

УДК [658.8+637]:004.9

Олена Михайленко, старший викладач
кафедри інженерної механіки та комп'ютерного проектування,
Галина Антонова, старший викладач
кафедри інженерної механіки та комп'ютерного проектування,
Іван Водяницький, асистент
кафедри інженерної механіки та комп'ютерного проектування,
Таврійський державний агротехнологічний університет імені Дмитра Моторного,
м. Запоріжжя, Україна

ПРОГНОЗУВАННЯ ПОПИТУ НА ПРОДУКЦІЮ ТВАРИННИЦТВА З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ DATAMINING

Анотація. У роботі досліджено технології DataMining як інструмент для вирішення бізнес-задач, що виникають у процесі функціонування підприємства. Наведено стислий опис програмного забезпечення, розробленого з метою прогнозування попиту на продукцію підприємства. Зазначене ПЗ базується на методах кластерного та регресійного аналізу. Результатом використання програми є прогнозований обсяг продажів продукції на наступний місяць, представлений у певному інтервалі значень. Для поліпшення візуалізації результатів передбачено відображення графіків, що ілюструють роботу програми. Програмне забезпечення розроблено в середовищі програмування Delphi для операційних систем Windows та Android.

Ключові слова: DataMining, кластерний аналіз, регресійний аналіз, прогнозування, кореляція.

Abstract. The paper explores Data Mining technologies as a tool for addressing business challenges that arise during enterprise operations. A concise description of the software developed for forecasting product demand is provided. The aforementioned software is based on cluster and regression analysis methods. The result of using the program is a predicted volume of product sales for the next month, presented within a specific range of values. To enhance the visualization of results, the display of graphs illustrating the program's operation is included. The software is developed in the Delphi programming environment for Windows and Android operating systems.

Keywords: DataMining, cluster analysis, regression analysis, forecasting, correlation.

Вступ. Сучасні організації накопичують значні обсяги даних, проте часто не можуть отримати від їх аналізу відчутну користь. Використовуючи історичні дані,

можна вирішувати критично важливі для бізнесу завдання, такі як оптимізація процесів, управління ризиками, підвищення прибутковості та утримання клієнтів. Обґрунтованість та професіоналізм прийняття управлінських рішень визначають ефективність діяльності компанії. У цьому контексті сучасні інформаційні системи надають керівнику суттєву підтримку.

Для автоматизації бізнес-процесів на виробництві можуть застосовуватися різноманітні технології аналізу даних, серед яких OLAP, KDD та DataMining. OLAP використовується для підготовки бізнес-звітів у сферах продажів, маркетингу, управління, прогнозування, фінансової звітності та подібних областях, забезпечуючи можливість оперативного формування описових та порівняльних зведених даних, а також отримання відповідей на різноманітні аналітичні запити. KDD являє собою процес виявлення знань у базах даних. DataMining (DM) – це метод виявлення у первинних даних раніше невідомих, неочевидних, практично цінних та зрозумілих знань, необхідних для прийняття рішень у різних сферах людської діяльності. Важливим критерієм при виборі технології аналізу даних є швидкість та точність отриманих результатів.

Виклад основного матеріалу. Розглянемо найбільш поширені технології статистичного аналізу даних, які допомагають приймати обґрунтовані управлінські рішення:

1. OLAP (оперативна аналітична обробка) являє собою технологію опрацювання інформації, що забезпечує оперативне отримання відповідей на складні аналітичні запити, які оперують багатьма вимірами даних. OLAP є частиною більш широкої концепції бізнес-аналітики та застосовується для створення бізнес-звітів.

2. KDD (виявлення знань у базах даних) – це процес ідентифікації цінних відомостей у неструктурованих даних. KDD охоплює такі етапи, як підготовка даних, відбір значущих характеристик, очищення даних, застосування методів Data Mining, подальша обробка результатів та їх інтерпретація.

3. Data Mining (DM), або глибинний аналіз даних, є методом виявлення в первинних даних раніше невідомих, неочевидних, практично цінних та зрозумілих знань, які необхідні для прийняття рішень у різноманітних галузях людської діяльності. Після дослідження предметної області було встановлено, що для досягнення вагомих результатів у діяльності підприємства необхідно використовувати методи поглибленої аналітики DataMining.

Розглянемо детальніше основні складові DataMining.

Кластерний аналіз.

