МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Запорізький національний технічний університет

Інформаційні системи та технології в управлінні

МЕТОДИЧНІ ВКАЗІВКИ

теоретичні відомості і завдання до лабораторних робіт

для студентів та магістрів денної форми навчання
спеціальності 7.803060101
Менеджмент організацій і адміністрування
Частина 2
Кластерний аналіз у бізнес-аналітіці.

2014

Інформаційні системи та технології в управлінні. Методичні вказівки, теоретичні відомості і завдання до лабораторних робіт для студентів та магістрів денної форми навчання спеціальності 7.803060101 Менеджмент організацій і адміністрування. Частина 2. Кластерний аналіз у бізнес-аналітіці. / Укл.: Біла Н.І. – Запоріжжя: ЗНТУ, 2014. – с. 38.

Містить теоретичні відомості, індивідуальні завдання та приклади за темою «Кластерний аналіз» із курсу «Інформаційні системи та технології в управлінні»

Укладачі: Біла Н.І. доцент

Рецензенти: Пінчук В.П., доцент
Вишневська В.Г., доцент.

Відповідальний за випуск Корніч Г.В., зав. кафедрою, професор

Затверджено на засіданні кафедри
обчислювальної математики,
протокол № 6 від 28.02.2014

ЗМІСТ

4 КЛАСТЕРНИЙ АНАЛІЗ У БІЗНЕС-АНАЛІТИЦІ

4.1. Теоретичні основи кластерного аналізу

Кластерний аналіз являє собою статистичні методи, що використовуються для класифікації багатомірних об'єктів або подій у відносно однорідні групи, які називають кластерами. Об'єкти в кожному кластері повинні бути схожі один на одного більше, ніж на об'єкти інших класів, і відрізнятися від об'єктів інших кластерів сильніше, чим від об'єктів власного класу.

В економіці кластерний аналіз використовується для досягнення таких цілей: сегментації ринку, вивчення поведінки покупців, визначення конкурентоспроможності нового товару, скорочення розмірності даних і ін.

Кластерний аналіз застосовується в різних областях. Він корисний, коли потрібно класифікувати велику кількість інформації.

Так, у медицині використовується кластеризація захворювань, лікування захворювань або їх симптомів, а також таксономія пацієнтів, препаратів і т.д. В археології встановлюються таксономії кам'яних споруджень і прадавніх об'єктів і т.д. У маркетингу це може бути задача сегментації конкурентів і споживачів. У менеджменті прикладом задачі кластеризації буде розбивка персоналу на різні групи, класифікація споживачів і постачальників, виявлення схожих виробничих ситуацій, при яких виникає брак. У медицині - класифікація симптомів. У соціології задача кластеризації - розбивка респондентів на однорідні групи.

У маркетингових дослідженнях кластерний аналіз застосовується досить широко - як у теоретичних дослідженнях так і маркетологами, що практикують і вирішують проблеми угруповання різних об'єктів. При цьому вирішуються питання про групи клієнтів, продуктів і т.д. Так, однією з найбільш важливих задач при застосуванні кластерного аналізу в маркетингових дослідженнях є аналіз поведінки споживача, а саме: угруповання споживачів в однорідні класи для одержання максимально повної картини про поведінку клієнта з кожної групи й про фактори, що впливають на його поведінку.

Важливою задачею, яку може розв'язати кластерний аналіз, є позиціонування, тобто визначення ніші, у якій слід позиціонувати новий продукт, пропонований на ринку. У результаті застосування

кластерного аналізу будується карта, по якій можна визначити рівень конкуренції в різних сегментах ринку й відповідні характеристики товару для можливості влучення в цей сегмент. За допомогою аналізу такої карти можливе визначення нових, незайнятих ніш на ринку, у яких можна пропонувати існуючі товари або розробляти нові.

Кластерний аналіз також може бути зручний, наприклад, для аналізу клієнтів компанії. Для цього всі клієнти групуються в кластери, і для кожного кластера виробляється індивідуальна політика. Такий підхід дозволяє суттєво скоротити об'єкти аналізу, і, у той же час, індивідуально підійти до кожної групи клієнтів.

Кластеризацію використовують, коли відсутні апріорні відомості щодо класів, до яких можна віднести об'єкти досліджуваного набору даних, або коли число об'єктів велике, що утрудняє їхній ручний аналіз.

Постановка задачі кластеризації складна й неоднозначна, тому що:

- оптимальна кількість кластерів у загальному випадку невідома;
- вибір міри «подібності» або близькості властивостей об'єктів між собою, як і критерію якості кластеризації, часто носить суб'єктивний характер.

Розповсюдженою мірою оцінки близькості між об'єктами є метрика, або спосіб завдання відстані. Найбільш популярні метрики – евклідова відстань і відстань Манхеттена.

Важливо розуміти, що сама по собі кластеризація не приносить яких-небудь результатів аналізу. Для одержання ефекту необхідно провести змістовну інтерпретацію кожного кластера. Така інтерпретація припускає присвоєння кожному кластеру ємної назви, що показує його суть. Для інтерпретації аналітик детально досліджує кожний кластер: його статистичні характеристики, розподіл значень ознак об'єкта в кластері, оцінює потужність кластера – число об'єктів, що потрапили в нього.

Звичайно в задачах кластерного аналізу вхідні дані представляють у формі прямокутної таблиці, кожний рядок якої представляє результат виміру p ознак на відповідному об'єкті:

$$X = \begin{pmatrix} x_{11} x_{12} \cdots x_{1p} \\ x_{21} x_{22} \cdots x_{2p} \\ \vdots \\ x_{n1} x_{n2} \cdots x_{np} \end{pmatrix}, \qquad (4.1)$$

де n - число об'єктів, що підлягають кластеризації.

Числові значення ознак, що входять у матрицю, можуть відповідати трьом типам змінних: якісним (або категорійним), ранговим і кількісним. *Якісні змінні*, як правило, приймають декілька різних значень, яким, хоча й можна поставити у відповідність деякі числа, але ці числа не будуть відбивати яку-небудь упорядкованість значень якісних змінних. І це потрібно враховувати при визначенні близькості. Значення *рангових змінних*, на відміну від якісних, упорядковані. Їх можна пронумерувати натуральними числами. Однак арифметичні операції над цими числами не мають сенсу. *Кількісні змінні* мають властивість упорядкованості, і над ними, на відміну від інших, можна виконувати арифметичні операції.

Бажане, щоб уся таблиця вхідних даних відповідала одному типу змінних. Якщо це не так, то різні типи змінних намагаються звести до якогось одного типу змінних. Найпростішою є процедура зведення до якісних змінних. Суть цієї процедури в наступному. Якщо є кількісні дані, то вони спочатку зводяться до рангових, для чого область значень кількісних змінних розбивається на інтервали, які нумеруються числами натурального ряду. Рангові змінні можна вважати якісними, якщо не враховувати впорядкованість їх значень. У свою чергу, якісні змінні переводяться в дихотомічні за наступним правилом. Кожне з можливих значень якісної змінної заміняється на 1, якщо якісна змінна прийняла це значення, і 0 - а якщо ні.

У тих випадках, коли всі показники кількісні, часто виникає проблема їх нормування, оскільки відмінність в одиницях виміру робить ці показники непорівнянними. Так, наприклад, при класифікації промислових підприємств за результатами фінансово-господарчої діяльності в опис включаються такі показники, як прибуток, рентабельність, собівартість, коефіцієнт поточної ліквідності і т.д. По прибутку підприємства можуть різнитися на десятки й сотні тисяч одиниць, а по рентабельності - на одиниці, а то й

десяті частки одиниці. Така непорівнянність практично перекреслює ідею багатомірної класифікації, тому що вона автоматично буде здійснюватися по більш масштабному показникові. Тому процедурі безпосереднього рознесення об'єктів по класах повинна передувати процедура приведення всіх показників до порівнянного виду, яку прийнято називати нормуванням. У практичних розрахунках частіше за інші використовуються два підходи до нормування. Один з них пов'язаний з ідеєю статистичної стандартизації, здійснюваної по формулі:

$$x_{ij}^{H} = \frac{x_{ij} - \bar{x}_j}{\sigma_i}, \qquad (4.2)$$

де x_{ij}^{H} - нормований j -ий показник і -го об'єкта;

x_{ij} – значення j -го показника і -го об'єкта;

$\bar{x}_j$ - середнє значення j-го показника по всій множині об'єктів, що кластеризуються;

σ_i - середньоквадратичне відхилення j -го показника.

При використанні такого нормування всі показники, що описують об'єкт, приводяться до виду, коли середнє дорівнює 0, а розкид навколо середнього - 1.

Другий підхід передбачає перетворення показників шляхом відображення інтервалу їх можливих значень на проміжок [0;l]. Це здійснюється за допомогою формули:

$$x_{ij}^{H} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}, \qquad (4.3)$$

де $x_j^{\min} = \min_i x_{ij}$; $x_j^{\max} = \max_i x_{ij}$.

Таким чином, за допомогою нормування вдається позбутися небажаного впливу різномасштабності показників на ступінь схожості між об'єктами.

Вибір виміру подібності є одним з вузлових моментів у задачах кластеризації, тому що від неї, в основному, залежить при даному алгоритмі кластеризації остаточний варіант розбивки об'єктів на класи. У кожному конкретному випадку цей вибір здійснюється

залежно від мети дослідження й природи самих об'єктів, що поділяються на групи.

4.2 Виміри близькості в алгоритмах кластеризації

Відстані між об'єктами можна уявити, якщо об’єкти представити у вигляді точок m-мірного простору R^m. У цьому випадку можуть бути використані різні підходи до обчислення відстаней.

