

МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК, ФІЗИКИ ТА МАТЕМАТИКИ
КАФЕДРА ІНФОРМАТИКИ, ПРОГРАМНОЇ ІНЖЕНЕРІЇ ТА
ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ

СЕГМЕНТАЦІЯ ЦІЛЮВИХ ГРУП СПОЖИВАЧІВ ДЛЯ РОЗРОБКИ
ІНВЕСТИЦІЙНИХ ПЛАНІВ З ПЕРСОНАЛІЗОВАНИМИ
РЕКОМЕНДАЦІЯМИ

Дипломна робота

на здобуття ступеня вищої освіти магістр

Виконавець: студентка 5 курсу 261М групи
спеціальності
126 «Інформаційні системи та
технології»
Килинич Дарія Олегівна

Науковий керівник: доцент, д.е.н. Кобець В.М.

Херсон – 2019

ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1	
ІНВЕСТИЦІЙНІ ПЛАНИ ТА РОЗРОБКА ПЕРСОНАЛІЗОВАНИХ РЕКОМЕНДАЦІЙ	
1.1. Інвестиційні плани.....	6
1.1.1. Критерії інвестиційної привабливості проектів.....	8
1.2. Персоналізовані рекомендації.....	10
ВИСНОВКИ ДО РОЗДІЛУ 1.....	13
РОЗДІЛ 2	
ОГЛЯД МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ В ЕКОНОМІЦІ ТА БІЗНЕСІ	
2.1. Методи класифікації.....	14
2.1.1. Кластерний аналіз.....	14
2.1.2. Опис застосовуваного алгоритму.....	20
2.1.3. Розбиття на кластери.....	21
2.1.4. Будування дендритів та визначення зв'язків	22
2.2. Дискримінантний аналіз.....	25
2.3. Метод головних компонент.....	34
ВИСНОВКИ ДО РОЗДІЛУ 2.....	36
РОЗДІЛ 3	
СЕГМЕНТАЦІЯ ЦІЛЬОВИХ ГРУП СПОЖИВАЧІВ ДЛЯ РОЗРОБКИ ІНВЕСТИЦІЙНИХ ПЛАНІВ ПЕРСОНАЛІЗОВАНИМИ РЕКОМЕНДАЦІЯМИ	
3.1. Формування моделей оптимального інвестиційного портфеля....	43
3.2. Модель оптимального портфеля Марковіца.....	46
3.3. Оптимальний портфель Марковіца максимальної ефективності...	51
3.4. Ризиковано-ефективна модель.....	53

3.5. Оптимальний портфель за моделлю Тобіна.....	54
ВИСНОВКИ ДО РОЗДІЛУ 3.....	56
ВИСНОВКИ.....	58
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60
Додаток А.....	66

ВСТУП

Одним із найбільш важливих маркетингових рішень підприємства є визначення і вибір сегментів ринку з цільовими групами споживачів, для яких підприємство планує працювати виробляти товар або надавати послуги. Обираючи сегменти ринку, підприємство має визначити профіль своїх потенційних споживачів і диференціювати їх на сегменти для здійснення ціноутворення.

Основна мета сегментування - забезпечити адресність та реалізацію товарів, що виготовляються. В результаті сегментування реалізується основний принцип клієнтоцентрованості - орієнтація на споживача.

Новизна дослідження полягає в застосуванні методів кластеризації і методу розробки диверсифікованого інвестиційного портфелю засобами RStudio та інших платформ до сегментації цільових груп споживачів для визначення найбільш прийнятних для них інвестиційних планів.

Об'єкт дослідження: ринки продукції для різних категорій споживачів.

Предмет дослідження: розробка інвестиційних планів для цільових категорій споживачів з персоналізованими рекомендаціями ІТ засобами.

Метою дослідження є проведення сегментування цільових груп споживачів для розробки інвестиційних планів із персоналізованими рекомендаціями.

Для досягнення поставленої мети визначені наступні завдання:

1. Виявити ознаки, що впливають на уподобання споживачів щодо інвестиційних планів.
2. Провести сегментацію споживачів на ринку інвестиційних послуг ІТ засобами.
3. Розробити автоматизовані персональні рекомендації для різних сегментів споживачів щодо інвестування.

Методи і технології дослідження: використання кластерного аналізу для сегментації споживачів; розробка типових інвестиційних планів за допомогою обробки великих масивів даних засобами RStudio; створення персоналізованих інвестиційних планів з використанням аналізу даних засобами пакету прикладних програм MS Excel.

Апробація дослідження проведена на міжнародній конференції ICTERI 2019, яка відбулася 12-15 червня в м.Херсон, Workshop: Information Technologies in Economic Research. За результатами конференції опублікована стаття «Support of Investors' Decision Making in Economic Experiments Using Software Tools» в електронному виданні CEUR-WS, що індексується в наукометричній базі Scopus. Тези роботи були опубліковані у видавництві Львівського національного університету імені Івана Франка, за результатами участі в конкурсі наукових робіт з економічної кібернетики, що відбувся 21-22 березня 2019.

Структура дослідження: вступ, три розділи, висновки, список використаних джерел. У розділі 1. Розділ 2. У розділі 3 (програмне забезпечення RStudio).

РОЗДІЛ 1

ІНВЕСТИЦІЙНІ ПЛАНИ ТА РОЗРОБКА ПЕРСОНАЛІЗОВАНИХ РЕКОМЕНДАЦІЙ

1.1. Інвестиційні плани

Інвестиційний проєкт (план) — це поєднані наміри та дії в одне ціле задля здійснення інвестицій та забезпечення певних фінансових, соціальних, виробничих чи економічних заходів з подальшим отриманням прибутку. Замовником інвестиційного проєкту являється суб'єкт юридична чи фізична особа, котра має фінансові заощадження (кошти) кошти й за вимогою підрядника надає підтвердження власної платоспроможності. Замовником може являтися приватна особа, міністерство, асоціація, відомство, тощо[1].

Інвестиційні плани поділяються на:

1. фінансові
2. комерційні
3. науково-технічні
4. економічні
5. виробничі
6. соціально-економічні

У контексті фінансового посередництва важливо поділяти фінансові та реальні інвестиції, різницею між ними є об'єкт інвестування. Реальні інвестиції включають інвестиції матеріальні (зазвичай це – майнові: споруди, обладнання, будинки, земля тощо) та нематеріальні. Узагальнено можна дати наступне визначення реальним інвестиціям – це капітальні вкладення у не фінансові активи зі строком корисного використання більше року.

Фінансові інвестиції мають місце тоді, коли мова йде про вкладення фінансових ресурсів у фінансові активи. Фінансові інвестиції можуть мати або орієнтований на довгострокове володіння фінансовими активами або спекулятивний характер. Потрібно звернути увагу на те, що в теорії та

практиці як синонім терміна «реальні інвестиції» – «прямі інвестиції», а як синонім терміна «фінансові інвестиції» зазвичай використовують термін «портфельні інвестиції».

Основними об'єктами фінансових інвестицій є [2]:

- боргові зобов'язання (облігації, казначейські зобов'язання, векселі тощо);

- інвестиційні сертифікати та інші інструменти, емітовані фінансовими посередниками

- депозити в банківських установах;

- акції та інші корпоративні права, що засвідчують внески до статутного фонду підприємств ;

Під терміном інвестиційне середовище зазвичай розуміють ринкову підсистему, котра забезпечує рух особливого виду товару – інвестицій. У контексті ж діяльності фінансових посередників, які опосередковують рух саме фінансових інвестицій, інвестиційне середовище, як ринкову підсистему, можна ототожнювати з фінансовою структурою, основними елементами якої є фінансові ринки, фінансові інструменти та професійні учасники цього ринку, призначенням яких і є сприяння вільному руху фінансових ресурсів між основними агентами інвестиційного процесу – інвесторами та емітентами).

Одним з найважливіших факторів збалансування попиту та пропозиції на інвестиції є часовий чинник, тобто врахування бажаних термінів надання й залучення інвестиційних ресурсів з боку інвесторів та реципієнтів, відповідно. Зокрема, населення (приватний інвестор) значну частину коштів акумулює в короткостроковій, високоліквідній формі, оскільки воно має бути впевненим, що в будь-який момент ці кошти можуть бути конвертовані в готівку. З іншого боку, економіка зацікавлена в довгострокових вкладеннях. Відповідно, вірогідність того, що вимоги, які висуваються до термінів використання інвестицій з боку особи, зацікавленої в залученні коштів, повною мірою задовольнятимуть потенційного інвестора, є надзвичайно малою.

Часовий фактор хоч і є найважливішим, але не єдиним, що потребує збалансування між інвесторами та реципієнтами. Серед інших характеристик інвестицій, які потребують узгодження, слід виділити такі: бажані обсяги індивідуально пропонованих та затребуваних фінансових ресурсів; розмір доходу на інвестиції та притаманний їм ризик; ліквідність тощо.

Отже, інвестиційне середовище (як фінансова структура) передбачає наявність наступних головних елементів:

- фінансових інструментів або активів, основними з яких як за обсягами, так і за значенням являються цінні папери;
- суб'єктної інфраструктури, центральне місце в ній посідають фінансові інститути;
- фінансових ринків, найважливіше значення з-поміж яких для фінансових посередників мають фондові ринки або ринки цінних паперів.

1.1.1. Критерії інвестиційної привабливості проектів

В першу чергу, критерій це показник. Але якщо показник це мірило якоїсь характеристики об'єкта або процесу, то критерій це показник, на основі якого формується вибір шляху досягнення мети і рівень досягнення мети. Метою інвестицій є збільшення капіталу інвестора. Тому інвестори, перед тим як вкладати свої заощадження, розглядають платформу для інвестування саме під кутом інвестиційної привабливості.

Важливою складовою аналізу інвестиційної діяльності проекту є її інвестиційна привабливість - це характеристика певних проектів як об'єктів майбутнього інвестування з позицій обсягів прибутку, ефективності використання активів, перспективності розвитку, платоспроможності, їх ліквідності та фінансового стану[3].

Фінансово-інвестиційний аналіз є способом перетворення, накопичання та використання інформації фінансового характеру. Він має кілька цілей:

- виявлення просторово-часових змін у фінансовому стані та результатах діяльності;
- прийняття рішення щодо фінансування інвестиційного проекту і розміру фінансування малих підприємств власними коштами;
- прогноз основних тенденцій зміни фінансового стану;
- визначення фінансового стану та фінансових результатів діяльності проекту.

У практиці інвестиційного проектування найпоширеніші такі критерії ефективності інвестицій:

- термін окупності інвестицій;
- чиста приведена вартість;
- внутрішня норма прибутку;

Чиста приведена вартість (NVP) - це сумарна вартість чистих грошових потоків на сьогоднішній день, які забезпечують проект[4]. В економічній літературі існують досить багато модифікації формули обчислення чистої приведеної вартості, однак найзагальнішим видом формули для розрахунку NVP є

$$NVP = -I_0 + \sum_{t=1}^n CF_t / (1 + r)^t \quad (1.1)$$

де I_0 - початкові інвестиції; CF_t - грошовий поті t-того року від реалізації інвестицій; r - норма дисконтування грошового потоку; n - кількість років у розрахунковому періоді.

Внутрішня норма прибутку (IRR) — це відсоткова ставка котра описує рентабельність інвестиції. Термін «внутрішня» підкреслює той факт, що ця відсоткова ставка є характеристикою інвестиції і не залежить від вартості капіталу, інфляції, ринкових відсоткових ставок.

$$\sum_{t=0}^n \frac{S_t}{(1 + IRR)^t} = 0 \quad (1.2)$$

де S_t - чистий грошовий потік у період (найчастіше рік) t , тобто сума всіх доходів мінус сума всіх витрат за цей період; t - номер останнього досліджуваного періоду (горизонт інвестиції).

Термін окупності інвестицій (PP) показує, через який період часу вкладені інвестиції повернуться до інвестора.

$$PP = \sum_{t=1}^n CF_t \geq I_0 \quad (1.3)$$

де PP – термін окупності інвестицій; I_0 - величина вихідних інвестицій; CF_t - дохід від реалізації інвестицій в t -ий рік; n – кількість років життя проекту.

1.2. Персоналізовані рекомендації

За останні двадцять років було проведено багато досліджень на тему як автоматично надавати рекомендації користувачу. За цей час було запропоновано багато різних методів та значно збільшився інтерес до рекомендаційних систем. Рекомендаційні системи мають співпрацювати з користувачем, отримуючи точні дані, щоб встановити його вподобання, зменшити непотрібні дані у переліку і надати рекомендації стосовно контенту. Одним із перших кроків щодо впровадження рекомендаційних систем було створення комп'ютерного бібліотекаря «Grundy» [6]. Це була досить примітивна система, яка групувала користувачів у так звані стереотипи на основі невеликого інтерв'ю користувача та застосовуючи інформацію стосовно вподобань кожного зі стереотипів, яка в свою чергу була завантажена до системи і надавала рекомендації користувачам.

Однією з перших напівавтоматичних систем колаборативної фільтрації стала Tapestry. Вона дозволяла користувачу робити запити до інформаційного домену, та надавала рекомендації за досвідом попередніх запитів або дій користувача. Це потребувало зусиль зі сторони користувача, але завдяки цьому стало можливим використовувати досвід попередніх користувачів.

Незабаром були розроблені нові системи фільтрації, які вже автоматично визначали відповідні думки та узагальнювали їх задля надання рекомендацій. Нова система GroupLens використовувала цей метод для ідентифікації статей, які б могли бути цікаві конкретному користувачу. Користувачам було необхідно надавати рейтинги до кожної статті, далі ці рейтинги система поєднувала з рейтингами інших користувачів й надавала персоналізовані рекомендації.

