Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 23, Issue 3, 2023
PRINT ISSN 2284-7995, E-ISSN 2285-3952

THE ROLE OF BIG DATA IN DIGITALIZING INFORMATION

Elena COFAS

University of Agronomic Sciences and Veterinary Medicine Bucharest of Bucharest, 59 Marasti Boulevard, District 1, 011464, Bucharest, Romania, E-mail: cofas.elena@managusamv.ro

Corresponding author: cofas.elena@managusamv.ro

Abstract

In a world increasingly shaped by data, its exponential growth demands global organizations to swiftly embrace and synchronize with the prompt evolution of our lives. Innovations in storage technology, the advent of IoT (Internet of Things), and the emerging regulations of the European Union, such as the General Data Protection Regulation (GDPR), all underscore how "Big Data" propels economic transformation. Amid the rapid proliferation of artificial intelligence and technology, Europe's digital overhaul assumes paramount importance, as recent crises underscore the urgency for more effective alternatives that fuel the imperative shift towards sustainability. The concept of "Big Data" has been integrated as a central pillar in the EU's digital transformation strategy, set for 2030, and consequently, within the ambit of its green strategy. This sector's inherent opportunities contribute to the EU's pursuit of climate neutrality by 2050. At its core, "Big Data" involves the amalgamation of extensive and diverse information, subjected to algorithmic analysis to drive decision-making. The data's significance extends beyond economic implications, permeating diverse domains such as safety, health, agriculture, environment, law, and even individual contexts, thereby accentuating the intrinsic essence of "Big Data". This paper addresses the intricate demands posed by the rapid expansion of this type of data, which is experiencing exponential growth in terms of accessibility and automated integration within digital landscapes. Its efficacy is contingent upon not merely the escalating capabilities of technology to facilitate the accumulation and retention of substantial data quantities, but also on its proficiency to conduct thorough analysis, comprehension, and effective utilization of the data's complete worth.

Key words**:** *"Big Data", analysis, cloud, cluster*

INTRODUCTION

The end of the last decade finds us at a juncture where technological advancement, the internet, and interconnected networks have seamlessly woven themselves into the fabric of our existence. The influence of digitization and the importance of cyberspace have witnessed an exponential surge in the recent years. A substantial portion of the world's population, businesses, and nations have tapped into this realm, progressively depending on increasingly intricate information and communication technology systems. The unrestricted dissemination of information transcends geographical limits, and the magnitude of data and information has experienced a substantial expansion.

In broad terms, the concept of "Big Data" is defined as "large volumes of data rapidly produced by a diverse array of sources" [6]. Furthermore, it pertains to the exponential growth, both in availability and automated utilization of diverse digital information, as well as their analysis through algorithms to underpin decision-making. Definitions concerning "Big Data" exhibit a degree of subjectivity due to the lack of a clear method for quantifying the size of a dataset. The exponential increase in data generated and collected by interconnected technologies, alongside the influx of information from sensors, voice, multimedia, and more, holds a pivotal role in the digital transformation processes across all fields. Simultaneously, beyond the collection and storage of vast data volumes, understanding the potential for analysis and comprehending the value of this data are equally significant [14].

At the heart of the EU's digital transition strategy lies the foundational concept of "Big Data," coinciding with the establishment of a unified European data market through the introduction of the "European Data Act" in February 2022 [24]. The overarching goal of the EU's single data market is to cultivate a

thriving data-driven economy, which is going to be:

- **dynamic,** by allowing the free flow of data within the EU, accessible to all, aimed at fostering a technology-based future and enhancing regional cooperation. Each EU member state can contribute to this realm by establishing what are known as common and interoperable "data spaces", ensuring that the essential data collected within these spaces is collaboratively exchanged;
- **attractive,** through investments in new data storage and processing tools and infrastructures, especially in cloud technology, which will facilitate European convergence in terms of research and the modernization of key sectors of activity;
- **steady**, being backed by well-defined regulatory norms concerning data privacy and protection, as well as in the field of competition law. Additionally, measures are in place to empower data subjects to retain complete authority over their own data;
- **cyclical**, in order to support the goals of the Green Deal. It is also essential to filter data for processing and storage, with an emphasis on data that truly holds value within these innovation and research processes. Similarly, GDPR underscores the importance of setting retention periods for data, enabling the reclamation of storage space and relieving the burden on servers and services that are responsible for their management.

It's worth noting that the European Union already houses vast quantities of qualitative, non-personal data that remain untapped, and this volume is continually expanding. In 2018, the European region stored 33 Zettabytes (Zb) of data (with 1 Zb equivalent to 1 trillion Gigabytes (Gb)). It's estimated that by 2025, this figure will escalate to 175 Zb. This scenario prompts the need to find ways to unlock the value of these data and grant access to European businesses and researchers, all within the boundaries of the law.

