

# **Análisis de métodos y modelos de pronóstico de variables económicas aplicables al estudio del empleo y de las políticas y reformas laborales**

**Documento de Trabajo\***

**Jesuswaldo Martínez Soria**  
**Santiago Licea Becerril**

**Dirección General de Investigación Estratégica**  
**Instituto Belisario Domínguez**  
**Senado de la República**

\* Este documento de trabajo se ha realizado bajo la dirección de Jesuswaldo Martínez Soria, investigador C de la Dirección General de Investigación Estratégica (DGIE) del Instituto Belisario Domínguez, y presenta diversos contenidos elaborados durante el servicio social de Santiago Licea Becerril, estudiante de la carrera de Ciencia de Datos impartida en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM. Los autores agradecen los comentarios del Dr. Juan Manuel Rodríguez de la DGIE y de la Dra. Verónica Arredondo y de Lucía Lozano de la Unidad de Matemáticas de la Universidad Autónoma de Zacatecas, que amablemente revisaron versiones preliminares de este documento, en el marco del convenio de colaboración con el IBD suscrito el 20 de noviembre de 2023.

## Introducción

El objetivo de este documento de trabajo es analizar modelos basados en el conocimiento de la econometría, las matemáticas y los métodos estadísticos que realicen pronósticos de variables económicas y que tengan utilidad en la evaluación de las políticas económicas. En particular, se trata de estudiar algunos métodos y modelos que estimen el desempeño y el pronóstico de variables de ocupación y empleo y que, eventualmente, permitan la valoración de los impactos que tienen las políticas y las reformas legislativas sobre el comportamiento del mercado de trabajo en México.

Esta línea de investigación surge a raíz del trabajo de Martínez y Vargas (2023), en donde se analiza la evolución y las perspectivas del empleo formal en México, de cara a la formulación de un balance sexenal de los resultados de las políticas de empleo. En ese estudio se sostiene la posibilidad de que, tanto la crisis de la pandemia en 2020 como la implementación de las políticas económicas de los años recientes, hayan repercutido en las capacidades productivas y en el mercado de trabajo propiciando un fuerte debilitamiento en la tendencia de largo plazo del número de trabajadores formales y, con ello, se hayan deteriorado sus perspectivas de crecimiento.

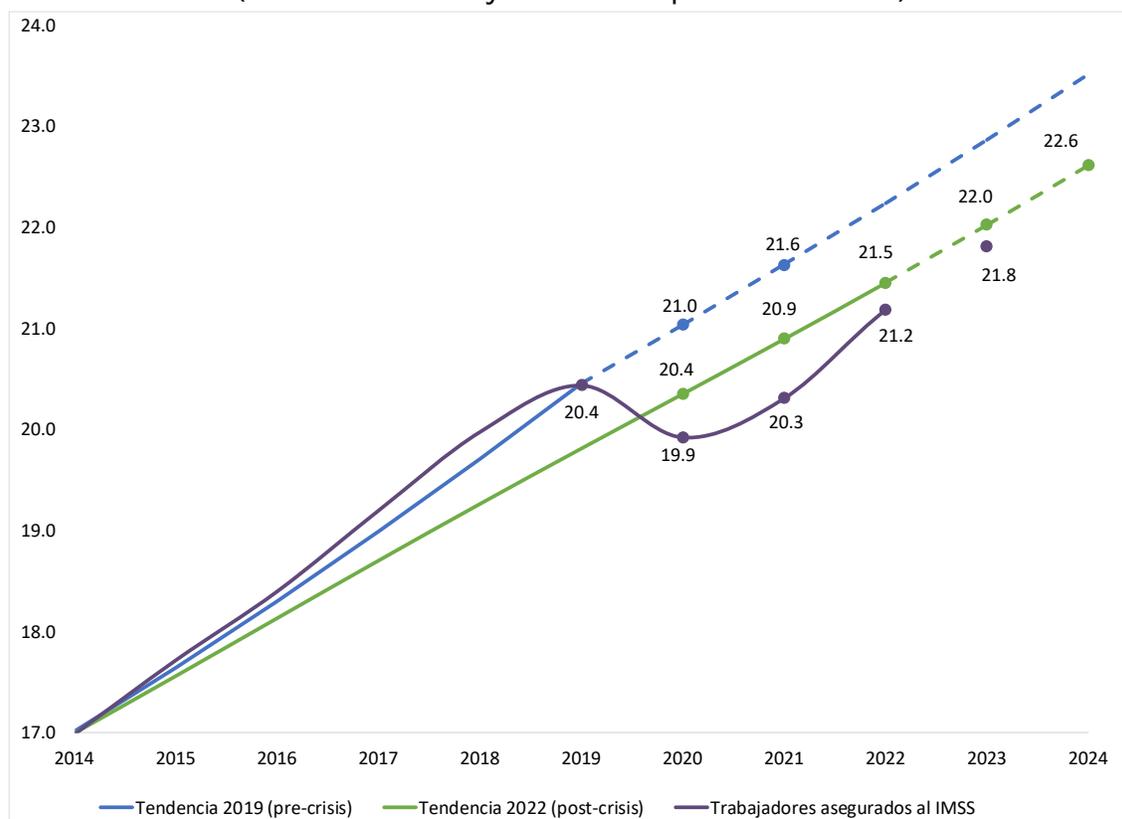
Para probar esta hipótesis se estimaron las tendencias de largo plazo y el ciclo del empleo formal para proyectar los niveles de empleo en 2023 y 2024, teniendo en cuenta que la información estadística disponible hasta ese momento llegaba a julio de 2023. La estimación de las tendencias de largo plazo se obtuvo a través de la descomposición de la serie histórica de los trabajadores asegurados en el Instituto Mexicano del Seguro Social (IMSS). Mediante la aplicación del filtro Hodrick-Prescott (1997) se obtuvieron dos variables subyacentes: la tendencia de largo plazo y las variaciones que constituyen el ciclo del empleo (ver gráfica 4 de Martínez y Vargas, 2023).

De acuerdo con la teoría y los resultados obtenidos, la tendencia del empleo formal sigue un comportamiento ascendente y estable, puesto que representa la evolución de los niveles de ocupación formal determinados por las condiciones estructurales de la economía y del mercado laboral. El componente cíclico constituye las variaciones determinadas por la situación coyuntural de la economía y del mercado laboral, por lo que indica los momentos de expansión, auge, recesión y crisis del empleo.

Este ejercicio permitió realizar proyecciones de crecimiento del empleo formal, extrapolando la trayectoria de las tendencias de largo plazo. En la figura 1 —tomada de Martínez y Vargas (2023)— se presentan los pronósticos de los niveles de empleo formal para 2023 y 2024 (línea punteada de color verde), con base en la tendencia de largo plazo —denominada post-crisis— cuya trayectoria fue afectada por el desplome en los niveles

de empleo en 2020. El pronóstico para el cierre de 2023 fue de 22.0 millones de trabajadores asegurados y para 2024 fue de 22.6 millones.

Figura 1. Proyecciones del empleo formal  
(Millones de trabajadores incorporados al IMSS)



Nota: El número de asegurados al IMSS en 2023 es la cifra acumulada a julio. El número de trabajadores es un promedio de los datos mensuales desestacionalizados con el método Census X-11.

Fuente: Martínez y Vargas (2023).

La eficacia de estos pronósticos se puede evaluar teniendo en cuenta los valores reales registrados por el IMSS al cierre de cada año. En el caso de 2023, la cifra puntual pronosticada fue de 22,030,022 trabajadores y la cifra real registrada por el IMSS fue de 22,024,386; es decir, la diferencia en la estimación fue de sólo 5,636 trabajadores, que equivale a una desviación del 0.026% con respecto al valor real, lo que indica un margen de error lo suficientemente reducido como para confiar en el método aplicado.

Más allá de la interpretación de los resultados y de la utilidad que tienen en la valoración de los efectos de las políticas públicas sobre el empleo —cuestiones que se analizan en el texto de Martínez y Vargas (2023)—, con la obtención de estas cifras y la aplicación de esta metodología se abre una reflexión técnica importante a propósito del estudio: ¿Qué

tan eficiente es ese método para predecir el comportamiento de las variables de empleo en comparación con otros métodos y modelos utilizados en el campo de las ciencias exactas y la econometría?

Para resolver esta pregunta de investigación, en este documento de trabajo se presenta la revisión de la literatura especializada que explica el uso de modelos de pronóstico y una serie de ejercicios que buscan evidencias sobre la eficiencia de los métodos. En particular, en este trabajo se obtuvieron los siguientes resultados:

- Se llegó a una metodología cuantitativa que busca minimizar la varianza de la media del error del pronóstico de múltiples modelos con el fin de compararlos.
- Se construyeron diferentes arquitecturas para cada uno de los modelos aplicados que pueden servir para modelar otras variables económicas en trabajos futuros.
- Tras construir una serie de modelos basados en prácticas del estado del arte para la predicción de variables económicas, se determinó que para los datos con los que se trabajó, la complejidad de los modelos no contribuyó a una mejora sustantiva en la precisión de los pronósticos.
- La extrapolación de la tendencia a largo plazo mostró una capacidad predictiva comparable con modelos de aprendizaje automático.

Los esfuerzos realizados en este documento van en línea con trabajos previos, como el de Guerrero (2013), en el que se concluye que los métodos propuestos por la Organización para la Cooperación y el Desarrollo Económico (OCDE), para estimar tendencias y ciclos y para generar los índices cíclicos coincidente y adelantado, resultan ser los más precisos en comparación con otros métodos del estado del arte. En particular, para estimar tendencias y ciclos se comparó el método PAT (tendencia promedio de fase) de la Oficina Nacional de Investigación Económica (NBER por sus siglas en inglés), el filtro de Hodrick-Prescott, propuesto por la OCDE, y el filtro de paso de banda de Christiano y Fitzgerald, que utiliza la Oficina Europea de Estadística (EUROSTAT). Por otro lado, para estimar los índices cíclicos se compararon los métodos del NBER, de la OCDE y de Stock-Watson.

Tanto en las estimaciones del índice adelantado de la OCDE como en la obtención de la tendencia a largo plazo en el trabajo de Martínez y Vargas (2023) se utiliza el filtro de Hodrick-Prescott; lo que invita a reflexionar sobre el potencial de sus capacidades predictivas y de otros métodos que logren modelar de mejor forma los ciclos y fluctuaciones de variables económicas. De aquí que se opta por recurrir a modelos econométricos y de aprendizaje automático empleados en el estado del arte.

En cualquier caso, el uso de métodos estadísticos y modelos matemáticos ha sido recurrente en la planeación y la evaluación de las políticas públicas. En distintos entornos

públicos se aplican modelos para tomar decisiones de política económica pero también para instrumentar programas, regulaciones y acciones gubernamentales de diversa índole, por lo que el estudio de las herramientas que aquí se desarrolla, constituye un insumo técnico que puede ser empleado para robustecer las líneas de la investigación legislativa.

