

**Мойсеєнкова Д.А.**  
**Жуковська О. А.**

канд. фіз.-мат. наук, доцент  
Національний технічний університет України «КПІ»

## **МОДЕЛЬ ПРИЙНЯТТЯ КОМПЛЕКСНОГО СКОРИНГОВОГО РІШЕННЯ**

### **МОДЕЛЬ ПРИНЯТИЯ КОМПЛЕКСНОГО СКОРИНГОВОГО РЕШЕНИЯ**

#### **MODEL OF INTEGRATED DECISION OF SCORING MAKING**

*Актуальність створення, впровадження та використання скорингових систем для управління кредитними ризиками сьогодні не викликає сумніву. Такі системи використовують характеристики клієнта, який бажає отримати кредит, і оцінює ризик шляхом передбачення манери погашення боргу позичальником. В основу таких систем зазвичай покладена модель прийняття рішення, яка побудована на основі одного з підходів: байєсівський, множинна регресія, дискримінантний аналіз, генетичні алгоритми, дерева класифікації, логістичної регресії, нейронні мережі та інші. У кожного з підходів є свої переваги та недоліки. У статті проведено аналіз існуючих алгоритмів побудови скорингових систем. Наведено метод ROC-аналізу, завдяки якому можливо визначити найбільш ефективну модель прийняття рішення про надання чи відмові кредиту. Зазначається, що на даний час, остаточне рішення приймається експертом. Однак, за одних і тих же умов різні люди приймають різні рішення, що пояснюється особистісними факторами, що впливають на процес прийняття управлінських рішень. Тому у статті запропоновано оцінювати позичальника на основі декількох моделей та у випадку протилежних рішень кожної моделі, остаточне рішення приймати ґрунтуючись на побудованій моделі прийняття колективного рішення.*

**Ключові слова:** скоринг, скорингова модель, різновиди скорингових моделей, модель прийняття комплексного рішення.

*Актуальность создания, внедрения и использования скоринговых систем для управления кредитными рисками сегодня не вызывает сомнения. Такие системы используют характеристики клиента, который желает получить кредит, и оценивает риск путем предвидения манеры погашения долга заемщиком. В основу таких систем обычно положена модель принятия решения, которая построена на основе одного из подходов: байесовский, множественная регрессия, дискриминантный анализ, генетические алгоритмы, деревья классификации, логистической регрессии, нейронные сети и другие. У каждого из подходов есть свои преимущества и недостатки. В статье проведен анализ существующих алгоритмов построения скоринговых систем. Приведен метод ROC-анализа, благодаря которому можно определить наиболее эффективную модель принятия решения о предоставлении или отказе кредита. Отмечается, что в настоящее время, окончательное решение принимается экспертом. Однако, при одних и тех же условиях разные люди принимают различные решения, что объясняется личностными факторами, влияющими на процесс принятия управленческих решений. Поэтому в статье предложено оценивать заемщика на основе нескольких моделей и в случае противоположных решений*

каждой модели, окончательное решение принимать основываясь на построенной модели принятия коллективного решения.

**Ключевые слова:** скоринг, скоринговая модель, разновидности скоринговых моделей, модель принятия комплексного решения.

*The relevance of designing, implementing and using scoring systems for credit risk management today is not in doubt. These systems use the characteristics of the client who wants to get a loan, and assesses risk by predicting manners repayment by the borrower. The basis of such systems are usually based on the model of decision making that is based on one of the approaches bayyesivskyy, multiple regression, discriminant analysis, genetic algorithms, classification tree, logistic regression, neural networks, and others. Each approach has its advantages and disadvantages. This article provides an analysis of the existing algorithms scoring systems. The method ROC-analysis, by which is possible to determine the most efficient model of the decision to grant or denial of credit. It is noted that at present, the final decision is made by an expert. However, the same conditions different people make different decisions, because of personal factors that influence the decision making process. Therefore, the paper proposes to assess the borrower based on multiple models and in case of conflicting decisions of each model, to make the final decision based on the constructed model of collective decision.*

**Keywords:** scoring, scoring model, variety of scoring models, model of collective decision.

**Вступ.** Кредитні ризики є невід’ємною частиною діяльності банківських установ, так як, більшість банків формують значну частину своїх доходів за рахунок кредитних операцій. Тому особливу уваги приділяють оцінці та мінімізації кредитних ризиків.

