

Економічні науки

Роговий Андрій Владиславович

студент Навчально-наукового комплексу
«Інститут Прикладного Системного Аналізу»
НТУУ "КПІ", Україна, м. Київ

МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ КРЕДИТНОГО РИЗИКУ

Анотація. У даній роботі проведено огляд найбільш поширених моделей оцінки банківських ризиків. Проведена їхня порівняльна характеристика.

Ключові слова: кредитний ризик, прогнозування, кредитоспроможність, модель.

На сьогодні побудова систем для оцінювання кредитного ризику не має чітко алгоритмізованого підходу. У більшості комерційних банків системи оцінки та менеджменту кредитних ризиків є власними розробками з різними методиками на основі даних про позичальників конкретного банку минулих років, або є готовими рішеннями спеціалізованих фірм на основі даних про позичальників декількох банків або фінансових інститутів. І в першому, і другому випадку методики побудови систем оцінки кредитного ризику, як правило, становлять комерційну таємницю. Тому дана предметна область потребує окрім необхідного багажу знань, також наявність евристичних ідей та методів [3, с1.].

Тому метою моєї роботи був огляд на найбільш актуальні для даної предметної області моделі та проведення їх порівняльного аналізу.

В якості даних для побудови прогнозу мною були взяті демографічні показники, оскільки для території України вони є найбільш показовими. А

саме, були взяті реальні дані про 1600 клієнтів банку. Інформація була представлена у розрізі, який наведено у таблиці 1.

Таблиця 1–Список змінних:

Ім'я змінних	Пояснення	Тип змінних
Age years	Вік позичальника в роках	integer
Sex	Стать особи	Female/Male
Marital status	Сімейний статус	string
Credit	Сума взятого позичальником кредиту	real
Region	Регіон проживання	integer
PrevEmployerSector	Сектор зайнятості особи	string
OccupationStatus	Професія	string
WorkExperienceMonth	Кількість місяців стажу	integer
Good/Bad	Статус повернення кредиту	Good/Bad

Модель логістичної регресії

Логістична регресія - окремий випадок узагальненої лінійної регресії. Передбачається, що залежна змінна приймає два значення і має біноміальний розподіл.

Задана вибірка - безліч m пар (x_i, y_i) , в яких опис i -го елемента $x_i \in \mathbb{R}^n$, і значення залежної змінної $y_i \in \{0, 1\}$.

Прийнята модель логістичної регресії, згідно з якою вільні змінної x і залежна змінна y пов'язані залежністю

$$y = \text{logit}^{-1}(z) + \varepsilon = \frac{1}{1 + \exp(-z)} + \varepsilon \quad (1)$$

$$\text{де } z = b_0 + \sum_{j=1}^n b_j x_j$$

де p_i - ймовірність настання дефолту по кредиту для i -го позичальника; $x_i^{(j)}$ - значення j -ої незалежної змінної; b_0 - незалежна константа моделі, b_j - параметри моделі; ε_i - компонент випадкової помилки [5].

Результати

В результаті побудови логістичної моделі були отримані наступні коефіцієнти:

$$Z = -0.0308 + 0.0231 \times \text{Age years} - 0.6233 \times \text{Sex} + 0.6047 \times \text{Education} + 0.0616 \times \text{Marital status} - 0.00002 \times \text{Credit} - 0.2673 \times \text{Region} + 0.09256 \times \text{PrevEmployerSector} - 0.103 \times \text{OccupationStatus} + 0.0064 \times \text{WorkExperience}$$

На наступному етапі вони були підставлені до логістичної кривої:

$$y = \frac{1}{1 + e^{-z}} \quad (2)$$

В результаті були отримані такі характеристики даної моделі для тестової вибірки:

- AUC=0,846;
- GINI=0,691.

На наступному кроці виникла ідея з приводу знаходження іншої моделі, яка буде давати кращі якісні показники. Тому було вирішено побудувати нейронну мережу зі зворотнім поширенням похибки.

