

## CAPÍTULO I

## Digitalización financiera pospandemia: una aproximación basada en aprendizaje automático

Santiago Carbó-Valverde  
Pedro J. Cuadros-Solas  
Francisco Rodríguez-Fernández

Este trabajo examina cuáles son los factores que predicen el grado de digitalización financiera en un contexto pospandémico. Usando una encuesta realizada en España a 2.121 usuarios *online* –entre 18 y 70 años– se emplea el *random forest* y *extreme gradient boosting* para predecir el proceso de digitalización financiera. El estudio muestra que las técnicas de aprendizaje automático predicen con un elevado grado de precisión el grado de digitalización financiera. Los principales predictores son la necesidad/utilidad percibida sobre la banca digital para consultar saldos, transferir dinero y relacionarse con el banco.

*Palabras clave:* digitalización financiera, aprendizaje automático, banca digital, pospandemia.

## 1. INTRODUCCIÓN

El uso de técnicas de aprendizaje automático (*machine learning*) en el campo de la banca y de las finanzas puede ayudar a examinar y simplificar la respuesta ante cuestiones que, hasta este momento, eran complejas. El potencial de los algoritmos para revelar patrones en un amplio conjunto de datos hace que el aprendizaje automático comience a ser utilizado dentro del análisis financiero. En concreto, la utilización de algoritmos de *machine learning* puede ser de interés en el campo de la digitalización financiera. Dicho proceso de transformación digital de los consumidores es especialmente complejo, ya que intervienen una combinación amplia de factores –sociales, conductuales, aptitudinales, perceptuales y relaciones– (Carbó-Valverde, Cuadros-Solas y Rodríguez-Fernández, 2020; Lin, Wang y Hung, 2020). En relación a las técnicas econométricas más usadas en la literatura para estudiar fenómenos como el grado de adopción de la banca digital o de determinados medios de pago digitales, el aprendizaje automático ofrece la posibilidad de revelar patrones de comportamiento de los consumidores que, hasta el momento, no eran obvios (Cui, Wong y Lui, 2006; Bajari *et al.*, 2015).

La aplicación de estas técnicas sobre el proceso de digitalización financiera ha ganado relevancia desde la pandemia de COVID-19. Como se ha venido observando desde la irrupción de la pandemia, todo el proceso de transformación digital de las sociedades se ha acelerado significativamente (Fu y Mishra, 2022). La pandemia ha provocado una transición acelerada de las empresas, consumidores y administraciones hacia un mundo más digital y tecnológico. En la mayor parte de las sociedades se evidencia un cambio en la forma en la que los distintos agentes sociales se relacionan o interactúan. Esta digitalización acelerada también ha impregnado sustancialmente a las finanzas. Por una parte, las instituciones financieras han visto la necesidad de intensificar sus planes de transformación tecnológica asumiendo para ello la adopción de nuevas tecnológicas bancarias –*blockchain*, *cloud computing*, biometría, inteligencia artificial o *big data*–. Por lo que respecta a la demanda, los consumidores han experimentado un salto digital en la gestión de sus finanzas. El acceso a servicios financieros *online* se ha popularizado, incluso en segmentos de población que estaban menos digitalizados antes de la pandemia (Carbó-Valverde, Cuadros-Solas y Rodríguez-Fernández, 2021). Además, lejos de ralentizarse una vez superada la fase más dura de la pandemia, parece seguir avanzando. Por ello, desde el punto de vista de la demanda, es importante comprender cómo las personas están respondiendo, a raíz de la pandemia, a dicha transformación tecnológico-financiera resulta más relevante aún. Además, como han puesto en evidencia algunos estudios, la inclusión digital genera una serie de beneficios para los propios clientes. En general, el acceso a servicios financieros digitales permite a los clientes obtener unas mejores condiciones –menos comisiones– y la posibilidad de reducir los costes de transacción para acceder al crédito.

Este capítulo pretende aprovechar las ventajas que ofrecen las técnicas de aprendizaje automático para examinar el grado de digitalización financiera de los consumidores en un contexto pospandémico. En concreto, el presente trabajo pretende avanzar en el conocimiento del proceso de digitalización bancaria de dos formas. En primer lugar, usando las técnicas de *machine learning* para mejorar la fiabilidad de las predicciones sobre el grado de digitalización bancaria de los consumidores. Y en segundo lugar, esclareciendo qué tipos de factores son los que inciden en mayor medida en dicho grado de digitalización.

Para tratar de conseguir los objetivos propuestos, el trabajo examina el proceso de digitalización financiera considerando cuatro dimensiones diferentes de este fenómeno: la adopción de la banca digital, el alcance de uso de la banca digital, la conversión en cliente de un banco exclusivamente digital y la adopción de medios de pagos no bancarios. El examen diferenciado de cada una de estas dimensiones ayuda a dar una respuesta más específica y detallada a las diferentes aristas que plantea el proceso de digitalización financiera. Desde un punto de vista empírico, este trabajo emplea una encuesta sobre el uso de los servicios bancarios digitales, los medios de pago digitales y la adopción de los criptoactivos realizada en diciembre de 2021 por 2.121 individuos de entre 18 a 70 años residentes en España. Sobre la base de dicha encuesta, representativa de la población española adulta, se emplean un conjunto de técnicas de aprendizaje automático. En concreto, el estudio utiliza dos de los algoritmos más populares en el campo del aprendizaje automático supervisado: *random forest* y *extreme gradient boosting*. Ambas técnicas presentan un conjunto de ventajas metodológicas para tratar de responder a la pregunta de investigación.

A modo de resumen, el presente trabajo concluye que las técnicas de aprendizaje automático son capaces de predecir con un elevado grado de precisión el grado de digitalización financiera de los consumidores. Además, muestra que los principales factores que predicen el grado de digitalización financiera son la necesidad/utilidad percibida de los consumidores sobre la banca digital para consultar saldos, transferir dinero y comunicarse con el banco. Sin embargo, los factores sociales –edad, población, nivel de estudios, nivel de ingresos, situación laboral– son los más relevantes para predecir quiénes usan medios de pago no bancarios y quiénes son clientes de bancos 100 % digitales. Estos resultados sugieren que existen diferencias relevantes en cómo se produce el proceso de digitalización financiera dependiendo de si el proveedor de dichos servicios es una entidad financiera tradicional o si se trata de un nuevo proveedor de servicios financieros (neobanco, *fintech* y/o *bigtech*).

El capítulo se estructura en cinco secciones que siguen a esta introducción. En la segunda sección se hace un repaso por el potencial que tiene el aprendizaje automático para examinar el proceso de digitalización financiera. En la sección tercera se describen los datos que van a emplear para examinar la pregunta de investigación. La metodología empleada, con una descripción pormenorizada de las técnicas de *machine learning* usadas, se analiza en la cuarta sección. La quinta sección se centra en presentar los resultados empíricos obtenidos y en discutir las implicaciones de dichos resultados en el entorno pos-COVID. Finalmente, la sexta sección presenta las conclusiones.

## 2. LA DIGITALIZACIÓN FINANCIERA A TRAVÉS DEL *MACHINE LEARNING*

Numerosos estudios han tratado de examinar cómo se produce el proceso de digitalización de los individuos. Es decir, qué determina que los individuos adopten el canal digital para realizar actividades que anteriormente realizaban de una manera presencial. Si bien, examinar el proceso de digitalización es complicado, ya que existen varios y complejos factores

que impulsan la digitalización de las personas (Pousttchi y Dehnert, 2018). En lo que respecta a la digitalización financiera, varios estudios empíricos han corroborado que dicho proceso es multifactorial (Carbó-Valverde, Cuadros-Solas y Rodríguez-Fernández, 2020; Lin, Wang y Hung, 2020), donde factores socioeconómicos, conductuales, aptitudinales, perceptuales y/o relacionales juegan, todos ellos, un papel relevante.

