

Dimplomado en Ciencia de Datos

Modulo II - Modelación Supervisada.


Examen 1

Gustavo Blas Duran

La calificación crediticia es el proceso de evaluar la solvencia de un cliente, respondiendo básicamente a la pregunta:

"¿Qué probabilidad hay de que esta persona pague un préstamo o crédito a tiempo?"

Para respaldar esto, las organizaciones recurren cada vez más a modelos predictivos que analizan los datos de los clientes y les asignan una puntuación o nivel de riesgo.



A continuación se presenta el desarrollo y la evaluación de modelos de calificación crediticia mediante técnicas de clasificación y regresión. Mediante el análisis de datos de los clientes, como su historial de pagos, la utilización del crédito y su comportamiento financiero, los modelos predicen con precisión la probabilidad de impago y estiman el comportamiento de pago. Se identificaron las características clave que contribuyen al riesgo, lo que permitió tomar decisiones crediticias más informadas.

- CUSTOMER ID: Identificador de cada cliente
- LIMIT BAL: Monto de la línea de crédito
- SEX: Genero (1=masculino, 2=femenino)
- EDUCATION: (1=preparatoria, 2=universidad, 3=secundaria, 4=otros, 5=desconocido, 6=desconocido, 0=nulo)
- MARRIAGE: Estatus marital (1=casado, 2=soltero, 3=otros, 0=nulo)
- AGE: Edad en años.
- PAY 0: Estatus de pago en Septiembre, 2005 (-2=no uso (non use), 0=uso, -1=pago correcto, 1=pago retrasado por un mes, retrasado por nueve o mas meses).
- PAY 2: Estatus de pago en Agosto, 2005 (La escala es igual que en PAY 0).
- PAY 3: Estatus de pago en Julio, 2005 (La escala es igual que en PAY 0).
- PAY 4: Estatus de pago en Junio, 2005 (La escala es igual que en PAY 0).
- PAY 5: Estatus de pago en Mayo, 2005 (La escala es igual que en PAY 0).
- PAY 6: Estatus de pago en Abril, 2005 (La escala es igual que en PAY 0).
- BILL AMT1: Monto por pagar en Septiembre, 2005
- BILL AMT2: Monto por pagar en Agosto, 2005
- BILL AMT3: Monto por pagar en Julio, 2005
- BILL AMT4: Monto por pagar en Junio, 2005
- BILL AMT5: Monto por pagar en Mayo, 2005
- BILL AMT6: Monto por pagar en Abril, 2005
- PAY AMT1: Cantidad pagada en Septiembre, 2005
- PAY AMT2: Cantidad pagada en Agosto, 2005
- PAY AMT3: Cantidad pagada en Julio, 2005
- PAY AMT4: Cantidad pagada en Junio, 2005
- PAY AMT5: Cantidad pagada en Mayo, 2005
- PAY AMT6: Cantidad pagada en Abril, 2005
- default.payment.next.month: Incumplimiento de pago (1=si, 0=no)

- El conjunto de datos original cuenta con 25 variables, a partir de las cuales se crearon 18 variables nuevas, no todas son usadas en los modelos. Pues se han seleccionado de acuerdo al al modelo, dependiendo que cuales nos proporcionaron un mejor ajuste del modelo implementado.

