

Tipo de documento: Tesis de maestría



Escuela de Negocios. Maestría en Finanzas

Regulación Financiera y datos masivos: una revisión del uso actual de big data en las entidades financieras internacionales

Autoría: Baez, Joanna

Año: 2024

¿Cómo citar este trabajo?

Baez, J.(2024). *Regulación Financiera y datos masivos: una revisión del uso actual de big data en las entidades financieras internacionales*. [Tesis de maestría. Universidad Torcuato Di Tella].

Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/13014>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial- Compartir igual 4.0 Argentina
Dirección: <https://repositorio.utdt.edu>

Trabajo Final de Graduación

Maestría en Finanzas UTDT

Año Académico 2024

Alumna: Joanna Baez

Tutor: Mariano Parnás

Regulación Financiera y datos masivos: una revisión del uso actual de big data en las entidades financieras internacionales

Índice

Resumen.....	4
Introducción	5
3. Cambios tecnológicos y surgimiento de big data.....	7
3.1 Definición de Big Data	7
3.2 Tipos de Datos.....	11
3.3 Técnicas de procesamiento de datos masivos	12
3.3.1 Machine Learning o Aprendizaje Automático	13
3.3.1.1 Aprendizaje Automático Supervisado	14
3.3.1.2 Aprendizaje Automático no Supervisado	14
3.3.1.3 Aprendizaje Automático por Refuerzo.....	14
3.3.1.4 Aprendizaje Profundo o Deep Learning.....	15
3.3.1.5 Minería de Textos.....	15
3.3.1.6 Análisis de Redes	15
3.3.1.7 Árboles de Decisión.....	16
3.3.1.8 Bosques Aleatorios.....	16
3.3.1.9 Redes Neuronales.....	16
3.3.1.10 Procesamiento Lenguaje Natural y Modelos de Lenguaje Grandes.....	17
3.4 ¿Qué es big data para los bancos centrales?	18
4. ¿Por qué es importante el uso de Big Data en la regulación financiera internacional?	19
5. Aplicación de Big Data en la Regulación Financiera.....	23
5.1 Estabilidad Financiera.....	26
5.2 Gestión de Riesgos.....	28
5.3 Investigación económica y predicción.....	31
5.4 Supervisión y Detección de Fraudes	37
5.4.1 Supervisión Micro.....	38
5.4.2 Entidades no supervisadas	39
5.4.3 Políticas de estabilidad macro financieras	39
5.4.4 Regulación antilavado de dinero y financiamiento del terrorismo	40
5.4.4.1 Herramientas usadas por los supervisores de Antilavado de Dinero y Financiamiento del Terrorismo	42
5.4.4.2 Herramientas usadas por las Unidades de Inteligencia Financiera	43
5.4.4.3 Identificación de patrones y tendencias en actividades delictivas	43
6. Desafíos y limitaciones	45
6.1 Regulación de IA privada.....	49

6.2 Regulación de IA pública.....	49
7. Conclusiones.....	51
8. Referencias bibliográficas.....	53
9. Anexo.....	56

Resumen

El uso de big data en la regulación financiera internacional se ha convertido en una herramienta esencial para abordar los complejos desafíos del sistema financiero global. El presente trabajo ofrece una revisión sobre el uso actual de big data por parte de las entidades financieras reguladoras en las distintas áreas que son de su competencia. A través de un repaso de la literatura se analizan diversas aplicaciones de big data que van desde la garantía de la estabilidad financiera y la gestión de los distintos tipos de riesgos financieros, la investigación económica, así como la prevención del fraude y el desarrollo de políticas regulatorias. En conjunto, este estudio contribuye al cuerpo existente de conocimiento al proporcionar una comprensión más profunda de cómo el uso de big data está impactando en la regulación financiera internacional y ofreciendo nuevas oportunidades para fortalecer la estabilidad y eficiencia de los mercados financieros a nivel global, así como también, dejando de manifiesto cuáles son los desafíos que tiene por delante.

Introducción

Para comprender la realidad actual es imprescindible reconstruir nuestro pasado y, para eso, es necesario reconocer que el desarrollo material de nuestra sociedad ha estado conducido por los cambios generados en la producción. A lo largo de la historia mundial se han sucedido profundas transformaciones económicas, políticas, sociales y culturales que modificaron radicalmente nuestra sociedad. Una forma de ejemplificar la periodicidad de dichos cambios es a través de las denominadas revoluciones industriales o, como llama Carlota Pérez, revoluciones tecnológicas. En su libro, Pérez (2003) instala el concepto de paradigma tecno económico para explicar que cada período es regido por una transformación tecnológica que da lugar a la explosión de nuevos productos, industrias e infraestructuras. El crecimiento económico desde fines del siglo XVIII hasta la actualidad está atravesado por 5 etapas distintas, asociadas a 5 revoluciones tecnológicas sucesivas.

La primera revolución tecnológica la identifica con la revolución industrial comenzada en Inglaterra a fines del siglo XVIII. En dicho período las nuevas tecnologías fueron la mecanización de la industria del algodón, el hierro forjado y el desarrollo de la maquinaria. Las infraestructuras nuevas fueron los canales y vías fluviales y la energía hidráulica. Este primer sistema tecnológico sirvió de base para el desarrollo de los siguientes, ya que, como menciona la autora, todos los procesos son interdependientes.

La segunda revolución es la era del vapor y los ferrocarriles a principios del siglo XIX donde identifica que el big bang iniciador de la revolución fue la prueba del motor a vapor *Rocket* para el ferrocarril Liverpool-Manchester. Aquí se identifica como nueva tecnología a las máquinas de vapor y maquinaria de hierro, movida con carbón. Con la construcción de los ferrocarriles, la producción de vagones y locomotoras fueron otras de las nuevas técnicas. Esto trajo consigo el desarrollo de infraestructuras como el servicio postal estandarizado de plena cobertura, el telégrafo y grandes puertos, depósitos y barcos para la navegación mundial.

Luego, la tercera revolución es conocida como la era del acero, la electricidad y la ingeniería pesada hacia fines del siglo XIX cuyo evento inaugurador fue la apertura de la acería Bessemer de Pennsylvania. Unas de las principales técnicas de esa etapa son el acero barato, ingeniería pesada química y civil, industria de equipos eléctricos, cobre y cables y alimentos embotellados y enlatados. En este caso, la infraestructura nueva que acompañó fue el desarrollo de la navegación mundial en veloces barcos de acero, redes trasnacionales de ferrocarril, grandes puentes y túneles, teléfono y redes eléctricas (para uso particular y para uso industrial).

La cuarta revolución es conocida como la era del petróleo, el automóvil y la producción en masa. Iniciada a principios del siglo XX su evento catalizador, según la autora, es la salida del primer modelo T de la planta Ford en Detroit. Las técnicas nuevas fueron la producción en masa de automóviles, petróleo barato y sus derivados, petroquímica, motor de combustión interna para automóviles, transporte de carga, tractores, aviones, tanques de guerra y generación eléctrica y electrodomésticos. La infraestructura asociada fueron las redes de oleoductos, redes de caminos y aeropuertos. También la electricidad de plena cobertura y la telecomunicación analógica mundial.

Por último, la quinta revolución es la era de la informática y las telecomunicaciones, comenzada a fines del siglo pasado y caracterizada por su inicio con el anuncio del microprocesador Intel en California. Para la autora, la técnica novedosa fue la revolución de la informática, la microelectrónica barata, computadoras, software, telecomunicaciones y desarrollo por computadora de biotecnología. La infraestructura subyacente en este caso fue la fibra óptica, radio y satélite, internet y correos electrónicos y transporte físico de alta velocidad (por tierra, mar y aire).

El avance de estas tecnologías es primordial para entender nuestro contexto actual. Dada la creciente disponibilidad de datos a gran escala, el término big data se convirtió en un concepto omnipresente. Como se menciona en IFC (2022) el desarrollo de internet de las cosas y el impacto de la digitalización y la mayor capacidad de procesar datos, como textos o imágenes, llevaron a la creación

de nueva información. En relación con esto, Forbes (2012) también remarca que la actividad humana está dejando una huella digital que proporciona información detallada sobre actividades, ubicaciones y hábitos personales como nunca. Esto ha generado un cambio de paradigma, vivimos en la era de big data y la revolución de los macrodatos puede convertirse en un punto de inflexión en la historia económica¹. Si bien el desarrollo de big data puede verse en diferentes industrias, nuestro enfoque será en el área financiera.

En este sentido, las crisis económicas y financieras han sido un factor clave para dar cuenta de la necesidad de avanzar con el uso de big data en el ámbito de la regulación financiera. Como se menciona en el trabajo de IFC (2022) luego de la crisis de 2007-2009 se produjo una considerable evolución de estrategias de recopilación de datos. Asimismo, según el reporte de FSB-IMF (2009), en abril 2009 los ministros de finanzas y gobernadores de Bancos Centrales del G-20 crearon un grupo de trabajo de cooperación para analizar las brechas de información y proporcionar propuestas para fortalecer la recopilación de datos. Según ellos, como en otras crisis financieras, la falta de información oportuna y precisa obstaculiza la capacidad de respuestas efectivas. Mencionan: "en efecto, la reciente crisis ha reafirmado una lección antigua: los buenos datos y análisis son la columna vertebral de una vigilancia efectiva y respuestas políticas tanto a nivel nacional como internacional" (FSB-IMF, 2009, p. 4).

Como menciona Kameda (2022), en el ámbito de las finanzas, los datos granulares ganaron mayor impulso en todo el mundo gracias a la crisis financiera de fines de la década del 2000. En esa época los bancos se encontraban excesivamente apalancados en derivados, este factor exacerbó la crisis y también hizo que los canales de transmisión de shocks sean más complejos. Debido a esto los bancos centrales, organizaciones internacionales, gobiernos y entidades de regulación han cooperado y tomado medidas para fortalecer el uso de datos granulares, especialmente en las pruebas de estrés que comenzaron a implementarse desde la crisis financiera. Por ejemplo, el uso de datos granulares sobre transacciones de repos han revelado que, cuando los mercados financieros se volvieron inestables en marzo 2020, el funcionamiento del mercado de repo de los bonos japoneses cayó debido a una notable disminución del número de proveedores de bonos, principalmente por un aumento en la demanda de las garantías.

A su vez, recientemente, hemos vivido un suceso que puso de manifiesto la necesidad de profundizar el uso de big data. La pandemia provocada por el COVID-19 también fue un disparador del inicio o intensificación del uso de fuentes alternativas de información por parte de las autoridades monetarias (Kameda, 2022). Según este autor el Banco Central de Japón (BCJ) ha empezado a incorporar datos de alta frecuencia en el *Outlook Report* desde abril 2020. Ejemplos de estos datos fueron tendencias de movilidad relacionadas con el comercio minorista y la recreación, número de visitantes a restaurantes a partir de un sistema de gestión de reservas proporcionado, evolución del consumo basado en los gastos de tarjeta de crédito y el número de personas en los aeropuertos de Japón. Por otro lado, además de datos cuantitativos se utilizaron datos textuales. El BCJ también ha publicado la creación un Índice de Sentimiento de Precios basado en la encuesta *Economy Watchers* conducida por el gobierno.

El crecimiento exponencial de la producción de datos y su consecuente necesidad de extracción de valor (ya que los datos por sí mismos no generan información sino su procesamiento) trajo la necesidad del desarrollo también de nuevas tecnologías de la información que pudieran hacer frente a un *input* más diverso y mucho más complejo. La implementación de herramientas más sofisticadas que permitan desarrollar y optimizar la gestión de riesgos financieros es necesaria, sobre todo en un contexto mundial de alta volatilidad.

En este marco, el presente trabajo tiene como propósito revisar en la literatura cuál es el estado actual del uso de big data en la regulación internacional financiera. Nos proponemos conocer cuáles son

¹ También debemos remarcar que hay versiones menos optimistas acerca del alcance de los macrodatos como, por ejemplo, se puede observar en Sosa Escudero (2019).

las áreas de la regulación en las que hay un desarrollo más profundo de estas tecnologías que en otras. Por último, pretendemos identificar cuáles son los desafíos que quedan por delante para implementar esa herramienta. El trabajo se divide de la siguiente manera: en la primera sección, se expondrá una explicación general del concepto de big data, de los tipos de datos que engloba y de las distintas herramientas que se utilizan para procesarla. En la segunda sección, hablaremos sobre las ventajas de su uso en la regulación financiera. En la tercera parte, analizaremos la aplicación en la práctica de los usos actuales de big data por parte de los bancos centrales. En la cuarta sección, abordaremos los desafíos y obstáculos pendientes de superar. Por último, se expondrán las conclusiones a las que arribamos.

3. Cambios tecnológicos y surgimiento de big data

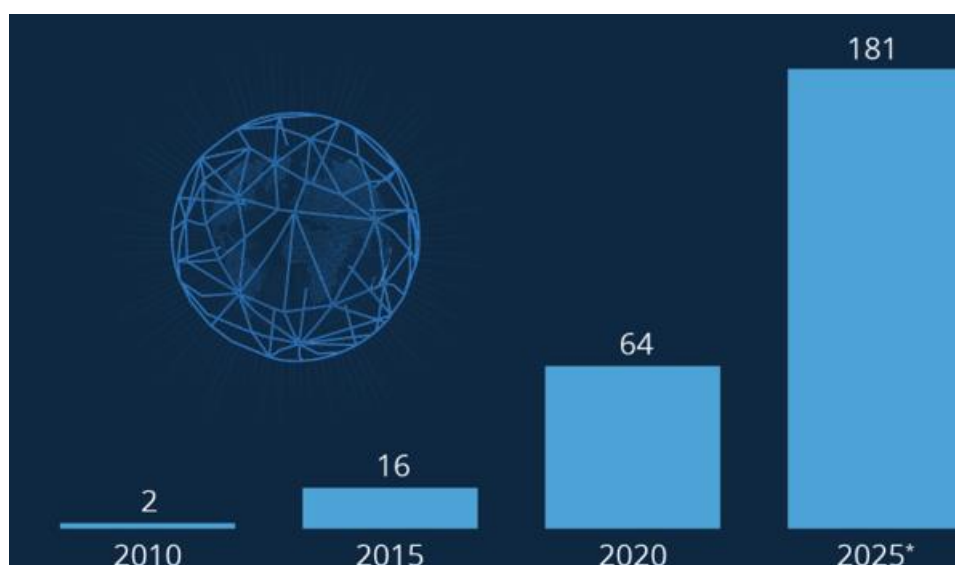
Tal como mencionamos en la introducción, antes de comenzar a desarrollar el concepto central de nuestro trabajo es necesario poder contextualizarlo. En este sentido, es importante situarnos en el momento histórico que vivenciamos traccionado de manera nunca vista por el desarrollo de las tecnologías, el avance creciente de la digitalización y la aceleración de las transformaciones que todos estos avances generan sobre nuestra sociedad. Teniendo esto presente, en los siguientes apartados de esta sección nos detendremos a explicar qué es big data, cuáles son los distintos tipos de datos de los cuales se compone y qué técnicas desarrolladas, están utilizando las entidades regulatorias financieras para procesarlo.

3.1 Definición de Big Data

Dado el rol central que tiene el concepto de big data en nuestro desarrollo, dedicaremos este apartado a desplegar las diferentes acepciones que existen sobre él. En este sentido, adelantamos que su definición no es única, sino que big data es un concepto dinámico y en movimiento. Dicho dinamismo proviene de la transformación que permanentemente le imprime la tecnología a la producción de datos. En otras palabras, lo que en un comienzo era considerado big data hoy parece ser ampliamente superado de la mano del avance de la tecnología. En consecuencia, sería más justo hablar sobre el paradigma de big data ya que este concepto nos permite amparar los desarrollos históricos que hasta el momento se han hecho de él. Para ello, expondremos las definiciones de diversos autores.

En primer lugar, nos interesa reflejar que, como comenta Mena Roa (2021) en su artículo, el crecimiento de la creación de datos ha sido significativa. En la Figura 1 podemos observar el crecimiento de la creación de datos a través de los últimos años y una previsión para el año 2025. Mientras que para el año 2010 el volumen de datos creados era de 2 zeta bytes (recordar que una zeta byte equivale a mil millones de gigabytes), para el año fueron de 15 zeta bytes, aumentando a 64 en el año 2020 estimando un nivel de 181 zeta bytes para el año 2025. En este artículo la autora también menciona que hay expertos que alertan que la creación y recopilación de datos crecerá más que su capacidad de almacenamiento mencionando que hasta 2021 solo el 2% de los nuevos datos creados se ha almacenado.

Figura 1 El Big Bang del Big Data



Nota: Adaptado de Mena Roa (2021)

La definición de big data considerada de referencia por la literatura fue acuñada por Laney (2001) y publicado por Gartner². En dicha publicación el autor visualizaba que el aumento del comercio electrónico, el incremento de fusiones y adquisiciones de firmas y el aprovechamiento de la información como recurso competitivo estaba llevando a las empresas a tener un mayor interés en la gestión de los datos en su nivel más básico. Dicha gestión podría permitir lograr una consistencia operativa, analítica y colaborativa, sin embargo, menciona el autor, las condiciones económicas cambiantes habían hecho más difícil ese camino. En particular, el comercio electrónico había aumentado los desafíos de gestión de datos en tres dimensiones: volumen, velocidad y variedad (3 V's).

Por un lado, según el autor, el volumen de datos se vio ampliado debido al desarrollo de canales electrónicos que aumentaron la disponibilidad y profundidad de datos disponibles sobre una transacción. El menor costo de los canales electrónicos permitió a las empresas además de llegar a más clientes o socios, generar hasta 10 veces más información almacenable. Conforme las firmas fueron tomando noción de la importancia de este activo intangible, comenzaron a comprar almacenamiento en línea adicional. En relación con esto, el autor destaca la importancia de una gestión eficiente en términos económicos sobre el almacenaje de los datos. Sobre este punto, nos explayaremos más adelante y será uno de los temas centrales alrededor de big data.

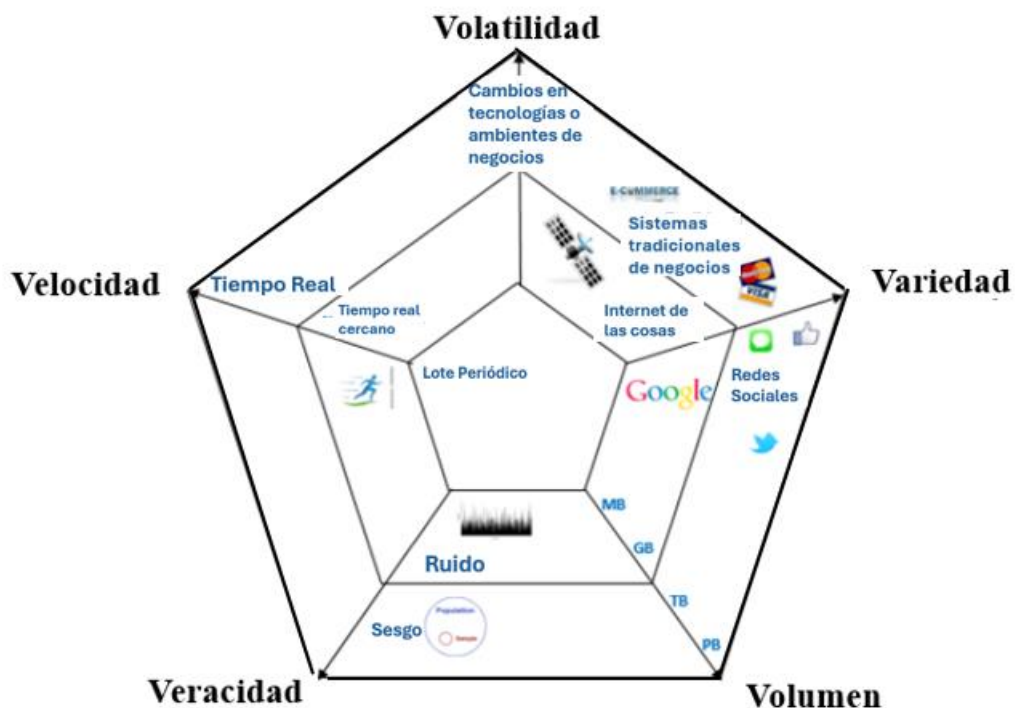
Además, Laney destaca que el comercio electrónico también aumentó la velocidad del uso de datos para respaldar cualquier tipo de interacción necesaria para sostener el comercio. Por ejemplo, la respuesta del sitio web, el análisis de disponibilidad de inventario, seguimiento de pedidos, entrega de productos, etc. Reconociendo la importancia de este elemento, las firmas están comenzando a implementar distintas soluciones a esta arista. El último elemento señalado por el autor es la variedad. Menciona que hasta el 2003 no habrá mayor barrera para la gestión de los datos que la variedad de los formatos incompatibles, estructuras de datos no alineadas y semántica de datos inconsistente. Para intentar resolver esto las firmas deberán complementar diferentes técnicas hasta que se logre consolidar alguna aplicación más general sobre dicha problemática.

² Fundada en 1979, es una empresa de investigación y consultoría reconocida a nivel mundial en el asesoramiento estratégico a empresas y la tecnología de la información.

A lo largo de los años se han ido agregando otros parámetros a la definición. Por ejemplo, en el artículo de IFC (2015) se toma la definición general de big data (es decir, las 3 V' s) agregando la dimensión de la veracidad. Esta cuarta dimensión no es más que replicar el principio estadístico de tener conocimiento sobre el conjunto de datos sobre el cual se realizarán posteriores análisis, estudios y, por lo tanto, conclusiones. La veracidad es necesaria para garantizar la transparencia en la cobertura y metodologías implementadas. Aunque los autores remarcan que, dado que la información bruta debe ser combinada y procesada para que tenga sentido, sería mejor hablar de *smart data* antes que big data.

Por su parte, en el trabajo de Hammer et al. (2017), se menciona una V adicional referida a la volatilidad presente en la tecnología y los entornos que producen los datos que podrían llevar a análisis y resultados inválidos. En la Figura 2 se pueden observar el detalle de las 5 dimensiones de big data a las que se hace referencia. En ella podemos observar que la volatilidad hace referencia a los constantes cambios en tecnologías o ambientes de negocio, la variedad hace referencia a las diversas fuentes de producción de datos como ser los sistemas tradicionales de negocios, las redes sociales, internet de las cosas. Volumen hace referencia a las distintas medidas de almacenamiento como ser megabyte, gigabyte, terabyte y petabyte. Veracidad hace referencia a la ausencia de ruidos y sesgos para que la información sea lo más fiel posible. Por último, la velocidad hace referencia a los datos que se obtienen en tiempo real, tiempo real más cercano o lotes de datos periódicos.

Figura 2 Las 5 Vs de Big Data – Volatilidad , Variedad, Volumen, Veracidad y Velocidad



Nota: adaptado de Hammer et al. (2017, p. 9)

Además, estos autores mencionan que, a diferencia de los datos estadísticos, que son recolectados ya con fines específicos, los big data, son un subproducto derivado de los sistemas administrativos y comerciales, redes sociales e internet de las cosas. Sin duda en los últimos años los macrodatos también conocidos como big data fueron tomando más relevancia dado su desarrollo y actualidad y es necesario poder abordarlos.

Para Doko y Miskovski (2019) el big data se refiere a la recopilación y análisis de grandes volúmenes de datos estructurados y no estructurados, potencialmente en tiempo real con el fin de crear valor para las empresas. De todas formas, destacan, por un lado, que no hay una sola definición de big data y, por el otro, que el concepto de big data en finanzas es distinto al contexto de otras industrias. Big data es en general definido siguiendo las siguientes propiedades llamadas “5 V’s”:

1. Volumen: se refiere a la enorme cantidad de datos creados por la conectividad en internet.
2. Variedad: los datos pueden ser estructurados y no estructurados. El ejemplo clásico de tipo de datos estructurales son los datos financieros (pagos de tarjeta de crédito, volumen comercializado en determinado mercado financiero, etc.). Mientras que los datos no estructurados se están volviendo cada vez más relevantes para descubrir nuevas conclusiones y hallazgos (periódicos, redes sociales).
3. Veracidad: refiere a la confiabilidad de los datos en big data especialmente si son obtenidos de fuentes públicas de terceros.
4. Velocidad: refiere a la velocidad de los datos entrantes provenientes de diferentes fuentes. Esta característica es necesaria para poder realizar análisis en tiempo real.
5. Valor: predecir nuevas tendencias a partir del análisis de datos permite a los bancos, instituciones financieras y empresas crear valor, comercializando nuevos servicios o anticipando dinámicas negativas de mercado.

