

ВИБІР ЕКОНОМЕТРИЧНОГО ІНСТРУМЕНТАРІЮ ПРИ ПОБУДОВІ СКОРИНГОВОЇ МОДЕЛІ НА ОСНОВІ ФІКТИВНИХ ЗМІННИХ

©2023 САВИНА С. С., ВОДЗЯНОВА Н. К., БІЛИК Т. О., КРАВЧЕНКО В. Л., СЕМАШКО К. А.

УДК 519.866
JEL: C19; G21

Савина С. С., Водзянова Н. К., Білик Т. О., Кравченко В. Л., Семашко К. А. Вибір економетричного інструментарію при побудові скорингової моделі на основі фіктивних змінних

Метою дослідження є вибір адекватного економетричного інструментарію для побудови скорингової моделі на специфічному масиві початкових даних, який містить переважно більшість фіктивних змінних. Незважаючи на значну кількість розробок, які присвячені побудові скорингових моделей, універсального методу, який дає змогу отримати високоефективний класифікатор для будь-яких даних, не виявлено. Тому актуальним залишається завдання вибору кращого методу для побудови скорингової моделі залежно від особливостей наявних даних. Найбільш вдалим підходом при виборі моделі для розв'язання задачі бінарної класифікації є застосування кількох видів економетричних моделей та вибір кращої з них за результатами класифікації. У наведеному дослідженні застосовано такі види моделей: дискримінантна модель, логіт- і пробіт-регресії та поліноміальна логістична регресія. Застосовувалися навчальні вибірки з різною структурою. Порівняння всіх отриманих моделей дає змогу зробити висновок, що кращою в даному випадку є поліноміальна логістична регресія. Ця модель демонструє високі показники класифікації для всіх введених класів об'єктів та має важливу перевагу порівняно з моделями, які здійснюють бінарний вибір. Перевагою поліноміальної логістичної регресії є також можливість вибору в кожному конкретному випадку зручної шкали для поділу позичальників більш ніж на два класи та визначення прийняттого для власних умов рівня ймовірності надійності позичальника, при якому слід відносити його до одного з виділених класів. Перспективами подальших досліджень у даному напрямку є застосування методів машинного навчання, які зможуть використовувати ансамблі кращих із розглянутих моделей. Крім того, запропоновані моделі можна використовувати і при розв'язанні аналогічних задач у інших сферах економічної діяльності.

Ключові слова: скорингова модель, логістична регресія, поліноміальна логістична регресія, бінарна класифікація.

Табл.: 5. **Формул.:** 1. **Бібл.:** 10.

Савина Світлана Станіславівна – кандидат економічних наук, доцент, доцент кафедри математичного моделювання та статистики, Київський національний економічний університет ім. В. Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: s.savina@kneu.edu.ua

ORCID: <https://orcid.org/0000-0003-0227-7081>

Researcher ID: <https://www.webofscience.com/wos/author/record/B-3367-2019>

Водзянова Наталія Костянтинівна – старший викладач кафедри математичного моделювання та статистики, Київський національний економічний університет ім. В. Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: vodzyanova.natalia@kneu.edu.ua

Білик Тетяна Олександрівна – кандидат економічних наук, доцент, доцент кафедри математичного моделювання та статистики, Київський національний економічний університет ім. В. Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: bilyk_t@kneu.edu.ua

ORCID: <https://orcid.org/0000-0002-4059-8794>

Researcher ID: <https://www.webofscience.com/wos/author/record/X-6915-2018>

Кравченко Вікторія Леонідівна – старший викладач кафедри математичного моделювання та статистики, Київський національний економічний університет ім. В. Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: kravchenko.viktoria@kneu.edu.ua

Семашко Катерина Анатоліївна – кандидат економічних наук, старший викладач кафедри математичного моделювання та статистики, Київський національний економічний університет ім. В. Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: kateryna.semashko@kneu.edu.ua

UDC 519.866
JEL: C19; G21

Savina S. S., Vodzyanova N. K., Bilyk T. O., Kravchenko V. L., Semashko K. A. Selection of Econometric Instruments when Building a Scoring Model Based on Dummy Variables