Conjunto de Datos

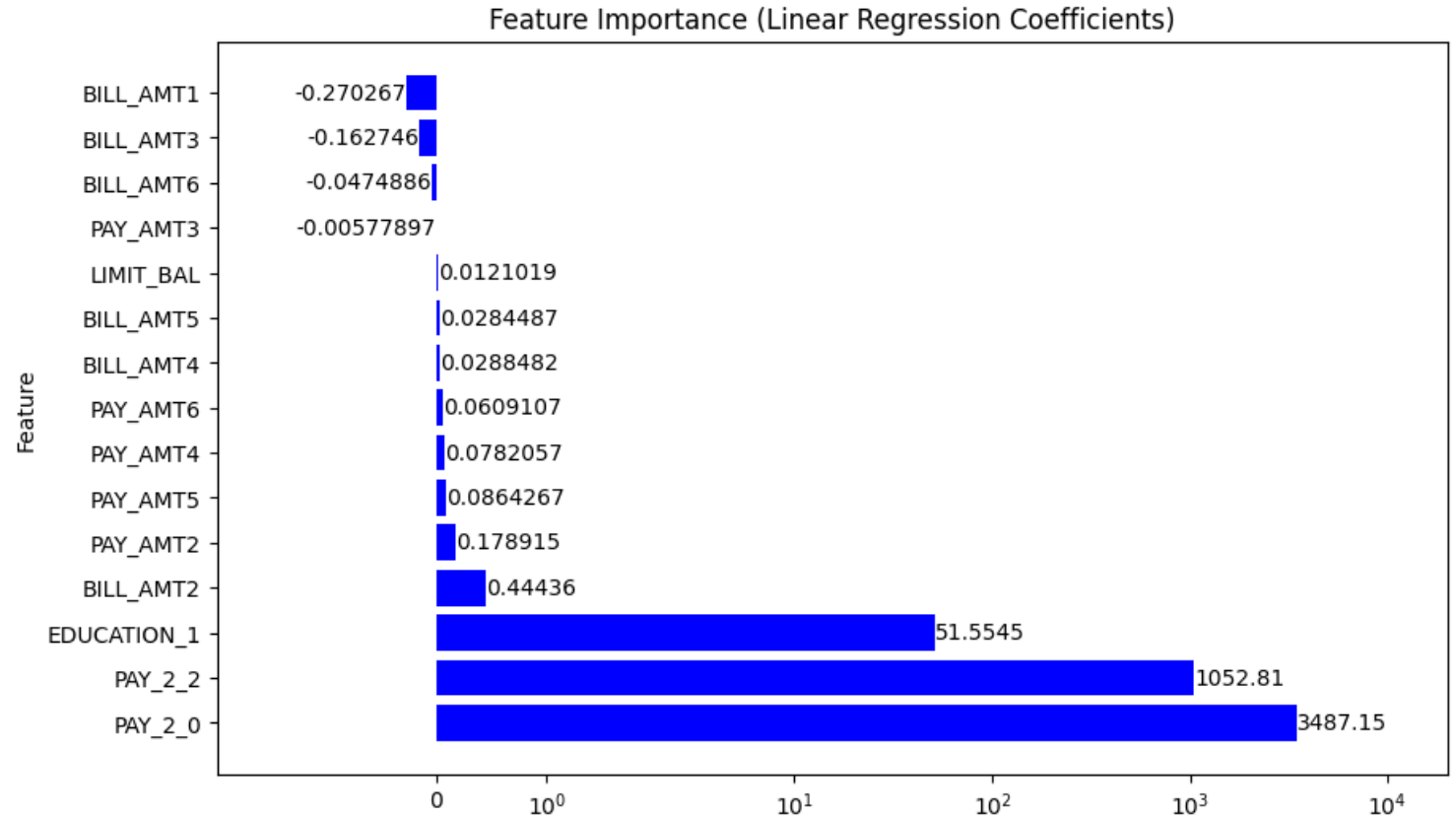
Regresión

- A continuación se presentan 3 modelos regresión en los cuales el objetivo es estimar el valor de las variables PAY_AMT1, APY_AMT2 y PAY_AMT3.
- Se realiza el ajuste y la validación de cada uno de los modelos utilizando la métrica de “ r^2 ”, el cual es un número que nos dice que tan bien un modelo explica los resultados.

PAY_AMT1

- Cada barra muestra las características que se consideran en el modelo, su longitud muestra que tan fuertemente afecta en la predicción.
- Las características con mayor impacto son PAY_2 y PAY_2, los cuales corresponden al status de pago del mes anterior (Agosto).
- La variable EDUCATION_1, También tiene un impacto en el modelo, (_1) corresponde a un nivel de Educación Universidad (Se cambio durante el procesamiento).
- En general, el status de los pagos del mes anterior, tienen un mayor impacto en la predicción.

El modelo tiene un poder predictivo moderado y puede ser útil para la estimación general, pero tiene margen de mejora. Aún puede facilitar la toma de decisiones, especialmente al combinarse con otras herramientas o con el criterio de expertos.



