PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD

Maestría en Sistemas de Información mención Data Science



TEMA:

Modelización del crecimiento empresarial ecuatoriano mediante ecuaciones diferenciales con retardo y algoritmos de machine learning.

AUTOR:

Giovanny Francisco Guanga Lanchimba

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN SISTEMAS DE INFORMACIÓN MENCIÓN DATA SCIENCE.

TUTOR:

Dr. Carlos Julio Mayorga Arias, Ph. D

Quito – Agosto 2026

Dedicatoria

A mi mami querida Carmen Guanga.

Agradecimiento

Agradezco al DIOS de mis padres Abraham, Isaac y Jacob por haberme dado la oportunidad de seguir está fascinante Maestría la cual no fue fácil pero con su bendición he podido culminarla.

A mi madre por su amor incondicional, por haber confiado siempre en mí y su esfuerzo que me inspiró a seguir adelante.

A Fer por haber confiado siempre en mí. Sus palabras de aliento en los momentos difíciles me motivaron a seguir adelante.

Agradezco a cada uno de mis profesores de la Maestría en Sistemas de Información por transmitir sus conocimientos y experiencias. A todos muchas gracias.

Un agradecimiento muy especial al Dr. Carlos Mayorga por haberme dado la oportunidad de realizar este proyecto de titulación, por su guía y paciencia. A usted doctor, Muchas Gracias.

Resumen.

La presente investigación tiene como objetivo modelar y predecir el comportamiento del crecimiento empresarial en Ecuador mediante la aplicación y comparación de modelos de aprendizaje automático y ecuaciones diferenciales con retardo. El estudio considera como variables objetivo las tasas de crecimiento de sociedades y personas naturales y el total de ventas y exportaciones. La información proviene de registros del Servicio de Rentas Internas y las variables macroeconómicas y financieras provienen de fuentes oficiales. Para el proceso de análisis se empleó la metodología CRISP-DM.

Para la predicción se implementaron los algoritmos XGBoost y Random Forest, tanto en configuraciones univariables como incorporando variables exógenas, además de un modelo basado en Ecuaciones Diferenciales con Retardo. Los modelos fueron evaluados mediante validación temporal Walk-Forward, respetando el orden cronológico de las observaciones y evitando fuga de información. Como métricas de desempeño se utilizaron el error cuadrático medio en su raíz (RMSE), el error absoluto medio (MAE) y el error porcentual absoluto medio (MAPE).

Los resultados muestran que el desempeño de los modelos depende de la variable analizada. Para la tasa de crecimiento de sociedades, XGBoost con variables exógenas presentó el mejor rendimiento global, con un RMSE de $0{,}134781$, MAE de $0,097265$ y un MAPE de $16{,}40\,\%$, aunque el modelo DDE obtuvo un MAE ligeramente inferior, evidenciando un desempeño competitivo. En el caso del total de ventas y exportaciones, el modelo DDE presentó los mejores resultados, alcanzando un RMSE de $2{,}790778 \times 10^9$, un MAE de $1{,}752034 \times 10^9$ y un MAPE de $6{,}96\,\%$, superando a XGBoost y Random Forest. Asimismo, el análisis de importancia de variables mostró que los rezagos de las propias variables objetivo constituyen una fuente relevante de información predictiva, mientras que las variables macroeconómicas aportan información adicional en determinadas configuraciones.

Palabras clave: Crecimiento Empresarial, Series Temporales, XGBoost, Random Forest, Ecuaciones Diferenciales con Retardo.

Abstract.

This research aims to model and predict the behavior of business growth in Ecuador through the application and comparison of machine learning models and delay differential equations. The study considers as target variables the growth rates of companies and natural persons, as well as total sales and exports. The information comes mainly from records of the Internal Revenue Service and from macroeconomic and financial variables obtained from official sources. The CRISP-DM methodology was used for the analysis process.

For prediction, the XGBoost and Random Forest algorithms were implemented in both univariate configurations and configurations incorporating exogenous variables, in addition to a model based on Delay Differential Equations. The models were evaluated using Walk-Forward temporal validation, preserving the chronological order of the observations and avoiding information leakage. The performance metrics used were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The results show that model performance depends on the variable being analyzed. For the growth rate of companies, XGBoost with exogenous variables achieved the best overall performance, with an RMSE of $0{,}134781$, an MAE of $0{,}097265$, and a MAPE of $16{,}40\,\%$, although the DDE model obtained a slightly lower MAE, demonstrating competitive performance. In the case of total sales and exports, the DDE model achieved the best results, with an RMSE of $2{,}790778 \times 10^9$, an MAE of $1{,}752034 \times 10^9$, and a MAPE of $6{,}96\,\%$, outperforming XGBoost and Random Forest. Likewise, the variable importance analysis showed that lags of the target variables themselves constitute a relevant source of predictive information, while macroeconomic variables provide additional information in certain configurations.

Keywords: Business Growth, Time Series, XGBoost, Random Forest, Delay Differential Equations.

Tabla de contenidos

DEDICATORIA ii

AGRADECIMIENTO iii

RESUMEN iv

ABSTRACT v

TABLA DE CONTENIDOS viii

1. INTRODUCCIÓN 1

1.1. Planteamiento del problema . . . 1

1.2. Justificación . . . 1

1.3. Objetivos . . . 2

1.3.1. Objetivo general . . . 2

1.3.2. Objetivos específicos . . . 2

1.4. Alcance . . . 3

2. MARCO TEÓRICO 4

2.1. Introducción. . . . 4

2.2. Crecimiento empresarial . . . 4

2.3. Demografía empresarial . . . 4

2.4. Ecuaciones diferenciales con retardo . . . 4

2.5. Método de diferencias finitas . . . 5

2.5.1. Aproximaciones de diferencias finitas . . . 5

2.6. Regla de Simpson . . . 6

2.7. Algoritmos de machine learning para series temporales . . . 6

2.7.1. Aprendizaje supervisado . . . 6

2.7.2. Series temporales como aprendizaje supervisado . . . 7
2.7.3. Estadísticas móviles . . . 7
2.7.4. Manejo de datos de series temporales . . . 8
2.7.5. Validación hacia adelante (Walk Forward Validation) . . . 8
2.8. Función de Autocorrelación Parcial (PACF) . . . 9
2.9. Prueba de Causalidad de Granger . . . 10
2.10. Estado del arte . . . 11

3. MARCO METODOLÓGICO **14**
3.1. Diseño de investigación . . . 14
3.2. Enfoque metológico . . . 14
3.3. Metodología CRISP-DM . . . 14
3.3.1. Comprensión del negocio . . . 15
3.3.2. Comprensión de los datos . . . 15
3.3.3. Preparación de datos . . . 24

4. MODELADO Y EVALUACIÓN **35**
4.1. Modelo ecuación diferencial con retardo . . . 35
4.1.1. Formulación matemática . . . 35
4.1.2. Tasa de Crecimiento de las Sociedades . . . 37
4.1.3. Tasa de Crecimiento de las Personas Naturales . . . 38
4.1.4. Total Ventas y Exportaciones . . . 39
4.2. Ingeniería de características en series temporales . . . 40
4.2.1. Selección de lags temporales de la variable objetivo: Tasa de crecimiento de sociedades . . . 40
4.2.2. Selección de lags temporales de la variable objetivo: Tasa de crecimiento de personas naturales . . . 42
4.2.3. Selección de lags temporales de la variable objetivo: Ventas y Exportaciones 43
4.3. Modelo XGBoost . . . 45
4.3.1. Tasa de Crecimiento de las Sociedades . . . 45
4.3.2. Tasa de Crecimiento de las Personas Naturales . . . 47

4.3.3. Total Ventas y Exportaciones . . . 49
4.4. Modelo Random Forest . . . 51
4.4.1. Tasa de Crecimiento de las Sociedades . . . 51
4.4.2. Tasa de Crecimiento de las Personas Naturales . . . 53
4.4.3. Total Ventas y Exportaciones . . . 55
4.5. Evaluación de los modelos . . . 58
4.5.1. Resultados del modelo de Ecuaciones Diferenciales con Retardo . . . 58
4.5.2. Resultados de XGBoost . . . 58
4.5.3. Resultados de Random Forest . . . 58
4.5.4. Comparación de los modelos . . . 59

5. DISCUSIÓN **61**
5.1. Interpretación del parámetro de retardo τ. . . . 61
5.2. Interpretación de los parámetros α y β. . . . 62
5.3. Contraste de los modelos de aprendizaje automático con el estado del arte . . . 62
5.4. Resultados . . . 63

6. CONCLUSIONES **65**

BIBLIOGRAFÍA **67**

A. ANEXOS **70**
A.1. Resultados modelo DDE . . . 70
A.1.1. Tasa de crecimiento de sociedades . . . 70
A.1.2. Tasa de crecimiento de personas naturales . . . 72
A.1.3. Total ventas y exportaciones . . . 73

CAPÍTULO 1.
INTRODUCCIÓN

1.1 Planteamiento del problema

La presente investigación se centra en el análisis del crecimiento temporal del número de empresas en Ecuador y de las ventas totales realizadas, observados a partir de los registros históricos del Servicio de Rentas Internas. La evolución del número de empresas constituye un indicador relevante de la actividad económica y del desarrollo productivo del país, debido a su relación con variables como empleo, inversión y dinamismo empresarial. Sin embargo, la dinámica de crecimiento empresarial no depende únicamente de las condiciones presentes, sino también de factores ocurridos en períodos anteriores, como cambios económicos, tributarios y financieros. Por ello, las ecuaciones diferenciales con retardo representan una alternativa matemática adecuada para incorporar efectos de memoria temporal en la modelación del sistema. Por otra parte, los algoritmos de aprendizaje automático, como XGBoost y Random Forest, han demostrado un alto desempeño en problemas de predicción y análisis de series temporales. No obstante, estos enfoques suelen priorizar la precisión predictiva sobre la interpretación matemática del comportamiento dinámico. En el contexto ecuatoriano, existe escasa investigación orientada a modelar el crecimiento del número de empresas mediante ecuaciones diferenciales con retardo utilizando datos oficiales del SRI y compararlo con técnicas de machine learning. En consecuencia, surge la necesidad de desarrollar modelos que permitan describir y predecir adecuadamente esta dinámica empresarial y evaluar el desempeño de ambos enfoques.

1.2 Justificación

El comportamiento y evolución del número de empresas y ventas totales registradas en Ecuador constituye un indicador relevante para comprender la dinámica del sector empresarial y apoyar la toma de decisiones relacionadas con la planificación económica y productiva país. Las variaciones en la creación, permanencia y salida de sociedades y personas naturales que realizan actividades con fines de lucro pueden estar asociadas con factores económicos como el crecimiento del producto interno bruto, la inflación, la tasa de desempleo y las tasas de interés, por lo que disponer de

herramientas que permitan analizar y anticipar su comportamiento resulta de interés para la planificación y formulación de estrategias. Adicionalmente, la dinámica actual de las sociedades, personas naturales y sus ventas puede verse influida no solo por el contexto económico contemporáneo, sino también por eventos de períodos anteriores. Por ello, resulta necesario emplear modelos capaces de incorporar explícitamente la dependencia respecto de valores pasados.

En este contexto, la presente investigación propone la aplicación de un modelo basado en una ecuación diferencial con retardo (DDE) para representar la dinámica mensual de las sociedades, personas naturales con fines de lucro en Ecuador y de las ventas totales realizadas, incorporando un término de memoria que permita considerar la influencia de estados anteriores de la variable. La utilización de este enfoque resulta pertinente debido a que las ecuaciones diferenciales con retardo permiten representar sistemas dinámicos en los que el estado presente depende de información histórica.

Como complemento, el desempeño del modelo DDE se contrasta con modelos de aprendizaje automático, específicamente XGBoost y Random Forest, que permiten analizar relaciones no lineales entre la variable objetivo, sus valores históricos y variables exógenas.

La investigación busca aportar, por tanto, una perspectiva que combine el enfoque de modelación matemática de sistemas dinámicos con memoria y las técnicas de aprendizaje automático, proporcionando evidencia sobre la capacidad de estos métodos para reproducir y predecir la evolución de las sociedades y personas naturales en Ecuador.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar y evaluar un modelo basado en ecuaciones diferenciales con retardo para describir y predecir la dinámica de crecimiento temporal del número de empresas y ventas totales en Ecuador, comparándolo con algoritmos de aprendizaje automático aplicados a series temporales.

1.3.2 Objetivos específicos

- Analizar los datos históricos del número de empresas registradas y las ventas totales reportadas en Ecuador con el fin de identificar patrones y tendencias temporales.

- Proponer un modelo de ecuaciones diferenciales con retardo que represente la dinámica del crecimiento empresarial.

- Implementar y comparar algoritmos de aprendizaje automático para evaluar su desempeño predictivo frente al modelo matemático propuesto.

1.4 Alcance

La presente investigación tiene como alcance el análisis y la predicción del comportamiento de las empresas en Ecuador, utilizando información histórica correspondiente al período comprendido entre 1950 y 2026. Para el período comprendido entre 1950 y 1999 se dispone de información sobre el total de sociedades y personas naturales activas, la cual se utiliza como referencia para analizar la evolución histórica del sector empresarial. Sin embargo, el modelado predictivo se desarrolla a partir de enero del año 2000, debido a que este período corresponde al inicio del proceso de dolarización de la economía ecuatoriana y, además, es a partir de este año que se dispone de información mensual de las variables exógenas consideradas en el estudio. Para el análisis de las ventas totales, se considera el período comprendido entre enero de 2000 y mayo de 2026.

El estudio utiliza datos del catastro provincial del Servicio de Rentas Internas, así como registros de ventas totales extraídos de SAIKU, la plataforma de análisis y estadística multidimensional de la institución y de variables económicas consideradas como factores exógenos, entre ellas el crecimiento del PIB, la inflación, la tasa de desempleo y las tasas de interés de los segmentos productivos.

El alcance de la investigación se limita a la modelación y predicción de la dinámica de las sociedades y personas naturales que realizan actividades con fines de lucro en Ecuador a partir de la información disponible para el período de estudio. Por tanto, los resultados no pretenden establecer relaciones causales entre las variables económicas ni explicar de manera exhaustiva los factores jurídicos, políticos o sociales que determinan la creación o cierre de sociedades. De igual manera, las predicciones se encuentran condicionadas por la calidad, disponibilidad y comportamiento histórico de los datos utilizados.

CAPÍTULO 2.
MARCO TEÓRICO

2.1 Introducción.

El presente capítulo tiene como propósito identificar los principales conceptos, teorías, modelos y antecedentes relacionados con el crecimiento empresarial, las ecuaciones diferenciales con retardo, el análisis de series temporales y los algoritmos de aprendizaje automático aplicados a la predicción de fenómenos económicos.

2.2 Crecimiento empresarial

Consiste en cambios de carácter cuantitativo como el incremento de capacidad productiva, número de empleados, volumen de producción, entre otros; cambios de carácter cualitativo como variación de la estructura organizativa, recomposición de las unidades de negocio o cambios en la cartera de productos; y, además, debe atender a un período de tiempo determinado (Hernández, 2014).

2.3 Demografía empresarial

La demografía empresarial abarca los eventos del ciclo de vida de una empresa, como la creación y disolución de otras empresas, su desaparición y cese de operaciones, y su proporción respecto al total de empresas; el seguimiento de las empresas a lo largo del tiempo, lo que proporciona información sobre su supervivencia o desaparición; la evolución a lo largo del tiempo de ciertas características, como el tamaño, lo que ofrece información sobre el crecimiento de las empresas, o de un grupo de empresas, según el tipo de actividad (Eurostat, 2025).

2.4 Ecuaciones diferenciales con retardo

Las ecuaciones diferenciales con retardo son una clase de ecuaciones diferenciales que incluyen retardos o efectos de memoria en sus formulaciones. A continuación, se presenta una ecuación

diferencial de primer orden con un único retardo constante de la forma

$$u'_t = f(t, u, w), \quad w = u(t-\tau), \quad t > t_0 \tag{2.1}$$

donde u es la función incógnita, t es el tiempo, f es una función continua dada, $\tau > 0$ es un retardo constante y t_0 es el tiempo inicial.

El problema de Cauchy para la ecuación diferencial con retardo (2.1) consiste en encontrar una solución de esta ecuación que satisfaga la condición inicial

$$u = \varphi(t) \quad \text{para} \quad t_0 - \tau \leq t \leq t_0, \tag{2.2}$$

donde φ es una función continua dada.

2.5 Método de diferencias finitas

Los métodos de diferencias finitas son técnicas numéricas que se utilizan para aproximar las derivadas de funciones. Supongamos que tenemos una función suave $f : [a, b] \to \mathbb{R}$. Discretizamos el intervalo en puntos de la malla

$$x_i = a + ih, \quad i = 0, 1, 2, \cdots, N,$$

donde $h = \frac{b-a}{N}$ es el espaciado uniforme de la malla.

2.5.1 Aproximaciones de diferencias finitas

- **Diferencias hacia adelante.** La aproximación de la primera derivada en x_i mediante diferencias hacia adelante es
$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_i)}{h}.$$
Esta fórmula tiene un error de truncamiento local de $\mathcal{O}(h)$.

- **Diferencias hacia atrás.** La aproximación de la primera derivada en x_i mediante diferencias hacia atrás es
$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{h},$$
también con error de truncamiento local de $\mathcal{O}(h)$.

- **Diferencia central.** La aproximación de diferencia central es más precisa
$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_{i-1})}{2h},$$

con error de truncamiento local de $\mathcal{O}(h^2)$.

2.6 Regla de Simpson

La Regla de Simpson resulta de integrar sobre el intervalo $[a, b]$ el segundo polinomio de Lagrange con nodos igualmente espaciados $x_0 = a$, $x_2 = b$ y $x_1 = a + h$, donde $h = (b - a)/2$.

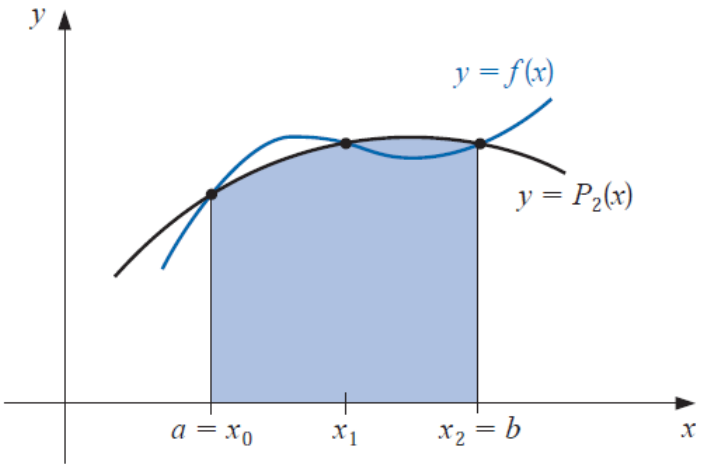


Figura 2.1: Integración numérica mediante la Regla de Simpson.
Nota: Tomado de *Numerical Analysis*, Burden and Faires (2010), Brooks - Cole Company.

Por lo tanto, la Regla de Simpson queda

$$\int_{x_0}^{x_2} f(x)dx = \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90}f^{(4)}(\xi),$$

para $\xi \in (x_0, x_2)$. El término $\frac{h^5}{90}f^{(4)}(\xi)$ es el error de truncamiento.

2.7 Algoritmos de machine learning para series temporales

2.7.1 Aprendizaje supervisado

La predicción de series temporales se debe plantear como un problema de aprendizaje supervisado. Esto permite acceder al conjunto de algoritmos estándar de aprendizaje automático, tanto lineales como no lineales.

Según (Brownlee, 2017), el aprendizaje supervisado consiste en tener variables de entrada X y una variable de salida y, y utilizar un algoritmo para aprender la función de mapeo de la entrada a la salida

$$y = f(X).$$

El objetivo es aproximar la función con tanta precisión que, al disponer de nuevos datos de entrada X, se puedan predecir las variables de salida y para dichos datos.

El pronóstico de series temporales presenta desafíos únicos debido a la naturaleza secuencial y temporal de los datos. A diferencia de las tareas tradicionales de aprendizaje supervisado, donde las observaciones son independientes e idénticamente distribuidas, los puntos de datos de las series temporales están intrínsecamente ordenados y suelen presentar tendencias (valores que aumentan o disminuyen con el tiempo), estacionalidad (patrones que se repiten según la época del año) y patrones cíclicos o incluso variaciones aleatorias e impredecibles.

2.7.2 Series temporales como aprendizaje supervisado

Según (Deka and Weiner, 2024), para aplicar los algoritmos XGBoost y Random Forest a la predicción de series temporales se requiere transformar los datos en un formato de aprendizaje supervisado que el modelo pueda procesar y esto implica varios pasos:

- Creación de características (features) de retardo,
- Incorporación de características estadísticas basadas en fechas móviles, y
- Garantizar que se conserve el orden temporal al dividir los datos en conjuntos de entrenamiento y prueba.

Una forma de codificar información temporal a partir de datos de series temporales es crear características de entrada que representen los valores en pasos de tiempo anteriores conocidas como características de retardo (lags).

2.7.3 Estadísticas móviles

Según (Deka and Weiner, 2024), las estadísticas móviles analizan la estabilidad de una serie temporal calculando estadísticas para los valores dentro de una ventana deslizante. Las estadísticas móviles proporcionan información adicional sobre la tendencia o la volatilidad en los datos de series

temporales. Estas estadísticas ayudan al modelo a capturar tendencias y variaciones a corto plazo, lo que lo hace más eficaz para realizar pronósticos.

2.7.4 Manejo de datos de series temporales

Al trabajar con datos de series temporales, es fundamental respetar el orden temporal de las observaciones. Para esto, los datos de series temporales deben dividirse secuencialmente, de modo que el conjunto de entrenamiento contenga las observaciones más antiguas, mientras que el conjunto de prueba contenga las más recientes. Esto garantiza que el modelo se evalúe como lo haría en producción, donde se predicen datos futuros basándose en patrones pasados.

Evaluar el rendimiento de los modelo es fundamental para comprender su capacidad de pronosticar valores futuros. Para esto, se utilizan las métricas error cuadrático medio (MSE), el error Raíz del Error Cuadrático Medio (RMSE) y Error Porcentual Absoluto Medio (MAPE).

2.7.5 Validación hacia adelante (Walk Forward Validation)

La validación hacia adelante (Walk-Forward Validation), es una técnica para evaluar modelos de pronóstico de series temporales. Divide el conjunto de datos en ventanas, entrenando y evaluando el modelo a medida que la ventana avanza a través de los datos, como se ilustra en la Figura 2.2 para el caso de una ventana deslizante.

Según (Brownlee, 2017) y (Subramanian, 2025), esta técnica evalúa de manera objetiva el rendimiento en el mundo real de un modelo de pronóstico de series temporales, simulando predicciones secuenciales a lo largo del tiempo y, así, capturar con precisión las dependencias temporales al evaluar el modelo con datos futuros no vistos inmediatamente después del entrenamiento.

