

УДК 65.012.004

Ілляшенко К.В., к.е.н., доцент

Таврійський державний агротехнологічний університет

## СУЧАСНІ ІНФОРМАЦІЙНІ МЕТОДИ АНАЛІЗУ ДАНИХ

У статті висвітлена проблематика обробки нагромадження сирової інформації на підприємствах з великими інформаційними потоками, досліджені особливості сучасних інформаційних методів аналізу даних, проаналізовані переваги методів Data Mining і Knowledge Discovery in Databases, визначені перспективи їх застосування на вітчизняних підприємствах

**Ключові слова:** інформація, аналіз, дані, Data Mining, Knowledge Discovery in Databases

Pyashenko K.

## MODERN INFORMATION METHODS OF DATA ANALYSIS

In the article the range of problems of processing of accumulation of raw information on enterprises with large information flows is lightened, the features of modern informative methods of analysis of data are investigated, advantages of methods of Data Mining and Knowledge Discovery in Databases are analyzed, the prospects of their application on domestic enterprises are certain.

Informatization of society has led to the fact that with the improvement of technologies for capturing and storing data collected huge amounts of diverse information. Work business involves the registration and documentation of all details of the current activity. Accumulated unstructured data form a chaotic clutter that without special treatment, have no beneficial use, moreover, complicate the process of finding the really necessary information. Modern requirements for such treatment as follows: data are unlimited and are dissimilar, the results should be concrete and understandable tools for their processing must be easy to use. One of the methods of this treatment is promising, but new to the majority of domestic enterprises data mining.

Now there are two ways of obtaining information: documentary and expert. In the first case, data is loaded from multiple data sources such as documents, databases, information systems, and more. Expert method involves extracting and structuring knowledge from memory of expert as specialist of which area.

A feature of the modern information processing there is a huge diverse amount of data, making untenable both statistical and expert approaches in many practical aspects, including economic. Therefore, to analyze the information stored in modern databases, methods must be efficient: easy to use, have a significant level of scalability and automaticity. This concept is the basis of modern information technologies of analysis: Data Mining and Knowledge Discovery in Databases (KDD).

Global implementation of information technology requires new approaches to the analysis of activity of enterprises that need to expand its analytical activities and to more effectively use modern information methods analysis. The use of techniques such as data mining and KDD, can allow enterprises to create a new paradigm of action in relation to modern conditions

**Keywords:** information, analysis, data, Data Mining, Knowledge Discovery in Databases

Ильяшенко Е.В.

## **СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ**

В статье высветлена проблематика обработки накопления сырой информации на предприятиях с большими информационными потоками, исследованы особенности современных информационных методов анализа данных, проанализированы преимущества методов Data Mining и Knowledge Discovery in Databases, определены перспективы их применения на отечественных предприятиях

**Ключевые слова:** информация, анализ, данные, Data Mining, Knowledge Discovery in Databases

**Постановка проблеми у загальному вигляді і її зв'язок з важливими науковими та практичними завданнями.** Інформатизація суспільства призвела до того, що з удосконаленням технологій запису та зберігання даних збираються величезні потоки самої різноманітної інформації. Робота підприємств супроводжується реєстрацією та документуванням усіх подробиць поточної діяльності. Накопичені неструктуровані дані утворюють хаотичні нагромадження, які без спеціальної обробки не мають корисного застосування, більш того, ускладнюють процес пошуку дійсно потрібної інформації. Сучасні вимоги до такої обробки наступні: дані мають необмежений обсяг і є різноманітними, результати повинні бути конкретні та зрозумілі, інструменти для їх обробки мають бути прості у використанні. Одним із методів такої обробки є перспективний, але новий для більшості вітчизняних підприємств інтелектуальний аналіз даних. Все це обумовлює доцільність і актуальність даного дослідження.

**Аналіз останніх досліджень, у яких започатковано вирішення проблеми.** Проблематикою аналізу даних, який би відповідав зазначеним вище вимогам, займалися багато провідних вчених, серед яких можна виділити А. Барсегіяна, В. Дюка,

М. Купріянова, Д. Макленнена, А. Самойленка, В. Степаненка, Ч. Танга, І. Холода, І. Чубукову тощо. Але це питання все ще мало досліджено.

**Цілі статті.** Ціллю даної статті є дослідження інформаційних методів аналізу даних та визначення перспектив їх застосування на вітчизняних підприємствах.

**Виклад основного матеріалу дослідження з обґрунтуванням отриманих наукових результатів.** Зараз існує два способи одержання інформації: документальний і експертний. У першому випадку дані отримуються з безлічі інформаційних джерел, таких як документи, бази даних, інформаційні системи тощо. Експертний спосіб припускає витяг і структурування знань із пам'яті експерта, фахівця з якоїсь галузі. Часто їх називають методами, що спрямовані на використання інтуїції та досвіду фахівців [1, с.32].

Серед методів першої групи в економіці поширені методи математичної статистики. Дані методи вирішують великий спектр завдань, однак не дозволяють знаходити та витягати знання з масивів даних. Крім того, високі вимоги до кваліфікації кінцевих користувачів обмежують їх використання.

Серед другої групи поширені так звані експертні системи, що представляють собою спеціальні комп'ютерні програми, що моделюють міркування людини. Наприклад, експертна система прийняття рішень на біржі, експертна система оцінки ефективності інвестицій тощо. Висока вартість створення та впровадження експертних систем, нездатність людей виявляти складні нетривіальні залежності, часта відсутність фахівців, що здатні кваліфіковано структурувати свої знання також заважають поширенню такого підходу [2, с.73].

Сучасний світ наскрізь пронизаний інформаційними потоками. Особливістю сучасних вимог до обробки інформації є величезний обсяг даних і їх різноманіття, що робить неспроможними як статистичні, так і експертні підходи в багатьох практичних аспектах, у тому числі й економічних. Тому для аналізу інформації, накопиченої у сучасних базах даних, методи повинні бути ефективними: простими у використанні, мати значний рівень масштабованості й автоматизму. Ця концепція полягає в основі двох сучасних інформаційних технологій аналізу: Data Mining і

Knowledge Discovery in Databases (KDD) [3, с.144 ].

Технологія Data Mining (видобуток даних) – це виявлення у вихідних необроблених сирих даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретацій знань. Тобто інформація, знайдена у процесі застосування методів Data Mining, повинна бути нетривіальною та раніше невідомою. Знання повинні описувати нові зв'язки між властивостями, прогнозувати значення одних ознак на основі інших [4, с.123].

Сфера застосування Data Mining нічим не обмежена. Найбільше цей метод необхідний комерційним підприємствам, які розгортають свої проекти на основі інформаційних сховищ даних (Data Warehousing). Досвід таких підприємств показує, що віддача від використання методу може досягати десятків разів [3, с.376].

Підприємства за допомогою подібних інформаційних методів аналізу можуть одержати відчутні переваги перед конкурентами та у стратегічному плануванні тощо.

Основною перевагою методів інтелектуального аналізу Data Mining є здатність виявлення нових знань, які неможливо одержати звичайними методами економічного аналізу. На наш погляд саме Data Mining буде найбільш затребуваним на підприємствах найближчим часом.

Виявлення знань у базах даних KDD – це послідовність дій, яку необхідно виконати для побудови моделі пошуку знання. Ця послідовність не описує конкретний алгоритм або математичний апарат, не залежить від предметної галузі. Це набір атомарних операцій, комбінуючи які, можна одержати потрібний розв'язок.

KDD складається з наступних етапів: підготовки вхідних даних, вибору інформативних ознак, очищення даних, застосування методів Data Mining, підсумкової обробки даних, інтерпретації отриманих результатів.

