

**Н. В. Кузнєцова, Ф. А. Чан, М. В. Самсонюк, М. В. Юрчук**

Інститут прикладного системного аналізу  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»,  
пр. Перемоги, 37, 03056 Київ, Україна  
e-mail: natalia-kpi@ukr.net, panh13@gmail.com,  
ms-0807@ukr.net, myurchuk13@gmail.com

## **Динамічне оцінювання ризиків і розробка антиризикових стратегій банківської діяльності**

*Статтю присвячено розробці єдиного підходу до динамічного оцінювання ризиків і прогнозування можливих втрат банку. Запропоновано методологію, яка передбачає виконання основних етапів дослідження і емпіричного аналізу вхідних даних, оцінювання ймовірності та прогнозування очікуваних втрат, пов'язаних з можливим відтоком клієнтів. Показано два підходи до оцінювання втрат від появи ризиків: на основі втраченого балансу та через математичне сподівання втрат, зважених на ймовірність настання ризику відтоку клієнтів. Запропоновано індекс ризику та оцінювання «ваги» (внеску) клієнта у сукупний банківський баланс, що дозволило визначити допустимий, критичний і катастрофічний рівні та ступені ризику. Виконано дослідження періоду відтоку клієнтів із застосуванням моделей виживання. Розроблені математичні моделі оцінювання ймовірності та прогнозування втрат на основі методу опорних векторів, градієнтного бустингу XGBoost, випадкового лісу, логістичної регресії, наївного байєсівського класифікатора дозволили отримати високі прогнозні оцінки точності і якості прогнозів можливих втрат за обома підходами до оцінювання втрат. Напрацьовані стратегії і рекомендації для банку дозволяють скоригувати стратегію його розвитку з урахуванням визначених і оцінених рівнів ризиків.*

**Ключові слова:** банківські ризики, індекс ризику, ступінь ризику, логістична регресія, випадковий ліс, XGBoost, моделі виживання.

### **Вступ**

Банківська система є доволі потужною системою, яка має велику кількість зв'язків і взаємовідношень з іншими системами, клієнтами, державними органами. Вона є чутливою до зовнішніх та інформаційних впливів, які можуть нести за-

© Н. В. Кузнєцова, Ф. А. Чан, М. В. Самсонюк, М. В. Юрчук

грозу як іміджу банку і його фінансовій стабільності, так і впевненості клієнтів установи. Основним джерелом банківських фінансів є надходження від клієнтів, зокрема депозитів, повернення виданих клієнтам кредитів, відсотків за кредитами. Негативні інформаційні впливи можуть похитнути довіру клієнтів і спонукати їх до «відтоку». Причинами відтоку також може бути незадоволення певними послугами, технічними збоями, комісіями або тарифами банків і дії конкурентів, направлені на контрагітацію, які рекламують і розвивають свої власні сервіси, заохочуючи нових клієнтів приєднуватися до новітніх послуг конкуруючих банків. Все це обумовлює необхідність періодичного моніторингу та перевірки задоволеності клієнтів банків, проведення спеціальних програм і акцій лояльності, спрямованих на утримання власних клієнтів.

### **Постановка задачі**

Проаналізувати основні типи ризиків і загроз, характерних для сучасної банківської діяльності. Визначити ключові характеристики та допустимі рівні ризиків, які пов'язані з можливим відтоком клієнтів, і на прикладі цього виду ризику розробити математичні моделі, що дозволять класифікувати клієнтів, які можуть у найближчий час забрати гроші з банку, а також спрогнозувати час і період таких дій. Розробити методологію аналізу і оцінювання ризиків через імовірність їхнього виникнення та можливі втрати. Запропонувати підходи до обчислення можливих втрат, пов'язаних з відтоком клієнтів, а також індикатори ризикованості таких клієнтів і їхній вплив на банківську систему в цілому. Застосувати апарат інтелектуального аналізу даних для прогнозування ймовірності відтоку клієнтів, застосувати моделі теорії виживання для прогнозування періоду відтоку. Визначити критичний, допустимий і катастрофічний рівні та ступені ризику, оцінити можливі втрати та кількість клієнтів, які здійснять відток. Запропонувати антиризикові стратегії, що пов'язані з виявленням клієнтів, схильних до відтоку, та напрацювати рекомендації для їхнього утримання та забезпечення стабільності банку.

### **Ризики банківської діяльності**

Банківська система зазнає значну кількість ризиків, зокрема через зовнішні зміни та впливи, які можуть знизити ефективність її діяльності. До таких ризиків можна віднести: ризик дефолту банку; ризик загострення економічної кризи в країні, що призведе до порушення фінансової стабільності банку; ризик втрати репутації (втрата довіри до банку з боку вкладників може призвести до відпливу коштів і неплатоспроможності); геолокаційні ризики (ризик місцезнаходження клієнта, специфічні закони тієї чи іншої країни); ринковий ризик (ризик зміни ринкових цін, а також зміни параметрів ринку, таких як відсоткові ставки, курси валют, ціни акцій чи товарів, кореляція між різними показниками ринку та зміна цих показників); ризик впровадження нових банківських продуктів, послуг (імовірність не досягти запланованого рівня окупності нових банківських продуктів, послуг, операцій чи технологій); ризик зловживань (несанкціоноване укладання угод з клієнтами); операційний ризик (порушення внутрішніх процесів); ризик порушення кредитоспроможності клієнта(-ів) [1]. Останній вид ризику є найбільш досліджений — існують спеціальні скорингові моделі [2, 3], методи, які розробляю-

ться для конкретного типу ризику, інструменти (скорингові карти) для опрацювання наданих клієнтом даних на робочих місцях, універсальні інформаційні технології, які дозволяють приймати рішення щодо видачі кредитів за лічені хвилини. Описані види ризиків ранжують за ймовірністю та наслідками у вигляді матриці ризиків, а також за ступенем їхньої небезпеки для банківської діяльності «за світлофором Базеля» у вигляді зеленої, жовтої і червоної зон.

