

# Modelos de Arquitectura de Big Data para identificar riesgos financieros en los bancos: Una Revisión Sistemática de la Literatura

## Big Data Architecture Models to Identify Financial Risks in Banks: A Systematic Literature Review

Gustavo Melgarejo<sup>1,a</sup>, José Santisteban<sup>1,b</sup>, Luis Rivera<sup>2,c</sup>

<sup>1</sup> Universidad Nacional Mayor de Marcos, Facultad de Ingeniería de Sistemas e Informática. Lima, Perú

<sup>2</sup> Universidade Estadual do Norte Fluminense, Centro de Ciências e Tecnologia - LCMAT. Rio de Janeiro, Brasil

<sup>a</sup> Autor de correspondencia: [gustavo.melgarejo1@unmsm.edu.pe](mailto:gustavo.melgarejo1@unmsm.edu.pe), <https://orcid.org/0000-0002-0080-0943>

<sup>b</sup> E-mail: [jsantistebanp1@unmsm.edu.pe](mailto:jsantistebanp1@unmsm.edu.pe), <https://orcid.org/0000-0003-4526-642X>

<sup>c</sup> E-mail: [rivera@uenf.br](mailto:rivera@uenf.br), <https://orcid.org/0000-0002-5029-2561>

### Resumen

El sector financiero enfrenta dificultades para gestionar riesgos debido a la creciente cantidad de datos, tanto estructurados como no estructurados, lo que complica la identificación de riesgos financieros, como el incumplimiento de pagos. Los modelos tradicionales no son suficientes para abordar este problema, por lo que se ha explorado el uso de Big Data. El objetivo de este estudio es revisar cómo los modelos de arquitectura de Big Data pueden mejorar la predicción y gestión de riesgos financieros en los bancos. Se realizó una revisión sistemática de la literatura, analizando 32 estudios relevantes publicados entre 2019 y 2023. Los resultados muestran que diversos marcos de trabajo y arquitecturas basadas en Big Data, como el uso de tecnologías como Apache Spark y Apache Storm, son eficaces para procesar grandes volúmenes de datos en tiempo real. Además, se destacaron técnicas de análisis de datos, como el aprendizaje automático, para mejorar la precisión en la identificación de riesgos. Este estudio concluye que la implementación de Big Data y técnicas avanzadas puede mejorar la toma de decisiones en la gestión de riesgos financieros, aunque aún persisten desafíos en la integración de estos modelos en las infraestructuras bancarias existentes.

Palabras clave: big data, riesgos financieros, análisis de datos, arquitectura

### Abstract

The financial sector faces difficulties in managing risks due to the increasing volume of both structured and unstructured data, complicating the identification of financial risks such as payment defaults. Traditional models are insufficient to address this problem, prompting the exploration of Big Data solutions. The objective of this study is to review how Big Data architecture models can enhance the prediction and management of financial risks in banks. A systematic literature review was conducted, analyzing 32 relevant studies published between 2019 and 2023. The results indicate that various Big Data frameworks and architectures, such as those utilizing technologies like Apache Spark and Apache Storm, are effective in processing large data volumes in real-time. Additionally, data analysis techniques like machine learning were highlighted to improve accuracy in risk identification. This study concludes that implementing Big Data and advanced techniques can improve decision-making in financial risk management, although challenges remain in integrating these models into existing banking infrastructures.

Keywords: big data, financial risks, data analysis, architecture

Recibido: 26-09-2024 - Aceptado: 16-12-2024 - Publicado: 30-12-2024

#### Citar como:

Melgarejo, G., Santisteban, J. & Rivera, L. (2024). Modelos de Arquitectura de Big Data para identificar riesgos financieros en los bancos: Una Revisión Sistemática de la Literatura. *Revista Peruana de Computación y Sistemas*, 6(2):87-94. <https://doi.org/10.15381/rpcs.v6i2.28877>

© Los autores. Este artículo es publicado por la Revista Peruana de Computación y Sistemas de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0) [<https://creativecommons.org/licenses/by/4.0/deed.es>] que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada de su fuente original.

## 1. Introducción

El sector financiero, bancos y otras instituciones de capital, siempre ha estado en una constante inversión en tecnología de información (TI), especialmente relacionada con el análisis de los datos. En los últimos años se ha evidenciado la presencia de nuevos tipos de datos, muchos de los que no presentan una estructura adecuada para procesarlos y almacenarlos en tiempo real, que el volumen de datos generados aumenta linealmente con el tiempo [1]. Todo esto imposibilita que las áreas de riesgos de los bancos puedan identificar con precisión los clientes que presentan inconvenientes en los pagos de sus préstamos y productos que adquieren. Esto es un desafío para el sector financiero, ya que la detección temprana de posibles riesgos es fundamental para garantizar la estabilidad financiera y de la economía en general. El uso y la correcta evaluación de los datos pueden ayudar a los bancos a comprender y combatir mejor los riesgos financieros ocultos [2]. La gestión de riesgos se convierte en un elemento importante de la gestión estratégica de las empresas y, en muchos casos, son cruciales para racionalizar las actividades comerciales y continuar las operaciones [3].

El sector financiero siempre ha tenido que lidiar con el riesgo crediticio y otros tipos de riesgos. Los modelos de riesgo tradicionales no son suficientes para manejar la gran cantidad de datos y variables que se deben considerar para la evaluación de riesgos. El volumen de bases de datos que manejan las empresas financieras es tan grande que se ha vuelto necesario abordar este problema. Una alternativa de abordar este desafío está en las técnicas de Big Data aplicadas a conjuntos masivos de datos financieros para la segmentación de grupos de riesgo [4]. El uso del Big Data permite la verificación de modelos y teorías existentes, así como en la mejora de la toma de decisiones [5].

