

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Факультет прикладної математики

Кафедра прикладної математики

«До захисту допущено»

Завідувач кафедри

_____ О. Р. Чертов

«___» _____ 2015 р.

Дипломна робота

на здобуття ступеня бакалавра

зі спеціальності 6.040301 «Прикладна математика»

на тему: Математичне та програмне забезпечення автоматизованої системи
визначення ступеня кредитоспроможності особи

Виконав: студент 4 курсу, групи КМ-11

Булах Олексій Романович

Керівник

асистент Тавров Д. Ю.

Консультант із нормоконтролю старший викл. Мальчиков В. В.

Рецензент

доцент, к.т.н., доцент Сапсай Т. Г.

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших
авторів без відповідних посилань.
Студент _____

Київ – 2015 року

**Національний технічний університет України
«Київський політехнічний інститут»**

Факультет прикладної математики

Кафедра прикладної математики

Рівень вищої освіти – перший (бакалаврський)

Спеціальність 6.040301 «Прикладна математика»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ О. Р. Чертов

«___» _____ 2015 р.

**ЗАВДАННЯ
на дипломну роботу студенту
Булаху Олексію Романовичу**

1. Тема роботи «Математичне та програмне забезпечення автоматизованої системи визначення ступеня кредитоспроможності особи»,
керівник роботи Тавров Данило Юрійович,
затверджені наказом по університету від «19» травня 2015 р. № 1039-С.
2. Термін подання студентом роботи: «12» червня 2015 р.
3. Вихідні дані до роботи: статистичні дані про фізичних осіб — позичальників.
4. Зміст роботи: виконати аналіз літературних джерел та існуючих програмних рішень у галузі побудови скорингових моделей, вибрати метод побудови нечіткої скорингової моделі, що не потребує залучення експерта, спроектувати автоматизовану систему для визначення ступеня кредитоспроможності позичальника та програмно її реалізувати, провести випробування розробленої системи.
5. Перелік ілюстративного матеріалу: схема взаємодії підсистем, екранні форми програми, блок-схеми розроблених алгоритмів.

6. Дата видачі завдання «23» лютого 2015 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	Вивчення літератури за тематикою роботи	10.11.2014	
2.	Проведення порівняльного аналізу математичних методів побудови скорингових моделей	08.12.2014	
3.	Розроблення математичного забез- печення нечіткої скорингової сис- теми, що не потребує залучення експерта	19.01.2015	
4.	Розробка мови управління програ- мами та проектування користува- цького інтерфейсу	16.02.2015	
5.	Розроблення програмного забезпе- чення системи	20.04.2015	
6.	Відлагодження програмного забез- печення, підготовка та розв'язування контрольних задач на комп'ютері	18.05.2015	
7.	Оформлення пояснівальної запис- ки до дипломної роботи	08.06.2015	

Студент _____

Булах О. Р.

Керівник роботи _____

Тавров Д. Ю.

АНОТАЦІЯ

Дипломну роботу присвячено розробці математичного та програмного забезпечення автоматизованої системи визначення ступеня кредитоспроможності особи.

У роботі проведено аналіз існуючих скрингових моделей, програмних систем, орієнтованих на застосування скрингових моделей до визначення кредитоспроможності позичальників. Обрано метод побудови нечіткої скрингової моделі на базі еволюційних обчислень, що не потребує заолучення експерта, котрий визначає правила для визначення кредитоспроможності. Запропоновано модифікацію методу у вигляді меметичного алгоритму, який дозволяє одержувати лішні результати з погляду точності класифікації.

Розроблено автоматизовану систему на основі системи нечіткого виведення для визначення ступеня кредитоспроможності особи. Роботу системи випробувано за допомогою реальних даних.

Роботу виконано на 50 аркушах, вона містить 2 додатки та перелік посилань на використані джерела з 24 найменувань. У роботі наведено 11 рисунків та 1 таблицю.

Ключові слова: система нечіткого виведення, еволюційні обчисlenня, меметичний алгоритм, скрингова система.

ABSTRACT

The thesis is devoted to developing mathematical models and software for the automated system for estimating the person's creditworthiness degree.

In this work, scoring models and software systems aiming at applying the scoring models to estimating the borrower creditworthiness are analyzed. An evolutionary based method for building the fuzzy scoring model is chosen. This method does not involve experts for defining the rules of creditworthiness estimation. A modification to the method is proposed, which allows to obtain results of a higher classification accuracy.

The fuzzy inference based automated system for estimating the person's creditworthiness degree is developed. The system is tested with the real data.

The work is presented in 50 pages. It contains 2 appendices and bibliography of 24 references. Eleven pictures and 1 table are given in the thesis.

Keywords: fuzzy inference system, evolutionary computing, memetic algorithm, scoring system.

Зміст

Вступ.....	8
1 Постановка задачі.....	10
2 Огляд існуючих скорингових моделей	11
2.1 Скорингова модель оцінки кредитоспроможності на основі різних типів статистичних моделей	11
2.2 Скорингова модель оцінки кредитоспроможності осіб на основі гіbridних експертних систем	14
2.3 Використання штучних нейронних мереж у скорингових системах оцінки ризиків.....	17
2.4 Нечітка логіка в роботі скорингових систем.....	18
2.5 Висновки до розділу	20
3 Математичне забезпечення	21
3.1 Системи нечіткого виведення.....	21
3.1.1 Поняття нечіткої множини.....	21
3.1.2 Поняття лінгвістичної змінної та нечіткого висловлювання	21
3.1.3 Загальний опис системи нечіткого виведення	23
3.1.4 Етапи побудови системи нечіткого виведення	25
3.2 Опис меметичного алгоритму побудови бази нечітких правил для СНВ	26
3.2.1 Загальна схема алгоритму	26
3.2.2 Представлення індивідів у популяції	27
3.2.3 Функція пристосованості	28
3.2.4 Відбір індивідів для виконання операцій варіації	29
3.2.5 Оператори варіації в алгоритмі	30
3.2.6 Меметична частина алгоритму	31
3.3 Висновки до розділу	32
4 Програмне забезпечення	33

4.1 Архітектура автоматизованої системи.....	33
4.1.1 Підсистема конвертації даних про клієнта	35
4.1.2 Підсистема нечіткого виведення	36
4.1.3 Підсистема генерування правил.....	38
4.1.4 Підсистема локального пошуку найпристосованіших правил ...	39
4.1.5 Підсистема визначення характеристик бази правил	40
4.1.6 Опис бази даних системи	42
4.2 Структура розроблених класів.....	43
4.3 Випробування роботи системи	44
4.4 Висновки до розділу	46
Висновки	47
Перелік посилань.....	48
Додаток А Лістинги програм	51
Додаток Б Ілюстративний матеріал.....	63



ВСТУП

На сьогоднішній день у людей та підприємств усіх форм власності все частіше виникає потреба залучення позикових коштів для здійснення своєї діяльності і отримання прибутку. Найрозповсюженнішою формою залучення коштів є отримання банківського кредиту. Потенційному позичальнику потрібно надати документи для того, щоб банк-кредитор здійснив оцінку його кредитоспроможності. За результатами проведеної оцінки банк вирішує питання щодо надання кредиту [1].

Кредитоспроможність — це [2] наявність у боржника передумов для проведення кредитної операції і його спроможність повернути борг у повному обсязі та в обумовлені договором строки.

Визначення оцінки кредитоспроможності позичальника передбачає поєднання об'єктивних даних, сформованих на основі бухгалтерської звітності, та суб'єктивних даних, які визначають на підставі інформації, що міститься в пакеті документів, який подає позичальник для одержання кредиту, а також інформації, отриманої з інших джерел.

Скоринг — це [3] метод класифікації позичальників для оцінки кредитного ризику, що являє собою математичну або статистичну модель, за допомогою якої на основі кредитної історії «минулих» клієнтів банк намагається визначити, наскільки велика ймовірність, що конкретний позичальник поверне кредит у відведеній термін. Задача скорингу полягає в тому, щоб перевести різномірну інформацію з розряду характеристик позичальника в категорію специфічних значень, привласнити відомостями певну кількість балів, а потім виявити комбінацію факторів, які дозволяють найліпше пояснити причини минулих неповернень кредитів [3].

Більшість існуючих скорингових моделей є чіткими, не враховуючи нечітку природу інформації про особу (вік, рівень доходу, кількість дітей й сім'ї, стаж роботи тощо). У літературі наведено приклади нечітких скорин-

гових систем, у яких використовується залучення експерта для визначення набору нечітких правил.

Дипломну роботу присвячено створенню автоматизованої системи побудови та використання скорингових моделей у вигляді систем нечіткого виведення для визначення кредитоспроможності позичальників. Для побудови скорингової моделі використовується меметичний алгоритм, що дозволяє будувати базу нечітких правил без залучення експерта.



1 ПОСТАНОВКА ЗАДАЧІ

Метою даної дипломної роботи є створення математичного та програмного забезпечення автоматизованої системи визначення ступеня кредитоспроможності особи.

Головними вимогами до математичного та програмного забезпечення є врахування нечіткої природи даних про потенційних позичальників та автоматизація процесу побудови бази нечітких правил, за якими здійснюють класифікацію.

Математичне забезпечення має забезпечувати процес автоматизованого генерування бази нечітких правил без залучення експерта. Вибір математичного апарату має бути простим у реалізації та найточніше описувати предметну область застосування.

Програмне забезпечення має давати користувачеві наступні функції:

- заповнення інформації про позичальників та її збереження в базі даних;
- уведення параметрів скорингової моделі як системи нечіткого виведення та меметичного алгоритму побудови бази її нечітких правил;
- автоматизовану генерацію бази нечітких правил скорингової моделі на базі навчальної вибірки та оцінювання класифікаційної здатності скорингової моделі на тестовій вибірці;
- оцінювання ступеня кредитоспроможності довільного позичальника та збереження результатів до бази даних позичальників.

Програмне забезпечення має виконуватися в операційній системі Windows 7 на IBM-сумісному комп’ютері, до складу якого входять:

- x86-сумісний процесор із тактовою частотою, не меншою за 1,5 ГГц;
- оперативна пам’ять обсягом, не меншим за 256 МБ.

