

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

PRONÓSTICO DE LAS TASAS DE CURA E INCUMPLIMIENTO: UNA  
APLICACIÓN DE LA PRUEBA DE ESTRÉS A LA CARTERA CREDITICIA

Trabajo final de investigación aplicada sometido a la consideración de la Comisión  
del Programa de Estudios de Posgrado en Estadística para optar al grado y título de  
Maestría Profesional en Estadística

FIGURELLA MARÍA LAURITO TORRES

Ciudad Universitaria Rodrigo Facio, Costa Rica

2025

## DEDICATORIA

A mi madre, Jenory, por inculcarme la importancia del estudio como clave para mi autosuficiencia y realización como mujer, y por impulsarme a soñar en grande, recordándome siempre que con perseverancia se logran las metas. Gracias por esa pregunta, a veces incómoda pero siempre motivadora e impulsadora: “¿Y la tesis?”.

A mi compañero de vida, Ignacio, ahora futuro esposo, por ser un apoyo clave en este proceso. Sin tus palabras y respaldo, la culminación de esta investigación habría sido mucho más difícil. Gracias por todas las palabras de ánimo, por recordarme que siempre hay soluciones, y por iluminar los momentos en los que todo parecía más desafiante. Definitivamente, eres un pilar fundamental en mi vida.

A mi padre, Francisco, por su incondicional apoyo y por ser un ejemplo constante de esfuerzo y dedicación, inspirándome a perseguir mis estudios en esta alma mater.

A la memoria de una de mis mejores amigas, Melissa, quien partió antes de tiempo de este plano terrenal, pero que sigue siendo una fuente de inspiración para seguir adelante. Una persona ejemplar, siempre de las primeras en emocionarse con mis logros. Sé que, donde sea que esté, se siente orgullosa de la culminación de esta meta.

## **AGRADECIMIENTO**

Quiero expresar mi especial agradecimiento a mi tutor, Erik Rodríguez Pacheco, por creer siempre en mis capacidades y permitirme asumir este reto en conjunto. Su apoyo y sabiduría han sido fundamentales para la realización de este proyecto, no solo desde una perspectiva de datos, sino también ayudándome a comprender la perspectiva bancaria. Sin duda, es una verdadera inspiración profesional.

A mis lectores, Gilbert Brenes Camacho y Simón Rodríguez Núñez, por ser un soporte clave a lo largo de este proyecto. Agradezco el tiempo que dedicaron a escucharme, motivarme a seguir adelante y brindarme consejos valiosos.

Este trabajo final de investigación aplicada fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.

---

Dr. Luis Barboza Chinchilla  
**Representante de la Decana  
Sistema de Estudios de Posgrado**

---

M.B.A. Erik Rodríguez Pacheco  
**Profesor Guía**

---

M.Sc. Simón Rodríguez Núñez  
**Lector**

---

Dr. Gilbert Brenes Camacho  
**Lector**

---

M.Sc. Ericka Méndez Chacón  
**Representante Programa de  
Posgrado en Estadística**

---

Fiorella Laurito Torres  
**Sustentante**

## TABLA DE CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTO.....	iii
HOJA DE APROBACIÓN.....	iv
TABLA DE CONTENIDO.....	v
RESUMEN EN ESPAÑOL.....	ix
ABSTRACT.....	x
LISTA DE CUADROS.....	xi
LISTA DE FIGURAS.....	xii
LISTA DE ABREVIATURAS.....	xiv
CAPÍTULO I: INTRODUCCIÓN.....	1
1.1    Contexto.....	1
1.2    Problemática.....	3
1.3    Objetivos.....	4
1.3.1    Objetivo General.....	4
1.3.2    Objetivos Específicos.....	4
1.4    Justificación.....	5
1.5    Plan de trabajo.....	7
CAPÍTULO II: MARCO TEÓRICO.....	8
2.1    Acuerdos Basilea.....	8
2.1.1    Basilea I.....	9
2.1.2    Basilea II.....	10
2.1.3    Basilea III.....	11
2.2    Regulación Bancaria Costarricense.....	12
2.3    Riesgo Crediticio.....	13
2.3.1    Indicadores de riesgo de crédito.....	14
2.3.1.1    Probabilidad de incumplimiento.....	14
2.3.1.2    Pérdida en caso de impago.....	15
2.3.1.3    Exposición en el momento de impago.....	16
2.3.1.4    Pérdida esperada.....	16
2.4    Gestión de riesgo.....	17
2.4.1    Pruebas de estrés.....	18

2.4.1.1	Escenarios hipotéticos por criterio de experto.....	19
2.4.1.2	Escenarios simulados.....	19
2.4.1.3	Algunas aplicaciones de pruebas de estrés.....	20
CAPÍTULO III: METODOLOGÍA.....		23
3.1	Materiales.....	23
3.1.1	Datos.....	23
3.1.1.1	Transformación de variables respuesta.....	26
3.1.1.2	Variables Macroeconómicas.....	26
3.1.1.3	Segmentos.....	29
3.1.1.4	Ventana de análisis.....	30
3.1.1.5	Herramienta de análisis y aplicación web.....	30
3.2	Métodos.....	31
3.2.1	Análisis de Series Temporales.....	31
3.2.1.1	Regresión con series temporales.....	32
3.2.1.2	Modelos Autorregresivos de Promedio Móvil.....	34
3.2.1.3	Redes Neuronales Autorregresivas.....	38
3.2.2	Modelos Aprendizaje Automático Supervisado.....	41
3.2.2.1	Método de k vecinos más cercanos.....	42
3.2.2.2	Máquinas de soporte vectorial.....	44
3.2.2.3	Bosques aleatorios.....	47
3.2.3	Validación de modelos.....	48
3.2.3.1	Validación Cruzada.....	48
3.2.3.2	Medidas de bondad de ajuste.....	50
3.2.3.3	Proceso selección del modelo.....	51
3.2.4	Prueba de estrés.....	53
3.2.4.1	Simulación de escenarios.....	54
3.2.4.2	Estimación de la distribución de pérdida.....	55
CAPÍTULO IV: RESULTADOS.....		57
4.1	Análisis Exploratorio.....	57
4.1.1	Tasas de cura.....	57
4.1.1.1	Cartera Total.....	58
4.1.1.2	Cartera Segmentada.....	60

4.1.1.2.1	Empresas medianas y grandes .....	60
4.1.1.2.2	Empresas pequeñas.....	62
4.1.1.2.3	Microempresas.....	63
4.1.2	Tasas de incumplimiento .....	65
4.1.2.1	Cartera Total.....	66
4.1.2.2	Cartera Segmentada.....	67
4.1.2.2.1	Empresas medianas y grandes .....	67
4.1.2.2.2	Empresas pequeñas.....	69
4.1.2.2.3	Microempresas.....	70
4.1.3	Variables Macroeconómicas.....	72
4.2	Modelado .....	75
4.2.1	Ejemplo proceso de modelado.....	76
4.2.1.1	Regresión con series temporales.....	76
4.2.1.2	Modelo Autorregresivo de Promedio Móvil .....	78
4.2.1.3	Redes Neuronales Autorregresivas.....	79
4.2.1.4	Método de k vecinos más cercanos .....	81
4.2.1.5	Máquinas de soporte vectorial .....	83
4.2.1.6	Bosques Aleatorios .....	85
4.2.2	Resultados Generales.....	87
4.2.3	Pruebas de estrés.....	93
4.2.3.1	Cartera Total.....	93
4.2.3.2	Cartera Segmentada.....	94
4.3	Dashboard: Aplicativo <i>Shiny</i> .....	96
CAPÍTULO V: CONCLUSIONES Y DISCUSIÓN.....		103
BIBLIOGRAFÍA .....		108
ANEXOS.....		118
Anexo 1.....		118
Anexo 2.....		119
Anexo 3.....		119
Anexo 4.....		120
Anexo 5.....		121
Anexo 6.....		121

Anexo 7.....	121
Anexo 8.....	122
Anexo 9.....	123
Anexo 10.....	124
Anexo 11.....	124
Anexo 12.....	124
Anexo 13.....	125
Anexo 14.....	126
Anexo 15.....	127
Anexo 16.....	127
Anexo 17.....	127
Anexo 19.....	129
Anexo 20.....	130
Anexo 21.....	130
Anexo 22.....	131
Anexo 23.....	132
Anexo 24.....	133
Anexo 25.....	134
Anexo 26.....	134
Anexo 27.....	134
Anexo 28.....	135
Anexo 29.....	136
Anexo 30.....	137
Anexo 31.....	137
Anexo 32.....	137
Anexo 33.....	138

## RESUMEN EN ESPAÑOL

La presente investigación ofrece una evaluación de la cartera crediticia del Banco Improsa, a través del pronóstico de las tasas de mora e incumplimiento en un contexto de cambios económicos significativos debido a crisis financieras recientes. Estas crisis han resaltado la necesidad de contar con sistemas financieros robustos, y las pruebas de estrés se han establecido como herramientas esenciales para evaluar de manera proactiva o anticipada, la estabilidad del sector financiero. Este enfoque es particularmente relevante para Banco Improsa, que se enfoca en el sector empresarial que puede experimentar efectos adversos de manera adelantada en razón de cambios en las principales variables económicas. En el marco teórico se abordaron las directrices internacionales del Comité de Basilea, destacando los Acuerdos de Basilea I, II y III, y su implementación en Costa Rica mediante normativas como el Reglamento sobre la Suficiencia Patrimonial de Entidades Financieras. Estas regulaciones detallan indicadores clave de riesgo de crédito, como la probabilidad de incumplimiento y la pérdida en caso de impago. La investigación implementó modelos de series temporales y técnicas de aprendizaje automático, incluyendo simulaciones de Monte Carlo, para llevar a cabo pruebas de estrés en la cartera crediticia del banco. Se evaluaron diferentes modelos de predicción, como ARMA, NNAR, KNN, SVR, y bosques aleatorios, utilizando métricas de precisión como RMSE y MAPE. Los modelos ARMA, NNAR y KNN se destacaron por su precisión y consistencia, mostrando MAPE inferiores al 3.2% en algunos segmentos empresariales. Estos resultados resaltan la importancia de las series temporales para mejorar la precisión de las predicciones, considerando los patrones y tendencias temporales. Adicionalmente, se identificaron variables macroeconómicas clave, como la inflación y la Tasa Básica Pasiva (TBP), que afectan significativamente el comportamiento crediticio de las empresas. La prueba de estrés también mostró que un enfoque segmentado permite una mejor identificación de escenarios de pérdidas extremas, alineándose con estándares internacionales y asegurando que el banco opere dentro de parámetros de riesgo aceptables establecidos por la SUGEF. Estos hallazgos se alinean al tipo de banco y enfoque estratégico, que se especializa en la identificación y atención de nichos específicos, estrategia que lo distingue del resto del sistema financiero y de los abordajes generales.

## **ABSTRACT**

This research provides an evaluation of Banco Improsa's loan portfolio by forecasting cure and default rates in a context of significant economic changes caused by recent financial crises. These crises have highlighted the need for robust financial systems, and stress tests have emerged as essential tools for proactively assessing the stability of the financial sector. This approach is particularly relevant for Banco Improsa, which focuses on the business sector that may experience early adverse effects due to changes in key economic variables. The theoretical framework addresses the international guidelines set by the Basel Committee, emphasizing the Basel I, II, and III Accords and their implementation in Costa Rica through regulations such as the Regulation on the Capital Adequacy of Financial Entities. These regulations detail key credit risk indicators, such as the probability of default and loss given default. The study implemented time series models and machine learning techniques, including Monte Carlo simulations, to conduct stress tests on the bank's loan portfolio. Various predictive models, including ARMA, NNAR, KNN, SVR, and random forests, were evaluated using precision metrics such as RMSE and MAPE. ARMA, NNAR, and KNN models stood out for their accuracy and consistency, achieving MAPE values below 3.2% in some business segments. These results highlight the importance of time series models in improving prediction accuracy by accounting for temporal patterns and trends. Additionally, key macroeconomic variables, such as inflation and the Basic Passive Rate were identified as significantly influencing the credit behavior of businesses. The stress test also demonstrated that a segmented approach allows for better identification of extreme loss scenarios, aligning with international standards and ensuring that the bank operates within acceptable risk parameters set by SUGEF. These findings align with the bank's strategic approach, which focuses on identifying and serving specific niches, a strategy that sets it apart from the broader financial system and more generalized approaches.

## LISTA DE CUADROS

Cuadro 1. Rangos de calificación de riesgo según porcentaje de pérdida esperada de la cartera crediticia .....	13
Cuadro 2. Notación matemática de modelos ARMA .....	37
Cuadro 3. Banco Improsa: Estadísticas descriptivas de la tasa de cura según tamaño de empresa, junio 2014 a julio 2022.....	58
Cuadro 4. Banco Improsa: Estadísticas descriptivas de la tasa de incumplimiento según tamaño de empresa, junio 2014 a julio 2022 .....	65
Cuadro 5. Estadísticas descriptivas de las variables independientes .....	73
Cuadro 6. Cartera Total: Resumen de medidas de rendimiento de las variantes del modelo de regresión longitudinal .....	77
Cuadro 7. Resumen de medidas de rendimiento de las variantes del modelo ARMA.....	79
Cuadro 8. Resumen de medidas de rendimiento de las variantes del modelo NNAR .....	81
Cuadro 9. Resumen de medidas de rendimiento de las variantes del modelo KNN .....	83
Cuadro 10. Resumen de medidas de rendimiento de las variantes del modelo SVR .....	85
Cuadro 11. Resumen de medidas de rendimiento de las variantes de los bosques aleatorios .....	87
Cuadro 12. Listado de variables macroeconómicas y sus .....	87
Cuadro 13. Cartera Total: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento .....	88
Cuadro 14. Medidas y Grandes empresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento .....	88
Cuadro 15. Pequeñas empresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento .....	89
Cuadro 16. Microempresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento .....	90
Cuadro 17. Resumen de los finales obtenidos y sus medidas de rendimiento .....	92

## LISTA DE FIGURAS

Figura 1. Distribución de pérdidas crediticias .....	17
Figura 2. Proceso de cálculo de la tasa de incumplimiento .....	24
Figura 3. Proceso de cálculo de la tasa de cura .....	25
Figura 4. Comparación entre una neurona biológica y una artificial .....	38
Figura 5. Ejemplo de una red neuronal.....	39
Figura 6. Funciones de activación más utilizadas .....	40
Figura 7. Comparación entre aplicación de ponderación uniforme o Epanechnikov en el modelo KNN.....	44
Figura 8. Variantes de la validación cruzada para series de tiempo .....	49
Figura 9. Etapas en el proceso de modelado de las tasas de incumplimiento y cura .....	53
Figura 10. Banco Improsa: Tasa de cura de la totalidad de la cartera crediticia, junio 2014 a julio 2022.....	59
Figura 11. Banco Improsa: Descomposición de la tasa de cura de la totalidad de la cartera crediticia, junio 2014 a julio 2022 .....	60
Figura 12. Banco Improsa: Tasa de cura de medianas y grandes empresas, junio 2014 a julio 2022 .....	61
Figura 13. Banco Improsa: Descomposición de la tasa de cura de medianas y grandes empresas, junio 2014 a julio 2022 .....	61
Figura 14. Banco Improsa: Tasa de cura de pequeñas empresas, junio 2014 a julio 2022 ..	62
Figura 15. Banco Improsa: Descomposición de la tasa de cura de pequeñas empresas, junio 2014 a julio 2022 .....	63
Figura 16. Banco Improsa: Tasa de cura de microempresas, junio 2014 a julio 2022 .....	64
Figura 17. Banco Improsa: Descomposición de la tasa de cura de microempresas, junio 2014 a julio 2022 .....	64
Figura 18. Banco Improsa: Tasa de incumplimiento de la totalidad de la cartera crediticia, junio 2014 a julio 2022.....	66
Figura 19. Banco Improsa: Descomposición de la tasa de incumplimiento de la totalidad de la cartera, junio 2014 a julio 2022 .....	67
Figura 20. Banco Improsa: Tasa de incumplimiento de las medianas y grandes empresas, junio 2014 a julio 2022.....	68
Figura 21. Banco Improsa: Descomposición de la tasa de incumplimiento de las medianas y grandes empresas, junio 2014 a julio 2022.....	68
Figura 22. Banco Improsa: Tasa de incumplimiento de las pequeñas empresas, junio 2014 a julio 2022.....	69
Figura 23. Banco Improsa: Descomposición de la tasa de incumplimiento de las pequeñas empresas, junio 2014 a julio 2022 .....	70
Figura 24. Banco Improsa: Tasa de incumplimiento de las micro empresas, junio 2014 a julio 2022 .....	71
Figura 25. Banco Improsa: Descomposición de la tasa de incumplimiento de las microempresas, junio 2014 a julio 2022.....	71

Figura 26. Matriz de correlaciones entre las tasas de cura según segmento y variables independientes .....	74
Figura 27. Matriz de correlaciones entre las tasas de incumplimiento según segmento y variables independientes .....	75
Figura 28. Banco Improsa, Cartera Total: Estimación del error mediante regresión longitudinal en validación cruzada de la tasa de incumplimiento según medida de precisión (k=29) .....	77
Figura 29. Banco Improsa, Cartera Total: Estimación del error mediante modelo ARMA en validación cruzada de la tasa de incumplimiento .....	78
Figura 30. Banco Improsa, Cartera Total: Estimación del error mediante modelo NNAR en validación cruzada de la tasa de incumplimiento .....	80
Figura 31. Banco Improsa, Cartera Total: Estimación del error mediante modelo KNN en validación cruzada de la tasa de incumplimiento .....	82
Figura 32. Banco Improsa, Cartera Total: Estimación del error mediante modelo SVR en validación cruzada de la tasa de incumplimiento .....	84
Figura 33. Banco Improsa, Cartera Total: Estimación del error mediante bosques aleatorios en validación cruzada de la tasa de incumplimiento .....	86
Figura 34. Resumen de la selección de variables según tipo de segmentación .....	91
Figura 35. Banco Improsa: Distribución de pérdidas simuladas .....	94
Figura 36. Banco Improsa: Distribución de pérdidas simuladas .....	95
Figura 37. Primera sección: Detalle del modelado, Parte I .....	97
Figura 38. Primera sección: Detalle del modelado, Parte II .....	98
Figura 39. Segunda sección: Predicción de la Pérdida .....	99
Figura 40. Segunda sección: Predicción de la Pérdida para .....	100
Figura 41. Tercera sección: Simulación de Pérdidas, metodología .....	101
Figura 42. Tercera sección: Simulación de Pérdidas, metodología segmentado .....	102

## LISTA DE ABREVIATURAS

BCBS:	Comité de Supervisión Bancaria de Basilea
BCCR:	Banco Central de Costa Rica
BPI:	Banco de Pagos Internacionales
BUST:	Bottom Up Stress Test
CEPAL:	Comisión Económica para América Latina y el Caribe
CNE:	Comisión Nacional de Emergencias
CONASSIF:	Consejo Nacional de Supervisión del Sistema Financiero
COVID-19:	Enfermedad por coronavirus 2019
EBA:	Autoridad Bancaria Europea
IMAE:	Índice Mensual de Actividad Económica
IPC:	Índice de Precios del Consumidor
LGD:	Pérdida en caso de impago
MEIC:	Ministerio de Economía, Industria y Comercio
Mipe:	Micro y pequeñas empresas
MTSS:	Ministerio de Trabajo y Seguridad Social de Costa Rica
OIT:	Organización Internacional del Trabajo
OMS:	Organización Mundial de la Salud
PD:	Probabilidad de incumplimiento
PESP:	Programas de Evaluación del Sistema Financiero
PIB:	Producto Interno Bruto
SUGEF:	Superintendencia General de Entidades Financieras
TBP:	Tasa Básica Pasiva
TCC:	Tipo de cambio compra
VAR:	Value at Risk

## CAPÍTULO I: INTRODUCCIÓN

### 1.1 Contexto

Los últimos años han estado marcados por cambios coyunturales que han tenido repercusiones en el ámbito económico y financiero (CEPAL, 2020a). La crisis financiera de 2008 tuvo secuelas negativas en la economía mundial, debido principalmente al impago de créditos en Estados Unidos (Juárez, Sánchez y Zurita, 2015). Una de las consecuencias de este evento coyuntural fue la crisis de la Eurozona, provocada por el endeudamiento de algunos países europeos tras la implementación de programas de rescate a bancos y empresas para combatir la crisis financiera internacional (Díaz, 2015). La crisis del euro evidenció la importancia de la regulación financiera, ya que los problemas en el sector bancario de algunos países condujeron a una crisis en toda la Eurozona (Weber, 2014). Más recientemente, la pandemia de COVID-19 originó una crisis económica a nivel mundial, debido a las medidas de restricción implementadas para evitar la propagación del virus, lo que causó un daño colateral en la actividad económica de los países (Banco de España, 2020). Adicionalmente, la crisis económica desencadenada por el conflicto entre Rusia y Ucrania, junto con el aumento en los precios del petróleo, ha exacerbado la incertidumbre económica global, afectando a los mercados financieros y el costo de la energía. A la luz de estos eventos de crisis, surge la necesidad de sistemas financieros sólidos para garantizar la estabilidad macroeconómica de un país, especialmente en países emergentes (Blanco y García, 2004).

Aunque la supervisión bancaria juega un papel primordial en promover la estabilidad macroeconómica, las instituciones financieras están expuestas a diferentes tipos de riesgo (Ware, 1996). Según el Acuerdo de Basilea II, existen tres principales fuentes de riesgo bancario: riesgo de mercado, operacional y crediticio (BCBS, 2004). Específicamente, el riesgo crediticio se refiere a la situación en la que los deudores incumplen con la obligación de pagar en una fecha establecida, entrando así en un estado de morosidad (Van Gestel y Baensens, 2009). En los últimos años, los niveles de morosidad en el sistema financiero costarricense han mostrado una tendencia al alza (SUGEF, 2019a). Esta situación genera preocupación, ya que incrementos en la morosidad impactan negativamente la rentabilidad

de las entidades bancarias (Chavarín, 2015) y como una derivada de ello pueden afectar los niveles de cobertura patrimonial para mitigación de esos riesgos.

Dado que los momentos de crisis exponen al sistema bancario a la materialización de riesgos, se han desarrollado diversos métodos de medición del riesgo crediticio con el fin de obtener mayor certeza sobre el nivel de pérdidas de capital que se podría experimentar debido a perturbaciones adversas. Las técnicas de estimación del riesgo crediticio se dividen en tres tipos de modelos: los financieros, los basados en el criterio de expertos, y los empíricos basados en datos históricos (Van Gestel y Baesens, 2009). Dentro de estos últimos se encuentran las pruebas de estrés, que son una herramienta para la gestión interna de riesgos en los bancos y cuya implementación es promovida en los Acuerdos de Basilea (BCBS, 2009). De acuerdo con la Autoridad Bancaria Europea, estas pruebas tienen como objetivo comprender qué sucede con los balances de las instituciones financieras ante una recesión económica (EBA, 2018).

Las pruebas de estrés no proporcionan la probabilidad de que ocurra una crisis determinada, sino que ayudan a identificar las vulnerabilidades de las instituciones bancarias (Borio, Drehmann, & Tsatsaronis, 2014). Estas pruebas consisten en modelos predictivos que evalúan una variable exógena en función de un conjunto de variables macroeconómicas y financieras (Blanco & García, 2004). Covas, Rump, y Zakrajšek (2014) afirman que el objetivo clave de las pruebas es determinar si la variable respuesta se mantiene por encima de un determinado umbral regulatorio, basándose en un escenario macroeconómico severo pero plausible.

Desde la perspectiva de la regulación bancaria, las pruebas de estrés son esenciales para medir la estabilidad y solidez del sector financiero, según los Programas de Evaluación del Sistema Financiero establecidos por el Fondo Monetario Internacional y el Banco Mundial (FMI, 2016). En el caso de Costa Rica, la SUGEF coordina anualmente la realización de la prueba de estrés denominada *Bottom Up Stress Test* (BUST), como parte del seguimiento y vigilancia de las entidades financieras supervisadas. Esta prueba se enfoca en el riesgo de crédito de las entidades y tiene como objetivo determinar el riesgo de pérdida de capital ante

escenarios macroeconómicos adversos. La realización de la prueba genera varios beneficios, ya que brinda una perspectiva de la capacidad de resistencia del sistema financiero en distintos escenarios y evalúa la habilidad técnica de las entidades en la implementación de pruebas de estrés (SUGEF, 2019b). Aunque importante, dicho ejercicio posee una limitante y es que los escenarios y segmentos a modelar son establecidos por el Regulador de forma genérica, por lo tanto, para obtener una mayor exactitud y utilidad en la evaluación del riesgo de pérdida de capital, es importante que las instituciones financieras adapten, de manera complementaria, la prueba de estrés a la naturaleza y contexto de su negocio.

## 1.2 Problemática

La prueba de estrés BUST se basa en una definición generalista del sistema financiero bancario. Se asume que los aspectos relacionados con el planeamiento de criterios para el cálculo de las variables y la segmentación de clientes se comportan de manera análoga entre las instituciones supervisadas, en un sistema financiero que además posee una particularidad importante dado el peso de la banca pública en relación a los activos y pasivos financieros que se concentran en ella. Por consiguiente, para alcanzar una estimación más precisa y útil del riesgo de pérdida de capital, las instituciones en general deben reformular los términos de la prueba acorde a sus especificaciones propias, hecho que se vuelve aun de mayor relevancia en el caso de los jugadores enfocados en banca de nicho, como es el caso de Banco Improsa que se dista del comportamiento habitual del sistema financiero total.

Particularmente, Banco Improsa se caracteriza por un enfoque empresarial con una clara orientación hacia el sector de micro y pequeñas empresas (MIPE), que representa aproximadamente el 70% de su cartera (Grupo Financiero Improsa, 2018). En los últimos siete años, el banco ha participado en las pruebas de estrés dirigidas por la SUGEF. Sin embargo, los resultados obtenidos en las pruebas generales dirigidas por el regulador, pueden ser más útiles para la planificación estratégica al hacer ajustes que lo lleven a pasar de modelos establecidos por el regulador, a modelos internos cuyo propósito es capturar mejor la realidad específica de Banco Improsa, por ejemplo, la segmentación de clientes implementada en los modelos del ejercicio BUST, no está alineada con su enfoque a nichos.

Por consiguiente, surge la necesidad de replicar la prueba de estrés utilizando una segmentación basada en el tamaño de la empresa, distinguiendo entre grandes, medianas, pequeñas y microempresas.

El presente trabajo expone una prueba de estrés en la que se evalúa el comportamiento de indicadores de riesgo crediticio, específicamente las tasas de cura e incumplimiento, de un banco privado costarricense. Se implementaron métodos analíticos de series de tiempo y algoritmos de aprendizaje automático supervisado para el pronóstico de estas tasas. De esta manera, se evaluaron las diferencias en los resultados de los indicadores de riesgo al aplicar ambos enfoques metodológicos.

### 1.3 Objetivos

#### 1.3.1 Objetivo General

- Aplicar pronóstico de tasas de cura e incumplimiento con la finalidad de determinar el riesgo de pérdida en la situación del capital de la entidad financiera, y su capacidad de resistencia ante escenarios macroeconómicos de estrés.

#### 1.3.2 Objetivos Específicos

- Efectuar un análisis descriptivo de las series de las tasas de cura e incumplimiento para cada uno de los segmentos bancarios.
- Particionar los datos en parte de entrenamiento y prueba mediante el método de validación cruzada para series de tiempo.
- Estimar modelos de vanguardia en el análisis de series de tiempo y modelos provenientes del aprendizaje automático supervisado para las tasas de cura e incumplimiento de cada uno de los segmentos.
- Contrastar los distintos modelos con base a medidas de rendimiento en la predicción de las tasas de cura e incumplimiento.
- Estimar la pérdida por riesgo de crédito por medio de los escenarios de las variables macroeconómicas simulados que representen un entorno de crisis financiera.

- Implementar una aplicación web en forma de dashboard para facilitar el análisis, la visualización, la interpretación y la toma de decisiones de los resultados de los pronósticos de las tasas de cura e incumplimiento.

#### 1.4 Justificación

En los últimos años, las entidades financieras se han visto afectadas por una serie de eventos económicos adversos. La crisis económica y financiera mundial de 2008 provocó una desaceleración del crecimiento de la economía a nivel mundial, causada en gran parte por los efectos de impago hacia el sector bancario, afectando el mercado financiero y produciendo una desaceleración de la actividad económica y bancaria mundial (FMI, 2018). En 2019 la economía mundial se desaceleró principalmente por el aumento de las disputas comerciales entre Estados Unidos y China, la incertidumbre sobre la salida del Reino Unido de la Unión Europea, y la disminución del comercio mundial (BCCR, 2019a). De acuerdo con la Comisión Económica para América Latina y el Caribe (CEPAL), la magnitud del comercio mundial disminuyó por primera vez desde la crisis financiera del 2008, debido a dichas barreras comerciales (CEPAL, 2020a).

En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró la enfermedad por coronavirus como pandemia (OMS, 2020), lo que desencadenó una paralización de la economía a nivel global. Las medidas adoptadas para contrarrestar la propagación del virus fueron comparables a situaciones de guerra (CEPAL, 2020a). Además, el secretario general de las Naciones Unidas, António Guterres, afirmó que la pandemia del coronavirus es la peor crisis que el mundo ha enfrentado desde la Segunda Guerra Mundial (citado en Sandford y Rodríguez, 2020). Como resultado, el coronavirus ha llevado a la economía mundial a una recesión de proporciones históricas (Goldman Sachs, 2020). Adicionalmente, los recientes conflictos bélicos entre Rusia y Ucrania, así como entre Israel y Hamas, han tenido un impacto significativo en las economías globales. Uno de los efectos más notorios ha sido el aumento en los precios del petróleo, lo que ha exacerbado las tensiones económicas a nivel internacional, afectando la estabilidad de los mercados y contribuyendo a la inflación en diversos países (Parrado y Garzón, 2024).

En relación con el contexto de Costa Rica, la economía está expuesta a los efectos económicos globales (BCCR y CONASSIF, 2020), especialmente de las condiciones económicas de Estados Unidos, ya que alrededor del 10% del producto interno bruto (PIB) de Costa Rica depende de la oferta y la demanda del país norteamericano (CEPAL, 2020a). En este sentido, la implementación de pruebas de estrés resulta crucial para la gestión de riesgos, ya que permite evaluar cómo podría comportarse la economía costarricense bajo condiciones adversas o inesperadas provenientes de shocks internacionales.

El sector empresarial costarricense está conformado por aproximadamente 133 mil empresas, de las cuales el 97% son micro, pequeñas y medianas empresas (mipyme) (CNE, MEIC, MTSS, 2020). La crisis sanitaria del COVID-19 puso en evidencia cómo, independientemente del tamaño, todas las empresas pueden verse afectadas por una crisis global, siendo las mipyme las más vulnerables a la interrupción de sus operaciones (OIT, 2020). Esta situación resulta especialmente importante en el banco, ya que la mayoría de sus clientes pertenecen a los sectores de comercio y servicios, los cuales fueron especialmente impactados por la pandemia. En este contexto, es crucial evaluar el comportamiento del riesgo crediticio mediante pruebas de estrés, las cuales permiten efectuar ajustes tácticos en beneficio del banco y de sus clientes, especialmente frente a la posibilidad de efectos adversos en la economía que afecten la capacidad de pago de sus clientes. La implementación de pruebas de estrés se vuelve aún más relevante al proporcionar una perspectiva sobre el capital del banco en momentos de crisis y al permitir la planificación estratégica adecuada.

Es importante señalar que el análisis presentado en esta tesis se limita a los datos y eventos ocurridos hasta julio de 2022. Los conflictos coyunturales y los desarrollos posteriores a esa fecha no se han considerado en el ámbito de este estudio. Por lo tanto, cualquier conclusión o recomendación derivada de este análisis debe interpretarse en el contexto temporal especificado, sin tener en cuenta los eventos ocurridos después de julio de 2022.

## 1.5 Plan de trabajo

Este trabajo pretende llevar a cabo una implementación de una prueba de estrés para el pronóstico de tasas de cura e incumplimiento, para determinar el riesgo de pérdida en la situación del capital de la entidad financiera y su capacidad de resistencia ante escenarios macroeconómicos adversos. El documento está dividido en cinco capítulos. El Capítulo II presenta el marco teórico, donde se profundizará sobre las normativas referentes a las pruebas de estrés en los Acuerdos Basilea, los conceptos del riesgo crediticio y los diversos tipos de evaluación de estabilidad financiera. El Capítulo III aborda la metodología necesaria para alcanzar los objetivos: datos, segmentos bancarios, descripción de las técnicas de series de tiempo y del aprendizaje automático supervisado, validación de los modelos y estimación de la pérdida esperada. En el Capítulo IV se describen los resultados de las pruebas de estrés para ambos panoramas de estudio. Finalmente, el Capítulo V concluye con los hallazgos obtenidos de la investigación y expone recomendaciones para futuras pruebas de estrés en entidades bancarias.

## CAPÍTULO II: MARCO TEÓRICO

Esta sección abarca aspectos de las directrices internacionales de la supervisión bancaria del Comité Basilea, el panorama de la regulación bancaria costarricense, los conceptos relevantes del riesgo de crédito y los métodos de gestión de riesgo. En primera instancia, se presenta la evolución del esquema internacional de gestión y evaluación de riesgos de los Acuerdos de Basilea, en el cual se enfatiza en el indicador de solvencia de capital. Asimismo, se expone la adaptación de las normas internacionales de gestión riesgo en la regulación del sistema bancario costarricense. Adicionalmente, se detallan los aspectos fundamentales en la gestión del riesgo de crédito. Por último, se presenta la conceptualización de métodos de gestión de riesgos bancarios y se abordan ejemplos de aplicaciones de pruebas de estrés en distintos países.

### 2.1 Acuerdos Basilea

El Banco de Pagos Internacionales (BPI) es la organización financiera internacional caracterizada por promover la cooperación entre bancos centrales. Se funda en 1930 en Basilea Suiza, como consecuencia de la Gran Depresión, convirtiéndola en la institución financiera internacional más longeva de la historia (Toniolo, 2005). En la sede del BPI, los países miembros del G10<sup>1</sup> en 1975 establecieron el Comité de Supervisión Bancaria de Basilea (BCBS, por sus siglas en inglés) con el fin de restaurar la confianza del sistema financiero internacional debido a las repercusiones causadas por la quiebra del Bankhaus Herstatt de Alemania Occidental en 1974 (BCBS, 2001).

En la actualidad, el Comité de Basilea es el principal creador de principios y estándares globales de la supervisión bancaria. Cabe destacar que, el apego a sus pautas no es mandatorio sino que el BCBS pone a disposición a los bancos centrales la implementación de sus lineamientos (Pascual, 2015). El Comité sobresale por una serie de normas acerca de principios básicos de cooperación entre supervisores, medidas de adecuación del capital y gestión de riesgos: Basilea I, Basilea II y Basilea III (Rodríguez, 2013). Según Sotelsek y

---

<sup>1</sup> Grupo de los Diez: Bélgica, Canadá, Francia, Italia, Japón, Países Bajos, Reino Unido, Estados Unidos y los bancos centrales de Alemania y Suecia.

Pavón (2012), los acuerdos surgen de países industrializados, sin embargo han tenido un impacto significativo en países emergentes.

### 2.1.1 Basilea I

En 1988, el BCBS publicó el primer acuerdo de Basilea, titulado “Convergencia Internacional sobre medición de Capital y Requisitos de Capital”, con el objetivo de alcanzar una estandarización en la manera de calcular la adecuación de capital de las entidades financieras (Gutiérrez, 2009). La adecuación de capital corresponde al nivel de recursos que las instituciones deberán resguardar para asegurar su funcionamiento ante los diversos riesgos que están inmersas (Soley y Rahnama, 2005). El acuerdo estipula que se debe cumplir con un nivel mínimo del coeficiente de solvencia del 8%. El método estándar para el cálculo del coeficiente de solvencia es mediante la razón del capital mínimo entre activos líquidos (Pascual, 2015).

Basilea I se focaliza inicialmente solo en el riesgo de crédito, ya que es el principal causante de quiebras y crisis bancarias (Gutiérrez, 2009). En consecuencia, el Comité de Basilea I publicó una enmienda del acuerdo en 1996, con el propósito de introducir al cálculo de adecuación de capital los riesgos de mercado: pérdidas causadas por la volatilidad en variables financieras como precio de acciones, tipo de cambio o tasas de interés (Rodríguez, 2013). Asimismo, la enmienda expone la posibilidad de implementar modelos internos para la estimación del coeficiente de solvencia, siempre y cuando estén basados en la metodología *Value at Risk* (VAR) y tengan una previa aprobación por las autoridades supervisoras de cada país (Gutiérrez, 2009).

Según Garcíandía (2009), los riesgos de crédito y mercado son importantes mas no suficientes para la estimación de adecuación de capital, por lo que se exigían cambios en incorporar otros tipos de riesgo. Por consiguiente, el BCBS planteó Basilea II estructurada en tres pilares complementarios (Avella, 2004).

### 2.1.2 Basilea II

El Comité publicó el segundo acuerdo de Basilea en junio de 2004, titulado “Marco revisado sobre convergencia internacional de medidas y normas de capital”. El objetivo de este acuerdo es proporcionar un método ampliado en riesgos bancarios en el cálculo de la adecuación de capital para así continuar promoviendo la estabilidad y seguridad del sistema financiero a través de la igualdad competitiva bancaria (BCBS, 2006).

Basilea II consta de tres pilares fundamentales y complementarios entre sí: requerimiento mínimo de capital, proceso de supervisión bancaria y transparencia en la difusión de información (Sotelsek y Pavón, 2012). Garcíandía (2009) afirma que en relación a Basilea I, el primer pilar tuvo algunas modificaciones mientras el segundo y tercer pilar son innovaciones en la normativa.

En relación al primer pilar de requerimiento mínimo de capital, el umbral mínimo del coeficiente de solvencia del 8% se mantuvo (Pascual, 2015). El cambio subyace en que se modificó el denominador del coeficiente, ya que se determinó ponderar los activos por los riesgos de crédito, mercado y operativo (Gutiérrez, 2009). Cabe mencionar que el riesgo operativo hace referencia a las pérdidas resultantes por fallos de personas o procesos internos de las instituciones (Van Gestel y Baesens, 2009).

El segundo pilar brinda, a las organizaciones supervisoras, reglas de revisión de los procesos internos de evaluación de riesgo y adecuación de capital (BCBS, 2006). Se espera que estas autoridades sean capaces de fijar los objetivos de capital de acuerdo al entorno y al perfil de riesgo de cada institución bancaria, por encima del 8% propuesto en el pilar I (Rodríguez, 2013).

El último pilar relacionado con la transparencia en la difusión de información, establece el deber de las entidades bancarias de exponer el nivel de capital y sus metodologías de gestión de riesgos, de manera pública (BCBS, 2006). Con el propósito de que las personas puedan realizar una valoración del estado financiero y el rendimiento de las instituciones

bancarias Por otro lado, se estimula el otorgamiento de incentivos a los bancos que gestionen de manera efectiva sus riesgos y penalizando a los que lo incumplan (Gutiérrez, 2009).

Debido a las consecuencias de la crisis financiera del 2008, el Comité de Basilea realizó esfuerzos por mejorar el marco internacional para la gestión de riesgos e incrementar los requerimientos de capital (Sotelsek y Pavón, 2012). En julio 2009 surge como solución provisional Basilea 2.5, el cual introduce como requisito implementar la metodología VAR con escenarios estresados (BCBS, 2011). Rodríguez resalta la importancia de incorporar a esta metodología simulaciones con el fin de apreciar la solvencia de capital ante cambios adversos del entorno económico (p. 83, 2013).

### 2.1.3 Basilea III

En 2010, el Comité de Basilea publicó el tercer acuerdo titulado “Marco regulatorio internacional para los bancos”, el cual está compuesto de una serie de medidas de urgencia respuesta a la crisis financiera del 2008 (BCBS, 2010). El objetivo de Basilea III es fortalecer el capital de las instituciones bancarias al aumentar la capacidad de absorber pérdidas en períodos de perturbaciones económicas (Warman, 2013).

Basilea III exige mayor calidad en la definición de capital para el cálculo del coeficiente de solvencia, otorgando mayor peso al capital ordinario y utilidades no distribuidas, es decir en la definición de capital predominan acciones seguras y perpetúas (Rodríguez, 2011). Además, el coeficiente de solvencia pasa de un 8% a 10.5%, adicionando un colchón de conservación de capital que podrá ser utilizado en circunstancias de estrés (BCBS, 2010).

Una de las críticas más frecuentes los primeros acuerdos de Basilea era la prociclicidad, la cual consiste en la reducción del otorgamiento de créditos debido al aumento de requisitos de capital durante recesiones económicas (Van Gestel y Baesens, 2009). Por lo tanto, se añade al coeficiente de solvencia un colchón de capital anticíclico de alta calidad, el cual corresponde a un porcentaje de los activos que puede variar entre 0% y 2.5% para tiempos de perturbaciones económicas o normales respectivamente (Rodríguez, 2011). En

tiempos favorables se recauda este porcentaje para utilizar en los momentos adversos, por lo que el coeficiente de solvencia de capital puede oscilar entre 10.5% a 13% dependiendo del entorno económico (BCBS, 2010).

## 2.2 Regulación Bancaria Costarricense

La adopción de los acuerdos de Basilea no es carácter obligatorio pero corresponde a un mecanismo transcendental para que las autoridades supervisoras puedan sobrellevar los cambios de los sistemas financieros (Rojas, 2008). En Latinoamérica, únicamente tres países están en proceso de implementar en su regulación los lineamientos de Basilea III: México, Argentina y Brasil (Warman, 2013).

En 2004, la SUGEF decretó postergar la implementación de Basilea II con el argumento de que el país no estaba preparado para ese abrupto cambio, puesto que se debían subsanar las deficiencias de la regulación existente (Rodríguez, 2012). No obstante, para no quedar estar al margen de los estándares internacionales, Costa Rica ha ido incorporando paulatinamente doctrinas de Basilea mediante su normativa (Gutiérrez, 2009). En relación, el superintendente de la SUGEF en el periodo 2004-2009, Óscar Rodríguez Ulloa, reafirma la importancia de importar a la regulación ideas inteligentes provenientes de los Acuerdos de Basilea acorde al entorno macroeconómico del país (2012, p.114). Camacho (2014) afirma que el país se caracteriza por una adopción parcial de las medidas de Basilea con un enfoque en la gestión de riesgo y cierre de brechas regulatorias.

En enero del 2006, el CONASSIF aprobó el Reglamento sobre la Suficiencia Patrimonial de Entidades Financieras (SUGEF 3-06), siendo una tropicalización de las ideas de Basilea, con el objetivo de estandarizar la metodología cálculo de la suficiencia patrimonial y el requerimiento mínimo de capital de las entidades financieras costarricenses (Gutiérrez, 2009). En la normativa el homólogo del coeficiente de solvencia de Basilea corresponde a la suficiencia patrimonial, la cual indica el capital suficiente para que una entidad financiera pueda resguardar las pérdidas por riesgo de mercado, operativo y de crédito; y estipula un nivel mínimo del 10% (CONASSIF, 2008; SUGEF, 2019c).

Por otro lado, el CONASSIF (2003) establece un indicador del nivel de pérdida esperada de capital ante una materialización del riesgo crediticio particularmente, se calcula según la siguiente fórmula:

$$\% \text{ Pérdida Riesgo Credito} = \frac{\text{Pérdida Esperada}}{\text{Cartera Total}} \times 100$$

El numerador denominado pérdida esperada de la cartera de crédito, corresponde al monto mínimo que las entidades financieras deben respaldar al final de cada cierre de mes, mientras que el denominador se compone del total de saldos por cobrar de operaciones de crédito (CONASSIF, 2003). De acuerdo al Acuerdo SUGEF 24-00: Reglamento para juzgar la situación económica financiera de las entidades fiscalizadas, este indicador debe ser menor a 1.7% para que una institución financiera sea catalogada fuera de riesgo como se observa el Cuadro 1 (SUGEF, 2019d).

**Cuadro 1. Rangos de calificación de riesgo según porcentaje de pérdida esperada de la cartera crediticia**

<b>Categoría</b>	<b>Rango</b>
Normal	Menor o igual a 1.7%
Riesgo 1	Mayor a 1.7% y menor o igual a 3.7%
Riesgo 2	Mayor a 3.7% y menor o igual a 8.6%
Riesgo 3	Mayor a 8.6%

Fuente: Elaboración propia según Acuerdo SUGEF 24-00.

En términos de esta investigación, el indicador de porcentaje de pérdida esperada por riesgo de crédito se implementa para poder clasificar el nivel de riesgo de pérdida de capital de la cartera en las dos alternativas de análisis, encaminado a alcanzar el objetivo general de estudio.

### 2.3 Riesgo Crediticio

Dado la naturaleza de la actividad de las entidades financieras, el riesgo crediticio es el que genera mayor pérdidas monetarias (Van Gestel y Baesens, 2009), puesto que existe la

posibilidad que una parte de los clientes bancarios no puedan hacer frente a pagar una deuda debido a la quiebra o insolvencia (Chorafas, 2000).

Los ciclos económicos son un factor importante en la ocurrencia de incumplimiento de obligaciones crediticias ya que en temporadas de expansión económica las tasas de morosidad tienden a disminuir (Saavedra y Saavedra, 2010). Por lo tanto, un pilar fundamental de la supervisión bancaria es la implementación de técnicas de modelado y análisis que permitan valorar el riesgo de crédito en distintos escenarios macroeconómicos (Fernández y Pérez, 2005). Asimismo, Delgado y Saurina (2004) afirman que la estabilidad de una entidad depende del monto resguardo ante las posibles pérdidas de las carteras crediticias y de la valoración de los factores relacionados a la materialización del riesgo de crédito.

Las metodologías de medición del riesgo crediticio se agrupan en dos familias: métodos estáticos y métodos dinámicos. El Acuerdo de Basilea II promueve los métodos dinámicos facilitando a las entidades financieras gestionar el riesgo de crédito implementando modelos bajo las políticas e infraestructura de cada entidad (Fernández y Pérez, 2007). En esta investigación se implementa un método dinámico para evaluar el riesgo de crédito, por lo cual se deben considerar los indicadores apropiados que se detallan en la siguiente sección.

### 2.3.1 Indicadores de riesgo de crédito

La evaluación de riesgo de crédito por métodos dinámicos para cumplir con los parámetros mínimos internacionales de Basilea debe contemplar distintos indicadores financieros: probabilidad de incumplimiento, pérdida en caso de impago y exposición al riesgo de crédito (BCBS, 2006).

#### 2.3.1.1 Probabilidad de incumplimiento

La probabilidad de incumplimiento (PD, por sus siglas en inglés) es cuando ocurre la materialización del riesgo de un evento de morosidad (Van Gestel y Baesens, 2009). Un aspecto importante en la definición de incumplimiento es establecer adecuadamente el

umbral de días en morosidad, el supuesto estándar considera que una operación bancaria incumple cuando han transcurrido más de 90 días en mora (McNeil, Frey y Embrechts, 2005). De acuerdo a Basilea II, las superintendencias de cada país orientan e inspeccionan a las entidades bancarias en la definición del incumplimiento de sus carteras (BCBS, 2006). En esta investigación, la manera de estimar la probabilidad de incumplimiento se detalla más adelante en la sección de Metodología.

### 2.3.1.2 Pérdida en caso de impago

La pérdida en caso de impago corresponde a una estimación de la fracción del portafolio que se perderá en caso de que se reincida en el incumplimiento y la proporción de la cartera respaldada por garantías o mitigadores de riesgo (McNeil, Frey y Embrechts, 2005; SUGEF, 2019b). La pérdida en caso de impago o LGD (*Loss Given Default*) puede tomar valores tanto valores negativos como positivos, donde las cifras negativas indican ganancias y el valor de cero significa que no hay pérdidas (Van Gestel y Baensens, 2009).

La SUGEF indica que la pérdida en caso de impago se calcula a partir de dos componentes: el valor de garantías de las operaciones crediticias y tasa de cura de la cartera (SUGEF, 2019b). Basilea II indica que la LGD se fundamenta en el valor de mercado estimado de las garantías (colateralización), es decir se consideran las operaciones crediticias cuyo saldo que está total o parcialmente cubierto por un mitigador de riesgo (BCBS, 2006). Asimismo, el cálculo de la pérdida en caso de impago se basa en la fracción de la cartera que recupera de un estado de morosidad (tasa de cura), puesto que se debe considerar la posibilidad que las operaciones que se curan recaigan en el incumplimiento de sus obligaciones (Van Gestel y Baensens, 2009). Por lo tanto, el cálculo de la LGD está dado por la siguiente fórmula:

$$LGD = (1 - \%Colateralización) * (1 - Tasa de Cura)$$

### 2.3.1.3 Exposición en el momento de impago

Basilea II insta a implementar en la gestión de riesgo la estimación de la exposición en el momento de impago (EAD, por sus siglas en inglés), la cual corresponde al monto de la cartera expuesto a que las operaciones de crédito incurran en el incumplimiento (BCBS, 2006). No obstante, el monto exacto de la exposición dependerá de los posibles sobregiros que no están contemplados en el portafolio (Van Gestel y Baesens, 2009). En relación, la SUGEF instruye que en el cálculo de la EAD se debe considerar un crecimiento de la cartera y cada entidad deberá ajustar acorde a su plan estratégico (SUGEF, 2019b).

Los indicadores de riesgo explicados (PD, LGD y EAD) cumplen un papel primordial en la gestión de riesgo de crédito ya que son los componentes en la determinación en las pérdidas esperada de la cartera de crédito (McNeil, Frey y Embrechts, 2005).

