

Моделювання в економіці, організація виробництва та управління проектами

УДК 004.4.22

А.О. Беккауер

Таврійський державний агротехнологічний університет, Мелітополь

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ DATA MINING ДЛЯ АВТОМАТИЗАЦІЇ БІЗНЕС-ПРОЦЕСІВ НА ВИРОБНИЦТВІ

Розглядаються технології Data Mining, на основі яких можна вирішувати бізнес-процеси, які виникають в процесі діяльності підприємства. Приводиться короткий опис програмного забезпечення (ПЗ), призначеного для прогнозування попиту на продукцію підприємства. ПЗ розроблено на основі кластерного та регресійного аналізу. В результаті роботи програми користувач отримує орієнтовану кількість продажів продукції в наступному місяці певному діапазоні. Для кращого візуального сприйняття результатів, на екрані виводяться графіки з результатами роботи програми. ПЗ розроблено в системі програмування Delphi для операційної системи Windows та Android.

Ключові слова: Data Mining, кластерний аналіз, регресійний аналіз, прогнозування, кореляція.

Вступ

Постановка проблеми. Організації накопичують величезні масиви даних, однак не в змозі отримати від цієї роботи реальну віддачу. Маючи історичні дані, можна вирішити критично важливі для бізнесу завдання: оптимізувати процеси, управляти ризиками, підвищувати прибутковість, утримувати клієнтів. Обґрунтованість і професійний рівень прийняття рішень визначає ефективність діяльності фірми. В цьому відношенні істотну допомогу керівнику надають сучасні інформаційні системи.

Аналіз інформації є невід'ємною частиною ведення бізнесу і одним з важливих чинників підвищення його конкурентоспроможності. В проведенні статистичного аналізу даних допомагають методи Data Mining. Існуючі статистичні системи через свою високу ціну ліцензії, складність у освоєнні та відсутність можливості працювати з портативних приладів стримують впровадження подібних систем у виробництво. Тому розробка та дослідження інформаційної системи статистичного аналізу даних на основі методів Data Mining є актуальним напрямком досліджень для подальшого впровадження на виробництві.

Аналіз останніх досліджень. Для автоматизації бізнес процесів на виробництві можна використовувати такі технології аналізу даних: OLAP, KDD, Data Mining. OLAP служить для підготовки бізнес-звітів з продажів, маркетингу, для потреб управління, для прогнозування, фінансової звітності та в схожих областях, дає можливість у реальному часі генерувати описові й порівняльні зведення даних і одержувати відповіді на різні інші аналітичні запити. KDD – добування знань із баз даних. Data Mining (DM) – «видобуток» даних [1].

Це метод виявлення в «сирих» даних раніше невідомих, нетривіальних, практично корисних і доступних для інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності. Важливим критерієм вибору технології аналізу даних є швидкість та точність отримання результатів аналізу.

Мета роботи. Провести аналіз існуючих технологій аналізу даних та обрати ту, яка найкраще підходить для створення програмного продукту автоматизації бізнес процесів на виробництві. Створити програмний продукт для прогнозування кількості продажів товару в наступному місяці.

Основна частина

Розглянемо найпоширеніші технології, які дозволяють проводити статистичний аналіз даних та допомагають приймати управлінські рішення:

1. OLAP (англ. online analytical processing, аналітична обробка в реальному часі) – це технологія обробки інформації, що дозволяє швидко отримувати відповіді на багатовимірні аналітичні запити. OLAP є частиною такого ширшого поняття, як бізнес-аналітика та служить для підготовки бізнес-звітів.

2. KDD (Knowledge Discovery in Databases) – добування знань із баз даних. Це процес пошуку корисних знань в «сирих даних». KDD містить у собі питання підготовки даних, вибору інформативних ознак, очищення даних, застосування методів Data Mining, постобробки даних і інтерпретації отриманих результатів.

3. Data Mining (DM) – «видобуток» даних. Це метод виявлення в «сирих» даних раніше невідомих, нетривіальних, практично корисних і доступних для інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності.

