

ІНТЕГРОВАНІЙ ПІДХІД ДО ОЦІНЮВАННЯ КРЕДИТНИХ РИЗИКІВ

Н.В.Кузнцова. **Інтегрований підхід до оцінювання кредитних ризиків.** Запропоновано інтегрований підхід для оцінювання кредитних ризиків на основі комбінації двох підходів — мережі Байєса і дерева рішень або мережі Байєса і бінарної логістичної регресії. Інтегрований підхід показав кращі результати і може застосовуватися у банківській діяльності.

Н.В.Кузнцова. **Интегрированный подход к оценке кредитных рисков.** Предложен интегрированный подход для оценки кредитных рисков на основании комбинации двух подходов — сети Байеса и дерева решений или сети Байеса и бинарной логистической регрессии. Интегрированный подход показал лучшие результаты и может быть использован в банковской деятельности.

N.V.Kuznyetsova. **Integrated approach to credit risks estimation.** The integrated approach to credit risks estimation is proposed. This approach is based on the combination of two approaches — the Bayesian network and decision tree or Bayesian network and binary logistic regression. The integrated approach has shown better results and can be used in bank activity.

Світова криза виявила і показала справжню ситуацію в банківському секторі України. Недостатньо ретельне вивчення позичальників і видача кредитів ненадійним особам спричинили виникнення великої кількості заборгованостей і неповернень по кредитах. Це було пов'язано ще й з тим, що частина банків взагалі не користувалася іноземними методиками оцінки клієнтів, а інша частина банків не намагалася їх вдосконалити та адаптувати до українських реалій. В Україні почала діяти велика кількість шахраїв, які, знаючи специфіку цієї методики, знали наперед “правильні” відповіді на питання працівників банку, часто обманюючи та застосовуючи велику кількість сфабрикованих документів і, врешті-решт, отримували кредит, який аж ніяк не збиралися повертати банку. В такій ситуації банкам слід було б змінювати модель оцінки позичальника кожні 2...3 роки, періодично включаючи нові дані для оцінки. Зазвичай відомими моделями, якими користуються найчастіше банки для оцінювання кредитування, є логістична регресія та дерева рішень [1, 2]. Застосування нових підходів і методів до оцінки кредитних ризиків дозволило б банкам швидко відновити кредитування в Україні та значно зменшити обсяги втрат від несумлінного кредитування.

Тому так важливо запропонувати інтегрований підхід для оцінки кредитоспроможності позичальника на основі комбінації двох підходів; порівняти результати моделей, побудовані на основі дерев рішень та інтегрованого підходу — комбінації дерев рішень і мереж Байєса, а також логістичної регресії та комбінації логістичної регресії і мереж Байєса; зробити висновки щодо доцільності використання інтегрованого підходу для вирішення фінансово-економічних задач.

Результати застосування моделей логістичної регресії, дерев рішень і мереж Байєса для задачі аналізу кредитоспроможності позичальників, а також порівняння якостей моделей розглядалось в багатьох роботах [3...5]. Суттєвим недоліком дерев рішень і логістичної регресії є те, що вони працюють лише з обмеженою кількістю вхідних даних. Тому при побудові моделей на основі цих методів застосовують невелику кількість факторів (близько 10), які вибираються або експертами, або програмою, але при цьому виникає загроза, що будуть не враховані і відкинуті суттєві фактори, які врешті-решт можуть спричинити серйозний вплив на результат. Саме тому необхідно запропонувати підхід, який унеможливить випадкове відкидання факторів.

Оскільки на етапі збору статистичних даних для будь-якої задачі завчасно не відомі суттєві фактори, які спричиняють вплив на результат, то фінансові установи намагаються зібрати якомога більш розширену інформацію. Маючи такий величезний набір даних щодо певної події, необхідно побудувати модель, яка визначить вплив факторів на результат. Очевидно, що на цьому етапі постає питання, як формалізувати отримані дані і виявити, які з них є суттєвими. Для цього пропонується побудувати мережу Байєса, яка встановить причинно-наслідкові

зв'язки між змінними, що відповідають факторам, силу зв'язків між цими змінними, а також дозволить виявити змінні, які взагалі не пов'язані з результируючою подією (“висячі змінні”). На основі побудованої мережі і встановлених зв'язків можна суттєво скоротити кількість факторів, які будуть включені на наступному етапі при побудові моделі. При цьому слід також пам'ятати про те, що для логістичної регресії зменшення кількості факторів, які включаються в модель, зазвичай спричиняє погіршення якості моделі. Тому необхідно застосовувати мережу Байєса лише як інструмент зменшення кількості факторів, які будуть включені в модель, а не як інструмент, який виявить найсуттєвіші фактори, а всі інші відкине.

