



UNIVERSIDAD
DE PIURA

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**Modelo de clasificación de locatarios para minimizar el
riesgo de impago del portafolio de inversiones
inmobiliarias**

Trabajo de Suficiencia Profesional para optar el Título de
Economista

Karol Iliana Esther Miranda Sánchez

Revisor(es):
Mgtr. Harry Omar Patrón Torres

Piura, junio de 2023

Declaración Jurada de Originalidad del Trabajo Final

Yo, Karol Iliana Esther Miranda Sánchez, egresada del Programa Académico de Economía de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Piura, identificado(a) con DNI 71130647

Declaro bajo juramento que:

1. Soy autor del trabajo final titulado:
“ Modelo de clasificación de locatarios para minimizar el riesgo de impago del portafolio de inversiones inmobiliarias ”
El mismo que presento bajo la modalidad de **Trabajo de suficiencia profesional**¹ para optar el **Título Profesional**² de Economista.
2. La asesoría del trabajo estuvo a cargo de:
 - Mgtr. Harry Omar Patrón Torres, identificado con DNI N° 07251849
3. El texto de mi trabajo final respeta y no vulnera los derechos de terceros o de ser el caso derechos de los coautores, incluidos los derechos de propiedad intelectual, datos personales, entre otros. En tal sentido, el texto de mi trabajo final no ha sido plagiado total ni parcialmente, para la cual he respetado las normas internacionales de citas y referencias de las fuentes consultadas.
4. El texto del trabajo final que presento no ha sido publicado ni presentado antes en cualquier medio electrónico o físico.
5. La investigación, los resultados, datos, conclusiones y demás información presentada que atribuyo a mi autoría son veraces.
6. Declaro que mi trabajo final cumple con todas las normas de la Universidad de Piura.

El incumplimiento de lo declarado da lugar a responsabilidad del declarante, en consecuencia; a través del presente documento asumo frente a terceros, la Universidad de Piura y/o la Administración Pública toda responsabilidad que pueda derivarse por el trabajo final presentado. Lo señalado incluye responsabilidad pecuniaria incluido el pago de multas u otros por los daños y perjuicios que se ocasionen.

Fecha: 19/06/2023.



Firma del autor³

¹ Indicar si es tesis, trabajo de investigación, trabajo académico o trabajo de suficiencia profesional.

² Grado de Bachiller, Título profesional, Grado de Maestro o Grado de Doctor.

³ Idéntica a DNI, no se admite digital salvo certificado.

A mis padres y hermano, por todo su apoyo y comprensión en cada decisión que he tomado, gracias por acompañarme en toda mi etapa universitaria, espero seguir compartiendo más logros con ustedes.



Resumen

El objetivo del presente modelo es clasificar potenciales locatarios como parte del proceso de evaluación y aceptación de contrapartes de operaciones inmobiliarias. Como parte del desarrollo, se busca determinar las variables relevantes asociadas a la probabilidad de atraso mayor a 2 meses por parte de los locatarios y la selección de un modelo *logit* que cumpla con los requisitos mínimos de aceptación estadística.



Tabla de contenido

Introducción	8
Capítulo 1. Informe sobre la experiencia laboral.....	9
1.1 Protecta security.....	9
1.2 Analista de inversiones	9
Capítulo 2. Trabajo de investigación.....	11
2.1 Contexto.....	11
2.2 Definición de la variable dependiente	11
2.3 Variables independientes.....	11
2.4 Estadísticos para medición del desempeño	17
2.4.1 Estadístico Kolmogorov-Smirnov.....	17
2.4.2 Curva de ROC.....	17
2.5 Selección de variables	17
2.6 Modelo de regresión logística	23
2.7 Interpretación de las variables independientes	24
2.7.1 Multi_Meses_mora_may_90_ult48m	24
2.7.2 Multi_Ult_6_Nor.....	24
2.7.3 Sectores.....	25
2.7.4 vtas_fis	25
2.7.5 dum_saldo_al_dia.....	25
2.8 Predictibilidad del modelo - Análisis de ajuste y discriminación	26
Conclusiones	27
Referencias.....	28
Notas a pie de página	29

Lista de tablas

Tabla 1. Descripción de las variables independiente.....	12
Tabla 2. Estadísticas descriptivas de variable dum trab	18
Tabla 3. Estadísticas descriptivas de variable status_sunat.....	18
Tabla 4. Transformación de variables	19
Tabla 5. Evaluación de variables independientes.....	19
Tabla 6. Evaluación del principio WOE en las variables independientes	21
Tabla 7. Matriz de correlaciones de variables independientes.....	22
Tabla 8. Resultados del modelo de regresión	24
Tabla 9. Tabla de ROC	26



Lista de figuras

Figura 1. Curva de ROC.....26



Introducción

Dentro del área de inversiones se cuenta también con una amplia cartera de inmuebles, la cual abarca el 35% del portafolio total siendo un área relevante para el estudio. Asimismo, nos enfocamos en los problemas que conlleva la operativa de cobranzas y atrasos por parte de un grupo de contrapartes.

Para tal, nos preguntamos cómo podríamos predecir aquellos operadores que logren obtener un mejor desempeño en cobranzas y cuáles deben ser las características óptimas para lograrlo.

El modelo tiene como objetivo optimizar el proceso de alquiler de locales mediante la clasificación de características relevantes que busquen potenciar el número de locatarios con mejor desempeño en los pagos de alquiler. Asimismo, cabe recalcar que buscamos operadores enfocados en tiendas *stand-alone*, es decir locales que no están ligados a centros comerciales y se encuentran ubicados en zonas específicas.

