

Departamento de Economía

Tipo de documento: Tesis de maestría



Maestría en Econometría

Predicción de la Actividad Económica Sectorizada a Corto Plazo en República Dominicana (2010–2025) mediante Técnicas Econométricas y de Machine Learning

Autoría: Burgos Castillo, María Tatiana

Fecha: 2025

¿Cómo citar este trabajo?

Burgos Castillo, M. (2025). "Predicción de la Actividad Económica Sectorizada a Corto Plazo en República Dominicana (2010–2025) mediante Técnicas Econométricas y de Machine Learning". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/13559>

El presente documento se encuentra alojado en el Repositorio Digital de la **Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



DEPARTAMENTO DE ECONOMÍA

MAESTRÍA EN ECONOMETRÍA

**Predicción de la Actividad Económica Sectorizada a Corto Plazo en
República Dominicana (2010–2025) mediante Técnicas
Econométricas y de Machine Learning**

Alumno: María Tatiana Burgos Castillo

Legajo: 23T3035

Tutor: Magdalena Cornejo

Fecha: junio 2025

Resumen

El presente estudio desarrolla un sistema de predicción sectorial de la actividad económica en la República Dominicana, combinando técnicas econométricas tradicionales y métodos avanzados de aprendizaje automático. El objetivo es ofrecer estimaciones mensuales actualizadas y precisas que permitan anticipar cambios coyunturales en los sectores clave del Producto Interno Bruto (PIB): industrial, manufactura local, zonas francas, agropecuario, servicios y construcción. Para ello, se emplearon modelos AutoARIMA, Support Vector Regression, Random Forest (básico y optimizado), y redes neuronales (CNN y RNN), entrenados con más de 150 series económicas transformadas, y evaluados mediante validación cruzada temporal y métricas de desempeño relativo fuera de la muestra. Los resultados revelan que la capacidad predictiva varía según el sector y el tipo de modelo, destacándose Random Forest optimizado por su solidez y estabilidad predictiva. Esta herramienta contribuye significativamente a mejorar el monitoreo económico sectorial, permitiendo una planificación de políticas públicas más eficiente, basada en evidencia. Se identifican también limitaciones asociadas a la disponibilidad de datos, lo que abre líneas de investigación futura centradas en el uso de nowcasting con variables externas y la implementación de técnicas de selección automática de predictores.

Palabras clave: Predicción económica, aprendizaje automático, AutoARIMA, Random Forest, redes neuronales, actividad sectorial, República Dominicana.

Abstract

This thesis develops a sectoral forecasting system for economic activity in the Dominican Republic, combining traditional econometric models with advanced machine learning techniques. The objective is to provide timely and accurate monthly estimates that anticipate short-term fluctuations in key sectors of the Gross Domestic Product (GDP): industrial, local manufacturing, free trade zones, agriculture, services, and construction. The models used include AutoARIMA, Support Vector Regression, Random Forest (basic and optimized), and neural networks (CNN and RNN), trained on over 150 transformed economic series and evaluated through time series cross-validation and out-of-sample performance metrics. Results indicate that predictive performance varies by sector and model, with optimized Random Forest demonstrating strong generalization and stability. This tool significantly enhances sectoral economic monitoring, enabling more effective and evidence-based policy planning. The study also recognizes limitations related to the forecasting horizon, data availability, and model interpretability, suggesting future research directions in nowcasting with external variables and automatic predictor selection techniques.

Keywords: Economic forecasting, machine learning, AutoARIMA, Random Forest, neural networks, sectoral activity, Dominican Republic.

Índice

I.	Introducción	5
II.	Antecedentes	7
III.	Metodología	10
	3.1 Diseño de la estrategia de predicción	10
	3.2 Especificación de los modelos	10
	3.2.1 Modelo ARIMA (<i>AutoRegressive Integrated Moving Average</i>)	10
	3.2.2 Modelo Random Forest (Especificación Base)	12
	3.2.3 Modelo Random Forest con Optimización de Hiperparámetros.....	13
	3.2.4 Support Vector Regression (SVR).....	14
	3.2.5 Modelo de Red Neuronal LSTM (Long Short-Term Memory)	16
	3.2.6 Modelo de Redes Neuronales Convolucionales (CNN).....	17
IV.	Datos	19
	4.1 Fuentes de Información, Estructura Temporal y Variables Predictoras.....	19
	Figura 1: Evolución temporal de la tasa de crecimiento trimestral del PIB por sector económico	20
	Tabla 1: Variables Predictoras Relevantes por Sector Económico.....	21
	4.2 Transformaciones Aplicadas	21
	Tabla 2: Variables Predictoras	23
	4.3 Estadísticos Descriptivos	26
	Tabla 3: Estadísticas descriptivas de las tasas de crecimiento anuales de las variables utilizadas (%)	27
	4.4 Visualización de Series Sectoriales de Actividad Económica (Figuras 2 a 7).....	28
	Figura 2: Actividad económica industrial	28
	Figura 3: Actividad económica manufactura	29
	Figura 4: Actividad económica de las zonas francas	29
	Figura 5: Actividad económica construcción.....	30
	Figura 6: Actividad económica agropecuaria	31
	Figura 7: Actividad económica servicios.....	31
V.	Resultados	32
	5.1 Evaluación de los Modelos.....	32
	5.2 Evaluación Out-of-Sample y capacidad predictiva	33
	Tabla 4: Evaluación de los Modelos en Out-of-Sample.....	33
	5.3 Interpretabilidad y Relevancia de los Predictores en los Modelos Random Forest Aplicados	34
	Figura 8: Importancia Relativa de los Predictores en Modelos Random Forest	35
	5.4 Pronósticos sectoriales a corto plazo (Figuras 9 a 14).....	35
	Figura 9: Actividad económica industrial a corto plazo.....	35
	Figura 10: Actividad económica manufactura a corto plazo	36
	Figura 11: Actividad económica de las zonas francas a corto plazo.....	37
	Figura 12: Actividad económica agropecuaria a corto plazo	37
	Figura 13: Actividad económica servicios a corto plazo	38

Figura 14: Actividad económica construcción a corto plazo	39
VI. Conclusión	40
VII. Referencias Bibliográficas	42
VIII. Anexos	44
Apéndice A: Transformaciones Aplicadas a las Variables Predictoras	44
Apéndice B: Pruebas.....	46

I. Introducción

El desarrollo económico sostenible de la República Dominicana depende en gran medida del desempeño de sectores estratégicos como la industria manufacturera, el agropecuario, las zonas francas, la construcción y los servicios. Estos sectores, aunque interrelacionados, presentan dinámicas productivas diferenciadas que responden a múltiples factores, incluyendo la inversión, la productividad, los flujos comerciales y la demanda interna y externa. Dado su peso relativo en la generación de empleo, ingresos fiscales y divisas, su evolución determina de forma directa la trayectoria del crecimiento económico nacional. En este sentido, comprender y anticipar el comportamiento de estas actividades no es solo un ejercicio académico, sino una necesidad crítica para el diseño de políticas públicas eficaces y orientadas al desarrollo.

En un entorno macroeconómico volátil y complejo, marcado por choques externos, tensiones inflacionarias, disrupciones logísticas y cambios tecnológicos acelerados, se vuelve imperativo contar con herramientas de predicción sectorial que permitan anticipar el comportamiento económico a corto y mediano plazo. La capacidad de descomponer y proyectar la evolución de la actividad económica por sectores no solo permite identificar patrones divergentes y puntos de quiebre, sino también estimar riesgos específicos, facilitando así una mejor asignación de recursos públicos y una planificación económica más eficaz.

Los métodos econométricos tradicionales, como los modelos ARIMA y VAR, han sido ampliamente utilizados para fines de predicción. No obstante, su capacidad de capturar no linealidades, interacciones de alta dimensión y eventos atípicos resulta limitada, especialmente en contextos de incertidumbre estructural. En contraste, los avances en ciencia de datos y aprendizaje automático han proporcionado nuevas herramientas que permiten procesar grandes volúmenes de información, adaptarse a cambios en la estructura de los datos y generar pronósticos con mayor precisión y resiliencia.

Ante este escenario, el presente estudio propone el desarrollo de un sistema de predicción sectorial de la actividad económica real de la República Dominicana que combine técnicas econométricas clásicas con modelos modernos de *machine learning*. Esta integración metodológica busca mejorar sustancialmente la calidad de las estimaciones, generar evidencia empírica relevante para la formulación de políticas públicas sectoriales.

Los resultados empíricos evidencian la utilidad de combinar modelos econométricos tradicionales y métodos modernos de *machine learning* para anticipar la evolución sectorial de la actividad económica en la República Dominicana. Aunque no existe un modelo único que se imponga sobre los demás para

todos los sectores, se destacan ciertos patrones: AutoARIMA muestra un buen desempeño en sectores con continuidad estructural, como zonas francas; los modelos basados en redes neuronales (RNN¹ y CNN²) sobresalen en sectores con estacionalidades o comportamientos no lineales marcados, como la agricultura y la construcción; mientras que el modelo de RF³ optimizado se presenta como el más robusto y versátil, logrando un equilibrio entre precisión, estabilidad y adaptabilidad. Estas evidencias sugieren que el enfoque híbrido adoptado representa una herramienta valiosa para la planificación económica sectorial.

En este tenor, los resultados del ejercicio predictivo muestran que no existe un único modelo óptimo para todos los sectores, pero sí un patrón claro de desempeño relativo. En el sector industrial, el modelo RF optimizado arrojó la mejor proyección con un crecimiento estimado de 6.1% y un RMSE de 0.72, ajustándose mejor que AutoARIMA (11.9%) que tendió al sobreajuste. En manufactura local, el modelo RNN ofreció una estimación más estable con un crecimiento de 3.2% y un MAE de 0.45, frente al 6.5% proyectado por AutoARIMA. Para zonas francas, donde predomina la volatilidad, AutoARIMA captó mejor los giros abruptos (proyección de 6.7%) aunque con mayor dispersión, mientras que RF básico logró menor varianza en la predicción (RMSE de 0.85). En el sector agropecuario, todos los modelos convergieron en un rango estrecho entre 3.4% y 4.8%, destacándose CNN por su precisión con un MAE de 0.36. En servicios, la versión del RF optimizado también lideró en estabilidad, estimando un crecimiento de 4.7% y un RMSE de 0.69. Finalmente, en construcción, el modelo RNN se comportó con mayor robustez, proyectando un crecimiento del 6.2%, con un RMSE de 0.78, frente a las proyecciones más extremas de AutoARIMA (13%) y SVR (-2.7%). Estos resultados confirman que los modelos de aprendizaje profundo, como RNN y CNN, tienden a suavizar las predicciones y adaptarse mejor en sectores con estacionalidad o alta sensibilidad a choques, mientras que el RF optimizado logra un balance notable entre precisión, estabilidad y capacidad de generalización en contextos diversos.

Finalmente, este estudio se organiza en cinco secciones principales. La introducción contextualiza el problema, define los objetivos y justifica la relevancia del estudio. La revisión de la literatura presenta los antecedentes conceptuales y metodológicos sobre la predicción sectorial de la actividad económica. En la sección de metodología, se detalla el enfoque cuantitativo adoptado, incluyendo la selección de variables, el procesamiento de datos y la especificación de los modelos. La sección de datos describe el conjunto informativo, los criterios de inclusión de predictores y las transformaciones realizadas.

¹ Modelo de Red Neuronal Recurrente (RNN)

² Modelo de Red Neuronal Convolutiva (CNN)

³ *Random Forest* (RF)

La sección de resultados desarrolla el análisis empírico de las predicciones por sector, comparando el desempeño relativo de los modelos aplicados mediante métricas estadísticas y pruebas de significancia. Finalmente, la sección de conclusiones sintetiza los principales hallazgos del estudio, discute sus implicaciones para el diseño de políticas económicas basadas en evidencia y expone las principales limitaciones de la investigación, así como posibles líneas de investigación futuras.

II. Antecedentes

Esta sección presenta los principales desarrollos teóricos y empíricos en torno a los modelos utilizados para predecir la actividad económica agregada y sectorial, tanto desde la perspectiva econométrica tradicional como desde enfoques más recientes basados en técnicas de *machine learning*. Se analizan trabajos fundamentales que introducen modelos de factores dinámicos, el uso del filtro de Kalman para imputación de datos y predicción en tiempo real (*nowcasting*), así como estudios que exploran el potencial predictivo de algoritmos como *Random Forest*, *Support Vector Machines* y Redes Neuronales en contextos económicos. Esta revisión permite identificar las fortalezas y limitaciones de cada enfoque, y sirve como base para justificar la elección de modelos híbridos en esta tesis, orientados a predecir la evolución de la actividad económica de seis sectores estratégicos del PIB real de la República Dominicana.

Los modelos de factores dinámicos (DFM) tienen sus orígenes en el trabajo de Geweke (1977) quien adaptó este tipo de modelos originalmente aplicados en estudios de corte transversal para el análisis de series de tiempo macroeconómicas. Posteriormente, Sargent & Sims (1977) demostraron empíricamente que un conjunto reducido de factores comunes podía explicar una proporción significativa de la varianza en variables clave como el producto interno bruto (PIB), los precios y el empleo.

El desarrollo formal del enfoque de factores dinámicos para la medición de la actividad económica se atribuye a Stock & Watson (1991) quienes propusieron un modelo en el cual se asume la existencia de un factor no observable que afecta simultáneamente a diversas series macroeconómicas. Utilizando datos mensuales de Estados Unidos entre 1959 y 1987 como producción industrial, ingreso real, ventas minoristas y empleo no agrícola, los autores encontraron que este factor latente estaba fuertemente correlacionado con el índice oficial de actividad publicado por el Departamento de Comercio.

Más adelante, Mariano & Murasawa (2003) extendieron el enfoque original al combinar series de diferentes frecuencias temporales (mensuales y trimestrales). En su modelo, el PIB real trimestral se combina con indicadores mensuales vinculados al ciclo económico, como el empleo o la producción

industrial. La estimación se realiza mediante máxima verosimilitud, tratando las series trimestrales como si fueran mensuales con observaciones faltantes, las cuales son completadas con simulaciones estocásticas bajo supuestos de independencia con respecto a los parámetros del modelo.

Inspirándose en estos desarrollos, Camacho & Pérez Quiros (2008) aplicaron un modelo de factores dinámicos para predecir el crecimiento del PIB en tiempo real en la Zona Euro. Para ello, combinaron series mensuales y trimestrales y utilizaron el filtro de Kalman para imputar los datos faltantes en las series cortas hacia atrás. Además, construyeron una base de datos en tiempo real, que incluye solo la información disponible en cada fecha de pronóstico, lo que les permitió evaluar el impacto de la nueva información sobre las proyecciones actualizadas del PIB.

Con base en ese trabajo, Camacho & Pérez Quiroz (2010) adaptaron el modelo al contexto español, diseñando un índice coincidente que resume el ciclo económico. Este indicador demostró ser útil para anticipar la desaceleración económica en España durante 2009, mostrando un buen desempeño predictivo en el corto plazo.

En una extensión regional, Camacho & Pérez Quiroz (2011) aplicaron esta misma metodología a varios países latinoamericanos (Argentina, Brasil, Chile, Colombia, México y Perú). Para cada economía, desarrollaron indicadores de crecimiento económico que mostraron una alta correlación con los principales índices de actividad nacionales, como el IMACEC en Chile y el EMAE en Argentina.

En el contexto latinoamericano, el trabajo de Paredes, Hawa, & Polanco (2015) constituye una referencia clave para el desarrollo de modelos de factores dinámicos aplicados a la predicción del PIB en la República Dominicana. Siguiendo el enfoque de Stock & Watson (1991) y las adaptaciones posteriores de Camacho y Quiroz, autores implementan un modelo que combina series temporales de distintas frecuencias mensuales y trimestrales, integrando indicadores como exportaciones, importaciones, crédito privado, ventas fiscales, remesas, consumo de cemento, y transacciones en tarjetas de crédito.

Una de las principales innovaciones de este estudio es el uso del Filtro de Kalman para la estimación recursiva del componente común no observado que resume la dinámica de la actividad económica sectorial.