Desde un punto de vista metodológico, para examinar las preferencias de los consumidores con respecto al uso de los medios digitales para realizar pagos u otros servicios financieros, la mayoría de los estudios anteriores han empleado técnicas econométricas estándares. Entre ellas, han predominado el uso de modelos de elección discreta (Hernández-Murillo, Llobet y Fuentes, 2010; Yusuf Dauda y Lee, 2015) y de ecuaciones estructurales (Aldás-Manzano *et al.*, 2009; Montazemi y Qahri-Saremi, 2015). A pesar de las bondades que presentan estos modelos econométricos, parece que la complejidad que caracteriza el proceso de digitalización financiera puede requerir del uso de otras técnicas que permitan ser capaces de aprovechar el uso de cantidades ingentes de datos para poder revelar patrones ocultos en el comportamiento digital de los individuos.

En este sentido, los métodos de aprendizaje automático se revelan como herramientas poderosas para aportar luz acerca del comportamiento de los consumidores (Cui, Wong y Lui, 2006; Witten, Frank y Hall, 2019; Lecun, Bengio y Hinton, 2015). La utilización de algoritmos de aprendizaje automático permitiría revelar los patrones complejos que impulsan el proceso de digitalización, ya que estos algoritmos pueden identificar patrones o conocimientos complejos y no obvios. El uso de estas técnicas puede ayudar a los modelos econométricos a revelar aquellos patrones de comportamiento que pueden estar más ocultos. Algunos estudios coinciden en que las técnicas de aprendizaje automático son adecuadas y efectivas para estudios de la demanda, ya que revelan patrones complejos (Bajari *et al.*, 2015).

Estos avances para examinar estas cuestiones relativas al comportamiento de los consumidores han motivado que recientes estudios las hayan comenzado a utilizar en diversos campos de la economía y de las finanzas: para encontrar patrones ocultos que pueden ser valiosos para la toma de decisiones en *marketing* bancario (Miguéis, Camanho y Borges, 2017) estimar las preferencias de los consumidores por productos tecnológicos (Chen, Honda y Yang, 2013), examinar las opciones de viaje (Hagenauer y Helbich, 2017) o modelar la respuesta del consumidor (Cui, Wong y Lui, 2006).

### 3. DATOS

#### 3.1. Encuesta

El estudio empírico se basa en una encuesta realizada por la empresa IMOP para Funcas<sup>1</sup> sobre el uso de los servicios bancarios digitales, los medios de pago digitales y la adopción

---

<sup>1</sup> El presente estudio se enmarca en las actividades realizadas por el Observatorio de la Digitalización Financiera (ODF) de Funcas.

de los criptoactivos<sup>2</sup>. Sigue la base metodológica de la *Survey of Consumer Payment Choice (SCPC)*, elaborada por el Banco de la Reserva Federal de Boston y usada ampliamente en la literatura académica para examinar cuestiones relativas al uso de medios de pagos (Schuh y Stavins, 2016; Kahn y Liñares-Zegarra, 2016). En concreto, fue realizada de forma *online* a 2.121 residentes en el territorio nacional –peninsular e insular– de entre 18 a 70 años.

Al tratarse de una encuesta *online*, es necesario advertir que el presente trabajo pretende examinar el grado de digitalización financiera de la población que ya tiene un determinado nivel de capacitación digital. Dicha elección metodológica no resta relevancia a las conclusiones que se puedan extraer del capítulo, ya que, según las estadísticas oficiales del Instituto Nacional de Estadística, el 91,8 % de la población española entre 16 y 74 años es usuaria, al menos semanalmente, de internet<sup>3</sup>.

El trabajo de campo tuvo lugar entre el 30 de noviembre y el 23 de diciembre de 2021, una vez superada la fase más dura de la pandemia. Para garantizar la representatividad de la muestra en el proceso de selección muestral se establecieron cuotas controladas por edad, sexo, comunidad de residencia y tamaño de la población. En términos agregados, el error muestral se estima en un  $\pm 2,2$  % para un nivel de confianza del 95,5 %.

La relevancia del presente estudio en España tiene su justificación. Por un lado, España es un país que, de acuerdo con las estadísticas nacionales, ha superado la fase inicial de implantación de la banca electrónica. Todas las entidades financieras que operan en el país ofrecen a sus clientes la posibilidad de acceder a los distintos servicios financieros a través del canal digital. Además, una amplia base de la población española, cerca del 70 % de la población, ha accedido en los últimos tres meses a la banca digital, ya sea a través de la web de su banco o accediendo con la aplicación bancaria instalada en el móvil. Por otra parte, es uno de los países que ha experimentado una mayor aceleración en la digitalización de la población a raíz de la pandemia (Comisión Europea, 2022). Por estos motivos parece oportuno examinar el grado de digitalización financiera en un país donde los servicios bancarios digitales juegan un papel relevante, y más especialmente, a raíz de la pandemia de COVID-19.

### 3.2. Variables

La encuesta completa recoge un conjunto amplio de preguntas que dan lugar a distintas variables que podemos clasificar en:

- **Demografía, economía y trabajo:** en ellas se recopila información socioeconómica de los individuos como el sexo, edad, nivel de estudios, nivel de ingresos, población y comunidad autónoma de residencia.

<sup>2</sup> Un resumen de los principales resultados de dicha encuesta puede encontrarse en <https://www.funcas.es/articulos/digitalizacion-financiera-en-la-pospandemia-que-ha-cambiado/>

<sup>3</sup> Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares. INE. Disponible en: <https://www.ine.es/dynt3/inebase/es/index.htm?padre=8321capsel=8324>

- **Educación financiera y habilidades digitales:** recoge información sobre el nivel de educación financiera de los individuos y acerca de su nivel de capacitación para realizar determinadas actividades usando el canal digital (por ejemplo, si el individuo es capaz de descargar la aplicación móvil de su banco sin ayuda de un familiar o amigo).
- **Frecuencia de uso del canal digital:** recoge un conjunto de preguntas sobre la frecuencia con la que se accede a la banca *online*. También existen preguntas sobre la frecuencia con la que se acude a la sucursal de su entidad financiera.
- **Cuentas y servicios bancarios:** incluye preguntas sobre el número de cuentas bancarias y de otros servicios financieros que tiene contratados el individuo. Además, también se pregunta si alguna de dichas cuentas bancarias tiene un acceso exclusivamente digital.
- **Relación con la entidad bancaria:** recoge preguntas sobre el grado de relación que tiene el individuo con la que considera su principal entidad financiera.
- **Satisfacción con la entidad bancaria:** recopila la opinión del individuo acerca del grado de satisfacción con su entidad financiera principal.
- **Percepción canal digital:** incluye un conjunto de preguntas sobre la percepción que tiene el usuario sobre el coste, dificultad y seguridad de la banca electrónica.
- **Grado de uso del canal digital:** recoge preguntas sobre el tipo de servicios financieros –consulta de saldos, pagos, realización de transferencias y comunicación con el banco– que el individuo realiza a través del canal digital.

### 3.3. Muestra

El **cuadro 1** recoge la distribución muestral. Dichos porcentajes están muy correlacionados con los proporcionados por las estadísticas oficiales lo que justifica su representatividad.

#### Cuadro 1.

#### Distribución muestral

	Número	Porcentaje
<i>Sexo</i>		
Varón	1.078	50,83
Mujer	1.043	49,17
<i>Edad</i>		
De 18 a 29 años	414	19,52
De 30 a 39 años	359	16,93
De 40 a 49 años	547	25,79
De 50 a 59 años	392	18,48
De 60 a 70 años	409	19,28

## Cuadro 1. (continuación)

**Distribución muestral**

	Número	Porcentaje
<i>Hábitat</i>		
Hasta 10.000 habitantes	316	14,90
De 10.001 a 50.000 habitantes	517	24,30
De 50.001 a 200.000 habitantes	528	24,89
De 200.001 a 500.000 habitantes	298	14,05
Más de 500.000 habitantes	462	21,78
<i>Nivel de estudios</i>		
Sin estudios	6	0,28
Estudios primarios o equivalentes	51	2,40
Enseñanza General Secundaria	366	17,26
Enseñanza de Formación Profesional	511	24,09
Universitarios	928	43,75
Posgrado o Doctorado	259	12,21
<i>Situación laboral</i>		
Trabaja	1.384	65,25
Pensionista/jubilado	297	14,00
En paro	212	10,00
Estudiante	182	8,58
Trabajo doméstico no remunerado	46	2,17
<i>Tamaño de la muestra</i>		
	2.121	100

**3.4. Dimensiones de la digitalización financiera**

Examinar el grado de digitalización financiera de un individuo es una tarea compleja, ya que el concepto de digitalización financiera es un concepto amplio. La interpretación de dicho concepto puede hacer que por “estar digitalizado” se entienda, simplemente tener acceso a la banca electrónica o hacer un uso intensivo del canal digital. Asimismo, la digitalización también puede englobar no solamente la frecuencia de uso, sino también la amplitud de las actividades realizadas *online*. De igual modo, con la revolución *Fintech* parte de dicho proceso de conversión en cliente financiero digital puede pasar por el uso de servicios financieros ofrecidos por entidades distintas a las entidades financieras tradicionales (neobancos, *fintech* o *bigtech*, entre otros). En definitiva, el proceso de digitalización financiera no es unidimensional, sino por el contrario, multidimensional. En un sentido amplio puede hablarse de varias dimensiones de la digitalización. Es por ello, que, en el presente estudio, para poder examinar con un mayor grado de profundidad el proceso de digitalización financiera se consideran explícitamente distintas dimensiones de la digitalización. En concreto, se examinan las siguientes cuatro dimensiones:

- *Adopción de la banca digital*: al examinar dicha dimensión se pretende conocer qué determina que los individuos se conviertan en usuarios frecuentes de los servicios bancarios a través del canal digital. El elemento clave que busca identificar dicha

dimensión es la frecuencia con la que el individuo hace uso de la banca digital. Dicha frecuencia de uso determinará su grado de adopción.

- *Alcance de uso de la banca digital*: con el alcance en el uso de la banca digital se busca conocer hasta qué punto los individuos llevan a cabo una pluralidad de actividades bancarias *online*. Realizar una distinción entre adopción y alcance es relevante, un individuo puede acceder muy frecuentemente a la banca electrónica únicamente para realizar una única actividad financiera. En este caso, hablaríamos de un individuo con un elevado grado de adopción de la banca digital, pero donde el alcance de su uso es muy limitado. En concreto, para medir dicha dimensión se consideran cuatro actividades financieras que pueden realizarse usando el canal digital: consultar saldos, realizar transferencias a terceros, realizar pagos de facturas o recibos y comunicarse con el banco.
- *Cliente de un banco exclusivamente digital*: con esta dimensión se pretende conocer qué determina que los individuos tengan contratada una cuenta bancaria con una entidad financiera que sea exclusivamente digital. En la medida en la que parte de la digitalización puede pasar por entidades nativas digitales, y no por las entidades financieras tradicionales, puede ser relevante explorar en profundidad dicha dimensión.
- *Adopción de medios de pagos no bancarios*: la actividad financiera donde se observa un mayor grado de digitalización es en la forma de pagar. Puesto que muchos de los nuevos medios de pagos son ofrecidos por entidades no bancarias (sin licencia bancaria), es relevante examinar cómo se produce la adopción de dichos pagos alternativos a los bancarios.

Para poder examinar cada una de las dimensiones en detalle, metodológicamente, los individuos son clasificados en diferentes categorías para cada una de las dimensiones analizadas. Respecto al grado de adopción de la banca digital, los 2.121 individuos de la muestra son clasificados en tres categorías. Aquel conjunto de individuos que no están digitalizados o lo están muy poco, ya que acceden a la banca *online* muy esporádicamente, es decir, una vez al año o casi nunca se consideran como “*No digitalizados (N)*”. Por su parte, identificamos como “*Medianamente digitalizados (M)*” a aquellos individuos que acceden a la banca *online* al menos una vez al mes o cada 15 días. Finalmente, aquellos individuos que acceden a la banca *online* casi todos los días o al menos una vez a la semana son categorizados como “*Ampliamente digitalizados (A)*”.

Por lo que respecta al alcance en el uso de la banca digital, los individuos son clasificados en cuatro categorías. Se consideran como “*No digitalizados (N)*” al conjunto de individuos que están fuera de la digitalización financiera porque no tienen acceso a la banca *online* o no realizan habitualmente (al menos semanalmente) ninguna actividad financiera. La categoría de “*Poco variado (P)*” la forman aquellos que solamente realizan una actividad financiera *online* habitualmente (al menos semanalmente). Aquellos individuos que realizan al menos dos actividades financieras *online* habitualmente (al menos semanalmente) son categorizados como “*Medianamente variado (M)*”. Por último, consideramos “*Omnidigital (O)*” a todo aquel

individuo que realiza al menos tres actividades financieras *online* habitualmente (al menos semanalmente).

La clasificación de los individuos en base a si son clientes de un banco exclusivamente digital es dicotómica. Se consideran “*No cliente (N)*” los que no son clientes de bancos exclusivamente digitales y “*Cliente (C)*” a los que sí son clientes. Del mismo modo, clasificamos a los individuos en función de si hacen uso de medios de pago a través de un proveedor no bancario (“*No usuario (N)*” vs. “*Usuario (S)*”).

#### 4. METODOLOGÍA

Como se ha apuntado anteriormente, para tratar de dar respuesta a la pregunta de investigación del presente estudio se va a utilizar una metodología basada en el uso de técnicas de *machine learning*. En concreto, desde un punto de vista metodológico el trabajo se estructura en las siguientes etapas:

1. Predicción del grado de digitalización financiera de los consumidores después de la pandemia mediante el uso de algoritmos supervisados:
  - ▶ *random forest*
  - ▶ *extreme gradient boosting*
2. Elección del algoritmo que sea capaz de predecir con mayor precisión el grado de digitalización financiera de los consumidores después de la pandemia.
3. Identificación/Discriminación de los factores que contribuyen en mayor medida a predecir (factores de primer orden) la digitalización financiera de los consumidores.
4. Predicción del perfil digital de un cliente bancario representativo

La selección de estos dos algoritmos –*random forest* y *extreme gradient boosting*– está justificada por su bondad predictiva y su explicabilidad. Por un lado, como han señalado diversos estudios, estos dos modelos tienen una elevada capacidad predictiva. Además, al ser algoritmos menos complejos como los de *deep learning* o redes neuronales gozan de una mayor explicabilidad.

##### 4.1. *Random forest*

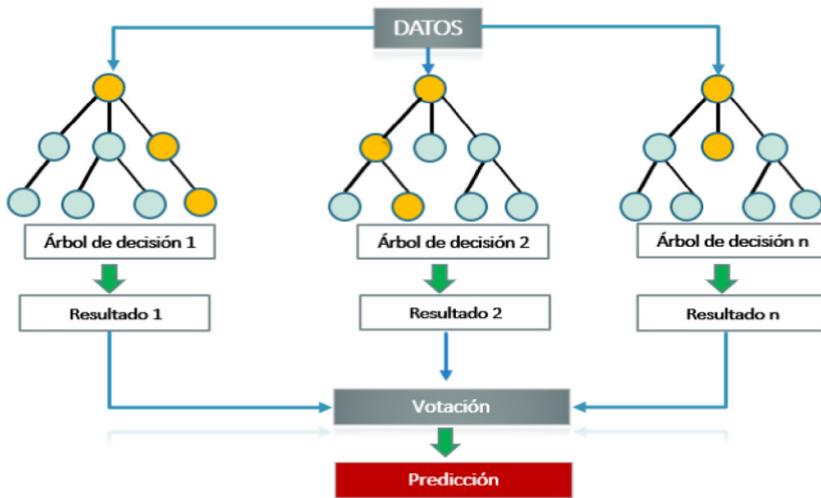
*Random forest* (bosque aleatorio) es un algoritmo de *machine learning* supervisado que se basa en la combinación de varios (muchos) árboles de decisión (Breiman, 2001). Dicha técnica combina una cantidad grande de árboles de decisión independientes probados sobre conjuntos de datos aleatorios con igual distribución. Para ello, el algoritmo sigue estos pasos:

1. Se genera un bosque de muchos árboles. Teniendo un conjunto  $N$  de observaciones diferentes, se elegirá una muestra  $N$  aleatoria con reemplazamiento.
2. Para cada nodo del árbol, se seleccionan  $m$  variables al azar de forma independiente entre todas las  $M$  variables posibles. Luego, en las  $m$  variables seleccionadas, encuentra la mejor división.
3. El algoritmo hace crecer cada árbol en la mayor medida posible.
4. Estos pasos se repiten en todos los árboles del conjunto, la predicción será la clase más votada entre todos los árboles del bosque.

La [figura 1](#) ilustra el funcionamiento que sigue dicho algoritmo.

Figura 1.

### Representación simplificada del *random forest*



Como han señalado estudios previos, la utilización del algoritmo del *random forest* presenta un conjunto de ventajas metodológicas con respecto a otros algoritmos de *machine learning* supervisados. Entre estas ventajas destacan que:

- Es una técnica fácil de interpretar, estable, que por lo general presenta buenas coincidencias con los resultados obtenidos con otras técnicas de *machine learning*.
- Se puede utilizar en tareas de regresión o de clasificación.
- El hecho de que pueda usarse para tareas de clasificación y regresión, combinado con su naturaleza no lineal, lo hace altamente adaptable a una variedad de datos y situaciones.

- Puede manejar cientos de predictores sin excluir ninguno. Además, es capaz de identificar cuáles son los predictores más importantes, por lo que esta técnica también se utiliza para reducir el número de variables empleadas en un estudio (reducción de dimensionalidad).
- Debido a la utilización de múltiples árboles de decisión, la varianza de la predicción disminuye respecto a un solo árbol de decisión, lo que reduce la posibilidad de obtener un modelo sobreadjustado.
- Se trata de un algoritmo cuyo entrenamiento es más simple en comparación con otras técnicas más complejas (redes neurales o *deep learning*), pero con un rendimiento similar.

Las ventajas que presenta el *random forest* para resolver problemas complejos ha motivado que sea utilizado en diversos campos:

- *Medicina*: para analizar historiales clínicos a fin de identificar enfermedades potenciales en los pacientes (Masetic y Subasi, 2016).
- *Sector bancario*: para detección de fraudes (Namiy Shajari, 2018) y clasificación de clientes para otorgamiento de crédito (Trivedi, 2020).
- *Sector financiero en sentido amplio*: para pronosticar comportamientos futuros de los mercados financieros (Ciner, 2019).
- *Marketing/Comportamiento del consumidor*: para pronosticar si un cliente comprará, o no, cierto producto (Miguéis, Camanho y Borges, 2017; Hagenauer y Helbich, 2017).

#### 4.2. *Extreme gradient boosting*

*Extreme gradient boosting* es un algoritmo de *machine learning* supervisado que se basa en la generación de múltiples modelos secuenciales de predicción, en los que cada nuevo modelo toma el resultado del modelo anterior, haciendo que los resultados obtenidos en cada secuencia sean cada vez más robustos, y, por tanto, obteniendo en cada iteración resultados más exactos. Se trata de una técnica de *machine learning* que surge de la idea de que, modelos sencillos mejorados pueden dar lugar a buenos modelos. Para ello, el algoritmo sigue estos pasos (ilustrados en la [figura 2](#)):

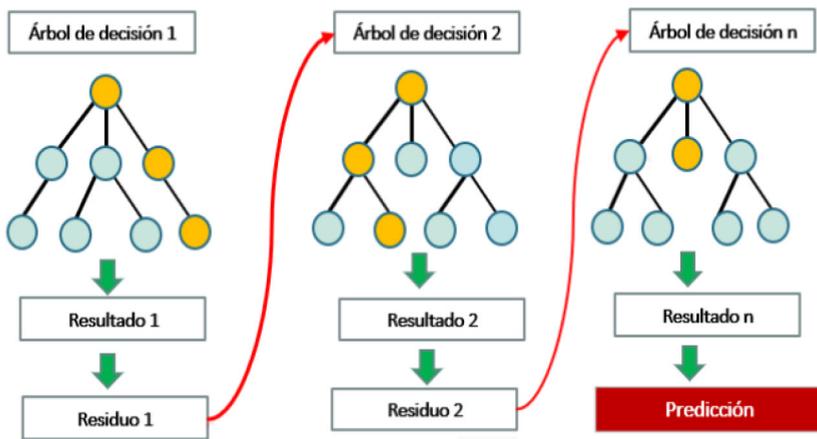
1. Durante el entrenamiento, los parámetros de cada modelo débil son ajustados iterativamente tratando de encontrar el mínimo de una función objetivo, que puede ser la proporción de error en la clasificación, el área bajo la curva (AUC), la raíz del error cuadrático medio (RMSE) o alguna otra.
2. Cada modelo es comparado con el anterior, Si un nuevo modelo tiene mejores resultados, entonces se toma este como base para realizar nuevas modificaciones. Si, por el

contrario, tiene peores resultados, se regresa al mejor modelo anterior y se modifica ese de una manera diferente.

- Este proceso se repite hasta llegar a un punto en el que la diferencia entre modelos consecutivos es insignificante, lo cual nos indica que hemos encontrado el mejor modelo posible, o cuando se llega al número de iteraciones máximas definido por el usuario.

Figura 2.

### Representación simplificada del *extreme gradient boosting*



Al igual que ocurre con el *random forest*, el *extreme gradient boosting* también presenta un conjunto de bondades metodológicas que motivan su uso en la literatura académica, entre ellas:

- Puede utilizarse con valores faltantes (*missing values*). Cuando encuentra un valor faltante en un nodo, intenta la división de la mano izquierda y derecha, y aprende el camino que conduce a una mayor pérdida para cada nodo. Luego hace lo mismo cuando trabaja en los datos de prueba.
- En comparación con otros métodos de aumento de gradiente (*gradient boosting*), el aumento de gradiente extremo utiliza una formalización de modelo más regularizada para controlar el sobreajuste.
- Es apropiado para base de datos grandes, donde el número de variables sea menor que el número de observaciones.
- Es más eficiente, requiriendo un menor tiempo de computación que otros algoritmos supervisados.

En los últimos años, también viene apreciándose como el uso del *extreme gradient boosting* también se está extendiendo a numerosos campos como:

- *Medicina*: para analizar historiales clínicos a fin de diagnosticar el cáncer (Ma *et al.*, 2020).
- *Sector financiero*: para predecir crisis bancarias (Carmona, Climent y Momparler, 2019).
- *Marketing*: para predecir la popularidad *online* (Nguyen, 2020).
- *Energía*: para predecir crisis en el mercado del gas natural (Zhang y Hamori, 2020).

## 5. RESULTADOS

### 5.1. *Random forest vs. extreme gradient boosting*

Con el fin de seleccionar la técnica con el mayor grado de precisión, siendo consistente con el estándar utilizado en la literatura de *machine learning*, se designa aleatoriamente al 70 % de los datos como datos de entrenamiento (1.485 observaciones) y al 30 % restante (636 observaciones) como de prueba. Utilizamos dicho umbral ya que la literatura empírica ha demostrado que los mejores resultados se obtienen si la división de los datos está en el marco del 20-30 % de los datos para pruebas y del 70-80 % restante para entrenamiento. Al hacerlo, podemos determinar la precisión del modelo asegurando que el algoritmo en realidad está encontrando patrones reales en los datos y que no se está sobreajustando. El grado de precisión de ambos modelos se calcula como el porcentaje de predicciones correctas (el número de puntos de datos pronosticados correctamente en relación al número total de datos).

Como se puede observar en el **cuadro 2**, para todas las dimensiones consideradas, el *extreme gradient boosting* es el algoritmo de *machine learning* que reporta los mayores porcentajes de predicciones correctas. Para aquellas dimensiones donde los individuos están clasificados en más de dos categorías –grado de digitalización bancaria y omnidigitaliza-

Cuadro 2.

### Precisión del *random forest* y del *extreme gradient boosting*

	<i>Random forest</i>		<i>Extreme gradient boosting</i>	
	% de aciertos	95 % I.C	% de aciertos	95 % I.C
Grado de digitalización bancaria	79,73	(76,30 – 82,88)	79,96	(78,19 – 81,65)
Omnidigitalización	79,01	(76,86 – 81,22)	80,01	(78,24 – 81,69)
Cliente banco excl. digital	81,57	(79,52 – 83,49)	85,38	(83,81 – 86,86)
Pagos no bancarios	66,11	(62,18 – 69,89)	71,52	(69,55 – 73,44)

ción– consideramos que el algoritmo ha acertado si ha sido capaz de predecir exactamente la categoría de entre todas las posibles. En concreto, el *extreme gradient boosting* predice con una precisión del 79,96 % el grado de digitalización bancaria y con una precisión del 80,01 % la diversidad del uso digital de la banca en línea. Además, es capaz de predecir con una precisión del 85,38 % qué usuarios son clientes de bancos exclusivamente digitales y con un 71,52 % quiénes usan medios de pagos no bancarios. Aunque no han sido reportados por cuestiones de espacio, las dos técnicas de *machine learning* empleadas presentan unos porcentajes de predicciones correctas significativamente superiores a la obtenidas usando modelos de elección discreta (*ordered logit and logit*). Al usar estos modelos econométricos, en la mayor parte de las dimensiones, el porcentaje de predicciones correctas se encuentra cercano al 65 %.

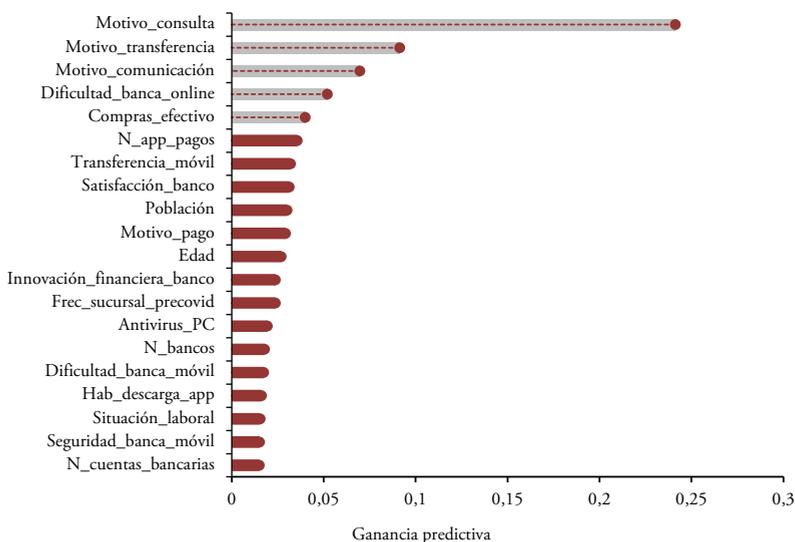
Puesto que el *extreme gradient boosting* es la técnica que presenta el mayor porcentaje de predicciones correctas, utilizaremos dicha técnica para identificar los factores que contribuyen en mayor medida a predecir (factores de primer orden) la digitalización financiera de los consumidores.

## 5.2. ¿Qué factores predicen mejor el grado de digitalización financiera?: importancia

Para tratar de discriminar cuáles son aquellos factores que predicen mejor el grado de digitalización financiera, en cada una de las dimensiones examinadas, vamos a calcular la

Figura 3.

### Importancia – Grado de digitalización financiera

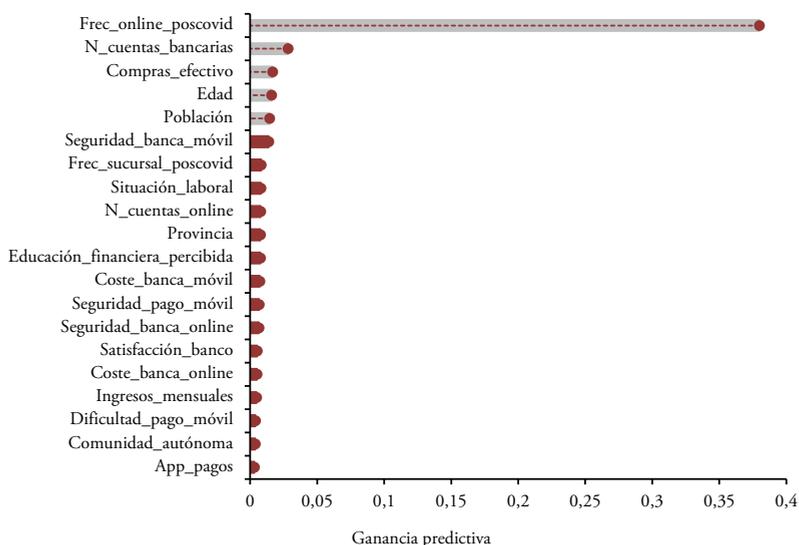


“ganancia predictiva” de cada uno de los factores (variables) considerados. Podemos definir la ganancia predictiva como la contribución relativa de cada factor a la capacidad predictiva del modelo. Un valor más alto de esta métrica implica que dicho factor es más importante para generar una predicción. En definitiva, con esta métrica pueden identificarse cuáles son las características que presentan una mayor importancia para predecir el grado de digitalización financiera.

La **figura 3** muestra los veinte factores que presentan una mayor ganancia predictiva (ordenados de mayor a menor) para predecir correctamente el grado de adopción de la banca digital. Como puede observarse, los principales factores que permiten clasificar a los individuos en función de su grado de adopción de la banca digital son la necesidad/utilidad percibida de los consumidores sobre la banca digital para consultar saldos, transferir dinero y comunicarse/relacionarse con el banco. En especial, destaca la consulta de saldos, lo que puede indicar que la adopción pasa primeramente por la necesidad de usar la banca electrónica para realizar tareas consultivas o informativas respecto a las finanzas personales. También puede observarse que, la percepción acerca de la dificultad de uso de la banca *online* es un factor que predice ampliamente el grado de digitalización financiera. Otros factores como las percepciones o ideas que se tienen acerca del coste y seguridad de la banca *online* no explican el grado de digitalización financiera. Asimismo, aquellos factores socioeconómicos como la edad, el lugar de residencia (rural vs. urbano), el nivel de ingresos o la situación laboral parecen revelarse como factores de segundo orden para predecir el grado de adopción de la banca digital por parte de los individuos.

Figura 4.

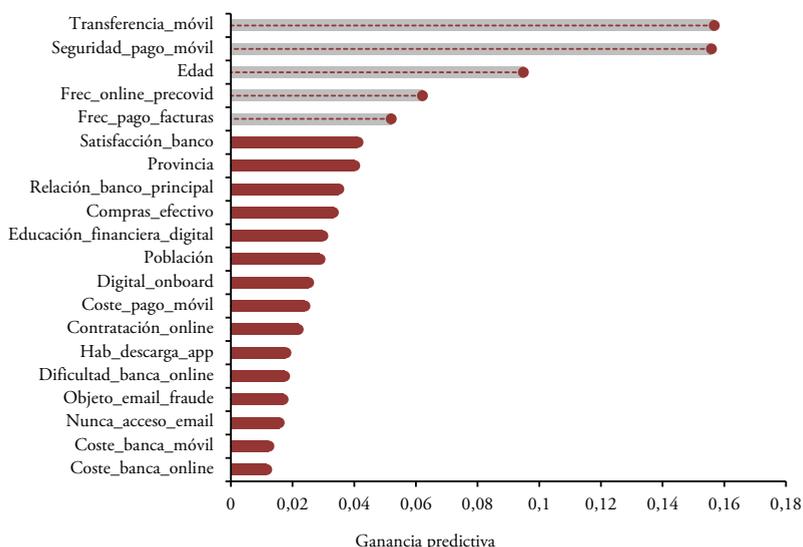
### Importancia – Alcance de la digitalización financiera



La **figura 4** muestra la importancia relativa para predecir correctamente el alcance en el uso de la banca digital. Como puede apreciarse, el principal factor, y casi único factor, que predice que un individuo pueda utilizar la banca *online* para realizar una pluralidad de actividades financieras es la frecuencia de acceso/uso de la banca *online*. Este resultado sugiere que la frecuencia en el uso de la banca *online* permitirá discriminar ampliamente quiénes son aquellos que acometen un uso más diversificado de la banca electrónica. En otras palabras, dicho resultado sugiere que no es posible alcanzar un elevado grado de diversificación de uso de la banca *online* si no existe previamente el hábito de acceder con una elevada frecuencia a la banca electrónica. Otros factores como las percepciones o ideas que se tienen acerca del coste, dificultad y seguridad de la banca digital no explican la diversificación en el uso de la banca *online*.

**Figura 5.**

### Importancia – Cliente de un banco exclusivamente digital



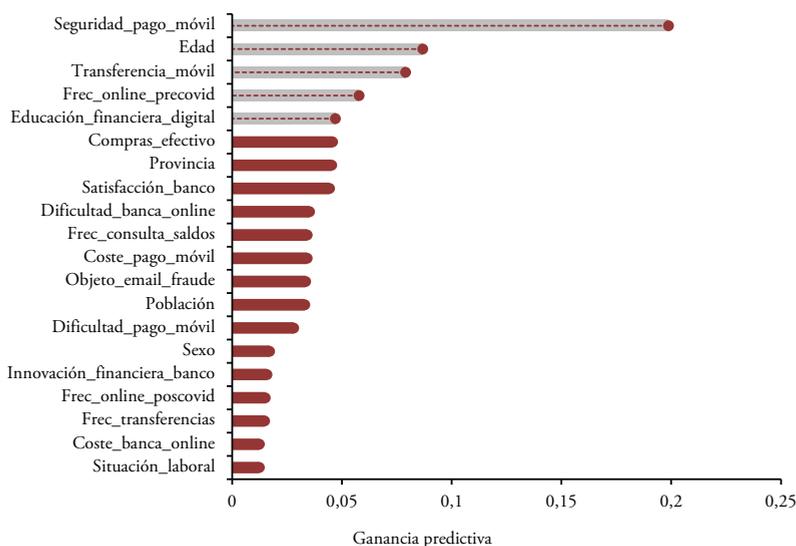
Por su parte, la **figura 5** recoge la importancia relativa de los factores que predicen que un individuo se convierta en un cliente de una entidad bancaria exclusivamente digital. Dicha figura revela que los principales factores que predicen que un consumidor sea cliente de un banco 100 % digital son, principalmente, la necesidad de transferir dinero a otros con el móvil y la percepción que se tenga acerca de la seguridad que tiene los pagos con el móvil. La relevancia que tienen los factores relacionados con el móvil y con los pagos parecen indicar que el proceso de conversión en cliente de un banco digital pasa por utilizar el móvil/*smartphone* como dispositivo para gestionar las finanzas, y especialmente para transferir dinero o para pagar. En este caso, factores como la edad sí parecen ser muy relevantes para predecir si un consumidor es cliente de un banco 100 % digital, lo que sugiere que los clientes de estos ban-

cos digitales pueden concentrarse en determinados segmentos de población. Además del grado de adopción de la banca digital, también aparecen como factores de segundo orden el grado de satisfacción que el consumidor tiene con su entidad financiera.

Los principales factores que predicen el uso de medios de pagos no bancarios están representados en la **figura 6**. Como se puede observar, la percepción acerca de la seguridad que existe para realizar pagos con el móvil es el primer factor que predice que un consumidor pague con medios de pagos ofrecidos por entidades no bancarias. Además, la edad también juega un papel importante para predecir el uso de medios de pagos no bancarios, lo que sugiere que puede existir una brecha en el uso de medios de pagos no bancarios por grupos de edades. También se observa cómo el nivel de educación financiera juega un papel relevante en el uso de medios de pagos no bancarios. Este resultado sugiere que carencias en el nivel de educación financiera pueden derivar en diferencias en el grado de uso de medios de pagos no bancarios.

Figura 6.

### Importancia – Usuario de pagos no bancarios



### 5.3. Tipología de factores

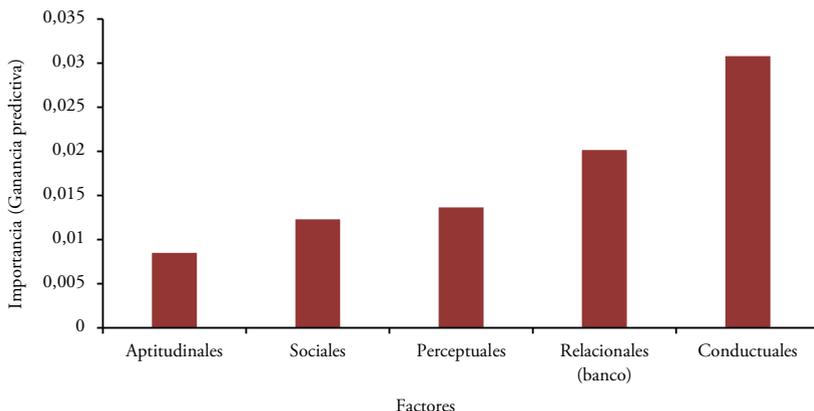
Para tratar de completar el ejercicio acerca de qué características presentan un mayor poder discriminatorio para predecir el grado de digitalización financiera, basándonos en la literatura académica, los distintos determinantes del grado de digitalización financiera los agrupamos en cinco tipologías de factores. El objetivo es proporcionar una imagen más amplia de qué está detrás de cada una de las diferentes dimensiones de la digitalización consideradas. En concreto, todos los factores (características) se agrupan en las siguientes:

- **Aptitudinales:** hacen referencia a aquellos determinantes referidos a la capacidad/habilidad que tiene cierta persona para poder desempeñarse competentemente en una actividad (ejemplo: habilidad para descargar una *app* bancaria sin ayuda de familiares o amigos).
- **Conductuales:** hacen referencia a aquellos determinantes referidos a la realización de determinadas conductas o actividades por parte de un individuo (ejemplo: verificar el cierre de la *app* bancaria o el porcentaje de compras que realiza en efectivo).
- **Relacionales (banco):** hacen referencia a aquellos determinantes referidos a la relación/afección que el individuo tiene con el sector bancario, en particular, con su entidad financiera de referencia (ejemplo: porcentaje de operaciones que realiza con su entidad financiera principal).
- **Sociales:** hacen referencia a aquellos determinantes referidos al ámbito socioeconómico y demográfico del individuo (ejemplo: sexo, población, nivel de ingresos).
- **Perceptuales:** hacen referencia a aquellos determinantes referidos a la percepción que el individuo tiene acerca del canal digital (ordenador y móvil) en el contexto bancario (ejemplo: percepción acerca del coste de la banca móvil).

Para cada una de las categorías, calculamos la importancia relativa –basándonos en la ganancia predictiva– con el objeto de conocer qué tipo de factores son los más relevantes para cada una de las dimensiones de la digitalización. Un mayor grado de ganancia predictiva indica que ese conjunto de factores es más relevante para predecir dicha dimensión.

Figura 7.

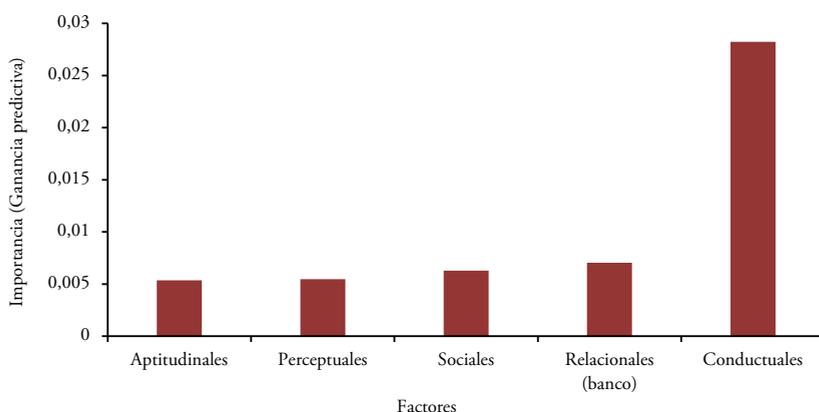
### Importancia por tipología de factores – Grado de digitalización financiera



Respecto a las dos dimensiones que tienen que ver con la digitalización bancaria (figuras 7 y 8) –adopción y el alcance de uso de la banca digital– se observa que los factores conductuales, es decir, los hábitos o costumbres adquiridas por los individuos sobre la forma de realizar actividades financieras, son los más relevantes seguidos de los relacionales. Por otra parte, la mayor relevancia de los factores relacionales frente a otros como los perceptuales y/o sociales sugieren que el grado de digitalización pasa en mayor medida por la intensidad de relaciones que los usuarios tienen con su banca en comparación a la percepción que puedan tener acerca de la banca electrónica o móvil. Estas figuras también revelan que, al menos para el grueso de la población, el nivel de aptitudes para usar el canal digital parece tener poca importancia para predecir el grado de adopción de la banca digital y la diversificación de su uso.

Figura 8.

### Importancia por tipología de factores – Alcance de digitalización financiera



Por otra parte, en relación a las otras dos dimensiones que hacen referencia al grado de digitalización financiera al margen del sector bancario tradicional –cliente exclusivo digital y uso de medios de pago no bancarios (figuras 9 y 10)– se observa que los factores sociales –edad, población, nivel de estudios, nivel de ingresos, situación laboral– son los más relevantes para predecir quiénes son los usuarios de medios de pago no bancarios como los clientes de bancos 100 % digitales.

Este resultado sugiere que tanto los usuarios de los bancos 100 % digitales, o también llamados “neobancos”, y de medios de pago no bancarios, ofrecidos por compañías *bigtech* (ApplePay, AmazonPay, etc.) o *fintech* (Verse, PayPal, etc.), pueden concentrarse en determinados segmentos de población muy concreta. Además, se puede apreciar cómo los factores perceptuales o relacionales juegan un papel secundario a la hora de explicar la transición de ser cliente de un banco tradicional a ser cliente/usuario de un neobanco o de una entidad no bancaria.

Figura 9.

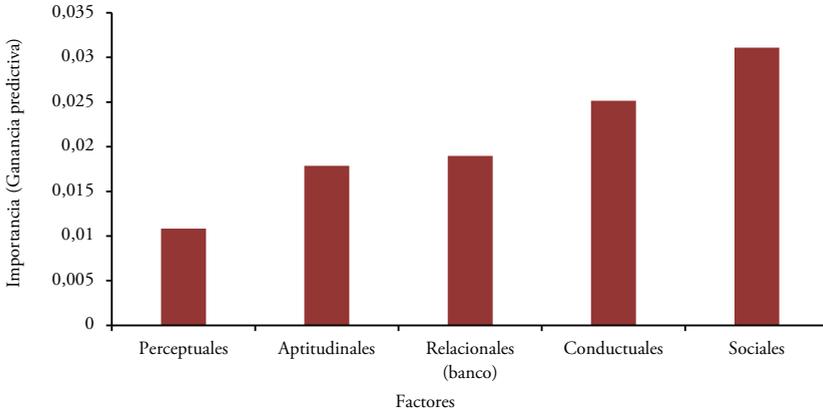
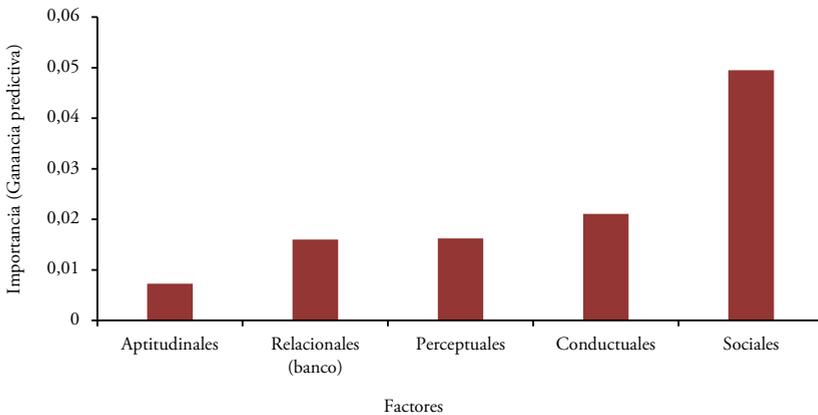
**Importancia por tipología de factores – Cliente de un banco exclusivamente digital**

Figura 10.

**Importancia por tipología de factores – Usuario de pagos no bancarios****5.4. El perfil digital de un cliente bancario representativo**

A partir de los datos de la encuesta, tratamos de predecir cuál es el perfil digital de un cliente bancario representativo. Para ello:

1. Para cada una de las variables empleadas se calcula la mediana para conocer el comportamiento de un individuo representativo.

2. Utilizando el *extreme gradient boosting* –que es el algoritmo con el mayor nivel de precisión– se predice cuál sería la categoría asociada a cada una de las cuatro dimensiones consideradas.

Tras seguir este procedimiento, se puede concluir que el perfil digital de un cliente bancario representativo se corresponde con un individuo:

- Ampliamente digitalizado: accede a la banca *online* casi todos los días o al menos una vez a la semana.
- Omnidigital: realiza al menos tres actividades financieras *online* habitualmente (al menos semanalmente).
- Que no acude a entidades exclusivamente digitales.
- Que es usuario de medios de pago no bancarios.

Este resultado confirma el elevado grado de digitalización bancaria del español medio, ya que hace un uso frecuente y variado de la banca digital. Además, también se aprecia una adopción significativa de medios de pagos no bancarios. Si bien, el consumidor representativo español no es cliente de entidades no bancarias, sino que los bancos tradicionales siguen siendo los principales proveedores de servicios financieros.

## 6. CONCLUSIONES

El salto digital de las sociedades se ha visto acentuado con la pandemia. El COVID-19 y las medidas de distanciamiento social adoptadas han provocado que las personas hayan adoptado el canal digital para realizar gestiones y actividades que, hasta hace poco, realizaban presencialmente. Este salto digital también ha impregnado el ámbito financiero, desde la realización de pagos hasta la contratación y gestión de las finanzas personales. La pandemia ha acelerado el grado de digitalización financiera de la sociedad. Si bien, a pesar del avance digital, existen diferencias en el uso del canal digital entre distintos segmentos de población. En este contexto, y una vez superada la fase más dura de la pandemia, resulta relevante examinar los factores que permiten predecir por qué algunos consumidores han avanzado más su grado de digitalización financiera. El conocimiento de dichos factores –sociales, conductuales, aptitudinales, perceptuales y relacionales– permite entender mejor cuáles pueden ser las trabas o barreras que pueden limitar la digitalización financiera de la sociedad.

Este estudio pretende profundizar en el conocimiento del proceso de digitalización financiera haciendo uso de técnicas de aprendizaje automático (*random forest* y *extreme gradient boosting*). Las ventajas que ofrecen estas técnicas de *machine learning* en relación con otras metodologías es que permiten conocer de una forma más detallada los patrones que afectan al comportamiento digital de los consumidores en lo que respecta a la gestión de sus finanzas personales. En concreto, el presente trabajo examina las distintas dimensiones del proceso de digitalización financiera. Por una parte, examina la digitalización banca-

ria al investigar los factores que predicen la adopción y el alcance de uso de la banca digital. Por otro lado, también analiza el grado de digitalización no-bancaria al examinar los determinantes que predicen la conversión en cliente de un banco exclusivamente digital y la adopción de medios de pagos no bancarios. El examen diferenciado de cada una de estas dimensiones ayuda a dar una respuesta más específica y detallada a las diferentes aristas que plantea el proceso de digitalización financiera. Para ello, el trabajo se basa en una encuesta sobre el uso de los servicios bancarios digitales, los medios de pago digitales y la adopción de los criptoactivos realizada en diciembre de 2021 por 2.121 individuos de entre 18 a 70 años residentes en España.

El trabajo concluye que, en comparación con otras técnicas, los dos algoritmos empleados ofrecen mejores predicciones acerca del grado de digitalización financiera de los consumidores. Además, los principales factores que predicen el grado de digitalización financiera son la necesidad/utilidad percibida de los consumidores sobre la banca digital para consultar saldos, transferir dinero y comunicarse/relacionarse con el banco. El principal factor, y casi único factor, que predice que un individuo pueda utilizar la banca digital para realizar una pluralidad de actividades financieras es la frecuencia de acceso/uso de la banca *online*. Por otra parte, los principales factores que predicen que un consumidor sea cliente de un banco 100 % digital son, principalmente, la necesidad de transferir dinero a otros con el móvil y la percepción que se tenga acerca de la seguridad que tiene los pagos con el móvil. Sin embargo, en lo que respecta a la adopción de medios de pago no bancarios, la percepción acerca de la seguridad que ofrece el móvil como método de pago es el primer factor que predice que un consumidor pague con medios de pagos ofrecidos por entidades no bancarias. Además, el trabajo evidencia el elevado grado de digitalización financiera de la sociedad española, ya que el cliente bancario representativo hace un uso frecuente y variado de la banca digital, a lo que se une, que paga con medios de pagos digitales no bancarios.

Estos resultados pueden tener importantes implicaciones para el conjunto de la industria financiera, ya que evidencian diferencias en el proceso de digitalización financiera en función de quién es el proveedor de los servicios financieros (banco tradicional vs. entidad bancaria digital). Para las entidades bancarias, estos resultados ponen de manifiesto la necesidad de avanzar hacia aplicaciones bancarias personalizadas que mejoren la actual experiencia de cliente con el objetivo de que éstos puedan en todo momento percibir la banca digital como un elemento que les ayuda a realizar sus actividades financieras básicas (consulta de saldos, realización de transferencias y comunicarse con su banco). Además, entender el proceso de digitalización bancaria es relevante, ya que dicho proceso puede generar externalidades positivas que van más allá del propio ámbito bancario. En muchos casos, se observa cómo algunos segmentos de población han avanzado en sus competencias digitales gracias al uso de la banca electrónica. El aprendizaje que supone la descarga de la aplicación móvil del banco y su uso ha facilitado que el mismo usuario pueda realizar otras actividades digitales –por ejemplo, pedir cita *online* a una Administración pública o realizar compras por internet.

## Referencias

- ALDÁS-MANZANO, J., LASSALA-NAVARRÉ, C., RUIZ-MAFÉ, C. y SANZ-BLAS, S. (2009). The role of consumer innovativeness and perceived risk in online banking usage. *International Journal of Bank Marketing*, 27, pp. 53–75.
- BAJARI, P., NEKIPELOV, D., RYAN, S. P. y YANG, M. (2015). Machine Learning Methods for Demand Estimation. *American Economic Review*, 105, pp. 481–485.
- BREIMAN, L. (2001). Random forests. *Machine learning*, 45, pp. 5–32.
- CARBÓ-VALVERDE, S., CUADROS-SOLAS, P. y RODRÍGUEZ-FERNÁNDEZ, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. *PlosOne*, 15(10).
- CARBÓ-VALVERDE, S., CUADROS-SOLAS, P. y RODRÍGUEZ-FERNÁNDEZ, F. (2021). Digitalización financiera y COVID-19: evidencia empírica. *Papeles Economía Española*, 170, pp. 143–156.
- CHEN, H. Q., HONDA, T. y YANG, M. C. (2013). Approaches for Identifying Consumer Preferences for the Design of Technology Products: A Case Study of Residential Solar Panels. *Journal of Mechanical Design*, 135, 061007 (2013).
- CINER, C. (2019). Do industry returns predict the stock market? A reprise using the random forest. *The Quarterly Review of Economics and Finance*, 72, pp. 152–158.
- COMISIÓN EUROPEA. (2022). Digital Economy and Society Index (DESI).
- CORREIA LOUREIRO, S. M., RÜDIGER KAUFMANN, H. y RABINO, S. (2014). Intentions to use and recommend to others. *Online Information Review*, 38, pp. 186–208.
- CUI, G., WONG, M. L. y LUI, H. K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, 52, pp. 597–612.
- ESTRELLA-RAMON, A., SÁNCHEZ-PÉREZ, M. y SWINNEN, G. (2016). How customers' offline experience affects the adoption of online banking. *Internet Research*, 26, pp. 1072–1092.
- FU, J. y MISHRA, M. (2022). Fintech in the time of COVID-19: Technological adoption during crises. *Journal of Financial Intermediation*, 50, 100945.
- HAGENAUER, J. y HELBICH, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, pp. 273–282.
- HERNÁNDEZ-MURILLO, R., LLOBET, G. y FUENTES, R. (2010). Strategic online banking adoption. *Journal of Banking and Finance*, 34, pp. 1650–1663.
- HITT, L. M. y FREI, F. X. (2002). Do Better Customers Utilize Electronic Distribution Channels? The Case of PC Banking. *Management Science*, 48, pp. 732–748.
- HONKA, E., HORTA, A. y VITORINO, M. A. (2017). Advertising, Consumer Awareness and Choice: Evidence from the U.S. Banking Industry. *RAND Journal of Economics*, 48, pp. 611–646.
- JÜNGER, M. y MIETZNER, M. (2020). Banking goes digital: The adoption of FinTech services by German households. *Finance Research Letters*, 34, 101260.
- KAHN, C. M. y LIÑARES-ZEGARRA, J. M. (2016). Identity Theft and Consumer Payment Choice: Does Security Really Matter? *Journal of Financial Services Research*, 50, pp. 121–159.
- KOULAYEV, S., RYSMAN, M., SCHUH, S. y STAVINS, J. (2016). Explaining adoption and use of payment instruments by US consumers. *RAND Journal of Economics*, 47, pp. 293–325.
- LECUN, Y., BENGIO, Y. y HINTON, G. (2015). Deep learning. *Nature*, 521, pp. 436–444.

- LIN, W. R., WANG, Y. H. y HUNG, Y. M. (2020). Analyzing the factors influencing adoption intention of internet banking: Applying DEMATEL-ANP-SEM approach. *PLoS One*, 15, pp. 1–25.
- MA, B. *et al.* (2020). Diagnostic classification of cancers using extreme gradient boosting algorithm and multinomics data. *Computers in Biology and Medicine*, 121, 103761.
- MASETIC, Z. y SUBASI, A. (2016). Congestive heart failure detection using random forest classifier. *Comput. Computer Methods and Programs in Biomedicine*, 130, pp. 54–64.
- MIGUÉIS, V. L., CAMANHO, A. S. y BORGES, J. (2017). Predicting direct marketing response in banking: Comparison of class imbalance methods. *Service Business*, 11, pp. 831–849.
- MONTAZEMI, A. R. y QAHRI-SAREMI, H. (2015). Factors affecting adoption of online banking: A meta-analytic structural equation modeling study. *Information and Management*, 52, pp. 210–226.
- NAMI, S. y SHAJARI, M. (2018). Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors. *Expert Systems with Applications*, 110, pp. 381–392.
- NGUYEN, L. T. K., CHUNG, H. H., TULIAO, K. V. y LIN, T. M. Y. (2020). Using XGBoost and Skip-Gram Model to Predict Online Review Popularity. *SAGE Open*, 10.
- POUSTTCHI, K. y DEHNERT, M. (2018). Exploring the digitalization impact on consumer decision-making in retail banking. *Electronic Markets*, 28, pp. 265–286.
- SCHUH, S. y STAVINS, J. (2016). How Do Speed and Security Influence Consumers' Payment Behavior? *Contemporary Economic Policy*, 34, pp. 595–613.
- STAVINS, J. (2020). Credit Card Debt and Consumer Payment Choice: What Can We Learn from Credit Bureau Data? *Journal of Financial Services Research*, 58, pp. 59–90.
- SZOPIŃSKI, T. S. (2016). Factors affecting the adoption of online banking in Poland. *Journal of Business Research*, 69, pp. 4763–4768.
- TRIVEDI, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, 63, 101413.
- WITTEN, I. H., FRANK, E. y HALL, M. A. (2019). *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*. *Morgan Kaufmann*, 53 (Morgan Kaufmann, 2019).
- YUSUF DAUDA, S. y LEE, J. (2015). Technology adoption: A conjoint analysis of consumers' preference on future online banking services. *Information Systems*, 53, pp. 1–15.
- ZHANG, W. y HAMORI, S. (2020). Do machine learning techniques and dynamic methods help forecast US natural gas crises? *Energies*, 13(9), 2371