

Predicción de Default en RD: un enfoque de Machine Learning para la evaluación del riesgo crediticio

Marcos García

Departamento de Estudios Económicos
Superintendencia de Bancos de la República Dominicana

Abstracto

En este estudio, se identificaron las variables que tienen mayor impacto en la predicción de *default* por parte de un deudor en el sistema financiero dominicano. Partiendo de los datos reportados en la Central de Riesgo de la Superintendencia de Bancos de República Dominicana y un criterio de *default*, se seleccionan 39 variables inspiradas por la literatura de predicción de *default*. Luego, se realiza la predicción, utilizando un modelo de aprendizaje automático denominado *catboost*. Para evaluar la interacción entre las variables e identificar la relevancia de cada una, se empleó la técnica de los valores *SHAP*. El resultado obtenido en cuanto al ajuste del modelo, medido mediante el área bajo la curva *ROC* (*ROC score*), fue del 86 % y con un *F1 score* de 50 %. Se aprecia que el poder predictivo disminuye significativamente en el periodo de pandemia correspondiente al COVID-19 (2020-2021), donde el F1 desciende a un valor mínimo de 44 %. Las cinco variables que tuvieron mayor impacto en la predicción de *default* fueron la utilización de tarjetas de crédito, la cantidad de días de atraso en la deuda de tarjetas de crédito, la vejez del crédito más antiguo del deudor, consumo del deudor en tarjetas de crédito y la cantidad de meses desde el último desembolso de un crédito al deudor. Adicionalmente, se observa que durante el periodo de pandemia, variables como el nivel de ingresos del deudor y la vejez de su crédito más antiguo, las cuales suelen tener un alto nivel predictivo, tuvieron menor importancia en la predicción de *default* en este periodo. En la sección de resultados se muestran los detalles en torno a estas y otras variables que inciden en la predicción de *default*.

Palabras clave: *catboost, machine learning, SHAP values, default, riesgo de crédito, República Dominicana*

Índice

1	Introducción	3
2	Revisión de literatura	5
3	Datos y pre-procesamiento	8
3.1	Naturaleza de la data	8
3.2	Variables del Modelo	9
4	Métodos y teoría	11
4.1	Un resumen de Catboost	11
4.2	Un resumen de SHAP Values	12
4.3	Procesamiento en el modelo de machine learning	12
4.4	Optimización de hiperparámetros	13
4.5	Validación cruzada k-fold	14
4.6	Métricas de evaluación	14
5	Resultados	15
6	Conclusión	32
	Referencias	34

1. Introducción

La identificación de los factores que inciden en el comportamiento de un deudor es crucial para minimizar el riesgo de crédito de un acreedor. De manera particular, al acreedor le interesa poder determinar si un potencial deudor podrá honrar su compromiso de deuda y no incurrir en *default*. Saber cuáles son las características que inciden en un potencial *default* permite tomar decisiones inteligentes basadas en la data disponible sobre el deudor. Esto permite otorgar créditos o accionar sobre créditos ya otorgados. Tradicionalmente, las instituciones que otorgan créditos han utilizado modelos estadísticos y ratios financieros para evaluar el riesgo de crédito. Estos modelos son altamente interpretables y por consecuencia aptos para un mercado altamente regulado. Al contraste, históricamente los modelos de aprendizaje automatizado (*Machine Learning*; ML por sus siglas en inglés) tienen la limitante que, aunque son superiores en poder predictivo, por su naturaleza son difíciles de interpretar. Sin embargo, las áreas de ML han generado técnicas, como la biblioteca SHapley Additive exPlanations (SHAP), que ahora se utilizan para identificar y analizar los factores de riesgo con mayor precisión. Esto permite tener lo mejor de ambos mundos, alto nivel predictivo e interpretabilidad de las decisiones que toma el modelo.

Los valores SHAP tienen su origen en la teoría de juegos cooperativos (Bussmann, Giudici, y Papenbrock, 2019) y se utilizan para aumentar la transparencia y la interpretabilidad de los modelos de aprendizaje automático. El algoritmo SHAP ayuda a explicar el resultado de los modelos de aprendizaje automático mediante la asignación de puntuaciones de importancia a cada característica o factor de riesgo. Estos puntajes indican cuánto contribuye cada característica al resultado de la predicción de *default*, lo que permite a los prestamistas identificar y priorizar los factores más importantes que afectan el riesgo crediticio.

Es importante tener presente que predecir *default* es una tarea muy compleja. El modelo a utilizar y la interpretación de los detalles son secundarios a la selección de las variables independientes que se utilizan para predecir. Según (“FICO”, 2023), algunas de las variables más importantes que se utilizan actualmente en la industria para predecir *default* son las siguientes :

- Calificación crediticia del deudor
- Relación deuda-ingreso (DTI, por sus siglas en inglés)
- Relación préstamo-valor (LTV, por sus siglas en inglés)
- Historial de empleo
- Historial de pago
- Antigüedad del historial crediticio
- Tipo de crédito
- Movimientos en cuentas bancarias
- Ahorros y certificados

Estas variables son fundamentales en la predicción de incumplimientos financieros. La calificación crediticia del deudor sirve como indicador principal de la capacidad de pago y solvencia. La relación deuda-ingreso (DTI) y la relación préstamo-valor (LTV) ofrecen perspectivas cruciales sobre la salud financiera del prestatario, evaluando la proporción entre sus deudas y ingresos, así como el valor del préstamo en relación con el valor del activo respaldado. El historial de empleo y de pago proporciona insights sobre la estabilidad laboral y la responsabilidad financiera del individuo. La antigüedad del historial crediticio y el tipo de crédito ofrecen una visión de la experiencia financiera y del comportamiento crediticio pasado. Además, examinar los movimientos en cuentas bancarias y la gestión de ahorros y certificados brinda información adicional sobre la gestión financiera. En conjunto, estas variables crean un perfil holístico que permite a las instituciones financieras evaluar de manera más precisa el riesgo crediticio y prever la probabilidad de incumplimiento.

En este contexto, esta investigación tiene como objetivo identificar los factores determinantes de riesgo de un deudor mediante el algoritmo de SHAP. Estos factores son clave para la predicción de *default*. Los factores o variables se ordenan de mayor a menor importancia y se cuantifica su impacto en la probabilidad de *default* que se predice.

2. Revisión de literatura

Un ejemplo de una investigación que usa algoritmos de aprendizaje automático en el desarrollo de modelos de riesgo de crédito es (Khandani, Kim, y Lo, 2010). Los autores comparan varios algoritmos de aprendizaje automático, incluida la regresión logística, los árboles de decisión, las redes neuronales y las máquinas de vectores de soporte, con los modelos tradicionales de calificación crediticia, como la calificación FICO. Descubrieron que los algoritmos de aprendizaje automático superaron a los modelos tradicionales de calificación crediticia en términos de precisión predictiva. Sin embargo, también notaron que los modelos de aprendizaje automático pueden ser más complejos y difíciles de interpretar, lo que puede ser un desafío para los reguladores y prestamistas que buscan comprender los factores que impulsan las decisiones crediticias. Esto podría haberse resuelto utilizando *Shapley values*.

En un estudio reciente realizado por (Schmitt, 2022), se compararon diversos modelos de redes neuronales con modelos de ensamblaje de árboles. Los resultados indicaron que, en el contexto del estudio, los modelos de ensamblaje de árboles obtuvieron mejores resultados que los modelos de redes neuronales, aunque también explican que esto puede ser debido a la limitante cantidad de datos que utilizaron y que con una base más amplia tal vez los resultados sean opuestos.

El documento de trabajo (Alonso y Carbo, 2021) desarrollado por el Banco de España extiende la muestra y también verifica que el algoritmo con mejor rendimiento fue de nuevo el *xgboost*. Su investigación también demuestra que utilizar modelos de *default* puede ahorrar a las entidades entre 12.4% y 17% en requerimientos de capital regulatorio.

En el contexto dominicano, (Bournigal y García, 2022) verifica que los modelos de árboles dan mejores resultados, específicamente el de *xgboost*, por lo cual para esta investigación se trabajó con una versión de este que maneja las variables categóricas de manera nativa. También se abordan de manera introductoria los valores SHAP, los cuales componen el mayor enfoque de esta investigación.

En (Bussmann y cols., 2019), el autor analiza la importancia de la facilidad de explicación en la gestión del riesgo de crédito y cómo se pueden utilizar las técnicas de aprendizaje automático para mejorar la transparencia y la interpretabilidad de los modelos de riesgo de crédito. El autor presenta un marco para el aprendizaje automático explicable, que incluye cuatro componentes principales: ingeniería de funciones, selección de modelos, explicación

de modelos y validación de modelos. El documento analiza cómo se puede implementar cada uno de estos componentes en el contexto de la gestión del riesgo crediticio y proporciona ejemplos de diferentes técnicas y herramientas que se pueden utilizar, siendo los valores SHAP de particular interés para este trabajo. El documento también destaca algunos de los desafíos y limitaciones del aprendizaje automático explicable en la gestión del riesgo crediticio, incluida la compensación entre precisión e interpretabilidad, la necesidad de experiencia en el dominio y el potencial de sesgos y discriminación en los datos.

La gestión de riesgo de crédito implica la consideración de múltiples variables cruciales que se pueden dividir en cinco categorías fundamentales. Estas categorías comprenden variables de buró de crédito, variables de transacción, variables de gastos, variables de depósito y variables del deudor (Khandani y cols., 2010).

- **VARIABLES DEL BURÓ DE CRÉDITO:** Esta categoría abarca aspectos relacionados con el historial crediticio del deudor y es esencial para evaluar su capacidad de pago y comportamiento financiero. Algunas de las variables clave incluyen la cantidad de líneas de crédito abiertas, el saldo total de la deuda, el número de créditos en atraso y otros indicadores similares. Estas variables proporcionan información valiosa sobre cómo el deudor ha gestionado previamente sus compromisos crediticios.
- **VARIABLES DE TRANSACCIÓN:** Estas variables se centran en el flujo de efectivo del deudor y proporcionan una visión de sus ingresos y gastos a lo largo del tiempo. Esto incluye el análisis de movimientos en cuentas bancarias y otras transacciones financieras que reflejan la capacidad del deudor para generar ingresos y gestionar sus finanzas.
- **VARIABLES DE GASTOS:** La categoría de gastos examina las obligaciones financieras y los gastos regulares del deudor, como vivienda, servicios públicos, transporte y otros gastos relacionados. Evaluar estas variables es esencial para determinar la capacidad del deudor para hacer frente a nuevos compromisos crediticios en función de su carga financiera existente.
- **VARIABLES DE DEPÓSITO:** Estas variables se relacionan con los activos líquidos y otras inversiones que el deudor posee en cuentas de ahorro, cuentas corrientes, certificados de depósito y otras formas de inversión. Estos activos pueden servir como colaterales en caso de incumplimiento, lo que es vital para calcular la exposición al riesgo en tiempos difíciles.

