

**АНАЛІЗ НА ОСНОВІ БАЗ ДАНИХ (DATA SCIENCE)
ДЛЯ УПРАВЛІНСЬКИХ РІШЕНЬ
З УРАХУВАННЯМ НЕВИЗНАЧЕНОСТІ МАКРО-
ТА МІКРОЕКОНОМІЧНОГО РІВНІВ**

Радіонова Ірина Федорівна,

*доктор економічних наук, професор,
ДВНЗ Київський національний університет імені Вадима Гетьмана,
ВНЗ «Університет економіки та права «КРОК»,
ORCID: <https://orcid.org/0000-0002-0941-2867>,*

Фаренюк Яна Валеріївна,

*аспірантка,
Київський національний університет імені Тараса Шевченка,
ORCID: <https://orcid.org/0000-0001-6837-5042>*

Анотація. Цей підрозділ монографії присвячено важливій науковій та прикладній проблемі пояснення та оцінювання економічної невизначеності для прийняття управлінських рішень. Сформульовано й перевірено припущення про іманентність для вирішення цієї проблеми інструментарію Data Science. Висновки щодо придатності такого інструментарію верифіковано із застосуванням модельних конструкцій макроекономічного та мікроекономічного рівнів.

Ключові слова: інструментарій Data Science, Data Mining, моделі нейронних мереж, регресійні моделі, моделі тактичного планування діяльності.

Актуальність досліджуваної проблеми пояснюється тим, що явище невизначеності національної економіки передбачає використання особливого інструментарію аналізу. Цей інструментарій має бути складнішим та витонченішим, ніж той, що придатний до застосування у рівноважних економіках зі стійким економічним зростанням. Українська економіка, за характером змін та за діючими механізмами відтворення, є економікою невизначеності. Тому управлінські рішення на загальнонаціональному рівні та на рівні окремих економічних суб'єктів передбачають адекватне реагування на фактори невизначеності, ідентифіковані з використанням особливого інструментарію.

Одним із головних *припущень* дослідження є ідея про те, що порівняно більш відповідним (релевантним) для аналізу економіки невизначеності є інструментарій Data Science. Саме це припущення верифікується у цьому дослідженні та на підставі верифікації зроблено певні управлінські узагальнення та висновки.

Метою дослідження є виявлення на конкретних прикладах аналізу та уточнення можливостей Data Science в оцінюванні економічної

невизначеності. Йдеться про невизначеність на мікроекономічному та макроекономічному рівнях. Ця мета досягається у процесі застосування інструментарію Data Science під час аналізу значних масивів економічної інформації з використанням доступного програмного забезпечення.

Достатньо поширеним та таким, що не викликає заперечень, є визначення Data Science як науки про роботу зі значними масивами (базами) даних з метою виокремлення неочевидної (прихованої за великою кількістю подій) інформації про існуючі зв'язки та залежності. Зрозуміло, що віднайдена неочевидна інформація має стати підґрунтям для прийняття більш обґрунтованих, і тому результативніших, управлінських рішень.

Для усвідомлення можливостей Data Science в аналізі невизначеності доцільно деталізувати зміст цієї науки з виокремленням її структурних елементів (складників) (рис. 1).

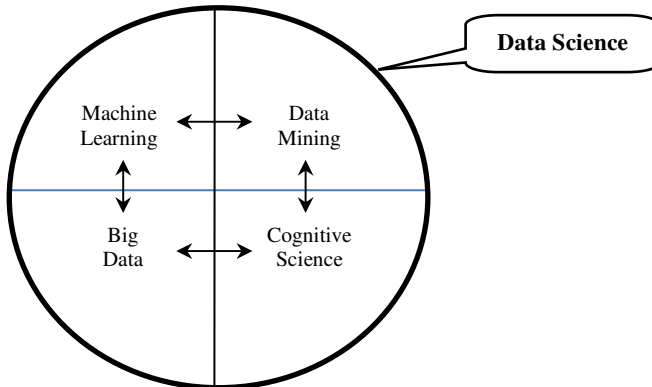


Рис. 1. Зміст Data Science за структурними елементами

Джерело: створено авторами самостійно на основі [1-5].

За рис. 1, Data Science є *інтегральною* аналітичною наукою. Вона формується при безпосередньому використанні здобутків принаймні чотирьох напрямків наукових досліджень, що ґрунтуються на певних ідеях та інструментарії, а саме:

- Cognitive Science – наука про загальні закономірності мислення та алгоритми пізнання світу, які реалізуються в різних галузях знань, до прикладу, в праві, фінансах, медицині, мистецтві, економіці тощо;

- Machine Learning – наука про ефективні способи навчання та самонавчання систем з використанням створених комп'ютерних програм, які застосовують алгоритми пізнання світу, включно з побудовою штучних нейронних мереж (Artificial Neural Network) [6];

- Big Data – наука про технології генерування та правила організації великих баз даних, їх візуалізацію та репрезентацію;

– Data Mining – наука про виокремлення з великих баз даних, за правилами програмового (машинного) навчання, неочевидних, але об'єктивно існуючих, отже, корисних для усвідомленого прийняття рішень, зв'язків (залежностей) та *закономірностей*.

Здатність Data Science забезпечувати відповідний (релевантний) аналіз економіки у стані невизначеності передусім пов'язана з виявленням *неочевидних* зв'язків (залежностей).

На наш погляд, неочевидні зв'язки (залежності) мають такі типові критеріальні ознаки:

– не перебувають «на поверхні» явищ, тому не виявляються емпіричним шляхом або «на практиці», як, наприклад, ідентифікується достатньо очевидний зв'язок між попитом та цінами конкретного ринку, між витратами виробництва та випуском тощо;

– не впливають з масово усвідомлених теоретичних конструкцій, які зазвичай ґрунтуються на логічному (дедуктивному) виведенні одного явища з іншого, як, наприклад, виведення інфляції з необґрунтованого (несподіваного) зростання окремих елементів сукупного попиту або вплив девальвації національної валюти на діяльність експортерів;

– спричиняються дискретно випадковими факторами («чорними лебедями»), які через обмеженість інтелектуальних можливостей спільнот і пересічних економічних суб'єктів, а також через обмеженість технічних можливостей аналізу не можуть бути передбачені з необхідною точністю;

– не є очікуваними, тобто такими, до яких економічні суб'єкти можуть пристосуватися адаптивним, раціональним, псевдораціональним або іншим шляхом;

– не можуть бути виявлені традиційними методами обробки інформації й потребують досконалішого інструментарію, який або ще не створено, або створено, але не використовується через неусвідомлення його переваг та можливостей.

Є підстави припускати, що вказані ознаки неочевидності зв'язків (залежностей) процесів та явищ формують осердя *економіки невизначеності*. В українській економіці до зафіксованих нами критеріальних ознак неочевидності додається й невизначеність зв'язків, породжена незавершеністю об'єктивно необхідних перетворень (реформ), хаотичністю в діях владних інститутів, відсутністю спадковості, отже, перервності (дискретності) позитивних традицій публічного управління економікою тощо.

Data Science-аналіз спирається на науковий фундамент різних шкіл та напрямків економіко-математичних досліджень, у межах яких вирішується завдання виявлення й оцінювання невизначеності.