Una evaluación poco precisa de los riesgos financieros puede llevar a decisiones equivocadas en la otorgación de créditos debido a que se desconoce si el cliente pueda pagar o no el préstamo reduciendo así los riesgos de morosidad de los clientes. En la lectura Wang [6] indica que una vez que ocurre el riesgo financiero representará una amenaza para la estabilidad financiera y el riesgo puede ser potencialmente dañino.

El objetivo del sistema de análisis de riesgo financiero es prevenir o reducir las deudas incobrables y las pérdidas de capital causadas por la deshonestidad de los prestatarios y los problemas de evasión de deuda [6]. El problema que se aborda en este estudio es la dificultad que enfrentan las áreas de riesgo de los bancos para identificar y gestionar eficazmente los riesgos financieros debido a la gran cantidad de datos estructurados y no estructurados generados en el entorno actual de negocios. Según Frommann [7], la ratio de morosidad bruta en Europa se sitúa en el 3,8% y el stock total en los balances ronda los 580.000 millones de euros. Aunque se trata de un descenso considerable respecto a la ratio del 8% y las cifras de 1 billón de euros registradas hace 5 años, Europa sigue estando bastante por detrás de otras

grandes economías avanzadas, como por ejemplo la tasa de morosidad en Estados Unidos y Japón es de sólo el 1,6% y el 1,1% respectivamente.

Por otro lado, en el artículo Evolución de mora y mora ajustada del Sistema Financiero [8] indica que el crecimiento de la cartera morosa en el Perú para finales del 2022 ha estado en línea con la expansión de las colocaciones y la brecha se ha mantenido relativamente estable (1.5%) muy similar a prepandemia (1.8%), debido a la crisis sociopolítica y los conflictos externos que impactan a la economía peruana. La morosidad de diciembre de 2019 a febrero de 2023 se incrementó principalmente por el uso de tarjetas de crédito de un 5,33% a un 7,01% respectivamente.

A pesar que las plataformas y herramientas tradicionales utilizadas para la gestión de riesgos en la industria bancaria, como los modelos estadísticos y las hojas de cálculo, pueden ayudar en algo en la identificación de los riesgos financieros, pero tienen limitaciones en cuanto a la capacidad de procesamiento y análisis de grandes cantidades de datos; estas herramientas no siempre pueden capturar la complejidad de los riesgos que están en constante evolución, y no siempre son capaces de proporcionar resultados precisos y oportunos. Al mejorar la integridad, oportunidad y sinergia de la información financiera, se mejora la exhaustividad y la capacidad de analizar problemas complejos en el sector financiero, logrando enfrentar cambios tan rápidos, y ayudar a mejorar las capacidades de gestión financiera proporcionando una guía de toma de decisiones valiosa para las operaciones comerciales [15].

Se debe de tener un marco de trabajo para la medición del riesgo crediticio basado en Big Data [9]. El algoritmo de machine learning con una precisión que varía entre el 73% y 75% para predecir los riesgos financieros [10]. En el artículo de Liu et al. [11] se aborda la identificación y gestión de riesgos financieros emergentes en la era del Big Data. Por su parte Xie et al. [12] discuten los riesgos y la supervisión de las aplicaciones de Big Data en el sector financiero. Guo et al. [13] proponen un modelo de evaluación de riesgos para créditos en línea. Finalmente, Li et al. [14] presentan un marco de trabajo para el análisis de Big Data en la fabricación inteligente, que puede ser adaptado para la gestión de riesgos financieros en bancos.

Los modelos de arquitectura de Big Data pueden proporcionar una solución eficaz para abordar este desafío, permitiendo a los bancos implementar soluciones tecnológicas que permitan analizar grandes volúmenes de datos de diferentes fuentes en tiempo real, lo que a su vez puede mejorar la precisión y la eficiencia en la toma de decisiones. Como indican Stockinger et al. [16], las grandes organizaciones financieras tienen cientos de millones de contratos financieros en sus balances, mientras que los mercados financieros son altamente volátiles, y los conjuntos de datos heterogéneos dentro y entre los bancos de todo el mundo ocasionan que el análisis financiero, casi en

tiempo real, sea muy desafiante. En tal sentido, se debe de determinar modelos de arquitectura que determinen estándares de modelado de datos y la creación de algoritmos de riesgo financiero estándares que permitan obtener información más escalable y consistente.

Por lo que el presente estudio plantea la cuestión de investigación de ¿Qué aspectos se han desarrollado sobre la predicción de riesgos financieros en los bancos? Teniendo en cuenta la pregunta se puede entender que la predicción de riesgos financieros en los bancos ha sido objeto de amplia investigación y desarrollo en los últimos años. Se han abordado diversos aspectos con el objetivo de mejorar la capacidad de los bancos para anticipar y mitigar posibles crisis financieras.

El objetivo de esta investigación es realizar una exhaustiva revisión sistemática de la literatura sobre los aspectos de cómo los modelos de arquitectura de Big Data pueden facilitar la predicción y gestión de riesgos financieros en bancos.

El resto del estudio se divide en 5 secciones, la sección 2 presenta la metodología de la revisión, la sección 3 muestra el análisis para responder a las preguntas planteadas, las secciones 4 y 5 muestran las conclusiones y referencias respectivamente.