Para el desarrollo de la tesis, se ha decidido separarla en dos capítulos. El primer capítulo se enfoca en la compañía en la cual labora el autor y sus funciones dentro de ella. El segundo capítulo nos explica el planteamiento del problema donde se detalla el problema a solucionar, los objetivos y metodología a desarrollar.

Finalmente se entregan las conclusiones de esta investigación siendo la conclusión principal, un modelo de variables relevantes para el uso de los especialistas con el fin de mejorar la toma de decisiones que permitan obtener mayores locatarios con menos atrasos en cobranzas superando un nivel de acierto del 76%.

Capítulo 1. Informe sobre la experiencia laboral

1.1 Protecta security

El autor de la presente investigación trabaja en una compañía de seguros llamada Protecta Security, desde el año 2020. Protecta Security (la cual será llamada la compañía en adelante), es una compañía de seguros vida que pertenece al grupo chileno Security con más de 15 años en el mercado peruano.

La compañía cuenta con respaldo de dos grandes grupo económicos Grupo Security y Grupo ACP. Actualmente tiene alrededor de 400 colaboradores y cuenta con más de 700,000 asegurados.

Adicionalmente, la compañía busca inculcar en todos sus colaboradores su cultura de trabajo, la misma que incluye los siguientes cuatro valores: cercanía, transparencia, profesionalismo e innovación.

A continuación, revisaremos los cargos ocupados por el autor en la institución financiera, así como los aprendizajes que obtuvo en cada uno de ellos.

1.2 Analista de inversiones

Dentro de las diversas áreas de la compañía, el autor se ubica dentro del área de inversión financieras como analista de inversiones.

El área de inversiones es una parte primordial de la compañía dado que es el encargado de generar la rentabilidad de la compañía y con eso poder asegurar a nuestros clientes un retorno de sus inversiones óptimo dentro del mercado. Por lo que, entre las principales funciones del autor están:

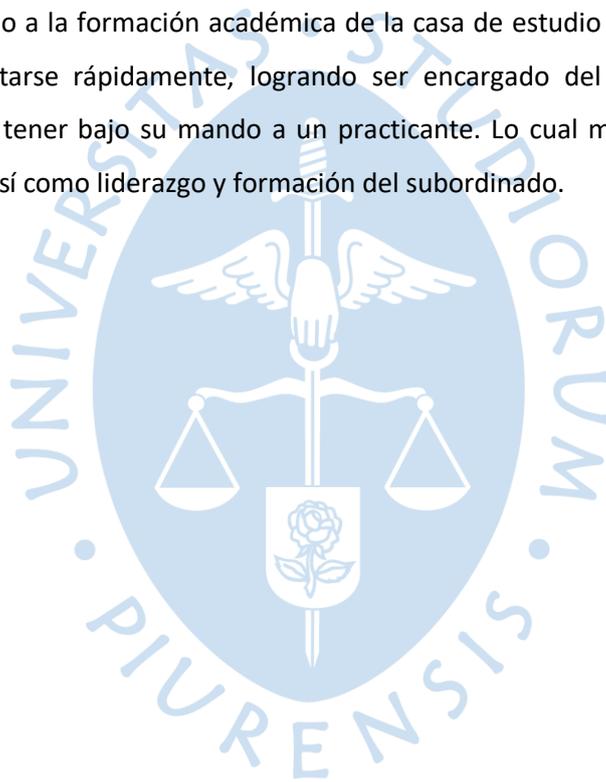
- Gestionar y supervisar el correcto registro de la información de las operaciones de inversiones en el sistema, y además realizar la revisión y validación para que pasen a la contabilidad de manera correcta, así como la comunicación con los custodios para su registro necesario.
- Apoyar la elaboración y análisis de conciliaciones bancarias y conciliaciones de las cuentas de inversiones u otras operaciones financieras, tanto locales como extranjeros
- Participar en el análisis y elaboración del portafolio de inversiones cumpliendo con las IFRS establecidas.
- Apoyar la atención de auditorías (interna, externa u otros entes reguladores).
- Elaborar informes, reportes e indicadores para gerencia y comité de inversiones
- Controlar que las inversiones realizadas en el día se encuentren enmarcadas dentro de la Política de Tesorería.
- Verificar y validar las operaciones en cuentas de banco tramitadas por el back office y administración efectivo, como la elaboración del flujo de caja diario para la atención de las operaciones diarias

- Actualizar en el sistema las calificaciones de riesgo de las instituciones financieras.
- Registrar pagos de inmuebles tales como compras y alcabala para el ingreso a la cartera de inmuebles, así como su registro dentro de las normativas de la SBS.
- Elaborar y enviar anexos y reportes a la SBS u otras entidades reguladoras.

Ser parte del área de inversiones en Protecta security, es una experiencia muy enriquecedora, ya que podemos ver la operativa del *front y back* de las inversiones, así como aprender nuevas plataformas para el registro y monitoreo del portafolio.

Las labores diarias del trabajo requieren de mucho contacto con las áreas de tesorería, contabilidad y riesgos de inversiones para poder conciliar y revisar el buen manejo de registro de las inversiones. Por lo cual, el aprendizaje dentro de esta compañía ha sido fructífero y desafiante.

Además, debido a la formación académica de la casa de estudio del autor este ha podido desenvolverse y adaptarse rápidamente, logrando ser encargado del área de back office de inversiones y logrando tener bajo su mando a un practicante. Lo cual mejoro sus habilidades de gestión y delegación, así como liderazgo y formación del subordinado.