Es decir, podemos observar cómo estos autores destacan al valor como una propiedad en lugar de la volatilidad como hacen en Hammer et al. (2017).

Según Nymand-Andersen (2016) el término big data se define como la habilidad de combinar información registrada y extraer inteligencia de múltiples fuentes. Es un término en constante movimiento y es necesario ponerlo en contexto de acuerdo con su temporalidad ya que lo que hace 10 años eran considerados grandes datos hoy en día no lo serían, sostiene el autor. Se remarca positivamente que la evolución de los servicios de datos genera una sociedad más conectada, con más cohesión, hecho que desemboca en el bienestar de la población. Además, identifica el valor económico que genera esta herramienta al proporcionar una ventaja competitiva siendo que esta información adicional permite aumentar la probabilidad de tomar decisiones más acertadas.

Como muestra de sus bondades, Kitchin (2015) menciona ciertas ventajas de big data como por ejemplo imputación de elementos de datos faltantes: sustituir por unidades similares o idénticas; edición: ayudar en la detección y tratamiento de anomalías en los datos de la encuesta; vinculación con otros datos: crear conjuntos de datos más ricos y/o perspectivas longitudinales; confrontación de datos: asegurar la validez y consistencia de los datos de la encuesta; mejora de la eficiencia operativa y efectividad de los Institutos Nacionales de Estadística mediante el uso de para data creada y capturada desde su estadística.

También big data podría reemplazar completamente fuentes estadísticas existentes como encuestas, proporcionar información estadística complementaria en el mismo dominio estadístico, pero desde otras perspectivas, mejorar las estimaciones de fuentes estadísticas (incluidas las encuestas) y proporcionar información estadística completamente nueva en un dominio estadístico particular.

Al parecer, en un momento, fue cuestión de quien ideaba más cantidad de V’s. Tal es así que en su artículo Shafer (2017) sostiene que la evolución de la definición de big data había llegado a completar las 42 V’s de big data. Algunos de los parámetros que menciona el autor son: ventaja, ya que big data proporciona una visión privilegiada de sistemas complejos; variabilidad, ya que modela fuentes de datos variables; vaticinación, debido al carácter predictivo que pueden extraerse de los datos masivos. También menciona a la visualización debido a que es la única manera que clientes interactúen con los modelos; vocabulario, ya que proporciona un vocabulario para afrontar una variedad de problemas. Como menciona Sosa Escudero (2019) “una definición que abarque 42 conceptos es cualquier cosa menos una definición” (p. 23).

Para Manyika et al. (2011) “Big data se refiere a conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas de software de bases de datos típicas para capturar, almacenar,

gestionar y analizar. Esta definición enfatiza su intencionalidad subjetiva; es decir, no se define big data en términos de ser mayor que un cierto número de terabytes. Suponemos que, a medida que la tecnología avance con el tiempo, también aumentará el tamaño de los conjuntos de datos que califican como big data. Tenga en cuenta también que la definición puede variar según el sector, dependiendo de qué tipos de herramientas de software están comúnmente disponibles y qué tamaños de conjuntos de datos son comunes en una industria en particular. Con esas salvedades, hoy en día el big data en muchos sectores oscilará entre unas pocas docenas de terabytes y varios petabyte (miles de terabytes)” (p. 6).

En resumen, uno de los rasgos en los que parece haber consenso entre los autores es en sostener que la definición de big data es una definición en movimiento, en este sentido es más bien un paradigma, y que puede variar según las distintas industrias. Otro de los rasgos en que coinciden generalmente es que la definición de big data debe incluir 3 dimensiones: velocidad, volatilidad y volumen. Luego, se desarrollan otras aristas que no son compartidas por todos como ser valor, veracidad, volatilidad, etc. Más allá de estas aristas, a juzgar por la centralidad del contenido, nosotros nos quedaremos con la definición acuñada por Sosa Escudero (2019): “big data se refiere a una copiosa cantidad de datos producidos espontáneamente por la interacción con dispositivos interconectados” (p. 23). También remarca el autor que, si bien big data comienza por el tamaño de los datos, muy rápido se va por caminos más interesantes.

3. 2 Tipos de Datos

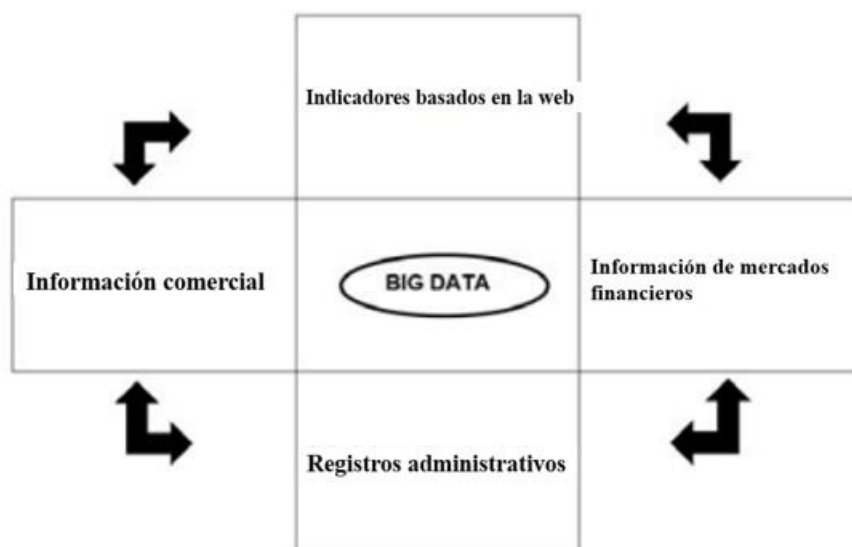
De la misma manera que comenzamos viendo la definición de big data, y muy ligada a la definición que demos de ella, en esta sección veremos cuáles son, a grandes rasgos, los distintos tipos de datos que conforman el paradigma de big data. Esto es muy importante ya que la naturaleza de los datos determinará qué tipos de técnicas podrán implementarse para procesarlos y a su vez, permitirá identificar cuáles son sus ventajas y desventajas.

Según el departamento de Asuntos Económicos y Sociales de las Naciones Unidas big data puede clasificarse en 3 grandes tipos de datos: redes sociales, sistemas comerciales tradicionales e internet de las cosas. El primer tipo refiere al registro de experiencias humanas por ejemplo Facebook, Twitter, Tumblr, blogs, Instagram, YouTube, etc. El segundo tipo refiere a eventos comerciales de interés, registro de clientes, transacciones, fabricación de productos, etc. Generalmente, este tipo de datos son altamente estructurados y están almacenados en sistemas de bases de datos relacionales. El tercer tipo de datos proviene del crecimiento fenomenal en el número de sensores y máquinas utilizadas para medir y registrar eventos. Por ejemplo, cámaras web, ubicación de teléfonos móviles, automatización del hogar, sensores de tráfico, videos de seguridad, vigilancia, etc.

De acuerdo con la interacción de bancos centrales, según encuestas realizadas y dentro su propio ámbito, el trabajo del IFC (2015) sostiene que el tercer tipo de datos debe dividirse entre sistemas administrativos y sistemas empresariales privados. Cada uno de ellos representa un problema distinto. En el cuadro 1 del Anexo se detallan una lista completa de los tipos de fuentes de datos, esto es, fuentes oficiales, base de datos internas públicas, información de internet, información de instituciones financieras, proveedores de datos, datos de posicionamiento móvil, registros de supermercados con sus respectivos ejemplos de cada uno de ellos.

Si bien los tipos de datos que se generan se pueden abrir en diversas categorías, Tissot (2019), define tal como se puede observar en la figura 3, 4 tipos de datos relacionados con los utilizados por los bancos centrales. Por un lado, se encuentra la información comercial, la información extraída de los registros administrativos, los datos extraídos de los mercados financieros y la información basada en la web.

Figura 3 Cuatro tipos de datos financieros de big data



Nota: Adaptado de Wibisono et al. (2019, p.5)

El autor menciona que, en contraste con el sector privado, la utilización de indicadores basados en la web por parte de los bancos centrales podría tener ciertas limitaciones, especialmente en lo que respecta a datos no estructurados como imágenes. No obstante, varios proyectos están en curso para aprovechar la información recopilada en Internet y respaldar la formulación de políticas monetarias y financieras. Además, es importante destacar el incremento en el acceso a información digitalizada, lo que refleja tanto el crecimiento de la disponibilidad de información textual en la web (por ejemplo, en redes sociales) como la capacidad de digitalizar, buscar y analizar documentos "tradicionales" impresos de manera similar a los indicadores basados en la web.

La mayoría de los conjuntos de datos financieros relevantes para los bancos centrales consisten en información detallada proveniente de registros extensos, cubriendo transacciones comerciales, desarrollos en los mercados financieros y operaciones administrativas. Este tipo de datos se ha expandido debido a la iniciativa de llenado de brechas de datos, respaldada por el G20, tras la crisis financiera de 2007-09. Como resultado, los bancos centrales tienen acceso a información detallada sobre el sistema financiero, a nivel de instituciones específicas, transacciones o instrumentos.

3.3 Técnicas de procesamiento de datos masivos

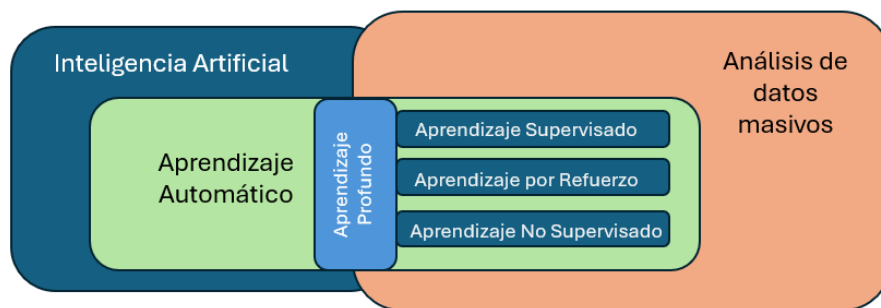
Tal como mencionamos en la introducción, a la par que destacamos la importancia del desarrollo de big data también es relevante hablar e identificar las técnicas que han surgido para procesar, transformar y generar conocimiento nuevo sobre la base de los datos masivos generados. Esto es, hacer foco en las herramientas que permiten extraer valor de los datos para tomar mejores decisiones. Por lo tanto, en esta sección daremos una visión general de las principales técnicas de procesamiento de big data que se utilizan en el ámbito de los bancos centrales y que, luego, nos permitirán comprender cuando hablemos del uso que se está haciendo en la práctica. Las técnicas que mencionaremos serán:

aprendizaje automático, aprendizaje profundo, redes neuronales, procesamiento de lenguaje natural, modelado de lenguaje grande, árboles de decisión, minería de datos y análisis de redes³.

Wibisono et al (2019) en su trabajo mencionan que se están desarrollando una serie de técnicas, a menudo referidas como "*big data analytics*" e "inteligencia artificial" (IA). El aumento en los volúmenes de datos ha ido de la mano con el avance de técnicas específicas para su análisis, gracias al campo de la "analítica de big data", que se refiere a la evaluación general de estos conjuntos de datos, y la IA, la cual se define como "la teoría y desarrollo de sistemas informáticos capaces de realizar tareas que normalmente requieren inteligencia humana" (FSB, 2017). Sin embargo, señalan, es importante mencionar que estos dos conceptos pueden tener diferencias sutiles entre sí (por ejemplo, es posible desarrollar herramientas para analizar conjuntos de datos de big data que no dependen necesariamente de técnicas de IA), tal como se ilustra en el Figura 4. En la misma se puede observar un detalle del desglose de las diferentes técnicas que se desprenden del aprendizaje automático.

Figura 4

Una visión esquemática de Inteligencia Artificial, Aprendizaje Automático y Análisis de datos masivos



Nota: Adaptado de Wibisono et al. (2019, p. 3)

La analítica de big data guarda muchas similitudes con las técnicas econométricas tradicionales, y en realidad, se basa en muchas metodologías y técnicas establecidas desarrolladas para la estadística en general, como, por ejemplo, el análisis de componentes principales, que surgió a principios del siglo pasado. No obstante, una característica distintiva es que estas técnicas se aplican a un conjunto de datos modernos que pueden ser muy extensos y complicados. La extracción de información relevante de estos conjuntos de datos no es sencilla y a menudo requiere un conjunto diferente de habilidades, según el tipo de información que esté en juego. Como resultado, tanto la analítica de big data como las técnicas de inteligencia artificial comprenden una variedad de enfoques estadísticos y de modelización, tales como el aprendizaje automático, técnicas de minería de texto, análisis de redes, modelización basada en agentes, entre otros que, desplegaremos a continuación.

3.3.1 Machine Learning o Aprendizaje Automático

Muchas de estas técnicas mencionadas previamente involucran el llamado aprendizaje automático. Según FSB (2017) éste es un subconjunto de técnicas de IA, que pueden definirse como un método para diseñar una secuencia de acciones para resolver un problema que se optimiza

³ Esta clasificación no pretende ser exhaustiva ni mucho menos, sino enunciar a modo general las técnicas que se explicarán en el desarrollo del trabajo

automáticamente a través de la experiencia y con intervención humana limitada o nula. Este enfoque es bastante similar a la econometría convencional, aunque con tres características distintas. Primero, el aprendizaje automático generalmente se enfoca en la predicción en lugar de identificar una relación causal (o, como menciona McKinsey & Company (2024 a), en lugar de recibir instrucciones explícitas). Segundo, el objetivo es elegir un algoritmo que se ajuste a los datos observados reales, en lugar de a un modelo teórico. Tercero, y vinculado al punto anterior, las técnicas se seleccionan observando su ajuste a los datos, y menos a las pruebas estadísticas más tradicionales utilizadas en econometría. Según los autores, dentro de las técnicas del aprendizaje automático podemos encontrar tres categorías.

3.3.1.1 Aprendizaje Automático Supervisado

Una de ellas es el aprendizaje automático supervisado. En él se proporciona un conjunto de datos de "entrenamiento" a un algoritmo, el cual contiene etiquetas en las observaciones. El propósito es categorizar puntos de datos individuales, determinando a qué clase (o categoría de observaciones) pertenece una nueva observación. Esto se deduce mediante el análisis de una muestra de observaciones previas, es decir, el conjunto de datos de entrenamiento, para el cual se conoce su categoría. El objetivo del algoritmo es predecir la categoría de una nueva observación en función de sus características. Por ejemplo, predecir si se aprobará un nuevo préstamo ("sí" o "no"), según sus características y comparándolas con un conjunto de datos históricos de préstamos aprobados o rechazados; o si una empresa es probable que incumpla en un futuro cercano.

3.3.1.2 Aprendizaje Automático no Supervisado

La segunda categoría se refiere al aprendizaje automático no supervisado, donde "los datos suministrados al algoritmo no tienen etiquetas". Esto implica que no se han definido categorías de antemano para un conjunto específico de observaciones, por lo que el algoritmo debe identificar los grupos, agrupando las observaciones que comparten características o "patrones" similares. Dos ejemplos sobresalientes de este enfoque son los algoritmos de agrupamiento y reducción de dimensionalidad. En el agrupamiento, el objetivo es identificar los grupos inherentes presentes en el conjunto de datos granulares, como, por ejemplo, identificar grupos de clientes o empresas con características similares, colocando las observaciones más similares en el mismo grupo de manera progresiva (enfoque ascendente). La reducción de dimensionalidad implica reorganizar la información original en un menor número de dimensiones, de manera descendente; el propósito es reducir el número de variables independientes de forma significativa, sin perder demasiada información.

3.3.1.3 Aprendizaje Automático por Refuerzo

Los autores también sostienen que hay otros tipos de algoritmos disponibles. Uno de ellos es el aprendizaje por refuerzo, que enriquece el aprendizaje no supervisado al recibir información adicional de retroalimentación, como la intervención humana. En este enfoque, el agente interactúa con su entorno tomando decisiones y recibiendo una evaluación (o recompensa) según esas decisiones. Luego, adapta su comportamiento para optimizar la recompensa esperada a lo largo del tiempo. Mediante este ciclo de acción, retroalimentación y ajuste, el agente desarrolla una estrategia óptima, permitiéndole tomar las mejores decisiones posibles en ese contexto específico.

3.3.1.4 Aprendizaje Profundo o Deep Learning

Dentro del aprendizaje automático también existen otras técnicas. McKinsey & Company (2024) destacan que el aprendizaje profundo es una versión más avanzada del aprendizaje automático y particularmente hábil en el procesamiento de una gama más amplia de recursos de datos (textos y datos no estructurados, por ejemplo, imágenes). Sostienen que esta técnica requiere de menos intervención humana y, a menudo, puede producir resultados más precisos que el aprendizaje automático tradicional. Utiliza redes neuronales (similares a las del ser humano) para ingerir y procesar datos a través de múltiples capas que reconocen características cada vez más complejas de los datos. Dentro de las redes neuronales utilizadas mencionan a 3 de ellas: red neuronal de retroalimentación, convolucional y red neuronal recurrente.

3.3.1.5 Minería de Textos

Wibisono et al. (2019) remarcan otra área en constante evolución dentro del campo de la analítica de big data es la minería de texto, que implica examinar información semántica mediante la automatización del análisis de grandes volúmenes de texto en lenguaje natural. El propósito es identificar patrones lingüísticos o léxicos con el fin de obtener conocimientos valiosos. Aunque la mayoría de la investigación económica se enfoca en indicadores numéricos, como precios o datos de ventas, se está generando una cantidad cada vez mayor de información textual relacionada con actividades económicas y financieras, tanto en internet (como publicaciones en redes sociales) como en una variedad de fuentes como informes financieros corporativos, artículos de prensa y debates de autoridades. El análisis de este tipo de información sin estructurar se ha vuelto esencial para los encargados de formular políticas, especialmente dada la importancia de indicadores "blandos" como la confianza y las expectativas durante la crisis financiera global.

Además de los algoritmos numéricos, los autores sostienen también que se pueden emplear métodos más simples que se basan en diccionarios para analizar datos de texto. Es posible elegir un conjunto de términos clave relevantes para el tema en cuestión, como por ejemplo una palabra relacionada con "confianza empresarial". Luego, se puede crear un indicador basado en la frecuencia con la que estas palabras clave seleccionadas aparecen en un documento determinado, lo que permite evaluar el aspecto temático (por ejemplo, la evolución del sentimiento empresarial). Un ejemplo notable es el Índice de Incertidumbre de Política Económica (IIPE), que mide el nivel de incertidumbre según la presencia de un grupo de términos clave relacionados con la economía, la política y la incertidumbre en los artículos de noticias.

3.3.1.6 Análisis de Redes

Un área importante de la analítica de big data se refiere al análisis de redes financieras, que puede considerarse como el análisis de las relaciones entre los elementos que constituyen el sistema financiero. Las percepciones sobre el funcionamiento de esta "red" se derivan de técnicas y representaciones gráficas. Este enfoque puede medir cómo los datos están conectados con otros datos, aclarar cómo estas conexiones importan y mostrar cómo los sistemas complejos se mueven con el tiempo. Esta técnica puede ser particularmente efectiva para conjuntos de datos de big data, permitiendo la descripción de sistemas complejos caracterizados por interacciones ricas entre sus componentes.

Este tipo de análisis resulta particularmente idóneo para representar la interrelación dentro de un sistema, como, por ejemplo, trazar la cadena de valor global entre países y sectores o los diferentes tipos de exposición asumidos por las instituciones financieras. Un ejemplo de esto es el trabajo reciente que busca evaluar la función de contrapartes centrales en el sistema financiero mediante la exploración

de las conexiones entre ellas y otras entidades financieras como grupos bancarios, prestando especial atención a las relaciones entre subsidiarias y matrices. Este enfoque puede ayudar a visualizar cómo una disrupción originada en una contraparte central individual afectaría a sus miembros de compensación y, a su vez, a otras contrapartes centrales.

Por otro lado, el trabajo de IFC (2024) proporciona una visión general de los métodos de aprendizaje automático. Por un lado, explican que el aprendizaje automático comprende técnicas diseñadas para extraer información de los datos con el objetivo principal de hacer predicciones. Además, se destaca que el aprendizaje automático es una evolución de técnicas estadísticas y econométricas tradicionales, pero no se basa en modelos preespecificados ni en supuestos estadísticos como la linealidad o la normalidad. De esta manera, el éxito en el entrenamiento de un modelo de aprendizaje automático se evalúa por su capacidad para predecir resultados en datos no vistos previamente, independientemente de cómo los modelos realicen esas predicciones. Las siguientes secciones describen algunas de las técnicas más comunes utilizadas en los bancos centrales, basadas en los ejercicios regulares de inventario organizados en la comunidad bancaria central bajo el paraguas del Comité Irving Fisher del Banco de Pagos Internacionales.

3.3.1.7 Árboles de Decisión

En primer lugar, se explican los métodos basados en árboles que son algoritmos versátiles en el ámbito del aprendizaje automático, capaces de segmentar los datos de manera secuencial en categorías más detalladas, lo que les otorga la capacidad de abordar diversas tareas. Estos árboles de decisión organizan los datos individuales en grupos basados en características específicas de interés, permitiendo un análisis jerárquico de cómo estas características se relacionan con un atributo determinado. Utilizando un ejemplo concreto, se muestra cómo un árbol de decisión puede dividir las casas según el número de habitaciones y el año de construcción para examinar cómo afectan estos factores al precio de las viviendas.

3.3.1.8 Bosques Aleatorios

El segundo lugar, los bosques aleatorios constituyen una técnica que fusiona múltiples árboles de decisión, cada uno entrenado con diferentes segmentos de los mismos datos. Este método busca mejorar la precisión de las predicciones en datos no vistos y reducir el riesgo de ajuste excesivo a los datos utilizados para el entrenamiento. Los autores resaltan que los bosques aleatorios, junto con modelos similares, proporcionan una manera más flexible de realizar análisis de regresión, al pronosticar los resultados a partir de las variables explicativas clave. Así mismo, se menciona que los enfoques basados en árboles, como los bosques aleatorios, pueden servir como herramientas exploratorias para identificar patrones en los datos sin imponer una estructura de modelo específica. Se ilustra cómo estos métodos pueden agrupar puntos de datos en categorías similares y también pueden ser efectivos para detectar valores atípicos mediante el uso de bosques de aislamiento.

3.3.1.9 Redes Neuronales

En tercer lugar, los autores resaltan a las redes neuronales como una herramienta esencial en el campo del aprendizaje automático, siendo ampliamente utilizadas incluso en los modelos más recientes. Estas redes están formadas por unidades de procesamiento llamadas neuronas artificiales, las cuales realizan transformaciones no lineales en múltiples valores de entrada para producir una sola salida, similar a lo que hacen las regresiones logísticas. Estas neuronas se agrupan en capas que pueden ser

apiladas, donde cada capa transforma la salida de la capa previa. Esta estructura se asemeja a la transmisión de señales eléctricas entre las neuronas en el cerebro humano. La profundidad de una red se determina por el número de capas, y los parámetros de las neuronas regulan la fuerza de las conexiones entre ellas. Durante el proceso de entrenamiento, estos parámetros se ajustan repetidamente para mejorar la precisión de las predicciones. Esta técnica es responsable de importantes avances en inteligencia artificial, como el reconocimiento facial y los asistentes de voz en dispositivos móviles.

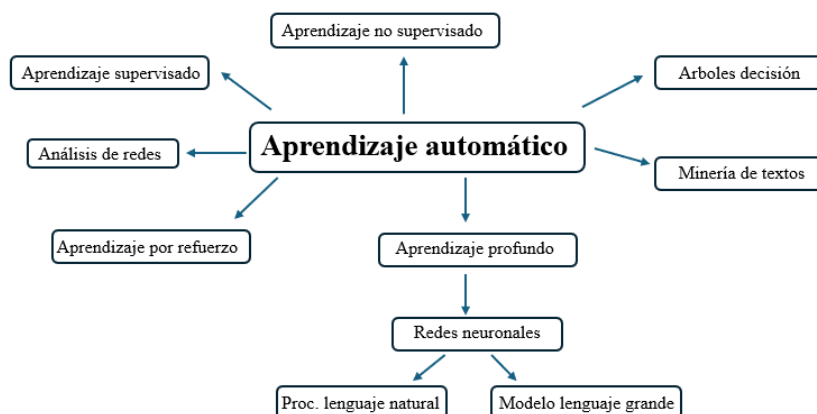
3.3.1.10 Procesamiento Lenguaje Natural y Modelos de Lenguaje Grandes

En cuarto lugar, el trabajo de IFC (2024) menciona a los transformadores que, presentados en 2017, han tenido un impacto significativo en el rendimiento de las redes neuronales en el procesamiento del lenguaje natural (PLN) y han facilitado el desarrollo de modelos de lenguaje más grandes (MLG). A diferencia de enfoques anteriores, los transformadores no se limitan a relacionar palabras cercanas, sino que buscan capturar las relaciones entre diferentes componentes de una secuencia de texto, incluso si están separados en la oración. Esto permite una mejor comprensión del contexto y de los diversos significados que una palabra puede tener según su contexto en la oración. Por ejemplo, la palabra "banco" puede tener diferentes significados en frases como "nadaré cruzando el río para llegar al otro banco" y "crucé la calle para ir al banco". Los transformadores han habilitado nuevas aplicaciones en PLN que implican textos extensos y han impulsado los avances más recientes en MLG, como ChatGPT.