## 2. Metodología de la Investigación

Se lleva a cabo una revisión sistemática de la literatura siguiendo el enfoque propuesto por Kitchenham & Charters [17], que consta de tres fases: planificación, desarrollo y resultados.

### 2.1. Planificación

Se identifica la necesidad de realizar una revisión sistemática de la literatura en función de la cuestión general de la investigación focalizada en la sección anterior. De esa cuestión, surgen otras cuatro sub cuestiones enfocadas en explorar los diferentes aspectos para identificar y gestionar los riesgos crediticio-financieros en los bancos:

- ¿Cuáles son los factores que influyen en la investigación?
- ¿Qué marcos de trabajo se han identificado?
- ¿Cuáles son los modelos de arquitectura de Big Data?
- ¿Que técnicas de análisis de datos utilizados en los modelos de arquitectura de Big Data?

Las palabras claves de búsqueda de fuentes de información surgen de las cuestiones. Así, se tienen: Factor, big data, financial risk, risk bank, big data architecture, architectural model, data analysis technique, credit risk analysis, data análisis. La cadena de búsqueda es la combinación de esas palabras claves, como ("big data") OR ("financial risk") OR ("financial risk") OR ("risk bank") OR ("big data architecture") OR ("architectural models") AND ("data analysis techniques" OR ("credit risk analysis" or "data analysis"). Con esa expresión se realizaron las búsquedas exhaustivas de fuentes relevantes para el contexto de la revisión sistemática de la literatura. Se examinan

minuciosamente artículos académicos y conferencias de renombre que están registrados en bases de datos reconocidas como Springer, ScienceDirect, Scopus, IEEE Xplore y Emerald. Los criterios de inclusión utilizados, como muestra la Tabla 1, se basaron en la temática del estudio, la publicación en inglés y el período de tiempo específico, abarcando desde el año 2019 hasta la fecha actual (junio 2023). Análogamente, los criterios de exclusión considerados son aquellos que no están relacionados con contexto de la investigación, fuentes con bajo índice de impacto y otros tal como es sintetizada por la Tabla 2. Se tomó como base la búsqueda de artículos a partir del 2019 debido, esta elección se respalda en los hallazgos del estudio realizado por Lasso Cardona [18], quienes señalaron que a partir de dicha fecha se ha observado un significativo crecimiento del 6% en el mercado de centros de datos en América Latina. Este crecimiento ha sido acompañado por un aumento en la demanda de servicios basados en la nube, Internet de las cosas (IoT) y el análisis de Big Data. En el marco de la revisión sistemática, se analizaron un total de 32 artículos seleccionados los cuales se clasificaron en áreas temáticas según su enfoque principal, lo cual permitió identificar tendencias clave en la literatura reciente, la distribución se presenta en la Tabla 3.

### 2.2. Desarrollo

Los artículos recopilados son evaluados en función de los títulos y resúmenes para determinar si cumplían con los criterios de inclusión establecidos previamente. Aquellos estudios que no se ajustaban a estos criterios fueron excluidos de la revisión asimismo se registraron los resultados de la búsqueda y se eliminaron los artículos duplicados, posteriormente se evaluó la relevancia de los estudios restantes en relación con el objetivo de

**Tabla 1**  
*Criterios de Inclusión*

| Criterios de Inclusión | Descripción   |
|------------------------|---|
| CI1                    | Estudios de investigación de artículos de contribución o aporte, sobre modelos marcos de referencia de arquitectura de Big Data y riesgos financieros |
| CE2                    | Se consideraron artículos solo en idioma inglés.  |
| CE3                    | Se consideraron los estudios que se publicaron desde enero de 2019 hasta junio de 2023  |

**Tabla 2**  
*Criterios de Exclusión*

| Criterios de Inclusión | Descripción  |
|------------------------|--|
| CE1                    | Artículos que no estaban claramente relacionados con los marcos de referencia de arquitectura de Big Data pero que aparecían en la búsqueda debido a la concordancia de palabras clave |
| CE2                    | Artículos que no tengan factor de impacto JSR (Journal Citation Report) o que no estén indexados a las bases de datos arbitradas   |
| CE3                    | Se excluyeron artículos que no estaban relacionados con la pregunta de investigación   |
| CE4                    | Artículos de investigación mayor a 5 años  |

**Tabla 3**

Áreas temáticas de big data

| Área Temática   | Cantidad |
|---|----------|
| Big Data y Análisis de Datos                          | 8        |
| Riesgo Financiero y Modelos de Predicción             | 9        |
| Modelos de Crédito y Análisis Financiero              | 4        |
| Aplicaciones en Telecomunicaciones y Gestión de Datos | 2        |
| Gestión de Riesgos en PYMES y Finanzas                | 1        |
| Gestión de Crisis y Tecnologías Financieras           | 1        |
| Teoría y Revisión de Literatura                       | 1        |
| Total   | 32       |

investigación. Estos resultados se pueden visualizar en la Figura 1.

