

DATA MINING – A WAY TO FIND NEW KNOWLEDGE

Klimets A.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Rogachevskaya A.I. – Senior Lecturer

Annotation. Data Mining is a relatively young branch of Data Science, which has taken a special place with the development of the Internet. This method is widely used in the modern world, in particular for analyzing users behavior and identifying their interests, since classical methods of data analysis are rather trivial and do not allow you to quickly respond to changes in users desires. Data Mining is not a specific technology, but a collection of different methods for solving specific problems.

Keywords. Data Mining, Data Science, knowledge discovery method, data analysis.

A large number of messages is generated every day for example, according to the research conducted in 2020 at Domo [1], 41,666,667 text messages are generated on Whatsapp, and YouTube receives 500 hours of video every minute. All created information flows are analyzed, processed and applied in various fields of activity: advertising, forecast of successful campaigns.

Data Mining is a research and discovery of hidden knowledge that was not previously known, non-trivial, practically useful, available for human interpretation in raw data [2]. In this area, both commercial for example, MatLab, Statistica and free for example, Weka, R, etc. specialized tools have been widely developed. However, they have both advantages designed for specific tasks and disadvantages: complicated interface with many parameters, operation only on a personal computer.

Data Mining is not the only method, it is a collection of a large number of different knowledge discovery methods. All tasks solved by Data Mining methods can be conditionally divided into six types (Figure 1):

- classification;
- regression;
- clustering;
- association;
- identification of anomalies;
- summation.

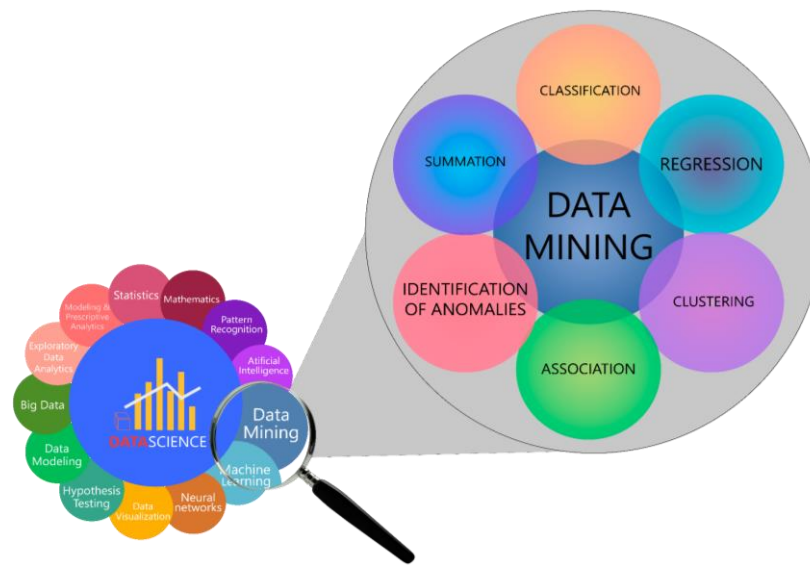


Figure 1 – The place of Data Mining in modern science

Main target of the Data Mining process is identification of trends and patterns that are impossible to find using classic analysis methods. Generally there are 4 steps of Data Mining process (Figure 2): problem definition, data gathering and preparation, model building and scoring of the result, knowledge deployment.

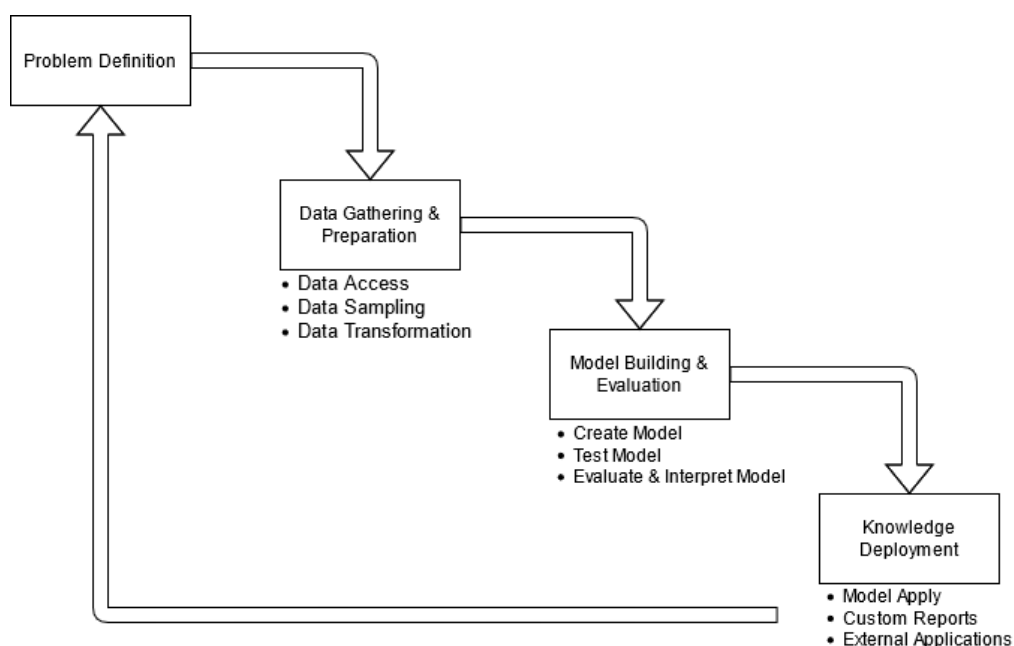


Figure 2 – Structure of Data Mining process

Data Mining is multidisciplinary in its origin, since it includes elements of numerical methods, mathematical statistics and probability theory, information theory and mathematical logic, artificial intelligence and machine learning. The scope of Data Mining is quite wide, however, it is not unlimited, despite the use of a large number of previously created data analysis and processing methods. Data Mining has the following limitations. First of all, Data Mining will not be beneficial without proper management. The Data Mining process will give results only in the case of clear understanding of the specific features of the business itself, data, analytics methods. Data mining allows to discover new ways of using company data, but direct analysis to find patterns is necessary for determination of their value. Secondly, it is important to remember that the predicted relationship between consumers and the brand is not the reason for consumers actions.

Data Mining, like any scientific method, has its advantages and disadvantages, some of which are revealed only while solving real problems. Advantages of Data Mining are as follows:

- Data Mining enables organizations to make lucrative modifications in operation and production;
- compared with other statistical data applications, Data Mining is cost-efficient;
- Data Mining helps in decision-making processes of an organization;
- it facilitates the automated discovery of hidden patterns as well as the prediction of trends and behaviors;
- it is a quick process that makes it easier for new users to analyze enormous amounts of data in a short time.

Disadvantages of Data Mining:

- most of Data Mining tools are difficult to operate and they need advanced training to work on;
- different Data Mining instruments operate in various ways due to the use of different algorithms, that's why the selection of the right Data Mining tools is a very challenging task
- the Data Mining techniques are not precise, so this may lead to severe consequences in certain conditions.

Data Mining is primarily used by organizations with high consumer demand such as retail, communications, finance, and marketing that need to determine prices, consumer preferences, position products, and impact on sales, customer satisfaction, and corporate profits. Data Mining enables a retailer to use point-of-sale records of customer purchases to develop and promote products. All this helps the organization to attract more customers.

References:

1. Data Never Sleeps 8.0 [Electronic resource]. – Mode of access: <https://www.domo.com/learn/data-never-sleeps-8>. – Date of access: 19.03.2021
2. Анализ данных и процессов: учеб. пособие / А.А. Барсегян [и др.]. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.