

**ОБ'ЄДНАННЯ МОДЕЛЕЙ LOGIT-РЕГРЕСІЙ
ЯК КОМІТЕТУ ЕКСПЕРТІВ ДЛЯ ОЦІНКИ
КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА**

С. С. Савіна

Канд. екон. наук, доцент,
доцент кафедри економіко-математичного моделювання
Державний вищий навчальний заклад «Київський національний
економічний університет імені Вадима Гетьмана»
проспект Перемоги, 54/1, м. Київ, 03680, Україна
s.savina@mail.ru

В. П. Бень

Провідний спеціаліст
Акціонерне товариство «МОТОР СІЧ»
проспект Моторобудівників, 15, м. Запоріжжя, 69068, Україна
benchik1985@yandex.ru

Дослідження присвячено побудові економіко-математичних моделей оцінювання кредитоспроможності позичальників-фізичних осіб. Для розв'язання даної задачі застосовано різноманітний математичний інструментарій, зокрема нейромережі типу багаточаровий перцептрон та логістичні регресії.

В роботі запропоновано новий методологічний підхід до відбору найбільш значущих чинників, на основі яких будуються моделі оцінки кредитоспроможності позичальників. Підхід ґрунтується на синтезі ймовірнісної нейромережі та генетичного алгоритму. Також розроблено методологічний підхід до реалізації «комітету експертів» за рахунок об'єднання кількох logit-регресійних моделей оцінювання кредитоспроможності позичальника. Проведено порівняльний експериментальний аналіз ефективності застосування однієї узагальненої logit-регресії та трьох моделей logit-регресій (комітету експертів), параметри яких оцінені окремо на основі масивів даних, що сформовані за виділеними якісними критеріями.

Тестування побудованих економіко-математичних моделей підтвердило високий рівень їх адекватності, що свідчить про можливість на їх основі істотно підвищити точність оцінювання кредитоспроможності потенційного позичальника та мінімізувати рівень кредитного ризику банківської установи. В результаті проведення модельних експериментів одержано висновок щодо

доцільності використання неймережі типу багат шарового перцептрон у разі необхідності розподілу позичальників лише за двома класами — надійних чи ненадійних; якщо ж дослідження передбачає необхідність більшої деталізації в оцінці кредитоспроможності, наприклад, визначення рівня ризикованості надання кредиту конкретній особі, то слід застосовувати комітет експертів (поєднання *logit*-регресій).

Результати дослідження можуть бути використані банківськими установами та всіма структурами, що можуть бути зацікавленими в адекватній процедурі оцінки кредитоспроможності фізичних осіб.

Ключові слова. *Оцінка кредитоспроможності, позичальник— фізична особа, якісні та кількісні чинники кредитного ризику, неймережі, logit-регресія, комітет експертів.*

ОБЪЕДИНЕНИЯ МОДЕЛЕЙ LOGIT-РЕГРЕССИЙ КАК КОМИТЕТА ЭКСПЕРТОВ ДЛЯ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ ЗАЕМЩИКА

С. С. Савина

Кандидат экономических наук, доцент,
доцент кафедры экономико-математического моделирования
Государственное высшее учебное заведение «Киевский национальный
экономический университет имени Вадима Гетьмана»
проспект Победы, 54/1, м. Киев, 03680, Украина
s.savina@mail.ru

В. П. Бень

Ведущий специалист
Акционерное общество «МОТОР СИЧ»
проспект Моторостроителей, 15, г. Запорожье, 69068, Украина
benchik1985@yandex.ru

Исследование посвящено построению экономико-математических моделей оценки кредитоспособности заемщиков-физических лиц. Для решения данной задачи применен разнообразный математический инструментарий, в частности нейросети типа многослойный перцептрон и логистические регрессии.

В работе предложен новый методологический подход к отбору наиболее значимых факторов, на основе которых строятся моде-

ли оценки кредитоспособности заемщиков. Подход основывается на синтезе вероятностной нейросети и генетического алгоритма.

Также разработан методологический подход к реализации «комитета экспертов» за счет объединения нескольких логит-регрессионных моделей оценки кредитоспособности заемщика. Проведен сравнительный экспериментальный анализ эффективности применения одной обобщенной логит-регрессии и трех моделей логит-регрессий (комитета экспертов), параметры которых оценены отдельно на основе массивов данных, сформированных по выделенным качественным критериям.

Тестирование построенных экономико-математических моделей подтвердило высокий уровень их адекватности, что свидетельствует о возможности на их основе существенно повысить точность оценки кредитоспособности потенциального заемщика и минимизировать уровень кредитного риска банковского учреждения. В результате проведения модельных экспериментов был получен вывод о целесообразности использования нейросети типа многослойный перцептрон в случае необходимости распределения заемщиков только по двум классам — надежных или ненадежных; если же исследование предполагает необходимость большей детализации в оценке кредитоспособности, например, определение уровня рискованности предоставления кредита конкретному лицу, то следует применять комитет экспертов (сочетание логит-регрессий).

Результаты исследования могут быть использованы банковскими учреждениями и всеми структурами, заинтересованными в адекватной процедуре оценки кредитоспособности физических лиц.

Ключевые слова. Оценка кредитоспособности, заемщик—физическое лицо, качественные и количественные факторы кредитного риска, нейросети, логит-регрессия, комитет экспертов.

INTEGRATION OF MODELS OF LOGIT-REGRESSIONS AS A COMMITTEE OF EXPERTS TO ASSESS THE CREDITWORTHINESS OF BORROWER

Svitlana Savina

PhD (Economic Sciences), Docent,
Associate Professor of Department of Economic and Mathematical Modeling

State Higher Educational Establishment
«Kyiv National Economic University named after Vadym Hetman»
54/1 Peremogy Avenue, Kyiv, 03680, Ukraine
s.savina@mail.ru

Vladyslav Ben'

Leading Specialist
The company "MOTOR SICH"

15 Motorobudivelnykiv Avenue, Zaporizhzhia, 69068, Ukraine
benchik1985@yandex.ru

The research is dedicated to the formation of economic and mathematical models of creditworthiness assessment of borrower-individual. To resolve this task different mathematical set of tools is applied, in particular neuronets of multilayer perceptron type and logistic regressions.

In this work new methodological approach to the selection of most important factors is proposed, on the basis of which new models of creditworthiness assessment of individuals are formed. This approach is based on synthesis of probabilistic neuronet and genetic algorithm.

Also, methodological approach to realization of "experts committee" due to merging of several *logit*-regression models of creditworthiness assessment of borrower-individual is developed. The comparative experimental analysis of effectiveness of implementation of one extended *logit*-regression and three *logit*-regression models (experts committee) is conducted, parameters of which were evaluated separately, based on data set formed according to distinguished qualitative criteria. Testing of formed economic-mathematical models confirmed the high level of their adequacy, that affirms about the ability on their basis to increase essentially the accuracy of creditworthiness assessment of potential borrower and minimize credit risk level of bank institution. As a result of implementation of modeling experiments the conclusion is received concerning the expediency of using neuronet of multilayer perceptron type in case of necessity of borrower division into two classes — reliable or unreliable; and if the research anticipates the necessity higher particularization in evaluation of creditworthiness, for instance, determination of risk level of giving credit to a particular person, then the expert committee (combination of *logit*-regressions) should be applied.

The investigation results can be used by bank institutions and all organizations that might be interested in adequate procedure of creditworthiness assessment of borrower-individuals.

Keywords. *Assessing the creditworthiness, an individual borrower, qualitative and quantitative factors of credit risk, neural network, logit-regression, the committee of experts.*

Постановка проблеми

Проблема кількісної оцінки та аналізу кредитних ризиків позичальників є актуальною як для закордонних, так і для вітчизняних банків, що займаються кредитуванням фізичних та юридичних осіб. До методик, які розробляються для кількісної оцінки кредитних ризиків, висувуються ряд умов, серед яких особлива вимога щодо прозорості [1].

Прозорість методики оцінки кредитного ризику — це можливість бачити не тільки явище в цілому, але і його деталі. Під прозорістю методики розуміють строгість використовуваних математичних методів, згладжування суб'єктивності експертних оцінок, наочність результатів оцінки та аналізу ризику, повне їх розуміння самими працівниками банків, відкритість методик для контролюючих органів і позичальників. Прозорість методики та результатів досягається обчисленням вкладів ініціюючих подій (критеріїв) до кредитного ризику. Отже, виникає необхідність застосування такого математичного апарату, що дасть змогу, по-перше, зменшити вплив суб'єктивного чинника при оцінці кредитоспроможності клієнтів, та, по-друге, визначити вплив кожного з факторів, що враховуються при аналізі, на кінцеву оцінку кредитного ризику.

Математична постановка задачі оцінки кредитоспроможності позичальника є задачею дослідження зміни деякої величини, що характеризує його кредитний ризик, під впливом зміни значень визначеного переліку чинників. У такій постановці маємо класичну задачу економетрики. Економетричні методи та моделі детально розроблені, обґрунтовані та мають широке розповсюдження в дослідженнях економічних процесів.