In this context, the aim of this research addresses the intricate demands posed by the rapid expansion of this type of data, which is experiencing exponential growth in terms of accessibility and automated integration within digital landscapes.

MATERIALS AND METHODS

Within the data processing and analysis process, "Big Data" combines the subsequent data types:

→ ***structured data*** (possess definitive properties, such as size and format, and can be processed using relational databases),

→ ***semi-structured data*** (do not adhere to formal data standards yet are not entirely disorganized), and

→ ***unstructured data*** (completely disorganized and cannot be stored or processed using relational databases).

Unstructured data is most commonly encountered in the form of audio files, images, video files, social media updates, as well as other textual formats like log files, interaction data, machine and sensor data. Graph databases are becoming increasingly significant due to their ability to display massive amounts of data in a manner that expedites and streamlines the analysis process [15]. "Big Data" encompasses large and intricate datasets that are so voluminous that traditional data processing software simply cannot manage them. Due to the exponential growth of "Big Data" volume (generated not only by human users but also by various sources), new strategies and technologies are needed to analyze "Big Data" sets at terabyte or even petabyte scale. With the advent of the *Internet of Things* (IoT), more and more objects and devices are connected to the internet, gathering data about customer usage patterns and product performance. The emergence of machine learning has further augmented the data generation. Cloud computing technology has greatly expanded the possibilities offered by "Big Data" [5].

A dataset becomes a cluster when data sharing the same properties are grouped together. The cloud storage system provides efficient scalability, enabling developers to create clusters for testing a subset of data (Fig. 1). As integration progresses, the data requires inputting, processing, formatting, and rendering it accessible in a practical format.

Following this, vital insights can be derived, serving the purposes of advancing *machine learning*, crafting *predictive models*, and identifying *behavioral patterns*.

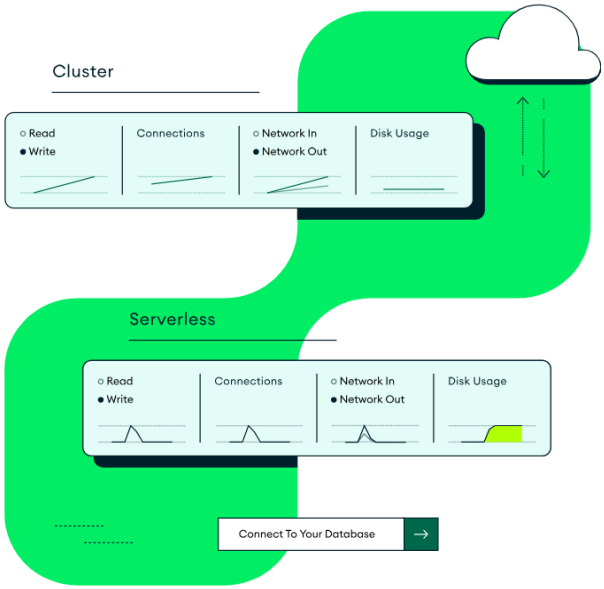

Fig. 1. Management of "Big Data" within the cloud environment
Source: https://www.mongodb.com/cloud/atlas.

In order to fully harness the potential of "Big Data," companies require the appropriate tools to process, analyze, and store the vital information they produce and gather on a daily basis for real-time outcomes. The four core components of any Big Data project include data storage (big data storage), data extraction (data mining), analysis, and visualization, with each element featuring innovative and high-tech instruments:

➢ **Data storage**: The storage of "Big Data" requires appropriate space, with storage solutions available in the form of cloud, on-site, or hybrid configurations. Data can be stored in diverse formats and integrated as needed, aligning with desired processing requirements and the essential processing engines within specific datasets. Cloud-based storage alternatives are progressively gaining favor due to their capacity to meet present computational needs, allowing flexible resource utilization, and ensuring secure and easily accessible data storage. As such they are essential to optimize the volume of information that can be stored, and it's worth noting that there already exist certain solutions specifically designed for this purpose, such as:

a)*HBase/Hadoop* - is an open-source platform that accommodates both structured and unstructured data, designed specifically for storing very large datasets using clusters [8].

b)*MongoDB* – proves to be highly beneficial for organizations employing a blend of semi-structured and unstructured data. For instance, it caters to businesses developing mobile applications or those requiring storage for product catalogs or data essential for real-time personalization [13].

➢ **Data mining**: Once the data is stored, tools need to be added to facilitate the discovery of information intended for analysis or visualization. The tools listed below aid in extracting relevant data without requiring manual tracing - a task that becomes impractical for humans, especially when dealing with thousands of records:

a)*IBM SPSS Modeler* can be employed to build predictive models using a visual interface, encompassing text analysis, entity analysis, decision management, and optimization. It enables the extraction of both structured and unstructured data within a comprehensive dataset [10].

b)*KNIME* is a scalable, open-source solution that provides an extensive array of algorithms and community contributions for data extraction and analysis, predictions, and key insights discovery. Text files, databases, documents, images, networks, and even Hadoop based data can be ingested, making it an ideal solution when dealing with combined data types [11].

c)*RapidMiner* is an open-source tool that empowers users to leverage templates instead of writing programming code, at the same time providing machine learning, data mining, predictive analytics, and business intelligence in order to support the entire process [18].