На сьогоднішній день існує багато способів мінімізації кредитного ризику. Одним із них є оцінка кредитоспроможності позичальника, яка здійснюється за допомогою скорингових систем, що спеціально розробляються для підвищення ефективності прийняття рішень щодо кредитних угод.

Дане питання розглядали багато дослідників, такі як Г.Н. Белоглазова, Н.І. Валенцева, В.С. Захаров, Г.Г. Коробова, А.М. Косой, Л.П. Кроливецка, О.І. Лаврушин, І.В. Ларіонова, І.Д. Мамонова, Ю.С. Масленченков, М. Альтмана, Л. Андерсена, К.Дж. Балтроп, Дж. К. Ван-Хорн, Т. Деглиша, Р.Г. Ольхова та інші, науковий вклад яких неможливо оцінити.

Але кожна банківська установа намагається удосконалити скорингові системи, тому в більшості банків використовують декілька скорингових моделей. Це в свою чергу призводить до появи нової проблема, а саме прийняття остаточного рішення щодо класифікації позичальника у ситуації невизначеності, коли рішення скорингових моделей суперечать одна одній. Тому в даній статті буде розглянута модель прийняття остаточного рішення. Обрана тема дослідження потребує детального вивчення та розгляду і вона є актуальною на сучасному етапі трансформації економіки України та має теоретичну і практичну цінність.

**Постановка завдання.** Метою даної статті є розробка математичної моделі для визначення ефективної скорингової стратегії в ситуації невизначеності з метою виявлення неблагонадійних потенційних позичальників.

**Методологія.** У роботі проаналізовані та використані математичні методи для вирішення поставленого завдання.

Існує безліч моделей скорингу. Кожна з них використовує свій набір факторів, що характеризує позичальника, і в результаті отримують свою власну оцінку ризику, що дозволяє розділяти позичальників на певні категорії. Для кращого розуміння скорингу, розглянемо найпоширеніші моделі оцінки.

*Байєсівський підхід.* Припустимо, що відомі наступні функції та величини:

- $p_G$ - частка «гарних» клієнтів,  $p_B = 1 - p_G$  – частка «поганих» клієнтів;
- $p(x|G)$ ,  $p(x|B)$  – ймовірність того, що відповідно «гарний» і «поганий» клієнт дадуть відповіді  $x$ ;
- $L$  – втрати від того, що «гарний» клієнт буде класифікований як «поганий»;
- $D$  – втрати від класифікації «поганого» клієнта як «гарного».

Згідно формули повної ймовірності,  $p(x) = p(x|G)p_G + p(x|B)p_B$  можна визначити ймовірність того, що клієнт надасть відповіді  $x$ , і потім за теоремою Байєса знайти

$$q(G|x) = \frac{p(x|G)p_G}{p(x)}, q(B|x) = \frac{p(x|B)p_B}{p(x)},$$

тобто ймовірність того, що клієнт з відповідними відповідями буде «гарним» чи «поганим».

Середні втрати в підрахунках на одного клієнта становить

$$EL = L \sum_{x \in A_B} q(G|x)p(x) + D \sum_{x \in A_G} q(B|x)p(x) = L \sum_{x \in A_B} p(x|G)p_G + D \sum_{x \in A_G} p(x|B)p_B$$

і є мінімальним при виборі

$$A_G = \{x | Dp(x|B)p_B \leq Lp(x|G)p_G\}.$$

Якщо припустити нормальність розподілу  $p(x|G)$ ,  $p(x|B)$  з загальною коваріаційною матрицею, то отримаємо лінійне правило

$$A_G = \{x | w_1x_1 + \dots + w_nx_n > c\}.$$

Таке правило фактично означає, що кожному набору відповідей  $x$  присвоюється «рахунок»  $s(x)$  за формулою  $s(x) = w_1x_1 + \dots + w_nx_n$ , і означає, що  $s(x)$  містить достатньо інформації для класифікації класів клієнтів. Таким чином, розмірність задачі зменшується з  $n$ , з розподілами  $p(x|G)$ ,  $p(x|B)$ , до 1