Розглянуті нижче виміри визначають відстані між двома точками, що належать простору вхідних змінних. Використовуються наступні позначення:

$X_Q \subseteq R^m$ — множина даних, що є підмножиною m-мірного простору;

$x_i = (x_{i1}, \ldots, x_{im}) \in X_Q$, $i = \overline{1,Q}$ — елементи множини даних;

Евклідова відстань - обчислюється в такий спосіб:

$$d_2(x_i, x_j) = \sqrt{\sum_{t=1}^{m} (x_{it} - \overline{x_{jt}})^2} \qquad (4.4)$$

Відстань по Хемінгу є просто середньою різниць по координатах. У більшості випадків даний вимір відстані приводить до таких же результатів, як і для звичайної відстані Евкліда, однак для неї вплив окремих великих різниць (викидів) зменшується (тому що вони не зводяться у квадрат). Відстань по Хемінгу обчислюється по формулі:

$$d_H(x_i, x_j) = \sum_{t=1}^{m} | x_{it} - x_{jt} | \qquad (4.5)$$

Цю відстань називають також *манхетенська відстань* (відстань міських кварталів), або " сіті-блок" відстань.

Відстань Чебишева. Ця відстань може виявитися корисною, коли бажають визначити два об'єкти як "різні", якщо вони різняться по якій-небудь одній координаті (яким-небудь одним виміром). Відстань Чебишева обчислюється по формулі

$$d_\infty(x_i, x_j) = \max_{1 \le t \le m} | x_{it} - x_{jt} | \qquad (4.6)$$

4.3 Методи кластеризації

При виконанні кластеризації важливо, скільки в результаті повинно бути побудовано кластерів. Передбачається, що кластеризація повинна виявити природні локальні згущення об'єктів. Тому число кластерів є параметром, що часто суттєво ускладнює алгоритм, якщо передбачається невідомим, й суттєво впливає на якість результату, якщо воно відомо.

Алгоритми кластеризації звичайно будуються як деякий спосіб перебору числа кластерів і визначення його оптимального значення в процесі перебору.

Методи кластерного аналізу можна розділити на дві групи:

- ієрархічні;
- неієрархічні.

Кожна із груп включає багато підходів і алгоритмів.

Використовуючи різні методи кластерного аналізу, аналітик може одержати різні розв'язки для тих самих даних. Це вважається нормальним явищем.

Ієрархічні методи кластерного аналізу

Суть ієрархічної кластеризації полягає в послідовнім об'єднанні менших кластерів у більші або поділі більших кластерів на менші.

Ієрархічні агломеративні методи (Agglomerative Nesting, AGNES) характеризуються послідовним об'єднанням вхідних елементів і відповідним зменшенням числа кластерів. На початку роботи алгоритму всі об'єкти є окремими кластерами. На першому кроці найбільш схожі об'єкти поєднуються в кластер. На наступних кроках об'єднання триває доти, поки всі об'єкти не будуть становити один кластер.

Ієрархічні дивизимні (розділюючи) методи (Divisive Analysis, DIANA) є логічною протилежністю агломеративним методам. На початку роботи алгоритму всі об'єкти належать одному кластеру, який на наступних кроках ділиться на менші кластери, у результаті утворюється послідовність груп, що розщеплюються.

Принцип роботи груп методів, що описані вище, показаний на рис. 4.1 у вигляді дендрограми.

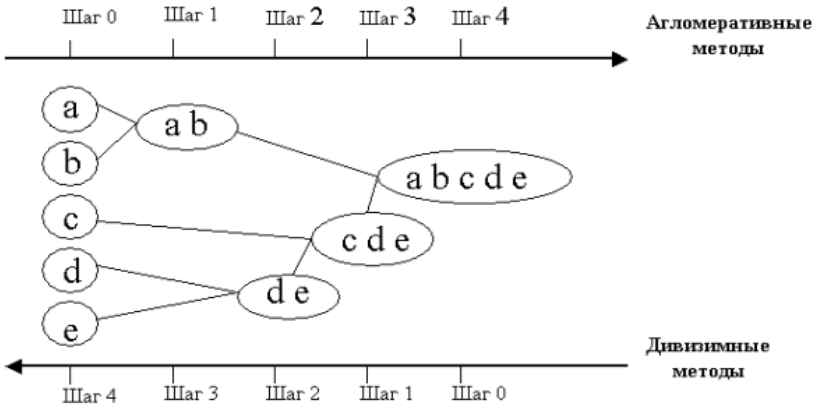


Рисунок 4.1 - Дендрограма агломеративних і дивизимних методів

Ієрархічні методи кластеризації різняться правилами побудови кластерів. У якості правил виступають критерії, які використовуються при вирішенні питання про "схожість" об'єктів при їхньому об'єднанні в групу (агломеративні методи) або поділу на групи (дивизимні методи).

Ієрархічні методи кластерного аналізу використовуються при невеликих об'ємах наборів даних. Перевагою ієрархічних методів кластеризації є їхня наочність.

Ієрархічні алгоритми пов'язані з побудовою дендрограм (від грецького dendron - "дерево"), які є результатом ієрархічного кластерного аналізу. Дендрограма описує близькість окремих точок і кластерів друг до друга, представляє в графічному вигляді послідовність об'єднання (поділу) кластерів. Дендрограма (dendrogram) - деревоподібна діаграма, що містить n рівнів, кожний з яких відповідає одному із кроків процесу послідовного укрупнення кластерів.

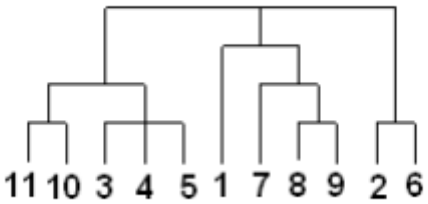


Рисунок 4.2 - Приклад вертикальної дендрограми

Числа 11, 10, 3 і т.д. відповідають номерам об'єктів або спостережень вхідної вибірки. Ми бачимо, що на першому кроці кожне спостереження представляє один кластер, на другому кроці спостерігаємо об'єднання таких спостережень: 11 і 10; 3, 4 і 5; 8 і 9; 2 і 6. На другому кроці триває об'єднання в кластери: спостереження 11, 10, 3, 4, 5 і 7, 8, 9. Даний процес триває доти, поки всі спостереження не об'єднаються в один кластер.

Неієрархічні методи кластерного аналізу

Однією із широко використовуваних методик кластеризації є розділова кластеризація, відповідно до якої для вибірки даних, що містить n записів (об'єктів), задається число кластерів k, яке повинне бути сформоване. Потім алгоритм розбиває всі об'єкти вибірки на k груп (k<n), які і являють собою кластери.

До найбільш простих і ефективних алгоритмів кластеризації відноситься k-means або в україномовному варіанті k-середніх. Він складається із чотирьох кроків.

Алгоритм k-means

Конструктивно алгоритм являє собою ітераційну процедуру, що складається з наступних кроків.

1. Задається число кластерів k, яке повинне бути сформоване з об'єктів вхідної вибірки.

2. Випадковим чином вибирається k записів, які будуть служити початковими центрами кластерів. Початкові точки, з яких потім виростають кластери, часто називають «насінням». Кожний такий

запис являє собою «ембріон» кластера, що складається тільки з одного елемента.

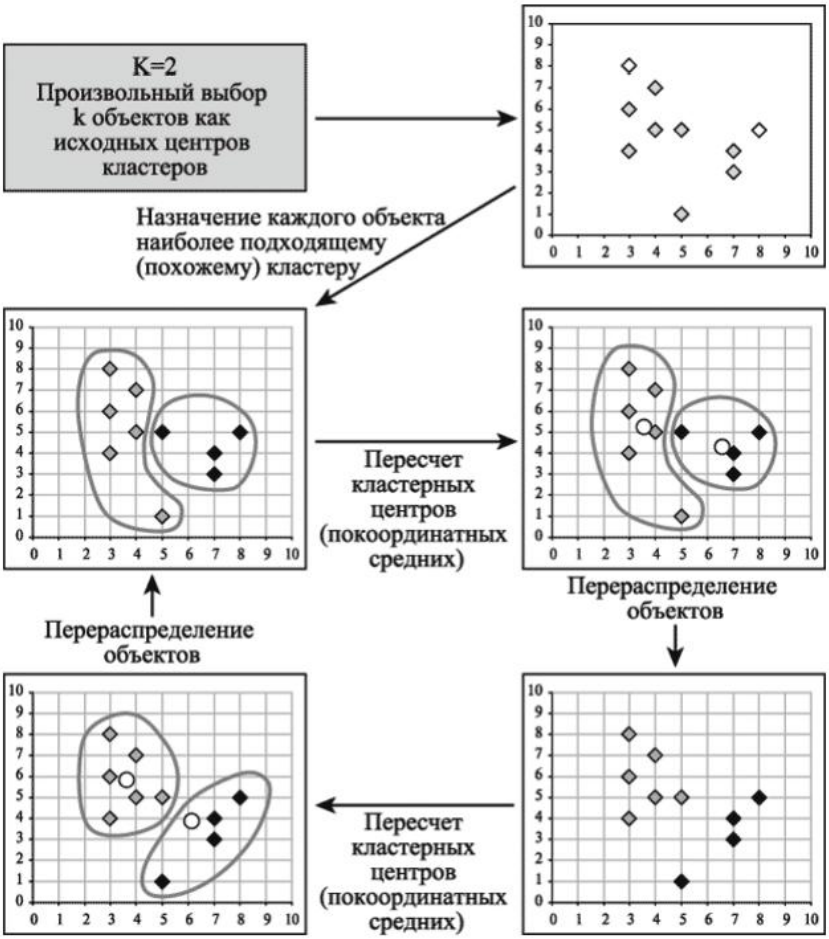


Рисунок 4.3 – Приклад роботи алгоритму k-means

3. Для кожного запису вхідної вибірки визначається найближчий до неї центр кластера.