У задачі оцінки кредитоспроможності позичальника необхідно встановити взаємозв'язок між певним переліком чинників і фактом повернення чи неповернення кредиту позичальником. Змінна, що вказує на факт виконання кредитних зобов'язань (повернення чи неповернення кредиту), може бути визначена за допомогою лише двох значень, зазвичай 0 та 1 (така змінна має назву бінарної). Таким чином приходимо до задачі моделювання бінарної змінної. Побудова звичайної множинної регресії в даній задачі не дасть потрібного результату, оскільки розраховані значення залежної змінної можуть не належати відріzkу $[0, 1]$ і інтерпретація таких результатів ускладнюється. Однак задачу по-

будови регресійної залежності в даному випадку можна представити не як передбачення значень бінарної змінної, а як моделювання деякої неперервної змінної, яка приймає значення з інтервалу $[0, 1]$. Такі задачі можуть бути описані за допомогою лінійних моделей ймовірності, або *logit*- та *probit*-моделей [2]. Завдяки способу побудови вказаних моделей прогнознi значення, які приймає досліджувана змінна, можуть не лише відповідати значенню 0 та 1, а бути інтерпретованими як ймовірність повернення (неповернення) кредиту позичальником.

Аналіз останніх джерел і публікацій

Проблеми оцінки кредитоспроможності позичальників, аналізу та оцінки кредитного ризику, а також створення системи ризик-менеджменту банків не втрачають своєї актуальності, тому дослідженням за вказаними напрямками присвячено багато наукових праць. Теоретико-методологічна база для аналізу та дослідження кредитного ризику розвинута вітчизняними вченими Вітлінським В. В. [3], Камінським А.Б. [4—6], Кишакевичем Б. Ю. [7—9], Пернарівським О. В. [10], Писанцем К. К. [11, 12]. Сучасний математичний інструментарій, зокрема методи нечіткої логіки та нейронних мереж, в управлінні діяльності комерційного банку застосовано та розвинуто в роботі Великоіваненко Г. І., Трокоз Л. О. [13].

Незважаючи на значну кількість досліджень, які стосуються моделювання кредитних ризиків, сьогодні невирішеною залишається низка задач, присвячених методам оцінювання кредитоспроможності позичальників-фізичних осіб, серед яких можна виділити проблеми формування та аналізу початкового масиву даних, відбору найбільш значущих факторів кредитного ризику, а також вибору найбільш адекватного математичного інструментарію відповідно до наявної інформаційної бази. Одним з важливих невирішених питань при оцінці кредитоспроможності позичальників-фізичних осіб автори вважають розробку адекватного підходу врахування чинників, що описуються якісними показниками.

Мета і завдання дослідження

Метою роботи є розробка адекватного математичного інструментарію оцінювання кредитоспроможності фізичних осіб-позичальників комерційних банків. Головними завданнями дослі-

дження є розробка методологічного підходу відбору з множини наявних чинників кредитного ризику найбільш важливих, а також побудова та застосування системи моделей, здатних адекватно враховувати як кількісні, так і якісні чинники.

Виклад основного матеріалу

Оцінка кредитоспроможності позичальника—фізичної особи проводиться у банку на основі інформації, що характеризує здатність клієнта отримувати дохід, якого буде достатньо для своєчасної сплати кредитних зобов'язань, а також наявність грошових та майнових активів, які за необхідності можуть слугувати забезпеченням виданого кредиту тощо. Крім того, при наданні кредиту необхідно також враховувати постійні зміни ринкової кон'юнктури, можливі ризики, що матимуть найбільший вплив на здатність клієнта розрахуватись за кредитними зобов'язаннями, а також економічні, соціальні та інші чинники [14].

Як правило чинники, які включаються до аналізу кредитних ризиків позичальників—фізичних осіб, складаються із соціальних, професійних, майнових і спеціальних банківських. До соціальних відносяться вік, сімейний стан, кількість членів сім'ї та кількість працюючих членів сім'ї. До професійних відносяться сфера діяльності, кваліфікація, посада, загальний стаж роботи, стаж на останньому місці роботи тощо. Майнові чинники характеризують фінансовий стан потенційного позичальника та включають дохід, додатковий дохід, розмір середнього залишку вкладу, стан рахунку, якість погашення попередніх кредитів, загальна сума збережень, наявність нерухомості, спосіб володіння майном тощо. Остання група чинників описує спеціальні банківські чинники, які визначаються кожним банком окремо і можуть включати, наприклад, час, протягом якого діє рахунок у даному комерційному банку, наявність випадків розглядання суперечок у судових інстанціях тощо.

При побудові економіко-математичної моделі оцінки кредитоспроможності необхідно обробляти досить значні обсяги інформації — кількість чинників може сягати кількох десятків, а перелік позичальників кількох тисяч і більше. Обробка таких масивів інформації є складним завданням. Неможливо однозначно виділити набір чинників, які будуть однаково важливими для розв'язання задачі класифікації позичальників за рівнем кредито-

спроможності. В наявній інформації завжди присутні як важливі дані, так і такі, цінність яких сумнівна.

Крім того, значна частина даних, що включається до розгляду при оцінці кредитоспроможності, представлена якісними показниками, які для врахування у математичних моделях необхідно перетворювати в кількісні. В більшості випадків якісні показники описують за допомогою фіктивних змінних. Для випадку, коли якісна змінна характеризується лише двома варіантами (наприклад, стать позичальника), вона може бути описана одною фіктивною змінною, що приймає значення 0 та 1 і кодує початкові якісні величини за принципом: 0 — дана якісна ознака відсутня, 1 — присутня.

Для випадку, коли якісна змінна описує кілька варіантів, введення фіктивних змінних є більш складною процедурою. Наприклад, такий показник, як освіта може приймати значення «середня», «середня спеціальна», «вища». Перетворення такої інформації в кількісну за допомогою однієї фіктивної змінної можливе шляхом кодування типу: «середня» — 1, «середня спеціальна» — 2, «вища» — 3. Але за такого кодування може виникати некоректне впорядкування значень. Автоматично встановлюється, що «середня спеціальна» освіта є гіршою за «вищу», оскільки кількісна оцінка однозначно фіксує перевищення однієї величини над іншою. Така ситуація не завжди відповідає дійсності і не для всіх випадків можна вважати такі заміни коректними. Більш точним у даному випадку є спосіб кодування якісних параметрів за допомогою кількох фіктивних змінних, кількість яких відповідає всім можливим варіантам, які описує якісна величина. Таким чином початковий масив інформації, що налічує десятки параметрів для дослідження, може зростати в кілька разів.

Також варто зважати на те, що за наявності великих сукупностей параметрів одна й та сама інформація може повторюватись у різних змінних, через що вони можуть бути сильно корельовані між собою. В таких випадках доцільно залишити тільки частину з кількох корельованих змінних. Однак у зв'язку з існуванням складних внутрішніх залежностей між досліджуваними чинниками може виникнути ситуація, коли дві чи більше взаємопов'язаних змінних разом несуть суттєву інформацію, яка не міститься в жодній з них окремо. Тому іноді доцільно включити ті вхідні змінні, які містять навіть незначну інформацію.

Зважаючи на вищенаведене, відбір початкових даних при розв'язанні задачі оцінки кредитоспроможності позичальника є важливим і досить складним завданням. Найбільш ефективним способом формування оптимального набору вхідних показників є повний перебір усіх можливих комбінацій чинників і можливих різновидів моделей, що на практиці реалізувати неможливо.

У такому разі першочерговим завданням в розв'язанні задачі оцінки кредитоспроможності позичальника стає пошук адекватного математичного інструментарію, а також зниження розмірності початкової сукупності даних.

Більшість соціально-економічних процесів є суттєво нелінійними, тому для їх дослідження доцільно використовувати відповідний математичний апарат. Саме таким інструментом для розв'язання задачі пошуку складних нелінійних функціональних залежностей у структурі пояснюючих змінних, що визначають кредитний ризик позичальників, є ймовірнісні нейронні мережі (*PNN — Probabilistic Neural Network*) [15]. Такі мережі, завдяки своїй радіально-базисній структурі, ефективно описують нелінійні залежності та дозволяють працювати з даними, що мають значний рівень шуму.

Побудова *PNN*-моделей при розв'язуванні задач класифікації має в основі ідею оцінки щільності ймовірності за допомогою ядерних функцій. Ймовірнісна нейронна мережа має не менше трьох шарів: вхідний, радіальний і вихідний. У вхідному шарі кількість нейронів відповідає кількості чинників, за якими буде проводитись класифікація. Кількість радіальних елементів обирається відповідно до обсягу навчальної вибірки. Вхідний шар і шар радіальних елементів утворюють повнозв'язну структуру. Кожному класу відповідає один вихідний елемент, який пов'язаний з усіма радіальними елементами, що відносяться до даного класу. Формальним правилом проведення класифікації є порівняння щільностей розподілу ймовірностей. Для оцінки функції щільностей розподілу ймовірностей, як правило, застосовують метод Парцена, згідно якого до кожного початкового елементу представляється деяка вагова функція, яка має назву функції потенціалу чи ядра. У більшості випадків за ядерні функції береться спрощена функція Гауса. В результаті значення вихідних сигналів будуть пропорційними ядерним оцінкам ймовірності належності відповідним класам конкретного вхідного елементу.

До найважливіших переваг ймовірнісної мережі слід віднести той факт, що вихідне значення має ймовірнісний зміст і є зручним для інтерпретації фахівцями банківських установ ризику невиконання позичальником кредитних зобов'язань. Крім того, така мережа дуже швидко навчається.

Ефективним засобом оптимізації параметрів нейронних мереж є генетичні алгоритми, що є адаптивним пошуковим методом, який спирається на селекцію найкращих елементів у популяції за аналогією з природним відбором і наслідуванням. Елементами в популяції у нашій задачі є перелік вхідних факторів, за якого точність класифікації позичальників буде найбільшою.