The aim of the study is to select adequate econometric instruments for building a scoring model on a specific array of initial data, which contains the vast majority of fictitious variables. Despite a significant number of developments devoted to the construction of scoring models, a universal method allowing to obtain a highly efficient classifier for any data has not been identified. Therefore, the task of selection of the best method for building a scoring model remains relevant, depending on the characteristics of the available data. The most successful approach when selecting a model for solving the problem of binary classification is the use of several types of econometric models and the choice of the best of them according to the results of classification. In the presented study, the following types of models were applied: discriminant model, logit and probit regressions, and polynomial logistic regression. Training samples with different structure were used. Comparison of all obtained models allows us to conclude that polynomial logistic regression is preferable in this case. This model demonstrates high classification rates for all introduced object classes and has an important advantage compared to models that make a binary selection. The advantage of polynomial logistic regression is also the possibility of selecting in each case a convenient scale for dividing borrowers into more than two classes and determining the level of probability of reliability of the borrower acceptable for its own conditions, at which it should be assigned to one of the selected classes. Prospects for further research in this direction are the use of machine learning methods that will be able to use ensembles of the best of the considered models. In addition, the proposed models can be used in solving similar problems in other spheres of economic activity.

Keywords: scoring model, logistic regression, polynomial logistic regression, binary classification.

Tabl.: 5. **Formulae:** 1. **Bibl.:** 10.

Savina Svitlana S. – PhD (Economics), Associate Professor, Associate Professor of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after V. Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: s.savina@kneu.edu.ua

ORCID: <https://orcid.org/0000-0003-0227-7081>

Researcher ID: <https://www.webofscience.com/wos/author/record/B-3367-2019>

Vodzyanova Natalia K. – Senior Lecturer of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after V. Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: vodzyanova.natalia@kneu.edu.ua

Bilyk Tetiana O. – PhD (Economics), Associate Professor, Associate Professor of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after V. Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: bilyk_t@kneu.edu.ua

ORCID: <https://orcid.org/0000-0002-4059-8794>

Researcher ID: <https://www.webofscience.com/wos/author/record/X-6915-2018>

Kravchenko Viktoriia L. – Senior Lecturer of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after V. Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: kravchenko.viktoria@kneu.edu.ua

Семашко Катерина Анатоліївна – PhD (Economics), Senior Lecturer of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after V. Hetman (54/1 Beresteiskyi Ave., Kyiv, 03057, Ukraine)

E-mail: katelyna.semashko@kneu.edu.ua

Кредитний скоринг – це система, яка розробляється для проведення оцінювання потенційного позичальника банку з метою встановлення рівня його надійності. Оцінка проводиться на основі анкетних даних клієнта та даних його кредитної історії, якщо вона існує. Зібрана інформація переводиться в бали, за якими встановлюється рівень надійності позичальника. Початковим підходом у кредитному скорингу було використання скорингових карт. У такій карті по кожному пункту інформації про позичальника наведено відповідну шкалу з балами. Сума всіх балів також за шкалою переводилася в оцінку надійності позичальника [1].

Сучасні підходи кредитного скорингу ґрунтуються на застосуванні математичних моделей. Задача розподілу позичальників на надійних і ненадійних належить до класичних задач математичного моделювання та відома як задача бінарної класифікації. На сьогодні розроблено значну кількість методів та моделей, що дозволяють розв'язувати задачу бінарної класифікації, проте актуальність розробок, присвячених пошуку методів для побудови адекватних скорингових моделей, продовжує зростати. Це обумовлено тим, що постійним є попит на кредитні ресурси, а отже, і необхідність точної оцінки надійності потенційних позичальників банку. Крім того, схожі за постановкою задачі виникають в інших сферах економічної діяльності. Наприклад, модель бінарного вибору дозволяє визначити, чи належить певне підприємство до класу банкрутів чи стабільно працюючих. У маркетингу задача бінарної класифікації може бути застосована для проведення лід-скорингу (*lead scoring*) – визначення готовності потенційного клієнта укласти угоду з надання послуг (чи придбати товар) [2].