R-squared: 0.3630

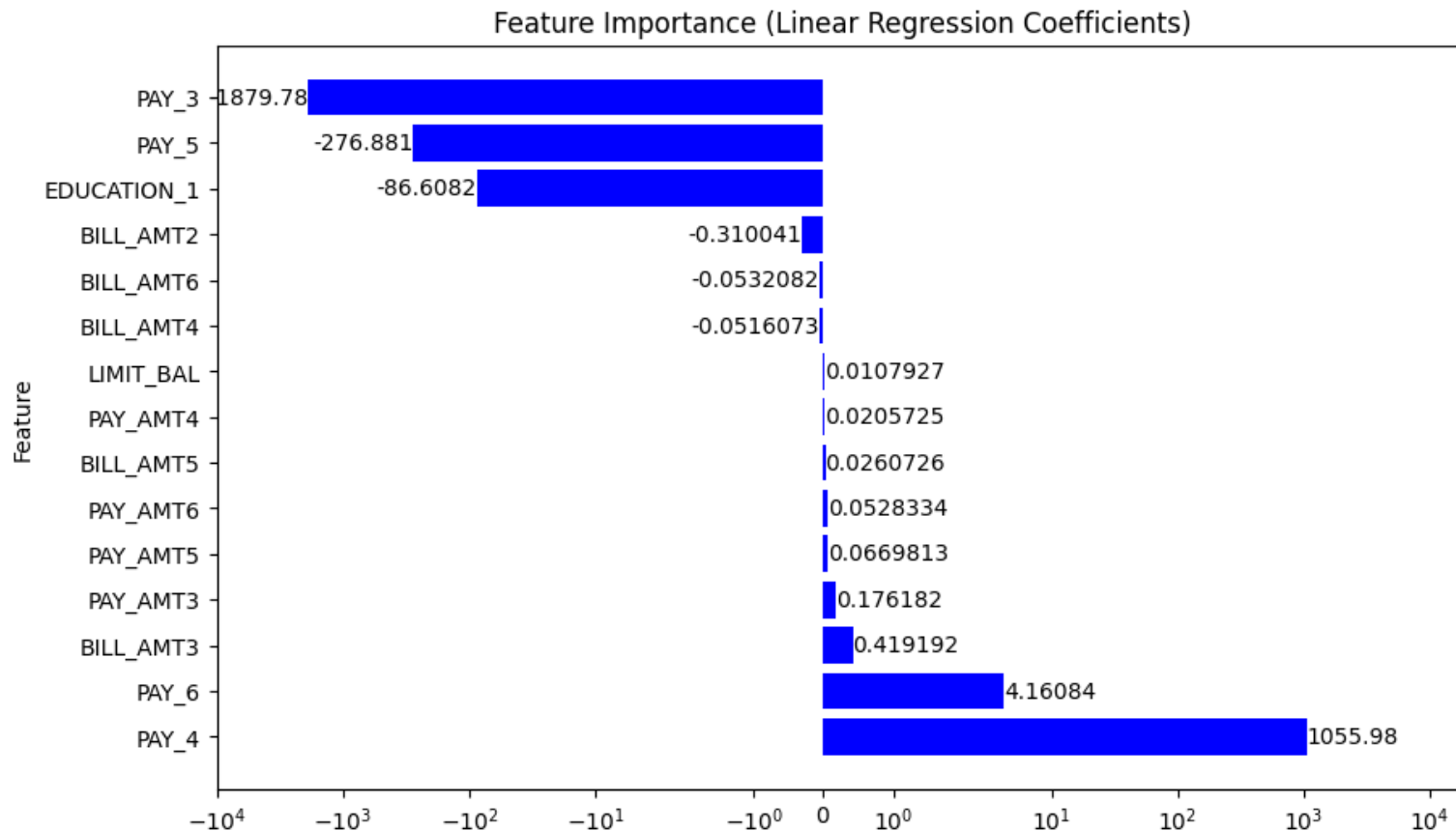
Mean squared error: 178446777.4620

Root mean squared error: 13358.3973

Mean Absolute Error: 5158.2651

PAY_AMT2

Los factores más importantes son el historial de pagos de una persona en meses específicos (especialmente los meses 3 y 4). Factores como el límite de crédito o el importe de las facturas tienen un efecto mínimo, mientras que algunos niveles educativos y el pago mensual tienen una influencia moderada.



R-squared: 0.3958

Mean squared error: 115516212.4118

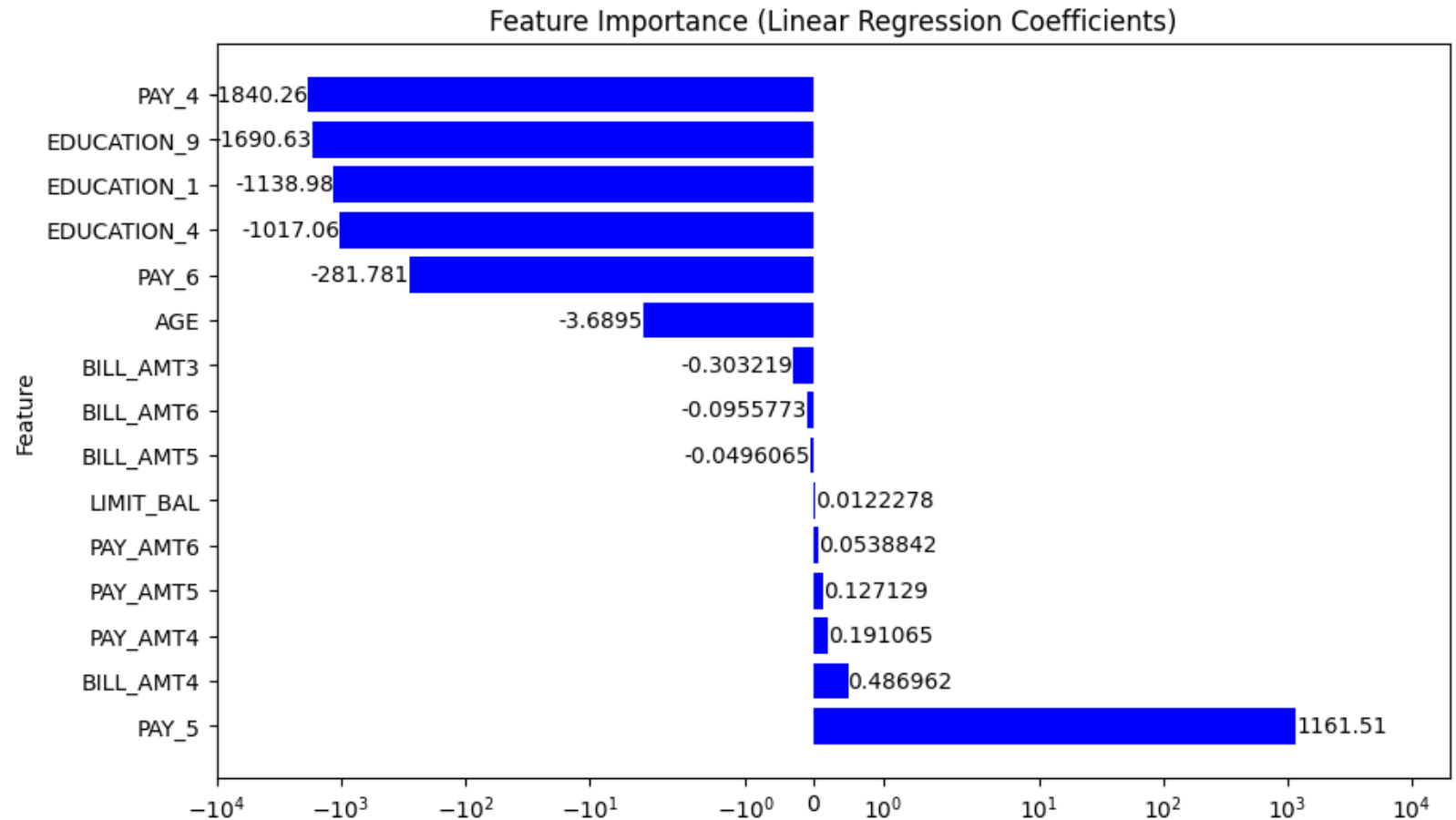
Root mean squared error: 10747.8469

Mean Absolute Error: 4862.3659

PAY_AMT3

El modelo prioriza el nivel educativo y el comportamiento de pago en meses específicos, especialmente los meses 4 y 5.