з розподілами  $p(s|G)$ ,  $p(s|B)$ . Знаходження правила класифікації зводиться до пошуку оптимальної межі  $c$  із задачі оптимізації

$$L \sum_{s < c} p(s|G)p_G + D \sum_{s \geq c} p(s|B)p_B \rightarrow \min.$$

*Множинна регресія.* Інший спосіб отримати лінійну скорингову функцію – використовувати одну із різновидів лінійної регресії. В простому випадку визначається залежна змінна  $Y$ , що приймає значення 1 в випадку «гарного» клієнта та 0 в випадку «поганого», і потім використовується метод найменших квадратів для знаходження ваг  $w_i$ , що забезпечує найкращі передбачення значення  $Y$ . При використанні лінійної регресії фактично відбувається спроба зв'язати ймовірність дефолту  $p$  зі значеннями відповіді на питання лінійної функції  $p = w_0 + w_1X_1 + \dots + w_nX_n$ .

Тут є очевидна невідповідність: ліва частина представляє собою ймовірність і повинна вимірюватися від 0 до 1, тоді як права може приймати будь-які значення. Для подолання цієї проблеми, значення ймовірності замінюється деякою функцією від неї, так щоб ця функція була монотонною та приймала всі значення [2].

Логістична регресія замінює ймовірність дефолту на логарифм шансів дефолту:

$$\log\left(\frac{p}{1-p}\right) = w_0 + w_1X_1 + \dots + w_nX_n = s(X).$$

Таким чином, «рахунок» пов'язаний тут не з вірогідністю дефолту  $p(B|x)$ , а з відношенням  $p(B|x)/p(G|x)$ :

$$s(x) = \log \frac{p(B|x)}{p(G|x)}.$$

*Дискримінантний аналіз.* Використовуючи лінійні скорингові функції, можна оцінити ймовірність належності клієнта тому чи іншому класу, визначивши

$$S_1(x) = w_0^1 + w_1^1X_1 + \dots + w_n^1X_n,$$

$$S_2(x) = w_0^2 + w_1^2X_1 + \dots + w_n^2X_n,$$

... ..

$$S_m(x) = w_0^m + w_1^mX_1 + \dots + w_n^mX_n.$$

Далі необхідно просто вибрати той клас, якому відповідає найбільший рахунок. В випадку двох класів, це дає ту ж відповідь, що й лінійна регресія.

*Лінійне програмування.* Нехай є набір даних, що включають в себе відповіді на питання  $x_{1n}, \dots, x_{in}$ , для кожного із  $N$  клієнтів і індикатори наявності дефолту  $Y_i$ . Ідеально було б знайти лінійну функцію  $s(x)$ , яка повністю розділяла «гарних» і «поганих» клієнтів. Однак, так як це звичайно неможливо, будемо шукати  $s(x)$  з міркувань мінімізації помилки  $a_i$ :

$$\begin{cases} a_1 + \dots + a_N \rightarrow \min, \\ w_1 x_{i1} + \dots + w_m x_{im} \geq c - a_i, Y_i = 0, \\ w_1 x_{i1} + \dots + w_m x_{im} \leq c + a_i, Y_i = 1, \\ a_i \geq 0. \end{cases}$$

Тут мінімізація відбувається по змінним  $(w_1, \dots, w_n, c, a_1, \dots, a_N)$ . Можна також мінімізувати максимально можливу помилку, замінив все  $a_i$  на одну змінну  $a$ .

Переваги методу лінійного програмування складається з можливості включити в програму додаткові обмеження. Наприклад, якщо кредитний продукт спрямований на людей молодого віку, можна зробити додаткове обмеження на відповідну вагу  $w_i \geq w_i^0$ , так що скорингова функція буде віддавати перевагу людям цього віку [6].