Нейромережа зі зворотнім поширенням похибки

Нейромережа - це статистична модель, яка складається з безлічі нейронів, згрупованих в шари, що створюють мережу. Кожен нейрон - це елемент із заданою одиничною функцією, який обробляє значення, які надходять до нього. Зв'язки між нейронами створюють мережу, що дозволяє визначити взаємозв'язку між окремими даними. [2, с7.] Типовий приклад структури нейромережі показаний на рисунку 1:

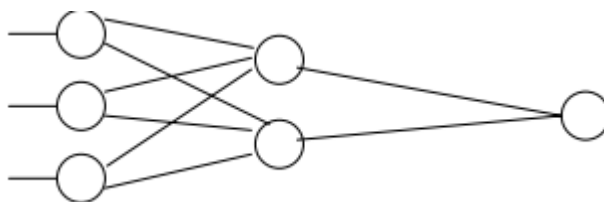


Рисунок 1–Нейронна мережа

Кожен нейрон складається з двох елементів.

Перший елемент - дендрит - додає вагові коефіцієнти до вхідних сигналів.

Другий елемент - тіло - реалізує нелінійну функцію, функцію активації нейрона.

Сигнал e - це зважена сума вхідних сигналів. Його ілюстрація наведена на рис. 2.

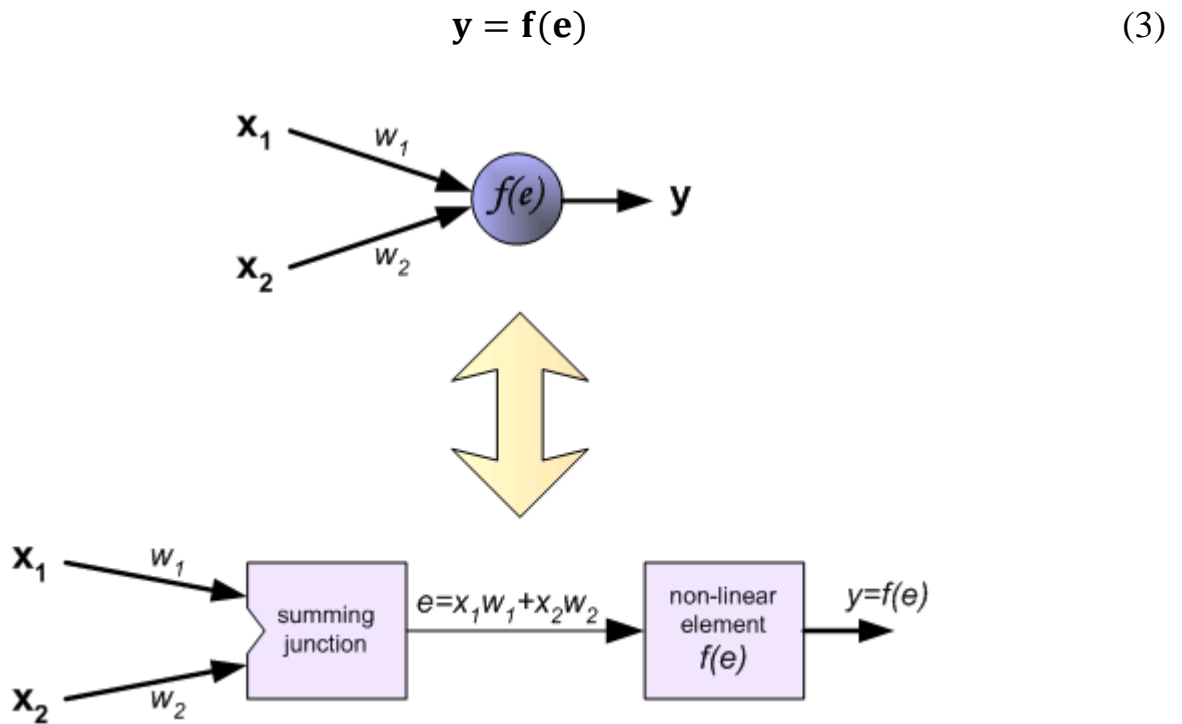


Рисунок 2–Вигляд нейрона

Щоб навчити нейронну мережу необхідно підготувати навчальні дані (приклади). Навчання - це послідовність ітерацій (повторень). У кожній ітерації вагові коефіцієнти нейронів підганяються з використанням нових даних з тренувальних прикладів [3].