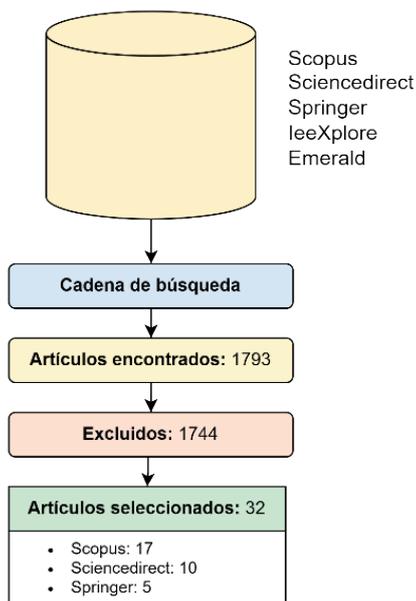
**2.3. Resultados**

El proceso de búsqueda inicial arrojó 1793 estudios relevantes. Luego de aplicar los criterios de inclusión y exclusión establecidos, se seleccionaron 378 estudios para su revisión. Durante esta etapa, se eliminaron los estudios duplicados y se revisaron los títulos y resúmenes de los estudios restantes. Como resultado de este proceso, se identificaron 31 estudios que cumplieran con los objetivos de la investigación y fueron incluidos en el proceso de revisión sistemática.

La Figura 1 ilustra visualmente el proceso de búsqueda y la obtención de las fuentes para esta investigación. Con relación a la categoría (cuartiles) Q1, Q2, Q3 y Q4 de las fuentes, son observados 11, 10, 9 y 1, respectivamente las cuales se visualizan en la Figura 2. Respecto a la intensidad de interés en la investigación sobre el tema tratado, la Figura 3 muestra que en el año 2022 se registró el mayor número de publicaciones. Este incremento evidencia un enfoque destacado en áreas relacionadas con la computación, las comunicaciones inalámbricas, la ciencia de datos, los sistemas informáticos y disciplinas afines, reflejando una

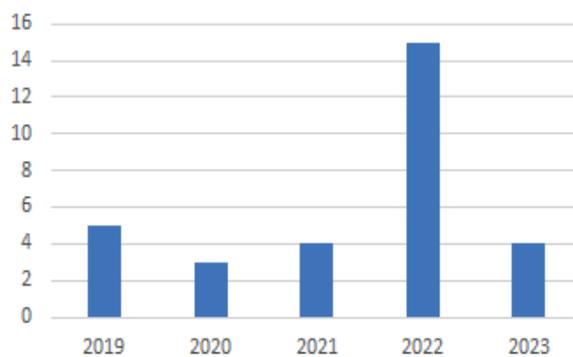
**Figura 1**

Búsqueda y selección de artículos



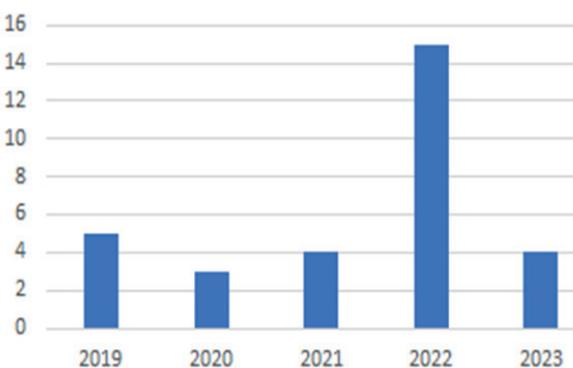
**Figura 2**

Artículos por Cuartiles



**Figura 3**

Artículos por Año



tendencia significativa hacia estos campos durante dicho periodo. Esto sugiere que estos temas tienen un impacto significativo en la actualidad y están siendo objeto de investigación en la comunidad académica.

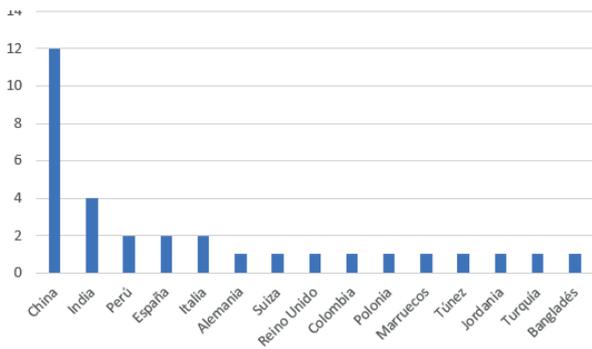
La Figura 4 presenta la distribución cuantitativa de artículos científicos por país, lo que permite identificar patrones de producción y contribución en la literatura académica sobre el tema analizado. Los datos evidencian un claro liderazgo de China, que acumula 12 artículos, representando la mayor participación global en el estudio.

En la Figura 5 se presenta la distribución de artículos recuperados a partir de diferentes bases de datos científicas. En total, se identificaron 32 artículos provenientes de las siguientes fuentes: Scopus, con la mayor representación, aportó 15 artículos (46.88%); seguida de ScienceDirect con 8 artículos (25.00%); IEEE Xplore con 4 artículos (12.50%); mientras que RPCS y Springer contribuyeron cada una con 2 artículos (6.25%). Finalmente, Scielo registró 1 artículo (3.12%). Estos resultados destacan la preponderancia de las bases de datos Scopus y ScienceDirect como fuentes primarias de información académica en la recopilación realizada.

