Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

Використання інструментів Data Science для прогнозування доходу рекламної компанії

студентки 2 курсу магістратури
спеціальності 051 «Економіка»
ОНП «Економічна кібернетика»
денної форми навчання
Сачко Вероніки Володимирівни

Науковий керівник:
докторка економічних наук, професорка
Затонацька Тетяна Георгіївна

Засвідчую, що в цій роботі немає запозичень із праць інших авторів без відповідних посилань
Вероніка САЧКО _____
(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 13 травня 2024 р., протокол № 13

Завідувачка кафедри:
докторка економічних наук, професорка
Олена ЛЯШЕНКО _____
(підпис)

КИЇВ – 2024

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 74 ст., 6 рис., 3 табл., 55 джерел, додатки.

Ключові слова: рекламний нетворк, маркетингові показники, прогнозування, машинне навчання, оцінка точності моделей.

Об'єкт дослідження: діяльність рекламної компанії, що займається закупівлею та продажем трафіку.

Мета дослідження: аналіз та побудова моделей для прогнозування доходу та розробка рекомендацій для збільшення ефективності рекламних компаній.

Методи дослідження: системний підхід, загальнонаукові методи дослідження, статистичний аналіз, моделі машинного навчання.

Наукова новизна, теоретична значимість дослідження: розроблено комплексні моделі прогнозування доходу на основі машинного навчання для моделювання рівня ефективності маркетингових стратегій.

Практична цінність: моделі, розроблені в ході дослідження, можуть бути впроваджені в практику бізнесів для забезпечення їх ефективності.

RESUME

Taras Shevchenko National University of Kyiv, Faculty of Economics, Department of Economic Cybernetics

Key words: advertising network, marketing indicators, forecasting, data analysis tools, machine learning, model accuracy evaluation.

The graduation research of student lies in the development of complex revenue forecasting ML models for modeling the level of effectiveness of marketing strategies and the overall stability of the company.

The work is interesting for using the results of work that the models developed during the study can be implemented in the practice of businesses to provide their effectiveness. Pages – 74, tables – 3, bibliog. – 55, append. – 3.

ЗМІСТ

ВСТУП

Актуальність теми дослідження. В умовах постійно зростаючої конкуренції в сфері реклами, точність прогнозування фінансових результатів стає критично важливою для успішної діяльності рекламних компаній. Завдяки стрімкому розвитку технологій обробки даних, інструменти Data Science відкривають нові можливості для оптимізації маркетингових стратегій та підвищення ефективності рекламних кампаній.

Застосування методів машинного навчання та аналітики дозволяє не тільки точно прогнозувати доходи від рекламних кампаній, але й виявляти найбільш ефективні канали реклами, розуміти поведінку споживачів, і тим самим оптимізувати розподіл бюджетів. Це не лише збільшує ROI (повернення інвестицій), але й сприяє більшій персоналізації рекламних повідомлень, що підвищує їхню ефективність.

Науковий інтерес до цієї теми також підживлюється швидким розвитком інноваційних технологій в області великих даних і аналітики. Вивчення та імплементація передових технологій в аналітиці і прогнозуванні дозволить рекламним компаніям не тільки виживати, але й процвітати в надзвичайно конкурентному середовищі. Таким чином, дослідження використання інструментів Data Science для прогнозування доходу рекламної компанії є не тільки актуальним, але й необхідним для подальшого розвитку ефективних маркетингових стратегій і забезпечення сталого зростання в індустрії.

Аналіз публікацій та досліджень за тематикою використання інструментів Data Science для прогнозування доходу виявив, що глибоке інтегрування технологічних інновацій та аналітичних методів є ключовим для підвищення ефективності та рентабельності підприємств різних галузей та напрямків. Це також сприяє зменшенню витрат і оптимізації процесів та бюджетів на основі використання моделей машинного навчання та інструментів аналізу даних. Помітний внесок у цьому напрямку зробили такі іноземні науковці, як Дж. Г. Шанахан [1, 2],

Ю.Шмідгубер [4], Й. Гудфеллоу [5], а серед українських вчених можна виділити І. Сергієнка [3], Р. Галагана [6] та Т. Затонацьку [7]. Вони досліджували застосування комплексних аналітичних моделей та алгоритмів машинного навчання для прогнозування результатів рекламної діяльності.

Незважаючи на значний прогрес у цій області, деякі аспекти прогнозування доходів від рекламних кампаній залишаються недостатньо дослідженими. Більшість наявних досліджень фокусуються на використанні конкретних технологічних рішень або окремих методів аналізу даних, в той час як інтегровані підходи, що об'єднують різні види аналітики, можуть забезпечити більш точні та універсальні результати. В цьому контексті особливо актуальним є застосування широкого спектру інструментів економетрики та машинного навчання для створення більш ефективних прогнозних моделей, що може відкрити нові можливості для оптимізації рекламних стратегій. Це підкреслює необхідність подальших досліджень зазначених аспектів у рамках обраної теми дипломної роботи, підтверджуючи її актуальність.

Метою даного дослідження є аналіз впливу інструментів Data Science на ефективність і дохідність рекламних кампаній, побудова моделей для прогнозування та розробка рекомендацій для оптимізації діяльності рекламного нетворку. В рамках дослідження були визначені такі основні завдання:

- дослідити характеристики та види рекламних компаній із урахуванням маркетингових показників та чинників впливу на них
- проаналізувати розвиток та застосування інструментів Data Science в рекламній індустрії, визначити ключові інструменти та підходи вимірювання їх ефективності.
- розробити моделі для прогнозування доходу рекламного нетворку, використовуючи дані з внутрішньої бази даних компанії.
- на основі отриманих результатів розробити рекомендації для оптимізації маркетингових рішень та стратегій та стимулювання зростання доходу.

Об'єкт дослідження – діяльність рекламної компанії, що займається закупівлею та продажем трафіку.

Предмет дослідження – методи та моделі, які використовуються для аналізу даних та прогнозування доходів рекламного нетворку.

Теоретико-методологічною базою дослідження для досягнення поставлених завдань включає системний підхід, методи порівняння, узагальнення та статистичний аналіз, основою ж є методи економіко-математичного моделювання, інструменти аналізу даних та машинне навчання.

Інформаційною базою є наукові праці, статті та посібники відповідної проблематики вітчизняних та зарубіжних дослідників, внутрішні дані з бази даних рекламної компанії, адаптовані для проведення дослідження.

Наукова новизна та теоретична значимість дослідження полягає у такому:

- теоретично обґрунтовано різні методики аналізу даних у контексті рекламної індустрії, що дозволяє покращити розуміння ключових чинників, які впливають на успіх рекламних кампаній;
- досліджено вплив маркетингових показників на дохідну складову рекламного нетворку, що допомагає визначити ефект їхніх змін на загальну стабільність компанії;
- проаналізовано й узагальнено різні підходи до вимірювання прибутковості діяльності рекламних компаній;
- розроблено комплексні інтегровані моделі прогнозування доходу, які інтегрують різноманітні інструменти Data Science для моделювання рівня ефективності маркетингових стратегій;

Практична цінність дослідження виражається через розробку рекомендацій для рекламних агентств - результати дослідження можуть бути використані для формулювання стратегій оптимізації маркетингової діяльності кампаній, що може підвищити дохідність і ефективність підприємства, застосування розроблених моделей для реальних рекламних компаній - моделі, розроблені в ході дослідження,

можуть бути впроваджені в практику рекламних агентств для забезпечення більш точного прогнозування результатів кампаній на основі аналізу даних та внесок у поліпшення інструментів аналізу в індустрії реклами - використання передових методів Data Science сприяє розвитку нових підходів до аналізу та управління рекламним доходом, що в кінцевому підсумку може змінити принципи розвитку цієї галузі.

Апробацією наукового дослідження став виступ на науковій конференції «Шевченківська весна 2024» на базі економічного факультету Київського національного університету імені Тараса Шевченка (секція «Моделювання та інформаційні технології в економіці: сучасні виклики та напрями розвитку»).

Структура роботи – робота складається зі вступу, трьох розділів, висновків та списку використаної літератури із 55 позицій та додатків. Загальний обсяг становить 74 сторінок, основний текст займає 65 сторінок.

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНИЙ ОГЛЯД ОРГАНІЗАЦІЇ РОБОТИ РЕКЛАМНИХ КОМПАНІЙ

1.1. Характеристика рекламної галузі та види рекламних компаній

Дослідження діяльності рекламної галузі та рекламних компаній є важливим аспектом сучасних маркетингових стратегій та економічних досліджень. Реклама відіграє ключову роль у формуванні споживацьких уподобань, культурних тенденцій та економічного розвитку. З огляду на швидкі зміни в технологіях і медіа-ландшафті, актуальність таких досліджень стає особливо важливою. Рекламна індустрія не тільки сприяє зростанню економіки за рахунок збільшення продажів та популяризації брендів, але й має значний вплив на суспільні цінності та культурні норми. Через свої повідомлення, реклама формує уявлення споживачів про те, що є бажаним і прийнятним у суспільстві, тому дослідження цієї галузі має значний соціальний, культурний та економічний вплив.

Рекламна галузь включає в себе широкий спектр учасників: від рекламних агенцій, які розробляють та виконують рекламні кампанії, до медіа-платформ, що розповсюджують рекламу (таких як телебачення, радіо, інтернет-ресурси, друковані видання). Також до галузі відносять компанії, що спеціалізуються на маркетингових дослідженнях, цифровому маркетингу, PR, і бренд-менеджменті. Основою для розуміння діяльності рекламних компаній слугує ряд теорій масової комунікації, психології споживача та маркетингових стратегій. Реклама як інструмент маркетингових комунікацій базується на принципах AIDA (Attention, Interest, Desire, Action), які описують етапи взаємодії споживача з рекламним повідомленням. Крім того, теорії медіапланування та покупки медійних послуг дозволяють ефективно розподіляти рекламні бюджети і визначати оптимальні канали комунікацій [8].

З розвитком технологій рекламна галузь значно трансформувалась. Цифровізація принесла інструменти, які дозволяють більш точно націлювати рекламу, аналізувати поведінку споживачів, та оцінювати ефективність рекламних

кампаній. Соціальні медіа, мобільний маркетинг, та інтерактивна реклама є лише кількома прикладами того, як нові технології змінили підходи до реклами. Одним з основних викликів для галузі є зростання скептицизму споживачів щодо реклами та підвищена увага до приватності та конфеденційності даних. Питання такі як "рекламна втома" та боротьба з блокуванням реклами вимагають від компаній більш креативних та етичних підходів. Загальносвітові законодавчі ініціативи, такі як GDPR в Європі, ставлять перед галуззю нові вимоги до прозорості та захисту інформації [13]. Стосовно перспектив розвитку, можна також зазначити, що майбутнє рекламної галузі обіцяє бути наповненим новими можливостями за рахунок розвитку штучного інтелекту, машинного навчання та доповненої реальності (VR). Ці технології дозволять створювати ще більш персоналізовані та залучаючі рекламні кампанії, які можуть взаємодіяти зі споживачами новими та інноваційними способах.

Рекламні агентства спеціалізуються на створенні та проведенні рекламних кампаній для клієнтів. Вони розробляють стратегії ефективного охоплення цільової аудиторії за допомогою різних медіа-каналів, створюють креативні оголошення та часто пропонують новий погляд на покращення маркетингових зусиль компанії. До сфери їхньої компетенції входять дослідження ринку, закупівля медіа та креативні послуги, що робить рекламні агентства цінними партнерами для компаній, які прагнуть просувати свої продукти чи послуги [10]. Діапазон типів рекламних бізнесів коливається від повного спектру послуг до вузькоспеціалізованих. До найбільш поширених бізнес-моделей можна віднести наступні види:

- Рекламні компанії повного циклу. Такі компанії пропонують інтегрований підхід до маркетингу та реклами, включаючи дослідження ринку, стратегічне планування, креативний дизайн, медіа-планування, PR та інші послуги. Вони працюють з клієнтами на всіх етапах рекламної кампанії, від ідеї до реалізації, що дозволяє забезпечити комплексне вирішення маркетингових завдань. Одна з ключових переваг агенції повного циклу полягає в здатності координувати

різні маркетингові канали для створення послідовної та ефективної кампанії. Вони забезпечують, що зв’язок на всіх платформах та медіа є узгодженим, і це підвищує загальну ефективність комунікації. Використання передових аналітичних інструментів дозволяє повносервісним агенціям точно вимірювати результативність кампаній та оптимізувати їх у реальному часі. Вони аналізують дані про взаємодію споживачів, конверсії та ROI (повернення інвестицій), що допомагає підвищити ефективність рекламних витрат [11]. В той же час, варто враховувати, що швидкий розвиток технологій вимагає від агенцій постійно оновлювати свої навички та інструменти, щоб залишатися конкурентоспроможними. Цифрова трансформація маркетингу постійно вносить нові вимоги, тому агенції в свою чергу повинні виконувати для забезпечення ефективної діяльності для своїх клієнтів.

- Медіа-агенції. Вони займають важливе місце в екосистемі маркетингу, спеціалізуючись на плануванні та купівлі рекламного часу та простору та аналізують, де найкраще розмістити рекламу, щоб досягти цільової аудиторії, використовуючи дані та аналітику для оптимізації витрат та ефективності рекламних кампаній. Медіа-агентства ведуть переговори з постачальниками медіа-простору для забезпечення найкращих цін та умов. Вони використовують свою експертизу та обсяги для досягнення масштабу та ефективності витрат. Слід зазначити, що одним із ключових аспектів роботи медіа-агентств є постійний моніторинг результатів рекламних кампаній та їх оптимізація в реальному часі. Це дозволяє агентствам коригувати кампанії з урахуванням реакції аудиторії, що забезпечує кращі результати та вищий ROI, при цьому адаптуючись до постійно мінливого медійного ландшафту та поведінки споживачів [16].
- Цифровий маркетинг. Агентства цифрового маркетингу відіграють ключову роль у сучасному рекламному просторі, зосереджуючись на використанні інтернету, мобільних пристроїв, соціальних медіа, пошукових систем та інших

цифрових каналів для просування продуктів та послуг. Ці агентства використовують різноманітні стратегії та інструменти для залучення цільової аудиторії та перетворення її у лояльних клієнтів. Ефективність цифрового маркетингу часто залежить від здатності збирати, аналізувати та правильно використовувати великі обсяги даних для розуміння поведінки споживачів та оптимізації кампаній. До основних функцій компаній можна віднести як оптимізацію веб-сайтів електронної торгівлі для забезпечення кращого користувацького досвіду та збільшення продажів через інтернет, стратегічне використання платформ соціальних медіа для просування брендів, продуктів чи послуг, забезпечення взаємодії з користувачами, залучення та взаємодії з аудиторією, так і керування рекламними кампаніями, де клієнти платять за кожне натискання на рекламу, що дозволяє швидко привертати трафік та інтерес до веб-сайтів [14]. Остання ознака характерна саме для рекламного нетворку (ad network). Він є важливою складовою цифрового маркетингу, який дозволяє рекламодавцям розміщувати рекламу на різних веб-сайтах та цифрових платформах через централізовану систему. Це сприяє більш ефективному дистрибутивному процесу, забезпечуючи рекламодавцям доступ до широкої мережі потенційних майданчиків для показу їхньої реклами, а видавцям — засіб монетизації своїх веб-ресурсів [17]. Окрім цього, в його основі лежить модель закупівлі та дистрибуції трафіку, а також можливість реалізації додаткових SaaS продуктів для підвищення ефективності бізнесу. Саме на базі такого типу компанії і побудовано дане дослідження.

Отже, рекламна індустрія охоплює широкий спектр спеціалізованих та багатопрофільних агенцій та компаній, кожна з яких відіграє свою унікальну роль у створенні та виконанні ефективних маркетингових стратегій. Розуміння цих відмінностей та можливостей є ключовим для оптимізації рекламних зусиль та досягнення комерційних цілей.

