



СЕКЦІЯ 6 СУЧАСНІ ПЕДАГОГІЧНІ ТЕХНОЛОГІЇ

УДК 004.89:378

ВИВЧЕННЯ ТЕХНОЛОГІЙ І ЗАСОБІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ DATA MINING В ОСВІТНЬОМУ ПРОЦЕСІ

Зелінська С.А., к. пед. н., докторант
Криворізький державний педагогічний університет

Засоби інтелектуального аналізу даних Data Mining дають змогу отримати раніше не відомі знання на базі використання дерев рішень, штучних нейронних мереж, генетичних алгоритмів, еволюційного програмування та нечіткої логіки. Використання інтелектуального аналізу даних в освітньому процесі дасть змогу отримати додаткові знання для підвищення керованості освіти. Використання відповідних інструментальних засобів, до яких можна віднести Oracle Data Mining, Microsoft Analysis Services, дасть змогу отримати максимально повні результати для прийняття ефективних рішень в освітньому процесі.

Ключові слова: інтелектуальний аналіз даних, освітній процес, прикладне програмне забезпечення, алгоритм, служба аналізу Analysis Services, Data Mining.

Средства интеллектуального анализа данных Data Mining позволяют получить ранее не известные знания на базе использования деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования и нечеткой логики. Использование интеллектуального анализа данных в образовательном процессе позволит получить дополнительные знания для повышения управляемости образования. Использование соответствующих инструментальных средств, к которым можно отнести Oracle Data Mining, Microsoft Analysis Services, позволит получить максимально полные результаты для принятия эффективных решений в образовательном процессе.

Ключевые слова: интеллектуальный анализ данных, образовательный процесс, прикладное программное обеспечение, алгоритм, служба анализа Analysis Services, Data Mining.

Zelinska S.A. STUDYING DATA MINING'S TECHNOLOGIES AND TOOLS IN THE EDUCATIONAL PROCESS

Means of data mining Data Mining allows you to get previously unknown knowledge based on the use of decision trees, artificial neural networks, genetic algorithms, evolution programming and fuzzy logic. The use of data mining in the educational process will provide additional knowledge to improve the manageability of education. Using the appropriate tools, which include Oracle Data Mining, Microsoft Analysis Services, will get the most complete results for making effective decisions in the educational process.

Key words: intellectual data analysis, educational process, application software, algorithm, analysis services Analysis Service, Data Mining.

Постановка проблеми. Актуальність виконання роботи зумовлена тим, що сучасні інформаційно-комунікаційні технології дають змогу ефективно збирати й накопичувати великі обсяги різномірних експериментальних даних, усебічний і грамотний аналіз яких є необхідним для проведення повноцінного дослідження. Стандартні методи статистичної й математичної обробки даних у більшості випадків не дають змогу виявити в даних існуючі нетривіальні та задалегідь не передбачувані закономірності, для виявлення яких розробляються й використовуються методи інтелектуального аналізу даних.

Інтелектуальний аналіз даних широко застосовується у світовій спільноті в процесі проведення досліджень у різних галу-

зях, наприклад у генетиці, біоінформатиці, астрономії, медицині, економіці, освіті, інформатиці та інших. Щоб проводити ефективні експерименти, отримуючи конкурентоспроможні й доказові результати, фахівці повинні володіти знаннями та навичками в галузі інтелектуального аналізу даних. Адаптація методів інтелектуального аналізу даних для обробки педагогічної інформації та широке їх упровадження в освітній процес можуть різко підвищити якість управління навчанням і контроль над ним.

Аналіз останніх досліджень і публікацій. Різні аспекти використання технологій інтелектуального аналізу даних як складової частини педагогічного управління були розглянуті в роботах Л.Ю. Овсяницької. Предметом дослідження таких авторів, як



Г.Ж. Солтан, С.С. Смаїлова, І.М. Увалієва, А.К. Томілін, був інтелектуальний аналіз даних у завданнях управління якістю освітнього процесу. У роботі А.А. Барсегян надано докладну характеристику методів і моделей аналізу даних Data Mining. У праці А.В. Кузнецової та О.В. Сенько описані можливості використання методів Data Mining під час лабораторних досліджень для виявлення закономірностей у масивах даних.