2 ОГЛЯД ІСНУЮЧИХ СКОРИНГОВИХ МОДЕЛЕЙ

2.1 Скорингова модель оцінки кредитоспроможності на основі різних типів статистичних моделей

У випадку використання статистичних моделей для розв'язання задачі скорингу, підготовка вхідних даних є [4] першим етапом. Велика база даних клієнтів є тією частиною автоматизованої скорингової системи, яка найбільше відповідає за достовірність кінцевого результату. Для статистичної оцінки даних потрібно мати базу даних за 2–5 років. При цьому з даної бази має бути видалено:

- записи, які містять нестандартні випадки (аномально великі суми кредиту, незвичайні цілі для займу, клієнти з нестандартними формами оплати тощо);
- записи, які містили відмову через порушення політики банку при подачі заявки (банкрутство, неповноліття, подвійні заяви тощо);
- інсайдерські кредити (кредити VIP, працівників банку);
- нестандартна поведінка клієнта після виплати кредиту;

Останнім пунктом підготовчого етапу є формалізація поняття дефолту («поганого» випадку) для банку, що працює зі скорингової системою.

Аналіз скорингових змінних — це [4] другий етап скорингової системи з використанням статистичної моделі. Основними джерелами для відalenня факторів аналізу є: заповнена анкета клієнта-позичальника, внутрішня кредитна історія банка та інформація, отримана в бюро кредитних історій на момент подачі анкети клієнта.

Отримані дані проходять етап перевірки на коректність. Так, наприклад, відсутність даних за певним показником можна тлумачити як помилку клієнта під час уведення або втрату даних системи. Якщо ж дані атрибути не є обов'язковими, то така ситуація є нормальнюю.

Кореляційний аналіз є важливим етапом оцінки скорингових характеристик. Усі змінні варто перевірити на наявність між ними кореляції. У разі виникнення проблеми мультикорельованості потрібно знайти [4] оптимальну комбінацію між видаленням статистичних характеристик та вибором однієї загальної змінної із кожного кореляційного кластера. Наступним підпунктом аналізу незалежності змінних є перевірка їх статистичної значущості. Суть аналізу полягає в перевірці наявності й сили зв'язку між залежною та незалежною змінними, що дозволяє визначити, які саме змінні є найважливішими для роботи даної моделі [4]. Методи, які найчастіше використовують на даному етапі:

- критерій χ^2 ;
- коефіцієнт Крамера V.

Основним етапом даної моделі є безпосередня побудова моделі. За основу статистичної моделі може бути взято такі моделі:

- багатофакторна лінійна регресія;
- багатофакторна логарифмічна регресія;
- штучні нейронні мережі;
- дерево прийняття рішень.

Багатофакторна лінійна регресія пов'язує поведінку залежної змінної (платоспроможність клієнта) із лінійною функцією низки незалежних змінних (скорингових характеристик). Шляхом мінімізації стандартних відхилень знаходять лінійну залежність, яку представляють у вигляді [5]:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im},$$

де y_i — залежна змінна (платоспроможність клієнта);

x_i — незалежні змінні (скорингові характеристики);

β_i — параметри моделі (скорингові ваги).

У результаті роботи одержують набір β_i , які визначають характер зв'язку між залежною змінною y_i та незалежними змінними x_i .

Багатофакторна логарифмічна регресія також пов'язує залежні та незалежні змінні, використовуючи скорингові ваги. Відмінність полягає в методі мінімізації відхилень. У логарифмічній регресії використовують принцип максимальної правдоподібності, а в лінійній — метод найменших квадратів. Формула логарифмічної регресії має вигляд [5]

$$\log \rho_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im},$$

де ρ_i — ймовірність позитивного випадку;

α — незалежний параметр логарифмічної регресії;

β_i — параметри моделі(скорингові ваги).

Дерево прийняття рішень — це модель, яку будують [5] на основі логічного ланцюга правил, що описує взаємозв'язки між змінними та очікуваним результатом. Структура дерева показує аргументацію правил, що дозволяє легко зрозуміти процес прийняття рішення. Графічно спрощену модель дерева прийняття рішень представлено на рисунку 2.1.



Рисунок 2.1 – Спрощена схема роботи дерева прийняття рішення

Вибір моделі базується на порівнянні статистичних показників якості моделі, серед яких виділяють [4]:

- статистику Колмогорова-Смірнова;
- коефіцієнт Джині;
- область під кривою ROC (Receiver Operating Curve).

Що більша кожна з характеристик, то якісніша збудована модель.

Методи роботи зі скоринговими системами можна порівняти таким чином:

- лінійна регресія краще працює з кількісним типом змінних;
- багатофакторна лінійна регресія будується на основі збалансованої вибірки (кількість «поганих» і «хороших» випадків рівна).
- для багатофакторної логарифмічної регресії не потрібно попередньо впорядковувати, змінювати тестову вибірку;
- на практиці лінійна та логарифмічна регресія дають схожі результати;
- вадою використання дерев являється не пристосованість системи до роботи з кількісними характеристиками, тільки з якісними;

2.2 Скорингова модель оцінки кредитоспроможності осіб на основі гібридних експертних систем

Проблема оцінки кредитоспроможності фізичної особи є неформалізованою задачею. Один із варіантів вирішення даної задачі є використання гібридних експертних систем.

Задачу оцінки кредитоспроможності представляють у вигляді [6]:

$$M = F(K, X),$$

де M — це комплексна оцінка об'єкта;

X — набір показників, які характеризують стан об'єкта;

K — набір критеріїв, за якими оцінюють показники та обчислюють M ;

F — деяка функція, за якою можна отримати загальну оцінку об'єкта.

Суть гібридної моделі полягає в декомпозиції задач на підзадачі та їх покрокове виконання. У скоринговій моделі виділяють 5 блоків [4]:

- a) соціальний стан особи;
- б) економічний стан;
- в) майнові права;
- г) параметри кредитної угоди;
- д) оцінка ділової репутації;

Кожний блок моделі характеризується набором показників, які визначають стан клієнта-позичальника з різних боків. Для кожного блока моделі використовується один із визначених методів розв'язку, а саме:

- a) аналітичний запис формули;
- б) штучна нейронна мережа;
- в) експертна система.

До блоку «соціальний стан», незалежно від типу скорингової системи, найчастіше включать наступні фактори аналізу [6]: стать, вік, сімейний стан, кількість членів сім'ї, освіта та наявність утриманців. У модулі «економічного стану» виділяють наступні атрибути [6]: сфера діяльності, рівень займаної особою посади, трудовий стаж, стаж останньої роботи, середній місячний дохід, організаційно-правова форма підприємства, термін проживання на місці реєстрації.

У блоках соціального та економічного стану в якості метода розв'язання використовують штучну нейронну мережу [6], оскільки ступінь впливу вхідних даних на кінцевий результат неможливо визначити однозначно. Проводять етап навчання мережі на тестових даних для визначення вагових коефіцієнтів для кожного атрибута блока. Фактори, які

отримують більший ваговий коефіцієнт, є важливішими для загальної оцінки позичальника від атрибутів із меншим коефіцієнтом.

У блоці «майнові права» варто відзначити наступні показники [6]: цінні папери, закордонний паспорт, наявність авто, нерухомість. Блок «оцінка ділової репутації» включає в себе наступні пункти [6]: кредитна історія, наявність фактів невиконання обов'язків перед іншими кредиторами, чи береться кредит для погашення іншого кредиту, фіктивність наданих клієнтом даних, наявність боргів, існування кредиту в іншому банку, схильність до алкоголізму та наркотичної залежності. Для цих модулів основним методом розв'язання є експертний підхід [6]: експерти складають набір правил, які визначають механізм відбору «хороших» та «поганих» клієнтів-позичальників.

У блоці «параметри кредитної угоди» виділяють такі найсуттєвіші фактори впливу [6]: вид кредиту, сума кредиту, відсоткова ставка, термін виплати кредиту, наявність поруки від фізичних осіб, застава. Методом розв'язання для даного блоку є формула, структура котрої залежить від особливостей скорингової системи. Блок служить для комплексної оцінки кредитоспроможності особи шляхом визначення його платоспроможності.

Згідно з даною гібридною схемою, загальну оцінку кредитоспроможності особи визначають за формулою [6]

$$Z = 0,15X_1 + 0,3X_2 + 0,25X_3 + 0,3X_4,$$

де Z — оцінка кредитоспроможності;

X_1 — соціальний стан;

X_2 — економічний стан;

X_3 — майнові права;

X_4 — оцінка ділової репутації;

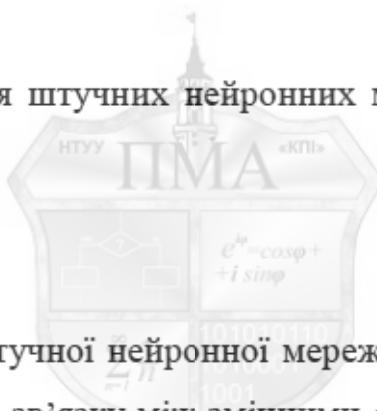
0,15, 0,3, 0,25, 0,3 — вагові коефіцієнти відповідних факторів ризику, які визначають кредитоспроможність клієнта-позичальника.

Запропонований механізм оцінки кредитоспроможності особи реалізовано в аналітичній інформаційній системі «Бізнес-Аналітик» [6].

Вадами наведеної гібридної моделі є:

- а) потреба у великому обсязі даних для навчання системи;
- б) не враховує макроекономічні показники під час аналізу кредитоспроможності позичальника;
- в) рішення, прийняті скоринговою системою раніше, впливають на рішення, які ця або інша система приймуть у майбутньому.

2.3 Використання штучних нейронних мереж у скорингових системах оцінки ризиків



Використання штучної нейронної мережі уможливлює [7] існування складного, нелінійного зв’язку між змінними скорингової моделі. Нейронна мережа являє собою систему взаємозв’язаних та взаємодіючих нейронів (вузлів). Двома основними характеристиками нейрона є функція активації та порогове значення. Перша визначає вихідне значення для нейрона, а поріг визначає мінімальне значення, за якого нейрон обчислює функцію активації. Графічно структуру штучного нейрона представлено на рисунку 2.2.

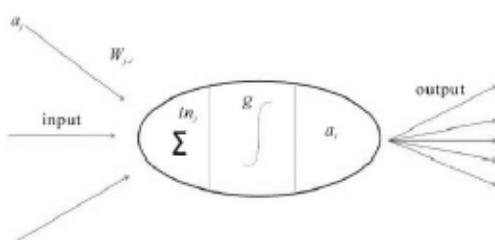


Рисунок 2.2 – Структура штучного нейрона

Архітектура нейронної мережі визначає кількість прихованих шарів мережі та кількість нейронів у кожному шарі.

