МРНТИ 06.01.11
УДК 338.2
DOI 10.58319/26170493_2024_3_68

MODERN METHODS OF MARKET PROCESSES ANALYSIS

KOLESNIKOVA K.V.*[1]
Doctor of Technical Sciences, Professor

MUKHAMEDIYEVA A.G.[1]
Candidate of Economic Sciences

ALPYSBAEV K.S.[1]
Candidate of Economic Sciences
[1]International Information Technologies University, Almaty, Republic of Kazakhstan

ABSTRACT. The article discusses the use of Data Mining methods for analyzing market processes. The main focus is on how data mining technologies can be used to identify hidden patterns and trends in the market, which contributes to making more informed management decisions and developing effective strategies. Examples of real applications demonstrate how the analysis of large amounts of data allows optimizing marketing campaigns and conducting market basket analysis. The conclusion emphasizes the importance of Data Mining for modern business and the need to integrate these methods into the daily practice of companies to increase their competitiveness in the market.

KEYWORDS: business process, customer segmentation, marketing research, data mining, data analysis.

НАРЫҚ ПРОЦЕСТЕРІН ТАЛДАУДЫҢ ЗАМАНАУЫ ӘДІСТЕРІ

КОЛЕСНИКОВА К.В.*[1]
техника ғылымдарының докторы, профессор

МУХАМЕДИЕВА А.Г.[1]
экономика ғылымдарының кандидаты

АЛПЫСБАЕВ К.С.[1]
экономика ғылымдарының кандидаты
[1]Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан Республикасы

АҢДАТПА. Мақалада нарықтық процестерді талдау үшін Data Mining әдістерін пайдалану қарастырылады. Деректерді алу технологиялары нарықтағы жасырын заңдылықтар мен үрдістерді анықтау үшін қалай қолданылуы мүмкін екендігіне аса назар аударылады, бұл неғұрлым негізделген басқару шешімдерін қабылдауға және тиімді стратегияларды әзірлеуге ықпал етеді.Нақты қосымшалардың мысалдары деректердің үлкен көлемін талдау маркетингтік науқандарды оңтайландыруға, нарықтық қоржындарға талдау жүргізуге мүмкіндік беретінін көрсетеді. Қорытындыда Data Mining-дің қазіргі заманғы бизнес үшін маңыздылығы және компаниялардың нарықтағы бәсекеге қабілеттілігін арттыру үшін осы әдістерді күнделікті тәжірибиеге еңгізу қажеттілігі аталады.

ТҮЙІН СӨЗДЕР: бизнес-процесс, клиенттерді сегменттеу, маркетингтік зерттеу, data mining, деректерді талдау

СОВРЕМЕННЫЕ МЕТОДЫ АНАЛИЗА РЫНОЧНЫХ ПРОЦЕССОВ

КОЛЕСНИКОВА К.В.*[1]
доктор технических наук, профессор

МУХАМЕДИЕВА А.Г.[1]
кандидат экономических наук

АЛПЫСБАЕВ К.С.[1]
кандидат экономических наук
[1]Международный университет информационных технологий, г. Алматы, Республика Казахстан

АННОТАЦИЯ. В статье рассматривается использование методов Data Mining для анализа рыночных процессов. Основное внимание уделяется тому, как технологии извлечения данных могут применяться для выявления скрытых закономерностей и тенденций на рынке, что способствует принятию более обоснованных управленческих решений и разработке эффективных стратегий. Примеры реальных приложений демонстрируют, как анализ больших объемов данных позволяет оптимизировать маркетинговые кампании, проводить анализ рыночных корзин. В заключении подчеркивается значимость Data Mining для современного бизнеса и необходимость интеграции этих методов в ежедневную практику компаний для повышения их конкурентоспособности на рынке.

КЛЮЧЕВЫЕ СЛОВА: бизнес-процесс, сегментирование клиентов, маркетинговое исследование, data mining, анализ данных.

INTRODUCTION. An important detail characterizes modern society:

- It has reached a point where vast amounts of information (in corporations, government departments, social security services, etc.) have been simultaneously concentrated in the form of databases.
- The power of computing equipment, systems, and communication and data transmission have increased significantly.
- The Internet has become humanity's only common repository of data and knowledge.

The demand for innovative ways to engage with data and its processing tools has become a pressing need in our evolving society. Modern business cannot function without information and analytical support, solving forecasting problems, risk management, etc.

The methods of mathematical statistics were unable to solve problems in the informational society and remained useful only for verification-driven data mining, testing pre-formulated hypotheses, and online analytical processing (OLAP) "rough" intelligence analysis, which forms the basis of operational, analytical data processing.

