

CAPÍTULO VIII

Predicción de series temporales basada en *Machine Learning*: aplicaciones económicas y financieras

Lorenzo Pascual
Esther Ruiz*

En este capítulo se realiza una breve revisión de aplicaciones empíricas de procedimientos de predicción basados en Machine Learning (ML) en el contexto de series temporales económicas y financieras. Dada la naturaleza reciente y cambiante de dichos procedimientos, esta revisión es limitada e incompleta. Nuestra contribución es la descripción de los ámbitos de aplicación donde ML ha sido utilizado con éxito.

Palabras clave: árboles de decisión, big data, LASSO, redes neuronales, regularización.

* Esther Ruiz agradece la ayuda financiera de la Agencia Estatal de Investigación (PID2019-108079GB-C21/AIE/10.13039/501100011033).

“So the question is, what is the real world? Is it simple or complex? Machine learning shows that there are examples of complex worlds. We should approach complex worlds from a completely different position than simple worlds. For example, in a complex world one should give up explain-ability (the main goal in classical science) to gain better predictability.”

Vapnik (2006)

1. INTRODUCCIÓN

La predicción es un instrumento fundamental para la toma de decisiones en un amplio abanico de aplicaciones entre las que naturalmente están áreas del ámbito económico como, por ejemplo, organización de empresas, macroeconomía, microeconomía y finanzas. La importancia de la predicción explica el gran interés académico y práctico por la metodología cuyo objetivo final es obtener procedimientos para predecir fenómenos futuros que sean lo más precisos posible.

Los procedimientos “tradicionales” de predicción se basan en modelos estadísticos relativamente sencillos que tratan de aproximar de la mejor forma posible el supuesto proceso generador de la variable que se quiere predecir. Estos modelos pueden ser lineales o no lineales, univariantes o multivariantes, paramétricos o no paramétricos, estacionarios o no estacionarios, pero, en cualquier caso, la estimación de sus parámetros puede realizarse utilizando datos con dimensiones relativamente pequeñas o medianas. Sin embargo, en las últimas décadas, dos fenómenos han propiciado la aparición de nuevos procedimientos de predicción basados en lo que se conoce como *Machine Learning*¹ (ML) supervisado². En primer lugar, la capacidad de ejecución en tiempo real de los ordenadores y, por lo tanto, su velocidad de procesamiento, han crecido de forma espectacular durante las últimas décadas, disminuyendo drásticamente el coste de computación. En segundo lugar, la arquitectura de sistemas permite el almacenamiento y tratamiento de cantidades ingentes de datos, lo que se conoce como *big data* (BD), que pueden clasificarse en datos estructurados, obtenidos de forma habitual mediante representaciones numéricas, y datos no estructurados, como textos, vídeos o imágenes; ver, por ejemplo, Garboden (2019) para una descripción de datos BD en el contexto de la predicción macroeconómica. La principal diferencia entre los procedimientos de predicción tradicionales y los basados en ML es que estos últimos utilizan

¹ Queremos pedir disculpas por la extensa utilización de términos en el idioma inglés a lo largo de este capítulo. La razón principal es que los conceptos asociados a dichos términos son conocidos cuando se utiliza el inglés, pero se crearían confusiones si utilizáramos traducciones literales al español que no son habituales en esta literatura. Nuestras disculpas por adelantado a la Real Academia Española.

² Supervisado se refiere a que la variable a predecir sirve para guiar el proceso de estimación. Aunque los procedimientos no supervisados quedan fuera de esta revisión, hay aplicaciones muy interesantes donde dichos procedimientos se han implementado para la predicción económica; ver, por ejemplo, Aromi (2020) que construye un indicador de incertidumbre basado en el contenido publicado en el *Wall Street Journal*, para predecir el ciclo económico en USA.

modelos estadísticos y algoritmos flexibles que permiten extraer información de datos de alta dimensión pero que, en general, no tratan de representar el mecanismo que ha generado los datos (únicamente se predice, no se trata de explicar). Sin embargo, es importante señalar que no existe una frontera bien definida entre los llamados procedimientos tradicionales de predicción y los procedimientos ML; ver, por ejemplo, las discusiones en Barker (2020), Januschowski *et al.* (2020) e Israel, Kelly y Moskowitz (en prensa) para excelentes descripciones de las diferencias entre procedimientos “tradicionales” y ML.

Además de en predicción, ML supervisado tiene importantes aplicaciones en problemas de clasificación. En algunos casos, no existe una clara frontera entre predicción y clasificación. Por ejemplo, cuando se clasifica a un cliente en función de si va a devolver o no un crédito, se podría considerar como una predicción de una variable cualitativa binaria. Sin embargo, en este capítulo, nos centraremos en la predicción de variables cuantitativas continuas. Existen también procedimientos no supervisados de ML que se utilizan principalmente para reducir la dimensión de los datos y para encontrar grupos homogéneos. Estos procedimientos de procesamiento de datos suelen utilizarse como paso previo a la implementación de procedimientos supervisados para la predicción. Los procedimientos más populares dentro de ML no supervisado son componentes principales, análisis factorial y análisis de conglomerados.

Los algoritmos de ML son computacionalmente intensivos y están diseñados para identificar de forma sistemática patrones y relaciones en los datos que sirvan de base para realizar la predicción. La mayoría de los procedimientos ML para la predicción se han desarrollado en el ámbito de variables independientes e idénticamente distribuidas (iid); véase, por ejemplo, Varian (2014) para regresiones supervisadas o Biau y Scornet (2016) para árboles de decisión. Sin embargo, en este capítulo nos centraremos en procedimientos de predicción para datos dinámicos en los que los propios retardos de la variable a predecir pueden aparecer entre los predictores. Nuestro objetivo es revisar brevemente las principales aplicaciones empíricas de procedimientos de predicción basados en ML en el contexto de series temporales económicas y financieras. Dada la naturaleza reciente y cambiante de dichos procedimientos, esta revisión es necesariamente limitada e incompleta. Por ejemplo, para limitar la discusión, no hemos incluido en esta revisión aplicaciones de procedimientos bayesianos para el tratamiento de BD; ver, por ejemplo, las descripciones en Varian (2014) y Scott y Varian (2014).

El resto del capítulo se organiza como sigue. La sección segunda describe brevemente los principales procedimientos de predicción con ML en el contexto de series temporales. La sección tercera describe aplicaciones empíricas de ML en el contexto de la predicción económica y financiera. Finalmente, la sección cuarta concluye.

2. PROCEDIMIENTOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE SERIES TEMPORALES

2.1. El problema de predicción

El objetivo es obtener predicciones de y_{T+h} , donde h es el horizonte de predicción, basadas en observaciones $\{y_1, \dots, y_T\}$, obtenidas normalmente en espacios equidistantes del tiempo. La predicción es univariante cuando y_{T+h} es un escalar, es decir, el objetivo es la predicción del valor futuro de una única variable, mientras que es multivariante cuando y_{T+h} es un vector. Para realizar la predicción de y_{T+h} se puede disponer además de observaciones de un conjunto de predictores, $x_{i,t}$, $i=1, \dots, N$, $t=1, \dots, T$. Al hablar de gran dimensión de los predictores, podemos encontrarnos con dos situaciones diferentes. En primer lugar, están los datos “anchos” en los que el número de predictores, N , es muy grande relativo al número de observaciones, T . Los datos pueden ser “largos” cuando hay muchas observaciones en relación al número de predictores; ver Diebold *et al.* (2019) que también definen los datos “densos” cuando la frecuencia de observación es muy elevada. El problema de predicción es obtener predicciones h -periodos-hacia-adelante mediante el siguiente procedimiento directo³:

$$\hat{y}_{T+h|T} = g_h(y_T, \dots, y_1, x_{1,T}, \dots, x_{1,1}, \dots, x_{N,T}, \dots, x_{N,1}), \quad [1]$$

dónde $g_h(\cdot)$ es una función, posiblemente no lineal y desconocida, de los datos disponibles en el momento T ; ver, por ejemplo, Gu, Kelly y Xiu (en prensa b), Huber y Stuckenschmit (en prensa), Varian (2014) y Caro y Peña (2021) para una detallada descripción del problema de predicción de series temporales basado en procedimientos ML. Para simplificar la descripción posterior, supondremos que las predicciones se realizan un-periodo-hacia-adelante, es decir, $h = 1$, y suprimiremos el subíndice de la función $g_h(\cdot)$. Los procedimientos ML para la predicción tratan de obtener predicciones de y_{T+1} lo más precisas posible cuando el modelo verdadero, es decir, la función $g_h(\cdot)$, es desconocido y/o muy complejo. Es por ello que estos procedimientos están especialmente diseñados para realizar predicciones cuando las bases de datos son de gran dimensión y, por lo tanto, es esperable que pueda haber no-linealidades y relaciones complejas difíciles de especificar *a priori*.