La técnica es ideal para datos de series temporales univariadas y adecuada para evaluar la adaptabilidad de un modelo a patrones cambiantes y realizar predicciones secuenciales a lo largo del tiempo. Es valiosa para evaluaciones realistas que consideran las dependencias temporales y la predicción del futuro como los pronósticos financieros, donde la identificación temprana de problemas es crucial.

De acuerdo con (Brownlee, 2017), esta técnica consiste en los siguientes pasos:

- Seleccionar el número mínimo de observaciones necesarias para entrenar el modelo, el cual se conoce como el ancho de la ventana si se está ocupando una ventana deslizante.
- Decidir si el modelo se entrenará con todos los datos disponibles (ventana expandible) o solo con las observaciones más recientes (ventana deslizante).

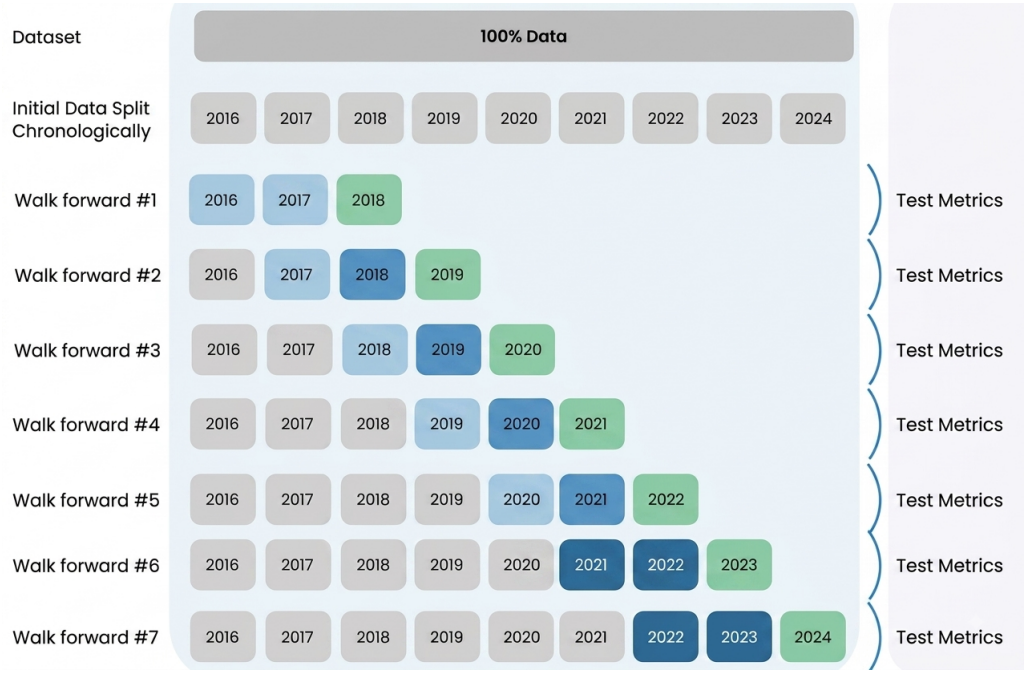


Figura 2.2: Validación hacia adelante (Walk Forward Validation).
Nota: Tomado de *Applied Machine Learning for Data Science Practitioners*, Subramanian (2025), John Wiley & Sons.

- Partiendo del inicio de la serie temporal, se utiliza el número mínimo de muestras de la ventana para entrenar el modelo.
- El modelo realiza una predicción para el siguiente paso temporal.
- La predicción obtenida se almacena o se compara con el valor real conocido.
- La ventana se expande para incluir el valor conocido y el proceso se repite.

2.8 Función de Autocorrelación Parcial (PACF)

La función de autocorrelación parcial mide la correlación única entre observaciones en diferentes rezagos (*lags*) eliminando los efectos de las variables intermedias. A diferencia de la función de autocorrelación (ACF), que captura la relación general entre observaciones, la PACF aísla la relación directa entre observaciones en retardos específicos.

Esta herramienta estadística es relevante en diversas disciplinas, como economía, finanzas, meteorología y otras, permitiendo a los analistas descubrir patrones ocultos y pronosticar tendencias futuras con mayor precisión. El área sombreada en azul que se observa en la Figura 2.3 representa

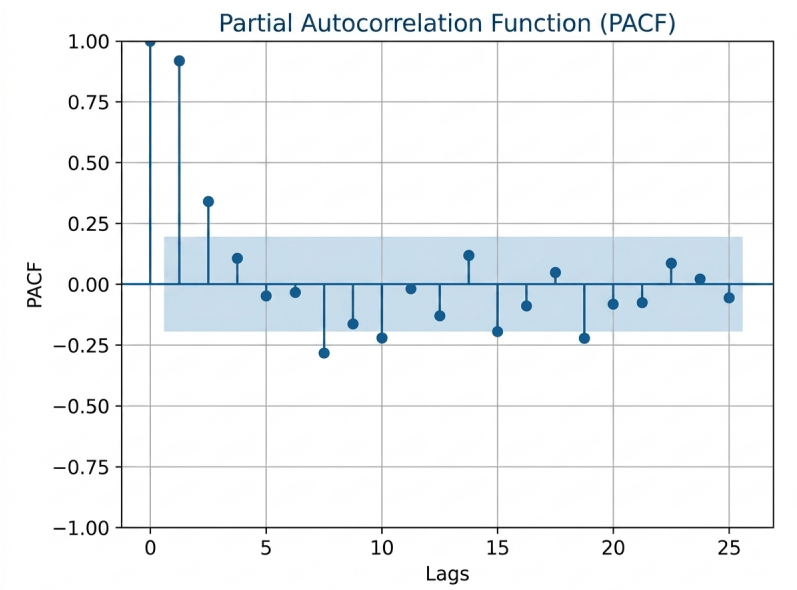


Figura 2.3: Gráfico PACF de datos de series temporales (conjunto de entrenamiento).
Nota. Tomado de *Essentials of Time Series Econometrics*, por Chopra (2025), Educohack Press.

el intervalo de confianza del $95\,\%$ o los límites de error; los retardos dentro del área son estadísticamente cercanos a cero. Por ejemplo, el pico en el retardo 0 muestra un valor de correlación de 1 e indica que una variable tiene correlación perfecta consigo misma. Adicionalmente, se puede observar que los retardos 1, 2 y 6 son los retardos significativos que se extienden fuera de los límites de error. Esta función es utilizada para determinar el número de *lags* necesarios para las variables objetivo. Según (Zakaria et al., 2023), en su investigación empleó la Función de Autocorrelación Parcial (PACF) como criterio técnico para seleccionar la estructura de retardos (*lags*) en la serie temporal del nivel del agua.

2.9 Prueba de Causalidad de Granger

Según (Sylvester et al., 2026), la prueba de causalidad de Granger se define como un procedimiento estadístico utilizado para determinar si los valores pasados de una variable Y pueden mejorar la predicción de otra variable X, es decir, que Y causa X en el sentido de Granger.

El criterio del valor p mínimo calcula la prueba de Granger para cada retardo (*lag*) candidato y selecciona aquel con el valor p más bajo, es decir, el de mayor significancia estadística. Este procedimiento permite identificar el número de *lags* en el que la variable exógena presenta la mayor capacidad predictiva sobre la variable dependiente u objetivo.

2.10 Estado del arte

El crecimiento empresarial constituye uno de los principales indicadores del desarrollo económico de un país. Diversas investigaciones han analizado la evolución de las empresas mediante enfoques estadísticos y modelos predictivos con el propósito de comprender los factores que influyen en su expansión y permanencia en el mercado. Van Witteloostuijn and Kolkman (2019) estudiaron la capacidad predictiva de diferentes técnicas para explicar el crecimiento empresarial. Utilizó una base de datos con $451{,}432$ registros correspondientes a $168{,}055$ empresas de Bélgica y los Países Bajos, abarcando de uno a seis años de historial por empresa. Se evaluó el crecimiento a través de la evolución de los activos totales y del número de empleados cuando no había datos financieros. A partir de 16 indicadores demográficos y financieros básicos, se generó un total de 113 variables predictoras. Se enfrentó al algoritmo de machine learning Random Forest contra dos herramientas tradicionales de la econometría, la regresión por mínimos cuadrados ordinarios y la regresión escalonada.

Los modelos lineales tradicionales mostraron un rendimiento con un R^2 de $0{,}05$ a $0{,}06$. En contraste, el modelo de Random Forest alcanzó un R^2 de $0{,}23$ en el conjunto de datos de entrenamiento y de $0{,}16$ en el conjunto de validación. Las variables más influyentes en orden de importancia para predecir el crecimiento fueron Activos Totales, Capital de Trabajo, Patrimonio Total, Ratio de Liquidez Corriente y Ratio de Solvencia.

El estudio concluye que el crecimiento empresarial no es un fenómeno aleatorio. La razón por la que la econometría tradicional falla no es la falta de datos, sino que el crecimiento se comporta como un sistema complejo. Este sistema posee sutiles relaciones no lineales e interacciones de alto orden que los modelos tradicionales son incapaces de ver, pero que el machine learning sí.

Posteriormente, Gangwani and Zhu (2024) analizaron de manera exhaustiva el estado del arte respecto al uso de la inteligencia artificial y el aprendizaje automático para modelar y predecir el éxito o fracaso empresarial. Los datos de los estudios analizados provienen de empresas que cotizan en bolsa en mercados de Estados Unidos de América, Europa y China, utilizando bases de datos comerciales masivas como Compustat y Orbis. Los datos obtuvieron de registros distribuidos en repositorios como Science Direct ($n = 3,303$), Springer ($n = 3,556$) e IEEE Xplore ($n = 988$). Agruparon las variables predictoras en tres vectores dimensionales: Variables de inversión como capital de riesgo, rondas de financiamiento, estrategias de innovación; variables endógenas del negocio tales como ratios financieros de balances contables, demografía corporativa, datos administrativos; y, variables exógenas de mercado como ciclos económicos, segmentación industrial y crecimiento sectorial.

La investigación desglosa el rendimiento y las limitaciones operativas de las tres algoritmos analizados. En primer lugar, al contrastar modelos lineales como la regresión logística frente a los modelos Random Forest y XGBoost se evidenció que los modelos de machine learning poseen una robustez significativamente superior para procesar matrices de datos corporativos estructurados.

Analizando enfoques como Análisis de Componentes Principales, Isomap y Local Linear Embedding, se destaca que reducir el número de variables de entrada optimiza drásticamente la convergencia de los modelos predictivos. Como métrica de referencia, se documentó que la aplicación de PCA permite contraer matrices complejas reduciendo, por ejemplo, de 14 ratios financieros a solo 5 componentes principales reteniendo el 91 % de la varianza total de los datos, lo que incrementa la estabilidad y exactitud de las proyecciones de quiebra o éxito de una empresa.

Por otro lado, los algoritmos de agrupamiento como k-Means y técnicas de detección de valores atípicos demostraron ser eficientes para emitir alertas tempranas sobre desviaciones en la salud financiera corporativa. Por su parte, las Redes Neuronales Convolucionales y la Memoria a Corto y Largo Plazo (LSTM) evidenciaron la mayor eficiencia cuando el objetivo predictivo involucra series temporales complejas de variables continuas o procesamiento de datos texturizados como minería de opiniones financieras y revisiones web.

Desde la perspectiva matemática, Guerrini et al. (2020) analizaron modelos de crecimiento económico con retardos temporales, demostrando que la incorporación de efectos de memoria puede modificar significativamente la dinámica del sistema y generar comportamientos imposibles de representar mediante modelos convencionales sin retardo.

De manera similar, Bobalová and Novotná (2021) desarrollaron un estudio orientado a describir y evaluar los procesos dinámicos con retraso temporal en la economía de las empresas. El propósito fue evidenciar que los fenómenos económicos reales como la producción, las inversiones y el consumo no generan impactos inmediatos, sino que se manifiestan con retrasos que deben ser resueltos mediante ecuaciones diferenciales funcionales. Para demostrar la aplicabilidad de estos sistemas dinámicos, analizaron dos escenarios prácticos: el primero, un sistema no lineal de inventarios y ventas adaptado a una empresa manufacturera del sector de servicios donde se demostró la importancia de la flexibilidad logística frente a la competencia; el segundo, un modelo de equilibrio entre oferta y demanda basado en la Ley de Walras aplicado al mercado avícola de la República Checa, donde se identificó un retraso estructural de un año en la capacidad de ajuste de la oferta frente a las fluctuaciones de precios. Las conclusiones principales destacan que la inclusión de retrasos temporales endógenos y exógenos en el modelado de negocios incrementa la precisión predictiva y dota a la gerencia de mejores herramientas de control y estabilización corporativa.

Más recientemente, Chen et al. (2025) propusieron enfoques híbridos entre aprendizaje auto-

mático y ecuaciones diferenciales con retardo, demostrando que es posible reconstruir modelos dinámicos complejos mediante técnicas basadas en datos, integrando interpretabilidad matemática y capacidad predictiva. En conjunto, estos estudios evidencian que tanto las ecuaciones diferenciales con retardo como los algoritmos de Machine Learning representan enfoques prometedores para modelar fenómenos económicos dinámicos. Sin embargo, existe una limitada aplicación de estas metodologías al contexto empresarial ecuatoriano, especialmente utilizando registros administrativos oficiales.

CAPÍTULO 3.
MARCO METODOLÓGICO

3.1 Diseño de investigación

La investigación adopta un diseño descriptivo, predictivo y comparativo. Es descriptivo porque analiza la evolución histórica del número de empresas y las ventas totales en Ecuador, identificando tendencias y patrones temporales. Es predictivo porque desarrolla modelos basados en ecuaciones diferenciales con retardo y algoritmos de machine learning para estimar el comportamiento futuro de las variables estudiadas. Además, posee un componente comparativo al evaluar y contrastar el desempeño de ambos enfoques mediante métricas de error y ajuste estadístico.

3.2 Enfoque metológico

El trabajo de titulación utiliza el método cuantitativo ya que se fundamenta en el análisis de datos numéricos correspondientes al crecimiento del número de empresas y venta totales de las mismas en Ecuador. El estudio busca evaluar la idoneidad predictiva y descriptiva de dos herramientas cuantitativas complejas, por un lado las ecuaciones diferenciales con retardo y por otro algoritmos de Machine Learning, aplicadas a una base de datos numéricos oficiales del Servicio de Rentas Internas. Se fundamenta en el procesamiento analítico de series temporales, la estimación de parámetros matemáticos y la validación a través de métricas estadísticas de error como MAE, MAPE y RMSE asegurando la total confiabilidad, objetividad y reproducibilidad de los resultados.

3.3 Metodología CRISP-DM

El proceso operativo en la Ciencia de Datos está estandarizado por la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), que se articula en seis fases principales, de las cuales tres se describen en este capítulo y las restantes en el siguiente.

3.3.1 Comprensión del negocio

El objetivo general del negocio es predecir el comportamiento crecimiento de sociedades, personas naturales con fines de lucro en Ecuador, utilizando información histórica y variables económicas relacionadas, con el propósito de disponer de una herramienta que permita analizar y anticipar cambios en la dinámica empresarial.

3.3.2 Comprensión de los datos

3.3.2.1. Recopilación de los datos

El conjunto de datos utilizado en esta investigación se recopiló del Servicio de Rentas Internas, con el fin de construir una base de datos y desarrollar modelos predictivos tanto del número de empresas activas como del total de ventas registradas.

Para analizar el número de empresas registradas en el país, los datos fueron obtenidos del Catastro del Registro Único de Contribuyente (RUC) por provincia, y los datos de las ventas totales se obtuvieron mediante la herramienta web de estadísticas multidimensionales del SRI llamada Saiku.

Según (Fernanda et al., 2024), considera que las fluctuaciones del tipo de cambio, los cambios en la política comercial y los eventos geopolíticos pueden afectar las ganancias, las operaciones de la cadena de suministro y el posicionamiento competitivo de las corporaciones multinacionales en los mercados internacionales.

Los factores macroeconómicos abarcan un amplio espectro de variables que reflejan la salud y el desempeño general de la economía. Estos factores incluyen el crecimiento del PIB, las tasas de inflación, las tasas de interés, los tipos de cambio, las tasas de desempleo y los indicadores de política fiscal. Cada una de estas variables ejerce influencias distintas sobre el desempeño financiero de las empresas a través de diversos canales. Por ejemplo, los cambios en las tasas de interés de los bancos centrales pueden afectar los costos de endeudamiento de las empresas, influyendo en el gasto en inversiones, los gastos de capital y los índices de apalancamiento financiero.

De esta forma, se recopilaron datos de variables macroeconómicas consideradas importantes para explicar el comportamiento de las variables objetivo como

- el Crecimiento del Producto Interno Bruto(en %), obtenido del Banco Central del Ecuador (BCE) ,
- la Inflación (en %), tomado del Banco Central del Ecuador y del Instituto Nacional de Estadística y Censos (INEC),

- las Tasas de Interés Productivo Corporativo, Empresarial, PYMES y Microcréditos (en %) obtenidos del Banco Central del Ecuador y del documento Noventa Años del BCE, y
- la Tasa de Desempleo (en %), obtenido del Banco Mundial, BCE e INEC.

3.3.2.2. Descripción del conjunto de datos

El conjunto de datos del Catastro del RUC comprende observaciones desde 1950 hasta 2026, correspondientes a las 24 provincias del Ecuador.

NUMERO_RUC	RAZON_SOCIAL	CODIGO_JURISDICCION	ESTADO_CONTRIBUYENTE	CLASE_CONTRIBUYENTE	FECHA_INICIO_ACTIVIDADES	FECHA_ACTUALIZACION	FECHA_SUSPENSION_DEFINITIVA	FECHA_REINICIO_ACTIVIDADES	OBLIGADO	...
17223	GUALOTUÃ□A BARAHONA	PICHINCHA	SUSPENDIDO	GEN	2019-03-20 00:00:00	2021-02-22 20:04:00	2021-02-22 20:04:00	NaN	N	...
17228	ANDRADE	PICHINCHA	SUSPENDIDO	GEN	2016-10-27 00:00:00	2019-07-03 11:50:33	2020-07-26 00:00:00	NaN	N	...
17261	CORDERO VINUEZA	PICHINCHA	ACTIVO	GEN	2023-01-27 00:00:00	2025-11-01 15:50:06	2024-10-29 16:10:18	2025-11-01 09:10:04	N	...
4500	PALMA ARBOLEDA	PICHINCHA	SUSPENDIDO	GEN	2019-03-20 00:00:00	2019-05-15 00:00:00	2020-12-12 19:10:02	NaN	N	...
17516	TUQUERRES CAISALETIN	PICHINCHA	SUSPENDIDO	GEN	2019-02-15 00:00:00	NaN	2020-03-22 00:00:00	NaN	N	...

Figura 3.1: Formato original de los registros del Catastro del RUC de la Provincia de Pichincha.
Nota. Tomado de *DataSets*, por el Servicio de Rentas Internas.

En el registro que se muestra en la Figura 3.1 se considera el Tipo de Contribuyente (clasificado en Personas Naturales y Sociedades de la provincia de Pichincha), las Fechas de Inicio y Fin de actividades, el Estado del Contribuyente, el Estado del Establecimiento y el Código CIIU. Con esa información se determinan las variables objetivo: el Número de Personas Naturales y Sociedades Activas con fines de lucro.

	YEAR	MES	TOTAL_VENTAS_EXPORTACIONES
0	2020	1	14895736557
1	2020	2	14264977512
2	2020	3	11867942042
3	2020	4	8825129434
4	2020	5	9929589750
5	2020	6	14902808913
6	2020	7	12539967052

Figura 3.2: Formato original de los registros del Total de Ventas y Exportaciones.
Nota. Tomado de *SAIKU*, por el Servicio de Rentas Internas.

Mediante la plataforma analítica y de inteligencia de negocios SAIKU del SRI se obtuvo los datos mensuales del total de ventas desde enero de 2020 a mayo de 2026, como se muestra en la Figura 3.2. De esta forma, se determina la tercera variable objetivo.

Las variables exógenas fueron obtenidas de fuentes oficiales y su incorporación permite capturar el efecto del entorno económico sobre la evolución del número de sociedades activas y ventas totales registradas en el país. En la Figura 3.3 se muestra la inflación del Ecuador desde 2021 hasta 2026.



Figura 3.3: Evolución de la inflación mensual, anual y acumulada en Ecuador (2021-2026).
Nota. Tomado de *Servicio de Información Económica: Inflación*, por el Banco Central del Ecuador

.

La información de las variables exógenas proveniente de las distintas instituciones contempla datos mensuales desde el año 2007. Basados en documentos históricos y otras bases de datos, se logró obtener información desde el año 2000 para completar la serie hasta 2007. De esta manera, se determinó el número de empresas activas desde 1950 hasta 1999; sin embargo, el periodo de estudio inicia a partir del año 2000, debido a la adopción de la dolarización en el país el 9 de enero de ese mismo año. De esta forma, el conjunto de datos comprende 316 observaciones mensuales correspondientes al período enero de 2000 a abril de 2026.

3.3.2.3. Calidad de los datos

Los registros del catastro del RUC de las 24 provincias no presentaron valores faltantes o nulos. Las omisiones presentes en la Figura 3.4 corresponden al año y mes de fin de actividades; es decir, los valores faltantes que aparecen se refieren a contribuyentes de la Provincia de Manabí que aún se mantienen activos. El siguiente código permitió identificar valores nulos en el catastro del SRI,

```
columnas=['NUMERO_RUC','ESTADO_ESTABLECIMIENTO','ESTADO_CONTRIBUYENTE',
   'TIPO_CONTRIBUYENTE','AÑO_INICIO_ACTIVIDADES',
```

```
    'MES_INICIO_ACTIVIDADES','AÑO_FIN_ACTIVIDADES','MES_FIN_ACTIVIDADES']
# Contamos cuántos NAs hay en cada columna
nas_por_columna = df_manabi2[columnas1].isna().sum()
# Contamos cuántos ceros reales hay en cada columna
ceros_por_columna = (df_manabi2[columnas] == 0).sum()
# Unimos ambos resultados en una tabla  para comparar
auditoria = pd.DataFrame({'Total NAs': nas_por_columna,
    'Total Ceros': ceros_por_columna})
print(auditoria)
```

	Total NAs	Total Ceros
NUMERO_RUC	0	0
ESTADO_ESTABLECIMIENTO	0	0
ESTADO_CONTRIBUYENTE	0	0
TIPO_CONTRIBUYENTE	0	0
AÑO_INICIO_ACTIVIDADES	0	0
MES_INICIO_ACTIVIDADES	0	0
AÑO_FIN_ACTIVIDADES	204196	0
MES_FIN_ACTIVIDADES	204196	0

Figura 3.4: Identificación de valores nulos y ceros en el catastro del RUC de la Provincia de Manabí.

Cabe mencionar que el SRI codifica con el valor 9999999 a aquellos contribuyentes de los cuales se carece de información específica sobre su actividad económica.

ESTADO_ESTABLECIMIENTO	DESCRIPCION_PROVINCIA_EST	DESCRIPCION_CANTON_EST	DESCRIPCION_PARROQUIA_EST	CODIGO_CIIU	ACTIVIDAD_ECONOMICA	AGENTE_RETENCION	ESPECIAL
CER	MANABI	MANTA	TARQUI	9999999	VERIFICAR	N	N

Figura 3.5: Registro del RUC con código 9999999 correspondiente a actividades no especificadas.