Por último, los autores destacan que los MLG han desempeñado un papel fundamental en el rápido avance de la inteligencia artificial generativa, conocida como "gen IA", la cual es capaz de producir contenido basado en instrucciones específicas y llevar a cabo tareas más allá del simple reconocimiento de lenguaje. Estos modelos, que son redes neuronales entrenadas para predecir la siguiente palabra en una secuencia de texto, adquieren una amplia base de conocimientos durante su proceso de entrenamiento. Por consiguiente, tienen la capacidad de hacer predicciones precisas incluso en textos que requieren matices o conocimientos especializados. Además, los MLG pueden ser adaptados para tareas específicas utilizando datos particulares, como en el caso de ChatGPT, que se beneficia de un MLG ajustado mediante retroalimentación humana para generar respuestas más útiles. El gen IA se distingue por ser accesible para una amplia variedad de usuarios y por su habilidad para extraer ideas de datos sin una estructura predefinida.

En resumen, el siguiente cuadro muestra lo abordado previamente, de la técnica del aprendizaje automático se desprenden diversas herramientas, tales como: aprendizaje no supervisado, aprendizaje supervisado, análisis de redes, aprendizaje por refuerzo, árboles de decisión, minería de textos y aprendizaje profundo. De este último, se desprenden las redes neuronales, las cuales pueden dividirse entre procesamiento de lenguaje natural y modelos de lenguaje grande.

Figura 5 Resumen de técnicas de Aprendizaje automático



Nota: elaboración propia en base a FSB (2017) y IFC (2024)

3.4 ¿Qué es big data para los bancos centrales?

Como dijimos previamente, el significado de big data no es el mismo para una industria que para otra. De acuerdo con el trabajo de IFC (2022) en más del 90% de los bancos centrales consultados, identifican como fuente de big data al conjunto de los datos no estructurados, como mensajes de texto, redes sociales, imágenes extraídas de internet. Esta fuente requiere herramientas específicas para ser limpiada y preparada adecuadamente y, en muchos casos, es necesario adquirir estos datos de proveedores privados. Tres ejemplos de esto pueden ser, en primer lugar, informes de movilidad (tendencias obtenidas a través de GPS). En segundo lugar, se relaciona con las búsquedas de internet, como Google Trends, que pueden ser extraídos para determinar ciertos factores económicos como, por ejemplo, expectativas sobre el mercado laboral. Una tercera fuente de big data para los bancos centrales es el texto en formato impreso, como artículos periodísticos, estados financieros de empresas, comunicados de prensa, etc.

Una segunda categoría refiere al conjunto de datos con un número grande de observaciones de series temporales. Esto incluye típicamente grandes conjuntos de datos financieros estructurados. Hay otros tipos de datos completamente nuevos que fueron recopilados en el contexto de las reformas regulatorias posteriores a la crisis financiera global: son las transacciones individuales de derivados que ahora se informan a los Repositorios de Transacciones y tienen un gran volumen y alta frecuencia de actualización.

Además, el informe refiere a otras dos categorías de datos estructurados mencionadas por los bancos centrales son, por un lado, las transacciones de pago. Aunque si bien no son nuevos, los bancos antes no las consideraban dentro de las estadísticas analizadas tradicionales. Remarcan que en los últimos años han intentado utilizarlos para llevar a cabo análisis. En segundo lugar, se refieren a conjuntos de datos transversales, que proporcionan observaciones para toda la población de interés, recopiladas con múltiples atributos, en un punto del tiempo. Un ejemplo de esto son los registros de créditos que, si bien no son nuevos, las nuevas técnicas informáticas disponibles están facilitando su análisis.

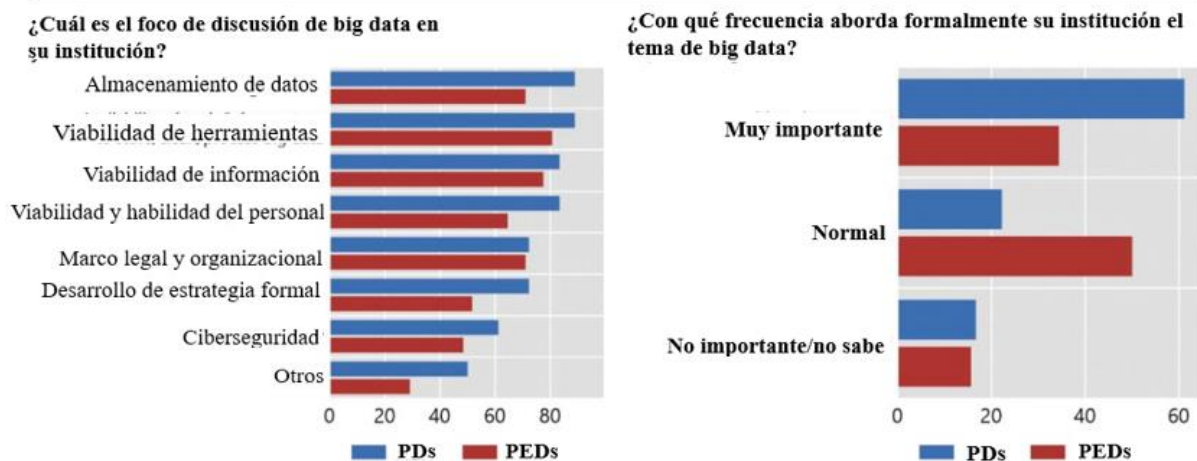
En la figura 6 se pueden observar los resultados de la encuesta actualizada en 2020 por el IFC. Por un lado, se visualiza cómo en los Países Desarrollados (PD's) la frecuencia con la que los bancos centrales abordan big data es muy importante para el doble, respecto de los Países en Desarrollo (PED's). Por otro lado, se observa que excepto en el foco de discusión sobre el marco legal y organizacional y en viabilidad de la información, en el resto de los focos, los PD's superan en porcentaje

a los PED's en todos los ámbitos (almacenamiento de datos, viabilidad de herramientas, habilidad del personal, desarrollo de estrategia formal, ciberseguridad y otros focos no especificados).

Figura 6

Discusiones de big data a través de países desarrollados y países en desarrollo

En porcentaje de respuestas



Nota: Adaptado de IFC. (2022, p. 8)

Según IFC (2022) una encuesta realizada en 2020 de actualización de una realizada en 2015 destaca que los bancos centrales tienen una visión integrada del big data que puede comprender desde los tipos de datos no estructurados que, a menudo se caracterizan por un gran volumen, velocidad y variedad y que deben procesarse utilizando tecnologías innovadoras. Para dos tercios de los encuestados, big data también incluye grandes datos tradicionales (es decir, bien estructurados) que son subproductos de operaciones comerciales (transacciones de pago), financieras (cotizaciones de precios en el mercado) y administrativas (archivos recopilados por instituciones públicas). Además, de la encuesta surgió que alrededor del 80% de los que respondieron utilizan big data con regularidad (ciertamente, la mitad lo hace solo con fines exploratorios), en comparación con solo un tercio, como sucedía en la encuesta del 2015. A su vez, que el interés político en el uso de esta herramienta aumentó a más del 60% de los encuestados, en comparación con el 10% de 2015.

Habiendo recorrido, a modo de introducción, por las diferentes nociones de qué es big data y cuáles son las técnicas puntuales utilizadas por los bancos centrales y otras entidades reguladoras, en el siguiente capítulo expondremos las principales razones por las cuales es importante el uso de big data en la regulación financiera internacional.

4. ¿Por qué es importante el uso de Big Data en la regulación financiera internacional?

Como mencionamos previamente las crisis financieras son un suceso clave donde queda de manifiesto la necesidad de contar con información precisa y en tiempo real por parte de los reguladores. En este sentido, el informe de FSB-FMI (2009) posterior a la crisis de 2008 es muy claro. La integración de economías y mercados, como lo evidencia la propagación de la crisis financiera a nivel mundial, resalta la importancia crítica de estadísticas relevantes que sean oportunas y coherentes internamente, así como comparables entre países. Además, el informe resalta lo siguiente:

Si bien la crisis financiera no fue el resultado de una falta de estadísticas económicas y financieras adecuadas, expuso una falta significativa de información, así como lagunas en los datos sobre vulnerabilidades clave del sector financiero relevantes para el análisis de la estabilidad financiera. Algunas de estas brechas afectaron la dinámica de la crisis, ya que los mercados y los encargados de formular políticas fueron sorprendidos desprevenidos por eventos en áreas poco cubiertas por fuentes de información existentes, como aquellas derivadas de exposiciones tomadas a través de instrumentos complejos y entidades fuera de balance, y de las conexiones transfronterizas de las instituciones financieras. En general, hay una necesidad de abordar las brechas de información en tres áreas principales que están interrelacionadas. (FSB-FMI, 2009, p. 9).

En el trabajo se destaca la necesidad de mejorar los datos de las conexiones de las redes financieras internacionales. Esto es, por un lado, mejorar la información sobre los vínculos financieros de entidades globales de importancia sistémica e instituciones financieras. Y, por el otro, fortalecer las iniciativas de recopilación de datos sobre flujos bancarios transfronterizos, inversiones y exposiciones, en particular, para identificar actividades de entidades financieras no bancarias.

Por último, para monitorear la vulnerabilidad de las economías nacionales a los shocks, remarcan la necesidad de promover estadísticas comparables entre países y trabajar para difundir datos más equiparables sobre precios de bienes raíces. Además, la crisis ha demostrado la necesidad de mejorar la comunicación de las estadísticas oficiales y promover la interacción entre las comunidades académicas, políticas y estadísticas. Si bien reconocen que reducir la brecha de datos puede resultar costoso, con el desarrollo continuo de los mercados es inevitable que se produzcan, por lo tanto, se incurren en costos significativos cuando la falta de información oportuna y precisa obstaculiza la intervención de los formuladores de políticas.

A su vez, debido a la fluctuación económica producida durante la pandemia del Covid-19 en una ventana de tiempo muy corta, de manera inédita, se puso de manifiesto la necesidad de poder comprender las condiciones económicas en los momentos oportunos. Tal como se evidenciaba, en ese momento las encuestas y fuentes de información sobre las que se basaban los banqueros no eran tan ágiles y eso entorpecía la toma de decisiones. El crecimiento del uso de datos de alta frecuencia y de mayor volumen se hizo cada vez más necesario ya que éstos podrían actualizarse diaria, semanal o mensualmente. Dado este contexto, las empresas han acelerado sus esfuerzos para ofrecer servicios de información digital. A su vez, se han desarrollado nuevas tecnologías que permiten el análisis de datos técnicos como el mayor uso de inteligencia artificial, *machine learning* y procesamiento de lenguaje natural (Kohei et al., 2021)

En la literatura en general hay cierto consenso en que el uso de big data trae consecuencias positivas para las tareas de los bancos centrales. Un documento de trabajo del Comité Irving Fisher en 2015 arribó a la conclusión que big data era percibida como una herramienta potente para el análisis de la estabilidad macroeconómica y financiera. La mayoría de los bancos centrales esperaban un mayor uso de big data para el área de pronósticos económicos (indicadores de inflación, desempleo, PBI, producción industrial, sector externo), ciclo de negocios (por ejemplo, indicadores de sentimiento) y análisis de estabilidad financiera (construcción de indicadores de riesgos, identificación de riesgo de crédito y mercado, monitoreo de flujos de capital, tareas de supervisión).

Por su parte, Hammer et al., (2017) mencionan que big data puede beneficiar directa o indirectamente a las estadísticas macroeconómicas y financieras y con eso, a la formulación de políticas a través de tres características:

- Producción de nuevos indicadores y respuesta a nuevas preguntas: big data rompe con el método tradicional de buscar causalidad. Más bien consiste en la búsqueda de patrones y correlaciones que, más que brindar un por qué de la situación, es alertarnos sobre lo que está ocurriendo. A su vez, big data puede permitir que el análisis de política vaya más allá de los agregados y examinar la verdadera causa subyacente para formular mejores políticas. Adicionalmente, big data podría permitir la vigilancia en países de bajos ingresos donde suelen haber problemas con la disponibilidad de datos.
- Reducción de los retrasos temporales de las estadísticas oficiales y apoyo a la predicción de indicadores existentes: sobre este punto los autores mencionan que una de las principales promesas de big data es obtener conocimientos de manera más rápida a través de la observación de variables en tiempo real. De esta manera, es posible extraer señales y poder realizar pronósticos en tiempo real antes de la publicación de los datos oficiales.
- Big data como fuente de datos e innovación en la producción oficial de estadísticas: las discusiones a nivel mundial sobre cómo deberían evolucionar las estadísticas oficiales en la era de big data demuestran que las agencias oficiales están comenzando a experimentar cambios sustanciales. Los compiladores están probando proyectos de big data para perfeccionar y complementar las fuentes de datos tradicionales. De esta manera, es posible pensar en reemplazos parciales o totales de ciertas estadísticas existentes. Por último, los autores sostienen que el big data podría ser una alternativa de bajo costo y alta calidad como fuente de datos.

Del mismo modo, el discurso de Benoît Cœuré, miembro del Comité Ejecutivo del BCE, en la conferencia sobre Regulación Económica y Financiera en la era del Big Data en 2017, hizo referencia a que el conjunto de datos estadísticos tradicionales resultó en ocasiones inadecuados para la toma rápida de decisiones durante la crisis económica vivida en el 2008. A partir de eso, los bancos centrales comenzaron a recopilar grandes cantidades de datos para ayudarlos a monitorear los mercados financieros y calibrar las políticas monetarias. Como muestra de eso menciona dos iniciativas: el proyecto AnaCredit y la recopilación de datos estadísticos del mercado monetario. El primero se refiere a la información de préstamos, concedidos por los bancos de la zona del euro a empresas. El segundo, a la información diaria sobre los préstamos en euro en el mercado monetario que, representan entre el 80% y 85% del balance total de los bancos de la zona euro. El autor menciona que ambas iniciativas proporcionan un marco de información muy valiosa para la toma de decisiones, y devela la necesidad de desarrollar distintas herramientas para poder procesar toda la información recolectada.

Por su parte, Nymand-Andersen (2016) sostiene que, si bien los bancos centrales no tienen por qué estar a la vanguardia del uso de big data, sí es una oportunidad para extraer señales económicas casi en tiempo real, mejorar sus previsiones económicas y poder medir de manera más precisa el impacto de sus políticas. El autor remarca la necesidad de desarrollar un marco conceptual y una hoja de ruta utilizando pruebas piloto. Dada las deficiencias observadas durante la última crisis financiera, big data podría proporcionar indicadores de alerta temprana para ayudar a identificar puntos del ciclo económico. De hecho, menciona que en la actualidad se están desarrollando técnicas de minería de textos, por ejemplo, para construir índices de sentimiento extraídos de las redes sociales y artículos de internet que, mediante la aplicación de algoritmos estadísticos adecuados, proporcionan información útil sobre el sentimiento del consumidor, la incertidumbre del mercado o la evaluación sistémica de los riesgos. Además, en relación con el potencial del *machine learning* sobre metodologías tradicionales mencionan lo siguiente:

Normalmente, estadísticos, econométricos e investigadores por igual utilizan técnicas bien establecidas (principalmente basadas en análisis lineales) para diversos propósitos, tales como: (i) generar estadísticas y métricas resumidas; (ii) realizar ahora pronósticos de indicadores económicos; (iii) estimar datos faltantes; y (iv) realizar pruebas de hipótesis. Sin embargo, las técnicas de aprendizaje automático pueden ser empleadas para descubrir patrones nuevos e interesantes en grandes conjuntos de datos, visualizar dichos conjuntos de datos, proporcionar estadísticas resumidas y predicciones, e incluso generar nuevas hipótesis y teorías derivadas de los nuevos patrones observados. Esto se debe a que estas nuevas herramientas pueden ayudar a identificar relaciones no lineales en los conjuntos de datos. (Nymand-Andersen, 2016, p. 7).

Y van más allá del aspecto regulatorio ya que mencionan que los macrodatos pueden tener un efecto de mejora del bienestar de la sociedad en general. Este marco crea una oportunidad para ajustar la teoría basada en modelos y reconocer la fragilidad de estos basados en suposiciones. El uso de big data puede facilitar el acceso al conocimiento procedente de los circuitos de retroalimentación entre los sectores económicos y las previsiones monetarias, la implementación de la política monetaria y los efectos de mercado. Los macrodatos también pueden proporcionar nuevas oportunidades para comprender el comportamiento involucrado en toma de decisiones y mapear vínculos de red entre grupos financieros específicos. Ellos presentan una oportunidad para extraer señales detalladas adicionales, contribuyendo a más opiniones matizadas y variadas, fomentando así una formulación de políticas sólida.

De manera adicional, el trabajo de Dicuonzo et al. (2019) sostiene que big data es relevante especialmente en las instituciones financieras en las que la gestión del riesgo es fundamental para su supervivencia y su éxito, asumiendo un papel estratégico. En su artículo investigan cómo los bancos pequeños enfrentan desafíos tecnológicos, mostrando el estado del arte sobre el uso real de las técnicas de recopilación y gestión de datos (por ejemplo, análisis de big data) para respaldar el proceso de gestión del riesgo. Además, el trabajo intenta identificar las habilidades requeridas del gestor de riesgos en la era digital. El documento contribuye al debate en curso sobre la utilidad y el uso de las innovaciones digitales en el sector bancario, discutiendo las perspectivas futuras de la gestión de riesgos 4.0 y el papel del análisis de big data en la gestión de riesgos.

Además, destacan que diversos investigadores mencionan que las tecnologías relacionadas con big data son aplicables en muchas áreas del sector bancario, incluido el comercio minorista (cobranzas bancarias, tarjetas de crédito, banca privada), comercial (análisis de riesgo de crédito, gestión de clientes y ventas, media préstamos de mercado), mercados de capitales (negociación y ventas, productos estructurados financieros) y gestión de activos (gestión patrimonial, gestión de inversiones de capital, informes globales de activos, análisis de depósitos de inversión). Sin embargo, la mayoría de los estudios muestran que las áreas de mayor uso de big data en el sector bancario se pueden agrupar en gestión de relaciones con clientes, detección y prevención de fraude, y gestión de riesgos y banca de inversión.

Por su parte, Doko y Miskovski (2019) coinciden en que big data y la ciencia de datos en la banca permite obtener más conocimiento para tomar decisiones de manera informada o para predecir riesgos e ingresos. Es decir, generar conocimiento a partir de datos existentes. Los autores listan una serie de ámbitos donde el uso de estas tecnologías tuvo mayor impacto en los bancos:

1. *Risk management*: big data puede aplicarse para mejorar el dominio de riesgos bancarios como el de crédito, riesgo de liquidez, riesgo operativo y riesgo de mercado.
2. Información y análisis del cliente: la generación de grandes volúmenes de datos generados a través de múltiples canales de servicio puede permitir descubrir patrones de comportamiento de clientes y aumentar el conocimiento de los consumidores.
3. Mejora en la detección del fraude: el análisis en tiempo real y la correlación de datos de múltiples fuentes permite determinar el fraude de manera más eficiente.
4. Previsión del mercado de valores: la combinación de varios conjuntos de datos de diferentes mercados y geografías puede generar una visión de mercado que genere señales de *trades*, de ganancias y de exposición al riesgo.
5. Análisis predictivo: los patrones encontrados en los datos pueden predecir un evento futuro a través de la comprensión de las redes sociales, tendencias en las noticias, predicción de precios y movimientos del mercado.
6. Segmentación del cliente: dentro de las estrategias de marketing, las tecnologías de big data permiten una clasificación más rápida y precisa de los clientes según segmentos donde compartan determinada característica o atributo.

Según Wibisono et al. (2019) las técnicas de análisis de big data como los árboles de decisión pueden arrojar información relevante sobre cómo los agentes económicos toman decisiones, por ejemplo, cómo se comportan los inversionistas en los mercados financieros. Otro ejemplo es el uso de indicadores de incertidumbre económica derivados de artículos de noticias para explicar cambios en los

indicadores macroeconómicos. Esto demuestra cómo el análisis de big data tiene el potencial de ofrecer percepciones no solo sobre eventos pasados, sino también sobre lo que podría ocurrir en el futuro y por qué.

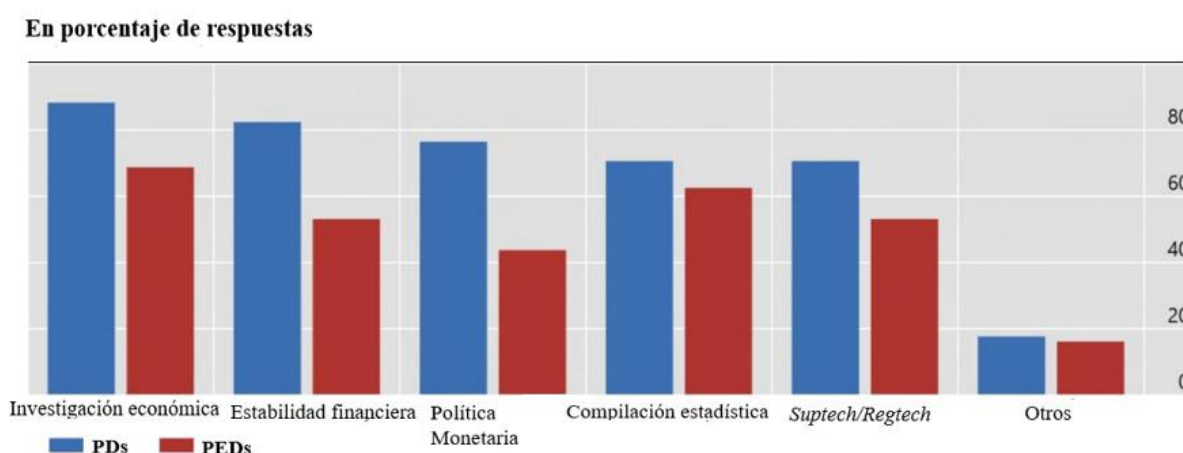
Estos nuevos conocimientos también pueden respaldar eficazmente las políticas de los bancos centrales en una variedad de áreas, como análisis de riesgo crediticio, pronósticos económicos, evaluaciones de estabilidad financiera y comunicación externa. Es notable que este enfoque puede ser altamente detallado, permitiendo una focalización precisa en mercados específicos, instituciones, instrumentos y ubicaciones, lo que resulta especialmente útil para respaldar políticas macroprudenciales. Además, los indicadores derivados del análisis de big data suelen estar disponibles de manera más oportuna que las estadísticas tradicionales; por ejemplo, los indicadores laborales pueden obtenerse a partir de anuncios de empleo en línea casi en tiempo real.

Como puede observarse, en la literatura revisada hay cierto consenso acerca del efecto positivo del uso de big data en la regulación financiera internacional. Los atributos reconocidos a los datos masivos para incrementar el conocimiento y la acción por parte de los reguladores, alienta a su uso, estudio y mayor implementación. En el siguiente apartado nos adentraremos en el uso práctico propiamente dicho de los datos masivos en la regulación financiera. Abordaremos su implementación en las diferentes áreas en las que tienen competencia las entidades reguladoras: estabilidad financiera, gestión de riesgos, investigación y predicción y supervisión y detección de fraude.

5. Aplicación de Big Data en la Regulación Financiera

A lo largo de la bibliografía consultada es visible que las áreas en las que se utiliza big data por parte de los bancos centrales son diversas. La Figura 6 se basa en el trabajo de IFC (2022) y explica el destino del uso de big data según el porcentaje de instituciones (divididas en PD's y PED's) que respondieron a la encuesta. En ella se puede observar cómo el mayor uso de big data está destinado a la investigación económica. El 80% de los PD's lo hace para respaldar la estabilidad financiera y la política monetaria. Por su parte, el trabajo micro financiero relacionado con supotech y regtech, así como la compilación estadística es utilizado por el 70% de los bancos centrales. En todas las áreas es notable la diferencia entre el uso de big data por parte de los PD's y los PED's.

Figura 6 ¿Con qué propósito usa big data tu institución?