Estudios como el que se presenta aquí cobran una particular importancia en el entorno legislativo, al menos por dos razones relacionadas con la función de control que tiene el Senado de la República sobre las acciones del Poder Ejecutivo. Una es la utilidad que tienen estos métodos y modelos en las labores de seguimiento, monitoreo y evaluación de las políticas gubernamentales y de las reformas legislativas, que buscan formular el balance de los esfuerzos públicos y conocer los efectos de las decisiones colectivas sobre el desarrollo económico y social del país. Saber los resultados de las políticas públicas de manera oportuna, por ejemplo, permite corregir errores en el diseño o la implementación de estas. Asimismo, anticipar la trayectoria de los indicadores económicos y sociales, a través de métodos estadísticos y modelos de pronóstico, ayudaría a reorientar o reformular los esfuerzos legislativos y las acciones gubernamentales.

La otra razón es la utilidad que tienen en el análisis de los presupuestos públicos, ya que todos los años se somete al Poder Legislativo el paquete económico realizado por la Secretaría de Hacienda y Crédito Público, que incluye los presupuestos públicos de los gastos e ingresos de la nación. En este paquete se entregan los Criterios Generales de Política Económica que contienen una serie de previsiones sobre el comportamiento futuro de diversas variables de producción, precios, empleo, etc. Con base en esas previsiones, la autoridad hacendaria toma las decisiones de política fiscal que son propuestas, en los presupuestos, a las cámaras legislativas. Como parte de sus responsabilidades, ambas cámaras se encargan de analizar, dictaminar y, eventualmente, modificar y aprobar los presupuestos de gasto e ingreso públicos. En ambos casos, el pronóstico de variables económicas es útil y convendría que fuese un esfuerzo sistemático y extendido en la investigación legislativa.

## 1. Estado del arte sobre el uso de modelos para el pronóstico de variables económicas

Los pronósticos de variables económicas tienen sus raíces en los esfuerzos históricos por comprender y predecir el comportamiento de la economía para que los responsables de la política pública y los agentes económicos puedan tomar decisiones informadas. Con el tiempo, estas predicciones se han refinado dados los avances científicos de la estadística, la economía y la tecnología, incluyendo el uso de modelos econométricos y técnicas de inteligencia artificial. En la actualidad, se utiliza una combinación de datos históricos, teorías económicas y algoritmos para proyectar tendencias futuras en variables como el Producto Interno Bruto (PIB), la inflación, las tasas de empleo, entre otros indicadores, lo que es crucial para la planeación de las políticas económicas, el comportamiento de las empresas en los mercados, la realización de inversiones, etc.

En los últimos años, en la estimación de pronósticos, se han tenido avances significativos con la incorporación de técnicas de *aprendizaje automático* o de máquina. Si bien esto supone una mejora para ciertas tareas, el aprendizaje automático también presenta desafíos específicos. Estos avances se destacan en diversos estudios que comparan los modelos tradicionales de series temporales con modernas arquitecturas de aprendizaje profundo y otras técnicas de aprendizaje automático.

Uno de los primeros modelos tradicionales es el ARIMA (acrónimo en inglés que significa Modelo Autorregresivo Integrado de Media Móvil, el cual se explica en la sección 2.1), que ha sido ampliamente utilizado para la predicción económica. ARIMA es efectivo para analizar series temporales estacionarias, descomponiendo las series en componentes autorregresivos, diferencias y medias móviles. Este modelo ajusta los datos a estructuras temporales específicas, permitiendo hacer pronósticos precisos a corto plazo. Su aplicación abarca desde predicciones de inflación hasta análisis de tendencias de desempleo y predicciones de tasas de interés.

En el estudio de Yamak et al. (2019) se encuentra que el modelo ARIMA superó a los modelos de redes LSTM (redes de memoria de corto y largo plazo) y GRU (redes de unidades recurrentes cerradas)<sup>1</sup> en la predicción de series temporales de una criptomoneda (Bitcoin). No obstante, en este estudio es preciso señalar que se limitaron a entrenar a los modelos en una ventana de tiempo corta. En cambio, en el trabajo de Siami et al. (2018) se observó que las redes LSTM fueron superiores al modelo ARIMA en la predicción de datos económicos mensuales. Cabe mencionar que se usaron registros mensuales desde 1985 hasta 2018, y aplicaron un *Rolling Forecast-Based Evaluation*, la

---

<sup>1</sup> Se revisan ambas arquitecturas de redes neuronales recurrentes en la sección 2.1, aunque también se puede consultar la sección 10 del capítulo 10 de Goodfellow (2016).

cual es una técnica que de forma secuencial va sumando más datos para entrenar<sup>2</sup>, y pronosticaron un único valor en cada caso. En resumen, el desempeño de los métodos tradicionales y el de los de aprendizaje de máquina tiende a variar dependiendo de la naturaleza de los datos que se desean modelar.

Otros estudios han explorado una gama más amplia de modelos de aprendizaje automático. En Ahmed et al. (2010) se compararon múltiples modelos en 1045 series temporales. Para cada caso obtuvieron el error mediante el error promedio del *cross validation*, también conocido como *backtesting* para el caso de las series de tiempo, la cual es una técnica utilizada para evaluar la precisión de los modelos predictivos que consiste en aplicar un modelo a distintos conjuntos de entrenamiento y prueba<sup>3</sup>, para al final promediar los errores obtenidos. En este estudio encontraron que el Perceptrón Multicapa y la Regresión por Procesos Gaussianos<sup>4</sup> ofrecían los mejores resultados, destacando la capacidad de estos modelos para manejar patrones complejos en los datos. Vale la pena señalar que en este estudio no se usaron modelos comunes como el ARIMA o las redes LSTM.

Recientemente, modelos como *Random Forest* y *Gradient Boosting*<sup>5</sup> han sido aplicados con éxito. Yoon (2020) utilizó estos modelos para predecir el crecimiento del PIB trimestral en Japón, obteniendo resultados prometedores al incluir poco más de 20 variables independientes. De forma similar, Muchisha et al. (2021) emplearon un enfoque de ensamble<sup>6</sup> mediante una regresión Lasso que combinaba *Random Forest* y redes neuronales, entre otros modelos, para predecir el crecimiento del PIB de Indonesia, destacando la eficacia de combinar múltiples técnicas de aprendizaje automático.

---

<sup>2</sup> En el aprendizaje automático el concepto de *entrenar* hace referencia a una serie de pasos estipulados en un algoritmo para que una computadora pueda aprender los patrones subyacentes de los datos con el fin de realizar un pronóstico hacia el futuro. Para más información consulte la introducción de Bishop (2006).

<sup>3</sup> El concepto de un conjunto de entrenamiento y prueba es habitual en el aprendizaje automático. Siguiendo con el concepto de *entrenar*; el conjunto de entrenamiento es aquel que se le da como entrada a un algoritmo para que la computadora aprenda los patrones subyacentes de los datos. El conjunto de prueba es aquel que contiene los valores reales que se espera que la computadora sea capaz de predecir. Por ejemplo, si se tienen datos mensuales del PIB de todo 2023 se puede *entrenar* un modelo con los datos de enero a octubre (siendo este el conjunto de entrenamiento) y hacer la predicción del PIB para noviembre y diciembre. Después se puede obtener el error (por ejemplo, la suma de las diferencias absolutas) entre esta predicción y los valores reales del PIB en noviembre y diciembre (el cual corresponde al conjunto de prueba).

<sup>4</sup> Tanto el modelo de Perceptrón Multicapa como la Regresión por Procesos Gaussianos se pueden consultar en la sección 5 del capítulo 16 y en la sección 2 del capítulo 15, respectivamente, de Murphy (2012).

<sup>5</sup> En la sección 2.1 se hablará de *Random Forest* y de una variante de *Boosting*, pero también se puede consultar en la sección 2.5 y 4, respectivamente, del capítulo 16 de Murphy (2012).

<sup>6</sup> Los métodos de ensamble de modelos buscan combinar las predicciones de múltiples modelos con el fin de mejorar las predicciones.

Se ha visto que la combinación de modelos en ensambles ha demostrado ser una estrategia efectiva. Esta técnica involucra el uso de múltiples modelos predictivos para luego combinar sus salidas, lo que generalmente resulta en predicciones más precisas y estables que las de cualquier modelo individual, además de reducir el riesgo de sobreajuste y aprovechar las fortalezas de diferentes métodos. Por ejemplo, Lu (2021) combinó el modelo ARIMAX (Modelo Autorregresivo Integrado de Media Móvil con Variables Exógenas) con una Red Neuronal de Retropropagación para mejorar las predicciones del PIB de Reino Unido. Su método consiste en hacer pronósticos preliminares con un modelo ARIMAX para valores del PIB, y después usar los errores de este modelo para entrenar una Red Neuronal de Retropropagación de modo que corrija las predicciones hechas por el primer modelo, ajustándose al componente no lineal de la serie de tiempo. Este método obtiene mejores resultados que usar alguno de estos dos modelos por sí mismo.

En conclusión, el estado del arte de los modelos para predecir variables económicas sigue en constante evolución. Actualmente se usan técnicas avanzadas de aprendizaje automático en conjunto con métodos tradicionales como ARIMA para mejorar la precisión de los pronósticos económicos. Aunque modelos como ARIMA siguen siendo relevantes por su robustez en datos estacionarios<sup>7</sup>, la incorporación de métodos más complejos como las redes neuronales y los ensambles de modelos han permitido abordar patrones más intrincados y dinámicos en los datos.

Los modelos de aprendizaje automático tienen la capacidad para captar relaciones complejas y no lineales, así como manejar grandes volúmenes y tipos de datos. No obstante, estos modelos también tienen sus limitaciones. Se tiene el problema de la interpretabilidad ya que los modelos más complejos, especialmente en aprendizaje profundo, pueden ser difíciles de interpretar, lo cual puede ser un problema en el contexto de las políticas económicas. Asimismo, se puede derivar en sobreajuste cuando el modelo es muy complejo para los datos que se tienen, ya que estos modelos dependen en gran medida de la calidad y la cantidad de los datos disponibles. Finalmente, las técnicas modernas de ensamblado, aunque también enfrentan desafíos como la interpretación y el sobreajuste, ofrecen nuevas posibilidades al incorporar una gran cantidad de variables y adaptarse mejor a la volatilidad económica. Los estudios citados ilustran que no existe un enfoque único superior para todos los escenarios, sino que la elección del modelo adecuado depende del contexto específico de los datos y de las necesidades de predicción.

---

<sup>7</sup> Una serie de tiempo se considera *estacionaria* cuando posee media y varianzas constantes a lo largo del tiempo.