*Генетичні алгоритми* представляють собою метод оптимізації, що не накладає стандартних обмежень на цільову функцію (гладкість, випуклість і т.д.). В використанні до задач скорингу генетичний алгоритм складається в генерації початкової множини скорингової функції, з наступним використанням до функцій з цього набору операцій «схрещування» і «мутації», а також виключенням найменш придатних функцій.

*Дерева класифікації* (рекурсивні алгоритми розбиття), на відміну від попередніх методів, не призначені для побудови скорингової функції, вони послідовно поділяють клієнтів на групи згідно однієї змінної так, щоб ці групи максимально можливо відрізнялися за величиною кредитного ризику. Даний процес триває до моменту, поки групи, що залишилися, не стають настільки малі, що наступне розбиття не призведе до значущих статистичних відмінностей у рівні ризику. Далі кожному «листу» дерева приписується певна категорія клієнтів (також йому може відповідати не певна категорія, а своя лінійна регресійна модель, що дозволяє спочатку грубо розділяти клієнтів на основні групи (рис. 1), а потім для кожної групи використовувати регресійну модель).



Рис.1. Дерево класифікацій

Метод найближчого сусіда (англ. k-nearest neighbor algorithm, *kNN*) – метод автоматичної класифікації об'єктів. Основним принципом методу найближчих сусідів є те, що об'єкт присвоюється до того класу, який є найбільш поширеним серед сусідів даного елемента. Клієнт класифікується залежно від того, до якого класу належить більшість його сусідів по обраній метриці (рис. 2).

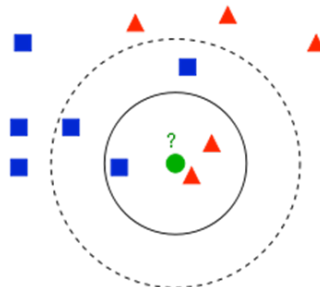


Рис. 2. Графічне представлення роботи методу *kNN*

Переваги методу - простота навчання моделі: при надходженні нових даних їх додають до існуючих, а по закінченні часу старі дані видаляють враховуючи зміну ситуації з часом.

Недолік - необхідність повного перегляду бази даних для класифікації нового клієнта (що може займати досить багато часу в разі великих обсягів інформації, скажімо, більш  $10^5 - 10^6$  записів). Крім того, вибір гарної метрики також є нетривіальним завданням, і якщо проводити його на основі історичних даних, то втрачається основна приваблива властивість алгоритму - відсутність необхідності попереднього аналізу даних та побудови моделі.

Нейронні мережі – це математичні моделі, а також їх програмні або апаратні реалізації, побудовані за принципом організації та функціонування

біологічних нейронних мереж - мереж нервових клітин живого організму. Нейронні мережі не програмуються в звичному сенсі цього слова, вони навчаються. Можливість навчання - одне з головних переваг нейронних мереж перед традиційними алгоритмами. У процесі навчання нейронна мережа здатна виявляти складні залежності між вхідними даними і вихідними, а також виконувати узагальнення. Це означає, що в разі успішного навчання мережа зможе повернути вірний результат на підставі даних, які були відсутні в навчальній вибірці, а також неповних та/або «зашумлених», частково перекручених даних.

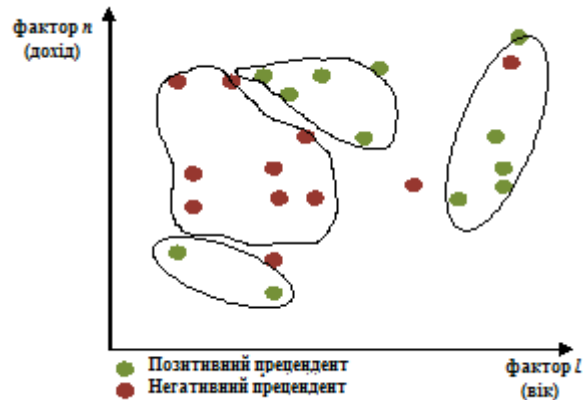


Рис. 3. Графічне представлення роботи нейронної мережі

Однак вони частіше застосовуються для скорингу юридичних осіб, ніж для скорингу фізичних осіб. Нейронна мережа дозволяє обробляти прецеденти навчальної вибірки з більш складним видом сегментів (рис. 3).