Наприкінці 90-х років XX ст. почали з'являтися перші комерційні рекомендаційні системи. Найбільш популярною системою вважається Amazon – система на підставі історії покупок та переглядів товарів, що надає користувачу відповідні рекомендації. Після цього рекомендаційні системи почали інтегрувати в системи електронної комерції, адже завдяки цим системам збільшувалися обсяги продажу товарів. Рекомендаційні системи знову привернули значну увагу у 2006, коли Netflix запропонував заохочувати користувачів задля поліпшення своєї рекомендаційної системи. Метою цього конкурсу було створення алгоритму рекомендацій, який міг би перевищувати свій внутрішній алгоритм CineMatch в автономних текстах на 10%. Це викликало значне збільшення активності як в академічних колах, так і серед звичайних користувачів.

Сьогодні існує декілька варіантів вирішення питання надання рекомендацій стосовно контенту. Для прикладу розглянемо два з них – рекомендації на основі контенту та колаборативна фільтрація .

Рекомендації від подібних користувачів – система визначає, наскільки схожий користувач з іншими користувачами із бази даних.

Контенті рекомендації – система за оцінками інших користувачів передбачає, яку саме оцінку поставив би користувач, враховуючи тих користувачів або продукти, які більш схожі на даний.

На початку 90-х XX ст. задля вирішення питань із навантаженням в онлайн-просторах почали використовувати колаборативну фільтрацію.

Основним припущенням цього методу є те, що думки інших користувачів можуть бути виділені та об'єднані таким чином, щоб забезпечити обґрунтоване прогнозування переваг активного користувача. Колаборативна фільтрація надає рекомендації, засновані на моделі попередньої поведінки користувача. Ця модель може бути побудована виключно на основі поведінки певного користувача або з урахуванням поведінки схожих користувачів зі схожими уподобаннями. У таких випадках коли колаборативна фільтрація бере до уваги поведінку інших користувачів вона використовує знання про цільові групи задля надання рекомендацій на підставі схожості поведінки та вподобань користувачів.

Існує два основних метода колаборативної фільтрації:

1. Заснований на сусідстві – для користувача підбирається група користувачів з найбільш схожими до нього вподобаннями, комбінуючи оцінки підгрупи, система надає рекомендації.

2. Заснований на моделі – цей підхід надає рекомендації, визначаючи параметри статистичних моделей задля оцінювання користувачів. Моделі можуть бути побудовані за допомогою кластеризації, дискримінантного аналізу тощо.

Підхід заснований на моделі є більш комплексним та надає точніші рекомендації. Цей підхід ефективніше обробляє великі набори даних, на відміну від попереднього[7]. Недоліком цього підходу є необхідність компромісу між точністю моделі та її розміром.

ВИСНОВКИ ДО РОЗДІЛУ 1

Після опрацювання матеріалу про інвестиції та персоналізовані рекомендації можна зробити наступні висновки:

Інвестиційний проект – це сукупність намірів та певний дій інвестора, що здійснює вклади, з метою забезпечення певних намірів, а саме: виробничих, соціальних, економічних чи фінансових вкладень з метою отримання прибутку.

Використання персоналізованих рекомендацій, а особливо персоналізованих рекомендацій в інвестуванні є актуальним питанням сьогодення. Інвестор, котрий має бажання інвестувати кошти, хоче знати, куди саме йому вигідно інвестувати, спираючись саме на його інвестиційну мету, дохід, наскільки він готовий ризикнути, тощо. Завдяки персоналізованим рекомендаціям, інвестор може обрати найбільш ефективний спосіб інвестування коштів.

РОЗДІЛ 2

ОГЛЯД МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ В ЕКОНОМІЦІ ТА БІЗНЕСІ

2.1. Методи класифікації

До класифікації відносяться такі задачі, відповідно до яких необхідно розподілити певні об'єкти або явища на однорідні групи (класи) при наявності сукупності властивостей, що описують дані об'єкти (коли властивість не одна, а їх багато). Задача ділиться на дві групи[9,50]:

- Визначення приналежності об'єкта до однієї з груп, які задані навчальними вибірками (дискримінантний аналіз)
- Розбиття множини об'єктів на однорідні групи за відсутності навчальних вибірок (кластерний аналіз)

Дуже часто важливим супутнім завданням є визначення мінімальної інформативної підмножини, котра описує об'єкт змінних, достатнього для розподілу об'єктів на однорідні групи.

До задач класифікації відносяться, наприклад, встановлення діагнозу за результатами аналізу та обстеження (дискримінантний аналіз), а також виявлення груп хворих, яких можна вважати однорідними за сукупністю індивідуальних особливостей (включаючи перебіг хвороби та процес лікування), що дозволить краще підібрати курс лікування (кластерний аналіз).

В останній час для вирішення задач класифікації намагаються застосувати штучні нейронні мережі. Однак їх застосування вимагає досить об'ємних навчальних вибірок, тому розглядати їх в даному дослідженні сенсу вважаємо недоцільно.

2.1.1. Кластерний аналіз

Для розподілу вибірки на однорідні підвибірки використовується метод кластерного аналізу. Кластерний аналіз – це сукупність багатовимірних

статистичних процедур, котрі дозволяють впорядкувати об'єкти за однорідними групами[10].

За допомогою кластерного аналізу досліджувану сукупність об'єктів, представлених у вигляді матриці «об'єкт-властивість», розподіляють на невелику кількість однорідних груп (їх кількість може бути завчасно відомою, чи ні). Навчаючої вибірки в кластерному аналізі не існує.

Матриця «об'єкт – властивість » має вигляд:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{pmatrix} \quad (2.1)$$

де x_{ij} – значення j -ї властивості для об'єкта з номером i . Тобто є n об'єктів та m властивостей, що описують ці об'єкти.

Класичним прикладом такого розподілу на кластери є різний кінцевий прибуток інвестованих коштів (отримає інвестор прибуток чи навпаки, втратить свої заощадження) залежності від того, куди саме обере вкласти фінанси інвестор. Розрив простору існування фактору може виникати при певній комбінації деякої сукупності незалежних змінних. Однак побудувати якісну та змістовну модель розривної області неможливо з об'єктивних причин. Потрібно знайти фактор або їх комбінацію, відповідальних за розрив, та будувати моделі для кожної із виділених підобластей окремо. Тобто одна модель описує позитивний кінцевий результат для інвестора (отримання прибутку), інша – негативний (втрати коштів), і т.п.

На рис. 2.1 показана діаграма розсіювання[11]. Дані розподіляються на дві розділені в просторі підвибірки (кластери), не зв'язані між собою.

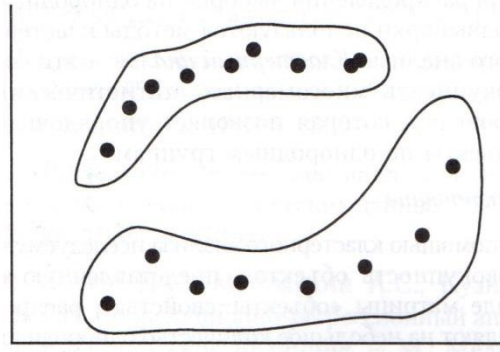


Рис.2.1. Приклад структування даних по кластерах

Загальна схема рішення задачі

Рішення задачі кластерного аналізу виконується в наступній послідовності:

1. Формування вибірки для аналізу.
2. Вибір сукупності ознак, котрі характеризують об'єкт.
3. Вибір міри подібності (відстані) між об'єктами та їх розрахунок.
4. Формування кластерів.
5. Аналіз отриманої інформації.

Найбільша кількість алгоритмів кластерного аналізу відносяться до так названих *алгомеративних процедур*, за допомогою яких спочатку в групи об'єднуються об'єкти, що знаходяться найближче, а потім до них приєднуються більш віддалені.

Можливі міри подібності (відстань між об'єктами)

Для визначення подібності між об'єктами використовують поняття відстані $d_{ij} (O_i, O_j)$ між об'єктами O_i та O_j . Чим менше відстань, тим більш схожими вважаються об'єкти. Аби бути придатним для визначення відстані, запропонована міра повинна володіти наступними властивостями:

- симетрією $d_{ij} (O_i, O_j) = d_{ji} (O_j, O_i)$;
- мінімальною відстанню об'єкта до самого себе $d_{ij} (O_i, O_j) = 0$;
- монотонною змінною d_{ij} в описуваному просторі;
- смисловою інтерпретацією міри (бажано).

Найбільш часто в кластерному аналізі використовують міри, котрі ґрунтуються на узагальненій *відстані Махаланобіса* [12], котре задається формулою:

$$d_{ij} = \sqrt{(X_i - X_j)^T \Lambda^T \Sigma^{-1} (X_i - X_j)} \quad (2.2)$$

де X – вектор спостереження, Λ - симетрична позитивно-визначена матриця вагових коефіцієнтів (зазвичай діагональна), Σ - коваріаційна матриця сукупності, з якої вилучені спостереження.

В реальності ж використовуються наступні приватні види відстаней:

Евклідова відстань. Це відстань між об'єктами, котре вираховується за формулою:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (2.3)$$

Евклідова відстань використовується при виконанні наступних умов:

- компоненти x взаємонезалежні, мають однакову дисперсію;
- компоненти однорідні в фізичному сенсі.

Якщо різні властивості, котрі характеризують об'єкт, мають різну важливість та її можливо оцінити (наприклад, за допомогою експертів), використовують *зважену евклідову відстань*:

$$d_{ij} = \sqrt{\sum_{k=1}^m \omega_k (x_{ik} - x_{jk})^2} \quad (2.4)$$

де ω_k – значимий коефіцієнт для k -ї властивості [13]. При цьому $0 \leq \omega_k \leq 1$ для всіх k .

Геммінгова відстань. Інколи її називають відстанню міських кварталів (тобто шлях від перехрестя до перехрестя не по прямій, а через вулиці). Визначається за формулою:

$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}| \quad (2.5)$$

Відстань між класами. Вказані міри визначають відстань між об'єктами. Для виконання кластерного аналізу необхідно встановити, що вважати відстанню між класами. Зазвичай використовують наступні міри визначення відстані між кластерами:

- Відстань обчислюється за принципом «найближчого сусіда», - мінімальна відстань між парою об'єктів, кожен з яких знаходиться в іншому кластері.

Обчислюється за формулою:

$$d(S_l, S_k) = \min_{X_i \in S_l, X_j \in S_k} d(X_i, X_j) \quad (2.6)$$

- Відстань обчислюється за принципом «дальнього сусіда», - максимальна відстань між парою об'єктів, кожен з яких знаходиться в іншому кластері.

Обчислюється за формулою:

$$d(S_l, S_k) = \max_{X_i \in S_l, X_j \in S_k} d(X_i, X_j) \quad (2.7)$$

- Відстань, що розраховується за «центром мас» кластерів.

$$d(S_l, S_k) = d(\bar{X}(l), \bar{X}(k)) \quad (2.8)$$

де $\bar{X}(l)$ - середнє арифметичне векторних спостережень, котрі входять до кластеру S_l . Тобто відстань між «центрами маси» відповідних кластерів.

- Відстань, що розраховується за принципом «середнього зв'язку», - арифметичне середнє всіх можливих пар комбінацій між об'єктами, котрі входять до різноманітних кластерів.

Обчислюється за формулою:

$$d(S_l, S_k) = \frac{1}{n_l n_k} \sum_{X_i \in S_l} \sum_{X_j \in S_k} d(S_i, S_j) \quad (2.9)$$

Якість розбиття на кластери.

Існує велика кількість різноманітних способів кластерного аналізу. Для порівняння якості розбиття на кластери був розроблено ряд функціоналів якості [14].

- Сума внутрішньокласових дисперсій відстані, що визначається за формулою:

$$Q = \sum_{k=1}^p \sum_{X_i \in S_k} d^2(X_i, \bar{X}_k) \quad (2.10)$$

де p – це кількість кластерів.

- Сума попарних внутрішньокласових відстаней, що визначається за формулою:

$$Q = \sum_{k=1}^p \sum_{X_i, X_j \in S_k} d^2(X_i, X_j) \quad (2.11)$$

де p – це кількість кластерів.

Різні процедури кластерного аналізу для одних і тих самих даних можуть давати різний розподіл на кластери (як за їх кількістю, так і за їх складом). Більшість методів кластерного аналізу не мають певного строгого статистичного обґрунтування.

2.1.2. Опис застосовуваного алгоритму

Розглянемо процедуру кластерного аналізу на прикладі.

Існує два різновиди дій, що можуть показувати різне розбиття на кластери після завершення роботи над об'єктами. Обирати той чи інший різновид дій потрібно з урахуванням постановки задачі. За можливості необхідно застосувати два способи та визначити, котрий з них найбільш відповідає фактично існуючим структурам даних. При *ізотонічному розбитті* група об'єктів складаються з однорідних за рівнем значень, а при *ізоморфному розбитті* в групи включаються об'єкти, близькі за своєю структурою. Це означає, що різні способи розбиття можуть давати різні групові утворення. Наприклад: є дані які характеризують розподіл прибутку фірм на розширення виробництва, наукові дослідження, соціальні виплати та інше. Тоді при ізотонічному розбитті групи будуть складатися з фірм, у який рівень прибутку фактично на рівні одна з одною, а при ізоморфному – в однорідні групи будуть включатись ті компанії, в яких структура розподілу прибутку будуть подібні між собою.

В обох випадках ознаки спочатку перетворюються таким чином, аби не було одиниць виміру та розмах шкали вимірювання був однаковим.

Ізотонічне розбиття

Для нормування шкал, необхідно виконати наступні перетворення.