Los procedimientos de predicción ML requieren, en general, seleccionar una arquitectura (por ejemplo, en los procedimientos de *Artificial Neural Networks (ANN)*, hay que seleccionar el número de nodos y capas y, en los procedimientos *Support Vector Machine (SVM)*, hay que seleccionar las funciones kernel) y/o algunos parámetros, denominados hiperparámetros (por ejemplo, en los procedimientos de regularización hay que selec-

³ Coulombe *et al.* (2020) señalan que el procedimiento directo es el más habitual en ML frente al procedimiento iterativo que se utiliza frecuentemente cuando las predicciones se basan en modelos “tradicionales” de series temporales.

cionar previamente los parámetros de regularización), que reducen la complejidad del modelo y que deben ser seleccionados con anterioridad a la propia estimación de los parámetros del modelo correspondiente. El procedimiento de selección de los hiperparámetros, estimación de los parámetros y evaluación de las predicciones se basa en dividir la muestra en tres submuestras: la submuestra de entrenamiento, en la que, dados los hiperparámetros, se estiman los parámetros; la submuestra de validación, en la que se eligen los hiperparámetros mediante la minimización de una determinada función de pérdida; y la submuestra de contraste, en la que se evalúan las predicciones. La figura 1 resume el procedimiento de estimación, validación y predicción de los procedimientos ML. Una vez seleccionados los hiperparámetros, los parámetros del modelo vuelven a estimarse utilizando conjuntamente las submuestras de entrenamiento y de validación. La submuestra de contraste es la que habitualmente se conoce como periodo “fuera-de-muestra”. En una gran mayoría de aplicaciones de predicción con ML, los hiperparámetros son seleccionados mediante validación cruzada y las predicciones son evaluadas utilizando el error de predicción cuadrático medio (EPCM); ver, por ejemplo, Barrow y Crone (2016b) sobre la validación cruzada y Coulombe *et al.* (2020) para una comparación con otras medidas de evaluación. Es importante señalar que, en los procedimientos de predicción ML, tanto la muestra de entrenamiento como la de prueba deben ser suficientemente grandes como para permitir encontrar patrones en los datos.

FIGURA 1

PROCEDIMIENTOS DE PREDICCIÓN ML

1	T_1	T_1+1	T	T_1+1	$T+H$
Muestra de entrenamiento (estimación): Dados los hiperparámetros, se estima el modelo.	Muestra de validación (selección del modelo): Selección de hiperparámetros que minimicen la función de pérdida de las predicciones obtenidas con el modelo estimado en la muestra de entrenamiento.		Muestra de contraste (predicción): Evaluación de las predicciones fuera-de-muestra: EPCM.		
Hiperparámetros: Regularización: Constantes Arboles: Nº de ramas Redes Neuronales: Función de activación	Función de pérdida: Error de predicción cuadrático medio (EPCM) Procedimiento: Validación cruzada				

Fuente: Elaboración propia.

Los procedimientos ML de predicción pueden clasificarse en dos grandes grupos dependiendo de que traten de reducir la dimensión de los predictores en [1] o de que traten de representar funciones $g(\cdot)$ complejas. Al primer grupo pertenecen los modelos de regresión con regularización y los procedimientos basados en SVM que se utilizan habitualmente para datos “anchos”. El segundo grupo está compuesto por los populares

modelos basados en árboles de decisión y por las redes neuronales o *deep learning*. A continuación se describen brevemente los procedimientos más populares dentro de cada uno de estos dos grandes grupos.

2.2. Reducción de la dimensión: regresión con regularización y Support Vector Machine

Cuando el número de predictores en un modelo de regresión es muy elevado, algunos de ellos pueden ser redundantes o no tener suficiente información como para ser incluidos en el modelo predictivo. En estos casos, antes de estimar, es necesario regularizar, es decir, forzar a los parámetros de algunos predictores a que tomen valores relativamente pequeños y/o seleccionar algunos predictores, para poder recuperar grados de libertad y mejorar la capacidad predictiva del modelo. Cuando la función $g(\cdot)$ es lineal, el modelo predictivo es el siguiente modelo de regresión

$$y_{T+1} = \beta' z_T + \varepsilon_{T+1}, \quad [2]$$

donde $Z_T = (y_{T-p}, \dots, y_{T-p}, x_{1,T-p}, \dots, x_{1,T-p}, \dots, x_{N,T-p}, \dots, x_{N,T-p})$ es el vector de predictores, con p siendo el número de retardos en el modelo; ver Eklund y Kapetanios (2008) para una descripción de los procedimientos para grandes bases de datos basados en el modelo [2]. Los $(N+1) \times p$ parámetros desconocidos, β , son estimados introduciendo una penalización convexa para forzar los parámetros en el vector β hacia cero en lo que se conoce como regresión con regularización. Entre los procedimientos de regresión con regularización más populares están la regresión *ridge* que restringe los parámetros hacia zero, reduciendo el sobreajuste, pero no selecciona predictores y, por lo tanto, no se genera un modelo más interpretable. Alternativamente, en el estimador conocido como *Least Absolute Shrinkage and Selection Operator (LASSO)*, se produce tanto la reducción de los parámetros como la selección de variables, mejorando tanto la precisión de las predicciones como la interpretabilidad del modelo al seleccionar solamente un subconjunto de los predictores disponibles. Cuando entre los predictores aparecen tanto un número alto de variables como sus retardos, Simon *et al.* (2013) introducen el *sparse group LASSO (sg-LASSO)*. El sg-LASSO reduce la dimensión tanto entre variables como entre sus retardos. Babii, Ghysels y Striankas (2020) consideran sg-LASSO cuando las variables en la regresión en [2] están observadas con distintas frecuencias y analizan las correspondientes propiedades del estimador y predictor. Otro procedimiento de regresión con regularización muy popular es *elastic net*, en el que se introduce una segunda penalización.

Finalmente, dentro de los modelos que tratan de estimar [2] está SVM, un procedimiento de minería de datos originalmente desarrollado por Vapnik en los Laboratorios Bell en 1995 para clasificación. Cuando se aplica a regresión, SVM se conoce como *Support Vector Regression (SVR)*; ver, por ejemplo, Smola y Scholkopf (2004) y Awad

y Khanna (2015) para descripciones detalladas de SVR no solo aplicadas a regresiones lineales si no también no-lineales.

2.3. Funciones de predicción desconocidas y/o complejas

En esta subsección, describimos procedimientos ML diseñados para obtener predicciones cuando la función $g(\cdot)$ en [1] es compleja y/o desconocida. Entre los procedimientos más populares están los árboles de decisión y los tres procedimientos propuestos para mejorar su comportamiento predictivo: 1) *Bagging*; 2) *Random Forest*; y 3) *Boosting*,⁴ y las redes neuronales.

2.3.1. Árboles de decisión

La idea básica de un árbol de decisión es segmentar el rango de valores de los predictores en un número finito de subregiones de tal forma que, dentro de cada subregión, los predictores son más homogéneos y se pueden utilizar procedimientos muy simples para la predicción como, por ejemplo, la media muestral de todas las observaciones en cada subregión. En los árboles de decisión, las particiones en regiones son binarias para todos los predictores y sus retardos, es decir, de cada nodo, solo pueden salir dos ramas. Los umbrales se determinan sucesivamente para minimizar el EPCM. Finalmente, la predicción de y_{T+1} se obtiene utilizando el valor de los predictores a través del árbol que nos va dirigiendo mediante decisiones binarias a una de las subregiones (nodos terminales del árbol).

En la construcción de un árbol de decisión, se seleccionan los mejores predictores y retardos para minimizar el EPCM. Sin embargo, es frecuente que, en este proceso, se obtengan árboles con muchas ramas que pueden generar sobreajuste y, consecuentemente, predicciones no óptimas. Por esta razón, en la práctica, se realiza una "poda" del árbol introduciendo en la función objetivo una penalización sobre el número de nodos terminales.

Bagging: el procedimiento *Bagging* (*Bootstrap Aggregation*) está diseñado para evitar el gran EPCM que se suele observar en las predicciones obtenidas mediante árboles de decisión debido a que el número de ramas que se utiliza finalmente para la predicción es relativamente pequeño. *Bagging* se basa en generar B muestras *bootstrap* de los predictores y construir el árbol de decisión correspondiente en cada una de las réplicas *bootstrap*. Es muy importante tener en cuenta que, en el caso de datos temporales, el remuestreo para obtener réplicas *bootstrap* de la muestra de entrenamiento ha de realizarse con especial cuidado para no destruir la dependencia temporal existente en los datos. Existen en la literatura diversas técnicas *Bootstrap* que tienen en consideración esta característica como, por ejemplo, el *wild bootstrap* o el *block bootstrap*; ver, por

⁴ Es importante señalar que *Boosting* puede utilizarse también en contextos diferentes de los árboles de decisión. Sirve para mejorar cualquier regla de predicción simple.

ejemplo, Pan y Politis (2016). La predicción final es la media de todas las predicciones obtenidas en cada una de las réplicas *bootstrap*.

