

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

«Моделювання маркетингової атрибуції в цифровому середовищі з використанням методів машинного навчання»

студентки 4 курсу

спеціальності 051 «Економіка»

ОПП «Економічна кібернетика»

денної форми навчання

Пустовіт Майї Григорівни

Науковий керівник:

д.е.н., професор

Чорноус Галина Олександрівна

Засвідчую, що в цій роботі немає запозичень із праць інших авторів без відповідних посилань

Студент _____

(підпис)

Роботу допущено до захисту перед ЕК

рішенням кафедри економічної кібернетики

від 12 червня 2025 р., протокол № 15

Завідувач кафедри:

доктор економічних наук, професор

Ляшенко Олена Ігорівна

(підпис)

КИЇВ – 2025

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 76 ст., 23 рис., 6 табл., 39 джерел, 2 додатки.

Ключові слова: маркетингова атрибуція, цифровий маркетинг, машинне навчання, значення Шеплі, ланцюги Маркова, алгоритмічні моделі, мультиканальність, Python.

Об'єкт дослідження: процес маркетингової атрибуції у цифровому середовищі.

Мета дослідження: реалізація алгоритмічної моделі маркетингової атрибуції на основі значень Шеплі для підвищення точності оцінки внеску каналів у багатоканальному середовищі цифрового маркетингу.

Методи дослідження: аналіз наукових джерел, теоретичне моделювання, порівняльний аналіз, стохастичне моделювання (ланцюги Маркова), теорія кооперативних ігор (значення Шеплі), програмна реалізація моделі в Python, візуалізація результатів.

Наукова новизна, теоретична значимість дослідження: розроблено та впроваджено модель атрибуції на основі значень Шеплі, що дозволяє враховувати маржинальний внесок каналів у всіх можливих комбінаціях взаємодій, на відміну від традиційних моделей маркетингової атрибуції, що засновані на правилах.

Практична цінність: результати дослідження можуть бути використані компаніями для точнішої оцінки ефективності маркетингових каналів, оптимізації бюджету та підвищення рентабельності інвестицій у цифрову рекламу.

RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Key words: marketing attribution, digital marketing, machine learning, Shapley values, Markov chains, algorithmic models, multichannel analytics, Python.

The graduation research of student Maiia Pustovit deals with the implementation of an algorithmic model of marketing attribution based on Shapley values to improve the accuracy of evaluating the contribution of marketing channels in a multichannel digital environment.

The work is relevant for marketing analysts and digital strategy teams, as it presents a comparative analysis of rule-based and algorithmic attribution models and offers practical solutions for campaign optimization. The Shapley Value model was implemented in Python and tested on real-world marketing data.

The practical value lies in the ability to use the results for better budget allocation, identifying undervalued channels, and increasing return on marketing investment (ROMI).

Pages 76, figures 23, tables 6, bibliog. 39, append. 2.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ МАРКЕТИНГОВОЇ АТРИБУЦІЇ У ЦИФРОВОМУ СЕРЕДОВИЩІ	8
1.1. Цифрове середовище як виклик для маркетингової аналітики.....	8
1.2. Концепція маркетингової атрибуції: цілі, виклики, технічна реалізація.....	14
1.3. Класифікація моделей маркетингової атрибуції	18
1.4. Недоліки традиційного підходу та роль машинного навчання	24
ВИСНОВКИ ДО РОЗДІЛУ 1	25
РОЗДІЛ 2. АЛГОРИТМИ МАШИННОГО НАВЧАННЯ ДЛЯ МОДЕЛЮВАННЯ АТРИБУЦІЇ.....	27
2.1. Загальна характеристика методів машинного навчання у маркетингу.	27
2.2. Ланцюги Маркова як основа стохастичного моделювання	28
2.3. Методи значень Шеплі для моделювання атрибуції	36
2.4. Порівняльний аналіз моделей атрибуції на основі ланцюгів Маркова і значень Шеплі	41
ВИСНОВКИ ДО РОЗДІЛУ 2	42
РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ АТРИБУЦІЇ	44
3.1. Постановка задачі та методологія моделювання маркетингової атрибуції	44
3.2. Характеристика обраного набору даних та попередня обробка для моделювання	46
3.3. Побудова моделі атрибуції на основі Shapley Values.....	55
3.4. Побудова моделей маркетингової атрибуції, що засновані на правилах та їхня оцінка	60
3.5. Порівняння результатів моделювання	63
3.6. Ключові висновки і рекомендації для бізнесу	65
ВИСНОВКИ ДО РОЗДІЛУ 3	68

	5
ВИСНОВКИ.....	69
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	71
ДОДАТКИ.....	75

ВСТУП

Актуальність дослідження. У сучасному цифровому середовищі маркетингові комунікації стали багатоканальними, персоналізованими та динамічними. При цьому зростає потреба в точному вимірюванні ефективності кожної взаємодії користувача з брендом. Традиційні моделі маркетингової атрибуції, що засновані на правилах, які надають усю цінність одній точці контакту вже не забезпечують належного рівня аналітичної точності. У відповідь на ці виклики дедалі ширше впроваджуються методи машинного навчання, зокрема моделі на основі ланцюгів Маркова та значень Шеплі. Вони дозволяють здійснювати справедливий і математично обґрунтований розподіл цінності між усіма каналами. Це дослідження є актуальним у зв'язку з посиленням вимог до аналітичної прозорості, ефективного управління бюджетом та складністю користувацьких шляхів.

Об'єктом дослідження є процес маркетингової атрибуції у цифровому середовищі.

Предметом дослідження є моделі маркетингової атрибуції, що засновані на правилах, алгоритмічні моделі на основі методів машинного навчання.

Метою дослідження є аналіз і практична реалізація моделей маркетингової атрибуції на основі методів машинного навчання для підвищення ефективності управління цифровими рекламними кампаніями. Зокрема, йдеться про побудову моделей на основі значень Шеплі та ланцюгів Маркова, їхню порівняльну оцінку та застосування до реальних даних. Для досягнення цієї мети були поставлені такі завдання:

- Узагальнити теоретичні підходи до маркетингової атрибуції;
- Систематизувати класифікацію моделей маркетингової атрибуції;
- Проаналізувати основні виклики сучасної маркетингової аналітики;
- Дослідити можливості застосування методів машинного навчання для атрибуції;
- Реалізувати модель Shapley Values Attribution у середовищі Python на основі реального набору даних;

- Здійснити порівняльний аналіз результатів моделей, що засновані на правилах і алгоритмічних моделях атрибуції;
- Сформулювати рекомендації для бізнесу на основі отриманих результатів;
- Оцінити можливості та обмеження використаних моделей у реальних умовах;

Методологічний інструментарій складається з методів теорії ймовірностей, елементи теорії ігор (значення Шеплі), методу стохастичного моделювання (ланцюги Маркова), економіко-математичного моделювання, аналізу даних у середовищі Python (pandas, numpy, matplotlib), а також методи порівняльного аналізу та візуалізації.

Наукова та практична новизна полягає у практичній реалізації моделі атрибуції на основі значень Шеплі для оцінки мультиканального шляху користувача, що дозволяє уникнути упередженості традиційних моделей та точніше відображає маржинальний внесок кожного каналу. У роботі також проведено порівняння моделей Шеплі та ланцюги Маркова у контексті цифрових кампаній із реальними даними.

Результати дослідження можуть бути використані компаніями для підвищення ефективності маркетингових кампаній, оптимізації рекламного бюджету та прийняття рішень на основі більш точних аналітичних моделей. Запропонована методика дозволяє ідентифікувати найбільш ефективні канали в складному середовищі цифрового маркетингу.

Інформаційною базою дослідження є наукові публікації в галузі маркетингової аналітики, офіційна документація платформ Google, Facebook, Kaggle, результати кейс-досліджень компаній, а також реальний набір даних цифрової рекламної кампанії.

Структура роботи. Кваліфікаційна робота містить 76 ст., складається зі вступу, трьох основних розділів, висновків, списку використаних джерел, та двох додатків.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ МАРКЕТИНГОВОЇ АТРИБУЦІЇ У ЦИФРОВОМУ СЕРЕДОВИЩІ

1.1. Цифрове середовище як виклик для маркетингової аналітики

У сучасному світі люди дедалі більше зазнають інформаційного навантаження з різних медіа-каналів, особливо в цифровому середовищі. Згідно дослідженню Statista, Bernard Marr & Co., понад 90% усіх наявних даних було згенеровано лише за останні два роки — це близько 2,5 квінтільйона байтів [1]. Найбільший відсотковий приріст було зафіксовано в період з 2019 по 2020 рік, що складав 56,59%. Очікується, що 120 зеттабайтів, згенерованих у 2023 році, збільшиться більш ніж на 150 % у 2025 році, досягнувши 181 зеттабайта, відбувається експоненціальне зростання, яке в майбутньому може бути важко контролювати, динаміка відображена на рис. 1.1.



Рис. 1.1. Динаміка обсягу згенерованих даних у світі, 2010-2025

Джерело: Складено автором на основі [1]

Цей інформаційний потік надає споживачам доступ до знань про безліч брендів і продуктів. Вони стикаються з різними форматами інформації на багатьох платформах — зокрема, у друкованих виданнях, на телебаченні, радіо, в пошукових системах, соціальних мережах, на контент-сайтах та блогах — ще до ухвалення рішення про покупку. Відео-формат відповідає за понад половину (53,72%) усього світового трафіку даних. Відповідно рис. 1.2., разом із соціальними мережами (12,69%) та іграми (9,86%) ці три категорії складають понад 3/4 (76,27%) усього інтернет-трафіку даних.

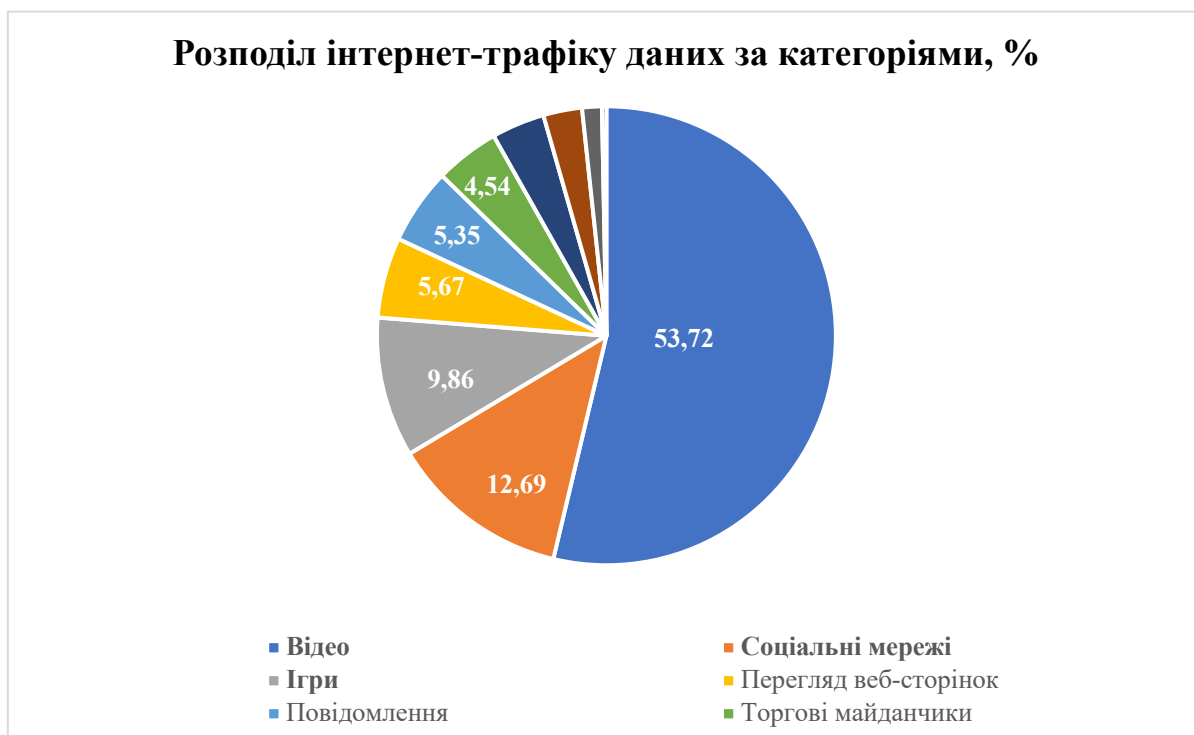


Рис. 1.2. Розподіл інтернет-трафіку даних за категоріями, %

Джерело: Складено автором на основі [1]

Бізнеси адаптуються під умови світової динаміки, яку створюють вподобання користувачів. Адаптація відбувається через створення різного типу контенту, у відповідності вимогам платформи, де відбувається запланована реалізація. Для брендів надзвичайно важливо взаємодіяти з клієнтами через ці численні платформи, адже майбутній покупець у щоденній рутині вже торкається потенційних точок входу. Це може відбуватись через телефонні дзвінки, маркетингові електронні листи, push-сповіщення в додатках, взаємодію в соціальних мережах тощо. Кожен канал створює дані на рівні

вражень, які можна пов'язати з конкретними клієнтами. Аналізуючи кожен етап клієнтського шляху, бренди можуть отримати значну конкурентну перевагу. Як зазначає Фаренюк Яна, аспірант кафедри економічної кібернетики у своїй науковій роботі «Principles Of Effective Planning Of Advertising Activity In The 21st Century» будь-яка рекламна кампанія буде успішною за умови правильного поєднання релевантного рекламного повідомлення (якісного креативу) та грамотного вибору комунікаційних каналів і інструментів донесення цього повідомлення до цільової аудиторії. На цьому етапі підприємство, у нашому випадку компанія, повинно чітко формувати мету кожної рекламної кампанії, яка може полягати у формуванні знань про бренд, створенні позитивного іміджу, донесенні інформації про продукт та його конкурентні переваги на ринку для того, щоб у перспективі зайняти міцні позиції й забезпечити зростання продажів. При цьому особливого значення набуває вирішення проблем визначення оптимального медіа-міксу, розміру рекламного бюджету та оптимізації процесів медіа-планування. [2]

Ці нові підходи потребують ефективного опанування. Структурні зміни створюють як і нові напрямки для розвитку, так і додаткові труднощі для спеціалістів. Згідно дослідження Forbes «18 Big Challenges Marketers And Advertisers Face This Year» сучасний маркетинг стикається з такими основними викликами, як [3]:

1. Перенасичення кількістю даних.

Враховуючи зростання обсягу даних, за останні періоди фахівці стикаються труднощами ефективного використання цих отриманих даних про споживачів. Через різноманітну структуру, цільове призначення, джерело виникає розмиття у розумінні їхньої цінності. Стає дедалі складніше визначити яка саме інформація, як результат когнітивної обробки даних, допомагає у створенні знань про клієнта і відповідного бізнес-рішення. Для спеціалістів існує важливість встановлення чітких цілей, визначення ключових показників та визначення відповідних джерел даних. Узгоджуючи зусилля зі збору даних з бізнес-цілями, відбувається зосередження на дійсно важливій інформації, зменшуючи тягар управління надмірними даними [4].

2. Поступова відмова від сторонніх файлів cookie.

У сучасному цифровому середовищі сторонні файли cookie відігравали ключову роль у відстеженні поведінки користувачів та реалізації моделей маркетингової атрибуції. Однак поступова відмова від їх використання зумовлена зростаючими вимогами до конфіденційності та змінами в регуляторному полі, суттєво ускладнює застосування традиційних методів відстеження та атрибуції [5]. Політика Mozilla щодо боротьби зі відстеженням призвела до того, що Firefox за замовчуванням блокує сторонні файли cookie від відомих трекерів. Apple також має подібну політику запобігання відстеженню; дотримання цієї політики призвело до появи аналогічного набору захисту файлів cookie третіх сторін, які ввімкнені за замовчуванням [6]. Google Chrome за замовчуванням блокує сторонні файли cookie в режимі інкогніто, також компанія додатково запустила тестування своїх нових функцій конфіденційності, де припиняється використання сторонніх файлів cookie у браузері Google Chrome для 1% користувачів [7]. Ця трансформація змушує маркетингологів адаптувати свої стратегії, зосереджуючись на розробці більш надійних підходів до збору та аналізу даних. Зокрема, акцент переноситься на використання власних (first-party) даних, які збираються безпосередньо від користувачів через взаємодію з веб-сайтами, мобільними додатками, програмами лояльності та іншими каналами. Вібхор Капур, Член ради Forbes зазначає: «Не маючи можливості орієнтуватися на людей, які використовують сторонні файли cookie, рекламодавцям потрібно буде набагато краще розуміти, що впливає на поведінку клієнтів та конверсії, і саме тут і знадобиться сильна атрибуція.»

3. Доведення рентабельності інвестицій

Згідно з аналітичними даними [8], обсяг глобальних витрат на цифрову рекламу за останні роки демонструє стабільне зростання. Рис. 1.3. відображає, що за період з 2019 по 2024 рік витрати на digital-рекламу зросли з \$392 млрд до \$790 млрд, подвоївшись протягом шести років. Щорічні темпи зростання коливалися в межах від 8,2% до 30%, що свідчить про сталу довіру бізнесу до цифрових каналів як джерела залучення та конверсії клієнтів.



Рис. 1.3. Глобальні витрати на цифрову рекламу (2019–2024)

Джерело: Складено автором на основі [8]

Така позитивна динаміка є логічним наслідком відносної легкості купівлі онлайн-реклами та поширення доступних інструментів самообслуговування, особливо в соціальних мережах. Це стало вирішальним фактором зростання витрат цифрової реклами, зокрема для малого та середнього бізнесу, який отримав змогу ефективно просувати свої продукти без великих початкових бюджетів і з доступом до поступового нарощення.

З іншого боку, зростання інвестицій у digital-маркетинг супроводжується і зростаючими очікуваннями з боку бізнесу щодо доведення їхньої ефективності. В умовах посиленої відповідальності маркетингових департаментів перед фінансовими підрозділами, ключовим викликом стає підтвердження рентабельності інвестицій (ROI). ROI (return on investment) — коефіцієнт повернення вкладених інвестицій, який відображає, наскільки бізнес-проект успішний і прибутковий. Цей фінансовий показник розраховують і для реклами, в такому разі фігурує термін ROMI (return on marketing investment [9]). Як зазначає Адріан Фальк, засновник агентства Believe Advertising & PR, «сьогодні компанії хочуть знати рентабельність інвестицій за кожен витрачений долар.

