Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 4 (2023)

Unified Data Analytics Platform For Financial Sector Using Big Data

[1]**Surendranadha Reddy Byrapu Reddy**
Sr. Analyst, Information Technology, Northeastern University, Lincoln Financial Group, Atlanta, GA USA
E-mail: suri.byrapu@gmail.com

Abstract: Organizations in today's data-driven digital economy are seeking ways to leverage the immense value of massive amounts of information so as to make more informed decisions. They are able to not only uncover new prospects, but also learn more and improve their performance thanks to big data analytics. Although many businesses have poured resources into big data analytics projects, most have failed to reap the benefits. While there has been a lot of research into the topic of big data analytics, relatively little is known about the strategies that organizations use to merge the various aspects of this discipline. The authors of this paper attempt to solve the problem by developing a comprehensive big data analytics maturity model that can be used by executives to evaluate their own proficiency levels and plan for future development. In this research, we combined traditional quantitative techniques with qualitative meta-synthesis. To begin, we conducted an extensive literature search to catalog the skills and procedures associated with big data analytics maturity. Later, using a quantitative survey method, experts' perspectives on the proposed fundamental skills and practices were analyzed and graded. Finally, a focus group was used to allocate the capabilities to maturity levels according to their priority of deployment, taking into account the architecture of the big data analytics maturity model. We propose a model with four primary capabilities, nine key dimensions (KDs), and five stages of development. Its framework is CMMI-based, which stands for "capability maturity model integration." Questionnaires and focus groups were used to illustrate the big data maturity model. As a guide for effective adoption of big data analytics, we provide the capabilities and KDs, together with their suggested deployment order and weight in the maturity model. Using the offered methodology, businesses may assess their current big data analytics capabilities and steer themselves toward the most effective paths for improvement. Because of this, it enables managers to assess their strengths and weaknesses and establish investment objectives. This work uses a meta-synthesis, which hasn't been done in this subject before, to create an extensive maturity model. The proposed method makes substantial contributions to large data research, and it is both descriptive and prescriptive. A framework is presented in this paper for evaluating different big data analytics projects and settling on a coherent plan of action for their future growth and development.
Keywords: Maturity Model, Big Data, Analytics, Finance, Investment, Digital Economy.

1. Introduction

Databricks' Unified Data Analytics Platform is designed to accelerate innovation by bridging the gap between data science, engineering, and business. Databricks is a unified data analytics platform that streamlines the process of cleaning and preparing massive amounts of data. The capability of the platform to train and deploy ML models in a continuous basis could be useful for all of your AI applications. Having a Unified Data Analytics Platform has three main benefits:

1. Innovate faster with big data
2. Make big data simple
3. Unifying data science and data engineering

1.1 Big Data Analytics in Finance

In recent years, "big data" has garnered a lot of attention, and for good reason. Companies that are able to leverage big data analytics to their advantage will benefit greatly from the amazing amount of data at their disposal. We'll discuss what "big data analytics in finance" is, how it's used, and whether or not it's advantageous for financial institutions in this blog post.

The term "big data in finance" describes the massive amounts of information that banks and other financial institutions produce every day. This information can originate from a wide range of channels, including financial transactions, the stock market, and even social media platforms. Collecting this data, understanding it, and drawing actionable insights is a major problem for businesses. Analytics for large amounts of data are crucial here.

1.2 Big Data Being Used in Finance

These are all important issues to consider when talking about financial applications of big data analytics.

Big data analytics is the practice of employing sophisticated analytic methods to interpret and draw conclusions from very massive data collections. This can be done for a variety of goals, including trend identification, fraud detection, and customer behavior prediction. Big data analytics has seen widespread adoption in sectors like retail and healthcare, and is just beginning to make headway in the financial sector.There are a number of compelling reasons why the financial sector is increasingly adopting big data analytics.

1.2.1 Volume of Data

There has been an explosion in the volume of information generated by financial institutions in recent years. This is because of the dramatic increase in consumer transactions made possible by the proliferation of digital channels like online banking and mobile banking.