En la figura 6, se puede observar la distribución porcentual de los artículos revisados, organizados en categorías relacionadas con el uso de Big Data para la identificación de riesgos financieros en bancos. Se observa que el 31.25% de los estudios están enfocados

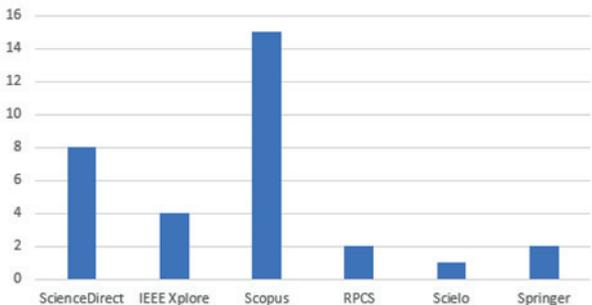
**Figura 4**

Artículos por País



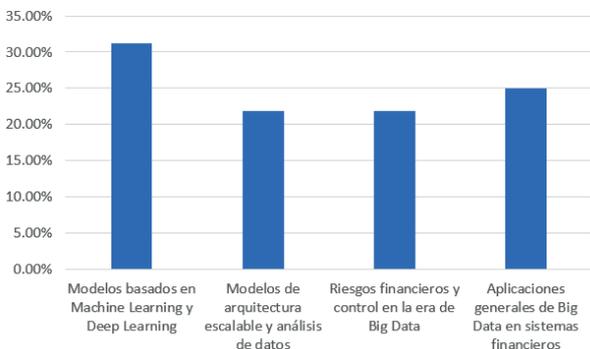
**Figura 5**

Artículos por base de datos



**Figura 6**

Categorías relacionadas al uso de big data



en Modelos basados en Machine Learning y Deep Learning, seguidos por Aplicaciones generales de Big Data en sistemas financieros con 25.00%. Por su parte, las categorías de Modelos de arquitectura escalable y análisis de datos y Riesgos financieros y control en la era de Big Data representan cada una el 21.88% del total de artículos revisados.

### 3. Análisis

Las cuestiones relacionadas a los factores influyentes, marcos de trabajo, arquitectura de big data, y técnicas de análisis de datos, considerados reto en este trabajo, son analizadas en función de las fuentes seleccionadas.

#### P1: ¿Cuáles son los factores que influyen en la investigación?

El trabajo de Stockinger et al. [16] considera los aspectos: evaluación de riesgos, análisis exhaustivo

y consistente, e infraestructura de gestión de riesgos financieros. Considera importante el análisis de sistema complejo de una perspectiva evaluación de riesgo. Este factor implica el análisis y comprensión de los posibles riesgos asociados al sistema en cuestión. El análisis completo y coherente es una tarea desafiante para obtener resultados confiables. Las infraestructuras no son suficientes en términos de gestionar los riesgos financieros de manera efectiva. Este factor señala la limitación de las herramientas y sistemas actuales para abordar los riesgos financieros de manera adecuada. Sin embargo, Chen y Jiang [19] indican que los factores influyentes desarrollo continuo de la tecnología de la información y Análisis científico de datos diversos y a gran escala. El primero es factor importante porque las finanzas son un producto del desarrollo continuo de la tecnología de la información. Indican que los avances tecnológicos han permitido la aplicación de la tecnología de Big Data en el campo de las finanzas. La tecnología de Big Data permite el análisis científico de datos de diferentes fuentes y a gran escala. Esto implica la capacidad de procesar y analizar grandes volúmenes de datos para obtener información valiosa.

Luo [20] considera otros factores basados en la acumulación de información histórica de transacciones, análisis de modelos y tendencias, y aprendizaje profundo y grandes volúmenes de datos históricos. Las organizaciones comerciales recopilan una gran cantidad de datos históricos de transacciones a medida que realizan negocios, destacando su disponibilidad. Los gerentes corporativos desean analizar los modelos disponibles y las tendencias en los datos históricos para descubrir oportunidades comerciales. Este factor resalta la importancia de utilizar el análisis para identificar patrones y oportunidades en los datos. Se menciona el uso de técnicas de aprendizaje profundo y el aprovechamiento de grandes volúmenes de datos históricos disponibles. Esto implica la utilización de enfoques avanzados y el acceso a una gran cantidad de datos para desarrollar modelos de pronóstico en tiempo real.

#### P2: ¿Qué marcos de trabajo se han identificado?

Son pocos los marcos de trabajo formalizados en gestión de riesgos crediticio-financieros encontrados en la literatura. Entre los trabajos más resaltantes se tienen a Łupiński [21], Xie et al. [12], Elia et al. [22], y Murugan [23]. Lupinski indica que el marco analítico es un complemento formalizado basado en el enfoque MMSMDL (modelos de conmutación de Markov multivariados con retrasos distribuidos), que combina datos micro prudenciales para la identificación de los indicadores de alerta temprana de estabilidad financiera de los bancos. En el trabajo de Xie et al. son mostrados los marcos basados en los aspectos calidad, capacidad, capital, garantía y condiciones para determinar el grado de riesgo. Es uno de los métodos comunes utilizados por las instituciones financieras para evaluar los riesgos de los clientes [REF]. Elia et al. proponen un marco basado en las reglas de asociación, que es una de las técnicas de minería de datos más utilizadas que se puede utilizar para descubrir interdependencias ocultas y correlaciones

existentes, permitiendo que las aplicaciones de Big Data puedan almacenar, gestionar y analizar en tiempo real todos los datos e información relacionada con el estado y condiciones tanto de la red, como de los distintos sistemas de información del banco. Sin embargo Murugan [23] indica que los marcos deben basarse en los Análisis de grandes volúmenes de datos, Implementación de plataformas inteligentes y el enfoque en datos en tiempo real todo esto empleando tecnologías de vanguardia.

La Tabla 4 destaca cómo los marcos ofrecen soluciones adaptadas a diferentes necesidades del sector financiero. Su análisis permite identificar herramientas y metodologías que pueden ser combinadas o implementadas según los requisitos específicos de instituciones financieras, fomentando un enfoque más integral y resiliente en la gestión de riesgos.

