

УДК 330.46

ЦЕСЛІВ О.В.

Київський національний технічний університет України "Київський політехнічний інститут"

МЕТОДИ ШТУЧНОГО ІНТЕЛЕКТУ ЯКІ ВИКОРИСТОВУЮТЬСЯ ДЛЯ КРЕДИТНОГО СКОРИНГУ

Мета. Дослідити можливість розв'язування задачі кредитного скорингу методами штучного інтелекту.

Методика. Використані загальновідомі методи штучного інтелекту: класифікації, побудова нейронної мережі прямого поширення та ймовірнісної нейронної мережі в системі Mathlab. В системі HUGIN розроблено експертну систему на основі байєсовської мережі довіри (БМД).

Результати. У ході реалізації методів доведена можливість розв'язування задачі скорингу засобами Mathlab та HUGIN. Проілюстровані переваги та недоліки наведених методів. Проаналізовано, що для даної задачі найкращим методом розв'язку є двошарова нейронна мережа прямого розповсюдження сигналу та ймовірнісна нейронна мережа. Доведено, що вибір методу залежить від стратегії банку та його пріоритетів.

Наукова новизна. Досліджено точність та ефективність методів кредитного скорингу в середовищі Mathlab та системі HUGIN.

Практична значимість. Наведеними методами в середовищах Mathlab та HUGIN можливо проводити кредитний скоринг в банках.

Ключові слова: штучний інтелект, кредитний скоринг, нейронні мережі, експертні системи, байєсовські мережі, логіко - ймовірнісні методи

Вступ. Призначення кредитного скорингу - автоматизоване прийняття рішень видачі кредитів приватним особам. Доступна інформація про потенційних позичальників міститься в заповнених ними анкетах. Такі фактори, як річний дохід, розмір непогашеного боргу, володіння нерухомістю або автомобілем, стаж роботи на останньому місці, вік - потенційно пов'язані з кредитоспроможністю і тому можуть виявитися вхідними змінними в скорингової моделі. Однак внаслідок кореляцій між цими факторами, у підсумковій моделі використовуються не всі з них. В ході побудови моделі можуть брати участь 50-60 змінних, а в кінцевій моделі залишається не більше 8-12 з них.

Методика оцінювання кредитоспроможності позичальників - фізичних осіб з використанням кредитного скорингу була розроблена американським економістом

Д.Дюраном на початку 40-х років ХХ ст. Ця методика є найбільш поширеною і дає можливість здійснити експрес - аналіз у присутності клієнта за його заявкою на кредит. Розробкою формалізованого підходу до оцінки кредитоспроможності позичальників було покладено такими зарубіжними вченими, як Е. Альтман, Дж. Синки, Е. Боді, А. Кейн, А. Маркус, Е. Найман. Серед вітчизняних вчених дану проблему розробляли такі вчені, як О.І. Лаврушин, Д.А. Ендовицький, Г.Н. Белоглазова, Є.В. Тихомирова, Н.А. Федорова.

Об'єктом дослідження є методи кредитного скорингу. В даний час для кредитного скорингу використовуються методи статистики (дискримінантний аналіз,

лінійна регресія, логістична регресія, дерева класифікації), дослідження операцій (лінійне програмування, нелінійна оптимізація) і штучного інтелекту (нейронні мережі, експертні системи, генетичні алгоритми, байєсовські мережі, логіко-імовірнісні методи). Зазначені методи можуть застосовуватися як окремо, так і в різних комбінаціях. Велика кількість абсолютно різних методів пояснюється суто прагматичним підходом: використовувати те, що працює найкраще для певної задачі.

Постановка завдання. Для побудови моделі береться вибірка даних по існуючих позичальникам, зазвичай не менше кількох тисяч записів [1].

Проблему кредитного скорингу можна розглядати як завдання класифікації: знаючи відповіді на питання анкети $x \in A$, можна визначити, до якої групи належить позичальник: $x \in A_G$ для «добрих» клієнтів, $x \in A_B$ для «поганих». При цьому необхідно розуміти, що абсолютно точна класифікація принципово неможлива.