4. Проводиться обчислення центроїдів – центрів ваги кластерів. Це робиться шляхом визначення середнього для значень кожної

ознаки всіх записів у кластері. Наприклад, якщо в кластер увійшли три записи з наборами ознак (x_1, y_1), (x_2, y_2), (x_3, y_3), то координати його центроїду будуть розраховуватися в такий спосіб:

$$(x, y) = \left(\frac{x_1 + x_2 + x_3}{3}, \frac{y_1 + y_2 + y_3}{3} \right).$$

Потім старий центр кластера зміщається в його центроїд. Таким чином, центроїди стають новими центрами кластерів для наступної ітерації алгоритму.

Кроки 3 і 4 повторяться доти, поки виконання алгоритму не буде перервано або поки не буде виконана умова відповідно до деякого критерію збіжності.

Зупинка алгоритму проводиться, коли границі кластерів і розташування центроїдів перестають змінюватися, тобто на кожній ітерації в кожному кластері залишається той самий набір записів. Алгоритм k-means звичайно знаходить набір стабільних кластерів за кілька десятків ітерацій.

4.4 Основні етапи кластерного аналізу

1. Вибір змінних – критеріїв для кластеризації.

Вибір об'єктів і показників для кластерного аналізу потребує одержання репрезентативної вибірки, що дозволяє переносити результати, отримані на обмеженій кількості об'єктів, на всю генеральну сукупність. Якщо ставиться завдання типології об'єктів, необхідно виділити систему інформативних ознак, які найбільше повно характеризують досліджувані об'єкти. Але, не обов’язково треба включати всі змінні до кластерного аналізу.

2. Вибір засобу виміру відстані між кластерами.

Найпоширенішою мірою для визначення відстані між двома точками на площині, утвореній координатними осями x і y, є евклідова міра. Часто за замовченням використовується квадрат евклідової відстані. Але поряд з евклідовою мірою відстані, можна обрати й інші дистанційні міри. Кластерний аналіз можна проводити не тільки зі змінними інтервальної шкали, але і з категорійними змінними, та у таких випадках застосовуються вже інші дистанційні міри.

3. Стандартизація (нормування) спостережень.

Рівні значень змінних часто дуже сильно відрізняються один від одного. Відповідно до формули евклідової міри, змінна, що має велике значення, практично цілком домінує над змінною з малими значеннями. Рішенням цієї проблеми є нормування (стандартизація) значень змінних за формулами (4.2) або (4.3).

Стандартизація приводить значення всіх перетворених змінних до єдиного діапазону значень, і якщо вона виконується за формулою (4.2), то середнє кожної приводиться до 0, а середнє відхилення – до 1. Тоді всі спостереження змінюються приблизно у діапазоні від –3 до +3.

4. Формування кластерів.

Існує дві основні групи методів формування кластерів : ієрархічні та неієрархічні. Перші поділяються на метод злиття та метод подрібнення. У першому випадку існуючі кластери розширюються шляхом об'єднання, доки не буде сформований один – єдиний кластер, що об'єднує всі спостереження. Метод подрібнення заснований на зворотній операції: спочатку всі спостереження об'єднуються у єдиний кластер, а потім починається процес розділення його на частини. Частіше використовують метод злиття .

З неієрархічних методів найбільш поширений метод k – середніх.

В прикладах будемо використовувати обидві групи методів. Причому, ієрархічні методи допоможуть нам визначити кількість кластерів.

5. Інтерпретація результатів .

Цей етап є достатньо складним і залежить від мети дослідника. На жаль, виразна картина відносин між змінними зустрічається не дуже часто. По-перше, структури кластерів, якщо отримуються, не так чітко розділені, особливо при наявності великої кількості спостережень. Скоріше навпаки: кластери розмиті і навіть проникають один в один. По-друге, як правило, кластерний аналіз проводиться з великою кількістю змінних, що ускладнює аналіз.

4.5 Вирішення задач кластеризації в програмі Deductor

Приклад 4.1. Кластерний аналіз країн – членів європейського союзу і України за критеріями податкової політики.

Постановка задачі

Розглянемо кластерний аналіз країн по ступеню їх подібності на основі певних показників, що характеризують критерії податкової політики держави. Результатом аналізу стане формування регіональних кластерів, що дозволить враховувати особливості розвитку країн-членів ЄС при здійсненні податкової реформи в Україні.

Стратегія формування ефективної податкової політики є неоднозначною в різних країнах, що обумовлене істотними відмінностями регіональних податкових систем за рівнем оподатковування, їх складом і структурою, досягнутим рівнем розвитку економіки, умовами входження в ринкове середовище, конкурентними перевагами, а також визначальними факторами гармонізації оподатковування в умовах глобалізації економічних систем. Однак податковий потенціал держави характеризується певною системою показників податкової політики, які можна об'єднати в кілька груп, виходячи з науково обґрунтованих критеріїв її реалізації. В окремі групи можна об'єднати показники, які характеризують фіскальну достатність, економічну ефективність і соціальну справедливість податкової політики окремої країни. Також аналізуються показники, які враховують тенденції соціально-економічного розвитку країни.

Разом з тим є очевидним, що всі ці показники, які характеризують політику в сфері оподатковування, тісно зв'язані між собою й мають деякий ступінь подібності закономірностей їх розвитку.

Метою роботи є об'єднання європейських країн у більші групи по ступеню їх подібності на основі певних показників, що характеризують фіскальну достатність, економічну ефективність, соціальну справедливість і гнучкість національної податкової політики. Кластерний аналіз європейських країн за основними критеріями податкової політики включає кілька основних етапів.

Допустимо, що є набір об'єктів O_i (країн), i=1,2,…,28 – податкової політики країн-членів ЄС і України, кожна з яких описується сукупністю ознак P_j, j=1,2,…,m – показниками критеріїв податкової політики. Остаточно для проведення кластерного аналізу були відібрані такі показники:

- загальне податкове навантаження, %
- ефективна ставка податків на споживання, %

- ефективна ставка податків з капіталу, %
- ефективна ставка податків на працю, %
- темп росту реального ВВП, % до попереднього періоду.

Таблиця 4.1 - Дані для кластерного аналізу країн за податками

	Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП
▶	UA	37,3	22,8	24,4	63,2	28,3	102,3
	BG	33,3	26,4	16,9	71,7	42,6	106
	EE	32,2	20,9	10,7	71,5	33,7	96,4
	LT	30,3	17,5	12,4	78,5	33	102,8
	LV	28,9	17,5	16,3	48,4	28,2	102,3
	RO	28	17,7	19,6	26,2	29,5	107,3
	SK	29,1	18,4	16,7	84,6	33,5	106,2
	AT	42,8	22,1	27,3	150,2	41,3	102,2
	BE	44,3	21,2	32,7	97,1	42,6	101
	DE	39,3	19,8	23,1	193,8	39,2	101
	ES	33,1	14,1	32,8	114,6	30,5	100,9
	FI	43,1	26	28,1	114,5	41,3	100,9
	FR	42,8	19,1	38,8	160,7	41,4	100,2
	IT	42,8	16,4	35,3	187,4	42,8	98,7
	NL	39,1	26,7	17,2	189,8	35,4	101,9
	PT	36,7	19,1	19,6	143,4	29,6	100,2
	SE	47,1	28,4	27,9	190,1	42,1	99,6
	CZ	36,1	21,1	21,5	127,1	39,5	102,5
	EL	32,6	15,1	15,8	102	37	102
	HU	40,4	26,9	19,2	98	42,4	100,6
	PL	34,3	21	22,5	108	32,8	105
	SI	37,3	23,9	21,6	121,7	35,7	103,5
	CY	39,2	20,6	36,4	110	24,5	103,6
	DK	48,2	32,4	43,1	267,8	36,4	99,1
	IE	29,3	22,9	15,7	153,1	24,6	97
	LU	35,6	27,1	39,6	173,3	31,5	100,1
	MT	34,5	20	42,1	197	20,2	101,7
	UK	37,3	17,6	45,9	180,2	26,1	99,9

Наступним етапом кластерного аналізу є обчислення деякого ступеня подібності або відмінності між вибраними для аналізу об'єктами. Це означає перехід від таблиці «об'єкт – ознака» до таблиці «об'єкт – об'єкт», де d_{ij} – вимір подібності або відмінності між об'єктами O_i і O_j. Як вимір подібності був використаний найпоширеніший показник – евклідова відстань:

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

де x_{ik} - значення k-го показника в і-ом об'єкті, і = 1,2,…, m.

Матриця евклідових відстаней відображає подібність і відмінність у податкових політиках різних країн. Чим менше значення, тим вище ступінь подібності двох країн і комбінацій у кластері. І навпаки, чим більше відповідне значення, тем більша відмінність між країнами.

У результаті цього етапу можна побудувати за допомогою пакетів STATISTICA 8 або MATLAB 7 деревоподібну діаграму, яка дає перше уявлення про кількість можливих кластерів (рис. 4.4).

У результаті побудови дендрограми можна сформувати гіпотезу про наявність як мінімум чотирьох кластерів.