Однак варто зазначити, що в межах досліджуваної теми існує велика кількість розрізних даних, кожен із дослідників акцентує увагу на цікавих йому характеристиках, що не дає змогу отримати цілісну систему практичного використання технологій інтелектуального аналізу Data Mining у предметній сфері. У зв'язку із цим необхідні додаткові дослідження технологій Data Mining та впровадження їх в освітній процес.

Постановка завдання. Мета роботи полягає у вивченні технологій і засобів інтелектуального аналізу даних Data Mining в освітньому процесі.

Відповідно до мети було визначено необхідність постановки й вирішення таких завдань:

- дати характеристику поняття «Data Mining»;
- проаналізувати сучасні підходи інтелектуального аналізу Data Mining в освітньому процесі;
- описати прикладні програмні продукти інтелектуального аналізу даних Oracle Data Mining та Microsoft Analysis Services.

Виклад основного матеріалу дослідження. Поняття «Data Mining» є збірною назвою, що використовується для позначення сукупності спеціалізованих методів виявлення в раніше не відомих даних нетривіальних, практично корисних і доступних для подальшої інтерпретації знань, необхідних для прийняття ефективних рішень у різних сферах людської діяльності.

Основою методів Data Mining є класифікація, моделювання та прогнозування, засновані на використанні дерев рішень, штучних нейронних мереж, генетичних алгоритмів, еволюційного програмування, асоціативної пам'яті, нечіткої логіки [3].

До методів Data Mining можна віднести широкий спектр статистичних методів, серед яких виділяються кореляційний і регресійний аналіз, дескриптивний аналіз, факторний аналіз, дисперсійний аналіз, компонентний аналіз, дискримінантний аналіз, аналіз часових рядів, аналіз виживаності, аналіз зв'язків.

Водночас необхідно зазначити, що реалізація таких методів передбачає деякі апіорні уявлення про аналізовані дані,

у свою чергу виникає певна розбіжність із цілями методів Data Mining, які полягатимуть у виявленні раніше не відомих знань.

Одним із найбільш важливих призначень Data Mining є візуалізація результатів, що дає змогу використовувати наданий інструментарій Data Mining людьми, які не мають спеціальної підготовки.

Інструменти Data Mining дають змогу виявити фактори, що впливають на результат. Тому, маючи велику кількість інформації, що була отримана в процесі моніторингу педагогічних показників, і заклавши їх в систему Data Mining, можна визначити, від яких причин та рушійних сил, заходів, вжитих дій більшою мірою залежить, наприклад, рівень компетенції випускників вищих навчальних закладів або їх можливість упродовж певного часу після завершення навчання знайти хорошу роботу за фахом [7].

Відповідно, можна визначити також фактори, які або не мають статистично значущого впливу, або негативно впливають на результат. Застосування інтелектуального аналізу статистичних даних в освітніх установах у процесі впровадження мережевих освітніх систем та електронного навчання сприяє якісним змінам традиційної системи освіти.

У роботі Л.Ю. Овсяницької обговорюються теоретичні та практичні аспекти застосування інтелектуального аналізу даних у педагогіці, продемонстровані приклади оцінювання якісних характеристик навчання й особистісних характеристик студентів за допомогою засобів Data Mining. Один із прикладів практичного застосування інтелектуального аналізу даних у педагогіці реалізований за допомогою використання інструменту засобів інтелектуального аналізу MS Excel у межах надбудови «Інтелектуальний аналіз даних», яка дає змогу використовувати переваги функцій прогнозного аналізу SQL Server.

На додаток до вищесказаного за допомогою інструменту «Аналіз ключових факторів впливу» може бути визначений зв'язок параметра з іншими показниками та залежність його від них. Якщо як результуючий фактор узяти знаходження випускниками роботи впродовж певного часу після закінчення вищого навчального закладу, тобто підтвердження того, що компетенції випускників задовольнили роботодавців, а як фактори, що впливають, – оцінки, отримані студентами на сесіях, державному іспиті, під час захисту дипломного проекту, участь у студентських наукових конференціях, наукових гуртках, творчих колективах, особистісні характеристики тощо, то можна зрозуміти, які чинники або їх комбіна-



ції зробили вирішальний вплив на результат [5].