**P3: ¿Cuáles son los modelos de arquitectura de Big Data?**

El modelo arquitectura para el análisis eficiente de las informaciones complejas es fundamental para el contexto del trabajo, que demanda operaciones en tiempo real. En ese sentido, se observa que Loiacono & Rulli [24] consideran un modelo de arquitectura que sea eficiente en la recopilación, procesamiento y visualización de datos. Por otro lado, Kastouni & Ait Lahcen [25] proponen el modelo de arquitectura basado en siete capas: conexión, integración, lotes, transmisión, servicio, interfaz, y tablero. Eso, según los autores, permite el procesamiento de datos por lotes y de transmisión, y la capacidad de proporcionar información en tiempo real. El modelo UDF de Spark es el recomendable para el procesamiento de consultas en paralelo [15]. Este enfoque tiene la ventaja de que el núcleo de cómputo existente solo requiere una reescritura mínima para aprovechar el entorno de cómputo paralelo de Spark. Kurni et al. [26], también indican una arquitectura SPARK empleando el enfoque de detección mediante la red

**Tabla 4**  
*Comparación de los marcos de trabajo en gestión de riesgos*

| Estudio | Marco/Enfoque  | Características Clave  | Aplicación Principal  |
|---------|--|--|---|
| [21]    | MMSMDL (Modelos de conmutación de Markov multivariados)                  | Identificación de indicadores de alerta temprana mediante datos microprudenciales                  | Predicción de estabilidad financiera de los bancos.           |
| [12]    | Modelo de las 5 C's (calidad, capacidad, capital, garantía, condiciones) | Evaluación integral del riesgo crediticio en cinco dimensiones.                                    | Análisis del riesgo crediticio de clientes financieros.       |
| [22]    | Reglas de asociación y Big Data  | Descubrimiento de correlaciones ocultas; análisis y gestión de datos en tiempo real.               | Identificación de riesgos en redes y sistemas de información. |
| [23]    | Marco basado en tecnologías avanzadas                                    | Análisis de grandes volúmenes de datos; plataformas inteligentes; enfoque en datos en tiempo real. | Mitigación de riesgos financieros en entornos dinámicos.      |

**Tabla 5**  
*Comparación de modelos de arquitectura de big data*

| Modelo/Enfoque                   | Estudio | Características Clave   | Aplicación Principal                              |
|----------------------------------|---------|---|---|
| Modelo de arquitectura eficiente | [24]    | Recopilación, procesamiento y visualización de datos en tiempo real.                  | Análisis general y generación de reportes.        |
| Arquitectura en capas            | [25]    | Siete capas (conexión, integración, lotes, transmisión, servicio, interfaz, tablero). | Procesamiento de datos en lotes y en tiempo real. |
| Spark con UDF                    | [15]    | Procesamiento paralelo mediante funciones definidas por el usuario (UDF).             | Consultas rápidas y eficientes en Big Data.       |
| Spark con Deep Maxout y MRPO     | [26]    | Detección de intrusiones con procesamiento en tiempo real mediante Apache Storm.      | Seguridad y detección de anomalías.               |

Deep maxout basada en MRPO propuesta para detectar intrusiones en un entorno de big data. En esos modelos de arquitectura, los trabajos se realizan en tiempo real gracias a Apache Storm, una tecnología de procesamiento de eventos que ayuda a integrar transacciones en tiempo real en la plataforma. La Tabla 5 permite identificar rápidamente qué modelo arquitectónico se adapta mejor a sus necesidades. Al detallar las características clave y aplicaciones principales, se destaca cómo cada enfoque aborda problemas específicos en la gestión de Big Data, desde visualización hasta seguridad y procesamiento en tiempo real.

**P4: ¿Qué Técnicas de análisis de datos son utilizados en los modelos de arquitectura de Big Data?**

En la literatura se observa el uso de técnicas de aprendizaje automático en los análisis de los datos de la big data. Así, Xie et al. [12] usa el algoritmo el árbol de decisión indicado para el análisis de datos, considerando fundamental, para el desarrollo de ese algoritmo, el crecimiento de sus nodos y establecimiento de criterios de juicio para que las decisiones de clasificación se puedan tomar de forma rápida y precisa. La técnica de análisis es el cálculo del bosque aleatorio [32], que es una ejecución del aprendizaje grupal en el árbol de elección [28]. Asimismo, indica que la regresión lineal, caracterizada por los datos de entrada como un modelo lineal, emplea estimación y resolver los parámetros del modelo utilizando el método de mínimos cuadrados bajo el principio de minimización del error cuadrático [29]. También se indican para este contexto de análisis, las técnicas de Naive Bayes con distribución gaussiana, Support Vector Machine con kernel lineal y Random Forest [31].