### Математичний апарат і методика вирішення задачі оцінювання та прогнозування ризиків

Постановка задачі передбачає вирішення декількох послідовних задач і напрацювання рекомендацій щодо менеджменту ризиків, тож методику вирішення задачі можна описати у вигляді наступних кроків.

**Етап 1.** Формалізація задачі, визначення ключових змінних і цільової змінної для прогнозування.

**Етап 2.** Вирішення задачі класифікації і прогнозування ймовірності настання відтоку як імовірності настання ризику.

**Етап 3.** Оцінювання очікуваних втрат для спрогнозованої групи клієнтів, що ймовірно здійснюватимуть відток.

Для прогнозування очікуваних втрат було використано два ідеологічно різні підходи. **Перший підхід** полягав у тому, що втрати банку оцінювались як «втрачені можливості», тобто сума коштів на балансах клієнтів, які найімовірніше здійснять відток:

$$Losses = \sum_i balance_i, \quad (1)$$

де  $balance_i$  — баланс на рахунку  $i$ -го клієнта, який здійснює відток.

На ці кошти у вигляді депозитів, вкладів, поточних рахунків банк розраховував і наводив у фінансовій звітності, підтверджуючи власну фінансову стабільність.

**Другий підхід** полягає в тому, що втрати банку розглядаються як математичне сподівання втрат, зважене на ймовірність відтоку кожного клієнта. За допомогою відповідних моделей оцінюються (прогнозуються) ймовірності втрати кожного клієнта, а обсяг втрат визначається за формулою

$$Losses = \sum_i balance_i \times P_i, \quad (2)$$

де  $P_i$  — ймовірність втрати  $i$ -го клієнта;  $balance_i$  — баланс на рахунку  $i$ -го клієнта.

**Етап 4.** Визначення «ваги» кожного клієнта для вартісної оцінки його відтоку та прогнозування втрат.

Для прогнозування ступеню ризику авторами було запропоновано оцінювання вагомості внеску кожного клієнта та його балансу у загальний баланс і фінансову стабільність банку.

Для визначення ймовірності та ступенів ризику, перш за все, необхідно оцінити ймовірність відтоку кожного клієнта. Для цього було використано модель

логістичної регресії (оскільки вона показала найточніший результат за втратами), а ваговий вклад оцінюється за наступною формулою:

$$w_i = \frac{b_i}{\sum_i b_i}, \quad (3)$$

де  $w_i$  — ваговий внесок  $i$ -го клієнта;  $b_i$  — баланс на картці(-ах)  $i$ -го клієнта,  $\sum_i b_i$  — сума вкладів за усіма клієнтами банку.

Далі визначається індекс ризику кожного клієнта за наступною формулою:

$$RI_i = w_i P_i, \quad (4)$$

де  $RI_i$  — індекс ризику  $i$ -го клієнта (ступінь ризикованості клієнта);  $w_i$  — ваговий внесок;  $P_i$  — ймовірність відтоку  $i$ -го клієнта.

**Етап 5.** Прогнозування періоду можливого відтоку клієнтів.

**Етап 6.** Напрацювання рекомендацій і стратегій для зменшення ймовірності відтоку, втрат і долі клієнтів, схильних до відтоку.

Оскільки на перших етапах вирішення задачі передбачається розв'язання задачі класифікації, то доцільно використати методи класифікації, а також застосувати логістичну регресію для перетворення і оцінювання ймовірності віднесення до кожного класу. Для оцінювання ймовірності відтоку рекомендується використовувати ймовірнісний байєсівський підхід. Було відібрано такі основні методи і алгоритми: випадковий ліс, метод  $k$ -найближчих сусідів, метод опорних векторів, алгоритм градієнтного бустингу XGboost, логістична регресія, наївний байєсівський класифікатор. Для прогнозування втрат було використано два підходи: на основі математичного сподівання відтоку та на основі «внеску» кожного клієнта в сукупний дохід банку. Прогнозування періоду відтоку клієнтів було виконано із використанням моделей з теорії виживання. Розглянемо коротко основні моменти згаданих методів та алгоритмів.

## Основні методи та моделі, використані у дослідженні

**Випадковий ліс** (random forest) — це ансамблевий метод машинного навчання, що використовується для задач регресії і класифікації і представляє собою множину дерев рішень [4]. У задачі регресії прогнози кожного дерева рішення усереднюються, а в задачі класифікації приймається рішення голосуванням за більшістю. Перевагами даного методу є його потужність, точність, можливість використання для широкого кола задач, включаючи лінійні та нелінійні задачі. Недоліками є складність інтерпретації, можливість перенавчання, необхідність задання критерію для підбору раціональної кількості дерев рішень.

**Метод опорних векторів** (SVM) — це набір алгоритмів «навчання з учителем», які досить ефективно використовуються в задачах класифікації. У цьому методі точка в просторі розглядається як вектор розмірності  $p$ . Необхідно визначити, чи можна розділити дані точки гіперплощиною розмірністю  $p-1$ . Таку гіперплощину називають *площиною класифікатора*. У випадку, якщо нова точка потрапляє в ту чи іншу область, які розділені отриманою гіперплощиною, вона

потрапляє у відповідний клас [5]. Звичайно, дані можуть бути розділені за допомогою різних гіперплощин. Найкраща гіперплощина — це така гіперплощина, при побудові якої розділення та розбіжність між двома класами відбувається максимально. Перевагами методу є його висока ефективність і нечутливість до перенавчання. Недоліком є те, що він не дуже добре працює з нелінійними задачами і задачами, в яких використовується велика кількість факторів.

