

РОЗДІЛ 9. МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ
ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІВИКОРИСТАННЯ ЕКОНОМІКО-МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ
У КОНСАЛТИНГОВІЙ ДІЯЛЬНОСТІ В УМОВАХ АСИМЕТРИЧНОЇ ІНФОРМАЦІЇ
THE USE OF ECONOMIC-MATHEMATICAL MODELLING IN CONSULTING
ACTIVITIES IN CONDITIONS OF ASYMMETRIC INFORMATION

У статті розглянуто використання економіко-математичного моделювання консалтинговими компаніями в умовах асиметричної інформації. Автор описує застосування стандарту CRISP-DM до задач консалтингової компанії та показує етапи підготовки даних для побудови математичних моделей та методи оцінювання прогнозованих значень. Такий підхід дозволить оптимізувати побудову моделей під конкретні бізнес задачі підприємств-клієнтів консалтингових компаній.

Ключові слова: асиметрична інформація, консалтинг, інформаційна нерівність, CRISP-DM, підготовка даних, оцінка моделей.

В статье рассмотрено использование экономико-математического моделирования консалтинговыми компаниями в условиях асимметричной информации. Автор описывает применение стандарта CRISP-DM к задачам консалтинговой компании и показывает этапы подготовки данных

для построения математических моделей и методы оценки прогнозируемых значений. Такой подход позволит оптимизировать построение моделей под конкретные бизнес задачи предприятий-клиентов консалтинговых компаний.

Ключевые слова: асимметричная информация, консалтинг, информационное неравенство, CRISP-DM, подготовка данных, оценка моделей.

The article discusses the use of economic-mathematical modelling by consulting companies in the conditions of asymmetric information. The author describes the application of the CRISP-DM standard to the tasks of a consulting company and shows the stages of data preprocessing for mathematical models and methods of evaluation predicted values. This approach will optimize building of models for specific business tasks of consulting companies' clients.

Key words: asymmetric information, consulting, information inequality, CRISP-DM, data preprocessing, model evaluation.

УДК 330.46

Бачало І.Б.аспірант кафедри економічної
кібернетикиЛьвівський національний університет
імені Івана Франка

Постановка проблеми. Консалтингові компанії щоденно стикаються з ситуацією збору та аналізу інформації з метою виробити для свого клієнта найкраще рішення та, за потреби, впровадити його у життя. Проте, трапляються обставини, коли можливе виникнення ризиків, пов'язаних з асиметричністю інформації – ситуацією, коли певні економічні суб'єкти, що ведуть економічну діяльність з клієнтом консалтингової компанії, володіють важливою інформацією, якою не володіє компанія-клієнт, та можуть приховувати або викривлювати цю інформацію. Така ситуація створює проблему неефективності функціонування ринку, що може завдати збитків від недобросовісних дій економічних суб'єктів та викликає витрати на збір додаткової інформації компанією клієнтом, що впливатиме на ціну його продукції в майбутньому.

Аналіз останніх досліджень і публікацій. Виявами асиметричності інформації в економіці почали займатися ще Ф. Хайєк, В. Вікрі та К. Ерроу, проте, значний поштовх дослідженню цією теми надали нобелівські лауреати Дж. Стігліц, Дж. Акерлоф та М. Спенс. Серед українських вчених асиметричність інформації на різних ринках у своїх роботах досліджували Л. Бабич, Н. Балдич, В. Огієнко, О. Пластун, О. Дудкін, К. Рожкова, В. Мазуренко, А. Маслов та інші. Проте, проблеми вико-

ристання економіко-математичного моделювання для усунення асиметричності інформації консалтинговими компаніями, які зумовлені нерівномірним розподілом інформації у різних його виявах, потребують подальшого дослідження.

Формулювання цілей статті. Метою статті є розроблення методики врахування асиметричності інформації консалтинговими компаніями із застосуванням економіко-математичного моделювання.

Виклад основного матеріалу дослідження. Ризик виникнення асиметричної інформації може виявлятися на будь-якому з етапів ведення консалтингового проекту та може бути наявним як до початку роботи з економічними суб'єктами – «ex-ante», так і під час роботи з ними «ex-post» [1].