Розглядається можливість застосування запропонованого інтегрованого підходу на конкретному прикладі та порівнюються отримані результати для стандартних моделей та моделей, побудованих за допомогою інтегрованого підходу.

Логістична регресія — це вид нелінійної множинної регресії, яка аналізує функціональну залежність між декількома незалежними змінними (регресорами) і залежною змінною [1, 2]. *Бінарна логістична регресія* застосовується у тому випадку, коли вихідна змінна може приймати тільки два значення.

У випадку застосування бінарної логістичної регресії ймовірність p входження клієнта у стан дефолту D пов'язана зі змінними (X_1, \dots, X_n) , як показано у рівнянні

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n.$$

Оцінка β_0 називається перетином, а інші оцінки — β_i — ще називають кутовими коефіцієнтами. Така модель має ту перевагу, що коректно представляє ймовірність між нулем і одиницею, незалежно від вхідних значень. Перепишучи попередній вираз і використовуючи вектори для представлення параметрів моделі і вхідних змінних, ймовірність дефолту можна представити у зручнішій формі

$$p = \frac{\exp(\beta X)}{1 + \exp(\beta X)}.$$

Дерева рішень — це один з методів автоматизованого аналізу даних, коли правила представляються у вигляді послідовної ієрархічної структури, в якій кожному вузлу відповідає єдиний вузол, що дає рішення [2, 3]. Під правилом розуміється конструкція, представлена у вигляді “якщо ..., то...”.

Формально *мережа Байєса* — це ймовірнісна модель у вигляді спрямованого ациклічного графа, кожний вузол якого представляє змінну модельованого процесу, а кожна дуга представляє причинне відношення між двома змінними [5, 6]. Змінні характеризуються розподілом ймовірності для кожного значення. На розподіл ймовірності кожного вузла впливають стани (для дискретних вузлів) або значення (для неперервних вузлів) кореневої вершини. Умовні ймовірності станів вузлів зберігаються у таблицях умовних ймовірностей.

Інтегрована модель на основі мережі Байєса і дерева рішень (ІМБД) — це модель, побудована на основі комбінації двох методів — мережі Байєса і дерева рішень, де на першому етапі для скорочення кількості змінних застосовується мережа Байєса, а для оцінки ймовірності дефолту — дерево рішень. Мережа Байєса для задачі кредитування, яка на основі статистичних даних банку (кредитних історій) встановлює причинно-наслідкові зв'язки між змінними — характеристиками клієнта та результатом — поверненням чи неповерненням ним кредиту, побудована на рис. 1. При відборі суттєвих факторів спочатку виключаються “висячі” змінні, які взагалі ні на що не впливають та є зайвими при побудові моделі, потім виключаються змінні, які не мають ніякого (ні прямого, ні опосередкованого) впливу на результируючу змінну — повернення кредиту.

На рисунку 1 можна побачити, що мережа Байєса встановила змінні (ліміт кредитної картки та щомісячне обслуговування), які ніяким чином не впливають на результируючу змінну — повернення кредиту. Тому вони можуть бути виключені зі списку змінних і на наступному етапі побудови дерева рішень кількість змінних буде меншою. На першому етапі мережа Байєса дозволила скоротити кількість факторів з 26 до 24.

Далі були скорочені фактори, які безпосередньо не впливають на результируючу змінну, такі як освіта, кількість дітей, вартість товару, початковий платіж, застава, сума інших доходів, стаж

на даному місці роботи, чи працює дружина. Серед суттєвих факторів (їх залишилось 17) були виявлені такі суттєві змінні, які на наступному етапі були включені у дерево рішень: вік, стать, сімейне становище, дохід, тип зайнятості, сума кредиту, тип місця проживання (своє чи орендоване), тощо.