### 2.3.1.4 Pérdida esperada

La pérdida esperada representa el monto promedio de pérdidas resultantes de la materialización del riesgo de crédito (Cardone y Trujillo, 2008) y corresponde al producto de la siguiente fórmula:

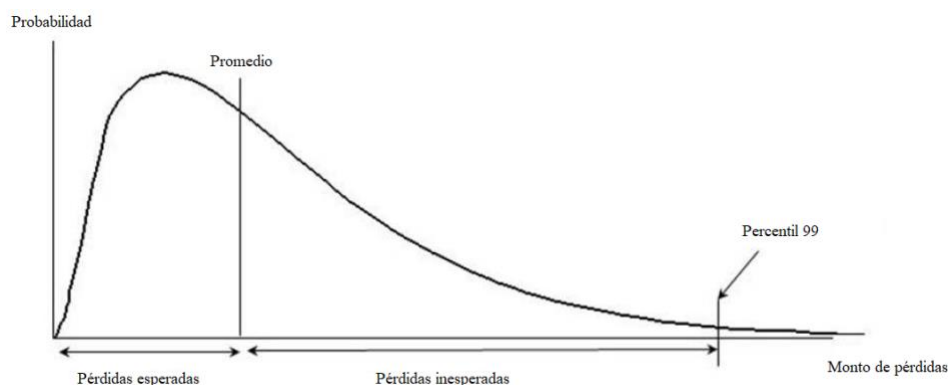
$$\text{Pérdida Esperada} = PD * LGD * EAD$$

De acuerdo con SUGEF, cada año la pérdida esperada se mitiga mediante las estimaciones mínimas (monto resguardado mandatoriamente), garantías retenidas y en última instancia con el patrimonio de cada entidad (SUGEF, 2019b). Por lo tanto, en el momento que una entidad bancaria no tenga resguardado el monto de para mitigar la pérdida esperada, el numerador del coeficiente de solvencia (propuesto por Basilea) se ve impactado (BCBS, 2006).

Cabe destacar que la pérdida esperada corresponde al promedio de la distribución de pérdidas de riesgo de crédito, además los montos que encuentren entre la pérdida esperada y el percentil 99 de las pérdidas se denominan pérdidas inesperadas, como se aprecia en la Figura 1 (McNeil, Frey y Embrechts, 2005). Las entidades bancarias pueden determinar la

máxima pérdida mediante la técnica *Value at Risk* (VAR), la cual se detalla en la sección de Metodología.

**Figura 1. Distribución de pérdidas crediticias**



Fuente: tomado del libro *Quantitative Risk Management: Concepts, Techniques and Tools*.

Los acuerdos de Basilea señalan que para complementar el cálculo de indicadores de riesgo de crédito, los bancos deberán llevar a cabo pruebas de tensión y de esta manera valorar posibles cambios coyunturales que puedan perjudicar la cartera crediticia y el capital de las entidades (BCBS, 2006).

## 2.4 Gestión de riesgo

La gestión de riesgo corresponde en una estimación del impacto que los riesgos pueden tener en la continuidad de alguna entidad con el objetivo de determinar la necesidad de adoptar medidas preventivas (García y Salazar, 2005). De acuerdo al Comité de Basilea (2004), los bancos no únicamente deben cubrir el capital necesario ante posibles pérdidas, sino que deben mejorar y utilizar técnicas internas de gestión de riesgos. Una manera de satisfacer esto es relacionar mediante modelos estadísticos el entorno macroeconómico con el riesgo de crédito (González y Uquillas, 2017), además los modelos predictivos han permitido determinar las pérdidas que se exponen los portafolios de crédito en coyunturas económicas de la mano con la implementación las pruebas de estrés (Fernández y Pérez,

2007). Las pruebas de tensión complementan otras técnicas y medidas con el objetivo de mitigar el aumento del nivel de riesgo (BCBS, 2009).

#### 2.4.1 Pruebas de estrés

Las pruebas de estrés son parte fundamental de la gestión de riesgos bancarios y su implementación es promocionada por los acuerdos de Basilea (BCBS, 2009). Según Blanco y García (2004) las pruebas de estrés corresponden a una estimación del impacto de circunstancias graves pero plausibles mediante variables macroeconómicas y financieras sobre la solvencia de instituciones bancarias.

Existen dos metodologías de análisis de las pruebas de estrés: el análisis de sensibilidad y el análisis de escenarios (Van Gestel y Baesens, 2009). El análisis de sensibilidad es una de las principales técnicas cuantitativas en la gestión de riesgos (Baker et al., 1999). La principal ventaja del análisis de sensibilidad es la facilidad y rapidez de su implementación, no obstante analiza cada variable predictiva por separado y utiliza valores extremos como un mecanismo para emular una perturbación económica (BCBS, 2009).

El análisis de escenarios es una metodología más sofisticada en la cual se mide la alteración simultánea de diversos factores del entorno macroeconómico (Blanco y García, 2004). De acuerdo Van Gestel y Baesens, existen distintas maneras de establecer los escenarios de las variables predictivas (2009):

1. Escenarios históricos: se pueden considerar escenarios macroeconómicos de recesiones económicas pasadas sin embargo se parte del supuesto de que el entorno económico se comporta de manera similar en las diferentes crisis.
2. Escenarios hipotéticos por criterio de experto: consiste en suponer escenarios adversos a partir del criterio de un experto en el entorno macroeconómico.
3. Escenarios simulados: el avance computacional ha permitido establecer distintos escenarios de variables macroeconómicas mediante métodos de simulación.

En esta investigación, se establecen los escenarios de las pruebas de estrés mediante simulaciones. Es importante destacar que en la prueba BUST de la SUGEF, se utilizan métodos basados en el criterio de expertos. A continuación, se detallarán ambos métodos de escenarios recientemente mencionados.

#### 2.4.1.1 Escenarios hipotéticos por criterio de experto

La selección de escenarios hipotéticos por criterio de experto se fundamenta en utilizar el juicio competente para evitar el riesgo de depender de choques macroeconómicos del pasado (Borio, Drehmann, Tsatsaronis, 2014). El Comité de Basilea, en la guía titulada “Principios para la realización y supervisión de pruebas de tensión”, afirma que la definición de escenarios debe combinar el conocimiento y la opinión de expertos del entorno macroeconómico, altos directivos de las entidades o supervisores bancarios (BCBS, 2009).

Según Basilea II, la determinación de los escenarios hipotéticos debe basarse en supuestos de perturbaciones poco probables pero verosímiles (BCBS, 2006). Las instituciones bancarias tienen la potestad de definir la severidad de los supuestos mientras sean evaluadas por las entidades supervisoras, en caso contrario pueden utilizar los escenarios determinados por los supervisores (BCBS, 2009).

#### 2.4.1.2 Escenarios simulados

La simulación de escenarios consiste en una recopilación de suposiciones de condiciones económicas poco probables que podrían inducir pérdidas crediticias elevadas (BCBS, 2005). Los escenarios son generados de manera reiterativa en un número suficiente de veces para cada una de las variables predictoras mediante el método de Monte Carlo (Dowd, 1998). De acuerdo Bellotti y Crook (2013), la simulación de las condiciones macroeconómicas se puede basar a partir de valores históricos de perturbaciones económicas pasadas, o bien de valores hipotéticos por criterio de experto.

El método de simulación de Monte Carlo se ha convertido en la metodología de medición de riesgos por excelencia (Feria y Oliver, 2004). Asimismo, el avance en la

capacidad computacional ha promovido la implementación de las simulaciones (Van Gestel y Baesens, 2009).

Los escenarios macroeconómicos simulados y los modelos predictivos permiten obtener la distribución de pérdidas que se podrían generar por el riesgos bancarios, lo cual complementariamente con la metodología VAR se obtiene una estimación de la pérdida máxima durante condiciones adversas (Dowd, 1998).

#### 2.4.1.3 Algunas aplicaciones de pruebas de estrés

Las pruebas de estrés se han implementado en estudios en diferentes países, los cuales varían según esquemas metodológicos, modelos matemáticos y evaluación de escenarios macroeconómicos.

En Austria se modelaron las probabilidades de incumplimiento de los préstamos bancarios con base a ocho variables macroeconómicas. Mediante modelos autorregresivos, se estimaron las pérdidas crediticias del sistema bancario y resultó que los bancos tienen la capacidad suficiente para cubrir las pérdidas crediticias (Boss, 2002).

Por otro lado, Wong et al. (2008) desarrolla una prueba de tensión para calcular el riesgo de la cartera de préstamos hipotecarios de Hong Kong, en el cual se incorporaron variables de intervención en relación a la crisis financiera asiática. Se modeló a partir de regresiones múltiples y modelos autorregresivos, la probabilidad de incumplimiento por distintas variables macroeconómicas. Aunado a esto, se efectuó una simulación de escenarios mediante Monte Carlo, con la cual se obtuvo la distribución de las probabilidades de incumplimiento y mediante de la metodología VAR, se observó que con un 99% de confianza, los bancos presentan un riesgo moderado en los escenarios estresados.

Covas, Rumpb y Zakrajšek (2013) realizaron una prueba de tensión en 14 bancos de Estados Unidos, en la cual se evaluó el riesgo de crédito estimando las pérdidas de capital mediante modelos de regresión cuantílica dinámica con un escenario macroeconómico severamente adverso. Se concluyó que la mayoría de estos bancos tienen la capacidad de absorber las pérdidas en escenarios coyunturales.

En Bellotti y Crook (2013) se evaluó el riesgo de crédito modelando la probabilidad de incumplimiento de tarjetas de crédito del Reino Unido mediante modelos de sobrevivencia discreta. Se implementó una simulación de Monte Carlo con el objetivo de obtener escenarios macroeconómicos extremos, además a partir de la descomposición de Cholesky se conservó la estructura de covariancias entre las variables predictoras. Con la metodología VAR, se obtuvo la distribución de las probabilidades de incumplimiento y asumiendo valores constantes de la pérdida en caso de impago y la exposición en el momento de impago, se estimó la pérdida crediticia en condiciones adversas.

En Latinoamérica, Benavidez y Jimenez (2015) realizaron una prueba de estrés en 65 entidades financieras mexicanas. Se modelaron las probabilidades de incumplimiento por medio de regresiones aparentemente no relacionadas (SUR, por sus siglas en inglés). De igual manera, se desarrolló una simulación de escenarios, donde se obtuvo que la pérdida estimada no aumenta significativamente en situaciones de estrés por lo cual el riesgo de crédito es moderado en las distintas instituciones financieras.

En 2018, la Autoridad Bancaria Europea coordinó una prueba de estrés en 48 bancos de la Unión Europea, donde se evaluó el coeficiente de solvencia de capital a partir de escenarios hipotéticos con un horizonte de tres años. En este ejercicio, cada uno de los bancos efectuó sus propios modelos mediante un marco metodológico común. La prueba de estrés contribuyó a que 25% de los bancos realizaran modificaciones en sus gestiones de riesgo de crédito (EBA, 2018).

En Costa Rica desde el año 2016, la SUGEF efectúa una prueba de estrés anualmente a las 16 entidades más grandes del sistema financiero costarricense. La prueba evalúa la resistencia de las entidades a escenarios hipotéticos que se fundamentan en las proyecciones del Programa Macroeconómico del Banco Central de Costa Rica y del Fondo Monetario Internacional. Los resultados evidencian que en un escenario no tan severo las instituciones tienen una capacidad razonable para seguir sus operaciones sin que su capital se vea comprometido. En contraste, ante un escenario adverso algunas entidades presentan oportunidades de mejora en relación al ajuste de nivel de capital (BCCR, 2019b).

En conclusión, las pruebas de estrés ofrecen una alternativa de evaluación del riesgo de pérdidas crediticias considerando situaciones de perturbación económica. Cabe destacar que, en Banco Improsa se ha efectuado una prueba estrés siguiendo el marco metodológico de la SUGEF y únicamente evaluando escenarios de estrés hipotéticos, por lo cual en términos de esta investigación es relevante implementar una metodología interna acorde a la necesidad del negocio e innovar en la simulación de escenarios macroeconómicos de estrés.

En el siguiente capítulo se muestran los materiales, métodos y procedimientos de análisis que se utilizan en este trabajo para la realización de la prueba de rendimiento ante una materialización del riesgo de crédito.

## CAPÍTULO III: METODOLOGÍA

En este capítulo se detallan aspectos importantes para llevar a cabo la prueba de estrés de la cartera de crédito de la institución financiera costarricense, como los datos utilizados, métodos analíticos de series de tiempo y modelos de Aprendizaje Automático Supervisado, técnicas de validación de modelos y especificaciones de la estimación de las pérdidas crediticias como los supuestos de los escenarios macroeconómicos y aspectos en la generación de la simulación de Monte Carlo.

### 3.1 Materiales

A continuación, se describen los insumos de información necesarios para la evaluación de los modelos predictivos: transformación de las tasas de cura e incumplimiento, variables de macroeconómicas, segmentos bancarios de interés y la ventana de análisis a considerar.

#### 3.1.1 Datos

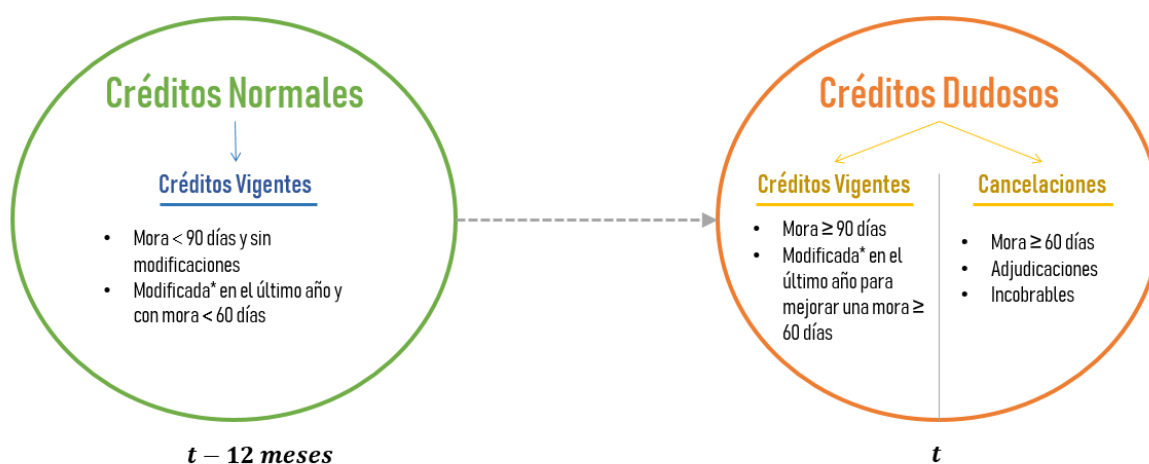
Los datos considerados para el pronóstico de las tasas de cura e incumplimiento corresponden a los referentes al reporte de información crediticia, que se envía mensualmente a SUGEF como parte de los procesos de supervisión bancaria. Por otro lado, las variables predictoras corresponden a distintos indicadores macroeconómicos calculados por el BCCR.

Como se mencionó anteriormente, la pérdida esperada crediticia basa su estimación tanto en los clientes que caen en un estado de morosidad como los que recuperan de este estado. Según SUGEF (2020), el cálculo de la probabilidad de incumplimiento, también conocida como tasa de incumplimiento, se refiere al porcentaje de los créditos normales que posterior al transcurrir un año pasan a créditos dudosos. Por lo tanto, el cálculo de la tasa de incumplimiento está dado por la siguiente fórmula:

$$PD = \frac{\text{Créditos Dudosos}_t}{\text{Créditos Normales}_{t-12}}$$

Para una mayor comprensión, en la Figura 2 se detalla el proceso de cálculo de la probabilidad de incumplimiento.

**Figura 2. Proceso de cálculo de la tasa de incumplimiento**



\*Modificado: operaciones crediticias prorrogadas, refinanciadas o readecuadas.

Fuente: Elaboración propia.

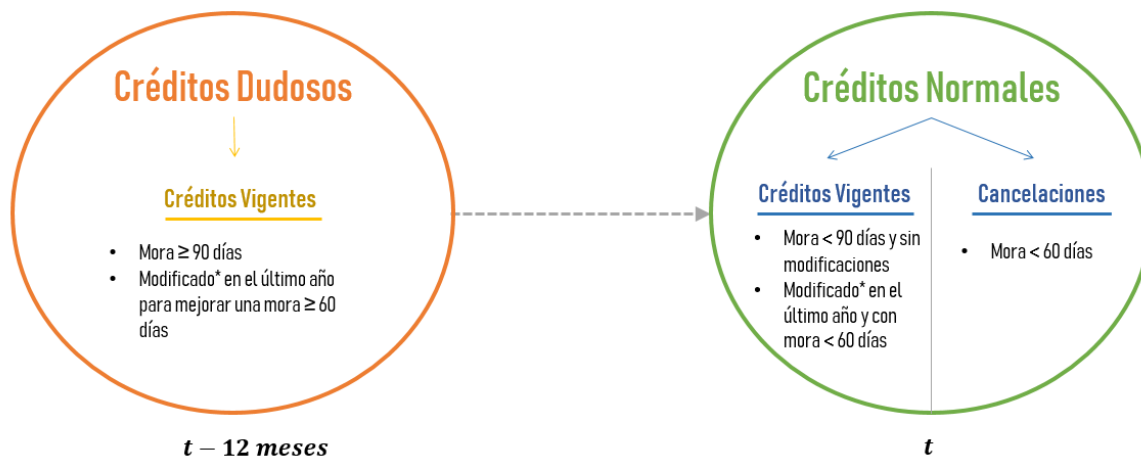
Por ejemplo para calcular la tasa de incumplimiento de diciembre 2020, se deben considerar las operaciones dudosas del periodo, las cuales se componen de los créditos vigentes que tengan una mora mayor o igual 90 días, las operaciones que hayan sido prorrogadas, refinanciadas o readecuadas, con el fin de mejorar su situación de atraso de 60 días o más, las operaciones adjudicadas, incobrables y cancelaciones al momento de la finalización tuviesen una mora de 60 días o más. Además, se deben tomar en cuenta los créditos normales del diciembre 2019, las cuales corresponden a las operaciones crediticias sin modificaciones que hayan tenido una morosidad menor a 90 días y operaciones modificadas que hayan tenido una mora menor a 60 días. Continuando con el ejemplo, se obtiene que diciembre 2019 hubo 2310 créditos normales de los cuales 97 operaciones en diciembre 2020 se clasificaron como créditos dudosos, por lo tanto la tasa de incumplimiento del mes de interés corresponde a 1.16% ( $97 \div 2310 \times 100$ ).

Por otro lado, la tasa de cura se refiere al porcentaje de créditos dudosos que posterior a 12 meses pasan a un estado de crédito normal o cancelado SUGEF (2020). La tasa de cura se calcula según la siguiente fórmula:

$$TC = \frac{\text{Créditos Normales}_t}{\text{Créditos Dudosos}_{t-12}}$$

Asimismo, en la Figura 3 se presenta el proceso de cálculo de la tasa de cura crediticia.

**Figura 3. Proceso de cálculo de la tasa de cura**



\*Modificado: operaciones crediticias prorrogadas, refinanciadas o readecuadas.

Fuente: Elaboración propia.

En caso particular de calcular la tasa de cura de diciembre 2020, se deben tomar en cuenta las operaciones normales del periodo, las cuales se componen de los créditos vigentes que tengan una mora menor 90 días, las operaciones que hayan sido prorrogadas, refinanciadas o readecuadas con un atraso menor a 60 días y las cancelaciones al momento de la finalización tuviesen una mora menor a 60 días. Por otro lado, se deben considerar los créditos dudosos del diciembre 2019, es decir a las operaciones crediticias con mora 90 días o más y operaciones modificadas para mejorar una situación de atraso de 60 días o más. Por lo tanto, si se tiene que en diciembre 2019 hubo 652 créditos dudosos de los cuales 167 operaciones en diciembre 2020 se clasificaron como créditos normales, por lo tanto la tasa de cura del mes de interés corresponde a 25.61% ( $167 \div 652 \times 100$ ).

Para el cálculo de tasas de cura e incumplimiento se consultan diversas tablas internas que alimentan al Sistema de Captura, Verificación y Carga de Datos (SICVECA), el cual corresponde a la plataforma tecnológica que permite el envío y validación preliminar de información cuantitativa entre las diversas entidades financieras y SUGEF. Por ende, se desarrolló una consulta avanzada mediante el *software Oracle SQL Developer*, para así obtener la información necesaria del respectivo cálculo de las tasas, cabe destacar que la

consulta fue debidamente validada por uno de los desarrolladores del banco. Posteriormente, se realizó el cálculo de tasas de cada uno de los meses mediante una función propia en el *software estadístico R*.

#### 3.1.1.1 Transformación de variables respuesta

En 1997, McKinsey & Company desarrolló un modelo de riesgo crédito mediante una prueba de estrés de diversos factores macroeconómicos (Boss, 2002). Con objetivo de estimar la pérdida por riesgo de crédito se modeló la probabilidad de incumplimiento con una transformación mediante la función *logit* (Wilson, 1997). No obstante, Wallis (1987) fue quien propuso este tipo de transformación debido a dos argumentos: i) permite estabilizar la variancia a lo largo del tiempo, y ii) dado que los valores transformados presentan un rango de  $-\infty$  a  $+\infty$ , asegura que las estimaciones y proyecciones de tasas se encuentren dentro del rango de 0 a 1.

La transformación por la función *logit* está dada de la siguiente manera:

$$y_t = \ln\left(\frac{p_t}{1 - p_t}\right)$$

Sea,  $p_t$ : tasa de cura o incumplimiento

En términos de esta investigación, se implementó la transformación por la función de *logit* tanto para la tasa de incumplimiento como para la tasa de cura, para así tener la posibilidad de implementar distintos modelos predictivos sin el inconveniente que las estimaciones de las variables respuesta presenten valores fuera del rango de 0 a 1.

#### 3.1.1.2 Variables Macroeconómicas

El Comité Basilea decreta que las entidades supervisoras tienen la potestad de modificar o imponer los escenarios macroeconómicos considerados en las pruebas estrés internas de las entidades financieras (BCBS, 2009). Por lo tanto, en esta investigación los

factores macroeconómicos a considerar se basan en los establecidos por SUGEF en la pruebas de estrés BUST 2019-2020 (SUGEF, 2019b; SUGEF, 2020).

Cabe destacar que la variable respuesta representa una variación del comportamiento de la cartera al transcurrir doce meses, por lo cual cada uno de las variables independientes corresponde a una variación interanual de indicadores macroeconómicos. Por ejemplo, la variación interanual del tipo de cambio compra (TCC) de mayo 2020 es:

$$\begin{aligned} \text{Variación interanual TCC} &= \left( \frac{TCC_{\text{mayo 2020}}}{TCC_{\text{mayo 2019}}} - 1 \right) \cdot 100\% \\ &= \left( \frac{570.22}{592.27} - 1 \right) \cdot 100\% = -3.72\% \end{aligned}$$

La fuente de origen de las variables predictoras corresponde al repositorio de indicadores económicos del BCCR, en el cual se presentan las actualizaciones de diversas series históricas. Debido a que se consultan varios indicadores, se desarrolló una función en R que se conecta mediante un API<sup>2</sup> al *Web Services* del BCCR, el cual permite descargar un archivo de datos en formato XML con los valores del indicador en un rango deseado de fechas.

A continuación cada de los indicadores financieros a considerar en el modelado de las tasas de cura e incumplimiento:

*Índice de Precios al Consumidor (IPC)*: el Instituto Nacional de Estadística y Censos (INEC) es la institución responsable del cálculo el este indicador (Ley N° 9694, 2019). El IPC es “un instrumento estadístico que permite medir la evolución, a través del tiempo, de los precios de un conjunto de bienes y servicios representativo del consumo final de los hogares, relacionando siempre el precio del mes actual con respecto al mes anterior” (INEC, 2020). Según Krugman, Wells y Olney (2006), tanto la inflación como la deflación pueden

---

<sup>2</sup> API es una abreviatura de *Application Programming Interfaces*, que en español significa interfaz de programación de aplicaciones.

ocasionar problemas en la economía, por lo es importante contar con indicadores que cuantifiquen las variaciones que se producen en el nivel general de los precios de la economía a lo largo de tiempo, y de acuerdo con el INEC el IPC tiene el objetivo de cumplir como indicador macroeconómico de la inflación, entendida esta como el incremento continuo y generalizado del nivel de precios de los bienes y servicios de una economía, en un período determinado (INEC, 2020).

*Índice Mensual de Actividad Económica (IMAE)*: según el BCCR es un indicador a corto plazo que tiene como finalidad aproximar la evolución de la economía, específicamente del valor agregado de diferentes industrias que conforman el PIB a precios básicos (BCCR, 2012). En la prueba de estrés BUST de SUGEF, se propone como variable predictora la variación interanual del PIB real (SUGEF, 2020), el cual corresponde al valor total de bienes y servicios producidos en la economía en un año (Krugman, Wells y Olney, 2006). En términos de esta investigación, dado que la periodicidad de las variables respuesta es mensual y no existe una estimación del PIB mensual, se utiliza el IMAE puesto que el BCCR afirma que la variación interanual del IMAE sirve como aproximación de la evolución mensual del Valor Bruto de Producción (VBP), que a su vez es una aproximación del PIB (BCCR, 2012).

*Tipo de cambio de compra dólar estadounidense (TCC)*: este indicador corresponde al precio en colones en que se cotiza el dólar en el mercado de divisas (Krugman, Wells y Olney, 2006). Por ejemplo, para obtener un dólar en diciembre del 2020, se necesitaban 617 colones.

*Tasa Básica Pasiva (TBP)*: el BCCR es la entidad encargada de la metodología y divulgación del cálculo de la Tasa Básica Pasiva (TBP), el cual corresponde a un promedio ponderado de las tasa de interés que los intermediarios financieros pagan de los recursos de dinero que reciben en colones, como cuentas corrientes, cuentas a la vista o certificados a plazo (BCCR, 2015).

*Crédito del sistema bancario nacional*: corresponde al monto, expresado en millones de colones, de los créditos formalizados al sector privado no financiero otorgados por los

intermediarios financieros del Sistema Bancario Nacional (SBN), que incluyen tanto bancos públicos como privados. Este indicador se presenta desglosado por tipo de moneda, lo que permite analizar la variación interanual del crédito del sistema financiero al sector privado no financiero en moneda nacional y extranjera. Es importante resaltar que, antes de calcular la variación interanual, el crédito del sistema bancario nacional se ajustó correctamente mediante el índice de precios para garantizar una comparación adecuada a lo largo del tiempo, ya que su unidad de medida original correspondía a una moneda

*Costos de construcción:* el BCCR en conjunto con el Colegio Federado de Ingenieros y Arquitectos realizan la Encuesta Trimestral de Avance de Proyectos de Construcción Privada (ETAPCP), la cual presenta un esfuerzo en la mejora de la calidad y alcance de la medición de las estadísticas de construcción; y permite realizar un seguimiento de avance a las construcciones desde el inicio de su proceso constructivo hasta su finalización con el fin de obtener lo efectivamente construido en el trimestre (Rivera, 2014). Se consideran dos indicadores: el costo de vivienda clase media y el costo de edificios; no obstante, dado que se caracterizan por una periodicidad trimestral, se implementó una interpolación *spline* cúbica para obtener una periodicidad mensual, la cual consiste en estimar valores intermedios de  $n-1$  polinomios cúbicos entre puntos consecutivos (Sheid, 1991).

### 3.1.1.3 Segmentos

Banco Improsa, al tener una clara orientación estratégica empresarial, segmenta clientes según su tamaño en grandes, medianas, pequeñas y micro empresas. En Costa Rica existen distintos criterios para clasificar el tamaño de una empresa, los cuales se basan en indicadores como el número de empleados, el monto de ingresos anuales y el valor de sus activos (Ver Anexo I). Por ejemplo la Caja Costarricense del Seguro Social (CCSS) clasifica a las empresas según el número de empleados: microempresas (1-5 trabajadores), pequeñas (6-30 trabajadores), medianas (31-100 trabajadores) y grandes (más de 100 empleados) (Cámara de Comercio de Costa Rica, 2012). Por otro lado, el Ministerio de Economía, Industria y Comercio (MEIC) clasifica las empresas según el número de empleados e

ingresos anuales dependiendo su actividad económica, este criterio es el que implementa en Banco Improsa para la segmentación de clientes (MEIC, 2017).

Internamente en el banco, debido a su enfoque empresarial existe un interés especial de las empresas MIPE, por lo cual la predicción de las tasas de cura e incumplimiento se realizó para la totalidad de la cartera crediticia y tres segmentos referentes al tamaño: micro empresas, pequeñas empresas y en conjunto las medianas y grandes empresas. Por lo tanto, el proceso de selección del mejor modelo de pronóstico se replica cuatro veces para cada uno de los tipos de las tasas.

#### 3.1.1.4 Ventana de análisis

Las tasas de cura e incumplimiento crediticio se caracterizan por tener una periodicidad mensual y se calcularon para el período de julio de 2014 a julio de 2022. Se limitaron las series temporales a este período debido a que la segmentación de los clientes según el tamaño de la empresa se registra a partir del año 2014.

#### 3.1.1.5 Herramienta de análisis y aplicación web

El programa utilizado para el análisis de datos fue *R Statistics*, versión 4.1.0, cuyo procesamiento abarca desde la preparación y visualización de los datos hasta el modelado de las tasas cura e incumplimiento para los segmentos de interés. El código se muestra en el anexo 33, donde se detallan los 26 paquetes implementados y se aprecia un total de 17 funciones creadas. Por otro lado, R brinda una variedad de paquetes que amplían al usuario diversas técnicas de procesamiento, modelado y visualización. El paquete Shiny facilita la generación de aplicaciones web que agilizan la comprensión de información de manera interactiva, al permitir cambiar distintos parámetros de los análisis (Chang et al, 2015). En términos de esta investigación, Shiny sofisticó la visualización de resultados de las pruebas de estrés, al permitir cambiar en tiempo real los escenarios de variables macroeconómicas para el pronóstico de las tasas de interés y estimación de la distribución de pérdidas debido a la materialización del riesgo de crédito.

## 3.2 Métodos

En esta sección, se describen los métodos analíticos de series de tiempo y los modelos de aprendizaje automático implementados para predecir las tasas crediticias. También se detalla el proceso de selección y validación de los modelos. Cabe destacar que, para cada uno de los modelos, se consideraron como variables predictoras las variables macroeconómicas descritas anteriormente. Además, se explica la metodología de simulación de escenarios de una prueba de estrés que permite obtener la distribución de pérdidas crediticias.

### 3.2.1 Análisis de Series Temporales

El primer enfoque para predecir las tasas de cura e incumplimiento correspondieron a los métodos analíticos de series de tiempo. Según Box y Jenkins (1976), una serie de tiempo es una secuencia de observaciones ordenadas en el tiempo. Existen dos tipos de análisis de series de tiempo: univariantes y multivariantes. En esta investigación, se focalizó en el análisis series de tiempo univariado o univariante, el cual tiene el objetivo de describir el comportamiento de la serie basándose únicamente en observaciones pasadas (Pankratz, 1983). Lo anterior es posible debido a que una característica intrínseca de las series temporales es que las observaciones adyacentes son dependientes (Box y Jenkins, 1976).

Existen dos tipos de series de tiempo: las continuas que corresponden a valores de la serie que se registran para todo momento  $t$  en un intervalo de tiempo; mientras que las series cronológicas discretas son observaciones que se obtienen sólo en momentos particulares (Hernández, 2011). En este estudio, se asume que las observaciones de la serie de tiempo son discretas y equiespaciadas en el tiempo. Por ejemplo en el caso de las tasa de cura ( $z_t$ ) del mes actual  $t$  y las tasas de meses previos ( $z_{t-1}, z_{t-2}, z_{t-3}, \dots$ ) son capaces de pronosticar las tasas para meses posteriores.

Las series cronológicas se componen cuatro patrones: tendencia, estacionalidad, movimiento cíclico y movimientos irregulares, los cuales pueden presentarse individualmente en una serie, o puede ocurrir combinados de distintas formas (Hansen y

Nelson, 2003). Hyndman y Athanasopoulos (2021) describen los componentes tendencia, estacionalidad y ciclo de la siguiente manera:

- Tendencia: es un componente que refleja un aumento o una disminución del comportamiento a largo plazo de los valores de la serie.
- Estacionalidad: es un patrón que se repite en intervalos de tiempo similares como: épocas del año o días de la semana.
- Ciclo: corresponde a fluctuaciones de crecimientos y caídas con duración de al menos de dos años, por lo general están relacionadas al ciclo económico presente en la serie cronológica.
- Movimiento irregular: son variaciones que no responden a un patrón reconocible, es decir son movimientos impredecibles se presentan de manera aleatoria (Hernández, 2011).

El análisis descriptivo de cada uno de los componentes ayuda a mejorar la comprensión de la serie temporal, y además se puede utilizar para mejorar la construcción de algunos modelos para el pronóstico de la serie (Hyndman y Athanasopoulos, 2021).

Las series temporales se encuentran en los distintos ámbitos de la ciencia, en Economía y Finanzas adicionalmente se considera el comportamiento de la serie temporal de interés en relación a otras variables macroeconómicas a lo largo del tiempo (González, 2009). Entre las técnicas que permiten el análisis dependencia de observaciones pasadas y las relación con variables exógenas están los modelos de regresión de series de tiempo, modelos autorregresivos integrados de promedio móvil y redes neuronales autorregresivas, las cuales se detallan en la siguiente sección.

### 3.2.1.1 Regresión con series temporales

Los modelos de regresión de series tienen el objetivo de pronosticar la serie de interés asumiendo que tienen una relación con otras series de tiempo e incluso con los componentes de la serie cronológica como la tendencia y estacionalidad (Hyndman y Athanasopoulos,

2021; Hernández, 2011). La notación matemática de la serie de interés  $z_t$  en términos de la relación lineal con  $p$  variables predictoras  $x_{t1}, x_{t2}, \dots, x_{tp}$  es la siguiente:

$$z_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_p x_{tp} + \varepsilon_t$$

Lo anterior en notación matricial es igual a  $\mathbf{Z} = \mathbf{X}\beta + e$ , donde dimensión de la matriz  $\mathbf{X}$  es  $n \times p$  (Harrell, 2015). La estimación de los coeficientes  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  se obtuvo mediante el método de mínimos cuadrados, en el cual consiste en elegir los coeficientes que minimicen la suma de errores al cuadrado ( $S(\beta) = \sum_{t=1}^T \varepsilon_t^2$ ), que también se puede expresar de la siguiente (Brockwell y Davis, 2016):

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (z_t - \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_p x_{tp})^2$$

En términos de notación matricial, se tiene que  $S(\beta) = e'e = (\mathbf{Z} - \mathbf{X}\hat{\beta})'(\mathbf{Z} - \mathbf{X}\hat{\beta})$ . Para obtener los valores de los coeficientes que minimicen la función objetivo se debe derivar la suma de errores al cuadrado (Barnard, 1963), de esta manera:

$$\frac{\partial S(\beta)}{\partial \beta} = \mathbf{X}'\mathbf{Z} - \mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

$$\hat{\beta} = \mathbf{X}'\mathbf{Z}(\mathbf{X}'\mathbf{X})'$$

Según Box y Jenkins (1976), en el modelo de regresión estándar se asume que los errores no están correlacionados con variancia ( $\mathbf{V} = \sigma_N^2 \mathbf{I}$ ) y la matriz de covariancias igual  $cov(\hat{\beta}) = \sigma_N^2 (\mathbf{X}'\mathbf{X})^{-1}$ , sin embargo en series de tiempo se asume que los errores están autocorrelacionados y normalmente distribuidos, por lo cual el estimador de la matriz de covarianza corresponde a:

$$cov(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \sigma_N^2 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Por lo tanto, el modelo de regresión para series de cronológicas con tendencia lineal y estacionalidad con periodicidad mensual, se denota de la siguiente manera:

$$z_t = \beta_0 + \beta_1 t + \beta_2 I_{Enero} + \beta_3 I_{Febrero} + \dots + \beta_{12} I_{Noviembre} + \beta_{13} x_{t1} + \beta_{14} x_{t14} \dots + \beta_p x_{tp} + \varepsilon_t$$

Donde,

$t$ : variable temporal que indica los meses de la serie, codificada como 1, 2, 3,..., n.

$I_{Enero}$  a  $I_{Noviembre}$ : conjunto de variables indicadoras, con diciembre como mes de referencia.

$x_{t1}$  a  $x_{tp}$ : conjunto de variables predictoras.

$\varepsilon_t$ : término de error.

En caso de un modelo de regresión con tendencia cuadrática, la notación matemática es la siguiente:

$$z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 I_{Enero} + \beta_4 I_{Febrero} + \dots + \beta_{13} I_{Noviembre} + \beta_{14} x_{t1} + \beta_{13} x_{t13} \dots + \beta_p x_{tp} + \varepsilon_t$$

### 3.2.1.2 Modelos Autorregresivos de Promedio Móvil

En particular, los modelos autorregresivos de promedios móviles asume que la serie cronológica es generada por una agregación lineal de choques aleatorios (Box y Jenkins, 1976). Además, estos modelos tienen el objetivo de pronosticar las series a partir de los valores pasados de la misma (Hyndman y Athanasopoulos, 2021). Para una mejor comprensión, es necesario la introducción del concepto de un proceso estocástico, el cual consiste en una sucesión de variables aleatorias  $Y_t$  ordenadas, donde  $t$  puede tomar un número infinito de períodos. (Arce, Mahía, 2009), como se ilustra a continuación:

$$\{Y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, y_2, \dots, y_T, y_{T+1}, \dots\}$$

La distribución de un proceso estocástico está caracterizada por la función de probabilidad conjunta de todas las variables aleatorias para cualquier valor de  $t$ , donde la función de marginal de cada variable o también conocida como densidad incondicional (Shumway y Stoffer, 2011), se denota de esta manera:

$$f_{Y_t}(y_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y_t^2}{2\sigma^2}}$$

De acuerdo con Pankratz (1983), la función de medias de un proceso estocástico proporciona las esperanzas de las variables aleatorias ( $E(Y_t) = \mu_t$ ), mientras que la función de varianzas se estima mediante  $Var(Y_t) = E(y_t - \mu_t)^2$ . Asimismo, la función de autocovariancias de un proceso estocástico describe las covariancias entre las variables, por ejemplo para un rezago de  $k$  la autocovariancia está dada por  $\gamma_k = cov(y_t, y_{t+k}) = E[(y_t - \mu_t)(y_{t+k} - \mu_t)]$  (Arce, Mahía, 2009).

El ruido blanco es un proceso estocástico que se compone de una secuencia de variables aleatorias independientes e idénticamente distribuidas, en el cual se asume que su media es cero, variancia constante y dado la independencia implica que las variables están no correlacionadas y su función de autocovariancia se simplifica a (Box y Jenkins, 1976):

$$\gamma_k = E(y_t y_{t+k}) = \begin{cases} \sigma_y^2 & \text{con } k = 0 \\ 0 & \text{con } k \neq 0 \end{cases}$$

Por otro lado, un proceso estocástico estacionario se caracteriza por una función de distribución conjunta invariante con respecto a al tiempo (Arce, Mahía, 2009). Según Hernández (2011), un es proceso estacionario de segundo orden o suave si se cumple las tres condiciones siguientes:

1. Media constante a lo largo del tiempo,  $E(Y_t) = \mu$ , independiente de  $t$ .
2. Variancia constante a lo largo del tiempo,  $Var(Y_t) = \sigma^2$ , independiente de  $t$ .
3. La autocovariancia depende únicamente de la diferencia de tiempo que los separa,  $\gamma_k = cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$ , independiente de  $t$ .

El teorema de Wold (1938) establece que cualquier proceso estacionario ( $y_t$ ) con media cero puede ser representado según la forma:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \eta_t$$

Donde:

$\psi \in \mathbb{R}$ ,  $\psi_j$  satisface que  $\psi_0 = 1$ ,

$\sum_{j=0}^{\infty} \psi_j^2 < \infty$ ,

$\varepsilon_t$  corresponde a un proceso estocástico de ruido blanco y presenta el error al predecir  $y_t$  con la función lineal de sus valores pasados ( $\varepsilon_t = y_t - E(y_t | y_{t-1}, y_{t-2}, \dots)$ ),

$\eta_t$  corresponde al componente lineal determinístico, cuya covarianza con  $\varepsilon_{t-j}$  es nula ( $cov(\eta_t \varepsilon_{t-j}) = 0$ ).

Al prescindir del componente lineal determinístico ( $\eta_t$ ), cualquier proceso estacionario puede ser explicado por la suma ponderada infinita de los valores pasados con media cero y variancia  $\sigma_\varepsilon^2$ , lo cual se denomina como una representación de media móvil infinita (Sargent, 1982). Una manera de presentación de estos procesos estacionarios es mediante los modelos autorregresivos de promedios móviles (ARMA).

Los modelos autorregresivos de promedios móviles constituyen representaciones teóricas de todas las posibles observaciones de una serie cronológica expresadas una forma algebraica que incorporan elementos probabilísticos (Hernández, 2011). La notación algebraica es la siguiente:

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j}$$

Donde:

$$\phi_0 = 1,$$

$\varepsilon_t$  es un proceso de ruido blanco (Hamilton, 2020).

Según Box y Jenkins (1976) los modelos autorregresivos de promedio móvil se pueden denotar por  $ARMA(p,q)$ , donde  $p$  es el número de periodos rezagados de la variable de interés y  $q$  el número de periodos anteriores de una sucesión de errores. Para una mayor comprensión de esta notación, en el Cuadro 2 se muestra la representación matemática de varios modelos ARMA.

**Cuadro 2. Notación matemática de modelos ARMA**

<b>Modelo ARMA</b>	<b>Representación matemática</b>
ARMA(0,0)	$y_t = \phi_0 + \varepsilon_t$
ARMA(0,1)	$y_t = \phi_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$
ARMA(1,0)	$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$
ARMA(1,1)	$y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

Fuente: Elaboración propia.

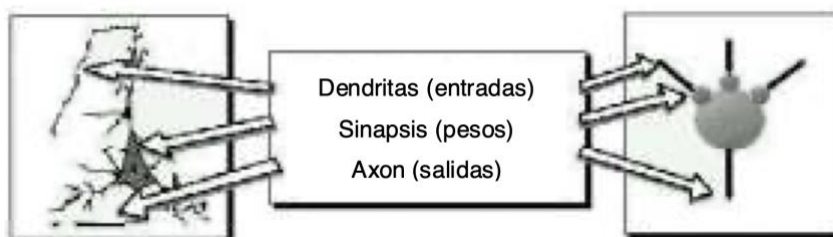
Para los fines de esta investigación, la elección del mejor modelo ARMA se aborda desde la perspectiva de lograr un mayor poder predictivo mediante la calibración de las posibles permutaciones de los valores de  $p$  y  $q$ . Por lo tanto, se prescinde del análisis exploratorio visual del comportamiento de la función de autocorrelación y la función de autocorrelación parcial (ACF y PACF, respectivamente, por sus siglas en inglés). Según Gamboa (2023), la selección de modelos a través de métodos tradicionales, como correlogramas, tiene la limitación de no abarcar un espectro amplio de coeficientes. Por ello, recomienda la estimación mediante la sobreparametrización con permutaciones, ya que tiende a capturar de manera más efectiva el comportamiento de una serie y a obtener pronósticos más precisos que los métodos tradicionales. Por otro lado, debido a la naturaleza de la variable respuesta, la cual se caracteriza por establecer el coeficiente de dos indicadores que hacen referencia a dos momentos separados por 12 meses, no se ha considerado ninguna

diferenciación para lograr una serie estacionaria. Este aspecto podría representar una posible área de mejora para investigaciones futuras.

### 3.2.1.3 Redes Neuronales Autorregresivas

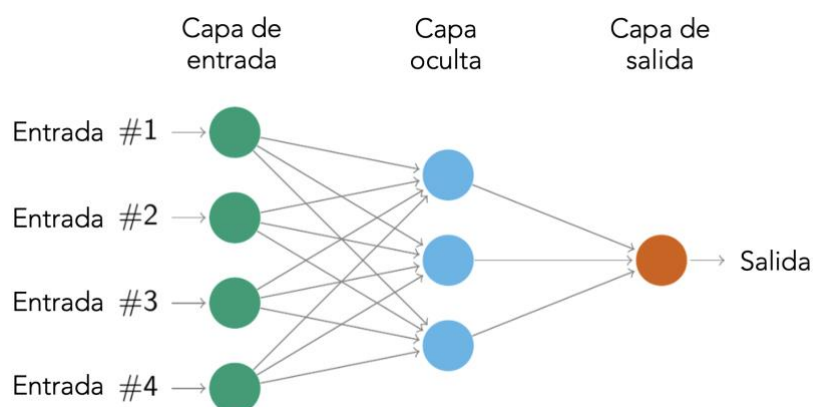
Los modelos matemáticos por redes neuronales surgen a raíz de la inspiración en los sistemas nerviosos del cerebro humano, que mediante a su estructura las conexiones de nodos o neuronas artificiales logran resolver problemas complejos de procesamiento de información (Izaurieta y Saavedra, 2000). Para mayor comprensión de esta analogía, en la Figura 4 se muestra una comparación de una neurona biológica y una artificial, las cuales procesan la información a través entradas (dendritas), establecen pesos (sinapsis) y generan salidas (axón) (Matich, 2001).

**Figura 4. Comparación entre una neurona biológica y una artificial**



Fuente: tomado de *Redes Neuronales: Conceptos Básicos y Aplicaciones*.

Al combinar múltiples neuronas artificiales se forman redes interconectadas masivamente capaces de resolver tareas específicas (Tang y Fishwick, 1993). Las conexiones tienen pesos sinápticos y se distribuyen en tres capas (o más capas), donde la información ingresa por una “capa de entrada”, viaja a través de una “capa oculta” y sale por una “capa de salida” (Acevedo et al, 2017). En la Figura 5, se ejemplifica una red neuronal artificial con cuatro nodos de entrada, una capa oculta con 3 nodos y un nodo de salida (Hyndman y Athanasopoulos, 2021).

**Figura 5. Ejemplo de una red neuronal**

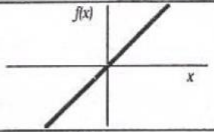
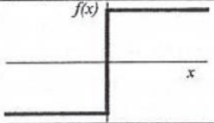
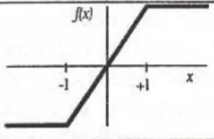
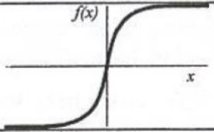
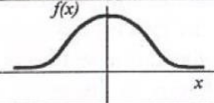
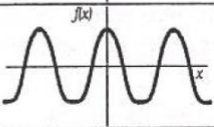
Fuente: tomado de *Forecasting: principles and practice* [Traducción al español].

Para poder obtener el valor estimado de nuestra variable respuesta, el cual equivale en el resultado de la capa salida, es necesario combinar cada una de las salidas de las neuronas con sus ponderaciones establecidas y transformaciones, lo cual es conocido como la regla de propagación (Acevedo et al, 2017). En detalle, en la capa inicial cada una de las entradas se combinan mediante función lineal con los pesos de cada nodo, la cual se representa la siguiente manera:

$$z_j = w_j + \sum_{i=1}^n w_{ij}x_i$$

Al llegar a la capa oculta, en cada uno de los nodos se pondera y se aplica una transformación no lineal o también conocido como función de activación y esto se repite para cada nodo y se envía a la siguiente capa (Karadzic y Pejovic, 2021). Según Hyndman y Athanasopoulos (2021) la función de activación al permitir relaciones no lineales entre la variable respuesta y sus predictoras, faculta que la red solucione relaciones complejas entre las entradas y salidas, y reduce el efecto de valores extremos. En la Figura 6 se presentan ejemplos de función de activación más utilizadas (Acevedo et al, 2017).

Figura 6. Funciones de activación más utilizadas

	Función	Rango	Gráfica
<b>Identidad</b>	$y = x$	$[-\infty, +\infty]$	
<b>Escalón</b>	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
<b>Lineal a tramos</b>	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq +l \\ +1, & \text{si } x > +l \end{cases}$	$[-1, +1]$	
<b>Sigmoidea</b>	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
<b>Gaussiana</b>	$y = Ae^{-Bx^2}$	$[0, +1]$	
<b>Sinusoidal</b>	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Fuente: tomado de *Desarrollo e innovación en Ingeniería*.

Entre las aplicaciones que se le han dado a los modelos de redes neuronales artificiales, se encuentra el procesamiento de lenguaje natural, reconocimiento de imágenes y análisis de datos de distintas áreas, no obstante Sharda (1994) menciona que una las principales áreas en las que se aplican las redes neuronales es en la predicción de series temporales. En la literatura, se han hecho comparaciones entre los métodos clásicos de series de tiempo y las redes neuronales y estas últimas se han caracterizado por estar a la altura y vanguardia de los clásicos (Mitreá, Lee, y Wu, 2009; Cruz, y Corrêa, C, 2017).

Los modelos por redes neuronales autorregresivas (NNAR) consisten en una red de con únicamente una capa oculta, donde sus entradas corresponden a rezagos de la serie y cuya combinación lineal es transformada por una función de activación sigmoidea (Maleki et al, 2018), es decir:

$$f(z) = \frac{1}{1 + e^{-z}}$$

Por lo tanto, su representación matemática corresponde a la siguiente:

$$y_t = w_0 + \sum_{j=1}^k w_j \cdot \frac{1}{1 + (w_{0,j} + \sum_{i=1}^p w_{ij} x_{t-i})} + \varepsilon_t$$

Donde:

$p$  es el numero rezagos (nodos de entrada),

$k$  es el numero nodos en capa oculta,

$\varepsilon_t$  es un proceso de ruido blanco (Maleki et al, 2018).