Минули ті часи, коли мільйони вкладалися у брендинг без конкретної звітності. Тепер маркетологи повинні бути прозорими у результатах, які можуть забезпечити» [3]. Відповідно працівники стикаються з необхідністю отримувати чіткі дані щодо ефективності кожної втіленої рекламної дії.

Також виникають труднощі з ефективним картуванням клієнтського шляху, побудові CJM (Customer journey map), який зазвичай базується на точних точках контакту та рівнях залучення, що ведуть споживачів до конверсії. CJM - це візуальне представлення шляху клієнта на платформі, від початкової взаємодії до остаточної цільової дії (покупки, реєстрації) і подальшої поведінки [10]. CJM клієнта узгоджується з основними етапами маркетингової воронки. Воронка є умовною моделлю, яка описує шлях користувача від першого дотику з брендом до конверсії (купівлі, підписки, замовлення) та лояльності в майбутньому. Концепція приписується Е. Сент-Елмо Льюїсу, який розробив модель AIDA у 1898 році. Модель AIDA (Attention, Interest, Desire, Action - увага, інтерес, бажання, дія) забезпечила фундаментальну основу для розуміння того, як споживачі просуваються через різні етапи перед покупкою. [11] . Основні етапи сучасної інтерпретації маркетингової воронки це: awareness (обізнаність), interest (інтерес), consideration (розгляд) , action (дія), loyalty (лояльність) [12].

CJM є деталізованим представленням цих етапів в контексті конкретної компанії чи продукту. Вона описує точки дотику користувача з брендом (канали, платформи, форми взаємодії), його мотивації, емоції та бар'єри на кожному з етапів воронки.

Однак, навіть при детальному плануванні воронки як CJM, на практиці маркетологи часто не можуть точно визначити, яка саме взаємодія мала вирішальне значення для переходу користувача між етапами. Прикладом може бути випадок, коли користувач бачить рекламний банер (етап Awareness), потім переглядає сайт (Interest), підписується на email-розсилку (Consideration), отримує знижку й купує товар (Conversion). Без якісної атрибуції важко визначити, який саме канал вплинув на рішення користувача найбільше: банер, email чи знижка?

І такі маркетингові труднощі як CJM та точна оцінка ROI найчастіше виникають не через неефективність кампаній, а через недостатньо якісні системи відстеження або вимірювання. І саме тут на перший план виходить моделювання маркетингової атрибуції.

1.2. Концепція маркетингової атрибуції: цілі, виклики, технічна реалізація

Маркетингова атрибуція — це процес виявлення та оцінювання впливу серії взаємодій користувача (точок контакту), які сприяють успішному залученню або конверсії клієнта. Кожній точці контакту надається певна цінність залежно від її внеску в загальне залучення клієнта на його шляху. Основна мета — допомогти брендам зрозуміти значущість кожної взаємодії в контексті всієї взаємодії з клієнтом. Визначаючи, які канали або точки контакту призводять до вищих коефіцієнтів конверсії, маркетингові команди можуть оптимізувати розподіл бюджету та стратегії комунікації. Це розуміння дає змогу маркетингологам чітко бачити, які послідовності дій користувачів найчастіше приводять до бажаної поведінки та, зрештою, до конверсій [4].

Сучасні користувачі можуть дізнаватися про бренд через різні джерела, серед яких органічні результати пошуку, медіа-кампанії з відображенням оголошень, посилення в соціальних мережах або ретаргетинг на сторонніх сайтах. Завдяки доступності численних платформ клієнти досліджують і купують товари з різних пристроїв і в різних місцях, тому фіксація кожної точки контакту є критично важливою для успішного моделювання атрибуції..

Маркетингова атрибуція надає компанії комплексне уявлення цифрового маркетингу і розширює простір для прийняття рішення. Поява великої кількості цифрових каналів — таких як пошукова реклама, соціальні мережі, email-маркетинг, банерна реклама, органічний пошук, відео тощо — зумовила необхідність системного аналізу впливу кожного з них. Класичні моделі оцінки ефективності базувалися здебільшого на загальних метриках, таких як вартість за клік (CPC) чи вартість за тисячу

показів (CPM), однак цього може бути недостатньо для повноцінного розуміння шляху користувача до конверсії. За рахунок правильної маркетингової атрибуції бізнес розуміє як кожна маркетингова кампанія і кожна одиниця вкладених грошей в неї впливає на шлях клієнта, починаючи від початкової обізнаності про бренд і закінчуючи остаточною покупкою продукту або послуги [13].

Чітке розуміння розподілу маркетингових витрат, яке забезпечує атрибуція, запобігає як переоцінці ефективності окремих каналів, так і недооцінці тих каналів, чия роль може бути неочевидною. Яскравим прикладом цього є медійна реклама яка здебільшого недооцінюється як канал. Вона не працює з великими доходами, оскільки їй бракує атрибуції конверсій, хоча вона може виступати основним рушієм впізнаваності бренду. [13] В конкретний період часу, наприклад, може здатися привабливим інвестувати кошти в розміщення банерної реклами на одному з медійних сайтів. Ймовірність того, що така реклама призведе до негайного продажу низька. Ефективна атрибуція дає змогу точно визначити необхідні кампанії до та після події, що дозволить генерувати продажі через кілька точок контакту з клієнтами протягом тривалого часу. Наприклад, якщо банерна реклама під час великої події може успішно підвищити впізнаваність бренду серед широкої аудиторії, то в кінцевому підсумку саме реклама в Instagram забезпечить вищий коефіцієнт конверсії.

Маркетингова атрибуція також сприяє покращенню клієнтського досвіду, оскільки дозволяє краще зрозуміти шлях клієнта, виділяти ключові точки контакту та визначати зони для вдосконалення. Знаючи, які взаємодії є критичними на різних етапах воронки продажів, спілкування з клієнтом стає більш цілеспрямованим та ефективним. Підвищується задоволеність і лояльність клієнта, з більшим впливом на різні сегменти аудиторії. Без якісної маркетингової атрибуції маркетологи можуть неправильно оцінити ефективність певних маркетингових зусиль та критичність проміжної точки контакту.

Вдосконалюється і стратегічне управління маркетингом. Аналіз поточних маркетингових цілей та реалізація відповідних заходів здатні впливати на стратегічний розвиток компанії. Підхід може гарантувати, що кожна попередня взаємодія

враховується та розглядається під час прийняття майбутніх рішень. З огляду на таку кількість точок дотику до маркетингу, які потрібно враховувати, та дедалі складніший шлях клієнта, вкрай важливо, щоб бізнеси робили крок назад, щоб переглянути свій поточний стан та способи оптимізації. У цьому випадку маркетингова атрибуція допомагає автоматизувати складний процес оцінки [14].

Маркетингова атрибуція, як технічний процес, передбачає систематичне збирання, обробку та аналіз даних про взаємодії користувачів. Етап збору даних реалізується за допомогою основних технологій відстеження, таких як файли cookie, пікселі відстеження, UTM-мітки та інші механізми, де відбувається ідентифікація джерела трафіку і поведінка користувача на сайті. UTM-мітки (Urchin Tracking Module) — це параметри, що додаються до URL-адрес для ідентифікації джерела трафіку, типу каналу та конкретної кампанії. Ці мітки дозволяють аналітичним системам, точно визначати, звідки прийшов користувач і яка кампанія спричинила його візит. Наприклад, параметри `utm_source`, `utm_medium` та `utm_campaign` допомагають розрізнити трафік з різних джерел і каналів. Ці дані зберігаються в системах веб-аналітики, таких як Google Analytics, Adobe Analytics або інших спеціалізованих платформах атрибуції [15].

Пікселі відстеження — це невидимі GIF-зображення розміром 1x1 піксель, які вбудовуються в веб-сторінки або електронні листи. Коли користувач завантажує сторінку або відкриває лист, піксель надсилає запит на сервер, фіксуючи інформацію про взаємодію, таку як IP-адреса, тип пристрою, браузер та геолокація. [16]

Файли cookie — це текстові файли, які зберігаються в браузері користувача для збереження інформації про його взаємодії з сайтом. Вони дозволяють відстежувати поведінку користувача на сайті, включаючи переглянуті сторінки, додані до кошика товари та інші дії. Файли cookie можуть бути як першої сторони (встановлені самим сайтом), так і третьої сторони (встановлені сторонніми сервісами), які дозволяють збирати дані про користувача на різних сайтах. [17]

Ця технічна реалізація маркетингової атрибуції базується на інтеграції основної технології відстеження, за рахунок якої збирається детальна інформація про взаємодії

користувачів з цифровими каналами. Ці дані вже виступають базою для побудови ефективних атрибуційних моделей.

Незважаючи на високий потенціал моделей маркетингової атрибуції для покращення бізнес-ефективності, їхнє практичне впровадження супроводжується рядом серйозних викликів. Ці труднощі стосуються як технічних, так і регуляторних аспектів, а також обмежень, пов'язаних із природою користувацької поведінки та багатоканальним середовищем.

Першим основним викликом є зниження доступності даних у зв'язку з посиленням конфіденційності. Поступова відмова від сторонніх файлів cookie, про які зазначалося в контексті глобальної проблематики, посилення світового регулювання конфіденційності, зокрема General Data Protection Regulation (GDPR) в Європейському Союзі, California Consumer Privacy Act (CCPA) у США, а також політика Apple щодо App Tracking Transparency (ATT) [18], значно ускладнюють використання традиційних підходів до вимірювання ефективності рекламних кампаній. Ці зміни обмежують можливості відстеження поведінки користувачів і зменшують обсяг доступних для аналізу даних, що безпосередньо впливає на точність атрибуційних моделей. Маркетологи змушені адаптуватися до нових умов, зокрема через активніше використання власних даних і співпрацю з партнерами з вимірювання мобільних технологій (MMP) [18].

Атрибуційні моделі також можуть бути схильними до ряду статистичних упереджень, зокрема позиційного упередження, це спричиняє хибні висновки про вплив каналів. Крім того, кореляційна природа деяких взаємозв'язків між поведінкою користувача та каналами може бути хибно інтерпретована як причинно-наслідкова. Такі неточності можуть призводити до неефективного розподілу маркетингового бюджету [bias].

Інформація про користувачів зазвичай зберігається у різних системах (аналітика сайтів, CRM, email-платформи, мобільні додатки тощо), що ускладнює отримання єдиного уявлення про повний шлях клієнта і робить дані фрагментованими та неповними. Крім того, такі канали як зовнішня реклама, івенти чи офлайн-покупки часто

не піддаються цифровому вимірюванню. Відстеження між пристроями (cross-device tracking) також ускладнене — користувач може взаємодіяти з брендом на мобільному телефоні, але конверсія може статися на комп'ютері, що створює розрив у даних [19].

1.3. Класифікація моделей маркетингової атрибуції

Перейдемо до опису класифікації найпопулярніших моделей. Атрибуційні моделі поділяються на дві основні категорії: детерміновані (моделі, що засновані на правилах) та алгоритмічні. У практиці цифрового маркетингу історично склалося, що найпоширенішими є детерміновані моделі атрибуції — моделі, засновані на фіксованих правилах, які однаково застосовуються до всіх користувачів. Ці моделі є простими у впровадженні та інтерпретації, однак мають обмежену точність у складних сценаріях взаємодії [20].

Моделі атрибуції одним дотиком (Single-Touch Attribution Models):

1. First Touch Attribution

Рис.1.4 ілюструє, що у цій моделі перша точка дотику, яка починає шлях конверсії, отримує повні 100% кредиту. Цей метод є технічно простим, легко реалізується без складних обчислень і часто використовується для оцінки каналів, що формують первинний попит, тобто для визначення, які джерела найкраще залучають нових потенційних клієнтів. Модель також не передбачає використання персональних ідентифікаційних даних, що є важливою перевагою в контексті дотримання норм конфіденційності. Однак він надто спрощує шлях користувача, ігноруючи подальші взаємодії, що впливають на рішення про покупку. Така атрибуція ґрунтується на кореляції, а не на причинно-наслідковому зв'язку. Це означає, що надання повної заслуги за конверсію першому дотику не обов'язково відображає його реальний внесок у досягнення цільової дії [21].



Рис. 1.4. Візуалізація моделі First Touch Attribution

Джерело: Складено автором на основі [21]

2. Last Click Attribution

Відповідно Рис. 1.5. модель останнього кліку приписує всю цінність конверсії останньому каналу, через який користувач взаємодіяв із сайтом перед здійсненням цільової дії. Наприклад, якщо користувач спочатку побачив рекламу в Instagram, потім перейшов за посиланням у розсилці, а згодом здійснив покупку після пошуку в Google, то вся конверсія буде приписана Google Ads. Її основний недолік — ігнорування ролі попередніх точок контакту, що створює викривлену картину ефективності. Модель останнього дотику часто використовується в кампаніях з коротким циклом продажів або коли основна мета — максимізувати негайні конверсії, тобто на фінальну дію користувача.

Веб-аналітика, така як Google Analytics, традиційно за замовчуванням використовує підхід атрибуції першого та останнього дотику для проведення аналізу атрибуції [21].



Рис. 1.5. Візуалізація моделі Last Click Attribution

Джерело: Складено автором на основі [21]

Моделі атрибуції кількома дотиками (Multi-Touch Attribution Models):

3. Linear Attribution

Згідно Рис. 1.6. лінійна модель рівномірно розподіляє цінність між усіма точками контакту користувача на його шляху до конверсії. Вона забезпечує збалансовану оцінку впливу каналів і комплексний огляд шляху клієнта, та гарантує, що кожна точка контакту отримує визнання за свій внесок. Такий підхід зручний для аналізу багатоканальних кампаній з тривалим життєвим циклом клієнта. Однак модель не враховує різну силу впливу каналів і може переоцінити менш значущі взаємодії. Крім того, вона не враховує крос-девайсну поведінку, офлайн-дотики та сезонні ефекти. Ефективна у випадках, коли всі точки контакту мають подібну вагу для досягнення бізнес-мети [21].

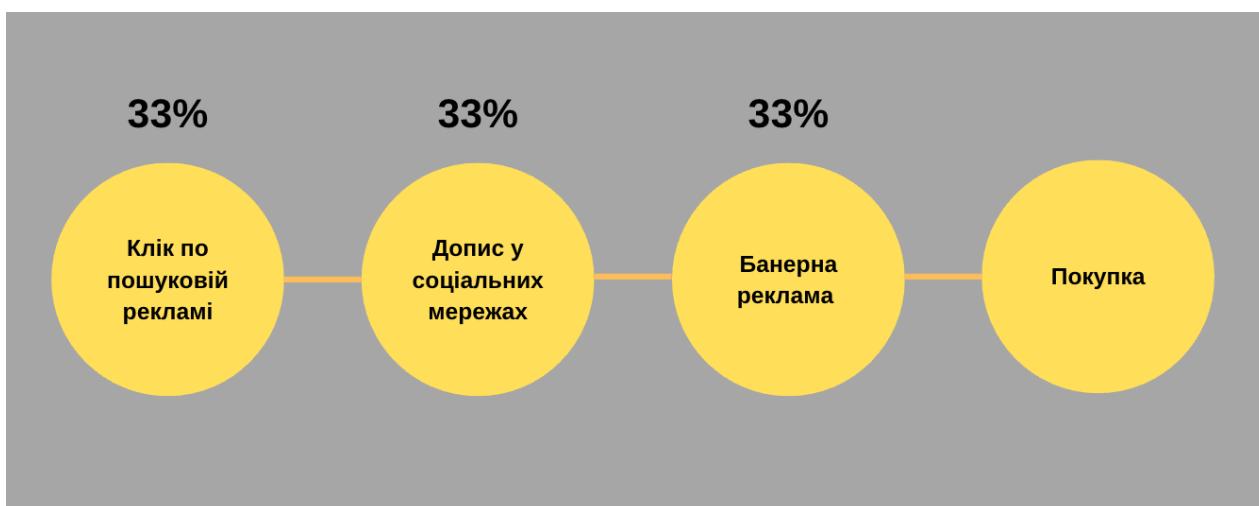


Рис. 1.6. Візуалізація моделі Linear Attribution

Джерело: Складено автором на основі [21]

4. Атрибуція з урахуванням часу спаду (Time-Decay Attribution):

На Рис. 1.7 зображено характер розподілу ваги між точками контакту відповідно до їхньої віддаленості в часі від конверсії. У такій моделі найбільше значення надається фінальним взаємодіям, наприклад, останній точці може бути приписано до 45 % загальної цінності, тоді як внесок попередніх взаємодій зменшується відповідно до часу. Це дозволяє виявити найбільш впливові канали на етапі прийняття рішення. Модель недооцінює початкові взаємодії, відповідальні за формування первинної зацікавленості. Вона складна у впровадженні, оскільки потребує вибору відповідної функції, вибір параметрів якої впливає на результати атрибуції, а отже вимагає аналітичного обґрунтування та тестування на реальних даних. Найдоцільніше використовувати в B2B-маркетингу, де рішення про покупку зазвичай приймається протягом тривалого часу та потребує багатьох дотиків і консультацій. Водночас модель є менш ефективною для бізнесів з короткими циклами продажів або імпульсивними покупками, де початкові точки контакту можуть мати вирішальне значення для формування попиту [21].



Рис. 1.7. Візуалізація моделі Time-Decay Attribution

Джерело: Складено автором на основі [21]

5. U-подібна модель (Position-Based Attribution):

U-подібна або позиційна модель (Рис. 1.8.) приписує по 40% цінності першій та останній точкам взаємодії, а решту 20% розподіляє між проміжними. Основна перевага моделі полягає у здатності відображати дві ключові точки взаємодії в клієнтській воронці: момент, коли користувач вперше дізнається про продукт, та момент, коли він ухвалює рішення про покупку або іншу цільову дію. Такий підхід дозволяє краще аналізувати коефіцієнт конверсії, повернення витрат на рекламу (ROAS) та життєву цінність клієнта (LTV). Проте модель є негнучкою — фіксований розподіл не завжди відображає реальний внесок каналів. Це означає, що визначені значення можуть не збігатися встановленим пропорціям, особливо в ситуаціях із нетиповими клієнтськими шляхами або динамічними змінами каналів. Вона також вимагає ретельного відстеження даних та аналітичної підтримки. Найкраще підходить для e-commerce та кампаній з чітко структурованими воронками, де шлях користувача зазвичай добре задокументований і природньо передбачає два основних етапи [21].



Рис. 1.8. Візуалізація U-подібної Position-Based Attribution моделі

Джерело: Складено автором на основі [21]

6. W-подібна модель (W-Shaped Attribution)

Ця модель розширює логіку U-подібної, додаючи вагу ключовому середньому дотику — зазвичай це точка кваліфікації користувача. Вона рівномірно розподіляє вагу між першою взаємодією, середньою точкою та останнім дотиком, призначає по 30% цінності. У моделі є глибше розуміння шляху клієнта і формат корисний для B2B-

компаній зі складними воронками продажів. Разом з тим, модель може нехтувати іншими важливими точками, що не потрапили до фокусної трійки [21].