Спочатку кожне значення ознаки замінюють на виражене за формулою:

$$V_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad (2.12)$$

де x_{ij} – значення j – тої ознаки для i – того об'єкта.

Після цього, для кожного об'єкта розраховують одне число за формулою [15]:

$$w_i = \sum_{j=1}^m V_{ij} \quad (2.13)$$

Відстань між двома об'єктами визначають за формулою:

$$d_{ij} = |\omega_i - \omega_j| \quad (2.14)$$

Ізоморфний розподіл

Для початку виконують нормування шкал за формулою:

$$Z_{ij} = \frac{\frac{x_{ij}}{\sum_{i=1}^n x_{ij}}}{\sum_{j=1}^m \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}} \quad (2.15)$$

В ізоморфному перетворенні відстань буде мінімальною тільки в тому випадку, коли вектори колінеарні, та максимальним – якщо вони перпендикулярні.

2.1.3. Розбиття на кластери

Примітка: Дендрит – це зв'язки між кластерами.

Після визначення відстані можливий розподіл на групи за допомогою *методу куль* [16]. Потім, побудувавши дендрити, можливо визначити форму сліду даних. Під слідом розуміється просторова форма, котру приймає сукупність експериментальних точок. В методі куль критичний радіус (відстань, що визначає, чи належить об'єкт до даного кластеру) розраховується за формулою:

$$r = \max_i \min_j d_{ij} \quad (2.16)$$

Тобто по кожному з об'єктів обирається мінімальна відстань до найближчого до нього об'єкту. Потім з цих мінімальних відстаней обирається максимальна. Тоді об'єкти, відстань між якими менша, ніж критична, належать до одного кластеру. На рис.2.2 показано розбиття на кластери за допомогою методу куль. В один кластер включаються об'єкти, відстань між якими менше критичної. Недолік методу полягає в тому, що кластери отримуються дещо штучними (форма кулі).

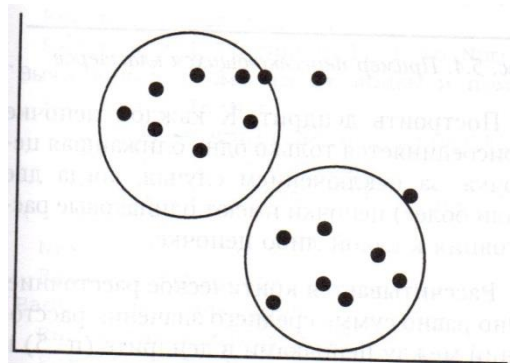


Рис. 2.2. Розбиття на кластери з використанням методу куль

2.1.4. Будівництво дендритів та визначення зв'язків

Первинне розбиття на кластери дозволяє отримати кластери круглої або еліпсоїдної форми. Оскільки при виконанні практичних задач такі кластери зустрічаються рідко (бувають випадки, коли слід має форму C-, S- образну форму), то на наступному етапі виконується побудова дендритів та визначення зв'язності в системі кластерів. [17] Це дозволяє об'єднувати початкові кластери в складніші кластери, котрі в більшій мірі відповідають їх реальному вигляду.

На цьому етапі визначають, чи є виділений кластер повністю незв'язним, чи утворює ту чи іншу пов'язану між собою структуру.

Для цього визначають відстань між кластерами, що дорівнює мінімальній відстані між двома об'єктами, що входять до цих кластерів.

$$C_{lk} = \min_{p \in g_l} \min_{q \in g_k} C_{lk}(p, q) \quad (2.17)$$

Критична відстань – це відстань, при перевищенні якої кластери вважаються незв’язними. Критичну відстань приймають рівною максимальній відстані між сусідніми елементами в одному кластері, та розраховують за формулою:

$$C_l(p) = \frac{1}{K} \sum_{l=1}^G \sum_{p=1}^{P_l} C_l(p) \quad (2.18)$$

де

$$C_{lk} = \min_{p \in g_l} C_{ll}(p, q), p = 1, 2 \dots P_l, K = \sum_{l=1}^G P_l$$

$C_{ll}(p, q)$ - відстань між елементами p та q , що належать до l -ї групи (кластеру); $C_l(p)$ – відстань від елемента p до сусіднього елемента в кластері l ; P_l – кількість елементів в l -й групі; G – кількість груп.

Після визначення відстаней кластери об’єднують в групи таким чином, аби в кінцевому результаті отримати дендрит, котрий об’єднує всю сукупність даних (кожен кластер з’єднується з найближчим до нього). Зазвичай, у більшості випадків отриманий дендрит має вигляд ланцюжка, але можливі й більш складні форми: «дерево», «розетка», «амеба», та інші[34].

Після побудови дендрита, що об’єднує всі дані, на основі аналізу критичної відстані, розривають зв’язки між тими кластерами, у яких відстань перевищує критичну. В результаті отримуємо набір дендритів, незв’язаних між собою, але кластери, котрі входять до кожного дендриту, утворюють зв’язану підвибірку. Дані, що входять до такої підвибірки, можна апроксимувати за допомогою регресійного аналізу.

Якщо кінцевий зв’язний дендрит має форму «розетка» чи «амеба» (рис.2.3), його слід розділити на простіші форми, оскільки отримати в такій області регресійну модель досить складно. Зазвичай в такій ситуації маємо

справу не з одним слідом, а з множиною декількох різних слідів. Область перетину в них спільна, а у всьому іншому – це окремі області простору (рис.2.4).

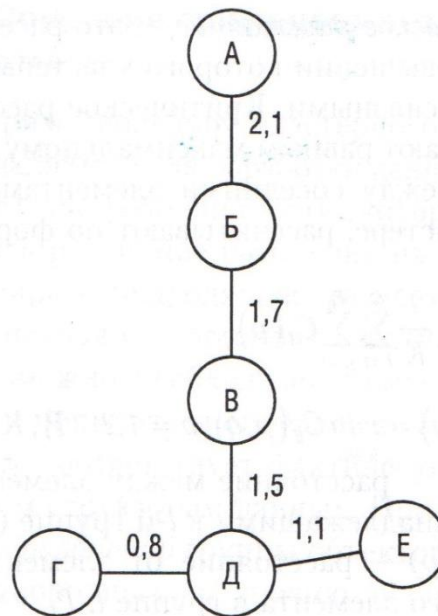


Рис.2.3. Типовий вигляд дендрита (цифри – відстань між відповідними кластерами)

Побудова дендрита виконується за наступним алгоритмом:

1. Взяти черговий об'єкт А. Якщо він входить хоч в один ланцюжок – перейти до наступного об'єкту.
2. Для чергового об'єкта знайти такий об'єкт В, відстань до якого від А мінімальна. Перевірити, чи не входить об'єкт В в одну із вже сформованих ланцюжків. Якщо так, то додати об'єкт А до цього ланцюжка. Якщо ні – утворити із А та В первинний ланцюг.
3. Якщо є ще об'єкти, котрі не включені в ланцюжки, то перейти до пункту №1 або №4.
4. Знати матрицю відстаней між ланцюжками. Відстань між ланцюжками дорівнює відстані між найближчими об'єктами в цих ланцюжках.

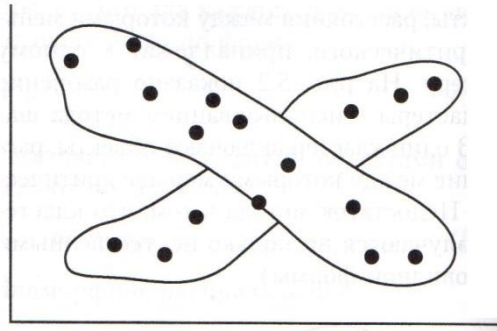


Рис. 2.4. Приклад кластерів, що перетинаються

5. Побудувати дендрит. До кожного ланцюжка приєднується лише один найближчий ланцюжок, за виключенням випадків, коли два чи більшеланцюжків мають однакові відстані до будь-якого ланцюжка.
6. Розраховується критична відстань, що дорівнює сумі середнього значення відстаней між ланцюжками в дендриті (пункт 5) та середньоквадратичного значення цих же відстаней.
7. Розриваються зв'язки, де відстань більше критичної.

В даному алгоритмі в якості об'єктів можуть бути використані як безпосередньо вихідні (початкові) об'єкти, так і кластери, котрі були виділені за допомогою методу куль. [25] При побудові алгоритму на основі початкових об'єктів враховується відстань між ними, а на основі кластерів – необхідно вирахувати відстань між кластерами.

2.2. Дискримінантний аналіз

Дискримінантний аналіз дає змогу вивчати відмінності між двома чи більше об'єктами за декількома змінними одночасно. Кожна група об'єктів називається класом. Класи розглядаються як значення деякої *класифікуючої змінної*, котра змінена за шкалою найменувань.

Якщо класифікуюча змінна залежить від дискримінантних, то в такому випадку дискримінантний аналіз є аналогом *багатофакторного регресійного аналізу*, коли відгук змінюється за шкалою найменувань.

У випадку, коли дискримінантна змінна залежить від класифікуючої, тобто належність об'єкта до певного класу викликає зміни одночасно в декількох змінних, дискримінантний аналіз є аналогом *узагальненого багатовимірного дисперсійного аналізу*. Наприклад, класифікуюча змінна – курс криптовалюти, а дискримінантні змінні, на основі котрих потрібно встановити кінцевий прибуток, характеризують інвестиції.

В дискримінантному аналізі розрізняють дві задачі: інтерпретація та класифікація. *Задача інтерпретації* – визначення кількості, значущості функцій та їх значень для відображення відмінності між класами.

Задача класифікації – визначення класу, до якого належить навий об'єкт.

В дискримінантному аналізі, на відміну від кластерного, є навчальна вибірка, в якій відомо, до якого з класів відносяться об'єкти. [20] За навчальною вибіркою необхідно побудувати правила, які в подальшому дозволять визначити, до якого класу відносяться нові об'єкти.

Вивчення відмінностей між двома та більше групами об'єктів одночасно за декількома змінними, що їх описують, дозволяє за навчальною вибіркою отримати правила (формули), за допомогою яких визначають належність об'єкта до певної групи.

Передумови визначення

1. Об'єкти (спостереження) належать до двох чи більше класів.
2. В кожному класі є принаймні два об'єкти.
3. Число дискримінантних змінних не має перевищувати кількості об'єктів.
4. Дискримінантні змінні вимірюються за шкалою інтервалів або шкалою відношень.
5. Дискримінантні змінні мають бути лінійно незалежні.
6. Дискримінантні змінні мають бути розподілені за багатовимірним нормальним законом розподілу (тобто кожна змінна має нормальний закон розподілу за інших фіксованих змінних).
7. Коваріаційні матриці класів приблизно рівні між собою.

Особливу увагу слід звернути на формування навчальних вибірок, оскільки часто ці вибірки не включають в себе змінних, за якими фактично відбувається класифікація об'єктів. Для перевірки необхідно об'єднувати навчальні вибірки в одну і намагатися розподілити їх на класи за допомогою кластерного аналізу. Якщо необхідної класифікації не вийшло, це означає, що підбір змінних виконано некоректно.

Вирішення задачі в випадку двох класів

Найбільш поширений є випадок двох класів. [43] Якщо є два класи, то об'єкти характеризують p змінних. Для першого класу сформована вибірка X обсягом n_1 .

1. Розраховується середнє значення по кожній змінній для кожної вибірки (класу):

$$\bar{x}_j = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{ij} \quad (2.19)$$

та

$$\bar{y}_j = \frac{1}{n_2} \sum_{i=1}^{n_2} y_{ij} \quad (2.20)$$

2. Визначаються оцінки коваріаційних матриць для кожного класу S_x і S_y :

$$S_{kj}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (2.21)$$

та

$$S_{kj}(y) = \frac{1}{n_2} \sum_{i=1}^{n_2} (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k) \quad (2.22)$$

3. Розраховується незміщена оцінка об'єднаної коваріаційної матриці:

$$S = \frac{1}{n_1 + n_2 - 2} (n_1 S_x + n_2 S_y) \quad (2.23)$$

4. Знаходиться зворотня матриця S^{-1} .
5. Розраховується вектор оцінок коефіцієнтів дискримінантної функції
 $A = S^{-1}(\bar{X} - \bar{Y})$
6. Визначаються оцінки векторів дискримінантних функцій вихідних змінних
 $U_x = XA$ та $U_y = YA$.
7. Обчислюється середнє значення оцінок дискримінантних функцій

$$\bar{u}_x = \frac{1}{n_1} \sum_{i=1}^{n_1} u_{xi} \quad (2.24)$$

та

$$\bar{u}_y = \frac{1}{n_2} \sum_{i=1}^{n_{12}} u_{yi} \quad (2.25)$$

8. Визначається дискримінантна константа $C = 1/2 (\bar{u}_x + \bar{u}_y)$

Важливими є пункти 4 і 5. Взагалі, ці два пункти описують рішення системи рівнянь $SA = (\bar{X} - \bar{Y})$ відносно A . Обчислення зворотної матриці потребує втричі більше розрахункових операцій, аніж рішення системи рівнянь.

Щоб визначити, до якого класу належить будь-яке спостереження Z (об'єкт), необхідно для початку розрахувати для нього оцінку дискримінантної функції:

$$u_z = \sum_{i=1}^p a_i z_i \quad (2.26)$$

Якщо це значення більше або дорівнює константі C , то новий об'єкт відноситься до класу X , якщо менше – до класу Y (при $\bar{u}_x > \bar{u}_y$).

Розширення алгоритму для числа класів більше двох

Є k класів, об'єкти характеризуються p змінними. Для кожного класу сформована вибірка $X^{(l)}$, об'ємом n_l .