## 2. Justificación de las tendencias de largo plazo como método de pronóstico

Para la medición de la ocupación formal e informal en México se tienen datos de diversas fuentes. Para medir la ocupación formal se cuenta con datos del Instituto Mexicano del Seguro Social (IMSS) sobre el número de trabajadores registrados, así como datos de la Encuesta Nacional de Ocupación y Empleo (ENOE) de personas asalariadas formales, y personas asalariadas con acceso al seguro social. Por otro lado, para la ocupación informal se tienen datos de la ENOE del número de personas asalariadas informales, personas ocupadas sin acceso al seguro social, y trabajadores por cuenta propia. Los datos de la ENOE empezaron a registrarse en 2005 y se miden con periodicidad trimestral. En cambio, los datos de ocupación formal del IMSS parten de julio de 1997 y se miden de forma mensual.

Con esto en cuenta, es importante averiguar qué modelo hace las mejores predicciones para cada una de estas series de tiempo. En particular, se quiere comparar el desempeño de la extrapolación con la tasa de crecimiento medio anual (TCMA<sup>8</sup>), de la tendencia de largo plazo obtenida con el filtro Hodrick-Prescott (HP) con respecto al desempeño de modelos estadísticos como el ARIMA y el Holt Winters, así como modelos de aprendizaje automático como el XGBoost y las redes neuronales LSTM y GRU. En esta sección se presenta una descripción detallada de los modelos utilizados, así como el método implementado para comparar las predicciones de cada uno, el cual consiste en la obtención de intervalos de confianza de los errores en los pronósticos.

A modo de resumen, en el desarrollo de esta sección se obtuvo lo que se muestra en la siguiente figura:

Figura 2. Intervalos de confianza para el Error Absoluto Medio de cada modelo

Modelos	Series de tiempo						Desempeño
	Datos del IMSS	Asalariados formales	Con acceso al Seguro Social	Trabajadores por cuenta propia	Asalariados informales	Sin acceso al Seguro Social	
TCMA	(182504, 220014)	(245570, 284954)	(226003, 262817)	(107236, 119579)	(141913, 158671)	(181985, 206043)	27
ARIMA	(143628, 157663)	(244870, 284954)	(200624, 262817)	(134830, 119579)	(157663, 158671)	(184481, 206043)	28

<sup>8</sup> La Tasa de Crecimiento Medio Anual se puede entender como la tasa que multiplica al valor inicial de una serie de tiempo para llegar exactamente al valor final de la misma. Esta misma tasa se puede usar con fines predictivos para extrapolar desde el valor final.

	178272)	280651)	229607)	148069)	175472)	208835)	
HoltWinters	(171133, 210401)	(539026, 653566)	(548373, 650371)	(254714, 289974)	(363166, 416275)	(333595, 396221)	10
LSTM	(254155, 290927)	(237882, 274704)	(275170, 311900)	(95593, 106877)	(157388, 175274)	(187084, 210453)	23
GRU	(220263, 248465)	(248906, 285003)	(233185, 264663)	(91571, 102959)	(168762, 187228)	(181722, 199275)	24
XGBoost	(157461, 189999)	(276000, 311851)	(281138, 320915)	(140114, 155366)	(217173, 238273)	(333741, 371552)	14
	Mejor desempeño: 6						
	Peor desempeño: 1						

Fuente: Elaboración propia.

En la Figura 2 se muestran los intervalos de confianza para el Error Absoluto Medio (EAM) de cada modelo; por ejemplo, para los datos del IMSS el modelo ARIMA tuvo el mejor desempeño, pues el error que genera se encuentra entre 143,628 y 178,272 personas más o personas menos del número real de personas (más adelante se muestra cómo se obtuvieron estos intervalos). En cambio, para la serie de Asalariados formales se ve que el mejor modelo fue el LSTM. Se asigna un número al desempeño de cada modelo en cada serie de tiempo, siendo el del mejor desempeño el que obtiene el número más alto. Al sumar el desempeño de cada modelo a través de todas las series de tiempo se obtuvo la columna de Desempeño. Se ve que el ARIMA obtuvo el mejor desempeño, seguido muy de cerca por la extrapolación mediante la TCMA.

Ahora bien, a continuación, se ofrecen los detalles técnicos de cada modelo, seguido de cómo y por qué se llega a los resultados de la Figura 2 y finalmente una serie de intentos para mejorar las predicciones hechas incorporando técnicas del estado del arte.

## 2.1 Resumen de los modelos

Cada uno de estos modelos se aplicó en el entorno de *Google Colab*, usando algunas de las librerías disponibles en *Python*. El código se puede encontrar en el repositorio de github de este trabajo<sup>9</sup>.

### **ARIMA:**

Sus siglas en inglés significan *Autoregressive (AR) Integrated (I) Moving Average (MA)*, componentes que comúnmente se representan con los parámetros  $p$ ,  $d$  y  $q$ , respectivamente. El componente  $p$  especifica el número de observaciones pasadas que el modelo considera para hacer la predicción. Por ejemplo, si  $p=6$  entonces para hacer una predicción el modelo hace la estimación mediante una combinación lineal de los 6 últimos valores de la serie de tiempo de interés.

Por otro lado, el componente MA indica cuántos errores en nuestras predicciones pasadas el modelo va a considerar. Por ejemplo, si  $q=1$  entonces el modelo primero hace un pronóstico preliminar para la última observación de la serie de tiempo mediante un modelo AR, obtiene el error  $e$  de esta predicción, y para el pronóstico usará este error  $e$  como otro parámetro que se añadirá a la combinación lineal del modelo AR. Es decir, el modelo ARMA le añade al modelo AR información sobre los errores que obtuvo en sus predicciones anteriores para que reajuste su predicción.

Finalmente, la I de ARIMA representa las diferencias de la serie de tiempo que se quiere aplicar. Por ejemplo, si se tiene una serie de tiempo [3,5,1], la primera diferencia de esta serie sería [5-3, 1-5] = [2, -4], lo que se puede interpretar como las distancias que hay entre datos consecutivos. Usualmente se aplica una diferencia ( $d=1$ ) para hacer la serie de tiempo estacionaria. Una vez hecha la predicción, se convierte esta predicción a las unidades originales de los datos.

Para poder construir varios modelos ARIMA se usa la función *auto\_arima* de la librería *pmdarima*, la cual de forma automática busca los mejores parámetros  $p$  y  $q$  mediante la optimización del *Akaike Information Criterion*, el cual busca un equilibrio entre el ajuste y la complejidad del modelo para una serie de tiempo dada. El parámetro  $d$  se fijó en 1 para volver las series de tiempo estacionarias.

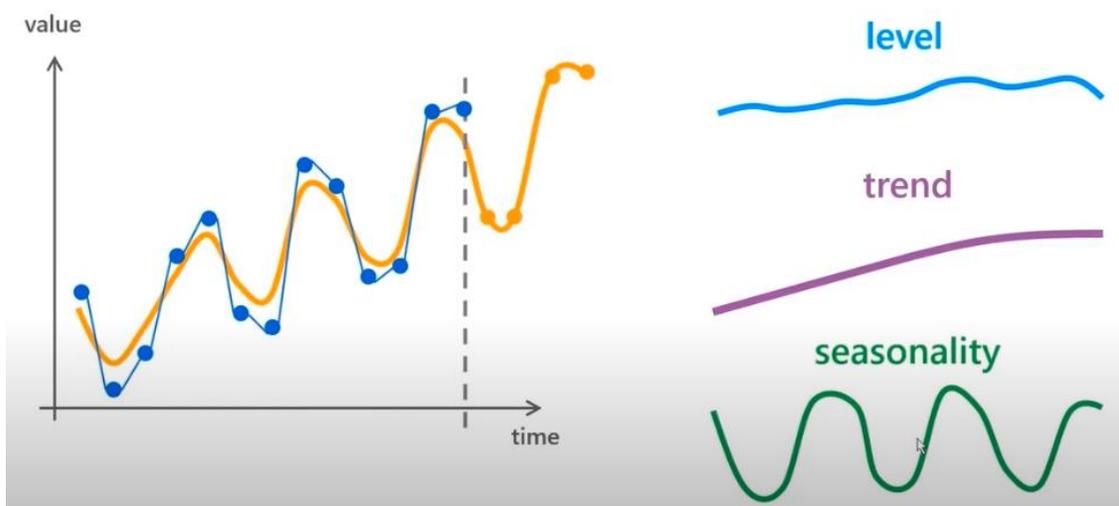
---

<sup>9</sup> Repositorio disponible en: <https://github.com/Slicea10/Analisis-de-metodos-y-modelos-de-pronostico-de-variables-de-empleo>

### Holt-Winters:

El método de Holt-Winters, o Suavizado Exponencial Triple, es una técnica de pronóstico de series temporales diseñada para capturar tendencias y estacionalidad. Consiste en tres componentes: nivel (que representa el valor promedio de la serie en ventanas de corto plazo), tendencia (que indica la dirección y tasa de cambio a largo plazo), y estacionalidad (que representa patrones repetitivos en un período específico). En la Figura 3 se ve un ejemplo de estos tres conceptos:

Figura 3: Ejemplificación del modelo Holt Winters.



Nota: La curva amarilla muestra la predicción del modelo.

Fuente: ArcGIS (2022).

El método implica actualizar estos componentes de manera iterativa. El nivel se actualiza en función de la observación actual, la tendencia se ajusta considerando el nivel actual y la tendencia pasada, y la estacionalidad se actualiza según los valores observados en un período estacional. Esta adaptabilidad permite al método acomodar cambios en los patrones subyacentes de la serie temporal.

Holt-Winters tiene dos variaciones principales: aditiva y multiplicativa. El modelo aditivo asume que los componentes se combinan linealmente, mientras que el modelo multiplicativo asume una combinación no lineal. La elección adecuada depende de la naturaleza de los datos. El pronóstico con Holt-Winters implica extrapolar valores futuros en función de los componentes (nivel, tendencia y estacionalidad) actualizados.

En este caso se usa la función *ExponentialSmoothing* de la librería *statsmodels.tsa.holtwinters*. Dentro de la función especificamos los siguientes parámetros: *seasonal='add'*, *seasonal\_periods=4*, *trend='add'*. Si bien *seasonal\_periods* se especifica como 12 para los datos del IMSS, para los datos de la ENOE se obtiene un mejor rendimiento con 4 ya que los datos originales tenían periodicidad trimestral. Como se ve, se opta por una estacionalidad y una tendencia aditiva. La estacionalidad aditiva significa que el componente estacional se suma al nivel o tendencia, mientras que una tendencia aditiva significa que el componente de tendencia se suma al nivel. Esto se determinó viendo que ninguna gráfica exhibió un comportamiento de crecimiento exponencial en su nivel o estacionalidad.

### **Redes LSTM:**

Las redes LSTM (*Long Short-Term Memory*) son un tipo de red neuronal recurrente diseñadas para capturar dependencias y patrones a largo plazo en datos secuenciales, lo que las hace particularmente efectivas para la predicción de series temporales. La intuición detrás de las LSTM radica en su capacidad para recordar y olvidar información selectivamente a lo largo de secuencias extensas.

