

УДК 519.226

П.І. Бідюк, Н.В. Кузнецова,
О.М. Терентьев

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ДАНИХ

Вступ

Установи, підприємства, банки і компанії, які виробляють або постачають певну продукцію чи послуги, тісно пов'язані з такими фінансовими процесами, як товарні розрахунки, оплата послуг, сплата податків, ведення рахунків, нарахування заробітної плати тощо. Для стійкого цілеспрямованого розвитку підприємства необхідно раціонально і завчасно планувати стратегії його розвитку, а також коректно розподіляти грошові потоки [1].

Ефективність управлінських рішень, пов'язаних із визначенням ціни продукту або розміру партії сировини, поставками продукції, заміною обладнання, впровадженням нової технології, наймом або звільненням працівників, має бути оцінена теоретично з точки зору загального успіху фірми, характеру її економічного та фінансового зростання в цілому. Якщо компанія має невеликий товарообіг і відповідно невеликі обсяги фінансових ресурсів, то вона не витрачає значні кошти на залучення спеціальних автоматизованих комп'ютерних систем для планування і аналізу фінансових процесів, а здійснює аналіз силами своїх фінансистів, бухгалтерів, директорів та інших спеціалістів. Потужні компанії, як правило, співпрацюють із великою кількістю партнерів і реалізують широку номенклатуру товарів. Вони мають великий щомісячний товарообіг і значні обсяги продажів. Такі компанії не можуть успішно працювати на ринку без серйозного інструментарію для збору, обробки та аналізу всієї вхідної інформації, а також коректного планування роботи окремих підрозділів чи розвитку компанії в цілому. Вони використовують різноманітні програмні пакети – Oracle, SAP, Nielsen, для збору і зберігання певної інформації, розробляють власні бази даних (БД), власні "надбудови" до існуючих систем, залучають інші компанії для проведення маркетингових чи соціологічних досліджень, рекламистів та піарщиків для забезпечення успішного просування компанії на ринку. Всі заходи і дії, які здійснюють компанії для свого успішного розвитку,

повинні бути вчасно реалізовані та спрямовані на просування конкретної продукції в поточний момент часу, а тому їх необхідність і ефективність слід попередньо проаналізувати з максимальним використанням відповідного сучасного математичного інструментарію.

Постановка задачі

Завданням статті є розробка імовірнісних моделей та моделей інших типів з метою створення узагальненої технології аналізу фінансово-економічних даних за допомогою інформаційної системи підтримки прийняття рішень (ІСППР), яка ґрунтується на оригінальному інтегрованому підході до її проектування. За допомогою створюваної ІСППР повинна розв'язуватись задача прогнозування розвитку фінансово-економічних процесів на фінансовому підприємстві, а також мають виконуватись аналіз можливості застосування запропонованої інформаційної технології обробки даних до аналізу і прогнозування об'ємів продажів компанії та оцінюватись (спрогнозуватись) платоспроможність позичальників кредитів банку.

Фінансовий аналіз підприємства

Основними функціями аналізу фінансових даних є об'єктивна оцінка фінансового стану підприємства, фінансових результатів його діяльності, ефективності та ділової активності суб'єкта аналізу і виявлення чинників та причин досягнення наявного стану і одержання результатів, а також підготовка і обґрунтування прийняття управлінських рішень у галузі фінансів.

Фінансовий аналіз здійснюється за допомогою множини математичних і статистичних моделей різних типів, які дають змогу структурувати та ідентифікувати причинні взаємозв'язки між основними процесами на підприємстві. Існує три основних типів моделей, які застосовують для виконання аналізу фінансового стану підприємства: *дескриптивні*, *предикативні* та *нормативні*.

Основа аналізу становлять дескриптивні моделі, які ґрунтуються на інформації з бухгалтерської звітності. До них належать: побудова системи звітних балансів; подання фінансової звітності в різних аналітичних розрізах; вертикальний та горизонтальний аналіз звітності; система аналітичних коефіцієнтів; аналітичні записки до звітності.

Предикативні моделі – це математичні і статистичні моделі для передбачення і обчислення оцінок прогнозів розвитку процесів. Вони використовуються для прогнозування доходів і прибутків підприємства та його майбутнього фінансового стану. Найбільш поширеними з них є розрахунки точки критичного обсягу продажів, побудова прогностичних фінансових звітів, моделі динамічного аналізу (жорстко детерміновані факторні, регресійні моделі).

Нормативні моделі – це такі моделі, які уможливають порівняння фактичних результатів діяльності підприємства з нормативними (розрахованими на підставі нормативу). Вони використовуються, як правило, для виконання внутрішнього фінансового аналізу. Їх суть полягає у встановленні нормативів на кожну статтю витрат стосовно технологічних процесів, видів виробів, а також при розгляді і з'ясуванні причин відхилень фактичних даних від цих нормативів.

Таким чином, у процесі аналізу фінансового стану підприємства можна використовувати найрізноманітніші прийоми, методи та моделі аналізу. Їх кількість і широта охоплення діяльності підприємства залежать від конкретних цілей аналізу та визначаються його завданнями в кожному випадку аналізу вибраного підприємства.

Фінансовий аналіз у сучасних умовах стає елементом керування. Фактично це єдиний інструмент оцінювання надійності будь-якого потенційного партнера (покупця, постачальника чи позичальника), оскільки взяті окремо один від одного числові дані бухгалтерської звітності не дають змоги отримати цілісну картину фінансового стану підприємства.

Фінансовий аналіз у системі керування фінансами господарюючого суб'єкта в найбільш загальному вигляді являє собою спосіб накопичення, трансформування та використання інформації фінансового характеру, що має за мету: 1) оцінити поточний і перспективний майновий і фінансовий стан господарюючого суб'єкта, в тому числі ризик його неплатоспроможності або банкрутства; 2) оцінити можливі і доцільні темпи розвитку господарюючого суб'єкта з позиції їх фінансового забезпечення; 3) виявити доступні джерела коштів і оцінити можливість та доцільність їх

мобілізації; 4) спрогнозувати стан господарюючого суб'єкта на ринку капіталу.

Організація збору фінансових статистичних даних підприємства

У процесі функціонування всередині компанії збирається і ведеться велика кількість різноманітної документації по підрозділах, департаментах, відділах і т.п. Це статистичні дані стосовно продажів, рахунки, звіти, договори з підрядчиками, оренди, послуг тощо. Для ефективної роботи компанії необхідно ретельно планувати її дії та виконувати спрямований (бажано оптимальний) розподіл ресурсів.

Нижче наведено загальну структурну схему інформаційної технології аналізу даних і прогнозування певних показників (рис. 1). На основі цієї схеми можна зобразити весь процес збору і попередньої обробки вхідної інформації та її використання для подальшого аналізу і прийняття рішень.