7. Користувацька модель (Custom Attribution)

Користувацька модель дозволяє бізнесу самостійно визначати правила розподілу цінності між точками контакту, з урахуванням специфіки клієнтських шляхів і стратегічних цілей. Вона гнучка і точна, адаптується під індивідуальні потреби бренду. Це особливо важливо для компаній із нестандартною структурою воронки або великою кількістю взаємодій у різних каналах. У таблиці 1.1 наведено порівняльну характеристику основних моделей маркетингової атрибуції. Разом з тим, розробка такої моделі потребує глибокої експертизи, ретельного аналізу даних і регулярного перегляду. Це ресурсозатратний процес, який не завжди виправданий для малого бізнесу [21].

Таблиця 1.1.

Порівняльна таблиця основних моделей маркетингової атрибуції

Attribution model	Розподіл цінності	Найкраще застосування
First Touch	Уся цінність приписується першій точці взаємодії.	Вимірювання ефективності початкової обізнаності про бренд.
Last Touch	Уся цінність приписується останній точці взаємодії.	Визначення остатнього впливу на прийняття рішення.
Linear	Цінність рівномірно розподіляється між усіма точками контакту.	Для оцінки всіх точок як однаково важливих у процесі конверсії.
Time-decay	Більше цінності надається останнім взаємодіям, менше - попереднім.	Для довших циклів продажу або у випадку високого впливу останніх взаємодій до конверсії.

Position-based	Цінність розподіляється між різними точками (залежно від моделі або платформи).	Коли важливими є і перша, і остання взаємодії.
Custom	Цінність надається відповідно до того, що є найважливішим для бізнесу.	Найкраще для компаній зі специфічними цілями або унікальними шляхами клієнтів.

Джерело: складено автором

1.4. Недоліки традиційного підходу та роль машинного навчання

Отже, традиційні моделі маркетингової атрибуції, зокрема моделі одного дотику (first-touch, last-touch) або навіть мультиточкові моделі, що засновані на правилах (linear, position-based), мають істотні обмеження в контексті сучасного цифрового середовища. Основна їхня слабкість полягає в застосуванні жорстко визначених правил, які однаково інтерпретуються для всіх користувачів незалежно від специфіки їх поведінки. Такі моделі не враховують складну природу клієнтських шляхів про яку неодноразово зазначалось, взаємозв'язки між каналами та змінність поведінки аудиторії у часі. Крім того, ці підходи не здатні адаптуватися до нових паттернів взаємодії або реагувати на нестандартні ситуації в середовищі цифрового маркетингу. [22]

Застосування машинного навчання в атрибуції дозволяє подолати ці обмеження за рахунок побудови адаптивних моделей, які вивчають історичні дані користувачів і виявляють закономірності у взаємодіях з маркетинговими каналами. Найбільш поширеною формою є модель атрибуції на основі даних (Data-Driven Attribution), яка використовує алгоритми класифікації, байєсівський аналіз або ансамблеві методи для розподілу цінності між каналами на основі їх фактичного внеску у конверсію [23].

На відміну від моделей, що засновані на правилах алгоритмічні моделі не спираються на ці заздалегідь визначені правила, а обчислюють ймовірнісний вплив кожного дотику з урахуванням повної картини клієнтських взаємодій. Зокрема, Data-Driven Attribution формує набір вагових коефіцієнтів, що динамічно оновлюються

залежно від зміни поведінки користувачів. Алгоритм враховує як кількісні, так і часові характеристики дотиків, дозволяючи краще ідентифікувати реальні драйвери конверсії.

Такі моделі дають вищу точність оцінки, зменшує упередження в аналізі та підтримує прийняття обґрунтованих маркетингових рішень. Разом з тим, для ефективного впровадження моделей на основі машинного навчання потрібен доступ до якісних та об'ємних даних, аналітична інфраструктура і відповідна технічна експертиза. Саме тому такі моделі найчастіше використовуються великими компаніями з розвинутою системою цифрової аналітики, зокрема у сферах e-commerce, фінансових послуг та B2B-маркетингу з довгими циклами прийняття рішень.

Роль машинного навчання в атрибуції продовжує зростати, оскільки бізнес прагне до більшої персоналізації, точнішого прогнозування результатів і оптимізації витрат на маркетингові кампанії в режимі реального часу.

ВИСНОВКИ ДО РОЗДІЛУ 1

У першому розділі було розкрито теоретичні засади маркетингової атрибуції як ключового інструменту цифрової аналітики. Визначено її сутність як процесу розподілу цінності конверсії між точками взаємодії користувача з брендом та сформульовано основні завдання: підвищення прозорості ефективності каналів, оптимізація бюджету й підкріплення управлінських рішень достовірними даними.

Окрема увага була приділена ключовим викликам, з якими стикається сучасний маркетинг: фрагментованість шляхів клієнта, мультиканальність, поява нових цифрових середовищ, складність у відстеженні крос-девайсних взаємодій та зростаючі вимоги до конфіденційності даних. У цьому контексті маркетингова атрибуція постає як відповідь на необхідність розуміння реального внеску кожного каналу в досягнення бізнес-результатів.

Розгляд традиційних моделей атрибуції (First Click, Last Click, Linear, Position-Based) дозволив виявити їхні переваги, проте також і значущі обмеження — передусім у здатності точно відображати складну природу поведінки споживачів. На цьому тлі все

більшої ваги набувають підходи на основі машинного навчання, які враховують часову динаміку, послідовність дотиків та їхню статистичну значущість у формуванні конверсії.

Подальше дослідження буде сфокусоване на аналізі алгоритмів машинного навчання, які будують більш точні та адаптивні атрибуційні моделі. У другому розділі будуть розглянуті стохастичні та імовірнісні підходи до моделювання впливу каналів, зокрема ланцюги Маркова, значення Шеплі та критерії їхньої ефективності.

РОЗДІЛ 2. АЛГОРИТМИ МАШИННОГО НАВЧАННЯ ДЛЯ МОДЕЛЮВАННЯ АТРИБУЦІЇ

2.1. Загальна характеристика методів машинного навчання у маркетингу.

Машинне навчання (ML) — це галузь штучного інтелекту, яка зосереджена на розробці алгоритмів, здатних вчитися на даних і покращувати свою продуктивність без явного програмування [24].

У маркетинговій практиці ML використовується для аналізу великих обсягів даних, виявлення закономірностей у поведінці споживачів, автоматизації прийняття рішень та прогнозування результатів різноманітної маркетингової діяльності. Цей інструмент використовується у вирішенні таких завдань, як сегментація клієнтів, персоналізація пропозицій, прогнозування відтоку користувачів, динамічне ціноутворення та, зокрема, атрибуція [25].

Методи машинного навчання поділяються на три основні групи: машинне навчання з учителем (Supervised machine learning), машинне навчання без учителя (Unsupervised machine learning) та навчання з підкріпленням (Reinforcement learning) [25]. Машинне навчання з учителем – це тип машинного навчання, де модель навчається на основі маркованих даних, де зв'язок вхід-вихід відомий. До алгоритмів такого типу відноситься регресія, класифікація, дерева рішень, наївні баєсівські класифікатори і нейронні мережі. У маркетингу вони використовуються для прогнозування дій користувача, ймовірності конверсії, або ж визначення відтоку клієнтів. У свою чергу, методи навчання без учителя, алгоритми, які працюють з немаркованими даними для виявлення закономірностей або груп, — наприклад, кластеризація — допомагають виявити схожі поведінкові паттерни серед користувачів. Сегментація аудиторії виступає базою для більш персоналізованої взаємодії. Навчання з підкріпленням характеризується тим, що алгоритми навчаються, взаємодіючи з середовищем та отримуючи зворотний зв'язок у вигляді винагород або штрафів. У маркетинговій сфері ця група використовується рідше, однак може бути ефективним для оптимізації бюджетів або керування стратегіями в умовах динамічного середовища [26].

Однією з ключових переваг машинного навчання в маркетингу є здатність обробляти великі обсяги як структурованих, так і неструктурованих даних. Алгоритми ML моделюють складні взаємозв'язки між каналами, і здатні обліковувати часову послідовність дотиків, а також виявляти нетипові, нелінійні шляхи до конверсії.

Відповідно така здатність методів ML може покращити моделі маркетингової атрибуції, які спрямовані на розуміння впливу кожної точки контакту з маркетингом на рішення клієнта про конверсію.

Реальний приклад компанії Choreograph демонструє ці переваги: кейс-дослідження показало, що рішення для впровадження багатомодельної атрибуції призвело до інкрементальних продажів на суму 12,9 мільйона доларів США та скорочення часу до отримання відомостей на 83% (щомісячне оновлення результатів порівняно з очікуванням понад 6 місяців при традиційному моделюванні маркетингового міксу). Можливість сегментувати відомості за різними вимірами (наприклад, за SKU, географічними регіонами, етапами шляху клієнта) та отримувати їх часто (щомісячні оновлення) перетворило атрибуцію з історичного звітування на динамічний інструментарій оптимізації. Компанія розпочала гнучкий маркетинг та постійне вдосконалення протягом усього життєвого циклу клієнта [27]. Розуміючи переваги методів ML, перейдемо до опису основних алгоритмічних методів, які використовуються безпосередньо для моделювання маркетингової атрибуції.

2.2. Ланцюги Маркова як основа стохастичного моделювання

Ланцюги Маркова – це тип ймовірнісної моделі. Вони представляють набір подій, що відбуваються послідовно та пов'язують кожну подію з кожною іншою подією за допомогою умовних ймовірностей. У контексті маркетингової атрибуції вони демонструють ймовірність того, що клієнт перейде від одного кроку (каналу або дії) у шляху клієнта до іншого, і залежать лише від попередніх кроків.

Ключові компоненти ланцюга Маркова включають:

- Стани (States): Це окремі умови або позиції в системі. У маркетинговій атрибуції стани зазвичай відповідають різним маркетинговим точкам взаємодії, наприклад, веб-сторінки, які користувач може відвідати під час сеансу перегляду (головна сторінка, сторінка “про нас”, сторінка контактів).
- Переходи (Transitions): Це рухи або зміни з одного стану в інший. Кожен перехід пов'язаний з імовірністю, яка кількісно визначає ймовірність переходу процесу з одного стану в інший. Ці ймовірності є вирішальними для розуміння динаміки модельованої системи та повинні задовольняти умову, що сума ймовірностей переходу з будь-якого заданого стану в усі інші стани (включаючи можливе перебування в тому самому стані) дорівнює 1 [28].

Модель функціонує на принципі відсутності пам'яті. Це той визначальний принцип, який стверджує, що майбутній стан залежить лише від поточного стану, а не від шляху, яким він був досягнутий. Така властивість дозволяє спростити моделювання складних процесів взаємодії користувача з каналами маркетингової комунікації, зберігаючи при цьому точність обчислення впливу кожного каналу на кінцеву конверсію.

Ланцюги Маркова характеризуються визначеним простором станів, матрицею переходів (що описує ймовірності між станами) та початковим станом або розподілом.

Побудований граф репрезентує точки взаємодії, між якими існують переходи, що відповідають певним ймовірностям. Побудова графа ґрунтується на історичних даних про взаємодії користувачів. Конверсія моделюється як кінцевий стан, а відмова (null) — як альтернативний завершальний стан. Для оцінки значущості кожного каналу застосовується підхід «removal effect», тобто аналіз того, як видалення конкретного вузла впливає на ймовірність досягнення конверсії (покупка, інша основна дія) [29].

Процес розрахунку складається з декількох етапів. Спочатку визначаються ймовірності переходів між усіма вузлами графа на основі частоти траєкторій, отриманих з даних, Рис. 2.1.

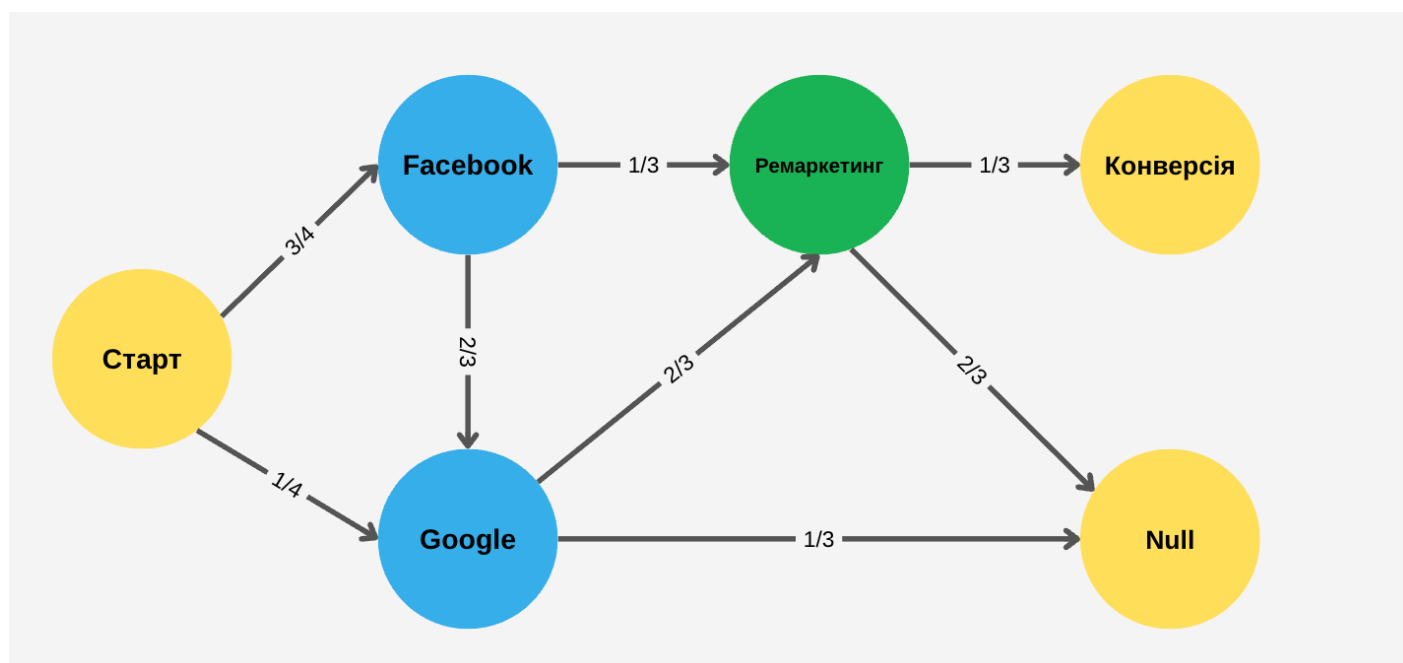


Рис.2.1. Ймовірнісний граф переходів у моделі ланцюгів Маркова

Джерело: Складено автором на основі [29]

Рис.2.1. ілюструє умову, де у користувача є 4 шляхи, 2 з яких ведуть до конверсії, отже загальний коефіцієнт конверсії становить 50%. Кожна вершина (вузол) у графі представляє канал або стан взаємодії користувача з брендом, а дуги (стрілки), як і зазначалось — ймовірність переходу між цими станами.

Початковим станом є вершина «Старт», з якої 75% користувачів переходять до каналу Facebook, а решта 25% — одразу до Google. Далі з Facebook користувачів переходять до Google, а $\frac{1}{3}$ потрапляє у процес ремаркетингу. З Google $\frac{2}{3}$ трафіку перенаправляється до ремаркетингу, а $\frac{1}{3}$ завершує шлях без конверсії (стан Null). Із ремаркетингу можлива або конверсія (з ймовірністю $\frac{1}{3}$), або завершення без результату (ймовірність $\frac{2}{3}$).

Розраховуючи загальну ймовірність конверсії для всіх можливих шляхів у графі, маємо:

$$P = \frac{3}{4} \times \frac{1}{3} \times \frac{2}{3} + \frac{3}{4} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} + \frac{1}{4} \times \frac{2}{3} \times \frac{2}{3} = \frac{1}{2}$$

Після оцінки базової ймовірності конверсії, поступово видаляється кожен вузол (канал) з графа. Для кожного такого видалення повторно обчислюється ймовірність досягнення стану конверсії, Рис. 2.2.

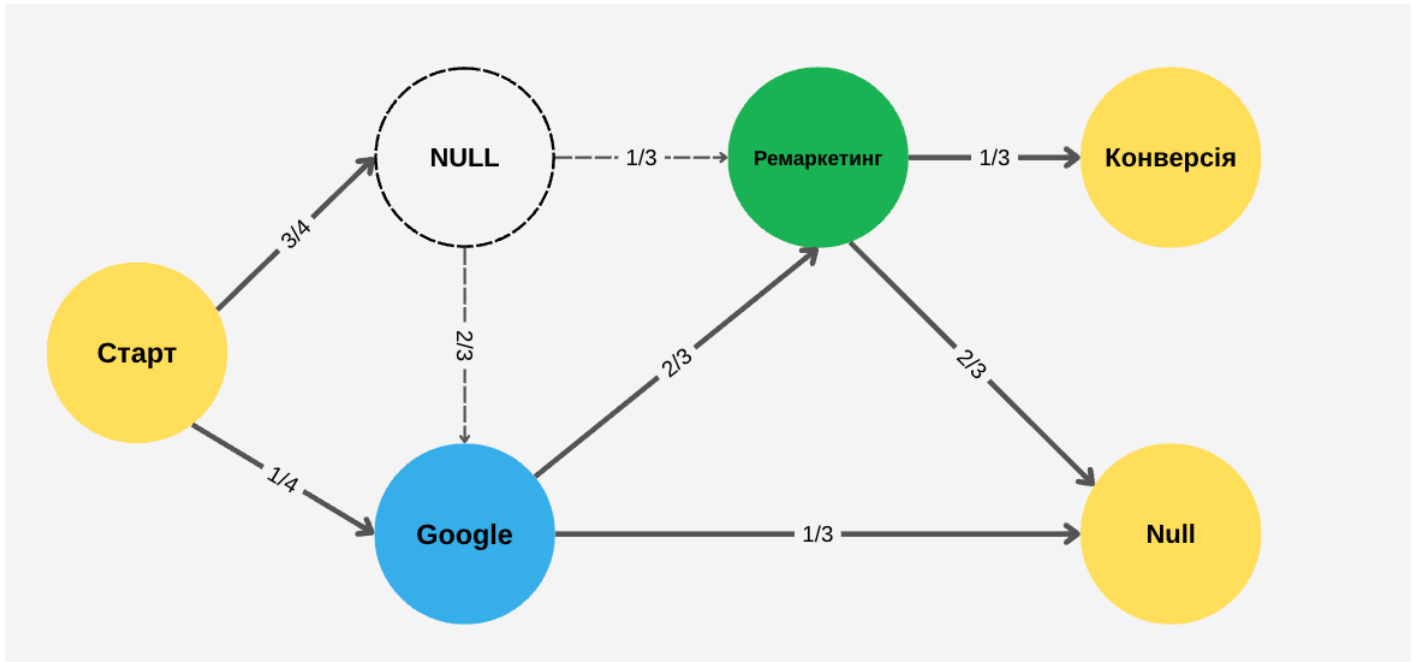


Рис.2.2. Граф переходів у моделі ланцюгів Маркова з видаленням каналу Facebook
Джерело: Складено автором на основі [29]

Розрахуємо ймовірність конверсії в усіченому графі, де видалено вузол взаємодії – Facebook. За таких умов дуги, що виходять з цього вузла перестають існувати. Маємо один можливий шлях користувача до конверсії, де ймовірність складає:

$$P = \frac{1}{4} \times \frac{2}{3} \times \frac{2}{3} = \frac{1}{9}$$

Аналогічні розрахунки проводяться для кожного вузла графа. Наступним етапом є обчислення ефекту видалення, який розраховується за формулою:

$$Removal\ effect\ (i) = 1 - \left(\frac{P(\text{conversion without node } i)}{P(\text{conversion with node } i)} \right)$$

Ефект видалення відображає відсоток усіх конверсій, які можуть будуть втрачені, за умови видалення даної точки контакту. Відповідний показник було обраховано для кожного каналу у таблиці 2.1.:

Таблиця 2.1.

