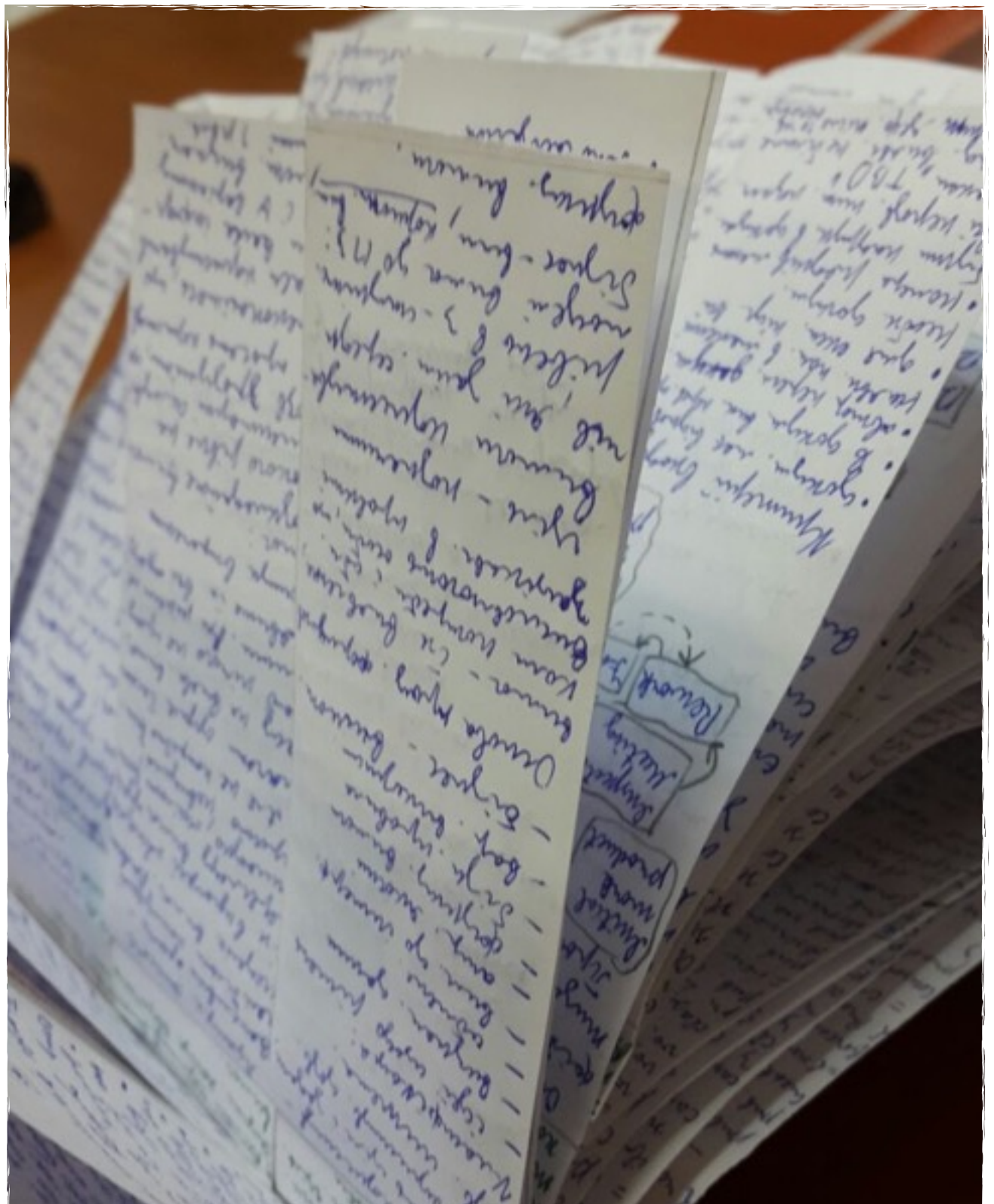


JUNE  
№6



# СИСТЕМНІ НАУКИ ТА КІБЕРНЕТИКА

## System Sciences & Cybernetics

### Електронний науковий журнал

Electronic Science Journal

No 6 June, 2017 ISSN 2221-7851

Університетський науковий збірник «Системні науки та кібернетика» створено для публікації студентських наукових статей при проведенні науково-дослідницької роботи у магістратурі.

Publisher: ESC «Institute for Applied System Analysis» NTUU «KPI»

[www.mmsa.kpi.ua/ssc](http://www.mmsa.kpi.ua/ssc)

---

# СИСТЕМНІ НАУКИ ТА КІБЕРНЕТИКА

Науковий електронний збірник НТУУ «КПІ»  
№ 6 • 2017

Виходить з вересня 2010 року

---

НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС «ІНСТИТУТ ПРИКЛАДНОГО  
СИСТЕМНОГО АНАЛІЗУ» НТУУ «КПІ» МОН ТА НАН УКРАЇНИ та  
ФАКУЛЬТЕТ ІНФОРМАТИКИ та ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ НТУУ «КПІ»

## **РЕДАКЦІЙНА КОЛЕГІЯ:**

### **Головний редактор**

В.Д. РОМАНЕНКО проф., д. т. н.

### **Заступник головного редактора**

П.І. БІДЮК проф., д. т. н.

### **Члени редколегії:**

Ю.В. БОГДАНСЬКИЙ проф., д. ф.-м. н.

В.Г. БОНДАРЕНКО проф., д. ф.-м. н.

Б.Н. ВАЙДА проф., Литва

В.Я. ДАНИЛОВ проф., д. т. н.

Ю.П. ЗАЙЧЕНКО проф., д. т. н.

Е. КРОПАТ проф., Німеччина

О.С. МАКАРЕНКО проф., д. ф.-м. н.

Н.Д. ПАНКРАТОВА проф., д. т. н.

А.І. ПЕТРЕНКО проф., д. т. н.

А.Ю. ПИЛИПЕНКО проф., д. ф.-м. н.

В.М. ПОДЛАДЧИКОВ проф., д. т. н.

О. ХАКАН проф., Туреччина

## **АДРЕСА РЕДАКЦІЇ:**

03056, м. Київ пр. Перемоги, 37, корп. 35, каб. 307

Навчально-науковий комплекс «Інститут прикладного системного аналізу»

НТУУ «КПІ» МОН та НАН України

Тел.: (044) 406-8359

<http://mmsa.kpi.ua/>

Відповідальний секретар: Н.В. Кузнєцова

Верстка: Н.В. Кузнєцова

## **Шановні читачі!**

Навчально-науковий комплекс «Інститут прикладного системного аналізу» НТУУ «КПІ» МОН та НАН України та факультет інформатики та обчислювальної техніки НТУУ «КПІ» видають університетський науковий електронний збірник:

### **«СИСТЕМНІ НАУКИ ТА КІБЕРНЕТИКА».**

Збірник публікує роботи теоретичного і прикладного характеру з широкого спектру проблем, пов'язаних із системними науками та кібернетикою.

Провідні тематичні розділи журналу:

- методи обробки інформації;
- прикладні методи системного аналізу;
- проектування кібернетичних систем;
- системні дослідження;
- математичне і програмне забезпечення;
- системи підтримки прийняття рішень;
- комп'ютерні системи та мережі;
- розв'язування задач оптимізації математичної кібернетики;
- теорія управління та обчислювальна математика;
- математичне моделювання та прогнозування економічних, технічних, екологічних і соціальних процесів;
- аналіз процесів і систем в умовах ризиків.

**Головний редактор наукового збірника** — заступник директора Навчально-наукового комплексу «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут» МОН та НАН України, доктор технічних наук, професор Віктор Демидович РОМАНЕНКО.

Університетський науковий електронний збірник «Системні науки та кібернетика» створено для публікації студентських наукових статей при проведенні науково-дослідної роботи в магістратурі.

Статті у збірнику публікуються українською, російською та англійською мовами у тому вигляді, в якому були надані авторами. Автори повністю несуть відповідальність за усі матеріали, наведені у статтях.

# СИСТЕМНІ НАУКИ ТА КІБЕРНЕТИКА

6 • 2017

## ЗМІСТ

<i>Бойко І. О., Бідюк П. І.</i> Застосування скорингового аналізу у кредитуванні.....	6
<i>Діденко Д. Г., Климковська К. Д.</i> Розподілене обчислення і прогнозування важливості брендів в Інтернеті .....	22
<i>Кангун Г. І., Бідюк П. І.</i> Моделювання військових конфліктів з використанням байєсівських мереж .....	34
<i>Діденко Д. Г., Климковська С. Д.</i> Релевантність посилань в Інтернеті за різними критеріями на рівні задачі оптимального розміщення контекстної реклами з нелінійною складовою .....	49
<i>Кобера В. В., Каніовська І. Ю.</i> Застосування реальних опціонів в управлінні ризиками інвестиційного проекту будівництва Великої кільцевої дороги навколо м. Київ .....	61
<i>Конєв Д. В., Бідюк П. І.</i> Моделювання і прогнозування нелінійного нестационарного процесу ціноутворення на нафту .....	76
<i>Костриця Д.В.</i> Розробка системної методології побудови скорингової карти.....	86
<i>Надточій О. В.</i> Моделювання і прогнозування демографічних процесів в Україні.....	99
<i>Пінчук В., Шишко Н.</i> Прогнозування ФЕП статистично-ймовірнісними методами та комбінування оцінок прогнозів .....	107
<i>Кузнєцова Н.В., Куца К.В., Штогрін С.Р.</i> Застосування методології аналізу виживання для дослідження споживчих ризиків .....	126

<i>Курін І. М., Тимощук О.Л. Методи нелінійної оптимізації в задачах проектування електричних двигунів .....</i>	136
РЕФЕРАТИ .....	148
ВІДОМОСТІ ПРО АВТОРІВ .....	160

## ЗАСТОСУВАННЯ СКОРИНГОВОГО АНАЛІЗУ У КРЕДИТУВАННІ

І.О. Бойко, П.І. Бідюк

### Вступ

Основною сучасною системною методологією прогнозування кредитних ризиків є кредитний скоринг, що полягає у розробці математичних моделей спеціального типу – скорингових моделей, які також називаються скоринговими картами. Метою їх побудови є прогнозування майбутнього стану обслуговування позичальником заборгованості або прогнозування довільних поведінкових показників за договором, клієнта, виходячи із соціально-демографічних характеристик, параметрів кредитного продукту, минулих поведінкових індикаторів.

Відомими сучасними теоретиками і практиками в галузі управління фінансовими ризиками є професори Ю.П. Зайченко, В.М. Подладчиков, Джонатан Н. Крук, Лін С. Томас, Девід Дж. Хенд, Л.М. Любчик, Елізабет Мейз, дослідники Наїм Сіддікі, Девід Б. Едельман та інші. Значний внесок у дослідження задач бінарної класифікації за допомогою логістичної регресії зробили Девід В. Хосмер, Стенлі Лемешоу, Пол Д. Елісон. Першим вченим, хто застосував підхід класифікації популяції на прикладі рослин був Рональд Ейлмер Фішер у 1936 р., а першим дослідником, який застосував таку методику для бінарної класифікації кредитів у 1941 р., будучи таким чином основоположником кредитного скорингу, є Девід Дюран. Він написав фундаментальну книгу «Елементи ризику у фінансуванні споживчої розстрочки» [1-7]. Системна методологія кредитного скорингу з точки зору системного аналізу відповідає фундаментальним принципам процедурної відкритості та раціональної доповнюваності, зокрема. До недоліків, подолання яких має найвищу актуальність, насамперед відносяться: відсутність чітких обмежень, визначеність ваг категорій вхідних змінних та їх інформаційної статистики, а також класичної бінарної логістичної регресії лише для випадку бінарної цільової змінної, визначеність ключових показників якості прогнозів – індексу Джині, відсутність рекомендацій щодо ініціалізації

початкового вектора коефіцієнтів у методі максимальної правдоподібності, низька застосовність популярних класичних методів чисельного інтегрування при оцінюванні рівнів статистичної значимості коефіцієнтів логістичної регресії та інші.

Дослідження присвячене аналізу кредитних ризиків за допомогою сучасних методів інтелектуального аналізу даних (логістична регресія, дерева рішень та байєсівська мережа) з використанням статистичних даних.

### **Постановка задачі**

Метою роботи є виконання таких завдань:

- зібрати статистичні дані стосовно характеристик позичальників кредитів для виконання обчислювальних експериментів;
- вибрати типи математичних моделей і методів інтелектуального аналізу даних для формального опису позичальників кредитів;
- побудувати вибрані типи моделей за допомогою статистичних даних;
- виконати порівняльний аналіз отриманих результатів.

У цьому дослідженні запропоновано новий підхід до прогнозування платоспроможності клієнтів фінансових установ шляхом комплексного застосування кількох методів інтелектуального аналізу даних та оцінки якості моделі.

### **Моделі для аналізу платоспроможності клієнтів і критерії їх якості**

Для аналізу якості моделей і встановлення найкращої моделі при розв'язанні певної задачі використовують декілька критеріїв для оцінювання адекватності моделей, зокрема такі [8]:

- загальна точність моделі;
- помилки I-го та II-го роду;
- індекс GINI та ROC-крива.

*Загальна точність моделі* (CA – common accuracy) визначається як:

$$CA = \frac{CorrectForecast}{N}, \text{ де } CorrectForecast \text{ – кількість вірно спрогнозованих}$$

випадків, а  $N$  – загальна кількість випадків. Загальна точність моделі є дещо суб'єктивною оцінкою, оскільки вона залежить від долі дефолтів в моделі, а також від порогу відсікання [9]. Для різних значень порогу точність моделі також буде приймати різні значення.

*ROC-крива* (Receiver Operation Characteristic – робоча характеристика приймача) показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих негативних прикладів. Перші називаються істинно позитивними, а інші – невірно негативними множинами. При цьому припускається, що у класифікатора існує певний параметр, варіюючи який, отримується певне розбиття на класи. Цей параметр часто називають порогом або точкою відсікання (cut-off), в залежності від якого будуть отримані різні величини помилок I-го та II-го роду.

Більш зрозумілим і частіше згадуваним в літературі параметром оцінювання адекватності моделі є індекс GINI, розроблений італійським статистиком Коррадо Джіні, який тісно пов'язаний з чисельним показником площі під ROC-кривою. *Індекс GINI* – це площа області між діагоналлю і кривою Лоренца, поділена на площу усієї області під діагоналлю. Індекс GINI широко використовується для аналізу роздільної здатності системи оцінювання при управлінні кредитними ризиками, тобто оцінки здатності моделі розділяти клієнтів на схильних та несхильних до дефолту. Якщо модель здатна оцінити клієнтів за ймовірністю дефолту, то більшість клієнтів, схильних до дефолту, мають отримати більшу ймовірність дефолту. Відповідно, найменша ймовірність дефолту має бути для клієнтів, несхильних до дефолту.

Діапазон значень індексу GINI складає  $0 \leq G \leq 1$ , а моделі з найвищою роздільною здатністю, тобто моделі, які роблять високоякісну класифікацію схильних до дефолту клієнтів і клієнтів, не схильних до дефолту, отримують найвищі значення коефіцієнтів. На практиці оцінка якості моделі суттєво залежить від даних, за якими вона будується. Для застосування на практиці скорингу (оцінки фінансового стану нових клієнтів) індекс GINI на рівні 55% є вже дуже високим, в той час як для скорингу поведінки (оцінки фінансового стану існуючих клієнтів) індекс GINI зазвичай приймає значення вищі 70%.



## Моделі на основі логістичної регресії

Логістична регресія – це різновид множинної регресії, загальне призначення якої полягає в аналізі зв'язку між декількома незалежними змінними (регресорами або предикторами) і залежною змінною. Бінарна логістична регресія, як випливає з назви, застосовується у випадку, коли залежна змінна є бінарною. Іншими словами, за допомогою логістичної регресії можна оцінювати вірогідність того, що подія наступить для конкретної випробовуваної (хворий/здоровий, повернення кредиту/ дефолт і т.д.) [10].

Розв'язання задачі класифікації за регресією може бути сформульоване інакше: замість прогнозу за бінарною змінною ми передбачаємо використання безперервної змінної із значеннями на відрізку  $[0,1]$  при будь-яких значеннях незалежних змінних. Це досягається застосуванням нелінійного регресійного рівняння (логіт-перетворення) (1):

$$P = \frac{1}{1 + e^{-x}}, \quad (1)$$

де  $P$  – ймовірність того, що відбудеться подія, яка нас цікавить;  $e$  – основа натурального логарифму;  $x$  – лінійне рівняння регресії.

Припустимо, що ми розглядаємо нашу залежну змінну в термінах основної ймовірності  $P$ , що лежить між 0 і 1. Перетворимо цю ймовірність  $P$  таким чином:

$$P' = \log_e(P/(1 - P)). \quad (2)$$

Це перетворення зазвичай називають логістичним або логіт-перетворенням. Теоретично  $P$  може приймати будь-яке значення. Оскільки логістичне перетворення вирішує проблему стосовно обмеження на інтервалі 0-1 для первинної залежної змінної (ймовірність), то перетворені значення можна використовувати у звичайному лінійному регресійному рівнянні. А саме, якщо провести логістичне перетворення обох частин описаного вище рівняння, ми отримаємо звичайну модель лінійної регресії.

## Методика побудови байєсівських мереж для скорингового аналізу

Мережі Байеса (МБ) – це ймовірнісно-статистична модель, яка дозволяє поєднати досить просте графічне представлення деякого процесу з його ймовірнісним характером, проаналізувати можливі варіанти розвитку ситуації,

відстежити правильність встановлення причинно-наслідкового зв'язку між окремими подіями і завдяки цьому підвищити обґрунтованість рішень при аналізі складних проблемних ситуацій [11].

Формально, *мережа Байєса* – це трійка  $N = \langle V, G, J \rangle$  першою компонентою якої є множина змінних  $V$ ; другою – спрямований ациклічний граф  $G$ , вузли якого відповідають випадковим змінним модельованого процесу;  $J$  — спільний розподіл ймовірностей змінних  $V = \{X_1, X_2, \dots, X_n\}$ . Виконується марковська умова: кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної.

Таким чином, *мережа Байєса (МБ)* – це модель представлення ймовірнісних залежностей (взаємозв'язків) між вершинами графа. Зв'язок  $A \rightarrow B$  називають причинним, якщо подія  $A$  є причиною виникнення  $B$ , тобто якщо існує механізм впливу значень змінної  $A$  на значення, які приймає змінна  $B$ . Мережу Байєса називають причинною (каузальною) тоді, коли всі її зв'язки є причинними.

При побудові мереж Байєса можливі випадки, коли попередня структура мережі задана емпірично (отримана від експертів, отримана за допомогою інших методів чи алгоритмів). Якщо структура невідома, то її потрібно побудувати. В обох випадках за наявності статистичної інформації зручно використовувати евристичний алгоритм побудови мережі Байєса, який у першому випадку скоригує структуру мережі, а у другому – побудує структуру мережі, яку потім можна буде модифікувати, використовуючи знання експертів.

*Евристичний метод побудови мереж Байєса* складається з двох етапів. На першому етапі виконується обчислення значень взаємної інформації між усіма вершинами. На другому етапі виконується цільовий пошук, який використовує в якості оціночної функції оцінку мінімальної довжини (ОМД), яка ґрунтується на принципі опису, що застосовується на кожній ітерації алгоритму навчання.

*Вхідні дані.* Множина навчальних даних  $D = \{d_1, \dots, d_n\}$ ,  $d_i = \{x_i^{(1)} x_i^{(2)} \dots x_i^{(N)}\}$  (нижній індекс – номер спостереження, а верхній – номер змінної),  $n$  – кількість спостережень,  $N$  – кількість вершин (змінних).

*Перший етап.* Для усіх пар вершин обчислюють значення сумісної інформації  $Set\_MI = \left\{ MI(x^i, x^j); \forall i, j \right\}$  за формулою:

$$MI(x^i, x^j) = \sum_{x^i, x^j} P(x^i, x^j) \cdot \log \left( \frac{P(x^i, x^j)}{P(x^i) \cdot P(x^j)} \right) \quad (3)$$

Елементи множини  $Set\_MI$  упорядковують за зростанням  $Set\_MI = \{MI(x^{m_1}, x^{m_2}), MI(x^{m_3}, x^{m_4}), MI(x^{m_5}, x^{m_6}), \dots\}$ .

*Другий етап . Крок 1.* Із множини значень взаємної інформації  $Set\_MI$  вибирають перші два максимальні значення  $MI(x^{m_1}, x^{m_2}), MI(x^{m_3}, x^{m_4})$  і будується множина моделей  $G$  виду:  $\{(m_1 \rightarrow m_2; m_3 \rightarrow m_4), (m_1 \rightarrow m_2; m_3 \leftarrow m_4), (m_1 \leftarrow m_2; m_3 \leftarrow m_4), (m_1 \leftarrow m_2; m_3 \rightarrow m_4), (m_1 \leftarrow m_2; m_3 \text{ не залежить від } m_4), (m_1 \rightarrow m_2; m_3 \text{ не залежить від } m_4), (m_1 \text{ не залежить від } m_2; m_3 \rightarrow m_4), (m_1 \text{ не залежить від } m_2; m_3 \leftarrow m_4), (m_1 \text{ не залежить від } m_2; m_3 \text{ не залежить від } m_4)\}$ .

Запис виду  $m_i \rightarrow m_j$  означає, що вершина  $x^{m_i}$  є прашуром вершини  $x^{m_j}$ .

*Крок 2.* Потім серед усіх моделей множини  $G$  здійснюється пошук. В параметрі  $g^*$  зберігається оптимальна мережева структура. Оптимальною буде та структура, у якій найменше значення функції  $L(g, x^n)$ .  $L(g, x^n)$  – це опис мінімальної довжини структури моделі  $g \in G$  при заданій послідовності спостережень з  $n$  спостережень  $x^n = d_1 d_2 \dots d_n$ , обчислюється за формулою

$$L(g, x^n) = H(g, x^n) + \frac{k(g)}{2} \cdot \log(n), \text{ де } k(g) \text{ – кількість незалежних умовних}$$

ймовірностей у мережній структурі  $g$ , а  $H(g, x^n)$  – емпірична ентропія, яка обчислюється за виразом:

$$H(g, x^n) = \sum_{j \in J} H(j, g, x^n), \quad k(g) = \sum_{j \in J} k(j, g), \quad (4)$$

де ОМД  $j$ -ї вершини обчислюється за формулою:

$$L(j, g, x^n) = H(j, g, x^n) + \frac{k(j, g)}{2} \cdot \log(n), \quad (5)$$

$k(j, g)$  – кількість незалежних умовних ймовірностей  $j$ -ї вершини:

$$k(j, g) = (\alpha^{(j)} - 1) \cdot \prod_{k \in \phi(j)} \alpha^k, \quad \text{де } \phi(j) \subseteq \{1, \dots, j-1, j+1, \dots, N\} \text{ – така множина,}$$

що  $\Pi^{(j)} = \{X^{(k)} : k \in \phi^{(j)}\}$ . Емпірична ентропія  $j$ -ї вершини обчислюється за формулою (6):

$$H(j, g, x^n) = \sum_{s \in S(j, g)} \sum_{q \in A^{(j)}} -n[q, s, j, g] \cdot \log \frac{n[q, s, j, g]}{n[s, j, g]}, \quad (6)$$

$$\text{де } n(s, j, g) = \sum_{i=1}^n I(\pi_i^{(j)} = s); \quad n[q, s, j, g] = \sum_{i=1}^n I(x_i = q, \pi_i^{(j)} = s),$$

а  $\pi^{(j)} = \Pi^{(j)}$  означає  $X^{(k)} = x^{(k)}, \forall k \in \phi^{(j)}$ , функція  $I(E) = 1$ , коли предикат  $E = true$ , в іншому випадку  $I(E) = 0$ .

Алгоритм навчання МБ з використанням ОМД виглядає таким чином:

1.  $g^* \leftarrow g_0 (\in G)$ .
2. Для  $\forall g \in G - \{g_0\}$  якщо  $L(g, x^n) < L(g^*, x^n)$ , то тоді  $g^* \leftarrow g$ .
3. На вихід подається  $g^*$  у якості рішення.

*Крок 3.* Після того як знайдена оптимальна структура (структури)  $g^*$  із  $G$ , із множини значень сумісної інформації  $Set\_MI$  вибирають наступне максимальне значення  $MI(x^{j\_next_i}, x^{j\_next})$ . За отриманим значенням

$MI(x^{i\_next}, x^{j\_next})$  і структурою (структурами)  $g^*$  будується множина моделей  $G$  виду  $\{ (g^*; i\_next \rightarrow j\_next), (g^*; i\_next \leftarrow j\_next), (g^*; i\_next$  не залежить від  $j\_next) \}$ . Після цього виконується крок 2.

*Умова закінчення роботи алгоритму.* Евристичний метод буде виконуватись до тих пір, поки не буде проаналізовано певне число елементів множини або усі  $\frac{N \cdot (N - 1)}{2}$  елементи множини  $Set\_MI$ . Як показує практика, у більшості випадків немає сенсу виконувати аналіз більше ніж половини (тобто  $\frac{N \cdot (N - 1)}{4}$ ) елементів множини  $Set\_MI$ .

*Дані на виході алгоритму.* У результаті застосування евристичного алгоритму отримують оптимальну структуру (структури)  $g^*$ .

Після того як мережа Байєса побудована, потрібно застосувати метод формування висновку. Якщо задані ймовірності певних подій, то висновок здійснюється шляхом перерахування ймовірностей за допомогою теореми Байєса. Якщо ж ймовірності подій не задані, але є певна статистика (дані), то можна використати алгоритм обчислення імовірнісного висновку на основі навчальних даних.

### **Методика побудови дерева рішень**

Дерева рішень – послідовні ієрархічні структури, що складаються з вузлів, які містять правила, тобто логічні конструкції вигляду "якщо ... то ..." [12]. Кінцевими вузлами дерева є "листя", відповідне знайденим рішенням і об'єднуючі деяку кількість об'єктів вибірки, що класифікуються. Це схоже на те, як положення листа на дереві можна задати, вказавши послідовність гілок, що веде до нього, починаючи від кореня і закінчуючи останньою гілочкою, на якій лист росте.

Нехай в таблиці довільних спостережень  $X$  розмірністю  $m > 1$  одна з ознак, виміряна у порядковій шкалі, визначає клас об'єкту і може приймати значення з деякого фіксованого набору  $\{y_1, y_2, \dots, y_k, \dots, y_p\}$ . Необхідно на основі

навчальної вибірки сформувати дерево класифікації (дерево рішень), що містить сукупність логічних умов, які дозволяють для довільного виміру  $x$  з  $X$  вказати клас якості  $U_k$ , до якого воно може належати.

Є цілий ряд причин, що роблять дерева класифікації гнучкішим засобом, ніж традиційні методи аналізу:

- схема одновимірного розгалуження, яка дозволяє вивчати ефект впливу окремих предикторних змінних і проводити послідовний аналіз їх внеску;
- можливість одночасно працювати із змінними різних типів, виміряних в безперервних і порядкових шкалах, або здійснювати будь-яке монотонне перетворення ознак;
- відсутність попередніх припущень про закони розподілу даних.

Область застосування дерева рішень сьогодні час широка, але всі завдання, що вирішуються за цим методом, можуть бути об'єднані в три такі групи.

*Опис даних:* дерева рішень дозволяють зберігати інформацію даних в компактній формі, тобто замість багатоелементних таблиць даних ми можемо зберігати дерево рішень, яке містить в концентрованій формі точний опис об'єктів;

*Класифікація:* дерева рішень відмінно справляються із завданнями класифікації, тобто віднесенням об'єктів до одного із заздалегідь відомих класів; при цьому цільова змінна повинна бути виміряна у порядковій шкалі.

*Регресія:* якщо цільова змінна має неперервні значення, дерева рішень дозволяють встановити залежність цільової змінної від незалежних (вхідних) змінних. Наприклад, до цього класу відносяться завдання чисельного прогнозування (прогнози значень цільової змінної).

Типовий алгоритм побудови топології (структури) дерев рішень складається з двох етапів: “побудова” або “створення” дерева (tree building) та “скорочення” дерева (tree pruning). При побудові дерева вирішуються питання вибору критерію розщеплення та зупинки навчання (якщо це передбачено

алгоритмом). У ході етапу скорочення дерева вирішується питання відсікання деяких його гілок.

### Приклади застосування методики

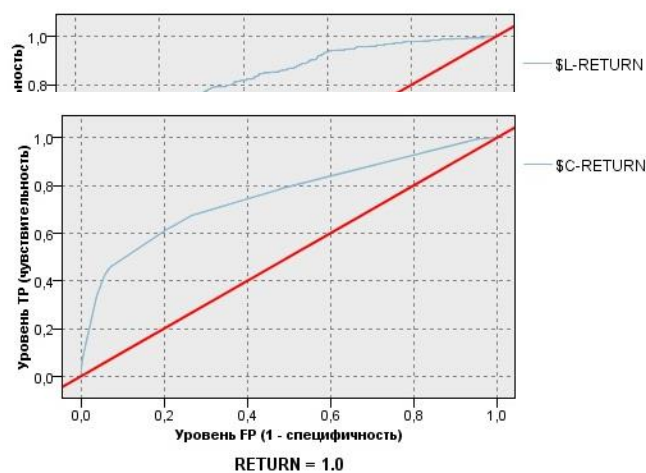
Побудовані нові моделі для оцінювання кредитоспроможності у формі логістичної регресії, мереж Байєса та дерев рішень для різних структур вибірок. У наведеній нижче таблиці 1 представлені основні показники точності даних моделей та якості оцінювання у вигляді індексу Джіні, значення AUC та помилок 1-го та 2-го роду для вибірки незміненої структури.

Таблиця 1 – Порівняння результатів точності моделей для вибірки незміненої структури

Назва методу	Індекс Джіні	Значення AUC	Помилка 1-го роду	Помилка 2-го роду	Точність моделі	Якість моделі
Логістична регресія	0,595	0,796	6,5%	16%	77,5%	Прийнятн а
Байєсівські мережі	0,736	0,868	7,6%	10,8%	81,56%	Дуже висока
Дерева рішень	0,51	0,755	4,6%	15,6%	79,81%	Прийнятн а

На наведених нижче рисунках 1, 2 і 3 представлені побудовані ROC-кривіна основі вибірки незміненої структури для різних моделей, які представляють рівень залежності чутливості моделі (TP) від її специфічності (FP).

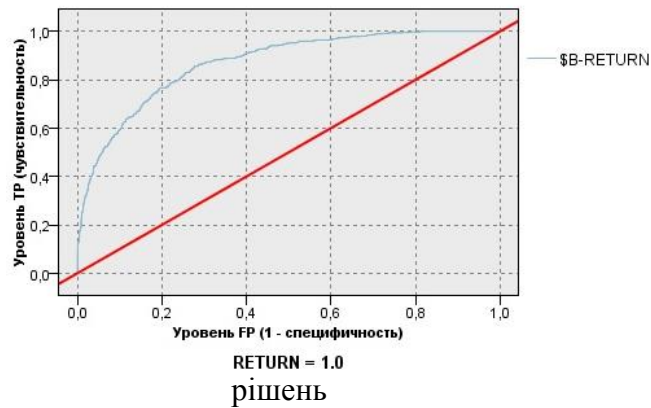
Рисунок 1 – ROC-крива для вибірки незміненої структури для логістичної



регресії

Рисунок 2 - ROC-крива для вибірки незміненої структури для мережі Байеса

Рисунок 3 - ROC-крива для вибірки незміненої структури моделі дерева



У наведеній нижче таблиці 2 представлені основні показники точності даних моделей та якості оцінювання у вигляді індексу Джіні, значення AUC та помилок 1-го та 2-го роду для вибірки зміненої структури (90%-повернень, 10%-дефолту).

Таблиця 2 – Порівняння результатів точності моделей для вибірки зміненої структури (90%-повернень, 10%-дефолту)

Назва методу	Індекс Джіні	Значення AUC	Помилка 1-го роду	Помилка 2-го роду	Точність моделі	Якість моделі
Логістична регресія	0,804	0,902	7,6%	1,6%	90,8%	Відмінна
Байєсівські мережі	0,928	0,964	2,8%	1,6%	94,6%	Відмінна
Дерева рішень	0	0,5	10%	-	90%	Незадовільна



На наведених нижче рисунках 4, 5 і 6 представлені побудовані ROC-криві на основі вибірки зміненої структури (90%-повернень, 10%-дефолту) для різних моделей, які представляють рівень залежності чутливості моделі (TP) від її специфічності (FP).

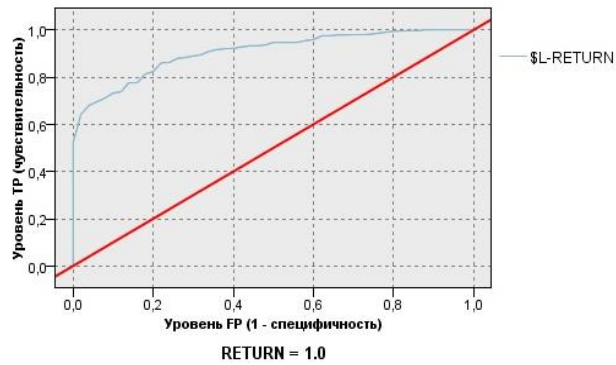


Рисунок 4 - ROC-крива для вибірки зміненої структури (90%-повернень, 10%-дефолт) для логістичної регресії

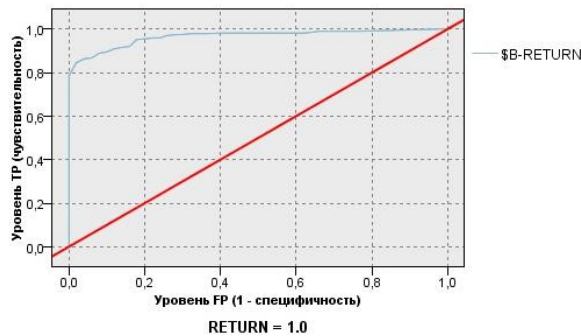


Рисунок 5 - ROC-крива для вибірки зміненої структури (90%-повернень, 10%-дефолт) для мережі Байєса

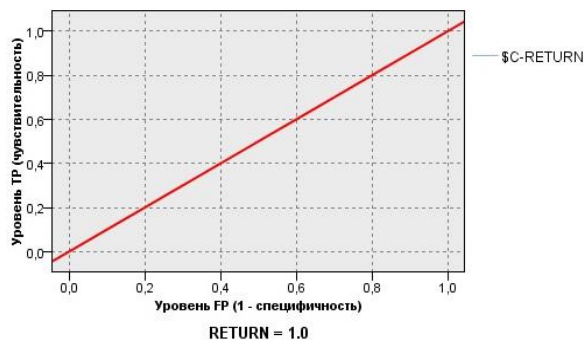


Рисунок 6 – ROC-крива для вибірки зміненої структури (90%-повернень, 10%-дефолт) для дерева рішень

У наведеній нижче таблиці 3 представлені основні показники точності даних моделей та якості оцінювання у вигляді індексу Джіні, значення AUC та помилок 1-го та 2-го роду для вибірки зміненої структури (50%-повернень, 50%-дефолту)

Таблиця 3 – Порівняння результатів точності моделей для вибірки зміненої структури (50%-повернень, 50%-дефолту)

Назва методу	Індекс Джіні	Значення AUC	Помилка 1-го роду	Помилка 2-го роду	Точність моделі	Якість моделі
Логістична регресія	0,671	0,835	12,4%	12,4%	75,2%	Дуже висока
Байєсівські мережі	0,807	0,904	7,6%	9,2%	83%	Відмінна
Дерева рішень	0,784	0,892	9%	6,6%	84,4	Дуже висока

На наведених нижче рисунках 7, 8 і 9 представлені побудовані ROC-кривіна основи вибірки зміненої структури (50%-повернень, 50%-дефолту) для різних моделей, які представляють рівень залежності чутливості моделі (TP) від її специфічності (FP).

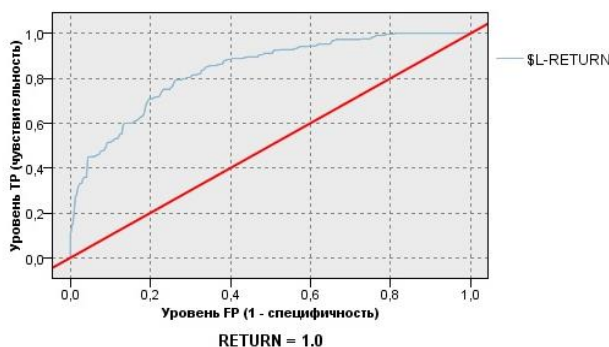


Рисунок 7 - ROC-крива для вибірки зміненої структури (50%-повернень, 50%-дефолт) для логістичної регресії

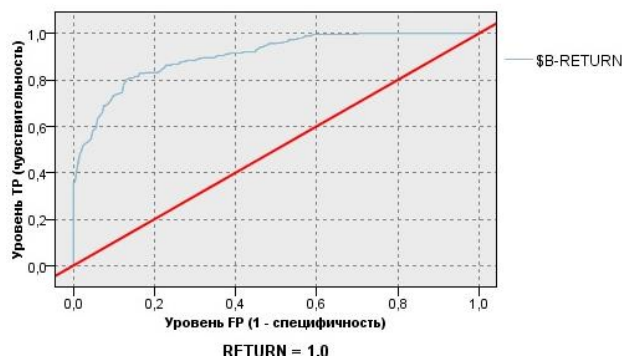


Рисунок 8 – ROC-крива для вибірки зміненої структури (50%-повернень, 50%-дефолт) для мережі Байєса

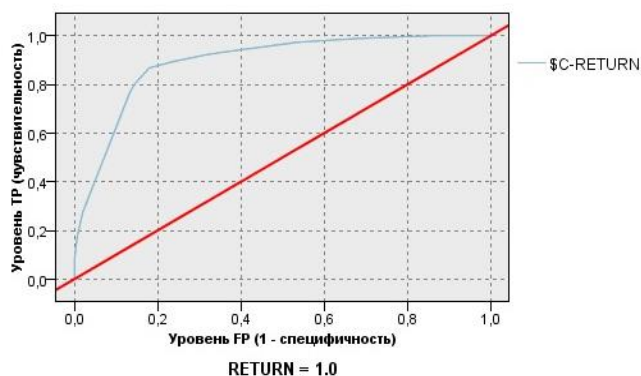


Рисунок 9 – ROC-крива для вибірки зміненої структури (50%-повернень, 50%-дефолт) для дерева рішень

### Аналіз отриманих результатів

В результаті комплексного застосування кількох методів інтелектуального аналізу даних, вдалося значно підвищити якість оцінювання – загальну точність платоспроможності у середньому на 2-3% та індекс Джіні на 9,6% з вибіркою зміненої структури (50%-повернень та 50%-дефолту); на 12% загальної точності та на 26% індекс Джіні з вибіркою зміненої структури (90%-повернень та 10%-дефолту) відповідно з найкращою побудованою моделлю мережею Байєса.

Як видно, показники точності інших моделей – логістичної регресії та дерев рішень, значно нижчі. Оскільки дерева рішень є грубшим інструментом прогнозування ніж логістична регресія та байєсівські мережі, для вибірки зміненої структури (90%-повернень, 10%-дефолту) він має найгірший результат показників точності та ROC-кривої.

## Висновки

Таким чином, у статті розглянуто задачу оцінювання кредитного ризику, визначені причини його виникнення та підходи до обчислення. Побудовані нові моделі для оцінювання кредитоспроможності у формі логістичної регресії, дерев рішень та мережі Байєса. Удосконалено системну методологію побудови моделей оцінювання кредитоспроможності клієнтів фінансових установ шляхом корегування структури вибірки даних. Представлено найкращу побудовану математичну модель у вигляді мереж Байєса з найвищою точністю як для вибірки з незмінною структурою так і для вибірок зі змінною структурою.

## Література

1. *Зайченко Ю.П.* Исследование операций: учебник / Ю.П. Зайченко. — 6-е изд., перераб. и доп. — К.: Издательский дом «Слово», 2003.
2. *Thomas L.C.* Credit Scoring and its applications: Monograph / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. — Philadelphia: SIAM, 2002. — 248 p.
3. *Siddiqi N.* Credit risk scorecards: developing and implementing intelligent credit scoring / Naeem Siddiqi. — Hoboken: John Wiley & Sons, Inc., 2006. — 196 p.
4. *Руководство по кредитному скорингу* / [Ванг Вэй, Влатса А. Димитра, Гленнон К. Деннис и др.]; под ред. Элизабет Мэйз; [пер. с англ. И.М. Тикота; науч. ред. Д.И. Вороненко]. — Минск: Гревцов Паблицер, 2008. — 464 с.
5. *Mok Jie-Men.* Reject Inference in Credit Scoring / Jie-Men Mok. — Amsterdam: ВМІ paper, 2009. — 38 p.
6. *Энциклопедия финансового риск-менеджмента* / [Барбаумов В.Е., Рогов М.А., Щукин Д.Ф. и др.]; под ред. А.А. Лобанова и А.В. Чугунова. — М.: Альпина Паблицер, 2003. — 786 с.
7. *Feelders A.J.* Credit scoring and reject inference with mixture models / A.J. Feelders // International Journal of Intelligent Systems in Accounting, Finance and Management. — 1999. — Vol. 8, № 4. — Pp. 271–279.
8. *Матрос Є.О.* Впровадження системного підходу до прогнозування обсягів втрат внаслідок реалізації банківських ризиків / Є.О. Матрос // Наук. вісті НТУУ «КПІ». — 2006. — № 3. — С. 37–44.
9. *Бідюк П.І.* Адаптивне прогнозування фінансово-економічних процесів на основі принципів системного аналізу / П.І. Бідюк // Наук. вісті НТУУ «КПІ». — 2009. — № 5. — С. 54–61.

10. *Бідюк П.І.* Теоретичні відомості щодо побудови моделей логіт та пробіт [електронний ресурс]. – Режим доступу: <http://sites.google.com/site/data4mining/pkis>.
11. *Кузнецова Н.В.* Системний підхід до аналізу кредитних ризиків з використанням мереж Байєса / Н.В. Кузнецова, П.І. Бідюк // *Наук. вісті НТУУ «КПІ»*. — 2008. — № 3. — С. 11–24.
12. *Rokach L.* Data mining with decision trees: theory and applications / Lior Rokach, Oded Maimon. — 2nd ed. — Singapore: World Scientific Publishing Co. Pte. Ltd., 2014. — 305 p.

## Розподілене обчислення і прогнозування важливості брендів в Інтернеті

Д.Г. Діденко, К.Д. Климковська

### Вступ

Щодня в мережі Інтернет публікується дуже багато контенту: відгуки про компанії та думки про них, фото і відео цікавих місць, опис нових товарів і багато іншого. Для того, щоб компанія ефективно функціонувала на просторах онлайн-спільнот, необхідний регулярний моніторинг видачі інформації в мережі Інтернет і його своєчасний аналіз. Активно працюючи з моніторингом соціальних мереж, компанії мають можливість швидше за конкурентів дізнаватися про нові тренди та очікувати потенційних клієнтів.