Будь-яка компанія, що працює на ринку, функціонує в рамках встановлених правил і користується офіційно встановленими документами, рейтингами, оцінками українських та міжнародних компаній. Наприклад, це – рейтинги компаній, продукції, міжнародні ціни на певні продукції, курси валют, нормативні акти,

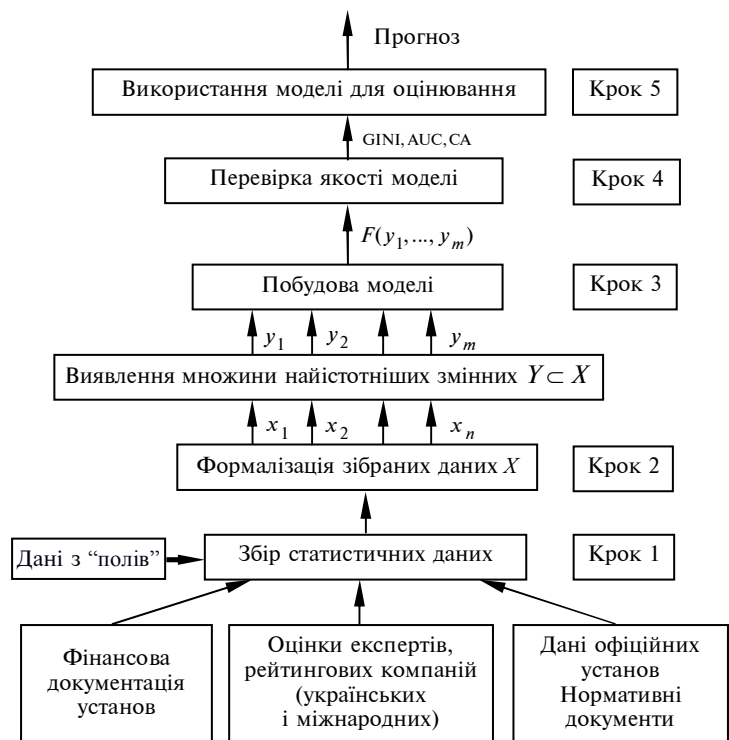


Рис. 1. Технологія аналізу даних і прогнозування в ІСППР

державні стандарти, вимоги, згідно з якими контролюється продукція, власні фінансові документи в зручному для компанії вигляді та форматі. Паралельно компаніями збирається так звана інформація стосовно "*первинних продажів*", тобто інформація, яка відображає обсяги продукції, законтрактовані та відвантажені дистриб'юторам, торговельним представникам, чи напрямок торговельним мережам для подальшої реалізації продукції споживачам. Слід зазначити, що складність і нагромадження інформації залежать від того, наскільки великою є деяка компанія і наскільки велику частку ринку вона займає. Первинні продажі компанії дистриб'юторам або "*ключовим клієнтам*" (великим торговельним мережам, які мають контракти на реалізацію напрямку, – типу Metro, Ашан тощо) подаються у вигляді великих упаковок, контейнерів тощо, які найчастіше закуповуються на місяць або більше і для всіх магазинів одночасно, тобто це сукупна оптова закупівля товару, яка за контрактами може містити певне навантаження і певні бонуси, і вона є обов'язковою закупівлею певної кількості контейнерів та певних видів товарів. За виконання цих обов'язкових умов контракту можуть нараховуватись певні бонуси, подарунки, дисконти тощо.

Ще однією градацією є інформація стосовно тих, кому здійснюється продаж товарів. Це можуть бути "*ключові клієнти*", великі дистриб'ютори, що цілком покривають певні регіони (кілька областей або навіть певну частину країни) і реалізують товари через оптові та роздрібні торговельні мережі.

Уся наведена статистична інформація обов'язково збирається поряд з *інформацією з "полів"*. Інформацією з "*полів*" називають інформацію, що надана дистриб'юторами, торговельними представниками, торговельними мережами, ринками щодо обсягів реалізованої продукції на вторинних ринках. Слід зазначити, що дані з "*полів*" цікаві для аналізу, оскільки саме вони містять у собі інформацію стосовно реального попиту та уподобання споживачів продукції. Ця інформація збирається кожного тижня і зводиться в єдину базу за регіонами таким чином, щоб наприкінці кожного місяця можна було зробити висновки стосовно виконання плану продажів та коригувати оцінки прогнозів на наступні періоди. Дана інформація має надходити централізовано в базу даних або в єдиний файл в однотипному вигляді та

бути перевіреною і проконтрольованою на місцях. Оскільки в даному випадку дистриб'ютори являють собою зацікавлену сторону, то таку інформацію необхідно контролювати з боку представників компанії на місцях (у магазинах, торговельних мережах). Ідеальною є організація такого збору інформації, щоб інформація надавалася з одного джерела і була вже уточненою і скоригованою, а не збиралася окремо дистриб'юторами та окремо представниками компаній, а потім довгий час не збігалася, що потребувало б уточнень і, відповідно, більших витрат часу. У наданій інформації мають міститися окремі поля – тип продукції, категорія товару, обсяг, місяць, тиждень, регіон, в якому вона реалізовувалася, назва дистриб'ютора тощо. Для внутрішнього фінансового аналізу обсяг реалізованої продукції може бути поданий у грошовому вигляді – гривні або іноземній валюті для формування звітів компанії, встановлення рентабельності підприємства та коригування в разі потреби цінової політики на певні види продукції. Для ретельного і коректного аналізу при економічній нестабільності і різкому коливанні курсу валют всередині країни та в світі є сенс разом з цим подавати обсяг реалізованої продукції ще й в ящиках (контейнерах) чи упаковках. Це дасть змогу коректно порівнювати обсяги продажів у різні періоди. Названа інформація є цілком достатньою для поверхневого аналізу діяльності компанії, але вона не дає глибокого розуміння причин і факторів, які впливають на коливання попиту товару, не дає інформації щодо покриття регіону, кількості товару на душу населення та не враховує регіональні і демографічні фактори. Для адекватного аналізу поряд із встановленими показниками і даними слід збирати інформацію стосовно чисельності населення в даному регіоні, рівня продажів (в ідеалі) або кількості видів конкурентних товарів на полиці тощо. Це дасть можливість не лише прогнозувати обсяги продажів на наступні періоди, а й аналізувати, наскільки успішно працює дистриб'ютор у даному регіоні, наскільки ефективно працює регіональна маркетингова компанія, наскільки відомою є компанія в даному регіоні тощо. У результаті такого аналізу аналітики отримують можливість дати розгорнуте пояснення щодо наявності прихованих резервів стосовно насиченості даною продукцією (або навіть певним видом), необхідності зміни маркетингової кампанії тощо.

Аналіз фінансових даних підприємства і прогнозування фінансових показників

Для поглибленого послідовного аналізу фінансової діяльності компанії і прогнозування обсягів продажів у наступні періоди необхідно визначити множину найістотніших характеристик, які слід ввести в модель. Скористаємось для цього інтегрованим підходом на основі мереж Байєса [1, 2], який передбачає використання кількох методів аналізу даних: на першому етапі – мереж Байєса, а на другому – дерев рішень або логістичної регресії.

На етапі збору статистичних даних для будь-якої задачі зазвичай *не відомі самі істотні фактори*, які спричиняють значний вплив на остаточний результат, а тому фінансові установи намагаються зібрати максимально розширену інформацію. Маючи великий набір даних щодо певної події, необхідно побудувати модель, яка *визначить вплив факторів на результат*. Слід формалізувати отримані дані і виявити, які з них є істотними. З цією метою пропонується *побудувати мережу Байєса, яка встановить причинно-наслідкові зв'язки між змінними, що відповідають факторам, силу зв'язків між вибраними змінними, а також дасть можливість виявити змінні, які взагалі не пов'язані з*

результуючою подією (“вісячі змінні”). На основі побудованої мережі і встановлених зв'язків можна істотно скоротити кількість факторів, які будуть включені на наступному етапі при побудові моделі. Процедура аналізу даних за допомогою запропонованого інтегрованого підходу наведено на рис. 2.

