



UNIVERSIDAD
DE PIURA

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**Diseño de *scoring* crediticio: una aplicación para
seguros de crédito**

Trabajo de Suficiencia Profesional para optar el Título de
Economista

Jimena Antonia Córdova Amaya

Revisor(es):

Mgtr. Harry Omar Patrón Torres

Lima, febrero de 2021



A mis padres, y a mi tío Paolo por todo su apoyo.





Resumen

Para una aseguradora de crédito, el monitoreo de la cartera de deudores para reducir el riesgo de impago, así como verificar la solvencia de la empresa de crédito son factores importantes para la compañía. La presente investigación, busca construir un modelo de seguimiento que sea capaz de identificar cuáles son los factores que predigan el cumplimiento de las obligaciones de pago.

Se han aplicaron los modelos de elección discreta, los cuales clasifican a los deudores en dos grupos *Default* y *No-Default*. Asimismo, se estimaron los modelos de regresión logística y el modelo de ensamblado de árboles.

De igual manera, la muestra se divide aleatoriamente en dos partes; una parte para ajustar los modelos y la otra para validar el poder de predicción de los modelos estimados.

El modelo de ensamblado de árboles se presenta como la herramienta efectiva para la predicción de *default* para esta investigación, no solo a nivel de capacidad de discriminación, sino como una herramienta de fácil entendimiento que permitiría potenciar su uso. A diferencia del modelo de regresión logística, el ensamblado de árboles no permite cuantificar la significancia de la variable explicativa en la predicción del incumplimiento.

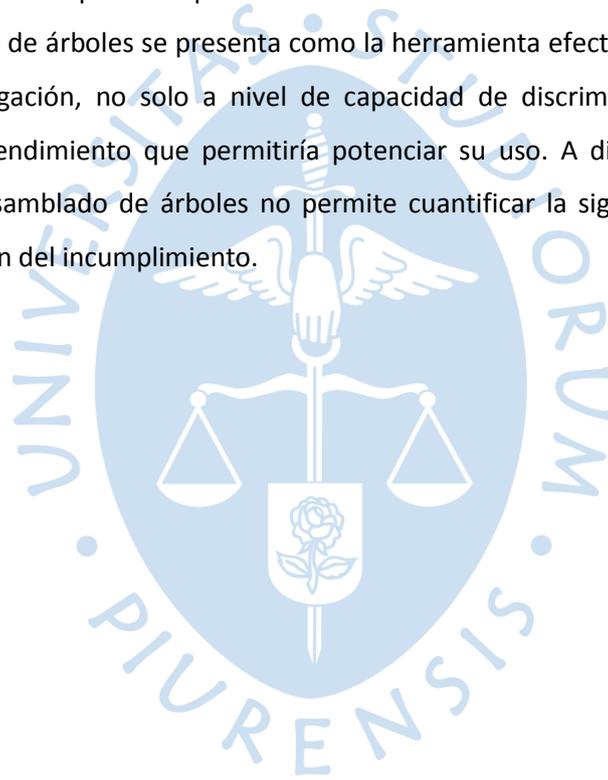




Tabla de Contenido

Introducción.....	13
Capítulo 1. Informe Experiencia Profesional.....	15
1.1. Experiencia profesional Compagnie Francaise d'Assurance pour le Commerce Extérieur (COFACE) Perú.....	15
1.1.1. Descripción de la empresa.....	15
1.1.2 Fundamentación.....	16
1.1.3 Aportes y Experiencias.....	17
1.2 Experiencia profesional en Avla Perú Compañía de Seguros.....	17
1.2.1 Descripción de la empresa.....	17
1.2.2 Fundamentación.....	18
1.2.3 Aportes y experiencias.....	19
Capítulo 2. Trabajo del módulo de Economía Financiera.....	21
2.1 Scoring de Crédito.....	21
2.1.1 Definición.....	21
2.1.2 Antecedentes.....	21
2.2. Seguro de Crédito.....	21
2.3 Datos y Variables.....	22
2.3.1Análisis Descriptivo de la Muestra.....	26
2.3.2 Influencia de cada variable con el Default.....	29
2.4. Estimación del Modelo.....	31
2.4.1 Estimación del Modelo Logístico.....	31
2.4.2 Ensamblado de árboles.....	32
2.4.3 Resultados.....	32
Conclusiones.....	35
Referencias.....	37
Apéndice.....	39
Apéndice A: Correlación de Variables-Mapa de Calor.....	39
Apéndice B: Relación entre pares de variables (Experiencia-Trabajo-Nivel de Bancarización).....	40
Apéndice C: Importancia de las variables- Ensamblado de árboles.....	41
Apéndice D: Validación Cruzada-Ensamblado de árboles.....	42
Apéndice E: Ensamblado de árboles: Comparación de Técnicas.....	43
Anexos.....	45
Anexo A: Constancia de Experiencia Laboral- COFACE Perú.....	45
Anexo B: Constancia de Experiencia Profesional-Avla Perú Compañía de Seguros.....	46



Lista de tablas

Tabla 1: Distribución de los deudores en Default y No-Default.....	23
Tabla 2: Variables disponibles en la muestra.....	24
Tabla 3: Variables Cuantitativas.....	26
Tabla 4: Variables Cualitativas.....	28
Tabla 5: Factores versus en <i>Default</i>	29
Tabla 6: Variables versus en <i>Default</i>	30
Tabla 7: Resultados Modelo Logístico.....	32
Tabla 8: Resultados Árbol de Decisión.....	33
Tabla 9: Comparativo de <i>Accuracy</i> y training time para las técnicas de árbol de decisión	43





Lista de Figuras

Figura 1: Presencia del Grupo Coface.....	15
Figura 2: Participación del Mercado Seguro de Crédito.....	18
Figura 3: Esquema de funcionamiento de Seguro de Crédito.....	22
Figura 4: Curva ROC de la Regresión Logística y Árbol de Decisión.....	33
Figura 5: <i>Accuracy vs Training time</i>	44





Introducción

Uno de los activos más importantes de una empresa es la cartera de cuentas por cobrar, el cual constituye uno de los flujos principales del negocio.

El monitoreo de la cartera de deudores hace posible verificar la solvencia, minimizando así el riesgo de impago, el cual es de vital importancia para una entidad aseguradora de crédito. Ahora bien, la interrogante que da inicio a la construcción del modelo de seguimiento es ¿cuáles son los factores (atributos) económicos, sociales que hacen que un cliente incumpla con sus obligaciones de pago?, es decir ¿Cuál son los indicadores más adecuados para predecir el impago de un deudor?.

La estructura de esta investigación se divide en dos capítulos; el primero se enfoca en el Informe de Experiencia Profesional y el segundo en el desarrollo de la investigación titulada *Diseño de Scoring crediticio: una aplicación para seguro de crédito*.

En esta investigación, la primera parte se compone de aspectos teóricos sobre el *scoring* de crédito y metodología de modelación. Seguidamente, se define el seguro de crédito y sus implicancias.

Luego, en la tercera parte, se estiman los modelos estadísticos utilizados como el modelo logístico y el modelo de ensamblado de árboles, mediante el uso y aplicación de funciones del paquete estadístico en Python. Siendo el caso, para este estudio empírico se contó con una base de datos proporcionada por una entidad aseguradora de crédito peruana.

Finalmente, se ha comparado los modelos utilizados teniendo en cuenta medidas de bondad de ajuste como AIC (Akaike, 1973), BIC (Schwartz, 1978). Concluyendo que, el resultado de los modelos fue similar en cuanto al poder de predicción siendo más riguroso el ensamblado de árboles para predecir si el cliente entrará en default.



Capítulo 1. Informe Experiencia Profesional

En este capítulo, se realiza una breve descripción de la experiencia profesional del autor desde el año 2017 a la actualidad. Asimismo, el autor menciona cómo los conocimientos adquiridos en el programa de Economía le han brindado las herramientas necesarias para ejecutar correctamente las funciones de los cargos ejercidos. De igual manera, se describen las experiencias y aportes realizados que le han permitido adquirir mayores conocimientos e intercambiar aprendizajes.