Розроблений метод вибору топології

Перед початком побудови нейромережі необхідно було підібрати для неї найкращу топологію. Вибір топології нейромережі досі не має чіткого сформульованого підходу. В результаті ознайомлення з рядом джерел у цьому питанні мною був визначений алгоритм на основі якого й відбувалася побудова топології нейронної мережі.

Нехай в мережі N прихованих шарів, тоді існує $N + 1$ способів модифікації. Перші N способів - додати 1 нейрон в кожен прихований шар, $N + 1$ -й спосіб - додати новий шар з двох нейронів між останнім прихованим шаром і вихідним, шар з одного нейрона не має сенсу. Можна додавати тільки один нейрон за раз, а потім тестувати, додавання більшої кількості нейронів потребує більше часу настройки. Кожен спосіб нарощування необхідно тестувати.

В результаті даного алгоритма було помічено, що якість моделі давала кращі показники при збільшенні нейронів в одному прихованому шарі (рис. 3), а при додаванні 2-го шару якість моделі різко погіршувалася.

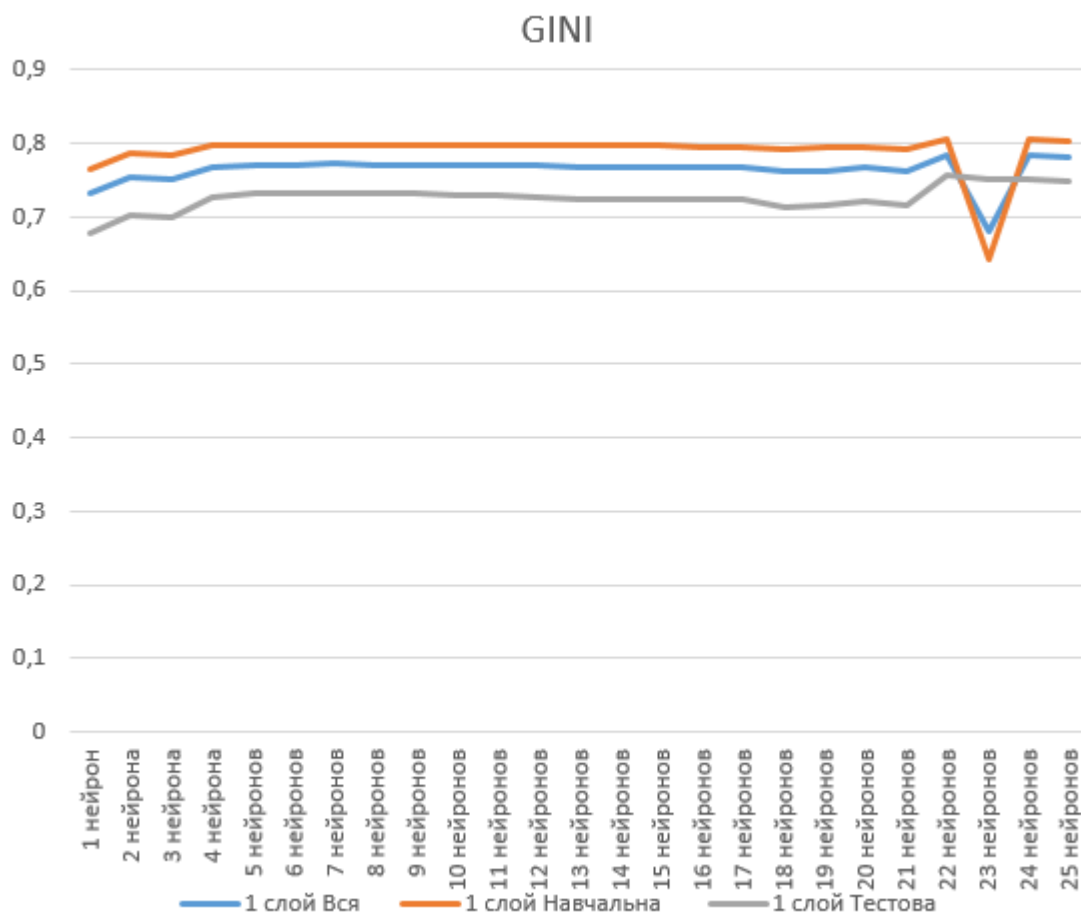


Рисунок 3—Залежність показника GINI від зміни кількості нейронів в одному прихованому шарі

Як бачимо найкращі показники для навчальної та всієї вибірки були отримані при 1-му прихованому слою та 24 нейронах в ньому, а для тестової при 1-му прихованому слою та 22 нейронах в ньому. Цікаво зазначити, що після 25-го нейрона в 1-у прихованому слою якість моделі різко погіршувалася.

Результати

В результаті цих спостережень була обрано топологія з 1-го прихованого слою та 22 нейронах у ньому. Оскільки при таких умовах тестова вибірка давала найкращі показники.

Були отримані такі характеристики даної моделі для тестової вибірки:

- AUC=0,878
- GINI=0,756

Загальний вигляд моделі наведено на рис. 4.