1. Розраховується середнє значення по кожній змінній для кожної вибірки (класу):

$$\bar{x}_j^{(l)} = \frac{1}{n_l} \sum_{i=1}^{n_l} x_{ij}^{(l)} \quad (2.27)$$

2. Визначаються оцінки коваріаційних матриць для кожного класу S_l :

$$S_{mj}^{(l)}(x) = \frac{1}{n_l} \sum_{i=1}^{n_l} (x_{ij} - \bar{x}_j)(x_{im} - \bar{x}_m) \quad (2.28)$$

3. Розраховується незміщена оцінка об'єднаної коваріаційної матриці.

$$S = \frac{1}{\sum_{i=1}^k n_i - k - 1} \sum_{i=1}^k n_i S_i \quad (2.29)$$

4. Знаходиться зворотня матриця S^{-1} .
5. Розраховуються вектори оцінок коефіцієнтів дискримінантної функції:

$$A^{(l)} = S^{-1} \bar{X}^{(l)}$$

6. Визначаються дискримінантні константи:

$$\lambda_l = \frac{1}{2} \bar{X}^{(l)} (S^{-1} \bar{X}^{(l)})$$

7. Визначається, чи належать нові об'єкти до класу. Об'єкт, що визначається вибіркою Z , належить до того класу, для якого $Z^T A^{(1)} - \lambda_1$ максимальне.

Оскільки в алгоритмах використовується розрахунок зворотної матриці, то при деяких наборах даних результат не може бути досягнутий наслідок

через вироджені матриці. В такому випадку слід змінити матрицю вхідних даних таким чином, щоб не було залежності між стовпчиками. [30] Це можливо перевірити, завдяки побудові кореляційної матриці (коефіцієнти кореляції не повинні бути близькими до 1).

Послідовність розрахунку задач для випадку k класів

Канонічна дискримінантна функція в такому випадку має вигляд:

$$f_{ki} = u_0 + \sum_{j=1}^p u_j X_{jki} \quad (2.30)$$

де f_{ki} – значення канонічної дискримінантної функції для i -того об'єкта в k -тому класі; u_j – шукані коефіцієнти дискримінантної функції; X_{jki} – значення дискримінантної змінної X_j для i -того об'єкта в класі k . Функцію будують таким чином, аби її середнє значення для різних класів відрізнялось якнайбільше. При цьому сукупність функцій має утворювати ортогональний простір, тобто функції мають бути незалежними одна від одної. З цього випливає, що кількість функцій не може перевищувати кількість класів мінус 1 або числа дискримінантних змінних (в залежності від того, яка з цих величин менша).

Задача вирішується в декілька етапів:

1. Побудова матриці \mathbf{T} , елементи якої визначаються за формулою:

$$t_{jl} = \sum_{k=1}^K \sum_{i=1}^n (X_{jki} - \bar{X}_j)(X_{jki} - \bar{X}_l) \quad (2.31)$$

де X_{jki} – значення дискримінантної змінної X_j для i -того об'єкта в класі k , \bar{X}_j – середнє значення для змінної j по всім класам, n – загальна кількість спостережень по всім класам, K – кількість класів.

5. Отримані коефіцієнти нормують за формулою: $u_i = v_i \sqrt{n - K}$.

При цьому: $\bar{x}_i = -\sum_{j=1}^p u_j \bar{X}_j$.

Коефіцієнти u_i називаються *нестандартними*, оскільки вони залежать від одиниць виміру змінних, тому досить часто здійснюється перехід до стандартних коефіцієнтів, які відображають відносний внесок змінної, що не залежить від шкали вимірювання. Перехід до стандартних коефіцієнтів здійснюється за формулою:

$$c_i = u_i \sqrt{\frac{W_{ii}}{n - K}} \quad (2.34)$$

де n – загальна кількість спостережень, K – число класів (груп), W_{ii} – діагональний елемент матриці оцінки розсіювання, що розраховується за формулою (5.32).

Величина стандартного коефіцієнта пропорційна його внеску в дискримінантні функцію. [28] При цьому якщо змінні скорельовані, то стандартні коефіцієнти не будуть відображати їх справжнього внеску.

Кількість дискримінантних функцій

Задачу дискримінації бажано вирішувати з використанням мінімальної кількості функцій. Кількість функцій в кожному конкретному випадку залежить від конфігурації класів в багатовимірному просторі дискримінантних змінних. Чим складніша конфігурація, тим більше функцій необхідно для їх розподілу. Для визначення кількості необхідних функцій використовується перевірка їх значимості. Для характеристики того, наскільки одна функція відрізняється від іншої, використовують відносний пропорційний зміст, який визначають для однієї функції за формулою:

$$\frac{\lambda_i}{\sum \lambda_i} 100\%$$

Але цей показник не може слугувати підставою для відкидання частини функцій. Для оцінки значущості використовують Λ -статистику Уїлкса. Критеріальне значення обчислюється за формулою:

$$\Lambda_k = \prod_{i=k+1}^K \frac{1}{1 + \lambda_i} \quad (2.35)$$

де K – кількість класів, а k – число розрахованих дискримінантних функцій. Чим ближче значення Λ до 0, тим краще відмінність класів. А чим ближче до 1, тим відмінність гірше (класи співпадають).

Класифікація без інтерпретації

При використанні дискримінантних функцій, окрім задач класифікації, вирішується і задача інтерпретації.[23] В деяких випадках немає необхідності вирішувати цю задачу. В цій ситуації використовують прості класифікуючі функції, що ґрунтуються безпосередньо на дискримінантних змінних:

$$h_k = b_{k0} + \sum_{i=1}^p b_{ki} x_i \quad (2.36)$$

де h_k - значення функції для k -го класу, x_i - значення дискримінантних змінних.

Значення b_k визначають за формулами:

$$b_{ki} = (n - K) \sum_{j=1}^p a_{ij} K_{jk} \quad (2.37)$$

$$b_{k0} = -0.5 \sum_{i=1}^p b_{ki} X_{jk} \quad (2.38)$$

де a_{ij} – елемент матриці, зворотної до W , яку розраховують за формулою (2.32).

Для класифікації використовують *квадрат відстані Махаланобіса*, чином за формулою:

$$D^2(X, G_k) = (n - K) \sum_{i=1}^p \sum_{j=1}^p a_{ij} (X_i - \bar{X}_{ik})(X_j - \bar{X}_{jk}) \quad (2.39)$$

В.Плота рекомендує використовувати скореговану незміщену оцінку даної величини:

$$D'^2 = \frac{n - p - 3}{n - 2} D^2 - \left(\sum_{i=1}^K \frac{p}{n_i} \right) \quad (2.40)$$

Об'єкт X зараховують до групи, відстань до якої ($D^2(X, G_k)$) найменша.

Ймовірність приналежності до класу

У випадку, коли має місце значне перекриття класів i , відповідно, відмінність між ними слабка, потрібно, окрім відстані, розраховувати ще й імовірність приналежності до класу за формулою:

$$p = \frac{1}{\sum_{k=1}^K e^{(f_{kx} - f_{max})}} \quad (2.41)$$

де f_{kx} - значення дискримінантної функції для об'єкта X в k -му класі, f_{max} - значення дискримінантної функції для об'єкта X для класу, відстань до якого мінімальна. Якщо ймовірність мала, то класифікувати об'єкт при даному способі розподілу неможна.

2.3. Метод головних компонент

Метод головних компонент – являється чи не одним з основних способів зменшити розмір набору даних, втративши при цьому найменшу кількість інформації. Даний метод був винайдений Карлом Пірсоном у 1901 році. Іноді метод головних компонент називають перетворенням Кархунена-Лоева або перетворенням Хотеллінга.

Даний метод складається з лінійного перетворення вхідного вектору x розмірності N у вихідний вектор y розмірності $M, N > M$. При цьому

компоненти вектору y являються некорельованими i , відповідно. спільна дисперсія після перетворення лишається незмінною. Матриця X складається з усіх прикладів зображення навчального набору.

ВИСНОВКИ ДО РОЗДІЛУ 2

Розглянувши вищезазначений матеріал можна зробити наступні висновки:

Кластерний аналіз – це сукупність багатовимірних статистичних процедур, котрі дозволяють впорядкувати об’єкти за однорідними групами.

За допомогою кластерного аналізу досліджувану сукупність об’єктів, представлених у вигляді матриці «об’єкт-властивість», розподіляють на невелику кількість однорідних груп (їх кількість може бути завчасно відомою, чи ні). Навчаючої вибірки в кластерному аналізі не існує.

Класичним прикладом такого розподілу на кластери є різний кінцевий прибуток інвестованих коштів (отримає інвестор прибуток чи навпаки, втратить свої заощадження) залежності від того, куди саме обере вкладати фінанси інвестор.

Дискримінантний аналіз дозволяє вивчати відмінності між двома чи більше об’єктами за декількома змінними одночасно. Кожна група об’єктів називається класом. Класи розглядаються як значення деякої класифікуючої змінної, котра змінена за шкалою найменувань.

Метод головних компонент – найпоширеніший спосіб зменшити розмір даних, втративши при цьому найменшу кількість інформації. Метод головних компонент використовують для аналізу складного набору даних, котрим притаманна мультиколінеарність, тобто присутність прихованих зв’язків між даними.

РОЗДІЛ 3

СЕГМЕНТАЦІЯ ЦІЛЬОВИХ ГРУП СПОЖИВАЧІВ ДЛЯ РОЗРОБКИ ІНВЕСТИЦІЙНИХ ПЛАНІВ З ПЕРСОНАЛІЗОВАНИМИ РЕКОМЕНДАЦІЯМИ

Дані для розрахунків отримані з відкритих джерел. Це вибірка зі 100 клієнтів італійських банків, котрих характеризує 4 показника, а саме: відношення до ризику, інвестиційна мета, стать та вік. (Рис.3.1)

	Відношення до ризику	Інвестиційна мета	Чоловік або жінка	Кількість років
1	3	3	0	55
2	3	3	1	60
3	3	3	0	52
4	3	3	1	55
5	4	3	0	95
6	3	3	0	76
7	3	3	1	65
8	4	4	0	61
9	3	3	1	82
10	3	3	1	75
11	3	3	1	93
12	4	4	0	75
13	3	3	0	57
14	3	3	0	51
15	3	3	1	57
16	3	3	0	72
17	3	3	0	67
18	3	3	0	62
19	3	3	1	59
20	3	3	1	59

Showing 1 to 20 of 100 entries

Рис.3.1. Дані для розрахунку

Для проведення сегментації клієнтів, було застосовано кластерний аналіз в платформі MS Excel [35]. За результатами проведенного кластерного аналізу видно, що дана вибірка розподілилась на два основні кластери: 1 – жінки працездатного віку, схильні до ризику та низької дохідності; 2 – жінки-пенсіонери, схильні до ризику та середньої дохідності (Рис.3.2)

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
90	0	58	3		4,452038	1,857429	1,857429	3,010377	3,015688	0,395834	55,80954	e2		89	1,126543279	18,52196		1	0	
91	1	78	3		22,19869	24,38184	22,19869	3,005189	3,007844	0,697917	66,90477	e1		90	19,43704711	1,516722		0	1	
92	0	66	3		1,142709	10,19816	1,142709	3,002594	3,003922	0,348958	66,45238	e1		91	7,500330842	10,53327		1	0	
93	1	72	3		5,585688	5,104186	5,104186	3,001297	3,001961	0,674479	69,22619	e2		92	13,43723942	4,515005		0	1	
94	1	71	3		1,803431	4,593983	1,803431	3,000649	3,00098	0,83724	70,1131	e1		93	12,43728951	5,513566		0	1	
95	1	54	3,5		16,14488	15,26238	15,26238	3,500324	3,00049	0,91862	62,05655	e2		94	4,671785126	22,52809		1	0	
96	1	82	3		19,94989	11,88802	11,88802	3,250162	3,000245	0,95931	72,02827	e2		95	23,43697361	5,499445		0	1	
97	0	86	3		14,00685	23,96629	14,00685	3,125081	3,000123	0,479655	79,01414	e1		96	27,45395455	9,521922		0	1	
	1	74	3									e2	Жінки-пенсіонери, схильні до ризику та середньої дохідності	97	15,4371587	2,521308		0	1	
98					5,042616	1,987949	1,987949	3,062541	3,000061	0,739827	76,50707			98	17,45858957	1,152754		0	1	
99	1	76	2,5		1,152754	3,220508	1,152754	3,03127	2,500031	0,869914	76,25353	e1		99	4,565217563	22,50866		1	0	
100	1	54	3		22,25955	22,50866	22,25955	3,015635	2,750015	0,934957	65,12677	e1								
	1	52	3									e1	Жінки працездатного віку, схильні до ризику та низької дохідності	100	6,564658838	24,50853		1	0	
101					13,12932	24,25906	13,12932	3,007818	2,875008	0,967478	58,56338									
102	0,44	66,75	3,125	СРЗНАЧ														55	45	
103								Відношення до ризику	інвестиційна мета	Чоловік або жінка	Кількість років							Жінки працездатного віку, схильні до ризику та низької дохідності	Жінки-пенсіонери, схильні до ризику та середньої дохідності	
104																				
105																				
106																				
107																				

Рис.3.2. Розрахунок кластерного аналізу

Наступним етапом проведення розрахунків було застосування методу головних компонент в платформі Rstudio [42]. За допомогою якого, можливо побачити, який з критеріїв найбільше впливає на рішення інвестора при інвестуванні власних заощаджень.