Також автором запропоновані підходи до складання шаблонів, які лежать в основі методів Data Mining, що в сукупності із застосуванням сучасних прикладних програмних засобів дають можливість здійснювати ефективну обробку матеріалів педагогічного моніторингу, виявляти чинники, які впливають на результат навчання, своєчасно коригувати подальшу роботу навчального закладу, тобто здійснювати педагогічне управління.

Г.Ж. Солтан, С.С. Смаїлова, І.М. Увалієва та А.К. Томілін запропонували модель системи інтелектуального аналізу даних стосовно освітнього процесу вищого навчального закладу. Ученими проаналізовано можливості інтелектуального аналізу даних з урахуванням специфіки освітньої сфери та описано досвід його застосування [6].

Специфіка інтелектуального аналізу даних у сфері освіти полягає в тому, що більшість подібних завдань мають на увазі побудову моделей, що розкривають закономірності в наявних даних. Тому найбільш затребуваними є описові моделі інтелектуального аналізу, які сприяють глибшому розумінню аналізованих даних [4]. Ключовим моментом у таких моделях є легкість отримання й прозорість результатів для сприйняття людиною. Можливо, виявлені закономірності є специфічною рисою саме конкретних досліджуваних даних і більше ніде не трапляються, проте це все одно може бути корисним, а тому має бути відоме. До такого виду завдань можна віднести кластеризацію та пошук асоціативних правил.

Водночас за результатами інтелектуального аналізу даних освітнього процесу можна вирішувати такий комплекс завдань:

- виявлення кластерів студентів, які є групою ризику щодо успішності;
- виконання оперативного аналізу великих обсягів даних (наприклад, результатів поточного й підсумкового контролю успішності студентів) та виявлення відхилень від нормального перебігу процесу навчання;
- виявлення проблемних ситуацій у навчальному процесі та їх можливих причин;
- аналіз накопичених даних для вдосконалення освітнього процесу.

Для ефективної реалізації інтелектуального аналізу даних існують вільно поширювані прикладні програмні продукти у вигляді вихідного коду реалізації популярних алгоритмів, самостійні прикладні програми (SAS Enterprise Miner, Poly Analyst, Deductor, RapidMiner), а також спеціалізовані засоби, інтегровані в систему управ-

ління базами даних (Oracle Data Mining, SQL Server Analysis Services) [1].

Oracle Data Mining призначений для аналізу даних методами, що належать до технології вилучення знань, або Data Mining. По-перше, основне завдання цієї технології полягає у виявленні у великих наборах даних прихованих закономірностей, залежностей і взаємозв'язків, корисних під час прийняття рішень на різних рівнях управління. По-друге, такі закономірності представляються у вигляді інтелектуальних або дескриптивних моделей різного типу, що дають змогу проводити класифікацію ситуацій чи об'єктів, прогнозувати їх поведінку, виявляти групи подібних об'єктів тощо. Істотно, що моделі будуються автоматично на основі аналізу наявних даних про об'єкти, спостереження й ситуації за допомогою спеціальних алгоритмів.

Основу опцій складають процедури, що реалізують різні алгоритми побудови моделей регресії, класифікації, кластеризації. Версія Data Mining 12c підтримує широкий спектр таких алгоритмів, у тому числі регресійний аналіз, класифікацію, кластеризацію, виявлення аномалій, пошук асоціацій, значимість атрибутів, виділення ознак [8].

Крім власне алгоритмів, до складу Oracle Data Mining входять спеціалізовані засоби підготовки даних, оцінки результатів, застосування моделей до нових наборів даних. Використовувати всі ці можливості можна як на програмному рівні за допомогою PL / SQL API, так і за допомогою спеціального графічного середовища Oracle Data Miner, що реалізоване у вигляді розширення для середовища SQL Developer та орієнтоване на роботу аналітиків, які працюють із завданнями прогнозування, виявлення тенденцій, сегментації тощо.