➢ **Data analysis**: Leveraging machine learning and artificial intelligence through visual analysis of diverse datasets enables the construction of data models, meaning that data can be thoroughly explored to make new discoveries and practically applied to each client's needs. The most powerful tools to

facilitate data analysis for obtaining essential business, customer, or global insights include:
a)Apache Spark - is one of the most well-known tools for data analysis, with users ranging from small businesses to government agencies and tech behemoths like Apple, Meta (Facebook), IBM, and Microsoft. It functions as a quick, effective, and open-source tool that is compatible with the main Big Data programming languages, such as Java, Scala, Python, R, and SQL. This tool also enables developers to extensively employ SQL, batch processing, stream processing, and machine learning all within a single location, alongside graph processing. Impressively versatile, it operates on platforms such as Hadoop (for which it was initially developed), Apache Mesos, Kubernetes, both as an independent framework and in the cloud [1].
b)*Presto* - is an open-source tool that employs distributed SQL queries, designed to function as a robust engine for interactive data analysis. This tool supports both non-relational sources like Hadoop Distributed File System (HDFS), Amazon S3, Cassandra, MongoDB, and HBase, as well as relational data sources including MySQL, PostgreSQL, Amazon Redshift, Microsoft SQL Server, and Teradata. It finds utility in massive corporations such as Meta, Netflix, Airbnb, and Groupon [16].
c)*SAP HANA* – is typically utilized to aid businesses in making prompt decisions, drawing upon extensive sets of data [19].
d)*Tableau* - combines data analysis and visualization tools, and can be used on a desktop, through a server, or online [23].
e)*Splunk Hunk* (*Analytics for Hadoop*) - serves as a comprehensive analysis tool capable of generating queries, charts, and visual representations of fed data, all manageable through a dashboard, swiftly created and shared via the Hunk interface. It also operates on other databases and stores, including Amazon EMR, Cloudera CDH, and the Hortonworks Data Platform [22].

- **Data visualization:** To ensure easy comprehension during presentations, data is transformed into data visualizations. The top visualization tools include:

a)*Plotly* supports the creation of charts, presentations, and dashboards from analyzed data, using JavaScript, Python, Matlab, Jupyter, or Excel. Utilizing a graphical user interface (GUI) for importing and analyzing data, along with an extensive visualization library and an online chart-building tool, it becomes incredibly easy for it to generate excellent graphics [15].
b)*DataHero* is a user-friendly visualization tool that can extract data from various cloud services and input them into charts and dashboards [4].
c)*QlikView* enables the creation of data visualizations from all data sources using self-service tools that eliminate the need for complex data models and can be shared with others, allowing collaborative decision-making based on revealed trends and data. Advanced capabilities allow visual analyses to be embedded in applications, while dashboards can guide individuals through the production of analytical reports without requiring an understanding of data science [17].

RESULTS AND DISCUSSIONS

a. Attributes of “Big Data”

Although the concept of "Big Data" is relatively recent, the origins of large datasets trace back to the 1960s and 1970s, during the nascent stage of the data universe, marked by the emergence of early data centers and the development of relational database systems. Around the year 2005, an increasing awareness emerged regarding the substantial volume of data being generated by users through platforms like Facebook, YouTube, and other online services. The advancement of open-source frameworks, such as Hadoop and more recently Spark, has played a pivotal role in the burgeoning of "Big Data," as they facilitate the more streamlined processing of massive data volumes and alleviate the cost burden of storage.
The etymology of the term "Big Data" dates back to the mid-1990s and pertained to the manipulation and analysis of extensive datasets. It was first introduced into discourse in 1998 by John R. Mashey, an IT specialist,

in his seminal work "Big Data and the Next Wave of Infrastress." Subsequently, in the year 2000, Peter Lyman and Hal Varian published the ground-breaking study "How Much Information?", representing the first endeavour to quantify the annual generation of new information on a global scale. In 2001, Douglas Laney elaborated on the distinct characteristics of "Big Data," which were later recognized and encapsulated as the "3 Vs."[12]:

- ***volume*** - pertains to size, likely the most widely recognized characteristic of "Big Data", especially considering that over 90% of all contemporary data has been generated in recent years; this aspect encompasses vast quantities of data, influenced by the proliferation of sources from which the data originates;
- ***velocity*** - refers to the speed at which data is generated, collected, updated, and processed in real-time, playing a significant role in determining the usefulness and potential of the data; typically, the highest data speed is achieved through direct in-memory transmission, as compared to disk writes; and
- ***variety*** - signifies the diversity of digital data, as data is acquired in a growing number of different formats, ranging from structured data (such as numeric data stored in databases) to unstructured data (such as text documents, emails, videos, audios, or financial transactions, depending on the specific nature of each company's activities, objectives, and strategies). In essence, data can be categorised based on its origin, source, and format (structured, semi-structured, or unstructured data).