Ретроспективний аналіз дає підстави для виокремлення певних етапів еволюції досліджень явища невизначеності:

– теорія ігор, ідеї якої сформульовано у 1940-х рр. Дж. фон Нейманом і О. Моргенштерном, а завданням визнано аналіз економічної

поведінки та прийняття рішень в умовах невизначеності та конфліктних ситуацій;

– економіко-математичне моделювання 1950-х – 1980 рр. з його новим для того періоду інструментарієм, який охоплює принаймні шість напрямків:

1) динамічні та стохастичні моделі, що враховують фактор випадковості та невизначеності (Р. Фріш, Я. Тінберген, С. Кузнець, К. Ерроу, Л. Клейн та ін.);

2) ітеративні методи для розв'язання задач великої розмірності (Н. Н. Боголюбов, Н. М. Крилов, М. Ф. Кравчук, А. Ю. Лучка);

3) методи сіткового планування (критичного шляху СРМ, оцінки та аналізу програм PERT), що дозволяють мінімізувати тривалість проєктів із урахуванням різноманітних чинників впливу;

4) моделі одночасних рівнянь, що передбачають охоплення системою окремих рівнянь з різним рівнем ідентифікації та різними залежними змінними, що пов'язані між собою (Т. Купманс, Т. Хаавельмо, Г. Гейл та ін.);

5) удосконалені моделі економічного розподілу та оцінки ресурсів, зокрема оцінки капіталовкладень з позицій теорії двоїстості (Г. Лемке, А. У. Таккер та ін.);

6) оптимізаційні моделі перспективного планування, які використовуються при розробці більш обґрунтованих перспективних планів господарського розвитку на основі пізнання найважливіших економічних пропорцій і співвідношень, темпів зростання виробництва й рівнів споживання, раціональної галузевої структури (Л. В. Канторович, В. С. Немчинов, В. Л. Макаров, А. Б. Горстко);

– моделі поведінки соціально-економічних систем та оцінювання їх керованості, що активно розробляються з кінця ХХ ст. і передбачають використання математичного інструментарію принаймні у таких двох особливих формах:

1) лінійних та нелінійних диференціальних рівнянь для ідентифікації ефектів запізнення в реакції систем на впливи (В. Вольтерра, А. Д. Мишкіс, Н. Н. Боголюбов, Д. Я. Хусаїнов, В. І. Фодчук);

2) нечіткої логіки з формалізацією у вигляді нечітких множин на основі експертного оцінювання (Л. Заде).

Покажемо, що певний потенціал для оцінювання невизначеності формують і достатньо прості інструменти економетричного аналізу. Це, зокрема, стосується множинної регресії – методу багатовимірного аналізу, за допомогою якого залежна змінна пов'язується з сукупністю незалежних змінних (регресорів). Дослідження інструментарію дає підстави стверджувати, що власне еволюція регресійного аналізу відбувалася в бік віднайдення способів оцінювання невизначеності. Йдеться про перехід від простої лінійної регресії до

нелінійної, а також до авторегресії з лаговими значеннями змінних для оцінювання тривалості імпульсних «відгуків» та до використання в регресійних рівняннях, так званих «фіктивних змінних».

Кожен згаданий інструмент економетричного аналізу, забезпечуючи певні переваги, має й обмеження. Таким типовим обмеженням є те, що за надмірно великих обсягів інформації виникає потреба в її обмеженні та спрощенні. Це спричиняє деформоване модельне відображення дійсності й, відповідно, недосконалість рішень, які спираються на модельні конструкції. Саме це обмеження частково долається, наприклад, з використанням штучних нейронних мереж (Neural Networks). Економетричні моделі та штучні нейронні мережі є інструментами Machine Learning, а, отже, складником Data Science (рис. 1).

За логікою моделей штучних нейронних мереж, вхідні сигнали складаються та, проходячи через передавальну (трансмісійну) функцію, генерують вихідний результат. Вхідні сигнали формують певні прошарки (шари, або «вузли») взаємодії, тому система цієї взаємодії є багаторівневою.

Штучні нейронні мережі вирішують багато завдань, мають більші, ніж інші інструменти, аналітичні можливості. Найважливішою, з огляду на мету нашого дослідження, є можливість їх використання для аналізу з неповною інформацією при присвоєнні кожній операції певного рівня ризику [7]. Саме у такий спосіб формуються додаткові можливості аналізу *невизначеності*.

Застосування інструментарію Data Science в економічних дослідженнях загалом та дослідженнях на макроекономічному рівні зокрема вже має власну історію і традицію. Найчастіше фокусом дослідження з використанням інструментарію Data Science ставали такі об'єкти:

- волатильність курсів валют у двосторонніх відносинах між країнами [8];
- прибутковість фінансового ринку, динаміка фондових індексів [9];
- динаміка ВВП, економічні шоки та «поворотні» точки у процесі економічних коливань [10; 11];
- фактори цінних коливань та динаміка інфляції [12; 13];
- ефективність окремих видів економічної політики, зокрема з використанням конструкцій, орієнтованих на агентів і творців політики (*agent-based modelling in economics and policy-making*) [14; 15].

Застосування Data Science на **макроекономічному рівні** ілюструватимемо на прикладі побудови нейронних мереж у дослідженні двох макроекономічних явищ – *економічного зростання* та *диференціації доходів*. Як і всі інші процеси та явища української економіки, обрані нами для дослідження змінні мають високий рівень невизначеності. Зокрема, тренд на економічне зростання у 2016 – 2018 рр. несподівано переривається гальмуванням у 2019 р. та значним спадом 2020 р.

Останній не пояснюється виключно факторами Covid-19 та світовою кризою, а має в українській економіці та суспільстві й інші особливі суттєві внутрішні причини.

У дослідженні феноменів економічного зростання та рівня диференціації доходів використовувалися дані за 31-им показником української економіки за період 1992 – 2018 рр. Для забезпечення однорідності статистичної інформації було максимально використано наявні в базах МВФ [16] дані по Україні. З метою продовження рядів даних за 1990-і рр., які не містилися в інформації МВФ, використані два інші джерела за шість років [17; 18].

Усі наявні показники розподілено на ендогенні, тобто ті, що пояснюються в моделях, та п'ять груп показників екзогенних, тобто тих, з використанням яких відбувається пояснення. Такими ендогенними показниками визначено два, а саме: темп економічного зростання та коефіцієнт Джинні. Відбір цих показників узгоджується з нашою метою пояснити економічне зростання та диференціацію доходів в українській економіці з використанням інструментарію Data Science. Ендогенні змінні розподілено на п'ять груп:

- 1) досягнутого рівня розвитку національної економіки;
- 2) соціальні;
- 3) фінансові;
- 4) монетарні;
- 5) ресурсного потенціалу економіки.

Розподіл екзогенних змінних моделей на п'ять груп є достатньо умовним і таким, що не має суттєвого технічного навантаження при побудові моделей та поясненні результатів.

Умовність поділу показників з бази досліджуваних даних виявляється, зокрема, у тому, що частина з них може бути віднесена як до однієї, так і до іншої групи. До прикладу, ми виходимо з того, що два показники фінансування освіти – G_{educ}/G та $Exp_{G/st}/Y_N$ – належать до показників досягнутого рівня розвитку. Підставою для цього є уявлення, що саме рівень і якість фінансування освіти, зокрема освіти вищої (університетської), найбільше віддзеркалює рівень розвитку і економіки, і суспільства. Натомість, можливий інший підхід, за яким показники фінансування освіти подаватимуться як такі, що належать виключно до групи фінансових. Підставою для такої їх приналежності є відображення через них пропорції розподілу державного бюджету. Отже, в основу поділу показників на групи можуть покладатися різні *теоретичні припущення*, тому будь-який існуючий поділ є «умовно правильним».