Microsoft Analysis Services (служби аналізу від компанії Microsoft) є частиною системи управління базами даних Microsoft SQL Server. Microsoft включила в SQL Server набір служб, пов'язаних із бізнес-аналізом і зберіганням даних. Ці служби включають у себе спеціалізовані служби інтеграції (Integration Services), служби аналізу (Analysis Services). У свою чергу Analysis Services включають у себе комплексний набір засобів для роботи з OLAP та інтелектуальним аналізом даних [2].

Microsoft Analysis Services підтримує різні набори прикладних програмних інтерфейсів (API) та об'єктних моделей для різних операцій у різних програмних середовищах. Так, Data Definition Language (DDL) являє собою сімейство комп'ютерних мов, які використовуються в прикладних комп'ютерних програмах для опису спеці-



алізованих структур баз даних, для моделей імпорту й експорту інтелектуального аналізу даних, у яких також підтримується PMML.

Служби Analysis Services надають такі функції й засоби для створення рішень з інтелектуального аналізу даних: набір стандартних алгоритмів інтелектуального аналізу даних; конструктор інтелектуального аналізу даних, призначений для створення й перегляду моделей інтелектуального аналізу даних, управління ними та побудови прогнозів; мова розширень інтелектуального аналізу даних (Data Mining Extensions to SQL, DMX).

Водночас використання готових алгоритмів у вигляді вихідного коду – це дуже багато роботи, а в разі застосування самостійних додатків інформаційна система виконує вибірку й очищення даних в оперативному режимі, експорт даних здійснює у форматі, який прийнятний для зовнішнього інструментального засобу. Далі всі маніпуляції з даними, такі як візуалізація та застосування методу аналізу, відбуваються окремо від інформаційної системи. На нашу думку, найбільш прийнятним способом є засіб, інтегрований у системі управління базами даних.

Отже, до переваг такого підходу можна віднести можливість об'єднання різнорідних даних, використання вбудованих засобів підготовки даних до інтелектуального аналізу, вибір і застосування різних алгоритмів інтелектуального аналізу даних, збереження й повторне використання моделей, можливість порівняльного аналізу ефективності різних моделей.

Висновки з проведеного дослідження. Підводячи підсумки, можна сказати, що технології й засоби інтелектуального аналізу даних Data Mining є відносно новими, проте вже зарекомендували себе з пози-

тивного боку. Застосування методів інтелектуального аналізу Data Mining у сукупності із сучасними програмними засобами дає змогу проводити обробку даних педагогічного моніторингу, виявляти чинники, що впливають на результат, враховувати отриману інформацію та коригувати подальшу роботу навчального закладу, здійснювати педагогічне управління.

ЛІТЕРАТУРА:

1. Барсебян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336 с.
2. Барсебян А.А., Куприянов М.С., Степаненко В.В., Холод И.И., Елизаров С.И. Анализ данных и процессов: учеб. пособие. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.
3. Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных. Врач и информационные технологии. 2005. № 2. С. 38–46.
4. Научно-методические и технологические основы информационной системы управления качеством учебного процесса / под ред. Л.И. Григорьева. М., 2008. 132 с.
5. Овсяницкая Л.Ю. Интеллектуальный анализ данных как составляющая педагогического управления. Образование и наука. 2013. № 1(10). С. 80–90.
6. Солтан Г.Ж., Смаилова С.С., Увалиева И.М., Томилин А.К. Интеллектуальный анализ данных в задачах управления качеством образовательного процесса. Инженерное образование. 2013. № 13. С. 36–43.
7. Sachin R.B., Vijay M.S. A Survey and Future Vision of Data Mining in Educational Field. Second International Conference on Advanced Computing & Communication Technologies. Rohtak, Haryana, India, 2012. P. 96–100.
8. Fanqjun Wu. Discussion on experimental teaching of data warehouse & data mining course for undergraduate education. The 7-th International Conference on Computer Science & Education. Melbourne, Australia, 2012. P. 1425–1429.