Таблиця 1

Покрокові етапи KDD

Етап	Послідовність кроків
Підготовка вхідного набору даних	Цей етап полягає у підготовці набору даних, у тому числі з різних джерел, вибору значимих параметрів тощо. Для цього повинні існувати розвинені інструменти доступу до різних джерел даних

Продовження табл. 1

<p>Попередня обробка даних</p>	<p>Для ефективного застосування методів Data Mining слід звернути увагу на запитання попередньої обробки даних. Дані можуть містити пропуски, шуми, аномальні значення, крім того, дані можуть бути надлишкові, недостатні тощо. У деяких завданнях потрібно доповнити дані якоюсь апріорною інформацією. Помилково припускати, що якщо подати дані на вхід системи в існуючому виді, то на виході будуть отримані корисні знання, тому вхідні дані повинні бути якісні та коректні</p>
<p>Трансформація та нормалізація даних</p>	<p>Цей крок необхідний для тих методів, які вимагають, щоб вихідні дані були в якомусь певному вигляді. Різні алгоритми аналізу вимагають спеціальним образом підготовлених даних, наприклад, для прогнозування необхідно перетворити часовий ряд за допомогою ковзного вікна або обчислення агрегованих показників. До завдань трансформації даних відносяться: ковзне вікно, приведення типів, виділення тимчасових інтервалів, перетворення безперервних значень у дискретні та навпаки, сортування, угруповання та інше</p>
<p>Data Mining</p>	<p>На цьому кроці застосовуються різні алгоритми для знаходження знань (нейронні мережі, дерева рішень, встановлення асоціацій тощо)</p>
<p>Підсумкова обробка даних</p>	<p>Інтерпретація результатів і застосування отриманих знань з додаткових джерел</p>

Центром усього цього процесу є методи Data Mining, які дозволяють виявляти закономірності та знання [5, с. 195 ].

Наприклад, є мережа магазинів і необхідно одержати прогноз обсягів продажу на наступний період. Першим кроком буде збір історії продажів у кожному магазині й об'єднання її в загальну вибірку даних. Далі буде попередня обробка зібраних даних: їх угруповання по місяцях, згладжування кривої продажів, усунення факторів, що негативно впливають на обсяги продажів. Далі будується модель залежності обсягів продажу від обраних факторів. Це можна зробити за допомогою лінійної регресії або нейронних мереж. За допомогою такої моделі можна одержати прогноз, у якому вході буде модель історії продажу. Знаючи прогнозне

значення, його можна використовувати, наприклад, для оптимізації розміщення товару на складі [6, с. 164].

Вважаємо, що найголовнішою перевагою KDD є те, що отримані таким чином знання можна помножувати, тобто вже побудовану модель зможуть застосовувати інші, без необхідності розуміння методик, за допомогою якої ці моделі побудовані. Знайдені знання повинні бути використані на нових даних з деяким ступенем вірогідності.

Сучасні технології аналізу стають все більш зручними та точними у розрахунках і прогнозуванні. Тому ми вважаємо, що вже найближчим часом інформаційні системи та інтелектуальні методи аналізу даних будуть використовуватися на більшості вітчизняних підприємств.

**Висновки.** Фінансові негаразди сучасної економіки вимагають від вітчизняних підприємств використання лише малозатратних методів аналізу, яким вважається традиційний економічний аналіз. Але глобальне впровадження інформаційних технологій потребує нових підходів до аналізу діяльності підприємств, яким необхідно розширювати свою аналітичну діяльність та більш ефективніше застосовувати сучасні інформаційні методи аналізу. Застосування таких методів, як Data Mining і KDD, може дозволити підприємствам створити нову парадигму діяльності відповідно до сучасних умов. Однак з наукової точки зору ці методи потребують подальших глибоких досліджень.

1. Knowledge Discovery Through Data Mining: What Is Knowledge Discovery? – Tandem Computers Inc. – 1996. – 54 p.
2. Дюк В.А. Data Mining: учебный курс [Текст] / В.Дюк, А.Самойленко. – СПб: Изд. Питер, 2001. – 368 с.
3. J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000. P.184.
4. Чубукова И.А. Data Mining: учебное пособие [Текст] / И.А.Чубукова. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.
5. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP [Текст] / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб: БХВ-Петербург, 2007. – 384 с.
6. Криват Б. Microsoft SQL Server 2008: Data Mining – интеллектуальный анализ данных [Текст] / Б.Криват, Д.Макленнен, Ч.Танг. – Минск: ВHV, 2009. – 720 с.