З метою запобігання виникнення таких ризиків консалтингова компанія повинна бути у змозі зібрати якнайбільше інформації про галузь, у якій вона надає консультацію, економічних суб'єктів, з якими вона має справу (конкуренти компанії клієнта, споживачі, підрядники, працівники) та бути в курсі трендів та очікувань зовнішнього середовища (можливі втручання держави в галузь, впливи науково-технічного прогресу, можливості та загрози з боку інших галузей). Отже, для того, щоб справитися з ризиками в умовах асиметрич-

ної інформації, консалтингова компанія має вирішити такі чотири задачі:

- 1) збір даних;
- 2) ідентифікація ризиків «ex-ante»;
- 3) ідентифікація ризиків «ex-post»;
- 4) розробка стратегії управління виявленими ризиками.

У залежності від галузі, в якій надається консультація консалтинговою компанією, кожна з вище перелічених задач може мати свої особливості, проте, на нашу думку, доцільно виділити їхні загальні характеристики.

Для першої задачі – збору даних (кількісних та якісних) консалтингові компанії можуть використовувати різні джерела: дані, що вже наявні у консалтингової компанії, дані, котрі є у наявності в компанії-клієнта та зовнішні джерела даних, у яких може бути інформація про ринок, споживачів тощо. Техніки збору даних теж можуть різнитися, проте, основними є інтерв'ю, анкети і опитування, спостереження, фокус-групи, історії та тематичні дослідження, документи, записи та бази даних [2].

Ці техніки можуть допомогти виявити відому вже інформацію, але в умовах асиметричної інформації вони є лише попереднім вказівником, де асиметричність інформації може з'явитися.

Наступними етапами є виявлення «ex-ante» та «ex-post» ризиків від асиметричності інформації, які також можуть різнитися в залежності від специфіки роботи консалтингової компанії. Такі ризики, можуть бути передбачені з допомогою експертів, які знають, де може виявлятися асиметричність інформації, а також з допомогою таких методів, як мозковий штурм, метод Дельфі, методика BERI тощо [3].

На етапі розроблення стратегії управління виявленими ризиками, на нашу думку, доцільно використовувати інструментарій економіко-математичного моделювання. Загальновідомі методи роботи з асиметрією в економіці: державне втручання, стандартизація роботи, вироблення репутації, надання гарантій та сигналінг [4; 5], дозволяють оцінювати асиметрію інформації тим економічним суб'єктам, що хочуть, щоб про них знали та зважали на їхні дії (продавці товарів, високопродуктивні працівники, держава, що регулює діяльність певних структур). У цьому випадку доцільно вико-

ристовувати методи математичного моделювання та статистичні методи машинного навчання з метою усунення проблем, які виникають за асиметричної інформації під час управління ризиками консалтингової діяльності – несприятливим відбором та ризиком недобросовісності [4; 6].

Під час використання статистичних методів важливо, щоб зібраної інформації було достатньо, вона була правдива та повна. Для того, щоб виконувалася перша умова – достатності інформації, консалтингова компанія може використовувати бази даних з необхідною інформацією від компанії-клієнта, власні або зовнішні бази даних. Необхідно, щоб дані були однотипними та структурованими, адже на їх основі буде побудовано математичну модель, яка б дозволила аналізувати наявні закономірності в даних та на їхній основі допомагати консалтинговій компанії боротися з асиметричністю інформації.

До зовнішніх джерел неструктурованих даних може належати також Всесвітня веб-мережа, де інформацію можна збирати з допомогою процедури веб-скрапінгу (*англ.* Web scraping – витягнення веб-даних).

Після збору необхідних даних консалтингова компанія може розпочати їхній аналіз. Для завдання аналізу даних однією з найбільш

