CALIBRACIÓN, MÉTRICAS Y EXPLICABILIDAD DE MODELOS DE IA

EJERCICIO DE FEEDBACK 1

Sergi Zarzuelo Abelló

Índice

Exploración Inicial, EDA y Limpieza de Datos

Se ha realizado una exploración inicial sobre el dataset *Default of Credit Card Clients*, del repositorio UCI, el cual contiene información de 30.000 clientes de una entidad de crédito en Taiwán, con las siguientes variables:

- **LIMIT_BAL**: Monto total del crédito otorgado (incluye préstamos personales y tarjeta de crédito).
- **SEX**: Género del cliente (1 = hombre, 2 = mujer).
- **EDUCATION**: Nivel educativo (1 = posgrado, 2 = universidad, 3 = secundaria, 4 = otro).
- **MARRIAGE**: Estado civil (1 = casado, 2 = soltero, 3 = otro).
- **AGE**: Edad del cliente (en años).
- **PAY_0** a **PAY_6**: Estado del pago del cliente en los últimos 6 meses:
    - -1 = pagó debidamente.
    - 0 = deuda totalmente pagada.
    - 1 = pago con 1 mes de retraso, 2 = con 2 meses de retraso, etc.
- **BILL_AMT1** a **BILL_AMT6**: Monto de la factura en NT$ (dólar taiwanés) para cada uno de los últimos 6 meses.
- **PAY_AMT1** a **PAY_AMT6**: Monto del pago realizado en NT$ para cada uno de los últimos 6 meses.
- *(variable objetivo)* **default payment next month**: Indica si el cliente impagó su deuda el mes siguiente (1) o no (0).

Variables categoricas

Se han realizado diversas pruebas para evaluar la calidad del dataset, así como análisis de valores nulos (no hay nulos), registros duplicados (no hay duplicados) y de consistencia de datos. Para esto último, se ha observado inconsistencias en algunos de los valores de algunas variables, concretamente:

- **EDUCATION**: se observan valores anómalos (0, 5 y 6) diferentes al rango de valores que indica la documentación del dataset (1-4)
- **MARRIAGE**: valores anómalos (0) diferentes al rango de valores que indica la documentación del dataset (1-3)
- **PAY_X**: valores anómalos (-2) diferentes al rango de valores que indica la documentación del dataset (>-1)

Como solución, todos estos valores anómalos han sido colocados en las respectivas categorías de "Otros". Adicionalmente, para el caso de PAY_X, el valor -2 se ha clasificado como -1.

SEX	n_clientes	tasa_default
1	11888	0.241672
2	18112	0.207763

EDUCATION_aj	n_clientes	tasa_default
1	10585	0.192348
2	14030	0.237349
3	4917	0.251576
4	468	0.070513

MARRIAGE_aj	n_clientes	tasa_default
1	13659	0.234717
2	15964	0.209283
3	377	0.236074

Tabla 1: Resumen variables categóricas

Análisis de PAY_0_aj:

PAY_0_aj	n_clientes	tasa_default
-1	8445	0.156187
0	14737	0.128113
1	3688	0.339479
2	2667	0.691414
3	322	0.757764
4	76	0.684211
5	26	0.500000
6	11	0.545455
7	9	0.777778
8	19	0.578947

Análisis de PAY_2_aj:

PAY_2_aj	n_clientes	tasa_default
-1	9832	0.168531
0	15730	0.159123
1	28	0.178571
2	3927	0.556150
3	326	0.616564
4	99	0.505051
5	25	0.600000
6	12	0.750000
7	20	0.600000
8	1	0.000000

Análisis de PAY_3_aj:

PAY_3_aj	n_clientes	tasa_default
-1	10023	0.167914
0	15764	0.174512
1	4	0.250000
2	3819	0.515580
3	240	0.575000
4	76	0.578947
5	21	0.571429
6	23	0.608696
7	27	0.814815
8	3	0.666667

Análisis de PAY_4_aj:

PAY_4_aj	n_clientes	tasa_default
-1	10035	0.173493
0	16455	0.183288
1	2	0.500000
2	3159	0.523267
3	180	0.611111
4	69	0.666667
5	35	0.514286
6	5	0.400000
7	58	0.827586
8	2	0.500000

Análisis de PAY_5_aj:

PAY_5_aj	n_clientes	tasa_default
-1	10085	0.177690
0	16947	0.188529
2	2626	0.541889
3	178	0.634831
4	84	0.607143
5	17	0.588235
6	4	0.750000
7	58	0.827586
8	1	1.000000

Análisis de PAY_6_aj:

PAY_6_aj	n_clientes	tasa_default
-1	10635	0.183921
0	16286	0.188444
2	2766	0.506508
3	184	0.641304
4	49	0.632653
5	13	0.538462
6	19	0.736842
7	46	0.826087
8	2	1.000000

Tabla 2: Resumen variables categóricas

Adicionalmente, se observa cierto desbalanceo entre las clases de la variable objetivo, con aproximadamente un 20% de eventos positivos (moras) frente al 80% de eventos negativos (no moras). Pese a que el porcentaje no se encuentra tan desbalanceado como podria ocurrir con un modelo de fraude (p.ej. 1%-99%), se realizarán pruebas más adelante sobre si ajustar el desbalanceo otorga mejores resultados al modelo o no.

