

ПРОГНОЗУВАННЯ ОПЕРАЦІЙНИХ РИЗИКІВ ПІДПРИЄМСТВА (ЕКОНОМЕТРИЧНИЙ АНАЛІЗ)

Д'яченко А.С.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹aniyamaniya1@gmail.com [0000-0003-0036-4405],

²pbidyuke_00@ukr.net [0000-0002-7421-3565]

Для визначення прогнозного рівня втрат від операційних ризиків (ОР) за принципом економетрики у роботі використані регресори, які пов'язані з типами ОР підприємства. Прогнозування реалізовано в пакеті Eviews. За результатами досліджень ризику «втрати від збоїв у постачанні обладнання» з використанням прийомів регресійного аналізу та статистики Дарбіна-Уотсона, створені адекватні авторегресійні прогнозні моделі 1-5 порядку. Висока точність та якість прогнозу у побудованих моделях визначені високим коефіцієнтом детермінації та низькою середньою абсолютною похибкою.

Ключові слова: операційний ризик підприємства, EVIEWS, тренд, регресійний аналіз, коефіцієнт Тейла, авторегресія.

1. ВСТУП

Фінансові кризи примушують сектори виробничої діяльності переглядати підходи, створювати нові способи збільшення рівня прибутків, впроваджувати сучасні програмні засоби прогнозування. Особливо гостро це питання стосується виробничих, видобувних та інших підприємств. Прикро, але на виробничих підприємствах поняття ОР недостатньо поширено. Такий підхід треба змінювати, бо ОР може привести до найбільших збитків. Для забезпечення ефективного реагування на кризові ситуації компаніям необхідно визначитися з управлінням, аналізом та прогнозом ОР.

2. ПОСТАНОВКА ПРОБЛЕМИ ТА МЕТОДИ ДОСЛІДЖЕНЬ

Мета дослідження полягає в аналізі та прогнозуванні операційних ризиків виробничого сектору економіки та практичному застосуванні інформаційних технологій – спеціалізованого економетричного пакету EVIEWS для вирішення задач математичного моделювання та прогнозування операційного ризику за допомогою кореляційного аналізу вихідних тимчасових даних. Виявлення стохастичного зв'язку та оцінка його сили – важливе завдання економетрики.

Методи дослідження стосовно розв'язання задач математичного моделювання та прогнозування операційного ризику: – а) казуальні (пошук зв'язку між певними критеріями і виявленою проблемою, застосування відповідної бази статистичних критеріїв для забезпечення високої якості даних і результатів); б) прогнозні статистичні (побудова адекватної моделі та генерування високоякісних прогнозів, передбачення майбутньої ситуації).

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

У поданому дослідженні аналіз ризиків проводився з урахуванням управлінської звітності за 2012 - 2021 рр. підприємства, що видобуває калійні солі. Проведений нами аналіз рівнів пріоритету ОР у формі рейтингування, показав, що основна причина настання ОР - погрози, пов'язані з втратами «від збоїв у постачанні систем та обладнання». Саме тому, цей тип ОР прийнятий для обчислювальних експериментів з побудови моделей прогнозування.

Обраний у дослідженні статистичний підхід ґрунтується на історичних даних ризикових подій підприємства. Кількісна оцінка ОР проводиться на підставі статистичних даних які мають поквартальні характеристики: величини втрат від ОР, кількість (од) і вартість обладнання (грн) та кількості зареєстрованих випадків що спричинили ризик. Прогнозування виконане на кінець 2022 року. Оскільки саме пакет Eviews [1] дозволяє працювати з часовими типами даних – це найкращий вибір для вирішення поставленого завдання та досягнення мети дослідження.

Для коректного проведення регресійного аналізу досліджено наявність можливих лінійних зв'язків між змінними, які в подальшому пояснюватимуть поведінку змінної (табл. 1), та визначити взаємозв'язок ризику з іншими змінними та змінних між собою.

Таблиця 1. Кореляція між змінними

	AMOUNT_O	AVERAGE_	COEFFICIENT	NUMBER_O	NUMBER_O	THE_NUMB
AMOUNT_O	1.000000	0.923658	0.161305	0.603975	0.970459	0.753462
AVERAGE_	0.923658	1.000000	0.197908	0.602374	0.970747	0.722572
COEFFICIENT	0.161305	0.197908	1.000000	0.121026	0.153841	0.090565
NUMBER_O	0.603975	0.602374	0.121026	1.000000	0.643044	0.396771
NUMBER_O	0.970459	0.970747	0.153841	0.643044	1.000000	0.764347
THE_NUMB	0.753462	0.722572	0.090565	0.396771	0.764347	1.000000

Визначено, що згідно з кореляційними показниками найбільше на ризик впливають середня вартість та кількість випадків, які спричиняють ризик. Це незалежні змінні, на основі яких зроблено прогноз для ОР на 2022 рік.

Один з етапів створення прогнозу моделі – розрахунок описових статистик (ОС) вибраних змінних (табл. 2), які для зручності ми перейменуємо так: кількість втрат від ризику – Y, середня вартість – X1, а кількість випадків, які зумовлюють ризик – X2. ОС необхідні для того, щоб більш детально чисельно охарактеризувати змінні, що вивчаються.

Таблиця 2. Описова статистика по усіх змінних

	AMOUNT_O	AVERAGE_	COEFFICIENT	NUMBER_O	NUMBER_O	THE_NUMB
Date: 10/26/22 Time: 21:52						
Sample: 2012Q1 2021Q4						
Mean	2160.953	7225.876	0.023800	13.97500	111.5750	4.125000
Median	2000.800	7582.000	0.020000	14.00000	110.5000	4.000000
Maximum	4000.900	10138.00	0.047000	19.00000	153.0000	8.000000
Minimum	732.1200	4589.000	-0.018000	8.000000	79.00000	2.000000
Std. Dev.	862.5884	1723.808	0.014879	2.759482	21.72661	1.712511
Skewness	0.551121	-0.290245	-1.268780	-0.277718	0.256783	0.548296
Kurtosis	2.548818	1.882380	4.450743	2.393544	1.907849	2.505043
Jarque-Bera	2.364172	2.348111	14.23878	1.127164	2.428301	2.412052
Probability	0.308538	0.308956	0.000809	0.582167	0.298802	0.299385
Sum	86678.10	289035.0	0.062000	559.0000	4463.000	165.0000
Sum Sq. Dev.	29018290	-1.18E+08	0.006634	296.9750	18409.78	114.3750
Observations	40	40	40	40	40	40

Примітка: Mean, Median, Maximum, Minimum – середнє, медіанне, максимальне та мінімальне значення відповідно; Std. Dev. – стандартне відхилення (корінь із дисперсії); Skewness – асиметрія (0 у нормального розподілу); Kurtosis – ексцес (3 у нормального розподілу); Jarque-Bera - статистика Жака-Бера (для перевірки гіпотези про те, що вибірка

із нормального розподілу; *Probability* – ймовірність помилитися, відкинувши гіпотезу про те, що вибірка з нормального розподілу (у даному випадку ймовірність нульова, тому ми не помилимося, якщо скажемо, що вибірка отримана не з нормального розподілу); *Sum* – сума значень усіх спостережень, *Sum Sq. Dev.* – сума квадратів відхилень рівнів ряду від середнього; *Observations* – кількість спостережень.

Оскільки у лінійній моделі незалежних змінних дві, модель множинної регресії з коефіцієнтами розраховуємо методом найменших квадратів (Least Squares) (табл. 3).

Таблиця 3. Модель множинної регресії

Dependent Variable: Y Method: Least Squares Date: 10/26/22 Time: 22:04 Sample: 2012Q1 2021Q4 Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.397134	0.043135	9.206663	0.0000
X2	90.70055	43.41496	2.089154	0.0436
C	-1076.827	226.1230	-4.762131	0.0000
R-squared	0.868639	Mean dependent var	2166.953	
Adjusted R-squared	0.861538	S.D. dependent var	862.5884	
S.E. of regression	320.9731	Akaike info criterion	14.45263	
Sum squared resid	3811879.	Schwarz criterion	14.57930	
Log likelihood	-286.0526	Hannan-Quinn criter.	14.49843	
F-statistic	122.3330	Durbin-Watson stat	0.393909	
Prob(F-statistic)	0.000000			

Оскільки ймовірність дуже мала, використовується альтернативна гіпотеза, яка говорить про нерівність коефіцієнтів нулю і їх статистичну значущість. Коефіцієнт детермінації (R-squared) на 87 % пояснює зміну Y в залежності від незалежних змінних. Таке значення є цілком прийнятним для побудованої моделі.

Примітка: *Std Error* – стандартна помилка; *t-статистика* (обраховується діленням коефіцієнту на стандартну помилку); *Prob* – вірогідність прийняття нульової гіпотези., *R-squared* – R-квадрат – це коефіцієнт детермінації. Метод оцінювання коефіцієнтів моделі – метод найменших квадратів.

Наступний крок прогнозування – визначення статистичних характеристик (СХ) адекватності моделі або однокрокове прогнозування. Для цього обчислюємо коефіцієнти множинної регресії для X1 та X2 – табл. 4 та отримуємо результати розрахунку СХ оцінок прогнозу і їх якості з оцінкою адекватності моделі (задовільний коефіцієнт детермінації).

Таблиця 4. Статистика та коефіцієнти множинної регресії

Dependent Variable: Y Method: Least Squares Date: 10/26/22 Time: 22:04 Sample: 2012Q1 2021Q4 Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.397134	0.043135	9.206663	0.0000
X2	90.70055	43.41496	2.089154	0.0436
C	-1076.827	226.1230	-4.762131	0.0000
R-squared	0.868639	Mean dependent var	2166.953	
Adjusted R-squared	0.861538	S.D. dependent var	862.5884	
S.E. of regression	320.9731	Akaike info criterion	14.45263	
Sum squared resid	3811879.	Schwarz criterion	14.57930	
Log likelihood	-286.0526	Hannan-Quinn criter.	14.49843	
F-statistic	122.3330	Durbin-Watson stat	0.393909	
Prob(F-statistic)	0.000000			

За статистикою Дарбіна-Уотсона (показник DW, яка стверджує, що похибки такої моделі є невисоко корельованими, приймаємо побудовану модель задовільною, але розуміємо, що модель можна покращити. Результат СХ оцінок прогнозу і їх якості в графічному вигляді (рис. 1, а) дозволяє визначити значення середньої абсолютної похибки (MAPE), середньоквадратичної похибки прогнозу (MSE) та коефіцієнту невідповідності Тейла (Theil) – рис. 1, б.

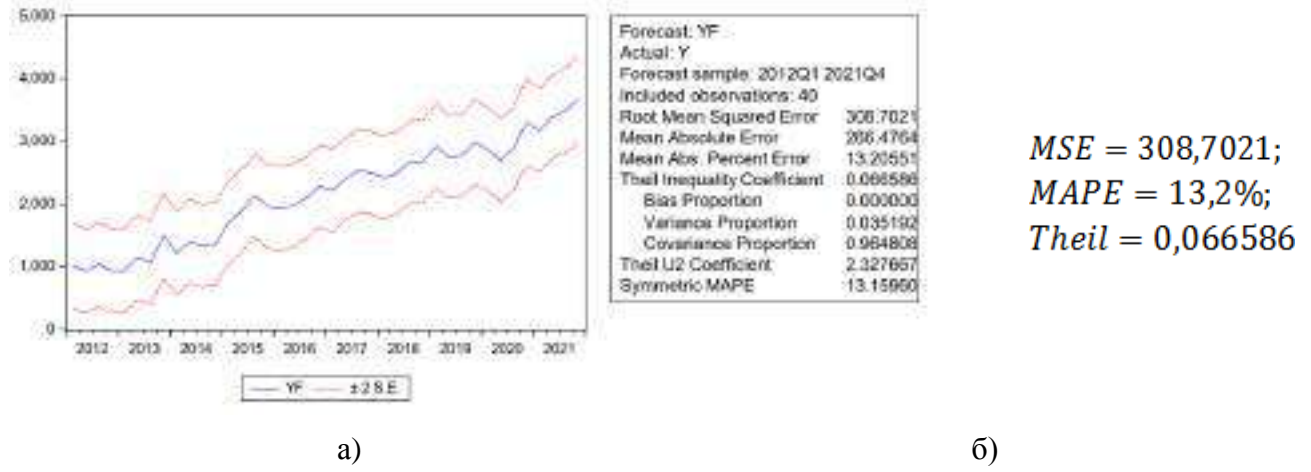


Рисунок 1. Результат СХ оцінок прогнозу і їх якості в графічному вигляді (а) та показники, що характеризують якість виконаного прогнозу (б)

Отриманий результат: прогнозні значення добре співвідносяться. Але значення «MAPE» свідчать про бажане покращення моделі, тому надалі обчислення робимо покрокове та підвищуємо порядок регресії. Проводимо тест на серію корельованих залишків Бройша-Годфрі, який дозволяє нам перевірити гіпотезу про те, що інші моделі описуються моделлю авторегресії. Це Serial Correlation LM Test - автокореляційна функція (АКФ) для змінної X1 у вигляді корелограми (рис. 2).

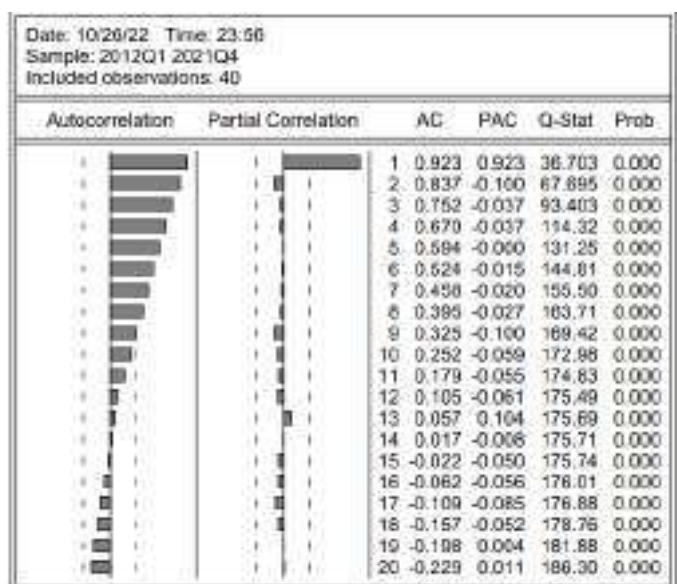


Рисунок 2. Корелограма значень автокореляційної та часткової АКФ

Аналіз отриманих даних дозволяє зробити висновок, що порядок авторегресії не перевищує 5. Тому аналізуємо авторегресію з покроковим підвищенням порядку - від першого до можливо максимального п'ятого. Результати розрахунку авторегресії першого та

п'ятого порядку подані в табл. 5. Спостерігається суттєве покращення оцінок прогнозів. Надалі застосовуємо апроксимацію моделі п'ятого порядку трендом (табл. 6). Рівняння побудованих автокореляційних моделей вказані в таблицях 5, 6. Отримана кінцева модель при порівнянні її з вхідними даними показала прийнятний результат: СХ прогнозу моделі з трендом – $R^2=0.99$, $DW=0.68797$, $\Sigma e^2=259668.8$; СХ прогнозу моделі без тренду – $R^2=0.8686$, $DW=0.3937$, $\Sigma e^2=3811879.1$. Дві статистичні характеристики моделі покращились, показник «МАРЕ» є цілком задовільним.

Таблиця 5. Результати розрахунку авторегресії першого та п'ятого порядків

Dependent Variable: X1 Method: Least Squares (Gauss-Newton / Marquardt steps) Date: 10/27/22 Time: 21:08 Sample (adjusted): 2012Q2 2021Q4 Included observations: 39 after adjustments X1=C(1)+C(2)*X1(-1)					Dependent Variable: X1 Method: Least Squares (Gauss-Newton / Marquardt steps) Date: 10/27/22 Time: 00:00 Sample (adjusted): 2016Q2 2021Q4 Included observations: 23 after adjustments X1=C(1)+C(2)*X1(-1)+C(3)*X1(-2)+C(4)*X1(-3)+C(5)*X1(-4)+C(6)*X1(-5)				
	Coefficient	Std. Error	t-Statistic	Prob.		Coefficient	Std. Error	t-Statistic	Prob.
C(1)	143.6036	147.7436	0.971977	0.3374	C(1)	737.6053	570.8157	1.292195	0.2136
C(2)	0.999815	0.020126	49.67690	0.0000	C(2)	1.439980	0.214029	6.728095	0.0000
R-squared	0.985226	Mean dependent var	7293.457		C(3)	-0.591378	0.267005	-2.214859	0.0407
Adjusted R-squared	0.984529	S.D. dependent var	1691.546		C(4)	-0.000758	0.051580	-0.008274	0.9935
S.E. of regression	208.3481	Akaike info criterion	13.56822		C(5)	0.056877	0.079673	0.713657	0.4850
Sum squared resid	1605130.	Schwarz criterion	13.65153		C(6)	0.034543	0.077091	0.448248	0.6596
Log likelihood	-262.5412	Hannan-Quinn criter.	13.59663		R-squared	0.979564	Mean dependent var	8476.870	
F-statistic	2467.794	Durbin-Watson stat	1.628976		Adjusted R-squared	0.973554	S.D. dependent var	846.9132	
Prob(F-statistic)	0.000000				S.E. of regression	139.0532	Akaike info criterion	12.91261	
					Sum squared resid	323997.7	Schwarz criterion	13.20883	
					Log likelihood	-142.4951	Hannan-Quinn criter.	12.98711	
					F-statistic	162.9743	Durbin-Watson stat	1.966286	
					Prob(F-statistic)	0.000000			
перший порядок					п'ятий порядок				

Таблиця 6. Результати оцінювання моделі для прогнозування Y з трендом

Dependent Variable: Y Method: Least Squares (Gauss-Newton / Marquardt steps) Date: 10/28/22 Time: 00:03 Sample: 2012Q1 2021Q4 Included observations: 40 Y = C(1)+C(2)*X1+C(2)*X2+C(3)*K+C(4)*K^2+C(5)*K^3				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1180.600	183.4514	6.435482	0.0000
C(2)	-0.158483	0.044878	-3.531432	0.0012
C(3)	191.7649	15.81718	12.12384	0.0000
C(4)	-7.452620	0.715458	-10.41685	0.0000
C(5)	0.138378	0.011377	12.16298	0.0000
R-squared	0.991052	Mean dependent var	2166.953	
Adjusted R-squared	0.990029	S.D. dependent var	862.5884	
S.E. of regression	86.13426	Akaike info criterion	11.86616	
Sum squared resid	259668.8	Schwarz criterion	12.07727	
Log likelihood	-232.3232	Hannan-Quinn criter.	11.94249	
F-statistic	969.0725	Durbin-Watson stat	0.687973	
Prob(F-statistic)	0.000000			

4. ВИСНОВКИ

Послідовність виконаних під час дослідження кроків по розрахунку прогнозних показників, дозволяє зробити деякі висновки, а саме: втрати від збоїв у постачанні систем та обладнання» на 4 квартали 2012 р. та прогнозних розрахунків 2022 р. показали, що показник втрат ОР збільшився майже в 4,68 разів, тобто зростання втрат складає практично 47 % на

рік. Випробувані методика оцінки та прогнозування має як переваги, так й недоліки. Залежно від цілей і наявності ПЗ, або при недостатньому статистичному полі треба комбінувати різні методики, а підсумкову оцінку операційних втрат ОР розраховувати як середньозважене.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Построение AR моделей для Индекса отребительских цен в пакете Eviews 5* (2022)
Студенческая библиотека онлайн: studbooks.net. Режим доступа:
https://studbooks.net/2273206/informatika/postroenie_modeley_indeksa_potrebitelskih_pakete_evie
[ws](https://studbooks.net/2273206/informatika/postroenie_modeley_indeksa_potrebitelskih_pakete_evie) (дата звернення 18.09.22).

АНАЛІЗ РИНКОВОГО РИЗИКУ ЙМОВІРНІСНО-СТАТИСТИЧНИМИ МЕТОДАМИ

Затірка В.В.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ lerazatirka@gmail.com,

² pbidyuke_00@ukr.net [0000-0002-7421-3565]

Робота будь-яких фінансових установ тісно пов'язана з виконанням різноманітних операцій з валютними ресурсами. У зв'язку з цим виникає необхідність високоякісного аналізу та управління ринковими ризиками. Багато широкоживаних моделей управління ризиками не підходять для використання українським фінансовим ринком, а отже необхідно розглянути та вдосконалити менш звичні методології. Метою цієї роботи є побудова адекватних моделей гетероскедастичних фінансових процесів для прогнозування дисперсії та оцінювання ринкових ризиків. Результатом дослідження є побудована модель, придатна для оцінювання ринкових ризиків. У роботі використано теоретичні та системні емпіричні методи дослідження.

Ключові слова: ринковий ризик, гетероскедастичний процес, дисперсія, VAR, метод Монте-Карло, аналіз ризиків.

1. ВСТУП

Будь-яким фінансовим рішенням завжди властива наявність ризику. Однією з основних проблем фінансових установ є оцінювання ринкових ризиків, які виникають в результаті випадкових змін цін, процентних ставок, курсів обміну валют та інших ринкових факторів. Більшість фінансово-економічних процесів економіки перехідного періоду мають нестационарний характер та містять нелінійності і через це виникає задача їх коректного моделювання, прогнозування та оцінювання супутніх ризиків.

Український фінансовий ринок в умовах перехідної економіки не є стабільним, таким, що динамічно розвивається і швидко реагує на зміни ринкових факторів. Отже, задача застосування і удосконалення методологій оцінки ризиків в Україні ставиться досить гостро.

2. МЕТОДИ ОЦІНЮВАННЯ РИНКОВОГО РИЗИКУ

Таксономія ризиків є важливим інструментом управління ризиками, оскільки вона полегшує процес ідентифікації ризиків і прискорює вибір того інструменту, як з ними боротися. Знаючи категорію ризику, можна чітко зрозуміти основний метод управління цим ризиком, оскільки для багатьох категорій ризику вже існують ефективні методи та принципи управління. НБУ запропонована класифікація ризиків, яка враховує їх залежність від доходів:

- Фінансові ризики – тобто ризики, що піддаються кількісній оцінці, а управління ними полягає в їх оптимізації.
- Нефінансові ризики – тобто ризики, що не піддаються кількісній оцінці, а управління ними зводиться до їх мінімізації [1].

Фінансові ризики в свою чергу поділяють на 4 категорії: кредитний ризик, операційно-технологічний ризик, ризик ліквідності та ринковий ризик [1]. Ринкові ризики включають такі види ризиків: ризик дефолту, процентний ризик торгової книги, ризик кредитного спреда, фондовий ризик, валютний ризик, товарний та ризик волатильності [2]. В роботі розглядаються саме ринкові ризики та методи його аналізу, оскільки вони є невід'ємною складовою діяльності будь-яких фінансових установ.

Ринкові фактори, а саме постійні випадкові зміни цін, курси обміну валют, процентні ставки та інші, є причиною виникнення ринкових ризиків. Оцінка ринкових ризиків є однією з головних проблем фінансових установ, і саме тому для цього існує множина методів. Єдиного рішення для ризик-менеджменту не існує, а тому різні фінансові організації намагаються знайти метод, який би краще підходив для їх конкретних запитів.

Кількісною мірою оцінки ринкового ризику є величина value-at-risk (VaR). Основою для оцінювання цієї величини є зміни курсу та цін за певний період часу в минулому. VaR – це величина збитків, яка з певною ймовірністю не буде перевищена. Цю ймовірність називають рівнем довіри. Загалом прийнято робити оцінку з рівнем довіри 95% або 99%. Простіше кажучи, величина VaR – це максимально ймовірна втрата. Її можна вимірювати у грошовому вираженні та за відсотковою ставкою.

Для того, щоб переконатися в адекватності моделі, необхідно верифікувати її, тобто провести «бек-тестування». Верифікація дозволяє встановити ступінь адекватності використаної моделі оцінки ризику у вигляді показника VaR реальним умовам ринку [3].

Існують дві групи методів обчислення VaR: параметричні та непараметричні. Широко відомими непараметричними методами є метод історичного моделювання та метод імітаційного моделювання Монте-Карло. В свою чергу, серед параметричних методів популярністю користуються дельта-нормальний метод та метод дельта-гама наближення.

Поширеною мірою ризику є дисперсія, або стандартне відхилення очікуваної доходності. Вона є доволі точною й широко визнаною мірою ризику, яку використовують у теоретичних моделях оцінювання фінансових інструментів. Це показник розпорошення доходності навколо очікуваної величини. Чим менше розпорошені ставки доходів навколо очікуваної величини, тим вищою є впевненість в одержанні цієї доходності у майбутній період [3]. Пошук волатильності доходності фінансових інструментів є однією з головних задач при використанні дельта-нормального методу для визначення величини VaR.

Дисперсія фінансових процесів є змінною у часі величиною, а отже нам необхідно побудувати математичну модель, яка б коректно описувала динаміку дисперсії та могла б робити прогнози на кілька кроків вперед. Для аналізу фінансових часових рядів часто використовують модель авторегресії з умовною гетероскедастичністю (АРУГ) [4]. Гетероскедастичність означає, що дисперсія процесу є змінною в часі на інтервалі дослідження і при створенні моделі процесу потрібно знайти закон зміни дисперсії. Побудувавши адекватну модель, ми можемо значно покращити якість рішень в управлінні ризиком.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для того щоб застосувати методологію оцінювання ринкового ризику на основі гетероскедастичних моделей динаміки його волатильності у роботі використовуються дані стосовно цін акцій, представлені на світових біржах. Для детального опису процесу побудови моделей динаміки дисперсії та оцінки ринкового ризику на фондовому ринку було обрано дані компанії Qualcomm (QCOM). Вхідні дані – це зміна динаміки цін закриття активів за період з 1 листопада 2021 року по 31 жовтня 2022 року (Рис.1). Загальна кількість значень у вибірці – близько 250.



Рисунок 1. Значення доходностей акцій компанії Qualcomm.

Для того щоб побудувати модель авторегресії (АР) необхідно спочатку побудувати автокореляційну та часткову автокореляційну функцію (АКФ та ЧАКФ відповідно). Модель АР першого порядку має такий вигляд: $y(k) = c(1) + c(2) * y(k-1)$. Після оцінювання рівняння АР(1), отримуємо модель такого вигляду: $y(k) = 0.002712 + 0.981243 * y(k-1) + e(k)$.

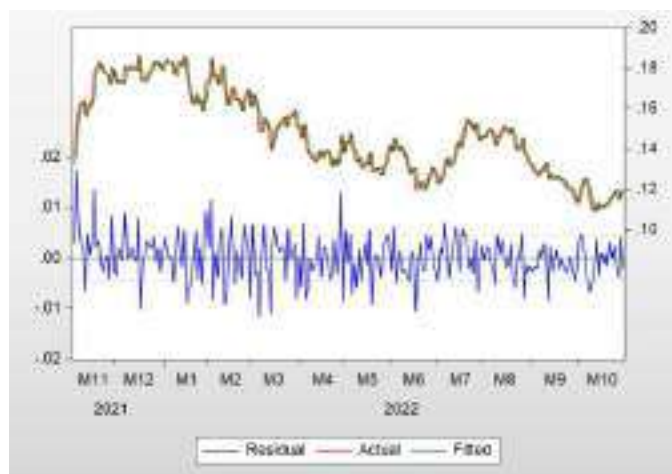


Рисунок 2. Графіки вихідного та оціненого рядів

Важливо підкреслити, що процес є гетероскедастичним, що підтверджується тестом Уайта щодо перевірки на гетероскедастичність залишків моделі. Наступний крок – це побудова моделі динаміки дисперсії. Для цього нам спершу потрібно визначити порядок моделі дисперсії, побудувавши корелограму квадратів залишків, а також корелограму вибіркової умовної дисперсії. Результати побудови корелограм дають нам зрозуміти, що порядки моделей для опису динаміки дисперсії рівні $p=1$ та $q=1$. Побудуємо такі моделі: ARCH(1), GARCH(1,1), EGARCH(1,1) та GJR(1,1).

Результати порівняння якості моделей зведено до Таблиці 1. Результати верифікації моделей для оцінювання ризику подано у Таблиці 2.

Загалом, за результатами моделювання можна зробити висновок, що кожна модель є задовільною для оцінювання ризиків. Втім, зважаючи на те, що інформаційні критерії моделей GARCH та EGARCH приймають мінімальні значення, вони є найкращими серед усіх оцінених.

Таблиця 1. Результати моделювання волатильності для вибраної акції

Тип моделі	LogL	AIC	SC	HQ
ARCH (1)	1806.562	-14.41250	-14.34207	-14.38415
GARCH(1,1)	2744.473	-21.90779	-21.82327	-21.87377
EGARCH(1,1)	2768.532	-22.09226	-21.99365	-22.05257
GJR(1,1)	1204.215	-9.577720	-9.479119	-9.538036

Таблиця 2. Результати процедури бек-тестування для оцінювання VaR

Тип моделі	VaR 95%	VaR 99%
ARCH (1)	(0) 100%	(0) 100%
GARCH(1,1)	(3) 98.81%	(0) 100%
EGARCH(1,1)	(8) 97.8%	(4) 98.43%
GJR(1,1)	(11) 95.58%	(3) 98.81%

Тепер порівняємо моделі на якість прогнозування. Нехай розмір навчальної вибірки нараховує 240 значень. Спрогнозуємо 15 наступних значень. Результати прогнозування подані у Таблиці 3.

Таблиця 3. Результати прогнозування волатильності за кожною з моделей

Тип моделі	MAE	MAPE	Theil Inequality
ARCH(1)	0.000728	36.96	0.0994
GARCH(1,1)	0.000547	1.68	0.0923
EGARCH(1,1)	0.000539	1.64	0.0983
GJR(1,1)	0.001551	4.84	0.0969

Порівняємо результати оцінювання VaR з результатами, отриманими за допомогою методу Монте-Карло, який є одним з найпопулярніших методів для оцінювання VaR.

Таблиця 4. Результати процедури бек-тестування для оцінювання VaR

Тип моделі	VaR 95%	VaR 99%
ARCH (1)	(0) 100%	(0) 100%
GARCH(1,1)	(3) 98.81%	(0) 100%
EGARCH(1,1)	(8) 97.8%	(4) 98.43%
GJR(1,1)	(11) 95.58%	(3) 98.81%
Монте-Карло	(7) 97.20%	(1) 99.96%

Загалом, можемо бачити, що усі моделі є у прийнятній мірі адекватними. Найкращими, як за статистичними критеріями, так і за практичною значимістю є моделі GARCH(1,1) та EGARCH(1,1). Модель експоненційної узагальненої авторегресійної умовно гетероскедастичності є найбільш адекватною для поставленої задачі, оскільки вона не завищує оцінки ризикової вартості та максимально правдоподібно відображає динаміку процесу. Порівнюючи результати, отримані за допомогою побудованих моделей з методом Монте-Карло, ми можемо бачити, що вони є достатньо близькими. Метод Монте-Карло є одним з найточніших для оцінювання ризиків, але складним в побудові. Використання гетероскедастичних процесів для оцінювання ризиків є простішою альтернативою.

4. ВИСНОВКИ

Фінансова організація чи виробниче підприємство за умов сучасної ринкової економіки зіштовхується з фінансовими ризиками, які можуть негативно позначитися на фінансову ефективність його діяльності. У зв'язку з цим велике значення мають наукові дослідження в галузі математичного моделювання процесів оцінювання фінансових ризиків та управління ними, що ґрунтуються на методах імовірнісної фінансової математики.

У ході дослідження побудовано адекватні моделі для опису динаміки поведінки та прогнозування дисперсії, яка використовується для оцінювання ринкового ризику. Для порівняння було побудовано кілька моделей – ARCH (1), GARCH(1,1), EGARCH(1,1) та GJR(1,1). В усіх випадках моделювання найгірші результати дає проста модель ARCH. Найкращі результати отримали за моделлю EGARCH.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ризики [Електронний ресурс] – Режим доступу до ресурсу: <https://a-bank.com.ua/uploads/media/default/0001/01/0dee1c4388ffda54e69e5fb7a98f5f948477edff.pdf>.
2. Про затвердження Положення про організацію системи управління ризиками в банках України та банківських групах: постанова НБУ від 11.06.2018 р. № 64. Дата оновлення: 06.12.2021. Режим доступу до ресурсу: <https://zakon.rada.gov.ua/laws/show/v0064500-18#Text>
3. Портфельне інвестування [Електронний ресурс] / А. А.Пересада, О. Г. Шевченко, Ю. М. Коваленко, С. В. Урванцева. – 2004. – Режим доступу до ресурсу: <https://buklib.net/books/21928/>.
4. Визначення величини ризику VaR на основі оцінок параметрів моделі стохастичної волатильності [Електронний ресурс] / П.І. Бідюк, М.М. Коновалюк // Систем. дослідж. та інформ. технології. — 2012. — № 3. — С. 85-94. – Режим доступу до ресурсу: <http://dspace.nbuv.gov.ua/handle/123456789/50180>
5. Taylor S.J. Modelling stochastic volatility: a review and comparative study // Mathematical Finance. — 1994. — No 4. — P. 183–204.

МОДЕЛЮВАННЯ СИСТЕМИ ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ РЕКЛАМНИХ НОСІЇВ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Каніовська І.Ю.¹, Бойко П.О.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ikaniovska@gmail.com,

²pjoidrivedisk@gmail.com

Продуктові компанії розміщують величезну кількість реклами і кожна реклама має свій рекламний носій. Ці носії були створені на основі досвіду дизайнерів і потреб компанії. Як наслідок, немає способу оцінити оголошення під час виробництва та дізнатися, наскільки вони добре працюють при використанні. Створення системи, яка буде прогнозувати ефективність рекламних носіїв суттєво допоможе правильно витратити рекламні бюджети та аналізувати основні складові успішних рекламних матеріалів. Метою є аналіз рекламних носіїв для розуміння їх складової ефективності. Результатом дослідження є розуміння основних елементів креативу, які впливають на ефективність та побудова системи прогнозування ефективності рекламних носіїв.

Ключові слова: рекламний носій, креатив, машинне навчання, ком'ютерне бачення, задача прогнозування.

1. ВСТУП

Продуктові компанії весь час витрачають мільйони на рекламу в інтернеті чи фізичному світі. В багатьох великих компаніях постає запитання, як зменшити неефективне використання рекламного бюджету. Однією із проблем “зливання” бюджету - маркетологи не можуть оцінювати ефективність рекламних носіїв до того, як вони використовують їх у рекламі.

Своєчасне припинення показу або не використання рекламних матеріалів є однією з найважливіших операцій з рекламою, яка може мати значний вплив на рекламні бюджети. Після попереднього аналізу 30 000 рекламних оголошень виявилось, що 50% рекламних креативів не принесли взагалі ні одної покупки. Це приблизно 5% рекламного бюджету, що були використані неефективно.

2. МЕТОДИ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ РЕКЛАМНИХ НОСІЇВ

Рекламний креатив — це зображеннями, або відео, які відображаються користувачам на веб-сторінці, у додатку чи іншому цифровому середовищі [1]. Є три типи оцінювання ефективності по часу: до запуску реклами, після закінчення транслявання рекламних носіїв та в реальному часі. На даний момент майже не досліджено перший тип - до запуску реклами.

Основні метрики по яким оцінюють ефективність рекламних носіїв – це коефіцієнт конверсії, коефіцієнт залучення, CTR, загальна вартість реклами.

Коефіцієнт конверсії: під час планування цифрової маркетингової кампанії коефіцієнт конверсії є одним із найважливіших KPI, на якому слід зосередитися. Цей показник відображає кількість відвідувачів, які перетворилися на клієнтів. Існує кілька різних типів коефіцієнтів конверсії, і можна використовувати будь-який із них, щоб визначити успіх рекламних кампаній.

Загальна вартість реклами: окрім коефіцієнтів конверсії, також можна переглянути загальну вартість кожного потенційного клієнта, що дасть вам уявлення про вартість ваших маркетингових зусиль.

CTR: відстеження того, скільки разів клієнт натискає на певне оголошення або відкриває електронний лист, може допомогти маркетологам визначити, наскільки ефективними є їхні стратегії цифрового маркетингу. Наявність точного звіту про ключові показники ефективності вашої кампанії може допомогти вам визначити, які кроки потрібно зробити, щоб збільшити продажі та покращити ефективність [2].

Залучення: вимірює кількість вподобайок, розповсюджень і коментарів, отриманих вашою рекламою. Цей показник показує, наскільки ваша реклама подобається людям. Це важливий показник, який слід перевіряти, бо реклами стає все більше, тому потрібно створювати таку рекламу, яка буде захоплювати користувачів взаємодіяти з нею.

Ключові показники ефективності рекламних носіїв можна покращити за допомогою виділення основних об'єктів в рекламних матеріалах, які є успішними, та застосування цих знань при розробці нових рекламних носіїв.

Основні об'єкти рекламних носіїв – це колір, текст, логотипи, предмети на зображенні або відео.

Для того щоб оцінювати рекламні кретиви до запуску можна застосувати методи комп'ютерного зору на основі глибокого навчання [3]. Це дозволить виділяти та проводити аналіз по основних складових рекламного носія:

- текст поданий у рекламі,
- основні кольори, текстури, яскравість,
- об'єкти (люди, тварини, тощо),
- розташування логотипу або кнопки, яка викликає дію, наприклад, покупка.

Для побудови моделі, яка буде обраховувати ефективність рекламних матеріалів, можна використовувати алгоритми машинного навчання. Один із прикладів – це використання навчання з вчителем. Тобто прогнозування числового значення, наприклад, CTR рекламних носіїв, маючи набір характеристик або ознак (проект, колір, текст тощо). Найбільш важливі алгоритми навчання з вчителем – це лінійна регресія (linear regression), логістична регресія (logistic regression), дерева прийняття рішень (decision trees), випадковий ліс (random forest), нейронні мережі (neural networks) [4].

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для дослідження було отримано доступ до даних по рекламних носіїв однієї з продуктових компаній України. В даних було отримано:

- рекламні носії,
- дані про дохід з рекламних носіїв,
- дані про витрати на показ рекламних носіїв.

Для отримання даних було написано два програмних кода на мові програмування Python 3. Перший на щоденній основі збирає дані витрат та доходів. Другий завантажує з Google Drive рекламні носії та зберігає локально на машину для подальшого використання.

Для розробки моделі, яка буде розраховувати ефективність рекламних носіїв було розроблено дві архітектури (Рис. 1 та Рис. 2).

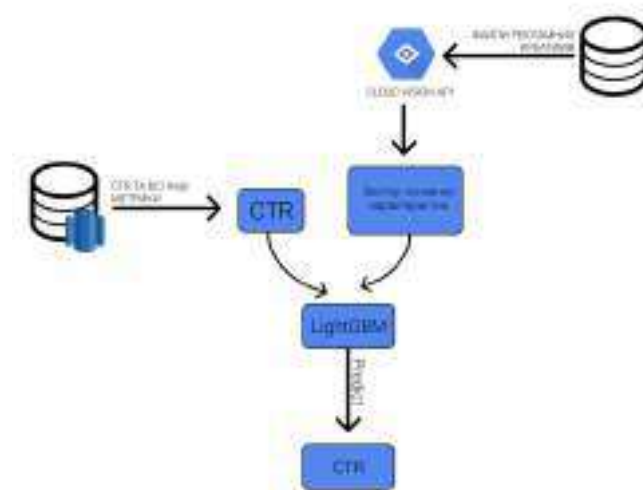


Рисунок 1. Архітектура №1

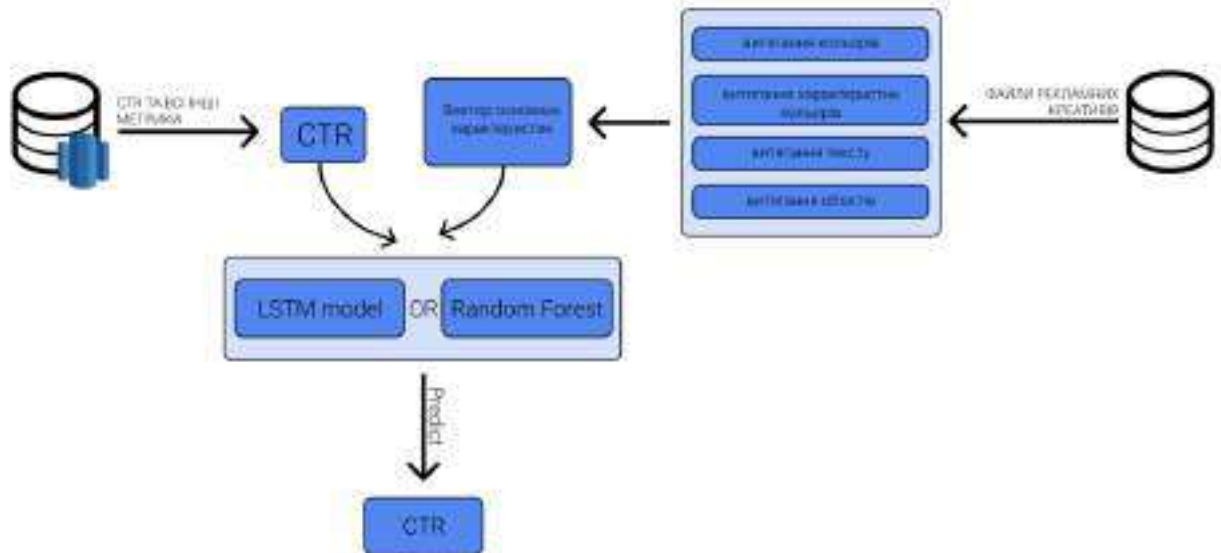


Рисунок 2. Архітектура №2

Архітектуру №1 було не реалізовано в повній мірі, бо один з елементів - Cloud Vision Арі не безкоштовний сервіс. Тому в Архітектурі №2 було замінено на 4 основні блоки: витягання кольорів з рекламних носіїв, розпізнавання тексту, розпізнавання об'єктів та отримання більш детальних характеристик по кольору.

Маючі рекламні креативи спочатку отримуємо інформацію про кольірну композицію зображень. За це відповідають дві функції: домінуючі кольори та показник барвистості. Для цього застосовувалась бібліотека extcolors. Також для розуміння барвистості було використано підхід, який Hasler and Sússtrunk розробили в своїй статті (шляхом опитування людей для оцінки барвистості за шкалою), бо люди можуть сприймати зображення як кольорове, дуже барвисте, не дуже кольорове або зовсім небарвисте (чорно-біле) [5].

Далі за допомогою бібліотеки `pytesseract` вилучаємо текст із рекламних носіїв. Та проводимо сентимент-аналіз для розуміння розподілу настрою в рекламному тексті. В результаті побачили більшість “позитивних” текстів.

Фінальний етап аналізу рекламних креативів був для вилучення типу об’єктів та його розташування на зображеннях за допомогою `Yolo`. Але не вийшло довести до успішного результату цей етап, бо алгоритм завжди виокремлював тільки людей (на рекламних носіях, які аналізувались в роботі, в більшості випадків були зображені люди), а всі інші елементи ігнорував на рекламних матеріалах.

