

N° 21-2019

Diciembre

INFORME ESPECIAL

**Nowcasting del PIB en Colombia
a partir de indicadores y
variables macro aplicando
inteligencia artificial**

Nowcasting del PIB en Colombia a partir de indicadores y variables macro aplicando inteligencia artificial

Diego Fernando Montañez Herrera

Grupo de Análisis de Coyuntura Económica. Universidad EAFIT

Abstract

El crecimiento del Producto Interno Bruto (PIB) es el principal indicador de desempeño económico de cualquier economía en el mundo, permite identificar los ingresos y la oferta agregada de un país en el corto, mediano y largo plazo, por consiguiente, el rumbo que puede tomar la actividad económica es fundamental para hacedores de política, autoridades económicas, agentes privados y el gobierno, no obstante, por su naturaleza macro, la publicación presenta rezagos de hasta más 4 semanas en Estados Unidos y Reino Unido; entre 6 y 7 en el área euro y de 50 días en el caso de Colombia a partir de la finalización de un trimestre. En consecuencia, es de vital importancia un diagnóstico de la economía en “tiempo real”, que en la literatura más moderna estos modelos se catalogan como de “nowcasting”, que es una palabra en inglés que resulta de la combinación de ahora (now) y futuro (casting) y que se puede definir como la predicción del presente, el futuro cercano, o el pasado muy reciente (Banbura, Giannone, Modugno y Reichlin, 2013).

Ahora bien, generalmente para el desarrollo del nowcasting se usan modelos de factores dinámicos y de ecuaciones puente, no obstante, la investigación tiene como valor agregado usar la inteligencia artificial, fundamentalmente el Machine Learning. Por consiguiente, se elabora un modelo inspirado en la teoría macroeconómica aplicando Machine Learning, teniendo como variables explicativas una combinación de 15 indicadores macroeconómicos adelantados que reflejan las condiciones cíclicas de la economía colombiana en periodicidad mensual, posteriormente se transforman a frecuencia trimestral desde el primer trimestre del año 2005 hasta el trimestre 2 del 2019.

Se encontró que el modelo de nowcasting para la economía colombiana, es consistente y preciso para pronosticar el desempeño económico en “tiempo real” y de forma oportuna, que supera notablemente los modelos tradicionales, pronosticando el trimestre 3 del PIB registrado por el DANE.

1. Introducción

El crecimiento del Producto Interno Bruto (PIB) es el principal indicador de desempeño económico tanto de economías avanzadas como de las emergentes, permite identificar los ingresos y la oferta agregada de un país en el corto, mediano y largo plazo, por consiguiente, el rumbo que puede tomar la actividad económica es fundamental para hacedores de política, autoridades económicas, los agentes privados y el gobierno, no obstante, por su naturaleza macroeconómica la publicación presenta rezagos de 4 semanas en Estados Unidos y Reino Unido; entre 6 y 7 semanas en el área euro; 8 semanas en Canadá; y entre 1 y 2 años para el Líbano (Ortega,2018) a partir de la finalización de un trimestre, en el caso colombiano son aproximadamente 50 días, por parte de la entidad estadística encargada el Departamento Administrativo Nacional de Estadística (DANE). Por tanto, resulta de vital importancia un diagnóstico de la economía en "tiempo real", que en los principales bancos centrales del mundo es conocido como "nowcasting", siendo esto una combinación entre el presente (now) y el futuro (forecasting) (Banbura, Giannone, Modugno y Reichlin, 2013).

Ahora bien, generalmente para el desarrollo del nowcasting se usan modelos de factores dinámicos (La Reserva Federal de Atlanta en Estados Unidos, Banco Central Europeo) y de ecuaciones puente (Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), la Reserva Federal de New York, Banco de Inglaterra la mayor parte de bancos centrales en el mundo), no obstante, la investigación tiene como valor agregado usar la inteligencia artificial, fundamentalmente el Machine Learning en este tipo de modelos. Por consiguiente, se elabora un modelo inspirado en la teoría macroeconómica aplicando Machine Learning, teniendo como variables explicativas una combinación de 15 indicadores macroeconómicos adelantados que reflejan las condiciones cíclicas de la economía colombiana en periodicidad mensual, posteriormente todos se transforman a la frecuencia trimestral desde el primer trimestre del año 2005 hasta el segundo trimestre del 2019.

Se proponen 2 modelos de Inteligencia Artificial enfocado en redes neuronales multicapa, conocidos en el campo como el MLP Regresor multivariado, que implementa un perceptrón multicapa (MLP) que se entrena usando la propagación hacia atrás sin función de activación en la capa de salida, siendo un modelo supervisado, dónde se tiene en cuenta para la inclusión de variables explicativas los principales determinantes de la literatura, indicadores líderes que

mayores correlaciones dinámicas presentan en el periodo en cuestión, en síntesis, el modelo 1 tiene como variable dependiente el PIB medido en porcentaje de trimestre anualizado (por ejemplo, tercer trimestre año t , contra tercer trimestre año $t-1$), mientras que el modelo 2 tiene como variable dependiente el PIB medido en niveles (miles de millones de pesos). En ambos modelos encontramos una proyección trimestral del PIB y que fácilmente se puede escalar hacia frecuencia anual, que mejora significativamente por la inclusión de los indicadores y variables macro líderes y el uso de redes neuronales, que de forma conjunta generan consistencia y precisión en nuestro modelo de nowcasting, convirtiéndose en una herramienta eficiente y de alta confiabilidad para medir el ritmo de la actividad económica del país de forma más oportuna para la toma de decisiones de los agentes de la economía con base en información menos incompleta.

El documento, adicionalmente a este apartado, se compone de una sección que presenta el uso de las redes neuronales en la macroeconomía, para continuar con una descripción del modelo de nowcasting para la economía colombiana y los principales indicadores y variables macro que se utilizan junto con la institución de dónde se alimenta los datos, con fuente de datos de fácil de acceso y de publicación más oportuna que se complementa con la presentación de los ciclos de cada uno de los indicadores macro y las correlaciones dinámicas de cada uno con respecto al PIB, para continuar, con la estimación y validación de los modelos mediante el programa más utilizado en el mundo para la aplicación de la inteligencia artificial: Python, la versión 3.7. Para mostrar lo bien que se comporta el modelo en lo que se refiere a la identificación de patrones y predicción, se realiza una breve sección respecto a resultados intra-muestrales y extra-muestrales incluso comparando con los resultados de un DSGE estándar, que se complementa con el recuento de la experiencia que se tuvo en la predicción del tercer trimestre del 2019 del PIB para Colombia, dónde el modelo nowcasting predijo lo observado en las cifras oficiales DANE, mostrando indicios de su buen comportamiento y uso como herramienta en economías emergentes. Finalmente, se presentará unas conclusiones y recomendaciones.

1. Redes Neuronales Artificiales (RNA) para macroeconomía

Las redes neuronales artificiales son un conjunto de técnicas perteneciente al campo de la Inteligencia Artificial (AI). Su estructura consiste en una red formada por nodos (o neuronas) y

conexiones, razón por la cual se asemejan al cerebro de los seres humanos, del cual procede su nombre (Moreno,2010). Las redes neuronales son de aplicación en diversidad de problemas, entre ellos podemos destacar en nuestro caso en primer lugar, el reconocimiento de patrones, fundamental en series macroeconómicas como el PIB, la inflación, el tipo de cambio y el desempleo, y en segundo lugar, para la aproximación de funciones, debido a su flexibilidad y facilidad de uso, que permite combinar la teoría macroeconómica y darle sentido económico a la predicción, pero respetando la evolución de los datos en la realidad económica.