Variables numéricas

Se analizan las variables numéricas, analizando en primer lugar los estadísticos más habituales:

	LIMIT_BAL	AGE	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6
count	30000.000000	30000.000000	30000.000000	30000.000000	3.000000e+04	30000.000000	30000.000000	30000.000000
mean	167484.322667	35.485500	51223.330900	49179.075167	4.701315e+04	43262.948967	40311.400967	38871.760400
std	129747.661567	9.217904	73635.860576	71173.768783	6.934939e+04	64332.856134	60797.155770	59554.107537
min	10000.000000	21.000000	-165580.000000	-69777.000000	-1.572640e+05	-170000.000000	-81334.000000	-339603.000000
25%	50000.000000	28.000000	3558.750000	2984.750000	2.666250e+03	2326.750000	1763.000000	1256.000000
50%	140000.000000	34.000000	22381.500000	21200.000000	2.008850e+04	19052.000000	18104.500000	17071.000000
75%	240000.000000	41.000000	67091.000000	64006.250000	6.016475e+04	54506.000000	50190.500000	49198.250000
max	1000000.000000	79.000000	964511.000000	983931.000000	1.664089e+06	891586.000000	927171.000000	961664.000000

Tabla 3: Resumen variables numéricas

Para complementar esta información de manera visual, se han graficado las distribuciones de las variables y su relación con la variable objetivo, tal y como se detalla a continuación.

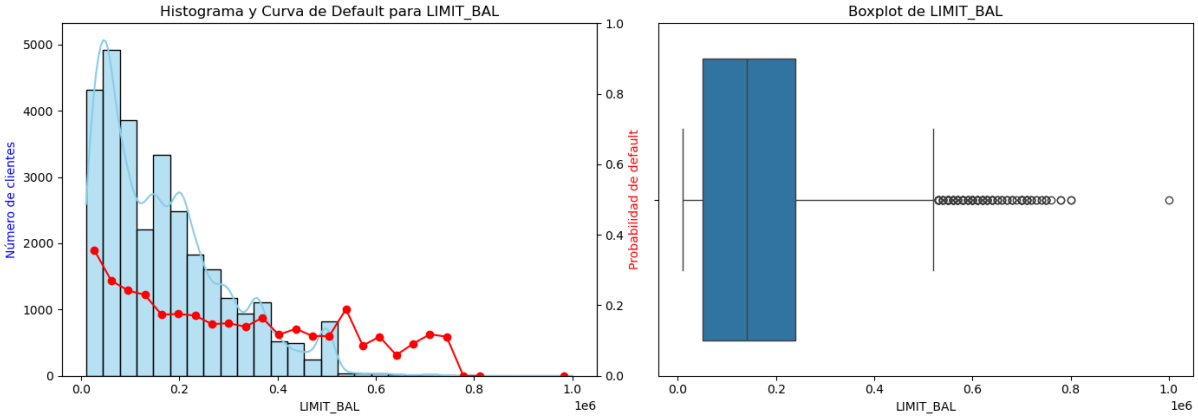

Ilustración 1: Distribución de LIMIT_BAL

Se observa una tendencia lógica de la distribución, donde se tiene un mayor volumen de préstamos con bajos importes y una menor materialidad de préstamos de importes elevados. Al haber más prestamos con bajos importes, también existe mayor probabilidad de impago, tal y como se evidencia en la curva roja decreciente.