У роботі [7] розглянуто роботу та архітектуру скорингової системи на основі нейронної мережі. Розроблені моделі оцінювали кредитні ризики для 507 італійських компаній.

Система передбачає використання 3 трьох прихованых шарів у нейронній мережі: 8 нейронів у першому шарі та по 4 — другому та третьому. У вигляді активуючої функції було обрано логістичну функцію

$$OUT = \frac{1}{1+e^{\alpha Y}},$$

де OUT — вихідне значення нейрона;

Y — вихідне значення для активуючої функції;

α — параметр нахилу.

У результаті застосування такої мережі було отримано результати [7], що демонструють відносно високу класифікаційну збіжність для визначення безпечних станів, і менш задовільні результати класифікації підприємств із високим ризиком.

Основною вадою нейронних мереж є відносно довгий підготовчий етап роботи, під час якого система навчається. До переваг належать гнучкість та універсальність під час роботи з будь-яким типом вхідних даних.

2.4 Нечітка логіка в роботі скорингових систем

Як було зазначено, робота зі скоринговими системами полягає в оцінці фінансового, динамічного чи соціального показника та присвоєння йому певної оцінки. У випадку використання чіткої логіки тому чи іншому

значенню присвоюють деяку однозначну величину. Наприклад, для атрибута «Вік» 35 років визначаються як «середній вік», а 32 відносять до «молодого віку». При цьому за виконання відповідної класифікації відповідає статистична база даних, або експерт.

Для уникнення однозначності використовують математичний апарат нечіткої логіки. Оцінка атрибута в такому разі залежатиме від функції належності, обраної для відповідної нечіткої множини. При цьому оцінка вибраного значення буде лежати в проміжку [0;1]. Наприклад, 35 років буде визначено зі ступенем 0,75 як «середній вік», і зі ступенем 0,25 – як «молодий вік». Нечіткість додає системі гнучкості, звільняючи її від суперечливих моментів вибору чітких меж для оцінки тих чи інших параметрів.

У скоринговій системі оцінці ризиків на основі нечіткої логіки [8] визначають 3 блоки характеризуючи атрибути:

- демографічний (вік, освіта, матеріальний статус, кількість дітей);
- фінансовий (місячний заробіток, стаж роботи, сфера роботи);
- фінансова безпека (wartість автомобіля, власного помешкання, решти нерухомості).

Наступний етап — визначення функцій належності для відповідних значень атрибутів. Найчастіше використовують стандартні функції: трикутну, трапецієву або гаусіану.

Останній етап для створення моделі — це визначення правил, за якими вона працюватиме. На практиці [8] таку систему складає експерт.

За результатами застосування система повинна дати відповідь: видавати особі кредит чи ні.

Вадою такого підходу є використання експерта, знання якого можуть бути неповними, а також неуніверсальність програми, яка під час зміни вибірки даних працюватиме за незмінним набором правил.

2.5 Висновки до розділу

За результатами огляду існуючих рішень за темою дипломної роботи можна зробити висновок, що використання нечіткого підходу для побудови скорингової моделі має низку переваг над чіткими:

- гнучка оцінка значень атрибутів, при якій одне й те саме значення можна віднести до різних блоків класифікації;
- належність до того чи іншого блоку оцінюється функціонально і з певним ступенем [0;1], що надає роботі системи наочності;
- зміна границь для оцінки значень атрибута потребує зміни тільки параметрів відповідної функції належності, а не переписування бізнес-логіки системи;
- неможливість чіткого описання складних залежностей типу значення атрибута — значення лінгвістичної змінної.

Серед існуючих скорингових моделей можна виділити системи, які використовують нечіткий підхід, проте в них базу правил для виконання класифікації складає експерт. Такий підхід має низку вад:

- залежність від наявності експерта для побудови бази правил;
- суб’єктивність системи, побудованої за допомогою експерта;
- відсутність динамічної зміни правил зі зміною вибірки без втручання експерта;
- відсутність автоматизованості роботи системи.

Для виправлення зазначених вад у роботі пропонується використання автоматизованого підходу для побудови бази правил нечіткої скорингової моделі на основі роботи еволюційних алгоритмів.

3 МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ

3.1 Системи нечіткого виведення

3.1.1 Поняття нечіткої множини

Нечітка множина [9] — це множина, елементи якої належать їй із певним ступенем. Позначимо через Ω універсальну множину об'єктів. Нечітка множина A в Ω характеризується своєю функцією належності $\mu_A(x)$, котра набуває значень у проміжку $[0;1]$ і показує ступінь належності елемента $x \in \Omega$ нечіткій множині A . Формально — це множина впорядкованих пар:

$$A = \{(x, \mu_A(x)); x \in \Omega, \mu_A(x) \in [0;1]\}.$$

Часто зручно записувати визначення нечіткої множини у списковому форматі. Так, наприклад, $A = \sum_{x_i \in \Omega} \frac{\mu_A(x_i)}{x_i}$ — для дискретного випадку, та $\int \frac{\mu_A(x_i)}{x_i}$ — для неперервного, де позначення суми та інтеграла виконують суто символічне значення.

Основна вимога, яку має задовольняти функція належності нечіткої множини — її значення мають належати проміжку $[0;1]$. Сама функція може являти собою будь-яку криву, форма якої визначається відповідно до поставленої проблеми з точки зору простоти, зручності та ефективності.

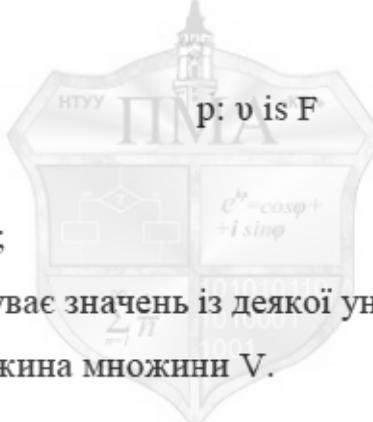
3.1.2 Поняття лінгвістичної змінної та нечіткого висловлювання

Лінгвістична змінна — це змінна, яка набуває значень у вигляді нечітких множин. Так, змінна «швидкість» може набувати значень «висока»,

«середня», «низька». Формально, лінгвістична змінна [10] — це п'ятірка $(L, T(L), \Omega, G, M)$, де L — назва змінної; $T(L)$ — терм-множина значень лінгвістичної змінної; Ω — універсальна множина; G — синтаксичне правило для визначення нових значень змінної L ; M — семантичне правило, яке визначає вигляд функції належності для кожного значення, створеного за допомогою правила G .

Фундаментальна відмінність між класичним поняттям висловлювання та нечітким висловлюванням полягає в діапазоні значень, які можуть набувати значення істинності висловлювання. У класичному випадку висловлювання може бути істинним або хибним, у нечіткому випадку істинність висловлювання оцінюється з певним ступенем [11]. Розглянемо некваліфіковані припущення, канонічна форма яких записується як

де p — висловлювання;



v — змінна, що набуває значень із деякої універсальної множини V ;

F — нечітка підмножина множини V .

Інший клас нечітких висловлювань — це умовні висловлювання, які інколи також називають нечіткими правилами. Кожне нечітке правило можна представити в канонічній формі:

$$p : \text{Якщо } x \in A \text{ тоді } Y \in B,$$

де x та Y — це змінні, які набувають значень у множинах X та Y відповідно. Правило складається з двох частин: антецедента — частини правила, яка задає умову, та консеквента — частини правила, яка показує наслідок.

3.1.3 Загальний опис системи нечіткого виведення

Задачу визначення ступеня кредитоспроможності особи можна розглядати як задачу нечіткої класифікації, де кожну особу можна віднести до класу кредитоспроможних або некредитоспроможних осіб. Система нечіткого виведення (СНВ) для такої задачі складається [12] з декількох вхідних лінгвістичних змінних $X_i, i = \overline{1, n}$, однієї вихідної змінної С — номера класу, до якого потрібно віднести особу (1 — для кредитоспроможних осіб, 0 — для некредитоспроможних), модуля нечіткого виведення та бази нечітких правил S . До вхідних лінгвістичних змінних належать характеристики особи про соціальний, фінансовий, демографічний стан особи. Кожна вхідна змінна відповідає певному атрибуту клієнта. Кожне правило $R_j \in S, j = \overline{1, s}$ в СНВ можна представити у вигляді

Якщо $x_1 = A_{j1}, \dots, x_n = A_{jn}$, то C_j із $CF = CF_j$,

де A_{j1}, \dots, A_{jn} — значення вхідних змінних (нечіткі множини, визначені своїми функціями приналежності $\mu_{j1}, \dots, \mu_{jn}$);

C_j — номер класу;

CF_j — ступінь достовірності правила R_j .

У такій СНВ базу нечітких формул S можна побудувати шляхом застосування еволюційного алгоритму, для чого потрібно сформувати навчальну вибірку з m n -вимірних шаблонів $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$, $p = \overline{1, m}$, де $x_{pi}, i = \overline{1, n}$ — значення атрибута p -го клієнта. Щодо кожного шаблону відомо, якому класу він належить. Тоді клас та ступінь достовірності кожного нечіткого правила можна детерміновано визначити згідно з наступною процедурою [12]:

- а) обчислити ступінь сумісності кожного навчального шаблону x_p з правилом R_j :

$$\mu_j(x_p) = \mu_{j1}(x_{p1}) \times \dots \times \mu_{jn}(x_{pn});$$

- б) для кожного класу обчислити суму ступенів сумісності навчальних шаблонів із R_j :

$$\beta_{c_j}(R_j) = \sum_{x_p \in C_j} \mu_j(x_p), \text{ де } C_j = \overline{0,1};$$

- в) знайти клас k , для якого досягнуто максимум $\beta_{c_j}(R_j)$:

$$\beta_{c_h}(R_j) = \max\{\beta_{c_0}(R_j), \beta_{c_1}(R_j)\}.$$

Якщо для обох класів досягнуто максимум $\beta_{c_j}(R_j)$, то визначити клас C_j неможливо, покладають $C_j = \emptyset$ і $CF_j = 0$. Якщо не виявлено шаблонів, які сумісні з правилом R_j ($\beta_{c_h}(R_j)$, для $h = 0,1$), то $C_j = \emptyset$;

- г) якщо можливо однозначно визначити максимальне значення $\beta_{c_j}(R_j)$ та клас k , який йому відповідає, то обчислити CF_j за формулою:

$$CF_j = \frac{\beta_k(R_j) - \beta_{1-k}(R_j)}{\beta_0(R_j) + \beta_1(R_j)}.$$

Якщо $\beta_0(R_j) = \beta_1(R_j)$, то $C_j = \emptyset$ та $CF_j = 0$ тому, що ми однозначно не можемо визначити клас C_j .