Основна мета кластерного аналізу полягає у виявленні груп подібних об'єктів у вибірці. Сфера застосування кластерного аналізу є надзвичайно широкою та охоплює багато галузей людської діяльності.

У кластерному аналізі для кількісної оцінки подібності використовується поняття метрики. Застосовуються різні міри відстані між об'єктами:

- евклідова відстань: $d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$ (1)
- зважена евклідова відстань: $d_{ij} = \sqrt{\sum_{k=1}^m \omega_k (x_{ik} - x_{jk})^2}$; (2)
- відстань city-block: $d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|$; (3)
- відстань Мінківського: $d_{ij} = (\sum_{k=1}^m |x_{ik} - x_{jk}|^p)^{1/p}$; (4)
- відстань Махаланобиса: $d_{ij} = (X_i - X_j)^T S_*^{-1} (X_i - X_j)$. (5)

Найчастіше на практиці використовується евклідова метрика.

Застосування різних алгоритмів кластерного аналізу може призвести до отримання відмінних результатів на одному й тому ж наборі даних, що є типовою ситуацією в аналітиці.

На основі технологій DataMining було створено програмне забезпечення, здатне автоматизувати значний обсяг обчислень та суттєво зменшити час, необхідний для прийняття управлінських рішень. Вхідними даними є групи товарів з інформацією про обсяги їх продажів за попередні періоди. Вихідними

даними є прогнозовані обсяги продажів кожного товару на наступний місяць, представлені у вигляді довірчого інтервалу з визначеною ймовірністю.

Програма функціонує в діалоговому режимі, забезпечуючи взаємодію між користувачем та обчислювальною системою. На кожному етапі розрахунків користувач має змогу відстежувати проміжні результати, а перехід до наступного кроку відбувається лише за його командою.

Інтерфейс розробленого програмного продукту відрізняється максимальною простотою та інтуїтивністю, враховуючи необхідність мінімальної комп'ютерної підготовки аналітика для ефективної роботи з ним.

Створено також додаток для операційної системи Android, що надає можливість виконання розрахунків на мобільних пристроях. Мобільна версія програми зберігає аналогічну логічну структуру з версією для персонального комп'ютера.

Для початку аналізу продукції необхідно імпортувати таблицю з первинними даними. Після цього слід ініціювати процес кластеризації, натиснувши відповідну кнопку.

Для наочного представлення результатів кластерного аналізу передбачено відображення графіку після натискання кнопки «Графік», де різними кольорами позначено товари, віднесені до різних кластерів (рис. 1).

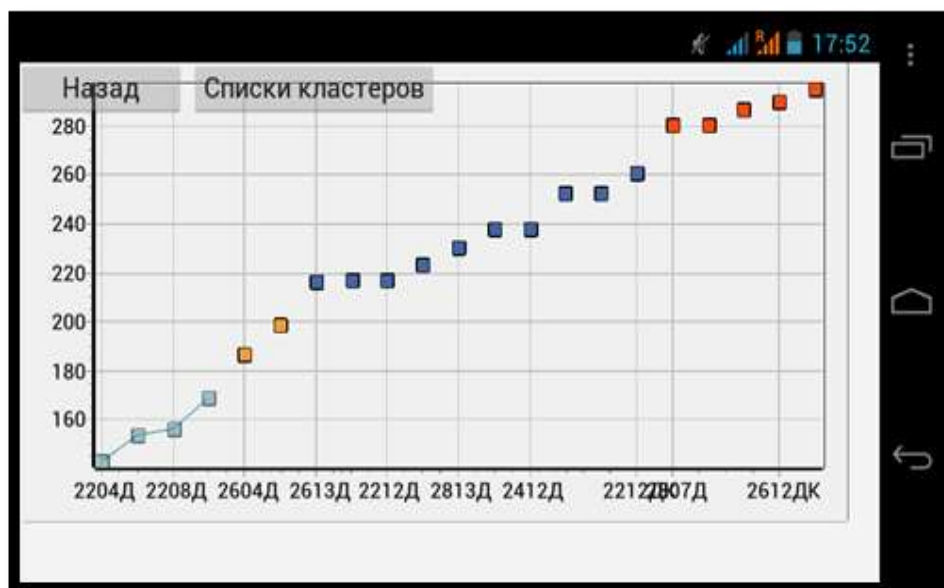


Рисунок 1 – Графік з результатами кластеризації

Для початку регресійного аналізу необхідно натиснути на кнопку «Регресія». На екрані з'явиться регресійна модель з точковим прогнозом та довірчим інтервалом (рис. 2).