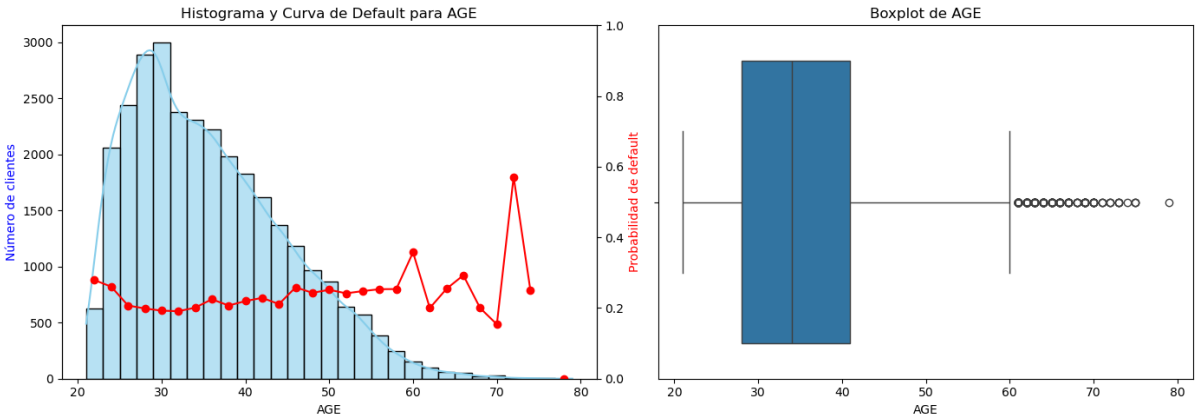

Ilustración 2: Distribución de AGE

En cuanto a la distribución de AGE, se encuentra una tendencia con sentido, con poca gente joven y mayor, y con gente adulta como predominante (son los que lógicamente piden más préstamos). En cuanto a la TD, se observa un leve incremento en las personas muy jóvenes (<25 años aproximadamente), pero la tendencia se mantiene relativamente estable de ahí en adelante, con una leve tendencia a incrementar con la edad, lo que también hace sentido.

En cuanto al análisis de las variables BILL_AMT y PAY_AMT, se evalúa la construcción de una variable ratio entre estas, con el objetivo de analizar el comportamiento en conjunto y capturar el porcentaje de la deuda que ha sido abonado por el cliente en cada mes. Esta nueva variable, acotada superiormente en 1 para evitar distorsiones, aporta una señal más clara y consistente que las variables originales por separado. Su comportamiento frente a la morosidad permite extraer conclusiones más precisas, reduciendo la volatilidad observada en las tendencias individuales de BILL_AMT y PAY_AMT.

