

УДК 004.6

DOI: 10.33813/2224-1213.23.2020.8



Чубукова Ольга Юріївна,

*доктор економічних наук, професор,
завідувач кафедри економічної кібернетики та маркетингу,
Київський національний університет технологій та дизайну
kek@knutd.com.ua*

Пономаренко Ігор Віталійович,

*кандидат економічних наук, доцент,
доцент кафедри економічної кібернетики та маркетингу,
Київський національний університет технологій та дизайну
igor_chip@ukr.net*



Недбайло Наталія Володимирівна,

*магістр кафедри економічної кібернетики та маркетингу,
Київський національний університет технологій та дизайну
tatakovtun@ukr.net*

ВИКОРИСТАННЯ DATA MINING ДЛЯ ОБРОБКИ БІЗНЕС-ДАНИХ

Предмет дослідження – підхід щодо можливості застосування методів data mining у рамках бізнес-аналітики з метою оптимізації прийняття управлінських рішень компанією. **Метою написання статті** є дослідження особливостей використання методів data mining для обробки первинних даних, які виступають у якості цінного ресурсу компанії, що може бути застосований для забезпечення

конкурентоздатності на певному ринку. **Методологія проведення роботи** – системно-структурного та порівняльного аналізів (для дослідження напрямів використання data mining для комплексного дослідження первинних даних); монографічного (при вивченні передумов зростання використання компаніями data mining у процесі дослідження даних); економічного аналізу (при оцінюванні доцільності застосування методів машинного навчання для забезпечення цілей бізнес-аналітики). **Результати роботи** – визначено особливості застосування data mining як одного з напрямів бізнес-аналітики, що дає можливість на основі первинних даних визначати неявні взаємозв'язки між відомими факторними та результативними ознаками. Розкрито основні напрями здійснення маніпуляцій з даними: класифікація та прогнозування, а також кореляційно-регресійний аналіз. Доведено важливість використання основних методів статистичного аналізу в процесі дослідження первинних даних. Наведено специфіку застосування нейронних мереж як одного з найважливіших методів машинного навчання. **Висновки** – використання data mining дозволяє компаніям збільшувати ефективність застосування наявних даних та відповідним чином оптимізувати стратегії розвитку. Наявність великої кількості методів машинного навчання та статистичних підходів розширює можливості комплексного аналізу даних. Важливу роль у даному випадку відіграють інноваційні технології та спеціалізовані мови програмування.

Ключові слова: data mining, дані, ефективність, економіка, машинне навчання, нейронні мережі, статистичний аналіз, управлінське рішення.

Чубукова Ольга Юрьевна,

доктор экономических наук, профессор, заведующий кафедры экономической кибернетики и маркетинга Киевского национального университета технологий и дизайна

Пономаренко Игорь Витальевич,

кандидат экономических наук, доцент, доцент кафедры экономической кибернетики и маркетинга Киевского национального университета технологий и дизайна

Недбайло Наталья Владимировна,

магистр кафедры экономической кибернетики и маркетинга Киевского национального университета технологий и дизайна

Использование data mining для обработки бизнес-данных

Предмет исследования – подход к возможности применения методов data mining в рамках бизнес-аналитики с целью оптимизации принятия управленческих решений компанией.

Целью написания статьи является исследование особенностей использования методов data mining для обработки первичных данных, которые выступают в качестве ценного ресурса компании, который может быть применен для обеспечения конкурентоспособности на определенном рынке. **Методология проведения работы** – системно-структурного и сравнительного анализов (для исследования направлений использования data mining при комплексном анализе первичных данных); монографического (при изучении предпосылок роста использования компаниями data mining в процессе исследования данных); экономического анализа (при оценке целесообразности применения методов машинного обучения для обеспечения целей бизнес-аналитики). **Результаты работы** – определены особенности использования data mining как одного из направлений бизнес-аналитики, которое дает возможность на основе первичных данных определять неявные взаимосвязи между известными факторными и результативными признаками. Раскрыты основные направления осуществления манипуляций с данными: классификация и прогнозирование, а также корреляционно-регрессионный анализ. Доказана важность использования основных методов

статистического анализа в процессе исследования первичных данных. Приведем специфику применения нейронных сетей как одного из важнейших методов машинного обучения. **Выводы** – использование data mining позволяет компаниям увеличивать эффективность использования имеющихся данных и соответствующим образом оптимизировать стратегии развития. Наличие большого количества методов машинного обучения и статистических подходов расширяет возможности комплексного анализа данных. Важную роль в данном случае играют инновационные технологии и специализированные языки программирования.

Ключевые слова: data mining, данные, эффективность, экономика, машинное обучение, нейронные сети, статистический анализ, управленческое решение.