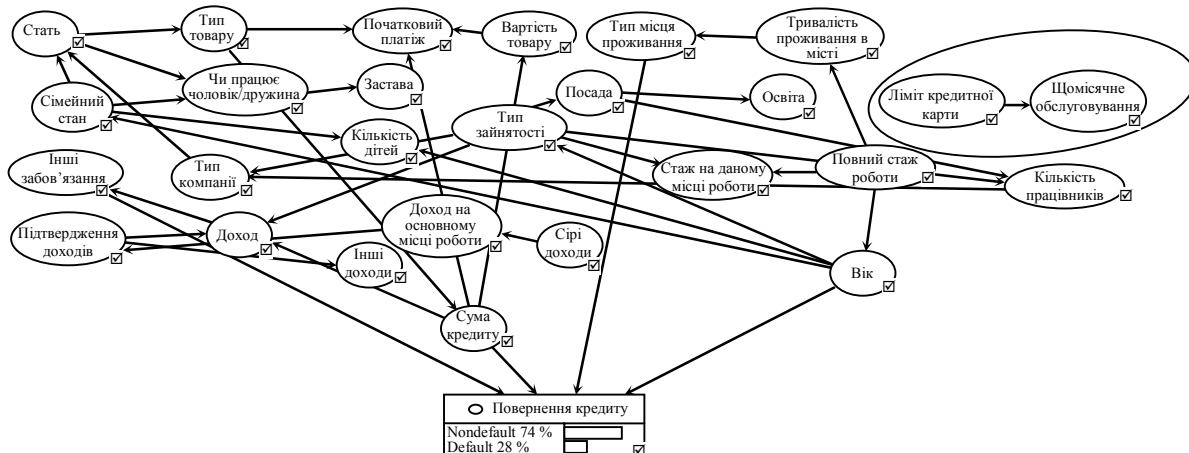


Рис. 1. Мережа Байєса для задачі оцінювання кредитоспроможності позичальника

Для побудованої моделі обчислюється загальна точність, помилки першого та другого роду для різних порогів відсікання (cut-off) (табл. 1). Поріг відсікання для моделі в даному випадку — це значення ймовірності дефолту, більше якого клієнт вважається ненадійним, і дається рекомендація не видавати йому кредит, або іншими словами, це нижнє значення ймовірності повернення кредиту. Для клієнтів, у яких значення ймовірності вище цього значення, дається рекомендація видати кредит.

Як видно з отриманих результатів, побудована модель показує кращі значення загальної точності для порогів відсікання 0,1 і 0,15, а також точніше визначає дефолти і менше допускає помилок у порівнянні з даними для моделі дерев рішень [3].

Більш зрозумілим і частіше згадуваним в літературі параметром оцінки якості моделі є індекс *GINI*, розроблений італійським статистиком Коррадо Джіні, який тісно пов'язаний з чисельним показником площі під *ROC*-кривою (Receiver Operation Characteristic — робоча характеристика приймача). Індекс *GINI* не залежить від кредитної політики банку та встановленого порога відсікання і може бути визначений через площу фігури, що знаходиться під *ROC*-кривою

$$GINI = 2AUC - 1.$$

Детальний опис побудови *ROC*-кривої та обчислення індексу *GINI* наведений у роботі [3]. *ROC*-крива для інтегрованої моделі на основі мережі Байєса і дерева рішень подається (рис. 2).

Для моделі, побудованої на основі дерева рішень, $AUC = 0,786$, а індекс *GINI* відповідно становить: $GINI = 2AUC - 1 = 0,572$ [3]. Значення площі під кривою для моделі на основі інтегрованого підходу $AUC = 0,799$, а $GINI = 2AUC - 1 = 0,598$, що є значно вищим результатом та, як відомо, найкращим результатом для країн з економікою, що розвиваються [1].

Таким чином, інтегрований підхід до оцінки кредитоспроможності позичальника дозволив покращити якість моделі, а також її здатність розрізняти надійних та ненадійних клієнтів, прибрав зайву пересторогу та зменшив кількість дарма відсіяних клієнтів.

Інтегрована модель на основі логістичної регресії і мережі Байєса (ІМЛБ) — це модель побудована на основі комбінації двох методів логістичної регресії і мережі Байєса. Оскільки при застосуванні логістичної регресії суворим обмеженням є невелика кількість факторів, то пропонується застосувати інтегрований підхід, де на першому етапі будується мережа Байєса, яка визначає суттєві змінні, а далі вже лише суттєві змінні включаються як фактори. Слід нагадати, що мережа Байєса (див. рисунок 1), дозволила скоротити кількість факторів з 26 до 17. В принципі можна було продовжити скорочувати кількість суттєвих факторів, але очевидно, що це спричинить погіршення якості моделі. Тому застосовується зворотний покроковий метод (Backward Stepwise), який по одній виключає змінні, що мають найменшу значущість. Результати застосування розглянутого інтегрованого підходу показано (табл. 2).