1.1. Experiencia profesional *Compagnie Francaise d'Assurance pour le Commerce Extérieur* (COFACE) Perú

1.1.1. Descripción de la empresa

El grupo *Compagnie Francaise d'Assurance pour le Commerce Extérieur*, en adelante COFACE, es uno de los líderes mundiales en seguro de crédito desde 1946, con presencia directa en 67 países e indirectamente - a través de sus socios - en más de 100 países, proporcionando apoyo a sus clientes con presencia global.

COFACE se crea en Francia como empresa especializada en el seguro de crédito de la exportación, incluyendo la gestión de garantías públicas en el nombre del Estado Francés e inicia su expansión internacional en 1992 con un desarrollo progresivo en numerosos países destacando Reino Unido, Alemania, Austria e Italia.

Figura 1: Presencia del Grupo COFACE Perú



Fuente: COFACE Perú. <https://www.coface.com.pe>

En 2006, el grupo COFACE se convierte en filial de Natixis segunda institución bancaria más grande de Francia conformada por banca corporativa, gestión de activos y servicios financieros de BPCE. Es así que desde 2014 COFACE cotiza en la Bolsa de París (Euronext) y, como elemento esencial para ejercer su actividad, COFACE mantiene su calificación AA por Fitch y A2 por Moody's.

La misión de COFACE es contribuir a que las empresas puedan fortalecer su capacidad de vender en el mercado doméstico e internacional, mediante la buena gestión de riesgos; es decir, maximizar las posibilidades de éxito, protegiendo del riesgo financiero inherente a su actividad. Toda vez que, el fortalecimiento de las empresas, contribuye también al buen funcionamiento de la economía. La visión COFACE es ser la aseguradora de crédito global más ágil de la industria.

Ahora bien, la historia de COFACE Perú como aseguradora de crédito comienza en 2016, ocupando el segundo puesto en participación de mercado hasta finales del 2018, año en el que se trasladaron las operaciones de Perú a COFACE Chile. Actualmente, en las oficinas de Perú se mantiene el centro de información de Latinoamérica, el cual ofrece servicios de reporte comerciales, homologaciones y recuperación de facturas vencidas.

La experiencia laboral del autor en COFACE Perú inicia en el año 2017 y culminó en diciembre 2018, bajo el cargo de *Account Executive* del área comercial. Siendo las principales funciones del autor, gestionar permanentemente el portafolio de asegurados locales, maximizar el índice de retención y rentabilidad de la cartera de seguro de crédito, así como también realizar ventas cruzadas de otros productos COFACE (reportes comerciales, recuperación de facturas vencidas, etc.).

1.1.2 Fundamentación

Para el ejercicio de las funciones asignadas, resultó de utilidad los cursos de finanzas corporativas y tópicos de la economía de la empresa; ya que al finalizar el contrato de seguro con cada cuenta, se debía revisar el resultado técnico como el índice de siniestralidad, riesgo de la cartera asegurada, prima devengada y ofrecer condiciones que aseguren la retención del clientes; es decir, obtener una propuesta competitiva y en paralelo proteger la rentabilidad de la cartera.

Para ello se debe realizar los análisis macroeconómicos enfocados en el sector de desarrollo tanto de la empresa asegurada como de la cartera de clientes. En el caso de asegurados con mercado de exportación, se debe revisar y ponderar el riesgo-país de la cartera de clientes. Asimismo, para determinar las condiciones a ofrecer se utilizaba un sistema de tarificación que recogía las variables macroeconomías, el resultado técnico y el riesgo de la cartera asegurada. De igual manera, en la interpretación y ajuste de condiciones las lecciones de Econometría y probabilidades fueron esenciales.

1.1.3 Aportes y Experiencias

Una de las bases fundamentales que tiene el grupo COFACE para lograr sus objetivos como empresa, es la constante capacitación de sus colaboradores.

Siendo el caso, desde la incorporación del autor a la institución, recibió formación sobre el seguro de crédito tanto capacitaciones *online* como presenciales en las cuales gerentes corporativos-tanto de la matriz en Francia como de la oficina regional en México u otras entidades de la región-constantemente actualizaban los conocimientos de la materia y la interpretación de los lineamientos de suscripción comercial. Incluso en el año 2018 COFACE Perú fue incluido en el programa *COFACE School*, plataforma que permitía acceder de lecciones sobre todos los productos y servicios del Grupo COFACE, tanto en el ámbito de suscripción, comercial y uso de las plataformas informáticas de COFACE.

Debido a las nuevas disposiciones y estrategias de negocio de la casa matriz, a finales del 2018, el corporativo decidió consolidar y transferir las operaciones de seguro de crédito peruanas a COFACE Chile; es decir, COFACE Perú se retiraba del mercado y el portafolio de clientes asegurados debía migrar a la cartera de Chile. Para ello, el autor tuvo que realizar una ardua gestión comercial, se debía comunicar cuidadosamente a los clientes el retiro físico de la empresa explicando con claridad los motivos sin que genere incertidumbre por el respaldo de la aseguradora. Asimismo, se debía alentar a los clientes a continuar con la relación comercial con el Grupo COFACE.

A nivel contractual, pese a pertenecer al mismo grupo, se tuvieron que emitir y negociar nuevas condiciones, por disposiciones legales de cada país, los aspectos del contrato diferían al vigente. En esta transición la formación académica e integra de la Universidad de Piura, en específico el curso de Ética para los Negocios tuvo un rol importante, debido a que se requería ser lo más transparente con el cliente y al mismo tiempo cumplir con los lineamientos de la empresa.

1.2 Experiencia profesional en Avla Perú Compañía de Seguros

1.2.1 Descripción de la empresa

Avla Perú Compañía de seguros, en adelante Avla, es un grupo asegurador pionero en ofrecer soluciones integrales en Latinoamérica a través de Seguros de Crédito, Seguros de Ingeniería, Garantías Financieras y Cartas Fianzas. Actualmente, cuenta con oficinas en Chile, México y Perú.

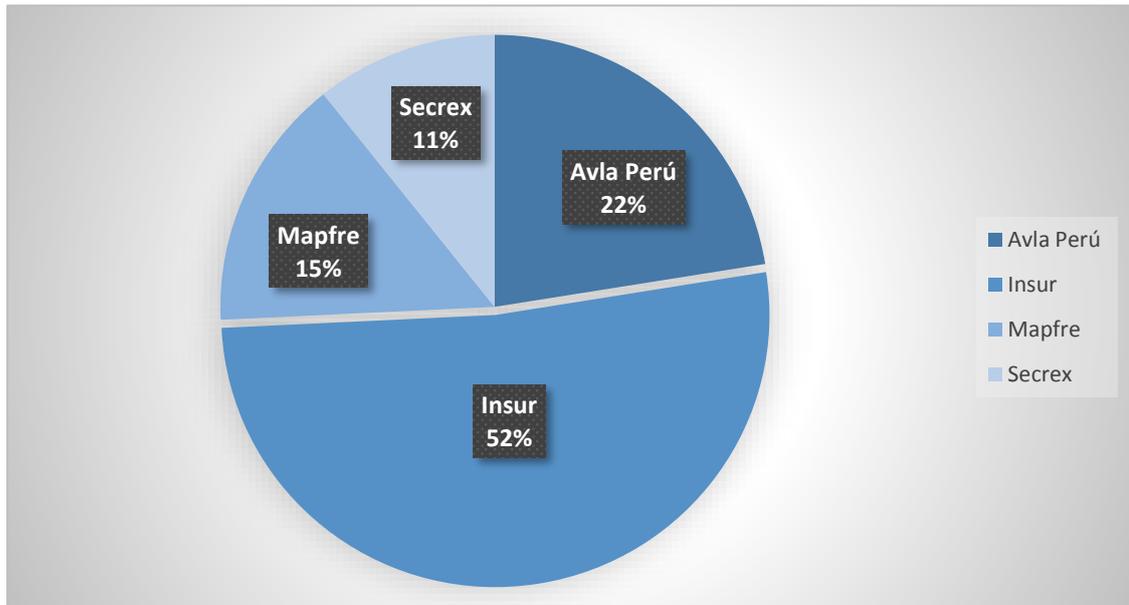
El objetivo de la empresa es respaldar a pequeñas, medianas y grandes empresas a través de un soporte integral brindando servicios financieros que contribuyan a su crecimiento y bienestar financiero. Avla busca ser el líder a nivel latinoamericano en la entrega de soluciones financieras a fin de contribuir con el desarrollo económico y social de la región.