На основі даного підходу побудовано такі моделі: *інтегровану модель на основі мережі Байєса та дерева рішень (ІМБД)*, а також *інтегровану модель на основі логістичної регресії та мережі Байєса (ІМЛБ)*.

Пунктиром відображено модуль, який не обов'язковий у даній схемі, тобто якщо існує можливість побудови кількох інтегрованих моделей, то він також функціонує, а якщо такої можливості немає або завчасно відомо, яка інтегрована модель краще підходить для конкретних даних, то може застосовуватись лише один модуль цієї схеми.

Необхідно зазначити, що інтегрований підхід передбачає можливість застосування методів у зворотному порядку, тобто спочатку використовується певний метод для виявлення істотних факторів (наприклад, логістична регресія або дерево рішень [3]), а потім для побудови моделі використовується мережа Байєса.

Покращення результатів прогнозування за допомогою інтегрованого підходу пояснюється таким чином: вибір факторів, що піддаються аналізу, не випадковий, а виконується на основі встановлених причинно-наслідкових зв'язків та певних характеристик (статистика Крамера, статистика χ^2 , інформативність IV (informational value)). При цьому кількість факторів, яку необхідно вводити в модель, значно зменшується, що дає можливість істотно знизити обчислювальні витрати на реалізацію методу. З іншого боку, кращу якість прогнозу можна пояснити формально через аналіз похибок оцінок прогнозу.

Для двох методів прогнозування середня оцінка визначається як

$$\hat{y}_c(k) = \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2}, \quad (1)$$

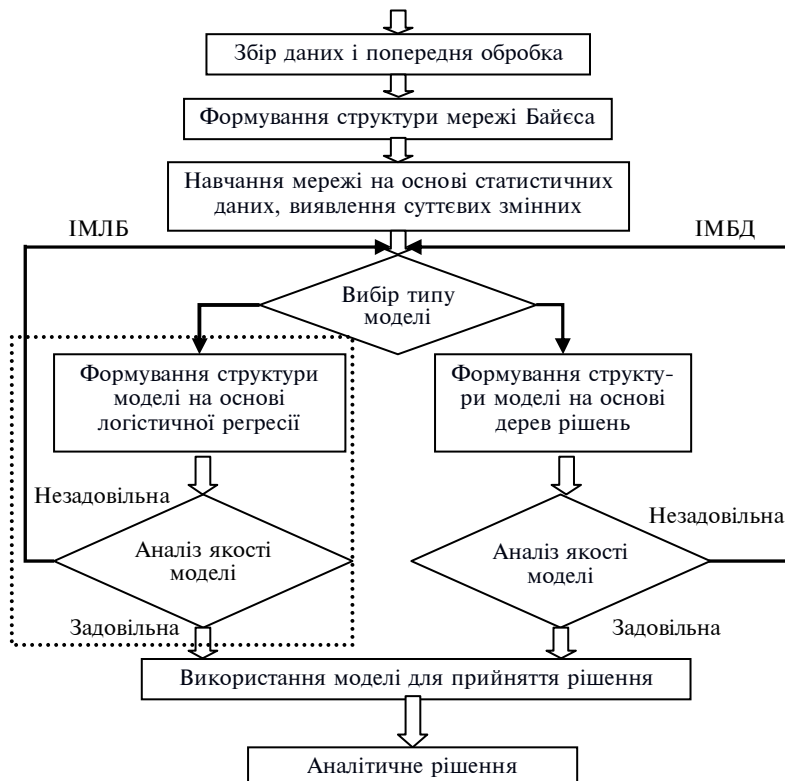


Рис. 2. Процедура інтегрованого підходу

де $\hat{y}_c(k)$ – комбінований прогноз; $\hat{y}_1(k), \hat{y}_2(k)$ – прогнози, отримані за двома різними методами. Якщо окремі прогнози не зміщені (це повинен забезпечувати метод прогнозування), то комбінований прогноз також буде незміщеним. Похибка комбінованого прогнозу обчислюється за формулою

$$e_c(k) = y(k) - \hat{y}_c(k) = y(k) - \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2} = \frac{e_1(k) + e_2(k)}{2}, \quad (2)$$

де $y(k)$ – фактичне значення прогнозованої змінної.

Дисперсія похибки комбінованого прогнозу має вигляд

$$\begin{aligned} \text{var} \left[\frac{e_1(k) + e_2(k)}{2} \right] &= E \left[\frac{e_1(k) + e_2(k)}{2} \right]^2 = \\ &= \frac{1}{4} E \left[e_1^2(k) + 2e_1(k)e_2(k) + e_2^2(k) \right] = \\ &= \frac{1}{4} \{ E[e_1^2(k)] + 2E[e_1(k)e_2(k)] + E[e_2^2(k)] \} = \\ &= \frac{1}{4} \left[\sigma_1^2 + 2 \frac{E[e_1(k)e_2(k)]}{\sigma_1\sigma_2} \sigma_1\sigma_2 + \sigma_2^2 \right] = \\ &= \frac{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}{4}. \end{aligned}$$

Таким чином, дисперсія комбінованого прогнозу обчислюється за виразом

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4}, \quad (3)$$

де ρ – коефіцієнт кореляції між похибками прогнозу. Якщо похибки прогнозування за двома моделями (методами) незалежні, тобто $\rho = 0$, то формула (3) спрощується:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2}{4}. \quad (4)$$

Якщо дисперсії близькі за значеннями і похибки оцінок прогнозів незалежні, то дисперсія комбінованої похибки буде значно меншою будь-якої з двох дисперсій. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$. Тоді

$$\sigma_c^2 = \frac{100 + 100}{4} = 50.$$

Однак навіть при існуванні досить високої кореляції між похибками прогнозування дисперсія похибки комбінованого прогнозу буде меншою, ніж дисперсія кожного методу окремо. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$ і $\rho = 0,8$. Тоді

$$\begin{aligned} \sigma_c^2 &= \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \\ &= \frac{100 + 100 + 2 \cdot 0,8 \cdot 10 \cdot 10}{4} = 90, \end{aligned}$$

тобто навіть у цій ситуації спостерігається зменшення дисперсії похибки прогнозу після усереднення оцінок, отриманих за двома методами.

Таким чином, інтегрована оцінка об'єднання випадкових величин має меншу дисперсію похибки порівняно з дисперсією похибок окремих оцінок. Іншими словами, похибка прогнозу на основі інтегрованого методу буде меншою порівняно з окремими похибками цих методів.

Розглянемо на прикладі можливість застосування інтегрованого підходу для аналізу фінансових показників підприємства.

Приклад задачі прогнозування продажів на наступні періоди. Нехай компанія з іноземним капіталом збирає статистику щодо рівня вторинних продажів кожного тижня і місяця по окремих регіонах, дистриб'юторах та видах товарів. Необхідно встановити ймовірність зростання продажів і спрогнозувати обсяг продажів у наступні періоди, а також оцінити, наскільки охоплений конкретний регіон продажами і чи добре працює дистриб'ютор у даному регіоні.

Задача описується такими змінними: назва регіону, дистриб'ютора або ключового клієнта, що працює в регіоні, тип (категорія) продукту, який продається, результат прогнозу – збільшення чи зменшення продажів у наступному місяці.

