tratamiento
de los Datos

B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Distribución Población:



Modelo 1: Clientes sin historial de ventas

Modelo 2: Clientes con historial de ventas

- Permite calcular con mayor exactitud las ventas de clientes empresa
- ¡¡ Marcó la diferencia en la competencia!!



A. Tratamiento
de los Datos

B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Operalización (Ejm)



Periodo: 202005

- Variables
- Cliente

¿Posee historial de
Ventas con nosotros?

Si

**M2: Clientes con historial
de ventas (80%)**

No

**M1: Clientes sin historial
de ventas (20%)**



A. Tratamiento
de los Datos

B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Operalización (Ejm:)



¿Posee historial de
Ventas con nosotros?

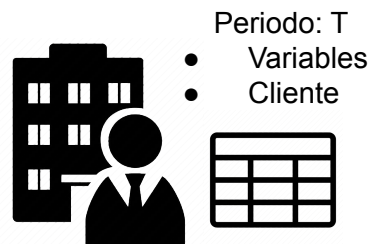
Si

**M2: Clientes con historial
de ventas**

No

**M2: Clientes sin historial
de ventas**

Ventas en 12M



Ventas en
el periodo T

$T \in [202004-201904]$



A. Tratamiento
de los Datos

B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Variables Categóricas

- Generación de Segmentos:
 - tipo_banca + tipo_segmento + tipo_cliente
 - tipo_segmento + categoria



A. Tratamiento de los Datos

B. Esquema de Validación

C. Generación de Variables

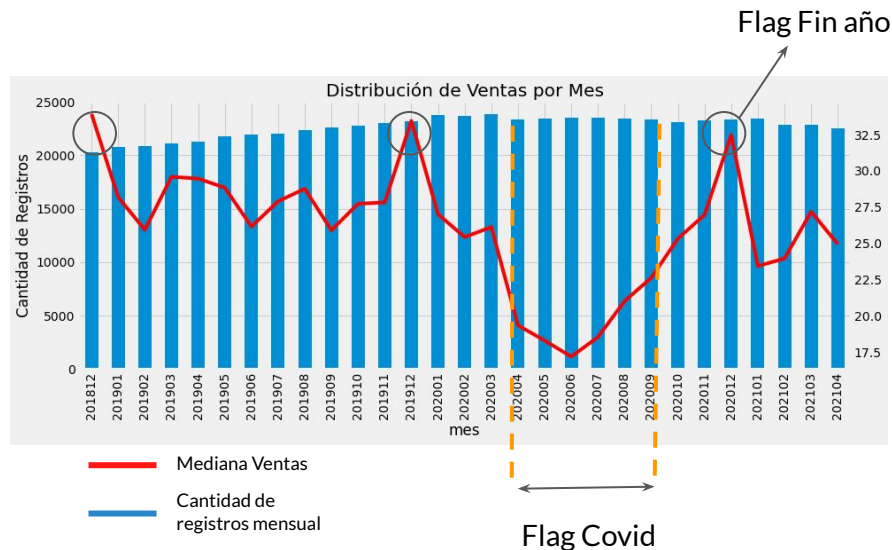
D. Modelado

Variables Categóricas

- Generación de Segmentos:
 - tipo_banca + tipo_segmento + tipo_cliente
 - tipo_segmento + categoria

Variables Numéricas

- Tendencia y Agregaciones de la variable ventas
- Flag Covid : 202004 - 202009
- Flag Fin año





A. Tratamiento
de los Datos

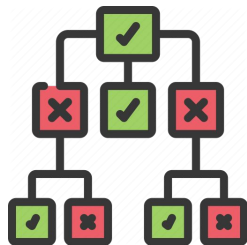
B. Esquema de
Validación

C. Generación
de Variables

D. Modelado

Algoritmos:

- Random Forest
- Redes Neuronales
- **LigthGBM**
- **Catboost**



Ensamblaje

**M1: Sin historial de ventas
(MAE 98.5)**

**M2: Con historial de ventas
(MAE 53.7)**





A. Tratamiento de los Datos

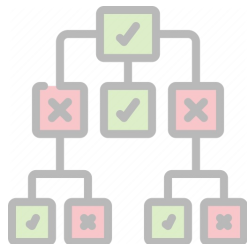
B. Esquema de Validación

C. Generación de Variables

D. Modelado

Algoritmos:

- Random Forest
- Redes Neuronales
- LigthGBM
- Catboost



M1: Sin historial de ventas
(MAE 98.5)

M2: Con historial de ventas
(MAE 53.7)



1er Lugar



Test Público

MAE: 77.5

Test Privado

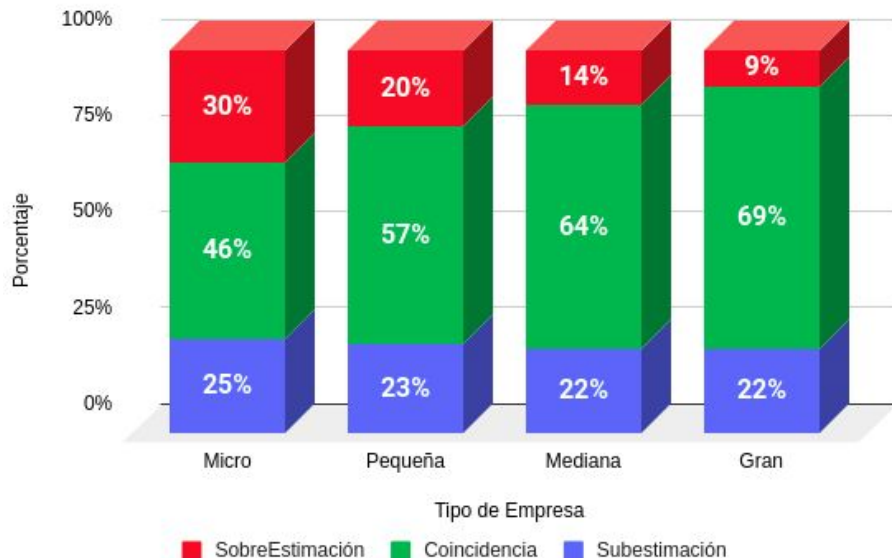
MAE: 78.7

- MAE: Error Absoluto Medio

A menor error, la solución es superior!!!