Недоліком нейронних мереж є те, що ваги зв'язків, визначені в результаті навчання, не мають ніякої інтерпретації в термінах кредитного ризику. Отже, практично неможливо пояснити пророкування, отримане за допомогою нейронної мережі, а також провести аналіз чутливості, щоб виділити найбільш значущі параметри.

*Байєсівські мережі* дозволяють представити багатомірні розподіли  $p(x|G)$ ,  $p(x|B)$  в вигляді комбінацій декількох більш низької розмірності. Про цьому в ході побудови моделі можуть бути враховані причино-наслідкові зв'язки.

Перевагою цього методу є можливість виводу по неповним даним. Якщо інформація про потенційного позичальника не є повною, то алгоритм виводу виявить ймовірність дефолту, базуючись лише на доступних даних [1].

*Логіко-ймовірнісний підхід* ґрунтується на поєднанні методів математичної логіки та теорії ймовірності. Цей підхід можна розглядати як спеціальний вид нелінійної регресії. В якості параметрів моделі тут виступають ймовірності дефолту, пов'язаного з тією чи іншою відповіддю на питання в анкеті. Оцінювальна величина – вірогідність дефолту. Для ідентифікації значень

параметрів використовується спеціально адаптовані алгоритми оптимізації [2].

Для ефективної оцінки позичальників, банківська установа найчастіше використовує дві або більше скорингових моделей. Тому дуже часто виникає проблема прийняття остаточного рішення, коли рішення моделей суперечить одна одній.

Існує безліч способів визначення ефективності моделей (загальна точність моделі, помилка I і II роду, ROC-крива і т.д.). Але не завжди вони є ефективними, тому банк найчастіше використовує експертну оцінку.

Як показує практика, експертна думка є доволі ефективним методом оцінки, але також має негативні сторони, що приводить до необ'єктивної оцінки клієнта:

- 1) у експертів мало часу на обдумування відповіді;
- 2) частина експертної групи може мати певний вплив на решту членів групи (Нерідко «криклива меншість» може подавити більшість, і ті вимушені будуть погодитися з цією пропозицією, розуміючи при цьому, що аргументи меншості помилкові);
- 3) неможливо перевірити достовірність оцінки окрім практики;
- 4) вплив особистих вподобань та упередженої думки експертів на прийняття рішення [4].

Тому було розроблено модель, що виключає вплив людського фактору, модель прийняття остаточного рішення.

**Результати дослідження.** Припустимо, що дві скорингові моделі ( $m = 2$ ), незалежно одна від одної, приймають рішення щодо надання кредиту, яке може знаходитися в одному з двох можливих станів ( $v = \{v_1, v_2\}$ ) з відомими ймовірностями  $p(v_m)$ ,  $p(v_1) + p(v_2) = 1$ . Нехай відомі ймовірності помилок першого та другого роду особистих рішень незалежних експертів  $p^{(i)}, i = \overline{1, 2}$ .

Тоді колективне рішення  $D = m, m = \overline{1, 2}$ , є оптимальним з точки зору мінімуму середньої ймовірності помилки на множині можливих ситуацій, а саме:

$$\Xi = \{S_{m_1 m_2} : (\delta_1 = m_1) \wedge (\delta_2 = m_2), m_1, m_2 = \overline{1, 2}\},$$

де  $v_i, i = \overline{1, 2}$  - функція, що показує яке рішення ( $v_m$ ) прийняла скорингова модель. Якщо в ситуації  $S_{m_1 m_2}, m_1 \neq m_2$  колективне рішення  $D = 1$  приймається на користь  $v_1$  за умови

$$p(v_1)p^{(m_2)}[1 - p^{(m_1)}] > p(v_2)p^{(m_1)}[1 - p^{(m_2)}] \quad (1)$$

і рішення  $D = 2$  на користь  $v_2$ , якщо

$$p(v_1)p^{(m_2)}[1 - p^{(m_1)}] < p(v_2)p^{(m_1)}[1 - p^{(m_2)}]. \quad (2)$$



Згідно з цим можна зробити висновок, що при виконанні нерівності в першому випадку, то кредит надається клієнту, якщо в другому – банк не надає кредит [3] – [4].