Capítulo 2. Trabajo de investigación

2.1 Contexto

Dentro de la cartera de inversiones, se encuentran la inversión inmobiliaria la cual representa una parte esencial dentro del portafolio por lo que es primordial poder seleccionar a las contrapartes que tengan un buen desempeño en sus pagos. En los últimos años, el mercado ha apostado por las tiendas de formato *stand alone*, los cuales hacen referencia a tiendas independientes que no se encuentran conectadas a un centro comercial, sino que se ubican en lugares estratégicos de manera individual, los cuales mayormente cuentan con su propio estacionamiento. Asimismo, en los últimos años ha habido un aumento de tiendas de conveniencia, tras la llegada de tiendas extranjeras como Tambo u Oxxo, las cuales manejan este formato debido a los bajos precios por metro cuadrado en comparación a las tiendas en centros comerciales. Por lo que tanto para la investigación, se contó con 58 registros de la base de inmuebles de la compañía de los cuales se recopiló la data relevante de los informes crediticios *Experian*.

Esta última es una compañía especializada en la recopilación del historial crediticio de más de mil millones de empresas y personas donde se encuentra la información sobre sus obligaciones financieras con entidades bancarias, compañías de financiamiento comercial, cooperativas, empresas de servicios como telefonía celular, televisión por suscripción, o burós de crédito.

2.2 Definición de la variable dependiente

La variable a explicar es de carácter binario, la cual toma valores de "1" cuando el locatario mantiene un atraso mayor o igual a 2 rentas al cierre del octubre 2021, los cuales recolecta datos de agosto, setiembre y octubre del año mencionado. La data cuenta con 29 valores tipo 1.

2.3 Variables independientes

Se construyeron 23 variables independientes en base a la información crediticia de los locatarios obtenida de *Experian*. Estas variables fueron creadas para describir el comportamiento de pago del cliente. En este sentido, las variables involucradas son potencialmente relevantes en la determinación de la probabilidad de incumplimiento y tienen un carácter tanto cuantitativo como cualitativo.

Tabla 1*Descripción de las variables independiente*

Variable	Descripción	Detalle
años_d_vida	Cuantifica el número de años de existencia. Se construye de la siguiente manera: (30/11/2021-fecha de inicio de actividades) / 365 días	variable multinomial - Menor a 10 años=1 - Menor a 20 años=2 - Mayor a 20 años =3
status_sunat	Es una variable binaria que coloca "1 " si el locatario es considerado un principal contribuyente (PRICO) o "0" si es mediano o pequeño (MEPECO).	variable binaria PRICO=1 MEPECO=0
dum_trab	Es una variable binaria que coloca "1 " si el locatario mantiene trabajadores formales o "0" en caso contrario.	variable binaria - Tiene trabajadores=1 - No tiene trabajadores=0
trabajadores	Es una variable multinomial que considera el número de trabajadores, para lo cual se segmenta el rango de datos en 3 categorías.	variable multinomial - No tiene=1 - Menor a 100=2 - Mayor a 100 =3
sectores	Es una variable multinomial que considera el sector donde se desempeña el locatario, para lo cual asigna distintas categorías.	variable multinomial - S. Comidas=1 - S. Industrial y otros=2 - S. Inmobiliario =3 - S. Mayorista =4 - S. Público =5 - S. Retail=6 - S. Salud=7 - S. Telecomunicaciones=8
vtas_fis	Es una variable multinomial que considera el nivel de ventas fiscales reportado por el informe <i>Experian</i> . Se asigna distintas categorías en base al nivel de ventas.	variable multinomial - No especifica=0 - Menores a S/. 547,500=1 - Entre S/. 547,500 y S/. 6,205,000 =2 - Entre S/. 6,205,000 y S/. 73,000,000 =3 - Mayores a S/. 73,000,000 =4
dum_deuda_SF	Es una variable binaria que coloca "1 " si el locatario presenta deuda o "0" en caso contrario.	variable binaria donde - No presenta deuda=0 - Presenta deuda=1
deuda	Es una variable multinomial que considera el nivel de deuda financiera del locatario, para lo cual se asigna distintas categorías.	variable multinomial - No tiene=1 - Menor a S/10 millones=2 - Mayor a S/10 millones =3
dum_saldo_al_dia	Es una variable binaria que coloca "1 " si el locatario mantiene el 100% del saldo de deuda al día o "0" en caso contrario.	variable binaria donde - Presenta saldo 100% al día=1 - De otra manera=0

Variable	Descripción	Detalle
Nro_d_obliga	Es una variable multinomial que considera el nivel de deuda financiera del locatario, para lo cual se segmenta el rango de datos en 3 categorías.	variable multinomial - Null=1 - Menor 3=2 - Mayor a 3=3
Multi_Ult_6_Nor	Es una variable multinomial creada para identificar si el locatario mantiene 6 clasificaciones del tipo Normal en los últimos 6 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 6 registros son Normales.	variable multinomial - Null=1 - No tiene=2 - Si tiene=3
Multi_Ult_12_Nor	Es una variable multinomial creada para identificar si el locatario mantiene 12 clasificaciones del tipo Normal en los últimos 12 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 12 registros son normales.	variable multinomial - Null=1 - No tiene=2 - Si tiene=3
Multi_Ult_18_Nor	Es una variable multinomial creada para identificar si el locatario mantiene 18 clasificaciones del tipo Normal en los últimos 18 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 18 registros son Normales.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene=3
Multi_Ult_24_Nor	Es una variable multinomial creada para identificar si el locatario mantiene 24 clasificaciones del tipo Normal en los últimos 24 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 24 registros son Normales.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene=3
Multi_Ult_36_Nor	Es una variable multinomial creada para identificar si el locatario mantiene 36 clasificaciones del tipo Normal en los últimos 36 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o	más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 36 registros son Normales.