Таблиця 1

Якісні характеристики моделі, побудованої на основі інтегрованого підходу

Показники моделі	Прогноз: Повернення кредиту (0)	Прогноз: Дефолт (1)	Процент точності
Cut-off=0,1			
Факт: Повернення кредиту (0)	74	76	0,493333333
Факт: Дефолт (1)	5	45	0,9
Загальна точність моделі			0,595
Cut-off=0,15			
Факт: Повернення кредиту (0)	83	67	0,553333333
Факт: Дефолт (1)	7	43	0,86
Загальна точність моделі			0,63
Cut-off=0,2			
Факт: Повернення кредиту (0)	89	61	0,593333333
Факт: Дефолт (1)	10	40	0,8
Загальна точність моделі			0,645
Cut-off=0,25			
Факт: Повернення кредиту (0)	102	48	0,68
Факт: Дефолт (1)	11	39	0,78
Загальна точність моделі			0,705
Cut-off=0,3			
Факт: Повернення кредиту (0)	106	44	0,706666667
Факт: Дефолт (1)	13	37	0,74
Загальна точність моделі			0,715

Таблиця 2

Якісні характеристики моделі, побудованої на основі інтегрованого підходу з застосуванням мережі Байєса і бінарної логістичної регресії

Показники моделі	Прогноз: Повернення кредиту (0)	Прогноз: Дефолт (1)	Процент точності
Cut-off=0,1			
Факт: Повернення кредиту (0)	61	89	0,406667
Факт: Дефолт (1)	2	48	0,96
Загальна точність моделі			0,545
Cut-off=0,15			
Факт: Повернення кредиту (0)	83	67	0,553333
Факт: Дефолт (1)	2	48	0,96
Загальна точність моделі			0,655
Cut-off=0,2			
Факт: Повернення кредиту (0)	94	56	0,626667
Факт: Дефолт (1)	4	46	0,92
Загальна точність моделі			0,7
Cut-off=0,25			
Факт: Повернення кредиту (0)	106	44	0,706667
Факт: Дефолт (1)	4	46	0,92
Загальна точність моделі			0,76
Cut-off=0,3			
Факт: Повернення кредиту (0)	117	33	0,78
Факт: Дефолт (1)	8	42	0,84
Загальна точність моделі			0,795

Для моделі, побудованої на основі бінарної логістичної регресії, значення площі під кривою $AUC = 0,838$, а $GINI = 2AUC - 1 = 0,676$ [3] (рис. 3). Значення площі під кривою для моделі на основі інтегрованого підходу становить $AUC = 0,832$, а $GINI = 2AUC - 1 = 0,664$.

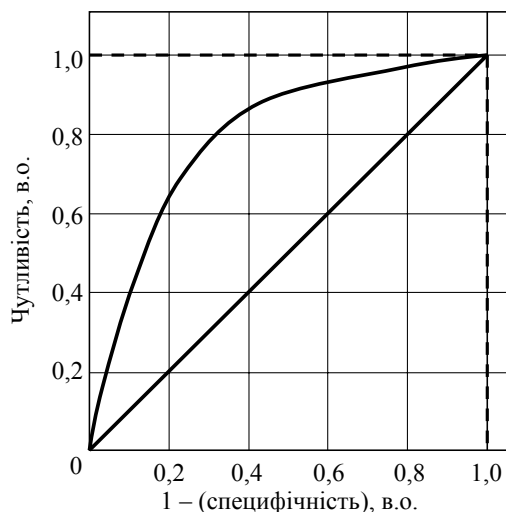


Рис. 2. ROC- крива для моделі на основі інтегрованого підходу

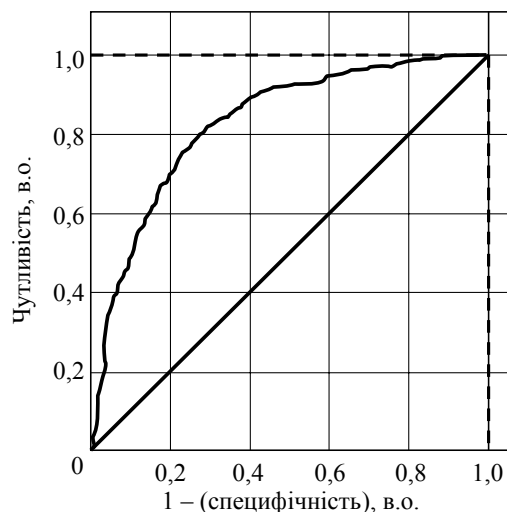


Рис. 3. ROC-крива для моделі на основі інтегрованого підходу з використанням бінарної логістичної регресії і мережі Байєса

Модель, побудована на основі інтегрованого підходу, показала краще виявлення дефолтів та сильнішу недовіру до надійних позичальників, тому загальна точність моделі та індекс GINI виявилися нижчими. Цей підхід буде актуальним для банку, який проводить жорстку політику відсікання ненадійних клієнтів, тобто відсіваючи позичальників, наприклад, для яких ймовірність повернення кредиту встановлена на рівні 0,8.

Результати порівняння якостей моделей на основі дерев рішень, бінарної логістичної регресії, мереж Байєса та інтегрованого підходу на основі мережі Байєса і дерева рішень та на основі мережі Байєса і логістичної регресії показано (табл. 3).

Таблиця 3

Порівняльна таблиця характеристик для різних моделей ІМБД та ІМЛБ

Назва методу	Індекс GINI	Значення AUC	Точність моделі	Якість моделі
Бінарна логістична регресія	0,676	0,838	0,78	Дуже висока
Дерева рішень	0,572	0,786	0,75	Прийнятна
Мережа Байєса	0,678	0,839	0,75	Дуже висока
Інтегрований підхід на основі мережі Байєса і дерева рішень	0,664	0,832	0,795	Дуже висока
Інтегрований підхід на основі мережі Байєса і логістичної регресії	0,58	0,799	0,715	Прийнятна

Інтегрований підхід дає можливість банку проводити більш збалансовану кредитну політику — при встановленні досить жорстких порогів відсікання (на рівні 0,1 та 0,15) інтегрована модель відсіює меншу кількість надійних позичальників, а тому застосування даного підходу буде особливо актуальним після прийняття банками рішення про відновлення масового роздрібно-кредитування, коли спочатку банкам прийдеться вести досить ретельну та обережну політику перевірки клієнта та видачі кредиту.

Запропонований у роботі підхід показав прийнятні результати при застосуванні на практиці, а тому у подальшому автором передбачається застосовувати його і для інших типів задач.

Література

1. Бідюк, П.І. Моделі оцінки ризиків кредитування фізичних осіб / П.І. Бідюк, Є.О. Матрос // Кібернетика та обчисл. техніка. — 2007. — № 153. — С. 87 — 95.
2. Kiss, F. Credit scoring processes from a knowledge Management perspective / F. Kiss // Periodica Polytechnica. Ser. Soc. Man. Sci. — 2003. — Vol. 11, № 1. — P. 95 — 110.
3. Кузнецова, Н.В. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування / Н.В. Кузнецова, П.І. Бідюк // Наук. вісті НТУУ “КПІ”. — 2010. — № 1. — С. 42 — 53.
4. Бідюк, П.І. Основні етапи побудови і приклади застосування мереж Байеса / П.І. Бідюк, Н.В. Кузнецова // Систем. дослідж. та інформ. технології. — 2007. — № 4. — С. 26 — 39.
5. Кузнецова, Н.В. Системний підхід до аналізу кредитних ризиків з використанням мереж Байеса / Н.В. Кузнецова, П.І. Бідюк // Наук. вісті НТУУ “КПІ”. — 2008. — № 3. — С. 11 — 24.
6. Сравнение методов интеллектуального анализа данных при оценивании кредитоспособности физических лиц / А.Н. Терентьев, П.И. Бидюк, А.В. Миронова [та др.] // Проблемы упр. и информатики. — К.: ИКИ НАНУ-НКАУ, 2009. — № 5. — С. 141 — 149.

Рецензент д-р техн. наук, проф. Одес. нац. політехн. ун-ту Римша В.В.

Надійшла до редакції 24 квітня 2010 р.