The complexity of huge datasets has been demonstrated in practice throughout time, which has led to the discovery of additional features beyond the three previously mentioned, leading to the emergence of the "10 Vs", which are as follows [2], [3]:

- ***variability*** - refers to:

- the inconsistency of typically unstructured data sequences, which need to be identified through anomaly detection methods, thus requiring data filtering and flow control.
- the fact that "Big Data" is also variable due to the multitude of data dimensions resulting from various types and sources of data;
- the inconsistent speed at which "Big Data" is loaded into the database.

- ***veracity*** (***accuracy)*** - refers to the quality of data and their sources ("trusted sources"), as well as the integrity and comprehensiveness of the data set. Essentially, it describes the discrepancies, inconsistencies, and uncertainties that come with data collection, and the quality and quantity of the data determine whether they can be effectively used to generate useful information.
- ***validity*** - similarly to veracity, it refers to how accurate and correct the data are for the intended use. According to Forbes, an estimated 60% of data scientists' time is spent on data cleaning before any analysis can take place. The effectiveness of "Big Data" analysis relies heavily on the quality of the underlying data, underscoring the importance of implementing appropriate practices to ensure consistent data quality, standardized definitions, and comprehensive metadata.
- ***vulnerability*** - "Big Data" poses new security concerns, a breach of their security constituting a significant violation.
- ***volatility*** - Before the emergence of the "Big Data" concept, organizations tended to store data indefinitely. It could even be kept in the live database without causing performance issues, but due to the speed and volume of "Big Data", its volatility needs to be carefully taken into consideration.
- ***visualization*** - entails presentations aimed at interpreting data and providing context. Current data visualization tools face technical challenges due to limitations in memory technology and reduced scalability, functionality, and response time. For instance, traditional charts cannot be relied upon when dealing with a billion data points, necessitating different data representation methods such as data clustering or using tree maps, parallel coordinates, circular network diagrams, or cones. Coupled with the multitude of variables stemming from data variety and high velocity, as well as the intricate relationships among them, it becomes

evident that crafting meaningful visualizations is no simple task.

➢ ***value*** – the most crucial characteristic is deriving value from data, rendering other aspects of "Big Data" meaningless without it. This encompasses how data analysis and processing can lead to their quantification, allowing for the correlation of these insights through processes that manage data accumulation.

All this parameters demonstrate that the significance of "Big Data" extends beyond the sheer volume of information that it covers, equally important being the speed at which it can be reached, as well as the numerous different categories of data involved. "Big Data" manifests in various formats including text analysis, social network analysis, web analysis, mobile analysis, multimedia analysis (including images, audio, and video), and data collected from the Internet of Things (IoT). However, despite the substantial data quantity, its intrinsic value to data holders is not guaranteed. Contemporary advanced technology operates within remarkably brief periods, processing an immense volume of data, often numbering in the millions. This process iteratively adjusts variables to discern patterns that can lead not only to problem resolution but also to a competitive edge. Consequently, a requisite technology is one that can adeptly collect, store, and process data through real-time analysis. In the context of a company's operations, it becomes imperative to consider aspects such as customer behavior, the inherent risks tied to the company's activities, its performance metrics, productivity levels, and its market valuation.

As depicted in Fig. 2, in order for Big Data utilization to hold relevance for an organization, it is imperative that relevant information is first extracted from data sources (with Big Data encompassing these storage sources). This entails employing appropriate data management techniques, while the data could be given in real-time or as archived information. Subsequently, depending on the type of data, dedicated software programs for analysis must be employed. This facilitates the straightforward determination of whether a specific process yields benefits for the respective company or if existing ones can be enhanced. Moreover, this study might present brand-new strategies for dealing with certain problems, favouring the organization's constant adaptation to its operational environment.

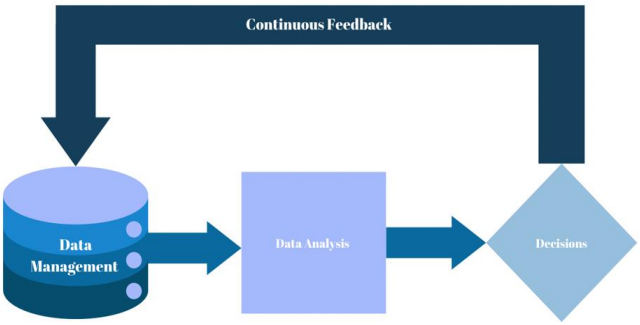

Fig. 2. The "Big Data" concept
Source: own contribution.

b. The importance of "Big Data"

Modern technology, particularly IoT, machine learning, and cloud computing, has significantly facilitated the exponential growth of data and has reshaped how companies comprehend the concept of "Big Data" and employ it to formulate development strategies. The advantages of "Big Data" are most apparent when a company takes the initiative to analyse the collected data and utilizes the insights to generate enhanced business outcomes. "Big Data" provides fresh perspectives that unveil new opportunities and propose novel business models [21].