**Модельний приклад.** Розглянемо на прикладі одного банку алгоритм розв'язку поставленої задачі. Для цього візьмемо дві скорингові моделі, що використовуються в банку, а саме – логістичну модель та метод найближчих сусідів, проаналізуємо їх за допомогою ROC-аналізу, а також на практиці застосуємо модель остаточного прийняття колективного рішення.

До банку звернувся клієнт, що бажає отримати кредит. Заповнивши анкету, ми отримали всі необхідні дані для оцінки його кредитоспроможності.

- 1) Вік – 23 роки;
- 2) Стать – чоловіча;
- 3) Перебуває у шлюбі – так;
- 4) Кількість утриманців – один;
- 5) Сумарний дохід – 5000 гривень;
- 6) Досвід роботи – 3 роки;
- 7) Термін проживання в регіоні – 23 роки;
- 8) Ринкова вартість нерухомості у власності – 20 тис. доларів;
- 9) Щомісячна плата по кредиту – 750 гривень.

Розглянемо першу скорингову модель, логістичну модель. Вона представлена у вигляді лінійної функції:

$$\log\left(\frac{y}{1-y}\right) = -2,872 - 0.024x_1 + 0.7983x_2 - 0.1408x_3 - 1.5141x_4 + \\ + 0.0007x_5 + 0.0033x_6 + 0.0094x_7 + 0.0109x_8 + 0.0009x_9$$

де  $x_1$  – вік;  $x_2$  – стать;  $x_3$  – перебуває у шлюбі(та/ні);  $x_4$  – кількість утриманців;  $x_5$  – підтверджений сукупний дохід;  $x_6$  – досвід роботи (років);  $x_7$  – термін проживання в регіоні (років);  $x_8$  – ринкова вартість нерухомості у власності (тис. доларів);  $x_9$  – щомісячна плата по кредиту (грн.).

Для того, щоб отримати ймовірність необхідно проранжувати отриманні дані та, підставивши ці значення у функцію, отримаємо:

$$\log\left(\frac{p}{1-p}\right) = -2,872 - 0.024 \cdot 23 + 0.7983 - 0.1408 - 1.514 + \\ + 0.007 \cdot 500 + 0.0033 \cdot 3 + 0.0094 \cdot 23 + 0.0109 \cdot 20 + 0.0009 \cdot 750 = 0,47$$

Якщо ймовірність менша 0.5, то клієнт є кредитоспроможним.

Так як в банку використовують дві скорингові моделі, то необхідно перевірити його на другій, методом найближчого сусіда. Для цієї моделі використовують базу даних позичальників банку та порівнюють клієнта з найближчим із них. З точки зору математики, ця модель представляє собою вибірку, яку можна зобразити на площині (рис. 4).

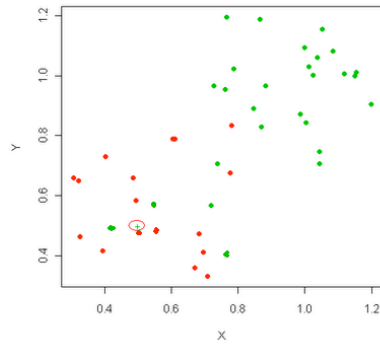


Рис. 4. Графічне представлення вибірки

На рис. 4 можна побачити, що «погані» позичальники позначені червоним, а «гарні» - зеленим. Згідно цього методу, клієнт відноситься до того класу, до якого належить найближчий сусід. А отже, клієнт не є кредитоспроможним.

В результаті оцінки позичальника скоринговими моделями отримуємо суперечність, для вирішення якої необхідно оцінити ефективність кожної з них та прийняти остаточне рішення щодо цього клієнта.

На основі минулого досвіду побудуємо *ROC* – криву для двох моделей та визначимо їх ефективність, а також визначимо специфічність та чутливість моделей.