Розробці підходів розв'язання задачі бінарної класифікації присвячена значна кількість літератури, аналіз якої дозволяє виділити кілька найбільш розповсюджених моделей. Найбільш поширеним є використання логістичної регресії. Її популярність обумовлена простотою реалізації, а також тим, що розрахункові значення моделі знаходяться в інтервалі $[0, 1]$ і можуть бути інтерпретовані як імовірність дефолту позичальника. Однією з найбільш ґрунтовних і глибоких робіт є праця N. Siddiqi [3], яка детально та поетапно розглядає процес побудови скорингової моделі. Більш складним й одночасно більш ефективним для багатьох випадків є використання дерев рішень [4]. Також значна кількість робіт присвячена побудові скорингових моделей на основі нейромереж, оскільки такі моделі демонструють високий рівень якості класифікації [5; 6].

Проте, незважаючи на значну кількість наробок, універсального методу, який дає змогу отримати високоефективний класифікатор для будь-яких даних, не виявлено. Так, наприклад, у роботі [5] розрахунки здійснювалися на основі штучно згенерованої бази даних. Зрозуміло, що результати застосування аналогічних моделей на реальних даних можуть суттєво відрізнятись. Тому актуальним залишається завдання вибору кращого методу для побудови скорингової моделі залежно від особливостей наявних даних. Якість моделі безпосередньо залежить від якості початкових даних, тому в кожному конкретному випадку, залежно від специфіки масиву початкових даних, слід розглядати кілька видів моделей з метою пошуку кращого.

Слід виділити також ще один не достатньо розроблений аспект у побудові скорингових моделей. Як

правило, розглядається варіант розподілу досліджуваних об'єктів лише на два класи: надійні та ненадійні позичальники. Тоді як більш доцільним є варіант, який використовувався у скорингових картах, а саме – наявність деякої шкали. Таким чином, можна отримати більш ґрунтовну інформацію про позичальників. Наприклад, виділити таких, які є абсолютно надійними та абсолютно ненадійними, але разом із тим виділити також групу позичальників, яких можна вважати надійними з певними застереженнями.

Метою даного дослідження є порівняння результатів застосування різноманітного економетричного інструментарію на специфічному масиві початкових даних для вибору найбільш ефективної моделі. Дана робота є узагальненням раніше проведених авторами досліджень [7–9].

Для побудови скорингової моделі було використано банківську інформацію стосовно клієнтів. Масив включає дані в такому співвідношенні: підвибірка із 300 значень – інформація про ненадійних позичальників та підвибірка із 700 значень – інформація про надійних позичальників.

Інформація про позичальників представлена за такими блоками. Блок інформації стосовно особи позичальника: вік, стать, рівень кваліфікації позичальника, тривалість зайнятості на момент кредитування. Блок, що стосується питань поточного кредиту: строк кредиту та цільове призначення. Інформація, що свідчить про рівень надійності позичальника: кредитна історія, наявність депозитів, наявність коштів на поточному картковому рахунку, наявність нерухомого майна.

Таким чином, отримана інформація містить лише три кількісні показники, всі решта десять є якісними. Для перетворення якісних змінних у кількісні було використано класичний підхід – застосування фіктивних змінних. Наприклад, такий показник, як цільове призначення кредиту має чотири рівні: автокредитування, соціальне кредитування, придбання побутової техніки та інші напрямки кредитування. Для кожного рівня вводиться фіктивна змінна. Сума всіх чотирьох змінних за одним спостереженням дорівнює одиниці, отримані змінні утворюють лінійну комбінацію, тому одночасно в модель можна вводити лише три фіктивні змінні четверта – є базовою [9, с. 130]. Аналогічним чином виконано перетворення всіх інших якісних змінних. У результаті отримано 19 фіктивних змінних.

