

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

«На правах рукопису»
УДК 004.942.3

«До захисту допущено»
Завідувач кафедри ММСА
_____ О.Л.Тимошук

«___» _____ 20__ р.

Магістерська дисертація
на здобуття ступеня магістра
зі спеціальності 124 Системний аналіз
на тему: «Марківські ланцюги для визначення маркетингового впливу на
різні групи людей»

Виконала:
студентка II курсу, групи КА-02мп
Тумасова Олександра Сергіївна _____

Керівник:
доцент кафедри ММСА,
к.ф.-м.н., доц.,
Каніовська Ірина Юріївна. _____

Рецензент:
доцент кафедри математичного аналізу
та теорії ймовірностей
КПІ ім. Ігоря Сікорського,
канд. фіз.-мат. наук, доцент
Ільєнко М.К. _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних посилань.

Студент _____

Київ
2021

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти – другий (магістерський)

Спеціальність – 124 «Системний аналіз»

ЗАТВЕРДЖУЮ

Завідувач кафедри

О.Л. Тимошук

(підпис)

«___» _____ 2021 р.

ЗАВДАННЯ

на магістерську дисертацію студенту
Тумасовій Олександрі Сергіївні

1. Тема дисертації: «Марківські ланцюги для визначення маркетингово впливу на різні групи людей», науковий керівник дисертації Каніовська Ірина Юріївна, доцент кафедри ММСА, к.ф.-м.н., доц., затверджені наказом по університету від «02» листопада 2021 р. № 3651-с.
2. Термін подання студентом дисертації 12.12.2021
3. Об'єкт дослідження: маркетингові комунікації.
4. Предмет дослідження: реакції споживачів, характеристики споживачів. методи машинного навчання для їх обробки, методи навчання з підкріпленням.
5. Перелік завдань, які потрібно зробити:
 - 1) Дослідити актуальність обраної теми;
 - 2) Здійснити огляд технічної літератури за темою дослідження;
 - 3) Зробити дослідження та аналіз існуючих методів та моделей;
 - 4) Здійснити огляд технічної літератури з теми марківських ланцюгів;
 - 5) Розробити архітектуру моделі для моделювання маркетингового впливу;
 - 6) Ввести метрику для оцінки результатів роботи моделі;
 - 7) Знайти та підготувати набори даних для навчання моделей та проведення аналізу;
 - 8) Розробити та протестувати систему, яка буде аналізувати вхідні дані про користувачів та моделювати їх реакції;
 - 9) Провести аналіз результатів;
 - 10) Провести порівняння створеної моделі з уже існуючими;
 - 11) Розробити концептуальні висновки;

- 12) Підготувати ілюстративний матеріал;
 - 13) Оформити пояснювальну записку.
6. Орієнтовний перелік ілюстративного матеріалу:
1. Методи оцінки реакції користувачів
 2. Алгоритм роботи моделі
 3. Приклад результатів використання системи
7. Дата видачі завдання: 1 вересня 2021 року.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Отримання завдання на магістерську дисертацію	01.09.2021 – 05.09.2021	Виконано
2.	Огляд технічної літератури за темою	06.09.2021 – 12.09.2021	Виконано
3.	Концептуальний вступ дисертації. Формулювання об'єкту, предмета, цілі, завдань, новизни, практичної значущості результатів. Дослідження актуальності обраної теми	13.09.2021 – 19.09.2021	Виконано
4.	Перший розділ. Дослідження актуальності моделювання маркетингового впливу. Огляд основних інструментів та підходів розв'язання даної задачі	20.09.2021 – 26.09.2021	Виконано
5.	Другий розділ. Математична основа розробки моделі	27.09.2021 – 03.10.2021	Виконано
6.	Третій розділ. Розробка системи для моделювання маркетингового впливу	04.10.2021 – 10.10.2021	Виконано
7.	Розширення функціоналу системи	11.10.2021 – 17.10.2021	Виконано
8.	Аналіз результатів	18.10.2021 – 24.10.2021	Виконано
9.	Проведення аналізу ринкових можливостей стартап – проекту	25.10.2021 – 31.10.2021	Виконано
10.	Підготовка ілюстративного матеріалу	01.11.2021 – 14.11.2021	Виконано
11.	Оформлення пояснювальної записки	15.11.2021 – 26.11.2021	Виконано

Студент

О. С. Тумасова

Науковий керівник
дисертації

І. Ю. Каніовська

РЕФЕРАТ

Магістерська дисертація: 94 с., 9 рис., 25 табл., 1 додаток, 23 джерела.

Дана робота присвячена дослідженню методів моделювання маркетингового впливу на цільові аудиторії споживачів та побудові нової моделі для вирішення подібних задач.

Об'єктом дослідження є маркетингові комунікації.

Предметом дослідження є моделювання маркетингового впливу на клієнтів компаній з певними характеристиками, методи навчання з підкріпленням, марківські ланцюги, метрики для оцінки реакції на взаємодію.

Метою дослідження є побудова системи для аналізу та моделювання реакції споживачів на маркетингові дії, що дають змогу визначати необхідність цільового впливу маркетингової кампанії на певну групу людей.

Актуальність роботи полягає в моделюванні реакції на маркетингову взаємодію, яка потрібна компаніям, що розвиваються або ставлять собі за мету розширення цільової аудиторії та збільшення обсягів продажу. Методи uplift моделювання дають змогу оцінити чистий ефект від взаємодії і не витратити кошти на клієнтів, яким реклама не потрібна.

Результатом роботи є побудована система для моделювання маркетингово впливу на споживачів на основі їх характеристик.

Новизною роботи є введення моделі Марківських ланцюгів для задач даного виду.

МАРКЕТИНГОВИЙ ВПЛИВ, UPLIFT МОДЕЛЮВАННЯ, МАРКІВСЬКІ ЛАНЦЮГИ, МАРКІВСЬКИЙ ПРОЦЕС ПРИЙНЯТТЯ РІШЕНЬ.

ABSTRACT

Master's thesis: 94 pp., 9 fig., 25 tables., 1 appendix, 23 sources.

Theme of the work is: "Markov chains for determining optimal marketing treatment of various groups of people"

Object of research is marketing communications.

The subject of research is modeling marketing impact on customers of companies with certain characteristics, Markov chains, metrics for uplift modeling.

Purpose of the work is to build a system for analyzing and modeling consumer response to marketing actions, which allows to determine the need for targeted impact of the marketing campaign on a particular group of people.

The relevance of the work is that modeling the response to marketing interactions is needed by all companies that are currently developing or just want to increase their audience. Uplift modeling techniques allow you to assess the net effect of the interaction and not spend money on customers who do not need advertising.

The result is a system for modeling marketing impact on consumers based on their characteristics.

The novelty of the work is the introduction of the model of Markov chains for this problem.

MARKETING INFLUENCE, UPLIFT MODELING, MARK'S CHAINS,
MARK'S DECISION PROCESS

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ	8
ВСТУП	9
РОЗДІЛ 1 АНАЛІЗ ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Актуальність визначення маркетингового впливу	11
1.2 Класифікація клієнтів	13
1.3 Основні компоненти моделювання маркетингово впливу та постановка задачі	14
1.4 Основні підходи до вирішення задачі	16
1.5 Стандартні метрики uplift моделювання	26
1.6 Висновки до розділу	33
РОЗДІЛ 2 ВИЗНАЧЕННЯ МАРКЕТИНГОВОГО ВПЛИВУ ЗА ДОПОМОГОЮ МАРКІВСЬКИХ ЛАНЦЮГІВ	34
2.1 Використання марківських ланцюгів для uplift моделювання	34
2.2 Марківські ланцюги	34
2.3 Марківський процес прийняття рішень	39
2.4.Визначення uplift моделювання для кількох дій	41
2.5 Метрика uplift моделювання для кількох дій	42
2.6 Зв'язок між uplift моделюванням та задачею “контекстних бандитів”	46
2.7 Метод градієнта політики	47
2.8 Модель марківського процесу прийняття рішень для uplift моделювання	50
2.9 Висновки до розділу	54
РОЗДІЛ 3 МОДЕЛЮВАННЯ МАРКЕТИНГОВОГО ВПЛИВУ	56
3.1 Розробка програмного забезпечення	56
3.2 Огляд та аналіз даних	56

3.3 Перевірка роботи моделі за умови двох дій	59
3.4 Перевірка роботи моделі при моделюванні кількох дій	61
3.5 Висновки до розділу	62
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ	63
4.1 Опис ідеї стартап проекту	63
4.2 Технологічний аудит ідеї проекту	65
4.3 Аналіз ринкових можливостей запуску стартап проекту	65
4.4 SWOT – аналіз	74
4.5 Розроблення ринкової стратегії проекту	75
4.6 Розроблення маркетингової програми стартап – проекту	78
4.7 Висновки	81
ВИСНОВКИ	82
ПЕРЕЛІК ПОСИЛАНЬ	84
ДОДАТОК А ЛІСТИНГ ПРОГРАМИ	87

ПЕРЕЛІК СКОРОЧЕНЬ

uplift - реакція на взаємодію.

SN-UMG - метрика для uplift моделювання

Qini - метрика для uplift моделювання

МППР - марківський процес прийняття рішень

ВСТУП

Проблемою моделювання поведінки клієнтів та прогнозуванням схильності до купівлі товару зацікавились ще в 50-х роках. Компанії прагнули збільшувати продажі своїх товарів та витратити кошти на маркетингове просування з більшою ефективністю. Створюючи перші моделі маркетологи намагалися визначити загальну схильність клієнтів купувати товар. Проте таке моделювання призводило до надлишкових, неефективних витрат - маркетингові дії також були направлені на людей, які купували б товар і без впливу рекламних кампаній та стимульованих цими рекламними кампаніями взаємодій.

Основна мета маркетингових дій - саме стимулювання потенційного клієнта до виконання цільових дій, які б він не здійснив без впливу рекламної кампанії, тобто змусити покупця змінити свою поведінку. Також, важливими є супутні цілі, такі як поінформованість про бренд створення позитивного іміджу компанії та підтримання лояльності клієнтів. Зазвичай ефективність маркетингових кампаній оцінюються на основі певного розрахунку рентабельності інвестицій вкладених у ці кампанії. У випадку якщо кампанія зосереджується на отриманні додаткового доходу, внаслідок росту продажів метрики для вимірювання ефективності стають нетривіальним, оскільки несуть в собі інтегровані показники, які не дають можливості розділити продажі внаслідок стимулювання та зміни поведінки клієнтів та продажі, які були б здійснені і без рекламної кампанії. Тобто частина коштів у таких рекламних кампаніях витрачається неефективно.

З метою зменшення витрат на маркетингові дії був введений клас моделей який оцінюють чистий ефект від взаємодії - uplift моделювання. Даний клас моделей знаходить саме тих клієнтів, які будуть купувати товар лише за наявності маркетингової взаємодії. Для вирішення цієї задачі використовуються різні підходи. Проте більшість з них надає не досить достовірні результати. Це можна пояснити тим, що задача нова та недостатньо досліджена.

В роботі досліджується моделювання маркетингового впливу на різні групи людей за допомогою uplift моделювання з методом заснованим на марківських ланцюгах та застосувати її на реальних даних. Даний метод дозволяє розглядати uplift моделювання для кількох дій. Для використання марківських ланцюгів потрібно переформулювати класичну задачу uplift моделювання та визначити її для кількох дій. Також вводиться метрика, якою можна буде оцінювати результати моделювання.

В першому розділі роботи розглянуто теоретичну базу моделювання маркетингового впливу та основні поняття, моделі та метрики uplift моделювання.

В другому розділі роботи розглянуто теоретичну базу марківських ланцюгів та марківського процесу прийняття рішень. Сформульовано визначення uplift моделювання для кількох дій та проведений взаємозв'язок з задачею “контекстуальних бандитів”. Для розв’язання цієї задачі введена нова метрика, з якою можна оцінювати результати моделювання саме марківськими ланцюгами.