Las personas con pagos atrasados o con niveles educativos menos favorables experimentan un fuerte impacto negativo, mientras que los pagos puntuales pueden mejorar significativamente el resultado.



R-squared: 0.3995

Mean squared error: 190039186.5633

Root mean squared error: 13785.4701

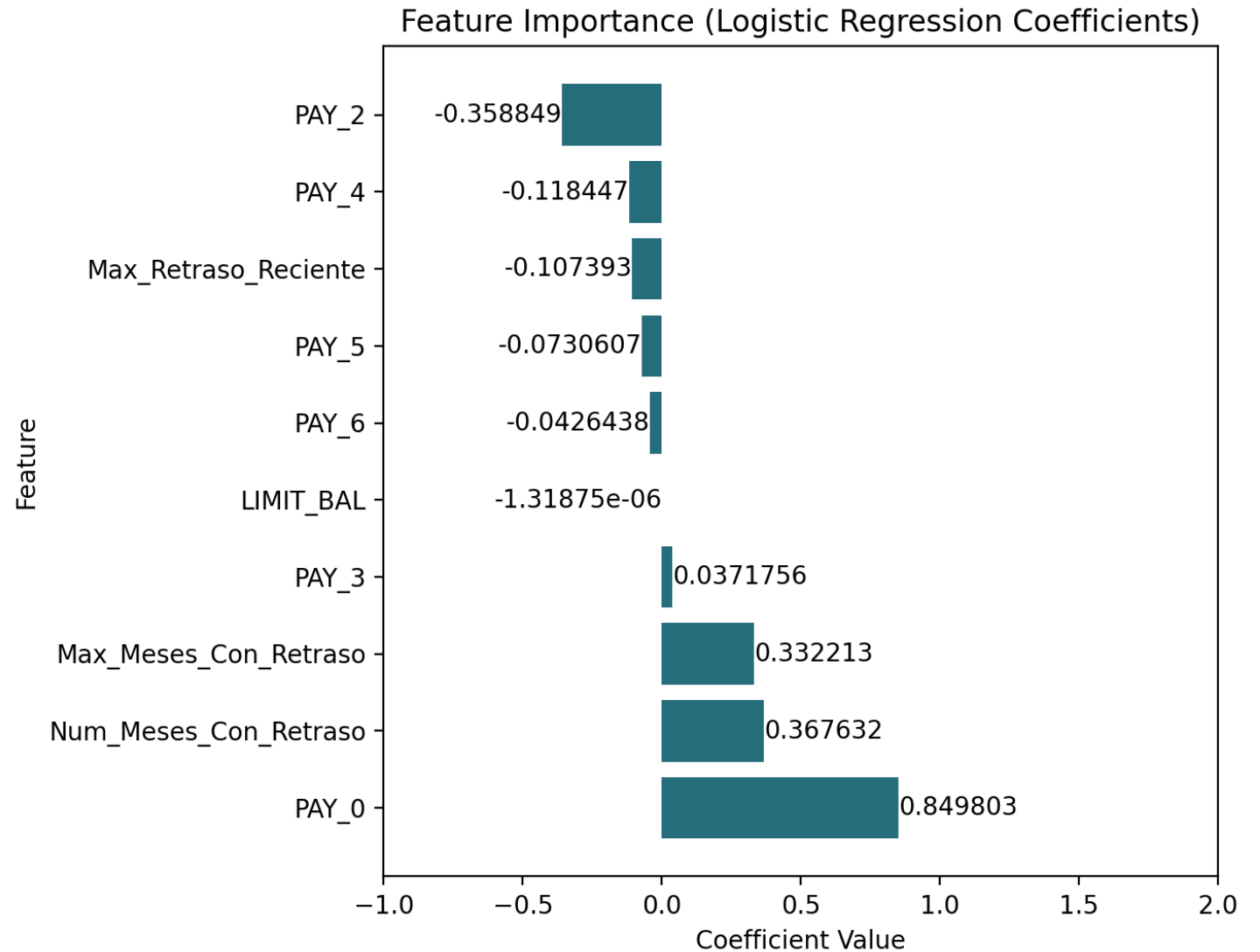
Mean Absolute Error: 5617.8604

DefaultPayment

El modelo se centra principalmente en el comportamiento de pago reciente, especialmente PAY_0 (estado del mes pasado), y en los patrones generales de retraso.

Los retrasos más recientes o frecuentes aumentan el riesgo (Num_Meses_Con_Retraso, Max_Meses_Con_Retraso).