Дослідження проводились поетапно. На першому етапі використовувався масив даних, який утворено за інформацією про надійних і ненадійних позичальників у рівних пропорціях [7]. Логістична регресія демонструє найвищий рівень якості класифікації в тих випадках, коли навчальна вибірка для моделі складається з однакової кількості об'єктів двох класів. Тому для навчання моделі було використано 300

значень, які стосувалися даних про ненадійних позичальників, та випадковим чином з інформації стосовно надійних позичальників було обрано також 300 значень. Для утвореного масиву застосовувалися два типи моделей: дискримінантна модель і логістична регресія. Реалізація моделей здійснювалась засобами ППП Statistica. За результатами моделювання за допомогою логістичної регресії виявлено, що з 23 параметрів моделі достовірними є 5. При цьому модель достовірна з надійністю понад 99%. Показник якості дискримінантної моделі Wilks'Lambda = 0,86 наближається до одиниці, що свідчить про досить розмиття межі між альтернативними групами та невисоку якість класифікації.

Ефективність застосування моделей можна оцінити за допомогою матриці класифікації. У матриці наведено у відсотках: TP (*true-positive*) – правильно класифіковані надійні позичальники; FP (*false-positive*) – неправильно класифіковані надійні позичальники; TN (*true-negative*) – правильно класифіковані ненадійні позичальники; FN (*false-negative*) – неправильно класифіковані ненадійні позичальники. Вказані показники для двох моделей наведено в табл. 1.

Таблиця 1

Характеристики якості класифікації моделей

Дискримінантна модель, %		Логістична регресія, %	
TP = 0,68	FP = 0,32	TP = 0,68	FP = 0,32
FN = 0,35	TN = 0,65	FN = 0,34	TN = 0,66

Джерело: авторська розробка.

Аналіз даних табл. 1 дозволяє зробити висновок, що якість класифікації для двох моделей можна вважати невисокою. Майже 70% правильно класифікованих надійних позичальників (однакова для двох моделей) та трохи нижчий показник для правильно класифікованих ненадійних клієнтів, причому незначно кращими є результати логістичної регресії. З метою підвищення якості класифікації на наступному етапі дослідження для побудови моделі було використано всі наявні дані, тобто навчальна вибірка містить повну підвибірку із 700 значень відносно надійних позичальників та підвибірку з 300 значень за ненадійними позичальниками [8]. Оскільки дискримінантна модель продемонструвала гірші результати, то для подальших досліджень було обрано логістичну регресію. Реалізація моделей здійснювалась у ППП EViews. У результаті побудови моделей, на відміну від попереднього випадку, достовірними виявились 9 параметрів моделей, тому для можливості вибору кращої моделі проведено побудову моделей, які включають лише ті чинники, параметри за якими виявились достовірними. Порівняння трьох отриманих моделей здійснювалось на основі розрахунку показників McFadden та LR statistic, а також трьох ін-

формаційних критеріїв: Akaike (AIC), Schwarz (BIC), Hannan-Quinn (HQIC). Результати за всіма показниками для трьох моделей наведено в *табл. 2* [8].

Таблиця 2

Основні характеристики моделей

Показник	Логістична регресія для повного переліку факторів	Логістична регресія для 9 факторів	Пробіт регресія для 9 факторів
McFadden R-squared	0,101875	0,095	0,095
LR statistic	124,4639	116,138	115,676
Akaike info criterion	1,143265	1,126	1,126
Schwarz criterion	1,256143	1,175	1,175
Hannan-Quinn criterion	1,186166	1,144	1,145

Джерело: авторська розробка.

Аналіз *табл. 2* дозволяє обрати одну з розглянутих моделей. За всіма трьома інформаційними критеріями кращі значення має модель з меншою кількістю чинників. Логістична та пробіт-регресія незначно відрізняються лише результатом LR statistic, тому для подальших розрахунків було обрано логістичну регресію, яка містить 9 чинників. У *табл. 3* наведено параметри зазначеної моделі та їх основні характеристики.

Таблиця 3

Коефіцієнти логістичної регресії та їх основні характеристики

Змінна	Коефіцієнт	Значення стандартного відхилення	p-value
C	-0,268	0,408	0,510
X ₃	-0,035	0,006	0,000
D ₃₂	0,427	0,157	0,007
D ₄₂	0,827	0,283	0,004
D ₆₁	0,889	0,355	0,012
D ₆₂	1,161	0,416	0,005
D ₆₃	1,726	0,379	0,000
D ₇₂	0,382	0,188	0,042
D ₇₃	0,885	0,240	0,000

Джерело: авторська розробка.