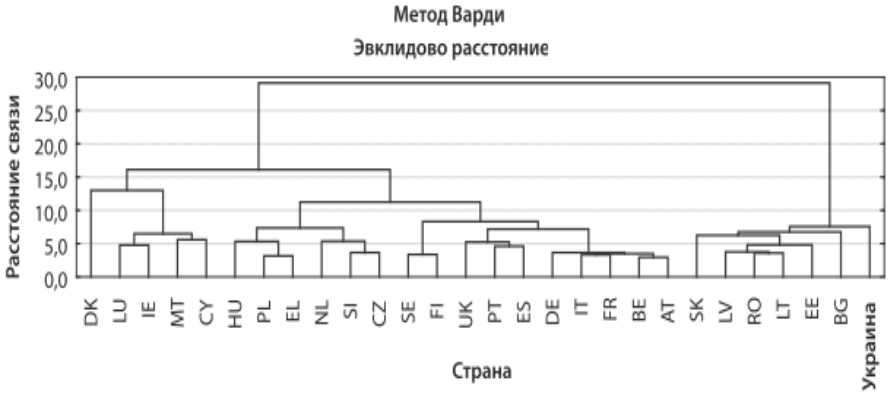


Рисунок 4.4 - Дендрограма відстаней між країнами за критеріями податкової політики

Наявність різкого стрибка середніх відхилень показників можна інтерпретувати як характеристику числа кластерів, які об'єктивно існують у досліджуваній сукупності. Тобто на кроці, де значення коефіцієнта збільшується стрибкоподібно, процес об'єднання в нові кластери необхідно зупинити, оскільки інакше були б об'єднані кластери, які знаходяться на відносно великій відстані один від іншого. Отже, гіпотеза про кількість кластерів (а саме, чотири) країн, які формуються на основі відповідних критеріїв податкової політики, підтверджується і є остаточною.

Кластерний аналіз у програмі Deductor

Спочатку необхідно здійснити імпорт розглянутих даних з файлу – Податки.txt. Результати імпорту показані в табл. 4.1.

Після цього вибираємо й запускаємо Майстер обробки "Кластеризація". При запуску Майстра необхідно настроїти призначення стовпців, тобто вибрати властивості, по яких буде відбуватися угруповання об'єктів. Укажемо стовпцю "Країна" призначення "Інформаційне", а іншим стовпцям – "Вхідне". (Рис. 4.5).

На наступному кроці Майстра необхідно настроїти спосіб поділу вхідної множини даних на тестове й навчальне, а також кількість прикладів у тій і іншій множині. Укажемо, що дані обох множин беруться випадковим чином, і визначимо всю множину як навчальну (100%).

Наступний крок пропонує настроїти параметри кластеризації, визначити на яку кількість кластерів буде розподілятися вхідна множина. Оберемо фіксовану кількість кластерів – чотири.

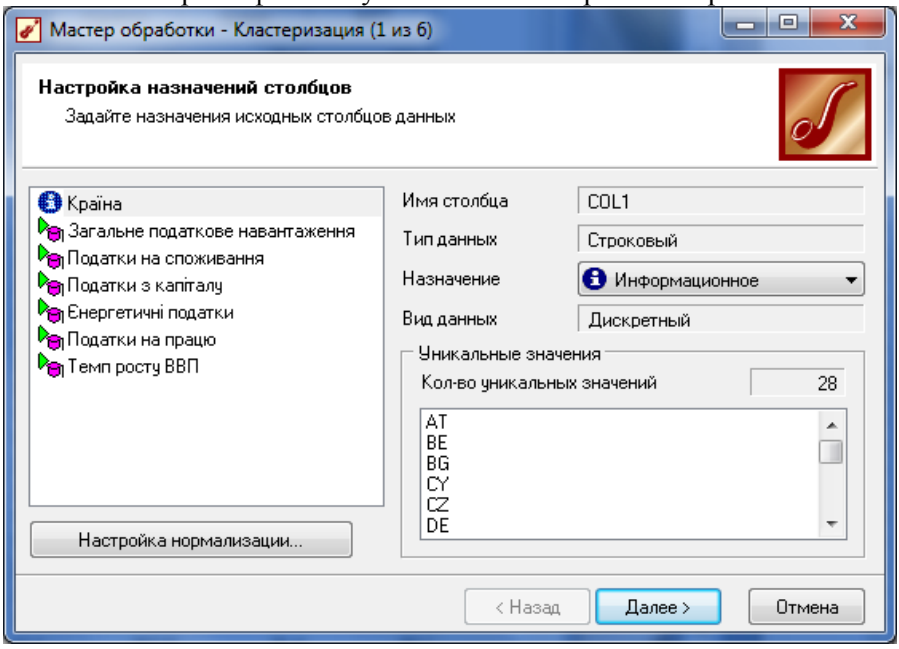


Рисунок 4.5 – Настроювання призначень стовпців

Для відображення отриманих груп кластерів виберемо в оброблювачі «Кластеризація» зі списку візуалізаторів такі способи відображення даних: 1) «Профілі кластерів» для визначення структури формування групи кластерів і 2) «Куб» для наочного перегляду отриманих результатів (рис. 4.6).

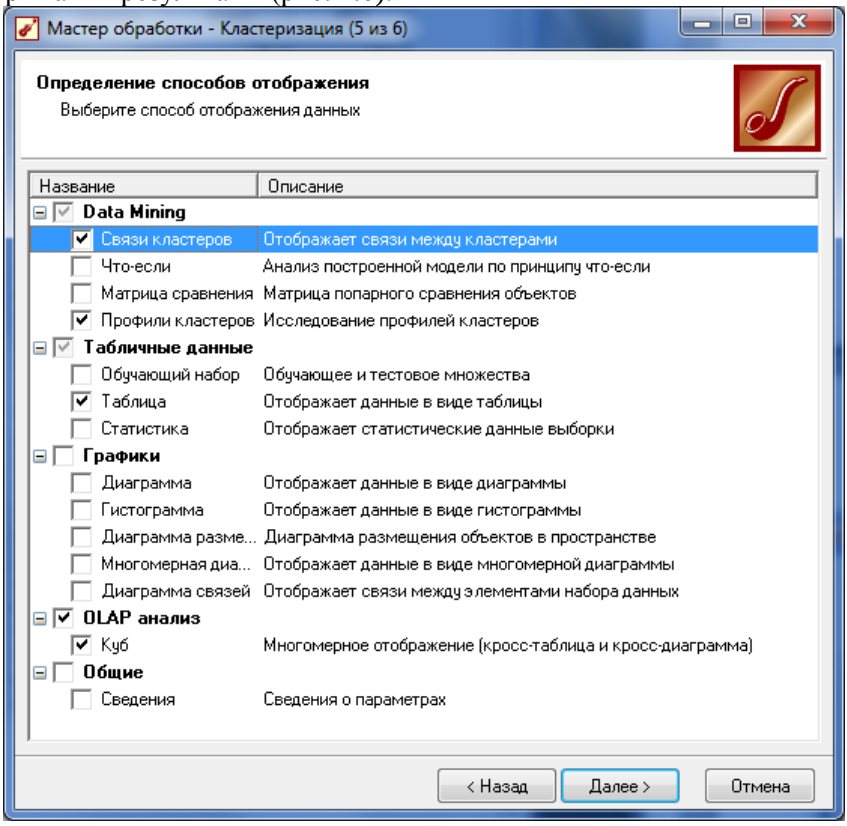


Рисунок 4.6 – Настроювання візуалізаторів

Для настроювання візуалізатора «Куб» необхідно вибрати розглянуті властивості як факти, а номер кластера як вимір (рис. 4.7).

Далі необхідно визначити як в таблиці розташовувати виміри і факти (рис. 4.8).

Найбільш правильно в подальших настроюваннях задати відображення фактів як середнє по розглянутій групі (рис. 4.9).