Однак, поділ екзогенних показників на певні групи, попри його умовність, є необхідним для так званого передмодельного аналізу. В його межах на рівні загальновідомих та усвідомлених теоретичних

конструкції вирішується питання про те, від чого «в принципі» може залежати досліджувана змінна. Зокрема, за новою класичною теорією, економічне зростання передусім залежить від ресурсного потенціалу економіки, від досягнутого рівня продуктивності, від заощадження та інвестування у технічний прогрес тощо. Відтак, ідеться про показники, охоплені, за нашою класифікацією, групами 1 і 5. Натомість, у кейнсіанській теорії економічне зростання пояснюється у зв'язку з фінансовими, кредитними показниками та показниками доходів, через які відбувається вплив на сукупні витрати. Отже, йдеться про показники виокремлених нами 2-ої, 3-ої, 4-ої груп.

Перелік показників, на основі яких будувалися дві нейронні мережі, подано у табл. 1.

Таблиця 1

Повний перелік показників, що використовувалися для побудови двох штучних нейронних мереж

Позначення показника латинськими літерами	Зміст показника	Порядковий номер показника в моделі
Ендогенні (пояснювані) змінні моделі		
g_Y	Темп економічного зростання (%)	6
k_{Gin}	Коефіцієнт диференціації доходів Джинні	2
Показники досягнутого рівня розвитку економіки		
Y/N	ВВП на одного громадянина (дол. США)	1
Im/Y	Частка імпорту у ВВП (%)	16
Ex/Y	Частка експорту у ВВП (%)	17
$S_{N/Y}$	Частка національних заощаджень у ВВП (%)	15
G_{educ}/G	Витрати на освіту в державних витратах (%)	26
$Exp_{G/st}/Y_N$	Частка державних витрат на студента (вища освіта) до ВВП на одного громадянина (%)	27
Соціальні показники		
Sh_{low20}	Частка найбіднішої квінтильної групи у доходах	3
Sh_{high20}	Частка найбагатшої квінтильної групи у доходах	4
k_{dif}	Квінтильний коефіцієнт диференціації доходів	5
Sh_{Pov}	Частка населення за межею бідності (частка тих, хто споживає на суму, меншу 1,9 дол. США)	31
Фінансові показники		
D_{Ext}/GNI	Частка зовнішнього боргу у валовому національному доході (%)	7
D_{Ext}	Обсяг зовнішнього боргу (дол. США)	8
T/Y	Частка податків на доходи у ВВП (%)	11
G/Y	Частка видатків у ВВП (%)	12
D_G/Y	Частка державного боргу у ВВП (%)	18
Cr_{IMF}	Використовувані кредити МВФ (у дол. США)	20

Продовження табл. 1

Монетарні показники		
π_{CPI}	Темп інфляції за індексом споживчих цін (%)	10
q_{USD}	Курс валюти (грн за дол. США)	13
π_{def}	Темп інфляції за показником дефлятора ВВП (%)	14
Res_{CB}	Резерви національного банку (в дол. США)	19
i^r	Реальна процентна ставка (%)	21
i_{dep}	Процентна ставка за депозитами (%)	22
g_{BM}	Зростання широких грошей (%)	23
Показники ресурсного потенціалу		
L	Робоча сила (осіб)	24
u_{ILO}	Рівень безробіття, за міжнародними методиками оцінювання (%)	25
u'	Рівень безробіття за оцінкою національних органів статистики (%)	9
N	Населення країни (осіб)	28
n	Темп зростання населення (%)	29
$migr_N$	Чиста міграція (осіб)	30

Джерело: побудовано авторами самостійно.

Перевірка зв'язків між змінними відібраної бази даних з використанням методів регресійного аналізу дала можливість з'ясувати, що найкращим для пояснення *економічного зростання* g_Y (6) виявився такий перелік змінних: π_{CPI} (10), q_{USD} (13), S_N/Y (15), Im/Y (16), Res_{CB} (19), Cr_{IMF} (20), i^r (21). Характеристики якості регресійної моделі подано далі у табл. 2.

Таблиця 2

Технічні характеристики якості моделі

Регрессионная статистика							
Множествен	0,91						
R-квадрат	0,84						
Нормирован	0,78						
Стандартная	4,10						
Наблюдения	28,00						
Дисперсионный анализ							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>		
Регрессия	7,00	1714,49	244,93	14,60	0,00		
Остаток	20,00	335,48	16,77				
Итого	27,00	2049,98					
Коэффициенты, стандартная ошибка статистики							
	<i>R-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>		
Y-пересечен	-71,20	12,39	-5,75	0,00	-97,04	-45,36	-45,36
10 π_{CPI} Inflat	0,01	0,00	4,86	0,00	0,00	0,01	0,00
13 q_{USD} Offic	0,32	0,16	1,92	0,07	-0,03	0,66	-0,03
15 S_N/Y Gross	0,61	0,17	3,51	0,00	0,25	0,97	0,25
16 Im/Y Impor	1,08	0,25	4,32	0,00	0,56	1,61	0,56
19 Res_{CB} Totz	0,00	0,00	4,55	0,00	0,00	0,00	0,00
20 Cr_{IMF} Use	0,00	0,00	-2,77	0,01	0,00	0,00	0,00
21 i^r Real inte	0,29	0,05	6,42	0,00	0,20	0,39	0,20

Джерело: побудовано авторами самостійно.

Наведені технічні характеристики моделі свідчать про таке:

- спостерігається висока якість моделі, оскільки подані у ній змінні пояснюють економічне зростання на 84% ($R^2 = 0,84$);
- імовірність того, що змінні моделі відібрано неправильно, є нульовою (F-критерій = 0), тобто змінні відібрано правильно;
- вплив усіх змінних моделі на економічне зростання, репрезентований коефіцієнтами при змінних, є значущим (P-значення $< 0,1$ для всіх змінних).

Попри високі технічні характеристики моделі, знаки при коефіцієнтах («+» чи «-») засвідчують суперечливі й такі, що важко пояснюються з позицій макроекономічної теорії, впливи. До таких суперечливих, з огляду на знаки при коефіцієнтах, впливів змінних, зокрема, належить таке:

- виявлено сильний позитивний вплив інфляції на економічне зростання (коефіцієнт при $\pi_{СРІ} = +0,01$), попри традиційний підхід до надмірної інфляції як до чинника з гальмівним впливом щодо зростання;
- ідентифіковано сильний позитивний вплив від збільшення частки імпорту у ВВП на економічне зростання (коефіцієнт при $Im/Y = +1,08$), попри традиційний підхід до імпорту як до явища «вилучення» з економічного кругообігу, а отже, як до фактора, що має гальмувати зростання;

- виявлено сильний позитивний вплив реальної ставки процента на економічне зростання (коефіцієнт при $i^r = +0,29$), попри уявлення про стримуючий вплив збільшення процентної ставки на інвестиції, а відтак, і на економічне зростання.

З'ясовані суперечності між характером (напрямоком) впливу показників, за якісною регресійною моделлю, з одного боку, та за логікою теоретичних пояснювальних моделей щодо цих впливів – з іншого, дають підстави принаймні для таких припущень:

- впливи економетрично відібраних показників є складнішими, ніж це передбачено в макроекономічній теорії;
- імовірно, що ці впливи є нелінійними, перехресними й такими, що ґрунтуються на багаторівневих залежностях.

Якщо наші припущення правильні, то в цьому випадку для моделювання та пояснення економічного зростання доречним є використання інструментарію Data Science з побудовою штучної нейронної мережі.