- Variables del Deudor: Esta categoría se enfoca en las características personales y financieras específicas del deudor, como su historial laboral, ingresos estables, nivel educativo y otras cualidades que pueden influir en su capacidad para cumplir con las obligaciones crediticias.

En conjunto, estas cinco categorías de variables proporcionan una evaluación integral del riesgo de crédito y permiten a las instituciones financieras tomar decisiones informadas sobre la concesión de crédito. Es importante recordar que el análisis de riesgo de crédito es dinámico y requiere una revisión constante para adaptarse a las cambiantes circunstancias financieras de los deudores.

En este estudio, se cuenta únicamente con acceso a un conjunto limitado de variables relacionadas con el buró de crédito y el deudor, por ejemplo el historial de empleo y antigüedad del historial crediticio. Es importante tener en cuenta que algunas variables no se incluirán en la investigación (por ejemplo, ninguna variable de transacción, de gasto o depósitos), lo que podría tener un impacto en el desempeño del modelo. En la **Tabla 1** podemos ver un comparativo entre otras investigaciones de ML y apreciar la diferencia en su poder predictivo.

Tabla 1: Variables Utilizada

Fuente	Modelo	AUC ROC Score
(Vincenzo Moscato, 2021)	Random Forest	71.70 %
(Khandani y cols., 2010)	CART	89.50 %
(Kruppa, Schwarz, Arminger, y Ziegler, 2013)	Random Forest	95.90 %
(Moscatelli, Parlapiano, Narizzano, y Viggiano, 2020)	Random Forest	75.70 %
(Moscatelli y cols., 2020)	Gradient Boosted Trees	75.40 %
(Bao, Lianju, y Yue, 2019)	Random Forest	93.60 %
(Bao y cols., 2019)	Gradient Boosted Trees	92.70 %
(Bussmann y cols., 2019)	Xgboost	93.00 %
(Alonso y Carbo, 2021)	Random Forest	82.60 %
(Alonso y Carbo, 2021)	Xgboost	83.70 %
(Schmitt, 2022)	Gradient Boosted Trees	88.85 %
(Schmitt, 2022)	Deep Learning	94.00 %

3. Datos y pre-procesamiento

La data utilizada para esta investigación proviene de la Central de Riesgos de la Superintendencia de Bancos de la República Dominicana (SB). Esta información se obtiene de manera mensual de todas las entidades de intermediación financiera reguladas por la SB a nivel de crédito individual.

3.1 Naturaleza de la data

La estructura de la data disponible se encuentra descrita en el Manual de Requerimientos de Información (MRI) 2021. Este documento contiene la información de todos los reportes que deben remitir las entidades de intermediación financiera a la Superintendencia de Bancos. Los reportes relevantes para la presente investigación son los siguientes:

- DE11 - Deudores comerciales de la entidad por entidad.
- DE21 - Detalle de créditos comerciales por deudor / cuentas contables.
- DE13 - Créditos de consumo de la entidad por deudor.
- DE23 - Detalle de créditos de consumo por deudor / cuentas contables.
- DE14 - Tarjetas de créditos personales.
- DE24 - Detalle de tarjetas de créditos.
- DE15 - Deudores hipotecarios de la entidad.
- DE25 - Detalle de créditos hipotecarios por deudor / cuentas contables.

Adicional a la Central de Riesgos se utilizan variables de la Tesorería de la Seguridad Social de la República Dominicana (TSS), de la cual tenemos acceso a la escala salarial en el que se encuentra un deudor y la actividad económica a la cual pertenece. De estos reportes se obtienen todas las variables utilizadas en el análisis, las cuales se detallan en la **Tabla 2**.

Se obtienen todos los deudores únicos y el objetivo es predecir si el deudor se encuentra en *default* dentro de los próximos 12 meses. En el actual Reglamento de Evaluación de Activos (REA) no existe una definición de cuándo un deudor se encuentra en *default*. Por lo tanto,

como decisión del investigador para los fines de este ejercicio, se procede a definir *default* de la siguiente forma.

Un deudor entrará en *default* al cumplir, por lo menos, con una de las siguientes condiciones:

- Que el deudor exceda en uno de sus créditos los 90 días de atraso.
- Que el deudor entre a condición de reestructurado o reestructurado temporal, según lo define el REA.

3.2 Variables del Modelo

La **Tabla 2** describe las variables que se identificaron como las más significativas para predecir *default*. Estas variables tienen como objetivo capturar la progresión en el deterioro de un préstamo. Por ejemplo, la variable *atraso_max_ultimo_ano_tc* mide si de las tarjetas de crédito de un deudor, alguna muestra indicios de dificultad de pago en el último año. De manera similar, la variable *atraso_tarjeta* ayuda a identificar si un deudor presenta dificultad en alguna de sus tarjetas al momento de la evaluación. La variable *vejez_deudor* busca evaluar la madurez del deudor.

En la **Tabla 2** se especifica la unidad de medida de las variables. Es relevante mencionar que más adelante en el análisis, el investigador optó por discretizar estas variables en intervalos específicos seleccionados por el algoritmo *catboost*. Esta elección fue realizada con el propósito de mejorar la comprensión de los valores SHAP asociados a cada una de las variables.

Tabla 2: Variables Utilizadas

Nombre	Definición	¿Qué Mide?	Unidad
utilizacion_tc	Es el ratio entre la deuda y el límite de la tarjeta	Deterioro debido a la sobre utilización de tarjeta	Porcentaje
atraso_max_ultimo_ano_tc	Días de atraso máximo del deudor en tarjetas de crédito en el último año	Deterioro debido a incapacidad de pago	Día
vejez_deudor	Es el tiempo transcurrido desde el desembolso del crédito activo más viejo	Madurez del deudor con facilidades crediticias	Meses
consumo_mes_tc	Consumo total del deudor en tarjetas de crédito	Gastos del deudor	Meses
meses_ultimo_desembolso	Tiempo desde el último desembolso	Deterioro debido a la utilización de nuevos créditos	Meses
entidad	La entidad del deudor (si tiene más de una se le asigna Entidad Múltiple)	Deterioro debido a la entidad	Meses
atraso_consumo	Días de atraso máximo del deudor en crédito de consumo	Deterioro debido a incapacidad de pago	Día
atraso_tarjeta	Días de atraso máximo del deudor en tarjetas de crédito	Deterioro debido a incapacidad de pago	Día
genero	Género	Riesgo asociado al género	Categoría
atraso_max_ultimo_ano_consumo	Días de atraso máximo del deudor en crédito de consumo en el último año	Deterioro debido a incapacidad de pago	Día
actividad	Es la actividad económica según reportado en la TSS	Riesgo asociado a la actividad económica	Categoría
Ingreso	Es el nivel de ingresos del deudor según la escala reportada por la TSS	Nivel económico del deudor	Pesos
entidad_se_finanza	Cantidad de entidades donde el deudor se financia con tarjetas	Deterioro debido a la utilización de nuevos créditos	Cantidad
deuda_tc	La deuda del deudor en tarjeta de crédito	Deterioro debido a la utilización de tarjeta	Pesos
limite_tc	El límite del deudor en tarjeta de crédito	Riesgo asociado a la capacidad de pago	Pesos
atraso_comercial	Días de atraso máximo del deudor en crédito comercial	Deterioro debido a incapacidad de pago	Día
meses_reportado_tss	Cantidad de meses el deudor fue reportado en TSS en el último año	Estabilidad laboral	Cantidad
atraso_max_ultimo_ano_comercial	Días de atraso máximo del deudor en crédito comercial en el último año	Deterioro debido a incapacidad de pago	Día
atraso_max_ultimo_ano_lineas	Días de atraso máximo del deudor en líneas en el último año	Deterioro debido a incapacidad de pago	Día
deuda_lineas	La deuda del deudor en líneas	Deterioro debido a la utilización de líneas	Pesos
tarjetas_desembolsada_ultimo_ano	Cantidad de tarjetas de crédito adquiridas en el último año	Deterioro debido a la utilización de nuevos créditos	Cantidad
cuota_comercial	Total de cuotas del deudor dirigidas a créditos comerciales	Riesgo asociado a una alta cuota	Pesos
atraso_lineas	Días de atraso máximo del deudor en líneas	Deterioro debido a incapacidad de pago	Día
creditos_desembolsada_ultimo_ano	Cantidad de desembolso de créditos en el último año (excluye TC)	Deterioro debido a la utilización de nuevos créditos	Cantidad
cuota_consumo	Total de cuotas del deudor dirigidas a créditos de consumo	Riesgo asociado a una alta cuota	Pesos
atraso_vehiculo	Días de atraso máximo del deudor en créditos de vehículo	Deterioro debido a incapacidad de pago	Día
atraso_max_ultimo_ano_hipotecario	Días de atraso máximo del deudor en créditos hipotecario en el último año	Deterioro debido a incapacidad de pago	Día
atraso_hipotecario	Días de atraso máximo del deudor en créditos hipotecarios	Deterioro debido a incapacidad de pago	Día
utilizacion_max_ultimo_ano	Es el ratio entre la deuda y el límite de la tarjeta en el último año	Deterioro debido a la sobre utilización de deuda	Porcentaje
cantidad_entidad	Cantidad de entidades a las que pertenece el deudor	Riesgo asociado a tener créditos en múltiples entidades	Cantidad
utilizacion_lineas	Es el ratio entre la deuda y el límite de línea	Deterioro debido a la sobre utilización de deuda	Porcentaje
desembolso_consumo	El desembolso total del deudor en créditos de consumo	Riesgo asociado a un alto desembolso	Pesos
tipo_cliente	Es el tipo de persona según la Tabla 94.0 "Tipo de Cliente" del MRI	Riesgo asociado al tipo de persona	Categoría
desembolso_comercial	El desembolso total del deudor en créditos comerciales	Riesgo asociado a un alto desembolso	Pesos
atraso_max_ultimo_ano_hipotecario	Días de atraso máximo del deudor en créditos hipotecarios en el último año	Deterioro debido a incapacidad de pago	Día
cuota_extra	Las cuotas extras que no se pagan mensualmente	Riesgo asociado a cuota extra	Pesos
carga_maxima	El ratio entre total de cuotas entre ingresos	Riesgo asociado a carga alta	Porcentaje
ingreso_anual	Son los ingresos anuales del deudor según reportado en la TSS	Nivel económico del deudor	Pesos
deuda_comercial	La deuda del deudor en créditos comerciales	Riesgo asociado a alta deuda en comercial	Pesos

4. Métodos y teoría

Nuestra meta en esta investigación es identificar cuáles son las variables que impactan la probabilidad de caer en *default* y comprender la importancia relativa de las variables en su incidencia sobre la probabilidad de *default*.