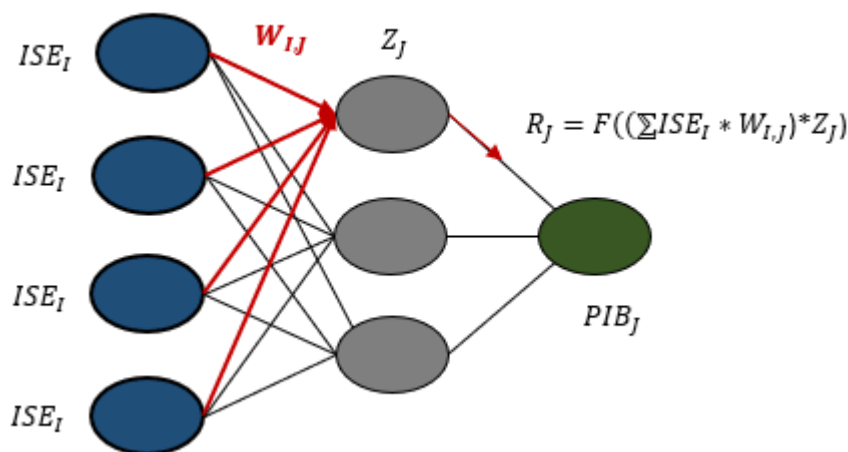
La importancia fundamental de una red neuronal en la macroeconomía consiste en la capacidad de detectar relaciones complejas y no lineales entre series de tiempo de variables macroeconómicas, que tanto teórica como en la práctica tiene algún tipo de relación que no necesariamente es lineal, a partir de unidades muy simples como las neuronas, puesto que, se pueden crear múltiples capas de forma paralela. La red se compone de variables entrada, que en nuestro caso son las series de tiempo de indicadores y variables macro entre el periodo 2005 hasta la más reciente publicación, por mencionar algunos que más adelante se explicarán de forma más detallada junto a las razones de teoría macro y estadística de su inclusión, son el Índice de Seguimiento a la Economía (ISE), la tasa de desocupados del mercado laboral, las obras nuevas de construcción en metros cuadrados del sector construcción, el índice de confianza industrial del sector empresarial doméstico, el Indicador de Bonos de Mercados Emergentes (EMBI) del país que capta en cierto grado el "spread" de Colombia en el mercado internacional, el uso de la capacidad instalada, entre otras, todas son recolectadas de forma mensual, su publicación es de fácil acceso y oportuna por las respectivas entidades competentes. Ahora bien, las variables salidas, tiene que ver con el crecimiento del PIB, ya sea en niveles (millones de pesos) o en porcentaje (trimestre año t, contra trimestre año t-1), solo que en este caso son datos trimestrales, por consiguiente, la necesidad de trimestralizar cada uno de los indicadores o variables entrada a usar en la red, que están relacionadas por algún tipo de correlación dinámica o dependencia, que reflejan las condiciones cíclicas de la economía colombiana.

Las neuronas en macroeconomía se pueden construir dependiendo del objetivo de investigación en diferentes capas. Las más sencillas, constan de una capa de entrada, una capa de neuronas (capa oculta) y una capa de salida, como se muestra en la figura 1. El

funcionamiento de una neurona consiste en la transformación de los valores de las entradas a través de las conexiones en una salida. Se obtiene la salida a partir de una función de propagación, una función de activación, y una función de transferencia (FED, 2019).

Para brindar un ejemplo, usemos el Indicador de Seguimiento de la Economía (ISE), en este caso particular, la ponderación de la variable entrada ISE, se denota de color rojo y se llama W , es conocida como la función de propagación siendo el resultado de la sumatoria de todas las entradas multiplicadas por los pesos de las conexiones (W), más un valor de sesgo, la función de activación en nuestro caso se denota Z_j , activa o desactiva la salida de esta neurona y, finalmente, la función de transferencias se denota como R_j se aplica al resultado de la función de propagación y normalmente consiste en una función de salida acotada, todo este análisis se sintetiza en la figura 1, que muestra el funcionamiento de la red con solo una capa oculta, además, de carácter univariado para un mejor entendimiento.

Figura 1: Ejemplo red neuronal artificial con una capa oculta para nowcasting en Colombia



Fuente: Elaboración propia.

En este documento, las redes que se usarán se encuentran en el marco de los modelos de aprendizaje supervisado, lo que implica, que los datos (entradas, es decir, los indicadores y variables macro) tienen una respuesta conocida (o salida), con la cual se ajusta o entrena la red neuronal.

La ventaja adicional dentro del campo de la macroeconomía del uso de redes neuronales tiene que ver con que no es necesario asumir una forma funcional que limita bastante la precisión de los pronósticos como en los modelos AR, MA, ARIMA, ARMA, SARIMA, VAR, sumado a que no es necesario asumir estacionariedad en las entradas, y no es relevante la multicolinealidad por la característica subyacente de las redes neuronales artificiales de detectar relaciones complejas y no lineales entre series de tiempo macro.

Posteriormente, de haber definido el número de capas suficientes de las variables entradas se procede a una fase fundamental y el corazón de este tipo de modelos, el entrenamiento de la red, en esta fase es dónde se incursiona en el mundo del Machine Learning propiamente. Como bien lo define la literatura estándar se tiene que *“El entrenamiento o aprendizaje, cuyo objetivo es que la red neuronal sea capaz de reproducir el comportamiento subyacente en los datos aportados, consiste básicamente en la minimización de una función de coste o error, lo que equivale a que la salida de la red se aproxima a la salida en los datos. La función de coste más común es la de promedio de errores al cuadrado (MSE)”*. De esta manera, existen diferentes tipos métodos de ajuste, se usarán los más estándares de la literatura, los métodos tipo gradiente (derivada multidimensional) y métodos basado en algoritmos genéticos, ambos métodos iterativos, que se repiten hasta cumplir alguno de los diferentes criterios de parada, entre los más importantes, la pertinencia de los coeficientes de la regresión lineal múltiple con la teoría macroeconómica, la consecución de un error mínimo, la consistencia con las correlaciones dinámicas, el número de iteraciones y la racionalidad económica de los datos y entrenamiento presentado.

Uno de las dificultades más comunes que se presentan al trabajar con redes neuronales en macro, consisten en modelos con “sobre aprendizaje”, por consiguiente, para controlar dicha problemática, en la programación se tiene en cuenta cada uno los datos de los indicadores y variables macro en el periodo en cuestión, se dividen en tres grandes grupos: Los datos de entrenamiento (x_{train}), como deben ser representativos del total de datos, se seleccionan aleatoriamente; los datos de validación, se emplean después de cada iteración, y por último, pero no menos importante, los datos test (x_{test}), se emplean una vez finalizado los x_{train} . La división de los datos para el nowcasting del PIB en Colombia fue un 80% de datos de entrenamiento, un 10% de validación y un 10% de “test”, aunque la elección de dichos

porcentajes depende del número de datos disponible, que para la macroeconomía y aún más si es un país emergente, son limitados teniendo en cuenta las otras ciencias que cuenta con un volumen de datos bastante significativo.

En síntesis, se presenta algunas ventajas y desventajas más comunes de las redes neuronales artificiales en macroeconomía:

Tabla 1: Algunas ventajas y desventajas de las redes neuronales en macroeconomía

Ventajas	Desventajas
No es relevante el problema de multicolinealidad entre variables entrada.	Sobre aprendizaje o pérdida de generalización.
Independencia entre complejidad del problema y dimensionamiento de la red.	Criterios de dimensionamiento de la red neuronal arbitrarios.
No es necesario supuesto de estacionariedad, se cuenta con rapidez de ajuste y simulación. No requiere forma funcional. Facilidad en la programación.	Comportamiento de caja negra, división aleatoria de datos en entrenamiento, validación y test, dan lugar a soluciones diferentes.
Laboratorio virtual, curvas de diseño.	Mínimos locales.

Fuente: Elaboración propia con revisión sistemática de la literatura.

La importancia de esta breve presentación de las redes neuronales y su aplicación a la macroeconomía tiene que ver, con brindar un mayor entendimiento de cuándo se pueden usar y cuando no se debería optar por este tipo de técnicas, por consiguiente, se ha seleccionado este tipo de modelación, teniendo en cuenta el interés de entrenar el patrón del comportamiento del PIB trimestral entre el periodo 2005-2019 con el menor MSE posible, es decir, con datos observados de los indicadores y variables macro de entidades competentes se pretende explicar los datos registrados del PIB según las cifras oficiales, lo que se conoce como NOW, pero, adicionalmente, para poder pronosticar de mejor manera el futuro de la senda del PIB, se brinda información lo más reciente posible (indicadores líderes mensuales) para orientar a la red neuronal artificial hacia dónde se debería dirigir el desempeño económico del país de la forma más oportuna y precisa posible, más conocido como FORECASTING, por consiguiente, se incursiona en un mundo un poco más nuevo para la economía, el llamado "Nowcasting" que se presentará en la siguiente sección.