Nota. Adaptado de (IFC, 2022, p.10)

En cuanto al uso de aplicaciones de big data por parte de los bancos se refiere a 4 tipos diferentes: el primer tipo se refiere al procesamiento de información textual a través del procesamiento del lenguaje natural. El objetivo es recopilar información cualitativa basada en texto y resumirla de manera cuantitativa para evaluar el grado de incertidumbre enfrentado por los agentes económicos. Estos índices se construyen básicamente configurando diccionarios que contengan la definición de términos relacionados con la incertidumbre y luego buscándolo en el texto considerado (por ejemplo, un artículo periódico). Luego, estos términos se cuentan y agregan para proporcionar un índice sintético.

Otro ejemplo es la evaluación de políticas. Por ejemplo, es posible cuantificar la postura de política monetaria comunicada al público. De manera similar, se pueden evaluar las expectativas de mercado sobre las decisiones de tasa de interés mediante el análisis de comentarios del mercado. El aspecto positivo de esto es que este análisis se puede actualizar frecuentemente, a diferencia de las encuestas tradicionales. Además, en mercados poco líquidos o desarrollados la información recopilada puede ser particularmente útil.

El segundo tipo se relaciona con ejercicios de pronóstico inmediato para producir análisis de la situación económica en tiempo real. Hasta un tercio de los encuestados respondió que utiliza big data para esto, particularmente estimaciones de inflación y crecimiento económico más temprano y/o con mayor frecuencia. Su horizonte temporal esencialmente, desde la compilación de índices semanales hasta estimaciones anticipadas de próximos lanzamientos como IPC o índice de producción y pronósticos de crecimiento del PBI. También estos modelos pueden ayudar a cubrir brechas estadísticas, cuando las series de referencia no existen o se interrumpen como sucedió en la pandemia. Para el caso, el mercado inmobiliario es un modelo ya que a menudo faltan los datos en las estadísticas oficiales, mientras que en internet los precios se pueden obtener relativamente fácil.

Una tercera categoría incluye las diversas aplicaciones desarrolladas por los bancos centrales para extraer perspectivas macroeconómicas a partir de conjuntos de datos financieros granulares, con el objetivo principal de respaldar políticas de estabilización macroeconómica. Estas aplicaciones básicamente buscan explotar los conjuntos de datos micro estructurados y amplios recopilados con fines regulatorios y estadísticos. Por ejemplo, los datos comerciales de los TR (Repositorios de Transacciones) pueden ser utilizados de diversas formas, como análisis de redes para identificar interconexiones entre participantes del mercado, así como riesgo sistémico. Y un último ejemplo se relaciona con datos sobre transacciones individuales de pagos, que tienen la ventaja de estar disponibles rápidamente y con frecuencia, en parte debido al aumento en las transacciones digitales observado en los últimos años. Estos datos pueden arrojar luz interesante sobre el funcionamiento del sistema financiero, ayudando a monitorear los pagos en tiempo real. Además, cada vez se utilizan más para respaldar ejercicios de predicción del ahora, como mencionaron aproximadamente una quinta parte de los encuestados

La cuarta categoría principal comprende la amplia gama de aplicaciones de suptech y regtech para respaldar políticas microsupervisivas (regulación y supervisión). En general, muchas de estas aplicaciones se centran en la evaluación de riesgos a nivel micro. Por ejemplo, la información a nivel de empresa recopilada de estados financieros o periódicos puede utilizarse para respaldar ejercicios de alerta temprana o mejorar la calificación crediticia (mencionado por un tercio y una quinta parte de los encuestados, respectivamente). Otra área importante se relaciona con la detección de fraudes (un cuarto de los casos), por ejemplo, al examinar contratos de crédito en busca de términos y condiciones sospechosos con el fin de mejorar la protección al consumidor. Por último, casi una quinta parte de los bancos centrales utilizan algoritmos de big data para fines de lucha contra el lavado de dinero y la financiación del terrorismo (AML/CFT) - por ejemplo, al analizar transacciones de pago para identificar patrones sospechosos.

Sobre esta misma temática el trabajo de Kathuria (2016) tiene como objetivo capturar cómo el análisis de big data se está utilizando con éxito en el sector bancario, con respecto a siguientes aspectos:

1. Patrón de gasto de los clientes
2. Usos del canal

3. Segmentación y elaboración de perfiles de clientes
4. Venta cruzada de productos basada en el perfil para aumentar la tasa de aciertos
5. Análisis de sentimientos y comentarios
6. Seguridad y gestión del fraude

Vimos una de las formas en que se capturan y utilizan los sentimientos de los clientes para evaluar el funcionamiento del banco. Hay muchas más formas en que los bancos y otras instituciones financieras han comenzado a capturar datos relacionados con los clientes para análisis de sentimiento, desde sitios web de redes sociales hasta varios canales de investigación de mercado. Observan análisis transaccionales y observan cómo los bancos hoy utilizan los patrones de gasto de sus clientes para seguir el comportamiento del consumidor basándose en el uso del canal y los patrones de consumo y segmentar a los consumidores dependiendo sobre los atributos antes mencionados e identificar clientes potenciales para la venta de productos financieros. La mayoría de estas indicaciones se pueden implementar fácilmente en los sistemas financieros utilizados en los bancos, lo que puede ayudar a los bancos a que refuercen la seguridad de los datos y previenen cualquier tipo de ataque. Aunque algunos de los controles, como los cables sospechosos las transferencias pueden dar lugar a “falsos positivos”; combinando varios de estos indicadores transaccionales y sentimentales para llegar adoptar un enfoque holístico de toma de decisiones y así implementar mecanismos sofisticados es ciertamente la necesidad de la hora para el sector bancario.

Más allá del consenso general sobre las bondades de big data en la regulación financiera observamos un impacto diferente según el área en la que ponga la lupa. Por ejemplo, en 2015 una encuesta⁴ realizada por el Comité Irving Fisher acerca de los usos e intereses sobre big data por parte de los bancos centrales, arrojó un gran interés por su uso, sin embargo, sólo un tercio de los consultados estaba usándolo efectivamente. Los proyectos de big data relacionados consisten principalmente en el uso de sistemas convencionales y conjuntos de datos estructurados (que se basan, por ejemplo, en fuentes “oficiales y administrativas” y basados en microdatos reportados por la industria bancaria). Sin embargo, también parece haber un gran interés en el uso de fuentes privadas de big data, como Google, datos de búsqueda, conjuntos de datos de proveedores de datos comerciales, datos de posicionamiento móvil, noticias en medios de comunicación, etc. (IFC, 2015).

Por su parte el FMI en 2017, declaró que los siguientes ítems fueron aprobados para comenzar su desarrollo: 1) uso de datos SWIFT para monitorear flujos financieros globales; 2) un sistema de alerta temprana basado en sentimientos/confianza; 3) pronóstico inmediato de PBI basado en las tendencias de búsqueda en Google; 4) agrupar flujos de efectivo del gobierno para mejorar el análisis de políticas; 5) aplicar análisis para una mejor administración tributaria y aduanera. Todas estas ideas se pensaron como complemento de los circuitos ya existentes.

En el mismo sentido, un trabajo de Binette y Tchegotarev (2019) analiza el lenguaje utilizado en el reporte de política monetaria con el fin de estudiar la forma en la que se comunica ya que eso es fundamental para la eficacia de la política monetaria. Mediante la utilización de *Deep learning*, específicamente una técnica llamada transfer learning, la red neuronal predice la siguiente palabra del cuerpo del texto y luego se entrena para diferenciar si las oraciones son positivas, negativas o neutrales. La nota también evalúa en qué se diferencian los reportes monetarios entre sí, lo que es un posible indicador de la incertidumbre percibida por parte de la autoridad de política monetaria en torno a las perspectivas económicas. El algoritmo destaca grandes perturbaciones en la historia macroeconómica de Canadá en los momentos donde los temas cambiaron drásticamente. Finalmente, con un novedoso algoritmo de aprendizaje profundo midieron el sentimiento (positivo o negativo) en la oración y agregaron los resultados para cada reporte. Los impactos excepcionalmente grandes de eventos clave como el 11 de septiembre, la crisis financiera mundial y otras son fácilmente reconocibles y han modificado significativamente el sentimiento.

⁴ Encuesta respondida por 69 bancos centrales y autoridades monetarias miembros del comité, representando un total de 83%.

Por su parte, D'Amuri y Marcucci (2017) utilizan a Google para intentar predecir la tasa de desempleo mensual de EE. UU. Para ello, sugieren un índice de intensidad de búsqueda de empleo en internet. Como resultado de su trabajo obtuvieron que los modelos basados en Google tuvieron un desempeño particularmente bueno en el punto de inflexión al comienzo de la Gran Recesión y su desempeño relativo se estabilizó a partir de entonces. Las especificaciones populares de series de tiempo que se complementan con este indicador mejoran su desempeño de pronóstico fuera de la muestra para horizontes de pronóstico de uno a doce meses.

5.1 Estabilidad Financiera

Uno de los objetivos más importantes de los bancos centrales es la de velar por la estabilidad financiera del sistema. En este apartado veremos cuál es el uso que le están dando actualmente las entidades financieras a big data en este aspecto.

Por ejemplo, Nyman et al. (2018) aplican el análisis algorítmico a grandes cantidades de datos basados en texto del mercado financiero para evaluar cómo las narrativas y los sentimientos desempeñan un papel a la hora de impulsar la evolución del sistema financiero. En este sentido, encuentran que los cambios en el contenido emocional de las narrativas de mercado están altamente correlacionados con las fuentes de datos. Se muestran claramente la formación (y posterior colapso) de niveles muy altos de sentimiento: un gran entusiasmo en relación con ansiedad antes de la crisis financiera mundial. Y desarrollaron una nueva metodología que intenta capturar el surgimiento del consenso de temas narrativos, lo que da una representación intuitiva de la creciente homogeneidad de creencias antes de la crisis. Los autores sostienen que con el aumento del consenso en torno a narrativas llenas de entusiasmo y sin ansiedad probablemente sea una importante señal de advertencia de una inminente crisis del sistema financiero.

Por su parte, Danielsson y Uthemann (2023) sostienen que una entidad reguladora del sistema financiero tiene dos metas distintas. La primera, conocida como micro regulación, aborda aspectos cotidianos como la gestión de riesgos, la protección al consumidor y la prevención de fraudes. En este caso, la integración de la inteligencia artificial promete ser de gran ayuda para las autoridades encargadas de la micro regulación. Por otro lado, la macro regulación, que se enfoca en cuestiones de amplio alcance como la estabilidad financiera, presenta un panorama menos claro. Esta perspectiva se concentra en el largo plazo, buscando prevenir crisis financieras sistémicas y grandes pérdidas en años y décadas futuras, así como en resolverlas si llegasen a ocurrir. Sin embargo, las políticas macro son más complicadas de implementar y menos precisas que las políticas micro. La eficacia de la inteligencia artificial para las autoridades financieras que supervisan tanto a nivel micro como macro, así como para abordar eventos de estrés y crisis, se ve directamente afectada por cuatro desafíos conceptuales: la disponibilidad de datos, incertidumbres desconocidas, interacciones estratégicas y la evolución de los objetivos.

En primer lugar, los autores mencionan el desafío de la disponibilidad de datos, lo cual puede parecer paradójico, considerando que el sistema financiero genera una gran cantidad de datos. Sin embargo, estos datos suelen medirse de manera inconsistente e incluso inexacta, y están limitados por autoridades que restringen su intercambio. Además, los datos sobre mitigación del estrés son escasos debido a la poca frecuencia de las crisis financieras. Es decir, la producción de datos varía según el aspecto de la realidad que se mire.

El segundo desafío radica en la singularidad de las crisis financieras, que comparten características comunes, pero también son únicas en función de factores específicos como la estructura institucional y las regulaciones vigentes en ese momento. Ese componente aleatorio y desconocido representa una complejidad para el uso de IA en la estabilidad financiera.

El tercer desafío conceptual que enfrenta la IA se relaciona con cómo responde el sistema financiero al control. La compleja retroalimentación entre las regulaciones y las respuestas a las

regulaciones frustra el seguimiento y control de los riesgos en el sistema. El riesgo exógeno supone que los factores de riesgo surgen fuera del sistema financiero y no están influenciados por él. La mayoría de las previsiones de riesgo y las micro regulaciones suponen que el riesgo es exógeno, fácil de analizar y, por lo general, no constituye un problema grave. El riesgo endógeno reconoce que la interacción de los agentes económicos no sólo impulsa los resultados, sino que también cambia la estructura del sistema financiero. El riesgo macro casi siempre es endógeno, y cualquier proceso basado en datos puede inducir a error a la IA porque los datos históricos pueden no ser informativos sobre la dinámica actual del riesgo endógeno.

El último desafío conceptual surge de los objetivos para los cuales la inteligencia artificial está diseñada para optimizar. En el ámbito de las micro regulaciones, se establece un conjunto de reglas conocido, que generalmente permanece estático en el período durante el cual se toman decisiones. Sin embargo, a medida que nos adentramos en escalas de tiempo más largas y eventos menos frecuentes, pero más severos, los objetivos se tornan cada vez más dinámicos. Aunque la inteligencia artificial puede adaptarse a entornos con objetivos en constante cambio, su eficacia disminuye y aumenta la probabilidad de errores a medida que la rareza de los eventos y las consecuencias de los errores se intensifican.

La inteligencia artificial seguramente será fundamental, según los autores, para asesorar a los encargados de tomar decisiones y será considerablemente más rápida en realizar análisis de escenarios que los expertos humanos, lo que contribuirá a obtener mejores resultados en el proceso de resolución. No obstante, no se puede delegar por completo la toma de decisiones a la IA, ya que su habilidad para construir escenarios e interpolar y extrapolar a partir de modelos estadísticos puede convertirse en una debilidad en comparación con el enfoque intuitivo de un grupo de personas que emplea la toma de decisiones distribuida basada en la comprensión intuitiva. En última instancia, esto implica que puede haber una delgada línea entre el asesoramiento de la IA y la toma de decisiones, especialmente en situaciones de crisis, donde un análisis de los eventos puede influenciar en gran medida las decisiones tomadas en respuesta a ellos.

Por su parte, el trabajo de FSB (2017) hace foco en la eficiencia que genera el uso de técnicas de inteligencia artificial. El uso de la inteligencia artificial y el aprendizaje automático en los servicios financieros puede traer beneficios clave para la estabilidad financiera en forma de eficiencias en la prestación de servicios financieros y en la supervisión de riesgos regulatorios y sistémicos. El procesamiento más eficiente de la información sobre riesgos crediticios y la interacción con el cliente a menor costo pueden contribuir a un sistema financiero más eficiente. Las aplicaciones internas (back-office) de inteligencia artificial y aprendizaje automático podrían mejorar la gestión de riesgos, la detección de fraudes y el cumplimiento de los requisitos regulatorios, potencialmente a un costo menor. En la gestión de carteras, el procesamiento más eficiente de la información de las aplicaciones de inteligencia artificial y aprendizaje automático podría ayudar a impulsar la eficiencia y la resiliencia de los mercados financieros, reduciendo desalineamientos de precios y (bajo supuestos benignos) reduciendo operaciones congestionadas. Finalmente, con los casos de uso por parte de los reguladores y supervisores, existe el potencial para aumentar la efectividad de la supervisión y realizar un mejor análisis de riesgos sistémicos en los mercados financieros.

Para la OECD (2021), el foco también está puesto en la optimización y la eficiencia ya que, sostienen, el big data ofrece una serie de ventajas para la estabilidad financiera:

1. Optimización en la prestación de servicios financieros: una mayor eficiencia en el procesamiento de datos sobre riesgos crediticios y en la interacción con los clientes, a un costo reducido, puede favorecer la eficacia del sistema financiero.
2. Supervisión de riesgos regulatorios y sistémicos: el análisis de grandes volúmenes de datos puede fortalecer la capacidad de los reguladores para monitorear los riesgos regulatorios y sistémicos en el ámbito financiero, lo que podría prevenir crisis y fomentar una mayor estabilidad.
3. Mejora en la gestión del riesgo: la implementación de herramientas de big data dentro de las instituciones financieras puede potenciar la gestión del riesgo, así como la detección de fraudes

y el cumplimiento de las normativas regulatorias, posiblemente con una reducción en los costos asociados.

4. Incremento en la eficiencia y resistencia de los mercados financieros: al procesar de forma más eficiente la información obtenida a través del big data, se puede mejorar la eficiencia y la resistencia de los mercados financieros, disminuyendo desalineaciones de precios y reduciendo el riesgo de operaciones congestionadas.

En resumen, el uso de big data en el ámbito de la estabilidad financiera parece estar enfocada en la mejora de la eficiencia de la regulación y la creación de indicadores de sentimiento de mercado que puedan conectarse con el riesgo sistémico. En este aspecto hay un punto de contacto cercano con el próximo apartado que abordaremos que es sobre la gestión de riesgos.

5.2 Gestión de Riesgos

Otra de las áreas donde se explotan los macrodatos de manera significativa es en la gestión de riesgos, sobre todo, en el riesgo de crédito. En este sentido, por ejemplo, un informe de IFC (2019) pone el foco en indicadores de riesgo de crédito. Tras la crisis financiera mundial de 2007-2008, los bancos centrales se han propuesto avanzar en iniciativas de estadísticas de datos con el fin de impulsar sus funciones supervisoras y monetarias. Como se menciona en Petropoulos et al. (2018) esto inevitablemente conducirá a que los bancos centrales posean grandes bases de datos, aumentando así, la necesidad de procesos robustos de extracción de datos y estadísticas financieras de modelado para apoyar una toma de decisiones más informada. Para los autores, los métodos de la econometría convencional no logran capturar eficientemente la información contenida en todo el espectro de los conjuntos de datos. Para abordar esos desafíos, en el trabajo investigaron el análisis de un gran conjunto de datos sobre préstamos de crédito corporativo y utilizaron técnicas de aprendizaje automático de vanguardia y redes neuronales de aprendizaje profundo.

La novedad del enfoque radica, sostienen los autores, en la combinación de la minería de datos y algoritmos que tienen como objetivo reducir la dimensionalidad de los datos y aumentar la precisión en predecir el comportamiento futuro de los préstamos corporativos, para facilitar una evaluación más efectiva en la supervisión micro y macro del riesgo de crédito en el sistema bancario griego. Para eso, se basan en el análisis de un gran conjunto de datos a nivel de préstamos, que abarca un período de 10 años de la economía griega con el fin de realizar una clasificación de la calidad crediticia del deudor y cuantificación de la probabilidad de incumplimiento bajo una configuración a lo largo del ciclo. Los resultados experimentales se compararon con otros métodos tradicionales, como métodos de regresión logística y análisis discriminante, y arrojaron resultados significativamente superiores. En la etapa final del análisis, desarrollaron una calificación crediticia que puede ofrecer un mecanismo de seguimiento proactivo de la dinámica del riesgo de crédito. Finalmente, el marco metodológico introducido puede respaldar un análisis más profundo de iniciativas de bases de datos como las del BCE AnaCredit (caso mencionado en el capítulo 4).

En relación con este mismo tema, un trabajo de Addo et al. (2018), sostiene que, debido a la tecnología hiperasociada al big data, la disponibilidad de datos y la potencia informática, la mayoría de los bancos están renovando sus modelos de negocio. Construyeron clasificadores binarios basados en modelos de *deep* y *machine learning* con datos reales para predecir la probabilidad de incumplimiento de préstamos. Las 10 características más importantes de estos modelos fueron seleccionadas y luego utilizadas en el proceso de modelado para probar la estabilidad de clasificadores binarios comparando su desempeño en datos separados. Los autores mencionan la importancia de que los reguladores tomen decisiones rápidas para ordenar el uso de técnicas de ciencia de datos que mejoren el rendimiento y ayude a comprender los límites de algunos de estos algoritmos. Por último, observaron que los modelos basados en árboles son más estables que los modelos basados en redes neuronales artificiales de múltiples capas.

En línea con este mismo riesgo, un trabajo de Huang et al. (2004), destaca la superioridad de los modelos de IA ya que lograron un mejor rendimiento que los métodos estadísticos tradicionales. El artículo introduce la técnica de aprendizaje automático basada en las máquinas de vectores de soporte y la red neuronal de retro-propagación, obteniendo altos porcentajes de precisión en la predicción. Luego de la crisis de 2008 y haciendo hincapié en el rol fundamental del consumidor en cada etapa de las crisis económicas, el estudio desarrolló un modelo de aprendizaje automático para el incumplimiento y la morosidad del crédito al consumo, cuyas previsiones permiten ahorrar entre el 6% y 22% de las pérdidas totales.

En este estudio, utilizaron dos conjuntos de datos para instituciones financieras de Taiwán y bancos comerciales de Estados Unidos. Los resultados mostraron que las máquinas de vectores de soporte lograron una precisión comparable a la de las redes neuronales de retropropagación. Aplicando los resultados de la investigación en la interpretación del modelo de redes neuronales, llevaron a cabo un análisis de contribución de variables financieras de entrada y determinaron la importancia relativa de las variables de entrada. Sostienen que esta información puede ayudar a los usuarios a comprender mejor el proceso de calificación de bonos. También utilizaron los resultados del análisis de contribución para comparar las características de los procesos de calificación de bonos en los mercados de Estados Unidos y Taiwán. Encontraron que los modelos óptimos que construyeron para los dos mercados utilizaban listas similares de variables financieras como entradas, pero que la importancia relativa de las variables era bastante diferente en los dos mercados.

En su trabajo sobre riesgo de crédito hipotecario, Sirignano et al. (2016) han desarrollado un modelo de aprendizaje profundo del riesgo hipotecario a lo largo de varios períodos y lo utilizaron para examinar un conjunto de datos que incluye registros de origen y desempeño mensual de más de 120 millones de hipotecas originadas en todo Estados Unidos entre 1995 y 2014. Los estimadores de las estructuras temporales de las probabilidades condicionales de prepagos, ejecuciones hipotecarias y diversos estados de morosidad capturan la dinámica de una amplia gama de variables específicas de los préstamos, así como variables macroeconómicas a nivel de código postal. Estos estimadores revelan la naturaleza altamente no lineal de la relación entre las variables y el comportamiento del prestatario, particularmente en lo que respecta al prepagó. Además, destacan los efectos de las condiciones económicas locales en el comportamiento del prestatario.

En este documento, el análisis se fundamenta en un modelo de aprendizaje profundo no lineal de transiciones de estado de prestatarios a lo largo de varios períodos, que tiene en cuenta la influencia en el comportamiento del prestatario de numerosas variables específicas del préstamo y del prestatario, así como variables económicas y demográficas a nivel nacional, estatal, de condado y de código postal.

Los autores sostienen que sus hallazgos empíricos ofrecen nuevas e importantes perspectivas sobre el comportamiento de los prestatarios de hipotecas. Se observa que la relación entre el comportamiento del prestatario y los factores de riesgo es altamente no lineal, lo que plantea interrogantes sobre muchos modelos lineales analizados en investigaciones previas. Los efectos de interacción, donde el impacto de una variable depende de los valores de otras variables, son comunes. En particular, encontraron evidencia que sugiere que los prepagos son los eventos más afectados, ya que involucran los efectos no lineales más fuertes entre todos los eventos. Los principales impulsores de los prepagos, que incluyen saldos pendientes originales y actuales del préstamo, interactúan de manera significativa, incluso antes de considerar factores estándar como las tasas de interés y los diferenciales de tasas de interés. Los prestatarios de préstamos junto con saldos pendientes relativamente pequeños son los más propensos a prepagar.

Por otro lado, los resultados obtenidos también resaltan la importancia de las condiciones económicas locales para el comportamiento del prestatario. Encontraron que el desempleo estatal tiene el mayor poder explicativo entre todos los factores, lo que sugiere una conexión más estrecha entre los mercados de financiamiento hipotecario y la macroeconomía de lo que se pensaba previamente. Además, se descubrió que la sensibilidad de un prestatario a los cambios en el desempleo no es constante, sino que depende fuertemente del desempleo actual. Esta sensibilidad también varía significativamente en toda la población de prestatarios, lo que destaca un efecto de interacción entre el

desempleo y muchas características estándar de los prestatarios que antes se estudiaban de forma individual.