При застосуванні генетичного алгоритму для відбору вхідних факторів виконується велика кількість експериментів з різними комбінаціями пояснюючих змінних. Для кожної комбінації будується ймовірнісна мережа і оцінюються її результати. Процедура з використанням методів генетичних алгоритмів повторюється ітераційно для пошуку найбільш релевантного набору вхідних показників, які застосовуються на наступних етапах дослідження [16].

Сутність методологічного підходу до відбору чинників для оцінювання кредитоспроможності позичальників, заснованого на поєднанні роботи нейромереж і генетичного алгоритму, полягає в такому. З повного переліку факторів впливу обирається випадковий набір чинників. Масив відібраних факторів кодується послідовністю з нулів та одиниць. Одиниця відповідає включенню до розгляду відповідного фактора, а при позначенні нулем відповідний фактор не враховується при побудові моделі. Наприклад, при дослідженні масиву з одинадцяти факторів, кодування рядка у двійковому форматі виду 00000000011 буде означати, що в модель буде включено лише два останні чинники.

Спочатку з повного масиву спостережень, що складається з інформації щодо позичальників, випадковим чином обирається набір рядків, які відповідають заданому шаблону, що в термінах теорії генетичних алгоритмів називається популяцією. Отримані спостереження використовуються для навчання та перевірки адекватності ймовірнісної нейронної мережі. Для оцінювання ефективності мережі визначається похибка моделювання — середньоквадратичне відхилення значень різниці фактичних і прогнозованих моделлю величин вихідного показника, який у даному випадку визначає належність до певного класу пози-

чальників. На основі таких похибок визначаються набори входних факторів, за яких нейромережа демонструє кращий результат класифікації.

Далі комбінації факторів, за якими була отримана низька точність класифікації, виключаються з розгляду, а з тих, що залишились, породжуються нові двійкові комбінації за допомогою штучних генетичних операцій мутації та схрещування, утворюючи нову популяцію, для якої також застосовується нова нейромережа та визначається точність класифікації за відібраним набором чинників. Після реалізації описаної процедури згідно запропонованого методологічного підходу обирається найкращий представник останнього покоління, а саме та двійкова комбінація факторів, врахування яких у нейромережі дає найбільш адекватний результат класифікації.

Статистичним джерелом проведеного дослідження є дані з кредитних заявок позичальників-фізичних осіб комерційного банку та відомостей щодо виконання ними зобов'язань за отриманими кредитами. Представлена інформація містить дані за 11 чинниками та нараховує 2175 спостережень.

З 11 чинників три є якісними (рівень освіти, статус працюючого, сфера діяльності), всі інші — кількісні показники (вік, стаж на останньому місці роботи, загальний стаж роботи, дохід, наявність депозитів, наявність виплачених у минулому кредитів, розмір сім'ї, кількість дітей).

Введемо позначення відібраних для аналізу чинників: *Var1* — вік позичальника, *Var2* — сфера діяльності, *Var3* — рівень освіти, *Var4* — стаж на останньому місці роботи, *Var5* — загальний стаж роботи, *Var6* — дохід, *Var7* — наявність депозитів, *Var8* — наявність раніше виплачених кредитів, *Var9* — розмір сім'ї, *Var10* — кількість дітей, *Var11* — статус працівника.

Описаний методологічний підхід до відбору найбільш значущих чинників для оцінювання кредитоспроможності позичальників реалізований засобами пакету *ST Neural Networks*.

Кількість можливих комбінацій факторів, що породжуються операціями мутації та схрещування, для масиву з 11 показників має велику розмірність, тому в табл. 1 представлено початковий фрагмент реалізації процедури перебору показників із відповідною вказівкою похибки моделювання ймовірнісною нейронною мережею.

Таблиця 1

ДОСЛІДЖЕННЯ ТОЧНОСТІ МОДЕЛЮВАННЯ КРЕДИТНОГО РИЗИКУ
ЗА РІЗНИХ ВАРИАНТІВ НАБОРУ ВХІДНИХ ФАКТОРІВ НЕЙРОМЕРЕЖІ

Похибка	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
1,951962	0	0	0	0	0	0	0	1	0	1	1
1,981826	0	0	0	0	0	0	0	0	1	0	1
1,982324	0	0	0	0	0	0	0	0	1	1	1
1,951611	0	0	0	0	0	0	0	1	0	0	1
1,950465	0	0	0	0	0	0	0	1	1	0	1

У табл. 1 у першому стовпчику відображено, яка похибка моделі властива відповідній комбінації включених факторів. Процедура відбору факторів продовжується на основі операцій мутації та схрещування в напрямку відбору комбінацій факторів, які покращують якість моделі, та в результаті обирається модель з найменшою похибкою. Кінцевий результат відбору наведено у табл. 2.

Таблиця 2

ПЕРЕЛІК ВХІДНИХ ФАКТОРІВ МОДЕЛІ ОЦІНКИ КРЕДИТНОГО РИЗИКУ,
ВІДБРАНИХ ІЗ ЗАСТОСУВАННЯМ ГЕНЕТИЧНОГО АЛГОРИТМУ

Похибка	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
1,907357	1	0	1	0	1	0	1	1	0	0	1

У результаті проведеного відбору визначено перелік найбільш впливових чинників, які дозволяють зменшити розмірність масиву даних, не втрачаючи при цьому ефективності моделі. Як видно з табл. 2, при безпосередній побудові математичних моделей оцінювання кредитоспроможності фізичних осіб необхідно використовувати як кількісні показники: вік, загальний стаж роботи, наявність депозитів і наявність раніше виплачених кредитів, так і якісні: рівень освіти та статус працівника.

Разом з описаним вище підходом до відбору вхідних факторів пропонується також застосування більш спрощених порівняно з генетичним алгоритмом процедур. На основі застосуван-

ня ймовірнісної мережі пропонується використовувати послідовні алгоритми з включенням чи виключенням чинників. Алгоритм послідовного включення починає з пошуку однієї вхідної змінної, яка найбільше визначає поведінку вихідної змінної. Далі визначається друга змінна, яка при додаванні її до першої максимально підвищує адекватність моделі. Процедура припиняється, коли або вичерпано всі змінні, або адекватність моделі перестає зростати.

Алгоритм послідовного виключення — це зворотній процес. Він починається з моделі, яка містить всі змінні і потім вилучає їх по одній, на кожному кроці визначаючи ту змінну, яка при вилученні мінімально знижує точність класифікації моделлю.

Результат застосування зазначених алгоритмів для досліджуваного масиву даних наведено в табл. 3.

Таблиця 3

**ПЕРЕЛІК ВХІДНИХ ФАКТОРІВ МОДЕЛІ ОЦІНКИ КРЕДИТНОГО РИЗИКУ,
ВІДБРАНИХ ЗА ДОПОМОГОЮ АЛГОРИТМІВ ВКЛЮЧЕННЯ
ТА ВИКЛЮЧЕННЯ ЗМІННИХ**

Алгоритм послідовного включення змінних											
Похибка	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
1,996958	1	0	0	1	1	0	0	1	0	1	1
Алгоритм послідовного виключення змінних											
Похибка	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
1,993235	1	0	1	1	1	1	1	1	0	1	1

Узагальнення проведеного за трьома підходами аналізу дозволяє виділити з одинадцяти початкових пояснюючих чинників шість, які доцільно залишити як вхідні змінні моделі оцінювання кредитоспроможності позичальників, а саме: вік (позначимо через x_1), стаж на останньому місці роботи (x_2), загальний стаж (x_3), наявність депозитів (x_4), наявність виплачених в минулому кредитів (x_5) і кількість дітей у сім'ї (x_6).

З метою підтвердження доцільності проведення описаної вище процедури попереднього відбору чинників порівнюються результати оцінки кредитоспроможності за допомогою двох нейромереж типу багат шаровий перцептрон. Перша використовує шість відібраних чинників, а друга мережа використовує всі початкові одинадцять чинників на тому самому масиві значень. Розрахунки проведено за допомогою пакету *STATISTICA Neural Networks*.

На рис. 1 наведено архітектуру та основні характеристики мережі для оцінки кредитоспроможності позичальників, яка використовує шість чинників і містить один прихований шар з 7 нейронів.

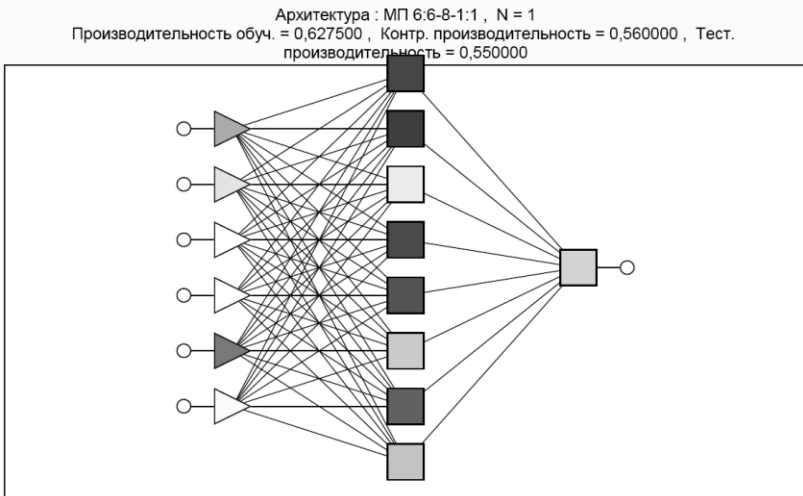


Рис. 1. Схема архітектури нейромережі для шести відібраних чинників кредитного ризику позичальників та її основні характеристики

До основних характеристик нейронної мережі в *ST Neural Networks* відносяться показники, які мають назву «продуктивність» і визначають частку вірно класифікованих прикладів окремо для навчальної, контрольної та тестової вибірок. Для наведеної на рис. 1 мережі точність класифікації для навчальної вибірки становила 62,8 %, для контрольної — 56 %, для тестової — 55 %.