Este procedimiento permite una estimación eficiente en presencia de datos incompletos o rezagados, condición común en economías como la dominicana. Así, el Filtro de Kalman resulta especialmente útil para generar proyecciones del PIB trimestral a partir de información mensual.

La capacidad de este modelo para actualizar pronósticos en tiempo real convierte al enfoque de Paredes, Hawa, & Polanco, (2015) en una contribución metodológica altamente relevante para estudios de *nowcasting* y *shortterm forecasting* en países emergentes.

La técnica de *nowcasting* ha adquirido relevancia en el análisis macroeconómico moderno debido a la necesidad de generar estimaciones de alta frecuencia para variables como el Producto Interno Bruto (PIB), cuyas publicaciones oficiales suelen tener un rezago temporal de entre 60 y 90 días. Esta brecha informativa puede limitar la capacidad de respuesta de los hacedores de política económica en contextos de alta volatilidad. En ese sentido, el *nowcasting* se presenta como una herramienta eficaz para anticipar el comportamiento de la actividad económica agregada, mediante el uso de indicadores que están disponibles con mayor frecuencia y oportunidad.

El estudio de Barrios, Escobar, Leslie, Martín, & Peña (2021) desarrollado por el Banco Interamericano de Desarrollo (BID), constituye un referente clave en esta materia. En él se implementan técnicas de aprendizaje automático supervisado para estimar el crecimiento del PIB trimestral en Belice y El Salvador, empleando un amplio conjunto de variables macroeconómicas con frecuencia mensual y trimestral. La base metodológica incluye modelos como *Random Forest*, Breiman (2001), *Gradient Boosting Machines*, Natekin & Knoll (2013), Redes Neuronales, así como regresiones penalizadas como *Lasso*, *Ridge* y *Elastic Net*.

El uso de *machine learning* en este contexto ofrece ventajas sustanciales frente a métodos tradicionales. En primer lugar, los modelos priorizan el desempeño fuera de muestra (*out-of-sample*), lo cual es crucial para generar pronósticos robustos. En segundo lugar, permiten modelar relaciones no lineales y complejas entre variables, mejorando la precisión predictiva incluso en condiciones de cambio estructural o eventos extremos como los provocados por la pandemia del COVID-19.

Además, los ejercicios realizados en ambos países muestran la adaptabilidad del enfoque. En el caso de Belice, el modelo predijo con notable precisión las caídas del PIB observadas en los trimestres de 2020 y principios de 2021, con errores mínimos (RMSE < 2 puntos porcentuales en promedio). En El Salvador, los modelos también anticiparon correctamente la dirección del crecimiento, superando en precisión a los modelos tradicionales basados en filtros de Kalman o modelos puente.

En términos metodológicos, las variables fueron transformadas en tasas de crecimiento trimestral interanual para enfrentar problemas de no estacionariedad. Las variables con frecuencia mensual se clasificaron según su naturaleza (flujo, *stock* o índice), aplicando técnicas de agregación trimestral coherentes con la estructura del PIB publicado. También se realizaron ejercicios de validación cruzada y prueba fuera de muestra, para confirmar la solidez de las predicciones.

El estudio concluye que los modelos de *machine learning* constituyen una alternativa metodológica potente para contextos de baja capacidad estadística, permitiendo producir estimaciones inmediatas y confiables del crecimiento económico.

III. Metodología

En este apartado se describe la estrategia metodológica empleada para la predicción de la actividad económica sectorizada en República Dominicana a corto plazo. El enfoque combina técnicas clásicas de series temporales con métodos modernos de aprendizaje automático, seleccionados por su capacidad para capturar dinámicas temporales, relaciones no lineales y estructuras complejas en los datos. Se especifican y aplican seis modelos predictivos: ARIMA, *Random Forest* (versión base y con optimización de hiperparámetros), Support Vector *Regression* (SVR), redes neuronales LSTM y redes convolucionales (CNN). Cada uno de estos modelos se adapta a la naturaleza de los datos económicos y permite evaluar distintas aproximaciones al problema de *forecasting* sectorial.

3.1 Diseño de la estrategia de predicción

La investigación adopta un enfoque cuantitativo y predictivo, combinando modelos econométricos tradicionales de series temporales con técnicas modernas de aprendizaje automático. El objetivo es modelar y predecir la evolución de seis sectores clave del Producto Interno Bruto (PIB) real de la República Dominicana: agropecuario, industrial, manufactura local, manufactura de zonas francas, construcción y servicios.

3.2 Especificación de los modelos

3.2.1 Modelo ARIMA (*AutoRegressive Integrated Moving Average*)

El modelo ARIMA (*Autoregressive Integrated Moving Average*), desarrollado por Box & Jenkins (1970) constituye una de las herramientas más consolidadas y empleadas en el análisis y pronóstico de series temporales. La representación general del modelo es:

$$ARIMA(p, d, q): \phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t \quad (1)$$

donde:

- B es el operador rezago,
- $\phi(B)$ y $\theta(B)$ son polinomios en B que representan los componentes AR y MA, respectivamente,
- d es el número de diferenciaciones requeridas para estacionariedad,
- ϵ_t es el término de error blanco.

El modelo ARIMA se denota como ARIMA (p,d,q) , donde:

- p : Número de rezagos en la parte autorregresiva (AR).
- d : Número de diferenciaciones aplicadas para lograr estacionariedad.
- q : Número de términos en la parte de media móvil (MA).

Combina tres componentes clave:

- AR (Autoregresivo): Modela la relación entre una observación y sus valores pasados mediante un proceso autorregresivo de orden P . Se expresa como:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \epsilon_t \quad (2)$$

donde ϕ_i son los coeficientes autorregresivos y ϵ_t es un término de error.

- I (Integrado): Representa el número de diferenciaciones necesarias para hacer estacionaria la serie. Si una serie no es estacionaria, se aplica una transformación de diferencias:

$$Y'_t = Y_t - Y_{t-1} \quad (3)$$

- MA (Media Móvil): Modela la relación entre una observación y los errores pasados mediante un proceso de media móvil de orden q .

$$Y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (4)$$

donde θ_i son los coeficientes de la media móvil.

Si los datos presentan estacionalidad, se extiende a SARIMA (*Seasonal* ARIMA), agregando componentes estacionales:

$$ARIMA(p,d,q) \times (P,D,Q)_s \quad (5)$$

donde $(P,D,Q)_s$ representa los términos estacionales de orden s .

3.2.2 Modelo Random Forest (Especificación Base)

El *Random Forest Regressor* es un algoritmo de aprendizaje supervisado basado en ensamblaje de múltiples árboles de decisión, diseñado para mejorar la precisión de las predicciones y mitigar el riesgo de sobreajuste. Introducido por Breiman (2001), este modelo se fundamenta en la técnica de *bagging* (*bootstrap aggregating*), que consiste en generar múltiples subconjuntos aleatorios de los datos originales con reemplazo para entrenar diferentes árboles de decisión de forma independiente.

Las predicciones finales se obtienen mediante el promedio de las predicciones individuales de cada árbol, lo que reduce la varianza del modelo sin incrementar sustancialmente el sesgo.

Una de sus principales fortalezas es su capacidad para capturar relaciones no lineales complejas y manejar interacciones de alto orden entre variables predictoras. Además, permite evaluar la importancia relativa de cada variable en el proceso de predicción, lo cual resulta especialmente útil en contextos de análisis económico.

El RF básico ha demostrado ser eficaz en aplicaciones de *nowcasting* del PIB, al ofrecer predicciones estables y robustas frente a estructuras de datos altamente volátiles y multidimensionales Bolhuis & Rayner (2020).

Dado un conjunto de datos con variables predictoras X y una variable objetivo y , el RF básico realiza múltiples particiones aleatorias y genera árboles de decisión que buscan minimizar el error cuadrático medio:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

donde y_i son los valores reales y \hat{y}_i las predicciones del modelo. El modelo también minimiza el error absoluto medio (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

y su versión en raíz cuadrada, el RMSE (Raíz del Error Cuadrático Medio):

$$RMSE = \sqrt{MSE} \quad (8)$$

Una de las ventajas clave del RF básico es su capacidad para medir la importancia de las características, evaluando cuánto contribuye cada variable en la reducción del error en los árboles de decisión.

Esta importancia se calcula mediante la reducción del error cuadrático cuando se excluye una variable en la predicción.

El código implementa un *Random Forest Regressor* con 100 árboles y ajusta el modelo con los datos de entrenamiento. Luego, predice sobre el conjunto de prueba y calcula las métricas MSE, MAE, RMSE.

Este enfoque garantiza que el modelo se entrene de manera eficiente, descartando variables irrelevantes y maximizando la interpretabilidad de los resultados, lo cual es clave en la producción de series temporales o modelos de *machine learning* en economía y finanzas.

3.2.3 Modelo Random Forest con Optimización de Hiperparámetros

La optimización del rendimiento del modelo *Random Forest* se logra a través del ajuste de hiperparámetros clave utilizando *BayesSearchCV*, una técnica basada en optimización bayesiana. A diferencia de enfoques tradicionales como el *grid search* o la búsqueda aleatoria, esta metodología permite explorar el espacio de hiperparámetros de manera más eficiente y dirigida, priorizando aquellas combinaciones que tienen mayor probabilidad de mejorar la métrica objetivo. Esto reduce significativamente el tiempo computacional y mejora la capacidad predictiva del modelo, al encontrar configuraciones óptimas que maximizan el desempeño fuera de muestra.

El modelo se optimiza ajustando los siguientes hiperparámetros clave:

- Número de árboles (*n_estimators*): Controla cuántos árboles se crean en el bosque.
- Profundidad máxima (*max_depth*): Limita la profundidad de los árboles para evitar sobreajuste.
- Número mínimo de muestras para dividir un nodo (*min_samples_split*): Determina cuándo un nodo puede dividirse.
- Número mínimo de muestras por hoja (*min_samples_leaf*): Define el tamaño mínimo permitido para los nodos hoja.

En la optimización se usa validación cruzada en series temporales *TimeSeriesSplit*, que divide la serie en múltiples conjuntos de entrenamiento y prueba manteniendo el orden cronológico.

La búsqueda de hiperparámetros se realiza con *BayesSearchCV*, que selecciona de forma iterativa las mejores combinaciones de parámetros evaluando el rendimiento a partir del coeficiente de determinación (R^2):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (9)$$

Una vez finalizada la optimización, se entrena el mejor modelo obtenido y se generan predicciones en el conjunto de prueba (X_{test}). Este enfoque permite optimizar el rendimiento del *Random Forest Regressor* de manera eficiente y escalable, asegurando un mejor ajuste del modelo a los datos sin riesgo de sobreajuste.

3.2.4 Support Vector Regression (SVR)

El modelo de *Support Vector Regression* (SVR) representa una extensión del algoritmo *Support Vector Machines* (SVM), diseñado para abordar problemas de regresión en lugar de clasificación. Su principal objetivo es estimar una función que aproxime los valores observados de la variable dependiente, permitiendo cierto margen de error (ϵ) dentro del cual las desviaciones no son penalizadas. Esta tolerancia controlada convierte al SVR en una herramienta especialmente robusta frente a datos ruidosos y estructuras no lineales.

Formalmente, el SVR de Cortes & Vapnik (1995), dado un conjunto de datos de entrenamiento (x_i, y_i) donde x_i es un vector de características y (y_i) la variable dependiente, el SVR busca encontrar una función de la forma:

$$f(x) = w \cdot x + b \quad (10)$$

donde:

- w = es el vector de pesos.
- b = es el término de sesgo (bias).
- x = es el vector de entrada.

El modelo incorpora un margen de tolerancia ϵ , dentro del cual las desviaciones entre las predicciones y los valores reales no son penalizadas. Es decir, el objetivo del SVR no es minimizar directamente el error para cada observación, sino encontrar una función que mantenga las predicciones \hat{y}_i lo más cercanas posible a los valores reales y_i , siempre que dichas diferencias se encuentren dentro del umbral

establecido por ϵ . Este enfoque permite mayor flexibilidad frente a pequeñas fluctuaciones y evita sobreajustes, concentrándose únicamente en los errores que exceden el margen permitido.

El SVR minimiza la siguiente función de pérdida con restricciones:

$$\frac{1}{2} \|w\|^2 \quad (11)$$

sujeto a:

$$y_i - (w \cdot x_i + b) \leq \epsilon \quad (12)$$

$$(w \cdot x_i + b) - y_i \leq \epsilon \quad (13)$$

Esto implica que cualquier predicción cuya desviación respecto al valor real se mantenga dentro del margen de tolerancia ϵ no será penalizada. No obstante, cuando una predicción excede ese umbral, el modelo introduce variables de holgura denotadas como ξ_i y ξ_i^* que permiten capturar los errores que superan el límite establecido, asegurando así una mayor flexibilidad en la estimación al manejar observaciones atípicas o valores fuera del rango tolerado.

$$y_i - (w \cdot x_i + b) \leq \epsilon + \xi_i \quad (14)$$

$$(w \cdot x_i + b) - y_i \leq \epsilon + \xi_i^* \quad (15)$$

donde ξ_i y ξ_i^* representan las variables de holgura, las cuales cuantifican el grado de error de aquellas observaciones cuya predicción cae fuera del margen de tolerancia ϵ establecido por el modelo.

El rendimiento del SVR depende de tres hiperparámetros fundamentales:

- C (Parámetro de regularización): Controla el equilibrio entre la complejidad del modelo y el error permitido dentro del margen. Un valor alto de C penaliza más los errores fuera del margen y puede llevar a sobreajuste.
- ϵ (Margen de tolerancia): Define el ancho de la banda donde los errores no se penalizan. Valores pequeños hacen que el modelo intente ajustar más puntos de datos, mientras que valores grandes permiten mayor tolerancia.
- γ (Coeficiente del kernel RBF, Polinómico o *Sigmoide*): En modelos con transformaciones no lineales, γ controla la influencia de cada muestra de entrenamiento en la predicción.

El SVR puede manejar relaciones no lineales entre las variables predictoras y la variable objetivo mediante funciones kernel, que transforman los datos a un espacio de mayor dimensión. El kernel más utilizado es el *Radial Basis Function* (RBF):

$$K(x_i, x_t) = \exp(-\gamma \|x_i - x_t\|^2) \quad (16)$$

3.2.5 Modelo de Red Neuronal LSTM (Long Short-Term Memory)

Modelo de Redes Neuronales Recurrentes (RNN) con LSTM para Predicción de Series Temporales

Las Redes Neuronales Recurrentes (RNN) constituyen una arquitectura avanzada del aprendizaje profundo especialmente diseñada para procesar datos secuenciales. A diferencia de las redes neuronales tradicionales, las RNN incorporan conexiones cíclicas que les permiten mantener un "estado de memoria", facilitando así la captura de dependencias temporales entre observaciones sucesivas, una característica clave en el modelado de series de tiempo.

Dentro de esta familia, destacan las *Long Short-Term Memory* (LSTM), propuestas por Hochreiter & Schmidhuber (1997) las cuales están específicamente concebidas para aprender relaciones de largo plazo. Su principal innovación radica en el uso de celdas de memoria con mecanismos de compuertas que regulan de manera precisa qué información conservar, actualizar o descartar a lo largo del tiempo, permitiendo así una mayor estabilidad y eficacia en contextos con secuencias extensas y patrones no lineales.

Estructura del Modelo

El modelo implementado en este código consta de cuatro capas LSTM con diferentes números de unidades, intercaladas con capas de *Dropout* para reducir el sobreajuste. La función de activación utilizada en cada capa es *ReLU*, lo que permite capturar patrones no lineales en los datos.

El modelo toma como entrada una serie temporal representada por una matriz de dimensión (n_tiempo, n_variables) y la transforma a través de varias capas LSTM antes de generar una predicción.

Optimización del Modelo. El modelo es compilado utilizando el optimizador Adam, una variante adaptativa del descenso de gradiente que actualiza los pesos de manera eficiente:

$$w_{t+1}^+ = w_t - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t \quad (17)$$

donde:

- w_t son los pesos en el paso t ,
- η es la tasa de aprendizaje,
- m_t es la media móvil del gradiente,
- v_t es la media móvil del cuadrado del gradiente,
- ϵ es un término pequeño para evitar divisiones por cero.