1.2. Характеристика основних показників та ключових метрик цифрового маркетингу

Сучасний світ маркетингу та реклами зазнав значних змін з появою цифрових технологій. Цифровий маркетинг стає все більш важливим елементом стратегій бізнесу, забезпечуючи брендам можливість досягнення ширшої аудиторії, покращення взаємодії з клієнтами та збільшення продажів. В епоху діджиталізації та глобалізації роль реклами та цифрового маркетингу в економічних процесах набуває нових вимірів. Розуміння ключових показників, що визначають ефективність рекламних стратегій, є критично важливим для компаній, які прагнуть оптимізувати свої маркетингові зусилля та максимізувати рентабельність інвестицій. На рис. А.1 додатку А наведена схема, що показує основні метрики, на які орієнтуються рекламні компанії, та їх взаємозв’язок.

Варто також окремо розглянути та охарактеризувати кожен індикатор, який свідчить про ефективність маркетингової діяльності.

1. *Advertising costs ma advertising revenue*. Дані показники є основними для розуміння того, наскільки прибутково функціонує компанія. Витрати на рекламу (ad costs) відображають собою загальну суму коштів, яку компанія витрачає на рекламні кампанії. Ця метрика є достатньо вагомою, оскільки вона впливає на можливість досягнення маркетингових цілей та збільшення впізнаваності бренду. В свою ж чергу дохід від реклами є показником, який вимірює грошове відшкодування, отримане компанією в результаті її рекламних зусиль [18]. Оцінка доходу від реклами дозволяє компаніям визначити ефективність своїх рекламних кампаній та стратегій. Високий дохід свідчить про те, що рекламні кампанії успішно залучають клієнтів та стимулюють продажі. Співвідношення між витратами на рекламу та доходом від неї важливе для розуміння ROI (повернення інвестицій) рекламних кампаній, характеристика якого буде надалі.

2. *ROI (Return on investments)*. Рентабельність інвестицій (ROI), показник, який використовується для оцінки результатів маркетингової кампанії порівняно із загальними витратами на неї, і більш ґрунтовним індикатором ефективності для подальшого планування та розробки стратегій. Формула для розрахунку даної метрики може виглядати наступним чином:

$$ROI = \left(\frac{\text{Витрати на кампанію} - \text{Чистий прибуток від кампанію}}{\text{Витрати на кампанію}}\right) \times 100\%.$$

Загалом, ROI допомагає підприємствам розуміти, наскільки ефективно їхні інвестиції у маркетинг перетворюються на прибутки, та які канали та стратегії є найбільш ефективними, і на основі цього розподіляти маркетинговий бюджет та оптимізовувати свою діяльність. Саме тому, коректне розуміння та використання даного індикатора може значно вплинути на стратегічне планування та загальний успіх компанії.

3. *CPC (Cost Per Click) та CPM (Cost Per Mile)*. Ці показники об'єднані кількома ключовими аспектами в контексті цифрового маркетингу. В першу чергу, обидві метрики є методами розрахунку вартості рекламних кампаній в онлайн середовищі. CPC фокусується на вартості індивідуального кліку, тоді як CPM розраховує вартість за кожну тисячу показів оголошення. CPC є основним індикатором ефективності та вартості привернення трафіку на веб-сайт або лендінгову сторінку [25]. Вона важлива для розрахунку бюджету рекламних кампаній та оптимізації витрат. Її перевагою є те, що CPC дозволяє спрямовувати рекламу на конкретні дії, замість простого показу реклами, і рекламодавці можуть оптимізувати свої кампанії для зниження CPC та підвищення ROI. А CPM в свою чергу виграє в аспекті того, що рекламодавці можуть легко розрахувати вартість кампаній на основі обсягів показів. Така модель є ефективною для досягнення для підвищення рівня обізнаності великої кількості людей, проте вона не гарантує взаємодію з рекламою, тому важко вимірювати конверсії або інші бажані дії. І може бути ситуація, що якщо

покази не призведуть до подальших дій користувачів, то витрати можуть бути неефективними. Загалом, CPC та CPM допомагають формувати стратегічне розуміння взаємодії між рекламними витратами та поведінкою споживачів, що є важливим для успішної реалізації цифрових рекламних кампаній, та за допомогою аналізу ефективності цих метрик, можна коригувати свої стратегії для підвищення ефективності та зниження витрат [26].

4. *Bounce rate (Коефіцієнт відтоку)*. Це метрика, яка вимірює частку відвідувачів веб-сайту, які залишають сайт після перегляду лише однієї сторінки, не взаємодіючи з сайтом далі (наприклад, без кліків по лінках, кнопкам тощо). Ця метрика часто використовується для оцінки релевантності та зацікавленості контенту сайту. Розрахувати даний показник можна наступним чином:

$$Bounce\ rate = \left(\frac{\text{Кількість сесій з однією сторінкою}}{\text{Загальна кількість сесій}}\right) \times 100\%.$$

Відсоток відмов використовується для оцінки якості трафіку на сайт, аналізу ефективності веб-дизайну та контенту, а також для визначення проблем зі зручністю навігації або швидкістю завантаження сторінок [19]. Низький відсоток відмов свідчить про те, що сайт відповідає очікуванням користувачів, тоді як високий показник може вказувати на потребу в оптимізації. На противагу даному індикатору варто розглянути коефіцієнт утримання користувачів.

5. *Retention rate (Коефіцієнт утримання)*. Є важливим метричним показником, який використовується для вимірювання відсотка користувачів, які залишаються з компанією протягом певного періоду часу після їх першої покупки або підписки. Цей показник відображає здатність компанії утримувати своїх клієнтів, що є ключовим аспектом для тривалого бізнес-успіху. Високий коефіцієнт утримання зазвичай вказує на задоволеність клієнтів, ефективність бізнес-моделі та продукту. Також, варто зазначити, що залучення нових клієнтів коштує значно дорожче, ніж утримання існуючих,

тому високий рівень утримання може значно знижувати загальні маркетингові витрати. Коефіцієнт утримання розраховується за формулою:

$$Retention\ rate = \left(\frac{\text{Кількість активних користувачів на кінець періоду} - \text{Кількість нових користувачів за період}}{\text{Кількість активних користувачів на початок періоду}}\right) \times 100\%.$$

Різні сегменти користувачів можуть мати різні рівні утримання, що вимагає сегментованого підходу до аналізу та вдосконалення.

6. *Conversions ma CTR (Click-Through Rate)*. Конверсія в контексті цифрового маркетингу відбувається, коли користувач виконує бажану дію, яка переводить його з потенційного клієнта в реального. Ці дії можуть включати покупку продукту, реєстрацію на подію, підписку на розсилку або завантаження матеріалів. Конверсія є критичним показником ефективності маркетингових кампаній, оскільки вона відображає реальний успіх в приведенні клієнтів до завершення бажаних дій. Їх вимірювання допомагає визначити, наскільки ефективно рекламні та маркетингові ініціативи стимулюють цільові дії користувачів [21]. Однією з найбільш важливих конверсії в площині цифрового маркетингу є клік-рейт. CTR — це відсоток користувачів, які клікають на рекламу або посилання відносно загальної кількості переглядів цього оголошення або посилання. Він розраховується за формулою:

$$CTR = \left(\frac{\text{Кількість кліків}}{\text{Кількість показів}}\right) \times 100\%.$$

CTR використовується для оцінки ефективності рекламних банерів, електронних листів, пошукових оголошень та інших цифрових маркетингових матеріалів. Високий показник метрики вказує на те, що реклама є привабливою та релевантною для аудиторії. Він дозволяє оцінити рівень зацікавленості аудиторії у рекламному контенті та допомогти виявити найбільш ефективні креативи та канали розміщення. Однак, CTR може бути заниженим або

завищеним через штучні кліки, що впливає на достовірність даних, тому треба уникати маніпуляцію даними та впроваджувати перевірку на фрод.

Розуміння показників цифрового маркетингу є фундаментальним для успішного управління та оптимізації маркетингових стратегій у сучасному бізнесі. Кожен показник, такий як конверсії, CTR, CPC, CPM, Retention Rate, та інші, відіграє критичну роль у вимірюванні та аналізі ефективності різних аспектів маркетингових кампаній. Вони дозволяють виміряти успіх рекламних кампаній, визначити, що працює добре, а що потребує змін. Вони є індикаторами того, наскільки добре маркетингові зусилля досягають бізнес-цілей [24]. Ретельне аналізування і використання цих показників може допомогти підвищити повернення інвестицій, мінімізувати витрати та збільшити прибуток. За допомогою метрик маркетологи можуть оптимізувати розподіл бюджету, коригувати стратегії та підвищувати загальну продуктивність кампаній. Також, такі індикатори, як-от Retention Rate, допомагають оцінити ступінь залученості та лояльності клієнтів. Вони вказують на можливість довгострокових відносин з клієнтами та ефективність стратегій утримання. Слід згадати, що у динамічному цифровому середовищі поведінка споживачів постійно змінюється, тому метрики допомагають слідкувати за цими змінами і адаптуватися до них, забезпечуючи відповідність контенту та реклами актуальним потребам та інтересам цільової аудиторії. Додатково до цього, аналіз показників дозволяє бізнесам планувати на основі даних, що в свою чергу знижує ризики та підвищує ймовірність успіху. Використання цих даних для прогнозування майбутніх тенденцій допомагає компаніям бути на крок попереду конкурентів [22].

Отже, глибоке розуміння та ефективне застосування метрик цифрового маркетингу є критичними для забезпечення успішної взаємодії з клієнтами, підвищення ефективності маркетингових ініціатив, та, в кінцевому підсумку, для досягнення стабільного зростання бізнесу.

1.3. Передумови та чинники, що впливають на коливання рекламних показників

Сучасний цифровий маркетинг знаходиться під впливом численних чинників, які можуть викликати значні коливання його показників. Важливість розуміння цих коливань полягає у здатності адаптуватися та оптимізувати маркетингові кампанії для досягнення кращих результатів.

Одним з основних чинників, що впливають на показники цифрового маркетингу, є швидкий розвиток технологій. Інновації, такі як штучний інтелект, машинне навчання, біг дані та інтеграція цифрових асистентів, змінюють способи, якими бренди взаємодіють зі своїми клієнтами. Наприклад, алгоритми персоналізації та автоматизація маркетингу можуть підвищити ефективність рекламних кампаній, водночас знижуючи вартість за клік (CPC) і покращуючи загальний Retention Rate [23]. На основі рішення щодо каналів залучення користувачів компанія може орієнтуватися на своїх потенційних клієнтів. Домен онлайн-маркетингу охоплює широкий спектр платформ, і з часом з’являється багато інших. Компаніям необхідно проводити регулярно аналіз витрат і доходів і ранжувати різні засоби на основі їхньої економічної ефективності. Додатково до цього, контент-маркетинг значною мірою сприяє покращенню конверсії та утримання споживачів. З тенденцій цілком очевидно, що контент є «королем» цифрового маркетингу. Він має бути унікальним і привабливим, містити в собі, наприклад ексклюзивне зображення, відео, веб-сайт, блог тощо. Контент-маркетинг, з точки зору впізнаваності, безпосередньо відповідає за те, щоб про бренд дізналося якомога більше людей. І тільки якщо контент буде якісним та будуть ефективно налаштовані всі джерела трафіку, алгоритми Google дозволять йому охопити маси [28].

Варто зауважити, що зміни у поведінці споживачів також можуть безпосередньо відображатися на показниках цифрового маркетингу. Користувачі стають більш обізнаними і вимогливими, вони очікують більшої персоналізації, швидкої взаємодії та високого рівня обслуговування. Ці зміни можуть впливати на

показники такі як CTR, конверсії та загальне задоволення клієнтів. Соціальні медіа і мобільні технології також істотно змінили способи, якими споживачі отримують інформацію та роблять покупки, що вимагає від брендів більшої адаптації своїх стратегій [31]. Інтернет-маркетинг є висококонкурентним, і дії конкурентів можуть впливати на показники діяльності інших учасників ринку. Нові стратегії, кампанії, просування та цінові війни можуть значно змінити ефективність рекламних кампаній. Моніторинг конкурентів і аналіз їхніх дій можуть допомогти компаніям прогнозувати зміни у показниках та своєчасно на них реагувати.

Проте, серед усіх цих чинників, не можна оминути й макроекономічні умови. Економічні коливання, такі як рецесія, інфляція, зміни у споживчих витратах та регуляторні зміни, мають великий вплив на діджитал-маркетинг. Наприклад, економічна нестабільність може знизити споживчі витрати, що в свою чергу позначиться на конверсіях та загальній ефективності маркетингових кампаній. Підприємства повинні бути готові до швидких коректив у своїх стратегіях, щоб ефективно реагувати на зовнішні економічні зміни. В контексті цього не слід забувати і про те, що законодавчі зміни, особливо ті, що стосуються конфіденційності даних та рекламних правил, можуть мати значний вплив на стратегії маркетингу. Нові регуляції, такі як GDPR у Європейському Союзі та CCPA в Каліфорнії, вимагають від компаній змін у способах збору та використання даних користувачів, що може змінити підходи до цільової реклами та персоналізації [32].

Окремо ще варто розглянути передумови змін показників рекламного нетворку, який забезпечує зв’язок між рекламодавцями та видавцями, і дозволяє рекламним кампаніям досягати більш широкої аудиторії з високою точністю таргетингу. Ці платформи спрощують процес покупки і продажу реклами, автоматизуючи багато аспектів, які раніше вимагали ручної роботи та значних часових витрат. Коливання показників у цих нетворках також можуть бути обумовлені додатковою низкою передумов та чинників. Одним із таких факторів є те, що рекламні платформи постійно оновлюють свої алгоритми для покращення

користувацького досвіду та ефективності рекламних кампаній [29]. Ці зміни можуть несподівано впливати на показники, такі як CTR, CPC і CPA, оскільки вони змінюють спосіб показу реклами користувачам. Важливо також зауважити, що технічні збої, такі як відмови серверів або помилки в трекінгу, можуть тимчасово спотворювати метрики, що в свою чергу може призвести до прийняття некоректних рішень. Наприклад, невірне відслідковування конверсій може викривити дані про ефективність рекламних кампаній, тому дуже важливо налагоджувати моніторинг основних показників, для швидкого виявлення проблем, в разі їх виникнення, та запобігання їх значного впливу.

Отже, висновок з цього полягає в тому, що успіх у цифровому маркетингу та управлінні рекламними нетворками вимагає не тільки розуміння чинників та передумов, які впливають на метрики, але й здатності гнучко адаптуватися до них. Моніторинг, аналіз та оптимізація показників на основі виявлених трендів і змін є ключовими для підтримання конкурентоспроможності та досягнення максимальної віддачі від інвестицій у рекламні кампанії.

Висновки до розділу 1

Рекламна галузь є важливою ланкою у модернізації економічних відносин між брендами та споживачами, дозволяючи компаніям ефективно просувати свої послуги та товари до цільової аудиторії. Різноманітність рекламних компаній, від агенцій повного циклу до спеціалізованих медійних та консалтингових фірм, підкреслює багатошаровість цієї сфери. Ці компанії розробляють стратегії, які включають використання цифрових каналів, таких як соціальні медіа, пошукові системи, і електронну пошту, що забезпечують ефективну взаємодію з потенційними покупцями.

Основні метрики, такі як конверсія, CTR (Click-Through Rate), CPC (Cost Per Click), CPM (Cost Per Mille) і ROI (Return on Investment), допомагають вимірювати ефективність рекламних кампаній. Ці показники дозволяють маркетологам

аналізувати, наскільки ефективно реклама перетворюється на реальні продажі або залучає увагу аудиторії.

Водночас, рекламна галузь і показники цифрового маркетингу схильні до коливань, які можуть бути викликані низкою зовнішніх та внутрішніх чинників. Зміни у технологіях, такі як розвиток штучного інтелекту та машинного навчання, мають потенціал радикально трансформувати підходи до рекламної діяльності, підвищуючи персоналізацію та ефективність рекламних звернень. Конкурентне середовище також вимагає постійного аналізу дій конкурентів, оскільки інноваційні маркетингові стратегії або зміни в ціновій політиці можуть впливати на власні рекламні показники.

Додатково, макроекономічні умови, такі як рецесії або інфляція, можуть змінити споживацькі витрати та поведінку, впливаючи на ефективність рекламних ініціатив. Регуляторні зміни, особливо у сфері захисту даних і приватності, вимагають налагодження рекламних стратегій, щоб відповідати новим вимогам без втрати залученості аудиторії.