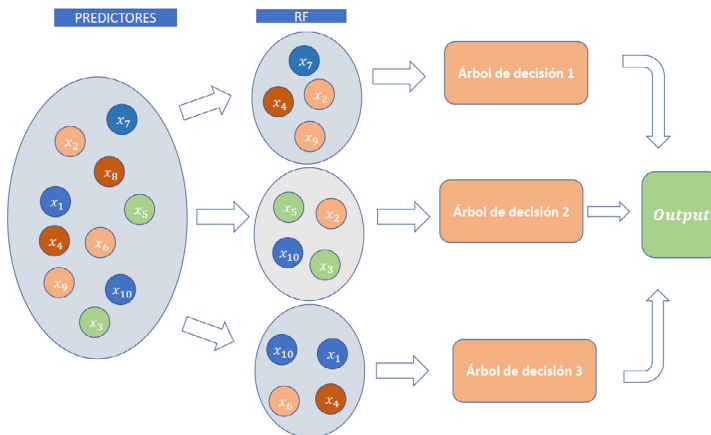
Bülmann y Yu (2002) han propuesto también el *Subbagging* basado en *subsampling* en lugar de Bootstrap y demuestran que aunque es computacionalmente más sencillo, tiene el mismo nivel de precisión que Bagging.

Random Forest: Random Forest (RF), propuesto originalmente por Breiman (2001), es un procedimiento diseñado para reducir la correlación entre árboles de decisión cuando estos son generados mediante Bagging, por el hecho de que los mismos predictores aparecen en todas las réplicas *bootstrap*. RF también se basa en construir árboles de decisión a partir de réplicas *bootstrap* de la muestra de entrenamiento. Sin embargo, mientras que Bagging utiliza en cada árbol todos los regresores existentes (p), RF selecciona de forma aleatoria, solamente $m < p$ de los p predictores, generalmente $m = \sqrt{p}$; ver la figura 2. Nótese que m vuelve a ser otro hiperparámetro que debe seleccionarse mediante validación cruzada en la muestra de entrenamiento. Junto con m , es necesario también determinar otros hiperparámetros como el número de réplicas Bootstrap o el número de árboles de decisión y el número mínimo de observaciones en los nodos terminales (poda). Tsay y Chen (2019) señalan que es necesario un estudio detallado de las propiedades de RF en el caso de que existan predictores fuertes dado que estos deben ser incluidos en cualquier modelo para obtener un buen ajuste.

Boosting: Mediante el procedimiento Boosting se generan árboles de forma secuencial a partir de un árbol inicial relativamente sencillo. En cada paso, se mejora el ajuste del

FIGURA 2

PROCEDIMIENTO DE SELECCIÓN DE ÁRBOLES MEDIANTE RANDOM FOREST



Fuentes: Elaboración propia.

árbol anterior mediante un nuevo árbol construido para los residuos en lugar de para la variable a predecir directamente. Con los ajustes a los residuos se mejora el árbol, mejorando áreas específicas donde el ajuste del árbol anterior no funciona correctamente. Al final del proceso, la predicción es una media ponderada de las predicciones de todos los árboles intermedios, dando más peso a los árboles últimos, que son los que están mejor calibrados. Nuevamente, el número de árboles, que es un hiperparámetro muy sensible que puede producir sobreajuste en el modelo, y la ratio de aprendizaje del Boosting, deben calibrarse mediante validación cruzada.

Existen variantes populares del Boosting, como el *Gradient Boosting*, que transcribe el método boosting como un algoritmo de optimización donde se minimiza una función de pérdidas adecuada y usa el método del gradiente en lugar de los residuos resultantes del ajuste previo para encontrar la siguiente mejora del árbol anterior. Hay que tener cierta precaución porque el Gradient Boosting tiende rápidamente al sobreajuste.

2.3.2. Redes neuronales artificiales

Las redes neuronales artificiales han ganado una gran popularidad en el área de predicción (a pesar del escepticismo sobre que las funciones de pérdida asociadas a estas redes son mayoritariamente no-convexas) debido a su excepcional comportamiento (ayudadas por procedimientos de gradiente descendente). Las ANN son procedimientos basados en datos con muy pocos supuestos previos sobre los modelos generadores de dichos datos. Las redes neuronales son modelos matemáticos computacionalmente intensivos que intentan imitar la forma en que los humanos aprenden cuando reciben información. Estos modelos se consideran "cajas negras" en términos de interpretabilidad con respecto a la relación entre las distintas variables. Sin embargo, su estructura permite capturar relaciones complejas entre los predictores y la variable a predecir sin que sea necesario especificar la forma correcta de dicha relación ya que la propia red neuronal tratará de identificar dichas relaciones a partir de los datos disponibles para su entrenamiento.

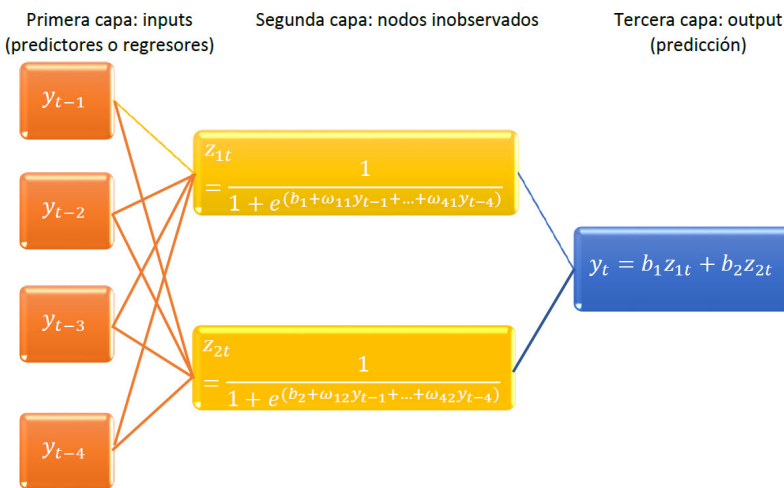
La primera aplicación de predicción utilizando ANN aparece en la tesis de Hu (1964) para la predicción meteorológica; ver Zhang, Patuwo y Hu (1998) para una revisión de la literatura sobre ANN a finales del siglo XX. Investigaciones recientes han validado los buenos resultados predictivos de las ANN mostrando que, si las redes son suficientemente grandes, casi todos los mínimos locales son muy similares al mínimo global; ver Choromanska *et al.* (2015). Una descripción detallada de las redes neuronales para la predicción de series temporales aparece en Palit y Popovic (2005).

Existen numerosos tipos de redes neuronales, siendo una de las más populares la conocida como "*multilayer feedforward networks*", que está definida por: (1) una capa de entrada, formada por varios nodos (o neuronas) que se limitan a recibir los valores de los

predictores; (2) una capa de salida relacionada con la variable que queremos predecir; y (3) entre ambas, una serie de capas ocultas cada una con un número de nodos que reciben información únicamente de los nodos de la capa inmediatamente anterior, y de cuyos nodos sale la información que servirá de *input* a cada uno de los nodos de la capa siguiente. Todos los nodos de una capa están conectados con todos y cada uno de los nodos de la capa siguiente, y los valores que los conectan y relacionan se conocen como “pesos”. Finalmente, en cada nodo es necesario definir una función de activación que transforma convenientemente la información que llega de cada uno de los nodos de la capa anterior. Entre las funciones de activación más populares están la lineal, la exponencial, y la logística o sigmoide⁵. En la figura 3 se representa una red neuronal autorregresiva con 4 retardos y una única capa con dos nodos.

FIGURA 3

RED NEURONAL AUTORREGRESIVA CON 4 RETARDOS Y UNA ÚNICA CAPA CON DOS NODOS



Fuente: Elaboración propia.

Para determinar todos los hiperparámetros que aparecen en la red se utiliza un procedimiento de estimación iterativo que utiliza los errores recursivamente para actualizar los pesos. En concreto, el error que se comete en el nodo final o de salida se distribuye hacia atrás a través de todos los nodos ocultos que llegan a dicho nodo final, de tal manera que a cada nodo se le asigna una parte de dicho error ajustando iterativamente

⁵ Cuando se utiliza la función de activación logística, las redes funcionan mejor cuando, tanto los predictores como la variable a predecir, están en la escala [0,1]. También se mejora considerablemente el resultado de una red si aquellos predictores que presenten una asimetría importante se transforman mediante la transformación logarítmica antes de convertirlos a la escala [0,1]. También es importante señalar que las redes no tienen un mecanismo de selección automática de los predictores que tienen que ser definidos y tratados previamente.

los pesos de cada nodo. Este procedimiento de optimización secuencial se conoce como "*back propagation of error*", y el hiperparámetro que controla el cambio o la actualización de los pesos de una iteración a la siguiente se conoce como ratio de aprendizaje, que suele frecuentemente fijarse en 0,1. El entrenamiento efectivo de una ANN sigue siendo una tarea difícil dado que el diseño de arquitecturas eficientes es más un arte que una metodología bien estructurada; ver De Stefani *et al.* (2019).