Розрахувавши кореляційну матрицю (Рис3.3.), ми побачили, що найбільший зв'язок спостерігається між критерієм «відношення до ризику» та «інвестиційна мета». А саме 0,69.

```

8
9
10 # Метод главных компонент
11
12
13 h <- data # набор данных по клиентам
14 glimpse(h) # посмотрим на набор данных
15 describe(h) # описательные статистики
16
17 # корреляционная матрица
18 cor(h)
19
20 # метод главных компонент с предварительной стандартизацией переменных
21 h.pca <- prcomp(h,scale=TRUE)
22
23 # извлекли первую главную компоненту:
24 pca1 <- h.pca$x[,1]
25 head(pca1)

```

20:1 (Top Level) R Script

Console Terminal x

C:/Users/Dawa/Desktop/ХДУ/data in R/Master2019/

```

> чоловік або жінка <dbl> 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1...
$ `кількість років` <dbl> 55, 60, 52, 55, 95, 76, 65, 61, 82, 75, 93, 75, 57, 51, 57, 72, 67, 62...
> describe(h) # описательные статистики

```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Відношення до ризику	1	100	3.16	0.47	3	3.10	0.00	2	5	3	1.12	2.42	0.05
Інвестиційна мета	2	100	3.08	0.34	3	3.00	0.00	2	4	2	1.34	4.51	0.03
чоловік або жінка	3	100	0.44	0.50	0	0.42	0.00	0	1	1	0.24	-1.96	0.05
кількість років	4	100	66.80	12.46	66	66.26	13.34	39	95	56	0.30	-0.56	1.25

```

> # корреляционная матрица
> cor(h)

```

	Відношення до ризику	Інвестиційна мета	чоловік або жінка	кількість років
Відношення до ризику	1.00000000	0.68689955	-0.04524813	0.03694263
Інвестиційна мета	0.68689955	1.00000000	-0.03108092	-0.09192771
чоловік або жінка	-0.04524813	-0.03108092	1.00000000	0.06631995
кількість років	0.03694263	-0.09192771	0.06631995	1.00000000

```

> |

```

Рис.3.3. Розрахунок кореляційної матриці

Наступним етапом було «вилучення» головної компоненти . Перша головна компонента (відношення до ризику) показала, що може пояснити 42,3% характеристик інвестора (Рис.3.4.). Тобто який фактор має найбільший вплив на прийняття рішення.

```

20 # метод главных компонент с предварительной стандартизацией переменных
21 h.pca <- prcomp(h, scale=TRUE)
22
23 # извлекли первую главную компоненту:
24 pca1 <- h.pca$x[,1]
25 head(pca1)
26 tail(pca1)
27 # извлекли веса, с которыми переменные входят в первую главную компоненту:
28 v1 <- h.pca$rotation[,1]
29 v1
30
31 # выборочная дисперсия каждой компоненты:
32 summary(h.pca)
33
34
35 # выборочная дисперсия каждой компоненты на графике:
36 plot(h.pca)
37
38 # исходный набор данных в новых осях

```

```

35:1 (Top Level)

```

```

Console Terminal
C:/Users/Dawa/Desktop/ХДУ/data in R/Master2019/
[1] -0.2731870 -0.4661670 -0.2577326 -0.4404096 1.0262334 -0.3813680
> tail(pca1)
[1] -0.5794995 -0.4328828 -0.5382877 -2.6319964 -0.4352581 -0.4249552
> # извлекли веса, с которыми переменные входят в первую главную компоненту:
> v1 <- h.pca$rotation[,1]
> v1
Відношення до ризику      Інвестиційна мета      Чоловік або жінка      Кількість років
                0.70060853                0.70574113                -0.08342530                -0.06416671
> # выборочная дисперсия каждой компоненты:
> summary(h.pca)
Importance of components:
              PC1      PC2      PC3      PC4
Standard deviation  1.3015  1.0316  0.9702  0.54837
Proportion of Variance 0.4235 0.2660 0.2353 0.07518
Cumulative Proportion 0.4235 0.6895 0.9248 1.00000
> |

```

Рис.3.4. Вилучення першої головної компоненти

Для наочності, буд побудований графік вибіркової дисперсії (Рис.3.5.).
Що показав, дійсно, перша головна компонента має найбільший вплив.

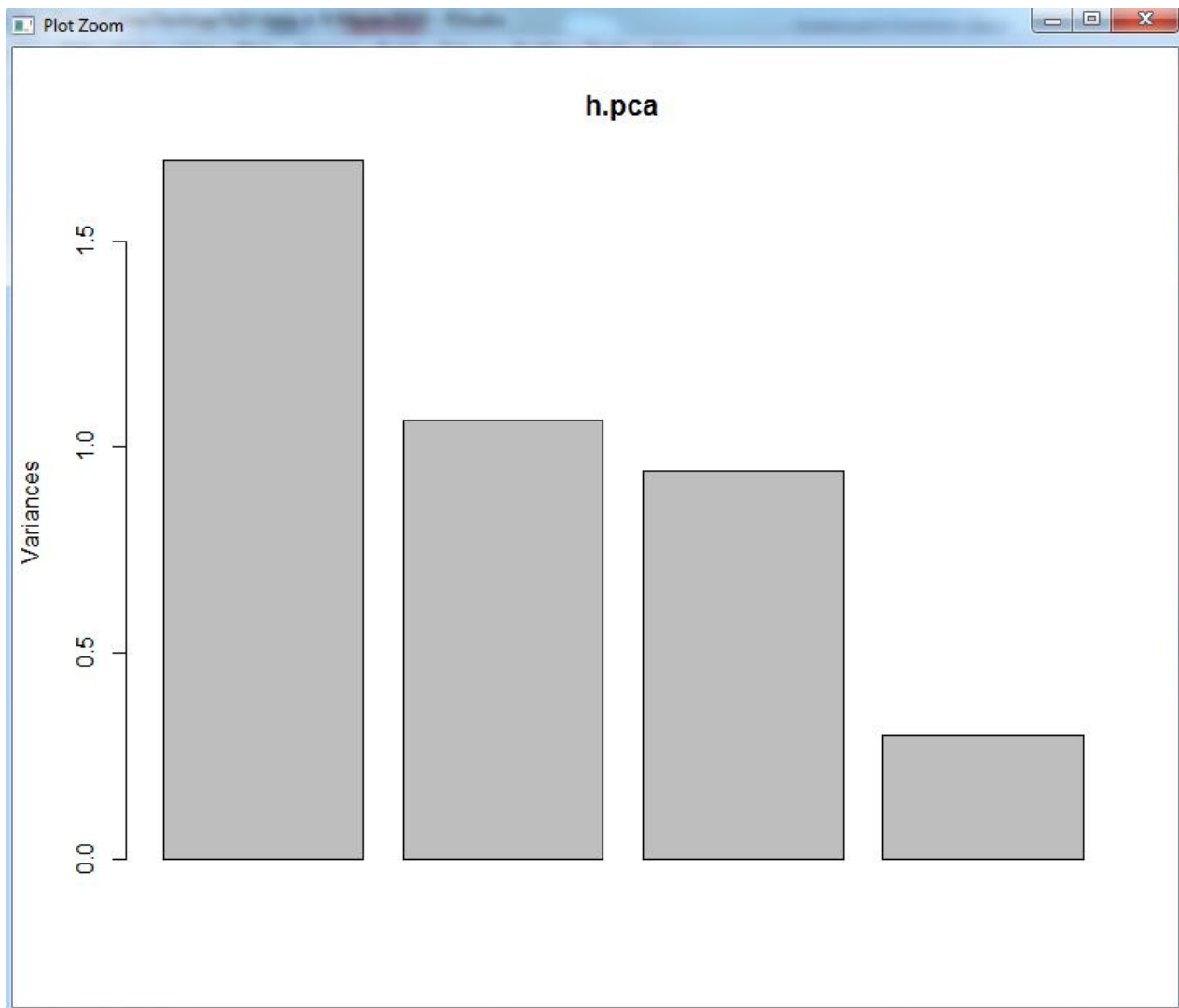


Рис.3.5. Графік вибіркової дисперсії

Дане спостереження було додатково візуалізовано в осях (Рис.3.6.). Де по осі X - буде перша головна компонента; на осі Y - друга головна компонента. За графіком видно, що перша головна компонента в себе включає інвестиційну мету та відношення до ризику. А друга стать та вік, що, як бачимо за графіком несуттєво буде впливати на рішення інвестора.

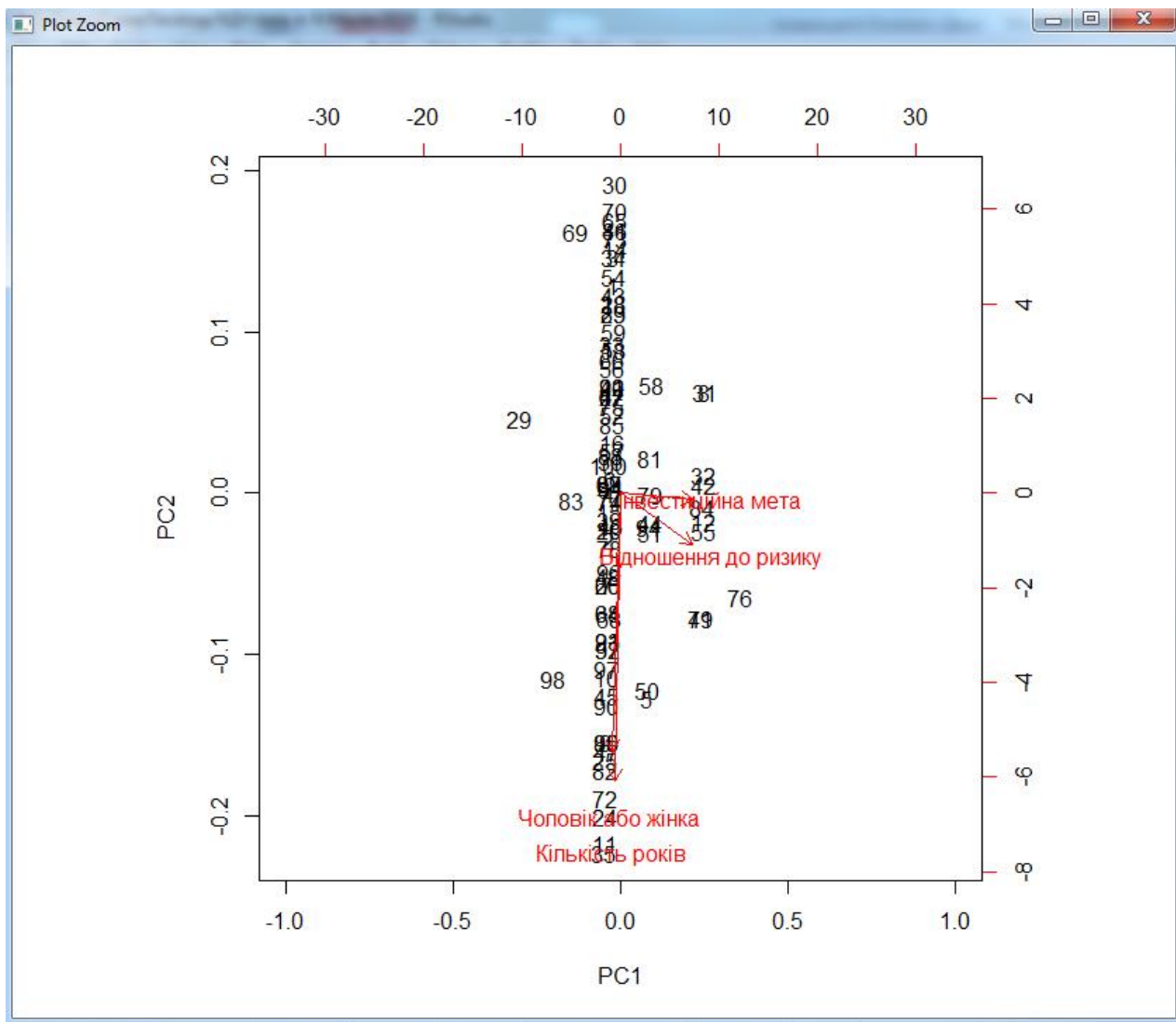


Рис.3.6. Візуалізація розрахунку на осях

Код розрахунку методу головних компонент в Rstudio (Рис.3.7.)

```

# Тип  | Пункт | Формат | Вид  | Строка
-----|-----|-----|-----|-----
library("dplyr") # манипуляції с даними
library("psych") # описательні статистики
library("lmtest") # тести для лінійних моделей
library("glmnet") # LASSO + ridge
library("ggplot2") # графіки
library("car") # vif

# Метод головних компонент

h <- data # набір даних по клієнтам
glimpse(h) # поглянемо на набір даних
describe(h) # описательні статистики

# кореляційна матриця
cor(h)

# метод головних компонент с попередньою стандартизацією змінних
h.pca <- prcomp(h,scale=TRUE)

# вивели першу головну компоненту:
pca1 <- h.pca$x[,1]
head(pca1)
tail(pca1)
# вивели ваги, з якими змінні входять в першу головну компоненту:
v1 <- h.pca$rotation[,1]
v1

# вибірка дисперсії кожної компоненти:
summary(h.pca)

# вибірка дисперсії кожної компоненти на графіку:
plot(h.pca)

# вихідні дані в нових осях
# по горизонталі --- pc1
# по вертикалі --- pc2
biplot(h.pca,xlim=c(-1,1))
|

```

Рис.3.7. Код методу головних компонент в Rstudio

3.1. Формування моделей оптимального інвестиційного портфеля.

Мета інвестора – вкласти власні кошти таким чином, аби не тільки зберегти свій капітал, але й за нагоди збільшити його.