En esencia, una celda LSTM está diseñada para mantener un estado de memoria que puede actualizarse o modificarse a medida que llega nueva información. Los componentes clave de una LSTM incluyen la celda de memoria, la compuerta de entrada, la compuerta de olvido y la compuerta de salida. A continuación, se da una explicación intuitiva sin entrar en tecnicismos:

La celda de memoria actúa como una cinta transportadora que se extiende a lo largo del tiempo, permitiendo que la información persista. Puede almacenar información relevante durante un periodo prolongado, capturando dependencias a largo plazo. La compuerta de entrada controla el flujo de nueva información hacia la celda de memoria, decide qué partes de los datos entrantes son importantes para almacenar en la celda de memoria. Por otro lado, la compuerta de olvido determina qué información en la celda de memoria debe descartarse, borra selectivamente información irrelevante u obsoleta, evitando que el modelo se vea abrumado por detalles innecesarios. Finalmente, la compuerta de salida decide qué información de la celda de memoria se utilizará para las predicciones.

En este caso, en la capa oculta de la arquitectura tenemos una capa de 16 celdas LSTM, seguida de otra de 8 celdas. Es una práctica común en redes neuronales probar diferentes arquitecturas, empezando por una simple y de ahí hacer modificaciones. La arquitectura propuesta es relativamente simple, haciendo que el tiempo de entrenamiento no sea tan prolongado. A continuación, se muestran las especificaciones de la red propuesta, usando la librería *keras*:

Figura 4: Arquitectura de la red LSTM.

```
model = Sequential()
model.add(LSTM(16, activation='relu', return_sequences=True, input_shape=(sequenceLength, 1)))
model.add(LSTM(8, activation='relu', return_sequences=False))
model.add(Dense(12))
model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(xTrain, yTrain, epochs=10, verbose=0)
```

Fuente: Elaboración propia.

Se usa un *sequenceLength* = 24, es decir, el modelo se entrenó mirando muchos ejemplos de 24 valores consecutivos en la serie de tiempo e intentando predecir los siguientes 12 valores. Se eligió este tamaño debido a que permite generar muchas secuencias de entrenamiento a la vez que le permite al modelo contar con dos años de datos para predecir el año siguiente. A cada una de estas secuencias se le conoce como un *batch*. Para cada serie de tiempo contamos con alrededor de 100 batches en el conjunto de entrenamiento. Fijamos *epochs*=10 para que el modelo recorra 10 veces todos los batches y así vaya actualizando sus pesos en repetidas iteraciones de aprendizaje.

### Redes GRU:

Las redes GRU (*Gated Recurrent Unit*) son una arquitectura de red neuronal recurrente (RNN) similar a las redes LSTM. Las GRUs buscan capturar dependencias a largo plazo en los datos siendo computacionalmente más eficientes que las LSTMs.

La característica clave de las GRUs es el uso de puertas que controlan el flujo de información dentro de la red. Estas puertas incluyen una puerta de actualización y una puerta de reinicio. La puerta de actualización decide cuánta información pasada debe ser llevada adelante, mientras que la puerta de reinicio determina cuánta información pasada olvidar, ambas basándose en la función sigmoide aplicada a una combinación lineal de los datos de entrada actuales y el estado anterior de la red. La presencia de estas puertas permite a las GRUs capturar de manera adaptativa información relevante a lo largo de diferentes escalas temporales.

A diferencia de las LSTMs, las GRUs tienen una arquitectura más simple con menos parámetros ya que no cuenta con el estado de celda ni con una compuerta de salida, lo que las hace computacionalmente más eficientes y rápidas de entrenar. Esta simplicidad a menudo hace que las GRUs sean adecuadas para tareas donde es crucial encontrar un equilibrio entre la complejidad del modelo y el rendimiento. Para la arquitectura, se usa una similar a la que utilizada para las redes LSTM:

Figura 5: Arquitectura de la red GRU.

```
model = Sequential()
model.add(GRU(16, activation='relu', return_sequences=True, input_shape=(sequenceLength, 1)))
model.add(GRU(8, activation='relu', return_sequences=False))
model.add(Dense(12))
model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(xTrain, yTrain, epochs=10, verbose=0)
```

Fuente: Elaboración propia.

### **XGBoost:**

El modelo '*eXtreme Gradient Boosting*', más conocido por sus siglas *XGBoost*, es un algoritmo de aprendizaje automático que destaca por su eficiencia y excepcional rendimiento predictivo. Este algoritmo es una evolución avanzada de los métodos tradicionales de *boosting* y ha sido diseñado específicamente para mejorar la velocidad y la precisión del aprendizaje automático. La esencia de *XGBoost* radica en su capacidad para construir un modelo predictivo robusto mediante la combinación de las predicciones de múltiples 'modelos débiles', que suelen ser árboles de decisión. Estos árboles se construyen de forma iterativa, y cada nuevo árbol se enfoca en corregir los errores cometidos por los árboles anteriores, prestando especial atención a las instancias que fueron clasificadas incorrectamente en iteraciones previas.

El proceso denominado '*boosting* con gradiente' hace referencia a esta técnica de optimización, en la que cada nuevo árbol se entrena para minimizar los errores residuales que el conjunto anterior no pudo resolver. *XGBoost* mejora significativamente sobre otros métodos de *boosting* gracias a varias técnicas clave:

- Regularización: *XGBoost* incorpora términos de regularización (L1 y L2), los cuales ayudan a prevenir el sobreajuste al penalizar modelos más complejos. L1 (Lasso) agrega una penalización basada en la suma de los valores absolutos de los coeficientes, promoviendo la esparsidad. L2 (Ridge) agrega una penalización basada en la suma de los cuadrados de los coeficientes, distribuyendo la penalización más uniformemente.
- Poda de árboles: A diferencia del método de parada temprana usado en otros algoritmos, *XGBoost* ofrece un control más refinado para el sobreajuste mediante la poda de los árboles una vez que dejan de aportar ganancias significativas.
- Manejo de datos faltantes: El algoritmo puede manejar internamente datos faltantes, asignando a estos una dirección en cada bifurcación de los árboles, dependiendo de lo que maximice la reducción de la pérdida.

- Optimización del rendimiento: Utiliza técnicas como el almacenamiento de estructuras de datos en forma de bloques para optimizar los recursos computacionales y manejar la escalabilidad a lo largo de múltiples núcleos.

Además, *XGBoost* ofrece una flexibilidad considerable al permitir la personalización de la función objetivo y la evaluación de criterios, haciendo posible adaptar el algoritmo a una amplia gama de necesidades específicas. Gracias a estas capacidades, *XGBoost* ha ganado una enorme popularidad y se ha convertido en una herramienta esencial en muchas competencias de ciencia de datos, como las organizadas en plataformas como *Kaggle*.

Similar a las redes, se utilizó un *sequenceLength=24*. Una vez que se entrenó el modelo, se busca predecir un solo ejemplo en lugar de directamente 12 como en las redes que se han implementado. Esto se debe a que *XGBoost* es un modelo que se usa para regresión y clasificación, no tanto para series de tiempo, por lo que no puede predecir múltiples valores a la vez. Una vez hecha esta predicción, se usa el modelo para hacer otra predicción, ahora enseñándole los datos de la secuencia original solo que omitiendo el primero y añadiendo al final el predicho. Se repite esto 12 veces para obtener la predicción de los 12 meses que se quieren.

Para este modelo no se especificó ningún parámetro pues se obtiene un buen rendimiento con las configuraciones predefinidas. Se utiliza la función *xgb* de la librería *xgboost* de *Python*.

## **2.2 Generación de intervalos de confianza para la evaluación de modelos**

Como se explicó anteriormente, para la predicción de la ocupación formal e informal en México, se evaluó la capacidad predictiva de una serie de modelos mediante la generación de intervalos de confianza para los errores. En particular, para las seis series de tiempo de ocupación formal e informal se evaluaron los modelos ARIMA, Holt-Winters, *XGBoost*, Redes LSTM, Redes GRU, y extrapolación mediante la tasa de crecimiento medio anual (TCMA).

Para los datos del IMSS se consideraron los datos mensuales desestacionalizados. En cambio, para los datos de la ENOE se hizo una *interpolación lineal* de los datos trimestrales. Esto último se hizo ya que mejoró considerablemente las predicciones para todos los modelos en comparación con los datos mensuales interpolados por el INEGI. Esto se debió probablemente a que la variabilidad incrustada siguió un comportamiento aleatorio que los modelos no lograron captar. Por ello, y ante la falta de información respecto a la metodología de la extrapolación hecha por el INEGI, se asumió una

variabilidad constante entre los trimestres. Para todas las series de tiempo se trabajó con los datos con periodicidad mensual para contar con suficientes datos para poder hacer una evaluación robusta del desempeño de los modelos, la cual se explica a continuación.

Todos los modelos predictivos para series de tiempo generan algún error a la hora de predecir. En este trabajo se calcula el error como el Error Absoluto Medio, el cual se calcula como:

Figura 6: Error absoluto medio.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Fuente: Hodson (2022).

Donde  $y_i$  corresponde al valor real,  $\hat{y}_i$  al valor predicho, y  $n$  al número de predicciones hechas (en este caso se puede ver como el número de meses en adelante que predecimos). Entonces, si se predicen 12 meses y se obtiene un error de 9, significa que los valores predichos de la serie de tiempo se alejan en promedio 9 unidades del valor real. Como nos interesa predecir desde enero del 2024 hasta diciembre de este año, se mantiene una  $n = 12$  en nuestro análisis.

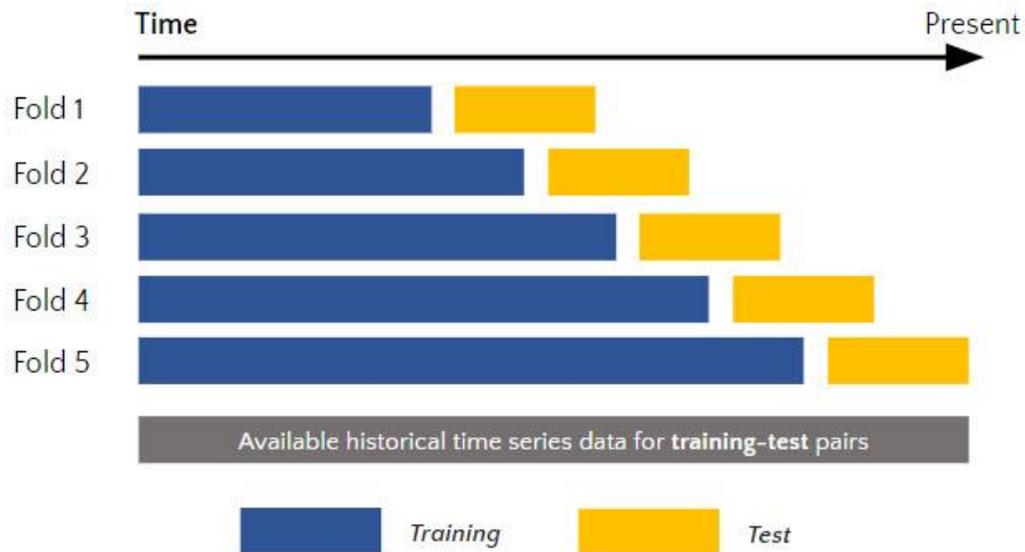
Para obtener el desempeño de un modelo de regresión o de series de tiempo, se suele fijar en qué tan pequeño es el error de sus predicciones. Dado que cuando se empezaron a evaluar los diferentes modelos no se contaba con los datos mensuales del 2024 para ninguna de las series de tiempo con las que se trabajó, no era posible calcular el error real de los modelos al realizar las predicciones para los 12 meses de este periodo de tiempo. Sin embargo, se usó cada modelo para predecir 12 meses para periodos de los que sí se tenían datos, y así poder evaluar su desempeño. Por ejemplo, prediciendo los meses del periodo de octubre del 2022 a octubre del 2023 y obteniendo el error de la predicción.