El grupo Avla cuenta con el apoyo de 19 prestigiosos reaseguradores internacionales que proporcionan respaldo, experiencia y estabilidad a la aseguradora. Evidencia de su fortaleza financiera,

es que cuenta con calificación A- por las clasificadoras de riesgo Feller, Humphreys e ICR. Asimismo, en el año 2020 Avla empezó a cotizar en la Bolsa de Comercio de Santiago.

Desde el inicio de las operaciones en Perú en el año 2017, Avla ha demostrado ser una empresa sólida con objetivos definidos; prueba de ello es que, actualmente cuenta con el segundo puesto en la participación de mercado de seguro de crédito. De igual manera, Avla ocupa puestos similares en los otros productos que ofrece.

Figura 2. Participación de Mercado de Seguro de Crédito



Fuente: SBS-Primas Netas Acumuladas Noviembre 2020
Elaboración propia

La participación de autor en la empresa dio inicio en el año 2019 desempeñando, hasta la actualidad, el cargo de Ejecutivo Post Venta. Algunas de las funciones a cargo son la retención de las pólizas de seguro de crédito; así como también salvaguardar la relación de Avla con sus clientes claves y corredores, velando por una suscripción comercial prudente y transparente tomando en cuenta el índice de siniestralidad de las cuentas claves a fin de obtener resultados rentables.

1.2.2 Fundamentación

La clave para tener un buen desempeño en las actividades asignadas al autor fue la formación analítica que forjo en la etapa académica, cursos como estadística, gestión de portafolio e informática, son pilares para el desarrollo óptimo del cargo. Asimismo, por la naturaleza del cargo, las habilidades interpersonales que priorizó la Universidad de Piura en la formación toman un rol importante en el desempeño del cargo; ya que, se mantiene comunicación directa con clientes y corredores.

1.2.3 Aportes y experiencias

El contar con experiencia en el rubro, le permite desenvolverse mejor en este cargo asumiendo nuevos retos. En esta etapa ha reforzado la gestión comercial, mejorando la capacidad de respuesta. Actualmente, maneja una vasta cartera de clientes de sectores diversificados. Asimismo, el autor puede conocer otros productos que ofrece la empresa tales como Cartas Fianzas, Seguro de Ingeniería, entre otros.

En la actualidad, venimos atravesado un escenario *sui generis* por motivo de la pandemia, dicha coyuntura ha elevado la exposición de riesgo del producto de seguro de crédito, la cual al ser alta afecta directamente a la cadena pagos entre empresas. A nivel general las condiciones de suscripción tuvieron ajustarse, significando un nuevo reto para la empresa, específicamente al equipo de seguro de crédito en términos de mitigar la exposición sin perjudicar al cliente.



Capítulo 2. Trabajo del módulo de Economía Financiera

2.1 Scoring de Crédito

2.1.1 Definición

Los modelos de *scoring* son instrumentos de clasificación o puntuación utilizados por las entidades financieras en la decisión de otorgar un crédito. La entidad financiera, dependiendo del riesgo que está dispuesto a asumir, decide si concede o no el crédito individual o empresarial; para ello, debe conocer la probabilidad que tiene cada solicitante de presentar morosidad (retraso en pagos).

2.1.2 Antecedentes

En general, los modelos predictivos aplicados al *credit scoring* incluyen métodos estadísticos como Altman (1968) que utilizó un análisis discriminante multivariado para el problema de predicción de la quiebra corporativa; el modelo Ohlson (1980) que aplicó la regresión logística la cual se convirtió en una metodología estándar para pronosticar la probabilidad de quiebra o el estado de *Default* de los créditos; el modelo West (2000) que aplicó métodos de inteligencia artificial; Huang, Chen y Wang (2007) emplearon una máquina de vectores de soporte y; Lee, Chiu, Chuo y Lu (2006) aplicaron el árbol de decisiones. Los modelos simples como el de discriminación lineal y regresión logística siguen siendo los métodos más usados debido a su implementación fácil y precisión, a pesar del gran avance de los otros modelos.

2.2. Seguro de Crédito

El Seguro de Crédito es un instrumento financiero que tiene como finalidad proteger a las empresas del riesgo de no pago de las cuentas por cobrar causado por créditos impagos vinculados a transacciones comerciales relacionadas al negocio, a través del monitoreo constante de la cartera.

La aseguradora otorga cobertura a las ventas a crédito, en caso de la insolvencia o el incumplimiento de pago por parte de su cliente (comprador) la aseguradora indemniza a la empresa.

Figura 3. Esquema de funcionamiento del Seguro de Crédito



Elaboración Propia

Para minimizar la probabilidad de impago se requiere de un monitoreo constante, preciso y efectivo; por ello, la aseguradora debe implementar herramientas que permitan mantener controlada la exposición de riesgo del portafolio.

Considerando que, el portafolio está conformado por empresas de diferente tamaño, estructura y variables predictivas (características) resulta necesario el uso de un modelo de seguimiento que priorice las variables relativas al comportamiento crediticio sobre las variables características propia de cada deudor.

2.3 Datos y Variables

Se cuenta con una muestra de 4,309 empresas que pertenecen a la cartera de deudores de una entidad aseguradora de crédito. Las predicciones de los modelos se interpretan en términos de la probabilidad de llegar a default una vez otorgado el crédito: *Scoring* de seguimiento.

Para la identificación de datos atípicos se examinó la distribución de las observaciones para cada variable, se estableció un umbral con valor estándar de 3 para la designación de caso atípico mediante el cálculo de puntuaciones tipificadas.

Como resultado se observaron que 58 casos surgieron por un error de codificación por lo que fueron retirados de la muestra a fin de no distorsionar el comportamiento estadístico de las variables. Asimismo, se identificaron 203 observaciones con ausencia de datos, los cuales fueron analizados con la finalidad de identificar la existencia de algún patrón no aleatorio en dicho proceso que pueda sesgar los resultados. La variable con datos ausentes es la de clasificación crediticia que otorga la compañía

de seguros. Esta variable es considerada, en el análisis preliminar, importante para la explicación del modelo, por lo cual se optó por excluirlos del análisis, evitando así algún sesgo.

La muestra contiene la variable de respuesta *Default* y 22 variables que describen atributos o características demográficas y financieras de los deudores, las cuales permiten explicar el fenómeno de *default*.

La variable *default* indica si el deudor presentó morosidad mayor a 60 días con alguno de los asegurados durante los últimos 12 meses. Es una variable dicotómica donde los deudores que no entraron en *default* se codificaron con el valor de 0 y los que cayeron en *default* con el valor de 1. En la muestra 1827 deudores cayeron el *default*.

Tabla 1. Distribución de los deudores en Default y No-Default

	No-Default	Default	Total
<i>Numero Deudores</i>	2,482	1,827	4,309
<i>Porcentaje</i>	58%	42%	100%

Fuente: Base de datos de la aseguradora
Elaboración Propia

Las variables independientes se seleccionaron de acuerdo a los factores que la entidad aseguradora considera para medir el nivel de riesgo de las operaciones de crédito que concede. En general, la información que las entidades examinan para evaluar el crédito son perfil socio-demográfico, la capacidad de endeudamiento, el comportamiento del cliente, situación financiera y patrimonial de la empresa.

Bajo la premisa que las empresas pertenecen a la cartera asegurada, la entidad cuenta con información directa respecto al comportamiento de crédito del deudor con el asegurado por lo que también serán consideradas para la estimación del modelo.