1.2.2 Risk Management

Analytics used to large datasets can improve banks' capacity to assess and control danger. In light of the current economic climate, risk management has assumed greater significance in banks and other financial institutions. These businesses can benefit greatly from the use of big data analytics to detect threats early on and implement countermeasures.

1.2.3 Detection of Fraud

Third, with the aid of big data analytics, fraud can be identified. As the volume of digital transactions rises, so too does the incidence of fraud. Financial organizations can benefit from big data analytics to detect and prevent fraud. Big data analytics can now accurately foresee how customers will act. Financial institutions can improve their marketing strategies and better serve their consumers if they have this information.

1.3 Benefits of Big Data in Finance

They bring a special set of potential and problems for big data analytics because they are one of the most data-sensitive business sectors. On the one hand, banks and other financial services providers have access to a wealth of information. However, this information is frequently private and subject to stringent rules.

Over the past two decades, people have relied more and more on computers to process and make sense of massive volumes of information. To handle massive amounts of data and understand them for improved decision-making, big data technologies have never had a more promising chance than in the financial sector. Financial institutions may use big data analytics to assist them face these difficulties and seize these opportunities.

1.3.1 Stocks Can Be Tracked Real-Time

Big data can track the stock market for precursors to price changes, for instance. This can provide financial firms with a substantial trading advantage. Improved financial products and services are another area where big data can be put to good use.

1.3.2 Off-Beat Financial Modeling

Big data allows businesses to create prediction algorithms that can determine which borrowers are most likely to stop making loan payments. Using this data, better lending products can be created that pose less risk to banks.

1.3.3 Analyzing Customer Behavior

Big data can also be utilized systematically to study how consumers act. Marketing and support efforts can be better tailored to specific demographics using this data.

1.3.4 Regulatory Compliance

Also, big data can aid in staying in line with the law. For instance, financial institutions can utilize big data to track and report potentially fraudulent activities. Fraud and money laundering can be avoided with these details.

At last, it can be stated that the financial institutions can benefit greatly from implementing big data analytics. Financial organizations may benefit from big data by making the most of the opportunities it presents and adapting to the challenges it poses.

2. Literature Review

In recent years, financial institutions, including banks and other banking institutions, have been quickening the pace of their digital transformation. Organizations in the financial sector are generating massive amounts of data about their financial and insurance processes as part of this transformation, and are using cutting-edge digital technologies (such as big data, AI, and the IoT) to collect, analyze, and fully capitalize on this data [1].

In addition, recent changes in legislative frameworks, such as Europe's 2nd Payment Services Directive (PSD2) [2,] make it easier for financial institutions to share data with one another, which in turn makes it possible for new kinds of digital financial services to be developed that are predicated on novel kinds of business models. The latter seek to remove barriers for new market players (such as PSPs in the context of PSD2) so that they can introduce innovative services that will delight their clients and bring in fresh revenue.

The banking and insurance industries have a lot of data-heavy digital transformation applications. This is because many different financial institutions share the same data resources. This is valid for a wide range of financial service applications, including but not limited to retail banking, corporate banking, payments, investment banking, capital markets, insurance services, and financial service security [3, 4].

Data from legacy banking systems (such as customer accounts, transactions, and investment portfolios) is combined with data from financial markets, regulatory datasets, social media, and real-time retail transactions in these applications. In addition, several FinTech and InsurTech applications (like personalized healthcare insurance based on medical devices and improved car insurance based on connected car sensors) take advantage of contextual data associated with finance and insurance services to offer better quality of service at a more competitive cost, thanks to the proliferation of IoT devices and applications (like Fitbits, smart phones, and smart home devices) [5]. Also, new opportunities for accurate, automated, and customized services [6] can be found in non-traditional data sources like social media and online news.

More money is being put into digital banking and insurance because of recent advancements in data storage and processing technologies (most notably, AI and blockchain [7]). Using innovations like big data and the internet of things, companies in the banking and insurance sectors are improving the quality and efficiency of their services and increasing the value they provide to customers. There have been early examples of deployment, but many challenges remain until the financial and insurance sectors can fully benefit from big data and AI.