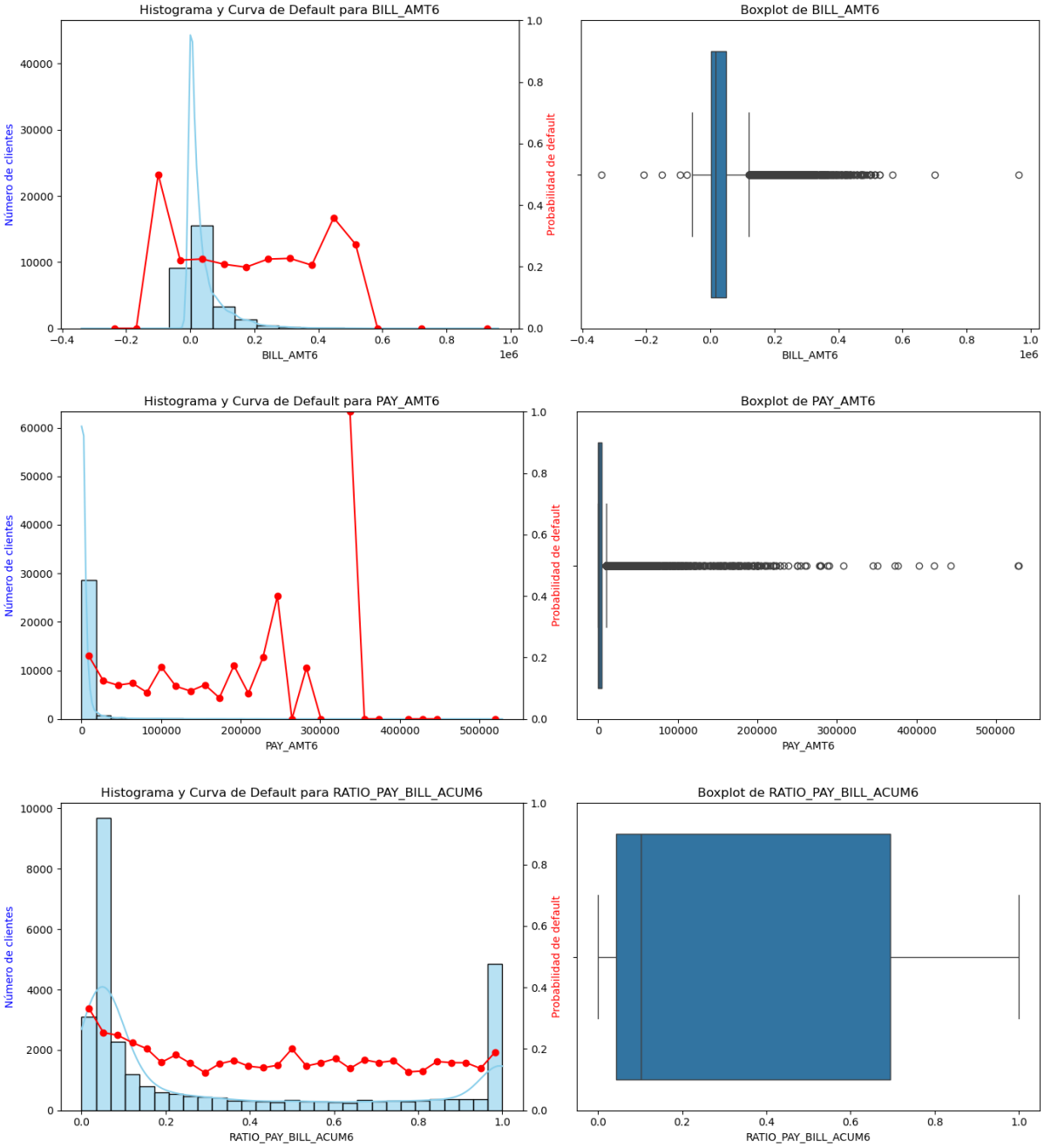

Ilustración 3: Distribución de LIMIT_BAL

De esta nueva manera, se puede tener una escala comparable para medir qué porcentaje de lo debido a podido pagar el cliente, observándose una peor morosidad en valores cercanos al 0% del ratio y una mejor en los cercanos al 100%. Adicionalmente, mencionar que gracias al boxplot se puede observar que la mayoría de los clientes va con cierto retraso en sus pagos mensualmente, estando el IQR entre el 10% y el 65% pagado.

Entrenamiento y Evaluación de Modelos

Previo al entrenamiento de los modelos, se realiza un ejercicio de tratamiento de variables, para obtener mejores resultados en el desempeño final. Aunque previamente ya se hayan realizado ciertos

ajustes obvios, en esta sección se evaluarán todo tipo de tratamientos para todas las variables disponibles.

Data preprocessing

Outliers

Para tratar los outliers se proponen dos metodologías estandarizadas:

- Acotar por IQR

Variable	Porcentaje IQR (%)	MAX Original	MIN Original	MAX después de Acotar IQR	MIN después de Acotar IQR
AGE	0.906667	79.0	21.0	60.5	21.0
LIMIT_BAL	0.556667	1000000.0	10000.0	525000.0	10000.0
RATIO_PAY_BILL_ACUM1	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM2	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM3	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM4	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM5	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM6	0.000000	1.0	0.0	1.0	0.0

Tabla 4: Resumen acotación IQR

- Acotar por percentiles 1 y 99

Variable	Porcentaje Percentil 1-99 (%)	MAX Original	MIN Original	MAX después de Acotar P1-P99	MIN después de Acotar P1-P99
AGE	1.130000	79.0	21.0	60.0	22.0
LIMIT_BAL	0.686667	1000000.0	10000.0	500000.0	10000.0
RATIO_PAY_BILL_ACUM1	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM2	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM3	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM4	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM5	0.000000	1.0	0.0	1.0	0.0
RATIO_PAY_BILL_ACUM6	0.000000	1.0	0.0	1.0	0.0

Tabla 5: Resumen acotación percentiles (1-99)

Analizando los resultados, se observa mejores resultados usando los percentiles 1-99, puesto que el IQR es demasiado agresivo para RATIO_PAY_BILL_ACUM1, por lo que se elige los percentiles como metodología final de acotación. No obstante, los resultados tampoco deberían de ser muy diferentes de haber elegido el IQR.

WOE (Numéricas)

Se crean agrupaciones en las variables numéricas en base a un WOE binning (agrupaciones por WOE diferenciado), asignando el valor del WOE a cada agrupación. Con esto último, se generalizan las linealidades al identificar cada agrupación con un valor logarítmico, y normalizado a una escala comparable de afectación sobre la variable objetivo (a diferencia de una categorización ordinal). Es decir, en vez de categorizar las agrupaciones como 1, 2 o 3, tomarán los valores 1.32, 0.56 y -1.81 (su WOE).

WOE binning para variable: LIMIT_BAL

Bin	Count	Count (%)	Event rate	WoE
(-inf, 25000.00)	2471	0.082367	0.362202	-0.692865
[25000.00, 45000.00)	1840	0.061333	0.358696	-0.677657
[45000.00, 75000.00)	4921	0.164033	0.269864	-0.263374
[75000.00, 125000.00)	4580	0.152667	0.242795	-0.121269
[125000.00, 165000.00)	3282	0.109400	0.198355	0.137923
[165000.00, 205000.00)	3284	0.109467	0.174482	0.2955
[205000.00, 245000.00)	2555	0.085167	0.167906	0.341854
[245000.00, 365000.00)	4591	0.153033	0.151165	0.466803
[365000.00, inf)	2476	0.082533	0.118740	0.74573
Special	0	0.000000	0.000000	0.0
Missing	0	0.000000	0.000000	0.0

Tabla 6: Ejemplo WOE para variable LIMIT_BAL

En el ejemplo de LIMIT_BAL, se puede observar cómo se crean agrupaciones con tasa de default diferenciada entre sí, o en otras palabras, con WOEs diferenciados.