Lo que busca el nowcasting básicamente es aprovechar la publicación periódica de otros indicadores y variables macro que generalmente en Colombia, se realizan con frecuencia diaria, mensual y otros de forma trimestral; la idea es utilizar dicha información generada con una frecuencia más alta y oportuna, para aproximar el cálculo del PIB trimestral y anual, a través de diversas metodologías como las ecuaciones puente, modelos de factores dinámicos, vectores autorregresivos, en nuestro caso el Machine Learning. Dichas estimaciones se realizan de forma recursiva a partir del siguiente mes del trimestre finalizado y publicado; se incorpora mes a mes (en Visual Code de Python) la nueva información disponible de los indicadores y variables macro hasta el tercer mes del trimestre a publicar, brindando una aproximación al valor real del PIB en miles de millones de pesos para dicho trimestre por lo menos con aproximadamente 50 días de anticipación según la publicación de las indicadores y variables macro de la economía colombiana.

2. Nowcasting en la economía colombiana

El diagnóstico de la economía en “tiempo real”, en la literatura se conoce comúnmente como “nowcasting”, siendo esto una combinación entre el presente (now) y el futuro (forecasting). Nowcasting en este documento se entenderá como el conjunto de indicadores y variables macro del desempeño económico, elaborados usando las fuentes de información de las entidades competentes. Teniendo en cuenta el rezago de la publicación de la información de las principales cifras macro, busca reflejar de forma oportuna el ritmo de la actividad económica del país, la novedad en el modelo tiene ver que no se utilizará lo estándar de factores dinámicos o ecuaciones puente, en cambio, se estimará con un modelo de inteligencia artificial, enfocado en Machine Learning, especialmente se aplicará redes neuronales multicapa.

El objetivo de investigación consiste en estimar el mejor modelo Nowcasting con los indicadores y variables macroeconómicas seleccionadas que permita pronosticar de forma consistente y precisa el desempeño de la economía colombiana (PIB) para la toma de decisiones de los agentes en el mediano y largo plazo aplicando inteligencia artificial. Para ello, se inicia con una fase de identificar los principales indicadores macroeconómicos líderes, de más fácil acceso y más adelantados que describan el comportamiento del PIB en Colombia, que se validan con las correlaciones dinámicas y teoría macroeconómica, para poder estimar modelos de Nowcasting usando Machine Learning. Posteriormente, se compara y selecciona el mejor

modelo que permita pronosticar el PIB de forma oportuna para la toma de decisiones de los agentes de la economía, se estiman dos modelos teniendo como variables explicativas todos los indicadores y variables macro con frecuencia mensual presentados en la Tabla 1 que se transforman en frecuencia trimestral, la única variación consiste en que la variable dependiente del modelo 1 (j) se estima en variación anual respecto al trimestre anterior, mientras que el modelo 2 (i) se estima como variable dependiente el PIB en niveles, medido en miles de millones de pesos.

Teniendo en cuenta las instituciones encargadas de divulgar la información macroeconómica del país se usará el DANE, Banco de la República, Fedesarrollo y Anif, ahora bien, para lo que se refiere a variables relevantes del mercado internacional que afectan la evolución de la actividad económica doméstica se usará la data del Banco Central de Reserva del Perú y la FRED de la Reserva Federal de St Louis de los Estados Unidos.

Para el caso de la economía colombiana, se presentan los principales indicadores macroeconómicos y su acrónimo útil para el entendimiento más adelante de las ecuaciones acompañado de su fuente de información, su periodicidad y medida de los datos que se sintetizan en la siguiente tabla:

Tabla 1: Indicadores y variables macro para el nowcasting del PIB en Colombia

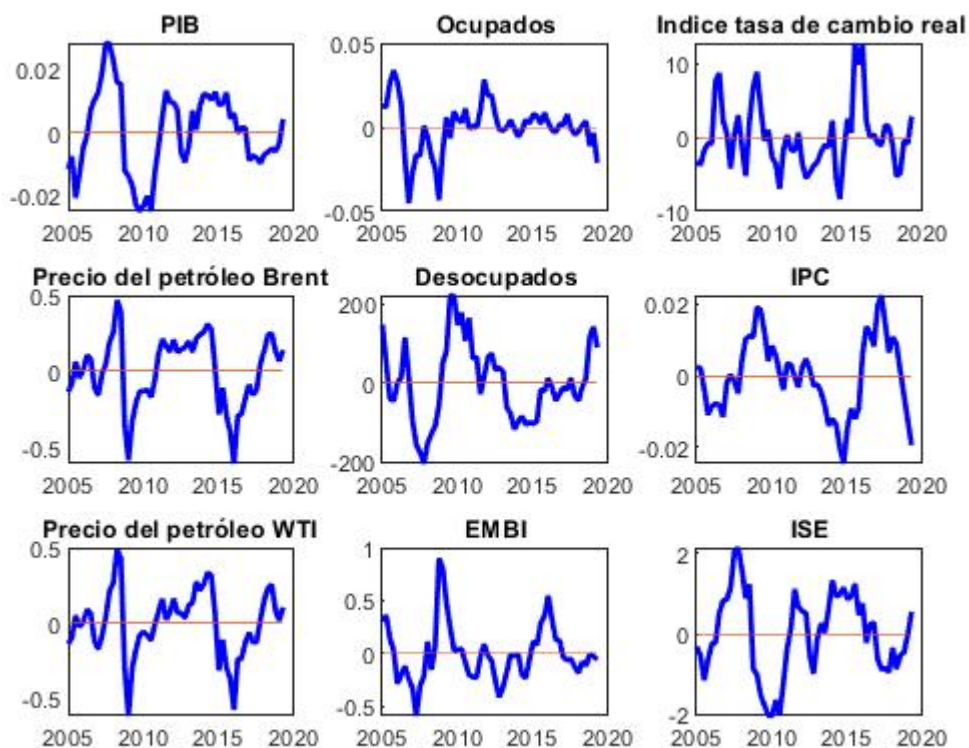
Indicador Macroeconómico de la economía colombiana	Fuente de información	Frecuencia de los datos	Unidad de los datos
Producto Interno Bruto (PIB) Base 2018.	DANE	Trimestral	i = Niveles j = Porcentaj
Indicador de Seguimiento de la Economía (ISE)	DANE	Mensual	Index
Precio Internacional del petróleo (ppb)	FRED	Diario	Dólares por barril
Tasa de desempleo (dese)	DANE	Mensual	Porcentaje
Índice de Precios al Consumidor (ipc) Base 2018	Banco de la República	Mensual	Index
Licencias aprobadas (licenses)	DANE	Mensual	Index
Obras de construcción nuevas (cnew)	Camacol DANE	Mensual	Metros cuadrados
Índice de Precios al Productor (ipp)	DANE	Mensual	Index
Número de ocupados (fw)	DANE	Mensual	Número de personas
Tasa de interés de política (tibr)	Banco de la República	Diario	Porcentaje

Indicador de Bonos de Mercados Emergentes (embi)	Banco de Perú	Mensual	Index
Ventas minoristas (mino)	Anif	Mensual	Index
Índice de confianza industrial (ici)	Fedesarrollo	Mensual	Index
Índice de confianza del consumidor (icco)	Fedesarrollo	Mensual	Index
Índice de confianza comercial (icc)	Fedesarrollo	Mensual	Index
Índice de uso de la capacidad instalada (capa)	Fedesarrollo	Mensual	Index

Fuente: Elaboración propia.

Los datos de las fuentes oficiales que se presentan en la tabla 1, se proceden a transformar de forma logarítmica para presentar de forma intuitiva su relación en el tiempo con el PIB, exceptuando las variables índices, para aplicarles el filtro estándar de Hodrick-Prescott para obtener los ciclos de los indicadores y variables macro de la economía colombiana con frecuencia trimestral desde el primer trimestre del año 2005 hasta el segundo trimestre del 2019, en la figura 2 se sintetiza un resumen.