3. Methodology

3.1 Qualitative Meta-Synthesis

Systematic reviews are distinguished from other types of literature reviews by the strict adherence to a set of established procedures for locating, evaluating, selecting, and summarizing relevant studies in order to address a particular subject. Finding all relevant published studies and using well-reasoned criteria to exclude some of them is necessary for a meaningful analysis and synthesis of the data. Hundreds, if not thousands, of potential study reports are often narrowed down to a few that are most like each other. Claims are made based on this process of exclusion, and they could create the appearance that they represent an entire body of research, even if they only comprise papers that share particular features, as determined by the design's complex network. In the quantitative world, where evidence of the effectiveness of an intervention is sought, this strictness makes perfect sense. Qualitative meta-synthesis is complicated by the fact that human subjectivities and experiences can distort the logic of an argument concerning what already exists and what can be derived from it.

There is a substantial divide between authors who view qualitive meta-synthesis as essentially critical, introspective, and integrative, on the one hand, and authors who view it as predominantly aggregative, on the other. More in-depth, cross-disciplinary understandings of phenomena over time and space can be attained through the critically interpretive method of meta-synthesis, which involves unpacking and peering into the impact of methodological choices, theoretical positionings, study contexts, and data sets. This requires a thorough investigation of the actions and motivations of the primary researchers who conducted each study and an understanding of why the results were reported in the ways that they were. Finding the results inside a written report, much alone classifying, evaluating, and integrating them, becomes far more complicated (and interesting) within this type of meta-synthesis than one might initially expect.4 Having a comprehensive and in-depth knowledge of formal and disciplinary culture, particularly the nuanced use of linguistic signifiers across research procedures and academic communities, is essential for comprehending the intellectual components that have produced each study report. This led to the belief, common in the early days of qualitative meta-synthesis, that the method functioned best when driven by groups of researchers from various areas of study.

3.2 Capability maturity model integration (CMMI) architecture

The CMM has been updated into the Capability Maturity Model Integration (CMMI). As can be seen in Figure 2, CMMI was released with five (5) progressive stages to effectively improve an organization's processes and performance. Here are the steps involved: The first is "initially managed," followed by "managed," "defined," "quantitatively managed," and "optimized."

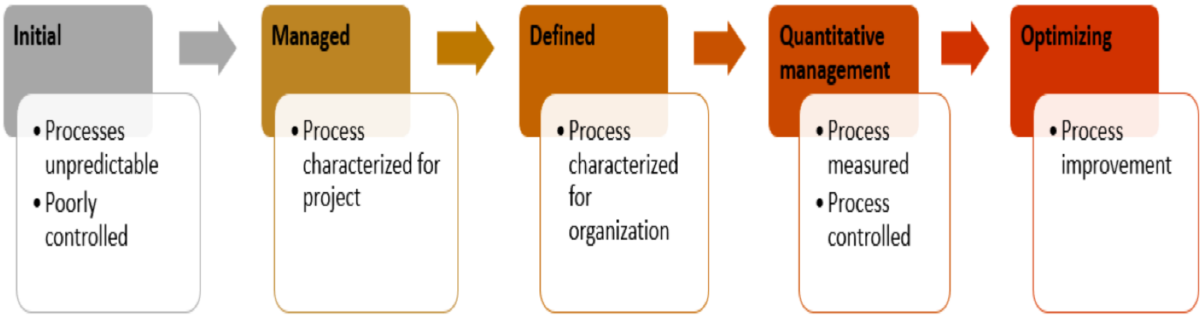

Fig 1: Capability maturity model integration (CMMI)

3.3 CMMI – Maturity Levels

The CMMI model categorizes the degree to which an organization's processes have matured into five distinct stages. Once an organization reaches a certain degree of maturity, its primary focus should shift to maintenance and steady improvement. The CMMI maturity levels are:

- Maturity Level 1: Initial

At this stage, the processes being modeled are being assumed.

- Maturity Level 2: Managed

This level is responsible for the planning, implementation, measurement, and monitoring of projects.

- Maturity Level 3: Defined

The establishment of the project's standards is the primary objective of this stage.

- Maturity Level 4: Quantitatively managed

At this stage, the project is in a controlled position, and the processes offer high-quality results with only a minimal amount of risk.