Принципи визначення кредитоспроможності приватного позичальника можна проілюструвати на прикладі моделі німецького банку, в якому підрахунок балів для рейтингу клієнта проводиться по 12 показниках:

1. Інформація. За відсутність несприятливої інформації кредитно довідкового бюро клієнт отримує 10 балів (5).
2. Здатність погашати заборгованість: до 60% - 0 балів; від 61 до 80% - 10 балів; від 81 до 100% - 20 балів.
3. Наявність забезпечення: від 0 до 25% - 1 бал; від 25 до 50% - 4; від 51 до 75% - 7; від 76 до 100% - 12; більше 100% - 20 балів.
4. Наявне майно. За наявне майно, будь то нерухомість, цінні папери або вклади в банках, клієнт отримує 10 балів.
5. Кредити, отримані в банку раніше. Клієнту не нараховуються бали, якщо він неакуратно користувався наданими позиками. Якщо клієнт не користувався раніше кредитом, це розцінюється в 5 балів. Якщо раніше отриманий клієнтом кредит погашався своєчасно або поточний погашається відповідно до договору, то він отримує 15 балів.
6. Кваліфікація. Ні кваліфікації - 0 балів; допоміжний персонал - 2; фахівці - 7; службовці - 9 балів; пенсіонери - 13; керівні працівники - 13 балів.
7. Трудова діяльність у останнього наймача: до одного року - 0 балів; до двох років - 3; до трьох років - 5; до п'яти років - 8; більше п'яти років - 12; пенсіонери - 0 балів.
8. Сфера зайнятості: держслужба - 10 балів; інші сфери - 6; пенсіонери - 0 балів.
9. Вік заявника: до двадцяти років - 0 балів; двадцяти п'яти - 2; тридцяти - 4; тридцяти п'яти - 8; п'ятдесяти - 9; шістдесяти - 11; більше шістдесяти років - 16 балів.
10. Сімейний стан: неодружений - 8 балів; одружений - 14; одружений, але живе окремо - 6; розлучений - 8; вдовець - 8 балів.
11. Спосіб найму житла. Що не має житла - 0 балів; має житло за наймом - 5; власну оселю - 10 балів.
12. Кількість утриманців: нуль - 10 балів; один - 7; два - 5; три - 2; більше трьох - 0 балів.

Таблиця 1. Вихідні дані

Інформація	Здатність погашання заборгованості	Наявність забезпечення	Майно	Кредити, отримані в банку раніше	Кваліфікація	Трудова діяльність в останнього працедавця	Сфера занятості: держслужба	Сімейний стан	Спосіб винайму житла	Кількість утриманців	Вік	Сума - S	Рейтинг
10	0	1	10	0	7	3	10	8	0	7	4	60	3
0	20	4	10	15	0	5	0	14	5	0	8	81	1
10	10	7	0	0	9	8	0	6	10	5	9	74	2
10	20	12	0	5	13	12	10	8	10	2	9	111	1
10	1	0	5	0	0	0	0	16	8	10	10	60	3
10	20	4	10	15	13	12	10	16	10	10	9	139	1
10	10	7	0	0	2	8	0	6	10	5	9	67	3
10	20	12	0	5	13	12	10	8	10	2	9	111	1
10	1	0	5	0	0	0	0	16	8	10	10	60	3
10	20	7	10	15	13	12	10	16	10	10	10	143	1
10	0	1	0	5	7	3	10	8	0	7	4	55	3
0	20	7	10	15	0	5	0	14	5	0	8	84	1
10	10	7	10	5	7	8	0	6	10	5	9	87	1
10	20	12	0	0	13	12	10	8	10	2	9	106	1
10	1	0	5	0	0	0	0	16	8	10	10	60	3
10	20	12	10	15	7	12	10	16	10	10	10	142	1
10	10	7	0	5	9	8	0	6	10	5	9	79	2
10	20	4	0	5	13	12	10	8	10	2	9	103	1
10	1	0	5	0	0	0	0	16	8	10	9	59	3
10	20	12	10	15	13	12	10	16	10	10	10	148	1

Технологія остаточного рішення про можливість кредитування така: якщо претендент набирає в сумі 81 бал економіст приймає позитивне рішення самостійно, при результаті від 61 до 80 балів потрібен дозвіл головного менеджера. При рейтингу нижче 60 балів у наданні кредиту клієнту відмовляють.