4.1 *Un resumen de Catboost*

Catboost es un algoritmo de *machine learning* perteneciente a la categoría de *boosting*. El *boosting* es una técnica que combina múltiples modelos débiles, obteniendo un modelo fuerte. Un modelo débil es aquel que tiene un desempeño apenas mejor que una predicción aleatoria. Con el *boosting*, se busca mejorar el rendimiento de los modelos débiles al combinarlos de manera inteligente. Con el tiempo, se han desarrollado diferentes algoritmos de *boosting*.

Uno de los primeros algoritmos de *boosting* fue AdaBoost, introducido en 1995 por (Freund y Schapire, 1995). Este método entrena iterativamente modelos débiles, por ejemplo, árboles de decisión de un solo nodo, y les da más peso a los ejemplos que fueron mal clasificados en cada iteración. El modelo final es una combinación ponderada de los modelos débiles, donde el peso de cada modelo es proporcional a su desempeño en los datos de entrenamiento.

Otro algoritmo de *boosting* es Gradient Boosting, presentado por (Friedman, 2001). Este método ajusta iterativamente árboles de decisión a los errores residuales del árbol anterior. El modelo final es una suma de los árboles individuales.

En 2014, se presentó Extreme Gradient Boosting (*xgboost*) por (Chen y Guestrin, 2016), que es una variante de Gradient Boosting que utiliza modelos de árboles regularizados y una implementación más eficiente. *Xgboost* es conocido por su escalabilidad y ha sido utilizado en muchas aplicaciones, incluyendo a los ganadores en varios concursos de Kaggle.

Catboost, presentado por investigadores de Yandex (Prokhorenkova, Gusev, Vorobev, Dorogush, y Gulin, 2018), es una mejora de *xgboost* en el manejo de variables categóricas como el *boosting* ordenado, que utiliza un orden en las categorías para entrenar árboles de decisión, y la regularización dinámica, que evita el sobre-ajuste al modificar los parámetros de regularización durante el entrenamiento.

Para el modelo utilizamos *catboost*, debido a que este maneja las variables categóricas automáticamente y también valores faltantes, evitando el sobre-ajuste o *overfitting*.

4.2 Un resumen de SHAP Values

Para explicar el modelo e identificar las variables con mayor incidencia en la predicción del *default*, utilizaremos Shapley Additive Explanations (SHAP, por sus siglas en inglés). Es un método para explicar el resultado de cualquier modelo de aprendizaje automático. El objetivo de SHAP es proporcionar explicaciones interpretables para las predicciones individuales realizadas por un modelo.

El marco SHAP se basa en la teoría de juegos, específicamente el valor de Shapley, que es un concepto utilizado para asignar pagos entre los jugadores en un juego cooperativo. En el contexto del aprendizaje automático, los jugadores son las características de entrada y el pago es la predicción del modelo.

Para calcular los valores SHAP para una predicción dada, el algoritmo SHAP considera todos los subconjuntos posibles de características de entrada y calcula la contribución de cada característica a la predicción del modelo. Esta contribución se calcula comparando la predicción del modelo con y sin la característica incluida, y tomando la diferencia promedio de todos los subconjuntos de características posibles.

Los valores SHAP resultantes representan la contribución de cada característica a la predicción para una instancia específica. Estos valores se pueden usar para clasificar la importancia de cada característica para una predicción dada, así como para explicar por qué el modelo hizo una predicción particular para una instancia específica.

Los valores SHAP se pueden visualizar utilizando varios tipos de gráficos, como gráficos de barras o gráficos de dispersión. Estos gráficos permiten a los usuarios interpretar fácilmente las contribuciones de cada característica a una predicción y comprender cómo llegó el modelo a su decisión.

4.3 Procesamiento en el modelo de machine learning

El proceso de preparación de datos implica limpiar, preprocesar y transformar la data en un formato que el algoritmo pueda utilizar. En nuestro caso, hemos discretizado las variables numéricas en intervalos para mejorar la interpretación de los valores SHAP y para

reducir la probabilidad de tener un modelo sobreajustado. En la sección de datos y procesamiento profundizaremos en las variables disponibles, seleccionadas y cómo se transformaron.

La información se recopila de manera anual, con un punto de corte en diciembre. Algunas de las variables se calculan utilizando datos desde enero del mismo año, y el objetivo es prever si el deudor entrará en incumplimiento durante el año siguiente. La data utilizada abarca de enero del 2015 hasta diciembre del 2022. Cabe destacar que cuando en este documento se menciona, por ejemplo, el análisis del 2016, se refiere a que se intenta predecir si los créditos activos a diciembre 2016 entraron en *default* en algún momento del 2017, utilizando la data del 2016 para calcular las variables independientes.

El próximo paso consiste en segmentar los datos en conjuntos de entrenamiento (*train*) y prueba (*test*). El conjunto de entrenamiento se emplea para el proceso de formación del modelo, que incluye la evaluación mediante validación cruzada *k-fold* y la optimización de hiperparámetros. Por otro lado, el conjunto de pruebas se reserva para medir el rendimiento real del modelo.

En la presente investigación hemos considerado 3 variantes del modelo, según la segmentación de los datos. En el primer caso, al cual haremos referencia como la **Variante 1**, tomamos la data de todos los años disponibles (2015 a 2022) y subdividimos esta en 70 % para el conjunto de entrenamiento y 30 % para el conjunto de prueba. En el segundo caso, con el objetivo de evaluar el cambio de relevancia en el tiempo de las características del deudor sobre su posible *default*, tomamos la data en segmentos bianuales, lo mínimo requerido para un análisis (por ejemplo, construimos un modelo para el periodo 2015-2016, donde se utilizan las características del 2015 para observar si el deudor cae en incumplimiento en el 2016; 70 % de cada conjunto de datos bianuales se utiliza para entrenar y el restante 30 % se utiliza para probar). A este se le denomina **Variante 2**. El último caso, al cual nos referimos como **Variante 3**, consiste en entrenar un modelo con la totalidad de los datos abarcando el periodo 2015-2021, para luego probar con los datos del periodo 2021-2022 (características de entrada al modelo del 2021 para evaluar el *default* durante el 2022).

4.4 Optimización de hiperparámetros

Los modelos de *machine learning* tienen hiperparámetros que se pueden ajustar para mejorar su rendimiento. Ajustar estos hiperparámetros implica probar diferentes valores y seleccionar el mejor conjunto de hiperparámetros que maximicen el puntaje F1 del modelo

bajo evaluación.

4.5 Validación cruzada *k-fold*

La validación cruzada *k-fold* es una técnica utilizada para evaluar el rendimiento de un modelo de aprendizaje automático. Implica dividir los datos en k subconjuntos o pliegues, y luego entrenar el modelo K veces, cada vez usando un pliegue diferente para evaluar y los pliegues restantes para entrenar. Luego, se promedian los resultados de cada ejecución para obtener una estimación más precisa del rendimiento del modelo.

4.6 Métricas de evaluación

Después de entrenar el modelo, evaluamos su rendimiento utilizando métricas como la curva AUC-ROC y F1 Score. Al modelo realizar una predicción donde el objetivo tiene dos posibles estados (en nuestro caso *default* y *no default*), el resultado será una de cuatro condiciones posibles: Verdaderos Positivos (*True Positives, TP*), Falsos Positivos (*False Positives, FP*), Falsos Negativos (*False Negatives, FN*), Verdaderos Negativos (*True Negatives, TN*). Teniendo estas posibilidades en cuenta, se pueden computar las siguientes estadísticas:

- Precisión: $\frac{TP}{TP+FP}$
- Exhaustividad (*Recall*): $\frac{TP}{TP+FN}$

En base a estas se calculan las siguientes medidas que serán las utilizadas para comparar los diferentes modelos:

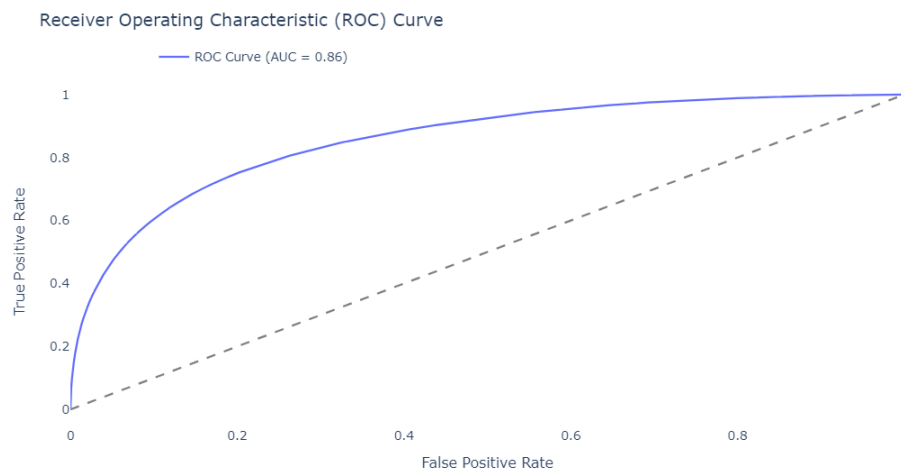
- F_1 score: $2\frac{precision \times recall}{precision + recall}$
- *Receiver Operating Characteristic (ROC) area under the curve (AUC)*: Se grafica el *recall* contra el *false positive rate*, que es el ratio de observaciones negativas que se clasifican como positivas incorrectamente. A partir de esta gráfica se calcula el área debajo de la curva. Mientras más cerca se encuentre esta área de 1, mayor valor predictivo ofrece el modelo.

5. Resultados

En esta sección, discutiremos los resultados de nuestra investigación. En primer lugar, examinaremos brevemente el poder predictivo del modelo, aunque este no es el enfoque principal del artículo. Nuestro principal objetivo es utilizar los valores SHAP para explorar las interacciones entre cada variable y obtener información sobre las variables más importantes y su impacto en el incumplimiento.

Para comenzar, examinamos la curva AUC-ROC, como se muestra en la **Figura 1**, donde la línea punteada representa un modelo *dummy* y la línea azul representa nuestra curva ROC. El área bajo la curva ROC es del 86 %, que es alta pero no perfecta, ya que el 100 % es el límite teórico. En el comparativo de la **Tabla 1** se observa que los resultados se encuentran dentro del promedio.

Figura 1. AUC ROC



En la **Tabla 3** se presenta la matriz de confusión para el modelo de la Variante 1. De la misma se puede calcular un puntaje F1 de 50 %. Es importante destacar que no existe un valor de referencia para determinar qué es un buen puntaje F1. Sin embargo, según prácticas comunes de la industria y los principios de evaluación de modelos, un rango que se considera razonable para un buen modelo es entre 60 % y 80 %, y mayor de 80 % se considera bueno. Un valor de 50 % se encuentra relativamente bajo en comparación a esto. Sin embargo, esta situación era esperada, debido a la limitación de contar únicamente con escasas variables relacionadas con el buró de crédito y la identificación del deudor. Es importante señalar que faltan variables significativas relacionadas con transacciones, gastos y depósitos, como

se mencionó en la Sección 2. Además, se debe tomar en cuenta que hemos tomado medidas para mejorar la interpretabilidad del modelo que afectan el poder predictivo.