variable multinomial
- - Null=1

- - No tiene=2
- - Si tiene =3



Variable	Descripción	Detalle
Multi_Ult_48_N or	Es una variable multinomial creada para identificar si el locatario mantiene 48 clasificaciones del tipo Normal en los últimos 48 meses. Null hace referencia a la falta de historial crediticio suficiente. La categoría "No tiene", toma en cuenta la existencia de una o más clasificaciones distintas a la Normal. La categoría "Tiene", significa que sus últimos 48 registros son Normales.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3
Multi_Meses_mora_9 _30_ult48m	Es una variable multinomial creada para identificar el número de meses con mora entre 9 y 30 días durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con mora entre 9 y 30 días. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con mora entre 9 y 30 días.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3
Multi_Meses_mora_3 1 _60_ult48m	Es una variable multinomial creada para identificar el número de meses con mora entre 31 y 60 días durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con mora entre 31 y 60 días. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con mora entre 31 y 60 días.	variable multinomial - - null=1 - - no tiene=2 - - si tiene =3
Multi_Meses_mora_6 1 _90_ult48m	Es una variable multinomial creada para identificar el número de meses con mora entre 61 y 90 días durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con mora entre 61 y 90 días. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con mora entre 61 y 90 días.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3

Variable	Descripción	Detalle
Multi_Meses_mora_may_90_ult48m	Es una variable multinomial creada para identificar el número de meses con mora mayor a 90 días durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con mora mayor a 90 días. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con mora mayor a 90 días.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3
Multi_Meses_cobr_jud_ult48m	Es una variable multinomial creada para identificar el número de meses con deuda financiera en cobranza judicial durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con deuda en cobranza judicial. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con deuda en cobranza judicial.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3
Multi_Meses_incobrab le_ult48m	Es una variable multinomial creada para identificar el número de meses con deuda financiera en incobrable durante los últimos 48 meses. Null hace referencia a que no se cuenta con historial crediticio. La categoría "No tiene", toma en cuenta existencia historial crediticio, pero sin la presencia de meses con deuda incobrable. La categoría "Tiene", significa que sus últimos 48 meses posee 1 o más meses con deuda en incobrable.	variable multinomial - - Null=1 - - No tiene=2 - - Si tiene =3
var_an_deuda	Es una variable multinomial que considera si la deuda del locatario presenta una variación anual positiva o negativa. Se crean 3 categorías para considerar los registros nulos (no se presentó deuda).	variable multinomial - - Null=1 - - Si es positiva=2 - - Si es negativa

Nota. Información crediticia obtenida de Experian

2.4 Estadísticos para medición del desempeño

Para la evaluación se usaron dos estadísticos principales con el fin de medir el desempeño de las variables y si estas eran relevantes para el modelo

2.4.1 Estadístico Kolmogorov-Smirnov

El estadístico Kolmogorov-Smirnov (en adelante, KS) o también conocido como “de separación” es una prueba no paramétrica utilizada principalmente para medir la bondad de ajuste de dos distribuciones entre sí. En efecto, el KS tiene como objetivo verificar que la distribución de los solicitantes buenos (aceptados) es lo más distinta posible a la distribución de los solicitantes malos (rechazados) (Tineo, 2019).

De acuerdo con Halim y Humira (2014), se espera que un modelo de puntuación posea un buen poder de discriminación cuando el valor del KS se encuentra entre 28% y 50%.

2.4.2 Curva de ROC

La curva ROC (Característica Operativa del Receptor) es un método estadístico para determinar la sensibilidad (1 - especificidad) en un sistema de clasificación binaria después de variar el umbral de discriminación. La curva superior representa la relación de verdaderos positivos (VPR = tasa de verdaderos positivos) a la tasa de falsos positivos (FPR = tasa de falsos positivos), teniendo en cuenta las variaciones en el umbral de discriminación, a partir del cual se determina que los casos son positivos (Tseng-Chung & Li-Chiu, 2005).

En este modelo, nos enfocamos en el área bajo la curva ROC (AUC), la cual es una representación escalar del rendimiento esperado del clasificador. AUC siempre está entre 0 y 1, cuanto mayor sea el valor, mejor será el clasificador.

2.5 Selección de variables

Se busca seleccionar las variables adecuadas (relevantes) para desarrollar el modelo logit. En ese sentido, las variables explicativas del apartado 3, fueron sometidas a tres filtros para verificar su idoneidad.

El primer filtro se enmarca en el análisis de concentración, el cual busca eliminar aquellas variables que poseen una cantidad abundante de valores de un solo tipo. El segundo principio de enfoca en el indicador información value¹ (IV), el cual es una medida ampliamente utilizada en modelos de clasificación. Según señala Zeng (2013), IV es una medida muy utilizada para la clasificación crediticia y su aporte se basa en asignar valores numéricos que determinan el poder predictivo de las variables independientes. estoy aquí. a dependencias binarias variable. Luego, se verifica el indicador *weight of evidence*² (WOE) y su respectiva gráfica, ya que de esta forma se mide la fuerza de cada atributo y la validación del principio de monotonicidad. Finalmente, tercer filtro, se basa en el análisis de correlación bivariada, dado que, Yoo et al., (2014) señalaron la relevancia de

este para detectar problemas de multicolinealidad y redundancia en la regresión multivariada, encontrándolo útil en la determinación de coeficientes de variables independientes, destacando el impacto negativo del problema.

Después de realizar el primer filtro, se elimina la variable “dum_trab” por tener más del 90% de concentración en un solo atributo. Posteriormente, se elimina la variable “status_sunat”, ya que cuenta con bajo nivel predictivo. Para el cual se utilizó el criterio de descarte: $IV < 0.02$, conforme con los criterios establecidos por Siddiqi (2006).