**Алгоритм градієнтного бустингу XGBoost.** Градієнтний бустинг — це техніка машинного навчання для задач класифікації і регресії, яка будує модель передбачення у формі ансамблю прогнозуючих моделей, зазвичай дерев рішень [6, 7]. Навчання ансамблю проводиться послідовно, на відміну, наприклад, від бегінга. На кожній ітерації обчислюються відхилення прогнозів уже навченого ансамблю на навчальній вибірці. Наступна модель, яка буде додана в ансамбль, буде передбачати ці відхилення. Таким чином, додавши передбачення нового дерева до пророкувань навченого ансамблю, можна зменшити середнє відхилення моделі. Нові дерева додаються до ансамблю до тих пір, поки помилка зменшується, або поки не виконується одне з правил «ранньої зупинки».

**Метод  $k$ -найближчих сусідів (KNearest)** — метричний алгоритм для автоматичної класифікації об'єктів або регресії до найближчих для нього об'єктів, значення яких уже відомі. Алгоритм може бути застосований до вибору з великою кількістю атрибутів (багатовимірних) [8].

Перевагами алгоритму є його швидкість й ефективність і легкість для розуміння, а недоліком є те, що необхідно визначити оптимальну кількість сусідніх точок  $k$ , які беруть участь у голосуванні.

**Логістична регресія (Logistic regression)** — метод побудови лінійного класифікатора, що дозволяє оцінювати апостеріорні ймовірності приналежності об'єктів класам. У логістичній регресійній моделі прогнозні значення (залежна змінна або відклик) не може бути меншим (або рівним) 0, або більше (або рівним) 1, за будь-яких значень незалежних змінних [9]. Ця модель часто використовується для аналізу бінарних залежних відкликів і прогнозування ймовірності одного з двох класів.

При цьому використовується наступне рівняння регресії:

$$y = x_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n,$$

де  $b_n$  — параметри, що оцінюються методом максимальної правдоподібності;  $x_n$

— фактори (незалежні змінні), а  $p = \frac{1}{(1 + e^{-y})}$ .

Перевагами методу є зручність його використання для прогнозу ймовірності настання події, можливість отримання інформації про статистичну значущість факторів. Недоліками є накладення припущення щодо наявності лінійних залежностей, вимушене використання методу максимальної правдоподібності, що може призвести до невисокого відсотку правильності прогнозу.

**Наївний байєсівський алгоритм класифікації (Naive Bayes)** — це алгоритм класифікації, заснований на теоремі Байєса з припущенням про незалежність ознак. Іншими словами, він передбачає, що наявність якої-небудь ознаки в класі не пов'язано з наявністю будь-якої іншої ознаки. Моделі на основі цього методу до-

суть прості та корисні при роботі з дуже великими наборами даних. Не зважаючи на свою простоту, алгоритм здатний перевершити точність складних алгоритмів класифікації.

Теорема Байєса дозволяє розрахувати апостеріорну ймовірність  $P(c|x)$  на основі  $P(c)$ ,  $P(x)$ ,  $P(x|c)$ :

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)},$$

де  $P(c|x)$  — апостеріорна ймовірність даного класу  $P(c)$  (тобто даного значення цільової змінної) при даному значенні ознаки  $x$ ;  $P(c)$  — апіорна ймовірність даного класу;  $P(x|c)$  — правдоподібність, тобто ймовірність даного значення ознаки при даному класі;  $P(x)$  — апіорна ймовірність даного значення ознаки.

Перевагами методу є його ефективність і здатність вирішувати нелінійні задачі, надавати оцінку у вигляді ймовірності, а недоліком вважається припущення щодо незалежності ознак.

Окрема частина дослідження була присвячена авторами розробці динамічної складової методології менеджменту ризиків, яка передбачає прогнозування ризиків з можливістю визначення завчасно моменту переходу ризику на більш високий ступінь або рівень. У попередніх дослідженнях [1, 10] було запропоновано застосування *теорії виживання* і розроблено динамічні моделі різних типів для прогнозування динаміки зміни ризиків у часі. Було запропоновано методи і алгоритми визначення моменту переходу ризиків на більш високий ступінь чи рівень та на конкретному прикладі показано можливість динамічного оцінювання ризиків у реальному часі [10, 11].

Для аналізу використовується вибірка (популяція), яка характеризується тим, що для кожного об'єкта відомий результат події (загибель чи виживання). Далі здійснюється один із видів цензурування (відсікання). Спостереження називаються цензурованими, якщо спостережувана залежна змінна представляє момент настання термінальної події, а тривалість дослідження обмежена за часом [1].

*Функція виживання* визначається:  $S(t) = P(T > t)$ , а *функція ризику*:

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta | T > t)}{\Delta}, \quad h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)}.$$

Найпростіша функція, яка визначає, що ризик є константою в часі:  $h(t) = \lambda$ , або що еквівалентно  $\log h(t) = \mu$  [11]. Оскільки  $S(t) = \exp[-\int_0^t h(u)du]$ , то після під-

становки та інтегрування отримуємо:  $S(t) = e^{-\lambda t}$ , а  $f(t) = \lambda e^{-\lambda t}$ . Це функція щільності ймовірності з відомим експоненційним розподілом з параметром  $\lambda$ . Таким чином, сталий ризик передбачає експоненційний розподіл до моменту часу, поки не наступить подія (або час між подіями) [12]. На практиці найбільш поширеними є динамічні моделі на основі пропорційних ризиків Кокса та їхні модифікації у вигляді лінійних і непараметричних моделей [1, 13, 14].