The potential advantages of "Big Data" [20] are:

✓ it assists businesses in comprehending the market and consumer behaviours, providing them with a clearer insight into the products that can be marketed in specific regions or demographic areas and, consequently, enabling them to outperform their competition.

✓ contributes to customer satisfaction and loyalty by discerning consumer patterns, enabling businesses to attract new customers and discover effective ways to meet customer expectations and maintain their loyalty.

✓ it helps in developing a results-oriented marketing strategy - data analysis streamlines marketing campaigns, allowing companies to

better understand their audience and implement more precise marketing tactics.

✓ it fosters innovation - prudent companies utilize insights from "Big Data" analysis to uncover weaknesses in their production process, aiding in the creation of superior quality products compared to their competitors.

✓ reduces operational costs and time - "Big Data" technologies, especially cloud-based mechanisms, store large volumes of data, which streamlines operational costs and enables faster business operations.

✓ equips companies with the capacity to develop competitive pricing strategies - In the past, conducting competitive analysis presented difficulties, but this is no longer the scenario. "Big Data" presents the chance to scrutinize competitors' approaches and can offer recommendations regarding actions to take or avoid while shaping your unique business strategy. Moreover, aided by "Big Data," it can be easier to identify price fluctuations, aiding in the creation of the most efficient pricing strategy. Ultimately, this grants a greater opportunity to establish a more balanced price in alignment with consumer purchasing behaviour and industry trends.

✓ assists companies in uncovering new revenue streams - an analysis of both consumers and competitors can lead to the discovery of untapped investment opportunities.

✓ serves as a risk analysis tool - "Big Data" analysis can aid in striking a balance between social and economic factors, alongside other external elements, by harnessing predictive insights.

The functioning of information technologies, critical information structures, and the security of data availability, integrity, and confidentiality are crucial for various aspects of modern society such as the business environment, government institutions, transportation, public safety, healthcare, communications, the banking and financial system, emergency services, utilities, and national defence. Even the largest organizations find "Big Data" to be a significant challenge that they cannot overlook. Its enormous potential to improve business decisions, achieve higher precision in customer targeting, and streamline internal processes is undeniable.

Applications of "Big Data"

"Big Data" presents significant opportunities across multiple domains (Fig. 3):

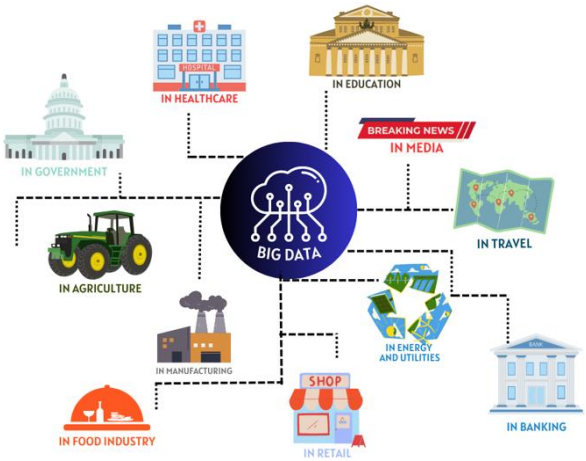

Fig. 3. "Big Data" Applicability in the society
Source: own contribution.

In Government/ public sectors:

Governments across the globe deal with vast amounts of data on a daily basis. This is largely due to the comprehensive updates and records they must maintain regarding their citizens, economic growth, energy resources, and more. This data necessitates thorough examination and analysis, serving as a valuable tool for government operations, primarily in two areas - welfare schemes and cyber security.

Figure 4 shows the fields of the public sector where BigData application is used.

Fig. 4. "Big Data" application in the public sector
Source: own contribution.

In the realm of welfare programs, this data is instrumental in expediting and informed decision-making for political initiatives, identifying areas requiring attention, monitoring agricultural landscapes, and tracking livestock. It also plays a crucial role in addressing national challenges such as terrorism, unemployment, and poverty. When it comes to cybersecurity, analytical tools are employed for tasks such as detecting fraud and apprehending tax evaders.

