

Інформаційна технологія розв'язання задачі скорингу на основі нейронної мережі

Магістерська кваліфікаційна робота
напряму 122 – “Комп'ютерні науки”

Виконала: Семенова Л.М.
ст. групи 1КН-18м

Науковий керівник:
Колесницький Олег Костянтинович

Актуальність теми дослідження

Тема магістерської кваліфікаційної роботи є актуальною, оскільки завжди існує потреба у точному визначенні скорингу для ефективнішої роботи з клієнтами. Використання при цьому нейронної мережі дозволить врахувати нелінійні залежності між відповідями клієнтів та їх кредитоспроможністю.

Дана технологія буде корисною для державних та приватних банківських та маркетингових систем, особливо в умовах її незначної вартості придбання, розробки та обслуговування.

Використання методів штучного інтелекту для розв'язання поставленої задачі надає великі можливості для модифікацій і налаштування різних методів під дану задачу, а отже дозволяє таким чином досягати кращий у порівнянні з традиційними існуючими методами результат.

Об'єкт, предмет та мета дослідження

Об'єктом дослідження є процес розв'язання задачі скорингу комп'ютерними засобами.

Предмет дослідження – це інформаційна технологія та програмні засоби розв'язання задачі скорингу з використанням штучних нейронних мереж та достовірність їх роботи.

Метою дослідження магістерської кваліфікаційної роботи є підвищення достовірності розв'язання задачі скорингу програмними засобами за рахунок використання штучних нейронних мереж.

Завдання дослідження

Для досягнення поставленої мети необхідно розв'язати такі наступні завдання:

- провести аналіз проблеми розв'язання задачі скорингу;
- розглянути існуючі методи вирішення задачі скорингу та обрати й обґрунтувати вибір методу, який задовольняє мету даної магістерської кваліфікаційної роботи;
- розробити математичну модель інформаційної системи розв'язання задачі скорингу та удосконалити її згідно з метою роботи;
- сформулювати стадії інформаційної технології та на їх основі розробити структуру та алгоритм роботи програмного засобу;
- виконати програмну реалізацію запропонованої інформаційної технології розв'язання задачі скорингу;
- провести тестування програмного продукту та виконати аналіз отриманих результатів.

Наукова новизна одержаних результатів

- знайшла подальшого розвитку інформаційна технологія розв'язання задачі скорингу, яка використовує нейронну мережу радіально-базисних функцій, що дозволяє підвищити достовірність розв'язання задачі скорингу;
- удосконалено модель навчання нейронної мережі РБФ за рахунок введення процедури підбору початкових значень вагових коефіцієнтів вихідного шару, що дозволило суттєво підвищити швидкість навчання нейронної мережі.

Апробація результатів роботи

Результати роботи були апробовані на Всеукраїнській науково-практичній інтернет-конференції та плануються до впровадження у виробництво.

За результатами дослідження опубліковано 3 тези доповідей конференцій.

Постановка задачі

- Завдання визначення скорингу ставиться таким чином. Є певний набір відповідей клієнта на питання анкети, що належать заздалегідь відомій кінцевій кількості класів $C = \{C_1, \dots, C_q\}$. Є деяка кінцева кількість об'єктів (навчальна вибірка), про кожен з яких відомо до якого класу він належить. Потрібно розробити програму, яка для будь-якого вхідного набору відповідей, який не обов'язково має належати навчальній множині, вирішував би, до якого класу цей об'єкт належить і знаходив його скоринговий бал.

Постановка задачі

У спрощеному вигляді скорингова модель – це зважена сума визначених характеристик клієнта.

$$S = a_1 * X_1 + a_2 * X_2 + \dots + a_k * X_k,$$

де S - значення скорингу,

X_1, X_2, \dots, X_k – параметри клієнта, які входять в оцінку його кредитоспроможності,

a_1, a_2, \dots, a_k – вагові коефіцієнти, які характеризують значимість відповідних параметрів клієнта для формування кредитного скорингу

Відомі методи реалізації задачі скорингу

Задача скорингу передбачає задачу класифікації.

Відомі класифікаційні методи:

- статистичні методи, що ґрунтуються на дискримінаційному аналізі;
- лінійне програмування;
- нейронні мережі;
- генетичні алгоритми;
- метод найближчих сусідів та ін.