## Імітаційне моделювання для аналізу та прогнозування ризиків

**Етап 1. Формалізація задачі, попередня обробка та аналіз вхідних даних**

Вхідними даними для розробки антиризикової стратегії було обрано реальні дані клієнтів банку. Дані містили інформацію щодо основних змінних клієнтів, таких як: ідентифікатор, прізвище, вік, стать, країна проживання клієнта; банківські характеристики: кредитний бал клієнта (отриманий за скоринговою моделлю), кількість років, коли він є клієнтом банку; фінансові змінні: баланс клієнта, приблизна зарплата клієнта, кількість послуг, до яких підключена картка клієнта; флагові змінні: чи має клієнт кредитну картку, чи веде активну банківську діяльність; цільову змінну (Exited) — чи задоволений клієнт обслуговуванням і залишається клієнтом, чи він здійснив відток. Оскільки певні дані (прізвище та ідентифікатор) є конфіденційними та захищеними банківською таємницею і не впливають на майбутні прогнози, то вони були видалені з подальшого аналізу. Слід зазначити, що вхідна вибірка, яка складалася з 10000 записів, є незбалансованою, відбиває реальну ситуацію в банку, де 79,6 % клієнтів залишаються і обслуговуються далі, а 20,4 % залишають банк. Такий відсоток клієнтів, що зупинили своє обслуговування в банку є критичним, тому банку необхідно терміново вжити заходів для виявлення клієнтів, що мають намір піти, і розробити стратегію для їхнього утримання.

Виконаємо попередній аналіз вхідних даних, зокрема проведемо дослідження залежності цільової змінної (Exited) від інших змінних (залежність від країни клієнта, його віку, статі). Основні частотні розподіли наведено в табл. 1.

Таблиця 1. Розподіл клієнтів залежно від різних ознак

	Франція	Німеччина	Іспанія
Загалом	5012	2510	2478
Здійснили відток	810	814	413
Залишились	2065	1696	4202
Відсоток відтоку	16,16 %	32,43 %	16,66 %
	Чоловіки		Жінки
Загалом	5457		4543
Здійснили відток	898		1139
Залишились	4559		3404
Відсоток відтоку	16,46 %		25,07 %
	Мають кредитну картку		Не мають кредитної карти
Загалом	7055		2945
Здійснили відток	1424		613
Залишились	5631		2332
Відсоток відтоку	20,18 %		20,81 %

Очевидно, що найбільшу кількість клієнтів банк має з Франції, а менше всього — з Німеччини. При цьому частка відтоку клієнтів обернено пропорційна кількості клієнтів у тій чи іншій країні. Скоріше за все, це пов'язано з тим, що

банк має певні проблеми в тих районах, де клієнтів менше. Загалом кількість клієнтів-чоловіків переважає в банку над кількістю клієток-жінок, проте частка відтоку жінок — вища, ніж у чоловіків.

Наступним кроком було проведено дослідження впливу віку клієнта на ймовірність відтоку. Всі клієнти були розподілені на такі вікові категорії (приблизно однакові за розміром): 1) від 18 до 32 років; 2) від 32 до 37 років; 3) від 37 до 44; 4) від 44 до 93 років.

Частотний розподіл відтоку клієнтів залежно від віку наведений на рис. 1,а. Як видно з рисунку, банк більше орієнтований на клієнтів молодшого віку, тому і кількість клієнтів груп 1–3 є більшою, порівняно з групою 4. Тому і відсоток клієнтів, які залишають банк, є більшою в групі 4. Ще одну залежність відтоку клієнтів від характеристики «кількість підключених послуг» показано на рис. 1,б. В основному, клієнти підключені до одного або двох додатків. Клієнти, які підключені лише до одного додатку, значно частіше залишають банк. Наступним кроком було проведення аналізу залежності відтоку клієнта від стану його активності, тобто використання додаткових послуг і сервісів (рис. 1,в). Згідно з отриманим частотним розподілом, клієнти, які активно користуються банківськими послугами, здійснюють відтік рідше, ніж клієнти, які послугами не користуються. Це пов'язано насамперед з тим, що «пасивні» клієнти не зацікавлені в користуванні послугами банку взагалі.