Для того самого масиву даних, що містить всі 11 чинників, використовується багат шаровий перцептрон з одним прихова-

ним шаром, який складається з 16 нейронів. На рис. 2 наведено архітектуру та основні характеристики цієї нейронної мережі.

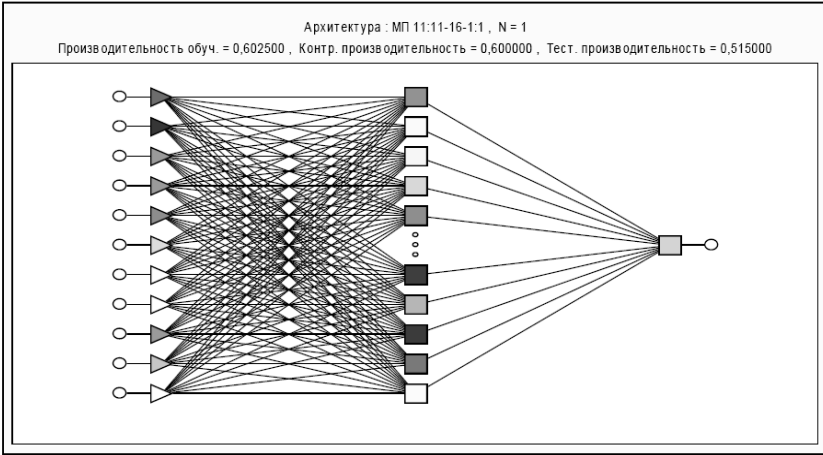


Рис. 2. Схема архітектури нейромережі для одинадцяти чинників оцінки кредитного ризику позичальників та її основні характеристики

Для наведеної на рис. 2 нейронної мережі точність класифікації на навчальній вибірці 60 %, контрольній — 60 %, тестовій — 51,5 %.

Аналіз характеристик адекватності двох побудованих мереж демонструє, що точність класифікації для навчальної та тестової вибірок на основі нейромережі, яка побудована з використанням меншої кількості чинників, є вищою порівняно з тою, яка містить всі одинадцять показників. Це є суттєвим результатом, оскільки друга нейронна мережа має майже на порядок більше параметрів за першу, а, відповідно, має бути здатною до більш ефективного виявлення складних нелінійних закономірностей. Отже, результати проведених експериментів засвідчують ефективність і доцільність застосування запропонованого методологічного підходу відбору чинників до моделі оцінювання кредитоспроможності позичальників—фізичних осіб.

Зауважимо, що застосування нейронної мережі у поєднанні з генетичними алгоритмами дозволило автоматизувати процес відбору факторів до моделі. Однак для забезпечення вимоги щодо прозорості методики оцінки кредитного ризику, про що йшлося

на початку статті, модель оцінки кредитоспроможності фізичних осіб має ґрунтуватись на економетричному інструментарії. Відповідно, відібрані на основі запропонованого методологічного підходу вхідні показники застосовуються для подальшої побудови *logit*- та *probit*-моделей оцінки кредитного ризику. Розв'яжемо задачу оцінки кредитоспроможності позичальників-фізичних осіб комерційного банку шляхом побудови лінійних моделей імовірності.

Позначимо змінну, яка відповідає стану повернення (неповернення) кредиту, через *Bad*. Змінна приймає значення 0, якщо кредит повернуто вчасно, та 1 — якщо зобов'язання за кредитом не виконано. Стосовно кожного клієнта відома інформація за *m* показниками. Вся сукупність інформації в результаті утворює масив $X = (x_{ji})$, ($j = \overline{1, m}; i = \overline{1, n}$). Індекс *j* визначає номер показника, що обрано для дослідження ($j = \overline{1, m}$) *i* — номер позичальника ($i = \overline{1, n}$), *n* — кількість спостережень в навчальній вибірці.

Класична лінійна модель регресії має вигляд

$$Bad_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + u_i, \quad i = \overline{1, n}, \quad (1)$$

де $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_m)$ — шукані параметри залежності, u_i — стохастична складова моделі.

Позначимо через $P_1 = P(Bad=1)$ ймовірність того, що величина *Bad* приймає значення одиниці. У такому разі модель (1) може бути записана у вигляді:

$$P_1 = P(Bad=1) = X\beta'. \quad (2)$$

Модель (2) називається лінійною моделлю ймовірності [2]. Головним недоліком у такій моделі є припущення лінійної залежності P_1 від β . Цей недолік усувають при використанні *logit*- і *probit*-моделей. Припускають, що

$$P_1 = F(X\beta'), \quad (3)$$

де *F* — деяка функція, область значень якої знаходиться на відрізьку [0, 1].

Функцією F може слугувати функція розподілу деякої випадкової величини. Тоді модель (3) можна представити, наприклад, у такій інтерпретації:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + u_i, \quad i = \overline{1, n}, \quad (4)$$

де y_i — деяка кількісна змінна, що має лінійний регресійний зв'язок із незалежними змінними X , для якого виконуються умови, що залишки моделі незалежні і однаково розподілені зі сталою дисперсією D та $M(u) = 0$.

За функцію F найбільш часто використовуються два типи функцій:

— функція логістичного розподілу

$$F(y_i) = \frac{e^{y_i}}{1 + e^{y_i}}, \quad (5)$$

за якої відповідна модель називається *logit*-моделлю,

— функція стандартного нормального розподілу

$$F(y_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-\frac{t^2}{2}} dt \quad (6)$$

при застосуванні якої відповідна модель називається *probit*-моделлю.

Для знаходження оцінок параметрів β моделі (3) використовується метод максимальної правдоподібності.

Логістичний розподіл досить схожий на нормальний і однозначний вибір щодо застосування одного з видів моделей є складним питанням. Відомо, що для малих вибірок і вибірок з незначним розкидом незалежних змінних висновки за *logit*- і *probit*-моделями будуть майже співпадати.

При оцінюванні якості моделі потрібно перш за все звертати увагу на значущість моделі та оцінок параметрів. Оцінити якість отриманої моделі можна також на основі параметру, який має назву відношення неузгодженості та обчислюється як відношення добутку кількості правильно класифікованих надійних і ненадійних позичальників до добутку неправильно класифікованих. Відношення неузгодженості повинно перевищувати одиницю.

Оскільки модель (3) нелінійна за параметрами β , то їх інтерпретація відрізняється від лінійного випадку. Диференціюючи (2) по X , отримуємо:

$$\frac{\partial P(\text{Bad} = 1)}{\partial X} = F'(X\beta')\beta = p(X\beta')\beta. \quad (7)$$

Таким чином, граничний ефект для кожного чинника x_j ($j = \overline{1, m}$) є змінною величиною і залежить від інших чинників. При використанні даної моделі певне уявлення щодо ефекту впливу незалежних змінних можна отримати, обчислюючи похідні (7) для середніх значень цих змінних.

Для випадку *logit*-моделі, коли використовується функція логістичного розподілу (5), щільність $p(X\beta')$ функції розподілу $F(X\beta')$ визначається за формулою:

$$p(X\beta') = p(y_i) = \frac{e^{y_i}}{(1 + e^{y_i})^2}. \quad (8)$$

Отже, для визначення наближеної оцінки граничного ефекту впливу кожного чинника x_j ($j = \overline{1, m}$) потрібно скористатись формулою:

$$\Delta x_j = \frac{(e^{\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \dots + \beta_m \bar{x}_m}) \beta_j}{(1 + e^{\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \dots + \beta_m \bar{x}_m})^2} \quad (j = \overline{1, m}), \quad (9)$$

де Δx_j — граничний вплив кожного чинника x_j , \bar{x}_j — середнє значення для кожного чинника x_j .

Як зазначалось вище, на етапі проведення попередньої обробки даних виникає необхідність включення в процес оцінки кредитоспроможності позичальників переліку якісних чинників. Проте безпосереднє їх застосування в *logit*- і *probit*-моделях неможливе.

З метою подолання цього обмеження автори пропонують застосувати такий методологічний підхід. Початкова сукупність

даних, яка містить значення кількісних та якісних чинників, розділяється на групи за якісними показниками. Тобто, з початкового масиву даних виділяються окремі масиви даних, що відповідають різним рівням кожного якісного показника. Наприклад, одним з якісних показників, що враховується при оцінюванні кредитоспроможності фізичних осіб, є «рівень освіти», який, у свою чергу, поділяється на два підрівні: «одна чи більше вищих освіти» та «середня і середня спеціальна освіта». Спочатку з усього масиву даних обираються лише ті записи, які відповідають позичальникам з певним підрівнем даного якісного показника, тобто мають одну чи більше вищих освіти. На основі відібраних даних знаходяться оцінки параметрів *logit (probit)*-регресії. Аналогічним чином оцінюються параметри іншої *logit (probit)*-регресії для другого підрівня якісного показника «рівень освіти».

Процедура продовжується з метою побудови низки *logit (probit)*-регресій, що описують всі рівні найбільш вагомих якісних показників. Кожна з моделей має характеристики, що дозволяють оцінити її адекватність. Таким чином, у результаті застосування описаного методологічного підходу утворюється набір найбільш адекватних моделей оцінювання кредитного ризику позичальника-фізичної особи відповідно до відібраних якісних характеристик. Проведення оцінки кредитоспроможності позичальника за таким підходом буде здійснюватись на основі не однієї моделі, а об'єднання результатів розрахунків кількома подібними моделями.