Іноземна компанія не має власного виробництва в Україні, і товар доводиться завозити з іншої країни. Тому якісне прогнозування обсягів продажів у наступні періоди дасть можливість правильно організувати логістику, тобто вчасне завезення товару в необхідних обсягах задля того, щоб увесь товар реалізувався і не було необхідності його десь зберігати.

У рамках загальної інформаційної системи, що створюється, треба виконати проектування і реалізацію бази даних для підприємства як основного сховища для зберігання і обробки статистичної інформації стосовно кількісних і вартісних показників бізнес-діяльності.

Таблиця 1. Приклад структуризації вхідних даних

Тип продукту	Місяць	Категорія продукту	Регіон	Назва дистриб'ютора	Обсяг продажів (в ящиках)	Обсяг продажів (у доларах США)	Прогноз на наступний місяць
Тип 1	Січень	Бренд 1	Київ	Дистриб'ютор 1	Зростання (1)
Тип 2	Лютий	Бренд 2	Черкаська область	Дистриб'ютор 2	Падіння (0)
...

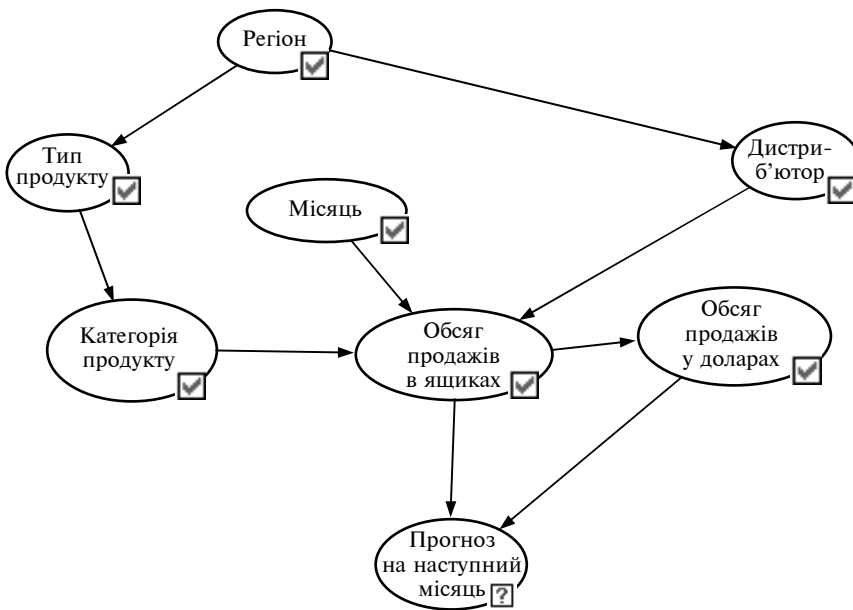


Рис. 3. Структура мережі Байєса для задачі аналізу продажів компанії

Основні типи операцій, які виконуються під час бізнес-діяльності і мають знайти відображення в автоматизованій системі, — це операції щодо закупівлі товару в постачальників, реалізації товару клієнтам, ведення обліку товарів, статистичні підрахунки і оформлення звітної документації (в тому числі у вигляді товарних чеків тощо, а також у вигляді звітів за певний період щодо результатів фінансової діяльності підприємства). Інформація, яка використовується під час діяльності, має бути збережена з можливістю її подальшого застосування в рамках аналізу чи видачі певних даних за звітний період [4].

Статистичні дані стосовно рівня продажів збираються в одиницях товару (контейнерах, упаковках) і грошових одиницях (доларах США, гривнях). Для об'єктивності, враховуючи коливання курсу національної валюти в Україні, слід оцінювати рівні продажів кількістю контейнерів або доларів США.

Вхідні дані, які використовуються для аналізу, зводяться до вигляду, наведеного в табл. 1.

Враховуючи статистичні дані, за допомогою інтегрованого підходу відбираємо істотні змінні, які застосовуємо для побудови моделі на основі мережі Байєса. Структуру побудованої мережі Байєса показано на рис. 3. Побудована структура мережі Байєса дає можливість визначити ймовірність зростання продажів у наступні періоди на основі вхідних даних.

Очевидно, що для стратегічного планування компанії слід збирати статистичні дані продажу товарів разом із широтою охоплення конкретним брендом регіонів України, щоб надалі ретельніше планувати маркетингову стратегію, враховуючи чисельність населення в

регіоні, співвідношення чоловічого та жіночого населення, середній рівень заробітної плати в регіоні тощо. Це сприятиме визначенню зв'язків і впливів між цими змінними і орієнтуванню продажу та акції на кінцевого споживача.

Архітектура системи підтримки прийняття рішень для аналізу фінансових даних підприємства

Для автоматизації процесу аналізу фінансових показників різних компаній, установ і підприємств необхідно створювати систему підтримки прийняття рішень, яка дасть можливість не лише отримувати і зберігати інформацію, а й аналізувати вхідні дані та подавати результати в певному вигляді.

Основні вимоги до інформаційної технології та інформаційної СППР такі:

- *адекватність* поставленій задачі;
- *своєчасність* задоволення інформаційних і обчислювальних потреб при проведенні аналізу фінансової діяльності;

- *мінімальний час відгуку* на аналітичні запити;
- можливість *зручного подання* вихідної інформації (наприклад, у табличному та графічному вигляді);
- можливість *внесення коректив* у методику розрахунку і форму відображення кінцевого результату;
- можливість *повторного вирішення* завдання з будь-якої стадії розрахунку;
- можливість роботи із системою *кільком користувачам* у мережі;
- *простота і дружність інтерфейсу* та легкість взаємодії людини–системи.

Загалом система підтримки прийняття рішень складається з таких елементів:

1) мовна система (МС) – забезпечує зв'язок між користувачем і всіма компонентами комп'ютерної системи. За допомогою МС користувач формулює проблему і керує процесом

її рішення, використовуючи запропоновані мовною системою синтаксичні та семантичні засоби;

2) система подання проміжних і кінцевих результатів – сукупність мовних і графічних конструкцій, яка вміщує в себе всі типи повідомлень, які отримує особа, що приймає рішення (ОПР) від СППР;

3) система попередньої обробки даних і генерування результатів – основна система, яка реагує на запити ОПР, виконує обробку даних і передає результати в систему подання результатів;

4) база знань і даних – числові дані, алгоритми обробки даних, лінгвістичні дані, правила формування висновку, можливі структури моделей тощо.