Esta estrategia tiene la ventaja de que se puede implementar fácilmente y se puede obtener una aproximación rápida sobre qué modelo es más confiable para la predicción que se quiere hacer. No obstante, es probable que el modelo con el error más pequeño en esta evaluación no sea el que mejores predicciones genere para el nuevo periodo que se quiere predecir. Por lo tanto, para cada modelo se opta por generar muchas predicciones para diferentes periodos de tiempo y así obtener el promedio de los errores.

Esta técnica es conocida como *backtesting* (o pruebas retrospectivas). Para una misma serie de tiempo la idea es entrenar muchos modelos para diferentes ventanas de tiempo para poder promediar los errores que se obtengan y así obtener una mejor aproximación

de la capacidad predictiva del modelo. En la siguiente figura se ilustra lo que se hizo:

Figura 7: Ejemplificación del backtesting en series de tiempo.



Nota: En este trabajo se optó por esta variante del backtesting por la cantidad limitada de datos. Sin embargo, se suelen tomar conjuntos de entrenamiento del mismo tamaño.

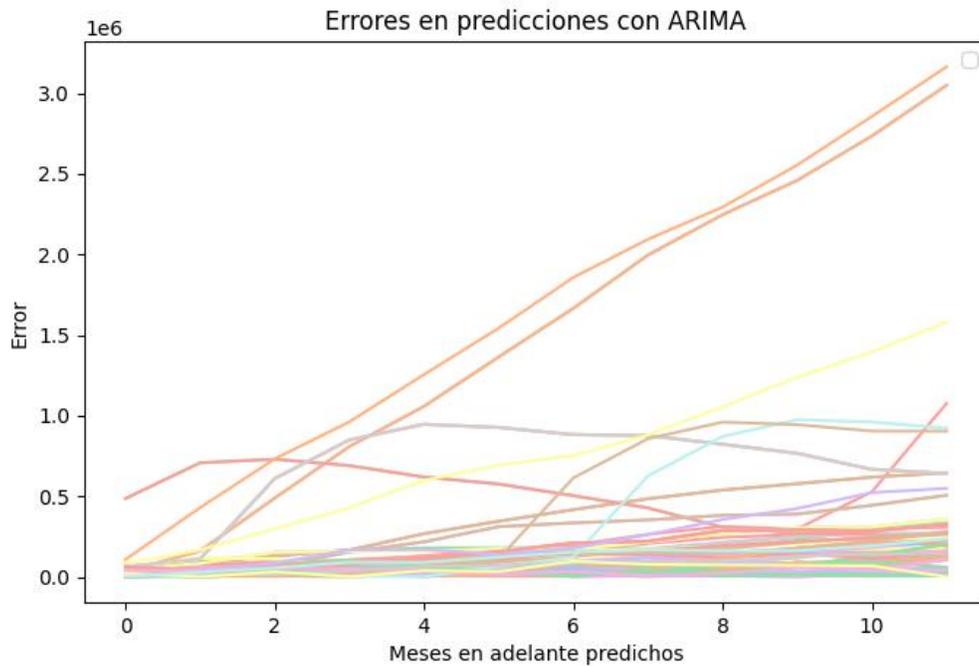
Fuente: Coupa (2023).

En este ejemplo, para una misma serie de tiempo, se crean 5 modelos. Cada uno se entrena y hace la predicción para diferentes periodos de tiempo. Al final se obtiene los errores de cada una de estas predicciones. Si se predicen 12 días y se evalúan 5 modelos, entonces se obtiene una lista de  $12 \times 5 = 60$  errores.

Se define una cantidad mínima de datos de entrenamiento de 120, para contar con 10 veces más datos que los que se quieren predecir. Se cuenta con datos de aproximadamente 220 meses para los datos de la ENOE y de 320 para los del IMSS. En vez de los cinco modelos que se entrenaron en la figura de arriba, se entrenan de 55 a 80 modelos (pues para algunas series de tiempo se contaba con más datos), para cada uno de los algoritmos y para cada serie de tiempo.

Por ejemplo, para la serie de tiempo de datos del IMSS y para el modelo ARIMA, se particionó los datos del 0 al 320 en 80 conjuntos de entrenamiento-prueba con un mínimo de 120 datos de entrenamiento y se obtuvieron los errores para las predicciones de todos estos modelos. Como en cada modelo se predicen 15 meses en este ejemplo, en total se obtienen  $80 \times 15 = 1200$  errores. En la siguiente gráfica cada línea representa un modelo, el eje X los meses predichos y el eje Y el error:

Figura 8: Errores de predicciones de modelos ARIMA.

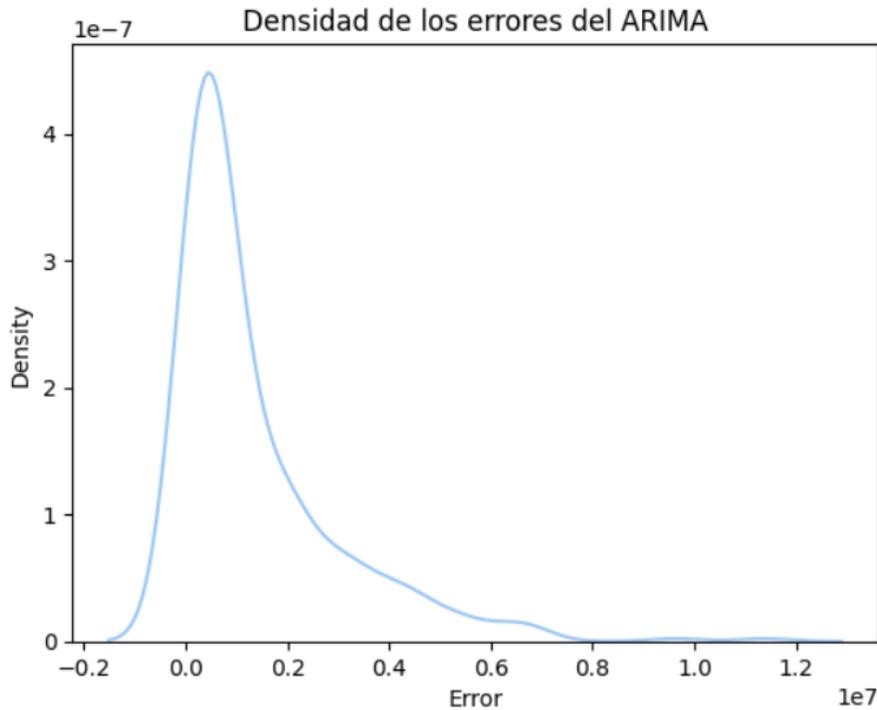


Nota: Se crearon 80 modelos para diferentes periodos de tiempo, cada uno haciendo un pronóstico de 15 meses.

Fuente: Elaboración propia.

Al graficar la *función de densidad* de estos 1200 errores, se obtiene una gráfica como la de la siguiente figura:

Figura 9: Función de densidad de errores de modelos ARIMA.



Fuente: Elaboración propia.

Es decir, la gráfica muestra que la probabilidad de que el error sea de aproximadamente  $0.15 \times 10^6$  es muy alta, mientras que la probabilidad de que sea un número menor o mayor es más pequeña. En un inicio se optó por tomar el promedio de los errores como una aproximación del error real de cada modelo. No obstante, al hacer el ejercicio en repetidas ocasiones se observó que esta aproximación era muy inestable, en parte porque la distribución de los errores, como se ve en la gráfica de la densidad, no sigue una distribución normal. Por lo tanto, una vez obtenidos los errores, de un modelo y para una serie de tiempo, se obtiene un muestreo *bootstrap*.

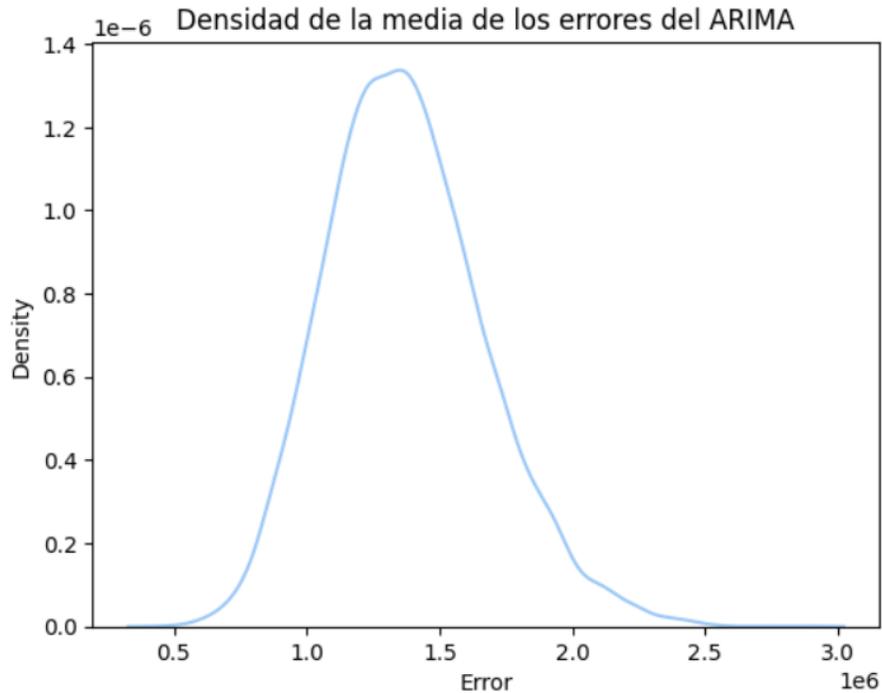
El muestreo *bootstrap* consiste en obtener una gran cantidad de muestras aleatorias con reemplazo de un conjunto de datos. Una vez obtenidas las muestras, cada una se promedia para así, por el *Teorema Central del Límite*<sup>10</sup>, obtener una distribución normal de estos promedios. En este caso, de los errores obtenidos se tienen 10000 muestras de tamaño 1200 (el número de errores/datos disponibles). Después de obtener el promedio

---

<sup>10</sup> El Teorema Central del Límite dice que, sujeto a ciertas condiciones, la suma de un conjunto de variables aleatorias, la cual es en sí misma una variable aleatoria, tiene una distribución que se asemeja cada vez más a una distribución normal a medida que el conjunto de variables aleatorias sea más grande. Su formulación se puede consultar en la sección 6.3 del capítulo 2 de Murphy (2012).

de cada una de estas muestras, para cada caso, se obtiene una distribución como la de la Figura 10.