3.3.2.4. Análisis Exploratorio de Datos

En la Figura 3.6 se observa una tendencia creciente de las Sociedades Activas y manteniendo un crecimiento continuo durante prácticamente todo el período analizado. Aunque el comportamiento de las Personas Naturales Activas es similar, a partir de 2016 empieza a decrecer, sufriendo una fuerte caída en 2020 y logrando una ligera estabilidad en 2022. Para abril de 2026 se contabilizan 214 924 sociedades activas y 1 421 033 personas naturales activas.

En la Figura 3.7 se observa que en el año 2020, el nivel base de ventas se situaba entre 8 y 15 mil millones de USD, registrando un mínimo notable alrededor en abril de 2020 debido al impacto de

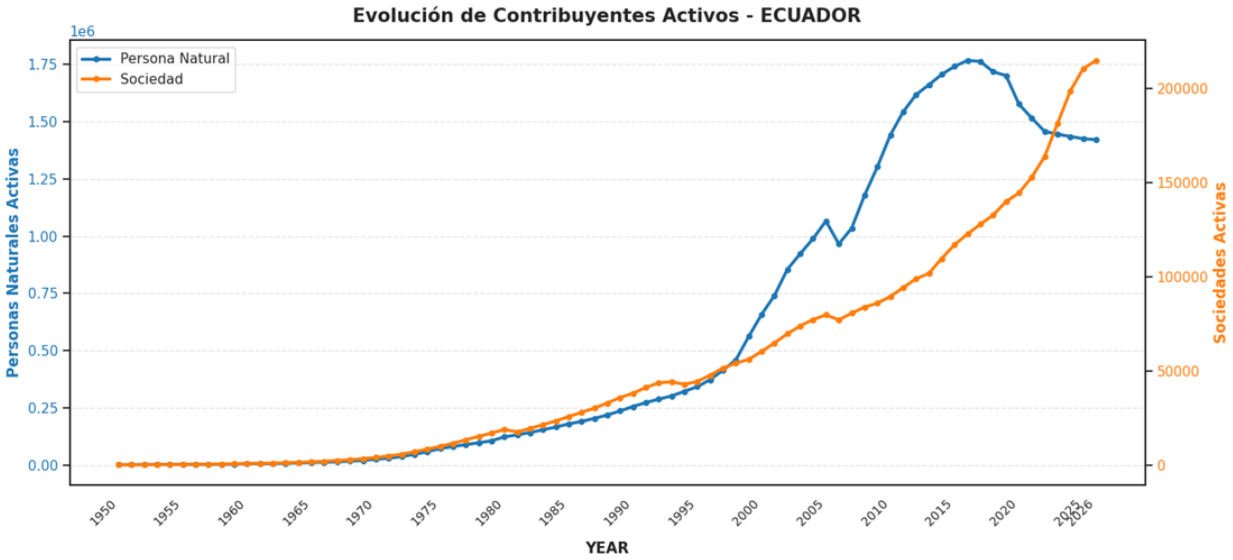


Figura 3.6: Evolución del número de contribuyentes activos en Ecuador (1950–2026).

la pandemia de COVID-19, donde cayó cerca de $8,8$ mil millones de USD. Los meses sin picos ha mostrado una clara tendencia ascendente, pasando de 13 a 15 mil millones de USD en 2020 a 2026.

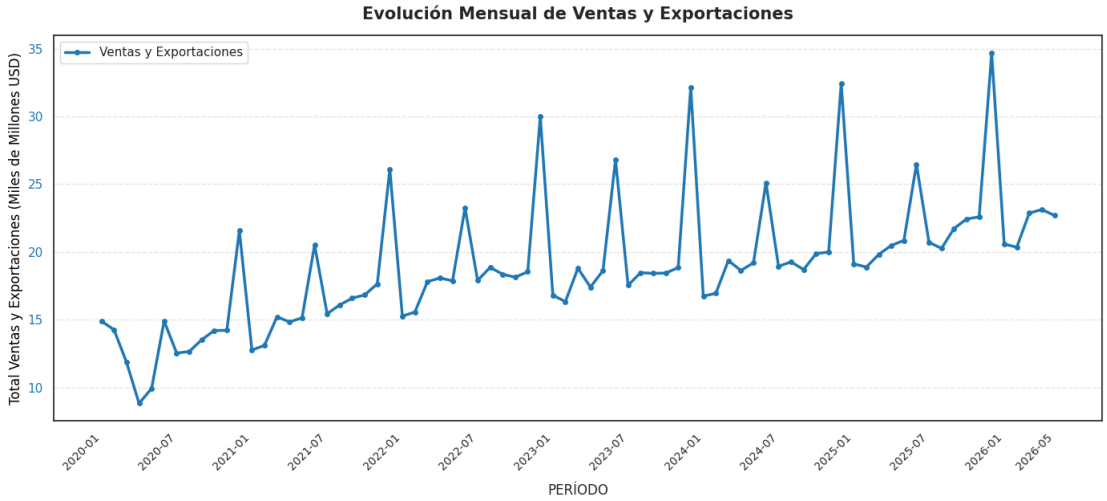


Figura 3.7: Evolución de las ventas totales y exportaciones.

La Figura 3.7 muestra una fuerte estacionalidad intraanual, en la que los picos ocurren en los meses de Diciembre y Junio; las ventas alcanzan entre 20 y 34 mil millones de USD en diciembre, mientras que en junio entre 15 y 26 mil millones de USD. Por otra parte, se registra una contracción en las ventas y exportaciones a inicios de cada año, durante Enero y Febrero, iniciando su recuperación a

partir del mes de Marzo.

Entre las magnitudes calculadas corresponde a los flujos netos anuales de sociedades y personas naturales, definidos por

$$F_t = I_t - C_t$$

para todo $t \in \{1950, \ldots, 2026\}$, donde:

- F_t representa el flujo neto de sociedades o personas naturales en el año t,
- I_t es la cantidad de sociedades o personas naturales que iniciaron actividades durante el año t, y
- C_t es la cantidad de sociedades o personas naturales que finalizaron actividades durante el año t.

Los estadísticos descriptivos para el flujo neto anual de Sociedades, Personas Naturales y Ventas se presentan en la Tabla 3.1.

Tabla 3.1: Estadísticos descriptivos.

Estadísticos	Flujo Sociedades	Flujo Personas Naturales	Ventas y Exportaciones
Años	77	77	7
Promedio	2 791,22	18 454,97	18 894 183 843,51 USD
Desviación estándar	3 547,58	45 119,74	4 742 565 109,29 USD
Mínimo	−2 603,00	−124 404,00	8 825 129 434,00 USD
Máximo	17 467,00	143 447,00	34 672 361 045,00 USD

El flujo de Sociedades presenta un promedio de $2\,791,22$ sociedades, con una desviación estándar de $3\,547{,}58$, lo que evidencia una alta variabilidad respecto a su valor medio. El valor mínimo es de $-2\,603$ sociedades indica que el número de sociedades dejaron de operar y superó al número de nuevas sociedades registradas. Por otro lado, el flujo de Sociedades alcanza un valor máximo de $17\,467$ sociedades. Esta diferencia considerable entre el mínimo y el máximo evidencia cambios importantes en el comportamiento de esta variable a lo largo del período analizado.

Para el flujo de Personas Naturales, se observa un promedio de $18\,454{,}97$, acompañado de una desviación estándar de $45\,119{,}74$, lo que indica una elevada dispersión y variabilidad interanual. El flujo mínimo registrado es de $-124\,404$, mientras que el máximo alcanza $143\,447$. La presencia

tanto de valores negativos como positivos muestra que esta variable puede experimentar importantes fluctuaciones entre períodos, lo que sugiere la existencia de años con una reducción significativa y otros con un crecimiento considerable en el número de personas naturales.

Por otra parte, las Ventas y Exportaciones presenta un promedio anual de aproximadamente 18 894 millones de USD, mientras que la desviación estándar alcanza 4 742 millones de USD. El valor mínimo es de aproximadamente 8 825 millones de USD y el máximo de 34 672 millones de USD. Estos valores muestran una diferencia importante entre los niveles mínimos y máximos de ventas y exportaciones, evidenciando una evolución considerable de la actividad económica durante el período analizado.

En la Figura 3.8 se evidencia que, a partir de la década de 1970, el flujo neto del número de empresas en el país comienza a crecer. Este comportamiento coincide con el conocido boom petrolero, el cual dio inicio a un cambio en la estructura económica ecuatoriana. Debido al incremento del precio y volumen de las exportaciones petroleras, varios sectores económicos se expandieron durante ese período (Banco Central del Ecuador, 1998).

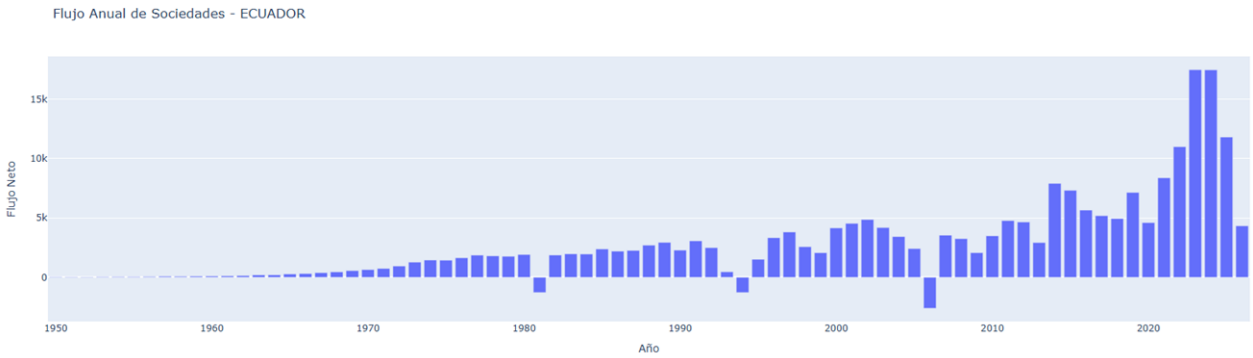


Figura 3.8: Evolución del flujo neto anual de las Sociedades en Ecuador.

Un decrecimiento en el flujo neto de Sociedades ocurre en 1981 y 1990 que podría atribuirse a la crisis de la deuda externa, la desaceleración económica y la implementación de políticas de ajuste estructural, factores que afectaron la inversión y la actividad empresarial (Oleas-Montalvo, 2017).

En las Figuras 3.8 y 3.9 se observa una disminución considerable del flujo neto en el año 2006, con 100 257 registros menos de Personas Naturales y 2 603 menos de Sociedades. Esta disminución puede asociarse a la aplicación de la Resolución NAC-DGER-2006-0253 publicada en el Registro Oficial del 11 de mayo de 2006 en el que resuelve

- Cancelar el registro único de contribuyentes de las sociedades que, de acuerdo a la información proporcionada por la Superintendencia de Compañías, se evidencie la inscripción de la resolución de cancelación en el correspondiente Registro Mercantil,

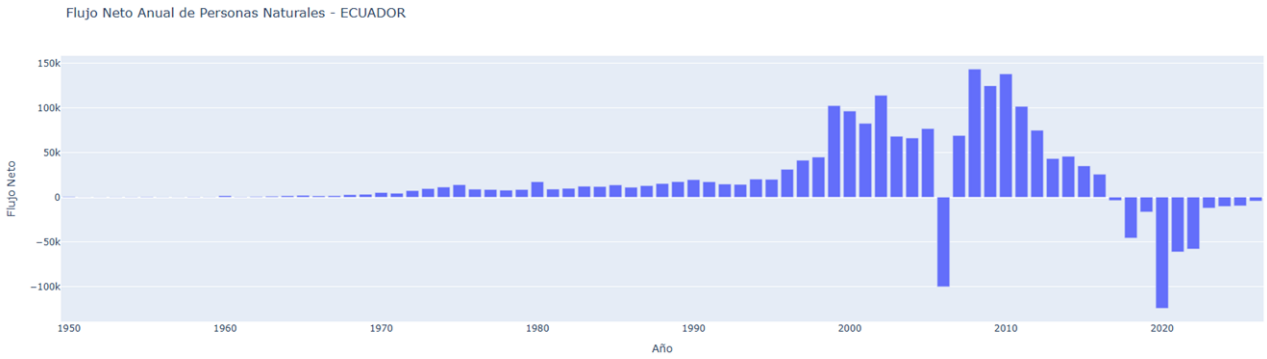


Figura 3.9: Evolución del flujo neto anual de las Personas Naturales en Ecuador.

- Cancelar el registro único de contribuyentes de las personas naturales que hayan fallecido,
- Suspender el registro único de contribuyentes de las personas naturales que se encuentren ausentes del país por un período superior a dos años continuos,
- Suspender el registro único de contribuyentes de las personas naturales que no hayan efectuado actividad económica en los últimos tres años; y,
- Cancelar el registro único de contribuyentes de las sociedades que no se encuentren sujetas al control de la Superintendencia de Bancos o Compañías (Corte Constitucional del Ecuador, 2006).

Esta depuración se confirma mediante el artículo periodístico de (El Universo, 2006) en el que menciona que el director del SRI, Alberto Cárdenas, aplicando la Resolución NAC-DGER-2006-0253, resolvió cancelar los RUC de 13 000 sociedades por no estar en el Registro Mercantil o por no realizar actividad económica en los últimos dos años y de 7 000 contribuyentes correspondiente a personas fallecidas.

De manera similar, actuaron en agosto de 2018; según el Servicio de Rentas Internas (2019) se realizó la depuración de 244 976 contribuyentes que no realizaron actividad económica.

En el año 2020 también se observa una disminución en el flujo de contribuyentes coincidiendo con los efectos económicos que produjo la pandemia de Covid-19. Según (Coba, 2021), un total de 3 494 sociedades económicas cerraron en el país y, de acuerdo con (Diario Expreso, 2020), alrededor del 25 % de emprendedores tuvieron que cerrar sus negocios.

Como se aprecia en la Figura 3.8, el flujo neto de las Sociedades registró un crecimiento acelerado, pasando de 4 597 en 2020 a 17 467 en 2023. Este comportamiento coincide con la implementación, en mayo de 2020, de un nuevo sistema de constitución de compañías a través de *Sociedad por*

Acciones Simplificadas (SAS) por parte de la Superintendencia de Compañías, Valores y Seguros (Franco, 2022). Además, la creación de sociedades ha crecido debido al lavado de activos que representa el $4\,\%$ del PIB y uno de los canales más frecuente es la inversión en bienes raíces. Según (Primicias, 2026), la compraventa de terrenos y el desarrollo de proyectos inmobiliarios posibilita la incorporación de importantes sumas de dinero al sistema económico, especialmente cuando se utilizan pagos en efectivo o estructuras societarias que obstaculizan la identificación del beneficiario final.

Asimismo, la Tasa de Crecimiento Mensual se calculó para el periodo comprendido entre enero de 2000 y mayo de 2026, definida como

$$\text{Tasa}(t) = \frac{\text{Sociedades Activas}(t) - \text{Sociedades Activas}(t-1)}{\text{Sociedades Activas}(t-1)}$$

donde t representa cada mes del periodo analizado.

La evolución de la tasa de crecimiento de contribuyentes activos en Ecuador desde 2000 a 2026 refleja una dinámica caracterizada por la estabilidad a largo plazo, interrumpida por eventos puntuales de depuración por parte del SRI como se observa en la Figura 3.10. El estudio se centra en este periodo debido a la información obtenida de las diferentes variables exógenas.

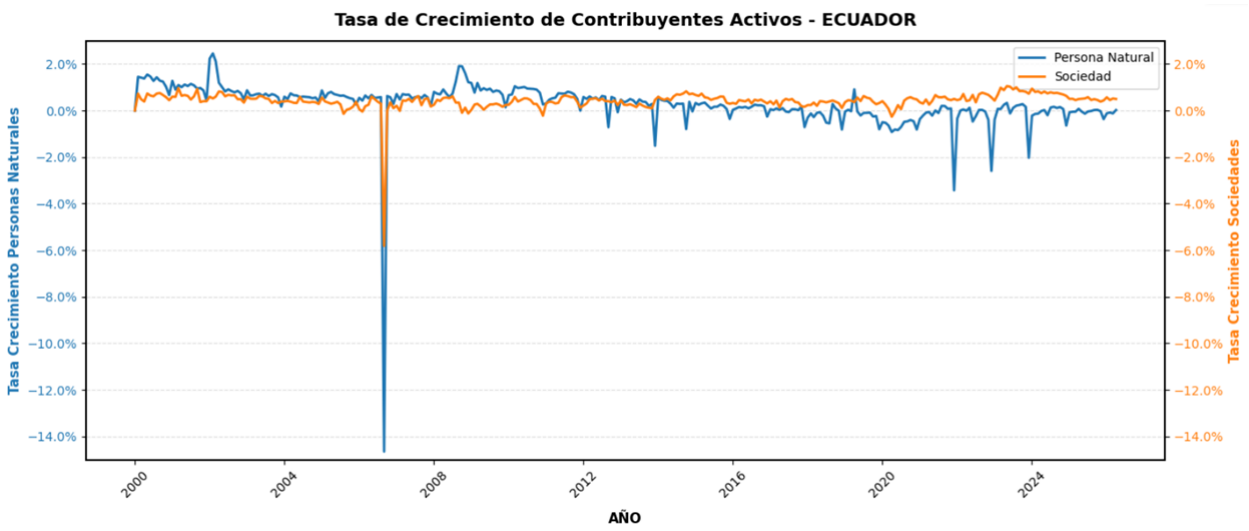


Figura 3.10: Tasa de crecimiento mensual de contribuyentes activos en Ecuador (2000–2026).

El segmento de Sociedades Activas mantiene un comportamiento estable con fluctuaciones leves, mientras que el grupo de Personas Naturales Activas evidencia una volatilidad superior. En el año 2006, se registró una caída de aproximadamente el $-15\,\%$ en Personas Naturales y del $-6\,\%$ en Sociedades, hecho atribuible a procesos masivos de depuración de registros inactivos del RUC.

Asimismo, en el periodo reciente (2020–2024), se observan contracciones puntuales periódicas en el segmento de Personas Naturales.

3.3.3 Preparación de datos

3.3.3.1. Limpieza y depuración de la información

Como se muestra en la Figura 3.1, el formato original de los registros del catastro del RUC consta de 21 columnas que son

- NUMERO_RUC: número único que identifica las personas y a las empresas que realizan actividades económicas.
- RAZON_SOCIAL: nombre oficial y legal de la empresa.
- CODIGO_JURISDICCION: provincia donde desarrolla sus actividades económicas.
- ESTADO_CONTRIBUYENTE: situación en la que se encuentra el contribuyente (Activo, Pasivo o Suspendido).
- CLASE_CONTRIBUYENTE: contribuyentes de regimen general (GEN), RIMPE (RMP), especial (ESP), etc.
- FECHA_INICIO_ACTIVIDADES: fecha en la que el contribuyente declara haber comenzado sus operaciones económicas.
- FECHA_ACTUALIZACION: fecha en la que el contribuyente realizó cambios o confirmó la información registrada en su RUC.
- FECHA_SUSPENSION_DEFINITIVA: fecha en la que el contribuyente notificó al SRI que dejó de realizar operaciones económicas temporal o definitivamente.
- FECHA_REINICIO_ACTIVIDADES: fecha en la que un contribuyente que se encontraba en estado pasivo o suspendido reactiva su RUC y empieza a operar económicamente.
- OBLIGADO: informa si el contribuyente está legalmente obligado a llevar contabilidad de su actividad económica.
- TIPO_CONTRIBUYENTE: naturaleza jurídica o personería del titular del registro (Persona Natural y Sociedad).

- NUMERO_ESTABLECIMIENTO: numérico que se asigna a cada sucursal.
- NOMBRE_FANTASIA_COMERCIAL: nombre de marca o rótulo comercial público.
- ESTADO_ESTABLECIMIENTO: situación operativa actual de un local (Abierto o Cerrado).
- DESCRIPCION_PROVINCIA_EST: provincia en la que se encuentra ubicado el establecimiento.
- DESCRIPCION_CANTON_EST: cantón en la que se encuentra ubicado el establecimiento.
- DESCRIPCION_PARROQUIA_EST: parroquia en la que se encuentra ubicado el establecimiento.
- CODIGO_CIIU: la Clasificación Industrial Internacional Uniforme es un código numérico utiliza el SRI para clasificar la actividad económica de los contribuyentes.
- ACTIVIDAD_ECONOMICA: consiste en el trabajo, negocio o servicio que realiza el contribuyente para generar ingresos.
- AGENTE_RETENCION: contribuyente al que la ley lo faculta y obliga a descontar un porcentaje de impuesto (IVA o Impuesto a la Renta).
- ESPECIAL: contribuyentes élite compuesto por un número reducido de personas naturales y empresas que, por su elevado volumen de ventas.

El siguiente código permitió crear las columnas de año y mes de inicio y fin de actividades mediante la extracción de esa información de las fechas correspondientes. Por ejemplo, para la Provincia de Tungurahua tenemos

```
df_tungurahua['AÑO_INICIO_ACTIVIDADES'] =
   pd.to_datetime(df_tungurahua['FECHA_INICIO_ACTIVIDADES'],
   errors='coerce').dt.year.astype('Int64')
#
df_tungurahua['MES_FIN_ACTIVIDADES'] =
   pd.to_datetime(df_tungurahua['FECHA_SUSPENSION_DEFINITIVA'],
   errors='coerce').dt.month.astype('Int64')
```

Seleccionamos los campos requeridos para calcular el número de personas naturales y sociedades con periodicidad anual y mensual,

```
df_tungurahua2 = df_tungurahua[['NUMERO_RUC','ESTADO_CONTRIBUYENTE',
        'TIPO_CONTRIBUYENTE', 'ESTADO_ESTABLECIMIENTO','CODIGO_CIIU',
        'AÑO_INICIO_ACTIVIDADES','MES_INICIO_ACTIVIDADES',
        'AÑO_FIN_ACTIVIDADES','MES_FIN_ACTIVIDADES']].copy()
```

Como resultado de este procedimiento, se obtiene un dataframe con las columnas de año y mes separadas.

TIPO_CONTRIBUYENTE	ESTADO_ESTABLECIMIENTO	CODIGO_CIIU	AÑO_INICIO_ACTIVIDADES	MES_INICIO_ACTIVIDADES	AÑO_FIN_ACTIVIDADES	MES_FIN_ACTIVIDADES
PERSONA NATURAL	CER	L682002	1973	1	2019	11
PERSONA NATURAL	CER	C161002	1984	11	<NA>	<NA>
PERSONA NATURAL	ABI	A012902	2009	5	<NA>	<NA>
PERSONA NATURAL	CER	G477205	1974	1	2006	9
PERSONA NATURAL	CER	S960907	2017	3	2018	2

Figura 3.11: Descomposición de las fechas en columnas independientes de año y mes.