Для реалізації інформаційної технології було обрано нейронні мережі.

Порівняння методів реалізації задачі скорингу



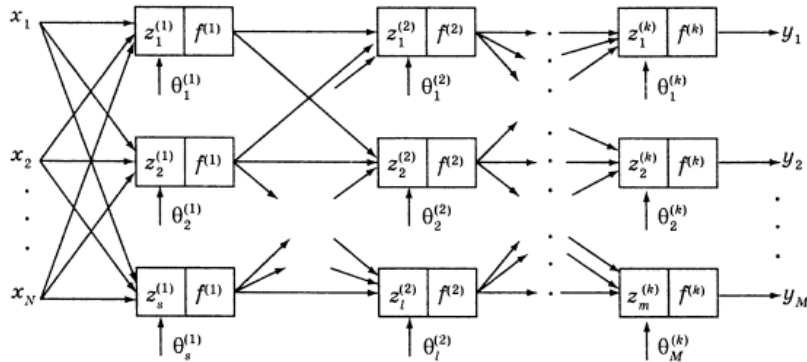
Вибір програми-аналогу

Відомі сервіси, які автоматизовують процес скорингу:

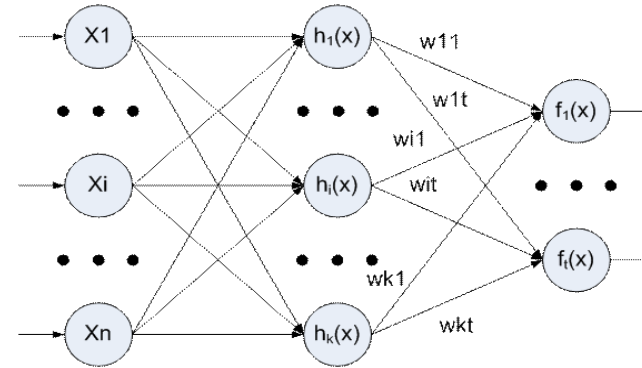
- SAS Credit Scoring
- EGAR Scoring
- Plug & Score Modeler
- Clementine (IBM SPSS Modeler)

В якості програми-аналога було обрано програму Clementine (IBM SPSS Modeler).

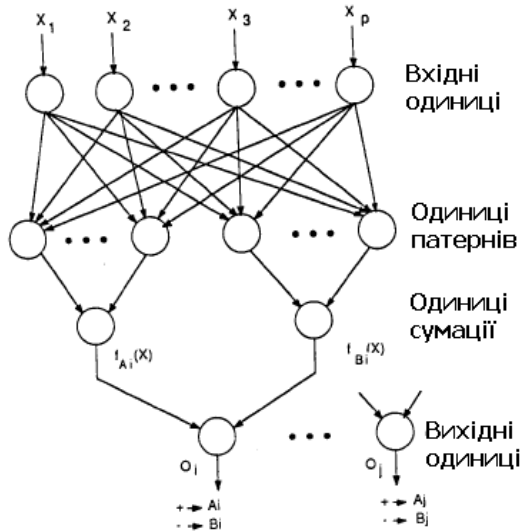
Обґрунтування вибору типу нейронної мережі