Figura 2: Ciclos de los principales indicadores y variables macro en la economía colombiana
2005q1-2019q2



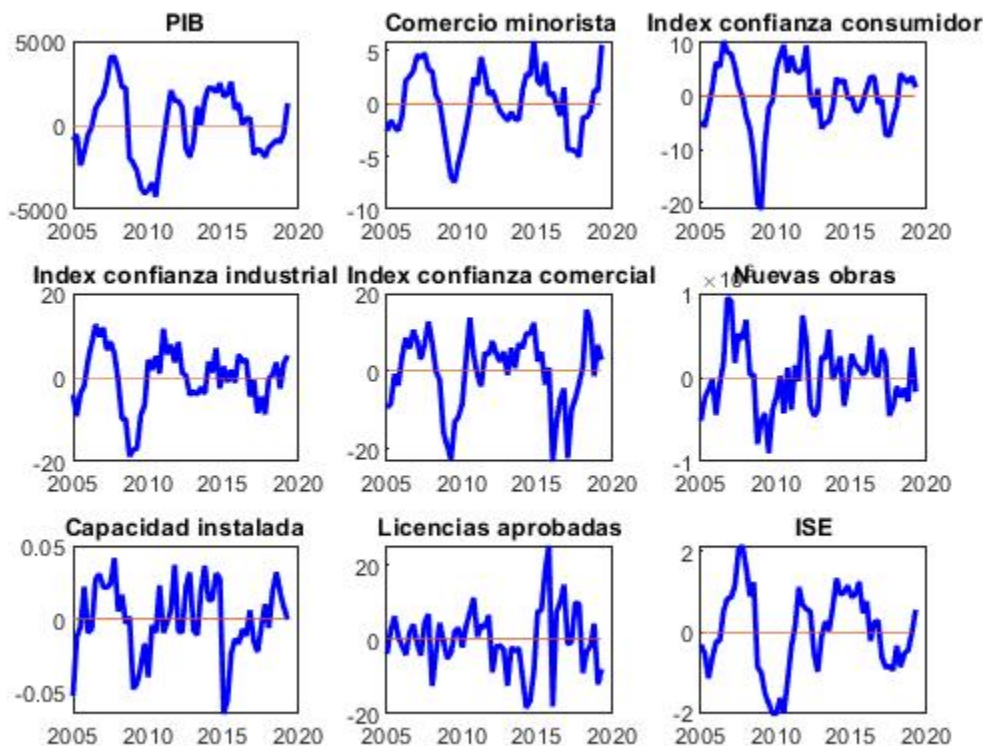
Fuente: Simulación realizada en MATLAB.

Se evidencia que el desempeño de la actividad económica del país (PIB) presentan una fuerte relación con el Indicador del Seguimiento de la Economía (ISE), algún tipo de relación con el precio internacional del petróleo tipo Brent y también el WTI, el Indicador de Bonos de Mercados Emergentes (EMBI), los desocupados y el índice de precios del consumidor (no hay evidencia de curva de Phillips en Colombia) y en menor medida con el índice de tasa de cambio real y el número de ocupados. La importancia radica, en que todas las variables explicativas del PIB mencionadas y presentadas son de publicación muy oportuna de forma mensual, incluso algunas de periodicidad diaria, que, sin duda alguna, contribuye a cerrar los rezagos de información en el país, respecto a la publicación de las cuentas nacionales por parte de las entidades competentes. Para el análisis se opta por el precio internacional del petróleo y no del

tipo WTI, por las propiedades químicas (pesado) de la producción de los barriles en el país y sus mayores correlaciones dinámicas que se muestran en la figura 4.

Ahora bien, existen otros indicadores macroeconómicos que muestran un ciclo similar al del Producto Interno Bruto para la economía colombiana, por ejemplo el índice de confianza del consumidor, el industrial y el comercial, la cuestión es que pueden tomar valores negativos, por consiguiente, en este caso particular, para el análisis de ciclos de la base de datos incluyendo el PIB no se le toma logaritmo para poder hacer análisis comparativo, simplemente se calcula el estándar filtro de Hodrick Prescott de forma trimestral para cada una de las variables, que indican una idea macro intuitiva respecto al PIB, los resultados de la simulación de los ciclos se muestran en la siguiente figura 3.

Figura 3: Ciclos de otros indicadores macro adelantados en la economía colombiana 2005q1-2019q2



Fuente: Simulación realizada en MATLAB.

Se presenta una importante correlación dinámica de nuevo con el ISE, no obstante, también es bastante notable con el índice de confianza industrial, índice de confianza del consumidor, índice de confianza comercial, comercio minorista y las obras de construcción nuevas en el país y en menor medida el índice del uso de capacidad instalada y las licencias aprobadas con el desempeño económico en el periodo de estudio. Todos indicadores y variables macro con frecuencia mensual que son de fácil acceso y de oportuna publicación.

Lo relevante de este breve estudio de las condiciones cíclicas de la economía colombiana, tiene que ver, con resaltar que todos los indicadores macro utilizados son de fácil acceso y oportuna publicación, y presentan patrones similares al ciclo del PIB, ahora bien, como las redes neuronales artificiales pronostica de forma consistente y precisa el comportamiento pasado hasta la información más reciente posible (now), se vuelve de vital importancia encontrar variables de entrada con correlaciones dinámicas bastante significativas para la inclusión en el modelo de Machine Learning, que permitan al modelo "aprender" la senda futura (Forecasting) del PIB con indicadores macro de fácil acceso para cualquier agente de la economía y de publicación oportuna, es decir, acortando los rezagos de información en el país. Finalmente, para contar con mayor rigurosidad estadística con los datos, se presentan las correlaciones dinámicas respecto al Producto Interno Bruto en Colombia de los principales indicadores y variables macro que muestran las condiciones cíclicas de la economía doméstica.

Figura 4: Correlaciones dinámicas de las condiciones cíclicas de la economía colombiana
2005q1-2019q2

Correlaciones dinámicas de los indicadores y variables macro respecto al PIB 2005-2019											
VARIABLES	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5
ISE	-0.0276	0.2099	0.4869	0.6838	0.8511	0.9898	0.8553	0.6893	0.4887	0.2223	-0.0154
Comercio Minorista	0.2618	0.4854	0.6809	0.7956	0.7941	0.7697	0.5674	0.3430	0.1041	-0.1656	-0.3742
Tasa de interés	-0.6293	-0.4949	-0.3065	-0.0725	0.1773	0.4109	0.6262	0.7621	0.7703	0.6894	0.5404
Desocupados	0.1034	-0.1232	-0.3801	-0.5603	-0.6624	-0.7224	-0.7521	-0.6679	-0.5596	-0.4136	-0.2455
IPC	-0.5200	-0.6084	-0.6836	-0.6756	-0.5975	-0.4890	-0.3327	-0.1624	-0.0015	0.1293	0.2217
Indice confianza industrial	0.5237	0.6264	0.6638	0.6332	0.5564	0.3915	0.1535	-0.1280	-0.3573	-0.5434	-0.6520
Indice confianza comercial	0.5491	0.6413	0.6599	0.5986	0.5258	0.4247	0.2428	0.0144	-0.2073	-0.3777	-0.5336
Obras nuevas	0.1800	0.3211	0.4492	0.5981	0.6392	0.5891	0.4719	0.3089	0.0893	-0.0905	-0.2415
EMBI	-0.5339	-0.5968	-0.5920	-0.4996	-0.4023	-0.2371	0.0257	0.2324	0.3637	0.4665	0.5379
Indice confianza consumidor	0.5732	0.5883	0.5305	0.4517	0.3202	0.1227	-0.1026	-0.3494	-0.5351	-0.6705	-0.7355
Capacidad instalada	0.2906	0.3806	0.5039	0.5038	0.3992	0.3254	0.2516	0.1077	-0.1132	-0.3002	-0.3942
IPP	-0.3170	-0.3755	-0.3787	-0.2776	-0.1925	-0.0253	0.0444	-0.0220	-0.0405	-0.0843	-0.2024
Tasa de cambio real	-0.2355	-0.2125	-0.1018	0.0218	0.0280	0.1371	0.1588	0.1931	0.2510	0.3042	0.3729
Precio del petróleo Brent	0.1852	0.2262	0.2630	0.2924	0.3730	0.3730	0.2562	0.0974	-0.0664	-0.2347	-0.3720
Precio del petróleo WTI	0.2250	0.2507	0.2635	0.2653	0.3168	0.3072	0.1999	0.0130	-0.1291	-0.2607	-0.3600
Ocupados	0.1925	0.0545	-0.0542	-0.1149	-0.1281	-0.1712	-0.1798	-0.1873	-0.1948	-0.2123	-0.1740
Licencias aprobadas	-0.2043	-0.0985	-0.0800	-0.0861	-0.1444	-0.1060	-0.0495	-0.1266	-0.0693	0.0116	0.0386

