

N

owcasting de la actividad económica en el Perú:

¿QUÉ TAN PRECISOS PODEMOS SER?

BRUNO GONZAGA*, WALTER RUELAS**,
JAIRO FLORES*** Y JUAN TANG****



* Jefe, Departamento de Análisis de Coyuntura del BCRP
bruno.gonzaga@bcrp.gob.pe



** Supervisor especializado, Departamento de Análisis de Coyuntura del BCRP
walter.ruelas@bcrp.gob.pe



*** Especialista en investigación económica, Departamento de Economía Mundial del BCRP
jairo.flores@bcrp.gob.pe



**** Especialista, Departamento de Indicadores de la Actividad Económica del BCRP
juan.tang@bcrp.gob.pe

En este artículo se explora el uso de modelos de *machine learning* para la predicción inmediata del PBI total y no primario. Estos modelos son comparados con un modelo de factores dinámicos (MFD). El preprocesamiento es considerado como un hiperparámetro adicional que se optimiza iterativamente en cada predicción. En el periodo de prueba, XGBoost destaca como el mejor modelo para la predicción del PBI total, y los modelos de *machine learning* presentan un desempeño similar al MFD para el PBI no primario.

INTRODUCCIÓN

Conocer el estado actual de la economía es una tarea clave para los responsables de la política económica, quienes deben tomar decisiones bajo el conjunto de información más actualizado posible. La predicción de corto plazo o *nowcasting* de la actividad económica es el método que permite tener una idea acertada del estado de la economía en tiempo real. Este artículo resume parte de la investigación realizada por Flores et al. (2024), en la que se emplean modelos de aprendizaje automático (*machine learning*, ML) para la predicción de corto plazo del PBI y del PBI no primario.

La literatura académica viene empleando crecientemente los modelos de ML para la predicción de variables macroeconómicas y financieras. Esto se debe a la capacidad que tienen estos modelos para capturar relaciones no lineales entre los predictores y la variable objetivo y, en un contexto con mayor disponibilidad de datos, de utilizar un abanico más grande de información que los modelos econométricos tradicionales.

Para el caso peruano, sin embargo, la única referencia cercana a esta nueva vertiente es la de Tenorio y Perez (2023), quienes aplican modelos de ML para la predicción del PBI total. Este trabajo amplía la referida literatura al explorar nuevas formas de construir modelos de ML para la predicción inmediata de la actividad económica.

METODOLOGÍA

Los modelos de ML dependen de hiperparámetros: variables que deben definirse previamente al ajuste de cada modelo a los datos. Encontrar un conjunto de hiperparámetros óptimo es una tarea crucial para asegurar la mejor generalización posible del modelo a datos no vistos previamente. Para ello, una estrategia común es dividir la muestra en tres secciones: entrenamiento, validación y prueba.