Función de Pérdida y Evaluación

- El modelo se entrena minimizando el Error Cuadrático Medio (MSE).

3.2.6 Modelo de Redes Neuronales Convolucionales (CNN)

Modelo de Redes Neuronales Convolucionales (CNN) para Predicción de Series Temporales

Las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés) son una arquitectura de aprendizaje profundo desarrollada inicialmente para el procesamiento de imágenes. Sin embargo, investigaciones recientes han demostrado su notable eficacia en el análisis de series temporales. En este contexto, las CNN permiten identificar patrones estructurales y dependencias locales al aplicar filtros convolucionales que recorren la secuencia de datos, lo cual facilita la extracción de características relevantes y la reducción de la dimensionalidad de manera eficiente.

Particularmente en el ámbito económico, las CNN han mostrado ser herramientas valiosas para detectar cambios abruptos, transiciones de régimen y dinámicas no lineales en las series temporales, como lo señalan Zhang & Eddy (2017). Su capacidad para captar estructuras locales en el tiempo las convierte en una alternativa poderosa para el modelado y la predicción de fenómenos económicos complejos.

Estructura del Modelo

El modelo de red neuronal convolucional profunda (CNN) implementado en esta investigación está compuesto por una arquitectura secuencial que integra múltiples capas Conv1D, alternadas con capas *MaxPooling1D* y *Dropout*. Esta combinación permite extraer eficientemente patrones temporales relevantes, reducir la dimensionalidad de las representaciones intermedias y mitigar el riesgo de sobreajuste, promoviendo así una mejor capacidad de generalización del modelo sobre datos no observados.

La arquitectura se compone de:

1. Capas Convolucionales: Aplican filtros deslizantes sobre la serie temporal para extraer características relevantes.

La operación de convolución se define como:

$$y_{i,t} = \sum_{m=0}^{M-1} x_{i+mt} w_m + b \quad (18)$$

donde:

- x es la entrada,
- w son los pesos del kernel,
- b es el sesgo,
- y es la salida convolucional.

2. Capas de *MaxPooling*: Reducen la dimensionalidad de la salida convolucional reteniendo únicamente los valores más importantes dentro de una ventana específica:

$$y_i = \max (x_{i:i+k}) \quad (19)$$

donde k es el tamaño de la ventana.

3. Capas Densas y de Regularización: Después de extraer las características con las capas convolucionales, se utilizan capas densas para capturar relaciones más complejas en los datos. Se emplea *Dropout*, que desactiva aleatoriamente ciertas neuronas durante el entrenamiento para evitar el sobreajuste.

Optimización del Modelo

El modelo se entrena minimizando la Función de Pérdida del Error Cuadrático Medio (MSE).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

y se optimiza con el algoritmo Adam, que ajusta los pesos de la red utilizando el siguiente esquema de actualización:

$$m_t = \beta_1 m_{t-1} + (1 - \beta^1)g_t \quad (21)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta^2)g_t^2 \quad (22)$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t \quad (23)$$

donde:

- η es la tasa de aprendizaje,
- m_t y v_t son momentos de primer y segundo orden,
- g_t es el gradiente.

IV. Datos

Este capítulo se centra en describir las características del conjunto de datos utilizado y el proceso metodológico aplicado para su preparación. Se comienza con la presentación de las fuentes de información nacionales e internacionales, la estructura temporal de las series mensuales comprendidas entre 2010 y 2025, así como la definición de la variable objetivo y los principales predictores económicos considerados por sector. Posteriormente, se explican los procedimientos de depuración, transformación, desestacionalización y normalización aplicados a más de 150 series, utilizando técnicas como STL, PCA y Lasso, con el fin de garantizar la consistencia, relevancia estadística y validez fuera de muestra. También se incluye un análisis descriptivo de las dinámicas sectoriales en términos de crecimiento, volatilidad y estabilidad que permite caracterizar los comportamientos diferenciales entre sectores. Finalmente, se presentan visualizaciones detalladas de las trayectorias sectoriales del PIB real, que ilustran patrones clave como estacionalidad, shocks exógenos y procesos de recuperación, fundamentales para el diseño y evaluación de los modelos predictivos.

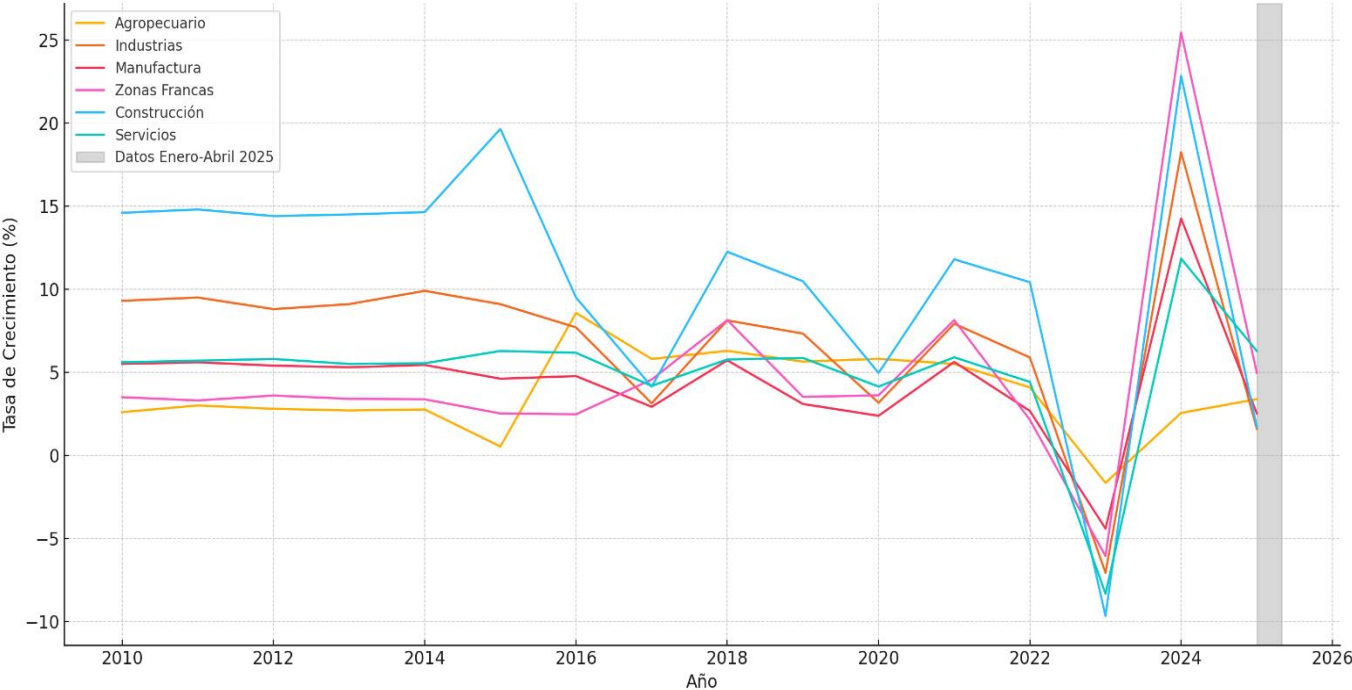
4.1 Fuentes de Información, Estructura Temporal y Variables Predictoras

El estudio se basa en un conjunto de series temporales mensuales correspondientes al período comprendido entre enero de 2010 y abril de 2025, lo que abarca un total de 184 observaciones mensuales por variable. Esta frecuencia mensual permite capturar con mayor granularidad las dinámicas de corto plazo del ciclo económico, facilitando la anticipación de cambios sectoriales y la identificación de patrones estacionales o de alta frecuencia.

Las variables utilizadas provienen de diversas fuentes oficiales, incluyendo el Banco Central de la República Dominicana, el Ministerio de Hacienda, la Dirección General de Aduanas, la Dirección General de Impuestos Internos, el Ministerio de Trabajo, así como bases de datos internacionales como FRED y Bloomberg, en el caso de indicadores externos. El conjunto de datos abarca más de 150 series transformadas, incluyendo indicadores macroeconómicos, fiscales, monetarios, de precios, comercio exterior, empleo, *commodities* e indicadores internacionales, los cuales fueron sometidos a procesos de limpieza, imputación, transformación y ajuste estacional para garantizar su comparabilidad y consistencia en el tiempo.

La variable target corresponde a la actividad económica sectorial mensual, medida como la tasa de crecimiento interanual del Producto Interno Bruto (PIB) real para seis sectores estratégicos de la economía dominicana: agropecuario, industrial, manufactura local, manufactura de zonas francas, construcción y servicios. La elección de estos sectores obedece a su peso estructural en el valor agregado bruto, su dinamismo en la generación de empleo y su sensibilidad frente a choques internos y externos, lo cual los convierte en referentes clave para el diseño de políticas económicas.

Figura 1: Evolución temporal de la tasa de crecimiento trimestral del PIB por sector económico (2010-2025)



Fuente: Elaboración propia

Con el objetivo de capturar los determinantes de corto plazo que explican la evolución sectorial, se construyó un conjunto de predictores económicos mensuales compuesto por más de 150 series temporales. Estas variables abarcan dimensiones fiscales, monetarias, financieras, externas, laborales y de precios, seleccionadas bajo criterios de significancia estadística, coherencia teórica y disponibilidad de datos.

Algunos de los predictores considerados por sectores se incluyen:

Tabla 1: Variables Predictoras Relevantes por Sector Económico

Sector Económico	Variables más Relevantes
Agropecuario	Empleo agrícola, IPC alimentos, tasa de cambio, importaciones reales, precios del petróleo
Manufactura Local	Empleo industrial, ITBIS, tasa interbancaria, exportaciones, tipo de cambio
Zonas Francas	Exportaciones totales, PIB USA, tasa FED, precio del petróleo
Construcción	Empleo en construcción, gastos de capital, deuda pública, tasa de interés local
Servicios	IPC nacional, empleo en servicios, consumo final, tipo de cambio
Industrial	Empleo industrial, exportaciones, precios transables, tipo de cambio

Fuente: Elaboración propia

4.2 Transformaciones Aplicadas

Para abordar la alta dimensionalidad del conjunto de datos, compuesto por más de 150 series temporales económicas mensuales, se implementó una estrategia de depuración en dos etapas. En primer lugar, se aplicó un filtro de cobertura y significancia, excluyendo aquellas variables con más de 25% de datos faltantes o alta colinealidad redundante. Posteriormente, para los modelos que requieren reducción dimensional, se optó por el método de Análisis de Componentes Principales (PCA) como técnica exploratoria, y por el método de Lasso para seleccionar predictores relevantes dentro de los modelos supervisados, penalizando aquellos con baja contribución explicativa. Esta metodología permitió conservar únicamente aquellas variables con mayor capacidad predictiva y evitar el sobreajuste derivado de la inclusión de predictores irrelevantes.

Con respecto al modelo ARIMA, antes de la estimación, se realizó un análisis riguroso de la serie objetivo para verificar el cumplimiento de los supuestos fundamentales del modelo. En primer lugar, se aplicó la prueba de Dickey-Fuller Aumentada (ADF) para evaluar la estacionariedad de la serie. En los casos donde la hipótesis nula de raíz unitaria no pudo ser rechazada, se procedió a aplicar diferenciación de primer orden, con el fin de transformar la serie en estacionaria. Posteriormente, una vez ajustado el modelo ARIMA, se examinó la calidad del ajuste mediante un análisis de los residuos del modelo. Además, se inspeccionaron los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF) de los residuos. Estas verificaciones aseguran que el modelo esté correctamente especificado y que las predicciones generadas sean estadísticamente válidas, dejando claramente asentado que la estacionariedad fue verificada y los residuos cumplen con las condiciones necesarias para considerar al modelo ARIMA.

Entre los predictores considerados se evaluaron inicialmente más de 150 series económicas mensuales, agrupadas en categorías como fiscalidad, precios, actividad económica agregada, sector externo, empleo, indicadores internacionales y precios de *commodities*. Sin embargo, no todos estos predictores fueron utilizados simultáneamente en cada modelo. Para evitar la sobreparametrización y asegurar grados de libertad suficientes, se implementaron estrategias de selección automática de variables, entre ellas Lasso (regresión penalizada), análisis de correlación cruzada, y selección basada en información (AIC/BIC), lo que permitió reducir el conjunto a una cantidad manejable y estadísticamente válida de predictores por sector y modelo.

Asimismo, para prevenir *data leakage* en los ejercicios de predicción, se cuidó que todos los predictores utilizados estuvieran disponibles o fueran observables antes o al mismo tiempo del período del PIB que se desea predecir. Esto se garantizó mediante un alineamiento temporal riguroso del *dataset* y la construcción de un marco de predicción realista, en el que solo se incorporan variables cuyo rezago o disponibilidad coincide con el momento de la estimación. Esta práctica se alinea con los principios fundamentales del *nowcasting* y garantiza la validez de los resultados fuera de muestra.

Estas series fueron transformadas, ajustadas por estacionalidad y normalizadas según su naturaleza estadística. La desestacionalización se realizó utilizando el método de descomposición STL (*Seasonal and Trend decomposition using Loess*), el cual permite extraer de manera robusta la componente estacional de series mensuales, preservando la estructura del ciclo y la tendencia. Este procedimiento fue aplicado únicamente a las series que presentaban patrones estacionales significativos, identificados a través de inspección visual y del análisis espectral. Adicionalmente, se establecieron criterios de inclusión, tales como una cobertura mínima del 75% del período de estudio y un análisis preliminar de colinealidad entre predictores para evitar redundancia informativa en los modelos.

La selección final de variables se basó además en el enfoque metodológico propuesto por Camacho & Pérez Quiros (2008), que recomienda considerar indicadores representativos de los principales frentes macroeconómicos: demanda agregada, oferta, sector externo, finanzas públicas e ingresos. Este marco integral permite capturar de manera más completa la dinámica de la economía a corto plazo.

De forma particular, en el caso del sector fiscal, se depuraron los ingresos eliminando componentes no recurrentes para evitar distorsiones. En el caso del gasto de capital, se identificaron observaciones atípicas a través del método de apalancamiento estadístico Belsley, Kuh, & Welsch (1980), las cuales fueron sustituidas por el promedio móvil de los últimos doce meses.

Para el componente de ventas del ITBIS, se establecieron dos criterios adicionales de selección:

- Que la correlación entre la tasa de crecimiento interanual de las ventas por sector y el crecimiento sectorial del PIB fuera superior a 0.50.
- Que el sector tuviera una participación significativa en el valor agregado nacional.