В третьому розділі описано результати моделювання запропонованим методом та проведено його порівняння з класичними підходами. Розглянуто два випадки uplift моделювання: для двох та кількох дій.

Четвертий розділ роботи присвячений опису стартап-продукту за темою.

РОЗДІЛ 1 АНАЛІЗ ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність визначення маркетингового впливу

Проблемою моделювання поведінки клієнтів зацікавились ще в 50-х роках. Підходи до цієї задачі поступово розвивалися та сформувалися 4 основні:

- Моделі, які намагаються знайти закономірності в історичних даних так охарактеризувати клієнтів, які вже купили товар. Такий підхід називається look-alike. Їх припущення ґрунтується на тому, що клієнти, які мають такі самі характеристики, як ті, хто вже купив товар теж його куплять. Тобто вони просто оцінюють схильність клієнта виконати цільову дію. Такі моделі будуть добре працювати на ринку, який далекий від насичення.

- “Моделі покупок” - за принципом схожі до попереднього підходу, проте надають більше уваги до більш нових даних. Вони більш чутливі до зміни характеристик клієнтів.

- Моделі “відповіді” - характеризує клієнтів, які здійснили покупку або виконали цільову дію у відповідь на певну маркетингову дію. Тобто вони прогнозують схильність клієнта купити товар, якщо була цільова дія. Для того, щоб відокремити тих, хто реагує на маркетингову дію, клієнтам надають код для акції або купони. Або в іншому випадку, вважають, що всі хто здійснив цільову дію після маркетингової кампанії - відреагували на неї. Ці моделі вважаються більш складними, ніж попередні дві, оскільки вони намагаються з'єднати факт покупки та маркетингову дію.

У контексті утримання клієнтів використовуються аналогічні підходи. Моделями прогнозується, які клієнти скоріш за все покинуть компанію, а в поєднанні з оцінкою вартості компанії можуть зосередитися на збереженні вартості.

Основна мета маркетингових дій - здійснення певних змін у поведінці клієнтів, стимулювання до виконання цільових дій. Також, є додаткові цілі, такі як поінформованість про бренд та створення доброзичливості клієнтів.

Більшість маркетингових кампаній оцінюються насамперед на основі певного розрахунку рентабельності інвестицій. Якщо компанія зосереджується на отриманні додаткового доходу, вимірювання успіху стає нетривіальним, оскільки важко зрозуміти, якого рівня продажів було б досягнуто, якби не здійснювалася відповідна маркетингова діяльність. Часто використовують контрольні групи (з якими комунікація не проводиться). Це вважається найкращим підходом для вимірювання поступового впливу діяльності прямого маркетингу - порівняння результатів цільової групи (з якими комунікація проводиться) з результатами контрольної групи, при тому, що групи вибираються випадковим чином з усієї сукупності.

Жоден з підходів, які зазначені вище не призначений для моделювання саме поступового впливу маркетингових дій, тобто змін в поведінці клієнта, оскільки вони не відрізняють клієнтів, які прийшли через маркетингову компанію від інших. В результаті цього компанія буде витратити багато грошей на маркетинг для клієнтів, які купували б товар і без реклами. Тому компанії зацікавлені в моделюванні реакції клієнта саме на маркетингову дію та розумінні того, який ефект від взаємодії вони мають. Наприклад, власник інтернет магазину може бути зацікавлений в оцінці ефекту від надсилання електронного листа на поштову адресу клієнта з певними характеристиками, на його схильність продовжувати покупки в тому ж магазині.

Uplift моделювання - це набір методів для прогнозування поступового впливу від взаємодії з об'єктом. Звичайні методи класифікації саме на прогнозуванні реакції клієнта, а uplift моделювання фокусується на оцінці чистого ефекту від взаємодії, шляхом моделювання різниці реакції людини до та після конкретної взаємодії.

Для вирішення цієї задачі спочатку намагалися використовувати звичайні методи машинного навчання, такі як k-nearest neighbors, метод опорних векторів та випадковий ліс. Проте, залишаються 2 нерозв'язані проблеми, які погіршують їх роботу при застосуванні, це:

- відсутність об'єктивної метрики для оцінки;

- індивідуальні реакції на наявність і відсутність взаємодії невідомі, що не дає змоги визначити реакцію на наявність конкретних дій.

Uplift модель оцінює чистий ефект від комунікації, тобто знаходить лише тих клієнтів, які виконують цільову дію лише за наявності комунікації. Модель оцінює різницю в поведінці клієнта за наявності та відсутності комунікації. Дану модель добре використовувати тоді, коли цільова дія виконується з високою ймовірністю за відсутності комунікації. Цей метод має декілька різних підходів до вирішення, про через їх різну формалізацію та причинно наслідковий вивід їх складно порівнювати.

1.2 Класифікація клієнтів

Клієнтів зазвичай розділяють на 4 групи:

1. Ті, яких краще не турбувати - вони будуть негативно реагувати за наявності комунікації.
2. Втрачені - це ті, які ніколи не виконують цільову дію, незалежно від того була комунікація чи ні.
3. Лояльні - ті, які будуть позитивно реагувати, не зважаючи на те була комунікація чи ні.
4. Переконаванні - це ті, які позитивно реагують на комунікацію, проте не виконують цільову дію за її відсутності.

Втрачені та лояльні клієнти не приносять доходу та, навіть, створюють додаткові витрати. Переконавальні клієнти - це саме ті, кого треба знайти за допомогою моделі. В залежності від компанії деякі типи клієнтів можуть бути відсутніми. Виконання цільової дії залежить від самої характеристики компанії, тому даний параметр теж треба строго сформулювати.

1.3 Основні компоненти моделювання маркетингово впливу та постановка задачі

Для класичних методів uplift моделювання необхідно сформулювати причинно-наслідковий зв'язок. Для побудови причинно-наслідкової моделі часто використовують модель Рубіна, в основі якої уявлення про потенційний прибуток та ефект від наслідку взаємодії.

Для того, щоб знайти ефект від взаємодії обчислюють різницю між реакціями людини за наявності комунікації та за її відсутності. Оскільки фактично це зробити неможливо, тому вводиться величина (дельта потенційно можливих реакцій) причинно-наслідковий ефект від наявності комунікації по відношенню до її відсутності[1]:

$$\tau_i = Y_i^1 - Y_i^0$$

де Y_i^1 - потенційна реакція людини, якщо була комунікація та Y_i^0 потенційна реакція, якщо не було комунікації, i - клієнт.

Знаючи певні ознаки кожного клієнта можна вводять усереднений ефект від взаємодії, його позначають CATE (умовний середній ефект від взаємодії) та рахують за формулою[1]:

$$CATE = E[Y_i^1|X_i] - E[Y_i^0|X_i]$$

де Y_i^1 - потенційна реакція людини, якщо була комунікація та Y_i^0 потенційна реакція якщо не було комунікації,

i - клієнт,

X_i - характеристики i -го клієнта.

Проте CATE та причинно-наслідковий ефект не спостерігають та, відповідно, не оптимізують, оскільки не можна спостерігати відразу Y_i^1 та Y_i^0

Тому вводиться бінарна змінна та використовують оцінку CATE - uplift:

$$uplift = E[Y_i^1 | X_i = x, W_i = w] - E[Y_i^0 | X_i = x, W_i = w]$$

де X_i - характеристики клієнта i ; Y_i спостережувана реакція клієнта в результаті маркетингової компанії, яка визначається як:

$$Y_i = W_i Y_{i1} + (1 - W_i) Y_{i0} = \begin{cases} Y_i^1, & \text{якщо } W_i = 1 \\ Y_i^0, & \text{якщо } W_i = 0 \end{cases}$$

де $W_i = 1$ - означає, що клієнт потрапив в цільову групу, де була комунікація, а

$W_i = 0$ - означає, що клієнт потрапив в групу, з якої комунікації не було (контрольну групу). Проте не завжди можна оцінити CATE за допомогою даної формули на основі спостережень. Дана формула для uplift буде працювати за умови, що розподіл на цільову та контрольну групу відбувається випадково, незалежно від будь-якої ознаки, а потенційна реакція об'єкта це наслідок з його ознак. Це називається умовне припущення незалежності [1] і записується як:

$$\{Y_i(1), Y_i(0)\} \perp W_i | X_i$$

Також вводять оцінку схильності клієнта, тобто того, що клієнт і потрапить в цільову групу:

$$p(X_i) = P(W_i = 1 | X_i)$$

Зважаючи на те, що задача полягає в тому щоб оцінити різницю між двома подіями, які є взаємовиключні для кожного об'єкта, до підбору даних. Тобто, для отримання вибірки даних для моделі потрібно:

1. Розділити випадковим чином репрезентативну клієнтської бази на цільову та контрольну групи.
2. На цільовій групі провести тестову маркетингову кампанію.

На основі даних отриманих з тестової маркетингової кампанії будується вибірка для моделі uplift. Загальна схема зображена на рис. 1.1

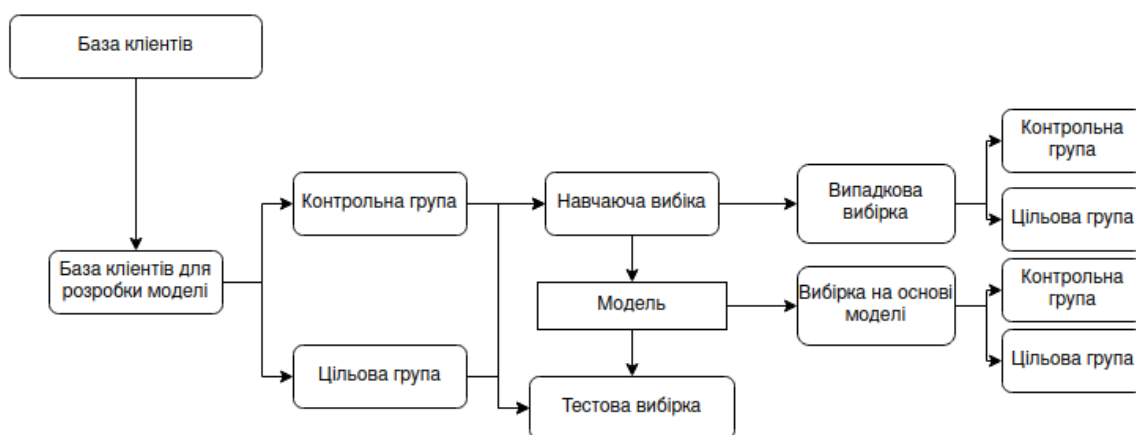


Рисунок 1.1 - Базова схема uplift моделювання

1.4 Основні підходи до вирішення задачі

Зазвичай виділяють п'ять класичних підходів до розв'язання задачі:

- двомодельний підхід - полягає в тому, щоб побудувати дві моделі прогнозування. Одна модель використовується на контрольній групі, інша на цільовій;

- трансформація класів - зазвичай використовується для випадків, коли цільова змінна бінарна. Проте є регресійна модифікація підходу, яка не має таких обмежень;
- підхід з однією моделью - найпростіше рішення, де модель одночасно навчається на контрольній та цільовій групі;
- багатокласова модель - знаходить співвідношення між 4 можливими класами;
- прогнозування напряду - модифікації добре відомих алгоритмів машинного навчання, наприклад, дерева рішень, метод опорних векторів або random forest.

Двомодельний підхід

Цей підхід вважаються базовою моделлю для uplift моделювання. Метод досить популярний і часто зустрічається в статтях. Наприклад, в статті [3] та [4] розглядається використання цієї моделі на різних реальних даних. Суть полягає в тому, що будується прогнози для двох умовних ймовірностей на цільовій та контрольній групі. Моделі, якими будується прогноз повинні бути незалежні. Алгоритм наступний:

1. Будується перша модель, яка оцінює ймовірність виконання цільової дії між клієнтами, з якими була взаємодія.
2. Будується друга модель, яка оцінює ймовірність виконання цільової дії, з якими не було комунікації.
3. Для кожного клієнта розраховується різниця оцінок двох моделей.