DM забезпечує розв'язок таких задач:

класифікація – встановлення функціональної залежності між вхідними й дискретними вихідними змінними; за допомогою класифікації вирішується задача віднесення об'єктів (спостережень, подій) до одного із задалегідь відомих класів.

кластеризація – це угруповання об'єктів (спостережень, подій) на основі даних (властивостей), що описують сутність об'єктів; об'єкти усередині кластера повинні бути «схожими» один на одного й відрізнятися від об'єктів, що ввійшли в інші кластери.

регресія – встановлення функціональної залежності між вхідними й неперервними вихідними змінними; прогнозування найчастіше зводиться до розв'язку задачі регресії.

асоціація – виявлення залежностей між зв'язаними подіями, що вказують, що з події X впливає подія Y; такі правила називаються асоціативними.

Після проведення аналізу предметної області, визначено, що для створення програмного модулю аналізу попиту на продукцію потрібно використовувати методи глибокої аналітики Data Mining.

Розглянемо більш докладно основні складові Data Mining, які використовуються в даній роботі.

Кластерний аналіз.

Основна мета кластерного аналізу – знаходження груп схожих об'єктів у вибірці. Спектр застосувань кластерного аналізу дуже широкий: його використовують в багатьох сферах людської діяльності. У кластерному аналізі для кількісної оцінки подібності вводиться поняття метрики. Використовуються різні міри відстані між об'єктами:

- евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}; \quad (1)$$

- зважена евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^m w_k (x_{ik} - x_{jk})^2}; \quad (2)$$

- відстань city-block:

$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|; \quad (3)$$

- відстань Мінківського:

$$d_{ij} = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^p \right)^{1/p}; \quad (4)$$

- відстань Махаланобиса:

$$d_{ij} = (X_i - X_j)^T S^{-1} (X_i - X_j). \quad (5)$$

В процесі кластеризації вхідні дані розділяються на групи (кластери) за певними ознаками. До основних характеристик кластерів можна віднести:

- центр кластера – це середнє геометричне місце точок у просторі змінних;

- дисперсія кластера – це міра розсіювання точок у просторі відносно центру кластера;

- радіус кластера – максимальна відстань точок від центру кластера.

Методи кластерного аналізу можна розділити на дві групи: ієрархічні; неієрархічні. Кожна з груп включає безліч підходів і алгоритмів. Використовуючи різні методи кластерного аналізу, аналітик може отримати різні рішення для одних і тих же даних. Це вважається нормальним явищем.

Аналітикові часто доводиться мати справу з даними, що представляють собою історію зміни різних об'єктів у часі. Такого роду дані називаються тимчасовими рядами (time series data). Часовий ряд являє собою послідовність спостережень за змінами в часі значень параметрів деякого об'єкта або процесу.

Проведене дослідження застосування моделей регресійного аналізу для розв'язання задачі прогнозування [2]. Досліджено різні види моделей часових рядів та методів знаходження ліній тренда:

- проста лінійна модель:

$$t_i = a + b \cdot i; \quad (6)$$

- поліноміальна модель:

$$t_i = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_n \cdot t^n; \quad (7)$$

- експоненціальна модель:

$$t_i = \exp(a + b \cdot i); \quad (8)$$

- логістична модель:

$$t_i = a / (1 + b \cdot e^{-k \cdot i}). \quad (9)$$

На основі методів Data Mining було розроблено програмний продукт, за допомогою яких можна автоматизувати велику кількість розрахунків, та значно зменшити час, необхідний для прийняття управлінських рішень. Як стаціонарну, так і мобільну версію програмного продукту розроблено за допомогою мови програмування Delphi [1].

На вході маємо групи товарів з кількістю продажів по кожному товару за попередні місяці. На виході маємо прогноз кількості продажів по кожному товару у наступному місяці. Спрогнозована кількість продажів представляється користувачу у вигляді довірчого інтервалу з певною вірогідністю.