Результати дослідження. Припустимо клієнти подали анкети з даними, які наведено в таблиці 1. Позначення в стовпці рейтинг наступні: 1- видати кредит; 2 - консультація з менеджером; 3 - не видавати.

Розв'язок цієї задачі проведемо, використовуючи методи штучного інтелекту: найпростіший шлях звичайна класифікація; другий створення нейронної мережі, третій створення ймовірнісної нейронної мережі. Класифікацію проведемо в пакеті Matlab.

```
data =[ 60 81 74 111 60 139 67 111 60 143 55 84 87 106 60 142 79 103
59 148]
```

```
training=data(1:10) %навчальні дані
```

```
sample=data(11:20) %реальні дані
```

```
group = [ 3 1 2 1 3 1 3 1 3 1 ]  
% класифікація  
[class, err, P, logp, coeff] = classify(sample, training, group);  
disp(class)% розподіл по класам  
3 2 2 1 3 1 2 1 3 1
```

Порівнюємо ці дані з реальними 3 1 1 1 3 1 2 1 3 1. Різниця в значеннях, яким відповідають суми 81 та 74, які близькі до граничних.

Коефіцієнти дискримінантної функції [2]

$A = \text{coeff}(1,2).linear,$

$A = 0.1174.$

Значення дискримінантної функції для даних training: $F_{train} = \text{training} * A,$

$F_{train} = 7.0416x_1 + 9.5062x_2 + 8.6847x_3 + 13.0270x_4 + 7.0416x_5$
 $+ 16.3131x_6 + 7.8631x_7 + 13.0270x_8 + 7.0416x_9 + 16.7825x_{10}.$

Значення дискримінантної функції інших даних sample:

$F = \text{sample} * A$

$F = 6.4548x_1 + 9.8583x_2 + 10.2104x_3 + 12.4402x_4 + 7.0416x_5$
 $+ 16.6652x_6 + 9.2715x_7 + 12.0881x_8 + 6.9243x_9 + 17.3693x_{10}$

Константа дискримінації:

$F_a = - \text{coeff}(1,2).const,$

$F_a = 11.2079.$

Розв'язати задачу можна використовуючи нейронну мережу [3]. Використовуємо ті самі дані: training; sample; group. Побудуємо мережу, використовуючи графічний інтерфейс користувача NNTool, тобто створюємо мережу прямого поширення, тобто без зворотних зв'язків. Основною властивістю таких нейронних мереж є їх стійкість. Графічний інтерфейс користувача NNTool дозволяє вибирати структури нейронних мереж з великого переліку і задавати різні алгоритми навчання для кожного типу мережі.

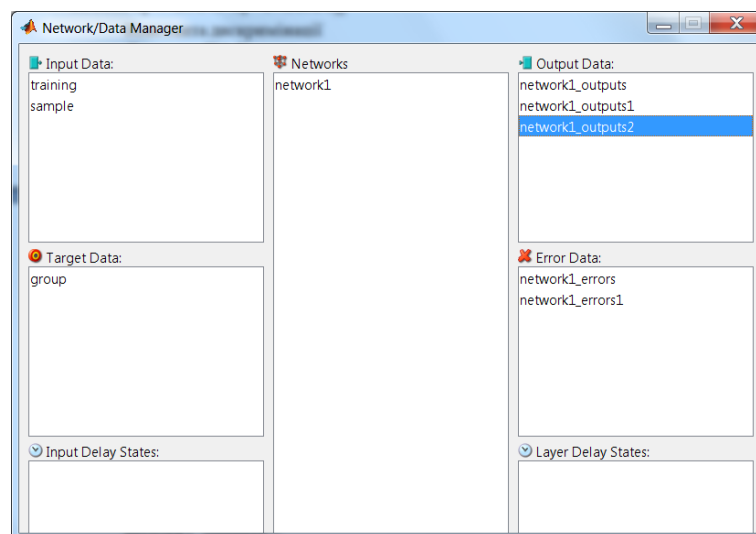


Рис. 1. Головне вікно NNTool

Значення результатів для навчальної вибірки повністю співпадають значенням таблиці 1. Якщо на вхід подається значення `sample`, отримані результати:

`network1_outputs2 =[3 1 1 3 1 1 3 1];`

Порівнюємо дані з таблицею `[3 1 1 3 1 2 1 3 1]`.

Помилку отримано в одному значенні, сума якого 79 балів. Результати показують, що маємо з десяти значень одну помилку, так звані середні клієнти, рішення по яким має приймати головний менеджер.

Досліджувалася двошарова нейронна мережа прямого розповсюдження сигналу. Модель має 20 входів, об'єднаних у вектор. Кожен елемент вектору відповідає клієнту. Розмірність закритого шару 20 нейронів. Функція активації нейронів гіперболічний тангенс. Критерій навчання середній квадрат помилки(MSE) Мережа навчається в пакетному режимі методом Левенберга - Марквардта.

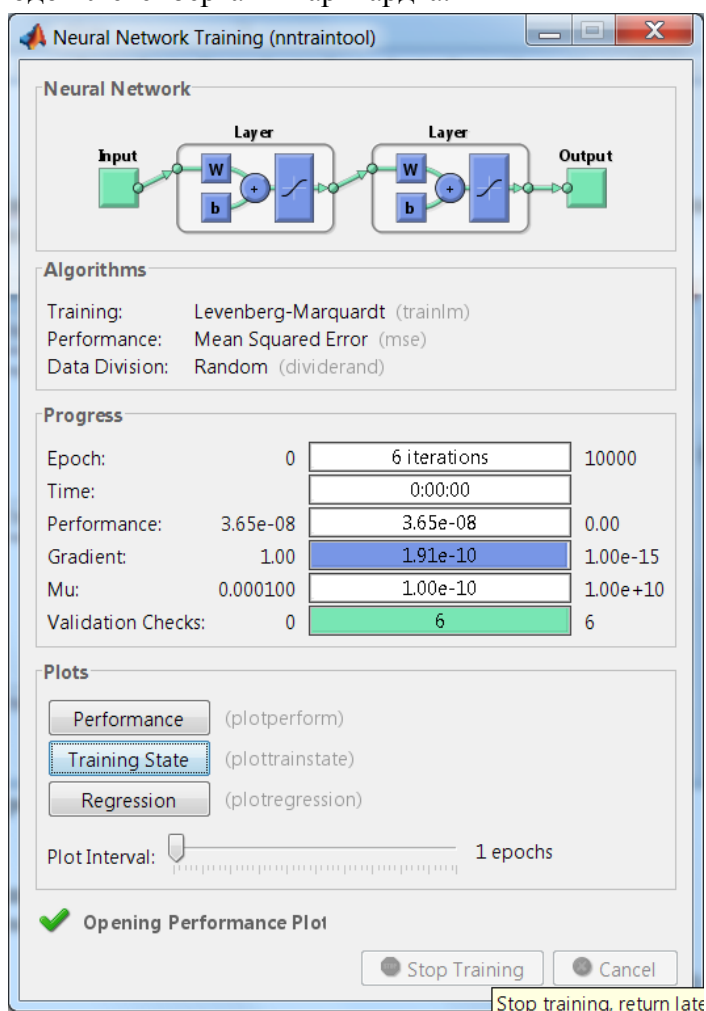


Рис. 2. Вікно параметрів "навчання" мережі

Третій варіант розв'язання даної задачі, це використання імовірнісної нейронної мережі [4]:

`P=[60 81 74 111 60 139 67 111 60 143]'; % вектори`

`Tc=[3 1 2 1 3 1 3 1 3 1]; % вектор цілій`

```

T= ind2vec(Tc);
T= full (T);           % - перетворення до повної матриці;
net= newpnn(P,T); % - ймовірнісна мережа
net.layers {1}.size   % - число нейронів 1—го шару;
net.layers {2}.size   % - число нейронів 2—го шару;
Y= sim (net, P);      % - моделювання мережі;
Yc= vec2ind(Y);       % - формування індексів класів;
Pt= [143 55 84 87 106 142 79 103 59 148]'; % - вектор для тестування;
A= sim (net, Pt);     % - тестування мережі;
Ac= vec2ind (A);
Ac = 1 3 1 1 1 3 1 2 1 3 1.
Реальні значення 1 3 1 1 1 1 1 2 1 3 1.
    
```

Теж одна помилка, тобто 10% від загальної кількості реальних даних, а саме при значенні суми 60, граничному значенні.

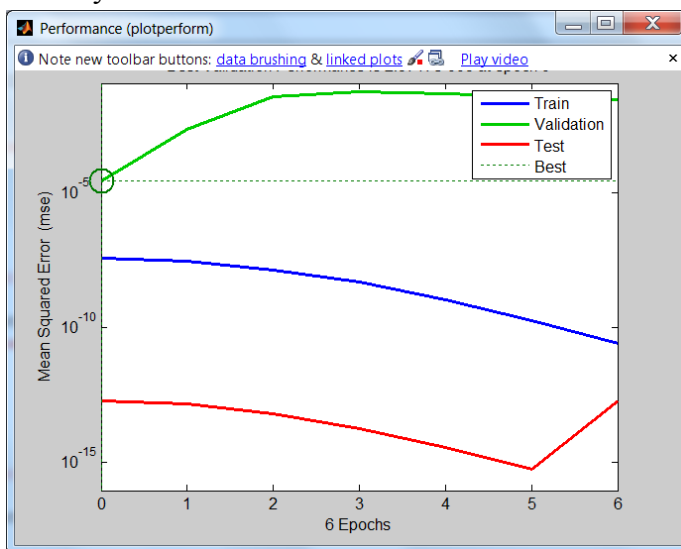


Рис. 3. Крива навчання

Останнім часом серед методів штучного інтелекту використовуються байєсовські мережі довіри [5]. Байєсовські мережі - це графічні структури для представлення ймовірнісних відносин між великою кількістю змінних і для здійснення ймовірнісного виведення на основі цих змінних. Припустимо, що відомі наступні величини: p_G - частка «добрих» клієнтів, відповідно, $p_B = 1 - p_G$ - частка «поганих» клієнтів; $p(x|G)$, $p(x|B)$, - ймовірності того, що відповідно «добрий» і «поганий» клієнт дадуть відповіді x ; L - втрати від того, що «добрий» клієнт буде класифікований як «поганий»;

D - втрати в разі дефолту, тобто класифікації «поганого» клієнта як «доброго» [4].

За формулою повної ймовірності можна обчислити:

$$p(x) = p(x|G)p_G + p(x|B)p_B \quad (1)$$

ймовірність того, що клієнт дасть відповіді x , по теоремі Байєса

$$q(G | x) = \frac{p(x | G)p_G}{p(x)}, \quad q(B | x) = \frac{p(x | B)p_B}{p(x)} \quad (2)$$

тобто ймовірності того, що клієнт з даними відповідями буде «добрим» або «поганим».

Середні втрати при розрахунку одного клієнта становлять:

$$EL = L \sum_{x \in A_B} q(G | x) p(x) + D \sum_{x \in A_G} q(B | x) p(x) = L \sum_{x \in A_B} p(x | G) p_G + D \sum_{x \in A_G} p(x | B) p_B \quad (3)$$

і є мінімальними при виборі:

$$A_G = \{x | Dp(x | B)p_B \leq Lp(x | G)p_G\} \quad (4)$$

Якщо далі припустити нормальність розподілів $p(x | G)$, $p(x | B)$, із загальною коваріаційною матрицею, то отримаємо лінійне правило

$$A_G = \{x | w_1 x_1 + \dots + w_n x_n > c\}. \quad (5)$$

Це правило фактично означає, що кожному набору відповідей x присвоюється «рахунок» $s(x)$ за формулою

$$s(x) = w_1 x_1 + \dots + w_n x_n, \quad (6)$$

тобто $s(x)$ містить достатньо інформації для розрізнення класу клієнта. Таким чином, розмірність задачі зменшується з n , з розподілами $p(x | B)$, $p(x | G)$, до 1 з розподілами $p(x | G)$, $p(s | B)$.