Tabla 3: Evaluación de Variante 1

		Valor Real		Total
		No Default	Default	
Predicción	No Default	2,416,745	171,837	2,588,582
	Default	215,218	196,106	411,324
Total		2,631,963	367,943	2,999,906

Para la Variante 3, la cual utiliza el periodo 2021-2022 como nuestro conjunto de prueba, observamos en la **Tabla 4** que el poder predictivo disminuye de manera notable, con el puntaje F1 reduciéndose a 44%. Esta disminución se debe a una ola de reestructuraciones que ocurrió durante este período, en respuesta a la crisis provocada por la pandemia de COVID-19.

Tabla 4: Evaluación de Variante 3

		Valor Real		Total
		No Default	Default	
Predicción	No Default	1,519,314	60,976	1,580,290
	Default	97,768	62,217	159,985
Total		1,617,082	123,193	1,740,275

Para continuar analizando la variación en el poder predictivo a lo largo de los años, se evaluó la capacidad del modelo descrito como **Variante 2**. Los resultados de estos modelos con datos bianuales se presentan en la **Tabla 5**. En esta tabla, se destaca que el año con la

Tabla 5: Evaluación de Variante 2

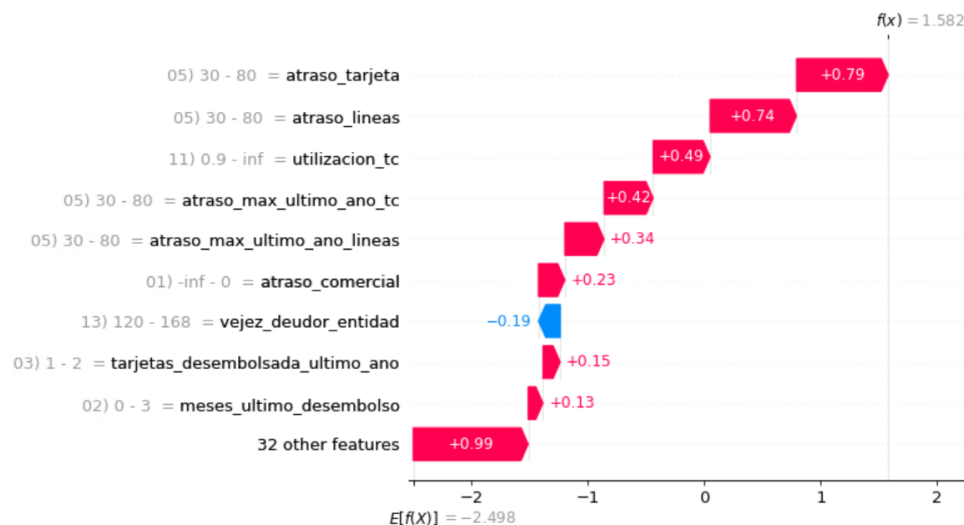
Año	Roc_score	F1_score
2015	86 %	54 %
2016	87 %	56 %
2017	87 %	52 %
2018	86 %	50 %
2019	87 %	57 %
2020	86 %	49 %

mejor capacidad de predicción fue 2019 con puntaje F1 de 57 %, mientras que el año con la capacidad de predicción más baja fue 2020 con F1 de 49 %, donde ya se empezaron a percibir los efectos de la pandemia de COVID-19.

Antes de profundizar en algunos gráficos que en principio muestran agregaciones de valores SHAP, creemos importante explicar claramente qué representa exactamente el valor SHAP. Para ilustrar esto, proporcionaremos un ejemplo a nivel de deudor. Específicamente, examinaremos cuatro casos que representan los cuatro resultados posibles al predecir *default*: verdadero positivo en la **Figura 2** (cuando predecimos correctamente *default*), verdadero negativo en la **Figura 3** (cuando predecimos correctamente el *no default*), falso positivo en la **Figura 4** (cuando predecimos incorrectamente *default*) y falso negativo en la **Figura 5** (cuando predecimos incorrectamente la condición de *no default*).

Podemos observar en la **Figura 2** que existe un valor base, que en nuestro caso es -2.498. A cada variable se le asigna un peso que puede mover este valor base, ya sea hacia el lado negativo (que indica que no hay incumplimiento) o hacia el lado positivo (que indica que hay incumplimiento). Cuando el valor base llega a 0, el modelo predice incumplimiento, y si permanece negativo, el modelo predice no incumplimiento.

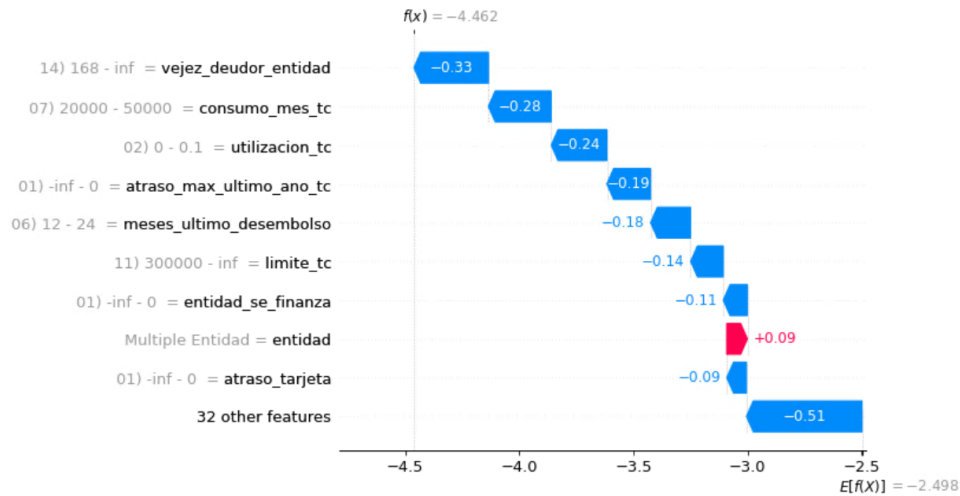
Figura 2. Ejemplo Verdadero Positivo



Como se puede apreciar en la **Figura 2**, la variable *atraso_tarjeta* actualmente presenta una morosidad que se encuentra entre los 30 y 80 días, lo que contribuye con un valor de +0.79 al valor base. En conjunto con las demás variables, el valor total resulta en 1.582,

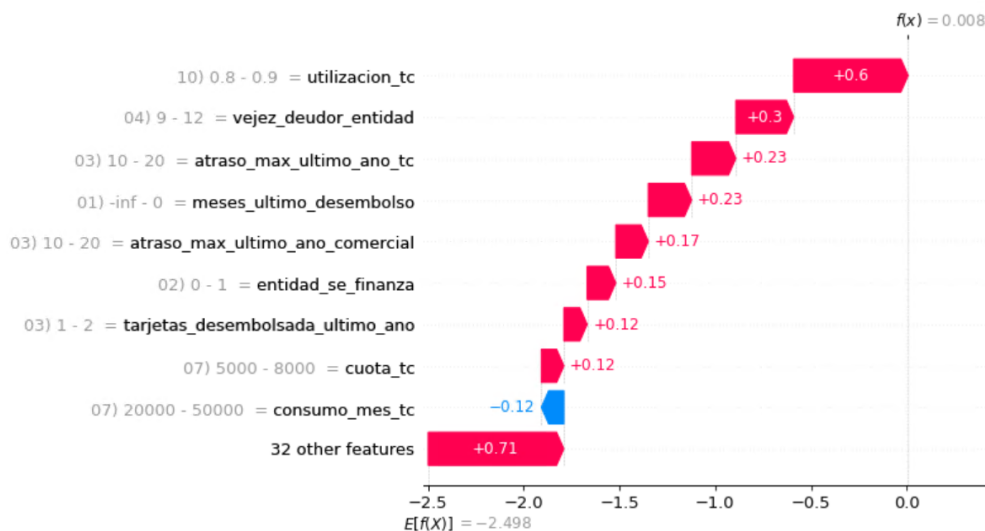
el cual es positivo. Por consiguiente, de acuerdo con el modelo, se predice la condición de incumplimiento para este deudor, y este pronóstico coincide con el resultado observado en la realidad.

Figura 3. Ejemplo Verdadero Negativo



En la **Figura 3**, se observa que el deudor cumple con las siguientes condiciones: tiene más de 168 meses como deudor, su consumo_mes_tc se encuentra entre 20,000 y 50,000, su tasa de utilización de tarjeta de crédito (utilizacion_tc) se encuentra en el rango de 0 a 0.1, lo cual implica una utilización menor al 10%, y no tiene ningún atraso en el último año en tarjetas de crédito. El modelo realiza una predicción de no incumplimiento en este caso, lo cual coincide con el resultado real.

Figura 4. Ejemplo Falso Positivo

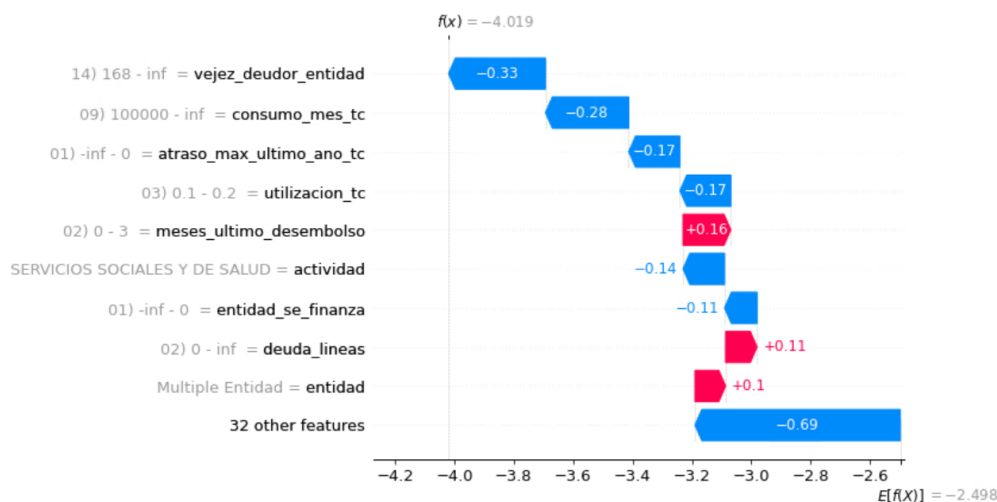


En la **Figura 4** se presenta un ejemplo en el que un crédito cumple con las siguientes condiciones: tiene una tasa de utilización de tarjeta de crédito (`utilizacion_tc`) en el rango de 0.8 a 0.9, una antigüedad de 9 a 12 meses y ha experimentado días de atraso en sus productos de tarjetas y créditos de consumo. Sin embargo, el valor Shap termina en 0.008, lo que implica que el modelo predice incumplimiento. No obstante, en la realidad, el resultado fue diferente, ya que el cliente no incurrió en incumplimiento.