Розрахунок атрибуції за моделлю ланцюгів Маркова

Канал	Ймовірність конверсії без даного каналу	Removal Effect	Частка в результаті
Facebook	$\frac{1}{9}$	$1 - \frac{\frac{1}{9}}{\frac{1}{2}} = 77,8\%$	$\frac{77,8}{244,4} = 0,32$
Google	$\frac{1}{6}$	$1 - \frac{\frac{1}{6}}{\frac{1}{2}} = 66,7\%$	$\frac{66,7}{244,4} = 0,27$
Ремаркетинг	0	$1 - \frac{0}{\frac{1}{2}} = 100\%$	$\frac{100}{244,4} = 0,41$
Усього	$\frac{1}{2}$	244,4%	1,00

Джерело: складено автором

При виключенні каналу Facebook ймовірність конверсії знижується з 50% до 11,1%, що дає ефект видалення на рівні 77,8%. Аналогічно, для Google цей показник становить 66,7%, а для Remarketing — 100%, що свідчить про його критичну роль у досягненні конверсії.

Однак сума всіх ефектів видалення не дорівнює 100% — у наведеному випадку вона становить 244,4%. Це пояснюється тим, що кожен канал бере участь у кількох шляхах одночасно, і їхні внески частково накладаються. Для уникнення дублювання та забезпечення пропорційного порівняння важливості каналів застосовується процедура нормалізації ефектів видалення. Нормалізацію виконується, діленням кожного ефекту видалення на суму всіх ефектів видалення. Цей підхід масштабує всі значення так, щоб їхня сума дорівнювала 1, що дозволяє кожному значенню представляти частину цілого.

Отже, найбільший внесок у досягнення конверсії демонструє канал Remarketing (41%), далі — Facebook (32%) і Google (27%).

При цьому кожна частка відображає відносний внесок каналу у загальну ймовірність досягнення конверсії. Відповідно, множення цих часток на фактичну кількість конверсій дає змогу отримати кількісну оцінку впливу кожного каналу в абсолютному вимірі. У розглянутому прикладі загальна кількість конверсій становить 2 тому кінцева атрибуція для кожного каналу розраховується шляхом множення його частки у результаті на значення 2. У таблиці 2.2. наведено прорахунок нормалізованої трибуційної частки конверсії за моделлю ланцюгів Маркова.

Таблиця 2.2.

Атрибуція конверсій за моделлю ланцюгів Маркова

Канал	Частка в результаті	Атрибуція конверсій
Facebook	0,32	$0,32 \times 2 = 0,64$
Google	0,27	$0,27 \times 2 = 0,54$
Ремаркетинг	0,41	$0,41 \times 2 = 0,82$
Усього	1,00	2,00

Джерело: складено автором

Ланцюги Маркова ефективно враховують не лише послідовність точок дотику, а й їх взаємозалежність, включаючи наявність циклів (петель), які виникають, коли користувач повторно взаємодіє з одним і тим самим каналом. Для складніших графів із великою кількістю циклів застосовуються ітеративні методи наближеного розрахунку, які оцінюють ймовірність конверсії навіть при нескінченній кількості можливих шляхів.

Окрему увагу привертають моделі Маркова вищого порядку, які надають системі «пам'ять»: у ланцюгах другого порядку ймовірність наступного стану залежить не тільки від поточного, а й від попереднього. Це дає змогу враховувати контекст, наприклад, ефективність ремаркетингу після попередньої взаємодії з Facebook або Google [30].

Незважаючи на те, що підхід до моделювання ґрунтуються на статистичних припущеннях і не гарантують виявлення причинно-наслідкових зв'язків, вони є надійним підходом для аналізу каналів у разі наявності великих масивів даних.

Практична реалізація атрибуції на основі ланцюгів Маркова потребує доступу до як конверсійних, так і неконверсійних шляхів користувача. Оскільки деякі інструменти аналітики, зокрема Google Analytics, обмежують доступ до неконверсійних даних, для повноцінного застосування моделі варто використовувати розширені джерела даних. За відсутності таких даних можливе спрощене використання Марковської атрибуції лише для конверсійних шляхів, однак це знижує точність оцінки каналів, що не ведуть безпосередньо до конверсії [30].

Загалом, використання ланцюгів Маркова особливо виправдане в умовах наявності довгих шляхів до конверсії та великої кількості взаємодій, де традиційні моделі формату last-click або linear не дають достатньої точності.

Основна її перевага полягає в тому, що модель враховує послідовність і взаємозв'язки між каналами. Вона дійсно дозволяє аналізувати не просто факт присутності каналу у шляху користувача, а й те, у якому порядку він взаємодіяв із каналами, щоб точніше оцінити реальний внесок кожного етапу у здійснення конверсії. Цінним є здатність виявляти критичні точки шляху. Завдяки показнику removal effect (ефекту видалення), модель показує, наскільки сильно зменшиться ймовірність конверсії, якщо виключити певний канал. Таким чином, маркетологи мають змогу виявити ключові точки впливу для бізнесу і посилити спрямування зусиль у відповідні канали. Ще однією важливою перевагою є гнучкість і здатність масштабуватися, добре адаптується до великої кількості каналів і складних шляхів користувачів. Варто виділити меншу чутливість до «шуму» - випадкових даних або клік-спаму: атрибуція за допомогою ланцюгів Маркова менш схильна до спотворення випадковими даними з низькою статистичною значущістю.

Проте, незважаючи на свої переваги, модель ланцюгів Маркова має і низку недоліків. Найбільш суттєвим з них є відсутність причинно-наслідкового аналізу. Модель базується на кореляційному аналізі, який вказує на потенційні зв'язки, але не доводить причинності. Вона показує зв'язки між каналами та конверсією, але не може довести, що саме певна взаємодія викликала цільову дію. Тобто весь алгоритм заснован

на спостережуваних кореляціях в історичних даних, а не справжній причинно-наслідковий висновок, який можна отримати з контрольованого експерименту (наприклад, А/В-тестування). Тому, навіть якщо гіпотетичне видалення каналу значно знижує ймовірність конверсії в моделі, це не доводить, що канал спричинив ці конверсії в реальному світі. Модель може приписувати заслуги каналам, які просто присутні на успішних шляхах, а не є рушіями успіху. Це внутрішнє обмеження означає, що моделі Маркова можуть не ефективно виявляти такі явища, як «викрадення конверсій», коли канал просто отримує заслуги за конверсії, які вже мали відбутися [31].

Ще одним обмеженням є те, що модель спрощує дійсність за рахунок принципу відсутності пам'яті. У класичному варіанті модель припускає, що наступний стан залежить лише від поточного, і не враховує попередні переходи. Це може спотворювати реальну поведінку, особливо у довгих або повторюваних шляхах.

Реалізація моделі стає складною у разі петель і великих графів, адже обчислення ймовірностей переходів та їх нормалізація у таких випадках є математично трудомісткими і потребують використання спеціалізованого програмного забезпечення.

До того ж модель може давати неточні результати для одноетапних шляхів, коли конверсія відбувається після взаємодії лише з одним каналом: у такому випадку нормалізований ефект видалення може несправедливо зменшити внесок цього каналу.

Для побудови надійної матриці ймовірностей переходів потрібен великий обсяг даних про переходи станів для побудови надійної матриці ймовірностей переходів, без якого результати моделі можуть бути нестійкими та ненадійними [32].

Відповідно попередніх пунктів, модель ланцюгів Маркова має міцний інструментарій для оцінки ефективності каналів у складному середовищі цифрового маркетингу і ряд вирішальних переваг, однак вимагає уважного ставлення до припущень, вхідних даних та інтерпретації результатів.

2.3. Методи значень Шеплі для моделювання атрибуції

Значення Шеплі – це концепція, що належить до математичної підгалузі теорії ігор. Основна проблема, яку вирішує значення Шеплі, полягає у справедливому розподілі прибутку в грі, в якій гравці можуть формувати коаліції.

У контексті маркетингової атрибуції теорія ігор використовується для моделювання взаємодії клієнтів з маркетинговими каналами як кооперативної гри, де кожен маркетинговий канал можна розглядати як гравця в грі. Сукупність усіх гравців/каналів працює разом для стимулювання конверсій та призначає кожній точці контакту справедливий кредит (використовуючи ці значення Шеплі) за конверсію на основі їхнього справжнього внеску [33].

Під справедливим розподіленням розуміємо відповідність чотирьом основним аксіомам: ефективність, симетрія, лінійність та фіктивний гравець .

1. Ефективність: сума значень Шеплі, розподілених між усіма учасниками коаліції, повинна дорівнювати загальному результату, досягнутому цією коаліцією. Жодна частка не губиться і не додається в процесі розподілу: весь прибуток (або всі конверсії) мають бути повністю розподілені між каналами.
2. Симетрія: якщо два гравці (канали) мають однаковий внесок у всі можливі коаліції, вони повинні отримати однакові значення Шеплі. Тобто, якщо вплив каналів А і В на конверсію є ідентичним у всіх сценаріях, їхній атрибуційний вклад також має бути рівним.
3. Лінійність: якщо ми маємо дві різні гри (наприклад, дві кампанії з різною конверсійною статистикою) і поєднуємо їх у єдину гру, то значення Шеплі для кожного гравця в об'єднаній грі має дорівнювати сумі його значень у кожній окремій грі. Ця властивість гарантує, що модель залишатиметься стабільною при зміні масштабу або при об'єднанні даних.
4. Фіктивний гравець: якщо певний гравець не змінює результат у жодній коаліції, тобто його присутність або відсутність не впливає на ймовірність виграшу, то

його значення Шеплі повинне дорівнювати нулю. Таким чином, тільки ті канали, які реально мають вплив, отримують відповідну частку [33].

У процесі розрахунку значень Шеплі для маркетингових каналів оцінюється граничний внесок кожного окремого каналу шляхом порівняння прогнозу моделі за наявності цього каналу та без нього. Різниця між цими двома прогнозами дозволяє кількісно визначити, яку додану цінність забезпечує відповідний канал для результату моделі — наприклад, ймовірності конверсії. Така оцінка проводиться для всіх можливих підмножин каналів, після чого обчислюється усереднене значення маржинального ефекту. Отримане усереднене значення і є значенням Шеплі для відповідного маркетингового каналу і відображає його справедливую частку впливу на загальний результат [34].

Розглянемо приклад роботи алгоритму Шеплі із двома конверсійними шляхами, які відрізняються наявністю або відсутністю певного маркетингового каналу.

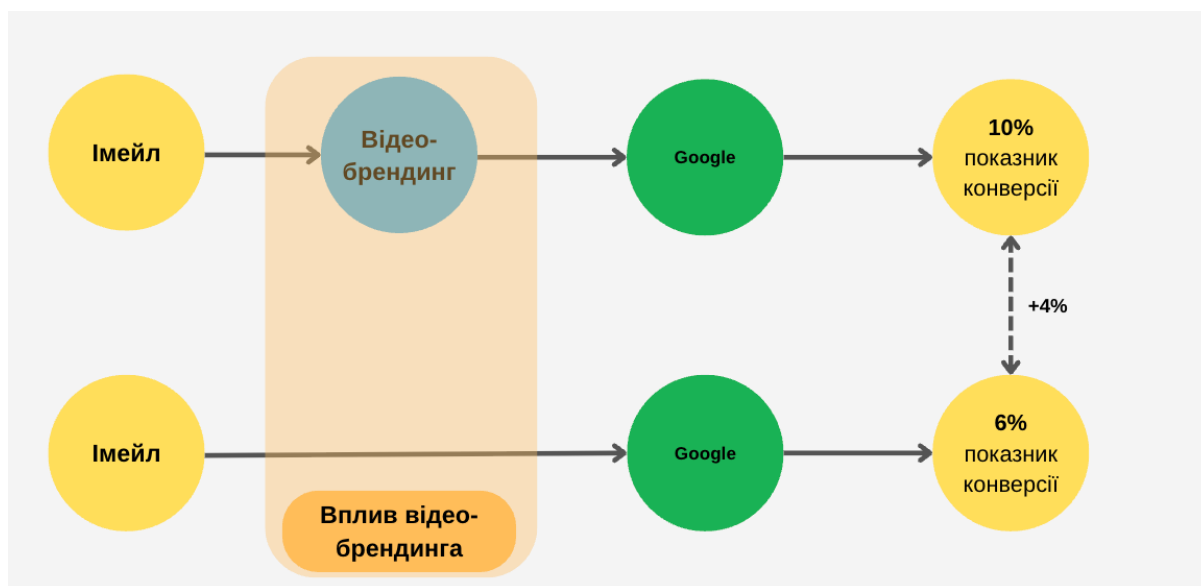


Рис.2.3. Оцінка граничного внеску каналу «Відео-брендинг» у моделі Шеплі

Джерело: Складено автором на основі [29]

Відповідно Рис.2.3. - канал «Відео-брендинг» присутній лише на одному з двох порівнюваних шляхів. Якщо шлях із цим каналом має коефіцієнт конверсії 10%, а без нього — лише 6%, то вплив цього каналу становить 4% у абсолютному вираженні або приблизно 67% у відносному. Такий внесок є суттєвим, і, відповідно, канал «Відео-

брендинг» отримає високе значення Шеплі. Це свідчить про те, що він справді створює значущу додану вартість у межах коаліцій каналів.

На Рис. 2.4. розглядається внесок каналу «Google», який збільшує ймовірність конверсії лише на 0,2% у абсолютному значенні (приблизно 2% у відносному). Вважаємо вплив незначним, йому присвоюється значно нижче значення Шеплі, порівняно з каналом «Відео-брендинг». Формально значення Шеплі для каналу i у множині каналів N із характеристичною функцією v визначається як:

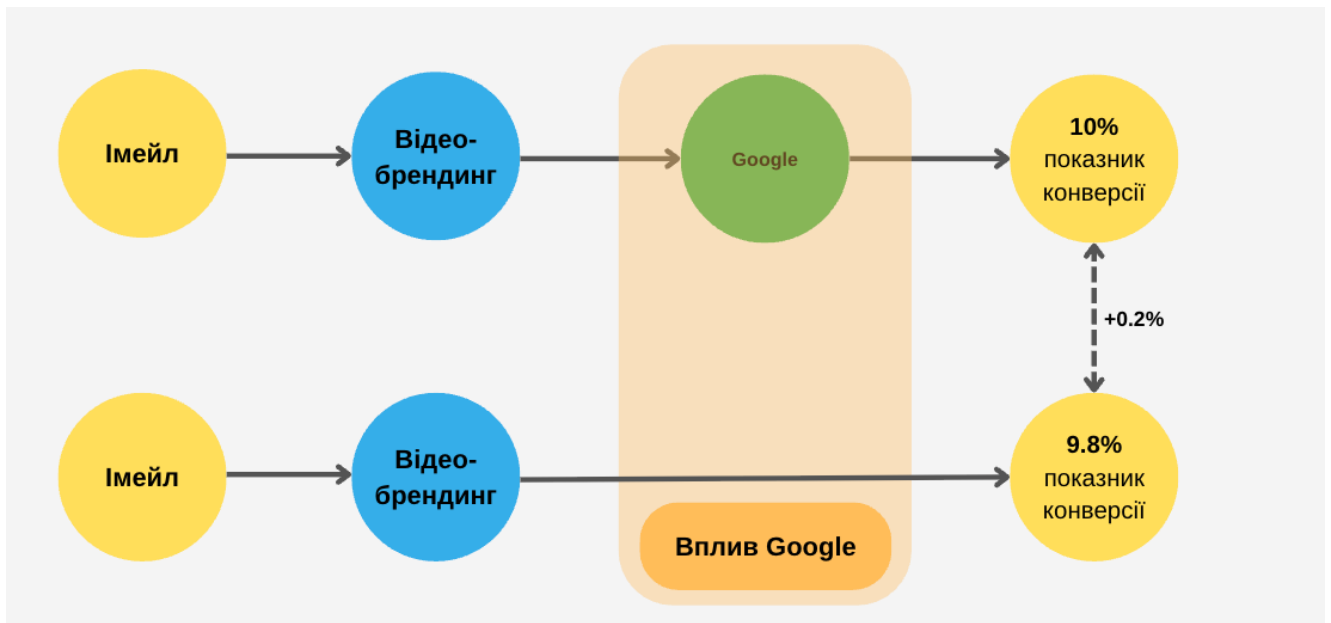


Рис.2.4. Оцінка граничного внеску каналу «Google» у моделі Шеплі

Джерело: Складено автором на основі [29]

$$\phi_i(N, v) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|! (|N| - |S| - 1)! \cdot [v(S \cup \{i\}) - v(S)]$$

Де:

- N – множина всіх каналів (гравців),
- $S \subseteq N \setminus \{i\}$ – підмножина каналів, що не включає канал i
- $v(S)$ – функція виграшу (коефіцієнт конверсії), який генерується коаліцією S
- $v(S \cup \{i\}) - v(S)$ - граничний внесок каналу i до коаліції S

- $|S|! (|N| - |S| - 1)!$ - вагові множники, що враховують кількість можливих перестановок каналів до та після включення каналу i

Шлях користувача у наведеному прикладі Імейл → Відео-брендинг → Google дає коефіцієнт конверсії 10%:

$$v(\text{Імейл, Відео – брендинг, Google}) = 10\%$$

Шлях без Google дає 9,8%:

$$v(\text{Імейл, Відео – брендинг, }) = 9,8\%$$

Граничний внесок каналу Google до цієї коаліції становить 0,2%. Значення Шеплі врахує цей внесок, а також аналогічні внески для всіх інших коаліцій без Google, і усереднить їх, формуючи підсумковий коефіцієнт атрибуції.