Tabla 2

Estadísticas descriptivas de variable dum trab

dum_trab	Conteo	%
0	5	8.6
1	53	91.4
Total	58	100

Nota. Elaboración propia

Tabla 3

Estadísticas descriptivas de variable status_sunat

Status_sunat	,0	1,0	no %	sí %	WOE	IV
,0	8	8	16	28	-	-
1,0	21	21	42	72	-	-
Total	29	29	58	100		-

Nota. Elaboración propia

Por otro lado, las variables “sectores” y “vtas_fis” se sometieron al proceso de transformación, toda vez que su forma original no permite el cálculo del indicador IV.

Tabla 4

Transformación de variables

Variable	Comentario	Nueva forma	IV
sectores	Se transforma ya que no se puede determinar el IV. *3,5 y 8 se fusionan, en la categoría 3 *La categoría 6 pasa a ser 5 * La categoría 7 pasa a ser 6	variable multinomial - s. comidas=1 - s. industrial y otros=2 - s. inmobiliario, publico y telecomunicaciones =3 s. mayorista =4 s. retail=5 s. salud=6	0.1817
vtas_fis	Se transforma ya que no se puede determinar el IV. *2 y 3 se fusionan, en la categoría 2 *La categoría 4 pasa a ser 3	variable multinomial - No especifica=0 - Menores a S/. 547,500=1 - Entre S/. 547,500 y S/. 73,000,000 =2 - Mayores a S/. 73,000,000 =3	0.1841

Nota. Elaboración propia

Luego, de validar el indicador IV, se procede a validar el cumplimiento del indicador WOE, con el objetivo de no vulnerar el principio de monotonicidad.

Tabla 5

Evaluación de variables independientes

Variable	WOE	Graf. WOE	Comentario
dum_saldo_al_dia	- 0.54		Cumple con WOE
	0.26		
Multi_Meses_cobr_jud_ult48m	- 0.51		Cumple con WOE
	0.11		
	1.10		
deuda	- 0.69		No cumple con WOE
	0.34		
	0.11		
dum_deuda_SF	- 0.69		Cumple con WOE
	0.23		
Nro_d_obliga	- 0.92		No cumple con WOE
	0.79		
	-		
sectores	0.29		No cumple con WOE
	- 0.51		
	- 0.69		
	0.59		
	- 0.22		
Multi_Ult_6_Nor	- 0.51		No cumple con WOE
	0.69		
	0.11		
vtas_fis	- 0.41		No cumple con WOE
	0.56		
	- 0.22		
	0.47		

Variable	WOE	Graf. WOE	Comentario
años_d_vida	- 0.08		No cumple con WOE
	1.10		
	- 0.07		

Variable	WOE	Graf. WOE	Comentario
Multi_Ult_1 2_Nor	- 0.61		No cumple con WOE
	0.22		
	0.25		

Variable	WOE	Graf. WOE	Comentario
Multi_Ult_48_Nor	- 0.37		No cumple con WOE
	0.41		
	0.25		

Variable	WOE	Graf. WOE	Comentario
Multi_Ult_1 8_Nor	- 0.45		No cumple con WOE
	0.69		
	- 0.08		

Variable	WOE	Graf. WOE	Comentario
Multi_Meses_mora_ 31_60_ult48m	- 0.51		No cumple con WOE
	0.17		
	0.29		

Variable	WOE	Graf. WOE	Comentario
Multi_Ult_2 4_Nor	- 0.49		No cumple con WOE
	0.61		
	-		

Variable	WOE	Graf. WOE	Comentario
var_an_deuda	- 0.92		No cumple con WOE

Variable	WOE	Graf. WOE	Comentario
Multi_Ult_3 6_Nor	- 0.17		No cumple con WOE

Nota. Elaboración propia

Las variables que no cumplieron con el criterio de WOE, se les aplicó el principio de transformación mediante la agrupación de atributos, con lo cual se logró cumplir con el indicador de WOE. No obstante, se eliminan las variables: “años_d_vida”, “deuda” y “Multi_Ult_36_Nor”, puesto que producto de la transformación para cumplir con WOE, las variables en mención pierden predicción (indicador $IV < 0.02$).

Tabla 6

Evaluación del principio WOE en las variables independientes

Variable	Principio WOE	Comentario
años_d_vida	Se transforma ya que no cumple con WOE. 2 y 3 se fusionan, en la categoría 2	variable binaria - Menor a 10 años=1 - Mayor a 10 años=2 Cumple con WOE pero se elimina por tener un IV<0.02
sectores	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE. 1,4 y 5 se fusionan, en la categoría 1 • 2,3 y 6 se fusionan, en la categoría 2 	variable multinomial - s. comidas, mayorista, retail=1 - s. industrial y otros, inmobiliario, público y telecomunicaciones y salud=2
vtas_fis	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 1 y 2 se fusionan, en la categoría 1 • 3 pasa a ser 2 	variable multinomial - No especifica=0 - Menores a S/. 73,000,000=1 - Mayores a S/. 73,000,000=2
deuda	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 1 y 2 se fusionan en=1 • 3 pasa a ser 2 	variable binaria donde - No presenta o es menor a S/10 millones=1, mayor a S/10 millones=2 Cumple con WOE pero se elimina por tener un IV<0.02
Nro_d_obliga	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 2 y 3 se fusionan en 2 	variable binaria donde - null=1 - tiene obligaciones=2
Multi_Ult_6_Nor	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 2 y 3 se fusionan en 2 	variable binaria donde - null=1 - diferente a null=2
Multi_Ult_36_Nor	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 2 y 3 se fusionan en 2 	variable binaria donde - null=1 - diferente a null=2 Cumple con WOE pero se elimina por tener un IV<0.02
var_an_deuda	<ul style="list-style-type: none"> • Se transforma ya que no cumple con WOE • 2 y 3 se fusionan en 2 	variable binaria donde - null=1 - diferente a null=2

Nota. Elaboración propia (Resultados en SPSS)

A continuación, se analizó la correlación bivariada para las variables restantes.