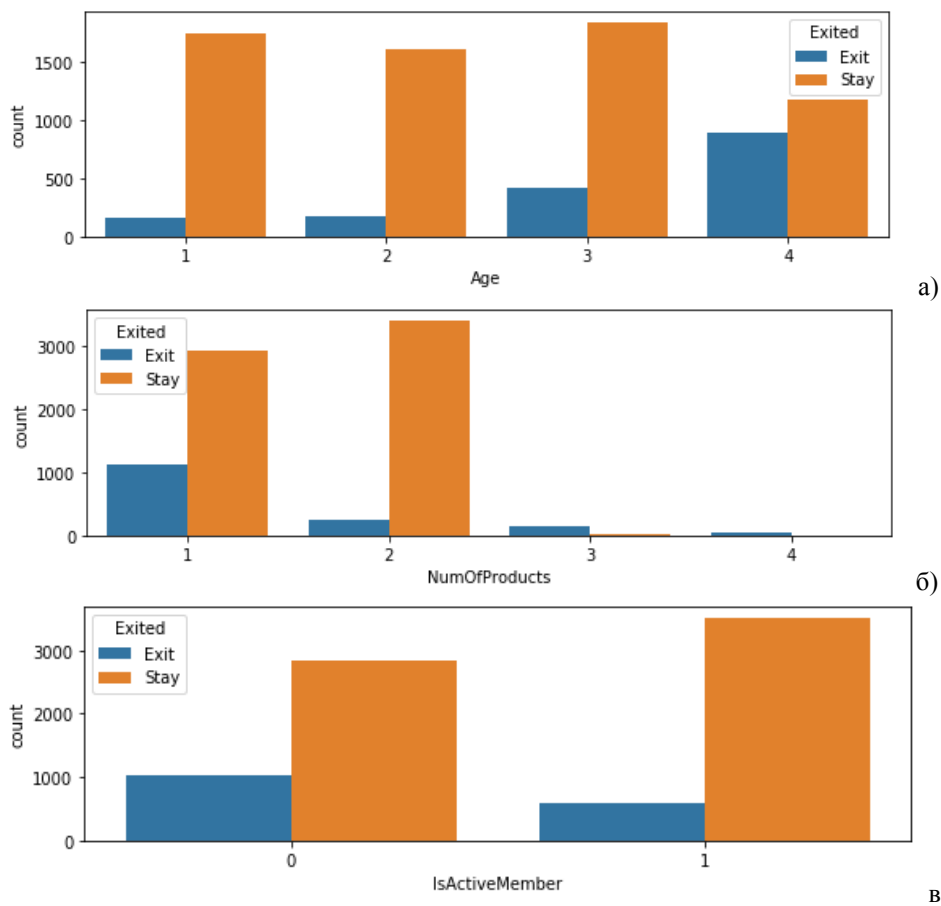


Рис. 1. Залежність цільової змінної від: а) віку; б) кількості підключених додатків; в) активності клієнта



Аналіз розсіювання балансу на рахунку клієнта залежно від його заробітної плати показав, що дані змінні розподілені доволі рівномірно, сильних відхилень не спостерігається.

Після аналізу усіх змінних набору даних для класифікації було вирішено обрати наступні фактори: CreditScore, Geography, Gender, Age, Tenure, Num of prod, HasCrCard, IsActive, EstimatedSalary, Balance та використати описані вище методи і алгоритми.

**Етап 2. Побудова моделей класифікації клієнтів, схильних до відтоку, та прогнозування ймовірності відтоку**

Навчання відбувалося на навчальній вибірці, що складала 80 % (8000 клієнтів) від усього вхідного набору даних, тестова вибірка становила 20 % (2000 клієнтів). Усі подальші оцінки якості методів і моделей проводилися саме на тестовій вибірці. Загалом реальні сумарні втрати для клієнтів з тестової вибірки склали 36,708,330\$, тоді як увесь сумарний баланс на картках даних клієнтів 153,809,233\$.

Точність побудованих моделей на основі визначених вище методів і алгоритмів представлено на рис. 2.

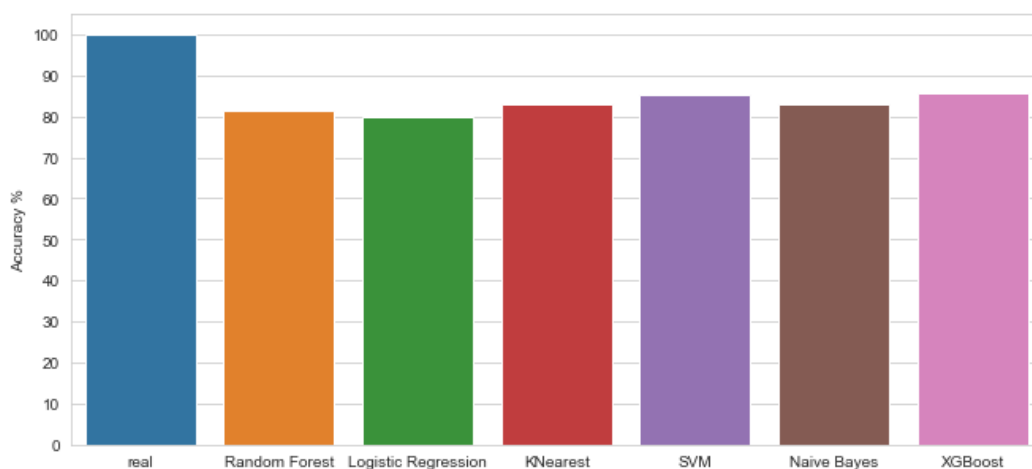


Рис. 2. Точність побудованих моделей на тестовій вибірці

**Етап 3. Прогнозування обсягу очікуваних втрат за двома підходами**

Спрогнозувавши тих клієнтів, які здійснять відток, просумуємо баланс на їхніх картках, і таким чином визначимо прогнозовані втрати за першим підходом (за формулою (1)). Точність очікуваних втрат, спрогнозованих різними моделями на тестовій вибірці, порівняно з реальними, представлено в табл. 2.

Отже, найбільш точними моделями за показником загальної точності виявилися моделі на основі опорних векторів (SVM) та XGBoost (табл. 2), проте найбільш точно прогнозує втрати модель логістичної регресії.

Наступним кроком спрогнозуємо очікувані втрати за другим підходом, тобто через математичне сподівання за формулою (2). Найбільш схожі до реальних значення надали ті ж моделі, які найкраще спрогнозували ймовірність відтоку клієнтів.

Результати прогнозування втрат як математичного сподівання різними моделями представлено на рис. 3.