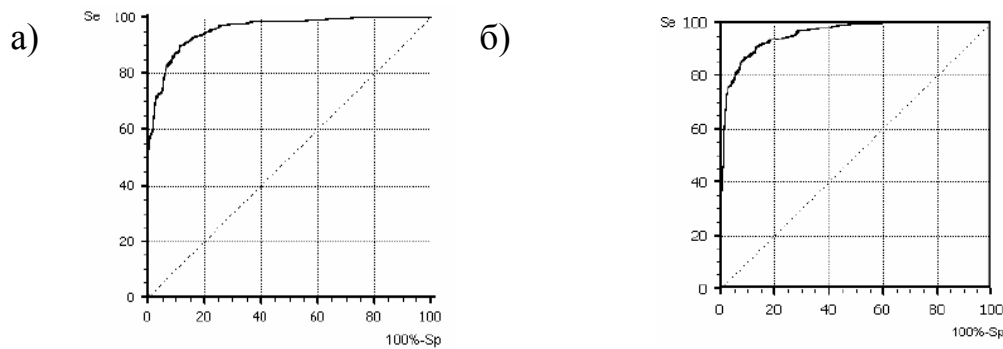


Рис.5. *ROC* – крива для: а) моделі логістичної регресії; б) моделі методом найближчого сусіда

Таблиця

Ефективність, чутливість та специфічність моделей

	Логістична модель	Метод найближчих сусідів
Ефективність:	0,958	0,961
Спецефічність (%)	88,6	91,1
Чутливість (%)	89,7	86,8

Як можна побачити з таблиці, дві моделі є доволі ефективними. Тому цю проблему необхідно вирішити завдяки моделі прийняття остаточного рішення. Перевіримо виконання нерівності (1)

$$0,53 \cdot 0,961 \cdot [1 - 0,958] > 0 \cdot 0,958[1 - 0,961],$$

і отримаємо, що  $0,214 > 0$ , тобто нерівність (1) виконується. Звідси робимо висновок про кредитоспроможність позичальника.

**Висновок.** У зв'язку з нестабільною ситуацією в країні, виникло зростання кредитних ризиків, пов'язаних з неповерненнями кредитів. Це призвело до підвищення процентних ставок за кредитами, що, в кінцевому рахунку знизило доступність кредитів. Тому розробники скорингових систем намагаються використати в своїх продуктах всі види моделей, що веде до ускладнення прийняття остаточного кредитного рішення у випадку протилежних оцінок.

Найбільш суттєвим науковим результатом, що відображає новизну та теоретичне значення дослідження є: оцінка позичальника на основі декількох моделей та у випадку протилежних рішень кожної моделі, остаточне рішення приймається ґрунтуючись на побудованій моделі прийняття колективного рішення. Окрім цього, отриманий результат може бути корисним підґрунтям для подальших теоретичних та практичних досліджень в даному напрямку.

### Література:

1. Churchill G. A., Nevin J. R., Watson R. R.: The role of credit scoring in the loan decision. Credit World. – March, 1977.; Myers J. H., Forgy E. W. The development of numerical credit evaluation systems // Journal of American Statistical Association. – September, 1963. – 15000 p – ISBN 0-87168-441-1.
2. Челноков В. А. Деньги, кредит, банки: учеб. пособие / В. А. Челноков. – М.: ЮНИТИ : ЮНИТИ–ДАНА, 2005. – 366 с. – Библиогр.: с. 362. – 30000 экз. – ISBN 5-238-00817-1.
3. Жуковская О. А., Файнзильберг Л. С. Интервальное обобщение байесовской модели принятия коллективногорешения в конфликтных ситуациях // Кибернетика и системный анализ. – 2005. – № 3. – 133 – 144 с.
4. Жуковская О. А. Интервальные модели принятия коллективных решений в конфликтных ситуациях //Международная конференция «Проблемы управления и приложения». Минск, 2005. . 32 с. – 1000 экз.– ISBN 5-215-01358-6.
5. Кім Дж. О. Факторний, дискриминантний і кластерний аналіз / Дж. О Кім, Ч. У. Мюллер. – М.: Фінанси й статистика, 1989. – 215 с. – 35000 пр. – ISBN 5-8135-0145-2.