

BIG DATA ANALYTICS

Vinnitsia National Technical University

Анотація

У статті йдеться про “великі дані” та їх важливість у сьогоденні, їх зародження, перші приклади використання, основні етапи аналітики, а також поєднання штучного інтелекту для роботи з ними.

Ключові слова: великі дані, штучний інтелект (ШІ), аналітика, інформація.

The article discusses big data and its importance today, its origins, the first examples of its use, the main stages of analytics, and the combination of artificial intelligence to work with it.

Abstract

Keywords: big data, artificial intelligence (AI), analytics, information .

Introduction

Each day, your customers generate an abundance of data. Every time they open your email, use your mobile app, tag you on social media, walk into your store, make an online purchase, talk to a customer service representative, or ask a virtual assistant about you, those technologies collect and process that data for your organization. And that’s just your customers. Each day, employees, supply chains, marketing efforts, finance teams, and more generate an abundance of data, too. Big data is an extremely large volume of data and datasets that come in diverse forms and from multiple sources. Many organizations have recognized the advantages of collecting as much data as possible. But it’s not enough just to collect and store big data—you also have to put it to use. Thanks to rapidly growing technology, organizations can use big data analytics to transform terabytes of data into actionable insights [1].

Big data origins

The earliest examples we have of humans storing and analyzing data are the tally sticks, which date back to 18,000 BCE! The Ishango Bone was discovered in 1960 in what is now known as Uganda and is thought to be one of the earliest pieces of evidence of prehistoric data storage. Paleolithic tribespeople would mark notches into sticks or bones, to keep track of trading activity or supplies. They would compare sticks and notches to carry out rudimentary calculations, enabling them to make predictions such as how long their food supplies would last. Then, in 2400 BCE came, the abacus. The first dedicated device constructed specifically for performing calculations. The first libraries also appeared around this time, representing our first attempts at mass data storage. The ancient Egyptians around 300 BC already tried to capture all existing ‘data’ in the library of Alexandria. Moreover, the Roman Empire used to carefully analyze statistics of their military to determine the optimal distribution for their armies. But, in more recent times it has revolutionized the modern business environment [3].

What is big data analytics?

Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques—like clustering and regression—and apply them to more extensive datasets with the help of newer tools. Big data has been a buzz word since the early 2000s, when software and hardware capabilities made it possible for organizations to handle large amounts of unstructured data. Since then, new technologies—from Amazon to smartphones—have contributed even more to the substantial amounts of data available to organizations. With the explosion of data, early innovation projects like Hadoop, Spark, and NoSQL databases were created for the storage and processing of big data. This field continues to evolve as data engineers look for ways to integrate the vast amounts of complex information created by sensors, networks, transactions, smart devices, web usage, and more. Even now, big data analytics methods are being used with emerging technologies, like machine learning, to discover and scale more complex insights [2].

How big data analytics works?

Big data analytics refers to the process of examining large and varied data sets, or "big data", to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other valuable information that can help organizations make more informed decisions. Here's a general overview of how big data analytics works:

1. **Data Collection:** the process starts with collecting vast amounts of data from diverse sources. These sources may include structured data from databases, unstructured data from social media, sensor data from IoT devices, transactional data, text data, multimedia content, and more.

2. **Data Storage:** after collecting data, it needs to be stored in a way that allows for efficient processing and analysis. Traditional relational databases may not be suitable for big data due to their limitations in scalability and flexibility. Therefore, organizations often use distributed storage systems like Hadoop Distributed File System (HDFS), cloud-based storage solutions, or NoSQL databases like MongoDB, Cassandra, or HBase [6].

3. **Data Processing:** once the data is stored, it undergoes processing to transform it into a usable format. This processing may involve cleaning the data to remove inconsistencies, errors, or irrelevant information. It may also include data normalization, integration, and aggregation to prepare it for analysis.

4. **Data Analysis:** the processed data is then analyzed using various techniques and algorithms to extract insights and patterns. This analysis can range from simple statistical analysis to more complex techniques such as machine learning, data mining, natural language processing, and predictive modeling.

5. **Visualization and Interpretation:** the insights derived from the data analysis are often visualized using charts, graphs, dashboards, and other visual aids to make them easier to understand. Visualization helps stakeholders interpret the findings and gain actionable insights from the data.

6. **Decision Making:** the final step involves using the insights gained from big data analytics to make informed decisions. These decisions could relate to business strategies, product development, marketing campaigns, risk management, resource allocation, and more.

What is importance of big data?

1. **Decision-Making:** big data analytics empowers organizations to make informed decisions based on data-driven insights rather than relying solely on intuition or past experiences. By analyzing large and diverse datasets, businesses can uncover hidden patterns, correlations, and trends that may not be apparent through traditional methods. This enables them to anticipate market demands, optimize operational processes, and identify new opportunities for growth.

2. **Enhanced Customer Experience:** in the era of hyper-personalization, understanding customer preferences and behavior is paramount. Big data analytics enables businesses to gain a comprehensive view of their customers by analyzing data from various touchpoints such as social media interactions, website visits, purchase history, and customer service inquiries. This insight allows organizations to tailor their products, services, and marketing strategies to meet the unique needs and preferences of individual customers, thereby enhancing the overall customer experience.

