

Cálculo de la Incertidumbre sobre el Retorno de Deuda

Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Guadalajara

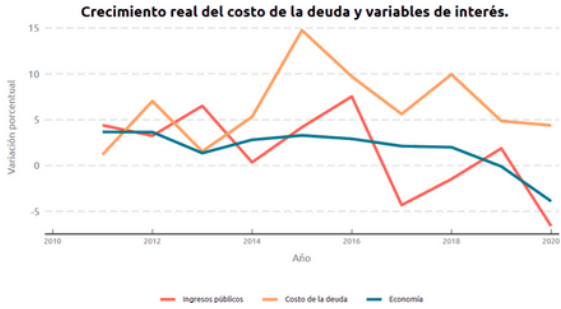


Motivación y Enfoque

- Recuperar deuda** de clientes morosos es uno de los mayores retos para las instituciones financieras.
- No solo es difícil predecir quién pagará, sino también **cuánto pagará**.
- El valor ideal del ratio se sitúa entre el 0.4 y 0.6. Si la ratio es superior, la empresa tiene excesivas deudas y está perdiendo su autonomía financiera.
- Análisis de datos** avanzado como herramienta para tomar decisiones estratégicas.

Objetivo del proyecto: ¿Qué queremos alcanzar?

- Segmentación** de la cartera de clientes por productos y grupos.
- Construir **modelos predictivos** para predecir cuánto pagará cada cliente y comparar estos modelos.
- Diseñar **estrategias de cobranza** personalizadas según segmento y predicción.
- Desarrollar una **interfaz** intuitiva para análisis y toma de decisiones.



Pre-procesamiento y Feature Selection

¿Por qué es importante limpiar y transformar las variables?

Buscamos la **probabilidad** de que se recupere cualquier porción de la deuda de un cliente. Es por ello que definimos las siguientes variables como objetivo o relevantes ya sea para el **feature engineering** o para el entrenamiento del modelo

Variable	Tipo de Dato	Descripción
Variable_Objetivo	int64	1 si el cliente pago > 0, 0 lo contrario
Producto	object	Tipo de producto asociado al cliente
Pago_M _i	float64	Cantidad que el cliente aportó a su deuda ese mes
Behavior_M _i	float64	Calificación creada por OSF del cliente
Ciclo_atraso_M _i	float64	Que tantos meses el cliente no ha pagado el mínimo
Saldo_mes_M _i	float64	Deuda parcial del mes correspondiente

¿Cual es la importancia de Feature Selection?

- Para el **feature selection** se modelaron las **combinaciones** posibles de las variables numéricas usando una regresión logística.
- Las variables que persisten a lo largo de los tres productos incluyen, **dias_deudados_m1** y la codificación de **Segmento ABC**. Algunas otras variables relevantes fueron, **dias_deudados_m2**.
- Realizar **feature selection** permite **reducir la complejidad** del modelo y **retener explicabilidad** de los modelos

Features	# Features	Accuracy	AUC
dias_deudados_m1 + dias_deudados_m2 + B + C	4	0.8025	0.8846
dias_deudados_m1 + dias_deudados_m2 + Variable_Objetivo_m1 + B + C	5	0.8004	0.8867
dias_deudados_m1 + dias_deudados_m2 + Variable_Objetivo_m1 + Variable_Objetivo_m2 + B + C	6	0.8003	0.8867
dias_deudados_m1 + dias_deudados_m2 + Variable_Objetivo_m2 + B + C	5	0.8002	0.8864
dias_deudados_m1 + dias_deudados_m2 + Variable_Objetivo_m1 + B	4	0.7950	0.8821

Flujo de Trabajo

Segmentación ABC

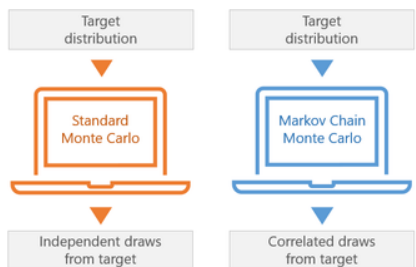
- La segmentación ABC se utilizó para **segmentar** a los clientes de **cada uno de los productos** (prestamos personales, tarjeta de créditos departamentales, y tarjetas de crédito visa).
- Como variable para medir la magnitud de la deuda para cada cliente se utilizó la norma **Saldo_mes_M_i**

$$L2_Saldo_Mes = \sqrt{\sum_{i=0}^6 (Saldo_Mes_M_i)^2}$$

Modelos Utilizados

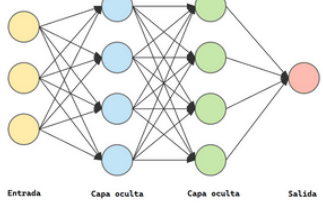
1. Modelos Bayesianos MCMC

Representación estadística que permite combinar información previa con datos observados para realizar inferencias sobre parámetros o hipótesis.