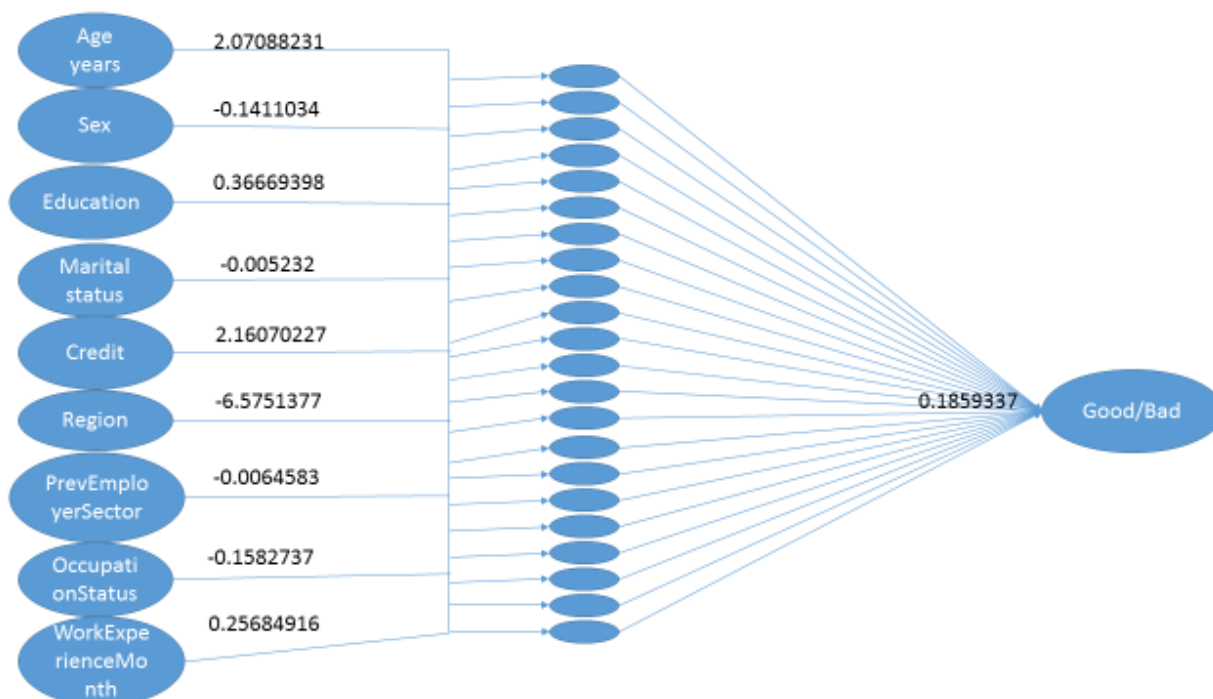


Рисунок 4–Визгляд побудованої нейронної мережі

Ця модель давала ще кращі показники ніж логістична регресія. Але при даній топології мала серйозний недолік у швидкодії. Оскільки оцінку ризику кредитоспроможності клієнта бажано постійно перераховувати через стрімкі зміни в економіці було вирішено також дослідити роботу нечіткої каскадної

нео-фазі нейронної мережі. На практиці такі моделі дають значно більші якісні показники через те, що вони більше пристосовані до нечіткості вхідної інформації [4]. Оскільки в Україні в якості вхідних параметрів для оцінки кредитоспроможності клієнта часто використовують демографічні показники, які заповнюються клієнтом власноруч, то врахування нечіткості такої інформації є дуже актуальним.

Нечітка каскадна нео-фазі нейронна мережа

Для початку розглянемо нео-фаззі нейрон з декількома входами і єдиним виходом, який зображений на рис. 5.