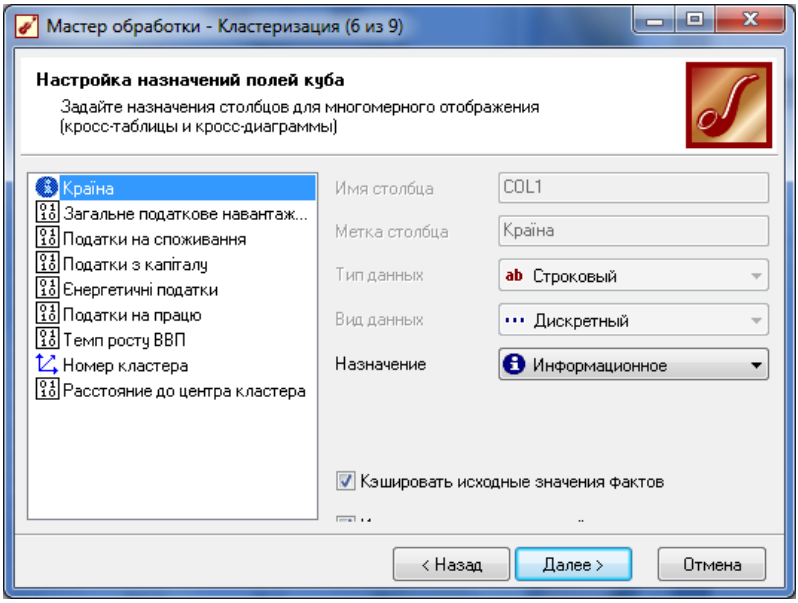


Рисунок 4.7 – Настроювання полів куба

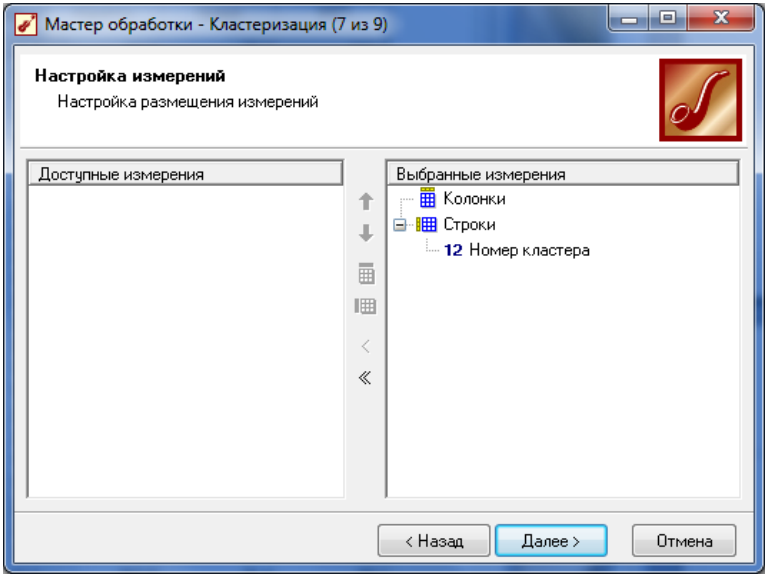


Рисунок 4.8 – Настроювання вимірів

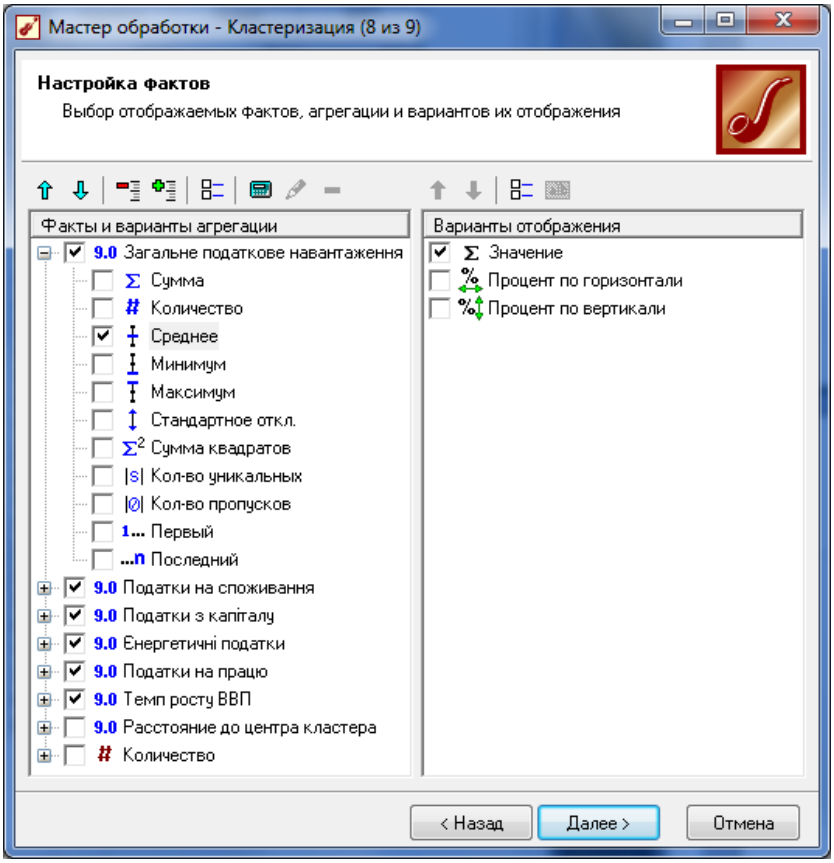


Рисунок 4.9 – Вибір фактів та їхньої агрегації

Загальну структуру сформованих алгоритмом кластерів можна переглянути у візуалізаторі «Профілі кластерів». У ньому представлені всі розглянуті властивості разом з характером впливу їх на склад кластера (рис. 4.10).

Основним фактором, що визначає склад кластера, є значимість властивостей, виражена у відсотках. Загальна значимість розглянутого поля визначається варіабельністю її розглянутих параметрів.

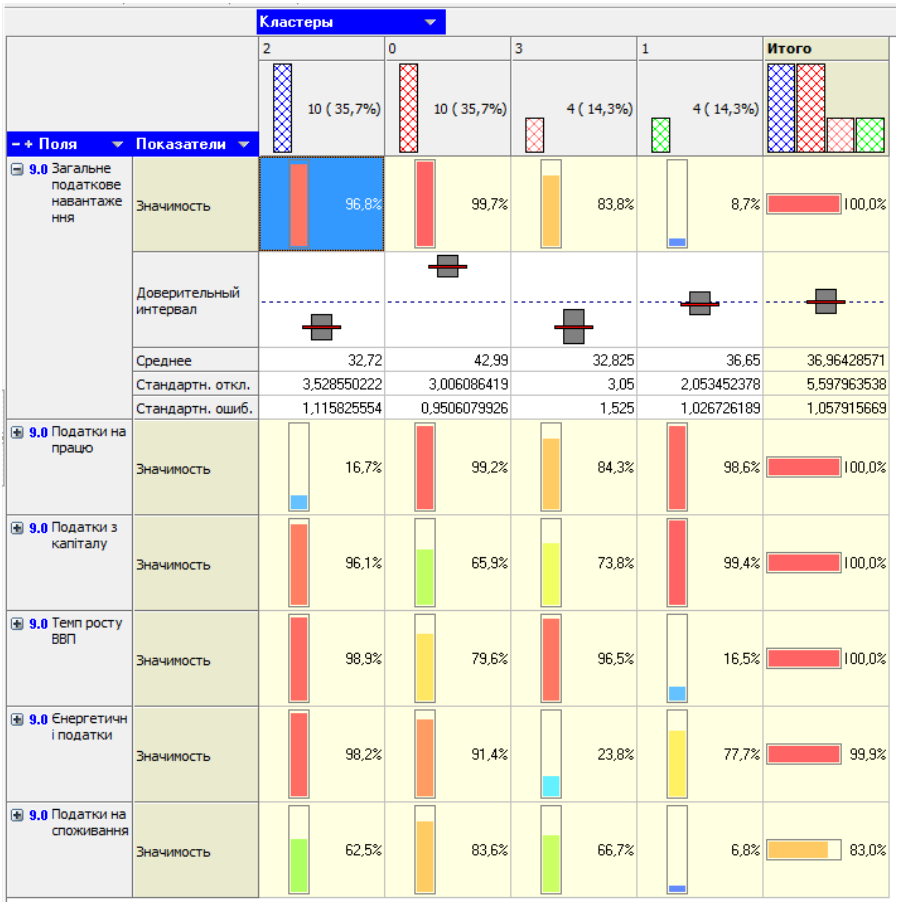


Рисунок 4.10 – Профілі кластерів

Значимість для безперервних і дискретних полів визначається по-різному. Значимість для безперервних полів установлюється залежно від відхилення середнього значення розглянутої групи кластерів від загального середнього всієї вибірки, чим більше виражене дане відхилення, тим більше його значимість. Значимість для дискретних полів визначається наявністю індивідуальних відмінностей, між розглянутими групами, чим більше виражені відмінності, тим більше значимість. Для кожної розглянутої властивості в кластері обчислюється: довірчий інтервал, середнє, стандартне відхилення й стандартна помилка (рис. 4.11).

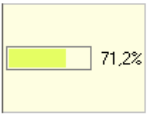
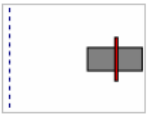
Показатель	Пример	Описание
Значимость		1 минус вероятность нулевой гипотезы. Значимость выражается в процентах. Для непрерывных полей используется *t*-критерий Стьюдента, а для дискретных полей – критерий хи-квадрат. Общая значимость поля определяется по F-критерию Фишера.
Доверительный интервал		Графическое изображение 95% доверительного интервала для среднего значения кластера (темно-серая область). Кроме этого, показываются: ▪ среднее значение по кластеру – красной линией; ▪ среднее значение по всей выборке – синей штрихпунктирной линией.
Среднее	–	Среднее значение по полю, рассчитанное для объектов, попавших в кластер.
Стандартное отклонение	–	Стандартное отклонение по полю, рассчитанное для объектов, попавших в кластер.
Стандартная ошибка	–	Стандартная ошибка по полю, рассчитанная для объектов, попавших в кластер.

Рисунок 4.11 – Визначення показників кластерів

Завершальним етапом кластерного аналізу є інтерпретація отриманих результатів. Результати аналізу свідчать, що навіть країни, які близько знаходяться на географічній карті, настільки сильно відрізняються за критеріями податкової політики, що не можуть бути віднесені до одного регіонального кластера. Усім кластерам треба надати умовні назви, які будуть відображати їх характеристики на основі 6 основних показників.