In agriculture: The utilization of "Big Data" analytics drives advancements in smart farming and precision agriculture practices, leading to cost savings and the emergence of new business opportunities. Crucial domains where big data is applied encompass facilitating the fulfilment of food demand by providing farmers with real-time updates on changes in rainfall, weather patterns, and factors influencing crop yield. It also contributes to enhancing the intelligent and precise utilization of pesticides, aiding farmers in making well-informed decisions about pesticide usage. Moreover, big data aids in efficiently managing farm equipment, optimizing supply chain operations, strategically planning seed planting and chemical application, and ensuring food safety through the collection of data on humidity, temperature, and chemical levels to monitor the health of growing plants.

In education: "Big Data" has revolutionized the education sector by harnessing vast amounts of data encompassing students, faculty, courses, and results. Analysing this data offers valuable insights that enhance educational institutions' operations and effectiveness. This ranges from personalized learning programs, redesigned course materials, and dynamic grading systems to predicting students' career paths. By scrutinizing individual student records, strengths, weaknesses, and interests can be understood, aiding in tailored guidance and suitable career predictions. "Big Data" has overcome the limitation of one-size-fits-all education through e-learning solutions, empowering administrators with analytics and data visualization to optimize university operations, recruitment, and student retention strategies (Fig. 5).

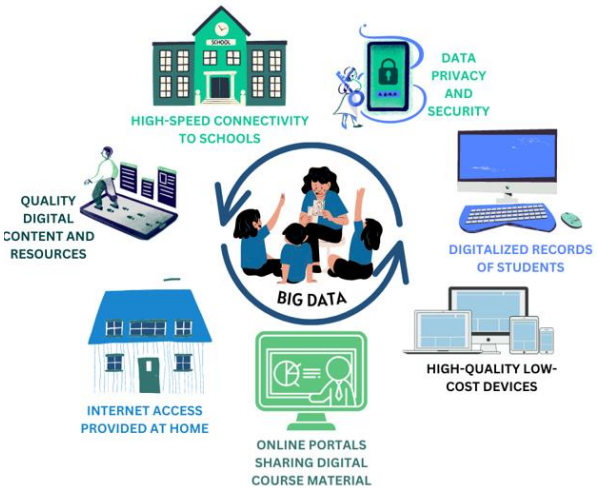

Fig. 5. "Big Data" Application in education
Source: own contribution.

In healthcare: "Big Data" has a pivotal role in advancing contemporary healthcare practices. It has transformed the healthcare sector comprehensively, encompassing cost reduction in treatments, anticipation of epidemic outbreaks, prevention of avoidable illnesses, improvement of overall quality of life, projection of daily patient income for staffing management, integration of electronic health records, implementation of real-time alerts for immediate care, utilization of health data for strategic planning enhancement, and mitigation of fraud and errors within this essential sector.

In media: The enthusiasm surrounding conventional methods of media consumption is gradually waning, giving way to contemporary practices of accessing online content through smart devices, which have emerged as the prevailing trend. These modern approaches are not only instrumental in predicting audience preferences, including genre, music, and content tailored to specific age groups, but also provide valuable insights into customer churn, as they manage to optimize the timing and cost-effectiveness of media streaming schedules, enhance the efficiency of product updates, and significantly contribute to precision in advertising targeting

In travel: Reduced wait times are the consequence of route planning that is

customised to the needs of each individual user thanks to the use of "Big Data". Tools like Google Maps, which provide routes with the least amount of traffic congestion, are examples of how it also helps with congestion management and traffic control. Additionally, "Big Data" data plays a crucial role in identifying accident-prone areas and consequently improving general traffic safety.

In financial institution/ banking: Covering a wide spectrum of functions including fraud detection, streamlining transaction processing, gaining deeper customer insights, optimizing trade execution, and delivering enhanced customer experiences, "Big Data" presents a diverse array of applications. At the same time, banks can use "Big Data" Analytics to gain valuable insights into customer trends, which can be shared with clients, along with the ability to conduct personalized assessments and swiftly generate comprehensive reports (Fig. 6).

Fig. 6. "Big Data" application in financial institution
Source: own contribution.

In manufacturing: "Big Data" has effectively contributed to the improvement of manufacturing processes, enabling tailored product design, ensuring robust quality maintenance, orchestrating efficient supply chain management, and conducting assessments to monitor potential risks.

I**n retail:** Big Data presents an extensive array of uses, participating in forecasting emerging trends, pinpointing relevant customers with precision timing, reducing marketing costs, and elevating the caliber of customer service. It encompasses maintaining holistic consumer perspectives, enabling personalized interactions, refining pricing strategies for optimal trend leverage, streamlining backend operations, and amplifying customer service excellence.

In energy and utilities: Energy and utility systems leverage a variety of Big Data sources, encompassing smart meters, grid infrastructure, weather information, power system metrics, storm data, and Geographic Information System (GIS) data. These platforms make use of this data to achieve cost reduction, enhance operational efficiency, minimize carbon emissions, and effectively manage the energy demand originating from end consumers.