Correlación positiva fuerte

Correlación negativa fuerte

Correlación negativa débil

Correlación positiva débil

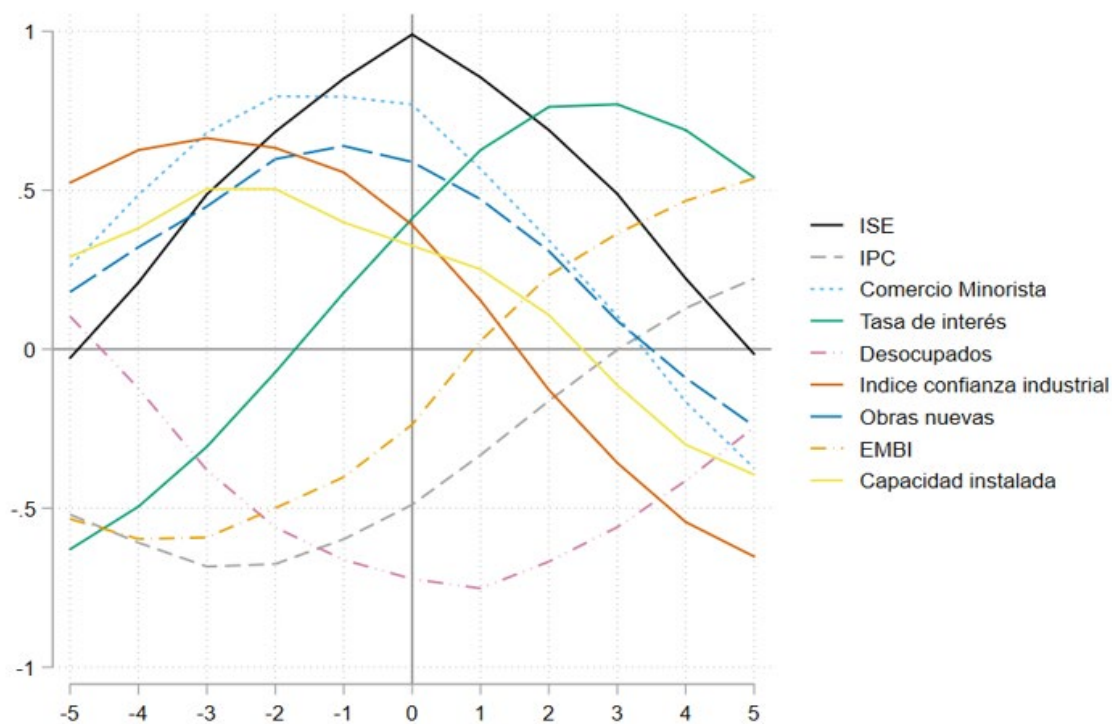
Fuente: Simulación realizada en STATA.

Como se evidencia, los valores más altos de las correlaciones respecto al Producto Interno Bruto en Colombia en el periodo 2005-2019 de cada uno de los indicadores macroeconómicos se muestran en color azul si es positiva (procíclica) y roja si es negativa (anticíclica), y los que registran correlaciones dinámicas de forma adelantada respecto al PIB, se presentan en los periodos antes del t, recordando que la frecuencia es trimestral. De esta forma, se valida la hipótesis que un conjunto de indicadores y variables macro líderes explica la evolución futura de la actividad económica del país, que serán el insumo fundamental para la construcción del modelo nowcasting. Cabe rescatar, las fuertes correlaciones respecto al PIB, de variables como el ISE, comercio minorista, la tasa de interés de política del banco de la República, las obras nuevas de construcción, el índice de confianza industrial y el Indicador de Bonos de Mercados

Emergentes (EMBI), que se muestran de forma ordenada de la mayor correlación presentada a la menor en la figura 4.

Otra forma de visualizar los coeficientes de correlación dinámica es mediante la figura 5. Las curvas dividen el plano cartesiano en 4 cuadrantes. La interpretación es que las curvas por encima del eje horizontal (dado por 0) corresponden a las variables que presentan una correlación positiva con el producto (variables procíclicas) en nuestro caso por ejemplo el ISE o el comercio minorista y por debajo de cero corresponden a las variables correlacionados negativamente (contracíclicas) en nuestro caso el Índice de Precios al Consumidor (IPC), el Indicador de Bonos de Mercados Emergentes (EMBI) o desocupados. A la izquierda del eje Y (dado por t) corresponden a las variables que presentan una correlación adelantada y a la derecha las variables rezagadas. Por ejemplo, uno de los más altos coeficientes de correlación, en valor absoluto, es entre el comercio minorista y el PIB que se encuentra encima del eje horizontal y en el eje vertical, es decir en t, por consiguiente, el indicador de ventas minoristas es una variable procíclica y contemporánea.

Figura 5: Coeficientes de correlación dinámica de los indicadores y variables macro



Fuente: Simulación hecha en STATA.

Teniendo en cuenta la validación tanto desde la teoría macroeconómica, estadística y lo observado en los datos de la economía real, se procede a estimar el modelo de nowcasting para la economía colombiana usando los indicadores y variables macro que muestran las condiciones cíclicas de la economía y permiten brindarles aprendizaje a las redes neuronales artificiales para la consistencia y precisión de la predicción del PIB en una economía emergente.

3. Modelo de Inteligencia Artificial: Machine Learning

La Inteligencia Artificial ha evolucionado hacia los temas de Machine Learning y Deep Learning durante la última década, siendo sin duda alguna uno de las temáticas más populares hoy en día. Sus avances se han extendido hacia modelos predictivos que bien compiten con la econometría tradicional de series de tiempo como los AR, MA, ARMA, ARIMA, VAR, incluso con los modelos dinámico-estocásticos de equilibrio general (DSGE) usados en macroeconomía en lo que refiere a un mejor reconocimiento del patrón del ciclo económico. No obstante, también se están revolucionando las técnicas de nowcasting usuales en economía: Factor dinámico y ecuación puente, por consiguiente, con la ayuda de la teoría macroeconómica y la identificación de los indicadores líderes y de las mayores correlaciones dinámicas presentadas en la sección anterior que serán elementos esenciales para complementar, brindar mayor precisión y darles sentido a las predicciones, permitirán el cierre de la brecha de los rezagos de información que se presentan respecto a la publicación oficial de los datos de cuentas nacionales y las entidades competentes.

Ahora bien, generalmente para el desarrollo del nowcasting se usan modelos de factores dinámicos (La Reserva Federal de Atlanta en Estado Unidos, Banco Central Europeo) y de ecuaciones puente (Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), la Reserva Federal de New York y la mayoría de bancos centrales) y el **Machine Learning** (Reserva Federal de Sant Louis, Estados Unidos). Por consiguiente, se elabora un modelo inspirado en la teoría macroeconómica aplicando Machine Learning, teniendo como variables explicativas una combinación de 15 indicadores y variables macroeconómicas adelantadas (ver tabla 1) que reflejan las condiciones cíclicas de la economía colombiana en periodicidad

mensual, posteriormente todas se transforman a la frecuencia trimestral desde el primer trimestre del año 2005 hasta el segundo trimestre del 2019.

Se proponen 2 modelos de Inteligencia Artificial enfocado en redes neuronales, es decir, se estiman dos modelos de redes neuronales multicapa que tiene que ver con el MLPRegressor multivariado, siendo uno de la gama de modelos de aprendizaje supervisado, dónde se tiene en cuenta la inclusión de variables independientes los principales determinantes de la literatura, indicadores macro líderes y los que mayores correlaciones dinámicas presentan en el periodo en cuestión, en síntesis, el modelo 1 tiene como variable dependiente el PIB medido en niveles (miles de millones de pesos), mientras que el modelo 2 tiene como variable dependiente el PIB medido en porcentaje de trimestre anualizado (por ejemplo, tercer trimestre año t, contra tercer trimestre año t-1).

Todas se transforman a frecuencia trimestral, para poder empalmar con la serie del PIB con ajuste calendario y estacional, por consiguiente, el modelo 1 (i) y modelo 2 (j) quedan de la forma:

$$PIB_{t,i,j} = \beta_0 + \beta_{1t}ise_t + \beta_{2t}ppb_t + \beta_{3t}dese_t + \beta_{4t}ipc_t + \beta_{5t}licenses_t + \beta_{7t}cnew_t + \beta_{8t}tibr_t + \beta_{9t}embi_t + \beta_{10t}mino_t + \beta_{11t}lic_t + \beta_{12t}ici_t + \beta_{13t}icco_t + \beta_{14t}icc_t + \beta_{15t}capa_t + \varepsilon_t.$$

Dónde i = Niveles; j = Porcentaje anual trimestralizado,
 ε_t es el ruido blanco estándar, con media cero y varianza constante.