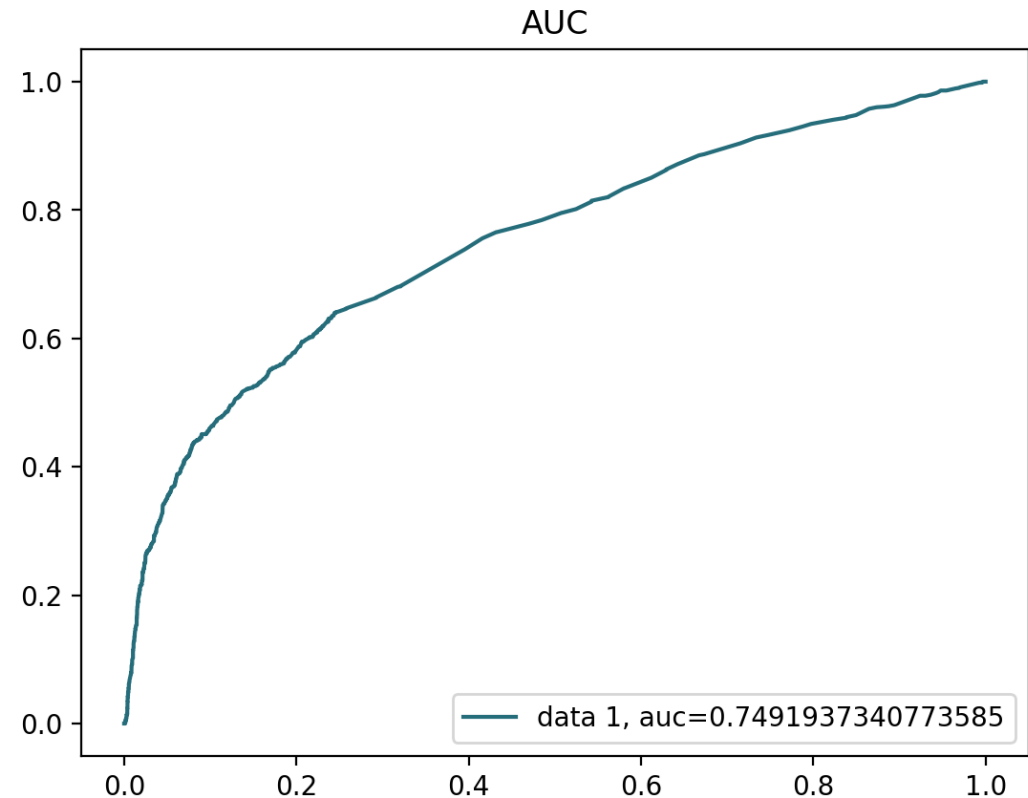
Sorprendentemente, el límite de crédito prácticamente un ligero impacto, y algunos retrasos en meses anteriores (como PAY_2) podrían interpretarse de forma más positiva, lo que podría indicar una recuperación.



DefaultPayment

El modelo puede diferenciar correctamente entre clientes de alto y bajo riesgo aproximadamente el 75 % del tiempo.

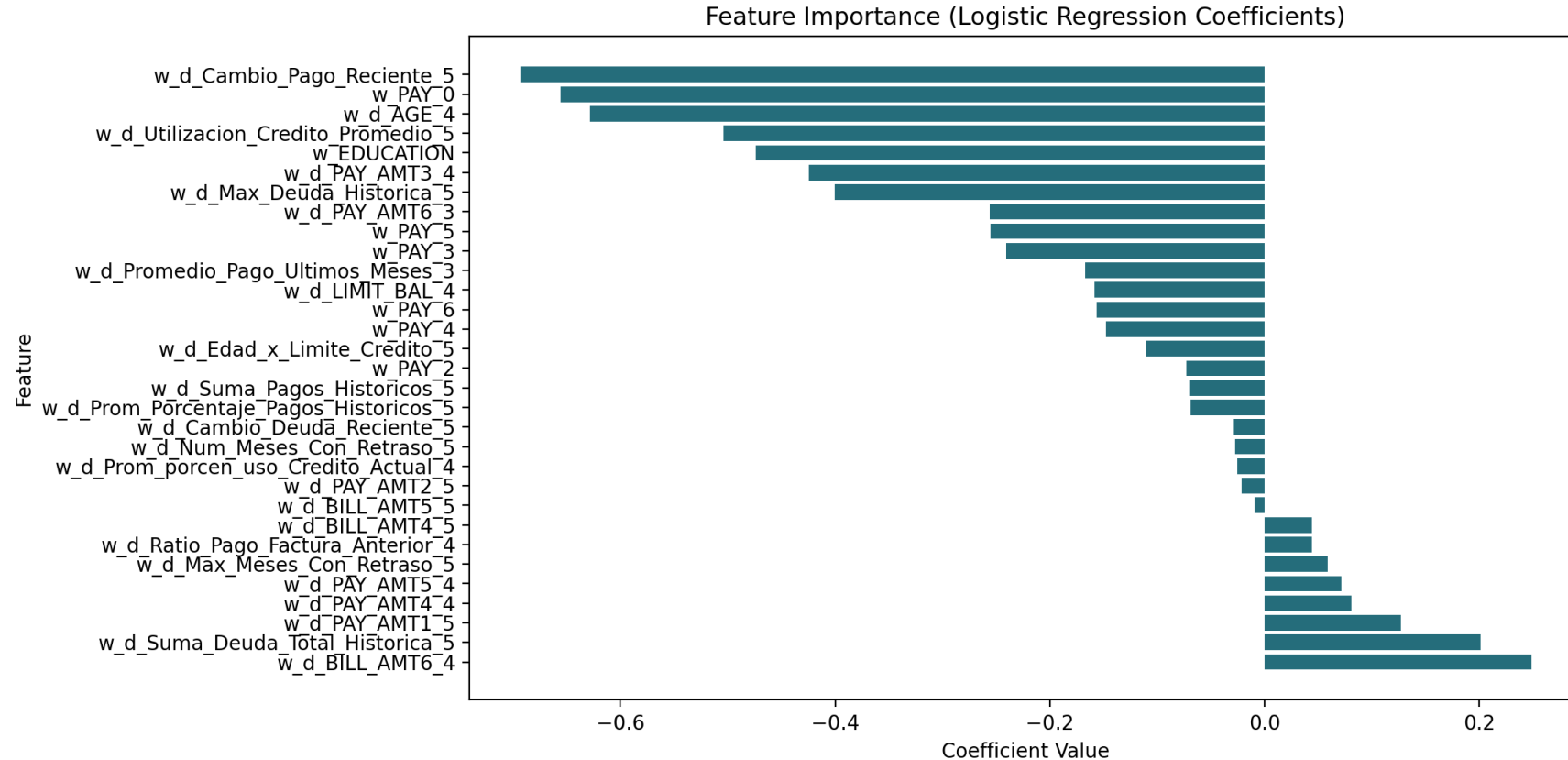
Este nivel de precisión proporciona una base sólida para la toma de decisiones basada en el riesgo, como la aprobación de créditos, la segmentación de clientes o las estrategias de intervención.