Значення p-value майже для всіх оцінок параметрів (крім константи) суттєво менше 0,05, отже, всі параметри моделі є достовірними з надійністю понад 95%. Таким чином, отримано модель вигляду:

$$p = \frac{1}{1 + e^{-(0,268 - 0,035X_3 + 0,427D_{32} + \dots + 0,885D_{73})}} \quad (1)$$

Результати класифікації за даною моделлю наведено в *табл. 4*.

Таблиця 4

Матриця класифікації за логістичною регресією на основі 9 чинників

Розраховані за моделлю результати та фактичні значення	Відсоток позичальників, віднесених за моделлю до класу надійних	Відсоток позичальників, віднесених за моделлю до класу ненадійних
Надійні позичальники	0,94	0,06
Ненадійні позичальник	0,23	0,77

Джерело: авторська розробка.

Порівняння результатів логістичних регресій, які наведено в *табл. 1* і *табл. 4*, дає можливість однозначно встановити кращу модель. Логіт-регресія, яку отримано на основі достовірних параметрів, демонструє значно вищі результати правильної класифікації за двома класами об'єктів: 94% надійних позичальників та 77% ненадійних.

Розподіл об'єктів на класи при застосуванні логістичної регресії відбувається за допомогою порівняння розрахованих за моделлю значень з деяким пороговим значенням. Як правило, за порогове приймається значення 0,5. Якщо розрахункове значення перевищує поріг, то об'єкт відносять до класу, який позначено 1 (у нашому випадку надійний позичальник), якщо значення менше 0,5 – об'єкт відносять до класу, який позначено 0 (у проведеному дослідженні – ненадійний клієнт). Однак встановлення такого жорсткого поділу може бути невиправданим. Розраховані за допомогою логістичної регресії значення завжди знаходяться в інтервалі від нуля до одиниці й тому інтерпретуються як імовірність того, що об'єкт належить до того класу, який було позначено 1. Отже, в нашому дослідженні при отриманні розрахункового значення, наприклад, 0,49 позичальника банку буде автоматично віднесено до ненадійних, хоча відхилення від порогового значення є незначним. На нашу думку, кращим варіантом буде розподіл позичальників за шкалою, яка дозволить для умов кожного конкретного випадку розділити потенційних клієнтів не на дві, а на більшу кількість груп, наприклад надійних, ненадійних і надійних при дотриманні додаткових умов. Такими умовами можуть бути вимоги, які підкріплюють інформацію стосовно рівня платоспроможності клієнтів. Наприклад, можна додатково вимагати наявності поручителя або певної суми коштів на картковому рахунку тощо. Крім того, залежно від специфіки банківської установи, доцільно передбачити умову гнучкого механізму побудови шкали розподілу позичальників.

Для розв'язання такої задачі на наступному етапі нашого дослідження було застосовано поліноміальну (*multinomial*) логістичну регресію [10]. Таку модель можна вважати узагальненим варіантом моделі бінарної класифікації. Як зазначалося, при розв'язуванні задачі бінарного вибору всі досліджувані об'єкти розподіляються лише на два класи (в нашому випадку надійні та ненадійні позичальники). У такому випадку модель представлена лише одним рівнянням. Якщо необхідно здійснити розподіл об'єктів на *m* класів, то модель буде містити *m* – 1 рівнянь.

У нашому дослідженні розглядається спрощений варіант розподілу позичальників на три класи [9]. За результатами застосування логістичної регресії (1) всі об'єкти розподілено на три класи за такою схемою:

- ✦ 1 клас – ненадійні позичальники, для яких значення за моделлю (1) менше 0,5;
- ✦ 2 клас – умовно надійні позичальники, для яких значення за моделлю (1) знаходиться в межах від 0,5 до 0,7;
- ✦ 3 клас – абсолютно надійні позичальники, для яких значення за моделлю (1) перевищує 0,7.

На основі трьох виділених класів засобами ППП EViews побудовано поліноміальну логістичну регресію. Результати класифікації за моделлю наведено в *табл. 5*.

а класична логістична регресія – лише один. Однак очевидно, що відсоток правильно визначених позичальників для двох класів (надійних та умовно надійних) є доволі високим і складає 90,6 та 81 відповідно.