Модель штучної нейронної мережі створено в середовищі Deductor. Результати побудови мережі подано на рис. 2.

Граф нейронної мережі засвідчує нелінійний вплив сімох змінних на економічне зростання, їх перехресну взаємодію та формування певних «вузлів» такої взаємодії. Найсильнішим, при формуванні одного з «вузлів», виявився вплив реальної процентної ставки та використовуваних кредитів МВФ, а найслабшим – вплив

частки імпорту у ВВП. При формуванні іншого «вузла», вплив усіх змінних виявився приблизно однаковим.

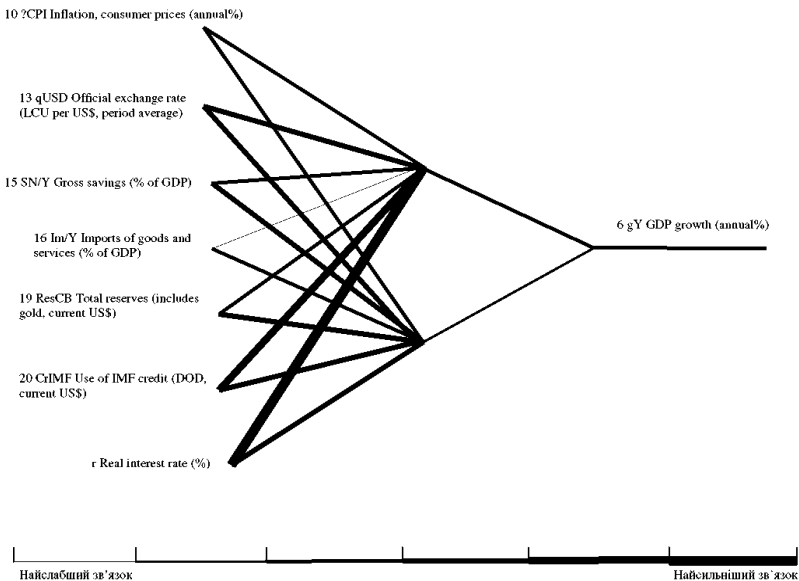


Рис. 2. Граф нейронної мережі, що ілюструє зв'язок відібраних змінних та економічного зростання

Джерело: побудовано авторами самостійно.

Перевірка нейронної мережі на тестовій вибірці подана на рис. 3.



Рис. 3. Перевірка якості нейронної мережі

Джерело: побудовано авторами самостійно.

Рис. 3 ілюструє ступінь наближення фактичних та змодельованих з використанням нейронної мережі значень економічного зростання. На тестовій вибірці (2017 – 2018 рр.) похибка прогнозу склала в середньому менше 1%. Це свідчить про високу якість нейронної мережі та можливість її використання для прогнозування й аналізу.

Перевірка зв'язків між змінними для побудови нейронної мережі, що моделює рівень диференціації доходів, за коефіцієнтом Джинні, методами регресійного аналізу виявила найкращі змінні для пояснення. Такими змінними стали: u^i (9), π_{CPI} (10), q_{USD} (13), S_N/Y (15), i^r (21), i_{dep} (22), n (29).

Характеристики якості регресійної моделі зв'язку між коефіцієнтом Джинні та іншими змінними подано далі у табл. 3.

Таблиця 3

Технічні характеристики моделі

Регрессионная статистика							
Множественный							
R-квадрат							
Нормированный							
Стандартная оц.							
Наблюдения							
Дисперсионный анализ							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>		
Регрессия	7,00	267,20	38,17	51,13	0,00		
Остаток	11,00	8,21	0,75				
Итого	18,00	275,42					
Кoeffициенты/стандартная ошибка							
	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение:	23,62	1,84	12,87	0,00	19,58	27,66	19,58 27,66
9 u Unemploy	-0,92	0,27	-3,41	0,01	-1,52	-0,33	-1,52 -0,33
10 π_{CPI} Inflation	-0,01	0,00	-7,66	0,00	-0,01	0,00	-0,01 0,00
13 q_{USD} Official	0,11	0,05	2,09	0,06	-0,01	0,22	-0,01 0,22
15 S_N/Y Gross sa	0,16	0,06	2,65	0,02	0,03	0,30	0,03 0,30
21 i^r Real interest	0,06	0,03	1,95	0,08	-0,01	0,13	-0,01 0,13
22 i_{dep} Deposit	0,24	0,03	7,17	0,00	0,17	0,32	0,17 0,32
29 n Population	-6,89	1,44	-4,78	0,00	-10,06	-3,72	-10,06 -3,72

Джерело: побудовано авторами самостійно.

Наведені технічні характеристики моделі у цьому випадку свідчать про таке:

- спостерігається дуже висока якість моделі, оскільки подані у ній змінні пояснюють диференціацію доходів, оцінену коефіцієнтом Джинні на 97% ($R^2 = 0,97$);

- фактично відсутня ймовірність того, що змінні моделі відібрано неправильно, (F -критерій = 0), тобто змінні відібрано правильно;

- вплив усіх змінних моделі на диференціацію доходів, репрезентований коефіцієнтами при цих змінних, є значущим (P -значення $< 0,1$ для всіх змінних).

За знаками коефіцієнтів при змінних («+» чи «-»), тобто за характером залежностей, вплив на диференціацію доходів відібраних змінних виглядає ще більш контроверсійно, ніж у попередній моделі економічного зростання. Характер впливу лише однієї змінної (з 7-ох)

може бути пояснений принаймні «з позицій здорового глузду». Йдеться про обернений зв'язок між темпом зростання населення та диференціацією доходів (коефіцієнт при $n = -6,89$). Така обернена залежність може пояснюватися хоча б тим, що зменшення темпу зростання населення, за інших однакових умов, збільшує частку людей старшого віку, доходи яких зазвичай є меншими. Виявлений обернений зв'язок між диференціацією доходів та двома змінними – безробіттям (коефіцієнт при $u' = -0,92$) й інфляцією (коефіцієнт при $\pi_{CPI} = -0,01$) – суперечить теорії. Адже усталеним є твердження про інфляцію як чинник збільшення розшарування громадян у суспільстві та про безробіття, що спричиняє зростання бідності. Позитивний вплив девальвації на збільшення диференціації доходів (коефіцієнт при $q_{USD} = +0,11$) може бути поясненим особливостями масового зберігання заощаджень українськими громадянами у валюті інших країн. Зв'язки ж диференціації доходів ще з трьома змінними – реальною ставкою проценту (коефіцієнт при $i^r = +0,24$), ставкою проценту за депозитами (коефіцієнт при $i_{dep} = -0,01$) та часткою заощаджень у ВВП (коефіцієнт при $S_N/Y = +0,16$) – є найбільш утаємниченими й мало зрозумілими. Саме вони, на наш погляд, найбільшою мірою виявляють невизначеність у сенсі того, що не підлягає поясненню, усвідомленню та очікуванню.

Усі зазначені особливості моделі диференціації доходів дають підстави припускати нелінійність та багаторівневість залежностей, отже – доцільність використання в моделюванні інструментарію штучних нейронних мереж.

Ілюстрація моделі штучної нейронної мережі, побудованої в середовищі Deductor, подана на рис. 4.

Граф нейронної мережі на рис. 4 засвідчує нелінійний, перехресний вплив сімох змінних на диференціацію доходів та факт формування «вузлів» взаємодії. Найсильнішим з усіх впливів виявився вплив частки національних заощаджень у ВВП при формуванні другого «вузла». При формуванні першого «вузла» взаємодії найсильнішим виявився вплив курсу національної валюти. Натомість, темп приросту населення здійснював у обох «вузлах» найслабший вплив на диференціацію доходів.