Говорити про важливість інформації, яку створюють користувачі соціальних мереж, стає загальноприйнятим. Через позитивні та негативні відгуки людей в Інтернеті компанії все частіше набувають і втрачають репутацію, а разом з нею і реальні гроші.

Користувачі соціальних мереж за одну добу створюють текст, який за обсягом перевершує всі твори Л.А. Толстого. Звичайно для текстів про споживчий досвід є спеціальні сервіси. Але далеко не кожен вважає своїм обов'язком писати відгуки тільки там, де це доречно. Багато хто віддає перевагу поділитися своїм досвідом зі своїми «друзями», а не з публікою того чи іншого сервісу.

Тобто в сучасному світі для вдалого розвитку компанії дуже важливо проводити моніторинг брендів та дізнаватися думку користувачів товару, наприклад через такий потужний механізм, як соціальні мережі.

Моніторинг соціальних мереж дозволяє безпосередньо почути клієнта в реальному часі, дізнатися його переваги, виявити цільову аудиторію, яка найбільш ймовірно стане споживачами продукту вибраної компанії.

У процесі відстеження згадувань та відгуків в різних соціальних медіа, компанія отримує інформацію, яка буде важлива для неї та її бізнесу [1].

Досить інформації компанія отримає також і про конкурентів, які перебувають на ринку. Це дозволить їй оперативно виправляти всі виникаючі проблеми, швидше реагувати на найменші зміни ринку та вподобань клієнтів. Саме на підставі звітів моніторингу соціальних мереж компанія зможе підібрати оптимальну стратегію подальшого розвитку свого бізнесу.

Крім того, моніторинг Інтернету та відстеження згадувань забезпечить компанію актуальними темами та інфо-приводами для додавання контенту на тематичні майданчики.

Зараз існує достатньо багато систем он-лайн моніторингу брендів. Майже кожна з них підтримує моніторинг у найпопулярніших соціальних мережах, таких як Facebook та

ВКонтакте. На ринку існують безкоштовні та платні системи. Далі наведено деякі загальні відомості про найпопулярніші системи он-лайн моніторингу.

### 1. Система IQBuzz.

IQBuzz – російський сервіс, який забезпечує збір інформації, аналізує частоту згадувань вибраної теми, виділяє тональність повідомлень, інформаційні сплески та їх причини. Він обробляє інформацію з Facebook, Twitter, ВКонтакте, LiveJournal, LiveInternet, Google+, YouTube, RuTube та безлічі інших джерел.

Цей сервіс крім негативних відгуків, дозволяє знаходити коментарі про конкурентів, цікаві новини з галузі бізнесу, знаходити нових клієнтів і тисячі інших можливостей, які відкриті перед компаніями, спрямованими на використання нових технологій.

Також сервіс дозволяє налаштувати автоматичні звіти, завдяки чому компанія-замовник завжди буде в курсі того, що пишуть про неї клієнти і отримує можливість оперативно реагувати на повідомлення в режимі реального часу. Додатковими можливостями IQBuzz є відображення різних графічних зрізів для розуміння всього, що відбувається в соціальних медіа.

IQBuzz пропонує безкоштовний доступ на 7 днів. Далі, в залежності від вибраного тарифного плану, користування сервісом коштує від 7500 до 70000 рублів на місяць.

Можливості IQBuzz:

- цілодобовий моніторинг найбільш популярних ресурсів соцмедіа;
- можливість колективної роботи;
- аналіз тональності повідомлень;
- таргетинг аудиторії за статтю, віком та іншими соціальними характеристиками;
- виділення фактичної інформації: відсотки, дати, персони, гроші;
- генерація періодичної звітності;
- понад 2,5 млрд. повідомлень в архіві;
- широкі можливості для аналітики;
- автоматичне визначення негативу і позитиву.

Технології IQBuzz:

- власна мова запитів і потужний пошук;
- аналіз соціально-демографічних характеристик авторів;
- можливість глибокого ретроспективного аналізу;
- власний пошуковий движок [2].

### 2. Система YouScan.

YouScan – український стартап, заснований у 2009 році. Сервіс відстежує згадки брендів, продуктів, конкурентів у блогах, форумах, соціальних мережах ВКонтакте, Twitter і

навіть в YouTube, і представляє результати моніторингу в зручному аналітичному інтерфейсі з функціями командної роботи.

Крім підписки на систему моніторингу, компанія також пропонує клієнтам послуги власних аналітиків і партнерів (оцінка тональності, категоризація згадок, складання аналітичних звітів).

Також у власному акаунті клієнт може самостійно створювати і видаляти теми в будь-який момент. Тема - це набір ключових слів, які описують кожен об'єкт моніторингу (наприклад, варіанти назви бренду, продукту, категорії).

YouScan пропонує безкоштовний доступ на 14 днів. Далі, в залежності від вибраного тарифного плану, користування сервісом коштує від \$1990 до \$19990 на рік.

Можливості YouScan:

- моніторинг Twitter, ВК, Facebook, Мій Світ, блогів, форумів;
- звіти за джерелами і авторами;
- функції командної роботи;
- експорт даних в Excel;
- автоматична розмітка тональності;
- відповіді в Twitter з системи;
- категоризація згадок (теги);
- послуги аналітика;
- розподіл прав доступу до тем;
- інтеграція з CRM;
- виділений менеджер [3].

### 3. Система BrandSpotter

BrandSpotter – російсько-український сервіс для моніторингу соціальних мереж, створений у 2009 році.

Розглянемо принцип роботи даного сервісу. Поряд з кожним висловлюванням формується його емоційне забарвлення (тональність). Воно фільтрується: по емоціям (тональностям), за темами, по конкурентам, за джерелами і всередині кожного з цих фільтрів знову по інших. Дані зі сторінки можна експортувати у вигляді xls-файлу.

За результатами аналізу будується графік. На графіку можна вибрати: динаміку по емоціям (тональностям), динаміку за темами, динаміку по конкурентам та будь-яку з цих кривих окремо. При цьому, всередині вибраного можна фільтрувати за джерелами і за емоціями. Для побудови звіту можна вибрати період даних і крок звіту (по днях, тижнях, місяцях).



Також можна встановлювати власні фільтри. Наприклад, це можуть бути автори, які згадували: всі ваші об'єкти пошуку, тільки вибрані вами теми, тільки вибраних вами конкурентів, тільки негативно або тільки одну тему і її - тільки позитивно і так далі.

Можна вибрати автора і переглянути всі його записи окремо. А можна відсортувати їх за кількістю згадок (хто говорить про бренд частіше за інших?). Або за кількістю передплатників (що говорять про вас впливові користувачі?).

Ціна користування даним сервісом складає від 19350 до 49560 рублів на місяць [4].

Всі вищезазначені сервіси виконують автоматичний моніторинг соціальних мереж, блогів та інших Інтернет майданчиків, автоматично визначаючи тональність відгуків користувачів про бренд та подаючи результати у зручному вигляді для користувачів системи. Метою даної роботи є створення аналогічної системи моніторингу соціальних мереж та блогів, поліпшення точності визначення тональності відгуків про бренд та створення додаткової можливості короткострокового прогнозування щоденної кількості позитивних та негативних відгуків про бренд, якої не має жодна з зазначених та інших аналогічних систем.

Дослідження присвячене створенню алгоритму визначення тональності тексту на основі наївного Баєсівського класифікатора та побудові нелінійної моделі, на основі якої розроблено алгоритм короткострокового адаптивного прогнозування значень часових рядів (щоденної кількості позитивних та негативних відгуків про бренд).

### **Постановка задачі**

Задача автоматичного моніторингу соціальних мереж та блогів є дуже популярною в наш час та розглядалася багатьма компаніями. Актуальність цієї задачі є очевидною так як компаніям необхідний регулярний моніторинг видачі інформації в мережі Інтернет і її своєчасний аналіз для того щоб дізнатися переваги й недоліки власного бренду, виявити цільову аудиторію, яка найбільш ймовірно стане споживачами бренду та оцінити власну PR кампанію.

Для вирішення задачі автоматичного моніторингу соціальних мереж та блогів необхідно розробити алгоритм автоматичного визначення тональності відгуків користувачів про бренд, також з метою поліпшення існуючих систем моніторингу розробити алгоритм короткострокового прогнозування щоденної кількості позитивних та негативних відгуків про бренд. Результат моніторингу подається як співвідношення щоденної кількості позитивних, негативних та нейтральних відгуків про бренд.

**Мета роботи.** Розробивши відповідні алгоритми створити систему автоматичного моніторингу соціальних мереж та блогів з можливістю прогнозування щоденної кількості позитивних та негативних відгуків про бренд.

**Викладення теорії:** Для визначення тональності відгуків користувачів був застосований метод машинного навчання з учителем. Машинне навчання з учителем є

найбільш поширеним методом, який використовується в дослідженнях. Його суть полягає в тому, щоб навчити машинний класифікатор на колекції заздалегідь розмічених текстів, а потім використовувати отриману модель для аналізу нових документів.

Процес створення системи аналізу тональності складався з таких кроків:

1. Була зібрана колекція документів для навчання класифікатора (500 реальних відгуків користувачів Facebook).
2. Кожен документ з навчальної колекції був представлений у вигляді вектору ознак (комбінація уніграм та біграм слів).
3. Для кожного документа була вказана «правильна відповідь», тобто тип тональності (наприклад, позитивна чи негативна), за цими відповідями і був навчений класифікатор.
4. В якості алгоритму класифікації був вибраний наївний баєсівський класифікатор (NB) та метод опорних векторів (SVM).
5. Використання отриманої моделі.

Для розв'язання задачі класифікації тональності відгуків було вибрано розділення можливої тональності на 3 класи: позитивна, негативна та нейтральна. Для навчання класифікатору була вибрана ієрархічна класифікація (рис. 1), так як у порівнянні з пласкою вона дає більшу точність.



Рисунок 1 – Ієрархічна класифікація тональності

У випадку ієрархічної класифікації класи діляться на групи, і навчаються кілька класифікаторів для визначення груп. В даній задачі тональність була поділена на 3 класи («позитивний», «негативний» та «нейтральний»). Тому, згідно з алгоритмом, спочатку був

навчений бінарний класифікатор, який відокремлює нейтральні тексти від суб'єктивних; потім навчався класифікатор, який відокремлює позитивні думки від негативних.

Для даного класифікатора було вибрано за ознаки уніграми, біграми та їх комбінацію. Це було зумовлено тим, що уніграми і біграми дають кращі результати ніж N-грами більш високих порядків (триграми і вище), тому що вибірка навчання в більшості випадків недостатньо велика для підрахунку N-грам вищих порядків.

Відповідний Баєсовський класифікатор - це функція classify, визначена формулою:

$$\text{classify}(f_1, \dots, f_n) = \underset{c}{\operatorname{argmax}} p(c = C) \prod_{i=1}^n p(F_i = f_i | C = c).$$

При короткостроковому прогнозуванні, а також у випадку зміни зовнішніх умов, коли найбільш важливими є останні реалізації досліджуваного процесу, найбільш ефективними є адаптивні методи, які враховують нерівноцінність рівнів часового ряду.

Для короткострокового прогнозування значень часових рядів (значення кількості позитивних та негативних відгуків про бренд за день) побудуємо нелінійну модель, яка дозволить відобразити розвиток експоненційної тенденції, а також рядів без тенденції:

$$Y(t + k) = A_0 + A_1 \exp(k) + A_2 \exp\left(\frac{1}{k}\right),$$

де  $t$  – поточний час;

$k$  – час випередження.

Дану модель було вибрано тому, що вона є точнішою за лінійну та краще описує зміни значень часового ряду.

Вона будується таким чином:

1. За декількома першими точками часового ряду методом найменших квадратів знаходяться значення параметрів  $A_0$ ,  $A_1$  та  $A_2$ .

$$Y_p(t) = A_0 + A_1 \exp(t) + A_2 \exp\left(\frac{1}{t}\right).$$

2. Використовуючи знайдені параметри, знаходиться прогнозне значення на наступному кроці.

$$Y_p(t + k) = A_0(t) + A_1(t) \exp(k) + A_2(t) \exp\left(\frac{1}{k}\right), k = 1.$$

3. Знайдемо похибку прогнозування.

$$e(t + k) = Y(t + k) - Y_p(t + k).$$

4. Згідно з похибкою змінюються значення параметрів моделі.

$$A_0(t + 1) = A_0(t) + A_1(t) + A_2(t) + (1 - \beta)^2 e(t),$$

$$A_1(t + 1) = A_1(t) + A_2(t) + (1 - \beta)^2 e(t),$$

$$A_2(t + 1) = A_2(t) + (1 - \beta)^2 e(t),$$

де  $\beta$  – коефіцієнт дисконтування даних, .

5. З моделі з виправленими параметрами знаходиться прогноз на наступний крок та відбувається перехід до п.3 якщо  $t < N$ . При  $t \geq N$  отримане значення використовується як прогнозне, не змінюючи параметри моделі [5].

**Приклад застосування методики:** На основі наведених алгоритмів був створений веб-додаток, який реалізує функціонал моніторингу соціальних мереж Facebook та ВКонтакте, блогів lifehacker.ru та businessinsider.com, автоматичного визначення тональності відгуків про бренд та короткострокового прогнозування кількості щоденної кількості позитивних та негативних відгуків про бренд. Продукт було створено мовою програмування Java з використанням Spring Framework для конфігурування компонентів додатка, управління життєвим циклом Java-об'єктів, реалізації моделі model-view-control та для доступу до API зазначених соціальних мереж та блогів. Системою управління базами даних була вибрана MySQL, а засобом відображення між об'єктами та реляційними структурами (ORM) був вибраний Hibernate. Сервером веб-додатку був вибраний Tomcat 7. Власний програмний продукт був порівняний з системою BuzzLook. Для ілюстрування роботи двох систем та їх порівняння був вибраний український бренд «Roshen». Був проведений моніторинг його згадувань за останній місяць (травень 2016 року). Після проведення аналізу системи були порівняні та були визначені їх переваги та недоліки.

Основні переваги системи BuzzLook:

- можливість проводити моніторинг більшої кількості ресурсів (twitter, Facebook, ВКонтакте, livejournal тощо);
- зручний інтерфейс та генерування звітів;
- можливість відповідати з системи в twitter;
- автоматичне визначення тональності відгуків.

Недоліки системи BuzzLook:

- невисока якість визначення тональності.

Результати моніторингу в системі BuzzLook наведені на рисунку 2.

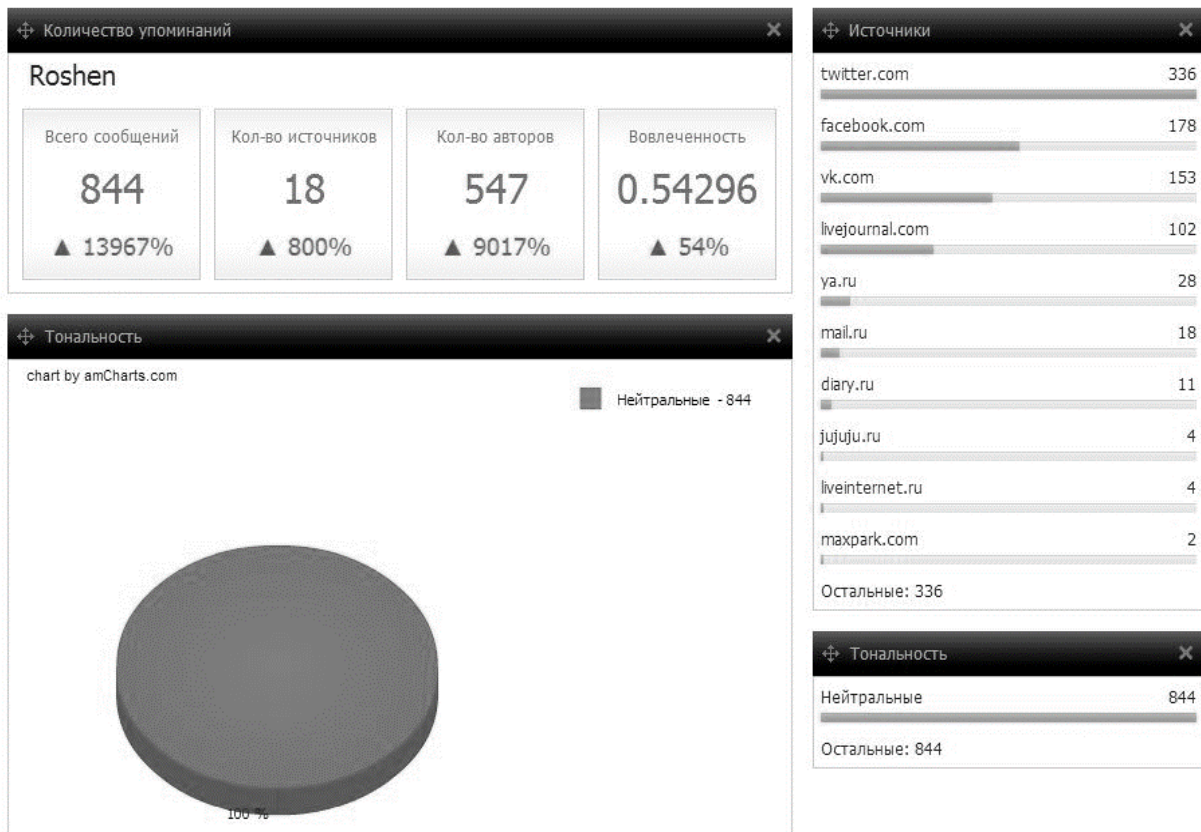


Рисунок 2 – Результати аналізу системи BuzzLook

Основні переваги власної системи:

- автоматичне визначення тональності відгуків;
- можливість фільтрувати відгуки за соціальними параметрами користувачів мережі;
- доволі висока точність визначення тональностей відгуків;
- можливість прогнозування щоденної кількості позитивних та негативних відгуків на наступні 7 днів.

Недоліки власної системи:

- проводиться моніторинг лише обмеженої кількості ресурсів: Facebook, ВКонтакте, Lifehacker.ru, Businessinsider.com.

Результати моніторингу в системі BuzzLook наведені на рисунку 3.

Brand's name \*

Time period

Max amount of feedbacks   Analyse All

Gender

Age range  -

**Proceed to analysis**

## Results of Analysis

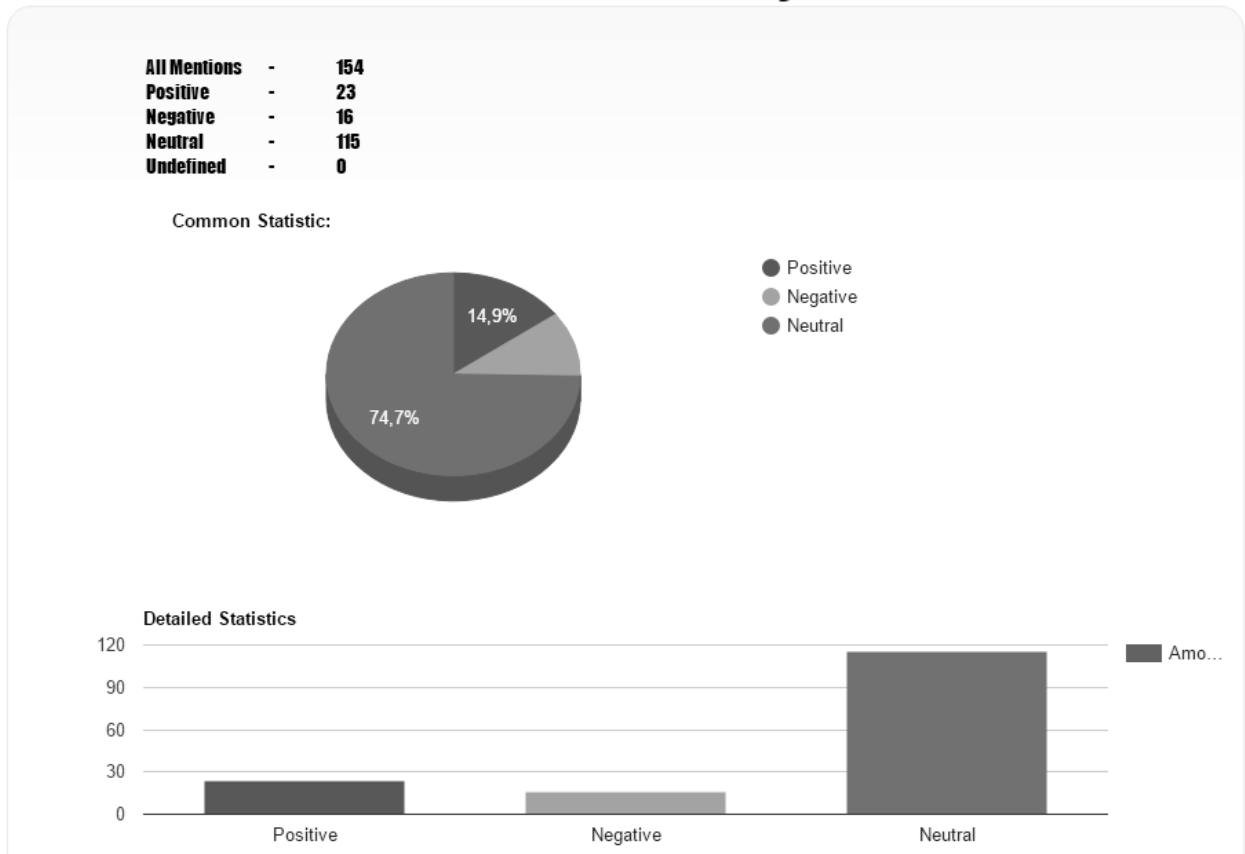


Рисунок 3 – Результати аналізу у власній системі

Результати короткострокового прогнозування наведені на рисунку 4.

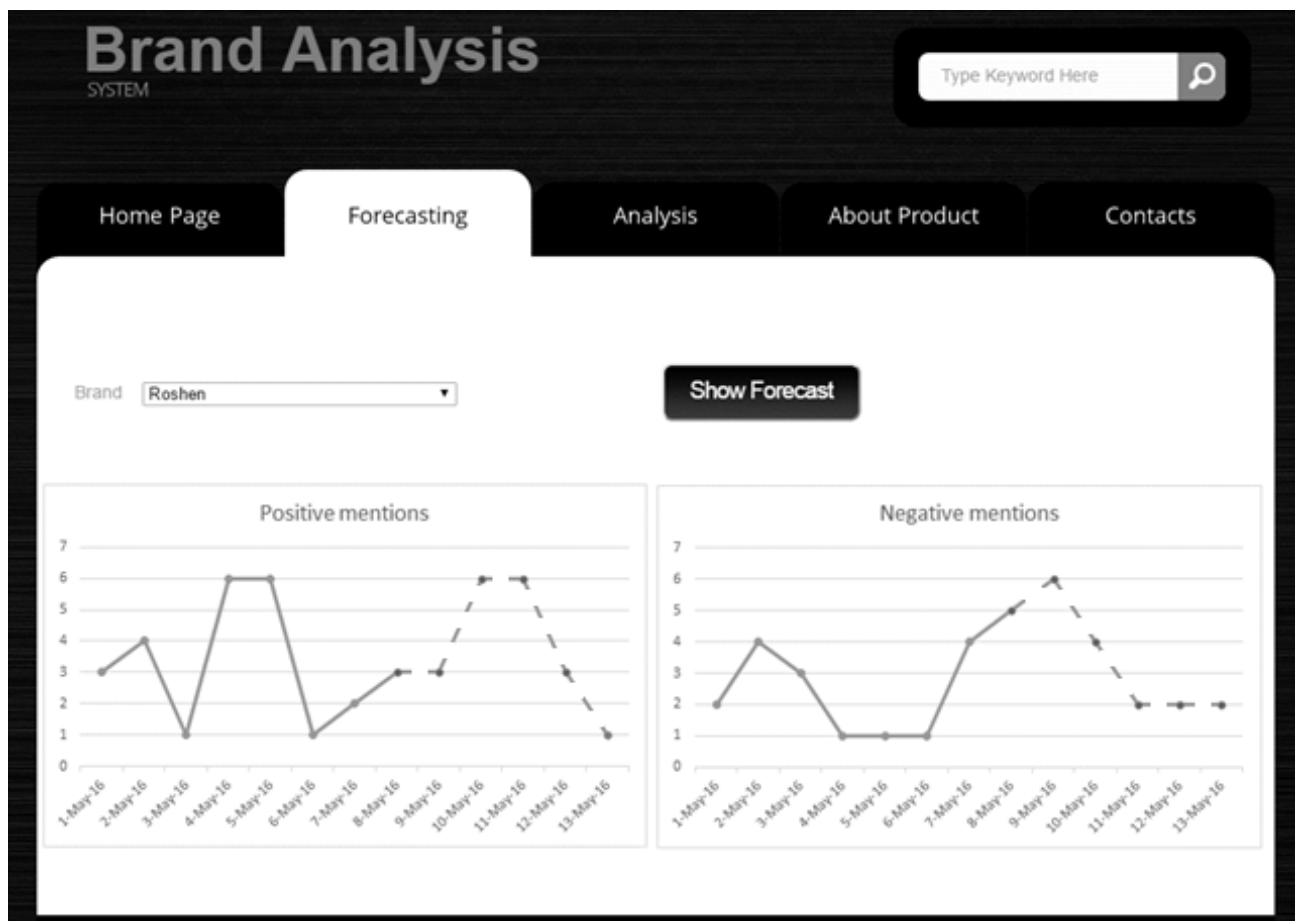


Рисунок 4 – Відображення результатів прогнозу

Суцільна лінія на графіку відображає зміну реальної кількості позитивних та негативних відгуків за попередній період, а штрих-пунктирна позначає зміну прогнозних значень кількості позитивних та негативних відгуків за наступні 6 днів.

Наведемо кількісне порівняння результатів роботи обох систем у таблиці 3.1.

Таблиця 3.1 – Порівняння роботи обох систем

Показник	BuzzLook	Власна система
Кількість знайдених згадувань бренду в соціальній мережі Facebook	178	154
Кількість позитивних згадувань	0	23
Кількість негативних згадувань	0	16
Кількість нейтральних згадувань	178	115
Кількість невстановлених згадувань	0	0

Тобто, як видно з результатів аналізу, власна система більш чутлива до емоційності відгуків, але в ній ще існує поняття невстановленої тональності, що рівносильне нейтральному згадуванню в системі BuzzLook.

## Аналіз отриманих результатів

Визначимо якість вирішення задачі. Для цього наведемо порівняння кількості визначених системою та експертом відгуків. Порівняння кількості визначених відгуків наведено в таблиці 2.

Таблиця 2 – Порівняння кількості відгуків, визначених системою та експертом

Тональність	Кількість відгуків, визначених системою	З них правильно визначених (думка експерта)
Позитивні відгуки	23	20
Негативні відгуки	16	15
Нейтральні відгуки	115	75

Тобто кількість відгуків, визначених системою складає:

$$\text{total number of opinions found by system} = 23 + 16 + 115 = 154.$$

З них правильно визначених:

$$\text{correctly extracted opinions} = 20 + 15 + 75 = 110.$$

Визначимо точність системи:

$$P = \frac{\text{correctly extracted opinions}}{\text{total number of opinions found by system}} = \frac{110}{154} \approx 0,71.$$

Програма, яка визначає тональність тексту з точністю 71%, робить це майже так само добре, як і людина. Тобто власна система достатньо точно визначає тональність тексту.

## Висновки

В результаті дослідження було реалізовано систему автоматичного моніторингу соціальних мереж та блогів. Реалізована система дозволяє проводити щоденний автоматичний моніторинг, отримувати статистику кількості позитивних, негативних та нейтральних відгуків про бренд та прогнозувати значення щоденної кількості позитивних та негативних відгуків про бренд на наступні 6 днів.

Для розв'язання поставленої задачі використовувались такі математичні методи: метод машинного навчання з учителем для автоматичного визначення тональності відгуків користувачів, наївний баєсівський класифікатор (NB) та метод опорних векторів (SVM) для класифікації тональності відгуків, ієрархічна класифікація для навчання класифікатору та адаптивний алгоритм короткострокового прогнозування на базі розробленої в роботі нелінійної моделі для прогнозування щоденної кількості позитивних та негативних відгуків про бренд, яка застосовується вперше.

Для оцінки результатів роботи методу власний програмний продукт був порівняний з системою моніторингу соціальних мереж BuzzLook. Було встановлено, що продукт визначає тональність тексту з точністю 71%, що свідчить про те, що він робить це майже так само добре,



як і людина. Тобто власна система достатньо точно визначає тональність тексту. Перспективою подальших досліджень є використання нейронної мережі, яка навчається самостійно для обчислення тональності відгуків про бренд.

### **Література**

1. *Халіров Д.* Маркетинг в соціальних мережах / Дамір Халіров. – М.: Манн, Іванов і Фербер, 2013. – 240 с.
2. *Система моніторингу брендів IQBuzz* [Електронний ресурс]. – Режим доступу: <http://iqbuzz.ru/>.
3. *Система моніторингу брендів YouScan* [Електронний ресурс]. – Режим доступу: <http://youscan.ru/>.
4. *Система моніторингу брендів BrandSpotter* [Електронний ресурс]. – Режим доступу: <http://brandspotter.ru/>.
5. *Марина Семененко, Инга Князева.* Модель Брауна / И. Князева // Сборник при финансовом университете при правительстве РФ, филиал в г. Калуге. – 2013. – №25. – С. 113-118.

Г.І. Кангун, П.І. Бідюк

## МОДЕЛЮВАННЯ ВІЙСЬКОВИХ КОНФЛІКТІВ З ВИКОРИСТАННЯМ БАЙЄСІВСЬКИХ МЕРЕЖ

### Вступ

Життя в умовах нестабільності міжнародних відносин та переходу до нових, складніших методів ведення конфліктів вимагає від України відповідного рівня системи національної безпеки та оборони, що зможе гарантувати можливість реалізації обраного політичного курсу і мінімізувати вплив інших держав на рішення, що приймаються, заради обрання курсу, оптимального саме для нашої держави, а не інших зацікавлених сторін.

Побудова моделі процесів, що мають відношення до можливості появи військових конфліктів, дозволяє всебічно розглянути та спрогнозувати розвиток ситуації, що в результаті полегшує пошук оптимального або раціонального рішення. Складність моделювання військового конфлікту полягає в першу чергу в тому, що він є динамічною системою з багатьма невідомими або невимірюваними змінними та параметрами. Крім того, необхідно враховувати неповноту та нечіткість інформації. Аналіз та моделювання подібної системи має на меті не тільки виявлення існуючих взаємозв'язків та залежностей, а і передбачення та можливе попередження майбутніх конфліктів на основі результатів аналізу.

Одним з методів аналізу та моделювання військових конфліктів є метод експертних оцінок, приклади практичного застосування якого розглядаються у роботах [1, 2]. Метод експертних оцінок було ефективно застосовано авторами роботи [1] для цілей воєнно-політичного прогнозування на прикладі Росії. Недолік, пов'язаний із суб'єктивністю оцінок було ліквідовано значною кількістю експертів та мозковим штурмом. В свою чергу у роботі [2] для дослідження можливих стратегій Росії для вирішення проблеми Південних Курил було сформовано ієрархічну цільову структуру досліджуваної проблеми та застосовано метод аналізу ієрархій Т. Сааті. Представлені результати використання методу аналізу ієрархій показали його достатню ефективність. Формування ієрархічних

цільових структур забезпечило чітку структурування досліджуваної проблеми, що дозволить суттєво спростити роботу аналітиків та експертів. Кількісні оцінки елементів "дерева цілей" надають можливість приймати більш обґрунтовані рішення при виборі тієї чи іншої стратегії поведінки в рамках конфлікту.

Робота [3] окреслює підхід до формалізації конфлікту, побудови його математичної моделі та дослідження конфлікту як багатовимірного, неоднозначного та слабо передбачуваного явища. Отримана модель дозволяє бачити тенденції розвитку та можливості досліджуваних процесів, а спеціальні процедури дають можливість уникнути необхідності попереднього накопичення статистичного матеріалу. Методику формування множини факторів воєнного конфлікту та системи взаємозв'язків між ними наведено у [4]. Тут викладено метод формалізації негативних та різнознакових взаємо-зв'язків між елементами когнітивних карт з використанням логіки антонімів. З метою відображення негативних та різнознакових зв'язків між елементами когнітивних карт введені відповідні оператори та сформовані їх основні властивості.

Можливість застосування нейронних мереж та, зокрема, Байєсівської нейронної мережі досліджується в роботах [5, 6]. У роботі [6] розглядаються пари учасників, і відповідно для кожної з пар пропонується така множина параметрів:

- демократія (для кожної з країн визначається за шкалою від мінус десяти до десяти, де десять – надзвичайна демократія, а мінус десять – повне самодержавство, після чого обирається мінімальне з двох значень);
- союзники (бінарна величина, що приймає значення 1, якщо члени пари пов'язані будь-якою формою військового союзу, та 0, якщо військовий союз відсутній);
- сусідство (має значення 1, якщо члени пари мають спільний кордон, та 0, якщо спільні кордони відсутні);
- відстань (обраховується як десятковий логарифм від відстані в кілометрах між столицями членів пар);
- основна сила (дорівнює одиниці для наддержав і 0 для всіх інших);
- спроможність (розраховується як десятковий логарифм від такої величини: сума загальної кількості населення, кількості людей в міських районах, промислового споживання енергії, виробництва чавуну і сталі, числа

активних військовослужбовців, військових витрат у доларах за останні п'ять років (від сильнішої держави до слабшої);

- залежність (сума значень експорту та імпорту між країнами, поділена на ВВП сильнішої країни; цей параметр є неперервним і вимірює рівень економічної незалежності менш економічно залежного члену пари).

У якості набору даних у роботі було використано політично значимі пари часів перед холодною війною (1885 – 1945 рр.) та під час холодної війни. Було встановлено, що при використанні чотирьох параметрів: демократії, союзників, спроможності і залежності; або з використанням тільки залежності або тільки можливості дозволяє уникнути всіх передбачених конфліктів. Використання тільки демократії дозволило уникнути 90% прогнозованих конфліктів, тільки союзників – 77%.

У даній роботі увагу буде приділено задачі моделювання та прогнозування можливості виникнення конфлікту для окремих сутностей, а не пар сутностей, на базі параметрів їх стану.

### **Постановка задачі**

Задача даної роботи полягає у розробці моделі військового конфлікту у формі байєсівської мережі, яка дозволить всебічно розглянути військовий конфлікт та ймовірності варіантів його розвитку на основі відповідної вибірки статистичних даних.

Побудова моделі процесу дасть можливість всебічно розглянути процес та спрогнозувати розвиток ситуації, що в результаті полегшить пошук оптимального управлінського рішення стосовно дій держави. Аналіз та моделювання подібної системи має на меті не тільки виявлення існуючих взаємозв'язків та залежностей, а і передбачення та попередження майбутніх конфліктів на основі результатів аналізу.

### **Побудова моделі**

При побудові моделі воєнного конфлікту на базі байєсівської мережі первинним і одним з найважливіших завдань є вибір параметрів (вузлів) мережі та

опис взаємозв'язків між ними. Будемо вважати можливими один з двох можливих результуючих вузлів:

- внутрішній конфлікт (вірогідність громадянського конфлікту всередині країни – сутності; цей вузол включатиме такі типи конфліктів, як повстання або революція, громадянська війна, відсутність внутрішнього конфлікту);
- зовнішній конфлікт (вірогідність конфлікту з іншою країною, включатиме наявність та відсутність зовнішнього конфлікту).

Складнішою задачею є раціональний вибір вихідних параметрів моделі. Виходячи з результатів дослідження [4], але з урахуванням переходу від дослідження пар сутностей до дослідження окремих сутностей, вважаємо доцільним обрати такі змінні для моделі:

- індекс людського розвитку;
- рівень інфляції;
- контроль корупції;
- демократія;
- наявність поточного конфлікту;
- рівень міграції;
- військові витрати (% GPD);
- рента природних ресурсів (% GPD).

Вищеназвані змінні мають таку інтерпретацію. Індекс людського розвитку – інтегральний показник, що розраховується щорічно для міждержавного порівняння і вимірювання рівня життя, грамотності, освіченості і довголіття, як основних характеристик людського потенціалу досліджуваної території. Він є стандартним інструментом при загальному порівнянні рівня життя різних країн і регіонів. Індекс публікується в рамках програми розвитку ООН в звітах про розвиток людського потенціалу [7].

Інфляція — надмірне, проти потреб товарообігу, збільшення кількості паперових грошей з їх наступним швидким знеціненням [8]. Вона проявляється зростанням загального рівня цін та зниженням купівельної спроможності грошей.

Індекс контролю корупції відображає в якому ступені державна влада здійснюється з метою особистої вигоди, в тому числі і дрібних та великих форм корупції, а також захоплення держави з боку еліт і приватних інтересів [9].

Розрахунковий показник дає оцінку країни за сукупним показником в одиницях стандартного нормального розподілу, тобто в діапазоні приблизно від -2,5 до 2,5. Він також вимірює силу і ефективність політики країни та інституційних границь (обмежень) для запобігання та боротьби з корупцією.

Демократія – політичний режим, за якого єдиним легітимним джерелом влади в державі визнається її народ. При цьому управління державою здійснюється народом – безпосередньо (пряма демократія), або опосередковано через обраних представників (представницька демократія) [10]. Іноді демократію визначають також як набір ідей і принципів, що стосуються свободи, власне, вона і являє собою інституціональну свободу. У формулюванні 16-го президента США Авраама Лінкольна, демократія – врядування «іменем народу, силами народу і для народу».

Рівень міграції – якісний показник рівня міграції у країні, що прийматиме одне з трьох значень: переважна імміграція, переважна еміграція або приблизна рівність цих процесів.

Військовий бюджет (або військові витрати) – також відомий як оборонний бюджет – це обсяг фінансових ресурсів, що виділяються країною для нарощування і підтримки збройних сил або інших методів, необхідних для оборонних цілей. Військові бюджети часто відображають, наскільки високою суб'єкт дослідження вважає ймовірність загрози проти нього, або величину агресії, яку він планує проявити. Розмір бюджету також відображає здатність країни фінансувати свої військові дії.

Оцінка природної ренти ресурсів розраховується як різниця між ціною ресурсного товару і середньою вартістю його здобуття. Це робиться шляхом оцінки світових цін одиниць конкретних товарів і віднімання оцінки середніх питомих витрат по видобутку (промислових витрат). Ці одиниці множиться на фактичну кількість товару, що країна видобуває, з метою визначення ціни кожного товару, як частки валового внутрішнього продукту (ВВП).

Як було вказано раніше, необхідно формалізувати та, за необхідності, дискретизувати усі обрані параметри. Для початку, визначимо можливі значення результуючих параметрів.

1. Внутрішній конфлікт = {наявний; відсутність внутрішнього конфлікту}.
2. Зовнішній конфлікт = {наявний; відсутність зовнішнього конфлікту}.

Далі перейдемо до формалізації вихідних параметрів.

1. Індекс людського розвитку буде поділено на категорії відповідно до числового значення коефіцієнта таким чином:
  - a. дуже високий рівень (0,8 - 1);
  - b. високий рівень (0,7 – 0,8);
  - c. середній рівень (0,55 – 0,7);
  - d. низький рівень (0 – 0,55).
2. Рівень інфляції визначається у відсотках відносно попереднього року і, відповідно, може мати як від’ємні значення (дефляція), так і додатні (інфляція). Відповідно до числового значення коефіцієнту, буде використано таку шкалу:
  - a. інфляція відсутня ( $< 0$ );
  - b. низький рівень (0 – 5%);
  - c. середній рівень (5% - 10%);
  - d. високий рівень ( $> 10\%$ ).
3. Індекс контролю корупції набуває значення в проміжку [-2,5; +2,5]. Виходячи з наявних статичних даних, ця змінна прийматиме такі значення:
  - a. дуже низький рівень ( $< -1$ );
  - b. незадовільний рівень ( $-1 - 0$ );
  - c. задовільний рівень (0 – 1,5);
  - d. високий рівень ( $> 1,5$ ).
4. Параметр «демократія» набуватиме одне з таких значень:
  - a. відсутня;
  - b. гібридний режим;
  - c. перехідний період;
  - d. повна демократія.
5. Наявність поточного конфлікту – бінарний параметр, що буде рівним 1 за наявності конфлікту та 0 за відсутності.
6. Рівень міграції включає в себе як показник відтоку людей з країни, так і притоку. Можливі такі його значення:
  - a. переважна імміграція (більше населення виїжджає, ніж в’їжджає);
  - b. переважна еміграція (більше населення в’їжджає, ніж виїжджає);
  - c. значення приблизно рівні між собою.

7. Показник військових витрат вираховується у доларах США або % від ВВП.

Пропонуються такі значення параметра на основі статистичних даних:

- a. позитивний (спостерігається збільшення частки військових витрат);
- b. негативний (спостерігається зменшення частки військових витрат);
- c. стабільний.

8. Рента природних ресурсів показує рівень видобування та продажу корисних копалин. Виходячи з історії, величина даного показнику на пряму впливає на вірогідність зовнішніх конфліктів. Можливі наступні значення параметра:

- a. дуже високий рівень ( $> 20\%$ );
- b. високий рівень (10 – 20%);
- c. середній рівень (5 – 10%);
- d. низький рівень (0 – 5%).

Також встановлено взаємозв'язки між вибраними змінними (вузлами) моделі. Граф зв'язків подано на рис. 1.

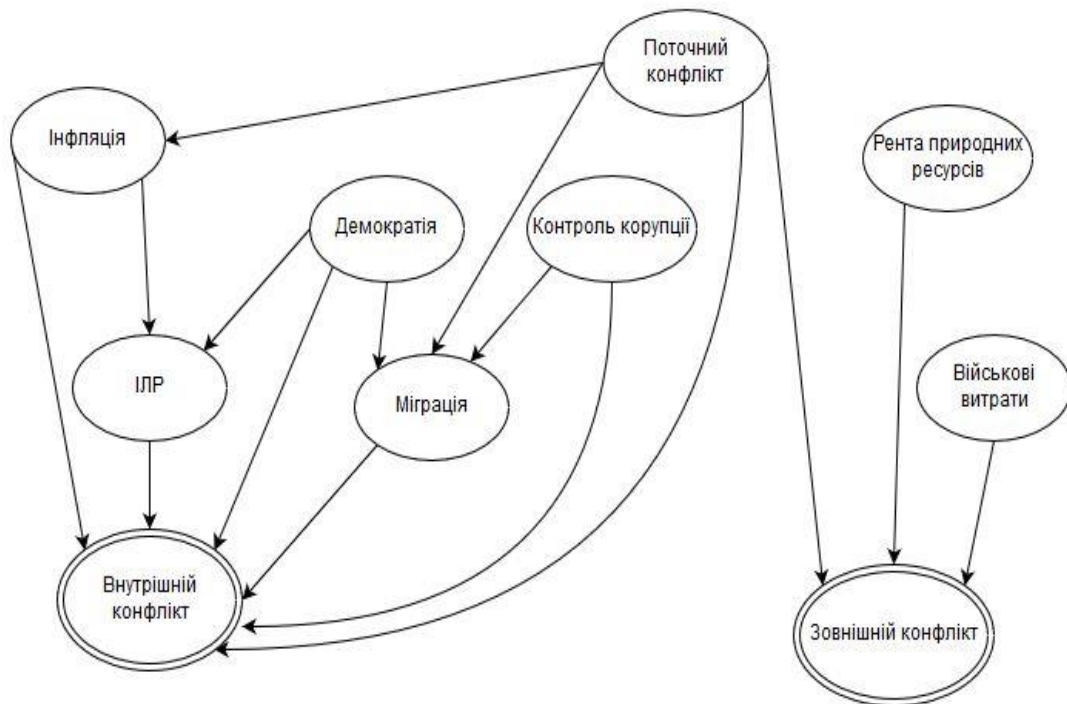


Рисунок 1 – Модель зв'язків параметрів мережі

## Моделювання та аналіз результатів



Для моделювання мереж байєса створено багато бібліотек на різних мовах програмування. У даному дослідженні використано бібліотеку SMILE та графічний інтерфейс GeNIe.

Статистичні дані для формування вибірки даних взято з відкритих джерел мережі Інтернет. Вибірка складається з даних про 51 країну з різними показниками (значеннями) стосовно обраних змінних моделі.

Для навчання параметрів мережі було використано такі алгоритми:

- Estimated Posterior Importance Sampling (EPIS);
- Adaptive Importance Sampling (AIS).

Аналіз чутливості мереж, що досліджувалися, дозволяє зробити такий висновок: найбільш чутливими змінними мережі виявилися такі: «Демократія», «Контроль корупції», «Рента природних ресурсів». Цей результат зберігається від мережі до мережі. Такий результат означає, що навіть невеликі зміни у даних вузлах істотно впливають на остаточний результат (внутрішній чи зовнішній конфлікт).

Далі проаналізуємо результати, отримані у процесі моделювання, для цільового параметра «Зовнішній конфлікт». З результатів першого експерименту видно, що за умови зростання державних витрат на військову сферу ймовірність зовнішнього конфлікту підвищується лише за наявності поточного військового конфлікту та низькому або високому значенні ренти природних ресурсів. Це можна пояснити тим, що модель не розрізняє зовнішні конфлікти за типом (нападаюча країна чи країна, що захищається). Виходячи з цього, отримані результати показують, що з ймовірністю трохи вищою за 0,50 зовнішні конфлікти країн, які вже приймають участь у зовнішньому конфлікті, будуть продовжуватися і надалі, причому цілком логічним припущенням буде вважати країни з низькою рентою природних ресурсів в даному випадку нападаючими, а з високою – тими, що захищаються.

За умови спадаючих витрат на військову сферу, ймовірність зовнішнього конфлікту коливається навколо 0,50 за умови низької та середньої ренти природних ресурсів. Це пояснюється тим, що зменшення військових витрат порівняно з попередніми роками збільшує ймовірність закінчення існуючих конфліктів через нестачу фінансування військових дій. Це зменшує ймовірність початку нових конфліктів зі сторони агресора, проте у той же час збільшує ймовірність стати жертвою конфлікту.

При збереженні сталого рівня витрат на військову сферу високою (0,80) є ймовірність зовнішнього конфлікту за умови низької ренти природних ресурсів та наявності поточного конфлікту – фактично, продовження існуючого конфлікту (для захоплення природних ресурсів) на поточному рівні. За інших умов, навіть за попередньої наявності конфлікту, ймовірність зовнішнього конфлікту майже напевно відсутня.

З результатів другого експерименту (після видалення з мережі найменш впливових зв'язків) можна зробити такі висновки. З представлених таблиць умовних ймовірностей параметра «Зовнішній конфлікт» видно, що, на відміну від попередньої моделі, за умови зростання державних витрат на військову сферу ймовірність виникнення зовнішнього конфлікту підвищується до 67% за високого значення ренти природних ресурсів, причому це значення вірне як за наявності поточного конфлікту, так і за його відсутності. За низької ренти природних ресурсів та відсутності конфлікту ймовірність зовнішнього конфлікту піднімається до 48%. Отже, ймовірність наявності зовнішнього конфлікту за другою моделлю вища за умови високої ренти природних ресурсів і, для таких країн, не залежить від наявності чи відсутності поточного конфлікту. Ці результати вважаємо більш достовірними за отримані у попередній моделі.

За умови спадаючих витрат на військову сферу, ймовірність виникнення зовнішнього конфлікту, так само як і в попередній моделі, наближається до 50% за умови низької та середньої ренти природних ресурсів.

При збереженні сталого рівня витрат на військову сферу, високою (75%) є ймовірність виникнення зовнішнього конфлікту за умови низької ренти природних ресурсів та, на відміну від попередньої моделі, відсутності поточного конфлікту. За інших умов, навіть за попередньої наявності конфлікту, ймовірність зовнішнього конфлікту майже напевно відсутня.

У зв'язку з кардинальною розбіжністю результатів першого та другого експерименту відносно параметра «Поточний конфлікт» при схожих результатах відносно інших параметрів, для подальших експериментів даний вузол було вирішено видалити з мережі.

За результатами третього експерименту, за умови зростання державних витрат на військову сферу та середньої ренти природних ресурсів ймовірність зовнішнього конфлікту майже повністю відсутня (< 1%). Це пояснюється відсутністю необхідності отримувати природні ресурси з інших країн. У той же

час, нарощування та підтримка військової сили гарантує захист від нападу на країну ззовні. Відсутність демократії за низької ренти природних ресурсів також показує низьку (10%) ймовірність зовнішнього конфлікту навіть за збільшення витрат на військову сферу. Для перехідної та повної демократії цей показник підвищується до 33% та 38% відповідно, а ймовірність зовнішнього конфлікту за гібридної демократії підвищується до 84%. За високої ренти природних ресурсів ймовірність зовнішнього конфлікту становить високі 67% для будь-якого типу демократії в країні, аналогічно попередньому експерименту.

За умови спадаючих військових витрат ймовірність зовнішнього конфлікту майже відсутня за умови високої та дуже високої ренти природних ресурсів (99% та 89% ймовірність відсутності конфлікту відповідно). Ситуація для середньої ренти природних ресурсів, тим не менш, значно відрізняється за зростаючих та спадаючих витрат на військову сферу. За спадання військових витрат, ймовірність зовнішнього конфлікту підвищується до 87% при відсутності демократії, спадаючи до 49% за перехідної та повної демократії та 12% за гібридної, оскільки відсутність якісної підтримки військової сфери робить країну вразливою перед нападниками. Відсутність демократії підвищує цю ймовірність, оскільки дозволить назвати воєнні дії «миротворчими» чи «такими, що допомагають населенню проти існуючого режиму».

За стабільних військових витрат ймовірність зовнішнього конфлікту дуже низька для середньої, високої та дуже високої ренти природних ресурсів. Лише за низької ренти при неповних різновидах демократії ймовірність зовнішнього конфлікту коливається між 50% для відсутності або гібридної та 88% для перехідної.

Отже, з отриманих під час проведення усіх трьох експериментів таблиць умовних ймовірностей для змінної «Зовнішній конфлікт» можна зробити такі висновки:

- середня рента природних ресурсів сама по собі мало впливає на ймовірність зовнішнього конфлікту, але ця ймовірність тим більша, чим менше грошей держава витрачає на підтримку та покращення у військовій сфері, і тим меншою, чим більш висока форма демократії створена у даній країні;
- збільшення військових витрат у країнах з низьким рівнем ренти ресурсів та перехідною або повною формою демократії підвищує ймовірність

зовнішніх конфліктів (більш ймовірно, зі сторони агресора), і досягає свого максимуму при гібридній демократії;

- стабільні військові витрати зазвичай означатимуть стабільність країни відносно військових конфліктів: підтримку існуючих та відсутність нових.

Розглянемо результати, отримані відносно параметра «Внутрішній конфлікт». Було проведено три аналогічні попереднім експерименти: з повною моделлю, після видалення найменш важливих зв'язків, після видалення вершини «Поточний конфлікт».

Через дуже велику кількість вхідних зв'язків параметра «Внутрішній конфлікт» у першому експерименті, повна таблиця умовних ймовірностей не була наведена. Тим не менш, можна зробити такі висновки.

По-перше, низький або дуже низький рівень контролю корупції у поєднанні з не дуже високим рівнем ІЛР та низьким рівнем демократії дає майже 90% ймовірність внутрішнього конфлікту. Це легко пояснюється невдоволеністю мешканців країни у рівні життя. По-друге, спостерігається низька ймовірність виникнення конфлікту за «перехідного» та «повного» показників демократії, а також – високого та дуже високого рівню ІЛР у поєднанні з низькою інфляцією та високим рівнем контролю корупції. Тим не менш, переважна імміграція у країну збільшує ймовірність виникнення внутрішнього конфлікту навіть за високих показників рівню життя та демократії. Причиною цьому є як невдоволеність корінних мешканців засиллям мігрантів, так і спробі мігрантів «встановити свій лад» у країні, до якої вони мігрували, а також ведення нечесного життя.

Виходячи з представлених у другому експерименті даних, висока ймовірність внутрішнього конфлікту за низького або середнього значення індексу людського розвитку призводить до переважної еміграції з країни і становить близько 83% для низького рівня ІЛР та 50% для середнього. За високого та дуже високого значення індексу людського розвитку, відповідно до другої моделі, ймовірність внутрішнього конфлікту майже відсутня, незначно зростаючи (33%) за високого ІЛР та переважної еміграції з країни (яка, зазвичай, пов'язана з низьким рівнем демократії та високим рівнем контролю корупції).

Третій експеримент показав, що ймовірність внутрішнього конфлікту за переважної імміграції дуже низька (17% та 4%) за високого та дуже високого значення індексу людського розвитку. За низького та середнього значення даного показника при переважній імміграції ймовірність внутрішнього конфлікту

збільшується (25% та 50%). За переважної еміграції з країни, ймовірність внутрішнього конфлікту також зменшується зі збільшенням значення індексу людського розвитку, від 79% за низького ІЛР до 25% за дуже високого. За приблизно рівного рівня імміграції та еміграції у країні, що досліджується, ймовірність виникнення конфлікту, навпаки, збільшується зі збільшенням рівня ІЛР, від 12% за низького рівня до 75% за високого, з падінням ймовірності до 30% за дуже високого ІЛР. Це пояснюється тим, що саме у країнах з високим рівнем ІЛР найбільше невдоволення викликають потоки мігрантів. Такі країни вже стали принадними для мігрантів з гірших за ІЛР країн, але, на відміну від країн з дуже високим ІЛР, ще недостатньо окріпли для того, щоб необхідність витрат ресурсів країни на мігрантів не впливала на рівень життя корінних мешканців.

Отже, з отриманих під час проведення усіх трьох експериментів таблиць умовних ймовірностей для параметра «Внутрішній конфлікт» можна зробити наступні висновки:

- високий рівень контролю корупції зменшує ймовірність внутрішнього конфлікту у країні;
- низький або дуже низький рівень контролю корупції у поєднанні з не дуже високим рівнем ІЛР та низьким рівнем демократії дає майже 90% ймовірність внутрішнього конфлікту;
- ймовірність виникнення конфлікту за «перехідного» та «повного» показників демократії, а також – високого та дуже високого рівню ІЛР у поєднанні з низькою інфляцією та високим рівнем контролю корупції майже напевно рівна нулю;
- переважна імміграція у країну збільшує ймовірність виникнення внутрішнього конфлікту навіть за високих показників рівню життя та демократії, особливо при середній та високій інфляції.

Далі було виконано прогнозування конфліктів для України на основі побудованої моделі. Ймовірність внутрішнього конфлікту було прогнозовано у 49%. Ймовірність зовнішнього – у 33%. Така низька ймовірність зовнішнього конфлікту (низька, оскільки реально ймовірність зовнішнього конфлікту в Україні дорівнює одиниці) пояснюється тим, що у контексті розробленої моделі нинішній конфлікт пояснюється слабко, адже він слабко пов'язаний як з рентою природних ресурсів, так і з рівнем демократії або військових витрат.

Подальші експерименти було спрямовано на пошук взаємозв'язків у вибірці даних за допомогою алгоритмів «вивчення структури», що дозволяє генерувати модель на основі мережі Байєса з вибірки статистичних даних.

Перший експеримент використовував Байєсів пошук. Було виявлено, що найбільш зв'язними є дані за такими змінними:

- «Контроль корупції»,
- «ІЛР»,
- «Демократія»,
- «Міграція».

При цьому «Контроль корупції» виявився параметром впливу відносно інших трьох параметрів. З отриманих таблиць умовних ймовірностей можна зробити такі висновки:

- підвищення рівня контролю корупції призводить до підвищення індексу людського розвитку;
- підвищення рівню контролю корупції означає перехід до вищої форми демократії;
- низький та дуже низький рівень контролю корупції призводить до переважної еміграції з країни, тоді як високий рівень – до переважної імміграції у країну.

Другий експеримент представив результат генерування мережі за методом РС. Отриманий граф включав у себе підграф, отриманий у попередньому експерименті. На відміну від попереднього методу, у згенерованій мережі були також наявні параметри «Індекс людського розвитку» та «Внутрішній конфлікт». Інфляція, за даною моделлю, впливає на ймовірність внутрішнього конфлікту прямо, тоді як міграція та демократія – опосередковано, через параметр «Індекс людського розвитку». Виходячи з таблиці умовних ймовірностей вузла «Внутрішній конфлікт», за високого та дуже високого показника індексу людського розвитку ймовірність внутрішнього конфлікту підвищується лише за низького та середнього рівню контролю корупції і становить близько 50%. За низького та середнього показника ІЛР ймовірність внутрішнього конфлікту коливається від 30% до 52%, падаючи до 10% за відсутності інфляції.

На даній моделі було отримано такий прогноз для України: ймовірність внутрішнього конфлікту становить 48%.

## **Висновок**

В дослідженні розглянуто проблему моделювання військових конфліктів, визначення та формалізації параметрів ймовірнісної моделі.

Результатом виконання дослідження стала побудована модель військового конфлікту на основі мережі Байєса. Модель було навчено на сформованій з реальних статистичних даних вибірці, проведено серію експериментів та проаналізовано отримані результати.

Побудована модель оцінила ймовірність внутрішнього конфлікту в Україні у 48%, що цілком відповідає реальності. Ймовірність зовнішнього було оцінено у 33%. Оскільки реально ймовірність зовнішнього конфлікту в Україні дорівнює одиниці, цей прогноз не є достатньо точним. Це пояснюється тим, що у контексті розробленої моделі нинішній конфлікт пояснюється слабко, адже він слабко пов'язаний як з рентою природних ресурсів, так і з рівнем демократії або військових витрат України.

Отже, модель можна вважати адекватною та придатною для практичного використання.

## **Література**

1. Богатирьов Е.Я., Макієв Ю.Д., Малішев В.П. Аналіз методів військово-політичного прогнозування // Стратегія громадянського захисту: проблеми та дослідження. – 2013. – Том 3. – № 2(5).
2. Абаєв Л.Ч. Моделювання криз і конфліктів у міжнародних відносинах: досвід практичного застосування // Проблеми національної стратегії. – 2012. – № 1. – С. 57–72.
3. Рома О.М., Толубко В.Б., Ленков С.В. Спосіб формалізації конфлікту та його системне моделювання // Ukrainian Information Security Research Journal. – 2009. – Том 11. – № 3(44).

4. Адаменко А. А. Методика формування множини факторів воєнного конфлікту та системи взаємозв'язків між ними // Наука і техніка Повітр. сил Збройн. сил України. – 2011. – № 1. – С. 141–145.
5. Beck N., King G., Langche Z. Improving Quantitative Studies of International Conflict: A Conjecture / Beck N. // American Political science review. – 2000. – Vol. 94. – №1. – Pp. 21–35.
6. Marwala, T. and M. Lagazio, Modelling and Controlling Interstate Conflict, Budapest / Marwala T. // Hungary: IEEE International Joint Conference on Neural networks. – 2004.
7. The Human Development concept [Електронний ресурс]. – Режим доступу: <http://hdr.undp.org/en/humandev/>
8. Словник української мови: в 11 томах. – 1973. — Том 4. — С. 42.
9. Lambsdorff J. How Corruption Affects Persistent Capital Flows // Economics of Governance. – 2003. – Vol. 4. – Pp. 229-243.
10. Етимологічний словник української мови в 7 томах. —1985. – Том 2. – С. 31.



## **Релевантність посилань в Інтернеті за різними критеріями на рівні задачі оптимального розміщення контекстної реклами з нелінійною складовою**

### **Вступ**

В даній статті розглядається проблема автоматизації оптимізації сайтів. Задача оптимізації сайтів є актуальною проблемою, а задача автоматизації цього процесу є дуже важливою.

Звичайні біржі (Sape, Blogun, MainLink, XAP) просто надають можливість користувачам-оптимізаторам закуповувати посилання централізовано, не витрачаючи час на самостійні пошуки. На прикладі біржі Sape можемо побачити, що процес закупівлі влаштований таким чином: є список посилань доступних до закупівлі з певними характеристиками. Метою користувача є закупівля певної кількості посилань з погляду на їх характеристики. В такому разі користувач знає суму, яку він згоден витратити на закупівлю, але не може це зробити гарантовано найкращим чином, робить це, покладаючись на інтуїтивні розрахунки [1].

Оптимізація сайтів складається з двох факторів – внутрішніх і зовнішніх. Стаття присвячена вивченню відомих методів підвищення релевантності сайту, впливаючи на зовнішні фактори, та підвищення ефективності існуючих методів за допомогою підвищення рівня автоматизації.

### **Постановка задачі**

Метою є створення біржі посилань, яка самостійно буде закуповувати необхідну кількість посилань з певними характеристиками таким чином, щоб ефективність від такої закупки була максимальною в сенсі підвищення релевантності сайту.

Суть автоматизації біржі посилань полягає в програмному підключенні до Google AdWords API. За допомогою сервісів даного API таких, як TargetingIdeaService і TrafficEstimatorService витягуються дані про доступні посилання – значення їх PR і тІЦ. Далі вирішується завдання оптимізації. У цьому завданні оптимізації цільової функцією, яку необхідно максимізувати, є лінійний функціонал (1):

$$R^2 \rightarrow R: L(n, r, t) \rightarrow \max \square, \quad (1)$$

де  $n$  – кількість посилань;

$r$  – значення PR;

t — значення тЦ.

I, відповідно, обмеження оптимізації – це обмеження сумарної ціни по вхідному бюджету і невід’ємність всіх змінних. Формалізована задача оптимізації являє собою задачу цілочисельного програмування [5-6].

### **Викладення теорії**

Біржа посилань – спеціалізований інтернет-ресурс, на якому відбувається купівля та продаж посилань. Сенс діяльності біржі посилань в наступному: на майданчику біржі веб-майстер продає місце під посилання на своєму сайті, а оптимізатор купує право розмістити посилання на свій сайт на сторінці веб-майстра [1].

Всі фактори, що впливають на положення сайту у видачі пошукової системи, можна розбити на зовнішні та внутрішні [2]. Робота по оптимізації включає роботу з внутрішніми чинниками (знаходяться під контролем власника веб-сайту) – приведення тексту і розмітки сторінок у відповідність з вибраними запитами, поліпшення якості та кількості тексту на сайті, стилістичне оформлення тексту (заголовки, жирний шрифт), поліпшення структури і навігації, використання внутрішніх посилань, а також зовнішніми чинниками.

В роботі розглянуто 2 основних зовнішніх фактори – PR і тЦ.

PR являє собою число від 1 до 10, яке ставиться у відповідність сторінці в Інтернеті. Його значення залежить від кількості і якості посилань, які ведуть з інших сторінок на даний документ. Чим більше посилань, тим більше буде значення PR.

Але також дуже важливим є те, яке значення статичної ваги має сторінка, з якої веде посилання на документ. По посиланню передається лише частина PR сторінки-донора.

Також якщо зі сторінки-донора проставлено декілька посилань, одна з яких на цей документ, то передана статична вага буде поділена між всіма цими посиланнями порівну.

Існує 2 шкали вимірювання PR [3]:

- Перший із варіантів є цілим дійсним числом і має лінійний характер, тобто збільшення цього значення буде пропорційно збільшенню статичної ваги, яка передається на даний документ по посиланням з інших ресурсів. Це число оновлюється в реальному часі і постійно враховується при ранжуванні.

- Другий варіант являється похідним від першого. Це значення називається тулбарним значенням і має діапазон від 0 до 10.

В таблиці 1 подано інтервали значень першої шкали у співвідношенні зі значеннями другої шкали [4].

Дійсне число	Тулбарне значення
0,00000001 – 5	1
5 – 25	2
26 – 125	3
126 – 625	4
625 – 3125	5
3125 – 15625	6
15626 – 78125	7
78126 – 390625	8
390626 – 1953125	9
1953126 – нескінченість	10

Таблиця 1 – Співвідношення шкал для PR.

Тулбарне значення являє собою  $\log_5$  від верхньої відповідного інтервалу.

тІЦ – тематичній індекс цитування, алгоритм пошукової системи Яндекс, призначений для визначення авторитетності сайтів шляхом підрахунку кількості ресурсів, які посилаються на нього та є схожої тематики. Використовується для оцінки релевантності (міри відповідності пошуковим запитам) сайтів в Яндекс.Каталозі, що дозволяє визначити значимість того чи іншого проекту.

Яндекс розраховує тІЦ сайтів за спеціальним алгоритмом, який враховує як подібність сайтів, що посилаються з тематики, так і рівень їх власного тІЦ і являє собою додатне ціле число (округлене додатне дійсне – аналогічне першій шкалі PR) [4].

#### **Приклад застосування методики**

Починаємо закуповувати посилання. Вхідними даними маємо поточний бюджет, ціну кожного посилання та значення PR і тІЦ для кожного посилання. Необхідно закупити

певну кількість посилань, беручи до уваги їх ціну, значення PR і тІЦ та обмежений бюджет. Значення PR і тІЦ беремо за допомогою підключення до AdWords API.

AdWords API – це набір веб-послуг, які використовуються для створення додатків, призначених для управління обліковими записами AdWords і пов'язаними з ними даними. Основною технологією обміну повідомленнями, використовуваної в AdWords API, є протокол SOAP версії 1.1. Крім того, для спрощення і прискорення розробки додатків надаються клієнтські бібліотеки [5-6].

Будемо шукати такий набір посилань, для яких сума їх PR і тІЦ буде максимальною. Для цього нам необхідно брати значення обох характеристик за лінійною шкалою. тІЦ є лінійним, PR має дві шкали, але доступне тільки тулбарне значення від 1 до 10. Відомо, що тулбарне значення являє собою логарифм з основою 5 від верхньої межі відповідного інтервалу по лінійній шкалі. Отже далі будемо брати до уваги верхню межу

$$5^{PR}$$

кожного інтервалу, тобто [3].

Для даної задачі доречно взяти алгоритм гілок та границь для розв'язку задач цілочисельного програмування з булевими змінними [7].

Розглянемо наступну задачу цілочисельного програмування.

Необхідно максимізувати функції

$$L_1 = \sum_{j=1}^n p_j x_j \rightarrow \max, \quad (2)$$

$$L_2 = \sum_{j=1}^n t_j x_j \rightarrow \max, \quad (3)$$

при обмеженнях

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m, \quad (4)$$

$$x_j \in \{0; 1\}, j = 1, \dots, n, \quad (5)$$

при чому  $a_{ij} \geq 0$ ,  $p_j \geq 0$ ,  $t_j \geq 0$ ,

де  $p_j = 5^{PR}$  – значення PR лінійною шкалою для  $j$ -го посилання;

$t_j$  – значення тПЦ для  $j$ -го посилання;

$x_j$  – булева змінна, яка показує, чи включене посилання до результуючого набору посилань для закупки (5);

$a_{ij}$  – ціна кожного  $j$ -го посилання;

$b_i$  – бюджет на закупку.

Метод гілок та границь застосовується для оптимізації однієї функції, тобто нам необхідно звести функції  $L_1$  і  $L_2$  до однієї, враховуючи їх різномірність. Для цього застосуємо метод зважених сум для нормованих неоднорідних критеріїв [8].

Формула зваженої суми критеріїв має вигляд:

$$F(L|w) = w_1 L_1 + w_2 L_2, \quad (6)$$

де  $w_1, w_2$  – відповідно ваги критеріїв – коефіцієнти важливості.

Щоб функції (2) і (3) були адитивними, їх необхідно нормувати. Нормовані функції (критерії)  $\tilde{L}_1, \tilde{L}_2$  безрозмірні і їх значення знаходяться в однакових межах від 0 до 1.

$$F(L|w) = w_1 \tilde{L}_1 + w_2 \tilde{L}_2, \quad (7)$$

де  $w_1, w_2$  – відповідно ваги критеріїв.

Таким чином ми отримали узагальнення формули для однорідних критеріїв. Але вона має недолік, який називається «Інтелектуальна помилка», яка породжена незалежністю процедур нормалізації критеріїв і вибору їх ваг [8].

Для вирішення цієї проблеми введемо адитивні функції цінності, які мають вигляд:

$$v(L) = v_1(L_1) + v_2(L_2), \quad (8)$$

Оскільки всі критерії необхідно максимізувати, то кожна часткова функція цінності  $v_i$  є зростаючою. Будемо розглядати лінійні функції з коефіцієнтами  $k_1$  і  $k_2$ .

Нехай  $L_i^*$  і  $L_i^*$  - найбільше та найменше значення критерію  $L_i(x)$  на множині варіантів

Функцію цінності можна подати у такому вигляді:

$$\hat{v}(L) = \frac{v_1(L_1^*)}{v_1(L_1^*) + v_2(L_2^*)} \frac{v_1(L_1)}{v_1(L_1^*)} + \frac{v_2(L_2^*)}{v_1(L_1^*) + v_2(L_2^*)} \frac{v_2(L_2)}{v_2(L_2^*)}, \quad (9)$$

$$\hat{v}(L) = w_1 \hat{v}_1(L_1) + w_2 \hat{v}_2(L_2), \quad (10)$$

де  $\hat{v}_i(L_i) = \frac{v_i(L_i)}{v_i(L_i^*)};$

$$w_i = \frac{v_i(L_i^*)}{\mu};$$

$$\mu = v_1(L_1^*) + v_2(L_2^*).$$

Отже цільова функція буде приймати вид:

$$L = \sum_{j=1}^n c_j x_j \rightarrow \max, \quad (11)$$

Метод гілок і границь використовує послідовно-паралельну схему побудови дерева можливих варіантів. Спочатку шукають допустимий план і для кожного можливого варіанту визначають верхню межу цільової функції. Гілки дерева можливих варіантів, для яких верхня межа нижче наближеного рішення, з подальшого розгляду виключаються.

Ефективність обчислювальних алгоритмів залежить від точності і простоти способу визначення верхньої межі можливих рішень і точності визначення наближеного рішення. Чим точніше спосіб визначення верхньої межі цільової функції, тим більше безперспективних гілок відсікається в процесі оптимізації. Однак збільшення точності розрахунку верхніх меж пов'язано із зростанням обсягу обчислень. Наприклад, якщо для оцінки верхньої межі використовувати симплекс-метод, то результат буде досить точним, але зажадає великого обсягу обчислювальної роботи. Розглянемо алгоритм розв'язку задачі (2) – (4) методом гілок і границь з простим і ефективним способом оцінки верхньої границі цільової функції.

Позначимо:  $U$  – множина змінних  $x_j$ ,  $S$  – множина фіксованих змінних, які увійшли в допустимий розв'язок;  $E_S$  – множина залежних змінних, які не можуть бути включеними в множину  $S$ , так як для них виконується нерівність:

$$a_{ij} > b_i - \sum_{x_j \in S} a_{ij} x_j, \quad i = 1, \dots, m; \quad (12)$$

Нехай  $G_S$  – множина вільних змінних, з яких проводиться вибір в  $S$  чергової змінної.

Розглянемо одновимірну задачу, де  $m = 1$  і задача (2) – (4) має тільки одне обмеження вигляду (3). Саме такий частковий випадок описує формулювання задачі оптимізації послань.

Позначимо  $h_{1j} = \frac{c_j}{a_{1j}}$  і припустимо, що  $x_j \in S$ ,  $j = (1, \dots, k < n)$  і виконуються умови

$$h_{1,k+1} \geq h_{1,k+2} \geq \dots \geq h_{1l}, l < n, \quad (13)$$

$$\sum_{j=k+1}^l a_{1j} > b_1 - \sum_{x_j \in S} a_{1j} x_j, \quad (14)$$

$$\sum_{j=k+1}^{l-1} a_{1j} \leq b_1 - \sum_{x_j \in S} a_{1j} x_j. \quad (15)$$

Умови (14), (15) означають, що в множину S без порушення нерівності (3) можна додатково внести елементи  $x_{k+1}, x_{k+2}, \dots, x_{l-1}$ . При введенні в множину S елементів  $x_{k+1}, x_{k+2}, \dots, x_l$  нерівність (3) не виконується.

Для визначення верхньої границі розв'язку може бути використано вираз:

$$H_S = \sum_{x_j \in S} c_j x_j + L'_S, \quad (16)$$

$$L'_S = \sum_{j=k+1}^{l-1} c_j + h_{1l} \Delta b_1, \quad (17)$$

$$\Delta b_1 = b_1 - \sum_{x_j \in S} a_{1j} x_j - \sum_{j=k+1}^{l-1} a_{1j}. \quad (18)$$

З умов (13) – (15) випливає, що  $L'_S$  не менше максимального значення величини

$\sum_{x_j \in G_S} c_j x_j$  при обмеженнях

$$\sum_{x_j \in G_S} a_{1j} x_j \leq b_1 - \sum_{x_j \in S} a_{1j} x_j = b'_1, \quad x_j \in \{0; 1\}, x_j \in G_S. \quad (19)$$

Пояснимо процес визначення  $L'_S$  графіком (рисунок 1).

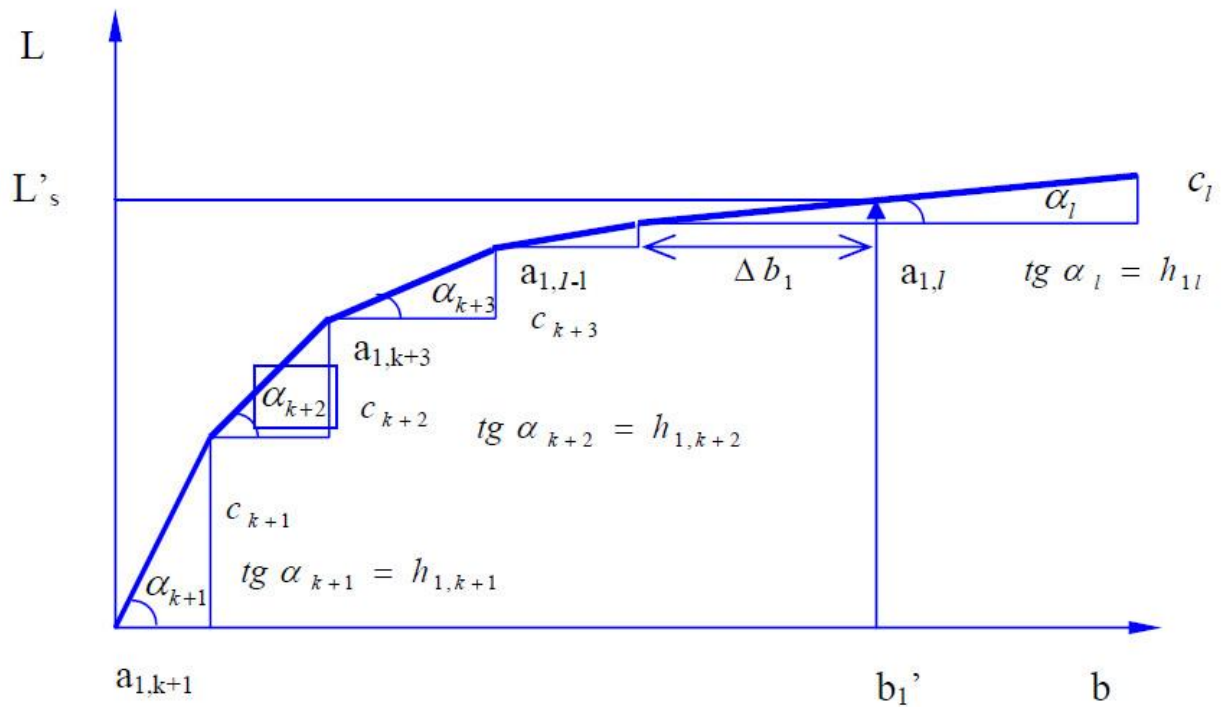


Рисунок 1 – Графічна інтерпретація побудови верхньої границі

Будуємо кусково-лінійну функцію  $L(b)$  зі спадаючим значенням градієнта  $h$ . Обчислюємо  $b_1'$  і за графіком знаходимо  $L'_S$ .

Вибір чергової змінної для включення в множину  $S$  проводиться за допомогою умови

$$h_{1r}(x_r) = \max_{x_j \in G_S} h_{1j}(x_j) \quad (20)$$

Для вибраної змінної  $x_r$  визначаються величини  $H_S(x_r)$  і  $H_S((x_r)')$ , тобто в  $S$  включається  $x_r = 1$  або  $x_r = 0$ .

Якщо в процесі розв'язку трапляється, що в множині  $G_S$  немає елементів,

Які можуть бути введені в множину  $S$  без порушення обмеження

(3), то отриманий розв'язок

$$L_S = \sum_{x_j \in G_S} c_j x_j \quad (21)$$

приймається в якості першого наближеного розв'язку  $L_0$ .



Всі вершини дерева можливих варіантів, для яких виконується умова  $H_S \leq L_0$ , з подальшого розглядання виключаються.

Із гілок, які залишились, вибирається гілка з максимальним значенням  $H_S$ , і процес пошуку оптимального варіанту продовжується. Якщо в процесі розв'язку буде знайдено:

$$L_S = \sum_{x_j \in G_S} c_j x_j > L_0, \quad (22)$$

то отриманий розв'язок приймається в якості нового наближеного результату. Обчислювальна процедура завершується, якщо для всіх гілок, які залишились, виконується умова  $H_S \leq L_0$  [7].

Для вимірності вектора  $x$ , звісно, є обмеження. При певній кількості посилань алгоритм починає працювати дуже довго, тому при великій кількості зареєстрованих посилань необхідно брати не всі доступні посилання на вхід алгоритму, а лише їх частину. Для цього достатньо розділити всі посилання на групи за однаковим значенням PR і з кожної вибрати однакову кількість посилань  $m$  так, щоб у суммі отримати таку кількість посилань  $n$ , яку можна обробити за реальний час. Ця кількість встановлюється емпіричним шляхом. Вибір проводиться за принципом найменших цін в групі. Далі від введеної суми береться частина:

$$p_i = p^* m, \quad (23)$$

де  $p^*$  – середнє арифметичне серед цін отриманої вибірки.

Для отриманої суми прогоняється метод гілок та границь на цій вибірці. Після закупки відхилені посилання повертаються до інших доступних, решта грошей додається до залишеної суми і ітерація повторюється. На введену суму накладається обмеження

$$P_{input} \leq P^* t, \quad (24)$$

де  $P^*$  - середнє арифметичне по всім цінам;

$t$  - можлива кількість ітерацій з погляду на час однієї ітерації.

Конкретно в програмному продукті реалізовані такі константи: кількість посилань у вибірці  $n = 200$ , максимальна кількість ітерацій  $t = 10$ . Цього більш ніж достатньо, адже зазвичай посилання закупаються по 50-80 посилань за один раз.

## Обговорення отриманих результатів

Розглянемо процес вибору посилань для закупівлі на прикладі розробленого веб-додатку. Для максимальної ефективності результату закупівлі користувач повинен тільки ввести необхідну суму коштів, яку він планує витратити на оптимізацію і система самостійно запропонує необхідну кількість посилань з найкращими характеристиками. На приклад, як можна побачити на рисунку 2, є 10 посилань з відповідними значеннями PR і тЦ:



Link Address	PR	тЦ	Price	Valid From	Valid To	Amount Of Places
www.google.com.ua	7	19000	220.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://www.wikipedia.org/	9	150000	100.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
www.facebook.com	9	330000	80.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
http://vk.com	9	360000	50.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://ipsa.kpi.ua	5	130	1.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	3
http://studsovet.ipsa.kpi.ua/studsovet	3	20	2.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://www.gismeteo.ua	6	5700	30.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
http://www.gismeteo.ua/weather-kyiv-4944	0	5700	9.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://vk.com/id47393806	0	360000	1.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
developers.google.com	7	2400	130.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1

Рисунок 2 – Вибір посилань

На рисунку 3 можна побачити список посилань, які система пропонує для купівлі.

HOME  
MY BUDGET  
LOG OUT



### Bought Links

Link Address	PR	тИЦ	Price
http://www.gismeteo.ua	6	5700	30.0
http://ipsa.kpi.ua	5	130	1.0
http://vk.com	9	360000	50.0
www.facebook.com	9	330000	80.0
http://studsovet.ipsa.kpi.ua/studsovet	3	20	2.0
www.google.com.ua	7	19000	220.0

### On Account

4617.0

total: 383

### Рисунок 3 – Закуплені посилання

Посилання і їх кількість обрані автоматично з урахуванням значень релевантності і допустимих витрат, це гарантує кращий результат, ніж вибір посилань користувачем власноруч.

#### **Висновок**

В результаті дослідження було реалізовано автоматизовану біржу посилань. Реалізована біржа дає змогу користувачу підібрати посилання для закупівлі найоптимальнішим чином в сенсі максимального поліпшення оптимізації сайту. Оптимізація проводиться за зовнішніми чинниками і є багатокритеріальною. Для розв'язання поставленої задачі використовувалися такі математичні методи: метод зважених сум нормованих неоднорідних критеріїв та метод гілок та границь для розв'язку задач цілочисельного програмування з булевими змінними. Також до уваги був прийнятий факт нелінійного росту тулбарного значення PR, що робить оптимізацію більш точною.

Новизною реалізованої біржі у порівнянні з існуючими є автоматизація закупки по обмеженому бюджету, тобто користувач повинен лише задати кількість коштів, які він

згоден витратити на закупку посилань і натиснути кнопку, далі система вирішить питання оптимальної закупки самостійно.

### **Література**

1. Байков В. Д. Интернет. Поиск информации. Продвижение сайтов / – СПб: БХВ-Петербург, 2000. – 288с.
2. Ашманов И. Продвижение сайта в поисковых системах / Ашманов И., Иванов А. – Вильямс, 2007. – 304 с.
3. Вигерс К. И. Разработка требований к программному обеспечению. / – М.: Русская редакция, 2004. – 535 с.
4. Яковлев О. О. Раскрутка и продвижение сайтов: основы, секреты, трюки. / – СПб: БХВ-Петербург, 2007. – 336 с.
5. Исследование: оценка качества ссылочного профиля [Электронный ресурс]. – Режим доступа до ресурсу: <http://pro100blogger.com/2013/05/Google-TrustRank.html>.
6. AdWords API руководство для разработчиков [Электронный ресурс]. – Режим доступа до ресурсу: <https://developers.google.com/adwords/api/?hl=ru>.
7. Ковалев М. М. Дискретная оптимизация (Целочисленное программирование). / – Мн: БГУ, 2003. – 191 р.
8. Метод взвешенной суммы критериев в анализе многокритериальных решений: PRO ET CONTRA./ – М., 2013. - 48 с.

**ЗАСТОСУВАННЯ РЕАЛЬНИХ ОПЦІОНІВ В УПРАВЛІННІ РИЗИКАМИ  
ІНВЕСТИЦІЙНОГО ПРОЕКТУ БУДІВНИЦТВА ВЕЛИКОЇ КІЛЬЦЕВОЇ  
ДОРОГИ НАВКОЛО М. КИЇВ**

**В.В. Кобера, І.Ю. Каніовська**

**Вступ**

Українські компанії поступово починають цікавитися методом реальних опціонів. Впровадження методу реальних опціонів дозволить фінансовим менеджерам швидше реагувати на зміну ринкової кон'юнктури, фокусувати увагу на різних сценаріях і не упустити нові можливості для розвитку. В теж час потрібен певний досвід застосування та інтерпретації результатів, крім того застосування реальних опціонів веде до істотної зміни культури прийняття інвестиційних рішень [1].

**Теоретична значимість** дослідження полягає в розвитку аспектів теорії державних і корпоративних фінансів; обґрунтуванні доцільності використання реальних опціонів в сфері управління фінансовими ризиками, збільшенні наукових знань теоретичного характеру, розкривають сутність, зміст і роль інвестицій в інфраструктурні проекти.

**Практична значимість** отриманих результатів полягає в доказі ефективності застосування моделі управління фінансовими ризиками з використанням методу реальних опціонів при оцінці інфраструктурних проектів.

**Постановка завдання**

Довести актуальність використання методу реальних опціонів для аналізу ризиків управління проектами та застосувати даний метод для проекту будівництва Великої кільцевої дороги навколо міста Київ.

### **Викладення теорії**

**Інвестиційний проект** являє собою комплекс взаємопов'язаних заходів, що передбачає певні вкладення капіталу протягом обмеженого часу з метою отримання доходів у майбутньому.

Інвестиційні проекти можуть бути дуже різноманітними за своєю спрямованістю і змістом об'єкта інвестування. Останнім часом особлива увага приділяється інфраструктурним проектам, впровадження яких забезпечує економічний і соціальний розвиток інфраструктури організації, громадського інституту, регіону і держави.

Інвестиційна діяльність пов'язана з високими ризиками і невизначеністю майбутнього розвитку проекту, що реалізовується. У зв'язку з цим необхідність створення оптимальної структури управління ризиками є одним із головних завдань для успішного виконання планів інвестування. Однак на практиці виникає безліч спірних моментів. Пов'язано це в першу чергу з відсутністю цілісної теорії управління ризиком, неоднозначністю використання різних методів оцінки ризику при побудові системи управління і нездатністю компаній в короткі терміни адаптуватися до нових умов на ринку.

Під **ризиком** розуміється фактор невизначеності, здатний зробити негативний вплив на реалізацію проекту в зв'язку зі зміною його параметрів у порівнянні з прогнозованими значеннями (збільшення вартості проекту, збільшення термінів введення в експлуатацію, зниження показників ефективності проекту і інші).

**Невизначеність** - це неповне або неточне уявлення про значення різних параметрів в майбутньому, породжуваних різними причинами і перш за все неповнотою або неточністю інформації про умови реалізації рішення, в тому числі пов'язаних з ними витрати і результати.

Одним з інструментів управління ризиками в умовах невизначеності є укладення між двома сторонами термінових контрактів, що передбачають право

однієї зі сторін відмовитися від їх виконання при настанні несприятливих умов. Для цього використовуються опціонні контракти.

**Опціонний контракт** (або просто опціон) – це цінний папір, який дає одній стороні угоди право придбати (або продати) базовий актив за фіксованою ціною протягом встановленого періоду часу (до встановленого терміну погашення), а іншу сторону зобов'язує продати (або, відповідно, придбати) цей актив за цією ціною в цей термін [3].

Основна привабливість опціонів для покупця пояснюється тим, що йому заздалегідь відомий максимально можливий розмір збитків – це величина премії, сплаченої за опціон, тоді як потенційний прибуток теоретично необмежений – в разі значного зростання ціни базових акцій в період дії опціону, покупець може розраховувати на високий прибуток. Особливо привабливими є опціони на акції, ринок яких відрізняється різкими та сильними ціновими коливаннями.

**Реальний опціон** – опціон, базовим активом за якого є реальний актив: завод, запаси нафти, машини, виробничі інвестиції і т.д.

**Метод реальних опціонів** розглядає методологію хеджування портфельів, які включають не тільки цінні папери, а й інвестиційні проекти, які передбачають майнові вкладення в різні активи. Реальні активи - це будь-які речові активи, такі, як дорогоцінні метали і камені, стратегічна сировина, колекційні предмети.

Реальні опціони використовуються при аналізі бізнес-планів по інвестиційним проектам, що передбачають вкладення в реальні активи.

Великі інфраструктурні проекти часто характеризуються високою ступенем невизначеності на етапі попередньої розробки, істотним ризиком прийняття інвестиційних рішень, множинністю джерел фінансування, обмеженістю фінансових ресурсів, а також наявністю великих обсягів інформації, яка в повному обсязі формалізована і не може бути врахована при застосуванні тільки кількісних методів. Тому необхідна розробка комплексної моделі, що передбачає використання математичних і емпіричних методів оцінки фінансового ризику і що дозволяє приймати обґрунтовані управлінські рішення.

Використання механізму державно-приватного партнерства в процесі розробки та реалізації інфраструктурних проектів створює таку ситуацію, де проект може бути здійснений єдиним учасником, виявленим на конкурсній основі, через юридичних обмежень або інших бар'єрів на вхід для конкурентів. У цьому випадку зміна вартості проекту з плином часу дає певні переваги.

Незалежно від того, що проект може бути реалізований тільки одним замовником, зміни у вартості проекту з плином часу можуть повідомити йому характеристики реального опціону на відстрочку. Якщо протягом кількох наступних років приведена вартість грошових потоків збільшується через зміни або вартості проекту, або ставки дисконтування, то проект, що має негативну чисту наведену вартість, може стати привабливим.

За результатами традиційного аналізу інвестиційне рішення приймається на підставі того, що організації не повинні інвестувати в проект з негативною чистою приведеною вартістю. Але це не означає, що реальні опціони на даний проект не мають вартості. Проекти, що наразі мають негативну чисту теперішню вартість, в майбутньому здатні володіти позитивною чистою приведеною вартістю, ймовірність цього є прямим результатом зміни показників спрямованості грошових потоків від проекту. Тому використання реальних опціонів дає можливість управляти фінансовим ризиком з урахуванням змін майбутньої вартості і прибутковості інфраструктурного проекту, які не можуть бути враховані в умовах використання традиційних методів дисконтування грошових потоків та розрахунків економічного ефекту.

Традиційні методи оцінки активів слабо піддаються змінам в ході реалізації проекту і незручні для прийняття рішення в разі зміни кон'юнктури ринку, або відстрочки реалізації інвестицій, в той час як метод реальних опціонів видається більш динамічним і дозволяє швидко реагувати на нові умови, які склалися як в економіці, так і всередині підприємства. Багато інвестиційні плани потребують перегляду з плином часу, так як не всі рішення прийняті на початку проекту є ефективними. У зв'язку з цим, метод реальних опціонів є сучасним інструментом щодо прийняття управлінських рішень та оцінки проектів. Однак одним з ключових моментів є процедура розрахунку і виявлення можливості застосування реального опціону в проекті.

Зазвичай ринкові моделі засновані на ідеї створення імітує портфеля і на арбітражі. У таких моделях, як модель Блека-Шоулза і Біноміальна модель, робиться припущення, що, використовуючи базовий актив і безризикові позики або позики, можна створити портфель-імітатор з грошовими потоками, які ідентичні платежах за опціоном. Крім того, в цих моделях допускається, що оскільки інвестори здатні створювати безризикові позиції шляхом покупки опціону і продажу імітує портфеля, то вони повинні торгуватися за тією ж ціною.



Якщо цього не відбувається, то інвестори отримують можливість створити безризикові позиції і піти з гарантованим прибутком, в чому і полягає суть арбітражу.

В рамках інфраструктурних проектів слід направляти інвестиції в акції компаній, що мають невеликий капітал, але в перспективних секторах економіки. Зарубіжний досвід свідчить про те, що слід направляти інвестиції в інфраструктурні проекти з довгостроковим періодом окупності.

Коли термін опціону зростає, вартість опціону збільшується. Державі слід відігравати активну роль в управлінні і прагнути отримати частку участі в тих компаніях, в які проводиться інвестування, - можливо, за допомогою конвертації боргових зобов'язань в пайові цінні папери.

Застосування методики реальних опціонів до оцінки інвестиційних проектів є обґрунтованим у випадку виконання таких умов:

- 1) результат проекту має високий ступінь невизначеності;
- 2) менеджмент компанії здатний приймати гнучкі управлінські рішення при появі нової інформації щодо проекту;
- 3) фінансовий результат проекту багато в чому залежить від рішень, які приймаються менеджерами [3].

В роботі розроблено модель управління фінансовими ризиками інфраструктурного проекту з використанням методу реальних опціонів на підставі комплексного використання формули Блека-Шоулза, методу Датар-Метьюз і застосування методу нечітких множин.

Найкращим чином формула Блека-Шоулза підходить для оцінки простих реальних опціонів, що мають єдине джерело невизначеності і єдину дату рішення. Використання методу

Вартість реального опціону для оцінки справедливої вартості активу, в свою чергу, розраховується на основі моделі Блека-Шоулза – формула (1).

$$\begin{aligned} P_0 &= X * (e^{-r*T}) * N(-d_2) - S_0 * N(-d_1) \\ C_0 &= S_0 * N(d_1) - X * (e^{-r*T}) * N(d_2), \end{aligned} \quad (1)$$

$$\text{де } d_1 = \frac{\ln \frac{S_0}{K} + \left(r + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t},$$

P - ціна акції або вартість активів компанії, S - ціна виконання або номінальна вартість боргу, t - період до виконання опціону (дюрація боргу), r - безризикова процентна ставка, яка відповідає дюрації,  $\sigma$  - середньоквадратичне відхилення вартості активу.

Датара-Метьюз замінює формулу Блека-Шоулза, а також значно спрощує процедуру розрахунку за допомогою імітаційного моделювання очікуваної чистої приведеної вартості проекту. Це робить спосіб знаходження вартості реального опціону наближеним до більш широкого кола користувачів. Перевага методу Датара-Метьюз для оцінки реальних опціонів полягає в тому, що на відміну від інших способів в даній моделі не потрібно знаходження значення середньоквадратичного відхилення і вартості проекту на поточний момент. У методі Датара-Метьюз використовується дві ставки дисконтування:

- Необхідна ставка прибутковості для тих грошових потоків, які схильні до ринкових ризиків;
- Безризикова ставка для тих інвестицій, які досить надійні і компанія може їх контролювати. Застосування двох різних ставок дозволяє скорегувати ризики проекту.

Відповідно до даного методу вартість реального опціону розраховується за такою формулою (2):

$$C_0 = E_0 [\max (S_0 e^{-\mu t} - X T e^{-rt}, 0)], \quad (2)$$

де  $ST$  - випадкова змінна, що представляє собою операційний дохід в момент  $T$ , яка дисконтується за необхідною ставкою прибутковості  $\mu$ ;  $XT$  - випадкова змінна, що представляє собою ціну виконання опціону (витрати на запуск проекту), яка дисконтується за безризиковою ставкою  $r$ ;  $C_0$  - вартість

реального опціону;  $E_0$  - середнє значення. У разі якщо  $\mu > r$ , то опціон не

розташований до ризику, а якщо  $r < \mu$ , то опціон схильний до ризику.

Перед початком імітаційного моделювання необхідно продисконтовані операційний дохід по необхідній ставкою прибутковості і інвестиції (витрати на запуск проекту) за безризиковою ставкою на дату початку проекту. Далі проводиться побудова ймовірності розподілу NPV за допомогою методу Монте-Карло. негативні результати розподілу не враховуються для розрахунку опціону, так як в цьому випадку витрати на запуск проекту перевищують доходи і менеджер змушений відхилити проект. Середнє значення позитивної частина розподілу буде вартість опціону.

Логічне продовження методи Датари-Метьюз є модель, представленої М. Коллана. На відміну від методу, запропонованого В. Датар і С. Метьюз, в даному методі використовується теорія нечітких множин, а не теорія ймовірності. В основі методу також лежить побудова розподілу NPV проекту. У методі нечітких множин вартість реального опціону розраховується за допомогою ймовірності розподілу NPV проекту, яке будується з використанням трьох або чотирьох сценаріїв грошових потоків, тобто при використанні трьох сценаріїв застосовується трикутне нечітке число, а при чотирьох - трапециевидное нечітке число. Таким чином, побудова розподілу здійснюється без імітаційного моделювання. Вартість реального опціону в даному методі визначається за такою формулою (3):

$$C = A(Pos) A(Pos) + A(Neg) \times E[A+], \quad (3)$$

де С - вартість реального опціону; А (Pos) - значення позитивної частини нечіткого розподілу; А (отр) - значення негативної частини нечіткого розподілу; E [А +] - середнє значення в позитивній частині нечіткого розподілу.

Таким чином, вартість реального опціону знаходиться шляхом розрахунку позитивної частини розподілу, яка ділиться на всю площу трикутного розподілу і потім отримане значення множиться на можливе середнє значення позитивної частини розподілу. Метод розрахунку, заснований на нечітких множинах, дає додаткові можливості компаніям з технологічними та інноваційними перевагами, оскільки дозволяє враховувати взаємозв'язок між ризиком і можливостями, які відкриваються перед підприємством.

Види опціонів та ризиків пов'язаних з ними ризиків представлена в таблиці 1 [6].

Таблиця 1. Види реальних опціонів і пов'язаних з ними фінансових ризиків

Види опціонів	Характеристика	Види ризиків
Опціон на припинення проекту (Вихід інвестора з проекту)	Можливість залишити діючий проект при несприятливому збігу подій і реалізувати його ліквідаційну вартість, яку можна розглядати як ціну виконання опціону "Put"	Ризик недооцінки вартості активу; Ризик упущеної вигоди
Опціон на відстрочку реалізації проекту	Можливість на відстрочку інвестицій в розвиток матеріальної бази, що еквівалентно опціону "Call"	Ризик упущеної вигоди; Ризик зростання витрат; Інфляційний ризик
Опціон на розширення проекту	Можливість розширення масштабів проекту, що формально еквівалентно опціону "Call"	ризик недооцінки вартості активів; ризик зростання витрат; ризик негативної прибутковості

Опціон на скорочення інвестицій в проект	Можливість скорочення масштабів проекту, що формально еквівалентно опціону "Put"	ризик упущеної вигоди; ризик негативної прибутковості
Опціон на перемикання, переклад інвестицій в інший проект	Портфель опціонів, що складається з "Call" і "Put" опціонів	ризик упущеної вигоди; ризик зростання витрат

### **Застосування методики**

Відповідно до виникаючими ризиками пропонується використовувати інструментарій реальних опціонів для мінімізації небажаних фінансових наслідків реалізації інфраструктурних проектів. Схема дій з управління фінансовими ризиками представлений на рис. 1.



Представлена модель наочно свідчить, що впровадження управлінської гнучкості щодо майбутніх можливих рішень впливає на вартість і привабливість проекту. Однак використання традиційних підходів, заснованих на методі дисконтування грошових потоків, не дозволяє проявити управлінську гнучкість при аналізі ефективності інвестиційного проекту. Рішенням даної проблеми може стати застосування методу реальних опціонів, що дозволить учасникам інфраструктурного проекту більш ефективно реагувати на зміну ринкової кон'юнктури і не упустити можливостей подальшого розвитку.

Послідовність етапів їх застосування представлена на рисунку 2

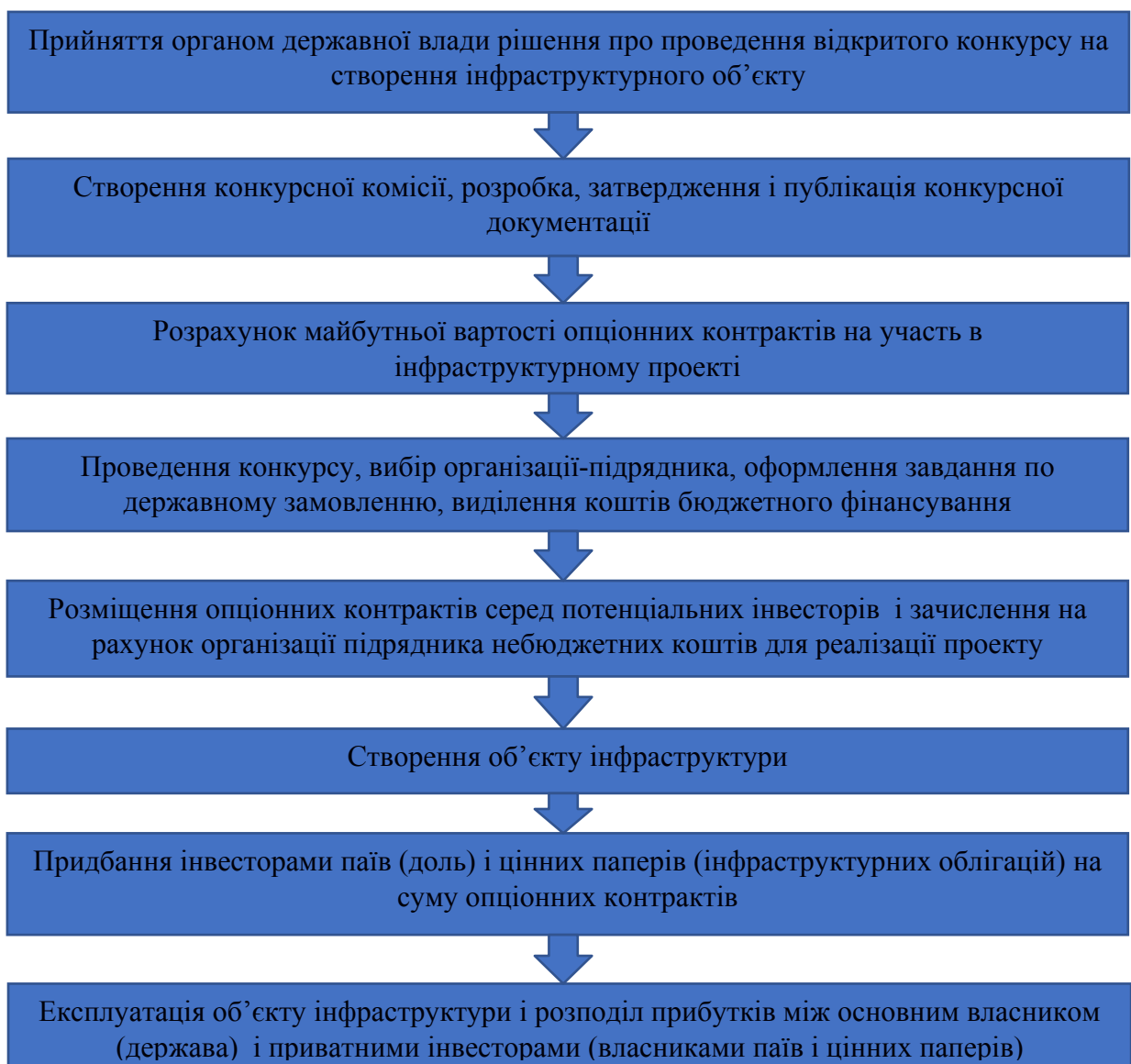


Рисунок 2 - Схема застосування реальних опціонів в інфраструктурних проектах з державною участю.

## Обговорення результатів методу

Даний метод був застосований для проекту будівництва Великої кільцевої дороги навколо Києва.

Метою побудови моделі було продемонструвати як впровадження управлінської гнучкості щодо майбутніх можливих рішень впливає на вартість проекту. Ігнорування гнучкості може спричинити за собою прийняття невірних рішень.

Завданням дослідження було виявити найбільш підходящі види реальних опціонів для такого типу проектів.

Етап будівництва проекту ділиться на 4 ділянки: перша ділянка будується за рахунок коштів держави, а решта 3 за рахунок приватних інвестицій. Для обґрунтованого прийняття управлінського рішення використовується метод реальних опціонів.

На етапах реалізації проекту обрано опціони, представлені на рисунку 3.

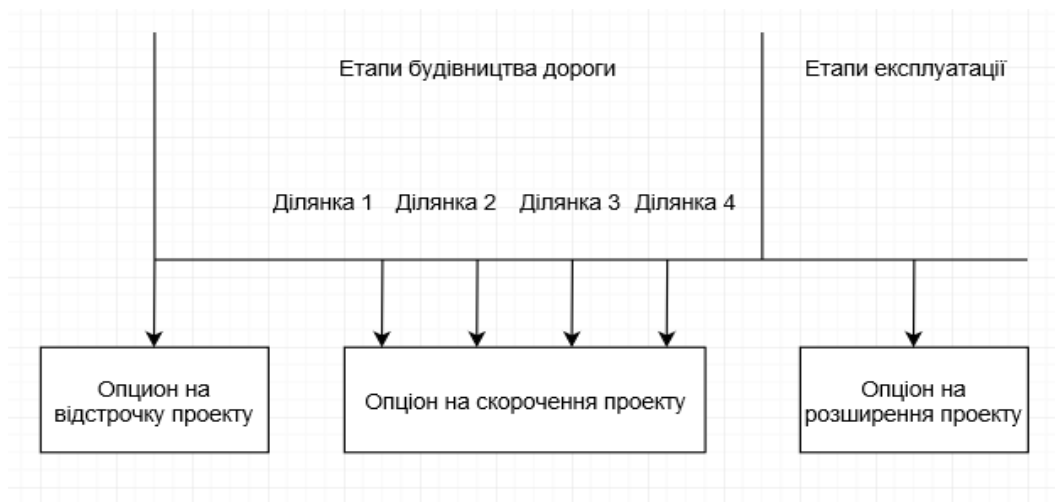


Рисунок 3. Види реальних опціонів на етапі організації проекту

Ділянка 1 - з'єднує дорогу Київ – Ковель – Ягодин та дорогу Київ – Чернігів – Нові Яриловичі.

Ділянка 2 - з'єднує дорогу Київ – Чернігів – Нові Яриловичі та дорогу Київ – Харків – Довжанський.



Ділянка 3 - з'єднує дорогу Київ – Харків – Довжанський та дорогу Київ – Одеса.

Ділянка 4 - з'єднує дорогу Київ – Одеса та дорогу Київ – Ковель – Ягодин.

Ми вважаємо, що найбільш підходящий вид опціону є на скорочення проекту на етапі будівництва ділянок, на розширення проекту на етапі експлуатації і опціон на відстрочку проекту.

Опціон на відстрочку обраний в силу того, що реалізацію проекту призупинили на час, через ризик невизначеності, тому даний проект буде досить підходящим прикладом для дипломної роботи. Можливі різні причини відстрочки проекту, але найбільш реалістична - фінансування будівництва може бути відкладено на більш ранніх періодах отримання прав на контракт з будівництва, ніж в більш пізній час; обсяги фінансування будуть зростати в міру зростання втрат в приведеної вартості внаслідок фактору очікування.

Теоретично можна передбачити опціон на відмову від проекту до початку старту будівництва ділянки 1, але це є економічно недоцільним зважаючи на специфіку проекту і того факту, що замовником по суті виступає держава.

Розраховано опціони було за алгоритмом, представленим на рисунку 4.



Рисунок 4. Алгоритм розрахунку вартості реального опціону проекту

Аналіз даного проекту свідчать, що, незважаючи на високу вартість будівництва і тривалий термін окупності інфраструктурного проекту Великої кільцевої дороги, він може стати привабливим для вітчизняних і зарубіжних інвесторів.

## **Висновки**

Доведено, що реальний опціон може застосовуватися в якості інструменту управління фінансовими ризиками і його використання дозволяє розширити діапазон прийняття рішень фінансовими менеджерами компаній.

Показані переваги використання методу реальних опціонів в порівнянні з методом дисконтування грошових потоків для аналізу і оцінки фінансових результатів в умовах високої невизначеності. визначено умови використання реальних опціонів в інфраструктурних проектах.

Проведено розрахунок вартості інвестиційного проекту будівництва великої кільцевої дороги навколо Києва із застосуванням формули Блека-Шоулза, методів Датара-Метьюз і нечітких множин.

Запропоновано економіко-математичну модель, яка включає чотири основні аспекти: сукупність фінансових відносин учасників проектів (органів державної влади, спеціалізованих фінансових інститутів, фінансових посередників і приватних інвесторів); етапи інвестиційної діяльності (від прийняття органом державної влади рішення про проведення відкритого конкурсу на створення інфраструктурного об'єкта до експлуатації готового об'єкта з розподілом доходів між основним власником (державою) і приватними інвесторами); дерево рішень, що відображає вчинення учасниками фінансово-значущих дій на різних етапах в залежності від рівня прибутковості проекту і виду використовуваних реальних опціонів; розподіл ризику і відповідальності між учасниками.

## Література

1. Воронцовский, А. В. Управление рисками с помощью условных и безусловных срочных контрактов на продажу продукции / А. В. Воронцовский // Финансы и бизнес. - 2006. - № 3. – С. 130-141
2. Грачева, М. В., Управление рисками в инновационной деятельности / М. В. Грачева, С. Ю. Ляпина. - М. : Юнити-Дана, 2010. – 352 с.
3. Маевский, Ф. В. Риск-менеджмент. Превентивное управление : учебно-практическое пособие / Ф. В. Маевский, В. Н. Вяткин, В. А. Гамза. – Днепропетровск: Наука и образование, 2013. – 264 с.
4. Семенкова, Е.В. Операции с ценными бумагами / Е. В. Семенкова. - М. : Дело, 2009. – 488 с.
5. Халл, Дж. К. Опционы, фьючерсы и другие производные финансовые инструменты / Дж. К. Халл. - Москва-Санкт-Петербург-Киев: ИД "Вильямс" - 2008. - 890 с.
6. Чернова, Г. В., Управление рисками / Г. В. Чернова, А. А. Кудрявцев - М.: ТК Велби, Изд-во Проспект., 2008. – 190 с.
7. Малюга К.А. Реальные опционы как инструмент управления финансовыми рисками инвестиционного проекта. [Электронный ресурс] — Режим доступа: <http://www.docme.ru/download/1412457>

УДК: 519.766.4

## **Моделювання і прогнозування нелінійного нестационарного процесу ціноутворення на нафту**

**Конєв Д.В., Бідюк П.І.**

### **Вступ**

В умовах загальної економічної нестабільності, коли у світі відбуваються фінансово-економічні кризи, виникає необхідність удосконалення світової фінансової системи, а також методів і технологій управління ризиками в даній сфері. Актуальною задачею є виконання поглиблених наукових досліджень в напрямі математичного моделювання і прогнозування фінансових та економічних процесів. Це стосується, безумовно, і українського фінансового ринку, на якому фінансовий ризик-менеджмент знаходиться на ранній стадії свого становлення.

Прагнення побудувати модель, яка б найбільш точно відповідала реальній поведінці фінансових ринків, та необхідність підвищення якості прогнозів ведуть до появи нових класів моделей, а також модифікацій вже існуючих [1]. Таким чином, побудова і дослідження адекватних математичних моделей, які описують та прогнозують динаміку таких фінансових інструментів як акції, облігації, котирування валют та інших, є актуальним напрямом сучасних досліджень.

Дослідження присвячене аналізу нелінійних нестационарних фінансово-економічних процесів, побудові та застосуванню регресійних моделей досліджуваних процесів до їх прогнозування. Зокрема, розглядається нестационарний процес ціноутворення на нафту.

### **Постановка задачі**

Метою роботи є виконання наступних завдань:

- вибрати типи математичних моделей для опису фінансово-економічних процесів;

- обрати статистичні дані та побудувати на їх основі вибрані моделі у формі регресійних рівнянь;
- оцінити статистичні показники побудованих моделей для процесу ціноутворення на нафту;
- провести порівняльний аналіз отриманих результатів та обрати найкращу модель.

Економічна динаміка є нелінійним явищем і саме виявлення нелінійностей не завжди є простою задачею. Нелінійність – це неочікувані зміни у напрямку розвитку процесів. В динамічних економічних системах вона виявляється, як підвищена на реакція на зміну одних факторів та повна нечутливість до інших [2].

Нестационарні процеси є поширеним підкласом нелінійних процесів, до якого, в свою чергу, входять процеси: з трендом та змінною дисперсією. Тренд – це загальний довгостроковий напрям розвитку процесу, що фактично співпадає з поточним середнім значенням. Виділяють два типи тренду: стохастичний та детермінований. А процеси зі змінною дисперсією називають гетероскедастичними [3].

Дані про фінансово-економічні процеси надаються у вигляді часових рядів. Зручним інструментом для їх аналізу та прогнозування є різницеві рівняння. Застосування різницевих рівнянь має деякі переваги: простота оцінювання структури і параметрів математичних моделей за допомогою стандартних пакетів прикладних програм; відносна простота знаходження та аналізу розв'язків отриманих рівнянь; можливість використання різницевих рівнянь для коротко- та довгострокового прогнозу та інші [4].

В загальному випадку прогноз може бути представлений одним (точковим) значенням змінної, інтервалом, в який попадає випадкова змінна, а також ймовірністю прийняття змінною деякого значення у вибраному інтервалі. Якщо для опису процесу застосовують лінгвістичні змінні, то прогнозом буде нечітке значення, але його також можна перетворити в чітке число. Можна по-різному ставити задачу прогнозування в залежності від рівня прийняття рішення та конкретної поставленої задачі управління чи контролю. Прогнозування може стосуватись наступних складових процесу:

- детермінованого тренду, як індикатора довгострокових змін процесу;
- стохастичного тренду, як показника коротко- та середньострокових змін;
- коливань, що накладаються на тренд;
- сезонних ефектів;
- приростів (швидкості) зміни процесу, які визначаються першими різницями;
- дисперсії або стандартного відхилення, як міри розсіювання процесу.

Відповідно до того, які складові процесу необхідно прогнозувати, ставиться задача побудови математичної, ймовірнісної (статистичної) або логічної моделі, що має меті забезпечити високу якість прогнозу на заданий часовий горизонт. Розглянемо деякі можливості створення математичного опису складових процесів різної природи [5].

#### Детермінований тренд

Якщо описати детермінований тренд за допомогою полінома від часу довільного порядку:

$$y(k) = a_0 + a_1 k + a_2 k^2 + \dots + a_p k^p + \varepsilon(k), \quad E[\varepsilon(k)] = 0,$$

то визначення прогнозу тренду зводиться до підстановки в це рівняння бажаного значення часу  $k$  і застосування безумовного математичного сподівання. Прогнозування значень тренду вважається довгостроковим прогнозом, оскільки детермінований тренд вказує на довгострокові зміни процесів. Обмеження на випадковий процес  $E[\varepsilon(k)] = 0$  необхідне для коректного застосування методів оцінювання моделей, а також для виконання подальшого аналізу якості оцінок прогнозів [6].

#### Стохастичний тренд

Для опису і прогнозування стохастичного тренду можна скористатись, наприклад, відомим рівнянням випадкового кроку із перетином (константою):

$$y(k) = a_0 + y(k-1) + \varepsilon(k),$$

розв'язок якого має вигляд:

$$y(k) = y_0 + ka_0 + \sum_{i=1}^k \varepsilon(i).$$

Сума  $\sum_{i=1}^k \varepsilon(i)$  в правій частині останнього рівняння описує випадкову складову тренду. Цю складову називають нерегулярною.

Коливання, що накладаються на тренд, або короткострокові зміни, можна описати рівняннями авторегресії з ковзним середнім:

$$y(k) = \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k).$$

При переході до перших (та вищих) різниць порядків з процесу вилучається тренд відповідного порядку. Наприклад, якщо процес містить лінійний тренд, то перші різниці вилучають його і після переходу до різниць ми маємо справу з коливаннями, які накладаються на тренд[6].

### **Прогнозування дисперсії**

Якщо дисперсія процесу змінюється в часі, то для її описання можна вибрати рівняння для формування відповідної функції прогнозування та обчислення в подальшому оцінки прогнозованого значення. Процеси зі змінною дисперсією отримали назву гетероскедастичних. Методика побудови моделей гетероскедастичних процесів передбачає такі кроки:

- математичний опис самої змінної рівнянням авторегресії невисокого порядку (наприклад, першого);
- математичний опис умовної дисперсії, як динамічної змінної, за допомогою рівняння прийнятної (за якістю прогнозу) структури.

### **Обчислювальні експерименти з побудови моделей**

Для оцінки параметрів моделі будемо використовувати метод найменших квадратів (МНК). В якості статистичних даних для виконання обчислювальних

експериментів використовувалися щомісячні показники цін на нафту впродовж п'яти років. Це обумовлено тим, що дані показники є досить нестабільними.

Побудуємо моделі для опису ціни на нафту. На рисунку 1 зображено значення ЧАКФ.

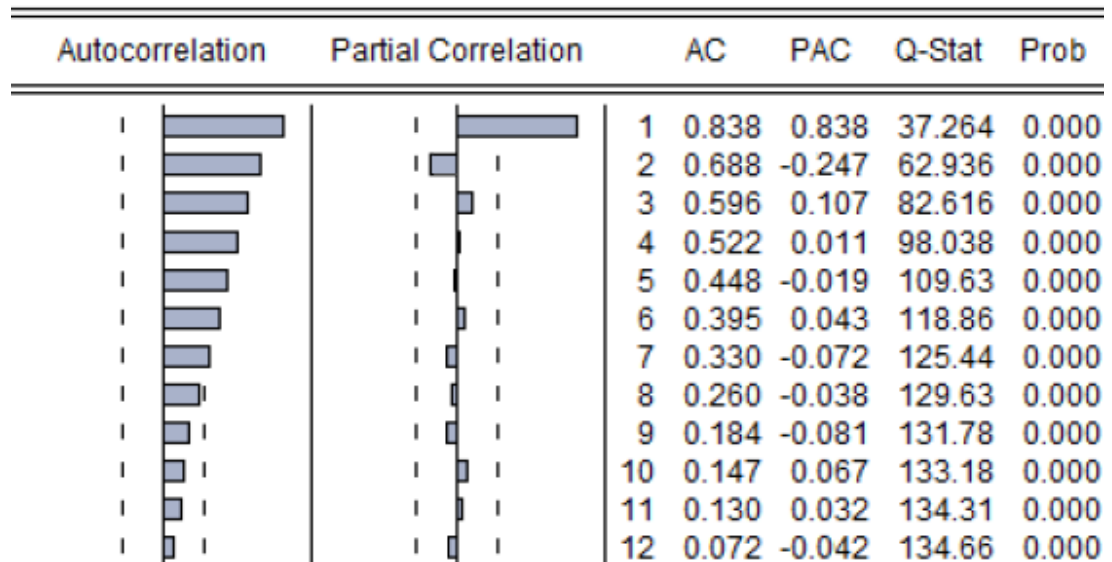


Рисунок 1. Значення ЧАКФ для показників цін на нафту

Далі побудуємо модель  $AR(1)$ , а також спробуємо ускладнити її до  $AR(2)$ . На рисунку 2 зображені статистичні показники якості моделі  $AR(1)$ . Як бачимо, коефіцієнт множинної детермінації та статистика Дарбіна-Уотсона мають хороші показники. На рисунку 3 – графічне зображення моделі з величиною помилки.



Dependent Variable: TRAFFIC  
Method: Least Squares  
Date: 06/05/17 Time: 21:00  
Sample (adjusted): 2 60  
Included observations: 59 after adjustments  
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.048987	0.010195	102.8925	0.0000

R-squared	0.985218	Mean dependent var	237.8305
Adjusted R-squared	0.985218	S.D. dependent var	183.2784
S.E. of regression	22.28298	Akaike info criterion	9.062327
Sum squared resid	28798.81	Schwarz criterion	9.097540
Log likelihood	-266.3387	Hannan-Quinn criter.	9.076073
Durbin-Watson stat	1.944284		

Inverted AR Roots	1.05
-------------------	------

Estimated AR process is nonstationary

Рисунок 2. Модель AP(1)

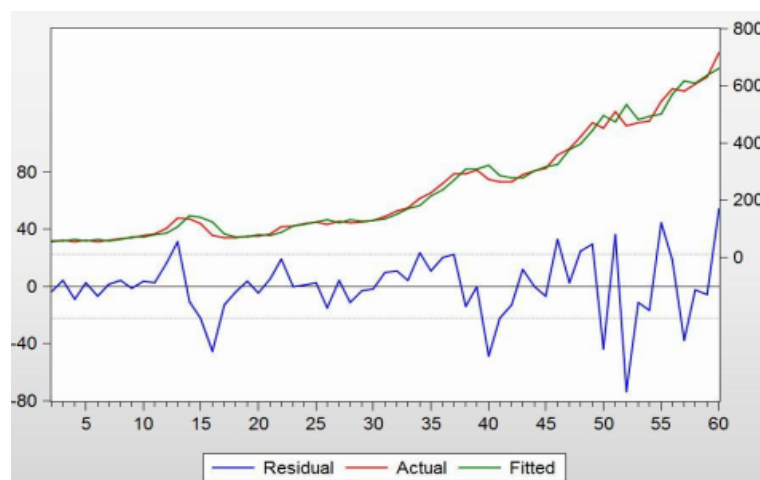


Рисунок 3. Графічне зображення моделі AP(1)

Далі модель була ускладнена до AP(2), а також були побудовані моделі AP(1) з включенням тренду та квадратичного тренду. На рисунку 4 зображені показники моделі AP(1) з лінійним трендом, на рисунку 5 – з квадратичним трендом, а на рисунку 6 – графічне зображення моделі з квадратичним трендом.

Dependent Variable: TRAFFIC  
 Method: Least Squares  
 Date: 06/05/17 Time: 21:07  
 Sample (adjusted): 2 60  
 Included observations: 59 after adjustments  
 TRAFFIC=C(1)+C(2)\*TRAFFIC(-1)+C(3)\*K

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-5.039655	6.702047	-0.751958	0.4552
C(2)	1.000539	0.045012	22.22846	0.0000
C(3)	0.520031	0.454823	1.143370	0.2577
R-squared	0.985556	Mean dependent var		237.8305
Adjusted R-squared	0.985041	S.D. dependent var		183.2784
S.E. of regression	22.41660	Akaike info criterion		9.106990
Sum squared resid	28140.23	Schwarz criterion		9.212628
Log likelihood	-265.6562	Hannan-Quinn criter.		9.148227
F-statistic	1910.567	Durbin-Watson stat		1.904447
Prob(F-statistic)	0.000000			

Рисунок 4. Модель AP(1) + лінійний тренд

Dependent Variable: TRAFFIC  
 Method: Least Squares  
 Date: 06/05/17 Time: 21:09  
 Sample (adjusted): 2 60  
 Included observations: 59 after adjustments  
 TRAFFIC=C(1)+C(2)\*TRAFFIC(-1)+C(3)\*K+C(4)\*K\*K

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	6.965417	4.735013	1.471045	0.1482
C(2)	0.994527	0.090116	10.37023	0.0000
C(3)	-0.409386	0.225287	-1.817171	0.0759
C(4)	0.011450	0.004657	2.458770	0.0179
R-squared	0.986864	Mean dependent var		67.47143
Adjusted R-squared	0.942148	S.D. dependent var		21.89539
S.E. of regression	5.266380	Akaike info criterion		6.238671
Sum squared resid	1248.064	Schwarz criterion		6.393105
Log likelihood	-148.8474	Hannan-Quinn criter.		6.297263
F-statistic	261.5674	Durbin-Watson stat		1.790029
Prob(F-statistic)	0.000000			

Рисунок 5. Модель AP(1) + квадратичний тренд

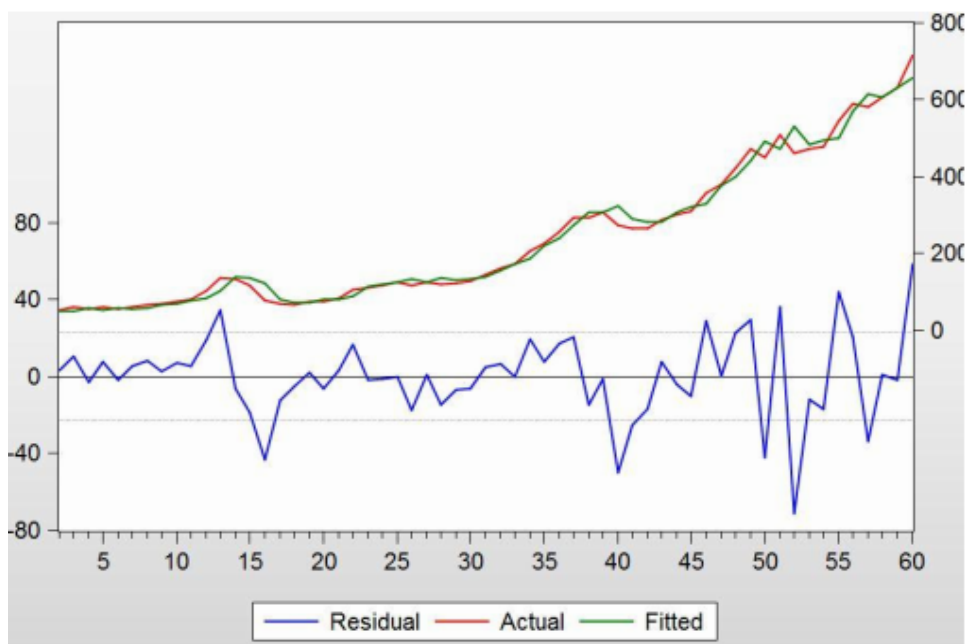


Рисунок 6. Графічне зображення моделі AP(1) + квадратичний тренд

Отже, були побудовані такі моделі:

1) AP(1):  $y(k) = c(1) + c(2) * y(k - 1)$ ;

2) AP(2):  $y(k) = c(1) + c(2) * y(k - 1) + c(3) * y(k - 2)$ ;

3) AP(1) + лін. тренд:  $y(k) = c(1) + c(2) * y(k - 1) + c(3) * k$ ;

4) AP(2) + квадр. тренд:  $y(k) = c(1) + c(2) * y(k - 1) + c(3) * k + c(4) * k^2$ .

Для вибору моделі прогнозування динаміки ціноутворення нафти, порівняємо отримані моделі за статистичними характеристиками та виберемо найкращу. Результати, отримані для побудованих моделей, занесені до таблиці 1.

Таблиця 1. Порівняння характеристик моделей

Модель	SEE	R2	DW	СКП	САПП	U
AP(1)	28798,81	0,9852	1,9444	22,093	7,8317	0,0369
AP(2)	28764,70	0,9850	1,8877	22,270	7,8811	0,0369
AP(1) + лін. тренд	28140,23	0,9856	1,9044	21,839	7,9930	0,0365

AR(1) + квадр.	25755,31	0,9868	1,7900	20,893	7,8562	0,0349
тренд						

З таблиці бачимо, що найкращою, стосовно однокрокового прогнозу, виявилася модель AR(1) з квадратичним трендом. Про це свідчать показники СКП і САПП (7,856%). Отримана модель має вигляд:

$$y(k) = 6,9654 + 0,9945y(k - 1) - 0,4093 * k + 0,0145 * k^2.$$

Коефіцієнт при затриманому на крок значення основної змінної у правій частині (0,9945) свідчить про наявність квадратичного тренду. Загалом побудовані моделі придатні для коротко- і довгострокового прогнозування, оскільки середня абсолютна похибка у процентах не перевищує значення 8,0.

### Висновки

У даній роботі були розглянуті деякі підходи до побудови моделей та оцінювання прогнозів фінансово-економічних процесів. Побудована модель для опису ціноутворення на нафту. Найкраще ряд описує модель AR(1) + квадратичний тренд. Про наявність квадратичного тренду свідчить також візуальний аналіз процесу ціноутворення, поданого у графічному вигляді. У останні роки на графіку видно чітко виражений спад, а потім поступовий підйом. Саме вплив таких неконтрольованих збурень впливає на побудову моделі та прогнозування, що набагато ускладнює цей процес.

У подальших дослідженнях доцільно застосувати для опису фінансових процесів альтернативні моделі, наприклад, поліноміальну модель, побудовану за методом групового урахування аргументів, нейронну та динамічну байєсівську мережі. Для підвищення якості оцінок прогнозів також доцільно застосувати комбінування оцінок, обчислених за альтернативними методами. Як правило, дисперсія похибок оцінок комбінованих прогнозів не перевищує або є меншою ніж дисперсія похибок оцінок найкращого прогнозу.

## Література

1. *Бідюк П.І.* Аналіз та моделювання економічних процесів перехідного періоду. – Київ: Інститут прикладного системного аналізу при НТУУ «КПІ», 2000.
2. *Бідюк П.І.* Системный подход к построению регрессионной модели по временным рядам / Бідюк П.І., Баклан И.В. // Системні дослідження та інформаційні технології, 2002 – с. 114-131.
3. *Priestley M.B.* Non-linear and Non-stationary Time Series Analysis. – London: Academic Press, 1989. – 140 p.
4. *Бідюк П.І.* Аналіз часових рядів (навчальний посібник) / Бідюк П.І., Романенко В.Д., Тимошук О.Л. – К.: Політехніка, 2010. – 317 с.
5. *Вітлінський В.В.* Моделювання економіки: Навч. посібник. — К.: КНЕУ, 2003. — 408 с.
6. *Присенко Г.В., Равікович Є.І.* Прогнозування соціально-економічних процесів: Навч. посібник — К.: КНЕУ, 2005. — 378 с
7. *Бідюк П.І., Баклан И.В.* Системный подход к построению регрессионной модели по временным рядам // *Автоматика. Автоматизация. Электротехн. комплексы и системы*, – 2003. – № 2. – с. 11-19.
8. *Молчанов И.Н.* Компьютерный практикум по начальному курсу эконометрики (реализация на Eviews): Практикум / Ростовский государственный экономический университет. / Молчанов И. Н., Герасимова И. А. – Ростов-н/Д., –2001.