Este escenario ejemplifica una situación en la que el modelo tenía todas las condiciones típicas de incumplimiento, pero factores externos o inesperados llevaron a un resultado distinto en la vida real. Esto destaca la importancia de considerar no solo los datos históricos y las variables del modelo, sino también variables de cuentas de ahorros y las circunstancias individuales y los eventos imprevistos que pueden influir en el comportamiento de los deudores.

En la **Figura 5**, se presenta un ejemplo de un falso negativo en el que las variables indican que el deudor tiene características que normalmente apuntarían al no incumplimiento. Estas características incluyen una alta madurez crediticia del deudor, un alto consumo y la ausencia de cualquier tipo de atraso en los pagos. Sin embargo, a pesar de estas señales positivas, el deudor efectivamente termina incurriendo en incumplimiento.

Figura 5. Ejemplo Falso Negativo



Este ejemplo ilustra una situación en la que el modelo no logró predecir correctamente el incumplimiento debido a circunstancias excepcionales o factores imprevistos que llevaron al resultado contrario de lo que se esperaba. Esto subraya la importancia de reconocer que, aunque los modelos de riesgo de crédito son herramientas valiosas, no son infalibles y deben considerarse junto con otros factores y análisis adicionales para una toma de decisiones más precisa.

Ahora, exploremos la importancia de cada variable. En el contexto del análisis, dos gráficos resultan particularmente relevantes. En primer lugar, la **Figura 6** muestra la media absoluta del valor SHAP, proporcionando una estimación de la importancia de cada variable al indicar cuánto afecta, en promedio, la predicción de incumplimiento o no incumplimiento. La **Figura 8** revela que la variable `utilizacion_tc` es la más significativa en términos de influencia en la predicción.

La **Figura 9** resalta que, en el rango de 0.4 a 0.5 de utilización de tarjeta de crédito, en lugar de reducir la probabilidad de caer en incumplimiento, esta variable la incrementa. Es evidente que, a medida que aumenta la `utilizacion_tc`, también aumenta la probabilidad de entrar en incumplimiento. Este hallazgo confirma lo que ya se ha documentado en la literatura.

La segunda variable más importante, según la media absoluta del valor SHAP, es `atraso_max_ultimo_ano_tc`. Esta importancia es coherente, ya que la presencia de atrasos

en el pasado sugiere una mayor probabilidad de que continúen en el futuro. La **Figura 10** proporciona una visualización de esta relación, mostrando que tener días de atraso incrementa la probabilidad de entrar en incumplimiento. Sin embargo, se observa que, a partir de 31 días de atraso, el efecto marginal en la probabilidad de incumplimiento es mínimo.

La **Figura 7** utiliza el valor máximo en lugar del promedio, lo que permite identificar casos en los que una variable puede no tener un poder predictivo alto en general, pero sí tener un valor específico que resulte significativo. En el caso de `atraso_lineas`, esta variable no proporciona mucha información cuando su valor es 0, ya que la mayoría de la población cae en esta categoría. Por lo tanto, en promedio, el valor resultante es bajo. Sin embargo, cuando la variable presenta algún día de atraso, su influencia en la predicción se vuelve significativa.

Esta observación resalta la importancia de considerar no solo los promedios, sino también los valores específicos de las variables en el análisis de riesgo de crédito, ya que ciertos valores pueden tener un impacto desproporcionadamente alto en la predicción.

Las figuras **Figura 6** y **Figura 7** muestran el orden de importancia bajo las dos medidas.

Figura 6. Promedio absoluto de valores SHAP

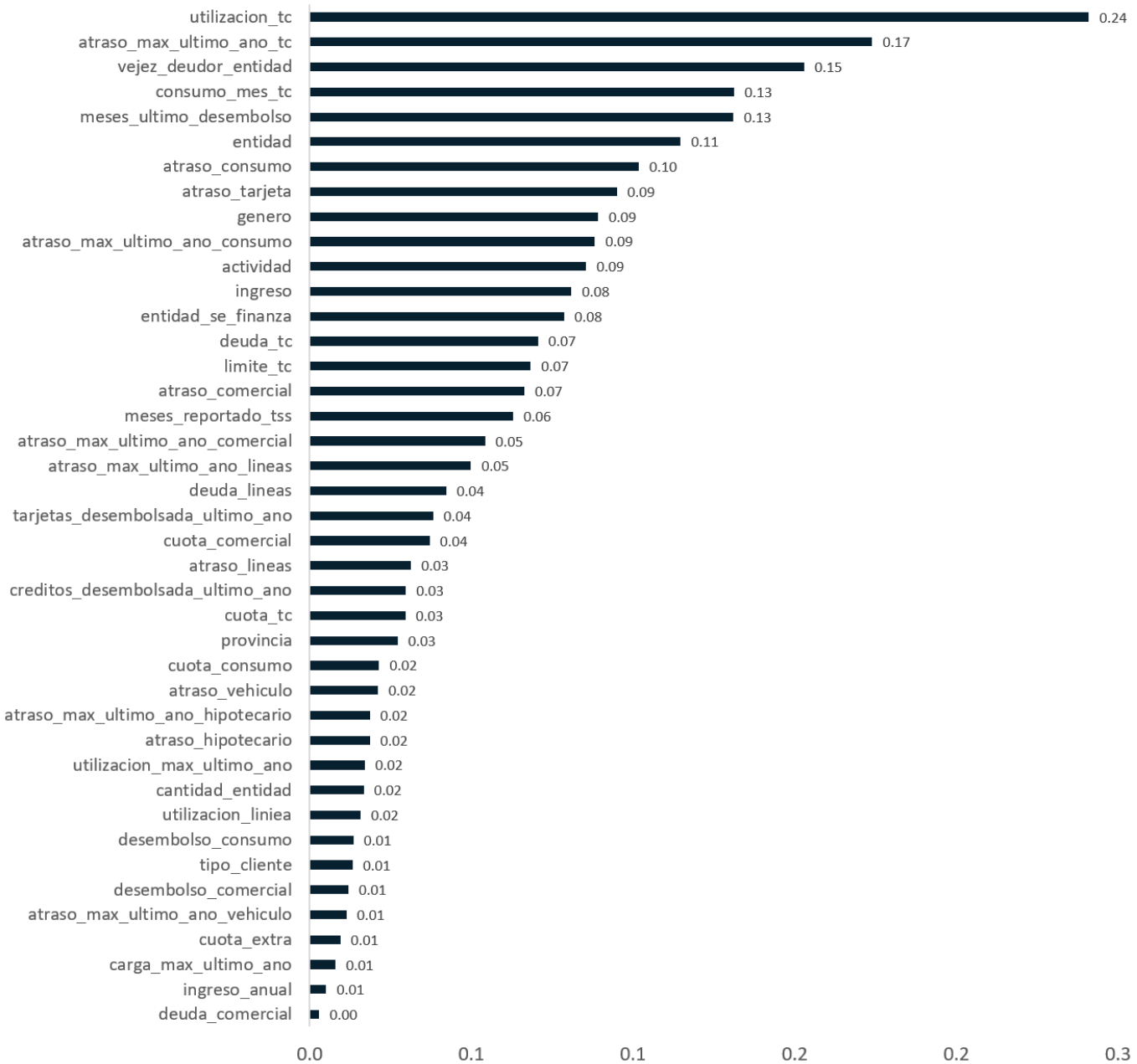
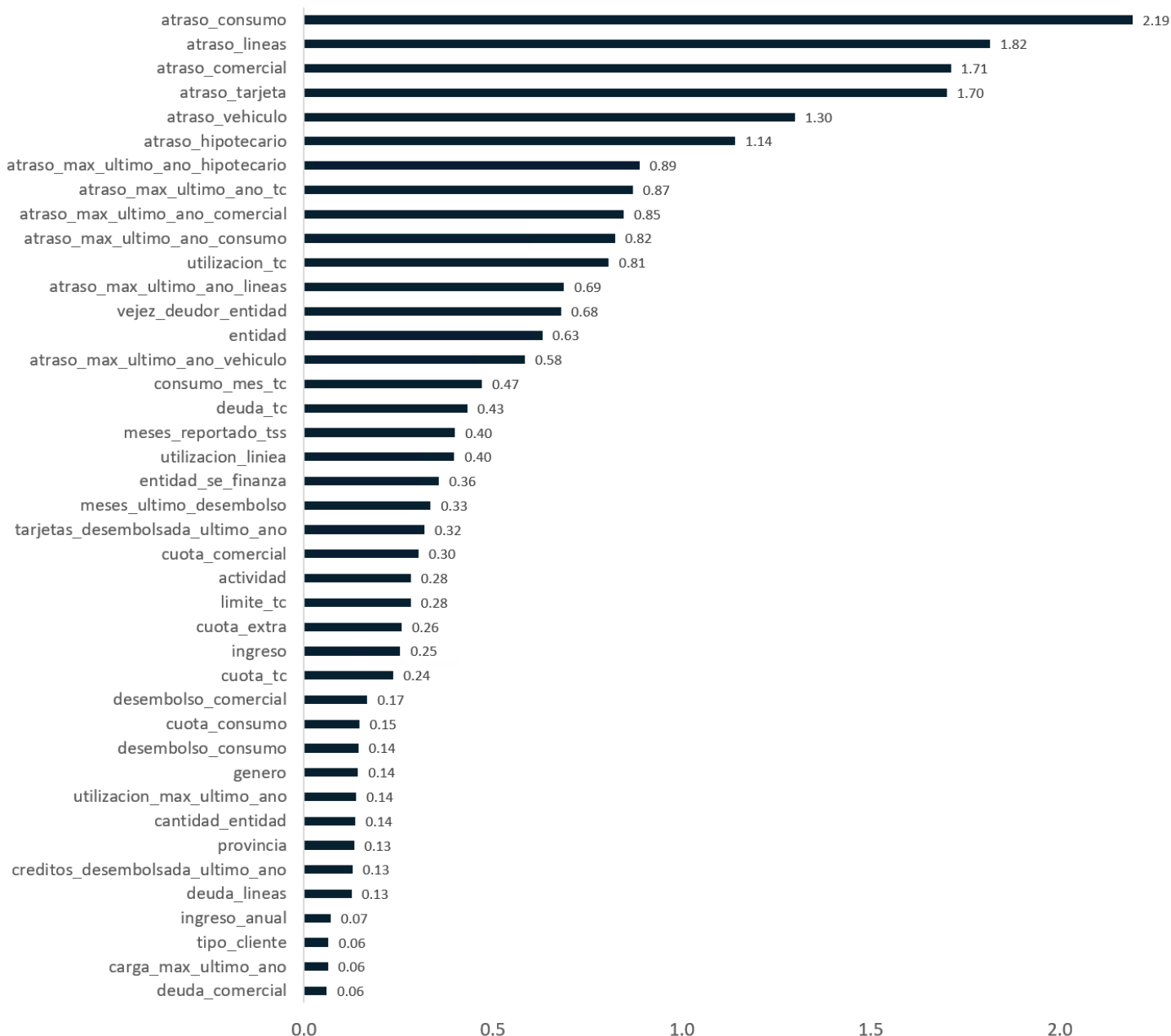