На практиці модель Шеплі активно використовується у системах аналітики, зокрема у великих рекламних платформах, таких як Google Ads Attribution або Facebook Attribution, для оцінки впливу каналів на конверсійний результат [36]. У межах маркетингової аналітики її реалізація відбувається за допомогою аналізу повного масиву шляхів користувача до конверсії та визначення граничного внеску кожного каналу для усіх можливих коаліцій каналів. Оскільки загальна кількість можливих комбінацій каналів зростає експоненційно залежно від кількості учасників, практична реалізація потребує значних обчислювальних ресурсів. Саме тому на практиці часто використовують апроксимації на основі вибіркового перестановок або інструменти SHAP (SHapley Additive exPlanations), які дозволяють скоротити обчислювальні витрати та масштабувати модель до великих наборів даних [34].

До основних переваг моделі значень Шеплі належить її здатність забезпечувати справедливий розподіл цінності між каналами, адже вона гарантує пропорційний розподіл на основі реального внеску кожного з них і розглядає весь шлях користувача як цілісну кооперативну гру. Це дає змогу уникнути упередженості, притаманної моделям із фіксованими правилами, такими як Last Click чи First Click.

Важливою перевагою є також те, що модель враховує взаємодію каналів, оцінюючи їхній спільний вплив на конверсію, завдяки чому краще відображається синергія між

ними. Підхід є інтерпретованим, адже пропонує універсальне розуміння внеску ознак і дає можливість аналітикам обґрунтувати свої висновки перед бізнесом.

Ще одна сильна сторона цієї моделі полягає в її високій прогностичній точності: дослідження Тильбургського університету показали, що рішення на основі значень Шеплі досягають найвищою прогностичної здатності порівняно з іншими моделями атрибуції на основі даних, такими як ланцюги Маркова та логістична регресія [35].

Разом з тим, модель має і низку недоліків, серед яких однією з головних проблем є її обчислювальна складність і висока вартість. Обчислення точних значень Шеплі є надзвичайно ресурсозатратним процесом при великій кількості каналів, оскільки кількість можливих коаліцій зростає експоненційно, що ускладнює практичне застосування моделі без використання апроксимаційних методів і потребує значних обчислювальних потужностей. Для невеликих компаній це часто виявляється надто витратним.

Крім того, значення Шеплі можуть бути чутливими до змін у даних, тобто значення можуть значно змінюватися при додаванні або видаленні каналів, і призводить до нестабільності в атрибуції та ускладнює інтерпретацію результатів експертами.

Ще одним обмеженням, як і у ланцюгах Маркова, є потреба у великому обсязі даних, адже для точного обчислення значень Шеплі потрібні деталізовані набори про шляхи користувачів, що є проблематичним для компаній із обмеженими ресурсами або для нових проєктів, де таких даних бракує.

Класична модель Шеплі не враховує порядок, у якому користувачі взаємодіють з каналами, тому якщо деякі маркетингові сценарії використовують фіксовану воронку продажів, то варто розробляти розширені версії моделі.

Отже, модель Шеплі дозволяє об'єктивно та математично обґрунтовано оцінити внесок кожного каналу у досягнення конверсії. Її головна перевага – справедливий розподіл цінності, який базується на реальному впливі каналів у різних комбінаціях. Незважаючи на високу обчислювальну складність, її практичне застосування стає значно простішим завдяки розвитку технологій обробки даних, широкій підтримці в

аналітичних платформах і використанню ефективних алгоритмічних наближень. У Google Cloud (Vertex AI, BigQuery ML) реалізовано підтримку SHAP для інтерпретації моделей [38].

2.4. Порівняльний аналіз моделей атрибуції на основі ланцюгів Маркова і значень Шеплі

У таблиці 2.3. наведено порівняльний аналіз основних характеристик моделей атрибуції на основі ланцюгів Маркова і значень Шеплі.

Таблиця 2.3.

Порівняльний аналіз моделей атрибуції на основі ланцюгів Маркова і значень Шеплі

Ознака порівняння	Модель на основі ланцюгів Маркова	Модель на основі значень Шеплі
Тип моделі	Ймовірнісна, послідовна	Теоретико-ігрова, кооперативна
Аналіз контексту порядку каналів	Враховується порядок переходів	Порядок не є критичним
Залежність від часової структури	Висока	Мінімальна
Урахування взаємодій між каналами	Обмежене, базується на переходах	Повне, враховуються всі коаліції
Стійкість до зміни розміру вибірки	Висока при великих вибірках	Залежить від обчислювальних ресурсів
Обчислювальна складність	Помірна	Висока (експоненційне зростання)

Гнучкість в адаптації до складних сценаріїв	Обмежена моделлю 1-го порядку	Висока завдяки коаліційному підходу
Інтерпретованість результатів	Висока, але потребує пояснення матриці ймовірностей переходів	Висока, особливо при візуалізації SHAP
Залежність від специфікації функції прибутку (бізнес-цілі)	Не потребує явного визначення	Необхідна характеристична функція
Актуальність у сучасних інструментах	Поширена в web-аналітиці	Вбудована в сучасні ML-платформи.

Джерело: складено автором

Порівняльний аналіз показав, що модель ланцюгів Маркова ефективна, коли важливий порядок дій користувача. Вона моделює ймовірності переходів між каналами та допомагає виявити ключові точки впливу. Модель легко масштабується та потребує менше ресурсів, що зручно для класичної веб-аналітики. Проте вона враховує лише переходи між каналами й не фіксує складних взаємодій між ними.

Натомість модель Шеплі аналізує всі можливі комбінації каналів і точно оцінює внесок кожного в конверсію. Підхід особливо корисний при персоналізованих кампаніях і складних шляхах користувача. Проте потребує більше обчислювальних ресурсів і точного налаштування, що ускладнює її використання в реальному часі. Також є ризик недооцінити рідкісні або малопомітні канали.

Отже, вибір між моделлю ланцюгів Маркова та підходом Шеплі залежить від аналітичних потреб бізнесу, складності користувацьких шляхів, наявних ресурсів та рівня точності, необхідного для стратегічного прийняття рішень. [37]

ВИСНОВКИ ДО РОЗДІЛУ 2

У другому розділі було досліджено сучасні підходи до моделювання маркетингової атрибуції на основі методів машинного навчання. Спершу було розглянуто загальні

можливості ML у сфері цифрового маркетингу, зокрема його здатність обробляти великі обсяги даних, виявляти поведінкові паттерни та підвищувати ефективність аналітики завдяки прогнозуванню і автоматизації. Алгоритмічні методи машинного навчання відкривають нові можливості для побудови обґрунтованої системи розподілу маркетингових ресурсів і стратегічного управління каналами комунікації. Далі детально проаналізовано модель атрибуції на основі ланцюгів Маркова, яка моделює ймовірності переходів між точками контакту, враховує послідовність дій користувача та дає змогу оцінити критичні канали за допомогою ефекту видалення.

Особливу увагу приділено моделі значень Шеплі. Спираючись на теорію кооперативних ігор, модель забезпечує справедливий розподіл внеску між усіма каналами шляхом оцінки їхнього граничного ефекту в усіх можливих коаліціях.

У завершальному підпункті проведено порівняльний аналіз обох моделей, який показав, що модель Маркова краще враховує часову структуру взаємодій, тоді як модель Шеплі - глибше аналізує взаємодію каналів незалежно від порядку. Обидві моделі мають практичне застосування, але за умов складної воронки та високої потреби у точності атрибуції перевагу варто надати моделі Шеплі.

РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ АТРИБУЦІЇ

3.1. Постановка задачі та методологія моделювання маркетингової атрибуції

Популярні моделі маркетингової атрибуції мають свої переваги та недоліки, і вибір найкращої моделі залежить від індивідуальних потреб бізнесу. Однак основна спільна характеристика популярних моделей полягає в тому, що вони базуються на правилах, і бізнес повинен заздалегідь вирішити, як він хоче, щоб кредит за конверсії продажів розподілявся між каналами. Відповідно постає потреба у розробці моделі атрибуції, яка здатна більш справедливо розподіляти значення конверсії між усіма каналами, залученими до шляху користувача.

Цей розділ присвячено практичній реалізації моделі атрибуції для аналізу наявного набору даних, що містить послідовності маркетингових контактів користувачів та інформацію про конверсії. Метою реалізації є порівняння ефективності класичних підходів (рівномірної, останньої взаємодії) з більш складними методами, зокрема Shapley Values. Практична частина роботи ґрунтується на аналізі маркетингових даних, отриманих із цифрової кампанії, яка охоплює понад 10 тисяч взаємодій користувачів з рекламними повідомленнями через різні канали.

Побудована модель атрибуції дозволяє відрізнити канали, які мають високий вплив на здійснення конверсії, від тих, чий внесок є низьким або незначним. Це дає змогу маркетингологам визначити, яким саме каналам слід приділяти більше уваги при плануванні та реалізації рекламних кампаній, а які - переглянути або оптимізувати.

Щоб підвищити ефективність цифрових кампаній та забезпечити раціональний розподіл маркетингового бюджету, компанії прагнуть отримати відповіді на такі запитання:

- Які канали забезпечують найбільший внесок у досягнення конверсій і є ключовими для взаємодії з аудиторією?
- Які комбінації каналів працюють найкраще разом, створюючи синергію на шляху користувача до конверсії?

- Чи існують канали, чий вплив недооцінюється традиційними моделями (наприклад, останнім кліком), але які мають вагомий маржинальний ефект у багатоканальному середовищі?

Відповідно створена модель атрибуції, дозволить:

- Оцінити внесок кожного маркетингового каналу в досягнення конверсії.
- Справедливо розподілити цінність між усіма точками взаємодії в багатоканальному шляху користувача. У майбутніх проєктах сформувані кампанії, що базуються на перевірених взаємодіях.
- Надати базу для розробки аналітичних рекомендації щодо оптимізації маркетингового бюджету. Знизити витрати на неефективні канали та підвищити ROI загальної стратегії просування.

Для вирішення задачі маркетингової атрибуції в цифровому середовищі з використанням моделі Shapley Values запропоновано наступну методологію.

Першим етапом передбачається виконання попереднього аналізу даних. За допомогою бібліотек `pandas`, `numpy` та `matplotlib` здійснюється завантаження і первинний огляд структури набору даних (загальний огляд даних і EDA), необхідно визначити основні зміни та їх частоту, виявити пропуски і аномалії, здійснити первинну візуалізацію взаємодій між змінними. Особливий акцент приділяється змінним `user_id`, `date_served`, `marketing_channel` та `conversion`, які є критично важливими для побудови шляхів користувача. Формуються гіпотези про характер взаємодій між каналами й поведінку користувачів.

Далі здійснюється підготовка даних для моделювання. Цей етап включає агрегацію даних до рівня користувача з формуванням впорядкованих послідовностей каналів (`user journeys`), що дозволяє відобразити індивідуальні шляхи взаємодій із кампанією. Для кожного користувача формується набір каналів, з якими він взаємодіяв, а також фіксується факт здійснення конверсії. Створюється таблиця коаліцій, яка містить унікальні комбінації каналів та кількість відповідних конверсій, що забезпечує основу для подальших розрахунків.

Наступним кроком є побудова моделі атрибуції на основі значень Шеплі. Методологія передбачає генерацію повного набору можливих коаліцій каналів, обчислення маржинального внеску кожного каналу в усіх можливих комбінаціях, а також нормалізацію результатів для отримання відносних часток впливу. Отримані значення візуалізуються за допомогою графічних інструментів Python, що підвищує інтерпретованість результатів та спрощує порівняння внеску окремих каналів.

Для підвищення обґрунтованості аналізу додатково реалізуються моделі маркетингової атрибуції, що засновані на правилах: модель рівномірного розподілу (Linear Attribution) та модель останнього кліку (Last Touch Attribution). Це дозволяє зіставити результати алгоритмічного підходу з традиційними методами та виявити ключові розбіжності в оцінці внеску каналів, оцінити якість моделі через порівняння з базовими методами.

Завершальним етапом є порівняльний аналіз отриманих результатів, під час якого проводиться оцінка ефективності каналів у різних моделях атрибуції. Аналіз дозволяє виділити ключові канали, що демонструють стабільний вплив у різних підходах, а також виявлено канали, чий внесок суттєво зростає лише в рамках мультиканальних взаємодій.

Завершальний етап - формування висновків та практичних рекомендацій щодо оптимізації маркетингового бюджету, підвищення ефективності каналів і побудови багатоканальних кампаній на основі результатів атрибуції. Це дозволяє компаніям отримати об'єктивну й обґрунтовану картину впливу каналів і приймати стратегічно важливі рішення на базі даних.

Відповідно описаних кроків побудови моделей, перейдемо до практичної реалізації.

3.2. Характеристика обраного набору даних та попередня обробка для моделювання

У рамках побудови моделі атрибуції на основі значень Шеплі використано реальний набір даних, який моделює поведінку користувачів у відповідь на різні маркетингові комунікації. Даний датасет містить 10 037 записів та включає інформацію про

користувачів, яким було показано маркетингові повідомлення через різні канали протягом січня 2018 року (Додаток Б). Набір даних складається з характеристик таблиці 3.1.:

Таблиця 3.1.

Опис атрибутів набору даних

Назва характеристики	Опис	Тип
user_id	Унікальний ідентифікатор користувача	Категоріальний/дискретний
date_served	Дата показу маркетингового повідомлення	Часовий/дискретний
marketing_channel	Канал, через який було доставлено повідомлення (Email, Facebook тощо)	Категоріальний/дискретний
variant	Тип креативу, що був використаний (control або personalization)	Категоріальний/дискретний
converted	Факт здійснення конверсії після взаємодії з повідомленням	Категоріальний/бінарний
language_displayed	Мова, якою було показано рекламне повідомлення	Категоріальний/дискретний
language_preferred	Мова, яку користувач вказав як бажану	Категоріальний/дискретний
age_group	Вікова категорія користувача	Категоріальний/дискретний

date_subscribed	Дата підписки користувача на сервіс	Часовий/дискретний
date_canceled	Дата скасування підписки (якщо є)	Часовий/дискретний/пропущені дані
subscribing_channel	Канал, через який користувач підписався	Категоріальний/дискретний
is_retained	Чи залишився користувач активним після певного періоду	Категоріальний/бінарний

Джерело: складено автором

Для початку практичної частини дослідження було здійснено імпорт основних бібліотек Python для роботи з табличними даними (pandas, numpy) та візуалізацій (matplotlib, seaborn, plotly.express). Кожен рядок датасету представляє окрему взаємодію користувача з маркетинговим повідомленням через певний канал комунікації. У таблиці 3.2 наведено результати перевірки на наявність нульових значень у наборі даних.

Таблиця 3.2.

Перевірка пропущених значень у наборі даних

Назва змінної	Кількість пропущених значень
user_id	0
date_served	16
marketing_channel	15
variant	0
converted	15
language_displayed	0
language_preferred	0
age_group	0
date_subscribed	8 181

date_canceled	9 460
subscribing_channel	8 181
is_retained	8 181

Джерело: складено автором

При побудові моделі атрибуції не всі змінні несуть цінність. Критично важливі змінні для побудови атрибуційної моделі - `user_id`, `marketing_channel`, `date_served` та `converted` мають незначну кількість пропусків (до 15-16 записів), що становить менше 0,2% від загального обсягу даних. Це дозволяє застосувати стратегію прямого видалення відповідних рядків без суттєвого впливу на якість аналітичних висновків. Ідея важливості змінних полягає в тому, що, наприклад, завдяки `user_id`, можливо відслідковувати повну історію взаємодій конкретного користувача для побудови персоналізованого шляху до конверсії (`user journey`) [39].

Найбільша кількість пропущених значень спостерігається у змінних, пов'язаних із підпискою користувача: `date_subscribed`, `subscribing_channel`, `is_retained`, а також `date_canceled`, які не є критичними у випадку побудови моделі.

Зосередимось на тих змінних, що мають пряме відношення до шляху користувачів і результативності каналів, у наборі даних зменшено кількість змінних до `user_id`, `marketing_channel`, `date_served`, `converted`.

3.2.1. Дослідницький аналіз даних

Метою цього аналізу є заглиблення в інсайти, приховані в даних, перевірка розподілу даних та підготовка до подальшого моделювання. Фокусуємось на наступних дослідницьких питаннях:

1. Яка частка користувачів здійснила конверсію? Як виглядає дисбаланс класів між конверсіями (`converted = True`) та неуспішними випадками (`converted = False`)?
2. Які канали комунікації найчастіше використовувались у кампанії? Чи існує перевага певних каналів?

3. Які комбінації каналів найчастіше зустрічаються у шляху користувача до конверсії?

Розпочнемо з візуалізації розподілу цільової змінної - conversion:

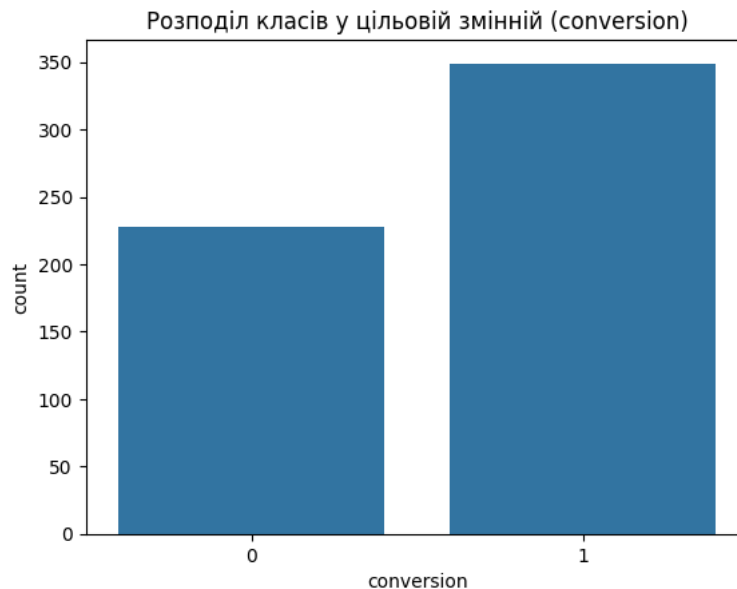


Рис.3.1. Розподіл класів у цільовій змінній

Джерело: Складено автором у середовищі Python

Відповідно Рис.3.1 клас 1 (успішна конверсія) домінує у вибірці - таких прикладів більше, ніж неуспішних, близько 60% усіх спостережень. Клас 0 (відсутність конверсії) становить меншу частину - 40% записів. У вибірці класовий дисбаланс є помірним. На відміну від багатьох реальних маркетингових кейсів, де конверсії - це 2-10% від загалу, тут обидва класи представлені достатньо для навчання моделей без застосування спеціальних методів корекції (наприклад, oversampling, SMOTE тощо).