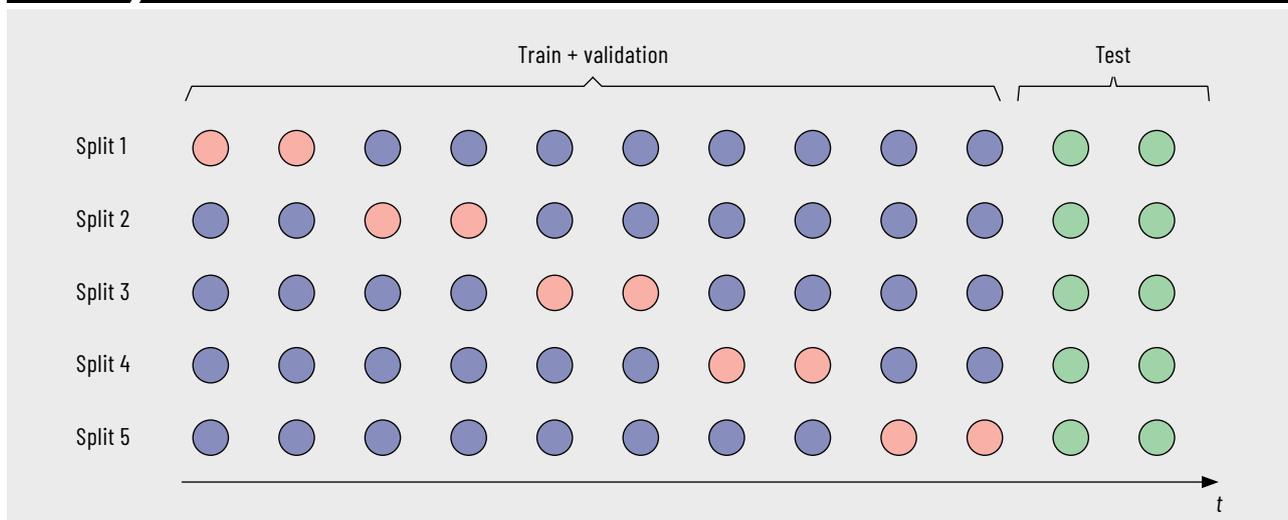
La predicción de corto plazo o *nowcasting* de la actividad económica **es el método que permite tener una idea acertada del estado de la economía en tiempo real.**



Al entrenar el modelo bajo distintos hiperparámetros en la sección de entrenamiento y observar su desempeño en la sección de validación, es posible obtener un conjunto de hiperparámetros que, potencialmente, podrán desempeñarse bien en la sección de prueba. En el Gráfico 1 se muestra una de las estrategias de validación cruzada más utilizadas en la literatura, que consiste en ajustar el modelo múltiples veces (tantas combinaciones de hiperparámetros como deseemos probar) en “K” submuestras de entrenamiento y evaluar su desempeño en las “K” submuestras de validación.

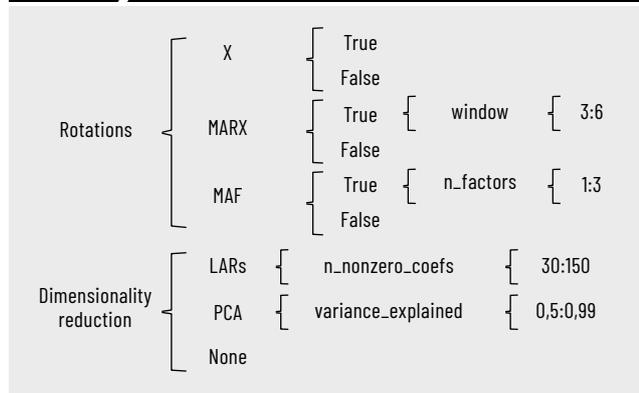
De manera más general, un hiperparámetro también es todo preprocesamiento a los datos antes de

GRÁFICO 1 ■ Validación cruzada K-Fold (5 folds)



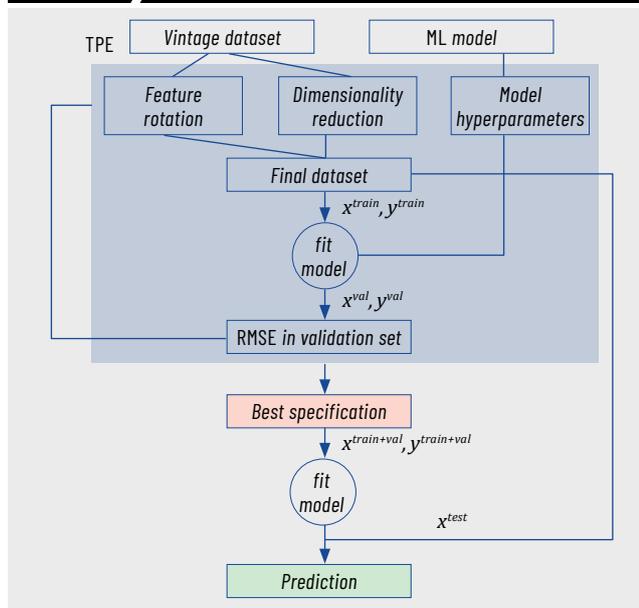
NOTA: ESTA ESTRATEGIA CONSISTE EN CREAR “K” MUESTRAS FICTICIAS A PARTIR DE UN SOLO CONJUNTO DE INFORMACIÓN (TRAIN + VALIDATION). LUEGO, PARA DISTINTAS CONFIGURACIONES DE HIPERPARÁMETROS, SE AJUSTA EL MODELO SOBRE LOS PUNTOS CELESTES Y SE EVALÚA EL ERROR SOBRE LOS PUNTOS ROJOS. LA CONFIGURACIÓN QUE CONSIGA EL MENOR ERROR ES LA ESCOGIDA PARA PREDECIR LOS PUNTOS VERDES.