Tabla 7

Matriz de correlaciones de variables independientes

Matriz de Correlaciones	trabajadores	sectores	vfas_fis	dum_deuda_SF	dum_saldo_al_dia	Nro_d_obliga	Multi_Ult_6_Nor	Multi_Ult_12_Nor	Multi_Ult_18_Nor	Multi_Ult_24_Nor	Multi_Ult_48_Nor	Multi_Meses_mora_9_30_ult48m	Multi_Meses_mora_31_60_ult48m	Multi_Meses_mora_61_90_ult48m	Multi_Meses_mora_may_90_ult48m	Multi_Meses_cobr_jud_ult48m	Multi_Meses_incobrable_ult48m	var_an_deuda
trabajadores	1																	
sectores	-0.14	1																
vfas_fis	0.55	-0.06	1															
dum_deuda_SF	0.26	0.08	0.15	1														
dum_saldo_al_dia	0.39	-0.03	0.24	0.76	1													
Nro_d_obliga	0.24	0.13	0.16	0.96	0.81	1												
Multi_Ult_6_Nor	0.22	0.11	0.18	0.87	0.72	0.91	1											
Multi_Ult_12_Nor	0.18	0.06	0.17	0.83	0.68	0.88	0.96	1										
Multi_Ult_18_Nor	0.20	0.09	0.16	0.80	0.64	0.84	0.92	0.96	1									
Multi_Ult_24_Nor	0.32	0.10	0.26	0.70	0.54	0.75	0.82	0.86	0.89	1								
Multi_Ult_48_Nor	0.33	0.12	0.19	0.55	0.45	0.60	0.66	0.69	0.72	0.81	1							
Multi_Meses_mora_9_30_ult48m	0.12	0.16	0.23	0.74	0.45	0.76	0.83	0.81	0.80	0.75	0.65	1						
Multi_Meses_mora_31_60_ult48m	0.22	0.11	0.18	0.87	0.72	0.91	1.00	0.96	0.92	0.82	0.66	0.83	1					
Multi_Meses_mora_61_90_ult48m	0.12	0.15	0.15	0.75	0.42	0.78	0.85	0.83	0.81	0.75	0.61	0.80	0.85	1				
Multi_Meses_mora_may_90_ult48m	0.12	0.15	0.14	0.77	0.48	0.80	0.88	0.85	0.82	0.76	0.60	0.76	0.88	0.95	1			
Multi_Meses_cobr_jud_ult48m	0.19	0.11	0.18	0.78	0.54	0.81	0.89	0.86	0.83	0.76	0.59	0.74	0.89	0.93	0.97	1		
Multi_Meses_incobrable_ult48m	0.22	0.11	0.18	0.87	0.72	0.91	1.00	0.96	0.92	0.82	0.66	0.83	1.00	0.85	0.88	0.89	1	
var_an_deuda	0.24	0.13	0.16	0.96	0.81	1.00	0.91	0.88	0.84	0.75	0.60	0.76	0.91	0.78	0.80	0.81	0.91	1

Nota. Elaboración propia (Resultados en SPSS)

Con respecto a la matriz de correlaciones, se puede apreciar que existen variables que mantienen un alto grado de correlación (>70%) con la gran mayoría del resto variables, tales como:

- dum_deuda_SF
- Multi_Meses_mora_9_30_ult48m
- Nro_d_obliga
- Multi_Meses_mora_31_60_ult48m
- Multi_Ult_6_Nor
- Multi_Meses_mora_61_90_ult48m

- Multi_Ult_12_Nor
- Multi_Meses_cobr_jud_ult48m
- Multi_Ult_18_Nor
- Multi_Meses_incobrabable_ult48m
- Multi_Ult_24_Nor
- var_an_deuda

Dichas variables se eliminaron en línea con lo expuesto. Por otro lado, se optó por dejar dentro del set de variables relevantes a Multi_Meses_mora_may_90_ult48m, puesto que al mantiene un indicador IV superior al resto de variables eliminadas. Finalmente, las variables resultado de la metodología de la selección de variables son:

- Trabajadores
- Sectores
- Vtas_fis
- Dum_saldo_al_dia
- Multi_ult_45_Nor
- Multi_Meses_mora_may_90_ult48m
- Mult_Ult_6_Nor

Sobre la variable “Multi_Ult_6_Nor”, esta se agregó en su forma original al set final de variables, decisión guiada por el *expertise* del modelador en cuanto a la necesidad y relevancia de discriminar en base a una variable multinomial que identifique si el locatario mantiene seis clasificaciones del tipo normal en los últimos seis meses.

2.6 Modelo de regresión logística

Tomando en cuenta las variables obtenidas en el apartado anterior, se procedió a realizar regresiones logísticas, aplicando la técnica *stepwise*, la cual es la construcción iterativa paso a paso de un modelo de regresión que implica la selección automática de variables independientes, este busca identificar relaciones entre las variables del modelo para tomar mejores decisiones y al mismo tiempo se procedió a eliminar las variables con signos contrarios a lo esperado, con lo cual se encontró un modelo con cinco variables explicativas (4 significativas) y con los siguientes coeficientes:

Tabla 8*Resultados del modelo de regresión*

Variable	Coefficient	Std. Error%	Z value	Pr(> z)1/
Multi_Meses_mora_may_90_ult48	2.7535	1.196	2.301	0.0214*
Multi_ult_6_Nor	-2.769	1.370	-2.021	0.0433*
Dum_saldo_al_dia	3.645	1.936	1.992	0.0598
Sectores	-0.918	0.551	-1.665	0.0959
Vtas_fis	0.499	0.381	1.308	0.1908

Nota. 1/Nivel de significancia: 0.05(*); 0.1(.). Elaboración propia (Resultados en SPSS)

Como se puede observar, las 2 primeras variables son significativas al 5%, las 2 siguientes son significativas al 10%, mientras que la última variable (vtas_fis) no es significativa. Sin embargo, se decidió dejar dentro del modelo, ya que muestra un aporte importante en la predictibilidad del modelo medido a través de la curva de ROC, KS y AIC.

2.7 Interpretación de las variables independientes

2.7.1 Multi_Meses_mora_may_90_ult48m

Es una variable multinomial creada para identificar el número de meses con mora mayor a 90 días durante los últimos 48 meses. Esta variable está compuesta de 3 categorías: el atributo 1 recoge los valores "Null" o nulos, que hace referencia a que no se cuenta con historial crediticio. La categoría 2 engloba los valores de "No tiene", el cual se refiere a que, si cuenta con historial crediticio, pero sin la presencia de meses con mora mayor a 90 días. Finalmente, la categoría 3 comprende aquellos valores de "Tiene", el cual significa que sus últimos 48 meses posee 1 o más meses con mora mayor a 90 días.

Sobre lo mencionado, se espera una relación positiva con la variable dependiente, puesto que cuando la variable alcance un valor de 3 (referente a 1 o más meses con mora mayor a 90 días en los últimos 48 meses), es más probable que el locatario muestre atraso mayor o igual a 2 meses.

2.7.2 Multi_Ult_6_Nor

Es una variable multinomial creada para identificar si el locatario mantiene 6 clasificaciones del tipo Normal en los últimos 6 meses. Esta variable está compuesta de 3 categorías: el atributo 1 recoge los valores "Null" o nulos, que hace referencia a la falta de historial crediticio suficiente en los 6 meses. Por otro lado, la categoría 2 engloba los valores de "No tiene", lo que se traduce en la existencia de una o más clasificaciones distintas a la Normal. Finalmente, la categoría 3 comprende aquellos valores de "Tiene", que significa que sus últimos 6 registros son normales.

Sobre lo mencionado, se espera una relación negativa con la variable dependiente, puesto que cuando la variable alcance un valor de 3 (que significa que sus últimos 6 registros son Normales), es más

probable que el locatario muestre un atraso menor a 2 meses.

2.7.3 Sectores

Es una variable binaria que posee 2 categorías. Se agrupan aquellos locatarios que mantienen negocios de comida, venta mayorista y minorista en la categoría 0, mientras que los locatarios que se desenvuelven en los sectores industrial, inmobiliario, público y telecomunicaciones, salud y otros son agrupados en la categoría 1.

En ese sentido, el signo negativo del coeficiente hace referencia a que los locatarios que pertenecen al grupo 1 (sectores industrial, inmobiliario, público y telecomunicaciones, salud y otros) son más probable que muestren un atraso menor a 2 meses. En efecto, los locatarios que desempeñan en la venta mayorista, retail y negocios de comida se encuentran más expuesto a un atraso mayor o igual a 2 meses.

2.7.4 *vtas_fis*

Es una variable multinomial que posee 3 categorías. La primera (0) engloba todos los datos que no se especifican por la Sunat, mientras que la segunda (1) categoría comprende aquellos locatarios que han reportado ventas fiscales menores a 73 millones de soles. Finalmente, la última categoría (2) recoge aquellos locatarios que mantienen ventas mayores a 73 millones de soles.

Sobre la razonabilidad de la relación positiva de esta variable con la variables dependiente, es relevante señalar que la variable se construye con información desfasada conforme el registro de ventas reportadas el año anterior, es comprensible que exista una relación positiva con la variable dependiente, toda vez que aquellas empresas con un nivel de facturación mayor a 73 millones de soles sostuvieron una fuerte carga para mantener la cadena de pagos y pasivos el año pasado o por otro lado, se podría argumentar que los locatarios con esta característica mantienen márgenes de ganancia más ajustados (como los mayoristas) o un descalce en los días de pago de pasivos que conlleva al atraso. En términos generales, la relación muestra que aquellas empresas que una facturación mayor a 73 millones de soles tienen una mayor probabilidad de mostrar un atraso mayor o igual a 2 meses.

2.7.5 *dum_saldo_al_dia*

Es una variable binaria que toma el valor de 1 cuando las acreencias financieras se encuentran 100% al día, por otro lado, coloca 0 cuando el resultado es diferente (nulo o saldo con atraso).

Al respecto, es preciso mencionar que la muestra utilizada para crear el presente modelo se corresponde con operadores que en su momento han pasado por un proceso de evaluación, el cual se corresponde con locatarios que usualmente son cadena a nivel nacional, pertenecen a un grupo accionarial sólido y que muy probablemente mantengan como principal fuente de fondeo al sistema financiero. Sin embargo, lo mencionado no los excluye de mostrar un atraso mayor a 2 meses en el

pago del arriendo, por el contrario, la obligación del pago del arriendo puede contar con desfases asociados al descalce en el flujo de efectivo de la empresa o por temas operativos (desconocidos).

De acuerdo con lo mencionado, no resulta absurdo que el coeficiente tenga signo positivo, dejando como inferencia que aquellos locatarios que mantiene un saldo al día en sus obligaciones financieras, es muy probable que muestren un atraso mayor o iguala a 2 meses. Finalmente, es relevante señalar que el uso de esta variable de forma aislada podría desnaturalizar su relación, ya que los coeficientes y signos obtenidos son congruentes con la regresión multivariada (de forma conjunta) expuesta en el apartado anterior.