Перейдемо до визначення найпопулярніших каналів охоплення. Рис. 3.2. відображає частоту показів кожного маркетингового каналу в межах вибірки. Враховує скільки разів користувачі взаємодіяли з кожним каналом протягом аналізованого періоду. Канал House Ads є найбільш використовуваним, з кількістю понад 250 показів, внутрішні рекламні інструменти компанії мали найвищу інтенсивність охоплення серед усіх. Далі за частотою йдуть Instagram, Facebook (105 показів), які також демонструють суттєву

присутність у маркетинговій комунікації, що є класичним для сучасних компаній.

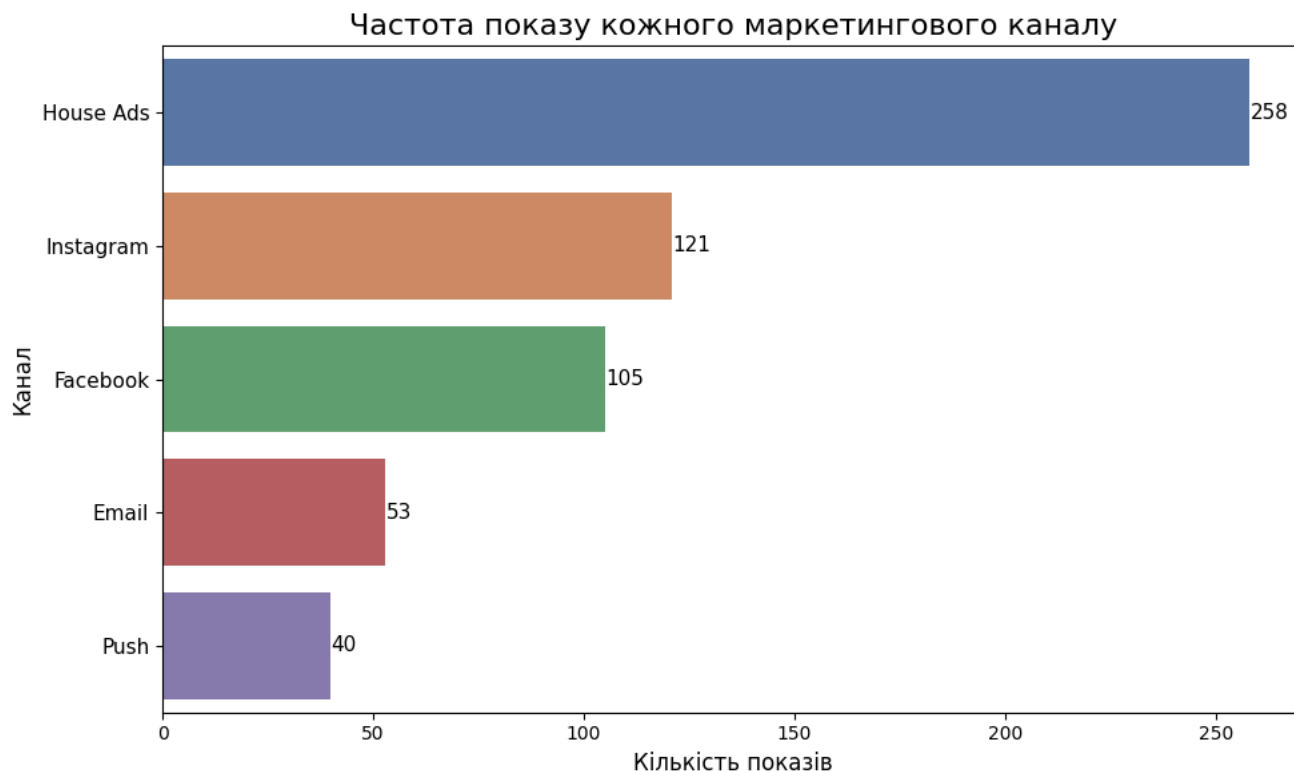


Рис.3.2. Частота показів кожного маркетингового каналу

Джерело: Складено автором у середовищі Python

Присутня нерівномірність розподілу маркетингових зусиль між каналами. Ця інформація є важливою в контексті атрибуційного аналізу: надмірна присутність одного каналу може штучно збільшувати його вплив на конверсії.

Дослідимо, чи існує залежність між частотою показів каналу та кількістю здійснених конверсій. Для цього побудовано графік розсіювання, де по осі X відображено кількість показів, а по осі Y - відповідну кількість конверсій для кожного з каналів. Така візуалізація дозволяє оцінити, наскільки частота використання каналу корелює з його результативністю. З'ясуємо, чи збільшення кількості показів певного каналу автоматично веде до зростання конверсій, або ж існують канали, які, навіть за меншої кількості показів, демонструють високий рівень ефективності.

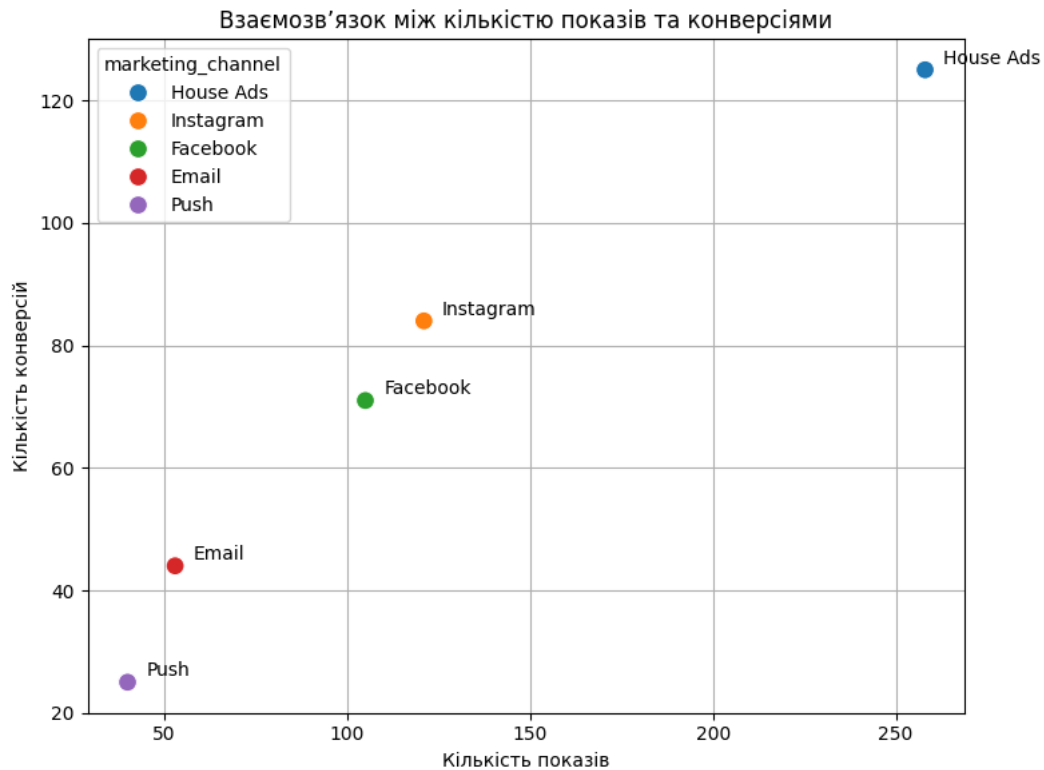


Рис.3.3. Взаємозв'язок між кількістю показів та конверсіями

Джерело: Складено автором у середовищі Python

На основі Рис.3.3. спостерігаємо, що канал House Ads має як найвищу кількість показів, так і найбільшу кількість конверсій, займає провідну роль у кампанії. Водночас Instagram та Facebook займають проміжні позиції - вони менш охоплюючі, але також демонструють високий абсолютний рівень конверсій. Email і Push мають обмежену кількість показів і порівняно меншу кількість успішних взаємодій.

Початковий візуальний аналіз демонструє наявність прямого лінійного зв'язку. Тобто, зростання кількості взаємодій на пряму пов'язане зі зростанням абсолютної кількості конверсій. Перевіримо коефіцієнт кореляції Пірсона:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Де:

- x_i — значення кількості показів для каналу i ;
- y_i — значення кількості конверсій для каналу i ;

- \bar{x} — середнє значення по стовпцю impressions;
- y — середнє значення по стовпцю conversions;
- r — коефіцієнт кореляції Пірсона.

У нашому випадку:

$$r = 0,98$$

Це свідчить про наявність сильного позитивного лінійного зв'язку між цими показниками. Однак, така залежність не дає повної відповіді на питання про ефективність каналів, оскільки не враховує їхню відносну результативність. Тому подальший аналіз буде зосереджений на вивченні конверсійної частки (conversion rate), яка дозволить об'єктивно оцінити внесок кожного каналу незалежно від його охоплення.

У традиційних підходах (моделях останнього або першого кліку) домінування каналу з великою кількістю показів дійсно може призвести до перекосів: йому буде приписано надмірну кількість конверсій лише через часту присутність у ланцюгах взаємодії. У таких випадках модель не оцінить внесок каналу по суті, а лише зафіксує його частоту.

Однак, у просунутих атрибуційних моделях, на основі значень Шеплі, цей ефект нейтралізується. Shapley Value оцінить його маржинальний внесок у кожній можливій комбінації каналів.

Продовжуючи аналіз ефективності каналів, наступним етапом стало вивчення взаємозв'язків між каналами, які спільно з'являються у взаємодіях одного користувача. Адже частота показів та рівень конверсій, навіть у поєднанні з оцінкою конверсійної частки, не дозволяють повною мірою врахувати мультиканальну природу цифрових кампаній. Важливо виявити, які канали найчастіше поєднуються між собою.

Для цього було побудовано матрицю співпадінь каналів, яка відображає, скільки користувачів бачили одночасно (або послідовно) два різні канали. Результати представлено у вигляді теплової карти (heatmap), де інтенсивність кольору ілюструє кількість таких співпадінь. По осі X та Y розміщено назви каналів, а кожна клітинка вказує, скільки разів відповідна пара каналів зустрічалась у рамках одного користувача.

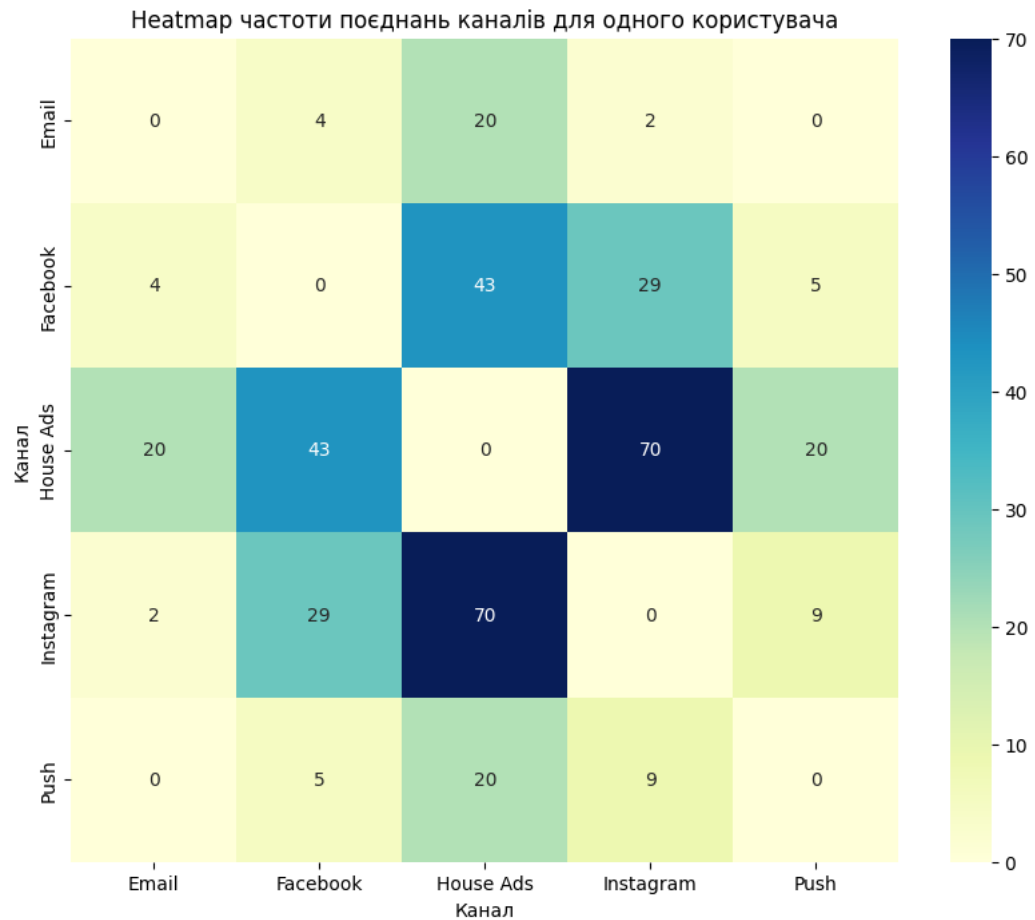


Рис.3.4. Heatmap частоти поєднань каналів для одного користувача

Джерело: Складено автором у середовищі Python

Відповідно Рис. 3.4. House Ads є найбільш інтегрованим каналом: він найчастіше поєднується з Instagram (70 випадків), Facebook (43) і Push (20). Це свідчить про його центральну роль, як було й попередньо досліджено. Instagram активно поєднується з Facebook (29), формуючи стабільний кластер взаємодій Meta. Натомість Email і Push рідше з'являються у комбінаціях, їх здебільшого використовують в рамках окремих, вузьких сценаріїв або як додаткові канали.

Отже, дані свідчать про помірний дисбаланс класів: близько 60% взаємодій завершилися конверсією, що є нетиповим для більшості реальних бізнес-кейсів і дає змогу уникнути додаткових процедур балансування вибірки при моделюванні.

У результаті аналізу частоти каналів було виявлено чітке домінування House Ads, який має найбільшу кількість як показів, так і конверсій. Instagram і Facebook посідають

друге й третє місце відповідно, демонструючи значну присутність та стабільну результативність. Email і Push мають менше охоплення, однак Email показав доволі високу частку конверсій, він ефективний при цільовому застосуванні.

Аналіз каналів у поєднаннях показав, що найбільш частими є комбінації House Ads - Instagram, House Ads - Facebook, а також Instagram - Facebook. Це бізнес-розуміння допоможе при оцінці вихідних результатів атрибуційної моделі Shapley Value і моделей маркетингової атрибуції, що засновані на правилах.

3.3. Побудова моделі атрибуції на основі Shapley Values

Перейдемо до побудови моделі маркетингової атрибуції на основі значень Шеплі (Shapley Values). Модель має визначити маржинальний внесок кожного маркетингового каналу у досягнення конверсії, враховуючи всі можливі комбінації його появи у користувацьких шляхах. (Додаток А)

3.3.1. Побудова впорядкованих шляхів та формування таблиці коаліцій

Перед розрахунком значень Шеплі, дані було агреговано на рівень користувача: Для кожного user_id з'єднуються унікальні значення marketing_channel, які зустрічалися в послідовності показів. Функція унікальності гарантує, що кожен канал у шляху користувача буде врахований лише один раз, але в тому порядку, в якому він з'являвся. Результатом є рядок на зразок 'Email,Push,Facebook', який представляє шлях користувача. Також для кожного користувача було зафіксовано, чи завершився його шлях конверсією.

Наступним етапом стало групування користувачів за унікальними комбінаціями каналів (channel combinations), що відображають усі можливі коаліції каналів, які зустрічаються в реальних шляхах. Для кожної з таких коаліцій було обраховано кількість конверсій, здійснених користувачами, які пройшли цей конкретний шлях. В результаті сформовано агреговану таблицю коаліцій (Рис.3.5.), де кожен запис представляє певну унікальну комбінацію каналів разом з відповідним числом успішних конверсій.

	channel_combo	total_conversions
11	House Ads	118
18	Instagram,House Ads	34
5	Facebook	34
0	Email	29
6	Facebook,House Ads	22
15	Instagram	22
3	Email,House Ads	11
16	Instagram,Facebook	10
21	Push	10
23	Push,House Ads	10
7	Facebook,Instagram	9
17	Instagram,Facebook,House Ads	6
19	Instagram,Push	6
13	House Ads,Instagram	3
14	House Ads,Push	2
8	Facebook,Instagram,House Ads	2
1	Email,Facebook	2
24	Push,Instagram	2
22	Push,Facebook	2
12	House Ads,Facebook	2

Рис.3.5. Агреговані шляхи користувачів

Джерело: Складено автором у середовищі Python

3.3.2. Генерація всіх можливих коаліцій каналів

Для переходу до розрахунку значень Шеплі було сформовано повний набір можливих коаліцій каналів, які теоретично могли б виникати в межах взаємодії користувача. Це означає генерацію всіх можливих непорожніх підмножин із доступного списку каналів, включно з окремими каналами, парами, трійками тощо. Такі підмножини розглядаються як коаліції, а кожна з них може має свою «вартість» - тобто кількість конверсій, що виникли за участі саме цієї комбінації каналів. [40]

Для цього етапу реалізовано функцію, яка за заданим списком каналів генерує всі можливі комбінації (підмножини) довжиною від 1 до n , де n - кількість унікальних каналів у вибірці. Ці комбінації відображають всі варіанти взаємодій, у яких канали можуть виступати спільно.

Додаткова функція перетворює кожен підмножину у відсортований текстовий рядок, наприклад 'Email,Facebook', уніфіковано зберігаємо значення в подальших

словниках Сформовано аналітичний простір усіх можливих коаліцій - базис для обчислення маржинального внеску кожного каналу.