In food industry: Big data assists food platforms in elevating their marketing strategies, curating innovative and highly desirable products, and empowering businesses to monitor competitors' growth rates while maintaining quality control and scrutinizing procurement and pricing choices. This data is also proving valuable for owners by enabling them to monitor factors such as product quality. It allows them to discern whether the product has undergone modifications, such as ingredient substitutions or adjustments in measurements, as well as determine if changes are minor, significant, or influenced by external factors such as seasonal variations or shifts in storage methods.

d. The risks of "Big Data"

While the potential of these datasets is immeasurable, "Big Data" poses risks concerning the protection of personal data and the right to privacy of individuals concerned; among these, we can mention:

- the extent of data collection, tracking, and profiling, considering that data is typically aggregated from various sources, leading to an increased level of detail;
- data protection embedded in products and services (privacy by design/by default), aiming to integrate confidentiality and data protection into the design specifications and architecture of information systems. Innovative and responsible engineering can enable individuals' rights (such as access, objection, restriction, rectification, and data portability) to be exercised effectively;

- data security faces obstacles due to the swift expansion of dataset volume, which can be effectively managed by developing adaptable systems at the data level, where filtering plays a vital role in this context;
- transparency, which may diminish in the absence of proper informing of the individuals subject to automated decisions, as they lack comprehension of the process they are exposed to and have limited control over their data. Individuals should receive clear information about what data is processed, including observed or inferred data about them, how and for what purposes their information is used, including the logic used in algorithms to determine assumptions and predictions. The National Supervisory Authority for Personal Data Processing (ANSPDCP) should have access to details about these automated mechanisms and the potential impact they may have on the rights of the individuals concerned;
- the absence of internal standards/ procedures in accordance with the current legislation regarding personal data protection. Data controllers/data processors are accountable for the decisions they were supposed to make, considering the prevailing legal regulations in this field;
- heightened possibilities of government surveillance and their potential abuses, such as blatant violations of fundamental human rights. Additionally, the volumes of generated data contain extensive and diverse quantities of information about our personal lives, a situation that can clearly lead to deducing behavioral tendencies as well as other individual details, including sensitive data (information about health, sexual orientation, religious beliefs, political affiliations, etc.).;
- discrimination based on data, which is evidently linked to the previously mentioned issues, highlights the inequalities that can arise if this field is not well regulated. When all aspects about an individual are known, legislators must ensure that this will not negatively impact the lives of the individuals concerned.

Companies rely on the data they collect about their customers, so how efficiently employees utilize this data is of paramount importance. The modern era of "digital" companies (such as Google, Facebook, Uber, and Airbnb), centers more around how they use the data they collect rather than what they sell or produce [7]. A significant debate surrounds the relationship between these types of companies and their users. Regulatory frameworks like General Data Protection Regulation (GDPR) have emerged as a recognition of the value users' data holds when accessing such services. However, many users are unaware of the extent of personal data they provide. Most companies collect information about their customers, and whether users feel uneasy about this depends on how that data is used and what they receive in return.

Therefore, "Big Data" must be used in a responsible and sustainable manner, safeguarding the fundamental rights and freedoms of individuals, while also applying and adhering to data protection legislation. Within the European Union, efforts are directed towards regulating and documenting "Big Data" in a fair and ethical manner, ensuring the attainment of maximum value from this dataset while respecting human rights [9]. Data related legislation aims to establish clear rules regarding the utilization of data generated by Internet of Things (IoT) devices, as well as how products are designed ("privacy by design"), enabling the facilitation of rights concerning the processing of personal data on a large scale, and fostering the creation of data value.

CONCLUSIONS

In broad and concise terms, "Big Data" embodies a wide-ranging effort to optimize our interconnected world by grasping and anticipating influential factors simultaneously. it delves into understanding humans reasoning, which, though unique from individual to individual, still exhibits certain patterns. The presence of high-quality and interoperable data from various domains enhances competitiveness and innovation, ensuring sustainable economic growth. The same dataset can be endlessly utilized and

reused for a multitude of purposes without any degradation its quality or quantity.

Moreover, "Big Data" pertains to data that involves greater *variety*, received in increasingly larger *volumes* and at higher *velocity* (“the three Vs”). This type of data is subject to exponential growth, both in its availability and in the automated utilization of digital information. It doesn’t rely only on technology's increasing capability to support the collection and storage of vast data volumes but also on its capacity to analyse, comprehend, and harness the entire value of the data. Despite their potential to address the previously mentioned challenges that may arise, these massive data volumes are entirely unstructured and cannot be stored or processed using relational databases. Analysing "Big Data" contributes positively to organizational performance. The power centers of global companies remain at the forefront of their respective industries by harnessing the potential of "Big Data".