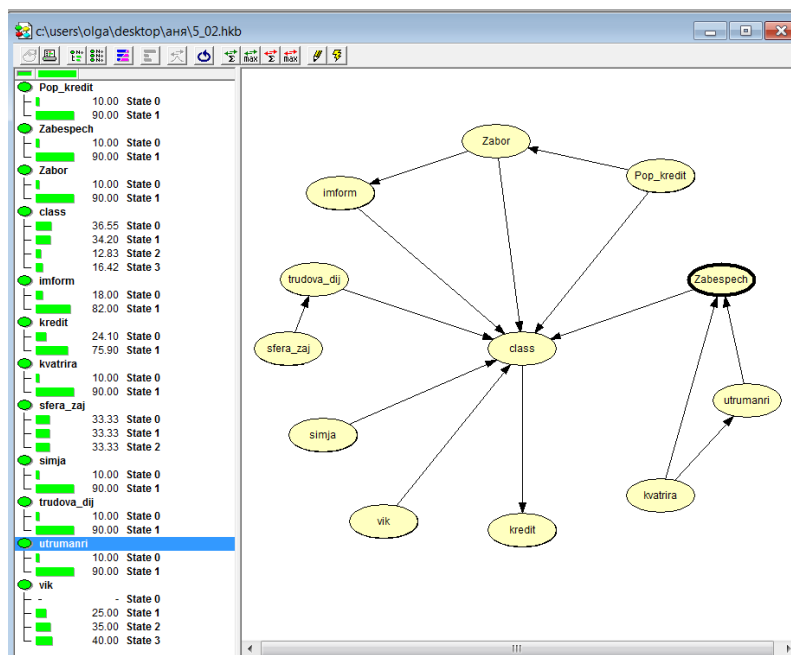


Рис. 4. Байєсовська мережа довіри в системі HUGIN

Створена байєсовська мережа довіри в програмі HUGIN. Перед тим, як назвати це повністю байєсовською мережею довіри необхідно визначити кількісне уявлення, тобто розрахувати велику кількість таблиць апіорних і умовних ймовірностей (1)-(6). Розрахунок наведено на Рис.4.

Висновки. В даній роботі досліджено методи штучного інтелекту до розв'язування задач кредитного скорингу. Зокрема розглядалися методи: класифікації,

побудови нейронної мережі прямого поширення та ймовірнісної нейронної мережі в системі Matlab. Всі методи мають право на існування. Вибір методу залежить від стратегії банку та його пріоритетів. Точність класифікації перевіряється поділом вибірки на дві частини, на одній з яких будується модель, а на іншій – перевіряється. Розрахунки проводилися в середовищі Matlab.

В системі HUGIN побудована байєсовська мережа довіри. Приклади розрахунків наведено. Безумовно це не єдиний підхід до побудови байєсовської мережі ф це необхідно досліджувати.

Список використаної літератури

1. Кадыров А.Н. Методика определения категории риска заемщика для управления уровнем риска кредитного портфеля банка // Финансы и кредит. 2002.
2. Мещеряков В.А. Применение нейросетевых технологий для решения задачи дискриминантного анализа в MATLAB // Проектирование инженерных и научных приложений в среде MATLAB: материалы V Международной научной конференции.— Харьков: «БЭТ», 2011.— С.360 – 366.
3. Медведев В. С., Потемкин В. Г. Нейронные сети. MATLAB 6. – М.: Диалог-МИФИ, 2002. – 496 с.
4. MATLAB Statistics Toolbox User's Guide. – The Mathworks, Natick, MA, 2009.
5. В. Baesens et al., "Learning bayesian network classifiers for credit scoring using Markov chain Monte Carlo search", in Proc. 16th Int. Conf. Pattern Recognition, Québec, Canada, August 2002, pp. 49–52.