GRÁFICO 2 ■ Mapa de especificaciones posibles para el preprocesamiento de datos



NOTA: SE MUESTRAN LAS POSIBLES ELECCIONES DE DOS HIPERPARÁMETROS COMÚNMENTE USADOS EN LA LITERATURA: LA REDUCCIÓN DE DIMENSIONALIDAD (DIMENSIONALITY REDUCTION) Y LAS TRANSFORMACIONES AL CONJUNTO DE VARIABLES (ROTATIONS).

GRÁFICO 3 ■ Esquema del procedimiento de predicción



NOTA: SE MUESTRA EL PROCEDIMIENTO UTILIZADO PARA REALIZAR PREDICCIONES A PARTIR DE UN CONJUNTO DE INFORMACIÓN (VINTAGE DATASET) Y UN MODELO (MODEL). EL EJERCICIO ASUME QUE EL PREPROCESAMIENTO DE LAS VARIABLES EXPLICATIVAS Y LA TÉCNICA DE REDUCCIÓN DE DIMENSIONALIDAD SON HIPERPARÁMETROS ADICIONALES A LOS PROPIOS DEL MODELO DE ML, LOS CUALES SON ENCONTRADOS DE MANERA ÓPTIMA PARA LA PREDICCIÓN.

que estos ingresen a un modelo. Ello incluye las transformaciones hechas a los datos, la selección de rezagos, la preselección de variables, la omisión de observaciones con extrema variabilidad, entre otros. En principio, no es claro cuáles son las estrategias a utilizar o si se debería utilizar alguna. En este documento, se consideran como hiperparámetros adicionales las transformaciones de los datos y la técnica de reducción de dimensionalidad. Por un lado, siguiendo a Coulombe et al. (2021), las transformaciones incluyen agregar variables en niveles (X), promedios móviles de

las variables (*moving average rotation of X*, MARX) y factores por cada grupo de variable-rezagos (*moving average factors*, MAF). Por el otro, la reducción de dimensionalidad implica el uso de *least angle regression* (LARS), análisis de componentes principales (PCA) o ningún método (*none*). Este mapa de especificaciones se resume en el Gráfico 2.

Bajo estas posibles especificaciones, se entrena un conjunto de 12 modelos de ML para la predicción mensual del PBI total y del PBI no primario. Con el objetivo de evaluar la mejor metodología, esta predicción se hace de manera directa (cada variable objetivo de manera individual) e indirecta (con un *nowcast* por cada uno de los sectores¹ que componen cada variable objetivo).

Así, para cada variable objetivo, modelo de ML y periodo de predicción, se escoge la mejor transformación a los datos y técnica de reducción de dimensionalidad, así como los hiperparámetros óptimos de cada modelo de ML, a través del algoritmo de *tree-structured Parzen of estimators* (TPE). Este algoritmo, propuesto en Bergstra et al. (2011), permite buscar de manera eficiente la combinación de hiperparámetros que minimiza una función objetivo —que en este caso es el error cuadrático medio en el conjunto de validación (Gráfico 1)— dado un espacio de búsqueda para dichos hiperparámetros.