WOE (Categóricas)

Se realiza el mismo proceso para las variables categóricas, explorando agrupaciones de las categorías en base al WOE.

WOE binning para variable: MARRIAGE_aj

Bin	Count	Count (%)	Event rate	WoE
[2]	15964	0.532133	0.209283	0.070563
[1, 3]	14036	0.467867	0.234753	-0.077025
Special	0	0.000000	0.000000	0.0
Missing	0	0.000000	0.000000	0.0

Tabla 7: Ejemplo WOE para variable MARRIAGE

En el ejemplo de MARRIAGE, se puede observar que las categorías 1 y 3 son agrupadas, lo que tiene sentido tal y como se ha podido observar en Tabla 1, donde se observaba una tasa de default similar entre ambas categorías.

Comparativa de variables numéricas

A partir de este punto, las variables numéricas tienen las siguientes formas:

1. La variable original
2. La variable acotada
3. La variable original convertida con WOE
4. La variable acotada convertida con WOE

Como no es óptimo incluir todas estas al entrenamiento del modelo, se realiza una comparativa previa para ver qué versión tiene un mejor poder predictivo entre las 4 opciones, realizando una comparativa bivariante de cada variable con la variable objetivo y concluyendo según el AUC.

VARIABLE	AUC_og	AUC_og_WOE	AUC_acot	AUC_acot_WOE	Mejor versión
LIMIT_BAL	0.617803	0.618294	0.617801	0.618294	WOE
AGE	0.503579	0.541413	0.503575	0.541413	WOE
RATIO_PAY_BILL_ACUM1	0.561582	0.573281	0.561582	0.573281	WOE
RATIO_PAY_BILL_ACUM2	0.565143	0.579578	0.565143	0.579578	WOE
RATIO_PAY_BILL_ACUM3	0.571382	0.580262	0.571382	0.580262	WOE
RATIO_PAY_BILL_ACUM4	0.572966	0.580826	0.572966	0.580826	WOE
RATIO_PAY_BILL_ACUM5	0.575170	0.584450	0.575170	0.584450	WOE
RATIO_PAY_BILL_ACUM6	0.577405	0.586032	0.577405	0.586032	WOE

Tabla 8: Comparativa de las versiones de las variables numéricas

Se observa que, para todas ellas, la versión con mayor poder predictivo son aquellas transformadas con el WOE, observándose un rendimiento similar sin acotar o acotando, por lo que se escoge la versión sin acotar.

Entrenamiento de Modelos

Para el entrenamiento, se exploran diversos tipos de modelos. En todos ellos, se plantea un enfoque stepwise, en el que se van incluyendo / eliminando las variables una a una según el poder predictivo que sumen / resten al modelo final, obteniendo finalmente la configuración de variables que otorguen un mayor rendimiento.

De manera resumida, el proceso seguido en cada uno de los siguientes desarrollos incluye los siguientes puntos:

- Se utiliza una estrategia de selección de variables tipo stepwise forward, añadiendo variables solo si mejoran una métrica objetivo (AUC, F1, recall, etc.) por encima de un umbral (threshold).
- Permite corregir el desbalance de clases mediante dos enfoques: SMOTE (use_smote=True) o class_weight='balanced' en el modelo.
- Realiza una división estratificada de los datos en entrenamiento (70%) y test (30%).
- Aplica escalado estándar a las variables predictoras mediante StandardScaler.
- Evalúa el rendimiento con métricas ajustables: AUC, F1, precisión, recall o accuracy.
- Implementa validación cruzada estratificada (K-Fold) para evaluar el rendimiento del modelo.

Random Forest

Se han obtenido los siguientes resultados en modelos de tipo random forest.

METODOLOGIA	STEPWISE	AUC	ACCURACY	PRECISION	RECALL	F1	TN	FP	FN	TP	FEATURES_USADAS
Balanced	AUC	75.4%	77.1%	48.4%	55.8%	51.8%	3882	790	586	741	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE
Balanced	RECALL	57.9%	47.5%	25.1%	69.4%	36.9%	1930	2742	406	921	RATIO_PAY_BILL_ACUM2_WOE
Balanced	F1	74.0%	75.9%	46.5%	60.2%	52.5%	3753	919	528	798	PAY_0_aj_WOE, PAY_3_aj_WOE, PAY_2_aj_WOE, PAY_4_aj_WOE
SMOTE	AUC	75.3%	77.0%	48.3%	56.2%	51.9%	3874	798	581	745	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE
SMOTE	RECALL	57.9%	47.5%	25.1%	69.4%	36.9%	1930	2742	406	921	RATIO_PAY_BILL_ACUM2_WOE
SMOTE	F1	73.8%	75.8%	46.3%	60.3%	52.4%	3745	927	526	800	PAY_0_aj_WOE, PAY_3_aj_WOE, PAY_2_aj_WOE, PAY_4_aj_WOE

Tabla 9: Resumen de los resultados de los modelos Random Forest

Se puede ver que entre usar SMOTE o el balanceo de clases los resultados obtenidos son similares, por lo que no hay una opción predominante sobre la otra en términos generales.