En primer lugar, se realizó un pronóstico intra-muestral, eliminando información del 2019 para la serie de tiempo del PIB tanto en porcentaje como en niveles, para evaluar el nivel de predicción de nuestro nowcasting, posteriormente, se presentará la experiencia del pronóstico del PIB con frecuencia trimestral para el tercer trimestre del año 2019, que fue exitosa, generando un modelo robusto, preciso y consistente, por último, se presenta lo que el modelo de nowcasting predice de crecimiento anual para la economía colombiana.

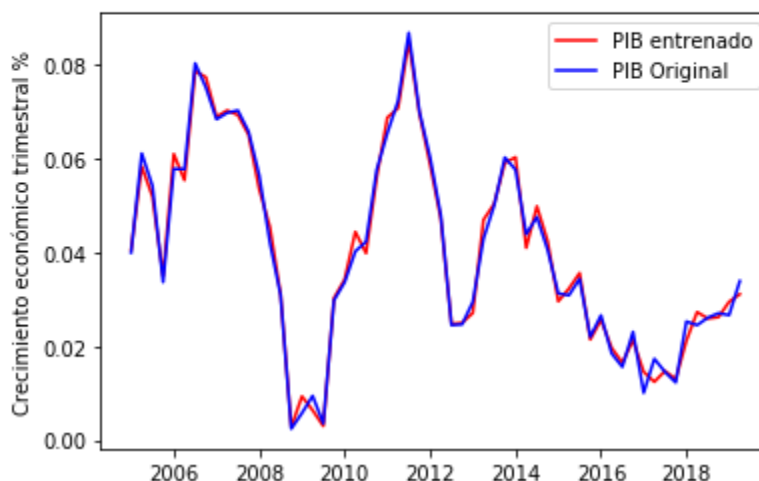
4. Resultados

5.1 Multi-layer Perceptron (MLP): Modelo crecimiento del PIB (%) real en Colombia

Todos los datos son trimestrales iniciando en el primer trimestre del año 2005, medidos en variación anual respecto al trimestre anterior que permite un mejor análisis y se estima con Python versión 3,7 usando la librería especializada para Machine Learning (scikit-learn), en la figura 6 se presentan los resultados del pronóstico intramuestral.

Figura 6: Resultados del modelo Nowcasting del PIB para Colombia

```
Predicción del PIB en el primer trimestre del 2019 es:  
[0.02944512]  
Predicción del PIB en el segundo trimestre del 2019 es:  
[0.03130405]
```

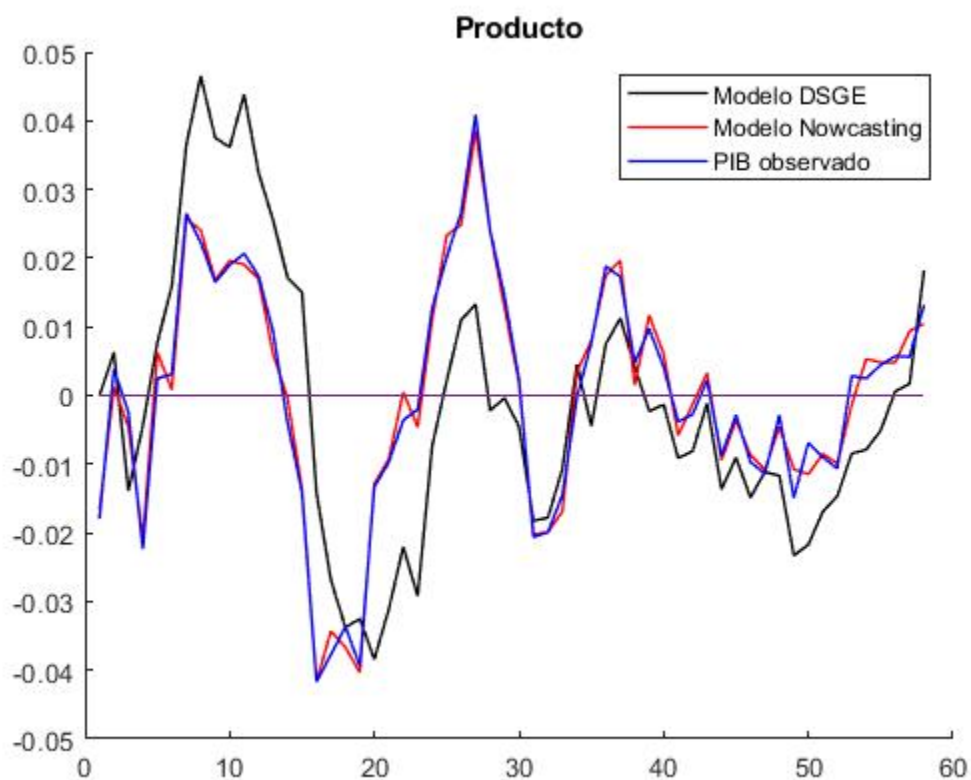


Fuente: Simulación realizada en Python

En primer lugar, para una validación del nowcasting se estimó una predicción intra-muestral, es decir, se eliminaron todos los datos del primer y segundo trimestre del PIB del 2019 para evaluar la capacidad predictiva del modelo, los resultados se muestran en la figura 6 que son prácticamente idénticos a los registrados en el primer y segundo trimestre de cuentas nacionales 2,7 y 3,2 respectivamente, es importante notar la capacidad predictiva del modelo y la importancia de las redes neuronales de "aprender" el patrón del PIB con indicadores y variables macro dado las altas correlaciones dinámicas encontradas, para mostrar la precisión obtenida, se presenta un gráfico comparativo con los modelos macro más estándar de

cualquier banco central en el mundo, se toma los resultados del modelo DSGE con un RBC estándar elaborado por Granda & Montañez (2019), con una función de utilidad Hercowitz–Huffman (**GHH**), con gobierno y con progreso técnico aumentativo con la misma base de datos utilizada para el nowcasting (desde el primer trimestre del 2005 hasta el segundo trimestre del 2019, en total 58 trimestres) y a simple vista se nota la diferencia en la figura 7, para poder hacer análisis comparativo frente al modelo DSGE para la economía colombiana, los resultados del nowcasting se le aplican los filtros de Hodrick y Prescott.

Figura 7: Resultados modelo Nowcasting vs modelo DSGE



Fuente: Simulación realizada en Python integrando con Dynare

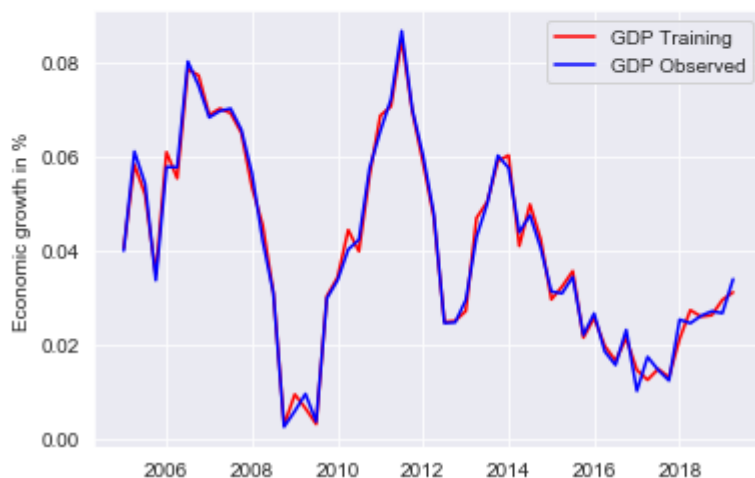
La inteligencia artificial combinada con el nowcasting al menos para la economía colombiana parece explicar mejor los ciclos del PIB que estimando por un modelo DSGE estándar, un resultado interesante, puesto que, probablemente vendrá una revolución de la macroeconomía

cuando se integre completamente este tipo de técnicas, pero es una discusión que va más allá del objetivo de este documento.

Ahora bien, regresando a la estimación del pronóstico trimestral del PIB, se vuelven a incorporar los datos del PIB hasta el más reciente (segundo trimestre del 2019) para estimar el modelo 1 con los indicadores y variables macro con la información más reciente posible, y se encuentran los resultados que se muestran en la figura 8.