Процес навчання можна записати наступним чином:

$$model^T = fit \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{p1} & \dots & x_{pk}, & y_p \end{pmatrix}, model^C = fit \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{q1} & \dots & x_{qk}, & y_q \end{pmatrix}$$

де $x_{11} \dots x_{pk}$ - цільова група;

$x_{11} \dots x_{qk}$ - контрольна група;

$y_1 \dots y_q, y_1 \dots y_p$ - цільові змінні для контрольної та цільової групи;

$model^C, model^T$ - відповідно побудовані моделі для контрольної та цільової групи.

Процес застосування моделі виглядає наступним чином:

$$\begin{aligned} model^T predict &= fit \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & y_m \end{pmatrix} - model^C predict = \\ &= fit \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & y_m \end{pmatrix} = \begin{pmatrix} u_1 \\ \dots \\ u_m \end{pmatrix} \end{aligned}$$

де $x_{11} \dots x_{mk}$ - клієнти для яких рахується uplift,

$u_1 \dots u_m$ - результат моделювання.

Перевага даного методу в його простоті. Завдяки тому, що висновок по кожній групі робиться окремо, алгоритми машинного навчання можна використовувати в їх звичайному вигляді, незалежно від того це будуть регресійна задача чи класифікація. Обидві моделі окремо можуть дати високу точність прогнозування. Проте, незважаючи на те, що даний підхід часто показує хороші результати, в деяких роботах зазначається, що інші моделі

дають набаго кращий результат. Наприклад, в статті [5] розглядаються такі підходи, як трансформація класів та методи на основі дерев рішень.

Існує модифікації такого підходу, де будуються дві залежні моделі. Ідея метода запропонованого в [6] полягає в тому, що використовують два бінарних класифікатора, один з яких використовує передбачення попереднього класифікатора в якості додаткових ознак об'єкта.

В [7] пропонується спочатку навчати дві моделі на різних групах. Потім цільові змінні модифікуються на основі передбачень моделей (перехресно) і потім моделі навчаються на оновлених даних, тобто спочатку моделі так само навчаються на тестовій та контрольній вибірці:

$$model^T = fit \left(\begin{matrix} x_{11} & \dots & x_{1k} & y_1 \\ \dots & \dots & \dots & \dots \\ x_{p1} & \dots & x_{pk} & y_p \end{matrix} \right), model^C = fit \left(\begin{matrix} x_{11} & \dots & x_{1k} & y_1 \\ \dots & \dots & \dots & \dots \\ x_{q1} & \dots & x_{qk} & y_q \end{matrix} \right)$$

$x_{train\ control}$
 $x_{train\ treatment}$

де $X_{train_treat}, X_{train_control}$ - цільова та контрольні вибірки;

$Y_{train_treat}, Y_{train_control}$ - цільові змінні для обох груп.

Далі кожна модель робить передбачення на вибірці, на якій навчалась інша модель, та формується нова цільова змінна, яка буде різницею передбачення моделі та попередньої цільової змінної для групи відповідної до моделі:

$$\hat{Y}^c = model_{predict}^C \left(\begin{matrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{p1} & \dots & x_{pk} \end{matrix} \right); \quad \hat{Y}^T = model_{predict}^T \left(\begin{matrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{q1} & \dots & x_{qk} \end{matrix} \right)$$

$$\tilde{D}^T = Y_{\text{train treat}} - Q^c; \tilde{D}^c = \hat{Y}^T - Y_{\text{train control}}$$

Наступний крок - побудова нових моделей, з тими самими вхідними даними, проте з оновленою цільовою змінною. Зважена сума передбачень цих моделей і буде передбаченням uplift:

$$(1 - g) \cdot \text{model}^T \text{ predict } \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & y_m \end{pmatrix} + g \cdot \text{model}^C \text{ predict } \begin{pmatrix} x_{11} & \dots & x_{1k}, & y_1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & y_m \end{pmatrix} =$$

$$= \begin{pmatrix} u_1 \\ \dots \\ u_m \end{pmatrix}$$

Трансформація класів

Метод запропонований в 2012 році [8], для випадку бінарної цільової змінної. Ідея метода побудова нової цільової змінної, наступним чином:

$$Z_i = Y_i^{obs} W_i + (1 - Y_i^{obs})(1 - W_i)$$

де Z_i - нова цільова змінна, Y_i - спостережувана цільова змінна, W_i - наявність комунікації. Тобто, якщо нова цільова змінна дорівнює 1, це означає, що результат при взаємодії був би таким же успішним, як і в контрольній групі, за умови, якби ми могли знати результати в обох групах. Тобто тоді коли цільова змінна дорівнює 1 та об'єкт знаходиться в цільовій групі або коли цільова змінна дорівнює 0, а об'єкт знаходиться в контрольній групі:

$$Z_i = \begin{cases} 1, & W_i=1 \text{ та } Y_i = 1 \\ 1, & W_i=1 \text{ та } Y_i = 1 \\ 0, & \text{в інших випадках} \end{cases}$$

Враховуючи, що $p(X_i = x) = 1/2$, тобто контрольні групи збалансовані та те, що розподілення груп відбувається випадково, було доведено [8], що:

$$uplift = 2 \cdot P(Z = 1) - 1$$

Таким чином задача зводиться до моделювання $P(Z_i = 1 | X_i)$. Цей метод популярний, оскільки він дає кращі результати ніж двомодельний підхід і при цьому залишається досить простим. Для цієї задачі може бути використаний будь-який стандартний класифікатор.

Проте дві умови на вхідні дані сильно звужують кількість випадків, коли даний підхід може бути застосований. Існує модифікована версія цього підходу немає таких обмежень. Цільова змінна будується таким чином[5]:

$$Z_i = Y_i \frac{W_i - p}{p(1-p)}$$

де Z_i - нова цільова змінна, W_i - флаг наявності комунікації, p - ймовірність потрапляння в цільову групу. Ймовірність p можна оцінити як долю об'єктів з флагом $W=1$ у вибірці або як функцію від властивостей об'єкта. Для другого способу потрібно навчити класифікатор на даних, де цільова змінна буде флаг W . На новій цільовій змінній Z можна навчити регресійну модель, з використанням середньоквадратичної помилки - тому, що саме в такому випадку передбачення моделі будуть умовним математичним сподіванням цільової змінної. Умовне математичне сподівання нової цільової змінної і є причинно-наслідковим ефектом[10]:

$$E[Y_i^* | X_i] = \tau(X_i)$$

де Z_i - нова цільова змінна, X_i - характеристики клієнтів, $\tau(X_i)$ - причинно-наслідковий вивід.

У випадку збалансованого розподілення об'єктів на групи, враховуючи рівності 5 та 16, рівність 14 переписується як:

$$Z_i = \frac{1}{2}Y_1^* + (1 - W_i)$$

що еквівалентно рівності 15 - тобто регресійний підхід зводиться до початкової формули.

Багатокласова модель

Оскільки з клієнтами можна взаємодіяти та спостерігати їх реакцію(виконання цільової дії), виокремлюють 4 класи клієнтів. Ці класи пов'язані з типами клієнтів з попереднього розділу, проте різниця їх в тому, що ці класи утворюються після безпосередньо спостереження. Кожен клас клієнта містить у собі 2 типи клієнтів одночасно.

- Ті які виконали цільову дію та з ними не було взаємодії (CR). Це можуть бути клієнти типу “не турбувати” або лояльні.
- Ті, які не виконали і з ними не було взаємодії (CN). Це можуть бути клієнти типу “втрачені” або “переконувані”.
- Ті, з якими була комунікація і вони виконали цільову дію (TR). Це можуть бути клієнти типу “лояльні” або “переконувані”.
- Ті, з якими була комунікація і вони не виконали цільову дію(TN). Це можуть бути клієнти типу “втрачені” або “не турбувати”.

Даний підхід полягає і прогнозуванні ймовірності віднесення клієнта до певного типу та побудові моделі для багатокласової класифікації. Тоді uplift буде рахуватися наступним чином:

$$uplift = [P(TR|X = x) + P(CN|X = x)] - [P(TN|X = x) + P(CR|X = x)]$$

Сумуються ймовірності потрапляння до класів TR та CN, так як вони містять у собі переконуваний тип клієнтів та віднімаємо ймовірність потрапляння у класи TN та CR так вони містять у собі, тип клієнтів яких краще не турбувати. Якщо вибірка не сильно збалансована пропонується рахувати uplift наступним чином:

$$uplift = \left[\frac{P(TR|X=x)}{r(T)} + \frac{P(CN|X=x)}{r(C)} \right] - \left[\frac{P(TN|X=x)}{r(T)} + \frac{P(CR|X=x)}{r(C)} \right]$$

де $r(T)$ - доля клієнтів в тестовій групі, $r(C)$ - доля клієнтів в контрольній групі.

Методи на основі однієї моделі

Найпростіше рішення задачі побудова моделі яка буде навчатись одночасно на двох групах, при цьому додається бінарний флаг наявності комунікації. Кожен об'єкт записується двічі - з флагом комунікації 1 та 0.

$$fit \left(\begin{matrix} x_{11} & \cdots & x_{1k} & w_1 & y_1 \\ \vdots & \ddots & \vdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{nk} & w_n & y_n \end{matrix} \right)$$

$X_{train} \quad W_{train} \quad Y_{train}$

де $x_{11} \dots x_{nk}$ - усі клієнти,

$w_1 \dots w_n$ - флаг належності до цільової групи,

$y_1 \dots y_n$ - цільова змінна.

Далі за стандартною схемою рахується різниця для передачень з флагом 1 та 0.

$$\begin{aligned} model \begin{pmatrix} x_{11} & \dots & x_{1k}, & 1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & 1 \end{pmatrix} + model \begin{pmatrix} x_{11} & \dots & x_{1k}, & 0 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mk}, & 0 \end{pmatrix} = \\ = \begin{pmatrix} u_1 \\ \dots \\ u_m \end{pmatrix} \end{aligned}$$

Також, можна збільшувати кількість ознак об'єкта, додавши добуток кожної ознаки на флаг комунікації.

Прогнозування напрямку

В даному методі пропонується модифікувати існуючі алгоритми машинного навчання. Найпопулярнішими методами для такого застосування є методи засновані на деревах рішень, тому тут буде наведено приклад саме для них [1]. Для задачі uplift моделювання в деревах змінюється критерій розподілення. Дерево повинно будуватися так, щоб максимізувати різницю між цільовою змінною у контрольній та цільовій групах.

Для кожного об'єкта ефект від взаємодії рахується наступним чином:

$$\hat{\tau} = \frac{\sum_i Y_i^{obs} W_i}{\sum_i W_i} - \frac{\sum_i Y_i^{obs} (1-W_i)}{\sum_i (1-W_i)}$$

де Y_i - реакція, що спостерігається,

$w_1 \dots w_n$ - флаг належності до цільової групи,

τ - ефект від комунікації.

Один із можливих критеріїв розділення дерева записується наступним чином [8]:

$$\Delta_{gain} = D_{after-split}(P^T, P^C) - D_{before-split}(P^T, P^C)$$

де P^T та P^C розподіл цільової змінної в контрольній та цільовій групі, а D - дивергенція між двома розподілами. Зазвичай використовують дивергенцію Кульбака-Лейблера, евклідову відстань або хі-квадрат:

$$KL(P: Q) = \sum_{k=Left.Right} p_k \log \frac{p_k}{q_k}$$

$$E(P: Q) = \sum_{k=Left.Right} (p_k - q_k)^2$$

$$\chi^2(P: Q) = \sum_{k=Left.Right} \frac{(p_k - q_k)^2}{q_k}$$

де k індикатор листка, а розподіли представлені як p_k . Якщо на вершині залишається представники однієї групи, то критерії стають стандартними для дерев. Також важливо, щоб при розділенні різниця між кількістю об'єктів яка потрапила в кожну "гілку" була мінімальна. Для того, щоб контролювати кількість об'єктів, які потрапляють у різні групи існує ще один критерій розбиття [9]:

$$D_{weighted}(P^T, P^C) = \frac{n_L \cdot D(P_L^T, P_L^C) + n_R \cdot D(P_R^T, P_R^C)}{n_L + n_R}$$

де n_L та n_R - кількість об'єктів, які потрапили в різні вершини, P_L^T , P_R^C - розподіл цільової та контрольної групи для різних вершин.

Дерево з такими критеріями розбиття можна використовувати в якості базового алгоритму в ансамблях, а також використовувати для них стандартні методи уникнення перенавчання. Існують також більш складні модифікації для не бінарної класифікації [8].

1.5 Стандартні метрики uplift моделювання

Для задачі uplift моделювання не можна використовувати звичайні метрики. Стандартно в машинному навчанні для оцінки моделі використовується кросс-валідація: розділяють дані на навчаючу вибірку та тестову. Модель навчається на першій вибірці, прогнозує цільову зміну тестової, потім результат прогнозу порівнюється з реальними значеннями. Для uplift моделювання немає реальних значень uplift, тобто передбачення моделі нема з чим порівняти. В цьому і полягає основна особливість uplift моделювання: не можна одночасно взаємодіяти і не взаємодіяти з клієнтом і подивитись різницю його реакцій. Класичні метрики зазвичай рахуються з групуванням об'єктів або на основі кривих.

Метрика “uplift@k”

Найпростіша метрика для цієї задачі - величина uplift на топ k відсотках вибірки. Тобто, нехай за допомогою uplift моделі нам потрібно відібрати k% клієнтів, з якими буде комунікація. Тоді нам цікаві результати прогнозу не на усій тестовій вибірці, а тільки на об'єктах з найбільшим передбаченням. Для застосування цієї метрики потрібно відсортувати вибірку за величиною передбаченого uplift і подивитись різницю середніх значень таргету в цільовій та контрольній групах.

$$\begin{aligned} uplift@k &= response\ rate@k_{(treatment)} - response\ rate@k_{(control)} \\ response\ rate@k &= mean(Y@k) \end{aligned}$$

де Y_k - таргет на k% групи,

uplift - результат моделювання реакції на маркетингову дію.

Дана метрика може приймати значення від -1 (коли в цільовій групі немає реакції, а в контрольній групі всі виконали цільову дію) до 1 (коли в цільовій групі всі виконали цільову дію, а в контрольній групі не було реакції). Проте, на практиці дана метрика зазвичай приймає значення від 0 до 1, в залежності від обраного k, особливостей вибірки та якості моделі.

Розраховувати метрику можна двома способами: спочатку сортувати вибірку за відсортованими значеннями uplift, а далі рахувати різницю таргету на двох групах, або навпаки сортувати клієнтів з цільової та контрольної групи окремо.

Метрика Uplift за процентилям

Дана метрика схожа на попередню, проте вона використовується в тих випадках, коли завчасно невідомо процент клієнтів k і потрібно подивитись, як модель буд себе вести при її різних значеннях. [11]

Алгоритм обрахунку цієї метрики наступний:

1. Дані сортуються за передбаченням uplift.
2. Відсортована вибірка ділиться на процентилях.
3. В кожному процентилях окремо оцінюється uplift як різниця між середніми значеннями цільової змінної в тестовій та контрольній групі.

Результатом даного алгоритму буде таблиця, в якій є наступні показники: розмір цільової та контрольної групи, перцентиль, середнє значення таргету в цільовій та контрольній групах, значення uplift. Якщо візуалізувати даний результат у графік, можна зручно оцінити ефективність моделі. Комунікація буде проводитись з тими клієнтами, які отримали високу оцінку uplift, тобто в тих випадках, коли різниця між цільовою та контрольною групою найбільша. Якщо різниця в оцінці двох груп велика, то там знаходяться саме переконувати тип клієнтів, який нам і треба знайти. Якщо значення uplift будуть від'ємні, це означає, що комунікація в цільовій групі дала негативний ефект і модель знайшла тип клієнтів, яких краще не турбувати. Добре побудована модель буде показувати високі значення в перших квантилях, та далі зменшення. Для прикладу, графік uplift зображено на рис. 1.2:

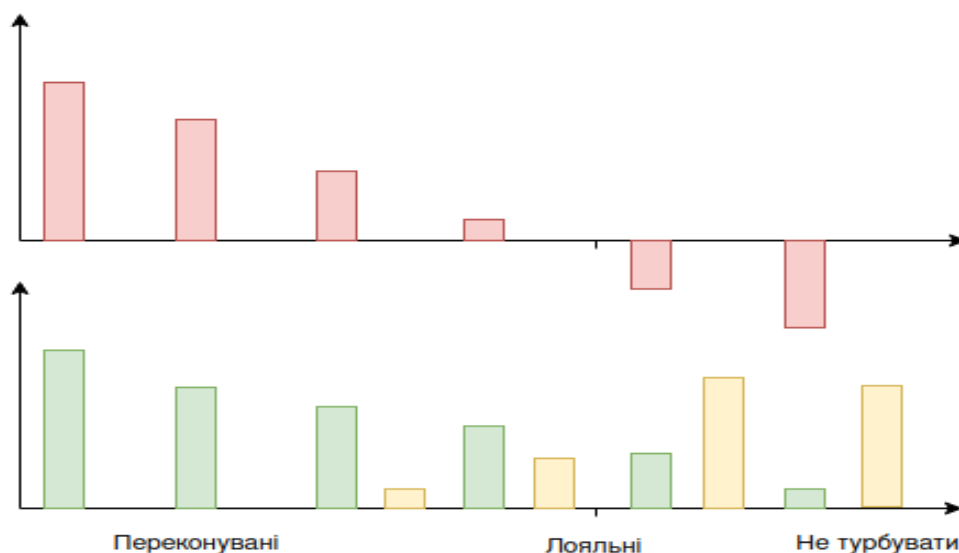


Рисунок 1.2 - Пояснення результатів моделювання для розподілення клієнтів

Зважений середній uplift

Якщо потрібно порахувати метрику в вигляді єдиного значення для всієї вибірки використовують зважене середнє значення uplift.[12] Для неї так само, як і в попередній спочатку створюють таблицю з наченнями uplift для різних персентилів, а потім рахують середньозважене значення, де ваги це процент об'єктів в цільовій групі:

$$weighted\ average\ uplift = \frac{1}{\sum_i N_i^T} \sum_i N_i^T uplift_i$$

де N_i^T - кількість клієнтів в групі i .

Дана метрика приймає значення від -1 до 1. Якщо метрика приймає значення 1, це означає що клієнти з контрольної групи ніколи не виконують цільову дію самостійно, а тільки при комунікації. За такого значення метрики

краще використовувати не uplift моделювання, а більш просту модель, наприклад look-alike.

Якщо значення метрики -1 , це означає, що в цільовій групі не було реакції. Тому значення метрики, які показують, що модель може дати хороші результати знаходяться в проміжку $[0,1)$.

Крива Uplift

Крива uplift - це функція від кількості об'єктів. Монотонне зростання кривої вказує на те, вплив усієї вибірки має позитивний ефект. Зміна напрямку кривої вказує на позитивні та негативні ефекти в даних. Якби їх не було, графік був би близький до базової лінії. Це корисні дані для маркетологів, щоб розуміти чи має їх кампанія глобальний позитивний чи негативний ефект і чи можуть вони отримати більше користі, працюючи з цією групою людей. Чим крива вище, тим краще. Функція кривої записується таким чином:

$$uplift\ curve(t) = \left(\frac{Y_t^T}{N_t^T} - \frac{Y_t^C}{N_t^C} \right) (N_t^T + N_t^C)$$

де Y_t^T, Y_t^C - реакція в цільовій та контрольній групах, t - накопичена кількість об'єктів, N_t^T, N_t^C - розмір цільової та контрольної групи.

Крива Qini

Uplift крива визначається різницею між значенням uplift на двох вибірках. Це не завжди найкращий метод оцінки моделі, оскільки немає гарантії, що результат на контрольній та цільовій вибірці співпадають. Інша крива для оцінки uplift моделі це Qini крива, яка визначається таким чином [13]:

$$qini\ curve(t) = Y_t^T - \frac{Y_t^C N_t^T}{N_t^C}$$

де Y_t^T, Y_t^C - реакція в цільовій та контрольній групах,

t - накопичена кількість об'єктів,

N_t^T, N_t^C - розмір цільової та контрольної групи.

Qini розраховується одразу для всієї вибірки даних. Чим вище крива, тим краще передбачення. Фізична сторона функції в тому, щоб крива збільшувалась не тільки на цільовій групі, для цього є множник N_t^T/N_t^C .

Між uplift та qini кривими є взаємозв'язок. Якщо контрольна та цільова групи будуть збалансовані, то крива qini буде в 2 рази нижче uplift:

$$uplift\ curve(t) = \frac{qini\ curve(t)(N_t^T + N_t^C)}{N_t^T}$$

де t - накопичена кількість об'єктів,

N_t^T, N_t^C - розмір цільової та контрольної групи

Метрики на основі кривих

За допомогою кривих не можна оцінити та порівняти моделі, для більш точного аналізу треба рахувати метрики. Для розрахунку метри заснованих на кривих, рахують як площу під кривою та нормувати її під площу ідеальної кривої [14]. При цьому часто з обох площ віднімається площа базової лінії. Якщо базова лінія зростаюча означає, що з цієї групи людей можна отримати користь від маркетингової кампанії.

Для qini кривої метрика називається AUQC, а для uplift кривої AUUQ. Схематично AUQC зображено на рис. 1.3:

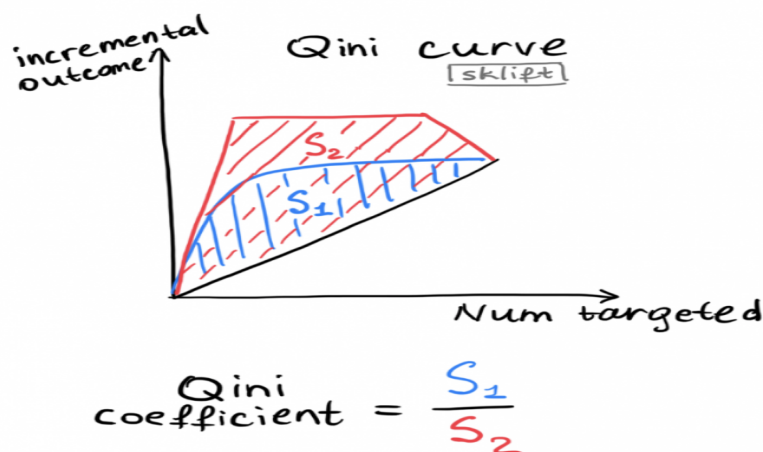


Рисунок 1.3 - Схематичне зображення розрахунку метрики на основі кривої.

Дані метрики добре підходять тільки для випадків однієї дії та бінарної відповіді.

Метрика зі зміненою цільовою

Для методу трансформації класів використовувалась модифікована цільова змінна, за формулою (13). В деяких випадках [1] рахують середньоквадратичну помилку між передбаченим причинно-наслідковий ефект та цільовою змінною, яка вже обчислена.

$$MSE(Y_i^*, \hat{\tau}) = \sum_i^n \frac{1}{n} (Y_i^* - \hat{\tau}_i)^2$$

де Y_i - реакція, що спостерігається,

τ - ефект від комунікації.

Формула є апроксимацією:

$$MSE(\tau_i, \hat{\tau}_i) = \sum_i \frac{1}{n} (\tau_i - \hat{\tau}_i)^2$$

Дану метрику можна розрахувати лише на змодельованих даних, коли ми знаємо величину причинно-наслідкового ефекту. Так, як звичайне MSE не можна порахувати, була запропонована оцінка. Оцінка може використовуватись тільки у випадку коли модель - дерево рішень [15].