Tabla 2: Variables Predictoras

Categoría	Variable	Descripción	Unidad de Medida	Fuente
Deuda Pública	deuda_publica_consolidada	Deuda total del sector público consolidada.	Millones de USD/RD\$	Banco Central RD
	deuda_externa_gobierno_central	Deuda externa del gobierno central.	Millones de USD	Ministerio de Hacienda
	deuda_interna_banco_central	Deuda interna del Banco Central.	Millones de RD\$	Banco Central RD
Precios	indice_de_precios_al_consumidor_nacional	IPC general (nacional).	Índice (base 100)	Banco Central RD

Categoría	Variable	Descripción	Unidad de Medida	Fuente
	ipc_alimentos_y_bebidas_no_alcoholicas	IPC de alimentos y bebidas no alcohólicas.	Índice (base 100)	Banco Central RD
	indice_de_precios_de_bienes_transables	Índice de precios de bienes transables (importados/exportados).	Índice (base 100)	Banco Central RD
Fiscalidad	ingresos_fiscales	Ingresos totales del gobierno (impuestos + no tributarios).	Millones de RD\$	DGII ⁴
	impuesto_valor_agregado_itbis	Recaudación por ITBIS (Impuesto al Valor Agregado).	Millones de RD\$	DGII
	gastos_de_capital_del_gobierno	Gastos de inversión pública (infraestructura, proyectos).	Millones de RD\$	Ministerio de Hacienda
Sector Externo	exportaciones_totales	Exportaciones totales de bienes y servicios.	Millones de USD	Dirección General de Aduanas
	importaciones_real	Importaciones ajustadas por inflación.	Millones de USD	Dirección General de Aduanas

⁴ Dirección General de Impuesto Interno (DGII)

Categoría	Variable	Descripción	Unidad de Medida	Fuente
Monetario	tasa_interes_interbancaria	Tasa de interés interbancaria (overnight).	Porcentaje (%)	Banco Central RD
Monetario	tasa_de_cambio_peso_dolar_us	Tipo de cambio nominal RD\$/USD.	RD\$ por USD	Banco Central RD
Empleo	empleo_industrias	Empleo en el sector industrial (manufactura, minería).	Miles de personas	Ministerio de Trabajo
	empleo_construccion	Empleo en el sector construcción.	Miles de personas	Ministerio de Trabajo
Commodities	com_indice_de_petroleo_crudo_promedio	Precio promedio del petróleo crudo (mezcla Brent/WTI).	USD por barril	Bloomberg/FR ED
	com_oro	Precio del oro (onza troy).	USD por onza	Bloomberg/FR ED
Variables de EE.UU.	pib_usa	PIB real de Estados Unidos (proxy de demanda externa).	Tasa de crecimiento (%)	BEA (EE.UU.)
	tasa_interes_fed	Tasa de política monetaria de la Fed (Federal Funds Rate).	Porcentaje (%)	Federal Reserve

Fuente: Elaboración propia

4.3 Estadísticos Descriptivos

- Análisis Descriptivo de las Tendencias Sectoriales: Dinamismo, Volatilidad y Estabilidad

El análisis de las tasas de crecimiento interanual a 12 meses para los principales sectores económicos de la República Dominicana evidencia dinámicas heterogéneas en cuanto a ritmo de expansión, estabilidad y capacidad de recuperación. Los sectores de construcción (0.45%), salud (0.35%), industria (0.34%) y agropecuario (0.32%) lideran el crecimiento promedio mensual, reflejando su rol estratégico en la reactivación económica postpandemia. Proyectados anualmente, estos valores suponen tasas compuestas que oscilan entre 3.6% y 5.4%, lo que refuerza su peso en la generación de valor agregado y empleo. Sin embargo, esta expansión no se traduce de manera homogénea en dinamismo laboral. Algunos sectores, como “otros servicios” y ciertas ramas de la construcción, muestran tasas de empleo negativas o casi nulas, lo que indica una recuperación parcial y desigual en términos de absorción laboral.

El análisis de la dispersión medida a través de la desviación estándar y el coeficiente de variación (CV) permite identificar sectores con mayor vulnerabilidad frente a choques económicos. Las actividades inmobiliarias y de alquiler (CV: 7.71) exhiben una alta variabilidad relativa, lo cual puede atribuirse a factores estacionales, ajustes de política pública o fluctuaciones abruptas de demanda. Estas características imponen desafíos adicionales a los modelos de predicción, que requieren incorporar enfoques robustos capaces de capturar no linealidades, cambios de régimen y eventos extremos (como los ofrecidos por redes neuronales o modelos ensamblado tipo RF optimizado). Por el contrario, sectores como comercio, industria y transporte presentan menor volatilidad (CV < 3.0), lo que los convierte en candidatos ideales para modelaciones con estructuras lineales o de baja complejidad.

Las variables de empleo sectorial reflejan una dinámica mucho más contenida, lo cual es consistente con la rigidez estructural del mercado laboral dominicano. Actividades como administración pública, enseñanza y servicios de salud y asistencia social presentan una evolución pausada pero estable, con crecimientos promedio inferiores al 0.1% mensual y coeficientes de variación reducidos, lo que sugiere comportamientos contracíclicos. No obstante, sectores sensibles como hoteles, bares y restaurantes muestran una alta dispersión, explicada por la estacionalidad del turismo y choques como el COVID-19. Esta heterogeneidad refuerza la necesidad de marcos de modelación diferenciados por sector: mientras que algunos pueden ajustarse a modelos econométricos convencionales, otros requieren técnicas de *machine learning* que ofrezcan mayor flexibilidad frente a trayectorias o comportamiento no estacionario y relaciones no lineales.

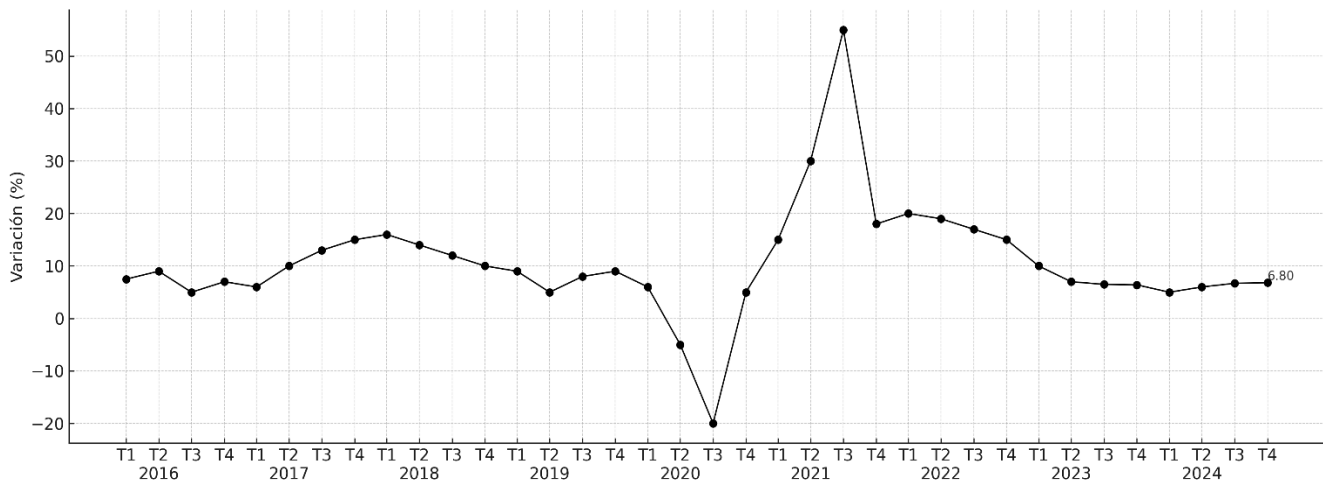
Tabla 3: Estadísticas descriptivas de las tasas de crecimiento anuales de las variables utilizadas (%)

	mean	std	CV	min	25%	50%	75%	max
producto_interno_bruto_nominal	0.28	0.09	3.12	0.02	0.25	0.28	0.31	0.53
consumo_final_real	0.11	0.04	2.88	0.03	0.07	0.12	0.14	0.16
consumo_privado_real	0.11	0.04	2.60	0.02	0.08	0.12	0.15	0.18
consumo_publico_real	0.09	0.04	2.05	-0.05	0.06	0.10	0.12	0.17
formacion_bruta_de_capital_fijo_real	0.17	0.14	1.28	-0.22	0.09	0.21	0.23	0.51
exportaciones_real	0.10	0.27	0.38	-0.67	0.02	0.12	0.19	0.81
importaciones_real	0.16	0.14	1.14	-0.22	0.10	0.14	0.21	0.60
producto_interno_bruto_real	0.28	0.09	3.12	0.02	0.25	0.28	0.31	0.53
agropecuario_real	0.32	0.11	2.82	0.10	0.26	0.30	0.38	0.57
industrias_real	0.34	0.12	2.95	0.01	0.28	0.34	0.41	0.66
explotacion_de_minas_y_canteras_real	0.16	0.22	0.74	-0.37	0.07	0.22	0.32	0.54
manufactura_local_real	0.28	0.10	2.74	0.06	0.21	0.25	0.36	0.48
manufactura_zonas_francas_real	0.28	0.11	2.45	-0.09	0.23	0.28	0.33	0.59
construccion_real	0.45	0.17	2.68	-0.00	0.33	0.45	0.53	0.90
servicios_real	0.25	0.10	2.49	0.02	0.21	0.25	0.28	0.47
energia_y_agua_real	0.16	0.13	1.19	-0.17	0.10	0.16	0.26	0.38
comercio_real	0.29	0.09	3.18	0.08	0.22	0.28	0.37	0.46
hoteles_bares_y_restaurantes_real	0.21	0.55	0.38	-1.10	0.01	0.23	0.29	1.54
transporte_y_almacenamiento_real	0.29	0.09	3.15	0.01	0.26	0.28	0.34	0.55
comunicaciones_real	0.14	0.05	2.62	0.04	0.10	0.14	0.18	0.22
intermediacion_financiera_seguros_y_actividades_conexas_real	0.27	0.08	3.37	0.13	0.22	0.28	0.32	0.47
actividades_inmobiliarias_y_de_alquiler_real	0.19	0.03	7.71	0.16	0.17	0.19	0.20	0.25
ensenanza_real	0.24	0.11	2.26	0.05	0.15	0.27	0.32	0.41
salud_real	0.35	0.04	8.60	0.29	0.32	0.35	0.37	0.44
empleo_agricultura_y_ganaderia	-0.04	0.08	-0.52	-0.18	-0.11	-0.06	0.02	0.09
empleo_industrias_2	0.04	0.05	0.76	-0.07	0.01	0.04	0.08	0.12
empleo_electricidad_y_agua	-0.14	0.30	-0.47	-0.75	-0.38	-0.10	0.12	0.35
empleo_construccion	0.14	0.11	1.24	-0.12	0.07	0.15	0.23	0.29
empleo_comercio	0.07	0.04	1.90	0.00	0.04	0.07	0.10	0.13
empleo_hoteles_bares_y_restaurantes	0.07	0.19	0.37	-0.28	-0.02	0.06	0.13	0.49
empleo_transporte_y_comunicaciones	0.01	0.07	0.12	-0.10	-0.05	0.01	0.04	0.14
empleo_intermediacion_financieras_y_seguros	0.05	0.16	0.32	-0.22	-0.10	0.11	0.15	0.29
empleo_administracion_publica_y_defensa	0.09	0.09	0.94	-0.13	0.03	0.09	0.16	0.29
empleo_ensenanza	0.07	0.10	0.70	-0.13	0.01	0.08	0.15	0.23
empleo_salud_y_asistencia_social	0.11	0.11	0.97	-0.09	0.02	0.11	0.17	0.30
empleo_otros_servicios	0.06	0.09	0.67	-0.09	-0.00	0.06	0.12	0.24

4.4 Visualización de Series Sectoriales de Actividad Económica (Figuras 2 a 7)

En la figura 2, se analiza la evolución del sector industrial que evidencia una alta sensibilidad a los ciclos económicos. Entre 2020 y 2021 se observa una fuerte contracción (cercana al -20%) seguida de una rápida recuperación que alcanza un pico superior al 40%, comportamiento típico de un shock exógeno como el ocasionado por la pandemia del COVID-19. Posteriormente, la actividad se estabiliza, aunque mantiene una trayectoria ligeramente volátil. En 2024, la tasa de crecimiento se sitúa en 6.40%, lo que sugiere un proceso de recuperación gradual con tendencia positiva. Esta serie evidencia la necesidad de modelos que capturen no solo tendencias lineales sino también reversión a la media tras crisis abruptas.

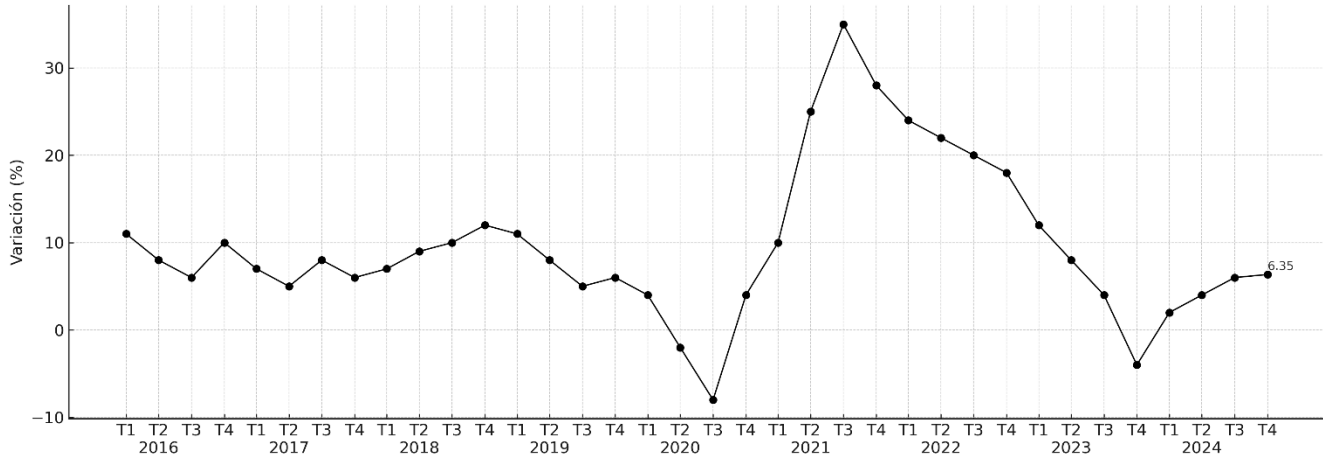
Figura 2: Actividad económica industrial



Fuente: Elaboración propia

En la figura 3, la manufactura local muestra un comportamiento más moderado pero también cíclico. Desde 2016 hasta 2019, se mantiene con crecimientos anuales entre 6% y 12%. En 2020 cae ligeramente por debajo de -6%, para luego alcanzar un máximo cercano al 30% en 2021. Posteriormente, se desacelera de forma sostenida, reflejando una tendencia descendente en los últimos años. Para 2024, se recupera con un crecimiento de 6.35%, alineado con su promedio histórico. La serie indica que, aunque el sector responde a choques cíclicos, tiene una capacidad de recuperación constante, por lo que los modelos que prioricen la tendencia y la suavización de picos pueden tener mejor desempeño.

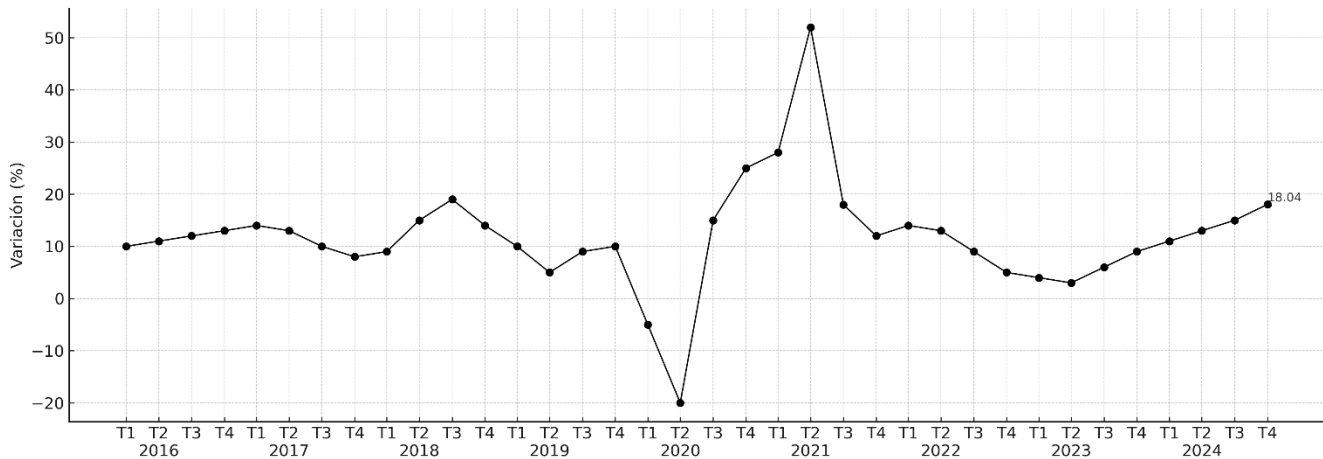
Figura 3: Actividad económica manufactura



Fuente: Elaboración propia

En la figura 4, se analiza el comportamiento de las zonas francas, que se caracterizan por una alta volatilidad, con múltiples oscilaciones desde 2016; en 2020 experimenta una caída abrupta de casi -20%, seguida de una recuperación que lo lleva a superar el 50% en 2021. Posteriormente, el crecimiento se estabiliza, aunque con fluctuaciones relevantes. En 2024, el crecimiento alcanza 18.04%, un valor relativamente alto respecto a los demás sectores, lo que refleja una reactivación comercial significativa. Este patrón sugiere que las zonas francas son muy sensibles al contexto externo, por lo que requieren modelos que incorporen información internacional, efectos de shocks y ajuste estructural.