Таблиця 4.1 – Середні показники по кожному кластеру

Номер кластера	Загальне пс	Податки на сп	Податки з капіт	Енергетичні под	Податки на пра	Темп росту ВВП
0	42,99	23,90	29,27	164,94	40,49	100,52
1	36,65	21,33	41,00	165,13	25,58	101,32
2	32,72	20,14	18,77	83,14	34,01	103,99
3	32,83	19,25	19,70	120,65	29,60	98,63
Итого:	36,96	21,53	25,83	129,43	34,49	101,60

Таблиця 4.2 – Країни, що віднесені до кластеру 0

	Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶	AT	42,8	22,1	27,3	150,2	41,3	102,2	0	0,20
	FI	43,1	26	28,1	114,5	41,3	100,9	0	0,25
	BE	44,3	21,2	32,7	97,1	42,6	101	0	0,35
	SE	47,1	28,4	27,9	190,1	42,1	99,6	0	0,36
	DE	39,3	19,8	23,1	193,8	39,2	101	0	0,37
	FR	42,8	19,1	38,8	160,7	41,4	100,2	0	0,38
	HU	40,4	26,9	19,2	98	42,4	100,6	0	0,46
	IT	42,8	16,4	35,3	187,4	42,8	98,7	0	0,49
	NL	39,1	26,7	17,2	189,8	35,4	101,9	0	0,51
	DK	48,2	32,4	43,1	267,8	36,4	99,1	0	0,82

Таблиця 4.3 – Країни, що віднесені до кластеру 1

	Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶	UK	37,3	17,6	45,9	180,2	26,1	99,9	1	0,29
	MT	34,5	20	42,1	197	20,2	101,7	1	0,30
	CY	39,2	20,6	36,4	110	24,5	103,6	1	0,36
	LU	35,6	27,1	39,6	173,3	31,5	100,1	1	0,43

Таблиця 4.4 – Країни, що віднесені до кластеру 2

	Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶	PL	34,3	21	22,5	108	32,8	105	2	0,20
	LT	30,3	17,5	12,4	78,5	33	102,8	2	0,29
	SK	29,1	18,4	16,7	84,6	33,5	106,2	2	0,29
	SI	37,3	23,9	21,6	121,7	35,7	103,5	2	0,36
	EL	32,6	15,1	15,8	102	37	102	2	0,37
	CZ	36,1	21,1	21,5	127,1	39,5	102,5	2	0,38
	LV	28,9	17,5	16,3	48,4	28,2	102,3	2	0,42
	UA	37,3	22,8	24,4	63,2	28,3	102,3	2	0,44
	RO	28	17,7	19,6	26,2	29,5	107,3	2	0,51
	BG	33,3	26,4	16,9	71,7	42,6	106	2	0,55

Таблиця 4.5 – Країни, що віднесені до кластеру 3

	Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶	PT	36,7	19,1	19,6	143,4	29,6	100,2	3	0,26
	IE	29,3	22,9	15,7	153,1	24,6	97	3	0,42
	EE	32,2	20,9	10,7	71,5	33,7	96,4	3	0,44
	ES	33,1	14,1	32,8	114,6	30,5	100,9	3	0,51

Результати по сформованих кластерах найбільш зручно розглядаються за допомогою візуалізатора "Куб", у який вбудована

крос-діаграма, що зображує отримані кластери в графічному виді, що суттєво спрощує аналіз (рис. 4.12).

При побудові крос-діаграми на панелі інструментів вікна крос-діаграми натисніть кнопки «Нормалізація, приведення графіків до єдиного масштабу».

Додайте в крос-діаграму всі параметри, по яких проводилася кластеризація, і легенду, яка вкаже яким кольором який параметр відображається.

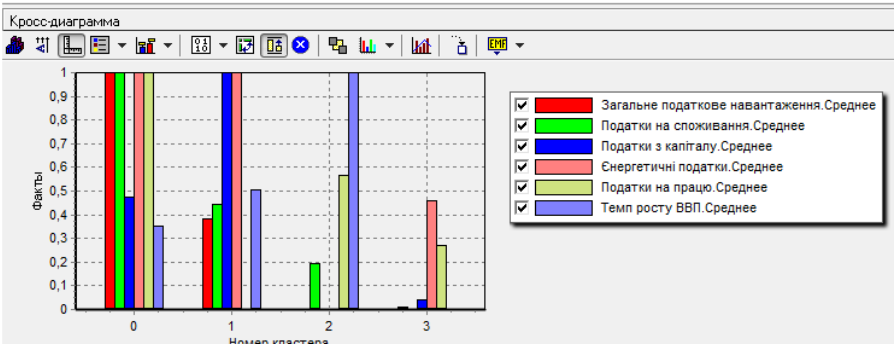


Рисунок 4.12 – Крос-діаграма – графіки приведені до єдиного масштабу

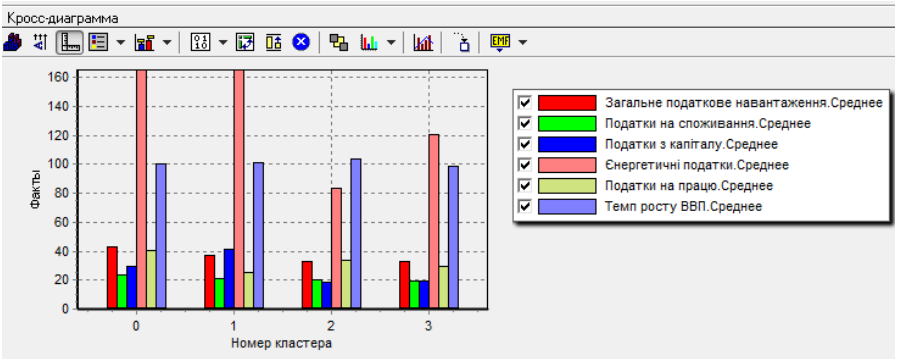


Рисунок 4.13 – Крос-діаграма – графіки не приведені до єдиного масштабу

Тепер починається головна робота менеджера-аналітика. Треба охарактеризувати кожен кластер та надати йому змістовне ім’я.

Кластер 0 - "Політика високого оподатковування" утворюють найбільша кількість країн, і повністю представлений "старими" країнами-членами ЄС (Австрія, Бельгія, Німеччина, Іспанія, Фінляндія, Франція, Італія, Нідерланди, Португалія й Швеція). Рівень оподатковування в середньому по кластеру є найбільшим серед європейських країн. Забезпечуючи середній рівень оподатковування споживання, більшість країн даної групи мають досить високі бюджетні надходження екологічних податків і податків з капіталу, а відповідно високі ставки цих податків. До того ж, середня ефективність оподатковування праці в кластері є найбільшою серед країн-членів ЄС, що обумовлене прогресивністю шкали оподатковування трудових доходів і високими соціальними гарантіями громадянам у європейських країнах даного кластера. Гнучкість податкової політики країн, які поєднані у кластер, забезпечується стабільними показниками соціально-економічного розвитку.

Кластер 1 - "Політика помірного оподатковування – стримуюча" поєднує 4 найбільш динамічні як в оподатковуванні, так і у власному економічному розвитку європейські країни (Кіпр, Люксембург, Мальта й Великобританія). Незважаючи на середній рівень оподатковування, більшість країн даної групи обмежують ефективність накопичення капіталу, обкладаючи його досить високими податками з капіталу. Вилучення екологічних податків і податків на споживання товарів і послуг у більшості країн цього кластера є найбільшими в Європі, тоді як ефективність оподатковування праці – найменшою. Останнє може бути пояснене відносно низькими ставками індивідуальних прибуткових податків із громадян і внесків на соціальне страхування, які компенсуються високими ставками корпоративних податків на прибуток і податків на споживання. У цілому ж, податкові реформи більшості країн даного кластера, а особливо Кіпру й Мальти, спрямовані на задоволення вимог, які пред'являються ЄС до країн-членів даного об'єднання, а також на збереження привабливості податкового режиму щодо міжнародного бізнесу і є логічними кроками на шляху розвитку країн як надійних європейських фінансових центрів.

У **Кластер 2 -** "Політика низького оподатковування" увійшли 10 країн (Україна, Болгарія, Естонія, Литва, Латвія, Румунія й

Словаччина). Загальне податкове навантаження в більшості країн є значно меншим в порівнянні із середньоєвропейським. Незважаючи на те, що Україна й Болгарія мають вищий рівень оподатковування своїх економік у межах даного кластера, ефективність вилучення податків з капіталу, використання інструментів екологічного й споживчого оподатковування в порівнянні із середнім значенням по ЄС є низьким, що характерно для всіх європейських країн даного кластера. У той же час, рівень вилучення податків на працю відповідає майже середньому значенню по ЄС, при цьому більшість податків платяться роботодавцями. У цілому ж, економічні системи даного кластера схожі за рівнем соціально-економічного розвитку, динамічно ростуть по основних макроекономічних показниках і активно здійснюють реформування власних національних податкових систем, вектори якого спрямовані на підвищення економічного росту в країні.

Кластер 3. Чехія, Греція, Угорщина, Польща й Словенія належать до кластера 3 "Політика помірного оподатковування – стимулююча". Податковий тиск на економіки цих країн відповідає середньому рівню по ЄС, але в порівнянні з іншими "новими" країнами-членами ЄС, держави даного кластера мають високу ефективність оподатковування споживання, капіталу й праці з мінімальною кількістю цих податків. Однак податки з капіталу й екологічні податки не обмежують активність економічних суб'єктів, а стимулюють їх до ефективного використання капіталу й енергозберігаючих технологій за рахунок низьких ставок корпоративних податків на прибуток і податків на енергетичну продукцію. Особливістю даного кластера є те, що країни даного об'єднання мають досить високий рівень оподатковування праці з відносно низькими ставками індивідуальних податків на трудові доходи й соціальних внесків з роботодавців і працівників у порівнянні із середньоєвропейськими показниками. У цілому ж, економічні системи країн близькі по своєму розвиткові по більшості соціально-економічним показникам і мають більш інтенсивну економічну політику.

Висновки. За допомогою кластерного аналізу європейських країн і України за ступенем подібності показників, які характеризують критерії ефективності податкової політики, виділено чотири регіональні кластери. Отримані результати можуть бути використані при розробці основних напрямків реформування податкової політики

й підвищення рівня соціально-економічного розвитку України, враховуючи регіональні особливості розвитку країн-членів ЄС, а разом з комплексом ефективних законодавчих рішень здатні обґрунтувати адекватність тієї або іншої стратегічної альтернативи в контексті інтеграції України до європейського економічного простору.