En cuanto a la métrica utilizada para el stepwise, se puede ver que la opción del F1 ofrece un equilibrio entre si se usa el AUC, donde solamente se mide el AUC, y la segunda, donde solamente se mide el

Recall. Con la opción del F1, se obtiene el mejor equilibrio entre Precision y Recall, y el AUC resultante es muy similar a si solamente se optimiza el AUC, por lo que se considera la solución óptima.

Sobre esta elección, finalmente quedaría acordar con el cliente para alinear el sentido de negocio, es decir, si se quiere dar mayor / menor peso a los verdaderos / falsos positivos / negativos, y ajustar el modelo en consecuencia.

Gradient Boosting

Se realiza el mismo proceso para un modelo de Gradient Boosting, obteniendo los siguientes resultados:

METODOLOGIA	STEPWISE	AUC	ACCURACY	PRECISION	RECALL	F1	TN	FP	FN	TP	FEATURES_USADAS
No SMOTE	AUC	76.9%	82.0%	67.0%	36.5%	47.2%	4433	239	843	484	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE, RATIO_PAY_BILL_ACUM5_WOE, PAY_2_aj_WOE, PAY_6_aj_WOE, RATIO_PAY_BILL_ACUM2_WOE, MARRIAGE_aj_WOE, PAY_4_aj_WOE, SEX_WOE
No SMOTE	RECALL	64.9%	79.7%	56.0%	37.2%	44.7%	4285	387	833	494	PAY_2_aj_WOE
No SMOTE	F1	73.0%	81.9%	66.8%	36.2%	46.9%	4430	242	846	481	PAY_0_aj_WOE, PAY_4_aj_WOE
SMOTE	AUC	76.0%	77.5%	49.3%	55.7%	52.3%	3911	761	587	739	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE, PAY_6_aj_WOE, PAY_4_aj_WOE, RATIO_PAY_BILL_ACUM5_WOE
SMOTE	RECALL	57.9%	47.5%	25.1%	69.4%	36.9%	1930	2742	406	921	RATIO_PAY_BILL_ACUM2_WOE
SMOTE	F1	73.2%	76.1%	46.8%	59.2%	52.2%	3777	895	541	785	PAY_0_aj_WOE, PAY_3_aj_WOE, PAY_2_aj_WOE

Tabla 10: Resumen de los resultados de los modelos Gradient Boosting

Para el GBM no existe la función class_weight = balanced, por lo que las dos opciones que se han explorado han sido usar SMOTE o simplemente no ajustar el desbalanceo. En esta ocasión, a diferencia del random forest, sí que se observan mejores resultados al utilizar SMOTE, principalmente si se analiza el Recall o, mejor, el F1-Score.

Sobre estas opciones, el modelo más completo parece ser el GBM - SMOTE - AUC, puesto que es el que mejor AUC tiene y tiene un F1 prácticamente similar al modelo GBM - SMOTE - F1.

Regresión Logística

Por último, se analizan modelos de regresión logística, obteniendo los siguientes resultados:

METODOLOGIA	STEPWISE	AUC	ACCURACY	PRECISION	RECALL	F1	TN	FP	FN	TP	FEATURES_USADAS
Balanced	AUC	75.9%	77.7%	49.7%	55.8%	52.6%	3923	749	586	741	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE, SEX_WOE, MARRIAGE_aj_WOE, PAY_2_aj_WOE, PAY_4_aj_WOE
Balanced	RECALL	58.0%	47.5%	25.1%	69.4%	36.9%	1930	2742	406	921	RATIO_PAY_BILL_ACUM2_WOE
Balanced	F1	75.3%	77.3%	48.9%	56.2%	52.3%	3892	780	580	746	PAY_0_aj_WOE, PAY_3_aj_WOE, LIMIT_BAL_WOE
SMOTE	AUC	75.9%	77.8%	49.9%	55.8%	52.7%	3928	744	586	741	PAY_0_aj_WOE, LIMIT_BAL_WOE, PAY_3_aj_WOE, SEX_WOE, MARRIAGE_aj_WOE, PAY_4_aj_WOE, PAY_2_aj_WOE
SMOTE	RECALL	58.0%	47.5%	25.1%	69.4%	36.9%	1930	2742	406	921	RATIO_PAY_BILL_ACUM2_WOE
SMOTE	F1	75.6%	77.4%	49.1%	56.6%	52.6%	3893	779	576	750	PAY_0_aj_WOE, PAY_3_aj_WOE, LIMIT_BAL_WOE, PAY_4_aj_WOE

Tabla 11: Resumen de los resultados de los modelos Regresión Logística

Por último, para el LOGREG vuelve a estar disponible la opción de ajustar los pesos de la clase desbalanceada, y nuevamente la conclusión es la misma que con el RF: no hay diferencia significativa entre un método u otro..