En términos simples, el TPE reduce el número de evaluaciones necesarias a la función objetivo para encontrar el mejor conjunto de hiperparámetros en un tiempo determinado respecto a otros algoritmos de búsqueda, al proponer candidatos prometedoros cada vez que se realiza una nueva evaluación. Esto se logra almacenando evaluaciones previas y construyendo funciones de densidad para cada hiperparámetro iterativamente, a partir de las cuales se extraen candidatos con una mayor probabilidad de minimizar la función objetivo. Una vez que se obtiene la configuración de hiperparámetros óptima, se realiza un ajuste del modelo con estos hiperparámetros sobre el conjunto de entrenamiento y validación, y se predice sobre el conjunto de prueba. El esquema del procedimiento que se aplica para la predicción de cada variable objetivo puede observarse en el Gráfico 3.

DATOS

El conjunto de datos comprende una amplia gama (170) de variables macroeconómicas y del mercado financiero, lo que proporciona una base integral para la predicción del PBI peruano. Estas variables incluyen indicadores generales de la actividad interna, de la balanza de pagos, del mercado financiero nacional e internacional, datos de encuestas del Banco Central de Reserva del Perú, índices de precios e indicadores climáticos. De acuerdo con Tenorio y Perez (2023), se

¹ En particular, se realizan *nowcasts* de los sectores comercio, servicios, manufactura primaria, manufactura no primaria y agropecuario. Los sectores pesca, construcción y minería e hidrocarburos no necesitan predecirse, pues su información es conocida con antelación a la publicación del PBI, mientras que para el sector de electricidad y agua se asume el crecimiento de la producción de electricidad a nivel nacional.

mejora el conjunto de información integrando datos no estructurados de Google Trends. Esta fuente proporciona información en tiempo real sobre el comportamiento del consumidor y el sentimiento económico.

Las variables son ajustadas estacionalmente —en caso de que exista evidencia de que presentan tal componente— para remover cualquier ruido que pueda afectar la identificación de los patrones subyacentes que las relacionen con la variable objetivo.

RESULTADOS

El Cuadro 1 muestra que aplicar técnicas de reducción de dimensionalidad es preferible a no aplicar este paso, y que el uso de LARS es preferible a utilizar PCA. Asimismo, para la mayoría de los modelos y predicciones realizadas, se prefiere la inclusión de promedios móviles de las variables independientes (MARX), lo que evidencia la utilidad de aplicar transformaciones a los datos para la reducción del error de predicción.

El Cuadro 2 resume la precisión de los métodos de ML para el periodo desde enero 2022 hasta agosto 2024. Se observa que es preferible utilizar una estrategia *bottom-up* para la predicción del PBI total y del PBI no primario: realizar ejercicios de *nowcast* individuales por sectores y luego agregar estas proyecciones conlleva a una mayor reducción del error cuadrático medio. Ello estaría asociado a ganancias de insumir el proceso generador de datos “por fuera” y no dejar que los modelos de ML intenten aprenderlo. Por otro lado, XGBoost resulta ser el modelo más preciso para la predicción *bottom-up* del PBI total, y los modelos de ML presentan un desempeño similar al MFD para la predicción *bottom-up* del PBI no primario.



Por otro lado,
XGBoost resulta ser el modelo
**más preciso para la predicción
bottom-up del PBI total, y los
modelos de ML presentan un
desempeño similar al MFD para
la predicción *bottom-up* del PBI no
primario.**



CUADRO 1 ■ Frecuencia de uso de hiperparámetros en los ejercicios fuera de la muestra*

Reducción de dimensionalidad			Rotaciones		
LARS	PCA	None	MARX	X	MAF
96,0	3,6	0,3	69,6	44,4	31,8

* SE MUESTRA, EN PORCENTAJE, LA FRECUENCIA DE USO DE LOS DISTINTOS MÉTODOS DE REDUCCIÓN DE DIMENSIONALIDAD Y ROTACIONES CONTEMPLADOS. EN TOTAL, SE CONDUJERON 12(NÚMERO DE MODELOS)*7(NÚMERO DE VARIABLES OBJETIVO)*32(PERIODOS DE PREDICCIÓN)=2688 EJERCICIOS FUERA DE LA MUESTRA.