Розуміння і адаптація до цих умов і динамічних чинників є ключем до підтримання ефективності та реалізації потенціалу цифрового маркетингу. Це дозволяє компаніям не тільки виживати в умовах постійної зміни, але й досягати успіху, розвиваючи більш осмислені і результативні маркетингові стратегії.

РОЗДІЛ 2. АНАЛІЗ ВИКОРИСТАННЯ ІНСТРУМЕНТІВ АНАЛІТИКИ ДАНИХ ТА МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ

2.1. Характеристика інструментів Data Science та їх використання

Для прогнозування рекламного доходу, Data Science використовує цілий арсенал інструментів, кожен з яких має свої особливості, що дозволяють максимізувати ефективність аналізу та висновків. Використання цих інструментів включає збір, обробку, моделювання та візуалізацію даних. Деякі із них розглянемо нижче:

- Python є однією з найбільш популярних мов програмування в області Data Science завдяки своїй гнучкості та широкому спектру бібліотек. Для прогнозування доходів рекламних нетворків використовуються бібліотеки як Pandas для обробки даних, Matplotlib та Seaborn для візуалізації результатів, а Scikit-learn або TensorFlow для побудови предиктивних моделей на основі історичних даних.
- R має ряд спеціалізованих пакетів для статистичного аналізу та моделювання, які можуть бути особливо корисними при роботі з комплексними моделями прогнозування. ggplot2 дозволяє створювати складні візуалізації, що є важливим для розуміння тенденцій та закономірностей у доходах від реклами.
- SQL необхідний для ефективної роботи з базами даних, де зберігаються дані про рекламні кампанії, кліки, перегляди та інші метрики, що впливають на доходи. Знання SQL дозволяє ефективно екстрагувати необхідні дані для подальшого аналізу.
- Tableau інтенсивно використовується для створення дашбордів та репортів, що дозволяє керівництву компанії швидко оцінювати ефективність рекламних кампаній та приймати обґрунтовані рішення на основі візуальної інформації.
- Apache Spark використовується для аналізу великих обсягів даних у режимі реального часу, що може бути корисним для оптимізації рекламних кампаній і швидкого реагування на зміни в поведінці користувачів або на ринку.

- Завершується процес використанням Jupyter Notebook для демонстрації результатів досліджень та моделей, забезпечуючи прозорий та інтерактивний спосіб подання результатів [55].

Кожен з цих інструментів вносить свій вклад у здатність компаній точно прогнозувати доходи та адаптуватися до змін у рекламному середовищі, дозволяючи максимізувати ефективність рекламних бюджетів та стратегій.

Прогнозування рекламного доходу за допомогою моделей Data Science вимагає ретельного вибору і застосування відповідних статистичних моделей та машинного навчання, які можуть адекватно обробити великі обсяги даних і визначити ключові фактори, що впливають на доходи. Варто проаналізувати основні із них, які найчастіше використовуються для прогнозування різних показників як на локальному, так і на більш глобальному рівнях.

1. *Логістична регресія*. Логістична регресія є одним з основних алгоритмів машинного навчання, що широко використовується для рішення задач класифікації. Незважаючи на свою простоту, цей метод демонструє високу ефективність в задачах, де необхідно прогнозувати категоріальні результати з великої кількості чисельних або бінарних змінних. Основна ідея логістичної регресії полягає в тому, щоб моделювати ймовірність настання події (наприклад, чи клікне користувач на рекламу) як логістичну функцію від змінних, які характеризують цю подію. Це відрізняє її від лінійної регресії, де пряма залежність між змінними і результатом не завжди прийнятна або реалістична, особливо коли результат має обмежений діапазон, як у випадку бінарного виходу. Формально логістична регресія моделює ймовірність настання події за допомогою логістичної функції, яка є S-подібною кривою, виведеною за допомогою логіт-перетворення (рис. Б.1 додатку Б). Її формула виглядає наступним чином:

$$Logit(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n, \qquad (2.1)$$

де p відображає ймовірність настання події (наприклад, клік по рекламі), $\beta_0, \beta_1, \dots, \beta_n$ – коефіцієнти моделі, а $x_1, \dots, x_n$ – вхідні змінні.

Ймовірність p обчислюється як $p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$ (2.2). Структура моделювання логістичної регресії наведена на рис. Б.2 додатку Б.

Застосування логістичної регресії для прогнозування доходу може бути реалізоване через моделювання ймовірностей різної рекламної активності. Наприклад, модель може оцінити ймовірність того, що користувач клікне на рекламу, яка є важливим індикатором ефективності рекламної кампанії. Це дозволяє рекламодавцям налаштувати свої стратегії, виходячи з прогнозованих результатів, аналізувати ефективність різних рекламних каналів і оптимізувати рекламні бюджети [33]. Логістична регресія відрізняється своєю здатністю до швидкої адаптації та інтерпретації результатів, роблячи її незамінним інструментом в арсеналі інструментів машинного навчання для аналізу та прогнозування, проте не робить її універсальною. Логістична регресія використовується для моделювання ймовірності настання певної події, зокрема, для прогнозування, чи користувач клікне по рекламному оголошенню. Стандартна логістична регресія ефективна для бінарної класифікації (так/ні). Хоча існують її варіанти для багатокласової класифікації (наприклад, мультиноміальна логістична регресія), вони можуть бути менш ефективними порівняно з іншими сучасними методами.

2. *Дерева рішень*. Дерева рішень є одним із фундаментальних алгоритмів машинного навчання, використовуваним як для класифікації, так і для регресійних задач. Цей алгоритм моделює рішення та їх можливі наслідки у вигляді деревоподібної структури, з легко інтерпретованими правилами та рішеннями, які дозволяють зрозуміти логіку, за якою модель приходить до певних висновків. Оскільки дерево рішень дотримується контрольованого підходу, алгоритм навчається на попередньо оброблених даних [37]. Дерева рішень будуються зверху вниз, причому кореневий вузол завжди знаходиться

у верхній частині структури, а листя дерева представляють результати. Після кореневого вузла кожен вузол ділиться на кілька вузлів.Дерева рішень будуються шляхом рекурсивного розбиття даних на підгрупи, використовуючи найбільш інформативні атрибути вибірки. Ці рішення ґрунтуються на метриках, які вимірюють рівень "нечистоти" або варіативності в даних, зокрема:

- *Індекс Джині* для класифікації, який вимірює ступінь того, наскільки часто випадково вибраний елемент буде неправильно класифікований:

$$G = 1 - \sum_{i=1}^{n} p_i^2, \qquad (2.3)$$

де p_i – ймовірність того, що елемент належить до класу i.

- *Ентропія*, що використовується для оцінки рівня неоднозначності чи випадковості набору даних, а її значення представляє собою ступінь випадковості певного вузла. Цей параметр дозволяє оцінити, наскільки результати розподілені випадковим чином, особливо коли є мало даних для чіткого прогнозування [42]. Вища ентропія свідчить про більшу випадковість у наборі даних. При побудові дерева рішень краще вибирати вузли з меншою ентропією. Формула виглядає наступним чином:

$$\mathrm{H} = -\sum_{i=1}^{n} p_i log_2(p_i), \qquad (2.4)$$

де p_i – аналогічно є ймовірністю класу i.

Додатково слід розглянути такий підвид, як випадковий ліс. Random forest складається з великої кількості окремих дерев рішень, які разом утворюють ансамбль. Зазвичай ці дерева навчаються за допомогою методу беггінга. Кожне дерево в ансамблі випадкового лісу робить свій прогноз щодо класу, а модель приймає клас, який набрав найбільшу кількість голосів, як кінцевий прогноз. Для беггінгу кожна модель створюється на основі зразків, отриманих із повного набору даних з повторенням. Кожна модель навчається незалежно, а остаточний прогноз формується на основі голосування між усіма моделями. Головна відмінність між випадковим лісом та беггінгом полягає в тому, що

випадковий ліс випадковим чином вибирає найкращу функцію для розбиття з підмножини доступних функцій, тоді як беггінг використовує всі можливі функції для визначення оптимального розбиття [40].

3. *Бустингові моделі*. Бустинг є однією з передових технік ансамблевого машинного навчання, яка полягає у побудові сильного класифікатора з послідовної комбінації слабких класифікаторів. Цей метод працює шляхом покращення прогнозу слабких моделей, насамперед дерев рішень, що дозволяє зменшити як упередженість, так і дисперсію в остаточному моделюванні. Основні види бустингу включають AdaBoost, Gradient Boosting та XGBoost, кожен із яких має унікальні характеристики та можливості. AdaBoost починає з тренування слабкого класифікатора на всіх доступних даних і наступно присвоює більшу вагу прикладам, які були неправильно класифіковані. Нові класифікатори фокусуються на складніших випадках, створюючи серію моделей, які спеціалізуються на різних аспектах даних. Формула оновлення ваги для AdaBoost визначається як:

$$D_{t+1}(i) = \frac{D_t(i)\exp\left(-\alpha_t y_i h_i(x_i)\right)}{Z_t}, \qquad (2.5)$$

де $D_t(i)$ є вагою і-го прикладу на t-ому кроці, α_t - вагою прийнятого класифікатора, y_i - істинним міткам, $h_i(x_i)$ - прогнозом класифікатора, Z_t - нормалізуючим чинником. Gradient Boosting використовує градієнтний спуск для мінімізації помилок між реальними та передбаченими значеннями. Він послідовно додає нові моделі, кожна з яких виправляє помилки, зроблені попередніми, але замість фокусу на помилково класифікованих прикладах, кожен новий класифікатор намагається зменшити загальну помилку системи. Формула оновлення для Gradient Boosting може бути представлена як:

$$F_{t+1}(x) = F_t(x) + \nu \sum_{i=1}^{N} \gamma_i h_t(x_i), \qquad (2.6)$$

де $F_t(x)$ - поточна комбінована модель, ν – швидкість навчання, γ_i - коефіцієнт для кожного дерева, що визначається через оптимізацію втрат [47]. XGBoost є

розширенням Gradient Boosting з оптимізацією для швидкості та продуктивності. Він включає регуляризацію для зменшення перенавчання та оптимізовану обробку даних, що робить його дуже популярним у промислових застосуваннях. XGBoost може автоматично керувати відсутніми даними і використовує блочну структуру для більш ефективних обчислень.

$$Obj = \sum_{i=1}^{N} l(y_i, \hat{y}_i^t) + \sum_{k=1}^{K} \Omega(f_k), \qquad (2.7)$$

де l – функція витрат, а Ω - регуляризуючий член, який впливає на складність моделі. CatBoost є ще одним сучасним алгоритмом бустингу, який спеціалізується на обробці категоріальних даних. Він включає передові методи для ефективної обробки категоріальних ознак, що дозволяє виконувати бустинг без попереднього перетворення цих ознак у числовий формат. Це забезпечує значне підвищення продуктивності та точності моделей. CatBoost використовує градієнтний спуск для оптимізації, але з особливістю — кожен етап оновлення включає обчислення для "випадково відкладених" даних, щоб запобігти перенавчанню:

$$F_{t+1}(x) = F_t(x) + \alpha \sum_{i=1}^{N} g_i \cdot h_t(x_i), \qquad (2.8)$$

де $F_t(x)$ – поточна модель, α – крок навчання, g_i - градієнти втрат, визначені відповідно до цільової функції, і $h_t(x_i)$ - внесок кожного нового дерева. Бустингові моделі, зокрема XGBoost, ефективні для прогнозування доходу від рекламних кампаній або продажів продуктів завдяки їхній здатності до моделювання складних нелінійних взаємозв'язків у великих обсягах даних. Вони здатні виявити важливі ознаки, що впливають на доходи, і надавати передбачення, які можуть допомогти компаніям оптимізувати маркетингові стратегії та покращити фінансові результати. Їхня висока точність і здатність до узагальнення роблять бустинг ідеальним вибором для складних аналітичних задач, де потрібно максимально точне прогнозування [44]. CatBoost може бути також особливо корисним для аналізу рекламних доходів, продажів продукції або будь-яких інших бізнес-даних, де категоріальні ознаки відіграють важливу

роль. Завдяки ефективному кодуванню та обробці категоріальних даних, CatBoost здатен виявити складні шаблони в даних, які інші моделі можуть пропустити. Це робить його ідеальним для сценаріїв, де точність прогнозування доходу критично важлива для стратегічного планування та прийняття рішень.

Отже, застосування цих моделей забезпечує не лише точність прогнозування, але й гнучкість у прийнятті рішень, дозволяючи рекламним компаніям оптимізувати свої стратегії та максимально використовувати бюджети. Правильний вибір та налаштування моделі, а також здатність адаптувати моделі під змінні умови ринку, відіграють ключову роль у забезпеченні успіху рекламних кампаній.

2.2. Визначення основних етапів моделювання для прогнозування доходу рекламного нетворку

Застосування технологій машинного навчання для аналізу доходів та витрат є ключовим елементом стратегічного планування та ефективного керування бізнесом. Моделювання на основі машинного навчання включає кілька ключових етапів, кожен з яких є важливим для розробки ефективної та функціональної моделі. Основні етапи включають в себе збір та очищення даних, вибір алгоритму, навчання моделі, її тестування або оцінка. Варто розглянути кожну складову окремо.

Першими і одним із найважливіших етапів йдуть збір та очищення даних, що є критично важливими кроками в процесі моделювання доходу на основі машинного навчання та потребують високої точності та особливої уваги. Починається цей процес зі збору необхідних даних, що можуть походити з різноманітних джерел, включаючи внутрішні бази даних компанії, відкриті джерела даних, соціальні мережі тощо. Він може відбуватися як ручним способом, так і за допомогою автоматизованих систем, що є більш поширеним та зручним. Важливо зібрати достатньо даних, щоб забезпечити репрезентативність, актуальність та статистичну значимість вивчених моделей [45].

Після збору даних слідує їх очищення, що включає видалення або корекцію неповних, невідповідних або помилкових записів. Це може бути автоматичне видалення дублікатів, коригування формату даних або заміна відсутніх значень. Очищення також може включати нормалізацію даних, де значення приводяться до одного масштабу, що особливо корисно для алгоритмів, чутливих до масштабу даних. Крім того, варто зазначити, що в процесі очищення важливо врахувати видалення або обробку викидів (аутлайєрів), тобто даних, що значно відрізняються від більшості інших значень [49]. Вони можуть спотворювати результати аналізу та навчання моделей, тому їх потрібно або коригувати, приводячи до середнього значення, або видаляти.

Після очищення даних, наступним кроком є їх трансформація, яка може включати створення нових змінних або адаптацію існуючих для кращого представлення особливостей даних, які є важливими для моделей машинного навчання. Також важливим є розділення даних на навчальні та тестові датасети, що дозволяє оцінити ефективність моделі на невідомих їй раніше даних. У сукупності, процеси збору та очищення даних формують фундамент для подальшого успішного моделювання та аналізу даних, які дозволяють розробляти точні та ефективні моделі.

Етап вибору алгоритму є ключовим у процесі моделювання на основі машинного навчання, оскільки від правильності вибору алгоритму залежить ефективність та точність кінцевих прогнозів. Вибір алгоритму залежить від ряду факторів, включаючи природу та складність даних, об'єм доступних даних, необхідність інтерпретації результатів, обчислювальні ресурси та специфічні цілі проекту [48].

Спершу визначається тип задачі машинного навчання, яка може бути регресією, класифікацією, кластеризацією або змішуванням декількох типів. Залежно від цього, вибираються алгоритми, які найкраще підходять для вирішення цих задач. Наприклад, для задач класифікації часто використовуються алгоритми

на основі дерев рішень, нейронних мереж або підтримувальних векторних машин. Далі важливо оцінити об'єм та характеристики доступних даних. Для великих наборів даних можуть бути підходящими потужні алгоритми, які вимагають значних обчислювальних ресурсів, такі як глибоке навчання. У випадку з меншими наборами даних або коли потрібна висока швидкість обробки, можуть бути більш ефективними простіші алгоритми, такі як лінійна регресія або логістична регресія.

Кінцевий вибір алгоритму часто супроводжується експериментуванням, де кілька алгоритмів можуть бути порівняно за їх ефективністю на частині доступних даних через процес крос-валідації та оцінки точності. Це дозволяє обрати алгоритм, який найкраще вирішує задачу в контексті конкретних умов та обмежень.