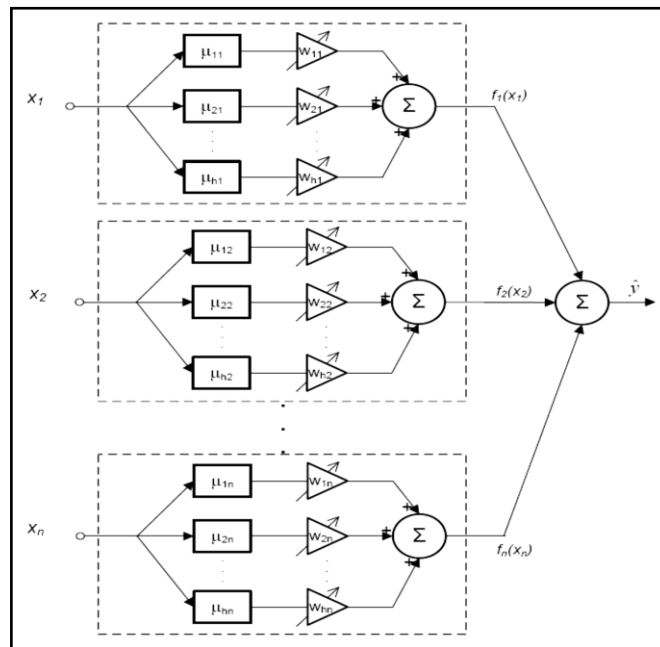


Рисунок 5–Архітектура нео-фаззі нейрона

Вихід реалізується таким відображенням:

$$\hat{y} = \sum_{i=1}^n f_i(x_i), \quad (4)$$

де x_i -і-й вхід ($i=1,2,\dots,n$),

\hat{y} -вихід системи.

Структурні блоки нео-fuzzy нейрона є нелінійним синапсом NSi, який переводить i -й вхідний сигнал в форму:

$$f_i(x_i) = \sum_{j=1}^h w_{ji} \mu_{ji}(x_i) \quad (5)$$

і виконує нечіткий висновок: Якщо $x_i \in x_{ji}$ те вихід є w_{ji} , де - нечітке число, функція належності якого μ_{ji} , w_{ji} -синаптична вага. Очевидно, що нелінійний синапс фактично реалізує нечіткий висновок Такагі-Сугено нульового порядку.

$$\hat{y}^k = \sum_{i=1}^n f_i(x_i(k)) = \sum_{i=1}^n \sum_{j=1}^k w_{ji}(k-1) \mu_{ji}(x_i(k)) \quad (6)$$

Тепер розглянемо загальну архітектуру каскадної нео-фаззі нейронної мережі (CNFNN) показана на рис. 6.

