

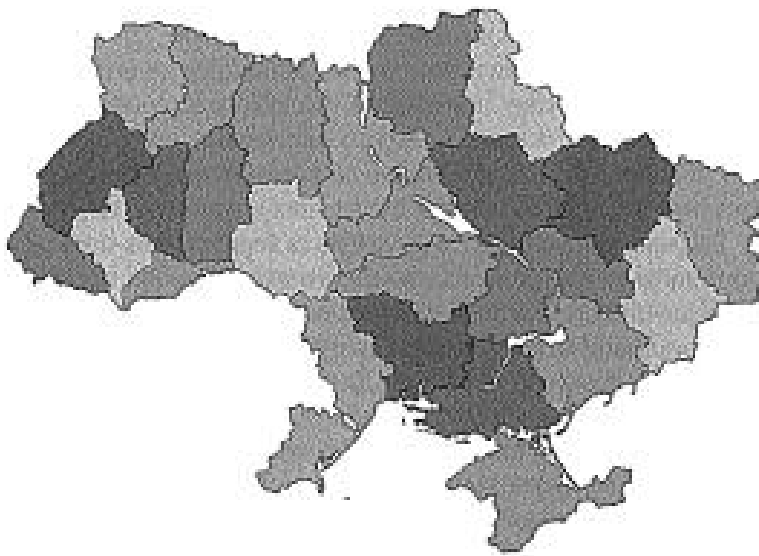
Інститут проблем реєстрації інформації
Національної академії наук України

О. Г. Додонов, А. І. Кузьмичов

Датамайнінг в Excel

Розвідувальний аналіз даних та прогнозування
з використанням надбудови

Analytic Solver Data Mining



Київ
Видавництво Ліра-К
2023

УДК 004.67

Д60

*Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України.
(протокол № 8 від 5 липня 2022 р.)*

Додонов О. Г., Кузьмичов А. І.

Датамайнінг в Excel. Розвідувальний аналіз даних та прогнозування з використанням надбудови Analytic Solver Data Mining. – Київ : Видавництво Ліра-К, 2023. – 240 с.

ISBN 978-617-520-424-5

Розвідувальний аналіз даних та датамайнінг (Exploratory Data Analysis & Data Mining) – міждисциплінарна методологія на тлі «великих даних», новітні інформаційні технології і процедури, зорієнтовані на виявлення вад в наборах табличних даних великих обсягів про об'єкти, згідно поставлених цілей.

Зазвичай ці набори «сирі», якщо отримані із зовнішніх джерел і, скоріше за все, невідомого походження, чи вони відомі і робочі, що регулярно застосовуються, але пошкодженні, випадково чи штучно. З-за їх великих розмірів шукану інформацію «видобувають»/«майнять» із даних досконалими, потужними й дорого вартісними комп'ютерними засобами, аби швидко й із найбільшою достовірністю зрозуміти їх природу і визначити наслідки виявлених негативних впливів на дані.

Видання призначене для студентів та користувачів-початківців з математичною та програмістською підготовкою на рівні повної середньої освіти, що цікавляться проблематикою Data Science, але у яких для практичної реалізації методів єдиним доступним і зрозумілим обчислювачем є табличний процесор MS Excel. Для них, без потреби щось програмувати, використана надбудова Analytic Solver Data Mining у складі ASPE (Analytic Solver Platform for Education, www.solver.com). Її інструменти разом зі стандартними засобами Excel застосовуються для підготовки і розвідки отриманих наборів даних й подальшого розв'язання задач кластеризації, класифікації та передбачення за технікою машинного навчання.

УДК 004.67

| 978-617-520-424-5

© Додонов О. Г., Кузьмичов А. І., 2023
© Видавництво Ліра-К, 2023

ЗМІСТ

Вступ	4
Набори даних: отримання, підготовка і аналіз	12
Вибірковий аналіз	18
Візуалізація даних	29
Відбір впливових змінних/властивостей (Feature Selection)	47
Перетворення (Transform)	58
Інструмент Missing Data Handling	60
Інструменти Transform Continuous Data	64
Інструменти Transform Categorical Data	74
Інструмент Principal Components	77
Кластеризація	89
Інструмент K-Means Clustering	90
Інструмент Hierarchical Clustering	96
Інструмент Text	99
Короткострокове передбачення (Time Series)	117
Інструмент Partition	120
Метод ARIMA	121
Інструмент Lag Analysis	121
Інструмент ARIMA Model	123
Згладжування (Smoothing)	124
Датамайнінг. Машинне навчання	135
Інструменти Partition	136
Датамайнінг. Класифікація	143
Інструмент Discriminant Analysis	143
Інструмент Logistic Regression	149
Інструмент k-Nearest Neighbors	154
Інструмент Classification Tree	158
Інструмент Naive Bayes	168
Інструменти Neural Network	172
Інструменти Ensemble	183
Інструмент Find Best	194
Датамайнінг. Передбачення	204
Інструмент Linear Regression	204
Інструмент k-Nearest Neighbors	210
Інструмент Regression Tree	211
Інструмент Neural Network	213
Інструменти Ensemble	216
Інструмент Find Best	221
Пошук правил асоціації	227
Додатки	233