Figura 7. Valores SHAP Máximos



Considerando los modelos de la Variante 2, podemos ver en la **Figura 8** un mapa de calor con las variables de mayor poder predictivo a través de los años. En el mismo se aprecia que `utilizacion_tc` y `atraso_max_ultimo_ano_tc` siempre tienen alto poder predictivo. En el caso del 2020, las variables `vejez_deudor_entidad` e `ingresos`, las cuales suelen tener alto nivel

predictivo, no son tan relevantes para este año. Esto se debe a los efectos de la pandemia del COVID-19. A continuación, veremos el detalle de algunas de las variables más relevantes. El resto puede ser encontrado en el repositorio de Github del Departamento de Estudios Económicos de la Superintendencia de Bancos, mediante el enlace https://github.com/SB-ESTUDIOS/paper_prediccion_default/tree/main

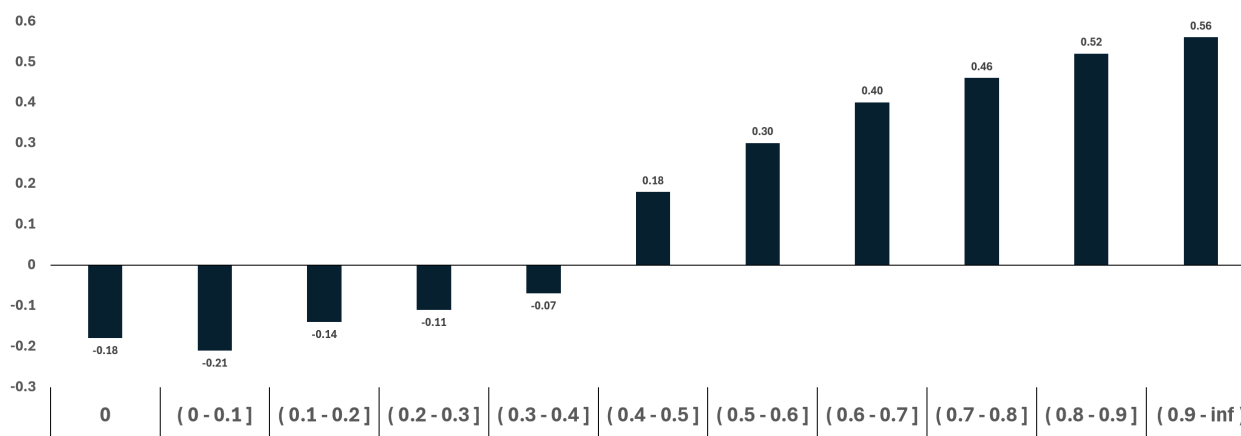
Figura 8. *Ranking* de importancia de variables por año

Feature	2015	2016	2017	2018	2019	2020
utilizacion_tc	1	1	1	1	2	2
atraso_max_ultimo_ano_tc	3	3	2	2	3	1
consumo_mes_tc	7	8	3	3	5	3
entidad	4	4	5	8	1	4
vejez_deudor_entidad	2	2	4	6	4	25
atraso_consumo	8	9	7	5	9	7
genero	11	6	8	4	11	17
ingreso	6	10	6	7	10	23
atraso_tarjeta	9	7	10	11	13	8
atraso_max_ultimo_ano_consumo	13	11	13	15	8	5
meses_ultimo_desembolso	5	5	12	9	17	21
entidad_se_finanza	16	15	16	13	12	12
meses_reportado_tss	14	16	15	12	24	11
atraso_comercial	12	14	14	14	19	13
actividad	18	18	11	10	6	24
atraso_max_ultimo_ano_comercial	10	12	17	20	14	18
tarjetas_desembolsada_ultimo_ano	19	13	9	16	18	26
atraso_max_ultimo_ano_lineas	20	17	18	17	22	14
deuda_lineas	17	19	35	19	26	6
limite_tc	21	21	24	24	16	20
deuda_tc	15	24	25	18	21	19
provincia	26	30	30	23	23	15
creditos_desembolsada_ultimo_ano	33	22	33	28	15	9
utilizacion_max_ultimo_ano	22	32	20	21	25	22
atraso_lineas	23	20	22	26	31	27
cuota_tc	30	26	19	32	28	29
cuota_consumo	28	29	27	35	30	10
cuota_comercial	32	23	23	22	29	28
utilizacion_linea	24	31	31	30	32	30
desembolso_consumo	29	25	28	31	40	16
tipo_cliente	36	27	21	29	27	35
atraso_max_ultimo_ano_hipotecario	25	28	29	27	38	31
atraso_vehiculo	34	33	26	25	34	33
desembolso_comercial	27	39	40	40	7	37
atraso_hipotecario	31	35	32	34	37	34
cantidad_entidad	37	38	34	36	20	36
ingreso_anual	35	34	37	37	39	38
atraso_max_ultimo_ano_vehiculo	39	36	38	33	36	41
carga_max_ultimo_ano	41	41	36	39	35	32
deuda_comercial	38	40	39	41	33	40
cuota_extra	40	37	41	38	41	39

La **Figura 8** destaca que `utilizacion_tc` es la variable más importante en el análisis. Esta variable es un elemento clásico en la literatura de riesgo de crédito y generalmente muestra resultados significativos. Se puede observar en la **Figura 9** que cuando el porcentaje de utilización es inferior a 0.4, la probabilidad de entrar en incumplimiento disminuye. No obstante, si el porcentaje supera el valor de 0.4, la probabilidad de incumplimiento aumenta.

Este hallazgo refuerza la idea de que el nivel de utilización de tarjetas de crédito es un indicador crítico en la evaluación del riesgo crediticio, y su impacto es coherente con lo reportado en la literatura especializada en el campo.

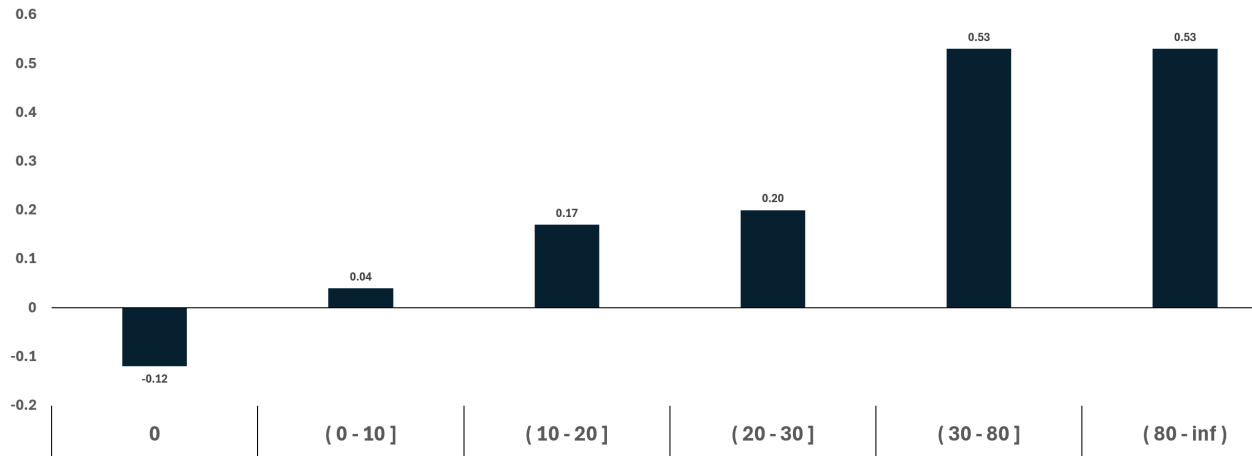
Figura 9. Shap `utilizacion_tc`



La segunda variable más importante, según la media absoluta del valor SHAP, es `atraso_max_ultimo_ano_tc`, la cual se desglosa en la **Figura 10**. Esto sugiere que si un deudor ha mostrado incumplimiento en el pasado, existe una alta probabilidad de que continúe incumpliendo en el futuro. Es interesante observar que tener de 30 a 80 días de atraso contribuye de manera similar a la probabilidad de incumplimiento que tener más de 80 días de atraso.

Este hallazgo indica que, en términos de riesgo crediticio, la presencia de atrasos en los pagos, ya sean de 30 a 80 días o más de 80 días, representa un factor significativo en la evaluación de la probabilidad de incumplimiento.

Figura 10. Valores SHAP de atraso_max_ultimo_ano_tc



La variable `vejez_deudor_entidad` en la entidad, como se muestra en la **Figura 11**, se destaca como una de las variables más importantes en el análisis. Esta variable puede ser de gran interés para las instituciones financieras, ya que indica que a medida que un deudor permanece más tiempo dentro de la entidad, su probabilidad de incumplimiento tiende a disminuir, y hay un punto donde el valor de esta variable revierte el efecto de la misma sobre la predicción de *default* del deudor.