Згідно з описаними вище вимогами спроектована інформаційна технологія аналізу фінансових даних всередині банку (оцінювання

фінансового стану компаній, окремих позичальників, прогнозування обсягів депозитів тощо). Технологія являє собою сукупність методів, програмних і технічних засобів, об'єднаних в єдиний технологічний ланцюжок, що забезпечує збір, збереження, редагування, обробку, подання і поширення інформації. Запропонована інформаційна технологія передбачає спільне використання певним чином модифікованої бази даних клієнтів, програмного комплексу GeNIe 2.0, зручного інтерфейсу для менеджерів, які виконують роботу з клієнтами, та додаткових програмних модулів побудови спеціальних оцінок (наприклад, індексу GINI та ін.). Прийняття рішення, наприклад стосовно можливості надання клієнту кредиту, здійснюється менеджером банку на основі отриманого значення ймовірності повернення кредиту. Всі етапи формальної перевірки документуються і зберігаються в базі даних

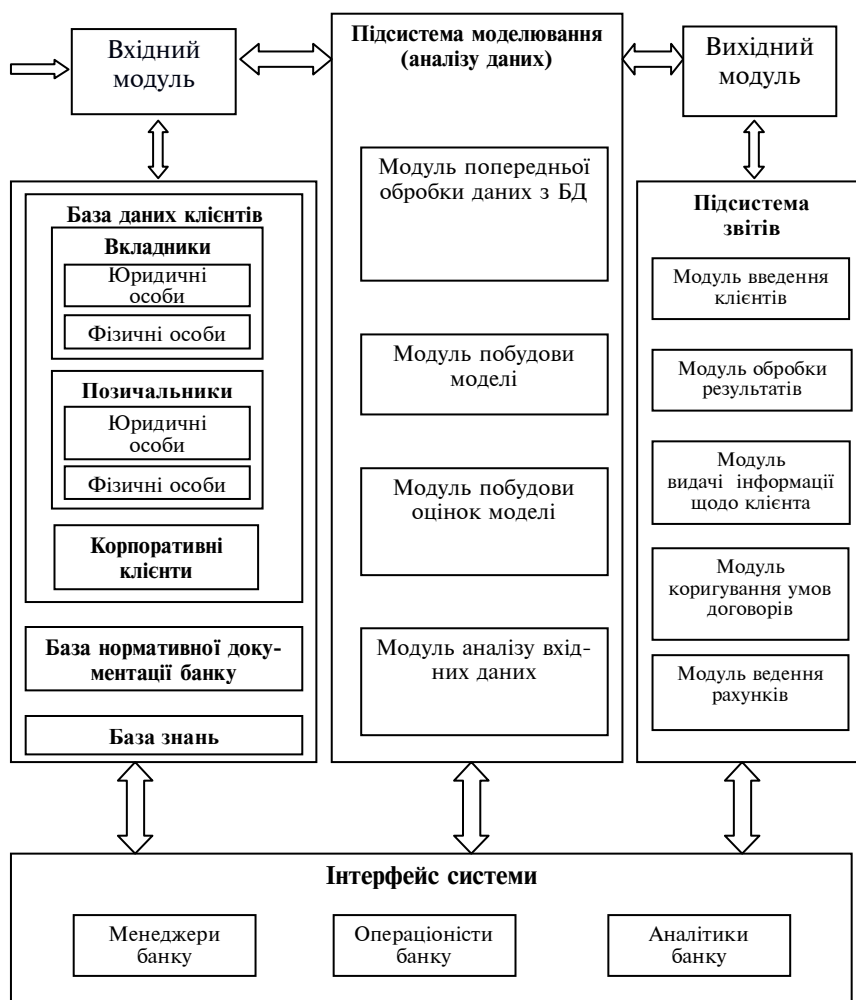


Рис. 4. Архітектура інформаційної СППР у комерційному банку

банку. Інформаційна технологія реалізована у вигляді інформаційної СППР, архітектуру якої наведено на рис. 4.

Застосування інтегрованого методу для аналізу кредитоспроможності позичальника

Дерева рішень. Метод дерев рішень (decision trees або answer tree) – один із найпопулярніших методів розв'язання задач класифікації та прогнозування. Іноді цей метод ІАД називають також деревами вирішувальних правил, деревами класифікації та регресії.

Вперше дерева рішень запропоновані Ховілендом і Хантом наприкінці п'ятдесятих років минулого століття. Першою працею, в якій викладається суть дерев рішень, була праця Ханта "Експерименти в індукції" [5], опублікована в 1966 р.

Серед методів побудови дерев рішень найпопулярніші такі:

- **CHAID** (CHi-squared Automatic Interaction Detector). Це найбільш відомий метод побудови дерев рішень, в якому для одержання оптимального розділення використовується критерій зв'язку між категоріальними змінними χ^2 (у випадку, якщо цільова змінна кількісна, то використовується F -критерій). Дані стосовно аналізу цільової змінної і змінні-фактори можуть бути як кількісними, так і категоріальними, однак кількісні змінні-фактори при побудові дерева перетворюються в категоріальні;

- **CART** (Classification and Regression Trees). Метод відомий також як "метод побудови дерев регресії і класифікації". На відміну від описаного вище методу він ґрунтується не на статистичних критеріях, а на зменшенні неоднорідності сегментів (вузлів). Він добре працює в тому випадку, коли всі змінні мають кількісний характер. В цьому методі можуть бути використані як кількісні, так і категоріальні цільові змінні і змінні-фактори;

- **QUEST** (Quick, Unbiased and Efficient Statistical Tree). В даному методі для вибору факторів застосовуються різні критерії залежно від типу

потенційно можливого фактора. Він дає можливість уникати зміщення, пов'язаного з вибором факторів з більшою кількістю категорій, але цільова змінна в цьому випадку повинна бути категоріальною. Змінні-фактори можуть бути як кількісними, так і категоріальними.

Якщо залежна, тобто цільова, змінна набуває дискретних значень, то за допомогою методу дерева рішень розв'язується задача класифікації. Якщо ж залежна змінна набуває безперервних значень, то дерево рішень установлює залежність цієї змінної від незалежних змінних,