De acuerdo con lo mencionado en los párrafos anteriores, las variables independientes seleccionadas pueden ser clasificadas en:

- Factores del Comportamiento de Crediticio: variables relacionadas a la situación actual del cliente en el sistema financiero como son la calificación en centrales de riesgos, promedio de días de mora. Monto de protestos y moras reportadas.
- Factores Socio-demográficos: variables relativas a las características de la empresa como tipo y estado de contribuyente, tamaño de la empresa, años de constitución, número de trabajadores.
- Factores de Comportamiento con la aseguradora: relativas a la comportamiento del deudor con el asegurado desde el otorgamiento del crédito, tiempo del crédito, moras vigentes, número máximo de días en mora.

La tabla 2 describe las variables con su respectivo bloque, y especifica la fuente de información de la muestra.

Tabla2. Variables disponibles en la muestra

Factor	Variable	Nivel de Medición	Fuente
Comportamiento Entidad	AVLA_CLAS Clasificación interna de los deudores	Numérica Ordinal: De 1 a 8, siendo 1 la mejor calificación y 8 la peor.	Primaria
	AVLA_EXPERIENCIA Número de meses transcurridos desde la aprobación del crédito	Numérica	Primaria
	AVLA_ACT Monto de prórrogas vigentes	Numérica	Primaria
Socio- Demográficos	TAM_EMPR_CRED_SBS Tamaño de la empresa	Numérica Ordinal:1 Gran Empresa, 2 Mediana Empresa 3 Pequeña Empresa 4 Micro Empresa	Secundaria
	TIP_PER Tipo de Empresa	Cualitativa Nominal: Jurídica o Natural	Secundaria
	ESTADO_CONTRIBUYENTE	Cualitativa Nominal: Activa o de Baja	Secundaria
	EXPERIENCIA Años de constitución	Numérica en años	Secundaria
	TRABAJADORES Número de trabajadores	Numérica	Secundaria
	Comportamiento Crédito	SCORE_EQUI Calificación Central de Riesgos	Numérica Ordinal: De 0 a 999, siendo 999 lo mejor
Factor	Variable	Nivel de Medición	Fuente
	DIASATRASO_6M Promedio de días de atraso en los últimos 6 meses.	Numérica	Secundaria
Comportamiento Crédito	MONTO_PROT_NA_ULT24M	Numérica	Secundaria

	Monto de Protestos de en los últimos 24 meses		
	MONTO_MOROSO_ULT24M	Numérica	Secundaria
	Monto de Moras de en los últimos 24 meses		
	DEUDA_DIRECTA_TOTAL	Numérica	Secundaria
	Nivel de Bancarización		
	MAX_DEUDDIREC_ULT12M	Numérica	Secundaria
	Máximo nivel de endeudamiento financiero en los últimos 12 meses		
Financieros	DEUDA_INDIRECTA_TOTAL	Numérica	Secundaria
	Nivel de endeudamiento con entidades no financieras en los últimos 12 meses		
	MAX_DEUDINDIREC_ULT12M	Numérica	Secundaria
	Máximo nivel de endeudamiento con entidades no financieras en los últimos 12 meses		
	RCC_CALF_NORMAL	Numérica	Secundaria
	Porcentaje de calificación normal en centrales de riesgos		
	RCC_CALF_CPP	Numérica	Secundaria
	Porcentaje de calificación cliente con problema potencial en centrales de riesgos		
Comportamiento Crédito	RCC_CALF_DEFICIENTE	Numérica	Secundaria
	Porcentaje de calificación deficiente en centrales de riesgos		
	RCC_CALF_DUDOSO	Numérica	Secundaria
	Porcentaje de calificación dudosa en centrales de riesgos		

RCC_CALF_PERDIDA

Porcentaje de calificación pérdida
en centrales de riesgos

Numérica

Secundaria

Fuente Primaria: Base de datos de la aseguradora

Fuente Secundaria: Centrales de Riesgos

2.3.1 Análisis Descriptivo de la Muestra

La tabla 3 proporciona el análisis descriptivo de las variables numéricas. En general, se observa que para la variable *Experiencia* (años) los deudores que cayeron en *default* no difieren de los que no cayeron en esta situación. En cuanto a *Número de trabajadores*, los deudores que cayeron en *default* tienen medianas y desviación estándar mayor a los que no cayeron en *default*. Los resultados demográficos de la muestra representan una situación contraria a lo que se esperaría. Normalmente se suele asociar la solidez de la empresa con los años de trayectoria o experiencia y el número de trabajadores.

Las variables relacionadas con el comportamiento crediticio en el sistema financiero, monto de protestos, moras, deuda directa e indirecta tienen valores mayores para los que cayeron en *default* sugiriendo que podría haber una influencia sobre la variable dicotómica estudiada. La calificación en centrales de Riesgo muestra mediana y media ligeramente mayor para los *No-Default*.

Tabla 3. Variables Cuantitativas

Variable	No-Default (2482 observaciones)			Default (1827 observaciones)		
	\bar{X}	X med	σ	\bar{X}	X med	σ
Score Centrales de Riesgo	538	803	169	530	788	184
Experiencia desde otorgamiento del crédito (años)	2	1	1	61	1	3
Experiencia(años)	45	16	11	43	17	12
Número de Trabajadores	5,178	109	483	9,112	256	1,064

Monto de protestos últimos 24 meses	2,711,711	26,298	171,125	13,561,415	69,739	842,469
Monto de moras últimos 24 meses	911,570	6,085	64,219	4,521,455	22,599	294,724
Nivel de Bancarización (Deuda Directa)	212,986,450	6,030,634	23,543,690	1,552,706,000	15,482,660	93,086,860
Máximo Nivel de Bancarización de los últimos 12 meses	248,625,418	9,063,173	33,423,230	1,641,656,036	23,343,490	114,518,600
Deuda Indirecta Total	469,218,400	2,341,768	29,393,210	832,495,000	7,162,662	60,743,650
Calificación en centrales de riesgo: Normal (%)	50	97	16	50	96	16
Calificación en centrales de riesgo: CPP (%)	50	2	14	50	3	14
Calificación en centrales de riesgo: Deficiente (%)	42	0	3	50	1	6
Calificación en centrales de riesgo: Dudoso (%)	50	0	3	50	0	3
Calificación en centrales de riesgo: Pérdida (%)	50	0	5	12	0	1

Fuente: Base de datos de la aseguradora- Elaboración Propia

La tabla 4 contiene el resumen de las variables cualitativas y la representación del default en cada una de ellas. La variable *Clasificación Avla* indica que las empresas con calificación A5 representan el 47.5% de las empresas que cayeron en *default*. Mientras que las empresas con menos reporte de

default fueron las A1. Es importante acotar que la variable *Clasificación Avla* es determinada en la admisión del crédito. Las variables que influyen en la clasificación son *Experiencia, Tamaño de la Empresa, Número de trabajadores, pertenencia a grupo económico, Nivel de Bancarización*, entre otros factores demográficos de la empresa y financieros (en caso hubiese). La clasificación es revisada periódicamente y/o cuando se solicita una revisión o nueva línea de crédito para el deudor. Para efectos del modelo de seguimiento a pesar de ser una variable con poder explicativo, podría generar problemas de multicolinealidad con el resto de variables explicativas.

Respecto a la variable *Tamaño de la empresa*, los resultados arrojan que, indistintamente del tamaño de la empresa, los porcentajes de *default* y *No-Default* son similares; lo cual sugiere que, el hecho de que el tamaño de la empresa no podría ser un factor de riesgo para el modelo de seguimiento.

En cuanto al *tipo de empresa y estado de contribuyente* los resultados son coherentes con los lineamientos de suscripción de admisión el cual no otorga crédito a empresas de baja y tiene controles estrictos para la aprobación de línea de crédito para personas naturales con RUC.