Таблиця 2. Порівняльна таблиця прогнозованих втрат різними моделями

Модель	Точність моделі	Спрогнозовані втрати
Random Forest	0,8155	41032548,81
Logistic Regression	0,798	35720191,14
Knearest	0,8295	35009989,8
SVM	0,8535	23916986,61
Naive Bayes	0,828	33241787,14
XGBoost	0,858	32572783,93
Реальні очікувані втрати		36708330,81

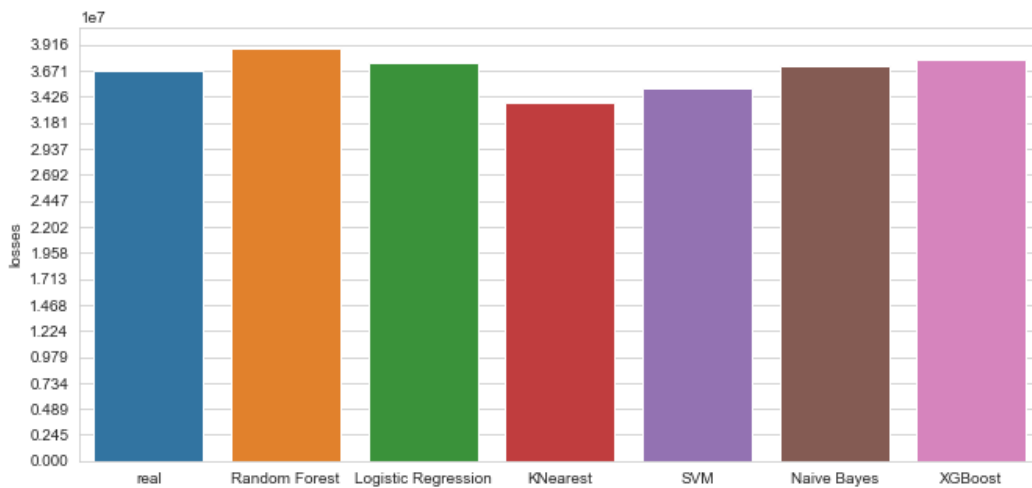


Рис. 3. Спрогнозовані втрати (як математичне сподівання) різними моделями

Оскільки візуальне порівняння таких великих втрат не є зручним, отримані результати були представлені у процентному вигляді абсолютних похибок прогнозування. Зведені результати наведені у табл. 3. Очевидно, що другий підхід оцінювання втрат на основі математичного сподівання виявився набагато точнішим, а найкращими методами виявилися наївний байесівський класифікатор і логістична регресія.

Таблиця 3. Середні абсолютні похибки прогнозування у процентах (САПП) обсягу втрат за обома підходами для кожної моделі

Модель	САПП (підхід 1), %	САПП (підхід 2), %
Random Forest	11,77993634	5,71816
Logistic Regression	2,691865792	2,176258
KNearest	4,62658193	8,049154
SVM	35,21616678	4,676671
Naive Bayes	9,443477607	1,313089
XGBoost	11,26596037	2,835378

#### Етап 4. Оцінювання ступеня ризику, ваги та індексу ризику кожного клієнта

За побудованими на попередніх кроках моделями для прогнозування ймовірності відтоку визначимо ймовірність відтоку кожного клієнта та його ваговий внесок за формулою (3). Далі розраховуємо індекс ризику для кожного клієнта за формулою (4) (результати представлено у вигляді табл. 4).

Таблиця 4. Зведені результати оцінювання та прогнозування для 11 перших клієнтів з тестової вибірки

№	Спрогнозована ймовірність відтоку клієнтів	Вагові внески клієнтів	Індекси ризику клієнтів
1	12,039	0	0
2	20,3955	0,000851859	0,0173741
3	12,8481	0	0
4	21,7295	0,000906273	0,0196929
5	12,2439	0,000505996	0,000619534
6	14,7657	0	0
7	5,94153	0,000849638	0,00504815
8	28,5868	0,00120633	0,034485
9	39,0981	0,000520133	0,0203362
10	21,9636	0	0
11	37,0804	0,000721734	0,0267622

Загальна сума індексів ризиків за всіма клієнтами визначає загальний сукупний ризик банку на даний момент. Для нашої тестової вибірки сума індексів ризику досягнула 24,385532, тобто це рівень максимального ризику, що може виникнути (якщо всі клієнти, які були спрогнозовані моделлю, як імовірні до відтоку, підуть з банку). Це значна частка клієнтів і, дійсно, є вже катастрофічною для банку. Для визначення допустимого, критичного та катастрофічного рівнів ризику було обрано такі частки (долі) з усього набору спрогнозованих ризикових клієнтів як 30 %, 60 % та 80 % від максимального ризику. За побудованими таблицями індексу ризику клієнтів і ймовірностями відтоку клієнтів, було визначено наступне.

1. Імовірність допустимого ризику складає 43 % (це відповідає ситуації, коли з банку підуть клієнти, ймовірність відтоку яких більша або рівна 43 %). При цьому сумарний рівень ризику такої ситуації складе 30 % від максимального ризику, тобто 7,315659573453298).

2. Імовірність критичного ризику складає 28 % (якщо підуть усі клієнти з імовірністю відтоку більшою або рівною 28 %, то сумарний рівень ризику складе 60 % від максимального, тобто 14,631319146906597).

3. Імовірність катастрофічного ризику складає 19 % (якщо підуть всі клієнти з імовірністю відтоку більшою або рівною 19 %, то сумарний рівень ризику складе 80 % від максимального, тобто 19,508425529208797).

Зведена таблиця ступенів і рівнів ризику залежно від кількості клієнтів, що здійснюють відток, а також сумарні втрати наведено в табл. 5.

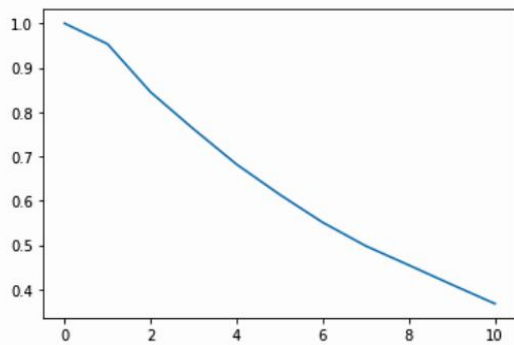
На попередньому етапі було спрогнозовано очікувані можливі втрати за обома підходами, які склали 37,190,344\$ (порівняно з реальними втратами за цією вибіркою 36,708,330\$). Отже, за нашими прогнозами, втрати будуть у межах між допустимим і критичним ризиками.