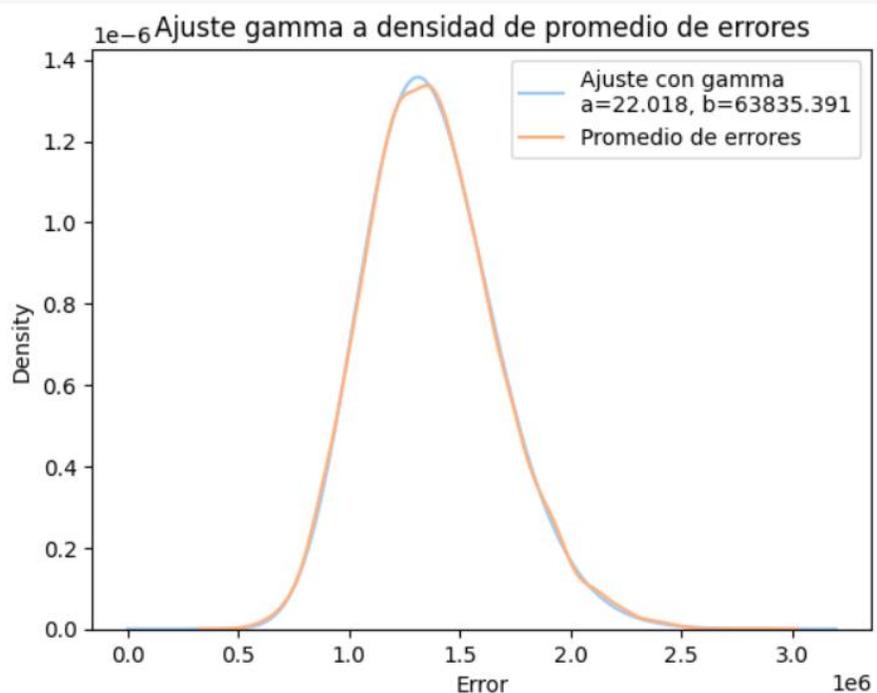
Figura 10: Función de densidad del promedio de errores de modelos ARIMA.



Fuente: Elaboración propia.

Aunque a simple vista parece que al hacer el muestreo *bootstrap* se obtuvo una distribución normal de los promedios de los errores, al aplicar una prueba de normalidad *Shapiro–Wilk*, se obtiene que esta distribución sigue sin ser normal pues se obtuvo un p-valor de 0.00046 lo cual es menor al nivel de significancia definido como  $\alpha = 0.05$ . No obstante, resultó ser más manejable que la de los errores brutos en el sentido de que luce menos asimétrica y por lo tanto se pueden generar intervalos de confianza más interpretables. Se opta por ajustar una distribución gamma en todos los casos. Haciendo el ajuste, se obtiene una aproximación muy parecida de la distribución del promedio de los errores, vista en la Figura 11.

Figura 11: Ajuste gamma a densidad del promedio de errores.

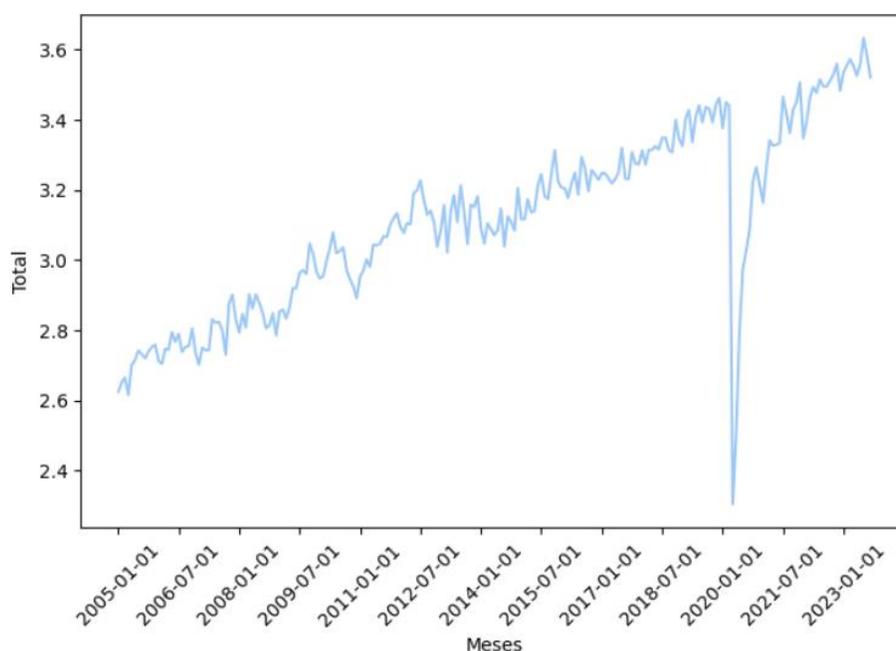


Fuente: Elaboración propia.

Para cada serie de tiempo y para cada modelo se aplica una prueba *Kolmogorov-Smirnov* para el ajuste gamma, obteniendo que, para todos los casos, se hizo un ajuste apropiado pues se obtuvieron p-valores mayores al nivel de significancia definido como  $\alpha= 0.05$ .

Por otro lado, cabe mencionar que las series de tiempo de ocupación informal tenían un aspecto como el que se ve en la siguiente figura:

Figura 12: Outlier en los datos de ocupación informal.



Fuente: Elaboración propia.

Incluso con técnicas como el suavizado exponencial o de media móvil no fue posible aminorar el impacto que los datos atípicos de la pandemia tuvieron en nuestros modelos. Dado que no se puede esperar que un modelo pueda predecir, con base en una variable, una crisis de esta magnitud, y como la caída en el total de personas es lo suficientemente grande como para no poder incorporarse como se hizo con las series de ocupación formal debido a que sesga las predicciones de todos los modelos, se opta por evaluar los modelos propuestos con datos hasta 2020/01 para estos casos. Esto no representará un problema a la hora de querer hacer predicciones para 2024 pues los modelos pueden tomar datos después del periodo atípico para hacer el pronóstico.

Una vez obtenidos los errores de cada modelo para cada serie de tiempo, y después de hacer el muestreo *bootstrap* en cada caso, se procede a obtener los intervalos de confianza para el promedio de los errores con distribución gamma. Se obtienen los resultados de la Figura 2.

Se observa que no hay un modelo que haya tenido el mejor desempeño a través de todas las series de tiempo. Por la simplicidad de la extrapolación con la TCMA se esperaba que los demás modelos tuvieran un mejor desempeño. No obstante, dada la naturaleza de los datos, se ve que esta técnica tiene un muy buen desempeño a través de todas las series

de tiempo de ocupación formal e informal en contraste con modelos más robustos y quizás menos interpretables.

### **2.3 Mejoras a los modelos mediante selección de variables y creación de ensambles**

Por último, se intenta aprovechar la capacidad de los modelos de aprendizaje automático para incorporar muchas variables dependientes en un intento por mejorar los pronósticos. Únicamente se consideran los datos del IMSS. Se obtiene el intervalo de confianza del error de cada modelo usando el método expuesto en la sección anterior. Se consideran 16 variables obtenidas del Banco de Información Económica del INEGI:

1. Ocupación.
2. Actividad industrial.
3. Actividad turística.
4. Actividad minera.
5. Formación bruta de capital fijo.
6. PIB.
7. Ahorro.
8. Indicador global de la tasa económica.
9. Oferta global.
10. Demanda global.
11. Consumo privado.
12. Formación bruta de capital físico.
13. Índice de precios al consumidor.
14. Índice de precios al productor.
15. Balanza comercial - exportaciones.
16. Balanza comercial - importaciones.

La ruta exacta para encontrar cada variable dentro del Banco de Información Económica del INEGI se encuentra en el Anexo A. Para todos los casos se usaron los datos desestacionalizados pues no se obtuvo ninguna mejora en la predicción con los datos originales.

Dado que para algunas variables sólo se contaban con los datos trimestrales se decidió hacer una interpolación lineal<sup>11</sup> para pasarlas a la frecuencia mensual. Asimismo, para las variables que empezaban después de la fecha del primer registro de los datos de

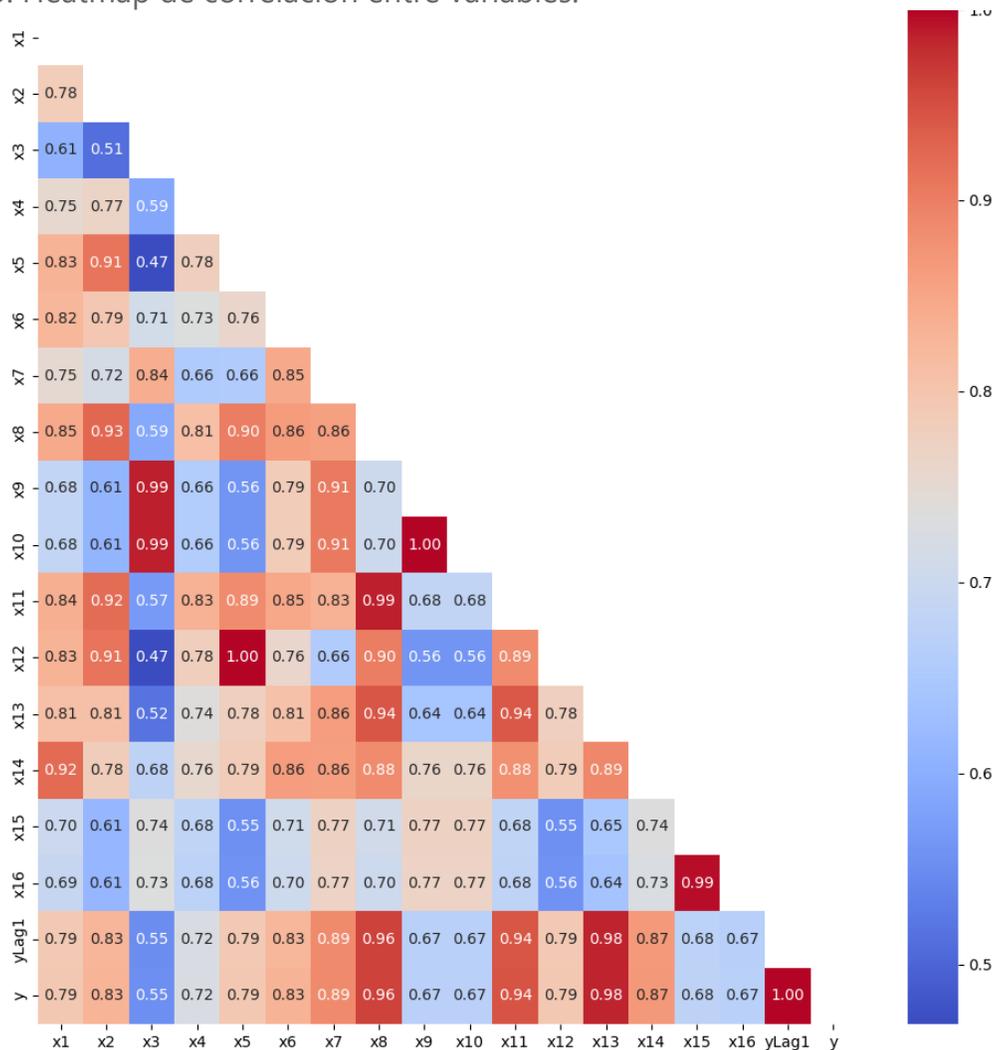
---

<sup>11</sup> En la interpolación lineal se traza una línea recta entre dos datos conocidos para obtener un tercero desconocido que se encuentre sobre esta línea.

ocupación del IMSS, la cual es la variable dependiente, o para aquellas que terminaban antes de la última fecha de registro de esta variable, se decidió llenar los datos faltantes con -1. Esto es una práctica común sobre todo para modelos como los que se aplicaron: redes neuronales LSTM, redes GRU, y XGBoost, debido a que pueden aprender a ignorar los valores -1 que en este caso corresponden a valores nulos.