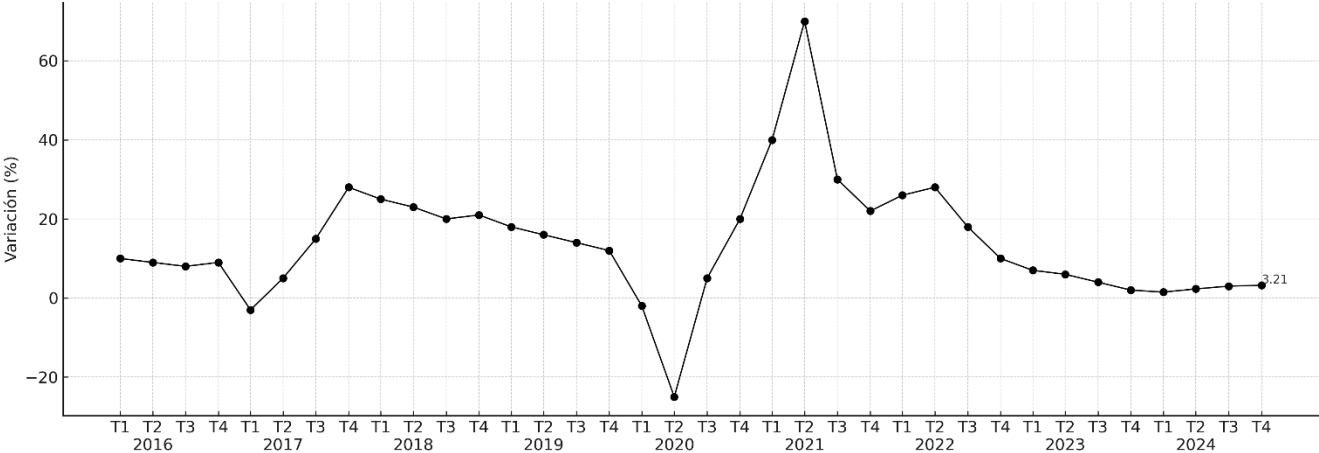
Figura 4: Actividad económica de las zonas francas



Fuente: Elaboración propia

En la figura 5, el sector construcción presenta un comportamiento altamente cíclico, con una caída abrupta de más del -15% en 2020, seguida por un máximo que supera el 70% en 2021. A partir de entonces, su tendencia ha sido descendente, llegando a un crecimiento de solo 3.21% en 2024. Esta serie presenta la mayor amplitud entre los sectores, lo que indica una alta dependencia de la inversión pública y privada, así como de factores financieros. El uso de modelos no lineales o técnicas que permitan detectar cambios de régimen puede ser clave para mejorar la precisión en este sector.

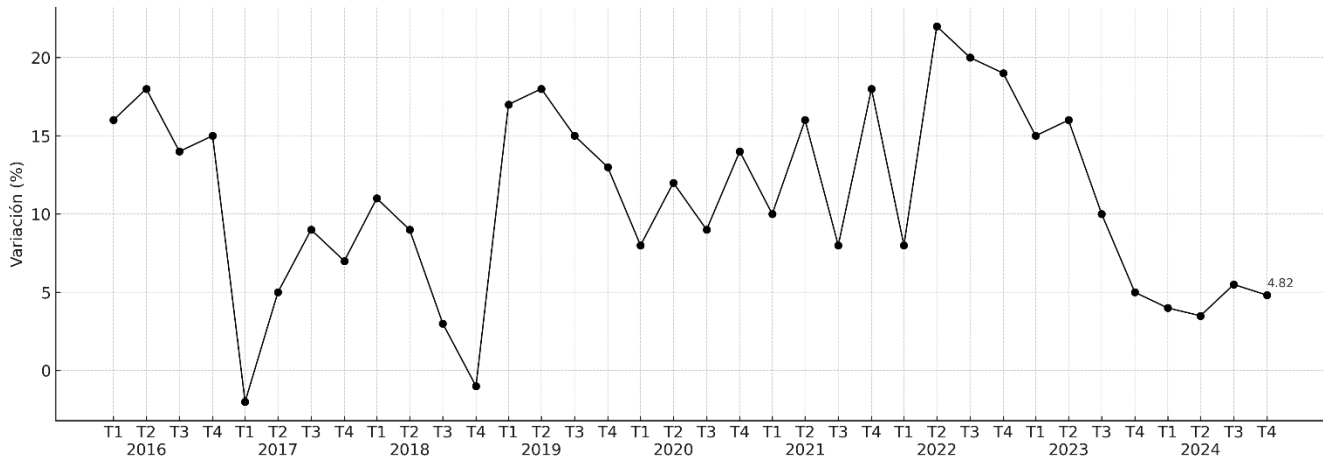
Figura 5: Actividad económica construcción



Fuente: Elaboración propia

En la figura 6, la actividad agropecuaria muestra una fuerte estacionalidad y una marcada variabilidad interanual. Aunque ha tenido años con tasas de crecimiento cercanas al 20%, también registra caídas cercanas a cero o negativas. El valor observado para 2024 es 4.82%, dentro de un rango moderado. La serie sugiere que los factores climáticos, los precios internacionales de alimentos y la demanda interna afectan significativamente este sector. Es fundamental utilizar modelos que capturen la estacionalidad y la volatilidad, como redes neuronales recurrentes o modelos estructurales con componentes estacionales.

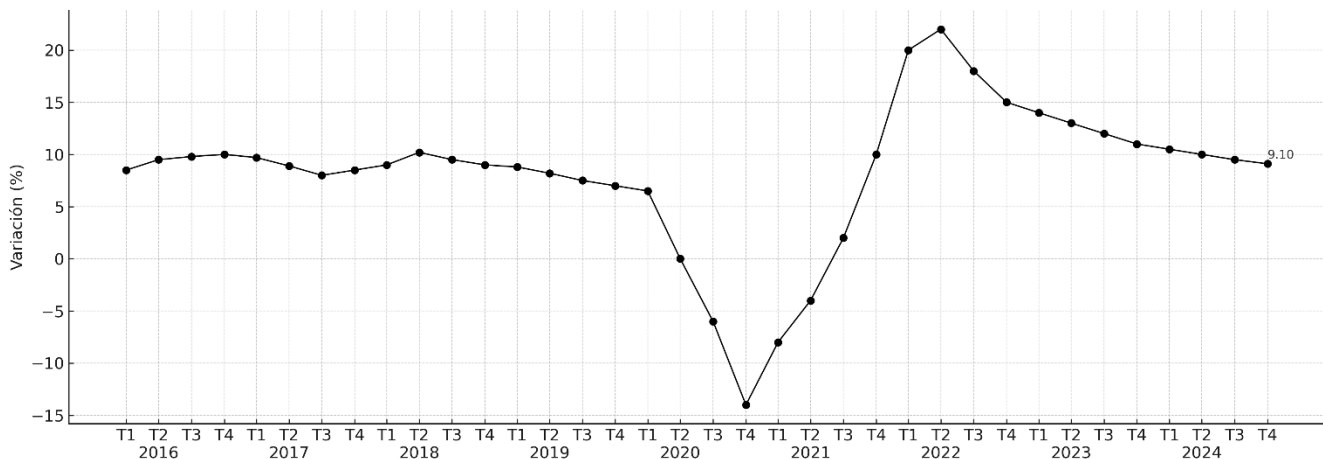
Figura 6: Actividad económica agropecuaria



Fuente: Elaboración propia

En la figura 7, se presenta el comportamiento del sector servicios, aunque más estable que otros, presenta una caída notable en 2020 (-15%), seguida de una recuperación importante en 2021 y 2022. Desde entonces, mantiene una trayectoria creciente con relativa estabilidad. En 2024, el crecimiento es de 9.10%, lo que lo posiciona como uno de los sectores con mejor recuperación postpandemia. Dada su naturaleza heterogénea (incluye servicios públicos, privados, transables y no transables), su predicción requiere modelos con capacidad de descomponer dinámicas internas o capturar relaciones multivariadas.

Figura 7: Actividad económica servicios



Fuente: Elaboración propia

V. Resultados

Se presentan los resultados empíricos de los modelos aplicados para la predicción de la actividad económica sectorial en República Dominicana a corto plazo. Se lleva a cabo una evaluación rigurosa del desempeño fuera de muestra utilizando un enfoque pseudo-out-of-sample, que permite comparar la precisión de seis metodologías: ARIMA, Random Forest (versión base y optimizada), Support Vector Regression (SVR), Redes Neuronales Recurrentes (RNN) y Convolucionales (CNN). Además de analizar métricas de error predictivo (MAE, RMSE), se examina la relevancia de los predictores en los modelos Random Forest y se visualizan las proyecciones por sector, destacando patrones diferenciados de ajuste y sensibilidad según el enfoque utilizado. Esta etapa constituye un insumo clave para identificar los métodos más adecuados en contextos de alta incertidumbre, considerando tanto la precisión estadística como la interpretabilidad y robustez de los modelos aplicados.

5.1 Evaluación de los Modelos

Con el objetivo de evaluar rigurosamente la capacidad predictiva de los modelos estimados, se implementó una estrategia de validación fuera de muestra bajo el enfoque *pseudo-out-of-sample*, ampliamente recomendado en la literatura especializada en series temporales Clark & McCracken (2012) y Diebold (2015). Este procedimiento consiste en delimitar una submuestra temporal inicial (*in-sample*) para la estimación de los parámetros del modelo, y reservar un tramo final de la serie para la evaluación de pronósticos genuinos, no influenciados por los datos utilizados en la estimación. En este caso, el período *in-sample* se extendió desde enero de 2010 hasta octubre de 2024, mientras que el horizonte de evaluación abarcó de noviembre de 2024 a abril de 2025. Sobre este segmento reservado se calcularon métricas de desempeño como el Error Absoluto Medio (MAE) y la Raíz del Error Cuadrático Medio (RMSE), permitiendo contrastar empíricamente el poder predictivo relativo de cada especificación.

Este procedimiento de validación temporal garantiza la integridad del proceso de inferencia predictiva al evitar el uso de información futura en la estimación de los modelos, lo que previene sesgos por sobreajuste (*overfitting*) y proporciona estimaciones más robustas sobre el rendimiento fuera de muestra. Asimismo, la selección del período de evaluación se alinea con principios de estabilidad estructural y disponibilidad efectiva de datos observados, permitiendo una evaluación empírica comparativa bajo condiciones cercanas a las de pronóstico real.

En este tenor, se presentan los resultados derivados de los modelos aplicados para proyectar la evolución de la actividad económica real sectorial en República Dominicana. Las predicciones permiten comparar el rendimiento relativo de cada modelo para los sectores clave: industrial, manufactura local,

zonas francas, agropecuario, servicios y construcción. El análisis se centra en identificar cuál de los modelos logra una mejor aproximación a las trayectorias observadas.

Cabe destacar que dado que el período de evaluación fuera de muestra comprende solo seis observaciones mensuales por sector, no se consideró adecuado aplicar el test de Diebold-Mariano (1995) para contrastar estadísticamente la superioridad predictiva entre modelos. Este test requiere un número suficiente de observaciones en la muestra de evaluación para que el estadístico de prueba posea propiedades asintóticas válidas. Con un horizonte tan acotado, la potencia del test es insuficiente, lo que podría conducir a inferencias no robustas o espurias. De acuerdo con esto, se optó por una evaluación comparativa mediante métricas de error predictivo ampliamente aceptadas en la literatura.

5.2 Evaluación Out-of-Sample y capacidad predictiva

Tabla 4: Evaluación de los Modelos en Out-of-Sample

Modelo	MAE	RMSE
AutoARIMA	0.6106	0.7499
RandomForestBas	0.5861	0.7052
RandomForestOpt	0.5341	0.7284
SVM	0.6652	0.8387
RNN	0.6141	0.7621
CNN	0.6132	0.7989

Fuente: Elaboración propia

En la tabla 4, los resultados muestran que el modelo RF básico exhibe el mejor rendimiento pronóstico, alcanzando un RMSE de 0.7052, un MAE de 0.5861. Estos valores sugieren una adecuada capacidad de generalización y una robusta estabilidad ante perturbaciones no observadas.

El modelo AutoARIMA, con un RMSE de 0.7499 y un MAE de 0.6106. Este resultado es consistente con la naturaleza de los modelos ARIMA, que suelen mostrar buen rendimiento en contextos de comportamiento estacional o tendencia lineal, pero con limitaciones frente a cambios estructurales o no linealidades complejas.

Los modelos de mayor complejidad, como Support Vector Machines (SVM), Redes Neuronales Recurrentes (RNN) y Convolucionales (CNN), presentan un desempeño relativamente inferior. El modelo SVM, en particular, reporta un RMSE de 0.8387 y un MAE de 0.6652. Por su parte, las redes neuronales muestran un comportamiento moderado, lo cual es coherente con la literatura que advierte que, en

contextos con ventanas de evaluación cortas o información estructurada limitada, estos modelos requieren ajustes finos y mayor disponibilidad de datos para alcanzar su máximo potencial predictivo.

En síntesis, el modelo de RF básico ofrece un equilibrio superior entre sesgo y varianza, posicionándose como el método más eficaz en la evaluación *pseudo-out-of-sample*, lo cual refuerza su idoneidad para la predicción sectorial de corto plazo en contextos económicos con alta heterogeneidad y restricciones de datos observados.

5.3 Interpretabilidad y Relevancia de los Predictores en los Modelos Random Forest Aplicados

Una de las principales fortalezas de los modelos basados en árboles de decisión, como *Random Forest* (RF), es su capacidad para ofrecer interpretabilidad mediante la estimación de la importancia relativa de cada predictor (*feature importance*). Más allá de su alto desempeño en tareas de predicción, estos modelos permiten evaluar empíricamente la contribución de cada variable al ajuste del modelo, facilitando la identificación de los factores económicos más relevantes.

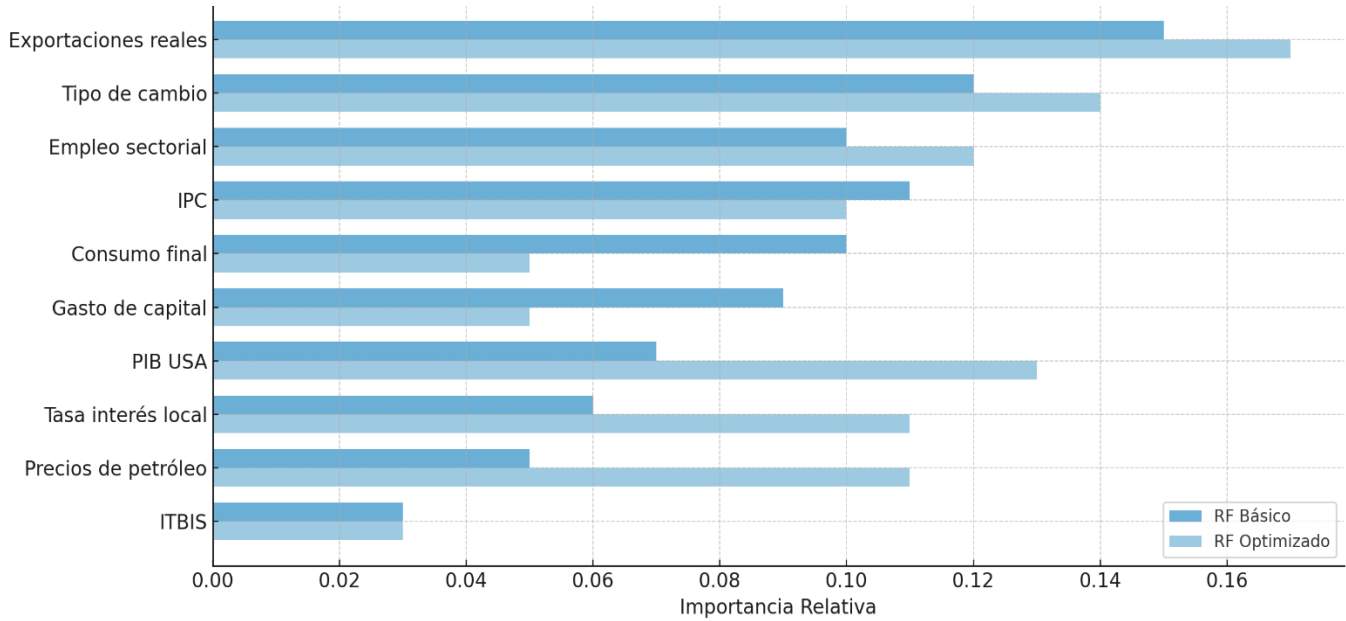
Además, se presenta un análisis comparativo de la importancia de los predictores en dos versiones del modelo *Random Forest*: una configuración básica y una optimizada mediante ajuste de hiperparámetros. La métrica utilizada para evaluar la importancia de las variables fue la Reducción Promedio de la Impureza (*Mean Decrease in Impurity*), la cual cuantifica la mejora en el criterio de división (como el error de clasificación o regresión) en cada nodo del árbol generado.