Tabla 4. Variables Cualitativas

Variable	<i>No-Default</i>		<i>Default</i>		Test χ^2 (Valor p)
	Número de deudores	Porcentaje	Número de deudores	Porcentaje	
Clasificación Avla					
A1	6	0.38%	7	0.30%	
A2	12	2.13%	39	1.18%	
A3	29	2.30%	42	1.65%	
					-6.340
A4	573	26.49%	484	24.53%	(0.338)
A5	1189	46.36%	847	47.25%	
A6	608	20.14%	368	22.65%	
B1	57	2.13%	39	2.23%	

Tamaño de Empresa					
Corporativo	64	2.58%	101	5.53%	
					3..376
Gran Empresa	437	17.61%	435	23.81%	(0.001)
Mediana Empresa	1271	51.21%	970	53.09%	
Pequeña Empresa	657	26.47%	292	15.98%	
Micro Empresa	53	2.14%	29	1.59%	
Tipo de Empresa					
Persona					
Natural con RUC	2	0.08%	5	0.27%	-0.884 (0.376)
Jurídica	2480	99.92%	1822	99.73%	
Estado Contribuyente					
					0.261
De Baja	1	0.04%	2	0.11%	(0.794)
Activo	2481	99.96%	1825	99.89%	

Fuente: Base de datos de la aseguradora -Elaboración Propia

2.3.2 Influencia de cada variable con el Default

Como una forma general de analizar la influencia de las variables agrupadas en los factores detallados anteriormente, en la tabla 5 se observa el seudo- R^2 de Mc-Faden (1974) para cada modelo ajustado de default contra las variables del factor.

Tabla 5. Factores versus el Default

Modelo: Default - variable del factor	Seudo- R^2
Factores del Comportamiento Crediticio: Días de atraso, Monto de Protestos y moras, calificación SBS	0.408
Factores Socio-Demográficos: Tamaño de la empresa, estado de contribuyente, tipo de deudor, experiencia, número de trabajadores.	0.386

Factores de Comportamiento en la Aseguradora:

Clasificación Avla, monto máximo de prórroga solicitada, monto de prórrogas vigentes, tiempo desde el otorgamiento del crédito	0.418
--	-------

Fuente: Base de datos de la aseguradora- Elaboración propia.

De acuerdo a la tabla 5 las variables de los componentes comportamiento crediticio y comportamiento en la aseguradora muestran un ajuste alto con un seudo de 0,408 y 0.418 respectivamente. Las variables socio-demográficas tienen un ajuste importante, pero bajo en comparación del resto, esto podría dar indicios de las variables que más aportan a la estimación del modelo, que a su vez guarda concordancia con el análisis descriptivo previo. Sin embargo, para la estimación de este modelo, las variables a utilizar serán seleccionadas independientemente al factor que pertenezcan ya que pueden estar fuertemente correlacionadas entre sí y alterar el grado de ajuste del modelo. Para ello, se aplicó la técnica de reducción de dimensionalidad y se obtuvo las variables que representan a cada factor. La tabla 6 muestra el pseudo- R^2 obtenido a través de regresiones multivariadas de las variables representativas.

Tabla 6. Variables versus la variable de Default

Factor	Variable	Seudo- R^2
Factores del Comportamiento Crediticio	Calificación normal SBS	0.453
	Monto de protestos	0.550
	Deuda Directa Total	0.227
Factores Socio- Demográficos	Estado de contribuyente	0.184
Factores de Comportamiento en la Aseguradora	Monto de prórroga Vigente	0.568
	Experiencia del deudor con la aseguradora	0.308

Fuente: Base de datos de la aseguradora- Elaboración propia.

2.4. Estimación del Modelo

El objetivo de esta investigación es construir un *scoring* crediticio aplicado al seguro de crédito a través de algunos modelos usualmente empleados como la regresión logística y el ensamblado de árboles. Para este estudio se pretende explicar la elección de cada deudor i entre dos alternativas posibles pagar o no pagar las facturas pendientes con el asegurado. $Y_i = \{0,1\}$, respectivamente, como función de unas variables que le caracterizan X_i 's y añadiendo un término de error que explique las diferencias entre los valores observados de Y_i y sus valores previstos.

Para ajustar los modelos se utiliza aleatoriamente el 40% de la muestra (1,720 deudores) y el 60% restante (2589 deudores) se deja para validar el poder de predicción de los modelos estimados. El punto de corte o umbral (C^*) establecido en la definición del evento estudiado fue de 0.50 para predecir $Y_i | X_i = 1$ si $F(\beta'X_i) > C^*$. Este valor del umbral se establece siguiendo la recomendación de Green (1992), quien sugiere tomar el umbral de 0,5 si la muestra cuenta con una proporción balanceada de 1 (*default*).

Para la selección de variables se tuvo en cuenta aspectos cualitativos y cuantitativas. En el análisis cuantitativo se observaron principalmente los niveles de significancia univariados. En cuanto al análisis cualitativos se verificó que las variables incluidas en los modelos correspondieran a la situación real de la entidad. Al tratarse de un modelo de seguimiento, a las variables asociadas a la demografía de la empresa (tamaño de empresa, experiencia, tipo de empresa) no se les asignó tanto peso en la estimación del modelo; ya que, estas variables fueron validadas para la admisión del crédito. En cambio, las variables como estado de contribuyente (activo, baja) sí fueron incluidas en el modelo ya que pueden cambiar en el tiempo.

Para la validación de los modelos se consideraron como criterios de información AIC, Seudo - R^2 Además, se validaron el poder discriminatorio y la calibración de cada modelo a través de la curva ROC y la matriz de confusión.

2.4.1 Estimación del Modelo Logístico

Para estimar el modelo logístico se utilizó la función *sklearn,linear_model* del paquete {stats} de Python. Las variables que resultaron significativas fueron *calificación normal en centrales de riesgo, monto de prórroga vigente en la aseguradora, deuda directa total, estado de contribuyente, monto de protestos y morosos en centrales de riesgo, y experiencia del deudor en la aseguradora*. Aunque la elección de variables es coherente puesto que el factor con mayor número de variables con influencia significativa en el modelo inicial fue el comportamiento crediticio y en el comportamiento en la aseguradora; fue necesario realizar un análisis para descartar la presencia de multicolinealidad, ello hizo descartar la variable de *Clasificación Avla* a pesar de mostrar alta significancia inicialmente.

2.4.2 Ensamblado de árboles

Para la elaboración del ensamblado de árboles se utilizó *xgb* del paquete *{xgboost}* de Python. El principio básico de este modelo es generar particiones recursivas por reglas de clasificación hasta llegar a una clasificación final, tal que es posible identificar perfiles (nodos terminales) en los que la proporción de clientes malos es muy alta (o baja) y de esta forma asignar su probabilidad.

Las premisas básicas para la construcción del modelo son la simplicidad, potencia y estabilidad. La simplicidad es un factor relevante en cuanto es muy importante para una entidad que cualquier colaborador en diferentes instancias sea capaz de comprender y entender por qué el modelo funciona y qué es lo que está prediciendo. La potencia medida en la capacidad de predecir y la estabilidad que el modelo funcione a través del tiempo y pueda detectar los cambios en la calidad de la cartera.

Las variables utilizadas para la elaboración del ensamblado de árboles, fueron las mismas que se establecieron para el modelo logístico.

2.4.3 Resultados

De acuerdo a la tabla 7 y 8, la capacidad predictivos del modelo logístico medida por el porcentaje de aciertos es del 62%. Por otro lado, la sensibilidad del modelo es de 10%, es decir, la capacidad para detectar los deudores que efectivamente entran en *default*. La especificad (proporción de deudores que efectivamente no llegaron a *default*) es de 98%, es decir, la capacidad para detectar a los clientes que no entraron en *default*. Todo indica que el modelo logístico es poco sensible para detectar los deudores efectivamente entrarán en *default*.

Tabla 7. Resultados Modelo Logístico

	Precision	Recall	F1-Score
No-Default	0.63	0.98	0.77
Default	0.74	0.10	0.17
accuracy			0.63
macro avg	0.69	0.54	0.47
weighted avg	0.67	0.63	0.53

Fuente: Base de datos de la aseguradora- Elaboración propia.