En otra vertiente están las técnicas de análisis descriptivo, basadas en encuestas para la recopilación de datos primarios, que miden los elementos de la variable independiente relacionados con las técnicas de análisis de Big Data (Volumen, Velocidad, Variedad y Veracidad), y el segundo enfoque es un enfoque aplicado

Tabla 6

## Técnicas de análisis de datos en big data

| Técnica                       | Descripción  | Estudio    | Aplicación Principal                                      |
|-------------------------------|--|------------|---|
| Árboles de Decisión           | Clasificación precisa mediante crecimiento de nodos y criterios de juicio.       | [12]       | Clasificación y análisis dinámico de grandes volúmenes.   |
| Bosque Aleatorio              | Aprendizaje grupal robusto basado en árboles de decisión.                        | [28], [32] | Reducción del sobreajuste y manejo de datos complejos.    |
| Regresión Lineal              | Ajuste de datos a modelos lineales minimizando el error cuadrático.              | [29]       | Modelado de relaciones continuas en datos financieros.    |
| Naive Bayes                   | Clasificación basada en probabilidades condicionales con distribución gaussiana. | [31]       | Análisis de datos no balanceados en sistemas financieros. |
| Support Vector Machines (SVM) | Clasificación mediante hiperplanos en espacios de alta dimensionalidad.          | [31]       | Separación precisa de clases en datos heterogéneos.       |
| Análisis Descriptivo          | Encuestas para medir impacto de volumen, velocidad, variedad y veracidad.        | [30]       | Reducción de la asimetría de información empresarial.     |

que mide la variable dependiente de la asimetría de la información con base en los estados financieros de las empresas industriales [30].

La Tabla 6 nos revela las técnicas más complejas (como Bosque Aleatorio y SVM) son las más adecuadas para análisis profundos y específicos en entornos dinámicos. Por otro lado, las técnicas más simples, como Regresión Lineal y Análisis Descriptivo, ofrecen soluciones rápidas y fáciles de implementar, especialmente útiles en aplicaciones generales.

#### 4. Conclusión

La revisión sistemática de la literatura confirma que el sector financiero ha invertido constantemente en tecnologías de información, particularmente en análisis de datos, para mejorar la identificación y gestión de riesgos financieros. Sin embargo, como señalan Cheng et al. [2], el aumento de datos no estructurados y la necesidad de análisis en tiempo real plantean desafíos significativos. La aplicación de modelos de arquitectura de Big Data permite analizar grandes volúmenes de datos de diversas fuentes en tiempo real, mejorando la precisión y eficiencia en la toma de decisiones, lo que facilita una detección temprana de riesgos financieros ocultos, esencial para la estabilidad económica y financiera. Los marcos de trabajo basados en Big Data, como destaca Pérez-Martín et al. [4], son eficaces para identificar indicadores de alerta temprana, aprovechando técnicas avanzadas como árboles de decisión, bosques aleatorios y regresión lineal para clasificar y predecir patrones en grandes conjuntos de datos. Estas herramientas permiten a las instituciones financieras procesar y analizar datos diversos y a gran escala, obteniendo información valiosa para predecir

y mitigar riesgos. Por otro lado, Stockinger et al. [16] enfatizan la necesidad de arquitecturas escalables, como Spark SQL, para manejar cálculos masivos y garantizar la transparencia en los análisis financieros

A pesar de que las herramientas tradicionales tienen limitaciones, Wu [29] y Wang [6] subrayan que la integración de modelos de arquitectura de Big Data mejora significativamente la integridad y capacidad de análisis, permitiendo evaluar riesgos de manera más precisa y oportuna. En América Latina, Lasso Cardona et al. [22] y Zelaya y Santisteban [5] resaltan cómo Big Data ha transformado la toma de decisiones en sectores como la logística y la seguridad, demostrando su impacto en la optimización de indicadores clave y la mejora de la satisfacción del cliente. Finalmente, Ma [9] destaca que la regulación y supervisión efectivas son cruciales para mitigar riesgos asociados con la concentración de datos y la privacidad.

En conclusión, la adopción de técnicas y modelos de arquitectura de Big Data es esencial para enfrentar los desafíos actuales en la identificación y gestión de riesgos financieros, contribuyendo a una mayor estabilidad financiera y a decisiones más informadas en el sector bancario. Estos enfoques representan un avance crítico para garantizar una evaluación precisa y adaptable en un entorno dinámico y complejo.

#### 5. Referencias

- [1] [Subramaniaswamy, V., Vijayakumar, V., Logesh, R., Indra Gandhi, V. (2019). Unstructured Data Analysis on Big Data Using Map Reduce. *Procedia Computer Science*, 50, 456-465. doi.org/10.1016/j.procs.2015.04.015
- [2] Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 1(5), 595-606. doi.org/10.1016/j.fmre.2021.08.017
- [3] Porras, J., Pastor, S., Alvarado, R. (2018). Modelo de gestión de riesgos de seguridad de la información para PYMES peruanas. *Revista peruana de computación y sistemas*, 1(1), Article 1. doi.org/10.15381/rpcs.v1i1.14856
- [4] Pérez-Martín, A., Pérez-Torregrosa, A., Vaca, M. (2019). Big Data techniques to measure credit banking risk in home equity loans. *Journal of Business Research*, 89, 448-454. doi.org/10.1016/j.jbusres.2018.02.008
- [5] Zelaya, L., Santisteban, J. (2024). La Influencia de Big Data en la definición de Indicadores Clave de Rendimiento Logístico para la Satisfacción del Cliente: Una revisión sistemática de la literatura. *Revista peruana de computación y sistemas*, 6(1), Article 1. doi.org/10.15381/rpcs.v6i1.2853
- [6] Wang, Y. (2020). Analysis of financial business model towards big data and its applications. *Journal of Visual Communication and Image Representation*, 71, 102729. doi.org/10.1016/j.jvcir.2019.102729
- [7] Frommann, J. (2023). 3 desafíos clave para hacer frente a los préstamos morosos (NPL). <https://www.recevee.com/es/blog/3-key-challenges-in-addressing-non-performing-loans-npls>
- [8] Evolución de mora y mora ajustada del Sistema Financiero. (2023). <https://alertaeconomica.com/evolucion-de-mora-y-mora-ajustada-del-sistema-financiero/>