Таблиця 5. Рівні, ступені ризику (ймовірності та втрат) у результаті моделювання

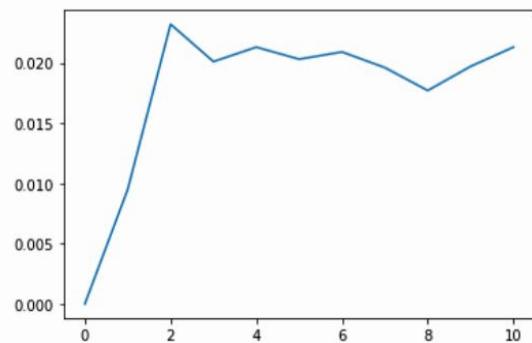
Рівень ризику	Відсоток втрат	Втрати	Кількість клієнтів, які здійснять відток	Частка клієнтів, які здійснять відток	Імовірність
Допустимий	30 %	19,847,611\$	176	7,32	43 %
Критичний	60 %	52,819,870\$	497	14,63	28 %
Катастрофічний	80 %	82,381,771\$	854	19,51	19 %

**Етап 5. Прогнозування періоду відтоку клієнтів**

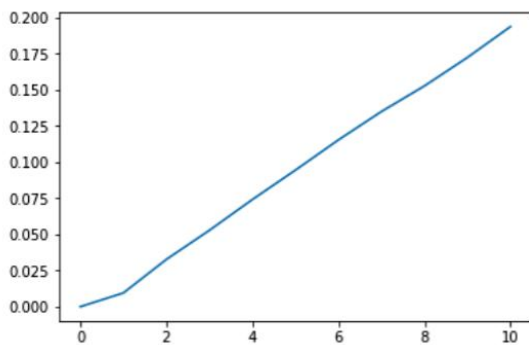
Для прогнозування можливого часу відтоку було зроблено вибірку з 10000 спостережень і виконано цензурування клієнтів. У результаті було відцензуровано 2037 клієнтів і побудовано модель виживання. На графіку (рис. 4) видно, що на першому році користування послугами відток клієнтів відбувається повільно (з імовірністю відтоку 0,05), далі ймовірність постійно зростає (на другому році до 0,16), на третьому 0,24 і т.д. Наступним етапом було оцінювання можливих втрат банку через відтік певної частки клієнтів за допомогою моделей виживання. Для цього було побудовано і оцінено функцію рівня ризику залежно від часу обслуговування клієнта. Також видно, що спочатку ризик сильно росте з піком на другому році, а потім залишається більш-менш стабільним. Сама частка клієнтів, які здійснюють відток, є не дуже великою у загальній вибірці, тому й частка ризику hazard rate становить 0,023 (рис. 4).



а)



б)



в)

Рис. 4. Моделювання періоду відтоку на основі теорії виживання:  
 а) функція виживання залежно від кількості років обслуговування;  
 б) частка ризику в часі;  
 в) сукупна частка ризику в часі

### **Етап 6. Напрацювання рекомендацій і стратегій для зменшення ймовірності відтоку, втрат і долі клієнтів, схильних до відтоку**

Найважливішою частиною після проведення оцінювання та прогнозування ризиків є розробка антиризикових стратегій і дій у разі настання ризику. Оскільки здійснювалося моделювання, то є можливість визначити та проаналізувати основні характеристики для найбільш ризикових клієнтів з метою виявлення групи найменш лояльних до банку клієнтів і проведення заохочуваних дій для того, щоб вони продовжили своє обслуговування в банку. Ризиковість клієнтів визначається за індексом ризику, обрахованим на четвертому етапі. У списку найбільш ризикових клієнтів частіше зустрічаються жінки, ніж чоловіки. Скоріш за все, вони почувуються недостатньо захищеними в банку, можливо, на них впливають компанії, що наносять втрати іміджу банку, або не мають такої ж кількості привілеїв, як чоловіки. Для запобігання цьому банк може спробувати ввести програми, які можуть зацікавити лише жінок. Наприклад, поширеною практикою є партнерство банків зі страховими, косметичними та іншими компаніями. У цьому випадку, наприклад, банк може спробувати домовитися з партнером-страховиком про вигідні умови страхування вагітних. Або, як варіант, банк може домовитися про додаткові знижки чи подарунки у найпопулярніших салонах і косметичних компаніях. Крім того, банк може запропонувати лояльніші умови для сімей, тобто коли у банку є картка тата, мами та їхніх дітей. За рахунок цього банк може збалансувати гендер своїх клієнтів.

Далі, бачимо, що більшість ризикових клієнтів живуть у Німеччині. У зв'язку з цим банк може спробувати націлити покращення обслуговування саме в Німеччині за рахунок введення окремої гарячої лінії німецькою мовою та збільшення штату особистих банкірів, які знають німецьку. Це зробить користування банкінгом для німців комфортнішим.

Також прослідковуємо те, що більший індекс ризику мають старші клієнти. Середній вік усіх тестових клієнтів — 38 років. Тим не менш, вік найбільш ризикових у середньому починається з 41 року і росте догори. Саме через це банку варто вводити лояльніші умови для старших клієнтів. Наприклад, запропонувати вигідні пропозиції і спеціальні програми з більшим трудовим стажем або більший відсоток на депозит для пенсіонерів.