Подібний підхід застосовувався для розв'язування задач із застосуванням нейромереж [17, 18], де об'єднання моделей іменовано комітетом (колективом, ансамблем) експертів. Доцільність застосування комітетів ґрунтується на тому, що точність колективного рішення переважає точність окремих моделей («експертів»). Автори цієї статті запропонували поширити ідею використання «комітету експертів» на альтернативний тип математичних моделей, а саме — *logit (probit)*-регресії.

Для оцінювання заявки позичальника за таким підходом поетапно використовується сукупність моделей, яку умовно можна вважати аналогом кредитних комітетів банківських установ, що приймає рішення про надання кредиту заявнику. Детально процедуру оцінки заявки на кредит пропонується реалізувати таким чином. На першому етапі оцінка позичальника здійснюється за моделлю, що має найвищий показник адекватності з усієї групи моделей, які отримані за виділеними для цього позичальника якіс-

ними критеріями. Далі проводиться аналіз за наступною по значущості моделлю, побудованою за відповідним даному позичальнику якісним критерієм. Якщо результати співпадають, можна вважати цей результат оцінкою позичальника, прийнятою комітетом експертів. Якщо результати за моделями протилежні, слід продовжити оцінювання, використовуючи наступну модель.

У нашому дослідженні для оцінювання рівня кредитоспроможності фізичної особи було відібрано три якісні показники: рівень освіти, статус працюючого та сфера діяльності. Тому рішення щодо позичальника за комітетом експертів буде відповідати однаково розрахунку хоча б двома з трьох моделей, побудованих для відповідних цьому позичальнику підгруп, виділених якісними показниками. Проте очевидно, що можна будувати моделі не лише за трьома групами якісних показників. Кількість груп поділу загальної сукупності даних, а отже і кількість моделей, буде різною для різних задач дослідження та для різних баз даних. Якщо потрібно буде враховувати результати розрахунків іншої кількості моделей, кінцевим результатом оцінки позичальника при такому підході можна вважати той, який буде надано більшістю моделей, що використовувались в розрахунках.

Крім того, для можливості врахування нерівнозначного відношення до результатів моделювання за виділеними групами (аналог рівня важливості рішення, прийнятого окремим експертом), можна встановлювати певні вагові коефіцієнти, які будуть визначати вагомість впливу кожної моделі на кінцеву оцінку кредитоспроможності позичальника. Вагові коефіцієнти можуть задаватись відповідно до показника точності класифікації кожної моделі.

Застосування запропонованого вище методологічного підходу до формування комітету експертів передбачає поділ початкової сукупності спостережень на групи. Отже, у дослідженні побудовано окремі моделі для кожної із груп спостережень відповідно до різних значень якісних показників. Тобто, за показником рівня освіти позичальника будувалось дві моделі: одна для фізичних осіб, які відносяться до підгрупи «вища освіта, дві та більше вищих освіти», та одна для підгрупи «середня та середня спеціальна освіта». За якісним показником статусу працівника було побудовано три моделі, які описують позичальників із власною справою, найманих робітників та з іншим статусом. Одним з головних чинників при оцінюванні кредитоспроможності позичальни-

ка спеціалісти виділяють наявність утриманців, тому окремі моделі побудовані для масивів, сформованих за критерієм кількості дітей у сім'ї — немає дітей, 1 дитина, 2 та більше дітей.

Для побудови кожної з моделей із загальної сукупності спостережень обиралась тільки та частина, яка відповідає певному якісному критерію. Тобто, спочатку зі всієї сукупності виділено лише масив даних, які відповідають позичальникам, що мають вищу або дві та більше вищих освіти. *Logit*- і *probit*-моделі оцінювання кредитних ризиків для цієї категорії позичальників, як і для всіх інших категорій, будувались на основі відібраних вище шести кількісних показників. На таких даних побудовано першу *logit*-модель *L1* та *probit*-модель *Pr1*. Інша частина масиву містить позичальників, що мають середню та середню спеціальну освіту. Ці дані використовувались для побудови моделей *L2* та *Pr2*. Аналогічно здійснювалось формування масивів для побудови моделей за показниками «Кількість утриманців» і «Статус працівника». Позначення моделей із характеристиками відповідних ним масивів статистичної інформації наведено в табл. 4.

Таблиця 4

**ПОЗНАЧЕННЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ,
ПОБУДОВАНИХ У ДОСЛІДЖЕННІ**

Критерій формування вибірки	Рівень освіти		Кількість утриманців			Статус працівника		
	Наявність вищої, двох та більше вищих освіти	Наявність середньої та спеціальної середньої освіти	Відсутність дітей	Одна дитина	Двоє та більше дітей	Власна справа	Найманий працівник	Інший статус
Позначення <i>logit</i> -моделі	<i>L1</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>	<i>L5</i>	<i>L6</i>	<i>L7</i>	<i>L8</i>
Позначення <i>probit</i> -моделі	<i>Pr1</i>	<i>Pr2</i>	<i>Pr3</i>	<i>Pr4</i>	<i>Pr5</i>	<i>Pr6</i>	<i>Pr7</i>	<i>Pr8</i>

У табл. 5 наведено оцінки параметрів $\beta = (\beta_0, \beta_1, \dots, \beta_6)$ для *logit*- і *probit*-моделей, обчислених для всіх сукупностей даних, сформованих за різними якісними критеріями. За знаком оцінок параметрів моделі можна визначити лише напрямок зміни залежної

величини від зміни відповідного вхідного показника. Безпосереднє значення, на яке змінюється залежна величина, обчислюється на основі (9). Відповідні розрахунки для всіх *logit*-моделей також наведено в табл. 5 у рядках, позначених Δx_j .

Всі моделі, крім *L6*, є статистично значущими з рівнем довіри 95 %. Модель *L6* стає статистично значущою лише при рівні довіри 26,44 %, що свідчить про її неадекватність. Тому результати розрахунків даної моделі у загальних висновках не використовувались.

Відношення неузгодженості для моделей *L1* і *Pr1* відповідно складають 1,8810 і 1,8438; відсоток коректно розпізнаних класів позичальників 60,73 % і 60,60 %.

Як зазначалось вище, розрахунки для *logit*- і *probit*-моделей майже співпадають як за характеристиками якості моделі, так і за значеннями отриманих параметрів регресій, що можна бачити з результатів розрахунків у табл. 5.

Значення оцінок параметрів кожної моделі вказують на чинники, що для даної групи позичальників визначають зменшення або зростання ризику неповернення кредиту. Так, наприклад, у моделях для першої підгрупи фізичних осіб, що мають вищу або дві та більше вищих освіти, збільшення стажу роботи працюючого на останньому місці роботи x_2 зменшує ризик неповернення кредиту, тоді як зростання величини загального стажу роботи x_3 , навпаки, збільшує цей ризик. У моделях підтверджується також цілком логічний висновок про те, що за наявності депозитів x_4 ризик неповернення кредиту значно зменшується, порівняно з іншими чинниками. Також зменшує даний ризик наявність у минулому повернених кредитів x_5 .

Значення, на які зростає чи зменшується залежна змінна при зростанні відповідного вхідного фактора на одиницю наведені в рядках з позначкою Δx_j . Так, найбільший вплив на зменшення ризику неповернення кредиту дає наявність у позичальника депозитів x_4 . При зростанні кількості депозитів на одиницю ризик неповернення кредиту для різних груп позичальників може зменшуватись до 13,9 %.

Таблиця 5

СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ, ПОБУДОВАНИХ У ДОСЛІДЖЕННІ

Модель	Значення оцінок параметрів							Значення функції втрат	Значення χ^2
	β_0	β_1	β_2	β_3	β_4	β_5	β_6		
<i>L1</i>	0,819	-0,026	-0,041	0,013	-0,343	-0,257	-0,234	481,9977	28,748
Δx_j	0,297	-0,009	-0,015	0,005	-0,124	-0,093	-0,085		
<i>Pr1</i>	0,495	-0,016	-0,025	0,008	-0,209	-0,159	-0,144	481,9726	28,798
<i>L2</i>	-1,52	0,026	0,035	0,006	-0,117	0,519	0,0003	950,11304	92,549
Δx_j	-0,41	0,007	0,010	0,002	-0,031	0,140	0,00007		
<i>Pr2</i>	-0,94	0,016	0,021	0,003	-0,074	0,320	0,0001	950,49762	91,780
<i>L3</i>	1,295	-0,027	-0,027	0,003	-0,421	-0,344	-0,100	654,80812	48,948
Δx_j	0,328	-0,007	-0,007	0,001	-0,106	-0,087	-0,025		
<i>Pr3</i>	0,805	-0,017	-0,016	0,002	-0,260	-0,211	-0,100	654,8697	48,825
<i>L4</i>	0,524	-0,032	-0,038	0,020	-0,075	-0,629	0,524	528,14569	30,766
Δx_j	0,161	-0,010	-0,012	0,006	-0,023	-0,193	0,161		
<i>Pr4</i>	0,309	-0,019	-0,022	0,011	-0,050	-0,389	0,309	528,14569	30,423
<i>L5</i>	0,109	0,030	0,080	0,028	0,583	0,325	-0,940	254,5499	52,296