Efectividad del Modelo por Rango de Ventas



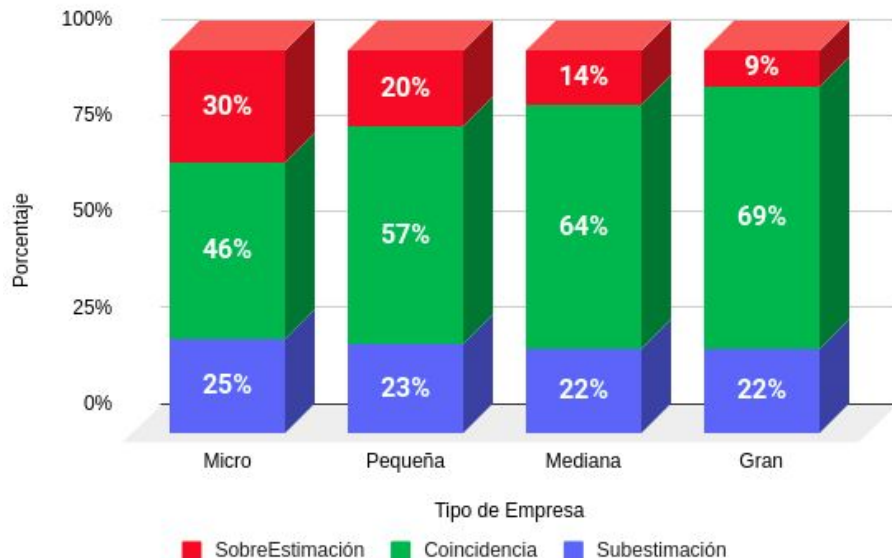
- La coincidencia supera el 45%
- La sobre estimación **no** supera el 30%
- El segmento “**Gran empresa**” se predice con mayor exactitud

¡ Disminución del Riesgo de Crédito!

- Rangos Sobre-Sub: (+/- 30%)
- Periodo de análisis (12M : 202105-202004)
- [Definición de tipo empresa](#)



Efectividad del Modelo por Rango de Ventas



- Rangos Sobre-Sub: (+/- 30%)
- Periodo de análisis (12M : 202105-202004)
- [Definición de tipo empresa](#)

- La coincidencia supera el 45%
- La sobre estimación no supera el 30%
- El segmento “Gran empresa” se predice con mayor exactitud

¡ Disminución del Riesgo de Crédito!

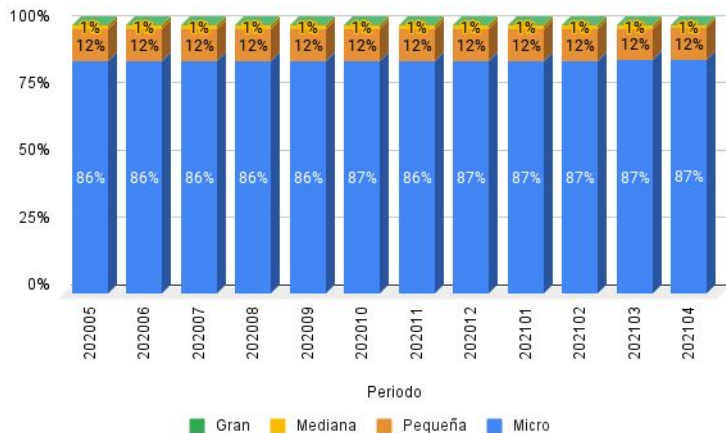
Oportunidad

- Definir de forma efectiva la Línea de Crédito
- Utilizar las ventas como input para modelos:
 - Fuga de Clientes
 - Propensión
 - Cross - Selling



Distribución Portafolio (12M)

Tipo Empresa	Venta Promedio Anual (uF)	Distribución
Micro	452	86.5%
Pequeña	6913	12%
Mediana	44251	1.3%
Gran	202335	0.2%



Estrategias de Negocio

- Establecer una estrategia diferenciada por cada segmento.
 - Micro y Pequeña: Promover asesorías
 - Mediana y Gran: Brindar mejores tasas
- Definir el perfil ideal del funcionario para cada segmento.
- Modelo de Valor - Rentabilidad
 - Alto Valor : Fidelización del cliente
 - Bajo Valor: No enfocarse



Conclusiones

- Los modelos desarrollados permitieron obtener la mejor solución **cuantitativa y cualitativa**.
- La solución presentada permite realizar **ofertas de crédito acertadas**
- Las ventas de los clientes que poseen historial en el banco son **2 veces mejor** predecidas que los clientes sin historia.



Conclusiones

- Los modelos desarrollados permitieron desarrollar la mejor solución **cuantitativa y cualitativa**.
- La solución presentada permite realizar **ofertas de crédito acertadas**
- Las ventas de los clientes que poseen historial de ventas en el banco son **2 veces mejor** predecidas que los clientes sin historia.

Trabajo Futuro

- Realizar la implementación y documentación del proyecto
- Segmentar el grupo Microempresa con variables de endeudamiento.
- Analizar la incorporación de variables macroeconómicas mensuales tales como:
 - Tasa de desempleo,
 - Crecimiento en el PIB





Muchas Gracias!



[Contacto](#)