Nótese que una red neuronal sin capas ocultas, es decir, con una capa de entrada con tantos nodos como predictores, y una capa de salida con un único nodo y una función de activación lineal, es equivalente al modelo de regresión lineal. Si la función de activación es la logística, la red se corresponde con el modelo de regresión logística. Sin embargo, los estimadores de los parámetros (pesos) obtenidos a partir de la red neuronal pueden diferir de los obtenidos en los modelos clásicos, porque el procedimiento de estimación difiere de mínimos cuadrados ordinarios (MCO) para el modelo de regresión lineal y de máxima verosimilitud (MV) para el modelo de regresión logística.

Finalmente, debemos mencionar las redes neuronales recurrentes (RNN), que son especialmente interesantes en la predicción de series temporales; ver Lipton, Berkowitz y Elkan (2015) para una revisión crítica. Este tipo de redes fueron concebidas en la década de los 80, pero no se popularizaron en aquel momento por la dificultad de entrenarlas debido a sus elevados requerimientos computacionales. Las redes descritas anteriormente tienen una función de activación que solo actúa en una dirección, hacia adelante, desde una capa hacia la siguiente, es decir, redes que no recuerdan valores previos; ver, por ejemplo, la figura 3. Alternativamente, una RNN incluye conexiones que apuntan hacia atrás, es decir, permiten una especie de retroalimentación entre las neuronas dentro de una misma capa. De esta forma, cada neurona recibe dos entradas, la entrada correspondiente a la capa anterior y a su vez la salida del instante anterior dentro de la misma capa. Esto implica que cada neurona recurrente tiene dos conjuntos de hiperparámetros, uno que se aplica a la entrada de datos que recibe de la capa anterior y otro conjunto aplicado a la entrada de datos correspondiente al vector salida del instante anterior. De esta forma la estructura de estas redes permite una conexión recurrente en el tiempo, y se podría decir que una neurona recurrente tiene en cierta forma memoria.

En general, es importante notar que las redes necesitan un gran número de observaciones para su aprendizaje y las predicciones fuera del rango de valores de las variables con las que han aprendido pueden tener malas propiedades. Además, la convergencia del proceso de estimación a un mínimo global suele ser complicada, por lo que el proceso de búsqueda debe inicializarse en distintos puntos aleatorios. Las redes neuronales pueden tener problemas relacionados con sobreajuste y un gran esfuerzo computacional; ver, por ejemplo, Aras y Kocakoc (2016). Para solucionar estos problemas se han propuesto procedimientos como el *Extreme Learning Machine (ELM)*; ver, por ejemplo, Wang *et al.* (2018) para una descripción.

3. ALGUNAS APLICACIONES EN ECONOMÍA Y FINANZAS

Existe un creciente interés por la utilización de ML en la predicción de variables relacionadas con el área económica, tanto a nivel macro, micro o finanzas. En esta sección revisaremos brevemente algunas de estas aplicaciones para ilustrar el potencial de ML en la predicción económica; ver también Chakraborty y Joseph (2017) para una excelente revisión de la literatura sobre predicción ML en el contexto económico con atención especial a tres casos de interés para bancos centrales. Además, revisaremos comparaciones empíricas de los procedimientos ML utilizados para la predicción de variables económicas con el objetivo de determinar las circunstancias en las que dichos procedimientos son ventajosos frente a procedimientos de predicción “tradicionales”.

3.1. Predicción macroeconómica

En el contexto de la predicción macroeconómica, los procedimientos ML descritos anteriormente, se han utilizado fundamentalmente en cuatro áreas de aplicación. En primer lugar, se han utilizado en el contexto de grandes conjuntos de predictores potenciales para reducir su dimensión. La segunda área de aplicación de procedimientos ML es para la predicción de magnitudes macroeconómicas cuando los predictores son no estructurados, como, por ejemplo, textos de noticias o consultas en Google. Los procedimientos ML también se han utilizado para la predicción en caso de relaciones y dependencias temporales no-lineales como las que se suelen observar en épocas de crisis. Finalmente, algunos autores han propuesto utilizar ML para la combinación de predicciones. A continuación, describiremos brevemente algunas aplicaciones empíricas realizadas dentro de cada una de estas áreas.

Reducción de dimensión

En el contexto de la predicción macroeconómica, los procedimientos de regularización se han utilizado frecuentemente para reducir el número de predictores; ver Diebold *et al.* (2020) quienes describen varios trabajos aplicados en los que la regresión con regularización se ha utilizado para la descripción y análisis de variables económicas.

Una de las áreas más populares de aplicación de procedimientos de regularización en el contexto de la predicción macroeconómica es en el contexto de modelos de factores dinámicos. Fan, Ke y Wang (2020) proponen una estrategia consistente basada en extraer factores comunes de los predictores y trabajar con las variables decorreladas. Jokubaitis, Celov y Leipus (en prensa) combinan LASSO y componentes principales para mejorar la predicción de los componentes del producto nacional bruto (PNB) en Estados Unidos (EE. UU.) y la Unión Europea (UE). Concluyen que los procedimientos de regularización mejoran a los procedimientos tradicionales de extracción de factores comunes basados en componentes principales, identificando conjuntos razonables de predictores. Sin embargo, Kim y Swanson (2018) concluyen que los métodos de regularización no mejoran con respecto a los modelos de factores, por lo que sugieren

que, en la práctica es razonable, reducir primero la dimensionalidad mediante la extracción de factores para luego utilizar las regresiones aumentadas con factores estimadas mediante ML; véase, por ejemplo, Bai y Ng (2008,2009), Schumacher (2010) y Umbach (2020) que también proponen utilizar procedimientos de regularización en el contexto de regresiones predictivas aumentadas con factores para la selección de los predictores. Panagiotelis *et al.* (2019) también concluyen, en el contexto de la predicción de magnitudes macroeconómicas en Australia que es difícil mejorar las predicciones de modelos sencillos de extracción de factores, con entre 20 y 40 predictores.

Los procedimientos de regularización también han sido implementados con éxito en el caso de modelos de predicción en los que los parámetros pueden cambiar en el tiempo; ver, por ejemplo, Kapetanios y Zikes (2018). Finalmente, Hillebrand, Lukas y Wei (en prensa) han propuesto una combinación de procedimientos de regularización junto con Bagging para reducir el EPCM en presencia de predictores débiles, definidos como aquellos predictores para los que el incremento de la varianza debido a la estimación de los parámetros correspondientes es mayor que el cuadrado del sesgo que se introduce si no se incorporaran dichos predictores en el modelos predictivo.

Regresores no estructurados

Las predicciones de magnitudes macroeconómicas se basan cada vez más en predictores no-estándar que frecuentemente requieren la utilización de procedimientos de ML; ver, por ejemplo, Clements y Fritsche (en prensa) quienes describen varias aplicaciones en las que se utiliza la información en textos para la predicción. Existe, por ejemplo, una literatura creciente en la que el ciclo económico se relaciona con consultas en Google, noticias y con otras variables no-estándar. Aprigliano, Ardizzi y Monforte (2019) predicen el PNB en Italia en función de los volúmenes y cantidades de diferentes medios de pago como, cheques, transferencias, tarjetas de crédito, etc. Como hemos comentado anteriormente, Babii, Ghysels y Striankas (2020) predicen el crecimiento del PNB en USA en función de datos basados en textos de noticias, mediante sg-LASSO. También Kalamara *et al.* (2020) utilizan datos basados en textos de noticias para predecir el crecimiento del PNB en Reino Unido. Concluyen que las ganancias en predicción son más pronunciadas en tiempos de crisis, cuando las predicciones son generalmente más importantes en la toma de decisiones. Ghirelli *et al.* (2021) describen también varios casos de éxito en la utilización de ML en la predicción económica dentro del Banco de España entre los que se encuentra la previsión a corto plazo de magnitudes macroeconómicas basada en Google trends y noticias de prensa, y la construcción de índices de sentimiento, entre otras.

Dependencias no-linealidades

Sin embargo, no es necesario utilizar datos de variables no-estándar y/o tener grandes bases de datos para recurrir a procedimientos de ML para la predicción. Coulombe *et*

al. (2020) realizan un análisis exhaustivo de las propiedades de los procedimientos de ML para la predicción macroeconómica y concluyen que el factor determinante del éxito de estos procedimientos es la existencia de no-linealidades (que, en el caso de las series macroeconómicas, suelen aparecer en presencia de crisis); ver también Kim y Swanson (2014) que concluyen que los procedimientos ML tienen ventaja predictiva en épocas de crisis⁶. Además también observan que la ventaja de los procedimientos de ML frente a procedimientos tradicionales aumenta con el horizonte de predicción.