Закінчення табл. 5

Модель	Значення оцінок параметрів							Значення функції втрат	Значення χ^2
	β_0	β_1	β_2	β_3	β_4	β_5	β_6		
Δx_j	0,026	0,007	0,019	0,007	0,139	0,077	-0,224		
<i>Pr5</i>	0,075	0,017	0,043	0,017	0,330	0,211	-0,549	254,54596	50,750
<i>L6</i>	0,445	0,016	-0,011	-0,03	-0,182	-0,569	-0,339	96,50678	7,6611
Δx_j	0,123	0,004	-0,003	-0,01	-0,050	-0,157	-0,094		
<i>Pr6</i>	0,271	0,010	-0,007	-0,02	-0,113	-0,352	-0,209	96,517363	7,6399
<i>L7</i>	-1,45	0,034	0,060	-0,01	0,375	0,414	0,072	1022,1565	97,025
Δx_j	-0,32	0,008	0,013	-0,003	0,083	0,092	0,016		
<i>Pr7</i>	-0,99	0,022	0,034	-0,01	0,228	0,256	0,047	1022,9439	95,450
<i>L8</i>	0,964	-0,030	0,034	0,004	0,237	-0,412	0,097	323,79315	21,826
Δx_j	0,251	-0,008	0,009	0,001	0,062	-0,107	0,025		
<i>Pr8</i>	0,602	-0,019	0,021	0,002	0,147	-0,255	0,060	323,7645	21,883

Наведені в табл. 5 параметри моделей описують схожі тенденції стосовно впливу різних факторів на ризик неповернення кредиту фізичними особами. Лише в деяких випадках для окремих груп позичальників можна виділити деякі специфічні відмінності. Наприклад, тоді як загалом для більшості випадків із збільшенням стажу роботи на останньому робочому місці ризик неповернення кредиту зменшується, то для групи позичальників, що сформована за критерієм рівня освіти (а саме середня та середня спеціальна освіта), спостерігається зворотний процес.

Отже, в процесі проведення дослідження отримано набір моделей, які утворюють описаний вище комітет експертів. Спочатку встановимо пріоритетність використання моделей на основі такої характеристики, як значущість. Аналіз показників адекватності *logit*- і *probit*-моделей дає можливість обґрунтовано залишити для моделювання кредитних ризиків фізичних осіб лише *logit*-моделі, оскільки суттєві відмінності із *probit*-моделями відсутні, а показник статистичної значущості χ^2 дещо вищий для *logit*-моделей майже у всіх випадках. Отже, використовуючи значення χ^2 , отримуємо наступний рейтинг *logit*-моделей (табл. 6).

Таблиця 6

**РЕЙТИНГ LOGIT-МОДЕЛЕЙ
НА ОСНОВІ ЇХ СТАТИСТИЧНОЇ ЗНАЧУЩОСТІ**

Модель	Значення χ^2	Рейтинг	Критерій формування сукупності спостережень при оцінюванні параметрів моделі
L7	97,025	1	Статус працівника — найманий працівник
L2	92,549	2	Рівень освіти — середня та середня спеціальна освіта
L5	52,296	3	Кількість дітей — двоє та більше
L3	48,948	4	Кількість дітей — діти відсутні
L4	30,766	5	Кількість дітей — одна дитина
L1	28,748	6	Рівень освіти — вища освіта, дві та більше вищих освіти
L8	21,826	7	Статус працівника — інший статус

Для оцінки кожної заявки на кредит потенційного позичальника застосовуватимуться кілька моделей із урахуванням наведеного рейтингу. Наприклад, якщо необхідно проаналізувати кредитну заявку позичальника з середньою освітою, який є найманним працівником і не має дітей (дані позичальника з ідентифікаційним номером (ID) 267, які не включались до жодної з навчальних вибірок при побудові моделей і використовуються лише для тестування результатів розрахунків), то така заявка може бути оцінена за моделями $L7$, $L2$ та $L3$. Тобто, умовно можна вважати, що комітет експертів складається з трьох осіб, рішення яких узгоджуються з цими моделями.

Спочатку оцінка даного позичальника здійснюється за моделлю, що має найвищий рейтинг серед усіх моделей, які отримані за виділеннями для цього позичальника якісними критеріями. Отже, слід починати з використання моделі $L7$. Розрахунок за цією моделлю здійснюється відповідно до (5) з урахуванням параметрів з табл. 5 за співвідношенням:

$$Y_{L7} = \frac{e^{-1,45+0,034x_1+0,06x_2-0,01x_3+0,375x_4+0,414x_5+0,072x_6}}{1 + e^{-1,45+0,034x_1+0,06x_2-0,01x_3+0,375x_4+0,414x_5+0,072x_6}}. \quad (10)$$

Отримане за моделлю значення для позичальника з ID 267 дорівнює $Y_{L7} = 0,378$.

Для інтерпретації отриманого результату необхідно його привести у відповідність до початкових значень вихідного показника (який приймав значення 0 для надійних позичальників та 1 для тих, що не виконали кредитних зобов'язань). З цією метою має бути встановлений певний поріг, при перевищенні якого значення залежної змінної прирівнюються до одиниці, якщо ж поріг не перевищено, то для залежної змінної приймається нульове значення. Як правило, в подібних задачах встановлюється поріг на рівні 0,5. Приймаючи таке порогове значення, доходимо висновку, що на основі моделі $L7$ даного позичальника слід вважати надійним (цей результат відповідає реальним даним, адже позичальником з шифром ID 267 кредит повернуто вчасно).

Аналогічно обчислюється кредитний ризик цього позичальника за наступною за рейтингом моделлю — $L2$. У результаті отримуємо значення $Y_{L2} = 0,318$, що також є нижчим за поріг поділу 0,5 та інтерпретується як надійність позичальника. В такому ви-

падку процедуру застосування комітету експертів можна зупинити, оскільки за двома з трьох моделей результат оцінки співпадає. Отже, можна вважати, що комітет експертів прийняв рішення на користь надання даному позичальнику кредиту.

Однак певну цікавість представляють собою ще й обчислені *logit*-моделлю значення до їх перетворення на 0 та 1. За логікою побудови моделі їх можна трактувати як імовірності неповернення певним позичальником кредиту, що характеризують ризиковість надання кредиту даній особі. Для отримання узагальненої оцінки комітету експертів щодо кредитного ризику позичальника та присвоєння йому кредитного рейтингу доцільно врахувати рішення за всіма моделями. Відповідно, результат розрахунку за моделлю $L3$ — значення $Y_{L3} = 0,638$.

Як зазначалось вище, з метою врахування нерівнозначного відношення до результатів моделювання за виділеними групами (аналог рівня важливості рішення, прийнятого окремим експертом), можна встановлювати певні вагові коефіцієнти, за якими будуть зважені результати розрахунку кожної моделі. Тоді загальний висновок на основі всіх трьох моделей шукатиметься не як середнє арифметичне їх розрахунків, а через коефіцієнти k_{L7} , k_{L2} та k_{L3} ($k_{L7} + k_{L2} + k_{L3} = 1$), які відобразатимуть рівень важливості кожного із значень Y_{L7} , Y_{L2} , Y_{L3} . Якщо встановити, наприклад, значення коефіцієнтів $k_{L7} = 0,5$, $k_{L2} = 0,25$ та $k_{L3} = 0,25$, то загальний результат згортки буде $Y = k_{L7}Y_{L7} + k_{L2}Y_{L2} + k_{L3}Y_{L3} = 0,428$, що вказує на досить невисоку міру ризикованості надання даному клієнту кредиту. Відповідно, за узагальненим розрахунком комітету експертів можна вважати такого позичальника надійним.

Для проведення більш детального тестування адекватності запропонованого методологічного підходу до формування комітету експертів використовувався набір тестових даних (що не використовувались для побудови моделей), який для спрощення розрахунків було впорядковано лише за критеріями моделей $L7$, $L2$, $L3$. У результаті утворено вибірку з 19 записів, перші десять з яких відповідають надійним клієнтам і дев'ять — ненадійним. У табл. 7 наведено результати розрахунку кредитоспроможності цих позичальників за трьома вказаними моделями, а також результат згортки отриманих розрахунків, що являє собою оцінку комітету експертів.

Таблиця 7

РЕЗУЛЬТАТИ МОДЕЛЮВАННЯ ЗА ОКРЕМИМИ МОДЕЛЯМИ ТА УЗАГАЛЬНЕННЯ КОМІТЕТОМ ЕКСПЕРТІВ

ID позичальника	Розрахунок за моделями			Розрахунок комітету експертів, \hat{Y}	Реальний стан позичальника
	Y_{L1}	Y_{L2}	Y_{L3}		
1122	0,531083	0,523884	0,478138	0,516047	надійний
1384	0,508094	0,513218	0,482396	0,502951	надійний
189	0,444824	0,393536	0,572667	0,463963	надійний
267	0,378051	0,317754	0,638451	0,428077	надійний
367	0,34411	0,28887	0,660561	0,409413	надійний
476	0,486715	0,455743	0,516863	0,486509	надійний
666	0,561778	0,515225	0,463929	0,525677	надійний
928	0,532061	0,497883	0,501888	0,515974	надійний
1054	0,557632	0,527396	0,460325	0,525746	надійний
1269	0,517989	0,488276	0,491985	0,50406	надійний
332061	0,569637	0,593182	0,42493	0,539347	ненадійний
345586	0,729851	0,62226	0,401131	0,620773	ненадійний
350555	0,587991	0,483444	0,507455	0,54172	ненадійний
331884	0,614601	0,600883	0,413561	0,560911	ненадійний
339671	0,787389	0,714581	0,345766	0,658781	ненадійний
346073	0,543752	0,50811	0,499423	0,523759	ненадійний
350728	0,600772	0,579699	0,444517	0,55644	ненадійний
331010	0,700957	0,636432	0,389468	0,606954	ненадійний
335256	0,70724	0,638001	0,370645	0,605782	ненадійний

Узагальнений розрахунок комітету з трьох моделей у цілому збігається з реальним виконанням позичальником кредитних зобов'язань (для надійних позичальників оцінка ризиковості знаходиться в межах від 0,4 до 0,52, для ненадійних — від 0,54 до 0,66). Зауважимо, що подібний тип моделей та результати розрахунків надають можливість встановлювати власні порогові значення, виділяти різні рівні надійності або ризиковості позичальників. Використовуючи дані табл. 7 і приймаючи термінологію зон ризику, яка наведена в [19], проілюструємо на рис. 3 можливий варіант аналізу кредитних заявок для тестових даних на прикладі розрахунків за моделлю *L7*.