2.8 Predictibilidad del modelo - Análisis de ajuste y discriminación

Al respecto, se consideraron los indicadores de Kolmogorov-Smirnov (KS), el área bajo la curva ROC y adicionalmente, el Criterio de Información de Akaike (AIC) para verificar el nivel de predicción del modelo. El modelo mostrado en el apartado 6, es aquel que muestra un mayor KS, curva de ROC y AIC. Como se puede apreciar en la Figura 1, el área bajo la Curva de ROC (AUCROC) nos indica que el modelo clasifica a los clientes de manera adecuada el 71.10% de las veces. Asimismo, su estadístico KS es mayor al 30%, lo cual permite señalar el buen nivel predictivo del modelo. De acuerdo con lo expuesto en los párrafos previos, se puede verificar a través de los estadísticos Curva de ROC (AUCROC), KS y AIC que el modelo de regresión logística propuesto presenta buenos niveles de ajuste y discriminación.

Tabla 9

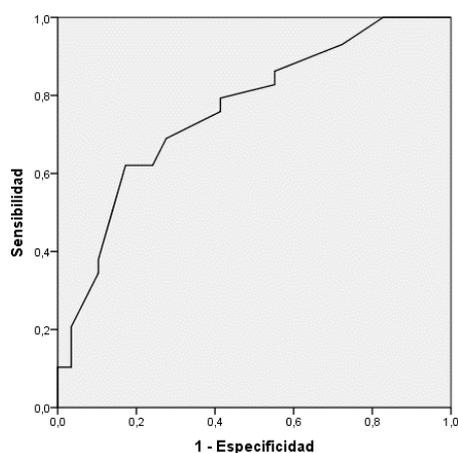
Tabla de ROC

	N° variables	AIC	AUC	KS	N° OBS
Modelo	5	77.746	76.10	44.84%	58

Nota. Elaboración propia (Resultados en SPSS)

Figura 1

Curva de ROC



Nota. Elaboración propia (Resultados en SPSS)

Conclusiones

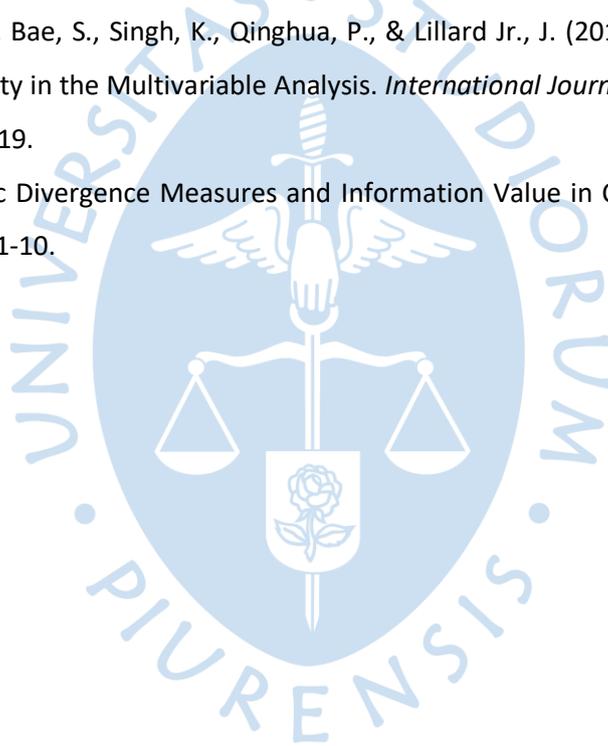
En el presente estudio intentamos crear un modelo que nos permita predecir si el nuevo locatario será una contraparte óptima dentro del portafolio, buscando encontrar las variables relevantes dentro de las características halladas en su historial crediticio, así como los datos de creación y detalle de la contraparte.

Esta investigación ha ayudado a entender a mayor profundidad la relación entre un nivel crediticio aceptable junto al historial de cobranzas de la cartera de inmuebles que mantiene la compañía. Pudimos encontrar cinco variables relevantes para la elección de locatarios óptimos para la cartera de inmuebles, así como para filtrar a potenciales operadores en la toma de decisión del comité de inversiones. Siendo una porción importante dentro del portafolio de inversiones, la gestión de cobranzas impacta directamente a los resultados por lo que el modelo de clasificación apoya el proceso de selección de locatarios.



Referencias

- Halim, S., & Humira, Y. (2014). *Credit Scoring Modeling*. Petra Christian University, repository.petra.ac.id. Surabaya - Indonesia: ISSN 1411-2485. Obtenido de <http://repository.petra.ac.id/id/eprint/16588>.
- Siddiqi, N. (2006). *Credit Risk Scorecards, Developing and Implementing Intelligent Credit Scoring*. Hoboken, NJ: John Wiley & Sons, Inc.
- Tineo, A. (2019). *Modelo de clasificación de clientes bancarizados de los niveles socioeconómicos "C" y "D"*. Tesis de Maestría en Finanzas, Universidad ESAN, Programa de la Maestría en Finanzas, Lima. Obtenido de <https://repositorio.esan.edu.pe/handle/20.500.12640/1566>.
- Tseng-Chung, T., & Li-Chiu, C. (2005). Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy Logic models using ROC curve analysis. *Expert Systems with Applications*, 547-556.
- Yoo, W., Mayberry, R., Bae, S., Singh, K., Qinghua, P., & Lillard Jr., J. (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. *International Journal of Applied Science and Technology*, 9-19.
- Zeng, C. (2013). Metric Divergence Measures and Information Value in Credit Scoring. *Journal of Mathematics*, 1-10.



Notas a pie de página

- ¹ Indicador information value = $\ln(\%good/\%bad) * (\%good - \%bad)$
- ² Indicador weight of evidence = $\ln(\%good/\%bad)$