De igual manera se puede utilizar la notación NNAR( $p,k$ ), donde si se tiene NNAR(4,7) quiere decir que es una red neural con 4 rezagos de la variable respuesta y 7 neuronas en la capa oculta. De acuerdo Hyndman y Athanasopoulos (2021), un modelo NNAR( $p,0$ ) equivale a un ARIMA( $p,0,0$ ) no obstante sin las restricciones en los parámetros para cumplir estacionariedad.

En las redes neuronales autorregresivas existen varios parámetros que pueden ser ajustados, pero para efectos de esta investigación se calibraran: la velocidad a la que los pesos de la red neuronal disminuyen durante el entrenamiento, denominado en inglés como *decay* y el número de nodos en la capa oculta de la red neurona denotado de igual manera como *size*.

### 3.2.2 Modelos Aprendizaje Automático Supervisado

El segundo enfoque empleado para predecir las tasas de cura e incumplimiento corresponde a los modelos de aprendizaje automático supervisado. El aprendizaje automático, o *Machine Learning* en inglés, combina conceptos y técnicas de diferentes campos tradicionales, como las matemáticas, la estadística y la informática (Conway y White,

2012). Según Friedman, Hastie y Tibshirani (2009), los problemas de aprendizaje pueden clasificarse aproximadamente como supervisado o no supervisado, en el presente trabajo se enfoca en el aprendizaje supervisado cuyo objetivo es estimar el valor de una medida en función de una serie de variables de entrada. Las técnicas de Aprendizaje Automático Supervisado que se consideraron son: k vecinos más cercanos, máquinas de soporte vectorial y bosques aleatorios.

### 3.2.2.1 Método de k vecinos más cercanos

De acuerdo Conway y White (2012) el modelo de k vecinos más cercanos (KNN, por su siglas en inglés) es considera el más intuitivo de todos los algoritmos de aprendizaje automático. Este algoritmo consiste en predecir la variable de interés según el promedio de los valores de los vecinos más cercanos dentro de un conjunto de entrenamiento de datos (Friedman, Hastie y Tibshirani, 2009), representado de la siguiente manera:

$$\hat{y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$

Donde  $N_k(x)$  es el vecindario de los valores  $x$  definidos por los k puntos más cercanos.

La cercanía entre las observaciones es medida a partir de la distancia Euclidiana, la cual es una función que permite asociar a un par de objetos mediante número real no negativo (Hernández y Brown, 1970) y está dada por:

$$d(X, Y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}$$

Donde,

$p$  es el número de variables del conjunto de datos,

$X$  y  $Y$  son vectores con los valores de las  $p$  variables de las dos observaciones.

La elección del valor óptimo de  $k$  vecinos depende en la dicotomía del sesgo-variancia o también conocido en inglés como el *bias-variance tradeoff*, puesto que a menor número de vecinos considerados se tiende a obtener un sesgo bajo, pero una variancia alta, es decir a una sobreestimación (James, Witten, Hastie y Tibshirani, 2013). Caso contrario de tomar un valor alto de  $k$  donde se comete una subestimación, por ejemplo, si  $k = n$  cada observación corresponde al promedio de la variable de interés obteniendo un alto sesgo y una baja variancia, por lo cual para términos de estudio se utilizó como valor máximo  $k$  igual a  $\sqrt{n}$ .

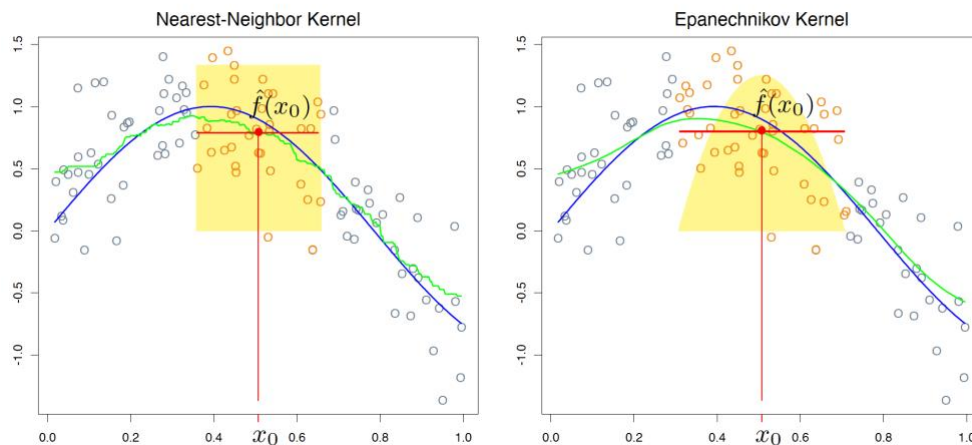
De acuerdo Kuhn y Johnson (2013), para mejorar el rendimiento del algoritmo de KNN se ponderan la contribución de los vecinos en la estimación según su distancia de la observación. En la Figura 7, se presenta una simulación de 100 pares de valores a partir de  $y = \sin(4x) + \varepsilon$ , donde  $x \sim U[0, 1]$ ,  $\varepsilon \sim N(0, \frac{1}{3})$ ; donde la línea azul es una representación de los valores reales y la línea verde corresponde a la estimación por medio de los 30 vecinos más cercanos y se observa como al ponderar las distancias (lado derecho del gráfico) se obtiene una estimación de  $y$  más suave e inclusive si se observa con detalle los valores a la derecha de  $f(x_0)$  tienden a acercar la curva azul de los valores observados. Este proceso de ponderación se realiza a través de dos etapas, la primera la ponderación en sí de las distancias y la segunda que consiste que aplicar una transformación mediante una función de *kernel* a las distancias ponderadas (Mukid et al, 2018). De acuerdo Hechenbichle y Schliep, (2004), las funciones de *kernel* más comunes corresponden a las siguientes:

- Rectangular,  $\frac{1}{2} I(|d| \leq 1)$
- Triangular,  $(1 - |d|) I(|d| \leq 1)$
- Epanechnikov,  $\frac{3}{4}(1 - d^2) I(|d| \leq 1)$
- Biweighth,  $\frac{15}{16}(1 - d^2)^2 I(|d| \leq 1)$
- Triweighth,  $\frac{35}{32}(1 - d^2)^3 I(|d| \leq 1)$
- Coseno,  $\frac{\pi}{4} \cos\left(\frac{\pi}{2}d\right) I(|d| \leq 1)$
- Guassiana,  $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{d^2}{2}\right)$

- Inversa,  $\frac{1}{|d|}$

Estas funciones de transformación de *kernel* corresponden al hiperpárametro se va ser probado para calibrar el modelo y encontrar el que mejor ajustara los datos.

**Figura 7. Comparación entre aplicación de ponderación uniforme o Epanechnikov en el modelo KNN**



Fuente: tomado de *The Elements of Statistical Learning*.

Una desventaja que presenta este algoritmo es que conforme el número de variables aumenta, el rendimiento del modelo tiende a disminuir puesto que computacionalmente se dificulta el cálculo de las distancias y por ende la identificación de los vecinos más cercanos, esto es conocido como la maldición de dimensionalidad (James, Witten, Hastie y Tibshirani, 2013).

### 3.2.2.2 Máquinas de soporte vectorial

Las máquinas de soporte vectorial (SVM, por sus siglas en inglés) fueron desarrolladas por Vladimir Vapnik en 1995, con el objetivo originalmente de resolver problemas de clasificación binaria. No obstante, han sido aplicado exitosamente en distintos casos como detección de rostros y caracteres escritos, categorización de texto y en la detección de objetos por visión artificial (Campbell y Cristianini, 1999). De acuerdo Vladimir Vapnik (1995), el algoritmo consiste en hallar un hiperplano de separación óptimo

entre clases de datos, maximizando el margen entre los puntos de datos más cercanos de distintas clases. Una extensión de las máquinas de soporte vectorial corresponde a la regresión de vectores de soporte (Support Vector Regression, SVR), que logra la estimación de una variable de respuesta continua, mediante la introducción de la función de pérdida  $\varepsilon$ -*insensitive* que permite calcular el hiperplano que contenga el mayor número de observaciones dentro de un margen  $\varepsilon$  (nivel de tolerancia) del valor observado de la variable respuesta (Boser, Guyon, y Vapnik, 1992; Zhang y O'Donnell, 2020). Según Hastie y Tibshirani (2009), este algoritmo es menos sensible a valores atípicos en comparación con una regresión.

Dado un conjunto de datos  $\{(x_i, y_i)\}$ , su objetivo es encontrar la función  $f(x)$  que esté a lo más  $\varepsilon$  desviaciones de los valores reales de la variable respuesta  $y_i$ , es decir que se espera las observaciones se encuentren en una banda o tubo entorno a nuestro hiperplano (Martín, 2016). Dado que el objetivo corresponde a minimizar el error de entrenamiento, se parte de la función (Goddard J et al, 2009):

$$f(x) = (w_1x_1 + \dots + w_px_p) + b = \langle w, x \rangle + b$$

Donde:

$$w_i \in \mathbb{R} \quad \forall i = 1, \dots, p \quad \text{y} \quad b \in \mathbb{R}.$$

La aproximación de la función  $f$  se realiza encontrando la banda lo más plano posible, es decir se busca minimizar  $w$ , lo cual puede ser alcanzado al minimizar la norma de los pesos (Zhang y O'Donnell, 2020):

$$\min_w \frac{1}{2} \|w\|^2$$

$$\text{Sujeto a } y_i - \langle w, x_i \rangle - b \leq \varepsilon \quad \text{y} \quad \langle w, x_i \rangle + b - y_i \leq \varepsilon.$$

No obstante para controlar los valores extremos que quedan fuera de la banda, se pueden agregar variables de holguras  $\xi$  y  $\xi^*$ , las cuales determinan cuantas observaciones

pueden ser toleradas fuera de la banda  $\varepsilon$ -insensitive (Vapnik, 1995), por lo tanto, se reformula el problema de optimización de la siguiente manera:

$$\min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

$$\text{Sujeto a } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Donde  $C > 0$  es llama la constante de regulación, ya que determina el equilibrio entre los errores de predicción y la planitud de la función, por lo cual se controla la complejidad evitando el sobreajuste (Goddard J et al, 2009). Por lo tanto, para resolver este problema de optimización se emplea programación cuadrática mediante multiplicadores de Lagrange (Díaz, Rangel y Martínez, 2006).

La regresión de vectores de soporte, diferencia de la regresión lineal múltiple, permite lidiar con datos no lineales introduciendo la función de *kernel*. En este caso, la función que se tiene corresponde a la siguiente:

$$f(x) = \langle w, \Phi(x) \rangle + b$$

Por lo cual el problema de optimización es escrito de esta manera:

$$\min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

$$\text{Sujeto a } \begin{cases} y_i - \langle w, \Phi(x_i) \rangle - b \leq \varepsilon + \xi_i \\ \langle w, \Phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

De acuerdo con Zhang y O'Donnell, (2020), SVR no lineal resulta ser más efectivo ya que reduce la complejidad de la optimización en algunos casos al encontrar una relación

no lineal sin la necesidad de buscar un hiperplano de alta dimensionalidad. Algunos de las funciones de *kernel* más comunes son: lineal, polinomial, base radial y tangente hiperbólico (Campbell y Cristianini, 1999). Estas funciones de kernel se va a probar para optimizar los resultados del modelo.

### 3.2.2.3 Bosques aleatorios

Los bosques aleatorios corresponden a un método de aprendizaje supervisado, que se basan en el método de agregación de Bootstrap o también conocido como *Bagging*, el cual es consiste en reducir la varianza de una función de predicción, por lo cual tienden a funcionar especialmente bien para procedimientos de alta varianza y bajo sesgo (Friedman, Hastie y Tibshirani, 2009). La idea fundamental consiste en combinar predicciones de modelos más simples, como lo son los árboles de regresión, para obtener así una predicción más precisa y robusta. Este método fue introducido por Leo Breiman como una mejora de los árboles de decisión, añadiendo el factor aleatorio que permite que cada árbol sea independiente entre sí y no correlacionados (Breiman, 2001).

Asumiendo un conjunto de datos con  $p$  variables predictoras y una respuesta, con  $N$  observaciones, es decir  $(x_i, y_i)$  donde  $i = 1, 2, 3, \dots, N$  con  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ . El algoritmo define las variables y puntos de división, suponiendo que la primera partición se obtienen  $M$  regiones  $R_1, R_2, \dots, R_m$ , se modela la respuesta como una constante  $c_m$  para cada región (Friedman, Hastie y Tibshirani, 2009):

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m)$$

Partiendo del criterio de minimización por la suma de cuadrados  $\sum (y_i - f(x_i))^2$  dentro de cada región. Cada árbol se construye de manera recursiva, dividiendo repetidamente los datos en dos subconjuntos aleatorios basados en el valor de una de las variables de entrada, es decir para cada variable de división  $j$  y punto de corte  $s$ , se busca resolver lo siguiente:

$$\min_{j,s} \left[ \min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

Este proceso se repite recursivamente hasta alcanzar una profundidad máxima del árbol o hasta tener menos de un número mínimo de observaciones en un nodo. Finalmente, la predicción de nuevos datos se realiza mediante la agregación por la media de las predicciones individuales de cada árbol de regresión. En caso de clasificación, se emplea la mayoría de los votos (Liaw y Wiener, 2002).

Con fin de optimizar el rendimiento del modelo, los hiperparámetros por calibrar en este algoritmo corresponden al número de árboles de regresión y el número de variables seleccionadas de manera aleatoria de las  $p$  posibles.

### 3.2.3 Validación de modelos

A continuación se describe el método utilizado para validar los modelos mencionados y las medidas de precisión empleadas para determinar su desempeño.

#### 3.2.3.1 Validación Cruzada

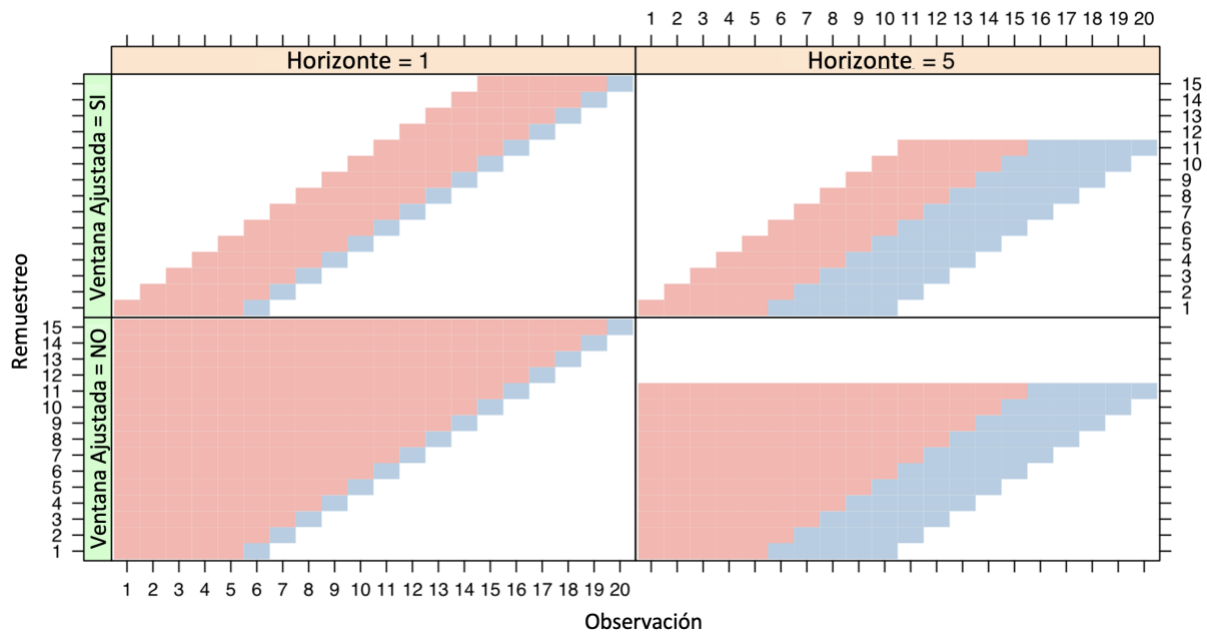
Con el propósito de evaluar el desempeño de los modelos, se empleó la validación cruzada, uno de los métodos más utilizados para estimar el error de predicción (Friedman, Hastie y Tibshirani, 2009). Este método implica llevar a cabo una partición de los datos para crear dos conjuntos distintos: uno de entrenamiento y otro de validación. El conjunto de entrenamiento se utiliza específicamente para entrenar y estimar diversos modelos, mientras que el conjunto de validación se emplea para verificar las predicciones obtenidas.

Debido a que a la variable respuesta que se está analizando corresponde a una serie de tiempo es importante respetar la estructura temporal inherente en los datos, por lo que se aplicó validación cruzada para series de tiempo. Este proceso consiste en que el conjunto de entrenamiento correspondiente consta únicamente de observaciones que ocurrieron antes del

conjunto de prueba, por lo cual, no se pueden utilizar observaciones futuras para construir el pronóstico (Dissanayake et al., 1999).

Existen diversas alternativas para implementar la validación cruzada en el contexto de series temporales, y para ello se ajustan tres parámetros distintos: el número de observaciones de la ventana inicial, el tamaño del horizonte y si se realiza o no un ajuste en la ventana de entrenamiento. Con el fin de proporcionar una comprensión más detallada de estos parámetros, se ilustran en la Figura 8. En esta investigación, se optó por un conjunto inicial de observaciones de entrenamiento con una duración de 70 meses, un horizonte de predicción de 1 mes y la ventana de entrenamiento sin ajuste. En otras palabras, el enfoque adoptado se representa en el cuadrante inferior izquierdo de la Figura 8. Por lo cual el procedimiento de validación cruzada se realizó un total de 29 veces, estas repeticiones también son denominadas como *fold*s.

**Figura 8. Variantes de la validación cruzada para series de tiempo**



Fuente: tomado de *The caret Package: Classification and Regression Training*.

### 3.2.3.2 Medidas de bondad de ajuste

La evaluación de la precisión de los modelos es esencial para garantizar la fiabilidad de las proyecciones y pronósticos. En este sentido, se calcularon dos medidas de bondad de ajuste ampliamente utilizadas para cuantificar la precisión de las predicciones: la raíz del error cuadrático medio (RSME, por sus siglas en inglés) y el error porcentual absoluto medio (MAPE, por sus siglas en inglés). Estas métricas ofrecen una visión cuantitativa de la discrepancia entre los valores observados y los valores predichos. Además, facilitan la comparación entre diferentes modelos y seleccionar el que presente un mejor rendimiento, es decir el modelo con la medida de ajuste de menor valor. Adhikari et al. (2013) menciona las principales características de estas medidas:

RSME: la raíz del error cuadrático medio proporciona una medida de la magnitud absoluta de los errores, penalizando los errores extremos que ocurrieron durante el pronóstico. Si bien la RMSE es una buena medida del error general de pronóstico, no es tan intuitiva ni fácil de interpretar como otras medidas. Se denota de la siguiente manera:

$$RSME = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2}$$

Donde:

$y_t$ : valor observado de la serie de tiempo

$f_t$ : Valor pronosticado

$n$ : tamaño del conjunto de prueba

MAPE: El error porcentual absoluto medio evalúa la precisión relativa de las predicciones y se caracteriza por ser independiente de la escala de medida. Es fácil de interpretar, ya que varía entre 0% y 100%, y cuanto menor sea, menor será el error de las predicciones. Su representación matemática corresponde a la siguiente:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_t - f_t}{y_t} \right| \times 100$$

Donde:

$y_t$ : valor observado de la serie de tiempo

$f_t$ : Valor pronosticado

$n$ : tamaño del conjunto de prueba

### 3.2.3.3 Proceso selección del modelo

En las secciones anteriores, se han detallado los modelos matemáticos implementados y las medidas de bondad de ajuste que se emplean para seleccionar el modelo más óptimo. Ahora, resulta fundamental abordar con claridad el proceso de modelado, incluyendo cada una de sus etapas y alternativas, con el objetivo de alcanzar una comprensión más profunda de la investigación.

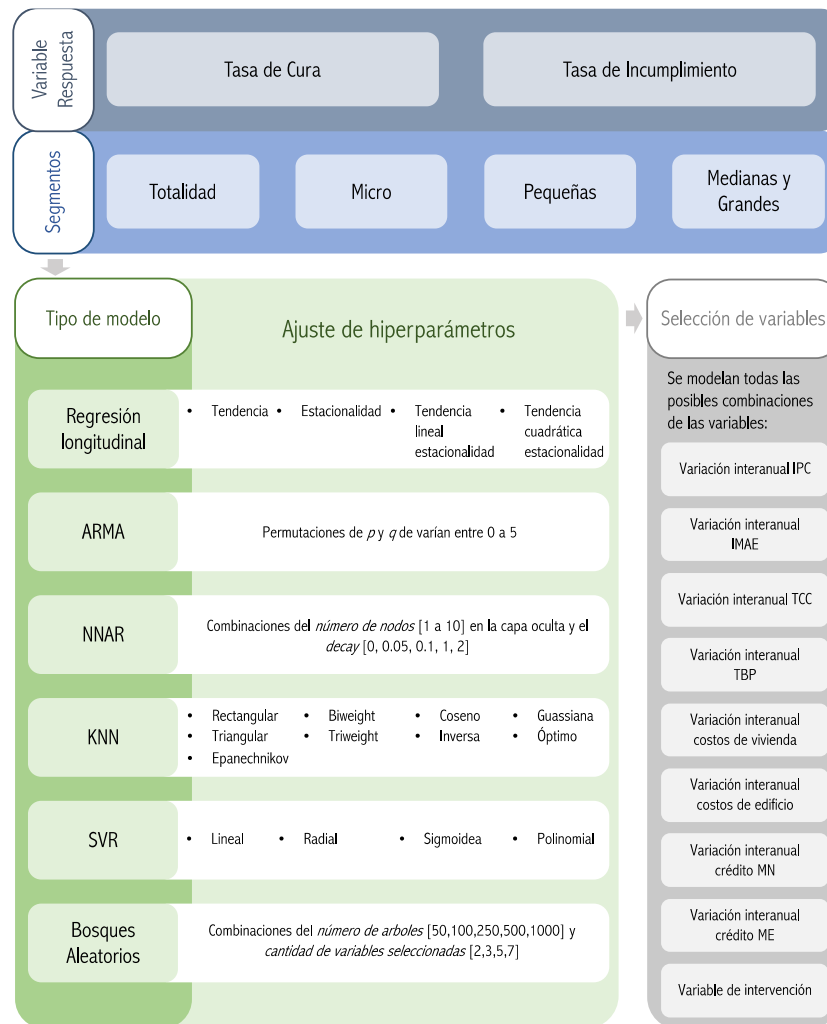
Como se detalló anteriormente, se modelaron dos variables de respuesta: la tasa de cura y la tasa de incumplimiento, abordando dos alternativas de análisis: la totalidad de la cartera y segmentando por tamaño (empresas corporativas, pequeñas empresas y microempresas). Esto conlleva a la necesidad de desarrollar ocho modelos para luego proceder a estimar la pérdida por riesgo de crédito de la entidad bancaria en ambos enfoques. En la Figura 9, se ilustra el proceso que se siguió para obtener estos ocho modelos, a continuación se explica cada etapa:

1. Se selecciona la variable independiente y segmentación de la cartera a modelar.
2. Con todas las variables predictoras se corren cada uno de los tipos de modelo.
3. En cada modelo, se prueban diferentes alternativas de hiperparámetros:
  - a. Regresión longitudinal: 4 variantes.
  - b. ARMA: 36 diferentes modelos debido a las permutaciones posibles de  $p$  y  $q$  (tomando valores entre 0 y 5, con repetición).
  - c. NNAR: 50 combinaciones de número de nodos ocultos y valor del *decay*.
  - d. KNN: 9 variantes del *kernel*.

- e. SVR: 4 variantes.
  - f. Bosques aleatorios: 45 combinaciones del número de árboles y cantidad de variables.
4. Partiendo del principio de parsimonia para cada tipo de modelo, una vez obtenido el óptimo ajuste de hiperparámetros, se procede con el proceso de selección de variables mediante el método de "fuerza bruta". Este enfoque exhaustivo implica evaluar todas las combinaciones posibles de variables para determinar cuál conjunto proporciona el mejor rendimiento del modelo. Por ejemplo, con tres variables (A, B, C), las combinaciones serían {A}, {B}, {C}, {A, B}, {A, C}, {B, C}, {A, B, C}. Es relevante señalar que este procedimiento se llevó a cabo para cada una de las ocho variables macroeconómicas. Sin embargo, en cada combinación, se incluyó una variable de intervención (tratando de reflejar los eventos coyunturales recientes), resultando en la evaluación de 255 alternativas.
  5. Debido a la validación cruzada, cada modelo elaborado se replica 29 veces.
  6. Este proceso se repite para los demás segmentos y cada tipo de tasa.

Dicho lo anterior, es factible calcular el número total de modelos individuales realizados, el cual asciende a 389 296 modelos (véase Anexo 2).

**Figura 9. Etapas en el proceso de modelado de las tasas de incumplimiento y cura**



Fuente: Elaboración propia.

### 3.2.4 Prueba de estrés

En la sección del Marco Teórico, se exploraron minuciosamente las pruebas de estrés, abordando su concepto, la relevancia de estas en la gestión de riesgos bancarios y las diversas metodologías de implementación: el análisis de sensibilidad y el análisis de escenarios. En el contexto de esta investigación, se llevó a cabo una prueba de estrés mediante el análisis de escenarios macroeconómicos. Dado que el enfoque de la misma se centra en la perspectiva de un especialista en estadística y programación, se optó por simular los escenarios

aprovechando los avances computacionales, descartando la opción de establecer los escenarios mediante el criterio de expertos en banca.

#### 3.2.4.1 Simulación de escenarios

Antes de explicar la creación de la simulación de escenarios, es crucial recalcar, tal como se mencionó previamente, que los Acuerdos de Basilea enfatizan la imperativa implementación, por parte de las entidades bancarias, de la metodología de *Value at Risk* (VaR). Esta metodología constituye una herramienta estándar para cuantificar los riesgos inherentes a las operaciones bancarias la cual estima la pérdida máxima esperada en el valor de una cartera crediticia, mediante la simulación de una distribución de pérdidas. Por ejemplo, un VaR de \$600,000 al 99% significa que hay un 1% de probabilidad de que la pérdida exceda los \$600,000 en el periodo dado. Para obtener la distribución de pérdidas, se utilizan simulaciones de Monte Carlo, que manejan efectivamente la complejidad y la aleatoriedad de los mercados financieros (Jorion, 2003), generando así diversos escenarios de las variables macroeconómicas para la predicción de tasas y, por consiguiente, de las pérdidas.

El método de Monte Carlo emplea álgebra básica para llevar a cabo la simulación, mediante la generación de números aleatorios de una distribución de probabilidad específica. Sin embargo, este procedimiento puede dar lugar a escenarios con variables independientes. (Kaczmarzyk, 2016). Dado que el propósito es simular variables macroeconómicas, resulta relevante incorporar las correlaciones intrínsecas entre ellas. En este sentido, se basó en la descomposición de Cholesky, una herramienta que se utiliza para generar variables correlacionadas a partir de variables independientes, ajustando según la matriz de varianza-covarianza (Paredes et al., 2022). En un conjunto de datos con  $p$  variables, sea  $\Sigma$  la matriz de varianza-covarianza. Al aplicar la descomposición de Cholesky, se genera una matriz triangular inferior  $A$ , tal que  $AA^T = \Sigma$ . Si  $z = (z_1, \dots, z_p)$  representa un vector de valores aleatorios independientes con una distribución normal estándar, entonces  $\theta = \mu + Az$  contiene valores aleatorios con una distribución normal multivariada cuya matriz de varianza-covarianza es igual a  $\Sigma$  (Gelman et al., 2021).

Por lo tanto, siguiendo este principio matemático, es posible generar escenarios utilizando la matriz de varianza-covarianza de las variables macroeconómicas de interés. Para ello, se siguieron los siguientes pasos:

- a. Se creó una tabla de 500 filas y 8 columnas, correspondientes a la cantidad de variables macroeconómicas consideradas, inicialmente rellena con ceros.
- b. Para cada fila:
  - i. Se generó un vector de 8 números aleatorios, utilizando una distribución normal estándar.
  - ii. Se llevó a cabo una multiplicación de matrices entre la matriz resultante de la descomposición de Cholesky de la matriz de varianza-covarianza de las variables macroeconómicas y el vector de números aleatorios previamente generado. El producto de esta multiplicación produjo una matriz de dimensiones  $8 \times 1$ .
  - iii. A este resultado se le sumó una matriz de  $8 \times 1$  que contenía las medias de las variables macroeconómicas.
  - iv. Se transpuso el resultado para obtener un vector de longitud 8.
  - v. Estos pasos se repitieron 500 veces para cada fila.
- c. Se añadió una novena columna correspondiente a la variable de intervención. Para generar escenarios más estresados o adversos, esta columna se llenó con valores de uno, asumiendo una continuidad de la situación coyuntural.

Esta metodología permitió la generación de escenarios que reflejan la interacción y comportamiento de las variables macroeconómicas bajo diferentes condiciones, facilitando así el análisis y la evaluación de riesgos en el contexto estudiado.

#### 3.2.4.2 Estimación de la distribución de pérdida

A partir, los escenarios simulados se realizaron predicciones de las tasas de cura e incumplimiento utilizando los modelos obtenidos para cada segmento. Además, retomando la fórmula para calcular la pérdida esperada (PE), como se describió en la sección del Marco Teórico:

$$PE = Tasa\ Incumplimiento * (1 - \%Colateralización) * (1 - Tasa\ de\ Cura) * EAD$$

Se establecieron los siguientes supuestos con respecto a la cartera crediticia con el fin de obtener una distribución de las 500 pérdidas estimadas:

- a. Se fijó el valor de la exposición en caso de impago (EAD) en 90,000,000 dólares. Es importante señalar que este valor no corresponde al real, sino que se estableció de esta manera por razones de confidencialidad de la información. La segregación en los segmentos de tamaño se estableció de la siguiente manera: 40% para empresas corporativas, 25% para pequeñas empresas y 35% para microempresas. Estos porcentajes se fijaron basado en la distribución porcentual de los saldos de la cartera crediticia.
- b. El porcentaje de colateralización se fijó en un 80% para todos los segmentos de la misma manera.

Con esta distribución de pérdidas, el método VAR establece que el promedio de estas corresponde a la pérdida esperada, mientras que el percentil 99% representa la pérdida inesperada. Esto proporciona a la entidad financiera una estimación del riesgo crediticio con un espectro más amplio de posibilidades en lugar de una estimación puntual.

Por último, y siguiendo la normativa de la Superintendencia, al obtener estas estimaciones de las pérdidas, también se permite calcular el indicador del porcentaje de riesgo de crédito, visto en la sección de Marco Teórico. Partiendo del supuesto de que el valor de la cartera es el mismo que el establecido para el EAD.

$$\% \text{ Pérdida Riesgo Credito} = \frac{\text{Pérdida}}{\text{Cartera Total}} \times 100$$

## CAPÍTULO IV: RESULTADOS

En este capítulo se describen los resultados obtenidos de las series históricas de las tasas de cura e incumplimiento, así como de las respectivas variables macroeconómicas descritas en el apartado metodológico. En primer lugar, se realiza un análisis exploratorio de estas variables, seguido por la presentación de los resultados de los modelos de predicción de las tasas para cada tipo de segmentación y su validación correspondiente. Además, se detallan los resultados de la distribución de pérdidas crediticias y se ofrece una interpretación sustantiva de los resultados obtenidos a partir de estos modelos. Por último, se incluyen capturas de pantalla del aplicativo desarrollado en Shiny, que constituye una herramienta de gran utilidad para la aplicación y comprensión de los modelos y la estimación de la distribución de pérdidas.

### 4.1 Análisis Exploratorio

A continuación, se presentan los análisis descriptivos de las series históricas de las tasas de cura e incumplimiento observadas en la cartera crediticia del Banco Improsa desde junio de 2014 hasta julio de 2022, así como de las principales variables utilizadas en el modelado.

#### 4.1.1 Tasas de cura

En el Cuadro 3 se presentan las estadísticas descriptivas de las tasas de cura en función del tamaño de las empresas atendidas por Banco Improsa durante el período de tiempo determinado. Se observa que, para la totalidad de la cartera, las tasas de cura varían en un rango del 2.09% al 20.33%, con una mediana del 9.09% y una media ligeramente superior, del 9.35%. Esto podría sugerir que algunos valores más elevados están incrementando el promedio, aunque la distribución parece ser bastante simétrica. Al segmentar las tasas de cura por tamaño de empresa, se aprecia que las tasas correspondientes a las microempresas oscilan entre el 0% y el 31.1%, con una mediana y una media de 6.2% y 8.3%, respectivamente. En el caso de las pequeñas empresas, las tasas varían desde el 0% hasta el 31.25%, con una mediana de 10.26%, superior a la de las microempresas. Las empresas

medianas y grandes presentan tasas de cura que van desde un mínimo del 0% hasta un máximo del 39.1%, siendo esta la tasa más alta entre las categorías. La mediana es del 9.9%, cercana a la mediana general, pero la media es más alta, del 11.5%, lo que podría indicar una distribución sesgada por valores superiores. En términos generales, se observa que, cuanto mayor es el tamaño de las empresas, más altas tienden a ser las tasas de cura, lo que podría sugerir que las empresas más grandes cuentan con más recursos o sistemas más eficientes para la recuperación de préstamos.

**Cuadro 3. Banco Improsa: Estadísticas descriptivas de la tasa de cura según tamaño de empresa, junio 2014 a julio 2022**

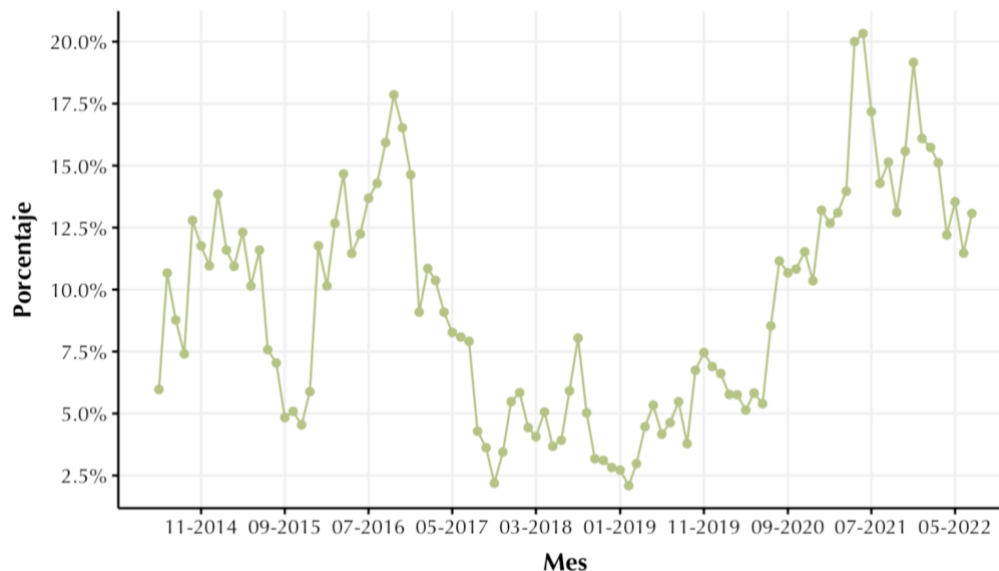
	Totalidad	Micro	Pequeñas	Medianas y Grandes
Mínimo	2.09%	0.0%	0.00%	0.0%
Percentil 25	5.35%	3.6%	4.23%	6.0%
Mediana	9.09%	6.2%	10.26%	9.9%
Promedio	9.35%	8.3%	10.04%	11.5%
Percentil 75	12.76%	10.1%	15.30%	13.1%
Máximo	20.33%	31.1%	31.25%	39.1%

Fuente: Cartera crediticia Banco Improsa

#### 4.1.1.1 Cartera Total

En cuanto a la serie temporal de la tasa de cura para la cartera total de la entidad financiera, en la Figura 10 revela una tendencia al alza que se inicia hacia finales de 2019, culminando en su punto más alto aproximadamente a mediados de 2021. Este incremento podría atribuirse a las fluctuaciones del contexto económico, precipitadas por la adopción de estrategias políticas a nivel nacional con el propósito de mitigar las repercusiones de la crisis sanitaria ocasionada por el coronavirus.

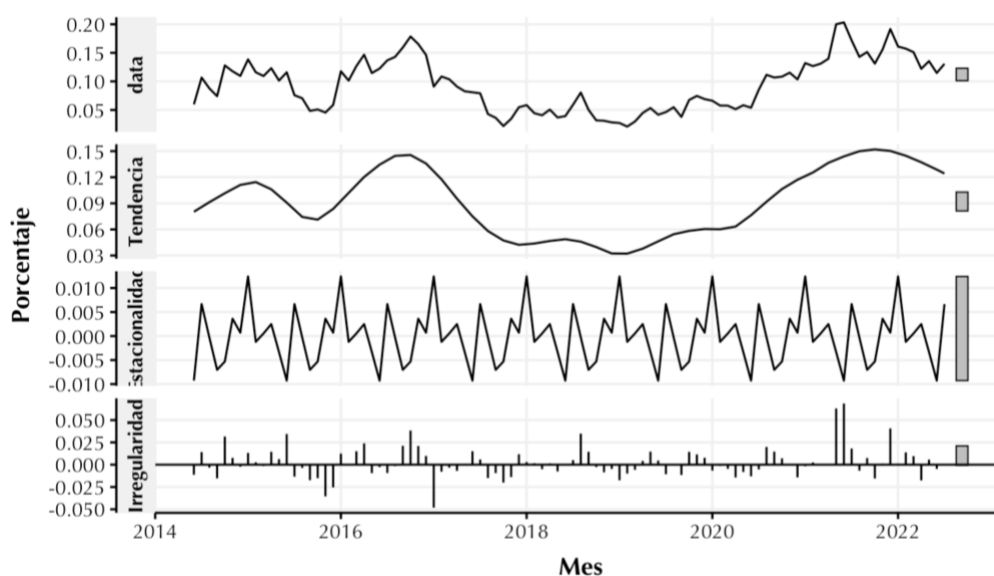
**Figura 10. Banco Improsa: Tasa de cura de la totalidad de la cartera crediticia, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

En la Figura 11 se ilustra la descomposición de la serie temporal correspondiente a la tasa de cura de la cartera total de créditos, desglosándola en tres componentes distintos: tendencia, estacionalidad e irregularidad (residuos). A partir de esta descomposición, se distingue una tendencia subyacente que refleja mejorías progresivas a lo largo del tiempo. Por otro lado, se aprecia un patrón que sube y baja en intervalos regulares, lo cual indica que hay componentes estacionales que ocurren de manera anual. El componente de irregularidad sugiere la presencia de fluctuaciones aleatorias o ruido en los datos, posiblemente atribuibles a variables exógenas imprevistas o a anomalías específicas dentro de la cartera de créditos.

**Figura 11. Banco Improsa: Descomposición de la tasa de cura de la totalidad de la cartera crediticia, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

#### 4.1.1.2 Cartera Segmentada

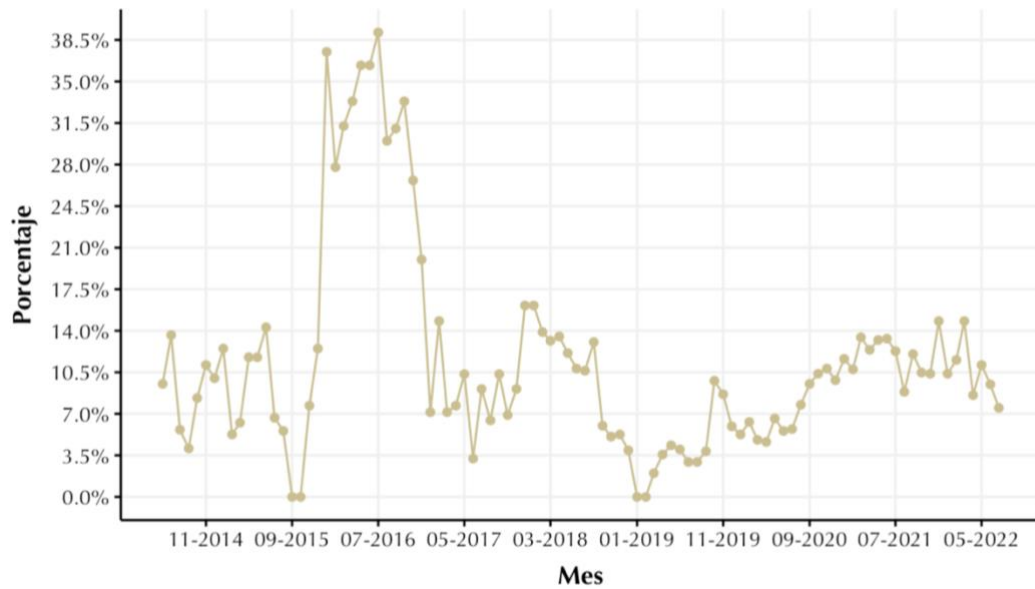
A continuación, se detallan las tasas de curas de la entidad bancaria segmentado según el tamaño de empresas: empresas corporativas, pequeñas empresas y microempresas:

##### 4.1.1.2.1 Empresas medianas y grandes

La Figura 12 evidencia una tendencia descendente en la tasa de cura para este segmento durante los primeros años, caracterizada por una volatilidad significativa que paulatinamente disminuye con el transcurso del tiempo. A partir de 2019, dicha tasa parece alcanzar una estabilidad, con fluctuaciones menos marcadas y una variabilidad más restringida. Es preciso señalar que en los meses de 2016 se observaron las tasas de cura más elevadas para este segmento, lo que podría estar vinculado a diversos cambios estratégicos implementados en la institución financiera en ese año específico. En lo que respecta a la descomposición de la serie temporal (Figura 13), se observa que el componente de tendencia indica una estabilización de la tasa de cura a lo largo del tiempo. Por otro lado, la estacionalidad revela patrones que oscilan con cierta regularidad, lo cual sugiere la influencia de factores cíclicos sobre dichas tasas de cura. En cuanto al componente irregular, se detectan

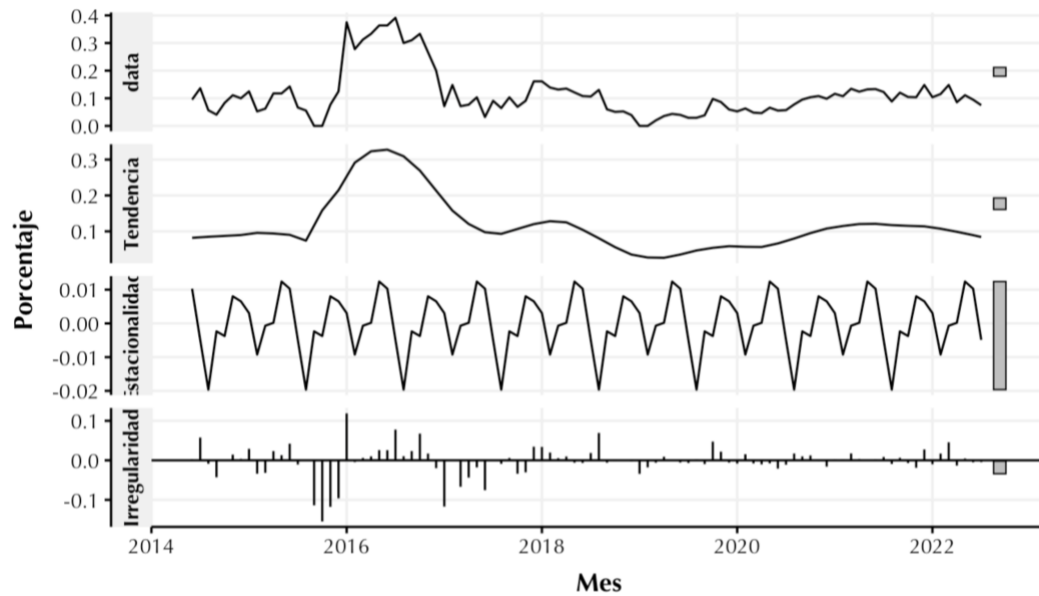
picos pronunciados que podrían atribuirse a variaciones específicas en las estrategias o políticas de crédito, las cuales habrían ejercido un impacto en las tasas de cura.

**Figura 12. Banco Improsa: Tasa de cura de medianas y grandes empresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 13. Banco Improsa: Descomposición de la tasa de cura de medianas y grandes empresas, junio 2014 a julio 2022**

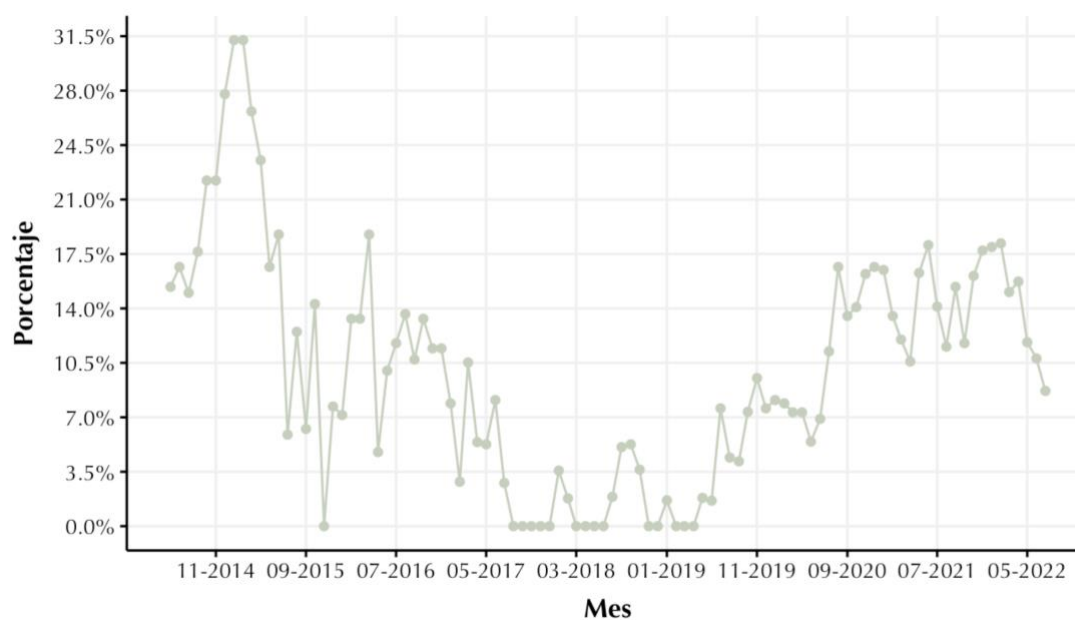


Fuente: Cartera crediticia Banco Improsa

#### 4.1.1.2.2 Empresas pequeñas

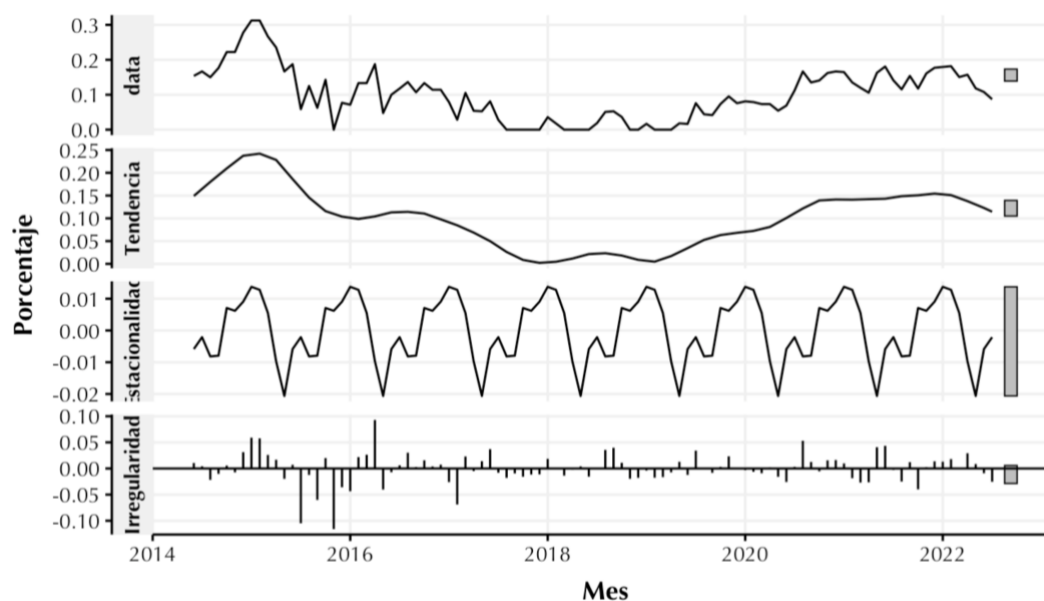
En lo que respecta a las empresas pequeñas, la tasa de cura presenta una considerable volatilidad con picos y caídas pronunciados. Sin embargo, a diferencia de las medianas y grandes empresas, la variabilidad no parece disminuir en el tiempo (Figura 14). Asimismo como se observa en la Figura 15, la tendencia muestra una ligera disminución a lo largo del tiempo. La estacionalidad sugiere la presencia de fluctuaciones que siguen un ciclo aproximadamente trimestral o, en algunos casos, semestral. Esto implica picos y caídas que ocurren regularmente cada 3 o 6 meses. Por otro lado el componente irregular muestra altos y bajos específicamente en los meses alrededor del 2016, lo que puede ser indicativo de la influencia de factores externos o eventos no previstos que afectan la tasa de cura de las pequeñas empresas.

**Figura 14. Banco Improsa: Tasa de cura de pequeñas empresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 15. Banco Improsa: Descomposición de la tasa de cura de pequeñas empresas, junio 2014 a julio 2022**

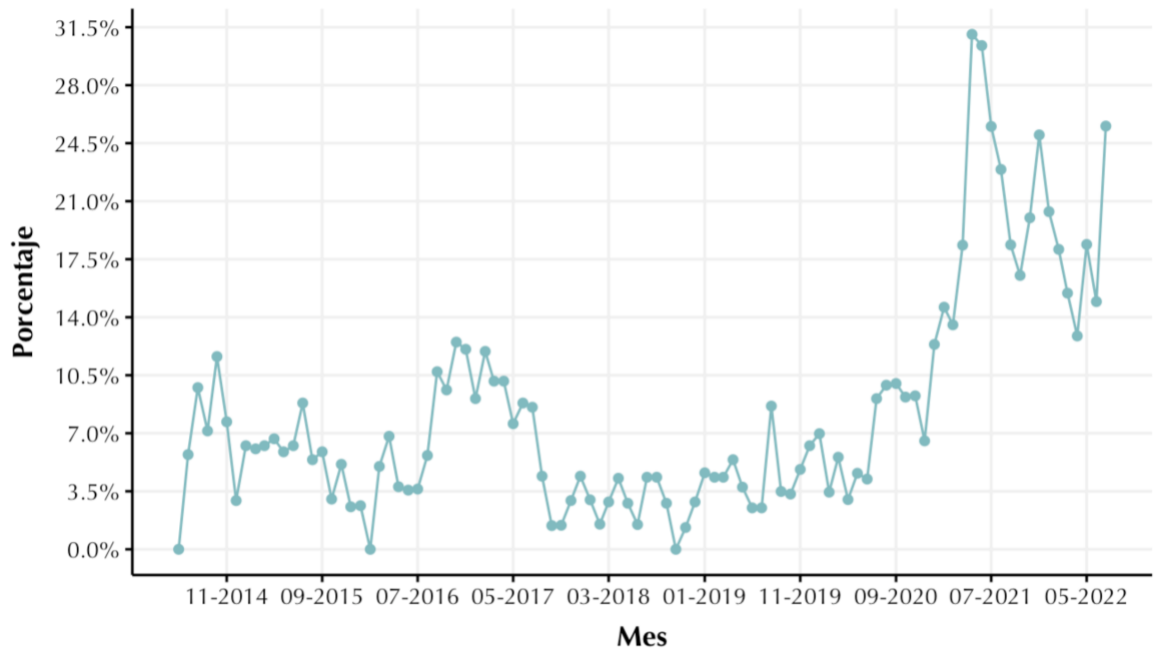


Fuente: Cartera crediticia Banco Improsa

#### 4.1.1.2.3 Microempresas

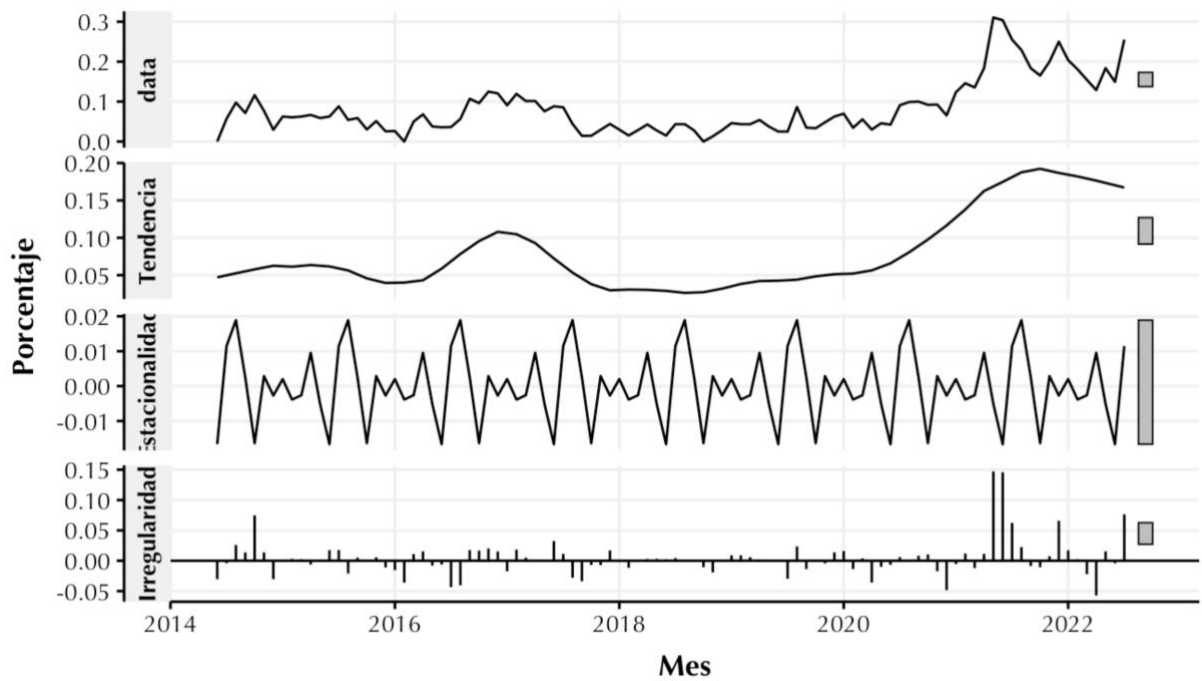
En la Figura 16 se representan las tasas de cura de las microempresas, las cuales evidencian una alta variabilidad, con algunos picos notablemente elevados en los últimos años de la serie. Estos picos están asociados a las políticas implementadas en 2020 para proteger a las empresas de los efectos provocados por la pandemia de COVID-19. Al descomponer la serie (Figura 17), se observa que la tendencia a largo plazo parece ser ascendente, especialmente a partir de finales de 2020, lo que podría señalar una mejora en las tasas de cura o un cambio en las políticas de crédito o recuperación del banco. El componente de estacionalidad exhibe movimientos consistentes, sugiriendo la presencia de patrones estacionales en la recuperación de las microempresas, lo cual quiere decir que en la misma época de cada año, se presentan incrementos o disminuciones que afectan las tasas de manera consistente. Por otro lado, el componente irregular muestra fluctuaciones, con picos y caídas, especialmente en los últimos meses, lo que indica que eventos aleatorios han impactado en la tasa de cura de las microempresas.

**Figura 16. Banco Improsa: Tasa de cura de microempresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 17. Banco Improsa: Descomposición de la tasa de cura de microempresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

#### 4.1.2 Tasas de incumplimiento

El Cuadro 4 proporciona un resumen estadístico de la tasa de incumplimiento según el tamaño de las empresas atendidas por el Banco Improsa desde junio de 2014 hasta julio de 2022. Con respecto a la totalidad de cartera, la tasa media de incumplimiento es del 1.879%, con una mediana ligeramente menor de 1.806%. Esto implica una distribución relativamente simétrica de las tasas de incumplimiento en toda la cartera de crédito. Separando la cartera por tamaño, las empresas corporativas tienen una tasa media más baja de 1.31% y una mediana aún más baja de 1.13%, con un valor máximo de 3.33%. Por otro lado, las pequeñas empresas tienen una tasa de incumplimiento media de 2.14%, muy similar a su mediana de 2.25%, lo que sugiere una distribución equilibrada. Las microempresas tienen una tasa media más alta de incumplimiento de 2.45%, con una mediana de 2.12%, y el valor máximo es considerablemente alto con un 5.19%. En otras palabras, las empresas más grandes presentan un mejor desempeño en el pago de sus créditos o que tienen más recursos para evitar el incumplimiento. Las micro y pequeñas empresas, en cambio, muestran una tendencia hacia tasas de incumplimiento más altas, lo que puede reflejar desafíos adicionales que enfrentan en su flujo de efectivo o acceso a financiamiento que las hace más susceptibles al incumplimiento.

**Cuadro 4. Banco Improsa: Estadísticas descriptivas de la tasa de incumplimiento según tamaño de empresa, junio 2014 a julio 2022**

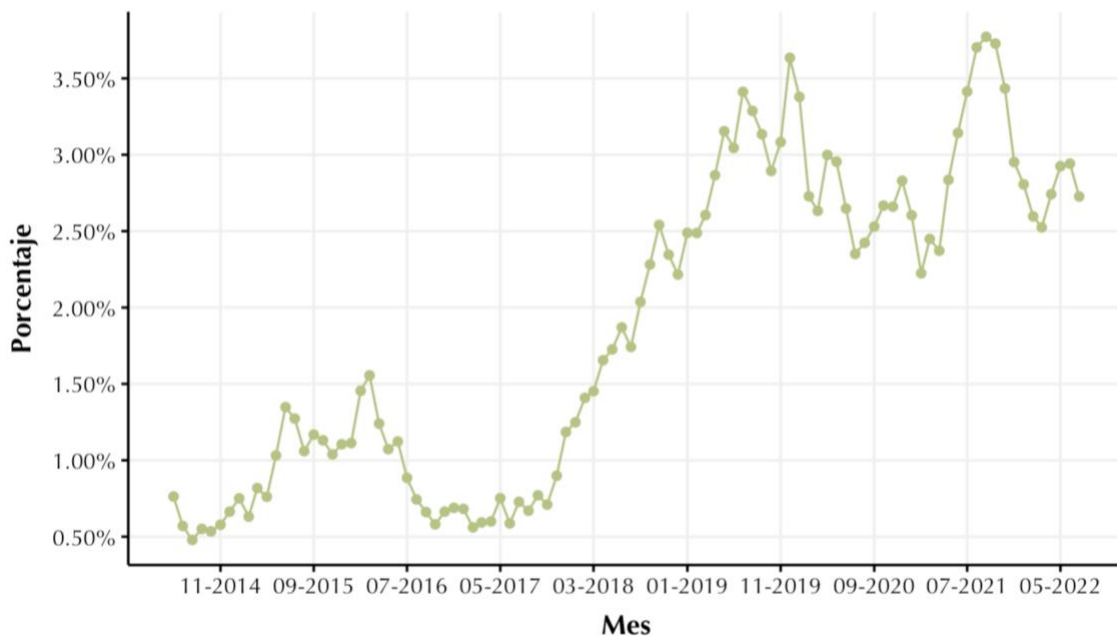
	Totalidad	Micro	Pequeñas	Medianas y Grandes
Mínimo	0.479%	0.68%	0.22%	0.22%
Percentil 25	0.783%	1.56%	0.78%	0.56%
Mediana	1.806%	2.12%	2.25%	1.13%
Promedio	1.879%	2.45%	2.14%	1.31%
Percentil 75	2.738%	3.40%	3.37%	1.91%
Máximo	3.772%	5.19%	4.66%	3.33%

Fuente: Cartera crediticia Banco Improsa

#### 4.1.2.1 Cartera Total

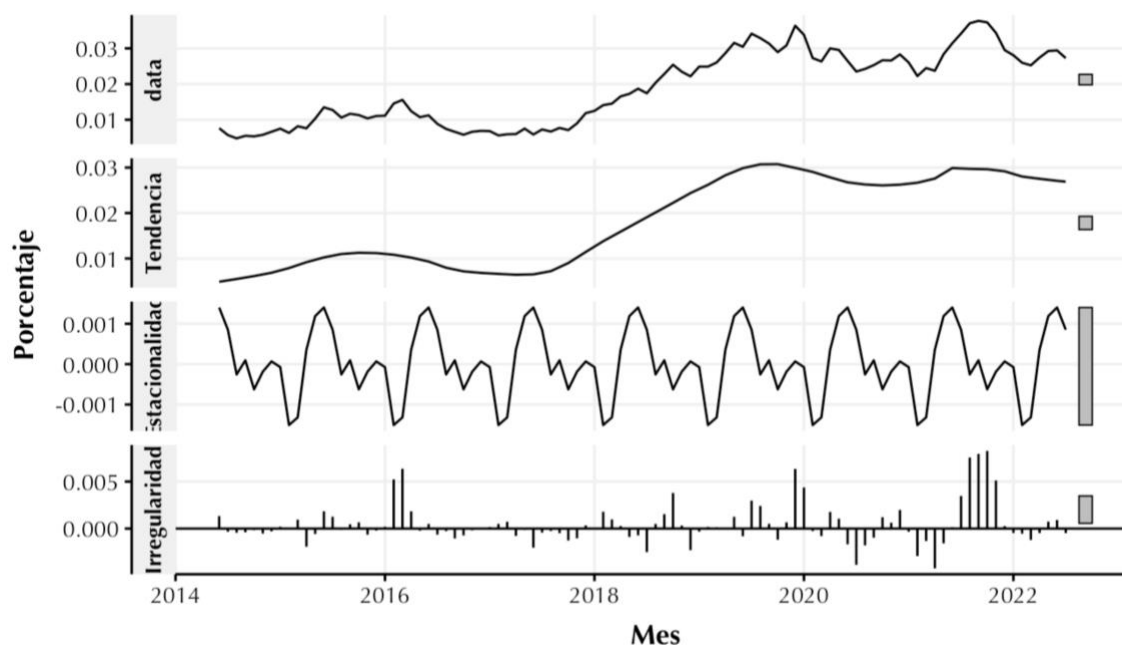
En la Figura 18 se presenta la tasa de incumplimiento de la totalidad de la cartera crediticia del Banco Improsa, la cual se caracteriza por un aumento gradual hasta principios de 2019, seguido de una variabilidad multiplicativa. A partir de este periodo, la tasa muestra una tendencia más plana con fluctuaciones dentro de un rango más definido, mostrando un cambio de nivel escalonado en la serie de tiempo. Con relación a la descomposición de la tasa de incumplimiento de este segmento (Figura 19), la línea de tendencia muestra una inclinación ascendente hasta 2019, sugiriendo un aumento en la tasa de incumplimiento durante ese período, seguido por una estabilización. El componente estacional muestra un patrón anual bien definido, con picos y disminuciones que se repiten cada 12 meses, mientras que la irregularidad muestra picos que representan fluctuaciones que no son explicadas por la tendencia ni la estacionalidad.

**Figura 18. Banco Improsa: Tasa de incumplimiento de la totalidad de la cartera crediticia, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 19. Banco Improsa: Descomposición de la tasa de incumplimiento de la totalidad de la cartera, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

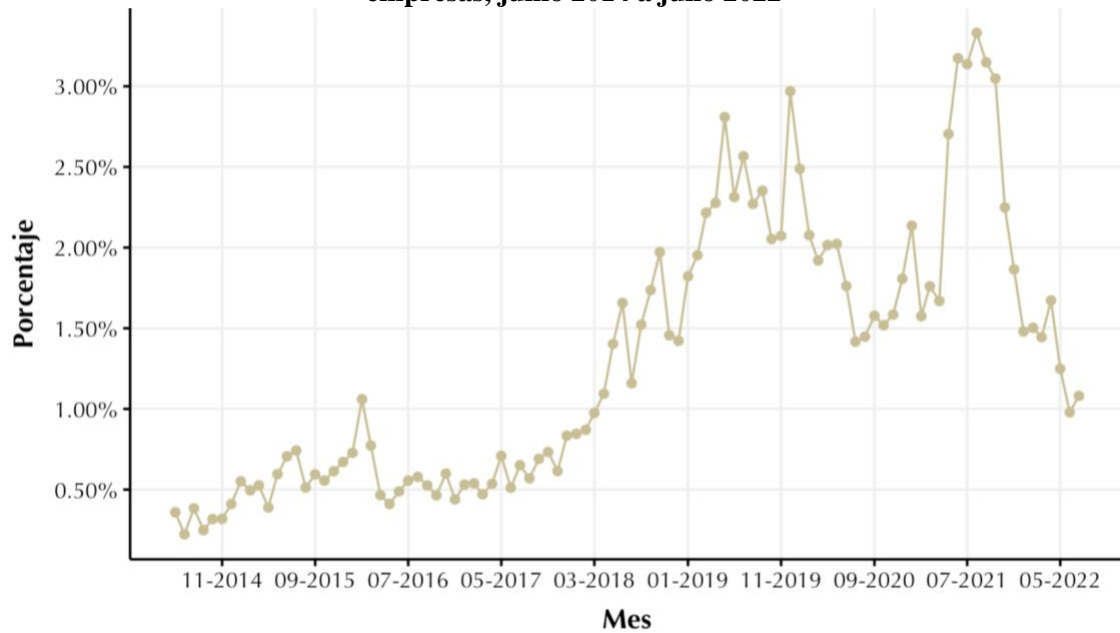
#### 4.1.2.2 Cartera Segmentada

A continuación, se detallan las tasas de incumplimiento de la entidad bancaria segmentado según el tamaño de empresas: empresas medianas y grandes, pequeñas empresas y microempresas:

##### 4.1.2.2.1 Empresas medianas y grandes

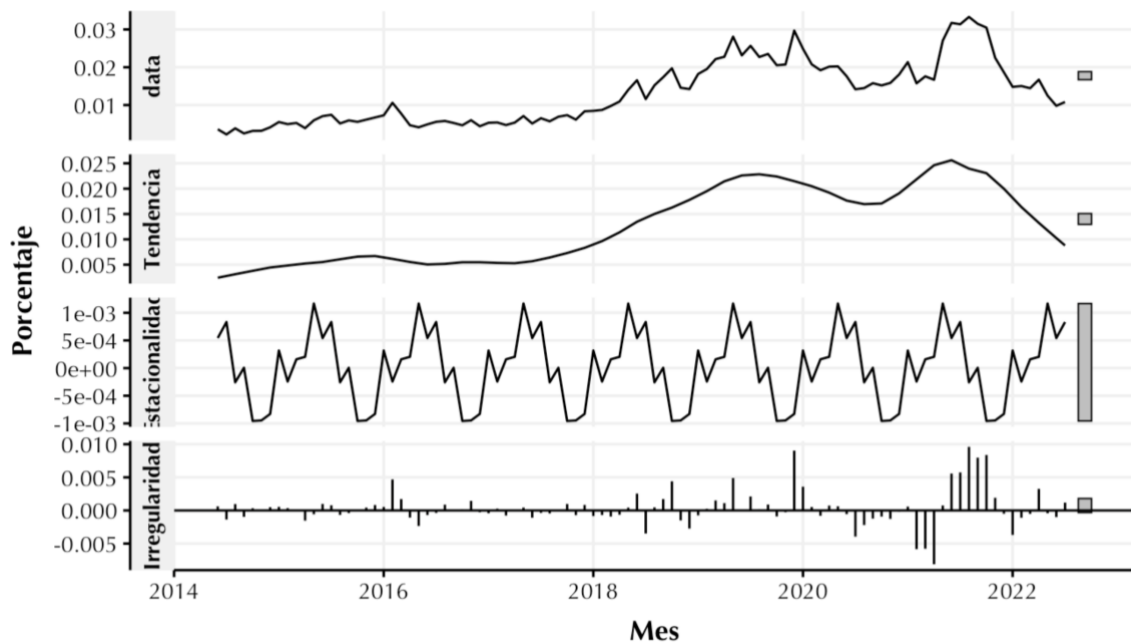
En las empresas corporativas se observa una tendencia ascendente en la tasa de incumplimiento hasta aproximadamente principios de 2019, seguida por una disminución. Esto puede sugerir que, después de cierto punto, factores como el entorno económico o las políticas bancarias han contribuido a una reducción en el incumplimiento (Figura 20). Por otro lado, el componente de estacionalidad parece mostrar patrones de fluctuación anuales. Finalmente, el componente de irregularidad se caracteriza por picos específicos en los meses posteriores de la serie (Figura 21).

**Figura 20. Banco Improsa: Tasa de incumplimiento de las medianas y grandes empresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 21. Banco Improsa: Descomposición de la tasa de incumplimiento de las medianas y grandes empresas, junio 2014 a julio 2022**

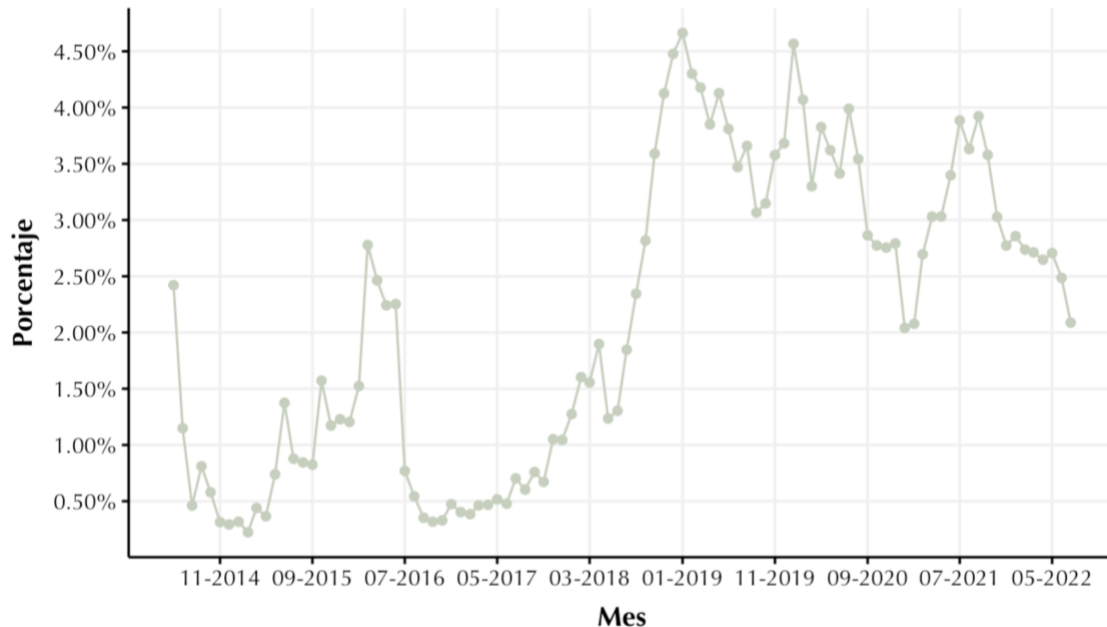


Fuente: Cartera crediticia Banco Improsa

#### 4.1.2.2.2 Empresas pequeñas

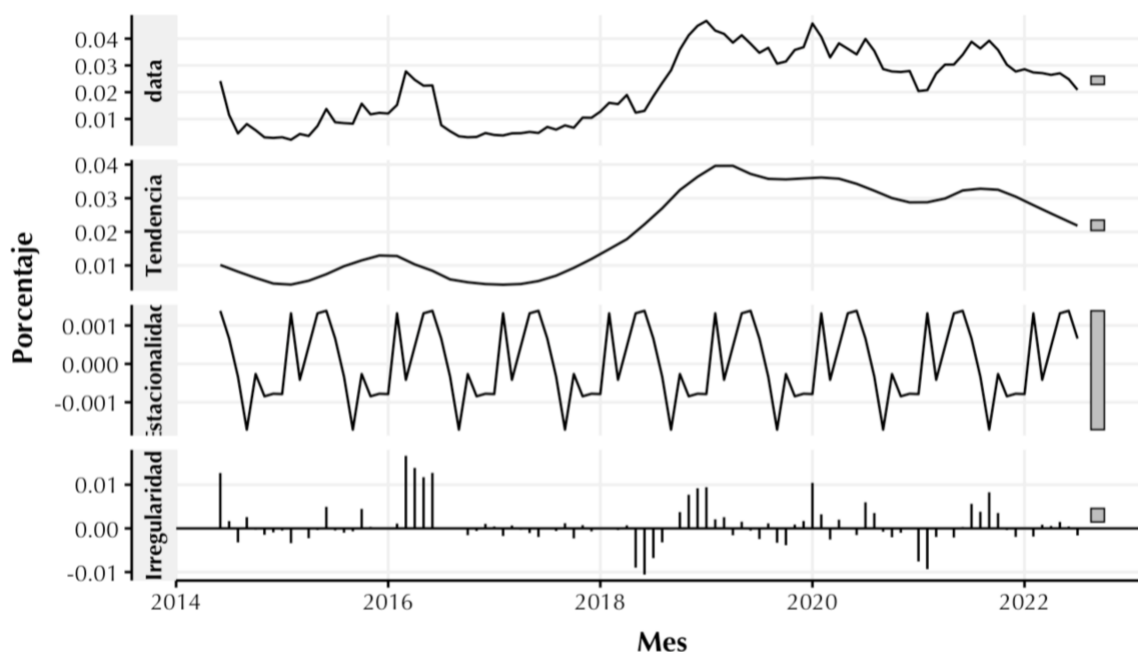
Las Figuras 22 y 23 presentan la tasa de incumplimiento de las pequeñas empresas del Banco Improsa y la descomposición de esta serie temporal, respectivamente, para el periodo comprendido entre junio de 2014 y julio de 2022. Se observa una tendencia general al alza en la tasa de incumplimiento hasta 2019, tras lo cual la tasa exhibe picos pronunciados pero con una tendencia general a la disminución. La estacionalidad muestra fluctuaciones periódicas, aunque su amplitud no es constante, lo que sugiere que factores estacionales podrían influir en la tasa de incumplimiento, pero no son el único determinante. El componente de irregularidad refleja picos y caídas en varios meses a lo largo de la serie, que podrían corresponder a eventos específicos o situaciones imprevistas que afectan a las pequeñas empresas.

**Figura 22. Banco Improsa: Tasa de incumplimiento de las pequeñas empresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 23. Banco Improsa: Descomposición de la tasa de incumplimiento de las pequeñas empresas, junio 2014 a julio 2022**

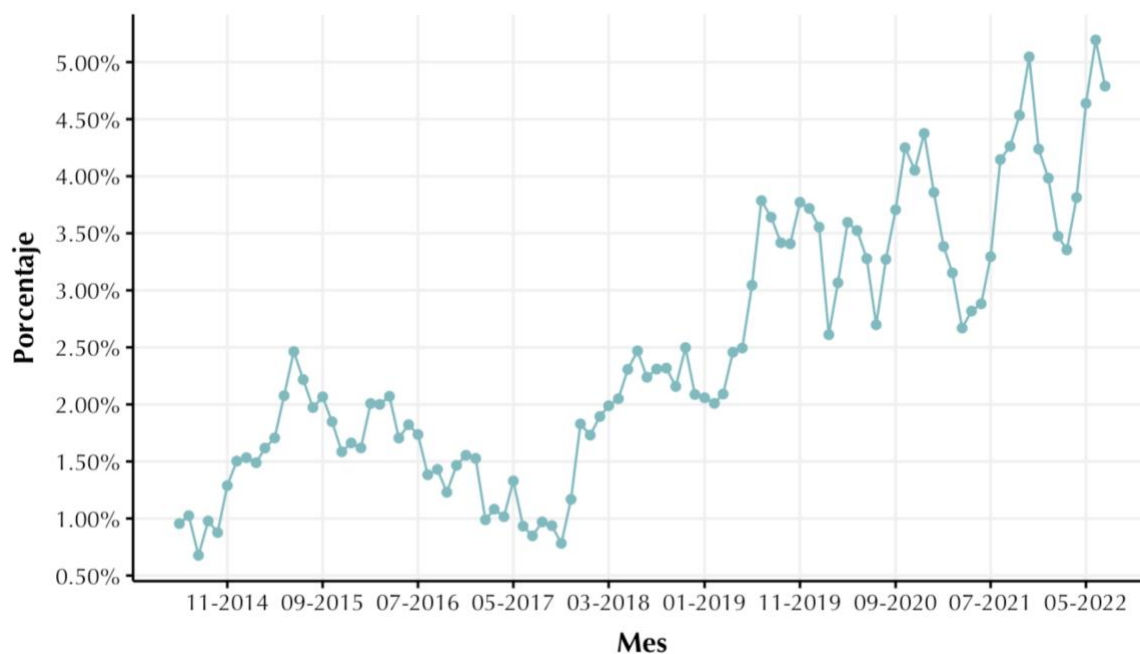


Fuente: Cartera crediticia Banco Improsa

#### 4.1.2.2.3 Microempresas

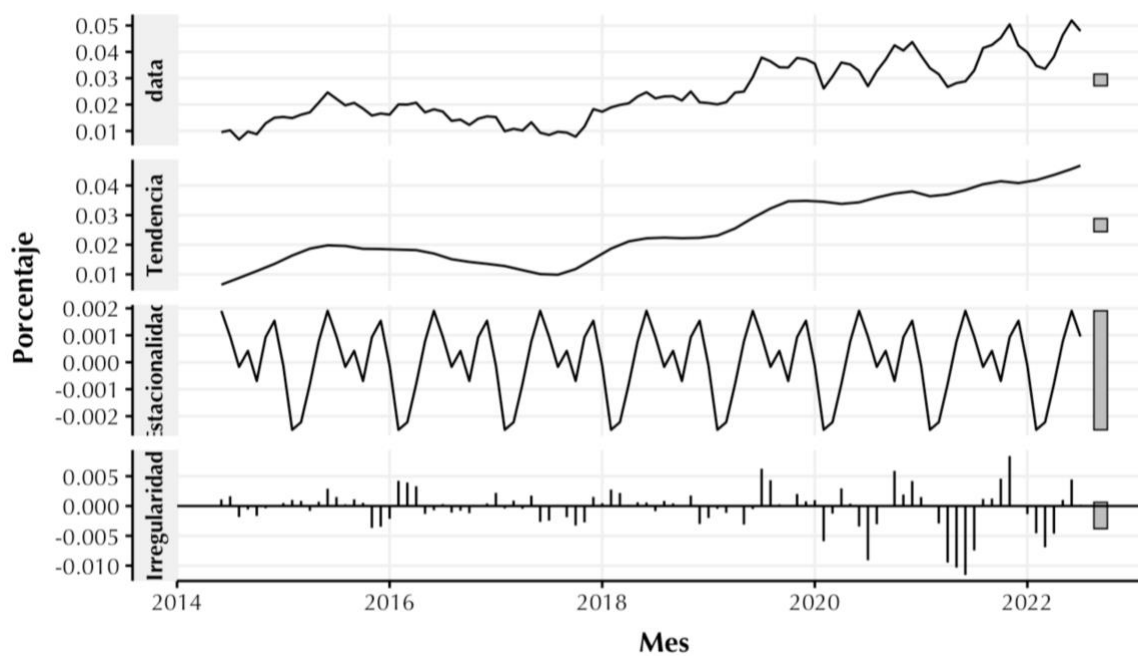
En la Figura 24, se representa la tasa de incumplimiento de las microempresas, que, al igual que las tasas anteriores, presenta una tendencia al alza a lo largo de la serie. Los componentes de esta serie temporal, que se desglosan en la Figura 25, muestran una tendencia con un incremento gradual hasta aproximadamente 2019, seguido de un nivel más estable con una ligera inclinación al alza, lo que sugiere que la tasa general de incumplimiento puede haberse estabilizado. La estacionalidad revela oscilaciones regulares, lo que podría indicar la presencia de factores cíclicos que inciden en la tasa de incumplimiento de las microempresas. Por otro lado, el componente irregular muestra picos significativos en distintos momentos, pero específicamente en los últimos meses de la serie.

**Figura 24. Banco Improsa: Tasa de incumplimiento de las micro empresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

**Figura 25. Banco Improsa: Descomposición de la tasa de incumplimiento de las microempresas, junio 2014 a julio 2022**



Fuente: Cartera crediticia Banco Improsa

#### 4.1.3 Variables Macroeconómicas

En el Cuadro 5 se presentan las estadísticas descriptivas de las variables predictoras incluidas en la investigación. Es importante destacar que estas variables se han calculado mediante variaciones interanuales, tal como se mencionó previamente en el apartado de Metodología. Se aprecia que la variación interanual del IPC presenta una distribución relativamente simétrica de los datos dado que los valores de la media y la mediana son equivalentes (2.0%); además se observa que la mayor de las variaciones del IPC han sido positivos evidenciando el aumento sostenido de inflación del periodo de estudio, lo implica que las personas y/o empresas necesitan más dinero para comprar la misma cantidad de bienes y servicios que antes o para cumplir sus obligaciones financieras.

De igual manera, al analizar la variación interanual del IMAE, se observa que los valores de esta métrica durante el periodo de interés se caracterizan mayoritariamente por aumentos, siendo el 75% de las variaciones mayores al 2.0% (Percentil 25). Además, la distribución de los datos es simétrica, con una mediana y una media ambas situadas en el 3%. En términos bancarios, un crecimiento económico continuo, reflejado en el aumento del IMAE, puede incrementar la capacidad de endeudamiento de las empresas, ya que sus ingresos tienden a aumentar, lo que les permite asumir y gestionar deudas adicionales de manera más efectiva.

En lo que respecta a la variación interanual del tipo de cambio de compra, se observan estadísticas descriptivas similares a las de la variación interanual del IMAE. Esto podría sugerir una relación entre ambas variables. Aunque se identifica una relación lineal positiva entre las variaciones interanuales del tipo de cambio de compra y el IMAE, esta es débil. Esto significa que, en general, cuando la variación interanual del tipo de cambio aumenta, la variación interanual del índice de actividad económica tiende a aumentar también, pero la relación entre estas variables es débil (coeficiente de correlación de Pearson: 0.148).

Por otro lado, al analizar la variación interanual de la tasa básica pasiva, se observa que los valores fluctúan en un rango del -41.0% al 43.0%, con una mediana del -8% y una media ligeramente superior, del -7.0%. Este comportamiento sugiere una alta volatilidad en

esta variable a lo largo del tiempo. Dicha volatilidad puede afectar la rentabilidad de los bancos, específicamente en el valor de los activos y pasivos. Asimismo, puede perjudicar a los prestatarios debido al riesgo de aumentos repentinos en sus pagos de intereses.

En lo que respecta a los costos de vivienda y edificios, la mayoría de los cambios en estas variables se caracterizan por ser incrementaciones, cuyos valores de la mediana y media coinciden en 3.0% y 4.0%, respectivamente. Sin embargo, se observa una discrepancia en el rango de las variables: para el costo de viviendas, el rango es del 15%, mientras que para el costo de edificios es del 18%, indicando una mayor volatilidad en este último.

Por último, la variación interanual del crédito en moneda nacional se caracteriza mayoritariamente por incrementos en el monto de los créditos formalizados, con una mediana y una media de 4.0% y 6.0%, respectivamente. En contraste, la variación interanual del crédito en moneda extranjera presenta tanto incrementos como disminuciones, oscilando entre -12% y 17%, con una mediana y una media de 2.0% y 4.0%, respectivamente. Esta última muestra una mayor volatilidad en comparación con las variaciones del crédito en moneda nacional.

**Cuadro 5. Estadísticas descriptivas de las variables independientes**

	Var. int. IPC	Var. int. IMAE	Var. int. TCC	Var. int. TBP	Var. int. costos vivienda	Var. int. costos edificios	Var. int. crédito MN	Var. int. crédito ME
Mínimo	-1.0%	-7.0%	-7.0%	-41.00%	-2.0%	-1.0%	-3.0%	-12.0%
Percentil 25	1.0%	2.0%	0.0%	-20.00%	1.0%	1.0%	2.0%	-1.0%
Mediana	2.0%	3.0%	4.0%	-8.00%	3.0%	3.0%	4.0%	2.0%
Promedio	2.0%	3.0%	3.0%	-7.00%	4.0%	4.0%	6.0%	4.0%
Percentil 75	3.0%	5.0%	6.0%	6.00%	5.0%	6.0%	11.0%	9.0%
Máximo	11.0%	12.0%	11.0%	43.00%	13.0%	17.0%	16.0%	17.0%

Fuente: BCCR

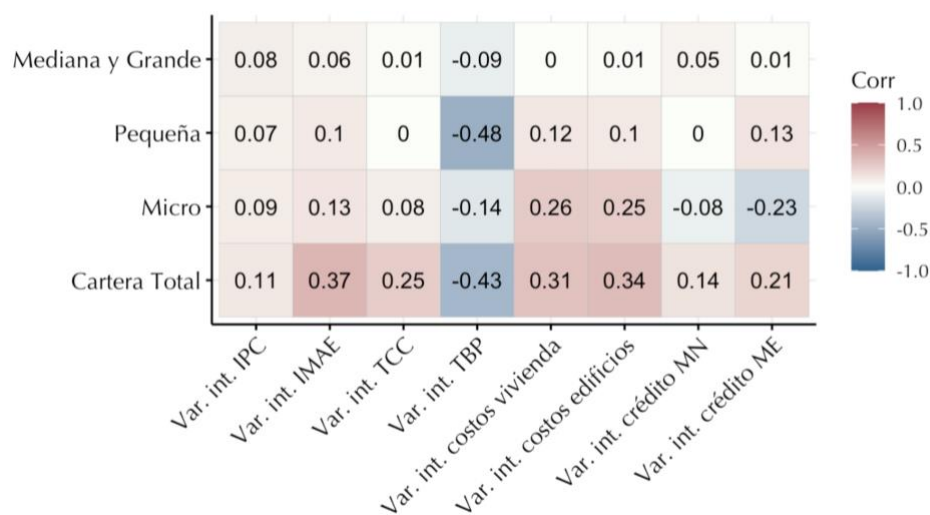
Además de observar el comportamiento propio de las variables macroeconómicas, es igualmente importante analizar cómo se relacionan dichas variables con las variables de respuesta según segmento.

En la Figura 26 se presenta la matriz de correlación entre las tasas de cura y las variables predictoras. La escala de colores del gráfico indica la fuerza de la correlación, con el rojo representando una correlación positiva fuerte (hasta 1.0), el azul una correlación negativa fuerte (hasta -1.0), y el blanco representando una correlación cercana a cero. Entre los resultados más relevantes,

se observa que, en el caso de la totalidad de la cartera, existe una correlación negativa moderada con la variación interanual de la Tasa Básica Pasiva (Var. int. TBP) de -0.43, lo que implica que las tasas de cura podrían disminuir cuando la TBP aumenta. Además, existe una correlación positiva moderada de 0.37 con la variación interanual del IMAE, lo que podría significar que las mejoras en la actividad económica general se asocian con una mayor tasa de cura en la cartera total.

Al segmentar la cartera y considerar las tasas de cura, se observa que en las empresas corporativas no existen correlaciones fuertes, lo que sugiere que las tasas de cura de estas empresas no se ven significativamente afectadas por las variaciones interanuales de las variables independientes proporcionadas. En el caso de las pequeñas empresas, la correlación más notable es con la variación de la Tasa Básica Pasiva (TBP) de -0.48, indicando que un aumento en la TBP está relacionado con una disminución en la proporción de préstamos de empresas pequeñas que salen del incumplimiento. En cuanto a las microempresas, las correlaciones en este segmento son en su mayoría débiles, con la excepción de las variaciones interanuales de los costos de vivienda y edificios (0.26 y 0.25, respectivamente), lo cual podría sugerir que un incremento en los costos de construcción tiene una influencia levemente positiva en la tasa de cura para las microempresas. Además, existe una correlación moderada negativa con la variación interanual del crédito en moneda extranjera, lo que implica que un aumento en los montos de los créditos formalizados en moneda extranjera está relacionado con una disminución en la proporción de préstamos de empresas pequeñas que salen del incumplimiento.

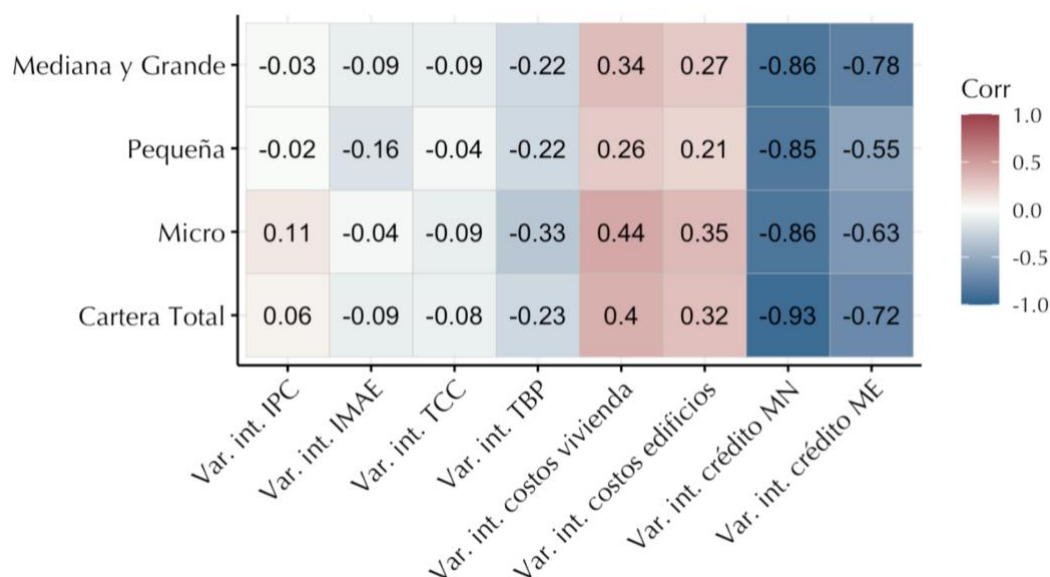
**Figura 26. Matriz de correlaciones entre las tasas de cura según segmento y variables independientes**



Fuente: Cartera crediticia Banco Improsa; BCCR

Con respecto a las tasas de incumplimiento, en la Figura 27 se observa que el comportamiento de las variables macroeconómicas es consistente a lo largo de todos los segmentos, es decir, que el tamaño de la empresa no excluye a que tengan afectaciones por las variaciones macroeconómicas. En general, se observa una correlación negativa moderada con la variación interanual del crédito en moneda extranjera y una correlación negativa fuerte con la variación interanual del crédito en moneda nacional. Esto sugiere que un aumento en la variación del crédito puede estar asociado con una disminución en las tasas de incumplimiento, siendo este efecto más pronunciado en los créditos en colones. Además, existe una correlación positiva moderada con los costos de construcción, lo que podría implicar que en estos costos están asociados con un aumento en las tasas de incumplimiento.

**Figura 27. Matriz de correlaciones entre las tasas de incumplimiento según segmento y variables independientes**



Fuente: Cartera crediticia Banco Improsa; BCCR

## 4.2 Modelado

En esta sección, se exponen los resultados obtenidos del proceso de modelado aplicado a las tasas de cura e incumplimiento observadas en la cartera crediticia de Banco Improsa. Se detallan los hallazgos en cada una de las etapas del modelado, culminando con la presentación de las medidas de precisión de los modelos obtenidos.

#### 4.2.1 Ejemplo proceso de modelado

El análisis de cada uno de los segmentos y tipos de tasas genera una gran cantidad de resultados. Por lo tanto, en esta sección se presentan únicamente los resultados para el caso en el que se estiman las tasas de incumplimiento para la totalidad de la cartera, con el fin de ejemplificar el proceso de modelado que se llevó para cada una de las otras segmentaciones y tipos de tasas. Para consultar en detalle los resultados de estos otros segmentos, se puede referir a la sección de Anexos.

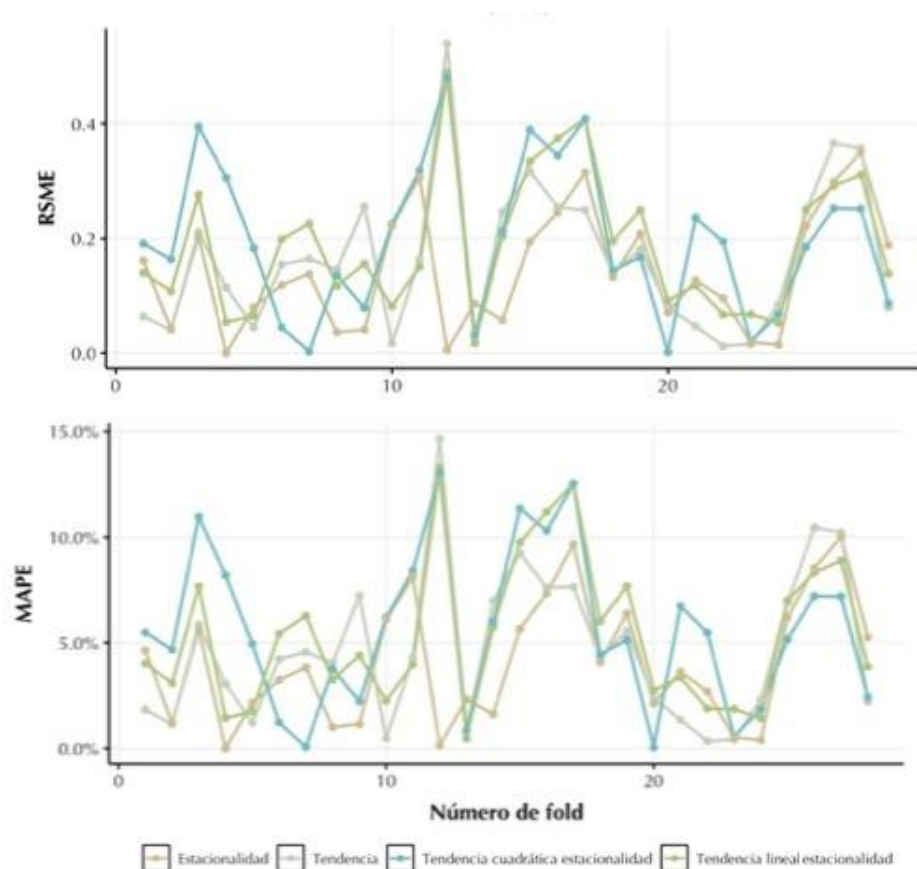
##### 4.2.1.1 Regresión con series temporales

La Figura 28 muestra una representación gráfica del proceso de ajuste de parámetros realizado para cada uno de los tipos de modelos implementados, evidenciando los resultados de una validación cruzada para estimar el error en la predicción de la tasa de incumplimiento mediante un modelo de regresión longitudinal, utilizando el RMSE y el MAPE como métricas de precisión. La implementación de este tipo de visualización facilita la obtención de una visión de la estabilidad del modelo, donde un modelo ideal mostraría una variación mínima en su rendimiento a lo largo de los distintos *folds* de la validación cruzada.

Teniendo esto en cuenta, se aprecia que, para ambas medidas (RMSE y MAPE), existe cierta variabilidad en los errores a través de los diferentes folds. Además, los valores de MAPE fluctúan entre 0% a 15%, lo cual se considera deseable. No obstante visualmente no es fácil determinar un modelo consistentemente mejor que los demás en todos los folds. (Figura 28).

Por lo tanto, en el Cuadro 6 muestra un resumen de las medidas de rendimiento para diferentes variantes de un modelo de regresión longitudinal aplicado a las tasas de cura e incumplimiento de la cartera total del Banco Improsa, donde los modelos resaltados en negrita corresponden a los seleccionados porque presentan un balance entre una mediana del error baja y una desviación estándar que no es excesivamente alta, lo que sugiere que son relativamente precisos y consistentes en sus predicciones. En resumen, los modelos seleccionados son: la regresión de series de tiempo considerando el componente de tendencia para la tasa de cura; y el componente de estacionalidad para la tasa de incumplimiento.

**Figura 28. Banco Improsa, Cartera Total: Estimación del error mediante regresión longitudinal en validación cruzada de la tasa de incumplimiento según medida de precisión (k=29)**



Fuente: Cartera crediticia Banco Improsa

**Cuadro 6. Cartera Total: Resumen de medidas de rendimiento de las variantes del modelo de regresión longitudinal**

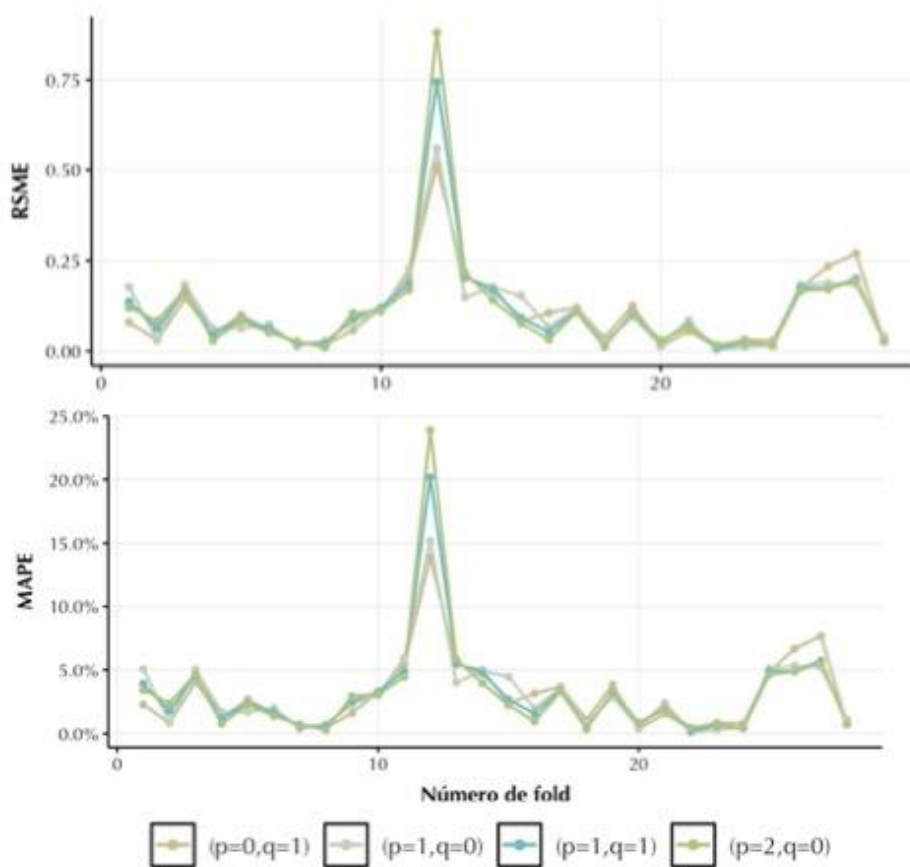
Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Tendencia</b>	<b>0.2710</b>	<b>0.6237</b>	<b>14.410%</b>	<b>32.78%</b>
Cura	Estacionalidad	0.2943	0.7116	16.080%	37.33%
Cura	Tendencia lineal estacionalidad	0.3397	0.7295	17.970%	38.38%
Cura	Tendencia cuadrática estacionalidad	0.3953	0.7349	18.960%	38.57%
<b>Incumplimineto</b>	<b>Estacionalidad</b>	<b>0.1294</b>	<b>0.1028</b>	<b>3.730%</b>	<b>2.97%</b>
Incumplimineto	Tendencia	0.1501	0.1286	4.260%	3.63%
Incumplimineto	Tendencia lineal estacionalidad	0.1530	0.1207	4.210%	3.51%
Incumplimineto	Tendencia cuadrática estacionalidad	0.1877	0.1316	5.330%	3.75%

Fuente: Cartera crediticia Banco Improsa

#### 4.2.1.2 Modelo Autorregresivo de Promedio Móvil

En relación con los modelos ARMA, se modelaron 36 permutaciones para cada segmentación y tipo de tasa. No obstante, las visualizaciones subsiguientes se han filtrado para mostrar únicamente los cuatro modelos con el mejor rendimiento. En Figura 29 se aprecia que los modelos ARMA  $(p=0, q=1)$  y  $(p=1, q=0)$  parecen presentar una mayor consistencia en su desempeño, mostrando un error de predicción más bajo y estable a través de los diferentes pliegues, en comparación con los modelos  $(p=1, q=1)$  y  $(p=2, q=0)$ , donde se observan picos más pronunciados, indicando una variabilidad mayor en el error de predicción.