УДК: 303.732.4:519.237

**Костриця Д.В.**

**Розробка системної методології побудови скорингової карти**

Науковий керівник: д.т.н., проф. Бідюк П.І.

Ризик втратити частину коштів – неминуча проблема усієї банківської системи, яка є пов’язаною з безпосередньою діяльністю банків. Уникнути цієї проблеми практично неможливо, навіть, можна сказати, нереально. Проте завжди можна звести ризики до мінімуму, що і є основною задачею, адже мінімізація ризику – один із кроків до збереження прибутку і забезпечення життєздатності банку або фінансової установи. Таким чином, високо актуальною є задача менеджменту фінансових ризиків з використанням сучасних методів математичного моделювання і прогнозування.

Першим пунктом на шляху ефективного управління кредитними ризиками є усвідомлення проблеми та знаходження правильного підходу до коректного і точного вимірювання ризику. Так, є досить багато методів оцінювання кредитних ризиків, розроблено велику кількість моделей, що і говорить про актуальність цієї проблеми. Одним із найбільш перспективних та широко застосованих методів є статистичний метод скоринг-аналізу. В основу даного методу покладено оцінку кредитоспроможності позичальника, що базується на історії минулих кредитних операцій – кредитній історії клієнта.

Дослідження присвячене розробці системної методології побудови скорингової моделі для оцінки ризиків кредитування фізичних осіб. Методологія подається у вигляді зручного алгоритму обробки історичної вибірки.

### **Постановка задачі**

Метою роботи є виконання таких завдань:

- зібрати статистичні дані стосовно характеристик клієнтів-позичальників кредитів фінансових установ;
- розробити удосконалену системну методологію побудови скорингу;
- обрати математичну модель для аналізу та прогнозування ризиків кредитування фізичних осіб;
- побудувати скорингову карту за розробленою методологією;

— виконати аналіз результатів.

При прогнозуванні дефолту часто використовують експертний підхід, так звана верифікація апікаційних даних. Неважко зрозуміти, що експертна оцінка є досить суб'єктивною. А якщо розглядати скоринговий підхід, то у його основі можуть лежати різні статистичні моделі: лінійна регресія, логістична регресія, пробіт-моделі, дерева рішень та інші – тобто відбувається зменшення рівня суб'єктивізму, що є немало важливим при прийнятті рішень. Найбільш часто при розробці скор-карт використовують саме логістичну регресію, адже вона дає більш точні результати в порівнянні, наприклад, з лінійною, і реалізувати її простіше, ніж пробіт-модель [1].

Логістична регресія – найбільш розповсюджена статистична модель для побудови скорингових карт при бінарній залежній змінній. Математична модель логістичної регресії являє собою залежність логарифму шансу від лінійної комбінації незалежних змінних:

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 x_i^{(1)} + \dots + b_k x_i^{(k)},$$

де  $p_i$  – ймовірність того, що  $i$ -й клієнт стане «поганим»;  $x_i^{(j)}$  – значення  $j$ -тої незалежної змінної;  $b_0$  – константа моделі;  $b_j$  – параметри моделі [2].

### **Розробка методики скорингового аналізу кредитоспроможності**

Незалежно від того, яка математична модель обрана для розрахунку скорингових балів (або ж коефіцієнтів регресії, якщо розглядати це в рамках математичної моделі), розроблено певні пункти, які є обов'язковими при побудові скорингу.

Крок 1. Опис та підготовка статистичних даних. Висунуто наступні вимоги до історичної вибірки:

1. Одноманітність вибірки. Розглядати необхідно клієнтів, які є схожими між собою, тобто це мають бути люди приблизно одного віку, сума кредиту повинна бути в межах невеликих відхилень, навіть місце проживання має досить велике значення – місто-мільйонер чи селище міського типу. Варто звернути

увагу, що із вибірки мають бути виключені дані про особливих клієнтів. До таких клієнтів можна віднести неповнолітніх, VIP-клієнтів або співробітників банку.

2. Велика кількість досліджуваних об'єктів. Для точного результату недостатньо взяти вибірку із, наприклад, 1000 елементів, необхідно мати в запасі великий об'єм інформації, тоді можна говорити про адекватність отриманого результату.

3. Обмеженість у часі. Хоча даний критерій можна віднести до правила одноманітності вибірки, проте на ньому варто зосередити особливу увагу. Під обмеженістю у часі мається на увазі розгляд кредитів, дата початку яких лежить у певному обмеженому періоді.

4. Розширена інформація про позичальників. Варто зазначити, що для аналізу та побудови кредитного скорингу, необхідно накопичувати якомога повніші дані по клієнтах.

5. Розподіл кредитних історій по різних кредитних продуктах. На даний момент розрізняють кілька кредитних ліній, а саме споживчий кредит, кредитна карта, кредит на авто, іпотека, кредит готівкою тощо. Для кожного з цих продуктів є свої умови кредитування, отже, і скорингова модель для кожного повинна будуватися окремо.

Крок 2. Визначення залежних та незалежних змінних. Коли статистичні дані зібрані за вказаними вимогами, визначаємось із залежною та незалежними змінними. В якості залежної змінної обираємо характеристику, яка показує наявність прострочених платежів по кредиту більш ніж на 90 днів. Незалежними змінними є всі характеристики, вказані позичальником у аплікаційній заявці. Наприклад, стать, сімейний стан, вік, посада, поточний стаж роботи, загальний стаж, строк проживання з останньою адресою, наявність утриманців та їх кількість, рівень освіти, дохід позичальника, дохід сім'ї, офіційність працевлаштування та ін. Також введено перетин параметрів та фінансові правила (відношення розходів до доходів та ін.).

Крок 3. Визначення об'єму вибірки. Можемо реалізовувати цей крок двома способами:

1. Правило 20 EPV (Event per Variable): на одного «поганого» клієнта має припадати 20 і більше предикторів.



2. Вибірка сформована випадковим чином із генеральної сукупності, розмірів якої ми не знаємо, тому можливе застосування такої формули для встановлення мінімального об'єму вибірки:

$$n = \frac{z_{\gamma}^2 \omega (1 - \omega)}{\Delta_{\omega}^2}, \quad (1)$$

де  $n$  – мінімальний об'єм вибірки,  $z_{\gamma}$  – значення стандартного нормального закону розподілу, яке визначається в залежності від обраного рівня надійності  $\gamma$ ,  $\omega$  – відсоток «поганих» клієнтів у тестовій вибірці,  $\Delta_{\omega}$  – максимально допустима гранична помилка оцінки долі «поганих» позичальників.

Крок 4. Описовий аналіз скорингових змінних. Окрему увагу варто звернути на наявність у вибірці даних, які мають допустимі значення, але які можна віднести до категорії екстремального характеру. Якщо таких елементів небагато, їх можна видалити із вибірки, та замінити їх елементами із середніми значеннями.

Робота з пропущеними значеннями – окремий етап аналізу даних перед побудовою моделі. Величезні масиви фінансових даних для побудови скорингової моделі завжди містять пропуски, причину яких необхідно аналізувати. Такими причинами можуть бути:

- помилки при вводі даних;
- помилки при зборі даних;
- неможливість отримання інформації;
- свідомо відмова позичальника від надання даних.

Можливим способом боротьби із відсутністю даних є заміна їх середніми значеннями.

Крок 5. Перевірка кількісних змінних на мультиколінеарність. Можна рахувати за допомогою побудови матриці кореляцій між елементами, але зручніше скористатись показником толерантності, а точніше величиною, оберненою до цього показника, яку ще також називають фактором «здуття» дисперсії, і розраховується за наступною формулою:

$$VIF = \frac{(SE)^2 \cdot (n - 1) \cdot Std.Dev^2}{OSE^2} \quad (2)$$

де SE – оцінка стандартної помилки; n – розмір вибірки; Std.Dev – стандартне відхилення; OSE – стандартна помилка. Якщо це значення перевищує 10, означає, що варто виключити змінну з аналізу.

Крок 6. Категоризація кількісних змінних. Категоризація дозволяє досягти наступних переваг при побудові скорингової карти: полегшення обробки екстремальних значень кількісних змінних, спрощення інтерпретації скорингової карти, відображення складних нелінійних зв'язків.

Категоризацію кількісних змінних будемо проводити за допомогою процедури біннінгу, тобто скористаємося таким алгоритмом:

1. Розбиваємо кількісні змінні на групи за близькими значеннями.

2. У кожній групі рахуємо відсотки «поганих» і «хороших» клієнтів, а також показник ваги категорій предиктора WOE, який допомагає знайти для змінної «границі чутливості» до появи змодельованої події ризику та провезти раціональним чином категоризацію кількісних змінних. Показники WOE для кожної категорії розраховуються за формулою:

$$WOE_i = \ln \left( \frac{a_i^{(1)}}{a_i^{(2)}} \right) \quad (3)$$

де  $a_i^{(1)}$  та  $a_i^{(2)}$  – відносні частоти «поганих» і «хороших» кредитів відповідно в і-тій групі категоризаційної змінної;  $i = 1, 2, \dots, k$ ,  $k$  – кількість категорій змінної.

3. Проведення аналізу, об'єднання сусідніх категорій та перерахунок показників WOE. При подальшому об'єднанні категорій керуються наступними правилами: в кожній групі має знаходитись не менше 5% від усіх спостережень змінної; не повинно бути груп, з кількістю «поганих» та «хороших» кредитів, рівних нулю; відсоток «поганих» позичальників і WOE повинні в достатній мірі відрізнятись один від одного за групами. При категоризації окрім статистичних

критеріїв необхідно керуватися логікою, можливістю та доцільністю об'єднання [4].

Крок 7. Визначення впливових характеристик для побудови скорингу. Для оцінки степенів взаємозв'язку між незалежними змінними та бінарною залежною змінною у кредитному скорингу прийнято використовувати показник інформаційного значення або IV, який розраховується за формулою:

$$IV = \sum_{i=1}^k [(d_i^{(1)} - d_i^{(2)}) * WOE_i], \quad (4)$$

де  $k$  – кількість категорій незалежної змінної, всі інші позначення із формули (3).

Крок 8. Застосування логістичної регресії для знаходження коефіцієнтів моделі та переведення коефіцієнтів моделі у скорингові бали. Завершальним етапом розробки скорингової моделі є перетворення коефіцієнтів логістичної регресії у скорингові бали. Якщо взяти оцінки коефіцієнтів логістичної регресії та помножити їх на значення незалежних змінних, отримаємо результуючий скоринговий бал у шкалі натуральних логарифмів:

$$score = b_1 x_1 + \dots + b_k x_k, \quad (5)$$

де  $x_j$  – значення параметрів для певного клієнта;  $b_j$  – оцінки коефіцієнтів логістичної регресії.

Крок 9. Оцінка моделі за критеріями якості. Введено наступні критерії якості:

1. Кількість помилок першого роду: спрогнозовано дефолт, а насправді клієнт «хороший». Критерій обчислюється досить просто, і дає чітке уявлення про те, на скільки модель якісна. Окрім абсолютних значень – кількостей помилок – додатково можна ввести метрики: відношення кількості помилок першого роду до загального числа дефолтів, або ж відношення кількості помилок першого роду до загального об'єму тестової вибірки.

2. Кількість помилок другого роду – схожий критерій на той, що описаний в першому пункті: дефолт не спрогнозовано, а насправді він був.

3. Середня абсолютна похибка у процентах. Розраховується у відсотках, і показує загальне відхилення спрогнозованих даних від фактичних:

$$САП_{\Pi} = \frac{1}{N} \sum_{k=1}^N \frac{|y(k) - \hat{y}(k)|}{|y(k)|} 100\%$$

де  $y(k)$  – значення змінної, що спостерігається,  $\hat{y}(k)$  – значення змінної, побудовані за параметрами, які знайшли [3].

### Побудова скорингової моделі за розробленою методикою та огляд результатів

В рамках розробленої системної методології було реалізовано програмний продукт, який дає змогу автоматизувати певні кроки розрахунків. Як було вже сказано, алгоритм передбачає перегляд результатів після кожного виконаного обчислення. Побудова скорингу не може бути віддана на автоматизацію повністю, оскільки навіть робота з пропущеними даними, екстремальними значеннями – це цілком ручна робота і потребує людської уваги.

Отже, за розробленим алгоритмом реалізовано збір статистичних даних, які відповідають всім вказаним вимогам. В якості залежної змінної обрано характеристику, яка показує наявність прострочених платежів по кредиту більше ніж на 90 днів. Незалежні змінні – параметри із аплікаційної заявки. Використовуючи формулу (1), розраховано мінімальний об'єм вибірки. Автоматизація цього кроку представлена на рисунку 1:

number of predictors	46	"bad" luck clients	2,59%
"bad" clients	740	value of standart normal distribution	1,96
"good" clients	27 790	max permissible error limit	0,14%
max permissible error	5,50%	min size (1)	35 470
reliability	95,00%	min size (2)	24 332

## Рисунок 1 – Визначення мінімального розміру навчальної вибірки

Як ми можемо бачити на рисунку 1, пораховано розмір двома способами, де «min size (1)» – розрахунок за правилом 20 EPV, а «min size (2)» – за формулою (1).

Окрім обробки даних, додатково провалі дуємо дані на наявність екстразначень та недостовірних значень:

- для характеристик, які вимірюються в місяцях, тривалість обмежимо від 0 до 120 місяців (всі значення, що знаходяться не в цьому діапазоні, з вибірки видаляємо);
- перевірка, щоб всі категоріальні поля мали уніфіковані словники;
- вік має бути більший за термін реєстрації, термін проживання, стаж, термін шлюбу;
- вік дружини/чоловіка має бути більший за термін шлюбу.

Користуючись вбудованим інструментом MS Excel, порахували мультиколінеарність між усіма кількісними змінними, в результаті довелось відмовитись від поля Total Income Confirmed (загальний підтверджений дохід), а також від Predict Rest (залишок від загального доходу після вирахування платежів по всіх відкритих кредитах та щомісячних витрат). Ці дві характеристики вводились нами додатково, і базуються на існуючих даних із заявки, саме тому не дивно, що їх довелось виключити через мультиколінеарність [4].

Для кожної кількісної змінної проведено категоризацію за алгоритмом, описаним в кроці 6. Розглянемо на прикладі характеристики «термін реєстрації». Зручно представити ці результати у вигляді графіків WOE, які дають розуміння, що динаміка не змінюється, а змінюється кількість категорій. На рисунках 2-3 показані графіки WOE до та після категоризації відповідно.

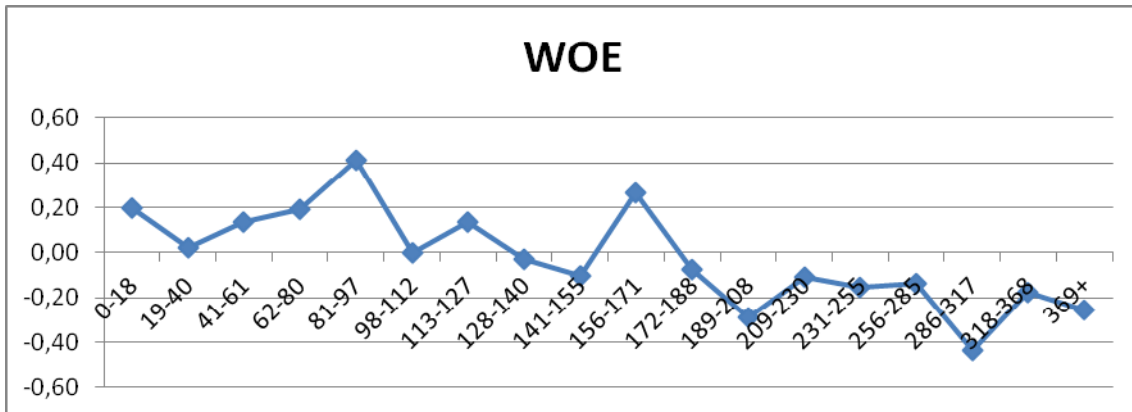


Рисунок 2 – Графік WOE до категоризації

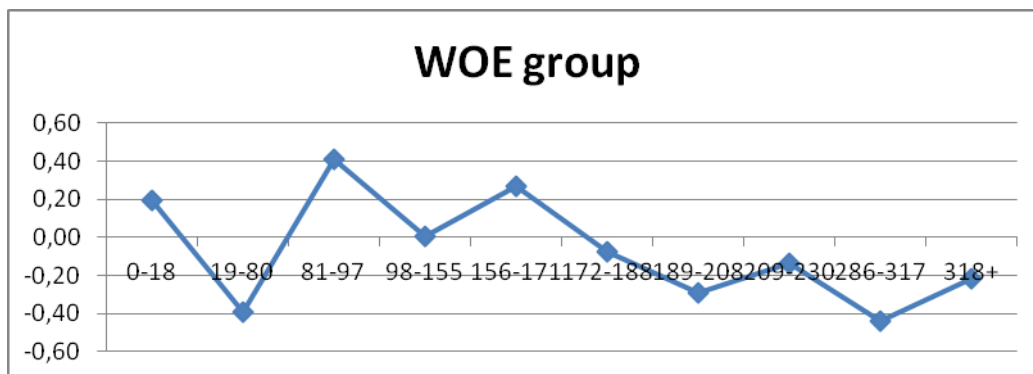


Рисунок 3 – Графік WOE після категоризації

Такі розрахунки проведені для всіх кількісних змінних. І наступним кроком є розрахунок значимості параметрів, або Information value. Він розраховується не тільки для кількісних змінних, а і для категоріальних, тому що необхідно розуміти, які параметри включати в модель, а які ні. Розглянемо для цієї ж характеристики розрахунок IV за формулою (4). У таблиці 1 подано результати розрахунку:

Таблиця 1 – Information value для категоризованої кількісної характеристики «термін реєстрації»

Group	IV
0-18	0,00
19-80	0,04

81-97	0,01
98-155	0,00
156-171	0,00
172-188	0,00
189-208	0,00
209-230	0,00
286-317	0,01
318+	0,00
TOTAL	0,07

Як бачимо з таблиці 1, фінальне значення метрики дорівнює 0,07. Загальноприйнятою вважається наступна градація: менше 0,02 – незалежна змінна не має прогностичних властивостей; між 0,02 та 0,1 – незалежна змінна має низьку прогностичну можливість; від 0,1 до 0,3 – середня прогностична можливість; від 0,3 до 0,5 – висока прогностична можливість; більше 0,5 – надзвичайно точний прогноз за цією незалежною змінною [5].

В нашому випадку характеристика «термін реєстрації» має низьку прогностичну властивість. Її включати до скорингової моделі не будемо.

Отже, обрано в якості математичної моделі, на основі якої побудована скорингова карта, логістичну регресію. Визначили, які параметри необхідно включати у модель. Проведемо ряд експериментів на різних навчальних вибірках. Для зручності опишемо всі критерії якості для кожної із побудованих моделей і покажемо все в таблиці 2:

Таблиця 2 – Результати експериментів

Навчальна вибірка	Середня абсолютна похибка, %	Помилки 1 роду, %	Помилки 2 роду, %
Навчальна вибірка 1. Об'єм 27 315 елементів, «поганих» 733 експериментів	29,4%	48/184 = 26,1%	51/184 = 27,7%
Навчальна вибірка 2. Об'єм 1 500	18,8%	45/184 =	35/184 =

елементів, «поганих» 733 експериментів		24,5%	19,02%
Навчальна вибірка 3. Об'єм 2 500 елементів, «поганих» 733 експериментів (випадковим чином відібрані інші елементи із повної вибірки, ніж для навчальної вибірки 2)	20,1%	47/184 = 25,5%	39/184 = 21,2%
Навчальна вибірка 4. Об'єм 3 000 елементів, «поганих» 702 експериментів (обмежено сумою кредиту 50 000 грн.)	27,6%	30/115 = 26,1%	28/115 = 24,3%

Як можемо бачити з таблиці 2, найкраща модель побудована на найменшій за об'ємом вибірці. Що, з одного боку, парадоксально. Але пояснюється це тим, що дана вибірка найбільш зважена, і кількість дефолтів та хороших експериментів приблизно однакова.

За формулою (5) пораховано скорингові бали для усіх введених в модель параметрів. Наведемо для прикладу скорингові бали для категоріальної об'єднаної характеристики «стать + сімейний стан». Окремо виведений параметр «стать» був статистично неважливим, і за недостатньо високим значенням метрики IV був виключений з подальшого аналізу. А от в поєднанні з розбитим на категорії кількісним параметром «сімейний стан», він увійшов до скорингової карти. Результати подано в таблиці 3.

Таблиця 3 – Скорингові бали для характеристики «вік + знак зодіаку»

Категорія	Скоринговий бал
F__ Ніколи у шлюбі не перебували	12
F__ Одружений / одружена	36
M__ Одружений / одружена	25
F__ Розведений / розведена	28
M__ Ніколи у шлюбі не перебували	-11
M__ Громадянський шлюб	-5
F__ Вдовець / вдова	14



F__Громадянський шлюб	25
M__Розведений / розведена	-36
M__Вдовець / вдова	10

Аналогічно побудовані скорингові розбивки для всіх інших параметрів. В такій шкалі зазвичай і рахуються скорингові бали для позичальників.

### **Висновки**

Спеціально для розробки моделі було підготовлено статистичну вибірку, яка складається близько із 37 000 елементів. Розглядалися тільки ті клієнти, які отримали позитивне рішення, а також накладено ряд інших обмежень, які зводять умови для позичальників до приблизно однакових. Було проведено процедури позбавлення пропусків даних та їх валідації.

Описано докладну покрокову інструкцію розробки скорингової моделі, що базується на логістичній регресії. Запропоновано модифіковані формули підрахунку скорингових балів. Розроблено програмний продукт, який обробляє статистичні дані та рахує скорингові бали за описаною методикою.

Перспективами для подальшого дослідження є градація на скорингові групи. Також доцільним є заміна логістичної регресії на нечіткі нейронні мережі, наприклад, ANFIS або TSK, які також дають якісні результати при оцінюванні ризику кредитоспроможності фізичних осіб на фінансових підприємствах.

### **Література**

1. Allen S. Financial risk management: A practitioner's guide to managing market and credit risk / Allen S. – Hoboken, N.J.: John Wiley & Sons, Inc., 2003.- 567р.
2. Костюченко Н. С. Анализ кредитных рисков / Н.С. Костюченко. – СПб.: ИТД «Скифия», 2010. – 440 с.
3. Бідюк П.І. Моделі оцінки ризиків кредитування фізичних осіб / Бідюк П.І., Матрос Є.О. // Кібернетика та обчислювальна техніка. — 2007. — № 153. — С. 87—95.
4. Стрельника Д.М. Скоринг як метод оцінки кредитоспроможності позичальника // Матеріали другої всеукраїнської наукової конференції

студентів та молодих вчених «Актуальні проблеми розвитку фінансово-кредитної системи України». — 2009. — 9–10 квітня. — С. 154–156.

5. Anderson, R. The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. [Text] / R. Anderson. – New York: Oxford University press, 2007. – 790 p. – ISBN 0199226407

**Надточій О. В.**

**Моделювання і прогнозування демографічних процесів в Україні**

Науковий керівник: д.т.н., проф. Бідюк П.І.

Моделі демографічних процесів і прогнози основані на них є важливими елементами в управлінні суспільними процесами. Вони дозволяють цілеспрямовано впливати на розвиток соціально-економічних явищ, коригувати їх у необхідний для країни бік. Демографічні прогнози дозволяють визначити ступінь розбіжності бажаних і можливих характеристик демографічних процесів. Отримані в результаті обчислень величини та параметри можна використовувати не тільки для відображення майбутніх характеристик демографічних процесів, а також використовувати при достроковому плануванні в галузях економіки, соціології та інших[1].

В сучасній демографічній ситуації спостерігається тенденція до зростання чисельності населення Землі. На сьогоднішній день чисельність населення складає 6,5 млрд. осіб. Але динаміка зростання кількості населення зумовлена насамперед збільшенням чисельності в найбідніших регіонах світу, в той час як в економічно розвинених країнах спостерігається зниження темпів природного приросту, а також часто зменшення загальної кількості населення у цілому.

Дослідження присвячене аналізу нестационарних демографічних процесів в Україні, побудові та застосуванню регресійних моделей досліджуваних процесів до їх коротко- та довгострокового прогнозування.

**Постановка задачі**

Метою роботи є виконання таких завдань:

- зібрати статистичні дані стосовно розвитку демографічних процесів в Україні для виконання обчислювальних експериментів;
- вибрати типи математичних моделей для опису нестационарних демографічних процесів і побудувати їх у формі регресійних рівнянь на основі статистичних даних;

— обчислити коротко- та довгострокові прогнози вибраних демографічних процесів.

В сучасній статистиці використовуються різні методи, які мають різну точність та беруть за основу різні моделі. Демографічні моделі призначені для опису стану населення та його змін, окремих елементів відтворення населення або процесу його відтворення в цілому.

Задача демографічної науки полягає не тільки в тому, щоб встановити і пояснити зміни в складі населення, але і передбачувати їх. Для цього виконують демографічні прогнози. Результати демографічних прогнозів потрібні органам державного управління і бізнес-структурам в багатьох країнах світу. Демографічні прогнози використовуються для визначення темпів розвитку економіки певних напрямків, при розробці планів соціального і економічного розвитку. На їх основі визначається попит на різні товари і послуги, оцінюються можливі зміни в навколишньому середовищі та міжнародній політиці. За допомогою прогнозів проводиться демографічна експертиза прийнятих рішень і розраховується кількість необхідних ресурсів для вирішення певних задач[1].

### **Моделі демографічних процесів та невизначеності моделювання**

Серед можливих невизначеностей, які можуть мати місце в моделюванні, прогнозуванні, оцінюванні ризиків та будіванні альтернатив, повинні бути вказані наступні: невизначеність даних; невизначеність структури моделі і її параметрів; невизначеність прогнозів, а також невизначеності пов'язані з обчислювальним процесом, наприклад, систематичні обчислення і помилки методології. На етапі оцінки структури і параметрів моделі ми зустрічаємо ще безліч невизначеностей, викликаних недостатньою кількістю інформації про нелінійні процеси, час затримки і типу розподілу даних. Для зменшення впливу невизначеностей вводяться адаптивні схеми оцінки для структури моделі, параметрів, а отже, прогнозу і оцінювання ризиків. Ключовий момент правильного оцінювання параметрів моделі полягає в правильному визначенні типу розподілу даних, що обробляються[2].

Кожний метод має свої недоліки і переваги щодо обчислювальних витрат та характеристик точності прогнозу. Так, висока точність прогнозу за допомогою МГВА або нейронної мережі іноді досягається за рахунок високих

обчислювальних витрат. Суттєвий вигравш можна досягти в такому випадку за допомогою простіших моделей авторегресії: АР, АРКС та АРІКС, перевагами яких є простота структури та можливості оперативної адаптації до характеристик процесу в реальному часі[5].

Зручним інструментом для аналізу та прогнозування часових рядів являються різницеві рівняння. Застосування різницевих рівнянь має деякі переваги: простота оцінювання структури і параметрів математичних моделей за допомогою стандартних пакетів прикладних програм на персональних комп'ютерах; відносна простота знаходження та аналізу розв'язків отриманих рівнянь; можливість використання різницевих рівнянь для коротко- та довгострокового прогнозу та інші[3].

Авторегресія (АР(p)): рівняння авторегресії описує пам'ять процесу, тобто вплив значень попередніх станів на його поточний стан:

$$y(k) = a_0 + a_1 y(k-1) + \dots + a_p y(k-p) = a_0 + \sum_{i=1}^p a_i y(k-i) + \varepsilon(k)$$

де  $a_i, i = 1, \dots, p$  – коефіцієнти моделі, які оцінюються на основі значень часового ряду;  $p$  – порядок авторегресії, який визначається числом затриманих в часі значень ряду, що використовуються в правій частині рівняння для описання динаміки змінної в момент  $k$ ;  $k = 1, 2, \dots$  – дискретний час;  $\varepsilon(k)$  – випадкова величина, поява якої зумовлена наступними причинами:

- вплив випадкових збурень на процес, що моделюється;
- похибки рівняння, зумовлені неточно вибраною структурою(можливо, що не враховано деякі регресори, введено непотрібні незалежні змінні або робиться спроба моделювати нелінійний процес за допомогою лінійного рівняння);
- методичні і обчислювальні похибки, які з'являються при обчисленні оцінок коефіцієнтів рівняння[4].

Парна регресія – вона включає в правій частині незалежну змінну(регресор):

$$y(k) = a_0 + a_1 x(k) + \varepsilon(k)$$

де  $x(k)$  – регресор (незалежна або екзогенна змінна); залежну змінну  $y(k)$  називають ендогенною змінною.

Авторегресія з ковзним середнім(АРКС(p,q)):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-i) + \varepsilon(k)$$

де  $q$  – порядок ковзного середнього,  $b_j$  – коефіцієнти ковзного середнього.

### Обчислювальні експерименти з побудови моделей

Для оцінки параметрів моделі будемо використовувати метод найменших квадратів(МНК).

Для виконання обчислювальних експериментів використовувалися такі вхідні дані: динаміка зміни чисельності всього населення України з 1991 по 2017 роки, динаміка зміни народжуваності, смертності, динаміка зареєстрованих шлюбів та розлучень за період з 1991 по 2015 роки.

Загальна тенденція зміни чисельності населення України зображена Рис. 1.

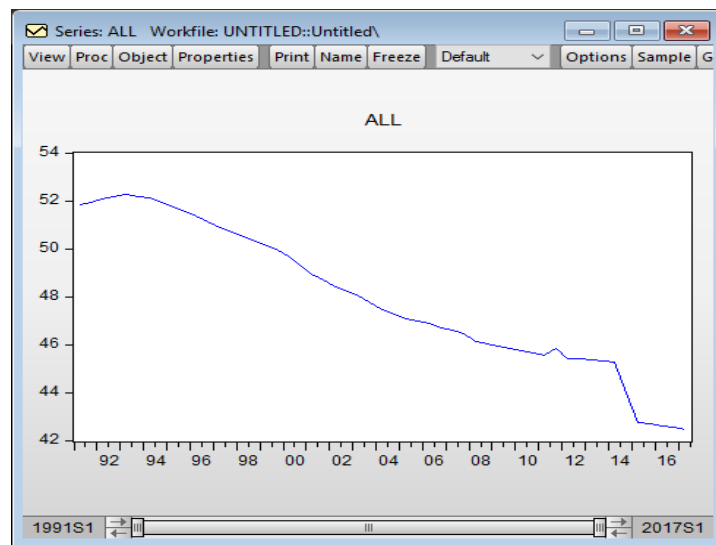


Рисунок 1. Загальна зміна динаміки чисельності населення в період з 1991 по 2017 рр.

Була розглянута можливість опису чисельності населення за допомогою моделі авторегресії з ковзним середнім. Авторегресійні моделі найпростіші за своєю структурою, але досить часто дають можливість досягти високої

адекватності досліджуваного процесу. Побудову моделей необхідно починати з моделей найнижчих порядків, які часто мають прийнятну адекватність процесу і забезпечують високу якість прогнозу. Використовуючи пакет Eviews було побудовано ряд моделей.

На рис. 2 зображена отримана модель AP(1).

Як бачимо, коефіцієнт множинної детермінації має досить гарне значення (0,992), сума квадратів похибок досить висока (3,219), статистика Дарбіна-Уотсона замала (1,242). Таким чином структуру моделі треба уточнити.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.803982	0.585555	-1.373027	0.1759
C(2)	1.012988	0.012172	83.21970	0.0000
R-squared	0.992832	Mean dependent var		47.83765
Adjusted R-squared	0.992689	S.D. dependent var		2.967405
S.E. of regression	0.253731	Akaike info criterion		0.132619
Sum squared resid	3.218973	Schwarz criterion		0.207667
Log likelihood	-1.448089	Hannan-Quinn criter.		0.161390
F-statistic	6925.519	Durbin-Watson stat		1.242308
Prob(F-statistic)	0.000000			

Рисунок 2. Модель AP(1) для чисельності всього населення України

Перевіримо характеристики якості однокрокового прогнозу – результати відображені на рис. 3.

СеКП=0,644; САП=0,517; САПП=1,070; U=0,0067.

Тобто середньоквадратична похибка (СеКП), середня абсолютна похибка (САП), середня абсолютна похибка в процентах (САПП) та коефіцієнт Тейла свідчать про загальну придатність моделі для прогнозування.

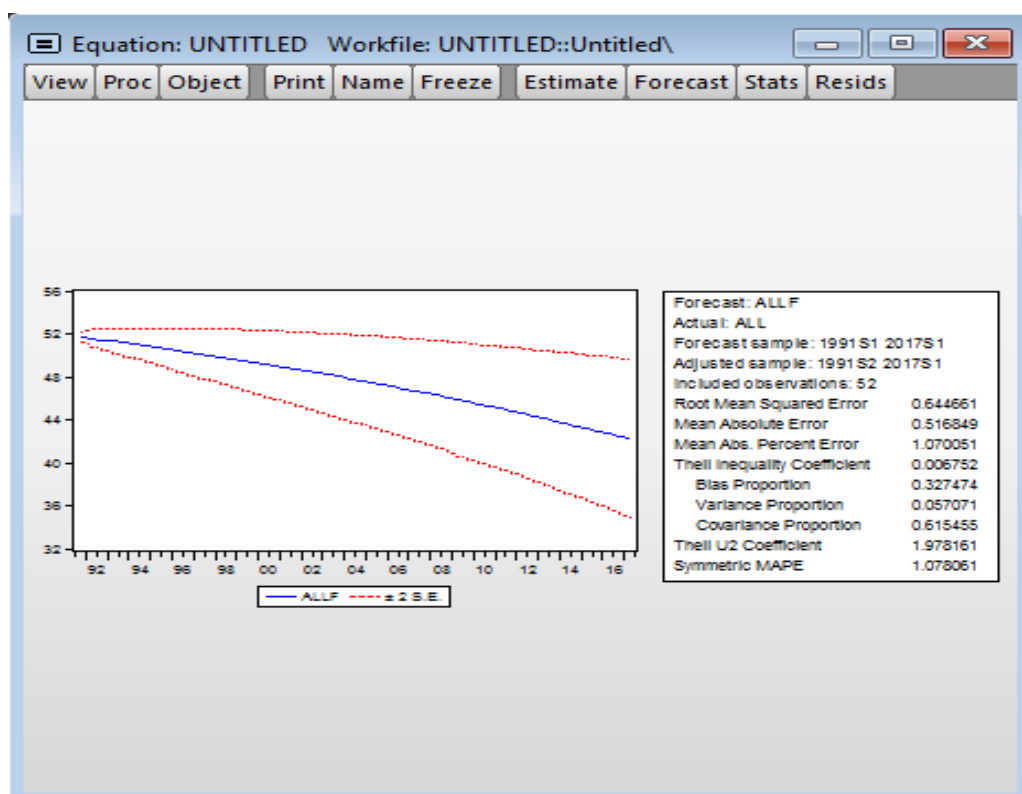


Рисунок 3. Характеристика якості прогнозу моделі AP(1)

Отже, були побудовані такі моделі:

- 1) AP(1):  $y(k) = c(1) + c(2) * y(k - 1)$ ;
- 2) AP(2):  $y(k) = c(1) + c(2) * y(k - 1) + c(3) * y(k - 2)$ ;
- 3) AP(3):  $y(k) = c(1) + c(2) * y(k - 1) + c(3) * y(k - 2) + c * y(k - 3)$ ;
- 4) AP(5):

$$y(k) = c(1) + c(2) * y(k - 1) + c(3) * y(k - 2) + c(4) * y(k - 3) + c(5) * y(k - 4) + c(6) * y(k - 5)$$

Для вибору моделі прогнозування динаміки народонаселення, порівняємо отримані моделі за статистичними характеристиками та виберемо найкращу.



Отримані результати, а також характеристики якості прогнозів занесені до Таблиці 1.

Таблиця 1

ALL	Коефіцієнт множинної детермінації, R <sup>2</sup>	Сума квадратів похибок, SSR	Статистика Дарбіна-Уотсона, DW	Середньо квадратична похибка, СеКП	Середня абсолютна похибка, САП	Середня абсолютна похибка в процентах, САПП	Коефіцієнт Тейла, U
AP(1)	0,992	3,219	1,242	0,644	0,517	1,070	0,0067
AP(2)	0,993	2,713	1,848	0,627	0,494	1,032	0,0066
AP(3)	0,994	2,545	1,990	0,555	0,435	0,916	0,0058
<b>AP(5)</b>	<b>0,994</b>	<b>2,384</b>	<b>1,944</b>	<b>0,511</b>	<b>0,385</b>	<b>0,825</b>	<b>0,0054</b>

Таким чином, отримані результати моделювання загальної чисельності населення України свідчать про високу адекватність отриманих моделей і якість прогнозів. Найкращий результат стосовно однокрокового прогнозування дає модель AP(5), для якої САПП=0,825.

Отримана модель:

$$y(k) = -0,462 + 1,431 * y(k-1) - 0,727 * y(k-2) + 0,39 * y(k-3) - 0,255 * y(k-4)$$

Для оцінювання якості прогнозу візьмемо значення за період з 1991 по 2015 роки і обчислимо прогноз на наступні три півріччя. Результати занесені до Таблиці 2.

Таблиця 2

Спостереження	Реальне значення	Прогноз	Похибка
2016S1	42,617345	41,40987	1,207475
2016S2	42,53978	41,05677	1,48301
2017S1	42,462218	40,69577	1,766448

САП	6,777653
САПП,%	15,96

Середня абсолютна похибка в процентах склала 15,96%, що свідчить про прийнятну адекватність побудованої моделі.

### **Висновки**

Таким чином, можна зробити висновок, що авторегресійна модель є досить простою з точки зору побудови та об'єму обчислювальних операцій і, при цьому, модель процесу та прогнозу являється адекватною

На основі вхідних статистичних даних було виконано прогнозування загальної зміни чисельності населення України, були розраховані критерії якості прогнозів і проведено їх аналіз. Найкращий результат було отримано для моделі AP(5). Чисельність населення України поступово скорочується. Для збільшення кількості населення, або хоча б підтримки кількості населення на стабільному рівні необхідно покращити економічний, політичний, соціальний та екологічний стани країни.

### **Література**

1. Новоселова С.В., Денисенко М.Б. Основи демографії. – Мінськ: Альтіора, 2012 – 133стор.
2. Tsay R.S. Analysis of financial time series. – New Jersey, Hoboken: CRC Press, 2010. – 715 p.
3. Бідюк П.І. Аналіз часових рядів (навчальний посібник) / Бідюк П.І., Романенко В.Д., Тимошук О.Л. – К.: Політехніка, 2010. – 317 с.
4. Бідюк П.І. Аналіз та моделювання економічних процесів перехідного періоду. – Київ: Інститут прикладного системного аналізу при НТУУ «КПІ», 2000. – 230 с.
5. Бидюк П.И. Системный подход к построению регрессионной модели по временным рядам / Бидюк П.И., Баклан И.В. // Системні дослідження та інформаційні технології, 2002 – с. 114-131.

**Прогнозування ФЕП статистично-ймовірнісними методами та комбінування оцінок прогнозів**

Пінчук В.О., Шишко Н.В.

Науковий керівник: д.т.н., проф. П.І. Бідюк

**Вступ**

Для прийняття коректних, об'єктивних рішень в економіці та фінансах необхідно використовувати сучасні методи математичного моделювання фінансово-економічних процесів (ФЕП) і чітко формувати альтернативи на основі оцінок прогнозів (та/або їх комбінування), отриманих за побудованими моделями. Більшість процесів сучасної економіки мають нелінійний нестационарний характер, що зумовлює високу складність обчислень та побудови моделей, незалежно від рівня розвитку економіки. Це пояснюється наявністю тісних структурних взаємозв'язків між економіками різних країн, значними коливаннями цін на енергоносії, нестабільністю курсів валют, значними коливаннями попиту на промислові товари внаслідок нестабільного рівня життя, впливом несприятливих погодних умов на сільськогосподарське виробництво та деякими іншими факторами. Залежність усіх сфер нашого життя від ФЕП є очевидною.

На даний момент вже існує чимало розробок у цій галузі, зокрема дослідженням присвячені роботи Бокса Дж., Дженкинса Г., Боровікова В.П., Івченко Г.І., Лукашина Ю.П., Оссовського С. [2]. Тож крім проблеми розробки нових якісних методів аналізу та прогнозування часових рядів фінансово-економічної сфери виникає проблема вибору методів дослідження з уже наявних, а також постає задача комбінування отриманих оцінок прогнозів.

Одним із найпопулярніших підходів до прогнозування динаміки процесів в економічних, фінансових, соціально-економічних та інших видах систем є прогнозування на основі моделей, побудованих за статистичними даними [2]. Воно використовується для оцінювання коротко- і середньострокових прогнозів інвестицій та доходів, формування бюджетів держави, регіонів і підприємств, об'ємів виробництва та нагромадження продукції, оцінювання альтернативних фінансово-економічних стратегій розвитку, прогнозування та менеджменту

ризиків довільної природи і розв'язання багатьох інших завдань планування й управління.

Дослідження присвячене аналізу альтернативних методів для прогнозування фінансово-економічних процесів, а також особливостям комбінування оцінок прогнозів, отриманих за цими методами.

## **Постановка задачі**

### **1. Огляд використаних методів прогнозування**

#### **Регресійний метод**

На сьогодні регресія отримала широке застосування, включаючи задачі прогнозування та керування. Метою регресійного аналізу є визначення залежності між основною змінною та множиною незалежних факторів (регресорів). При цьому коефіцієнти регресії можуть визначатися за методом найменших квадратів, його модифікаціями чи методом максимальної правдоподібності, залежно від структури моделі [1].

Модель авторегресії (АР) та модель ковзного середнього (КС) є одними із тих, що застосовуються найчастіше. Для досягнення більшої точності при моделюванні часто доцільно об'єднати їх в одній моделі - така модель позначається АРКС(р, q) та називається авторегресією із ковзним середнім. Саме вона використовується в дослідженні. Для довільного процесу АРКС(р,q) оцінку умовного прогнозу можна записати наступним чином:

$$E_k[y(k+s)] = a_0 + \sum_{i=1}^p a_i E_k[y(k+s-i)] .$$

Існує множина критеріїв, які визначають якість побудованої моделі і якість прогнозу. Нижче наведено деякі з них, які використано в даній статті.