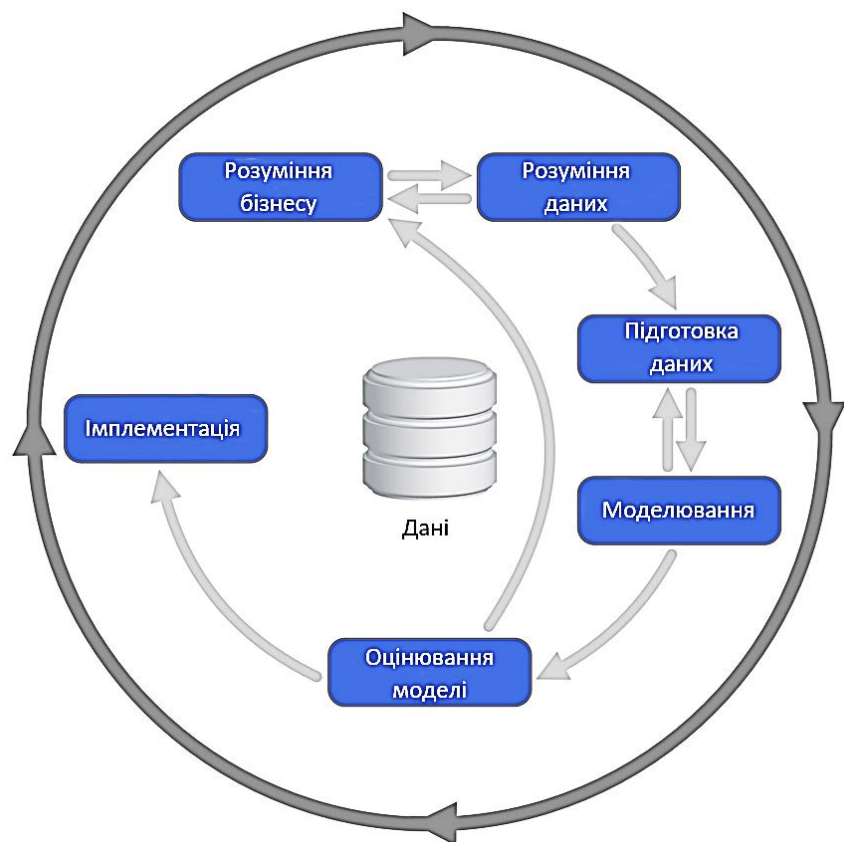


Рис. 1. Міжгалузевий Стандартний Процес Дослідження Даних CRISP-DM

Джерело: Shearer C., The CRISP-DM model: the new blueprint for data mining [8]

популярних методологій вважається Міжгалузевий Стандартний Процес Дослідження Даних CRISP-DM (*англ.* Cross-Industry Standard Process for Data Mining) [7], зображений на рис. 1

Методологія CRISP-DM включає 6 етапів, кожен з яких передбачає виконання набору дій, що ведуть до автоматизації прийняття рішень під час аналізу даних. Стандарт широко застосовується компаніями, які мають справу з великими об'ємами даних та хочуть автоматизувати прийняття рішень на їхній основі. Зокрема, такими суб'єктами можуть бути банки, страхові компанії, інтернет-магазини, рекламні компанії, що працюють у галузі інтернет маркетингу, соціальні мережі та інші сервіси, що можуть збирати та обробляти інформацію про своїх користувачів. Консалтингові компанії теж можуть використовувати стандарт CRISP-DM, який дозволив би справлятися з асиметрією інформації на ринку.

На першому етапі стандарту «Розуміння бізнесу» консалтинговій компанії потрібно з'ясувати особливості ведення бізнесу, які дані збирає клієнт, яку інформацію ще потрібно зібрати, щоб отримати оптимальний результат під час моделювання, під час прийняття яких рішень може виникати асиметричність.

На наступному етапі «Розуміння даних» консалтингова компанія повинна дослідити дані, що є в наявності, особливості збору та фіксації даних, простежити можливі взаємозв'язки та кореляції. У літературі цей підхід ще має назву розвідувальний та пояснювальний аналіз (*англ.* Exploratory and Explanatory Analysis) [8] та часто передбачає візуалізацію наявних даних для легшого їх сприйняття нетехнічними особами.