**Figura 29. Banco Improsa, Cartera Total: Estimación del error mediante modelo ARMA en validación cruzada de la tasa de incumplimiento según medida de precisión  $(k=29)$**



Fuente: Cartera crediticia Banco Improsa

En el Cuadro 7, se observa que para la tasa de incumplimiento, el modelo ARMA ( $p=0, q=1$ ) tiene la mediana del RSME y MAPE más baja con valores de 0.0804 y 2.330% respectivamente, junto con las desviaciones estándar más bajas para estas métricas. Esto sugiere un alto grado de precisión y consistencia en las predicciones de este modelo, condiciones preferibles ya que indican una mayor robustez del modelo. Además, se aprecia que para la tasa de cura el modelo ARMA ( $p=1, q=0$ ) tiene la mediana de MAPE más baja con un 8.730%, lo que sugiere que este modelo tiene el mejor desempeño en términos del error entre los modelos considerados. Sin embargo, también presenta una mayor desviación estándar en el MAPE (29.100%), lo que indica una cierta variabilidad en el rendimiento de este modelo entre los diferentes conjuntos de datos utilizados en la validación cruzada.

**Cuadro 7. Resumen de medidas de rendimiento de las variantes del modelo ARMA**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(<math>p=1, q=0</math>)</b>	<b>0.1576</b>	<b>0.5463</b>	<b>8.730%</b>	<b>29.100%</b>
Cura	( $p=5, q=4$ )	0.2018	0.5113	10.790%	27.050%
Cura	( $p=0, q=3$ )	0.2090	0.5495	10.340%	29.120%
Cura	( $p=5, q=3$ )	0.2328	0.3891	11.740%	20.600%
<b>Incumplimeto</b>	<b>(<math>p=0, q=1</math>)</b>	<b>0.0804</b>	<b>0.1074</b>	<b>2.330%</b>	<b>2.950%</b>
Incumplimeto	( $p=1, q=0$ )	0.0827	0.1108	2.350%	3.020%
Incumplimeto	( $p=2, q=0$ )	0.0837	0.1622	2.330%	4.400%
Incumplimeto	( $p=1, q=1$ )	0.0888	0.1387	2.460%	3.770%

Fuente: Cartera crediticia Banco Improsa

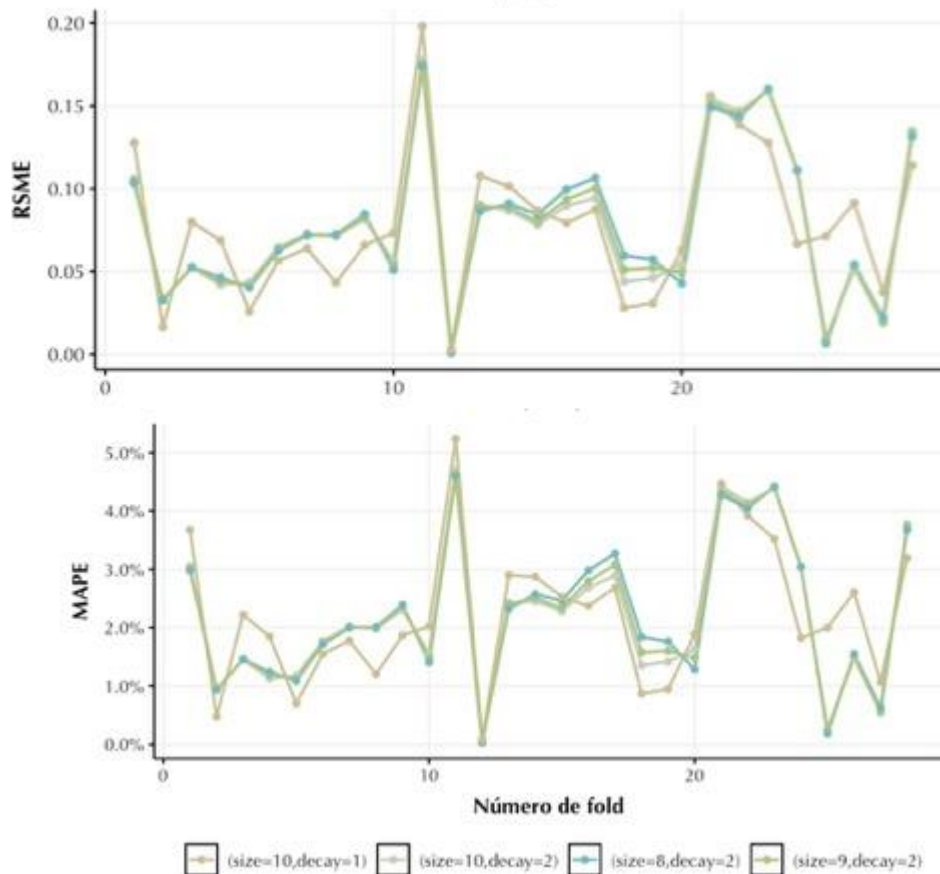
En comparación con los modelos de regresión basados en series temporales, se observa que los modelos ARMA presentan, en términos generales, un mejor rendimiento en cuanto a las medidas de error. Estos modelos obtienen resultados preferibles en ambos tipos de tasa, con valores de MAPE inferiores al 10%.

#### 4.2.1.3 Redes Neuronales Autorregresivas

En la Figura 30, las líneas de cada gráfico representan variantes del modelo NNAR con distintos tamaños del número de nodos en la capa oculta y parámetros de regularización de los pesos (decay). Se observa que en ambos gráficos, las variantes del modelo con ( $size=10, decay=1$ ) y ( $size=10, decay=2$ ) exhiben un comportamiento similar entre sí en términos de RMSE y MAPE a lo largo de los diferentes folds. Los modelos con ( $size=8, decay=2$ ) y ( $size=9, decay=2$ ) también presentan un comportamiento similar entre sí, aunque

con ligeras diferencias en comparación con los modelos de tamaño 10. Además, se aprecia cierta estabilidad en la variabilidad de los errores a través de los distintos folds, y los valores de MAPE fluctúan entre 0% y 5%, lo que se considera muy deseable en el análisis de series temporales.

**Figura 30. Banco Improsa, Cartera Total: Estimación del error mediante modelo NNAR en validación cruzada de la tasa de incumplimiento**



Fuente: Cartera crediticia Banco Improsa

De igual manera, se evidencia en el Cuadro 8 que, para la tasa de incumplimiento, el modelo con (size=10, decay=2) presenta la mediana de RMSE más baja (0.0720) y también el MAPE más bajo (2.000%), con las desviaciones estándar más bajas en ambas métricas (0.0457 para RMSE y 1.260% para MAPE). Esto indica un rendimiento consistente y con una alta precisión. Para la tasa de cura, el modelo con (size=4, decay=0.1) presenta la mediana de RMSE más baja (0.1122) y una desviación estándar de RMSE de 0.1478, que es

comparativamente más baja que las de otras variantes, lo que implica que sus errores son relativamente consistentes.

Cabe destacar que, al comparar estos resultados con modelos anteriores, se observa una mejora evidente en el rendimiento y, específicamente, en la estabilidad de los modelos en términos de la variabilidad del error, tanto para la tasa de cura como para el incumplimiento. En estos modelos, los valores de MAPE son inferiores al 7%.

**Cuadro 8. Resumen de medidas de rendimiento de las variantes del modelo NNAR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(size=4,decay=0.1)</b>	<b>0.1122</b>	<b>0.1478</b>	<b>6.3700%</b>	<b>8.950%</b>
Cura	(size=3,decay=0.1)	0.1228	0.1558	6.8700%	9.470%
Cura	(size=2,decay=0.1)	0.1259	0.1821	6.7500%	10.410%
Cura	(size=5,decay=0.1)	0.1538	0.1503	8.4200%	9.470%
<b>Incumplimineto</b>	<b>(size=10,decay=2)</b>	<b>0.0720</b>	<b>0.0457</b>	<b>2.0000%</b>	<b>1.260%</b>
Incumplimineto	(size=8,decay=2)	0.0721	0.0448	2.0000%	1.240%
Incumplimineto	(size=9,decay=2)	0.0722	0.0450	2.0100%	1.240%
Incumplimineto	(size=10,decay=1)	0.0723	0.0443	2.0100%	1.220%

Fuente: Cartera crediticia Banco Improsa

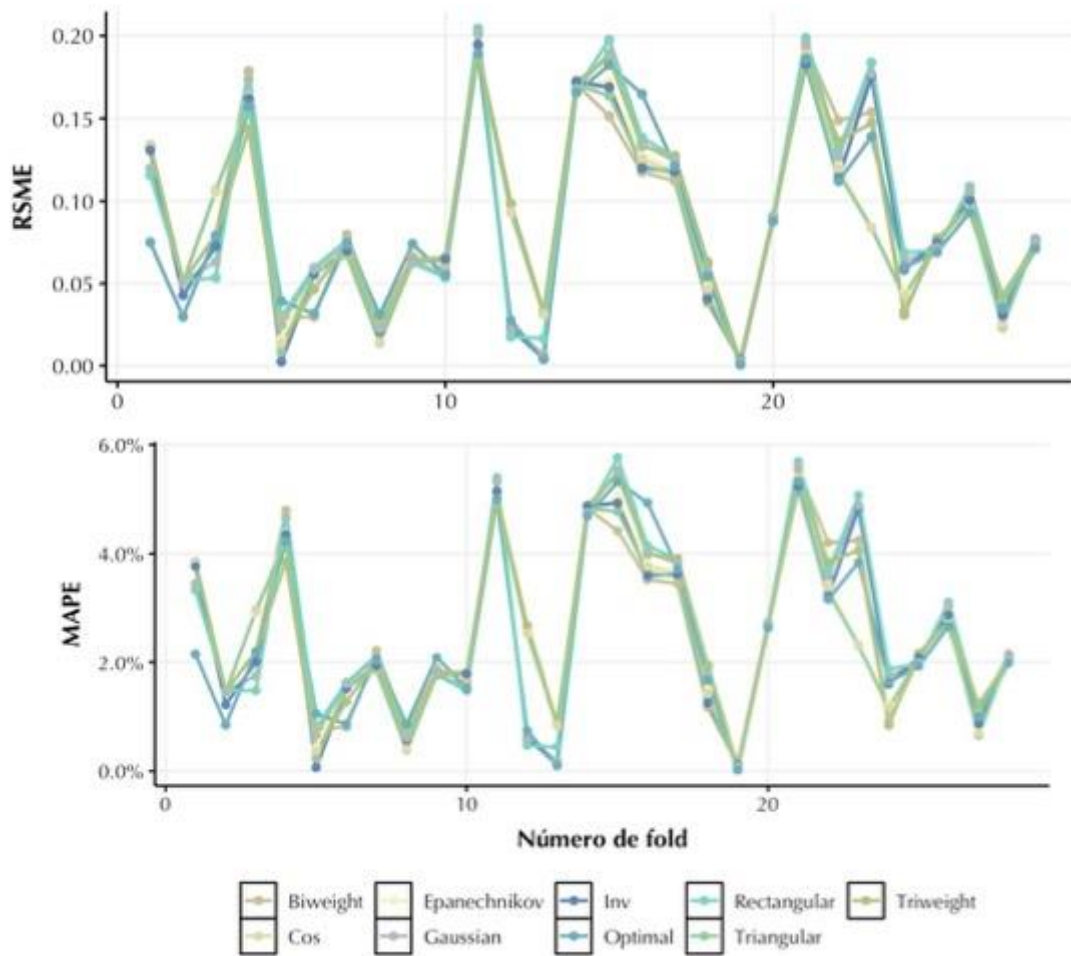
#### 4.2.1.4 Método de k vecinos más cercanos

En la Figura 31, se presenta la estimación del error del modelo KNN en la validación cruzada de la tasa de incumplimiento, donde cada línea representa una función de ponderación diferente utilizada en el modelo KNN. Se observa que a lo largo de los distintos folds, los errores mantienen una estabilidad, sin destacar picos donde el error aumente significativamente. Cabe resaltar que los valores de MAPE fluctúan entre 0% y 6%, lo cual se considera un resultado muy deseable.

En el Cuadro 9, se muestra que para el modelo que estima la tasa de incumplimiento, la función de ponderación gaussiana presenta la mediana de RMSE más baja (0.0727) y la mediana de MAPE más baja (2.030%), junto con las desviaciones estándar más bajas en ambas métricas (0.0603 para RMSE y 1.700% para MAPE). Esto implica un alto grado de precisión y consistencia en las predicciones de la tasa de incumplimiento. En contraste, para la tasa de cura, la función de ponderación inversa proporciona la mediana más baja en ambas métricas (0.1079 para RMSE y 4.760% para MAPE), aunque con una mayor desviación

estándar, lo que sugiere un buen ajuste general pero con cierta variabilidad entre los diferentes folds.

**Figura 31. Banco Improsa, Cartera Total: Estimación del error mediante modelo KNN en validación cruzada de la tasa de incumplimiento**



Fuente: Cartera crediticia Banco Improsa

**Cuadro 9. Resumen de medidas de rendimiento de las variantes del modelo KNN**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Inv</b>	<b>0.1079</b>	<b>0.1740</b>	<b>4.760%</b>	<b>9.830%</b>
Cura	Triweight	0.1082	0.1505	5.500%	8.570%
Cura	Rectangular	0.1084	0.1900	5.180%	10.760%
Cura	Triangular	0.1120	0.1398	5.550%	8.320%
Cura	Cos	0.1143	0.1431	5.890%	8.520%
Cura	Epanechnikov	0.1145	0.1447	6.270%	8.620%
Cura	Biweight	0.1173	0.1413	5.140%	8.320%
Cura	Gaussian	0.1186	0.2149	5.910%	13.470%
Cura	Optimal	0.1305	0.1600	6.310%	9.600%
<b>Incumplimineto</b>	<b>Gaussian</b>	<b>0.0727</b>	<b>0.0603</b>	<b>2.030%</b>	<b>1.700%</b>
Incumplimineto	Rectangular	0.0727	0.0614	2.030%	1.730%
Incumplimineto	Inv	0.0736	0.0588	2.050%	1.650%
Incumplimineto	Optimal	0.0743	0.0569	2.080%	1.610%
Incumplimineto	Triweight	0.0784	0.0552	2.190%	1.570%
Incumplimineto	Biweight	0.0794	0.0561	2.210%	1.570%
Incumplimineto	Cos	0.0801	0.0540	2.230%	1.520%
Incumplimineto	Epanechnikov	0.0802	0.0543	2.230%	1.530%
Incumplimineto	Triangular	0.0802	0.0535	2.230%	1.510%

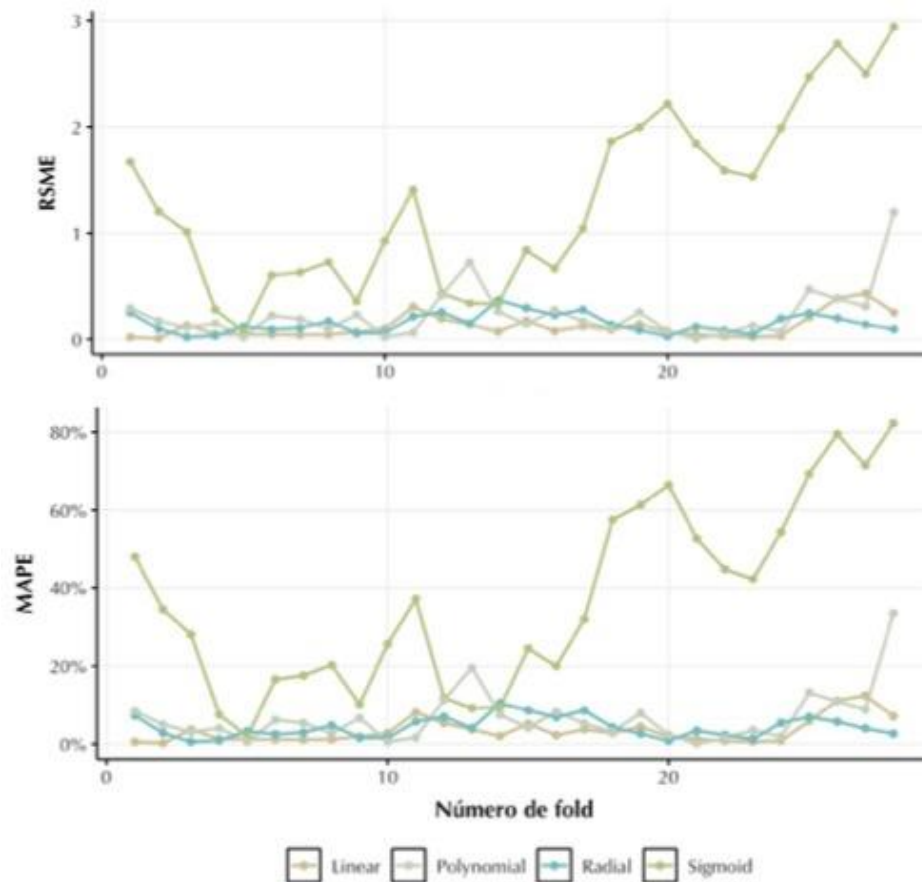
Fuente: Cartera crediticia Banco Improsa

Al comparar estos resultados con los modelos enfocados en series temporales, se observa que se logra un rendimiento similar. Incluso en la tasa de cura, se obtiene una magnitud de la mediana del error menor y un grado de variabilidad similar. En el caso de la tasa de incumplimiento, las medidas de error presentan valores semejantes.

#### 4.2.1.5 Máquinas de soporte vectorial

La Figura 32 muestra el desempeño de un modelo SVR utilizando diferentes tipos de kernel en la validación cruzada de la tasa de incumplimiento. Se observa que los *kernels* lineal, radial y polinomial tienen un desempeño más consistente y errores más bajos a través de los diferentes folds, comparados con el *kernel* sigmoidal, que presenta mayor variabilidad y picos más altos en varios puntos, indicando mayores errores de predicción en esas repeticiones específicas.

**Figura 32. Banco Improsa, Cartera Total: Estimación del error mediante modelo SVR en validación cruzada de la tasa de incumplimiento**



Fuente: Cartera crediticia Banco Improsa

En el Cuadro 10, se observa que para la tasa de incumplimiento, el modelo con la función kernel lineal presenta la mediana del RMSE y MAPE más baja, con valores de 0.0725 y 2.11% respectivamente, y también tiene las desviaciones estándar más bajas para estas métricas. Además, se aprecia que para la tasa de cura, el modelo con el kernel radial tiene la mediana de MAPE más baja, con un 7.43%, y también presenta la menor desviación estándar en el MAPE (12.39%). En general, los modelos SVR presentan valores de las medidas de rendimiento favorables e incluso mejores que los modelos ARMA y de regresión longitudinal, aunque no mejores que los modelos NNAR y KNN.

**Cuadro 10. Resumen de medidas de rendimiento de las variantes del modelo SVR**

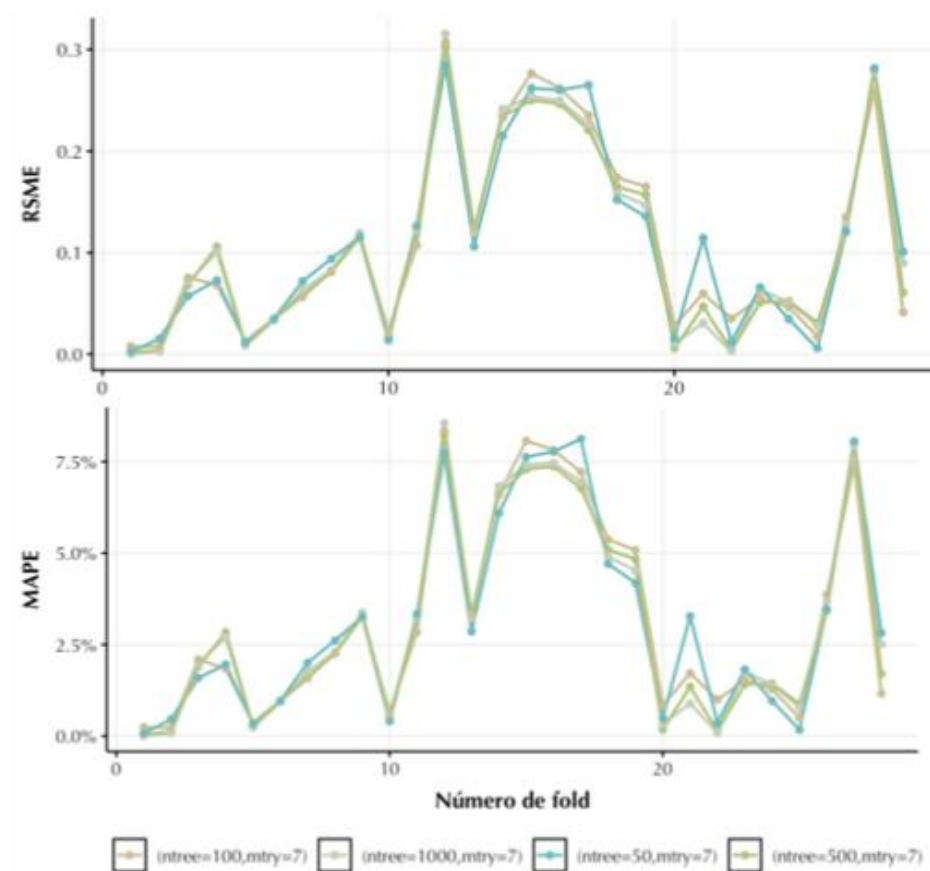
Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Radial</b>	<b>0.1587</b>	<b>0.2236</b>	<b>7.43%</b>	<b>12.39%</b>
Cura	Linear	0.3441	0.3365	18.68%	16.71%
Cura	Polynomial	0.4027	0.3773	18.23%	20.69%
Cura	Sigmoid	1.1043	0.6168	59.57%	31.97%
<b>Incumplimineto</b>	<b>Linear</b>	<b>0.0725</b>	<b>0.1120</b>	<b>2.11%</b>	<b>3.16%</b>
Incumplimineto	Radial	0.1289	0.0925	3.68%	2.66%
Incumplimineto	Polynomial	0.1716	0.2468	5.09%	6.85%
Incumplimineto	Sigmoid	1.1248	0.8298	33.29%	23.94%

Fuente: Cartera crediticia Banco Improsa

#### 4.2.1.6 Bosques Aleatorios

Con respecto a los bosques aleatorios, se modelaron 45 combinaciones para cada segmentación y tipo de tasa, y se han filtrado los cuatro mejores modelos según su rendimiento. En la Figura 33, las líneas de cada gráfico representan variantes del modelo de bosques aleatorios con distintos números de árboles de regresión y cantidades de variables seleccionadas de manera aleatoria de las ocho posibles. Se observa que el modelo con 50 árboles y 7 variables seleccionadas presenta picos más pronunciados, indicando una mayor variabilidad en el error de predicción. No obstante, en la figura no se logra distinguir de manera visual un modelo que sobresalga en su rendimiento entre las distintas variantes.

**Figura 33. Banco Improsa, Cartera Total: Estimación del error mediante bosques aleatorios en validación cruzada de la tasa de incumplimiento**



Fuente: Cartera crediticia Banco Improsa

Por lo tanto, en el Cuadro 11 se observa que, para la tasa de incumplimiento, el modelo con 100 árboles de regresión y 7 variables seleccionadas en cada nodo presenta la mediana del RMSE más baja (0.0721) y también el MAPE más bajo (1.97%), con desviaciones estándar de 0.0457 para RMSE y 1.260% para MAPE, que son magnitudes bajas, aunque no las más bajas. Para la tasa de cura, el mejor modelo corresponde al que contiene 100 árboles de regresión y 3 variables seleccionadas en cada nodo, donde la mediana de RMSE es igual a 0.2566 y la desviación estándar del RMSE es de 0.2304. En términos generales, para la tasa de incumplimiento en la totalidad de la cartera, los bosques aleatorios parecen presentar una mejora en el rendimiento en relación con otros tipos de modelos al analizar la mediana de las medidas de error, pero no en términos de la desviación estándar.

**Cuadro 11. Resumen de medidas de rendimiento de las variantes de los bosques aleatorios**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(ntree=100,mtry=3)</b>	<b>0.2566</b>	<b>0.2304</b>	<b>13.83%</b>	<b>14.040%</b>
Cura	(ntree=100,mtry=2)	0.2783	0.2108	14.09%	13.420%
Cura	(ntree=250,mtry=2)	0.2794	0.2366	12.41%	15.210%
Cura	(ntree=50,mtry=2)	0.3068	0.2344	14.90%	15.030%
<b>Incumplimineto</b>	<b>(ntree=100,mtry=7)</b>	<b>0.0721</b>	<b>0.0933</b>	<b>1.97%</b>	<b>2.720%</b>
Incumplimineto	(ntree=500,mtry=7)	0.0749	0.0925	2.08%	2.680%
Incumplimineto	(ntree=1000,mtry=7)	0.0864	0.0938	2.41%	2.720%
Incumplimineto	(ntree=50,mtry=7)	0.0974	0.0924	2.72%	2.700%

Fuente: Cartera crediticia Banco Improsa

#### 4.2.2 Resultados Generales

En el Cuadro 12 se presentan las distintas variables macroeconómicas consideradas en esta investigación y sus respectivas abreviaturas utilizadas en las figuras subsiguientes. Esto se hace con el fin de proporcionar un entendimiento más claro de los resultados.

**Cuadro 12. Listado de variables macroeconómicas y sus respectivas abreviaturas**

Abreviatura	Variable macroeconómica
var_inflacion	Variación interanual del IPC
var_imae	Variación interanual del IMAE
var_tcc	Variación interanual del TCC (Depresión del colón)
var_tbp	Variación interanual de la TBP
var_vivienda	Variación interanual de los costos de vivienda
var_edificio	Variación interanual de los costos de edificio
var_credito.mn	Variación interanual del crédito en moneda nacional (MN)
var_credito.me	Variación interanual del crédito en moneda extranjera (ME)

Fuente: Elaboración propia.

Dejando esto claro, se aprecia en el Cuadro 13 que para ambos tipos de tasa, los modelos con mejor precisión según el MAPE y el RSME son la red neuronal autorregresiva y el modelo KNN. Es interesante notar que el modelo con el enfoque de series temporales por redes neuronales autorregresivas se comporta consistentemente bien en ambos tipos de tasas.

Por otro lado, se observa que para ambas tasas se logran alcanzar modelos con menos variables, pero con mayor rendimiento. Se destaca que el comportamiento total de la cartera

en el pago de los créditos está relacionado principalmente con las variaciones en la inflación, la actividad económica y el crédito del sistema bancario en moneda extranjera.

**Cuadro 13. Cartera Total: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento**

Tasa	Modelo	Parámetros	Variables	RSME	MAPE
<b>Cura</b>	<b>RED</b>	<b>(size=4,decay=0.1)</b>	<b>var_inflacion, var_imae, var_tcc, var_credito.me, var.intervencion</b>	<b>0.1427</b>	<b>7.70%</b>
Cura	KNN	Inv	var_tcc, var_vivienda, var_edificio, var_credito.me, var.intervencion	0.1466	8.00%
Cura	SVR	Radial	var_tcc, var_vivienda, var.intervencion	0.1517	8.40%
Cura	ARMA	(p=1,q=0)	var_credito.me, var.intervencion	0.1603	8.70%
Cura	Regresión Longitudinal	Tendencia	var_imae, var_credito.me, var.intervencion	0.2367	12.40%
Cura	Bosques Aleatorios	(ntree=250,mtry=2)	var_imae, var_tcc, var_tbp, var_edificio, var.intervencion	0.2324	12.70%
<b>Incumplimiento</b>	<b>RED</b>	<b>(size=10,decay=2)</b>	<b>var_inflacion, var_imae, var_tbp, var_vivienda, var_credito.me, var.intervencion</b>	<b>0.0699</b>	<b>2.00%</b>
Incumplimiento	KNN	Gaussian	var_inflacion, var_imae, var_tcc, var_edificio, var.intervencion	0.0710	2.00%
Incumplimiento	ARMA	(p=0,q=1)	var_tcc, var_tbp, var_vivienda, var_credito.me, var.intervencion	0.0796	2.30%
Incumplimiento	Bosques Aleatorios	(ntree=100,mtry=7)	var_tcc, var_tbp, var_edificio, var_credito.mn, var_credito.me, var.intervencion	0.0850	2.40%
Incumplimiento	SVR	Linear	var_inflacion, var_edificio, var_credito.mn, var.intervencion	0.1106	3.10%
Incumplimiento	Regresión Longitudinal	Estacionalidad	var_inflacion, var_vivienda, var_edificio, var_credito.mn, var.intervencion	0.1259	3.60%

Fuente: Cartera crediticia Banco Improsa

En el Cuadro 14, se proporciona una evaluación comparativa de los modelos para predecir las tasas de cura e incumplimiento en empresas corporativas. Los modelos con el rendimiento más destacado están resaltados en negrita: el modelo K-Nearest Neighbors para la tasa de cura con un RMSE de 0.1596 y un MAPE del 7.3%; y para la tasa de incumplimiento, el modelo ARMA con un RMSE de 0.1246 y un MAPE del 3.2%. Se observa que varios factores macroeconómicos y de mercado específicos están relacionados con la salud financiera de las medianas y grandes empresas en su comportamiento crediticio, como la inflación y la tasa básica pasiva en ambos tipos de tasas.

**Cuadro 14. Medidas y Grandes empresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento**

Tasa	Modelo	Parámetros	Variables	RSME	MAPE
<b>Cura</b>	<b>KNN</b>	<b>Inv</b>	<b>var_inflacion, var_imae, var_tcc, var_tbp, var_credito.mn, var.intervencion</b>	<b>0.1596</b>	<b>7.300%</b>
Cura	SVR	Radial	var_vivienda, var_credito.me, var.intervencion	0.1651	7.400%
Cura	Regresión Longitudinal	Tendencia	var_imae, var_credito.mn, var.intervencion	0.2316	10.700%
Cura	RED	(size=1,decay=0.05)	var_tcc, var.intervencion	0.2959	13.100%
Cura	ARMA	(p=0,q=1)	var_imae, var_credito.mn, var.intervencion	0.3206	14.500%
Cura	Bosques Aleatorios	(ntree=50,mtry=2)	var_imae, var_edificio, var.intervencion	0.4671	22.400%
<b>Incumplimiento</b>	<b>ARMA</b>	<b>(p=4,q=1)</b>	<b>var_inflacion, var_tbp, var_vivienda, var_credito.me, var.intervencion</b>	<b>0.1246</b>	<b>3.200%</b>
Incumplimiento	KNN	Optimal	var_inflacion, var_tbp, var_credito.me, var.intervencion	0.1373	3.500%
Incumplimiento	RED	(size=7,decay=0.1)	var_inflacion, var_imae, var_tcc, var_vivienda, var_credito.me, var.intervencion	0.1487	3.700%
Incumplimiento	Bosques Aleatorios	(ntree=100,mtry=3)	var_inflacion, var_tbp, var_edificio, var_credito.mn, var.intervencion	0.1458	3.700%
Incumplimiento	Regresión Longitudinal	Tendencia	var_inflacion, var_imae, var_tbp, var_edificio, var_credito.mn, var_credito.me, var.intervencion	0.1510	3.800%
Incumplimiento	SVR	Linear	var_inflacion, var_edificio, var_credito.mn, var_credito.me, var.intervencion	0.1542	3.900%

Fuente: Cartera crediticia Banco Improsa

En el caso de las pequeñas empresas, el modelo KNN presenta un RMSE de 0.2136 y un MAPE de 11.60%, lo que indica un buen rendimiento dentro de los modelos presentados para predecir la tasa de cura de créditos. Este modelo utiliza las variables de la variación interanual del tipo de cambio de compra y la variación interanual del crédito en moneda nacional. Mientras tanto, para la tasa de incumplimiento de créditos en pequeñas empresas, el modelo de red neuronal autorregresiva con nueve nodos ocultos y un factor de decaimiento de 2 presenta un RMSE de 0.0880 y un MAPE de 2.50%, demostrando ser más preciso en predecir los incumplimientos de créditos. Este modelo incluye la variación interanual de la inflación, del tipo de cambio de compra, los costos de edificios y la variación interanual del crédito en ambas monedas (Cuadro 15).

**Cuadro 15. Pequeñas empresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento**

Tasa	Modelo	Parámetros	Variables	RSME	MAPE
<b>Cura</b>	<b>KNN</b>	<b>Inv</b>	<b>var_tcc, var_credito.me, var.intervencion</b>	<b>0.2136</b>	<b>11.60%</b>
Cura	SVR	Radial	var_imae, var_vivienda, var_credito.mn, var.intervencion	0.2344	12.80%
Cura	RED	(size=10,decay=1)	var_credito.mn, var.intervencion	0.2408	13.30%
Cura	ARMA	(p=0,q=1)	var_credito.mn, var.intervencion	0.3025	15.90%
Cura	Bosques Aleatorios	(ntree=250,mtry=3)	var_imae, var.intervencion	0.5070	24.10%
Cura	Regresión Longitudinal	Tendencia	var_imae, var_credito.mn, var.intervencion	0.8562	45.30%
<b>Incumplimiento</b>	<b>RED</b>	<b>(size=9,decay=2)</b>	<b>var_inflacion, var_tcc, var_edificio, var_credito.mn, var_credito.me, var.intervencion</b>	<b>0.0880</b>	<b>2.50%</b>
Incumplimiento	KNN	Rectangular	var_inflacion, var_tbp, var_vivienda, var_credito.me, var.intervencion	0.0890	2.50%
Incumplimiento	Bosques Aleatorios	(ntree=250,mtry=4)	var_inflacion, var_tcc, var_edificio, var_credito.mn, var.intervencion	0.1106	3.10%
Incumplimiento	ARMA	(p=5,q=2)	var_inflacion, var_vivienda, var_credito.mn, var.intervencion	0.1117	3.20%
Incumplimiento	SVR	Linear	var_inflacion, var_vivienda, var_credito.mn, var.intervencion	0.1354	3.90%
Incumplimiento	Regresión Longitudinal	Tendencia	var_inflacion, var_vivienda, var_credito.mn, var.intervencion	0.1439	4.20%

Fuente: Cartera crediticia Banco Improsa

En el Cuadro 16, se observa que los modelos destacados en negrita para las microempresas corresponden al modelo KNN para la tasa de cura y al modelo ARMA para la tasa de incumplimiento. El modelo KNN presenta un RMSE de 0.2785 y un MAPE de 18.00%, utilizando variables como la variación interanual del tipo de cambio, la tasa básica pasiva, la variación interanual de los costos de vivienda, y la variación interanual del crédito en moneda nacional. Para la tasa de incumplimiento, el modelo ARMA presenta un RMSE de 0.0991 y un MAPE de 3.10%, incorporando variables como la variación interanual del tipo de cambio de compra, la tasa básica pasiva, la variación interanual de los costos de edificación, y la variación interanual del crédito en moneda nacional y en moneda extranjera.

La principal diferencia entre estos modelos, además del rendimiento, es que el modelo para la tasa de cura considera el costo de vivienda, mientras que el modelo para la tasa de incumplimiento considera el costo de edificios.

**Cuadro 16. Microempresas: Resumen de las medidas de error por tipo de modelo, ajuste de parámetros y combinación de variables con mejor rendimiento**

Tasa	Modelo	Parámetros	Variables	RSME	MAPE
<b>Cura</b>	<b>KNN</b>	<b>Triweight</b>	<b>var_tcc, var_tbp, var_vivienda, var_credito.mn, var.intervencion</b>	<b>0.2785</b>	<b>18.000%</b>
Cura	ARMA	(p=2,q=0)	var_imae, var_credito.mn, var.intervencion	0.3390	21.600%
Cura	SVR	Radial	var_inflacion, var_tcc, var_edificio, var_credito.mn, var.intervencion	0.3058	21.700%
Cura	Regresión Longitudinal	Tendencia	var_inflacion, var_credito.mn, var.intervencion	0.3111	22.400%
Cura	Bosques Aleatorios	(ntree=500,mtry=2)	var_imae, var.intervencion	0.4445	29.700%
Cura	RED	(size=1,decay=1)	var_inflacion, var_tbp, var_vivienda, var_edificio, var_credito.mn, var.intervencion	0.4754	33.700%
<b>Incumplimiento</b>	<b>ARMA</b>	<b>(p=5,q=1)</b>	<b>var_tcc, var_tbp, var_edificio, var_credito.mn, var_credito.me, var.intervencion</b>	<b>0.0991</b>	<b>3.100%</b>
Incumplimiento	RED	(size=6,decay=0.05)	var_inflacion, var_tcc, var_vivienda, var_edificio, var_credito.mn, var_credito.me, var.intervencion	0.1002	3.100%
Incumplimiento	KNN	Optimal	var_inflacion, var_imae, var_tcc, var_edificio, var.intervencion	0.1129	3.500%
Incumplimiento	SVR	Radial	var_edificio, var_credito.me, var.intervencion	0.1308	4.100%
Incumplimiento	Bosques Aleatorios	(ntree=50,mtry=4)	var_vivienda, var.intervencion	0.1379	4.300%
Incumplimiento	Regresión Longitudinal	Tendencia	var_tbp, var_vivienda, var_edificio, var.intervencion	0.1544	5.000%

Fuente: Cartera crediticia Banco Improsa

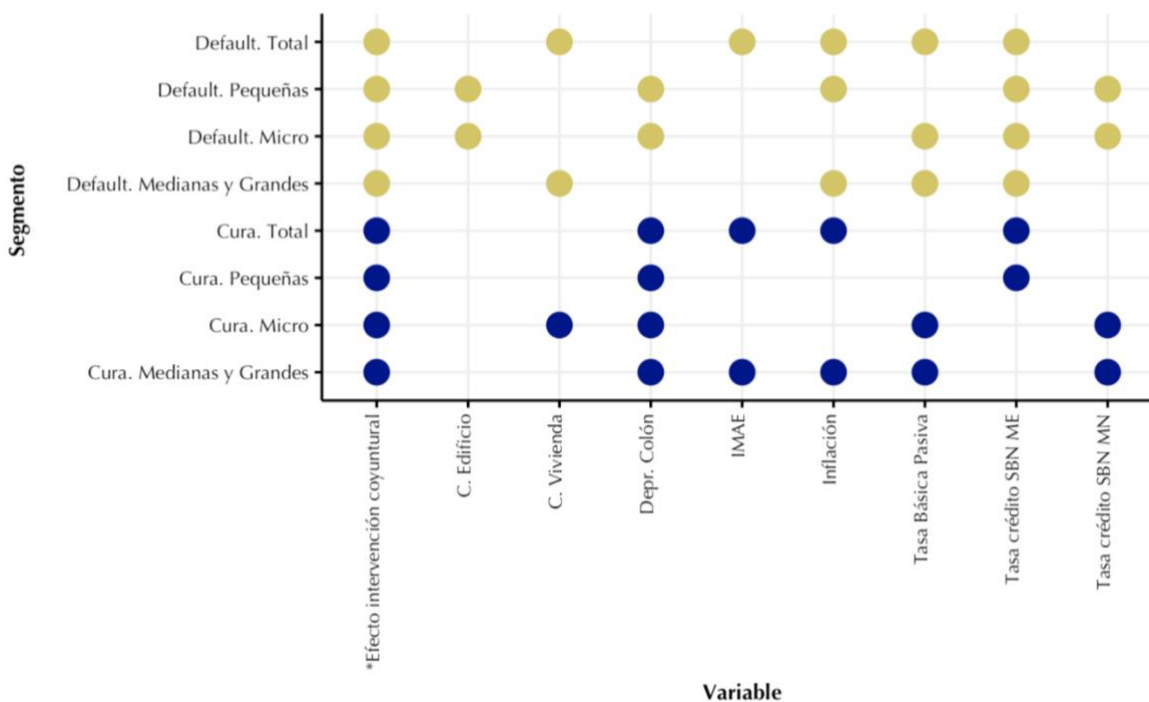
La Figura 34 muestra un gráfico de puntos que resume la selección de variables macroeconómicas utilizadas en modelos estadísticos para diferentes segmentos de empresas, con el objetivo de predecir tasas de incumplimiento y recuperación de créditos. Las variables se ubican en el eje horizontal y los segmentos empresariales en el eje vertical, incluyendo total, pequeñas, micro, y medianas y grandes empresas. Se observa que algunas variables son consistentemente seleccionadas a través de diferentes segmentos para ambos escenarios de recuperación e incumplimiento de créditos, sugiriendo que estas variables tienen un impacto generalizado y son consideradas predictores fuertes de la conducta crediticia en la cartera del banco.

Por ejemplo, la variación del crédito en ambas monedas parece ser de gran relevancia para la predicción de la tasa de cura de créditos, resaltando la importancia del acceso y las condiciones del crédito en la capacidad de las empresas para saldar sus deudas. La variación de la TBP y la variación de la inflación también son variables comunes en la modelación para ambos escenarios, reflejando cómo la situación económica general puede influir en el comportamiento crediticio. La selección de variables como la depreciación del colón sugiere

una consideración de cómo las políticas monetarias y la estabilidad de la moneda afectan la capacidad de pago de las empresas, especialmente en aquellas que manejan deudas o ingresos en moneda extranjera.

Por otro lado, los costos en construcción tienden a ser más seleccionados para las tasas de incumplimiento, lo que podría indicar una inclinación de los clientes a incurrir en gastos de construcción, como expandir su infraestructura y capacidad de almacenamiento, en vez de cumplir con sus obligaciones crediticias. Cabe destacar que los costos de vivienda solo se relacionan con la tasa de recuperación de las microempresas, lo cual tiene sentido ya que este segmento generalmente se caracteriza por contar con viviendas productivas, es decir, viviendas que son núcleos de sus giros de negocio. Además, el gráfico evidencia que la variable de intervención se mantuvo fija para los segmentos y tasas, como se detalló en el apartado de Metodología. Esto deja claro que este es un aspecto que se puede modificar y evaluar en futuras investigaciones.

**Figura 34. Resumen de la selección de variables según tipo de segmentación**



Fuente: Cartera crediticia Banco Improsa

En términos generales, el Cuadro 17 muestra un resumen de los resultados de modelos finales para predecir las tasas de recuperación y de incumplimiento de créditos, aplicando tanto un enfoque sin segmentar como uno segmentando por tamaño de empresa. Se aprecia a partir de los resultados obtenidos que la tasa de recuperación tiende a presentar medidas de error mayores en comparación a la tasa de incumplimiento, para la totalidad de la cartera, el modelo NNAR tiene un MAPE de 7.7%, lo cual es comparativamente bajo. Segmentando, los MAPEs son 7.3% para medianas y grandes empresas, 11.6% para pequeñas empresas y 18.0% para microempresas. Se observa una tendencia de aumento en el MAPE a medida que disminuye el tamaño de la empresa. Para la tasa de incumplimiento en el enfoque segmentar, el modelo NNAR arroja un MAPE de 2.0%, mostrando una precisión excepcional. Segmentando, el MAPE es de 3.2% para medianas y grandes empresas y 3.1% para microempresas. Aunque estos MAPEs son ligeramente más altos, aún reflejan una alta precisión. Para las pequeñas empresas, el MAPE es de 2.5%, que sigue siendo bajo y sugiere una precisión comparable con el enfoque sin segmentar.

**Cuadro 17. Resumen de los finales obtenidos y sus medidas de rendimiento**

Tasa	Enfoque	Segmento	Modelo	RSME	MAPE
Cura	Sin segmentar	Total	NNAR	0,1427	7,7
		Medianas y grandes	KNN	0,1596	7,3
	Segmentando	Pequeñas	KNN	0,2136	11,6
		Micro	KNN	0,2785	18,0
Incumplimiento	Sin segmentar	Total	NNAR	0,0699	2,0
		Medianas y grandes	ARMA	0,1246	3,2
	Segmentando	Pequeñas	NNAR	0,0880	2,5
		Micro	ARMA	0,0991	3,1

Fuente: Elaboración propia.

Analizando exclusivamente los valores de MAPE, el enfoque sin segmentar parece ofrecer una precisión general alta para la predicción de incumplimiento, lo que podría indicar que un modelo generalizado puede capturar adecuadamente la dinámica del incumplimiento de créditos para todas las empresas, independientemente de su tamaño. Para la tasa de cura de créditos, aunque el modelo sin segmentar presenta un buen rendimiento, el modelo segmentado para medianas y grandes empresas ofrece una precisión ligeramente mejor.

No obstante, los modelos segmentados son preferidos ya que permiten aplicar estrategias diferenciadas de gestión de riesgos y créditos adaptadas a las características únicas de cada segmento empresarial. Esto se debe, en parte, a que las microempresas presentan una alta variabilidad en sus tasas de recuperación, lo que justifica un MAPE mayor. Un modelo segmentado podría proporcionar predicciones más ajustadas a la realidad de estas empresas, mientras que el modelo de cartera total podría estar sesgando los valores y subestimar las medidas de error. Además, los modelos segmentados permiten obtener resultados más detallados y personalizados para intervenciones específicas o la simulación de escenarios de estrés en cada segmento de empresa, conduciendo a una mejor gestión de riesgos al identificar y mitigar factores de riesgo para diferentes tipos de empresas.

#### 4.2.3 Pruebas de estrés

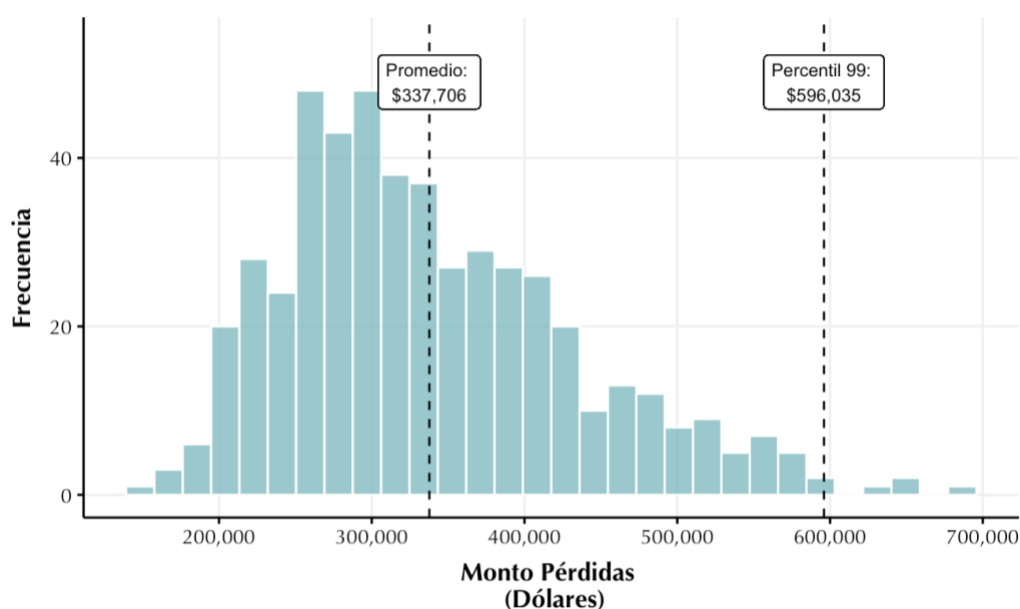
Esta sección se centra en los resultados de la prueba de estrés, analizando las tasas de cura e incumplimiento predichas a partir de escenarios simulados mediante el método de Monte Carlo. Además, se contrastan las distribuciones de pérdidas por riesgo de crédito y el indicador de porcentaje de crediticia de la cartera para los diferentes enfoques presentados. Es crucial recordar los supuestos establecidos en el apartado de Metodología, donde se definió un valor de cartera de \$90,000,000 y se estableció que el porcentaje de créditos cubiertos por garantías (colateralización) es del 80%.