Робота програми відбувається у діалоговому режимі між оператором та обчислювальною системою. На кожному етапі розрахунків, користувач має змогу спостерігати за результатами обчислень, а перехід до наступної стадії розрахунків відбувається тільки після команди оператора. Головне вікно програми представлено на рис. 1.

Інтерфейс програмного продукту розроблювався максимально простим та інтуїтивно зрозумілим з урахуванням мінімальної підготовки аналітика для роботи з комп'ютером.

Роздивимось роботу програми на основі додатку для операційної системи Android.

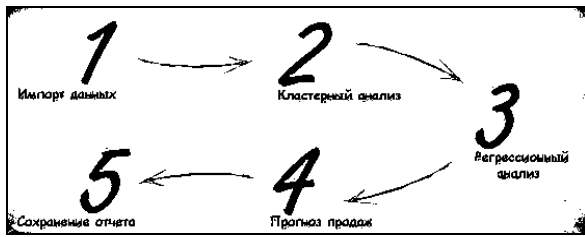


Рис. 1. Головне вікно програми

Розроблено додаток для операційної системи Android, який дозволяє проводити розрахунки за допомогою мобільних пристроїв. Мобільна версія програми має ту ж саму логічну структуру, що й програма для персонального комп'ютера. Для того щоб розпочати аналіз продукції, необхідно завантажити таблицю з вхідними даними. Вікно з завантаженою таблицею вхідних даних представлена на рис. 2.

Имя	Цена	2014				
		Январь	Февр...	Март	Апрель	Май
2204Д	143	642	626	620	613	621
2213Д	154	523	528	542	522	530
2208Д	156,1	431	431	428	432	422
2404Д	168,8	542	529	537	541	537
2604Д	186,6	547	559	352	592	439
2413Д	198,72	540	432	720	476	757
2613Д	216,72	552	615	772	686	389
2207Д	217	612	573	616	510	419
2212Д	217	622	770	469	655	595

Рис. 2. Додаток після завантаження таблиці вхідних даних

Для того щоб візуально побачити як було проведено кластеризацію, необхідно натиснути на кнопку «График» після чого на екрані з'явиться графік з результатами кластеризації, на якому різними кольорами показані товари, які потрапили до різних кластерів (рис. 3).

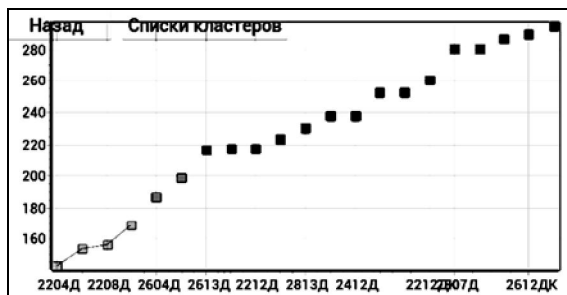


Рис. 3. Графік з результатами кластеризації

Створений програмний продукт дозволяє за допомогою регресійного аналізу та часових рядів розрахувати прогнозовану кількість продажів товарних груп [3, 4].

Для того щоб розпочати регресійний аналіз необхідно натиснути на кнопку «Регрессия». На екрані з'явиться регресійна модель з точковим прогнозом та довірчим інтервалом (рис. 4).

Для того, щоб продивитись текстове повідомлення з прогнозом, необхідно натиснути на кнопку «Прогноз». Результати прогнозування зображені на рис. 5.