- [9] Li, H., Zhang, H., Ma, Y. (2020). Credit Risk Measurement, Decision Analysis, Transformation and Upgrading for Financial Big Data. *Journal of Risk Analysis and Crisis Response*, 10(1), 30-40.
- [10] Zhang, Y., Zhu, X., Chen, J., Chen, L., Li, H. (2020). Bank Financial Risk Prediction Model Based on Big Data. *Journal of Intelligent & Fuzzy Systems*, 38(1), 1071-1083.
- [11] Liu, Z., Li, X., Liu, C., Zhang, Q. (2020). Combating emerging financial risks in the big data era: A perspective review. *Future Generation Computer Systems*, 110, 1027-1037.
- [12] Xie, X., Li, K., Li, J. (2020). Discussion of the Risks and Supervision of Financial Big Data Applications. In *2020 IEEE 2nd Conference on Financial Technology (Chongqing)* (pp. 57-61). IEEE.
- [13] Guo, W., Xue, Y., Chen, T., Lu, Y. (2020). Risk assessment of internet credit based on big data analysis. *IEEE Access*, 8, 126327-126335.
- [14] Li, Y., Liu, Y., Zhang, L., Shao, Z. (2021). Model-based Big Data Analytics-as-a-Service framework in smart manufacturing: A case study. *Journal of Intelligent Manufacturing*, 32(1), 185-199.
- [15] Yang, N. (2022). Financial Big Data Management and Control and Artificial Intelligence Analysis Method Based on Data Mining Technology. *Wireless Communications and Mobile Computing*, 2022. doi.org/10.1155/2022/7596094
- [16] Stockinger, K., Bundi, N., Heitz, J., Breymann, W. (2019). Scalable architecture for Big Data financial analytics: User-defined functions vs. SQL. *Journal of Big Data*, 6(1), 46. doi.org/10.1186/s40537-019-0209-0
- [17] Kitchenham, B., Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, EBSE Technical Report EBSE-2007-01.
- [18] Cardona, L., Ocampo, D., Esponda, R. (2022). Aplicaciones de la Datificación y Big Data en América Latina entre el 2015 y 2019. *Revista Logos Ciencia y Tecnología*, 14(2), 125-143. doi.org/10.22335/rict.v14i2.1594
- [19] Chen, L., Jiang, R. (2022). Internet Financial Risk Model Evaluation and Control Decision Based on Big Data. *Wireless Communications and Mobile Computing*, 2022. doi.org/10.1155/2022/8606624
- [20] Luo, T. (2018). Research on decision-making of complex venture capital based on financial big data platform. *Complexity*, 2018. doi.org/10.1155/2018/5170281
- [21] Łupiński, M. (2022). Forecasting the instability of polish banks. *Argumenta Oeconomica*, 2022(2), 115-134. doi.org/10.15611/aoe.2022.2.06
- [22] Elia, G., Polimeno, G., Solazzo, G., Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*, 90, 617-632. doi.org/10.1016/j.indmarman.2020.03.015
- [23] Murugan, M., T, S. (2023). Large-scale data-driven financial risk management & analysis using machine learning strategies. *Measurement: Sensors*, 27, 100756. doi.org/10.1016/j.measen.2023.100756
- [24] [24] Loiacono, G., Rulli, E. (2022). ResTech: Innovative technologies for crisis resolution. *Journal of Banking Regulation*, 23(3), 227-243. doi.org/10.1057/s41261-021-00154-4
- [25] Kastouni, M., Lahcen, A. (2022). Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part A), 2758-2770. doi.org/10.1016/j.jksuci.2020.11.024
- [26] Kurni, M., Md, M., Yannam, B. (2022). MRPO-Deep maxout: Manta ray political optimization based Deep maxout network for big data intrusion detection using spark architecture. *Advances in Engineering Software*, 174, 103324. doi.org/10.1016/j.advengsoft.2022.103324
- [27] Sfaxi, L., Aissa, M. (2021). Designing and implementing a Big Data benchmark in a financial context: Application to a cash management use case. *Computing*, 103(9), 1983-2005. doi.org/10.1007/s00607-021-00933-x
- [28] Wei, J. (2022). A Hybrid Intelligent System for Company Financial Risk Detection Based on Tree-Based Model and Deep Neural Network. *Wireless Communications and Mobile Computing*, 2022. doi.org/10.1155/2022/5630319
- [29] Peng, H., Lin, Y., Wu, M. (2022). Bank Financial Risk Prediction Model Based on Big Data. *Scientific Programming*, 2022, undefined-undefined. doi.org/10.1155/2022/3398545
- [30] Alshehadeh, A., Alia, M., Jaradat, Y., Injadat, E., Al-Khawaja, H. (2023). Big data analytics techniques and their impacts on reducing information asymmetry: Evidence from Jordan. *International Journal of Data and Network Science*, 7(3), 1259-1266. doi.org/10.5267/j.ijdns.2023.4.012
- [31] Hindistan, Y., Aiyakogu, B., Rezaeinazhad, A., Korkmaz, H., Dag, H. (2019). Alternative Credit Scoring and Classification Employing Machine Learning Techniques on a Big Data Platform. 731-734. doi.org/10.1109/UBMK.2019.8907113
- [32] Pradhan, M. R., Akter, S., & Al Marouf, A. (2020). Performance evaluation of traditional classifiers on prediction of credit recovery. *Lecture Notes in Electrical Engineering*, 672, 541-551. https://doi.org/10.1007/978-981-15-5558-9\_48