Наступним етапом було задано функцію $v(S)$, яка обчислює загальну "цінність", результативність певної коаліції – конверсійну силу. Для кожної коаліції (наприклад, ['Email', 'Facebook']) шукає всі її підмножини (['Email'], ['Facebook'], ['Email,Facebook']) і сумує їхні значення з `conversion_dict`.

$$v(S) = \sum_{T \subseteq S} \text{conversion}(T)$$

Де:

- S - коаліція каналів,
- T підмножина S , яка присутня в конверсійному словнику.

Наприклад,

$$v(['Email', 'Push']) = \text{conversion}('Email') + \text{conversion}('Push') + \text{conversion}('Email,Push')$$

Далі було реалізовано розрахунок значень Шеплі для кожного каналу i за класичною формулою:

$$\phi_i(N, v) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|! (|N| - |S| - 1)! \cdot [v(S \cup \{i\}) - v(S)]$$

Розрахунок проводиться у вигляді окремої функції, яка приймає на вхід агрегований датафрейм із комбінаціями каналів та відповідною кількістю конверсій для кожної з них. Із цього датафрейму формується словник, у якому ключами є рядки з назвами каналів (наприклад, "Email,Push"), а значеннями - кількість конверсій для відповідної коаліції. Це дозволяє швидко звертатися до потрібного значення при обчисленнях.

Далі визначаються усі канали, які зустрічаються у сольних комбінаціях, тобто в таких шляхах, де користувачі бачили лише один канал.

Наступним етапом є обчислення функції вартості коаліції, яку попередньо було задано, для всіх можливих підмножин каналів. Тобто, якщо коаліція складається з трьох каналів, її вартість буде визначена як сума конверсій усіх одиничних, парних та потрійних комбінацій, присутніх у наборі даних.

3.3.3. Розрахунок значень Шеплі

Для кожного каналу перебираються всі коаліції, у які він ще не входить, і оцінюється його маржинальний внесок у кожен з них. Маржинальний внесок визначається як різниця між вартістю коаліції після додавання даного каналу та її початковою вартістю:

$$\text{contrib}_i(S) = v(S \cup \{i\}) - v(S)$$

Цей внесок зважується за допомогою комбінаційного коефіцієнта, який враховує кількість можливих порядків входження каналу в коаліцію. Формула вагу у загальній формулі Шеплі:

$$\text{weight}_i(S) = \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!}$$

У результаті кожен канал отримує значення Шеплі, що інтерпретується як його середній маржинальний внесок у досягнення конверсії, з урахуванням усіх можливих коаліцій та порядків взаємодії.

3.3.4. Нормалізація результатів та візуалізація внесків каналів

Отримані значення Шеплі були візуалізовані на графіку. Вони демонструють середній маржинальний внесок кожного з каналів у досягнення конверсій.

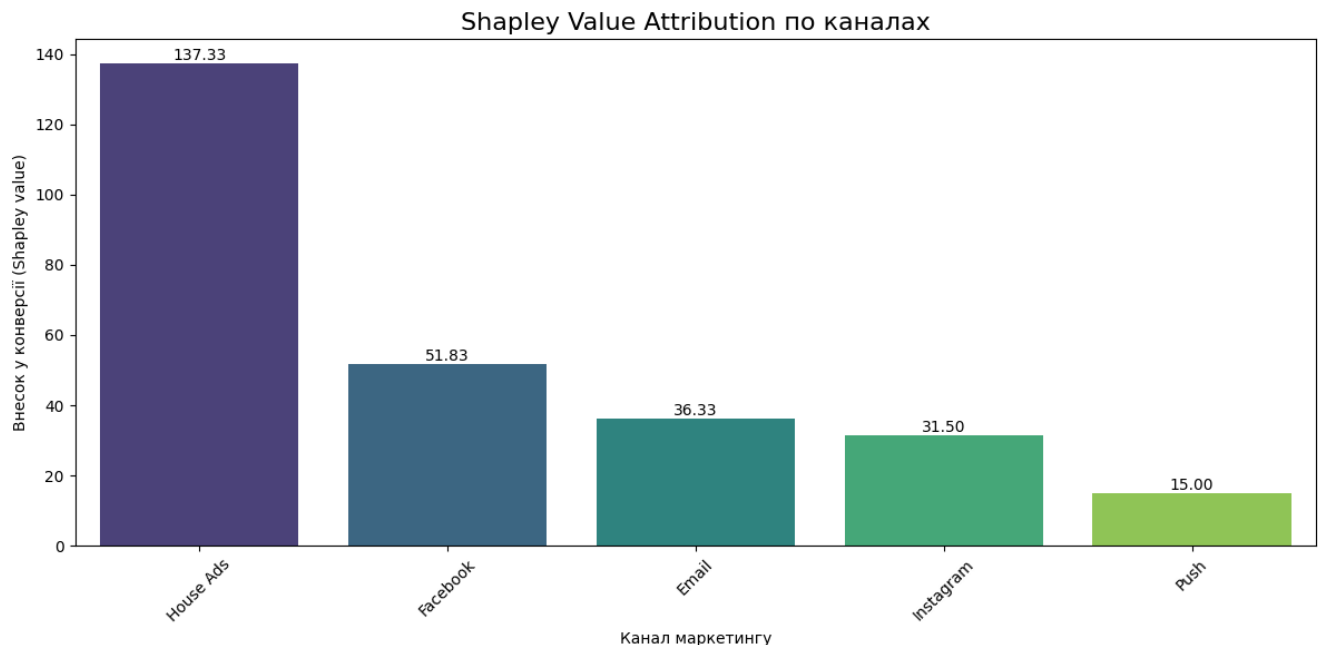


Рис.3.6. Shapley Value Attribution по каналах

Джерело: Складено автором у середовищі Python

Найвище значення має канал House Ads із внеском 137,33, який свідчить про його критичну роль у шляху користувача до цільової дії. Цей результат підтверджує не лише високу частоту його використання, але й значущість у поєднаннях з іншими каналами - House Ads забезпечує найбільший приріст цінності коаліцій, у які входить.

Другим за впливом є Facebook, його внесок - 51,83, це стабільний ефект каналу як у сольних комбінаціях, так і в мультиканальних шляхах. Email та Instagram продемонстрували схожий рівень впливу - 36,33 та 31, 50 відповідно. Їхня участь у конверсійних ланцюгах помірна, але системна. Найменший внесок має Push - 15,00 значення корелює з його менш активним використанням, і обмеженим маржинальним ефектом у контексті інших каналів.

Задля зручнішої оцінки результатів проведемо нормалізацію значень - переведення абсолютних внесків каналів у відносні частки. Процес полягає в діленні внеску кожного каналу на суму всіх внесків, що дає частку впливу каналу в загальному обсязі конверсій.



Рис.3.7. Нормалізовані внески по каналах на основі Shapley Values Attribution

Джерело: Складено автором у середовищі Python

На основі Рис. 3.7., отримуємо наступні спостереження:

House Ads відповідає за 50% загального впливу. Канал є функціонально ефективним, але при цьому є потенційна залежність загальної конверсійної стратегії від цього внутрішнього каналу. Це створює певний ризик надмірної концентрації, і водночас відкриває можливість масштабування успішного патерну на зовнішні або нові канали.

Канал Facebook з часткою 19.06%, може служити зв'язуючою ланкою між контактами користувача з іншими каналами. Це сигнал до подальших інвестицій у Facebook як у канал із потенціалом зростання за рахунок посилення взаємодії.

Email і Instagram мають порівняно схожі показники - 13,36% та 11,58% відповідно. Push-сповіщення становлять лише 5,51%. Такий канал комунікації має потенціал точкового впливу, наприклад, для реактивації «сплячих» користувачів або в рамках промо-кампаній з обмеженим терміном дії.

Щоб оцінити ефективність побудованої моделі і коректність розрахованих часток, порівнюємо її результати з класичними моделями, що засновані на правилах.

3.4. Побудова моделей маркетингової атрибуції, що засновані на правилах та їхня оцінка

3.4.1. Linear Attribution

Реалізуємо логіку лінійної атрибуції (Linear Attribution), де внесок у конверсію рівномірно розподіляється між усіма каналами, які були присутні в шляху користувача. Кожному каналу присвоюється частка внеску, яка розраховується як:

$$Value = \frac{1}{n}$$

Де:

n – кількість каналів у шляху

Модель не враховує порядок впливу каналів, а лише їхню присутність. Надає рівний внесок кожному з каналів, що брали участь у шляху до конверсії.

Нормалізувавши внесок кожного каналу у конверсію маємо, відповідно Рис. 3.8. Найвищий відсотковий внесок, як і в моделі Shapley Values отримав канал House Ads - 47,86%, на другому місці опинився Facebook (18,08%). Instagram - 16,81% показав. Email

має відсотковий внесок у 10,62%. Найменша частка залишається у Push сповіщень - 6,63%.



Рис.3.8. Нормалізовані внески по каналах на основі Linear Attribution

Джерело: Складено автором у середовищі Python

Побудуємо таблицю Touchpoints, щоб визначити на якому етапі шляху користувача кожен канал має найбільший вплив. Touchpoint вважаємо позиція каналу. Наприклад, якщо користувач бачив послідовність каналів: Email → Facebook → House Ads, то:

- Email - Touchpoint 1
- Facebook - Touchpoint 2
- House Ads - Touchpoint 3

Максимальна можлива кількість Touchpoint у наборі даних – 3. Згідно з Рис. 3.9. House Ads домінує у шляху користувача, зустрічається на всіх touchpoints, включно з останнім. Facebook і Instagram – це канали початку і середини шляху, зустрічаються переважно на 1-2 touchpoint. Це означає, що вони працюють як канали обмірковування, залучення або підтримки. Їхнє значення зростає в комбінації з іншими, а не як фінальний крок. Тому їхній внесок у кінцеву конверсію може бути недооцінений класичними моделями типу Last Touch.

Touchpoint Channel	Touchpoint 1	Touchpoint 2	Touchpoint 3	Total
House Ads	44.07	61.44	100.0	47.86
Facebook	18.68	16.49	0.0	18.08
Instagram	17.90	13.03	0.0	16.81
Email	13.18	0.00	0.0	10.62
Push	6.17	9.04	0.0	6.63

Рис.3.9. Значення Linear Attribution для всіх каналів у точках дотику 1-3.

Джерело: Складено автором у середовищі Python

Email - переважно перший touchpoint. Він ймовірно, тригерить початковий інтерес (наприклад, через розсилку або вітання) і діє як самостійний канал конверсії, де покупка відбувається в один етап.

Push - канал нагадування. 6,17% на Touchpoint 1, 9,04% на Touchpoint 2 - працює як допоміжний елемент. Може бути активований через мобільний застосунок, і має обмежений конверсійний вплив.

3.4.2. Last Touch Attribution

Модель Last Touch Attribution, працює за принципом які канали найчастіше завершують користувацький шлях перед конверсією. Модель не враховує жодної попередньої взаємодії, тому добре підходить лише для сценаріїв, де саме фінальний клік-взаємодія має вирішальне значення. Згідно з Рис. 3.10. - очевидним лідером є канал House Ads, який отримав 60,23% усіх конверсій, якщо оцінювати виключно за останньою взаємодією. Відповідно Рис. 19 - це і є єдиний канал, який присутній на Touchpoint 3, і має найбільшу частку на на Touchpoint 2. Внутрішні рекламні повідомлення часто завершують комунікаційний ланцюжок і фактично «закривають» угоду.

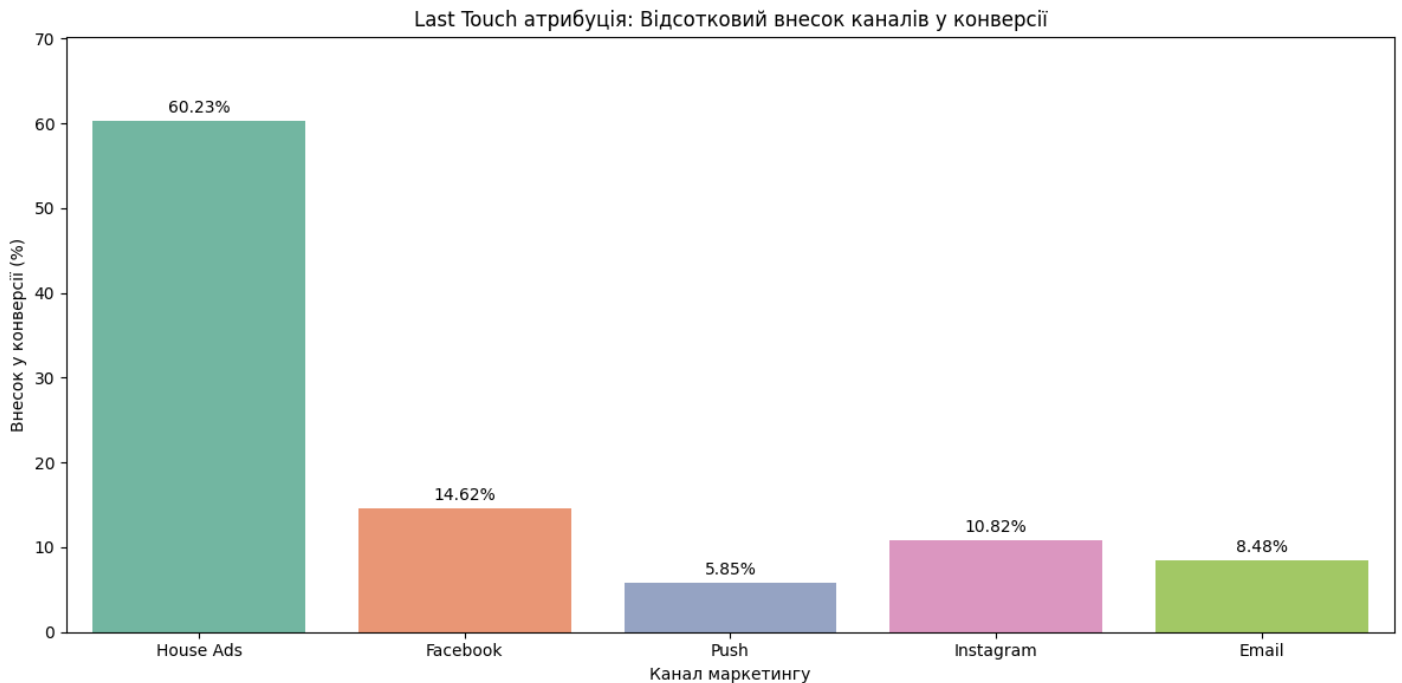


Рис.3.10. Нормалізовані внески по каналах на основі Last Touch Attribution

Джерело: Складено автором у середовищі Python

Facebook (14,62%), Instagram (10,82%) та Email (8,48%) відіграють менш домінуючу роль. Цікавим є спостереження, що Email взагалі не присутній на Touchpoint 2, виступає тільки на першому місці. Тому конверсії атрибутовані за Last Touch – працюють виключно в соло на Email.

3.5. Порівняння результатів моделювання

У межах порівняльного аналізу результатів трьох моделей - Shapley Values, Linear Attribution та Last Touch Attribution - виявлено відмінності у розподілі внеску між каналами, що дозволяє зробити низку висновків щодо поведінки користувачів та ефективності маркетингових зусиль, а також природі функціонування моделі. Результати порівняння внеску відображені на Рис. 3.11

	Channel	Shapley %	Linear %	Last Touch %
0	House Ads	50.49	47.86	60.23
1	Facebook	19.06	18.08	14.62
2	Email	13.36	10.62	8.48
3	Instagram	11.58	16.81	10.82
4	Push	5.51	6.63	5.85

Рис.3.11. – Порівняльна таблиця внесків по каналах на основі Shapley Values, Linear Attribution та Last Touch Attribution

Джерело: Складено автором

Основні відмінності в результатах можна пояснити таким чином:

1. Last Touch Attribution максимально концентрується на завершальному контакті, приписуючи весь внесок останньому каналу в шляху до конверсії. Саме тому House Ads отримує тут найвищу оцінку (60.23 %), адже за результатами heatmap-поєднань та таблиці touchpoints ми бачили, що House Ads часто фігурує саме на фінальних етапах. Це створює ефект переоцінки, оскільки ігноруються попередні етапи взаємодії.
2. Linear Attribution ділить внесок рівномірно між усіма каналами, присутніми в шляху користувача. Тому цей підхід зменшує вагу домінантного фінального каналу, натомість підвищує оцінку тих, які регулярно з'являються поряд з іншими, як Instagram або Email. Вони тут демонструють дещо вищий внесок, ніж у Last Touch, бо враховується їх підтримуюча роль у мультиканальних ланцюгах.
3. Shapley Values дає найбільш структурну й математично обґрунтовану оцінку: вона розподіляє внесок кожного каналу відповідно до його маржинального ефекту в усіх можливих коаліціях каналів.

Shapley Values Attribution дозволила:

- виявити канали, чий внесок зростає завдяки комбінаціям (наприклад, Email, який отримав 13,36 % у Shapley проти 8,48 % у Last Touch);
- уникнути ефекту “останнього дотику”, завдяки чому House Ads отримує реалістичнішу, хоч і досі високу оцінку (50,49 %);
- виключити канали, що часто фігурують, але не впливають критично на результат, як-от Push, чий внесок є стабільно низьким в усіх підходах.

У підсумку, розбіжності у внесках виникають через різні припущення моделей про причинно-наслідковий зв'язок між каналом і конверсією. Last Touch - спрощує його до останньої дії, Linear - розподіляє без урахування впливу, а Shapley - оцінює реальний внесок з урахуванням контексту, порядку і взаємодії каналів. Саме тому результати відрізняються, і саме тому Shapley краще пристосована для оптимізації бюджетів у складному мультиканальному середовищі.

Висновки та практичні рекомендації на основі практичної реалізації моделей атрибуції

Реалізована модель маркетингової атрибуції на основі значень Шеплі дозволила глибоко оцінити реальний внесок кожного каналу в досягнення конверсії з урахуванням усіх можливих коаліцій каналів у багатоканальному середовищі. Відповідно маркетологи можуть отримати цінну інформацію про ефективність своїх маркетингових зусиль і планування бюджету.

3.6. Ключові висновки і рекомендації для бізнесу

House Ads виступає найбільшим драйвером конверсій. Канал отримав найбільшу оцінку маржинального внеску, що підтверджує його вагому роль у шляху до конверсії не лише як фінального дотику, а й як частини ефективної взаємодії на всьому.