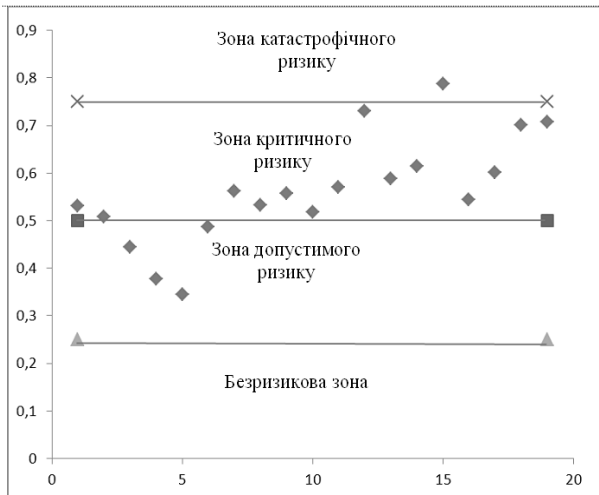


Рис. 3. Визначення зон ризику для оцінювання кредитних заявок за моделлю *L7*

Якщо здійснити поділ всіх позичальників лише на надійних і ненадійних, то результати моделювання досить точно співпадають з реальними даними. Для перших десяти клієнтів найбільшим значенням ризику за моделлю *L7* є 0,56, інші дев'ять, за винятком одного, перевищують 0,57. Однак у випадку, коли необхідно більш обережно приймати рішення, доцільно ввести зони ризику. На рис. 3 інтервал $[0, 1]$ поділений найпростішим чином — на рівні чотири частини, визначаючи однакові діапазони зон ризику.

Згідно рекомендацій, наданих у [19], для заявок, оцінка яких потрапляє в зону допустимого ризику, можна приймати рішення про надання кредиту без особливих пересторог; на групу клієнтів, що потрапили в зону критичного ризику, слід більше звернути увагу в процесі оцінювання; а клієнтів, оцінка яких знаходиться або в зоні катастрофічного ризику, або наближається до його межі, доцільно вважати потенційно ненадійними та відмовити у наданні кредиту.

Очевидно, що наведений поділ на зони ризику є досить суб'єктивний. У будь-якому разі, встановлення порогових значень зон ризику буде залежати від суб'єкта прийняття рішення і такий підхід надасть можливість в мінливих зовнішніх умовах якомога точніше налаштувати процес відбору кредитних заявок.

З метою підтвердження доцільності застосування методологічного підходу до формування комітету експертів у дослідженні було також проведено порівняння результатів моделювання для двох випадків: використання комітету моделей та однієї моделі, побудованої для всієї сукупності даних (без виокремлення підгруп за якісними критеріями). У табл. 8 наведено параметри та статистичні характеристики *logit*-моделі *L9*, налаштованої на повному масиві початкових даних.

Таблиця 8

**СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ МОДЕЛІ, ПОБУДОВАНОЇ
НА ПОВНОМУ МАСИВІ ПОЧАТКОВИХ ДАНИХ**

Модель	Значення оцінок параметрів							Значення функції втрат	Значення χ^2
	β_0	β_1	β_2	β_3	β_4	β_5	β_6		
<i>L9</i>	1,162	-0,020	-0,029	-0,004	-0,176	-0,539	-0,086	1831,494	128,93
Δx_j	0,304	-0,005	-0,008	-0,001	-0,046	-0,141	-0,023		

Як можна бачити з табл. 8, всі включені в узагальнену модель *L9* фактори зменшують ризик неповернення кредитів. Найбільш суттєво (на 14 %) зменшиться кредитний ризик при збільшенні на одиницю раніше виплачених кредитів. Порівняння результатів розрахунків кредитних ризиків позичальників із тестової множини, отриманих за моделлю *L9* та вище описаним комітетом експертів, наведено в табл. 9.

Таблиця 9

ПОРІВНЯННЯ РЕЗУЛЬТАТІВ РОЗРАХУНКІВ ЗА МОДЕЛЛЮ, ПОБУДОВАНОЮ НА ПОВНОМУ МАСИВІ ДАНИХ, ТА КОМІТЕТУ ЕКСПЕРТІВ

ID	$L9$	Розрахунок комітету експертів, Y	Реальний стан позичальника
1122	0,482322	0,516047	надійний
1384	0,491791	0,502951	надійний
189	0,580512	0,463963	надійний
267	0,639978	0,428077	надійний
367	0,664243	0,409413	надійний
476	0,534112	0,486509	надійний
666	0,487422	0,525677	надійний
928	0,500066	0,515974	надійний
1054	0,478876	0,525746	надійний
1269	0,50904	0,50406	надійний
332061	0,569199	0,539347	ненадійний
345586	0,603023	0,620773	ненадійний
350555	0,493881	0,54172	ненадійний
331884	0,577776	0,560911	ненадійний
339671	0,707294	0,658781	ненадійний
346073	0,543951	0,523759	ненадійний
350728	0,597568	0,55644	ненадійний
331010	0,643866	0,606954	ненадійний
335256	0,610892	0,605782	ненадійний

Як можна пересвідчитись з табл. 9, за моделлю $L9$ маємо дещо більші помилки класифікації, ніж отримані комітетом експертів. Так, з перших десяти значень, отриманих із застосуванням моделі $L9$, за умови встановлення лінії поділу на рівні 0,54 помилково можуть бути визнані ненадійними три позичальника, з ненадійних один може бути помилково визнаний надійним.

Аналізуючи проведені розрахунки, варто вказати на доцільність здійснення оцінки заявок на кредит на основі комплексу моделей, які представляють аналог групи експертів, оскільки використання однієї моделі, розрахованої за всією сукупністю даних, надто узагальнює результати та призводить до збільшення помилок класифікації.

Висновки

У роботі розглянуто деякі аспекти побудови економіко-математичних моделей оцінювання кредитоспроможності позичальників-фізичних осіб, що стосуються питань, недостатньо досліджених

на сьогодні в спеціалізованій літературі. Перше з них пов'язано з тим, що банківський аналіз кредитоспроможності позичальників ґрунтується на необхідності обробки значних обсягів інформації. В статті запропоновано методологічний підхід до відбору найбільш значущих чинників для оцінювання кредитоспроможності позичальників, що заснований на поєднанні роботи ймовірнісної нейромережі та генетичного алгоритму. Подібний підхід дозволяє на великих масивах даних ефективно здійснювати перебір можливих комбінацій вхідних факторів для побудови математичних моделей.

Також у дослідженні розроблено методику врахування якісних чинників оцінювання кредитоспроможності позичальників — фізичних осіб шляхом формування окремих масивів даних за виділеними якісними критеріями, що використовуються для побудови відповідних економіко-математичних моделей. Кожна з таких моделей дає оцінку позичальника з точки зору одного з важливих якісних чинників, що впливають на його кредитоспроможність. У статті також запропоновано методологічний підхід, який дозволяє узагальнити отримані оцінки кількох *logit*-регресійних моделей, які разом утворюють «комітет експертів».

Запропонований методологічний підхід до застосування комітету експертів придатний не тільки для оцінки кредитоспроможності фізичних осіб — він може бути узагальнений для будь-яких інших видів кредитування чи задач класифікації.

Також у статті проведено експериментальне дослідження ефективності двох запропонованих підходів до оцінювання кредитоспроможності позичальників. Остаточний вибір одного з них залежить від мети дослідження: використання нейромережі типу багатошарового перцептрону дозволяє ефективно здійснити розподіл усіх позичальників за двома класами — надійні та ненадійні клієнти, тоді як результат застосування «комітету експертів» *logit*-моделей може бути інтерпретований як імовірнісна оцінка надійності позичальника.

Тестування побудованих економіко-математичних моделей підтвердило високий рівень адекватності запропонованих методологічних підходів, покладених у їх основу. Їх застосування може дозволити істотно підвищити точність оцінювання кредитоспроможності потенційного позичальника та мінімізувати рівень кредитного ризику банківської установи. Результати дослідження можуть бути використані банківськими та іншими фінансовими установами, зацікавленими в адекватній процедурі оцінки кредитоспроможності фізичних осіб.