MATERIAL AND METHODS OF RESEARCH. By 2025, the volume of all data worldwide will be 163 zettabytes (ZB). The forecast was published in the report of the analytical firm IDC, "The Age of Data 2025" [1]. One zettabyte is 1021 bytes. That is, the total volume of information will be

163*1021 bytes. To understand and find something helpful in this ocean of information, it is necessary to apply analysis methods data or Data Mining widely.

Modern Data Mining technology is based on templates (patterns) that reflect fragments of multi-aspect relationships in data. These templates represent regularities inherent in sub-elections data that can be compactly expressed in an understandable form. The search for patterns is carried out by methods not limited to the framework of prior assumptions about the sample structure and in the form of distributions of the values of the analyzed indicators.

Scope of application Data Mining she is not limited by anything - she is applicable wherever there is data. The specifics of modern requirements for processing such a volume of information:

- The data has unlimited volume.
- The data are heterogeneous (quantitative, qualitative, textual).
- The results must be specific and clear.
- Tools for processing raw data should be easy to use.

The main task of the analyst is to generate hypotheses. He solves it based on his knowledge and experience. However, the accumulated data that is being analyzed also contains knowledge. Such knowledge is often called "hidden" since it includes gigabytes and terabytes of information that a person cannot explore independently. In this regard, there is a high probability of losing hypotheses that can bring significant benefits. To identify hidden knowledge, it is necessary to use unique methods of automatic analysis, with the help of which it is essential to extract knowledge from the "heaps" of information practically. The term data mining has firmly established itself in this direction. The classic definition of this term was given in 1996 by one of the founders of this direction - Grigory Pyatetsky–Shapiro: "... Data Mining is the study and discovery by a "machine" (algorithms, artificial intelligence) in raw data of hidden knowledge that was previously unknown, non-trivial, practically useful, and accessible to human interpretation".

The data accumulated by enterprises and organizations in databases and other sources (so-called business data) has its characteristics. Business data is rarely collected specifically to solve analytical problems. Companies and organizations collect data for commercial purposes: record keeping, financial analysis, reporting, decision-making, etc. In this way, business data differs from experimental data collected for research purposes. The primary consumers of business data are usually decision-makers in companies. Business data usually contains errors, anomalies, contradictions, and omissions. This is a consequence of companies not collecting data for analysis; they have errors of various natures, which reduces their quality.

An essential provision of data mining is the non-triviality of the patterns sought. This means the found patterns should reflect non-obvious, unexpected regularities in the data, constituting the so-called hidden knowledge.

Data Mining Standards (sometimes called standards for data mining) ensure compatibility, efficiency, and quality of processes and tools used in data analysis. Standards affect three main aspects of Data Mining. First, the unification of interfaces by which any application can access the functionality of Data Mining. Here, two directions have emerged. These are the standardization of interfaces for object programming languages (CWM Data Mining, JDM, OLE DB for Data Mining) and attempts to develop an add-on for the SQL language, allowing access to the Data Mining toolkit built directly into the relational database (SQL/MM, OLE DB for Data Mining) [2].

The second aspect of standardization is the development of a single agreement on storing and transferring Data Mining models. It is easy to guess that the basis for such a standard is the XML language. The standard itself is called PMML (Predicted Model Markup Language). Finally, there is the CRISPR standard, which provides recommendations for organizing the data mining process.

1. Standard CWM (Common Warehouse Metamodel is a standard developed by the OMG consortium for exchanging metadata between different software products and repositories involved in creating corporate DSS. It is based on open object-oriented technologies and standards and uses UML (Unified Modeling Language) as a modeling language, XML and XMI (XML Metadata Interchange) for exchanging metadata, and the Java programming language for implementing models and specifications [3]. Currently, there is only one officially recognized standard.

2. The CRISP standard was developed in 1996 by Daimler-Benz (now DaimlerChrysler), Integral Solutions Ltd. (ISL), NCR, and OHRA. A year later, a consortium was formed to develop a CRISP-DM standard independent of the industry, application area, and tools used.

3. The CRISP-DM standard is described in terms of a hierarchical process model. The model consists of tasks described at four levels of abstraction (from more general to more specific): phases, everyday tasks, specialized tasks, and process examples. Several development phases are distinguished at the top level of

the Data Mining process. Each includes several everyday tasks related to the second level of the hierarchy. The second level is called general because the functions that make it up need to consider the specifics of the application area for which they are solved. They are assumed to be complete and unchangeable. Completeness means coverage of both the entire process and possible Data Mining applications. In turn, immutability means the model should be relevant for previously unknown Data Mining methods [4].