4.6 Варіанти завдань для самостійної роботи

Завдання 4.1. На підприємстві «Прогрес» функціонують 16 науково-виробничих відділів, зайнятих випуском різної продукції, робіт і послуг. Оскільки види діяльності, кількість працюючих, рентабельність відділів, суттєво різняться між собою, було вирішено згрупувати відділи в кілька однорідних груп, а потім для кожної групи розробити свою систему преміювання. Після ретельного аналізу вибрали чотири ознаки, за допомогою яких описувалися найбільш важливі параметри кожного відділу: 1) вартість активної частини основних виробничих фондів, тис. руб. (X1); 2) середньомісячний обсяг робіт відділу, тис. руб. (X2); 3) питома вага робіт/послуг відділу по внутрішньо фірмовій кооперації, % (x3); 4) середньомісячний прибуток відділу, тис. руб. (x4). Дані по відділах наведені в табл.4.6.

Проведіть кластеризацію відділів у пакеті Deductor, використовуючи метод k-середніх (число кластерів задайте рівним 4). Знайдіть статистичні характеристики кожного кластера.

Таблиця 4.6

№ відділу	x1	x2	x3	x4	№ відділу	x1	x2	x3	x4
1	699	190	53	11	9	293	391	16	66
2	532	211	19	42	10	300	396	29	87
3	650	152	46	14	11	73	160	0	22
4	768	216	67	17	12	862	199	51	22
5	67	106	0	32	13	112	136	0	29
6	322	397	26	52	14	289	388	31	74
7	736	180	49	18	15	512	195	6	58
8	501	239	11	60	16	490	201	9	65

Завдання 4.2. Проведіть кластеризацію споживачів по їхньому відношенню до відвідування магазинів для покупки товарів на основі результатів дослідження, суть якого в тому, що споживачів попросили виразити їхній ступінь згоди з наступними твердженнями по 7-бальній шкалі (1 - не згодний, 7 - згодний): V1 - «Відвідування магазинів для покупки товарів - приємний процес»; V2 - «Відвідування магазинів для покупки товарів погано позначається на бюджеті»; V3 - «Я поєдную відвідування магазинів для покупки товарів з харчуванням поза помешканням»; V4 - «Я намагаюся зробити кращі покупки при відвідуванні магазинів»; V5 - «Мені не подобається відвідування магазинів для покупки товарів»; V6 - «Я можу заощадити багато грошей, порівнюючи ціни в різних магазинах». Результати цього дослідження наведені в табл. 4.7.

Таблиця 4.7

Споживач	V1	V2	V3	V4	V5	V6
1	6	4	7	3	2	3
2	2	3	1	4	5	4
3	7	2	6	4	1	3
4	4	6	4	5	3	6
5	1	3	2	2	6	4
6	6	4	6	3	3	4
7	5	3	6	3	3	4
8	7	3	7	4	1	4
9	2	4	3	3	6	3
10	3	5	3	6	4	6

Завдання 4.3. З метою адресної підтримки малого бізнесу Департаментом економічного розвитку міста А було вирішено побудувати комп'ютерну, що розпізнає систему на основі методів багатомірної класифікації, що дозволяє по певному перелікові показників ідентифікувати малі підприємства для визначення проведеної щодо них економічної політики. Дані для розв'язку поставленої задачі представлені в табл. 4.8. Розділите всю вибіркову сукупність підприємств на окремі групи й по середніх характеристиках груп, що вийшли, визначите, у які із класів увійшли підприємства, що потребують фінансової підтримки, які нормально функціонують, а які вже, можливо, стали банкрутами.

Таблиця 4.8

Підприємство	Коефіцієнт поточної ліквідності, X1	Коефіцієнт забезпеченості власними засобами, X2	Коефіцієнт втрати (відновлення) платоспроможності, X3
1	1,30	0,23	1,13
2	0,73	-1,36	0,59
3	2,02	0,24	1,46
4	0,64	-1,09	0,72
5	1,28	0,23	1,19
6	1,52	0,51	1,42
7	2,00	0,50	1,69
8	0,32	0,16	0,37
9	1,18	0,15	1,04
10	0,92	-1,10	0,51

Завдання 4.4. Проведіть класифікацію комерційних банків методом k-середніх на предмет оцінки їх надійності, установивши експертним шляхом оптимальне число кластерів. Визначите склад кожного кластера, його статистичні характеристики. Основні показники роботи банків наведені в табл. 4.9.

Таблиця 4.9

Банк	Чисті активи, тис. руб.	Ліквідні активи, тис. руб.	Сумарні зобов'язання, тис. руб.
1	728481,825	12731,458	1527149,283
2	43831,446	-24198,034	79374,219
3	19973,371	629,285	27452,437
4	26484,649	-16262,703	31193,252
5	20393,837	3483,837	29484,226
6	174967,000	6783,932	260847,887
7	137371,384	3197,923	12736,830
8	62763,913	6158,736	97264,837
9	183,837	-189,780	18373,803
10	11836,910	-414,712	19724,460

Завдання 4.5. У комерційний банк ВАТ «Друг» звернулися керівники 12 великих підприємств міста А з проханням про надання кредиту. Фахівці кредитного відділу банку з метою ухвалення надійного рішення (тобто видачі кредиту, що гарантує повернення) по задоволенню цих прохань вирішили в першу чергу спробувати розділити підприємства на групи відповідно до їхнього фінансового стану. У якості факторів, що визначають фінансовий стан підприємств, були обрані необоротні активи (X1), оборотні активи (X2), власний капітал (X3), довгострокові зобов'язання (X4), короткострокові зобов'язання (X5), виторг від реалізації (x6), собівартість (X7), чистий прибуток (X8). Значення цих показників наведені в табл. 4.10. Здійсните кластеризацію підприємств і зробіть висновки про доцільність надання кредиту тієї або іншій групі, що утворилась.

Таблиця 4.10 - Показники, що характеризують діяльність підприємств, що звернулися в банк за кредитом

x1	x2	x3	x4	x5	x6	x7	x8
5116652	1655737	4912417	619623	1240349	6391468	5820259	532581
1226241	1224983	1457028	93921	900275	5027062	3462529	499271
5851307	1460596	421161	395121	1295621	4489673	2291589	67368
86188	840198	93900	604792	227694	141282	122932	10
213652	289893	187876	138430	177239	474607	439172	8238
292249	410349	44432	14565	643601	684336	636529	-36067
107355	265899	132056	7656	233542	293423	302575	110
155221	797983	74255	860	878949	244337	249286	-133140
2852	69444	-27284	913	98667	173460	126278	-27697
292001	130363	129216	155051	138097	357466	312348	-5967
659633	1295344	132248	1650653	1172076	1671660	1626270	122137
170298	666081	616076	582	219721	1002735	807602	117997

Завдання 4.6. Керівництво філії регіональної телекомунікаційної компанії, що надає послуги мобільного зв'язку, поставило задачу сегментації абонентської бази. Її цілями є:

- побудова профілів абонентів шляхом виявлення їх схожої поведінки в плані частоти, тривалості й часу дзвінків, а також щомісячних витрат;

- оцінка найбільше й найменш дохідних сегментів.

Ця інформація може надалі використовуватися для:

- Розробки маркетингових акцій, спрямованих на певні групи абонентів;
- Розробки нових тарифних планів.
- Оптимізації витрат на адресне Sms-розсилання про нові послуги й тарифи;
- Запобігання відтоку клієнтів в інші компанії.

Дані за останні кілька місяців, що взяті з білінгової системи, перебувають у файлі mobile.txt..

Проведіть сегментацію клієнтів телекомунікаційної компанії, розбивши множину записів на 6 кластерів. Проведіть аналіз отриманого розбиття й дайте змістовну характеристику кожного кластера.

Завдання 4.7. Використання кластерного аналізу для соціально-економічної оцінки розвитку районів Забайкальського краю.

Райони Забайкальського краю сильно відрізняються по площі й кількості населення, що проживає в них. Самим маленьким по території є Калганський район, а найбільшим – Каларський. Співвідношення їх площ 1:19. Найменша чисельність населення в Тунгіро-Олекминському районі, а найбільш заселений – Читинський (з м.Чита). Співвідношення становить 1:203. Диференціація муніципальних районів краю по величині грошових доходів на душу населення становить 6,4 рази. Диференціація по величині середньомісячної заробітної плати на одного працівника становить 4,8 рази. Приблизно така ж картина спостерігається й по більшості інших показників, що характеризують рівень соціально-економічного розвитку районів краю. Таким чином, неоднорідність обумовлює можливість виділення серед усієї сукупності муніципальних районів краю певних груп (кластерів).

Для розробки комплексних програм соціально-економічного розвитку муніципальних районів була розроблена номенклатура показників по кожному муніципальному району для оцінки соціально-економічного розвитку краю. Представлена номенклатура показників у таблиці 1 містить у собі 9 узагальнених показників по різних

напрямках, що характеризують рівень соціально-економічного розвитку муніципального району.