Весь аналіз до цього зберігався в окремих для кожного етапу текстовому файлі. Далі в єдиний `Pandas DataFrame` збираються основні характеристики та метрики рекламних матеріалів і передаються на наступний етап машинного навчання. Основні колонки в фінальному `DataFrame` зображені на Рис. 3.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2289 entries, 0 to 2288
Data columns (total 29 columns):
#   Column                Non-Null Count  Dtype
---  -
0   color_1                2289 non-null   object
1   color_1_occurrence     2289 non-null   int64
2   color_2                2289 non-null   object
3   color_2_occurrence     2289 non-null   int64
4   color_3                2289 non-null   object
5   color_3_occurrence     2289 non-null   int64
6   color_4                2289 non-null   object
7   color_4_occurrence     2289 non-null   int64
8   color_5                2289 non-null   object
9   color_5_occurrence     2289 non-null   int64
10  click                  2289 non-null   int64
11  impression             2289 non-null   int64
12  post_engagement        2289 non-null   int64
13  ctr                    2285 non-null   float64
14  ID                     2289 non-null   object
15  Text                   2077 non-null   object
16  Text_Char_Length      2289 non-null   int64
17  Text_Word_Length      2289 non-null   int64
18  sentiment_score        2289 non-null   float64
19  sentiment              2289 non-null   object
20  Ad_color_diversity     2289 non-null   object
21  Ad_color_valence       2289 non-null   float64
22  Ad_color_arousal       2289 non-null   float64
23  Ad_color_dominance     2289 non-null   float64
24  Ad_hue_def             2289 non-null   float64
25  Ad_saturation_def      2289 non-null   float64
26  Ad_value_def           2289 non-null   float64
27  Ad_edge_pixel         2289 non-null   object
28  Ad_s_mean              2289 non-null   float64
dtypes: float64(9), int64(10), object(10)
memory usage: 517.7+ KB
```

Рисунок 3. Колонки в фінальному `DataFrame`

Було використано алгоритм випадковий ліс для оцінки ефективності рекламних кампаній по CTR. Результати Random Forest: Оцінка RMSE - 0.08, R2_Squared - 0.25, оцінка MAE - 0.06.

Для покращення роботи мережі Yolo треба самостійно відмітити наші об'єкти та контури, та натренувати на декількох кроків мережу, щоб покращити в результаті вилучення об'єктів та збільшити вектор ознак на вхід у моделі.

В подальшому можна використовувати більш новіші моделі, які основані на архітектурі трансформера. Також в поточній реалізації додали можливість отримання текстових даних з зображення і в наступних кроках можна використовувати ці тексти додаючи до поточних ознак. Та за допомогою конкатенації всіх ознак та подаючи їх на вхід на нейронну мережу або моделі машинного навчання отримати якісніший результат, оскільки ми матимемо більше інформаційних чинників з нашого ресурсу та покращенні вектори ознак, які можливо будуть краще лінеаризувати простір в якому ми будемо отримувати результати.

4. ВИСНОВКИ

На сьогоднішній день маємо одну з головних цілей маркетингових команд – це ефективне використання рекламних бюджетів. Виникає проблема – частина рекламного бюджету витрачається на неефективну рекламу. Тому є необхідність створення системи, яка буде обчислювати ефективність рекламних носіїв до того, як їх будуть використовувати у рекламних кампаніях.

У ході дослідження було реалізовано таку систему, яка дозволяє виконувати оцінку ефективності рекламних носіїв. Для цього було використано дані реальної продуктової компанії, написана системи отримання цих даних, були застосовані алгоритми комп'ютерного зору та машинного навчання.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. О.І. Вікарчук, О.О. Калініченко, С.М. Ніколенко, І.О. Поїнта Рекламний креатив, навч. посібник. Київ 2018, 7 с.
2. Jin Chen, Ju Xu, Gangwei Jiang, Tiezheng Ge, Zhiqiang Zhang, Defu Lian, Kai Zheng. Creative Optimization for E-Commerce Advertising. – 2021. Р. 1–3. – [Електронний ресурс]. – Режим доступу : <https://arxiv.org/pdf/2103.00436.pdf>
3. Ted Vrontas. 5 Examples That Show How Machine Learning is Changing Digital Advertising, 2020. [Електронний ресурс]. – Режим доступу : <https://instapage.com/blog/machine-learning-in-advertising>
4. Aurelien Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol; O'Reilly Media, 2019. 25-290 p.
5. David Hasler, Sabine E. Suesstrunk, Measuring colorfulness in natural images, 2003. [Електронний ресурс]. – Режим доступу : <https://www.spiedigitallibrary.org/proceedings/Download?urlId=10.1117%2F12.477378>

АНАЛІЗ РОЗПОВСЮДЖЕННЯ COVID-19 НА ТЕРИТОРІЇ УКРАЇНИ З ВИКОРИСТАННЯМ МОДИФІКОВАНОЇ SEIRD-МОДЕЛІ

Клименко А.І.¹, Подколзін Г.Б.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹asja653@gmail.com [0000-0001-9595-8155],

²podkolzin.gleb@iit.kpi.ua [0000-0002-7120-2772]

Метою роботи є розширення базової SEIRD-моделі за рахунок використання додаткових факторів. Метод дослідження програмна реалізація математичної моделі SEIRD на базі мови програмування: Python. Результатом дослідження є аналіз отриманих результатів з прогнозування розвитку епідемії COVID-19 із використанням статистичних даних перебігу захворювання на території України. Новизна дослідження полягає в урахуванні додаткових факторів при моделюванні таких, як народжуваність, смертність, вакцинованість так невакцинованість населення.

Ключові слова: епідеміологічні моделі, COVID-19, математичне моделювання, пандемія.

1. ВСТУП

Протягом усієї історії людське суспільство жило з періодичними епідеміями та пандеміями. Численні спалахи захворювань призводили до смерті, суспільних потрясінь і економічних розладів. Таким чином, прогнозування того, як спалах може прогресувати, важливо для пом'якшення його наслідків, чим і займається область епідеміологічного моделювання.

В епідеміології для опису епідемічних процесів і розуміння динаміки інфекційних захворювань використовуються різні математичні моделі. Багато сучасних математичних моделей, які використовуються для опису передачі інфекційних захворювань, базуються на компартментальних моделях Кермака та Маккендріка.

Епідеміологічні моделі відіграли центральну роль у пандемії COVID-19, особливо коли потрібні були термінові рішення, а доступних доказів було мало. Вони використовувалися для прогнозування еволюції захворювання та для інформування при розробці алгоритму дії запобіжних заходів.

2. МОДИФІКОВАНА SEIRD-МОДЕЛЬ

Пандемія COVID-19 вплинула на життя людей у всьому світі. Дослідники пропонують різні моделі та підходи для вивчення того, як можна пом'якшити розповсюдження хвороби. Найпоширенішими у епідеміологічному моделюванні є SIR, SEIR та SEIRD моделі. Найбільш актуальною вважається модель, що охоплює перебіг захворювання протягом усіх його етапів: період здоров'я людини, інкубаційний період, період клінічного перебігу захворювання, одужання та смерті. SIR модель не містить параметра, що відповідає за інкубаційний період захворювання, SIR та SEIR не можуть бути використані, якщо епідемія передбачає смертність та народжуваність. Саме тому для модифікації була обрана SEIRD-модель, що враховує всі необхідні параметри для дослідження всіх етапів захворювання.

Класична SEIRD-модель не розглядає наявність вакцинованих осіб у популяції. В даній роботі була запропонована модифікована SEIRD-модель, яка враховує цей фактор:

$$\left\{ \begin{array}{l} \frac{dS_{unvac}}{dt} = l - \mu S_{unvac} - \frac{S_{unvac}(\beta_{uu}I_u + \beta_{uv}I_v)}{N} \\ \frac{dS_{vac}}{dt} = -\mu S_{vac} - \frac{S_{vac}(\beta_{vu}I_u + \beta_{vv}I_v)}{N} \\ \frac{dE_{unvac}}{dt} = \frac{S_{unvac}(\beta_{uu}I_u + \beta_{uv}I_v)}{N} - (\mu + \alpha_{unvac})E_{unvac} \\ \frac{dE_{vac}}{dt} = \frac{S_{vac}(\beta_{vu}I_u + \beta_{vv}I_v)}{N} - (\mu + \alpha_{vac})E_{vac} \\ \frac{dI_{unvac}}{dt} = \alpha_{unvac}E_{unvac} - (\gamma_{unvac} + \mu + \theta_{unvac})I_{unvac} \\ \frac{dI_{vac}}{dt} = \alpha_{vac}E_{vac} - (\gamma_{vac} + \mu + \theta_{vac})I_{vac} \\ \frac{dR_{unvac}}{dt} = \gamma_{unvac}I_{unvac} - \mu R_{unvac} \\ \frac{dR_{vac}}{dt} = \gamma_{vac}I_{vac} - \mu R_{vac} \\ \frac{dD}{dt} = \theta_{vac}I_{vac} + \theta_{unvac}I_{unvac} + \mu(S_{vac} + S_{unvac} + E_{unvac} + E_{vac} + I_{unvac} \\ + I_{vac} + R_{unvac} + R_{vac}) \end{array} \right.$$

де S_{unvac} — сприятливі для вірусу невакциновані особи населення, які не інфіковані, але можуть заразитися при контакті з інфікованою особою (невакцинованою або вакцинованою); S_{vac} — сприятливі для вірусу вакциновані особи населення, які не інфіковані, але можуть заразитися при контакті з інфікованою особою (невакцинованою або вакцинованою); E_{unvac} — кількість невакцинованих людей із захворюванням у латентному режимі (контактували з інфікованою особою); E_{vac} — кількість вакцинованих людей із захворюванням у латентному режимі (контактували з інфікованою особою); I_{unvac} — кількість невакцинованих хворих людей, що передають вірус невакцинованим і вакцинованим сприятливим особам; I_{vac} — кількість вакцинованих хворих людей що передають вірус невакцинованим і вакцинованим сприятливим особам; R_{unvac} — кількість невакцинованих перехворівших, що сприйнятливі до повторного зараження, хоча ймовірність менша; R_{vac} — кількість вакцинованих перехворівших, що сприйнятливі до повторного зараження, хоча ймовірність менша; D — люди, що померли і від вірусу, і з інших причин; θ_{unvac} — смертність від вірусу інфікованих невакцинованих; θ_{vac} — смертність від вірусу інфікованих вакцинованих; β_{uu} — ймовірність передачі вірусу від інфікованих невакцинованих до невакцинованих; β_{uv} — ймовірність передачі вірусу від інфікованих невакцинованих до вакцинованих; β_{vu} — ймовірність передачі вірусу від інфікованих вакцинованих до невакцинованих; β_{vv} — ймовірність передачі вірусу від інфікованих вакцинованих до вакцинованих; α_{unvac} — ймовірність переходу захворювання із латентної фази у відкриту у невакцинованих; α_{vac} — ймовірність переходу захворювання із латентної фази у відкриту у вакцинованих; γ_{unvac} — одужання інфікованих невакцинованих від вірусу; γ_{vac} — одужання інфікованих вакцинованих від вірусу; μ — смертність не від інфекції; l — народжуваність.

Основна відмінність цієї моделі від класичної SEIRD-моделі - це поділення популяції на вакцинованих та невакцинованих. Ймовірність зараження вакцинованих осіб (S_{vac}) набагато нижча, ніж у невакцинованих осіб (S_{unvac}). Хворі вакциновані особи (I_{vac}) менш заразні, у них менше ймовірність померти, ніж у невакцинованих інфекційних осіб (I_{unvac}).

3. РЕЗУЛЬТАТИ МОДЕЛЮВАННЯ З УКРАЇНСЬКИМИ ДАНИМИ

Змоделюємо ситуацію з COVID-19 в Україні. Для розрахування коефіцієнтів скористуємося інформацією за 2021 рік.

Кількість українців за 2021 (без урахування окупованих територій) становить 41 588 354 людей [1, 3]. Смертність за 2021 рік становить 714 263 людей [2, 3], з яких COVID-19 становить 86 015 випадків [2]. Народжуваність за 2021 рік становить 271 983 дитини [2, 3].

Кількість усіх смертей за день у середньому становить $714\,263/365 = 1\,956.88$ людей. Тобто, за день від загальної кількості населення помирають $1\,956.88/41\,588\,354 = 0.0000470537$. У цю цифру також входить смертність від COVID-19. Смертність від COVID-19 становить $(86\,015/365)/41\,588\,354 = 0.0000056664$.

На жаль, статистики смертності вакцинованих та невакцинованих саме по Україні не було знайдено. Проте можна розрахувати цей коефіцієнт, якщо врахувати, що дані, що були зазначені в інтерв'ю з професоркою Школи громадської охорони здоров'я Леанни Вен [4], правдиві: що вакциновані люди мають у шість разів менше шансу заразитися, ніж у невакцинованих, і мають ймовірність померти від коронавірусу у 11 разів менше. Леанна Вен посилається на дослідження CDC (Centers for Disease Control) [5].

У цьому випадку смертність вакцинованих за день становить 0.0000004722, а невакцинованих 0.0000051942. Відповідно, маємо $\Theta_{unvac} = 0.0000051942$, $\Theta_{vac} = 0.0000004722$.

Відповідно, смертність за день не від COVID-19 становить $0.0000470537 - 0.0000056664 = 0.0000413872$, тобто $\mu = 0.0000413872$. Народжуваність за день становить $(271\,984/365)/41\,588\,354 = 0.0000179176$, тобто, $l = 0.0000179176$.

Так як для модулювання треба і вакцинована, і невакцинована популяція, треба навести дані стосовно цього. Станом на 2021 рік в Україні вакциновано двома вакцинами 15 201 112 людей [6]. Тобто, $15\,201\,112/41\,588\,354 = 0.3655136724$. У даному дослідженні моделюється популяція, що складається зі ста людей, серед яких відповідно є і вакциновані, і невакциновані.

З них буде два хворих (один вакцинований та один невакцинований), 35.5279289582 — це ще не хворілі вакциновані, а 61.8132941437 — ще не хворілі невакциновані.

Інкубаційний період — це кількість днів між моментом зараження чимось і моментом появи симптомів. Для розрахунку коефіцієнту, що відповідає за швидкість переходу вірусу із латентного періоду у повноцінний інфікований, наступною формулою:

де α — швидкість переходу захворювання із латентної фази у відкрити; T_{inc} — середній час інкубаційного періоду вірусу.

Віруси постійно змінюються, що іноді призводить до появи нових штамів. Різні штами COVID-19 можуть мати різні інкубаційні періоди. У середньому симптоми з'явилися у новоінфікованої людини приблизно через 5.6 дня після контакту [7]. Тобто, $T_{inc} = 5.6$, відповідно, $\alpha = 1/5.6 = 0.17858$. Відповідно до досліджень, інкубаційний період у вакцинованих та невакцинованих складає однакову кількість днів [8], тобто $\alpha_{vac} = \alpha_{unvac} = 0.17858$.

Дослідження показали, що організму може знадобитися 2 тижні, щоб подолати легку хворобу, або до 6 тижнів у важких або критичних випадках [9]. Отже, візьмемо час одужання за два тижні у середньому.

Розрахуємо γ відповідно до формули розрахунку коефіцієнту швидкості одужання:

$$\gamma = \frac{1}{T_{rec}}$$

де γ — коефіцієнт одужання інфікованих людей від вірусу; T_{rec} — середній час одужання.

Тобто, $T_{rec} = 14$ днів. Відповідно, $\gamma = \gamma_{unvac} = 1/14 = 0.0714$. У випадку, коли населення поділяється на вакцинованих та невакцинованих, час одужання буде різний. Дослідження 2021 року, що було проведено Центрами з контролю та профілактики захворювань, зазначено, що вакциновані учасники проводили в ліжку в середньому від двох до шести днів хвороби менше, ніж невакциновані [10]. Візьмемо в середньому, що вакциновані хворіють на шість днів менше, тобто $\gamma_{vac} = 1/8 = 0.125$.

Розрахуємо тепер ймовірності передачі захворювання. Для розрахування даних для вакцинованих людей, знадобиться статистика по ефективності вакцини (візьмемо вакцину Pfizer). Відповідно до даних [11], Pfizer має захист від легкої форми COVID-19 95%. Тобто, ймовірність захворіти, якщо ти вакцинований, $p_{заразитися, якщо ти вакцинований} = 0.05$.

Також треба зазначити, що вакциновані люди менш схильні до передачі хвороби, навіть якщо вони інфікуються. На листопадовій прес-конференції Тедрос Гебреесус, генеральний директор ВООЗ, заявив, що вакцини на 60 відсотків захищають від поширення вірусу до появи дельта-варіанту [12][13]. Тобто $p_{передати хворобу, якщо ти вакцинований} = 0.4$

Відповідно до того ж дослідження [12][14], у вакцинованих людей у десять разів менше шансів заразитися [13], а ймовірність заразити мене вдвічі менша, судячи з наведених вище цифр.

Тобто, $p_{заразитися, якщо ти невакцинований} = p_{захворіти, якщо ти вакцинований} \times 10 = 0.5$, а $p_{передати хворобу, якщо ти невакцинований} = 0.4 \times 2 = 0.8$.

Тепер можемо розрахувати коефіцієнти передачі захворювання.

$$\beta_{uu} = p_{передати хворобу, якщо ти невакцинований} \times p_{заразитися, якщо ти невакцинований} = 0.8 \times 0.5 = 0.4$$

$$\beta_{uv} = p_{передати хворобу, якщо ти невакцинований} \times p_{заразитися, якщо ти вакцинований} = 0.8 \times 0.05 = 0.04$$

$$\beta_{vu} = p_{передати хворобу, якщо ти вакцинований} \times p_{заразитися, якщо ти невакцинований} = 0.4 \times 0.5 = 0.2$$

$$\beta_{vv} = p_{передати хворобу, якщо ти вакцинований} \times p_{заразитися, якщо ти вакцинований} = 0.4 \times 0.05 = 0.02$$

Змодельюємо з цими параметрами модель. На рис. 1 можна побачити результати для ста днів:

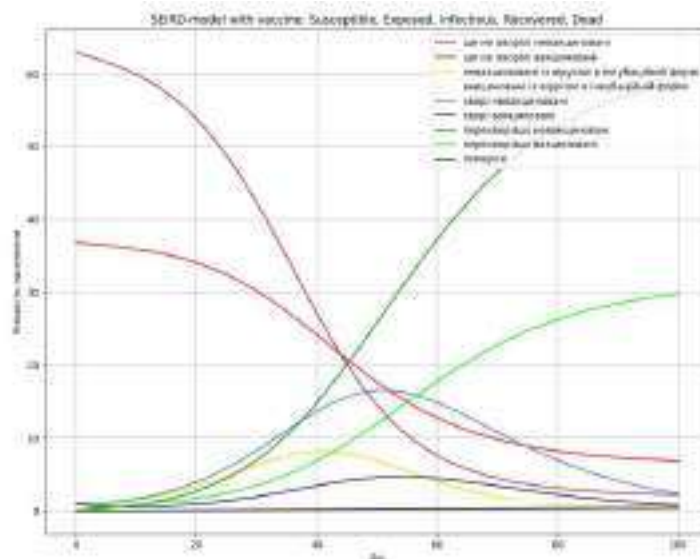


Рисунок 1. Модифікована SEIRD-модель для 100 днів

З рис. 1 можна побачити, як кількість ще не хворівших людей зменшується і одночасно кількість перехворівших зростає протягом часу (100 днів). Кількість перехворівших невакцинованих зростає швидше, і з урахуванням вакцинації [10] результат є достовірним. Кількість померлих відносно повільно зростає. Кількість не хворівших невакцинованих спадає швидше, бо у них ймовірність підхопити хворобу вище, у той час як вакциновані хворіють повільніше.

Якщо ж розглянути більший часовий проміжок (наприклад, 8 000 днів), то з рис. 2 можна побачити, як зростає кількість померлих, і зменшується для всіх інших груп дослідження. Наприклад, кількість перехворівших людей зменшується, бо модифікована SEIRD-модель враховує також природну смертність, у яку у даному випадку не входить смертність від COVID-19.

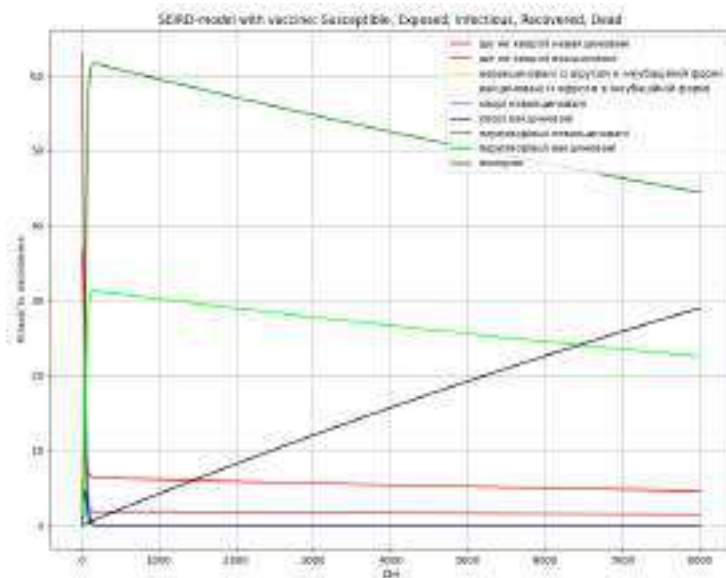


Рисунок 2. Модифікована SEIRD-модель для 8000 днів

Таке зростання смертності відносно всіх інших показників на рис.2 обумовлюється тим, що в Україні народжуваність менше, ніж смертність [15]. Якщо ж збільшити коефіцієнт народжуваності, щоб він перевищив смертність, то ситуація буде виглядати інакше (Народжуваність $l = 0,0500181008$) (Рис. 3):

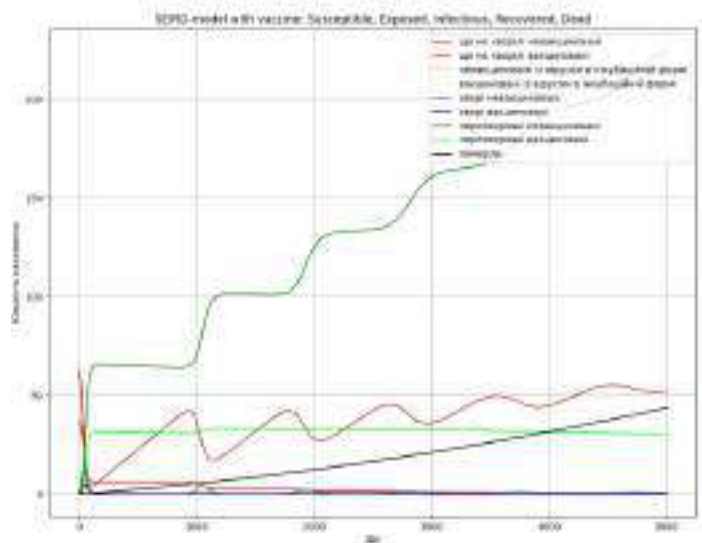


Рисунок 3. Модифікована SEIRD-модель для 5000 днів з народжуваністю $l = 0,0500181008$

Згідно рис. 3, зростає лише кількість ще не перехворілих невакцинованих, бо дана модель не передбачає вакцинацію і відповідно перехід від невакцинованих до вакцинованих. Уся популяція, що народжується, за замовчуванням є невакцинованою. З рис. 3 ще можна побачити, як періодично йде процес захворювання: на тисячному і двохтисячному дні можна побачити хвилі з хворих невакцинованих, і відповідно періодами зменшуються ще не перехворілі невакциновані.

4. ВИСНОВКИ

За результатами даної роботи було проаналізовано та досліджено модифіковану SEIRD-модель, на базі якої було зроблено наступні висновки, що корелюються з реальними статистичними даними України, а саме:

- Кількість захворілих невакцинованих в період піку епідемії приблизно в 3.5 рази більше, ніж захворілих вакцинованих;
- Смертність в Україні перевищує народжуваність;
- Розвиток епідемії має періодичний характер з тенденцією до затухання з часом.

Отримані результати підтверджують коректність роботи розробленої моделі, яка дозволяє моделювати ситуацію розповсюдження захворювання COVID-19 з урахуванням таких додаткових факторів, як розподілення населення на вакцинованих та невакцинованих, та використання додаткових коефіцієнтів, які дозволяють також врахувати вплив народжуваності та смертності населення. Використання перелічених додаткових факторів при моделюванні дозволяє прогнозувати більш реалістичну картину як розповсюдження захворювання, так і прорахування піків епідемії та її спадів. Результати даного дослідження можуть використовуватися в якості прогнозу розвитку епідемії для організації відповідних запобіжних заходів в медичних закладах, громадських місцях та ін.

Представлені розрахунки на базі модифікованої SEIRD-моделі використовували статистичні дані з українських джерел.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Населення України. URL: <https://index.minfin.com.ua/ua/reference/people/> (дата звернення: 18.11.2022)
2. Смертність України за 2021. URL: <https://index.minfin.com.ua/ua/reference/people/deaths/2021/> (дата звернення: 18.11.2022)
3. Держстат України. Демографічний щорічник населення України за 2021 рік: статистичний збірник. Київ. 2022. с.27, 53.
4. Why vaccinated people dying from Covid-19 doesn't mean the vaccines are ineffective. URL: <https://edition.cnn.com/2021/10/18/health/covid-19-vaccination-colin-powell-death-wellness/index.html> (дата звернення: 19.11.2022)
5. COVID-19-associated hospitalizations among vaccinated and unvaccinated adults ≥ 18 years – COVID-NET, 13 states, January 1 – July 24, 2021. URL: <https://www.medrxiv.org/content/10.1101/2021.08.27.21262356v1.full#T3> (дата звернення: 13.11.2022)
6. 6 Вакцинація в Україні. URL: <https://index.minfin.com.ua/ua/reference/coronavirus/vaccination/ukraine/> (дата звернення: 13.11.2022)
7. Coronavirus Incubation Period. URL: <https://www.webmd.com/lung/coronavirus-incubation-period#1> (дата звернення: 18.11.2022)
8. SARS-Cov-2 incubation period according to vaccination status during the fifth COVID-19 wave in a tertiary-care center in Spain: a cohort study. URL:

<https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-022-07822-4>

<https://www.webmd.com/lung/coronavirus-incubation-period#1> (дата звернення: 18.11.2022)

9. Coronavirus Recovery. URL: <https://www.webmd.com/lung/covid-recovery-overview#2> (дата звернення: 15.11.2022)

10. It's not just severity—the types of Covid symptoms you get depend on the vaccines you've received, new data says. URL: <https://www.cnbc.com/2022/10/26/covid-symptoms-differ-based-on-vaccination-status-zoe-health-study.html> (дата звернення: 14.11.2022)

11. COVID-19 vaccine efficacy explained. URL: <https://www.nebraskamed.com/COVID/covid-19-vaccine-efficacy-explained> (дата звернення: 12.11.2022)

12. Modelling of COVID-19. URL: <https://www.doherty.edu.au/our-work/institute-themes/viral-infectious-diseases/covid-19/covid-19-modelling/modelling> (дата звернення: 10.11.2022)

13. Vaccinated NSW residents to be allowed into Victoria as the state records 2,179 COVID-19 cases and six deaths. URL: <https://www.skynews.com.au/australia-news/coronavirus/watch-live-vic-health-officials-to-provide-covid19-update/news-story/863cdc24d57dd787251a8cfff26b5ec5> (дата звернення: 10.11.2022)

14. Your unvaccinated friend is roughly 20 times more likely to give you COVID. <https://theconversation.com/your-unvaccinated-friend-is-roughly-20-times-more-likely-to-give-you-covid-170448> (дата звернення: 12.11.2022)

15. Народжуваність в Україні продовжує падати і готується поставити новий антирекорд останніх 30-ти років. URL: <https://opendatabot.ua/analytics/deropulation-2021> (дата звернення: 10.11.2022)

ЗАСТОСУВАННЯ ФРАКТАЛЬНИХ МОДЕЛЕЙ НЕОБМЕЖЕНОЇ ВАРІАЦІЇ ДО АНАЛІЗУ ЧАСОВИХ РЯДІВ

Кравченко А.А.¹, Бондаренко В.Г.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹kravchenkoann1602@gmail.com, ²bondareng@gmail.com

Прогнозування майбутнього завжди було актуальною проблемою, не тільки через невизначеність, але й через те що для багатьох компаній прогнозування майбутніх цін акцій чи цінних паперів є невід’ємною складовою бізнес процесу. Метою роботи є аналіз часових рядів за допомогою моделі фрактального броунівського руху, їх основних властивостей та методів прогнозування. Результатом роботи є порівняльний аналіз прогнозу, побудованого за допомогою фрактального броунівського руху та класичної моделі для роботи з часовими рядами. У роботі було використано теоретичні та емпіричні методи дослідження.

Ключові слова: часовий ряд, фрактальний броунівський рух, прогноз, авторегресійні моделі, аналіз.

1. ВСТУП

Математичні моделі національної, регіональної та глобальної економіки є дуже потужними інструментами для теоретичних досліджень реальних макроекономічних процесів. Важливою перевагою макроекономічних моделей є їхня доступність для детального математичного аналізу та можливість їх вивчення за допомогою невеликої кількості вхідних даних [1].

Особливо цікавими є моделі що описують процеси з довгою пам’яттю, в багатьох випадках це моделі, що основані на дробовому численні, адже якщо брати моделі на основі інтегральних або диференціальних функцій цілого порядку, вони будуть описувати економічні процеси, в яких агенти фактично мають тотальну амнезію. Процеси з довгою пам’яттю можна описати як статистичні методи, що дозволяють аналізувати поведінку часових рядів. Через те що процеси з довгою пам’яттю мають ряд нестандартних властивостей, застосування вже звичних моделей Маркова та АРКС в більшості випадків не будуть давати бажаних результатів [2].

2. ЗАСТОСУВАННЯ ФРАКТАЛЬНИХ МОДЕЛЕЙ ДЛЯ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Оскільки при побудові стохастичних моделей часових рядів в ряді випадків доцільно використовувати в якості базового процесу фрактальний броунівський рух, то він і є метою цього дослідження. Модель фрактального броунівського руху, це процес $B_H t, t \geq 0, B_H 0 = 0$, де H – параметр Харста, що визначається як гаусівський випадковий процес з нульовим середнім та кореляційною функцією (1).

$$R(t, s) = E B_H(t) B_H(s) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}) \quad (1)$$

Варто також зауважити, що значення параметру Харста визначає поведінку часового ряду. Цей скаляр представляє відносну тенденцію часового ряду або до сильного регресу до

середнього, або до групування у певному порядку, а також визначає фрактальні властивості процесу. Оскільки параметр Харста є мірою довгострокової пам'яті часового ряду, то на основі його значення часовий ряд можна класифікувати як:

- $H < 0.5$ – антиперсистентний ряд. Чим ближче значення до 0, тим сильнішою є реверсія до середнього. На практиці це означає, що за високим значенням слідує назьке значення, і навпаки.

- $H = 0.5$ – геометричне випадкове блукання, тобто стандартний вінерівський процес.

- $H > 0.5$ – трендовий, або ж стійкий ряд. Чим ближче значення параметру Харста до 1, тим сильніший тренд, тобто на практиці, за високим значенням йде ще більше.

В силу автономності фрактального броунівського руху [5], параметр Харста можна знайти за допомогою однокрокового коефіцієнта кореляції приростів, визначених рівністю (2).

$$\rho_1 = 2^{2H-1} - 1 \quad (2)$$

Також варто відмітити, що прогноз значень фрактального броунівського руху можливий лише для персистентного випадку, тобто коли параметр Харста є більшим за 0.5. З основними характеристиками моделі фрактального броунівського руху детальніше можна ознайомитися в роботах [3,4].

Наступною задачею є оцінка параметрів фрактального броунівського руху. Тут варто зазначити, що основними об'єктами дослідження і оцінки є не значення фрактального броунівського руху, а саме їх прирости $y_k = X \frac{k}{n} - X \frac{k-1}{n}$, $k = 1, 2, \dots, n$, де $X t = \sigma B_H(t)$, що утворюють гаусівську послідовність. Постає гіпотеза: чи можна вважати побудовані раніше прирости, приростами фрактального броунівського руху? Перевірка цієї гіпотез полягає у визначенні значень наступних статистик (3):

$$\begin{aligned} A_n &= \frac{1}{n} v_k y_k^3, \text{ при } H \in 0, \frac{1}{2}, \\ B_n &= \frac{1}{n^{1+H}} v_k^2 y_k^3, \text{ при } H \in 0, \frac{1}{2}, \\ D_n &= \frac{1}{n^{2H}} v_k y_k^3, \text{ при } H \in \frac{1}{2}, 1, \\ F_n &= \frac{1}{n^H} y_k^3, \text{ при } H \in \frac{1}{2}, 1, \end{aligned} \quad (3)$$

де v_k визначаються як сума приростів $v_k = \sum_{j=1}^{k-1} y_j$.

Оскільки в даному дослідженні розглядаються ряди з довгою пам'яттю, то гіпотеза вважається прийнятною, якщо статистики D_n та F_n задовольняють наступним умовам:

$$0 < D_n < 6, -8.4 < F_n < 8.4 \quad (4)$$

Після того, як ми прийняли гіпотезу, що прирости y_k дійсно є приростами фрактального броунівського руху, тепер можемо побудувати прогноз на їх основі. Як відомо, найкращим прогнозом, або ж екстраполяцією траєкторії випадкового процесу $X(t)$ у точці t є величина $X(t + \tau)$, що задовільняє наступній умові $X t + \tau = E(X(t + \tau)|D_\tau)$, де D_τ - σ -алгебра, утворена випадковими величинами $X t, 0 \leq t \leq \tau$. Загалом у випадку фрактального броунівського руху визначення цього умовного середнього є доволі складною задачею, але в деяких випадках допускаються спрощення. В нашому випадку розглянемо ряд, побудований на основі гаусівського вектора:

$$\begin{aligned} \xi &= \xi_1, \dots, \xi_n, \xi = \xi_{tr}, \xi_{est}, \\ \dim \xi_{tr} &= m, \dim \xi_{est} = n - m, \\ \xi &\approx \aleph 0, S, \xi_{tr} \approx \aleph 0, A, \xi_{est} \approx \aleph 0, D \end{aligned}$$

Тобто в такому випадку кореляційний оператор ξ це блочна матриця наступного вигляду:

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

де елементи матриці B, C є взаємною кореляцією координат ξ_{tr} та ξ_{est} . Якщо ж розглядати перший вектор як вектор, що спостерігається, а другий вектор – той, що оцінюється відповідно, то за рахунок гаусовості найкращий прогноз збігатиметься з лінійною оцінкою:

$$\xi_{est} = E \xi_{est} \xi_{tr} = CA^{-1}\xi_{tr} \quad (5)$$

Тобто в нашому випадку, оскільки

$$\xi_k = y_k = B_H \frac{k+1}{n} - B_H \frac{k}{n},$$

Маємо змогу розбити прирости y_k на навчальну y_1, \dots, y_m та прогнозовану y_{m+1}, \dots, y_n вибірки, та за рухонек гіпотези, що ці прирости є приростами фрактального броунівського руху з відомою кореляційною функцією (1), формула (5) набуває наступного вигляду:

$$y_{m+j} = \sum_{k=1}^m \sum_{i=1}^m \left(\frac{m+j-k+1}{2} {}^{2H} + \frac{m+j-k-1}{2} {}^{2H} - \frac{m+j-k}{2} {}^{2H} \right) a^{ki} y_i \quad (6)$$

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для дослідження було використано дані щодо цін акцій відомої соціальної мережі “Twitter Inc” розмірністю 372 значення, що відповідає відрізку ледь більшому за календарний рік. Для застосування моделі фрактального броунівського руху була проведена необхідна первинна обробка даних, а саме приведення рядку динаміки зміни цін акцій з нестационарного ряду до стаціонарного, шляхом визначення та віднімання тренду, для досягнення нового ряду з нульовим середнім, а також переходом до приростів оброблених значень, аби досягнути однієї з умов стаціонарності – коваріація між будь-якими двома точками на однаковому проміжку є констатним значенням. Задля того, аби підтвердити гіпотезу, що побудований «новий» ряд можна описати за допомогою фрактального броунівського руху за допомогою формули (2) був обрахований параметр Харста, що показав, що даний ряд є часовим рядом з довгою пам'яттю, а також значення статистик (3), що належать довірчому інтервалу (4). Після цього припускаючи, що значення ряду утворюють гаусівський вектор(що було визначено за допомогою середнього значення та гістограми приростів), за рахунок його лінійності було побудовано прогноз, який в силу гаусовості збігається з лінійною оцінкою, за формулою (6). За допомогою цього були отримані результати прогнозу (Рис. 1).

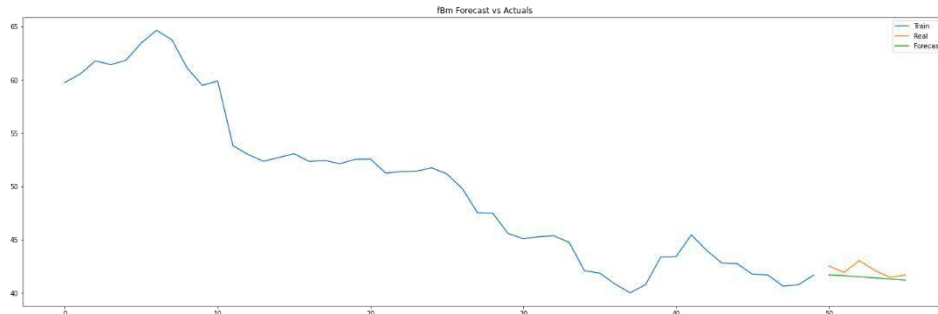


Рисунок 1. Графічне зображення прогнозу майбутніх значень за допомогою фрактального броунівського руху

Аби оцінити якість побудованого прогнозу, в якості порівняльного методу, також було побудовано прогноз для того самого вікна спостережень за допомогою моделі АРІКС [6], а саме АРІКС(1, 1, 1) (Рис. 2).

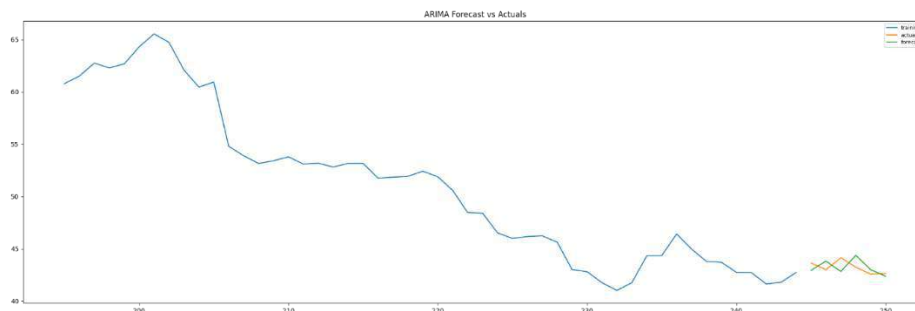


Рисунок 2. Графічне зображення прогнозу майбутніх значень за допомогою моделі АРІКС(1, 1, 1)

Результати порівняльного аналізу якості роботи прогнозування обома методами наведено у таблиці нижче (Табл. 1).

Таблиця 1. Характеристики адекватності моделей та оцінки якості прогнозів

Тип моделі	Характеристики моделі					
	R^2	SSE	DW	MSE	MAE	$Theil\ index$
ARIMA(1, 1, 1)	0.763	4.419	3.118	0.736	0.017	0.0005
fBm	-	3.882	-	0.647	0.015	-

Можна помітити, що за характеристиками, що відповідають за оцінку якості моделі, тобто середньою квадратичною похибкою, середньою абсолютною похибкою та накопиченням квадратів похибок (MSE , MAE та SSE) більш якісним прогнозом є прогноз за допомогою фрактального броунівського руху. Інші ж характеристики коефіцієнт детермінації R^2 , індекс Дарбіна-Вотсона DW та індекс Тейла, що відповідають за оцінку адекватності побудованої моделі, обраховані лише для моделі АРІКС, адже вони застосовуються лише для лінійних моделей і були наведені для того, аби показати, що модель АРІКС, з якою порівнюється робота моделі фрактального броунівського руху, дійсно є дійсно валідною моделлю для порівняння.

4. ВИСНОВКИ

З кожним днем економічні та фінансові системи стають все більш невизначеними та залежними від подій у зовнішньому світі, адже 21 сторіччя це вік персоналій, а не продуктів, тому багато економічних процесів, таких як ціна акцій чи цінних паперів, залежать не лише від успіхів самої компанії, але й від дій та висловлювань облич, що представляють цю компанію, а отже мають великий і доволі довгостроковий вплив на успіхи компанії.

Тому одним з найбільш ефективних методів у цій галузі є аналіз, оснований на моделях часових рядів із довгою пам'яттю. У ході дослідження на практичному прикладі реальних цін акцій протягом календарного року, було проведено аналіз за допомогою моделі фрактального броунівського руху. А також для порівняльного аналізу для того самого набору даних була використана одна з класичних лінійних моделей - АРІКС.

Результати дослідження показали, що навіть при точно підібраних порядках моделі АРІКС, та побудованого на її основі прогнозу, більш ефективним на заданій вибірці вхідних

даних все ж виявилася апроксимація майбутніх значень за допомогою моделі фрактального броунівського руху.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. R.G.D. Allen, *Mathematical Economics*, 2nd edn., Macmillan, London, 1960, 812p., ISBN 978-1-349-81549-4, 10.1007/978-1-349-81547-0.
2. W.Palma *Long-Memory Time Series: Theory and Methods*, Willey-InterScience, Hoboken, 2007, 304p., ISBN 978-0-470-11402-5.
3. F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, “Stochastic Calculus for Fractional Brownian Motion and Applications”, Springer, 329 p., 2013. doi: 10.1007/978-1-84628-797-8.
4. R.F. Peltier and J. Levy Vehel, “A new method for estimating the parameter of fractional Brownian motion”, *Rapport de recherché de l’INRIA*, no. 2396, 27 p., 1994.
5. I. Nourdin, “Noncentral convergence of multiple integrals”, *Ann. Probab.*, vol. 37, no. 4, pp. 1412–1426, 2009. doi: 10.1214/08-AOP435.
6. Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed). New York, USA: Springer

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ПОКАЗНИКІВ СМЕРТНОСТІ ВІД ЗАБРУДНЕННЯ ПОВІТРЯ

Кулина А.П.¹, Данилов В.Я.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹kulyna.anisiia@lil.kpi.ua

²danilov1950@ukr.net [0000-0003-3389-3661]

Задача побудови моделей та прогнозу є важливою складовою розв'язання проблеми забруднення та смертності у зв'язку із нею. Маючи модель, що здатна оцінити ризик смертності, можна зрозуміти доцільність покращення природних умов, відстеження та контролю викидів у атмосферне повітря. Для побудови таких моделей краще за все використовувати методи машинного навчання. Метою роботи є порівняння методів системи прогнозування показників смертності від забруднення повітря. Об'єкт дослідження - набір реальних даних, що описує рівень забруднення та рівень смертності від забруднення.

