МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ім. Петра Могили

М.Т. ФІСУН, І.О. КРАВЕЦЬ, П.П. КАЗМІРЧУК, С.Г. НІКОЛЕНКО

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ:

ПРАКТИКУМ

для освітніх напрямів підготовки 6.050101 «Комп'ютерні науки» і 6.040303 «Системний аналіз»

Львів
«Новий Світ – 2000»
2020

УДК 004.31(075.8)
ББК з973.3я7
Рекомендовано Вченою радою Чорноморського державного університету
ім. Петра Могили, протокол № 12 від 02 липня 2014 р.

Рецензенти:
Бідюк Петро Іванович – професор кафедри математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут», доктор технічних наук, професор.
Пасічник Володимир Володимирович – професор кафедри інформаційних систем та мереж, доктор технічних наук, професор.
Ходаков Віктор Єгорович – завідувач кафедри інформаційних технологій Херсонського національного технічного університету, доктор технічних наук, професор, заслужений діяч науки і техніки України.

Автори: **Фісун** Микола Тихонович, докт. техн. наук, професор, завідувач. кафедри інтелектуальних інформаційних систем Чорноморського державного університету ім. Петра Могили.
Кравець Ірина Олександрівна, канд. техн. наук, доцент, доцент кафедри інтелектуальних інформаційних систем Чорноморського державного університету.
Казмірчук Петро Петрович, старший програміст ТОВ «Глобал Лоджик Україна».
Ніколенко Світлана Григорівна, старший викладач кафедри інтелектуальних інформаційних систем Чорноморського державного університету ім. Петра Могили.

Інтелектуальний аналіз даних: практикум [Комплект] / Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. – Л.: «Новий Світ – 2000», 2020. – 162 с. 1 електр. опт. диск (CD-R): додаток.

ISBN 978-966-418-287-7

Практикум з інтелектуального аналізу даних призначений для формування практичних навичок та їх використання при побудови аналітичних інформаційних систем та аналітичних модулів у системах підтримки прийняття рішень. Він складається з двох частин. У першій частині практикуму розглянуто моделі, задачі і алгоритми інтелектуального аналізу даних з подальшою їх реалізацією студентами у програмному середовищі однієї з універсальних мов програмування. Після засвоєння і розуміння сутності моделей й алгоритмів DM студенти переходять до опанування функцій та розв'язку задач DM, реалізованих в середовищі СКБД MS SQL Server 2008, що складають другу частину практикуму. Кожна з частин має комплект завдань для аудиторної і самостійної роботи.

ISBN 978-966-418-287-7

ЗМІСТ

Частина 2. Застосування алгоритмів Datamining у середовищі MS SQL Server 2008

ВСТУП

Виникнення і розвиток моделей і методів інтелектуального аналізу даних, який відомий також як Data mining (видобування/розкопка даних), пов'язане з новим витком у розвитку засобів і методів обробки даних. Спроби застосування методів традиційної математичної статистики для обробки великих обсягів накопиченої інформації привели спочатку до появи напрямку OLAP (On-Line Analytical Processing) – оперативної аналітичної обробки даних. Розвиток цього напрямку, в свою чергу, сприяв появі багатовимірної (хоча б на логічному рівні) моделі даних. Багатовимірні куби або просто OLAP-куби зручні для застосування методів математичної статистики як по вимірах кубів, так по їх сукупностях, що разом з можливостями створення ієрархій за вимірюваннями зробили OLAP-технології затребуваними, а це призвело до розробки відповідних модулів (служб, сервісів) в складі популярних СКБД. Однак методи OLAP не дозволяли здійснювати більш глибокий аналіз даних. Моделі та методи інтелектуального аналізу даних дозволяють досягти більш високого рівня знань, видобутих з даних. Існують різні визначення інтелектуального аналізу даних, серед яких найбільш поширеним є наступне. *Data mining* - це процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних і доступних для інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності. До таких сфер відносяться, наприклад: маркетинг, торговельна діяльність (продажі, закупівлі, ціноутворення), фінансова діяльність (грошовий обіг, бухгалтерський аналіз, бюджетування), соціологічні опитування, використання різноманітних ресурсів, логістика та багато інших.