Перевірка якості нейронної мережі на тестовій вибірці подана на рис. 5.

На рис. 5 ілюстровано високий ступінь наближення фактичних та змодельованих з використанням нейронної мережі значень диференціації доходів, яку оцінено за коефіцієнтом Джинні. На тестовій вибірці (2017 – 2018 рр.) похибка прогнозу склала, в середньому, менше 1%. Це свідчить про дуже високу якість нейронної мережі та про те, що вона може використовуватися для прогнозування, аналізу й прийняття управлінських рішень.

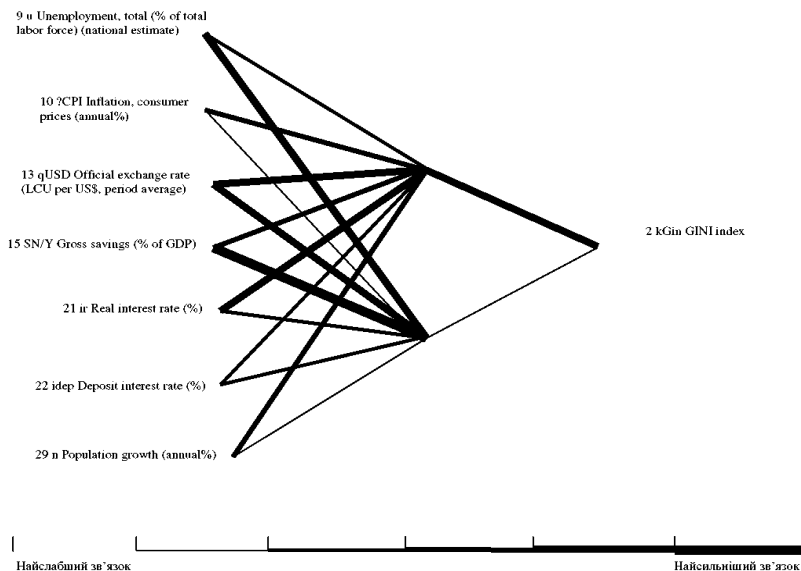


Рис. 4. Граф нейронної мережі зв'язку між відібраними змінними та диференціацією доходів, за коефіцієнтом Джинні

Джерело: побудовано авторами самостійно.

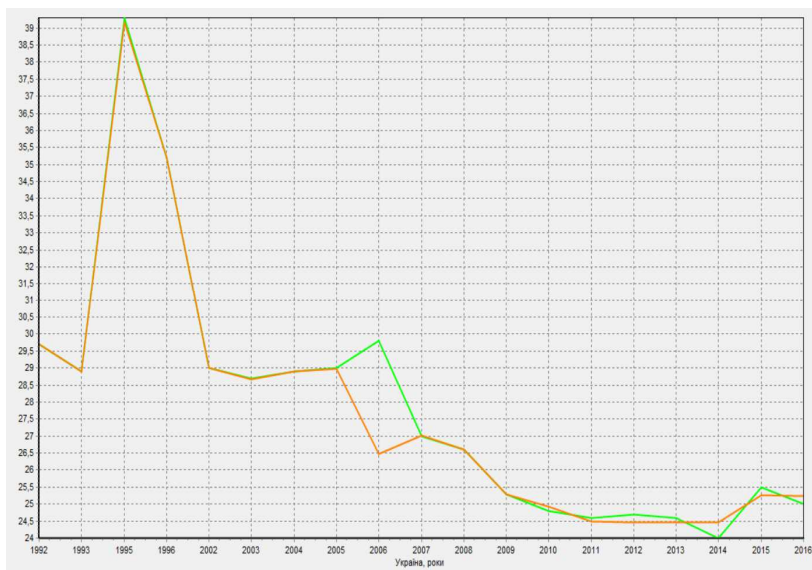


Рис. 5. Перевірка якості нейронної мережі

Джерело: побудовано авторами самостійно.

Узагальнюючи результати використання інструментарію Data Science в аналізі на макроекономічному рівні, ми робимо такі проміжні висновки:

– у процесі економетричного аналізу макроекономічних явищ – економічного зростання та диференціації доходів – виявлено суттєві, але «непояснювані» й непередбачені зв'язки (залежності). За нашим припущенням, у цьому виявляється економічна неочевидність, так звані «приховані смисли», отже, економічна невизначеність;

– використання інструментарію Data Science дає можливість робити висновки про фактично існуючі залежності та якісно прогнозувати майбутній перебіг подій навіть у випадку існування нечітко ідентифікованих впливів. Попри нечітку ідентифікацію, автоматичне формування «вузлів» взаємодії між змінними засвідчує існування складних залежностей.

Метою цього дослідження є пошук відповіді на питання, в який спосіб використання інструментарію Data Science може сприяти прийняттю більш обґрунтованих управлінських рішень. Зрозуміло, що на макроекономічному рівні йдеться про рішення органів публічної влади й про так званих національних регуляторів. Розглянутий нами приклад може використовуватися в контексті принаймні таких рекомендацій щодо управлінських рішень:

– прогнозування динаміки змінних, які можуть бути метою впливу національних регуляторів (економічне зростання та диференціація доходів є саме такими цільовими показниками), не може спиратися лише на очевидні фактори впливу. В прогнозуванні мають враховуватися нелінійні багаторівневі зв'язки між змінними;

– для обґрунтування конкретних результатів (відгуків) імпульсів, що йдуть в економіку від органів публічної влади, доцільно формувати більш повне й досконале уявлення про трансмісію імпульсів. Остання охоплює як очевидні, так і неочевидні елементи трансмісії, тобто зв'язки між змінними.

Обидві рекомендації щодо вдосконалення управлінських рішень на макроекономічному рівні можуть досягатися з використанням інструментарію Data Science.

Актуальність використання Data Science на **мікроекономічному рівні** – рівні окремих економічних суб'єктів – у сучасній українській економіці визначається принаймні такими обставинами. По-перше, спостерігається низька ефективність інвестицій загалом та інвестицій у банківську сферу зокрема. Тому виникає потреба аналізу та використання таких інструментів дослідження, які б сприяли виявленню не лише очевидних, а й глибоких, прихованих (неочевидних) причин низької ефективності. По-друге, при прийнятті управлінських рішень, наразі, фактично використовується мізерна частка всієї інформації, пов'язаної

з діяльністю окремих економічних суб'єктів. За експертним оцінюванням, ця частка становить лише 0,5%. Тому розширення бази досліджуваних даних може сприяти більш обгрунтованій відповіді на питання «Чому ефективність є низькою та яким чином підвищити віддачу від інвестицій» для кожного конкретного випадку.

Практичне використання технологій Data Science на мікроекономічному рівні ми ілюструватимемо на конкретному прикладі, який дає можливість виявити переваги цього інструментарію.

Об'єктом дослідження став конкретний банк*, що входить до 10-и топбанків України у 2018 році. Завданням (метою) аналізу став пошук та обгрунтування способів підвищення ефективності інвестицій. Ішлося про медіа-інвестиції у рекламу банківських послуг. Метою оптимізації інвестиційних витрат на рекламу було визначено збільшення продажів банківських послуг та зростання доходності (прибутковості) банку, тобто підвищення ROMI (Return of Marketing Investments).

Логіка модельної конструкції у цьому випадку була такою, яка представлена на рис. 6 і відображає, так звану «воронку продажів» банку через один з каналів продажів – кол-центр.