Ключові слова: машинне навчання, показник смертності, прогноз, забруднення повітря, PM_{2.5}, CO₂.

1. ВСТУП

Глобальне потепління, вирубка лісів, деградація екосистем і втрата життєво важливих середовищ – це лише деякі з багатьох проблем, які постали перед людством через широкомасштабну урбанізацію, економічний та промисловий розвиток. Немає сумніву, те, що відбувається сьогодні, визначатиме, наскільки ефективно ми будемо справлятися з ризиками та потрясіннями в майбутньому. І все ж забруднення повітря все ще залишається однією з найбільших проблем сьогодення, яка є наслідком вищеперерахованих чинників. Ці чинники безумовно полегшують умови існування людини, але є «але». Навколишнє середовище об'єднує у собі велику різноманітність важливих складових, однією з яких є атмосферне повітря [1]. Його забруднення підвищує ризик розвитку неінфекційних захворювань. Ці захворювання поділяються на три основні категорії: серцево-судинні захворювання, респіраторні захворювання та рак, які є основними причинами глобальної смертності. Окрім цього, це також збільшує ризик захворювань від інфекцій нижніх дихальних шляхів, передчасних пологів та інших причин смерті дітей і немовлят [2].

Враховуючи серйозність негативного впливу забруднення повітря, задача пошуку, аналізу та порівняння методів прогнозування показників смертності є актуальною проблемою. Маючи модель, що здатна оцінити ризик смертності, можна зрозуміти доцільність покращення природних умов, лікування хворих, розширити спектр прогнозу в дослідженнях для того, щоб зрозуміти важливість відстеження та контролю викидів у атмосферне повітря, а також способи їх зниження. Оскільки використання машинного навчання для прогнозування смертей від конкретних захворювань викликає все більший інтерес у науковій літературі [3], метою даного дослідження є порівняльний аналіз різних моделей машинного навчання (МН) для прогнозування показників смертності від забруднення повітря.

2. ПОСТАНОВКА ЗАДАЧІ

Метою роботи є розробка програмного забезпечення системи прогнозування показників смертності від забруднення повітря. Буде виконано аналіз і обробку набору даних та порівняння обраних методів машинного навчання для прогнозування показників смертності від забруднення повітря. Методи машинного навчання будуть застосовуватися на наборі реальних даних. За вхідними параметрами необхідно буде визначити рівень смертності. Об'єктом дослідження є набір реальних даних, що описує рівень забруднення (показник PM2.5 та CO2) та рівень смертності від забруднення.

3. ОПИС НАБОРУ ДАНИХ

У задачах машинного навчання вибір та правильна попередня обробка набору даних є вкрай важливими для отримання точних результатів. У даній роботі використовуються наступні набори даних: «Death Due to Air Pollution» (зібрано організацією Our World in Data), «PM2.5 Global Air Pollution» (The World Bank Group), «CO2 Emission by countries» (зібрано організацією Our World in Data). Ці дані було об'єднано в один спільний набір даних. Кінцевий датасет містить 6468 записів. Параметрами є країна, рік, рівень CO2 та PM2.5. Вихідним параметром є показник смертності населення. Дані продемонстровано на рис. 1.

	Entity	Year	Air pollution (total) (deaths per 100,000)	CO2 emission (Tons)	PM25
0	Afghanistan	1990	289.477309	59182404.0	0.000000
1	Afghanistan	1991	281.277967	61609770.0	0.000000
2	Afghanistan	1992	278.963056	62989232.0	0.000000
3	Afghanistan	1993	278.790915	64322287.0	0.000000
4	Afghanistan	1994	287.162923	66603889.0	0.000000
...
6463	Zimbabwe	2013	143.850145	701597967.0	23.111088
6464	Zimbabwe	2014	138.200536	713503854.0	22.805960
6465	Zimbabwe	2015	132.752553	725729475.0	21.733709
6466	Zimbabwe	2016	128.692138	736467042.0	21.726947
6467	Zimbabwe	2017	125.028843	746048675.0	22.251671

6468 rows × 5 columns

Рисунок 1. Вигляд датасету

Також нижче представлено на рис. 2 рівень смертності в усьому світі від забруднення повітря.

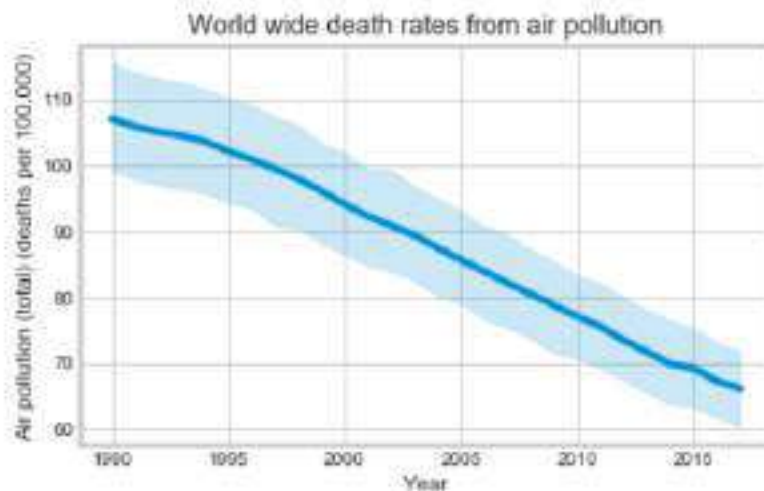


Рисунок 2. Рівень смертності в усьому світі від забруднення повітря

Для оцінки роботи моделей дані були випадковим чином розділені на дві вибірки. Поділ даних на навчальну та тестову вибірки було виконано у співвідношенні 70/30.

4. ОПИС АЛГОРИТМУ

Для виконання прогнозу смертності від забруднення повітря було обрано методи машинного навчання. Машинне навчання (МН) — це область штучного інтелекту, заснована на обчислювальних алгоритмах, які дозволяють комп'ютерам навчатися безпосередньо з даних, без явного програмування [4]. Алгоритми МН за прийнятний час аналізують великі обсяги даних, що представлені багатьма характеристиками (прогнозуючими змінними), і можуть виявляти та моделювати складні зв'язки між даними [5]. Таким чином, МН можна визначити як набір інструментів і методів для ідентифікації шаблонів у даних. Ці шаблони можна використовувати для покращення розуміння конкретної проблеми, прогнозування майбутнього та сприяння прийняттю рішень. Кажуть, що алгоритм навчається на даних. Використовуючи різні конфігурації, мета полягає в тому, щоб знайти модель, яка краще описує набір даних [6].

Моделі прогнозування в медицині історично спиралися на логістичну регресію [7]. Однак більш гнучкі методи машинного навчання призвели до розробки високоточних моделей. Наприклад, нещодавні дослідження продемонстрували, що алгоритми машинного навчання, такі як випадкові ліси та глибокі нейронні мережі, можуть передбачати гострі стани у пацієнтів точніше, ніж традиційна логістична регресія [8,9]. Тим не менш, важко заздалегідь визначити, який метод буде найкращим для конкретної проблеми [10].

Саме тому з метою виявлення оптимального методу для даної проблеми, порівняно кількох алгоритмів машинного навчання, які коротко описані нижче. Проведено порівняння

Лінійна регресія.

Лінійна регресія – це алгоритм машинного навчання, заснований на керованому навчанні, виконує завдання регресії. Регресія моделює цільове прогнозоване значення на основі незалежних змінних. Здебільшого використовується для з'ясування зв'язку між змінними та прогнозом.

eXtreme Gradient Boosting.

Метод посилення градієнту використовує дерева рішень як базові оцінювачі. Деревя будуються з використанням залишків, а не фактичних міток класу. Тобто, базовими оцінювачами в цих алгоритмах є дерева регресії, а не дерева класифікації. Це тому, що залишки неперервні, а не дискретні. Метод посилення градієнту будує дерева послідовно,

щоб кожне дерево покращувало модель [11]. У даній роботі використовувалася популярна реалізація GBM під назвою eXtreme Gradient Boosting (XGBoost).

Випадковий ліс.

Випадковий ліс подібний до XGBoost тим, що він створює ансамбль дерев рішень, але замість того, щоб будувати їх послідовно, він будує кожне дерево окремо на основі випадкової вибірки навчальних даних [12]. У кожному дереві доступна лише випадкова кількість змінних предикторів для кожного розділення «так-ні», що призводить до того, що дерева відрізняються одне від одного. Загальний результат враховує рішення усіх дерев. У регресійних задачах використовується середнє значення серед прогнозів по всіх деревах.

Метод опорних векторів.

Метод опорних векторів (SVM – support vector machine) проектує дані в N-вимірний простір на основі значень змінних, що є входними параметрами (N-кількість параметрів); а потім створює межу, яка чітко класифікує точки даних [12].

Метод k-найближчих сусідів.

Алгоритм k-найближчих сусідів, також відомий як KNN або k-NN, — це непараметричний метод навчання з учителем. Використовує сусідство для класифікації або прогнозування щодо групування окремої точки даних, виходячи з припущення, що подібні точки можна знайти поруч. Тобто, якщо k найближчих сусідів належать до певного класу, то і точка, яка підлягає класифікації, буде відноситися до цього ж самого класу. Даний метод можна використовувати як для регресії, так і для задач класифікації [12].

5. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Методи, що були описані вище, застосовано на наборі даних. Результати обчислення точності на тренувальній та тестовій вибірках представлено нижче в таблиці 1.

Таблиця 1. Результати обчислення точності для реалізованих методів

Метод	Точність тренувальної вибірки	Точність тестової вибірки
Лінійна регресія	0.9579	0.7980
eXtreme Gradient Boosting	0.9943	0.9372
Випадковий ліс	0.9967	0.9262
Метод опорних векторів	0.2925	-0.1378
Метод k-найближчих сусідів	0.9878	0.9219

Відповідно до таблиці 1 найкращий результат показує метод посилення градієнту (eXtreme Gradient Boosting). Але метод випадкового лісу та метод k-найближчих сусідів не сильно поступаються у точності моделі. Трохи гірший результат на тестовій вибірці показала лінійна регресія. Метод опорних векторів для даного набору даних зовсім не підходить, адже точність тренувальної вибірки досить низька, а точність тестової менше нуля. Від'ємне значення вказує на те, що лінія регресії не відповідає тренду даних, іншими словами, це означає, що модель погана.

6. ВИСНОВКИ

З кожним роком проблема забруднення повітря стає все більш важливою та необхідною для вирішення, щороку помирають мільйони людей через високу концентрацію небезпечних для живих організмів речовин. Необхідно мати системи що здатні оцінити ризик смертності та побудувати прогноз. У даній роботі було розглянуто низку методів машинного навчання для прогнозування показників смертності від забруднення повітря. Методи машинного навчання було застосовано на наборі реальних даних. Результати

показали, що для прогнозу смертності найкращою моделлю є метод посилення градієнту. Точність прогнозу складає 0.94.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Борисенко А. В. Проблематика громадського здоров'я [Електронний ресурс] / А. В. Борисенко // ДУ «Вінницький ОЛЦ МОЗ України». – 2020. – Режим доступу до ресурсу: http://cgz.vn.ua/problematika-gromadskogo-zdorovya/problematika-gromadskogo-zdorovya_455.html.
2. WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.
3. Ludermir TB. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. *Estudos Avançados*. 2021, v. 35, n. 101, pp. 85–94.
4. Minchole A, Rodriguez B. Artificial intelligence for the electrocardiogram. *Nature medicine*. 2019;25(1):22–3. Epub 2019/01/09. pmid:30617324
5. Deo, Rahul C. Machine learning in medicine. *Circulation*, v. 132, n. 20, p. 1920–1930, 2015. pmid:26572668
6. Wiens Jenna; Shenoy Erica S. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, v. 66, n. 1, p. 149–153, 2018. pmid:29020316
7. Goldstein BA, Navar AM, Pencina MJ, et al. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *J Am Med Inform Assoc*. 2017; 24:198–208
8. Churpek MM, Yuen TC, Winslow C, et al. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*. 2016; 44:368–374
9. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019; 572:116–119
10. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019; 110:12–22
11. Saini A. Gradient Boosting Algorithm: A Complete Guide for Beginners [Електронний ресурс] / Anshul Saini // *Analytics Vidhya*. – 2021. – Режим доступу до ресурсу: <https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/>.
12. Hastie T, Tibshirani R, Friedman JH: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, Springer, 2009, pp xxii, 745

МЕТОДИ ОЦІНКИ І ПРОГНОЗУВАННЯ ФІНАНСОВИХ РИЗИКІВ ДЛЯ БАНКІВСЬКОГО СЕКТОРА

Мельник І.А.¹, Гуськова В.Г.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ Imelnik406@gmail.com, ² guskovavera2009@gmail.com

Метою дослідження є створення адекватних математичних моделей та розробка ефективних методів вирішення питань моделювання та прогнозування фінансових процесів, а також визначення ризиків потенційних втрат на основі використання методів інтелектуального аналізу даних, прикладної статистики та аналізу часових рядів. Внаслідок того, що управління фінансовими ризиками є проблемою, з якою стикаються всі учасники ринку ця проблема є надзвичайно актуальною станом на зараз.

Ключові слова: орієнтування доходів менеджменту; бюджетування; планування; прогнозування; управління ризиками

1. ВСТУП

Людство завжди намагалося зрозуміти навколишній світ, шукаючи в усьому зв'язки та закономірності, накопичуючи величезні обсяги даних. Технології інтелектуального аналізу (ІАД) даних допомагають аналітику зрозуміти інформативність та повноту даних, виявити їх зв'язок та створити модель для вирішення реальних проблем[1]. Інтелектуальний аналіз даних спрямований на виявлення раніше невідомих закономірностей у масивах даних, реалізацію практичної та доступної інтерпретації знань, а також побудову аналітичних моделей, оптимізованих для вирішення певних класів прикладних задач при прийнятті рішень у різних сферах людської діяльності.

Стандартні технології аналізу даних базуються на підходах математичної статистики, що не у всіх випадках описують реальний стан речей та стандартні статистичні методи відкидають нетипові спостереження — так звані піки та сплески. За основу інтелектуальний аналіз даних бере концепцію шаблонів, щоб знайти широкоаспектний зв'язок між даними. Саме ці шаблони дозволяють зобразити неочікувані зв'язки та закономірності і залучити «приховані знання» («hidden knowledge») в наборах даних[2].

2. МЕТОДИ ОЦІНКИ ФІНАНСОВИХ РИЗИКІВ БАНКІВСЬКОГО СЕКТОРУ

Для застосування методологій та моделей ІАД, машинного навчання моделі та досягнення бажаних успішних результатів необхідно підготувати та сформувати набір даних. Визначення очікуваного фінансового ризику буде можливим за допомогою створеної моделі та доступу користувача до економічних даних. Також, важливо порівняти алгоритми та підходи ІАД та дослідити їх вплив на точність прогнозу моделі, відсоток розподілу масиву даних, що використовується для навчання та тестування, а також вплив пропусків у масиві даних.

Постановка та визначення задачі прогнозування полягають у наступному:

- 1) сформувати набори даних, використовуючи дані з відкритих сайтів (наприклад kaggle.com), які включають економічні показники та певний фінансовий ризик для кожного набору;
- 2) оцінити сучасні інструменти та програми для проведення ІАД;
- 3) провести огляд методологій, підходів та моделей інтелектуального аналізу даних для обробки набору даних та виконання моделювання і прогнозування фінансового ризику;
- 4) надати обґрунтування вибору середовища та мови програмування для вирішення питання створення програмного комплексу;
- 5) виконати аналіз даних (перший аналіз даних, візуалізація даних);
- 6) виконати обробку набору даних (приведення категоріальних стовпців до числових, заповнення пропусків, нормалізація даних);
- 7) створення математичних моделей для прогнозування фінансового ризику;
- 8) для навчання моделі прогнозування ризиків використовувати необхідні методи, наприклад лінійна регресія;
- 9) потім слід розглянути результати навчання та дослідження.

При створенні математичної моделі для представлення об'єкта використовуються теорія множин, теорія графів, теорія ймовірностей та інші математичні методи. До моделі застосовуються такі умови:

1. Модель повинна належним чином відповідати процесу або об'єкту.
2. Якщо аналітичне рішення є неможливим, для кожного рівняння в моделі повинно існувати чисельне рішення.
3. Модель повинна бути достатньо адаптивною, щоб її можна було використовувати для пояснення групи пов'язаних процесів або функціонування процесу за різних обставин.
4. Вимога робастності (robust - міцний, довговічний). Під робастністю розуміється здатність моделі забезпечувати відповідний прогноз вихідної змінної для будь-якого відрізка часового ряду, що відповідає обраному режиму роботи, а не тільки для того відрізка, на якому вона побудована.
5. Необхідність адаптивності. Внаслідок цього критерію принаймні частина параметрів моделі (хоча б один) може бути покращена в міру надходження свіжих даних з об'єкта.

Придатність регресійних моделей може бути визначена шляхом аналізу послідовності залишків (похибок моделі), при цьому розрахункові значення отримують шляхом заміни реальних значень усіх компонентів, що входять до моделі, на розрахункові. Послідовність залишків досліджується на відповідність критеріям випадкової складової економічного часового ряду, зокрема близькості математичного сподівання до нуля, випадковості відхилень, відсутності автокореляції та нормальності закону розподілу.

В основі статистичних методів аналізу та прогнозування, як правило, лежить глибока обробка статистичних даних, що стосуються досліджуваного процесу. В результаті можна виділити наступні етапи прогнозування.

Етап 1 - постановка завдання - визначається перелік факторів і показників (змінних), опис зв'язків між ознаками, що цікавлять, а також ролі цих факторів і показників (тобто які з них в рамках конкретного завдання можна вважати вхідними (тобто повністю або частково регульованими або, принаймні, такими, що легко піддаються реєстрації і прогнозуванню)[3].

Етап 2 - апріорний, передмодельний аналіз - ґрунтується на первинному вивченні змістовної сутності досліджуваного явища або процесу. На цьому етапі розробляються і формалізуються різноманітні гіпотези і припущення, які представляють собою апріорні знання про ці явища.

Етап 3 - інформаційно-статистичний - відомий також як збір даних, передбачає реєстрацію значень використовуваних у дослідженні компонентів і показників на різних часових і (або) географічних циклах функціонування модельованої системи.

Етап 4 - Специфікація моделі базується на передумовах і перших припущеннях, зроблених на другому етапі. Вона також передбачає створення фундаментальних зв'язків моделі, що з'єднують вхідні та вихідні змінні.

Етап 5 - Для того, щоб "підігнати" значення невідомих параметрів під вихідні статистичні дані, а саме - ідентифікація моделі - проводиться статистичне дослідження даних. Для реалізації цього етапу спочатку необхідно визначити, чи можливо теоретично відновити значення невідомих параметрів моделі за наявними вихідними статистичними даними для обраної на четвертому етапі структури моделі.

Етап 6 - верифікація моделі - порівняння результатів, оцінок і наслідків моделі з реальністю з використанням різних підходів. Інша назва цього етапу - статистичний аналіз точності та достатності моделі.

Об'єктивна оцінка якості прогнозу є важливим кроком у процесі прогнозування. Життєво важливо оцінити очікувані значення, оскільки вони є випадковими величинами. Обґрунтованість і точність прогнозів є двома важливими складовими прогнозування. На цьому етапі для оцінки якості прогнозу використовується набір стандартів, методів і процесів[4].

3. АНАЛІЗ РЕЗУЛЬТАТІВ МОДЕЛЮВАННЯ ФІНАНСОВИХ РИЗИКІВ

У зв'язку з тим, що визначення кредитоспроможності є погано структурованою проблемою з об'єктивною наявністю як якісних, так і кількісних показників, моделювання фінансових ризиків, зокрема кредитного ризику, ускладнюється. Крім того, багато банківських даних можуть бути представлені у нечіткому вигляді, а не піддаватися точному, чіткому визначенню[5].

Крок 1 - База знань та пояснення даних експерименту. Для первинного аналізу та розробки математичних моделей визначення кредитоспроможності було обрано вибірку з 3350 клієнтів. Ці клієнти були ідентифіковані за такими ознаками: x_1 - стать позичальника; x_2 - вік; x_3 - сімейний стан; x_4 - кількість дітей; x_5 - наявність власного житла; x_6 - працює чи не працює; x_7 - освіта; x_8 - вид діяльності; x_9 - наявність поручителя-фізичної особи; x_{10} - сума кредиту. Вихідна змінна - оцінка кредитного ризику - позначена у (результат аналізу).

Крок 2 - Заповнення пропусків у даних. Для застосування до генеральної сукупності дані, отримані в результаті вибіркового обстеження, мають бути оброблені. Значення узагальнюючих показників встановлюються на основі зведеного набору всіх даних, зібраних під час обстеження.

Виключення часткових спостережень з прогалинами і подальший аналіз отриманих "повних" даних є найпростішим методом інтерпретації неповних даних. Штучні пропуски у співвідношенні 5% до 10% були введені у вихідних даних дослідження кредитоспроможності.

Для заповнення пропусків було використано регресійну модель та середнє значення вибірки. Значення похибки MAPE для різних методів заповнення пропусків даних про кредитоспроможність наведено нижче (Табл. 1).

Таблиця 1. Значення похибок MAPE при заповненні пропусків

Метод заповнення пропусків	Часовий ряд	Пропущено значень, %	
		5	10
Модель регресії	Заробітна плата	16,12	17,16
	Сума кредиту	4,92	5,01
Середнє за вибіркою	Заробітна плата	22,15	24,19
	Сума кредиту	28,35	27,87

Як видно з табл.1, в експерименті використовуються часові ряди з 5% та 10% фіктивних пропусків даних відносно загальної кількості значень часового ряду. Як видно з таблиці 1, похибка MAPE значно варіюється для розглянутих часових рядів. Однак найбільш успішними виявилися методи, які використовували для заповнення пропусків прогнози регресійної моделі (MAPE = 4,92%).

Крок 3 - у табл. 2 наведено точність моделей, створених на основі критеріїв інформативності оцінювання змінних. В якості моделей використовувалися модель LOGIT та нечіткі системи на основі правил з трикутною та трапецієподібною функціями належності.

Таблиця 2. Показники якості моделі для всіх змінних та меншого набору атрибутів

Тип моделі		Точність (загальна кількість)
Logit model		0,89
FRBS	TRIANGLE	0,851
	TRAPEZOID	0,940

З таблиці бачимо, що найвищу точність результатів показала система на основі нечітких правил з трапецієподібною функцією належності (0,940).

Крок 4 - співвідношення (%) навчальної та тестової вибірок, що використовувалися для прогнозування, становило 70/30, 80/20 та 90/10. Результати наведені в таблиці 3.

Таблиця 3. Інструменти прогнозування атрибутів

Співвідношення навчальної та перевіркової вибірок, %		70/30	80/20	90/10
Modeltype		Значення MSE		
Logitmodel		0,76658	0,76838	0,7612
FRBS	TRIANGLE	0,6104	0,61050	0,6191
	TRAPEZOID	0,23062	0,23001	0,2303

Системи, що використовують нечіткі правила з трапецієподібною функцією належності та співвідношенням навчальної та тестової вибірки 80/20, показали найкращий результат MSE - 0,23001.

4. ВИСНОВКИ

Початкова інформація та характеристики, що визначають кредитоспроможність позичальника, можуть бути суперечливими або відсутніми. На кредитоспроможність позичальника, як правило, впливає його фінансовий стан, який включає власні кошти, ліквідні активи та навички управління грошовими коштами. Під час двох етапів оцінки кредитоспроможності необхідно враховувати бізнес-ризик, а також фінансовий стан позичальника, для чого розглядаються фінансові коефіцієнти та грошові потоки. За цих обставин бажано обирати стратегію, яка поєднує в собі низку методів аналізу даних та прогнозування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Методи інтелектуального аналізу даних [Електронний ресурс] – Режим доступу до ресурсу: <https://buklib.net/books/24506/>
2. Statistical Modeling [Електронний ресурс] – Режим доступу до ресурсу: <https://www.heavy.ai/technical-glossary/statistical-modeling>
3. Theodore M. Barnhill and William F. Maxwell Journal of Banking & Finance, 2002, vol. 26, issue 2-3, pp. 347-374.
4. Forecasting in manufacturing management [Електронний ресурс] – Режим доступу до ресурсу: https://link.springer.com/referenceworkentry/10.1007/1-4020-0612-8_363
5. Qualitative and quantitative analysis of creditworthiness of the companies [Електронний ресурс] – Режим доступу до ресурсу: <https://js.ugd.edu.mk/index.php/JE/article/view/2736>

АНАЛІЗ ПРОЦЕСІВ ТА ПРИЙНЯТТЯ РІШЕНЬ НА РИНКУ КРИПТОВАЛЮТ

Мельников А.А.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ melnykovandrew@gmail.com

Об'єктом дослідження даної роботи є ринок криптовалют. Предметом дослідження є часові ряди фінансового ринку та, зокрема, ринку криптовалют, а також методи машинного навчання, які спрямовані на більш якісне та ефективне прийняття рішень. Метою роботи є розробка системи прийняття рішень (СППР), що базується на методах штучного інтелекту, та відтворення схеми прийняття рішень трейдера відповідними моделями. Результатом дослідження є модель побудована на базі методу штучного інтелекту та машинного навчання, а також висновок про придатність нейронних мереж до вирішення поставленої проблеми за умови постійного додаткового навчання. В роботі використано емпіричні й теоретичні методи дослідження.

Ключові слова: криптовалюта, машинне навчання, згорткові нейронні мережі, трейдинг, глибоке навчання, бінарна класифікація.

1. ВСТУП

На сьогодні криптовалютні операції стають дедалі більш поширеним методом розрахунків між фізичними особами. Також зростає інтерес до операцій з деривативами. Зокрема вже зараз для українців доступні два типи ф'ючерс контрактів: безстроковий та кварталний [1]. У розріз із типовим біржовим трейдингом для торгівлі криптовалютами або для заключення контрактів на їх купівлю продаж не потрібно спеціальної ліцензії, а правове регулювання в цій сфері доволі м'яке, що робить проведення операцій доступним для будь кого у віці від 16 років.

Однак є і перепони на шляху популяризації біржових операцій на основі криптовалют, зокрема відсутність прив'язки до фізичних активів. Це призводить до важко прогнозованості поведінки ринку криптовалют – ціна формується лише очікуваннями користувачів. Оскільки поведінка ринку залежить від рішень користувачів, то в роботі було вирішено відтворити процес прийняття рішень за допомогою класифікації деякого відрізка часового ряду - аналогічно до роботи трейдера.

Під час виконання роботи поставлено та виконано такі задачі:

1. Розробити систему прийняття рішень, що базувалась на методах штучного інтелекту;
2. Відтворити схему прийняття рішень трейдера відповідними моделями;
3. Провести порівняння з загально відомими моделями.

Побудована система повністю відповідає поставленим цілям і відтворює процес прийняття рішень трейдером (тобто підхід до задачі іде з боку задачі класифікації а не прогнозування часового ряду). В якості ядра системи було обрано нейронну мережу та логістичну функцію втрат.

Також виконано аналіз моделей на старіння та ефективність підходу на 16 криптовалютах. В якості візуалізації ефективності роботи моделі було побудовано РОК криві для моделей в періоді навчання та на ринкових даних після зміни загальної тенденції.

2. РОЗГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ ТА РОЗКРИТТЯ ЗАДАЧІ

Розробка моделі спирається на факт домінування на ринку візуального сприйняття ситуації трейдерами. Часткове розкриття структури ринку криптовалют було розглянуте в попередніх роботах [2]. Спираючись на попередні дослідження очікування користувачів є найбільш впливовим фактором для ринку криптовалют.

Однак з іншого боку одним з важливих факторів що впливають на формування ціни криптовалют також виступає ціна підтримки мережі як платіжної системи. Ця величина може бути оцінена і закладена як мінімальна ціна активу. Відповідно, з кожним роком вона зростає через ускладнення мережі, контрольованої інфляції цін на електроенергію та залучення більшої кількості користувачів до мережі. Загалом це створює тенденції до зростання ціни криптовалют що робить прибутковими навіть стратегії з нульовим математичним середнім (частково використання цих стратегій призводить до зростання капіталу що знаходиться безпосередньо в криптовалюти як інвестиція – провокуючи подальше зростання). Однак основною ціллю даної роботи є прогнозування змін курс відповідно до першого чинника - очікування користувачі, звісно в майбутньому цей чинник буде зменшувати свій вплив (рис. 1).

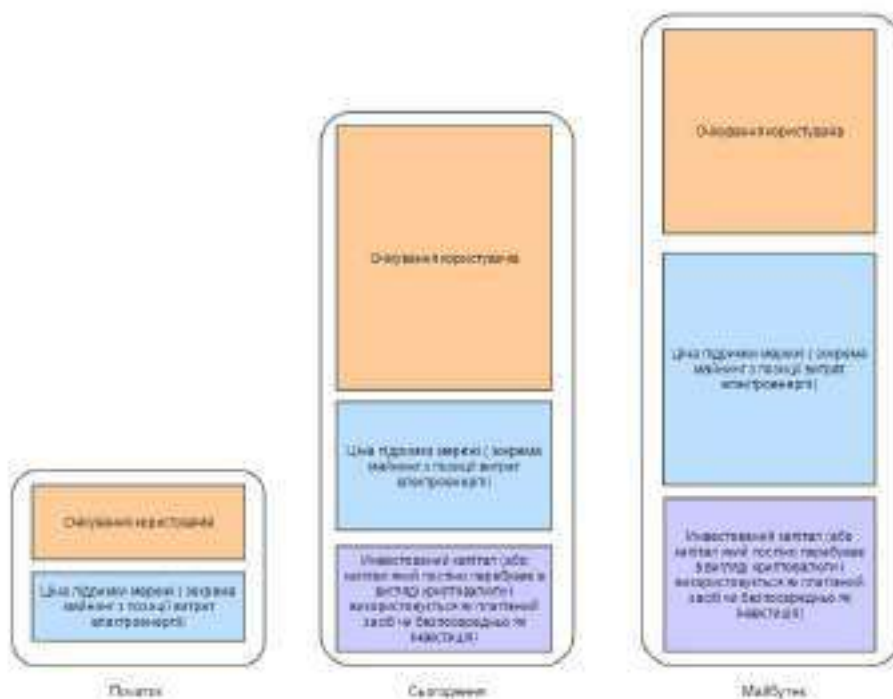


Рисунок 1. Основні ціноутворюючі чинники для криптовалют (представлено динаміку зростання ціни підтримки системи та інвестиційної привабливості криптовалют)

Через неможливість оцінити сподівання безпосередньо було вирішено використовувати нейронну мережу як засіб що дозволяє зробити висновки аналогічні до висновків трейдера та знайти приховані закономірності в їх поведінці.

Оскільки ситуація на ринку не може характеризуватись постійно одними величинами, то побудована модель матиме тенденції до старіння. Як пізніше буде розглянуто в розділі 3. Цей проміжок займатиме лише 3 дні. Після цього часу модель необхідно перевірити на якість та прийняти рішення про додаткового навчання чи продовження роботи (рис. 2).



Рисунок 2. Життєвий цикл моделі

Модель безпосередньо відповідає за прийняття рішень. Користувач моніторить загальну роботу системи лише за тижневими звітами. Відсторонення користувача від роботи системи є свідомий крок. На жаль, людина не може бути абсолютно неупередженим, а з вигляду на ефективність прийняття рішень упередженість користувача може бути фатальною (оскільки мова йде безпосередньо про гроші, людина схильна до ризику при ймовірності високої винагороди, навіть при незначному шансі на виграш).

Також революційним є відмова від безпосереднього прогнозування курсу. Саме перехід на навчання системи приймати рішення дозволяє збільшити точність. Адже з практичної точки зору не має сенсу наскільки зростає курс активу, якщо це приносить прибуток то є успіх. Таким чином модель безпосередньо навчається діям аналогічним до дій користувача.

Отже, у розділі було розглянуто основні характеристики розробленої системи та її особливості. Було наведено характеристику досліджуваного об'єкта, а саме ринків криптовалют. Розглянуто складові ціноутворення.

На основі викладених фактів було прийнято рішення про розробку системи, що базується на нейронній мережі і відповідає безпосередньо за прийняття рішень.

3. РОЗГЛЯД МЕТОДІВ ВИРІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

Трейдинг доволі консервативна сфера прийняття рішень. Хоча на сьогодні на будь-якому ринку трейдеру доводиться обробляти величезні масиви інформації та шукати приховані залежності. Однак прийняття рішень все відповідальний процес який неможливо перекласти з людини через правове регулювання сфери. Відповідно до законів алгоритм прийняття рішень повинен бути чітким і зрозумілим, однак нейронні мережі працюють за принципом чорного ящика і не дають чіткого розуміння роботи в критичних ситуаціях. Однак криптовалюта не має чітких норм щоб регулювали операції з нею, а отже в немає жодних обмежень на модель.

З іншого боку доволі масове використання криптовалют почалось лише 4-6 років тому - а отже сфера має незначну вибірку даних, що вже накладає обмеження на методи побудови моделей.

Одним з найважливіших моментів при побудові дослідження є обрання метрики для оптимізації роботи моделі, тобто формалізація поставленої задачі за допомогою функції на обраному просторі (для формалізації очікувань користувача/дослідника). Було розглянуто наступні метрики:

1. Binary cross entropy (бінарна перехресна ентропія):

$$H(y, \underline{y}) = \frac{1}{n} \sum_{i=1}^n y_i \log(y_i) + (1 - y_i) \log(1 - y_i),$$

де y – вектор дійсних значень розмірності n , де n – кількість тренувальних даних,

\underline{y} – вектор отриманих прогнозів моделі,

y_i, \underline{y}_i – відповідні координати векторів.

Дозволяє враховувати як правильно класифіковані об'єкти так і неправильно з однаковою вагою і тих і тих класів. Приводить до більш чіткого розділення класів. Зокрема на рисунку 3 подано зміну значення похибки при дійсному значенні 1 та 0.

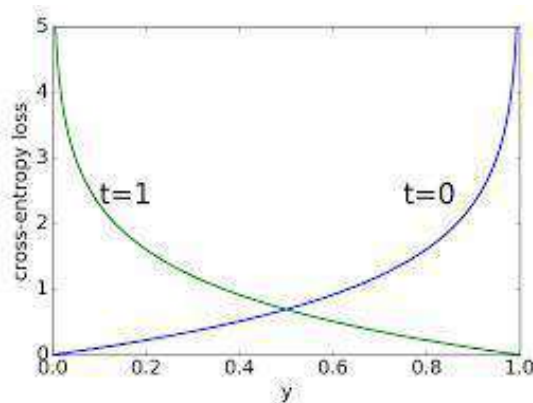


Рисунок 3. Зміни значення втрат при в залежності від значення що було визначено моделлю (y) та реальних значень (t - реальне значення належності до класу, синім кольором - немає приналежності до класу, зеленим - приналежність є)

Це працює лише у випадку бінарного індикатора класу. У випадку не бінарної класифікації результат може бути представлено як середнє значення по кожному з векторів класу. Метрику можна використовувати для задач с несиметричною кількістю об'єктів класів;

2. ROC AUC (Area Under the Receiver Operating Characteristic Curve) площа під кривою помилок, побудованій на різних порогах відсікання рисунок 4 [3]. Може використовуватись лише для визначення якості бінарного класифікатора. Головною особливістю є врахування кількості пар об'єктів одночасно класу що правильно класифіковані та які були відбраковані (тобто помилку першого та другого роду). Також за формою кривої можна визначити оптимальний рівень для прийняття рішень. Оскільки метрика враховує ефективність розділення класів моделлю то у випадку незначної кількості об'єктів класу буде давати хибний результат (точніше призведе до навчання моделі ігнорувати цей клас).

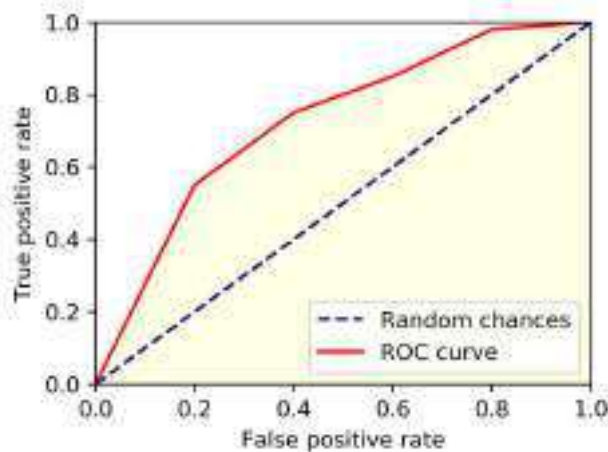


Рисунок 4. Типова РОК крива. Синім зазначено випадкові значення. Червоним - робота моделі: чим більше червона лінія відрізняється від діагоналі тим краща якість моделі. Аналогічно чим більша площа під графіком тим краще.

3. F-score. Метрика враховує кількість правильно класифікованих об'єктів до всіх класифікованих та кількість некласифікованих. Обидва критерії мають однакову вагу.

$$Fscore = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

де *precision* – відношення кількості правильно класифікованих об'єктів до кількості всіх класифікованих (рисунок 5),
recall – відношення кількості правильно класифікованих об'єктів до кількості всіх об'єктів (рисунок 5).

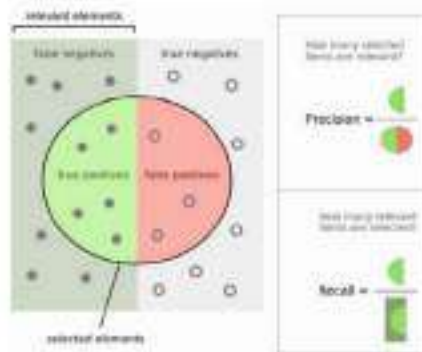


Рисунок 5. Схема до пояснення термінів

Однак при незначній кількості об'єктів класу є схильність до їх занулення. Метрику не можна використовувати для незначної кількості даних. Через несиметричність даних та вимогу в автоматизації було обрано бінарний перехресний критерій як якість моделі (точніше його мінімізація). Однак для оцінки моделей на старіння було вирішено використовувати РОК криву - так як вона має доволі просту інтерпретацію і гарно візуалізує якість моделей. Також може показати різні варіанти помилок моделі та порівняти їх із випадковими результатами.

Отже, в третьому розділі було розглянуто типові метрики для проведення оптимізації та типи моделей. Було докладно розглянуто причини та передумови обрання конкретних засобів.

4. РЕЗУЛЬТАТИ ЗАСТОСУВАННЯ МОДЕЛЕЙ

Поведінка моделі була досліджена на трьох відрізках:

1. Навчання базової моделі.
2. Тестування навчальної моделі та донавчання.
3. Тестування моделі отриманої після навчання на попередніх відрізках.

Для кожного активу використовувались моделі одного типу, що відрізняються лише процесом тренування. В процесі навчання точність моделі (крок динамічно змінюються, якщо точність почне коливатись і не буде зменшуватись).

Для первинної моделі було обрано інтервали у 10000 вимірів (через технічні обмеження системи) Зокрема нижче наведено навчання моделі на проміжку з 2020-10-10 по 2020-10-15.

Виміри відбуваються кожні 30 секунд. В векторі вимірювань присутні змінні що характеризують: об'єм торговельних операцій, ціна на 2 біржах, зміна ставки фінансування для деривативів. Для прийняття рішень використовувався інтервал що відповідає максимальному торговельному вікні що може спостерігати трейдер - 6 годин (або 720 вимірів). В якості умови прийняття рішень було обрано підвищення ціни активу на 2% і не зниження її менш ніж на 1%.

Нижче наведено лог навчання моделі:

```
start=2020-10-10 16:55:14.182461 end=2020-10-15 11:55:00.786176 long:70 short:137
train leng 15873, test leng 1500
long:0 short:33
Epoch 1/10
249/249 - 156s - loss: 0.0465 - val_loss: 0.0636
Epoch 2/10
249/249 - 154s - loss: 0.0377 - val_loss: 0.0650 Epoch 3/10
249/249 - 156s - loss: 0.0330 - val_loss: 0.0637 Epoch 4/10
249/249 - 156s - loss: 0.0302 - val_loss: 0.0625 Epoch 5/10
249/249 - 155s - loss: 0.0273 - val_loss: 0.0621 Epoch 6/10
Epoch 00006: ReduceLROnPlateau reducing learning rate to 1.0000001111620805e-07.
249/249 - 156s - loss: 0.0255 - val_loss: 0.0611 Epoch 7/10
249/249 - 155s - loss: 0.0246 - val_loss: 0.0612 Epoch 8/10
249/249 - 155s - loss: 0.0248 - val_loss: 0.0608 Epoch 9/10
249/249 - 155s - loss: 0.0240 - val_loss: 0.0614 Epoch 10/10
Epoch 00010: ReduceLROnPlateau reducing learning rate to 1.000000082740371e-08.
249/249 - 156s - loss: 0.0245 - val_loss: 0.0608.
```

З логів програми видно, що модель показує доволі непогані результати. Однак потрібно врахувати, що у валідаційній вибірці відсутні представники одного з класів, а отже можливе несиметричне навчання. Однією з переваг обраної моделі є можливість донавчання моделі в процесі її використання. Таким чином можна вирішити проблему старіння моделі без перенавчання по всій вибірці. Так через деякий час моделі починають псуватись.

Для ілюстрації цього факту було обрано ще один 5 денний проміжок часу через 6 днів після навчання моделі. Навчання моделей було проведено на доволі тихому ринку, були відсутні тенденції зростання. Очікувано, що після зміни тенденцій ринку моделі показують гірші результати або будуть потребувати перенавчання.

Однак по невдачах можна теж судити про внутрішні якості моделі та потенціал моделі для подальшого навчання. Зокрема, можна побачити доволі непогані результати, якщо взяти обернену модель (оскільки модель де-факто бінарний класифікатор, то достатньо змінити клас). Це свідчить про недостатнє навчання моделі (потрібно збільшити кількість епох).

Також це свідчить про те що дані та модель підібрані правильно і є потенціал для продовження дослідження.

Отже в розділі було розглянуто основні особливості обраної моделі. Було виявлено тенденції до старіння моделі. Навіть у випадку старіння модель на поданому активі дала задовільні результати точності. Це дозволяє зробити висновки що:

1. Поставлена проблема може бути розв'язана за допомогою обраних засобів.
2. Обрана модель дозволяє отримати непогані результати, однак наявне старіння моделі і потрібне донавчання моделі.
3. Модель може працювати без донавчання протягом 1-3 днів. На 5 день ефективність роботи моделі значно зменшується.
4. Модель та кількість даних було обрано правильно і подальший розвиток СППР на її основі мають потенціал до подальшого розвитку.
5. На 25% криптовалютах було досягнуто комерційний рівень точності на тестових даних.

5. ВИСНОВКИ

Отже, у поданій роботі розглянуто варіант трейдингової системи прийняття рішень що базується на класичному підході оперуванням з активами. Було розглянуто модель нейронної мережі, на базі якої приймалось рішення на

ринку. Було доведено можливість її застосування до поданої задачі. Також результати засвідчили її ефективність для окремих пар навіть після значного проміжку часу, а отже з постійним донавчанням розглянута система може бути цілком автоматичною. В роботі було досліджено періоди донавчання. Також було проведено тестування роботи моделі на реальних даних, що переконливо свідчать про правильний вибір типу моделі для даної проблеми, однак для деяких валютних пар необхідно провести додаткове донавчання моделі (а саме збільшити кількість епох навчання).

Успішність всіх попередніх кроків свідчить про правильно обраний напрям роботи, а саме застосування моделей класифікації для ринків з метою прогнозування напряму торговельних операцій інших трейдерів (уникаючи прогнозування власне значень часового ряду, прогнозуючи лише тенденцію до росту чи до падіння). Даний підхід є унікальним, однак застосування його, на мою думку, обмежене лише сферами де ціноутворення залежить більшою мірою від очікувань користувачів (інших трейдерів), що і є характерним для ринку криптовалют.

Наведено повний лог програми навчання моделі, дані що використовувались для навчання їх візуалізацію та ефективність моделі на цих проміжках. Дослідження охоплювало як стабільний ринок (з умовно незмінним трендом) так і швидко зростаючий. Така зміна умов мала б жахливі наслідки для моделей, що сфокусовані на прогнозуванні ціни. Однак з погляду прийняття рішень трейдером це має незначний ефект, що і було показано моделлю на прикладі ринку EOS. Однак для інших ринків спостерігалась значна зміна поведінки, що вплинуло на модель, зокрема обернене рішення моделі було більш наближеним до розміченого – це свідчить про недостатня навченість моделі та існування глобальних патернів поведінки користувачів.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Романенко В.Д., Милявский Ю.Л., Канцедал Г.О. Стабилизация неустойчивого курса криптовалюты на основе модального управления импульсным процессом когнитивной карты. XXVI Міжнародна конференція з автоматичного керування. Автоматика 2020: матеріали 26 наукової конференції, м. Київ, 13-15 жовтня 2020р. С. 75-76.

2. Бінанс відкриває гривневий шлюз через Advcash. URL: <https://www.binance.com/ru/support/articles/360035986971> (дата звернення: 22.11.2020).
3. Nielsen. M. Neural Networks and Deep Learning: free online book, 2019. chapter 3. URL: <https://static.latexstudio.net/article/2018/0912/neuralnetworksanddeeplearning.pdf> (дата звернення: 22.11.2020).
4. Fawcett T. An introduction to ROC analysis: Pattern Recognition Letters / USA, 2006, pages 861-874.
5. Xiao-Yuan Jing, Fei Wu, Dong Xiwei, Baowen Xu. An Improved SDA Based Defect Prediction Framework for Both Within-Project and Cross-Project ClassImbalance Problems. IEEE Transactions on Software Engineering. 2016. P. 43. DOI: <http://dx.doi.org/10.1109/TSE.2016.2597849>.
6. Heorhii Kantsedal. USE OF DECISION TREES IN THE LIMITED CRYPTOCURRENCY MARKETS FOR ADAPTIVE DECISION MAKING. XX International Students R&D Conference 2019 “Science and Technology of the XXI century”: Kyiv, Ukraine, 22 October 2019. P. 72-73.
7. Канцедал Г.О. Градієнтний бустинг. Травень 2019. URL: https://www.researchgate.net/publication/335567928_gradientnij_busting (дата звернення: 22.11.2020).
8. Канцедал Г.О. Проміжні результати. Ризики при проведенні торговельних операцій в криптовалюти. 2020. DOI: <http://dx.doi.org/10.13140/RG.2.2.14023.85922>.

МОДЕЛІ І ПРОГНОЗИ ВОЛАТИЛЬНОСТІ ФІНАНСОВИХ ПРОЦЕСІВ З ВИКОРИСТАННЯМ ЙМОВІРНІСНО-СТАТИСТИЧНИХ МЕТОДІВ

Морозов Р.Д.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹morozov.roman@lil.kpi.ua

Практично всі процеси, у будь-якій сфері діяльності – економічні, соціальні, екологічні, технічні, фінансові потрібно розглядати як нелінійні та нестационарні. Моделі фінансового ринку являються типовим представником нелінійних нестационарних процесів, адже вони, як правило, нестабільні, тобто мають досить складну динаміку процесу. Відповідно, для того щоб врахувати всі параметри, які впливають на розвиток нелінійного нестационарного процесу, тривіальних моделей буде недостатньо, оскільки отриманий результат буде некоректним. Тому потрібно використовувати нові, складніші (адекватні) моделі для того щоб забезпечити високу якість прогнозу.

Ключові слова: нелінійні процеси, нестационарні процеси, волатильність, мережа байєса, нейронна мережа.

1. ВСТУП

Моделювання цін акцій є основою фінансової економіки і, зокрема, теорії управління інвестиційним портфелем і оцінки фінансових інструментів.

Прогнозування і моделювання волатильності – важливе питання дослідження фінансових ринків. Велика емпірична робота була пророблена для поліпшення моделей волатильності, оскільки більш точні прогнози приводять до кращого ціноутворення фінансових активів і кращого управління ризиками. З іншого боку, волатильність фондового ринку інтенсивно вивчалася в останні три десятиліття, і було пророблено великий обсяг емпіричної роботи.

Статистично-ймовірнісні методи є одними з найпопулярніших для прогнозування розвитку економічних процесів, а також об'ємів торгових операцій, об'ємів виробництва, формування бюджетів підприємств та держави, оцінювання альтернативних економічних стратегій, прогнозування та менеджменту економічних і фінансових ризиків. Сьогодні без економічних прогнозів не обійтись також для визначення цілей розвитку суспільства, обґрунтування напрямів розвитку економіки або передбачення наслідків прийнятих рішень.

2. ОПИС ОБРАНИХ МЕТОДІВ

Більшість процесів у сучасному нестабільному світі є нелінійними та нестационарними. Останнім часом виникає все більше нових, непередбачуваних процесів, що негативно впливають на діяльність суспільства.

У різних напрямках досліджень встановлено, що часові ряди, які описують поведінку економічних і фінансових змінних у сучасній економіці, показують нерегулярні закономірності, які часто характеризуються коливаннями, що передбачають різке зниження та збільшення деяких параметрів [1].

З тих пір як економісти почали працювати над цими питаннями, основна увага зосередилася на спробі зрозуміти причини коливання економічних змінних, а також природу та причини цієї нерегулярної поведінки. Потім особливий акцент було зроблено на з'ясуванні того, чи є нестабільність явищем, що виникає ендогенно в економічній системі, чи вона продиктована екзогенними стохастичними збуреннями. Ці дослідження спрямовані на те, щоб пояснити ринкову динаміку, і мають на меті показати, що поява хаосу узгоджується з поведінкою оптимізуючих агентів, широко поширених у процесі прийняття рішень як фірмами, так і споживачами. Ці розробки природним чином привели до побудови нелінійних динамічних моделей [2].

Причини виникнення волатильності – перехідні процеси в економіці, суспільних формаціях, внутрішні та зовнішні військові конфлікти (втрати), глобальні та локальні фінансові кризи, недостатньо кваліфікований менеджмент на всіх рівнях управління економікою, застарілі технології у промисловості та сільському господарстві, зміна режимів функціонування соціально-економічних систем (наприклад, закінчення перехідного процесу і перехід до усталеного), природні катаклізми (вулкани, повені, потепління, ...).

Останнім часом мережі Байеса почали набирати популярність серед способів моделювання процесів різної природи та стали окремим напрямом в області інтелектуального аналізу даних. Мережі Байеса широко застосовуються для обробки часових рядів (статистичні дані), та інших все можливих типів представлення даних, наприклад, експертні оцінки, лінгвістичні змінні тощо [3].

Можливість врахування водночас кількісних та якісних показників, правильно використати надходження нової інформації у динаміці, використання явної залежності між існуючими факторами (що в свою чергу впливають на фінансові показники) є, мабуть, основною перевагою мережі Байеса. При використанні мережі Байеса часто виникають труднощі при знаходженні методу формування точного імовірнісного висновку базуючись на навчальній вибірці: для того щоб порахувати ймовірності вершин вузлів, потрібно використати матрицю спільних ймовірностей замість значень умовних ймовірностей емпіричних значень. Основною метою є те, що це дозволяє не розробляти структуру мережі заздалегідь. Також мережу просто реалізувати і при цьому досягти бажаної обчислювальної швидкості (це залежить від розміру вибірки) [4]; також не можна нехтувати складністю (чи навіть проблемою) вибору змінних для побудови моделі.

Нейронні мережі є одним із найпопулярніших методів (математичних моделей) вирішення задачі прогнозування часових рядів, вони включають в себе розвинену методологію структурного моделювання й методів навчання, що базуються на добре розвинутій теорії нелінійного програмування. Метою прогнозування є зменшення ризику при прийнятті рішень. Прогноз зазвичай виходить помилковим, але похибка залежить від прогнозуючої системи, яка використовується. Безумовно програмні реалізації розглянутих алгоритмів повинні мати дружній інтерфейс для користувача, дозволяти фахівцеві конкретної предметної області моделювати процес самостійно, без участі фахівця-математика.

Експоненційне згладжування є одним з методів прогнозування часових рядів для одновимірних даних. Цей метод створює прогнози, які є середньозваженими показниками минулих спостережень, де вагові показники старих спостережень експоненційно зменшуються. Форми експоненційного згладжування розширюють аналіз до даних моделі з тенденціями та сезонними компонентами [2].

Методи прогнозування експоненційного згладжування подібні тим, що вони прогнозують ряд на основі попередніх значень. Проте дані моделі використовують

експоненційно спадні вагові коефіцієнти. Це означає, що минулі значення впливають менше, ніж останні.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для порівняння оцінок двокрокового прогнозу використано ряд значень індексу NYSE Composite. Для прогнозування методом групового врахування аргументів (Рис. 1) було використано дворядний алгоритм із розміром навчальної вибірки 45 значень (60% від загальної кількості вимірів), за даними отриманої моделі відновлено 2975 значень.

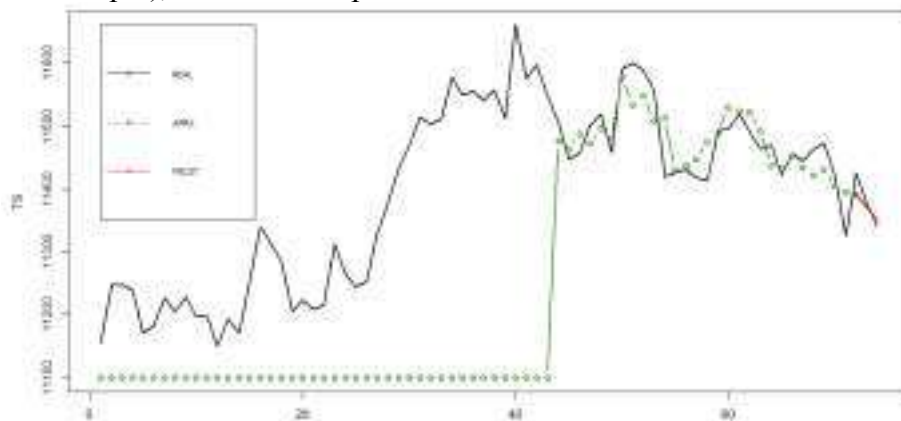


Рисунок 1. Двокроковий прогноз МГВА

На (Рис. 2, 3) зображено результати прогнозування методом експоненційного згладжування та за допомогою нейронної мережі відповідно.

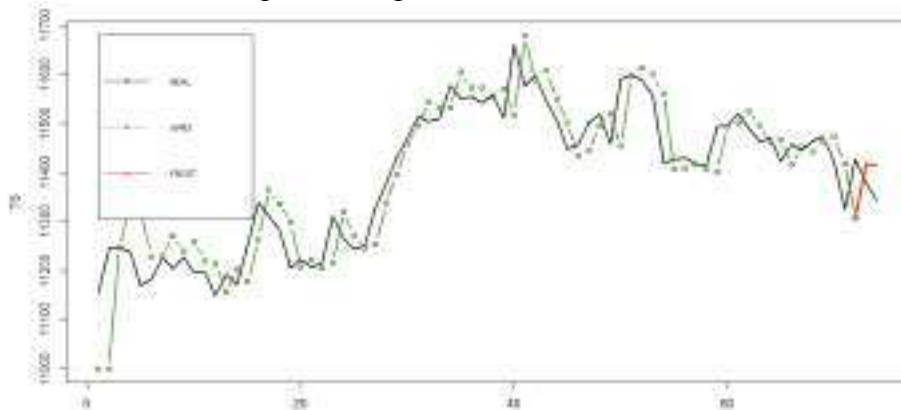


Рисунок 2. Двокроковий прогноз методом експоненційного згладжування

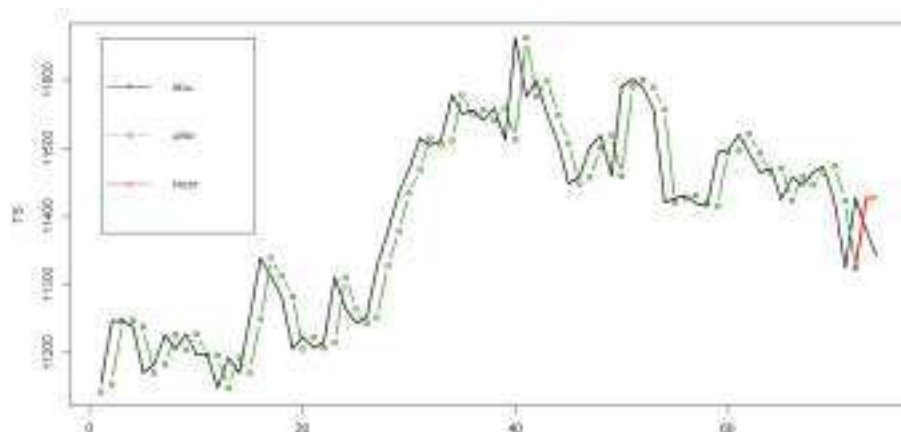


Рисунок 3. Двокроковий прогноз, отриманий за допомогою нейронної мережі

Результати експерименту подані нижче (Табл. 1). Як видно з таблиці, найкращі результати на ділянці прогнозування показав поліноміальний дворядний МГВА: прогнозоване значення відрізняється від реального лише на 0.7%, що є дуже хорошим результатом, при цьому, варто відмітити, що при відновленні даних метод показав значно гірші результати: СКП для відновленої моделі – 37,11; СКП для прогнозу – 8,57.

Таблиця 1. Результати прогнозування

NYSE Composite	модель			прогноз			
	СКП (RMSE)	МАП	КД	СКП (RMSE)	САПП (MAPE)	МіАПП	МАПП (MAE)
ЕЗ	56.740	73.14	0.86	58,84	4,97	3,49	6,45
МГВА	37.110	92.70	0.91	8,57	0,74	0,62	0,87
НМ	52.065	148.83	0.86	68,99	5,86	4,26	7,46

Діаграма зміни прогнозного, з використанням мережі Байєса, і реального відхилення індексу на останніх 50 часових інтервалах подана нижче (Рис. 4).



Рисунок 4. Відхилення індексу від значення на попередньому кроці

Після аналізу результатів бачимо, що у понад 90% випадків було точно передбачено зростання або падіння значення обраного індексу. У 86% випадків прогнозне і реальне значення відхилення індексу від значення в попередній момент часу мають відхилення менше 10%. І лише 15% експериментів мають відхилення більше 10% – така ситуація характерна для малих різниць та може бути пов’язана з варіантом дискретизації змінної.

Перш за все слід підкреслити, що метод експоненційного згладжування дає прогноз лише на 1 крок, тому його результати в цьому аналізі враховуватись не будуть. Експоненційне згладжування є найпростішим з розглянутих методів. Цей метод дає достатньо якісний для практичного використання прогноз на 1 крок за умови стаціонарності ряду або чіткої тенденції без суттєвих коливань. Тому область можливого використання метода досить обмежена. Прогнозування за допомогою нейронної мережі в цілому дало хороший результат. Проте при прогнозуванні на більшу кількість кроків якість суттєво погіршується. Це можна пояснити використанням простої мережі- з одним прихованим шаром. Метод групового врахування аргументів видає дуже якісні прогнози при короткостроковому прогнозуванні. Також зазначимо, що порівняно з усіма іншими методами МГВА значно програє в швидкості роботи, однак дозволяє отримати високоякісну модель при малій кількості вихідних даних, що дозволяє побудувати якісний прогноз.

При побудові мережі Байєса встановлено, що у 86% випадків прогнозне і реальне значення відхилення індексу від значення в попередній момент часу мають відхилення менше 10%. При цьому, при прогнозуванні на більшу кількість кроків якість прогнозу не погіршується. Для покращення результату варто звернути увагу на вибір кількості проміжків дискретизації та, за можливості, збільшити навчальну вибірку для мережі.

4. ВИСНОВКИ

При прогнозуванні фінансово-економічних показників необхідно раціонально обрати метод, який якісно прогнозує необхідний показник. При цьому слід також визначити параметри метода відповідно до значень ряду, що прогнозується. Таким чином найбільш "вузким" місцем є вибір методу та параметрів моделі в залежності від особливостей вхідних даних для знаходження найкращого варіанту. Перевагами методу експоненційного згладжування є швидкість обчислень та обґрунтованість результатів. Використання МГВА має перевагу у відсутності потреби створювати структуру моделі. Мережа Байєса допомагає описати процес шляхом статистичного опису ймовірностей та взаємодії між ними. Метод групового врахування аргументів дозволяє вирішувати широке коло задач, з якими звичайні регресійні моделі не справляються. Перевагою є якість прогнозу на довгих рядах без ресурсоємних обчислень. До недоліків можна віднести побудова нової моделі при зміні якогось параметра. Застосування нейронних мереж для прогнозування процесів в економіці, фінансах, екології, медицині вирішує багато актуальних задач. Перевагами такого підходу є незалежність дослідника від поведінки моделі, гнучкість і точність обчислень. До недоліків можна віднести те, що мережа сама приймає рішення і вивести явну модель поведінки ми не можемо.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кулявець В. О. Прогнозування соціально-економічних процесів. Київ: Кондор, 2009. – 194 с.
2. Зайченко Ю.П. Основи проектування інтелектуальних систем: навч. посіб. Київ: Слово, 2004. – 352 с.
3. Бідюк П.І., Терентьев О.М. та Коновалюк М.М. Байєсівські мережі в технологіях інтелектуального аналізу даних. Київ: Штучний інтелект, 2010. – 168 с.
4. Левченко Л.О. Огляд методів прогнозування фінансового стану підприємства на основі економетричних моделей / Л.О. Левченко, Д.С. Белова // Управління розвитком складних систем. - 2013. - Вип. 14. – С. 164-169.

СИМУЛЯЦІЯ МОНТЕ КАРЛО СИСТЕМ КЕРУВАННЯ ЗАПАСАМИ

Наконечна В.А.¹, Тимошук О.Л.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹nakonechnavlada@gmail.com,

²o.tymoshchuk@kpi.ua [0000-0003-1863-3095]

Запаси є невід’ємною частиною будь-якого підприємства, і вміння керувати ними безперечно грає велику роль в успіху та розвитку фірми, тому важливо розрізняти основні системи керування запасами, їх переваги та недоліки за тих чи інших обставин і параметрів. Метою роботи є аналіз систем керування запасами за допомогою імітаційного моделювання методом Монте Карло. Результатом дослідження є демонстрація переваг та недоліків найрозповсюдженіших систем керування запасами. У роботі було використано теоретичні та емпіричні методи дослідження.

Ключові слова: логістика запасів, керування запасами, система керування запасами, симуляція Монте Карло.

1. ВСТУП

Запаси різних видів продукції створюються для досягнення різноманітних цілей, які визначаються керівництвом промислової або торгової фірми (підприємством, організацією). До найбільш поширених цілей створення запасів належить підвищення ефективності підприємства, ефективне обслуговування споживачів, страхування перебоїв у постачанні, захист від підвищення цін закупівлі, економія на оптових знижках та транспортуванні [1].

Питання управління запасами займає важливе місце у системі управління поточними активами кожного підприємства. Воно є ключовою логістичною функцією, на долю якої, згідно з декількома джерелами, разом з транспортуванням приходиться від 60 до 80% загальних логістичних витрат [2]. За даними консалтингової компанії Deloitte, у 2020 році в умовах пандемії COVID-19, показник запасоемності в загальному по всіх індустріях США перевищив історичний максимум за останні 20 років. Аналогічний тренд простежується в усьому світі, у тому числі в Україні. [3] За останні два роки ситуація не покращилася, з повномасштабним вторгненням навряд чи залишилися підприємства, чії логістичні ланцюги не постраждали від нових перешкод. Але навіть незважаючи на це можна впевнено сказати що система управління запасами низького рівня з часом стає слабкістю підприємства – воно починає зазнавати збитків від поганого обслуговування через незадовільне накопичення важливих запасів і підвищені інвестиції, що зумовлене затриманням строку виконання замовлень.

2. СИСТЕМИ КЕРУВАННЯ ЗАПАСАМИ

Основним призначенням функції управління запасами є забезпечення збалансованості між обслуговуванням споживачів та інвестиціями в товарно-матеріальні запаси, особливо з точки зору концепції оптимізації запасів. Виділяють багато моделей керування запасами. Деякі джерела [4] розрізняють моделі за кількістю товару в системі, за постійністю таких параметрів як, наприклад, попит, за часом, за стратегією керування, за динамічністю

закупівлі, а інші виділяють ще більше категорій: за попитом, за поповненням запасів (за затримкою і за об'ємом), за витратами, обмеженнями та в залежності від стратегії керування. [1] З такого величезного розмаїття на практиці використовується лише обмежена кількість моделей - в основному ті, які надають відносно прості способи регулювання параметрів замовлення, поставок і рівня запасів на складі, а також не вимагають великих обсягів вихідної інформації і складних методів контролю.

Управління запасами передбачає організацію контролю їх фактичного стану. У теорії управління запасами було розроблено основні системи управління ними, з яких найпоширенішими є модель з фіксованим обсягом (названа також модель економічного розміру замовлення) і модель з фіксованим періодом (названа також періодичною моделлю, моделлю періодичного контролю) [5, 6]. Розглянемо ці системи детальніше.

Система управління з фіксованим розміром замовлення (Рис. 1). У даній системі розмір замовлення на поповнення запасу є постійною величиною і здійснюється за умови зменшення наявного на складах запасу до встановленого мінімального рівня, який називають "точкою замовлення". Залежно від інтенсивності споживання матеріальних ресурсів інтервали між поставками можуть бути різними, а величина замовлення зафіксована і не змінюється ні за яких умов роботи системи [5, 7]

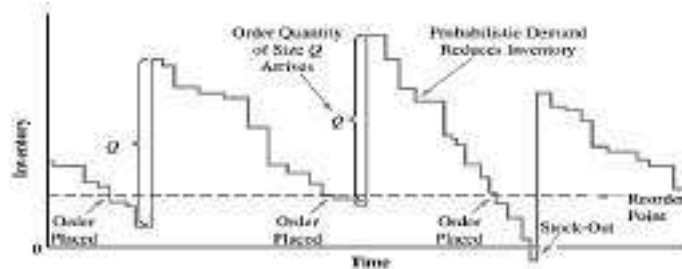


Рисунок 1. Шаблон інвентаризації для системи управління з фіксованим розміром замовлення з стохастичним попитом [8]

Система управління з фіксованим інтервалом часу між поставками (Рис. 2). У цій системі вхідним параметром є період часу між замовленнями. Вона зорієнтована на роботу за умови значних коливань споживання. Щоб запобігти завищенню обсягів запасів, які знаходяться на складі, або їхньому дефіциту, замовлення подаються не тільки у встановлені моменти часу, але і за умови досягнення запасом граничного рівня. Таким чином, рівень матеріального запасу регулюється як зверху, так і знизу. У тому випадку, якщо розмір запасу знижується до мінімального рівня раніше настання терміну подачі чергового замовлення, то робиться позачергове замовлення. В інший час система функціонує як система з фіксованою періодичністю замовлення. [5, 7]

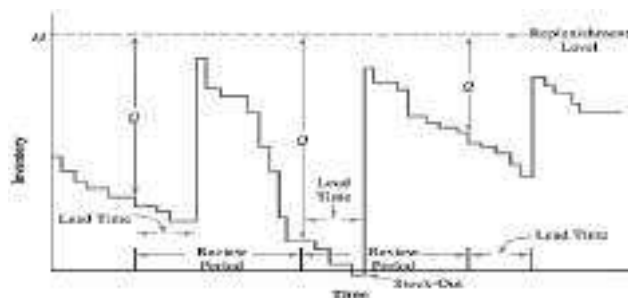


Рисунок 2. Шаблон інвентаризації для системи управління з фіксованим інтервалом часу між поставками з стохастичним попитом [8]

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

3.1 Вхідні дані

Для дослідження було сформовано вибірку з чотирьох продуктів. Параметри цих продуктів представлено у Табл 1.

Таблиця 1. Відомі параметри продуктів

№	Час виконання замовлення, дні	Об'єм, м3	Ціна закупівлі, грн	Ціна продажу, грн	Початковий запас
1	9	0.57	12	16.01	2750
2	6	0.052	7	8.6	22500
3	16	0.53	6	6	5200
4	22	1.05	37	37	1400

Попри те що змодельовати поведінку споживача досить важко, можна дати приблизну оцінку того, що клієнт має ймовірність p розмістити замовлення в будь-який день. Це p можна просто обчислити, поділивши кількість замовлень минулого року на кількість робочих днів.

Якщо у підприємця немає контракту з конкретним клієнтом виникає ще одна невизначеність – розмір замовлення. Для цього прикладу було зроблено припущення, що розмір замовлення відповідатиме логнормальному розподілу, параметри розподілу якого невідомі (що часто трапляється). Тому є важливим фіксувати історичні продажі продукту. Метою є максимізація очікуваного прибутку. За допомогою XYZ аналізу даних кожного товару можна побачити чітку різницю у попиті (Рис. 3). Наприклад, продукт 2 – це продукт великого обсягу, тобто його купують щодня ($p_2 = 1$), а середній розмір його замовлення становить 649. В порівнянні продукт 4 купують лише у 24% випадків, а його середній розмір замовлення становить близько 150.

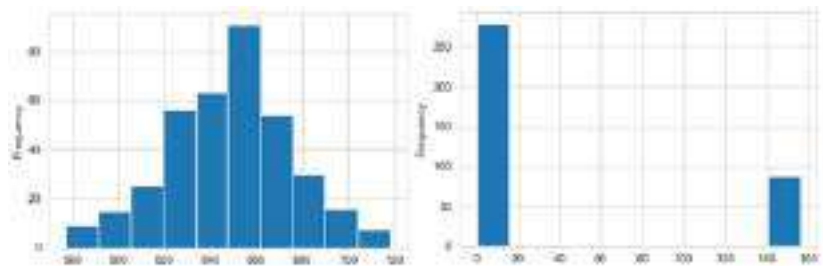


Рисунок 3. Графіки попиту продукту 2 (ліворуч) та 4 (праворуч)

У Табл. 2 надано короткий опис параметрів кожного продукту, які можна розрахувати виключно на основі минулих даних про продажі

Таблиця 2. Короткий опис параметрів кожного продукту

Продукт	1	2	3	4
Середнє значення	103.5	648.6	201.7	150
Стандартне відхилення	37.32	26.45	31.08	3.22
Очікувана пропорція днів	0.76	1	0.7	0.24
Очікуваний попит (Під час виконання замовлення)	705	3891.3	2265.8	784.8
Стандартне відхилення (Під час виконання замовлення)	165	64.8	383.3	299.7

Для моделювання поведінки попиту та розрахунку прибутку для однієї реалізації було проведено моделювання за методом Монте-Карло. Під час моделювання алгоритм повторює кожен день, намагаючись визначити рівень запасів продукту. Це необхідно для створення попиту на продукт на цей день. Це моделювання проводилося 10000 разів, щоб отримати кілька реалізацій прибутку та частки втрачених замовлень кожного разу.

3.2 Результати симуляції системи з фіксованими інтервалами часу між замовленнями

За результатами симуляції було побудовано гістограми та розраховано середнє значення і середнє відхилення прибутку та частки втрачених замовлень для конкретної точки замовлення (M). Так само цей експеримент можна виконати для діапазону значень M, щоб визначити значення, яке дає нам найбільший очікуваний прибуток. Як ви можете бачити на малюнку нижче для продукту 1, ми змодельовали результати для діапазону значень від 1000 до 3000. Це дало оптимальний очікуваний прибуток у розмірі 87 863 грн. Рисунок 4 нижче демонструє результати цих симуляцій для продукту 1 до точки 2071.

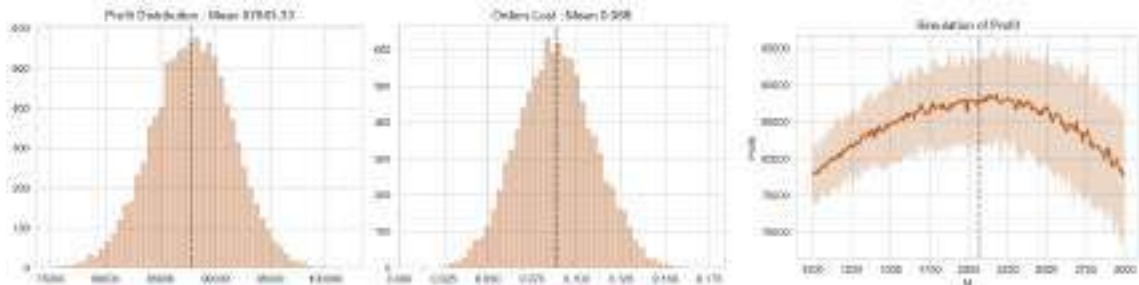


Рисунок 4. Гістограми для прибутку (ліворуч) та частки втрачених замовлень (посередині), графік моделювання для M (праворуч) для продукту 1

Використовуючи місячний інтервал перевірки, у таблиці 3 нижче наведено оптимальні точки замовлень, очікуваний річний прибуток і частку втрачених замовлень протягом року для кожного з продуктів.

Таблиця 3. Результати симуляції системи з фіксованим інтервалом замовлень

	1	2	3	4
Оптимальна точка замовлення	2 071	18 424	4 154	1 305
Очікуваний річний прибуток, грн	87 992	372 188	165 959	320 507
Очікуване стандартне відхилення, грн	5 035	1 454	6 561	29 864
Очікувана частка втрачених замовлень	0.077	0.01	0.05	0.074

3.3 Результати симуляції системи з фіксованими розміром замовленнями

За результатами симуляції було побудовано гістограми та розраховано середнє значення і середнє відхилення прибутку та частки втрачених замовлень для кількості замовлення 2002 і точки повторного замовлення 812 (Рис. 5) Виконуючи Grid Search між діапазоном значень для кожного продукту, можна розробити комбінацію кількості замовлення та точок повторного замовлення, які оптимізують функцію прибутку. Як можна бачити на рисунку 5 для продукту 1, значення були змодельовані для кількості замовлень від 1000 до 3000 і точки повторного замовлення між 500 і 1100. Схоже, що на графіку відображається опукла вгору функція, що означає максимальну точку прибутку.

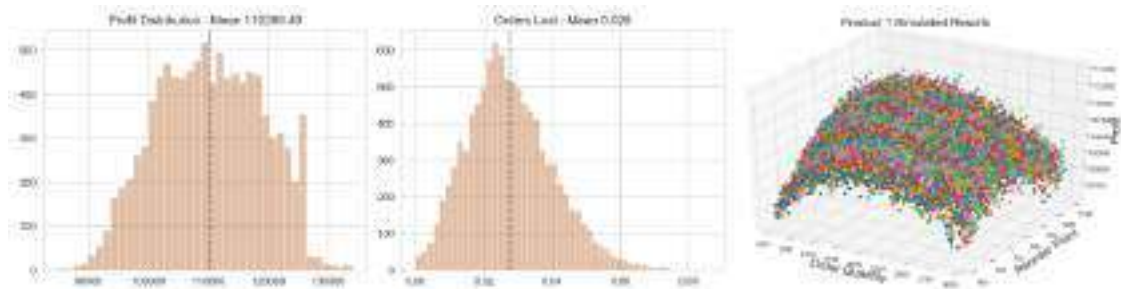


Рисунок 5. Гістограми для прибутку (ліворуч) та частки втрачених замовлень (посередині), результати Grid Search (праворуч) для продукту 1

Використовуючи цю політику, у таблиці 4 нижче наведено оптимальні точки повторного замовлення, оптимальну кількість замовлення, очікуваний річний прибуток і частку втрачених замовлень протягом року для кожного продукту.

Таблиця 4. Результати симуляції системи з фіксованим розміром замовлень

	1	2	3	4
Оптимальна точка замовлення	2002	21150	3318	1099
Оптимальна точка повторного замовлення	812	4036	2877	1221
Буферний запас	107	145	611	436
Очікуваний річний прибуток, грн	110,174	501,621	196,235	378,896
Очікуване стандартне відхилення, грн	8,814	25,860	10,439	40,724
Очікувана частка втрачених замовлень	0.028	0.01	0.004	0.019

3.4. Фінальний огляд результатів

В завершення огляду результатів складемо таблицю 5, в якій можна чітко помітити, що система з фіксованим розміром замовлення перевершує систему з фіксованим інтервалом з точки зору очікуваних прибутків для кожного продукту.

Таблиця 5. Порівняння прибутку між Системою з фіксованим інтервалом замовлення та Системою з фіксованим розміром

	Фіксований інтервал	Фіксований розмір
1	87,992.00	110,174.00
2	372,188.00	501,621.00
3	165,959.00	196,235.00
4	320,507.00	378,896.00
Разом	946 646.00	1,186,926.00

Загальний очікуваний прибуток від системи з фіксованим розміром замовлення на 25% вищий, ніж від системи з фіксованим інтервалом часу між замовленнями.

4. ВИСНОВКИ

Порівняння двох найрозповсюдженіших систем управління запасами призводить до висновку про наявність у них недоліків та переваг.

Система з фіксованим розміром замовлення потребує безперервного обліку поточного запасу складі. Це призводить до підвищення витрат. Однак максимальний бажаний запас у цій системі завжди буде меншим, ніж у системі з фіксованим інтервалом часу між замовленнями. Це призводить до економії витрат утримання запасів складі з допомогою

скорочення площ, займаних запасами, а також забезпечує гнучкість щодо часу розміщення замовлення. Оскільки в системі з фіксованим розміром є дві важливі змінні тоді, коли в періодичній системі є лише одна, простір рішень для системи з постійним розміром потенційно буде більшим.

Система з фіксованим інтервалом часу між замовленнями потребує лише періодичного контролю рівня запасу. Це призводить до підвищення рівня максимального бажаного запасу в порівнянні з системою з фіксованим розміром замовлення, але скорочує витрати у зв'язку з використанням системи. Крім того вона дозволяє підприємству краще прогнозувати замовлення, зроблені протягом певного періоду часу.

Якщо вибір системи полягає суто в максимізації очікуваного загального прибутку, за результатами дослідження можна зробити висновок що система з фіксованим розміром замовлення є найкращим вибором з двох.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Корпоративная логистика. 300 ответов на вопросы профессионалов / В. І. Сергіїв та ін. ; ред. В. І. Сергіїв. Москва :ИНФРА-М, 2005. 975 с. URL: <https://www.zagorskaya.info/wp-content/uploads/2016/09/Сергеев-логистика-300-ответов.pdf> (дата звернення: 25.10.2022).

2. Тюнін Д. К. Эволюция развития систем управления запасами. Управление экономическими системами: электронный научный журнал. 2013. URL: <https://cyberleninka.ru/article/n/evolyutsiya-razvitiya-sistem-upravleniya-zapasami/viewer> (дата звернення: 26.06.2022).

3. Найпоширеніші помилки в управлінні запасами, які припускають більшість підприємств - KLS. KLS. URL: <https://kyivlogisticsschool.com/news/najposhirenishi-pomilki-v-upravlinni-zapasami/> (дата звернення: 19.11.2022).

4. Бродецький Г. Л. Управление запасами. Москва : Эксмо, 2007. 282 с.

5. Зінченко Є. В. Переваги та недоліки основних систем управління запасами підприємства / Є. В. Зінченко // Розвиток європейського простору очима молоді: економічні, соціальні та правові аспекти : матеріали Всеукраїнської науково-практичної конференції докторантів, молодих учених та студентів, м. Харків, 28 квітня 2017 року. – Х. : ФОП Панов А. М., 2017. – С. 538 – 543.

6. Ткаченко Н. М. Бухгалтерський фінансовий облік на підприємствах України. Київ : А.С.К., 2001. 784 с.

7. Управління запасами. Міжнародний університет бізнесу та нових технологій. URL: <http://www.studfiles.ru/preview/3000222/page:2/>(дата звернення: 25.10.2022).

8. An introduction to management science: quantitative approaches to decision making / D. R. Anderson та ін. 14-те вид. Cengage Learning, 2015. 912 с.

ІНТЕГРУВАННЯ ПУАССОНІВСЬКИХ ВИПАДКОВИХ МІР ТА ЗНАХОДЖЕННЯ ЇХ АСИМПТОТИК ТА ОЦІНКА ПАРАМЕТРІВ ПУАССОНІВСЬКИХ ВИПАДКОВИХ ПОЛІВ

Пилипенко А.Ю.¹, Горбенко Н.О.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

²sobolonia@gmail.com

Багато прикладних задач, які зустрічаються сьогодні застосовують моделі процесів Пуассона. Проте розробка багатьох важливих моделей і методів ще не завершена до кінця. Мета цієї роботи полягає в розв'язанні наступних задач для пуассонівських точкових випадкових мір: знаходженні максимальних значень, побудові консистентних оцінок параметрів.

Ключові слова: пуассонівські точкові випадкові міри, рекорди, статистика випадкових процесів.

Теорія екстремальних значень – це елегантна та математично захоплююча теорія, а також наука, який пронизує величезну різноманітність застосувань. Розглянемо такі приклади: станції моніторингу забруднення повітря, які розташовані в різних місцях міста, урядові постанови вимагають, щоб концентрація забруднення, виміряна на кожному місці, була нижче певного рівня, тому якраз ці максимальні значення нас і цікавлять; гірничодобувна компанія відбирає зразки керна в певному регіоні, продовження буріння відбуватиметься в напрямку максимального зосередження руди; в спортивних змаганнях нецікаві другі, треті місця, а цікавлять в першу чергу рекорди України, світові і т.п. [1].