Багатошаровий перцептрон



Нейронна мережа РБФ



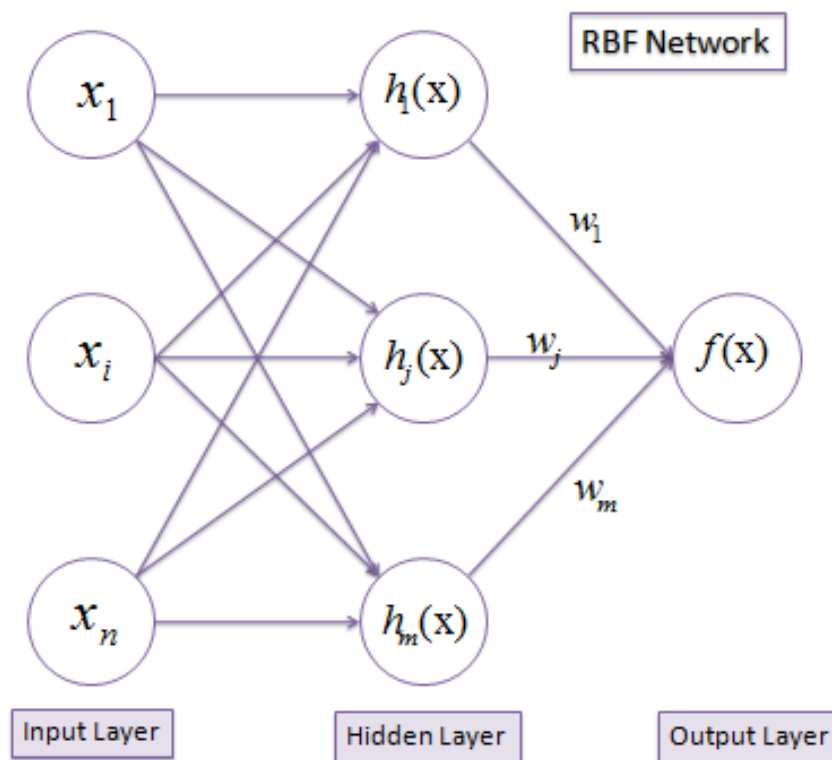
Ймовірнісна нейронна мережа

Було обрано нейронну мережу радіально-базисних функцій, яка навчається за методом зворотного поширення помилки (мережа зворотного поширення).

Архітектура РБФ-мережі

- Архітектура мережі складатиметься з трьох шарів: вхідний шар, прихований шар з нелінійною RBF-функцією активації та лінійний вихідний рівень
- Кількість нейронів на вхідному шарі 20, що визначається кількістю тестових запитань для клієнта.
- У прихованому шарі кількість нейронів становить 50 – оптимальна кількість, яка буде достатня для досягнення потрібної точності та достовірності, і не дуже велика, щоб не витратити великої кількості ресурсів.
- На вихідному рівні буде 2 нейрони, оскільки мережа має розділяти клієнтів на два класи.
- Виходи нейромережі є цифровими, оскільки вони показують скорингову оцінку клієнта. Діапазон значень складає 0..1. Чим ближче значення до 1, тим вища скорингова оцінка.

Архітектура та математична модель РБФ-мережі



$$f(x) = \sum_{j=1}^m w_j h_j(x)$$

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right)$$

Удосконалення методу навчання нейронної мережі

У роботі пропонується удосконалити модель навчання нейронної мережі радіально-базисних функцій за рахунок введення процедури підбору початкових значень вагових коефіцієнтів вихідного шару, що дозволить суттєво підвищити швидкість навчання та достовірність визначення скорингової оцінки.

Навчання проходить у три етапи: обирається центр вектору RBF функції у прихованому шарі, загальна цільова функція у вигляді функції найменших квадратів, метод зворотного поширення помилки.