CUADRO 2 ■ Raíz del error cuadrático medio enero 2022-agosto 2024

Modelo	PBI Total ¹	PBI Total ²	PBI No Primario ¹	PBI No Primario ²
Lasso	0,636	1,100	0,763	0,907
Ridge	0,621	1,123	0,737	0,907
Elastic Net	0,597	1,120	0,720	0,885
SVR	0,630	1,142	0,718	0,919
Decision Tree	0,696	1,302	0,719	1,035
KNN	0,773	1,127	0,742	0,952
RF	0,683	1,178	0,722	0,951
AdaBoost	0,647	1,154	0,768	0,966
GBoost	0,610	1,187	0,742	0,950
XGBoost	0,590	1,150	0,721	0,977
Bagging	0,613	1,181	0,731	0,979
MLP	0,634	1,108	0,743	0,885
Promedio ML	0,612	1,114	0,726	0,927
MFD	0,840	1,245	0,672	0,841

1/ MÉTODO INDIRECTO.
2/ MÉTODO DIRECTO.

GRÁFICO 4 ■ PBI total y nowcasts de modelos (método indirecto)

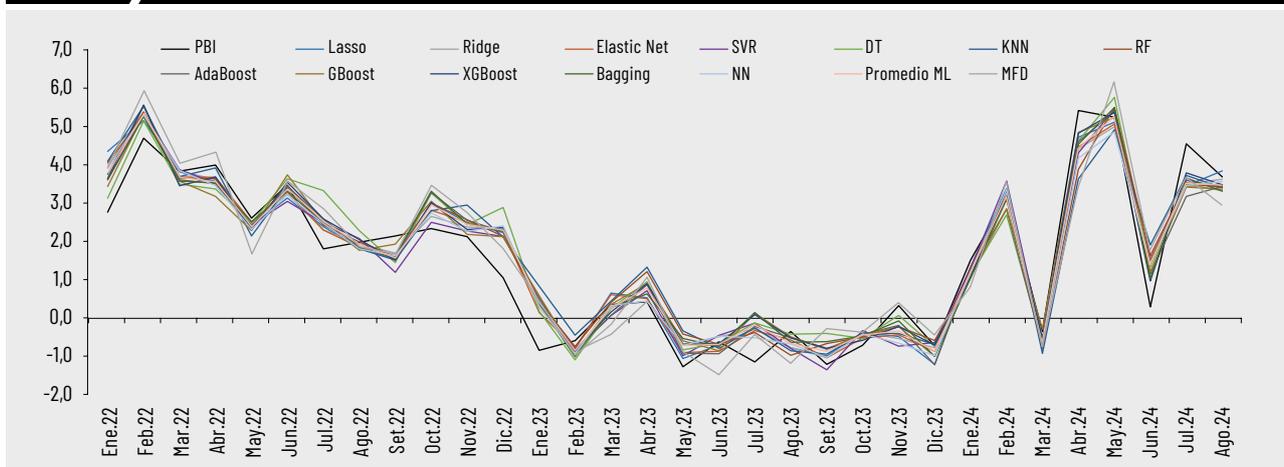
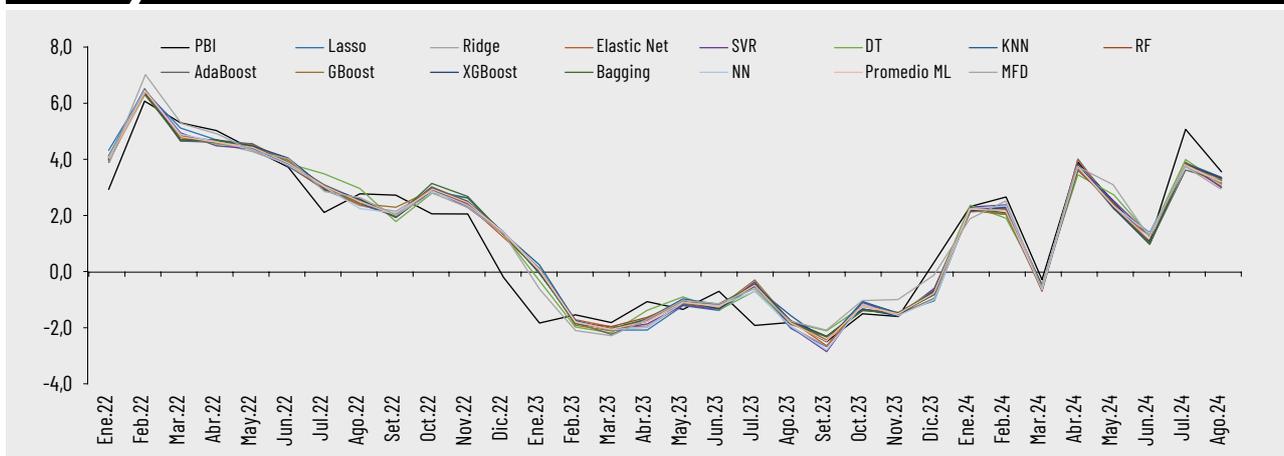