Як і OLAP-технології, моделі та методи Data mining (DM) також реалізуються у складі сучасних СКБД, однак вони ще не знайшли широкого застосування в інформаційних системах та системах підтримки прийняття рішень. Зокрема, це можна пояснити тим, що реалізовані в програмних продуктах з Data mining алгоритми є «чорною скринькою», тому структурно посібник побудовано з двох частин.

У перший частині наведено алгоритми основних задач Data mining та розглянуто їх реалізацію різними мовами програмування (Pascal, C++, Java), для чого студентам пропонується реалізувати наведені в практикумі алгоритми однією із наведених мов програмування. У цій частині лабораторного практикуму, що складається з 5 лабораторних робіт, передбачено виконання практичних занять з таких тем: класифікація; кластеризація; пошук асоціативних правил; тренд-аналіз (декомпозиція) часових рядів та прогнозування часових рядів з детермінованим трендом; адаптивні методи прогнозування нестаціонарних часових рядів.

У другій частині практикуму студенти мають навчитися будувати моделі і застосовувати методи Data mining у середовищі СКБД MS SQL Server 2008 Вона містить всі необхідні відомості про засоби інтелектуального аналізу даних, що представлені в даній СКБД, включаючи мову DMX. В цих лабораторних роботах використовується навчальна бази даних, яка представляє собою

частину бази даних Adventure Works DW 2008 (компанія Microsoft). База даних стосується продажів велосипедів та аксесуарів до них. До цієї частини практикуму включено чотирі лабораторних роботи, в яких виконуються такі завдання, як: доступ до вхідних даних проекту, встановлення зв'язку з сервером, створення й налаштування нового проекту Analysis Services, додавання нових моделей до структури DM, аналіз і інтерпретація отриманих результатів.

Робота №1 пов'язана з організацією доступу до вхідних даних та створенням проекту **Analysis Services** у середовищі **Business Intelligence Studio.**

В роботі №2 розв'язується задача формування розсилок поштової реклами потенційним покупцям за допомогою методів: дерева рішень, кластеризації, спрощеної моделі мережі Баєса.

Робота №3 стосується прогнозування загальних обсягів продажів на основі часових рядів та прогнозуванню продажів у часі за допомогою побудованої моделі.

В роботі №4 розв'язується задача виявлення асоціативних правил та створення й аналізу на їх основі моделі кошику покупця з використанням алгоритму **Microsoft Association Rules**.

Для виконання лабораторних робіт використовується програмне забезпечення MS SQL Server 2008 R2 та MS Visual Studio 9 (**MS Visual Studio 2008)**. Студенти повинні бути зареєстровані на сервері Database Engine як користувачі з повним переліком прав (папка Security – Logins). Можна надавати групові права, якщо в структуру логіна для Widows входить номер групи. Для роботи з Analysis Server у властивостях сервера вибрати команду Security та додати користувачів з переліку зареєстрованих на Database Engine.

Навчальна база даних та проекти для виконання лабораторних робіт надаються на компакт-диску, що входить до складу практикуму, у вигляді файлів, опис яких наведений в Додатку. Вона, як вже зазначалося, є однією з переліку баз великої виробничої гіпотетичної компанії Adventure Works Cycles, розроблених фірмою Microsoft для демонстрації роботи можливостей Analysis Server. Компанія має основне й допоміжні виробництва, що випускають такі товари як запасні частини, одяг для велосипедистів, аксесуари для велосипедів. Компанія має регіональну торговельну мережу. База має великий обсяг даних.

Склад навчальної бази і зв'язки наведені у файлах **Опис_навчальної_бази.xlsx** та **AdventureWorksDW2008-schema.PDF.**

Лабораторні роботи можна використовувати і з версіями MS SQL Server 2012 або MS SQL Server 2014, враховуючі незначні відхилення в користувацькому інтерфейсі, але навчальну базу відповідної версії тоді треба завантажити з сайту http://sqlserversamples.codeplex.com/.