У сукупності, вибір алгоритму вимагає ретельного аналізу та може включати значну кількість тестування та тонкого налаштування, щоб забезпечити найкращу можливу ефективність моделі [50].

Етап навчання моделі є наступним у розробці систем на основі машинного навчання, який полягає у використанні даних для тренування алгоритму розпізнавати певні шаблони та закономірності. На цьому етапі вже сформований певний алгоритм, який було обрано на попередньому етапі, і визначена мета — ефективно «навчити» цей алгоритм робити точні прогнози на основі наданих даних.

Так як спочатку дані, з якими працюватиме модель, мають бути належним чином підготовлені, що включає їх очищення та трансформацію, це дозволить усунути виникнення помилок, заповнити відсутні значення та конвертувати дані у формат, зручний для обробки алгоритмом.

Під час навчання моделі алгоритм машинного навчання аналізує тренувальний набір даних, намагаючись знайти шаблони, що дозволять йому робити прогнози. Залежно від типу моделі, процес навчання може включати адаптацію ваг, налаштування параметрів або формування правил рішень. Модель ітеративно перевіряє свої прогнози проти фактичних результатів і коригує свої внутрішні параметри для мінімізації помилок [52]. Також, вона може проходити через

багаторазові цикли тренувань і коригувань, поки не досягне задовільного рівня точності або поки не буде вичерпано встановлений ліміт ітерацій. Ефективність цього процесу значною мірою залежить від кількості, якості та різноманітності даних, доступних для навчання, а також від складності самої задачі, яку модель має вирішувати.

По завершенню навчання моделі вона готова до тестування та оцінки на тестовому наборі даних, що дозволяє перевірити її здатність узагальнювати знання на нових, раніше невідомих даних. Такий підхід дозволяє зрозуміти, наскільки ефективно модель може працювати в реальних умовах.

Далі йде тестування моделі, яке забезпечує верифікацію якості та ефективності моделі перед її впровадженням у реальні умови. Перед початком тестування дані зазвичай розділяють на навчальний, валідаційний та тестовий набори. Навчальний набір використовується для тренування моделі, валідаційний — для налаштування параметрів та вибору кращої моделі, а тестовий — для оцінки кінцевої ефективності [53].

Після навчання моделі на валідаційному наборі даних вибирається найкраща модель, яка потім перевіряється на тестовому наборі даних. Цей крок дозволяє зрозуміти, як модель буде працювати в реальному світі. Важливо, що тестовий набір повинен бути абсолютно незалежним від набору даних, який використовувався під час навчання, щоб забезпечити об'єктивність оцінки. Тестування моделі включає в себе оцінку різних метрик ефективності, таких як точність, повноту, F1-скор та область під кривою ROC. Вибір конкретних метрик залежить від специфіки завдання та вимог проекту. Також можуть бути використані додаткові методи оцінки, такі як крос-валідація, щоб забезпечити стійкість моделі до різноманітності даних.

Після завершення тестування і аналізу результатів може бути проведена оптимізація моделі, щоб покращити її продуктивність. Це може включати налаштування гіперпараметрів, використання більш складних моделей або

додавання нових функцій до даних. Кінцевий етап перед реалізацією моделі на практиці — це оцінка її стабільності та надійності, що може зажадати подальших ітерацій тестування та налагодження. Проте, тестування та оцінка моделі не завершуються з моментом її розгортання [52]. Важливо проводити постійний моніторинг та оновлення моделі, враховуючи зміни в даних і умовах її використання. Такий підхід дозволяє забезпечити стабільність та актуальність системи машинного навчання на протязі її життєвого циклу.

Отже, процес моделювання для прогнозування доходу є комплексним та багатоетапним. Зі збору та підготовки даних до тестування та впровадження моделі, кожен етап має ключове значення для розробки ефективної моделі. Особливу увагу слід звертати на якість даних, вибір адекватної моделі, точність прогнозування та здатність моделі адаптуватися до змін у зовнішньому середовищі. Ефективне виконання цього процесу не лише підвищує точність прогнозів доходів, але й сприяє кращому розумінню динаміки ринку, що є важливим для стратегічного планування та прийняття обґрунтованих комерційних рішень у сфері рекламної діяльності.

2.3. Методи перевірки точності використання моделей та побудованого прогнозу

Методи перевірки точності та оцінювання моделей і побудованих прогнозів є важливою складовою в аналітиці даних і машинному навчанні. Точність моделі визначає, наскільки добре модель передбачає або класифікує дані.

Для кількісної оцінки моделі використовуються різні метрики. Для задач регресії часто використовують певні види похибок ($MSE, RMSE, MAE, MAPE$) та коефіцієнт детермінації (R^2). Більш детально далі окремо про кожний показник.

- *MSE (Mean Squared Error, середньоквадратична помилка)*. MSE вимірює середнє квадратів різниць між фактичними та прогнозованими значеннями. Це дає уявлення про те, наскільки близько прогнози моделі до фактичних значень даних. Вона є досить чутливою до високих похибок. Це важливо, коли великі

помилки в прогнозі відіграють значущу роль у плануванні та витратах, наприклад, у бюджетуванні рекламних кампаній. Формула обрахунку виглядає наступним чином:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (2.9)$$

де $\hat{y}_i$ – прогнозовані значення, y_i – фактичні значення, n – кількість спостережень.

Вищі значення MSE вказують на гіршу якість моделі та на те, що модель може давати значно неправильні прогнози, що особливо критично для планування рекламного бюджету. Менші значення MSE свідчать про те, що модель точніше прогнозує доходи, що допомагає зменшити фінансові ризики.

• *RMSE (Root Mean Squared Error, корінь з середньоквадратичної помилки)*. RMSE є квадратним коренем із MSE, що робить його більш інтерпретованим, оскільки має ті ж одиниці виміру, що й залежна змінна.

$$RMSE = \sqrt{MSE} \qquad (2.10)$$

Як і MSE, великі значення RMSE вказують на гіршу прогнозну силу моделі. У контексті рекламного доходу, низький RMSE вказує на те, що помилки між фактичними та прогнозованими доходами невеликі, що сприяє точнішому бюджетуванню та кращому розподілу рекламних витрат.

• *MAE (Mean Absolute Error, середня абсолютна помилка).* MAE вимірює середнє абсолютне відхилення між фактичними значеннями та прогнозами. Це проста та інтуїтивно зрозуміла метрика, яка не залежить від масштабу та має ті ж одиниці, що й вимірювані дані. Розраховується за формулою:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \qquad (2.11)$$

де $\hat{y}_i$ – прогнозовані значення, y_i – фактичні значення, n – кількість спостережень.

MAE можна використовувати для порівняння різних моделей або підходів до прогнозування. Модель із нижчим значенням похибки вважається кращою,

оскільки вона дає точніші прогнози. Менші значення MAE свідчать про те, що модель стабільно дає точніші прогнози, що є критично важливим для планування бюджету та стратегічних рішень у сфері маркетингу.

• *MAPE (Mean Absolute Percentage Error, середня абсолютна відсоткова помилка)*. MAPE вимірює середнє абсолютне відхилення між фактичними та прогнозованими значеннями у відсотковому виразі. Ця метрика корисна для порівняння прогнозів на різних масштабах даних і дає легко інтерпретоване значення помилки. Має наступний вигляд розрахунку:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|, \qquad (2.12)$$

Для рекламних доходів, високий MAPE може свідчити про нестабільність прогнозів або про недооцінку значущих трендів чи сезонних змін. Низький MAPE вказує на те, що модель забезпечує стабільні та надійні прогнози доходів, що дуже важливо для ефективного планування рекламних кампаній.

• *R^2 (R-squared, коефіцієнт детермінації)*. R^2 вимірює частку варіативності в залежній змінній, яку можна пояснити за допомогою незалежних змінних у моделі. Значення R^2 близьке до 1 вказує на те, що модель добре пояснює залежність, а значення близьке до 0 означає, що модель не пояснює варіативність залежної змінної взагалі.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}, \qquad (2.13)$$

де $\bar{y}_i$ – середнє значення y.

У контексті рекламного доходу, високий R^2 свідчить про те, що модель адекватно враховує фактори, що впливають на доходи. Це дозволяє рекламодавцям більш впевнено робити прогнози та оптимізувати свої стратегії на основі вивчених залежностей [47]. Кожна з цих метрик на свій лад допомагає розуміти ефективність та точність прогнозування в різних аспектах, а також допомагає у вдосконаленні рекламних стратегій за допомогою точного планування та виконання бюджету.

Для задач класифікації використовуються такі метрики, як $Accuracy$, $Recall$, $Precision$, $F1-score$ та $AUC-ROC$. Варто коротко розглянути кожну з них.

• *Precision*. Дана метрика вимірює точність передбачення позитивних класів. Точність визначається як відсоток правильно передбачених позитивних випадків від усіх випадків, які модель визначила як позитивні. Висока точність означає, що значна частина позитивних передбачень дійсно є позитивними, але може втрачатися деяка кількість істинно позитивних випадків, які не були ідентифіковані. Ця метрика особливо важлива, коли наслідки хибно позитивних результатів важливіші або більш критичні, ніж наслідки хибно негативних [48]. Метрика розраховується наступним чином:

$$Precision = \frac{TruePositive}{TruePositive+FalsePositive}$$

Обрахунок точності (Precision) відіграє ключову роль у різних сферах, де важливо мінімізувати кількість хибно позитивних результатів. Точність може бути вирішальною у ситуаціях, де витрати чи ризики, пов'язані з неправильним позитивним класифікаційним рішенням, є особливо значними. У контексті діяльності рекламних компаній, якщо модель матиме низьку точність, це може призвести до втрати рекламних бюджетів через таргетування непідходящих користувачів, які не перетворюються на клієнтів. Висока точність допомагає зменшити такі випадки, забезпечуючи, що реклама показується тим, хто більш імовірно згенерує доходи. Підвищення точності прогнозування може також вказати на потребу вдосконалення аналітичних методів або використання більш деталізованих даних, щоб краще розуміти потреби і поведінку користувачів та краще підбирати рекламні банери, що підвищує релевантність і користь реклами для людини [44].

• *Recall*. Повнота (recall) оцінює здатність моделі правильно ідентифікувати усі актуальні випадки певного класу. Ця метрика показує, яка частка з усіх реальних позитивних випадків була вірно класифікована як позитивна. Висока

повнота означає, що модель ефективно виявляє більшість позитивних випадків, але це може включати ризик збільшення кількості помилково позитивних результатів (хибно позитивних). Повнота визначає, яку частину реальних позитивних випадків наша модель змогла коректно ідентифікувати. Цей показник є важливим у ситуаціях, де хибно негативні результати є більш проблематичними, ніж хибно позитивні. Формула розрахунку визначається як:

$$Recall = \frac{TruePositive}{TruePositive+FalseNegative}$$

Розрахунок повноти (Recall) для моделі прогнозування рекламного доходу має велике значення, оскільки він дозволяє оцінити ефективність моделі в ідентифікації всіх потенційно прибуткових рекламних взаємодій. У контексті рекламних кампаній, висока повнота означає, що модель ефективно виявляє більшість ситуацій, коли реклама може згенерувати дохід, мінімізуючи кількість "втрачених можливостей" або ситуацій, коли потенційно дохідна реклама не використовується належним чином. Висока повнота допомагає забезпечити, що модель не пропускає рекламні взаємодії, які могли б принести прибуток. Це критично для ефективної оптимізації бюджетів реклами і збільшення ROI (повернення інвестицій). Проте покладання надмірного акценту на повноту може призвести до зростання хибно позитивних результатів, коли реклама показується користувачам, які не зацікавлені в продукті. Це може вести до втрати коштів і зниження ефективності рекламних кампаній. Тому, важливо знайти баланс між повнотою і точністю, щоб оптимізувати загальну продуктивність рекламних стратегій.

• *Accuracy*. Метрика визначає відсоток загальної кількості правильних передбачень, які здійснила модель, у порівнянні з усіма оціненими випадками. Ця метрика є однією з найпростіших для розуміння, оскільки вона вимірює частку правильно ідентифікованих випадків (як позитивних, так і негативних) від загальної кількості випадків у датасеті. Точність може бути не надто

інформативною, якщо класи незбалансовані, тобто коли один з класів значно перевищує інший за кількістю зразків. Точність визначається як частота правильних передбачень, які робить класифікатор. Її можна обчислити як відсоток вірних передбачень від усіх зроблених прогнозів [46]. Наприклад, якщо модель показує точність 99%, може здатися, що вона працює відмінно, однак такий висновок може бути оманливим. У деяких випадках висока точність може вказувати на таку проблему, як перенавчання моделі. Розраховується метрика за формулою:

$$Accuracy = \frac{TruePositive+TrueNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative}$$

Обрахунок точності (Accuracy) для моделі прогнозування рекламного доходу має важливе значення, оскільки він дозволяє оцінити загальну здатність моделі правильно класифікувати випадки, в яких реклама буде або не буде приносити дохід. Це особливо важливо у контексті ефективного розподілу рекламного бюджету та оптимізації рекламних стратегій. Точність допомагає визначити, наскільки добре модель виконує свої завдання у загальному вимірі. Високі показники свідчать про те, що модель ефективно відрізняє прибуткові та неприбуткові рекламні сесії, що є критично важливим для розуміння загальної продуктивності кампаній. Accuracy надає чіткі показники для ухвалення рішень на основі аналізу даних, які в подальшому можна використовувати для коригування стратегій розміщення реклами або зміни таргетингу, щоб збільшити доходи від рекламних кампаній.

• *F1-score*. F1-оцінка є гармонічним середнім між точністю і повнотою, і часто використовується для визначення загальної ефективності моделі при роботі з незбалансованими датасетами, де один з класів може мати більшу представленість за інший. Ця оцінка допомагає збалансувати потребу в зменшенні хибно позитивних і хибно негативних результатів і є корисною, коли важливі обидва типи помилок. Високі значення вказують на те, що модель

має як високу точність, так і високу повноту, що свідчить про її ефективність в ідентифікації позитивних класів з мінімальними помилками. Оцінка F1 більш чутлива до екстремальних варіантів і досягає свого максимуму, коли точність (Precision) і повнота (Recall) є рівними. F1-оцінка може виявитися корисною метрикою в наступних ситуаціях, коли хибно позитивні (FP) і хибно негативні (FN) результати мають однакову вагу, і коли збільшення кількості даних не змінює результат, а кількість істинно негативних випадків (True Negative) залишається високою. Формула для розрахунку може виглядати наступним чином:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Обрахунок F1-оцінки для моделі прогнозування рекламного доходу має важливе значення, оскільки ця метрика забезпечує збалансоване оцінювання як точності (Precision), так і повноти (Recall) моделі. У контексті рекламних кампаній, де важливо не лише ідентифікувати потенційно прибуткові можливості, а й уникнути зайвих витрат на нецільові аудиторії, метрика допомагає визначити, наскільки ефективно модель виконує обидва завдання. Модель із високою F1-оцінкою забезпечує ефективне використання рекламного бюджету, мінімізуючи витрати на неефективні рекламні покази, що не приносять доходу, і водночас максимізуючи охоплення потенційно дохідних можливостей.

• *AUC-ROC.* Area Under the Curve - Receiver Operating Characteristics є однією з ключових метрик у машинному навчанні для оцінювання якості бінарних класифікаторів. Ця метрика забезпечує інтегральне розуміння того, наскільки добре модель може розрізняти між двома класами — позитивним (цільовим) і негативним. Характеристика оператора приймача (ROC) представляє собою ймовірнісну криву, яка демонструє залежність частоти істинно позитивних результатів ($TPR\ (Sensitivity) = \frac{TruePositive}{TruPositive + FalseNegative}$) від частоти хибно

позитивних результатів ($FPR = 1 - Specificity = \frac{FalsePositive}{TrueNegative+FalsePositive}$) за різними порогами, допомагаючи розрізнити «сигнал» від «шуму». Площа під кривою (AUC) вимірює здатність класифікатора розрізняти між класами [35]. На рис. Б.1 додатку Б наведений графік площі між кривою та осями X та Y. Із зазначеного графіка очевидно, що вищі значення AUC вказують на кращу продуктивність моделі при різних порогах між позитивними та негативними класами. AUC — це площа під ROC-кривою. Це числове значення від 0 до 1, де:

- AUC = 1 вказує на ідеальну модель, яка досконало розрізняє між класами;
- AUC = 0.5 означає, що модель не має дискримінаційної здатності, тобто вона працює на рівні випадкових здогадок;
- AUC < 0.5 вказує на модель зі зворотними передбаченнями, що є гірше, ніж випадкове вгадування.

AUC-ROC не залежить від конкретного порогу, що використовується для класифікації позитивних чи негативних класів, що робить її корисною для оцінки загальної якості моделі. ROC-крива та площа під нею забезпечують стійкі оцінки навіть для датасетів, де кількість прикладів одного класу значно переважає інший. Обчислення AUC-ROC для моделі прогнозування рекламного доходу є особливо важливим, оскільки ця метрика надає цінне розуміння загальної здатності моделі розрізняти між класами. У контексті рекламного доходу, модель повинна ефективно ідентифікувати випадки, що приносять дохід (позитивні), та випадки, що не приносять дохід (негативні). В порівнянні з іншими метриками, які можуть залежати від специфічного порогового значення, AUC-ROC надає узагальнене оцінювання продуктивності моделі. Це особливо корисно в умовах, коли складно визначити фіксований поріг для класифікації [54].

Отже, методи перевірки точності та оцінювання моделей є критично важливими в процесі розробки будь-якої аналітичної моделі, особливо в контексті прогнозування. Ці методи забезпечують засоби для вимірювання і порівняння ефективності моделей перед тим, як використовувати їх у реальних застосуваннях.

Висновки до розділу 2

Інструменти Data Science, такі як машинне навчання, статистичний аналіз та обробка даних, відіграють важливу роль у зборі, аналізі та інтерпретації даних для бізнес-аналітики. Використання цих інструментів дозволяє компаніям отримувати значні переваги через здатність виявляти тенденції, передбачати економічні показники і покращувати стратегічне планування. Моделі, зокрема логістична регресія, випадковий ліс та алгоритми бустингу, дозволяють ефективно аналізувати великі обсяги даних і знаходити значущі залежності та прогнозувати майбутні доходи від рекламних кампаній.

Процес моделювання для прогнозування доходу рекламного нетворку починається зі збору та підготовки даних, включаючи очищення та структурування інформації. Наступним кроком є вибір і тренування моделі, залежно від особливостей даних та бізнес-завдань. Після тренування, модель оптимізується і валідується через тестування на нових даних, щоб забезпечити її точність і надійність перед використанням у реальних умовах. Точність прогнозів, отриманих з використанням моделей Data Science, перевіряється через різні методи, зокрема, перехресну валідацію та тестування на окремій вибірці. Метрики, як-от точність, відгук, F1-схожість і AUC-ROC, забезпечують кількісну оцінку ефективності моделей. Застосування цих методів дозволяє ідентифікувати і усунути можливі недоліки в моделях, забезпечуючи більш високу надійність та точність прогнозів.

Використання інструментів Data Science в моделюванні для прогнозування доходу рекламних нетворків виявляється ефективним способом підвищення продуктивності та оптимізації маркетингових витрат. Процес моделювання вимагає

ретельного аналізу та правильного використання статистичних методів і машинного навчання. Правильно побудовані та перевірені моделі можуть значно підвищити точність бізнес-прогнозів, що є ключем до успішного рішення стратегічних завдань.

РОЗДІЛ 3. ПОБУДОВА МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ ДОХОДУ РЕКЛАМНОЇ КОМПАНІЇ

3.1. Характеристика основних факторів для моделювання

У попередніх підрозділах дослідження були розглянуті основні аспекти рекламної галузі, вплив на її основні показники, характеристики моделей аналізу даних та машинного навчання, тому далі вже можна переходити до безпосереднього моделювання та прогнозування доходу. Для цього будуть використані окремі моделі машинного навчання. Проте в першу чергу необхідно визначити фактори для побудови моделі та описати їх.

Для побудови моделі прогнозування рекламного доходу необхідно мати історичні дані та мати розуміння, що означає кожна змінна. У роботі використовується реальні дані з внутрішньої бази компанії, тому для збереження конфіденційності за правилами NDA їхні значення були модифіковані. Дані у вигляді статистики по кожному рекламному партнеру та сабсорсу трафіку, який він використовує, та містять інформацію про кількість івентів, що призводять до отримання доходу.

Опис даних:

- Джерело: внутрішня база даних рекламної компанії.
- Період збору: 2024-01-01 - 2024-01-10
- Кількість спостережень: 30,764,006
- Кількість змінних: 21
- Опис змінних:
    - date - дата (тип даних: object, рекомендовано конвертувати в datetime): дата події або запису даних;
    - partner_id - ідентифікатор партнера (тип даних: int64): унікальний номер, що ідентифікує партнера;
    - subsource_id - ідентифікатор джерела (тип даних: int64): унікальний номер, що ідентифікує джерело або субджерело трафіку;

- current_LT - поточний LT (lifetime) (тип даних: int64): позначає час життя користувача на продукты на конкретну дату;
- visits - візити (тип даних: int64): кількість візитів;
- is_react - реакція (тип даних: int64): бінарна змінна, яка показує наявність реакції;
- is_webview - веб-перегляд (тип даних: int64): бінарна змінна, що вказує чи була активність в веб-перегляді;
- is_rtb_group - група RTB (тип даних: int64): показує приналежність до групи аукціонів реального часу;
- en_language - мова (тип даних: int64): визначення мови, чи це трафік для англомовних країн;
- impressions - покази (тип даних: int64): кількість показів реклами або контенту;
- clicks - кліки (тип даних: int64): кількість кліків по рекламним банерам;
- unq_clicks - унікальні кліки (тип даних: int64): кількість унікальних кліків;
- cpa_revenue - дохід CPA (тип даних: int64): дохід від певної дії за ціновою моделлю «Cost Per Action»;
- redirects - переадресації (тип даних: int64): кількість переадресацій зі сторінки;
- autoredirects - автоматичні переадресації (тип даних: int64): кількість автоматичних переадресацій;
- messages - повідомлення (тип даних: int64): кількість повідомлень;
- likes - лайки (тип даних: int64): кількість лайків;
- skips - пропуски (тип даних: int64): кількість пропущених подій або дій;
- has_push_click - клік по push-повідомленню (тип даних: int64): чи був клік по push-повідомленню;

- has_mail_visit - візит по email (тип даних: int64): чи був візит з електронної пошти;
- revenue - дохід (тип даних: float64): загальний дохід компанії.

Далі в ході виконання практичної частини досладження буде визначено за допомогою статистичних методів, наскільки дані є консистентними та чи всі фактори варто використовувати при побудові моделей в залежності від їхніх значень. Також можна зазначити основні бібліотеки, які будуть використані при моделюванні:

- pandas – застосовується для обробки даних.
- numpy – використовується для маніпуляцій з багатовимірними масивами та матрицями.
- sklearn.ensemble – модуль для роботи з ансамблями дерев рішень.
- sklearn.tree – модуль для маніпуляцій з деревами рішень.
- sklearn.model_selection – модуль для різних операцій з даними, включаючи їх розбиття на тренувальні та тестові набори, крос-валідацію та налаштування гіперпараметрів.
- sklearn.metrics – модуль, що надає функції для оцінки ефективності моделей.
- xgboost – модуль для градієнтного бустінгу.
- catboost – оптимізований для категоріальних ознак модуль градієнтного бустінгу.
- random – модуль для генерації псевдовипадкових чисел та виконання різноманітних випадкових операцій.

Отже, розглянувши структуру даних датасету рекламних показників та описавши основні бібліотеки, що будуть застосовані при моделюванні, можна вже безпосередньо переходити до аналізу взаємозв’язків між факторами, побудови моделей та їх оцінки.

3.2. Моделювання доходу та аналіз результатів

Раніше був розглянутий набір даних із показниками рекламного доходу та інших маркетингових метрик, що мають на нього безпосередній вплив. Ключовим етапом у процесі обробки даних є ідентифікація та вилучення ознак, які не відображають загальний характер даних. Наявність високої кореляції між змінними може свідчити про зайву або повторювану інформацію. В таких випадках рекомендується усунути одну з корелюючих змінних.

Кореляційна матриця служить інструментом для виявлення таких ознак. Вона вимірює ступінь взаємозв'язку між декількома змінними у наборі даних. Змінні, що демонструють значну кореляцію, можуть вказувати на наявність надмірної або дублюючої інформації, і в такому разі одну з корелюючих змінних потрібно вилучити.

Застосування кореляційної матриці при побудові моделей для виявлення зайвих змінних має кілька важливих переваг, які покращують якість та ефективність аналітичного процесу. Першою з них є те, що кореляційна матриця допомагає ідентифікувати високу кореляцію між змінними, що є ознакою мультиколінеарності. Мультиколінеарність може призвести до ненадійних та нестабільних оцінок коефіцієнтів у регресійних моделях, а її виявлення дозволяє підвищити точність моделі. Також, коли модель використовує менше змінних, ризик перенавчання знижується. Перенавчання відбувається, коли модель надто добре адаптується до тренувальних даних і погано прогнозує на нових даних. Кореляційна матриця допомагає вибрати тільки ті змінні, які дійсно важливі, що підсилює здатність моделі генералізувати. Видалення зайвих змінних допомагає уникнути проблем зі стабільністю в моделях, коли невеликі зміни в даних можуть призвести до великих змін у прогнозах. Стабільність моделі є ключовим фактором для її надійності і точності. Тому, застосування кореляційної матриці є ефективним інструментом для підготовки даних перед моделюванням, що забезпечує більшу чистоту, точність та ефективність аналітичного процесу.

На основі всіх факторів, що були присутні у наборі даних була побудована кореляційна матриця, представлена нижче на рис. 3.1. Значення кореляції варіюються від -1 до +1, де значення близьке до +1 вказує на сильну позитивну кореляцію, значення близьке до -1 вказує на сильну негативну кореляцію, а значення близьке до 0 означає відсутність лінійної кореляції.

	partner_id	subsource_id	current_LT	visits	is_react	is_webview	is_rtb_group	en_language	impressions	clicks	unq_clicks	cpa_revenue	redirects	autoredirects	messages	likes	skips	has_push_click	has_mail_visit	revenue
partner_id	1.00	0.21	-0.00	-0.00	0.01	0.00	-0.01	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.01	-0.00
subsource_id	0.21	1.00	-0.00	-0.04	-0.05	-0.02	-0.04	-0.04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02	-0.03	-0.01
current_LT	-0.00	-0.00	1.00	0.00	0.00	-0.00	0.00	0.00	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02	-0.01	-0.01	-0.01	-0.04	-0.03	-0.02
visits	-0.00	-0.04	0.00	1.00	0.55	0.27	0.99	0.99	0.12	0.07	0.07	0.03	0.05	0.07	0.13	0.09	0.10	0.17	0.09	0.04
is_react	0.01	-0.05	0.00	0.55	1.00	0.34	0.52	0.53	0.09	0.10	0.10	0.06	0.09	0.11	0.09	0.09	0.06	0.29	0.18	0.07
is_webview	0.00	-0.02	-0.00	0.27	0.34	1.00	0.23	0.27	0.05	0.07	0.07	0.02	0.06	0.05	0.05	0.04	0.04	0.14	0.10	0.03
is_rtb_group	-0.01	-0.04	0.00	0.99	0.52	0.23	1.00	0.99	0.12	0.06	0.06	0.03	0.05	0.07	0.13	0.09	0.10	0.15	0.08	0.04
en_language	-0.01	-0.04	0.00	0.99	0.53	0.27	0.99	1.00	0.12	0.07	0.07	0.03	0.05	0.07	0.13	0.09	0.09	0.16	0.09	0.04
impressions	-0.00	-0.01	-0.02	0.12	0.09	0.05	0.12	0.12	1.00	0.71	0.71	0.39	0.62	0.61	0.96	0.78	0.85	0.11	0.31	0.45
clicks	-0.00	-0.01	-0.02	0.07	0.10	0.07	0.06	0.07	0.71	1.00	1.00	0.51	0.93	0.91	0.57	0.63	0.51	0.14	0.42	0.58
unq_clicks	-0.00	-0.01	-0.02	0.07	0.10	0.07	0.06	0.07	0.71	1.00	1.00	0.51	0.93	0.91	0.57	0.63	0.52	0.14	0.43	0.59
cpa_revenue	-0.00	-0.01	-0.01	0.03	0.06	0.02	0.03	0.03	0.39	0.51	0.51	1.00	0.66	0.57	0.34	0.41	0.28	0.07	0.52	0.95
redirects	-0.00	-0.01	-0.02	0.05	0.09	0.06	0.05	0.05	0.62	0.93	0.93	0.66	1.00	0.90	0.51	0.60	0.44	0.13	0.53	0.71
autoredirects	-0.00	-0.01	-0.02	0.07	0.11	0.05	0.07	0.07	0.61	0.91	0.91	0.57	0.90	1.00	0.48	0.57	0.42	0.15	0.41	0.66
messages	-0.00	-0.01	-0.01	0.13	0.09	0.05	0.13	0.13	0.96	0.57	0.57	0.34	0.51	0.48	1.00	0.77	0.85	0.11	0.28	0.38
likes	-0.00	-0.01	-0.01	0.09	0.09	0.04	0.09	0.09	0.78	0.63	0.63	0.41	0.60	0.57	0.77	1.00	0.73	0.09	0.31	0.46
skips	-0.00	-0.01	-0.01	0.10	0.06	0.04	0.10	0.09	0.85	0.51	0.52	0.28	0.44	0.42	0.85	0.73	1.00	0.07	0.21	0.32
has_push_click	0.00	-0.02	-0.04	0.17	0.29	0.14	0.15	0.16	0.11	0.14	0.14	0.07	0.13	0.15	0.11	0.09	0.07	1.00	0.21	0.09
has_mail_visit	-0.01	-0.03	-0.03	0.09	0.18	0.10	0.08	0.09	0.31	0.42	0.43	0.52	0.53	0.41	0.28	0.31	0.21	0.21	1.00	0.53
revenue	-0.00	-0.01	-0.02	0.04	0.07	0.03	0.04	0.04	0.45	0.58	0.59	0.95	0.71	0.66	0.38	0.46	0.32	0.09	0.53	1.00

Рис. 3.1. Кореляційна матриця на основі початкового набору даних.

Джерело: розроблено автором.

Оцінюючи значення кореляції для *revenue* (рекламного доходу), можна виділити декілька ключових спостережень:

1. Змінні з вищою кореляцією:

   a. ‘*cpa_revenue*’ має найвищу кореляцію з доходом, з значенням 0.95. Це вказує на тісний зв'язок між ‘*cpa_revenue*’ та загальним доходом, що може свідчити про те, що ‘*cpa_revenue*’ є значущим предиктором для ‘revenue’.

   b. ‘*has_mail_visit*’ має кореляцію 0.53 з ‘*revenue’*, що також є значимим, але менш сильним зв'язком.

2. Змінні з помірною кореляцією:

   a. ‘*clicks’* та ‘unq_clicks’ мають кореляцію з доходом на рівні 0.58 та 0.59 відповідно, що вказує на позитивний вплив цих змінних на доходи.

b. ‘*redirects’* та ‘*autoredirects’* мають кореляції з ‘*revenue’* на рівні 0.71 та 0.66, відповідно, вказуючи на можливу значимість цих ознак у генерації доходу.

3. Змінні з низькою або мінімальною кореляцією:

a. Багато змінних, як наприклад ‘*partner_id’*, ‘*subsource_id’*, ‘*current_LT’*, мають дуже низьку кореляцію з доходом, що вказує на їх малу або відсутню роль у прогнозуванні доходів.

Ці результати можуть допомогти визначити, які змінні варто включити в подальші аналітичні моделі для прогнозування доходу. Зокрема, змінні з високою кореляцією з доходом, можуть бути основними кандидатами для включення до моделі, тоді як змінні з дуже низькою кореляцією можуть бути виключені для спрощення моделі та зменшення ризику колінеарності.