##### 4.2.3.1 Cartera Total

Siguiendo la metodología de las pruebas de estrés de simulación de escenarios mediante Monte Carlo, los resultados expuestos en la Figura 35 muestran la distribución de pérdidas en la cartera total del banco, sin distinción por tamaño de empresa. El promedio de pérdidas es de \$337,706, lo cual ofrece una idea de la pérdida típica, pero no captura el riesgo de eventos extremos, que son especialmente relevantes en la gestión del riesgo. Por otro lado, según la metodología VaR, el percentil 99 se sitúa en \$596,035. Esto indica que el 1% de las pérdidas en la cartera excede este valor, lo cual sugiere que el banco debería asegurarse de tener suficiente capital para cubrir estas pérdidas potenciales extremas.

Al calcular el indicador del porcentaje de riesgo de crédito, tanto para la pérdida esperada (promedio) como para la pérdida inesperada (percentil 99), se obtiene que, siguiendo el método clásico, el indicador de la pérdida esperada es del 0.375%, lo cual se encuentra dentro de la clasificación normal de riesgo según los parámetros establecidos por SUGEF y presentados en el Cuadro 1. En el caso de la pérdida inesperada, aunque el valor es considerablemente mayor que la pérdida típica, el indicador de riesgo de crédito alcanza un valor de 0.662%, ubicándose también dentro del rango de riesgo aceptable.

**Figura 35. Banco Improsa: Distribución de pérdidas simuladas por riesgo crediticio según enfoque sin segmentar cartera**

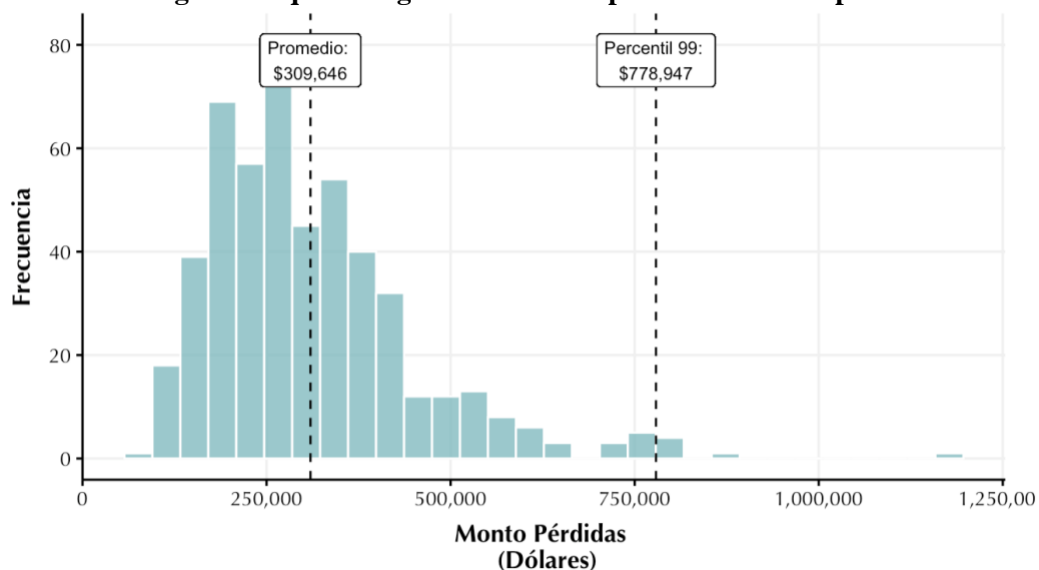


Fuente: Cartera crediticia Banco Improsa

#### 4.2.3.2 Cartera Segmentada

En la Figura 36, se muestra la distribución de pérdidas considerando la segmentación por tamaño de empresa. El promedio es ligeramente menor que en el enfoque sin segmentar (\$309,646), pero el percentil 99 es significativamente más alto (\$778,947), lo que indica que las pérdidas extremas pueden ser mayores en este enfoque. En el caso del indicador del porcentaje de riesgo de crédito según la pérdida esperada es igual a 0,344% y para la pérdida inesperada corresponde a 0,865%. De igual manera, ambos valores se encuentran dentro de la clasificación normal de riesgo según los parámetros establecidos por SUGEF.

**Figura 36. Banco Improsa: Distribución de pérdidas simuladas según enfoque de segmentar cartera por tamaño de empresa**



Fuente: Cartera crediticia Banco Improsa

Al contrastar ambas alternativas de análisis, los indicadores del porcentaje de riesgo de crédito, según las pérdidas esperadas, presentan valores cercanos al 0.3%. Esto sugiere que para estimar la pérdida esperada, representada por el promedio de pérdidas, el enfoque sin segmentar podría ser suficiente y más eficiente. Como se observa en las figuras, el promedio en el enfoque sin segmentar es ligeramente mayor, pero no de manera significativa. La ventaja de utilizar un enfoque sin segmentar es que simplifica el análisis y reduce la complejidad operativa, evitando la necesidad de desarrollar y mantener múltiples modelos para diferentes segmentos de tamaño de empresa, lo que puede ahorrar tiempo y recursos.

Sin embargo, es importante considerar que el promedio de pérdidas no captura la totalidad del riesgo, ya que ignora la variabilidad y la posibilidad de pérdidas extremas, que son cruciales tanto para el cálculo del capital económico como para la gestión de riesgos. El enfoque segmentado muestra un riesgo más alto de pérdidas extremas (percentil 99) en comparación con el enfoque sin segmentar, lo que podría indicar que segmentar por tamaño de empresa permite una identificación más clara de los escenarios de pérdidas extremas, esencial para la gestión efectiva del riesgo y la adecuación del capital. En el enfoque sin segmentar, el riesgo extremo parece ser menor, lo que podría llevar a subestimar el capital necesario para cubrir las pérdidas inesperadas.

Por lo tanto, si el objetivo es meramente estimar la pérdida esperada, el enfoque sin segmentar proporciona estimaciones suficientes. Sin embargo, si se requiere estimar la pérdida inesperada utilizando la metodología VaR, como recomiendan los organismos internacionales, el enfoque sin segmentar subestima el nivel de pérdida, lo que podría resultar en que el banco no reserve el capital suficiente para cubrir las pérdidas inesperadas.

Para esta investigación, se recomienda el enfoque de segmentación de la cartera, considerando tanto la precisión de los modelos como la estimación de los indicadores financieros y la variabilidad en los datos. Este enfoque no solo cumple con los acuerdos internacionales y con las pruebas de estrés que la Superintendencia implementa anualmente, sino que genera valor agregado al alinearse con la estrategia de negocio de la entidad financiera.

#### 4.3 Dashboard: Aplicativo *Shiny*

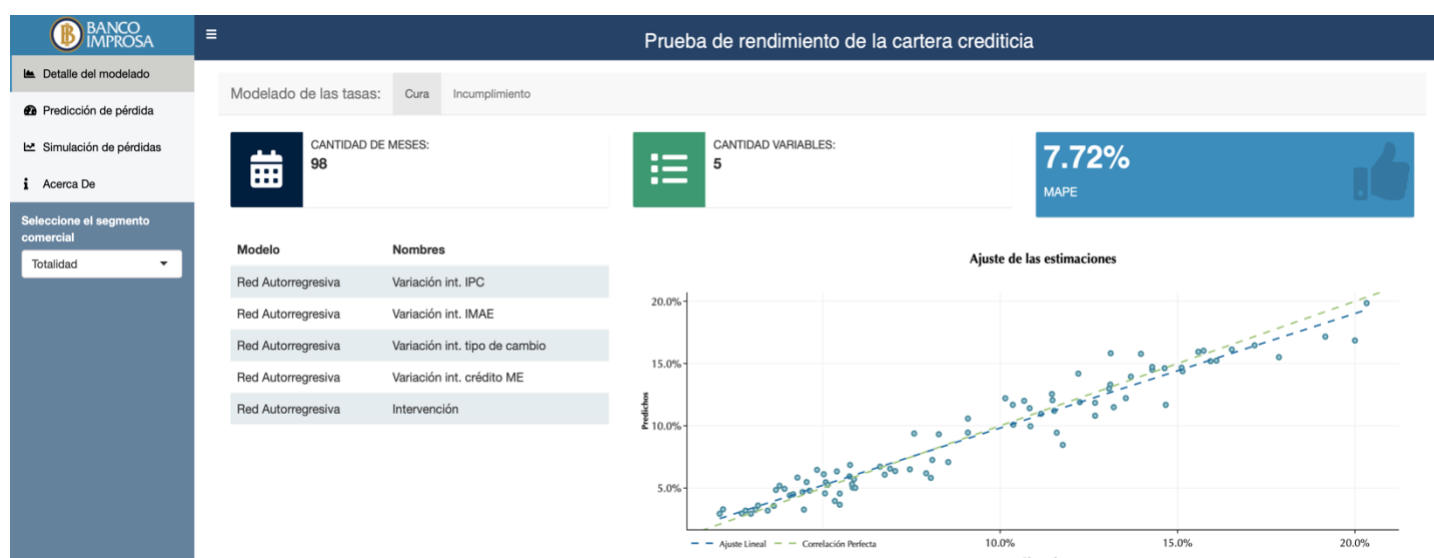
Uno de los objetivos de esta investigación es facilitar la usabilidad de los resultados de las pruebas de estrés a través de una aplicación web utilizando *Shiny*, que incluye visualizaciones interactivas permitiendo realizar una predicción a tiempo real de la pérdida esperada a partir de un escenario macroeconómico dado y simular distintos escenarios para obtener una distribución de pérdidas por riesgo de crédito. Al agregar valor a las pruebas de esta manera, los resultados ahora pueden ser utilizados eficazmente para la toma de decisiones dentro de la entidad bancaria. Es importante destacar que, para esta investigación, el aplicativo desarrollado no se ha publicado en ninguna plataforma de difusión, cumpliendo así con los lineamientos de seguridad de la información establecidos por la entidad bancaria.

Para un mayor entendimiento del aplicativo, se procede a detallar una guía de navegación del Dashboard titulado: “Prueba de Rendimiento de la Cartera Crediticia”. El cual se compone de 3 secciones de análisis: Detalle del modelado, Predicción de la Pérdida y Simulación de Pérdidas.

En esta primera sección (Figura 37), se presenta un resumen del modelo predictivo para cada segmento y tipo de tasa. En el panel lateral, el usuario puede seleccionar el segmento,

mientras que el tipo de tasa deseado se puede seleccionar desde las pestañas del panel principal. Como se muestra en el ejemplo, se incluyen resultados para la totalidad de la cartera y la tasa de cura. Se detallan los meses y variables utilizados en la modelación, la precisión del MAPE, el tipo de modelo matemático y las variables predictoras. Además, en la parte derecha del panel principal, un gráfico de dispersión compara las estimaciones con los valores observados, facilitando la visualización del rendimiento del modelo mediante líneas de ajuste lineal y de correlación perfecta.

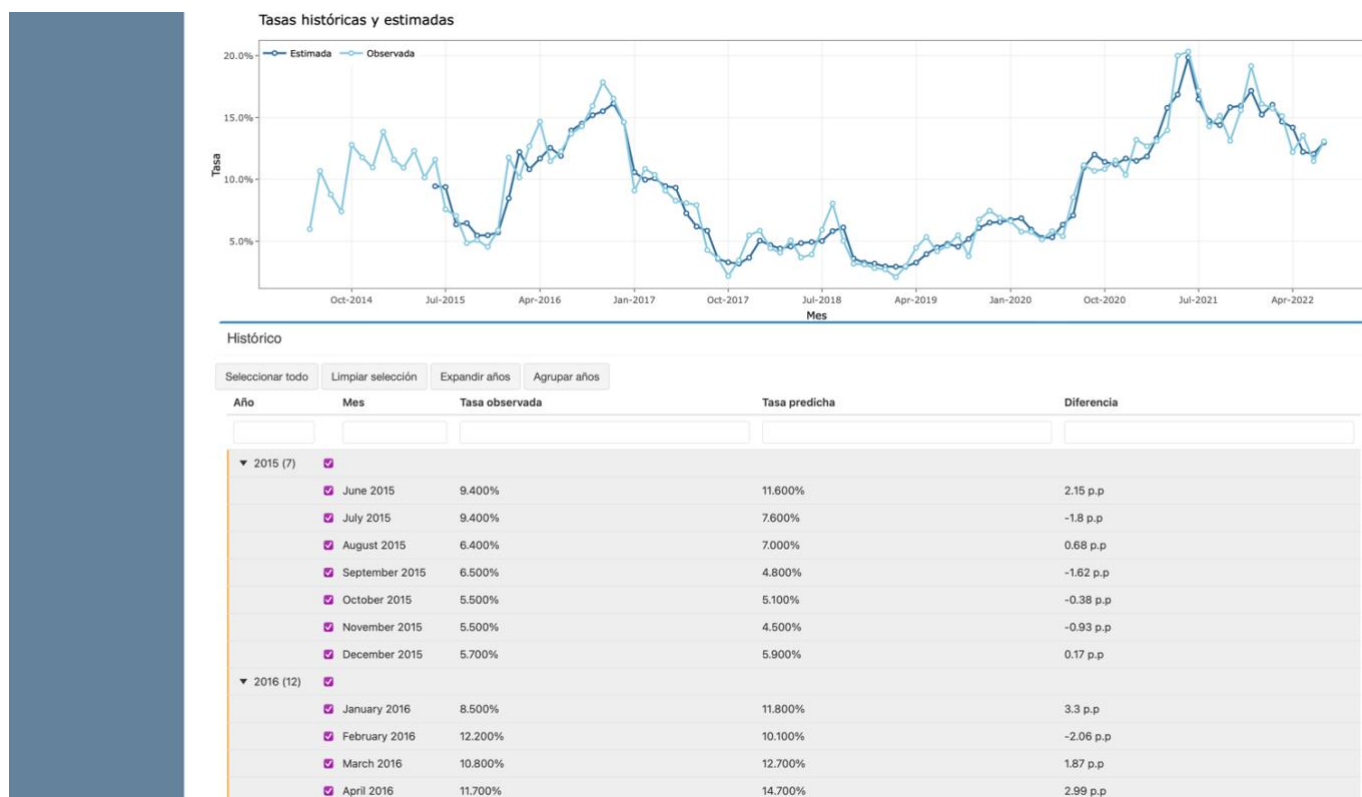
**Figura 37. Primera sección: Detalle del modelado, Parte I**



Fuente: Cartera crediticia Banco Improsa

En la Figura 38, desplazándose hacia abajo en la misma sección, se encuentra un descriptivo de tasas históricas y estimadas. En la parte superior, un gráfico de líneas compara las tasas estimadas con las observadas, permitiendo visualizar cómo las predicciones del modelo se alinean con los datos históricos. Debajo del gráfico, una tabla detalla las tasas observadas y predichas mes a mes, junto con la diferencia entre ambas. Esta tabla cubre varios años, desde 2015 hasta 2022, facilitando un análisis detallado de la precisión del modelo. Cabe destacar que la tabla incluye un filtro que permite al usuario seleccionar los meses específicos que desea visualizar en el gráfico superior.

**Figura 38. Primera sección: Detalle del modelado, Parte II**



Fuente: Cartera crediticia Banco Improsa

Por otro lado, la segunda sección, denominada 'Predicción de la pérdida esperada', se muestra en la Figura 39. En el panel izquierdo, se aprecian los distintos parámetros de entrada que el usuario puede modificar para obtener la predicción de las tasas de cura e incumplimiento, así como la pérdida esperada. Los parámetros de entrada disponibles incluyen: tipo de moneda (dólares o colones), monto de la exposición en caso de impago, porcentaje de colateralización, y las variables económicas necesarias para la predicción, las cuales pueden variar dependiendo del segmento que se desee analizar. En el ejemplo presentado en esta figura para las pequeñas empresas, las estimaciones de las tasas se muestran en el panel derecho en la parte superior. Además, en la parte inferior se muestran dos gráficos: uno ilustra el valor de la pérdida esperada y el porcentaje de riesgo de crédito, lo cual, dependiendo del valor, permite visualizar el tipo de riesgo que se presentaría; el

segundo gráfico muestra la pérdida esperada desglosada por sus componentes. Donde se observa que el monto de la pérdida esperada se clasifica dentro del nivel 1 de riesgo de crédito, correspondiente a un indicador de 2.34%.

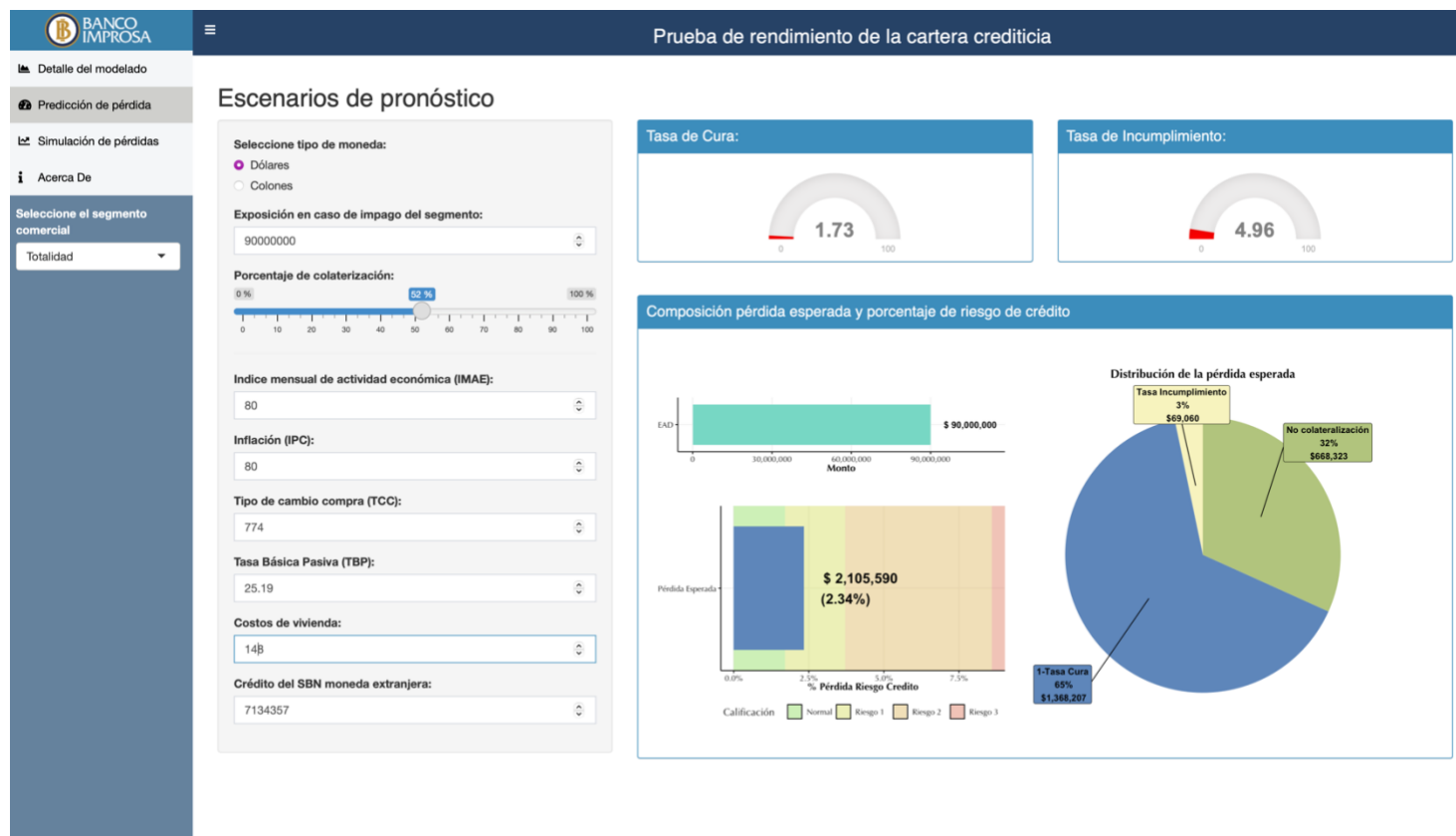
**Figura 39. Segunda sección: Predicción de la Pérdida para pequeñas empresas**



Fuente: Cartera crediticia Banco Improsa

De la misma manera, la Figura 40 muestra esta sección del dashboard, pero específicamente para la totalidad de la cartera crediticia. En el escenario proyectado, se observa que el indicador de pérdida de riesgo corresponde a 0.35% por ende se clasifica dentro del nivel de riesgo de crédito normal.

**Figura 40. Segunda sección: Predicción de la Pérdida para la totalidad de la cartera**



Fuente: Cartera crediticia Banco Improsa

La tercera sección del dashboard, denominada ‘Simulación de Pérdidas’, se muestra en la Figura 41. Esta sección permite a los usuarios simular la distribución de pérdidas de la cartera total del banco mediante el uso de Monte Carlo y un modelo de tasas sin segmentar. Los usuarios pueden ajustar varios parámetros según su interés, incluyendo el número de escenarios, la exposición en caso de impago, el porcentaje de colateralización, el tipo de moneda (dólares o colones), y el método para estimar los escenarios, que puede ser a través de promedios históricos o valores ingresados manualmente. El resultado se presenta en forma de un histograma que muestra la distribución de los montos de pérdida esperada por nivel de riesgo de crédito, facilitando la visualización de la variabilidad en las pérdidas esperadas bajo diferentes escenarios. En la simulación ejemplificada, se observa que el promedio de pérdidas es de \$323,994, mientras que el percentil 99 alcanza los \$577,734.

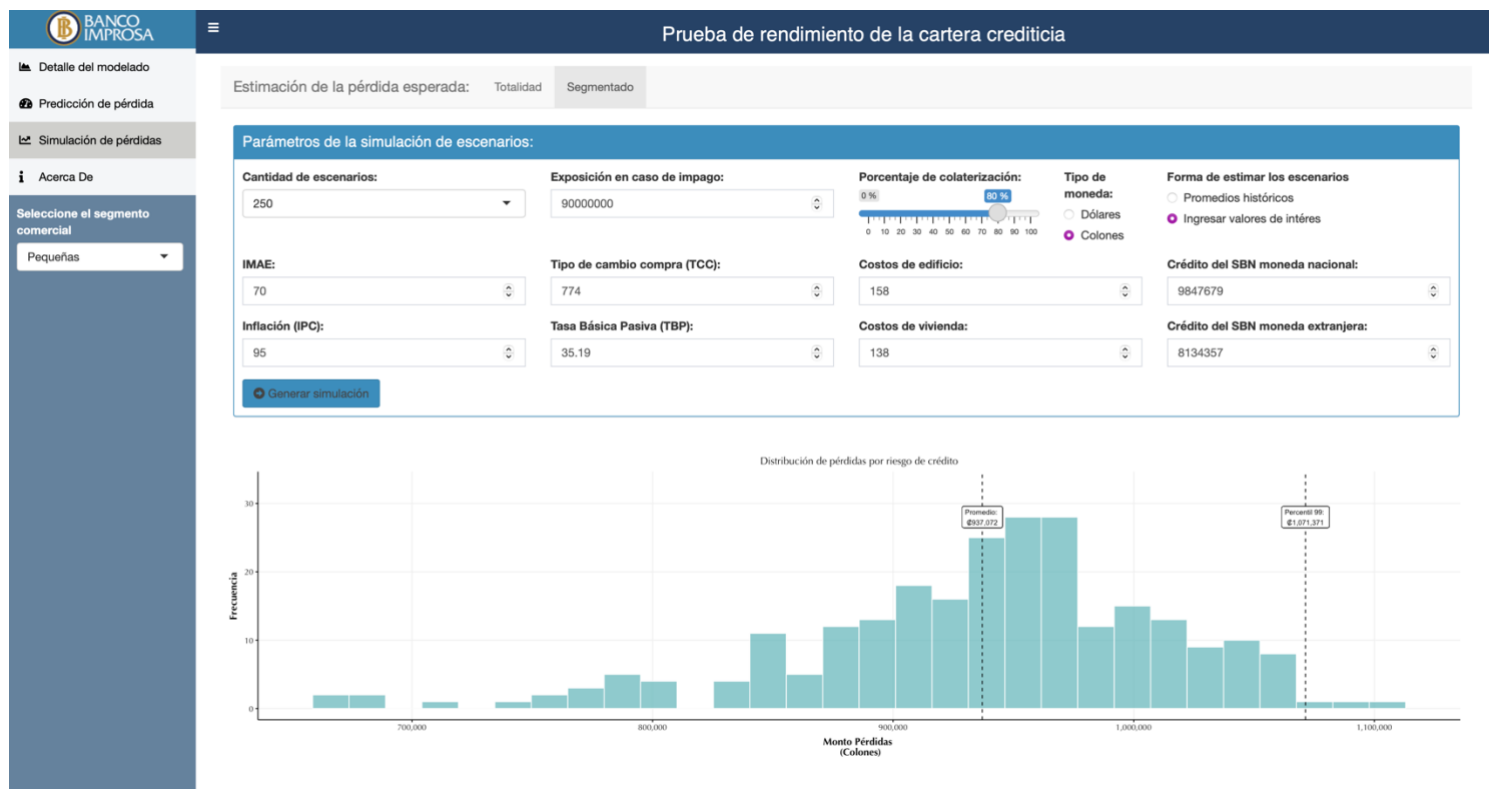
**Figura 41. Tercera sección: Simulación de Pérdidas, metodología sin segmentar la cartera**



Fuente: Cartera crediticia Banco Improsa

Por otro lado, la Figura 42 se presenta la distribución en este caso implementando un modelado segmentando por el tamaño de empresa e ingresando valores de las variables macroeconomicas manualmente. El histograma resultante muestra una distribución detallada de los montos de pérdida esperada, proporcionando una representación visual del riesgo de crédito donde el promedio de pérdidas es de \$937,072, mientras que el percentil 99 alcanza los \$1,071,731.

**Figura 42. Tercera sección: Simulación de Pérdidas, metodología segmentado la cartera por el tamaño de empresa**



Fuente: Cartera crediticia Banco Improsa

## CAPÍTULO V: CONCLUSIONES Y DISCUSIÓN

En general, la presente investigación ofrece una prueba de estrés a la cartera crediticia mediante el pronóstico de las tasas de mora e incumplimiento, en el contexto de recientes cambios económicos provocados por crisis financieras que han marcado la necesidad de contar con sistemas financieros robustos. Las pruebas de estrés se han consolidado como herramientas esenciales para evaluar la estabilidad y solidez del sector financiero, y su adaptación a las características específicas de cada institución es crucial para obtener resultados precisos. Este enfoque es especialmente relevante para Banco Improsa, que centra sus operaciones en el enfoque empresarial, lo que requiere una atención particular a los riesgos específicos de su cartera crediticia.

El capítulo del marco teórico abordó las directrices internacionales de supervisión bancaria del Comité de Basilea, la regulación bancaria en Costa Rica y los conceptos de riesgo de crédito y métodos de gestión de riesgos. Se destacaron los Acuerdos de Basilea I, II y III, que establecen estándares globales para la supervisión bancaria. En Costa Rica, las doctrinas de Basilea se han incorporado gradualmente mediante normativas como el Reglamento sobre la Suficiencia Patrimonial de Entidades Financieras.

Dentro de estas regulaciones, se detallaron los indicadores clave de riesgo de crédito, como la probabilidad de incumplimiento, la pérdida en caso de impago y la exposición en el momento del impago. Se enfatizó la importancia de las pruebas de estrés para evaluar la capacidad de las entidades financieras de enfrentar situaciones adversas, destacando que la implementación de metodologías internas que simulen escenarios macroeconómicos de estrés es fundamental para gestionar eficazmente los riesgos bancarios y garantizar la estabilidad y seguridad del sistema financiero.

La metodología implementada demostró cómo, a través de modelos de series temporales y técnicas de aprendizaje automático, se puede incorporar para llevar a cabo una prueba de estrés en la cartera de créditos de una institución financiera costarricense. Se describieron los datos utilizados, los métodos analíticos aplicados, las técnicas de validación de modelos y las especificaciones para estimar las pérdidas crediticias. Se detalló cómo se

generaron los escenarios macroeconómicos mediante simulaciones de Monte Carlo para estimar las tasas de cura e incumplimiento y se calculó la pérdida crediticia esperada, considerando supuestos sobre el porcentaje de colateralización y la exposición en caso de impago.

Este enfoque permitió obtener estimaciones de las pérdidas que ayudan a la institución financiera a evaluar su riesgo crediticio y calcular indicadores clave de riesgo, ofreciendo una alternativa innovadora y de vanguardia mediante la incorporación de un proceso de simulación. Cabe destacar que este tipo de metodología no se considera en la Prueba BUST de la SUGEF, que basa sus escenarios en el criterio de expertos, lo cual puede introducir sesgos en el ejercicio. Al utilizar simulaciones, se logra una mayor objetividad y precisión en la evaluación del riesgo, reduciendo la dependencia de interpretaciones subjetivas.

Entre los resultados más importantes del análisis de las tasas de cura e incumplimiento en la cartera crediticia del Banco Improsa, se observó que las tasas de cura varían a lo largo del tiempo según el tamaño de las empresas, siendo más altas en las empresas más pequeñas en los últimos meses. En contraste, las tasas de incumplimiento tienden a ser más elevadas en las micro y pequeñas empresas en comparación con las empresas corporativas. El análisis de la tendencia de estas tasas a lo largo del tiempo reveló fluctuaciones y cambios en su estabilidad. En general, se identificaron patrones estacionales y fluctuaciones irregulares que podrían deberse a factores económicos y políticas bancarias. Estos hallazgos resaltan la importancia de considerar el tamaño de las empresas al evaluar el riesgo crediticio y adaptar las estrategias de gestión de riesgo en función de estas diferencias. Para Banco Improsa, es importante no descartar la segmentación por tamaño de empresa al analizar su cartera, ya que permite abordar de manera efectiva las discrepancias observadas.

Con respecto a la interacción de las tasas de cura e incumplimiento con las variables macroeconómicas utilizadas en la investigación, se aprecia que en relación a las tasas de cura, la tasa básica pasiva y el IMAE tienen impactos significativos en la cartera total, mientras que en las empresas corporativas no hay correlaciones fuertes. Las pequeñas empresas se ven

afectadas por la TBP, y las microempresas pueden beneficiarse de aumentos en los costos de construcción. Por otro lado, un aumento en la variación de los créditos en ambas monedas puede estar relacionado con una disminución en las tasas de incumplimiento. Estas correlaciones pueden ayudar a prever el comportamiento futuro de las tasas de cura e incumplimiento, lo que puede ser fundamental para la formulación de estrategias de gestión del riesgo crediticio.

En relación con la predicción de las tasas de cura e incumplimiento, se desarrollaron diversos modelos utilizando técnicas como la regresión con series temporales, el modelo autorregresivo de promedio móvil (ARMA), redes neuronales autorregresivas (NNAR), el método de k vecinos más cercanos (KNN), máquinas de soporte vectorial (SVR) y bosques aleatorios. Estos modelos se evaluaron implementando diferentes métricas de precisión, como el RMSE y el MAPE, mediante validación cruzada. Cada modelo presentó un rendimiento distinto; sin embargo, los modelos ARMA, NNAR y KNN se destacaron por su precisión y consistencia en las predicciones.

En general, todos los modelos presentaron MAPE deseables inferiores al 18%. No obstante, los modelos que predijeron las tasas de incumplimiento mostraron un rendimiento destacado en diferentes segmentos de empresas, como medianas, grandes, pequeñas y micro, siendo los más precisos con MAPE inferiores al 3.2%. Estos resultados subrayan la efectividad de los modelos para proporcionar predicciones confiables en diversos contextos empresariales.

Es importante resaltar como los modelos de series de tiempo mostraron un rendimiento superior en la predicción de las tasas de cura e incumplimiento debido a su capacidad para capturar patrones y tendencias temporales. La inclusión del factor temporal en los modelos de pronóstico es esencial para mejorar la precisión y adaptarse a los cambios en el entorno financiero. Por lo tanto, es fundamental considerar las series de tiempo en los pronósticos para una gestión más efectiva del riesgo crediticio y una mejor toma de decisiones estratégicas.

Adicionalmente, se identificó que variables macroeconómicas como la inflación, la TBP, la depreciación del colón y el crédito en moneda extranjera tienen un impacto significativo en el comportamiento crediticio de las empresas. Los costos de construcción están más relacionados con las tasas de incumplimiento, mientras que los costos de vivienda impactan principalmente las tasas de las microempresas. Estos hallazgos sugieren que diferentes factores económicos y de gasto pueden influir en la capacidad de pago de las empresas y en su comportamiento crediticio, reafirmando la importancia de considerar la segmentación por tamaño de empresa en el análisis de la cartera crediticia de la institución financiera.

Al analizar en detalle los modelos segmentados por tamaño de empresa, se observa que son preferidos porque permiten aplicar estrategias específicas de gestión de riesgos y créditos adaptadas a las características únicas de cada tipo de empresa. Estos modelos proporcionan predicciones más ajustadas a la realidad de las microempresas, evitando sesgos y subestimación de las medidas de error. Además, ofrecen resultados detallados y personalizados para cada segmento, lo que conduce a una mejor gestión de riesgos al identificar y mitigar factores de riesgo específicos para cada tipo de empresa.

Se analizaron los resultados de la prueba de estrés utilizando el método de Monte Carlo para simular diferentes escenarios. Se compararon las tasas de cura e incumplimiento predichas, así como las distribuciones de pérdidas por riesgo de crédito y el indicador de porcentaje de riesgo crediticio en la cartera del banco. La prueba de estrés implementada se centró en analizar las pérdidas esperadas y extremas en la cartera del banco, comparando los resultados entre un enfoque sin segmentar y otro segmentado por tamaño de empresa. Se encontró que, aunque el enfoque sin segmentar simplifica el análisis, el enfoque segmentado permite identificar con mayor claridad los escenarios de pérdidas extremas, lo cual es esencial para la gestión del riesgo. Se recomienda utilizar el enfoque segmentado para estimar tanto la pérdida esperada como la inesperada, ya que se alinea con los estándares internacionales y con la estrategia del negocio del banco.

Cabe destacar que, al implementar ambas metodologías (segmentada y no segmentada) en el cálculo, tanto la pérdida esperada como la inesperada se encontraron dentro de los parámetros de riesgo aceptables establecidos por la SUGEF. Este resultado es de suma importancia para el banco, ya que indica que se encuentra dentro de rangos seguros de riesgo. No obstante, esta estimación debe estar en constante actualización para poder corroborar que se cuenta con suficiente capital para cubrir posibles pérdidas extremas.

Por último, la aplicación web desarrollada utilizando Shiny es una herramienta que facilita la usabilidad de los resultados de la presente prueba de estrés. Permite predecir pérdidas por riesgo de crédito, simular distintos escenarios de riesgo y facilitar la toma de decisiones en la entidad bancaria. El aplicativo sigue los lineamientos de seguridad de la información del banco, asegurando la protección de los datos y la integridad de los procesos de análisis.

El presente trabajo constituye un primer abordaje en la construcción de una prueba de estrés interna para Banco Improsa, que no solo satisface las necesidades estratégicas del negocio en términos de segmentación por tamaño de empresa, sino que también incorpora soluciones de vanguardia. Al integrar modelos robustos de series de tiempo, como los modelos autorregresivos y las redes neuronales, junto con simulaciones de Monte Carlo para la estimación de pérdidas, se amplían y mejoran los resultados en comparación con las metodologías previamente implementadas en la institución.

Para futuras investigaciones, se identifican varias oportunidades de mejora. En primer lugar, incorporar más variables exógenas relacionadas con el tiempo, como días hábiles y feriados, además de considerar otras posibles variables de intervención, podría enriquecer el análisis. También sería beneficioso incluir el rezago de las variables predictoras para capturar posibles efectos retardados de los indicadores macroeconómicos. Finalmente, aumentar la granularidad del análisis más allá del tamaño de empresa permitiría una comprensión más detallada de los diferentes segmentos de la cartera crediticia. Estas mejoras podrían aportar a un modelo más preciso y adaptable a las dinámicas cambiantes del entorno económico.

## BIBLIOGRAFÍA

- Acevedo, E., Serna, A., & Serna, E. (2017). *Principios y características de las redes neuronales artificiales*. Desarrollo e innovación en ingeniería, 173.
- Adhikari, R., K. A. R., & Agrawal, R. K. (2013). *An introductory study on time series modeling and forecasting* (pp. 42–45). Lap Lambert Academic Publishing GmbH KG. <https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf>
- Arce, R., Mahía R., (2003) *MODELOS ARIMA*. Academia. Recuperado de: [shorturl.at/jFGHJ](http://shorturl.at/jFGHJ)
- Arce, R., y Mahía, G. (2009). *Modelos Programa Citius*.- Técnicas de Previsión de variables financieras. (M. d. Citius, Ed.).
- Avella, M. (2004). *Los acuerdos bancarios de Basilea en perspectiva*. Revista del Banco de la República.
- Baker, S., Ponniah, D, & Smith, S. (1999). *Survey of Risk Management in Major UK Companies*. Journal of Professional Issues in Engineering Education and Practice, 94-102.
- Banco de España. (2020). *Escenarios macroeconómicos de referencia para la economía española tras el covid-19*. BOLETÍN ECONÓMICO.
- Banco Mundial. (2020). *El Banco Mundial en Costa Rica. Costa Rica: panorama general*. Recuperado de: <https://www.bancomundial.org/es/country/costarica/overview>
- Barnard, G. A. (1963). *The logic of least squares*, J. R. Stat. Soc., B25, 124--127.
- BCBS. (2001). *History of The Basel Committee and its membership*. Banco de Pagos Internacionales. Suiza.
- BCBS. (2005). *Studies on the validation of internal rating systems*. Banco de Pagos Internacionales. Suiza.
- BCBS. (2006). *Convergencia internacional de medidas y normas de capital*. Banco de Pagos Internacionales. Suiza.
- BCBS. (2009). *Principios para la realización y supervisión de pruebas de tensión*. Banco de Pagos Internacionales. Suiza.
- BCBS. (2010). *Marco regulador internacional para los bancos*. Banco de Pagos Internacionales. Suiza.

- BCBS. (2011). *Revisiones del marco de riesgo de mercado de Basilea II*. Banco de Pagos Internacionales. Suiza.
- BCCR, CONASSIF. (2020). *BCCR y Conassif acuerdan medidas de política monetaria y financiera para mitigar impacto económico del covid-19*. Comunicado de Prensa. San José, Costa Rica.
- BCCR. (2012). *Cambio de año base. Índice Mensual de Actividad Económica (IMAE) Cuentas Nacionales de Costa Rica. Año Referencia 2012*. San José, Costa Rica.
- BCCR. (2015). *La Junta Directiva del Banco Central de Costa Rica en el artículo 8 del acta de la sesión 5703-2015, celebrada el 13 de octubre del 2015*. San José, Costa Rica.
- BCCR. (2019a). *Revisión Programa Macroeconómico 2019-2020*. San José, Costa Rica.
- BCCR. (2019b). *Informe Anual de Estabilidad Financiera*. San José, Costa Rica.
- Bellotti, T., & Crook, J. (2013). *Forecasting and stress testing credit card default using dynamic models*. *International Journal of Forecasting*, 563–574.
- Benavides, G., & Jiménez, L. (2015). *Stress-Testing para carteras de crédito del Sistema Bancario Mexicano*. *Revista Mexicana de Economía y Finanzas*, 117–140.
- Blanco, R., & García, A. (2004). *Las pruebas de estrés en los programas de evaluación del sistema financiero*. *Estabilidad Financiera*, 107-122.
- Borio, C., Drehmann, M., & Tsatsaronis, K. (2014). *Stress-testing macro stress testing: Does it live up to expectations?*. *Journal of Financial Stability*, 3-15.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 44e152). New York, NY, USA: ACM.
- Boss, M. (2002). *Macroeconomic Credit Risk Model for Stress Testing the Austrian Credit Portfolio*. Financial Stability Report 4, Oesterreichische Nationalbank.
- Box, G., y Jenkins, G. (1976). *Time series analysis: Forecasting and control*. New Jersey, Wiley.
- Box, G., y Tiao, G. (1975). *Intervention analysis with applications to economic and environmental problems*. *Journal of the American Statistical Association*, No 70, 70-79
- Breiman, L (2001). *Random forests, Machine Learning*. 45: 5–32.

- Brockwell, P. J., Brockwell, P. J., Davis, R. A., & Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.
- Camacho, A. (2014). *El sistema financiero costarricense en los últimos 25 años*. San José, Costa Rica: Academia de Centroamérica.
- Cámara de Comercio de Costa Rica. (2012). *Estadísticas Económicas. Cámara De Comercio De Costa Rica*. San José, Costa Rica.
- Campbell, C.; Cristianini, N. (1999) *Simple learning algorithms for training support vector machines*, Department of Engineering Mathematics, University of Bristol.
- Cardone, C., & Trujillo, A. (2008). *Mitigación del riesgo de crédito en Basilea II y financiación de las PYME. El caso del aval de las SGR*. Madrid, España: Ministerio de Educación y Ciencia.
- CEPAL. (2020a). *América Latina y el Caribe ante la pandemia del COVID-19*. Efectos económicos y sociales. Informe Especial COVID-19.
- CEPAL. (2020b). *Dimensionar los efectos del COVID-19 para pensar en la reactivación*. Efectos económicos y sociales. Informe Especial COVID-19.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). *Shiny: web application framework for R*. R package version, 1(5), 2017.
- Chavarín R. (2015). *Morosidad en el pago de créditos y rentabilidad de la banca comercial en México*. Revista Mexicana de Economía y Finanzas, 71-83.
- Chorafas, D. (2000). *Managing credit risk, analysis rating and pricing the probability of default*. Euromoney Institutional Investor PLC, Londres.
- CNE, MEIC, MTSS. (2020). *Guía para la prevención, mitigación y continuidad del negocio por la pandemia del covid-19 en los centros de trabajo*. Mesa de Infraestructura Productiva Sub-mesa de trabajo “Continuidad del Negocio”. San José, Costa Rica.
- CONASSIF. (2003). *Información de entidades supervisadas dispuesta al público por medio del sitio web*. San José, Costa Rica.
- CONASSIF. (2008). *Reforma Reglamento sobre Apertura y Funcionamiento de Entidades Autorizadas y Funcionamiento de Fondos de Pensiones, Capitalización Laboral y Ahorro Voluntario Previstos en Ley de Protección al Trabajador y Reglamento de Inversiones Entidades Reguladas*. San José, Costa Rica.

- Conway, D., & White, J. (2012). *Machine Learning for Hackers*. Sebastopol, CA: O'Reilly Media.
- Covas, F., Rumpb, B., & Zakrajšek, E. (2014). *Stress-testing US bank holding companies: A dynamic panel quantile regression approach*. *International Journal of Forecasting*, 714-716.
- Cruz, R., & Corrêa, C. (2017). *Previsión de demanda intermitente con métodos de series de tiempo y redes neuronales artificiales: Estudio de caso*. *Dyna*, 84(203), 9-16.
- Delgado, J., & Saurina, J. (2004). *Riesgo de crédito y dotaciones a insolvencias. Un análisis con variables macroeconómicas*. Dirección General de Regulación Banco de España. Madrid, España.
- Díaz E., Rangel P. & Martínez F. (2006) *Adaptive Support Vector Machines para predicción de series de tiempo*. XI Simposio de tratamiento de señales, imágenes y visión artificial. STSIVA 2006, Colombia.
- Díaz, A., (2015) *Crisis de la zona euro: fallas estructurales y políticas de austeridad*. ECONOMÍA UNAM, Universidad Autónoma Metropolitana, 45-68.
- Dissanayake, B., Hemachandra, O., Lakshitha, N., Haputhanthri, D., & Wijayasiri, A. (2021). *A comparison of ARIMAX, VAR and LSTM on multivariate short-term traffic volume forecasting*. In Conference of Open Innovations Association, FRUCT (No. 28, pp. 564-570).
- Dowd K., (1998) *Beyond Value at Risk. The New Science of Risk Management*. John Wiley & Sons, Reino Unido.
- EBA. (2018). *Prueba de estrés 2018 en toda la Unión Europea*. Autoridad Bancaria Europea.
- Feria J., & Oliver, M. (2004) *Valor en riesgo relativo (VeR): más allá de la teoría de carteras*. *Revista Electrónica de Ciencia Administrativa (RECADM)*. Sevilla, España.
- Fernández, H., & Pérez, F. (2005). *El modelo logístico: Una herramienta estadística para evaluar el riesgo de crédito*. *Revista Ingenierías Universidad de Medellín*, 55-73.
- Fernández, H., & Pérez, F. (2007). *Las redes neuronales y la evaluación del riesgo de crédito*. *Revista Ingenierías Universidad de Medellín*, 77-91.
- FMI. (2016). *El Programa de Evaluación del Sector Financiero (PESF)*. FMI Hoja de hechos. Washington, DC.

- FMI. (2018). *Costa Rica Revisión de la Estabilidad del Sector*. Informe de país del FMI. Washington, DC.
- Gamboa, C. (2023). *La sobreparametrización en el ARIMA: una aplicación a datos costarricenses*. (Tesis de posgrado). Universidad de Costa Rica, San José, Costa Rica.
- García, J., & Salazar, P. (2005). *Métodos de Administración y Evaluación de Riesgos*. (Tesis de grado). Universidad de Chile, Santiago, Chile.
- Garciandía, R. (2009). *Los Acuerdos de Basilea: algunas reflexiones sobre su naturaleza jurídica a la luz de la reciente crisis financiera*. *Revista española de derecho internacional*, 551-556.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2021). *Bayesian Data Analysis* (3rd ed.). Editorial.
- Goddard J., de los Cobos S., Pérez B., & Gutiérrez M., (2009) *Un algoritmo para el entrenamiento de máquinas de vector soporte para regresión*, *Revista de Matemática: Teoría y Aplicaciones* Vol. 7 Num.1-2.
- Goldman Sachs. (2020). *Roaring into recession*. Top of Mind, 87. Recuperado de: <https://www.goldmansachs.com/insights/pages/roaring-into-recession-f/report.pdf>
- González, C., & Uquillas, A (2017). *Modelo Macro para Pruebas de Tensión de Riesgo de Crédito de Consumo en el Sistema Financiero Ecuatoriano*. *Analitika*, *Revista de análisis estadístico*, 75-99.
- González, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. Sarriko-On. País Vasco.
- Grupo Financiero Improsa. (2018). *Memoria Anual*. San José, Costa Rica.
- Gutiérrez, M. (2009). *Basilea II: logros y retos para la legislación bancaria costarricense*. (Tesis de grado). Universidad de Costa Rica, San José, Costa Rica.
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- Hansen, J. & Nelson, R. (2003). *Forecasting and recombining time-series components by using neural networks*, *Journal of the Operational Research Society*, 54:3, 307-317.
- Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed). Springer

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer.
- Hechenbichler, K., & Schliep, K. (2004). *Weighted k-nearest-neighbor techniques and ordinal classification*.
- Hernández F., Brown, M. (1970). *El Concepto de Distancia y su Aplicación en Estadística Multivariada*. Datos Diagnosticos y Tendencias.
- Hernández, O. (2011). *Introducción a las series cronológicas*. Editorial UCR.
- Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.
- INEC. (2020). *Índice de Precios al Consumidor base diciembre 2020. Metodología*. San José, Costa Rica.
- Izaurieta, F., & Saavedra, C. (2000). *Redes neuronales artificiales*. Departamento de Física, Universidad de Concepción Chile.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York, NY: Springer.
- Jorion, P. (2003). *Financial Risk Manager Handbook*. Third Edition. New Jersey: Wiley.
- Juárez G., Sanchez A., & Zurita J. (2015) *La crisis financiera internacional de 2008 y algunos de sus efectos económicos sobre México*. Contaduría y Administración. Universidad Autónoma Metropolitana, 128-146.
- Kaczmarzyk, J. (2016). *Reflecting interdependencies between risk factors in corporate risk modeling using Monte Carlo simulation*. *Ekonometria*, (52), 98-107.
- Karadzic, V., & Pejovic, B. (2021). *Inflation forecasting in the Western Balkans and EU: a comparison of Holt-Winters, ARIMA and NNAR Models*. *Amfiteatru Econ*, 23, 517-532.
- Krugman, P., Wells, R. y Olney, M. (2011) *Introducción a la Economía*, Editorial Reverte. Barcelona, España.
- Kuhn, M. (2019). *The caret Package: Classification and Regression Training*. Recuperado de <https://topepo.github.io/caret/>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer.

- Ley N° 9694. *Ley del Sistema de Estadística Nacional*, aprobada el 30 de mayo de 2019 y publicada en el Alcance N° 133 de La Gaceta N° 110 del 13 de junio de 2019. San José, Costa Rica.
- Liaw, A., & Wiener, M. (2002). *Classification and regression by randomForest*. R news, 2(3), 18-22.
- Maleki, A., Nasser, S., Aminabad, M. S., & Hadi, M. (2018). *Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics*. KSCE Journal of Civil Engineering, 22, 3233-3245.
- Martín, J. J. (2016). *Support Vector Regression: Propiedades y aplicaciones*. In Universidad de Sevilla. Departamento de Estadística e Investigación Operativa. Universidad de Sevilla.
- Matich, D. J. (2001). *Redes Neuronales: Conceptos básicos y aplicaciones*. Universidad Tecnológica Nacional, México, 41, 12-16.
- McNeil, A., Frey, R., & Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press. Nueva Jersey, Estados Unidos.
- MEIC. (2017). *Informe PND, Año 2017. Sector Economía, Industria y Comercio Análisis de las metas de los objetivos sectoriales*. San José, Costa Rica.
- Mitrea, C. A., Lee, C. K. M., & Wu, Z. (2009). *A comparison between neural networks and traditional forecasting methods: A case study*. International journal of engineering business management, 1, 11.
- Mukid, M. A., Widiari, T., Rusgiyono, A., & Prahutama, A. (2018). *Credit scoring analysis using weighted k nearest neighbor*. In Journal of Physics: Conference Series (Vol. 1025, No. 1, p. 012114). IOP Publishing.
- OIT. (2020). *El COVID-19 y el mundo del trabajo: Repercusiones y respuestas*. Observatorio de la OIT, Ginebra, Suiza.
- OMS. (2020). *Enfermedad por coronavirus 2019 (COVID-19). Informe de situación - 51*. . Ginebra, Organización Mundial de la Salud.
- Pankratz, A. (1983) *Forecasting with Univariate Box-Jenkins Models*, New York: John Wiley and Sons.