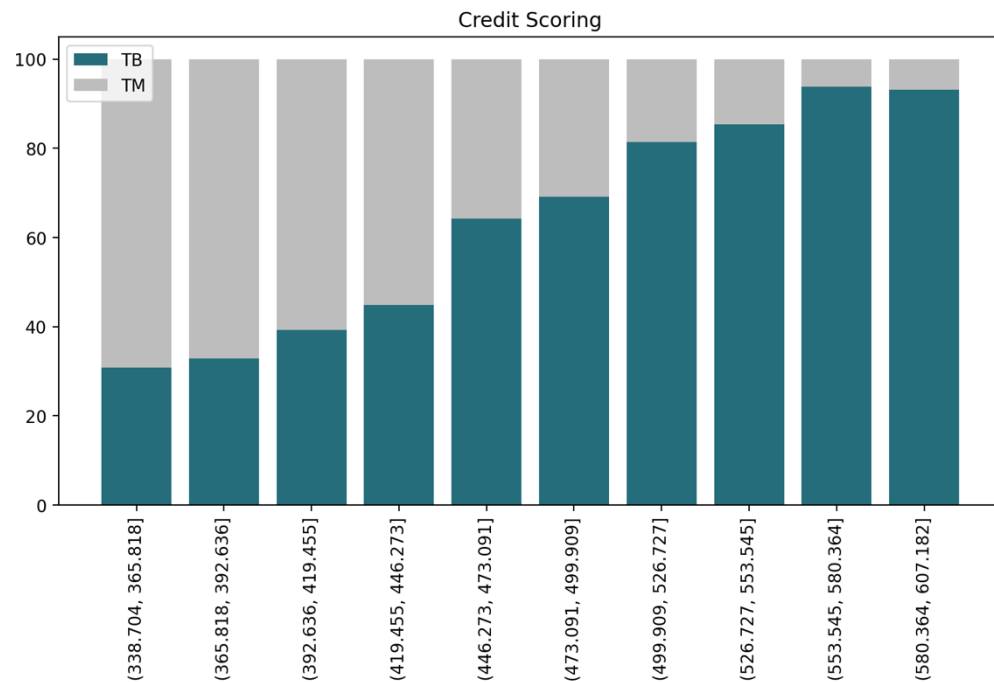
	Model	Accuracy	Precision	Recall	F1 Score	ROC
0	Logistic Regression	0.82	0.71	0.28	0.40	0.62

CreditScoring

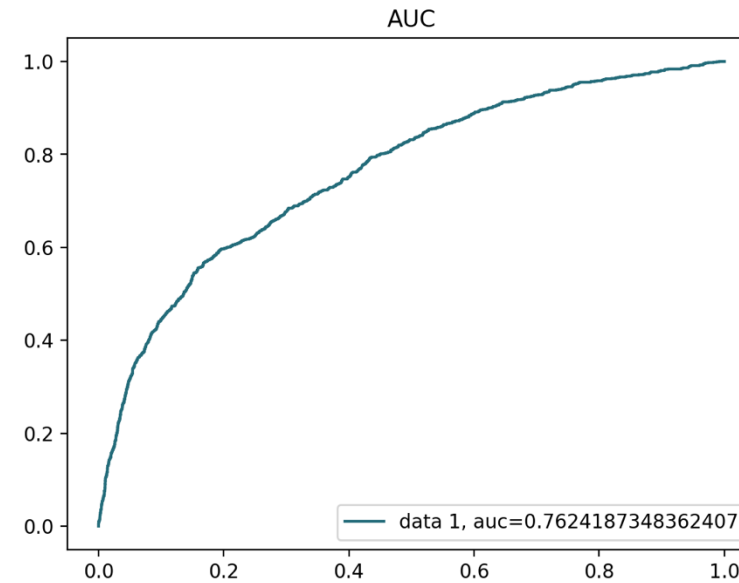
- Este modelo enfatiza el comportamiento de pago reciente, la edad, el uso del crédito y la educación como factores clave.
- Ser mayor, tener educación y mantener un bajo uso del crédito con pagos constantes reduce el riesgo previsto.
- Por el contrario, un alto nivel de deuda pasada o pagos bajos recientes aumentan la predicción del riesgo.