Враховуючи, що змінні ‘*visits’*, ‘*is_react*’, ‘*is_webview*’, ‘*is_rtb_group*’ та ‘*en_language*’ мають достатньо низьку кореляцію з цільовою метрикою доходу, було прийнято рішення виключити їх із подальшого моделювання. Тому кореляційна матриця факторів для побудови моделі після підготовчого етапу даних представлена на рис. 3.2.

	partner_id	subsource_id	current_LT	impressions	clicks	unq_clicks	cpa_revenue	redirects	autoredirects	messages	likes	skips	has_mail_visit	revenue
partner_id	1.00	0.21	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.00
subsource_id	0.21	1.00	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.01
current_LT	-0.00	-0.00	1.00	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02	-0.01	-0.01	-0.01	-0.03	-0.02
impressions	-0.00	-0.01	-0.02	1.00	0.71	0.71	0.39	0.62	0.61	0.96	0.78	0.85	0.31	0.45
clicks	-0.00	-0.01	-0.02	0.71	1.00	1.00	0.51	0.93	0.91	0.57	0.63	0.51	0.42	0.58
unq_clicks	-0.00	-0.01	-0.02	0.71	1.00	1.00	0.51	0.93	0.91	0.57	0.63	0.52	0.43	0.59
cpa_revenue	-0.00	-0.01	-0.01	0.39	0.51	0.51	1.00	0.66	0.57	0.34	0.41	0.28	0.52	0.95
redirects	-0.00	-0.01	-0.02	0.62	0.93	0.93	0.66	1.00	0.90	0.51	0.60	0.44	0.53	0.71
autoredirects	-0.00	-0.01	-0.02	0.61	0.91	0.91	0.57	0.90	1.00	0.48	0.57	0.42	0.41	0.66
messages	-0.00	-0.01	-0.01	0.96	0.57	0.57	0.34	0.51	0.48	1.00	0.77	0.85	0.28	0.38
likes	-0.00	-0.01	-0.01	0.78	0.63	0.63	0.41	0.60	0.57	0.77	1.00	0.73	0.31	0.46
skips	-0.00	-0.01	-0.01	0.85	0.51	0.52	0.28	0.44	0.42	0.85	0.73	1.00	0.21	0.32
has_mail_visit	-0.01	-0.03	-0.03	0.31	0.42	0.43	0.52	0.53	0.41	0.28	0.31	0.21	1.00	0.53
revenue	-0.00	-0.01	-0.02	0.45	0.58	0.59	0.95	0.71	0.66	0.38	0.46	0.32	0.53	1.00

Рис. 3.2. Кореляційна матриця на після видалення зайвих факторів набору даних.
Джерело: розроблено автором.

Було розглянуто декілька моделей машинного навчання на базі початкового набору даних для того, щоб обрати найбільш точну та підходящу для подальшого

моделювання, навчання та покращення. У табл. 3.1 наведено перелік проаналізованих моделей з показниками оцінки їхньої якості:

Таблиця 3.1

Результати якості моделей прогнозування

Метрика якості моделі	MAE	MSE	R^2	RMSE	MAPE
Linear regression	0.120928	0.802393	0.650822	0.892311	0.359451
Lasso-regression	0.100923	0.820824	0.670028	0.913423	0.300982
Random forest	0.009012	0.653482	0.712387	0.810044	0.278913
XGBoost	0. 006274	0.542250	0.750907	0.736377	0.243594
CatBoost	0.008239	0.602398	0.729833	0.771248	0.230012

Джерело: складено автором на основі власних розрахунків.

Для подальшого моделювання була обрана модель XGBRegressor, так як вона має найнижчі показники по MAE, MSE, RMSE та MAPE, а також найвищий R^2.

XGBoost ефективно обробляє великі обсяги даних і характеризується своєю високою точністю в різних задачах прогнозування, що робить його ідеальним для комплексних наборів даних. Оскільки кореляційна матриця вказує на наявність значних взаємозв'язків між деякими змінними і доходом, XGBoost може використовувати цю інформацію для покращення точності прогнозу. Також, XGBoost добре підходить для ситуацій, коли дані містять пропущені значення або коли в наборі даних є велика кількість ознак, які можуть впливати на результат. Він автоматично обробляє пропущені дані та забезпечує вбудовану перехресну валідацію, що дозволяє краще налаштувати модель і забезпечити її узагальнювану здатність. Однією з ключових переваг XGBoost є його здатність до регуляризації, яка допомагає запобігти перенавчанню моделі. Це особливо важливо в бізнес-задачах, де модель повинна бути стійкою до різних випадковостей у даних. Крім того, важливість ознак, яку надає XGBoost, може бути використана для розуміння, які змінні найбільше впливають на доходи, дозволяючи більш цілеспрямовано оптимізувати маркетингові стратегії та бізнес-процеси. Тому, виходячи з усіх

основних ознак моделі та характеристики її використання, можна сказати, що вона є оптимальною для прогнозування та підходить для даної проблематики дослідження та даних.

Була зроблена перша ітерація моделювання без підбору параметрів (*baseline model*), щоб оцінити загальні показники моделі. Проста модель без підбору гіперпараметрів може виявити, чи є базові алгоритми адекватними для розв'язання задачі з наявними даними. Якщо така базова модель показує неприйнятну продуктивність, це може вказувати на потребу в більш складних методах або на необхідність перегляду вхідних даних. Початкове моделювання допомагає виявити можливі проблеми з даними, такі як неправильна обробка, потреба в інженерії ознак, аномалії та викиди. Це дозволяє здійснити необхідні корективи перед тим, як витрачати час на тонке налаштування моделі. Використання базової моделі як вихідної точки дає змогу метрично оцінити вигоду від подальших удосконалень і допомагає у виборі наступних кроків у процесі розробки моделі. Це допомагає зосередитись на тих аспектах моделювання, які дійсно можуть призвести до значного покращення продуктивності.

В рамках першої ітерації для моделі XGBoost без тюнінгу гіперпараметрів отримали наступні дані:

Таблиця 3.2

Результати якості базової моделі XGBoost

Метрика якості моделі	Результати моделювання
Mean Absolute Error (MAE)	0.006274
Mean Squared Error (MSE)	0.542250
R-squared (R^2)	0.750907
Root Mean Squared Error (RMSE)	0.736377
Mean Absolute Percentage Error (MAPE)	0.243594

Джерело: складено автором на основі власних розрахунків.

Метрика MAE 0.006274 свідчить про те, що середні помилки моделі є дуже низькими, що є позитивним показником, оскільки вказує на малі помилки в абсолютних вимірах. Значення MSE 0.542250, яке вище за MAE, вказує на наявність декількох більших помилок у даних, хоча в цілому помилки залишаються відносно низькими. Показник R^2 0.750907 свідчить про те, що близько 75% варіативності доходу можна пояснити за допомогою вхідних ознак у моделі. Це високий показник для R^2, що говорить про добру адаптацію моделі до даних. Значення RMSE 0.736377 вказує на те, що стандартне відхилення помилок моделі від фактичних значень є помірним, що дозволяє краще оцінити розподіл помилок. Результати MAPE 0.243594 (24.36%) можуть вважатися прийнятним для багатьох додатків, але також свідчить про наявність значного розкиду в прогнозах по відношенню до фактичних значень.

Ці результати показують, що модель XGBRegressor без тюнінгу гіперпараметрів вже показує гарну продуктивність і має потенціал для подальшого покращення, якщо провести налаштування гіперпараметрів. Значення R^2 в 75% є особливо обнадійливим, оскільки воно вказує на те, що модель ефективно пояснює більшу частину варіативності відповідних даних.

Для покращення результатів можна виконати підбір параметрів моделі за допомогою RandomizedSearchCV - метод з бібліотеки Scikit-learn, який використовується для автоматизації процесу підбору гіперпараметрів моделі машинного навчання. Відмінно від традиційного пошуку по сітці (GridSearchCV), який систематично перевіряє всі комбінації параметрів, RandomizedSearchCV вибирає випадкові комбінації параметрів для випробувань, що робить процес пошуку швидшим та часто більш ефективним, особливо при великій кількості гіперпараметрів і/або коли деякі параметри впливають на продуктивність моделі непропорційно більше за інші. За допомогою отриманих параметрів ***XGBRegressor (learning_rate=0.09266162953207084, max_depth=4, n_estimators=173,***

subsample=0.517032198914886) знову відбувається навчання моделі. Після цього отримали наступні результати:

Таблиця 3.3

Результати якості моделі XGBoost із додаванням гіперпараметрів

Метрика якості моделі	Результати моделювання
Mean Absolute Error (MAE)	0.009832
Mean Squared Error (MSE)	0.302280
R-squared (R^2)	0.861141
Root Mean Squared Error (RMSE)	0.549800
Mean Absolute Percentage Error (MAPE)	0.669809

Джерело: складено автором на основі власних розрахунків.

У порівнянні з попередніми показниками якості значення MAE збільшилось з 0.006274 до 0.009833. Це означає, що середня абсолютна помилка зросла, що може вказувати на збільшення помилок у деяких прогнозах. Збільшення MAE може бути результатом компромісу заради покращення інших метрик, таких як R^2. MSE значно знизилась з 0.542250 до 0.302281, що свідчить про зменшення великих помилок в прогнозах. Це важлива зміна, оскільки MSE більш чутливий до великих помилок, ніж MAE. R^2 значно зріс з 0.750907 до 0.861142, що є значним покращенням. Це показує, що модель тепер пояснює значно більшу частину варіативності цільової змінної, і є більш точною та ефективною у загальному контексті. Зниження RMSE з 0.736377 до 0.549800 також позитивно, оскільки нижче значення RMSE означає менші квадратичні відхилення прогнозів від реальних значень. MAPE зросло з 0.243594 до 0.669810, що показує збільшення середнього відсоткового відхилення між прогнозами та дійсними значеннями. Це може вказувати на те, що після тюнінгу, хоча модель і стала краще загалом, вона може бути менш точною при особливо малих значеннях даних.

Покращення в MSE та R^2 після тюнінгу гіперпараметрів свідчать про значне покращення загальної продуктивності моделі, хоча певне збільшення в MAE і

MAPE вказує на необхідність додаткової уваги до деталей у певних аспектах прогнозування. Загалом, модель стала більш ефективною у поясненні варіативності доходу, що є ключовим для більшості аналітичних застосувань.

3.3. Оцінка отриманих результатів та надання рекомендацій щодо покращення діяльності рекламного підприємства

Аналіз та результати моделювання рекламного доходу з використанням XGBRegressor дали змогу отримати важливі висновки та виявити можливості для оптимізації рекламної мережі. Спочатку модель була налаштована без тюнінгу гіперпараметрів, де вона показала задовільні результати, з R^2 на рівні 0.7509, що свідчило про здатність моделі пояснювати близько 75% варіативності доходу. Однак, після налаштування гіперпараметрів, продуктивність моделі значно покращилася, з R^2, що зріс до 0.8611, вказуючи на більш високу точність прогнозів.

Незважаючи на покращення в загальній продуктивності, модель показала деяке збільшення середньої абсолютної помилки (MAE) та середньої абсолютної відсоткової помилки (MAPE), що може вказувати на більшу помилку в окремих випадках прогнозування. Це збільшення помилок може бути пов'язане з перенавчанням або неоптимальним вибором гіперпараметрів для конкретних аспектів даних.

На основі отриманих результатів, рекомендації для покращення діяльності рекламного нетворку можуть включати кілька напрямків. Перш за все, важливо продовжувати моніторинг та аналіз впливу різних ознак на доходи для подальшої оптимізації моделі. Це включає регулярне оновлення даних та перевірку на нових даних для забезпечення стабільності та надійності прогнозів.

Також, рекомендується розглянути можливість додаткового тюнінгу гіперпараметрів з використанням більш просунутих методів, таких як байєсовська оптимізація, для знаходження балансу між продуктивністю та надійністю моделі. Важливо також враховувати нові змінні або модифікувати існуючі ознаки для

поліпшення точності прогнозів, що може включати більш глибокий аналіз поведінки користувачів та реакції на рекламу. Важливо проаналізувати, які ознаки найбільше впливають на доходи, і зосередити увагу на їхній оптимізації та можливому розширенні. Наприклад, можна досліджувати нові типи ознак, такі як географічні дані або поведінкові фактори користувачів, для того, щоб визначити якісь паттерни та впроваджувати для окремих сегментів окремі стратегії показу реклами для більшої залученості

Заохочення інновацій та тестування нових підходів у керуванні рекламними кампаніями можуть допомогти підвищити ефективність рекламних ініціатив та, відповідно, збільшити доходи. Крім того, активне використання аналітичних звітів та візуалізацій допоможе краще розуміти тенденції та візуально оцінювати результати різних рекламних стратегій. Оптимізація рекламного нетворку вимагає комплексного підходу, що включає як технічні аспекти моделювання та аналізу даних, так і стратегічне планування на основі отриманих інсайтів. Вдосконалення прогнозування рекламного доходу сприятиме кращому розумінню та оптимізації рекламних кампаній, підвищуючи їх ефективність і прибутковість.

Висновки до розділу 3

Характеристика та визначення основних аспектів впливу на маркетинговий дохід та побудова кореляційної матриці допомогли визначити, які фактори варто включати до моделі. Ці фактори допомагають визначити потенційний прибуток від рекламних кампаній та оптимізувати витрати.

Моделювання доходу проводиться шляхом створення прогностичних моделей, які базуються на згаданих індикаторах, для оцінки можливих доходів від рекламної діяльності. Результати моделювання після їх отримання аналізуються з метою виявлення ключових драйверів доходу та витрат. Важливим є порівняння прогнозованих результатів з фактичними, що допомагає зрозуміти ефективність поточних стратегій.

Аналіз отриманих результатів включає вивчення відповідності прогнозів реальним показникам та виявлення областей для покращення. На основі аналізу можуть бути надані рекомендації щодо коригування рекламних бюджетів, цільових аудиторій або рекламних повідомлень для підвищення рентабельності інвестицій. Також важливо розглянути технологічні інновації та можливості автоматизації для оптимізації рекламних процесів..

ВИСНОВКИ

В умовах зростаючої конкуренції у сфері реклами, здатність точно прогнозувати фінансові показники стає ключовою для ефективної діяльності рекламних компаній. Завдяки прискореному прогресу в області технологій аналізу даних, інструменти Data Science створюють нові шляхи для вдосконалення маркетингових тактик та підвищення результативності рекламних проєктів. Використання аналітики та машинного навчання не лише дозволяє вірогідно прогнозувати дохід, а й ідентифіковувати найбільш результативні рекламні канали, розуміти споживчу поведінку і тим самим раціонально розподіляти бюджети.

Власне, за мету було визначено аналіз впливу інструментів Data Science на ефективність і дохідність рекламних кампаній, побудова моделей для прогнозування та розробка рекомендацій для оптимізації діяльності рекламного нетворку. У ході виконання визначених завдань для досягнення цієї мети були сформульовані наступні висновки.

1.1. Визначено, що аналіз діяльності рекламного сектору та окремих рекламних компаній стає ключовим елементом сучасних маркетингових стратегій і економічного аналізу. Рекламна індустрія вносить вагомий внесок в економічне зростання через збільшення обсягів продажів та популяризацію брендів, водночас формуючи суспільні цінності, та відіграє значну роль на загальноекономічному рівні. Таким чином, розуміння специфіки різних типів рекламних компаній дозволяє вибирати найбільш ефективні стратегії для оптимізації їхньої діяльності та оптимізації рекламного сектору загалом.

1.2. Розглянуто ключові метрик, що оцінюють ефективність рекламних стратегій, що є критично необхідним для компаній, які прагнуть поліпшити свої маркетингові активності та збільшити прибутковість вкладень. Зроблено висновок, що показники допомагають оцінити результативність рекламних кампаній, визначити ефективні та неефективні стратегії. Вони служать маркерами досягнення бізнес-цілей та дозволяють планувати на основі даних,

що знижує ризики і підвищує шанси на успіх. Аналіз цих даних і прогнозування майбутніх трендів дозволяє компаніям завжди бути на крок попереду конкурентів.

1.3. Показано, що на рекламну галузь, маркетингові та фінансові показники безпосередньо впливають прямі та опосередковані чинники, кожен з яких певним чином змінює динаміку та ефективність діяльності компаній. До таких факторів можна віднести як зміну споживчої поведінки внаслідок діджиталізації та зовнішніх факторів впливу, так і макроекономічні умови, бо економічні коливання, такі як рецесія, інфляція та зміни у споживчих витратах, мають велике значення для цифрового маркетингу. Тому, важливо розуміти, що моніторинг, аналіз та оптимізація показників на основі виявлених трендів і змін є ключовими для підтримання конкурентоспроможності та досягнення максимальної віддачі від інвестицій у рекламні кампанії.