- Maturity Level 5: Optimizing:

At this level, the processes are both consistent and adaptable. At this point, the organization should optimize its resources in order to achieve continual improvement.

3.4 CMMI – Capability Levels

Organizational performance and process growth in relation to a particular process area can be evaluated using the CMMI's capability levels in addition to the maturity levels. The levels of capability are as follows:

- Capability Level 0: Incomplete: At this level of capability, the process is incomplete, goals are undefined, and standards are not satisfied.

- Capability Level 1: Initial: Concerns about a process area's performance are dealt with here.
- Capability Level 2: Managed: At this stage, the effects of previous efforts to enhance the process are beginning to show.
- Capability Level 3: Defined: At this stage, the organization's norms have been maintained and process standardization is a top priority.
- Capability Level 4: Quantitatively Managed: Statistics and other quantitative methods are being used to keep tabs on the process.
- Capability Level 5: Optimizing:At this point in time, the procedure is constantly being refined.

4. Results And Discussion

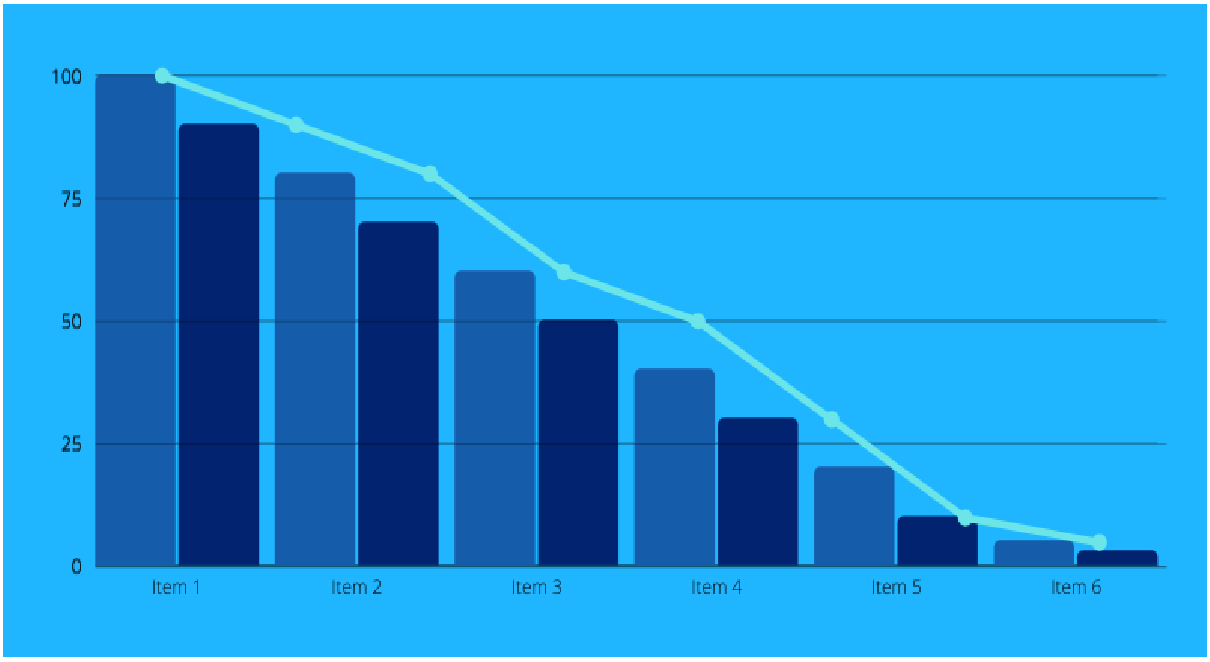

Fig 2: Businesses under beginning to move their decision-making processes.

Machine learning algorithms are gradually replacing humans in corporate decision-making roles. These boost productivity while decreasing the potential for human error. Financial services companies can make speedy judgments with the use of models that estimate the risk associated with specific investments or endeavors. They can still undertake thorough study and due diligence on investments without compromising quality.