Combinación de predicciones

Finalmente, se han realizado varias aplicaciones en las que los procedimientos ML se han utilizado para la combinación de predicciones. Recientemente, Diebold y Shin (2019) proponen utilizar LASSO para estimar los pesos de la combinación de predicciones⁷. Es conocido que, cuando se combinan predicciones, la combinación que parece tener mejores propiedades es la media. Sin embargo, al combinar un número muy grande de predicciones, algunas de ellas pueden ser redundantes o, al menos, no tener suficiente información como para ser incluidas. Diebold y Shin (2019) proponen estimar los coeficientes de la combinación de predicciones utilizando LASSO y centrando la restricción alrededor de $\frac{1}{K}$, donde K es el número de predicciones que se combinan, de forma que algunas de las predicciones de la combinación desaparezcan y los coeficientes del resto se reduzcan hacia la media. El procedimiento es denominado *partially-egalitarian LASSO* (pe-LASSO). Diebold y Shin (2019) analizan el comportamiento de pe-LASSO para predecir el crecimiento en la eurozona basado en la combinación de predicciones del *Survey of Professional Forecasters* del Banco Central Europeo y concluyen que obtienen EPCMs menores que con la media muestral de las predicciones.

Recientemente, Montero-Manso *et al.* (2020) también han propuesto utilizar ML, en concreto, Gradient Boosting para combinar las predicciones de diversos modelos.

3.2. Predicción microeconómica

Los procedimientos ML también han sido utilizados a nivel microeconómico para la predicción de la demanda de productos concretos; ver, por ejemplo, Huber y Stuckenschmit (en prensa) que utilizan Boosting para la predicción de la demanda de una distribuidora de pan y Fildes, Ma y Kolassa (en prensa) para una revisión de la literatura sobre demanda al por menor. Bose *et al.* (2017), dentro de Amazon, proponen una plataforma para la predicción probabilística de la demanda que permite trabajar con catálo-

⁶ Véase también la lista de referencias con aplicaciones de procedimientos de ML para la predicción de series macroeconómicas.

⁷ Bajo una función de pérdida cuadrática, el problema de la combinación óptima de predicciones es un problema de regresión lineal.

gos de millones de productos. Taieb y Hyndman (2014) predicen la demanda de energía eléctrica mediante un procedimiento híbrido en el que se utiliza Boosting en los residuos de un modelo lineal. Finalmente, Camacho, Ramallo y Ruiz Marín (2021) predicen los precios de la vivienda en Madrid basados en árboles de decisión en un contexto de datos de sección cruzada.

Además, Choi y Varian (2012) ilustran cómo utilizar los índices diarios y semanales publicados en tiempo real por Google sobre el volumen de consultas que los usuarios hacen en internet (Google trends) para predecir variables como, por ejemplo, las ventas de vehículos a motor, las reclamaciones del subsidio de desempleo, el nivel de confianza de los consumidores o el número de turistas en Australia.

En un contexto microeconómico, ML también se ha utilizado con éxito para la predicción de quiebras. En un artículo muy interesante, Petropoulos *et al.* (2020) analizan un panel de datos bancarios para determinar los determinantes de quiebra y, simultáneamente, proporcionar una señal adelantada para una quiebra potencial. En este análisis utilizan redes elásticas para seleccionar las variables determinantes de la quiebra y comparan las predicciones de varios procedimientos ML para la predicción, concluyendo que RF y ANN tienen un comportamiento parecido y superior a los demás métodos considerados.

Finalmente, en el contexto de predicción microeconómica, los procedimientos ML también se han implementado para la combinación de predicciones. Por ejemplo, Barrow y Crone (2016a) utilizan ML para la combinación de predicciones y analizan el comportamiento del procedimiento propuesto en el contexto de un subconjunto de series de ventas de la competición M3.

3.3. Predicción financiera

Lo mismo que en el caso de la predicción macroeconómica, las principales aplicaciones de predicción ML en el contexto de variables financieras, pueden agruparse en aquellas que tratan de reducir la dimensión, las que utilizan predictores no estándar y las que tratan de funciones predictivas no lineales.

Regularización

Swanson, Xiang y Yang (en prensa) utilizan componentes principales con *targeted predictors* (basado en LASSO y elastic net) para predecir la estructura de los tipos de interés utilizando índices económicos. Concluyen que ML puede ser importante en la predicción cuando esta se realiza de manera individual. Sin embargo, cuando consideran la combinación de diferentes modelos predictivos, los modelos que sólo utilizan el pasado de los tipos de interés tienen EPCMs menores que aquellos que incorporan información de variables macroeconómicas.

Predictores no estándar

También existen aplicaciones muy interesantes de datos no estructurados para la predicción financiera, intentando relacionar el comportamiento de los mercados con noticias o búsquedas en Google. Una de las primeras contribuciones en esta dirección es Tetlock, Saar-Tsechansky y Macskassy (2008) que analizan la relación entre los medios de difusión y el mercado financiero. Recientemente, ha aparecido un concepto nuevo conocido como *Internet concern*, que trata de la utilización de datos de búsquedas para la cuantificación de la especulación en un mercado determinado. Wang *et al.* (2018) utilizan *Internet concern* para predecir la volatilidad en el mercado del crudo. Perlin *et al.* (2017) también han analizado la relación de las búsquedas en Google con mercados financieros. Chen *et al.* (en prensa) utilizan ANN y RF para predecir los tipos de cambio del bitc oin en funci on de varios factores econ omicos y tecnol ogicos entre los que est a, por ejemplo, el n umero de búsquedas en Google y en la Wikipedia. El reciente art culo de Gentzkow, Kelly y Taddy (2019) realiza una excelente revisi on sobre procedimientos ML aplicados a la utilizaci on de textos en la predicci on econ mica y financiera.

En el caso de datos de alta frecuencia (intradarios), Yan *et al.* (2020) utilizan SVM para predecir volatilidad.

Finalmente, Arratia (2021) propone modelizar rendimientos financieros en funci on de sentimientos. Limongi y Ravazzolo (2019) tambi en investigan c omo los sentimientos de los inversores afectan a los rendimientos financieros y eval an el poder predictivo de los  ndices de sentimiento sobre rendimientos de mercados financieros en EE.UU. y UE.

Dependencias no lineales

En el  rea de finanzas, Israel, Kelly y Moskowitz (en prensa) describen las posibilidades de ML en relaci on a distintos problemas. En concreto, estos autores exponen los retos a los que se enfrentan los procedimientos ML en la predicci on de rendimientos financieros, caracterizados por: (1) no tener (relativamente) grandes cantidades de datos, al estar estos limitados por el hecho de que no se puede experimentar para generar nuevas observaciones, y_i , de las que se pueda aprender; y (2) por tener la se al relativamente peque a en relaci on al ruido, consecuencia de la eficiencia de los mercados financieros. Sin embargo, los procedimientos ML se han utilizado con  xito en algunos problemas de predicci on de rendimientos financieros como, por ejemplo, cuando existen no-linealidades en las relaciones din micas entre los rendimientos financieros y variables de estado o din micas complejas en la exposici on a los factores de riesgo; ver, por ejemplo, Gu, Kelly y Xiu (en prensa a, en prensa b) que muestran ganancias significativas, tanto desde el punto de vista financiero como estad stico, cuando se utilizan autoencoders (redes neuronales correspondientes a componentes principales),  rboles o redes neuronales, respectivamente, para la predicci on de precios de activos financieros individuales. Sin embargo, recientemente, De Nard, Hediger y Leippold (2020)

proponen utilizar procedimientos más sencillos basados en *subsampling*, con mejores resultados.