Для нових клієнтів банк може запропонувати перший період безкоштовного обслуговування особистим банкіром, який буде підказувати та пояснювати суть і вигоди тих чи інших банківських послуг — користувач має бути зацікавленим. Крім того, можна ввести цікаві нагороди у вигляді банківського мерчу, наприклад, за першу покупку, за перший депозит, перший місяць, квартал, півроку та рік, проведений разом з банком. Для старих же клієнтів потрібно ввести додаткові послуги вищого рівня. Наприклад, створити спеціальну картку для найвідданіших клієнтів (8+ чи 9+ років обслуговування в банку), яка надаватиме кращі умови для клієнтів.

Важливим показником також є баланс на карті клієнтів. Найбільш ризиковими є клієнти з дійсно високим балансом, тому банк має вводити відповідні політики, які будуть ставити лояльніші умови зі збільшенням коштів на рахунок. Наприклад, при перетині певних порогів, клієнту можна виділити безкоштовного

особистого банкіра, спеціальне обслуговування безпосередньо в банку, кращі умови та акції у партнерів банку тощо.

Явно прослідковується, що клієнти, які не мають активної банківської історії, більш схильні до відтоку з банку, що є доволі очевидним. Тому банку при фіксації зменшення активності клієнта за його банківською картою слід спробувати нагадати про можливості та банківські послуги через мейл або смс-розсилку. Якщо ж користувач не почне знову користуватися карткою, то спробувати зв'язатися з ним у телефонному режимі для вирішення проблеми в індивідуальному порядку.

Ще один важливий висновок, який було отримано за допомогою моделей виживання, що в банку досить стабільний відтік з роками користування продуктами. Це може свідчити про те, що банківські продукти та можливості відстають від конкурентів або ще не розроблені зручні послуги. В такому випадку банку слід провести аналіз конкурентів і наявних у них банківських сервісів і продуктів та визначити, що потрібно покращити.

## Висновки

Запропоновано методику оцінювання та прогнозування ризиків, яка дозволяє спрогнозувати ймовірність настання ризику, визначити обсяги можливих ризиків, внесок клієнта у сукупний баланс банку, індекс ризику кожного клієнта та сукупний ризик для банківської діяльності. Вона дозволяє банку коригувати рівні та ступені ризику в оперативному порядку залежно від змін середовища та зовнішніх впливів. Проаналізовано основні показники клієнтів, що обслуговуються в банку, та виявлено суттєві фактори, які впливають на відтік клієнтів. На основі цих факторів було вирішено задачу класифікації, яка полягала у прогнозуванні клієнтів, які залишать банк. Розраховано пов'язані з цим втрати банку. На основі отриманої класифікації було розглянуто різні види ризиків, обраховані ймовірності їхнього настання. Виконані в роботі дослідження та моделювання дозволили визначити як найбільш ефективний — підхід до оцінювання можливих втрат через оцінювання математичного сподівання втрат. На базі проведеного моделювання та прогнозування було запропоновано основні антиризикові стратегії, окремі підходи для різних груп клієнтів досліджених банків, рекомендації для зниження ймовірності виникнення ризиків. Розроблені стратегії можуть легко поєднуватися та комбінуватися між собою для досягнення кращого результату.

1. Кузнецова Н.В., Бідюк П.І. Теорія і практика аналізу фінансових ризиків: системний підхід/монографія. Київ: Вид-во «Ліра-К», 2020. 400 с.

2. Сиддики Н. Скоринговые карты для оценки кредитных рисков: пер. с англ. Е. Ильичева. Москва: Манн, Иванов и Фербер, 2014. 268 с.

3. Кузнецова Н.В., Бідюк П.І. Моделирование кредитных рисков на основе теории выживания. *Проблемы управления и информатики*. 2017. № 6. С. 33–46.

4. Breiman L. Random forests. *Machine Learning*. 2001. Vol. 45, No. 1. P. 5–32. <https://doi.org/10.1023/A:1010933404324>

5. Cortes C., Vapnik V. Support-vector networks. *Machine Learning*. 1995. Vol. 20, No. 3. P. 273–297.

6. Friedman J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. 2001. Vol. 29, No. 5. P. 1189–1232. URL: <https://www.jstor.org/stable/2699986?origin=JSTOR-pdf&seq=1>

7. Friedman J. H. On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*. 1997. Vol. 1, No. 1. P. 55–77.
8. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы системной идентификации. Тольятти: ИЭВБ РАН, 2003. 463 с.
9. Kuznietsova N.V., Seebauer M., Zabelin S. Some methods for estimating financial risks in banking. *IEEE 1st Conf. on System Analysis and Intelligent Computing, SAIC 2018*. 2018. P. 271–274. doi: <https://ieeexplore.ieee.org/document/8516873>.
10. Кузнецова Н.В. Динамічний метод оцінювання ризиків у системі фінансового менеджменту. *Реєстрація, зберігання і оброб. даних*. 2019. Т. 21. № 3. С. 85–98. <https://doi.org/10.35681/1560-9189.2019.21.3.183724>
11. Кузнецова Н.В., Бідюк П.І. Динамічне моделювання фінансових ризиків. *Індуктивне моделювання складних систем*. 2017. Вип. 9. С. 122–137. URL: [http://nbuv.gov.ua/UJRN/Imss\\_2017\\_9\\_15](http://nbuv.gov.ua/UJRN/Imss_2017_9_15)
12. Stepanova M., Thomas L.C. Survival analysis methods for personal loan data. *Operations Research*. 2002. Vol. 50, No. 2. P. 277–289.
13. Havrylovych M., Kuznietsova N. Survival analysis methods for churn prevention in telecommunications industry. *CEUR Workshop Proceeding*. 2020. Vol. 2588. P. 47–58. URL: <http://ceur-ws.org/Vol-2577/paper5.pdf>
14. Cox D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*. 1972. Vol. 34, No. 2. P. 187–220.

Надійшла до редакції 26.02.2021