A continuación, se determina el período a partir del cual se inicia el conteo del número de sociedades y personas naturales,

```
df_tungurahua3 = df_tungurahua2[(
    df_tungurahua2['AÑO_INICIO_ACTIVIDADES'] >= 1950)&
    ((df_tungurahua2['AÑO_FIN_ACTIVIDADES'] >= 1950)
       |(df_tungurahua2['AÑO_FIN_ACTIVIDADES'].isna())) &
    (~df_tungurahua2['CODIGO_CIIU'].astype(str).str.contains('999999'))]
```

El procedimiento anterior se llevó a cabo debido a que se identificaron registros con años inconsistentes, como se observa en la Figura 3.12.

NUMERO_RUC	RAZON_SOCIAL	CODIGO_JURISDICCION	ESTADO_CONTRIBUYENTE	CLASE_CONTRIBUYENTE	FECHA_INICIO_ACTIVIDADES
1860000640001	GOBIERNO AUTONOMO DESCENTRALIZADO MUNICIPAL DE...	TUNGURAHUA	ACTIVO	GEN	1860-07-31 00:00:00
1860000210001	GOBIERNO AUTONOMO DESCENTRALIZADO MUNICIPALIDA...	TUNGURAHUA	ACTIVO	GEN	1756-09-20 00:00:00
1860000210001	GOBIERNO AUTONOMO DESCENTRALIZADO MUNICIPALIDA...	TUNGURAHUA	ACTIVO	GEN	1756-09-20 00:00:00

Figura 3.12: Detección de anomalías en la fecha de inicio de actividades.

Cabe mencionar que existen contribuyentes que tienen un único RUC, pero varios establecimientos asociados. Por ejemplo, el campo NUMERO_ESTABLECIMIENTO asigna el valor 1 a la matriz,

el valor 2 a la primera sucursal y así sucesivamente. Como consecuencia, un mismo contribuyente puede estar representado por múltiples registros en el catastro, lo que incrementa el número total de registros. Debido a esto, mediante el siguiente código se obtiene un único registro por cada RUC, dando prioridad a los contribuyentes que no tienen fecha de fin de actividades y cuyo establecimiento se encuentra abierto,

```
    df_tungurahua_ruc_unico=(df_tungurahua3.assign(
        ColumnaExtra=df_tungurahua3['AÑO_FIN_ACTIVIDADES'].isna()&
        (df_tungurahua3['ESTADO_ESTABLECIMIENTO'] == 'ABI'))
        .sort_values(by=['NUMERO_RUC','ColumnaExtra'],ascending=[True,
           False]).drop_duplicates(subset=['NUMERO_RUC'],keep='first').
        drop(columns=['ColumnaExtra']).sort_index())
```

Dado que el objetivo del estudio es analizar el crecimiento empresarial, únicamente se consideraron las personas naturales y las sociedades que desarrollan actividades con fines de lucro. Debido a esto, se excluyeron entidades sin fines de lucro, tales como fundaciones, partidos políticos, instituciones educativas, entre otras. La clasificación se realizó con base en las secciones de la Clasificación Nacional de Actividades Económicas (CIIU). El siguiente código implementa dicha clasificación,

```
letras_con_lucro = ['A','B','C','F','G','H','I','J',
    'K','L','M','N','R']
df_tungurahua_con_lucro = df_tungurahua_ruc_unico[
    (df_tungurahua_ruc_unico['AÑO_INICIO_ACTIVIDADES'] >= 1950) &
    ((df_tungurahua_ruc_unico['AÑO_FIN_ACTIVIDADES'] >= 1950) |
       (df_tungurahua_ruc_unico['AÑO_FIN_ACTIVIDADES'].isna())) &
    (df_tungurahua_ruc_unico['CODIGO_CIIU'].astype(str).str[0].str.upper().
    isin(letras_con_lucro))]
```

Finalmente, el siguiente código permite contabilizar mensualmente el número de personas naturales y sociedades que iniciaron y finalizaron sus actividades.

```
# Filtramos los datos
df_base_tungurahua = df_tungurahua_con_lucro[
    df_tungurahua_con_lucro['TIPO_CONTRIBUYENTE'].isin(['PERSONA
       NATURAL', 'SOCIEDAD'])
    ].copy()
# Realizamos el conteo de Inicio de Actividades (>= 2000)
inicios = df_base_tungurahua[df_base_tungurahua[
    'AÑO_INICIO_ACTIVIDADES']>=2000].groupby(['AÑO_INICIO_ACTIVIDADES',
```

```
    'MES_INICIO_ACTIVIDADES','TIPO_CONTRIBUYENTE']).size().
    unstack(fill_value=0)
inicios.columns = [f'{col.replace(" ", "_")}_INICIO' for col in
   inicios.columns]
inicios.index.names = ['YEAR', 'MES']
# Realizamos el conteo de Fin de Actividades (>= 2000)
cierres = df_base_tungurahua[df_base_tungurahua[
    'AÑO_FIN_ACTIVIDADES'] >= 2000].groupby(['AÑO_FIN_ACTIVIDADES',
    'MES_FIN_ACTIVIDADES','TIPO_CONTRIBUYENTE']).size().
    unstack(fill_value=0)
cierres.columns = [f'{col.replace(" ", "_")}_FIN' for col in
   cierres.columns]
cierres.index.names = ['YEAR','MES']
# Unimos ambas tablas y formamos en un solo DataFrame
df_final_tungurahua = pd.concat([inicios,cierres],
   axis=1).fillna(0).astype(int).reset_index()
# Ordenamos por año y mes
df_final_tungurahua =df_final_tungurahua.sort_values(by=['YEAR',
    'MES']).reset_index(drop=True)
# Mostramos el resultado
df_final_tungurahua.head(12)
# Guardamos el dataframe
from google.colab import drive
drive.mount('/content/drive')
df_final_tungurahua.to_csv('/content/drive/MyDrive/Tesis_PUCE/Programas/
    df_tungurahua_con_lucro_2000.csv',index=False,encoding='utf-8-sig')
```

A continuación, en la Figura 3.13 se presenta una parte del DataFrame correspondiente a la provincia de Tungurahua.

Este procedimiento se aplicó a las 24 provincias del Ecuador para generar el DataFrame de cada una de ellas y, posteriormente, consolidar la información en un único DataFrame a nivel nacional. Adicionalmente, para cada provincia se calcularon el número de personas naturales y sociedades activas, las tasas de crecimiento y el flujo neto. Las gráficas correspondientes se presentan en los anexos.

	YEAR	MES	PERSONA_NATURAL_INICIO	SOCIEDAD_INICIO	PERSONA_NATURAL_FIN	SOCIEDAD_FIN
0	2000	1	456	10	30	3
1	2000	2	241	5	25	1
2	2000	3	258	11	26	2
3	2000	4	264	10	26	2
4	2000	5	317	7	32	0
5	2000	6	282	5	36	1
6	2000	7	338	8	30	2
7	2000	8	373	7	34	2
8	2000	9	352	19	24	0
9	2000	10	295	10	35	3
10	2000	11	383	9	31	4
11	2000	12	315	10	58	6

Figura 3.13: Consolidado mensual del inicio y cese de actividades económicas según la naturaleza del contribuyente (Provincia de Tungurahua).

3.3.3.2. Tratamiento de valores nulos

Los registros del catastro del SRI correspondientes a cada provincia no presentaron valores nulos. Esto se evidencia en el caso de la provincia de Manabí, como se mostró en la Figura 3.4.

3.3.3.3. Integración de las fuentes de datos

Los 24 DataFrames fueron unificados mediante el siguiente código

```
from google.colab import drive
drive.mount('/content/drive')
ruta_base = '/content/drive/MyDrive/Tesis_PUCE/Programas/'
# Lista de las provincias
archivos_provincias = [
    'tabla_azuayCFL.csv',              #01
    'tabla_bolivarCFL.csv',            #02
    'tabla_canarCFL.csv',              #03
    'tabla_carchiCFL.csv',             #04
    'tabla_chimborazoCFL.csv',         #05
    'tabla_cotopaxiCFL.csv',           #06
    'tabla_elOroCFL.csv',              #07
    'tabla_esmeraldasCFL.csv',         #08
```

```
    'tabla_galapagosCFL.csv',          #09
    'tabla_guayasCFL.csv',             #10
    'tabla_imbaburaCFL.csv',           #11
    'tabla_lojaCFL.csv',               #12
    'tabla_los_riosCFL.csv',           #13
    'tabla_manabiCFL.csv',             #14
    'tabla_moronaCFL.csv',             #15
    'tabla_napoCFL.csv',               #16
    'tabla_orellanaCFL.csv',           #17
    'tabla_pastazaCFL.csv',            #18
    'tabla_pichinchaCFL.csv',          #19
    'tabla_santaElenaCFL.csv',         #20
    'tabla_santoDCFL.csv',             #21
    'tabla_sucumbiosCFL.csv',          #22
    'tabla_tungurahuaCFL.csv',         #23
    'tabla_zamoraCFL.csv',             #24
]
# Cargamos y unimos todos los archivos en una sola lista
lista_df_provincias = []
#
for archivo in archivos_provincias:
    ruta_completa = f"{ruta_base}{archivo}"
    try:
        df_prov = pd.read_csv(ruta_completa)
        lista_df_provincias.append(df_prov)
    except FileNotFoundError:
        print(f"Advertencia: No se encontró el archivo {archivo} en la
           ruta especificada.")
# Combinamos los datos de todas las provincias en un solo DataFrame
df_consolidado = pd.concat(lista_df_provincias,ignore_index=True)
```

El código anterior, junto con varias líneas adicionales, permitió consolidar el DataFrame correspondiente al Ecuador. A partir de este, se calcularon los flujos netos, el número total de personas naturales y sociedades activas, así como las respectivas tasas de crecimiento. Los resultados obtenidos se presentan en la Figura 3.14.

El DataFrame nacional anual permitió analizar la información acumulada hasta 1999. A partir del año 2000, la información se encuentra disponible con periodicidad mensual, lo que permitió realizar

YEAR	FLUJO_NETO_PN	PERSONA_NATURAL_ACTIVAS	TASA_CRECIMIENTO_PN_%	FLUJO_NETO_SOCIEDAD	SOCIEDAD_ACTIVAS	TASA_CRECIMIENTO_SOCIEDAD_%
1950	604	604	0	37	37	0
1951	126	730	20	34	71	91
1952	250	980	34	31	102	43
1953	312	1292	31	43	145	42
1954	363	1655	28	49	194	33
1955	496	2151	29	47	241	24
1956	281	2432	13	63	304	26
1957	235	2667	9	85	389	27
1958	464	3131	17	79	468	20
1959	329	3460	10	93	561	19
1960	1547	5007	44	99	660	17
1961	427	5434	8	113	773	17

Figura 3.14: DataFrame nacional anual.

un análisis con mayor nivel de detalle temporal. La Figura 3.15 presenta el DataFrame correspondiente a la información mensual.

FECHA	YEAR	MES	PERSONA_NATURAL_INICIO	PERSONA_NATURAL_FIN	SOCIEDAD_INICIO	SOCIEDAD_FIN	FLUJO_NETO_PERSONA_NATURAL	PERSONA_NATURAL_ACTIVAS	FLUJO_NETO_SOCIEDAD	SOCIEDAD_ACTIVAS
2000-01-01	2000	1	10705	756	468	242	9949	571493	226	56336
2000-02-01	2000	2	8961	659	596	187	8302	579795	409	56745
2000-03-01	2000	3	8944	731	599	321	8213	588008	278	57023
2000-04-01	2000	4	8901	798	595	373	8103	596111	222	57245
2000-05-01	2000	5	10048	813	641	221	9235	605346	420	57665
2000-06-01	2000	6	9760	921	661	289	8839	614185	372	58037
2000-07-01	2000	7	8563	742	701	345	7821	622006	356	58393
2000-08-01	2000	8	9765	862	662	236	8903	630909	426	58819

Figura 3.15: DataFrame nacional con periodicidad mensual (2000-2026).

Finalmente, en la Figura 3.16 se presenta el DataFrame correspondiente a las ventas y exportaciones totales registradas. Los datos se encuentran desagregados mensualmente y comprenden el período desde enero de 2000 hasta mayo de 2026.

3.3.3.4. Tratamiento de valores atípicos

La Figura 3.10 evidencia la presencia de valores atípicos en la variable correspondiente a la tasa de crecimiento.

El análisis de valores atípicos mediante el criterio IQR para la tasa de crecimiento de sociedades

	ANIO_FISCAL	MES_FISCAL	TOTAL_VENTAS_EXPORTACIONES
0	2020	1	14895736557
1	2020	2	14264977512
2	2020	3	11867942042
3	2020	4	8825129434
4	2020	5	9929589750
5	2020	6	14902808913
6	2020	7	12539967052
7	2020	8	12654059265
8	2020	9	13524216727
9	2020	10	14201241581
10	2020	11	14216475455
11	2020	12	21580304571

Figura 3.16: Registro de las ventas totales.

refleja una baja variabilidad general, con umbrales normativos entre $-0{,}06\,\%$ y $0{,}96\,\%$. Como se observa en la Figura 3.17 y la Tabla 3.2, se identifica un valor atípico extremo correspondiente a septiembre de 2006, con una variación de $-5{,}82\,\%$.

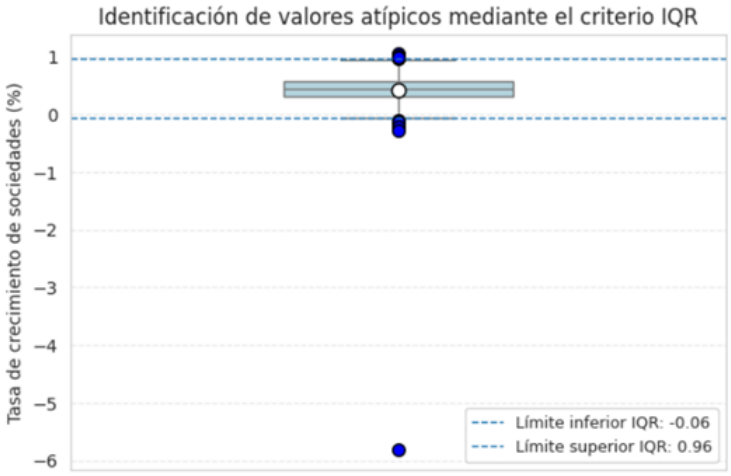


Figura 3.17: Detección de valores atípicos en la tasa de crecimiento de sociedades.

Como se mencionó en el análisis exploratorio, este comportamiento podría estar asociado con la aplicación de la Resolución NAC-DGER-2006-0253, mediante la cual el SRI dispuso la cancelación de los Registros Únicos de Contribuyentes de varias sociedades y personas naturales.

El análisis de valores atípicos para la tasa de crecimiento de personas naturales refleja una banda

Tabla 3.2: Valores atípicos identificados mediante el criterio IQR para la tasa de crecimiento de sociedades.

Periodo	Tasa de Crecimiento (%)
Mar - 2001	0,97
Ago - 2005	−0,14
Sep - 2006	−5,82
Oct - 2008	−0,10
Sep - 2008	−0,13
Dic - 2008	1,90
Dic - 2010	−0,22
Abr - 2020	−0,27
Mar - 2023	0,98
May - 2023	1,05
Jun - 2023	1,03
Ago - 2023	0,99

normal de comportamiento entre $-1{,}06\,\%$ y $1{,}76\,\%$. Como se evidencia en la Figura 3.18 y la Tabla 3.3, se presenta una anomalía negativa en septiembre de 2006 ($-14{,}65\,\%$), explicada por procesos masivos de depuración de catastro tributario.

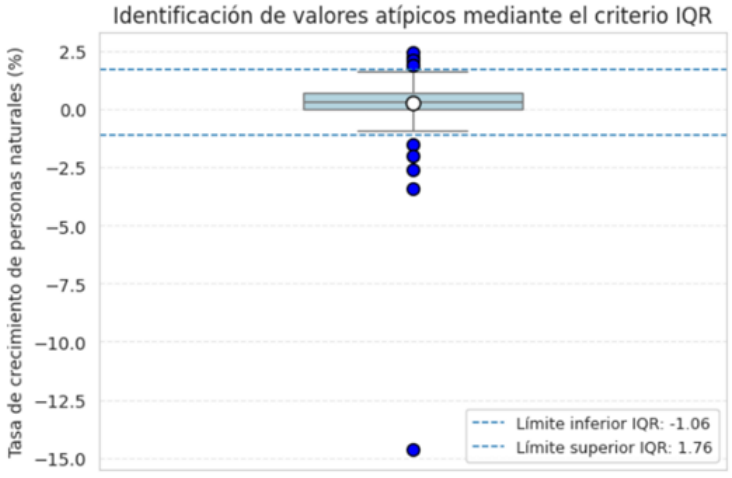


Figura 3.18: Detección de valores atípicos en la tasa de crecimiento de personas naturales.

Debido a que los valores atípicos responden a procesos administrativos reales de depuración por parte del SRI, se optó por imputarlos reemplazándolos mediante interpolación lineal a los valores atípicos correspondientes al mes de Septiembre 2006.

Tabla 3.3: Valores atípicos identificados mediante el criterio IQR para la tasa de crecimiento de personas naturales.

Periodo	Tasa de Crecimiento (%)
Ene - 2002	2,24
Feb - 2002	2,46
Mar - 2002	2,12
Sep - 2006	−14,65
Sep - 2008	1,91
Oct - 2008	1,90
Dic - 2013	−1,51
Dic - 2021	−3,43
Dic - 2022	−2,59
Dic - 2023	−2,03

CAPÍTULO 4.
MODELADO Y EVALUACIÓN

En este capítulo se desarrolla la etapa de Modelado y Evaluación de la metodología CRISP-DM, en la cual se construyen y analizan los modelos predictivos propuestos para estimar el comportamiento de las variables objetivo: las tasas de crecimiento de las sociedades y personas naturales, así como las ventas totales y exportaciones en Ecuador. Para ello, se emplean modelos de aprendizaje automático y basados en ecuaciones diferenciales con retardo.

4.1 Modelo ecuación diferencial con retardo

4.1.1 Formulación matemática

El problema de Cauchy para la ecuación diferencial con retardo viene dado por

$$\begin{aligned} x'(t) &= \alpha x(t) + \beta x(t-\tau), \quad t > 0 \\ x(t) &= \varphi(t), \quad -\tau \leq t \leq 0. \end{aligned} \tag{4.1}$$

y cuya solución se establece en el siguiente lema (Mayorga, 2024)

Lema 4.1. *Sea $h > 0$ tal que $Nh = \tau$, para algún número entero $N \geq 1$. Escribiendo $t_n \equiv nh$ y $x_n \equiv x(t_n)$, para $n \geq -N$, la solución numérica dada por $x_n = \varphi(t_n)$, para $-N \leq n \leq 0$, y, para $(m-1)\tau \leq nh \leq m\tau$ y $m \geq 1$ por*

$$\begin{aligned} x_{n+1} = x(t_n + h) = e^{\alpha h} \sum_{k=0}^{m-1} \frac{\beta^k h^k}{k!} x_{n-kN} \\ + \frac{\beta^m}{(m-1)!} \int_{t_n - m\tau}^{t_n - m\tau + h} (t_n - m\tau + h - s)^{m\tau} e^{\alpha(t_n - m\tau + h - s)} \varphi(s) ds \end{aligned} \tag{4.2}$$

define un esquema numérico exacto del problema (4.1)*.*

Con el propósito de evaluar la integral presente en el Lema 4.1, se propone aproximarla numéricamente mediante la Regla de Simpson sobre el intervalo $[t_n - m\tau, t_n - m\tau + h]$. Esta aproximación facilitará la implementación posterior de la optimización bayesiana para la estimación de los pará-

metros α y β. Así, se tiene que

$$\begin{aligned}
&\int_{t_n-m\tau}^{t_n-m\tau+h}(t_n-m\tau+h-s)^{m\tau}e^{\alpha(t_n-m\tau+h-s)}\varphi(s)ds\\
&=\frac{(t_n-m\tau+h)-(t_n-m\tau)}{6}\Big[(t_n-m\tau+h-(t_n-m\tau))^{m\tau}e^{\alpha(t_n-m\tau+h-(t_n-m\tau))}\varphi(t_n-m\tau)\\
&+4(t_n-m\tau+h-(t_n-m\tau+h/2))^{m\tau}e^{\alpha(t_n-m\tau+h-(t_n-m\tau+h/2))}\varphi(t_n-m\tau+h/2)\\
&+(t_n-m\tau+h-(t_n-m\tau+h))^{m\tau}e^{\alpha(t_n-m\tau+h-(t_n-m\tau+h))}\varphi(t_n-m\tau+h)\Big]-\frac{h^5}{90}\psi^{(4)}(\xi)\\
&=\frac{h}{6}\left[h^{m\tau}e^{\alpha h}\varphi(t_n-m\tau)+4\left(\frac{h}{2}\right)^{m\tau}e^{\alpha h/2}\varphi(t_n-m\tau+h/2)\right]-\frac{h^5}{90}\psi^{(4)}(\xi)\\
&=\frac{h^{1+m\tau}}{6}e^{\alpha h}\left[\varphi(t_n-m\tau)+2^{2-m\tau}e^{-\alpha h/2}\varphi(t_n-m\tau+h/2)\right]-\frac{h^5}{90}\psi^{(4)}(\xi),
\end{aligned}\tag{4.3}$$

con esto, la expresión (4.2) queda

$$\begin{aligned}
x_{n+1}=e^{\alpha h}&\sum_{k=0}^{m-1}\frac{\beta^k h^k}{k!}x_{n-kN}\\
&+\frac{\beta^m}{(m-1)!}\cdot\frac{h^{1+m\tau}}{6}e^{\alpha h}\left[\varphi(t_n-m\tau)+2^{2-m\tau}e^{-\alpha h/2}\varphi(t_n-m\tau+h/2)\right]-\frac{h^5}{90}\psi^{(4)}(\xi),
\end{aligned}\tag{4.4}$$

donde

$$\psi(\xi)=(t_n-m\tau+h-\xi)^{m\tau}e^{\alpha(t_n-m\tau+h-\xi)}\varphi(\xi),$$

para $\xi\in(t_n-m\tau,t_n-m\tau+h)$.

Para aplicar la Optimización Bayesiana mediante la librería Optuna de Python se requiere establecer un intervalo o espacio de búsqueda para los parámetros α y β. A partir de estos intervalos, el algoritmo explora distintas combinaciones de parámetros con el objetivo de minimizar el error cuadrático medio (RMSE) entre los datos observados y las predicciones del modelo.