Chubukova Olha,

doctor of economics, professor, head of Economic Cybernetics and Marketing Department Kyiv National University of Technologies and Design

Ponomarenko Ihor,

candidate of economic sciences, associate professor of Economic Cybernetics and Marketing Department Kyiv National University of Technologies and Design

Nedbailo Natalia,

master of Economic Cybernetics and Marketing Department Kyiv National University of Technologies and Design

Using data mining to process business data

The subject of the research is the approach to the possibility of applying data mining methods in the framework of business analytics in order to optimize the adoption of management decisions by the company. **The purpose of writing this article** is to study of data mining methods features use of primary data, which act as a valuable resource of the company, which can be used to ensure competitiveness in a particular market. **Methodology.** The research methodology is system-structural and comparative analyzes (to study the approaches of data mining data for the complex analysis of first data); monograph (studying the preconditions for the growth of data mining companies' use in the process of data research); economic analysis (when assessing the feasibility of using machine learning methods to ensure the goals of business intelligence). **The scientific novelty** consists the peculiarities of data mining application as one of the directions of business analytics are determined, which makes it possible to determine implicit relationships between known factor and result characteristics on the basis of primary data. The main directions of data manipulation are revealed: classification and forecasting, as well as correlation-regression analysis. The importance of using the basic methods of statistical analysis in the process of studying primary data is proved. The specifics of the use of neural networks as one of the most important methods of machine learning are given. **Conclusions.** The use of data mining allows companies to increase the efficiency of the use of available data and optimize development strategies accordingly. The presence of a large number of machine learning methods and statistical approaches expands the possibilities of comprehensive data analysis. Innovative technologies and specialized programming languages play an important role in this case.

Keywords: data mining, data, efficiency, economy, machine learning, neural networks, statistical analysis, management decision.

Постановка проблеми. Сучасні компанії в умовах посилення конкуренції у офлайн та онлайн-середовищі вимагають викорис-

тання ефективних інструментів, що дозволяють оптимізувати їх функціонування як у цілому, так і на рівні окремих структурних підрозділів. Активний розвиток інноваційних технологій надав можливості накопичення компаніями великих масивів даних, що можуть розглядатись як цінний ресурс у процесі прийняття відповідних управлінських рішень. Розвиток реляційних баз та застосування таких мов як SQL дозволяє здійснювати не лише накопичення, але й широкий перелік маніпуляцій з наявними даними, формуючи відповідні вибіркові сукупності. Поряд з цим, у процесі обробки даних можливо застосовувати такі спеціалізовані мови програмування як Python та R, що дозволяють здійснювати дослідження за допомогою різноманітних методів машинного навчання [1, 2].

В окреслених умовах доцільно комплексно дослідити особливості використання методів data mining, які можуть бути застосовані для великих масивів даних, а також для малих вибірових сукупностей. Проте необхідно забезпечити репрезентативність використовуваних сукупностей, а також застосувати лише ті методи машинного навчання, які дозволяють отримати високі результати за умови дотримання науково обгрунтованих підходів.

Аналіз основних досліджень і публікацій. Дослідженням data mining займаються такі видатні вчені як Ч. Аггарвал, П. Брюс, Я. Віттен, Ф. Провост, М. Стейнбах, П. Тан, Т. Фавсет, Е. Франк, Г. Шмуелі та ін.

Мета дослідження. Мета статті – розглянути особливості використання методів data mining у сучасних умовах для комплексного аналізу «сирих» даних за допомогою відповідних методів та формування на основі отриманих результатів комплексних управлінських рішень, які дозволять підвищити ефективність функціонування компаній.

Виклад основного матеріалу. В сучасних умовах у процесі роботи з даними застосовується підхід data mining, який представляє собою видобуток та інтелектуальний аналіз даних. Передбачається використання сукупності методів, які дозволяють всебічно дослідити наявні дані та визначити приховані взаємозв'язки між певними факторними та результативними ознаками. У певних випадках нетривіальне використання методів машинного навчання дає можливість ідентифікувати сутність досліджуваного явища та розробити відповідні управлінські рішення, які дозволяють ефективність діяльності компанії [3].

До базових методів data mining відносяться різноманітні методи класифікації, моделювання та прогнозування, що реалізуються завдяки використанню асоціативної пам'яті, генетичних алгоритмів, дерев рішень, еволюційного програмування, нечіткої логіки, штучних нейронних мереж. У процесі реалізації підходу data mining необхідно приділяти увагу використанню статистичних методів: аналіз взаємозв'язків, аналіз часових рядів, описативний аналіз, дискримінантний аналіз, дисперсійний аналіз, компонентний аналіз, кореляційний і регресійний аналіз, факторний аналіз. Отже, методи data mining слід розглядати як комплексний підхід, що поєднує між

собою бази даних, статистику та штучний інтелект. Завдяки застосуванню потужного набору методів для візуалізації даних можливо наочно продемонструвати різним групам користувачів отримані результати [4, 5].

У відповідності з концепцією data mining передбачається пошук у великих обсягах «сирих» даних певних «прихованих знань». Проте слід відмітити, що наявність неявних взаємозв'язків не завжди вдається ідентифікувати у зв'язку з їх відсутністю. До основних груп прихованих знань слід віднести:

- невідомі раніше знання про певні соціально-економічні явища та процеси;
- нетривіальні знання неможливо ідентифікувати при простому ознайомленні з певними даними, тому виникає потреба у їх комплексному дослідженні за допомогою різноманітних статистичних методів, досягти значних результатів також можливо за допомогою методів машинного навчання [6].