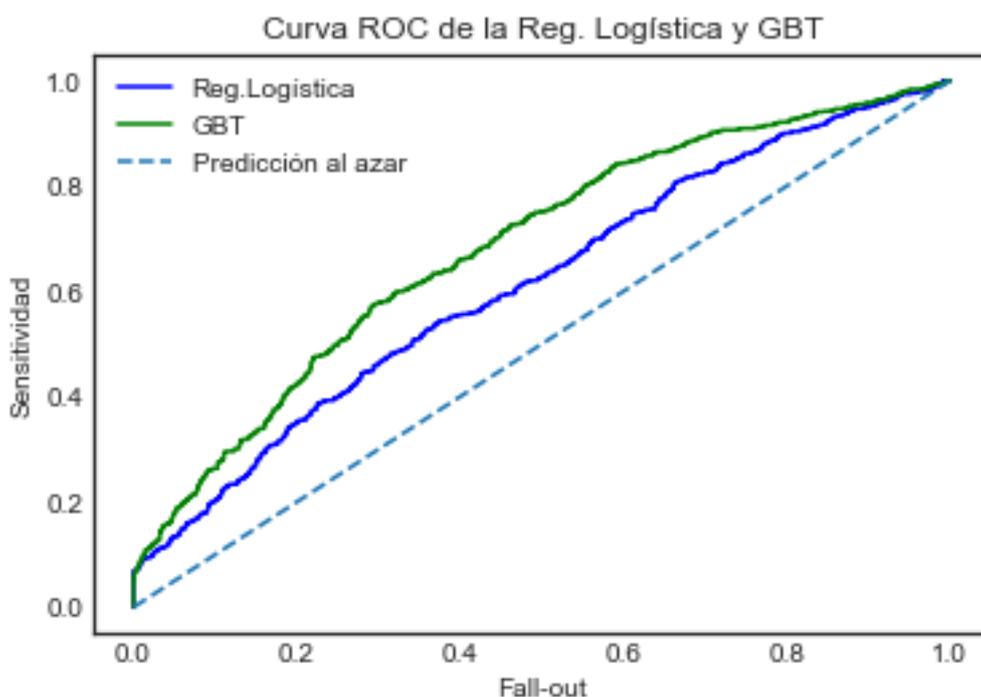
Respecto al modelo de ensamblado de árboles, la capacidad predictiva es de 68%, la capacidad para detectar los deudores que efectivamente entran en *default* es de 51% y la capacidad para detectar a los clientes que no entraron en *default* es de 72%.

Tabla 8. Resultados Ensamblado de árboles

	Precision	Recall	F1-Score
No-Default	0.70	0.74	0.72
Default	0.56	0.51	0.53
accuracy			0.65
macro avg	0.63	0.63	0.63
weighted avg	0.65	0.65	0.65

Fuente: Base de datos de la aseguradora- Elaboración propia.

Con base a las predicciones arrojadas por el modelo, se estimó la curva ROC, ilustrada en el gráfico 4. El AUC del modelo logístico es de 0.61 y el del ensamblado de árboles es de 0.68%.

Figura 4. Curva ROC de la Regresión Logística y ensamblado de árboles

Fuente: Base de datos de la aseguradora- Elaboración propia.

El modelo logístico tiene un desempeño aceptable en la calificación crediticia; sin embargo, algunas características propias de la muestra hacen que el modelo presente algunas limitaciones. El modelo logístico supone parámetros fijos o constantes, por lo tanto, no admite variaciones aleatorias en las preferencias de los individuos. Es decir, asume que todos los deudores tienen la misma respuesta ante un cambio en las variables explicativas. Asimismo, el modelo no admite correlación entre observaciones.

Por otr

o lado, el modelo de ensamblado de árboles tiene como beneficio clave la capacidad de revisar los datos e identificar patrones y tendencias que podrían no ser captados por otros modelos. Es esta capacidad del modelo, de identificar con éxito una relación causal entre dos eventos, la que podría explicar la mayor capacidad predictiva en relación al modelo logístico.



Conclusiones

Los dos modelos estimados logístico y ensamblado de árboles tienen similar poder discriminatorio; sin embargo, el de ensamblado de árboles resultó ser el de mayor sensibilidad para detectar deudores que efectivamente entrarán en *default*.

En cuanto a las variables que determinaran que un deudor entre en *default*, resultaron significativas las relacionadas con el factor de comportamiento de crediticio y con la aseguradora. Se descartaron algunas variables por multicolinealidad en el caso de *Clasificación Avla* o por problemas de tipificación, posiblemente existe error en registro. Las variables que explican el evento de llegar a *default* en los modelos ajustados, resultaron con signos acordes a los esperados.

El ensamblado de árboles se presenta como una herramienta efectiva para la predicción de este trabajo, no solo a nivel de capacidad de discriminación sino como una herramienta de fácil entendimiento que permitiría potenciar su uso. Sin embargo, una desventaja del modelo- en comparación a la regresión logística- es que no es posible cuantificar la magnitud con la que una variable aporta para la predicción del incumplimiento.

Para una entidad aseguradores de crédito es vital contar con una herramienta estadística adecuada para la predicción del comportamiento de los deudores, puesto que la rentabilidad de la misma en gran medida corresponde al correcto pago de las obligaciones comerciales contraídas por parte de los deudores hacia los asegurados.

La optimización y aplicación del modelo de seguimiento de riesgo, tendrá un impacto positivo en los siguientes aspectos:

- Incrementa la automatización del proceso de seguimiento del crédito.
- Disminuye los tiempos de revisión, el cual reduce el costo promedio de procesamiento por crédito.
- Permite tomar decisiones con mayor precisión, debido a la capacidad del modelo para mejorar con el tiempo.

Para futuras investigaciones es necesario disponer de una muestra que contenga la información tanto de los deudores con línea de crédito vigente como también a los que en algún periodo se canceló la línea de crédito.

Además, se debería consolidar en una base de datos la información financiera de las empresas, sobre todos las relativas a la disponibilidad de ingreso como son ingresos y nivel de endeudamiento. Ello con la finalidad de aumentar el poder discriminatorio del modelo y corregir el sesgo de la muestra.

Respecto al modelo, se podría combinar el modelo de *XGBoost* con otros modelos de ensamblado de árboles para generar un clasificador más potente. Asimismo, aplicar métodos de

optimización de hiperparámetros bayesianos para evitar el error de susceptibilidad que presenta el modelo.