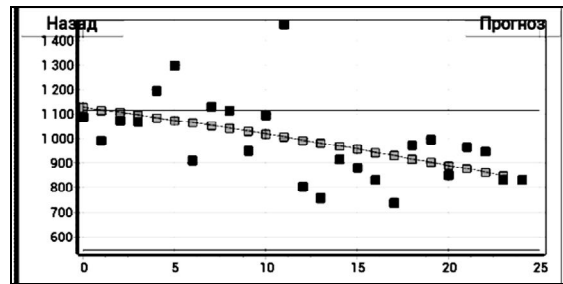


Рис. 4. Графік прогнозу продажів

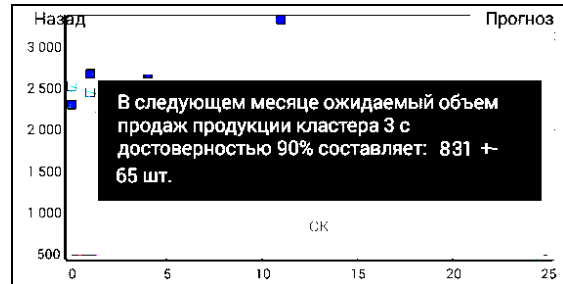


Рис. 5. Результати прогнозування

Висновки

В даній роботі було проведено аналіз існуючих технологій аналізу даних. На основі проведеного аналізу було визначено, що для проведення статистичного аналізу та автоматизації бізнес процесів на виробництві, найкраще використовувати методи Data Mining для отримання реальної віддачі у роботі підприємства.

На основі технологій Data Mining розроблено програмні продукти, які направлені на автоматизацію бізнес процесів, які ведуть до збільшення ефективності роботи підприємства та зменшення навантаження на аналітиків. Розроблена Android версія програми дозволяє проводити розрахунки з мобільних пристроїв. Використовувати мобільну версію програми дуже зручно, коли працівник не має змоги працювати за персональним комп'ютером.

Список літератури

1. Митчелл К. Программирование и отладка в Delphi: Учебный курс / К.: Митчелл. – М.; СПб.: К., 2003.
2. Радченко Станислав Григорьевич. Методология регрессионного анализа: Монография / Станислав Григорьевич Радченко. – К.: «Корнийчук», 2011. – С. 376.
3. Саяпова А.Р. Математические методы прогнозирования экономических показателей: учебное пособие / А.Р. Саяпова, Е.А. Гусельникова, И.А. Лакман, Н.М. Шамуратов. – Уфа, 2000. – 128 с.
4. Фёрстер Э. Методы корреляционного и регрессионного анализа = Methoden der Korrelation - und Regressionsanalyse / Э. Фёрстер, Б. Рёнц. – М.: Финансы и статистика, 1981. – 302 с.

Надійшла до редколегії 6.11.2015

Рецензент: д-р техн. наук, проф. О.М. Леженкін, Таврійський державний агротехнологічний університет, Мелітополь.

**ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ DATA MINING ДЛЯ АВТОМАТИЗАЦИИ
БИЗНЕС-ПРОЦЕССОВ НА ПРОИЗВОДСТВЕ**

А.А. Беккауер

Рассматриваются технологии Data Mining на основе которых можно решать бизнес процессы, которые возникают в процессе деятельности предприятия. Приводится краткое описание программного обеспечения (ПО), предназначенного для прогнозирования спроса на продукцию предприятия. ПО разработано на основе кластерного и регрессионного анализа. В результате работы программы пользователь получает ориентировочное количество продаж продукции в следующем месяце в определенном диапазоне. Для лучшего визуального восприятия результатов, на экране выводятся графики с результатами работы программы. ПО разработано в системе программирования Delphi для операционной системы Windows и Android.

Ключевые слова: Data Mining, кластерный анализ, регрессионный анализ, прогнозирование, корреляция.

**APPLICATION OF DATA MINING TECHNOLOGIES FOR AUTOMATION
OF BUSINESS PROCESSES IN THE INDUSTRY**

А.А. Bekkauer

Data mining technologies, which can be the basis for business processes occurred in the process of an enterprise activity are considered. Brief description of software for production demand forecasting is presented. The software is developed on the basis of cluster and regression analysis. As the result of software operation a user gets approximate number of production sales for the next month in the definite range. For better visual perception graphs with results of work are displayed on the screen. The software is developed in Delphi programming system for Windows and Android operating systems.

Keywords: Data Mining, cluster analysis, regression analysis, forecasting, correlation.