ВИСНОВКИ

Побудова скорингової моделі здійснювалася на основі специфічного масиву початкових даних, який складається переважно з фіктивних змінних. У процесі пошуку адекватного економетричного інструментарію для побудови якісної скорингової моделі було застосовано такі види моделей: дискримінантна модель, логіт- і пробіт-регресії та поліноміальна логістична регресія. Застосовувалися навчальні вибірки з різною структурою. Порівняння всіх отриманих моделей дає змогу зробити висновок, що кращою в даному випадку є поліноміальна логістична регресія. Ця модель демонструє високі показники класифікації для всіх уведених класів об'єктів та має важливу перевагу порівняно з моделями, які здійснюють бінарний вибір. Використовуючи у практичній діяльності для розробки скорингової моделі поліноміальну логістичну регресію, конкретна банківська установа має можливість самостійно обирати шкалу, за якою позичальники будуть розподілені на класи, а також обирати прийнятний для власних умов рівень імовірності надійності позичальника, за якого слід відносити його до одного з виділених класів.

Таблиця 5

Результати класифікації на основі поліноміальної логістичної регресії

Номер класу	Кількість об'єктів у класі	Кількість правильно класифікованих об'єктів	Кількість неправильно класифікованих об'єктів	Правильно класифіковані об'єкти, %	Неправильно класифіковані об'єкти, %
1	107	84	23	78,505	21,495
2	363	295	68	81,267	18,733
3	530	480	50	90,566	9,434
Усього	1000	859	141	85,900	14,100

Джерело: авторська розробка.

Дані табл. 5 деталізують якість проведення класифікації поліноміальною логістичною регресією. Наведено загальну кількість об'єктів у кожній групі та кількість об'єктів, які були правильно та неправильно класифіковані, а також відповідні значення у відсотках.

Відсоток правильно класифікованих позичальників 1 класу (ненадійні позичальники) є більшим порівняно з тими моделями, що були розглянуті раніше. Слід зазначити, що при застосуванні скорингових моделей досить важливою характеристикою є саме правильне визначення ненадійних позичальників. Неможливо коректно порівняти рівень класифікації надійних позичальників, оскільки поліноміальна модель містить два класи такого виду позичальників,

Як зазначалося вище, задачі зі схожою постановкою виникають не лише в процесі банківської діяльності, тому запропонований вид моделі можна буде застосовувати і при розв'язанні аналогічних задач у інших сферах економічної діяльності. ■

БІБЛІОГРАФІЯ

- Кузнецова Н. В. Розробка скорингових карт для аналізу ризиків банківської діяльності. *Реєстрація, зберігання і обробка даних*. 2017. Т. 19. № 4. С. 35–44. DOI: <https://doi.org/10.35681/1560-9189.2017.19.4.142920>
- Sydorenko N. A Beginner's Guide To B2B Lead Scoring. URL: <https://snov.io/blog/guide-to-lead-scoring/>

3. Siddiqi N. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, 2005. 208 p.
4. Bastos J. A. Predicting Credit Scores with Boosted Decision Trees. *Forecasting*. 2022. Vol. 4. Iss. 4. P. 925–935. DOI: <https://doi.org/10.3390/forecast4040050>
5. Васильєв О. М. Скорингове моделювання на основі нейронних мереж для визначення рейтингу позичальника банку. *Економіка України*. 2020. № 10. С. 54–62. DOI: <https://doi.org/10.15407/economyukr.2020.10.054>
6. Garcin M., Stéphan S. Credit scoring using neural networks and SURE posterior probability calibration. 2021. URL: https://hal.science/hal-03286760v1/file/Article_VF.pdf
7. Савіна С. С., Водзянова Н. К. Побудова скорингової моделі для масиву даних на основі фіктивних змінних // Technologies and strategies for the implementation of scientific achievements : collection of scientific papers «SCIENTIA» with Proceedings of the III International Scientific and Theoretical Conference. April 28, 2023. Stockholm, Kingdom of Sweden: European Scientific Platform. P. 58–59. DOI: <https://doi.org/10.36074/scientia-28.04.2023>
8. Савіна С. С., Водзянова Н. К. Побудова моделі оцінки ймовірності дефолту позичальників на основі якісних показників // Розвиток наукової думки постіндустріального суспільства: сучасний дискурс : матеріали III Міжнародної наукової конференції (м. Львів, 28 квітня 2023 р.). Вінниця : Європейська наукова платформа, 2023. С. 103–105. URL: <https://archive.mcmd.org.ua/index.php/conference-proceeding/issue/view/28.04.2023/25>
9. Савіна С. С., Водзянова Н. К. Застосування поліноміальної логістичної регресії при побудові скорингових моделей // Актуальні питання розвитку галузей науки : матеріали I Міжнародної наукової конференції (м. Чернігів, 12 травня 2023 р.). Вінниця : Європейська наукова платформа, 2023. С. 130–132. DOI: [10.36074/mcmd-12.05.2023](https://doi.org/10.36074/mcmd-12.05.2023)
10. Brooks Ch. *Introductory Econometrics for Finance*. 3rd ed. Cambridge University Press, 2014. 740 p.