На третьому етапі «Підготовки даних» дані, що є наявними для аналізу потрібно очистити та доповнити. У літературі цей етап часто називають етапом попередньої обробки (*англ.* Data Preprocessing). Це важливий етап в моделюванні, адже інформація, яка буде використовуватися у математичній моделі, повинна мати цифровий вигляд та бути повною для кожного спостереження, що використовуватиметься під час моделювання. Основні активності, які можуть виконуватися на цьому етапі є такими [9]:

1) робота з пропущеними (відсутніми) змінними. Часто під час внесення інформації людиною у спеціалізовані програми інформація може вписуватися некоректно або не записуватися взагалі. Наприклад, повертаючись до ситуації з аналізом кредитних ризиків, оператор, що вносить дані позичальника в банківську програму для обслуговування клієнтів, може забути внести інформацію про вік боржника, у полі з поштовим індексом увести «00000», якщо програма не дозволяє залишити це поле пустим, а боржник не знає свого індексу тощо. У такому випадку, під час процедури

підготовки даних таку проблему можна вирішити кількома способами:

А) не враховувати спостереження з некоректними та пустими значеннями під час статистичного аналізу. Часто це не найкраща практика, адже що більше даних використовується для аналізу, то точнішою буде економіко-математична модель.

Б) у поля з пропущеними даними ввести відсутні дані. Часто вручну це зробити неможливо, адже незаповнених полів може бути дуже багато. Тому у поле з пропущеною змінною можуть вносити середнє значення, моду або медіану по групі, якій належить спостереження.

2) Поділ наявних даних на тренувальну та тестову вибірки, для того, щоб на тренувальній вибірці навчати математичну модель, а на тестовій вибірці перевірити її точність.

3) Приведення змінних до однієї шкали. Такі перетворення повинні виконуватися для математичних моделей та алгоритмів машинного навчання, які чутливі до великих різниць у значеннях між різними змінними. Щоб привести змінні до однієї шкали можна використати процеси нормалізації або стандартизації.

4) Проведення процедури регуляризації. Якщо на етапі оцінювання побудованої моделі було виявлено, що модель показує набагато кращі результати на тренувальній вибірці, ніж на тестовій, то це можливий індикатор можливого перенавчання моделі (*англ.* Overfitting).

Можливими шляхами вирішення цієї проблеми можуть бути збір більшої кількості тренувальних даних; уведення покарання за складність, використовуючи процедуру регуляризації; обрання простішої моделі, з меншою кількістю параметрів; зменшення розмірності даних.

Оскільки побудова вхідної вибірки більшої розмірності часто не є можливою, то одним з можливих вирішень проблеми перенавчання є використання регуляризації. Концепція регуляризації передбачає введення штрафів для надто великих ваг параметрів моделі. Загалом виділяють L1 регуляризацію [9], в якій вектор ваг w для кожної j -тої змінної записують так:

$$L1: \|w\|_1 = \sum_{j=1}^m |w_j| \quad (1)$$

та L2 регуляризацію, вектор ваг якої w записується так:

$$L2: \|w\|_2^2 = \sum_{j=1}^m w_j^2 \quad (2)$$

5) зменшення розмірності даних (*англ.* Dimensionality Reduction). У випадку, якщо існує багато змінних, які можна використати для побудови математичної моделі, то з метою зменшення ефекту перенавчання або для зменшення витрат комп'ютерного часу на підбір параметрів застосовують методи зменшення розмірності даних.

Існує два основні методи зменшення розмірності – відбір атрибутів (*англ.* Feature Selection) та виділення атрибутів (*англ.* Feature Extraction). Під час процедури відбору атрибутів обирають підмножину наявних змінних і на ній будують математичну модель. Під час використання методу виділення атрибутів з наявних змінних утворюється підмножина нових змінних.

Хоча існує багато методів відбору ознак [9; 10], які застосовують в аналізі текстів та природньої мови, розпізнаванні образів, аналізі генів тощо, для економіко-математичного моделювання доцільно використовувати алгоритм послідовного зворотного відбору (*англ.* Sequential Backward Selection), який дозволяє зменшити кількість змінних до наперед заданого рівня, відкидаючи по одному атрибуту з множини змінних та дивлячись, як це впливає на точність моделі. Також, щоб визначити важливість ознак, найважливіші з яких можна використовувати у моделі, використовують алгоритм побудови багатьох дерев рішень (*англ.* Random Forest), який надає оцінку важливості кожної змінної.

Для виділення атрибутів можуть вводити нові змінні, що базуються на вже наявних, наприклад, створювати відношення та коефіцієнти, брати квадратні корені або логарифми від змінних, а також застосовують методи аналізу головних компонент (*англ.* Principal Component Analysis – PCA), дискримінантний аналіз (*англ.* Linear Discriminant Analysis – LDA), ядерні методи аналізу головних компонент (*англ.* Kernel Principal Component Analysis) та інші [9].