Para esto, los valores iniciales de α y β se estiman mediante regresión lineal. Como los datos tienen periodicidad mensual, $h=1$. En consecuencia, $N=\tau$ y aplicando el método de diferencias finitas hacia adelante a la ecuación diferencial con retardo (4.1), se obtiene el siguiente esquema de discretización

$$x'(t_n)\approx\frac{x(t_{n+1})-x(t_n)}{h}=\alpha x(t_n)+\beta x(t_n-\tau)$$

$$\underbrace{x(t_{n+1}) - x(t_n)}_{=y_n} = \alpha x(t_n) + \beta x(t_{n-N})$$

obteniéndose la regresión lineal múltiple

$$y_n = \alpha x(t_n) + \beta x(t_{n-N}). \tag{4.5}$$

4.1.2 Tasa de Crecimiento de las Sociedades

La Tabla 4.1 nos muestra que el modelo DDE que presenta su mejor desempeño aquel cuya función inicial corresponde a un polinomio de cuarto grado, obteniendo el menor MAE ($0{,}096479$) y MAPE ($16{,}557778\,\%$). Por su parte, el polinomio de primer grado alcanza el menor RMSE ($0{,}134762$), aunque la diferencia respecto al polinomio de cuarto grado ($0{,}135760$) es reducida.

Tabla 4.1: Métricas del modelo DDE con $\tau = 54$.

Función Inicial ($\tau = 54$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$0{,}099369$	$0{,}134762$	$17{,}115244$
Polinomio 2do grado	$0{,}104125$	$0{,}140459$	$17{,}958865$
Polinomio 3er grado	$0{,}097306$	$0{,}135841$	$16{,}581087$
Polinomio 4to grado	$0{,}096479$	$0{,}135760$	$16{,}557778$
Cubic Spline	$0{,}102772$	$0{,}139217$	$17{,}476408$

Por lo tanto, el problema de Cauchy para la ecuación diferencial con retardo para la tasa de crecimiento de las sociedades está dado por

$$x'(t) = -0{,}17211684\, x(t) + 0{,}22037720\, x(t-54), \quad t > 0$$

y la función inicial se define como

$$\varphi : [-54, 0] \to \mathbb{R}$$
$$t \mapsto \varphi(t) = -4{,}557 \times 10^{-7} t^4 - 3{,}9191 \times 10^{-5} t^3 - 0{,}00120 t^2 - 0{,}02074 t + 0{,}34232.$$

La Figura 4.1 muestra como el modelo DDE con $\tau = 54$ y cuya función inicial es un polinomio de cuarto grado logra seguir la tendencia de la tasa de crecimiento de sociedades. Las principales diferencias aparecen en períodos de variaciones abruptas, donde el modelo tiende a subestimar o sobreestimar algunos máximos y mínimos.

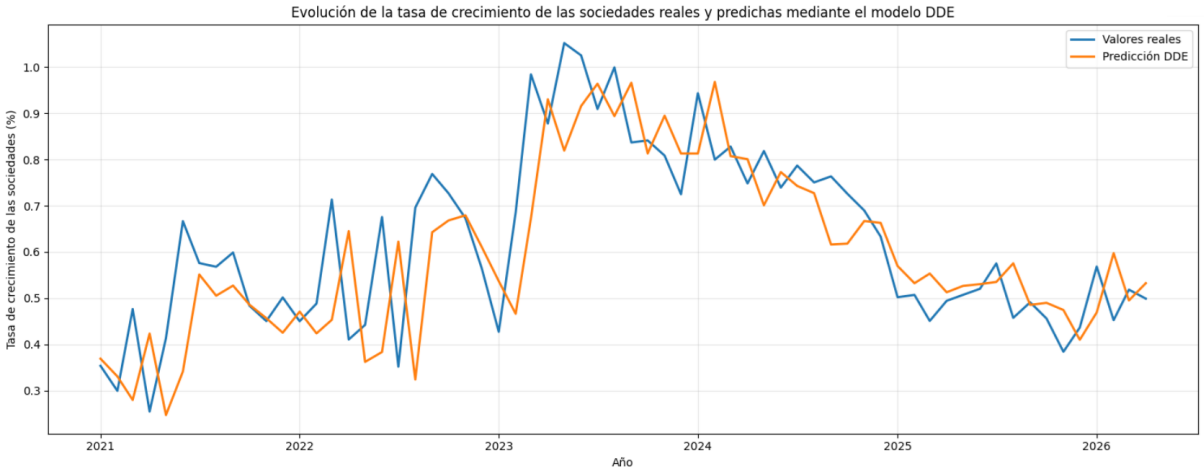


Figura 4.1: Evolución temporal de las tasas de crecimiento de sociedades reales frente a la predicción con DDE ($\tau = 54$ meses).

4.1.3 Tasa de Crecimiento de las Personas Naturales

La Tabla 4.2 muestra que el mejor rendimiento del modelo DDE se obtiene al utilizar un polinomio de segundo grado como función inicial, al alcanzar los menores valores de MAE (0,132005) y RMSE (0,187698). Aunque el polinomio de primer grado obtiene el menor MAPE (252,29 %), el polinomio de segundo grado muestra una mayor precisión global en términos de error absoluto y error cuadrático medio.

Tabla 4.2: Métricas del modelo DDE con $\tau = 36$.

Función Inicial ($\tau = 36$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,134712	0,191207	28,2502
Polinomio 2do grado	0,132005	0,187698	27,5124
Polinomio 3er grado	0,135954	0,192478	29,2998
Polinomio 4to grado	0,133697	0,189326	27,8669
Cubic Spline	0,136490	0,193421	27,9664

Así, el problema de Cauchy para la ecuación diferencial con retardo para las personas naturales viene dado por

$$x'(t) = -0{,}55375342\, x(t) + 0{,}17102876\, x(t-36), \quad t > 0$$

y la función inicial se define como

$$\varphi : [-36, 0] \to \mathbb{R}$$
$$t \mapsto \varphi(t) = -0{,}01083\, t^2 - 0{,}74821\, t - 11{,}87788.$$

La Figura 4.2 muestra que el modelo DDE con un retardo de $\tau = 36$ meses reproduce parcialmente la dinámica temporal de la tasa de crecimiento de las personas naturales. El modelo tiende a subestimar tanto los máximos positivos como las caídas pronunciadas de la serie. Captura adecuadamente la tendencia y parte de la dependencia temporal de la serie, pero presenta limitaciones para representar perturbaciones o cambios abruptos de corto plazo.



Figura 4.2: Evolución de personas naturales reales frente a la predicción con DDE ($\tau = 36$ meses).

4.1.4 Total Ventas y Exportaciones

La Tabla A.12 nos muestra que el modelo DDE presenta su mejor desempeño al utilizar un polinomio de primer grado como función inicial, al alcanzar el menor MAE, de $1{,}752034 \times 10^9$, y el menor MAPE, de $6{,}962525\,\%$. Aunque, el polinomio de segundo grado obtiene el menor RMSE, con $2{,}745917 \times 10^9$ la diferencia respecto al polinomio de primer grado es reducida.

De esta forma, el problema de Cauchy para la ecuación diferencial con retardo para las ventas totales y exportaciones está dado por

$$x'(t) = -0{,}87611721\, x(t) + 0{,}89611007\, x(t-5), \quad t > 0$$

y la función inicial se define como

$$\varphi : [-5, 0] \to \mathbb{R}$$
$$t \mapsto \varphi(t) = -1{,}537 \times 10^9\, t + 7{,}345 \times 10^9.$$

Tabla 4.3: Métricas del modelo DDE con $\tau = 5$.

Función Inicial ($\tau = 5$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$1{,}752034 \times 10^9$	$2{,}790778 \times 10^9$	6,962525
Polinomio 2do grado	$1{,}776612 \times 10^9$	$2{,}745917 \times 10^9$	7,240085
Polinomio 3er grado	$2{,}196790 \times 10^9$	$3{,}574591 \times 10^9$	8,798902
Polinomio 4to grado	$1{,}870063 \times 10^9$	$3{,}082022 \times 10^9$	7,415137
Cubic Spline	$1{,}845871 \times 10^9$	$3{,}013210 \times 10^9$	7,346500

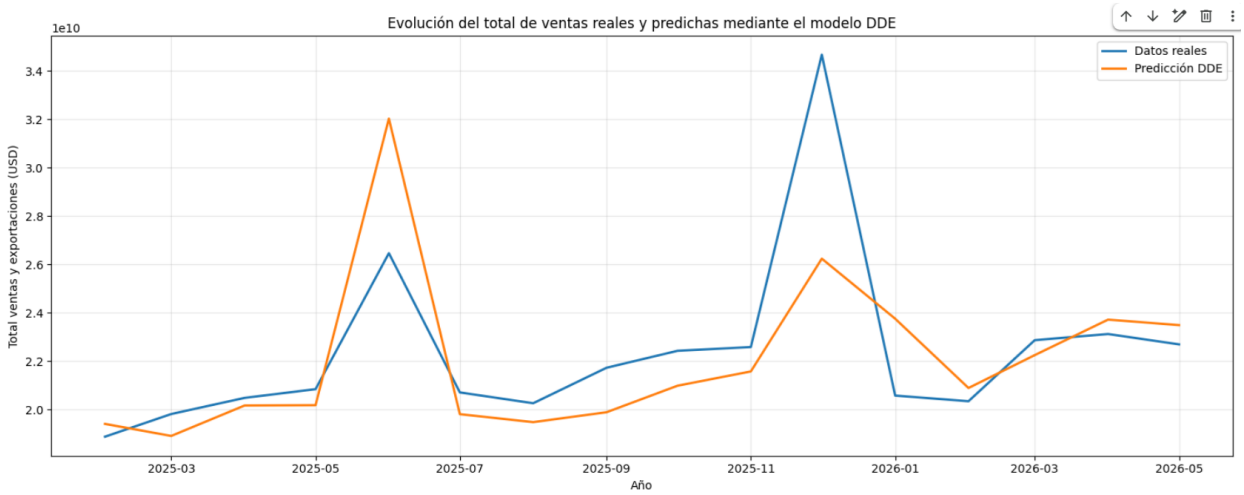


Figura 4.3: Evolución temporal del total de ventas y exportaciones reales frente a la predicción con DDE ($\tau = 5$ meses).

La Figura 4.3 muestra que el modelo DDE con un retardo de $\tau = 5$ y polinomio de primer grado reproduce la evolución del total de ventas y exportaciones, manteniendo las predicciones próximas a los valores reales durante gran parte del período de prueba. Sin embargo, se observan ciertas diferencias en los períodos caracterizados por variaciones abruptas.

4.2 Ingeniería de características en series temporales

4.2.1 Selección de lags temporales de la variable objetivo: Tasa de crecimiento de sociedades

La Figura 4.4 muestra que *Lag 1* presenta un pico negativo que sobresale significativamente por debajo de la banda de confianza (aproximadamente en $-0{,}28$) lo que indica una relación inversa

estadísticamente significativa con el periodo inmediatamente anterior.

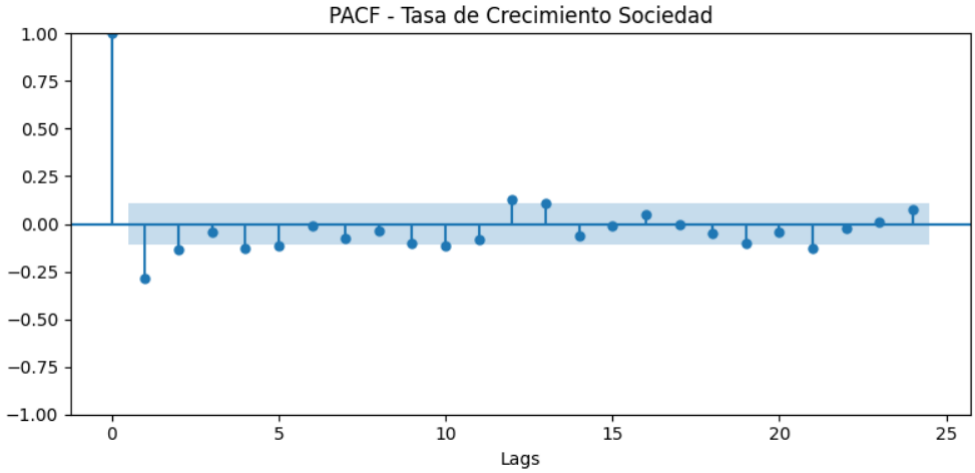


Figura 4.4: Función de autocorrelación parcial (PACF) de la serie temporal de tasa de crecimiento de sociedades. El número de retardos se expresa en meses.

La mayoría de los retardos se mantienen dentro o ligeramente sobresaliendo de la banda de confianza, lo que sugiere que no hay un impacto directo significativo a esos plazos una vez controlado el efecto de los periodos intermedios.

4.2.1.1. Selección de lags temporales de las variables exógenas

Las tres variables exógenas, Interés Corporativo, Interés Empresarial e Inflación, presentan evidencia de aportar información predictiva adicional a la contenida en los valores históricos de la tasa de crecimiento de las sociedades, como se observa en la Tabla 4.4.

Tabla 4.4: Selección de rezagos de las variables exógenas mediante la prueba de causalidad de Granger para la variable Tasa de Crecimiento de Sociedades.

Variable	Lags	p - Valor
Interés Productivo Corporativo	Lag 1	0.0234
Interés Productivo Empresarial	Lag 1	0.0030
	Lag 9	0.0205
Inflación	Lag 1	0.0018
	Lag 3	0.0242
	Lag 4	0.0158

4.2.2 Selección de lags temporales de la variable objetivo: Tasa de crecimiento de personas naturales

La Figura 4.5 indica que *Lag 1* muestra un pico negativo pronunciado (alrededor de $-0{,}36$), sobrepasando el límite inferior de la banda de confianza del $95\,\%$. Los *Lags 2, 4, 5, 9* y *11* presentan valores negativos leves que rebasan ligeramente la sombra azul, lo que sugiere cierta persistencia en el ajuste de corto plazo.

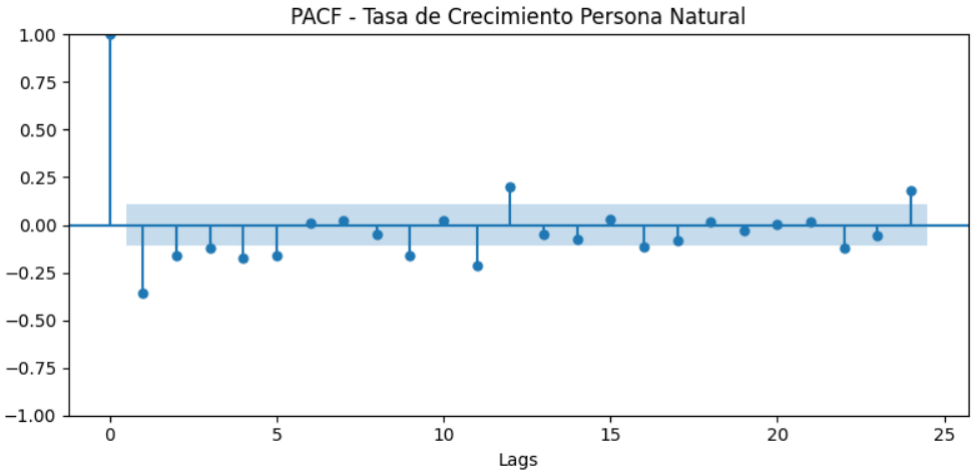


Figura 4.5: Función de autocorrelación parcial (PACF) de la serie temporal de Tasa de Crecimiento de Personas Naturales. El número de retardos se expresa en meses.

El *Lag 11* muestra un pico negativo (aproximadamente $-0{,}21$), mientras que el *Lag 12* presenta un pico positivo (aproximadamente $+0{,}20$) superando la banda azul. Finalmente, el *Lag 24* junto con el *Lag 12* confirman un comportamiento estacional.

4.2.2.1. Selección de lags temporales de las variables exógenas

La Tabla 4.5 indica que los *Lags 1* y *2* de la tasa de interés corporativo tienen un impacto inmediato y de muy corto plazo (1 a 2 meses). El *Lag 1* de la tasa de interés empresarial con ($p = 0{,}0037$) influye de manera directa e inmediata en el siguiente periodo ($t - 1$). Los *Lags 1* y *2* del interés productivo PYMES muestra una alta significancia en el plazo inmediato.

Tabla 4.5: Selección de rezagos de las variables exógenas mediante la prueba de causalidad de Granger para la variable Tasa de Crecimiento de Personas Naturales.

Variable	Lags	p - Valor
Interés Productivo Corporativo	Lag 1	0.000
	Lag 2	0.0184
Interés Productivo Empresarial	Lag 1	0.0037
Interés Productivo PYMES	Lag 1	0.0000
	Lag 2	0.0038
Interés Productivo Microcrédito	Lag 19	0.0015
	Lag 20	0.0024
	Lag 21	0.0036
	Lag 22	0.0041
Inflación	Lag 1	0.0000
	Lag 5	0.0038
	Lag 8	0.0004
	Lag 9	0.0007
	Lag 12	0.0000
	Lag 23	0.0000
	Lag 24	0.0000

Los *Lags 19, 20, 21* y *22* del interés productivo de microcrédito presenta un efecto diferido a largo plazo (alrededor de 1.5 a 2 años después). Finalmente, los *Lags 1, 5, 8, 9, 12, 23* y *24* correspondientes a la tasa de inflación impacta directamente en los pequeños negocios o microempresas.

4.2.3 Selección de lags temporales de la variable objetivo: Ventas y Exportaciones

De la Figura 4.6 podemos observar que el *Lag 1* con ($\approx -0{,}47$) es altamente significativo ya que sobresale por debajo de la banda y al igual que el *Lag 2* con ($\approx -0{,}39$). El *Lag 5* con ($\approx -0{,}73$) es el pico negativo más profundo y pronunciado de todo el gráfico. Y por último, el *Lag 11* muestra una fuerte relación inversa con el mismo periodo del ciclo anterior menos un mes.

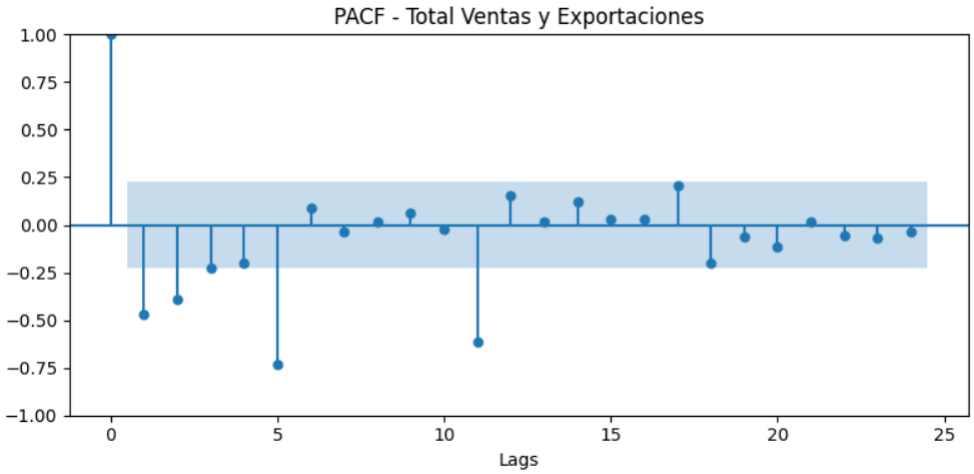


Figura 4.6: Función de autocorrelación parcial (PACF) de la serie temporal de Ventas y Exportaciones. El número de retardos se expresa en meses.

4.2.3.1. Selección de lags temporales de las variables exógenas

La Tabla 4.6 muestra las variables exógenas que tienen capacidad predictiva estadísticamente significativa sobre la variable objetivo Total Ventas y Exportaciones.

Tabla 4.6: Selección de rezagos de las variables exógenas mediante la prueba de causalidad de Granger para la variable Ventas.

Variable	Lags	p - Valor
Interés Productivo Empresarial	Lag 1	0.0043
Interés Productivo PYMES	Lag 16	0.0366
Interés Productivo Microcrédito	Lag 12	0.0146
Crecimiento del PIB	Lag 12	0.0300
Inflación	Lag 12	0.0001
	Lag 13	0.0163
Tasa de Desempleo	Lag 1	0.0000
	Lag 2	0.0000
	Lag 3	0.0001
	Lag 4	0.0007
	Lag 5	0.0009

4.3 Modelo XGBoost

4.3.1 Tasa de Crecimiento de las Sociedades

Para el modelado de la tasa de crecimiento de las sociedades se consideran dos enfoques. En el primero, el modelo se construye únicamente a partir de los rezagos de la variable objetivo. En el segundo, se incorporan tanto los rezagos de la variable objetivo como los rezagos de las variables exógenas.

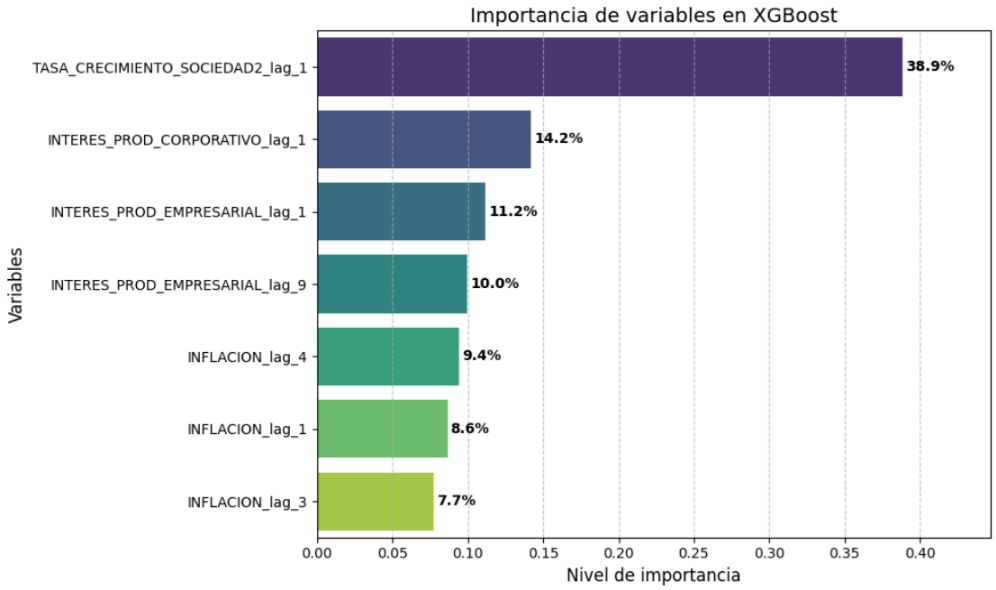


Figura 4.7: Importancia relativa de las variables en el modelo XGBoost.

En la Figura 4.7 se observa que la TASA_CRECIMIENTO_SOCIEDAD2_lag_1 es la variable más dominante con 38,9 %. Esto nos indica que la dinámica de crecimiento reciente es el predictor más fiable del crecimiento futuro inmediato.

Respecto al entorno financiero, se tiene que las variables INTERES_PROD_CORPORATIVO_lag_1 con (14,2 %), INTERES_PROD_EMPRESARIAL_lag_1 con (11,2 %) e INTERES_PROD _EMPRESARIA_lag_9 (10,0 %) indican que el costo del dinero tanto corporativo como empresarial a corto plazo impacta fuertemente.

Y en un entorno macroeconómico, las variables INFLACION_lag_4 (9,4 %), INFLACION_lag_1 (8,6 %) e INFLACION_lag_3 (7,7 %) muestran que el comportamiento de los precios influye significativamente; y, el rezago de 4 periodos indica que el impacto de la inflación tarda un tiempo en absorberse en la actividad económica real.