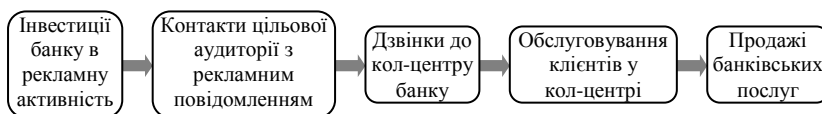


Рис. 6. Логіка модельної конструкції при обгрунтуванні рішень для комерційного банку з використанням інструментарію Data Science

Джерело: побудовано авторами самостійно.

Інформаційна база дослідження охоплювала дані за більше ніж 30-ма показниками у різних часових проміжках, що, за джерелом (походженням) даних, поділялися на три групи, а саме:

- показники діяльності банку щодо «воронки продажів»;
- показники щодо рекламної активності банку та його конкурентів;
- дані про макроекономічне та соціальне середовище підприємницької діяльності.

Використання перших двох груп джерел даних у цьому дослідженні є апіорі зрозумілим. Адже йдеться про проєкт використання рекламної активності для збільшення продажів банківських продуктів. Третя група джерел потребує детальнішого пояснення.

У період економічного спаду внаслідок політичної кризи та початку російсько-української війни у 2014 – 2015 рр. банківський сектор

* Через зобов'язання про нерозголошення інформації назва банку не оприлюднюється.

загалом і даний конкретний банк зокрема зазнали суттєвих втрат від змін макроекономічного середовища. У цей період відбулася девальвація національної валюти у 3-4 рази (з 8 до 26 – 32 грн за дол. США), значно скоротився індекс довіри до банків [19], депозити у валютах інших країн зменшилися на 57%, депозити в гривнях – на 8%, кредити – на 38% [20]. Тому зрозуміло, що рекомендації для банку в 2018 р. мали спиратися на припущення про існування певних трендів очікувань та реакцій споживачів банківських послуг, що сформувалися у 2014 – 2015 рр. Необхідним було й урахування нових макроекономічних тенденцій економічного зростання у 2016 – 2018 рр.

Акцентування уваги на інвестуванні у рекламну діяльність визначалося тим, що під час кризового спаду національної економіки рекламна діяльність банків у країні мінімізувалася. У медійному просторі була лише інформація про декілька великих банків, які намагалися підтримувати довіру українців [21; 22]. Досліджуваний банк інвестував лише в одну рекламну кампанію у 2016 р., але рівень реакції був невисоким. Тому кампанія була згорнута, а медіареклама не використовувалася впродовж 2-х років.

При здійсненні рекламної кампанії досліджуваного банку постала загальна бізнес-задача: інвестування у медіарекламу та розробка медійної стратегії у такий спосіб, який би спричинив максимальний відгук у вигляді дзвінків до кол-центру банку. Конкретним завданням інвестиційного проекту стало формування «найкращого медіаміксу» з різних рекламних інструментів за оптимального розподілу рекламного бюджету. Критерієм оптимізації визначено досягнення позитивної віддачі від інвестицій (ROI). Відповідно, мета Data Science-проекту полягала в дослідженні, моделюванні та прогнозуванні завантаженості кол-центру банку у відповідності з рекламною діяльністю банку в засобах масової інформації. При цьому використовувався Data Mining підхід і технології машинного навчання.

Проект здійснювався згідно з міжгалузевим стандартом обміну даними CRISP-DM [23] – аналітичною моделлю, що дістала найбільше використання в подібних дослідженнях [24]. Для створення необхідної бази даних зібрано інформацію щодо показників попередніх рекламних кампаній банку впродовж 2013 – 2016 рр., а також дані щодо соціально-економічного розвитку. Була врахована класична практика медіапланування (врахування частки голосу в медіаканалі, частоти контактів зі споживачами та рівня охоплення цільової аудиторії). Проте основна увага була зосереджена на бізнес-показниках (дзвінках до кол-центру банку, продажах послуг банку й рівні конверсії). У такий спосіб був розроблений новий інтегральний підхід для прийняття рішень, що базується на Data Science, технологіях машинного навчання й максимальному використанні даних.

За допомогою програмного забезпечення Excel та R-Studio була побудована економетрична модель з ключовою метрикою «Вхідні дзвінки до кол-центру». Для реалізації підходу використовувалися математичні методи аналізу й прогнозування на основі бази даних з історичними даними самого банку, медіа-агенції та зовнішнього середовища (соціальних і макроекономічних показників). До моделі були додані параметри, які впливали на конверсію (трансмісійний механізм) від медіа-активності до дзвінків, а також від дзвінків до замовлень та продажів. Параметри було оптимізовано саме для отримання найкращого коефіцієнту конверсії.

Загальна економетрична модель поділялася на підмоделі для моніторингу бізнес-завдань на кожному конкретному етапі, а саме:

1) модель короткострокового щотижневого планування, що дозволяло фіксувати та оцінювати вплив реклами на досягнення результатів діяльності банку (як позитивних, так і негативних) у будь-який момент часу;

2) модель тактичного планування, що дозволяє планувати погодинну інтенсивність трафіку кол-центру банку у відповідь на зміни медіа-активності. Зокрема, було виявлено залежність між вхідними дзвінками потенційних клієнтів банку до кол-центру та обсягом розміщеної реклами на телебаченні протягом дня. Це дало можливість визначити ефективність телевізійної активності в кожен годину дня і впродовж тижня. Модель створила підґрунтя для операційної оптимізації роботи кол-центру банку.

Модель множинної регресії з більш ніж 30 факторами зі щоденними та погодинними характеристиками має такий вигляд:

$$\text{Calls_by_hours} = \text{hours_coefficient} \times \text{day_coefficient} \times (\text{Constant} + a_1 \times \text{Adstock}(\text{TV}_1) + a_2 \times \text{Adstock}(\text{TV}_2) + \dots + a_n \times \text{Adstock}(\text{TV}_n) + b \times \text{Radio_activity} + c_i \times \text{billboards_i} + d_i \times \text{Integrated_economic_indicator_i}),$$

де Calls_by_hours – кількість дзвінків до кол-центру з погодинною розбивкою; hours_coefficient – ефективність телевізійної активності за певну годину; day_coefficient – ефективність телевізійної активності за певний день тижня; Constant – базовий (органічний) рівень дзвінків за певну годину; a_i – ефективність телевізійної активності i -го виду, $i = 1 \dots n$; Adstock – відповідно, миттєвий, тривалий і відстаючий ефект впливу реклами на телебаченні на поведінку покупців банківських послуг упродовж певного часу; $\text{Adstock}(\text{TV})_t = \text{TV}_t + a \times \text{Adstock}(\text{TV})_{t-1}$; b – ефективність активності на радіо; Radio_activity – активність на радіо; c_i – ефективність активності в зовнішній рекламі i -го виду, $i = 1 \dots n$; billboards_i – активність у зовнішній рекламі i -го виду, $i = 1 \dots n$; $\text{Integrated_economic_indicator}$ – інтегральний показник, який одночасно віддзеркалює рівень ВВП, доходів та динаміку банківських продуктів.

Модель є достатньо складною, з технічного погляду, тому що втілює поєднання субмоделей з урахуванням перебігу подій упродовж кожного дня і кожного тижня. Базою для розрахунків стали дані з джерела даних*. Ці ж дані використано для побудови рисунків, поданих далі.

Технічні характеристики однієї з субмоделей загальної моделі подано у табл. 4.