- Paredes, O. C., La Paz, L. M., & Martínez, Á. Q. (2022). *Valor en Riesgo y Correlaciones entre Plazos de la Curva de Rendimientos de la Deuda Pública Interna Venezolana: Un Estudio Para el Periodo 2013-2018*. Cuadernos UCAB, (20).
- Parrado L. J., & Garzón, A. E. (2024). *Análisis de los precios del petróleo derivados de la guerra de Ucrania-Rusia y el impacto en el mercado local e internacional para el primer semestre del 2022*. Tecnología en Gestión del Comercio Internacional.
- Pascual, H. (2015). *Gestión de riesgos, acuerdos de Basilea y crisis: evidencia internacional* (Tesis de pregrado). Universidad Pública de Navarra. Pamplona, España.
- Rivera, M. (2014). *CFIA participa en el cálculo de estadísticas nacionales*. Bitácora. Contenido especializado en ingeniería y arquitectura, Octubre 2014.
- Rodríguez, J. (2013). *Crisis financiera y regulación de la solvencia bancaria: una reflexión crítica sobre los acuerdos de Basilea*. Revista de Economía Crítica, 65-95.
- Rodríguez, O. (2012). *Supervisión bancaria en Costa Rica: un camino difícil*. San José, Costa Rica: Academia de Centroamérica.
- Rojas, D. (2013). *La Implementación de los principios de supervisión bancaria del Comité de Basilea en la República de Colombia*. Revista De La Maestría En Derecho Económico, 11-61.
- Saavedra, M. L., & Saavedra, M. (2010). *Modelos para medir el riesgo de crédito de la banca*. Cuadernos de Administración, 295-319.
- Sandford, A., & Rodríguez, M. (1 de Abril de 2020). *La pandemia de coronavirus es la mayor crisis del mundo desde la II Guerra Mundial, dice la ONU*. Euronews. Recuperado de: <https://es.euronews.com/2020/04/01/la-pandemia-de-coronavirus-es-la-mayor-crisis-del-mundo-desde-la-ii-guerra-mundial-dice-la>
- Sargent, T. (1982). *Teoría macroeconómica*. Boston Academic Press, pp 286-290
- Sharda, R. (1994). *Neural networks for the MS/OR analyst: An application bibliography*. Interfaces, 24(2), 116-130.
- Sheid, F. (1991). *Métodos numéricos*. 2a edición, México, Mc Graw Hill, 709 p.
- Shumway, R. & Stoffer, D. (2011) *Time series Analysis and its applications*. Third Edition. Springer.

- Soley, J. & Rahnema, A. (2005). *Basilea II: una nueva forma de relación banca-empresa*. Segunda Edición. Mc Graw Hill. España.
- SUGEF (2019a). Morosidad mayor a 90 días y cobro judicial / Cartera Directa. [Información Financiera Contable]. Recuperado de: [https://www.sugef.fi.cr/reportes/informacion\\_financiera\\_contable/Indicadores%20Financieros.aspx](https://www.sugef.fi.cr/reportes/informacion_financiera_contable/Indicadores%20Financieros.aspx)
- SUGEF (2019b). 2019 *Bottom Up Stress Test*. Guía metodológica. San José, Costa Rica.
- SUGEF (2019c). *Acuerdo SUGEF 3-06: Reglamento sobre la suficiencia patrimonial de entidades financieras*. San José, Costa Rica.
- SUGEF (2019d). *Acuerdo SUGEF 24-00: Reglamento para juzgar la situación económica financiera de las entidades fiscalizadas*. San José, Costa Rica.
- SUGEF (2020). 2020 *Bottom Up Stress Test*. Guía metodológica. San José, Costa Rica.
- Tang, Z., & Fishwick, P. A. (1993). *Feedforward neural nets as models for time series forecasting*. *ORSA journal on computing*, 5(4), 374-385.
- Toniolo, G. (2005). *Central Bank Cooperation at the Bank for International Settlements, 1930-1973*. Reino Unido: Cambridge University Press.
- Van Gestel, T., & Baesens, B. (2009). *Credit Risk Management Basic Concepts: financial risk components, rating analysis, models, economic and regulatory capital*. New York: Oxford University Press Inc.
- Wallis, K. F. (1987). *Time series analysis of bounded economic variables*. *Journal of Time Series Analysis*, 8(1), 115-123.
- Ware, D. (1996) *Principios básicos de supervisión bancaria*. Centro de Estudios Monetarios Latinoamericanos, México.
- Warman, F. (2013). *Documentos de investigación en países latinoamericanos y efectos de Basilea III*. Centro de Estudios Monetarios Latinoamericanos, Ciudad de México.
- Weber C. (2015) *La Euro Crisis. Causas y síntomas*. *Estudios Fronterizos*, 150-172.
- Wilson, M. (1997) *Portfolio credit risk (I)*, *Risk*, (10-9), 111-17.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist and Wiksell.

- Wong, J., Choi, K., & Fong, T. (2008). *A framework for macro stress testing the credit risk of banks in Hong Kong*. Hong Kong Monetary Authority Quarterly Bulletin.
- Zhang, Fan & O'Donnell, Lauren. (2020). *Support vector regression*. 10.1016/B978-0-12-815739-8.00007-9.
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). *Forecasting with artificial neural networks: The state of the art*. International journal of forecasting, 14(1), 35-62.

## ANEXOS

## Anexo 1

## Metodologías para la clasificación de empresas según tamaño

Institución	Variables	Criterio	Clasificación	Observaciones	Fuente
MAG	Ingresos	$x \leq 155.000 \$$	Micro	Aplica solo para la actividad agrícola.	Gaceta N° 84 <a href="http://www.gaceta.go.cr/pub/2010/05/03/comp_03_05_2010.pdf">http://www.gaceta.go.cr/pub/2010/05/03/comp_03_05_2010.pdf</a>
		$x \leq 540.000 \$$	Pequeña		
		$x \leq 1500.000 \$$	Mediana		
CCSS	N° de trabajadores	$x \leq 5$	Micro	No separan por actividades económicas	La CCSS se basa en la Ley de Fortalecimiento a las Pequeñas y Medianas Empresas, pero adapta los parámetros para definir el tamaño de la empresa según número de trabajadores.
		$5 < x \leq 30$	Pequeña		
		$30 < x \leq 100$	Mediana		
		$x > 100$	Grande		
Procomer	N° de trabajadores	$x < 6$	Micro	No separan por actividades económicas	Sitio web <a href="http://www.procomer.com/uploads/downloads/empleo_exportador_2010.pdf">http://www.procomer.com/uploads/downloads/empleo_exportador_2010.pdf</a>
		$6 \leq x \leq 99$	Pymes		
		$x \geq 100$	Grande		
MEIC	Ingresos y n° de trabajadores	$P = 10$	Micro	<b>Industria:</b> $P = [(0,6 \times pe/100) + (0,3 \times van/\$1.738.000.000) + (0,1 \times afe/\$1.086.000.000)] \times 100$ <b>Comercio y servicios:</b> $P = [(0,6 \times pe/30) + (0,3 \times van/\$2.979.000.000) + (0,1 \times ate/\$931.000.000)] \times 100$ <b>Tecnologías de Información y Comunicación:</b> $P = [(0,6 \times pe/50) + (0,3 \times van/\$2.979.000.000) + (0,1 \times ate/\$931.000.000)] \times 100$	Decreto ejecutivo N°: 30857, 33111, 33747, 35585, 37121 Resolución N°: 0039/2014MEIC, DIGEPYME 029-15
		$10 < P = 35$	Pequeña		
		$35 < P = 100$	Mediana		
Banco Nacional	Ingresos y n° de trabajadores	$x \leq 5$ y $x \leq 175.000 \$$	Micro		Sitio web <a href="https://www.bncr.fi.cr/BNCR/Pymes/SegAtendidos.aspx">https://www.bncr.fi.cr/BNCR/Pymes/SegAtendidos.aspx</a>
		$5 < x \leq 30$ y $x \leq 600.000 \$$	Pequeña		
		$30 < x \leq 100$ y $x \leq 1500.000 \$$	Mediana		

Fuente: Ministerio de Economía, Industria y Comercio (2017).

## Anexo 2

### Estimación del número total de modelos elaborados

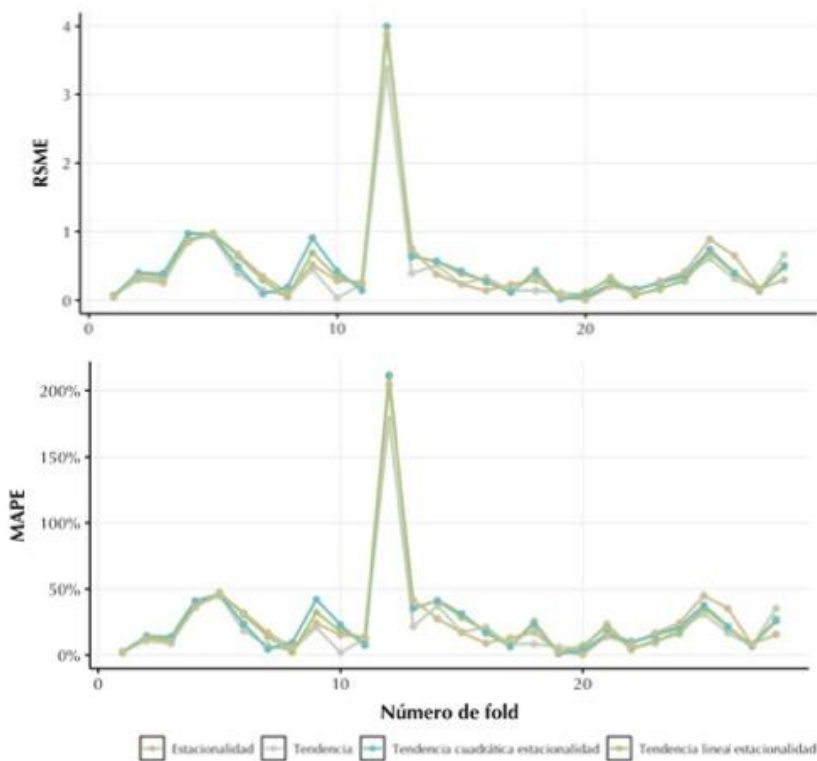
Modelo	Proceso Individual		Cálculo	Repeticiones		
	Hiperparámetros	Combinaciones de variables		Tasa	Segmentos	Grupos en VC <sup>1</sup>
Regresión longitudinal	4	255	60.088	2	4	29
ARMA	36		67.512			
NNAR	50		70.760			
KNN	9		61.248			
SVR	4		60.088			
Bosques Aleatorios	45		69.600			
<b>Total</b>	<b>148</b>	<b>1530</b>	<b>389.296</b>			

<sup>1</sup>VC: validación cruzada

Fuente: Elaboración propia

## Anexo 3

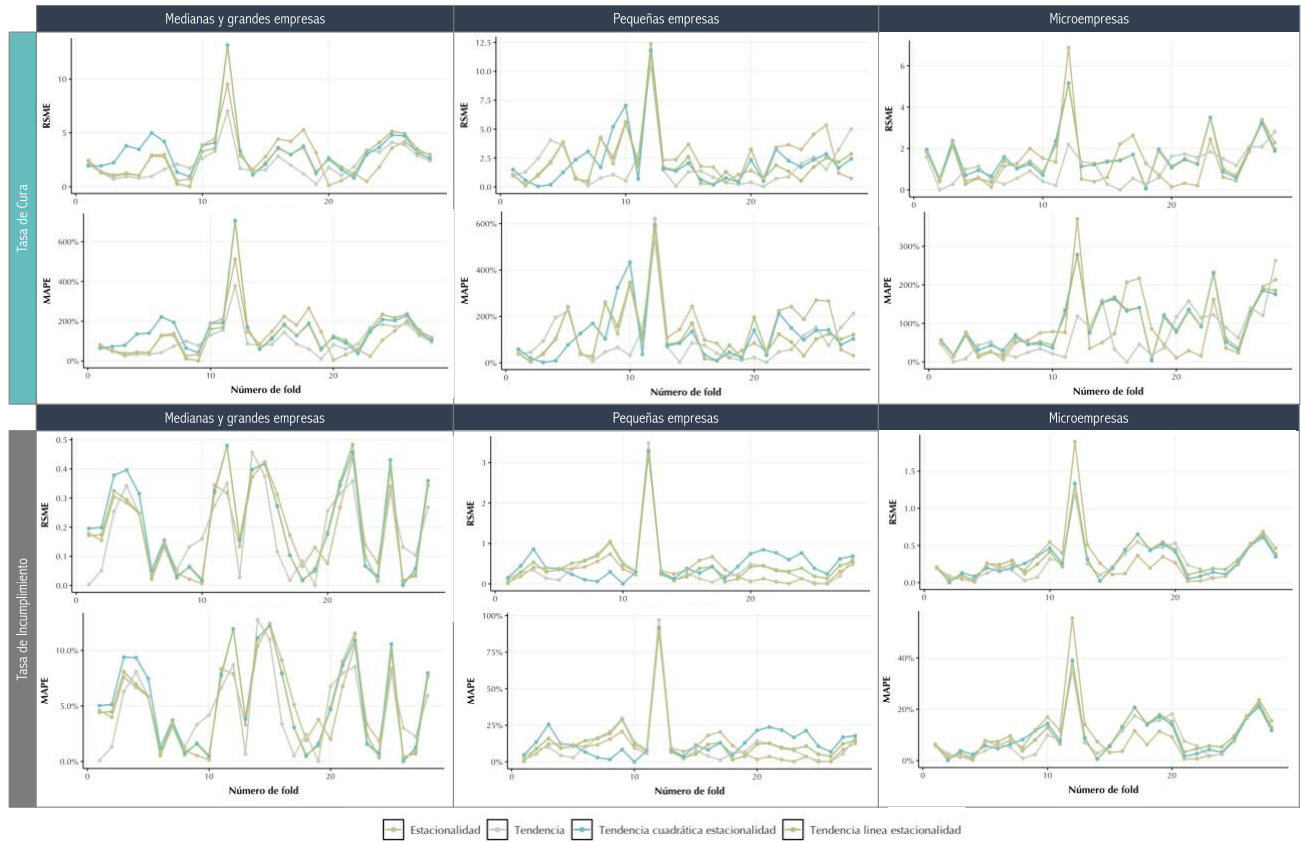
### Banco Improsa, Cartera Total: Estimación del error mediante regresión longitudinal en validación cruzada de la tasa de cura según medida de precisión (k=29)



Fuente: Cartera crediticia Banco Improsa

Anexo 4

**Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante regresión longitudinal en validación cruzada según tipo de tasa por tamaño de empresa (k=29)**



Fuente: Elaboración propia

### Anexo 5

#### Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo de regresión longitudinal

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Tendencia</b>	<b>1.8054</b>	<b>1.3760</b>	<b>84.33%</b>	<b>72.960%</b>
Cura	Tendencia lineal estacionalidad	2.8132	2.4052	125.64%	129.340%
Cura	Estacionalidad	2.8246	1.9827	124.63%	103.950%
Cura	Tendencia cuadrática estacionalidad	3.0613	2.2865	137.23%	123.150%
<b>Incumplimineto</b>	<b>Tendencia</b>	<b>0.1384</b>	<b>0.1357</b>	<b>3.41%</b>	<b>3.540%</b>
Incumplimineto	Estacionalidad	0.1635	0.1362	4.28%	3.540%
Incumplimineto	Tendencia lineal estacionalidad	0.1728	0.1575	4.45%	4.020%
Incumplimineto	Tendencia cuadrática estacionalidad	0.1859	0.1619	4.85%	4.100%

### Anexo 6

#### Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo de regresión longitudinal

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Tendencia</b>	<b>1.1835</b>	<b>2.1448</b>	<b>62.22%</b>	<b>106.6900%</b>
Cura	Tendencia cuadrática estacionalidad	1.6700	2.4274	91.67%	132.6400%
Cura	Tendencia lineal estacionalidad	1.7141	2.2941	95.71%	123.0200%
Cura	Estacionalidad	1.8999	2.4571	99.72%	133.6700%
<b>Incumplimineto</b>	<b>Tendencia</b>	<b>0.2795</b>	<b>0.6455</b>	<b>7.91%</b>	<b>17.9400%</b>
Incumplimineto	Estacionalidad	0.3321	0.5850	9.45%	16.3600%
Incumplimineto	Tendencia lineal estacionalidad	0.3664	0.5946	10.74%	16.5500%
Incumplimineto	Tendencia cuadrática estacionalidad	0.3822	0.5945	11.46%	16.5600%

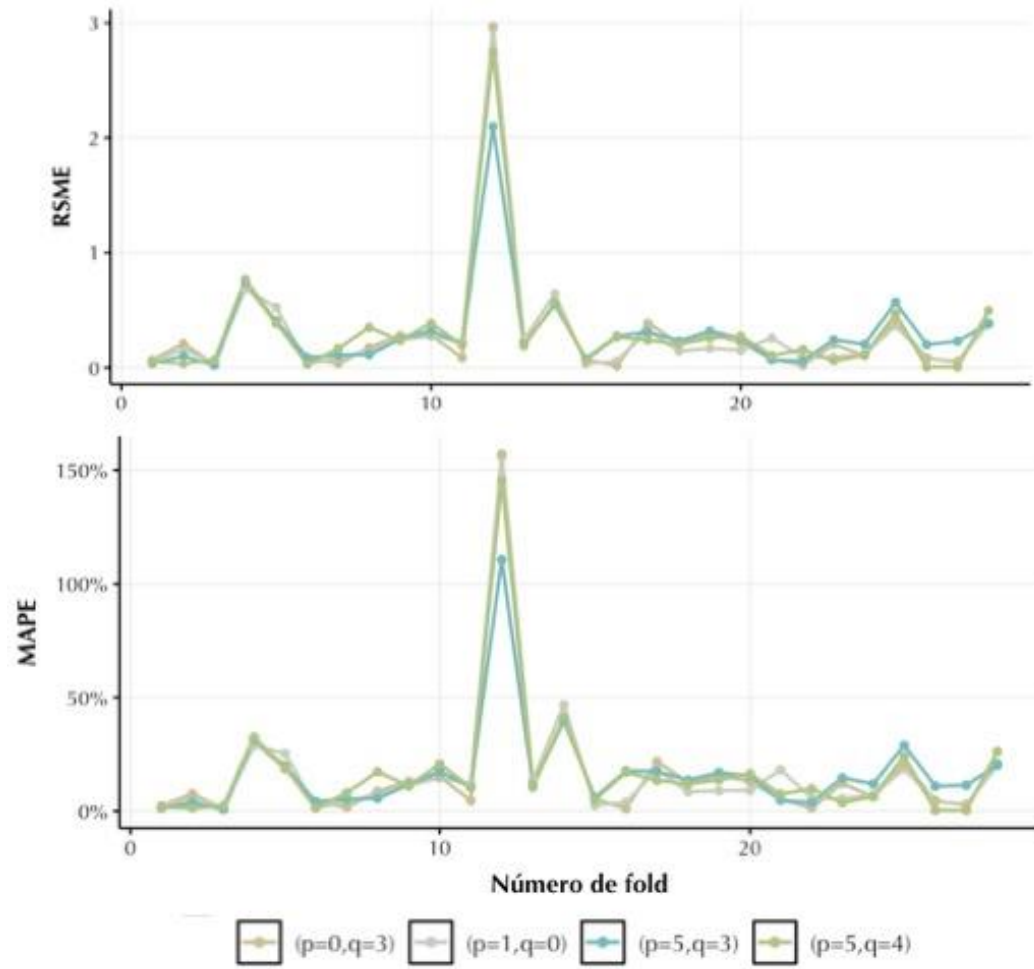
### Anexo 7

#### Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo de regresión longitudinal

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Tendencia</b>	<b>1.0675</b>	<b>0.7627</b>	<b>45.58%</b>	<b>62.720%</b>
Cura	Estacionalidad	1.2089	1.3896	53.81%	85.470%
Cura	Tendencia cuadrática estacionalidad	1.3006	1.0487	78.47%	68.180%
Cura	Tendencia lineal estacionalidad	1.3868	1.0740	76.59%	70.540%
<b>Incumplimineto</b>	<b>Tendencia</b>	<b>0.2206</b>	<b>0.2501</b>	<b>6.65%</b>	<b>7.790%</b>
Incumplimineto	Estacionalidad	0.2451	0.3593	7.49%	10.730%
Incumplimineto	Tendencia lineal estacionalidad	0.2508	0.2672	7.76%	8.230%
Incumplimineto	Tendencia cuadrática estacionalidad	0.2598	0.2674	7.98%	8.210%

## Anexo 8

**Banco Improsa, Cartera Total: Estimación del error mediante modelo ARMA en validación cruzada de la tasa de cura según medida de precisión (k=29)**



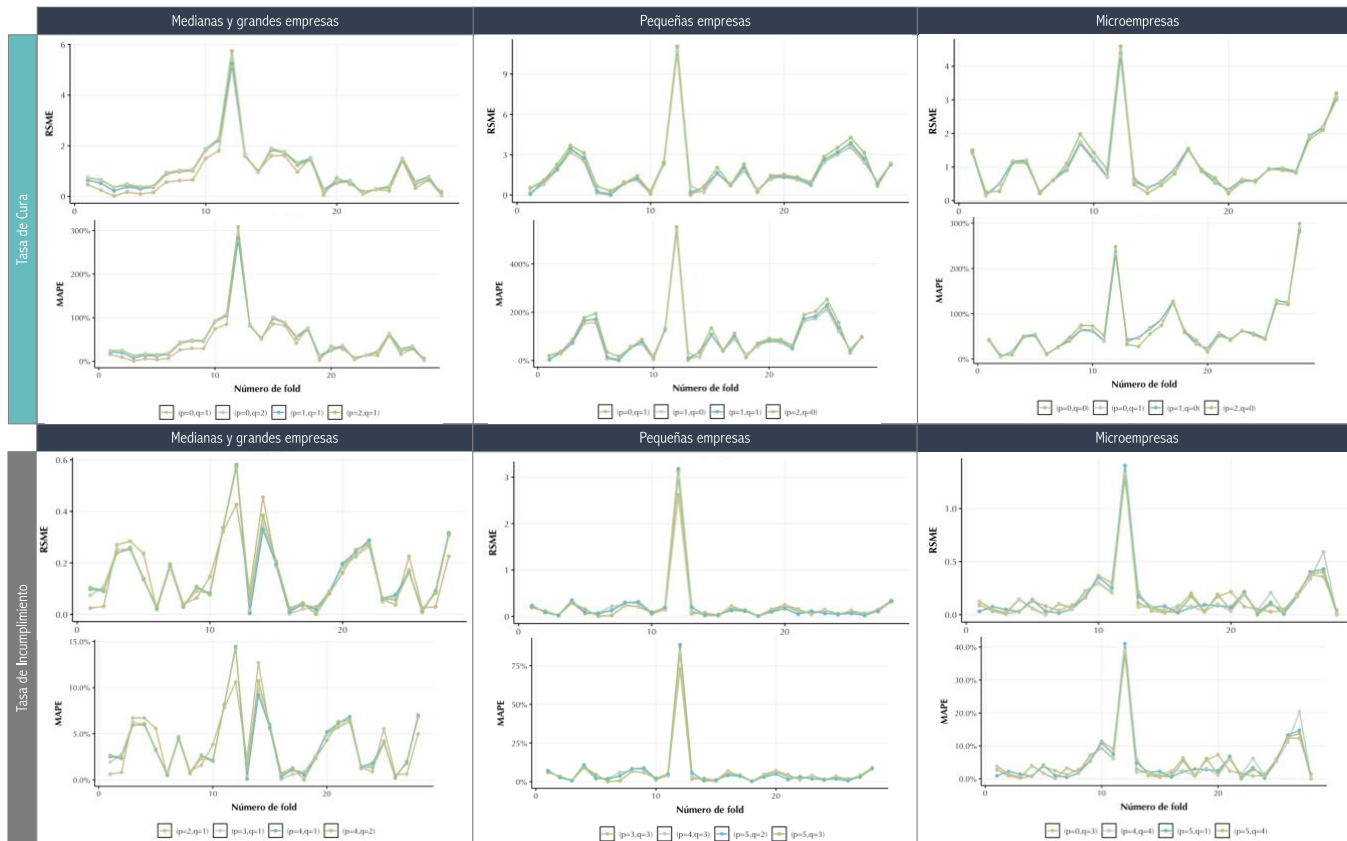
Fuente: Cartera crediticia Banco Improsa

Anexo 9

Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante modelo

ARMA en validación cruzada según tipo de tasa por tamaño de empresa

(k=29)



Fuente: Cartera crediticia Banco Improsa

### Anexo 10

#### Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo ARMA

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(p=0,q=1)</b>	<b>0.5994</b>	<b>1.0185</b>	<b>28.140%</b>	<b>54.310%</b>
Cura	(p=2,q=1)	0.7089	1.1001	31.830%	59.000%
Cura	(p=1,q=1)	0.7135	1.0221	35.410%	54.630%
Cura	(p=0,q=2)	0.7613	1.0562	35.460%	56.610%
<b>Incumplimineto</b>	<b>(p=3,q=1)</b>	<b>0.0992</b>	<b>0.1349</b>	<b>2.690%</b>	<b>3.370%</b>
Incumplimineto	(p=4,q=1)	0.1026	0.1342	2.600%	3.340%
Incumplimineto	(p=4,q=2)	0.1036	0.1343	2.630%	3.370%
Incumplimineto	(p=2,q=1)	0.1195	0.1284	3.150%	3.280%

### Anexo 11

#### Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo ARMA

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(p=1,q=0)</b>	<b>1.1537</b>	<b>2.0190</b>	<b>71.870%</b>	<b>103.85%</b>
Cura	(p=0,q=1)	1.2051	2.0341	70.180%	104.74%
Cura	(p=1,q=1)	1.3117	2.1101	76.020%	108.73%
Cura	(p=2,q=0)	1.4307	2.1255	87.120%	110.43%
<b>Incumplimineto</b>	<b>(p=5,q=2)</b>	<b>0.1064</b>	<b>0.5850</b>	<b>3.080%</b>	<b>16.31%</b>
Incumplimineto	(p=5,q=3)	0.1139	0.5736	3.340%	15.99%
Incumplimineto	(p=4,q=3)	0.1180	0.5492	3.300%	15.31%
Incumplimineto	(p=3,q=3)	0.1319	0.4778	3.700%	13.32%

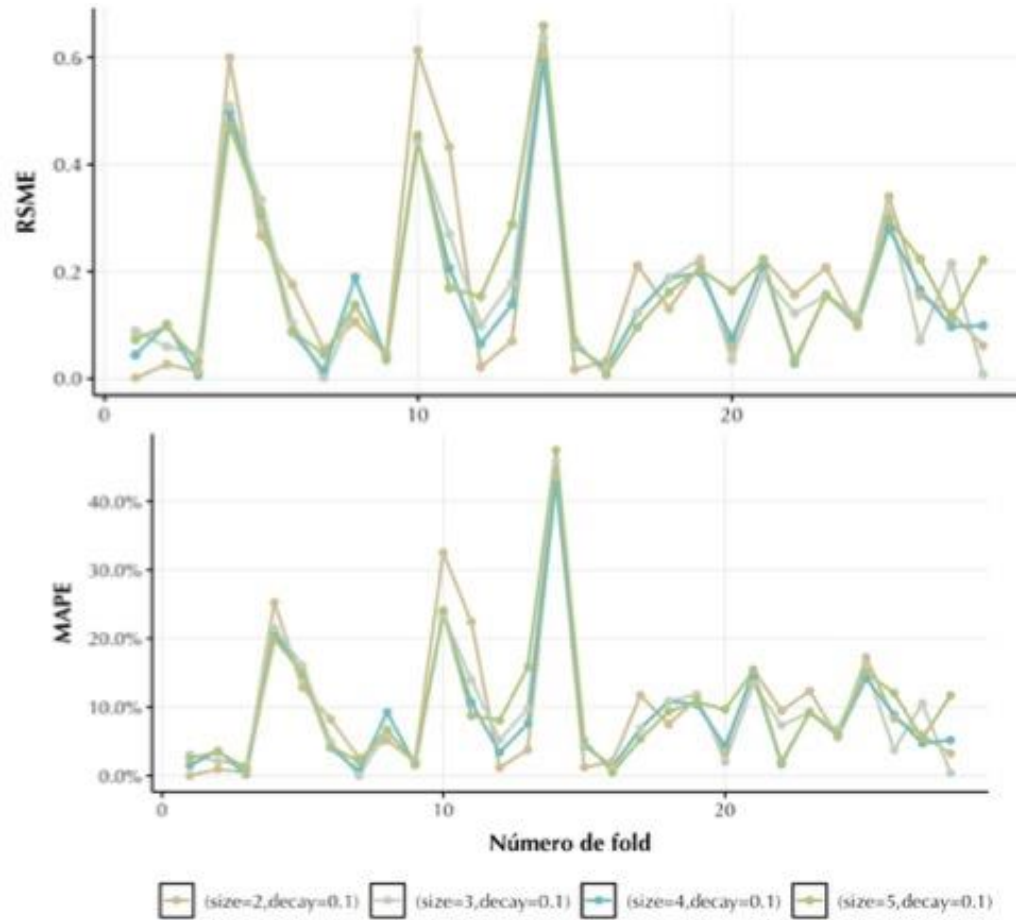
### Anexo 12

#### Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo ARMA

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(p=0,q=0)</b>	<b>0.8871</b>	<b>0.9055</b>	<b>51.58%</b>	<b>62.34%</b>
Cura	(p=2,q=0)	0.9153	0.9618	50.79%	65.75%
Cura	(p=1,q=0)	0.9228	0.9009	52.42%	62.69%
Cura	(p=0,q=1)	0.9302	0.9008	52.82%	62.94%
<b>Incumplimineto</b>	<b>(p=5,q=1)</b>	<b>0.0778</b>	<b>0.2691</b>	<b>2.37%</b>	<b>8.01%</b>
Incumplimineto	(p=4,q=4)	0.0814	0.2601	2.50%	7.84%
Incumplimineto	(p=0,q=3)	0.0853	0.2469	2.62%	7.34%
Incumplimineto	(p=5,q=4)	0.0991	0.2537	2.94%	7.56%

## Anexo 13

**Banco Improsa, Cartera Total: Estimación del error mediante modelo NNAR en validación cruzada de la tasa de cura según medida de precisión (k=29)**



Fuente: Cartera crediticia Banco Improsa

Anexo 14

**Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante modelo NNAR en validación cruzada según tipo de tasa por tamaño de empresa (k=29)**



Fuente: Cartera crediticia Banco Improsa

## Anexo 15

**Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo NNAR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(size=1,decay=0.05)</b>	<b>0.3764</b>	<b>1.2478</b>	<b>17.710%</b>	<b>48.420%</b>
Cura	(size=1,decay=0.1)	0.5138	1.1640	24.720%	48.140%
Cura	(size=1,decay=2)	0.7314	0.5577	32.030%	29.200%
Cura	(size=2,decay=2)	0.7928	0.5954	33.700%	31.110%
<b>Incumplimineto</b>	<b>(size=7,decay=0.1)</b>	<b>0.1184</b>	<b>0.1277</b>	<b>2.770%</b>	<b>3.370%</b>
Incumplimineto	(size=8,decay=0.1)	0.1202	0.1278	2.820%	3.370%
Incumplimineto	(size=3,decay=0.1)	0.1210	0.1250	2.820%	3.300%
Incumplimineto	(size=9,decay=0.1)	0.1236	0.1285	2.900%	3.390%

## Anexo 16

**Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo NNAR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(size=10,decay=1)</b>	<b>0.6654</b>	<b>0.7923</b>	<b>36.350%</b>	<b>41.830%</b>
Cura	(size=9,decay=1)	0.6700	0.7920	36.600%	41.790%
Cura	(size=8,decay=1)	0.6770	0.7903	37.380%	41.690%
Cura	(size=7,decay=1)	0.6814	0.7940	38.510%	41.770%
<b>Incumplimineto</b>	<b>(size=9,decay=2)</b>	<b>0.0666</b>	<b>0.0877</b>	<b>1.960%</b>	<b>2.400%</b>
Incumplimineto	(size=8,decay=2)	0.0704	0.0877	2.070%	2.400%
Incumplimineto	(size=7,decay=2)	0.0729	0.0877	2.140%	2.400%
Incumplimineto	(size=6,decay=2)	0.0771	0.0877	2.270%	2.400%

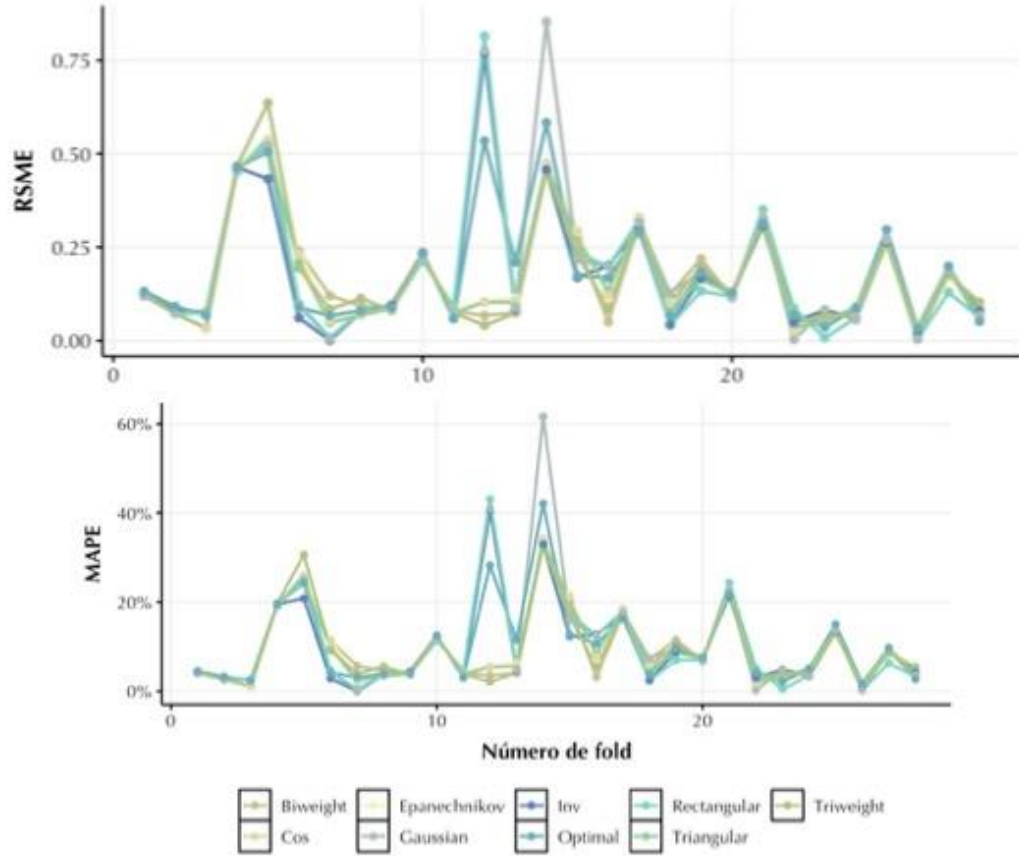
## Anexo 17

**Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo NNAR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(size=2,decay=1)</b>	<b>0.4832</b>	<b>0.4412</b>	<b>28.720%</b>	<b>34.140%</b>
Cura	(size=3,decay=1)	0.5348	0.4379	30.400%	33.360%
Cura	(size=1,decay=1)	0.5512	0.4113	27.800%	38.090%
Cura	(size=3,decay=2)	0.5969	0.4467	31.400%	45.450%
<b>Incumplimineto</b>	<b>(size=6,decay=0.05)</b>	<b>0.1171</b>	<b>0.0853</b>	<b>3.540%</b>	<b>2.750%</b>
Incumplimineto	(size=5,decay=0.05)	0.1193	0.0840	3.660%	2.720%
Incumplimineto	(size=4,decay=0.05)	0.1260	0.0864	3.790%	2.790%
Incumplimineto	(size=10,decay=1)	0.1298	0.0866	3.770%	2.790%

Anexo 18

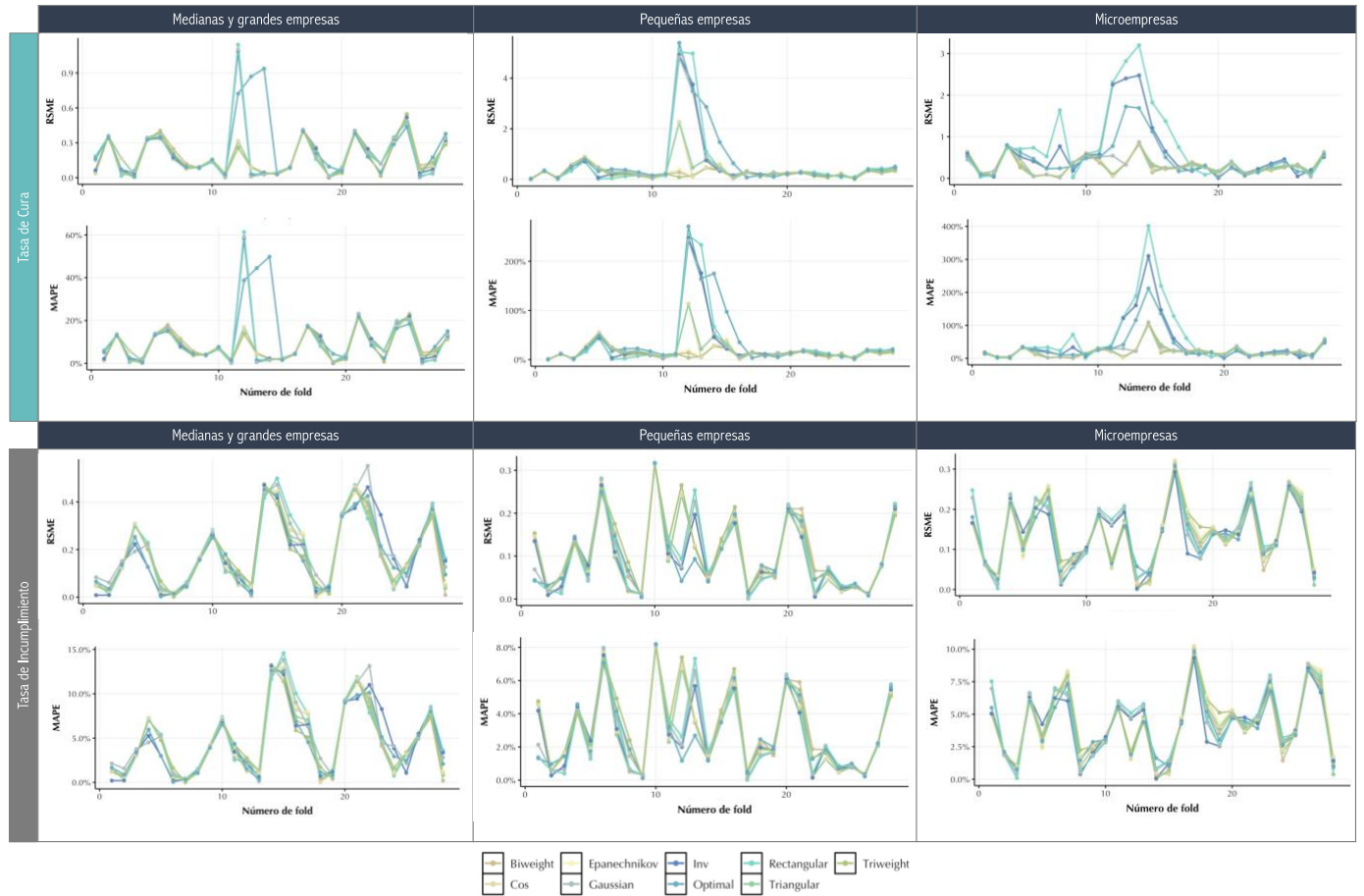
**Banco Improsa, Cartera Total: Estimación del error mediante modelo KNN en validación cruzada de la tasa de cura según medida de precisión (k=29)**



Fuente: Cartera crediticia Banco Improsa

Anexo 19

**Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante modelo KNN en validación cruzada según tipo de tasa por tamaño de empresa (k=29)**



Fuente: Cartera crediticia Banco Improsa

## Anexo 20

**Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo KNN**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Inv</b>	<b>0.1051</b>	<b>0.2274</b>	<b>5.240%</b>	<b>11.7500%</b>
Cura	Triweight	0.1142	0.1508	5.280%	6.9200%
Cura	Biweight	0.1238	0.1496	5.770%	6.9100%
Cura	Rectangular	0.1269	0.2356	5.960%	12.2000%
Cura	Gaussian	0.1291	0.2296	5.430%	11.8600%
Cura	Epanechnikov	0.1337	0.1452	5.770%	6.7800%
Cura	Cos	0.1347	0.1449	5.830%	6.7600%
Cura	Triangular	0.1365	0.1451	5.860%	6.7700%
Cura	Optimal	0.1775	0.2468	8.050%	12.9000%
<b>Incumplimineto</b>	<b>Optimal</b>	<b>0.1435</b>	<b>0.1413</b>	<b>3.620%</b>	<b>3.7300%</b>
Incumplimineto	Epanechnikov	0.1518	0.1535	3.740%	4.0900%
Incumplimineto	Cos	0.1540	0.1538	3.790%	4.0900%
Incumplimineto	Inv	0.1552	0.1493	3.650%	3.9000%
Incumplimineto	Biweight	0.1569	0.1463	3.860%	3.8700%
Incumplimineto	Triangular	0.1580	0.1488	3.890%	3.9400%
Incumplimineto	Rectangular	0.1655	0.1498	3.910%	4.0400%
Incumplimineto	Triweight	0.1662	0.1392	4.140%	3.6700%
Incumplimineto	Gaussian	0.1680	0.1525	4.090%	4.0200%

## Anexo 21

**Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo KNN**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Triweight</b>	<b>0.2074</b>	<b>0.1891</b>	<b>11.530%</b>	<b>11.0800%</b>
Cura	Inv	0.2152	1.0971	11.510%	53.7000%
Cura	Biweight	0.2176	0.1891	11.540%	11.1200%
Cura	Rectangular	0.2230	1.2696	12.270%	61.7800%
Cura	Epanechnikov	0.2463	0.2051	13.190%	12.2400%
Cura	Cos	0.2498	0.4269	13.070%	22.1400%
Cura	Triangular	0.2589	0.4289	13.630%	22.3700%
Cura	Gaussian	0.2903	1.0519	15.040%	51.6900%
Cura	Optimal	0.2951	1.2250	17.000%	63.3000%
<b>Incumplimineto</b>	<b>Optimal</b>	<b>0.0661</b>	<b>0.0807</b>	<b>1.930%</b>	<b>2.2200%</b>
Incumplimineto	Rectangular	0.0674	0.0921	1.910%	2.5500%
Incumplimineto	Epanechnikov	0.0703	0.0870	2.010%	2.4100%
Incumplimineto	Gaussian	0.0705	0.0887	2.060%	2.4500%
Incumplimineto	Cos	0.0721	0.0866	2.060%	2.4000%
Incumplimineto	Biweight	0.0750	0.0874	2.180%	2.4200%
Incumplimineto	Inv	0.0751	0.0839	2.080%	2.3200%
Incumplimineto	Triangular	0.0759	0.0860	2.170%	2.3800%
Incumplimineto	Triweight	0.0837	0.0875	2.430%	2.4300%

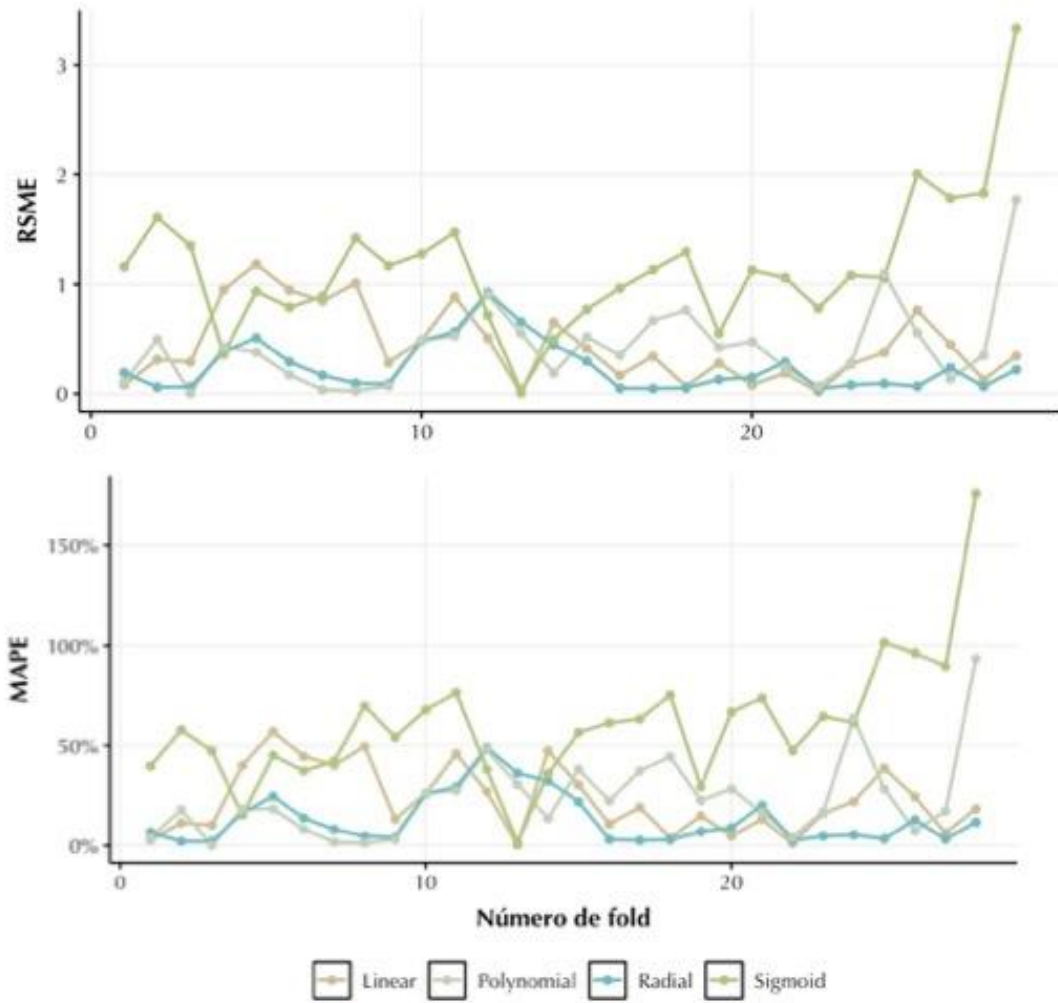
## Anexo 22

**Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo KNN**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Triangular</b>	<b>0.2549</b>	<b>0.2217</b>	<b>16.0600%</b>	<b>20.7700%</b>
Cura	Cos	0.2572	0.2232	16.9800%	21.0300%
Cura	Epanechnikov	0.2582	0.2211	17.2000%	21.0000%
Cura	Triweight	0.2612	0.2182	15.5800%	20.7200%
Cura	Biweight	0.2708	0.2238	15.9500%	20.8300%
Cura	Gaussian	0.2936	0.2099	18.7200%	20.9400%
Cura	Optimal	0.3069	0.4349	18.5100%	46.6900%
Cura	Inv	0.3812	0.6836	21.5500%	66.6600%
Cura	Rectangular	0.5326	0.8559	26.5000%	87.5100%
<b>Incumplimineto</b>	<b>Optimal</b>	<b>0.1306</b>	<b>0.0730</b>	<b>4.1600%</b>	<b>2.3300%</b>
Incumplimineto	Triangular	0.1335	0.0830	4.2300%	2.6500%
Incumplimineto	Cos	0.1366	0.0830	4.3600%	2.6600%
Incumplimineto	Biweight	0.1381	0.0796	4.3700%	2.5200%
Incumplimineto	Epanechnikov	0.1389	0.0841	4.4100%	2.7000%
Incumplimineto	Gaussian	0.1392	0.0824	4.3900%	2.5800%
Incumplimineto	Inv	0.1410	0.0776	4.3100%	2.4200%
Incumplimineto	Rectangular	0.1464	0.0860	4.6000%	2.6900%
Incumplimineto	Triweight	0.1488	0.0786	4.5400%	2.5100%