Аналіз головних компонент є методом стиснення даних шляхом їхнього лінійного перетворення. Під час застосування цього методу враховується максимальна дисперсія у багатовимірних даних і проводиться їхнє проектування на нову підмножину даних з такою ж або меншою розмірністю [11]. Загальний вигляд даного методу зменшення розмірності передбачає побудову матриці W розмірністю $d \times k$, яка б дозволила перетворити вектор x на новий k вимірний простір ознак, що має менше ознак, ніж оригінальний d вимірний вектор змінних ($d > k$). У результаті повинен бути отриманий новий набір змінних z для кожного спостереження x . Математично це можна записати так:

$$x = [x_1, x_2, \dots, x_d], x \in \mathbb{R}^d, \quad (3)$$

$$\downarrow xW, W \in \mathbb{R}^{d \times k}, \quad (4)$$

$$z = [z_1, z_2, \dots, z_k], z \in \mathbb{R}^k. \quad (5)$$

Дискримінантний аналіз (LDA) подібний до методу PCA, алгоритм якого намагається знайти підмножину ознак, які б оптимізували розділення на класи. Цей алгоритм є методом навчання з вчителем, тобто йому відомо, які комбінації змінних утворюють певні класи [9]. Можна подумати що використання LDA краще працюватиме у задачах

класифікації, проте, вченими було показано, що використання зменшення розмірності без вчителя у PCA алгоритмі дає кращі результати у задачах класифікації ніж LDA метод [12].

Після підготовки даних переходять до етапу «моделювання». Загалом, у інформатиці, під час моделювання станів, запис алгоритму передбачення для яких не є можливим, застосовують методи машинного навчання, яких виділяють 3 види – навчання з вчителем, навчання без вчителя та навчання з підкріпленням. Для економіко-математичного моделювання, для передбачень майбутніх станів явищ як методу врахування асиметричності інформації в межах трьох вищезазначених методів можна виділити такі 4 основні типи задач: побудова регресії та класифікація (навчання з учителем), кластеризація (навчання без вчителя) та моделі обмеженої раціональності (навчання з підкріпленням) [9; 13].

Прикладами алгоритмів класифікації можуть бути: лінійна та логістична регресія, метод опорних векторів, метод k -найближчих сусідів тощо. Регресію пробують дослідити за допомогою методів лінійної регресії, нелінійної регресії, непараметричної регресії, Байєсівської лінійної регресії та інших. У кластерному аналізі застосовують методи k -середніх, ієрархічної кластеризації, алгоритмів OPTICS та DBSCAN тощо.

На наступному етапі циклу «оцінювання моделі» проводиться оцінювання якості моделі, її здатність до вірних передбачень. На цьому етапі потрібно обрати метрику, яка б відображала якість моделі, та дозволяла б порівнювати різні моделі між собою. У залежності від моделі, метриками можуть бути [11]:

Для задач регресії:

– сума квадратів залишків (*англ.* Residual Sum of Squares – RSS, або Sum of Squared Errors – SSE):

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (6)$$

де: y_i – реальне значення результуючого показника,

\hat{y}_i – теоретичне значення прогнозоване моделлю,

n – кількість спостережень.

– стандартна похибка (*англ.* Residual Standard Error – RSE):

$$RSE = \sqrt{\frac{1}{n-2} RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (7)$$

– коефіцієнт детермінації R^2 :

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}, \quad (8)$$

де: $TSS = \sum_{i=1}^n (y_i - \bar{y})^2,$ (9)

\bar{y} – середнє значення результуючого показника.

Також для задач регресії обчислюють середньоквадратичну похибку, логарифмічну середньоквадратичну похибку, медіанну абсолютну похибку тощо.

Для задач класифікації для оцінювання якості моделі можуть бути взяті значення обчислені з таблиці помилок (англ. Confusion Matrix), яка для задач бінарної класифікації має вигляд матриці 2×2, в якій наведені вірно класифіковані та помилково класифіковані значення (рис. 2) [14].