En cuanto al modelo más equilibrado, podría ser tanto LOGREG - Balanced - AUC como LOGREG - SMOTE - AUC, debido que tienen el AUC más alto y el F1 es prácticamente similar a los modelos construidos optimizando el F1.

Comparativa

Comparando todos estos modelos, se observa que el modelo con mayor AUC es aquel con GBM – No SMOTE - AUC. No obstante, si analizamos el recall puede verse que este es muy bajo (36.5%), por lo que no resultaría optimo en un modelo de detección de riesgo crediticio como este (estaríamos dejando sin identificar muchos eventos positivos).

Por el contrario, lo más interesante en estos modelos es tener un buen F1, pudiendo mantener paralelamente un buen AUC. Por esa razón, ordenando los modelos por F1, se puede ver que hay otros modelos mejores. Concretamente, **el modelo LOGREG – BALANCED - AUC tiene uno de los F1 más altos (52,6%), y también uno de los mayores AUC (75,9%), por lo que se elige como el mejor candidato**.

Explicabilidad del Modelo con SHAP

Una vez escogido el mejor modelo, se analiza la explicabilidad de este utilizando la librería SHAP. Para ello, se repara en el modelo de regresión logística con los features:

- PAY_0_aj_WOE
- LIMIT_BAL_WOE
- PAY_3_aj_WOE
- SEX_WOE
- MARRIAGE_aj_WOE
- PAY_2_aj_WOE
- PAY_4_aj_WOE

Sobre este modelo, se aplica la función *Explainer* de la librería, que otorga información sobre la importancia de cada una de estas variables del modelo.

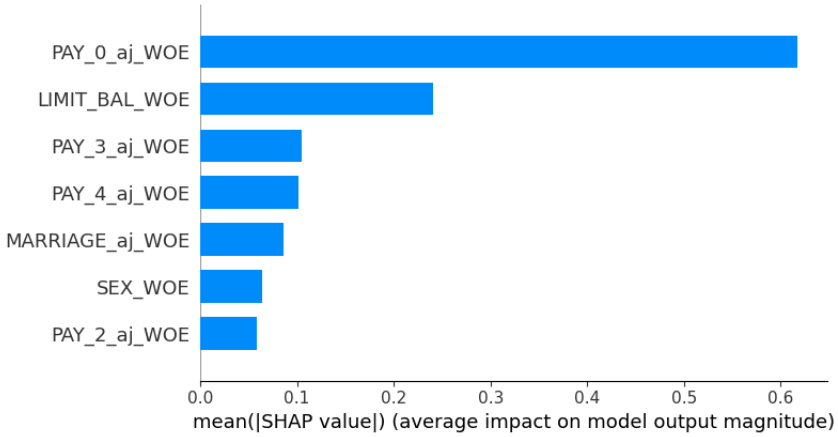

Ilustración 4: Importancia de las variables del modelo final

Entre estas, se observa que la más relevante es el comportamiento del cliente en el último mes disponible (PAY_0: septiembre de 2005)[1], lo que tiene sentido para determinar la morosidad en el siguiente mes.

Otra variable relevante es el LIMIT_BAL, lo que también hace sentido, ya que el límite otorgado a cada cliente está relacionado con su comportamiento y, en consecuencia, con su morosidad (bajos limites están relacionados con más morosidades, tal y como se ha analizado en el EDA inicial).

[1] Según la documentación del dataset: PAY_0: estado del pago en sep05 ; PAY_1: estado del pago en aug05 ; PAY_2: estado del pago en jul05; etc.

Por otro lado, se analiza el detalle de una predicción especifica, para ver como el modelo estima la probabilidad de impago y la relevancia de cada variable en dicha predicción. Para ello, se usa una gráfica tipo waterfall.