Figura 11. Valores SHAP de vejez_deudor_entidad_shap

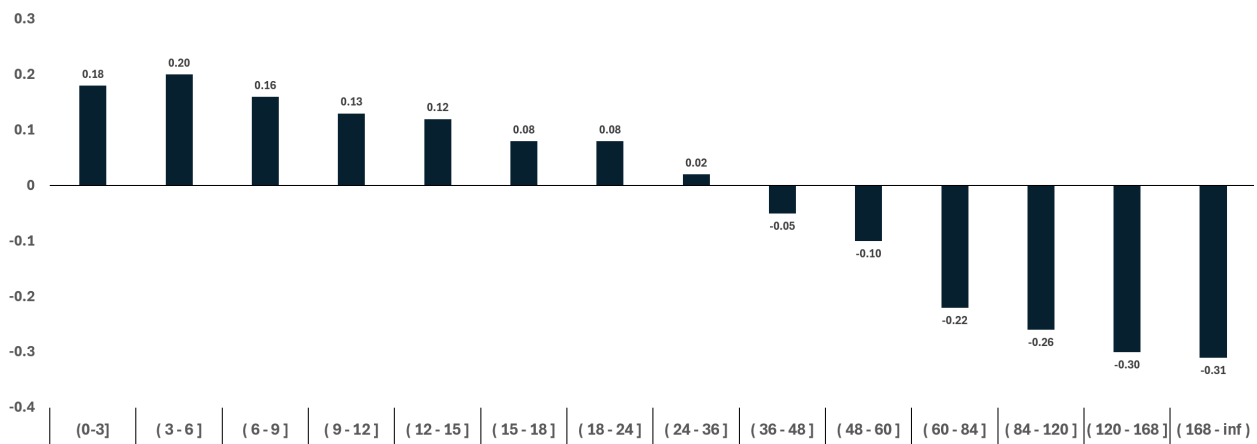
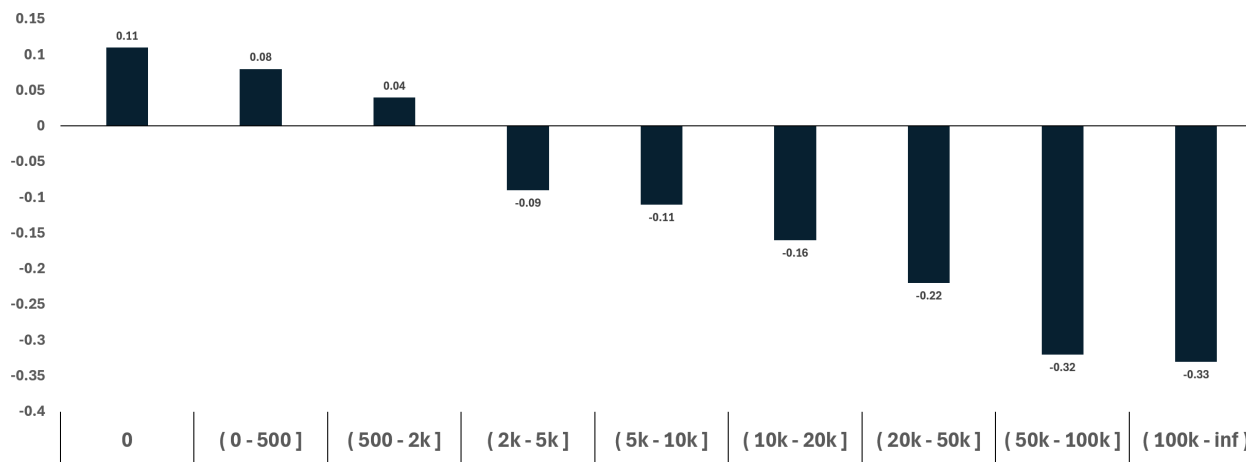


Figura 12. Valores SHAP de consumo_mes_tc

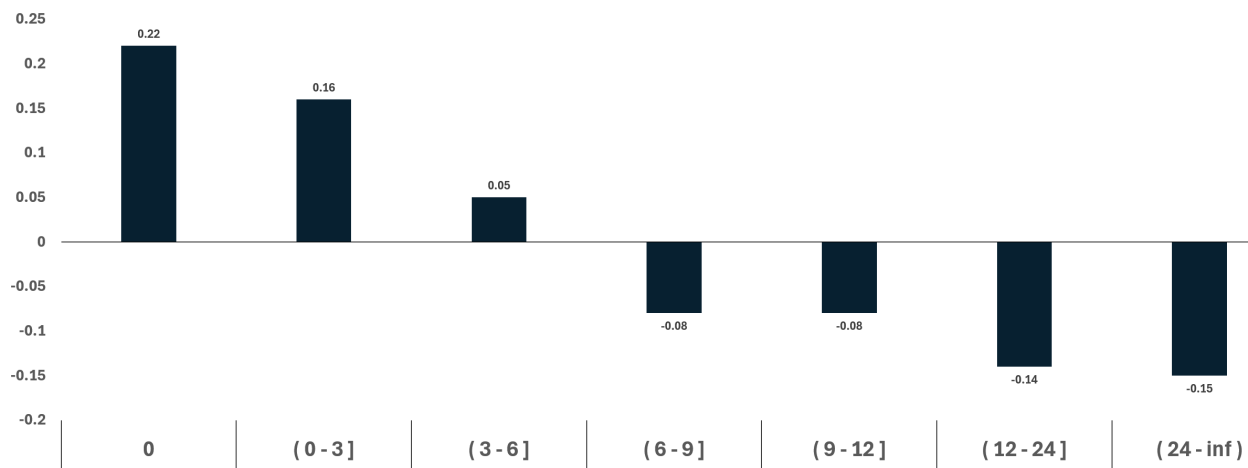


La variable consumo_mes_tc, como se muestra en la **Figura 12**, presenta un comportamiento inesperado en el cual, a medida que el consumo aumenta, la probabilidad de caer en incumplimiento disminuye. Este patrón, contrario a la intuición, pudiera explicarse de varias maneras:

- Proxy de Ingreso: es posible que el modelo esté considerando esta variable como un proxy de ingresos. En otras palabras, los deudores con un mayor consumo mensual en tarjeta de crédito podrían ser percibidos como aquellos con mayores ingresos y, por lo tanto, menor riesgo de incumplimiento.
- Relación con la Utilización de la Tarjeta: también es importante notar que existe una relación inversa entre la utilización de la tarjeta de crédito y el consumo. Cuando el deudor tiene una alta utilización de su tarjeta de crédito, es decir, está utilizando una gran parte de su límite de crédito, el mismo tiene menor margen para realizar consumos adicionales. Esto podría influir en un consumo mensual menor.

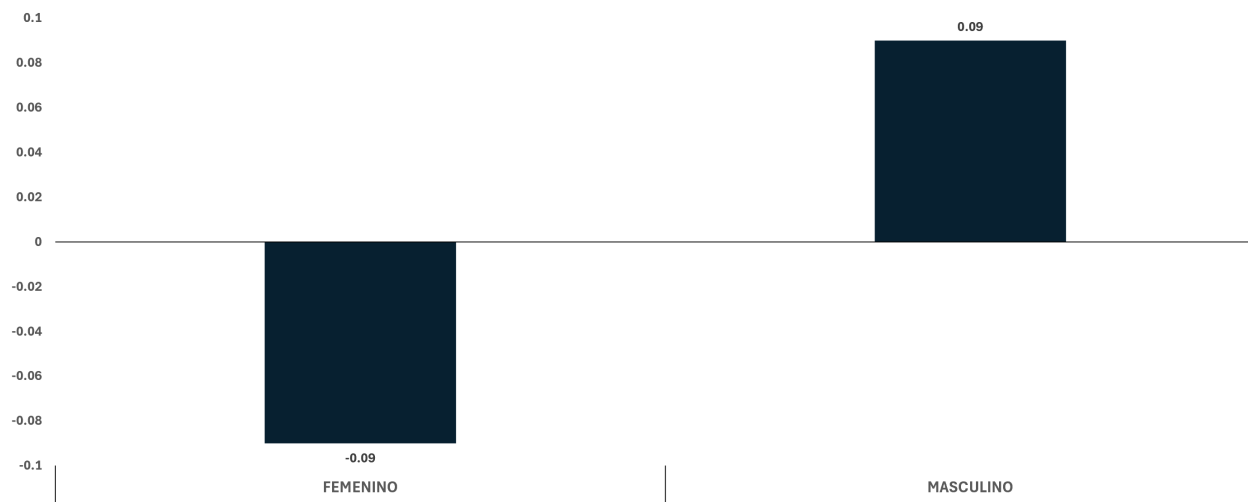
En cualquier caso, este hallazgo subraya la importancia de comprender las relaciones entre las variables en un modelo de riesgo de crédito y considerar múltiples factores para interpretar correctamente sus resultados.

Figura 13. Valores SHAP de meses_ultimo_desembolso



La variable meses_ultimo_desembolso, como se muestra en la **Figura 13**, indica que tener un desembolso en los últimos 6 meses contribuye a un aumento en la probabilidad de caer en incumplimiento. Este hallazgo sugiere que los desembolsos recientes pueden estar relacionados con un mayor riesgo crediticio.

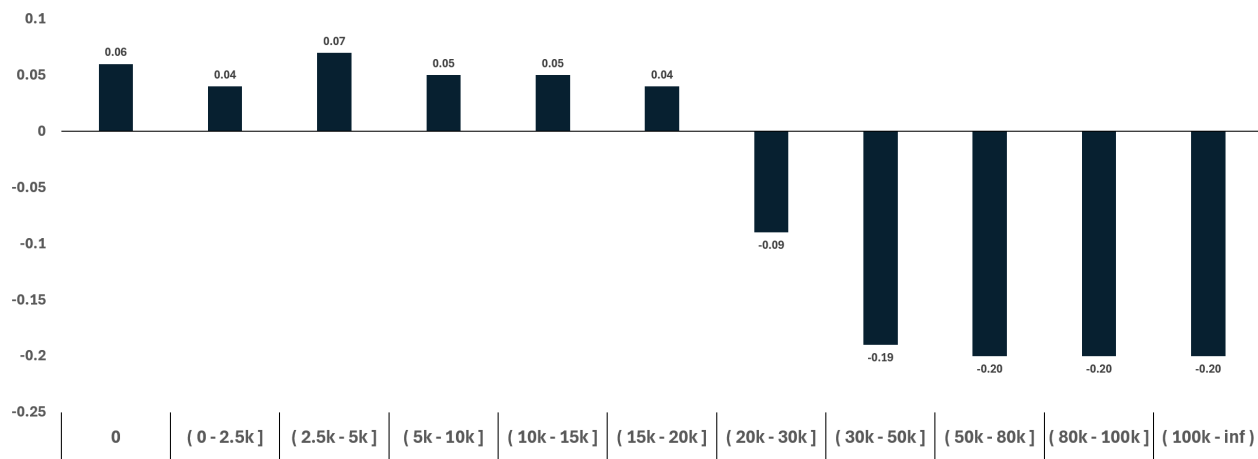
La presencia de desembolsos recientes como un factor que aumenta la probabilidad de incumplimiento puede ser indicativa de varios escenarios. Por ejemplo, podría reflejar la situación en la que los deudores que han recibido nuevos créditos en los últimos meses están enfrentando dificultades financieras que los hacen más propensos a incumplir sus obligaciones.

Figura 14. Valores SHAP de genero

El género, como se muestra en la **Figura 14**, también ha arrojado resultados significativos en el análisis. En este caso, los resultados indican que los hombres son más propensos a caer en incumplimiento en comparación con las mujeres.

Es importante destacar que la inclusión del género como variable en un modelo de riesgo de crédito puede generar discusiones sobre equidad y sesgo. Es fundamental que las instituciones financieras y los responsables de la toma de decisiones tengan en cuenta las implicaciones éticas y legales al utilizar variables como el género en sus modelos y que tomen medidas para garantizar la equidad y la no discriminación en sus prácticas crediticias.