Набір цінних паперів, котрі перебувають в учасника ринку, називається його портфелем. Вартість портфеля – це сумарна вартість всіх складових частин його паперів. Якщо на сьогоднішній день його вартість є P , а через рік вона буде дорівнювати \hat{P} , то $(\hat{P} - P)/P$ потрібно назвати прибутковістю портфеля у відсотках річних. Отже прибутковість портфеля – це дохідність на одиницю його вартості.

Нехай x_i – частка капіталу, котру витратили на купівлю цінних паперів i -го виду. Весь вкладений капітал беремо за одиницю.[41] Нехай d_i - дохідність у частках річних паперів i -го виду з розрахунку на одну грошову одиницю.

Знайдемо прибутковість всього портфеля d_p . З одного боку, через рік капітал портфеля буде дорівнювати $1 + d_p$, з іншого – вартість паперів i -го виду збільшиться з x_i до $x_i + d_i x_i$, так що сумарна вартість портфеля буде

$$\sum_i x_i + \sum_i x_i d_i = 1 + \sum_i x_i d_i$$

Прирівнюючи обидва вирази для вартості портфеля, отримуємо:

$$1 + d_p = 1 + \sum_i x_i d_i$$

Звідси:

$$d_p = \sum_i x_i d_i \tag{2.32}$$

Отже, задача збільшення капіталу портфеля еквівалентна такій же задачі про дохідність портфеля, котра виражена через дохідність паперів і їх частини формулою (2.32).

Як правило, дохідність може змінюватись в часі, так що можна вважати її випадковою величиною. Нехай m_i - середня очікувана прибутковість, тобто $m_i = M[d_i]$ - математичне очікування дохідності і $D_i = \sigma^2$ - дисперсія i -ї дохідності. Позначимо $\sigma^2 = D_i = r_i$ і будемо називати m_i і r_i відповідно, ефективністю та ризиком i -го цінного паперу. Через v_{ij} позначимо коваріації дохідностей цінних паперів i -го і j -го видів.

Оскільки дохідність частин, що входять в портфель цінних паперів випадкова, то й дохідність даного портфеля являється також випадковою величиною.

Математичне очікування дохідності портфеля позначимо через m_p і позначимо як ефективність портфеля:

$$m_p = M[d_p] = x_1 \cdot M[d_1] + \dots + x_n \cdot M[d_n] = \sum_i x_i m_i$$

Дисперсію дохідності портфеля визначимо через r_p і назвемо ризиком портфеля.

$$r_p = D[d_p] = \sum_i \sum_j x_i x_j ij$$

Вихідні дані на листі Microsoft Excel наведені на Рис.3.8.

	A	B	C	D	E
1	BTC-USD	ETH-USD	LTC-USD	NEO-USD	BCH-USD
2	6377,78	200,63	50,27	16,164	424,09
3	6388,44	200,19	51,33	15,9823	462,01
4	6361,26	207,49	50,97	16,6304	477,99
5	6376,13	218,45	53,61	16,2908	560,18
6	6419,66	218,56	55,95	16,9895	558,47
7	6461,01	217,18	55,96	16,2942	628,51
8	6530,14	212,23	52,37	15,941	617,26
9	6453,72	210,07	51,9	15,9742	617,32
10	6385,62	212,53	51,96	15,9851	544,42
11	6411,27	211,34	52,4	15,7727	544,45
12	6411,27	210,42	51,36	15,3724	533,34
13	6371,27	206,83	50,65	15,3724	516,31
14	6359,49	181,4	49,7	13,52	439,31
15	6359,49	175,18	43,61	12,6967	421,32
16	5738,35	174	42,46	12,3735	421,65
17	5648,03	177,07	41,93	12,5928	388,82
18	5575,55	149,18	42,33	12,5928	386,25
19	5554,33	130,34	36,64	10,1746	386,54
20	5623,54	136,7	33,04	8,9536	336,95
21	4871,49	126,71	32,2	9,2109	223,7

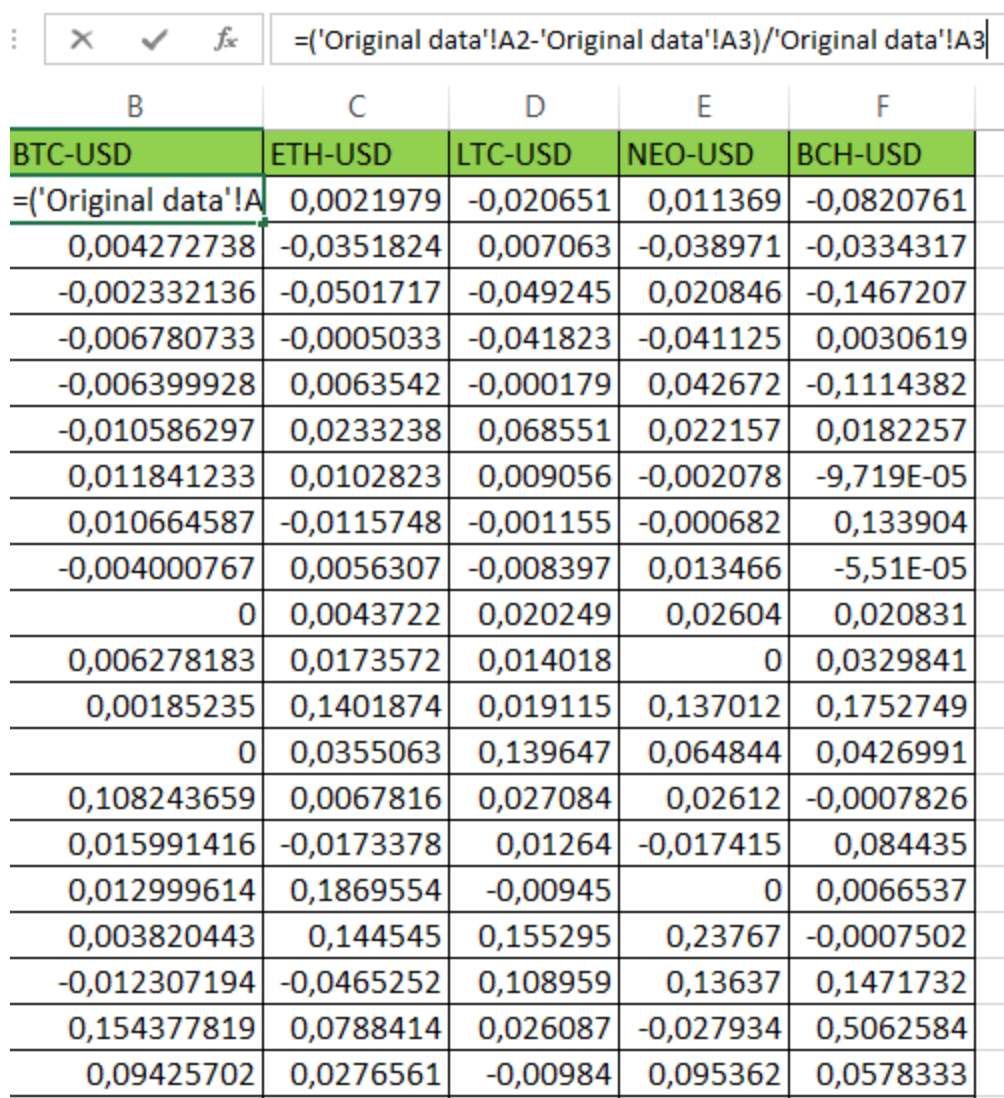
Рис.3.8. Вихідні дані для розрахунку

Перш ніж побудувати модель необхідно визначити дохідність кожного з виду цінного паперу $d_i = (\dot{P}_i - P_i)/P_i$ для кожного моменту часу, тобто побудувати матрицю D .

Формули розрахунку наведені на Рис.3.9.

Необхідно визначити частки капітали, котрі були витрачені на закупівлю кожного виду криптовалют, тобто елементи матриці $X = \{x_1; x_2, \dots, x_5\}$.

У даному випадку задамося початковим наближенням $X = \{0,1; 0,1; 0,1; 0,1; 0,1\}$.



	B	C	D	E	F
	BTC-USD	ETH-USD	LTC-USD	NEO-USD	BCH-USD
=('Original data'!	A2	-'Original data'!	A3)	/'Original data'!	A3]
	0,004272738	-0,0351824	0,007063	-0,038971	-0,0334317
	-0,002332136	-0,0501717	-0,049245	0,020846	-0,1467207
	-0,006780733	-0,0005033	-0,041823	-0,041125	0,0030619
	-0,006399928	0,0063542	-0,000179	0,042672	-0,1114382
	-0,010586297	0,0233238	0,068551	0,022157	0,0182257
	0,011841233	0,0102823	0,009056	-0,002078	-9,719E-05
	0,010664587	-0,0115748	-0,001155	-0,000682	0,133904
	-0,004000767	0,0056307	-0,008397	0,013466	-5,51E-05
	0	0,0043722	0,020249	0,02604	0,020831
	0,006278183	0,0173572	0,014018	0	0,0329841
	0,00185235	0,1401874	0,019115	0,137012	0,1752749
	0	0,0355063	0,139647	0,064844	0,0426991
	0,108243659	0,0067816	0,027084	0,02612	-0,0007826
	0,015991416	-0,0173378	0,01264	-0,017415	0,084435
	0,012999614	0,1869554	-0,00945	0	0,0066537
	0,003820443	0,144545	0,155295	0,23767	-0,0007502
	-0,012307194	-0,0465252	0,108959	0,13637	0,1471732
	0,154377819	0,0788414	0,026087	-0,027934	0,5062584
	0,09425702	0,0276561	-0,00984	0,095362	0,0578333

Рис.3.9. Формули розрахунку дохідностей криптовалюти

3.2. Модель оптимального портфеля Марковіца, яка забезпечує мінімальний ризик і задану ефективність, має вигляд:

$$\begin{cases} r_p = D[d_p] = \sum_i \sum_j x_i x_j v_{ij} \rightarrow \min \\ \sum_i x_i d_i = m_p \\ \sum_i x_i = 1, \quad x_i \geq 0 \end{cases} \quad (2.33)$$

У даній моделі існує обмеження – лінійна функція. Обмеження в матричному вигляді можна написати як $X\bar{D} = m_p$, де \bar{D} - матриця-рядок середніх значень дохідностей i -х цінних паперів, які визначаються за формулою середнього арифметичного:

$$d_i \approx \frac{1}{N} \sum_{k=1}^N d_{ik} \quad (2.34)$$

Для матриці \bar{D} використано функція СРЗНАЧ (число1; число2;...), наприклад СРЗНАЧ(B16:B24).

Результати розрахунків дохідностей кожної з криптовалют, максимального значення, середніх значень дохідностей (функція МАКС(число1;число2...)) дохідності з всіх середніх наведені на Рис.3.10.

14		0,017488484	-0,03857	-0,015505	-0,030429	-0,0551938	максимальна дохідність	
15	Середня дохідність	0,0157%	0,1670%	8,6904%	0,3914%	0,5239%	8,69%	план X

Рис.3.10. Результати розрахунку

Цільова функція у моделі Марковіца (2.33) r_p - нелінійна, і являється нічим іншим як квадратична форма, яка в матричній формі запишеться як $r = X^T V_{ij} X$, де X – це матриця-стовпець змінних, а X^T – матриця-рядок змінних, котру отримаємо за допомогою транспонування.

Коваріація або кореляційний момент доходностей розраховується як:

$$V_{ij} = M \left\{ \frac{\Delta_{ik}}{(d_{ik} - d_i)} \cdot \frac{\Delta_{jk}}{(d_{jk} - d_j)} \right\} \approx \frac{1}{N} \sum_{j=1}^N \Delta_{ik} \cdot \Delta_{jk}$$

де Δ_{ik} і Δ_{jk} - відхилення доходностей i -го та j -го паперів від середньої арифметичної доходності.


Матрицю коваріації доходностей криптовалюти V_{ij} знайдемо за допомогою функції КОВАР(массив1; массив2), а саме КОВАР(\$B\$16;\$B\$24;B16;B24). Результати усіх проміжних розрахунків, а також компоненти початкового плану наведені на Рис.3.11.

Розглянута модель Марковіца [25] відноситься до моделей нелінійного програмування. Для її вирішення використовується надбудова «Пошук рішення», робота з якою була детально вивчена під час засвоєння навчальної дисципліни «Оптимізаційні методи та моделі». Результат заповнення вхідної надбудови наведено на Рис.3.12.


	0,01010000	0,0257007	0,025192	0,03074	0,0307740				
	0,000719144	0,054956	-0,007963	-0,005409	0,0611008				
	0,017488484	-0,03857	-0,015505	-0,030429	-0,0551938	максимальна доходність			
Середня доходність	0,0157%	0,1670%	8,6904%	0,3914%	0,5239%	8,69%	план X		
матриця коваріацій	0,001814935	0,0003918	0,00028	4,07E-05	0,0008476		29%	BTC-USD	
	0,000391825	0,0026115	0,000405	0,000255	0,000722		0%	ETH-USD	
	0,00027994	0,0004049	1,503373	0,0017	-0,0030485		7%	LTC-USD	
	4,07429E-05	0,000255	0,0017	0,003337	0,0003452		25%	NEO-USD	
	0,00084756	0,000722	-0,003049	0,000345	0,0065633		39%	BCH-USD	
X'	29%	0%	7%	25%	39%	100%			
X'V	0,000888516	0,000488	0,101939	0,001096	0,002696	Ефективність	75% від max доходності		
xi*di	4,56786E-05	0	0,005933	0,000976	0,0020528	0,009007215	6,52%		
						X'VX	0,00855		
						ризик			

Рис.3.11. Результат розрахунків і початковий план

Параметры поиска решения X

Оптимизировать целевую функцию: 

До: Максимум Минимум Значения:

Изменяя ячейки переменных:
 

В соответствии с ограничениями:

\$G\$332 = 1
 \$G\$334 >= \$H\$334

Сделать переменные без ограничений неотрицательными

Выберите метод решения:

Метод решения

Для гладких нелинейных задач используйте поиск решения нелинейных задач методом ОПГ, для линейных задач - поиск решения линейных задач симплекс-методом, а для негладких задач - эволюционный поиск решения.