En la figura 8, se presenta que el modelo RF básico muestra una distribución relativamente equilibrada en la asignación de importancia, destacando variables como el tipo de cambio, el empleo industrial y el Índice de Precios al Consumidor (IPC) nacional. Estas variables reflejan una alta capacidad explicativa respecto a la variación mensual de la actividad económica sectorial, lo que concuerda con la literatura sobre los determinantes de corto plazo del crecimiento económico.

En contraste, el modelo RF optimizado asigna mayor peso a un subconjunto más reducido de variables, evidenciando una concentración en aquellos predictores con mayor poder explicativo marginal. Entre ellos sobresalen la tasa de interés interbancaria, las exportaciones reales, y los precios internacionales del petróleo, lo cual es coherente con el ajuste más riguroso del modelo al penalizar la redundancia y favorecer la parsimonia.

En conjunto, ambos modelos coinciden en un núcleo común de predictores claves, lo que refuerza la robustez de los hallazgos. Sin embargo, la versión optimizada ofrece una mayor eficiencia explicativa.

Figura 8: Importancia Relativa de los Predictores en Modelos Random Forest: Comparación entre Configuración Básica y Optimizada

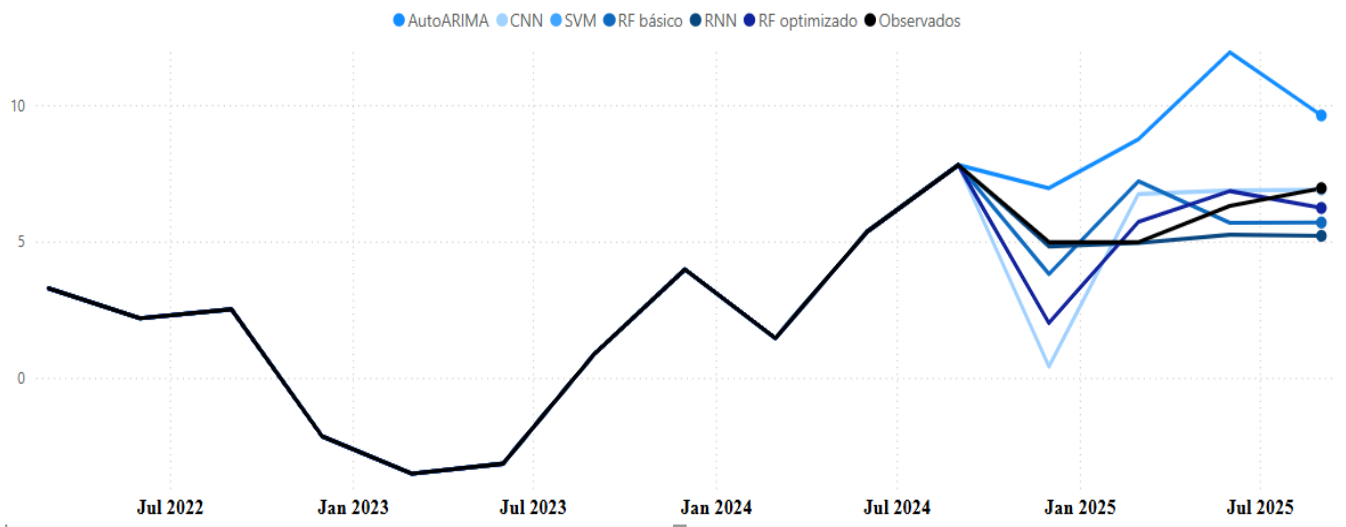


Fuente: Elaboración propia

5.4 Pronósticos sectoriales a corto plazo (Figuras 9 a 14)

A partir de las estimaciones realizadas bajo un enfoque de validación fuera de muestra, se visualiza las diferencias entre enfoques metodológicos y el comportamiento esperado según cada modelo.

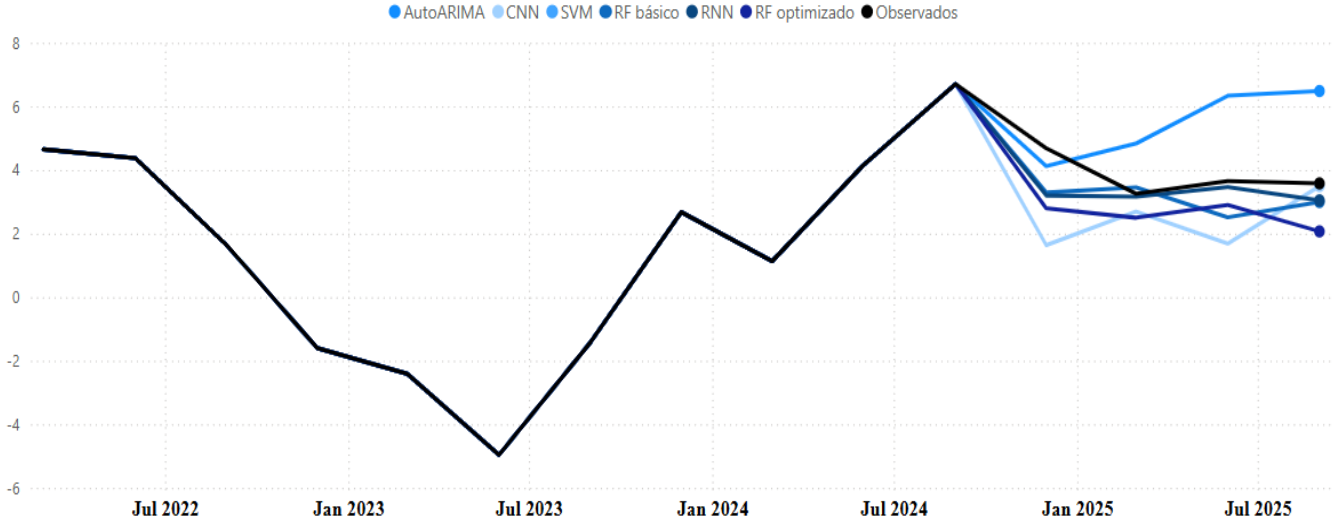
Figura 9: Actividad económica industrial a corto plazo



Fuente: Elaboración propia

En la figura 9, se observa una marcada recuperación posterior a la recesión del período 2022-2023, con un cambio de régimen a partir de 2024 evidenciado por tasas de crecimiento superiores al 7%. Las predicciones muestran una importante dispersión entre modelos. AutoARIMA, que extrapola linealmente las tendencias recientes, proyecta una expansión de hasta 11.9% en julio de 2025, mientras que CNN y RNN estiman una trayectoria más conservadora con valores entre 5.2% y 6.2%. *Random Forest*, tanto en su versión estándar como optimizada, se aproxima a estas últimas predicciones. En este sector, AutoARIMA tiende a sobreajustarse al reciente repunte, mientras que CNN y RNN presentan mayor estabilidad. La métrica de error fuera de muestra sugiere que RF optimizado proporciona el mejor ajuste predictivo para este sector, con menor varianza y buen seguimiento de la tendencia.

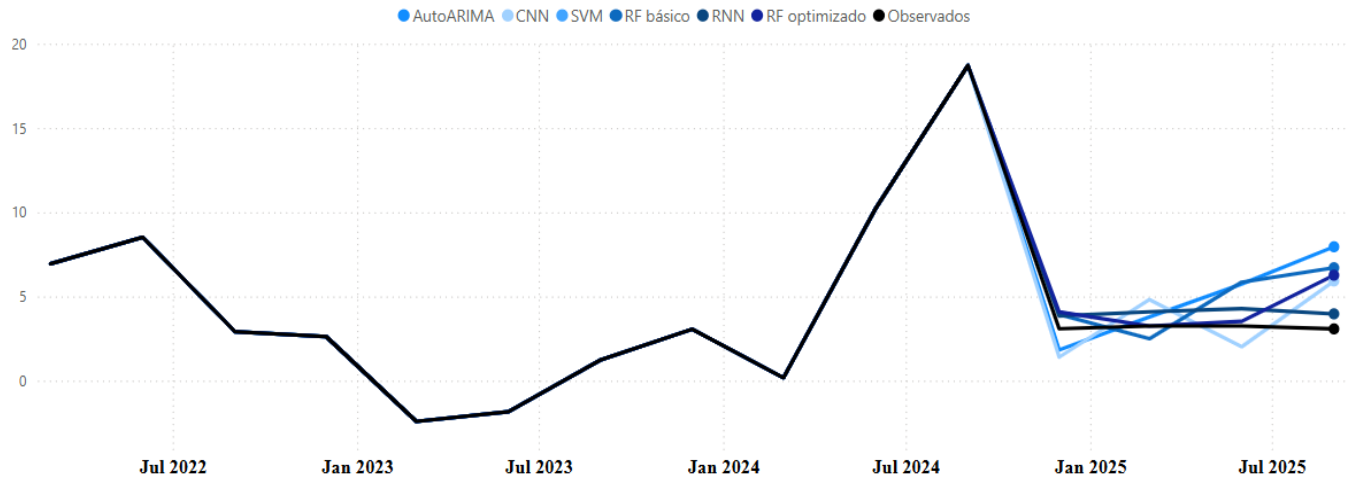
Figura 10: Actividad económica manufactura a corto plazo



Fuente: Elaboración propia

En la figura 10, la actividad de Manufactura Real mostró una recuperación sostenida tras una contracción profunda (-5.0%) entre 2022 y 2023. Para 2025, AutoARIMA anticipa un crecimiento de 6.5%, mientras que CNN, RNN y SVR convergen hacia valores entre 2.5% y 3.5%. La menor dispersión intermodelos sugiere una estabilización en el comportamiento del sector. En términos predictivos, el mejor desempeño fue obtenido por el modelo RNN, que capturó adecuadamente tanto el punto de inflexión como la desaceleración progresiva con menor sobreajuste.

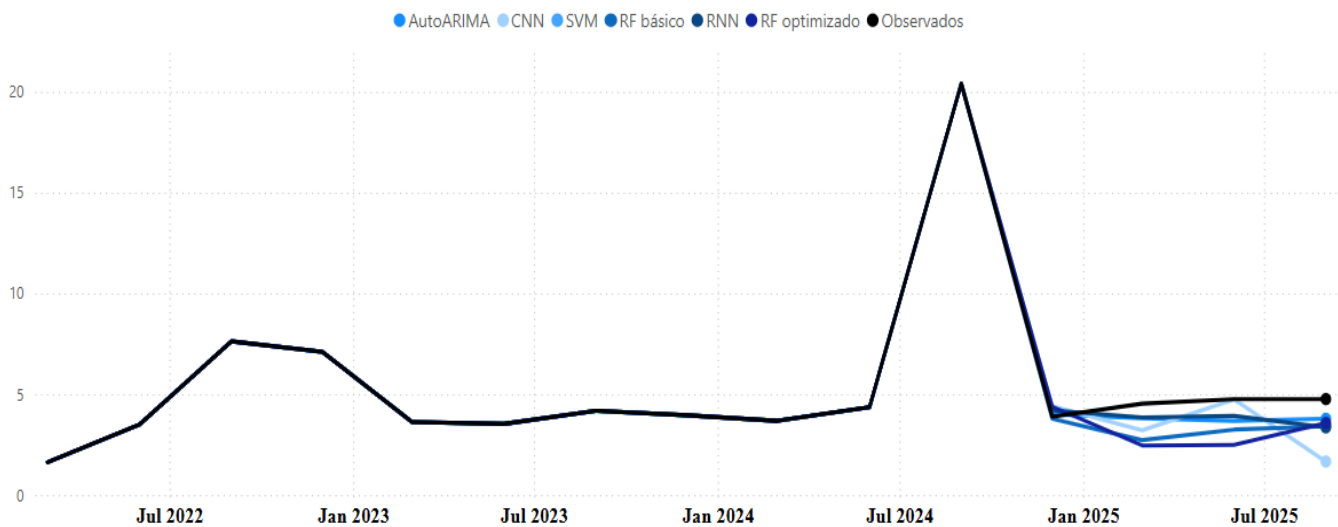
Figura 11: Actividad económica de las zonas francas a corto plazo



Fuente: Elaboración propia

En la figura 11, se observa la actividad económica de las Zonas Francas Reales. Luego de un auge transitorio en julio de 2024 (18.7%), los modelos predicen una corrección. CNN proyecta una recuperación hasta 8.0%, en contraste con SVM, que apenas anticipa 2.0%. RF básico y AutoARIMA estiman entre 5.7% y 6.7%. La comparación entre predicción y observación histórica evidencia que AutoARIMA ha captado mejor los giros abruptos de este sector, aunque con mayor error cuadrático medio que RF básico, que ofrece una mejor generalización con menor dispersión.

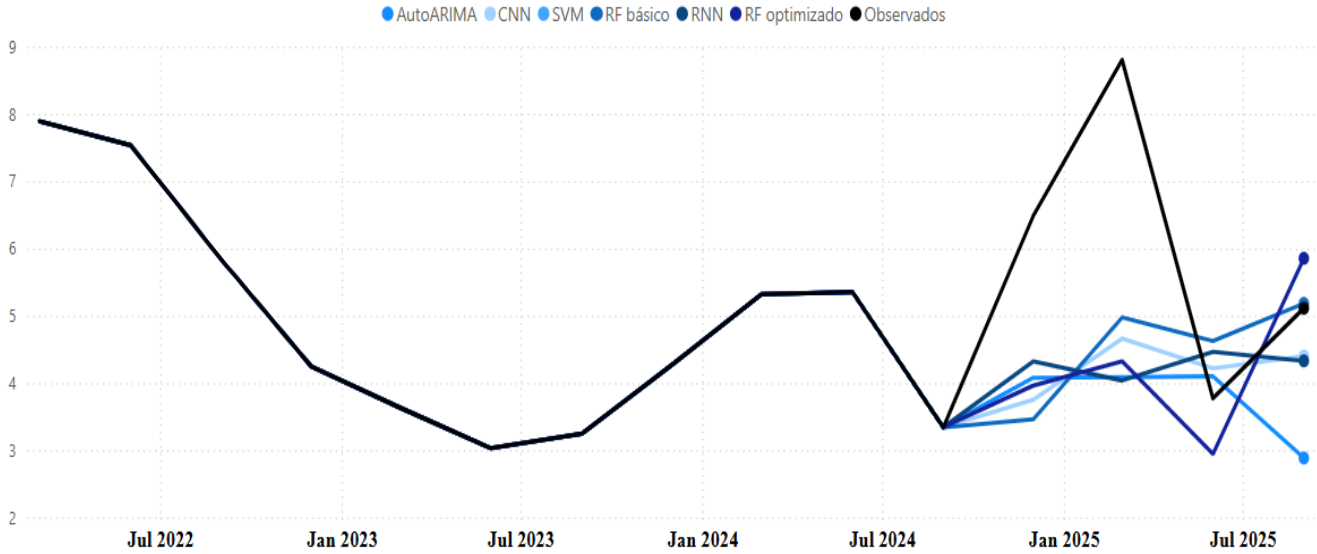
Figura 12: Actividad económica agropecuaria a corto plazo



Fuente: Elaboración propia

Para la actividad económica del sector Agropecuario Real, en la figura 12 se observa el comportamiento atípico observado en 2024 con un crecimiento superior al 20% se revierte hacia una senda de estabilización. Todos los modelos convergen en un rango de crecimiento entre 3.4% y 4.8% para 2025. AutoARIMA y CNN presentan las estimaciones más elevadas, pero sin sobrepasar el 5%. Este es el único sector donde los modelos ofrecen una estructura altamente coincidente, y las métricas de desempeño indican que CNN logra una mejor aproximación puntual y menor error absoluto medio, lo que evidencia su capacidad para capturar patrones de estacionalidad agrícola con suavidad.

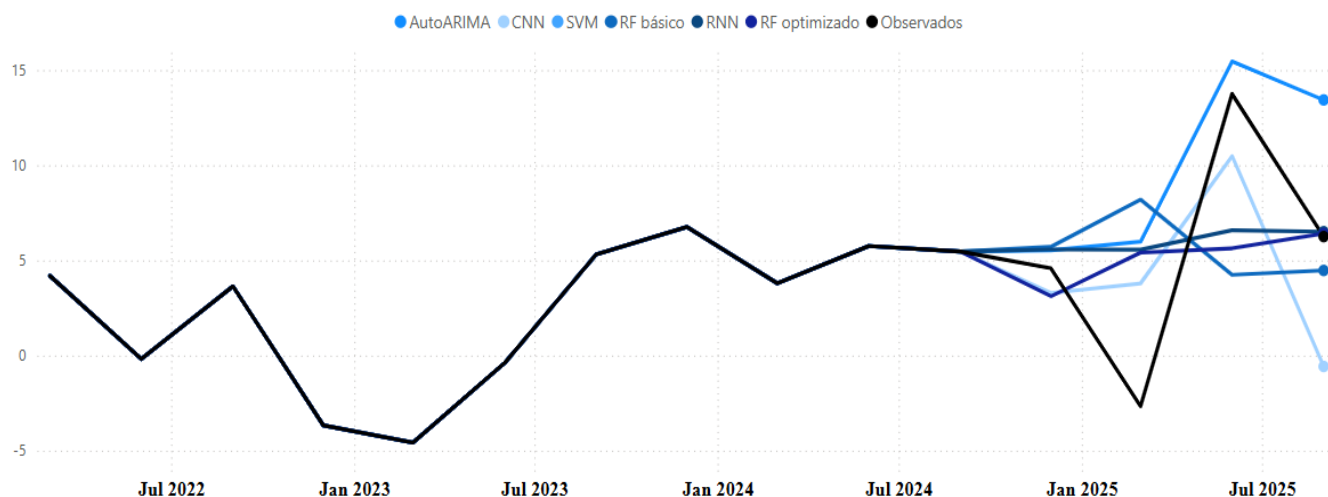
Figura 13: Actividad económica servicios a corto plazo



Fuente: Elaboración propia

En la figura 13, respecto a la actividad de Servicios, el comportamiento proyectado para 2025 revela heterogeneidad significativa entre modelos. SVM muestra un comportamiento errático con un salto hasta 8.8%, mientras CNN prevé apenas 2.9%. RF básico, RNN y AutoARIMA convergen en valores entre 4.3% y 5.2%. Dada la alta composición heterogénea del sector, que incluye tanto servicios transables como no transables, la comparación de desempeño sugiere que RF optimizado ofrece una mejor estabilidad predictiva y captura la dinámica multicomponente con menor sesgo.

Figura 14: Actividad económica construcción a corto plazo



Fuente: Elaboración propia

En la figura 14, la actividad económica del sector Construcción evidencia divergencias importantes en las predicciones. AutoARIMA y CNN anticipan un crecimiento por encima del 13%, mientras que SVM proyecta una contracción significativa (-2.7%). RNN y RF básico promedian entre 5.7% y 6.6%. Considerando la sensibilidad del sector a políticas de inversión pública y choques exógenos, se evidencia que el modelo RNN ofrece una mayor robustez, con menor sensibilidad a extremos y mejor desempeño en validación cruzada.

En conjunto, los resultados empíricos revelan que los modelos de aprendizaje profundo (RNN, CNN) tienden a suavizar las trayectorias, mientras que AutoARIMA y SVR presentan mayor sensibilidad a extremos o ciclos recientes. RF básico y optimizada ofrecen un balance entre capacidad predictiva, estabilidad y sensibilidad sectorial, siendo especialmente efectivos en sectores con alta exposición a factores externos o multicolinealidad estructural entre predictores.

VI. Conclusión

La identificación oportuna de cambios en la dinámica económica es esencial para mitigar los efectos de los ciclos económicos y fomentar un desarrollo sostenible. Para lograrlo, se requiere contar con herramientas analíticas capaces de generar estimaciones precisas y actualizadas sobre la evolución de la actividad económica en el corto plazo. En contextos donde los indicadores oficiales, como el IMAE, se publican con rezago, resulta crucial desarrollar mecanismos de proyección sectorial que permitan anticipar tendencias, facilitar la toma de decisiones informadas y proveer insumos valiosos para el diseño e implementación de políticas económicas más oportunas y focalizadas.

Con este propósito, en esta investigación se construyó un sistema de predicción de la actividad económica sectorizada que integra enfoques tradicionales de series temporales y técnicas avanzadas de aprendizaje automático. A través de modelos como AutoARIMA, *Random Forest* (en sus versiones básica y optimizada), *Support Vector Regression*, redes neuronales recurrentes (RNN) y convolucionales (CNN), se logró capturar tanto la estructura temporal de las series como patrones no lineales complejos característicos de los distintos sectores económicos.

Los hallazgos empíricos revelan que no existe un único modelo óptimo para todos los sectores. Por ejemplo, en industrias con dinámicas persistentes, como las zonas francas, AutoARIMA logra una buena aproximación, mientras que en sectores más inestables o con patrones estacionales marcados, como la agricultura o la construcción, los modelos basados en redes neuronales muestran mayor capacidad de adaptación. En general, RF optimizado se posiciona como una herramienta confiable, mostrando consistencia predictiva y bajo error fuera de muestra en diversos sectores, por lo que constituye una herramienta idónea para su uso en entornos institucionales.

Una de las principales fortalezas del sistema propuesto radica en su capacidad de actualización flexible. A medida que se incorpora nueva información, las proyecciones se ajustan de forma automática, facilitando el monitoreo continuo y proporcionando alertas tempranas para la formulación de políticas adaptadas al contexto. Asimismo, el modelo permite identificar la contribución relativa de cada variable predictora, enriqueciendo el análisis económico y mejorando la toma de decisiones basada en datos.

Aunque las predicciones ganan precisión en la medida en que se completa el conjunto informativo disponible, los modelos desarrollados ofrecen un desempeño superior al de metodologías convencionales, lo que reafirma su utilidad para anticipar escenarios económicos. En este marco, el sistema de predicción desarrollado en esta tesis representa una herramienta de gran valor para la planificación, permitiendo intervenciones más precisas, oportunas y focalizadas por sector económico.

Sin embargo, esta investigación no está exenta de limitaciones estructurales que deben ser consideradas. Una de las principales reside en la incorporación de técnicas que capturen cambios estructurales en los regímenes económicos, como modelos de cambio de régimen o de volatilidad estocástica, que podrían mejorar el desempeño predictivo en entornos de alta incertidumbre.

A partir de estas limitaciones, se abren diversas líneas de investigación futuras. Una dirección prometedora es la incorporación de modelos multivariados que integren relaciones dinámicas entre sectores, como los VAR estructurales o los modelos dinámicos bayesianos. También se sugiere explorar arquitecturas híbridas que combinen componentes lineales (para la captura de tendencias y estacionalidades) con capas profundas no lineales, lo que permitiría capturar tanto la estructura temporal como las no linealidades complejas de forma más robusta. Además, integrar indicadores alternativos de alta frecuencia como consumo eléctrico, movilidad o datos de comercio exterior en tiempo real podría enriquecer la base informativa y fortalecer la capacidad de *nowcasting*.

Finalmente, la implementación de técnicas de detección de puntos de cambio (*change point detection*) y el uso de simulaciones contrafactuales permitirían evaluar con mayor profundidad la sensibilidad del sistema de predicción ante shocks inesperados, mejorando así su utilidad como instrumento anticipativo en la formulación de política económica.

VII. Referencias Bibliográficas

- Barrios, J., Escobar, J., Leslie, J., Martín, L., & Peña, W. (2021). *Nowcasting para predecir actividad económica en tiempo real: los casos de Belice y El Salvador*. Banco Interamericano de Desarrollo. Obtenido de file:///C:/Users/ciuda/Downloads/Nowcasting-para-predecir-actividad-economica-en-tiempo-real-los-casos-de-Belice-y-El-Salvador.pdf
- Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*.
- Bolhuis, D., & Rayner, J. (2020). *Nowcasting GDP using machine learning algorithms: A realtime assessment*. Reserve Bank of Australia Bulletin.
- Box, J., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco.
- Breiman, L. (2001). *Random Forest, Machine Learning*. Vol.45. pp.5-32.
- Breiman, L. (2001). *Random forests. Machine Learning*.
- Camacho, M., & Pérez Quiroz, G. (2011). *Latin STINGS: Indicadores de Crecimiento a Corto*. Santiago de Chile: Serie Macroeconomía del Desarrollo 108, Comisión Económica para América Latina y el Caribe (CEPAL), División de Desarrollo Económico .
- Camacho, M., & Pérez Quiros, G. (2008). *Introducing the Euro-STING: Short Term INDicator*. Banco de España Working Paper No.0807.
- Camacho, M., & Pérez Quiroz, G. (2010). *Spain-STING: Spain Short Term Indicator of Growth*.
- Cortes, C., & Vapnik, V. (1995). *Support-vector networks. Machine Learning*.
- Dickey, D., & Fuller, W. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association .
- Géron, A. (2019). *Hans~on Machine Learning with Sakit-Learn, Keras and Tensorflow*. O'Reilly.
- Geweke, J. (1977). *The Dynamic Factor Analysis of Economic Time Series*. Amsterdam.
- Goulet, C., Leroux, M., Stevanovic, D., & Stéphane, S. (2019). *How is Machine Learning Useful for Macroeconomic Forecasting*.
- Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory. Neural Computation*.
- Mariano, R., & Murasawa, Y. (2003). *new coincident index os business cycles based on*. Journal of Applied Econometrics 18, pages 427-443.
- Natekin, A., & Knoll, A. (2013). *Gradient boosting machines*.
- Paredes, E., Hawa, N., & Polanco, J. (2015). *Indicador Adelantado del Ciclo Económico para República Dominicana*. Banco Central de la República Dominicana, Doc. Trabajo 15-03.
- Sargents, T., & Sims, C. (1977). *Business Cycle Modelling Without Pretending*. Federal.
- Sims, C. (1980). *Macroeconomics and reality*. Econometrica.

- Stock, J., & Watson, M. (1991). *Probability model of the coincident economic indicators*. In. Cambridge University.
- Stock, J., & Watson, M. (2001). *Vector autoregressions*. Journal of Economic Perspectives.
- Stock, J., & Watson, M. (2012). *Generalized Shrinkage Methods for Forecasting Using Many Predictors*. Journal of Business Economics and Statistics.
- Zhang, G., & Eddy , B. (2017). *Forecasting with artificial neural networks*. Journal of Forecasting.

VIII. Anexos

A. Tablas Adicionales

Apéndice A: Transformaciones Aplicadas a las Variables Predictoras

Categoría	Variable	Transformación Aplicada
Deuda Pública	deuda_publica_consolidada	Logaritmo natural y diferencia de primer orden
	deuda_externa_gobierno_central	Logaritmo natural y diferencia de primer orden
	deuda_interna_banco_central	Logaritmo natural y diferencia de primer orden
Precios	indice_de_precios_al_consumidor_nacional	Diferencia de primer orden
	ipc_alimentos_y_bebidas_no_alcoholicas	Diferencia de primer orden
	indice_de_precios_de_bienes_transables	Diferencia de primer orden
Fiscalidad	ingresos_fiscales	Logaritmo natural y diferencia de primer orden
	impuesto_valor_agregado_itbis	Logaritmo natural y diferencia de primer orden
	gastos_de_capital_del_gobierno	Logaritmo natural y diferencia de primer orden
Sector Externo	exportaciones_totales	Logaritmo natural y diferencia de primer orden
	importaciones_real	Logaritmo natural y diferencia de primer orden
Monetario	tasa_interes_interbancaria	Logaritmo natural, se ajustó la estacionalidad.
	tasa_de_cambio_peso_dolar_us	Diferencia de logaritmos (variación porcentual)

Categoría	Variable	Transformación Aplicada
Empleo	empleo_industrias	Logaritmo natural y diferencia de primer orden
	empleo_construccion	Logaritmo natural y diferencia de primer orden
Commodities	com_indice_de_petroleo_crudo_promedio	Logaritmo natural y diferencia de primer orden
	com_oro	Logaritmo natural y diferencia de primer orden
VARIABLES DE EE.UU.	pib_usa	Logaritmo natural y diferencia de primer orden mensual por interpolación
	tasa_interes_fed	Logaritmo natural y diferencia de primer orden

Apéndice B: Pruebas

Pruebas (ADF)

	Variable	ADF	p-value (Nivel)	ADF_dif	p-value_dif	Orden_Integración	constante
0	log_deuda_publica_consolidada_growth	-1.358	0.000	-5.035	0.000	1	False
1	log_deuda_publica_externa_consolidada_growth	-1.619	0.000	-4.788	0.000	1	False
2	log_deuda_publica_interna_consolidada_growth	-2.884	0.047	NaN	NaN	0	False
3	log_deuda_sector_publico_no_financiero_growth	-1.773	0.000	-4.828	0.000	1	False
4	log_deuda_total_gobierno_central_growth	-1.893	0.000	-4.835	0.000	1	False
5	log_deuda_externa_gobierno_central_growth	-1.616	0.000	-5.007	0.000	1	False
6	log_deuda_interna_gobierno_central_growth	-3.366	0.012	NaN	NaN	0	False
7	log_deuda_intragubernamental_growth	-3.933	0.002	NaN	NaN	0	False
8	log_deuda_intragubernamental_con_bcrd_growth	-3.366	0.012	NaN	NaN	0	False
9	log_deuda_total_resto_del_spnf_growth	-2.795	0.003	-3.814	0.003	1	False
10	log_deuda_externa_resto_del_spnf_growth	-3.977	0.002	NaN	NaN	0	False
11	log_deuda_interna_resto_del_spnf_growth	-2.790	0.003	-3.819	0.003	1	False
12	log_deuda_sector_publico_financiero_growth	-2.768	0.002	-3.906	0.002	2	False
13	log_deuda_total_banco_central_growth	-2.768	0.002	-3.906	0.002	2	False
14	log_deuda_externa_banco_central_growth	-2.383	0.000	-5.224	0.000	1	False
15	log_deuda_interna_central_banco_central_growth	-2.623	0.000	-6.816	0.000	2	False
16	log_indice_de_precios_al_consumidor_nacional_g...	-0.944	0.006	-3.564	0.006	1	False
17	log_ipc_alimentos_y_bebidas_no_alcoholicas_growth	-2.068	0.000	-4.474	0.000	1	False
18	log_ipc_bebidas_alcoholicas_y_tabaco_growth	-4.186	0.001	NaN	NaN	0	False
19	log_ipc_prendas_de_vestir_y_calzado_growth	-1.474	0.003	-3.793	0.003	4	False
20	log_ipc_vivienda_growth	-2.372	0.000	-4.816	0.000	1	False
21	log_ipc_muebles_growth	-2.274	0.000	-6.760	0.000	2	False
22	log_ipc_salud_growth	-1.037	0.001	-4.004	0.001	2	False
23	log_ipc_transporte_growth	-2.715	0.002	-3.871	0.002	1	False

24	log_ipc_comunicaciones_growth	-1.548	0.001	-4.209	0.001	1	False
25	log_ipc_recreacion_y_cultura_growth	-0.670	0.019	-3.224	0.019	1	False
26	log_ipc_educacion_growth	-2.333	0.000	-6.338	0.000	2	False
27	log_ipc_restaurantes_y_hoteles_growth	-1.250	0.003	-3.772	0.003	3	False
28	log_ipc_bienes_y_servicios_diversos_growth	-2.962	0.039	NaN	NaN	0	False
29	log_indice_de_precios_al_consumidor_subyacente...	-1.245	0.016	-3.285	0.016	1	False
30	log_indice_de_precios_de_bienes_transables_growth	-2.391	0.000	-5.453	0.000	2	False
31	log_indice_de_precios_de_bienes_no_transables_...	0.742	0.001	-4.262	0.001	1	False
32	log_ingresos_fiscales_growth	-4.163	0.001	NaN	NaN	0	False
33	log_impuestos_growth	-2.424	0.000	-5.142	0.000	1	False
34	log_impuestos_ingreso_utilidades_y_ganancias_d...	-4.400	0.000	NaN	NaN	0	False
35	log_pagaderos_por_personas_fisicas_growth	-2.317	0.000	-5.097	0.000	1	False
36	log_pagaderos_por_sociudades_y_otras_empresas_...	-4.197	0.001	NaN	NaN	0	False
37	log_otros_impuestos_ingreso_utilidades_y_ganan...	-4.622	0.000	NaN	NaN	0	False
38	log_impuestos_sobre_la_propiedad_growth	-4.119	0.001	NaN	NaN	0	False
39	log_impuestos_sobre_los_bienes_y_servicios_growth	-2.694	0.000	-5.314	0.000	1	False
40	log_impuestos_generales_sobre_los_bienes_y_ser...	-2.354	0.000	-5.031	0.000	1	False
41	log_impuesto_valor_agregado_itbis_growth	-3.891	0.002	NaN	NaN	0	False
42	log_impuestos_sobre_transacciones_financieras_...	-2.943	0.041	NaN	NaN	0	False
43	log_impuestos_selectivos_sobre_bienes_growth	-3.147	0.023	NaN	NaN	0	False
44	log_impuestos_sobre_servicios_especificos_growth	-5.254	0.000	NaN	NaN	0	False
45	log_impuestos_uso_de_bienes_y_permiso_para_usa...	-2.295	0.000	-6.421	0.000	1	False
46	log_impuestos_comercio_y_transacciones_interna...	-2.975	0.037	NaN	NaN	0	False
47	log_contribuciones_sociales_growth	-5.127	0.000	NaN	NaN	0	False
48	log_otros_ingresos_gobierno_growth	-5.330	0.000	NaN	NaN	0	False
49	log_gastos_del_gobierno_growth	-3.352	0.013	NaN	NaN	0	False
50	log_remuneracion_empleados_growth	-5.054	0.000	NaN	NaN	0	False

52	log_contribuciones_sociales_de_los_empleadores...	-4.529	0.000	NaN	NaN	0	False
53	log_uso_de_bienes_y_servicios_growth	-5.413	0.000	NaN	NaN	0	False
54	log_intereses_pago_gobierno_growth	-1.871	0.027	-3.089	0.027	1	False
55	log_a_no_residentes_growth	-1.851	0.000	-4.452	0.000	1	False
56	log_a_residentes_growth	-4.167	0.001	NaN	NaN	0	False
57	log_subsidios_growth	-3.448	0.009	NaN	NaN	0	False
58	log_subsidiosempresaspublicas_growth	-3.007	0.034	NaN	NaN	0	False
59	log_donaciones_del_gobierno_growth	-4.526	0.000	NaN	NaN	0	False
60	log_a_organismos_internacionales_growth	-4.879	0.000	NaN	NaN	0	False
61	log_a_otras_unidades_del_gobierno_general_growth	-4.344	0.000	NaN	NaN	0	False
62	log_gastos_corrientes_del_gobierno_growth	-4.421	0.000	NaN	NaN	0	False
63	log_gastos_de_capital_del_gobierno_growth	-7.192	0.000	NaN	NaN	0	False
64	log_prestaciones_sociales_growth	-4.659	0.000	NaN	NaN	0	False
65	log_otros_gastos_del_gobierno_growth	-7.239	0.000	NaN	NaN	0	False
66	log_transferencias_no_clasificadas_en_otra_par...	-7.239	0.000	NaN	NaN	0	False
67	log_transferencias_gubernamentales_corrientes_...	-2.848	0.000	-5.928	0.000	1	False
68	log_transferencias_gubernamentales_de_capital_...	-7.494	0.000	NaN	NaN	0	False
69	log_gastos_gubernamentales_para_proyectos_de_i...	-3.486	0.008	NaN	NaN	0	False
70	log_inversion_bruta_en_activos_no_financieros_...	-3.656	0.005	NaN	NaN	0	False
71	log_activos_fijos_gobierno_growth	-3.715	0.004	NaN	NaN	0	False
72	log_gasto_de_capital_manual_efp_fmi_1986_growth	-1.745	0.000	-5.721	0.000	1	False
73	log_erogacion_gastos_gobierno_growth	-5.023	0.000	NaN	NaN	0	False
74	log_tasa_interes_interbancaria_growth	-1.854	0.000	-4.855	0.000	1	False
75	log_tasa_interes_activa_rd_growth	-0.300	0.000	-4.632	0.000	1	False
76	log_tasa_interes_deposito_ahorros_growth	-1.646	0.034	-3.006	0.034	1	False
77	log_tasa_interes_pasiva_rd_growth	-2.311	0.000	-4.322	0.000	2	False
78	log_tasa_interes_deposito_p_us_growth	-2.571	0.018	-3.237	0.018	1	False

79	log_tasa_interes_deposito_ahorros_us_growth	-2.043	0.004	-3.677	0.004	1	False
80	log_tasa_interes_pasiva_us_growth	-2.573	0.000	-4.280	0.000	1	False
81	log_tasa_de_cambio_peso_dolar_us_growth	-3.352	0.013	NaN	NaN	0	False
82	log_indicador_mensual_de_actividad_economica_g...	-2.515	0.000	-7.454	0.000	2	False
83	log_base_monetaria_restringida_growth	-3.355	0.013	NaN	NaN	0	False
84	log_base_monetaria_ampliada_growth	-2.684	0.000	-5.905	0.000	2	False
85	log_medio_circulante_M1_growth	-2.120	0.000	-4.385	0.000	2	False
86	log_oferta_monetaria_M2_growth	-2.173	0.000	-5.718	0.000	1	False
87	log_dinero_sentido_ampliado_M3_growth	-2.853	0.001	-4.161	0.001	2	False
88	log_empleo_empleototal_growth	-2.023	0.000	-4.770	0.000	1	False
89	log_empleo_agricultura_y_ganaderia_growth	-2.335	0.000	-5.978	0.000	1	False
90	log_empleo_industrias_2_growth	-3.514	0.008	NaN	NaN	0	False
91	log_empleo_electricidad_y_agua_growth	-2.063	0.000	-4.534	0.000	1	False
92	log_empleo_construccion_growth	-3.698	0.004	NaN	NaN	0	False
93	log_empleo_comercio_growth	-1.510	0.000	-5.185	0.000	1	False
94	log_empleo_hoteles_bares_y_restaurantes_growth	-4.090	0.001	NaN	NaN	0	False
95	log_empleo_transporte_y_comunicaciones_growth	-3.259	0.017	NaN	NaN	0	False
96	log_empleo_intermediacion_financieras_y_seguro...	-4.432	0.000	NaN	NaN	0	False
97	log_empleo_administracion_publica_y_defensa_gr...	-3.072	0.029	NaN	NaN	0	False
98	log_empleo_ensenanza_growth	-2.522	0.000	-5.967	0.000	1	False
99	log_empleo_salud_y_asistencia_social_growth	-1.661	0.000	-5.735	0.000	1	False
100	log_empleo_otros_servicios_growth	-2.015	0.000	-6.255	0.000	1	False
101	log_com_indice_de_commodities_growth	-1.784	0.000	-5.823	0.000	1	False
102	log_com_indice_de_precios_de_alimentos_growth	-2.249	0.000	-5.522	0.000	1	False
103	log_com_indice_de_metales_base_growth	-2.836	0.000	-4.279	0.000	1	False
104	log_com_indice_de_metales_preciosos_growth	-2.313	0.038	-2.972	0.038	1	False
105	log_com_indice_de_metales_sin_oro_growth	-2.824	0.005	-3.674	0.005	1	False

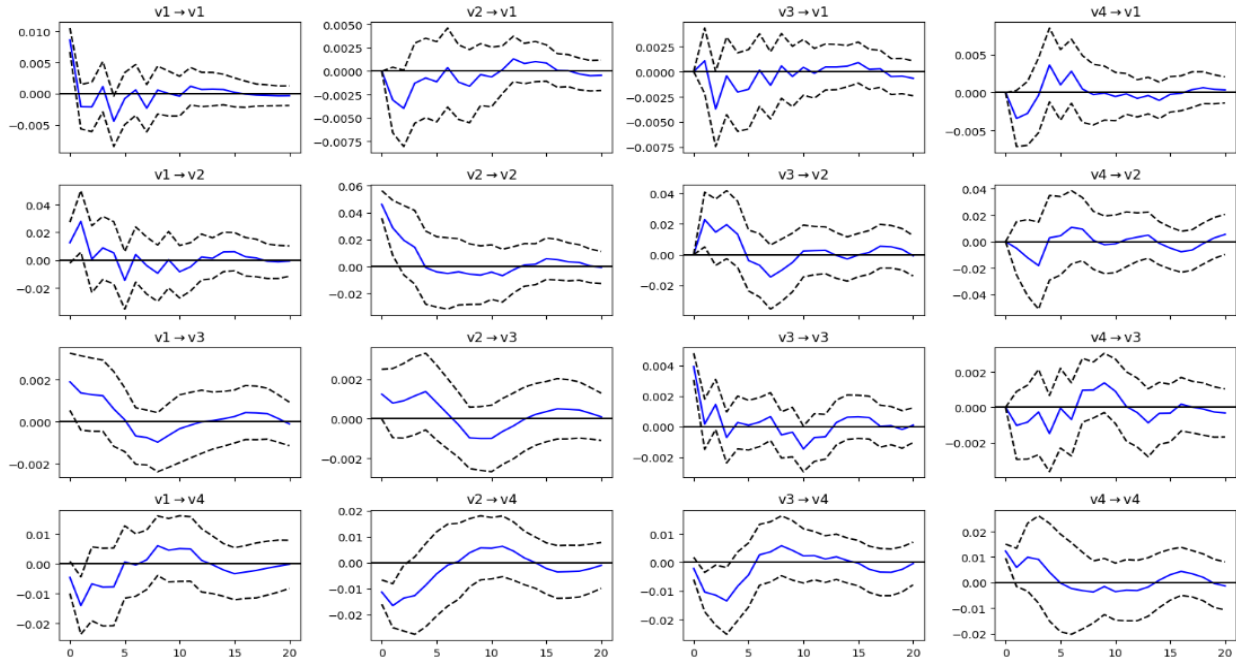
106	log_com_indice_de_energia_growth	-1.925	0.000	-5.602	0.000	1	False
107	log_com_indice_promedio_de_petroleo_crudo_wit...	-2.416	0.001	-4.257	0.001	1	False
108	log_com_indice_promedio_de_gas_natural_growth	-4.680	0.000	NaN	NaN	0	False
109	log_com_indice_de_carbon_growth	-1.942	0.000	-5.189	0.000	1	False
110	log_com_carbon_australiano_growth	-1.766	0.000	-5.257	0.000	1	False
111	log_com_carbon_sudafricano_growth	-3.071	0.029	NaN	NaN	0	False
112	log_com_cobre_growth	-3.156	0.023	NaN	NaN	0	False
113	log_com_gas_natural_ttf_growth	-5.799	0.000	NaN	NaN	0	False
114	log_com_gas_natural_lng_japon_growth	-4.660	0.000	NaN	NaN	0	False
115	log_com_gas_natural_henry_hub_growth	-3.244	0.018	NaN	NaN	0	False
116	log_com_indice_de_petroleo_crudo_promedio_growth	-2.446	0.001	-4.244	0.001	1	False
117	log_com_petroleo_brent_growth	-2.434	0.000	-4.315	0.000	1	False
118	log_com_petroleo_dubai_fateh_growth	-2.620	0.000	-4.275	0.000	1	False
119	log_com_petroleo_wti_growth	-2.320	0.001	-4.080	0.001	1	False
120	log_com_harina_de_soja_growth	-2.293	0.000	-5.254	0.000	1	False
121	log_com_aceite_de_soja_growth	-2.299	0.033	-3.021	0.033	1	False
122	log_com_soja_growth	-2.052	0.000	-5.667	0.000	2	False
123	log_com_azucar_mercado_libre_growth	-2.109	0.000	-6.521	0.000	1	False
124	log_com_azucar_importacion_eeuu_growth	-1.641	0.000	-5.707	0.000	1	False
125	log_com_trigo_hard_red_growth	-2.479	0.000	-5.705	0.000	1	False
126	log_com_oro_growth	-2.598	0.043	-2.919	0.043	1	False
127	log_com_plata_growth	-2.647	0.005	-3.643	0.005	1	False
128	log_com_platino_growth	-2.233	0.004	-3.715	0.004	1	False
129	log_pib_usa_growth	-2.927	0.042	NaN	NaN	0	False
130	log_exportaciones_usa_growth	-2.601	0.003	-3.791	0.003	1	False
131	log_importaciones_usa_growth	-1.819	0.000	-4.417	0.000	1	False
132	log_gasto_federal_growth	-3.768	0.003	NaN	NaN	0	False
133	log_ingresos_federales_corrientes_growth	-2.238	0.001	-4.109	0.001	1	False

133	log_ingresos_federales_corrientes_growth	-2.238	0.001	-4.109	0.001	1	False
134	log_ipc_usa_precio_produccion_growth	-1.234	0.000	-5.323	0.000	1	False
135	log_ipc_usa_urbano_growth	-2.273	0.003	-3.820	0.003	1	False
136	log_ipc_usa_urbano_combustibles_growth	-2.736	0.001	-4.088	0.001	2	False
137	log_ipc_usa_urbano_energia_growth	-0.990	0.027	-3.096	0.027	1	False
138	log_tasa_interes_fed_growth	-2.721	0.001	-4.153	0.001	1	False
139	log_empleos_usa_vacantes_growth	-2.177	0.000	-5.325	0.000	1	False
140	log_empleos_usa_contrataciones_growth	-3.479	0.009	NaN	NaN	0	False
141	log_tasa_politica_monetaria_growth	-2.600	0.009	-3.474	0.009	1	False
142	log_depositos_growth	-3.068	0.029	NaN	NaN	0	False
143	log_prestamos_growth	-2.743	0.005	-3.671	0.005	1	False
144	log_ied_flujos_totales_growth	-4.523	0.000	NaN	NaN	0	False
145	log_minerales_growth	-2.557	0.000	-5.499	0.000	1	False
146	log_agropecuarios_growth	-2.313	0.000	-6.069	0.000	1	False
147	log_industriales_growth	-3.733	0.004	NaN	NaN	0	False
148	log_exportaciones_totales_growth	-4.103	0.001	NaN	NaN	0	False
149	log_producto_interno_bruto_nominal_growth	-3.320	0.014	NaN	NaN	0	False
150	log_agropecuario_real_growth	-3.047	0.031	NaN	NaN	0	False

Funciones Impulso-Respuesta

{'log_indice_de_precios_al_consumidor_nacional_growth_dif_1': 'v1', 'log_producto_interno_bruto_real_growth_dif_0': 'v2', 'tasa_politica_monetaria_growth_dif_0': 'v3', 'log_tasa_de_cambio_peso_dolar_us_growth_dif_0': 'v4'}

Funciones Impulso Respuesta



Descomposición de la Varianza del Error de Pronóstico (FEVD)

```
{'log_indice_de_precios_al_consumidor_nacional_growth_dif_1': 'v1', 'log_producto_interno_bruto_real_growth_dif_0': 'v2', 'tasa_politica_monetaria_growth_dif_0': 'v3', 'log_tasa_de_cambio_peso_dolar_us_growth_dif_0': 'v4'}
```

