Journal of the Knowledge Economy (2024) 15:13374–13393
https://doi.org/10.1007/s13132-023-01607-z

Knowledge Management and Data Analysis Techniques for Data-Driven Financial Companies

Galena Pisoni[1] · **Bálint Molnár**[2] · **Ádám Tarcsi**[2]

Received: 2 April 2021 / Accepted: 12 November 2023 / Published online: 29 November 2023

Abstract

In today's fast-paced financial industry, knowledge management and data-driven decision making have become essential for the success of financial technology (FinTech) companies. Big data (BD) is a prevalent phenomenon that can be found across many industries, including finance. Despite its complexity and difficulty to comprehend, big data is a critical component of financial services enterprises and technology architectures. We examine BD from various aspects, considering data science (DS) techniques and methodologies that can be applied during the operation of an enterprise. Our aim is to provide an overview of knowledge management (KM) practices and data analysis (DA) strategies and techniques in the daily operations of financial companies. We address the role of knowledge management, data analytics in a financial institution. The paper demonstrates financial institutions' enablement for new services resulting from technological advancements.

Keywords Financial-technology knowledge management · Business intelligence · Organizational science · Data science · Human–computer interaction · Digital transformation

Introduction

Due to digitalization and digital transformation (DT), the financial industry has undergone a massive change during the previous decade (Scardovi, 2017; Chanias et al., 2019; Goldstein et al., 2021). This change is characterized by increased agility in

✉ Galena Pisoni
g.pisoni@yorksj.ac.uk

Bálint Molnár
molnarba@inf.elte.hu

Ádám Tarcsi
ade@inf.elte.hu

1 York Business School, York St John University, Lord Mayor's Walk, YO31 7EX York, UK

2 ELTE, Eötvös Loránd University, Pázmány Péter 1/C, IK 1117 Budapest, Hungary

the back-office operations of the bank and increased connectivity and information processing speed on the customer side too (Buckley & Webster, 2016; Joshi et al., 2017). There has been a movement toward the usage of data in financial organizations recently, partly owing to DT, and partly due to improved regulation and standardization (Payment Services Directive Two PSD2). Big data has brought an abundance of new forms of commercial operations, the main structural drivers for these trends are new patterns of mobile device and media use, as well as reduced fear of or unwillingness to utilize the Internet, especially among the different ages and consumer groups (Lee, 2015; Dapp & Slomka, 2015; Aslanian & Fischer, 2019; Gai et al., 2018).

Financial technology as FinTech refers to financial industry disruptors and innovators who employ ubiquitous technologies, the Internet, and automated data processing to create new products and services (Gomber et al., 2017). These new solutions might come from a variety of places, including new businesses (most typically startups), current financial service providers, and even established technological firms (Zavolokina et al., 2016a, b). Because the finance business has traditionally relied on a large amount of data, computer science has always been favored in this area, and financial companies are always on the lookout for new ways to support existing operations (De Brentani & Cooper, 1992; Trelewicz, 2017). SMEs (small- and medium-size enterprises, including micros) have insufficient resources (financial, human, organizational, etc.) to seriously engage in the adoption of new, frequently disruptive technology, while financial companies may have the resources, but may not be able to implement the variety of e-services based on data analytics, due to strong technical competence needed.

Improvement of processes is one of the most recent key objectives of knowledge management inside an organization; furthermore, the obtained information resulting from the improvement experiment should be efficiently and effectively distributed across business processes stakeholders (Haddad & Hornuf, 2019). Nowadays, there are numerous tools derived from information and communication technologies (ICT) that can be used to acquire knowledge through process mining, utilizing traditional transactional systems such as enterprise resource planning, customer relationship management, and human capital management, and applying a set of algorithms from data science (Jones, 2017; Vercellis, 2009). Data can be further collected from social media, Internet of Things (IoT) sensors, e-mail, and instant messaging. The data science toolboxes that are available and exploitable published and proposed in this reference (Pisoni et al., 2021), for instance, allows the companies to obtain knowledge and, as a result, support strategic and operational decision-making inside their company. A combination of existing technologies—both internal and external to the company—can be used as source systems to extract, prepare, ingest, and then store data (Molnar et al., 2020). The data is reorganized and converted so that it may be analyzed to obtain information and knowledge using advanced data science algorithms that are tailored to the unique requirements for new company concepts. Large data sets with heterogeneous structures may be analyzed fast using data science and modern information architectures (Fang, 2015). Like this, one can construct data-intensive workflows for gathering, analyzing, interpreting, and reviewing the results and can put new business and technical approaches and solutions in place. Exploratory data analysis gives financial industry companies the ability to develop new ways of generating revenues, based on data. Data coming from a variety of sources is collected, and a data analytics platform is established to meet the needs of financial companies (Ceaparu, 2020). Social media data, sensor data, and internet-based statistics are some examples of such potential sources of data. Different kinds of services can benefit from such data, digital financing, including algorithms related to digital factoring, invoicing, and loan calculations, as well as the technology supporting it, digital investments, trading and crowdfunding, digital money, virtual currencies, and digital payments (Gomber et al., 2017; Pisoni, 2021; Molnar et al., 2020).

Knowledge management (KM) involves the systematic collection, organization, and dissemination of information and knowledge within an organization. KM helps to ensure that the right information is available to the right employee/decision maker at the right time. By implementing KM, financial companies can ensure that their employees have access to the latest information and insights, which can help to improve decision-making processes and support the development of new services and products. This is particularly important in a rapidly changing industry where new technologies and business models are constantly emerging. By capturing and sharing the collective knowledge and skills of its employees, a FinTech company can build a culture of innovation and continuous improvement. Knowledge management also helps companies to maintain a competitive edge by allowing them to respond quickly to changes in the market and capitalize on new opportunities.