Загальною рисою цих ситуацій є те, що найбільший інтерес представляють найбільші або найменші значення спостережень. Рішення в прикладних задачах іноді приймаються на основі поведінки екстремальних значень [2].

Сам процес Пуассона є одним із найважливіших стохастичних процесів, і тому його часто вважають першою математичною моделлю в багатьох випадках. Існує велика кількість літератури про застосування моделей процесів Пуассона в різних областях (астрономії, біології, аналізі зображень, медицині, фізиці, теорії надійності тощо). Водночас ідентифікація багатьох важливих пуассонівських моделей (як і загальна теорія оцінювання) ще недостатньо розроблена, і така спроба допомогла б покрити цю прогалину [3].

1. ПОСТАНОВКА ПРОБЛЕМИ

Мета роботи полягає в дослідженні пуассонівських випадкових мір, а саме: знаходження асимптотичної поведінки екстремальних значень, побудові консистентних оцінок параметрів полів, тощо.

2. ПРИКЛАДИ ДОСЛІДЖЕННЯ

Наведемо приклад знаходження розподілу максимуму від випадкової кількості доданків.

Приклад 1.

Нехай ξ_k і η_k – випадкові незалежні величини.

$\{\xi_k\} \sim \text{Geom}(\alpha)$, а $\{\eta_k\} \sim \text{Geom}(\beta)$.

Знайти функцію розподілу $\max_{1 < k \leq \eta} \xi_k$.

Розв'язання

$$\begin{aligned}
 P \max_{1 < k \leq \eta} \xi_k \leq x &= \sum_{n=0}^{\infty} P \eta = n, \max_{1 < k \leq n} \xi_k \leq x = \sum_{n=0}^{\infty} P(\eta = n) P(\max_{1 < k \leq n} \xi_k \leq x) \\
 &= \sum_{n=0}^{\infty} P(\eta = n) F_{\xi}^n(x) \\
 P \eta = n &= q^{n-1} p, n \geq 1; \\
 P \xi = k &= \beta^{k-1} \alpha, k \geq 1.
 \end{aligned}$$

$$\begin{aligned}
 F_{\max_{1 < k \leq \eta} \xi_k}(x) &= P \max_{1 < k \leq \eta} \xi_k \leq x = \sum_{n=1}^{\infty} q^{n-1} p F_{\xi}^n(x) = \begin{matrix} \text{геометрична} \\ \text{прогресія} \end{matrix} = \frac{p F_{\xi}(x)}{1 - q F_{\xi}(x)} \\
 &= \frac{p(1 - \beta^x)}{1 - (1-p) 1 - \beta^x}.
 \end{aligned}$$

Відповідь: $F_{\max_{1 < k \leq \eta} \xi_k}(x) = \frac{p(1 - \beta^x)}{1 - (1-p) 1 - \beta^x}$.

3. ОСНОВНІ ЗАДАЧІ

Задача 1 (Про забруднення).

Пуассонівське випадкове поле на $[0; T] \times R$ з інтенсивністю $\frac{c}{|x|^\alpha} dt dx, \alpha > 1, t \in [0; T], x \in R$. Нехай t – моменти забруднення, x – величина забруднення.

Побудувати консистентну оцінку для c та α , тобто якусь функцію, яка прямує до c та α при $T \rightarrow \infty$.

Розв'язання

Для кращої візуалізації наведемо рисунок 1.

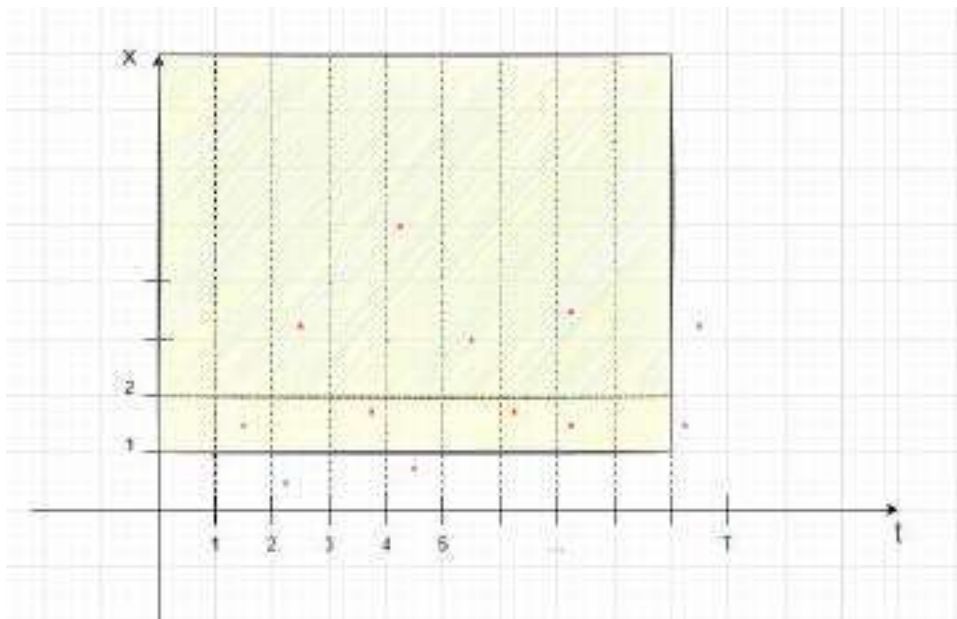


Рисунок 1. Ілюстрація розв'язку

Позначимо через ν_1 - інтенсивність на полосі $0; 1 \times [1; \infty)$ та ν_2 - інтенсивність на полосі $0; 1 \times [2; \infty)$ та знайдемо їх:

$$\nu_1 = c \int_0^1 \int_1^\infty \frac{1}{|x|^\alpha} dx dy = c \int_0^1 \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_1^\infty dx = \text{при } \alpha > 1 = c \int_0^1 \frac{1}{\alpha-1} dx = \frac{cx}{\alpha-1} \Big|_0^1 = \frac{c}{\alpha-1}.$$

$$\nu_2 = c \int_0^1 \int_2^\infty \frac{1}{|x|^\alpha} dx dy = c \int_0^1 \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_2^\infty dx = \text{при } \alpha > 1 = c \int_0^1 \frac{2^{-\alpha+1}}{\alpha-1} dx = \frac{c2^{-\alpha+1}x}{\alpha-1} \Big|_0^1 = \frac{c2^{-\alpha+1}}{\alpha-1}.$$

Нехай $E_1 = E_1(n)$ - це кількість точок, які потрапляють в полосу $0; 1 \times [1; \infty)$, $E_2 = E_2(n)$ - це кількість точок, які потрапляють в полосу $0; 1 \times [2; \infty)$.

$$\text{За законом великих чисел } \frac{E_1}{n} \rightarrow \frac{c}{\alpha-1}, \frac{E_2}{n} \rightarrow \frac{c2^{-\alpha+1}}{\alpha-1}.$$

$$\frac{E_2}{E_1} \rightarrow 2^{-\alpha+1},$$

$$\log_2 \frac{E_2}{E_1} \rightarrow -\alpha + 1,$$

$$\alpha^* \rightarrow 1 - \log_2 \frac{E_2}{E_1}, (1.1)$$

$$c^* \rightarrow \frac{(\alpha-1)E_1}{n}, \text{ де } \alpha \text{ беремо з (1.1).}$$

Відповідь: $\alpha^* \rightarrow 1 - \log_2 \frac{E_2}{E_1}, c^* \rightarrow \frac{E_1 \log_2 \frac{E_2}{E_1}}{n}$ - консистентні оцінки.

Задача 2.

Спостерігається Пуассонове випадкове поле на $[0; T] \times \mathbb{R}$ з щільністю

- 1) $\frac{1}{x^\alpha}, \alpha > 1,$
- 2) $e^{-|x|},$

криві $x = t^\beta, x = a \ln t.$

Яка кількість екстремальних точок скінченна чи нескінченна буде потрапляти в заштриховану зону кривої? Позначимо $X(T)$ - максимальна точка пуассонівського поля на відрізьку $[0, T]$. Знайти $\lim_{T \rightarrow \infty} \frac{X(T)}{x(T)}$.

Розв'язання

Для кращої візуалізації наведемо рис.2 та рис. 3:

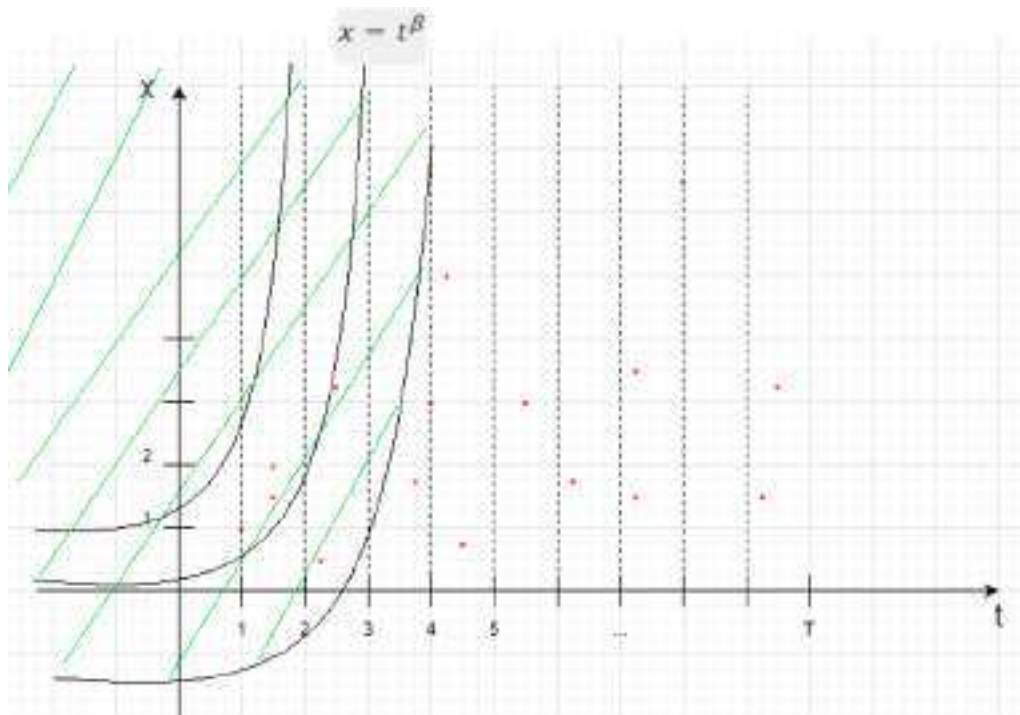


Рисунок 2. Ілюстрація для кривої $x = t^\beta$

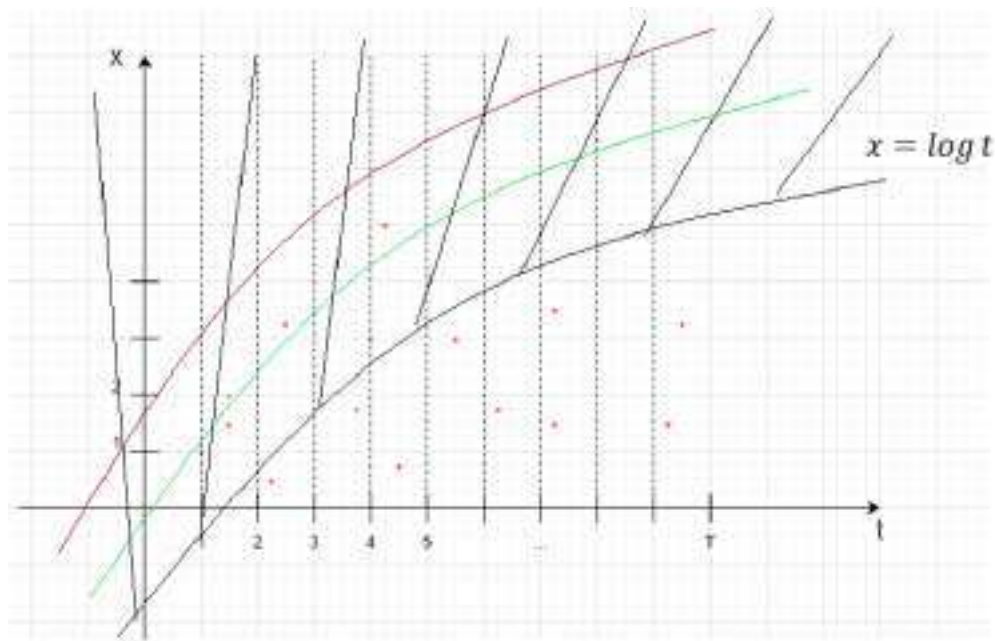


Рисунок 3. Ілюстрація для кривої $x = a \log t$.

Наведемо два розв'язки з чотирьох, де буде скінченна і нескінченна кількість точок.

$$1) \int_1^\infty \int_{a \ln t}^\infty e^{-|x|} dx dt = \int_1^\infty -e^{-x} \Big|_{a \ln t}^\infty dt = \int_1^\infty \frac{1}{t^a} dt = \frac{t^{a-1}}{a-1} \Big|_1^\infty = \frac{1}{1-a}, \quad \text{при}$$

$a < 1$ скінченна кількість точок лежить в цій зоні, отже max теж буде скінченна кількість точок і нескінченна при $a > 1$;

2)
$$\int_1^{\infty} \int_t^{\infty} \frac{1}{x^\alpha} dx dt = \int_1^{\infty} \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_t^{\infty} dt = \int_1^{\infty} \frac{t^{(-\alpha+1)\beta}}{-\alpha+1} dt = \frac{t^{-\alpha+1} \beta+1}{(-\alpha+1)(-\alpha+1) \beta+1} \Big|_1^{\infty} = \frac{1}{(\alpha-1)(-\alpha+1) \beta+1}$$
, при $-\alpha\beta > \beta + 1$. Отже, скінченна кількість точок лежить в цій зоні, отже max теж буде скінченна кількість точок.

Відповідь: 1) при щільності $e^{-|x|}$ та кривій $\ln t$ скінченна кількість точок лежить вище від кривої та $\lim_{T \rightarrow \infty} \frac{X(T)}{\ln t} = 1$ при $a < 1$ і нескінченна кількість при $a > 1$; 2) при щільності $\frac{1}{x^\alpha}$ та кривій t^β скінченна кількість точок лежить вище від кривої для $\beta > \frac{1}{1+\alpha}$ та $\lim_{T \rightarrow \infty} \frac{X(T)}{x(T)} = 0$, а для $\beta < \frac{1}{1+\alpha}$ – нескінченна кількість точок $\lim_{T \rightarrow \infty} \frac{X(T)}{x(T)} = \infty$.

4. ВИСНОВКИ

Наведені методи досліджень найбільших значень є дуже важливими, адже ця галузь не є дослідженою до кінця. Саме ця тема застосовується дуже часто для забезпечення високоякісних результатів, щоб застосувати їх в багатьох сферах людської діяльності, наприклад для дослідження рівня забруднень, як було наведено вище, а це може бути дуже корисним для збереження нашої екології.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Yu. A. Kutoyants *Statistical Inference for Spatial Poisson Processes*. – 233 Spring Street, New York: Springer, 1998, 278 p.
2. Sidney I. Resnick *Extreme Values, Regular Variation, and Point Processes* – 233 Spring Street, New York: Springer, 2007, 320 p.
3. Sidney I. Resnick *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling* – , 233 Spring Street, New York: Springer, 2007, 404 p.

ДОСЛІДЖЕННЯ СТОХАСТИЧНИХ СИСТЕМ ЗА ЧАСТКОВИМ СПОСТЕРЕЖЕННЯМ

Пилипенко А.Ю.¹, Ковальчук О.О.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ pilipenko.ay@gmail.com [0000-0002-8786-1617],
² olelchik1770@gmail.com

Об'єктом дослідження є параметри, які описують популярність блогера. Ми припускаємо, що популярність блогера модулюється ланцюгом Маркова народження і загибелі. Популярність ми не можемо спостерігати, на відмінну від кількості лайків та дизлайків, що мають залежний від популярності пуассонівський розподіл. Результатом роботи є побудовані консистентні оцінки для параметрів популярності. Отримані результати можна використовувати при роботі з маркетингом впливу.

Ключові слова: ланцюги Маркова, умовні розподіли, консистентна оцінка, закон великих чисел.

1. ВСТУП

Кількість бізнесів постійно зростає, і усі вони конкурують за увагу людей. Маркетинг є однією з ключових організацій будь-якого бізнесу, і саме вони відповідальні за залучення нових клієнтів. Для цього вони використовують різноманітні джерела трафіку, але кожен з них, на жаль, не безмежний. Тому з появою соціальних мереж і популярністю блогерів, маркетологи побачили потенціал для створення нового типу реклами. Цей спосіб залучення клієнтів отримав назву маркетинг впливу.

Працюючи з маркетингом впливу маркетолог повинен обрати правильного блогера серед тисячі інших для розміщення реклами бренду. Зазвичай для цього маркетологи збирають певну статистику соціальних мереж блогера, а саме, кількість підписників, лайків, коментарів, охоплення публікації. Усі ці статистики є випадковими і постійно змінюються у часі, тому важко спрогнозувати, які ж будуть показники на момент виходу реклами бренду.

Для оцінки блогера ми вирішили досліджувати саме його популярність, і з якою інтенсивністю вона змінюється, спостерігаючи таку статистику як кількість лайків та дизлайків під останніми публікаціями у соціальних мережах.

2. ПОСТАНОВКА ЗАДАЧІ

Розглянемо наступну задачу: потрібно обрати найкращого блогера для реклами бренду. Кожен блогер має певну популярність – кількість людей, які знають даного блогера, зацікавлені у ньому і потенційно готові відвідати його сторінки у соціальних мережах. Чим більше людей відвідає соціальні мережі блогера, тим більше буде охоплення у реклами бренду.

Припустимо, що популярність блогера випадковим чином змінюється у часі. Наприклад, певний час популярність зростає, а потім спадає.

Дані про блогера, які ми можемо спостерігати, є кількість переглядів, лайків чи дизлайків минулих його публікацій. Маючи ці дані, потрібно прийняти рішення, чи купувати рекламу бренду у певного блогера.

Припустимо, що популярність блогера – це випадковий процес $\xi(t)$, що описується ланцюгом Маркова народження і загибелі з параметром p . Параметр p визначає, з якою інтенсивністю зростає популярність блогера, $q = 1 - p$ – інтенсивність спадання. Різниця лайків та дизлайків у момент t – це випадкова величина $\eta(t)$, що має розподіл Пуассона з випадковим параметром $\xi(t)$: $\eta \sim Pois(\xi(t))$. Якщо $\xi(t)$ набуває від’ємне значення $-n$, вважатимемо, що $\eta \sim -Pois(n)$ (якщо кількість дизлайків більша ніж кількість лайків). При додатніх ξ величина η відображає наскільки кількість лайків більша за кількість дизлайків.

Зауважимо, що пуассонівський розподіл випадкової величини η є цілком природнім явищем і відповідає граничній теоремі Пуассона, якщо кількість лайків і дизлайків не дуже багато. При великій кількості лайків, скоріше за все, η матиме нормальний розподіл: $\eta \sim \mathcal{N}(\xi(t), \xi(t))$.

Потрібно, спостерігаючи випадкові величини $\eta \sim Pois(\xi(t))$, оцінити популярність блогера $\xi(t)$, а саме параметр p , за яким популярність $\xi(t)$ збільшується у часі.

Розглянемо три різні варіанти розподілів популярності $\xi(t)$:

- a) Ланцюг Маркова народження і загибелі на $\{0, \dots, m\}$ – зі скінченною кількістю станів (Рис. 1);

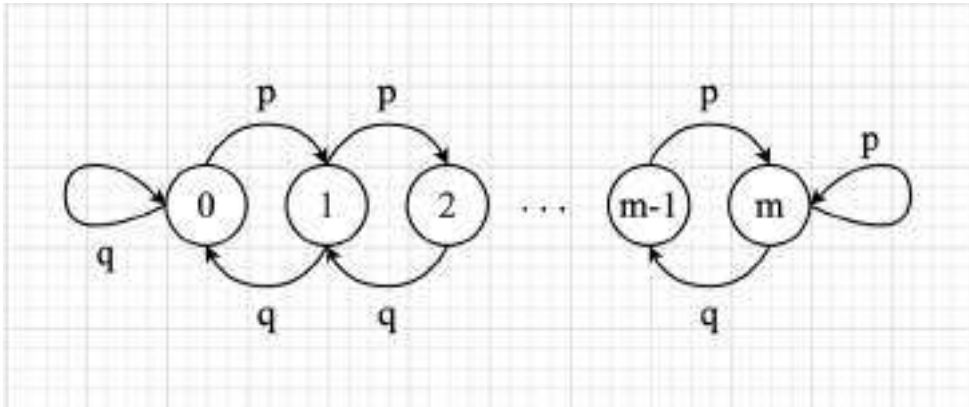


Рисунок 1. Ланцюг Маркова народження і загибелі на $\{0, \dots, m\}$

- b) Ланцюг Маркова народження і загибелі на $\{0, \dots, +\infty\}$ – з необмеженою кількістю станів, якщо $p < \frac{1}{2}$ (Рис. 2);
- c) Ланцюг Маркова народження і загибелі на $\{-\infty, \dots, +\infty\}$ – з необмеженою кількістю станів, з притяганням до середнього значення популярності $\alpha = 0$ (Рис. 3).

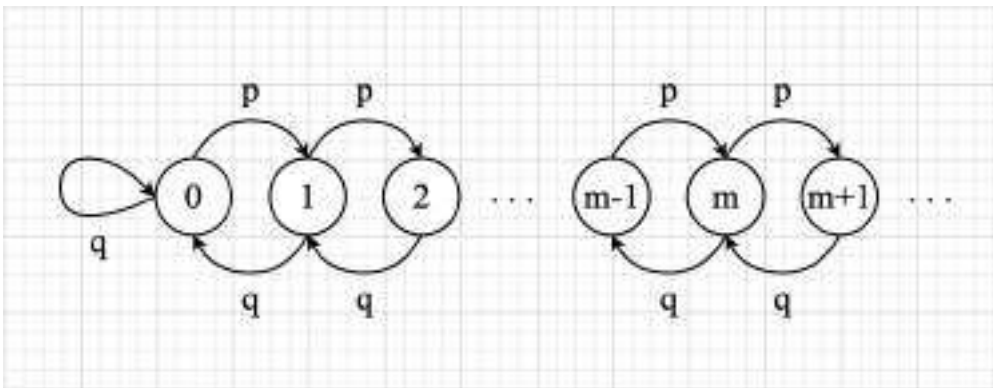


Рисунок 2. Ланцюг Маркова народження і загибелі на $\{0, \dots, +\infty\}$

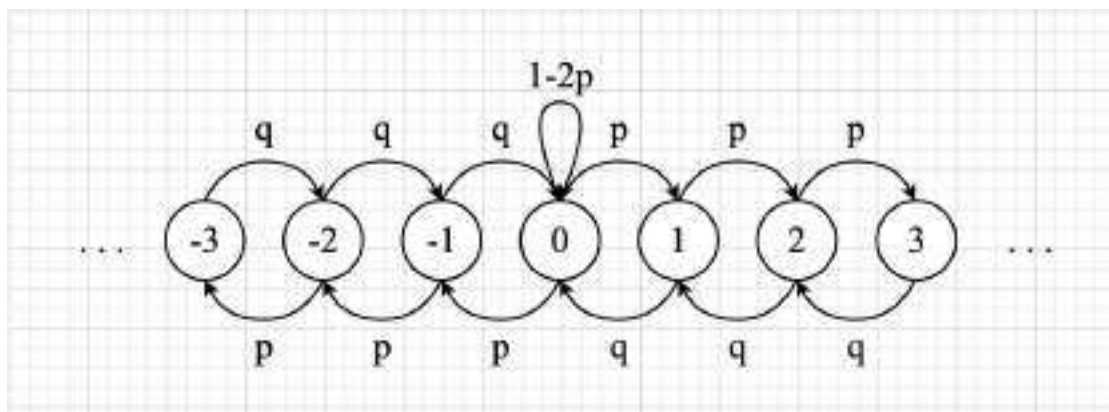


Рисунок 3. Ланцюг Маркова народження і загибелі на $\{-\infty, \dots, +\infty\}$ – з притягненням до значення популярності $\alpha = 0$

3. РОЗВ'ЯЗОК ЗАДАЧІ

Було проведено дослідження випадкової величини $\eta \sim Pois \xi(t)$ за допомогою умовних розподілів. Для цього нехай у деякий момент t $\xi(t) = i$. Тоді умовний розподіл:

$$P \eta = k \mid \xi = i = \frac{e^{-i} i^k}{k!}.$$

Виразили розподіл η через формулу повної ймовірності:

$$P \eta = k = \sum_i P \xi = i P \eta = k \mid \xi = i\}.$$

Розписали математичне очікування η :

$$\begin{aligned} E_{\eta} &= \sum_k k \cdot P \eta = k = \sum_k k \cdot \sum_i P \xi = i P \eta = k \mid \xi = i\} = \\ &= \sum_i P \xi = i \cdot \sum_k k \cdot P \eta = k \mid \xi = i\} = \sum_i \pi_i \cdot E_{\eta \mid \xi=i}, \end{aligned}$$

де:

π_i – стаціонарний розподіл Ланцюга Маркова, що відповідає розподілу $\xi(t)$,

$E_{\eta \mid \xi=i}$ – умовне математичне сподівання η .

Знайшли, що $E_{\eta \mid \xi=i} = E_{\eta \mid \eta \sim Pois i} = i$.

Отже, $E_{\eta} = \sum_i \pi_i \cdot i$

Дослідили збіжність середнього арифметичного випадкових величин η . Зауважимо, що ми не можемо застосувати закон великих чисел, оскільки η_k розподілені неоднаково.

Нехай кількість випадкових величин η , що ми спостерігаємо, рівна n :

$$\frac{\sum_k \eta_k}{n} = \frac{\sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{n} = \frac{\sum_i \sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{n}.$$

Нехай N_i – кількість разів відвідування стану i за n кроків:

$$\frac{\sum_i \sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{n} = \frac{\sum_i \frac{N_i \sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{N_i} \cdot N_i}{n} = \sum_i \frac{N_i}{n} \frac{\sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{N_i}.$$

Для ланцюгів Маркова, у яких усі стани сполучні і період рівний 1, справедливе твердження $\frac{N_i}{n} \rightarrow \pi_i$ майже напевно (м.н.). Також за законом великих чисел $\frac{N_i \sum_k \eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{N_i} \rightarrow$

$E_{\eta \mid \xi=i}$ майже напевно. Отже,

$$\sum_{i=0}^{N_i} \frac{\eta_k \cdot \mathbb{I}_{\{\xi=i\}}}{N_i} \cdot \frac{N_i}{n} \cdot \mathbb{E}_{\eta} \xi=i \cdot \pi_i$$

Загальний результат:

$$\frac{k \eta_k}{n} \cdot \mathbb{E}_{\eta} \xi=i \cdot \pi_i = i \cdot \pi_i = \mathbb{E}_{\pi_i}(\xi)$$

Для розв'язання задачі знайшли стаціонарний розподіл ланцюга Маркова народження і загибелі, підставили його у рівняння $\mathbb{E}_{\eta} \xi=i \cdot \pi_i$ і виразили консистентну оцінку ρ для параметру ρ .

- а) Стаціонарний розподіл для ланцюга Маркова народження і загибелі на $\{0, \dots, m\}$ знаходиться з системи рівнянь:

$$\begin{aligned} \pi_1 &= \pi_0 \cdot \frac{p}{q} \\ \pi_2 &= \pi_0 \cdot \left(\frac{p}{q}\right)^2 \\ &\vdots \\ \pi_i &= \pi_0 \cdot \frac{p^i}{q^i} \\ \sum_{i=0}^m \pi_i &= 1 \end{aligned}$$

Після заміни $\frac{p}{q} = \beta$ і розв'язання системи, знайшли, що стаціонарний розподіл:

$$\pi_i = \frac{\beta^i - 1}{\beta^{m+1} - 1} \cdot \beta^i$$

Для знаходження оцінки ρ потрібно розв'язати рівняння відносно параметра ρ :

$$\mathbb{E}_{\pi_i} \xi = \sum_{i=0}^m i \cdot \pi_i = \frac{\beta-1}{\beta^{m+1}-1} \cdot \sum_{i=0}^m i \cdot \beta^i = \frac{\beta^{m+1} - m + 1 - \beta + \beta}{\beta-1 \cdot \beta^{m+1}-1}, \text{ де } \beta = \frac{p}{1-p}.$$

Це можна зробити за допомогою чисельних методів обчислення.

- б) Стаціонарний розподіл для ланцюга Маркова народження і загибелі на $\{0, \dots, +\infty\}$ – з необмеженою кількістю станів, при $\rho < \frac{1}{2}$ знаходиться з системи рівнянь:

$$\begin{aligned} \pi_1 &= \pi_0 \cdot \frac{p}{q} \\ \pi_2 &= \pi_0 \cdot \left(\frac{p}{q}\right)^2 \\ &\vdots \\ \pi_i &= \pi_0 \cdot \frac{p^i}{q^i} \\ \sum_{i=0}^{\infty} \pi_i &= 1 \end{aligned}$$

Після заміни $\frac{p}{q} = \beta$ і розв'язання системи, знайшли, що стаціонарний розподіл:

$$\pi_i = (1 - \beta) \cdot \beta^i$$

Для знаходження оцінки ρ для параметр ρ потрібно розв'язати рівняння:

$$\mathbb{E}_{\pi_i} \xi = \sum_{i=0}^{\infty} i \cdot \pi_i = 1 - \beta \sum_{i=0}^{\infty} i \cdot \beta^i = \frac{\beta}{(1 - \beta)}$$

Отримали, що $\beta = \frac{\mathbb{E}_{\pi_i} \xi}{1 + \mathbb{E}_{\pi_i} \xi}$. Нехай $\frac{\mathbb{E}_{\pi_i} \xi}{1 + \mathbb{E}_{\pi_i} \xi} = M$.

Тоді консистентна оцінка для p :

$$\rho = \frac{M}{1 + M}, \text{ де } M = \frac{\mathbb{E}_{\pi_i} \xi}{1 + \mathbb{E}_{\pi_i} \xi}.$$

с) Стационарний розподіл для ланцюг Маркова народження і загибелі на $\{-\infty, \dots, +\infty\}$ – з необмеженою кількістю станів, з притягненням до значення популярності $\alpha = 0$ знаходиться з системи рівнянь:

$$\begin{aligned} \pi_1 &= \pi_0 \cdot \frac{p}{q} \\ \pi_2 &= \pi_0 \cdot \left(\frac{p}{q}\right)^2 \\ &\vdots \\ \pi_i &= \pi_0 \cdot \frac{p^i}{q^i} \\ &\dots \\ \sum_{i=-\infty}^{+\infty} \pi_i &= 1 \end{aligned}$$

Після заміни $\frac{p}{q} = \beta$ і розв'язання системи, знайшли, що стационарний розподіл:

$$\pi_i = \frac{1 - \beta}{1 + \beta} \cdot \beta^i$$

Для знаходження оцінки ρ для параметр ρ потрібно розв'язати рівняння:

$$\mathbb{E}_{\pi_i} \xi = \sum_i |i| \cdot \pi_i = 2 \sum_{i=0}^{\infty} i \cdot \beta^i = \frac{2\beta}{(1 - \beta^2)}$$

Це квадратне рівняння, додатним коренем якого є $\beta = \frac{-2 - 2 \sqrt{1 + \mathbb{E}_{\pi_i} \xi^2}}{2\mathbb{E}_{\pi_i} \xi}$. Нехай $\frac{-2 - 2 \sqrt{1 + \mathbb{E}_{\pi_i} \xi^2}}{2\mathbb{E}_{\pi_i} \xi} = M$. Тоді консистентна оцінка для p :

$$\rho = \frac{M}{1 + M}, \text{ де } M = \frac{-2 - 2 \sqrt{1 + \mathbb{E}_{\pi_i} \xi^2}}{2\mathbb{E}_{\pi_i} \xi}.$$

4. ВИСНОВКИ

У роботі було проведено дослідження 3 видів ланцюгів Маркова народження та загибелі, які моделюють популярність блогера. Параметр p відображає, з якою інтенсивністю зростає популярність. Спостерігаючи такі випадкові величини як кількість лайків та дизлайків, різниця яких розподілена за залежним від популярності пуассонівським розподілом, було знайдено консистентні оцінки ρ для параметра p .

Для дослідження були використані знання з тем ланцюги Маркова народження та загибелі, стационарні розподіли, гранична теорема Пуассона, закон великих чисел та умовні розподіли.

Отриманні результати можна використовувати в маркетингу, а саме при роботі з маркетингом впливу.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gut A. An intermediate course course in probability, 2nd edn. NewYork: Springer New York, 2009. 319 с.
2. Гринь А.Г. Цепи Маркова: учеб. пособ. Омск: Омский гос. ун-т им. Ф.М. Достоевского, 2019. 42 с.
3. Феллер В. Введение в теорию вероятностей и её приложения: в 2-х т. / пер. с англ. Ю.В. Прохоров. Москва: Мир, 1967. Т. 2. 752 с.
4. Brémaud P. Recurrence of markov chains. *Discrete probability models and methods*. Cham, 2017. С. 145–183.
5. Турчин В. М. Теорія ймовірностей і математична статистика. Дніпропетровськ: ІМА-прес, 2014. 556 с.
6. Чжун К.Л. Однородные цепи Маркова / ред. С.Х. Сираждинов; пер. В.Ф. Колчин. Москва: Мир, 1964. 426 с.

ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ОПЕРАЦІЙ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Позняк Д.І.¹, Кузнєцова Н.В.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹pozniakdi@gmail.com,

²n.kuznietsova@kpi.ua [0000-0002-1662-1974]

Виникнення шахрайських операцій з платіжними картками є одним з небажаних сценаріїв використання платіжних систем, тому для їх своєчасного виявлення доцільно використовувати методи машинного навчання, одним із результатів застосування яких може стати інформаційна технологія з класифікації транзакцій. Метою роботи є дослідження і розробка методів виявлення шахрайських операцій. Результатом дослідження є інформаційна технологія з виявлення шахрайств. У роботі було використано теоретичні та емпіричні методи дослідження.

Ключові слова: шахрайські операції, машинне навчання, інформаційна технологія, налаштування гіперпараметрів.

1. ВСТУП

Шахрайство - це заволодіння чужим майном або придбання права на майно шляхом обману чи зловживання довірою. Із розповсюдженням електронних грошових переказів та оплати товарів банківськими картками, на жаль, зростає також і кількість шахрайських операцій у фінансовій сфері [1]. Своєчасне виявлення та протидія шахрайським операціям є значущим питанням серед сучасних надавачів платіжних послуг по всьому світу, адже це один із ключових аспектів підтримки роботи систем гарантування платіжних операцій.

Зважаючи на це, актуальним є дослідження транзакцій у платіжних системах для виявлення факторів впливу на появу шахрайських операцій, моделювання систем виявлення шахрайств, і відповідно зниження ризику проведення шахрайських операцій у платіжних системах. При цьому, є важливим врахування усіх особливостей вихідних даних під час їх аналізу та ретельної підготовки їх до використання у алгоритмах машинного навчання. Цього можна досягнути шляхом комбінування відомих технік аналізу даних із емпіричними дослідженнями. Використання даних підходів дозволяє збільшувати прогностну здатність моделей, що будуються, а отже - створювати ефективні інформаційні технології для виявлення шахрайських операцій.

2. ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Основною задачею дослідження є розробка інформаційної технології для виявлення шахрайських транзакцій. Дане питання містить в собі ряд підзадач, серед яких: розвідковий аналіз даних, підготовка даних до навчання моделей, навчання моделей наряду з налаштуванням їх гіперпараметрів та аналіз результатів моделювання. Варто зазначити, що виявлення шахрайських транзакцій у масиві даних методами машинного навчання являє собою задачу бінарної класифікації. Тобто приймаючи на вхід вибірку з записів-транзакцій, результатом роботи моделі буде спрогнозоване значення змінної-маркера шахрайства

(назвемо її isFraud) для кожної транзакції. При цьому змінна isFraud є бінарною - такою, що приймає значення з множини $\{0,1\}$.

3. ОГЛЯД ОСНОВНИХ МЕТОДІВ ДОСЛІДЖЕННЯ

Виявлення шахрайських операцій, будучи актуальною проблемою сучасної банківської та фінансової сфери, за останній час було досліджено у багатьох роботах. Одними із найбільш використовуваних методів у даному питанні є інтелектуальний аналіз даних, наряду із активним залученням автоматизації процесів [2]. Зокрема деякі дослідники демонструють ефективність методів машинного навчання з учителем [3].

Навчання з учителем має на увазі алгоритми, що навчаються на спостереженнях типу “стимул-відповідь” таким чином, щоб потім визначати “відповідь” для “стимулів”, що не входять до навчальної множини спостережень. Популярними алгоритмами навчання з учителем для вирішення задачі класифікації є випадковий ліс та варіації градієнтного бустингу. Дані алгоритми, також відомі як ансамблеві, дозволяють боротись із проблемою перенавчання моделей - явищем, за якого модель не здатна здійснювати надійні передбачення на загальних даних, крім тих, що використовувались для навчання моделі. Ідея роботи ансамблевих алгоритмів полягає у комбінуванні декількох алгоритмів машинного навчання, що прийнято називати базовими алгоритмами, або слабкими моделями. Так, бустинг та випадковий ліс в якості базових алгоритмів зазвичай використовують дерева рішень. Основною відмінністю бустингу від випадкового лісу є послідовне навчання слабких моделей таким чином, щоб вони виправляли помилки попередніх і підвищували якість усього ансамблю, в той час як у випадковому лісі дерева рішень навчаються незалежно, після чого класифікуючи об’єкт шляхом голосування.

Варто зазначити, що дані, на яких навчаються алгоритми виявлення шахрайських операцій, у різних дослідженнях на цю тему є досить схожими - містять схожі набори відомостей про транзакції [4]. Це пояснюється тим, що дані, зібрані під час операції з кредитною картою, мають відповідати міжнародним стандартам фінансової звітності. Наприклад, моделі для виявлення шахрайств з картками у банківських установах часто є чутливими до зміни такого параметру як збільшення кредитного ліміту а також ліміту оплати в інтернеті та дозвіл на платежі за кордоном [5]. На жаль, у наборах даних у відкритому доступі дана інформація зазвичай не представлена, оскільки це суперечило б договору про захист персональних даних у банківських установах. Для побудови більш точних моделей дослідники зазвичай використовують feature engineering - конструювання нових ознак на основі існуючого датасету. Так, наприклад, знаючи час транзакції можна створити додаткові категоріальні змінні, що вказуватимуть на належність транзакції до певного дня тижня, місяця, свята, тощо. Метою агрегації змінних є виявлення зв’язків між вихідними змінними, які може бути важко виявити іншим чином. Конструювання змінних потрібно проводити з обережністю, щоразу валідуючи роботу моделі. До feature engineering також відносять стандартизацію даних, label та one-hot encoding.

Щоб порівняти роботу навчених класифікаторів між собою, за прогнозами моделей на тестових вибірках можна обчислити різноманітні метрики. Одним із показників якості роботи моделі є криві ROC (Receiver operating characteristic), за якими можна оцінити відношення істинно позитивних прогнозів моделі до хибно позитивних [6]. На графіку за віссю абсцисс відкладається $FPR \left(\frac{\text{істинно позитивні}}{\text{хиб.негат.} + \text{іст.позт.}} \right)$, а за віссю ординат - $TPR \left(\frac{\text{хибно позитивні}}{\text{істинно негат.} + \text{хибно позит.}} \right)$. Крива ROC може слугувати надійним методом ідентифікації класифікаторів, що потенційно є оптимальними [7]. При цьому можна також обчислити

метрику AUC, яка власне є площею під кривою ROC (Area Under Curve). Показник AUC, що дорівнює одиниці, свідчатиме, що модель класифікує дані ідеально вірно.

Ефективна робота алгоритмів машинного навчання залежить зокрема від налаштування гіперпараметрів моделей. Розглянемо A - алгоритм з N гіперпараметрів, що належать простору $\theta = \theta_1 \times \theta_2 \times \dots \times \theta_h$, при цьому конкретний набір гіперпараметрів для A має наступний вигляд (1). Функція втрат $L(Y, f(X, \theta))$ дає можливість оцінити роботу алгоритму із заданим набором гіперпараметрів θ , навченого на вибірці X з цільовою змінною Y . Таким чином можна формалізувати задачу до пошуку θ^* (2).

$$\theta = \langle \theta_1, \theta_2, \dots, \theta_h \rangle, \theta_i \in \Theta_i \quad (1)$$

$$\theta^* = \arg \min L(Y, f(X, \theta)) \quad (2)$$

Дана процедура потребує залучення методів нелінійної оптимізації. Одним із широко використовуваних методів прямого пошуку є метод симплексного спуску, також відомий як метод Нелдора-Міда [8]. Сутність методу полягає у послідовному переміщенні та деформуванні (зовнішньому та внутрішньому скороченні, стягуванні) симплексу навколо точки екстремуму.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для дослідження було використано набір даних з архіву компанії-надавача платіжних послуг Vesta Corporation, доступний для завантаження з ресурсу Kaggle. Навчальний набір початково складався з 590540 рядків-транзакцій та 434 змінних-інформації по кожній транзакції.

Перш ніж навчати моделі, було проведено розвідковий аналіз даних. Завдяки йому зокрема було встановлено часовий проміжок, у який відбувались усі транзакції. Так, завдяки аналізу моделей пристроїв, що було використано для здійснення транзакцій та відкритих даних про купівельну активність від компанії Google, було прийнято рішення вести відлік транзакцій у датасеті, починаючи із 1 листопада 2017 року. Це в свою чергу дало можливість сконструювати нові змінні для датасету - день тижня та час транзакції.

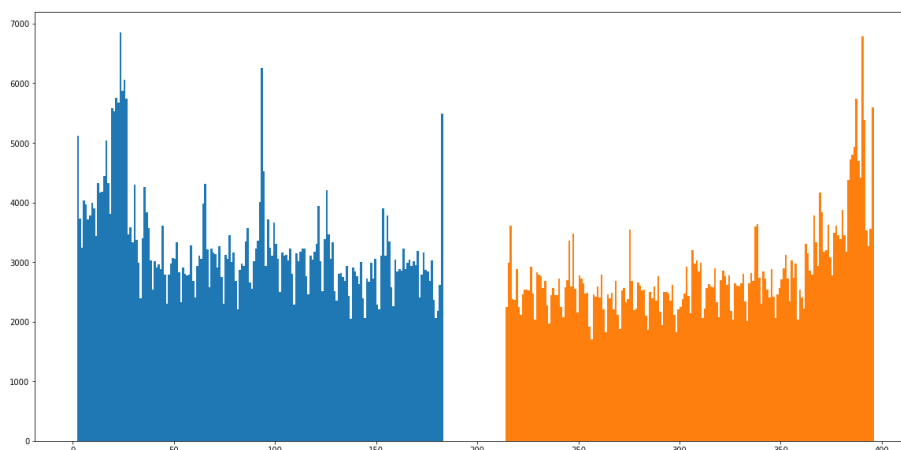


Рисунок 1. Розподіл транзакцій за днями у навчальній та тестовій вибірках

Розвідковий аналіз датасету також показав наявність мультиколінеарності серед ряду змінних. За допомогою їх групування на основі розподілу пропущених значень з вихідних 339 змінних з цією особливістю було отримано 128, майже на 60% менше. Враховуючи великий обсяг даних, крім можливості подолати проблему незначущості параметрів моделі, це позитивно вплинуло також і на швидкість навчання алгоритмів.

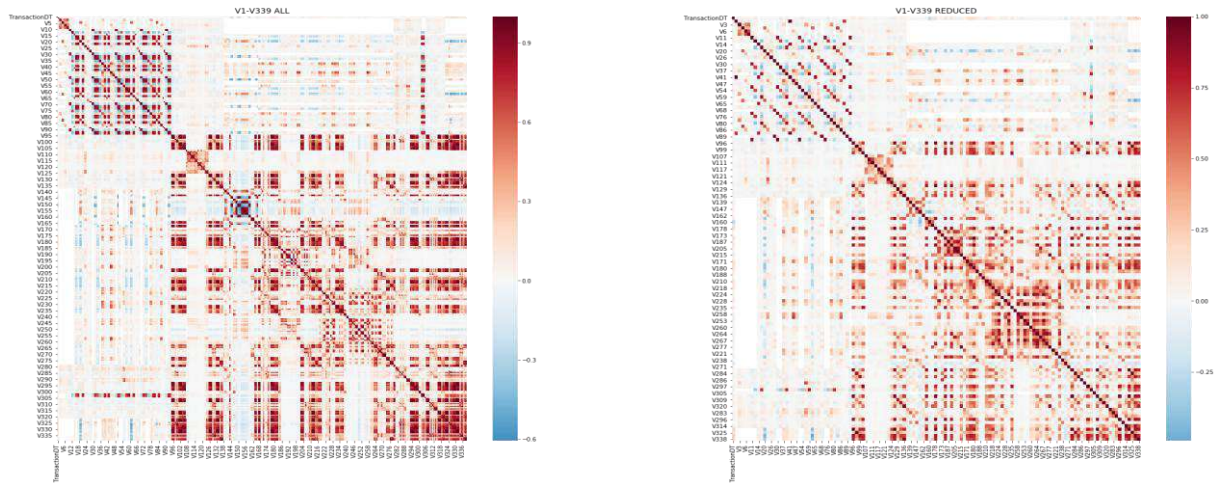


Рисунок 2. Кореляційні матриці вихідної і згрупованої вибірок

Також перш ніж навчати алгоритми, було заповнено пропуски у датасеті. Вибіркове середнє було використане для неперервних змінних, у той час як для категорійних було використано позначку NA як окрему категорію. Через виявлену незбалансованість класів у вибірці, для випадків шахрайств було використано метод генерації додаткових екземплярів мінорного класу SMOTE. Дані також було закодовано за допомогою label encoding, щоб категорійні змінні були представлені числовими значеннями. Під час навчання було застосовано стратегію 5-кратної крос-валідації. При внесенні змін у навчальні дані або модель, ефективність тих чи інших дій було оцінено за середнім значенням AUC по усіх фолдах крос-валідації.

У ході дослідження було навчено наступні моделі: випадковий ліс та модифікації градієнтного бустингу XGBClassifier та LGBM. Для градієнтного бустингу було виконано налаштування гіперпараметрів за допомогою пошуку оптимальних параметрів на заданому просторі з використанням алгоритму Нелдора-Міда. Результати роботи моделей на тестовій вибірці подано у таблиці 1.

Таблиця 1. Результати роботи моделей

№	Алгоритм	AUC
1	Random Forest Classifier	0.887
2	XGBoost Classifier	0.897
3	LGBM	0.919

Як видно з таблиці, Light Gradient Boosting Machine найкраще працює на тестових даних з-поміж трьох розглянутих алгоритмів. У подальшому можна працювати над покращенням моделей, проводячи експерименти з конструювання змінних та, за можливості, працюючи над агрегацією транзакцій за користувачами, щоб таким чином розглядати саме поведінку користувача платіжною картою і знаходити аномалії у ній, а не у кожній окремій транзакції.

Використовуючи результати проведеного аналізу та моделювання, було створено інформаційну технологію з виявлення шахрайських транзакцій. Ідея роботи даної системи полягає в попередній обробці транзакції (послідовність дій, використану також і для навчання моделей), поданні її на вхід до попередньо навченої моделі та, як результат, її класифікації. Інформаційна технологія надає можливість обирати для класифікації

транзакцій одну із моделей: випадкового лісу, екстремального градієнтного бустингу або модифікацію стохастичного градієнтного бустингу LGBM, а також виконує порівняння результатів, отриманих під час застосування вищевказаних моделей. Провівши експериментальні обчислення у спроектованій СППР, можна не лише визначити розмір можливих втрат, а й зробити висновок про те, який з методів обчислення є найкращим для використання на практиці.

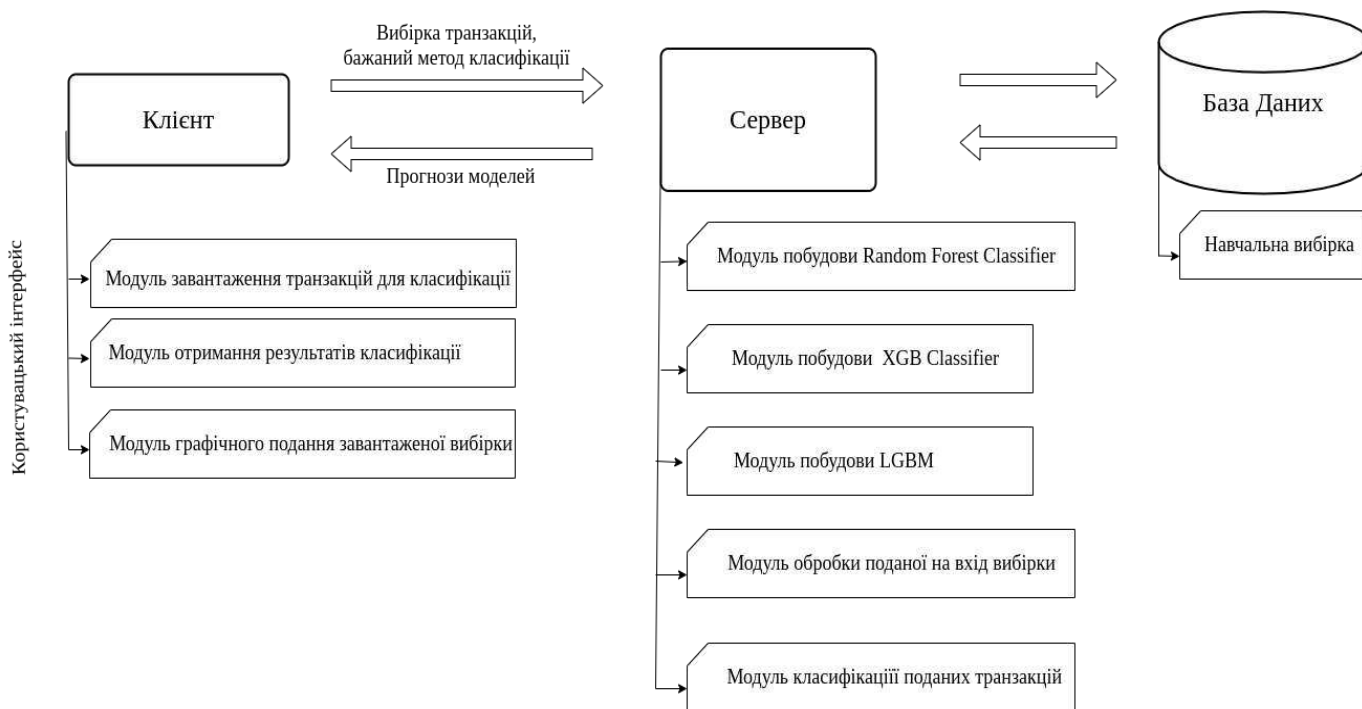


Рисунок 3. Клієнт-серверна архітектура інформаційної технології

5. ВИСНОВКИ

На сьогоднішній день шахрайські операції становлять загрозу безпеці платіжних систем, тому повсякчас виникає необхідність знаходження нових ефективних підходів до їх виявлення. Одним із перспективних підходів у вирішенні цього питання є інтелектуальний аналіз даних. Для ефективної класифікації транзакцій методами ІАД доцільним є створення інформаційної технології.

У ході дослідження було проведено розвідковий аналіз даних, підготовку вихідного датасету до навчання моделей - зокрема проведено feature engineering, а також label encoding та балансування класів. На основі підготованих даних було навчено ряд моделей машинного навчання, що порівнювались між собою за показником AUC. Таким чином, у результаті досліджень було створено інформаційну технологію з виявлення шахрайських операцій, у якій використано отримані у ході експериментів бінарні класифікатори.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Інтернет-представництво Національного банку України [Електронний ресурс]. Доступно: <https://bank.gov.ua/ua/news/all/startuye-informatsiy-na-kampaniya-natsionalnogo-banku-z-platijnoyi-bezpeki-shahraygudbay>. Дата звернення: Жов. 2022.
2. C. Phua, V. Lee, K. Smith, and R. Gayler “A Comprehensive Survey of Data Mining-based Fraud Detection Research”, School of Business Systems, Faculty of Information Technology, Monash University, Clayton, Australia, 2010 [Online]. Available: https://www.researchgate.net/publication/302557906_A_Comprehensive_Survey_of_Data_Mining-based_Fraud_Detection_Research. Accessed on: Oct 2022.

3. W.F. Yu, and N. Wang “Research on Credit Card Fraud Detection Model Based on Distance Sum”, in *International Joint Conference on Artificial Intelligence*, Hainan, China, 2009, pp. 353-356, DOI: 10.1109/JCAI.2009.146.

4. A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, “Feature engineering strategies for credit card fraud detection”, *Expert Systems with Applications*, vol. 51, pp. 134-142, June 2016, DOI: 10.1016/j.eswa.2015.12.030.

5. N. V. Kuznietsova “Scoring Technology for Risk Assessment of Fraud in Banking”, *Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2016)*, CEUR Workshop Proceedings, vol. 1813, pp. 54–61, 2016.

6. F. Provost., and T. Fawcett, “The case against accuracy estimation for comparing induction algorithms”, In *Proc. 15th Intl. Conf. On Machine Learning*, pp. 445–453, Madison, WI, USA, 1998.

7. F. Provost., and T. Fawcett, “Robust Classification for Imprecise Environments”, *Machine Learning* vol.42, pp. 203–231, 2001, DOI: 10.1023/A:1007601015854.

8. J.A. Nelder, and R. Mead, “A simplex method for function minimization” *The Computer Journal*, vol. 7, pp. 308–313, 1965.

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ АНАЛІЗУ РОЗВИТКУ ФІНАНСОВИХ ПРОЦЕСІВ

Радіо О.В.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹oliaradi007@gmail.com