Por último, destacan que estos resultados empíricos tienen importantes implicancias para los inversores en valores respaldados por hipotecas. El papel predominante del desempleo resalta la exposición de los prestatarios a los ciclos económicos, que pueden ser una fuente significativa de correlación entre préstamos. Esta correlación destaca la necesidad de que los inversores en hipotecas diversifiquen el riesgo hipotecario geográficamente, más allá de las características convencionales de los prestatarios resaltadas en la literatura. La naturaleza no lineal de la influencia del desempleo y otras variables en el comportamiento del prestatario tiene implicaciones para la calificación y cobertura de dichos préstamos. Los inversores deben tener en cuenta los efectos no lineales al construir posiciones de cobertura contra la volatilidad macroeconómica, mientras que las agencias de calificación deben considerar el comportamiento no lineal al evaluar la exposición de los inversores en hipotecas ante condiciones macroeconómicas adversas.

El trabajo de Chakraborty y Joseph (2017) modela hipotéticos escenarios de alerta en los balances contables de las entidades en un entorno de información incompleta. Esto puede ser útil para detectar potenciales riesgo de crédito, liquidez y de mercado ya que lo que hacen en dicho trabajo es, mediante el uso de aprendizaje automático, como los árboles de decisión, crear un umbral de desvío para ciertas líneas de balance. De esta forma, un caso de alerta puede dispararse por una variación por arriba o por debajo de ciertos umbrales de los activos de una empresa en un trimestre.

Este mismo trabajo también se aborda un posible riesgo sistémico dado por la irrupción de alguna empresa que cause un cambio de lógica en la industria que opera como sucedió con Uber y Airbnb. En el tercer caso de estudio, mediante un análisis de *clustering* se analiza la relación de financiamiento entre empresas e inversores dentro del sector tecnológico con el objetivo de preseleccionar empresas disruptivas en sus industrias que puedan generar cambios relevantes para el marco. Los autores remarcan que esto mismo puede suceder en la industria financiera y, por su relevancia en la economía, los bancos centrales pueden anticiparse y estar preparados ante posibles riesgos sistémicos.

Sobre el riesgo sistémico, también el trabajo de Nyman et al. (2021) aplica un análisis algorítmico basado a datos en texto del mercado financiero para evaluar cómo las narrativas y los sentimientos podrían impulsar la evolución del sistema financiero. Encontraron que los cambios en el contenido emocional de las narrativas están altamente correlacionados entre las fuentes de datos y muestran la formación (y el posterior colapso) de la exuberancia antes de la crisis financiera global. Sostienen que las métricas también tienen poder predictivo para otros indicadores de sentimiento comúnmente utilizados y parecen influir en las variables económicas. Una novedosa aplicación de aprendizaje automático también apunta a un mayor consenso en torno a la narrativa fuertemente positiva anterior a la crisis. En conjunto, argumentan que sus métricas podrían ayudar a advertir sobre inminentes dificultades en el sistema financiero.

Este artículo utiliza un enfoque narrativo de convicción para intentar identificar el sentimiento y el riesgo sistémico en los sistemas financieros utilizando big data. Específicamente, aplicaron análisis algorítmicos motivados teóricamente a grandes cantidades de datos no estructurados basados en texto que abarcan tres fuentes de interés potencial relacionadas con el sistema financiero: comentarios diarios internos del Banco de Inglaterra sobre noticias y eventos del mercado, artículos de noticias de Reuters en el Reino Unido e informes de investigación de corredores. Identificaron métricas cuantitativas que intentan capturar cambios en el sentimiento del sistema financiero junto con el grado de consenso sobre el tema narrativo en el mercado. Las dos nuevas medidas que desarrollaron sugieren que los cambios en el sentimiento narrativo tuvieron una influencia significativa en el sistema financiero en el período previo a la crisis financiera global y posteriormente.

En resumen, la medida de sentimiento parece dar señales de alerta temprana de acontecimientos financieros importantes, y los cambios clave se identifican mediante pruebas de ruptura estructural. En particular, el sentimiento general se encontraba en niveles muy altos y estables a mediados de la década de 2000, lo que posiblemente sea indicativo de exuberancia en el sistema financiero. Desde mediados

de 2007, un aumento de la ansiedad provocó rápidas caídas del sentimiento que continuaron hasta poco después del colapso de Lehman Brothers. Y hubo nuevas caídas en la confianza antes del inicio de la crisis soberana de la zona del euro en 2011-2012. También, para medir la solidez de las métricas de sentimiento, las compararon con medidas económicas y financieras estándar de confianza del consumidor y volatilidad del mercado, y con algunas medidas de sentimiento e incertidumbre basadas en textos comúnmente utilizadas, pero más atóricas. Los resultados resaltan cómo las medidas de sentimiento propuestas y el consenso narrativo se correlacionan bien con ciertas variables económicas y financieras, y en algunos casos incluso parecen "causarlas". Dependiendo de la fuente del texto, algunos funcionan mejor con variables financieras, otros con variables macroeconómicas.

5.3 Investigación económica y predicción

Otra de las áreas más relevantes para los bancos centrales es la de investigación y predicción, ya que de esta manera pueden generar nuevos saberes a aplicar en base a sucesos pasados o modelos teóricos. En esta sección veremos cómo, a través del uso de datos no estructurados generalmente, los bancos centrales pueden generar predicciones, por ejemplo, sobre el bienestar de la sociedad, mediciones como el consumo y el PBI y sentimientos de mercado.

En relación con esto, el trabajo de Chakraborty y Joseph (2017) explica el potencial que tiene el aprendizaje automático para los bancos centrales respecto de los enfoques tradicionales. En el caso de estudio planteado por los autores, analizan un marco de modelación para pronosticar la inflación del Reino Unido en un plazo de 2 años, utilizando series temporales de variables como PBI, desempleo, deuda interna, tasa de interés y precios de materias primas.

Por otro lado, hay autores que demuestran cómo la información textual no estructurada en el mayor diario de negocios en Noruega se puede descomponer en temas de noticias diarias y utilizarse para pronosticar en tiempo actual el crecimiento trimestral del PBI (Thorsrud, 2016). La hipótesis de este autor es la siguiente: en la medida que el periódico proporcione una descripción relevante en la economía, cuanto más intensamente este representado un tema determinado en un momento dado, más probabilidad que ese tema represente algo de importancia para el desarrollo actual y futuro de la economía. El autor demuestra que, en comparación con un gran banco de expertos aquí representado por los bancos centrales oficiales y un sistema de combinación de pronósticos de última generación, la metodología propuesta en ocasiones funciona hasta un 15 por ciento mejor, y es especialmente competitiva en torno a puntos de inflexión importantes del ciclo económico. Además, si la agencia estadística que produce las estadísticas del PIB hubiera utilizado la metodología basada en noticias, habría resultado en un proceso de revisión menos ruidoso. Por lo tanto, concluye que las noticias reducen el ruido.

En la economía, los agentes utilizan una amplia gama de información de alta frecuencia, incluidos los medios de comunicación, para orientar sus acciones y, de esta manera, influir en las fluctuaciones económicas agregadas. En dicho trabajo, se presenta cómo los datos de texto no estructurados obtenidos de un periódico de negocios importante pueden mejorar la precisión de la predicción ahora del crecimiento trimestral del PIB, tanto en promedio como en momentos críticos. Específicamente, al desglosar los datos de texto en temas de noticias diarias y emplear un modelo de factor dinámico con frecuencia mixta y variable en el tiempo, se demuestra que es posible generar pronósticos que superan hasta en un 50 por ciento a los de modelos de series temporales simples, y en ocasiones hasta un 15 por ciento mejor que un sistema de predicción ahora de última generación.

Un elemento clave para los resultados obtenidos, mencionan, en este análisis es la aplicación de un modelo de Asignación de Dirichlet Latente (LDA) para desglosar el texto no estructurado en temas de noticias. En comparación con los enfoques textuales existentes en economía y finanzas, como el recuento de palabras, la clasificación de sentimientos y la recolección de términos y tendencias de búsqueda web, el enfoque adoptado aquí es considerablemente intensivo en datos y requiere un alto

poder computacional. Sin embargo, se podría argumentar que el enfoque LDA ofrece varias ventajas en términos de automatización e interpretación.

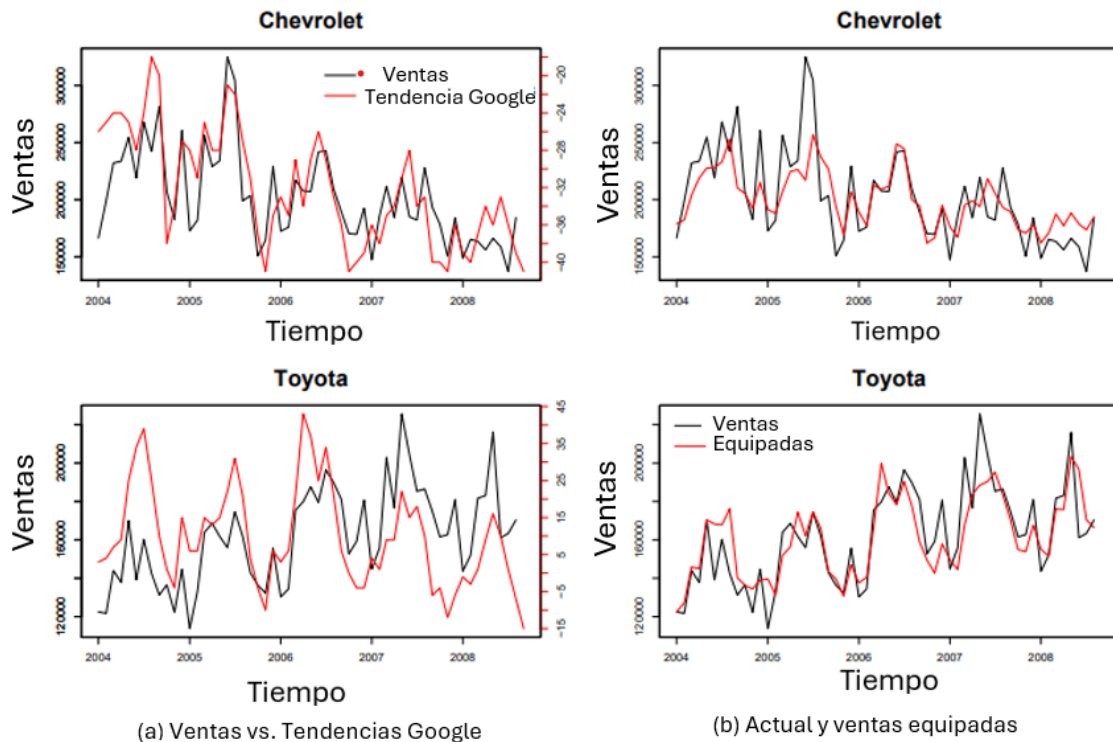
Simoni (2019) en su trabajo estudia el uso de datos de búsqueda de Google para predecir la tasa de crecimiento del PIB de la zona euro. Su principal objetivo es evaluar la utilidad de los datos de búsqueda de Google para la predicción inmediata cuando no se dispone de datos oficiales, en el contexto de un pseudo análisis en tiempo real. Debido a que los datos de búsqueda de Google son de alta dimensión, en el sentido de que el número de variables es grande en comparación con la dimensión de la serie temporal, hay un precio para pagar por usarlos: primero, se debe reducir su dimensionalidad de ultra alta a alta mediante el uso de un procedimiento de selección y, en segundo lugar, se necesita utilizar un estimador regularizado para tratar las variables preseleccionadas.

Del análisis empírico realizado surgen cuatro hechos destacados. Primero, en el contexto de un análisis en tiempo real, destacan la utilidad de los datos de búsqueda de Google en estimación inmediata de la tasa de crecimiento del PIB de la zona del euro durante las primeras cuatro semanas del trimestre, cuando no hay información sobre el estado de la economía. Mostraron que, al comienzo del trimestre, los datos de Google proporcionan una imagen precisa de la tasa de crecimiento del PIB. En segundo lugar, tan pronto como estén disponibles los datos oficiales, es decir, a partir de la semana 5 con la publicación de las encuestas de opinión, el poder relativo de predicción inmediata de los datos de Google se desvanece instantáneamente. En tercer lugar, demuestran que preseleccionar los datos de Google antes de ingresar a los modelos de predicción inmediata parece ser una estrategia pertinente en términos de precisión de la predicción inmediata. Estos resultados argumentan a favor del uso de datos de búsqueda de Google al comienzo del trimestre, cuando no hay información oficial información disponible para la formulación de políticas en tiempo real.

Respecto de la utilización de la información no estructurada, hay diversos estudios que explotan este tipo de fuentes de información para extraer predicciones que superan a las mediciones tradicionales. Por ejemplo, Choi y Varian (2009) destacan que, a diferencia de las publicaciones gubernamentales de datos sobre las condiciones de la economía que suelen tener cierto atraso, las tendencias de Google proveen diaria o semanalmente reportes de varias industrias. Los autores hipotetizan que las consultas realizadas en dicho motor de búsqueda pueden tener cierta correlación con el nivel actual de la actividad económica, y, por lo tanto, pueden ser útiles para predecir las publicaciones de datos posteriores. Por ejemplo, el volumen de consultas sobre una marca particular de autos durante la segunda semana de junio puede ser útil para predecir el reporte de ventas del mes de junio cuando recién se conozca en Julio. Para ello utilizan un índice de volumen de consultas de Google por ubicación geográfica y categoría y lo aplican modelos de auto regresión estacionales simples y modelos de efectos fijos para demostrar que son superiores a los modelos que excluyen estos predictores.

En los siguientes gráficos se puede ver la relación encontrada por los autores entre las búsquedas en Google y las ventas de las marcas Chevrolet y Toyota. Explican que el índice de tendencias de Google en las dos últimas semanas del mes pasado esta positivamente asociado con las ventas actuales de cada marca.

Figura 7 Ventas y tendencia de Google – Chevrolet y Toyota



Nota: Adaptado de Choi y Varian (2009 p.9)

Por su parte, Algan et al. (2015) utilizando también las tendencias de Google construyeron un indicador de bienestar individual en EE. UU. El modelo que proponen predice con éxito la evolución fuera de la muestra de la mayoría de las medidas subjetivas de bienestar. Mediante este trabajo los autores proponen mostrar cómo el big data puede mejorar la comprensión de los fundamentos del bienestar humano al encontrar que palabras clave como empleo, seguridad financiera, vida familiar y ocio son fuertes predictores de la variación en el bienestar subjetivo en EE. UU. Esta medición permite tener una visión íntegra de una variable tan completa compleja como lo es el bienestar y podría lograrse casi en tiempo real.

En el trabajo de Tobback et al. (2018) se expone que, recientemente se ha intentado medir la incertidumbre de la política económica utilizando desacuerdos de pronóstico y referencias de noticias, lo que resulta en el frecuentemente mencionado “Índice de Incertidumbre de la Política Económica”. En la configuración original, se supone que un artículo de noticias aborda la incertidumbre política si contiene ciertas palabras clave predefinidas. En este trabajo intentan mejorar las noticias parte del índice IPE utilizando técnicas de minería de texto. Durante la crisis, ha quedado claro que los efectos económicos negativos dentro del zona del euro son mucho más fuertes de lo que se había previsto anteriormente y los mecanismos detrás de estos efectos de contagio aún no están completamente identificados. En este contexto los autores exploran qué parte de la incertidumbre política en Bélgica se debe por contagio en la zona del euro.

La contribución de este trabajo es triple. Primero, se intenta mejorar el índice ya existente gubernamental resolviendo algunos de sus problemas más importantes. En segundo lugar, intentan demostrar cómo las técnicas de minería de datos se pueden aplicar para resolver un problema relacionado con políticas. En este caso particular, el problema relacionado con las políticas es encontrar

una medida para la incertidumbre de política económica. Evalúan la incertidumbre política detectando patrones automáticamente en artículos de noticias, utilizando técnicas de minería de textos de última generación en un gran conjunto de datos de artículos periodísticos. Finalmente, este es el primer estudio que estima un índice de incertidumbre de política económica para Bélgica, teniendo en cuenta al mismo tiempo los efectos de contagio dentro de la zona del euro.

El objetivo del trabajo además es predecir y pronosticar el consumo privado trimestral. La selección de indicadores mensuales en tiempo real se centra en variables estándar (indicadores “duros” / “blandos”) y variables menos estándar. Entre este último grupo los autores analizan: i) indicadores indirectos de incertidumbre económica y política; ii) tarjetas de pago transacciones, medidas en “puntos de venta” y retirios en cajeros automáticos; iii) indicadores basados en consultas de búsqueda relacionadas con el consumo obtenidas a través de Google. Se destacan algunos resultados: 1) los indicadores “duros” y de tarjetas de pago son los que obtienen mejores resultados cuando se toman individualmente y aún más cuando se combinan; 2) sin embargo, los indicadores “suaves” son útiles para detectar señales cualitativas en el horizonte de predicción inmediata; 3) los indicadores de incertidumbre y basados en Google añaden valor cuando se combinan con indicadores tradicionales, más notablemente en horizontes de estimación más allá del pronóstico inmediato.

Por su parte, un trabajo de Bodas et al. (2018) en España presenta un índice de comercio minorista de alta dimensionalidad construido para pronosticar el desempeño económico del sector de comercio minorista. Para ello utilizan técnicas y recursos de big data, basados en las operaciones de los clientes del banco, con tarjetas de crédito y débito terminales punto de ventas. Además, analizaron la dinámica del consumo de alta frecuencia utilizando el comportamiento minorista de BBVA y un modelo estructural de series temporales. Los resultados de este trabajo muestran que los datos de transacciones con tarjeta replican con gran precisión la evolución del mercado español y es un buen predictor de la actividad comercial.

El trabajo de Gil et al. (2018) propone un nuevo indicador de sentimiento económico basado en artículos periodísticos que permite evaluar la actividad económica en España en tiempo real. Este indicador se puede construir diariamente y resulta más eficaz que el Indicador de Sentimiento Económico de la Comisión Europea, al tiempo que muestra tres ventajas principales en comparación con este tipo de indicadores: 1) no sufre sesgos a lo largo de los ciclos económicos; 2) se puede construir diariamente y en tiempo real –cada día, el índice se puede calcular utilizando datos que se refieren al día anterior–, mientras que las encuestas se refieren, en general, a las primeras semanas del mes de referencia y se publican en los últimos días del mes; 3) no depende de la tasa de respuesta de la encuesta, lo que podría generar problemas de muestreo.

Según IFC (2024) ante la pregunta sobre cuáles son los casos de uso actuales de machine learning e IA en los bancos centrales, se pueden organizar mejor por alcance: (i) recolección de información y compilación estadística; (ii) análisis macroeconómico y financiero para respaldar la política monetaria; (iii) supervisión de sistemas de pago; y (iv) supervisión y estabilidad financiera. Esta sección proporciona ejemplos relevantes en cada área.

Recolección de información es garantizar la disponibilidad de datos de alta calidad como insumos para el análisis económico y para la compilación y producción de estadísticas es un desafío importante para los bancos centrales. Los problemas incluyen la limpieza de datos, el muestreo, la representatividad y la correspondencia de nuevos datos con fuentes existentes. El volumen y la complejidad crecientes de los datos requieren herramientas de calidad de datos eficientes y flexibles.

Para proporcionar microdatos de alta calidad, los bancos centrales están utilizando progresivamente técnicas de machine learning. Los bosques de aislamiento son particularmente adecuados para los conjuntos de datos grandes y granulares típicos de los bancos centrales, debido a su escalabilidad y capacidad para identificar valores atípicos independientemente de la forma de distribución de los datos. También hay beneficios en un enfoque de dos pasos: inicialmente, un modelo identifica de forma autónoma valores atípicos potenciales, que luego son revisados por expertos que proporcionan retroalimentación para refinar el algoritmo. Este enfoque equilibra el valor de la experiencia en el dominio con los costos de los aportes humanos. Al analizar diferentes métodos para

explicar la clasificación de valores atípicos, este enfoque puede superar el problema de los modelos de machine learning "caja negra" que carecen de "explicabilidad". Además, los métodos de machine learning explicables proporcionan orientación a los expertos sobre qué puntos de datos merecen verificación manual.

Con respecto al análisis macroeconómico y financiero para respaldar la política monetaria, los bancos centrales dependen ampliamente de dicho análisis para construir una explicación sobre su política. En un entorno complejo, un desafío importante es extraer eficientemente información de una amplia variedad de fuentes de datos tradicionales y no tradicionales. Para los autores, el machine learning ofrece herramientas valiosas en esta área.

Las redes neuronales, por ejemplo, pueden descomponer la inflación de servicios en diferentes componentes, revelando cuánta inflación se debe a aumentos de precios pasados, expectativas de inflación, la brecha de producción o precios internacionales. Dichos modelos pueden procesar más variables de entrada que los econométricos tradicionales, lo que permite a los bancos centrales utilizar conjuntos de datos granulares en lugar de datos más agregados. Otra ventaja es la capacidad de las redes neuronales para reflejar complejidades no lineales en los datos, lo que puede ayudar a los modeladores a capturar mejor las no linealidades, desde el límite inferior cero hasta las tenencias de activos desiguales y cambios en la dinámica de la inflación.

Otros casos de uso son la obtención de estimaciones en tiempo real de las expectativas de inflación o la síntesis de condiciones económicas a lo largo del tiempo. Por ejemplo, los modelos de bosques aleatorios pueden identificar publicaciones en redes sociales relacionadas con los precios y luego alimentarlas en otro modelo de bosques aleatorios que clasifica cada publicación como reflejo de inflación, deflación u otras expectativas. La diferencia en los recuentos diarios de publicaciones en redes sociales para indicadores de inflación más altos versus más bajos indica las expectativas de inflación. De manera similar, las publicaciones en redes sociales pueden usarse para rastrear la credibilidad de la política monetaria del banco central con el público en general.

En relación con el avance la mayor cantidad de datos es necesario establecer marcos de calidad de la información. En este sentido, el trabajo de Araujo et al. (2022) sostiene que un ejemplo reciente es el proyecto de detección de anomalías llevado a cabo por el Banco Central Europeo, el cual tiene como objetivo respaldar la verificación de la calidad de los datos utilizados en la producción de estadísticas sobre las tasas de interés a corto plazo del euro. Este proyecto se basa en un conjunto de datos detallados sobre transacciones individuales observadas en los mercados monetarios, conocido como el Informe Estadístico sobre el Mercado Monetario, que implica la participación de 47 bancos distribuidos en 10 países y abarca aproximadamente 50.000 transacciones diarias.

Se enfrentaron desafíos significativos, especialmente relacionados con la presencia de variables no numéricas, asimetrías en la distribución de datos, la necesidad de realizar verificaciones rápidas de la calidad de los datos para respaldar un proceso de producción diario y la complejidad en la interpretación de los resultados obtenidos para los usuarios. Estos desafíos fueron abordados mediante el empleo de diversas técnicas de aprendizaje automático, que incluyeron la conversión de variables categóricas a numéricas, la utilización de correlaciones observadas y la detección de anomalías mediante diferentes modelos y algoritmos, como el análisis de regresión estándar, bosques de aislamiento, agrupamiento jerárquico y la identificación de anomalías basada en el aprendizaje supervisado con XGBoost.

Otro ejemplo es la iniciativa del Banco de Pagos Internacionales para desarrollar un flujo de trabajo de validación altamente automatizado que se basa en herramientas de aprendizaje automático y en una capacidad informática mejorada. Se informa que el enfoque de validación de datos implementado es adecuado para un gran volumen de indicadores, aproximadamente 3.000 series temporales diarias de datos del mercado financiero. Por lo tanto, claramente supera a métodos más tradicionales, como controles gráficos o advertencias basadas en umbrales. Además, la nueva solución parece ser mejor para abordar el riesgo de errores que son "Tipo I" o "falsos positivos" (es decir, el rechazo erróneo de la existencia de una anomalía que de hecho existe) así como "Tipo II" o "falsos negativos" (es decir, la falta de rechazo de la existencia de una anomalía mientras que los datos son de hecho correctos).

Uno de los aspectos clave de los nuevos enfoques de calidad que emplean técnicas de aprendizaje automático es la detección de valores atípicos en la gran cantidad de observaciones recopiladas en tiempo real en los repositorios de big data, lo que está haciendo que las acciones manuales tradicionales sean cada vez más ineficientes o incluso imposibles de llevar a cabo. Para abordar esta situación, el Banco de Israel ha desarrollado paneles de control utilizando una aplicación específica (R Shiny), que aprovecha el lenguaje de programación R para análisis estadístico y gráficos, para verificar todas las transacciones diarias en los mercados de derivados reportadas por las instituciones financieras. Una vez que los datos se cargan en el panel de control, los usuarios pueden seleccionar las variables a analizar, aplicar filtros, explorar los datos visualmente y utilizar algoritmos específicos para detectar valores atípicos, como el Bootlier Plot basado en histogramas de densidad, el bosque de aislamiento, entre otros.