Para intentar mejorar las predicciones se construye un mapa de correlación entre las variables, así como visualizaciones de estas para descartar aquellas que poseen una alta correlación entre ellas. Esto se aprecia en las figuras 13 y 14.

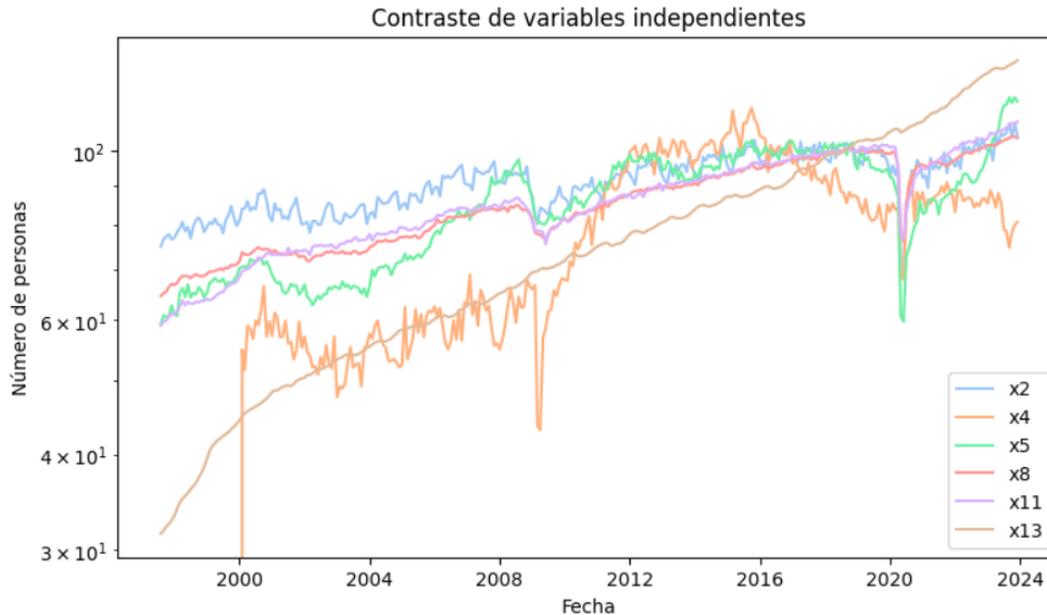
Figura 13: Heatmap de correlación entre variables.



Nota: Un 1 representa una correlación perfecta, una correlación de 0 nos dice que no se puede determinar ningún sentido de covarianza, y un -1 representa una correlación negativa perfecta. La enumeración de las variables corresponde al de las 16 variables elegidas del Banco de Información Económica.

Fuente: Elaboración propia.

Figura 14: Contraste de gráficas de algunas variables independientes.



Fuente: Elaboración propia.

Se intenta quitar todas las variables que tuvieran correlación con otras, aquellas con una gráfica similar y aquellas con muchos valores nulos, creando diferentes conjuntos de variables independientes. Por ejemplo, en la Figura 14 se muestra uno de los conjuntos de variables que se eligió, ya que se ve que en sus gráficas no muestran una correlación tan marcada. No obstante, ninguna reducción del número de variables contribuyó a que los modelos tuvieran un mejor desempeño.

Como se menciona anteriormente, se aplican redes neuronales LSTM, redes GRU, y XGBoost. El modelo GRU tuvo el mejor desempeño, generando un intervalo de confianza del error de (225865.55291536974, 262815.32947021053), lo cual no supuso una mejora de lo que se obtuvo anteriormente sin incluir variables independientes.

De igual manera, se intenta crear modelos de ensamble mediante un promedio de las predicciones de los modelos que tuvieron el mejor desempeño para los datos del IMSS. Se obtienen los siguientes intervalos para los errores:

Ensamble TCMA-ARIMA:  
(188444.61774184505, 227069.0692692241)  
Ensamble Holt-XGBoost:

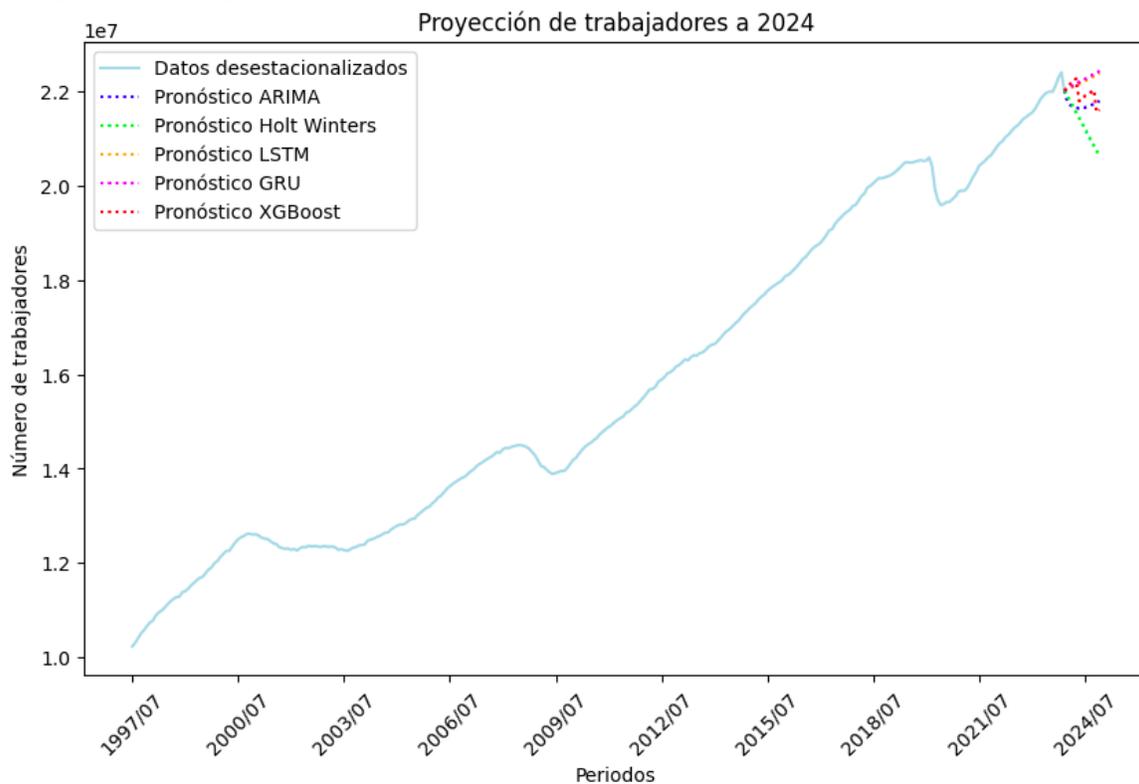
(170838.68535836056, 212453.83573792587)

Lo cual tampoco supuso una mejora significativa respecto a lo que se tenía anteriormente.

### 3. Pronóstico del empleo formal al cierre de 2024 con los distintos modelos

En esta sección se aplican los modelos ARIMA, Holt Winters, LSTM, GRU, y XGBoost para hacer el pronóstico del 2024 para los datos mensuales desestacionalizados de ocupación formal del IMSS actualizados hasta diciembre del 2023. Dado que no se obtienen mejoras al incorporar más variables, se opta por aplicar los modelos univariados de la sección 4.2. Se obtiene lo siguiente:

Figura 15: Proyección de empleo formal de 2024.

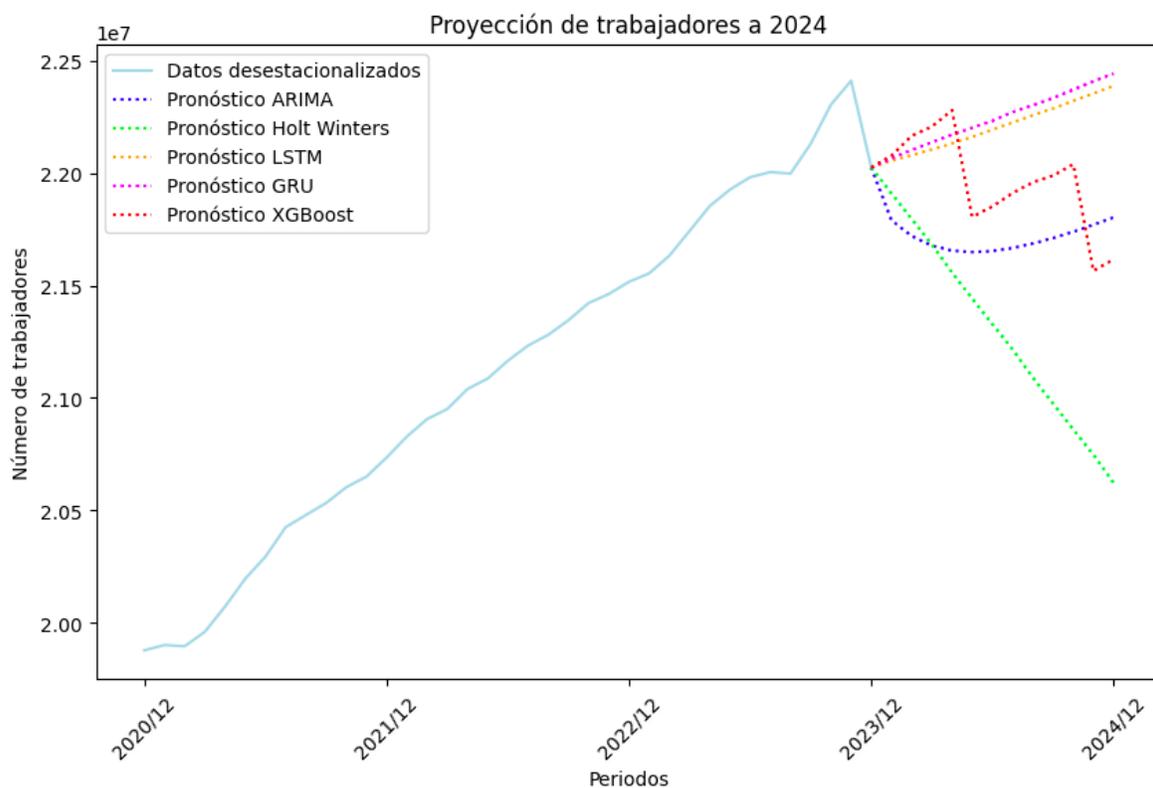


Nota: Se basa en los datos del IMSS y se aplica cinco modelos diferentes para la predicción.

Fuente: Elaboración propia.

Haciendo un acercamiento gráfico a los pronósticos:

Figura 16: Proyección de empleo formal de 2024, con datos históricos desde el 2020/12.



Fuente: Elaboración propia.

Se ve que los pronósticos de las redes LSTM y GRU son prácticamente idénticas y más similares a lo que se obtendría usando la TCMA. Por otro lado, la predicción de Holt Winters luce un tanto pesimista, mientras que la del XGBoost sugiere una tendencia escalonada a la baja. Finalmente, vemos que el ARIMA parece haber pronosticado una caída seguida de un alza como en el caso del 2009 y del 2020. En la Figura 17 se muestran los valores puntuales del pronóstico de los modelos para el periodo 2024/12.

Figura 17: Pronóstico de los modelos para diciembre del 2024.

Pronósticos para 2024/12	
Modelo	Pronóstico
ARIMA	21,802,269.66
Holt Winters	20,623,514.31
LSTM	22,452,806.27
GRU	22,397,719.13
XGBoost	21,615,589.55

Fuente: Elaboración propia.

## Conclusiones

En conclusión, en este trabajo se hizo un análisis sobre las diferentes metodologías para el pronóstico del empleo formal e informal en México. Se destaca la relevancia de dichos pronósticos, ya que resultan cruciales para el diseño de políticas públicas, para el ajuste de programas y reformas del mercado laboral, para mejorar la planificación fiscal, beneficiando el desarrollo económico y la estabilidad social del país. Se introduce el método de extrapolación mediante la tasa de crecimiento medio anual como un modelo lineal simple que permite proyectar la tendencia de una serie de tiempo para el pronóstico de periodos futuros. Si bien su uso se encontraba justificado por la simpleza y la alta interpretabilidad de sus pronósticos, hacía falta contrastar su eficacia con modelos más robustos para descartar alguna posible falta de precisión en las predicciones de este modelo.

Para atender esta inquietud, se exploró dentro de la literatura especializada el uso de modelos para la predicción de variables económicas y se vio el potencial que tienen los modelos de aprendizaje automático. No obstante, también se constató que actualmente se suelen usar en conjunto con modelos estadísticos tradicionales para la predicción de variables económicas, de modo que se aprovechen las fortalezas de cada uno.

Posteriormente, se dio una descripción tanto de los modelos estadísticos como de los de aprendizaje automático que se decidió usar para la predicción de la ocupación. Seguidamente se describió la metodología usada para comparar el desempeño de los

modelos mediante la creación de intervalos de confianza para los errores. Si bien la metodología propuesta hace uso de técnicas comunes en el estado del arte como el *backtesting*, sumándole a este la creación de intervalos de confianza, se pudieron hacer más esfuerzos para reducir la varianza de los errores de los modelos, como la incorporación de regularizaciones a los modelos o separar los errores de forma mensual, aunque esto último no se hizo por motivos de simplicidad de los resultados pues se deseaba evaluar a los modelos de forma anual, agregando su desempeño a través de los doce meses.

Finalmente, se intentó mejorar los pronósticos de nuestros modelos mediante la creación de ensambles de aquellos con el mejor desempeño. Asimismo, se hizo una selección de variables independientes para que los modelos de aprendizaje automático contaran con más información para el pronóstico. No obstante, aunque en todos los casos se obtuvieron buenos resultados, ninguno de estos modelos supuso una mejora respecto a los modelos univariados que se tenían.

En resumen, al ver la comparación de los modelos en la Figura 2, se observa que la extrapolación mediante la tasa de crecimiento medio anual resultó ser un método con una buena capacidad predictiva que es comparable con los modelos de aprendizaje automático, incluso después de buscar mejorarlos mediante la incorporación de más variables o técnicas de ensamble, como se hace en el estado del arte para la predicción de variables económicas. Además, al ser un modelo más simple e interpretable, su uso para la predicción de la ocupación formal e informal en México resultó estar completamente justificado.

## **Anexos:**

**Anexo A:** Detalles sobre rutas de búsqueda dentro del Banco de Información Económica para las 16 variables independientes.

Observaciones: La numeración corresponde al orden en que se muestran las variables en la sección 4.3.

1: Ocupación, empleo y remuneraciones > Población ocupada, subocupada y desocupada (resultados trimestrales de la ENOE, 15 años y más) > Valores absolutos > Nacional > Total > Población ocupada > Por sector de actividad económica > Total /a /f1

2: Indicadores económicos de coyuntura > Actividad industrial, base 2018 > Series desestacionalizadas y tendencia-ciclo > Total de la actividad industrial > Serie original corregida por efectos del calendario > Índice /f1

3: Indicadores económicos de coyuntura > Actividad turística, base 2018 > Series desestacionalizadas y tendencia-ciclo > Valor agregado bruto Nacional > Serie desestacionalizada > Índice /f1

4: Indicadores económicos de coyuntura > Industria minerometalúrgica > Series desestacionalizadas y tendencia-ciclo > Serie desestacionalizada > Índice /f1

5: Indicadores económicos de coyuntura > Formación bruta de capital fijo, base 2018 > Series desestacionalizadas y tendencia-ciclo > Total > Serie desestacionalizada > Índice /f1

6: Indicadores económicos de coyuntura > Producto interno bruto trimestral, base 2018 > Series desestacionalizadas y tendencia-ciclo > A precios de 2018 > Total > Serie desestacionalizada > Valores absolutos /f1

7: Indicadores económicos de coyuntura > Indicador trimestral del ahorro bruto, base 2018 > Series desestacionalizadas y tendencia-ciclo > Valores absolutos > Ahorro total > Serie desestacionalizada > Valores absolutos /f1

8: Indicadores económicos de coyuntura > Indicador global de la actividad económica, base 2018 > Series desestacionalizadas y tendencia-ciclo > Total > Serie desestacionalizada > Índice /f1

9: Indicadores económicos de coyuntura > Oferta y demanda global de bienes y servicios, base 2018 > Series desestacionalizadas y tendencia-ciclo > A precios constantes > Oferta > Total > Serie desestacionalizada > Valores absolutos /f1

10: Indicadores económicos de coyuntura > Oferta y demanda global de bienes y servicios, base 2018 > Series desestacionalizadas y tendencia-ciclo > A precios constantes > Demanda > Total > Serie desestacionalizada > Valores absolutos /f1

11: Indicadores económicos de coyuntura > Indicador mensual del consumo privado en el mercado interior, base 2018 > Series desestacionalizadas y tendencia-ciclo > Total > Serie desestacionalizada > Índice /f1

12: Indicadores económicos de coyuntura > Formación bruta de capital fijo, base 2018 > Series desestacionalizadas y tendencia-ciclo > Total > Serie desestacionalizada > Índice /f1

13: Indicadores económicos de coyuntura > Índices de precios > Índice nacional de precios al consumidor. Base segunda quincena de julio de 2018=100 > Mensual > Índice > Índice general /f1

14: Indicadores económicos de coyuntura > Índices de precios > Índice nacional de precios productor. Producción total según actividad económica. Base Julio 2019=100 (SCIAN 2013) > INPP con Petróleo y con Servicios /f1

15: Indicadores económicos de coyuntura > Balanza comercial de mercancías de México > Series desestacionalizadas y tendencia-ciclo > Exportaciones (FOB) > Total > Serie desestacionalizada > Valores absolutos /f1

16: Indicadores económicos de coyuntura > Balanza comercial de mercancías de México > Series desestacionalizadas y tendencia-ciclo > Importaciones (FOB) > Total > Total > Serie desestacionalizada > Valores absolutos /f1

## Referencias:

- (n.a). (2023). *Using backtesting for model selection and evaluation*. Coupa. Recuperado de: [https://help.llama.ai/release/platform/doc-center/demand\\_topics/dem\\_modeler\\_engine\\_backtesting.htm](https://help.llama.ai/release/platform/doc-center/demand_topics/dem_modeler_engine_backtesting.htm)
- Ahmed , Nesreen K., Amir F. Atiya , Neamat El Gayar & Hisham El-Shishiny (2010) An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Recuperado de: <https://sci-hub.se/10.1080/07474938.2010.481556>
- ArcGIS. (2022). *Time Series Forecasting*. Youtube. Recuperado de: <https://www.youtube.com/watch?v=gxoZ-vWUlh4&t=802s>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Guerrero, V. M. (2013). *Capacidad predictiva de los índices cíclicos compuestos para los puntos de giro de la economía mexicana*. Scielo. Recuperado de: [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-20452013000100002](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-20452013000100002)
- Hodrick, R. J. y E. C. Prescott (1997), "Postwar U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29 (1), pp. 1-16.
- Hodson, T. (2022). *Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not*. Geoscientific Model Development. Recuperado de: <https://gmd.copernicus.org/preprints/gmd-2022-64/gmd-2022-64.pdf>
- Lu, S. (2021). *Research on GDP Forecast Analysis Combining BP Neural Network and ARIMA Model*. Cardiff University Business School. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8604606/pdf/CIN2021-1026978.pdf>
- Martínez S., J. y Vargas F., A. (2023). *Evolución y perspectivas del empleo formal. Datos y reflexiones iniciales para la formulación de un balance sexenal*. Instituto Belisario Domínguez, Senado de la República. Recuperado de: <http://bibliodigitalibd.senado.gob.mx/bitstream/handle/123456789/6045/Reporte%20TE%20108%20Empleo%20formal.pdf?sequence=1&isAllowed=y>
- Muchisha, N., Tamara, N., Andriansyah., Soleh, A. (2021). *Nowcasting Indonesia's GDP Growth Using Machine Learning Algorithms*. IPB University. Recuperado de: <https://stat.ipb.ac.id/journals/index.php/ijsa/article/view/824/311>
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

- Siami, S., Tavakoli, N., Siami, A. (2018). *A Comparison of ARIMA and LSTM in Forecasting Time Series*. Recuperado de: <https://sci-hub.se/10.1109/icmla.2018.00227>
- Yamak, P., Yujian, L., P, Gadosey. (2019). *A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting*. Recuperado de: <https://sci-hub.se/10.1145/3377713.3377722>
- Yoon, J. (2020). *Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach*. Springer Science. Recuperado de: <https://sci-hub.se/10.1007/s10614-020-10054-w>