В нинішній час фондовий ринок стає популярним інвестиційним майданчиком як для інституційних, так і для індивідуальних інвесторів. Волатильність ринку та інновації у фінансовій торгівлі зробили прийняття інвестиційних рішень все більш складним та ризикованим. У таких умовах системи підтримки прийняття рішень (СППР) грають важливу роль та сприяють підвищенню якості рішень. Метою роботи є аналіз фінансових процесів фондового ринку з метою покращення прийнятих рішень. Результатом дослідження є СППР, що виконує прогноз волатильності та оцінює ринковий ризик із застосуванням методології VAR. У роботі використані експериментальні дослідження та методи математичного моделювання.

Ключові слова: волатильність, прогнозування, математичні моделі, VAR, ринковий ризик.

1. ВСТУП

В нинішній час фондові ринки широко розповсюджені та в певній мірі впливають на макроекономічні показники будь-якої сфери чи в цілому країни та їх можна віднести до ефективних. Тобто фондовий ринок – це ринок, на якому активно конкурує велика кількість інвесторів, що прагнуть до максимізації прибутку компаній, кожна з яких намагається передбачити майбутню ринкову вартість окремих цінних паперів, і де важлива поточна інформація майже вільно доступна для всіх учасників. Іншими словами, на ефективному ринку будь-якої миті часу фактична ціна цінного паперу буде гарною оцінкою її внутрішньої вартості. Денний оборот фондового ринку налічується мільярдами доларів та мільйонами торгових доручень. З кожним роком учасників ринку стає дедалі більше, серед них фінансові організації, хеджфонди, банки та приватні інвестори. Трейдери використовують різні стилі торгівлі та системи прийняття рішень, кожен стиль чи система мають різну економічну ефективність та ступінь ризику.

Однак, доступність надійних даних фінансового ринку та потужних програмних інструментів є найвищою. У таких умовах системи підтримки прийняття рішень (СППР) грають найважливішу роль і підвищують якість рішень, прийнятих як професійними, так і індивідуальними інвесторами. Отже, користувачам, які прагнуть підвищити ефективність ухвалення рішень на фондовому ринку, необхідне застосування СППР, що й пояснює актуальність обраного дослідження.

2. МАТЕМАТИЧНІ МОДЕЛІ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ПРОЦЕСІВ

Система підтримки прийняття рішень, або СППР, являє собою комп'ютерну інформаційну систему, яка організовує, збирає й аналізує дані, які можуть

використовуватися під час прийняття рішень для управління, експлуатації та планування [1]. Системи підтримки прийняття рішень призначені для того, щоб прискорено і високоякісно допомогти людям аналізувати великі обсяги даних для прийняття обґрунтованих рішень. Іншими словами, коли потрібно ухвалити рішення, можна звернутися до системи підтримки прийняття рішень і встановити, що «кажуть» дані про це рішення або вибір, який необхідно зробити [2,3]. СППР можуть створювати складні моделі або уявлення того, що відбувається з даними [4]. Приклади таких систем включають ручні системи, гібридні системи, різноманітні типи аналітики, а також складне програмне забезпечення для підтримки прийняття рішень. Фактором, який відрізняє новіші комп'ютерні системи від ранніх систем підтримки ухвалення рішень, є їхня здатність аналізувати надзвичайно великі набори даних, надаючи засновані на даних рекомендації, що виключають здогади під час ухвалення рішень.

Традиційний підхід до розробки СППР полягає в тому, щоб визначити математичну модель, встановити, як система приймає рішення, і використовувати програмне забезпечення для оптимізації, щоб визначити та порівняти результати різних сценаріїв. Цей метод ґрунтується на розпорядчій аналітиці і є надзвичайно ефективним.

Як зазначається в роботі [5], комп'ютеризація фондових ринків і доступність детальних електронних записів потоку заявок і динаміки цін на фондових ринках за останнє десятиліття вивільнила терабайти високочастотних даних про транзакції, потоки заявок і динаміку портфеля заявок на фондових ринках, які дають детальне уявлення про високочастотну динаміку попиту, пропозиції та ціни на цих ринках. Ці дані можуть бути використані для дослідження природи механізму ціноутворення, який описує, як ринкові ціни реагують на коливання попиту та пропозиції. На високому рівні "механізм ціноутворення" - це карта, яка відображає взаємозв'язок між ринковою ціною та такими змінними, як ціна, історія та потік замовлень [12]:

$$Price_{t + \Delta t} = F(Price\ history\ 0 \dots t, Order\ Flow\ 0 \dots t, Other\ Information) + F(X_t, \varepsilon_t) \quad (1)$$

де X_t – множина змінних стану (наприклад, лагові значення ціни, волатильності та потоку заявок), наділених певною динамікою; ε_t – випадковий "шум" або інноваційний член, що представляє надходження нової інформації та інші ефекти, які не повністю охоплюються змінними стану.

Ціна на фондовому ринку формується на основі рівноваги пропозиції та попиту, а також на основі залежності кількості угод від ціни, як показано на рис. 1.

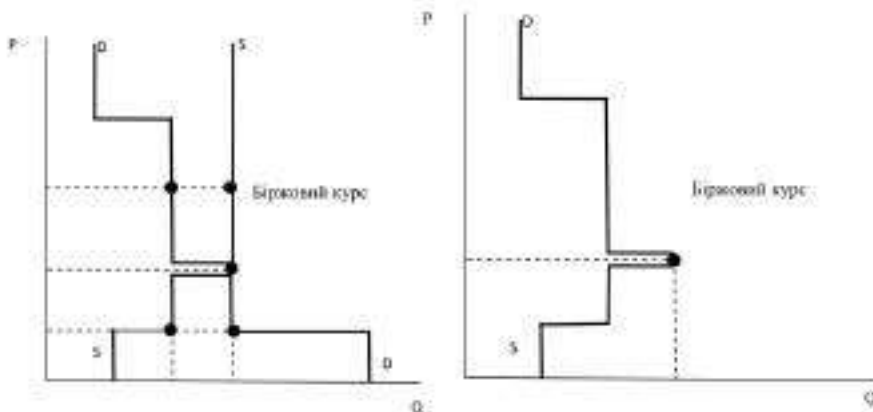


Рисунок 1. Формування ціни на фондовому ринку: а) на основі рівноваги пропозиції та попиту; б) на основі залежності кількості угод від ціни

Існує низка методів оцінювання ціноутворення на фондовому ринку, метою яких є побудова максимально адекватної математичної моделі вибраного процесу. Якщо розглядати прогнозування волатильності та оцінку ринкового ризику до таких моделей

можна віднести ARCH/GARCH моделі. Загалом, ARCH/GARCH методологію можна охарактеризувати як методологію моделювання дисперсії досліджуваного показника. Оскільки дисперсія – момент другого порядку, модель дисперсії нелінійна, отже, її не можна оцінювати методами, розробленими для лінійних моделей, таких як ARIMA. Також під час досліджень велику увагу привернула нейронна мережа LSTM, саме своїми результатами. Основними компонентами мережі LSTM є рівень введення послідовності та рівень LSTM. Шар введення послідовності вводить дані послідовності або часового ряду в мережу. Шар LSTM вивчає довгострокові залежності між тимчасовими кроками даних послідовності.

Великої популярності на сьогоднішній день набула методологія Value at Risk (VaR). Це зумовлено тим, що така методологія володіє рядом безсумнівних переваг, а саме:

- дає змогу оцінити ризик в межах можливих втрат, співвіднесених з ймовірністю їх виникнення;
- вимірювати ризики на різних ринках універсальним чином;
- агрегувати ризики за окремими позиціями в єдину величину (інтегральний ризик) для всього портфеля активів, враховуючи при цьому інформацію про кількість позицій, і волатильності на ринку.

Волатильність або абсолютна/відсоткова дисперсія цін часто вважається гарним заходом для оцінювання ринкового ризику. Професійні аналітики також схильні використовувати такі методи, як моделювання вартості під ризиком (VaR) для виявлення потенційних збитків за допомогою статистичного управління ризиками.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

У процесі виконання роботи розроблено і реалізовано систему підтримки прийняття рішень, яка призначена для аналізу ціноутворення на фондовому ринку. Вона виконує такі задачі: побудова моделей гетероскедастичних процесів для прогнозування волатильності, оцінювання ринкових ризиків методом VaR та аналіз якості отриманої прогнозованої моделі. Головною метою створення даної програми було підвищення рівня ефективності прийняття рішень на фондовому ринку, а також проілюструвати роботу методів моделювання фінансових процесів, застосувати їх на реальних даних.

Для моделювання у даній роботі обрано процес ціноутворення акцій міжнародної компанії TESLA, взятий з інтернет ресурсу *finance.yahoo.com*. Обраний датасет складається з таких колонок: «High», «Low», «Open», «Close», «Volume», «Adj Close» та містить 1259 записів, а саме з 13.11.2017 – 11.11.2022. Для подальшої роботи потрібно обрати характеристику, яка буде найбільш коректною для моделювання. Після дослідження значення кожної колонки було визначено взяти саме колонку «Adj Close», адже вона показує скориговану ціну закриття, тобто ціну закриття після всіх застосованих коригувань. Двома основними з них є розбиття та дивіденди. Після візуального аналізу графіків бачимо, що розподіл цін схожий на нормальний, проте являється трохи вищим, а от розподіл Стюдента краще описує ряд, особливо на хвостах розподілу (Рисунок 2, 3).

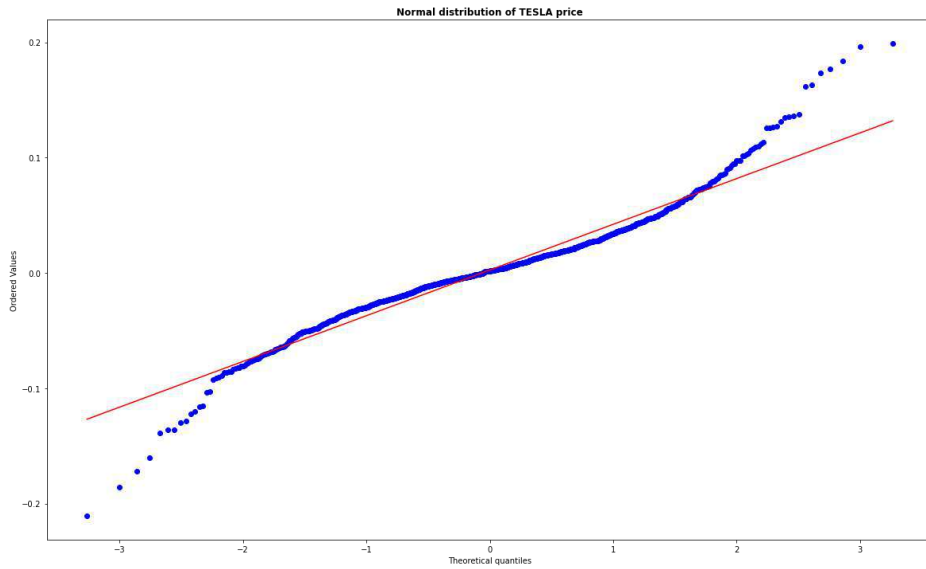


Рисунок 2. Графік нормального розподілу даних компанії TESLA

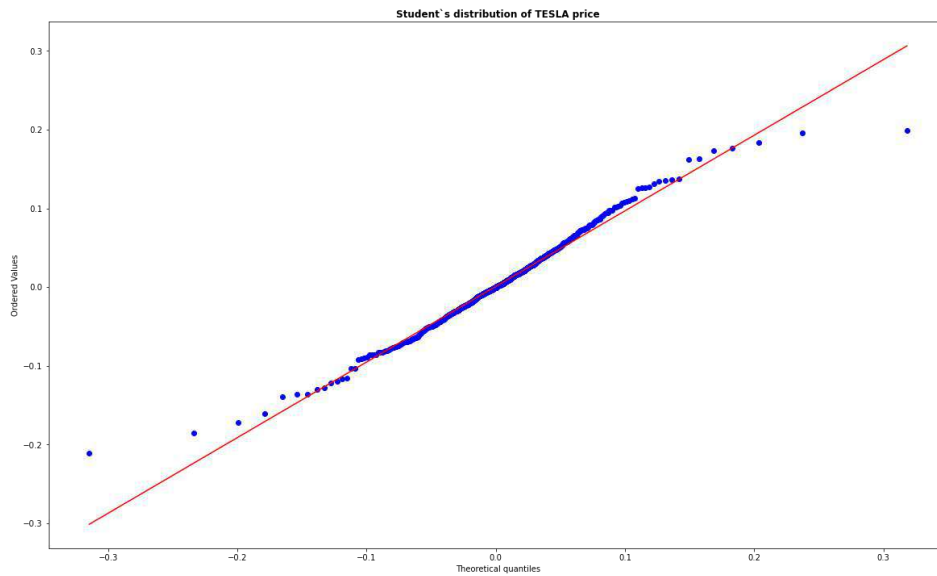


Рисунок 3. Графік розподілу Стьюдента компанії TESLA

Для побудови відповідних математичних моделей було проведено підготовку даних. Спочатку розбиваємо вибірку на тестову й тренувальну, для очистки ряду від випадкових шумів застосовуємо фільтр Калмана, а також проведено тест на гетероскедастичність. Можна звернути увагу, що ймовірність, отримана в результаті, явно менша за поріг 0,05 (Рис. 4). Це свідчить про те, що у даному випадку необхідно відхилити нуль-гіпотезу, яка припускає незмінність дисперсії у часі. Отже для залишків даної моделі характерна наявність ефекту ARCH.

Dep. Variable:	Adj_Close	R-squared:	0.000	
Mean Model:	Constant Mean	Adj. R-squared:	0.000	
Vol Model:	GARCH	Log-Likelihood:	-5505.34	
Distributions:	Normal	AIC:	11018.7	
Method:	Maximum Likelihood	BIC:	11039.1	
Date:	Wed, Nov 16 2022	No. Observations:	1208	
Time:	14:05:47	Df Residuals:	1207	
	Mean Model	Df Model:	1	

	coef	std err	t	P> t

mu	21.0683	9.064e-02	232.432	0.000
				[20.891, 21.246]

Volatility Model				

	coef	std err	t	P> t

omega	0.1408	3.216e-02	4.377	1.201e-05
				[7.774e-02, 0.204]
alpha[1]	1.0000	1.051e-02	95.116	0.000
				[0.979, 1.021]
beta[1]	1.7790e-11	6.167e-06	2.885e-06	1.000
				[-1.209e-05, 1.209e-05]

Рисунок 4. Описова статистика та результати тесту Engle

Наступним кроком для моделювання волатильності дисперсії процесу було визначення порядку самої моделі. Для цього побудували графіки АКФ та ЧАКФ для квадратів залишків моделі та побудували декілька моделей для порівняння. Результати подано в таблиці 1.

Таблиця 1. Результати застосування GARCH моделі

Model	MAE	MAPE	THEIL
GARCH (1,1)	45.503398699012	11.60310685149867	0.25971130284802
GARCH (2,4)	43.702794762397	11.51956147319662	0.25971123179284
GARCH (4,8)	43.259252938935	11.73202941090763	0.25708675806683
GARCH (4,12)	42.703393537471	11.60310954523587	0.25711302848347

Отримані результати прогнозів в таблиці 1 виявилися задовільними. В процесі дослідження також перевірялися моделі EGARCH і TARARCH, які показали майже такі ж оцінки, тому було вирішено не додавати їх у дану роботу. Як найкращу модель було обрано саме GARCH (4,12).

Також у поданій магістерській роботі проведено дослідження щодо прогнозування часових рядів за допомогою рекурентних нейронних мереж. Для моделювання справжньої волатильності було обрано саме нейронну мережу LSTM, яка працює по принципу системи довгої короткочасної пам'яті. Представимо результати роботи даного методу в таблиці 2.

Таблиця 2. Результати роботи LSTM моделі

Model	MAE	MAPE	THEIL
LSTM	7.3834419570547	0.42678149293463	0.06897247272352

Останній крок дослідження – це розрахунок ринкового ризику методом Value-at-Risk. Для розрахунку використано прогнози, отримані саме за допомогою моделі LSTM. Згадуємо, що на початку розділу було зроблено дослідження про схильність до розподілу Стюдента, тобто розраховуємо показник саме для цього випадку. Результати подано в таблиці 3.

Таблиця 3. Порівняння результатів застосування моделей

Показник	Результат
95%	478.834
99%	435.057

4. ВИСНОВКИ

Враховуючи складність фінансового ринку та нелінійну характеристику фінансових змінних, штучна нейронна мережа, як важлива форма штучного інтелекту, стала великим досягненням у багатьох фінансових галузях та може застосовуватися для аналізу фінансових процесів на фондовому ринку.

Згідно з літературними джерелами найбільшу ефективність при прогнозуванні цін на фондовому ринку продемонстрували нейронні мережі LSTM, саме тому вони обрані для моделювання в межах магістерської дисертації. Окрім цього, для моделювання дисперсії при аналізі фінансових показників на фондовому ринку обрано методологію ARCH/GARCH, яка є найбільш ефективною серед аналогів.

У ході дослідження було реалізовано таку систему підтримки прийняття рішень, яка дозволяє виконувати моделювання волатильності певного процесу та розраховувати ринковий ризик. Ця система використовує моделі GARCH та LSTM, а також методологію VaR для обчислення ринкового ризику. Результати дослідження показали, що найбільш ефективним методом виявилася саме нейронна мережа LSTM. Окрім того, було визначено значення показника VaR для 95% і 99% ймовірностей.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Конспект лекцій з курсу “Системи прийняття рішень” для студентів напряму підготовки 6.030502 “Економічна кібернетика”, спеціальності 051 “Економіка” / С. В. Гринчуцька – Тернопіль, ТНТУ імені І. Пулюя, 2017. – 130с.

2. Системи і методи підтримки прийняття рішень. [Електронний ресурс] навч. посіб. для здобувачів ступеня магістра за освітніми програмами «Системний аналіз та управління», «Системний аналіз фінансового ринку» спеціальності 124 «Системний аналіз» / П.І.Бідюк, О.Л.Тимошук, А.Є.Коваленко; Л.О.Коршевнік КПП ім. Ігоря Сікорського ;. Електронні текстові дані. Київ : КПП ім. Ігоря Сікорського, 2020. – 259 с.

3. Системи підтримки прийняття рішень [Текст]: навчальний посібник для самостійного вивчення дисципліни / [уклад.: С. М. Братушка, С. М. Новак, С. О. Хайлук] ; Державний вищий навчальний заклад “Українська академія банківської справи Національного банку України”. – Суми : ДВНЗ “УАБС НБУ”, 2010. – 265 с.

4. Проектування інформаційних систем підтримки прийняття рішень: навч. посіб. / [Бідюк П.І., Коршевнік Л.О., Коваленко А.Є., Гожий О.П.]. — К.: НТУУ «КПІ», 2013. – 608с.

5. Justin Sirignano, Rama Cont. Universal features of price formation in financial markets: perspectives from Deep Learning. 2018. – 21 p.

6. Башкіров О.В. Порівняльний аналіз VAR-методів оцінки ризику активів банку О.В. Башкіров, Проблеми і перспективи розвитку банківської системи України : зб. наук. праць ДВНЗ «УАБС НБУ». – Вип. 14. – С. 302–309.

7. What is a decision support system (DSS)? [Електронний ресурс]: Режим доступу: <https://www.techtarget.com/searchcio/definition/decision-support-system> (дата звернення: 15.10.2022) НТУУ «КПІ», 2010. – 310с.

8. Market Risk [Електронний ресурс]: Режим доступу: <https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/market-risk/> (дата звернення: 17.10.2022).

9. Long Short-Term Memory Networks [Електронний ресурс]: Режим доступу: <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html> (дата звернення: 19.10.2022).

МЕТОДИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СТАНУ ПІДПРИЄМСТВА ЗА ДОПОМОГОЮ РЕГРЕСІЙНИХ МОДЕЛЕЙ.

Сандига Я.С.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

На сьогоднішній день кількість підприємств збільшується кожного дня. Для аналізу релевантності та напрямів розвитку підприємства потрібно постійно розуміти ступінь ризику банкрутства самого підприємства. Основною проблематикою в даній сфері прогнозування є вибір методу, який буде використано для вказаних цілей. Далі в роботі буде розглянуто дещо нестандартний підхід до аналізу стану підприємства за допомогою логістичної регресії. Метою роботи визначено створення програмного забезпечення (ПЗ), за допомогою якого користувач може подати на вхід початкові дані фінансової звітності і отримати оцінку ризику банкрутства підприємства.

Ключові слова: фінансовий стан підприємства, логістична регресія, прогнозування ризиків банкрутства.

1. ВСТУП

Перетворення фінансового аналізу в статистичний починається з розрахунків великої кількості показників, що з часом не відображають актуальність проблеми (такі як коефіцієнти трансформації та іммобілізації). Зазвичай, аналіз фінансового стану підприємства зводиться до оцінювання одного коефіцієнта, який на думку користувача даних є найбільш впливовим в конкретній постановці задачі. Якщо ж перейти до правильної побудови зв'язків між економічними факторами та математичними методами, то можна зробити висновок, що в даному випадку нам не потрібна така велика кількість показників. Навпаки, основною метою побудови моделей аналізу стану підприємства є використання мінімальної кількості показників для максимально правильного розрахунку коефіцієнта, що характеризує стан підприємства. В залежності від значення цього коефіцієнту ми отримуємо відповідь на економічне запитання щодо того, наскільки близько підприємство знаходиться до банкрутства.

2. МЕТОДИ ОЦІНЮВАННЯ ФІНАНСОВОГО СТАНУ ПІДПРИЄМСТВА

Першим методом буде розглянуто побудову модель логістичної регресії. Логістична регресія – це статистичний регресійний метод, що використовується у випадку коли пояснювальна (залежна) змінна може набувати тільки двох значень, зазвичай 0 та 1 (бінарна логіт-регресія) або, більш загально, скінченну множину значень. Її призначення полягає у тому, щоб прокласти зв'язок між декількома незалежними змінними та однією залежною змінною. Таким чином, результатом роботи даного методу буде вірогідність тієї чи іншої події, а це якраз те, що нам потрібно, тобто необхідно оцінити вірогідність події банкрутства підприємства.

Для початку, розглянемо рівняння множинної регресії:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Якщо використовувати такий вигляд моделі для оцінки ймовірності події, ми можемо отримати проблему, коли значення y буде більшим за 1 або меншим за 0. Така ситуація можлива через те, що змінна відгуку бінарна за своєю природою при множинній регресії. Для вирішення задачі ми змінюємо завдання регресії і будемо прогнозувати не значення бінарної змінної, а безперервну змінну на інтервалі $[0;1]$. Таку заміну можна зробити за допомогою формули:

$$P = \frac{1}{1 + e^{-y}}$$

де P – ймовірність події; Y – визначається рівнянням регресії; e – основа натурального логарифма.

Оцінювання коефіцієнтів логістичної регресії може виконуватись різними методами, але зазвичай використовується метод максимальної правдоподібності (ММП). Він застосовується для оцінювання параметрів генеральної сукупності по даним вибірки. Основу метода виступає функція правдоподібності, яка виражає щільність імовірності сумісної появи результатів вибірки:

$$L(Y_1, Y_2, \dots, Y_k; \theta) = p(Y_1; \theta) \cdot \dots \cdot p(Y_k; \theta)$$

Згідно до методу, в якості оцінки невідомого параметра приймається таке значення $\Theta = \Theta(Y_1, \dots, Y_k)$, що максимізує функцію L . Максимізувати значення функції правдоподібності еквівалентно максимізації її логарифму:

$$L^*(Y, \theta) = \ln(L(Y, \theta)) \rightarrow \max$$

Так виглядає логарифмічна функція правдоподібності:

$$L^* = \sum_{i=1}^k \ln P_i(w) + \sum_{i=0} \ln(1 - P_i(w)) = \sum_{i=1}^k [Y_i \ln P_i(w) + (1 - Y_i) \ln(1 - P_i(w))]$$

де $0, 1$ – множина спостережень, для яких Y може приймати значення 0 або 1.

3. ПРАКТИЧНЕ ВИКОРИСТАННЯ МОДЕЛІ ЛОГІСТИЧНОЇ РЕГРЕСІЇ В ОЦІНКАХ ФІНАНСОВОГО СТАНУ ПІДПРИЄМСТВА

Розглянемо приклад використання логістичної регресії для визначення фінансового стану підприємства. Важливим недоліком бінарної логістичної регресії є те, що для обрахунків вона може отримувати обмежену кількість вхідних факторів. Тому для використання такого методу прогнозування потрібно залучити експерта, який визначить найважливіші фактори в фінансовій звітності підприємства. В прикладі використано такі показники фінансової звітності:

- коефіцієнт автономії;
- коефіцієнт абсолютної ліквідності;
- коефіцієнт поточної ліквідності;
- коефіцієнт оборотності;
- коефіцієнт рентабельності.

Маючи інформацію про показники 160 підприємств, перевіримо наскільки точно працює логістична регресія. На основі 154 підприємств побудуємо модель логістичної регресії та спробуємо спрогнозувати дані інших 6 підприємств. За допомогою системи EViews побудуємо модель (рис. 1):

Dependent Variable: CRITICAL_CONDITION
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 05/17/14 Time: 21:59
Sample: 1 154
Included observations: 154
Convergence achieved after 7 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
KOEF_ABS_LIKVID	-21.45350	7.413021	-2.894029	0.0038
KOEF_AUTONOMY	-8.887599	1.448821	-6.134365	0.0000
KOEF_OBOROT_AKTYVIV	-1.718040	0.375649	-4.573531	0.0000
KOEF_POTOCH_LIKVID	0.693932	0.223439	3.105683	0.0019
RENTAB_CAPITALU	-5.891127	1.983635	-2.969864	0.0030
C	5.340537	0.969530	5.508374	0.0000

McFadden R-squared	<u>0.663470</u>	Mean dependent var	0.500000
S.D. dependent var	0.501631	S.E. of regression	<u>0.284025</u>
Akaike info criterion	<u>0.544452</u>	Sum squared resid	11.93923
Schwarz criterion	0.662774	Log likelihood	-35.92277
Hannan-Quinn criter.	0.592514	Deviance	71.84554
Restr. deviance	213.4893	Restr. log likelihood	-106.7447
LR statistic	141.6438	Avg. log likelihood	-0.233265
Prob(LR statistic)	0.000000		

Obs with Dep=0	77	Total obs	154
Obs with Dep=1	77		

Рисунок 1. Основні характеристики якості моделі

Отримаємо таке рівняння логістичної регресії для нашої моделі:

$$P_{condition_{critical}} = 1 / (1 + \exp(-(5.340537 - 21.45350 * koef_abs_likvid - 8.887599 * koef_autonomy - 1.718040 * koef_oborot_aktyviv + 0.693932 * koef_potoch_likvid - 5.891127 * rentab_capitalu)))$$

На рисунку 2 показано прогноз за моделлю логістичної регресії та його порівняння з реальними значеннями по 6 підприємствах. Фактично отримано високу якість прогнозування стану підприємств.

Таблиця 1. Характеристики якості моделі

Назва	Формула	Значення
McFadden R-squared (коефіцієнт детермінації)	$= 1 - \frac{\log L^{max}(full)}{\log L^{max}(empty)}$	Відношення логарифмічної функції правдоподібності повної моделі до нульової Значення від 0 до 1 Чим більше значення, тим краще модель описує дані
Інформаційний критерій Акайке (AIC)	$= 2k - 2\ln(L)$ <ul style="list-style-type: none"> • k – кількість параметрів статистичної моделі; • L – максимізоване значення функції правдоподібності 	Для кращої моделі $AIC \rightarrow \min$
S.E. of regression (Standard Error of the Regression)	$= \frac{\sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2}}{N}$ <ul style="list-style-type: none"> • N – розмір вибірки • y – фактичні значення • \hat{y} – значення оцінені за моделлю. 	Для кращої моделі $S.E. \text{ regression} \rightarrow \min$

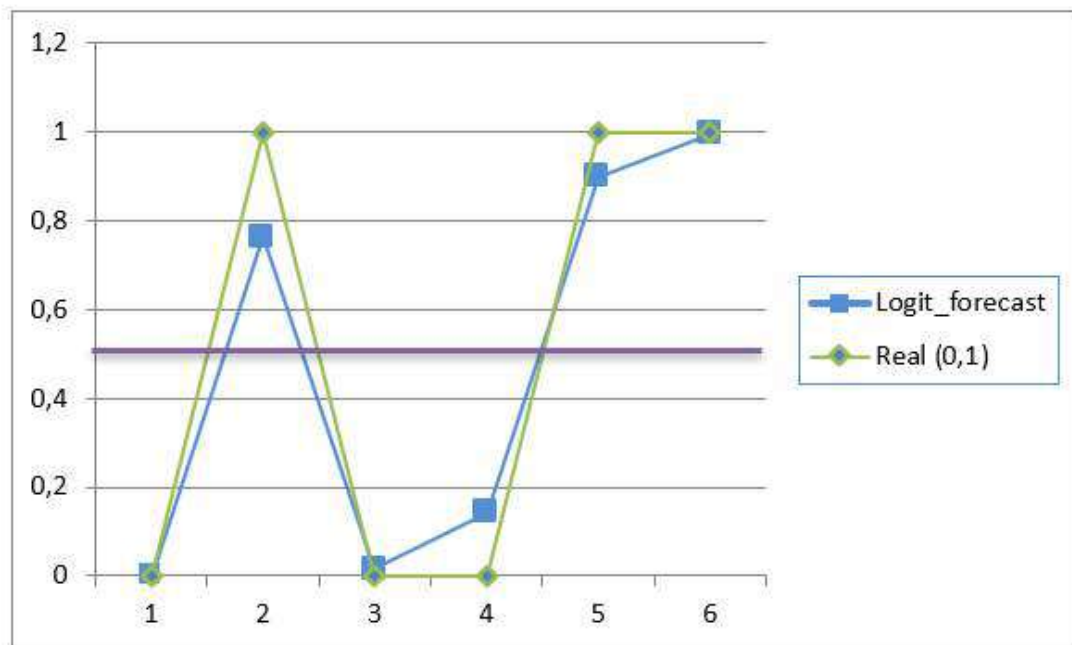


Рисунок 2. Прогноз за моделлю логістичної регресії та порівняння її з реальними значеннями по 6 підприємствах

4. ВИСНОВОК

Логістична регресія має високий ступінь точності визначення фінансового стану підприємства за допомогою правильно обраних коефіцієнтів. Однак, складність моделі заперечує її просте використання звичайним користувачем. Коректне використання прогнозуючої змінної допомогло нам використати модель для наших цілей та є дуже хорошим, що і намагались ми зробити в даній роботі. Вважаємо, що використання описаного методу знайде свого споживача та цілком може бути автоматизовано для кінцевого користувача, що спростить роботу та розуміння отриманих результатів і обов'язково допоможе в майбутньому.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Поддєрьогін А.М. Фінанси підприємств / Поддєрьогін А.М. – М.: Финансы и статистика, 1999. – 261с.
2. Ковалев В.В. Введение в финансовый менеджмент / Ковалев В.В. – М.: Финансы и статистика, 2000. – 234 с.
3. Sahakian С.Е. The Delphi Method [Text] / Sahakian С.Е. – The Corporate Partnering Institute, 1997. – 11 p.
4. А.Н.Борисов. Модели принятия решений на основе лингвистической переменной / А.Н.Борисов и др. – Рига: Зинатне, 1982. – 256 с.
5. Bollerslev Т. Generalized Autoregressive Conditional Heteroskedasticity / Bollerslev Т. // Journal of Econometrics. – 1986. - Vol. 31 – p. 307-327.
6. Engle Robert F. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation / Engle Robert F. // Econometrica. - 1982. - Vol. 50 – p. 987-1007.

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ

Селін Ю.М.¹, Жук В.М.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

²volodymyr.zhuk.gp@gmail.com

Вивчення та прогнозування часових рядів економічної природи потребує впровадження широкого кола засобів математичного апарату, інформаційних систем, систем керування сховищами даних тощо. Розробка та викроистання інтелектуальних систем (ІС) для прогнозування нелінійних нестационарних процесів в економіці дозволить частково автоматизувати створення прогнозів, підвищити якість прогнозів використаних алгоритмів, знизить планку необхідних знань і навичок для використання та розуміння поведінки того чи іншого економічного процесу. Метою роботи є створення ІС для прогнозування нелінійних нестационарних процесів в економіці із застосуванням системного підходу. Результатом дослідження є ІС, що виконує прогнозування часових економічних рядів з використанням алгоритму лінгвістичного моделювання.

Ключові слова: інтелектуальна система, системний підхід, лінгвістичне моделювання, часові ряди.

1. ВСТУП

Зберігання та аналіз даних економічної природи – один з ключових кроків для детального розуміння виникнення та поведінки будь-якого економічного процесу. Правильна робота з цими даними дозволяє комплексно прогнозувати процеси економічної природи.

Саме тому створення інформаційної системи, що дозволяє комплексно проводити моделювання часових рядів є одним з найактуальніших завдань, яке дозволяє покращити точність прогнозів, зменшити обсяг витраченого часу, автоматизувати необхідні процеси тощо. Застосовуючи різні методології прогнозування, можливо покращити загальну точність інформаційної системи, в залежності від природи даних, які моделюються. Подальше порівняння результатів дасть змогу краще налаштувати необхідні параметри для отримання більш точного прогнозу.

2. МЕТОДИ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Існує багато методів прогнозування часових рядів. Найбільш розповсюджені з них – методи регресійного аналізу. Ці методи зазвичай дозволяють побудувати точний прогноз із лінійними або експоненційними залежностями. Хоча існують і поліноміальні види регресії для побудови більш складних моделей. Підхід побудови регресії базується на знаходженні відношення між залежною змінною y та однією або декількома незалежними змінними x . Для збільшення ефективності моделі, рівняння може містити значення шуканої змінної, затриманої у часі на визначену кількість кроків(лаги). Тому, мішане рівняння регресії виглядає наступним чином : $y(k) = a_0 + a_1y(k - 1) + a_2y(k - 2) + \dots + a_n(k)y(k - n) + b_1x_1(k) + b_2x_2(k) + \dots + b_px_p(k) + e(k)$ [1]. Також може бути застосований

сегментований підхід, коли для кожного окремого сегменту намагаються побудувати регресійну модель [2].

Метод групового урахування аргументів іноді вважають узагальненням підходу регресійного аналізу. Ця методика використовується у дуже різноманітних галузях аналізу даних (наприклад, розпізнавання образів, моделювання різноманітних систем, оптимізації тощо). У ході виконання алгоритму відбувається перебір різних варіантів моделювання вхідного процесу, при цьому автоматично обирається найкращий варіант моделі за спеціальними критеріями адекватності (коефіцієнт детермінації, суми квадратів похибки тощо). Постановка завдання для застосування методу повинна включати наступні елементи :

- вибір додаткових описів, які створюють основу для можливого остаточного вибору моделі;
- відбір та адаптація параметрів моделі для конкретного застосування;
- розробка нової або застосування відомої моделі оцінки;
- вибір критерії якості моделі для використання в проміжних кроках обчислень і для остаточного вибору моделі;

Моделі, побудовані з відповідним чином налаштованим МГУА, зазвичай забезпечують середню або високу якість короткострокових прогнозів [3].

Серед методів, що будуть використовуватися в ІС для прогнозування нелінійних нестационарних процесів слід зазначити лінгвістичного моделювання. Цей метод належить до родини статистичних методів прогнозування, що зумовлює широку можливість його використання з даними абсолютно різної природи. Даний тип методів прогнозування відрізняється універсальністю, проте трохи гіршою точністю.

Для побудови системи, необхідно розв'язати задачу пошуку лінгвістичного образу для досліджуваного часового ряду, яка складається з декількох етапів:

- обрахунок різницевого ряду;
- вибір значення критерію інтервального розбиття різницевого ряду;
- співставлення часовому інтервалу ряду певної літери алфавіту;
- побудова матриці переходу між будь-якими двома літерами;

Задача побудови першої різниці часового ряду – аналогу першої похідної [4]. Цей спосіб дозволяє прибрати тренд часового ряду та зробити його стаціонарним для подальшої побудови моделі для прогнозу, оскільки майже все методі прогнозування дають успішний результат при використанні їх зі стаціонарними часовими рядами. Слід зазначити, що використання перших різниць не завжди зробить ряд стаціонарним. Для визначення порядку тренду можна проаналізувати коефіцієнти автокореляції часового ряду. Маємо : вектор з цілих чисел X потужністю: $n = X$. Отримаємо : вектор з цілих чисел D потужністю: $k = |D|$.

Тобто, $\forall d_i \in D: d_i = x_{i+1} - x_i$.

Задача вибору оптимального значення критерія інтервального розбиття різницевого ряду. Розв'язок задачі дозволяє побудувати алфавіт користувача шляхом розділення відсортованого ряду першої (зазвичай, якщо порядок тренду лінійний) різниці на безліч інтервалів, в якому кожний елемент характеризує певну літеру заданого алфавіту [4]. Варто розуміти, що дуже велика або дуже маленька кількість інтервалів розбиття негативно вплине на отриманий результат, тому що при побудові матриці переходу або дуже багато значень будуть потрапляти в один проміжок (при малому розбитті) або значення ймовірності переходу від одної літери до іншої будуть майже однаковими (при великому розбитті).

Лінгвістична задача. Метою цієї задачі є формування лінгвістичного ланцюга шляхом співставлення відповідної букви алфавіту для кожного спостережуваного значення різницевого ряду. Кожна буква у визначеному алфавіті однозначно відповідає певному

інтервалу з набору інтервалів, які були отримані при розв'язанні минулої задачі розбиття ряду відповідними інтервалами. Лінгвістична задача. Метою цієї задачі є формування лінгвістичного ланцюга шляхом співставлення відповідної букви алфавіту для кожного спостережуваного значення різницевого ряду. Кожна буква у визначеному алфавіті однозначно відповідає певному інтервалу з набору інтервалів, які були отримані при розв'язанні минулої задачі розбиття ряду відповідними інтервалами[4].

Задача побудови матриці переходу між будь-якими двома літерами. Отримана послідовність аналізується на наявність граматичних конструкцій. На виході ми отримуємо список граматичних конструкцій з ймовірностями їх присутності в процесі, а також матрицю ймовірностей переходу від одного символу до іншого символу. Цей етап тісно пов'язаний із моделюванням (прихованих) процесів Маркова, а також методом подібних траєкторій.

ІС складається таких основних модулів: інтерфейсу користувача, підсистеми аналізу, системи прогнозування та модуля генерації звіту. Користувач завантажує файл з даними, система сама проводить оцінку часового ряду, визначає необхідні параметри для методів, що будуть використовуватися та будує прогнози. Далі отриманий прогноз звіряється з реальним, похибка зберігається у базу даних, а наступне обчислення робиться з урахуванням дійсного результату. Далі описана поверхнева блок-схема структури ІС (Рис. 1).



Рисунок 1. Структура інформаційної системи

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для цілей прогнозування був обраний часовий ряд акції компанії Netflix. У якості часового горизонту було обрано період в один торговельний рік – 252 дні. Період прогнозування було взято в 5 днів.

Візуалізація часового ряду дає інформацію щодо його нестационарності (Рис.2).

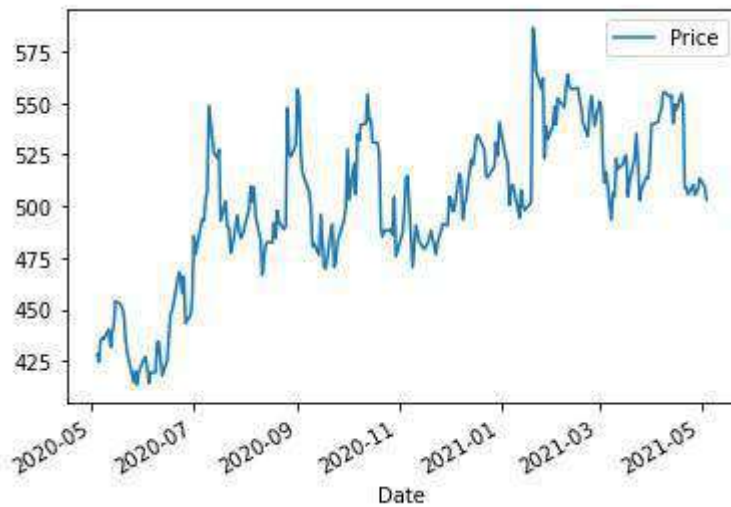


Рисунок 2. Графічне зображення часового ряду “Netflix”

Метод лінгвістичного моделювання дозволяє застосовувати мінімальну необхідну попередню обробку даних перед використанням даного алгоритму. Необхідність диференціювання ряду для зведення його до стаціонарного є єдиною необхідною умовою для подальшої побудови математичної моделі.

Після приведення ряду до стаціонарного ми визначаємо критерій розбиття, тобто на скільки інтервалів нам розділити значення часового ряду та крок з яким ми будемо це робити. Крок визначається як різниця максимального та мінімального значення диференційованого ряду, поділеного на кількість інтервалів. Зазвичай оптимальною є кількість 25-35 інтервалів, де добре підходять літери англійського алфавіту. Далі ми будемо відповідні інтервали, співставляючи кожній літері відрізок. Після чого ми перетворюємо наш числовий ряд у лінгвістичну послідовність. Далі побудуємо матрицю переходів яка містить ймовірнісні значення переходу від однієї літери до іншої. Після отримання прогнозу, ми рахуємо отриману помилку та використовуємо справжнє значення для створення наступного прогнозу.

Результати прогнозування часового ряду різними методами наведено нижче (Табл. 1).

Таблиця 1. Порівняння результатів статистичних тестів використаних методів

Модель часового ряду	R^2	Sum squared resid	Durbin – Watson
AP(1)	0.87447	43540.31	2.21994
AP(4)	0.872836	42237.67	1.95032
AP(5)	0.87669	40409.32	1.95289
APKC(4,4)	0.90509	39770.35	2.00993
ARIMA(4,1,4)	0.955309	40037.91	2.005514
ARIMA(5,1,1)	0.968889	41037.91	2.017372
Linguistic modeling	0.9348136	11.637035	1.425621

Також ми можемо побачити результати прогнозування, виконане за допомогою методу лінгвістичного моделювання на 5 кроків вперед (Рис. 3).

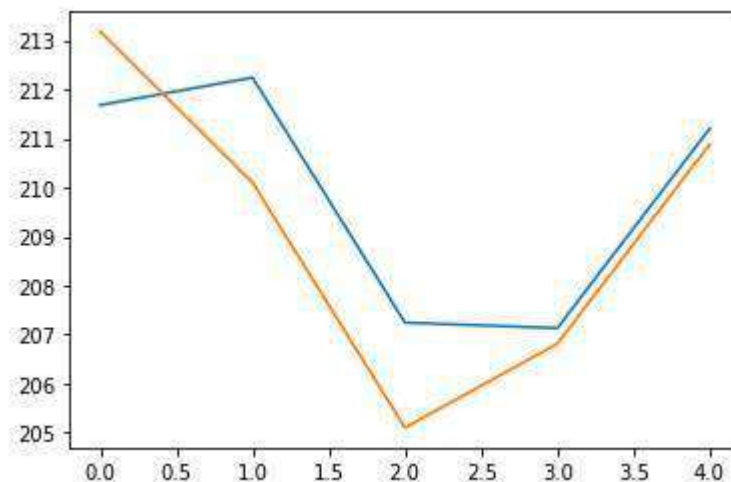


Рисунок 3. Прогноз на 5 кроків вперед з використанням методу лінгвістичного моделювання

4. ВИСНОВКИ