En el trabajo de Araujo et al. (2022) también se destaca que las técnicas de aprendizaje automático, y en particular el aprendizaje profundo, están siendo cada vez más utilizadas para apoyar los ejercicios de modelado macroeconómico. Sostienen que la experiencia de los bancos centrales muestra que pueden beneficiarse de la disponibilidad de fuentes de datos no convencionales y nuevos métodos basados en aprendizaje automático, como el aprendizaje profundo. Esta lección está en línea con la literatura existente, especialmente cuando se trata de casos en los que los datos/expertise son limitados. Los autores sostienen que se ha entrenado una herramienta con datos de EE. UU. y propuesto una estrategia de transferencia de aprendizaje para identificar las fases del ciclo económico en Brasil y la zona del euro. Un interés de este enfoque es aprovechar los conocimientos adquiridos por los expertos económicos de una región y aplicarlos a otras áreas geográficas, por ejemplo, en ausencia de un comité de fechas reconocido para el ciclo económico.

En este trabajo también mencionan otro caso de uso importante de aprendizaje automático es comprender de manera agnóstica cuáles son los impulsores de las variables macroeconómicas, es decir, siguiendo un enfoque puramente basado en datos en lugar de depender de suposiciones previas. Por ejemplo, afirman, se han aplicado la técnica de bosque aleatorio para analizar los impulsores de la inflación y en particular el papel de los factores financieros que suelen ser ignorados en el conjunto de herramientas de los modeladores macroeconómicos. Más en general, los modelos basados en aprendizaje automático parecen especialmente adecuados para descubrir factores explicativos de una multitud de variables candidatas.

Es más, sostienen, dadas sus crecientes contribuciones al análisis económico y financiero y al modelado del comportamiento de los agentes, no debería sorprender que se recurra cada vez más al aprendizaje automático para apoyar ejercicios de pronóstico que abarcan desde el corto plazo, es decir, ejercicios de predicción en tiempo real que intentan predecir el pasado reciente y el presente, hasta el horizonte a más largo plazo, incluidos los escenarios de riesgo. El enfoque ha sido principalmente en mejorar la precisión de los ejercicios de pronóstico "estándar" de los bancos centrales, que típicamente se centran en el PIB real y la inflación como variables clave que influyen en sus decisiones de política.

En relación con la inflación, el Banco Central de la Federación Rusa y la Escuela Superior de Economía han examinado cómo diversas técnicas de aprendizaje automático contribuyen a prever el índice de precios al consumidor (IPC) utilizando tanto datos en tiempo real como datos ajustados. Se llevó a cabo una comparación entre cuatro algoritmos de aprendizaje automático populares: bosques aleatorios, aumento de gradiente, red neuronal bayesiana y regresión regularizada, esta última siendo un tipo de regresión lineal adaptada para manejar un gran número de variables y evitar el sobreajuste, como la red elástica. Se determinó que todos estos algoritmos superaban a un modelo autorregresivo, lo que refuerza la evidencia sobre la utilidad de los métodos de aprendizaje automático en la predicción. Sin embargo, la elección del modelo con mejor desempeño variaba según el horizonte de predicción. Por ejemplo, se observó que el aumento de gradiente y las redes neuronales tenían un mejor desempeño para pronósticos a un mes, mientras que la red elástica presentaba el mejor rendimiento en un horizonte de seis meses.

En cuanto a la política monetaria, Araujo et al. (2022) sostienen que las decisiones de política monetaria pueden apartarse de la influencia "pura" de los desarrollos macroeconómicos en cuanto a la producción e inflación debido a factores adicionales. Sin embargo, estas relaciones suelen ser difíciles

de analizar, principalmente debido a la no linealidad y problemas de dependencia temporal, lo que representa un caso de uso importante para las técnicas de aprendizaje automático.

Un ejemplo de esto es el enfoque basado en aprendizaje automático desarrollado por el Banco de Indonesia para considerar mejor el impacto del comportamiento de los inversores extranjeros en los movimientos del tipo de cambio y, por lo tanto, en las decisiones de política monetaria. Este enfoque implicó el análisis de una gran cantidad de variables y el uso de algoritmos de clasificación basados en árboles para seleccionar las variables más relevantes. Luego, se utilizaron diversas técnicas de aprendizaje automático para predecir los montos de inversión diarios de los inversores individuales. Los resultados demostraron la importancia de los rendimientos de los bonos como predictores de los flujos de inversión externa, dependiendo del horizonte de inversión de los inversores. Se planea expandir este enfoque para analizar otros mercados financieros y facilitar su integración en las operaciones de política monetaria a través de un panel de control.

Además de facilitar la captura de una gama más amplia de factores determinantes, las herramientas de aprendizaje automático parecen contribuir a una mejor comprensión del proceso de toma de decisiones de política monetaria en sí mismo. Un estudio realizado por el Fondo Monetario Internacional, que retoman los autores, se enfocó en las discusiones de política monetaria en los Estados Unidos y empleó una técnica de procesamiento de lenguaje natural específica: la Asignación Latente de Dirichlet, para analizar los temas tratados por los miembros del Comité Federal de Mercado Abierto (CFMA) durante el período de 2003 a 2012. Las transcripciones de las reuniones se dividieron en aproximadamente 45.000 entradas de texto, que consistían en oraciones expresados por los gobernadores. El algoritmo fue diseñado con el propósito de segmentar todo el conjunto de datos de texto en ocho temas: pronósticos, modelado económico, lenguaje de declaraciones, riesgos, banca, decisiones de votación, actividad económica y comunicación. Este estudio reveló los temas que recibieron mayor atención por parte del CFMA al tomar decisiones de política. Por ejemplo, se encontró que durante la crisis financiera mundial predominaban las discusiones sobre el modelado económico, pero posteriormente el tema principal fue la banca, y más adelante, la comunicación.

5.4 Supervisión y Detección de Fraudes

En cuanto a la supervisión y regulación bancaria, existe un claro impulso para que las autoridades reguladoras obtengan más datos a nivel micro. Después de la crisis financiera, los reguladores se han mostrado interesados en ampliar su recopilación de datos para monitorear mejor los riesgos y vulnerabilidades financieras. Nuevas fuentes de big data pueden respaldar estas tareas de supervisión; tales fuentes incluyen, por ejemplo, operaciones en línea en plataformas de negociación, transacciones de pago con tarjeta de crédito, datos de banca móvil, registros relacionados con sistemas de liquidación de valores y pagos en efectivo, cámaras de compensación y operaciones de reporto y liquidación de derivados, así como transacciones comerciales y minoristas y transacciones de consumo. compras por Internet, por mencionar sólo algunas. Esto dará lugar a nuevas oportunidades para los reguladores, particularmente en Europa con el lanzamiento del Mecanismo Único de Supervisión, pero también a nivel más global con la recopilación de datos liderada por el Consejo de Estabilidad Financiera sobre bancos de importancia sistémica global (G-SIB) y aseguradoras (G-SIB). SII), con una creciente demanda de análisis de datos a nivel micro sobre contrapartes e instrumentos. (Andersen, 2016)

El ámbito de la supervisión financiera (principalmente de entidades bancarias para aquellos bancos centrales encargados de su supervisión, pero también para otras instituciones financieras no bancarias cuando corresponda) ha experimentado un notable aumento en el uso de aprendizaje automático, especialmente para habilitar la supervisión tecnológica (*suptech*), es decir, el uso de nuevas tecnologías y análisis de grandes datos para respaldar la supervisión. Estas técnicas pueden mejorar la eficiencia de los supervisores en: (i) llevar a cabo tareas de supervisión tradicionales (por ejemplo, informes de calidad, detección de anomalías, envío de instrucciones); (ii) facilitar la evaluación de

fragilidades a nivel micro; y (iii) identificar y abordar nuevos temas emergentes, como los riesgos financieros relacionados con el clima, las vulnerabilidades derivadas de la pandemia de Covid-19 o las consecuencias de la creciente digitalización en las finanzas (por ejemplo, el desarrollo de las *Fintech*).

Por un lado, el flujo de información supervisora comienza con la presentación de registros individuales por parte de las empresas monitoreadas a las autoridades. En cuanto a los ejercicios estadísticos macroeconómicos, el aprendizaje automático puede hacer que esta presentación sea más eficiente al fortalecer la calidad de los datos en cuestión. Por ejemplo, el Banco de España ha desarrollado una herramienta impulsada por aprendizaje automático que imputa información faltante y también detecta valores atípicos en los estados financieros de empresas no financieras. Paralelamente, los valores faltantes se imputaron mediante análisis de regresión, de modo que la información ausente para una variable del informe de una empresa podría estimarse a partir de los valores de sus otras variables. En general, este proyecto resaltó la importancia de seleccionar las características apropiadas del modelo que se está considerando, teniendo en cuenta adecuadamente el conocimiento experto del dominio y considerando el impacto de los costos de cómputo durante la fase de entrenamiento.

Según Araujo et al. (2022), en una línea similar, el Banco de Canadá ha desarrollado un método novedoso basado en aprendizaje automático para detectar anomalías en los datos reportados por instituciones financieras. El objetivo era mejorar la eficiencia y calidad del proceso existente que maneja millones de puntos de datos por mes y que puede ser sensible para apoyar la política económica. El proyecto se basó en un procedimiento de dos pasos, donde las instituciones financieras designadas como "similares" se agruparon en un grupo y luego se analizaron conjuntamente utilizando un algoritmo de aprendizaje automático supervisado. Al superar el enfoque tradicional basado en reglas seguido anteriormente, el nuevo método ayudó a detectar anomalías que no se habían encontrado antes, al tiempo que ahorra tiempo significativo. Además, los procedimientos se diseñaron para ser escalables, completamente explicables y capaces de ejecutarse tanto en un entorno en la nube como en un lago de datos propietario.

Por otro lado, la comunicación supervisora (una subtarea del área de supervisión) también incluye lo que se denomina redacción de cartas supervisoras. Este proceso suele ser laborioso y requiere habilidades analíticas y de comunicación avanzadas. Además, mantener la coherencia en los diversos mensajes preparados por todo el equipo de supervisión y transmitidos a un gran número de empresas puede ser desafiante y engorroso. Ante estos problemas, el Banco Central de Malasia ha desarrollado una herramienta supotech que apoya la comunicación con las entidades supervisadas, con el objetivo de mejorar tanto la eficiencia del proceso como la coherencia de los mensajes transmitidos. Esta herramienta contaba con dos funcionalidades principales que se complementaban entre sí: Análisis de Tono y Búsqueda de Oraciones. Por último, sostiene el proyecto ilustró bien el flujo de trabajo completo que respalda la implementación de una solución basada en aprendizaje automático.

5.4.1 Supervisión Micro

La evaluación de las debilidades a nivel micro es otra parte fundamental de la supervisión donde, dada la gran cantidad y complejidad de datos, el uso de aprendizaje automático ha demostrado, mencionan los autores, ser especialmente adecuado. Por ejemplo, en un estudio realizado por el Banco de Japón se destacó cómo las técnicas de aprendizaje automático pueden mejorar la calidad de los datos empresariales utilizados en la supervisión. Este estudio comparó las predicciones realizadas por analistas profesionales de una agencia de crédito japonesa sobre el riesgo de que una empresa salga del mercado, con las predicciones hechas por algoritmos de aprendizaje automático, específicamente un bosque aleatorio. Se encontró que, en general, los algoritmos superaron las predicciones de los expertos, aunque estos últimos podrían tener un mejor desempeño al evaluar empresas con menos datos disponibles. Esto sugiere que las técnicas de aprendizaje automático no deben reemplazar por completo el juicio experto, sino utilizarse en conjunto.

Además, las técnicas de aprendizaje automático pueden ser empleadas para potenciar los modelos utilizados con el objetivo de asegurar la estabilidad financiera al integrar fuentes adicionales de información. Por ejemplo, el Banco Central de Rusia ha creado algoritmos de aprendizaje automático (como la combinación de regresión logística con bosques aleatorios) que incorporan datos adicionales sobre los pagos diarios para mejorar los modelos tradicionales de probabilidad de incumplimiento que son fundamentales en los ejercicios de supervisión financiera y que suelen basarse en datos contables a nivel empresarial. En este contexto, también se observó que es crucial analizar detenidamente el grado de precisión de las respectivas técnicas, prestando especial atención a las diferencias entre los sectores económicos identificadas.

5.4.2 Entidades no supervisadas

Aunque la supervisión micro se enfoca en las empresas específicas que deben ser monitoreadas por las autoridades, es igualmente importante considerar otros sectores menos regulados para evitar el arbitraje regulatorio, que ocurre cuando empresas no reguladas compiten en la prestación de servicios similares a las ofrecidas por entidades reguladas. En este sentido, el uso de técnicas de aprendizaje automático puede proporcionar a los supervisores ideas útiles y ayudarles a aplicar la "proporcionalidad" al considerar nuevos participantes en el sistema financiero. Un ejemplo es la identificación de entidades involucradas en *Fintech*, definido como la innovación tecnológica utilizada para respaldar o proporcionar servicios financieros. El banco de Francia y el banco alemán han desarrollado herramientas basadas en aprendizaje automático para identificar y monitorear estas entidades, superando la falta de información suficiente debido a su rápido desarrollo.

5.4.3 Políticas de estabilidad macro financieras

Independientemente de si el banco central tiene la responsabilidad de supervisar o no las actividades microfinancieras, uno de sus principales objetivos políticos se relaciona con la dimensión macro de la estabilidad financiera. Como mencionamos previamente el impacto de la crisis financiera global ha generado un mayor interés en desarrollar un enfoque integral para monitorear los riesgos financieros, que se centra en la situación de diversas instituciones simultáneamente en un momento dado y en cómo estos riesgos evolucionan a lo largo del tiempo con el ciclo financiero. Esta doble perspectiva requiere la recopilación y el análisis de grandes cantidades de datos que abarcan una amplia variedad de empresas y períodos prolongados. Por lo tanto, no es sorprendente que la función de estabilidad financiera de los bancos centrales pueda beneficiarse del aumento en el uso de aprendizaje automático. Por ejemplo, Fouliard et al. (2021) detallan cómo puede mejorar la capacidad para prevenir crisis financieras mucho antes de que ocurran. Dos aspectos importantes que merecen ser destacados desde esta perspectiva son: (i) el respaldo del aprendizaje automático para integrar información de diversas fuentes, a menudo no relacionadas, lo que puede ayudar a identificar vulnerabilidades a nivel del sistema y su evolución en el tiempo; y (ii) la capacidad para respaldar otras tareas de política que también son relevantes desde una perspectiva de estabilidad financiera.

A su vez, apoyar la función macro financiera implica la recolección de datos confiables de diversos sectores económicos, lo que enfatiza la importancia de procesos sólidos de garantía de calidad para manejar bases de datos que no son producidas directamente por el banco central. Un ejemplo de esto es la experiencia del Banco de Portugal en el uso de información del registro crediticio portugués. Esta fuente de datos se caracteriza por su alta granularidad, lo que resulta en un gran número de observaciones complejas (más de 200 atributos) y requiere controles estrictos de calidad de datos para detectar anomalías e identificar cambios sutiles. Para abordar posibles anomalías, se han implementado dos filtros automáticos. El primero es una prueba de consistencia de informes, que evalúa si todos los instrumentos financieros se informan de manera coherente hasta su vencimiento. El segundo es una

verificación de concentración, que examina la coherencia de la presentación de variables categóricas a nivel de agente. La detección de anomalías se basa en un algoritmo de bosque de aislamiento, lo que facilita la identificación de lagunas en la presentación de informes, patrones de datos inusuales y rupturas estructurales, mejorando así la calidad de la información disponible para apoyar los análisis macro prudenciales.

Además, la amplia escala de los conjuntos de datos detallados de interés para los responsables de políticas respalda las iniciativas basadas en aprendizaje automático para desarrollar un marco más estructurado. Esto se debe a que, incluso aparte de los problemas de calidad de datos, las limitaciones analíticas y computacionales pueden obstaculizar el uso efectivo de estos datos para monitorear la estabilidad financiera. Para abordar este desafío, el Banco Central Europeo (BCE), Deloitte y Google han colaborado en el desarrollo de una solución que consiste en una red dinámica de múltiples capas. Esta herramienta ayuda a los supervisores a analizar las estadísticas disponibles de manera integral y a realizar diversas operaciones respaldadas por herramientas de ciencia de datos, como la agregación, el filtrado y el análisis detallado desde el nivel de datos individuales. Los autores sostienen que esta solución ha facilitado las tareas de monitoreo de la estabilidad financiera frente a eventos de riesgo sistémico, como la agitación causada por la pandemia de Covid-19 en los mercados financieros en marzo de 2020.

Estas estrategias pueden ser fundamentales para analizar las interconexiones observadas en distintos momentos en los diversos segmentos del sistema financiero, lo cual es de gran interés para las autoridades macro-prudenciales. Por ejemplo, un proveedor de servicios australiano, Quant Property Solutions, analizó las relaciones entre los sectores bancario y de vivienda. Un mecanismo relevante en este contexto es que los prestamistas que ejecutan hipotecas sobre bienes raíces pueden influir significativamente en los precios de la vivienda, lo que puede tener graves implicaciones para la estabilidad financiera debido a la información limitada disponible. Las técnicas de aprendizaje automático se utilizaron para apoyar el proceso de descubrimiento de precios de mercado, ayudando a comprender las contribuciones de múltiples factores (como la situación del banco que vende una propiedad, características geo específicas del mercado inmobiliario y perspectivas económicas).

En cuanto a la dimensión temporal del riesgo sistémico, los métodos de aprendizaje automático pueden ayudar a manejar grandes y complejos conjuntos de datos que cambian con el tiempo. Por ejemplo, el Banco Central de la Federación Rusa (CBRF) ha adoptado este enfoque para facilitar el trabajo con bases de datos a nivel micro que involucran préstamos bancarios, lo que ha resultado en beneficios significativos en términos de garantía de calidad de datos, escalabilidad y automatización de operaciones, y una mayor interpretación de resultados. Además, esta solución también permitió la integración de la base de datos en cuestión con otras fuentes de información, como los servicios fiscales federales y las estadísticas estatales.

5.4.4 Regulación antilavado de dinero y financiamiento del terrorismo

Según el trabajo de Coelho et al. (2019) el conjunto de herramientas avanzadas de recolección y análisis de datos utilizadas por las autoridades financieras gracias a las nuevas tecnologías se conoce como *suptech*. Estos avances en inteligencia artificial, especialmente en aprendizaje automático, procesamiento de lenguaje natural y otras capacidades analíticas avanzadas, han facilitado el desarrollo de tales herramientas en el área de análisis de datos. Estas herramientas han abierto oportunidades para mejorar la capacidad de supervisión y regulación las autoridades financieras.

Tal como expresan en su trabajo, la detección de posibles infracciones de lavado de dinero (LD) y la lucha contra el financiamiento del terrorismo (FT) es un campo donde las herramientas de análisis de datos parecen estar más avanzadas. Las funciones de supervisión de LD/FT y de inteligencia financiera tienen mandatos distintos. Las autoridades encargadas de la supervisión de LD deben garantizar el cumplimiento de las instituciones financieras con los requisitos para combatir el lavado de dinero y el financiamiento del terrorismo, mientras que las unidades de inteligencia financiera (UIF) se

encargan de recibir y analizar informes de transacciones sospechosas y otra información relevante para el lavado de dinero, así como de difundir los resultados de dichos análisis. En ocasiones, las UIF también pueden tener funciones de supervisión de LD.

Tanto los supervisores de LD/FT como las UIF requieren herramientas avanzadas de análisis de datos para examinar los grandes volúmenes de información disponibles. Las autoridades de LD/FT suelen recibir cantidades considerables de datos transaccionales y no transaccionales, y algunas están ampliando su alcance colaborando con otras entidades gubernamentales y privadas. Además de las fuentes de datos tradicionales, algunas autoridades están explorando el uso de fuentes de información no convencionales, como artículos de periódicos y redes sociales, para obtener análisis más completos.

Los autores también remarcan que, a pesar de las diferencias en los mandatos, las autoridades de LD/FT tienden a buscar herramientas similares de análisis de datos, como análisis de redes, procesamiento de lenguaje natural, minería de texto y aprendizaje automático. Estas herramientas mejoran su capacidad para detectar redes de transacciones relacionadas, identificar comportamientos inusuales y convertir grandes cantidades de datos en información útil.

Es decir, los supervisores de LD/FT y las UIF utilizan herramientas analíticas de datos innovadoras similares para alcanzar sus objetivos. Esto incluye el análisis de redes, el aprendizaje automático, el procesamiento de lenguaje natural y la minería de texto. Aunque con objetivos diferentes, ambos tipos de autoridades utilizan estas herramientas para aumentar su capacidad para detectar redes de transacciones relacionadas, identificar comportamientos inusuales y, en general, transformar cantidades significativas de datos estructurados y no estructurados en información útil que contribuye a sus respectivos procesos.

Las analíticas avanzadas empleadas por los supervisores de LD/FT están principalmente orientadas a mejorar la evaluación de cumplimiento de las instituciones financieras individuales. El mandato habitual de las autoridades supervisoras de LD/FT es supervisar o monitorear las instituciones financieras y asegurar su cumplimiento con los requisitos para combatir el LD/FT. Para hacerlo, tales autoridades deben evaluar la gobernanza, la gestión de riesgos, los controles internos, los procesos y sistemas implementados en las instituciones financieras individuales para asegurarse de que tengan la capacidad suficiente para disuadir el flujo de fondos ilícitos. Esta evaluación de riesgos de integridad generalmente resulta de una combinación de actividades en el lugar y fuera del lugar. Las herramientas analíticas avanzadas desarrolladas y/o utilizadas por las autoridades supervisoras de LD/FT buscan mejorar la evaluación fuera del lugar del perfil de riesgo de las instituciones individuales, lo que también contribuye a la priorización basada en riesgos de las inspecciones en el lugar.

Otro aspecto muy relevante para los reguladores es la supervisión de sistemas de pagos. Los sistemas de pagos que funcionan correctamente son fundamentales para la estabilidad del sistema financiero, sin embargo, la vasta cantidad de datos de transacciones, a menudo con una distribución altamente sesgada, plantea desafíos para distinguir transacciones anómalas de las regulares. Identificar correctamente los pagos anómalos es crucial para abordar problemas como posibles quiebras bancarias, ciberataques o delitos financieros de manera oportuna. El lavado de dinero, en particular, socava la integridad y seguridad del sistema financiero global.

El Proyecto Aurora del BIS Innovation Hub utiliza datos de lavado de dinero sintéticos para comparar la identificación de pagos fraudulentos por varios modelos tradicionales y de aprendizaje automático (BISIH, 2023). Los modelos, que incluyen bosques de aislamiento y redes neuronales, se entrenan con transacciones de lavado de dinero conocidas (sintéticas) y luego predicen la probabilidad de lavado de dinero en datos no vistos. Según los autores, los modelos de aprendizaje automático superan a los métodos basados en reglas predominantes en la mayoría de las jurisdicciones o a las regresiones logísticas tradicionales. Las redes neuronales gráficas, que toman relaciones de pago como entrada, identifican redes de transacciones sospechosas particularmente bien. Estos modelos pueden funcionar eficazmente incluso con el agrupamiento de datos que garantiza la confidencialidad, lo que sugiere que la cooperación para analizar conjuntamente múltiples bases de datos puede ser segura y beneficiosa. Esto ilustra el potencial de una mayor cooperación entre las autoridades.

Otro enfoque para supervisar las transacciones de pago implica el uso de métodos de aprendizaje no supervisado para identificar automáticamente transacciones que merecen una inspección más detenida. Por ejemplo, los modelos de autoencoders, redes neuronales donde tanto las capas de entrada como de salida observan los mismos datos, distinguen pagos típicos de anómalos y pueden detectar dinámicas no lineales como las corridas bancarias. En simulaciones, estos modelos identificaron eficazmente, sostienen, patrones de retiros significativos de depósitos bancarios durante varios días. Los autoencoders también identificaron una variedad de anomalías reales en los sistemas de pago, incluidas las interrupciones operativas entre bancos nacionales importantes.

5.4.4.1 Herramientas usadas por los supervisores de Antilavado de Dinero y Financiamiento del Terrorismo

El análisis avanzado de datos, sostienen los autores, en primer lugar, puede emplearse para evaluar el riesgo general de las instituciones financieras, asignándoles una calificación según su probabilidad de no cumplir con los requisitos de LD/FT. Por ejemplo, el Centro Canadiense de Análisis de Transacciones e Informes Financieros (CCATIF) ha creado un modelo heurístico que considera diversos factores de riesgo relacionados con el perfil de la institución, su historial de cumplimiento, su comportamiento de reporte y otra información de inteligencia. Estos factores se determinan mediante métodos como el análisis de componentes principales y el análisis geoespacial (por ejemplo, su proximidad a fronteras) y se basan en una amplia cantidad de datos recibidos de entidades reportantes y otras fuentes. Basándose en estos factores de riesgo, el modelo clasifica a las instituciones según su probabilidad de incumplir con las regulaciones de LD/FT. Además, cada año se elabora un modelo de desafío utilizando una combinación de aprendizaje automático y métodos heurísticos para comparar el rendimiento del modelo en uso y actualizarlo si es necesario.