1.6. Висновки до розділу

В першому розділі розділі розглянуто актуальність моделювання маркетингового впливу. Розглянути різні підходи до цієї задачі: моделі схожості,

моделі відповідей, uplift моделювання. Перші два типи моделей оцінюють ймовірність того, що клієнт буде схильним до покупки товару. Проте для впровадження маркетингових дій цього замало - компанії більш вигідно витратити гроші тільки на клієнтів, які б без маркетингової компанії не купили товар.

Була проведена класифікація усіх типів клієнтів. Та пояснено, які типи клієнтів є цільовою аудиторією для маркетингових дій. Для оцінки чистого ефекту від взаємодії використовують uplift моделювання. Даним методом можна знайти саме клієнтів, яких можна переконати.

Наведені основні класичні підходи до uplift моделювання: двух-модельний підхід, одномодельний, трансформація класів, методи на основі дерев рішень. Також розглянуті класичні метрики для uplift моделювання.

Розділ 2 ВИЗНАЧЕННЯ МАРКЕТИНГОВОГО ВПЛИВУ ЗА ДОПОМОГОЮ МАРКІВСЬКИХ ЛАНЦЮГІВ

2.1 Використання марківських ланцюгів

Найбільш популярні підходи до uplift моделювання це двомодельний підхід та методи на основі дерев рішень. Проте, на реальних даних такі моделі показують не дуже хороший результат, а існуючі метрики не дають змоги моделювати uplift за наявності декількох дій (більше двох).

Використовуючи марківські ланцюги маємо змогу створити нове бачення на задачу та отримати кращі результати. Оскільки такий підхід відрізняється від попередніх, необхідно ввести деякі нові поняття та метрики, які будуть працювати для нової побудованої моделі.

2.2 Марківські ланцюги

Випадковий процес це набір випадкових величин, індексованих множиною T , яка зазвичай позначає різні моменти часу. Часто, множина індексів T буває множиною натуральних чисел (дискретний час) або дійсних чисел (неперервний час). Випадкові величини можуть бути залежні та незалежні один від одного, простір станів випадкових величин теж може бути дискретним або неперервним. Різні сімейства випадкових процесів мають певні властивості, які дозволяють краще їх досліджувати та розуміти. Марківські процеси - випадкові процеси, які задовольняють Марківську властивість. Перехід процесу з одного стану в інший називають кроком процесу.

Марківська властивість: в будь-який момент часу умовний розподіл наступних станів процесу з заданими поточними і минулими станами залежить тільки від поточного стану. Тобто:

$$\begin{aligned} P(\text{майбутній стан} \mid \text{поточний стан, минулий стан}) = \\ = P(\text{майбутній стан} \mid \text{поточний стан}) \end{aligned}$$

Однорідні ланцюги Маркова - марковський процес з дискретним часом та простором станів. Формально, ланцюг маркова позначають наступним чином:

$$X = (X_n)_{n \in N} = (X_0, X_1, X_2, \dots)$$

де в кожний момент часу процес бере свої значення з дискретної множини E , такої що $X_n \in E$.

Тоді марківська властивість означає наступне:

$$\begin{aligned} P(X_{n+1} = S_{n+1} \mid X_n = S_n, X_{n-1} = S_{n-1}, X_{n-2} = S_{n-2}, \dots) = \\ = P(X_{n+1} = S_{n+1} \mid X_n = S_n) \end{aligned}$$

де $S_1 \dots S_n$ - стани ланцюга,

$n_1 \dots$ - момент часу.

Завдяки марківській властивості, динаміку ланцюга маркова досить просто визначити. Для цього потрібні початковий розподіл ймовірностей та перехідні ймовірності. Початковий розподіл ймовірностей безумовний та позначається, як:

$$P(X_0 = s) + q_0(s) \quad \forall s \in E$$

де s - стан в якому ланцюг перебуває в початковий момент часу,

q - ймовірність такої події.

Умовну ймовірність того, що $P\{X_{k+1} = j | X_k = i\}$ називають однокроковою перехідною ймовірністю з стану i в стан j .

$$P(X_{n+1} = S_{n+1} | X_n = S_n) = p(S_n, S_{n+1}) \quad \forall (S_{n+1}, S_n) \in E \times E$$

де $s_1 \dots s_n$ - стани ланцюга, $n_1 \dots$ - момент часу.

Умовну ймовірність називають $P\{X_{k+n} = j | X_k = i\}$ n -кроковою з стану i в стан j . Зазвичай перехідні ймовірності залежать не тільки від початкового та кінцевого стану, а і від моменту переходу. Ланцюги в яких перехідні ймовірності залежать тільки від початкового та кінцевого стану називають стаціонарними. За визначенням для них буде вірно наступне:

$$P_{ii}(0) = 1, P_{ij}(0) = 0, j \neq i$$

Матриця переходу ймовірностей, складається з ймовірностей того, що в момент часу $n+1$ стан є наступним для стану в момент часу n для будь-якої пари станів) позначають так:

$$\mathbb{P} = [P_{ij}] = \begin{vmatrix} P_{00} & P_{01} & \dots \\ P_{10} & P_{11} & \dots \\ \dots & \dots & \dots \end{vmatrix}$$

де P_{ij} - ймовірність переходу з i в j .

Матриці переходу можуть бути однокроковими та багатокроковими. Якщо відомі початковий розподіл та матриця переходу, тоді можна визначити ймовірність будь-якого результату процесу. Матриці переходу стохастичні і в

них є дві основні властивості: невід'ємність усіх елементів та сума по рядку або стовпчику дорівнює 1:

$$1^{\circ} P_{ij}(n) \geq 0, i, j = 0, 1, \dots,$$

$$2^{\circ} \sum_j P_{ij}(n) = 1, i = 0, 1, \dots$$

Наслідком Марківської властивості є рівняння Колмогорова - Чепмена, яке виконується для будь-яких станів i, j з множини X та будь-яких цілих моментів часу r, s :

$$P_{ij}(r + s) = \sum_{k \in X} P_{ik}(r) P_{kj}(s)$$

де P_{ij} - ймовірність переходу з i в j , r, s - момент часу, k - індекс, який проходиться по усім странам.

В матричному вигляді:

$$P(r + s) = P(r)P(s)$$

З цього рівняння можна отримати, що за допомогою матриці однокрокових ймовірностей завжди можна отримати матрицю багатокрокових перехідних ймовірностей:

$$P(n) = (P)^n$$

Тобто матриця в момент часу n буде початковою матрицею в степені n .

Для того, щоб продемонструвати, як саме будуються прогнози марківськими ланцюгами, розглянемо приклад. Нехай є система, яка може

приймати два стани: S та R. Відомо, що перші 6 станів системи були такими: E = 'SRRSRS' і треба визначити які стани будуть наступними. Початковий розподіл ймовірностей, враховуючи, що відомо що останній стан був S буде таким $v(0) = [P[S] \ P[R]] = [1 \ 0]$. Ймовірності переходу із стану i в j позначатимуться через P_{ij} , де i, j належать множині $\{S, R\}$, а матриця ймовірностей позначимо через P . Тоді, наступний стан системи буде визначатись як $v(1) = v(0) * P = [P1(S) \ P1(R)]$. Для прогнозу обирається той стан у якого буде більша ймовірність. Усі наступні стани рахуються аналогічним чином. Проте виникає питання скільки кроків можна передбачити таким чином.

Ланцюги пояснюють поведінку випадкового процесу, тому з їх допомогою добре прогнозувати майбутні дії будь-якої системи. Проте, якщо ланцюг з часом починає збігатись до стаціонарного вектора, сила прогнозування зменшується. Прикладом таких ланцюгів наведений у 18, та зображено на рис. 2.1:

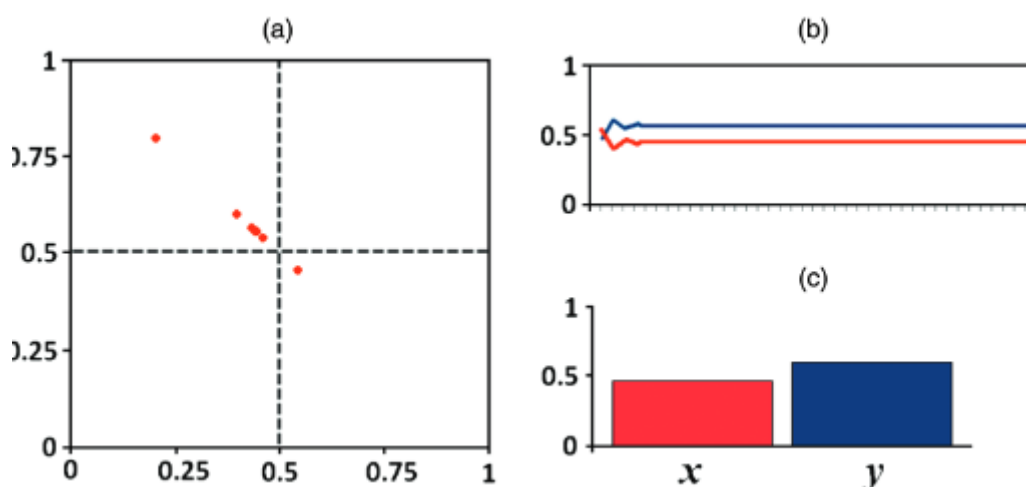


Рисунок 2.1 - Приклад марківського ланцюга, який має схильність збігатися до одного значення[18].

на (a) вертикальна вісь позначає значення компоненти x, а горизонтальна значення для компоненти y. Рисунок (б) показує поведінку вектора ймовірностей на 40 кроках. Червона лінія відповідає за компоненту x, а синя за компоненту y.

Ланцюги маркова охоплюють широкий спектр практичних задач та використовуються в різних областях. Зокрема зараз марківські ланцюги

використовуються задачах моделювання пов'язаних з медициною, прогнозуванням часових рядів, прогнозування погоди та багато іншого.

2.3 Марківський процес прийняття рішень

Марківський процес прийняття рішень це розширення марківського ланцюга. Це математична основа для моделювання ситуацій прийняття рішень де результати частково випадкові, частково контролюються. В широкому сенсі марківський процес прийняття рішень охоплює велику кількість моделей. Вводиться поняття агент - це сутність, яку в процесі моделювання навчають вибрати правильні дії.

Марківський процес з скінченним набором станів та дискретним часом має такі компоненти:

- S - скінченна множина всіх можливих станів системи, яка також називається простором станів або середовищем. Агент взаємодіє з середовищем, про не може їм маніпулювати.
- A - це скінченний набір усіх дій, які може здійснити агент. Дія це вибір який робить агент. На будь-якій ітерації агент повністю знає стан системи.
- D - скінченна або нескінченна послідовність натуральних чисел які позначають момент часу.
- $T: S \times A \times S \times D \rightarrow [0, 1]$ - перехідна функція, відображення, що визначає ймовірність $T(s_1, a, s_2, t)$ переходу в стан s_2 , якщо дія a виконується, коли агент знаходиться в стані s_1 в момент часу крок t ;
- $R: S \times A \times S \times D \rightarrow R$ — функція винагороди, яка дає кінцеве числове значення винагороди $R(s_1, a, s_2, t)$, отримане при переході системи зі стану s_1 у стан s_2 в результаті виконання дії a на момент часу t . При виконанні дії в системі отримується винагорода.

Відображення T та R залежать лише від поточного стану системи, та не залежать від попередніх кроків. Для того, щоб вирішити дану задачу, треба обрати спосіб вибору дій при якому система буде контролюваною. Агент не вибирає стан, до якого перейде система. Спосіб поведінки агента в певний момент часу визначає політика - відображення від станів середовища до дій які треба виконати. Формально, політика є розподілом ймовірностей за набором дій, враховуючи поточний стан.

Функція винагороди визначає мету задачі. Ця функція переводить стан системи в одне число, що вказує на те чи є цей стан хорошим чи ні в контексті поставленої задачі. Мета моделювання визначити оптимальну політику, за якої буде максимізована загальна сума винагород. Обравши дію, яка приносить більшу винагороду можна втратити більшу загальну винагороду в довгостроковій перспективі. Тому вводиться змінна яка позначає “знижку” $\gamma \in [0, 1]$. Її вводять для того, щоб зменшити внесок подальших винагород в загальну суму.

Також вводяться дві функції вартості, які визначають наскільки вигідно бути в певному стані або виконувати певну дію з визначеною політикою. Вони визначають довгостроковий виграш від стану або дії. Формально, функція вартості записується так[21]:

$$v_{\pi}(S) = E_{\pi}[R_{t+1} + \gamma v_{\pi}(S_{t+1}) | S_t = s]$$

де s - поточний стан,

R_{t+1} - виграш від перебування в наступних станах,

v_{π} - функція вартості,

γ - фактор знижки.

Для функцій вартості часто використовують рівняння Беллмана, яке отримується з визначення функції вартості. Розв'язок цього рівняння знаходять в такому в такому вигляді:

$$v_{\pi}(s) = \sum_{a \in A} \pi(a|s) \left(R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_{\pi}(s') \right)$$

де s - поточний стан,

a - дія,

R_{t+1}^a - виграш від наступних дій в відповідних станах,

v_{π} - функція вартості,

γ - фактор знижки,

π - політика вибору дій.

Таким чином ми можемо знайти значення довгострокового виграшу для всіх можливих політик, проте нам потрібно знайти найбільш оптимальне рішення.

За допомогою марківського процесу прийняття рішень можна вирішити багато проблем навчання з підкріпленням, для цього треба визначити множину станів та дій. Для отримання оптимального рішення марківського процесу прийняття рішень використовують такі алгоритми, як: Value Iteration, Policy Iteration, SARSA, Q-Learning.

2.4 Визначення uplift моделювання для кількох дій

Для того, щоб використовувати марківські ланцюги для uplift моделювання, потрібно переоформлювати дану задачу як марківський процес та ввести метрику для неї. В такому формулюванні задачі не потрібно явно ділити дані на 2 вибірки, для цього використовуються негативне та позитивне винагородження, для того щоб визначати яка дія буде правильною в конкретному середовищі.

Для початку треба ввести основні елементи та їх позначення uplift моделювання, у випадку кількох дій щодо окремих об'єктів. Отже, маємо:

- X : змінна індивідуального вектора ознак. Зазвичай, це ознака одного клієнта.
- $a \in \{0, 1, \dots, K\}$: закодована дія. Зокрема, $a = 0$ означає відсутність дії.
- $\pi(\cdot)$: політика вибору дій для кожної ознаки. Використовуємо $\pi(x) = a$ позначення відображення того, що політика обирає дію a для x , а $\pi(a | x)$ відповідно ймовірність такої реалізації.
- $Y(x, a)$: спостережувана реакція на дію, коли x отримує дію a . Реакція буде цілим числом.
- $B(x)$: реакція X , коли він не отримує жодної дії.
- $L(x, a)$: реакція коли X отримує дію a .

З такими позначеннями адитивна модель для uplift моделювання записується наступним чином:

$$L(x, a) = Y(x, a) - B(x)$$

Тобто uplift це різниця реакції на комунікацію та реакції без комунікації. Коли взаємодії не було $Y(x, 0) = B(x)$ і $L(x, 0) = 0$. Задача uplift моделювання знайти політику $\pi(\cdot)$ для максимізації uplift, тобто:

$$\max_{\pi} \mathbb{E}_x[L(X, \pi(X))]$$

2.5 Метрика uplift моделювання для кількох дій

Для знаходження політики для максимізації uplift потрібно ввести незміщену метрику для реакції на будь-яку дію. Часто для оцінки реакції

використовується A/B тестування, проте в реальних задачах це може бути дорого. Тому пропонується використовувати метрику, яка спирається на дані з попереднього експерименту. Нехай є вибірка з N записів та K груп з різними діями і контрольна група без дій. Кожен елемент був призначений до однієї з $K+1$ групи за політикою p . Наприклад, в попередній дії експерименту для об'єкта X була обрана дія A , і його реакція була записана. Ймовірність того, що політика обрала дію A для об'єкта X буде позначатись так: $p(a | x)$, де $\sum p(a | x) = 1$. В даному випадку політика $p(\cdot)$ не повинна бути незалежною від $Y(x, a)$.

На основі даних вводиться незміщена оцінка реакції uplift для конкретної політики π . Це робиться шляхом оцінки очікуваних реакцій на дії для політики (лема 1) [15] та реакції без взаємодії (лема 2) [15] окремо. Надалі, для позначення ймовірності події буде використовуватись $\Pr(\cdot)$, $I\{\cdot\}$ - функція індикатор. Також, $\pi(x) = a$ означає, що політика обрала дію a для об'єкта x , а $p(x) = a$ означає, що дія a була обрана для x в реальних даних. Отже, для введення метрики потрібно розглянути дві леми та теорему, яка з них слідує.

Лема 1: Маючи політику дій π , для кожної дії $a \in \{0, 1, \dots, K\}$, нехай існує змінна:

$$Z^T(a, X) = \frac{Y(X, a) I\{p(X)=a\}}{p(a|X)}$$

Тоді умовне очікування цієї змінної можна обчислити наступним чином:

$$E_x[Z^T(a, X) | \pi(X) = a] = E_x[Y(X, \pi(X)) | \pi(X) = a]$$

Доведення леми наведено в [19] і воно майже повністю повторює доведення для показник зворотної схильності з 20. З даної леми можна отримати оцінку загальної реакції для об'єкта X :

$$\sum_{a=0}^K E_x[Z^T(a, X)|\pi(X = a)] = E_x[Y(X, \pi(X))]$$

Зокрема, характер реакції клієнтів щодо конкретної політики можна аналогічно оцінити. Не зважаючи на те, що загальна реакція всіх клієнтів є константою та не залежить від політики, її все одно треба оцінити, щоб ввести метрику для uplift реакції.

Лемма 2: Маючи політику дій π , для кожної дії $a \in \{0, 1, \dots, K\}$, нехай існує змінна:

$$Z^C(X) = \frac{Y(X, 0)I\{p(X)=0\}}{p(0|X)}$$

Тоді її можна оцінити, як умовне очікування початкової реакції на продукт:

$$E_x[Z^C(X)|\pi(X) = a] = E_x[B(X)|\pi(X) = a]$$

та

$$E_x[Z^C(X)] = E_x[B(X)]$$

Доведення лемми 2 схоже до доведення лемми 1, окрім використання того факту, що $Y(x, 0) = B(x)$ для $\forall x$. Початкова реакція розглядається для кожної дії окремо, оскільки тому що кожна з них залежить від політики. Щоб порахувати середнє за вибіркою її реалізації допоможе у визначенні градієнта політики дій.

Наслідком лемми 2 та лемми 1 є теорема про незміщену оцінку політики. Як відомо реакцію uplift можна рахувати як різницю між реакцією на дію та натуральною реакцією.

Теорема 1: Маючи політику дій π , очікувана реакція uplift від політики π буде:

$$E_x[L(X, \pi(X))] = E_x[Y(X, \pi(X))] - E_x[B(X)]$$

Нехай x_i ознаки i -го об'єкта та N кількість об'єктів. Тоді незміщена оцінка $E_{X, \pi(X)}[L(X, \pi(X))]$ буде:

$$\bar{Z}(\pi) = \frac{1}{N} \sum_{i=1}^N \sum_{a=0}^K z^T(a, x_i) I\{\pi(x_i) = a\} - \frac{1}{N} \sum_{i=1}^N z^C(x_i)$$

Надалі, незміщена оцінка $\bar{Z}(\pi)$ буде називатись загальна метрика uplift моделювання (UMG). Враховуюч теорему 1 на нерівність чебишева, якщо дисперсія UMG σ^2 , тоді:

$$|\bar{Z}(\pi) - E_{x, \pi(X)}[L(X, \pi(X))]| \leq \frac{\sigma}{N\sqrt{\epsilon}}$$

де $\Pr(\cdot)$ позначає ймовірності випадкової події, а $I\{\cdot\}$ - це 0/1 функція індикатора. І $\pi(x) = a$ являє собою реалізацію, яку вибирає політика π дія a для x , тоді як $p(x) = a$ представляє той факт, що дія a виконується над x .

Така незміщена оцінка може бути метрикою для будь-якої моделі uplift з декількома діями. Іншими словами, мета задачі знайти політику, яка будем максимізувати uplift.

Дану метрику можна нормалізувати, це може бути необхідно для контролю дисперсії. Нормалізована метрика буде мати наступний вигляд:

$$\hat{Z}^{SN}(\pi) = \frac{\sum_{i=1}^N \sum_{a=0}^K z^T(a, x_i) I\{\pi(x_i)=a\}}{\sum_{i=1}^N \frac{I\{\pi(x_i)=a_i\}}{p(a_i|x_i)}} - \frac{\sum_{i=1}^N z^C(x_i)}{\sum_{i=1}^N \frac{I\{a_i=0\}}{p(0|x_i)}}$$

де x_i – це ознака об'єкта, a_i - дія, яка була виконана над ним. Така оцінка буде мати зміщення $1/N$, яке можна ігнорувати для великих N , і оцінка буде асимптотично незміщена. Така метрика буде позначатись SN-UMG.

Обидві метрики, введені в цьому розділі, можуть бути метриками для оцінки різних моделей uplift моделювання. Також, SN-UMG може використовуватись для оцінки значення винагороди в марківському процесі прийняття рішень.

2.6 Зв'язок між uplift моделюванням та задачею “контекстних бандитів”

Задача багаторуного бандита це класична задача начання з підкріпленням. Вона полягає в тому, щоб на кожному кроці обирати одну з дій, в кожній з яких свій розподіл ймовірності успіху. Проблема в тому, що розподіл ймовірностей успіху невідомий завчасно тому треба балансувати між тим, щоб постійно вибрати найкращу дію та тим, що обирати інші дії з метою дізнатись більше про їх розподіл. Мета - поступово обирати дії так, щоб максимізувати загальний прибуток від дій. В звичайній задачі багаторуких бандитів вибір робиться випадковим не спираючись на інформацію про стан середовища. В задачі контекстних бандитів вибір дії робиться на основі інформації про стан середовища. Така модель оптимізує свої рішення на основі попередніх спостережень та персоналізує їх для комножного стану середовища. Таку модель можуть використовувати для того, щоб вирішувати яку саме рекламу показувати клієнтам на сайті.

Враховуючи схожість формулювань задач треба перевірити їх еквівалентність. Політика π^* яка максимізує uplift, буде максимізувати і реакцію на певну дію, тобто:

$$\begin{aligned}
\pi^* &= \operatorname{argmax}_{\pi} E_{x, \pi(X)} [L(X, \pi(X))] = \\
&= \operatorname{argmax}_{\pi} E_{x, \pi(X)} [Y(X, \pi(X)) - B(X)] = \\
&= \operatorname{argmax}_{\pi} E_{x, \pi(X)} [Y(X, \pi(X))]
\end{aligned}$$

В задачах контекстуального бандита також розглядається така проблема. В ній оптимізується політика для виконання певних дій на об'єкт на основі даних з попереднього кроку моделювання. Так само, як і в uplift моделюванні, для кожного об'єкта (з певним набором ознак) виконується одна дія, знаючи відповідну реакцію. З точки зору, пошуку рішення, задача контекстуального бандита та uplift моделювання еквівалентні, за умови, якщо відсутність взаємодії вважається окремою дією. Оскільки знайти оптимальну оцінку неможливо, шукається наближена оптимальна оцінка. За теоремою 2 з [17]: якщо початкової реакції немає, тобто $E[B(X)] = 0$, то наближена максимізація uplift еквівалентна максимізації реакції на дію і задачу uplift моделювання можна розглядати, як задачу контекстуального бандита. Проте в більшості випадків початкова реакція буде наявною, а за цієї умови задачі не еквівалентні. Тому краще одразу розглядати моделювання саме реакції uplift.

2.7 Метод градієнта політики

Велика кількість теорії ґрунтується на гіпотезі винагороди, яка стверджує, що ціль навчання з підкріпленням може бути пояснена одним скаляром, який називається винагородою. Більш формально гіпотеза формулюється так: “все мається на увазі під цілями та задачами можна вважати максимізацією очікуваного значення сукупної суми винагороди”.

Необхідно максимізувати винагороду за умови, що дії обираються певною політикою, тому для вирішення задачі змінюють параметри політики дій. Як і в

будь-якій задачі машинного навчання, вводиться набір параметрів θ (як наприклад коефіцієнти поліномів або ваги нейронної мережі) для параметризації політики - π_θ . Позначимо загальну винагороду для послідовності певного набору дій та станів τ як $r(\tau)$. Тоді задачу максимізації очікуваної винагороди можна записати наступним чином:

$$J(\theta) = E_\pi[r(\tau)]$$

Кожен скінченний марківський процес прийняття рішень має хоча б одну оптимальну політику дій(тобто ту, за якої можна досягти максимальної винагороди). Та серед оптимальних політик хоча б одна стаціонарна та детермінована.

Як і в будь-якій задачі машинного навчання, якщо знайдено параметри θ^* , які максимізують J задача вирішена. Стандартний підхід до такої задачі максимізації це використання методу градієнтного спуску. В цьому методі параметри перебираються використовуючи наступне правило:

$$\theta_{t+1} = \theta_t + \alpha \nabla J(\theta_t)$$

В цьому методі виникає питання, як знайти градієнт цільової функції. Оскільки інтеграли зазвичай погано обчислюються, шукаються способи обійти їх. Для початку можна переформулювати градієнт:

$$\begin{aligned} \nabla E_\pi[r(\tau)] &= \nabla \int \pi(\tau) r(\tau) d\tau = \\ &= \int \nabla \pi(\tau) r(\tau) d\tau = \\ &= \int \pi(\tau) \nabla \log \pi(\tau) r(\tau) d\tau \end{aligned}$$

Звідси можна отримати теорему про градієнт політики дій: “похідною від очікуваної винагороди є математичне сподівання добутку винагороди та градієнта логарифма політики дій π_θ ”, тобто:

$$\nabla E_{\pi_\theta}[r(\tau)] = E_{\pi_\theta}[r(\tau) \nabla \log \pi_\theta(\tau)]$$

Познимо ймовірність того, що початковий стан буде s_0 , як $P(s_0)$. Оскільки ймовірність наступних дій не залежить від попередніх, використовується правило добутку. Параметризовану політику дій визначається таким чином:

$$\pi_\theta(\tau) = P(s_0) \prod_{t=1}^T \pi_\theta(a_t | s_t) p(s_{t+1}, r_{t+1} | s_t, a_t)$$

На кожному кроці робиться певна дія, яку обирає політика π_θ та стан змінюється. Це все перемножується T раз (стільки скільки робиться кроків). Якщо логарифмувати, отримаємо такий результат:

$$\begin{aligned} \nabla \log \pi_\theta(\tau) &= \sum_{t=1}^T \nabla \log \pi_\theta(a_t | s_t) \Rightarrow \\ \Rightarrow \nabla E_{\pi_\theta}[r(\tau)] &= E_{\pi_\theta} \left[r(\tau) \left(\sum_{t=1}^T \nabla \log \pi_\theta(a_t | s_t) \right) \right] \end{aligned}$$

З цього слідує, що немає потреби знати початковий розподіл станів та транзитивні ймовірності. Цей результат дуже важливий для практичного застосування, оскільки моделювати ці дві змінні дуже складно. Усі алгоритми, які використовують даний результат відомі як “алгоритми без моделі”, тому що середовище не моделюється.

Проте математичне сподівання все ще має місце. Простий, проте ефективний підхід - вибірка дуже великої кількості послідовностей марківського ланцюга та їх усереднення. Це незміщена оцінка, подібна до апроксимації

інтеграла в неперервному просторі з дискретним набором точок в області визначення. Цей метод відомий, як ланцюг Маркова Монте-Карло і часто використовується в байесівських нейромережах та імовірнісних графічних моделях. Незважаючи, на те, що градієнт параметризованої політики не залежить від винагороди, вона збільшує дисперсію при вибірці в методі Монте-Карло.

Якщо $r(\tau)$ замінити на дисконтовану оцінку G_t , отримаємо класичний алгоритм Policy Gradient. Тобто градієнт політики буде рахуватись наступним чином:

$$\nabla E_{\pi_{\theta}}[r(\tau)] = E_{\pi_{\theta}} \left[\left(\sum_{t=1}^T G_t \nabla \log \pi_{\theta}(a_t | s_t) \right) \right],$$

де G_t визначається наступним чином:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$$

2.8 Модель марківського процесу прийняття рішень для uplift моделювання

Позначимо марківський процес як (S, A, P, R) , де S - множина станів, A - множина дій, P - перехідна ймовірність з стану s в стан s' за умови, якщо була виконана дія a : $P_{ss'}^a = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$. $R = E\{r_{t+1} | s_t = s, a_t = a\}$ - миттєва винагорода за перехід з стану в стан при виконанні дії.

В задачі uplift моделювання треба максимізувати значення uplift реакції. Тому в кожен момент часу t спостерігається стан s_t , де стан - характеристики одного об'єкта ($S=X$), та за допомогою політики обирається дія з множини A та

отримується винагорода $R = L(x, a)$ (uplift реакція). Перехід зі стану в стан позначає перехід від одного об'єкта до іншого. Ймовірність переходу від стану s до стану s' не залежить від стану та дії: перехідні ймовірності рівні для всіх пар (s, s') і визначаються як $P_{ss'}^a = 1/|S|$. Оскільки всі ймовірності рівні, то стаціонарний розподіл побудованого марківського ланцюга буде еквівалентним розподілу випадкової з усіх даних, та не буде залежати від вибору дії. Політика буде обирати дії так, щоб максимізувати загальну винагороду. Тобто після вибору дії, результати оцінюються і політика оновлюється. Кожного моменту часу t система з стану $s_t = x_i$, може перейти в будь-який стан $s_{t+1} = x_j$ з рівною ймовірністю.

Схематично модель відображено на рис. 2.2. Синіми квадратами позначені стани а кольоровими обрані дії для кожного з станів. Результати оцінюються метрикою та параметри політики оцінюються, рахується градієнт, а потім параметри оновлюються.

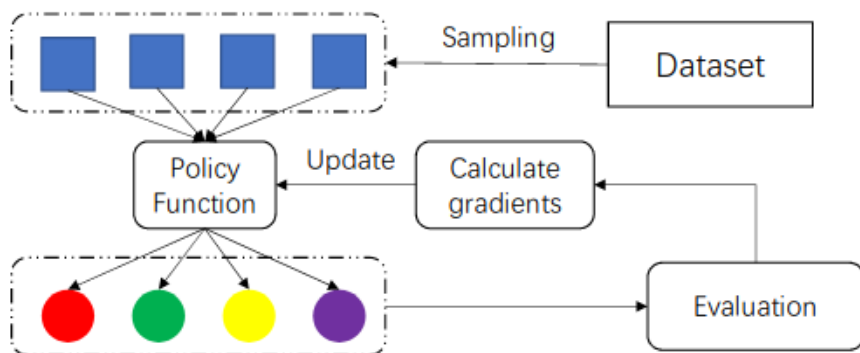


Рисунок 2.2 - Загальний алгоритм роботи моделі

На рисунку 2.3 зображено марківський ланцюг для цієї задачі, для випадку п'яти об'єктів:

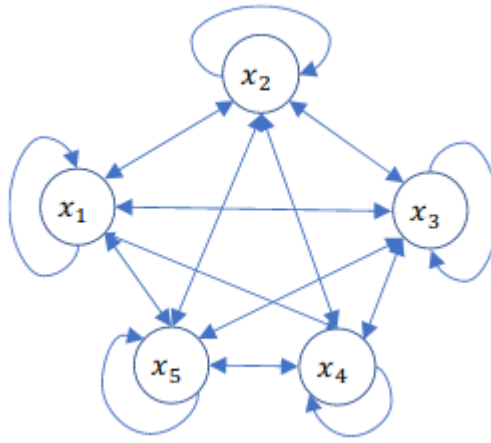


Рисунок 2.3 - Приклад марківського ланцюга для задачі

Кожен раз можна обрати дію a , знаходячись в стані s відповідно до політики $\pi(\theta)$, $a \sim \pi(s, \theta)$, де θ - параметри політики. Політика оцінюється відповідно до довгострокової очікуваної винагороди:

$$J(\theta) = E_{x, \pi(X)} [L(X, \pi(X))]$$

Отже, щоб максимізувати J , можна безпосередньо оптимізувати параметр θ . Відповідно до теореми про градієнт політики, можна обчислити градієнт таким чином:

$$\nabla_{\theta} J(\theta) = E_{\pi(\theta)} [\nabla_{\theta} \log(\pi(a|s, \theta)) Q^{\pi(\theta)}(s, a)]$$

де $Q^{\pi}(s, a)$ це винагорода за дію та стан за умови слідування політиці π та визначається, так:

$$Q^{\pi}(s, a) = R_s^a + \sum_{s'} p_{ss'}^a V^{\pi}(s')$$

де s' наступний стан та $V^{\pi}(s')$ винагорода за стани починаючи з s' , яка визначається як середнє очікування усіх подальших станів:

$$V^{\pi}(s) = \lim_{n \rightarrow \infty} E \left[\frac{R_{s_1}^{\pi(s_1)} + \dots + R_{s_n}^{\pi(s_n)}}{n} \right] = E_t [L(s_t, \pi(s_t)) | s_0 = s], \forall s \in S$$

З цього слідує, що для пошуку градієнта необхідно знати Q^{π} . Для цього необхідно визначити специфічне Q^{π} саме для uplift моделювання. З визначення, вона складається з двох частин: винагороди за дії Ras та за значення в певному стані $V^{\pi}(s^*)$. За допомогою метрики введеної в параграфі 2.3 можна порахувати $V^{\pi}(s^*)$. На кожному кроці навчання випадковим чином обирається M вибірок даних для оцінки значення. Значення uplift важко дізнатись, тому використовується наближення Ras . Відповідно з теореми 1 можна отримати, що якщо $p(s) = \pi(s) = a$, тоді відповідь $Y(s, a)$ має позитивний вплив на результат від дії $p(s)$, а якщо $p(s) = 0$ і $\pi(s) = a > 0$, тоді $Y(s, 0)$ негативно впливає на результат від дії $p(s)$, для інших випадків $Y(s, a)$ не впливає на результат, тому вони не розглядаються. Таким чином, Ras можна вважати оцінкою реакції uplift за метрикою UMG, тобто:

$$R_s^a = Y(s, a) \frac{1}{p(a|s)}, \text{ якщо } p(s) = \pi(s) = a$$

$$R_s^a = -Y(s, 0) \frac{1}{p(0|s)}, \text{ якщо } p(s) = 0 \text{ and } \pi(s) = a$$

Для зменшення дисперсії необхідно відняти базові лінії. Вводиться базова лінія, яка залежить від дій. Тоді значення $Q^{\pi}(s, a)$ буде рахуватись наступним чином:

$$Q^{\pi}(s, a) = (Y(s, a) - \overline{Y^C}(a)) / p(a|s) + (V^{\pi}(s^*) - \overline{V^{\pi}}(s^*)),$$

якщо $p(s) = \pi(s) = a$

$$Q^{\pi}(s, a) = (Y^C(a) - Y(s, 0)) / p(0|s) + (V^{\pi}(s^*) - \overline{V^{\pi}}(s^*)),$$

якщо $p(s) = 0 \text{ and } \pi(s) = a > 0$

де $Y^C(a) = \sum Y^C(a)/M$, $Y^C(a)$ - середнє за вибіркою умовне очікування $ZC(X)$, як було зазначено в лемі 2. $V \pi(s^*) = \sum_{m=1}^M V \pi_m(s^*)/M$ середнє значення $V \pi(s^*)$ в декількох вибірках. Таким чином, параметр θ можна оптимізувати для кожної вибірки:

$$\theta \leftarrow \theta + \alpha \sum_{i=1}^{|\Gamma_m|} \nabla_{\theta} \log \pi(s_i, a_i) Q^{\pi}(s_i, a_i), s_i \in \Gamma_m$$

Загальний алгоритм методу Policy Gradient для uplift моделювання буде таким:

- на вхід системі подається усі дані, розмір вибірок bs , кількість епох навчання, швидкість навчання. Результатом буде параметризована політика;
- на кожній епосі обирається M вибірок $\Gamma = \{\Gamma_1, \dots, \Gamma_M\}$ з даних, де кожна вибірка містить bs записів;
- для кожної вибірки виконуються дії $A = \{a^1, \dots, a^m\}$, де $a \sim \pi(s, i, \theta)$ та рахуються значення $\{Y^C(a)\} = SN - UMG(\Gamma_m, A)$;
- після цього обраховується загальне значення $V \pi(s^*) = \sum V \pi(s^*)/M$;
- рахується $Q\pi(s, i, a), \forall s \in \Gamma_m$, за рівнянням 9 ;
- параметри політики змінюються відповідно до рівняння.