У результаті дослідження «сирих» даних необхідно досягти:

- практичної корисності, яка проявляється в цінності для окремих структурних підрозділів та компанії у цілому;
- інтерпретованості результатів, оскільки в процесі комплексного аналізу даних та використання спеціалізованих методів машинного навчання необхідно забезпечити легкість формування висновків та наочність для різноманітних груп користувачів.

Data mining за своєю сутністю є складним явищем, що розвивається під впливом комплексу факторів, серед яких доцільно використати наступні:

- інтенсивний розвиток програмного та апаратного забезпечення;
- зменшення вартості, збільшення надійності та розмірів накопичувачів даних;
- вдосконалення методів машинного навчання;
- накопичення великих обсягів даних [7, 8].

Завдяки інтенсивному розвитку інноваційних технологій та їх активному впровадженню у діяльність компаній відбувається постійне зростання попиту на фахівців у сфері обробки даних. У широкий вжиток поряд з data mining увійшли такі поняття як big data, data science, deep learning та ін. Міжнародний ринок обробки даних характеризується високими темпами зростання, поряд з цим, в Україні лише починається усвідомлення приватними компаніями доцільності застосування сучасних методів обробки інформації.

Наявність великої кількості програмного забезпечення дає можливість залучати фахівців різного рівня кваліфікації, досягаючи прийнятних результатів у ході комплексного дослідження структурованих та неструктурованих даних. Проте, слід відзначити, що застосування підходів data science збільшує ймовірність досягнення позитивного результату, який може бути монетизований приватними компаніями, але в певних випадках реалізація різноманітних алгоритмів не дає можливості отримати прийнятний результат, що пов'язано зі специфікою первинних даних.

Висновки.

1. Data mining дає можливість проводити комплексні дії з «сирими» даними, що сприяє високому попиту на відповідних фахівців серед компаній. Отримані у результаті дослідження закономірності є цінними в процесі формування відповідних управлінських рішень, оскільки дають можливість чітко усвідомити основні тенденції та взаємозв'язки у певних соціально-економічних явищах та процесах. Тобто завдяки застосуванню data mining зростає цінність накопичуваної компанією інформації.

2. В процесі обробки даних необхідно приділяти належну увагу методам статистичного аналізу, а також підходам машинного навчання. Для їх реалізації доцільно використовувати такі мови програмування як Python та R. Завдяки гнучкості побудови алгоритмів існує можливість адаптувати наявні методи до певної специфіки даних, а також комбінувати різноманітні підходи, отримуючи оптимальну модель.

Список використаних джерел

1. **5 IMPORTANT FUTURE TRENDS IN DATA MINING** [Електронний ресурс]. – Режим доступу: <https://www.flatworldsolutions.com/data-management/articles/data-mining-future-trends.php>
2. **The SQL Tutorial for Data Analysis** [Електронний ресурс]. – Режим доступу: <https://mode.com/sql-tutorial/introduction-to-sql/>
3. **Difference of Data Science, Machine Learning and Data Mining** [Електронний ресурс]. – Режим доступу: <https://www.datasciencecentral.com/profiles/blogs/difference-of-data-science-machine-learning-and-data-mining>
4. **Clearly Explained: How Machine learning is different from Data Mining** [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/clearly-explained-how-machine-learning-is-different-from-data-mining-4ee0e0c91bd4>
5. **10 Machine Learning Methods that Every Data Scientist Should Know** [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0eeee9>
6. **A Tour of Machine Learning Algorithms** [Електронний ресурс]. – Режим доступу: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
7. **BIG DATA TRENDS 2020** [Електронний ресурс]. – Режим доступу: <https://www.analyticsinsight.net/big-data-trends-2020/>
8. **4 Exciting Trends In Data Science To Watch Out For In 2020** [Електронний ресурс]. – Режим доступу: <https://www.expresscomputer.in/big-data-analytics/4-exciting-trends-in-data-science-to-watch-out-for-in-2020/51997/>

References

1. **5 IMPORTANT FUTURE TRENDS IN DATA MINING**
<https://www.flatworldsolutions.com/data-management/articles/data-mining-future-trends.php>

2. The SQL Tutorial for Data Analysis

<https://mode.com/sql-tutorial/introduction-to-sql/>

3. Difference of Data Science, Machine Learning and Data Mining

<https://www.datasciencecentral.com/profiles/blogs/difference-of-data-science-machine-learning-and-data-mining>

4. Clearly Explained: How Machine learning is different from Data Mining

<https://towardsdatascience.com/clearly-explained-how-machine-learning-is-different-from-data-mining-4ee0e0c91bd4>

5. 10 Machine Learning Methods that Every Data Scientist Should Know

<https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0eeee9>

6. A Tour of Machine Learning Algorithms

<https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

7. BIG DATA TRENDS 2020

<https://www.analyticsinsight.net/big-data-trends-2020/>

8. 4 Exciting Trends In Data Science To Watch Out For In 2020

<https://www.expresscomputer.in/big-data-analytics/4-exciting-trends-in-data-science-to-watch-out-for-in-2020/51997/>