Було створено інформаційну систему, яка дозволяє прогнозувати нелінійні нестационарні процеси в економіці. Окрім того, виконані наступні поставлені задачі:

- виконано аналіз деяких методів дослідження нелінійних нестационарних процесів;
- виявлено нелінійності і нестационарності у сучасних фінансово-економічних процесах;
- вибрано процеси для дослідження та зібрано необхідні статистичні дані;
- виконано порівняльний аналіз отриманих результатів і виробити рекомендації стосовно їх практичного застосування;

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Юрченко М. Є. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 “Економіка” освітня програма “Економічна кібернетика”, “Економічна аналітика”. 2018. 88с.

2. Бідюк П. І, Романенко В. Д., Тимошук. О. Л. Аналіз часових рядів: підручник / НТУУ КПІ ім. Ігоря Сікорського. Київ: ВПК "Політехніка", 2013. 599 с.

3. Зайченко Ю.П., Кебкел О.Г., Крачковський В.Ф. Нечіткий метод групового врахування аргументів та його застосування в задачах прогнозування макроекономічних показників. //Наукові вісті НТУУ «КПІ», №2, 2000р. с.18-26

4. Селін Ю.М. Системний аналіз екологічно небезпечних процесів різної природи / Ю.М. Селін // Системні дослідження та інформаційні технології. — 2007. — № 2. — С.22–32

ТЕОРЕТИКО-ІГРОВИЙ ПІДХІД ДЛЯ ТРЕЙДИНГОВИХ СТРАТЕГІЙ НА КРИПТОРИНКУ

Семіконь Я.В.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

semikonur@gmail.com

Мета дослідження - виявити кращу стратегію для отримання прибутку у криптовалюті, використовуючи фундаментальний та технічний аналіз для коректного визначення можливих вигравів та ймовірностей. Більше того, до дослідження залучається теорія ігор, а саме ігри з природою. Економіка або економічний ринок непередбачувані, тому у даній роботі такі ігри розглядаються з використанням декількох критеріїв: критерій Севіджа, Байєса, Лапласа.

Ключові слова: теорія ігор, трейдинг, криптовалюти, фундаментальний аналіз, ігри з природою.

1. ВСТУП

Криптовалюта або електронна готівка – це віртуальна грошова одиниця, яка вимірюється у цифрових монетах або коїнах (від англ. coin). Ці монети складаються з зашифрованих пакетів даних, які неможливо скопіювати або підробити. Щоб згенерувати інформацію, яка лежить в основі цифрових монет, знадобиться потужний комп'ютер та нехитре програмне забезпечення. Емісія, або іншими словами виробництво монет, відбувається децентралізовано, тобто немає єдиного емітента. Технічний сенс майнінгу – добувати зашифровані коди за допомогою підбору числових варіацій. Сам процес майнінгу – це розрахунок ланцюжка даних для блоку інформації, який вносять у блокчейн. За розрахунок та формування нового блоку даних, який відповідає всім вимогам системи, майнер (власник цього або цих комп'ютерів) отримує певну кількість одиниць криптовалюти, які можуть після конвертуватися у звичні нам гроші (гривні, долари, євро тощо).

Першою криптовалютою у світі є Біткоїн (Bitcoin), саме він стоїть біля джерел формування ринку криптовалют і всього, що з ним пов'язано. Біткоїн, а точніше технологія його виробництва, з'явився на світ у 2008 році. Його «батьком-засновником» є Сатоші Накамото, який опублікував файл з описом протоколу та принципу роботи платіжної системи за допомогою однорангової мережі. Виробляти біткоїни може кожен, хто має необхідне обладнання, але при цьому ніхто не може самовільно змінювати кількість існуючих біткоїнів іншим шляхом, окрім як самостійно створювати їх за всіма правилами системи.

Торгівля криптовалютою – це купівля-продаж криптовалют з метою отримання прибутку. Торгова стратегія в торгівлі криптовалютою, сформульована інвестором, – це алгоритм, який визначає набір заздалегідь визначених правил для купівлі та продажу на ринках криптовалют.

Теорія ігор – це математична дисципліна, що сформувалася у середині минулого століття, за часів великих війн. Гра – частина нашого життя, і конфлікти, які виникають і вирішуються через воєнні дії, скандали, проблеми – це все є елементами теорії ігор. Коли виникла математична теорія ігор, математики показали, що завдання теорії ігор –

формалізувати реальну життєву ситуацію, що є грою; навчитися за допомогою математичних методів отримувати додаткові переваги у цій грі, вибирати оптимальні стратегії.

2. ТЕОРІЯ ІГОР З ВИКОРИСТАННЯМ АНАЛІЗУ КРИПТОРИНКУ

Теоретичний ігровий підхід до фінансових ринків є не менш «науковим», ніж будь-який інший; це просто нешаблонно. Якщо гру на фінансових ринках можна розглядати як форму гри, то чому б не використати науку про ігри для їх аналізу?

Прийняття рішень в умовах невизначеності відбувається, коли кожне рішення призводить до конкретного результату, який відбувається з невідомою ймовірністю. Якщо ви йдете додому в годину пік і хочете мінімізувати час, витрачений на дорогу, чи варто вибирати «звичайний» маршрут чи спробувати альтернативний маршрут? Будь-яке рішення призводить до конкретного результату (прибуття додому), але ви можете не знати який маршрут є найшвидшим у конкретний день.

Щоб визначити ціну на акції або валюту, або на будь-який інший фінансовий інструмент, існує два основних типи аналізу фінансових ринків. Перший називається технічний аналіз (ТА). Технічний аналіз передбачає, що для прогнозування ціни на актив у майбутньому достатньо знати статистику як у минулому ринок реагував на якісь певні фактори і швидше за все з великою ймовірністю і в майбутньому поводитиметься так само. Другий – фундаментальний аналіз – це метод аналізу та прогнозування ринку з урахуванням усіх фундаментальних фактів: політичні, переговори, вибори, зміна влади, всі можливі форс-мажори, природні катаклізми, війни та революції, психологічні фактори, якісь чутки, очікування. Це все також може впливати на фінансові ринки. Інша група факторів – це економічні чинники і їх потрібно розділити на дві категорії. Перша, – якщо ми говоримо про економічні чинники, пов'язані з якоюсь конкретною компанією, де необхідно враховувати всі публікації, звітності, засідання ради директорів, можливо прийняття якихось важливих рішень. Другий, – якщо ми говоримо про фундаментальні фактори в рамках світової економіки або економіки конкретної держави.

У дослідженні розглядається практичне застосування знаходження найкращої стратегії у заданій задачі з критеріями Севіджа, Байеса і Лапласа, а також аналіз отриманих результатів. У роботі використовуються реальні дані, отримані з відкритих доступів сайтів Yahoo Finance та Trading view. Для задачі візьмемо три криптовалюти: одна, яка відома усім індивідам, що займаються трейдингом, або просто знайомі з цією темою; другою буде криптовалюта нової української біржі, що запустилась порівняно нещодавно та випустила свій криптокоїн; третя – теж відносно нова криптовалюта – токен управління проектом STEPN, грою, яка працює за принципом «ходи і заробляй»; GMT використовується для розрахунків в ігровому пулі, дозволяючи гравцям заробляти гроші просто ідучи або прогулюючись. Це валюти Ethereum, Qmall coin та STEPN (GMT) відповідно. Надалі будемо розглядати їх як потенційні криптовалюти для інвестування. Сама задача буде поставлена таким чином: трейдеру необхідно вибрати одну з трьох криптовалют, в яку він потенційно може вкластися з 10 тисячами гривень, якщо прибуток залежить від того, яким буде ринок: ріст, консолідація (стабілізація курсу) або падіння. Також важливо встановити рамки інвестування; в нашій задачі період інвестування буде місяць - з 28 березня 2022 року до 28 квітня 2022 року. Використовуючи критерій Байеса для встановлення ймовірностей, необхідно перш за визначити можливі стани ринку в найближчій час. Важливо розуміти, що Біткоїн є засновником, «батьком» усіх інших існуючих коїнів, тому треба враховувати, що якщо Біткоїн знаходиться на стадії падіння, то і всі інші монети не будуть рости (принаймні так швидко та з великими іксами, як при Біткоїні у стадії консолідації чи росту). По-друге,

використовуючи технічний аналіз необхідно проаналізувати графік Біткоїну за періоди: місяць, півроку та рік для бачення усїєї картини та спрогнозувати вірогідності стану ринку через місяць.

Необхідно проаналізувати графіки таких періодів:

1. 12 місяців (28.03.21 - 28.03.22)
2. 6 місяців (28.09.21 - 28.03.22)
3. Місяць (28.02.22 - 28.03.22)



Рисунок 1. Графік біткоїну за період 28.03.2021 - 28.03.2022, таймфрейм 1 тиждень

Бачимо головну картину, яку маємо з Біткоїном, а саме те, що на графіку є фігура «подвійна вершина», що виступає провідником розвороту тренду (Рис. 1). Тобто, з ймовірністю 90% тренд піде вниз. Тепер необхідно дивитись більш детально період за пів року (Рис. 2).



Рисунок 2. Графік біткоїну за період 28.09.2021 - 28.03.2022, таймфрейм 1 день

На даному графіку (Рис. 2) бачимо, що тренд дійсно йде все вниз та вниз, але з лютого спостерігаємо деяку стабільність, тобто маємо етап консолідації. Консолідація – це термін

для акцій або цінних паперів, які не продовжують і не змінюють більшу цінову тенденцію. Консолідовані акції зазвичай торгуються в обмеженому діапазоні цін і пропонують відносно мало торгових можливостей, доки не з'явиться інша модель. Наступний етап може бути будь-яким, але так як ми розуміємо ситуацію загалом (дивлячись на річний графік) можемо припустити, що графік піде вниз. У той час на 28 березня маємо відмітку у 47 152 долара за 1 біткоїн (Рис. 3).



Рисунок 3. Графік біткоїну за період 28.02.2022 - 28.03.2022, таймфрейм 4 години

На графіку (Рис. 3) за період у місяць бачимо, що нібито графік йде наверх, пробиває максимум за останній місяць, та, можливо, така ціна протримається або піде вгору на дуже короткий час. Та, дивлячись трохи далі та спираючись на загальний графік, можемо не передбачити значний ріст ціни. Тому маємо ймовірність того, що буде падіння цін протягом зазначеного періоду - 50%, консолідації - 35% та росту - 15% відсотків.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Необхідно залучити не тільки технічний, а й фундаментальний аналіз. Наразі Qmall та STEPN (GMT) нові монети, тому важливо розуміти, що на нових проєктах на початку є багато новин стосовно них (наприклад безкоштовна комісія на вивід/залиття грошей або підключення нових функцій, розіграші і т.п.), а, отже, і запит та інтерес багатьох людей інвестувати дуже великий. Тому якщо буде ріст, то він буде стрімким, дасть багато іксів (але недовгим). Щодо Ethereum, то ця криптовалюта вже давно у використанні на ринку, та вона «йде» за біткоїном, тому якщо біткоїн піде вниз, то і ціни на Ethereum також підуть вниз.

Таблиця 1. Таблиця вигравів монет та ймовірностей на кожному стані ринку

State/ Cryptocurrency	Growth	Consolidation	Loss
Ethereum	2	1	- 3
Qmall	3	2	- 1
STEPN (GMT)	4	2	- 2
probability	0,15	0,35	0,5

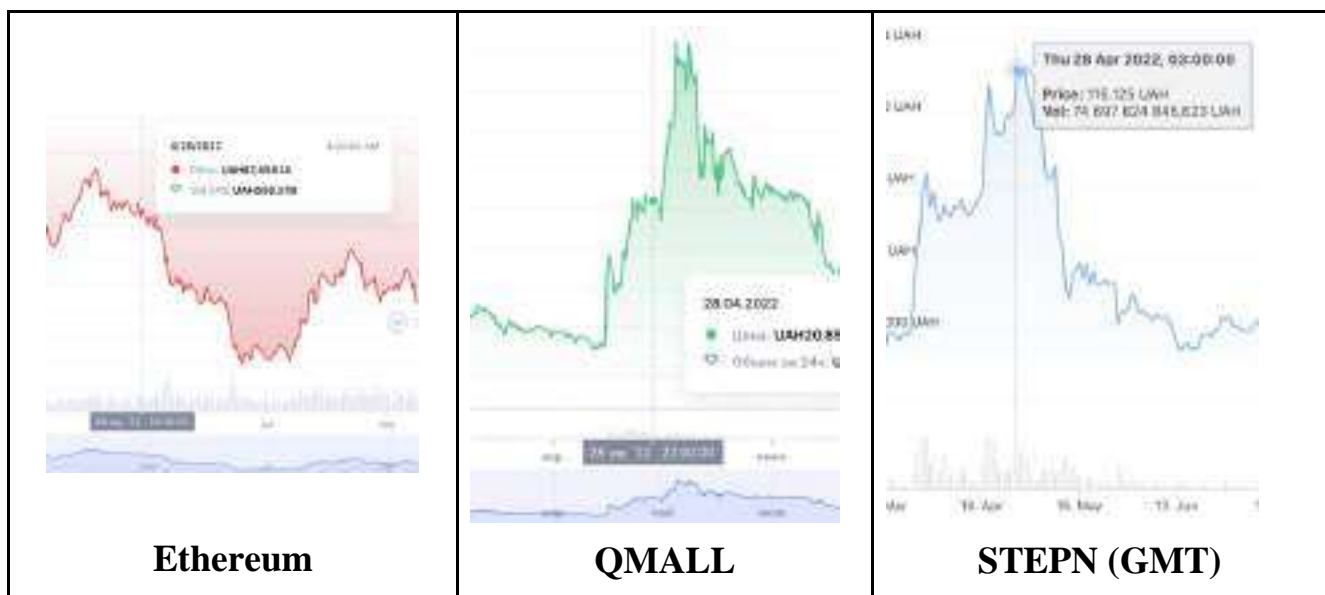
Таблиця 2. Порівняльна таблиця отриманих результатів з використанням критеріїв Лапласа, Байеса та Севіджа

Criterion	Laplace	Bayes	Savage
Result	Qmall, STEPН (GMT)	Qmall	Qmall, STEPН (GMT)

З результату бачимо, що непогано було б інвестувати в монети Qmall та STEPН (GMT), але є результат за критерієм Байеса, який показує лише один варіант інвестування – це монета Qmall.

Нижче наведена таблиця з вже відомими цінами на наш період з 28 березня по 28 квітня.

Таблиця 3. Таблиця даних з вартості криптовалют на заданий період



Для розрахунків була реалізована програма, яка знаходить кращу стратегію за матрицею виграшів за критеріями Лапласа, Севіджа та Байеса. Програма розроблялася в середовищі IntelliJ IDEA CE та була реалізована на мові Kotlin.

4. ВИСНОВКИ

У даному дослідженні була описана задача про виявлення кращої стратегії для інвестування на крипторинку з завчасно обраних варіантів: монети QMALL, STEPН (GMT) та Ethereum. Також були обрані дані вартості біткоїну у графіках за різний період часу та потім проаналізовані, використовуючи технічний аналіз. Більше того, був задіяний фундаментальний аналіз для виявлення наскільки швидко або стрімко монета або токен може вирости.

Програма, що виявляє найкращий шлях для інвестування та заробляння гарного прибутку, була написана на мові JVM Kotlin у середовищі IntelliJ IDEA. Продукт ще може бути вдосконалений, а саме додаванням більшої кількості можливих монет для інвестування та реалізування юзер френдлі віконечка для вводу та виводу результату.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Сергей Бондаренко. Теория игр Инструмент, благодаря которому блокчейн-революция стала возможной. URL: <https://www2.deloitte.com/ua/uk/pages/press-room/deloitte-press/2018/game-theory.html> (дата звернення: 16.11.2022).
2. Филиппов Е. Криптовалюта. М.: STFOREX, 2017. URL: http://el.tfi.uz/images/Kriptoalyuta_ot_a_do_ua.pdf (дата звернення: 16.11.2022).
3. Круглик, Р. И., Брыкова, А. Л. Обзор бирж криптовалют для трейдерской деятельности. Постулат. 2018. № 1.
4. Shelton R. B. Gaming the market: Applying game theory to create winning trading strategies. John Wiley & Sons. 1997. Vol. 69.
5. Богоявленский С.Б. Критерий Лапласа. 2014. URL: http://risking.ru/materials/risktheory/part2_9.html (дата звернення: 16.11.2022).

СИСТЕМНИЙ ПІДХІД ДО АНАЛІЗУ КРЕДИТНОГО ТА РИНКОВОГО РИЗИКІВ

Смиківська Д.В.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹smykovska.daryna@lil.kpi.ua

Кредитний ризик і ринковий ризики є двома найбільш поширеними і критичними фінансовими ризиками. Для більш цілісного та ефективного аналізу цих типів ризику варто застосовувати системний підхід. Побудова системи підтримки прийняття рішень (СППР) є досить розповсюдженим фактичним впровадженням системного підходу, що значно полегшує процес аналізу ризиків. Метою роботи є аналіз ризику контрагента та ризику процентної ставки, а також їх кореляції. Результатом дослідження є побудована СППР, що моделює процентну ставку, оцінює ймовірність дефолту та взаємозв'язок між двома типами ризиків за допомогою моделювання CVA. У роботі використано теоретичні та емпіричні методи дослідження.

Ключові слова: системний підхід, кредитний ризик, ринковий ризик, СППР, CVA, метод Монте-Карло.

1. ВСТУП

Слід зазначити, що будь-яка фінансова дія передбачає певний рівень ризику. Ризики є наслідком наявності невизначеності та є випадковими величинами, оскільки визначаються існуванням різноманітних факторів, вплив яких на результати фінансової діяльності неможливо достовірно передбачити заздалегідь.

Ринковий і кредитний ризики часто розглядалися як не пов'язані джерела ризику: види ризиків вимірювалися окремо та управлялися окремо. Проте кредитний і ринковий ризики переплітаються. По-перше, кредитний ризик залежить від факторів ринкового ризику, оскільки ймовірність дефолту, вартість застави та вартість вимог можуть залежати від процентних ставок, обмінних курсів або інших ринкових цін. По-друге, ринковий ризик залежить від факторів кредитного ризику, оскільки дефолт контрагента може відкрити раніше закрити позицію.

Неповне розуміння взаємозв'язку між кредитним і ринковим ризиком все ще є джерелом недоліків в управлінні ризиками фінансових установ. Інтегроване управління кредитними та ринковими ризиками може допомогти окремим установам досягти кращого розуміння їхньої загальної ризикової позиції, таким чином сприяючи покращенню стабільності фінансової системи в цілому.

Наразі досить критичною є тема якісних методів аналізу ризиків. Це пояснює чому популярним є використання методології, яка передбачає детальне вивчення усіх ознак того чи іншого типу ризику. Саме такою можна розглядати системну методологію, адже вона вимагає поетапності дій у вивченні ризиків та допомагає в удосконаленні процесів збору та обробки інформації, і оцінювання того, що використовується в процедурах моделювання, і розробці управлінських рішень. Тому можна зазначити, що саме системний підхід надасть змогу визначити рівень ризику та його кореляцію з іншим типом ризику.

2. МЕТОДИ ОЦІНЮВАННЯ КРЕДИТНОГО ТА РИНКОВОГО РИЗИКУ

Для того щоб ефективно керувати ризиком потрібно його ідентифікувати та провести аналіз для загальної структури розуміння його виникнення і трендів притаманних цьому ризику. Наступним етапом є оцінювання ризику, що дозволяє корпораціям, урядам та інвесторам оцінити ймовірність того, що несприятлива подія може негативно вплинути на бізнес, економіку, проект або інвестиції. Хоч ризик не завжди являє собою саме настання несприятливих подій при його оцінюванні завжди визначаються саме можливі втрати.

Істотним є не тільки оцінювання можливих втрат, а й структурний та цілісний підхід до огляду чинників виникнення ризику і його залежності від факторів, що впливають на подальші можливі відхилення від очікуваних результатів. Тому слід акцентувати увагу на важливості імплементації системного підходу.

Системний підхід до аналізу ризиків можна визначити як метод дослідження інтегральних залежностей і взаємодій факторів, що впливають на появу або прояв ризиків, прогнозування та оцінювання наслідків їх реалізації [1].

Як було зазначено раніше, одним із результатів імплементації системного підходу є саме побудова СППР. СППР генерує рішення за допомогою різноманітної множини алгоритмів і методів, включаючи алгоритми перетворення даних, ігрові методи пошуку та обробки даних, ситуаційний аналіз, методи прийняття рішень на основі нейронних мереж, імітаційне моделювання, метод когнітивної карти, байєсівські мережі та інше.

СППР може бути використана в діяльності малих підприємств з низьким ступенем забезпеченості та розвитку інформаційних технологій. Архітектура СППР розділена на такі рівні: - інтерфейс; - обчислювальний модуль (система обробки даних); - база знань і даних; - система отримання результатів (система прийняття рішень) [2, 3].

Аналіз кредитного і ринкового ризиків відбувається із застосуванням методів, заснованих на побудові певних математичних моделей. Існує множина таких методів і моделей, де серед найбільш популярних для аналізу ринкового та кредитного ризиків можна визначити моделі АРУГ, УАРУГ, VaR, лінійну регресію, дерева рішень, випадковий ліс, нейронні мережі та мережі Байєса, моделювання CVA (ринкова вартість кредитного ризику контрагента), модель Vasicek та модель Hull-White. Моделювання CVA для процентного свопу складається з: - моделювання процентної ставки; - обчислення ймовірності дефолту, очікуваного ризику; - підрахунку CVA.

Для прогнозування процентних ставок, які використовуються для процентного свопу використано модель Vasicek. Модель Vasicek вже давно використовується як техніка для прогнозування еволюції цін на фінансові інструменти. Ринковий ризик, час, волатильність і середній коефіцієнт повернення є компонентами цього підходу [4, 5].

Однофакторна модель – це така, яка має лише один фактор ризику, це свідчить про те, що зміни процентної ставки пояснюються одним компонентом (фактором ринкового ризику).

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Розрахунки дослідження базувалися на таких інструментах, як процентні свопи та кредитні дефолтні свопи (CDS). Для дослідження була обрана така інформація:

- 3-місячна середньозважена процентна ставка за міжбанківськими кредитами Libor;
- кредитний дефолтний своп контрагента JPMorgan Chase дефолтного свопу.

У цій роботі використано однорічний процентний своп з датою розрахунків від 17.09.2019 і терміном дії до 17.09.2020.

Процентні ставки прогнозувались з використанням моделі Vasicek. Усі симуляції починаються з періоду 0, який є датою початку обміну. Це вказує на те, що протягом усіх симуляцій процентна ставка періоду 0 буде встановлена на рівні 2,7859%. Довгострокове середнє значення становитиме 3,7337%, що є історичним середнім значенням 3-місячної ставки Libor у доларах США від початку до періоду 0. Волатильність визначається як історичне стандартне відхилення тримісячної ставки Libor у доларах США, яке становить 0,0427.

Завдяки моделюванню 10000 симуляцій процентної ставки методом Монте-Карло отримуємо наступний графік (Рис.1):

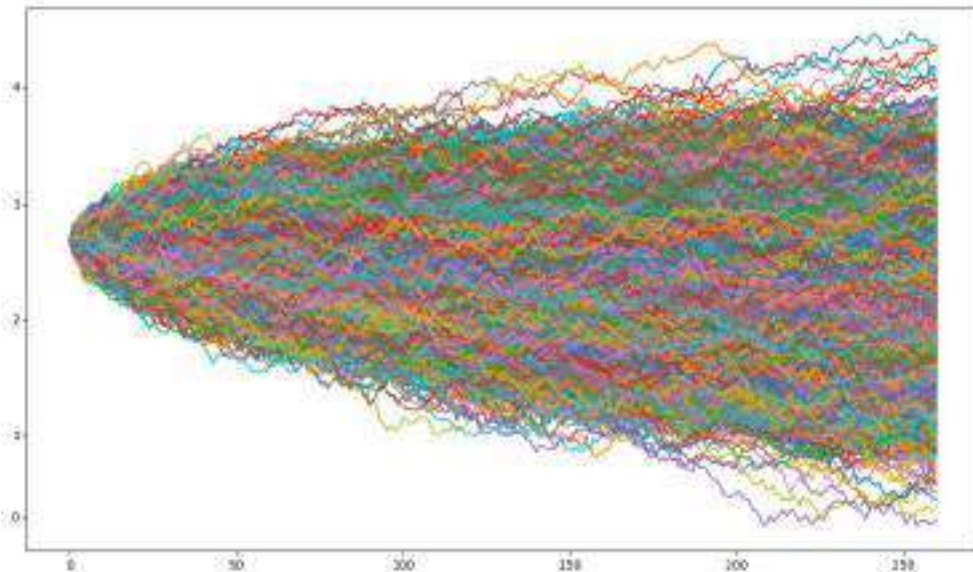


Рисунок 1. 10000 симуляцій процентної ставки

Найбільше значення у моделюванні становить близько 4,5%, тоді як найменше значення у моделюванні становить приблизно 0%. Модель Васічека враховувала від'ємні відсоткові ставки, і, як можна побачити, кілька симуляцій опускалися нижче 0% у деякі моменти часу.

Симуляції негативної ставки невеликі в основному через реверс середнього значення. Важко завершити моделювання негативним результатом, коли процентні ставки відхиляються в бік довгострокового середнього (що є позитивним). Однак, дивлячись на історичні 3-місячні ставки Libor в доларах США, це не обов'язково викликає занепокоєння, оскільки ці ставки ніколи не опускалися нижче 0% протягом тривалих періодів часу, але регулярно торкалися нульової нижньої межі під час Великої рецесії.

Далі ми розраховуємо щоденні шляхи впливу та моделюємо Expected Exposure (Рис. 2):

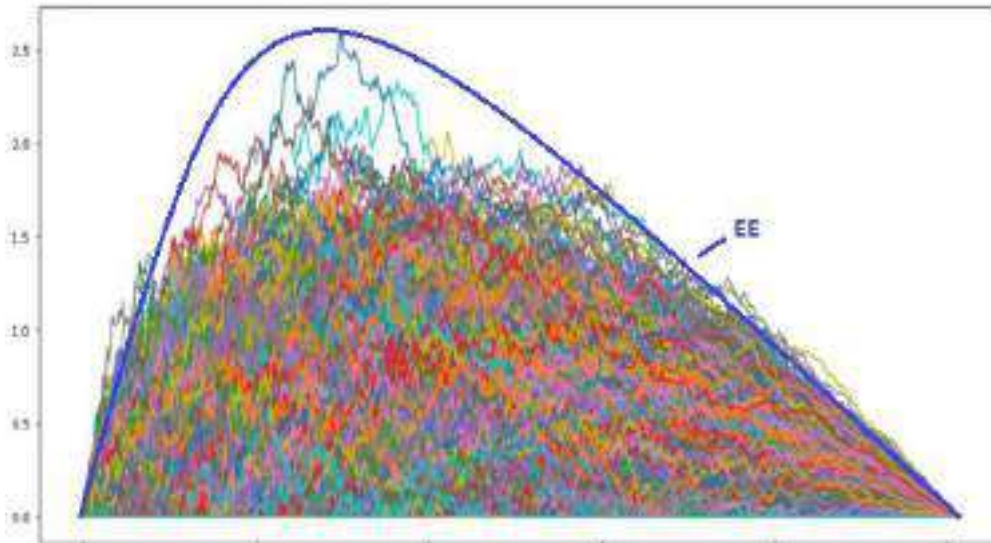


Рисунок 2. Шляхи впливу та Expected Exposure

LGD або втрата в разі дефолту є фіксованим параметром, і в цьому дослідженні дорівнює 61%.

JPMorgan Chase є контрагентом у цьому дослідженні, і оскільки це велика фірма, вона має відповідні CDS, які будуть використовуватися в наступних обчисленнях.

Розраховуємо ймовірність дефолту, як ймовірність дефолту між двома періодами. Отримані значення подані в таблиці 1.

Таблиця 1. Граничні ймовірності дефолту для JPM

Дата погашення	Гранична ймовірність дефолту
6 місяців	0.0024039308084044123
1 рік	0.0032476259533171154
2 роки	0.004158892130961667
3 роки	0.004967525197541266
4 роки	0.006329247322708333
5 років	0.007743749056949012
7 років	0.011023452687434012
10 років	0.014031132505716224

У таблиці нижче показано граничний ризик дефолту для термінів погашення на кривій спреда CDS. Однак нам потрібні ці цифри на щоденній основі для того щоб розрахувати оцінку CVA для процентних свопів. Щоденні значення будуть знайдені за допомогою інтерполяції поліномами методом Ерміта для того щоб гарантувати, що щоденні ймовірності за замовчуванням обчислюються на основі таблиці 4.

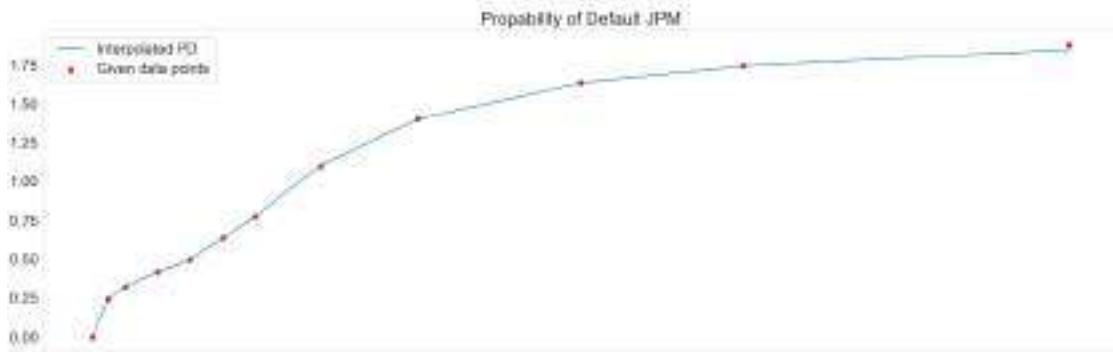


Рисунок 3. Интерпольована ймовірність дефолту

Ми бачимо, що ймовірність дефолту з часом зростає. Це цілком очікувано, оскільки компанія з хорошим кредитним рейтингом має низький ризик дефолту в короткостроковій перспективі, який з часом зростатиме.

Таблиця 2. Значення CVA

CVA	31479
SE для CVA	0.0251 %

Потім, для того щоб зафіксувати кореляцію між зміною спреду CDS та зміною процентної ставки, ми моделюємо фактичні зміни 3-місячної ставки Libor у доларах США протягом періоду свопу, а також обчислюємо ймовірність дефолту на основі фактичних спредів CDS.

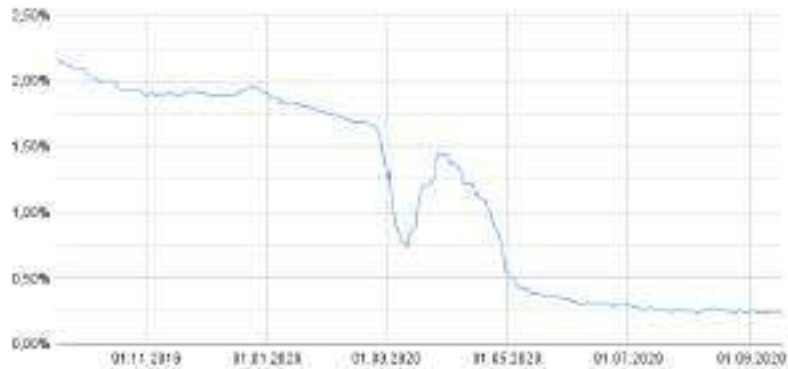


Рисунок 4. Ставка Libor



Рисунок 5. Ймовірність дефолту та ставка Libor

Рис. 5 показує, що існує певна відносно послідовна залежність між зміною спредів CDS та зміною процентних ставок. Процентні ставки знизилися в той же час, коли ймовірність дефолту різко зросла. У результаті, оскільки процентні ставки залишалися низькими, ймовірність дефолту залишалася високою. По суті, тенденція, візуалізована на Рис. 5, вказує на те, що зі зниженням процентних ставок знизилися і кредитна якість контрагента.

4. ВИСНОВКИ

На сьогоднішній день через величезну важливість деривативів для сучасної економіки всім, хто цікавиться міжнародним бізнесом, важливо якісно та ефективно аналізувати невизначеності та володіти знаннями про можливості цих інструментів. Як було зазначено раніше, системний підхід є ефективним інструментом методології досліджень для структурованого дослідження об'єкта як цілісної множини елементів в сукупності відношень і зв'язків між ними. Для ефективного аналізу кредитного та ринкового ризику із застосуванням системного підходу рекомендується СППР. У ході виконання дослідження реалізовано таку систему, яка дозволяє моделювати процентну ставку для подальшої імітації змін на основі єдиного фактору ринкового ризику. За допомогою моделювання значень процентної ставки та ймовірності дефолту ми мали змогу наочно побачити кореляцію між ризиком процентної ставки та ризиком контрагента.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Згуровський М. З., Панкратова Н. Д. Основи системного аналізу: підруч. для студ. вищ. навч. закл. К. : Вид. група ВНУ, 2007. 544 с.
2. Белевець М. С., Бідюк П. І., Кузнецова Н. В. Розробка системного підходу до менеджменту ризиків. Системні науки та кібернетика. 2015. № 1. С. 30–48. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.
3. Бідюк П. І., Кузнецова Н. В., Терентьев О. М. Система підтримки прийняття рішень для аналізу фінансових даних. Наукові вісті НТУУ "КПІ". 2011. № 1. С. 48–61.
4. Brace, A., Gatarek, D., and Musiela, M., .The market model of interest rate dynamics. *Mathematical Finance*, 7(2), (1997), 127–47.
5. Theodore Barnhill and William Maxwell. Modeling correlated market and credit risk in fixed income portfolios. *Journal of Banking and Finance*, 26:347–374, 2002.

МЕТОДИ ПОБУДОВИ ПРОГНОЗУ ФІНАНСОВИХ ДАНИХ. АНАЛІЗ ТА ПОРІВНЯННЯ

Собкович В.С.¹, Тимошук О.Л.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ sobk2015@gmail.com, ² o.tymoshchuk@kpi.ua [0000-0003-1863-3095]

В дослідженні розглянуто задачу прогнозування ціни для акцій компаній, які тісно пов'язані із фінансовими ринками. Для аналізу методів побудови прогнозу часових рядів пропонуються методики для роботи з фінансовими даними. Побудована програмна система прогнозу дозволяє отримати дані прогнозування для фінансових рядів на кілька кроків вперед, що дає можливість безпосереднього використання результатів дослідження для роботи на торговельних фінансових біржах.

Ключові слова: прогнозування фінансових даних, фінансові ринки, методи побудови прогнозу, фінансові часові ряди, порівняння даних.

1. ВСТУП

Економічна стійкість різних компаній залежить від різноманітних показників. Одними із найважливіших є фінансові характеристики. Можливість побудови прогнозу є важливою складовою для ефективної роботи компанії.

Для аналізу обираємо акції компаній JPYX, AABV, SFIX і PNC що мають різні тенденції на ринку. Відповідно вони утворюються на фондових ринках. Також дані ціни є волатильними та будуть постійно змінюватись. Залежить це від багатьох причин, наприклад сезонність, або загальні економічні процеси, світова криза, війна, пандемія, погодні чинники.

Сьогодні галузь торгівлі на фондових ринках залишилась інструментом, що не зазнав значних втрат та зберіг можливість працювати, тому є актуальною побудова прогнозу показників для фінансового ринку.

2. МАТЕМАТИЧНІ МОДЕЛІ ДЛЯ ПОШУКУ РІШЕНЬ В ЗАДАЧАХ ПРОГНОЗУВАННЯ

2.1 Алгоритм ARIMA

ARIMA це клас моделей, які «пояснюють» певний часовий ряд на основі його власних минулих значень, тобто особистих лагів та похибок прогнозу з лагом, щоб можна було використовувати рівняння [1]. Будь-який «несезонний» часовий ряд, який демонструє закономірності і не є випадковим білим шумом, можна моделювати за допомогою моделей ARIMA, що задані формулою (1):

$$\Delta^d y_t = a_1 \Delta^d y_{t-1} + \dots + a_p \Delta^d y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}, \quad (1)$$

де y - є деяким процесом; Δ^d - оператор різниці часового ряду порядку d ; параметри a, β - параметри моделі; ε_t - білий шум. Першим кроком для побудови моделі ARIMA важливо зробити часовий ряд стаціонарним. Найпоширеніший підхід це віднімання. Тобто від поточного значення відняти попереднє значення. Наступним кроком буде

визначення, чи потрібні моделі в цілому лаги для ковзного середнього (КС). Ми можемо дізнатися потрібне значення лагів КС, перевіривши графік часткової автокореляції.

Часткову автокореляційну функцію можна уявити як кореляцію між графіком та її лагом. Відповідна математична формула (2) для часткової автокореляційної функції для третього лагу має вигляд:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + a_3 Y_{t-3}. \quad (2)$$

Тобто у випадку, якщо Y_t — поточний ряд, а Y_{t-1} — це перший лаг від Y , то часткова автокореляція лагу 3 (Y_{t-3}) — це коефіцієнт a_3 у наведеному рівнянні (2).

Шукану автокореляційну функцію в стаціонарному ряді можна виправити, додавши достатню кількість термінів авто регресії. Отже, спочатку ми приймаємо порядок лагів авто регресії рівним стільки, скільки перетинають межу достатності на графіку часткової автокореляційної функції.

Також можна знайти на графіку автокореляційної функції відхилення для кількості лагів ковзного середнього. Це дозволить побудувати більш точне прогнозування.

2.2 Алгоритм LSTM

Проаналізуємо роботу мережі довгої короткочасної пам'яті [2], що показано на Рис. 1.

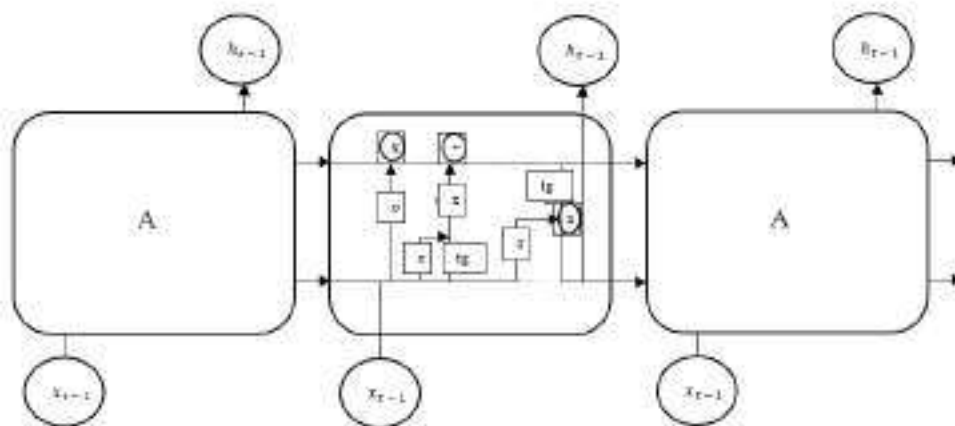


Рисунок 1. Принцип роботи алгоритму LSTM

Основною частиною в алгоритмі роботи є стан горизонтальної лінії, яка проходить через верхню частину блок-схеми. Вона може бути змінена сигмоподібним шаром, котрий передає значення від 0 до 1. Числа будуть нести інформацію про те, скільки попередніх даних ми візьмемо до наступного кроку [3].

Алгоритм роботи мережі. Першим кроком мережа з використанням сигмоїдального шару забирає чи відкидає частину значень. Відповідно, формула для роботи першого рівня описана виразом (3):

$$f_t = \sigma (w_f * [h_{t-1}, x_t] + b_f), \quad (3)$$

де f_t - це значення, котрі передаються до другого рівня роботи; σ - сигмоїдальна функція, що повертає значення від 0 до 1; w_f - вага відповідного нейрона; h_{t-1} - значення, отримані з попереднього кроку роботи мережі; x_t - дані для поточного значення часу; b_f - зміщення. На другому кроці мережа визначає нову інформацію, котру буде зберігати у стані комірки. Спершу сигмоїдальний рівень вирішить, які значення ми оновлюватимемо, а потім шар, що містить тангенс, створить вектор нових значень \hat{c}_t , котрі будуть додаватися до стану.

Значення, які передаються до наступного рівня, обчислюються відповідно до формули (4), а вектор нових значень показаний у формулі (5):

$$i_t = \sigma (w_i * [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\hat{C}_t = \text{tg} (w_c * [h_{t-1}, x_t] + b_c) \quad (5)$$

Надалі буде оновлюватися останній стан C_{t-1} . Він передаватиметься до C_t . Попередній стан буде множитись на f_t (функція воріт, які забувають) та до знайденого числа додаватиметься $i_t * \hat{C}_t$ (дані поточного значення часу помножені на отримані значення попереднього кроку роботи мережі). Отримані значення будуть станом комірки у даний проміжок часу. Формула (6) буде відображати алгоритм:

$$C_t = (f_t * C_{t-1} + i_t * \hat{C}_t). \quad (6)$$

На останньому кроці мережа буде визначати, які значення будуть шуканими. Ці дані будуть залежні від стану клітинки. Перш за все буде працювати сигмоїдальна функція. З її допомогою ми вирішимо, які частини стану однієї клітинки ми виведемо. Далі перевіриться стан клітинки (будуть значення в проміжку між -1 та 1). Тепер ці значення помножаться на результат сигмоїдального шару, це треба для виведення тих значень, які мали залишитися. Відповідні значення для вихідного шару зображені в формулі (7) та отримані результати роботи відповідно у формулі (8):

$$O_t = \sigma (W_o * [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = (O_t * \text{tg} (C_t)). \quad (8)$$

3. ДОСЛІДЖЕННЯ РЕЗУЛЬТАТІВ

Початкові дані представляють собою вибірку часових рядів, акції різних компаній, які мають різні тенденції на ринку. Зокрема, ці тенденції розуміють під собою ведмежий, бичачий ринок, та горизонтальні тренди. Були вибрані акції компаній JPYX, AABV, SFIX і PNC.

Для роботи моделі ARIMA значення p та q, які відповідають за порядок авторегресії і порядок ковзного середнього, використовувалися від 1 до 12. Характеристика d, яка означала даний порядок диференціювання дорівнювала 1 та 2. Модель працює гарно на невеликих проміжках. Прогноз побудований для останніх 15 параметрів. Для результату було отримано значення критерію Акайке та статистика Дарбіна-Уотсона.

Проаналізуємо найкращі моделі для усіх часових рядів відповідно до критерію Акайке (для моделей АРІКС).

Наступним методом прогнозу є модель LSTM.

Для SFIX і JPYX число даних що подається на вхід дорівнює 260 значень (число робочих днів у році), також для AABV обрано 840 значень і для PNC подано 1270 значень (взяті за останні 3 та 5 років відповідно). З них 50 значень будуть в перевірочній вибірці. Розмір партії = 50. Кількість епох дорівнює 25.

Порівняння результатів роботи для SFIX показано у таблиці 1.

Таблиця 2 демонструє порівняння результатів роботи для JPYX.

У таблиці 3 представлено порівняння результатів роботи для даних AABV.

Порівняння результатів для PNC показано у таблиці 4.

Таблиця 1. Результати роботи для SFIX

	ARIMA(8,2,1)	ARIMA(8,1,4)	ARIMA(9,2,2)	ARIMA(6,2,1)	LSTM
R2	0,78228624	0,790338	0,773869	0,781441	0,559485
Sum squared Resid	11,36	11,73	11,72	11,26	13,85260
AIC	165,687634	197,6674	268,3034	262,6322	
DW	2,08357188	1,757118	2,014487	2,023688	
MAE	0,03793743	0,114221	0,080607	0,039669	0,695109

Таблиця 2. Результати роботи для JPYX

	ARIMA(4,1,4)	ARIMA(5,1,3)	ARIMA(5,1,5)	ARIMA(6,1,4)	ARIMA(6,1,5)	LSTM
R2	0,874897018	0,87812228	0,87897618	0,88848263	0,878978	0,654425
Sum squared Resid	45,49595539	44,28559090	44,25381031	40,46509896	44,13075778	201,7806
AIC	108,9511491	109,334338	110,460998	111,432053	113,073335	
DW	2,039194292	2,05487416	2,0020572	2,0266948	1,97412352	
MAE	0,02544935	0,02245648	0,0282241	0,02262397	0,02190825	0,596936

Таблиця 3. Результати роботи для AABV

	ARIMA(2,2,2)	ARIMA(10,2,2)	ARIMA(11,2,4)	LSTM
R2	0,775330093	0,75601934	0,74938959	0,821961
Sum squared Resid	0,00110278	0,00119473	0,00123537	0,02249
AIC	626,9193942	633,469462	634,631295	
DW	2,033178966	1,99047297	2,02201218	
MAE	0,000178533	0,00019655	0,00018912	0,00010725

Таблиця 4 – Результати роботи для PNCD

	ARIMA(9,1,8)	ARIMA(10,1,6)	ARIMA(10,1,10)	ARIMA(11,1,10)	ARIMA(11,1,11)	LSTM