Cabe mencionar que, como se observa en la Figura 4.4, únicamente se obtuvo un rezago de la variable objetivo como variable predictora. Por esta razón, al existir una sola variable de entrada, no se ha incluido el gráfico de importancia de variables, dado que su importancia corresponde al 100 %.

La Tabla 4.7 muestra que el modelo XGBoost con variables exógenas presenta un rendimiento ligeramente superior al modelo que utiliza únicamente los rezagos de la variable objetivo.

Tabla 4.7: Comparativa del desempeño de XGBoost según la presencia de variables exógenas en el conjunto de prueba para tasa de crecimiento de sociedad.

Configuración	RMSE	MAE	MAPE (%)
XGBoost (sin variables exógenas)	0,134830	0,099381	17,19
XGBoost (con variables exógenas)	0,134781	0,097265	16,40

En la Figura 4.8 se presenta el gráfico de la evolución de las tasas de crecimiento de las sociedades reales y predichas mediante el modelo XGBoost con inclusión de variables exógenas y mejor rendimiento. Se observa que en los valles y picos de la línea de predicción hay un desfase de 1 mes después de la línea azul. Esto se debe a la dependencia de un 38,9 % de la variable TASA_CRECIMIENTO_SOCIEDAD_lag_1, el modelo espera ver el movimiento real para ajustar su predicción del siguiente mes.

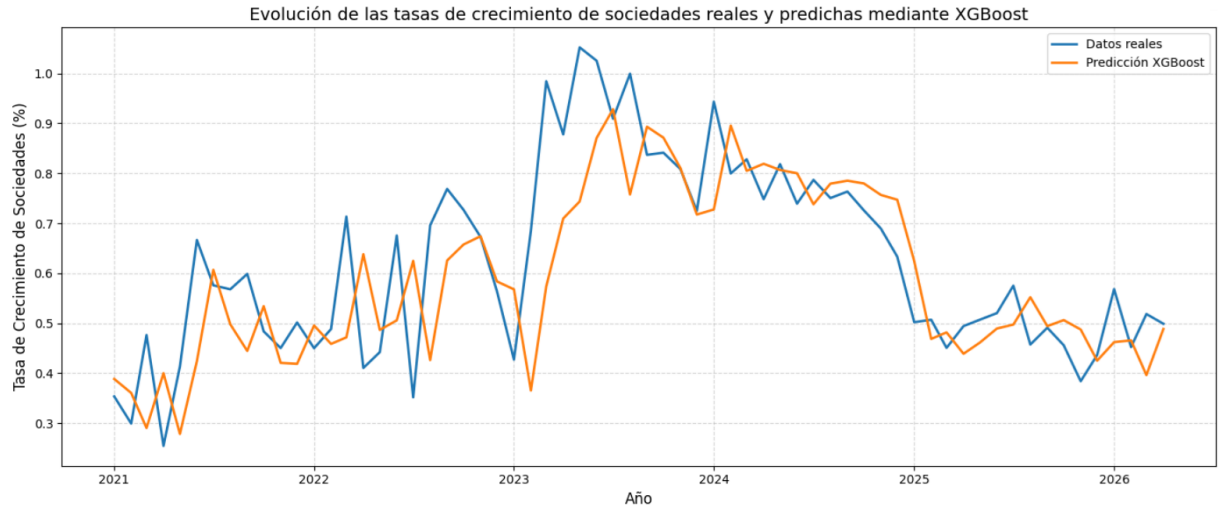


Figura 4.8: Evolución temporal de la tasa de crecimiento de sociedades reales frente a la predicción con XGBoost.

4.3.2 Tasa de Crecimiento de las Personas Naturales

El modelado de la tasa de crecimiento de las personas naturales se construye a partir de los rezagos de la variable objetivo y, en segundo enfoque, se consideran tanto los rezagos de la variable objetivo como los rezagos de las variables exógenas.

La Figura 4.9 muestra que la variable que domina el modelo es TASA_CRECIMIENTO_PN2_lag_5 con (31,3 %), es decir, casi un tercio de las divisiones en los árboles de XGBoost dependen de lo que ocurrió hace 5 meses. Por otro lado, las variables TASA_CRECIMIENTO_PN2_lag_1 lag_2 en conjunto tienen un (27,9 %) lo que muestra que la inercia reciente suele ser un fuerte predictor del comportamiento futuro. Los rezagos intermedios lag_9 con (14,3 %) y lag_4 con (12,1 %) sugieren que existen patrones periódicos o dinámicas intermedias importantes. Aunque el aporte de los rezagos lag_11 y lag_12 son bajos, son útiles para capturar efectos de estacionalidad anual.

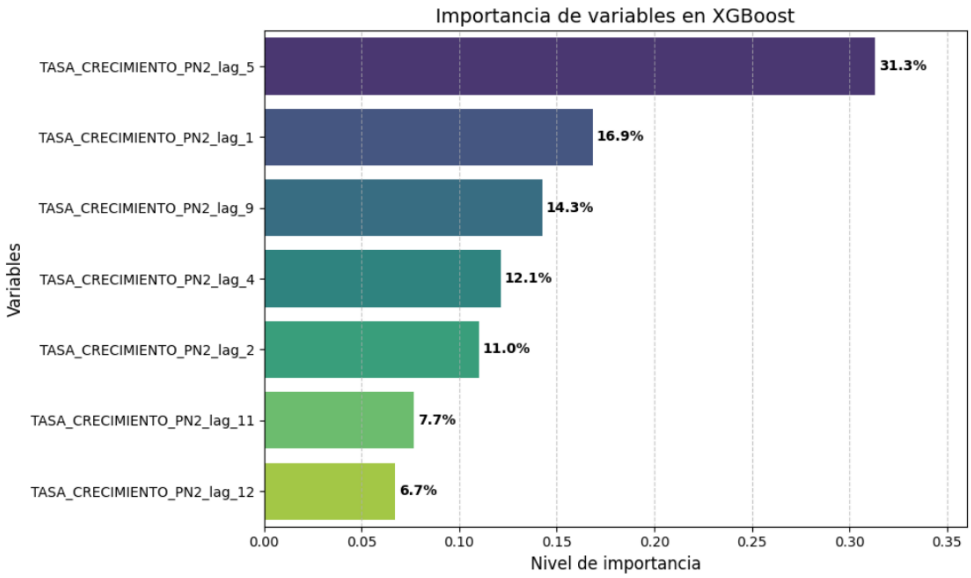


Figura 4.9: Importancia relativa de las variables en el modelo XGBoost (sin variables exógenas).

La Figura 4.10 muestra que la variable predictora TASA_CRECIMIENTO_PN2_lag_12 con(23,3 %) es la más importante del modelo. Casi una cuarta parte de la capacidad predictiva reside en lo que sucedió hace exactamente un año, lo que señala un patrón de fuerte estacionalidad anual. Los rezagos TASA_CRECIMIENTO_PN2_lag_1 con (11,1 %) y lag_2 (10,1 %) confirman que la inercia de corto plazo sigue siendo vital.

Por otro lado, Tasas de Interés de Microcrédito con rezagos lejanos como lag_19 (5,8 %), lag_20 (4,6 %), lag_21 (1,9 %) y lag_22 (1,5 %) sugiere que las variaciones en las tasas de microcrédito tardan entre 1.5 y 2 años en impactar completamente. Finalmente, las tasas de interés influyen de

forma más directa e inmediata en el corto plazo.

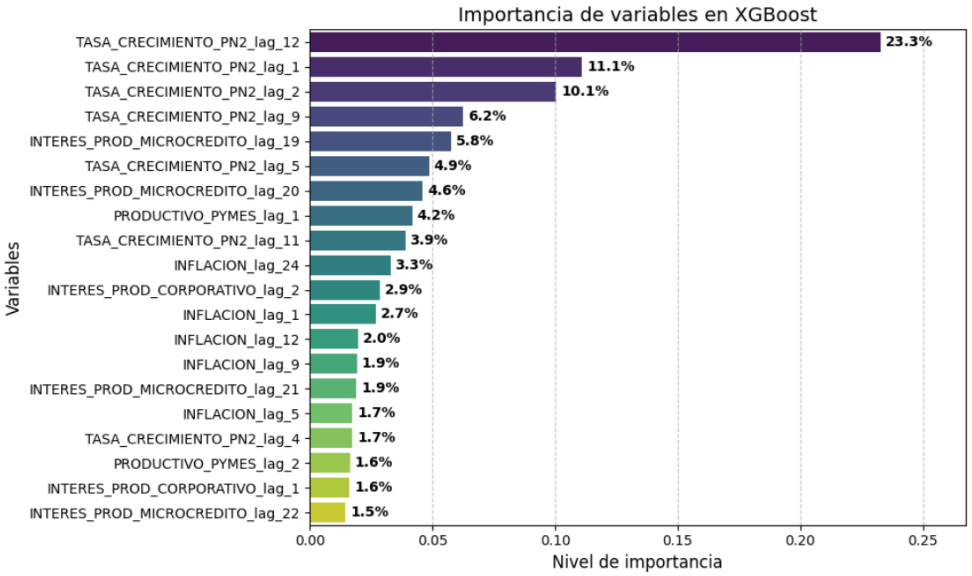


Figura 4.10: Importancia relativa de las variables en el modelo XGBoost.

Al observar la Tabla 4.8, se evidencia que el modelo XGBoost con variables exógenas presenta un rendimiento ligeramente superior, al superar al modelo univariable en las tres métricas evaluadas en los conjuntos de entrenamiento y prueba. La incorporación de información macroeconómica y crediticia contribuyó a mejorar ligeramente la capacidad de generalización del modelo.

Tabla 4.8: Comparativa del desempeño de XGBoost según la presencia de variables exógenas en el conjunto de prueba para la tasa de crecimiento de personas naturales.

Configuración	RMSE	MAE	MAPE (%)
XGBoost (sin variables exógenas)	0,221012	0,155922	40,21
XGBoost (con variables exógenas)	0,219364	0,144632	39,76

La Figura 4.11 muestra que el modelo XGBoost con variables exógenas logra reproducir de manera razonable la dinámica general de la tasa de crecimiento de las personas naturales. El modelo tiende a suavizar los valores extremos, subestimando la magnitud de algunas caídas y picos registrados en la serie real. Se evidencia un ligero desfase entre los datos observados y los predichos.

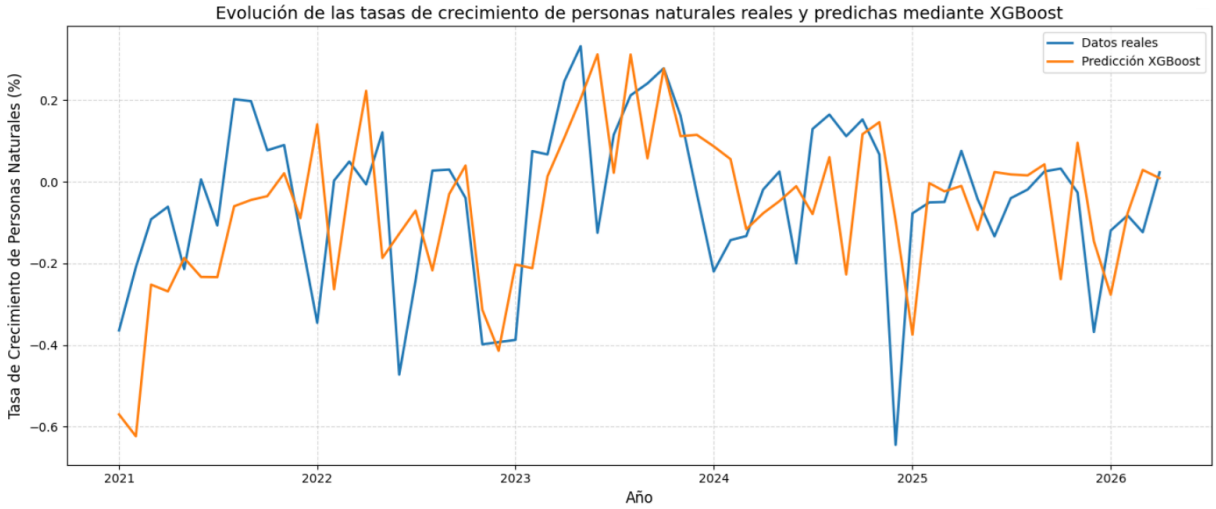


Figura 4.11: Evolución temporal de la tasa de crecimiento de personas naturales reales frente a la predicción con XGBoost.

4.3.3 Total Ventas y Exportaciones

La Figura 4.12 evidencia que el modelo XGBoost sin variables exógenas se basa en el comportamiento histórico de las ventas y exportaciones. El *lag_11* presenta la mayor importancia con un (54,9 %), seguido de los rezagos de 5 y 1 mes, con 24,3 % y 17,0 %, respectivamente. Estos resultados sugieren que la serie presenta una dependencia temporal tanto de corto como de mediano plazo, destacándose especialmente la información correspondiente a aproximadamente un año atrás.

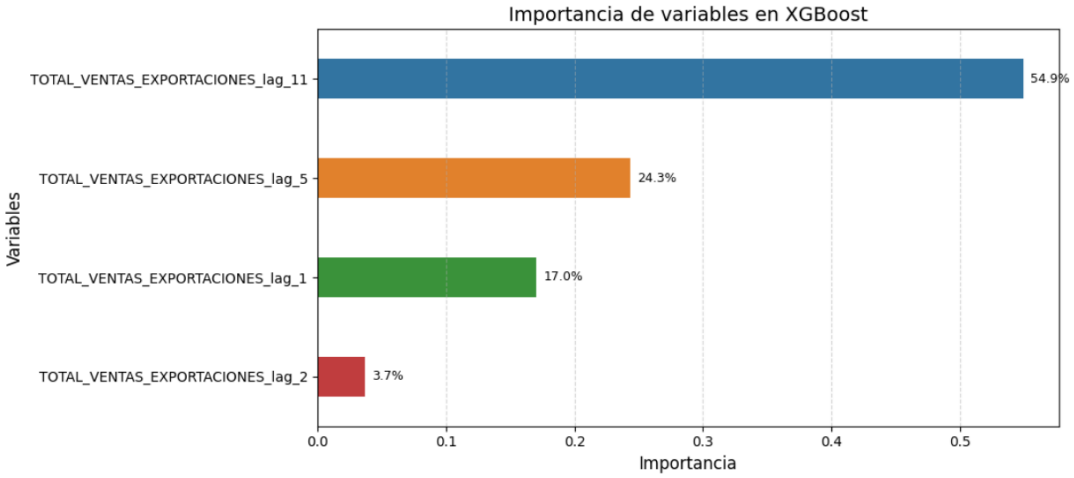


Figura 4.12: Importancia relativa de las variables en el modelo XGBoost (sin variables exógenas).

En la Figura 4.13 se puede observar que el rezago 11 de la variable objetivo es la más relevante,

con 17,8 %, seguido del rezago de 5 meses, con 14,9 %, y de la tasa de desempleo rezagada un mes, con 14,3 %. Otro rezago relevante de un mes de la variable objetivo alcanza un 13,1 %. Estos resultados evidencian que el modelo no depende exclusivamente de la dinámica histórica de las ventas y exportaciones, sino que también incorpora información proveniente del entorno macroeconómico.

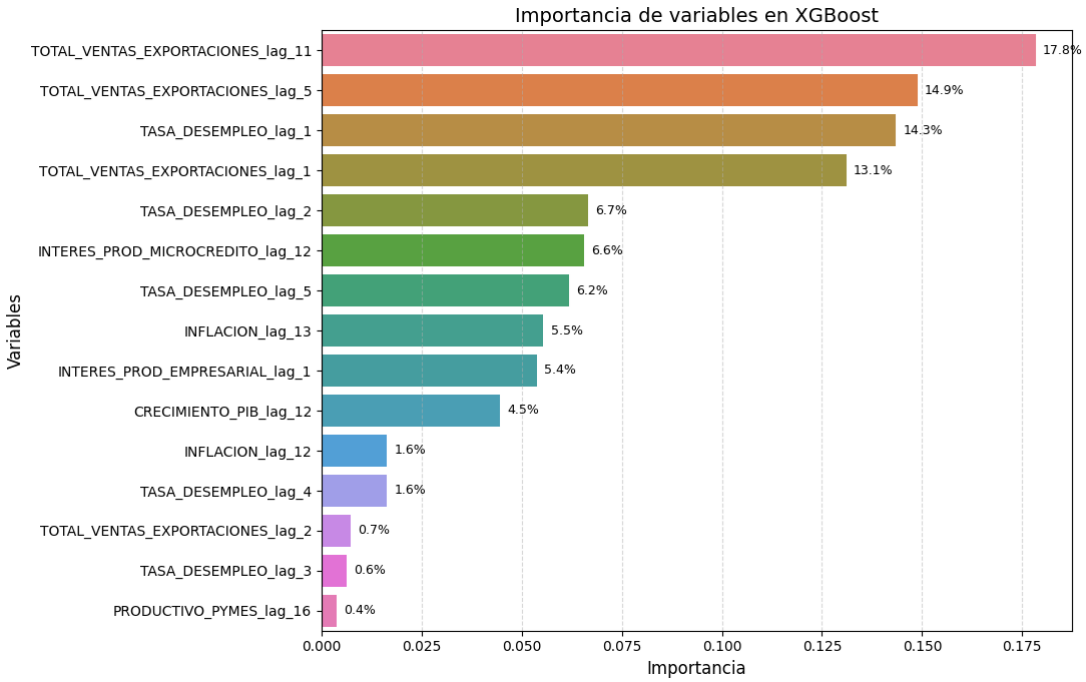


Figura 4.13: Importancia relativa de las variables en el modelo XGBoost.

La Tabla 4.9 muestra que la incorporación de variables exógenas no mejoró considerablemente el ajuste del modelo XGBoost sobre ninguno de los conjuntos. El modelo sin variables exógenas presenta menores valores de RMSE, MAE y MAPE en los datos no observados, con un MAPE de 8,12 % frente a 10,03 % del modelo con variables exógenas. Por tanto, el modelo sin variables exógenas presenta un mejor rendimiento predictivo y una mayor capacidad de generalización.

Tabla 4.9: Comparativa del desempeño de XGBoost según la presencia de variables exógenas en el conjunto de prueba para total de ventas.

Configuración	RMSE	MAE	MAPE (%)
XGBoost (sin variables exógenas)	$3{,}539 \times 10^9$	$2{,}064 \times 10^9$	8,12
XGBoost (con variables exógenas)	$4{,}06 \times 10^9$	$2{,}521 \times 10^9$	10,03

La Figura 4.14 presenta la evolución temporal de los valores reales y predichos del total de

ventas y exportacione mediante el modelo XGBoost sin variables exógenas. En términos generales, el modelo reproduce adecuadamente la tendencia de la serie y mantiene las predicciones próximas a los valores observados durante la mayor parte del período de prueba. Sin embargo, se observan diferencias importantes en los meses caracterizados por variaciones abruptas, por ejemplo, finales de 2025 se observa un pico en las ventas reales y que es subestimado por el modelo.

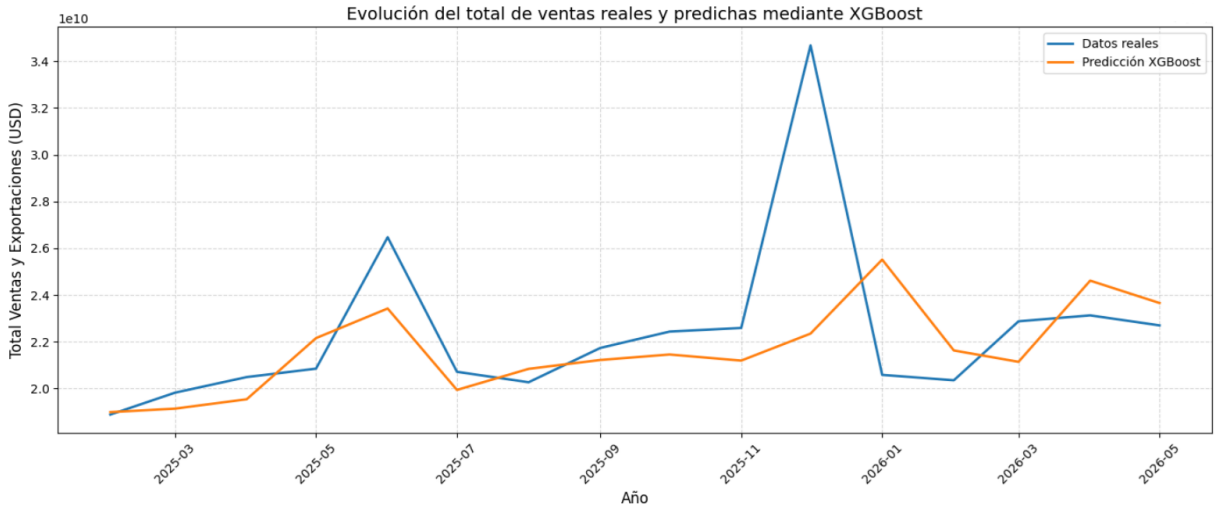


Figura 4.14: Evolución temporal del total de ventas y exportaciones reales frente a la predicción con XGBoost.

4.4 Modelo Random Forest

4.4.1 Tasa de Crecimiento de las Sociedades

En la Figura 4.15 se muestra que el rezago TASA_CRECIMIENTO_SOCIEDAD2_lag1 constituye la variable de mayor importancia dentro del modelo con (41,2 %) y evidencia una fuerte dependencia de corto plazo de la variable objetivo respecto de su comportamiento inmediatamente anterior. Por otro lado, las variables exógenas que destacan son la INFLACION_lag_4 con (11,4 %), el INTERES_PROD_EMPRESARIAL_lag_9 con (10,2 %), y el INTERES_PROD_CORPORATIVO _lag_1 con (10,1 %). Esto indica que el modelo complementa la información histórica de la variable objetivo con factores macroeconómicos para realizar sus predicciones.

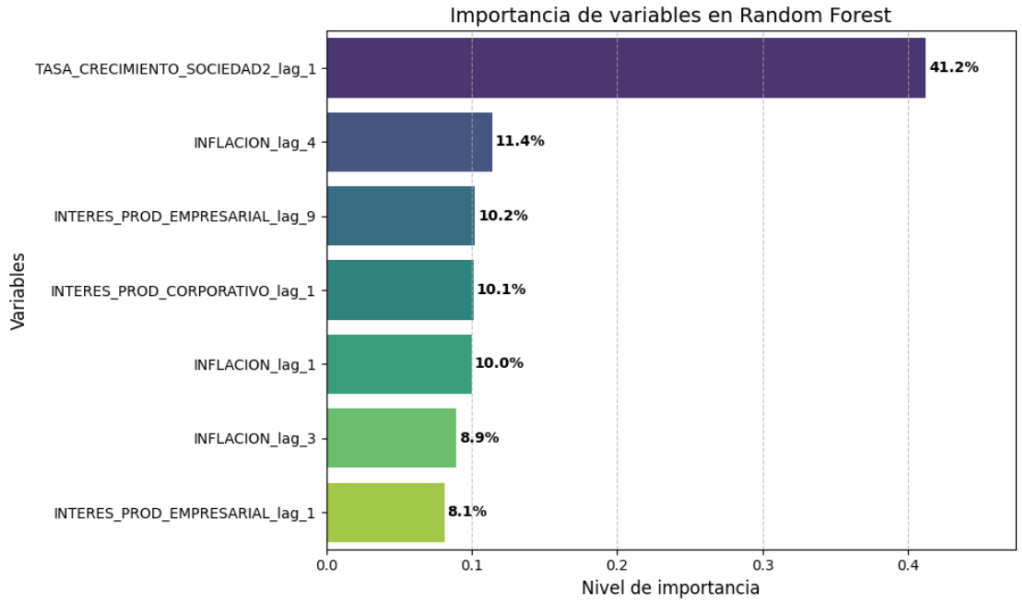


Figura 4.15: Importancia relativa de las variables en el modelo Random Forest.