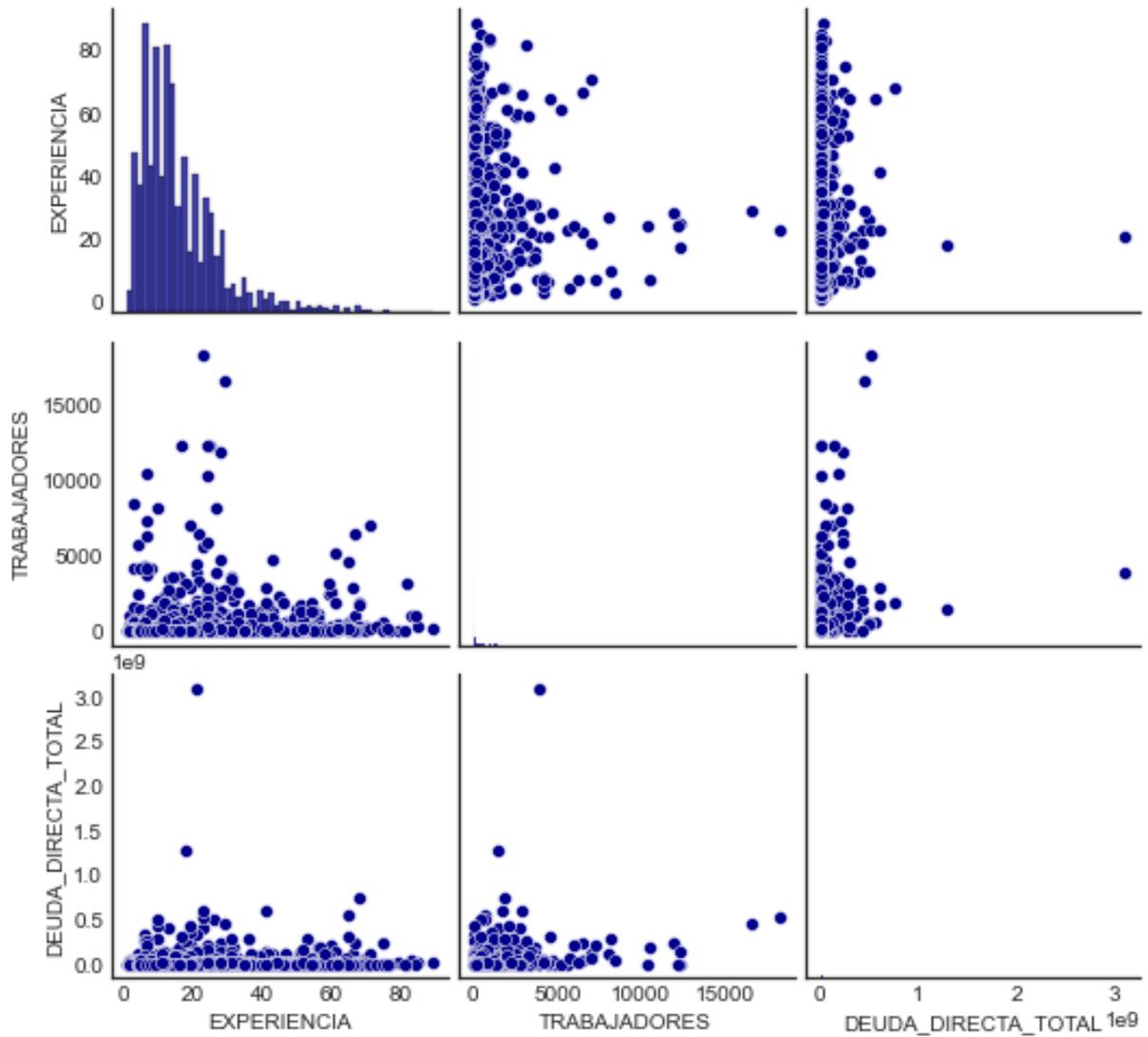
Referencias

- Boyle, M., Crook, J. N., Hamilton, R. and Thomas, L. C. (1992) *Methods for credit scoring applied to slow payers. en Credit Scoring and Credit Control* (pp. 75-90). Oxford: Clarendon.
- Balzarotti, V. y Castelpoggi, F. (2009). Modelos de puntuación crediticia: la falta de información y el uso de datos de una central de riesgos. Banco Central de la República de Argentina. *Ensayos Económicos*, 56, 95 – 156.
- Breiman, L., Friedman, J., Olsen, R & Stone, C. (1984), *Classification and Regression Trees*, Wadsworth. Belmont.
- Elith, J., Leathwick, J.R, & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802-813
- Greene, W. (1992). A statistical model for credit scoring. *Department of Economics Stern School of Business. New York University*.
- Grunert, J., Norden, L. and Weber, M., (2005). The role of non-financial factors in internal credit ratings, *Journal of Banking and Finance*, Vol. 29, No. 2, pp. 31–45
- Hsia, D. C. (1978) Credit scoring and the equal credit opportunity. *Hast. Law J.*, 30, 371-448.
- Joos, P., Vanhoof, K., Oeghe, H & Sierens (1998), Credit classification: a comparison of logit models and decision trees. *Faculteit Economie en Bedrijfskunde*
- Kolesar, P and Showers, J. L. (1985) A robust credit screening model using categorical data. *Management Sci.*, 31, 123-133.
- Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50, 1113–1130.
- Orgler, Y. E. 1970. A credit scoring model for commercial loans. *Journal of Money, Credit and Banking* II (4), 435-445.
- Safavian, S. R. and Landgrebe, D. (1991) A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.*, 21 660-674
- Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 26, 61–68.
- Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, 15, 757–770.
- Wu, T.-C., & Hsu, M.-F. (2012). Credit risk assessment and decision making by a fusion approach. *Knowledge-Based Systems*, 35, 102–110.

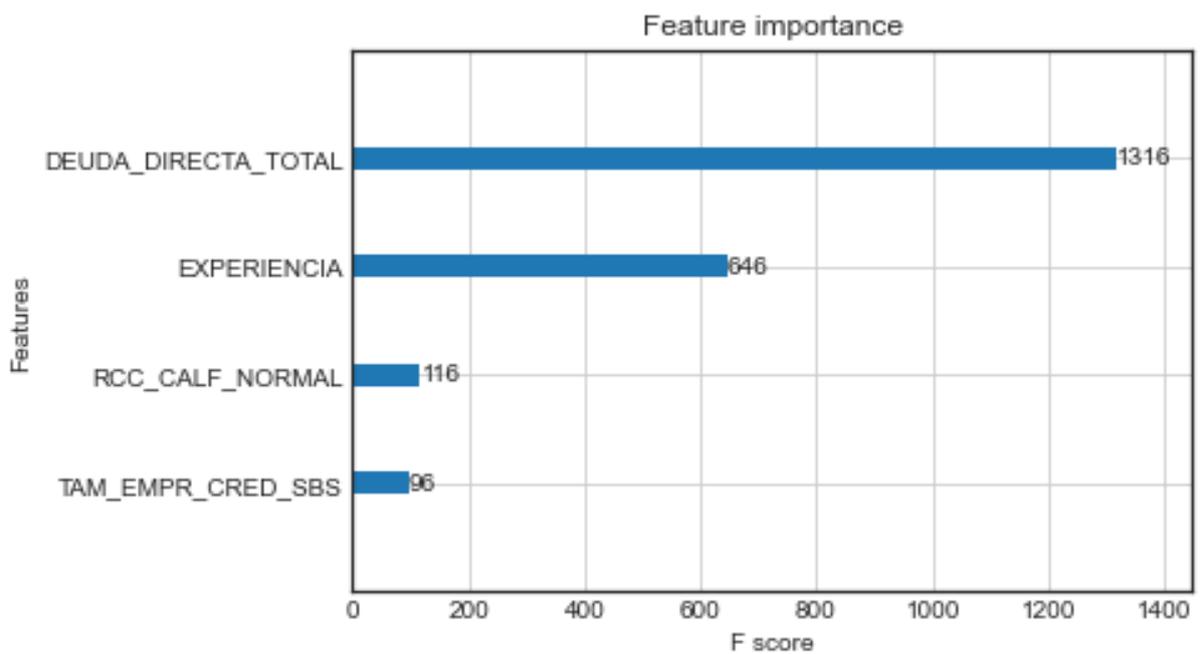
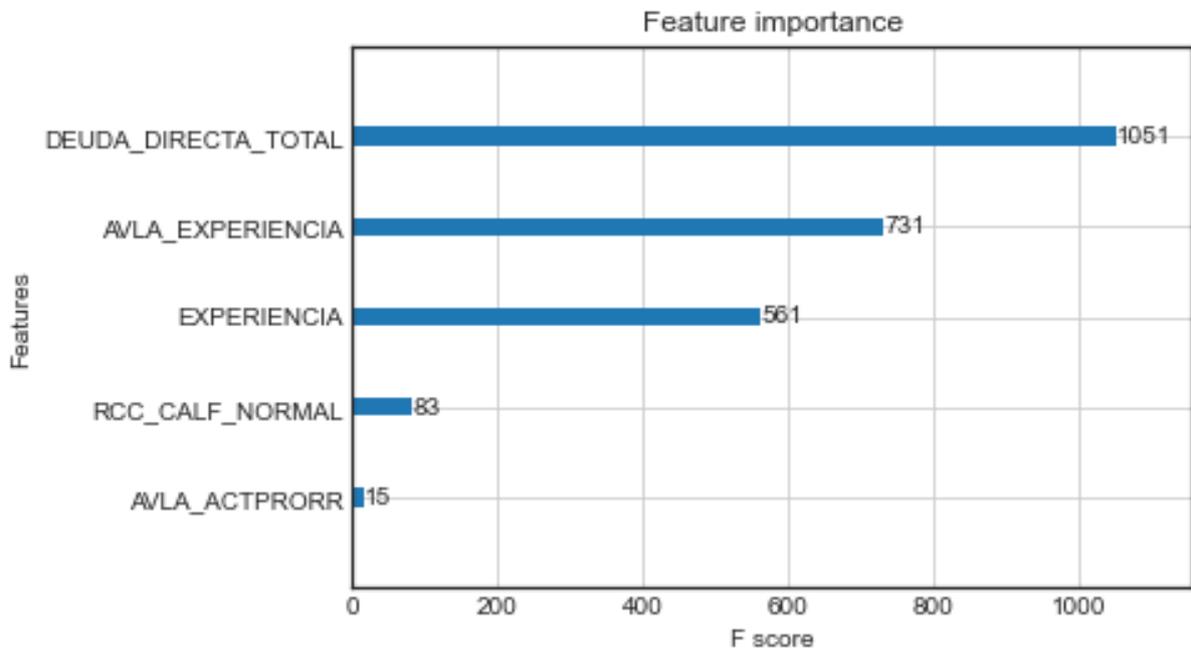