3.4. Comparación entre procedimientos

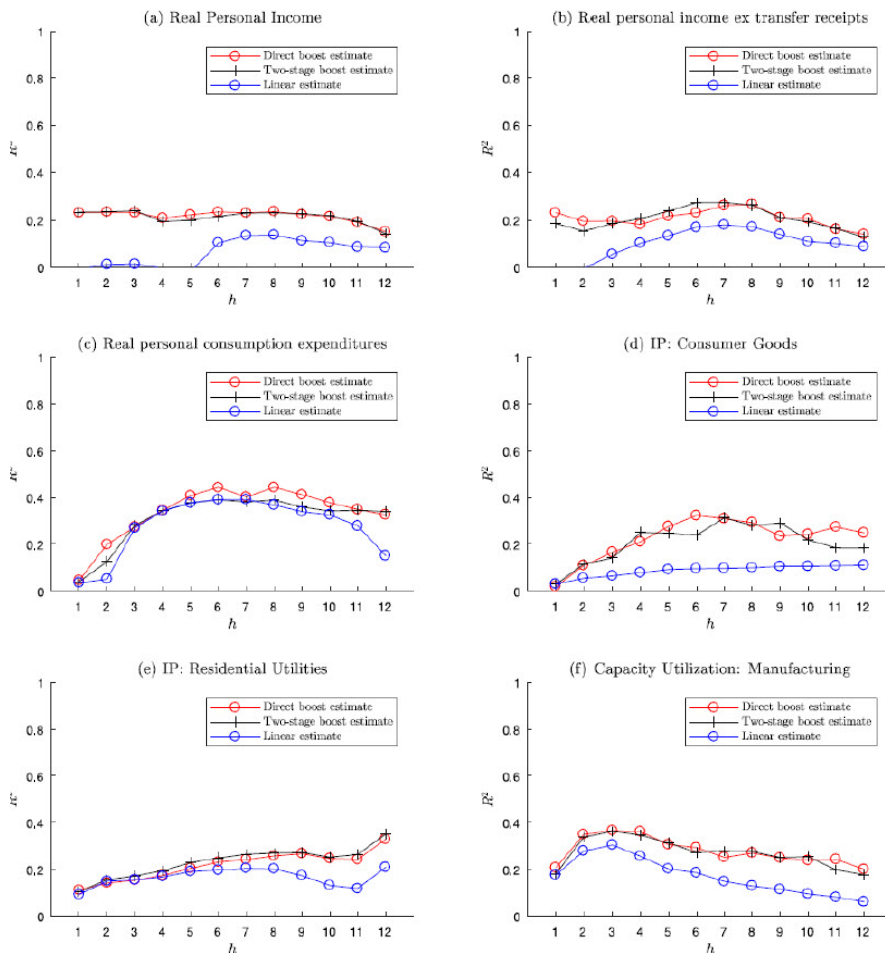
Los procedimientos ML han sido comparados entre sí y con otros procedimientos tradicionales en el contexto de la predicción de variables económicas. Las comparaciones se han realizado tanto en el contexto de competiciones de procedimientos predictivos en las que se comparan los procedimientos en la predicción de un gran número de variables como en el contexto de la predicción de una única o de un número reducido de variables.

Entre las comparaciones de predicciones en el contexto de un gran número de variables, una de las primeras comparaciones sistemáticas entre un amplio abanico de procedimientos ML fue realizada por Ahmed *et al.* (2010) en el contexto de las variables mensuales relacionadas con la economía y negocios de la competición M3; ver Makridakis y Hibon (2000) para una descripción de dicha competición. En dicho contexto y realizando predicciones un periodo-hacia-adelante univariantes, Ahmed *et al.* (2010) concluyen que los mejores procedimientos, en términos de minimizar el EPCM, se obtienen utilizando las redes neuronales y la regresión de procesos Gaussianos. Más recientemente, Smyl (2020) es el procedimiento ganador de la última competición M4. El procedimiento propuesto por Smyl (2020) se basa en una combinación entre suavizamiento exponencial y redes neuronales recurrentes. Kim y Swanson (2014) también comparan distintos procedimientos ML para la predicción de 144 variables macroeconómicas de EE.UU. en el contexto de los modelos de difusión basados en factores. Nuevamente, concluyen que la ventaja de los procedimientos ML se encuentra en presencia de no-linealidades con procedimientos basados en la combinación de factores y regularización obteniendo menores EPCMs. Kauppi y Virtanen (en prensa) también realizan un análisis exhaustivo comparando predictores lineales y Boosting para las superanalizadas 128 series macroeconómicas de McCracken y Ng (2016). La ventaja comparativa de Boosting frente a procedimientos lineales depende del tipo de no-linealidad que tengan los datos y, por lo tanto, puede cambiar dependiendo de la variable concreta que se vaya a predecir. Por ejemplo, la figura 4 que se ha tomado de Kauppi y Virtanen (en prensa) muestra los coeficientes de determinación en la predicción de varias magnitudes macroeconómicas para $h = 1, \dots, 12$ cuando se utiliza Boosting y modelos lineales en la predicción. Las magnitudes que se predicen son ingresos y consumo individuales, precios de bienes de consumo y de residencias y la capacidad de utilización. Esta figura ilustra que las conclusiones sobre la comparación entre procedimientos tradicionales y ML pueden depender tanto de las propiedades de la variables que se predice como del horizonte de predicción. Por ejemplo, en el caso de la predicción del consumo individual, las diferencias solo se observan en el caso de predicciones 12 meses hacia adelante. Sin embargo, en el caso de las variables de ingresos, las diferencias se observan en las predicciones a corto plazo y se hacen menores cuando el horizonte de predicción se incrementa.

Como hemos comentado anteriormente, el comportamiento de los procedimientos de predicción ML también se ha comparado con procedimientos “tradicionales” en el contexto de la predicción de una única o un número reducido de variables. Muy recientemente, Joseph *et al.* (2021) comparan varios procedimientos para la predicción de la inflación en Reino Unido basada en la desagregación de los precios a nivel granular (581 precios desagregados). Los modelos que utilizan en su comparación se basan en: (1) Modelos de factores dinámicos para la reducción de la dimensionalidad; (2) modelos de regresión con regularización (*Ridge* y *LASSO*); y (3) ML (*Elastic Net*, *SVM*, *ANN* y

FIGURA 4

PREDICIONES DE VARIABLES MACROECONÓMICAS TOMADAS DE KAUPPI Y VIRTANEN (EN PRENSA)



RF). Las predicciones obtenidas con estos procedimientos son comparadas en términos de EPCMs con respecto a un modelo autorregresivo. Concluyen que los procedimientos no-lineales de ML y, en particular, SVM, tienen un gran potencial en la predicción de los datos granulares y muy volátiles que se obtienen a nivel de precios microeconómicos, abriendo la puerta a que puedan ser utilizados para modelizar precios muy desagregados a nivel de productos o incluso de precios obtenidos de las páginas web. Los mejores resultados de los procedimientos de ML se observan al aumentar el horizonte de predicción y, en concreto, en la predicción a medio plazo (un año hacia adelante). La conclusión sobre que la utilización de procedimientos ML es más útil cuando el horizonte de predicción se incrementa también se ha obtenido por muchos otros autores como, por ejemplo, Kim y Swanson (2014) en el contexto de su comparación y D'Amuri y Marcucci (2017) en la predicción de desempleo en EE.UU. Kotchoni, Leroux y Stevanovic (2019) también concluyen que la utilización de procedimientos de regularización para predecir varias variables económicas tiene ventajas al aumentar el horizonte de predicción. Colombo y Pelagatti (en prensa) llegan a la misma conclusión al predecir tipos de cambio. La mejora en la capacidad predictiva de los procedimientos ML aparece con $h > 3$.

Fornaro y Luomaranta (2020) utilizan procedimientos Boosting para proporcionar predicciones de la economía finlandesa. Comparan dichas predicciones con otras obtenidas con procedimientos "tradicionales" y concluyen que las predicciones basadas en BD se adelantan en el tiempo a las tradicionales lo que, evidentemente es una gran ventaja cuando dichas predicciones van a ser utilizadas en la toma de decisiones.

Finalmente, Medeiros *et al.* (2021) también predicen la inflación en USA y concluyen que los procedimientos ML con un gran número de predictores producen predicciones más precisas que otros modelos tradicionales de series temporales. Entre los procedimientos ML, ellos concluyen que el que merece más atención es RF, cuyo buen comportamiento es debido no únicamente a cómo se seleccionan los predictores si no también a la incorporación de relaciones dinámicas no lineales entre variables macroeconómicas relevantes y la inflación. Mediante RF se pueden reducir hasta un 30% el EPCM cuando se compara con predicciones basadas en procedimientos tradicionales de series temporales.

4. CONCLUSIONES

A lo largo de este capítulo, hemos ilustrado cómo los procedimientos ML pueden ser muy útiles en distintos contextos de la predicción relacionada con variables económicas. Los procedimientos basados en regularización han mostrado su ventaja en la predicción basada en factores comunes como una forma automática de seleccionar los predictores. Los procedimientos basados en árboles de decisión y redes neuronales se han mostrado especialmente ventajosos en el caso de la predicción de variables con dependencias no-lineales desconocidas. Además, los procedimientos de predicción ML tienen gran potencial cuando los datos son granulares y volátiles, así como en aplica-

ciones con predictores no estructurados. Finalmente, varios autores concluyen que las predicciones basadas en ML pueden adelantarse en el tiempo a las basadas en procedimientos tradicionales y que tienen ventajas comparativas a medida que el horizonte de predicción se incrementa.