Figura 8: Nowcasting del crecimiento del PIB (%) en Colombia

```
The error is: 0.0021812533905280836
The value of r2 is: 0.98928624069948
The parameters are:
' [ 2.21318827e-04 -7.35049543e-05 -1.39845145e-03 -1.02930174e-04
-1.48106511e-04 1.57459366e-09 4.75203455e-04 -3.26055830e-06
-9.41557990e-02 -1.69051980e-05 -9.52615525e-05 2.36257701e-04
-2.34104930e-04 -2.56604029e-05 -2.86474092e-02 2.39778961e-03
9.36275883e-01]
GDP forecast in the third quarter 2019 is:
[0.03216562]
GDP forecast in the fourth quarter 2019 is:
[0.03659138]
```



Fuente: Simulación realizada en Python.

Según el modelo en porcentaje de crecimiento trimestral anualizado con un score de 98,9% se espera un crecimiento trimestral cercano al 3,2% y 3,6% para los últimos trimestres del año respectivamente. No obstante, los resultados obtenidos por nowcasting, aunque son más precisos que los modelos tradicionales, y se encuentra con un score por encima del 98%, es

posible mejorarlo. Para ello, se tomarán el PIB en niveles dónde la integración de redes neuronales y señales de hacia donde se puede dirigir la economía en los próximos periodos recogidos en los indicadores y variables macro con mayores correlaciones dinámicas, y que su publicación es de fácil acceso y de periodicidad mensual, permitirá disminuir el nivel de varianza en los pronósticos del PIB para la economía colombiana.

5.2 Modelo 2: Multi-layer Perceptron (MLP): Modelo en niveles del PIB real en Colombia.

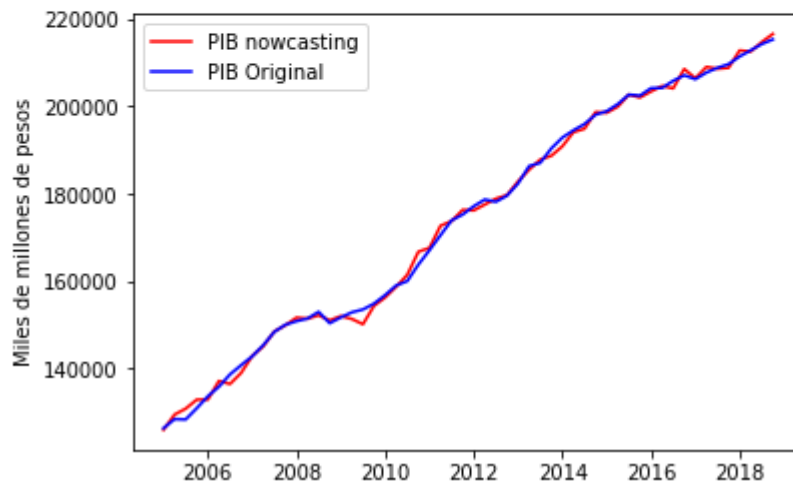
El modelo 2 usa redes neuronales que se alimentan de series temporales de los indicadores y variables macro presentados en la Tabla 1, en primer lugar, se evalúa de nuevo la capacidad predictiva del modelo nowcasting eliminado dos trimestres del PIB en niveles (miles de millones de pesos), es importante, notar que la precisión mejora considerablemente incluso con un score del 99.8%. Los parámetros que se presentan van en consonancia con la descripción del modelo previamente presentado. La serie que predice el nowcasting para la economía colombiana es la línea color rojo, mientras que, la serie de los datos registrados por el DANE se muestran de color azul, ambas series medidas en miles de millones pesos.

Figura 9: Nowcasting del PIB Colombia vs el PIB observado medido en niveles

```

El error es: 1195.320563305269
The value of r2 is: 0.9980839634709674
The parameters are:
' [ 5.91330443e+01  2.83611100e+01 -1.22251623e+03  5.30833833e+02
  1.25692118e-03 -1.42393048e+02  1.82588906e+00 -1.66275875e+04
  2.50862568e+00  7.55403847e+02  2.75995166e+01 -1.46734535e+02
 -4.08386074e+01]
Predicción del PIB en el primer trimestre del 2019 es:
[217069.79017885]
Predicción del PIB en el segundo trimestre del 2019 es:
[220097.02975559]

```



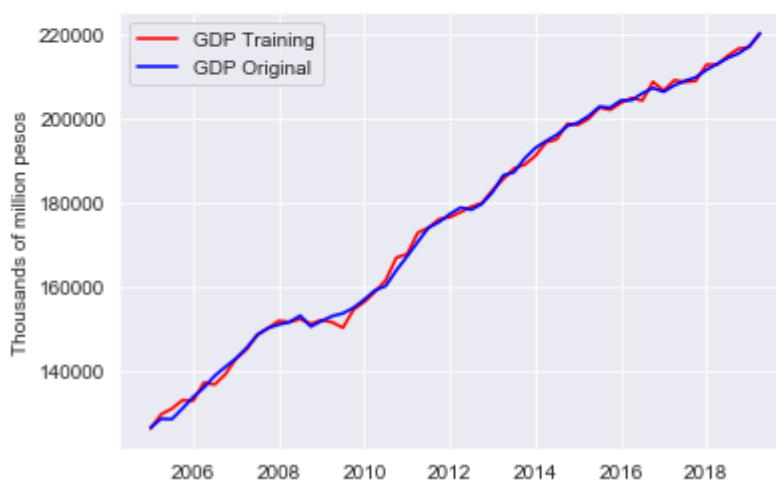
Fuente: Simulación realizada en Python.

Dato observado PIB del DANE	Resultado del Nowcasting con AI
217087.3919	217069.7902
220072.3731	220097.029

Finalmente, se reincorpora los datos más recientes del primer y segundo trimestre del 2019 acompañado de la última publicación de los indicadores y variables macro utilizados, para pronosticar el tercer y cuarto trimestre del 2019 del desempeño de la actividad económica en colombiana, los resultados se presentan en la siguiente figura 10.

Figura 10: Nowcasting predice el tercer trimestre del PIB registrado del DANE

```
The error is: 1160.1565098442327
The value of r2 is: 0.9982873788477529
The parameters are:
' [ 7.00025707e+01  3.53813782e+01 -1.10884954e+03  5.42979647e+02
  1.34544429e-03 -1.73002916e+02  1.91292850e+00 -1.55128567e+04
  4.35775252e+00  7.53650424e+02  3.32527859e+01 -1.37785577e+02
 -4.55696350e+01]
GDP forecast in the third quarter 2019 is:
[221205.44139665]
GDP forecast in the fourth quarter 2019 is:
[222120.79113358]
```



Fuente: Simulación realizada en Python.

Por consiguiente, cuando se calcula el crecimiento del tercer trimestre del año 2018 frente al trimestre proyectado por el modelo nowcasting se estima en un 3,3% que fue efectivamente lo registrado por el DANE, con una cifra que rondaba los 221979 miles de millones de pesos.

Es importante destacar, que el modelo 2, es decir, el modelo del PIB en niveles es más preciso que el modelo 1 del PIB en variación trimestral anualizada, en lo que se refiere al nowcasting, no obstante, ambos son mejores no solamente en la capacidad predictiva, sino también en explicar el patrón de comportamiento de las series de tiempo macroeconómicas que los modelos estándar DSGE. Por consiguiente, se cuenta con un modelo nowcasting que nos permitirá estimar la evolución futura del PIB en Colombia, usando Machine Learning y los indicadores y variables macro con mayores correlaciones dinámicas y adelantadas, de esta manera, el modelo nowcasting con indicadores macro permite realizar un diagnóstico de la

economía más oportuno, de fácil aplicación y elimina en cierto grado los rezagos de información que se presentan en economías emergentes como Colombia.