2.1. Здійснено огляд основних статистичних моделей та машинного навчання, що мають застосування в різних галузях економіки та дозволяють адекватно обробляти великі обсяги даних і визначати ключові фактори, що впливають на доходи. До таких моделей можна віднести регресії, дерева рішень та бустинги. Визначено, що кожна з них має свої особливості застосування для моделювання доходу в залежності від набору та представлення даних, а також мети прогнозування, тому для рекламного сектору правильний вибір та налаштування моделі, а також здатність адаптувати моделі під змінні умови ринку, відіграють ключову роль у забезпеченні успіху рекламних кампаній.

2.2. Визначено основні етапи моделювання для побудови прогнозу рекламного доходу, а саме визначення та вигрузка даних за певний період, їх аналіз, перевірка на консистентність та якість, в разі необхідності – очистка даних або їх корекція у випадку невідповідності; трансформація даних та розділення їх на навчальний, валідаційний та тестовий набори; вибір алгоритму та на основі найбільш точної моделі проводиться її навчання, тестування та оцінка.

Ефективне виконання всіх етапів моделювання не лише підвищує точність прогнозів доходів, але й сприяє кращому розумінню динаміки ринку, що є важливим для стратегічного планування та прийняття обґрунтованих комерційних рішень.

2.3. Розглянуто основні метрики оцінки якості та точності побудованих моделей. Вони не тільки дозволяють оцінити ефективність моделі в прогнозуванні результатів, але й забезпечують можливість порівняння різних моделей та вибору найкращої. З їхньою допомогою можна визначити, як добре модель працює з реальними даними, ідентифікувати потенційні недоліки в моделі, та оптимізувати її для кращої продуктивності. Таким чином, метрики точності виступають як фундаментальний інструмент у розробці надійних та ефективних моделей прогнозування.

3.1. Практична частина роботи виконувалася за допомогою ПЗ Jupyter Notebook та Python. Для аналізу були використані дані внутрішньої бази рекламної компанії за період 2024-01-01 - 2024-01-10 та 21 змінною. Додатково до цього, були визначені бібліотеки програмного продукту, що використовувались під час моделювання. Аналіз цих аспектів дав розуміння про структуру даних, необхідність їх трансформації або додаткового інженірингу, а також попередньо дав можливість сформулювати гіпотезу – чи вплив всіх наявних маркетингових показників безпосередньо впливає на рекламний дохід.

3.2. Відповідно до визначених етапів моделювання було побудовано декілька моделей, зокрема лінійна та ласо-регресія, випадковий ліс та бустинги, і оцінена їхня точність. Для побудови моделей використовувалися показники, що були попередньо відібрані на основі кореляційного аналізу та тісноти зв’язку із рекламним доходом. Для прогнозування була обрана модель XGBRegressor на основі найкращих показників якості із проаналізованих моделей. Її точність була оцінена як при першій ітерації, без тюнінгу, так і з

додаванням до моделі гіперпараметрів на основі методу RandomizedSearchCV та при перенавчанні. В цілому, покращення моделі після додавання гіперпараметрів є свідченням того, що процес налаштування моделі став більш адаптованим і цілеспрямованим, відповідаючи специфічним потребам даних і задачі.

3.3. На основі отриманих результатів, рекомендації для покращення діяльності рекламного нетворку можуть включати кілька напрямків. Особливо важливо продовжувати моніторинг та аналіз впливу різних ознак на доходи для подальшої оптимізації моделі. Це включає в себе регулярне оновлення даних та перевірку на нових даних для забезпечення стабільності та надійності прогнозів. Також, враховуючи гарну кореляцію реакції на поштові листи та кількості перенаправлень з сайту з метрикою доходу, можна додатково формулювати гіпотези та розглядати моделі класифікаторів, що в свою чергу дасть додаткове джерело інформації для інсайтів.

Заохочення інновацій та тестування нових підходів у керуванні рекламними кампаніями можуть допомогти підвищити ефективність рекламних ініціатив та, відповідно, збільшити доходи. Оптимізація рекламного нетворку вимагає комплексного підходу, що включає як технічні аспекти моделювання та аналізу даних, так і стратегічне планування на основі отриманої інформації. Вдосконалення прогнозування рекламного доходу сприятиме кращому розумінню та оптимізації рекламних кампаній, підвищуючи їх ефективність і прибутковість.

Отже, впровадження сучасних цифрових рішень та інструментів моделювання допомагають рекламній галузі бути більш конкурентоспроможною та більш якісно оцінювати свої витрати на основі маркетингових показників, і тим самим, прогнозувати прибуток, який вони отримують від певних кампаній. Проте для того, щоб моделювання було якісним та результативним, варто ґрунтовно підходити до підбору та вибору факторів прогнозування, а також моделі, щоб на їх основі отримати точні та добре інтерпретовані результати.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Shanahan J. Estimating the Expected Effectiveness of Text Classification Solutions under Subclass Distribution Shifts. 2012 IEEE 12th International Conference on Data Mining. 2012. URL: https://www.academia.edu/85149216/Estimating_the_Expected_Effectiveness_of_Text_Classification_Solutions_under_Subclass_Distribution_Shifts
2. Silva R., Zhang J., Shanahan J. Probabilistic workflow mining. Proceedings of the eleventh ACM SIGKDD. 2005. URL: https://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/rbas/papers/f211-silva.pdf
3. Sergienko I.V., Deineka V.S. Optimal Control of Distributed Systems with Conjugation Conditions. New York: Kluwer Academic Publishers. 2005.
4. Eck D., Schmidhuber J. Finding temporal structure in music: blues improvisation with LSTM recurrent networks. Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing. 2002. URL: https://ieeexplore.ieee.org/document/1030094
5. Goodfellow I., Bengio Y., and Courville A. Deep learning, Cambridge, MA: MIT Press. Adaptive computation and machine learning series | Includes bibliographical references and index. 2017. URL: https://books.google.com.ua/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=deep+learning&ots=MON3bpnGQR&sig=JI5AJXz5-Fk-vjxWNxgEeVbqD0U&redir_esc=y#v=onepage&q=deep%20learning&f=false
6. Galagan R.M. Analysis of application of neuralnetworks to improve the reliability of active thermal NDT / R. M. Galagan, A. S. Momot // KPI Science News. – Kyiv. – 2019. – № 1 (2019). – P. 7-14 DOI: https://doi.org/10.20535/kpisn.2019.1.157374
7. Zatonatska T., Dluhopolskyi O., Chyrak, I., & Kotys, N. The internet and e-commerce diffusion in European countries (modeling at the example of Austria, Poland, and Ukraine). Innovative Marketing. 15(1). 66-75. 2019. URL: http://dx.doi.org/10.21511/im.15(1).2019.06

8. Argueta J. K., Pérez-Latre F. J. The Transformation of Advertising Agencies in a Digital World. Handbook of Media Management and Economics. 2nd Edition, Routledge. 2018. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315189918-25/transformation-advertising-agencies-digital-world-j%C3%BCrg-kaufmann-argueta-francisco-p%C3%A9rez-latre
9. Behboudi M., Hanzaee K.H., Koshksaray A.A. A Review of the Activities of Advertising Agencies in Online World, International Journal of Marketing Studies Vol. 4, No. 1, 2012. URL: http://dx.doi.org/10.5539/ijms.v4n1p138
10. Shen F. Banner advertisement pricing, measurement, and pretesting practices: perspectives from interactive agencies. Journal of Advertising. Vol XXXI. No. 3. 2002.
11. Rodgers S., Thorson E. The Interactive Advertising Model: How Users Perceive and Process Online Ads. Journal of Interactive Advertising. 1(1). 2000. URL: http://jiad.org/vol1/no1/radgers/index.html
12. Bergen M., Dutta S., Walker O.C. Agency relationships in marketing: a review of the implications and applications of agency and related theories. Journal of Marketing. 56(July). 1-24. 1992. URL: http://dx.doi.org/10.2307/1252293
13. Hanafizadeh P., Behboudi M. Online Advertising and Promotion: Modern Technologies for Marketing. pp. 1-430. 2012.
14. Harvey M.G., Rupert J.P. Selecting an industrial advertising agency. Journal of Industrial marketing Management. 17(2). 119-27. 1988.
15. Rabindranath M., Singh A.K. Advertising Agencies. In: Advertising Management. Palgrave Macmillan. Singapore. 2024. URL: https://doi.org/10.1007/978-981-99-8657-6_5
16. Drossos D.A., Fouskas K.G., Kokkinaki F., Papakyriakopoulos D. Advertising on the internet: perceptions of advertising agencies and marketing managers. pp. 244-264. 2011. URL: https://doi.org/10.1504/IJIMA.2011.038238

17. Мальчик М. В., Адасюк І. П. Реклама в інтернеті: теоретичний аналіз та особливості. Journal of Lviv Polytechnic National University Series of Economics and Management Issues. Vol. 5. No. 1. 2021. URL: https://science.lpnu.ua/sites/default/files/journal-paper/2021/may/23590/210488verstka-77-87_0.pdf
18. Wang K. Y., Shih E., Peracchio L. How banner ads can be effective: Investigating the influences of exposure duration and banner ad complexity. International Journal of Advertising. Vol. 32 Issue 1. pp.121-141. 2013. URL: http://dx.doi.org/10.2501/IJA-32-1-121-141
19. Saura J.R. Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge. Volume 6. Issue 2. 2021. pp. 92-102. URL: https://doi.org/10.1016/j.jik.2020.08.001
20. Palos-Sánchez P., Suárez L.M. Understanding the Digital Marketing Environment with KPIs and Web Analytics. Future Internet. 9(4). 76. 2017. URL: https://doi.org/10.3390/fi9040076
21. Fiorini P.M., Lipsky L.R. Search marketing traffic and performance models. Comput. Stand. Interface. 34, 517–526. 2012. URL: https://doi.org/10.1016/j.csi.2011.10.008
22. Saura J. R. Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge. 6(2). 92-102. 2021. https://doi.org/10.1016/j.jik.2020.08.001
23. Ghahremani-Nahr J., Nozari H. A Survey for Investigating Key Performance Indicators in Digital Marketing. International Journal of Innovation in Marketing elements. VOL. 1 NO. 1: AUTUMN / Original Research. 2021. URL: https://doi.org/10.59615/ijime.1.1.1
24. Piñeiro-Otero T., Martínez-Rolán X. Understanding Digital Marketing—Basics and Actions. pp 37–74.

25. Antevenio Diferencias entre CPM, CPC, CPL, CPA y CPI 2015. http://www.antevenio.com/blog/2015/01/diferencias-entre-cpm-cpc-cpl-cpa-cpi/
26. Farris P.W., Bendle N.T. Marketing metrics: The Definitive Guide to Measuring Marketing Performance. 2006. URL: https://books.google.com.ua/books?hl=uk&lr=&id=7Ptw4nBoGmkC&oi=fnd&pg=PR7&dq=main+indicators+and+metrics+of+digital+marketing&ots=2bod2YcdXx&sig=wPbNXOqisnp9CejOquEGXyRmF_k&redir_esc=y#v=onepage&q&f=false
27. Ambler T. Marketing metrics: The Marketing Book. 6th Edition. Routledge. Pages 14. 2007. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9780080942544-30/marketing-metrics-tim-ambler
28. Zahay D., Griffin A. Marketing strategy selection, marketing metrics, and firm performance. Journal of Business & Industrial Marketing. Vol. 25 No. 2. pp. 84-93. URL: https://doi.org/10.1108/08858621011017714
29. Narkulova S. Methods Of Assessment And The Main Indicators Of The Effectiveness Of Advertising On The Internet. Journal Of Marketing, Business And Management (JMBM), VOLUME 1. ISSUE 4 (June) ISSN: 2181-3000.
30. Пономаренко І. В. Цифровий маркетинг як ефективний інструмент підвищення рівня конкурентоспроможності компанії / І. В. Пономаренко // Проблеми інноваційно-інвестиційного розвитку. – 2018. – № 15. – С. 57-65. URL: https://er.knutd.edu.ua/handle/123456789/12447
31. Conversion Metrics for Digital Marketing. URL: https://www.profit.co/blog/kpis-library/top-10-conversion-metrics-for-digital-marketing/
32. Factors Influencing the Use of Digital Marketing by Small and Medium-Sized Enterprises during COVID-19. Informatics 9(4):86. 2022. URL: https://www.researchgate.net/publication/365510632_Factors_Influencing_the_Use_of_Digital_Marketing_by_Small_and_Medium-Sized_Enterprises_during_COVID-19

33. Logistic Regression in Machine Learning URL: https://www.geeksforgeeks.org/understanding-logistic-regression/
34. Yanwu Yang1 , Panyu Zhail Click-Through Rate Prediction in Online Advertising: A Literature Review, School of Management, Huazhong University of Science and Technology. Wuhan. China 6-50. URL: https://doi.org/10.1016/j.ipm.2021.102853
35. Narkhede S. Understanding AUC - ROC Curve. Towards Data Science. 2018. URL: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5
36. Гнот Т.В., Негрей М.В. Алгоритми Data Science у моделюванні бізнес-процесів. ЕКОНОМІКА І СУСПІЛЬСТВО. № 12 / 2017. URL: https://economyandsociety.in.ua/journals/12_ukr/125.pdf
37. Pavlyshenko B.M. Linear, machine learning and probabilistic approaches for time series analysis. Data Stream Mining & Processing (DSMP). IEEE First International Conference on. IEEE. 2016. pp. 377–381.
38. Фаренюк Я. Аналіз на основі баз даних (Data Science) для управлінських рішень з урахуванням невизначеності макро та мікроекономічного рівнів, Економіка невизначеності: зміст, оцінювання, регулювання (pp. 99-121). URL: https://www.researchgate.net/publication/352753009_ANALIZ_NA_OSNOVI_BAZ_DANIH_DATA_SCIENCE_DLA_UPRAVLINSKIH_RISEN_Z_URAHUVANNAM_NEVIZNACENOSTI_MAKROTA_MIKROEKONOMICNOGO_RIVNIV
39. Cielen D., Meysman A., Ali M. Introducing Data Science: Big data, machine learning, and more, using Python tools. Simon and Schuster. Computers. pp. 320. 2016. URL: https://books.google.com.ua/books?hl=en&lr=&id=bTozEAAAQBAJ&oi=fnd&pg=PT14&dq=data+science+tools+and+its+using&ots=-jiWBtff3J&sig=gR_Kx9mp2iTM0L_y1ROsALWmCbw&redir_esc=y#v=onepage&q=data%20science%20tools%20and%20its%20using&f=false

40. Bloice M., Holzinger A. A Tutorial on Machine Learning and Data Science Tools with Python, Lecture Notes in Computer Science ((LNAI,volume 9605)). pp. 435-480. 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-50478-0_22
41. George N. Practical Data Science with Python: Learn tools and techniques from hands-on examples to extract insights from data. Packt Publishing. 2021. URL: https://books.google.com.ua/books?hl=en&lr=&id=4eRFEAAAQBAJ&oi=fnd&pg=PP1&dq=data+science+tools+and+its+using&ots=ob6jvetFat&sig=etHpPnGrD-ofMd-FVUc8GQ22ZLA&redir_esc=y#v=onepage&q=data%20science%20tools%20and%20its%20using&f=false
42. Jordan M.I., Mitchell T.M.: Machine learning: trends, perspectives, and prospects. Science 349(6245). 255–260. 2015. URL: https://www.science.org/doi/abs/10.1126/science.aaa8415
43. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. (JMLR) 12(10). 2825–2830. 2011. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/
44. Grover P., Kar A.K. Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature, Global Journal of Flexible Systems Management. Volume 18. pages 203–229. 2017. URL: https://link.springer.com/article/10.1007/s40171-017-0159-3
45. Bradlow E. T., Gangwar M., Kopalle P., Voleti S. The role of big data and predictive analytics in retailing. Journal of Retailing. 93(1). 79–95. 2017. URL: https://www.sciencedirect.com/science/article/abs/pii/S0022435916300835?via%3Dihub
46. Недашківська Н. І. Інтелектуальний аналіз даних. Практикум [Електронний ресурс] : навчальний посібник для студентів, які навчаються за спеціальністю

