

## Presentación

Los nuevos datos masivos han abierto la posibilidad de incrementar nuestro conocimiento en todos los campos y, en particular, mejorar las decisiones económicas y empresariales. Para ello, un componente imprescindible es ampliar nuestra capacidad de predicción. Con este objetivo, este libro analiza varias contribuciones que se presentaron en las jornadas del mismo nombre celebradas en Funcas el 7 de noviembre de 2023. Las jornadas fueron organizadas por los tres editores de esta monografía y tuvieron una amplia participación presencial y *online*. El lector interesado puede visualizar las presentaciones orales de los trabajos incluidos en esta monografía en el canal de Funcas en YouTube: <https://www.youtube.com/watch?v=5EW1vVhdga0youtube>

Los rápidos cambios económicos y sociales de las últimas décadas han generado una creciente incertidumbre respecto al futuro. Tenemos ejemplos muy recientes, como la epidemia del COVID-19, que hundi6 las economías de todo el mundo, o las guerras en Ucrania y Gaza, que están teniendo importantes repercusiones económicas. Estos cambios han generado la necesidad de incluir más información en los modelos de predicción, así como de desarrollar procedimientos resistentes a posibles no linealidades, como cambios estructurales. Las predicciones racionales se basan siempre en extrapolar de manera inteligente el pasado hacia el futuro, pero la evolución de la economía depende mucho de factores sociales y políticos que resultan muy difíciles de prever y, además, pueden tener un gran impacto en los resultados económicos. Resulta importante, por tanto, incorporar estas nuevas variables a los modelos de predicción.

El primer bloque de trabajos de esta monografía incluye los capítulos I, II y III, que presentan contribuciones donde se analizan nuevos tipos de datos que pueden mejorar la predicción. Entre ellas están las obtenidas con el análisis del lenguaje natural o de sentimientos reflejando cuestiones como la percepción subjetiva de la conjuntura económica o la importancia del cambio climático y que pueden tener impacto en las decisiones económicas. También se analiza el importante problema de cómo el análisis con variables desagregadas complementa el análisis global sobre el conjunto de variables originales.

El capítulo I está escrito por **Álvaro Ortiz** y **Tomasa Rodrigo** y estudia la incorporación de información obtenida con el lenguaje escrito para mejorar las decisiones empresariales. Los autores muestran cómo gracias al desarrollo de los modelos de lenguaje natural, los

textos digitales se están convirtiendo en una fuente muy importante de información para guiar estrategias económicas. Entre los modelos de generación de lenguaje, ChatGPT, que se ha convertido en un referente en los métodos de inteligencia artificial aplicados al análisis automático de datos no estructurados, es un ejemplo destacado.

En el capítulo II, **María Alló** y **María Loureiro** analizan la evolución del sentimiento social hacia el cambio climático en España y su posible efecto sobre las decisiones económicas. Aunque es generalmente aceptado que el cambio climático representa un desafío crucial para la sociedad actual, la percepción de su importancia ha sido poco estudiada. En este trabajo se utiliza información obtenida de redes sociales para medir la percepción sobre su evolución temporal, incluyendo técnicas de procesamiento de lenguaje natural. Los resultados analizan también el efecto de cambios estructurales, como la pandemia debida al COVID-19.

Otra de las formas de incorporar información adicional es utilizar datos más desagregados de los habituales en los modelos económicos. **Antonio Montañés** estudia, en el capítulo III, las diferencias en la evolución regional de los precios en España utilizando datos desagregados tanto en el espacio, por provincias, como en el tiempo, por periodos. Se realiza un estudio para contrastar si existe un patrón de comportamiento único, o varios diferentes en el espacio y en el tiempo, detectando la presencia de grupos o clústeres de series temporales. Tener en cuenta estas variables es importante para obtener predicciones más detalladas y precisas.

El segundo bloque se centra en el análisis de datos masivos. Por un lado, en el capítulo IV, **Josep Mestres Domènech** presenta el portal *Economía en Tiempo Real*, de CaixaBank Research. A partir de información de CaixaBank, como los datos de tarjetas de gasto y reintegro de tarjetas de crédito, ingresos (nóminas y subsidios de paro) o los pagos efectuados en España en sus TPV con las tarjetas emitidas por entidades extranjeras, se describen nuevos indicadores de consumo, acceso a la vivienda, salarios, turismo o desigualdad.

En los capítulos V y VI, el enfoque se centra en la utilización de técnicas capaces de aprovechar las oportunidades de los datos masivos. Por un lado, en el capítulo V, **Isabel Casas**, **Juan Miguel Marín** y **Helena Veiga** exploran la eficacia de distintos modelos en la predicción de la varianza diaria realizada a partir de datos intradiarios de Bitcoin, NASDAQ y S&P500. Este capítulo muestra cómo la disponibilidad de datos, a frecuencias muy altas, constituye una gran ventaja, posibilitando la estimación de la volatilidad de forma consistente.

Por otro lado, en el capítulo VI, **María Ángeles Carnero**, **Ángel León** y **Trino Níguez** abordan la capacidad del acceso y análisis de datos en tiempo real para predecir la rentabilidad de una cartera de inversión y cómo usar dichas predicciones para seleccionar los activos y su ponderación dentro de la cartera de inversión. El análisis se realiza sobre el índice Russell 1000, que refleja de forma continua las 1.000 empresas más grandes en el mercado de valores estadounidense.

El último bloque está comprendido por los capítulos VII y VIII y está dedicado al mercado eléctrico, de gran relevancia económica y social y cuyo análisis se está sofisticando debido a la actual abundancia de datos. En ambos capítulos se analizan, desde distintos puntos de vista, las curvas de oferta de los productores eléctricos.

En el capítulo VII, **Jorge Arias Martí** y **Andrés M. Alonso Fernández** estudian las ofertas realizadas por los productores de electricidad del mercado español. El objetivo es agrupar los conjuntos de oferta mediante técnicas de clasificación no supervisada utilizando la distancia de Hausdorff para realizar los grupos. Una vez obtenidos los distintos grupos, se relacionan con variables de producción energética, además de las habituales variables temporales, como hora, día de la semana y mes.

En el último capítulo de la monografía, **Antonio Muñoz**, **José Portela**, **Eugenio Sánchez-Úbeda** y **Guillermo Mestre** introducen modelos de series temporales funcionales inspirados en la metodología Box-Jenkins para modelar y predecir las curvas de oferta del mercado eléctrico. Esta propuesta permite introducir un enfoque probabilístico en las estrategias de oferta de los agentes del mercado. La caracterización obtenida con estas técnicas de las curvas de oferta permite captar su dinámica, que incluye varias estacionalidades en el mismo modelo. La propuesta se aplica al mercado eléctrico diario italiano.

Queremos agradecer a todos los autores su contribución a este trabajo y al director general de Funcas, Carlos Ocaña, todo su apoyo a las actividades de *big data* y, en particular, a la realización de las jornadas que han dado lugar a esta monografía. Por último, agradecemos a Myriam González su eficiencia en su publicación.

**Daniel Peña, Pilar Poncela y Eva Senra**

Marzo 2024