Модуль нечіткого виведення визначає клас шаблону x_p згідно з наступним підходом [12]. Для кожного шаблона модуль визначає єдине правило- переможець R_j за формулою:

$$\mu_p \cdot CF_j = \max_{R_j \in S} (\mu_j(x_p) \cdot CF_j), \quad (1)$$

тобто правило, котре має максимальний добуток ступеня сумісності для шаблону x_p $\mu_j(x_p)$ та ступеня достовірності CF_j . Якщо одночасно декілька правил є переможцями для одного шаблону або $\mu_j(x_p) = 0 \forall R_j \in S$, то вважають, що шаблон x_p не належить жодному класу.

3.1.4 Етапи побудови системи нечіткого виведення

У загальному випадку можна виділити такі етапи розроблення будь-якої системи нечіткого виведення [11]:

- ідентифікувати релевантні входні та вихідні лінгвістичні змінні системи та діапазон їх значень;
- вибрати найважливіші нечіткі значення дляожної змінної;
- сформулювати систему правил, які представляють знання системи необхідних для вирішення поставленої задачі;
- вибрати відповідний метод конвертації нечіткої змінної у чітке представлення.

У даній роботі входні змінні відповідають характеристикам позичальника, вихідна змінна відповідає класу позичальника (кредитоспроможний або некредитоспроможний).

Функції належності формують шляхом рівномірного розбиття області визначенняожної змінної [12]. Даний підхід не погіршує класифікаційну здатність системи. СНВ не включає механізму налагодження параметрів функцій належності, оскільки формування складних меж між класами (не паралельних координатним вісям) здійснюється за рахунок визначення оцінки достовірності CF_j кожного нечіткого правила R_j .

Формування системи правил у пропонованій системі здійснюється автоматизовано, із застосуванням меметичного алгоритму, який буде описано в написаному розділі.

3.2 Опис меметичного алгоритму побудови бази нечітких правил для СНВ

У літературі виділяють два методи побудови бази нечітких правил систем нечіткого виведення (СНВ) за допомогою еволюційних обчислень. За піттсбурзького підходу (Pittsburgh approach), кожний індивід у популяції в алгоритмі являє собою повну базу правил [13]. За мічиганського підходу (Michigan approach) кожний індивід є окремим правилом, а популяція являє собою повну базу правил [12]. У даній роботі пропонується меметичний алгоритм, який працює за мічиганським підходом, оскільки піттсбурзький підхід обчислювально значно більш трудомісткий [13, с. 616].

3.2.1 Загальна схема алгоритму

Меметичний алгоритм побудови бази правил для СНВ, пропонований в даній роботі, є узагальненням алгоритму з [12] та передбачає виконання наступних кроків:

- а) згенерувати початкову популяцію (базу нечітких правил) P з μ правил, вибираючи елементи правил випадковим чином за рівномірним законом;
- б) оцінити кожне правило з поточної популяції, обчислюючи значення функції пристосованості $f(r)$ $\forall r \in P$;

- в) вибрати λ пар батьківських індивідів;
- г) застосовувати до кожної батьківської пари оператор схрещування з ймовірністю p_c та помістити результат у множину P' ;
- д) застосовувати до кожного нашадка оператор мутації з імовірністю p_m та помістити результат у множину P'' ;
- е) застосовувати до множини P'' оператор локального пошуку з ймовірністю p_{mem} ;
- ж) замінити k найменш пристосованих індивідів у P на індивідів із P'' ;
- з) перевірити виконання умови завершення: якщо вона виконується, зупинити алгоритм, або повернутися до кроку в в протилежному випадку.



3.2.2 Представлення індивідів у популяції

Першим кроком у визначенні будь-якого еволюційного алгоритму [14] є визначення відображення початкової задачі на задачу, яка розглядається в рамках еволюційному алгоритму. Об'єкти, які складають розв'язок поставленої початкової задачі, називають фенотипом, а їх представлення в еволюційному алгоритмі — генотипом. Відповідне відображення фенотипу на генотип називають представленням.

Вибір представлення даних є найскладнішою частиною та відіграє одну з найважливіших ролей під час проектування будь-якого еволюційного алгоритму, у тому числі меметичного.

Кожний індивід (хромосома) у популяції відповідає деякому правилу та являє собою вектор $r = (r_1, \dots, r_n)$, де r_i — номер функції належності, що відповідає i -тій лінгвістичній змінній. Якщо $r_i = 0$, то в якості функції

належності для i -тої лінгвістичної змінної береться функція, тотожна одиниці. Іншими словами, значення відповідної змінної не впливають на виведення з відповідного правила.

Множина можливих значень, яких може набувати r_i , складається з усього набору значень лінгвістичних змінних, визначених на даній нечіткій множині, та додатково — відсутності значення як такого (число 0). Імплементація відсутності значення є корисним, оскільки на кінцевий результат деяка змінна може не впливати, або її вплив незначний.

Розгляньмо для прикладу хромосому $r = (1,0,3)$, де 1 — кодування значення «молодий» лінгвістичної змінної, що описує вік людини, 0 — кодування відсутності значення для лінгвістичної змінної, що описує дохід людини, 3 — кодування значення «вища» лінгвістичної змінної, що описує рівень освіти людини. Виведення з такого правила враховуватиме тільки значення віку та рівня освіти клієнта-позичальника.

Варто зазначити, що в пропонованому меметичному алгоритмі еволюційний процес має місце тільки на рівні нечітких правил, і не впливає на параметри відповідних функцій належності. Такий підхід обрано, виходячи з потреби збереження простоти інтерпретовності отримуваних правил із боку користувачів системи.

3.2.3 Функція пристосованості

Функція пристосованості показує [14] ступінь якості кожного індивіда як розв'язку поставленої задачі. Для того, щоб оцінити пристосованість кожного нечіткого правила в базі правил S , спочатку потрібно класифікувати кожний шаблон із навчальної вибірки згідно з процедурою, описаною в розділі 2.1.3. Тоді, функція пристосованості $f(r)$ визначається як:

$$f(r) = NCP(r),$$

де $NCP(r)$ — кількість тестових шаблонів, яких правильно класифіковано за допомогою правила r .

Якщо більше за одне нечітке правило має один і той самий консеквентний клас C_i з однаковим максимальним значенням $\mu_p \cdot CF_j$ отриманим у (1), то правило-переможець визначають як правило з найменшим індексом. Таким чином, у наступному поколінні з правил, які дублюють інформацію (мають одинаковий антецедент і консеквент правила) залишається лише правило з мінімальним індексом.

3.2.4 Відбір індивідів для виконання операцій варіації

Перш ніж перейти до виконання операцій варіації (схрещування та мутації) для генерування правил з бази правил S , потрібно відібрати λ пар так званих батьківських правил. Відбір правил відбувається за допомогою метода рулетки:

- обчислити масив імовірностей $probability$ для метода рулетки;
- сформувати рулетку;
- згенерувати псевдовипадкове число r ;
- знайти індекс i у масиві $probability$, такий, що $r \leq probability_i$;

Обчислення масиву $probability$ відбувається за наступною схемою:

- a) оцінити пристосованість правила з використанням функції пристосованості $f(r)$;
- б) якщо загальна пристосованість для всієї бази правил S дорівнює 0, то в), інакше г);

- в) задати $probability_i = 1/length(probability)$;
- г) задати $probability_i = (f(r)_i - \min\{f(r)_i\})/d$, де $d = \sum(f(r)_i - \min\{f(r)_i\})$.

Формування рулетки — це етап, під час якого створюють масив, який складається з інтервалів $[0; k_j]$, де $k_j \leq 1$, та визначається як $k_j = \sum_{i=1}^j probability_i$. Оскільки початок інтервалу завжди є 0, то доцільно зберігати лише кінець інтервалу.

Алгоритм рулетки — це один із найпростіших підходів для відбору індивідів із популяції. У цьому методі всі хромосоми в популяції розміщуються на рулетці. Кожна хромосома являє собою елемент рулетки. Розмір кожного сегмента є пропорційним до значення функції пристосованості $f(r)$ кожного індивіда [15, 16, 17]. Тому, що більше $f(r)$ для відібраного правила, то більший розмір сегмента на рулетці, який йому відповідає. Правила, котрі мають більшу $f(r)$, а відповідно й більший сегмент на рулетці мають більшу ймовірність для потрапляння у множину батьків [18].

3.2.5 Оператори варіації в алгоритмі

До операторів варіації, які використовують у будь-якому еволюційному алгоритмі, належати оператор схрещування та оператор мутації.

Оператор схрещування (кросовер) — це бінарний оператор, який [14] об'єднує інформацію з генотипу двох батьків в один або два генотипи нашадків. Схрещування — це одна з найважливіших особливостей в еволюційних алгоритмах, головною задачею якої є створення неоднорідності в популяції.

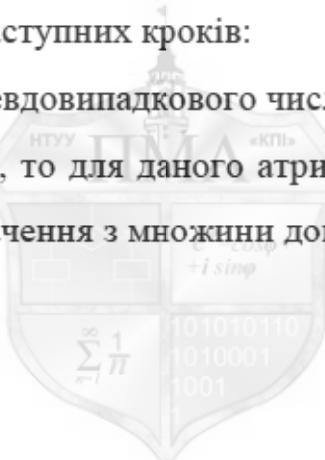
Оператор схрещування застосовують до двох правил R_i та R_j з високою ймовірністю p_c . Суть схрещування полягає в обміні значень відповід-

них атрибутів між двома правилами. У даній роботі використано рівномірне схрещування [19], яке працює наступним чином:

- випадковим чином згенерувати вектор-рядок l довжиною, рівною довжині правила, який складається з бінарних значень (0 або 1);
- правила R_i та R_j обмінюються значеннями атрибутів у тих місцях, де значення $l_k = 1$, а у випадку $l_k = 0$, атрибути правил залишаються незмінними.

В еволюційних алгоритмах оператор мутації — це унарний оператор варіації, який повертає модифікованого індивіда. Оператор мутації застосовують до правила R_i з низькою ймовірністю p_m . У даній роботі використано оператор випадкової заміни [14], який для кожного атрибута правила передбачає виконання наступних кроків:

- генерація псевдовипадкового числа r ;
- якщо $r \geq p_m$, то для даного атрибута правила випадковим чином визначаємо інше значення з множини допустимих для даного атрибута (лінгвістичної змінної).



3.2.6 Меметична частина алгоритму

Меметичний алгоритм [20] (від слова «мем», уведеного Річардом Докінзом у [21]) — різновид еволюційного алгоритму, особливістю якого є використання всіх допустимих знань, пов'язаних із поставленою задачею. Використання набору знань в меметичних алгоритмах відбувається шляхом включення алгоритмів апроксимації, методів локального пошуку, використання рекомбінованих операторів тощо.

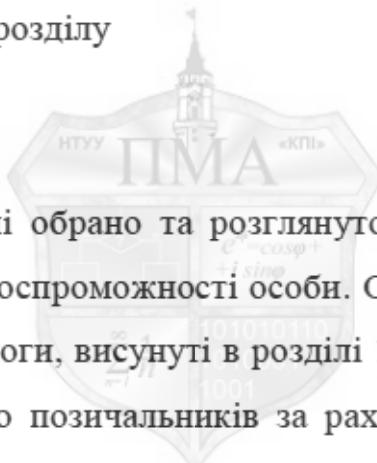
У даній роботі пропонується реалізувати процедуру локального пошуку в розроблюваному меметичному алгоритмі наступним чином [22].

Після виконання схрещування та мутації індивідів (правил) та одержання множини P'' потрібно:

- вибрати випадковим чином індекс q правила з P'' та індекс елемента t ньому;
- шляхом перебору всіх можливих значень елемента t вибрати значення, котре дає найбільшу пристосованість правила q .

Таким чином отримується поліпшена множина правил, у якій сумарне значення функції пристосованості більше, аніж у випадку відсутності процедури локального пошуку.

3.3 Висновки до розділу



У даному розділі обрано та розглянуто математичне забезпечення для визначення кредитоспроможності особи. Обране математичне забезпечення задоволяє вимоги, висунуті в розділі 1, зокрема врахування нечіткої природу даних про позичальників за рахунок використання системи нечіткого виведення та автоматизації процесу побудови нечітких правил за допомогою меметичного алгоритму.

Запропоновано модифікацію класичного еволюційного методу побудови бази нечітких правил нечіткого класифікатора у вигляді меметичного алгоритму шляхом додавання етапу локального пошуку найпристосованіших правил системи, що дозволяє зменшити пошуковий простір алгоритму.

4 ПРОГРАМНЕ ЗАБЕЗПЕЧНЯ

4.1 Архітектура автоматизованої системи

Розроблена автоматизована система складається з таких підсистем:

- підсистема конвертації даних про клієнта;
- підсистема нечіткого виведення;
- підсистема генерування правил;
- підсистема локального пошуку найбільш пристосованих правил;
- підсистема визначення характеристик бази правил;
- база даних системи визначення кредитоспроможності особи.

Вхідними даними для системи є дані, які характеризують клієнта з демографічного, соціального та фінансового боку. Дані вносяться шляхом заповнення запропонованих полів клієнтської частини інтерфейсу програми. Для системи було вибрано наступні атрибути клієнта:

- сімейний статус;
- вік;
- місячний дохід;
- отриманий рівень освіти;
- грошова одиниця, у якій отримується заробітна плата;
- сфера роботи клієнта;
- стаж роботи;
- наявність фінансової допомоги;
- наявність утриманців;
- кількість записів у медичній картці;
- кількість дітей;

- кількість повнолітніх дітей;
- відношення зарплати до загального обсягу майна;
- оцінка нерухомості клієнта.

На виході система видає оцінку кредитоспроможності клієнта, яка приймає два значення: клієнт кредитоспроможний, клієнт некредитоспроможний. Відповідно до вимог, висунутих у розділі 1, система забезпечує:

- уведення характеристик користувача за вказаними пунктами;
- навчання системи в режимі адміністратора;
- прогнозування кредитоспроможності клієнта;
- збереження результуючого прогнозу до БД.

Описані підсистеми є взаємозв'язаними, передаючи інформацію про дані роботи системи між собою. Схему взаємодії підсистем представлено на рисунку 4.1.

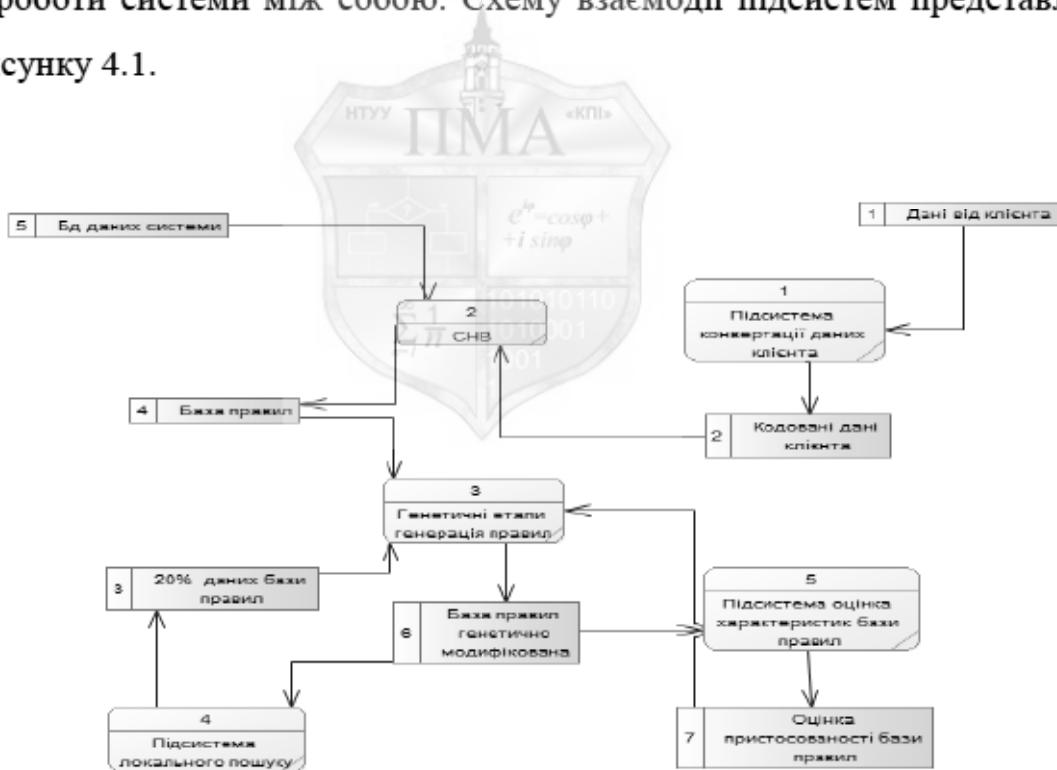


Рисунок 4.1 – Схема взаємодії підсистем автоматизованої системи визначення кредитоспроможності клієнта

4.1.1 Підсистема конвертації даних про клієнта

Підсистема конвертації даних про клієнта забезпечує трансформацію даних у формат, зручний для аналізу та виконання над ним різного роду арифметичних та логічних операцій. Зазначені значення атрибутів користувачів буде переведено у числове представлення. Для атрибутів, які приймають дискретний набір значень, дані трансформуються в ціличислове представлення Z , а для атрибутів із дійсним значенням змінюється лише діапазон значень ($[0;1]$).

Вхідними значеннями для даної підсистеми є введені значення атрибутів по клієнту. При цьому клієнт має можливість не вибирати значення для атрибуту. Невибране значення не буде вважатися за помилку. Вихідними значеннями підсистеми буде чисрова матриця, яка дає повну характеристику про клієнта (-ів).

Даний етап є підготовчим етапом у роботі системи. Основна його задача — сформувати тестову матрицю для навчання системи. Виділяють наступні кроки в роботі даної підсистеми:

- ініціалізація розміру вхідної та вихідної матриці;
- зчитування навчальних даних із БД або тестового файлу;
- конвертація даних про клієнта в масив строкових даних;
- переведення строкових даних масиву в числове представлення залежно від логіки представлення атрибутів;
- збереження вихідної матриці.

Блок-схему роботи даної підсистеми представлено на рисунку 4.2.

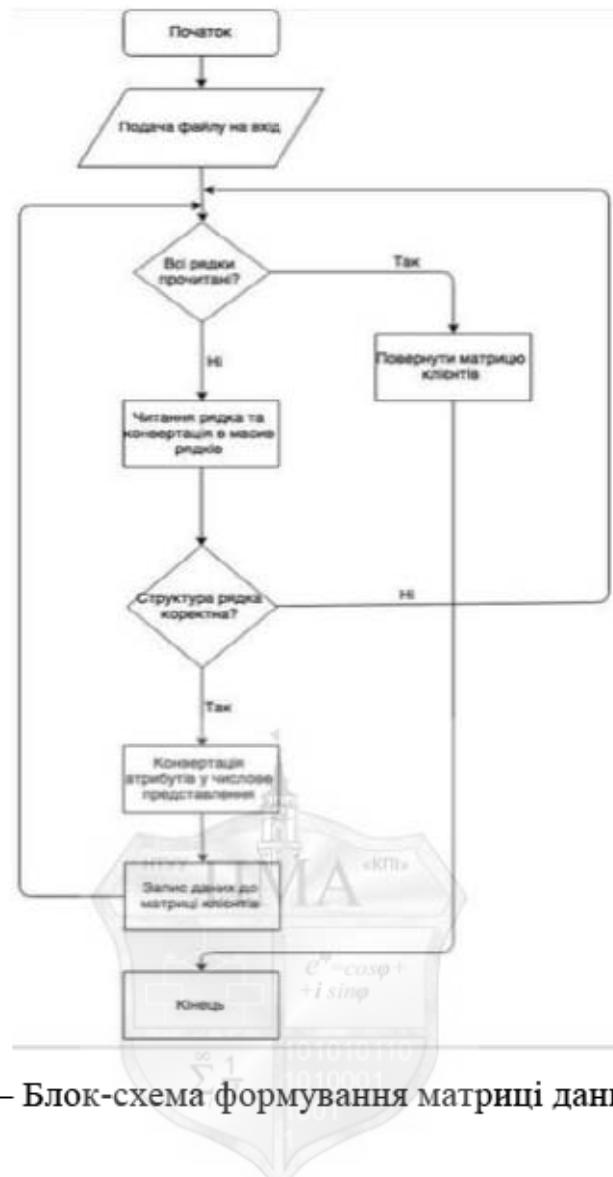


Рисунок 4.2 – Блок-схема формування матриці даних про клієнта

Після формування матриці даних про клієнта її потрібно нарізати для блоки для дальнішої роботи. Блок-схему відповідного алгоритму представлено на рисунку 4.3.

4.1.2 Підсистема нечіткого виведення

Дана підсистема забезпечує налаштування каркасу системи нечіткого виведення, за допомогою якої здійснюється визначення ступеня кредитоспроможності клієнта. До налаштувань системи належать:

- кількість атрибутів, які оцінюються в системі;
- тип та обмеження за кожним атрибутом;
- набір функцій належності для кожного атрибута на заданому інтервалі;
- тип та параметри для кожної функції належності.

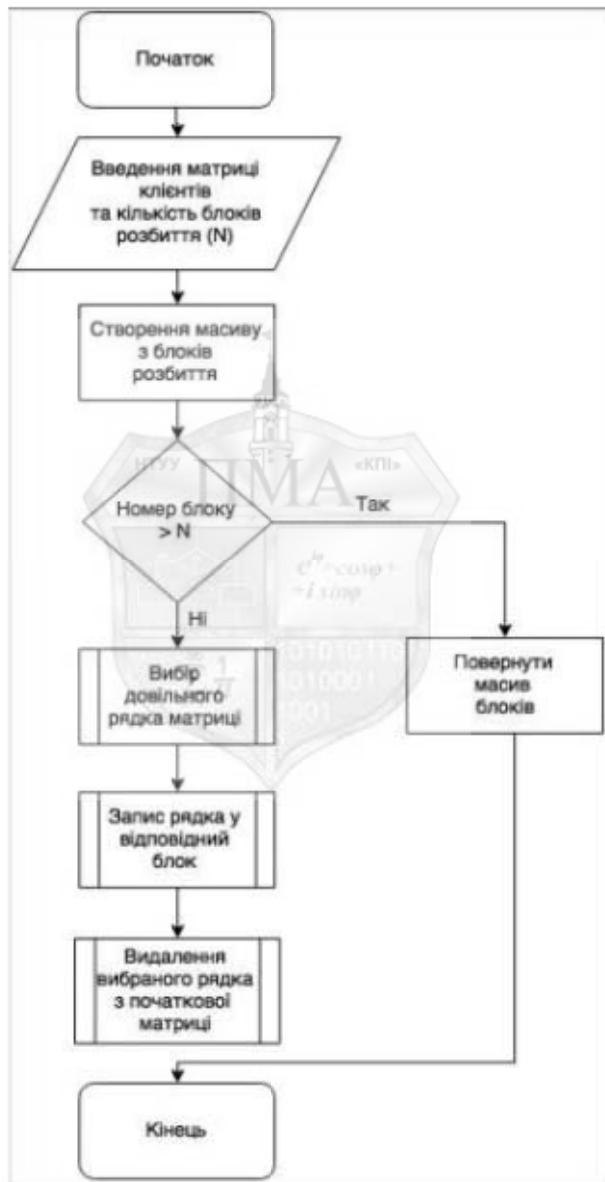


Рисунок 4.3 – Блок-схема алгоритму нарізання початкової матриці на блоки

Таким чином, отримується набір даних та їхніх характеристик, які визначатимуть роботу системи. Вхідні налаштування вносяться самостійно

розробником системи або можуть модифікуватися у процесі навчання для конкретної задачі. Вихідними даними для даної підсистеми виступають дані отримані за допомогою характеристик, вказаних у налаштуваннях.

Нечіткість даної системи забезпечує той факт, що на неявно визначених значеннях атрибутів може прийматися два рівноцінних значення атрибуту. Так, щодо значення атрибуту вік (23 роки) можна сказати, що на 0,75 вік даного клієнта є «молодим», а на 0,45 — «середнім». Нечіткість системи забезпечує гнучкий аналіз даних та високу збіжність результатів роботи практичних завдань.

Схематично дану систему зображенено на рисунку 4.4.

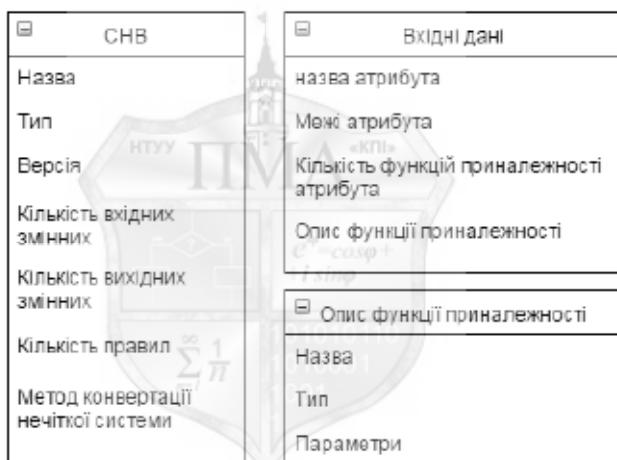


Рисунок 4.4 – Схематичне представлення системи нечіткого виведення

4.1.3 Підсистема генерування правил

Правила, за якими відбувається класифікація клієнта, — це основні дані підсистеми. Їх відсутність або некоректність нівелює роботу всієї системи. Оскільки генерування правил у системі виконується шляхом виконання генетичних операцій, то основу даної системи складає реалізація відповідного меметичного алгоритму.

Дана підсистема складається з наступних блоків:

- блок вибірки даних;
- блок рекомбінації батьківських даних;
- блок мутації нащадків.

Вхідними даними для підсистеми є повний набір правил, що являє собою двовимірну матрицю дійсних чисел. Данна матриця правил генерується на етапі ініціалізації роботи даної підсистеми. Дані під час ініціалізації генеруються випадковим чином, беручи межі довільного значення конкретного атрибута з налаштувань підсистеми СНВ, описаної вище.

Вихідними даними підсистеми виступає частина правил, модифікована в процесі використання еволюційних операцій. Таким чином, відбувається заміна найменш пристосованих правил на ті правила, які були згенеровано.

Блок-схему роботи підсистеми представлено на рисунку 4.5.



4.1.4 Підсистема локального пошуку найпристосованіших правил

Підсистема локального пошуку додає до набору евристичних алгоритмів точність пошуку правил, що дають можливість точнішу класифікацію.

Суть роботи підсистеми полягає в тому, щоб за вибраним довільним чином правилом та довільним атрибутом вона знаходить те значення атрибута, яке дає найбільший ступінь пристосованості. Іншими словами, вибране правило повинно правильним чином визначати якнайбільше число тестових шаблонів.

Вхідними даними для підсистеми є набір правил, згенерований підсистемою генерування правил. Таким чином, сумарне значення пристосо-

ваності зростає після виконання локального пошуку. На виході отримаємо оновлений набір правил, який потрапляє до загального списку правил.

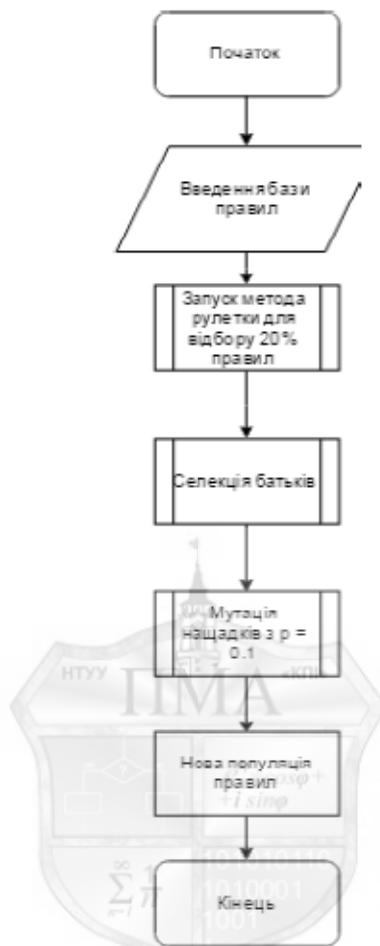


Рисунок 4.5 – Блок-схема роботи підсистеми генерування правил

Блок-схему роботи підсистеми локального пошуку найпристосованіших правил представлено на рисунку 4.6.

4.1.5 Підсистема визначення характеристик бази правил

Оцінка правила стосовного того, наскільки воно гарно підходить для характеристики клієнтів, виконує підсистема визначення характеристик бази правил. Основними її характеристиками є:

- пристосованість правила;
- сумарна пристосованість набору правил;
- оцінка достовірності правила;
- оцінка сумісності тестового шаблона для кожного правила;
- сумарна оцінка сумісності по кожному правилу.

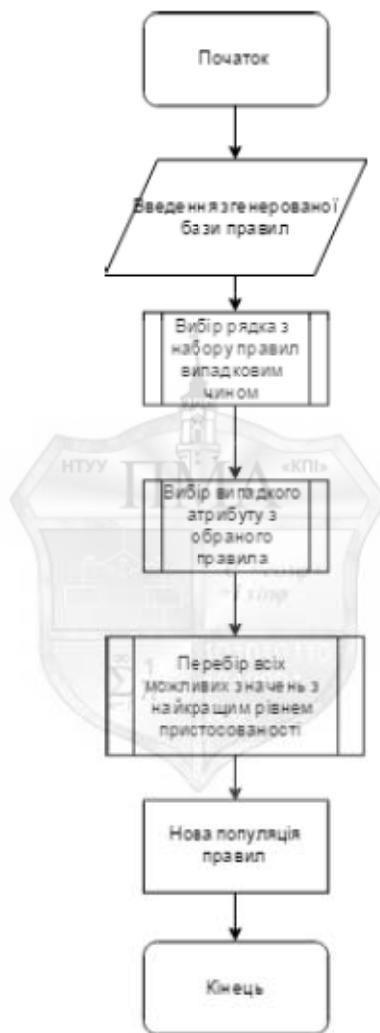


Рисунок 4.6 – Блок-схема роботи підсистеми локального пошуку

Вхідними даними для підсистеми є набір правил, який потрібно охарактеризувати.

Вихідними даними є оцінена характеристика правила, яка найчастіше представляється дійсним числом (для одного правила) або масивом дійсних чисел (для набору правил).