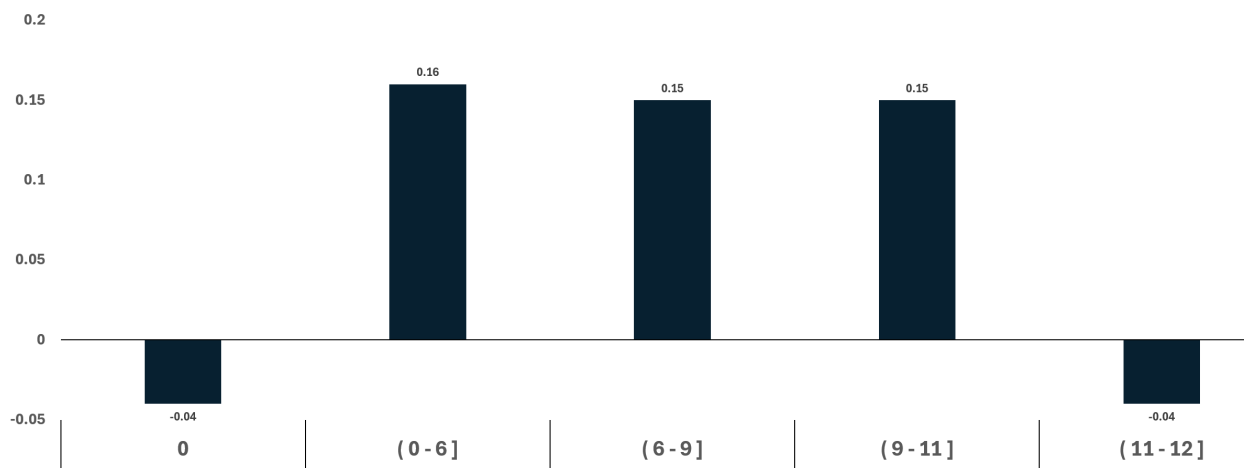
Figura 15. Valores SHAP de ingresos



Los ingresos del deudor, como se muestra en la **Figura 15**, es otra variable de suma importancia en el análisis. Esta variable proporciona información sobre el nivel socioeconómico del deudor, y se observa que, a partir de un umbral de ingresos de 20,000 pesos, el deudor comienza a ser menos propenso a caer en incumplimiento.

Este hallazgo sugiere que los deudores con ingresos más altos tienden a tener un menor riesgo de incumplimiento. La relación entre los ingresos y el riesgo crediticio es un patrón comúnmente observado en el análisis de riesgo de crédito, ya que los ingresos suelen estar relacionados con la capacidad de pago de un individuo.

Figura 16. Valores SHAP de meses_reportado_tss



La cantidad de meses que un deudor ha sido reportado a la TSS (Tesorería de la Seguridad Social), como se muestra en la **Figura 16**, revela un resultado interesante. En particular, se observa que no tener ningún reporte o tener 12 meses continuos de reporte son los dos casos en los que el deudor es menos propenso a caer en incumplimiento.

Este patrón puede estar relacionado con la informalidad laboral y la estabilidad del empleo. La ausencia de reportes en la TSS podría indicar que el deudor trabaja en la economía informal y tener los 12 meses de reporte sugiere una mayor estabilidad laboral en un trabajo formal. Los casos particulares son cuando el deudor tiene algún reporte, pero no los doce, lo que se asocia con una mayor inestabilidad laboral y financiera.

6. Conclusión

Este estudio se propuso identificar las variables críticas para la evaluación del riesgo crediticio en la República Dominicana, y ha logrado este objetivo de manera efectiva. La utilización de *catboost* y, posteriormente, *xgboost* junto con la biblioteca SHAP, ha demostrado ser una herramienta valiosa para automatizar y mejorar la eficiencia en la evaluación de riesgos, proporcionando una comprensión más profunda de los factores que inciden en el riesgo crediticio.

Se destaca la influencia de variables como la tasa de utilización de tarjetas de crédito y el máximo atraso en créditos existentes, las cuales consistentemente muestran significancia

en la predicción de incumplimientos. Además, se observa la importancia de factores como los ingresos, la estabilidad laboral y la consideración de eventos externos, como la pandemia de COVID-19.

Este estudio contribuye significativamente a la literatura al identificar las variables más relevantes en el contexto específico de la República Dominicana, fortaleciendo la comprensión existente y proporcionando una base para investigaciones futuras. Las implicaciones prácticas son notables, ofreciendo una base para la construcción de variables críticas que faciliten el monitoreo y la prevención de incumplimientos, destacando la importancia tanto del poder predictivo como de la interpretabilidad de los modelos.

Es crucial reconocer la limitación principal de este estudio en cuanto al acceso limitado a algunas variables clave, lo que sugiere la necesidad de mejorar la recopilación de datos en futuras investigaciones. En conjunto, este análisis subraya la importancia de un enfoque integral en la evaluación del riesgo crediticio, considerando tanto las variables tradicionales como el contexto económico y laboral.

En conclusión, este estudio ofrece una perspectiva integral sobre el riesgo crediticio en la República Dominicana, subrayando la necesidad de considerar múltiples factores y la importancia continua de explorar y refinar la utilización de modelos avanzados en el sector bancario. La combinación de enfoques analíticos avanzados, la identificación de variables clave y la conciencia de las limitaciones actuales proporcionan una base sólida para futuras investigaciones y mejoras en la gestión del riesgo crediticio en la región.

Los resultados de este trabajo pueden ser encontrados en el repositorio de Github del Departamento de Estudios Económicos de la Superintendencia de Bancos, mediante el enlace https://github.com/SB-ESTUDIOS/paper_prediccion_default/tree/main

Referencias

- A.A., D., N., A., y S., Q. (2019). Estimating probabilities of default of different firms and the statistical tests. *Glob Entrepr Res*, 9(27). doi: <https://doi.org/10.1186/s40497-019-0152-8>
- Abellán, J., y Mantas, C. J. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 41(8), 3825-3830. Descargado de <https://www.sciencedirect.com/science/article/pii/S0957417413009676> doi: <https://doi.org/10.1016/j.eswa.2013.12.003>
- Alonso, A., y Carbo, J. M. (2021). Understanding the performance of machine learning models to predict credit default: A novel approach for supervisory evaluation. (215). Descargado de <https://dx.doi.org/10.2139/ssrn.3774075>
- Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., y Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627-635. Descargado de <https://doi.org/10.1057/palgrave.jors.2601545> doi: 10.1057/palgrave.jors.2601545
- Bao, W., Lianju, N., y Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. , 128, 301-315. Descargado de <https://doi.org/10.1016/j.eswa.2019.02.033>
- Bellotti, T., y Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36(2, Part 2), 3302-3308. Descargado de <https://www.sciencedirect.com/science/article/pii/S0957417408000857> doi: <https://doi.org/10.1016/j.eswa.2008.01.005>
- Bournigal, G., y Díaz, M. (2021). Riesgo de crédito en la república dominicana: Estudio de matrices de transición. Descargado de https://sb.gob.do/sites/default/files/nuevosdocumentos/SB_01-21_Riesgo_de_Credito.pdf
- Bournigal, G., y García, M. (2022). Predicción del default en carteras de crédito: Un enfoque de machine learning (ml). Descargado de <https://cdn.bancentral.gov.do/documents/biblioteca/documents/2022-cuarto.pdf>
- Brown, I., y Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453. Descargado de <https://www.sciencedirect.com/science/article/pii/S095741741101342X> doi: <https://doi.org/10.1016/j.eswa.2011.09.033>
- Bussmann, N., Giudici, P., y Papenbrock, J. (2019). Explainable machine learning in credit risk management. Descargado de <https://dx.doi.org/10.2139/ssrn.3506274>

- Chen, T., y Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *arXiv preprint arXiv:1603.02754*. Descargado de <https://arxiv.org/abs/1603.02754>
- Económicos, E. (s.f.).
 Descargado de https://github.com/SB-ESTUDIOS/paper_prediccion_default/tree/main
- Fico. (2023). Descargado de <https://www.myfico.com/>
- Finlay, S. (2011). Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research*, 210(2), 368-378. Descargado de <https://www.sciencedirect.com/science/article/pii/S0377221710006272> doi: <https://doi.org/10.1016/j.ejor.2010.09.029>
- Freund, Y., y Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119–139. Descargado de <https://doi.org/10.1006/jcss.1997.1504> doi: 10.1006/jcss.1997.1504
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 29(5), 1189–1232. Descargado de <https://doi.org/10.1214/aos/1013203451> doi: 10.1214/aos/1013203451
- Hand, D., Sohn, S., y Kim, Y. (2005, octubre). Optimal bipartite scorecards. *Expert Systems with Applications*, 29(3), 684–690. (Funding Information: The work of So Young Sohn on this project was supported by Chevening Scholarship number KOR3500200 from the British Council.) doi: 10.1016/j.eswa.2005.04.032
- He, J., Shi, Y., y Xu, W. (2005). classifications of credit cardholder behavior by using multiple criteria non-linear programming. , 154–163.
- Khandani, A., Kim, A., y Lo, A. (2010). Consumer credit risk: Individual probability estimates using machine learning. , 34, 2767-2787. Descargado de <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- Kruppa, J., Schwarz, A., Armingier, G., y Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. , 40. Descargado de <https://doi.org/10.1016/j.eswa.2013.03.019>
- Lee, T.-S., y Chen, I.-F. (2003, 01). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Proceedings of the International Conference on Information and Knowledge Engineering*, 2, 533-538.
- Leo, M., Sharma, S., y Maddulety, K. (2019). Machine learning in banking risk management: A literature review. , 7. Descargado de <https://doi.org/10.3390/risks7010029>

- Lessmann, S., Baesens, B., Seow, H.-V., y Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136. Descargado de <https://www.sciencedirect.com/science/article/pii/S0377221715004208> doi: <https://doi.org/10.1016/j.ejor.2015.05.030>
- Moscatelli, M., Parlapiano, F., Narizzano, S., y Viggiano, G. (2020). Corporate default forecasting with machine learning. , 161. Descargado de <https://doi.org/10.1016/j.eswa.2020.1135673>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., y Gulin, A. (2018). Catboost: unbiased boosting with categorical features. , 6638–6648. Descargado de <https://proceedings.neurips.cc/paper/2018/file/9045e5a2e02e7c86e2090d3a70e1bec0-Paper.pdf>
- Schmitt, M. (2022). Deep learning vs. gradient boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring. *Journal of Financial Services Research*. Descargado de <https://doi.org/10.1007/s10693-009-0075-6>
- Tsai, C.-F., y Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. Descargado de <https://www.sciencedirect.com/science/article/pii/S0957417407001558> doi: <https://doi.org/10.1016/j.eswa.2007.05.019>
- Vincenzo Moscato, G. S., Antonio Picariello. (2021). A benchmark of machine learning approaches for credit score prediction. *ELSEVIER*. Descargado de <https://www.elsevier.com/locate/eswa>
- Yang, Y. (2007). Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research*, 183(3), 1521-1536. Descargado de <https://www.sciencedirect.com/science/article/pii/S037722170601191X> doi: <https://doi.org/10.1016/j.ejor.2006.10.066>
- Zhang, D., Zhou, X., Leung, S. C., y Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*, 37(12), 7838-7843. Descargado de <https://www.sciencedirect.com/science/article/pii/S0957417410003544> doi: <https://doi.org/10.1016/j.eswa.2010.04.054>