Для перегляду текстового повідомлення з прогнозом необхідно натиснути на кнопку «Прогноз». Результати прогнозування зображено на рис. 3.



Рисунок 2 – Графік прогнозу продажів

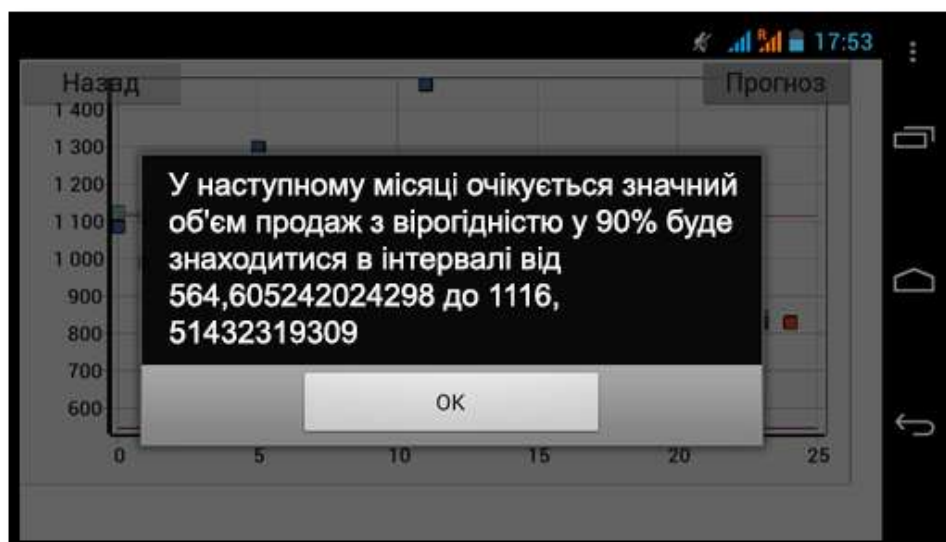


Рисунок 3 – Результати прогнозування

Висновки: У представлений роботі здійснено огляд існуючих підходів до аналізу даних. За результатами проведеного дослідження встановлено, що для статистичного аналізу та автоматизації виробничих бізнес-процесів найбільш доцільним є застосування методів DataMining, що дозволяє отримати відчутну практичну користь для підприємства.

На базі технологій DataMining розроблено програмні рішення, спрямовані на автоматизацію бізнес-процесів, що сприяє підвищенню продуктивності підприємства та оптимізації робочого навантаження аналітиків. Створена Android-версія програмного забезпечення надає можливість виконання розрахунків за допомогою мобільних пристроїв, що є особливо зручним для співробітників, які не мають постійного доступу до стаціонарних комп'ютерів.

Список використаних джерел

1. Радченко Станислав Григорьевич. Методология регрессионного анализа: Монография. К.: «Корнийчук», 2011. С. 376.
2. Вершков О. О., Мацулевич О. Є., Дереза О. О. Загальні налаштування системи MASTERCAM для виконання завдань з розробки управляючих програм токарної обробки валів. *Розвиток сучасної науки та освіти: реалії, проблеми якості, інновації : матеріали V Міжнародної наук.-практ. інтернет-конф.* (м. Запоріжжя, 29-31 травня 2024 р.). Запоріжжя : ТДАТУ, 2024. С. 100-105.
3. Дереза О.О., Антонова Г.В., Тетервак І.Р., Валієва К.М. Аналітичні дослідження методики інтелектуального аналізу даних. *Розвиток сучасної науки та освіти: реалії, проблеми якості, інновації: матеріали IV Міжнародної наук.-практ. інтернет-конф.* (м. Запоріжжя, 29-31 травня 2023 р.) / [за наук. ред. С. В. Кюрчев, В. О. Радкевич, В. М. Кюрчев та інш.]. Запоріжжя : ТДАТУ, 2023. С. 147-153
4. Мацулевич О.Є. Застосування спеціалізованої PLM-системи Technologi CS при розробці автоматизованої системи ведення конструкторсько-технологічних баз даних підприємства сільськогосподарського машинобудування. *Праці Таврійського державного агротехнологічного університету: Наукове фахове видання / ТДАТУ; Запоріжжя : ТДАТУ, 2024. Вип. 24, том. 1.* <https://doi.org/10.32782/2078-0877-2024-24-1-13>.