R2	0,847332831	0,846372419	0,853333074	0,849185742	0,842075675	0,9308607
Sum squared Resid	578,49704726	574,00493475	547,0556718	557,37438677	592,903753	7437,0384
AIC	1512,748571	1468,7141449	1468,705809	1471,181233	1472,931583	
DW	2,006802687	1,990320241	1,993016397	1,98793682	1,988749209	
MAE	0,1090303	0,118914	0,114430183	0,092760017	0,099412983	0,020977

У фінансах коефіцієнт R2 вище 0,7 означає високий рівень кореляції. Сума квадратів залишків визначає наскільки добре регресійна модель пояснює або представляє дані в моделі. Інформаційний критерій Акайке буде використовуватись для порівняння

відповідності різних регресійних моделей. Він кращий для моделі де є найнижче значення AIC. В роботі для значень SFIX якщо порівнювати з моделями ARIMA, краще працює модель ARIMA (8, 2, 1). Для значень JPYX найкраща ARIMA (4, 1, 4). Для акцій AABV найкраще буде ARIMA (2, 2, 2). Для значень PNCD найкращою серед авторегресійних інтегрованих ковзних середніх є ARIMA (10, 1, 10). Наступною характеристикою показана Статистика Дарбіна Уотсона. Це є тест на автокореляцію залишків із статистичної моделі або регресійного аналізу. Статистика вимірюється від 0 до 4, також 2 означає, що у вибірці не виявлено автокореляції. Середня абсолютна похибка це показник оцінки моделі, який є середнім абсолютних значень окремих помилок передбачення для значень. Значення найменше для SFIX для ARIMA (8, 2, 1), для JPYX даних для ARIMA (5, 1, 3), для AABV а також для даних PNCD найменше для алгоритму LSTM.

4. ВИСНОВКИ

В роботі було досліджено нейронну мережу LSTM та метод ARIMA.

Порівнявши результати роботи моделей для часового ряду SFIX найкраще працює модель ARIMA (8, 2, 1). Вона дає найкращі значення у порівнянні з моделями ARIMA. Також вона має менші значення середньої абсолютної помилки. Відповідно для часового ряду JPYX найкращі значення дає модель ARIMA (4, 1, 4). Вона має найкращі значення критерію Акайке порівнюючи з іншими моделями та найменшу середню помилку в порівнянні з мережею LSTM. У AABV найкращою є LSTM. Її MAE менша за значення усіх моделей. Для PNCD найкраща модель також LSTM. Де значення середньої абсолютної помилки менші за значення інших моделей відповідно.

Порівнюючи результати зростаючого та спадного ринку, а також часового ряду без тренду можна зазначити що алгоритм ARIMA добре працює на невеликому розмірі даних, в той час, коли алгоритм LSTM дає хороші результати для великих часових рядів. Також точність роботи залежить від початкового набору значень. Ринок має бути зі чітко визначеним трендом, і також для правильності роботи не потрібно обирати нестабільний ринок. Ці тенденції дозволяють покращити отримані значення.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Модель ARIMA [Електронний ресурс] – Режим доступу до ресурсу: <https://www.machinelearningplus.com/time-series/arma-model-time-series-forecasting-python/>.
2. Introduction to LSTM network [Електронний ресурс] – Режим доступу до ресурсу: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/>.
3. Алгоритм LSTM [Електронний ресурс] – Режим доступу до ресурсу: <https://habr.com/ru/company/wunderfund/blog/331310/>.

СИСТЕМНИЙ ПІДХІД ДО АНАЛІЗУ ФІНАНСОВИХ РИЗИКІВ

Танчук М.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

mariya.kovalenko.2000@gmail.com

Ризики є елементами ведення фінансової діяльності та є наслідком невизначеності, тому для ефективного управління ними доцільно використовувати системний підхід, одним із результатів застосування якого може стати створення Системи підтримки прийняття рішень (СППР), яка дасть змогу суттєво полегшити аналіз ризиків. Метою роботи є аналіз фінансових ризиків для покращення їх оцінок із застосуванням системного підходу. Результатом дослідження є СППР, що виконує оцінку ринкового ризику фінансового портфелю із застосуванням методології VAR. У роботі було використано теоретичні та емпіричні методи дослідження.

Ключові слова: фінансові ризики, системний підхід, VAR, CVAR, аналіз ризиків, ринковий ризик.

1. ВСТУП

Категорія ризику займає значне місце в економічній науці й практиці. Ризик - це елемент ведення будь-якої фінансової діяльності. Ризики є результатом невизначеності і являють собою випадкові величини, оскільки зумовлюються наявністю ряду факторів, вплив яких на результати фінансової діяльності не можна заздалегідь точно визначити.

Зважаючи на це, одним із найактуальніших завдань, що обумовлює доцільність дослідження фінансових ризиків та механізмів управління ними, є виявлення економічної сутності фінансових ризиків, їх моделювання та визначення впливу на ведення фінансової діяльності. При цьому, важливим є врахування усіх особливостей та властивостей ризику та розгляд його як цілісної системи взаємопов'язаних елементів, що визначають причини та наслідки появи ризиків [1]. Це можна досягнути шляхом застосування системної методології, яка, до того ж, дозволяє встановити міру і залежність фінансових ризиків від інших видів ризиків, тим самим дає можливість зробити оцінку ризику найбільш точною.

2. МЕТОДИ ОЦІНЮВАННЯ ФІНАНСОВИХ РИЗИКІВ

За одним із варіантів класифікації, фінансові ризики можна поділити на чотири великі категорії: ринковий ризик, кредитний ризик, ризик ліквідності та операційний ризик. Під час розрахунку ризикового капіталу портфелю традиційно розділяють ринковий та кредитний ризики та працюють з ними незалежно [6]. Для визначення рівня можливих втрат та забезпечення інвесторів від потенційних збитків при настанні несприятливих ситуацій на ринку надзвичайно важливим є запровадження ефективного процесу управління ризиками.

Загалом, процес управління ризиками можна поділити на такі етапи. Першим етапом є ідентифікація ризику. Цей етап має на меті виявити існуючі ризики, а також ті, що можуть потенційно виникнути у майбутньому. Наступним етапом є аналіз ризику, метою якого є одержання необхідної інформації про структуру та властивості ризику, що виникає. Оцінювання ризиків являє собою кількісний опис виявлених ризиків, у ході якого

визначаються ймовірність та розмір можливого збитку від конкретного ризику. Останнім етапом управління ризиками є контроль ризику та результатів процесу управління.

Оцінювання ризику полягає у визначенні втрат, пов'язаних з видами ризику, які були виявлені на першому етапі управління ризиком, шляхом побудови економіко-математичних моделей і проведення відповідних розрахунків [1].

Оскільки фінансовий ризик за своєю природою передбачає невизначеність щодо здійснення певного виду фінансової операції у майбутньому, то не завжди вдається отримати ідеальні оцінки [2]. Саме тому, на сьогодні, одним із найактуальніших напрямів розвитку економічної науки є пошук ефективних, науково обґрунтованих підходів та методів аналізу та оцінювання фінансових ризиків.

Одним із таких підходів є системний підхід, який можна визначити як напрям методології досліджень, що полягає у дослідженні об'єкта як цілісної множини елементів в сукупності відношень і зв'язків між ними, тобто розгляд об'єкта як модель системи.

Як зазначалось раніше, оцінювання ризиків відбувається із застосуванням методів, заснованих на побудові певних математичних моделей, які у подальшому використовуються для вимірювання визначених на попередніх етапах видів ризику.

Існує низка методів оцінювання фінансових ризиків, метою яких є побудова функціональної, ефективною та максимально точною моделі для певних категорій ризиків з метою створення точних прогнозів та визначення стратегії мінімізації ризику. До них можна віднести методологію VaR, мережі Байєса, а також деякі математичні моделі для прогнозування втрат, такі як ARCH та GARCH модель.

Великої популярності на сьогоднішній день набуває використання методології Value at Risk (VaR). Це зумовлено тим, що ця методологія володіє рядом безсумнівних переваг, а саме:

- дає змогу оцінити ризик в межах можливих втрат, співвіднесених з ймовірністю їх виникнення;
- вимірювати ризики на різних ринках універсальним чином;
- агрегувати ризики за окремими позиціями в єдину величину (інтегральний ризик) для всього портфеля активів, враховуючи при цьому інформацію про кількість позицій, і волатильності на ринку [4].

Загалом, VaR відображає мінімальний збиток інвестиції, який буде понесений на певному рівні ймовірності (довірчому інтервалі) або максимальний збиток, який буде реалізований на рівні ймовірності. Однак, чим менш стабільна інвестиція, тим більша ймовірність того, що VaR не дасть повної картини ризиків, оскільки ця міра байдужа до всього, що виходить за межі її власного порогу. Саме тому для оцінки екстремальних подій (найгірших випадків) поруч із показником VaR також розраховують показник очікуваних втрат Expected Shortfall, найбільш відомий як CVaR або Conditional Value at Risk (Рис. 1).

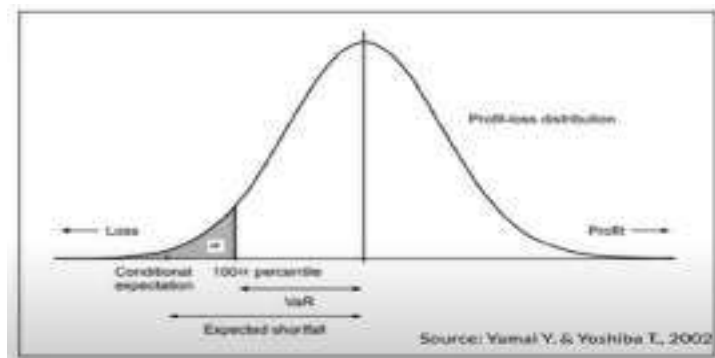


Рисунок 1. Візуалізація VaR та CVaR на графіку розподілу портфелю

До основних методів оцінювання VaR та CVaR відносяться: історичний метод, параметричний метод та метод моделювання Монте Карло, який відрізняється від попередніх більшою складністю обчислень, однак показує більш точні результати. Для полегшення обчислення рівня ризику та уникнення можливих помилок особою, що приймає рішення, було запропоновано розробити систему прийняття рішень (СППР), яка виконуватиме обрахування вартості під ризиком для певного фінансового портфелю.

СППР складається таких основних підсистем: підсистеми інтерфейсу користувача, підсистеми управління даними та основної підсистеми, яка тісно пов'язана із підсистемою функціональних модулів. Найважливішими модулями є модуль попереднього аналізу даних, що виконує завантаження даних, проводить оцінку доходності портфелю та обраховує волатильність. Надалі ці підрахунки використовуються у модулі виконання обчислень, що виконує, власне, оцінку вартості під ризиком для різних методів та заданих глибини аналізу та рівня довіри. Рівень довіри або довірчий інтервал показує, з якою ймовірністю економічний капітал повністю покриє наслідки реалізації ризиків, які будуть меншими за розмір економічного капіталу [4]. Для проектованої СППР було реалізовано можливість встановлення рівнів довіри: згідно з рекомендаціями Базеля II, що складає 99%, а також згідно із системою “Risk Metrics” – 95% [5]. СППР надає можливість обирати для аналізу один із наступних методів: метод параметричної оцінки, найбільш поширений у формі варіаційно-коваріаційної моделі, метод історичного моделювання та метод імітаційного моделювання Монте-Карло, а також виконує порівняння результатів, отриманих під час використання вищевказаних методів. Провівши експериментальні обчислення у спроектованій СППР, можна не лише визначити розмір можливих втрат, а й зробити висновок про те, який з методів обчислення є найкращим для використання на практиці.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для дослідження було сформовано інвестиційний портфель з цін акцій компаній Tesla, Amazon, Apple, Cisco та Microsoft Corp. У якості часового горизонту було взято період у 800 днів. Початкова інвестиція була обрана у розмірі 10 000 \$, а в якості періоду для прогнозування обрано один торговельний рік – 252 дні.

Для обчислення VaR необхідно розглянути розподіл прибутковості портфелю (Рис.2).

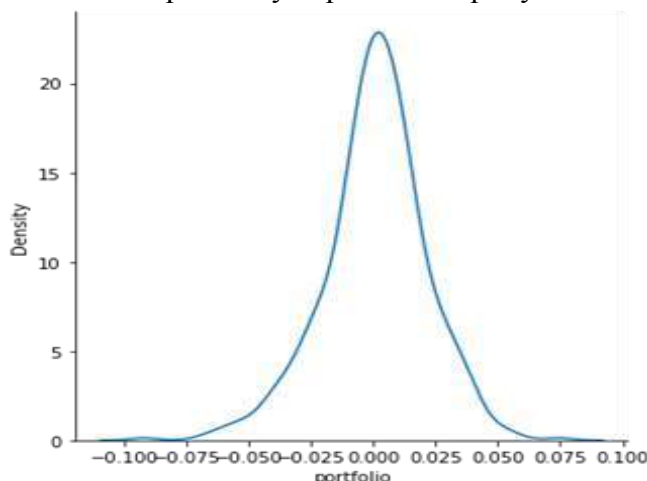


Рисунок 2. Розподіл прибутковості портфелю

Для розрахунку VaR припущення полягає в тому, що доходи розподілені нормально. Перший погляд на прибутковість показує, що щоденні доходи для портфелю розподіляються

приблизно нормально, але хвости більш товсті. Це натякає на доцільність використання поряд із нормальним розподілом розподілу Стюдента під час реалізації параметричного методу оцінки VaR. Використання розподілу Стюдента-t для оцінки VaR може значно покращити оцінку VaR [7].

Найпростішим методом розрахунку VaR є історичний метод, для якого припущення про розподіл доходів не робиться, а моделюється на основі історичних емпіричних даних. Візуалізуємо для цього методу на гістограмі розподілу доходностей портфелю (Рис. 3) підраховані значення VaR та CVaR для рівня довіри 95%.

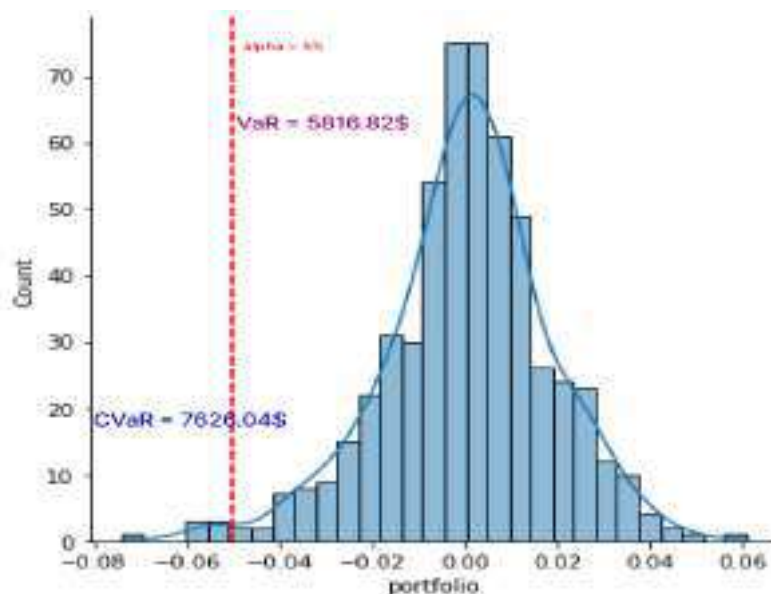


Рисунок 3. Значення VaR та CVaR для $\alpha = 0.05$ для історичного методу

Для параметричного методу робиться припущення щодо розподілу доходів і використовуються історичні доходи портфелю та стандартні відхилення (оцінки) для визначення параметрів моделі. Для порівняння було реалізовано дві параметричні моделі VaR: нормальний розподіл та t -розподіл (жирні хвости).

Найскладнішим у реалізації виявився метод Монте-Карло, який хоч і відрізняється високою обчислювальною складністю, проте вважається найбільш точним та гнучким серед усіх методів обчислення VaR. У даному методі кожен фактор ризику моделюється за допомогою псевдовипадкових чисел. Метод Монте-Карло використовує для побудови траєкторії цін геометричний броунівський рух та дозволяє імітувати будь-який розподіл факторів ризику, включаючи важкі хвости розподілів. При цьому, збільшення кількості симуляцій підвищує точність обчислень. Застосувавши 400 симуляцій, було визначено вартість під ризиком для заданих довірчих інтервалів, а також визначено показники очікуваних втрат (CVaR) за межами інтервалів (Рис. 4).

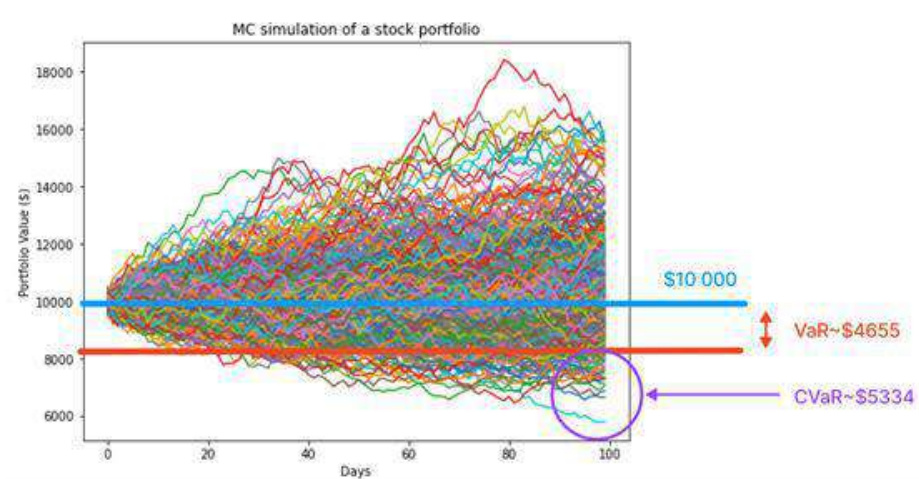


Рисунок 4. Значення VaR та CVaR для методу Монте Карло $\alpha = 5\%$

Результати обчислень VaR та CVaR наведені у таблиці нижче (Табл. 1).

Таблиця 1. Значення VaR та CVaR для різних рівнів довіри та різних методів обчислення

Довірчий інтервал	Метод Знач. VaR/CVaR	Історичний	Параметричний $N(0,1)$	Монте Карло
		95%	VaR	5816.82
	CVaR	7626.04	4935.01	4174.38
99%	VaR	8690.92	5796.53	4754.52
	CVaR	10515.36	6903.9	5258.14

Зазначимо, що величини VaR та CVaR, оцінені за параметричним методом та методом Монте-Карло схожі, порівняно з історичним методом. Історичний метод припускає, що майбутні зміни у курсі акцій будуть подібні до минулих, у той час як метод Монте-Карло додає у передбачення випадковість шляхом застосування випадкової вибірки, що додає аргумент на користь використання саме цього методу для обчислення VaR.

Загалом, VaR надає хороше уявлення про втрати, які відбуваються менше заданого довірчого інтервалу, але не про ті, що лежать за його межею.

Наприклад, для досліджуваного портфелю значення VaR на рівні 95%, підраховане за допомогою методу Монте-Карло, становить близько 3300 доларів США. Це означає, що інвестор повинен мати достатньо капіталу, аби покрити ці втрати. Проте у випадку отримання збитків більшого розміру за межами заданого довірчого інтервалу (95%), ймовірно, інвестор не буде готовий до таких збитків, якщо розглядатиме виключно VaR.

Саме тому для кількісного визначення усіх збитків за межею VaR та врахування втрат у випадку настання найгіршого сценарію, варто застосовувати CVaR. Таким чином інвестор зможе отримати краще уявлення про потенційні найгірші втрати.

4. ВИСНОВКИ

На сьогоднішній день фінансові системи стають все більш складними та невизначеними, тому виникає необхідність знаходження більш ефективних інструментів аналізу та прогнозування фінансових ризиків для якісного управління ризиками.

Одним із найбільш ефективних підходів до оцінювання ризиків на сьогоднішній день є системний підхід, який можна визначити як напрям методології досліджень, що полягає у дослідженні об'єкта як цілісної множини елементів в сукупності відношень і зв'язків між ними. Для ефективно оцінки ризиків із застосуванням системного підходу доцільним є створення інтелектуальної системи підтримки прийняття рішень.

У ході дослідження було реалізовано таку систему, яка дозволяє виконувати оцінку вартості під ризиком для певного інвестиційного портфелю. Ця система використовує методологію VaR та CVaR та реалізує низку методів обчислення значень VaR та CVaR. Результати дослідження показали, що найбільш ефективним методом виявився імітаційний метод Монте-Карло, який дозволяє найбільш точно підрахувати величину можливих втрат. Окрім того, було визначено, що для кращої оцінки ризику для найгірших сценаріїв варто застосовувати CVaR.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кузнєцова Н.В., Бідюк П.І. Теорія і практика аналізу фінансових ризиків: системний підхід, монографія. Київ 2020, 400 с.
2. Куцик П.О., Васильців Т.Г., Сороківський В.М., Стефаняк В.І., Сокорівська М.В. . Управління фінансовими ризиками навч. посібник. Львів: Растр-7, 2016. 318 с.
3. Башкіров О.В. Порівняльний аналіз VAR-методів оцінки ризику активів банку О.В. Башкіров, Проблеми і перспективи розвитку банківської системи України : зб. наук. праць ДВНЗ «УАБС НБУ». – Вип. 14. –С. 302–309.
4. Moraux F. How valuable is your VaR? Large sample confidence intervals for normal VaR. F. Moraux. Journal of risk management in financial institutions. – 2011. – № 4.2. – P. 189–200. – [Електронний ресурс]. – Режим доступу : <https://perso.univ-rennes1.fr/franck.moraux/research/JRMFI.pdf>
5. Longerstaey J. Risk Metrics TM – Technical Document . J. Longerstaey, M. Spencer. Morgan Guaranty Trust Company of New York: NewYork, 1996. [Електронний ресурс]. – Режим доступу : <http://yats.free.fr/papers/td4e.pdf>
6. Theodore Barnhill and William Maxwell. Modeling correlated market and credit risk in fixed income portfolios. Journal of Banking and Finance, 26:347–374, 2002
7. Lin, C. and Shen, S. (2006), "Can the student- t distribution provide accurate value at risk?", Journal of Risk Finance, Vol. 7 No. 3, pp. 292-300. <https://doi.org/10.1108/15265940610664960>

ПРОГНОЗУВАННЯ ІНДЕКСУ АКЦІЙ НА ОСНОВІ ХМАРНИХ ОБЧИСЛЮВАЛЬНИХ СЕРВІСІВ

Тараненко О.С.¹, Бідюк П.І.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹alloha.t.99@gmail.com

Торгівля акціями та цінними паперами набрала великої популярності в останній час. Але такий вид діяльності може нести за собою ризики втрати фінансів. Метою роботи є збір та попередня обробка вхідних даних та розробка моделі для прогнозування індексів акцій на основі хмарних обчислювальних сервісів. Результатом дослідження є навчена модель у вигляді багатошарової, згорткової нейронної мережі та мережі довгої короткочасної пам'яті із використанням хмарного сервісу AmazonWebServices. В роботі було використано емпіричні й теоретичні методи дослідження.

Ключові слова: індекс акції, машинне навчання, згорткові нейронні мережі, багатошарова нейронна мережа, мережа довгої короткочасної пам'яті, AmazonWebServices

1. ВСТУП

Будь-які дії, пов'язані з цінними паперами, корелюють з певними ризиками, тому вони потребують точних розрахунків і точного аналізу. Хмарні сервіси забезпечують швидке та гнучке розгортання інфраструктури, забезпечують їх повсякчасну доступність. В якості архітектури доцільно використовувати нейронні мережі, бо вони не обмежуються типами і кількістю даних.

Із цього випливає необхідність у розробці моделей прогнозування індексу акцій, побудованих на основі штучних нейронних мереж, що використовують хмарні сервіси.

2. СХЕМА РЕАЛІЗАЦІЇ МОДЕЛІ СИСТЕМИ ПРОГНОЗУВАННЯ ІНДЕКСУ АКЦІЙ В РЕАЛЬНОМУ ЧАСІ НА БАЗІ ХМАРНИХ ОБЧИСЛЮВАЛЬНИХ СЕРВІСІВ

Розглянемо схему реалізації моделі системи прогнозування індексу акцій на базі хмарних обчислювальних сервісів, функції її компонентів та логіку їх взаємодії (рис. 1). Система базуватиметься в Amazon Web Services, тому усі умовні позначення на схемі відповідають умовним позначенням із документації Amazon Web Services.

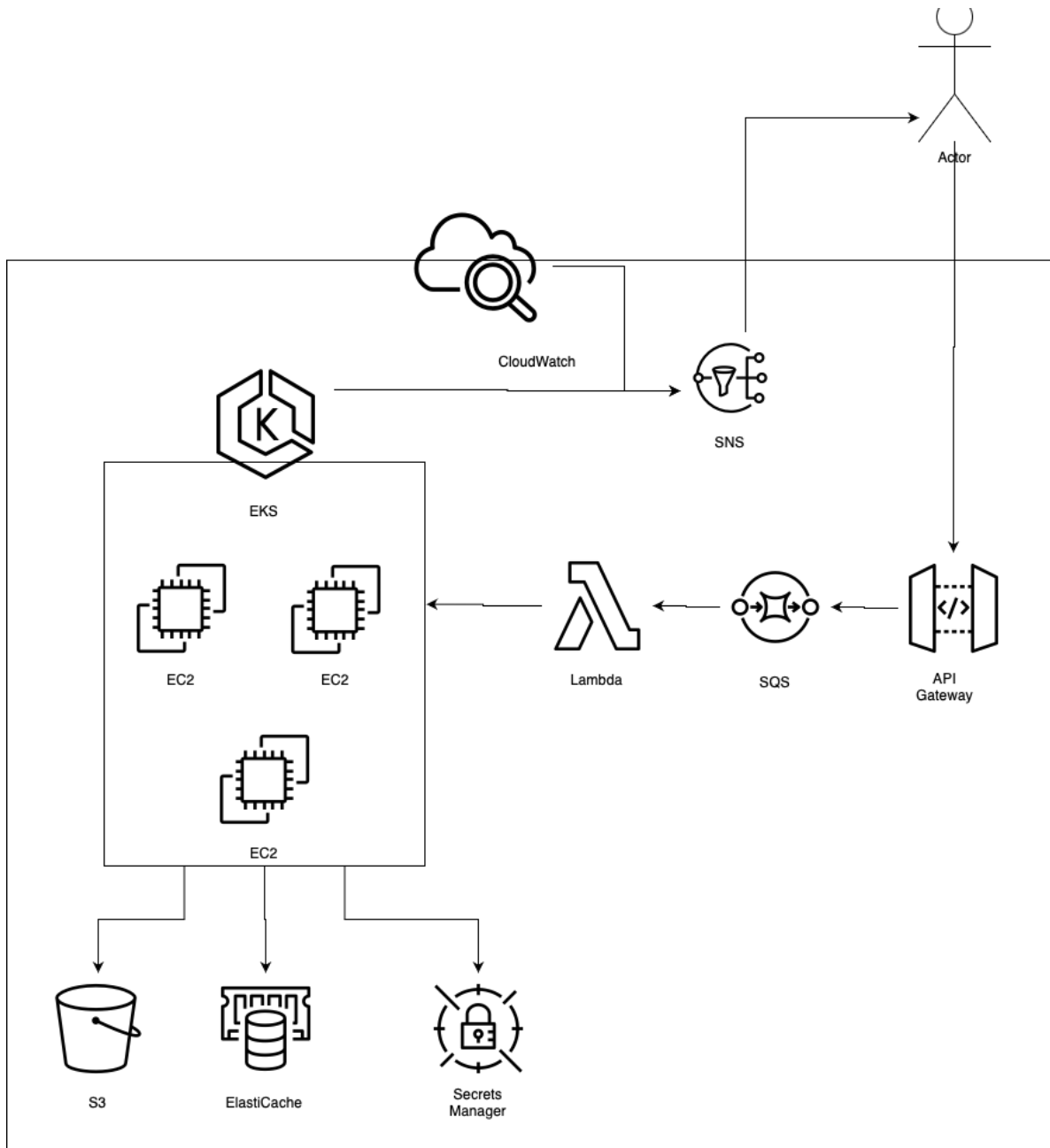


Рисунок 1. Схема реалізації моделі системи прогнозування індексу акцій на базі хмарних обчислювальних сервісів

Розглянемо взаємодію компонентів системи.

1. Користувач за допомогою командного рядка, або графічного інтерфейсу надсилає запит на зміни до системи на ендпоінт Amazon API Gateway.

2. На рівні Amazon API Gateway налаштоване правило фільтрації кількості запитів, щоб унеможливити DDoS атаку на сервіс.

3. Від Amazon API Gateway повідомлення складається у чергу Amazon SQS, яку налаштовано працювати за принципом FIFO, що забезпечує збереження порядку надісланих повідомлень.

4. Подія наявності у Amazon SQS черзі повідомлень запускає роботу Amazon Lambda, яка робить перевірку запитів з точки зору безпеки. Це може досягатись за допомогою

використання різних паролів, 2FAключів, або підтверджуючими запитами до джерела, яке має право надсилати запити.

5. Окрім того, у разі декількох одночасних запитів можливо сортувати їх за пріоритетом. До прикладу запити, які відповідають за фундаментальні зміни в системі, такі як її термінове вимкнення, будуть опрацьовуватись позачергово.

6. Потім запит направляється до Kubernetes кластера (Amazon EKS), де безпосередньо опрацьовується і вносить зміни по роботі системи.

7. Kubernetes кластер використовуватиме безпосередньо збір даних про індекс акцій, їх обробку, насичення даних додатковою інформацією, передбачення майбутньої ціни та реалізовуватиме стратегію купівлі та продажу акцій.

8. Для зберігання великих обсягів даних, які не потребуватимуть миттєвого доступу, використовуватиметься Amazon S3 (Simple Storage Service). А саме такі дані, як давньо історичної ціни на акції, застарілу додаткову інформацію про акції та про компанії, результати старих передбачень та інше.

9. Для зберігання даних, до яких потрібен миттєвий доступ буде використовуватись ElastiCache for Redis, що забезпечить миттєвий доступ до інформації, необхідної для роботи системи.

10. Для того, щоб Kubernetes кластер міг безпечно отримати доступ до ключів доступу до необхідних для роботи сервісів, використано Amazon Secrets Manager, в IAM полісі якого буде прописано, що доступ до секретних значень матиме лише необхідний Kubernetes кластер.

11. Для простого поверхневого моніторингу системи використано Amazon CloudWatch, який дозволить своєчасно виявити та проаналізувати проблеми функціонування системи.

12. В разі виникнення проблеми (сервіс Amazon CloudWatch) або в разі потреби сповістити користувача про певну подію (сервіс Kubernetes) направить повідомлення до сервісу Amazon SNS, який доставить його користувачу одним з обраних методів (Email, SMS, HTTPS, Push).

Для швидкого розгортання системи буде використовуватись IaC (Infrastructure as Code) інструмент Terraform, розроблений компанією HashiCorp. Це декларативний безкоштовний програмний продукт, який за допомогою стандартизованого синтаксису дозволяє відправити API запит до провайдера (Amazon Web Services) та за лічені хвилини створити усі інфраструктурні елементи з необхідними налаштуваннями. Це дозволить прописавши усі необхідні компоненти лише раз розгорнути дану інфраструктуру в будь-якому обліковому записі Amazon Web Services.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для реалізації компоненту системи, який відповідає за прогнозування, використано декілька поширених варіацій нейронних мереж, а саме багатошарова нейронна мережа, згортова нейронна мережа та мережа довгої короткої пам'яті. Для реалізації цих мереж використовувалась мова прогнозування Python та бібліотеки Pandas, Numpy, Sklearn, Keras та Tensorflow, а для візуалізації використовувалась бібліотеки Matplotlib та Seaborn. Метою даного дослідження є реалізувати декілька поширених архітектур нейронних мереж та обрати найбільш придатну для задачі прогнозування в реальному часі, а саме ту, яка має оптимальну точність та швидко працює. Для прогнозування використовувались по хвилинні дані індексу S&P500.

Результати роботи багатошарової нейронної мережі подано на рисунках 2-4.

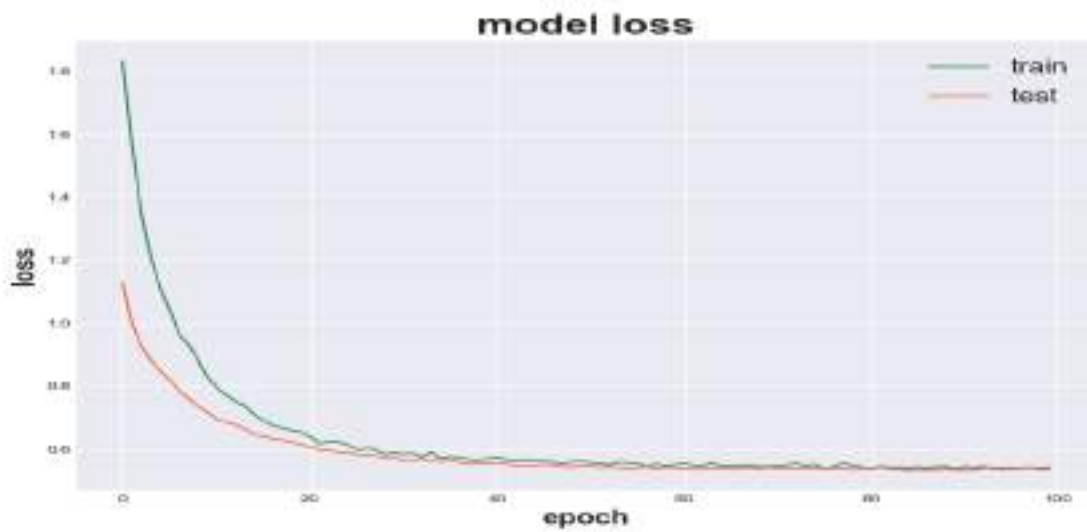


Рисунок 2. Динаміка зміни похибки багатосарової нейронної мережі

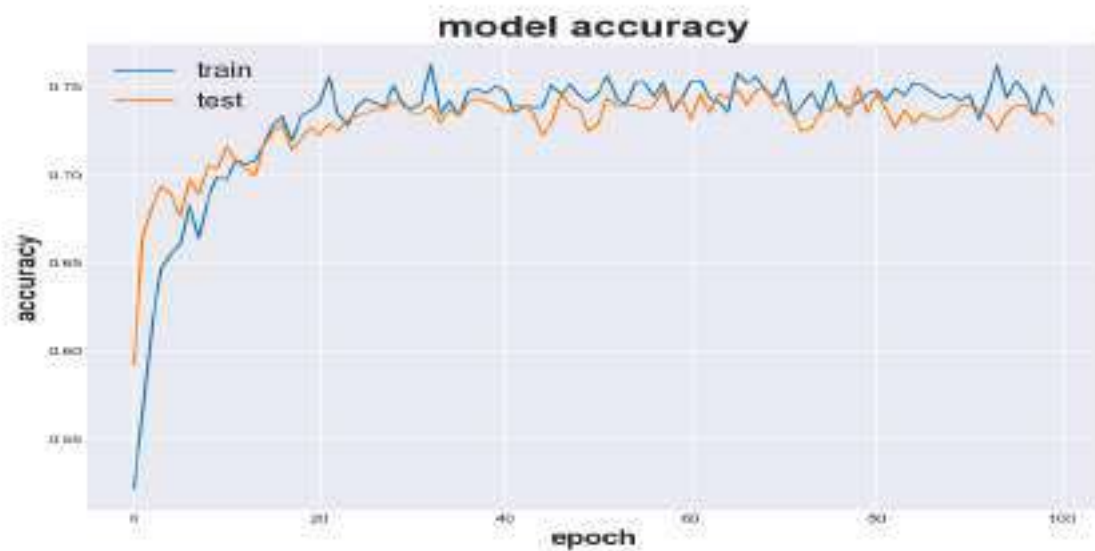


Рисунок 3. Динаміка зміни точності багатосарової нейронної мережі

	precision	recall	f1-score	support
0	0.73	0.76	0.75	242
1	0.74	0.71	0.73	234
accuracy			0.74	476
macro avg	0.74	0.74	0.74	476
weighted avg	0.74	0.74	0.74	476
True positive =	0.76			
False positive =	0.248			
True negative =	0.714			
False negative =	0.277			

Рисунок 4. Результати роботи багатосарової нейронної мережі

4. ВИСНОВКИ

В поданій роботі проведено огляд засобів аналізу та прогнозування індексу акцій та можливості хмарних обчислювальних сервісів, а також було розроблено модель системи прогнозування індексу акцій в реальному часі на базі хмарних обчислювальних сервісів та реалізовано прототип компоненту прогнозування індексу акцій. В результаті модель системи передбачає автономну роботу системи з можливістю віддаленого керування цією системою за допомогою вхідних API запитів. Базуватиметься система на хмарному сервісі Amazon Web Services.

Також розроблено і протестовано компоненту для прогнозування індексу акцій. Були побудовані багатошарова нейронна мережа, мережа довгої короткої пам'яті та згортована нейронна мережа, які аналізували по хвилинні дані індексу S&P500. Вони передбачали, як чи зросте або впаде індекс S&P500 через 5 хвилин після моменту прогнозування на основі даних за останні 30 хвилин. Звичайна багатошарова нейронна мережа показала найкращий час та трохи кращий результат, тож для даного компонента доцільно обрати її.

В майбутньому можна вдосконалити систему реалізувавши інші компоненти моделі, такі як Engicler, що дасть системі додаткові дані про компанії, акції яких вона намагається передбачити.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про цінні папери і фондовий ринок: Закон України від 23.02.2006 р. – №3480-IV. URL: <https://zakon.rada.gov.ua/laws/show/3480-15#Text>.
2. Про акціонерні товариства : Закон України № 514-VI від 17.09.2008 // Відомості Верховної Ради України. 2008. № 50–51.
3. AWS Services List - Top 10 AWS Services. URL: <https://mindmajix.com/top-aws-services>
4. Amazon API Gateway features. URL: <https://aws.amazon.com/api-gateway/features/>
5. Top programming languages. Most popular and fastest choices for developers. URL: <https://www.zdnet.com/article/top-programming-languages-most-popular-and-fastest-growing-choices-for-developers/>
6. Amazon Lambda features. URL: <https://aws.amazon.com/lambda/features/>
7. Serverless computing with AWS Lambda. URL: <https://www.infoworld.com/article/3210726/serverless-computing-with-aws-lambda.html>.
8. Amazon Leads \$150-Billion Cloud Market. URL: <https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/>
9. Google Cloud vs AWS: Difference Between AWS and GCP. URL: <https://www.guru99.com/google-cloud-vs-aws.html>.
10. Засоби прогнозування індексу акцій на основі апарату штучних нейронних мереж, 2020. URL: <https://ela.kpi.ua/handle/123456789/37035>

ІНСТРУМЕНТИ АНАЛІЗУ ПРОЕКТНИХ РИЗИКІВ

Гогуля О.О.¹, Тимошук О.Л.²

Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна

¹aleksandragogulia@gmail.com,

²o.tymoshchuk@kpi.ua [0000-0003-1863-3095]

Можливість визначати та управляти ризиками є досить актуальною задачею сьогодення. Багато процесів наразі є досить непередбачуваними, тому мати можливість передбачити та мінімізувати ризики є досить важливим процесом. Метою роботи є дослідження інструментів для аналізу ризиків та їх порівняння. Результатами роботи є застосування методів на конкретному процесі однієї із компаній.

Ключові слова: проектні ризики, ризик-менеджмент, аналіз ризиків.

1. ВСТУП

Управління ризиками є досить актуальною задачею сьогодення. В сьогоденнішніх реаліях існує багато нестабільних та неконтрольованих процесів, на які досить важко повпливати. Тому можливість передбачати та контролювати ризики є важливою сферою, яка активно розвивається.

Невизначеність та пов'язані з цим ризики важлива частина управління проектами на сьогодні. В роботі було проведено задачу ідентифікації ризиків для конкретного проекту та аналіз вибору ефективних методів та стратегії реагування на них.

2. МЕТОДИ АНАЛІЗУ РИЗИКІВ

На практиці ризик-менеджмент виділяється в окремий специфічний напрям діяльності. Під цим поняття варто розуміти систему управління ризиками проекту, яка передбачає вибір та застосування інструментів та методів, спрямованих на знаходження ризиків, обрахунок імовірності їх настання, визначення фінансових величин втрат (або прибутку у випадку позитивного впливу ризику на проект) та нейтралізацію самих ризиків [1].

Головною ціллю управління ризиками є мінімізація впливу ризику на фінансові втрати. Існують наступні етапи управління ризиком [2]:

- Ідентифікація та аналіз ризику.
- Аналіз альтернативних методів управління ризиком.
- Вибір методів управління ризиком.
- Виконання обраного методу управління ризиком.
- Моніторинг результатів і вдосконалення системи управління ризиком.

Розглянемо кожен із них детальніше.

Під ідентифікацією та аналізом ризиків можна розуміти процес виявлення та опису ризиків, розбір їх специфіки, яка виражається певними притаманними рисами, за допомогою яких їх можна охарактеризувати та віднести до певної групи. [3] Також це включає дослідження розмірів збитку, потенційною зміною ризиків в часі, взаємозв'язками між ними та вивчення факторів, які на них впливають. Даний етап забезпечує інформаційну базу для

проведення подальших етапів ризик-менеджменту. Для реалізації цього етапу процесу використовуються наступні методи:

- метод Дельфі;
- мозковий штурм;
- метод номінальної групи;
- аналіз схожих проектів в тій ж, або суміжній областях.

Наступним етапом є аналіз альтернативних методів управління. Метою цього етапу є дослідження інструментів, які можна застосовувати для усунення реалізації та впливу ризику на процеси проекту та різні його вкладові. Варто зазначити, що методи управління будуть обиратись індивідуально для кожного з типів ризиків.

Етапом вибору методів управління ризиком дозволяє сформулювати загальну стратегію по боротьбі з ризиком для всього життєвого циклу проекту. Вибір буде робитись на основі доцільності та ефективності застосування того чи іншого методу управління до кожного конкретного ризику з урахуванням обмежень у фінансових ресурсах та загальним впливом ризику на фінальний результат проекту.

Наступний етап передбачає пряме застосування обраних методів у випадку настання тієї чи іншої ризикової ситуації. Тому на даному етапі важливо буде максимально готовим швидко та ефективно застосовувати обрану стратегію, адже зазвичай присутня обмеженість в часових ресурсах. Також важливо визначитись з відповідальними за кожне рішення та з тими, хто буде контролювати процес реалізації стратегії управління ризиком.

Останнім етапом є моніторинг результатів та вдосконалення системи управління ризиком. Це передбачає ітеративний підхід з кожним наступним застосуванням обраної стратегії управління ризиком. Важливо розуміти, що цей етап може повпливати та внести зміни в усі попередні. Ми отримуємо та аналізуємо інформацію про ризики, наскільки ефективним був обраний метод для їх управління, який збиток був заданий проекту.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Було запропоновано алгоритм аналізу ризиків при плануванні процесу розширення компанії та переходу її на новий рівень. Хоч алгоритм було застосовано на даному проекті, він може бути валідним і для багатьох інших.

Розглянемо кроки, завдяки яким даний алгоритм реалізується:

- складення плану проекту;
- ідентифікація можливих ризиків на кожному з етапів;
- за допомогою методу Дельфі обраховуються ймовірності настання ризику;
- побудова матриці ризиків дозволяє визначити вплив ризику на проект в разі його настання;
- створення методів реагування на кожен із ризиків.

Було проведено аналіз ризиків в різних сферах управління компанією. Було побудовано таблицю ризиків, розглянемо один із них для прикладу.

Таблиця 1. Опис ризику, сфери його впливу та відповідальних експертів

Сфера ризику	Ризик	Номер ризику	Відповідальні
Найм спеціалістів	Найм некваліфікованого спеціаліста призведе до затримки роботи над проектом та грошових втрат на конкретному проекті.	1	- Recruiting Team - СТО

Після проведення серії опитувань серед експертів було складено таблицю з оцінкою ймовірності кожного з ризиків.

Таблиця 2. Оцінка ймовірності настання ризику кожним із експертів

Номер ризику	Відповідальний	Компетентність експерта	Оцінка ймовірність
1	Recruiting Team	0.8	0.05
	СТО	1	0.02

Після проведення всіх обчислень за формулами отримуємо наступну таблицю результатів

Таблиця 3. Проста оцінка, середньозважена оцінка настання кожного із ризиків

Номер ризику	Проста оцінка	Середньозважена оцінка
1	0.035	0.033



Рисунок 1. Зображення побудованої матриці ймовірностей для проекту компанії

Матриця ризиків зображає ступінь впливу ризиків на проект та показує які з ризиків мають найбільший вплив та потребують нагального реагування та чіткого плану протидії [4].

На основі попередньо проведених досліджень на складених було розроблено рекомендації для кожного з ризиків та визначено критичність їхнього впливу.

4. ВИСНОВКИ

Кожного дня кількість невизначеностей в сфері управління ризиками зростає, тому є велика необхідність створення та застосування ефективних алгоритмів для цього.

У ході написання роботи було розроблено алгоритм, який дозволяє зробити аналіз проектних ризиків, визначити ймовірність їх настання та їх впливу на подальший розвиток

проекту. Результатом роботи стали дослідження, які наводять ймовірність настання того чи іншого ризику, визначені відповідальні експерти, які допоможуть з керуванням ризиків та розроблено рекомендації по впливу на ризики.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Фінансовий менеджмент Г.Г. Кірейцев, Житомир: ЖІТІ, 2001. – 440 с.
2. Етапи управління ризиком. URL: https://stud.com.ua/19859/strahova_sprava/etapi_upravlinnya_rizikom.
3. Бабаєв В.М. Навчальний посібник для студентів спеціальності 87 «Управління проектами». Харків: ХНАМГ, 2006. 244 с.
4. Guide to the Project Management Body of Knowledge (PMBOK® Guide)