Література

1. *Соложенцев Е. Д.* Прозрачность методик оценки кредитных рисков и рейтингов / Соложенцев Е. Д., Степанова Н. В., Карасев В. В. — СПб.: С.-Петербургский ун-т, 2005. — 197 с.
2. *Магнус Я. Р.* Эконометрика / Магнус Я. Р., Катывшев П. К., Пересецкий А. А. — М.: Дело, 2004. — 576 с.
3. Кредитний ризик комерційного банку: Навч. посіб. / [Вітлінський В. В., Пернарівський О. В., Наконечний Я. С., Великоіваненко Г. І.; За ред. В. В. Вітлінського]. — К.: Знання, 2000. — 251 с.
4. *Камінський А. Б.* Скорингові технології в кредитному ризик-менеджменті / А. Б. Камінський, К. К. Писанець // Бізнес Інформ. — 2012. — № 4. — С. 197—201.
5. *Камінський А. Б.* Моделювання фінансових ринків: монографія / А. Б. Камінський. — К.: Видавничо-поліграфічний центр «Київський університет», 2006. — 304 с.
6. *Камінський А. Б.* Модель кредитного скорингу заснована на концепції виживання / А. Б. Камінський, К. К. Писанець // Формування ринкової економіки в Україні. — 2012. — Вип. 27. — С. 136—42.
7. *Кишакевич Б. Ю.* Моделювання та оптимізація кредитних ризиків банку: монографія / Б. Ю. Кишакевич. — Дрогобич: Коло, 2011. — 412 с.
8. *Кишакевич Б. Ю.* Стрес-тестування кредитного портфеля банку на основі багатофакторних моделей / Б. Ю. Кишакевич // Економічний простір: Збірник наукових праць. — Дніпропетровськ: ПДАБА, 2011. — № 45. — С. 161—171.
9. *Кишакевич Б. Ю.* Модель оцінки кредитоспроможності та ймовірності дефолту позичальника банку / Б. Ю. Кишакевич // Вісник Львівської державної фінансової академії. — 2010. — № 19. — С. 230—239.
10. *Пернарівський О. В.* Аналіз, оцінка та способи зниження банківських ризиків / О. В. Пернарівський // Вісник НБУ. — 2004. — № 4. — С. 44—48.
11. *Писанець К. К.* Кредитный скоринг и принятие решений в финансовых моделях риск-менеджмента банков / К. К. Писанець // Уральский вестник: Астана. — 2013. — № 5. — С. 105—112.
12. *Писанець К. К.* Проблема вибору моделі кредитного скорингу для оцінки кредитного ризику позичальника у споживчому сегменті [Електронний ресурс] / К. К. Писанець // Ефективна економіка. — 2013. — № 10. — Режим доступу до журн.: <http://www.economy.nayka.com.ua/?op=1&z=2416>.
13. *Великоіваненко Г. І.* Нейро-нечітка модель оцінювання прострочених позик комерційного банку / Г. І. Великоіваненко, Л. О. Трокоз // Нейро-нечіткі технології моделювання в економіці. — 2014. — № 3. — С. 23—66.

14. Ковтун Р. С. Значимость методико-практического аспекта оценки кредитоспособности физических лиц в рамках потребительского кредитования. [Электронный ресурс] / Р. С. Ковтун // Успехи современного естествознания. — 2007. — № 12. — Ч. 3. — С. 74—75. — Режим доступа: http://www.rae.ru/use/?section=content&op=show_article&article_id=7778703.

15. Кветный Р. Н. Вероятностные нейронные сети в задачах идентификации временных рядов / Р. Н. Кветный, В. В. Кабачий, О. О. Чумаченко // Наукові праці ВНТУ. — 2010. — № 3. — С. 2—6.

16. Нейронные сети. STATISTICA Neural Networks: Методология и технологии современного анализа данных / [Под ред. В. П. Боровикова]. — 2-е изд., перераб. и доп. — М.: Горячая линия — Телеком, 2008. — 392 с.

17. Царегородцев В. Г. Оптимизация экспертов boosting-коллектива по их кривым обучения. Материалы XIII Всеросс. семинара «Нейроинформатика и ее приложения» / В. Г. Царегородцев. — Красноярск, 2004. — С. 152—157.

18. Научная сессия МИФИ-2007. IX Всероссийская научно-техническая конференция «Нейроинформатика-2007»: Лекции по нейроинформатике. — М.: МИФИ, 2007. — Ч. 2. — 148 с.

19. Вітлінський В. В. Аналіз, моделювання та управління економічним ризиком: [Навч.-метод. посібник для самост. вивч. дисц.] / В. В. Вітлінський, П. І. Верченко. — К.: КНЕУ, 2000. — 292 с.

References

1. Solozhentsev, E. D., Stepanova, N. V., & Karasev, V. V. (2005). *Prozrachnost' metodyk otsenky kredytnykh ryiskov y rejtynhov*. SPb: S-Peterburgskij universitet [in Russian].

2. Magnus, Ja. R., Katyshev, P. K., & Pereseckij, A. A. (2004). *Ekonometrika*. Moskva: Delo [in Russian].

3. Vitlins'kyj, V. V., Pernarivs'kyj, O. V., Nakonechnyj, Ya. S., & Velykoivanenko, H. I. (2000). *Kredytnyj ryzik komertsijnoho banku*. Kyiv: Znannia [in Ukrainian].

4. Kamins'kyj, A. B., & Pysanets', K. K. (2012). Skorynhovi tekhnolohii v kredytnomu ryzik-menedzhmenti. *Biznes Inform (Business Inform)*, 4, 197—201 [in Ukrainian].

5. Kamins'kyj, A. B. (2006). *Modeliuvannia finansovykh rynkiv*. Kyiv: Vydavnycho-polihrafichnyj tsentr «Kyivs'kyj universytet» [in Ukrainian].

6. Kamins'kyj, A. B., & Pysanets', K. K. (2006). Model' kredytnoho skorynhu zasnovana na kontseptsii vyzhvannia. *Formuvannia rynkovoi ekonomiky v Ukraini (The formation of market economy in Ukraine)*, 27, 136—142 [in Ukrainian].

7. Kyshakevych, B. Yu. (2011). *Modeliuvannia ta optymizatsiia kredytnykh ryzykiv banku*. Drohobych: Kolo [in Ukrainian].

8. Kyshakevych, B. Yu. (2011). Stres-testuvannia kredytnoho portfelia banku na osnovi bahatofaktornykh modelej. *Ekonomichnyj prostir (Economic space)*, 45, 161—171 [in Ukrainian].

9. Kyshakevych, B. Yu. (2010). Model' otsinky kredytopromozhnosti ta jmovirnosti defoltu pozychal'nyka banku. *Visnyk L'vivs'koi derzhavnoi finansovoi akademii (Bulletin LSAF)*, 19, 230—239 [in Ukrainian].

10. Pernarivs'kyj, O. V. (2004). Analiz, otsinka ta sposoby znyzhennia bankivs'kykh ryzykiv. *Visnyk NBU (Bulletin of the National Bank of Ukraine)*, 4, 44—48 [in Ukraine].

11. Pisanets, K. K. (2013). Kreditnyj skoring i prinjatje reshenij v finansovyh modeljah risk-menedzhmenta bankov. *Ural'skij vestnik (Ural Bulletin)*, 5, 105—112 [in Russian].

12. Pysanets', K. K. (2013). Problema vyboru modeli kredytnoho skorynhu dlja otsinky kredytnoho ryzyku pozychal'nyka u spozhyvchomu sehmenti. *ECONOMY.NAYKA.COM.UA. Efektyvna ekonomika (Efficient economy)*, 10. Retrieved November 11, 2014, from <http://www.economy.nayka.com.ua/?op=1&z=2416> [in Ukrainian].

13. Velykoivanenko, H. I., & Trokoz, L. O. (2014). Nejro-nechitka model' otsiniuvannia prostrochennykh pozyk komertsijnogo banku. *Nejro-nechitki tekhnologii modeljuvannia v ekonomitsi (Neuro-Fuzzy Modeling Techniques in Economics)*, 3, 23—66 [in Ukrainian].

14. Kovtun, R. S. (2007). Znachimost' metodiko-prakticheskogo aspekta ocenki kredytoposobnosti fizicheskikh lic v ramkah potrebitel'skogo kreditovanija. *RAE.RU. Uspehi sovremennogo estestvoznanija (The success of modern science)*, 12(3), 74—75. Retrieved Desember 24, 2014, from http://www.rae.ru/use/?section=content&op=show_article&article_id=7778703 [in Russian].

15. Kvetnyj, R. N., Kabachij, V. V., & Chumachenko, O. O. (2010). Verojatnostnye nejronnye seti v zadachah identifikacii vremennykh rjadov. *Naukovi praci VNTU (Scientific works of VNTU)*, 3, 2—6 [in Ukrainian].

16. Borovikov, V. P. (2008). *Nejronnye seti. STATISTICA Neural Networks: Metodologija i tehnologii sovremennogo analiza dannyh*. Moskva: Gorjachaja linija — Telekom [in Russian].

17. Caregorodcev, V. G. (2004). Optimizacija jekspertov boosting-kollektiva po ih krivym obuchenija. *Materialy XIII Vseross. seminaru «Nejroinformatika i ee prilozhenija» (Proceedings of XIII All-Russian Seminar «Neuroinformatics and its applications»)*, 152—157 [in Russian].

18. *Nauchnaja sessija MIFI-2007. IX Vserossijskaja nauchno-tehnicheskaja konferencija «Nejroinformatika-2007»: Lekcii po nejroinformatike*. (2007). Moskva: MIFI [in Russian].

19. Vitlins'kyj, V. V., & Verchenko, P. I. (2000). *Analiz, modeljuvannia ta upravlinnia ekonomichnym ryzykom*. Kyiv: KNEU [in Ukrainian].

Стаття надійшла до редакції 30.04.2015