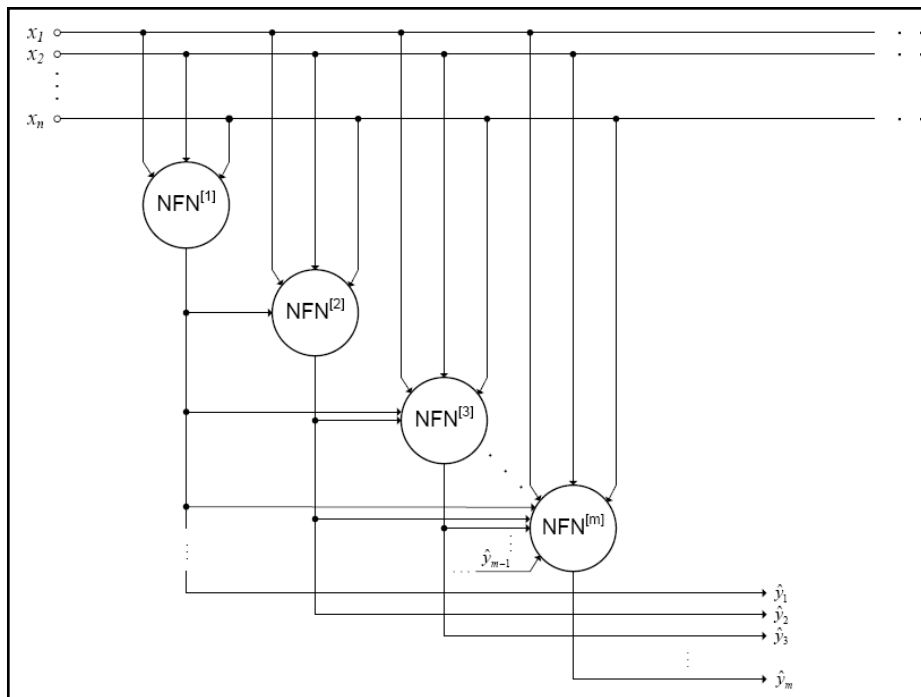


Рисунок 6–Архітектура каскадної нео-фаззі нейронної мережі

-нео-фаззі нейрон першого каскаду:

$$\hat{y}^{[1]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[1]} \mu_{ji}(x_i) \quad (7)$$

-нео-fuzzy нейрон m-го каскаду:

$$\hat{y}^{[m]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[m]} \mu_{ji}(x_i) + \sum_{i=n+1}^{n+m-1} \sum_{j=1}^h w_{ji}^{[m]} \mu_{ji}(\hat{y}^{[i-n]}) \quad (8)$$

Серед найбільш важливих переваг нео-фаззі-нейрона можна відзначити високу швидкість навчання, обчислювальну простоту, можливість знаходження глобального мінімуму критерію навчання в режимі реального

часу. Критерієм навчання (цільовою функцією) є стандартна локальна квадратична функція помилки [1]:

$$E(k) = \frac{1}{2} (y(k) - \hat{y}(k))^2 = \frac{1}{2} e^2 = \frac{1}{2} (y(k) - \sum_{i=1}^n \sum_{j=1}^h w_{ji} \mu_{ji} (x_i(k)))^2 \quad (9)$$

Результати

На початку розробки даної моделі постала задача у виборі необхідної кількості термів та каскадів для найкращої роботи мережі. При збільшенні кількості каскадів та варіюванні каскадами було помічено, що модель дає кращі результати при 4-х термах (таблиця 2).