124 «Системний аналіз». Київ : КПІ ім. Ігоря Сікорського, 2021. URL: https://ela.kpi.ua/handle/123456789/53763

47. Medium Why exclude highly correlated features when building regression model?? URL: https://towardsdatascience.com/why-exclude-highly-correlated-features-when-building-regression-model-34d77a90ea8e
48. Agrawal R. Know The Best Evaluation Metrics for Your Regression Model. Analytics Vidhya. 2023. URL: https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/
49. Provost F., Fawcett T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. Vol. 1. No. 1. 2013. URL: https://doi.org/10.1089/big.2013.1508
50. Armstrong J.S. Selecting Forecasting Methods, Principles of Forecasting. International Series in Operations Research & Management Science ((ISOR,volume 30)). 2001. URL: https://link.springer.com/chapter/10.1007/978-0-306-47630-3_16
51. Armstrong J. S., Brodie A. Hypotheses in marketing science: Literature review and publication audit. Marketing Letters. 12. 171–187. 2001. URL: https://link.springer.com/article/10.1023/A:1011169104290
52. Januschowski T., Gasthaus J., Wang J. Criteria for classifying forecasting methods, International Journal of Forecasting. Volume 36. Issue 1. pp. 167-177. 2020. URL: https://doi.org/10.1016/j.ijforecast.2019.05.008
53. Wellens A.P., Boute R.N., Udenio M. Simplifying tree-based methods for retail sales forecasting with explanatory variables, European Journal of Operational Research, Volume 314. Issue 2. pp. 523-539. 2024. URL: https://doi.org/10.1016/j.ejor.2023.10.039
54. Mishev K., Gjorgjevikj A., Vodenska I., Chitkushev L. Forecasting Corporate Revenue by Using Deep-Learning Methodologies, International Conference on

Control. Artificial Intelligence. Robotics & Optimization (ICCAIRO). 2019. URL: https://ieeexplore.ieee.org/abstract/document/9057159

55. Medium Data Science Toolbox: Essential Tools and Techniques Every Data Scientist Should Know. 2023. URL: https://nsworldinfo.medium.com/data-science-toolbox-essential-tools-and-techniques-every-data-scientist-should-know-a0c31bd0f976

ДОДАТКИ

Додаток А

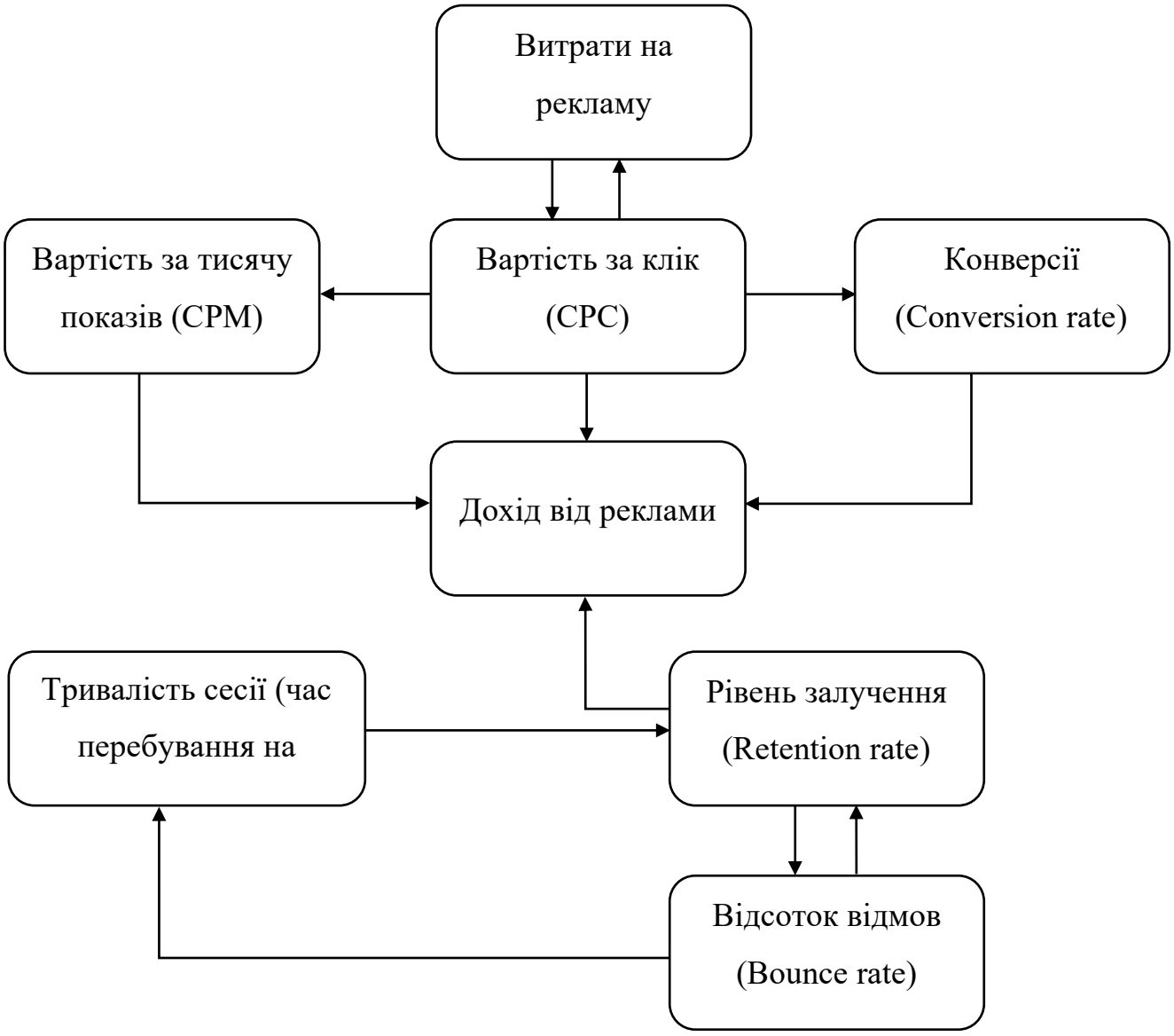

Рис. А.1. Основні метрики цифрового маркетингу

Джерело: складено на основі [31].

Додаток Б

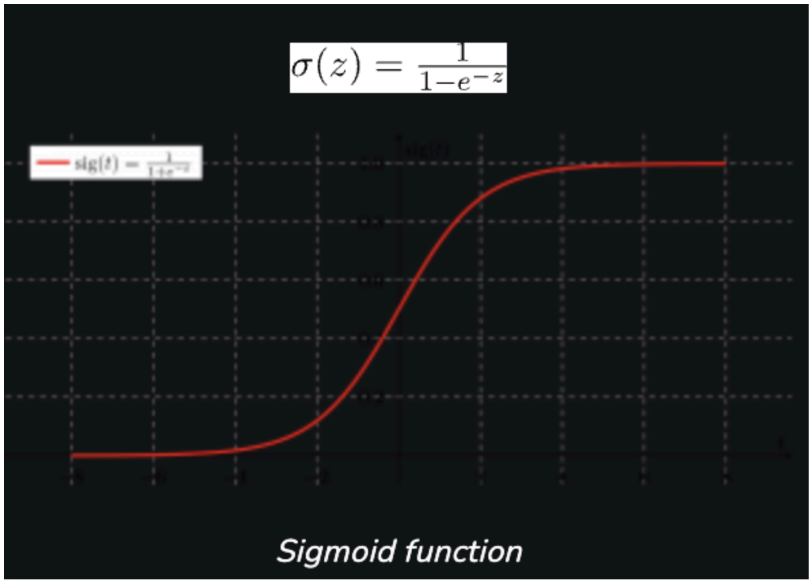

Рис. Б.1. Сигмоїдна функція логістичної регресії.

Джерело: [33].

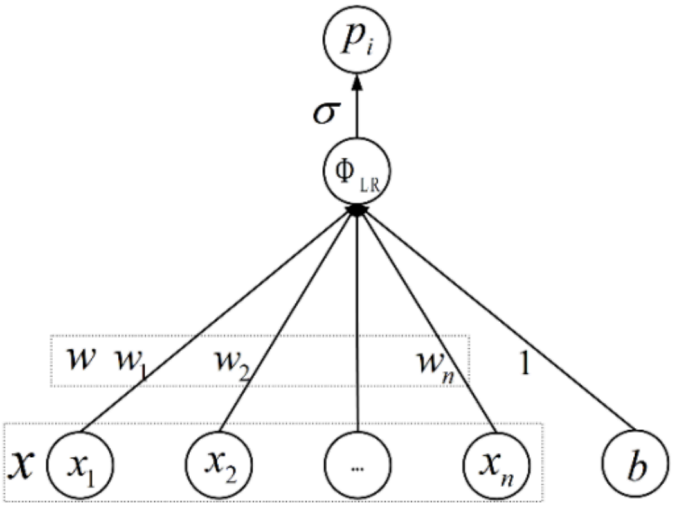

Рис. Б.2. Структура моделювання логістичної регресії.

Джерело: [34].

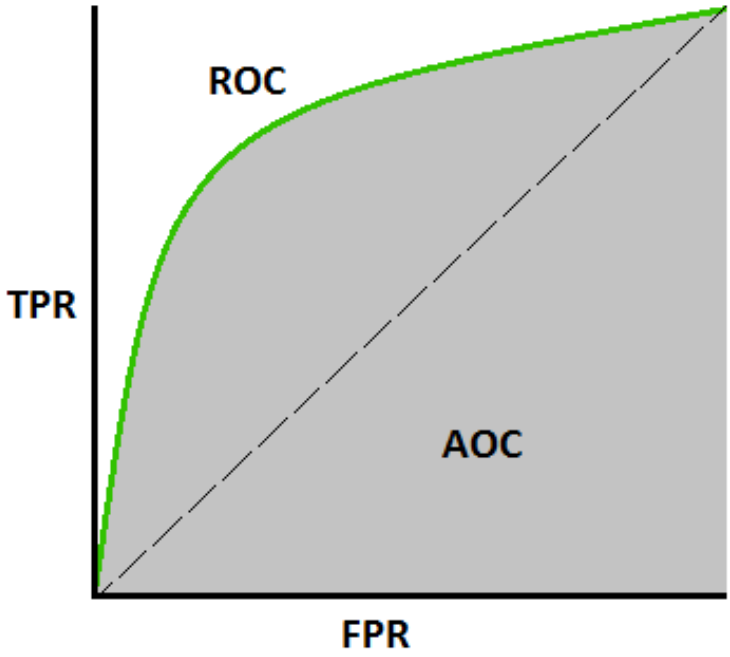

Рис. Б.3. Графік ROC-кривої.

Джерело: [35].

Додаток В

Скрипт для побудови моделі

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

from xgboost import XGBRegressor
from catboost import CatBoostRegressor
from sklearn.ensemble import GradientBoostingRegressor, RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.metrics import mean_absolute_percentage_error as mape
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.model_selection import RandomizedSearchCV

df = pd.read_csv('diploma_dataset.csv')

df_new = df.iloc[:, 1:].copy()

df_new['revenue'] = df_new['revenue'].replace(0.0, 0.01)

corr_t = df_new.corr()

corr_t.style.background_gradient(cmap='coolwarm').format(precision=2)

df_new = df.drop(columns=['date', 'visits', 'is_react', 'is_webview', 'is_rtb_group',
'en_language', 'has_push_click'])
```

```
df_new['revenue'] = df_new['revenue'].replace(0.0, 0.01)

corr = df_new.corr()

corr.style.background_gradient(cmap='coolwarm').format(precision=2)

X, y = df_new.iloc[:, :-1], df_new.iloc[:,-1:]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)

y_test = y_test['revenue'].values
y_train = y_train['revenue'].values

def test_model(model_name):
    model_name.fit(X_train, y_train)
    y_pred = model_name.predict(X_test)

# Calculate evaluation metrics
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    r_squared = r2_score(y_test, y_pred)
    rmse = np.sqrt(mse)
    m_mape = mape(y_test, y_pred)

# Print the evaluation metrics
    print(model_name, 'results')
    print("Mean Absolute Error (MAE):", mae)
    print("Mean Squared Error (MSE):", mse)
```

```
    print("R-squared (R²):", r_squared)
    print("Root Mean Squared Error (RMSE):", rmse)
    print("MAPE:", m_mape)

model = XGBRegressor()

test_model(model_name)

def hyperParameterTuning(X_train, y_train, m_name):
    param_tuning = {
        'learning_rate': stats.uniform(0.01, 0.1),
        'max_depth': stats.randint(3, 10),
        'min_child_weight': stats.randint(1, 5),
        'subsample': stats.uniform(0.5, 0.7),
        'colsample_bytree': [0.5, 0.7],
        'n_estimators':stats.randint(50, 200),
        'objective': ['reg:squarederror']
    }

    gsearch = RandomizedSearchCV(m_name, param_distributions=param_dist,
n_iter=20, cv=5, scoring='r2', verbose=10)

    gsearch.fit(X_train,y_train)

    return gsearch.best_params_, gsearch.best_estimator_

model_tun = XGBRegressor()
```

```
print("Best set of hyperparameters: ", random_search.best_params_)
print("Best score: ", random_search.best_score_)

best_params, best_estimator = hyperParameterTuning(X_train, y_train)

test_model(best_estimator)
```

Календарний план виконання кваліфікаційної роботи магістра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи магістра	24.10.2023	
2	Розробка та затвердження завдання кваліфікаційної роботи магістра	30.10.2023	
3	Збір та опрацювання списку джерел для кваліфікаційної роботи	05.12.2023	
4	Підготовка теоретичного розділу 1	08.01.2024	
5	Збір даних матеріалів для проведення аналізу в розділі 2	22.01.2024	
6	Підготовка розділу 2	19.02.2024	
7	Збір бази даних та її аналіз перед побудовою моделей	04.03.2024	
8	Побудова моделей	18.03.2024	
9	Оформлення отриманих результатів моделювання у розділ 3	25.03.2024	
10	Написання висновків	01.04.2024	
11	Остаточне оформлення результатів	29.04.2024	
12	Перевірка на плагіат	08.05.2024	
13	Попередній захист роботи	10.05.2024	
14	Рецензування	13.05.2024	
15	Подача роботи на кафедру	17.05.2024	
16	Захист роботи	27.05.2024	

Науковий керівник: Тетяна ЗАТОНАЦЬКА_____
(підпис)

Студент: Вероніка САЧКО_____
(підпис)

Київський національний університет імені Тараса Шевченка

Економічний факультет
Кафедра економічної кібернетики

ЗАВДАННЯ
на кваліфікаційну роботу магістра
студентки 2 курсу спеціальності 051 «Економіка» ОПП «Економічна кібернетика»
Сачко Вероніки Володимирівни

1. Тема роботи: «Використання інструментів Data Science для прогнозування доходу рекламної компанії».
2. Термін завершення роботи: 29.04.2023 р.
3. Попередній захист роботи: 10.05.2023 р.
4. Об’єкт дослідження: діяльність рекламної компанії, що займається закупівлею та продажем трафіку.
5. Предмет дослідження: методи та моделі, які використовуються для аналізу даних та прогнозування доходів рекламного нетворку.
6. Мета дослідження: аналіз впливу інструментів Data Science на ефективність і дохідність рекламних кампаній, побудова моделей для прогнозування та розробка рекомендацій для оптимізації діяльності рекламного нетворку.
7. Завдання дослідження:
   - 7.1. дослідити характеристики та види рекламних компаній із урахуванням маркетингових показників та чинників впливу на них;
   - 7.2. проаналізувати розвиток та застосування інструментів Data Science в рекламній індустрії, визначити ключові інструменти та підходи вимірювання їх ефективності;
   - 7.3. розробити моделі для прогнозування доходу рекламного нетворку, використовуючи дані з внутрішньої бази даних компанії;

Науковий керівник: докторка економічних наук, професорка Затонацька Тетяна Георгіївна

Тетяна ЗАТОНАЦЬКА _____

Студент: Вероніка САЧКО _____

Затверджено на засідання кафедри економічної кібернетики
протокол № 13 від 13 травня 2024 р.