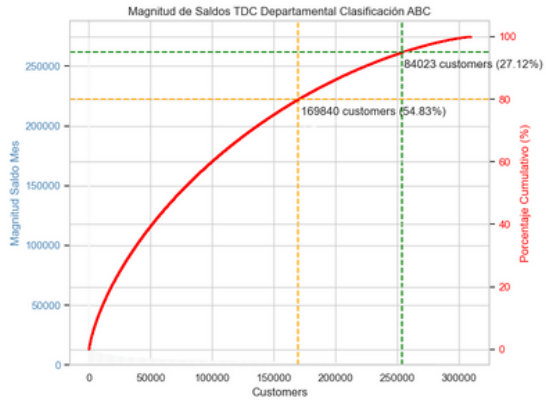
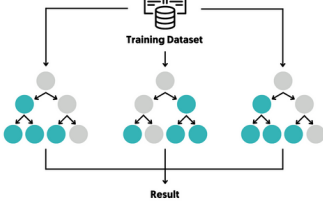
2. Redes neuronales

Se usó un perceptrón multicapa (MLP) como modelo de predicción de deuda porque tiene la capacidad de captar relaciones no lineales y complejas entre las variables de entrada y el comportamiento de los clientes.

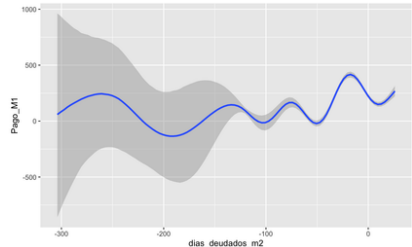


3. Regresión: Random Forest

Algoritmo de aprendizaje automático que utiliza una colección de árboles de decisión para realizar predicciones.



Ventajas y desventajas de modelos Determinista vs. Probabilístico



Modelo Bayesiano

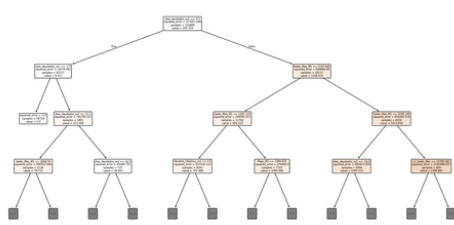
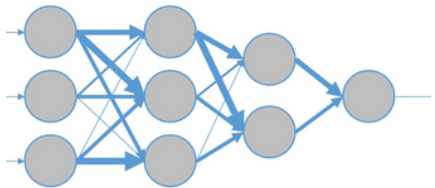
Red Neuronal

Ventajas

- Alta capacidad de aprendizaje
- Son muy flexibles
- Capaces de crear modelos no lineales

Desventajas

- Difícil interpretabilidad
- Tienden a sobre entrenarse con facilidad
- Tienen una función de pérdida en general no convexa con más de un mínimo local



Random Forest

Ventajas:

- Ofrece **robustez** de modelo
- Resistente al **overfitting**
- Son **interpretables**; Simple
- Permite ver las **features más importantes** para predecir.

Desventajas

- Alto costo computacional** y de memoria
- Puede quedar corto frente a arquitecturas más potentes como redes neuronales

Resultados y Estrategias

Predictores Deterministas

1. Red Neuronal

Métrica	PP	TDC Dept	TDC Visa
R ²	0.93	0.85	0.81
MAE	38.49	8.31	39.81
MSE	20421	3474.35	34956.93

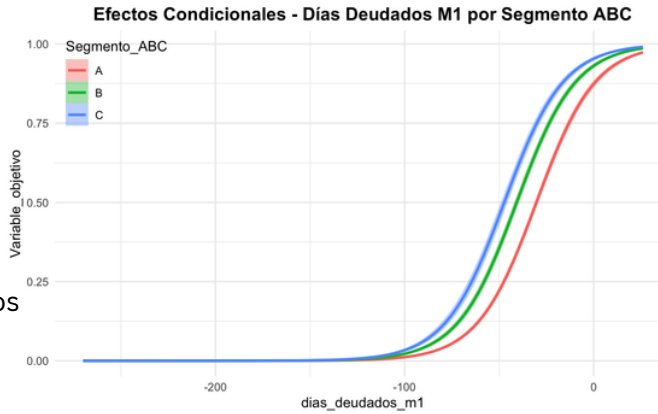
2. Random Forest

Métrica	PP	TDC Dept	TDC Visa
R ²	0.98	0.84	0.89
MAE	18.25	7.83	44.93
MSE	5792.81	3460.21	12323.49

- Los modelos deterministas obtienen un rendimiento satisfactorio **R² > 0.9**, e incluso similar en el pronóstico de la cantidad de pago para cada cliente
- Si al final se desea seleccionar solo un modelo, es necesario considerar que el método probabilístico trabaja bajo diferentes métricas que los modelos convencionales (deterministas), mientras que para los predictores continuos se podrían aplicar pruebas estadísticas como la **prueba de Wilcoxon** que compara las mediciones obtenidas

Predictores Probabilísticos

- Los resultados del **MCMC** proporcionan estimaciones puntuales y la capacidad de realizar inferencias estadísticas
- Se entrenaron **dos** modelos;
 - Modelo discreto**, que determina la probabilidad de 0 o 1 de pago
 - Modelo Continuo**, que determina la cantidad de pago
- En la grafica de efectos podemos observar que el segmento “A” seguido del “B” y luego “C” muestra el comportamiento de pago por segmento que indica la **probabilidad de pagar más del mínimo**



Estrategias Efectivas

Segmentación de Clientes:

Dividir a los clientes en grupos según su comportamiento de pago y adaptar la estrategia de cobranza a cada segmento.

Conclusión y Producto Final

- Se desarrolló una **interfaz interactiva** que integra los tres modelos predictivos, accesible a través de una página web intuitiva.
- Comparador de clientes:** permite visualizar diferencias clave entre perfiles de pago, mostrando predicciones y comparaciones de variables de comportamiento financiero.
- Simulador interactivo:** posibilita ingresar características demográficas y financieras de un cliente simulado para predecir su comportamiento de pago.



Demo de dashboard