Таблиця 2–Залежність показника GINI від варіювання кількості термів та каскадів

		1 касад	2 касада	3 касада	4 касада	5 касадів
2 терма	Тестова	0.699113	0.701226	0.700932	0.700521	0.700189
	Навчальна	0.772138	0.773694	0.773943	0.773427	0.773455
3 терма	Тестова	0.644472	0.673308	0.678121	0.669846	0.703652
	Навчальна	0.748364	0.768606	0.769803	0.773625	0.791113
4 терма	Тестова	0.707701	0.698232	0.747552	0.75839	0.748843
	Навчальна	0.799406	0.801147	0.814003	0.816977	0.814638
5 термів	Тестова	0.626297	0.653823	0.665972	0.680097	0.689859
	Навчальна	0.746324	0.760064	0.766023	0.773197	0.777341

Тому в подальшому дослідженні було обрано розбиття на 4 терми.

При збільшенні каскадів вже менш стрімко збільшувалися якісні характеристики моделі та одночасно суттєво погіршувалася швидкодія моделі. Тому була обрана модель з 4-а термами та 40 каскадами.

Аналіз результатів роботи моделей

Мною були обрані такі моделі: логістична регресія, нейронна мережа зі зворотнім поширенням похибки, каскадна нечітка нео-фаззі нейронна мережа

Були визначені параметри при яких дані моделі давали найкращі показники AUC та GINI. Для нейронної мережі зі зворотнім поширенням похибки найкращі результати показала модель з топологією, яка складалася з 1-го прихованого шару та 22-х нейронів у ньому. Також було встановлено, що при кількості нейронів в одному прихованому шарі більшій за 25 або при збільшенні кількості прихованих шарів якісні показники моделі різко зменшуються. Для нечіткої каскадної нео-фазі-нейронної мережі найкращі якісні характеристики показала модель з 4-х термів та 40 каскадів.

Найкращі показники AUC, яких вийшло досягнути занесені до табл. 3

Таблиця 3 – Індекс AUC для кожної з моделей

	Тестова	Навчальна
Логістична регресія	0.845575	0.887478
Нейромережа	0.902321	0.877933
Нечітка нейромережа	0.931058	0.943658

Найкращі показники GINI наведені у таблиці 4:

Таблиця 4–Індекс GINI для кожної з моделей

	Тестова	Навчальна
Логістична регресія	0.69115	0.774956
Нейромережа	0.804641	0.755867
Нечітка нейромережа	0.862116	0.887317

Як бачимо, найкращі якісні показники дала нечітка каскадна нео-фазі-нейронна мережа. Тому для нашої вибірки найкраще взяти саме таку мережу з 4 термами та 40 каскадами.

Література:

1. Агаї Аг Гаміш Ові Нафас. Прогнозування ризику банкрутства в промисловій та банківській сфері з використанням нечітких моделей та

- алгоритмів: дис. кандидата технічних наук : 01 04 2016 р / Агаї Аг Гаміш Ові Нафас К., 2016. – 276 с.
2. Ковальов М.С. Методика построения банковских скоринговых моделей для оценки кредитоспособности физических лиц [Электронный ресурс] / Ковальов М.С. // Интернет-журнал «Науковедение» Выпуск 2, март – апрель 2014.— Режим доступа: <http://www.bsu.by/Cache/pdf/49623.pdf>
 3. Л.Г.Комарцова, А.В. Максимов Нейрокомпьютеры. - Москва : МГТУ им. Н.Э.Баумана, 2002.
 4. Кузнецова Н.В., Бідюк П.І. [Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування](#) [Электронный ресурс] / Н.В. Кузнецова, П.І. Бідюк //Вестник ХНТУ №1(34), 2010 г. Экономика и менеджмент. — Режим доступа: http://www.nbu.gov.ua/old_jrn/natural/Vkhdtu/2010_2/02_003.pdf
 5. Паклин Н. Логистическая регрессия и ROC-анализ — математический аппарат [Электронный ресурс] – Режим доступа: <http://www.basegroup.ru/library/analysis/regression/logistic/>