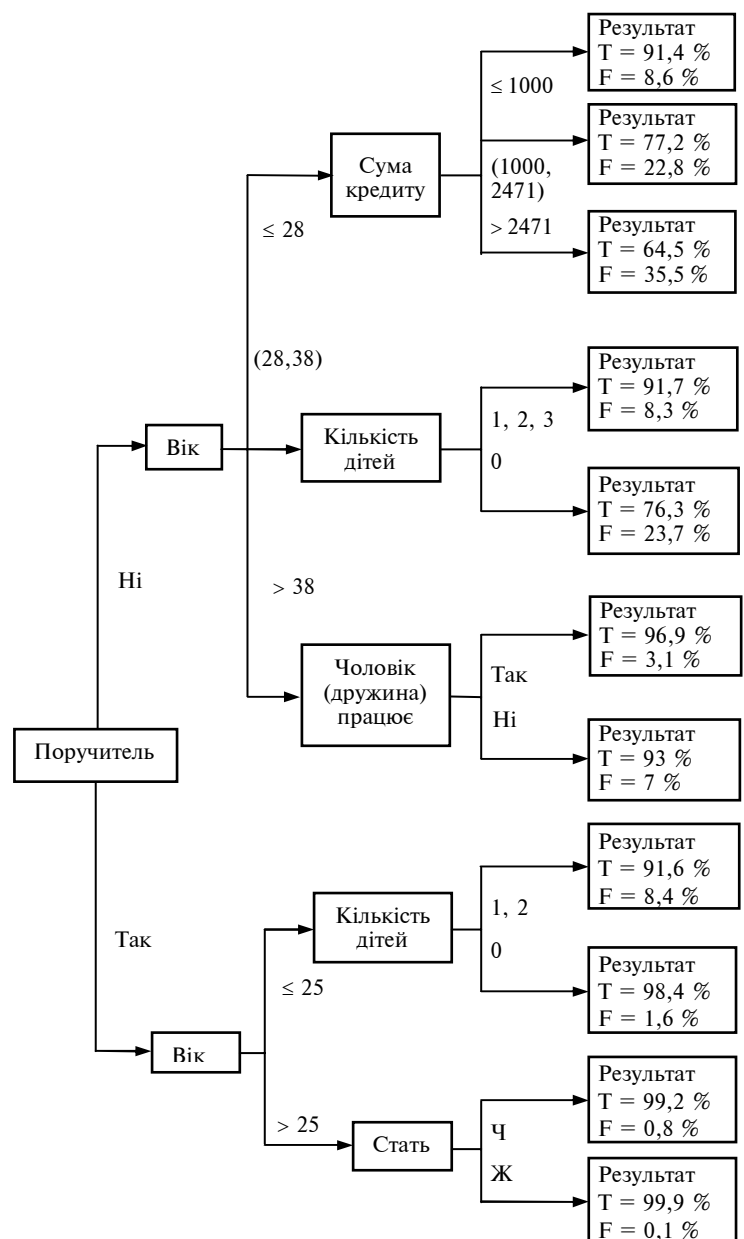


Рис. 5. Система кредитного скорингу у вигляді дерева рішень за методом CHAID

тобто розв'язується задача чисельного прогнозування. У простому вигляді дерево рішень – це спосіб подання правил в ієрархічній послідовній структурі. Основа такої структури – це відповіді “Так” або “Ні” на ряд можливих запитань.

На рис. 5 наведено скорингову систему у вигляді дерева рішень за методом CHAID, задачею якої є отримання відповіді на запитання: “Чи видавати кредит?” Щоб розв'язати цю задачу, необхідно визначити ймовірність дефолту клієнта банку. Для цього треба відповісти на ряд запитань, що містяться у вузлах (вершинах) цього дерева – починаючи з його кореня. Вершина “Поручитель” – вершина перевірки, тобто умова. Якщо відповідь позитивна – “Так”, то здійснюється перехід до верхньої частини дерева – до вершини “Вік”, при негативній – до нижньої частини дерева. Таким чином, внутрішній вузол (вершина) дерева – це вузол перевірки певної умови. Далі йде наступне питання і так далі, поки не буде досягнуто кінцевого вузла (вершини) дерева, який є вузлом рішення. Для нашого дерева кінцевим вузлом є вершина “Результат”, що набуває двох значень: “Т” і “F” (Т – true і F – false), де “F” – ймовірність неповернення клієнтом кредиту, виданого банком (дефолту).

У результаті проходження від кореня дерева “Поручитель” до його кінцевої вершини “Результат” розв'язується задача класифікації. На основі отриманих значень ймовірностей “Т” і “F” визначається ступінь кредитоспроможності клієнта та приймається рішення стосовно видачі або відмови в наданні кредиту.

Побудова скорингових моделей здійснена за трьома описаними методами в системі статистичних розрахунків SPSS 16. Як видно з табл. 4, найкращий результат серед дерев рішень отримано за методом CHAID.

Кластерний аналіз. У 1939 р. Тріон ввів новий термін “кластерний аналіз”. На відміну від завдань класифікації кластерний аналіз не потребує апріорних припущень стосовно даних. Задачу кластеризації об'єктів можна розглядати як процес виявлення істотно упорядкованих даних у багатовимірній матриці, завдяки чому стає можливим виділення кластерів – “щільних” скупчень об'єктів, що досліджуються. Кластерний аналіз розвивався паралельно в кількох напрямках: біологія, психологія та інші науки, тому більшість методів мають по дві й більше назв.

На сьогодні кластерний аналіз містить у собі більше 100 різноманітних алгоритмів [6]. Для розв'язання задачі побудови скорингових моделей використано 49 методів кластерного аналізу:

- ієрархічні: 1 – зв'язку між групами; 2 – зв'язку всередині груп; 3 – ближнього сусіда; 4 – віддалених сусідів; 5 – зважений центроїдний; 6 – медіан; 7 – Варда.

- ітераційні: 1 – *l*-середніх; 2 – *k*-медіан; 3 – двокроковий алгоритм BIRCH.

За критерій визначення міри відстані між кластерами серед наведених методів взято такі міри: 1 – квадратична евклідова; 2 – проста евклідова; 3 – коефіцієнт кореляції Пірсона; 4 – Чебишева; 5 – Мінковича.

Для розв'язання задачі скорочення кількості змінних у задачі кластеризації клієнтів банку застосовано найпопулярніший метод факторного аналізу – метод головних компонентів. При цьому для розв'язання задачі обертання факторів застосовано метод варімаксу. У табл. 2 наведено отримані значення оберненої матриці. За цими значеннями можна визначити, з яких змінних (атрибутивів) складається кожен із факторів. Фактор-1 складається із змінних “стать”, “вік”, “сімейний стан” і “тип трудової зайнятості”, фактор-2 – із змінних “сімейний стан”, “кількість дітей” і “чоловік (дружина) працює”, фактор-3 – із змінних “освіта”, “поручитель” і “сума кредиту”.

Таблиця 2. Значення оберненої матриці, отримані з використанням нормалізації Кайзера

Вхідні дані	Фактор		
	1	2	3
Стать	-0,538	0,085	-0,223
Вік	0,811	0,171	-0,036
Сімейний стан	0,577	0,566	-0,024
Кількість дітей	-0,153	0,769	-0,033
Чоловік (дружина) працює	0,093	0,697	-0,014
Освіта	-0,158	-0,075	0,537
Тип трудової зайнятості	0,543	-0,266	-0,369
Поручитель	0,121	-0,141	0,529
Сума кредиту	0,085	0,181	0,626

За отриманими результатами обчислювальних експериментів і з використанням критерію загальної точності моделі для класифікації кредитоспроможності клієнтів банку встановлено, що кращим виявився метод ближнього сусіда,

коли за міру відстані береться коефіцієнт кореляції Пірсона, або з попереднім факторним аналізом – за методом головних компонент. Також прийнятна загальна точність моделі, одержана при застосуванні методів: зв'язку між групами, найбільш віддалених сусідів, зваженого центроїдного, медіан і k -середніх. Низьку точність показали методи зв'язку всередині груп і Варда, а найгіршим виявився двокроковий метод.

Штучні нейронні мережі. Одним із широко відомих методів ІАД є штучні нейронні мережі (ШНМ). Незважаючи на велику розмаїтість варіантів існуючих нейронних мереж, вони мають загальні риси. Так, усі вони складаються з великої кількості зв'язаних між собою однотипних елементів – нейронів, які імітують нейрони головного мозку. На рис. 6 показано схему нейрона для задачі оцінювання кредитоспроможності, де x_i – значення i -го входу нейрона; w_i – вага i -го синапсу.

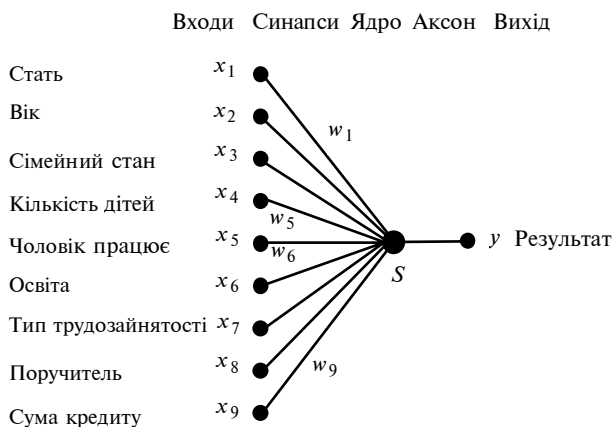


Рис. 6. Схема нейрона

Для побудови скорингових моделей використано такі методи побудови ШНМ [7, 8], як зворотне поширення (back propagation) та алгоритм навчання з відновленням (resilient propagation). Моделювання виконано на основі програмної аналітичної платформи Deductor. Серед моделей, побудованих у вигляді ШНМ, найкращими виявилися два типи ШНМ: коли у шарі 9 вхідних, 3 прихованих та 1 вихідний нейрони і 9 вхідних, 5 прихованих та 1 вихідний нейрони.