Apéndice B: Relación entre pares de variables (Experiencia-Trabajo-Nivel de Bancarización)

Relación entre pares de variables



Apéndice C: Importancia de las variables- Ensamblado de árboles



Apéndice D: Validación Cruzada-Ensamblado de árboles

Se construyó un *dataframe cv* a partir de 100 interacciones de validación cruzada. El siguiente gráfico muestra el score AUC para cada interacción. El promedio del *score AUC* es de 0.66.



Apéndice E: Ensamblado de árboles: Comparación de Técnicas

Como se mencionó en el documento los árboles de decisiones son algoritmos populares de aprendizaje automático aplicado para regresiones y clasificaciones.

Para el presente trabajo se utilizó la técnica de XGBoost debido a la robustez del modelo. Sin embargo, con el objetivo de comparar las actuaciones en términos de precisión y tiempo de entrenamiento presentamos la misma regresión aplicando la técnica de Bosques aleatorios (*Random Forest*) y *Gradient Boosting*.

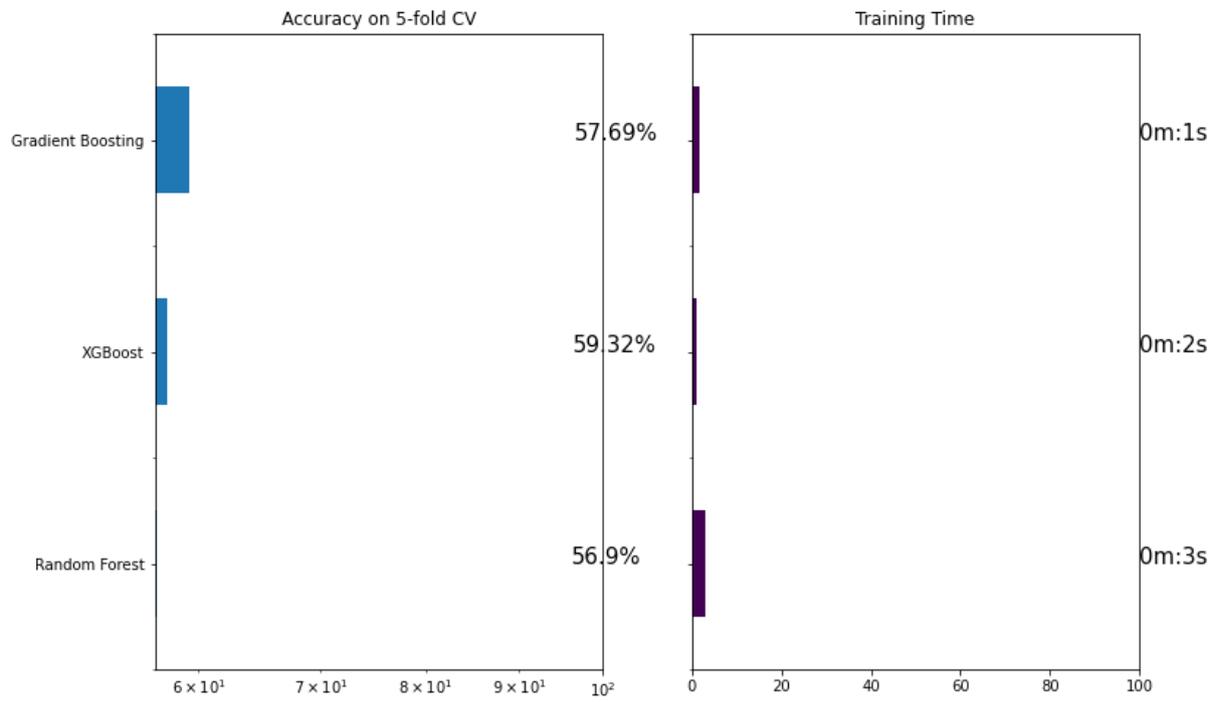
Random Forest Esta técnica usa el promedio de diferentes conjuntos de datos a fin de reducir la variación de los resultados. Algunas de las características del modelo es que predice eficientemente en base de datos grandes y tiene un método eficaz para estimar datos perdidos manteniendo la exactitud cuando una gran proporción de los datos está perdida. Sin embargo, en bases de datos con variables categóricas con diferentes niveles, este modelo se parcializa a favor de los atributos con alto nivel.

Gradient Boosting El funcionamiento de este modelo es construir una serie de árboles, en el que cada uno es una versión actualizada del anterior. Es decir, en cada interacción, se construye un árbol en el conjunto de datos (X, r) en lugar de (X, y) , donde “r” indica los residuos obtenidos por el árbol anterior. Posteriormente, se agrega una versión reducida de este clasificador a la anterior, y el proceso continuo hasta concluir el ciclo o cuando alcanza una determinada condición de ruptura.

XGBoost Es una versión “extrema” de *Gradient Boosting* en el sentido que es más eficiente, flexible y portátil y sobre todo más preciso.

Tabla 9. Comparativo de Accuracy y training time para las técnicas de árbol de decisión

	Accuracy (on 5-fold CV)	Training Time (seconds)
<i>XGBoost</i>	57.688551	1.12241
<i>Random Forest</i>	56.897176	2.945781
<i>Gradient Boosting</i>	59.316674	1.59768

Figura 5: Accuracy vs Training Time

Anexo A: Constancia de Experiencia Laboral- COFACE Perú



Lima, 08 de enero de 2019

CERTIFICADO DE TRABAJO

Por la presente certificamos que el(la) Sr(a). **CORDOVA AMAYA, JIMENA ANTONIA**, identificado(a) con DNI N° **76339119** ha trabajado en nuestra Empresa COFACE SEGURO DE CREDITO PERU S.A. con RUC N° 20600524233 desde el 16 de octubre de 2017 hasta el 08 de enero de 2019, desempeñando el cargo de **EJECUTIVO DE CUENTAS** en el área de **COMERCIAL**.

Se expide el presente certificado a solicitud del interesado para los fines que crea conveniente.

Atentamente,


COFACE
José Luis Rendón
Gerente General


Jimena Cordova Amaya
76339119

Anexo B: Constancia de Experiencia Profesional-Avla Perú Compañía de Seguros



CONSTANCIA DE TRABAJO

15 de enero de 2021

A QUIEN CORRESPONDA:

AVLA PERÚ COMPAÑÍA DE SEGUROS S.A. identificado con RUC Nro. 20600825187, mediante la presente se deja constancia que el (la) Señor(ita) CORDOVA AMAYA JIMENA ANTONIA, identificado(a) con DNI 76339119 labora en nuestra empresa desde el 09 de enero del 2019; desempeñando el cargo de EJECUTIVO DE POST VENTA.

Se expide la presente constancia a solicitud del (la) interesado(a) para los fines que estime conveniente.

Atentamente.

AVLA PERÚ COMPAÑÍA DE SEGUROS S.A.

AVLA PERÚ COMPAÑÍA DE SEGUROS S.A.