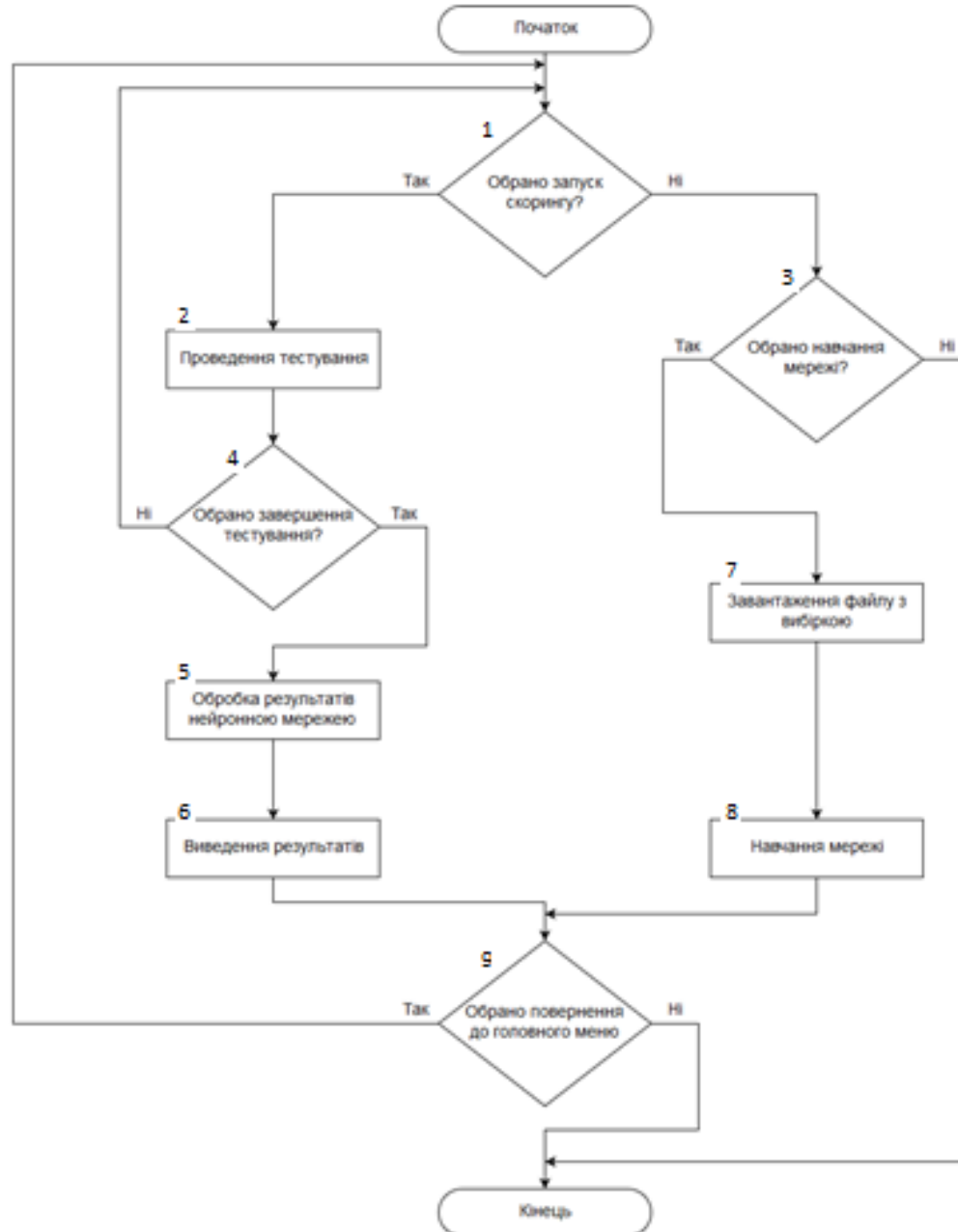
$$K(\mathbf{w}) \stackrel{\text{def}}{=} \sum_{t=1}^T K_t(\mathbf{w})$$

При процесі навчання буде додатково врахований вектор початкових значень вагових коефіцієнтів вихідного шару w_j .