Los procedimientos de predicción ML han sido originalmente diseñados para observaciones de sección cruzada independientes. Sin embargo, dichos procedimientos están aún en su infancia en el caso de datos temporales, con muchos temas importantes sin explorar o solo explorados parcialmente; ver la discusión en Barker (2020). Por ejemplo, algunos de los procedimientos ML diseñados para datos independientes, se han extendido a datos temporales estacionarios. Sin embargo, la predicción mediante ML en el caso de variables no estacionarias todavía está sin tratar en profundidad con muy pocas menciones en la literatura. Ahmed *et al.* (2010) han considerado el efecto de la diferenciación en el contexto de las variables económicas observadas mensualmente de la competición M3 cuando se utiliza ML para la predicción univariante y concluyen que los resultados son peores (mayores EPCM) que cuando se predicen las variables en niveles. Bontempi, Ben Taieb y Le Borgue (2013) han considerado la utilización de ML en el contexto de variables no estacionarias y concluyen que los procedimientos de aprendizaje local pueden aplicarse a variables no estacionarias. Finalmente, en relación con la presencia de no-estacionariedad, Colombo y Pelagatti (en prensa) han considerado la utilización de VSM en el contexto de modelos con mecanismo de corrección del error mientras que Escribano y Wang (en prensa) predicen inflación del precio de la gasolina en España mediante una combinación de procedimientos de series temporales clásicos basados en cointegración no-lineal y árboles aleatorios, mejorando a cada uno de estos procedimientos por separado en términos de EPCM.

En este capítulo, por cuestiones de espacio, nos hemos centrado en predicción un-periodo-hacia-adelante de una única variable. La predicción multivariante está todavía desarrollándose. Recientemente, De Stefani *et al.* (2019) han considerado predicción multivariante proponiendo una versión ML de los modelos de factores dinámicos (DFMs). Aunque la mayoría de la literatura se ha centrado en predicciones un-periodo-hacia-adelante, existen también varias aportaciones donde se han considerado predicciones varios-periodos-hacia-adelante; ver, por ejemplo, Bontempi, Taieb y Le Borgue (2013), Taieb y Hyndman (2014), De Stefani *et al.* (2019) y Kauppi y Virtanen (en prensa).

Las predicciones son realmente útiles para la toma de decisiones cuando van acompañadas de medidas de incertidumbre, es decir, cuando aparecen en términos probabilísticos. Evidentemente, cualquier predicción lleva asociada una incertidumbre por lo que no es sólo importante que dicha incertidumbre sea lo más pequeña posible sino también que, sea cuál sea su nivel, la podamos medir adecuadamente. La toma de decisiones será diferente dependiendo de la incertidumbre asociada con las predicciones que se utilicen. La predicción de las probabilidades asociadas con fenómenos futuros y la consideración de diversas formas de incertidumbre está siendo el foco de atención

de muchos trabajos relacionados con los procedimientos de ML descritos en la sección anterior; ver las referencias en la discusión de Januschowski *et al.* (2020) y Salinas *et al.* (2020). Sin embargo, este es un área en la que todavía es necesario realizar una investigación intensa. Hyndman y Athanasopoulos (2018) proponen utilizar bootstrap basado en residuos para construir intervalos de predicción en el caso de redes neuronales autorregresivas.

Finalmente, queremos mencionar que, a pesar de que los procedimientos ML han demostrado su utilidad empírica, en muchos casos, se carece de una explicación analítica del porqué de estos buenos resultados. Solamente existen algunas explicaciones parciales que ayuden a entender mejor las propiedades de estos procedimientos. Por ejemplo, recientemente, Scornet, Biau y Vert (2015) han probado la consistencia del procedimiento Bagging mientras que Athey, Tibshirani y Wager (2019) proporcionan resultados asintóticos para RF; ver también Bühlmann y Yu (2002) para algunos resultados asintóticos del Bagging y Subbagging.

Referencias

- AHMED, N. K., ATIYA, A. F., EL GAYAR, N. y EL SHISHINY, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), pp. 594-621.
- APRIGLIANO, V., ARDIZZI, G. y MONFORTE, L. (2019). Using payment system data to forecast economic activity. *International Journal of Central Banking*, 15(4), pp. 55-80.
- ARAS, S. y KOCAKOC, I. D. (2016). A new model selection strategy in time series forecasting with artificial neural networks: IHTS. *Neurocomputing*, 174, pp. 974-987.
- AROMI, J. D. (2020). Linking words in economic discourse: Implications for macroeconomic forecasts. *International Journal of Forecasting*, 36(4), pp. 1517-1530.
- ARRATIA, A. (2021). Predicciones financieras basadas en análisis de sentimiento de textos y minería de datos. En: D. PEÑA, P. PONCELA y E. RUIZ (eds.), *Nuevos Métodos de Predicción Económica con Datos Masivos*. Madrid: Funcas.
- ATHEY, S., TIBSHIRANI, J. y WAGER, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), pp. 1148-1178.
- AWAD, M. y KHANNA, R. (2015). *Efficient Learning Machines*. Berkeley: Springer.
- BABII, A., GHYSELS, E. y STRIANKAS, J. (2020). Machine learning time series regressions with an application to nowcasting, arXiv:2005.14057v2[econ.EM].
- BAI, J. y NG, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146, pp. 304-317.

- (2009). Boosting diffusion indexes. *Journal of Applied Econometrics*, 24, pp. 607-629.
- BARKER, J. (2019). Machine learning in M4: What makes a good unstructured model? *International Journal of Forecasting*, 36(1).
- BARROW, D. K. y CRONE, S. F. (2016a). A comparison of AdaBoost algorithms for time series forecast combination. *International Journal of Forecasting*, 32, pp. 1103-1119.
- (2016b). Cross-validation aggregation for combining autoregressive neural network forecasts. *International Journal of Forecasting*, 32(4), pp. 1120-1137.
- BIAU, G. y SCORNET, E. (2016). A random forest guided tour, with discussion. *TEST*, 25, pp. 197-227.
- BONTEMPI, G., BEN TAIEB, S. y LE BORGUE, Y.-A. (2013). Machine learning strategies for time series forecasting. En: M.-A. AUFURE y E. Zimanyi (eds.), *Business Intelligence*. Berlin: Springer-Verlag.
- BOSE, J.-H., FLUNKERT, V., GASTHAUS, J., JANUSCHOWSKI, T., LARGE, D., SALINAS, D., SCHELTER, S., EGER, M. y Wang, Y. (2017). Probabilistic demand forecasting at scale. *Proceedings of the VLDB Endowment*, 10(12).
- BREIMAN, L. (2001). Random forest. *Machine Learning*, 45(1), pp. 5-32.
- BÜLMANN, P. y YU, B. (2002). Analyzing Bagging. *Annals of Statistics*, 30(4), pp. 927-961.
- CAMACHO, M., RAMALLO, S. y RUIZ MARÍN, M. (2021). Árboles de decisión en economía: una aplicación a la determinación del precio de la vivienda. En: D. P. PEÑA, P. PONCELA y E. RUIZ (eds.). *Nuevos Métodos de Predicción Económica con Datos Masivos*. Madrid: Funcas.
- Caro, A. y PEÑA, D. (2021). Predicción de series temporales económicas con modelos factoriales dinámicos y de Machine Learning. En: D. PEÑA, P. PONCELA y E. RUIZ (eds.). *Nuevos Métodos de Predicción Económica con Datos Masivos*. Madrid: Funcas.
- CHAKRABORTY, C. y JOSEPH, A. (2017). Machine learning at central banks. *Staff Working Paper*, no. 674.
- CHEN, W., XU, H., JIU, L. y GAO, Y. (2021). Machine learning model for Bitcoin exchange rate predictions using economic and technology determinants. *International Journal of Forecasting*.
- CHOI, H. y Varian, H. (2012). Predicting the present with Google trends. *Economic Records*, 88, pp. 2-9.
- CHOROMANSKA, A., HENAFF, M., MATHIEU, M., AROUS, G. B. y LeCUN, L. (2015). The loss surfaces of multilayer networks. *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- CLEMENTS, M. P. y FRITSCHKE, U. (2020). Text-based data and forecasting: Editors introduction. *International Journal of Forecasting*, 36(4), pp. 1476-1477..

COLOMBO, E. y PELAGATTI, M. (2020). Statistical learning and exchange rate forecasting. *International Journal of Forecasting*, 36(4), pp.1260-1289..

COULOMBE, P. G., LEROUX, STEVANOVIC, M. D. y SURPRENANT, S. (2021). How is machine learning useful for macroeconomic forecasting? *International Journal of Forecasting*.

D'AMURI, F. y MARCUCCI, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33, pp. 801-816.

DE NARD, G., HEDIGER, S. y LEIPPOLD, M. (2020). *Subsample factor models for asset pricing: The rise of VASA*. Manuscrito.