Для надання загальної інформації про те, яка кількість спостережень є класифікованою неправильно використовують коефіцієнт помилок:

$$ERR = \frac{FP + FN}{FP + FN + TP + TN} \quad (10)$$

Для показу якості класифікатора загалом, використовують метрику точності (англ. Accuracy):

$$ACC = \frac{TP + TN}{FP + FN + TP + TN} = 1 - ERR \quad (11)$$

Для специфічних задач, де присутні незбалансовані дані, використовують метрики [9, 14]:

– відгуку / чутливості (англ. Recall/Sensitivity):

$$REC = \frac{TP}{P} = \frac{TP}{FN + TP} \quad (12)$$

– специфічності (англ. Specificity), яка ще вказує на помилку першого роду:

$$SPC = \frac{TN}{N} = \frac{TN}{FP + TN} \quad (13)$$

– точності (англ. Precision):

$$PRE = \frac{TP}{TP + FP} \quad (14)$$

Також обчислюють коефіцієнт похибки першого роду, яку ще називають помилковим спрацюванням класифікатора (англ. False Positive Rate), та при перевірці статистичних гіпотез можуть позначати α :

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC \quad (15)$$

Та похибку другого роду, яку позначають β та ще іноді називають пропуском події (англ. False Negative Rate):

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - REC \quad (16)$$

На основі похибки другого роду обчислюють «потужність критерію» (англ. Power), яка, по суті, є метрикою чутливості:

$$Power = 1 - \beta = REC \quad (17)$$

Проте, досить часто використовуються комбінації з вищенаведених метрик, прикладом тут може бути значення індексу F1 (англ. F1-score):

$$F1 = 2 \frac{PRE \times REC}{PRE + REC} \quad (18)$$

Також для бінарних класифікацій будують ROC – криві (англ. Receiver Operating Characteristic), що є співвідношенням між часткою спостережень правильно класифікованих ознак серед усіх спостережень, що мають прогнозовану ознаку – TPR та часткою спостережень, що були помилково класифіковані – FPR [11]. Коефіцієнт TPR обчислюють:

$$TPR = \frac{TP}{P} = \frac{TP}{FN + TP} \quad (19)$$

Для кількісної оцінки ROC – кривих обчислюють показник AUC (англ. Area Under ROC Curve), що є площею під кривою ROC.

Для задач мультикласової класифікації метрики для оцінювання якості моделей можуть використовуватися методи мікро- та макро-усереднення [9].

За методом мікро-усереднення якість моделі обчислюють, як відношення кількості вірно класифікованих представників кожного класу до кількості вірно класифікованих та помилково класифікованих представників цього класу:

$$PRE_{macro} = \frac{PRE_1 + \dots + PRE_k}{k} \quad (20)$$

Цей метод дозволяє надати ваги кожному класу, в залежності від того, скільки спостережень у ньому було.

За методом макро-усереднення беруть середнє значення точності для кожного класу:

$$PRE_{macro} = \frac{PRE_1 + \dots + PRE_k}{k} \quad (21)$$

де: k – кількість прогнозованих класів.

Даний підхід передбачає, що ваги кожного класу в моделі є рівними, що дозволяє оцінити загальну якість моделі для усіх класів.

Для оцінювання моделей класифікації також використовуються такі методи, як побудова кривих

		Фактичний клас	
		Позитивне значення класу (P)	Негативне значення класу (N)
Прогнозований клас	Позитивне передбачення класу	Вірно передбачений клас (TP)	Невірно позитивне значення (FP)
	Негативне передбачення класу	Невірно негативне значення (FN)	Вірно знехтуваний клас (TN)

Рис. 2. Таблиця помилок для бінарної класифікації

Джерело: узагальнено автором

Колмогорова – Смірнова, розрахунок коефіцієнта Джині, коефіцієнт подібності Жаккара та інші [15].

Для задач кластеризації неможливо точно вказати на якість проведеної процедури формування кластерів, адже кількість кластерів наперед невідома і може обиратися або експертним методом, або на основі кількісних показників певних параметрів кластеризації. Проте, можна виділити такі види показників [16]:

Внутрішні індекси – використовують для оцінювання якості алгоритму кластеризації без врахування зовнішньої інформації. Часто ці показники опираються на різноманітні відношення відстані між точками всередині кластерів та відстанями між різними кластерами. Сюди входять такі показники, як байєсівський інформаційний критерій, індекс Калінські – Харабаша, Дейвіс – Боулдін індекс, індекс силуету, Данн індекс та інші.

Зовнішні індекси – використовують для оцінювання якості алгоритму кластеризації, коли дані про те, якому кластеру яка точка повинна відповідати будуть відомі після проведення кластеризації. Тут можуть використовуватися методи оцінювання якості моделей, що застосовують під час класифікації, а також такі показники, як індекс взаємної інформаційності, індекс чистоти та показник ентропії.

Відносні індекси – індекси, що використовують для порівняння двох різних алгоритмів кластеризації. Тут для оцінювання можуть застосовуватися як зовнішні так і внутрішні індекси.

Опісля проведення етапу оцінювання моделей, за умови отримання задовільних результатів, консалтингові компанії можуть застосовувати їх у своїй діяльності для вирішення конкретних бізнес задач в умовах асиметричної інформації. Етап «імплементції» моделей, як правило, передбачає застосування побудованих алгоритмів на даних клієнта, із застосування обчислень на комп'ютерах або серверах та збір статистики про якість моделей у реальному житті та постійне їхнє навчання на нових даних.

Висновки з проведеного дослідження.

Основними типами задач, що можуть бути вирішені консалтинговими компаніями за допомогою економіко-математичного моделювання є проведення класифікації, побудова регресії, кластерний аналіз та побудова моделей обмеженої раціональності. Проте, дані, що використовуються для побудови моделей, можуть бути не повними або спотвореними, тому перед етапом моделювання доцільно здійснити процедуру очистки та доповнення даних. Варто зазначити, що для управління ризиками в умовах нерівномірно розподіленої інформації, консалтингові компанії повинні визначити джерела даних, в яких асиметричність інформації може виникати, та чітко

сформулювати завдання від бізнесу, враховуючи його особливості.

Застосування методології дослідження даних CRISP-DM та математичних методів аналізу та обробки інформації дозволить консалтинговим компаніям надавати якісніші послуги своїм клієнтам та керувати ризиками, що виникають при роботі в середовищі з нерівномірно розподіленою інформацією.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Луцишин З., «Moral Hazard» зростання зовнішньої заборгованості України в умовах доларизації економіки / З. Луцишин, Т. Цвігун. – Міжнародна економічна політика. – 2012. – Вип. 1–2. – С. 83–106.
2. Guidelines for the routine collection of capture fishery data. – FAO Fisheries Technical Paper. No. 382. – Rome, FAO. 1999. – P. 113.
3. Донець Л., Економічні ризики та методи їх вимірювання: Навчальний посібник. – К.: Центр навчальної літератури. – 2006. – 312 с.
4. Akerlof G., The Market for «Lemons»: Quality Uncertainty and the Market Mechanism. – The Quarterly Journal of Economics, v. 84. – 1970. – Pp. 488–500.
5. Spence M., Job Market Signaling. – The Quarterly Journal of Economics, v. 87. – 1973. – Pp. 355–374.
6. Stiglitz J., Arnott R., The Basic Analytics of Moral Hazard. – National bureau of economic research, working paper # 2484. – 1988. – P. 56.
7. Shearer C., The CRISP-DM model: the new blueprint for data mining, Journal of Data Warehousing, v. 5. – 2000. – Pp. 13–22.
8. Nussbaumer K., Storytelling with Data: A Data Visualization Guide for Business Professionals. – John Wiley & Sons. – 2015. – P. 288.
9. Raschka S., Python Machine Learning. – Packt publishing, Birmingham, UK. – 2015. – P. 454.
10. Hastie T., The Elements of Statistical Learning. Data Mining, Inference, and Prediction. – Springer, 2nd edition. – 2016. – P. 745.
11. Gareth J., An Introduction to Statistical Learning with Applications in R. – Springer. – 2013. – P. 426.
12. Martínez A., PCA versus LDA. – Journal IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 23. – 2001. – Pp. 228–233.
13. Barak A., A decision support method, based on bounded rationality concepts, to reveal feature saliency in clustering problems. – Decision Support Systems, v. 54. – 2012. – Pp. 292–303.
14. Haibo He, Learning from Imbalanced Data. – IEEE Transactions on knowledge and data engineering, v. 21. – 2009. – Pp. 1263–1284.
15. Řezáč, M., How to Measure the Quality of Credit Scoring Models. – Czech Journal of Economics and Finance, v. 61. – 2011. – Pp. 486–507.
16. Rendon E., A comparison of internal and external cluster validation indexes. – In Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications, v. 11. – 2011. – Pp. 158–163.