4.1.6 Опис бази даних системи

База даних системи визначення ступеня кредитоспроможності особи складається з наступних сутностей:

- клієнт;
- демографічні характеристики клієнта;
- фінансова характеристика клієнта;
- соціальна характеристика клієнта.

Сутність «клієнт» включає такі атрибути:

- унікальний ідентифікатор клієнта;
- ім'я клієнта;
- прізвище клієнта;
- реальний клас кредитоспроможності клієнта;
- клас клієнта, визначений системою.

Для сутності «клієнт» необов'язковими атрибутами є тільки ім'я та прізвище клієнта. Дані атрибути не є обов'язковими в силу того, що навчальна вибірка працює з закодованими даними, тому конкретні відомості про клієнта невідомі. Зв'язок сутностей «клієнт» та його характеристик здійснюється за полем «унікальний ідентифікатор», що явно характеризує унікальність користувача.

Відповідну ER-діаграму представлено на рисунку 4.7.

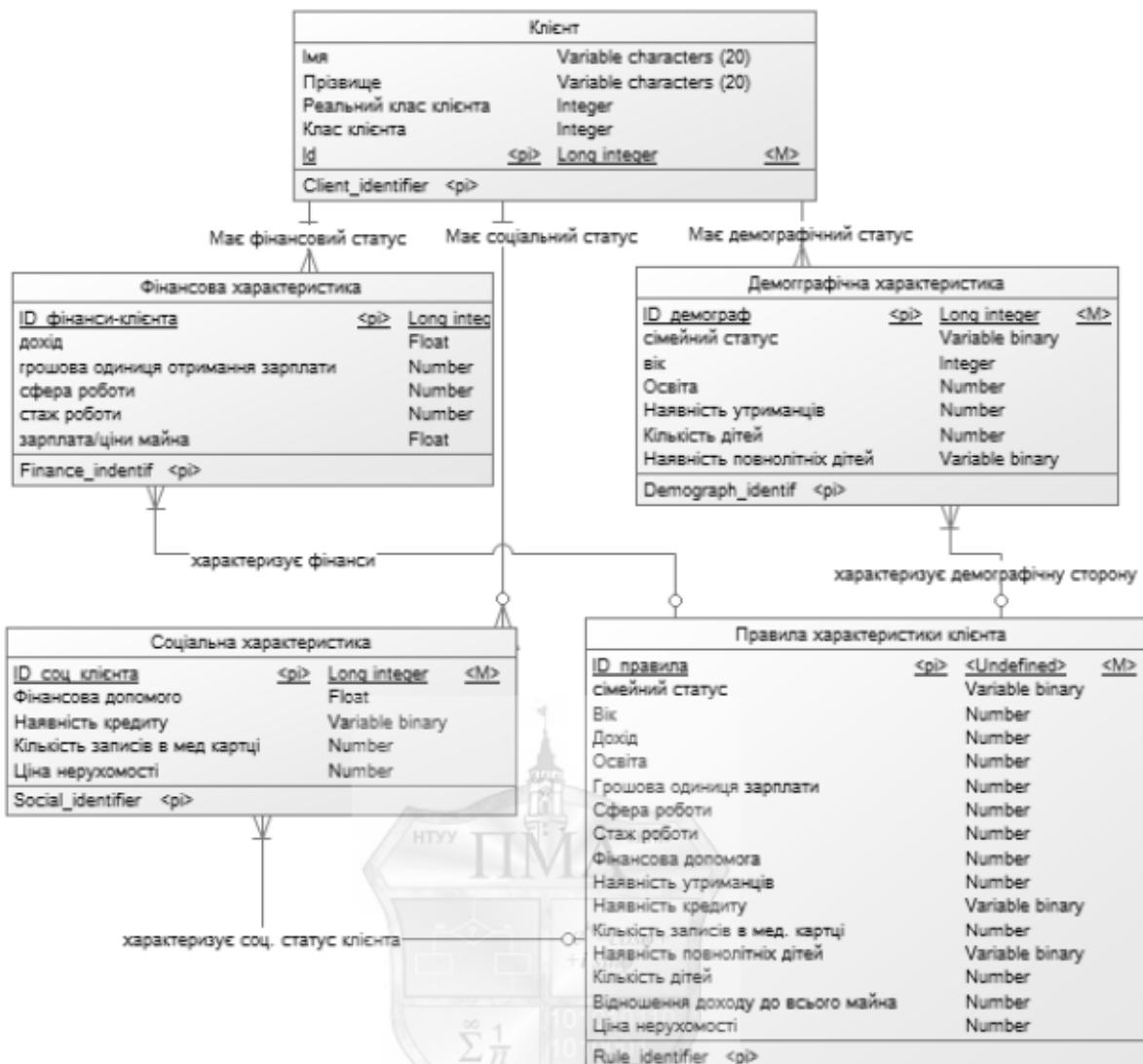


Рисунок 4.7 – ERD автоматизованої системи визначення
ступеня кредитоспроможності особи

4.2 Структура розроблених класів

Під час розробки програмного забезпечення було створено взаємозв'язок між мовою програмування C# та математичним пакетом MATLAB 2012b. Математичний функціонал програми реалізовано в MATLAB, а взаємодія з користувачем та GUI — за допомогою C#. Дані про клієнта та результати навчання системи (набір правил) зберігаються у

відведеній базі даних, реалізованої за допомогою мови T-SQL в середовищі SQL Server Management Studio.

Структуру класів представлено на рисунку 4.8.

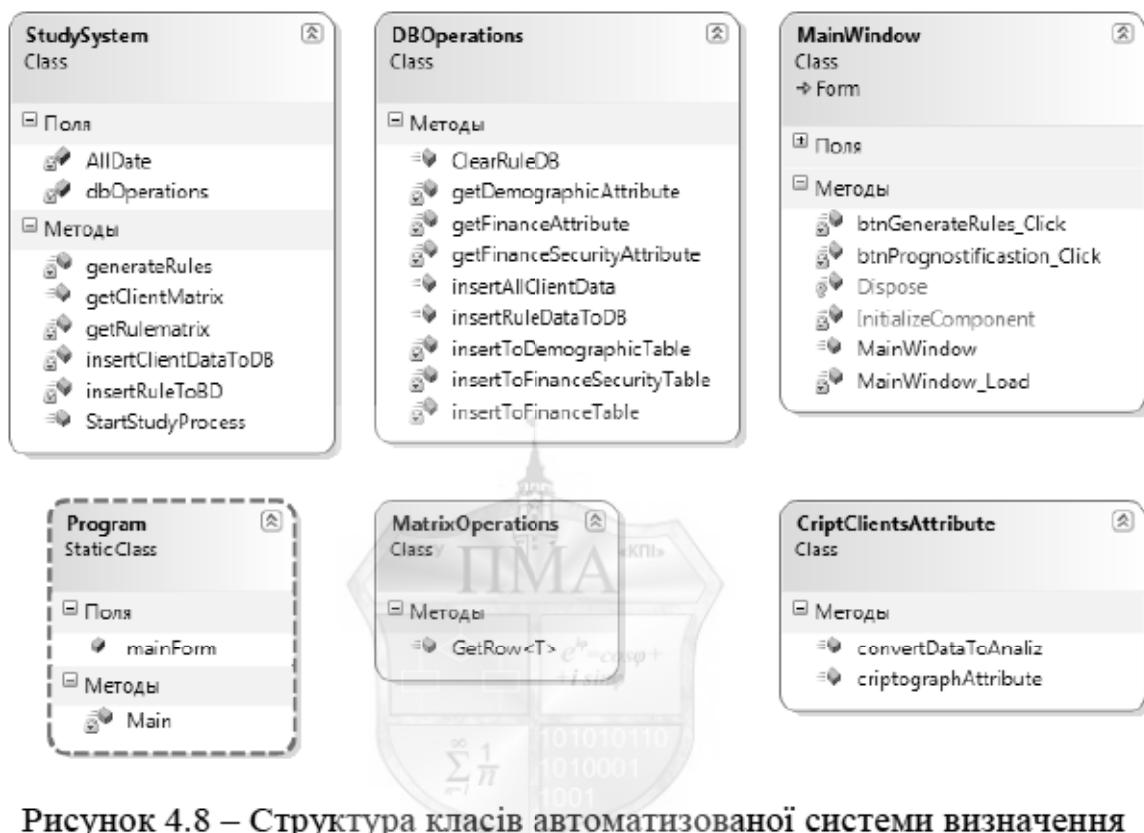


Рисунок 4.8 – Структура класів автоматизованої системи визначення ступеня кредитоспроможності особи

4.3 Випробування роботи системи

У рамках виконання дипломної роботи було перевірено працездатність та ефективність роботи розробленої автоматизованої системи. Зображення графічного інтерфейсу користувача програми представлено на рисунку 4.9.

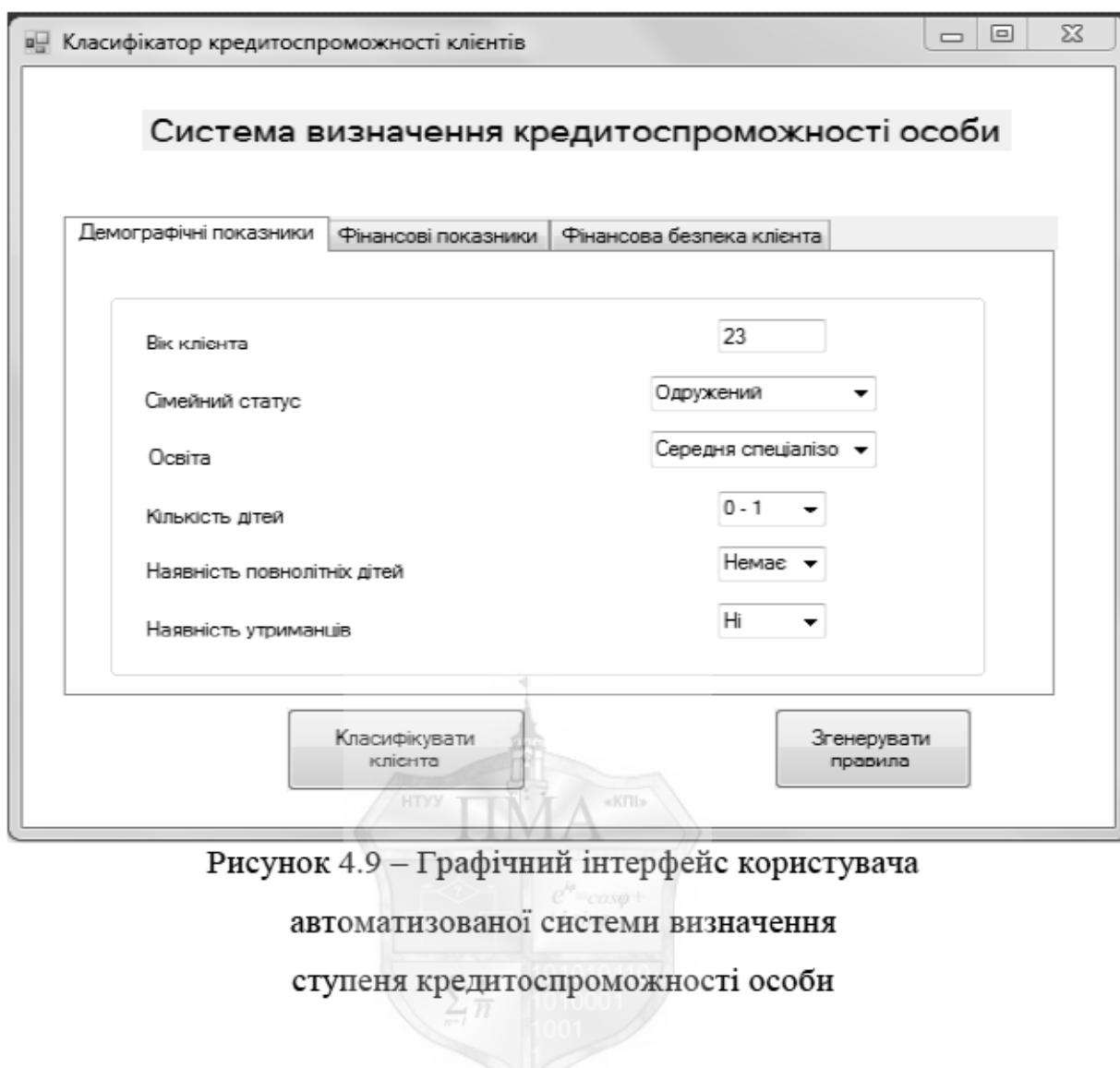


Рисунок 4.9 – Графічний інтерфейс користувача
автоматизованої системи визначення
ступеня кредитоспроможності особи

Роботу системи було випробувано на тестовій вибірці, взятої з UCI Machine Learning Repository, CA: University of California [23]. Характеристики даної вибірки наступні:

- обсяг навчальної вибірки — 620 клієнтів;
- обсяг тестової вибірки — 69 клієнтів;
- кількість атрибутів клієнта — 15;
- кількість категорій класифікації атрибутів — 3;
- кількість категорійних атрибутів — 9;
- кількість порядкових атрибутів — 6;

Систему було оцінено з погляду якості класифікації клієнтів стосовно їхньої кредитоспроможності. Результати роботи трьох алгоритмів класифікації наведено в таблиці 4.1.

Таблиця 4.1 – Порівняльна характеристика випробуваних алгоритмів

Джерело	Метод розв'язання	Середня частка правильно класифікованих клієнтів
[24]	Алгоритм C4.5	84,3%
[12]	Класичний еволюційний алгоритм	86,7%
Запропонована система	Меметичний алгоритм	90,1%

4.4 Висновки до розділу

Спроектовано автоматизовану систему визначення ступеня кредитоспроможності особи, розроблено алгоритми роботи кожної підсистеми. Описано роботу кожної підсистеми та представлено блок-схеми відповідних алгоритмів. Надано структуру самої системи та БД, яка зберігає потрібні дані. Розроблена система надає графічний інтерфейс користувача та задоволяє всі вимоги, поставлені в розділі 1.

За результатами випробування системи на контрольному прикладі на основі реальних даних перевірено роботу програми та відсоток правильно класифікованих користувачів. Середня частка правильно класифікованих клієнтів за 20 запусків еволюційного алгоритму складає 86,7%. Використання меметичного алгоритму, описаного в 3.2, дає можливість одержати середню частку правильно класифікованих шаблонів, рівною 90,1%.

ВИСНОВКИ

У дипломній роботі проведено аналіз використовуваних на практиці скорингових моделей та вибрано математичний апарат нечіткої логіки як такий, що найточніше представляє дані про потенційних позичальників. Для побудови бази нечітких правил вибрано метод на основі еволюційних обчислень як такий, що не вимагає залучення експерта.

На основі вибраних методів спроектовано та реалізовано автоматизовану систему визначення ступеня кредитоспроможності особи. Випробування розробленої системи на контрольному прикладі на основі реальних даних показало, що середня частка правильно класифікованих клієнтів за 20 запусків еволюційного алгоритму складає 86,7%. Використання меметичного алгоритму, описаного в 3.2, дає можливість одержати середню частку правильно класифікованих шаблонів, рівною 90,1%.

Основні ідеї, викладені в роботі, опубліковано в тезах доповіді «Меметичний метод побудови системи нечіткого виведення для визначення ступеня кредитоспроможності особи» на 17-їй Міжнародній науково-технічній конференції SAIT 2015.

ПЕРЕЛІК ПОСИЛАНЬ

1. Скорба О. А. Оцінка кредитоспроможності позичальника / О. А. Скорба // Проблеми і перспективи розвитку банківської системи України: Збірник наукових праць. — 2005. — 6. — С. 162–168.
2. Положення про порядок формування та використання банками України резервів для відшкодування можливих втрат за активними банківськими операціями : затв. Постановою Національного банку України від 25.01.2012 №23.
3. Верба А. В. Скоринг, як інструмент управління кредитним ризиком банку / А. В. Верба [Електронний ресурс]. — 2013. — Режим доступу: www.bulletin.uabs.edu.ua/store/eco/2013/b3f3e2db4abbb5d6c09b81bb8caeae5.pdf
4. Ковалев М. Методика построения банковской скоринговой модели для оценки кредитоспособности физических лиц / М. Ковалев, В. Корженевская // Банки Казахстана. — 2008. — 1. — С. 43–48.
5. Nurlybaeva K. Algorithmic Scoring Models / K. Nurlybaeva, G. Bakayeva // Applied Mathematical Sciences. — 2013. — С. 571–586.
6. Пятковський О. И. Скоринговая система оценки кредитоспособности физических лиц на основе гибридных экспертных систем / О. И. Пятковский, Д. В. Лепчугов, В. В. Бондаренко // Алтайский государственный технический университет им. И. И. Ползунова. — 2008. — С. 127–129.
7. Pacelli V. An Artificial neural network approach for credit risk management / V. Pacelli, M. Azzolini // Journal of Intelligent Learning Systems and Applications. — 2011. — Р. 103–112.

8. Tomasz K. Fuzzy Logic in Financial Management / K. Tomasz // Fuzzy Logic – Emerging Technologies and Applications [ed. Elmer Dadios]. — 2012. — P. 259–286.
9. Zadeh L. A. Fuzzy sets / L. A. Zadeh // Information and Control. — 1965. — 8. — P. 338–353.
10. Zadeh L. A. The concept of a linguistic variable and its application to approximate reasoning—II / L. A. Zadeh // Information Science. — 1975. — 8. — P. 301–357.
11. Klir G. J. Fuzzy Sets and Fuzzy Logic. Theory and Applications / G. J. Klir, B. Yuan. — Prentice Hall, 1995. — 574 p.
12. Ishibuchi H. Performance Evaluation of Fuzzy Classifier Systems of Multidimensional Pattern Classification Problem / H. Ishibuchi, T. Nakashima, T. Murata // IEEE Transactions on Systems, Man, and Cybernetics. — 1999. — C.601 – 608.
13. Ishibuchi H. Comparison of the Michigan and Pittsburgh Approaches of the Design of the Fuzzy Classification Systems / H. Ishibuchi, T. Nakashima, T. Murata // Electronics and Communications in Japan (Part III: Fundamental Electronic Science). — 1997. — Vol. 80, Issue 12. — P. 10–19.
14. Eiben A. E. Introduction to Evolutionary Computing / A. E. Eiben, J. E. Smith. — Berlin, Heidelberg : Springer-Verlag, 2007. — 316 p
15. Goldberg D. E. Genetic Algorithms in Search, Optimization, and Machine Learning / D. E. Goldberg. — Addison-Wesley, 1989. — 432 p.
16. Goldberg D. E. A comparative analysis of selection schemes used in genetics algorithms / D. E. Goldbert, K. Deb // Foundation of genetic algorithms. — 1991. — P. 69–93.
17. De Jong K. A. An analysis of the behavior of a class of genetic adaptive systems / K. A. De Jong. — University of Michigan Ann Arbor, 1975. — 268 p.

18. Kumar R. Blending Roulette Wheel Selection and Rank Selection in Genetic Algorithms / R. Kumar, Jyotishree // International Journal of Machine Learning and Computing. — 2012. — Vol. 2, No. 4. — P. 365–370.
19. Syswerda G. Uniform crossover in genetic algorithms / G. Syswerda // Proceedings of the 3rd International Conference on Genetic Algorithms [ed. J. D. Schaffer]. — Morgan Kaufmann Publishers Inc., 1989. — P. 2–9.
20. Moscato P. On evolution, search, optimization, genetic algorithms and martial arts: Toward memetic algorithms / Pablo Moscato // C3P Report 826: Caltech Concurrent Computation Program. — Caltech, CA, 1989. — P. 33–48.
21. Dawkins R. The Selfish Gene / Richard Dawkins. — [3rd ed.]. — Oxford, New York : Oxford University Press, 2006. — 360 p.
22. Тавров Д. Ю. Меметичний метод побудови системи нечіткого виведення для визначення ступеня кредитоспроможності особи / Д. Ю. Тавров, О. Р. Булах // Системний аналіз та інформаційні технології : матеріали 17-ї Міжнародної науково-технічної конференції SAIT 2015, Київ, 22–25 червня 2015 р. / ННК «ПІСА» НТУУ «КПІ». — К. : ННК «ПІСА» НТУУ «КПІ», 2015. — С. 173–174.
23. Credit Approval Data Set [Електронний ресурс]. — Режим доступу: <https://archive.ics.uci.edu/ml/datasets/Credit+Approval>.
24. Quinlan J. R. C4.5: Programs for Machine Learning / J. R. Quinlan. — San Mateo : Morgan Kaufmann, 1993. — 302 p.