Financial planners, on a more consumer-facing level, evaluate a person's lending and credit history to see if they are eligible to get a mortgage. This kind of investigation required a lot of time and individual effort before the advent of machine learning. Approvals are now speedier and more trustworthy than ever with the help of powerful pre-built data models are shown in figure 2.

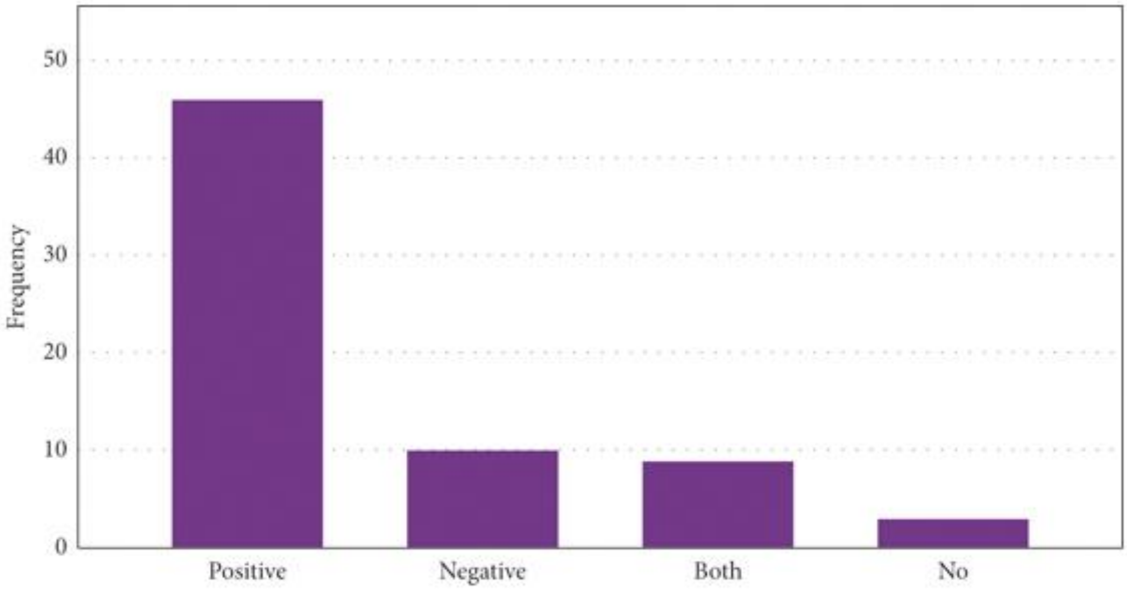

Fig 3: Effects on the capability of programmers.

Figure 3 displays the responses we received from our survey of software development professionals to determine how much they believe CMMI affects programmer capability.