Таблиця 4

Технічні характеристики субмоделі

Indicator	Coefficient	Stand. Error	t-statistics	P-value
Constant	19,78	5,97	3,31	0,0017
Economic indicator	-3,82	0,08	-50,44	0,0000
Billboard	32,98	0,42	77,77	0,0000
Radio	65,24	4,45	14,67	0,0000
TV1	158,53	0,75	211,77	0,0000
TV2	140,34	1,08	130,09	0,0000
TV3	178,96	1,45	123,61	0,0000
TV4	110,27	7,70	14,32	0,0000
Multiple R-squared	0,97		Adjusted R-squared	0,97
F-statistics	11894,423		p-value	0,0000

Джерело: розроблено авторами самостійно.

Основними критеріями технічної оптимізації моделі були: підвищення R^2 , уникнення автокореляції, гетероскедастичності та мультиколінеарності. Досягнуті результати моделювання є такими: модель оцінює вплив факторів з вірогідністю на рівні 97% ($R^2 = 97\%$), існує гомоскедастичність (тобто відсутня гетероскедастичність), відсутня автокореляція.

Побудована економетрична модель дозволила визначити вплив кожного значущого фактора та розробити рекомендації щодо найбільш ефективного використання медіаресурсів. Основні п'ять рекомендацій є такими.

По-перше, доцільність дотримання певної тривалості медіакампанії для мінімізації «ефекту зносу» на основі даних про те, що після досягнення так званої ваги флайту в X TRPs (основний показник вимірювання телевізійної активності), протягом Y тижнів, ефективність телевізійної активності знижується за рахунок ефекту зносу (рис. 7). Було рекомендовано підтримувати тривалість флайту на необхідному рівні TRPs для максимальної ефективності реклами.

По-друге, важливість здійснення ротації (заміни) роликів під час флайту для додаткового зростання кількості дзвінків до кол-центру

* Внутрішня база даних українського банку (конфіденційна інформація).

банку та часткового зменшення ефекту зносу. Висновок зроблено на підставі оцінки того, що заміна рекламного оголошення дозволяє збільшити кількість дзвінків на 19%, але це не компенсує ефекту зносу. З урахуванням наявності коротких флайтів рекомендовано використовувати різні рекламні матеріали.

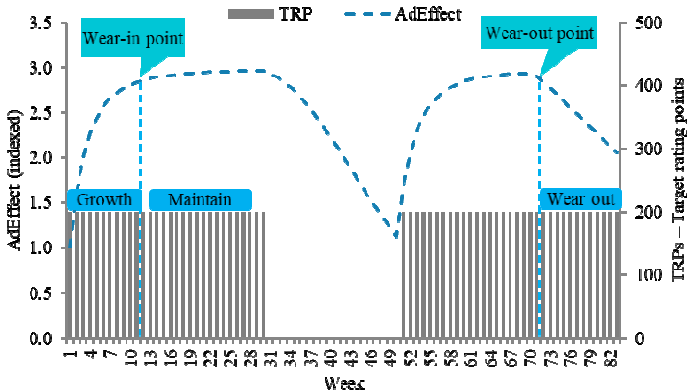


Рис. 7. «Wear-out» ефект («ефект зносу» рекламного повідомлення)

Джерело: розроблено авторами самостійно.

По-третє, доцільність розміщення виключно конкретних медіаресурсів, наприклад, лише ролика X" (рис. 8), уникаючи використання інших. З'ясовано, що, враховуючи ціну, розміщення реклами, наприклад, з X" ролик, є найбільш ефективним. Тому для реалізації KPIs рекомендовано 100% використання лише певної версії відеореклами.

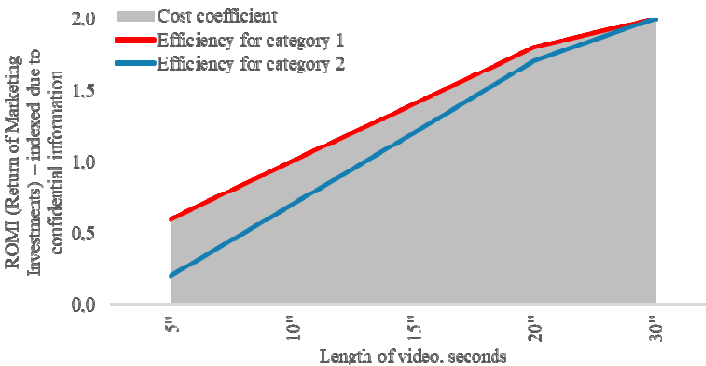


Рис. 8. Ефективність різної тривалості відеоматеріалів

Джерело: розроблено авторами самостійно.

По-четверте, обґрунтованість додаткової активності в інших (окрім уже застосованих) каналах комунікації за останні тижні телевізійної кампанії для додаткового охоплення потенційних клієнтів та зростання кількості дзвінків до кол-центру банку (рис. 9). Доведено, що інший медіаканал допомагає генерувати додаткові дзвінки, а одночасне застосування декількох каналів комунікації дає приріст дзвінків за кожен день активності в розмірі +20% порівняно з дзвінками, забезпеченими лише однією телевізійною рекламою.

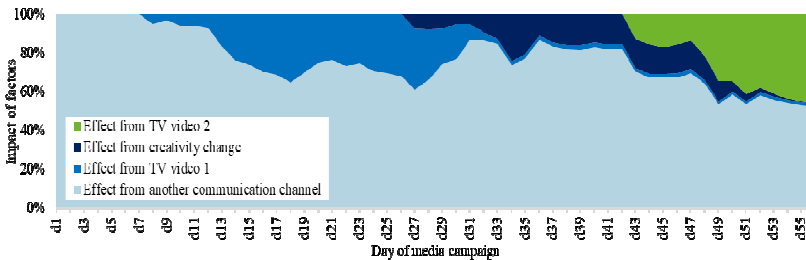


Рис. 9. Декомпозиція моделі з застосуванням різних каналів комунікації

Джерело: розроблено авторами самостійно.

По-п'яте, доцільність втілення тактичних організаційних змін, а саме: а) обмеження розміщення реклами у вихідні та святкові дні на телебаченні, що підтверджується графіком (рис. 10);

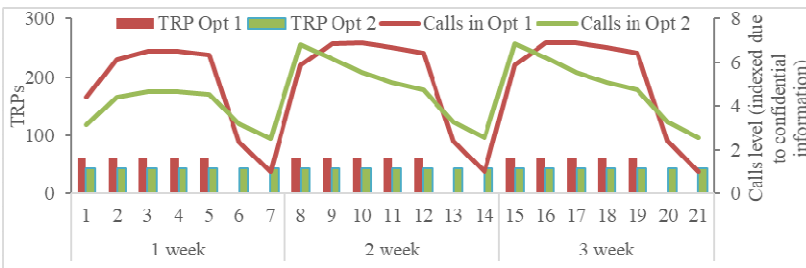


Рис. 10. Дзвінки до кол-центру банку за різних сценаріїв розподілу телевізійної активності протягом тижня

Джерело: розроблено авторами самостійно.

б) рівномірний розподіл активності впродовж дня, за обмеження частки вечірнього розміщення. Доведено, що ефективність розміщення у вечірній час є нижчою, ніж денна та ранкова, а ефект вечірнього прайм-тайму виявився таким самим, як і для денного розміщення реклами (рис. 11).