#### **1) Інформаційний критерій Акайке (Akaike information criterion, AIC)**

Це міра відносної якості статистичних моделей для заданого набору даних. Для лінійної моделі множинної регресії значення критерію розраховується за такою формулою:

$$AIC = \ln \left( \frac{\sum_{i=1}^k \varepsilon_i^2}{n} \right) + \frac{2k}{n},$$

де  $n$  – кількість спостережень;  $k$  – кількість параметрів моделі,  $\sum_{i=1}^k \varepsilon_i^2$  – сума квадратів залишків моделі, отримана при оцінці коефіцієнтів моделі методом найменших квадратів.

При збільшенні кількості пояснюючих змінних перший доданок у правій частині зменшується, а другий збільшується. Таким чином, критерій не тільки винагороджує за якість наближення, але і штрафує за використання зайвої кількості параметрів моделі. Серед кількох альтернативних моделей перевага віддається тій, у якої значення  $AIC$  менше.

## 2) Коефіцієнт детермінації

Коефіцієнт детермінації розглядають, як правило, в якості основного показника, що відображає міру якості регресійної моделі, яка описує зв'язок між залежною і незалежними змінними моделі. Коефіцієнт детермінації показує, яка частка варіації пояснюючої змінної  $y$  врахована в моделі і обумовлена впливом на неї факторів, включених в модель:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})}{\sum_{i=1}^n (y_i - \bar{y})},$$

де  $y_i$  – значення спостережуваної змінної;  $\bar{y}$  – середнє значення по спостережуваних даних;  $\hat{y}$  – модельні значення, побудовані за оціненими параметрами. Чим ближче  $R^2$  до 1, тим вища якість моделі.

## 3) Критерій Дарбіна-Уотсона

Критерій Дарбіна-Уотсона (чи DW-критерій) — статистичний критерій, що використовується для знаходження автокореляції залишків першого порядку регресійної моделі. Обчислюється за формулою:

$$DW = 2 - 2\rho,$$

де  $\rho$  – коефіцієнт автокореляції першого порядку і дорівнює:

$$\rho = \frac{\frac{1}{N-1} \sum_{k=2}^N [e(k) - \bar{e}][e(k-1) - \bar{e}]}{\sigma_e^2},$$

при цьому:

$$\sigma_e^2 = \frac{1}{N-1} \sum_{k=1}^N [e(k) - \bar{e}]^2,$$

$$e(k) = y(k) - \overline{y(k)}.$$

### Метод експоненційного згладжування

В основі методів експоненційного згладжування (ЕЗ) закладена ідея постійного перегляду прогнозних значень по мірі надходження нових даних. Модель ЕЗ присвоює експоненційно спадаючі ваги спостереженням по мірі їх «старіння». Таким чином, останні доступні спостереження мають більший вплив на прогнозне значення, ніж старі спостереження.

Функція ЕЗ має вигляд

$$y(k) = \hat{y}(k) + \varepsilon(k),$$

$$\hat{y}(k) = \alpha y(k-1) + (1-\alpha) \hat{y}(k-1),$$

де  $\alpha$  – коефіцієнт згладжування. Слід звернути особливу увагу на вибір даного коефіцієнта, адже він характеризує швидкість реакції моделі, а також її можливість згладжувати випадкові відхилення. Коефіцієнт слід обирати, опираючись на властивості конкретного ряду.

Якісним вибором початкових умов не варто нехтувати – є кілька підходів до вибору, проте, зазвичай, значення обчислюються на основі апріорних даних ряду. В нашому випадку, початкова умова задається так:  $\hat{y}(1) = y(1)$ .

Для моделювання процесів із трендом використовується модель Хольта або ж, як іще називають, подвійне ЕЗ. У цьому випадку в моделі потрібно розглядати рівень та тренд як дві окремі складові [6]:

$$y(k) = \hat{y}(k) + \varepsilon(k),$$

$$\hat{y}(k) = \alpha y(k-1) + (1-\alpha) [\hat{y}(k-1) + z(k-1)],$$

$$z(k) = \gamma [\hat{y}(k-1) - \hat{y}(k-2)] + (1-\gamma) z(k-1).$$

де  $\alpha$  – коефіцієнт згладжування рівня,  $\gamma$  – коефіцієнт згладжування тренду.

## Нейронні мережі (НМ)

Найбільш перспективним кількісним методом прогнозування є використання нейронних мереж. Однією із основних переваг НМ є гнучка структура – для її зміни у рамках визначеної архітектури достатньо регулювати кількість шарів та нейронів. Лише ці незначні перетворення надають можливість значно підвищити точність прогнозування.

Ще одна серйозна перевага НМ полягає в тому, що експерт не вибирає структуру математичної моделі поведінки часового ряду. Побудова нейромережевої моделі відбувається адаптивно під час навчання, без участі експерта. При цьому нейронній мережі надаються приклади з бази даних і вона сама налагоджується під ці дані.

Недоліком нейронних мереж є їхня недетермінованість – після навчання мережа є «чорним ящиком», який якимось чином працює, але логіка прийняття рішень прихована від експерта.

Прогнозування за допомогою НМ складається з кількох основних етапів: – збір початкових даних; – аналіз і перетворення даних; – відбір змінних.

Якість навченої НМ може бути оцінена такими показниками: – ефективність моделі; – ефективність роботи мережі; – стійкість; – стабільність.

Ефективність моделі визначає якість навченої мережі, ґрунтуючись на характеристиці, що досягається при необхідному або ідеальному виході мережі. Для обчислення ефективності моделі спочатку визначають оцінку поведінки мережі при необхідному або ідеальному виході й порівнюють результат з тим, що отриманий на виході мережі при подачі на її вхід елементів навчальної вибірки.

Ефективність роботи мережі є мірою потенціалу роботи навченої мережі. Цей показник дає оцінку дії нейронної мережі на тестовій вибірці й порівнює отриману оцінку з тією, котра досягається при використанні необхідного (ідеального) виходу.

Стійкість – характеристика навченої мережі в різні часові періоди, якими можуть бути вікна, що перекриваються або не перекриваються, та змінюють розміри. Використання ковзних вікон дає розподіл значень характеристики НМ.

Стабільність – характеристика навченої НМ, заснованої на вікнах змінних розмірів. Основна ідея зводиться до використання методу Монте-Карло для оцінки характеристики навченої мережі на багатьох тимчасових інтервалах, кожний з яких має випадкові дати початку і закінчення [5].

На сьогодні існує безліч нейропакетів, проте повністю універсальних не існує, оскільки неможливо заздалегідь передбачити всі нейронні структури, які можуть знадобитися для розв'язання тих або інших прикладних задач.

### Комбінування оцінок прогнозів

В роботі використовуються наступні способи комбінування оцінок прогнозів:

- усереднення прогнозів (вагові коефіцієнти при цьому однакові):

$$\hat{F}_k(t) = \frac{1}{n} \sum_{i=1}^n \hat{F}_i(t)$$

де  $\hat{F}_k(t)$  - комбінований прогноз;  $\hat{F}_i(t)$  - прогнози, отримані різними методами. Якщо дисперсії близькі за значенням та похибки оцінок прогнозів незалежні, то дисперсія комбінованої похибки буде значно меншою будь-якої з двох дисперсій [1].

- зважене усереднення прогнозів.

Якщо інформація щодо характеристик індивідуальних прогнозів відсутня, то можна присвоїти різні вагові коефіцієнти окремим прогнозам на основі суб'єктивних або експертних суджень:

$$\hat{F}_k(t) = \sum w_i \hat{F}_i(t)$$

де  $w_i$  – вагові коефіцієнти,  $\sum w_i = 1$ .

Але оскільки моделі, які дають менші суми квадратів похибок прогнозів, генерують якісніші прогнози, то логічно прийняти цю міру за основу для визначення вагових коефіцієнтів, що й було реалізовано. Позначимо суму квадратів похибок прогнозування (для історичного прогнозу) через

$$SSR = \sum_{k=1}^N e^2(k)$$

Тоді вираз для вагових коефіцієнтів окремих прогнозів:

$$w_i = \frac{1/SSR_i}{\sum_{k=1}^n 1/SSR_k}$$



Існує множина критеріїв, які визначають якість побудованої моделі і якість прогнозу. В роботі використовуються наступні:

- середньо-квадратична похибка (СКП) :

$$СКП = \sqrt{\frac{1}{s} \sum_{i=1}^s [y(k+s) - \hat{y}(k+s, k)]^2};$$

- середня похибка в процентах (СПП) – середнє значень похибок оцінок прогнозу в процентах відносно фактичного значення показника:

$$СПП = \frac{1}{s} \sum_{i=1}^s \frac{[y(k+s) - \hat{y}(k+s, k)]}{y(k+s)} \times 100\%;$$

- абсолютна середня похибка в процентах (АСПП) – середнє абсолютних значень похибок оцінок прогнозу в процентах відносно фактичного значення показника:

$$АСПП = \frac{1}{s} \sum_{i=1}^s \frac{|y(k+s) - \hat{y}(k+s, k)|}{|y(k+s)|} \times 100\%.$$

Оскільки ця міра характеризує відносну якість прогнозу, то її використовують, в основному, для порівняння точності прогнозів різнорідних об'єктів (процесів) прогнозування. Однак, вона є завжди корисною при виконанні порівняльного аналізу якості прогнозування одного й того ж процесу різними методами, оскільки відносна міра є чіткою і зрозумілою для дослідника і практичного користувача [2].

### **Мережа Байєса**

Також в роботі досліджується ймовірнісне прогнозування за допомогою дискретної мережі Байєса. Вона використовується для моделювання ситуацій, в яких присутня структурна чи статична невизначеність, тобто немає достатньої інформації про стан процесу в момент прийняття рішення. Мережа визначається двома компонентами – ациклічним графом та множиною параметрів, які визначають мережу. Вершинами графу є досліджувані змінні, ребра – умовні залежності між ними. Теорему Байєса розглядаємо як механізм формування висновку [3].

## 2. Результати виконання обчислювальних експериментів

Навчальна вибірка була сформована за даними, представленим на сайті з різними економічними показниками – <http://www.investing.com>. Для дослідження було обрано такі біржові індекси:

1. S&P 500 - фондовий індекс, у кошик якого включено 500 акціонерних компаній США, що мають найбільшу капіталізацію. Список належить компанії Standard & Poor's і нею ж складається.
2. Промисловий індекс Доу-Джонса (англ. Dow Jones Industrial Average) - біржовий індекс цінних паперів (акцій) 30 найбільших американських підприємств. Індекс був створений для зведення інформації по акціях індустріальних підприємств на американській біржі цінних паперів. Один з найстаріших індексів цінних паперів у США.
3. NASDAQ-100 - біржовий індекс, який відображає динаміку зміни вартості акцій 100 найбільших компаній, які присутні у лістингу біржі NASDAQ. В індексі відсутні компанії фінансового сектору.
4. Nasdaq Composite (досить часто трейдери називають його просто Nasdaq) - включає більше трьох тисяч компаній (як американських, так і іноземних), які входять до лістингу біржі NASDAQ. Акції будь-якої з цих компаній впливають на індекс пропорційно своїй ринковій вартості. І їх кількість значно перевищує кількість іноземних компаній, представлених в AMEX і NYSE. Через специфіку біржі NASDAQ індекс вважається важливим показником динаміки курсу паперів високотехнологічних і швидко зростаючих компаній.
5. NYSE Composite — біржовий індекс американських цінних паперів (акцій підприємств), є одним з найпопулярніших фондових індексів у світі. Відображає зміну вартості всіх акцій, які торгуються на Нью-Йоркській фондовій біржі. На сьогодні включає приблизно 2 000 компаній, 1 600 з яких є американськими (загальна капіталізація цих компаній становить близько 20 трильйонів доларів), а 360 представляють інші країни.

Для аналізу статистичними методами використовувався ряд NYSE Composite тому що він є має нестационарні статистичні характеристики, які є цікавими з точки зору формування структури моделі. Для побудови БМ використовувались усі 5 індексів. Варто відзначити, що в первісному аналізі брали участь більша кількість індексів, але в моделі представлені найбільш значущі, за якими і будувалася БМ. Дослідження проводилось на основі вищеписаних показників з 1 січня 2015 року по 19 квітня 2017. Робоча вибірка складається з 578 значень (у вихідні дні біржа не працює), рис. 1. В роботі імітувались та досліджувались прогнози на 2 кроки. Тому для зручності та можливості перевірки часовий ряд був штучно зменшений і використовувався без двох останніх значень. Саме це дозволяє нам перевірити якість прогнозу та побачити це наочно. Усі розрахунки проводились за допомогою мови статистичної обробки даних *R*.

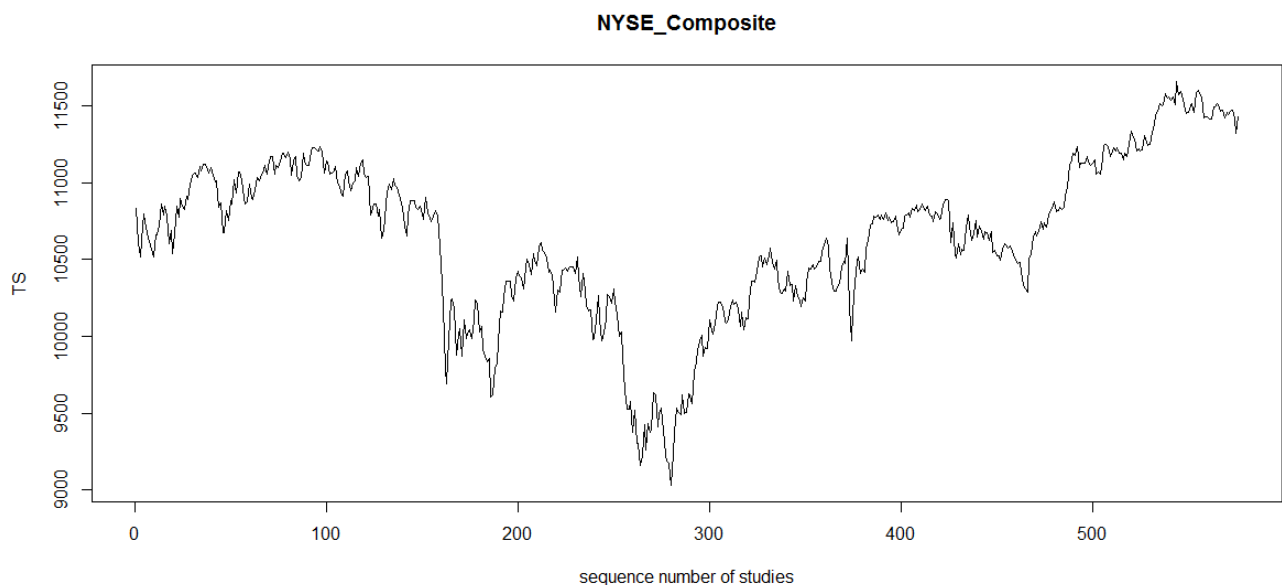


Рисунок 1. Досліджуваний ряд NYSE\_Composite

Одним із способів перевірки стаціонарності процесу є тест Дікі-Фуллера, котрий належить до виду тестів на «одичні корені». Ряд має одиничний корінь (порядок інтегрованості 1), якщо його перші різниці утворюють стаціонарний ряд. Це можна формально записати так  $y(t) \sim I(1)$ . Суть тесту полягає у перевірці гіпотези  $H_0$  про рівність одиниці коефіцієнта  $\alpha$  рівняння:

$$\Delta y(k) = (\alpha - 1)y(k-1) + \varepsilon(k).$$

де  $\Delta y(k) = y(k) - y(k-1)$  – оператор різниці першого порядку (різницю порядку  $d$  можна отримати послідовним застосуванням оператора першої різниці  $d$  разів). Дана статистика має розподіл Дікі-Фуллера. Якщо значення статистики лежить лівіше критичного значення при заданому рівні значущості, то гіпотеза про одиничний корінь відхиляється і процес визнається стаціонарним (в сенсі даного тесту). В таблиці 3.1 наведені результати проведеного тесту Дікі-Фуллера для ряду NYSE\_Composite.

Таблиця 3.1 – Результати тесту Дікі-Фуллера для ряду NYSE\_Composite

DF-статистика	$H_0$	<i>p-value</i>
-1,5753	процес нестаціонарний	0,7581

Таким чином, імовірність зробити помилку I-го роду відхиливши гіпотезу про нестаціонарність становить 0,7581. При побудові регресійної моделі було отримано наступні варіанти моделей:

Таблиця 1 - Запропоновані моделі для часового ряду

Модель	Критерій Акайке
ARIMA(2,1,2)	6802,148
ARIMA(0,1,0)	6804,734
ARIMA(1,1,0)	6801,528
ARIMA(0,1,1)	6806,747
ARIMA(0,1,0)	6802,796
ARIMA(2,1,0)	6802,489
ARIMA(1,1,1)	6802,941
ARIMA(2,1,1)	6804,389
ARIMA(1,1,0)	6799,653
ARIMA(2,1,0)	6800,654
ARIMA(1,1,1)	6801,089
ARIMA(2,1,1)	6802,537

Варто відмітити тип отриманих моделей – ARIMA(p,d,q), де p – авторегресійні доданки; d – різниці; q – доданки ковзного середнього. Вони дають змогу значно розширити множину процесів для моделювання та отримати

повнішу інформацію про досліджуваний ряд. Найкращу модель ARIMA(1,1,0) вибираємо виходячи з умови мінімуму інформаційного критерію Акайке. Розраховуємо для неї інші статистичні критерії.

Таблиця 2 – Значення критеріїв для запропонованої моделі

Модель	Статистика Дарбіна-Уотсона	Коефіцієнт детермінації
ARIMA(1,1,0)	1,9816	0,9707

Бачимо, що значення обох критеріїв близькі до ідеальних, тому дали сміливо використовуємо запроповану модель ARIMA(1,1,0). Результати прогнозування бачимо на рисунку нижче.

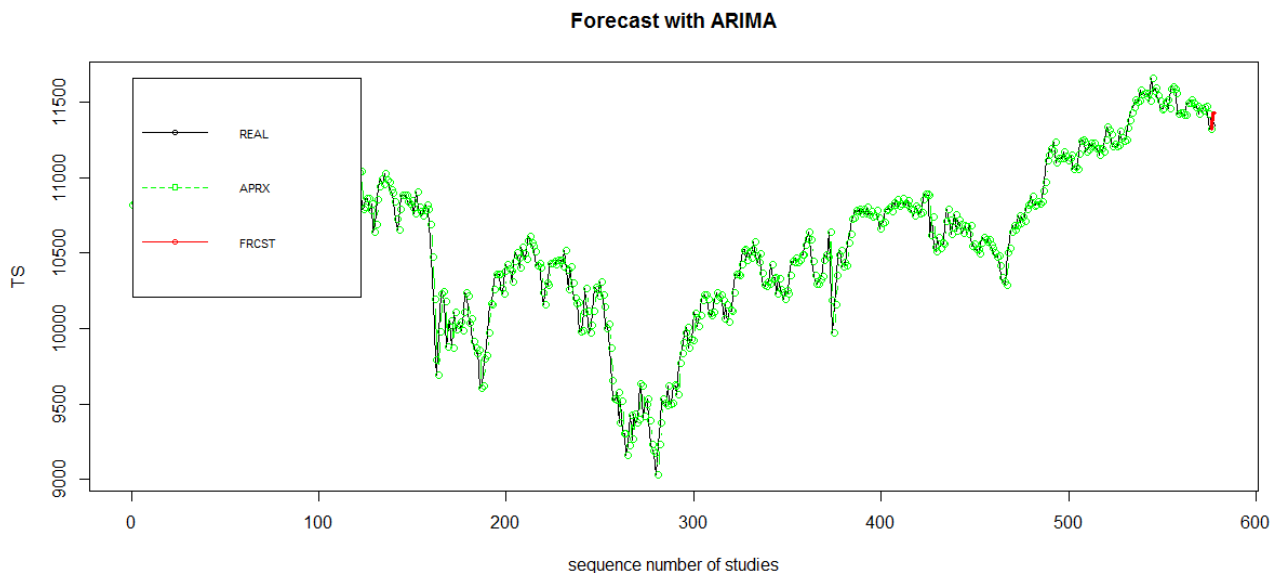


Рисунок 2 - Прогнозування на основі регресійного аналізу

Отримані такі значення критеріїв: СКП= 68,63; СПП = 0,69; САПП = 0,69.

Результати застосування експоненційного згладжування

Отримані результати прогнозування на основі цього методу подано нижче.

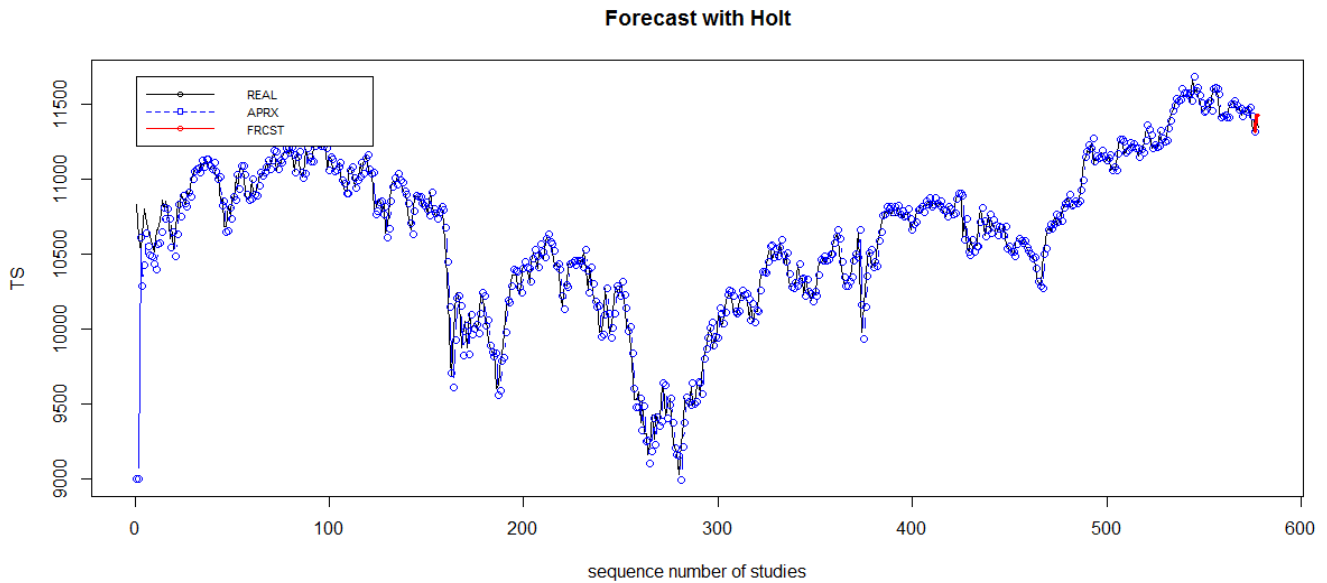


Рисунок 3– Прогноз методом експоненційного згладжування

Варто відзначити, що перші два значення ряду використані для обчислення початкового значення, саме тому, відновлення ряду починається з третього значення, а першим двом точкам, для зручності зображення присвоєно нульові значення. Отримані такі значення критеріїв: – СКП= 49,42; – СПП = 0,43; – САПП =0,43.

#### Результати застосування нейромережі

Для прогнозування за допомогою НМ була створена НМ прямого поширення (перцептрон) з одним прихованим шаром та заданою кількістю вузлів. За замовчуванням кількість вузлів обирається як одна друга розмірності вхідного вектору плюс ще одне значення., проте при багаторазових дослідженнях, в нашому випадку, це виявилось неоптимальним. Проте ми взяли 45 вузлів, оскільки при більшій кількості якість прогнозу практично не змінювалась, а швидкодія помітно знижувалась. Початкові ваги обираються випадково. Далі, при обчисленні прогнозу, вони усереднюються. Мережа постійно навчається на один крок прогнозу. Модель, яка використовується, називається NNAR (p, P). Вона є аналогом ARIMA (p,0,0) (P, 0,0) моделі, але з нелінійними функціями. Отримані такі результати:

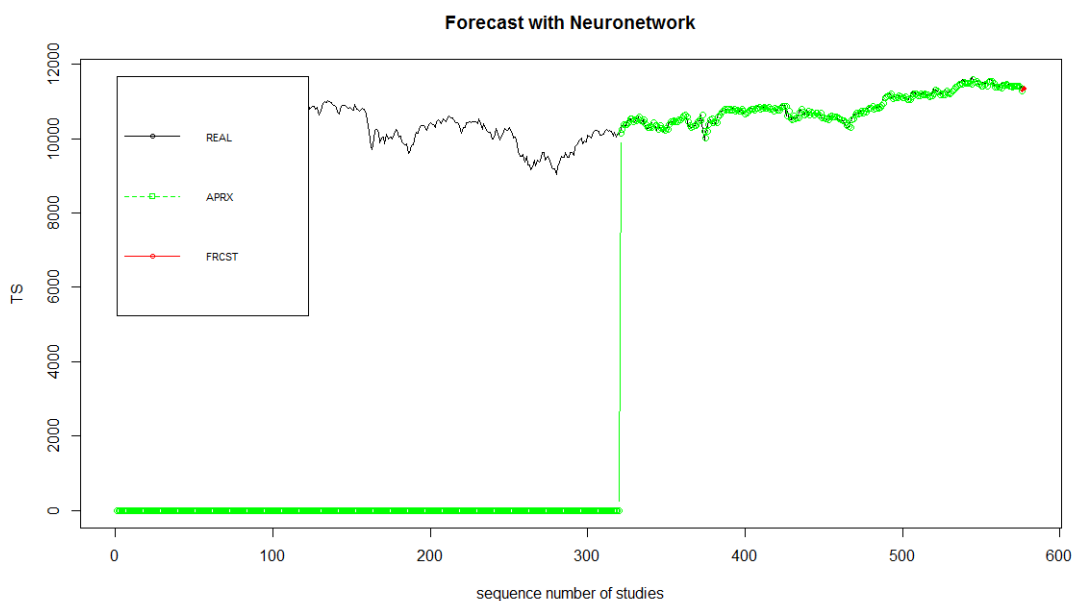


Рисунок 4 – Прогнозування за допомогою НМ

Отримані такі значення критеріїв: – СКП= 65,26; – СПП = -0,29; – САПП = 0,49.

На основі отриманих результатів прогнозування розглядались усі можливі комбінації прогнозів та порівнювались критерії якості. Фактичні величини (взяті з ряду вимірів) для прогнозованих значень:

- 18.04.2017 – 11378,58;
- 19.04.2017 – 11342,42.

Результати першого способу комбінування подано у таблиці 3:

Таблиця 3 - Усереднення отриманих прогнозів

№	Комбінації методів	Y1	Y2	СКП	СПП	САПП
1	ARMA	11440,11	11438,67	80,78	-0,69	0,69
2	ES	11436,53	11381,49	49,42	-0,43	0,43
3	NN	11355,83	11431,86	65,26	-0,29	0,49
4	ARMA+ES	11438,32	11410,08	63,82	-0,56	0,56
5	ARMA+NN	11397,97	11435,27	67,07	-0,49	0,49
6	ES+NN	11396,18	11406,68	47,11	-0,36	0,36
7	ARMA+ES+NN	11410,82	11417,34	57,67	-0,47	0,47

З розрахованої вище таблиці бачимо, що найкращий прогноз досягається при комбінуванні методів прогнозування за допомогою методу експоненційного згладжування та НМ (САПП = 0,36). Також високоякісний прогноз отримуємо за допомогою НМ. Варто відзначити, що комбінування оцінок будь-яких з наведених методів дає кращі результати, аніж використання лише регресійної моделі або моделі експоненційного згладжування.

Далі розглядаємо метод зваженого середнього прогнозів. Розраховані значення ваг оцінок методів для усіх можливих варіантів комбінування представлені нижче в таблиці 4.

Таблиця 4 - Розраховані ваги оцінок методів прогнозування

№	Комбінації методів	ARMA	ES	NN
1	ARMA+ES	0,39	0,61	
2	ARMA+NN	0,44		0,56
3	ES+NN		0,57	0,43
4	ARMA+ES+NN	0,15	0,45	0,40

На основі цих розрахунків проведено комбінування, а результати подані в таблиці 5.

Таблиця 5 - Зважене комбінування прогнозів

№	Комбінації методів	Y1	Y2	СКП	СПП	САПП
1	ARMA+ES	11437,93	11403,79	60,36	-0,53	0,53
2	ARMA+NN	11392,91	11434,86	66,14	-0,47	0,47
3	ES+NN	11401,83	11403,15	45,98	-0,36	0,36
4	ARMA+ES+NN	11404,79	11410,22	51,39	-0,41	0,41

Аналізуючи отримані критерії якості, бачимо, що в переважній більшості випадків комбінування оцінок прогнозів значно покращує якість прогнозу. Зокрема в нашому випадку найякісніший прогноз отримуємо, зважено скомбінувавши метод експоненційного згладжування та прогноз за допомогою



НМ: похибки при цьому менші за всі інші комбінації та окремі прогнози: – СКП = 45,99; – СПП = -0,37; – САПП = 0,37.

Також варто взяти до уваги хороші результати, отримані за рахунок комбінування усіх розглянутих способів прогнозування. Отримані критерії якості значно кращі ніж без використання комбінування, при чому для всіх методів.

Проте бачимо, що порівнюючи між собою два способи комбінування не можна однозначно виділити кращий. В більшості випадків точніші результати отримані за рахунок зваженого комбінування оцінок прогнозів (вагові коефіцієнти обчислюються за допомогою сум квадратів похибок оцінок прогнозів для кожного методу), проте в деяких – за використанням способу усереднення оцінок. Це насамперед, залежить від методів прогнозування, які комбінуються, та коректності і доцільності їх використання для конкретного ряду.

### **Прогнозування за допомогою МБ**

Для побудови мережі використовувалися статистичні дані за проміжок часу, аналогічний до попередніх досліджень для п'яти індексів, які описані вище. Вибірка була поділена на навчальну (527 випадків) та перевірочну (50 випадків).

Дослідження зручно проводити, розглядаючи величину відхилення індексу від його попереднього значення [4]. Саме такий підхід використаний в роботі.

$$Y_{k,t}^{MБ} = Y_{k,t} - Y_{k-1,t}$$

де  $i = \overline{1,5}$  – номер ряду,  $Y_{k,t}$  – історичні дані за період  $k$  для ряду  $i$ ,  $k = \overline{2,n}$ , де  $n$  – кількість досліджень (розмірність вибірки).

Перед початком побудови моделі проводимо процедуру дискретизації даних. В нашому випадку проводиться ієрархічна дискретизація. Кількість проміжків дискретизації обиралась емпіричним шляхом. Результат дискретизації на прикладі одного індексу зображено нижче, на рисунку 5.

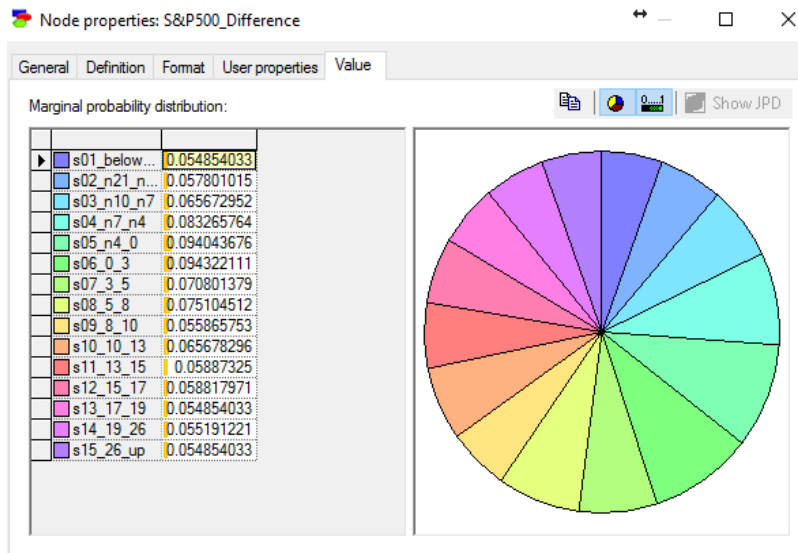


Рисунок 5 – Дискретизація значень ряду різниць індексу S&P500

Далі вибирається відповідний алгоритм навчання мережі у разі необхідності використовується експертне знання і виконується навчання мережі.

При побудові структури мережі Байеса в програмі GeNIe слід пам'ятати, що обраний алгоритм впливає на швидкість і на саму побудовану структуру. Найшвидшим реально є алгоритм Greedy Thick Thinning, його і будемо використовувати для аналізу прикладів. У результаті роботи алгоритму отримуємо тільки одну структуру, яка є логічною і оптимальною (рис 6).

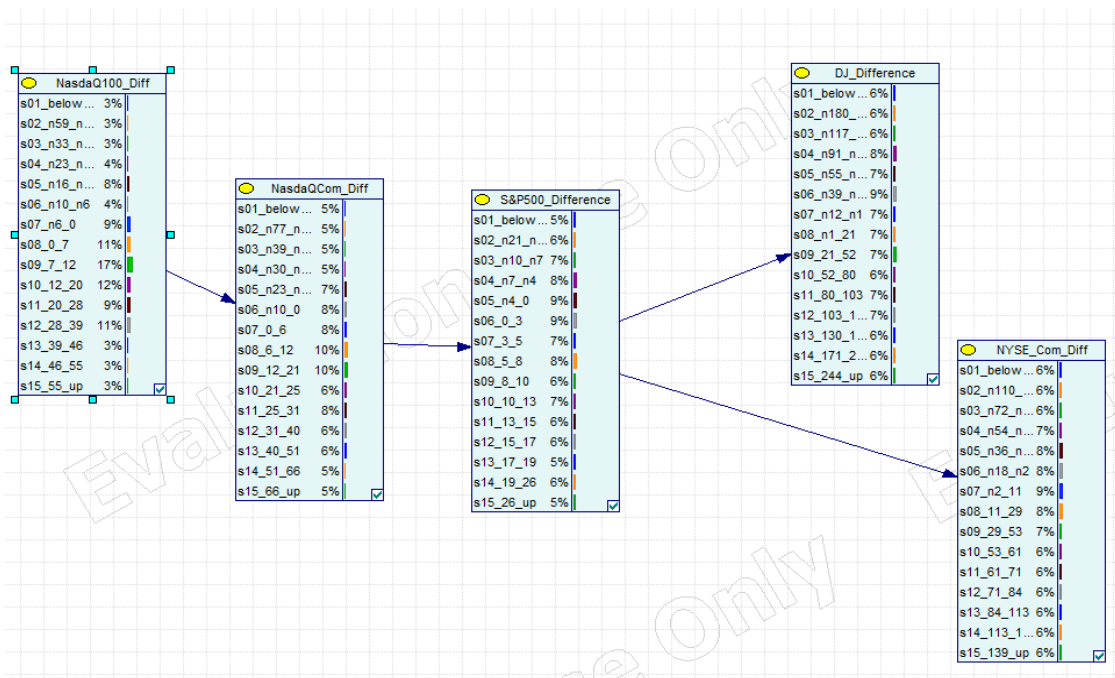


Рисунок 6 – Структура мережа Байеса у вигляді гістограм вибірки

По структурі отриманої мережі робимо висновок щодо індексу для прогнозування – Nasdaq100. На рисунку подано результат одного з експериментів.

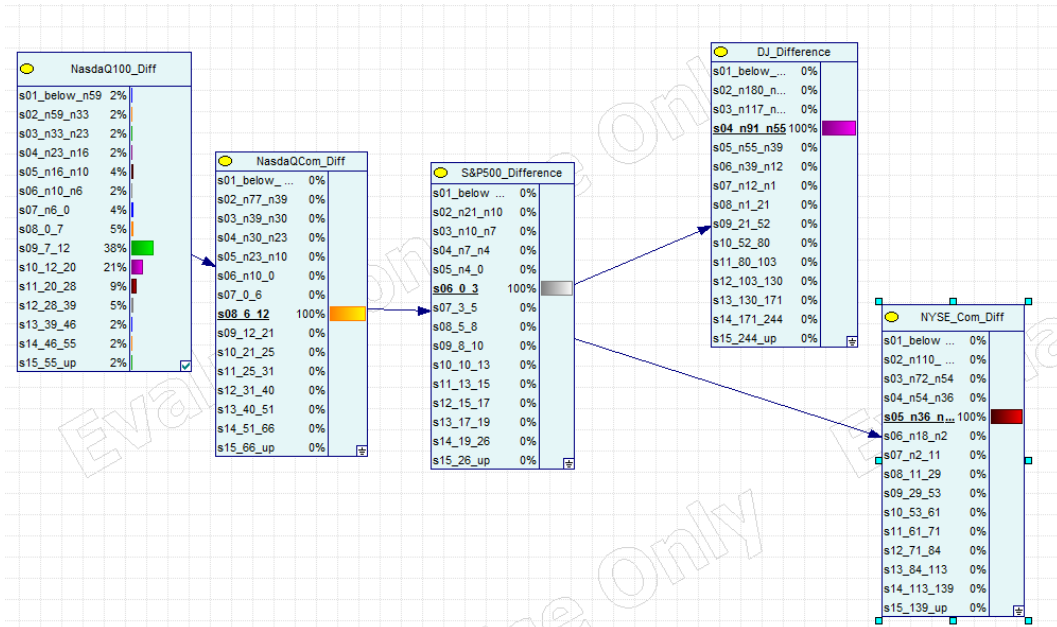


Рисунок 7 - Результат прогнозування зміни індексу Nasdaq100 для 19.04.2017р

На рис. 8 показана діаграма зміни прогнозного і реального відхилень індексу на останніх 50 часових інтервалах.

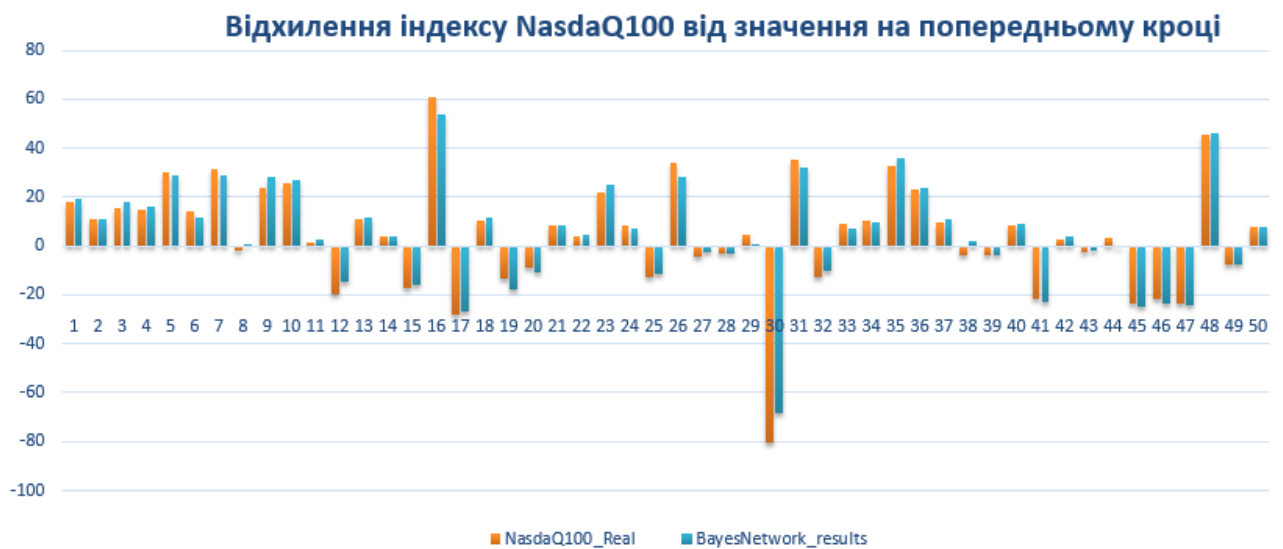


Рисунок 8 – Відхилення індексу Nasdaq100 від значення на попередньому кроці

Аналіз отриманих результатів свідчить, що у 94% випадків було точно передбачене зростання або падіння значення обраного індексу. В 86% випадків прогнозне і реальне значення відхилення індексу Nasdaq100 від значення в попередній момент часу мають відхилення менше 10%. При цьому 10% експериментів мають відхилення більше 15% – така ситуація характерна для малих відхилень та може бути пов'язана з варіантом дискретизації змінної.

### **Висновки**

Отже, в роботі показано актуальність та важливість знаходження високоякісного розв'язку загальної задачі прогнозування на сьогодні. Виконано огляд таких методів прогнозування: на основі регресійних рівнянь, методи експоненційного згладжування, прогнозування за допомогою нейронних мереж та мереж Байєса. Виконано комбінування оцінок прогнозів, отриманих статистичними методами за допомогою двох різних підходів до комбінування. Усі експерименти проводились на основі показників фінансових індексів S&P 500, Доу-Джонса, NASDAQ100, Nasdaq Composite, NYSE Composite в період з січня 2015 року по квітень 2017.

Встановлено, що для комбінування оцінок прогнозів найефективніше використовувати метод зваженого середнього, комбінуючи при цьому оцінки, отримані за допомогою нейронної мережі та методу експоненційного згладжування. Саме при такому способі комбінування досягаються найкращі значення критеріїв якості прогнозу. Проте не можна гарантувати, що для іншого часового ряду зазначена комбінація буде найкращою. При використанні моделей у формі МБ отримали, що у 86% випадків прогнозне і реальне значення відхилення індексу Nasdaq100 від значення в попередній момент часу мають відхилення менше 10%, що є дуже хорошим результатом.

Для удосконалення роботи та можливостей подальших досліджень необхідно реалізувати й інші методи моделювання і прогнозування часових рядів та, можливо, модифікувати обрані з метою розширення класів процесів, до яких їх можна застосовувати. А також необхідно звернути увагу на вибір кількості проміжків дискретизації та, за можливості, збільшити навчальну вибірку для мережі.