CreditScoring



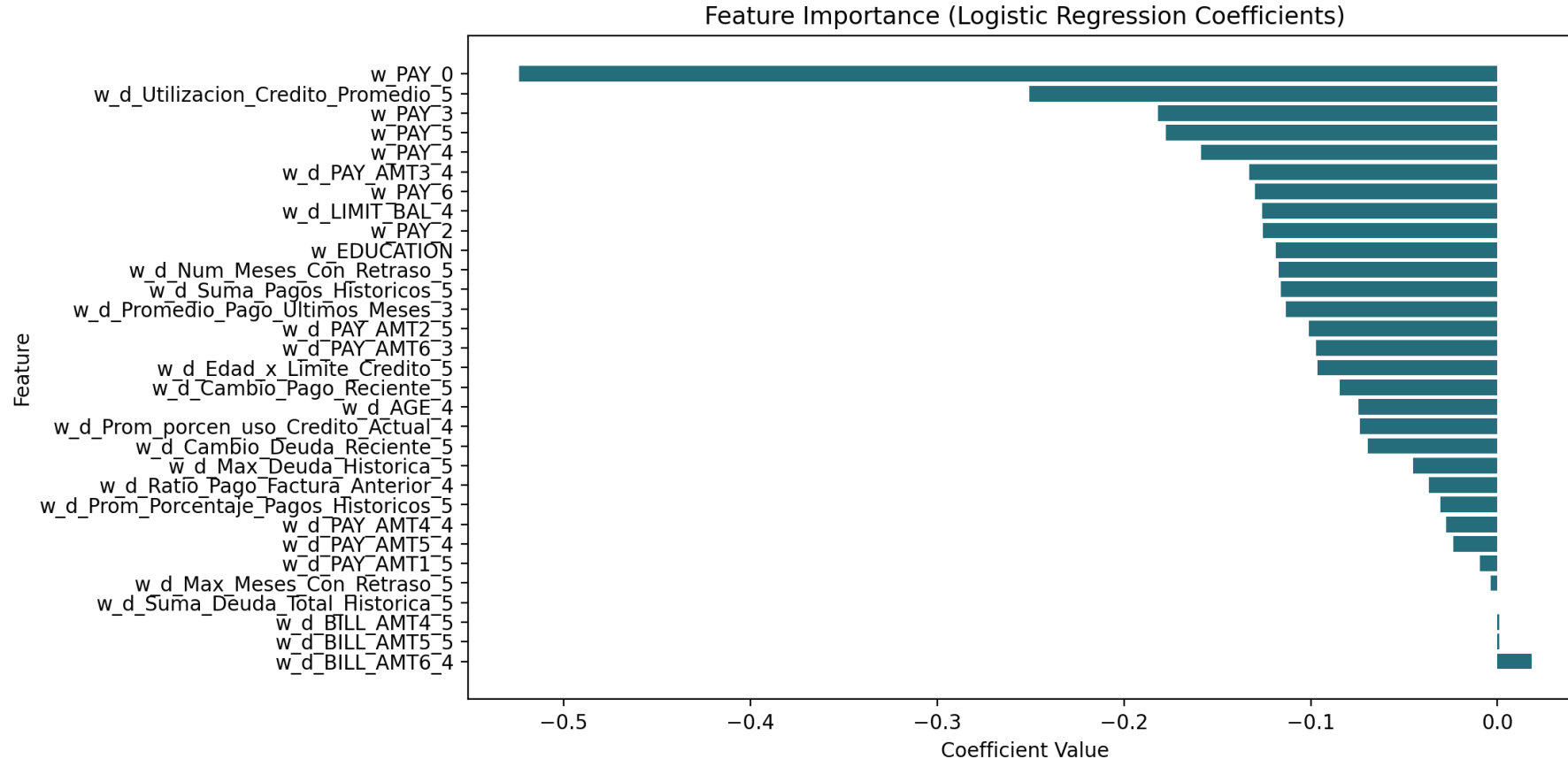
- Un AUC de 0.762 implica que existe un 76.2% de probabilidad de que el modelo asigne una puntuación de riesgo más alta a un cliente que incumplirá sus pagos en comparación con uno que no lo hará. Este nivel de desempeño sugiere que el modelo es fiable para apoyar decisiones crediticias



	Model	Accuracy	Precision	Recall	F1 Score	ROC
0	Logistic Regression	0.81	0.65	0.28	0.39	0.62