Рис.3.12. Результат заполнения окна надстройки «Поиск решения»
 Отриманий оптимальний план наведений на Рис.3.13.

	0,017488484	-0,03857	-0,015505	-0,030429	-0,0551938	максимальна дохідність			
Середня дохідність	0,0157%	0,1670%	8,6904%	0,3914%	0,5239%	8,69%	план X		
матриця коваріацій	0,001814935	0,0003918	0,00028	4,07E-05	0,0008476		0%	BTC-USD	
	0,000391825	0,0026115	0,000405	0,000255	0,000722		0%	ETH-USD	
	0,00027994	0,0004049	1,503373	0,0017	-0,0030485		73%	LTC-USD	
	4,07429E-05	0,000255	0,0017	0,003337	0,0003452		0%	NEO-USD	
	0,00084756	0,000722	-0,003049	0,000345	0,0065633		27%	BCH-USD	
X'	0%	0%	73%	0%	27%	100%			
X'V	0,000430948	0,0004893	1,10261	0,00134	-0,0004914	Ефективність	75% від max дохідності		
xi*di	0	0	0,063784	0	0,0013937	0,065177901	6,52%		
						X'VX	0,80914		
						ризик			

Рис.3.12. Оптимальний план

Звіти за результатами та стійкістю подані на Рис.3.13 та Рис.3.14.

Ячейка целевой функции (Минимум)			
Ячейка	Имя	Исходное значение	Окончательное значение
\$H\$336	X'VX 75% від max дохідності	0,008546703	0,809144299

Ячейки переменных				
Ячейка	Имя	Исходное значение	Окончательное значение	Целочисленное
\$H\$326	матриця коваріацій план X	29%	0%	Продолжить
\$H\$327	план X	0%	0%	Продолжить
\$H\$328	план X	7%	73%	Продолжить
\$H\$329	план X	25%	0%	Продолжить
\$H\$330	план X	39%	27%	Продолжить

Ограничения					
Ячейка	Имя	Значение ячейки	Формула	Состояние	Допуск
\$G\$332	X' максимальна дохідність	100%	\$G\$332=1	Привязка	0
\$G\$334	xi*di Ефективність	0,065177901	\$G\$334>=\$H\$334	Привязка	0

Рис.3.13. Звіт за результатами

Ячейки переменных

Ячейка	Имя	Окончательное Значение	Приведенн. Градиент
\$H\$326	матриця коваріацій план X	0	0,139116967
\$H\$327	план X	0	0,098364779
\$H\$328	план X	0,733963273	0
\$H\$329	план X	0	0,039454078
\$H\$330	план X	0,266036727	0

Ограничения

Ячейка	Имя	Окончательное Значение	Лагранжа Множитель
\$G\$332	X' максимальна доходність	1	-0,142503923
\$G\$334	$x_i \cdot d_i$ Ефективність	0,065177901	27,01517697

Рис.3.14. Звіт зі стійкості отриманого плану


Таким чином, для мінімізації ризику капітал необхідно розподілити між криптовалютами так: найбільшу частку – 73 % – вкласти в криптовалюту LTC, 23 % капіталу вкласти в криптовалюту BCH. В акції BTC, NEO та ETH в цьому випадку вкладати капітал не потрібно взагалі.

3.3. Оптимальний портфель Марковіца заданого ризику й максимальної можна представити в наступному вигляді:


$$\begin{cases} m_p = \sum_i x_i d_i \rightarrow \max \\ \sum_i \sum_j x_i x_j v_{ij} = r_p \\ \sum_i x_i = 1, \quad x_i \geq 0 \end{cases} \quad (2.35)$$

Для вирішення даної задачі нелінійного програмування потрібно використати надбудову "Пошук рішення". Прийнятний ризик задано на рівні 125 % від мінімального рівня ризику, знайденого на попередньому етапі. Результат заповнення вікна надбудови "Пошук рішення" для моделі Марковіца з максимізацією ефективності наведено на Рис.3.15.

Параметры поиска решения X

Оптимизировать целевую функцию: 

До: Максимум Минимум Значения:

Изменяя ячейки переменных:
 

В соответствии с ограничениями:

\$G\$332 = 1
 \$H\$336 <= \$H\$339

Сделать переменные без ограничений неотрицательными

Выберите метод решения:

Метод решения

Для гладких нелинейных задач используйте поиск решения нелинейных задач методом ОПГ, для линейных задач - поиск решения линейных задач симплекс-методом, а для негладких задач - эволюционный поиск решения.

Рис.3.15. Результат заполнения окна надстройки «Поиск решения»

Отриманий оптимальний план наведений на Рис.3.16.

	0,017488484	-0,03857	-0,015505	-0,030429	-0,0551938	максимальна дохідність		
Середня дохідність	0,0157%	0,1670%	8,6904%	0,3914%	0,5239%	8,69%	план X	
матриця коваріацій	0,001814935	0,0003918	0,00028	4,07E-05	0,0008476		0%	BTC-USD
	0,000391825	0,0026115	0,000405	0,000255	0,000722		0%	ETH-USD
	0,00027994	0,0004049	1,503373	0,0017	-0,0030485		100%	LTC-USD
	4,07429E-05	0,000255	0,0017	0,003337	0,0003452		0%	NEO-USD
	0,00084756	0,000722	-0,003049	0,000345	0,0065633		0%	BCH-USD
σ'	0%	0%	100%	0%	0%	100%		
σ'V	0,00027994	0,0004049	1,503373	0,0017	-0,0030485	Ефективність		
σi*di	0	0	0,086904	0	0	0,086903868		

Рис.3.16 Оптимальний план

Таким чином, для максимізації ефективності капітал необхідно розподілити між криптовалютою так: всі 100% коштів вкласти в криптовалюту - LTC. В криптовалюту BTC, ETH, NEO, BCH вкладати капітал не потрібно.

3.4. Ризиковано - ефективна модель.

Зазначені вище моделі Марковіца в обмеженнях використовують визначені рівні ризику та дохідності, що зазначаються попередньо. Ці моделі залежать від думок і знань експертів ринку цінних паперів і не дають інвестору відповідь на головне запитання, куди вкласти власні заощадження – не ризиковану (малодохідну) або ризиковану (прибуткову). [37]Для подолання цього непорозуміння була запропонована ризиковано-прибуткова модель, що дозволяє не розмірковувати над проблемою визначення кращого рівня ризику для кожного портфеля і має такий вигляд:

$$f(x_i) = \frac{\sqrt{\sum_i x_j^2 v_j^2 + \sum_i \sum_j x_i x_j v_{ij}}}{\sum_i x_i d_i} \quad (2.36)$$

$$\sum_i x_i = 1; \quad \forall x_i \geq 0$$

де v_i – варіація дохідності цінних паперів i -го виду; v_{ij} – коваріація дохідності цінних паперів i -го та j -го видів.

Цільова функція в матричній формі запишеться як:

$$f = \frac{\sqrt{X^2 V_i^2 + X^T V_{ij} X}}{X \bar{D}}$$

де V_i^2 - матриця варіацій дохідностей цінних паперів.

3.5. Оптимальний портфель за моделлю Тобіна. Після появи робіт Дж.Тобіна, вплив "портфельної теорії", котра була запропонована Марковіцом значно посилюється. Дж. Тоббін зауважив, що ринковий портфель, а саме сукупність усіх наявних у даний момент часу у інвесторів цінних паперів, є ефективним. Більш того будь-яка комбінація ринкового портфеля з безризиковим активом дає такий же ефективний портфель, який має менший ризик, хоча і з меншим очікуваним доходом. Дж. Тобін запропонував включити державні облігації в аналіз безризикових активів. Саме тому моделі оптимального портфеля Тобіна відрізняються від моделей оптимального портфеля Марковіца лише наявністю безризикових паперів.

Модель оптимального портфеля Тобіна, яка забезпечує мінімальний ризик і задану ефективність, має такий вигляд:

$$\begin{cases} \sum_i \sum_j x_i x_j v_{ij} \rightarrow \min; \\ x_0 m_0 + \sum_i x_i d_i = m_p; \\ x_0 + \sum_i x_i = 1, \quad x_i \geq 0. \end{cases} \quad (2.37)$$

де m_0 - ефективність безризикових паперів;

x_0 - частка капіталу, вкладена в без ризикові папери;

$x_i x_j$ - частка капіталу, вкладена в цінні папери i -го і j -го видів;

d_i – математичне сподівання (середнє арифметичне) дохідності i -го цінного паперу;

v_{ij} - кореляційний момент між ефективністю паперів i -го і j -го видів.

Оптимальний портфель Табіна максимальної ефективності й заданого (прийнятого) ризику можна представити у вигляді:

$$\begin{cases} x_0 m_0 + \sum_i x_i d_i \rightarrow \max; \\ \sum_i \sum_j x_i x_j v_{ij} = r_p; \\ x_0 + \sum_i x_i = 1, \quad x_i \geq 0. \end{cases} \quad (2.38)$$

де r_p - заданий ризик портфеля.

Розрахункові формули та методи вирішення аналогічні формулам задачі формування оптимального портфеля цінних паперів Марковіца, наведені вище.

ВИСНОВКИ ДО РОЗДІЛУ 3

Отже, після проведеного мною дослідження, можна зробити наступні висновки:

Результат кластерного аналізу, котрий був проведений задля сегментації вибірки, котра складалась зі 100 клієнтів італійських банків, що характеризувалися за чотирма критеріями (відношення до ризику, інвестиційна мета, стать, вік) показав, що набір клієнтів розподілився на 2 найбільші кластери: 1 – жінки працездатного віку, схильні до ризику та низької дохідності; 2 – жінки-пенсіонери, схильні до ризику та середньої дохідності.

За допомогою платформи Rstudio та методу головних компонент, була отримана відповідь на запитання: «Який з критеріїв має найбільший вплив на рішення інвестора?». По-перше, розрахувавши кореляційну матрицю, ми побачили, що найбільший зв'язок спостерігається між критерієм «відношення до ризику» та «інвестиційною метою», а саме 69%. По-друге, за результатом вилучення першої головної компоненти («відношення до ризику»), було зрозуміло, що найбільший вплив на рішення інвестора має саме відношення до ризику. Даний показник може пояснити 42,3% рішення інвестора, на другому місці показник інвестиційна мета, котрий може пояснити 26,6% рішення інвестора. Для наочності було побудовано графік та візуалізовано результати на осях, що підтверджує розрахунки.

Останнім етапом проведення розрахунків, було за допомогою моделі Марковіца, Тобіна визначити яку частку капіталу потрібно вкласти інвестору задля отримання максимальному прибутку при мінімальному ризику. фінансовими інструментами для інвестування обрано 5 найбільш популярних криптовалют, а саме BTC(Bitcoin), ETH(Ethereum), BCH(Bitcoin Cash), LTC(Litecoin), NEO., для моделі визначення інвестиційного плану застосовувалися темпи приросту цін криптовалют в проміжок часу з 1 листопада 2018 року по 1 листопада 2019 року. Дані для розрахунку були

отримані з сайту, який знаходиться у відкритому доступі, <https://finance.yahoo.com/cryptocurrencies> (Додаток А).

Таким чином, за результатами розрахунків, для мінімізації ризику капітал необхідно розподілити між криптовалютами так: найбільшу частку – 73 % – вкласти в криптовалюту LTC, 23 % капіталу вкласти в криптовалюту BCH. В NEO, BTC та ETH в цьому випадку вкладати капітал не потрібно взагалі.

В свою чергу, для максимізації ефективності капітал необхідно розподілити між криптовалютами так: всі 100% коштів вкласти в криптовалюту - LTC. В інші, а саме в BTC, ETH, NEO, BCH вкладати капітал не потрібно.

ВИСНОВКИ

Результати проведеного дослідження дають підстави зробити такі висновки:

Інвестиційний проект – це сукупність намірів та певний дій інвестора, що здійснює вклади, з метою забезпечення певних намірів, а саме: виробничих, соціальних, економічних чи фінансових вкладень з метою отримання прибутку.

Використання персоналізований рекомендацій, а особливо персоналізований рекомендацій в інвестуванні є актуальним питанням сьогодення. Завдяки персоналізованим рекомендаціям, інвестор може обрати найбільш ефективний спосіб інвестування коштів.

Кластерний аналіз – це сукупність багатовимірних статистичних процедур, котрі дозволяють впорядкувати об'єкти за однорідними групами.

Класичним прикладом такого розподілу на кластери є різний кінцевий прибуток інвестованих коштів (отримає інвестор прибуток чи навпаки, втратить свої заощадження) залежності від того, куди саме обере вкласти фінанси інвестор.

Метод головних компонент – один з основних способів зменшити розмір даних, втративши при цьому найменшу кількість інформації. Основними передумовами для того, щоб ефективно понизити розмірність масиву даних, є «сильний» зв'язок між початковими змінними, внаслідок чого інформація, що міститься в даних, дублюється, що дозволяє об'єднання декількох показників в один, тим самим зменшити вибірку даних.

Проведення розрахунків задля сегментації груп споживачів для розробки інвестиційних планів з персоналізованими рекомендаціями показав, що:

Вибірка зі 100 клієнтів італійських банків розподілилася на два основні кластери 1 – жінки працездатного віку, схильні до ризику та низької дохідності; 2 – жінки-пенсіонери, схильні до ризику та середньої дохідності.