## Література

1. Бідюк П.І. Аналіз часових рядів – К.: ННК ІПСА, 2006. –188 с.
2. Бідюк П.І. Системний підхід до прогнозування на основі моделей часових рядів // *Системні дослідження та інформаційні технології*, 2003, №3, с. 88 – 110.
3. Бидюк П.И., Терентьев А.Н., Коршевнюк Л.А. Прогнозирование экономических показателей при помощи сетей Байеса
4. Бидюк П.И., Терентьев А.Н., Гасанов А.С. Построение и методы обучения Байесовских сетей // *Кибернетика и системный анализ*. – 2005. – 4. – С. 133-147.
5. Хайкин С. Нейронные сети: полный курс. / Хайкин С.– М.: ООО «И. Д. Вильямс», 2006.– 1104 с.
6. Prajakta S.K. Time series Forecasting using Holt-Winters Exponential Smoothing / Kanwal Rekhi School of Information Technology Journal [Електронний ресурс]. 2004. – 13 р. режим доступу: [http://www.it.iitb.ac.in/~praj/acads/seminar/04329008\\_ExponentialSmoothing.pdf](http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf)

**ЗАСТОСУВАННЯ МЕТОДОЛОГІЇ АНАЛІЗУ ВИЖИВАННЯ ДЛЯ  
ДОСЛІДЖЕННЯ СПОЖИВЧИХ РИЗИКІВ**

**Кузнєцова Н.В., Куца К.В., Штогрін С.Р.**

**Вступ**

Останнім часом в Україні значно зросла вимогливість банків і необхідність застосування ними нових моделей, інструментів та технологій для напрацювання оптимальної кредитної політики. Це зумовлено фінансовою кризою, істотною девальвацією національної валюти, вибірковістю клієнтів, які все менше звертаються за споживчими кредитами та намагаються отримати найвигідніші умови. Банками все частіше поряд з задачами аплікаційного скорингу (видачі кредитів) вирішуються задачі прогнозування поведінки клієнта до моменту повернення ним кредиту [1, 2]. Це задача поведінкового (біхевіоріального) скорингу, де в якості вхідних даних використовуються кредити, по яким ще не закінчився термін повернення, тобто виконується прогнозування поведінки та дій клієнта у наступні періоди [3]. Такий аналіз дозволяє банками вчасно приймати управлінські рішення, наприклад, збільшуючи кредитний ліміт по кредитним картам клієнтів, що вчасно повертають кредитні кошти, і навпаки, застосовувати обмеження по клієнтам, у яких наявні прострочки та заборгованості.

**Постановка задачі**

Метою даної статті є дослідження можливості застосування методів інтелектуального аналізу даних для біхевіоріального (поведінкового) скорингу, побудова моделей на основі існуючих методів та розробка нових поведінкових моделей з можливістю прогнозування часу настання дефолту для споживчих кредитів на основі обраного математичного апарату з теорії аналізу виживання.

**Загальні припущення теорії аналізу виживання**

Доступною інформацією для моделювання ймовірності дефолту (PD) за кредитом є вибірка з  $n$  незалежних однаково розподілених випадкових величин  $\{(Y_1, X_1, \delta_1), \dots, (Y_n, X_n, \delta_n)\}$ , випадкового вектора  $\{Y, X, \delta\}$ , де  $Y = \min\{T, C\}$  - спостережувана зрілість,  $T$  - час до дефолту,  $C$  - час до кінця дослідження або

передбачуваного анулювання кредиту,  $\delta = (T \leq c)$  - індикатор не цензурування і  $X$  являє собою вектор пояснюючих коваріат. У цих умовах аналізу виживаності ми будемо припускати, що існує невідоме співвідношення між  $T$  і  $X$ . Ми також будемо вважати, що випадкові величини  $T$  і  $C$  є умовно незалежними для даного  $X$  [4].

У попередніх припущеннях можна повністю охарактеризувати умовний розподіл випадкової величини  $T$ , використовуючи деякі загальні співвідношення в аналізі виживання. Таким чином, функція умовного виживання,  $S(t | x)$ , умовний ступінь небезпеки,  $\lambda(t | x)$ , умовна кумулятивна функція ризику,  $\Lambda(t | x)$ , умовна функція розподілу  $F(t | x)$  пов'язані наступним чином:

$$S(t | x) = P(T > t | X = x) = \int_t^{\infty} f(u | x) du$$

$$\lambda(t | x) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, X = x)}{\Delta t} = \frac{f(t | x)}{S(t | x)}$$

$$\Lambda(t | x) = \int_0^t \lambda(u | x) du = \int_0^t \frac{f(u | x)}{S(u | x)} du$$

$$S(t | x) = e^{-\Lambda(t|x)}$$

$$F(t | x) = 1 - S(t | x).$$

Відповідно до вимог Базеля II [6] моделі кредитного скорингу використовуються для вимірювання ймовірності дефолту в горизонті часу  $t + b$  з часом зрілості  $t$ . Типове значення  $b = 12$  (в місяцях). Таким чином, повинна бути обчислена наступна ймовірність:

$$PD(t | x) = P(t \leq T < t + b | T \geq t, X = x) = \frac{P(T < t + b | X = x) - P(T \leq t | X = x)}{P(T \geq t | X = x)} = \frac{F(t + b | x) - F(t | x)}{1 - F(t | x)} = 1 - \frac{S(t + b | x)}{S(t | x)} \quad (1)$$

де  $t$  - спостережуваний термін погашення кредиту, а  $x$  - значення коваріаційного вектору  $X$  для цього кредиту.

### Моделі пропорційних ризиків

У нашій статті ми використовували підхід пропорційних ризиків Кокса до моделі функції умовного виживання  $S(t | x)$ . Головною ідеєю цього методу є оцінка сукупної умовної функції ризику,  $L(t | x)$ , з використанням максимальної

правдоподібності. Метою є розробка умовної моделі для індивідуального  $S(t|x)$ , яка визначена в термінах  $L(t|x)$ . Для того, щоб описати  $\hat{PD}^{PHM}$  (ймовірність дефолту при моделі пропорційних ризиків) ми визначимо такі вирази по відношенню до теорії регресії Кокса [2].

Оцінка функції умовної швидкості небезпеки визначається наступним чином:

$$\hat{\lambda}(t|x) = \hat{\lambda}_0(t) \exp(x^T \hat{\beta}),$$

де  $\hat{\lambda}_0(t)$  є оцінкою базової функції швидкості небезпеки  $\lambda_0(t)$ , а  $\hat{\beta}$  являє собою оцінку вектора параметрів  $\beta$ .

Таким чином, при припущенні про модель пропорційних ризиків,  $PD$  оцінюється як:

$$\hat{PD}^{PHM}(t|x) = \frac{\hat{F}_{\hat{\beta}}(t+b|x) - \hat{F}_{\hat{\beta}}(t|x)}{1 - \hat{F}_{\hat{\beta}}(t|x)} = 1 - \frac{\hat{S}_{\hat{\beta}}(t+b|x)}{\hat{S}_{\hat{\beta}}(t|x)},$$

де  $1 - \hat{F}_{\hat{\beta}}(t|x) = \hat{S}_{\hat{\beta}}(t|x) = \exp(-\hat{\Lambda}(t|x))$ .

Метод оцінювання для цієї моделі складається з двох кроків. На першому етапі інтегральна функція базового ризику  $\Lambda_0(t)$ , оцінюється наступним чином:

$$\hat{\Lambda}_0(t) = \frac{\sum_{i=1}^n \mathbb{1}\{Y_i \leq t, \delta_i = 1\}}{\sum_{j=1}^n \mathbb{1}\{Y_j \geq Y_i\}},$$

Тоді параметр  $\beta$  оцінюється як

$$\hat{\beta}^{PHM} = \arg \max_{\beta} L(\beta),$$

де часткова функція правдоподібності задається

$$L(\beta) = \prod_{i=1}^n \frac{\exp(x_i^T \beta)}{\left( \sum_{j=1}^n \mathbb{1}\{Y_j > Y_i\} \exp(x_j^T \beta) \right)}.$$

Таким чином, оцінка умовної інтегральної функції ризику обчислюється за формулою:

$$\hat{\Lambda}(t|x) = \int_0^t \hat{\lambda}(s|t) ds = \exp(x^T \hat{\beta}^{PHM}) \hat{\Lambda}_0(t) \quad [4].$$



### Узагальнена лінійна модель

Для визначення «часу життя» кредитної картки можна прийняти узагальнену лінійну модель:

$$P(T \leq t | X = x) = F_{\theta}(t | x) = g(\theta_0 + \theta_1 t + \theta^T x),$$

де  $\theta = (\theta_2, \theta_3, \dots, \theta_{p+1})^T$ ,  $p$  – вимірний вектор,  $g$  – відома функція зв'язку, така як логістична чи пробіт-функція. Таким чином, ця модель характеризує умовний розподіл часу життя кредиту  $T$  в термінах невідомих параметрів. Як тільки параметри будуть оцінені, отримуємо оцінку функції умовного розподілу,  $F_{\hat{\theta}}$  і, нарешті, оцінка PD може бути обчислена шляхом включення цієї оцінки у рівняння (1), тобто

$$PD^{\hat{\theta}^{GLM}}(t | x) = \frac{F_{\hat{\theta}}(t+b | x) - F_{\hat{\theta}}(t | x)}{1 - F_{\hat{\theta}}(t | x)} = 1 - \frac{S_{\hat{\theta}}(t+b | x)}{S_{\hat{\theta}}(t | x)},$$

де  $\hat{\theta} = \hat{\theta}^{GLM}$  є оцінкою максимальної правдоподібності вектору параметрів.

Розглянемо одновимірний випадок коваріантності. У такому випадку  $\theta = \theta_2$  і умовний розподіл задається моделлю  $F(t | x) = g(\theta_0 + \theta_1 t + \theta_2 x)$ , зі щільністю  $f(t | x) = \theta_1 g'(\theta_0 + \theta_1 t + \theta_2 x)$ . Оскільки зазвичай задана випадкова цензурована справа вибірка, то умовна функція правдоподібності представляє собою добуток членів, що включають умовну щільність, для нецензурованих даних та умовної функції виживання для цензурованих даних:

$$L(Y, X, \theta) = \prod_{i=1}^n f(Y_i | X_i)^{\delta_i} (1 - F(Y_i | X_i))^{1-\delta_i},$$

де  $Y_i$  – строк погашення  $i$ -го кредиту і  $\delta_i$  є індикатором дефолту для  $i$ -го кредиту.

Таким чином, логарифмічна функція правдоподібності визначається:

$$\begin{aligned} l(\theta) &= \ln(L(Y, X, \theta)) = \sum_{i=1}^n [\delta_i \ln(f(Y_i | X_i)) + (1 - \delta_i) \ln(1 - F(Y_i | X_i))] = \\ &= \sum_{i=1}^n [\delta_i \ln(\theta_1 g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i)) + (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] = \\ &= \sum_{i=1}^n \delta_i [\ln(\theta_1) + \ln(g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] + \sum_{i=1}^n (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i)) \end{aligned}$$

І, нарешті, оцінка знаходиться як максимізація функції логарифмічної правдоподібності:

$$\hat{\theta}^{GLM} = \arg \max_{\theta} l(\theta)$$

## **Експериментальне моделювання**

### **Опис вхідних даних**

Дані про поведінку клієнтів-власників кредитних карт були отримані із кредитної історії одного з провідних банків України. Вони включають в себе інформацію про кредитні карти, що видавались цим банком з вересня 2013 по квітень 2016. В цю інформацію входить детальний опис заявника, тобто дані його заявки на отримання кредитної карти та поведінка зі щомісячною деталізацією. Також сюди входять певні ідентифікатори (ідентифікатор контрагента, угоди тощо), які були модифіковані для збереження банківської таємниці.

Відповідно до побудованої моделі кредитної карти відомості про користування кредитними картами зводились до інформації в термінах заборгованості, або балансу, простроченої заборгованості, ліміту, зняття (користування лімітом) та платежами за договором.

При цьому варто відмітити, що постала проблема уніфікації даних, оскільки кредитна карта у виборці представлені двома продуктами, що відрізнялися методом нарахування щомісячних мінімальних платежів: абсолютне значення у вигляді комісії в першому варіанті та процентна ставка відносно балансу у другому. Однак обраний підхід обробки та побудована модель дозволяють розглядати ці два продукти, як один, що в свою чергу розв'язує наступні задачі:

- збільшення розміру вибірки;
- абстрагування відносно продуктів.

Це є важливою особливістю, оскільки, по-перше, розмір вибірки дуже впливає на предикативність побудованих на її основі моделей. По-друге, з точки зору пересічних людей кредитна карта – це певне загальне ім'я, і вони, в переважній більшості, не звертають увагу на такі особливості кредитування. Тому в цьому контексті уніфікація є певним природнім етапом.

Для формування цільового поля був обраний наступний критерій [3]:

- прострочка більше 100 грн.;
- кількість днів прострочки не менше 90 (3 прострочених платежі підряд).

Друга умова є загальноприйнятою в міжнародній практиці. Після настання такого моменту у більшості випадків банки розпочинають процедуру списання, вважаючи такі карти дефолтними. Накладання ж першої умови спричинено вже самими даними: в процесі попередньої обробки даних було виявлено випадки погашення основної заборгованості і нарахування комісії в один операційний

місяць. Досить часто на таку комісію ніхто не звертав увагу і вона переходила в прострочену заборгованість, залишаючись такою досить тривалий час (іноді понад 90 днів). Зрозуміло, що такі випадки аж ніяк не можна вважати прикладами дефолту.

У результаті попереднього аналізу можна виділити декілька типових кредитних карт. В більшості випадків за кредитними картами, які не перейшли у стан дефолту, все одно можна спостерігати прострочені платежі. Значення прострочених днів в межах від 1 до 7 є характерним саме для «хороших» клієнтів.

### Порівняльний аналіз результатів різних методів

Авторами виконувалось моделювання споживчих ризиків декількома підходами. Зокрема, були використані методи інтелектуального аналізу даних для оцінювання ймовірності дефолтів для кредитних карток. Був обраний апарат нейронних мереж, лінійної регресії та логістичної регресії.

Для нейронних мереж був застосований алгоритм побудови багатосарового перцептронну з трьома прихованими шарами. Для логістичної мережі було застосовано логіт-перетворення. За різними статистичними критеріями було виконане порівняння якості моделей на навчальній та перевірочній вибірці (таблиці 1 та 2). Кращою моделлю за сукупністю статистичних критеріїв виявилась нейронна мережа, однак результати навіть на навчальній вибірці говорять про необхідність покращення моделей.

Табл.1. Результати моделювання на навчальній вибірці

Модель	Misclassification rate	IKA	SBC	GINI	KS
Нейронна мережа	0.26	5772.87	5997.73	0.634	0.481
Логістична регресія	0.272	5922.28	5988.42	0.609	0.465
Лінійна регресія	0.273	9357.37	9291.24	0.609	0.462

Табл.2. Результати моделювання на перевірочній вибірці

Модель	Misclassification rate	GINI	KS
Нейронна мережа	0.271	0.615	0.461
Логістична регресія	0.278	0.6	0.449
Лінійна регресія	0.279	0.6	0.448

Побудовані моделі дозволяють визначити ймовірність переходу у стан дефолту на основі деякої інформації (параметрів) кредитної карти та, відповідно, історії платежів та обслуговування кредиту за попередні періоди. Такий підхід є класичним аплікаційним скорингом, що використовується для прийняття рішень щодо видачі кредиту (кредитної карти). Більш цікавим є налаштування певних параметрів кредитної карти з урахуванням інформації щодо інших подібних кредитних карток, тобто встановлення кредитного ліміту, терміну погашення, можливої процентної ставки. Саме ці параметри дозволяють банкам отримувати вигоду навіть за тими кредитами, які в певний період можуть перейти у стан проблемних або навіть безнадійних.

### **Побудова моделей аналізу виживання**

Для виконання поведінкового скорингу [3, 5] та прогнозування періоду настання проблемності кредиту була застосована методологія аналізу виживання. Для прогнозування періоду проблемності кредиту були побудовані два типи моделей: параметричні та напівпараметричні.

За нашою задачею було сформовано вибірку з 55286 спостережень за 4597 кредитними картами зі щомісячною деталізацією. За весь період було зафіксовано 714 випадків дефолту, у 3837 спостерігались прострочки, було здійснено цензування 760 кредитних карток, що склало 16,53% із загального обсягу кредитних карт.

Було побудовано 2 типи моделей аналізу виживання: параметричні (з використанням процедури Lifereg: експоненційним, логістичним та розподілом Вейбулла) та напівпараметрична модель (пропорційних ризиків Кокса). Серед параметричних моделей кращою виявилась модель з використанням розподілу Вейбулла. На практиці найбільш відомою є модель пропорційних ризиків Кокса, оскільки допускає мінімальну кількість припущень, яка за нашими експериментами також показала високу предикативну здатність.

Далі здійснювався аналіз періоду, коли виникають проблеми з обслуговуванням кредиту. На рис.1 видно, що перші 3 місяці кредитні картки не переходять у стан дефолту, проте з ростом періоду «життя» кредитної карти наступають все частіше випадки проблемності обслуговування кредитної карти. До 12 місяців ймовірність «виживання» кредитної карти становить більше 0,6, проте через 24 місяці ймовірність виживання падає аж до 0,1. Також

спостерігається досить різке падіння проценту ідеального обслуговування кредиту через 18 місяців.

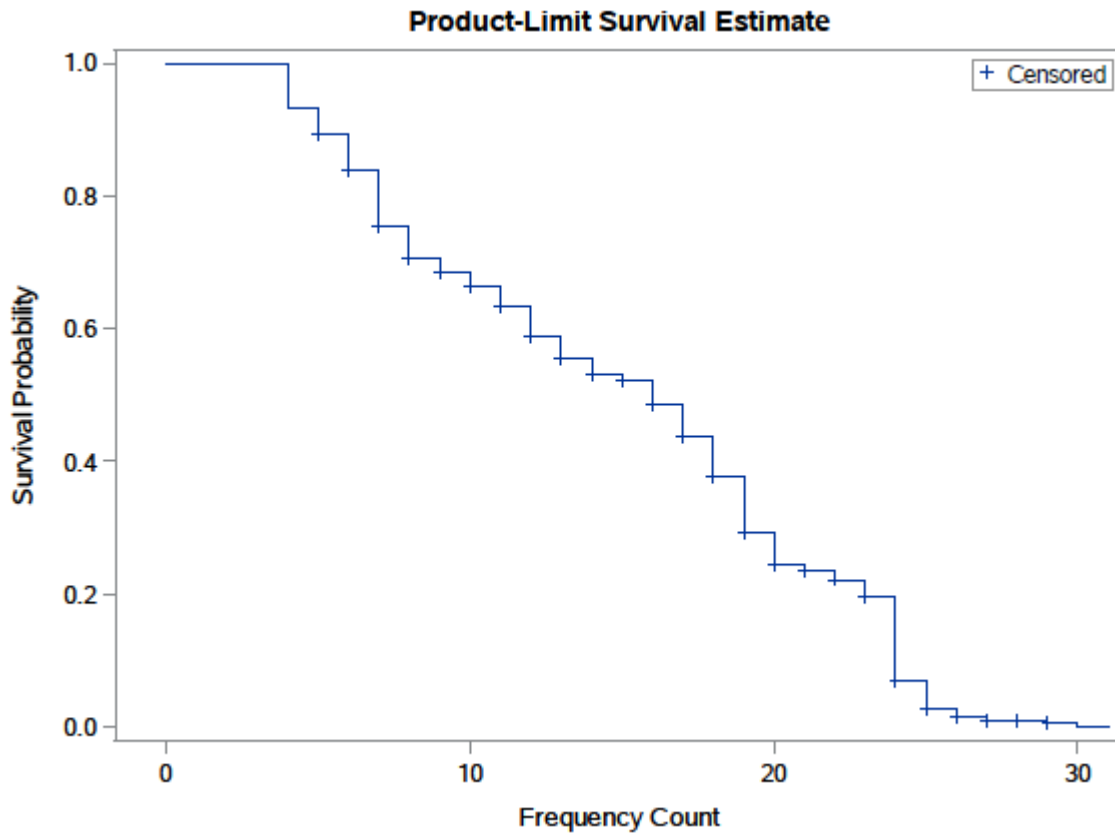


Рис.1 Результати моделювання для аналізу ймовірності виживання кредитної карти з часом

Наступним етапом є оцінка функції ризику в часі, тобто визначення, на якому ступені ризику знаходиться кредитна карта в залежності від часу обслуговування. Як видно з рис.2, до 18 місяців рівень ризику є прийнятним (до 0,1), є лише невеликий сплеск через 8 місяців, який потім вирівнюється, проте з часом ризик значно зростає. Це може стати своєрідним компромісом для політики банку. Визначивши допустимий рівень ризику, який банк може прийняти на себе, можна визначити оптимальний період та ліміт кредитування для певного типу клієнтів.

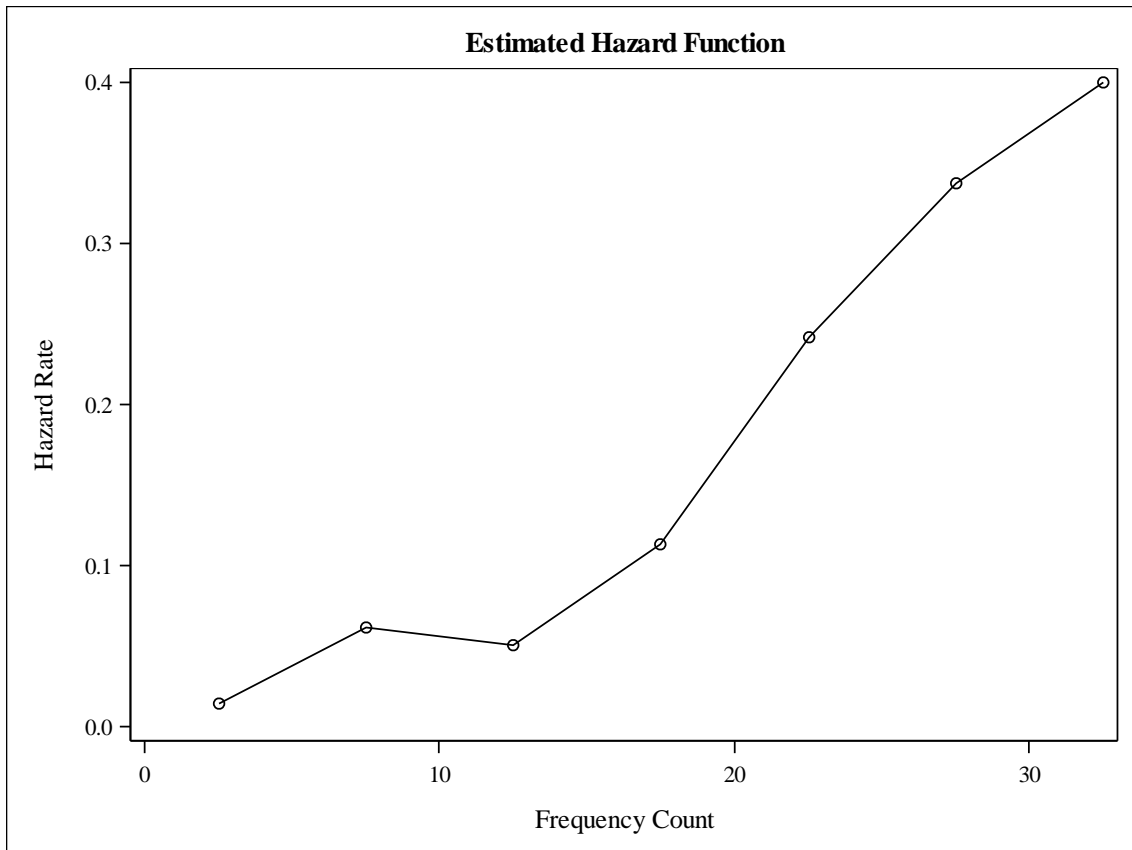


Рис. 2 Оцінена функція ризику в залежності від часу

### Висновки

Проведене дослідження можливості застосування теорії аналізу виживання для біхевіоріального скорингу показало доцільність та перспективність цього підходу. Побудовані моделі (параметричні та напівпараметричні) дозволили визначити період, в який настає проблемність обслуговування кредитних заборгованостей за картками, а також визначити ймовірність настання такої події на певному етапі кредитування. Це дозволяє говорити про обсяги можливих втрат банків в результаті негативної динаміки по кредитним карткам, а також застосовувати реальні засоби боротьби з про строчками на більш ранніх етапах (ще до закінчення терміну кредиту). Також моделі пропорційних ризиків Кокса, які показали найкращі результати для нашої прикладної задачі, можуть бути використані банками для розробки нових кредитних продуктів (в залежності від потреб споживачів на різні періоди обслуговування та з різними процентними ставками) та планування фіскальних дій в разі несвоєчасного обслуговування кредитів.

## Література

1. Siddiqi N. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. – 2005. – 196p.
2. Cox D. R. Regression models and life-tables / D. R. Cox, S. Society, S. B. Methodological // 2007. — Vol. 34, No. 2. — P. 187–220.
3. Фомін О.В. Скорингові моделі поведінки клієнтів-власників кредитних карток для оцінки їх платоспроможності/ Фомін О.В., Кузнєцова Н.В. // Системні науки та кібернетика. — 2016. — №5. — с.56-67. [Електронний ресурс]. — Режим доступу : [http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc\\_5\\_2016.pdf](http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_5_2016.pdf).
4. Cao R., Vilar J.M., Devia A. Modelling consumer credit risk via survival analysis / SORT 33 (1) January-June 2009, p.3-30.
5. Marimo M. Survival analysis of bank loans and credit risk prognosis / M. Marimo // [Електронний ресурс]. — Режим доступу : [http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis\\_Survival%20Analysis\\_28.03.%202015\\_v1.pdf](http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis_Survival%20Analysis_28.03.%202015_v1.pdf).
6. Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework – Comprehensive Version // [Електронний ресурс]. — Режим доступу: <http://www.bis.org/publ/bcbs54.htm> .

# Методи нелінійної оптимізації в задачах проектування електричних двигунів

І.М. Курін, О.Л. Тимошук

## Вступ

Науково-технічний прогрес в різних галузях промисловості викликає зростання потреб в електричних машинах малої потужності. Як показує світовий досвід, виробництво мікромашин представляє собою самостійну галузь, яка динамічно розвивається, і темпи зростання якої перевищують темпи зростання електромашинобудівництва в цілому.

Розробка нових технологічних процесів виробництва електричних машин відкриває широкі можливості для підвищення продуктивності праці та ефективності виробничої діяльності. Виробництво електричних машин малої потужності характеризується багатомоделлюваністю, різноманітністю конструкцій, підвищеними вимогами до надійності та технічного рівня, технологічності конструкцій, підвищенням рівня спеціалізації та кооперації виробництва [1].

Технологічність конструкції надає істотного впливу на техніко-економічні показники серійного та масового виробництва електричних машин малої потужності. Конструкція спроектованої електричної машини має бути технологічною, тобто новий двигун має мати певну з технологічних міркувань величину повітряного зазору, коефіцієнт заповнення пазу, ширину зубців, відношення довжини активної зони до зовнішнього діаметру тощо [2].

Сучасні електричні машини є досить проблемними у тепловому відношенні. Тепловий розрахунок дозволяє контролювати вибір електромагнітних навантажень, забезпечуючи високе використання активних матеріалів без зниження надійності двигуна.

Проблема створення електричних машин, які мають оптимальне відношення електромагнітних навантажень та геометричних розмірів, з'явилася із зародженням електромашинобудівництва і залишається актуальною до сьогоднішнього часу. Праці багатьох вчених були присвячені пошукам закономірностей, які б дозволили із меншими витратами праці та часу вирішити цю проблему відповідно до конкретно заданих умов [3, 4]. Проте необхідність



врахування великої кількості факторів, які є часто суперечливими, ускладнює отримання та використання таких закономірностей при вирішенні прикладних задач.

Оптимізація проекту є головною ціллю кожного інженера, який намагається створити окремий елемент, пристрій чи систему для забезпечення певних потреб. Проте здійснення цієї цілі зазвичай є досить проблематичним, адже у розпорядженні конструкторів знаходиться невелика кількість стандартизованих методів, здатних допомогти в оптимізації.

Поява швидкодіючих ЕОМ відкрила нові можливості для дослідження та оптимального проектування електричних машин. Задача оптимізації електричних машин зводиться до пошуку локального, а іноді, глобального, чи достатньо близького до нього екстремуму функції якості в багатовимірному просторі із нелінійною системою обмежень.

Вибір критерію якості залежить від призначення електричної машини і пред'явлених до неї вимог [1, 3, 5]. Для деяких електричних машин доцільно вибрати мінімум маси чи мінімум приведених витрат, для інших максимум енергетичних показників – ККД,  $\cos \phi$ , іноді визначними критеріями є перенавантажувальна здатність –  $K_m$  і підвищені пускові властивості –  $K_p$ . Таким чином, не можливо знайти універсальний критерій оптимальності.

Дослідження присвячене аналізу методів нелінійного програмування в прикладних технічних задачах, створенні програмного забезпечення для пошуку оптимальних геометричних параметрів однофазного асинхронного двигуна.

### **Постановка задачі**

Розглянути існуючі методи нелінійної оптимізації і створити програмний продукт для пошуку оптимальних геометричних параметрів однофазного асинхронного двигуна із різними технічними обмеженнями.

### **Метод штрафних функцій**

Основна ідея даного методу нелінійного програмування полягає у використанні штрафної функції – цільової функції допоміжної задачі безумовної оптимізації [3, 6]. За допомогою штрафної функції вихідна задача умовної мінімізації перетворюється в послідовність задач безумовної мінімізації.

Штрафна функція визначається виразом

$$P(x, R) = f(x) + Q(R, g(x)), \quad (1)$$

де  $R$  – набір штрафних параметрів;

так званий штраф  $Q$  є функцією  $R$  і функцій, які задають обмеження.

Алгоритм, який було використано в програмному продукті:

- 1) задати початковий вектор  $x^0$ , початковий параметр  $R=0$ ,  $\varepsilon_1$  – параметр закінчення одновимірної оптимізації,  $\varepsilon_2$  – параметр закінчення процедури безумовної оптимізації та  $\varepsilon_3$  – параметр завершення роботи алгоритму;
- 2) побудувати штрафну функцію  $P(x^t, R^t) = f(x^t) + Q(R^t, g(x^t))$ ;
- 3) знайти  $x^{t+1}$ , який є мінімумом функції  $P(x, R)$  при фіксованому  $R = R^t$ . У якості початкової точки використовувати  $x^t$ , а в якості параметра закінчення кроку -  $\varepsilon_1$ . В програмному продукті безумовна мінімізація відбувається методом спряжених градієнтів;
- 4) перевірити виконання умови  $|P(x^{t+1}, R^t) - P(x^t, R^{t-1})| \leq \varepsilon_2$ . У випадку, якщо вона виконується, то покласти  $x^{t+1} = x^t$  і завершити процес розв'язання. Інакше, перейти до наступного кроку;
- 5) покласти  $R^{t+1} = R^t + \Delta R$  у відповідності з обраним правилом перерахунку та перейти до пункту 2.

В програмному продукті в якості штрафу було використано штраф типу квадрату зрізання який описується такою формулою:

$$Q(R, g(x)) = \begin{cases} 0, & g(x) \geq 0 \\ (g(x))^2, & g(x) < 0 \end{cases} \quad (2)$$

### Модифікований метод множників Лагранжа

Розглянемо функцію

$$P(x, \sigma) = f(x) + R \sum_{j=1}^m ((g_j(x) + \sigma_j)^2 - \sigma_j^2) \quad (3)$$

де  $R$  – постійний ваговий коефіцієнт;

$\langle \cdot \rangle$  - функція зрізання, яка має вид:

$$\langle x \rangle = \begin{cases} 0, & x \geq 0 \\ x, & x < 0 \end{cases} \quad (4)$$

Параметри виконують іншу роль, ніж  $R$ , виконуючи зсув штрафних доданків. Даний зсув ітеративно уточнюється, в результаті чого процес збігається до розв'язку при не жорстких обмеженнях на задачу [3].

Позначимо через  $x^t$  точку мінімуму штрафної функції, яка використовується на t-ій ітерації. При переході до (t+1)-ої ітерації множники перераховуються за формулою:

$$\sigma_j^{t+1} = (g_j(x^t) + \sigma_j^t), j = 1, 2, \dots, m \quad (5)$$

Через наявність у формулі оператора зрізання вектор  $\sigma$  не має додатних компонент. Кожен новий вектор  $\sigma^t$  визначає зсув аргумента відповідного штрафу, причому формула перерахунку така, що в результаті зміни зсуву при переході до нової під задачі штраф за порушення обмежень збільшується, і, як наслідок цього, стаціонарні точки  $x^t$  наближаються до допустимої області.

Для контролю збіжності метода використовують послідовності  $x^t, \sigma^t, f(x^t)$  та  $g(x^t)$ . При цьому припускається, що алгоритм безумовної оптимізації, який використовується, кожний раз дозволяє знайти стаціонарну точку і після цього завершити ітерацію на основі певного критерію. Зупинка основного алгоритму відбувається тоді, коли хоча б одна з даних послідовностей перестане значно змінюватись при перерахунку множників і наступної безумовної оптимізації.

### Метод комплексів

Розглянемо один із методів прямого пошуку, тобто пошук оптимальної точки буде проводитися лише обчисленням значень цільової функції та обмежень, не обчислюючи їх градієнти. Даний тип методів був розроблений у зв'язку з тим, що в прикладних технічних задачах цільова функція та обмеження можуть бути недиференційованими.

Одним із прикладів методів даного типу є метод комплексів, який базується на прямому пошуку оптимальної точки вздовж симплексу. Було запропоновано будувати множину пробних P точок випадково і послідовно. Маючи попередньо задані інтервали для змінних x, можна на ЕОМ будувати їх випадкові реалізації, а саме, маючи верхні та нижні межі  $x_i^U$  і  $x_i^L$ , координати точок розраховуються за формулою:

$$x_i = x_i^L + r_i (x_i^U - x_i^L), i = 1, 2, \dots, n. \quad (6)$$

де  $r_i$  — випадкові числа, рівномірно розподілені на інтервалі (0,1).

Кожна отримана точка перевіряється на допустимість, і якщо одне із обмежень порушується, то точка зміщується до центру мас вже побудованих

точок до тих пір, поки дана точка не стане допустимою. Загальна кількість точок  $P$  має бути не менше, ніж  $n+1$ , але, зазвичай, вона є набагато більшою.

Після того як множина із  $P$  точок отримана, в кожній з них обчислюється цільова функція, і точка, яка має найбільше значення, відкидається. Нова точка отримується шляхом відображення точки, яка видаляється, через центр мас інших точок. Якщо  $x^R$  – точка, яку відкидаємо, а  $\bar{x}$  – центр мас точок, які залишилися, то нова точка визначається так:

$$x^m = \bar{x} + \alpha(\bar{x} - x^R) \quad (7)$$

Параметр  $\alpha$  є відстанню відображення: при  $\alpha=1$  має місце рівність  $\|x^m - \bar{x}\| = \|\bar{x} - x^R\|$ ,  $\alpha > 1$  відповідає розтягу, а  $\alpha < 1$  – стиску.

Коли знайдена нова точка, обчислення значення цільової функції та обмежень, можливі наступні випадки:

а) нова точка допустима і значення цільової функції в ній не співпадає з максимальним значенням серед усієї множини точок. В такому випадку обираємо точку з максимальним значенням функції та знову виконуємо операції відображення;

б) нова точка допустима і значення цільової функції в ній співпадає з максимальним значенням. Замість того, щоб виконувати відображення, змістимо дану точку на половину відстані до центру мас;

в) нова точка недопустима. В цьому випадку зменшуємо вдвічі відстань до обчисленого раніше центру мас. Процедура пошуку продовжується до тих пір, поки багатогранник стягнутий до центру мас в межах ладанної точності чи поки різниця між значеннями цільової функції в вершинах не стане досить малою.

Були проведені числові експерименти по описаному алгоритму та на основі аналізу було виявлено, що краще обирати  $\alpha > 1$  та  $P > 2n$ . Вибір  $\alpha$  більшим за одиницю компенсує стиск комплексу, який спричинений зменшенням відстані до центру мас. Велике число вершин використовується для попередження виродження комплексу, коли пошук виконується біля границі допустимої області.

Хоча метод комплексів не потребує диференційованості функцій, які визначають задачу, так як не потрібно шукати похідні від них, необхідною умовою є опуклість допустимої області. Дана умова є істотною при знаходженні центру мас і зміщенні допустимої, але не задовольняючої щодо значення цільової функції, точки.

## Метод допустимих напрямків

Розглянемо ЗНП з обмеженнями у вигляді нерівностей:

$$\begin{cases} f(x) \rightarrow \min \\ g_j(x) \geq 0, j = 1, \dots, m \end{cases} \quad (8)$$

Нехай  $x^0$  – початкова точка, яка задовольняє обмеженням, тобто  $g_j(x^0) \geq 0, j = 1, \dots, m$ . Припустимо, що деякі обмеження виконуються як рівності для  $x^0$ . Нехай вектор  $d$  визначає підходящий напрям для пошуку, якщо  $d$  – напрямок спуску, тобто  $\nabla f(x^0)d < 0$  і якщо точки променя  $x(\alpha) = x^0 + \alpha d, \text{де } \alpha \geq 0$  є допустимими принаймні на невеликій відстані від  $x^0$ . З точністю, яка забезпечується лінійною апроксимацією, точки  $x(\alpha)$  вздовж  $d$  будуть допустимими, якщо для всіх зв'язуючих в  $x^0$  обмежень виконується умова

$$g_j(x^0) + \nabla g_j(x^0)(x - x^0) \geq 0 \quad (9)$$

Так як за припущенням  $g_j(x^0) = 0$  і  $x - x^0 = \alpha d, \text{де } \alpha \geq 0$ , тоді остання умова еквівалентна умові для  $\nabla g_j(x^0)d \geq 0$  для всіх  $g_j(x)$ , рівних нулю в точці  $x^0$ . Напрямок  $d$ , який задовольняє записаним вище нерівностям, отримав назву допустимого напрямку.

Основна ідея даного методу полягає в тому, щоби на кожному кроці алгоритму визначати вектор  $d$  як допустимий напрямок і скалярний параметр  $\theta > 0$ , що задовольняють умови  $\nabla f(x^k)d \leq -\theta$  і  $\nabla g_j(x^k)d \geq \theta$ , а значення  $\theta$  обирається по можливості великим.

Після того як вектор напрямку обрано, наступне наближення може бути визначено пошуком по  $\alpha$  вздовж прямої  $x = x^k + \alpha d^k$  до тих пір, поки або  $f(x)$  не досягне оптимума, або яке-небудь з обмежень не стане порушеним. Зазвичай спочатку відбувається обчислення значення  $\bar{\alpha}$ , при якому яке-небудь з обмежень  $g_j(x) \geq 0$  вперше виявляється зв'язуючим. При відомому  $\bar{\alpha}$  можна використати будь-яку процедуру одномірного пошуку для визначення  $\alpha$ , яке мінімізує функцію  $f(x^k + \alpha d^k)$  на відріжку  $[0, \bar{\alpha}]$ .

Проте у випадку неопуклої допустимої області необхідно обережно проводити одномірний пошук, бо, рухаючись вздовж спочатку допустимого напрямку  $d$ , можна вийти за межі області допустимих значень, а потім знову повернутися до неї [7].

## Оптимальне проектування однофазного асинхронного двигуна

Для оптимального проектування електричного двигуна було написано програмний продукт на мові програмування C# із зручним для користування інтерфейсом, написаним за допомогою системи побудови клієнтських додатків Windows Presentation Foundation (WPF). Головне вікно програми зображено на рисунку 1.

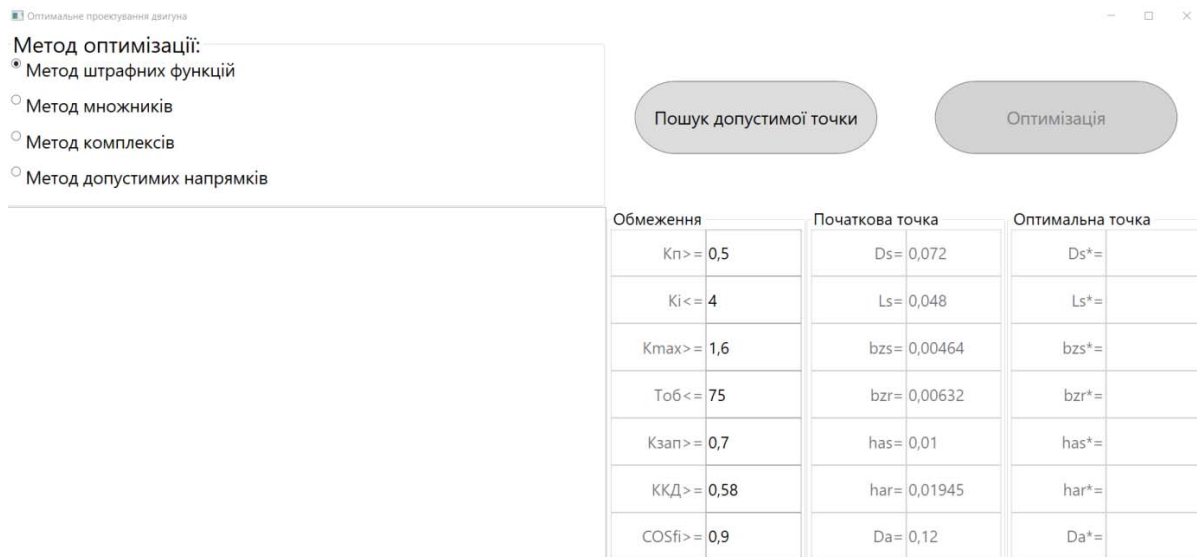


Рисунок 1 – Головне вікно програми

В якості критерію оптимальності двигуна було обрано сумарну масу електротехнічної сталі.

Спочатку, проведемо проектування двигуна із такими допустимими величинами обмежень:

- кратність початкового пускового моменту ( $K_p$ )  $\geq 0.5$ ;
- кратність початкового пускового струму ( $K_i$ )  $\leq 4$ ;
- кратність максимального моменту ( $K_{max}$ )  $\geq 1.6$ ;
- перевищення температури обмоток статора ( $T_{об}$ )  $\leq 75$ ;
- коефіцієнт заповнення пазу ( $K_{зап}$ )  $\geq 0.7$ ;
- коефіцієнт корисної дії (ККД)  $\geq 0.58$ ;
- коефіцієнт потужності (COSfi)  $\geq 0.9$ .