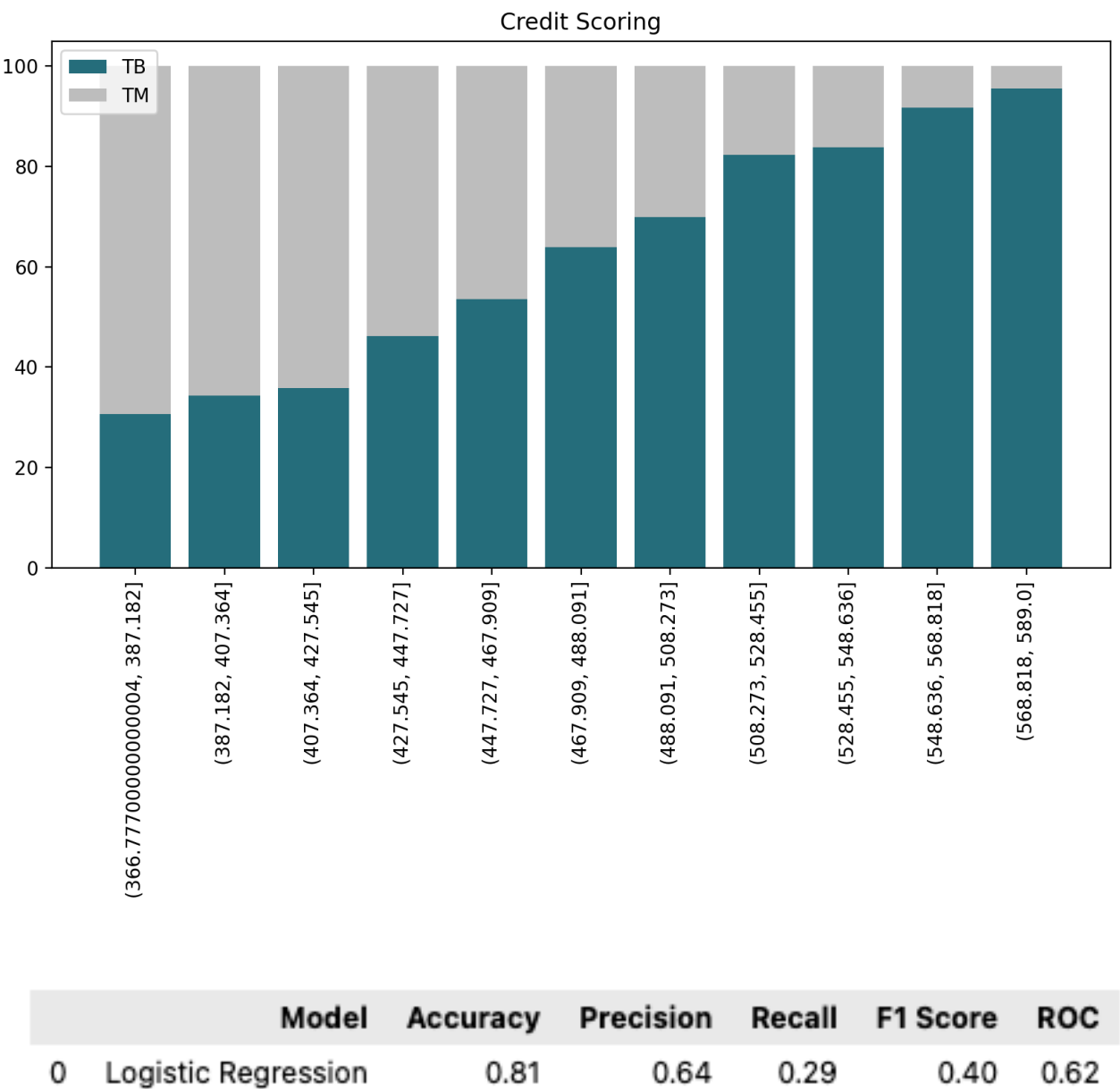
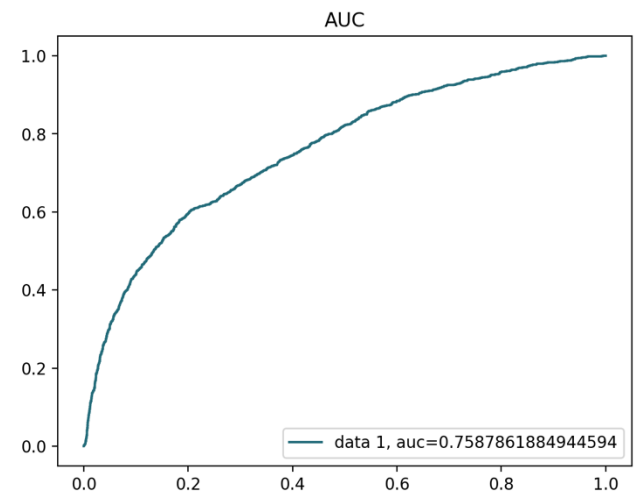
CreditScoringPP

- El modelo prioriza el historial de pagos reciente y el uso responsable del crédito.
- Pagar a tiempo, mantener saldos bajos y usar el crédito con prudencia reduce el riesgo previsto.
- Otros elementos, como la edad, la educación y los patrones de pago más antiguos, influyen en cierta medida, pero no tanto.



CreditScoringPP

Esta distribución sugiere que el modelo de clasificación está funcionando correctamente al asignar mayores puntajes a perfiles con menor riesgo crediticio (TB) y menores puntajes a perfiles con mayor riesgo (TM).



Conclusiones

Se desarrolló y evaluó una serie de modelos de clasificación y regresión para respaldar la calificación crediticia, con el objetivo de predecir el comportamiento de pago de los clientes e identificar segmentos de riesgo.

Se analizaron indicadores clave de rendimiento (KPI), como los coeficientes del modelo, las puntuaciones ROC AUC y la importancia de las características, para evaluar su eficacia.

El aprovechamiento de los modelos de clasificación y regresión en la calificación crediticia permite a las organizaciones tomar decisiones crediticias más inteligentes, rápidas y precisas, al tiempo que gestionan de forma proactiva el riesgo de la cartera y mejoran la rentabilidad.