La Tabla 4.10 muestra que el modelo Random Forest con variables exógenas presenta un mejor desempeño en el conjunto de prueba debido a que los valores de RMSE, MAE y MAPE son menores respecto al modelo que utiliza únicamente los rezagos de la variable objetivo. Aunque las diferencias en RMSE y MAE son moderadas, la mejora simultánea en las tres métricas indica que la incorporación de variables exógenas aporta información predictiva adicional y permite una mejor generalización del modelo para la tasa de crecimiento de sociedades.

Tabla 4.10: Comparativa del desempeño de Random Forest según la presencia de variables exógenas en el conjunto de prueba para la tasa de crecimiento de la sociedad.

Configuración	RMSE	MAE	MAPE (%)
Random Forest (sin v. exógenas)	0,152063	0,110978	19,49
Random Forest (con v. exógenas)	0,149068	0,108103	17,46

La Figura 4.16 corresponde al modelo Random Forest con variables exógenas. Presenta la tendencia general de la tasa de crecimiento de sociedades, manteniendo las predicciones cercanas a los valores observados durante gran parte del período analizado. A pesar de eso, se identifican diferencias más importantes en períodos caracterizados por variaciones abruptas, especialmente durante 2023, cuando el modelo subestima algunos de los valores máximos registrados.

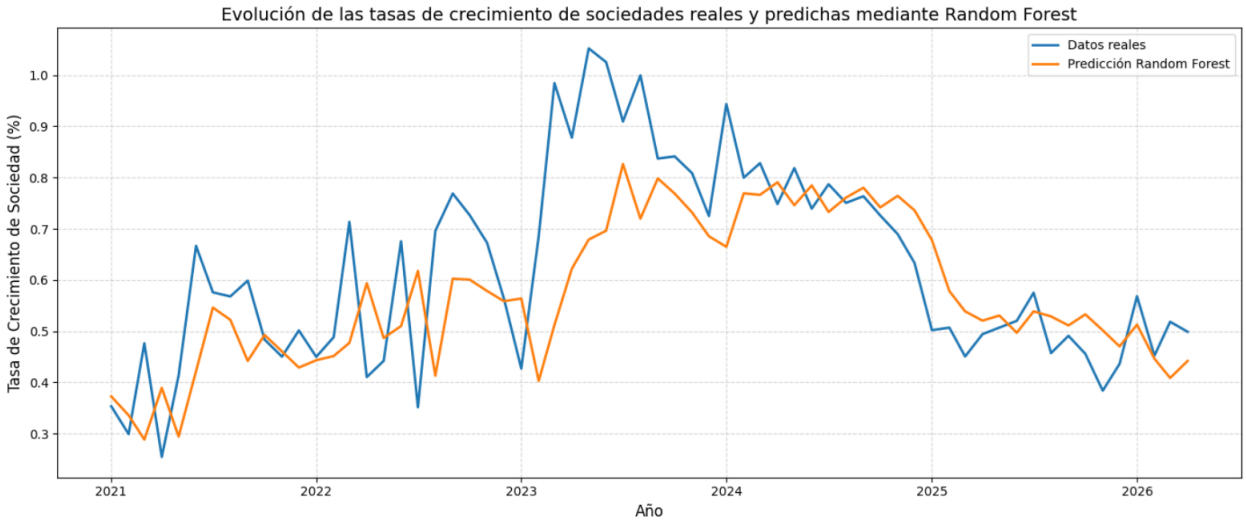


Figura 4.16: Evolución temporal de la tasa de crecimiento de sociedades reales frente a la predicción con Random Forest.

4.4.2 Tasa de Crecimiento de las Personas Naturales

La Figura 4.17 muestra que el rezago más recientes de la tasa de crecimiento de personas naturales presentan la mayor importancia relativa con (24,7 %). Le sigue el rezago de dos períodos con (21,7 %).

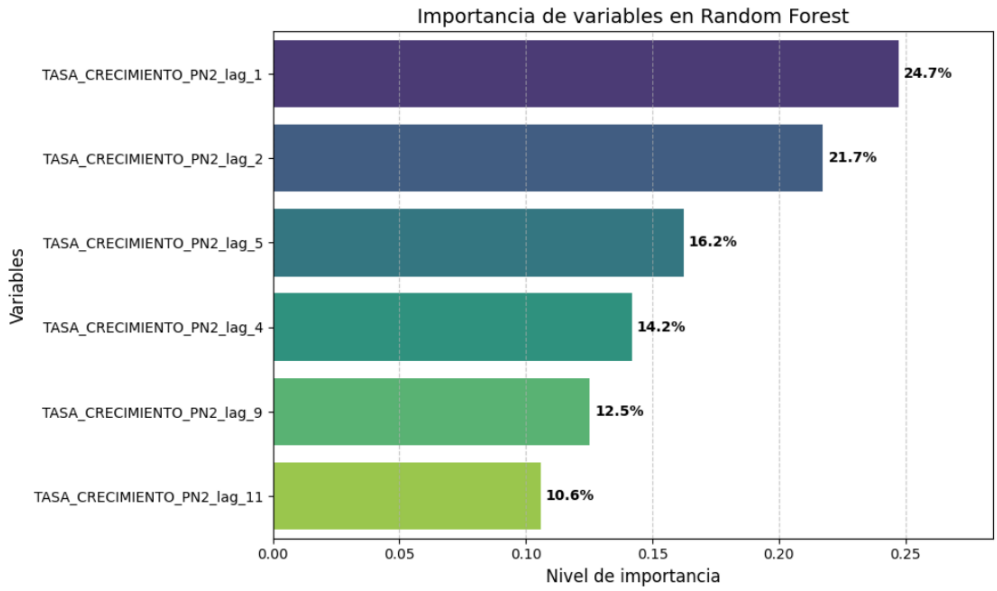


Figura 4.17: Importancia relativa de las variables en el modelo Random Forest (sin variables exógenas).

En conjunto, ambos representan el (46,4 %) de la importancia total, lo que evidencia una marcada dependencia de corto plazo de la variable objetivo. De manera similar, los *lags* de 4, 5, 9 y 11 períodos mantienen una participación relevante indicando que el modelo también incorpora información histórica a largo plazo.

La Figura 4.18 muestra que el rezago más reciente de la variable objetivo se mantiene como la variable predictora más relevanta con una contribución menor de (12,8 %). Y en el mismo orden de contribución que en el caso anterior, se mantienen los rezagos de la propia tasa de crecimiento de personas naturales. En conjunto, estos rezagos concentran aproximadamente el (59,6 %) de la importancia total, lo que evidencia que la dinámica histórica de la variable objetivo constituye la principal fuente de información predictiva del modelo.

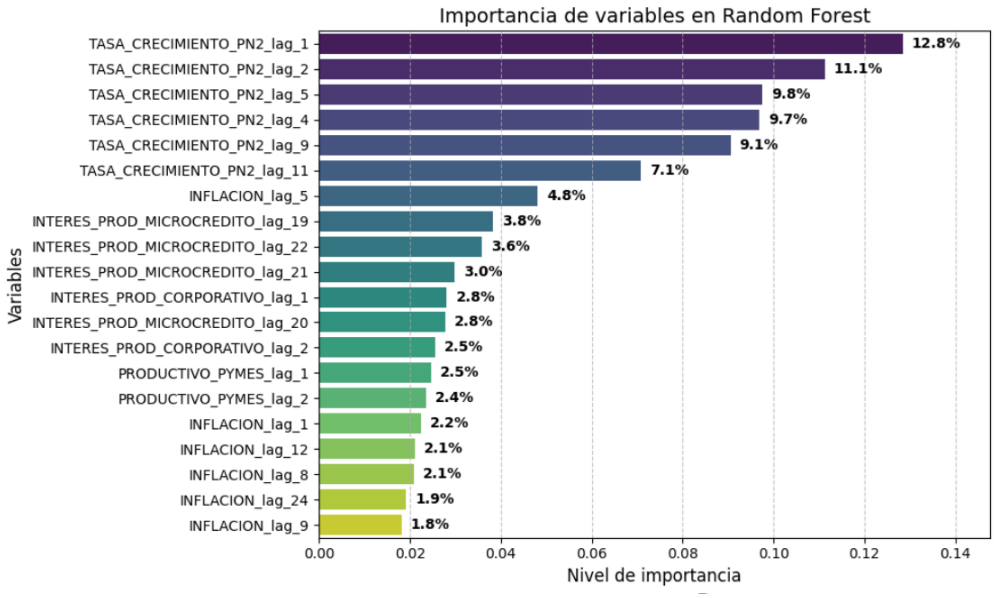


Figura 4.18: Importancia relativa de las variables en el modelo Random Forest.

La Tabla 4.11 muestra que el modelo Random Forest sin variables exógenas presenta un mejor desempeño predictivo para la tasa de crecimiento de personas naturales. Esta configuración obtiene menores valores de RMSE, MAE y MAPE en el conjunto de prueba.

Por tanto, la incorporación de variables macroeconómicas no mejoró la capacidad predictiva del modelo en este caso.

Tabla 4.11: Comparativa del desempeño de Random Forest según la presencia de variables exógenas en el conjunto de prueba para la tasa de crecimiento de personas naturales.

Configuración	RMSE	MAE	MAPE (%)
Random Forest (sin v. exógenas)	0,197226	0,154260	29,49
Random Forest (con v. exógenas)	0,217020	0,0,171596	30,04

La Figura 4.19 muestra que el modelo Random Forest reproduce de manera general la evolución temporal de la tasa de crecimiento de personas naturales, siguiendo varios de los cambios de dirección observados en la serie real.

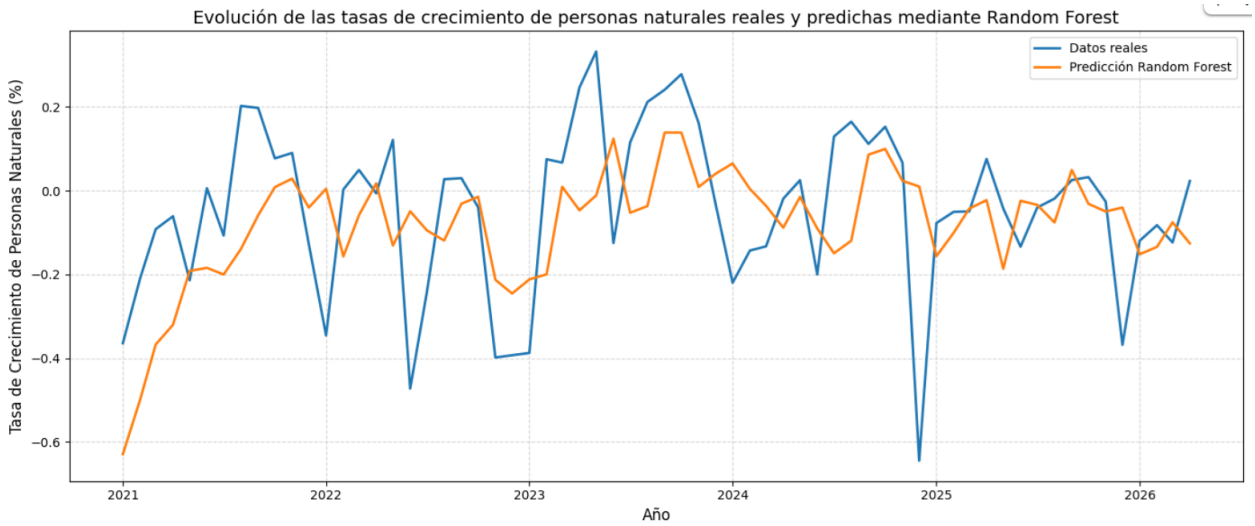


Figura 4.19: Evolución temporal de la tasa de crecimiento de personas naturales reales frente a la predicción con Random Forest.

4.4.3 Total Ventas y Exportaciones

La Figura 4.20 nos indica que el rezago TOTAL_VENTAS_EXPORTACIONES_lag_11 es la variable más relevante del modelo con (50,6 %) y junto con el rezago de 5 meses, con (28,7 %) representan el (79,3 %) de la importancia total. Esto sugiere a la estacionalidad que presentan las ventas y exportaciones a mediados y fin de año.

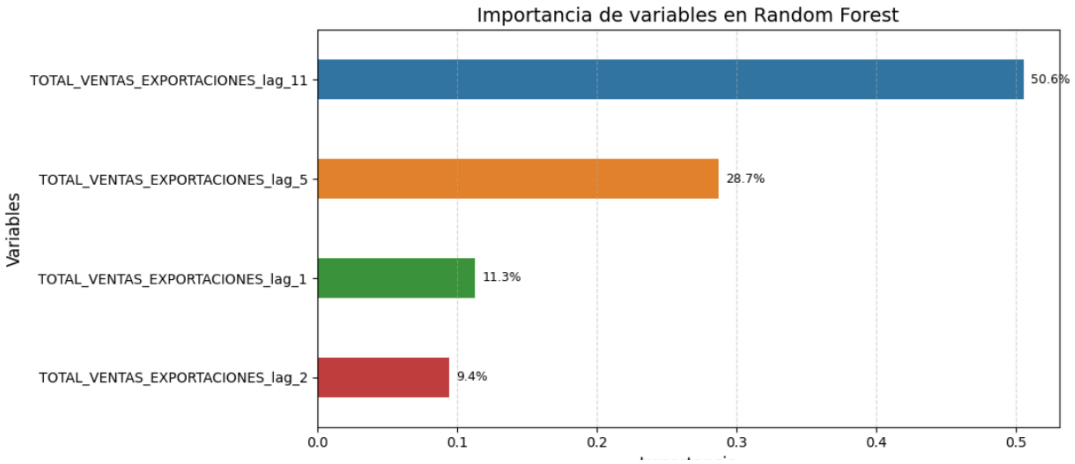


Figura 4.20: Importancia relativa de las variables en el modelo Random Forest (sin variables exógenas).

En la Figura 4.21 se puede notar que el rezago TOTAL_VENTAS_EXPORTACIONES_lag_11 continúa siendo la variable predictora más relevante dentro del modelo con (23,8 %), seguido de la inflación rezagada 13 meses, con (20,7 %), y del rezago de 5 meses de la variable objetivo, con (15,4 %). La elevada participación de los rezagos de la variable objetivo evidencia que la información histórica de las ventas y exportaciones continúa desempeñando un papel importante en la predicción.

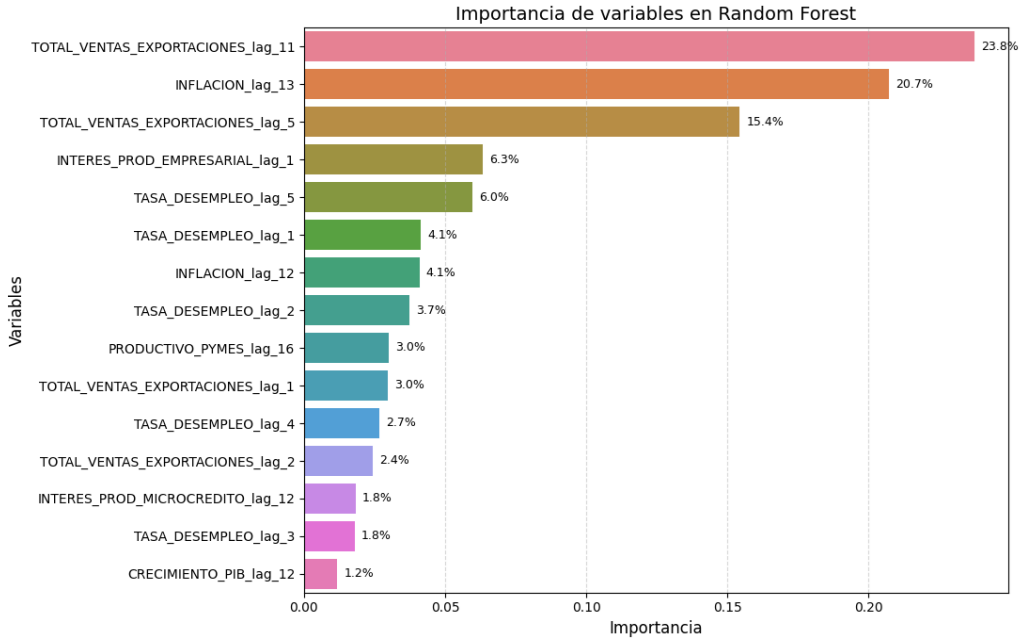


Figura 4.21: Importancia relativa de las variables en el modelo Random Forest.

En la Tabla 4.12 muestra que el modelo Random Forest con variables exógenas presenta un mejor rendimiento predictivo respecto a la configuración que utiliza únicamente los rezagos de la variable objetivo. La incorporación de variables exógenas reduce el RMSE de $4{,}194 \times 10^9$ a $3{,}817 \times 10^9$, el MAE de $2{,}663 \times 10^9$ a $2{,}431 \times 10^9$ y el MAPE de $10{,}78\,\%$ a $9{,}88\,\%$. Esta disminución indica que las variables macroeconómicas incorporadas aportan información predictiva adicional mejorando la capacidad de generalización del modelo para el total de ventas y exportaciones.

Tabla 4.12: Comparativa del desempeño de Random Forest según la presencia de variables exógenas en el conjunto de prueba para total de ventas.

Configuración	RMSE	MAE	MAPE (%)
Random Forest (sin v. exógenas)	$4{,}194 \times 10^9$	$2{,}663 \times 10^9$	10,78
Random Forest (con v. exógenas)	$3{,}817 \times 10^9$	$2{,}431 \times 10^9$	9,88

La Figura 4.22 nos muestra diferencias en períodos caracterizados por incrementos abruptos de las ventas. En particular, el modelo subestima los máximos registrados a mediados y finales de 2025 y posteriormente presenta valores superiores a los observados una vez que estos picos desaparecen.

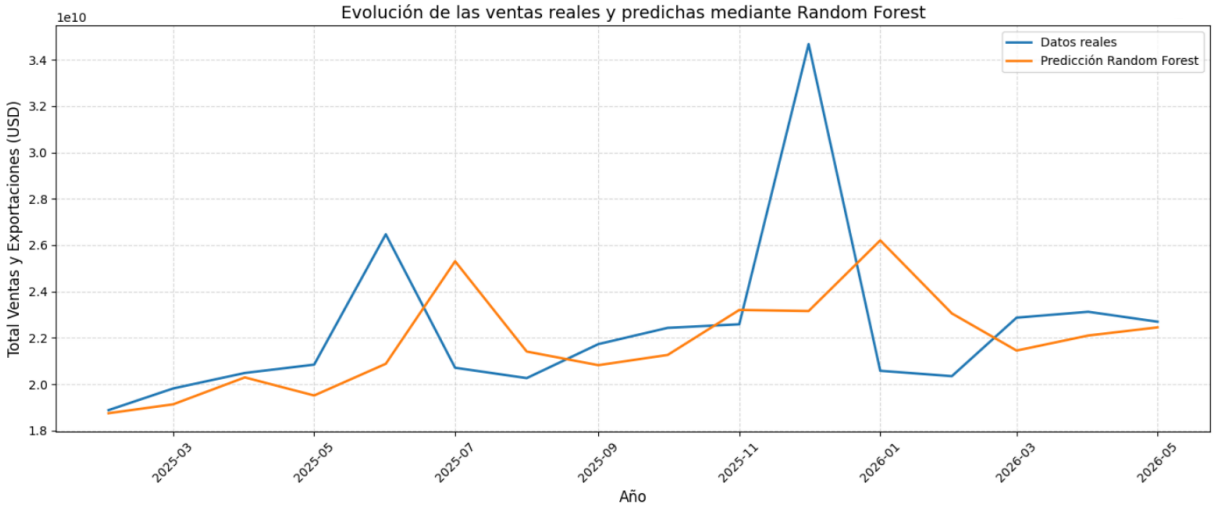


Figura 4.22: Evolución temporal de ventas y exportaciones reales frente a la predicción con Random Forest.

4.5 Evaluación de los modelos

4.5.1 Resultados del modelo de Ecuaciones Diferenciales con Retardo

Tabla 4.13: Resultados de las métricas de evaluación del modelo DDE para la tasa de crecimiento de sociedades, personas naturales y el total de ventas.

Variable objetivo	RMSE	MAE	MAPE (%)
Sociedad ($\tau = 54$, φ 4to. grado)	0,135760	0,096479	16,557778
Persona Natural ($\tau = 36$, φ 2do. grado)	0, 187698	0, 132005	27,5124
Total ventas ($\tau = 5$, φ 1er. grado)	$2{,}790778 \times 10^9$	$1{,}752034 \times 10^9$	6,962525

4.5.2 Resultados de XGBoost

Tabla 4.14: Resultados de las métricas de evaluación del modelo XGBoost para la tasa de crecimiento de sociedades, personas naturales y el total de ventas.

Variable objetivo	RMSE	MAE	MAPE (%)
Sociedad (con variables exógenas)	0,134781	0,097265	16,40
Persona Natural (con variables exógenas)	0,219364	0,144632	39,76
Total ventas (sin variables exógenas)	$3{,}539 \times 10^9$	$2{,}064 \times 10^9$	8,12

4.5.3 Resultados de Random Forest

Tabla 4.15: Resultados de las métricas de evaluación del modelo Random Forest para la tasa de crecimiento de sociedades, personas naturales y el total de ventas.

Variable objetivo	RMSE	MAE	MAPE (%)
Sociedad (con variables exógenas)	0,149068	0,108103	17,46
Persona Natural (sin variables exógenas)	0,197226	0,154260	29,49
Total ventas (con variables exógenas)	$3{,}817 \times 10^9$	$2{,}431 \times 10^9$	9,88

4.5.4 Comparación de los modelos

4.5.4.1. Tasa de Crecimiento de Sociedades

De acuerdo a la Tabla 4.16, el **modelo XGBoost** (con variables exógenas) presenta el mejor rendimiento global, aunque la diferencia respecto al DDE es muy pequeña.

Tabla 4.16: Comparativa del rendimiento de los modelos DDE, XGBoost y Random Forest para la tasa de crecimiento de sociedades.

Modelo	RMSE	MAE	MAPE (%)
DDE	0,135760	0,096479	16,56
XGBoost	0,134781	0,097265	16,40
Random Forest	0,149068	0,108103	17,46

4.5.4.2. Tasa de Crecimiento de Personas Naturales

De acuerdo a la Tabla 4.17, los resultados muestran que el **modelo DDE** presenta el mejor desempeño predictivo, al obtener los menores valores de RMSE (0,187698), MAE (0,132005) y MAPE (27,5124 %).

Tabla 4.17: Comparativa del rendimiento de los modelos DDE, XGBoost y Random Forest para la tasa de crecimiento de personas naturales.

Modelo	RMSE	MAE	MAPE (%)
DDE	0,187698	0,132005	27,5124
XGBoost	0,219364	0,144632	39,76
Random Forest	0,197226	0,154260	29,49

4.5.4.3. Total Ventas y Exportaciones

Según lo mostrado por la Tabla 4.18, el **modelo DDE** presenta el mejor rendimiento predictivo para el total de ventas y exportaciones, ya que obtiene los valores más bajos en las tres métricas de evaluación.

Tabla 4.18: Comparativa del rendimiento de los modelos DDE, XGBoost y Random Forest para el total de ventas y exportaciones.

Modelo	**RMSE**	**MAE**	**MAPE (%)**
DDE	$2{,}790778 \times 10^9$	$1{,}752034 \times 10^9$	6,96
XGBoost	$3{,}539 \times 10^9$	$2{,}064 \times 10^9$	8,12
Random Forest	$3{,}817 \times 10^9$	$2{,}431 \times 10^9$	9,88

CAPÍTULO 5.
DISCUSIÓN

5.1 Interpretación del parámetro de retardo τ.

De acuerdo a (Guerrini et al., 2020), en economía, muchos procesos dependen de eventos pasados, por lo que es natural utilizar ecuaciones diferenciales con retardo temporal para modelar fenómenos económicos como el ciclo económico de Kaldor. En su modelo, las decisiones de inversión realizadas en el pasado afectan posteriormente al stock de capital y pueden generar fluctuaciones en las variables económicas. Por otro lado, según (Minniti et al., 2006), la probabilidad de que un emprendedor individual sobreviva en el mercado durante más de 42 meses varía entre países; y, cuando poseen y administran un negocio que ha estado en funcionamiento durante más de 42 meses son considerados dueños de negocios establecidos. En este contexto, como las sociedades y las personas naturales con fines de lucro representan distintos segmentos de la actividad empresarial y productiva del Ecuador, entre ellos emprendedores, microempresas, pequeñas, medianas y grandes empresas, pueden presentar dependencia de eventos pasados y los resultados obtenidos mediante el modelo con ecuaciones diferenciales con retardo permiten identificar diferentes horizontes temporales de las variables analizadas. En particular, los resultados sugieren lo siguiente:

- En el caso de las sociedades, el retardo $\tau = 54$ meses nos da a conocer que su sostenibilidad y permanencia en el mercado nacional pueden estar asociadas con un proceso de consolidación de aproximadamente cuatro años y medio.

- Para las personas naturales, el retardo de $\tau = 36$ meses sugiere que se requieren aproximadamente tres años para consolidar el negocio y superar la etapa inicial de su actividad empresarial.

- Y el retardo de $\tau = 5$ meses para las ventas y exportaciones sugiere que se conserva información relevante de las condiciones observadas aproximadamente cinco meses atrás, evidenciando una dependencia temporal o estacionalidad.

5.2 Interpretación de los parámetros α y β.

De acuerdo a (Ning and Wang, 2016), el modelo de crecimiento neoclásico diferencial con retraso

$$x'(t) = -\alpha x(t) + \beta x^{\gamma}(t-\tau)e^{-\delta x(t-\tau)}$$

muestra el proceso de acumulación de capital, donde x es el capital por trabajo, α, β, γ y δ son parámetros positivos, δ refleja la fuerza de un efecto negativo causado por el incremento del capital, γ es un indicador para medir los rendimientos a escala. En este contexto, el primer término del modelo de crecimiento neoclásico un efecto reductor del capital por trabajor; mientras que, el segundo término una tiene contribución positiva a la acumulación de capital actual.

A partir de la interpretación planteada por (Ning and Wang, 2016), es posible establecer una analogía con los parámetros estimados en el presente estudio. En los modelos DDE correspondientes a las tres variables objetivo se obtuvo $\alpha < 0$ y $\beta > 0$. Por tanto, el término $\alpha x(t)$ introduce un efecto contemporáneo de reducción o ajuste sobre la dinámica de la variable, análogo, en términos de signo, al componente $-\alpha x(t)$ considerado en el modelo de crecimiento neoclásico.

Por otra parte, el valor positivo de β indica que el estado observado τ períodos atrás ejerce un efecto de incremento sobre la dinámica presente, evidenciando un mecanismo de dependencia temporal o memoria.

5.3 Contraste de los modelos de aprendizaje automático con el estado del arte

Los resultados obtenidos mediante XGBoost y Random Forest permiten contrastar la capacidad de los algoritmos de aprendizaje automático para modelar la dinámica del crecimiento empresarial ecuatoriano con los hallazgos reportados en la literatura. (Van Witteloostuijn and Kolkman, 2019) evidenciaron que Random Forest presentó una mayor capacidad predictiva que los modelos econométricos lineales tradicionales para explicar el crecimiento empresarial, atribuyendo esta ventaja a la capacidad del algoritmo para identificar relaciones no lineales e interacciones complejas entre las variables financieras y demográficas de las empresas. En este sentido, los resultados de la presente investigación respaldan la utilización de algoritmos de aprendizaje automático para representar un fenómeno cuya evolución no necesariamente responde a relaciones estrictamente lineales.

(Van Witteloostuijn and Kolkman, 2019) trabajaron con información individual correspondiente a empresas de Bélgica y los Países Bajos, considerando características como activos totales, capital

de trabajo, patrimonio, liquidez y solvencia. En cambio, la presente investigación analiza el crecimiento empresarial ecuatoriano y exportaciones observando la evolución mensual o anual de las variables objetivo: las tasas de crecimiento de sociedades y personas naturales y el total de ventas. Por esta razón, mientras que en el estudio citado la capacidad predictiva depende principalmente de características financieras propias de cada empresa, en esta investigación la predicción se sustenta en la información histórica de la propia variable objetivo y en variables relacionadas con el entorno macroeconómico y financiero.

Asimismo, los resultados se relacionan con lo señalado por (Gangwani and Zhu, 2024), quienes clasifican los factores empleados para estudiar el éxito o fracaso empresarial en variables de inversión, variables endógenas del negocio y variables exógenas relacionadas con las condiciones del mercado. En concordancia con este planteamiento, en la presente investigación se incorporaron variables macroeconómicas y financieras, entre ellas tasas de interés, inflación, pib y desempleo. Su inclusión permite considerar que la dinámica empresarial no depende únicamente de su comportamiento histórico, sino también de las condiciones económicas bajo las cuales operan las empresas.

Según (Vukovic et al., 2023), en su estudio, los autores incorporaron la rentabilidad del año anterior, ROA_{t-1}, como variable explicativa de la rentabilidad actual y encontraron un efecto positivo y estadísticamente significativo. Su inclusión incrementó el R^2 ajustado del modelo de $0{,}198$ a $0{,}418$, lo que evidencia que una proporción importante del comportamiento actual de la rentabilidad se encuentra relacionada con sus valores previos. Además, los autores señalan que la rentabilidad rezagada constituye uno de los factores más relevantes para la predicción y que su incorporación mejora la precisión de los modelos. En la presente investigación se observa un comportamiento similar, aunque aplicado a una variable distinta y con una frecuencia temporal mensual. En los modelos XGBoost y Random Forest se incorporaron rezagos de las variables objetivo seleccionados mediante el análisis PACF que nos indica que el comportamiento inmediatamente anterior contiene información relevante para estimar la variable objetivo en estudio.

5.4 Resultados

En términos generales, los tres modelos presentaron resultados competitivos; no obstante, las métricas de evaluación permitieron identificar diferencias en su capacidad predictiva según la variable objetivo analizada. Para la predicción de la tasa de crecimiento de las sociedades, XGBoost con variables exógenas presentó el mejor rendimiento, lo que evidencia que la incorporación de información adicional relevante contribuye a mejorar la capacidad predictiva del modelo. En el caso de la tasa de crecimiento de las personas naturales, el mejor desempeño correspondió al modelo

DDE con un retardo de $\tau = 36$ meses y una función inicial definida mediante un polinomio de segundo grado. Finalmente, para la predicción del total de ventas y exportaciones, el modelo DDE con $\tau = 5$ meses y una función inicial polinómica de primer grado presentó el mejor rendimiento predictivo.

CAPÍTULO 6.
CONCLUSIONES

- El análisis de los datos históricos permitió identificar que la evolución del número de empresas (sociedades y personas naturales con fines de lucro) y del total de ventas en Ecuador presenta una estructura temporal caracterizada por variaciones, cambios de tendencia, estacionalidad y dependencia respecto a observaciones anteriores (*lags*). Este comportamiento de los datos justificó el uso de modelos para series temporales que incorporan explícitamente información histórica, ya sea mediante rezagos en los algoritmos de aprendizaje automático o mediante un término de retardo en el modelo de ecuaciones diferenciales con retardo.

- La propuesta del modelo de ecuaciones diferenciales con retardo de la forma

  $$x'(t) = \alpha x(t) + \beta x(t - \tau),$$

  logró representar la dinámica empresarial mediante la interacción entre el estado actual del sistema y su comportamiento en el pasado. Los resultados obtenidos mostraron que el parámetro de retardo (τ) permite incorporar de forma explícita la memoria temporal de cada variable objetivo. Asimismo, en los modelos estimados se obtuvieron valores de ($\alpha < 0$) y ($\beta > 0$), lo que puede interpretarse como la existencia de un efecto reductor asociado al estado actual y una contribución positiva proveniente del estado observado (τ) períodos atrás.

  La selección del retardo y de la función inicial influyó en el desempeño del modelo DDE. Para la tasa de crecimiento de personas naturales, la mejor configuración correspondió a ($\tau = 36$) meses con una función inicial polinómica de segundo grado, mientras que para el total de ventas la mejor configuración se obtuvo con ($\tau = 5$) y una función inicial polinómica de primer grado. Esto evidencia que la escala temporal de memoria no es necesariamente la misma para todas las variables empresariales analizadas.

- Los algoritmos XGBoost y Random Forest se implementaron bajo un esquema de validación temporal Walk-Forward Validation, con el propósito de evaluar su capacidad predictiva respetando el orden cronológico de las observaciones y separando los datos en conjuntos de entrenamiento ($80\,\%$) y de prueba ($20\,\%$). Los resultados mostraron que ambos algoritmos son capaces de representar relaciones no lineales entre los valores históricos de las variables objetivo y las variables macroeconómicas y financieras incorporadas al modelo.

La incorporación de variables exógenas mostró que, en determinadas configuraciones, la información macroeconómica y financiera aporta capacidad predictiva adicional a la contenida en los propios rezagos de las series. Esto indica que la evolución empresarial no depende únicamente de su trayectoria histórica, sino también de las condiciones externas bajo las cuales se desarrolla la actividad económica.

- En cumplimiento del objetivo general, se desarrolló y evaluó un modelo basado en ecuaciones diferenciales con retardo para describir y predecir la dinámica temporal del crecimiento empresarial y del total de ventas en Ecuador, y se comparó su desempeño con XGBoost y Random Forest. Los resultados muestran que la DDE constituye una alternativa competitiva frente a los algoritmos de aprendizaje automático y que, dependiendo de la variable analizada, puede incluso presentar un mejor desempeño predictivo.

BIBLIOGRAFÍA

Banco Central del Ecuador (1998). Memoria anual 1997: Capítulo 8. Technical report, Banco Central del Ecuador, Quito, Ecuador.

Bobalová, M. and Novotná, V. (2021). Modeling of time delayed processes in business economics. In *Proceedings of the 15th International Management Conference "Managing People and Organizations in a Global Crisis*
, pages 79–89, Bucharest, Romania.

Brownlee, J. (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.

Burden, R. L. and Faires, J. D. (2010). *Numerical analysis*. Brooks - Cole Publishing Company Pacific Grove, CA:.

Chen, X., Liu, Z., Wang, Q., and Han, Z. (2025). Dynamic-informed machine learning: Data-driven reconstruction of delay differential equations models. *Applied Soft Computing*, 184:113743.

Chopra, R. (2025). *Essentials of Time Series Econometrics*. Educohack Press.

Coba, G. (2021). SRI: 3.494 empresas cerraron en el país durante la pandemia. *Primicias*.

Corte Constitucional del Ecuador (2006). Registro Oficial. Publicación oficial en el sistema ESACC, Quito, Jueves 11 de Mayo del 2006 – Nº 268.

Deka, P. P. and Weiner, J. (2024). *XGBoost for Regression Predictive Modeling and Time Series Analysis: Learn How to Build, Evaluate, and Deploy Predictive Models with Expert Guidance*. Packt Publishing.

Dhanaraj, R. K., Samuel, P., Sathyamoorthy, M., Balusamy, B., and Ravi, V. (2024). *Self-powered Sensors: A Path to Wearable Electronics*. Elsevier.

Diario Expreso (2020). El 25 % de los emprendedores en Ecuador cierra su negocio en época de COVID-19. *Diario Expreso*.

El Universo (2006). SRI eliminó a veinte mil contribuyentes del RUC. *El Universo*.

Eurostat (2025). Business demography statistics.

Fernanda, J., Marley, R., and Suhendra, F. (2024). Exploring the impact of macroeconomic factors on company financial performance: Insights into economic indicators and sectoral heterogeneity. *Journal on Economics, Management and Business Technology*, 2(2):73–83.

Franco, P. T. (2022). La constitución de compañías creció un 49 % en el 2021, con 16.338 nuevas empresas. *El Universo*.

Gangwani, D. and Zhu, X. (2024). Modeling and prediction of business success: A survey. *Artificial Intelligence Review*, 57:44.

Guerrini, L., Krawiec, A., and Szydłowski, M. (2020). Bifurcations in an economic growth model with a distributed time delay transformed to ODE. *Nonlinear Dynamics*, 101:1263–1279.

Hernández, M. (2014). *Administración de empresas*. Ediciones Pirámide.

Mayorga, C. J. (2024). *Soluciones numéricas exactas y esquemas en diferencias finitas no estándar para ecuaciones diferenciales con retardo*. Tesis doctoral, Universidad de Alicante. https://rua.ua.es/bitstream/10045/148146/1/tesis_carlos_julio_mayorga_arias.pdf.

Minniti, M., Bygrave, W. D., and Autio, E. (2006). Global entrepreneurship monitor: 2005 executive report. Executive report, Babson College and London Business School.

Ning, Z. and Wang, W. (2016). The existence of two positive periodic solutions for the delay differential neoclassical growth model. *Advances in Difference Equations*, 2016(1):266.

Oleas-Montalvo, J. (2017). Ecuador 1980-1990: crisis, ajuste y cambio de régimen de desarrollo. *América Latina en la historia económica*, 24(1):210–242.

Primicias (2026). La ruta del lavado de activos: en Ecuador circulan al menos usd 5.000 millones ilícitos. *Revista Gestión - Primicias*.

Servicio de Rentas Internas (2019). Informe de Labores 2018. Informe institucional, Servicio de Rentas Internas (SRI), Quito, Ecuador. Accedido: 1 de agosto de 2026.

Subramanian, V. (2025). *Applied machine learning for data science practitioners*. John Wiley & Sons.

Sylvester, J. H., Thornton, M. A., Larson, E. C., and Thornton, M. A. (2026). Lag order selection for granger causality estimation. In *Intelligent Computing: Proceedings of the 2026 Computing Conference, Volume 3*, volume 3, page 102. Springer Nature.

Van Witteloostuijn, A. and Kolkman, D. (2019). Is firm growth random? a machine learning perspective. *Journal of Business Venturing Insights*, 11:Artículo e00107.

Vukovic, D. B., Spitsina, L., Gribanova, E., Spitsin, V., and Lyzin, I. (2023). Predicting the performance of retail market firms: Regression and machine learning methods. *Mathematics*, 11(8):1916.

Zakaria, M. N. A., Ahmed, A. N., Malek, M. A., Birima, A. H., Khan, M. M. H., Sherif, M., and Elshafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in muda river, malaysia. *Heliyon*, 9(7).

CAPÍTULO A.
ANEXOS

A.1 Resultados modelo DDE

A.1.1 Tasa de crecimiento de sociedades

Tabla A.1: Métricas del modelo DDE con $\tau = 36$.

Función Inicial ($\tau = 36$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,099419	0,138092	17,250083
Polinomio 2do grado	0,099030	0,138190	17,092440
Polinomio 3er grado	0,097244	0,140496	16,627742
Polinomio 4to grado	0,099068	0,136741	17,211762
Cubic Spline	0,099699	0,137494	17,290101

Tabla A.2: Métricas del modelo DDE con $\tau = 42$.

Función Inicial ($\tau = 42$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,099940	0,137366	17,416901
Polinomio 2do grado	0,100933	0,137885	17,380756
Polinomio 3er grado	0,101437	0,136892	17,685021
Polinomio 4to grado	0,103197	0,141137	17,876242
Cubic Spline	0,101101	0,137383	17,593956

Tabla A.3: Métricas del modelo DDE con $\tau = 48$.

Función Inicial ($\tau = 48$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,112527	0,142713	18,521350
Polinomio 2do grado	0,114975	0,145922	18,907049
Polinomio 3er grado	0,115244	0,145611	18,875278
Polinomio 4to grado	0,113233	0,143933	18,693878
Cubic Spline	0,117435	0,148075	19,450230

Tabla A.4: Métricas del modelo DDE con $\tau = 54$.

Función Inicial ($\tau = 54$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,099369	0,134762	17,115244
Polinomio 2do grado	0,104125	0,140459	17,958865
Polinomio 3er grado	0,097306	0,135841	16,581087
Polinomio 4to grado	0,096479	0,135760	16,557778
Cubic Spline	0,102772	0,139217	17,476408

Tabla A.5: Métricas del modelo DDE con $\tau = 60$.

Función Inicial ($\tau = 60$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,099317	0,133441	17,276211
Polinomio 2do grado	0,104605	0,141112	18,216869
Polinomio 3er grado	0,100999	0,137541	17,643648
Polinomio 4to grado	0,102200	0,139960	17,915955
Cubic Spline	0,100406	0,137103	17,492661

A.1.2 Tasa de crecimiento de personas naturales

Tabla A.6: Métricas del modelo DDE con $\tau = 36$.

Función Inicial ($\tau = 36$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,134712	0,191207	28,2502
Polinomio 2do grado	0,132005	0,187698	27,5124
Polinomio 3er grado	0,135954	0,192478	29,2998
Polinomio 4to grado	0,133697	0,189326	27,8669
Cubic Spline	0,136490	0,193421	27,9664

Tabla A.7: Métricas del modelo DDE con $\tau = 42$.

Función Inicial ($\tau = 42$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,146496	0,198370	29,7000
Polinomio 2do grado	0,131841	0,181831	27,3323
Polinomio 3er grado	0,154117	0,205461	31,0978
Polinomio 4to grado	0,129192	0,182333	30,7245
Cubic Spline	0,153500	0,205990	31,8136

Tabla A.8: Métricas del modelo DDE con $\tau = 48$.

Función Inicial ($\tau = 48$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,135375	0,185610	33,7161
Polinomio 2do grado	0,138145	0,187033	34,2461
Polinomio 3er grado	0,136846	0,187038	33,0057
Polinomio 4to grado	0,134840	0,185280	33,8835
Cubic Spline	0,136115	0,186634	34,1210

Tabla A.9: Métricas del modelo DDE con $\tau = 54$.

Función Inicial ($\tau = 54$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,139787	0,199512	28,3620
Polinomio 2do grado	0,139379	0,198757	29,1584
Polinomio 3er grado	0,126572	0,175148	22,2611
Polinomio 4to grado	0,138736	0,196870	28,0221
Cubic Spline	0,139963	0,198480	29,4708

Tabla A.10: Métricas del modelo DDE con $\tau = 60$.

Función Inicial ($\tau = 60$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	0,138995	0,200551	30,7870
Polinomio 2do grado	0,136949	0,198161	31,4861
Polinomio 3er grado	0,137415	0,197912	31,6771
Polinomio 4to grado	0,137463	0,197626	30,2673
Cubic Spline	0,136775	0,197972	30,7642

A.1.3 Total ventas y exportaciones

Tabla A.11: Métricas del modelo DDE con $\tau = 4$.

Función Inicial ($\tau = 4$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$3{,}172559 \times 10^9$	$4{,}934302 \times 10^9$	13,057098
Polinomio 2do grado	$3{,}056844 \times 10^9$	$4{,}810985 \times 10^9$	12,397067
Polinomio 3er grado	$2{,}935077 \times 10^9$	$4{,}438390 \times 10^9$	11,905247
Polinomio 4to grado	$3{,}065618 \times 10^9$	$4{,}655440 \times 10^9$	12,441793
Cubic Spline	$3{,}155638 \times 10^9$	$4{,}773711 \times 10^9$	12,964002

Tabla A.12: Métricas del modelo DDE con $\tau = 5$.

Función Inicial ($\tau = 5$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$1{,}752034 \times 10^9$	$2{,}790778 \times 10^9$	6,962525
Polinomio 2do grado	$1{,}776612 \times 10^9$	$2{,}745917 \times 10^9$	7,240085
Polinomio 3er grado	$2{,}196790 \times 10^9$	$3{,}574591 \times 10^9$	8,798902
Polinomio 4to grado	$1{,}870063 \times 10^9$	$3{,}082022 \times 10^9$	7,415137
Cubic Spline	$1{,}845871 \times 10^9$	$3{,}013210 \times 10^9$	7,346500

Tabla A.13: Métricas del modelo DDE con $\tau = 6$.

Función Inicial ($\tau = 6$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$3{,}672981 \times 10^9$	$5{,}868770 \times 10^9$	15,550199
Polinomio 2do grado	$3{,}070377 \times 10^9$	$4{,}689383 \times 10^9$	12,543217
Polinomio 3er grado	$3{,}084604 \times 10^9$	$4{,}788193 \times 10^9$	12,856445
Polinomio 4to grado	$3{,}782980 \times 10^9$	$5{,}659562 \times 10^9$	16,148521
Cubic Spline	$3{,}331398 \times 10^9$	$5{,}299914 \times 10^9$	13,973723

Tabla A.14: Métricas del modelo DDE con $\tau = 7$.

Función Inicial ($\tau = 7$)	**MAE**	**RMSE**	**MAPE (%)**
Polinomio 1er grado	$3{,}210036 \times 10^9$	$5{,}384308 \times 10^9$	13,677634
Polinomio 2do grado	$3{,}711993 \times 10^9$	$4{,}982708 \times 10^9$	15,902792
Polinomio 3er grado	$3{,}079125 \times 10^9$	$5{,}133887 \times 10^9$	13,154636
Polinomio 4to grado	$3{,}014946 \times 10^9$	$5{,}068953 \times 10^9$	12,844687
Cubic Spline	$3{,}483670 \times 10^9$	$4{,}853634 \times 10^9$	14,768144