Anexo 23

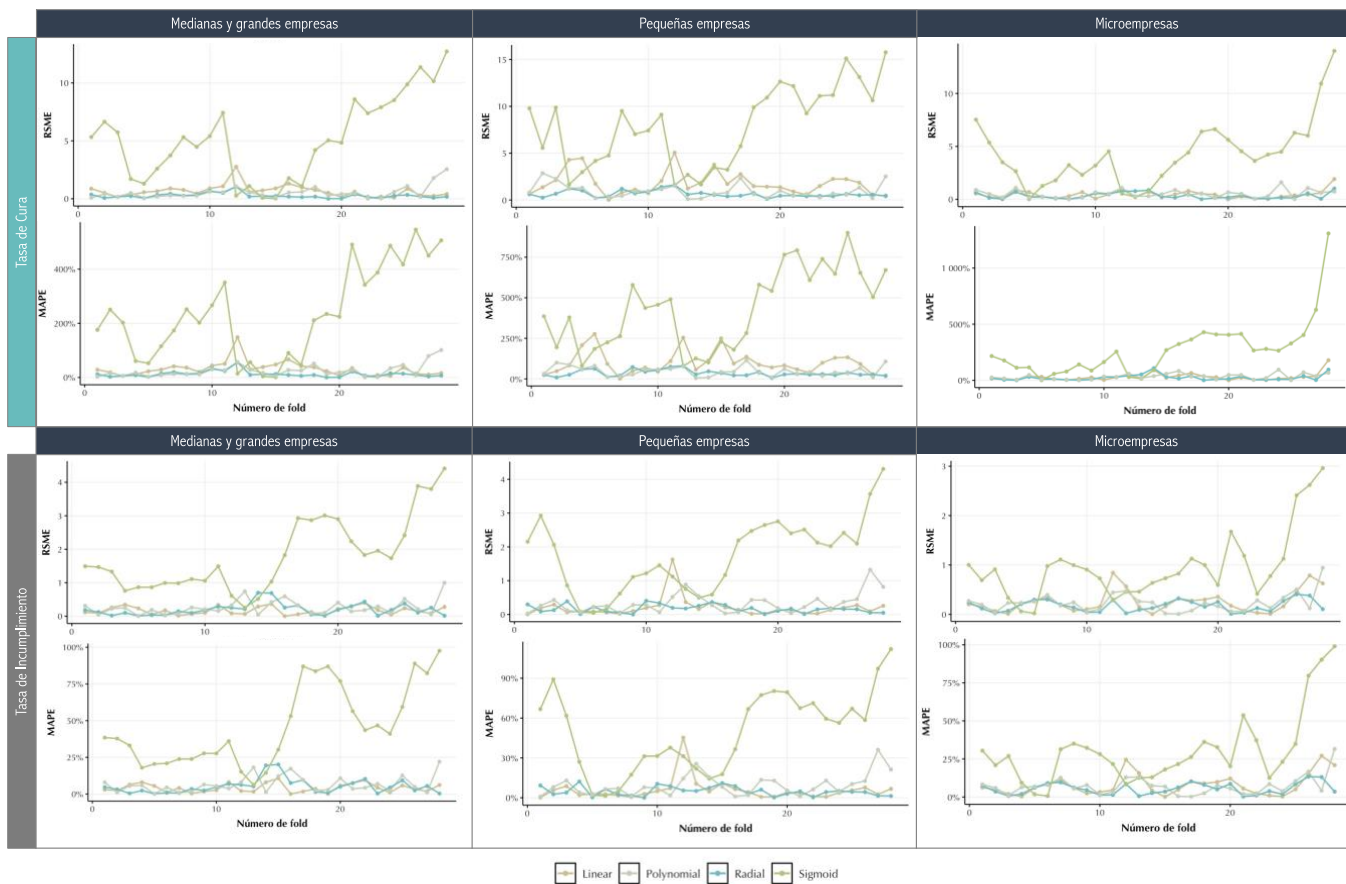
**Banco Improsa, Cartera Total: Estimación del error mediante modelo SVR en validación cruzada de la tasa de cura según medida de precisión (k=29)**



Fuente: Cartera crediticia Banco Improsa

## Anexo 24

**Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante modelo SVR  
en validación cruzada según tipo de tasa por tamaño de empresa  
( $k=29$ )**



Fuente: Cartera crediticia Banco Improsa

## Anexo 25

**Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo SVR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Radial</b>	<b>0.2067</b>	<b>0.2153</b>	<b>9.52%</b>	<b>11.34%</b>
Cura	Polynomial	0.4151	0.5539	17.36%	23.76%
Cura	Linear	0.5959	0.5243	28.85%	28.00%
Cura	Sigmoid	5.1767	3.5661	217.77%	167.89%
<b>Incumplimineto</b>	<b>Linear</b>	<b>0.1331</b>	<b>0.1187</b>	<b>3.43%</b>	<b>3.01%</b>
Incumplimineto	Radial	0.1830	0.1827	4.58%	5.10%
Incumplimineto	Polynomial	0.2106	0.2291	5.16%	5.62%
Incumplimineto	Sigmoid	1.4911	1.0987	38.12%	27.09%

## Anexo 26

**Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo SVR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Radial</b>	<b>0.5582</b>	<b>0.3585</b>	<b>27.53%</b>	<b>20.05%</b>
Cura	Polynomial	0.7368	0.7582	42.41%	31.19%
Cura	Linear	1.4837	1.2630	84.77%	71.50%
Cura	Sigmoid	9.1791	4.2288	446.93%	241.63%
<b>Incumplimineto</b>	<b>Linear</b>	<b>0.1356</b>	<b>0.2979</b>	<b>3.86%</b>	<b>8.32%</b>
Incumplimineto	Radial	0.1575	0.1191	4.38%	3.47%
Incumplimineto	Polynomial	0.2694	0.2966	7.64%	8.17%
Incumplimineto	Sigmoid	2.0423	1.0793	57.47%	30.34%

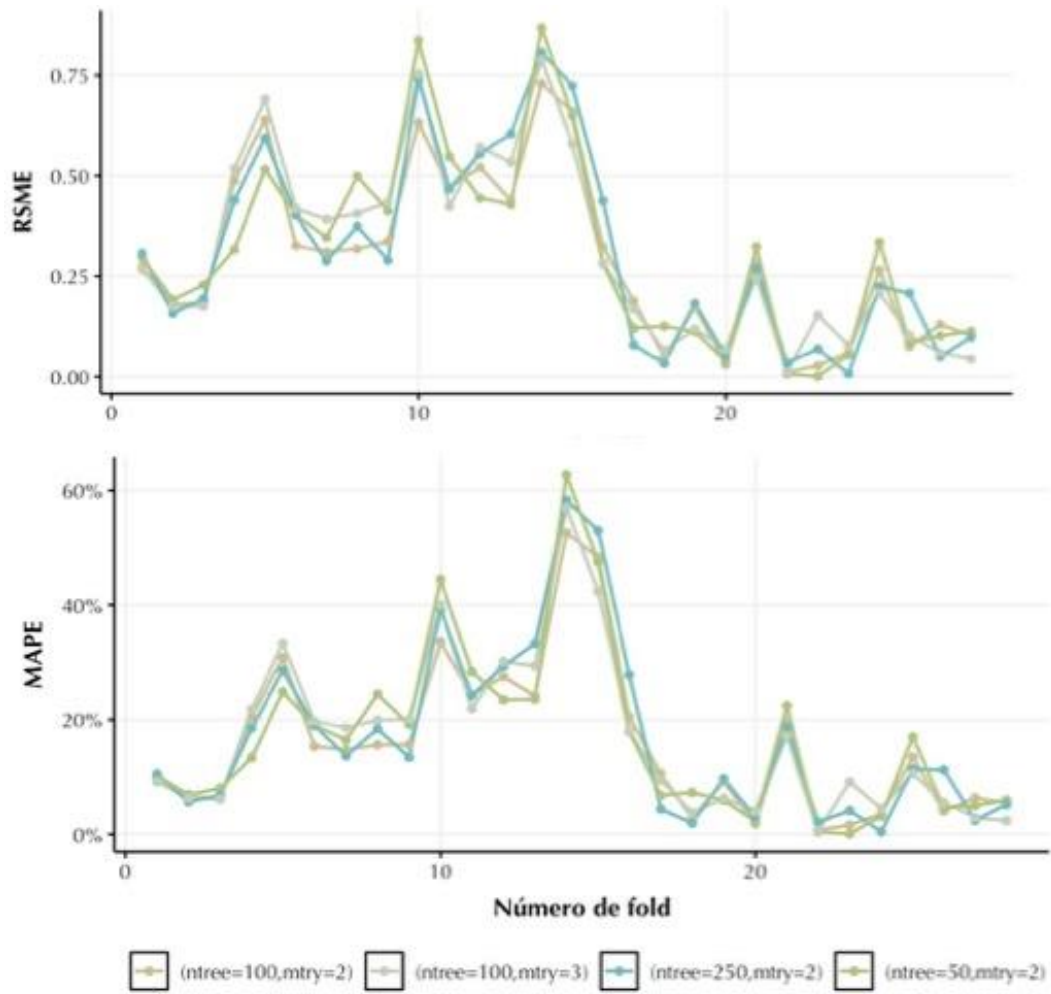
## Anexo 27

**Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo SVR**

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>Radial</b>	<b>0.2287</b>	<b>0.2922</b>	<b>12.39%</b>	<b>26.40%</b>
Cura	Linear	0.4132	0.3901	23.36%	35.74%
Cura	Polynomial	0.4937	0.3914	31.56%	26.19%
Cura	Sigmoid	3.9294	3.1075	260.40%	255.00%
<b>Incumplimineto</b>	<b>Radial</b>	<b>0.1460</b>	<b>0.1161</b>	<b>4.76%</b>	<b>3.81%</b>
Incumplimineto	Linear	0.1814	0.2233	5.77%	7.14%
Incumplimineto	Polynomial	0.2300	0.1949	6.65%	6.33%
Incumplimineto	Sigmoid	0.8662	0.7020	26.62%	23.99%

Anexo 28

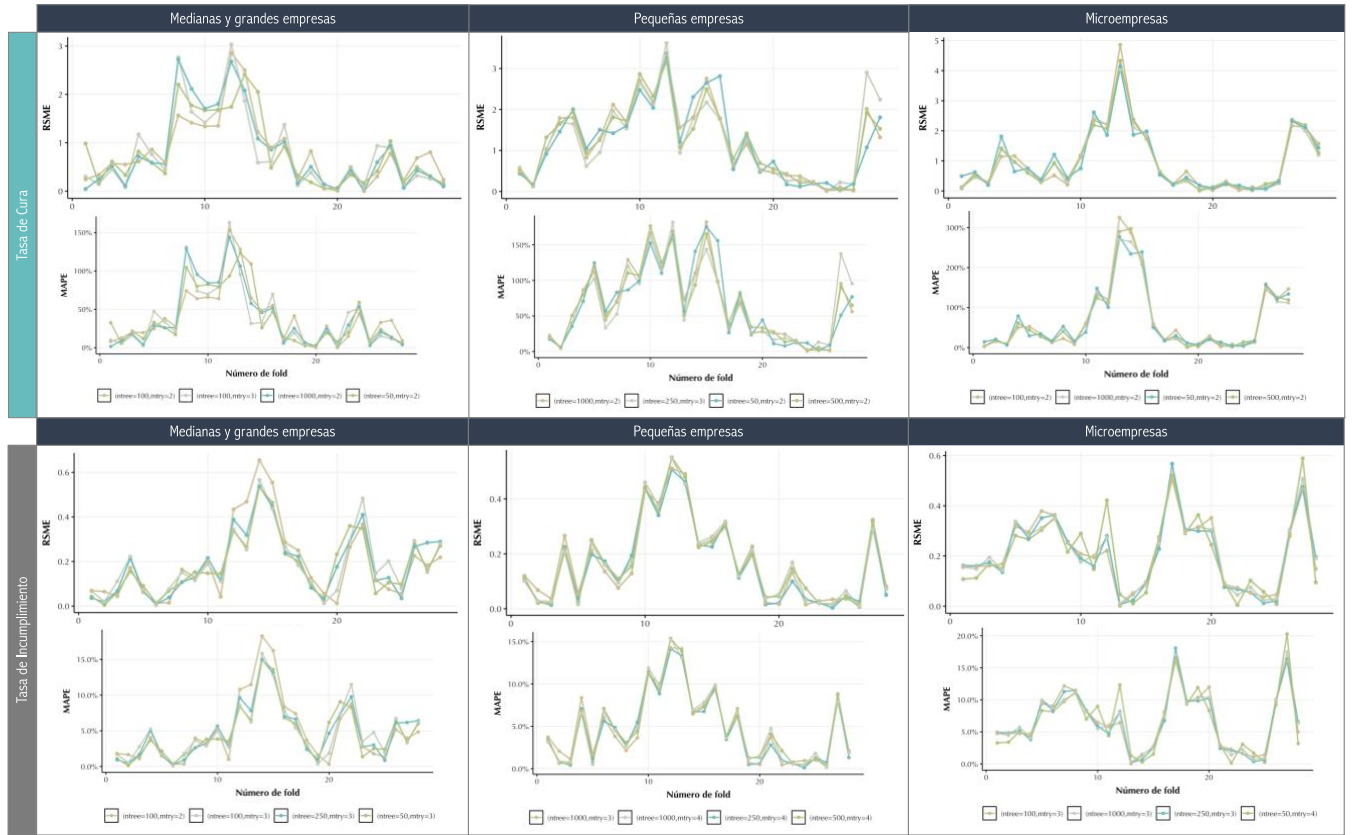
Banco Improsa, Cartera Total: Estimación del error mediante bosques aleatorios en validación cruzada de la tasa de cura según medida de precisión (k=29)



Fuente: Cartera crediticia Banco Improsa

Anexo 29

**Banco Improsa, Cartera Total: Estimación de medidas de rendimiento mediante bosques aleatorios en validación cruzada según tipo de tasa por tamaño de empresa (k=29)**



Fuente: Cartera crediticia Banco Improsa

### Anexo 30

#### Banco Improsa, Medianas y grandes empresas: Resumen de medidas de precisión de variantes del modelo bosques aleatorios

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(ntree=100,mtry=3)</b>	<b>0.4734</b>	<b>0.8074</b>	<b>24.4200%</b>	<b>40.980%</b>
Cura	(ntree=50,mtry=2)	0.4909	0.7259	22.7300%	36.920%
Cura	(ntree=1000,mtry=2)	0.5406	0.8127	25.6000%	40.870%
Cura	(ntree=100,mtry=2)	0.6469	0.6876	30.4900%	36.270%
<b>Incumplimineto</b>	<b>(ntree=100,mtry=2)</b>	<b>0.1469</b>	<b>0.1702</b>	<b>3.8400%</b>	<b>4.670%</b>
Incumplimineto	(ntree=50,mtry=3)	0.1501	0.1366	3.6000%	3.730%
Incumplimineto	(ntree=100,mtry=3)	0.1520	0.1422	3.5900%	3.820%
Incumplimineto	(ntree=250,mtry=3)	0.1526	0.1432	3.9500%	3.820%

### Anexo 31

#### Banco Improsa, Pequeñas empresas: Resumen de medidas de precisión de variantes del modelo bosques aleatorios

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(ntree=250,mtry=3)</b>	<b>0.9960</b>	<b>0.9645</b>	<b>48.440%</b>	<b>53.0000%</b>
Cura	(ntree=50,mtry=2)	1.1419	0.9413	56.670%	55.0500%
Cura	(ntree=1000,mtry=2)	1.2192	0.9215	61.860%	53.7300%
Cura	(ntree=500,mtry=2)	1.2853	0.8869	57.930%	51.6800%
<b>Incumplimineto</b>	<b>(ntree=250,mtry=4)</b>	<b>0.1101</b>	<b>0.1487</b>	<b>3.390%</b>	<b>4.1500%</b>
Incumplimineto	(ntree=1000,mtry=3)	0.1262	0.1507	3.730%	4.2200%
Incumplimineto	(ntree=500,mtry=4)	0.1366	0.1520	3.990%	4.2300%
Incumplimineto	(ntree=1000,mtry=4)	0.1388	0.1526	4.040%	4.2500%

### Anexo 32

#### Banco Improsa, Microempresas: Resumen de medidas de precisión de variantes del modelo bosques aleatorios

Tasa	Modelo	Mediana RSME	Desviación Estándar RSME	Mediana MAPE	Desviación Estándar MAPE
<b>Cura</b>	<b>(ntree=1000,mtry=2)</b>	<b>0.5307</b>	<b>0.9991</b>	<b>28.400%</b>	<b>79.950%</b>
Cura	(ntree=500,mtry=2)	0.5627	1.0261	26.930%	83.490%
Cura	(ntree=50,mtry=2)	0.5810	0.9984	29.510%	79.030%
Cura	(ntree=100,mtry=2)	0.5906	1.0987	28.300%	87.220%
<b>Incumplimineto</b>	<b>(ntree=250,mtry=3)</b>	<b>0.1931</b>	<b>0.1420</b>	<b>6.230%</b>	<b>4.660%</b>
Incumplimineto	(ntree=50,mtry=4)	0.1931	0.1541	6.020%	5.050%
Incumplimineto	(ntree=1000,mtry=3)	0.1973	0.1372	6.170%	4.530%
Incumplimineto	(ntree=100,mtry=3)	0.2016	0.1343	6.130%	4.440%

## Anexo 33

```

suppressMessages(library(ggplot2))
suppressMessages(library(tidyverse))
suppressMessages(library(readxl))
suppressMessages(library(scales))
suppressMessages(library(knitr))
suppressMessages(library(kableExtra))
suppressMessages(library(ggcorrplot))
suppressMessages(library(tseries))
suppressMessages(library(timeSeries))
suppressMessages(library(forecast))
suppressMessages(library(TTR))
suppressMessages(library(lubridate))
suppressMessages(library(plotly))
suppressMessages(library(smooth))
suppressMessages(library(astsa))
suppressMessages(library(fpp))
suppressMessages(library(fpp2))
suppressMessages(library(VGAMdata))
suppressMessages(library(GGally))
suppressMessages(library(seasonal))
suppressMessages(library(fpp2))
suppressMessages(library(fpp2))
suppressMessages(library(scales))
suppressMessages(library(GGally))
suppressMessages(library(kknn))
suppressMessages(library(e1071))

datos<-read_excel("~/Documents/TesisBUSTinterno/Analisis/Datos/datos_completo_bust_v4.xlsx")
str(datos)
datos$mes <- as.Date(datos$mes)
datos <- datos %>% filter(mes>="2014-06-01")
datos.mod <- datos %>%
  select(starts_with("var"))
seleccion.colores<-c("#CBBF92", "#F7F3BE", "#5F87BA", "#76D7C4", "#B1C47D")
colores<-colorRampPalette(seleccion.colores)
logistic <- function(p) log(p / (1-p) )
logistic.inv<- function(y) exp(y) / (1+exp(y))
theme_Publication <- function(base_size=12, base_family="helvetica") {
  library(grid)
  library(ggthemes)
  (theme_foundation(base_size=base_size, base_family=base_family)
   + theme(plot.title = element_text(face = "bold",
                                     size = rel(1.2), hjust = 0.5),
           plot.subtitle = element_text(hjust = 0.5),
           text = element_text(family="Optima"),
           panel.background = element_rect(fill=NA,color = NA),
           plot.background = element_rect(fill=NA,color = NA),
           panel.border =element_rect(fill=NA,color=NA),

```

```

axis.title = element_text(face = "bold",size = rel(1)),
axis.title.y = element_text(angle=90,vjust =2),
axis.title.x = element_text(vjust = -0.2),
axis.text = element_text(),
axis.line = element_line(colour="black"),
axis.ticks = element_line(),
panel.grid.major = element_line(colour="#f0f0f0"),
panel.grid.minor = element_blank(),
plot.margin=unit(c(10,5,5,5),"mm"),
strip.background=element_rect(colour="#f0f0f0",fill="#f0f0f0"),
strip.text = element_text(face="bold")
))
}
grafico.rsme <- function(data){
  niv <- length(levels(data$validacion))
  ggplot(data)+
    geom_line(aes(x=iteracion,y=rsme,group=validacion,color=validacion),
              size=0.8)+
    geom_point(aes(x=iteracion,y=rsme,group=validacion,color=validacion))+
    theme_Publication()+
    ylim (min(data$rsme),
          max(data$rsme) )+
    theme(legend.title = element_blank(),
          legend.position = "bottom"
    )+
    scale_color_manual(values=colores(niv))
}
grafico.mape <- function(data){
  niv <- length(levels(data$validacion))
  ggplot(data)+
    geom_line(aes(x=iteracion,y=mape,group=validacion,color=validacion),
              size=0.8)+
    geom_point(aes(x=iteracion,y=mape,group=validacion,color=validacion))+
    theme_Publication()+
    ylim (min(data$mape),
          max(data$mape) )+
    theme(legend.title = element_blank(),
          legend.position = "bottom"
    )+
    scale_color_manual(values=colores(niv))+
    scale_y_continuous(labels = percent)
}
cuadro.resumen.error <-
  function(res.cura=t_cura_regr,res.incump=t_incump_regr,segmento="Segmento",nom.modelo="Modelo"){

  n.variantes <- length(levels(res.cura$validacion))

  cuadro <- data.frame(Tasa=rep("",2*n.variantes),

```

```

        Modelo=rep("",2*n.variantes),
        m.rsme=rep(0,2*n.variantes),
        v.rsme=rep(0,2*n.variantes),
        m.mape=rep(0,2*n.variantes),
        v.mape=rep(0,2*n.variantes)

for (i in 1:n.variantes) {

cuadro$Tasa[i] <- "Cura"
cuadro$Modelo[i] <- levels(res.cura$validacion)[i]

cuadro[i,"m.rsme"] <- (res.cura %>%
                      filter(validacion==levels(res.cura$validacion)[i]) %>%
                      summarise(m.rsme=median(rsme)))[1,1]
cuadro[i,"v.rsme"] <- (res.cura %>%
                      filter(validacion==levels(res.cura$validacion)[i]) %>%
                      summarise(v.rsme=sd(rsme)))[1,1]
cuadro[i,"m.mape"] <- (res.cura %>%
                      filter(validacion==levels(res.cura$validacion)[i]) %>%
                      summarise(m.mape=median(mape)))[1,1]
cuadro[i,"v.mape"] <- (res.cura %>%
                      filter(validacion==levels(res.cura$validacion)[i]) %>%
                      summarise(v.mape=sd(mape)))[1,1]

cuadro$Tasa[i+n.variantes]<- "Incumplimineto"
cuadro$Modelo[i+n.variantes] <- levels(res.incump$validacion)[i]

cuadro[i+n.variantes,"m.rsme"] <- (res.incump %>%
                                  filter(validacion==levels(res.incump$validacion)[i]) %>%
                                  summarise(m.rsme=median(rsme)))[1,1]
cuadro[i+n.variantes,"v.rsme"] <- (res.incump %>%
                                  filter(validacion==levels(res.incump$validacion)[i]) %>%
                                  summarise(v.rsme=sd(rsme)))[1,1]
cuadro[i+n.variantes,"m.mape"] <- (res.incump %>%
                                  filter(validacion==levels(res.incump$validacion)[i]) %>%
                                  summarise(m.mape=median(mape)))[1,1]
cuadro[i+n.variantes,"v.mape"] <- (res.incump %>%
                                  filter(validacion==levels(res.incump$validacion)[i]) %>%
                                  summarise(v.mape=sd(mape)))[1,1]
}

cuadro <- cuadro%>%
  mutate_if(is.numeric, function(x) round(x,4)) %>%
  arrange(Tasa,m.rsme) %>%
  mutate(m.mape=scales::percent(m.mape),
         v.mape=scales::percent(v.mape))
names(cuadro) <- c("Tasa","Modelo","Mediana RSME","Desviación Estándar RSME","Mediana
MAPE","Desviación Estándar MAPE")
cuadro %>%
  kbl(caption = paste0("Cuadro ",num.cuadro,". ",segmento,

```

```

      ": Resumen de medidas de precisión de variantes del modelo de ",
      nom.modelo),
      align = c("l", "l", "r", "r", "r", "r")) %>%
kable_classic(full_width = F, html_font = "Optima", font_size = 12)%>%
column_spec(3:6, width = "7em")%>%
column_spec(2, width = "20em") %>%
row_spec(c(1,n.variantes+1), bold = T)
}
CV.reg.longitudinal <- function(data=datos.mod,
  nom.var.respuesta="s.cura.micro",var.pred=predictoras){
n <- dim(data)[1]
data <- data %>%
  mutate(tendencia=seq(1:n),
          tendencia2=tendencia^2,
          estacionalidad= relevel(as.factor(data$mes),"12") )
grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F) #65,horizon
  = 1,fixedWindow = F)
n.folds <- lapply(grupos,length)$train
rsme.tend <- rep(0, rep(n.folds))
rsme.tend.estac <- rep(0, rep(n.folds))
rsme.tend2.estac <- rep(0, rep(n.folds))
rsme.estac <- rep(0, rep(n.folds))
mape.tend <- rep(0, rep(n.folds))
mape.tend.estac <- rep(0, rep(n.folds))
mape.tend2.estac <- rep(0, rep(n.folds))
mape.estac <- rep(0, rep(n.folds))
for (k in 1:n.folds) { # se repita la cantidad de folds 45
  taprendizaje <- data[grupos$train[[k]], ]
  ttesting <- data[grupos$test[[k]], ]
  actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
  modelo <- lm(as.formula(paste0(nom.var.respuesta,"~tendencia+",paste(var.pred,collapse =
    "+"))),data=taprendizaje)
  prediccion <- predict(modelo, ttesting)
  rsme.tend[k]<-MLmetrics::RMSE(prediccion,actual)
  mape.tend[k]<-MLmetrics::MAPE(prediccion,actual)
  modelo <- lm(as.formula(paste0(nom.var.respuesta,"~tendencia+estacionalidad+",
    paste(var.pred,collapse = "+"))),data=taprendizaje)
  prediccion <- predict(modelo, ttesting)
  rsme.tend.estac[k]<-MLmetrics::RMSE(prediccion,actual)
  mape.tend.estac[k]<-MLmetrics::MAPE(prediccion,actual)
  modelo <- lm(as.formula(paste0(nom.var.respuesta,"~tendencia+tendencia2+estacionalidad+",
    paste(var.pred,collapse = "+"))),data=taprendizaje)
  prediccion <- predict(modelo, ttesting)
  rsme.tend2.estac[k]<-MLmetrics::RMSE(prediccion,actual)
  mape.tend2.estac[k]<-MLmetrics::MAPE(prediccion,actual)
  modelo <- lm(as.formula(paste0(nom.var.respuesta,"~estacionalidad+",
    paste(var.pred,collapse = "+"))),data=taprendizaje)
  prediccion <- predict(modelo, ttesting)
  rsme.estac[k]<-MLmetrics::RMSE(prediccion,actual)

```

```

    mape.estac[k]<-MLmetrics::MAPE(prediccion,actual)
  }
errores<-data.frame(iteracion=c(rep(1:n.folds,4)),
                    rsme=c(rsme.tend,rsme.tend.estac,rsme.tend2.estac,
                          rsme.estac),
                    mape=c(mape.tend,mape.tend.estac,mape.tend2.estac,
                          mape.estac),
                    validacion=as.factor(c(rep("Tendencia",n.folds),
                                           rep("Tendencia lineal estacionalidad",n.folds),
                                           rep("Tendencia cuadrática estacionalidad",n.folds),
                                           rep("Estacionalidad",n.folds)
                                           )))
  )
return(errores)
}
CV.knn <- function(data=datos.mod, nom.var.respuesta="s.incump.total",var.pred=predictoras){
  n <- dim(data)[1]
  grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
  n.folds <- lapply(grupos,length)$train
  rsme.rectangular <- rep(0, rep(n.folds))
  rsme.triangular <- rep(0, rep(n.folds))
  rsme.epanechnikov <- rep(0, rep(n.folds))
  rsme.biweight <- rep(0, rep(n.folds))
  rsme.triweight <- rep(0, rep(n.folds))
  rsme.cos <- rep(0, rep(n.folds))
  rsme.inv <- rep(0, rep(n.folds))
  rsme.gaussian <- rep(0, rep(n.folds))
  rsme.optimal <- rep(0, rep(n.folds))
  mape.rectangular <- rep(0, rep(n.folds))
  mape.triangular <- rep(0, rep(n.folds))
  mape.epanechnikov <- rep(0, rep(n.folds))
  mape.biweight <- rep(0, rep(n.folds))
  mape.triweight <- rep(0, rep(n.folds))
  mape.cos <- rep(0, rep(n.folds))
  mape.inv <- rep(0, rep(n.folds))
  mape.gaussian <- rep(0, rep(n.folds))
  mape.optimal <- rep(0, rep(n.folds))
  for (k in 1:n.folds) { # se repita la cantidad de folds 45
    taprendizaje <- data[grupos$train[[k]], ]
    ttesting <- data[grupos$test[[k]], ]
    actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
    modelo <- train.kknn(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                       data=taprendizaje,kmax=floor(sqrt(nrow(data))),kernel = "rectangular")
    prediccion <- predict(modelo, ttesting)
    rsme.rectangular[k]<-MLmetrics::RMSE(prediccion,actual)
    mape.rectangular[k]<-MLmetrics::MAPE(prediccion,actual)
    modelo <- train.kknn(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                       data=taprendizaje,kmax=floor(sqrt(nrow(data))),kernel = "triangular")
    prediccion <- predict(modelo, ttesting)
  }
}

```



```

        rep("Triweight",n.folds),
        rep("Cos",n.folds),
        rep("Inv",n.folds),
        rep("Gaussian",n.folds),
        rep("Optimal",n.folds)
    ))
)
returnerrores)
}
CV.svr <- function(data=datos.mod, nom.var.respuesta="s.incump.total",var.pred=predictoras){
  n <- dim(data)[1]
  grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
  n.folds <- lapply(grupos,length)$train
  rsme.linear <- rep(0, rep(n.folds))
  rsme.radial <- rep(0, rep(n.folds))
  rsme.sigmoid <- rep(0, rep(n.folds))
  rsme.polynomial <- rep(0, rep(n.folds))
  mape.linear <- rep(0, rep(n.folds))
  mape.radial <- rep(0, rep(n.folds))
  mape.sigmoid <- rep(0, rep(n.folds))
  mape.polynomial <- rep(0, rep(n.folds))
  for (k in 1:n.folds) { # se repita la cantidad de folds 16
    taprendizaje <- data[grupos$train[[k]], ]
    ttesting <- data[grupos$test[[k]], ]
    actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
    modelo <- svm(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                  data=taprendizaje, kernel = "linear")
    prediccion <- predict(modelo, ttesting)
    rsme.linear[k]<-MLmetrics::RMSE(prediccion,actual)
    mape.linear[k]<-MLmetrics::MAPE(prediccion,actual)
    modelo <- svm(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                  data=taprendizaje, kernel = "radial")
    prediccion <- predict(modelo, ttesting)
    rsme.radial[k]<-MLmetrics::RMSE(prediccion,actual)
    mape.radial[k]<-MLmetrics::MAPE(prediccion,actual)
    modelo <- svm(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                  data=taprendizaje, kernel = "sigmoid")
    prediccion <- predict(modelo, ttesting)
    rsme.sigmoid[k]<-MLmetrics::RMSE(prediccion,actual)
    mape.sigmoid[k]<-MLmetrics::MAPE(prediccion,actual)
    modelo <- svm(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                  data=taprendizaje, kernel = "polynomial")
    prediccion <- predict(modelo, ttesting)
    rsme.polynomial[k]<-MLmetrics::RMSE(prediccion,actual)
    mape.polynomial[k]<-MLmetrics::MAPE(prediccion,actual)
  }
  errores<-data.frame(iteracion=c(rep(1:n.folds,4)),
                      rsme=c(rsme.linear,rsme.radial,rsme.sigmoid,
                              rsme.polynomial),

```

```

mape=c(mape.linear,mape.radial,mape.sigmoid,
       mape.polynomial),
validacion=as.factor(c(rep("Linear",n.folds),
                       rep("Radial",n.folds),
                       rep("Sigmoid",n.folds),
                       rep("Polynomial",n.folds)
                      ))
)

returnerrores)
}

CV.arima <- function(data=datos.mod, nom.var.respuesta="s.incump.total",var.pred=predictoras){
  n <- dim(data)[1]
  grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
  n.folds <- lapply(grupos,length)$strat
  comb <- expand.grid(p = 0:5, q = 0:5)
  for (k in 1:n.folds) { # se repita la cantidad de folds 16
    comb[paste("rsme", k, sep="")] <- rep(0,dim(comb)[1])
    comb[paste("mape", k, sep="")] <- rep(0,dim(comb)[1])
    for (i in 1:dim(comb)[1]){
      taprendizaje <- data[grupos$train[[k]], ] %>% mutate_all(as.numeric) # para que el arima
        pueda leer la var de intervencion
      ttesting <- data[grupos$test[[k]], ] %>% mutate_all(as.numeric)
      actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
      serie = ts(na.omit(as.numeric(as.matrix(taprendizaje[,nom.var.respuesta]))), frequency = 12)
      modelo <- try(Arima(serie, order=c(comb$p[i],0,comb$q[i]), xreg =
        as.matrix(taprendizaje[,var.pred]),
        include.mean = TRUE ),silent = T)
      if(class(modelo) == "try-error"){
        comb[i,paste("rsme", k, sep="")]<-NA
        comb[i,paste("mape", k, sep="")]<-NA
      }else{
        prediccion <- forecast::forecast(modelo, h = dim(ttesting)[1],
          xreg = as.matrix(ttesting[,var.pred]) )
        comb[i,paste("rsme", k, sep="")]<-MLmetrics::RMSE(as.numeric(prediccion$mean),actual)
        comb[i,paste("mape", k, sep="")]<-MLmetrics::MAPE(as.numeric(prediccion$mean),actual)
      }
    }
  }
}

res.mejores.comb <- comb %>% mutate(orden= comb %>%
  select(starts_with("rsme")) %>%
  rowSums()/dim(comb)[1] ) %>%
  mutate(pq=paste0("p=",p,"q=",q,"")) %>%
  arrange(orden) %>% head(4)

mejores.comb <- res.mejores.comb$pq[rep(seq_len(nrow(res.mejores.comb)), each = n.folds)]
rsmes <- res.mejores.comb %>% select(starts_with("rsme")) %>% t() %>% c
mapes <- res.mejores.comb %>% select(starts_with("mape")) %>% t() %>% c
errores<-data.frame(iteracion=c(rep(1:n.folds,4)),
  rsme=c(rsmes),

```

```

        mape=c(mapes),
        validacion=as.factor(c(mejores.comb
        ))
    )
    returnerrores)
}
CV.nnetar <- function(data=datos.mod, nom.var.respuesta="s.incump.total",var.pred=predictoras){
  n <- dim(data)[1]
  set.seed(1234)
  grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
  n.folds <- lapply(grupos,length)$strat
  comb <- expand.grid(size = c(1:10),
                    decay = c(0, 0.05, 0.1, 1, 2))
  for (k in 1:n.folds) { # se repita la cantidad de folds 16
    comb[paste("rsme", k, sep="")] <- rep(0,dim(comb)[1])
    comb[paste("mape", k, sep="")] <- rep(0,dim(comb)[1])
    for (i in 1:dim(comb)[1]){
      taprendizaje <- data[grupos$train[[k]], ] %>% mutate_all(as.numeric) # para que el arima
        pueda leer la var de intervencion
      ttesting <- data[grupos$test[[k]], ] %>% mutate_all(as.numeric)
      actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
      serie = ts(as.numeric(as.matrix(taprendizaje[,nom.var.respuesta])), frequency = 12)
      modelo <- try(nnetar(serie, size = comb$size[i], decay= comb$decay[i] ,
                        xreg = as.matrix(taprendizaje[,var.pred]) ),silent = T)
      if(class(modelo) == "try-error"){
        comb[i,paste("rsme", k, sep="")]<-NA
        comb[i,paste("mape", k, sep="")]<-NA
      }else{
        prediccion <- forecast::forecast(modelo, h = dim(ttesting)[1],
                                         xreg = as.matrix(ttesting[,var.pred]) ,PI = F )
        comb[i,paste("rsme", k, sep="")]<-MLmetrics::RMSE(as.numeric(prediccion$mean),actual)
        comb[i,paste("mape", k, sep="")]<-MLmetrics::MAPE(as.numeric(prediccion$mean),actual)
      }
    }
  }
  res.mejores.comb <- comb %>% mutate(orden= comb %>%
                                select(starts_with("rsme")) %>%
                                rowSums()/dim(comb)[1] ) %>%
    mutate(pq=paste0(" (size=",size," ,decay=",decay," )")) %>%
    arrange(orden) %>% head(4)
  mejores.comb <- res.mejores.comb$pq[rep(seq_len(nrow(res.mejores.comb)), each = n.folds)]
  rsmes <- res.mejores.comb %>% select(starts_with("rsme")) %>% t() %>% c
  mapes <- res.mejores.comb %>% select(starts_with("mape")) %>% t() %>% c
  errores<-data.frame(iteracion=c(rep(1:n.folds,4)),
                    rsme=c(rsmes),
                    mape=c(mapes),
                    validacion=as.factor(c(mejores.comb
                    )))
}

```

```

    return(errores)
  }
CV.rf <- function(data=datos.mod, nom.var.respuesta="s.incump.total",var.pred=predictoras){
  n <- dim(data)[1]
  set.seed(1234)
  grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
  n.folds <- lapply(grupos,length)$strat
  p <- length(var.pred)
  comb <- expand.grid(ntree = c(50,100,250,500,1000),
                    mtry = unique(round(c(p*0.25,p/3,p*0.5,p*0.75),0)) )
  for (k in 1:n.folds) { # se repita la cantidad de folds 16
    comb[paste("rsme", k, sep="")] <- rep(0,dim(comb)[1])
    comb[paste("mape", k, sep="")] <- rep(0,dim(comb)[1])
    for (i in 1:dim(comb)[1]){
      taprendizaje <- data[grupos$strat[[k]], ] %>% mutate_all(as.numeric) # para que el arima
      pueda leer la var de intervencion
      ttesting <- data[grupos$test[[k]], ] %>% mutate_all(as.numeric)
      actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
      modelo <- randomForest::randomForest(as.formula(paste0(nom.var.respuesta,"~",paste(predictoras, collapse = "+"))),
                                           data=taprendizaje,importance=F, ntree=comb$ntree[i],
                                           mtry=comb$mtry[i])
      prediccion <- predict(modelo, newdata=ttesting)
      comb[i,paste("rsme", k, sep="")]<-MLmetrics::RMSE(as.numeric(prediccion),actual)
      comb[i,paste("mape", k, sep="")]<-MLmetrics::MAPE(as.numeric(prediccion),actual)
    }
  }
  res.mejores.comb <- comb %>% mutate(orden= comb %>%
                                   select(starts_with("rsme")) %>%
                                   rowSums()/dim(comb)[1] ) %>%
  mutate(pq=paste0("ntree=",ntree,"",mtry="",mtry, " ")) %>%
  arrange(orden) %>% head(4)
  mejores.comb <- res.mejores.comb$pq[rep(seq_len(nrow(res.mejores.comb)), each = n.folds)]
  rsmes <- res.mejores.comb %>% select(starts_with("rsme")) %>% t() %>% c
  mapes <- res.mejores.comb %>% select(starts_with("mape")) %>% t() %>% c
  errores<-data.frame(iteracion=c(rep(1:n.folds,4)),
                      rsme=c(rsmes),
                      mape=c(mapes),
                      validacion=as.factor(c(mejores.comb
                                              )))
  return(errores)
}
CV.general <- function(data=datos.mod,
                       nom.var.respuesta="s.incump.total",var.pred=predictoras,metodo=train.kknn,...){
  n <- dim(data)[1]
  rsme<-0
  mape<-0

```

```

grupos<-caret::createTimeSlices(y=1:n,initialWindow = 70,horizon = 1,fixedWindow = F)
n.folds <- lapply(grupos,length)$train
n.folds2 <- n.folds
for(k in 1:n.folds) {
  taprendizaje <- data[grupos$train[[k]], ]
  ttesting <- data[grupos$test[[k]], ]
  actual <- as.numeric(as.matrix(ttesting[,nom.var.respuesta]))
  if( identical(metodo,Arima) | identical(metodo,nnetar) ){
    taprendizaje <- data[grupos$train[[k]], ]%>% mutate_all(as.numeric)
    ttesting <- data[grupos$test[[k]], ]%>% mutate_all(as.numeric)
    serie = ts(as.numeric(as.matrix(taprendizaje[,nom.var.respuesta])), frequency = 12)
    modelo <- try(metodo(serie, xreg = as.matrix(taprendizaje[,var.pred]), ...),silent = T)
    if(class(modelo) == "try-error"){
      n.folds2 <- n.folds2-1
    }else{
      if(identified(metodo,Arima)){
        prediccion <- forecast::forecast(modelo, h = dim(ttesting)[1],
                                          xreg = as.matrix(ttesting[,var.pred]) )
        rsme <- rsme + MLmetrics::RMSE(as.numeric(prediccion$mean),actual)
        mape <- mape + MLmetrics::MAPE(as.numeric(prediccion$mean),actual)
      }else{
        prediccion <- forecast::forecast(modelo, h = dim(ttesting)[1],
                                          xreg = as.matrix(ttesting[,var.pred]) ,PI = F )
        rsme <- rsme + MLmetrics::RMSE(as.numeric(prediccion$mean),actual)
        mape <- mape + MLmetrics::MAPE(as.numeric(prediccion$mean),actual)
      }
    }
  }else{
    if( identical(metodo,xgboost.reg.propia) | identical(metodo,svm) ){
      taprendizaje <- data[grupos$train[[k]], ]%>% mutate_all(as.numeric)
      ttesting <- data[grupos$test[[k]], ]%>% mutate_all(as.numeric)
      modelo <- metodo(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse = "+"))),
                      data=taprendizaje, ... )
      prediccion <- predict(modelo, as.matrix(ttesting[,var.pred]))
      rsme <- rsme + MLmetrics::RMSE(prediccion,actual)
      mape <- mape + MLmetrics::MAPE(prediccion,actual)
    }else{
      if( identical(metodo,lm) ){
        data2 <- data %>%
          mutate(tendencia=seq(1:n),
                 tendencia2=tendencia^2,
                 estacionalidad= releval(as.factor(data$mes),"12") )
        taprendizaje <- data2[grupos$train[[k]], ]%>% mutate_all(as.numeric)
        ttesting <- data2[grupos$test[[k]], ]%>% mutate_all(as.numeric)
        modelo <- lm(as.formula(paste0(nom.var.respuesta, ... ,
                                       paste(var.pred,collapse = "+"))),data=taprendizaje)
        prediccion <- predict(modelo, ttesting)
        rsme <- rsme + MLmetrics::RMSE(prediccion,actual)
        mape <- mape + MLmetrics::MAPE(prediccion,actual)
      }
    }
  }
}

```

```

    }else{
      modelo <- metodo(as.formula(paste0(nom.var.respuesta,"~",paste(var.pred,collapse =
""))),
                      data=taprendizaje, ... )
      prediccion <- predict(modelo,ttesting)
      rsme <- rsme + MLmetrics::RMSE(prediccion,actual)
      mape <- mape + MLmetrics::MAPE(prediccion,actual)
    }
  }
}
RSME <- rsme/ n.folds2
MAPE <- mape/ n.folds2
return(list("RSME"=RSME,"MAPE"=MAPE,"Modelo"=modelo))
}
FuerzaBruta.selec.var <- function(data=datos.mod, nom.var.respuesta="s.incump.total",
                                var.pred=predictoras,metodo=train.kknn,
                                nom.intervencion="var.intervencion",...){
var.pred2 <- var.pred[var.pred != nom.intervencion] #se quita para poder agregar a todas las
combinaciones la intervencion
lista <- list()
for (i in seq_along(var.pred2)) {
  lista <- c(lista, combn(var.pred2, i, simplify = F))
}
for (i in 1:length(lista)) { #aqui se agrega la intervencion a cada obligacion
  lista[[i]] <- c(lista[[i]],nom.intervencion)
}
validacion<-CV.general(data,nom.var.respuesta, var.pred= lista[[1]],metodo, ...)
rsme<-validacion[[1]]
mape<-validacion[[2]]
modelo<-validacion[[3]]
combinacion<-gsub(" ",", ",do.call(paste,as.list( lista[[1]] ) ) )
for(i in 2:length(lista)){
  validacion<-CV.general(data,nom.var.respuesta, var.pred= lista[[i]],metodo, ...)
  rsme.i<-validacion[[1]]
  mape.i<-validacion[[2]]
  modelo.i<-validacion[[3]]
  if(rsme.i < rsme){
    rsme<-rsme.i
    combinacion<- gsub(" ",", ",do.call(paste,as.list( lista[[i]] ) ) )
    mape<-mape.i
    modelo<-modelo.i
  }
}
}
return(
  list(
    Solucion=paste("La mejor combinación es: ",combinacion,", con un RSME de",round(rsme,4),
                  "y MAPE del",scales::percent(mape),sep = " "),
    Combinacion=combinacion,

```

```

    RSME=rsme,
    MAPE=mape,
    Modelo=modelo)
  )
}
estimacion.perdidas <- function(mod.cura=t_cura_modelo_final, mod.incump=t_incump_modelo_final,

    res.fb.cura=t_res_fb_cura_nnetar,res.fb.incump=t_res_fb_incump_nnetar,

    escenarios=ESC.SIM,EAD=EAD,COLATERIZACION=COLATERIZACION,segmento,num.graf){
var.cura <- gsub(",","",strsplit(res.fb.cura$Combinacion," ")[[1]])
var.incump <- gsub(",","",strsplit(res.fb.incump$Combinacion," ")[[1]])
if (class(mod.cura)[1]== "lm") {
  if(length(mod.cura$model$tendencia)!=0){
    tend.p <- mod.cura$model$tendencia[length(mod.cura$model$tendencia)]+1:12
  }else{tend.p <-NA}
  if(length(mod.cura$model$tendencia2)!=0){
    estac.p <- (sqrt(mod.cura$model$tendencia2[length(mod.cura$model$tendencia2)])+1:12)^2
  }else{tend2.p <-NA}
  if(length(mod.cura$model$estacionalidad)!=0){
    estac.p <- 1:12
  }else{estac.p <-NA}
  escenarios$tendencia <- sample(tend.p,500,replace = T)
  escenarios$tendencia2 <- sample(tend2.p,500,replace = T)
  escenarios$estacionalidad <- sample(estac.p,500,replace = T)
  tasas.cura <- as.numeric(predict(mod.cura,escenarios))
}else{
  if(class(mod.cura)[1]== "nnetar" | class(mod.cura)[1]== "forecast_ARIMA"){
    escenarios2 <- escenarios %>% mutate_all(as.numeric)
    tasas.cura <- as.numeric((forecast::forecast(mod.cura, h = 500,
                                                    xreg = as.matrix(escenarios2[,var.cura]) ,PI =
                                                    F ))$mean)
  }else{
    if(class(mod.cura)[1]== "xgb.Booster" | class(mod.cura)[1]== "svm"){
      escenarios2 <- escenarios %>% mutate_all(as.numeric)
      tasas.cura <- as.numeric(predict(mod.cura,as.matrix(escenarios2[,var.cura])) )
    }else{
      tasas.cura <- as.numeric(predict(mod.cura,escenarios[,var.cura]))
    }
  }
}
if (class(mod.incump)[1]== "lm") {
  if(length(mod.incump$model$tendencia)!=0){
    tend.p <- mod.incump$model$tendencia[length(mod.incump$model$tendencia)]+1:12
  }else{tend.p <-NA}
  if(length(mod.incump$model$tendencia2)!=0){
    tend2.p <- (sqrt(mod.incump$model$tendencia2[length(mod.incump$model$tendencia2)])+1:12)^2
  }else{tend2.p <-NA}
  if(length(mod.incump$model$estacionalidad)!=0){

```

```

    estac.p <- 1:12
  }else{estac.p <-NA}
  escenarios$tendencia <- sample(tend.p,500,replace = T)
  escenarios$tendencia2 <- sample(tend2.p,500,replace = T)
  escenarios$estacionalidad <- sample(estac.p,500,replace = T)
  tasas.incump <- as.numeric(predict(mod.incump,escenarios))
}else{
  if(class(mod.incump)[1]== "nnetar" | class(mod.incump)[1]== "forecast_ARIMA"){
    escenarios2 <- escenarios %>% mutate_all(as.numeric)
    tasas.incump <- as.numeric((forecast::forecast(mod.incump, h = 500,
                                                    xreg = as.matrix(escenarios2[,var.incump])
                                                    ,PI = F ))$mean)
  }else{
    if(class(mod.incump)[1]== "xgb.Booster" | class(mod.incump)[1]== "svm"){
      escenarios2 <- escenarios %>% mutate_all(as.numeric)

      tasas.incump <- as.numeric(predict(mod.incump,as.matrix(escenarios2[,var.incump])))
    }else{
      tasas.incump <- as.numeric(predict(mod.incump,escenarios[,var.incump]))
    }
  }
}
dat.per <- data.frame(perdidas = EAD*(1-COLATERIZACION)*(1-
  logistic.inv(tasas.cura))*logistic.inv(tasas.incump))
max.count <- max(hist(dat.per$perdidas, breaks=seq(min(dat.per$perdidas), max(dat.per$perdidas),
  length.out=30), plot=FALSE)$counts)+5
grf.perdidas <- ggplot(dat.per, aes(x=perdidas)) +
  geom_histogram(fill=colores(4)[3], colour="white",alpha=0.8)+
  ylim(c(0,max.count))+
  theme_Publication()+
  labs(title = paste0("Figura ",num.graf),
       subtitle =paste0(segmento,": Distribución de pérdidas por riesgo de crédito"),
       y="Frecuencia", x="Monto Pérdidas \n (Dólares)")+
  scale_x_continuous(labels = comma)+
  geom_vline(xintercept = mean(dat.per$perdidas),
            linetype="dashed")+
  geom_vline(xintercept = quantile(dat.per$perdidas,0.99),
            linetype="dashed")+
  geom_label(x = mean(dat.per$perdidas), y = max.count-5, label = paste0("Promedio:
  \n",dollar(mean(dat.per$perdidas)) ), fill = "white",size = 3)+
  geom_label(x = quantile(dat.per$perdidas,0.99), y = max.count-5, label = paste0("Percentil 99:
  \n",dollar(quantile(dat.per$perdidas,0.99)) ), fill = "white",size = 3)
return(list(grafico=grf.perdidas, perdidas=dat.per$perdidas,
  promedio.p=mean(dat.per$perdidas),perc99.p=quantile(dat.per$perdidas,0.99))

```