Для даних клієнтів найбільший вплив на прийняття рішення щодо інвестування має їх відношення до ризику та інвестиційна мета, вік та стать не майже не мають значення. Фінансовим інструментом для інвестування обрана криптовалюта для моделі визначення інвестиційного плану застосовувалися темпи приросту цін криптовалют за останні 365 днів. Розрахунок показав наступне: для мінімізації ризику капітал необхідно розподілити між криптовалютами так: найбільшу частку – 73 % – вкласти в криптовалюту LTC, 23 % капіталу вкласти в криптовалюту BCH. В NEO, BTC та ETH вкладати капітал не потрібно взагалі. В свою чергу, для максимізації ефективності – всі 100% коштів вкласти в криптовалюту - LTC. В інші, а саме в BTC, ETH, NEO, BCH вкладати капітал не потрібно.

СПИСОК ВИКОРИСТАНИ ДЖЕРЕЛ

1. Інвестиційний менеджмент : навч. посіб. / [Гриньова В. М., Коюда В. О., Лепейко Т. І. та ін.]. — [2-ге вид., доопрац. і доп.]. — Х. : ВД «ІНЖЕК», 2005. — 664 с.
2. Гукалюк А. Ф. Інвестиційна привабливість як чинник підвищення конкурентоспроможності території / А. Ф. Гукалюк, І. М. Іванович // Актуальні проблеми економіки. — 2010. — № 6(108). — С. 167—174.
3. Власов М. П. Моделирование экономических процес сов / М. П. Власов, П. Д. Шимко. — Ростов н/Д : Феникс, 2005. — 409, [1] с.
4. Кузнецова С. В. Управление портфелем ценных бумаг: модели Марковица и Тобиана : учебн. пособ. / С. В. Кузнецова, Н. Р. Стронгина ; М-во образования Рос. Федерации ; Нижегород. гос. ун-т им. Н. И. Лобачевского. — Н. Новгород : Изд-во Нижегород. госуниверситета, 2002. — 102 с.
5. Портфельне інвестування : навч. посібн. / А. А. Пересада, О. Г. Шевченко, Ю. М. Коваленко та ін. — К. : КНЕУ, 2004. — 408 с.
6. Нурминский Е. А. Оптимальный портфель инвестора в модели Марковица (На примере рынка ГКО, 1994 – 1995 гг.) / Е. А. Нурминский, А. В. Пономаренко. — Владивосток : ИПМ, 1996. — 20 с.
7. Воронин. А.В. Использование кластерного анализа для выбора локальных стратегий[Текст] / А.В. Воронин // Проблемы и перспективы управления экономикой и маркетингом ворганизации. — №1. — 2001.
8. Енюков И.С. Факторный, дискриминантный и кластерный анализ [Текст] / Енюков И.С. — М.: Финансы и статистика, 1989.
9. Алгоритми та програми відновлення залежностей / Під ред.. В.М. Вапника. — М.: Наука, 1984, - 186с.
10. Плюта В. Порівняльний багатовимірний аналіз в економетричному моделюванні / Пер. з польськ. — М.: Фінанси та статистиа, 1989. — 175с.
11. Факторний дискримінантний та кластерний аналіз: Пер. з англ.. — М.:Фінанси та статистика, 1989. — 215 с.

12. Лапач С.М., Чубенко А.В., Бабія П.М. Статистика в науці та бізнесі. – К.: МОРІОН, 2002. – 640с.
13. Бромвич М. Аналіз економічної ефективності капіталовкладень: Пер. з англ.. – М.: ИНФРА-М, 1996., - 424с.
14. Олійник А. О. Інтелектуальний аналіз даних: навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя: ЗНТУ, 2012. – 271 с.
15. Бахрушин В. Є. Методи аналізу даних: навчальний посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2011. – 268 с.
16. Багрова І.В. Економічний механізм забезпечення надійності інвестиційних проектів: методологічні засади та їх практичне застосування / І.В. Багрова, Т.С. Яковенко // Вісник економічних наук України. – 2009. – № 1(15). – С. 29-33.
17. Дука А.П. Теорія та практика інвестиційної діяльності: навч. посібн. / А.П. Дука. – К.: Вид-во «Каравела», 2007. – 424 с
18. Meyer K. E. Investment strategies in emerging markets. – UK: Edward Elgar Publishing Ltd, 2004.
19. Lam J. W. Robo-Advisors: A Portfolio Management Perspective / J. W. Lam. – Yale College New Haven, Connecticut, 2016.
20. Авакян Н. Инвестиционные портфели – как получить 40% годовых при практически нулевых рисках [Електронний ресурс] / Нарек Авакян // utmagazine. – 2014. – Режим доступу до ресурсу: <https://utmagazine.ru/posts/5285-post-547aefcf30174>.
21. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д.Мешалкин. – М.: Финансы и статистика. 1989. – 607с.
22. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистический методы. – М.: Финансы и статистика, 1998. – 352 с.

23. Алгоритмы и программы восстановления зависимостей / Под ред. В.И.Вапника. – М.: Наука, 1984. – 816с.

24. Горохов М.Ю., Малев В.В. Бизнес-планирование и инвестиционный анализ. -М.: Информационно-издательский дом Филинь. - 1998. - 670с.

25. Timirova A. N. Svoystva modeli Markovitsa pri zadanii parametrov sredstvami teorii nechetkih mnozhestv [Markovits model properties when specifying parameters using the fuzzy set theory]. Available at: [http://www.intsys.msu.ru/magazine/archive/v16\(1-4\)/timirova-203-214.pdf](http://www.intsys.msu.ru/magazine/archive/v16(1-4)/timirova-203-214.pdf)

26. Шарп У., Александер Г., Бэйли Д.. Инвестиции. - М.: Инфра-М. - 1997. -1027с.

27. Zverev V. A., Zvereva A. V., Evsyukov S.G. Ryinok tsennyih bumag [Stocks and bonds market]. Moskva, Dashkov i K Publ., 2016. 496 p.

28. Tereschenko A. O. Finansovaya deyatel'nost' sub'ektov hozyaystvovaniya [Financial activities of business entities]. Kiev, KNEU Publ., 2003. 554 p.

29. Крамаренко Г.О. Фінансовий аналіз та планування. – Дніпропетровськ: Видавництво ДАУБП, 2001. – 224 с.

30. Лабораторні та практичні роботи з методики викладання математики. / Под ред. Лященко Є.І. - М: Освіта, 1988.

Інтернет ресурси:

31. Брюховецька Н. Ю. Оцінка інвестиційної привабливості підприємства: визначення недоліків деяких існуючих методик / Н. Ю. Брюховецька, О. В. Хасанова. [Електронний ресурс] – Режим доступу: http://www.nbuv.gov.ua/portal/Soc_Gum/eprom/2009_44/st_44_15.pdf

32. Дані про темпи росту криптовалют [Електронний ресурс] – Режим доступу: (<https://finance.yahoo.com/cryptocurrencies>)

33. Офіційний сайт програми Stata [Електронний ресурс] – Режим доступу: (<https://www.stata.com/why-use-stata/>)
34. Вікіпедія – поняття «інвестиційний проект» [Електронний ресурс] – Режим доступу: (https://uk.wikipedia.org/wiki/%D0%86%D0%BD%D0%B2%D0%B5%D1%81%D1%82%D0%B8%D1%86%D1%96%D0%B9%D0%BD%D0%B8%D0%B9_%D0%BF%D1%80%D0%BE%D0%B5%D0%BA%D1%82)
35. Вікіпедія – поняття «кластерний аналіз» [Електронний ресурс] – Режим доступу: (https://uk.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%BD%D0%B8%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7)
36. Medina J. The Chilean Business Cycles Through the Lens of a Stochastic General Equilibrium Model [Електронний ресурс] / J. Medina, С. Soto // Working Papers Central Bank of Chile. – 2007. – Режим доступу до ресурсу: <https://ideas.repec.org/p/chb/bcchwp/457.html>.
37. Інтегровані підходи до управління. [Електронний ресурс] – Режим доступу: https://pidruchniki.com/12461220/menedzhment/integrovani_pidhodi_upravlinnya
38. Моделювання купівельної поведінки організації-споживача . [Електронний ресурс] – Режим доступу: (<http://library.if.ua/book/42/2936.html>).
39. Що таке клієнтоорієнтованість? [Електронний ресурс] – Режим доступу:(<https://xsreality.org/что-такое-клиентоориентированность-printsipy-klientoorientirovannostiklientoorientirovannost-eto-odin-iz-naibolee-vaznyh-voprosov-kotorye-dolzhna-reshitlyubaya-kompaniya-vo-vremya-svoego-stanvolen/>).
40. «Метод Главных Компонент (РСА)» [Електронний ресурс] – Режим доступу: <http://bourabai.kz/cm/pca.htm>

41. Авакян Н. Модель портфеля ценных бумаг Гарри Марковица - суть и принципы построения [Электронный ресурс] / Нарек Авакян // utmagazine. – 2014. – Режим доступа до ресурсу: <https://utmagazine.ru/posts/5365-model-portfelya-cennyh-bumag-garri-markovica-sut-i-principyu-postroeniya>.

42. Метод головних компонент в Rstudio. [Электронный ресурс] – Режим доступа: <https://www.youtube.com/watch?v=SObiC3JzHxw>

43. Сегментація споживачів [Электронный ресурс] – Режим доступа: (https://stud.com.ua/129898/marketing/segmentatsiya_spozivachiv)

44. Стратегія підприємства: поняття, еволюція, концепція [Электронный ресурс] – Режим доступа: http://dspace.uccu.org.ua/bitstream/123456789/2179/1/СП_Безпарточний.doc

45. Фінансові інвестиції, їх відмінність від інвестицій у реальний капітал. Поняття “портфельних інвестицій”. [Электронный ресурс] – Режим доступа: https://studopedia.ru/1_112581_Investitsiyne-seredovishche-ta-yogo-elementi.html

46. Інвестиції як економічна категорія та їх роль в економіці. [Электронный ресурс] – Режим доступа: https://knowledge.allbest.ru/economy/2c0a65635a2ac79a4c43b89421216d27_0.html

47. Фінансові інвестиції – це про вкладення фінансових ресурсів у фінансові активи. [Электронный ресурс] – Режим доступа: <https://studopedia.org/2-129321.html>

48. Дискримінантний аналіз при нормальному розподілі показників. [Электронный ресурс] – Режим доступа: <http://ukrdoc.com.ua/text/39882/index-1.html>

49. Формування портфеля інвестицій та його моделей. [Электронный ресурс] – Режим доступа:

https://studme.com.ua/15060913/investirovanie/formirovanie_portfelya_investitsiy_ego_modeli.htm

50. Фінансовий ринок як елемент інвестиційного середовища. [Електронний ресурс] – Режим доступу: <http://osvita-plaza.in.ua/index/0-1490>

51. Обґрунтування інвестиційних проектів. [Електронний ресурс] – Режим доступу: <https://works.doklad.ru/view/fEuojogDJOc.html>

Додаток А

yahoo! finance Search for news, symbols or companies [Sign in](#) [Mail](#)

Finance Home Watchlists My Portfolio Screeners Premium Markets Industries Personal Finance Videos News Tech

Cryptocurrencies Calendars Trending Tickers Stocks: Most Actives Stocks: Gainers Stocks: Losers Top ETFs Commodities World Indices Currencies Top Mutual Funds ...

Results List heatmap view

Matching Cryptocurrencies 1-22 of 118 results [Add to Portfolio](#) [Share](#)

ⓘ Results were generated a few mins ago. Pricing data is updated frequently. Currency in USD

<input type="checkbox"/> Symbol	Name	Price (Intraday)	Change	% Change	Market Cap	Volume in Currency (Since 0:00 UTC)	Volume in Currency (24Hr)	Total Volume All Currencies (24Hr)	Circulating Supply	52 Week Range	1 Day Chart
<input type="checkbox"/> BTC-USD	Bitcoin USD	7,453.24	-116.42	-1.54%	134.733B	18.835B	18.835B	18.835B	18.077M	3,191.30 - 13,796.49	
<input type="checkbox"/> ETH-USD	Ethereum USD	151.44	-0.77	-0.51%	16.471B	7.191B	7.191B	7.191B	108.76M	82.83 - 361.40	
<input type="checkbox"/> XRP-USD	XRP USD	0.2232	-0.0033	-1.44%	9.666B	1.141B	1.141B	1.141B	43.3B	0.21 - 0.51	
<input type="checkbox"/> USDT-USD	Tether USD	1.0013	-0.0006	-0.06%	4.113B	21.05B	21.05B	21.05B	4.108B	0.97 - 1.06	
<input type="checkbox"/> BCH-USD	Bitcoin Cash USD	218.05	-1.11	-0.51%	3.956B	1.264B	1.264B	1.264B	18.143M	75.08 - 522.09	
<input type="checkbox"/> LTC-USD	Litecoin USD	47.96	+0.53	+1.11%	3.057B	2.806B	2.806B	2.806B	63.746M	22.82 - 148.43	
<input type="checkbox"/> EOS-USD	EOS USD	2.8159	+0.0631	+2.29%	2.654B	1.684B	1.684B	1.684B	942.499M	1.59 - 8.59	
<input type="checkbox"/> BNB-USD	Binance Coin USD	15.48	-0.21	-1.32%	2.407B	214.709M	214.709M	214.709M	155.537M	4.19 - 39.57	
<input type="checkbox"/> XLM-USD	Stellar USD	0.0594	+0.0018	+3.12%	1.191B	227.108M	227.108M	227.108M	20.055B	0.05 - 0.17	