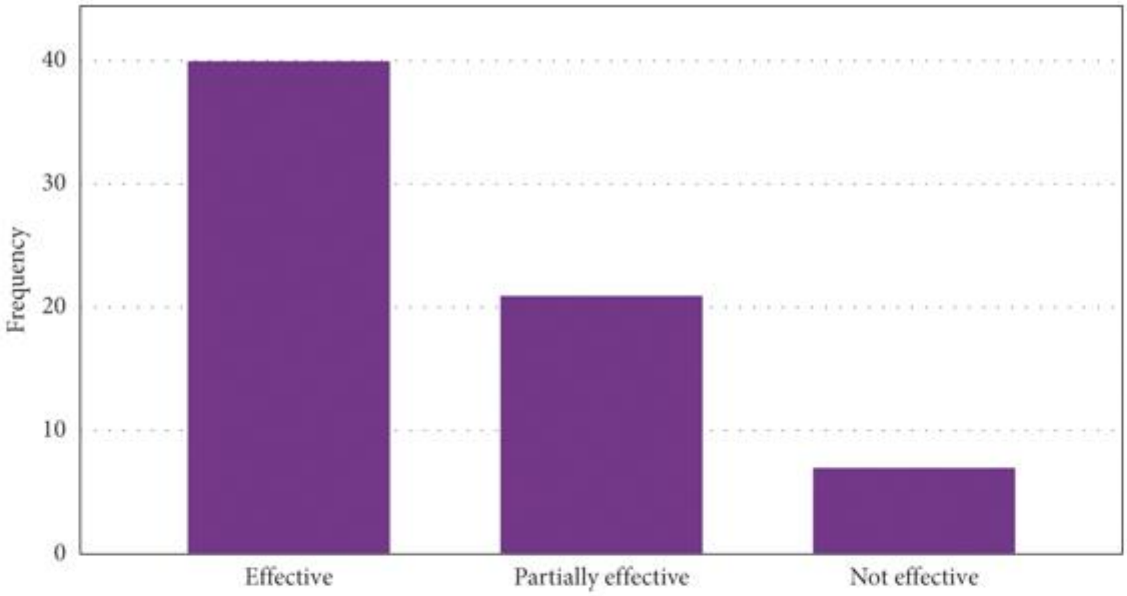

Fig 4: Effectiveness of our proposed prototype.

There are five stages in the Capability Maturity Model Integration (CMMI) framework: inception, management, definition, quantitative management, and optimization. We propose modifying Level 3 to include new activities like rapid prototyping and joint requirement planning. Software engineers and businesses are receiving the prototypes. As can be seen in Figure 4, they implemented the offered remedy and gave positive feedback.

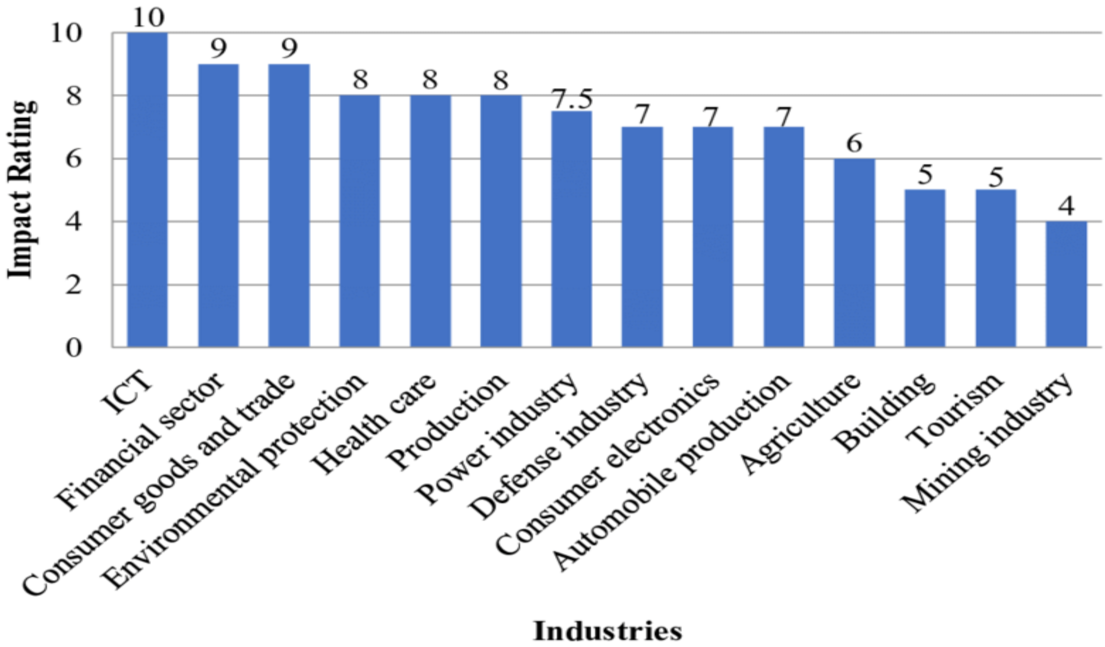

Fig 5: Big Data growth rate

Numerous records are created by the financial sector. Figure 5 displays the strong Big Data growth rates in five different industries, one of which being the financial technology sector.