REFERENCES

- Bastos, J. A. "Predicting Credit Scores with Boosted Decision Trees". *Forecasting*, vol. 4, no. 4 (2022): 925-935. DOI: <https://doi.org/10.3390/forecast4040050>

- Brooks, Ch. *Introductory Econometrics for Finance*. Cambridge University Press, 2014.
- Garcin, M., and Stephan, S. "Credit scoring using neural networks and SURE posterior probability calibration". 2021. https://hal.science/hal-03286760v1/file/Article_VF.pdf
- Kuznietsova, N. V. "Rozrobka skorynhovykh kart dlia analizu ryzykiv bankivskoi diialnosti" [Development of Scoring Cards for Analyzing the Risks of Banking Activity]. *Reiestratsiia, zberihannia i obrobka danykh*, vol. 19, no. 4 (2017): 35-44. DOI: <https://doi.org/10.35681/1560-9189.2017.19.4.142920>
- Savina, S. S., and Vodzianova, N. K. "Pobudova modeli otsinky imovirnosti defoltu pozychalnykiv na osnovi yakisnykh pokaznykiv" [Building a Model for Estimating the Probability of Borrower Default Based on Qualitative Indicators]. *Rozvytok naukovoï dumky postindustrialnoho suspilstva: suchasnyi dyskurs*. 2023. 103-105. <https://archive.mcmd.org.ua/index.php/conference-proceeding/issue/view/28.04.2023/25>
- Savina, S. S., and Vodzianova, N. K. "Pobudova skorynhovoi modeli dlia masyvu danykh na osnovi fiktyvnykh zminnykh" [Building a Scoring Model for a Data Array Based on Dummy Variables]. *Technologies and strategies for the implementation of scientific achievements : collection of scientific papers «SCIENTIA» with Proceedings of the III International Scientific and Theoretical Conference*. April 28, 2023. Stockholm, Kingdom of Sweden: European Scientific Platform. 58-59. DOI: <https://doi.org/10.36074/scientia-28.04.2023>
- Savina, S. S., and Vodzianova, N. K. "Zastosuvannia polinomialnoi lohistrychnoi rehresii pry pobudovi skorynhovykh modelei" [Application of Polynomial Logistic Regression in Building Scoring Models]. *Aktualni pytannia rozvytku haluzei nauky*. Vinnytsia: Yevropeiska naukova platforma, 2023. 130-132. DOI: [10.36074/mcmd-12.05.2023](https://doi.org/10.36074/mcmd-12.05.2023)
- Siddiqi, N. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, 2005.
- Sydorenko, N. "A Beginner's Guide To B2B Lead Scoring". <https://snov.io/blog/guide-to-lead-scoring/>
- Vasyliiev, O. M. "Skorynhove modeliuвання na osnovi neuronnykh merezh dlia vyznachennia reitynhu pozychalnyka banku" [Scoring Modeling Based on Neural Networks for Determining a Bank Borrower's Rating]. *Ekonomika Ukrainy*, no. 10 (2020): 54-62. DOI: <https://doi.org/10.15407/economyukr.2020.10.054>