Необхідно продовжувати інвестувати в внутрішню рекламу (in-app рекламу, банери, рекомендаційні блоки в інтерфейсі). Її висока ефективність пояснюється точним таргетингом, нативністю, а також тим, що вона з'являється на пізніх етапах користувацького шляху.

Дії, які дозволять ще більше розкрити потенціал каналу:

- Посилити персоналізацію внутрішньої реклами: адаптувати креативи залежно від минулих взаємодій користувача;
- Використовувати House Ads як ретаргетинговий інструмент для користувачів, які вже взаємодіяли з іншими каналами (наприклад, Email або Instagram);
- Розробити серії сценаріїв на основі поведінки (наприклад, «House Ads only after email open»).

Facebook (19,06 %) - ефективний канал підтримки та залучення. Відіграє роль як у верхній, так і в середній частині воронки, формуючи первинну зацікавленість та підтримуючи динаміку переходів. Його внесок у коаліціях значний, особливо в поєднанні з House Ads.

Рекомендації по оптимізації каналу:

- Запускати рекламні кампанії з деталізованим аудиторним таргетингом (Lookalike Audiences на основі конверсійної поведінки); [41]
- Використовувати Facebook як канал «нагадування», особливо в серіях взаємодій, де він є не першим дотиком;
- Тестувати відеоформати й інтерактивні пости для підвищення глибини взаємодії перед появою House Ads.

Email (13,36 %) – стабільний канал, проте потребує системного розширення взаємодій. Не працює самостійно як фінальний тригер, але суттєво підсилює ефект комбінацій. Внесок у Шеплі-моделі вищий, ніж у Last Touch, . Необхідно доопрацювати інтеграцію Email-маркетингу на фінальних етапах шляху користувача. Адже функція CRM, де Email є основним інструментом, догнати аудиторію, яка не зробила конверсійну дію на початковому етапі.

Рекомендовані дії:

- Визначити сегменти аудиторії з високою ймовірністю відповіді на email (наприклад за даними Retention);
- Застосовувати тригерні розсилки: покинуті кошики, повторне залучення, онбординг;

- Провести A/B тести з темами листів, часом відправлення, форматами СТА (call-to-action).

Instagram (11,58 %) мав вищі оцінки в класичних моделях, але нижчу в Шеплі. Характер каналу вказує на його видимість, але відносно слабкий маржинальний ефект, свідчить про обмежений вплив на остаточне прийняття рішення і показ в поєднанні з більш впливовими каналами. Він здебільшого формує емоційний контакт і підвищує залучення.

Практичні рекомендації:

- Використовувати Instagram переважно на ранніх етапах воронки (awareness), з креативами, що формують брендове сприйняття;
- Інтегрувати його з House Ads або Facebook для підтримки багатоканальної комунікації;
- Проводити кампейни з посиланнями на LP (landing page), а не просто на профіль бренду, щоб збільшити ймовірність фінальної конверсії;

Push-повідомлення (5,51 %) - канал з низьким впливом, який потребує перезапуску. Кагал має найменший внесок за результатами всіх моделями. Їхній вплив недостатній і неправильно структурований. Канал з найменшою кількістю показів, відповідно для проведення тесту є потреба в збільшенні трафіку, щоб визначити чи є канал ефективним . Є сенс змінити тактику або переглянути аудиторії, що їх отримують.

Рекомендації по оптимізації каналу:

- Перейти до більш персоналізованих push-повідомлень (наприклад, із згадкою попередніх дій користувача);
- Обмежити частоту й додати A/B тестування: чи варто надсилати push до чи після email або інших каналів;
- Вимірювати час відкриття push і його вплив на повернення користувачів до взаємодії.

Отже, враховуючи гнучкість і точність моделі Шеплі, компанія може використати її для оцінки розрахунку маржинального ефекту кожного дотику. Такий підхід дозволить зменшити залежність від традиційних моделей, які системно переоцінюють фінальні етапи взаємодії користувача з каналами. Водночас він сприятиме ефективнішому розподілу маркетингового бюджету, оскільки канали з незначним впливом більше не отримуватимуть непропорційного фінансування. На основі фактичного впливу каналів, а не лише їхньої частоти чи охоплення, можна буде точніше планувати багатоканальні кампанії. У результаті компанія формуватиме дійсно чітку маркетингову стратегію, де кожен канал працює як частина злагодженої системи, орієнтованої на ефективну взаємодію з користувачем.

ВИСНОВКИ ДО РОЗДІЛУ 3

У третьому розділі було проведено практичну реалізацію моделі атрибуції з використанням значень Шеплі на основі даних цифрової маркетингової кампанії. Детально розглянуто етапи підготовки даних, включаючи побудову шляхів користувача та формування коаліцій каналів. Реалізована модель Шеплі дала змогу математично обґрунтовано розподілити внесок кожного каналу у конверсіях, що дозволило виявити канали з недооціненим маржинальним ефектом. У межах порівняння було побудовано також моделі маркетингової атрибуції, що засновані на правилах: рівномірного розподілу (Linear Attribution) та останнього кліку (Last Touch), результати яких продемонстрували значні відмінності у розподілі внесків. Проведений аналіз підтвердив, що модель Шеплі є більш справедливою та інформативною в багатоканальному середовищі, оскільки враховує всі можливі варіанти взаємодії каналів. Отримані результати лягли в основу практичних рекомендацій для оптимізації маркетингової стратегії компанії.

ВИСНОВКИ

У рамках кваліфікаційної роботи було комплексно досліджено проблему маркетингової атрибуції в умовах сучасного цифрового середовища, а також запропоновано ефективні підходи до моделювання впливу каналів з використанням методів машинного навчання.

По-перше, узагальнено теоретичні основи маркетингової атрибуції та проведено класифікацію її основних моделей. Встановлено, що традиційні моделі, що засновані на правилах — такі як моделі першого кліку, останнього кліку, рівномірного розподілу або позиційна (U-shaped) — хоча й залишаються поширеними через свою простоту, мають суттєві обмеження в умовах складних багатоканальних взаємодій. Зокрема, вони не враховують послідовність, контекст і взаємозалежності між точками контакту, що призводить до викривленого уявлення про реальний внесок каналів у досягнення конверсій.

По-друге, проаналізовано алгоритми машинного навчання, які застосовуються в моделюванні атрибуції. Особливу увагу було приділено ланцюгам Маркова як стохастичному методу оцінки впливу каналів, а також моделі на основі значень Шеплі, що походить з теорії кооперативних ігор. Обидва підходи продемонстрували здатність глибше враховувати взаємодію каналів, контекст шляхів користувача та мультиканальні ефекти, чим значно перевершують моделі маркетингової атрибуції, що засновані на правилах в умовах цифрового маркетингу.

Третім кроком стало практичне впровадження моделі Shapley Values Attribution у середовищі Python на основі реального набору даних. Було виконано побудову шляхів користувачів, формування коаліцій каналів, обчислення граничного внеску кожного каналу у всіх можливих комбінаціях та візуалізацію результатів. Реалізована модель дозволила виявити канали з найбільшим маржинальним ефектом у досягненні конверсій та виключити переоціненість каналів із великою частотою, але незначною реальним внеском.

Завдяки порівняльному аналізу результатів моделей, що засновані на правилах (Last Click, Linear Attribution) із алгоритмічним підходом на основі значень Шеплі було встановлено, що моделі, що засновані на правилах мають тенденцію або до надмірної концентрації (наприклад, Last Click), або до спрощення впливу (наприклад, Linear Attribution). Модель Шеплі, навпаки, дозволила об'єктивно оцінити внесок каналів у різних комбінаціях, підтвердило її вищу точність, гнучкість і прогностичну силу цього підходу для маркетингової аналітики.

Було сформульовано практичні рекомендації для бізнесу на основі отриманих результатів. Зокрема, запропоновано змістовну інтерпретацію внесків каналів, виявлено ролі кожного з них у шляху користувача, визначено ті канали, які є ключовими на різних етапах воронки та виявлено точки для оптимізації рекламного бюджету. Результати дослідження можуть бути використані в аналітичних відділах компаній для планування багатоканальних кампаній, коригування креативних стратегій та підвищення ефективності взаємодії з цільовою аудиторією.

Отже, результати дослідження не лише підтвердили переваги алгоритмічних моделей атрибуції, а й продемонстрували, як точна аналітика може стати стратегічною перевагою для бізнесу у світі, де кожен дотик до клієнта має значення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Statista. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028. URL: <https://www.statista.com/statistics/871513/worldwide-data-created/>
2. Farenjuk Y., Chornous G. Principles of effective planning of advertising activity in the 21st century // Economy of Ukraine. – 2023. – No. 2. – P. 55-72.. URL: <https://nasu-periodicals.org.ua/index.php/economyukr/article/view/2023-02-3>
3. Forbes. 18 Big Challenges Marketers And Advertisers Face This Year. URL: <https://www.forbes.com/councils/forbesagencycouncil/2024/05/23/18-big-challenges-marketers-and-advertisers-face-this-year/>
4. Dataiku, Inc Marketing Attribution High-Value Marketing AI Use Cases . URL: <https://pages.dataiku.com/hubfs/Marketing-AI-Attribution.pdf>
5. Forbes. Navigating The Future Of Attribution: Beyond Third-Party Cookies. URL: <https://www.forbes.com/councils/forbesbusinesscouncil/2024/05/13/navigating-the-future-of-attribution-beyond-third-party-cookies/>
6. Developer Mozilla. Third-party cookies. URL: https://developer.mozilla.org/en-US/docs/Web/Privacy/Guides/Third-party_cookies
7. Google ending third-party cookies in Chrome. URL: <https://www.cookiebot.com/en/google-third-party-cookies/>
8. Statista. Change in digital advertising spending worldwide from 2022 to 2028. URL: <https://www.statista.com/statistics/325726/worldwide-digital-ad-spend-growth>
9. eSputnik. ROI: визначення в маркетингу. URL: <https://esputnik.com/uk/slovyk-email-marketologa/roi>
10. Cases Media. Невидимий шлях до успіху: оволодіння Customer journey map. URL: <https://cases.media/article/nevidimii-shlyakh-do-uspikhu-ovolodinnya-customer-journey->

[map?srsId=AfmBOopkxn61DzTKydniKEehEc%20noiWZZoXIN16VgGM4pnCcJFcQ8hMf](https://er.knutd.edu.ua/bitstream/123456789/14560/3/Sapian_A_Vyshnevskaja_M.pdf)

11. Sapian A. Vyshnevskaja M., 2019. Scientific journal «ΛΟΓΟΣ. The art of scientific mind» . The marketing funnel as an effective way of the business strategy. URL: https://er.knutd.edu.ua/bitstream/123456789/14560/3/Sapian_A_Vyshnevskaja_M.pdf
12. Kyivstar Business Hub. Як працювати з маркетинговою воронкою продажів: етапи, правила створення. URL: <https://hub.kyivstar.ua/articles/shho-take-voronka-prodazhiv-ta-yak-yiyi-pravylnno-stvoryty>
13. Amazon Ads. Marketing attribution guide. URL: <https://advertising.amazon.com/library/guides/marketing-attribution>
14. Nogood. Using Responsive Display Ads. URL: <https://nogood.io/2023/12/15/why-should-an-advertiser-consider-using-responsive-display-ads/>
15. DOMO. Understanding Marketing Attribution: A Step-by-Step Guide. URL: <https://www.domo.com/blog/understanding-marketing-attribution-a-step-by-step-guide/>
16. LinkedIn. How to Store UTM Parameters in Cookies for Better Attribution, Personalization, and Analytics. URL: <https://www.linkedin.com/pulse/how-store-utm-parameters-cookies-better-attribution-analytics-malik-rmjnc>
17. Intuit Mailchimp. Pixel Tracking vs Cookies: Key Differences Explained. URL: <https://mailchimp.com/resources/pixel-tracking/>
18. Mapflo. How Tracking works: Pixels, Cookies, UTM Parameters, gclid, wbruid & fclid. URL: <https://www.mapflo.io/data-set-up-and-tracking/tracking-code-pixels-cookies-utm-parameters/>
19. Leadsrx. Marketing attribution in the age of data privacy. URL: <https://leadsrx.com/resources/blog/marketing-attribution-in-the-age-of-data-privacy/>
20. Rebid. Data fragmentation in marketing. URL: <https://www.rebid.co/data-fragmentation-in-marketing/>

21. Banzai. Top Marketing Attribution Models and How to Choose Which To Use. URL: <https://www.banzai.io/blog/marketing-attribution-models>
22. Stape. What is a marketing attribution model? URL: <https://stape.io/blog/marketing-attribution-models#what-is-a-marketing-attribution-model>
23. NoGood. Marketing attribution. URL: <https://nogood.io/2024/08/07/marketing-attribution/>
24. AppsFlyer. Data-driven attribution (glossary). URL: <https://www.appsflyer.com/glossary/data-driven-attribution/>
25. Google Cloud. What is machine learning. URL: <https://cloud.google.com/learn/what-is-machine-learning>
26. IBM. Machine learning types. URL: <https://www.ibm.com/think/topics/machine-learning-types>
27. Choreograph. Uncovering \$12.9 M in incremental sales with multi-model attribution & causal impact studies. URL: <https://www.choreograph.com/resource/uncovering-12-9m-in-incremental-sales-with-multi-model-attribution-causal-impact-studies>
28. TripleWhale. Markov chain attribution. URL: <https://www.triplewhale.com/blog/markov-chain-attribution>
29. Adequate Digital. Markov chain attribution modeling: complete guide. URL: <https://adequate.digital/en/markov-chain-attribution-modeling-complete-guide/>
30. Weberlo. Markov chain attribution model. URL: <https://www.weberlo.com/guides/markov-chain-attribution-model>
31. WallStreetMojo. Markov model. URL: <https://www.wallstreetmojo.com/markov-model/>
32. Man Group. Shining light on machine learning. URL: <https://www.man.com/insights/shining-light-machine-learning>
33. The Modern Scientist (Medium). What is the Shapley value? URL: <https://medium.com/the-modern-scientist/what-is-the-shapley-value-8ca624274d5a>

34. Treasure Data. Multi-touch attribution (MTA) with Shapley values tells marketers what works best. URL: <https://www.treasuredata.com/blog/multi-touch-attribution-mta-with-shapley-values-tells-marketers-what-works-best/>
35. Tilburg University. Evaluating attribution models on predictive accuracy, interpretability, and robustness Joep van der Plas. URL: <https://arno.uvt.nl/show.cgi?fid=147278>
36. Google Ads Data Hub. Guide to Shapley. URL: <https://developers.google.com/ads-data-hub/guides/shapley>
37. Corvidae.ai. Corvidae vs Shapley and Markov. URL: <https://corvidae.ai/blog/corvidae-vs-shapley-and-markov/>
38. Google Cloud. Explainable AI overview (Vertex AI). URL: <https://cloud.google.com/vertex-ai/docs/explainable-ai/overview>
39. Shapley Value Methods for Attribution Modeling in Online Advertising. Kaifeng Zhao. Seyed Hanif Mahboobi. Saeed R. Bagheri. URL: <https://arxiv.org/pdf/1804.05327>
40. Aidan Cooper. How Shapley values work. URL: <https://www.aidancooper.co.uk/how-shapley-values-work/>
41. Facebook Business. Optimization tips to help maximize your campaign results. URL: <https://www.facebook.com/business/news/optimization-tips-to-help-maximize-your-campaign-results>

ДОДАТКИ

Додаток А

Фрагменти коду Python у побудові шляхів користувача та реалізації методу Шеплі

Побудова шляхів користувача

```

user_paths = (
    df
    .groupby('user_id')
    .agg({
        'marketing_channel': lambda x: ', '.join(pd.unique(x)), # зберігаємо порядок появи
        'conversion': 'max'
    })
    .reset_index()
    .rename(columns={'marketing_channel': 'channel_combo', 'conversion':
'conversion_flag'})
)

user_paths
    .groupby('channel_combo')
    .agg(total_conversions=('conversion_flag', 'sum'))
    .reset_index()
    .sort_values(by='total_conversions', ascending=False)
)

```

coalition_data.head(20)

Реалізація внесків Шеплі

```

def calculate_shapley_from_dataframe(df, channel_col='channel_combo',
conversion_col='total_conversions'):

```

Returns:

```

c_values = df.set_index(channel_col)[conversion_col].to_dict()
df['is_solo'] = df[channel_col].apply(lambda x: x if len(x.split(',')) == 1 else np.nan)
unique_channels = list(df['is_solo'].dropna().unique())
v_values = {}
for coalition in generate_channel_combinations(unique_channels):
    key = ",".join(sorted(coalition))
    v_values[key] = v_function(coalition, c_values)
shapley_values = defaultdict(float)
n = len(unique_channels)
for channel in unique_channels:
    for A_str in v_values:
        A = A_str.split(",")
        if channel not in A:
            A_with_channel = sorted(A + [channel])
            A_with_channel_key = ",".join(A_with_channel)
            weight = (factorial(len(A)) * factorial(n - len(A) - 1)) / factorial(n)
            contrib = v_values.get(A_with_channel_key, 0) - v_values[A_str]
            shapley_values[channel] += weight * contrib
    shapley_values[channel] += v_values[channel] / n
return dict(shapley_values)

```

Приклад очищеного вхідного набору даних

user_id	date_served	marketing_channel	converted
a100000029	01.01.2018	House Ads	TRUE
a100000030	01.01.2018	House Ads	TRUE
a100000031	01.01.2018	House Ads	TRUE
a100000032	01.01.2018	House Ads	TRUE
a100000033	01.01.2018	House Ads	TRUE
a100000034	01.01.2018	House Ads	TRUE
a100000035	01.01.2018	House Ads	TRUE
a100000036	01.01.2018	House Ads	TRUE
a100000037	01.01.2018	House Ads	TRUE
a100000038	01.01.2018	House Ads	TRUE

Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи бакалавра	До 1 листопада 2024 року	
2	Розробка та затвердження завдання кваліфікаційної роботи бакалавра	До 1 грудня 2024 року	
3	Збір інформації, її аналіз, обробка, консультації з науковим керівником	До 10 квітня 2025 року	
4	Підготовка роботи відповідно до вимог оформлення	До 20 квітня 2025 року	
5	Подача роботи для попереднього захисту	До 5 червня 2025 року	
6	Отримання відгуку наукового керівника	До 11 червня 2025 року	
7	Отримання рецензії на кваліфікаційну роботу бакалавра	До 17 червня 2025 року	
8	Перевірка роботи на плагіат	До 18 червня 2025 року	

Науковий керівник: д.е.н, професор

Чорноус Галина Олександрівна

Студент: Пустовіт Майя Григорівна