DE STEFANI, J., LE BORGNE, Y.-A., CAELEN, O., HATTAB, D. y BONTEMPI, G. (2019). Batch and incremental dynamic factor machine learning for multivariate multi-step-ahead forecasting. *International Journal of Data Science and Analytics*, 7(4), pp. 311-329.

DIEBOLD, F. X. y SHIN, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, 35, pp. 1679-1691.

DIEBOLD, F. X., GHYSELS, E., MYKLAND, P. y ZANG, L. (2019). Big data in dynamic predictive econometric models (editorial). *Journal of Econometrics*, 212(1), pp. 1-3.

EKLUND, J. y KAPETANIOS, G. (2008). A review of forecasting techniques for large data sets. *National Institute Economic Review*, 203(1), pp. 109-115.

ESCRIBANO, A. y WANG, D. (en prensa). Forecasting gasoline prices with mixed random forest error correction models. *International Journal of Forecasting*.

FAN, J., KE, Y. y WANG, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, 216, pp. 71-85.

FILDES, R., MA, S. y KOLASSA, S. (en prensa). Retail forecasting: Research and practice. *International Journal of Forecasting*.

FORNARO, P. y Luomaranta, H. (2020). Nowcasting Finish real economic activity: a machine learning approach. *Empirical Economics*, 58(1), pp. 55-71.

GARBODEN, P. M. E. (2019). Sources and types of big data for macroeconomic forecasting. En: P. FULEKY (ed.). *Macroeconomic Forecasting in the Era of Big Data*. Springer.

GENTZKOW, M., KELLY, B. T. y TADDY, M. (2019). Text as data. *Journal of Economic Surveys*, 57(3), pp. 535-574.

GHIRELLI, C., HURTADO, S., PÉREZ, J. J. y URTASUN, A. (2021). Desarrollos con Big data para el análisis coyuntural en los bancos centrales. En: D. PEÑA, P. PONCELA y E. RUIZ (eds.). *Nuevos Métodos de Predicción Económica con Datos Masivos*, Madrid: Funcas.

- GU, S., KELLY, B. y XIU, D. (en prensa a). Autoencoder asset pricing models. *Journal of Econometrics*.
- (en prensa b). Empirical asset pricing via machine learning. *Review of Financial Studies*.
- HILLEBRAND, E., LIKAS, M. y WEI, W. (en prensa). Bagging weak predictors. *International Journal of Forecasting*.
- HU, M. J. C. (1964), Application of the adaline system to weather forecasting. *Master Thesis, Technical report, 6775-1*. Stanford, CA: Stanford Electronic Laboratories,.
- HUBER, J. y STUCKENSCHMIT, H. (en prensa), Daily retail demand forecasting using machine learning with emphasis on calendar special days. *International Journal of Forecasting*.
- HYNDMAN, R. y ATHANASOPOULOS, G. (2018). *Forecasting: Principles and Practice*, 2ª edición, OTexts, Melbourne (Australia). OTexts.com/fpp2. Descargado en 7/29/2020.
- ISRAEL, R., KELLY, B. y MOSKOWITZ, T. (en prensa). Can machines “learn” finance? *Journal of Investment Management*.
- JANUSCHOWSKI, T., GASTHAUS, J., WANG, Y., SALINAS, D., FLUNKERT, V., BOHLKE-SCHIEDER, M. y LALLOT, C. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36, pp. 167-177.
- JOSEPH, A., KALAMARA, E., POTJAGAILO, G. y KAPETANIOS, G. (2021). Forecasting UK inflation bottom up. *International Journal of Forecasting*.
- JOKUBAITIS, S., CELOV, D. y LEIPUS, R. (en prensa). Sparse structures with LASSO through principal components: Forecasting GDP components in the short run. *International Journal of Forecasting*.
- KALAMARA, E., TURRELL, A., REDL, C., KAPETANIOS, G. y KAPADIA, S. (2020). Making text count: economic forecasting using newspaper text. *Staff Working Paper*, no. 865. Bank of England.
- KANUPPI, H. y VIRTANEN, T. (2020). Boosting nonlinear predictability of macroeconomic time series. *International Journal of Forecasting*, 37(1), pp. 151-170.
- KAPETANIOS, G. y ZIKES, F. (2018). Time varying LASSO. *Economics Letters*, 169, pp. 1-6.
- KIM, H. H. y SWANSON, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178, pp. 352-367.
- (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage. *International Journal of Forecasting*, 34(2), pp. 339-354.
- KOTCHONI, R., LEROUX, M. y STEVANOVIC, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, 34(7), pp. 1050-1072.
- LIMONGI CONCETO, C. y RAVAZZOLO, F. (2019). Optimism in financial markets: Stock market returns and investor sentiments. *Journal of Risk and Financial Management*, 12(141).

- LIPTON, Z. C., BERKOWITZ, J. y ELKAN, C. (2015). A critical review of recurrent neural networks for sequence learning, arXiv:1506.00019
- MAKRIDAKIS, S. y HIBON, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16, pp. 451-476.
- MCCRACKEN, M. W. y NG, S. (2016), FRED-MD: A monthly data base for macroeconomic research. *Journal of Business & Economic Statistics*, 34, pp. 574-589.
- MEDEIROS, M. C., VASCONCELOS, G. F. R., VEIGA, A. y ZILBERMAN, E. (2021). Forecasting inflation in a data-rich environment: The benefits of ML methods. *Journal of Business and Economic Statistics*, 39(1).
- MONTERO-MANSO, P., ATHANASOPOULOS, G., HYNDMANN, R. J. y TALAGALA, T. S. (2020). FFORMA: feature-based forecast model averaging. *International Journal of Forecasting*, 36(1), pp. 86-92.
- PALIT, A. K. y POPOVIC, D. (2005). Computational Intelligence in Time Series Forecasting. *Theory and Engineering Applications*, Springer-Verlag, Secaucus.
- PAN, L. y POLITIS, D. N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions (with discussion). *Journal of Statistical Planning and Inference*, 177, pp. 1-27.
- PANAGIOTELIS, A., ATHANASOPOULOS, G., HYNDMAN, R. J., JUANG, B. y VAHID, F. (2019). Macroeconomic forecasting for Australia using a large number of predictors. *International Journal of Forecasting*, 35(2), pp. 616-633.
- PERLIN, M. S., CALDEIRA, A., SANTOS, A. P. y PONTUSCHKA, M. (2017), Can we predict the financial markets based on Google's search queries? *Journal of Forecasting*, 36, pp. 454-467.
- PETROPOULOS, A., SIAKOULIS, V., STAVROULAKIS, E. y VLACHOGIANNAKIS, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3), pp. 1092-1113.
- SALINAS, D., FLUNKERT, V., GASTHAUS, J. y JANUSCHOWSKI, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), pp. 1181-1191.
- SCHUMACHER, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters*, 107(2), pp. 95-98.
- SCORNET, E., BIAU, G. y VERT, J.-P. (2015). Consistency of random forests. *Annals of Statistics*, 43, pp. 1716-1741.
- SCOTT, S. L. y VARIAN, H. (2014), Predicting the present with Bayesian structural time series. *Journal of Mathematical Modelling and Numerical Optimization*, 5(1-2).
- SIMON, N., FRIEDMAN, J., HASTIE, T. y TIBSHIRANI, R. (2013), A sparse-group LASSO. *Journal of Computational and Graphical Statistics*, 22(2), pp. 231-245.
- SMOLA, A. J. y SCHOLKOPF, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 4, pp. 199-222.

- SMYL, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1), pp. 75-85.
- SWANSON, N. R., XIANG, W. y YANG, X. (en prensa). Predicting interest rates using shrinkage methods, real-time diffusion indexes, and model combinations. *Journal of Applied Econometrics*, 35(5), pp. 587-613.
- TAIEB, S. B. y HYNDMAN, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, 30(2), pp. 382-394.
- TETLOCK, P. C., SAAR-TSECHANSKY, M. y MACSKASSY, S. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3), pp. 1437-1467.
- TSAY, R. y CHEN, R. (2019). *Nonlinear Time Series Analysis*. Wiley.
- UMBACH, S. L. (2020). Forecasting with supervised factor models. *Empirical Economics*, 58(1), pp. 169-190.
- VAPNIK, V. N. (2006). *Estimation of Dependences Based on Empirical Data*. New York: Springer.
- VARIAN, H. R. (2014). Big Data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), pp. 3-18.
- WANG, J., ATHANASOPOULOS, G., HYNDMAN, R. J. y WANG, S. (2018). Crude oil price forecasting based on internet concern using an extreme learning machine. *International Journal of Forecasting*, 34(4), pp. 665-677.
- YAN, R., YU, L., YU, H., XU, G., WU, Y. y LIU, Z. (2020). Big data analytics for financial market volatility forecast based on support vector machine. *International Journal of Information Management*, 50, pp. 452-462.
- ZHANG, G., PATUWO, B. E. y HU, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), pp. 35-62.