3. **Improved Operational Efficiency:** big data analytics can drive efficiencies across all facets of business operations. By analyzing operational data in real-time, organizations can optimize processes, streamline supply chains, minimize downtime, and reduce costs. Predictive analytics can also help businesses anticipate maintenance needs, identify potential bottlenecks, and optimize resource allocation, leading to greater efficiency and productivity.

4. **Innovation and Product Development:** big data serves as a catalyst for innovation by providing valuable insights into market trends, consumer behavior, and emerging technologies. By analyzing market data and customer feedback, businesses can identify unmet needs and develop innovative products and services that resonate with their target audience. Moreover, big data analytics can facilitate rapid prototyping, testing, and iteration, allowing organizations to bring new products to market more quickly and effectively.

5. **Risk Management and Fraud Detection:** in an increasingly complex and interconnected world, managing risks and detecting fraudulent activities is a critical concern for businesses across industries. Big data analytics enables organizations to identify potential risks, anticipate market fluctuations, and detect anomalies in real-time. By leveraging advanced analytics techniques such as machine learning and anomaly detection, businesses can proactively mitigate risks, safeguard against fraud, and ensure regulatory compliance [5].

6. Scientific Discovery and Research: big data is revolutionizing scientific research and discovery across various fields, from genomics and climate science to particle physics and astronomy. By analyzing massive datasets generated by experiments, simulations, and observations, researchers can uncover new insights, validate hypotheses, and advance our understanding of the world around us. Big data analytics also facilitates collaboration and knowledge sharing among researchers, accelerating the pace of scientific discovery and innovation.

Artificial Intelligence for Big Data

Artificial Intelligence (AI) plays a central role in Big Data, contributing in several essential ways. AI-driven algorithms automate the data analysis process, resulting in significant time savings and reduced human error. These algorithms efficiently handle vast datasets, unveiling hidden patterns and trends that might otherwise remain unnoticed.

It also excels in predictive analytics, utilizing historical data to make informed predictions. Whether forecasting customer behavior, equipment failures, or market trends, AI empowers decision-making with actionable insights. They are adept at detecting anomalies within datasets, a critical capability for tasks like fraud detection, network security, and quality control.

AI-powered recommendation systems leverage Big Data to offer personalized content and product suggestions, as exemplified by Netflix and Amazon. Lastly, Natural Language Processing (NLP) in AI enables organizations to analyze and comprehend customer sentiment, feedback, and textual opinions, contributing to product and service enhancements.

AI for Big Data has made significant impacts across various industries:

1. **Healthcare:** AI is used to analyze patient data, assist in diagnosing diseases, predict patient outcomes, and even personalize treatment plans based on individual health records.
2. **Finance:** Financial institutions utilize AI for fraud detection, algorithmic trading, credit risk assessment, and customer service chatbots.
3. **Retail:** AI-driven recommendation engines personalize shopping experiences, optimize inventory management, and provide dynamic pricing strategies.
4. **Manufacturing:** Predictive maintenance powered by AI reduces downtime by forecasting equipment failures, while quality control systems enhance product quality.
5. **Marketing:** AI enhances marketing campaigns by analyzing customer behavior, segmenting audiences, and optimizing ad targeting [4].

Conclusion

In summary, the importance of big data analytics lies in its ability to unlock valuable insights from vast and varied datasets, driving informed decision-making, enhancing customer experiences, improving operational efficiency, fostering innovation, mitigating risks, and advancing scientific discovery. By harnessing the power of big data analytics, organizations can gain a competitive edge, adapt to changing market dynamics, and propel growth in the digital age.

REFERENCES

1. What is Big Data. URL: <https://www.oracle.com/big-data/what-is-big-data/>
2. Big Data Analytics. URL: <https://www.tableau.com/learn/articles/big-data-analytics>
3. The history of Big Data. URL: <https://www.lightsondata.com/the-history-of-big-data/>
4. Artificial Intelligence for Big Data. URL: <https://www.questionpro.com/blog/artificial-intelligence-for-big-data/>
5. Importance of Big Data. URL: <https://www.techtarget.com/searchdatamanagement/definition/big-data>
6. The best Big Data tools for storage. URL: <https://www.turing.com/resources/best-big-data-platforms#b.-data-storage>

Савелко Ростислав Олегович – студент 2 курсу Вінницького національного технічного університету, факультету інформаційних технологій та комп'ютерної інженерії, групи ІІІ-226, Вінниця, e-mail: rostiksavelko@gmail.com.

Науковий керівник: *Кот Сергій Олександрович* — викладач кафедри іноземних мов, Вінницький національний технічний університет, м. Вінниця, e-mail: kot.sergii@vntu.edu.ua.

Savelko Rostyslav Olehovich — second year student of Vinnytsia National Technical University, Faculty of Information Technology and Computer Engineering, Group 1PI-22b, Vinnytsia, e-mail: rostiksavelko@gmail.com.

Supervisor: **Kot Serhiy Oleksandrovykh**— an Assistant Professor of Foreign Languages Department, Vinnytsia National Technical University, Vinnytsia, e-mail: kot.sergii@vntu.edu.ua.