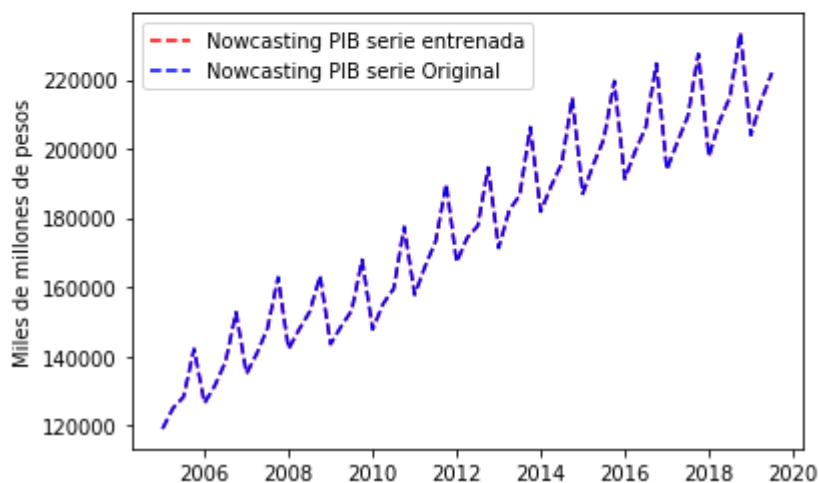
Finalmente, teniendo en cuenta el mejor modelo nowcasting para la economía colombiana, se realiza la predicción trimestral teniendo el PIB en niveles en miles de millones de pesos, no obstante, en esta ocasión tomando la serie original del DANE (no desestacionalizada) que es la serie con que se registra el crecimiento económico en Colombia y se estima la predicción para el cuarto trimestre del PIB serie original como se muestra en la figura 11.

Figura 11: Nowcasting predice el tercer trimestre del PIB registrado del DANE

```

El error es: 4.170133942914751e-10
El score del modelo es: 1.0
Los coeficientes del modelo son:
' [ 2.01173000e+03  1.91700692e-13 -1.20830090e-13 -2.47184213e-12
  -9.27879410e-14 -8.52181349e-16 -2.27996190e-12  1.30065141e-14
   8.22580712e-10  3.71610400e-13  1.93922951e-12 -1.36126485e-12
   4.41758023e-12  2.28063218e-13 -4.50587360e-10]
Predicción del PIB tercer trimestre:
[221978.54494955]
Predicción del PIB cuarto trimestre:
[241910.21162907]

```



Fuente: Simulación realizada en Python.

En síntesis, el modelo nowcasting estima un crecimiento del PIB para Colombia con frecuencia trimestral que ronde el 3,6%, (241.910 miles de millones de pesos sobre la serie original), lo que implica, que la economía colombiana estaría crecimiento 3,27% en el año 2019 con los indicadores y variables macro hasta del 13 de diciembre del 2019. La ventaja del nowcasting es

que a medida que se publica nueva información respecto a los indicadores y variables macro, disminuye gradualmente el error de predicción y elimina los rezagos de publicación de la información macroeconómica en el país.

Conclusiones

La evolución del PIB en cualquier economía del mundo es fundamental para hacedores de política, autoridades económicas, los agentes privados y para el gobierno, no obstante, por su naturaleza macroeconómica su publicación generalmente presenta rezagos de hasta más de 50 días (en el caso de Colombia) a partir de la finalización de un trimestre por la entidad estadística encargada. Por consiguiente, resulta de vital importancia un diagnóstico de la economía en “tiempo real”, que en el documento se denominó “**nowcasting**”, siendo esto una combinación entre el presente (now) y el futuro (forecasting).

La integración del nowcasting y el Machine Learning, especialmente las redes neuronales, sumado a la selección adecuada de los principales indicadores y variables macro como el Indicador de Seguimiento de la Economía (ISE), ventas de comercio minorista, el índice de confianza industrial y comercial, el número de desocupados y las construcciones de obras nuevas, entre otros, teniendo en cuenta cada uno sus correlaciones dinámicas más fuertes desde el primer trimestre del año 2005 hasta el segundo trimestre del año 2019, forman un modelo de diagnóstico del PIB en “tiempo real” consistente, preciso y de fácil aplicación por parte de los agentes económicos, eliminando las brechas de información que se presentan en la divulgación de información macroeconómica por parte de las entidades competentes.

En ambos modelos encontramos una proyección trimestral y anual del PIB que mejora significativamente por la inclusión de los indicadores macroeconómicos líderes (Tabla 1) y el uso de redes neuronales, que de forma conjunta generan consistencia, precisión y fortalece nuestro modelo de nowcasting, siendo un mecanismo eficiente para medir el ritmo de la actividad económica del país de forma más oportuna, lo que implica un gran avance para la economía en la toma de decisiones de los agentes.

Referencias

- Banbura, M., D. Giannone, M. Modugno, and L. Reichlin. (2013). "Nowcasting and the Real-Time Data Flow." In G. Elliott and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier-North Holland.
- Baquero, D. & González Manuel. (2019). A nowcasting model for Ecuador: Implementing a time-varying mean output growth. *Journal Economic Modelling Volume 82, November 2019, Pages 250-263*
- Banco central de reserva del Perú (2019). SPREAD - EMBIG Colombia (PBS). <https://estadisticas.bcrp.gob.pe/estadisticas/series/diarias/resultados/PDO4709XD/html>
- Banco de la República (2019). Estadísticas. Recuperado <https://www.banrep.gov.co/es/estadisticas>
- Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti. (2017). "Macroeconomic Nowcasting and Forecasting with Big Data." Federal Reserve Bank of New York Staff Reports, no. 830, November.
- Bonaldi, P., Prada, J. D., González, A., Rodríguez, D., & Rojas, L. E. (2011). Método numérico para la calibración de un modelo DSGE. *Revista Desarrollo y Sociedad*, (68), 119-156.
- Casares, F. (2017). NOWCASTING: Modelos de factores dinámicos y ecuaciones puente para la proyección del pib del ecuador. *COMPENDIUM, ISSN Impresa 1390-8391, ISSN Online 1390-9894, Volumen 4, N° 8, Agosto, 2017, pp 25 – 46*
- Chakraborty, C. & Joseph, A. (2017). Machine learning at central banks. *Staff Working Paper*, 674, Bank of England
- Chaudhuri, T.D. & Ghosh, I. (2016). Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework. *Journal of Insurance and Financial Management*, 1 (5), 92-123
- Comisión para América Latina y el Caribe (CEPAL) (2018). La Inversión Extranjera Directa en América Latina y el Caribe. pp 9-131
- Departamento Administrativo Nacional de Estadística (DANE) (2019). Estadísticas por tema. Recuperado de: <http://www.dane.gov.co/index.php/estadisticas-por-tema/comercio-internacional>

- Departamento Nacional de Planeación (DNP) (2019). Estadísticas comercio exterior. Recuperado de: <https://www.dnp.gov.co/programas/desarrollo-empresarial/comercio-exterior-e-inversion-extranjera/Paginas/estadisticas.aspx>
- Fedesarrollo (2019). Encuestas de opinión empresarial, consumidor y financiera. <https://www.fedesarrollo.org.co/encuestas/empresarial-oe>
- García-Cicco, Javier, Roberto Pancrazi, and Martín Uribe. 2010. Real Business Cycles in Emerging Countries?. *American Economic Review*, 100 (5): 2510-31.
- Hodrick, R & Prescott, E (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, Vol. 29, No. 1 (Feb. 1997), pp. 1-16.
- International Monetary Fund, Global price of Brent Crude [POILBREUSDQ] & price of WTI, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/POILBREUSDQ>, December 11, 2019.
- León, C. & Ortega, F. (2018). Nowcasting economic activity with electronic payments data: A predictive modeling approach. *Borradores de economía*, Núm.1037, 2018.
- MathWorks (2019). MATLAB to machine learning. <https://es.mathworks.com/machine-learning.html>
- Parra-Alvarez, J. C. (2008). Hechos estilizados de la economía colombiana: fundamentos empíricos para la construcción y evaluación de un modelo DSGE. *Borradores de Economía*; No. 509
- Sarmiento, P., & David, J. (2005). A real business cycle model for Colombia.
- Scikit-Learn (2019). Machine Learning in Python. Recuperado de: https://scikit-learn.org/stable/modules/neural_networks_supervised.html
- Tiffin, A. (2016). Seeing in the dark: A machine-learning approach to nowcasting in Lebanon. IMF Working Paper, WP/16/56, International Monetary Fund.
- World Bank Group. (2019). DataBank. Recuperado de: <https://datos.bancomundial.org/>



Escuela de Economía y Finanzas
Centro de Investigaciones Económicas y Financieras
Grupo de investigación en Estudios en Economía y Empresa
Línea de Macroeconomía Aplicada

Carrera 49 N° 7 Sur-50, Medellín - Colombia
Teléfono: (057-4) 261 9500 Ext 9532 - 2619532
cief@eafit.edu.co

Vigilada Mineducación