REFERENCES:

1. Lutsyshyn Z. (2012), «Moral Hazard» zrostantia zovnishnoi zaborhovanosti Ukrainy v umovakh dolaryzatsii ekonomiky [«Moral Hazard» growth of Ukraine's external debt in the conditions of dollarization of the economy. *International economic policy*, v. 1–2, pp. 83–106.
2. FAO Fisheries Technical Paper (1999). Guidelines for the routine collection of capture fishery data. FAO Fisheries Technical Paper, No. 382.
3. Donets L. (2006), *Ekonomichni ryzyky ta metody yikh vymiriuvannia* [Economic risks and methods of their measurement]. Kyiv: Center for Educational Literature (in Ukrainian).
4. Akerlof G. (1970), «The Market for «Lemons»: Quality Uncertainty and the Market Mechanism», *The Quarterly Journal of Economics*, vol. 84, pp. 488–500.
5. Spence M. (1973), «Job Market Signaling», *The Quarterly Journal of Economics*, vol. 87, pp. 355–374.
6. Stiglitz J., Arnott R. (1988), «The Basic Analytics of Moral Hazard», National bureau of economic research, working paper No. 2484, pp. 1–56.
7. Shearer C. (2000), The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing*, v. 5, pp. 13–22.
8. Nussbaumer K. (2015), *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons.
9. Raschka S. (2015), *Python Machine Learning*. Packt publishing, Birmingham, UK.
10. Hastie T. (2016), *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer-Verlag New York, 2nd edition.
11. Gareth J. (2013), *An Introduction to Statistical Learning with Applications in R*. Springer-Verlag New York.
12. Martínez A. (2001), PCA versus LDA. *Journal IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, pp. 228–233.
13. Barak A. (2012), A decision support method, based on bounded rationality concepts, to reveal feature saliency in clustering problems. *Decision Support Systems*, v. 54, pp. 292–303.
14. Haibo He (2009), Learning from Imbalanced Data. *IEEE Transactions on knowledge and data engineering*, v. 21, pp. 1263–1284.
15. Řezáč M. (2011), How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance*, v. 61, pp. 486–507.
16. Rendon E. (2011), A comparison of internal and external cluster validation indexes. *Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications*, v. 11, pp. 158–163.

Bachalo I.B.

Postgraduate Student at Department of Economic Cybernetics
Ivan Franko National University of Lviv

THE USE OF ECONOMIC-MATHEMATICAL MODELLING IN CONSULTING ACTIVITIES IN CONDITIONS OF ASYMMETRIC INFORMATION

In the article, the author describes using economic-mathematical modelling by consulting companies in the conditions of asymmetric information.

Every day consulting companies face the situation of collecting and analysing information with the purpose to make an optimal decision for their clients and, if it needs, to implement them in the real life. However, it can happen that information is distributed between economic agents uniformly, so it can lead to the appearance of new risks associated with asymmetric information. For instance, the suppliers or customers of consulting company's client can have important information, which does not have the company-client and they can hide or distort this information. This situation produces the problem of market inefficiency when the client has to spend time and money on additional data collection and analysis and it reflects on a price of its product in the future. In some cases, the information asymmetry can be the reason for the market crash.

Whereas different researches propose to deal with unequal information distribution with «Signalling» and «Screening» methods, the author describes the application of economic-mathematical modelling to solving the problem of information asymmetry. The key technique that can be used by consulting companies is Cross-Industry Standard Process for Data Mining (CRISP-DM). The article describes the key steps of the methodology and demonstrates theoretical concepts of data preprocessing for mathematical models and methods of models' evaluation. But, for the effective use of the proposed technique with a purpose of managing company risks in conditions of asymmetric information, consulting companies should identify sources of information asymmetry and clearly formulate tasks from a business.