Інші, складніші, архітектури за наявною навчальною вибіркою клієнтів банку навчалися погано або не навчалися зовсім. Для алгоритму навчання з відновленням задано 2000 епох навчання.

Бінарні регресійні моделі дискретного вибору. В 1805 р. було опубліковано першу працю, присвячену методу найменших квадратів (МНК), з якої розпочався формальний розвиток регресійного аналізу, хоч насправді метод запропонував Гаусс у 1795 р. Моделі дискретного вибору призначені для пояснення дискретного вибору результату, тобто розв'язання задачі оцінювання кредитоспроможності позичальника за моделями дискретного вибору означає, що $y = 1$ (клієнт кредитоспроможний), якщо $probability \geq 0,5$, або $y = 0$ (дефолт) – інакше. Коефіцієнти наведених нижче рівнянь обчислені з використанням системи моделювання "Eviews 3".

Лінійна ймовірнісна модель. Це різновид множинної регресії, загальне призначення якої полягає в аналізі зв'язку між кількома незалежними змінними (регресорами або предикторами) і залежною змінною:

$$probability = 0,7545 - 0,0358x_1 + 0,0024x_2 + 0,01x_3 + 0,0121x_4 + 0,0214x_5 + 0,0142x_6 + 0,0037x_7 + 0,1112x_8 - 1,6732x_9.$$

Основний недолік лінійної ймовірнісної моделі полягає в тому, що значення ймовірності виходить за діапазон $[0, 1]$, а тому замість неї найчастіше використовуються логіт- і пробіт-моделі.

Логіт-модель. Ще в 1838 р. бельгійський математик Верхулст запропонував ідею використання логістичної функції для демографічних досліджень, але широкої популярності вона набула в 1925 р. завдяки практичним дослідженням Юли, Перла та Ріда. Логіт-модель оцінювання кредитоспроможності фізичних осіб має вигляд

$$probability = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + (\exp(z))^{-1}} = \frac{1}{1 + \exp(-z)},$$

де $z = -0,9563x_1 + 0,0695x_2 + 0,1426x_3 + 0,1875x_4 + 0,2264x_5 + 0,2727x_6 + 0,0509x_7 + 3,0315x_8 - 0,0005x_9$.

Пробіт-модель. Ця модель була запропонована в 1935 р. біологом Бліссом, але великого поширення набула лише на початку сімдесятих років минулого століття завдяки поширенню мейнфреймів, які здатні розв'язувати задачі

нелінійної максимізації. Для задачі, що розв'язується, логіт-модель записується так:

$$probability = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt,$$

де $z = -0,2677 - 0,4469x_1 + 0,035x_2 + 0,0918x_3 + 0,1097x_4 + 0,1148x_5 + 0,1436x_6 + 0,1539x_7 + 1,4139x_8 - 0,0002x_9$.

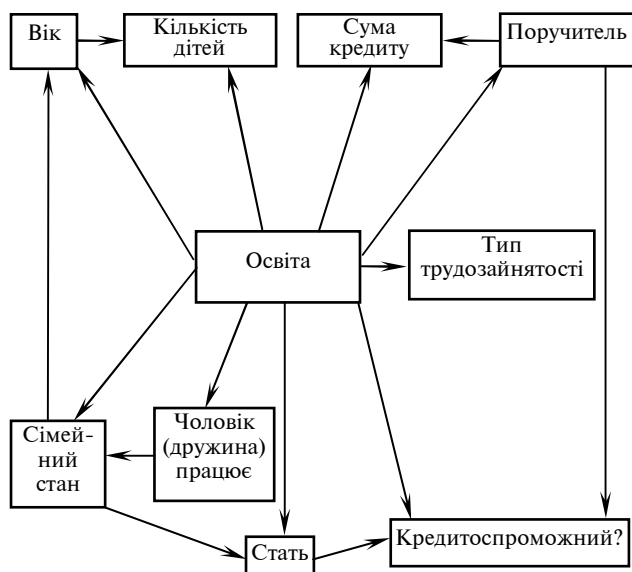


Рис. 7. Система кредитного скорингу у вигляді БМ

Байєсівські мережі (БМ). Байєсівські мережі з'явилися в 1985 р. на перетині теорії ймовірностей і теорії графів завдяки американському вченому Джуді Перлу, хоч теорема Байєса, покладена в основу побудови ймовірного висновку за БМ, запропонована була ще в 1763 р. Ідея використання БМ полягає в зображенні причинно-наслідкових зв'язків процесу у вигляді графа, в якому кожна змінна (фактор, атрибут) процесу є вершиною, а існуючі причинно-наслідкові зв'язки між змінними процесу являють собою спрямовані дуги, що з'єднують відповідні вершини.

На рис. 7 показано скорингову модель у вигляді БМ, побудовану за ітераційним евристичним методом, який ґрунтується на використанні оцінки взаємної інформації між вершинами та функції опису мінімальної довжини [9].

У табл. 3 наведено десять змодельованих ситуацій із використанням методу побудови точного ймовірного висновку в БМ за навчальними даними.

Побудованій скоринговій моделі у вигляді БМ (див. рис. 7) відповідають такі статистичні характеристики: похибка першого роду – 115 %, похибка другого роду – 157, загальна похибка – 272, загальна точність моделі – 0,918, похибка класифікації – 15 %.

Аналіз результатів. У табл. 4 і 5 наведено статистичні і прогнозные характеристики, отри-

Таблиця 3. Змодельовані ситуації

Номер ситуації	Інстанційовані вершини (апріорна інформація стосовно клієнта)	Ймовірність повернення клієнтом кредиту, %
1	Стать = "Чоловік"	92,08
2	Стать = "Жінка"	97,55
3	Поручитель = "Так"	99,06
4	Поручитель = "Ні"	87,98
5	Вік < 32 років і Сімейний стан = "Самотній" та Сума кредиту > 5000	76,92
6	Тип трудової зайнятості = "Працівник банку" і Сімейний стан = "Одружений"	94,66
7	Освіта = "Вища" і Кількість дітей = "один" та Чоловік (дружина) працює = "Так"	97,39
8	Освіта = "Середня" і Кількість дітей = "немає" та Чоловік (дружина) працює = "Ні" і Поручитель = "Ні" та Сума кредиту > 2500	69,78
9	Стать = "Чоловік" і Сімейний стан = "Удівець" та Освіта = "Середня спеціальна"	78,95
10	Стать = "Жінка" і Сімейний стан = "Удівець" та Освіта = "Середня спеціальна"	98,81

Таблиця 4. Статистичні характеристики отриманих моделей оцінювання кредитоспроможності