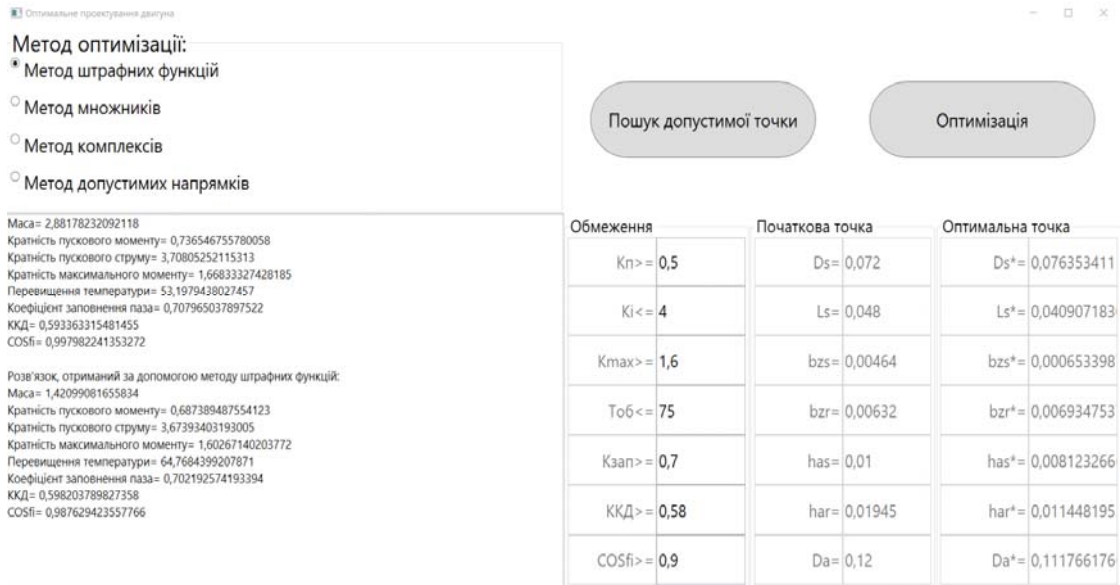


Рисунок 2 – Проектування першого двигуна за допомогою методу штрафних функцій

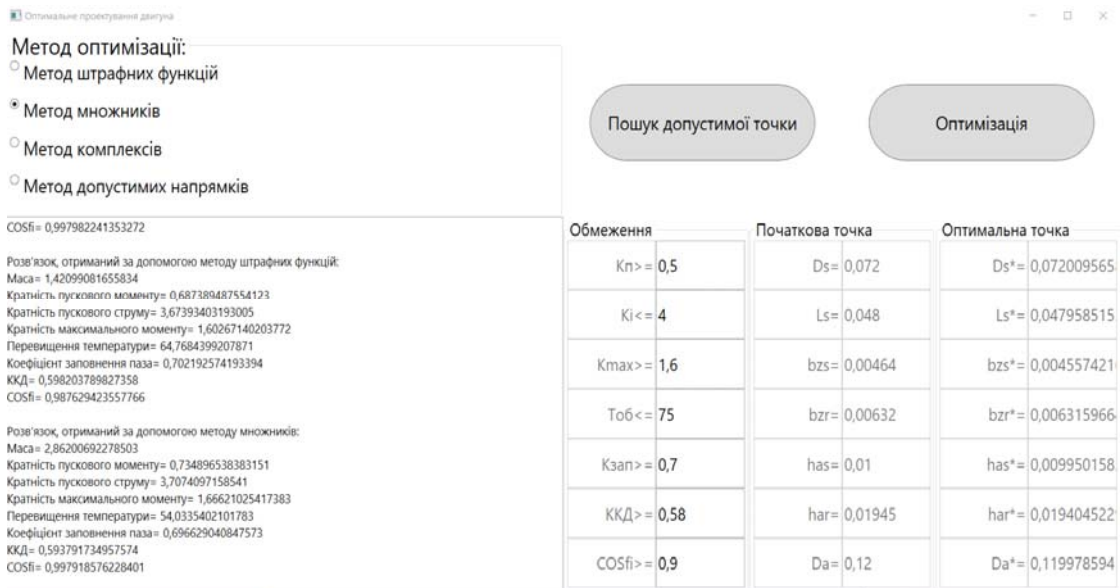


Рисунок 3– Проектування першого двигуна за допомогою методу множників

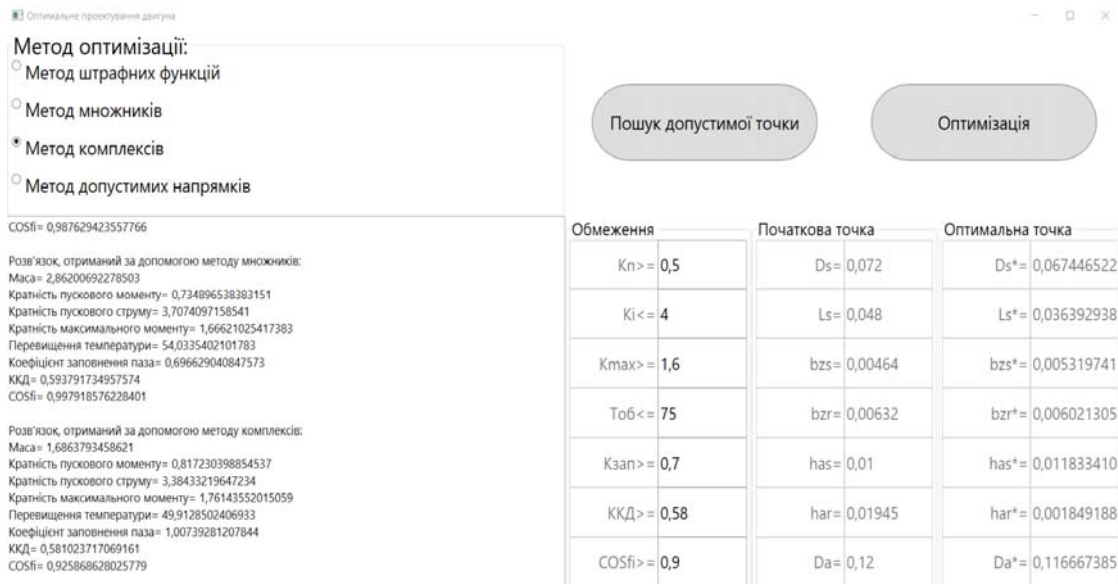


Рисунок 4 – Проектування першого двигуна за допомогою методу комплексів

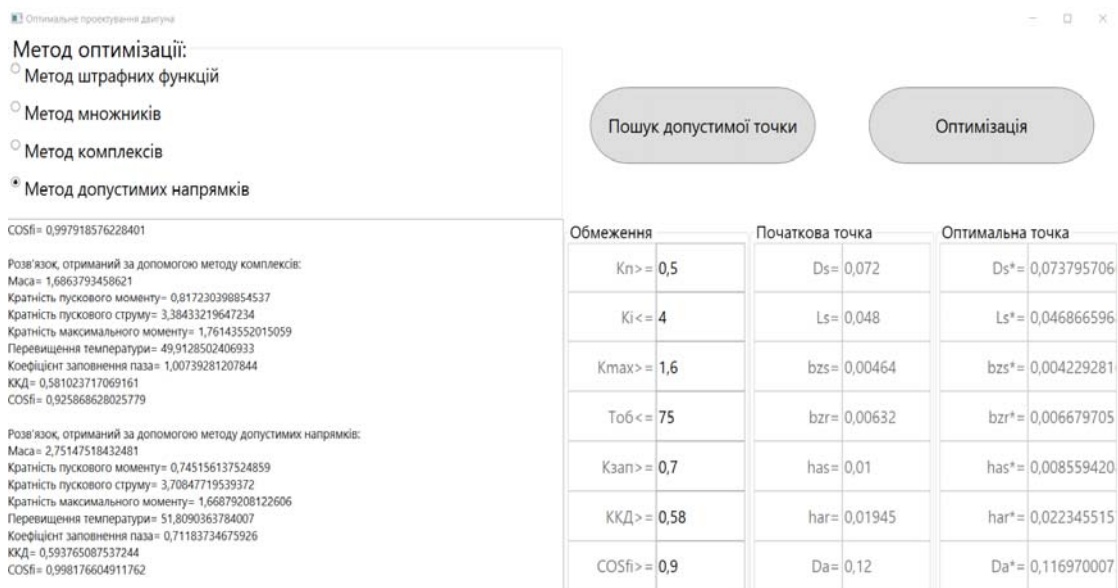


Рисунок 5– Проектування першого двигуна за допомогою методу допустимих напрямків

Отже, для проектування першого двигуна було розглянуто усі чотири методи оптимізації. Серед них, точку із найменшим значенням цільової функції дав метод штрафних функцій.

Тепер проведемо проектування іншого двигуна із наступними величинами обмежень:

– кратність початкового пускового моменту ( $K_{п} \geq 0,6$ ;



- кратність початкового пускового струму ( $K_I$ )  $\leq 3.7$ ;
- кратність максимального моменту ( $K_{max}$ )  $\geq 1.6$ ;
- перевищення температури обмоток статора ( $T_{об}$ )  $\leq 75$ ;
- коефіцієнт заповнення пазу ( $K_{зап}$ )  $\geq 0.72$ ;
- коефіцієнт корисної дії (ККД)  $\geq 0.6$ ;
- коефіцієнт потужності (COSfi)  $\geq 0.9$ .

Метод оптимізації:

- Метод штрафних функцій
- Метод множників
- Метод комплексів
- Метод допустимих напрямків

Маса= 2,68464810979293  
Кратність пускового моменту= 1,24372254607748  
Кратність пускового струму= 3,31782026339079  
Кратність максимального моменту= 1,9500596855124  
Перевищення температури= 6,00011825140506  
Коефіцієнт заповнення паза= 4,99648395933728  
ККД= 0,608569799365023  
COSfi= 0,911806836987974

Даний метод не дає результат.

Обмеження	Початкова точка	Оптимальна точка
$K_p \geq 0,6$	$D_s = 0,066312295$	$D_s^* =$
$K_I \leq 3,8$	$L_s = 0,040570225$	$L_s^* =$
$K_{max} \geq 1,6$	$bzs = 0,003645967$	$bzs^* =$
$T_{об} \leq 75$	$bzr = 0,006696380$	$bzr^* =$
$K_{зап} \geq 0,72$	$has = 0,059974122$	$has^* =$
ККД $\geq 0,6$	$har = 0,019402069$	$har^* =$
COSfi $\geq 0,9$	$Da = 0,109394688$	$Da^* =$

Рисунок 6 – Проектування другого двигуна за допомогою методу штрафних функцій

Метод оптимізації:

- Метод штрафних функцій
- Метод множників
- Метод комплексів
- Метод допустимих напрямків

Маса= 2,68464810979293  
Кратність пускового моменту= 1,24372254607748  
Кратність пускового струму= 3,31782026339079  
Кратність максимального моменту= 1,9500596855124  
Перевищення температури= 6,00011825140506  
Коефіцієнт заповнення паза= 4,99648395933728  
ККД= 0,608569799365023  
COSfi= 0,911806836987974

Розв'язок, отриманий за допомогою методу множників:  
Маса= 2,62203235588968  
Кратність пускового моменту= 1,24639250631121  
Кратність пускового струму= 3,3040669458262  
Кратність максимального моменту= 1,95004734627675  
Перевищення температури= 7,47435086028344  
Коефіцієнт заповнення паза= 2,92764939667127  
ККД= 0,60750302740546  
COSfi= 0,910881498437215

Обмеження	Початкова точка	Оптимальна точка
$K_p \geq 0,6$	$D_s = 0,066312295$	$D_s^* = 0,066317695$
$K_I \leq 3,8$	$L_s = 0,040570225$	$L_s^* = 0,040524501$
$K_{max} \geq 1,6$	$bzs = 0,003645967$	$bzs^* = 0,003837573$
$T_{об} \leq 75$	$bzr = 0,006696380$	$bzr^* = 0,006698347$
$K_{зап} \geq 0,72$	$has = 0,059974122$	$has^* = 0,059999186$
ККД $\geq 0,6$	$har = 0,019402069$	$har^* = 0,019363694$
COSfi $\geq 0,9$	$Da = 0,109394688$	$Da^* = 0,109346461$

Рисунок 7– Проектування другого двигуна за допомогою методу множників

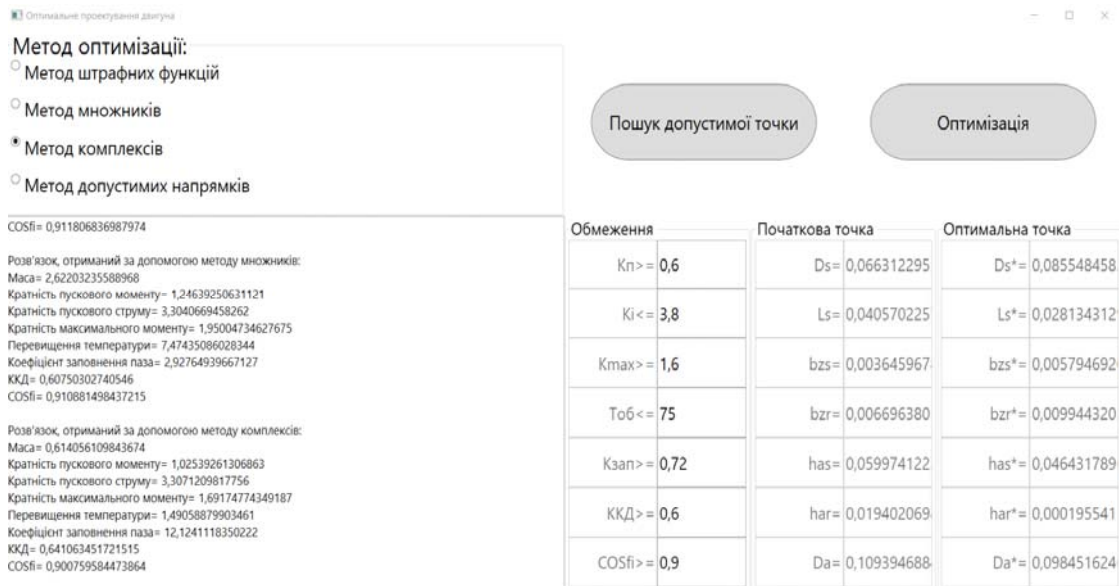


Рисунок 8 – Проектування другого двигуна за допомогою методу комплексів

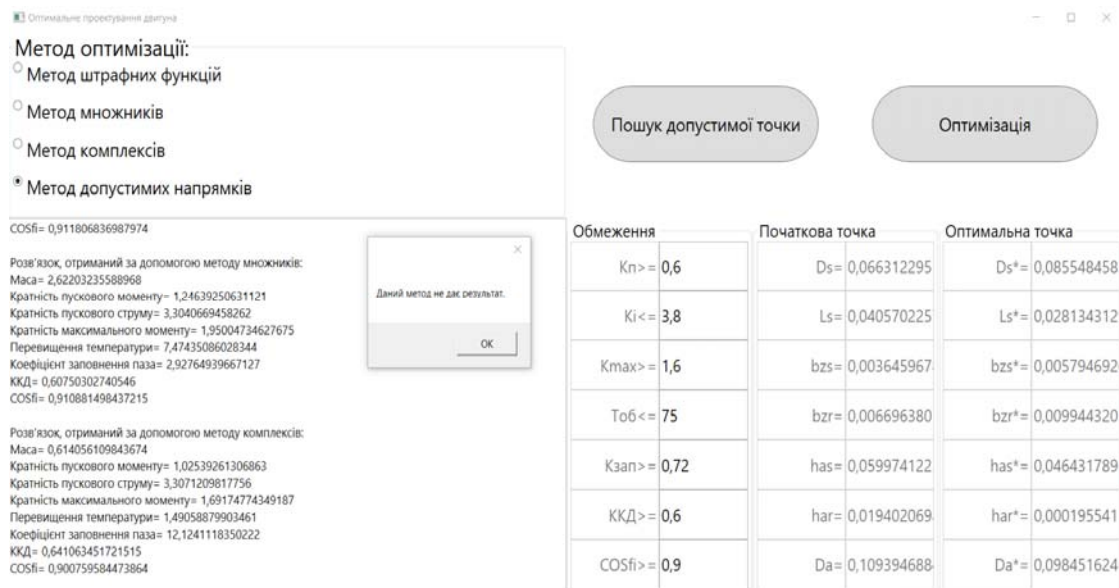


Рисунок 9 – Проектування другого двигуна за допомогою методу допустимих напрямків

Згідно з результатами роботи програми, проектування двигуна із потрібними технологічними обмеженнями було зроблено лише методами множників та комплексів. Це пояснюється тим, що в функціях, які описують обмеження, відбулося ділення на число, близьке до нуля, через що обчислення вивели алгоритм за межі допустимої області для методів штрафних функцій та допустимих напрямків.

## **Висновки**

Таким чином, оптимально спроектувавши двигуни, кожен з яких мав свої технічні обмеження, було встановлено, що методи штрафних функцій та допустимих напрямків є нестійкими щодо обчислень. Було продемонстровано ефективність використання методу комплексів та модифікованого методу множників Лагранжа для оптимального проектування електричних двигунів. Створено програмний продукт, який дозволяє проводити проектування однофазного асинхронного двигуна із різними технічними обмеженнями.

### Література:

1. Антонов М. В. Технология производства электрических машин / М. В. Антонов, Л. С. Герасимов. – Москва: Энергоизда, 1982. – 512 с.
2. Лопухина Е. М. Проектирование асинхронных микродвигателей с применением ЭВМ / Е. М. Лопухина, Г. А. Семенчуков. – Москва: Высшая школа, 1980. – 512 с.
3. Реклейтис Г. В. Оптимизация в технике. Кн. 1 / Г. В. Реклейтис, А. Рейвиндран, К. М. Рэгсдел ; пер. с англ. В. Я. Алтаева. – Москва : Мир, 1986. – 349 с.
4. Хог Э. Прикладное оптимальное проектирование / Эдвард Хог, Ясбир Арора ; пер. с англ. В. М. Картвелишвили. – Москва : Мир, 1983. – 478 с.
5. Акулич И. Л. Математическое программирование в примерах и задачах / Иван Людвигович Акулич. – Москва : Высшая школа, 1986. – 319 с.
6. Карманов В. Г. Математическое программирование / Владимир Георгиевич Карманов. – Москва : Наука, 1980. – 256 с.
7. Реклейтис Г. В. Оптимизация в технике. Кн. 2 / Г. В. Реклейтис, А. Рейвиндран, К. М. Рэгсдел ; пер. с англ. В. Я. Алтаева. – Москва : Мир, 1986. – 320 с.

## РЕФЕРАТИ

**І.О. Бойко, П.І. Бідюк**

### **ЗАСТОСУВАННЯ СКОРИНГОВОГО АНАЛІЗУ У КРЕДИТУВАННІ**

**Проблематика.** Стратегія управління ризиком займає важливе місце в забезпеченні ефективного функціонування банківських установ в умовах ринкової економіки та в умовах кризових явищ. Скорингові моделі є базовим індикатором кредитоспроможності потенційного позичальника. На їх основі експерт ухвалює остаточне рішення про видачу кредиту.

**Мета дослідження.** Зібрати статистичні дані стосовно характеристик позичальників кредитів для виконання обчислювальних експериментів. Вибрати типи математичних моделей і методів інтелектуального аналізу даних для формального опису позичальників кредитів. Побудувати вибрані типи моделей за допомогою статистичних даних. Виконати порівняльний аналіз отриманих результатів.

**Методика реалізації.** Для розв'язання поставлених задач використано метод оцінки кредитного ризику за допомогою побудови скорингових моделей. Для побудови моделей використовується логістична регресія, дерева рішень та мережі Байєса.

**Результати дослідження.** Основними результатами дослідження є: зібрані статистичні дані стосовно характеристик позичальників кредитів та виконані обчислювальні експерименти; вибрані типи математичних моделей і методів інтелектуального аналізу даних для формального опису позичальників кредитів у вигляді логістичної регресії, дерева рішень та мережі Байєса; побудовані вибрані типи моделей за допомогою статистичних даних; представлено найкращу побудовану математичну модель у вигляді мереж Байєса з найвищою точністю як для вибірки з незміненою структурою так і для вибірок зі зміненою структурою.

**Висновки.** В результаті комплексного застосування кількох методів інтелектуального аналізу даних, вдалося значно підвищити якість оцінювання – загальну точність платоспроможності у середньому на 2-3% та індекс Джіні на 9,6% з вибіркою зміненої структури (50%-повернень та 50%-дефолту); на 12% загальної точності та на 26% індекс Джіні з вибіркою зміненої структури (90%-

повернень та 10%-дефолту) відповідно з найкращою побудованою моделлю мережею Байеса.

Показники точності інших моделей – логістичної регресії та дерев рішень, значно нижчі. Оскільки дерева рішень є грубішим інструментом прогнозування ніж логістична регресія та байєсівські мережі, для вибірки зміненої структури (90%-повернень, 10%-дефолту) він має найгірший результат показників точності та ROC-кривої.

**Ключові слова:** кредитування; кредитний ризик; скоринг; кредитоспроможність; логістична регресія; мережі Байеса; дерева рішень.

Д.Г. Діденко, К.Д. Климковська

## **РОЗПОДІЛЕНЕ ОБЧИСЛЕННЯ І ПРОГНОЗУВАННЯ ВАЖЛИВОСТІ БРЕНДІВ В ІНТЕРНЕТІ**

**Проблематика.** Моніторинг соціальних мереж дозволяє безпосередньо почути клієнта в реальному часі, дізнатися його переваги, виявити цільову аудиторію, яка найбільш ймовірно стане споживачами продукту вибраної компанії. У процесі відстеження згадувань та відгуків в різних соціальних медіа, компанія отримує інформацію, яка буде важлива для неї та її бізнесу.

**Мета дослідження.** Головна мета дослідження: розглянути та порівняти принцип роботи найпопулярніших систем моніторингу соціальних мереж, розробити власний алгоритм визначення тональності тексту та алгоритм короткострокового прогнозування для прогнозування щоденної кількості позитивних та негативних відгуків про бренд, створити веб-додаток на основі вищезазначених алгоритмів для моніторингу соціальних мереж та блогів.

**Методика реалізації.** Для розв'язання поставлених задач використано такі методи: метод машинного навчання з учителем для визначення тональності відгуків користувачів, наївний баєсівський класифікатор (NB) та метод опорних векторів (SVM) для класифікації тональності відгуків, ієрархічна класифікація для навчання класифікатору та адаптивний алгоритм короткострокового прогнозування на базі розробленої в роботі нелінійної моделі для прогнозування щоденної кількості позитивних та негативних відгуків про бренд.

**Результати дослідження.** Основними результатами дослідження є такі: розроблено власний алгоритм визначення тональності тексту на основі наївного баєсівського класифікатору, побудовано нелінійну модель, яку було взято за

основу методу адаптивного короткострокового прогнозування щоденної кількості позитивних та негативних відгуків та створений веб-додаток для моніторингу соціальних мереж та блогів який реалізує запропоновані алгоритми.

**Висновки.** В роботі був реалізований веб-додаток для моніторингу соціальних мереж Facebook та ВКонтакте, блогів lifehacker.ru та businessinsider.com, автоматичного визначення тональності відгуків про бренд та короткострокового прогнозування щоденної кількості позитивних та негативних відгуків про бренд. Для оцінки результатів роботи методу власний програмний продукт був порівняний з системою моніторингу соціальних мереж BuzzLook. Було встановлено, що продукт визначає тональність тексту з точністю 71%, що свідчить про те, що він робить це майже так само добре, як і людина. Тобто власна система достатньо точно визначає тональність тексту. Перспективою подальших досліджень є використання нейронної мережі, яка навчається самостійно для обчислення тональності відгуків про бренд.

**Ключові слова:** моніторинг соціальних мереж, важливість бренду, тональність відгуку, алгоритм обчислення тональності тексту, алгоритм короткострокового прогнозування, бренд.

**Г.І. Кангун, П.І. Бідюк**

## **МОДЕЛЮВАННЯ ВІЙСЬКОВИХ КОНФЛІКТІВ З ВИКОРИСТАННЯМ БАЙЄСІВСЬКИХ МЕРЕЖ**

**Проблематика.** Побудова моделі процесу дозволяє всебічно розглянути процес та прогнозувати розвиток ситуації, що в результаті полегшує пошук оптимального рішення. Складність моделювання військового конфлікту полягає в першу чергу в тому, що він є динамічною системою з багатьма невідомими або невимірюваними параметрами. Крім того, необхідно враховувати неповноту та нечіткість інформації. Аналіз та моделювання подібної системи має на меті не тільки виявлення існуючих взаємозв'язків та залежностей, а і передбачення та попередження майбутніх конфліктів на основі результатів аналізу.

**Мета дослідження.** Розглянути загальні підходи до моделювання військових конфліктів. Розробити модель військового конфлікту на основі байєсівської мережі. Провести моделювання та наліз результатів.

**Методика реалізації.** Для розв'язання поставлених задач побудовано модель на основі мережі байєса. Було використано бібліотеку SMILE та

графічний інтерфейс GeNIe для моделювання. Статистичні дані для формування вибірки даних було взято з відкритих джерел.

**Результати дослідження.** Розглянуто існуючі підходи до моделювання військових конфліктів. Побудовано модель військового конфлікту на основі мережі байеса. Модель було навчено на сформованій з реальних статистичних даних вибірці, проведено серію експериментів та проаналізовано отримані результати.

**Висновки.** В роботі розглянуто проблему моделювання військових конфліктів, визначення та формалізації параметрів моделі.

Результатом роботи стала побудована модель військового конфлікту на основі мережі байеса. Модель було навчено на сформованій з реальних статистичних даних вибірці, проведено серію експериментів та проаналізовано отримані результати.

Побудована модель оцінила ймовірність внутрішнього конфлікту в Україні у 49%, що цілком відповідає реальності. Ймовірність зовнішнього було оцінено у 33%. Оскільки реально ймовірність зовнішнього конфлікту в Україні дорівнює одиниці, цей прогноз не є точним. Проте, це пояснюється тим, що у контексті розробленої моделі нинішній конфлікт пояснюється слабо, адже він слабо пов'язаний як з рентою природних ресурсів, так і з рівнем демократії або військових витрат України.

Отже, модель можна вважати адекватною та придатною для практичного використання.

**Ключові слова:** моделювання; військовий конфлікт; мережі байеса; прогнозування.

**Д.Г. Диденко, С.Д. Климковская**

## **РЕЛЕВАНТНІСТЬ ПОСИЛАНЬ В ІНТЕРНЕТІ ЗА РІЗНИМИ КРИТЕРІЯМИ НА РІВНІ ЗАДАЧІ ОПТИМАЛЬНОГО РОЗМІЩЕННЯ КОНТЕКСТНОЇ РЕКЛАМИ З НЕЛІНІЙНОЮ СКЛАДОВОЮ**

**Проблематика.** Задача оптимізації сайтів є актуальною проблемою, а задача автоматизації цього процесу є дуже важливою. Існуючі біржі посилань пропонують веб-майстрам самостійно обирати посилання для розміщення контекстної реклами по наведеним значенням зовнішніх факторів релевантності.

Але беручи до уваги декілька факторів і маючи обмежені кошти веб-майстру важко власноруч зробити оптимальний вибір закупки.

**Мета дослідження.** Метою є створення автоматизованої біржі посилань, яка на основі двох основних зовнішніх факторів релевантності – PR і тПЦ та обмеженого бюджету самостійно зможе підібрати необхідну кількість посилань найоптимальнішим чином в сенсі максимального поліпшення оптимізації сайту.

**Методика реалізації.** Для розв'язання поставленої задачі використовувалися такі математичні методи: метод зважених сум нормованих неоднорідних критеріїв та метод гілок та границь для розв'язку задач цілочисельного програмування з булевими змінними. Також до уваги бралась не лінійність росту тулбарного значення PR. Значення PR і тПЦ брались за допомогою підключення до AdWords API.

**Результати дослідження.** Основним результатом дослідження є реалізація автоматизованої біржі посилань. Реалізована біржа дає змогу користувачу підібрати посилання для закупівлі найоптимальнішим чином в сенсі максимального поліпшення оптимізації сайту. Оптимізація проводиться за зовнішніми чинниками і є багатокритеріальною.

**Висновки.** Новизною реалізованої біржі у порівнянні з існуючими є автоматизація закупки по обмеженому бюджету, тобто користувач повинен лише задати кількість коштів, які він згоден витратити на закупку посилань і натиснути кнопку, далі система вирішить питання оптимальної закупки самостійно.

**Ключові слова:** SEO, оптимізація, біржа посилань, посилання, автоматизація, гілка, верхня границя.

**В.В. Кобера, І.Ю. Каніовська**

## **ЗАСТОСУВАННЯ РЕАЛЬНИХ ОПЦІОНІВ В УПРАВЛІННІ РИЗИКАМИ ІНВЕСТИЦІЙНОГО ПРОЕКТУ БУДІВНИЦТВА ВЕЛИКОЇ КІЛЬЦЕВОЇ ДОРОГИ НАВКОЛО М. КИЇВ**

**Проблематика.** Фінансові ризики інвестиційного проекту і управління ними за допомогою методу та інструментарію реальних опціонів.

**Мета дослідження.** Фінансові відносини між державою і приватними інвесторами щодо реалізації інвестиційних проектів.



**Методика реалізації** - розробити модель управління фінансовими ризиками інвестиційного проекту, що передбачає використання методу реальних опціонів в умовах невизначеності.

**Результати дослідження.** В роботі показані переваги використання методу реальних опціонів в порівнянні з методом дисконтування грошових потоків для аналізу і оцінки фінансових результатів в умовах високої невизначеності.

**Висновки.** Було визначено умови використання реальних опціонів в інфраструктурних проектах та проведено розрахунок вартості інвестиційного проекту будівництва великої кільцевої дороги навколо Києва із застосуванням формули Блека-Шоулза, методів Датара-Метьюз і нечітких множин.

**Конєв Д.В., Бідюк П.І.**

## **МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕЛІНІЙНОГО НЕСТАЦІОНАРНОГО ПРОЦЕСУ ЦІНОУТВОРЕННЯ НА НАФТУ**

**Проблематика.** В умовах сучасної нестабільності фінансово-економічні процеси (ФЕП) відрізняються наявністю нелінійностей та нестационарності. Це приводить до необхідності в розробці та удосконаленні моделей даних процесів з метою їх прогнозування. За допомогою правильно побудованої моделі можна робити доволі точні прогнози економічних показників.

**Мета дослідження.** Головною метою дослідження є розгляд методів моделювання нелінійних нестационарних процесів (ННП) та їх застосування для побудови моделей фінансових процесів. Виконати оцінку побудованих моделей; застосувати їх для прогнозування розвитку процесу ціноутворення на нафту з метою виявлення найкращої з них.

**Методика реалізації.** Для розв'язання поставлених задач використано такі методи: розгляд існуючих рішень та методик для виявлення нелінійностей та нестационарності; підхід до по опису ННП на основі регресійних моделей; оцінювання якості побудованих моделей та прогнозів.

**Результати дослідження.** Основними результатами дослідження є такі: розглянуто основні типи ННП, за допомогою вибраних моделей регресії побудовані моделі ціноутворення на нафту, виконано прогнозування показників процесу, отримані статистичні характеристики, аналіз якості моделі та прогнозу.

**Висновки.** Таким чином, були розглянуті деякі підходи до побудови моделей та оцінок прогнозів в економіці та фінансах. Побудовано моделі авторегресії та авторегресії з трендом для опису ціноутворення на нафту. Статистичні характеристики отриманих моделей є доволі близькими, проте можна виділити найкращу з них: авторегресію першого порядку з квадратичним трендом. На графіках виділяється різкий спад, а потім поступовий під'йом. Вплив таких неконтрольованих збурень ускладнює побудову моделей та їх прогнозування.

**Ключові слова.** фінанси, економіка, нелінійність, нестационарність, регресія, нафта.

**Д.В. Костриця, П.І. Бідюк**

## **РОЗРОБКА СИСТЕМНОЇ МЕТОДОЛОГІЇ ПОБУДОВИ СКОРИНГОВОЇ КАРТИ**

**Проблематика.** Кредитоспроможність позичальника завжди була і залишається одним з основних критеріїв визначення доцільності встановлення кредитних відносин з клієнтом. Але високий рівень проблемних кредитів, що спостерігається в кредитних портфелях вітчизняних банків свідчить про недосконалість процесів визначення кредитоспроможності позичальників. У зв'язку з цим зростає актуальність питання дослідження сучасних методик та вдосконалення діючої методики оцінювання кредитоспроможності клієнтів банку. З огляду на важливість тієї ролі, яку відіграє саме оцінка кредитоспроможності позичальників в мінімізації кредитного ризику, в сучасних умовах постає актуальність питання вдосконалення існуючих методів.

**Мета дослідження.** Розробка скорингової системи на основі історичних даних для оцінювання ризику кредитування фізичних осіб.

**Методика реалізації.** Для побудови скорингової моделі розглянуто основні теоретичні відомості щодо математичних методів та підходів до інтелектуального аналізу даних, що застосовуються до моделювання і прогнозування банківських кредитних ризиків, та більш детально вивчено модель логістичної регресії.

**Результати дослідження.** Розроблено системну методологію скорингового аналізу, яка включає в себе попередню базову обробку даних, побудову математичної моделі і адаптацію результатів до банківської методики. За основу

скорингової моделі, яка розроблена у поданій роботі, взято логістичну регресію. Визначено коефіцієнти значимості аплікаційних параметрів фізичних осіб. Побудована скорингова карта використовується для поділу клієнтів банку на хороших та поганих. Виконано перевірку шляхом порівняння отриманих результатів із фактичними. Розроблено систему автоматизації певних кроків при побудові скорингу. Оскільки поставлена задача є досить специфічною, неможливо повністю автоматизувати побудову скорингової моделі, адже після кожного кроку необхідно переглядати результати.

**Висновки.** Описано докладну покрокову інструкцію розробки скорингової моделі, що базується на логістичній регресії. Запропоновано модифіковані формули підрахунку скорингових балів. Розроблено програмний продукт, який обробляє статистичні дані та рахує скорингові бали за описаною методикою. Серед подальших можливостей розвитку дослідження є розробка інструменту для інтерпретації скорингових балів. Які скорингові бали набирають зазвичай хороші клієнти, як провести градацію на скорингові групи – ось питання, над якими ще варто працювати. А також в перспективі доцільно застосувати деякі з нейронних мереж, наприклад, нейронні мережі ANFIS, TSK, які можуть вирішувати задачі оцінювання кредитоспроможності.

**Ключові слова:** кредитний ризик, логістична регресія, керування ризиками, скорингова модель, ймовірність дефолту, information value, аплікаційний скоринг, залежна змінна, показники ваги категорій.

**О.В. Надточій**

## **МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ В УКРАЇНІ**

**Проблематика.** Спроектвані та проаналізовані моделі демографічних процесів в Україні, побудовані прогнози на майбутні періоди. Це приводить до підвищення якості розроблюваних моделей та оцінок прогнозів.

**Мета дослідження.** Головна мета дослідження: зібрати статистичні дані стосовно розвитку демографічних процесів в Україні для виконання обчислювальних експериментів; вибрати типи математичних моделей для опису нестационарних демографічних процесів і побудувати їх у формі регресійних рівнянь на основі статистичних даних; обчислити коротко- та довгострокові прогнози вибраних демографічних процесів.

**Методика реалізації.** Для розв'язання поставлених задач були зібрані різні показники чисельності населення України, побудовані регресійні моделі, побудований прогноз на основі обраних моделей.

**Результати дослідження.** Основні результати дослідження: на основі вхідних статистичних даних було виконано прогнозування загальної зміни чисельності населення України, були розраховані критерії якості прогнозів і проведено їх аналіз.

**Висновки.** Таким чином, можна зробити висновок, що авторегресійна модель є досить простою з точки зору побудови та об'єму обчислювальних операцій і, при цьому, модель процесу та прогнозу являється адекватною. Чисельність населення України поступово скорочується. Для збільшення кількості населення, або хоча б підтримки кількості населення на стабільному рівні необхідно покращити економічний, політичний, соціальний та екологічний стани країни.

**Ключові слова:** демографічні показники чисельності населення; регресійні рівняння; авторегресійна модель; коротко- та довгостроковий прогноз.

**Пінчук В., Шишко Н.**

## **ПРОГНОЗУВАННЯ ФЕП СТАТИСТИЧНО-ЙМОВІРНІСНИМИ МЕТОДАМИ ТА КОМБІНУВАННЯ ОЦІНОК ПРОГНОЗІВ**

**Проблематика.** Для прийняття коректних рішень в економіці та фінансах необхідно використовувати сучасні методи математичного моделювання фінансово-економічних процесів (ФЕП) і чітко формувати альтернативи на основі оцінок прогнозів, отриманих за побудованими моделями. Сьогодні вже існує чимало розробок у цій галузі, отож крім проблеми розробки нових якісних методів прогнозування часових рядів виникає проблема вибору методів з уже наявних, а також постає задача комбінування отриманих оцінок прогнозів для покращення результатів.

**Мета дослідження.** Метою цієї роботи є аналіз кількох статистично-ймовірнісних методів прогнозування фінансово-економічних процесів та дослідження особливостей комбінування оцінок прогнозів, отриманих за цими методами.

**Методика реалізації.** Для розв'язання поставлених задач використано такі методи прогнозування: лінійна регресія, експоненційне згладжування, нейронна

мережа, мережа Байєса. Досліджено два способи комбінування – середнє арифметичне та зважене середнє оцінок прогнозів.

**Результати дослідження.** Основними результатами дослідження є такі: побудовано короткострокові прогнози для біржового індексу NYSE\_Composite за допомогою статистичних методів прогнозування та індексу Nasdaq100 за допомогою мережі Байєса з п'яти вузлів. Проведено комбінування оцінок прогнозів двома способами та порівняно результати. Отримані критерії якості виявились кращими ніж без використання комбінування, при чому майже для всіх методів. Проте, порівнюючи між собою два способи комбінування однозначно виділити кращий не можна.

**Висновки.** Отже, нами було виконано огляд наступних статистично-ймовірнісних методів прогнозування фінансово-економічних процесів: лінійна регресія, експоненційне згладжування, нейронна мережа, мережа Байєса. Проаналізовано їх переваги та недоліки. Розглянуті різні підходи до комбінування оцінок прогнозів. Експерименти та аналіз результатів проведено за допомогою програми GeNIe 2.0 та мови для статистичного аналізу даних R.

**Ключові слова:** фінансово-економічний процес, часовий ряд, прогнозування, коефіцієнт детермінації, мережа Байєса, нейромережа.

**Н.В. Кузнєцова, К.В. Куца, С.Р. Штогрін**

## **ЗАСТОСУВАННЯ МЕТОДОЛОГІЇ АНАЛІЗУ ВИЖИВАННЯ ДЛЯ ДОСЛІДЖЕННЯ СПОЖИВЧИХ РИЗИКІВ**

**Проблематика.** Аналіз споживчих ризиків банківського сектору є доволі популярною задачею. Зазвичай виконується аналіз кредитних заявок (аплікаційний скоринг) для з'ясування платоспроможності клієнта і прийняття рішення щодо видачі чи відмові у кредиту. Поряд з цим актуальною є задача оцінювання стану клієнта в процесі дії кредитного договору для визначення його поведінки щодо подальшого обслуговування кредиту.

**Мета дослідження.** На основі зібраних статистичних даних виконати порівняльний аналіз методів інтелектуального аналізу даних для прогнозування дефолту позичальника. Побудувати нові моделі для прогнозування поведінки клієнта і обрати кращу з них для моделювання споживчих ризиків.

**Методика реалізації.** Для аплікаційного скорингу побудувати моделі на основі нейронних мереж, логістичної регресії, лінійної регресії та виконати

порівняльний аналіз на предмет їх предикативної здатності. Побудувати моделі виживання на основі різних статистичних тестів для прогнозування критичного часу настання споживчого ризику.

**Результати дослідження.** Побудовані моделі для різних методів інтелектуального аналізу даних показали прийнятну точність прогнозування ймовірності переходу споживчого кредиту у стан дефолту. Найкращою моделлю за сукупністю статистичних критеріїв виявилась нейронна мережа багат шарового перцептронну з трьома прихованими шарами. За результатами побудови моделей виживання з використанням експоненційного, логістичного та розподілу Вейбулла, кращою виявилась модель з розподілом Вейбулла.

**Висновки.** Задача аналізу кредитоспроможності клієнтів, що звертаються за кредитними продуктами, має виконуватись у двох напрямках: аналізу клієнта на етапі прийняття рішення щодо видачі кредиту та аналізу його поведінки у процесі дії кредитної угоди. Для першої підзадачі доцільно будувати скорингові карти методами інтелектуального аналізу даних, а для поведінкового сокрингу доцільно будувати моделі виживання.

**Ключові слова:** споживчі ризики, аналіз виживання, нейронні мережі.

**І.М. Курін, О.Л. Тимощук**

## **МЕТОДИ НЕЛІНІЙНОЇ ОПТИМІЗАЦІЇ В ЗАДАЧАХ ПРОЕКТУВАННЯ ЕЛЕКТРИЧНИХ ДВИГУНІВ**

**Проблематика.** Побудова електричних двигунів із геометричними параметрами, які оптимізують економічні чи технічні показники і задовольняють систему обмежень інших показників двигуна, надає можливість ефективно та раціонально використовувати ресурси підприємств багатьох галузей промисловості.

**Мета дослідження.** Головна мета дослідження: розглянути існуючі числові алгоритми та методи оптимізації для проектування електричних двигунів. Розробити програмне забезпечення, яке знаходить оптимальні параметри для проектування однофазного асинхронного двигуна.

**Методика реалізації.** Для розв'язання поставлених задач використано такі методи: формалізація технічної задачі у вигляді задачі нелінійного програмування; чотири числові методи нелінійної оптимізації (метод штрафних

функцій, модифікований метод множників Лагранжа, метод комплексів, метод допустимих напрямків).

**Результати дослідження.** Основними результатами дослідження є такі: визначення найкращих методів нелінійного програмування серед обраних для оптимального проектування електричних двигунів; розробка програмного продукту, здатного шукати оптимальні геометричні параметри однофазного асинхронного двигуна.

**Висновки.** Таким чином, оптимально спроектувавши двигуни, кожен з яких мав свої технічні обмеження, було встановлено, що методи штрафних функцій та допустимих напрямків є нестійкими щодо обчислень. Було продемонстровано ефективність використання методу комплексів та модифікованого методу множників Лагранжа для оптимального проектування електричних двигунів. Створено програмний продукт, який дозволяє проводити проектування однофазного асинхронного двигуна із різними технічними обмеженнями.

**Ключові слова:** оптимальне проектування, однофазний асинхронний двигун, нелінійна оптимізація, математичне програмування.

## ВІДОМОСТІ ПРО АВТОРІВ

**Бідюк Петро Іванович** – д.т.н., професор Інституту прикладного системного аналізу Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Бойко Ірина Олександрівна** – студентка Інституту прикладного системного аналізу Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Діденко Дмитро Георгійович** – к.т.н., доцент кафедри ММСА Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Кангун Ганна Іллівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Каніовська Ірина Юріївна** – к.ф.-м.н., ММСА Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Кобера Валентина Василівна** – магістр ММСА Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Климковська Катерина Дмитрівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Климковська Станіслава Дмитрівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Конєв Дмитро Володимирович** – студент Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.



**Костриця Дарина Василівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Кузнєцова Наталія Володимирівна** – к.т.н., доцент кафедри ММСА Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Курін Ілля Миколайович** – студент Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Куца Каріна Володимирівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Надточій Олександра Вікторівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Пінчук Владислав Олександрович** – студент Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Тимошук Оксана Леонідівна** – завідувач кафедри ММСА, к.т.н., доцент кафедри ММСА Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Шишко Наталія Володимирівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ.

**Штогрін Світлана Романівна** – студентка Інституту прикладного системного аналізу НТУУ «КПІ ім. Сікорського», м. Київ