For efficient data collection, processing, and application of specialized analysis methods, organizations need advanced technological infrastructure encompassing hardware and software. In addition, they must engage analytical professionals ("data scientists") who blend programming and statistics expertise, while also possessing advanced domain-specific knowledge pertinent to the organization's sector. These experts must adeptly communicate and present information, often utilizing graphs and reports to provide context to the results.

In the realm of evolving technological advancements, "Big Data" has emerged as an invaluable asset, offering a myriad of benefits and opportunities. The world's most competitive companies harness insights from data analytics to maintain their competitive edge, setting a precedent that businesses of all sizes can emulate for expeditious growth and heightened customer satisfaction. As we navigate this era of information-driven progress, embracing the power of "Big Data" stands as a strategic imperative, propelling organizations toward a future characterized by innovation, efficiency, and strategic advantage.

Hence, datasets have the potential to:

- foster innovation and research by providing real-time access to data that would otherwise have been much harder to obtain and analyze, facilitating comparative processes;
- enable easy adaptation of the service sector to ever-changing customer preferences and needs, through the development of systems capable of analyzing multiple factors regarding consumer patterns over specific intervals of time;
- liberalize the economic sector and enhance productivity by making improved and up-to-date operational business information available to all. This will particularly benefit small and medium-sized enterprises, reducing their costs and helping them tailor their services and products to market demands;
- streamline activities in the public sector through process and communication digitalization;
- simplify people's lives in a digitally advancing world, with each passing day.

REFERENCES

[1]Apache Spark, 2023, Unified engine for large-scale data analytics, https://spark.apache.org/, Accessed on 02.08.2023.

[2]Bogdan, M., Borza, A., 2019, Big Data Analytics as A Strategic Capability: A Systematic Review, Proceedings of the International Management Conference, Vol. 13(1), 575-583. Faculty of Management, Academy of Economic Studies, Bucharest, Romania.

[3]Bogdan, M., Lungescu, D. C., 2018, Is strategic management ready for “Big Data? A review of the “Big Data” analytics literature”, Management Research. Managerial Challenges of the Contemporary Society. Proceedings, 11(2)

[4]DataHero- Crunchbase Company Profile & Funding, 2016,
https://www.crunchbase.com/organization/datahero, Accessed on 01.08.2023.

[5]Gandomi, A., Haider, M., 2015, Beyond the hype: “Big Data” concepts, methods, and analytics, International Journal of Information Management, 35(2), 137-144.

[6]George, G., Osinga, E. C., Lavie, D., Scott, B. A., 2016, Big Data and data science methods for management research, Academy of Management Journal, 59(5), 1493-1507.

[7]Google Cloud, 2023,

https://cloud.google.com/learn/what-is-big-data, Accessed on 25.07.2023.
[8]Hbase/Hadoop, 2023, https://hadoop.apache.org/, Accessed on 28.07.2023.
[9]Hearing on the fundamental rights implications on big data, 2016, https://www.europarl.europa.eu/committees /en/big-data/product-details/, Accessed on 30.07.2023.
[10]IBM SPSS MODELER, 2023, https://www.ibm.com/docs/en/spss-modeler/, Accessed on 28.07 2023.
[11]KNIME analytics platform, 2023, https://www.knime.com/knime-analytics-platform, Accessed on 01.08.2022
[12]Laney, D., 2001, 3d data management: controlling data volume, velocity and variety. Meta group research note, 6(70).
[13]MongoDB- a developer data platform, 2023, https://www.mongodb.com/cloud/atlas/m_source, Accessed on 01.08.2022
[14]ORACLE website, 2023, https://www.oracle.com/ro/big-data/what-is-big-data/, Accessed on 25.07.2022.
[15]Plotly website, 2023, https://plotly.com/, Accessed on 02.08.2022.
[16]Presto website, 2023, https://prestodb.io/docs/current/sql.html, Accessed on 31.07.2023.
[17]QlikView website, 2023, https://www.qlik.com/us/products/qlikview, Accessed on 01.08.2023.
[18]RapidMiner website, 2023, https://rapidminer.com/, Accessed on 31.07.2023
[19]SAP HANA website, 2023, https://www.sap.com/products/technology-platform/hana/, Accessed on 02.08.2023.
[20]Shan, S., Luo, Y., Zhou, Y., 2019, Big Data analysis adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories, Technology analysis & strategic management, Vol. 31(4), 406-420.
[21]Sharda, R., Delen, D., Turban, E., 2013, Business intelligence: a managerial perspective on analytics, Prentice Hall Press.
[22]Splunk Hunk website, 2023, https://www.splunk.com/en_us/blog/learn/splunk-hunk-analytics-for-hadoop.html, accessed on 01.08.2023.
[23]Tableau: business intelligence and analytics software, 2023, https://www.tableau.com/support/releases#main-content, Accessed on 01.08.2023.
[24]The European Data Act, 2022, https://www.eu-data-act.com/, Accessed on 30.07.2023.