4. Here, sufficient attention should be paid to the project objectives in terms of business prospects, knowledge definition in the formulation of the Data Mining problem, and the prospects for further development of the initial plan for achieving the objectives. It is essential to fully understand the business for which the solution is being sought to understand what data and how it should be analyzed in the future. This phase includes the following tasks:

- defining business goals;
- definition of the situation;
- defining Data Mining goals;
- creating a project plan.

The first crucial task in the data mining process is to thoroughly analyze the client's true goals. This step is paramount as it guarantees that the resulting system will effectively address the problems that the user is primarily concerned about. To achieve this, it is imperative to identify the initial business goals and correctly formulate the corresponding questions.

The next task is to formulate the Data Mining goals in business terms, for example, "based on three years of purchasing information and demographic information, predict how much of a product a consumer will buy". Success in achieving these goals should also be described in these terms. For example, success is reaching a certain level of prediction accuracy. If the business goals cannot be effectively translated into Data Mining goals, this may be a reason to review the problems being solved. The last task solved in this phase is to draw up a project plan that sequentially reveals the intentions for achieving the Data Mining goals, including an outline of specific steps, time intervals, an initial assessment of potential risks, and the necessary tools and methods to support the project in working order. It is generally accepted that 50-70% of the time and effort in developing a Data Mining project is spent on the data preparation phase; 20-30% on the data understanding phase; 10-20% on each of the modeling, assessment, and business understanding phases; and 5-10% on the deployment phase [5].

The PMML standard (Predicted Model Markup Language) plays a significant role in the data mining process. It is designed to facilitate the exchange of built mining models between Data Mining systems. This standard describes the form of representation of models in the form of an XML document, thereby enhancing interoperability.

The PMML standard (Predicted Model Markup Language) plays a significant role in the data mining process. It is designed to facilitate the exchange of built mining models between Data Mining systems. This standard describes the form of representation of models as an XML document, thereby enhancing interoperability.

A PMML document can describe multiple models. In addition, the list of models in a PMML document can be empty. Such a document can be used for purposes other than transferring models, such as transferring primary metadata for building a model.

Some types of PMML models, such as neural networks or regression, can be used for different purposes. Some implement numerical predictions, while others can be used for classification [6].

Other Data Mining Standards. In late 1991 - early 1992, developers of text search systems, acting under the auspices of the IEEE organization, implemented a specification of a language called SFQL (Structured full-text Query Language). The goal of SFQL was to describe an extension to the SQL language that could be used in full-text documents. The sixth part of this standard, SQL/MM Data Mining, is devoted to the Data Mining process. It attempts to provide a standard interface for Data Mining algorithms. They can represent the top level of any object-relational database system and an intermediate level. This standard supports four main Data Mining models [7]:

- rule model – allows you to find patterns (rules) in the relationships between different parts of the data;
- cluster model - helps to group data records that have common characteristics and identify the more important of these characteristics;
- regression model – helps the analyst predict the meaning of new numerical data based on known data;
- classification model – similar to a regression model but focuses on predicting categorical rather than numerical data (classes).

Microsoft developed the Microsoft Data Mining Extensions (DMX) standard. Like the SQL/MM language, it applies Data Mining methods to relational databases. This standard extends Microsoft's OLE DB and is included in SQL Server 2005 Analysis Services. The Microsoft Data Mining Extensions (DMX) interface is designed to be used as a data mining interface and user interface (UI) management tool. The solution proposed by Microsoft in SQL Server 2005 allows several data mining algorithms to be used as extensions of this interface [8].

The Java Data Mining standard aims to develop a Java API for data mining developers. It combines the efforts of two groups: JSR 74 and JSR 247. JSR 74's goal is the Java Data Mining API (JDMAPI) standard. It will specify API functions for building data mining models, extracting knowledge using these models, and creating, storing, accessing, and storing data and metadata that support the results of data mining and data transformation choices.

In e-business systems, where special meanings have questions about attracting and retaining customers, Data Mining technologies are often used to build recommendation systems for online stores and to solve the problem of personalizing website visitors. Recommendations for products and services, built based on patterns in customer purchases, have a substantial persuasive effect force. Statistics show that almost every visitor to the Amazon store takes the opportunity to look at what other customers have bought [9].

Personalization of clients or, in other words, automatic recognition of the client's belonging to a specific target audience allows the company to conduct a more flexible marketing policy. Since e-commerce money and payment systems are also electronic, ensuring security in transactions with plastic cards becomes an important task. Data Mining allows you to identify cases of fraud (fraud detection). In e-commerce, all Data Mining methodologies developed for conventional marketing remain valid. In addition, this area is closely related to Web Mining. The specificity of Web Mining lies in using traditional Data Mining technologies to analyze highly heterogeneous, distributed, and significant in-volume information on Web sites. The distinction can be made here between two directions: web content mining and web usage mining.