5. Conclusion

There are growing requirements for regulatory reporting, which calls for better procedures and infrastructure in the financial sector. Better capabilities for tasks like real-time simulation and model testing arise from increased CFO and CRO collaboration and integration. Businesses in the financial sector would do well to adopt a new data operational model that puts data's contribution to decision making front and center. As part of the data governance process, the new model should identify roles and responsibilities as well as requirements for data quality. Data Manager, Data Quality Steward, Data Requirement for Various Reports, etc. are all part of the job description. The goal of this research was to provide a working definition of the term "Metaverse" that is grounded on the existing literature. The tried-and-true meta-synthesis approach has been used to arrive at this conclusion. Minimizing project faults in developing countries while keeping customer input steady and expanding business requires using maximum CMMI to enhance project efficiency by assisting, overseeing, and monitoring. Project management consistency and productivity are both bolstered by CMMI, as are the programming abilities of those working on the projects.

References

[1] Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. Journal of Big Data, 7, 21. https://doi.org/10.1186/s40537-020-00291-z CrossRef

[2] Botta, A., Digiacomo, N., Höll, R., & Oakes, L. (2018). PSD2: Taking advantage of open-banking disruption. Mc Kinsey. https://www.mckinsey.com/industries/financial-services/our-insights/psd2-taking-advantage-of-open-banking-disruption. Accessed 20 July 2021.

[3] Troiano, E., et al. (2020). Security challenges for the critical infrastructures of the financial sector. In J. Soldatos, J. Philpot, & G. Giunta (Eds.), Cyber-physical threat intelligence for critical infrastructures security: A guide to integrated cyber-physical protection of modern critical infrastructures (pp. 2–12). Now Publishers. https://doi.org/10.1561/9781680836875.ch1 CrossRef

[4] Troiano, E., Soldatos, J., Polyviou, A., et al. (2019). Big data platform for integrated cyber and physical security of critical infrastructures for the financial sector: Critical infrastructures as cyber-physical systems. MEDES, 262–269.

[5] Soldatos, J. (2020, December). A 360-degree view of IoT technologies. Artech House. ISBN: 9781630817527.

[6] Dierckx, T., Davis, J., &Schoutens, W. (2020). Using machine learning and alternative data to predict movements in market risk. arXiv: Computational Finance.

[7] Chen, W.; Lai, K.; Cai, Y. Topic generation for Chinese stocks: a cognitively motivated topic modeling method using social media data. Quantitative finance and economics, 2018, 2(2), pp. 279–293.

[8] Chung, W. BizPro: Extracting and categorizing business intelligence factors from textual news articles. International Journal of Information Management, 34(2), 2014, pp. 272–284.

[9] Chye Koh, H.; Kee Low, C. Going concern prediction using data mining techniques. Managerial Auditing Journal, 2004, 19(3), pp. 462–476.

[10] Coussement, K.; Van den Poel, D. Integrating the voice of customers through call center emails into a decision support system for churn prediction. Information Management, 45(3), 2008, pp. 164–174.

[11] Dong, W.; Liao, S.; Liang, L. Financial Statement Fraud Detection using Text Mining: a Systemic Functional Linguistics Theory Perspective. In Proceedings of the Pacific Asia Conference on Information Systems (PACIS), Chiayi, Taiwan, 26th June, 2016, AISeL, p. 188.

[12] Duan, Z.; He, Y.; Zhong, Y. Corporate social responsibility information disclosure objective or not: An empirical research of Chinese listed companies based on text mining. Nankai Business Review International, 9(4), 2018, pp. 519–539.

[13] Ediger, D.; Jiang, K.; Riedy, J.; Bader, D.A.; Corley, C. Massive social network analysis: Mining twitter for social good. In Proceedings of 39th International Conference on Parallel Processing, San Diego, CA, USA, 13–16th September, 2010, IEEE, pp. 583–593.

[14] Eler, D.M.; Grosa, D.; Pola, I.; Garcia, R.; Correia, R.; Teixeira, J. Analysis of Document Pre-Processing Effects in Text and Opinion Mining. Information, 2018, 9(4), p. 100.

[15] Feuerriegel, S.; Gordon, J. Long-term stock index forecasting based on text mining of regulatory disclosures. Decision Support Systems, 112, 2018, 88–97.

[16] Feuerriegel, S.; Gordon, J. News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. European Journal of Operational Research, 272(1), 2019, pp. 162–175.