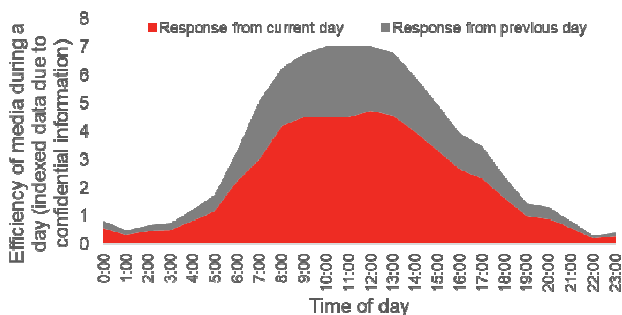


Рис. 11. Кількість дзвінків з рівномірним розподілом рейтингів на телебаченні протягом дня (погодинна ефективність телевізійної активності)

Джерело: розроблено авторами самостійно.

Такі рекомендації не можна приймати одночасно для всіх компаній на ринку, оскільки результати – це сукупність багатьох факторів та умов, які формуються кожного разу. Це вимагає індивідуального підходу в кожному конкретному випадку.

Основними досягнутими результатами Data Science-проекту з використанням технологій машинного навчання стали рекомендації щодо оптимізації інвестицій у рекламу, з огляду на показники бізнесової діяльності банку. Ці рекомендації, як уже зазначено, стосувалися формування найбільш ефективного поєднання (міксу) засобів масової інформації, найкращого періоду (часу доби чи дня тижня) для використання ЗМІ з метою рекламування діяльності банку, тривалості використання окремих інструментів реклами тощо. Були досягнуті такі конкретні економічні результати:

- вартість рекламної кампанії зменшено на 14%;
- забезпечено на 58% вищий рівень конверсії (ROMI) – зв'язку між інвестуванням у медіа-активність (рекламні кампанії) та показниками діяльності банку-рекламодавця – порівняно зі стандартним ринковим рівнем конверсії;
- оптимізовано діяльність кол-центру банку та мінімізована ймовірність «втрати» дзвінків клієнтів через неможливість результативного контакту;
- розширено можливості прогнозування результатів рекламної кампанії банку: в середньому відхилення від прогнозу становило не більше 11% за щоденного прогнозування результатів та не більше 8% – за щотижневого.

Показовим є те, що застосування Data Science сприяло виявленню суттєвого впливу чинників, які не можна було передбачити на початку дослідження та оцінити з використанням іншого інструментарію дослідження. Відтак, ми робимо висновок: Data Science є результативним

інструментом виявлення й оцінювання економічної невизначеності на мікроекономічному рівні.

Висновки:

За результатами дослідження можна зробити такі узагальнення:

1. Data Science є інтегральною аналітичною наукою, що гарантує кращі результати при використанні великих баз даних та вирішенні проблеми ідентифікації й оцінювання економічної невизначеності.

2. Застосування інструментарію Data Science під час аналізу на *макроекономічному* рівні сприяло виявленню таких проявів невизначеності:

- суттєвого впливу раніше не ідентифікованих, тобто прихованих, «неповерхневих» факторів впливу (змінних);
- існування неприродних, тобто таких, що не пояснюються ні дедуктивним, ні емпіричним шляхом зв'язків;
- формування випадкових «вузлів» взаємодії між змінними, що можуть взаємно посилюватися або послаблюватися.

3. Застосування інструментарію Data Science під час аналізу на *мікроекономічному* рівні сприяло виявленню таких проявів невизначеності:

- вплив на процеси та явища факторів з інших суміжних сфер, які змінюють настрої, очікування, реакції;
- нелінійна конверсія (передання) імпульсів, яка, ймовірно, має свої оптимуми, або періоди найкращих значень;
- непередбачуваний розподіл одних і тих же подій у часі, коли йдеться про часові піки та спади активності;
- несподівані поєднання (мікси) подій та інструментів, які або посилюють, або послаблюють імпульси;
- покращення управлінських рішень на підставі застосованого інструментарію Data Science можуть стосуватися надійнішого прогнозування, відбору способів управлінського впливу, більш обґрунтованого тактичного планування, організаційних змін у взаємодії окремих підрозділів економічних суб'єктів тощо.

Література:

1. Провост Ф., Фоусетт Т. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані / Пер. з англ. А. Дудченко. Київ : Наш Формат, 2019. 400 с.
2. Черняк О., Захарченко П. Інтелектуальний аналіз даних : підручник. Київ, 2014. 599 с.
3. Brin S. et al. Dynamic Itemset Counting and Implication Rules for Market Basket Data. N. Y. : ACM Press, 1997.
4. Einav L., Levin J. The Data Revolution and Economic Analysis. Stanford University and NBER. URL : <http://www.fas.org/irp/agency/dod/jason/data.pdf>.

5. Blum A., Hopcroft J., Kannan R. Foundations of Data Science. 2018. URL : <http://deeplearning.net/tutorial/deeplearning.pdf>.
6. Tom M. Mitchell Machine Learning. 1997. URL : <http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf>.
7. Чередніченко А. О., Шура Н. Застосування штучних нейронних мереж як дієвого механізму прийняття ефективних управлінських рішень. *Глобальні та національні проблеми економіки*. 2015. Вип. 4. С. 628-630.
8. Davis J. T., Episcopos A., Wettimuny S. Predicting direction shifts on Canadian–US exchange rates with artificial neural networks. *International Journal of Intel ligent Systems in Accounting, Finance & Management*. 2010. №10. P. 83-96.
9. Armano G., Marchesi M., Murru A. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*. 2005. №170(1). P. 3-33.
10. Tkacz G. Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*. 2001. №17. P. 57-69.
11. Vishwakarma K. P. Recognizing business cycle turning points by means of a neural network. *Computational Economics*. 1994. №7. P. 175-185.
12. Moshiri S., Cameron N. Neural network versus econometrics models in forecasting inflation. *Journal of Forecasting*. 2000. №19. P. 201-217.
13. Mitra S., Rupak T., Seema B., Mondal S. Artificial Neural Network Based Model for Forecasting of Inflation in India. *Fuzzy Information and Engineering*. 2016. Vol. 8. Issue 1. P. 87-100.
14. Hoog S. Deep Learning in (and of) Agent-Based Models: A Prospectus / Sander van der Hoog. 2017. URL : <https://arxiv.org > pdf>.
15. Chakraborty C., Joseph A. Machine learning at central banks / Chiranjit Chakraborty and Andreas Joseph. *Bank of England. Staff Working Paper*. 2017. №674. URL : <https://www.bankofengland.co.uk > working-paper>.
16. Офіційний сайт бази даних Світового банку. URL : <https://data.worldbank.org/indicator/>.
17. Kravchuk R. S. Ukrainian Political Economy: The First Ten Years. *Palgrave Macmillan*. NY. 2002. 352 p. URL : <https://books.google.com.ua/books?id=9BuBDAAAQBAJ&pg=PA147&lpg=PA147&dq>.
18. Офіційний сайт Державної служби статистики України. URL : <http://www.ukrstat.gov.ua>.
19. Сайт GfK Ukraine. URL : <https://www.gfk.com/uk-ua/>.
20. Сайт Національного банку України. URL : <https://bank.gov.ua/>.
21. Сайт Nielsen Ukraine. URL : <https://www.nielsen.com/ua/uk/>.
22. Сайт Індустріального телевізійного комітету. URL : <http://itk.ua/en>.
23. Brown M. S. What IT Needs To Know About The Data Mining Process. *Forbes*. 2015.
24. Shearer C. The CRISP-DM model: the new blueprint for data mining. 2000. №5. P. 13-22.