In the first case, we are talking about automatic search and extraction of quality information from Internet sources overloaded with "information noise", as well as all sorts of tools and automatic classification and annotation of documents. Web Usage Mining aims to identify patterns in the behavior of users of a specific Website (group of sites), mainly which pages are used in what time sequence and by which groups of users are asked [10].

Data Mining allows for solving problems of identifying groups of consumers with similar behavioral stereotypes, i.e., segmenting the market. In addition, it will enable market basket analysis and sequential analysis.

Knowing the relationship between purchases and time patterns allows you to optimally regulate supply. Sequential analysis helps retailers make decisions about stockpiling. It answers questions like, "If a customer buys a video camera today, is he likely to buy a new battery pack in the future?"

Using Data Mining technologies to analyze clients' profitability and risk (churn prevention) and protect against fraud saves telecommunications companies huge amounts of money and resources. One of the standard data mining methods is the analysis of records using the detailed characteristics of calls. The purpose of such analysis is to identify categories of clients with similar stereotypes of using services and to develop attractive sets of prices and services [11].

A classic example of the practical application of Data Mining is the solution to the problem of possible insolvency of bank clients. This issue, which worries any employee of the bank's credit department, can be resolved intuitively. If the client's image in the bank employee's mind corresponds to his idea of the client's creditworthiness, then the loan can be issued; otherwise - it is refused. According to a similar scheme, but more productively and fully automatically, the decision support systems (Decision System Support) with built-in Data Mining functionality installed in thousands of American banks work. Free from subjective bias, they rely on their work only on the bank's historical database, where detailed information about each client is recorded and, ultimately, the fact of his creditworthiness. Data Mining classification algorithms process this data, and the results are used for decision-making. Credit risk analysis consists, first, of assessing the borrower's creditworthiness. This problem is solved based on the analysis of accumulated information, i.e., the credit history of "old" clients. Using Data Mining tools (decision trees, cluster analysis, neural networks, etc.), a bank can obtain profiles of reliable and unreliable borrowers. In addition, it is possible to classify a borrower by risk groups, which means deciding on the possibility of lending and setting a credit limit, interest rates, and repayment period.

In insurance, as in banking and marketing, the task is to process large amounts of information to identify typical client groups (profiles). This information is used to offer specific insurance services with the least risk to the company and, possibly, with the greatest benefit to the client.

One of the most promising areas of data mining applications is using this technology in analytical CRM. CRM (Customer Relationship Management) – customer relationship management [12].

When these technologies are used together, knowledge mining is combined with "money mining" of customer data.

An essential aspect of the work of the marketing and sales departments is the formation of a holistic view of customers, information about their features, characteristics, and the structure of the customer base. CRM uses so-called customer profiling, which provides a complete view of all the necessary customer information. Customer profiling includes the following components: customer segmentation, customer profitability, customer content, and customer reaction analysis. Each of these components can be studied using Data Mining, and their analysis, as profiling components, can ultimately provide the knowledge that is impossible to obtain for each characteristic.

RESULTS AND THEIR DISCUSSION. Considering the above processes, we can highlight several key aspects that reflect the significance of the study. Modern companies face the challenges of forecasting, risk management, and business process optimization. Data Mining helps identify hidden patterns and trends, contributing to making informed management decisions and developing effective strategies. Integrating data mining methods into companies' daily practice is a prerequisite for increasing their competitiveness in the market. Companies that use analytical tools in a highly competitive environment gain a significant advantage. Data Mining methods can automate data analysis processes, which allows enterprises to use their resources more efficiently and implement innovations. This is especially important when automation is critical to success in the digital age. The methods discussed in the study are used to create recommendation systems and personalize marketing efforts. This allows companies to understand their customers' preferences better and develop more targeted marketing campaigns, increasing their efficiency and profitability.

Using data Mining solves the task of segmenting customers based on their profitability. The analysis identifies those segments of buyers that bring the most significant profit. Segmentation can be carried out based on customer loyalty. As a result of segmentation, the entire customer base will be divided into specific segments with common characteristics. Through these characteristics, the company can select a marketing policy for each group of customers individually.

Data mining technology can also be used to predict the reaction of a particular segment of customers to a specific type of advertising or promotion based on retrospective data accumulated in previous periods [13].