Data-driven decision-making is the process of using data and analytics to inform business decisions. Financial companies generate large amounts of data, which can be analyzed to gain valuable insights and make informed decisions; on the other hand, it revolutionized the way financial companies operate. By adopting a data-driven approach, financial companies can improve the accuracy and efficiency of their decision-making processes, reduce the risk of incorrect decisions, and gain a competitive advantage. This allows them to identify trends, forecast future performance, and optimize business processes in real time. For example, predictive analytics can be used to identify potential fraud, while customer behavior analysis can help companies to develop more personalized financial products and services.

One of the key benefits of data-driven decision making in FinTech is the ability to make decisions quickly and accurately. In a fast-paced industry, having access to real-time data can give companies a significant advantage over their competitors. By automating routine tasks and processes, companies can also increase efficiency, reducing costs and increasing profitability.

To implement KM and data-driven decision-making, financial companies need to have the right infrastructure and technology in place. This includes data storage and management systems, data analytics tools, and data visualization software. Additionally, it is important to have a clear and defined data strategy, as well as a dedicated team responsible for data management and analysis.

Given the above, in this paper we try to understand and conduct a literature review and perform case studies with companies in the finance domain regarding their knowledge management and data analysis techniques and processes as well as suited architectures for this aim. We set our research question:

RQ1: Which are knowledge management practices and data analytics techniques currently applied in the domain of finance?

Reconciled and different views regarding these aspects were collected from both, literature and interviews.

The outcome of the study is a consolidated view of how knowledge management and data-driven decision-making can be implemented in financial companies.

The rest of the paper is as follows. The second section presents the methodology, the third section presents the findings, that is the knowledge management practices and data analysis techniques in the financial industry. The fourth section critically examines the consequences of the use of data-driven decision-making in financial companies, and the fifth section closes the paper.

Methodology

To evaluate the underlying practices regarding knowledge management and data analysis, as well as the gathering and usage of consumer data in the generic data science finance ecosystem, we used a two-step qualitative process.

We conducted a literature review, identifying, and summarizing different knowledge management and data analysis current practices for future researchers or practitioners in the domain. To do so, we took SCOPUS as a database of reference, identified the occurrence, and scanned for publications containing the following keywords: "data-driven" and "finance," and "knowledge management" and "finance," respectively (Table 1). We searched for publications between 2015 and 2023 in both cases. The search produced 381 papers and 340 papers, respectively; out of these we eliminated all the false positives and kept only the relevant for the study, which amounted to 275 and 220, respectively.

The second step consisted of an exploratory multiple case study (Eisenhardt & Graebner, 2007), with a particular focus on SME-s (small- and medium-size enterprises, including micros). We used the data sources listed in Table 1 as a guide. We report the answers provided by the interviewers on the knowledge management and data analytics techniques employed or in process of setting in their companies.

The data were collected from the interviews with the case studies (Table 2).

Table 1 Table detailing the queries used in the literature review

Subdomain	Query	N. of results	N. of papers analyzed
Data analytics	TITLE-ABS-KEY ("data-driven" AND "finance")	381	275
Knowledge management	TITLE-ABS-KEY ("knowledge management" AND "finance")	340	220

Findings

The goal of our analysis is to outline the knowledge management practices and data analysis techniques already used or can be used in data-driven decision-making in the finance domain by the case studies. We first present the summarized findings from the literature review, on both knowledge management practices in contemporary finance company and data analysis techniques, and later we present the findings from the case studies, on the data analytics approaches already used, and show practitioners view on the subject matter.

Knowledge Management in Contemporary Finance Company

The competition and digital transformation enforce the use of knowledge management in the most modern form. These market forces motivate the utilization of knowledge that is stored and contained in an enterprise. Knowledge management in FinTech advances the productivity and innovation capability of the company. The disciplined data science methods provide a sound grounding in business decision-making (Stodden, 2020; Suryn et al., 2003; Zhang & Zhou, 2004).

There is a reference model for the business process of knowledge management (Yu et al., 2009; White & Miers, 2008) that supports the understanding of the nature of e-commerce. The methodologies of data science and the reference model of knowledge management can be reconciled. The objective is to buttress the business process of decision-making that aims at either rivalry or innovation. The integration of the CRISP-DM methodology Wirth and Hipp (2000, April), that is a proposed a comprehensive process model for carrying out data mining projects, and the knowledge management process lays the foundation for a reference model (Hotz, 2022). Knowledge management in the recent technology and competitive environment exploits intensively the available technologies that are designated under the umbrella of AI, and this set of technologies consists of computational intelligence, data science, machine learning, and decision theory. The business decision goals are formulated through strategic planning, e.g., balanced score card or business canvas with a focus on innovation in the initial phase of the knowledge management model (Kaplan, 2009; Osterwalder & Pigneur, 2010). Then the business goals are decomposed into sub-goals and thereby sub-tasks of decision making that take into account the specific requirements of AI/data science approaches considering the properties and structures of data collections in the phase of "investigation of the problem domain."

The data understanding is a significant step in CRISP-DM methodology and the use of data lake for the business goal realization of enterprises (Molnár et al., 2020; Ma & Molnár, 2020). Firstly, raw data should be ingested then transformed into a proper structure and placed into the appropriate zone. The selected data collections are required to assist in solving the problem identified in the previous phase. The important activities in this phase are as follows: data cleaning, data structuring, and transformation to fit the purpose, i.e., to adjust to the data structure demand of the potential algorithms and approaches (Ilyas & Chu, 2019; Cheon & Baek, 2016; Cleff, 2014).

Table 2 Sample overview for the case study part

Company	Founded	Number of employees	Value proposition	Data sources
Company case 1	2015	7	ReadID payments via phone	Interview with CTO
Company case 2	2008	16	Automated (robo) advisor	Interview with founder
Company case 3	2011	50	Data lakes technology provider for banks and insurance companies	Interview with the key manager of the area
Company case 4	2017	11	Fintech solution for transferring money (via mobile) between users	Interview with founder
Company case 5	2018	27	Company delivering solutions for FinTech regulation	Interview with tech lead
Company case 6	2017	35	Smart investments platform	Interview with the lead developer
Company case 7	2015	15	Crowdsourcing company for the collection of money from families and friends	Interview with manager and owner

In the phase of "*method selection and data preparation*," the goal is to find suitable algorithms and approaches in "AI" that supports knowledge acquisition from the data collections. There are different methods that can be contemplated to be used for knowledge discovery as follows: statistical learning, deep learning, machine learning including various neuron network solutions, soft computing, genetic algorithms, etc. There is an iteration to fulfill the data architecture requirements within this phase to approximate the expected objectives (see Fig. 1).

In the phase "*assessment of results*," the activity of "*discoveries of data and knowledge pattern*" is embedded into a positive feedback loop. The selected algorithms and approaches are made operational on the prepared data architecture. Depending on the problem and the capability of the selected algorithms, various patterns can be identified through classification, association rule, and interdependencies between "feature" variables or vectors and the target variable. There is again a positive feedback loop for *variable selection* to tune the model to achieve high performance in the light of the appropriate metrics.

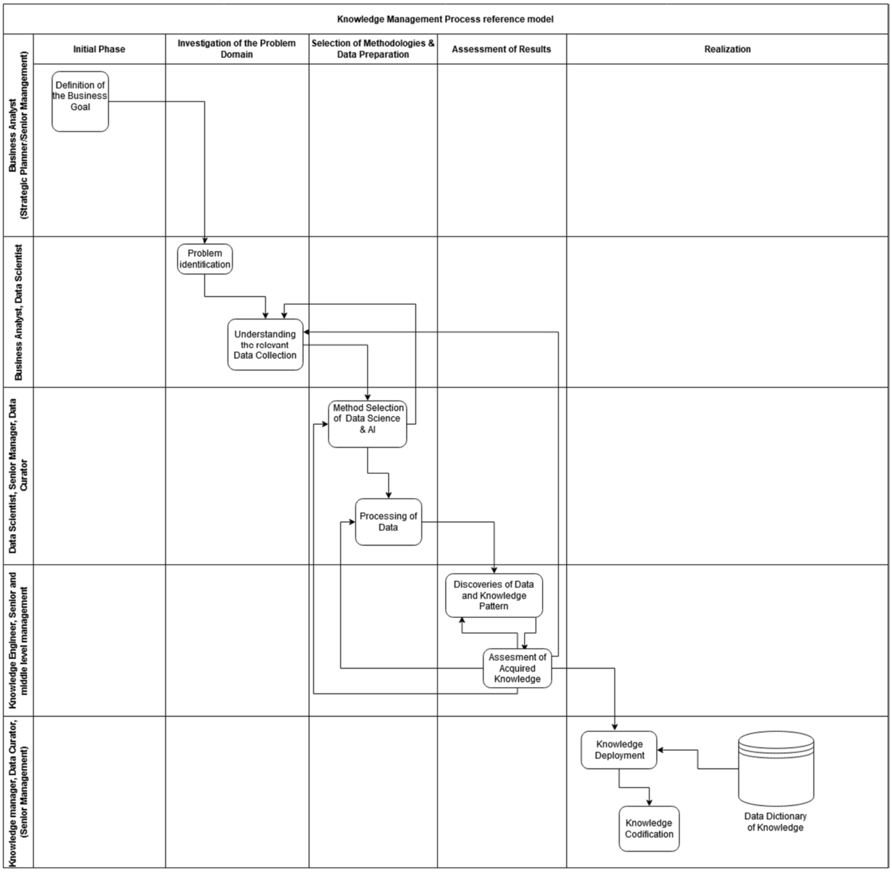

Fig. 1 Knowledge management reference model in the context of data science

In the activity of "*assessment of acquired knowledge*," the validation of the acquired knowledge should be carried out, since the verification of the elaborated model happened in the previous step. The validation means that the potential users of the acquired knowledge will compare it to empirical experiences. The potential users will overarch a wide spectrum of business units of an enterprise, e.g., marketing, finance, customer relationship, information logistics, production, and various level of management.

If the business environment and the acquired knowledge do not fit together then again, a feedback loop must be started. This refinement and fine tuning of models include amendment of models, restructuring the data models, modifying the data architecture, and adjusting the hyperparameters of the used algorithms.

In the "*realization*" phase, the acquired knowledge must be explicitly formalized and "*codified*" (Nonaka & Toyama, 2015) in the activity of "*knowledge codification*." Codification means that the knowledge discovered, validated, and approved is transformed into a form that can be read by both humans and machines, and stored in a repository or dictionary within a data lake.

The "*knowledge deployment*" activity involves the dissemination and proper use of knowledge. Some knowledge may be shared with partners such as customers, consumers, suppliers, and other stakeholders. The exploitation of the acquired knowledge emerges in solving business problems, making decisions, and handling the relationships with partners as finishing or initiating partnerships.

Data Analysis Techniques and Strategies

When planning the implementation of data-intensive flows, first, a suited storage structure should be employed that can store the following types of data: documents, images, videos, and any other unstructured format. Because metadata exist inside the data acquired, coupled with precise information on the entities and thus the data themselves, the concept of metadata plays an important role in the administration of information. Data science algorithms can generate metadata from unstructured, semi-structured, and even structured data since metadata is linked to aggregated data as well. The information items and collections have versions that allow the data engineers to reconstruct the results and re-run the analysis for your working partners. The first step lies in data collection, metadata definition, metadata maintenance, data quality, data gathering, and data entry into the system. This set of tasks is completed in a systematic manner to ensure data security, processing, and compliance (Witt et al., 2009; Baca, 2016; Bach et al., 2019).

An important aspect for finance companies is to include several business and IT objectives in this framework as follows: information security, monitoring, and availability, as well as the performance of the data management system, and the integrity and dependability of the data. Professional activities include assuring the information architecture's business continuity, saving/restoring, archiving, realizing the data structure during the physical design phase, and enforcing the organization's technical standards and rules. This method enables the creation of pipelines for data science methods. Data engineers are responsible data collecting design, development, testing, integration, management, and optimization. In the same time, information architecture may even define the enterprise architecture technology layer that generates the data. The suited databases to support data analytics (either relational or NoSQL) are programmed and built, so that the data later can be cleansed and prepared, and last, the data is analyzed. Suited data analysis models are chosen so that the users can comprehend and perceive the outputs of them (Beheshti et al., 2017; Beheshti et al., 2018).

One of the central systems at the various company decision-making levels is the cognitive information system (CIS), which supports the use of information and knowledge management and the creation of appropriate decisions (Molnár & Mattyasovszky-Philipp, 2019). One variable in decision-making is the time factor, as the right decision has to be made at the right time and with the right speed since companies can create a business competitive advantage by making fast decisions (Mattyasovszky-Philipp et al., 2022). In the context of decision-making, speed and effectiveness are essential during human–computer interaction (HCI). During planning architecture and the knowledge management techniques used, there is a need to consider how humans and computers work together. There are no differences in the case of FinTech companies either. The central knowledge management supporting the cognitive information system involves several additional subsystems and knowledge-based agents, which can employ this paradigm to solve problems and assess hypotheses. The ontology from the knowledge base describes the types of objects or concepts in the application domain, like the relationships between them (Tecuci et al., 2016). Ontologies are crucial in supplying shared knowledge models to semantic-driven applications, which are the focus of the semantic web. Due to their ability to evaluate and qualify an ontology, ontology metrics represent a significant methodology. From the perspective of ontology developers, by evaluating ontology quality, they can automatically identify areas that may need more work and identify some ontology components that may cause issues (García et al., 2010). The instances of these concepts, together with their properties and relationships, are included too. Deep learning (DL) and machine learning (ML) are important technologies to be used by agents, as they play an essential role in generating information for a given decision (Tecuci et al., 2016). They can also be grouped according to their role and function in basic learning strategies:

- Inference: deduction, induction, abduction, analogy
- Computational or representational mechanism: rules, trees, neural networks (NN)
- Learning goal: to learn a concept, discover a formula, acquire new facts, acquire new knowledge about an entity, and refine an entity (Tecuci et al., 2016).

Along with the development of big data lifecycle analysis, knowledge graphs (KGs) have also become a major area of artificial intelligence (Paulheim, 2017). Graphs have always been present in the broader AI literature, but with the emergence of big data on the web and other spheres, the need has arisen to enable machines to understand and use this data in a productive analytical way to enable further use and results. As an example, the search system is not optimal for other kinds of tasks that involve knowledge discovery, comprehension, and learning which require iteration, comparison, and reflection by a user (Karanam et al., 2017). A technological aid to this use of knowledge management is the knowledge graph and associated AI-based technologies, such as natural language processing (NLP) for natural language value representation. A knowledge graph is a graph-theoretic representation of human knowledge such that it can be ingested with semantics by a machine from this point. Knowledge graphs have become a popular data representation that sits at the intersection of knowledge discovery, data mining, semantic web, and the aforementioned natural language processing (Kejriwal, 2019).

We present what came out to be a generic data analysis process for data-empowered financial companies, based on the literature review, in Fig. 2. As Fig. 2 shows the flow of data in a finance company that can be perceived as a knowledge management cycle. The ever-changing market landscape influenced by digital transformation enforces a dynamic process and knowledge management (Bouafia & Molnár, 2018).

Understanding the current economic environment through information gathering provides the opportunity for companies to set up models for data analytics that are capable of forecasting and thereby dynamic modeling to adjust the strategic and tactical objectives of the enterprise. The results of the models support decision-making and selecting future actions from them. The major aim of the regular data analysis exercise is to collect experiences and enhance the situational knowledge of the people in the enterprise to aid critical thinking and the ability to create reasonable proposals exploiting the aggregated tacit knowledge (Grant et al., 2009). Thus, correct

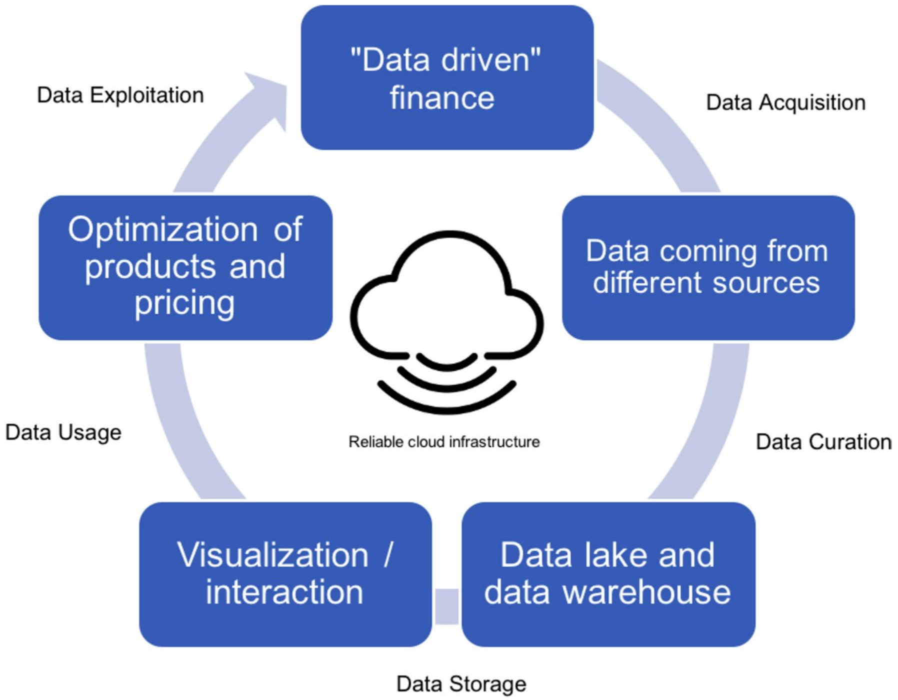

Fig. 2 Generic representation of the data analysis process of financial company

assessments of future actions are augmented. Another advantage for financial companies from using dynamic knowledge management is the overarching perception of the dynamically changing market environment that the enterprises can come across. In organizations the accumulated information and knowledge is disseminated among the members of the human resource of the enterprise that emerges in the form of organizational memory and is analyzed with specialized data mining tools and processes.

The shared image of the current environment, the history, and the outputs of previous activities as gathered experience, and the forecasted outcomes of future actions facilitate the staff of data analytics to collaborate in problem solving with the entire organization. This information and knowledge exchange is embedded into data-intensive workflows and dynamic business processes that concentrate on knowledge acquisition and data gathering; thereby, these processes facilitate tracking the current situation and adjusting to the changing business and security environment (Lui & Lamb, 2018). We derive the following important aspect to bear in mind when implementing enterprise architectures for big data use and analysis in financial companies:

- *Dynamic business processes for knowledge management.* The data-intensive workflows and business processes that can be modified dynamically maintain the situation awareness and the collection of correct data. Modern data architectures assist in developing the quality, accuracy, and frequency of updating to attain rigor control. The applicable toolset includes traditional statistics, quality management (Al-Hakim, 2007), data collection about the demand for services and products, and then stored in data lake, data warehouse, furthermore, the market trend prediction, too.
- *Critical, systems thinking.* The organizational-wide critical thinking underpins exploratory data analysis, the investigation of alternative hypotheses, options of decisions and actions, and the results. Computational finance and data science offer methods for market dynamics modeling, forecasting, predicting supply and demand, and cost-risk analysis based on data stored in the data lake. The recent computational intelligence algorithms enable the analysis of competing hypotheses and multiple courses of action.
- *Shared vision of operation.* It is a typical (knowledge) management task to disseminate and apply the acquired knowledge effectively. The IT solutions help to distribute the information and the acquired knowledge company-wide in proper format and content. Besides the simple e-mail systems, there are sophisticated collaborative work tools, modern database management systems such as SQL-NoSQL, and content management systems.
- *Knowledge creation.* The information architecture supports the logistics of information and the business processes that utilize the collected data in data warehouses and data lakes. The data-intensive workflows and business processes can be improved to aid the conversion of information into actionable knowledge concerning the handling of velocity, accuracy, uncertainty, and decision-making. The modern data management systems data warehouse, data marts, data lakes,

and the algorithms of computational intelligence lay the foundation for effective information and knowledge management.

- *Guarding of intellectual properties.* The protection of the data and information stored in the systems of the enterprise is crucially important against corruption, eavesdropping, and deterioration. Besides the traditional data backup, the toolset of cryptography can be used to ensure the security, safety, and protection of information and data.

Use Cases

Based on the research and the interviews we defined the following knowledge management practices and data-driven (data analysis) techniques that can be used or being used.

Case 1: ReadID Payments via Phone

Knowledge Management

- Creating a knowledge base or repository of information related to the company's products, services, and customers can help employees access information quickly and easily.
- Documentation of business processes, workflows, and policies.

Data-Driven Techniques

- Customer segmentation and behavior analysis are used to understand the purchasing behavior of customers, identify trends and patterns, and develop targeted marketing materials, strategies.
- Predictive analytics is applied to identify potential customer behavior based on past trends and patterns, allowing the company to develop personalized experiences for customers and improve their overall experience.
- Text mining and sentiment analysis can be used to process and analyze large amounts of unstructured data like customer feedback and reviews to gain insights into customer satisfaction and preferences.
- An enterprise architecture framework like Zachman can be used to organize the data and information within the company and ensure that the data collected and analyzed is aligned with the company's goals and objectives.

Case 2: Automated (Robo) Financial Advisory Services

Knowledge Management

- Document management: Organizing and storing financial advice documentation, such as financial plans, investment strategies, and client information.

- Collaboration and sharing: Encouraging cross-functional collaboration and sharing of information between teams to improve decision-making and productivity.
- Knowledge capture and reuse: Documenting and sharing best practices, experience, and lessons learned from previous projects and experiences.
- Competency management: Identifying and developing the skills, knowledge, and expertise required for financial advisory services.
- Performance support: Providing training, coaching, and guidance to support the performance of financial advisors in providing advice to clients.

Data-Driven Techniques

- Predictive modeling: Using algorithms to analyze historical data and predict future market trends and client behavior.
- Customer segmentation: Analyzing customer data to identify and group similar customers to personalize financial advice and investments.
- Sentiment analysis: Analyzing customer feedback and sentiment data to understand customer opinions and preferences.
- Market basket analysis: Identifying patterns and relationships between financial products and services to optimize product recommendations.
- Text mining: Analyzing and extracting valuable information from unstructured data sources, such as customer feedback and reviews.

Case 3: Data Lakes Technology Provider for Banks and Insurance Companies

Knowledge Management Practices

- Data governance: Establishing policies, processes, and standards for managing data assets: defining roles and responsibilities, creating data dictionaries, and monitoring data quality.
- Data cataloging: Creating a centralized repository of data assets, including descriptions of data elements, lineage, and usage: a metadata management tool to.
- Data sharing: Facilitating the sharing of data between different teams, departments, and stakeholders: a data sharing platform to enable data discovery, access, and collaboration.

Data-Driven Techniques

- Data visualization: Representing data graphically to enable insights and understanding. Business intelligence system to create interactive visualizations and reports for decision makers.
- Predictive modeling: Building models to make predictions about future outcomes based on historical data: machine learning algorithms to predict customer behavior or loan default rates.

- Data mining: Analyzing large datasets to identify patterns and relationships: to uncover hidden patterns in customer behavior or to offer new insurance packages to customers.
- Data warehousing: Storing data in a centralized repository optimized for querying and analysis.

Case 4: C2C Money Transferring Service

Knowledge Management Practices

- Document management: Store and organize relevant data and information related to the company's processes, systems, and services: including user manuals, process flows, and training materials.
- Knowledge sharing: Encourage and facilitate the sharing of knowledge and information among employees to ensure consistency and accuracy in all business operations: including regular team meetings, internal forums, and peer-to-peer mentoring.
- Collaboration platforms: Implement collaboration tools that allow employees to share information and work together on projects: project management platforms, chat apps, and shared document repositories.

Data-Driven Techniques

- Predictive analytics: machine learning algorithms to analyze past data and make predictions about future trends and patterns, helping the company to identify opportunities for growth and improve decision-making processes.
- Customer segmentation: Divide customers into different groups based on their behavior, preferences, and demographics in order to better understand customer needs and develop targeted marketing strategies.
- Data visualization: Present data and insights in a visual format that is easy to understand and interpret.
- Fraud detection: Utilizing machine learning algorithms and data analysis techniques to detect fraudulent activities.
- A/B testing: Test different marketing and product strategies by randomly dividing customers into two groups and comparing the results to determine which strategies are most effective and optimize their offerings.

Case 5: Company Delivering Solutions for FinTech Regulation

Knowledge Management Practices

- Document management: Creating and maintaining a centralized repository of all regulatory documents, guidelines, and policies.

- Collaboration platforms: Implementing collaboration tool (SharePoint) to facilitate cross-functional teamwork and information sharing.
- Knowledge retention: to retain institutional knowledge and avoid the loss of expertise when employees leave the company.

Data-Driven Techniques

- Predictive analytics: Building predictive models to analyze the impact of regulatory changes on the business and make proactive adjustments.
- Sentiment analysis: Conducting sentiment analysis on social media and other unstructured data sources to gauge public sentiment on regulatory issues.
- Data mining: Mining regulatory data to identify patterns, trends, and relationships that may inform future compliance strategies.

Case 6: Smart Investments Platform

Knowledge Management Practices

- Documentation and knowledge sharing: Documenting processes, workflows, and procedures to share knowledge among employees and stakeholders.
- Collaboration and communication: Encouraging collaboration and communication among teams to ensure that knowledge is shared and utilized effectively.
- Continuous learning and training: Providing ongoing training and development opportunities to employees to help them stay up-to-date with the latest trends and technologies in the industry.
- Data governance: Implementing data governance policies to ensure that data is collected, stored, and used in a consistent and secure manner.

Data-Driven Techniques

- Data warehousing: Storing large amounts of data in a centralized repository to support data analysis and decision-making.
- Predictive analytics: Using statistical models and machine learning algorithms to predict future outcomes based on historical data.
- Data mining: Analyzing large amounts of data to identify patterns and trends that can inform investment decisions.
- Data visualization: Using visual representations of data, such as charts and graphs, to help stakeholders understand complex data sets and make informed decisions.

The smart investment platform uses these techniques to gather and analyze data on market trends, investment strategies, and customer preferences to provide personalized investment recommendations to clients. The platform could also use data visualization tools to present this information in an easily understandable format, and use data governance policies to ensure that client data is securely stored and protected.

Case 7: Crowdsourcing Company for the Collection of Money from Families and Friends

Knowledge Management Practices

- Document management: maintaining organized, up-to-date, and secure documentation to ensure the effective management of information. Store all transaction records and legal agreements in a centralized database.
- Collaboration and communication: the company use instant messaging tools and an AI chatbot.
- Process management: creating a set of defined business processes, for example, pre-defined business process within the ERP system for accepting payments, verifying user profiles, and distributing funds.

Data-Driven Techniques

- Customer segmentation: dividing customers into groups based on common characteristics such as demographics, behavior, or interests to create targeted marketing campaigns or improve customer experience (different designs for each segment).
- Business intelligence: Dashboard reporting—interactive dashboards to visualize data and support decision-making.

Discussion

Stakeholders are seeing a significant shift in their work resulting from practices and strategies regarding knowledge management and data analysis based on big data. This trend requires a comprehensive understanding of how the ecosystem will evolve to implement new knowledge management and data analysis approaches.

This paper provides valuable new insights by precisely determining the role of knowledge management and data analytics approaches in the FinTech domain, the applications, and everyday use cases in which the companies use these technologies in the financial sector. While the previous literature has a generic character, this work presents a more in-depth analysis of the use of data analytics for FinTech companies and knowledge management practices that might be in place in finance companies.

These findings can be used by practitioners to analyze changing business informatics requirements and practices in the domain of financial technology. There is a need to understand how to translate stakeholder competencies into new enterprise architectures, suited for improved business operations based on new trends in knowledge management and data analytics. The problem arises from the reconciliation of knowledge related to different domains to provide integrated decision-making, strategy, and technique formulation support. This work provides a more in-depth examination instead of the general approach of the use of data-driven knowledge management and data analysis for FinTech companies. Finance companies' existing information architectures are complex structures, which only adds to the complexity and necessitates more attention to the development of new data science approaches. Practitioners can use these findings to analyze changing business informatics practices. Practitioners can then change their current ways of thinking and practices into methodologies that are already well suited to the application of modern data science approaches to their data.

Our work focused on knowledge management and data analysis practices that may need to be extended in future work examples of application domains in the different sub-domains of finance.

Future research into scenarios, deployment strategies, and corporate software could help to improve the current findings.

Conclusion

As a result of our research, we were able to better understand how financial institutions are currently using big data approaches for their work. A literature review of recent fields in digital finance was the first step in our research. We looked at data-driven analysis processes and the big data collection, analysis, and exploitation approach.

We examine practical applications of data-driven processes in finance and consequently, our paper contributes to the literature by creating the basis for companies to use data science to make value. “Big data” and “knowledge management” for financial businesses are based on providing unique or more efficient solutions. These knowledge management and data analysis procedures provide significant commercial value to the study’s participants and help them strengthen their market position.

Author Contribution Galena Pisoni and Bálint Molnár worked on the conceptualization of the raised issue; they write the original draft version, then they carried out editing the revision. Ádám Tarcsi proofread the draft and revision, and supervised the process. Bálint Molnár acquired funding to support the creation of the paper.

Funding The paper was supported by the “Application Domain-Specific Highly Reliable IT Solutions” project, implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the Thematic Excellence Programme TKP2021-NVA 29 (National Challenges Subprogramme) funding scheme, and by the COST Action CA19130—“Fintech and Artificial Intelligence in Finance Towards a transparent financial industry” (FinAI)https://www.cost.eu/actions/CA19130/, accessed on 2021–12–20.

Availability of Data and Materials Data available on request from the authors.

Code Availability N/A.

Declarations

Conflict of Interest The authors declare no competing interest.

References

Al-Hakim, L. (2007, January). Operations and information quality management: current trends and future directions. In Proceedings of the 8th International Conference on Operations and Quantitative Management.

Aslanian, H., & Fischer, F. (2019). *The future of finance: The impact of FinTech, AI, and Crypto on financial services*. Springer.

Baca, M. (2016). Introduction to metadata. Getty Publications.

Bach, M. P., Krstić, Ž., & Seljan, S. (2019). Big data text mining in the financial sector. In *Expert Systems in Finance* (pp. 80–96). Routledge.

Beheshti, A., Benatallah, B., Nouri, R., Chhieng, V. M., Xiong, H., & Zhao, X. (2017). Coredb: a data lake service. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2451–2454).

Beheshti, A., Benatallah, B., Nouri, R., & Tabebordbar, A. (2018). CoreKG: a knowledge lake service. *Proceedings of the VLDB Endowment, 11*(12), 1942–1945.

Bouafia, K., & Molnár, B. (2018, June). Dynamic business process: comparative models and workflow patterns. In The 11th conference of phd students ın computer scıence volume of short papers CS2 (Vol. 194).

Buckley, R. P., & Webster, S. (2016). FinTech in developing countries: charting new customer journeys. *Journal of Financial Transformation*, *44*.

Ceaparu, C. (2021). IT solutions for big data processing and analysis in the finance and banking sectors. In *Intelligent Methods in Computing, Communications and Control: Proceedings of the 8th International Conference on Computers Communications and Control (ICCCC) 2020 8* (pp. 133–144). Springer International Publishing.

Chanias, S., Myers, M. D., & Hess, T. (2019). Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. *The Journal of Strategic Information Systems, 28*(1), 17–33.

Cheon, M. K., & Baek, D. H. (2016). An assessment system for evaluating Big Data capability based on a reference model. *Journal of the Society of Korea Industrial and Systems Engineering, 39*(2), 54–63.

Cleff, T. (2014). *Exploratory data analysis in business and economics*. Springer.

Dapp, T., Slomka, L., AG, D. B., & Hoffmann, R. (2015). Fintech reloaded–traditional banks as digital ecosystems. *Publication of the German original*, 261–274.

De Brentani, U., & Cooper, R. G. (1992). Developing successful new financial services for businesses. *Industrial Marketing Management, 21*(3), 231–241.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*(1), 25–32.

Fang, H. (2015, June). Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)* (pp. 820–824). IEEE.

Gai, K., Qiu, M., & Sun, X. (2018). A survey on FinTech. *Journal of Network and Computer Applications, 103*, 262–273.

García, J., García-Peñalvo, F. J., & Therón, R. (2010). A survey on ontology metrics. *World Summit on the Knowledge Society (3rd 2010 Corfu Island, G.) Knowledge management, information systems, e-learning, and sustainability research.* Springer.

Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies, 34*(7), 3213–3225.

Gomber, P., Koch, J. A., & Siering, M. (2017). Digital Finance and FinTech: Current research and future research directions. *Journal of Business Economics, 87*(5), 537–580.
Grant, J., Markarian, G., & Parbonetti, A. (2009). CEO Risk-related incentives and income smoothing. *Contemporary Accounting Research, 26*(4), 1029–1065.
Haddad, C., & Hornuf, L. (2019). The emergence of the global fintech market: Economic and technological determinants. *Small Business Economics, 53*(1), 81–105. 589.
Hotz, N. (2022). What is CRISP DM?. Aug. 2022. https://www.datascience-pm.com/crisp-dm-2/, Access: 2022–08–15
Ilyas, I. F., & Chu, X. (2019). *Data cleaning*. Morgan & Claypool.
Jones, R. L. (2017). Enterprise FinTech: a metaphor for modern organizations. *ASBBS Proceedings, 24*(1), 353.
Joshi, S., Bhatia, S., Raikar, K., & Pall, H. (2017). Customer experience and associated customer behaviour in end user devices and technologies (smartphones, mobile internet, mobile financial services). *International Journal of High Performance Computing and Networking, 10*(1–2), 118–126.
Kaplan, R. S. (2009). Conceptual foundations of the balanced scorecard. *Handbooks of management accounting research*, *3*, 1253–1269.
Karanam, S., Jorge-Botana, G., Olmos, R., & van Oostendorp, H. (2017). The role of domain knowledge in cognitive modeling of information search. *Information Retrieval Journal, 20*, 456–479.
Kejriwal, M. (2019). *Domain-specific knowledge graph construction*. Cham: Springer International Publishing.
Lee, P. (2015). The fintech entrepreneurs aiming to reinvent finance. *Euromoney (UK)*, *46*(552), 42–48.
Lui, A., & Lamb, G. W. (2018). Artificial intelligence and augmented intelligence collaboration: regaining trust and confidence in the financial sector. *Information & Communications Technology Law, 27*(3), 267–283.
Ma, C., & Molnár, B. (2020). Use of ontology learning in information system integration: a literature survey. In *Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings 12* (pp. 342–353). Springer Singapore.
Mattyasovszky-Philipp, D., Putnoki, A. M., & Molnár, B. (2022). The unrepeatable human mind—Challenges in the development of cognitive information systems—What makes a machine human? *Electronics, 11*(3), 394. https://doi.org/10.3390/electronics11030394
Molnár, B., & Mattyasovszky-Philipp, D. (2019). An architectural approach to cognitive information system. *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 459–462). IEEE.
Molnár, B., Pisoni, G., & Tarcsi, Á. (2020). Data lakes for insurance industry: Exploring challenges and opportunities for customer behaviour analytics, risk assessment, and industry adoption. *ICETE, 3*, 127–134.
Nonaka, I., & Toyama, R. (2015). The knowledge-creating theory revisited: Knowledge creation as a synthesizing process. In: Edwards, J.S. (eds) The Essentials of Knowledge Management. OR Essentials Series. Palgrave Macmillan, London. https://doi.org/10.1057/9781137552105_4
Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers* (Vol. 1). John Wiley & Sons.
Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web, 8*(3), 489–508.
Pisoni, G. (2021). Going digital: case study of an Italian insurance company. *Journal of Business Strategy, 42*(2), 106–115.
Pisoni, G., Molnár, B., & Tarcsi, Á. (2021). Data science for finance: Best-suited methods and enterprise architectures. *Applied System Innovation, 4*(3), 69.
Scardovi, C. (2017). *Digital transformation in financial services* (Vol. 236). Springer International Publishing.
Stodden, V. (2020). The data science life cycle: a disciplined approach to advancing data science as a science. *Communications of the ACM, 63*(7), 58–66.
Suryn, W., Abran, A., & April, A. (2003). *ISO/IEC SQuaRE: The second generation of standards for software product quality*. Acta Press.
Tecuci, G., Marcu, D., Boicu, M., & Schum, D. A. (2016). *Knowledge engineering: building cognitive assistants for evidence-based reasoning*. Cambridge University Press.
Trelewicz, J. Q. (2017). Big data and big money: The role of data in the financial sector. *IT Professional, 19*(3), 8–10.

Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making*. John Wiley & Sons.
White, S. A., & Miers, D. (2008). *BPMN modeling and reference guide: understanding and using BPMN*. Future Strategies Inc.
Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).
Witt, M., Jacob Carlson, D., Brandt, S., & Cragin, M. H. (2009). Constructing data curation profiles. *International Journal of Digital Curation, 4*(3), 93–103.
Yu H, Huang X, Hu X, Wan C (2009) Knowledge management in e-commerce: A data mining perspective. In 2009 *International Conference on Management of e-Commerce and e-Government*.
Zavolokina, L., Dolata, M., & Schwabe, G. (2016a, December). FinTech transformation: How IT-enabled innovations shape the financial sector. In FinanceCom 2016 (pp. 75–88). Springer, Cham.
Zavolokina, L., Dolata, M., & Schwabe, G. (2016b). The FinTech phenomenon: Antecedents of financial innovation perceived by the popular press. *Financial Innovation, 2*(1), 1–16.
Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34*(4), 513–522.