Таблиця 4.11 - Номенклатура показників

	Узагальнений показник $К_i$	**Частковий показник (субіндекс)**
1.	Рівень матеріального забезпечення	Коефіцієнт матеріального добробуту
		Витрати населення
		Рівень пенсійного забезпечення
2.	Рівень житлово-комунального й культурного забезпечення	Забезпеченість населення житлом
		Забезпеченість населення упорядженим житлом
		Забезпеченість населення об'єктами культури й відпочинку
3.	Рівень охорони здоров'я	Народжуваність
		Рівень дитячої смертності
		Забезпеченість населення лікарями
		Забезпеченість населення середнім медичним персоналом
		Забезпеченість лікарняними ліжками
4	Рівень екологічної безпеки	Рівень ПДК шкідливих речовин у повітрі, ґрунті, водоймах
5	Будівництво	Обсяг робіт, виконаних по договорах будівельного підряду (у цінах, що фактично діяли, тисяч рублів)
6	Товарні ринки й надання послуг населенню	Оборот роздрібної торгівлі (у цінах, що фактично діяли), тисяч рублів
		Об'єм платних послуг населенню (у цінах, що фактично діяли; тисяч рублів)
7	Промисловість	Основні показники розвитку промисловості в розрізі галузей (по повному колу підприємств) (у т.ч. видобуток корисних копалин; обробні виробництва; виробництво й розподіл електроенергії, газу й води)

8	Сільське господарство	Продукція сільського господарства в господарствах усіх категорій (у цінах, що фактично діяли; тисяч рублів)
9	Щільність населення	-

Конкретні значення узагальнених показників надані у файлах «Забайкалье_2000.txt» і «Забайкалье_2005.txt» у нормованій формі й розраховані по даним Комітету економіки Адміністрації Читинської області за період 2000-2005 року включно. Нормування проводилось по формулах (4.3).

Провести кластеризацію районів на 3 кластера. Охарактеризувати кожний із кластерів.

Виконати цю процедуру для 2000 року, а потім для 2005 року. Подивитися як змінилися кластери за цей період і яким був розвиток районів за цей період.

Завдання 4.8. Сегментація ринку автомобілів.

Маркетингова діяльність підприємства відіграє вирішальну роль при формуванні ефективності підприємства в цілому. Саме за рахунок служби маркетингу формується складова ефективності підприємства, що відповідає за ефективність споживача. Тому однією з актуальних задач-досліджень на сучасному етапі є визначення множини параметрів продукції власного виробництва, що будуть включатись до товарного портфелю фірми, який би враховував існуючі виробничі потужності підприємства, враховував показник рівня задоволення очікувань споживачів, дав змогу зробити вибір маркетингової стратегії просування товарів на ринку в межах цільових сегментів та, визначення відповідного рівня витрат на маркетинг.

Попередній аналіз існуючої товарної політики ЗАТ «ЗАЗ» та ЗАТ «ІЗАА» на 2009 рік, що входять до корпорації «УкрАвто», вказав на достатньо високий рівень їхньої спеціалізації, зокрема, на орієнтацію випуску недорогих легкових автомобілів класу В та С. Проте, як зазначалось вище, підвищена концентрація на одному ринковому сегменті при ігноруванні інших, не менш перспективних категорій споживачів, сприяє підвищенню рівня економічного ризику такої діяльності. Тому для розробки маркетингового плану концентрації та оптимізації своїх ресурсів на ринку необхідно провести процедуру його сегментації.

Отже, метою завдання є розрахувати рівень спеціалізації на підприємствах об'єднання «УкрАвто» та провести сегментацію ринку та оцінку рівня задоволення очікувань споживачів, завдяки чому буде можливе своєчасне прийняття рішень щодо реалізації маркетингових заходів.

Аналіз технічних характеристик легкових автомобілів, присутніх на ринку України та рівня їхньої комплектації, дозволив сформувати систему показників, що відображають ступінь їхньої привабливості для споживачів.

Так, до показників сегментації ринку, які суттєво впливають на вибір споживачів, належать такі: рівень безпеки автомобілю; рівень та якість обладнання салону; технічні характеристики; економічні характеристики тощо. Експертним шляхом було визначено якісну оцінку важливості наведених показників. Так, найбільш суттєвим фактором для споживача при виборі автомобілю є економічні характеристики, тобто його ціна. Наступним визначальним фактором є технічні характеристики транспортного засобу (витрати пального, обсяг двигуна, обсяг багажного відділення тощо). Рівню безпеки та обладнанню салону споживачі віддають однакову перевагу. Отже, система переваг споживачів в якісному виді набуває вигляду:

економічні характеристики>технічні характеристики>безпека ~ обладнання салону.

Рівень безпеки автомобілю визначається на основі наявності наступних характеристик: антиблокувальна система гальмування (ABS); електронна система розподілу гальмівних зусиль (EBD); система курсової стійкості (VSA); подушки безпеки; активні підголівники передніх сидінь; ремені безпеки з переднатягувачем; кріплення дитячих крісел; центральний замок; імобілайзер двигуна тощо.

На основі експертних оцінок було визначено пріоритет зазначених характеристик для забезпечення рівня безпеки автомобілю. Так, всі засоби безпеки, окрім центрального замку, були оцінені однаковою мірою важливості.

В свою чергу, рівень обладнання салону характеризується наявністю наступних систем: підсилювач керма; підсилювач гальмів; кондиціонер або клімат-контроль; круїз контроль; регулювання керма; регулювання сидіння водія; обігрів передніх сидінь; електропривод

вікон; електропривод дзеркал з підігрівом; радіо підготовка, аудіо система; коректор світла фар; протитуманні фари тощо.

При визначенні міри важливості кожної з зазначених систем експерти виходили з наступних передумов: на першому місці повинні бути засоби та системи, що створюють комфортні умови керування автомобілем, на другому – інші допоміжні функції салону.

До технічних характеристик легкового автомобілю, важливих для споживачів, віднесено: наявність механічної трансмісії або автомату; обсяг двигуна; обсяг багажного відділення; максимальна швидкість; витрати пального; ємність паливного баку.

Найважливішими серед зазначених технічних характеристик, що впливають на вмотивованість покупки, обрано обсяг двигуна, обсяг багажного відділення та витрати пального.

В якості економічної характеристики виступає ціна придбання автомобілю.

Отже, визначено систему показників, на основі якої можна проводити побудову карти конкурентних переваг, сегментацію ринку та оцінку рівня задоволення очікувань споживачів.

Для побудови карти конкурентних переваг товарів на ринку необхідно для кожної товарної одиниці визначити узагальнюючий рівень безпеки, обладнання салону, технічних та економічних характеристик з урахуванням міри важливості кожної з них. Для цього всі характеристики легкових автомобілів повинні пройти процедуру нормування. За результатами узагальнюючої оцінки рівня безпеки, обладнання салону, технічних та економічних характеристик складено таблицю, наведену в файлі «Автомобілі.txt».

Виконати кластерний аналіз ринку автомобілів, використовувати 5 кластерів. Надати вичерпні характеристики кластерів.

Варіант 4.9. Згідно з моделлю Тафлера фінансова стійкість підприємства може бути оцінена за такими показниками:

- X1 - прибуток до виплат / поточні зобов'язання;
- X2 - поточні активи / зобов'язання;
- X3 - поточні зобов'язання / загальна вартість активів;
- X4 – інтервал кредитування.

У файлі «tafler.txt» перебувають зазначені дані для 103 підприємств. Провести кластерний аналіз даних, розбивши їх

спочатку на 2 кластера, а потім на 3. Дати змістовну характеристику кожного кластера.

Варіант 4.10. Згідно з моделлю Ліса фінансова стійкість підприємства може бути оцінена за такими показниками:

- X1 - обіговий капітал / сума активів;
- X2 - прибуток від реалізації / сума активів;
- X3 - нерозподілений прибуток / сума активів;
- X4 - власний капітал / позиковий капітал.

У файлі «lisy.txt» перебувають зазначені дані для 103 підприємств. Провести кластерний аналіз даних, розбивши їх спочатку на 2 кластера, а потім на 3. Дати змістовну характеристику кожного кластера.

4.7 Контрольні питання

1. У чому полягає задача кластеризації?
2. Приведіть приклади застосування кластерного аналізу в бізнесі.
3. Яке програмне забезпечення можна використовувати для розв'язку задачі кластеризації?
4. Які алгоритми кластеризації ви знаєте?
5. Опишіть ієрархічні методи кластеризації.
6. Які бувають типи змінних?
7. Які етапи кластерного аналізу ви знаєте?
8. Для чого необхідно проводити нормування даних?
9. Що являє собою метод k-means і які в нього недоліки?
10. Що називають дендрограмой?
11. Яка задача бізнес-аналітика після проведення кластерного аналізу?
12. Яку мету переслідує бізнес-аналітик при проведенні кластерного аналізу?

5 РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Чубукова И.А. Data mining: учебное пособие – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с. – ISBN 5-9556-0064-7.
2. Ситник В.Ф. Інтелектуальний аналіз даних. К.: КНЕУ, 2007. –
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – СПб.: БХВ-Петербург, 2007. – 384 с.
4. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям: Учеб. пособие. 2-е изд., перераб. и доп. - СПб.: Питер, 2010. – 704 с.

Додаток - Список скорочень

AT Австрійська Республіка	IT Італійська Республіка
BE Королівство Бельгія	LT Литовська Республіка
BG Республіка Болгарія	LV Латвійська Республіка
CY Республіка Кіпр	MT Республіка Мальта
CZ Чеська Республіка	NL Королівство Нідерландів
DE Федеративна Республіка Німеччина	LU Велике Герцогство Люксембург
DK Королівство Данія	PL Республіка Польща
EE Естонська Республіка	PT Португальська Республіка
EL Грецька Республіка	RO Республіка Румунія
ES Королівство Іспанія	SE Королівство Швеція
FI Фінляндська Республіка	SI Республіка Словенія
FR Французька Республіка	SK Словацька Республіка
HU Угорська Республіка	UK Сполучене Королівство Великої Британії та Північної Ірландії
IE Республіка Ірландія	