CONCLUSION. Thus, by determining customer behavior patterns using Data Mining technology, it is possible to significantly increase the efficiency of the marketing, sales, and distribution departments. When combining CRM and Data Mining technologies and competently implementing them into business, the company gains significant advantages over its competitors

REFERENCES:

1. Hornik, K., Gun, B. & Hahsler M.A. (2005). Rules – a computational environment for mining association rules and frequent item sets. *Journal of Statistical Software, 14*(15), 1-25.

2. Fayyad, U. & Piatetsky-Shapiro, G. (1996). *Advances in knowledge Discovery and Data Mining.* AAAI Press.

3. Hand, D. (2006). Classifier technology and illusion of progress. *Statistical Science, 21*, 1-14.

4. Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning.* Create Space Independent Publishing Platform. http://www.sthda.com/

5. Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 5,* 113-142.

6. Larose, D.T. (2006). *Data Mining methods and models.* New Jersey: Published by John Wiley & Sons.

7. Hastie, T., Tibshirani, R. & Friedman J. (2009). *The elements of statistical learning: Data Mining, inference and prediction.* NY: Springer-Verlag.

8. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59-69.

9. Leemans, S.J., Fahland, D. & van der Aalst, W.M.P. (2018). *Software & systems modeling, 17*(2), 599-631.

10. Redmore, S. (2019). *Machine learning for natural language processing*. https://www.lexalytics.com/lexablog/machine-learning-vs-natural-language-processing-part-1

11. Morozov, V., Mezentseva, O., Steshenko, G. & Proskurin M. (2020). *Product development of start-up through modeling of customer interaction based on Data Mining*. Communications in Computer and Information Science, 399-415.

12. Kolesnikova, K., Mezentseva, O. & Savielieva, O. (2021). Neural network simulation model of realization of the business analysis process. *Communication and Intelligent Systems Lecture Notes in Networks and Systems, 204*.

13. Van der Aalst, W.M.P. & van Dongen B.F. (2002). *Discovering workow performance models from timed logs.* International conference on engineering and deployment of cooperative information systems. Verlag, Berlin, 45-63.

14. Alpysbayev, K.S., Mukhamedyeva, A.G. & Kolesnikova, K.V. (2024). Digital economy: features, trends and guidelines. *Education. Quality Assurance, 2*(35), 72-78.

СВЕДЕНИЯ ОБ АВТОРАХ:

Kolesnikova Kateryna* - Doctor of Technical Sciences, Professor, Vice-Rector for Research Activities, JSC «International Information Technology University», Almaty, Republic of Kazakhstan
E-mail: kkolesnikova@iitu.edu.kz

Mukhamediyeva Ardak - Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, JSC «International Information Technology University», Almaty, Republic of Kazakhstan
E-mail: amukhamediyeva@iitu.edu.kz

Alpysbayev Kaisar - Candidate of Economic Sciences, Associate Professor of the Department of Economics and Business, JSC «International Information Technology University», Almaty, Republic of Kazakhstan
E-mail: kalpysbayev@iitu.edu.kz

Колесникова Катерина Викторовна* – техника ғылымдарының докторы, профессор, ғылыми жұмыстар жөніндегі проректоры, «Халықаралық ақпараттық технологиялар университеті» АҚ, Алматы қ., Қазақстан Республикасы
E-mail: kkolesnikova@iitu.edu.kz

Мухамедиева Ардак Габитовна – экономика ғылымдарының кандидаты, экономика және бизнес кафедрасының қауымдастырылған профессоры, «Халықаралық ақпараттық технологиялар университеті» АҚ, Алматы қ., Қазақстан Республикасы
E-mail: amukhamediyeva@iitu.edu.kz

Алпысбаев Кайсар Серикулы – экономика ғылымдарының кандидаты, экономика және бизнес кафедрасының қауымдастырылған профессоры, «Халықаралық ақпараттық технологиялар университеті» АҚ, Алматы қ., Қазақстан Республикасы
E-mail: kalpysbayev@iitu.edu.kz

Колесникова Катерина Викторовна* – доктор технических наук, профессор, проректор по научно-исследовательской деятельности, АО «Международный университет информационных технологий», г. Алматы, Республика Казахстан
E-mail: kkolesnikova@iitu.edu.kz

Мухамедиева Ардак Габитовна – кандидат экономических наук, ассоциированный профессор кафедры экономики и бизнеса, АО «Международный университет информационных технологий», г. Алматы, Республика Казахстан
E-mail: amukhamediyeva@iitu.edu.kz

Алпысбаев Кайсар Серикулы - кандидат экономических наук, ассоциированный профессор кафедры экономики и бизнеса, АО «Международный университет информационных технологий», г. Алматы, Республика Казахстан
E-mail: kalpysbayev@iitu.edu.kz