Назва методу	Похибка			Точність моделі	Відсоток похибок при класифікації
	1-го роду	2-го роду	Загальна		
За умови 90 % ймовірності повернення кредиту					
CHAID	76	275	351	0,895	10
CART	68	299	367	0,891	20
QUEST	47	546	593	0,823	15
Найкращі ієрархічні методи кластеризації з використанням квадратичної міри Евкліда					
Зважений центроїдний	174	93	267	0,92	15
Варда	146	296	442	0,868	15
Найкращі ієрархічні методи кластеризації з використанням коефіцієнта Пірсона					
Зв'язок між групами	177	26	203	0,939	20
Зв'язок всередині груп	171	370	541	0,838	20
Віддалених сусідів	177	71	248	0,926	20
Медіан	177	26	203	0,939	15
Неієрархічні методи кластеризації					
<i>k</i> -середніх для ітерацій і класифікації	116	564	680	0,797	10
<i>k</i> -середніх тільки для класифікації	172	119	291	0,914	15
Двокроковий	78	1245	1323	0,605	45
Найкращий ієрархічний метод з попередньою факторизацією					
Ближнього сусіда	178	1	179	0,947	15
Лінійна ймовірнісна модель	175	0	175	0,947	15
Логіт-модель	163	6	169	0,949	15
Пробіт-модель	168	2	170	0,948	15
Байєсівська мережа	115	157	272	0,918	15

Таблиця 5. Результати порівняння нейронних мереж

Назва алгоритму побудови ШНМ	Кількість нейронів у шарі			Точність моделі	Відсоток похибок при класифікації
	Вхідних	Прихованих	Вихідних		
Back propagation	9	3	1	0,9	20
Back propagation	9	5	1	0,91	15
Resilient propagation	9	3	1	0,91	15
Resilient propagation	9	5	1	0,9	15

мані при використанні відповідних методів ІАД для побудови скорингових моделей оцінювання кредитоспроможності фізичних осіб. При класифікації розглядалося значення порогу в 90 %, тобто якщо ймовірність повернення кре-

диту нижча 90 %, то клієнт класифікується як некредитоспроможний.

Найкращі результати отримано за методом дерев рішень CHAID і кластерного аналізу *k*-середніх. Для цих методів відсоток похибок класифікації дорівнює 10, а це означає, що із 100 виданих кредитів 10 було класифіковано невірно. Більшість скорингових систем, отриманих за іншими методами ІАД, дали похибку класифікації 15–20 %. За критерієм загальної точності моделі, найкращий результат із значенням 0,949 отримано за логіт-моделлю, а серед методів кластерного аналізу – ближнього сусіда з попередньою факторизацією.

Висновки

Аналіз фінансової діяльності компанії, установи або банку необхідно використовувати для забезпечення подальшого успішного розвитку підприємства, розширення його діяльності та ефективного керування ресурсами. Ана-

ліз фінансової звітності і розрахунок фінансових коефіцієнтів дають можливість оцінити поточний фінансовий стан підприємства і зробити прогноз на майбутнє. Оскільки стандартизованої системи оцінки фінансового стану не існує, то аналітики можуть використовувати альтернативні множини показників.

Запропонована в статті технологія аналізу даних дає можливість системно підійти до процесу побудови ІСППР для аналізу і прогнозування стану підприємства. Інтегрований підхід, який запропоновано використовувати для аналізу і прогнозування даних з використанням ІСППР, сприяє істотному покращенню точності і якості прогнозу, встановленню причинно-наслідкових зв'язків і поясненню причин коливань окремих змінних, а також виконанню експрес-аналізу мінімальної кількості найбільш важливих показників і поданню оперативної оцінки фінансового стану підприємства.

Серед методів ІАД, використаних для побудови скорингових моделей, найкращі резуль-

тати отримано за допомогою дерева рішень методу CHAID і методу кластерного аналізу k -середніх. Таким чином, підтверджено високу ефективність використання методів ІАД для аналізу і моделювання економетричних показників як необхідний крок розв'язання важливої проблеми – забезпечення стабільності та фінансової безпеки банківської системи України.

У подальших дослідженнях планується вивчити можливості підвищення точності моделей і впровадити механізм прогнозу ймовірностей дефолтів на основі скорингових моделей, розширити функціональні можливості ІСППР іншими сучасними методами обробки даних та альтернативними методами прийняття рішень. Досвід використання систем такого типу свідчить про те, що збільшення кількості альтернатив, які може згенерувати система, сприяє покращенню (підвищенню економічної ефективності) прийнятого рішення.

П.І. Бідюк, Н.В. Кузнецова, А.Н. Терентьев

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ АНАЛИЗА ФИНАНСОВЫХ ДАННЫХ

Рассмотрены основные особенности финансового анализа деятельности предприятия, предложена технология анализа финансовых данных предприятия на основе оригинального интегрированного подхода. На примерах анализа и прогнозирования уровня продаж компании, а также оценивания дефолта заемщиков кредитов показана возможность использования предложенной технологии. Приведено описание архитектуры разработанной информационной системы поддержки принятия решений для коммерческого банка.

P.I. Bidyuk, N.V. Kuznyetsova, O.M. Terentyev

DECISION SUPPORT SYSTEM BASED FOR ANALYSIS OF FINANCIAL DATA

In this paper, the basic features of financial analysis of an enterprise are considered, and the technology of the financial data analysis based on the integrated approach is proposed. We also show that this technology can be applied by giving the examples of sales analysis and performing the future forecasts as well as estimating defaults of credit borrowers. Moreover, we highlight the architecture of the proposed information system of decision-making support for the commercial bank.

1. *Kuznyetsova N.* Integrated approach to credit rating // Int. Scientific Conf. "Intelligence, Integrity, Reliability". – Kyiv: NTUU "KPI", 2010. – P. 30–31.
2. *Кузнецова Н.В.* Інтегрований підхід до оцінювання кредитних ризиків // Тр. Одес. политехн. ун-та. – Одеса, 2010. – № 1 (33). – С. 157–165.
3. *Кузнецова Н.В., Бідюк П.І.* Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування // Наукові вісті НТУУ "КПІ". – 2010. – № 1. – С. 42–53.
4. *Бідюк П.І., Коршевнюк Л.О.* Система для оцінювання і прогнозування стану підприємства на основі мереж Байєса // Наукові праці. – Миколаїв: Вид-во Чорномор. держ. ун-ту ім. Петра Могили. – 2010. – 134, вип. 121. – С. 60–74.
5. *Hunt E.B., Marin J., Stone P.* Experiments in Induction // The American Journal of Psychology. – N.Y.: Academic Press, 1966. – 80, N 4. – P. 651–653.
6. *Кривова О.А., Коваленко А.С.* Применение кластерного анализа для выявления соотношений индикаторов демографического развития // Кибернетика и вычис-

- лительная техника. – К.: Академперіодика, 2007. – 153. – С. 24–39.
7. Хайкин С. Нейронные сети: Полный курс. – М.: Изд. дом “Вильямс”, 2008. – 1104 с.
8. Riedmiller M., Braun H. A direct adaptive method for faster back propagation learning: the RPROP algorithm // IEEE International conference on image processing, Chicago, Illinois, USA, 3–6 May, 1993. – 1. – P. 586–591.
9. Терентьев А.Н., Бидюк П.И., Коршевнюк Л.А. Байесовская сеть – инструмент интеллектуального анализа данных // Пробл. управления и информатики. – К.: ИКИ НАНУ-НКАУ, 2007. – № 4. – С. 83–92.

Рекомендована Радою
Навчально-наукового комплексу
“Інститут прикладного системного
аналізу” НТУУ “КПІ”

Надійшла до редакції
1 лютого 2011 року