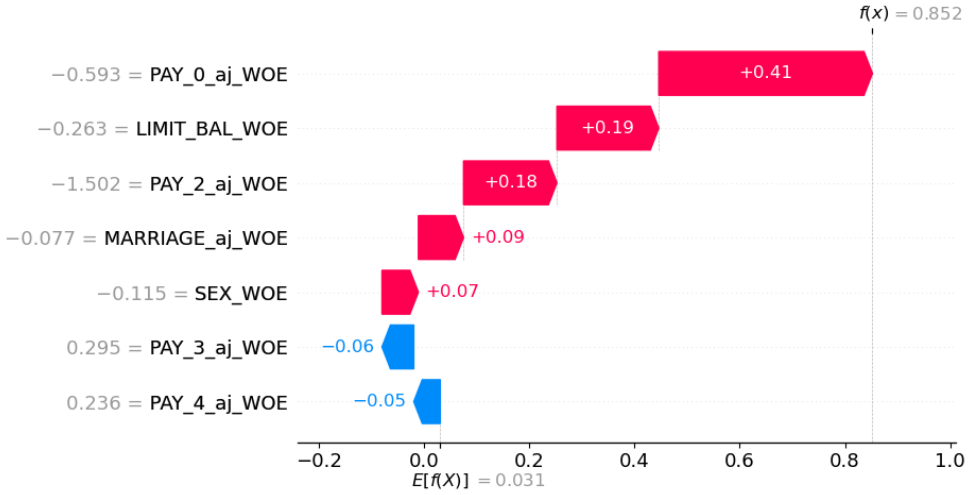

Ilustración 5: Aportación de cada variable a la predicción individual de un cliente

Tal y como se ha visto inicialmente, las dos variables que más influyen en la predicción son PAY_0 y LIMIT_BAL, en este caso para aumentar la morosidad. Como las variables han entrado al modelo con su valor WOE, la interpretación no es tan directa, pero para ello se muestra el detalle de los valores originales de las variables en la tabla de debajo, donde se comprueba que el cliente presentaba un mes de atraso en PAY_0 y dos meses en PAY_2 (ambos han contribuido a empeorar su predicción).

PAY_0_aj	PAY_0_aj_WOE	PAY_2_aj	PAY_2_aj_WOE	PAY_3_aj	PAY_3_aj_WOE	PAY_4_aj	PAY_4_aj_WOE	SEX	SEX_WOE	MARRIAGE_aj	MARRIAGE_aj_WOE	LIMIT_BAL	LIMIT_BAL_WOE	default payment next month
1	-0.593072	2	-1.501653	0	0.295164	0	0.235542	1	-0.115154	1	-0.077025	50000	-0.263374	1

Tabla 12: Detalle del cliente analizado

Por otro lado, como en PAY_3 y PAY_4 no se observan atrasos en los pagos, estos contribuyen a mejorar la estimación, aunque en menor medida puesto que tienen un menor peso (son más alejadas en el tiempo, el comportamiento reciente es más relevante).

Otras variables como SEX (1: Hombre) y MARRIAGE (1: casado) aumentan la probabilidad, ya que estas categorías están relacionadas con mayor morosidad, tal y como se ha visto en el EDA inicial.

Con todo ello, la predicción de la regresión es correcta, puesto que estima una probabilidad de impago de 85,2% para el mes siguiente y efectivamente la marca de default esta activada (el cliente impagó el mes siguiente).

Reflexión y Evaluación Crítica

El modelo de regresión logística desarrollado ha mostrado un desempeño razonable, alcanzando un AUC del 75,9%, posicionándose entre los mejores resultados obtenidos en la comparativa de modelos. Además, logró un F1-score de 52,6%, una métrica especialmente relevante en este contexto debido al desbalance de clases. Dado que la clase negativa es mayoritaria, métricas como la accuracy (77,7%) pueden resultar engañosas, al reflejar una alta capacidad predictiva solo sobre dicha clase, sin capturar adecuadamente el rendimiento sobre los clientes morosos.

Analizando el comportamiento del modelo frente a los impagos, se observan áreas claras de mejora. La clase positiva (clientes que impagan) representa aproximadamente el 22% de la muestra, y a pesar de haberse aplicado un ajuste por pesos balanceados, el recall (55,8%) y la precisión (49,7%) se mantienen en niveles moderados. Esto se traduce en un número elevado de falsos negativos (586), es decir, clientes que deberían haber sido rechazados, pero fueron clasificados como "buenos", lo cual representa un riesgo económico directo.

Una primera limitación es, por tanto, el desbalance de clases. Aunque se ha testado el uso de SMOTE, los resultados no han mejorado significativamente. Por ello, se sugiere explorar métodos alternativos de balanceo como ADASYN, así como la combinación de técnicas de sobremuestreo con selección de variables supervisada.

Otra vía de mejora es el ajuste del umbral de decisión. Actualmente se utiliza el valor por defecto (0.5), pero, dado que el coste asociado a un falso negativo es mayor, podría ser beneficioso desplazar este umbral hacia valores más conservadores, priorizando el recall y disminuyendo el riesgo de impago. Este sería el típico ejercicio a realizar con el cliente para alinear el rendimiento del modelo con las expectativas de negocio.

Asimismo, se recomienda investigar la creación de nuevas variables, por ejemplo, mediante binarizaciones por percentiles optimizadas según métricas como el AUC o el F1-score.

Por último, si bien modelos como Random Forest o Gradient Boosting no superaron la regresión logística, podrían evaluarse otras alternativas modernas como XGBoost, LightGBM o CatBoost. Estos modelos ofrecen mayor capacidad para capturar interacciones complejas y mantienen la interpretabilidad cuando se combinan con herramientas como SHAP.