GRÁFICO 5 ■ PBI no primario y nowcasts de modelos (método indirecto)



El Gráfico 4 y el Gráfico 5 muestran el ejercicio de *nowcast* del PBI total y del PBI no primario, respectivamente, de cada modelo de ML, del promedio de estos modelos y del MFD empleando el método indirecto o *bottom-up*. Resalta la dificultad de los métodos de ML para realizar el *nowcasting* del PBI y del PBI no primario durante diciembre de 2022 y enero de 2023, época que se caracterizó por la ocurrencia de conflictos sociales en el país.

CONCLUSIONES

Los resultados obtenidos muestran que los modelos de ML son herramientas útiles para la predicción inmediata del PBI total y del PBI no primario de Perú. Al ser provistos de un amplio conjunto de datos, los modelos son capaces de combinar la información contenida en estos de manera valiosa. Especificaciones asociadas al preprocesamiento son optimizables al concebirlas como hiperparámetros, lo que puede ayudar a reducir el error de predicción. Asimismo, utilizar una estrategia *bottom-up* para la predicción de las variables objetivo es preferible a realizar una predicción directa, lo que está asociado

a ganancias de insumir el proceso generador de datos por fuera. En el periodo de prueba, XGBoost y el MFD destacan como los mejores modelos para la predicción *bottom-up* del PBI total y no primario, respectivamente, si bien la mayoría de los modelos de ML presenta un desempeño similar al MFD para este último caso.

REFERENCIAS

- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyperparameter optimization. En Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., & Weinberger, K. Q. (eds.). Advances in Neural Information Processing Systems 24 (2546-2554). Recuperado de: https://proceedings.neurips.cc/paper_files/paper/2011/file_e/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf
- Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2021). Macroeconomic data transformations matter. *International Journal of Forecasting*, 37(4), 1338-1354.
- Flores, J., Gonzaga, B., Ruelas-Huanca, W., & Tang, J. (2024). *Nowcasting Peruvian GDP with Machine Learning Methods*. Documento de Trabajo 2024-019. BCRP. <https://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2024/documento-de-trabajo-019-2024.pdf>
- Tenorio, J., & Perez, W. (2023). *GDP nowcasting with machine learning and unstructured data to Peru*. Documento de Trabajo 2024-003. BCRP. <https://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2024/documento-de-trabajo-003-2024.pdf>