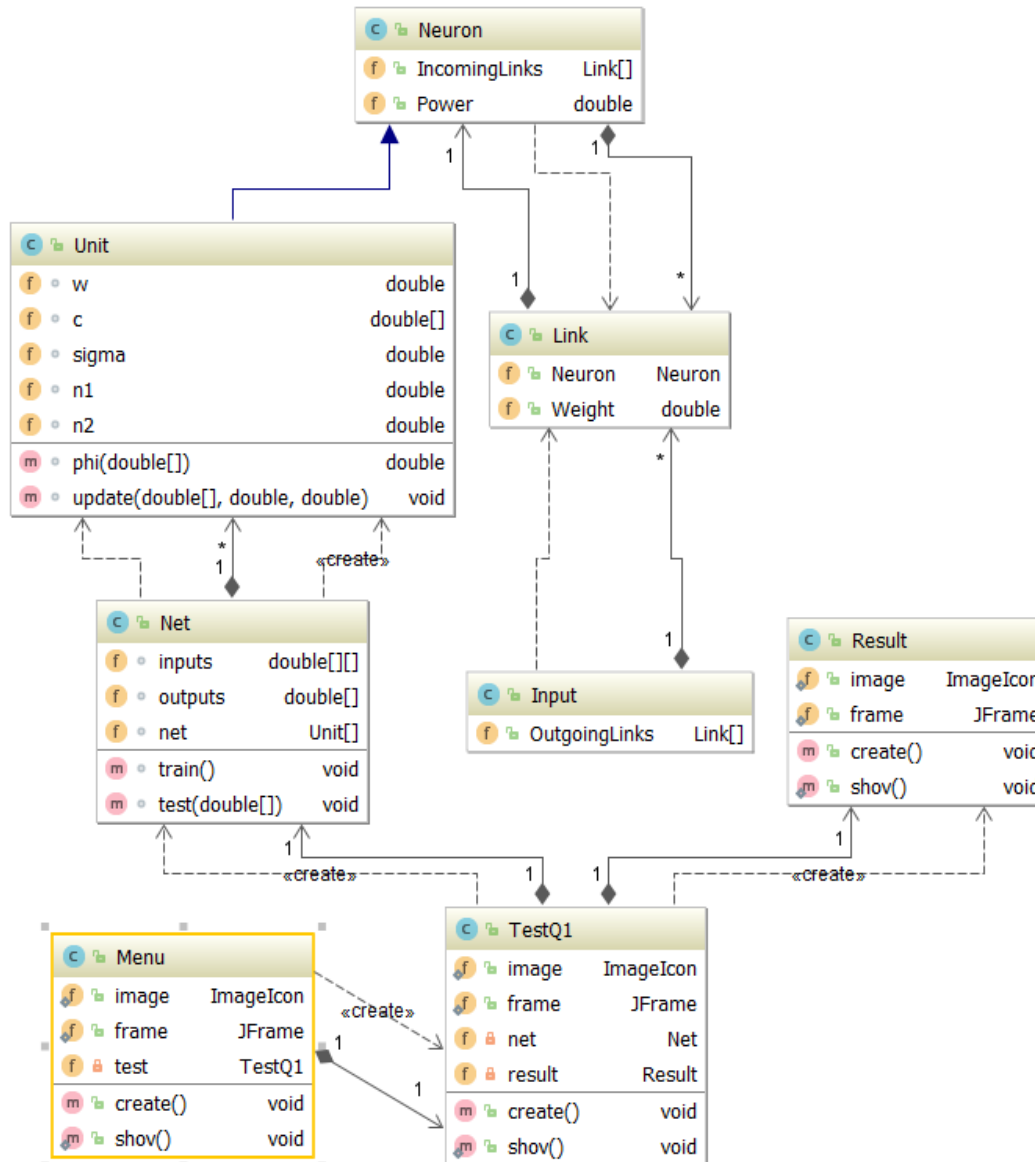
Структура інформаційної технології



Алгоритм роботи програмного забезпечення



UML-діаграма класів

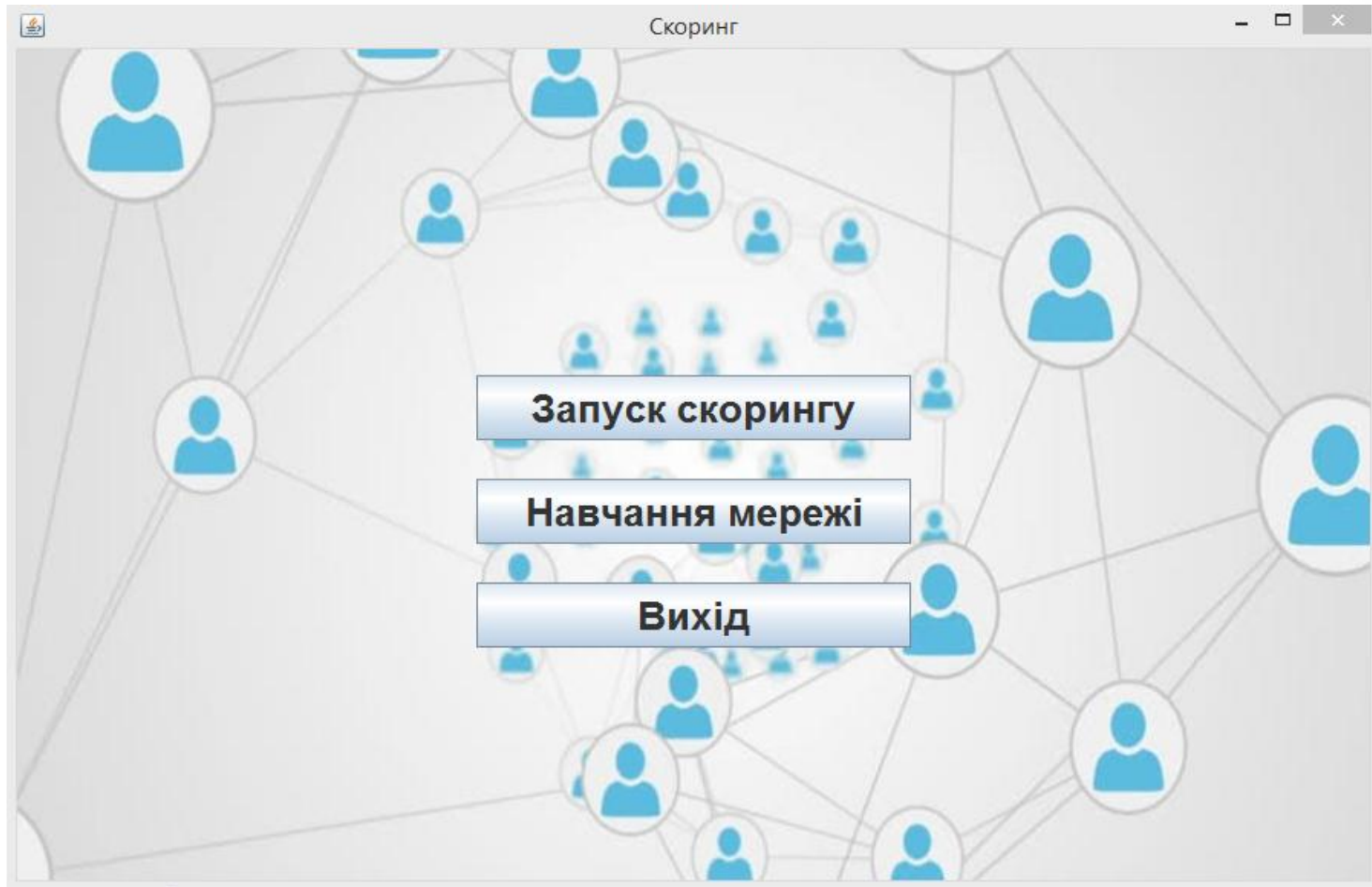


Вибір мови та середовища програмування

Було розглянуто мови програмування C++ та Java.

Для реалізації програмного модуля було обрано мову програмування Java з використанням технології Swing та середовище розробки IntelliJ IDEA.

Стартове вікно програми



Робочі вікна програми

Скоринг

Вкажіть Вашу стать?

Вкажіть Ваш вік?

Вкажіть Ваш місячний прибуток?

Який Ваш соціальний статус?

Який Ваш сімейний стан?

Скоринг

Результат: 0,837
Клієнту можна видати кредит

Тестування та аналіз результатів

Таблиця 3.1 – Результати тестування програмного забезпечення

Параметр	Програма-аналог	Розроблена програма
Розмір тестової вибірки	100	100
Кількість правильно визначених прогнозів	85	92
Кількість неправильно визначених прогнозів	15	8
Достовірність	85%	92%

Із таблиці видно, що розроблений програмний модуль має вищу достовірність роботи (92%), ніж аналогічна програма (85%), а значить достовірність визначення скорингу покращена на 7%, тобто мета роботи досягнута.

Висновки

- В ході виконання магістерської кваліфікаційної роботи було розроблено інформаційну технологію розв'язання задачі скорингу з використанням нейронної мережі.
- На основі аналізу переваг та недоліків методів реалізації скорингу було обрано нейромережевий метод для реалізації інформаційної технології.
- Як нейронну мережу для реалізації було обрано мережу радіально-базисних функцій.
- Запропоновано удосконалення моделі навчання за рахунок введення процедури підбору початкових значень вагових коефіцієнтів вихідного шару.
- За результатами тестування достовірність роботи програмного забезпечення становила 92%, що на 7% вище, ніж у програми-аналога.

Дякую за увагу!