A su vez, varias entidades supervisoras emplean el análisis de redes como una herramienta para examinar cómo las instituciones bajo su supervisión están expuestas al riesgo de lavado de dinero. La creación de estas redes no es un fin en sí mismo, sino un método para evaluar el grado de cumplimiento de las instituciones financieras con las regulaciones de prevención del lavado de dinero y financiamiento del terrorismo. Los resultados de estos análisis suelen utilizarse para priorizar las inspecciones en el terreno y, en ocasiones, se comparten con unidades de inteligencia financiera y fuerzas del orden para respaldar sus operaciones. Por ejemplo, el Banco Central de Países Bajos ha implementado una herramienta para evaluar cómo las instituciones financieras están expuestas a redes de transacciones sospechosas.

Las conexiones también pueden ser establecidas utilizando información no organizada. La Autoridad Monetaria de Singapur (AMS) emplea análisis de redes en los STRs. En la actualidad, este análisis se lleva a cabo en los campos estructurados de los STRs. AMS está investigando la viabilidad de expandir su solución actual de análisis de redes para incluir datos extraídos del texto no estructurado de los STRs. Esta expansión puede lograrse mediante la aplicación de PLN para extraer detalles como las relaciones entre las partes mencionadas en el texto no estructurado de un STR individual. De manera similar, Comisión Nacional Bancaria y de Valores (CNBV) está desarrollando una aplicación piloto que examina noticias que mencionan a personas y empresas específicas implicadas en el lavado de dinero, y luego busca estas entidades en diversas fuentes de datos adicionales (como redes sociales y datos transaccionales) para establecer conexiones entre personas o empresas que podrían estar involucradas en actividades de lavado de dinero.

También, las tecnologías innovadoras mejoran la capacidad de las autoridades para manejar grandes volúmenes de datos sin estructura. En el banco central neerlandés (BCN), se está desarrollando una herramienta con este propósito, la cual busca optimizar el análisis de los informes de análisis de riesgo de integridad sistémica (ARIS). Estos informes son autoevaluaciones presentadas por los bancos sobre sus riesgos de integridad. La herramienta permite a los supervisores formular preguntas estándar y luego buscar información en una amplia cantidad de informes (por ejemplo, número de clientes de

alto riesgo) para obtener respuestas. Su objetivo es determinar si el banco tiene un entendimiento suficiente de los riesgos a los que está expuesto y si cuenta con controles internos adecuados para gestionarlos. Para extraer las respuestas, la herramienta empleará una combinación de minería de texto y aprendizaje automático supervisado. Al mismo tiempo, CNBV está trabajando en el desarrollo de una herramienta para detectar transacciones inusuales no reportadas.

5.4.4.2 Herramientas usadas por las Unidades de Inteligencia Financiera

Por su lado, diversas UIF están empleando el análisis de redes para detectar actividades delictivas. Para lograrlo, estas UIF se basan principalmente en la base de datos de STR, aunque también pueden aprovechar otras fuentes de datos disponibles. Al utilizar esta información, las UIF pueden fusionar los datos proporcionados por distintas entidades reportantes en distintos periodos de tiempo para identificar conexiones ocultas y establecer la red. Por ejemplo, UIF ha creado una herramienta que genera redes utilizando los STR. Además, UIF está explorando una herramienta que aplicaría una combinación de análisis de redes y técnicas de mapas autoorganizados en su base de datos de declaraciones de oro. De manera similar, la Agencia Federal de Monitoreo Financiero (ROSFIN) está desarrollando una herramienta que utilizará STR y otra información disponible, como contratos públicos, declaraciones aduaneras y datos sobre directores o fundadores de empresas, para construir las redes CCATIF emplea STR, transferencias bancarias y transacciones en efectivo grandes para identificar redes criminales potenciales en Canadá. Por otro lado, el regulador de inteligencia financiera (AUSTRAC) ha estado utilizando el análisis de redes en su base de datos de remesas para detectar actividades sospechosas que traspasan fronteras.

Además, se pueden emplear herramientas innovadoras para evaluar la probabilidad de actividad de lavado de dinero en transacciones individuales. En Finlandia, RAP ha creado una herramienta que otorga una puntuación a cada (STR) utilizando diversos criterios, como la frecuencia con la que la parte ha sido reportada en los últimos seis meses o si la parte tiene antecedentes criminales o está siendo investigada por la UIF. Por otro lado, ROSFIN ha desarrollado una herramienta que emplea el aprendizaje automático para identificar empresas posiblemente ficticias, basándose en una muestra de empresas ficticias ya conocidas. UIF está empleando big data para monitorear transferencias bancarias hacia y desde países específicos. Al combinar datos estructurados y, en menor medida, datos no estructurados (como artículos de prensa), la herramienta puede calcular indicadores que ayudan a evaluar el nivel de anormalidad de cada flujo.

5.4.4.3 Identificación de patrones y tendencias en actividades delictivas

Las UIF también recurren a herramientas avanzadas de análisis de datos para detectar nuevas tendencias y patrones en el lavado de dinero. Dado que los métodos de lavado de dinero están en constante evolución, es crucial que las autoridades cuenten con herramientas que les permitan adaptarse rápidamente a este entorno dinámico. Identificar de manera temprana nuevas tipologías permite a las UIF compartir esta información con los participantes del mercado y otras autoridades, además de respaldar sus propias investigaciones.

Los mecanismos de clasificación automática son útiles para las autoridades en la identificación de nuevas tendencias y en el seguimiento de la evolución de las tipologías de lavado de dinero a lo largo del tiempo y en áreas geográficas específicas. Por ejemplo, en la actualidad, la UIF está desarrollando una herramienta para clasificar STR según el tipo de esquema de lavado de dinero que se haya perpetrado. En primer lugar, se extrae información de la parte no estructurada del STR mediante el PLN. Luego, se emplea una técnica de aprendizaje automático supervisado para categorizar los STR según la tipología identificada. De manera similar, FINTRAC está trabajando en una herramienta de minería de texto para clasificar los STR.

En su trabajo Guirelli et al. (2021) mencionan que el enfoque conocido como Latent Dirichlet Allocation o LDA puede emplearse para identificar temas en conjuntos de datos de texto. Se trata de un método de aprendizaje no supervisado, lo que implica que los datos no requieren etiquetas de tema y que la definición de los temas no es determinada por el modelador, sino que es el resultado de aplicar el modelo sobre los datos. Una característica atractiva de este método es que está fundamentado en un marco estadístico: supone que los documentos se generan conforme a un proceso estadístico generativo (la distribución de Dirichlet), de modo que cada documento puede ser descrito por una distribución de temas y cada tema puede ser descrito por una distribución de palabras. Estos temas son latentes, es decir, no son observados directamente, a diferencia de los documentos y las palabras contenidas en cada documento.

El primer paso consiste en construir un corpus con los datos de texto. En este caso, se trata de una amplia base de datos con más de 780.000 observaciones que incluye todas las piezas de noticias publicadas por El Mundo (un periódico español) entre 1997 y 2018, obtenidas del repositorio de prensa española de Dow Jones. Posteriormente, estos datos de texto deben ser procesados y limpiados para obtener una versión del corpus que carezca de puntuación, números o caracteres especiales y que esté completamente en minúsculas, además de excluir las palabras más comunes (como artículos y conjunciones). Luego, estos datos son alimentados a un *stemmer* específico del lenguaje, que elimina las variaciones de las palabras (por ejemplo, tiempos verbales) y las reduce a su forma básica o raíz, lo que se utiliza para crear una representación de "bolsa de palabras" del corpus: una tabla grande con una fila para cada pieza de noticia y una columna para cada posible palabra reducida, rellena con números que indican cuántas veces aparece cada palabra en cada pieza de noticia (es importante notar que esta matriz será muy dispersa, ya que la mayoría de las palabras de un diccionario extenso no aparecerán en la mayoría de las piezas de noticia).

Esta representación de "bolsa de palabras" del corpus se utiliza después como entrada para el algoritmo LDA, que se encarga de identificar 128 temas distintos discutidos en estos textos, y asignar a cada pieza de noticia una estimación de la probabilidad de pertenecer a cada uno de esos temas. El algoritmo analiza los textos y determina qué palabras tienden a aparecer juntas y cuáles no, asignándolas de manera óptima a diferentes temas con el objetivo de minimizar la distancia entre los textos asignados a cualquier tema dado, y maximizar la distancia entre textos asignados a diferentes temas.

El resultado es una base de datos que contiene, para cada trimestre desde 1997 hasta 2018, el porcentaje de piezas de noticias que están asociadas con cada uno de los 128 temas identificados por el modelo de aprendizaje no supervisado. Además, se aplica un diccionario de términos positivos y negativos a cada pieza de noticia, y los resultados se agregan en series trimestrales que indican qué tan positivas o negativas son las piezas de noticias relacionadas con cada tema.

Por otro lado, para Danielsson y Uthemann (2023) la IA sobresale y supera a los humanos en el modelado y la gestión de riesgos. Sostienen que está avanzando rápidamente en la detección de fraudes, la protección del consumidor y otras malas conductas, pero su uso a veces se ve frustrado por los silos de datos. Sin embargo, argumentan que existen soluciones tecnológicas que pueden superar tales limitaciones. La IA beneficia a este tipo de aplicaciones porque hay muchos datos con los que entrenar. Las decisiones de los supervisores humanos se incorporan a los algoritmos de aprendizaje por refuerzo de la que la IA hace tan buen uso. Los objetivos que la IA debe cumplir son claros e inmutables a lo largo del tiempo en que opera, y el costo de los errores es contenido y fácil de abordar.

Sin embargo, la IA también puede amplificar los micro problemas existentes, como los que surgen del sesgo algorítmico, tal vez con criterios raciales para las decisiones crediticias, pero una mejor capacitación debería aliviar la mayoría de esas preocupaciones. Varios factores frustran el uso de la IA con fines macroeconómicos y, lo que es peor, pueden hacer que engañe a los formuladores de políticas e incluso desestabilice el sistema financiero. Los datos son limitados y pueden ser engañosos a medida que el sistema financiero experimenta cambios estructurales continuos. Monitorear las vulnerabilidades del sistema y controlar los riesgos es difícil porque los factores de inestabilidad sólo surgen en tiempos de crisis. Los actores económicos amplifican endógenamente el estrés y cambian su comportamiento en respuesta a los intentos regulatorios de control. Además, el mayor uso de la IA, particularmente en la gestión de riesgos, puede potencialmente aumentar la prociclicidad y, por ende, el riesgo sistémico.

Será mejor encontrar técnicas de medición y gestión de última generación, que luego serían similares en todo el sistema, armonizando creencias y acciones e induciendo prociclicidad. Esto sólo se verá amplificado por una mayor subcontratación de la gestión de riesgos a un pequeño conjunto de instituciones con tecnologías y datos superiores.

6. Desafíos y limitaciones

A lo largo del trabajo hemos ido recorriendo los usos de big data por parte de los reguladores financieros haciendo foco en el impacto positivo que tienen las distintas herramientas para la gestión de riesgos, prevención del fraude, garantizar la estabilidad financiera e investigación económica y predicción. En este apartado abordaremos la contracara de lo que venimos discutiendo, resaltando cuáles son las limitaciones que tienen estas herramientas hoy en día y, por lo tanto, los desafíos que afrontan.

Habiendo remarcado la importancia de big data para la regulación financiera internacional, es necesario también destacar cuáles son los desafíos y obstáculos que aún persisten. Para empezar, hay mucha discusión sobre la calidad de los datos tratados, en cuanto a sesgo y representatividad. Por ejemplo, para el caso de las técnicas aplicadas en la red social Twitter, el trabajo de Nymand-Andersen (2016) menciona que no hay que dejar de considerar que no necesariamente todos los tweets representan al total de la población, ya que seguramente haya gente excluida de esa muestra. Es necesario comprender cómo y qué tipos de ajustes se deben realizar sobre la muestra de acuerdo con el contexto de la situación. Por otro lado, otro desafío a analizar es la constancia y solidez de los resultados obtenidos a lo largo del tiempo.

Por su parte, Wibisono et al. (2019) también sostiene que otra inquietud surge de que, dado que las muestras de big data suelen no ser representativas (por ejemplo, no todos los individuos están presentes en Facebook, y aún menos en Twitter), es posible que su fiabilidad no sea tan sólida como se podría suponer. Además, existe el peligro de que la recolección y el tratamiento de grandes volúmenes de datos se vean obstaculizados por las leyes de privacidad y/o cambios en la composición de los participantes del mercado. Por ende, resulta crucial que las autoridades pertinentes coordinen sus acciones para aprovechar los beneficios del análisis de big data sin poner en riesgo la privacidad ni la confidencialidad de la información.

Otros de los desafíos son los costos de almacenamiento y gestión de la infraestructura requerida. En este sentido, el trabajo Ghirelli et al. (2021) expone que la disponibilidad de gran cantidad de información plantea importantes desafíos en términos de gestión, capacidad y costes de almacenamiento, y seguridad y confidencialidad de la infraestructura requerida. Además, la gestión óptima de enormes conjuntos de datos estructurados y no estructurados requieren la integración de nuevos profesionales perfiles (científicos de datos e ingenieros de datos) en los bancos centrales y transmite la necesidad de una transformación digital en toda regla de estas instituciones. Además, la diversa naturaleza de las nuevas fuentes de información requiere la asimilación y desarrollo de técnicas que transformen y sintetizen datos, en formatos que puedan incorporarse a análisis económicos. Por ejemplo, las técnicas de análisis textual permiten que la información contenida en el texto para ser procesados y convertidos en datos estructurados, como en Google Tendencias, bases de datos de medios en línea, redes sociales (por ejemplo, Facebook y Twitter), web portales de búsqueda (por ejemplo, portales creados para la búsqueda de vivienda o empleo), datos de telefonía móvil, o datos satelitales, entre otros.

Por su parte, el trabajo de IFC (2020) sostiene que, si bien muchas instituciones han emprendido importantes iniciativas para desarrollar plataformas adecuadas que faciliten el almacenamiento y procesamiento de conjuntos de datos muy grandes y complejos, el progreso ha sido variado. Esto se debe en parte a la necesidad de contratar y capacitar personal, lo cual es difícil debido a la oferta limitada de candidatos con las habilidades necesarias (por ejemplo, científicos de datos). Por otro lado, los desafíos incluyen la base legal para el uso de datos privados y la seguridad, ética y las preocupaciones

de privacidad que esto implica, así como la “imparcialidad” y precisión de los algoritmos capacitados en conjuntos de datos preclasificados y/o no representativos. La calidad de datos y los problemas de gobernanza también son importantes, ya que gran parte de los nuevos macrodatos recopilados como subproducto de actividades económicas o sociales debe ser curado antes de se pueden realizar análisis estadísticos. Estos desafíos son en general considerados igualmente importantes entre los diferentes bancos centrales de todo el mundo.

El trabajo de IFC (2021), por su parte, también menciona que además de la infraestructura de Tecnologías de la Información (TI), hay muchos otros desafíos como la base legal para el uso de información privada y la preocupación sobre la ética y la justicia en la precisión de los algoritmos entrenados en base a datos preclasificados y/o no representativos. A su vez, resaltan que las predicciones no solo deben ser exactas sino también interpretables y representativas. Para esto, la transparencia respecto de la información es esencial para comprobar su calidad y que las decisiones de políticas estén cimentadas sobre una base sólida.

Además de lo ya mencionado, la accesibilidad de los datos continúa siendo un obstáculo. En este sentido, Einav y Levin (2014) destacan los desafíos por delante al hablar de big data en su trabajo, ya que mencionan que los economistas que desean aprovechar grandes conjuntos de datos nuevos enfrentan varios desafíos. Estos incluyen obtener acceso a los datos, desarrollar la gestión de datos y capacidades de programación necesario trabajar con conjuntos de datos a gran escala y, finalmente, pensar con creatividad enfoques para resumir, describir y analizar la información contenida en estos datos. Una observación adicional es que, al trabajar con grandes y ricos conjuntos de datos, puede no ser trivial simplemente descubrir qué preguntas los datos podrían responder convincentemente. Mientras que en el pasado un investigador podía simplemente abrir sus datos en la pantalla y tener una idea visual de las características clave, grandes conjuntos de datos requieren tiempo y esfuerzo para tareas triviales conceptualmente, como extraer y resumir diferentes variables y explorar relaciones entre ellos.

Para Dicuonzo et al. (2019) el desafío para el sector bancario es importar tecnologías no convencionales (*Internet of Things*, inteligencia de negocios, *Big Data* y *blockchain*) que permitan procesar enormes cantidades de datos de forma rápida y precisa, lo cual es útil no sólo en gestión de relaciones con el cliente, sino también en ciberseguridad, detección de fraude y la optimización de los procesos de toma de decisiones. La innovación tecnológica debe ir acompañada de un cambio cultural y, en este sentido, la formación de equipos compuestos por expertos estadísticos, matemáticos y habilidades tecnológicas y científicos de datos que puedan combinar habilidades de análisis de datos con habilidades funcionales para crear automáticamente procesos de valor. Este cambio disruptivo en curso necesariamente afecta a las personas e involucra a toda la organización, incluyendo TI, *marketing*, control empresarial y de gestión, planificación estratégica y servicio al cliente, para obtener importantes ventajas en términos de análisis de riesgos, prevención de fraude y análisis avanzado de clientes inteligencia a través del almacenamiento, organización, gestión y uso operativo de grandes cantidades de datos; esto será parte de una evolución futura.

Doko y Miskovski (2019) destacan dentro de los desafíos a superar al desarrollo de tecnologías de la información por parte de los bancos, el desarrollo de infraestructura, privacidad de la información y costo de almacenamiento. En cuanto a la privacidad y seguridad el manejo de big data es más escalable y flexible en la nube, pero la privacidad y las normas de seguridad a menudo restringen esta decisión de movimiento. También hay problemas de reputación dado que ha despertado críticas por traspasar límites de privacidad. El análisis de big data está limitado por numerosos cumplimientos normativos en esos aspectos que impactan en los análisis que se derivan de él. Otro de los desafíos es el problema en el almacenamiento y procesamiento. Si bien los datos estructurados bancarios crecen continuamente, los datos no estructurados están creciendo más rápido y se están convirtiendo en una fuente más importante para los clientes.

Por su parte, Hammer et al. (2017) y, en línea con los trabajos que venimos detallando, advierten tres grandes desafíos. En primer lugar, se encuentra la calidad de los datos. Sostienen que, para la toma de decisiones políticas, la evaluación de la calidad de indicadores derivados de grandes datos es crucial para minimizar ciertos riesgos como el reputacional, gubernamental y político. También sostienen que

muchos tipos de big data no representan muestras aleatorias de población, así como los indicadores basados en big data tienen un período corto de tiempo y contienen valores atípicos y no es posible garantizar su continuidad.

Una segunda categoría de desafíos, sostienen los autores, es la del acceso a big data. Refieren en primer lugar que, el origen y la creación de los grandes datos están en su mayoría fuera de control de las instituciones nacionales o internacionales (con la excepción de los datos administrativos). En segundo lugar, sostienen que la privacidad, confidencialidad y riesgos de ciberseguridad con una preocupación importante al utilizar big data. En este sentido, los procedimientos de protección de la privacidad y las técnicas de las tecnologías de la información para minimizar riesgos son cruciales. Por último, mencionan que el precio de big data no es necesariamente bajo y que, a pesar de una multitud de opciones técnicas, los procesos de selección de estas suelen ser bastantes complejos.

En tercer lugar, otro de los grandes desafíos de big data son los perfiles profesionales con habilidades y tecnologías nuevas. Para crear valor a partir de big data será necesario, sostienen los autores, contar con un equipo multidisciplinario. Las agencias de estadísticas, los bancos centrales y las agencias públicas no sólo tendrán que capacitar al personal existente, sino que tendrán que competir con el sector privado para reclutar o contratar nuevo personal.

Según Kitchin (2015) un primer desafío es el acceso a la big data. Aunque algunos de los datos son producidos por agencias públicas, como los datos meteorológicos, ciertos datos sobre transporte y sistemas administrativos, gran parte de ellos son generados por empresas privadas y generalmente no están disponibles para su uso público. Para que los Institutos Nacionales de Estadística accedan a estos datos deben formar asociaciones estratégicas vinculantes con estas empresas o crear o modificar ciertos instrumentos legales (como leyes de estadística) con el objetivo de obligar a las empresas a proporcionar dichos datos. Esto necesariamente conllevo una negociación política y sobre todo tiempo.

Una vez logrado esto, es necesario evaluar su idoneidad para implementar, reemplazar o agregar a las estadísticas oficiales. En primer lugar, dado que como dijimos previamente, los datos son creados en su mayoría por empresas privadas para sus necesidades específicas, es decir, nunca fueron pensados para ser utilizados para la producción de estadísticas oficiales. Por lo tanto, es necesario establecer en qué medida estos datos reutilizados proporcionan sustitutos adecuados, confiables y rigurosos para datos muestreados y más específicos.

Por otro lado, mientras que los Institutos Nacionales de Estadística establecen de manera cuidadosa sus marcos y parámetros de muestreo, los big data generalmente no son representativos de toda una población, ya que solo captan a quienes utilizan un servicio. Es decir, hay un desafío de este tipo de datos en el marco de ciertas metodologías existentes. El autor ejemplifica mencionando el caso de las tarjetas de crédito donde solo se relacionan con las personas que poseen una tarjeta de crédito o de las redes sociales donde solo interactúan aquellos que las utilizan. También sucede con el Índice de Precios al Consumidor, donde es necesario seguir el mismo conjunto de bienes y servicios con pesos estadísticamente definidos a lo largo del tiempo, en lugar de simplemente recopilar datos de la web indefinidos.

Otro interrogante sobre la idoneidad de los big data para las estadísticas oficiales se ve en las redes sociales Twitter y Facebook donde constantemente se están ajustando sus diseños y modos de interacción. Por lo tanto, los datos creados son inconsistentes entre usuarios y el tiempo. Una vez establecida la idoneidad de los datos, el autor menciona la necesidad de realizar una prueba sobre la viabilidad tecnológica relacionada con la transferencia, almacenamiento, limpieza, verificación de los datos, así como también la integración de éstos con las estadísticas oficiales ya existentes. Menciona la falta de herramientas cómodas para el usuario que dificultan su uso y su integración a los flujos de trabajo actuales.

Un desafío clave también mencionado por este autor es la gestión del cambio. Big data requiere de la preparación de la organización para asumir nuevos roles y responsabilidades. Esto es, requiere del establecimiento de nuevos sistemas de ciclo de vida de los datos, seguidos de la construcción y mantenimiento de una determinada infraestructura de tecnología de la información que pueda manejar,

almacenar y procesar el conjunto de grandes datos. A su vez, no hay que olvidarse que éstos deben garantizar la seguridad de los datos y el cumplimiento de la privacidad de los datos.

Otro de los riesgos a los que se exponen los Institutos Nacionales de Estadística es que su reputación y credibilidad pública sean debilitadas. Estos constituyen activos fundamentales para estas entidades con lo cual asociarse con un tercero comercial y utilizar sus datos aumenta la exposición al riesgo. La reutilización de los datos, donde no hay consentimiento explícito por parte de quienes representan esos datos, genera un riesgo de reacción negativa y de resistencia pública a dicho uso. El caso de los escándalos de WikiLeaks y Snowden y otras violaciones de datos, muestran cómo la confianza pública sobre el manejo de datos personales ya ha sido minada. Y el panorama es más general, cuando se observa cómo el robo de tarjetas de crédito o información personal de empresas privadas también ha reducido la confianza pública en éstas.

Por último, el autor menciona el riesgo de que las estadísticas oficiales sean privatizadas por los gobiernos o su sostenibilidad dependan cada vez más para sobrevivir de la comercialización de sus bases de datos. Es decir, la amenaza de que el conjunto de datos siga abierto.

En IFC (2024) los autores mencionan un conflicto entre la precisión y la comprensibilidad en los modelos de aprendizaje automático. Aunque los modelos avanzados pueden lograr predicciones altamente precisas, su complejidad puede dificultar su interpretación, especialmente en cuanto a la importancia de las variables de entrada. Esto puede llevar a la aceptación de un modelo complejo como una "caja negra", lo que dificulta la evaluación de sesgos y la explicación del comportamiento del modelo de manera comprensible para los humanos. Este desafío se vuelve más pronunciado en los modelos de inteligencia artificial generativa, que pueden generar respuestas incorrectas como si fueran correctas. Este fenómeno subraya la importancia de la supervisión humana, especialmente en tareas que requieren razonamiento lógico.

Además, en este trabajo los autores sostienen que el uso de datos no estructurados y personales plantea desafíos importantes en términos de marcos legales y privacidad de datos. En el pasado, la mayoría de los datos estaban almacenados en instituciones públicas con procedimientos claros de acceso y aseguramiento de calidad. Sin embargo, sostienen, en la actualidad, una gran cantidad de datos son generados por individuos y empresas del sector privado, generalmente sin documentación pública suficiente. Esto genera inquietudes sobre la claridad de los marcos legales que regulan el uso de estos datos. Además, el proceso de entrenamiento o ajuste de modelos de lenguaje puede requerir grandes volúmenes de datos de diversas fuentes, lo que complica aún más los aspectos legales y de privacidad. La disponibilidad de datos personales no estructurados también plantea dilemas éticos y de privacidad, ya que los ciudadanos tienen derecho a la privacidad y pueden sentirse incómodos con el acceso de los bancos centrales a sus datos.

Otro factor por considerar como desafío tiene que ver con el de las inversiones que deben realizar los bancos centrales en cuanto a tecnologías de la información y en capital humano. Los autores remarcan los elevados costos iniciales que provoca esto, así como también el reto que implica contratar o retener al capital humano que reúna los atributos combinados de la comprensión económica y habilidades de programación. Sumado a esto, se encuentra la realidad que no pueden competir con los salarios del sector privado.

En cuanto al gran debate surgido por el avance de la IA y el futuro del trabajo, los autores resaltan que los sistemas de IA podrían actuar de copilotos para los equipos de supervisión humanos. La IA podría mejorar el análisis al liberar tiempo de los economistas para interpretar datos en cambio de recolectarlos y limpiarlos. De esta manera, afirman que la IA no hará arcaicos a los humanos, sino más bien, los comentarios de los expertos permitirán mejorar los modelos y mitigar problemas de alucinación de estos.

Finalmente, otro de los desafíos que observan los autores es el de la concentración de la oferta en pocas manos. El hecho de que los modelos fundamentales más poderosos sean proporcionados por unas pocas grandes empresas no solo plantea riesgos en términos de innovación y de dinamismo económico, sino que la concentración de recursos podría generar riesgos importantes para la estabilidad

financiera, operativa y reputacional. Por ejemplo, se estaría mucho más expuestos a fallas de tecnología de información o ciberataques dirigidos a estos proveedores. También, si muchas instituciones adoptan los mismos pocos algoritmos, su comportamiento en momentos de episodios de estrés podría ser muy similar entre ellas y generar eventos masivos indeseables como corridas bancarias, retención de liquidez, ventas masivas, etc.

Por último, Danielsson y Uthemann (2023) mencionan por su parte y hacen una distinción entre el uso de la IA por parte de privados y por parte de entidades públicas. Cada una de ellas tiene una implicancia diferente y estudian a nivel general cuáles son los riesgos a los que se está expuesto. Respecto del uso de la IA por privados afirman que puede derivar en un mayor riesgo sistémico causado por la prociclicidad.

6.1 Regulación de IA privada

Los autores afirman que los mercados financieros plantean desafíos únicos para la inteligencia artificial, tanto en el nivel micro como macroeconómico, debido a la complejidad de sus características. La convergencia de objetivos entre participantes y la disponibilidad de recursos considerables hacen que los mercados financieros sean especialmente susceptibles a comportamientos coordinados. Muchas decisiones de inversión exhiben fuertes complementariedades, donde los beneficios de una acción específica aumentan con el número de participantes del mercado que toman esa acción, lo que puede dar lugar a resultados no deseados como colusiones, pánicos bancarios, ventas forzadas y movimientos hacia activos seguros. Aunque, afirman, estos resultados han sido una característica constante de los mercados financieros desde mucho antes de la era de la tecnología informática, se ven reforzados por el uso de computadoras y algoritmos para analizar datos y tomar decisiones. El aumento en el uso de inteligencia artificial solo intensificará este proceso. Dada la uniformidad en los objetivos de las inteligencias artificiales del sector privado, junto con los datos con los que se entrenan y su capacidad superior para encontrar soluciones óptimas, se espera que aprovechen las complementariedades de manera más eficiente que sus homólogos humanos, potencialmente incrementando el estrés en el sistema financiero. El resultado puede ser un mayor riesgo sistémico causado por la prociclicidad.

Además, la inteligencia artificial también es útil para aquellos que buscan explotar el sistema financiero para obtener ganancias financieras ilegales y causar daños, como delincuentes, terroristas y Estados-nación hostiles. Esto se debe a que el problema computacional al que se enfrentan estas entidades es mucho más sencillo que el de las autoridades, ya que solo tienen que encontrar un área para explotar, mientras que las autoridades tienen que patrullar todo el sistema.

A su vez, remarcan que el mayor uso de la inteligencia artificial en el sector privado también crea problemas legales para las autoridades financieras porque genera otra capa de negación. Asignar responsabilidad legal por malas conductas en el sector financiero ya es complicado y será más difícil cuando se utilice la inteligencia artificial para tomar decisiones.

6.2 Regulación de IA pública

Por otro lado, surgen diversos problemas cuando las autoridades financieras emplean la IA para la regulación. Uno de ellos es que, a medida que aumenta el uso de la IA para las micro regulaciones, esta tiende a seleccionar las técnicas óptimas para gestionar el riesgo. Aunque este fenómeno es inherente a todos los diseños regulatorios, es probable que la IA intensifique el empleo de métodos similares de medición y gestión de riesgos. Esta tendencia genera preocupación porque cuantos más actores del mercado y reguladores perciban el entorno de manera uniforme, más probable es que las regulaciones induzcan a los participantes del mercado a reaccionar de manera similar ante los cambios

abruptos, lo que amplifica el ciclo financiero, creando auges y caídas: es lo que se conoce como prociclicidad.

Además, este enfoque uniforme en la gestión del riesgo aumenta la vulnerabilidad del sistema a las mismas incertidumbres, lo que eleva aún más el riesgo sistémico. Debido a los elevados costos fijos asociados con la modelización y la gestión de riesgos, el desarrollo de sistemas de gestión de riesgos presenta un rendimiento creciente a escala para las empresas, similar a la computación en la nube. Esto podría llevar a que la gestión de riesgos se concentre en unos pocos proveedores de IA, lo que incrementa la homogeneidad en las creencias y acciones, y amplifica la prociclicidad.

Otro aspecto para tener en cuenta en el uso de la IA en el sector público es la cuestión de la rendición de cuentas. Si un supervisor humano comete un error, se puede responsabilizar a esa persona, brindarle más formación o despedirlo. Sin embargo, no es tan sencillo hacer que la IA tome decisiones equivalentes. De esta manera, cuando se cuestionan las decisiones de la IA, es posible que la IA regulatoria no pueda explicar su razonamiento o por qué considera que cumple con las leyes y regulaciones. En última instancia, esto implica que la supervisión interna del uso de la IA en las agencias reguladoras y su interacción con el sistema legal externo requerirá políticas distintas a las que se emplean para los supervisores humanos actuales.

El uso de IA amplifica riesgos que podrían afectar la seguridad y solidez de una institución financiera, dada la falta de explicabilidad o interpretabilidad de los procesos de los modelos de IA, con potencial de prociclicidad y riesgo sistémico en los mercados. La dificultad para comprender cómo el modelo genera resultados podría crear posibles incompatibilidades con los marcos de supervisión financiera existentes y los marcos de gobierno interno, incluso desafiar el enfoque tecnológicamente neutral para la formulación de políticas. La IA puede presentar riesgos particulares de protección al consumidor, como riesgos de resultados sesgados, injustos o discriminatorios para el consumidor, o preocupaciones sobre la gestión y el uso de datos. Aunque muchos de los riesgos potenciales asociados con la IA en finanzas no son únicos de la IA, el uso de la IA podría amplificar tales vulnerabilidades dada la complejidad de las técnicas empleadas, la adaptabilidad dinámica de los modelos basados en IA y su nivel de autonomía para las aplicaciones de IA más avanzadas.

Por último, en el cuadro 2 del Anexo se detalla la explicación que dan los autores sobre las diferentes aristas de esta problemática que fue desplegada también en el capítulo 5 del presente trabajo. Las tareas referidas al ámbito micro de la regulación como prevención de fraude, protección al consumidor, resolución de falla de un banco pequeño presentan mejores condiciones para la utilización de la IA. Mientras que las tareas referidas al ámbito macro de la regulación de grandes quiebras bancarias, administración de crisis sistémicas presentan peores condiciones para que la implementación de la IA por parte de los reguladores pueda caracterizarse como positiva.

Por su parte, el trabajo de McKinsey & Company (2024 b) sintetiza cuatro fuentes principales de riesgo entrante por la adopción de la IA:

- Amenazas a la seguridad, resultantes del mayor volumen y sofisticación de los ataques de *malware* habilitado para IA de generación.
- Riesgo de terceros, resultante de los desafíos para comprender dónde y cómo terceros pueden estar implementando IA genérica, creando posibles exposiciones desconocidas.
- Uso malicioso, resultante de la posibilidad de que los malos actores creen *deepfakes* convincentes de representantes o marcas de la empresa que resulten en un daño significativo a la reputación.
- Infracción de propiedad intelectual (PI), como resultado de que la propiedad intelectual (como imágenes, música y texto) se introduzca en motores de entrenamiento para modelos de lenguaje grandes subyacentes y se haga accesible a cualquier persona que utilice la tecnología.

Sostienen que la mayoría de las organizaciones se verán favorecidas por un esfuerzo enfocado en investigar cómo la inteligencia artificial genérica está alterando su entorno externo, con dos objetivos

principales. El primero es comprender las posibles vulnerabilidades a riesgos emergentes, basadas en el perfil de riesgo de la organización (por ejemplo, el número de terceros con acceso a datos sensibles o confidenciales que deben limitarse en el entrenamiento de modelos de inteligencia artificial generados externamente). El segundo objetivo es entender la madurez y la preparación del entorno de control: las habilidades técnicas y no técnicas que la organización tiene en marcha para prevenir, detectar y, en última instancia, responder a los riesgos emergentes. Estas incluyen medidas de seguridad cibernética y antifraude, la debida diligencia de terceros para identificar dónde pueden estar implementando inteligencia artificial generativa terceros críticos, y la capacidad de restringir la extracción de propiedad intelectual de la empresa por parte de motores utilizados para entrenar grandes modelos de lenguaje.

7. Conclusiones

La comprensión de nuestra realidad actual surge del análisis del devenir de nuestra sociedad. En este sentido, el presente trabajo inició haciendo un recorrido histórico sobre las distintas revoluciones tecnológicas que se han sucedido hasta la actualidad y que nos permitió tener un marco general sobre nuestra etapa contemporánea. De esta manera, pudimos situar el surgimiento de big data y la importancia de su estudio, para las industrias en general y para las finanzas en particular.

En la sección 3 abordamos las distintas definiciones de big data, resaltando su carácter dinámico. También desplegamos los diferentes tipos de datos que conforman los datos masivos para luego, concentramos en una explicación general de las distintas técnicas de procesamiento de big data. Esto nos permitió comprender algunas de esas técnicas cuando vimos su aplicación práctica en la actualidad. Por último, definimos qué era big data para los bancos centrales.

En la sección 4 se buscó resaltar la importancia del uso de big data en la regulación financiera internacional. En ese marco, observamos que la necesidad de su uso se veía más visible en momentos de crisis financieras y eventos disruptivos como lo fue la pandemia provocada por el COVID-19. Además, se ha dejado en evidencia el consenso general por parte de la literatura acerca la importancia del uso del big data en la regulación financiera, destacando su potencial para mejorar la supervisión y detección de riesgos sistémicos, así como para fomentar la transparencia y la integridad del mercado. Estas capacidades son vitales en un entorno financiero globalizado y cada vez más complejo, donde los riesgos pueden surgir inesperadamente y propagarse rápidamente.

Posteriormente, en la sección 5, se buscó brindar un detalle sobre cuáles son las áreas de la regulación en las que se está utilizando el big data. En ese sentido, detectamos que el área donde más se utiliza es en la de investigación económica y predicción, dada la centralidad que tienen los modelos y procesamiento de datos no estructurados que cada vez están ganando más relevancia. En cuanto otras de las principales aplicaciones del big data en la regulación financiera, se ha identificado una amplia gama de usos, que van desde la vigilancia del mercado y el cumplimiento normativo hasta la evaluación del riesgo y la detección de fraudes. Estas aplicaciones ilustran el potencial transformador del Big Data en la forma en que se supervisan y regulan los mercados financieros a nivel internacional.

En la sección 6 abordamos los desafíos y obstáculos que hoy en día siguen teniendo las diferentes técnicas y los inconvenientes que generan a la hora de sacar conclusiones. Sin embargo, a pesar de sus ventajas, también se han señalado varios desafíos que deben superarse para aprovechar plenamente el potencial del big data en la regulación financiera. Estos desafíos incluyen aspectos relacionados con la privacidad y la protección de datos, la interoperabilidad de sistemas y la capacidad para gestionar y analizar grandes volúmenes de datos de manera eficiente. A su vez, también mencionamos el desafío de afrontar un mayor riesgo cibernético y de prociclicidad.

En resumen, esta revisión ha destacado el papel cada vez más importante que desempeña el big data en la regulación financiera internacional. Aunque existen desafíos significativos por delante, el potencial transformador del big data para mejorar la supervisión y la estabilidad del sistema financiero

justifica un mayor esfuerzo para superar estos obstáculos y aprovechar plenamente sus beneficios. En última instancia, el uso efectivo del big data en la regulación financiera tiene el potencial de promover un sistema financiero más seguro, eficiente y transparente a escala global.

8. Referencias bibliográficas

- Addo P. M., Guegan D., and Hassani B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *Risks*, N.6, 2 (38): 2227-9091.
- Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. N. (2020). *Can news help to measure economic sentiment? An application in Covid-19 times*. Banco de España.
- Algan, Y., Beasley, F., Guyot, K., Higa, F. y Senik, C. N.(2016), *Big Data Measures of Well-Being: Evidence from a Google Well-Being Index in the United States*, (Working Paper No 1605) CEPREMAP.
- Araujo, D., Giuseppe, B., Marcucci, J., Schmidt, R. y Tissot, B.(noviembre 2022), Machine learning applications in central banking. *IFC Bulletin*, 57, 7-33.
- Binette, A. y Tchegotarev, D. (2019). *Canada's monetary policy report: If text could speak, what would it say?* (Staff Analytical Note 2019), Bank of Canada.
- Bodas, D., García, J., Murillo, J., Pacce, M., Rodrigo, T., Ruiz, P., Ulloa, C., Dios, J. y Velero, H. (2018). *Measuring retail trade using card transactional data* (Working Paper No. 18/03), BBVA Research.
- Chakraborty, C., & Joseph, A. (2017). *Machine learning at central banks* (Working paper No. 674), Bank of England.
- Coelho R., De Simoni M., Prenio J., (2019), *Suptech applications for anti-money laundering* (No 18). FSI Insights on policy implementation
- Coeuré, B. (2017). Policy analysis with big data, speech at the conference on “Economic and Financial Regulation in the Era of Big Data”, Banque de France, Paris, November.
- Committee on the Global Financial System and Financial Stability Board (2017), *Artificial Intelligence and Machine Learning in Financial Services*. Market Developments and Financial Stability Implications
- Danielsson J., Uthemann A., (2023) *On the use of artificial intelligence in financial regulations and the impact on financial stability*. (Discussion Paper No 125). Systemic Risk Centre.
- Dicuonzo, G., Galeone, G., Zappimbulso, E y Dell'Atti, V. (2019). Risk Management 4.0: The Role of Big Data Analytics in the Bank Sector. *International Journal of Economics and Financial Issues*, 9(6), 40-47.
- Doko, F. and Miskovski, I. An Overview of Big Data Analytics in Banking: Approaches, Challenges and Issues. (2019). *UBT International Conference*. 270.
- Einav, L., & Levin, J. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy*, 14, 1–24.
- Financial Stability Board and International Monetary Fund (FSB-IMF) (2009) *The financial crisis and information gaps*. Report to the G-20 Finance Ministers and Central Bank Governors.
- Ghirelli C., Hurtado S., Pérez J., Urtasun A., (2021), New Data Sources for Central Banks, S. Consoli et al. (eds.), *Data Science for Economics and Finance*, https://doi.org/10.1007/978-3-030-66891-4_8
- Gil, M., Pérez, J., Sánchez, J. y Urtasun, A. (2018). *Nowcasting private consumption: Traditional indicators, uncertainty measures, credit cards and some internet data*. (Working Paper No. 1842), Bank of Spain. <https://repositorio.bde.es/bitstream/123456789/8840/1/dt1842e.pdf>
- Hammer, C., Kostroch, D., Quiros, G., & Staff of the IMF Statistics Department (STA) Internal Group (2017). *Big data: potential, challenges, and statistical implications*, IMF Staff Discussion Note.
- Irving Fisher Committee on Central Bank Statistics (2015). *How central banks communicate on official statistics*, (No 15), IFC Report.
- Irving Fisher Committee on Central Bank Statistics (2015). *Central banks' use of and interest in "big data"* (No 3). IFC Report
- Irving Fisher Committee on Central Bank Statistics (2020). *Use of big data sources and applications at central banks* (No 13), IFC Report
- Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., y Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, 37(4), 543-558 DOI:[10.1016/S0167-9236\(03\)00086-1](https://doi.org/10.1016/S0167-9236(03)00086-1)
- Kameda, S. (2022). Use of Alternative Data in the Bank of Japan's Research Activities. *Bank of Japan Review*, E(1), 1-9. https://www.boj.or.jp/en/research/wps_rev/rev_2022/rev22e01.htm

- Kathuria, A. (2016), Impact of big data analytics on banking sector. *International Journal of Science, Engineering and Technology Research*, 5(11), 3138-3141. DOI:[10.1016/j.procs.2015.04.098](https://doi.org/10.1016/j.procs.2015.04.098)
- Kitchin, R. 2015. "Big Data and Official Statistics: Opportunities, Challenges and Risks." *Statistical Journal of the International Association of Official Statistics* 31 (3) 471-481. DOI 10.3233/SJI-150906
- Kohei, M., Yusuke, O., Tomohiro, S., Koji, T., Nowcasting Economic Activity with Mobility Data. *Bank of Japan Review*, E(2), 1-9. https://www.boj.or.jp/en/research/wps_rev/wps_2021/wp21e02.htm
- Laney D. (6 de febrero de 2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Gartner, file No. 949.
- Manyika J., Chui, M., Brown, B., Bughin J., Dobbs R., Roxburgh C., Hung Byers A., (2011) *Big data: the next frontier for innovation, competition, and productivity*. McKinsey & Company. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- McKinsey Explainers (2024 a). *What is Deep Learning?* McKinsey & Company. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-deep-learning>
- *McKinsey Quarterly* (2024 b). Implementing generative AI with speed and safety McKinsey & Company. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/implementing-generative-ai-with-speed-and-safety>
- Mena Roa, M. (22 de octubre de 2021). *El big bang del big data*. Statista. <https://es.statista.com/grafico/26031/volumen-estimado-de-datos-digitales-creados-o-replicados-en-todo-el-mundo/>
- Nyman R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P. & Smith, R. (2018). *News and narratives in financial systems: exploiting big data for systemic risk assessment*. (Staff Working Paper No. 704), Bank of England.
- Nyman, R., Kapadia, S. y Tuckett, D., News and narratives in financial systems: Exploiting big data for systemic risk assessment (2021). *Journal of Economic Dynamics and Control*, 127, 1-28
- Nymand-Andersen, P. (2016), *Big data: The hunt for timely insights and decision certainty*, (No 14). IFC Report. <https://www.bis.org/ifc/publ/ifcwork14.pdf>
- Organization for Economic Cooperation and Development (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance*.
- Pérez, C. (2003), *Revoluciones tecnológicas y capital financiero*. Ed Polarín.
- Peters B., (12 de julio de 2012) *The Age of Big Data*. Forbes <https://www.forbes.com/sites/bradpeters/2012/07/12/the-age-of-big-data/?sh=5f7283154f66>
- Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Klamargias, A. (2019). A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting. In Bank for International Settlements (Ed.), *IFC Bulletins chapters, The use of big data analytics and artificial intelligence in central banking* (vol. 50). Basel: Bank for International Settlements.
- See Choi, H. y H. Varian (2009), "Predicting the Present with Google Trends", *The Economic Record*, 88, 2-9. DOI:[10.2139/ssrn.1659302](https://doi.org/10.2139/ssrn.1659302)
- Shafer T. (Abril 2017) The 42 V's of Big Data and Data Science, *KDnuggets*. <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>
- Sirignano, J., Giesecke, K. and Sadhwani, A. (2016). *Deep learning for mortgage risk*. Stanford University <http://dx.doi.org/10.2139/ssrn.2799443>
- Sosa Escudero, W. (2019) *big data*, Siglo XXI editores.
- Thorsrud, L. A. (2016), *Nowcasting Using News Topics. Big Data versus Big Bank*, (Norges Bank Working Paper No 20). Norges Bank Research. <http://dx.doi.org/10.2139/ssrn.2901450>
- Tissot, B (2019): "Making the most of big data for financial stability purposes", in S Strydom and M Strydom (eds), *Big Data Governance and Perspectives in Knowledge Management*, IGI Global, pp 1-24.

- Tobback, E., Daelemans, W., Junqué, E., Naudts, H., y Martens, D. (junio 2018), *Belgian economic policy uncertainty index: Improvement through text mining*, International Journal of Forecasting, Volume 34, Issue 2, pag, 355-36 <https://doi.org/10.1016/j.ijforecast.2016.08.006>
- Wibisono, O., Dhin, H., Widjanarti, A., Andhika, A. y Tissot, B. (2019), The use of big data analytics and artificial intelligence in central banking, *IFC Bulletin*, 50, 6-25.

9. Anexo

Cuadro 1

Principales tipos de fuentes de datos	Ejemplos
Fuentes oficiales	Transacciones de inversión
	Archivos fiscales
	Oficinales centrales de balances
	Registro de préstamos
	Registro de préstamos hipotecarios
	Supervisores de Mercado financiero
	Estados financieros públicos
	Indicadores de actividad de mercado financiero
Base de datos internas públicas	Encuestas del Banco Central monetarias y financieras
	Oficinas de estadísticas Nacionales
	Información de bancos supervisores
	Información pública de hoja de balances
Información de internet	Baidu
	Google
	Portales de trabajo
	Otros múltiples sitios de internet
	Proveedores de información obtenida mediante web scraping
Información de instituciones financieras	Instituciones residentes de crédito
	Entidades no bancarias
	Repositorio de operaciones de derivados
	Conglomerados de fondos de inversión
Proveedores de datos	Servicios de análisis de Microsoft
	Proveedores comerciales
	Operaciones de liquidación estadísticas de valores
	Operaciones con tarjeta de crédito
Datos de posicionamiento móvil	Operadores móviles
	Proveedores de servicios basados en la ubicación
Registros de supermercados	Tesco (precios online)
	POS (punto de venta)

Nota: Adaptado de IFC (2015)

Cuadro 2

Tarea	Datos	Mutabilidad	Objetivos	Autoridad	Responsabilidad	Consecuencias
Fraude/Protección al consumidor	Amplio	Muy baja	Claro	Singular	Generalmente clara	Baja
Micro riesgo / Previsión de rutina	Amplio	Muy baja	Mayormente Claro	Singular	Claro	Moderada
Criminalidad/ Terrorismo	Limitado	Muy baja	Mayormente Claro	Múltiple	Moderada	Moderada
Ataques Estados Nacionales	Limitado	Muy alta	Complejo	Múltiple	Moderada	Muy severa
Resolución de falla de banco pequeño	Limitado	Parcial	Claro	Mayormente singular	Generalmente clara	Moderada
Resolución de grandes quiebras bancarias/ Turbulencia de mercado	Inusual	Muy alta	Complejo	Múltiple	A menudo poco clara	Severa
Administración de crisis sistémico	Inusual o no disponible	Muy alta	Complejo y contradictorio	Múltiple e internacional	Poco clara incluso ex-post	Muy severa

Nota: elaboración propia en base a Danielsson y Uthemann (2023)