Рекомендовано до публікації: д.е.н., професор Чубукова О.Ю., КНУТД
Стаття надійшла до редакції 21.02.2014

МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ИСПОЛЬЗУЕМЫЕ ДЛЯ КРЕДИТНОГО СКОРИНГА

ЦЕСЛИВ О.В.

Киевский национальный технический университет Украины "Киевский политехнический институт"

Цель. Исследовать возможность решения задачи кредитного скоринга методами искусственного интеллекта.

Методика. Используются общеизвестные методы искусственного интеллекта: классификация, построение нейронной сети прямого распространения и вероятностной нейронной сети. В системе HUGIN разработана экспертная система на основе байесовской сети доверия (БМД).

Результаты. В ходе реализации методов доказана возможность решения задачи скоринга средствами Matlab и HUGIN. Проиллюстрированы преимущества и недостатки приведённых методов. Проанализировано, что для данной задачи наилучшим методом решения является двухслойная нейронная сеть прямого распространения сигнала и вероятностная нейронная сеть. Доказано, что выбор метода зависит от стратегии банка и его приоритетов.

Научная новизна. Исследованы точность и эффективность методов кредитного скоринга в среде Matlab . и системе HUGIN.

Практическая значимость. Приведенными методами в средах Matlab и HUGIN возможно проводить кредитный скоринг в банках.

Ключевые слова: *искусственный интеллект, кредитный скоринг, нейронные сети, экспертные системы, байесовские сети, логико-вероятностные методы.*

ARTIFICIAL INTELLIGENCE METHODS USED FOR CREDIT SCORING

TSESLIV O.V.

Kyiv National Technical University of Ukraine "Kyiv Polytechnic Institute"

Purpose. To investigate the possibility of credit scoring task solving using the artificial intelligence methods.

Methodology. Common methods of artificial intelligence were used: classification, construction feed forward neural network and probabilistic neural network. Expert system based on Bayesian belief networks (BBN) was developed in HUGIN system.

Findings. During the methods implementation the possibility of scoring task solving by means of Matlab and HUGIN has been proved. Advantages and disadvantages of the given methods were illustrated. It has been analyzed that for this task the best method for solving is two-layer neural network direct signal propagation and probabilistic neural network. It has been proved that the choice of method depends on the bank's strategy and its priorities.

Originality. Accuracy and effectiveness of credit scoring methods in Matlab environment. and HUGIN system has been investigated.

Practical value. Given methods in Matlab and HUGIN environments can be used for credit scoring in banks.

Keywords: *artificial intelligence, credit scoring, neural networks, expert systems, Bayesian networks, logical - probabilistic methods*

УДК: 338.2:336

БРЕУС С. В.

Київський національний університет технологій та дизайну

ЗАХОДИ ФІНАНСУВАННЯ СОЦІАЛЬНО НЕЗАХИЩЕНИХ ВЕРСТВ НАСЕЛЕННЯ В ОСВІТЯНСЬКІЙ СФЕРІ

Мета. Розробка заходів щодо покращення фінансування сфери вищої освіти в контексті доступу до неї соціально незахищених верств населення.

Методика. Основними методами, які використані при написанні статті були такі: узагальнення теоретичних джерел; метод економічного аналізу; загально наукові прийоми аналізу і синтезу.

Результати. Системі вищої освіти в контексті фінансування притаманні певні проблеми, які негативно впливають на можливість її отримання населенням, яке не відноситься до соціально незахищених категорій, проте має право на навчання за рахунок коштів державного бюджету за результатами зовнішнього незалежного оцінювання (ЗНО).

Наукова новизна. Питання фінансового забезпечення вищої освіти, у тому числі державного фінансування, розглядались в працях багатьох авторів, зокрема, таких як: