## Розділ 8. Вибір і застосування методів інтелектуального аналізу даних (ІАД) для управління бізнес процесами.

### Опис методів ІАД та обґрунтування доцільності їх використання.

Кластерний аналіз (від англ. data clustering) представляє собою методологію сегментації набору об'єктів на групи (кластери), де елементи всередині групи демонструють високий ступінь схожості, а елементи різних груп суттєво відрізняються за визначеними параметрами.

Ключові переваги методології

Порівняно з іншими методами класифікації інформації, кластерний аналіз має декілька суттєвих переваг:

1. Гнучкість у виборі критеріїв – дозволяє здійснювати сегментацію як за

одиничним параметром, так і за комплексом характеристик, причому значущість кожного параметра легко регулюється введенням спеціальних вагових коефіцієнтів у розрахункові формули.

1. Універсальність щодо типу даних – не встановлює обмежень на природу аналізованих об'єктів, що дає можливість працювати з наборами даних практично будь-якого походження та структури.
2. Автономність функціонування – більшість алгоритмів кластеризації здатні самостійно встановлювати оптимальну кількість кластерів та визначати їх характеристики без додаткової участі дослідника.

**Методологічні основи**

Сутність процедури полягає в представленні класифікованих об'єктів у вигляді векторів індивідуальних ознак, організованих у форматі таблиці "об'єкт-властивість". На основі цих даних формується матриця відстаней або показників подібності, яка використовується для безпосередньої кластеризації. Таким чином задача класифікації вирішується з використанням чітко структурованого математичного апарату.

Принцип групування базується на концепції, що об'єкти з найбільш близькими значеннями характеристик формують окремий кластер – гомогенну групу елементів.

**Цільове призначення**

Залежно від специфіки конкретного прикладного завдання, кластеризація може бути спрямована на досягнення різних цілей:

Виявлення структурної організації даних шляхом ідентифікації природних груп, що спрощує подальшу обробку інформації в рамках кожного сегмента

Оптимізація обсягу аналізованої інформації через виділення типових представників кожного кластера, коли розгляд одного репрезентативного об'єкта забезпечує розуміння характеристик усіх споріднених елементів

Ідентифікація аномальних елементів, які не вписуються у жодну з визначених груп.

**Розробка систем класифікації й типології**

* Формування концептуальних схем організації об'єктів
* Генерування гіпотез на основі аналізу структури даних
* Валідація гіпотез через перевірку наявності у вихідних даних груп, виділених за певними критеріями

**Прикладне значення**

Кластеризація застосовується у випадках відсутності початкової інформації про можливу класову приналежність об'єктів досліджуваної вибірки, або коли масштаб вибірки робить неможливим її мануальний аналіз.

Якісна кластеризація забезпечує таку сегментацію, при якій внутрішньогрупова схожість об'єктів значно перевищує міжгрупову, а відмінності між кластерами виражені сильніше, ніж відмінності між об'єктами всередині одного кластера.

Додатковою перевагою кластерного аналізу є можливість редукції розмірності складних даних, що підвищує наочність їх представлення та інтерпретації.

Ключове прикладне завдання кластерного аналізу полягає у формуванні однорідних груп у багатовимірному просторі параметрів, що робить цей метод незамінним інструментом для структуризації великих масивів різнорідної інформації.

* 1. **Представлення результатів застосування обраного методу.**

Для проведення кластерного аналізу був обраний програмний продукт IBM SPSS Modeler.

IBM SPSS Modeler - програмне забезпечення для data mining, що поєднує в собі всі необхідні технічні та аналітичні інструменти для щоденної роботи з даними, розробки та впровадження ефективних прогностичних моделей.

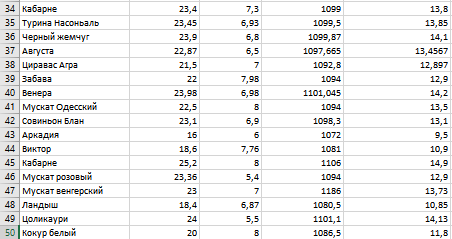
Економічний ефект від використання IBM SPSS Modeler в рішенні задач data mining забезпечується кількома складовими:

* простота доступу до даних, їх обробки, агрегування і зміни структури;
* швидку побудову та оцінка якості моделей на основі використання ефективних статистичних методів і алгоритмів машинного навчання;
* оперативне впровадження побудованих моделей в практику прийняття рішень;
* швидке отримання віддачі від інвестицій в програмне і апаратне забезпечення за рахунок високої продуктивності, інтегрованості і масштабованості;
* відповідність міжгалузевому стандарту data mining CRISP-DM.

Швидкість і ефективність процесу data mining.

Процес моделювання полягає у виявленні в даних стійких закономірностей, які можуть бути використані для прийняття рішень.

У IBM SPSS Modeler реалізована трирівнева архітектура обробки даних.

Завдання, які не потребують великих обсягів обчислень і доступу до великих масивів даних, можуть виконуватися в IBM SPSS Modeler на локальних робочих станціях.

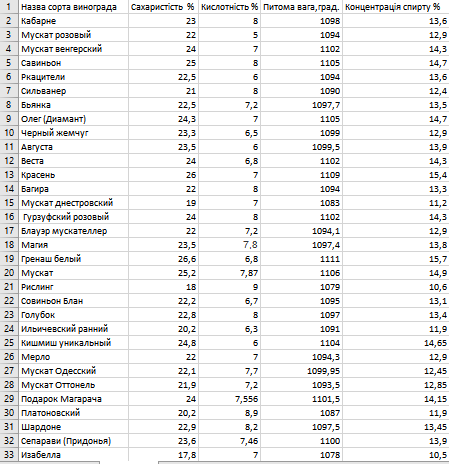


Рис.8.2.1- Представлення даних в файлі Exel.

Проведення кластерного аналізу даних

Створена модель потоку.



Рис.8.2.2 – Створена модель потоку

Налаштування джерела даних

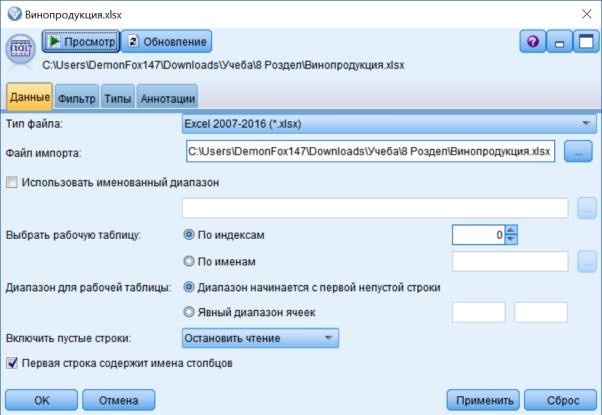
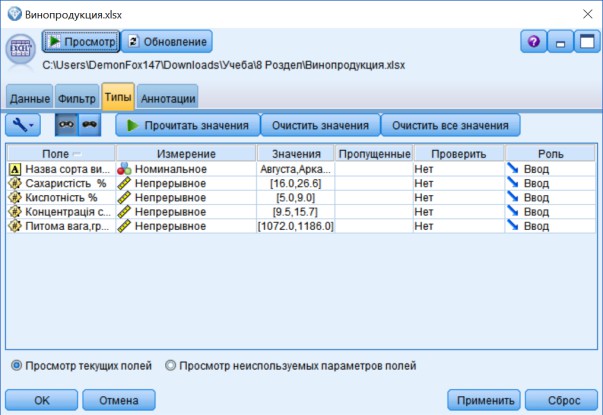
 

Рис.8.2.3 – Налаштування джерела даних Побудована діаграма з бульбашками за полями: «Сахаристість»,

«Кислотність», «Концентрація спирту».

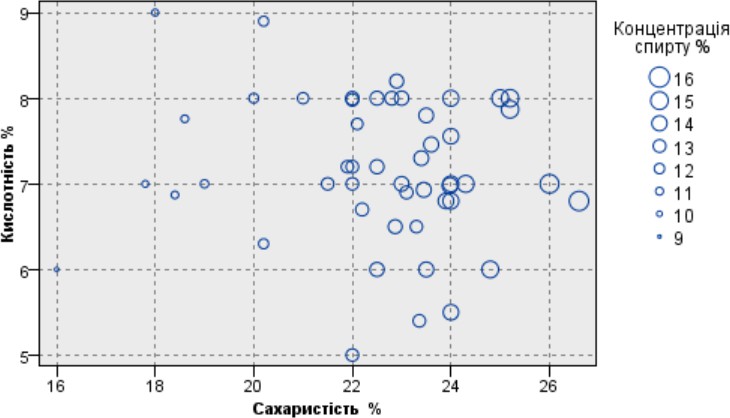


Рис.8.2.4 – Діаграма з бульбашками на основі отриманих даних Діаграма «Кислотність».

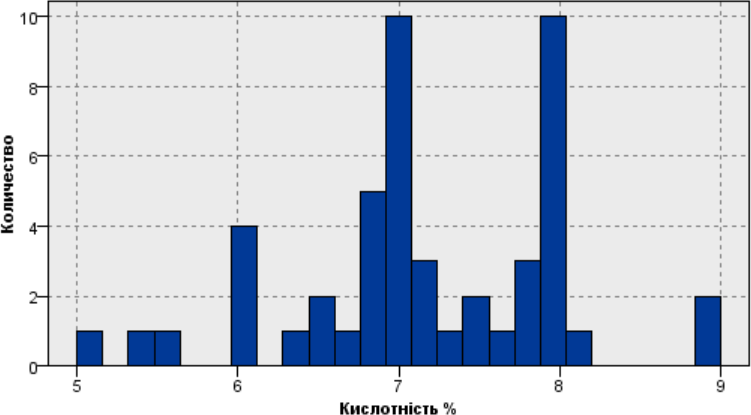


Рис.8.2.5 – Діаграма за полем «Кислотність» Мультіграфік за полями: «Сахаристість», «Кислотність»,

«Концентрація спирту».

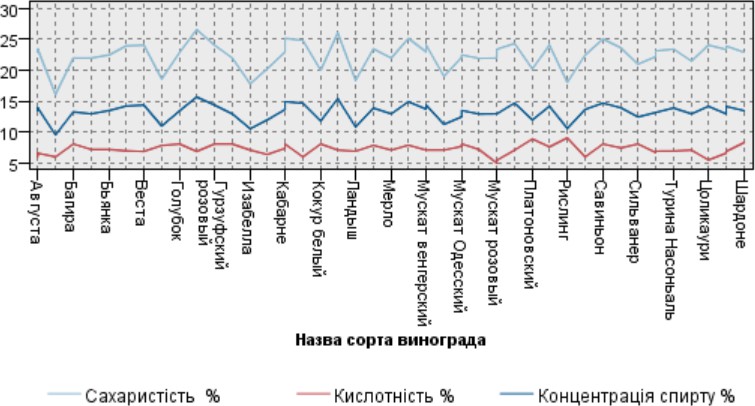


Рис.8.2.6 – Побудований мультіграфік за полями: «Сахаристість»,

«Кислотність», «Концентрація спирту».

Проведення налаштування методу К-середніх для кластерного аналізу. В ході проведення експерименту з визначення кількості кластерів, було визначено, що їх оптимальна кількість становить 5, так як при більшій їх кількості різниця між кластерами стає дуже малою.

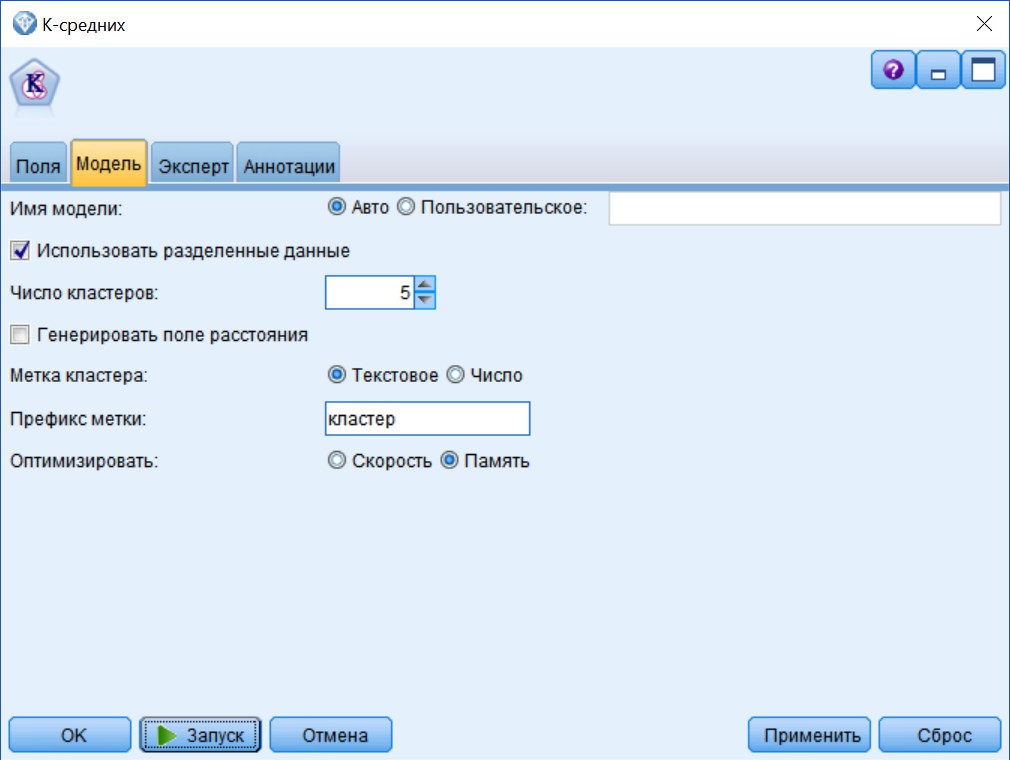


Рис.8.2.7 – Налаштування блоку К-середніх

Результат використання методу К-середніх.

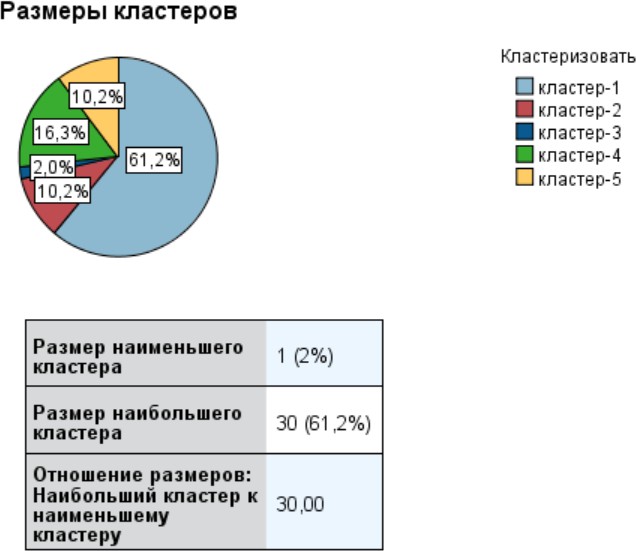




Рис.8.2.8– Результат проведення кластерного аналізу



Рис.8.2.9 – Параметри отриманих кластерів



Рис.8.2.10 – Табличні дані отриманих кластерів (результат відпрацювання моделі К-середніх)

### Розробка концепції алгоритму для автоматизації підтримки прийняття рішень.

Система підтримки прийняття рішень представляє собою інтерактивний програмний комплекс, розроблений для надання керівникам зручного доступу до аналітичних даних і моделювання ситуацій. Такі системи особливо ефективні при роботі з частково структурованими та неструктурованими задачами в різноманітних сферах діяльності.

Сучасні інформаційні платформи здатні задовольняти широкий спектр управлінських потреб. Характер інформації, яку отримує відповідальна за прийняття рішень особа, корелює з глибиною аналізу та рівнем структуризації проблемної ситуації. В ефективно функціонуючих організаціях інформаційні технології забезпечують підтримку на всіх рівнях управління – стратегічному, тактичному та оперативному.

На рівні операційного менеджменту виникають завдання різного ступеня структурованості. Інформаційні системи підприємства працюють з трьома типами проблем:

**Структуровані проблеми**

Це завдання, де ключові взаємозв'язки представлені чіткими формалізованими алгоритмами, що піддаються програмуванню. Такі проблеми охоплюють операції обліку та контролю, базовий динамічний і структурний аналіз, документообіг та його тиражування. Важливо розуміти, що структурованість проблеми визначається не її складністю, а можливістю формалізації для застосування математичних методів, включаючи дослідження операцій, що само по собі може бути технічно складним завданням.

**Слабкоструктуровані проблеми**

Характеризуються тим, що лише частина істотних залежностей піддається формалізації. Для таких задач відсутні методи прямого перетворення даних у рішення, а сама постановка завдання передбачає роботу в умовах інформаційної неповноти. В окремих випадках використання теорії нечітких множин дозволяє побудувати формальні схеми прийняття рішень. До цієї категорії відносяться проблеми розподілу інвестицій, відбору дослідницьких проектів та ін.

Слабкоструктуровані проблеми характеризуються:

* Множинністю варіантів рішення
* Орієнтацією на майбутні події та стани
* Неповною визначеністю часових і вартісних параметрів
* Потребою в комбінуванні різнорідних ресурсів
* Підвищеним ризиком при залученні значних ресурсів

Концептуальне моделювання таких завдань вимагає залучення додаткової експертної інформації від осіб, залучених до вирішення проблеми, що обумовлює суб'єктивність моделей і ускладнює застосування формальних математичних методів.

**Неструктуровані проблеми**

Для них визначені лише ключові ресурси, характеристики та параметри, однак кількісні залежності між ними невідомі. Ця категорія охоплює значну частину завдань прогнозування, стратегічного планування та організаційного управління. Вирішення таких проблем потребує застосування неформальних процедур, що ґрунтуються на інформації з високим ступенем невизначеності. Більшість неструктурованих проблем вирішується за допомогою евристичних підходів, ефективність яких безпосередньо залежить від особистісних якостей виконавця – рівня поінформованості, професійної підготовки, творчого потенціалу та інтуїції.

**Розподіл відповідальності серед працівників**

Різні категорії управлінських завдань розподіляються між працівниками відповідно до їх функцій:

Керівники зосереджуються на неструктурованих та, меншою мірою, слабкоструктурованих проблемах. Їхня діяльність максимально насичена творчими елементами при мінімізації рутинних операцій.

Фахівці переважно працюють із слабкоструктурованими завданнями. Рівень творчої складової в їхній роботі варіюється залежно від специфіки конкретних проектів. Саме спеціалісти забезпечують інформаційно-аналітичну підготовку для прийняття управлінських рішень.

Технічний персонал (коректори, касири, експедитори, друкарі, секретарі тощо) виконує регламентовані операції, що вимагають відповідного розуміння інформації та володіння виробничими навичками. Їхня діяльність пов'язана з вирішенням чітко сформульованих завдань і не потребує складного методологічного забезпечення, реалізуючись у межах стандартних інформаційних систем.