2.9 Висновки до розділу

В даному розділі розглянуто основні поняття марківських ланцюгів та марківських процесів прийняття рішень. Сформульовано основні поняття теми, розглянуто основні властивості. Процес моделювання буде складатися з двох частин - переходу із стану в стан та вибір дії. Такі методи дозволяють

моделювати маркетинговий вплив навіть з кількома різними діями, що відрізняється від більш класичних підходів.

В розділі введено визначення uplift моделювання для кількох дій та введено усі необхідні поняття. Було проведено порівняння задачі uplift моделювання та задачі “контекстних бандитів”. Зроблено висновок, що при певних умовах їх можна вважати еквівалентними. Оскільки метою роботи було побудова універсальної моделі, яка дозволяє розв’язувати широкий клас задач, визначено формулювання задачі uplift моделювання через “контекстних бандитів” не є раціональним.

Також, запропонована нова метрика для оцінки результатів моделювання, оскільки класичні метрики не підходять для визначення uplift моделювання для кількох дій.

РОЗДІЛ 3 МОДЕЛЮВАННЯ МАРКЕТИНГОВОГО ВПЛИВУ

3.1 Розробка програмного забезпечення

Для uplift моделювання була розроблена програма на мові Python. Зокрема, були використані бібліотеки pandas, numpy, tensorflow. Розроблений програмний продукт є кроссплатформним.

В програмі реалізовано два методи: марківський процес прийняття рішень (МППР) з алгоритмом навчання policy gradient та метод на основі двох моделей з моделлю випадковий ліс та з нейромережею. Для оцінки роботи методів реалізовані метрики qini та sn-umg, введена в другому розділі. Також є можливість моделювати uplift для двох або кількох маркетингових дій.

Програма представляє собою чотири модулі: реалізація методу policy gradient, реалізація двомодельного підходу, модуль з метриками та модуль з порівнянням методів.

3.2 Огляд та аналіз даних

Для моделювання використовується два набори даних: датасет з хакатону “MineThatData” та згенерований набір даних.

“MineThatData” відкритий набір даних, що містить результати розсилки електронною поштою з рекламою інтернет магазину. Він містить інформацію про 64000 клієнтів з основною маркетинговою інформацією, такою як: сума грошей витрачених в попередні періоди, дата останньої покупки та інше. Цільова змінна датасету - відвідування сайту.

Набір даних містить, зображено на рис. 3.1 такі ознаки об’єкта: період від останньої покупки, кількість грошей витрачена за минулий період (рік), тип реклами яку отримували клієнти (з чоловічими або жіночими товарами або

ніяку), поштовий індекс (ділиться на категорії міський, приміський та сільський), індикатор того, що клієнт новий та канали, на яких раніше купував клієнт.

	recency	history_segment	history	mens	womens	zip_code	newbie	channel
0	10	2) \$100 - \$200	142.44	1	0	Surburban	0	Phone
1	6	3) \$200 - \$350	329.08	1	1	Rural	1	Web
2	7	2) \$100 - \$200	180.65	0	1	Surburban	1	Web
3	9	5) \$500 - \$750	675.83	1	0	Rural	1	Web
4	2	1) \$0 - \$100	45.34	1	0	Urban	0	Web
63995	10	2) \$100 - \$200	105.54	1	0	Urban	0	Web
63996	5	1) \$0 - \$100	38.91	0	1	Urban	1	Phone
63997	6	1) \$0 - \$100	29.99	1	0	Urban	1	Phone
63998	1	5) \$500 - \$750	552.94	1	0	Surburban	1	Multichannel
63999	1	4) \$350 - \$500	472.82	0	1	Surburban	0	Web

Рисунок 3.1 - Приклад даних

З кореляційної матриці на рис. 3.2 вказує на те, що майже всі ознаки не мають високої кореляції між собою:

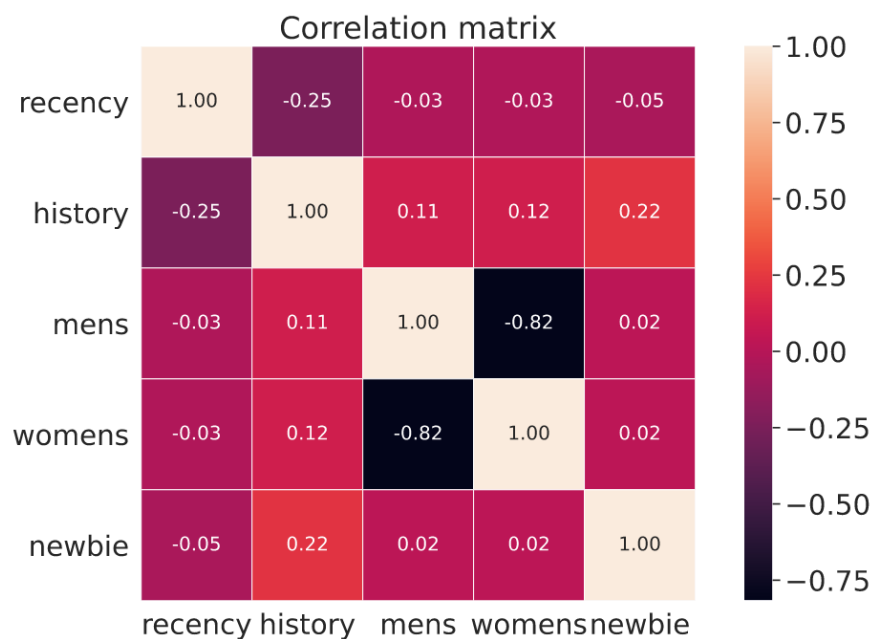


Рисунок 3.2 - Кореляційна матриця

Категоріальна ознака “history segment”, що позначає приблизні витрати споживача, розподілена як на рис. 3.3

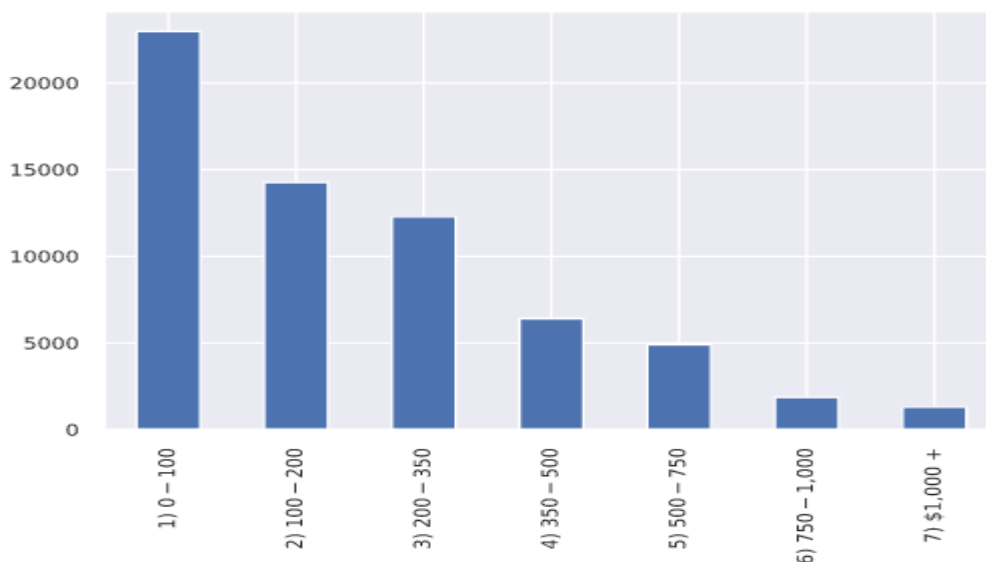


Рисунок 3.3 - Розподіл категоріальної ознаки “history segment”

Для моделювання дану ознаку краще перевести в числову, взявши середину кожного відрізка.

Цільова змінна “візити сайту” приймає значення 1, якщо клієнт зайшов на сайт та 0 в іншому випадку. З графіку видно на рис. 3.4, що цільова змінна не збалансована:

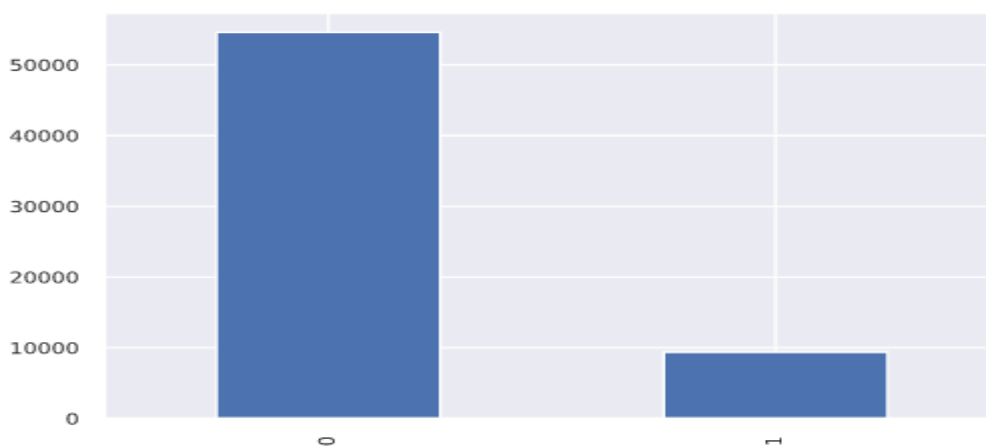


Рисунок 3.4 - Розподіл цільової змінної

Для перевірки uplift моделювання з кількома діями немає відкритого датасету, тому було згенеровано синтетичний набір даних. Алгоритм працює аналогічно до алгоритму запропонованого в [22]. Метод заснований на дереві рішень, тому значення uplift від різних дій залежать від одного атрибута.

Простір ознак - це 50-вимірний гіперкуб з довжиною 10, тобто $X_{50} = [0,10]^{50}$. Ознаки рівномірно розподілені в просторі, тому $X_d \leftarrow U[0,10]$, для $d = 1, \dots, 50$. Буде розглядатись чотири різні дії, $A = \{1,2,3,4\}$. Реакція на кожну дію:

$$Y(x) = 5f_1(X) + f_2(X) + \varepsilon$$

де f_1 - реакція на сам продукт f_1 ,

f_2 - реакція на маркетингову дію,

ε - шум.

Функції f_1 та f_2 мають однаковий вигляд, проте з різними параметрами:

$$f(x_1, \dots, x_{50}) = \sum_{i=1}^{50} a_i * \exp\left\{-\sum_{j=1}^{50} b_{ij}|x_i - c_{ij}|\right\}$$

де x_i - ознаки об'єктів,

a_i, b_i, c_i - параметри.

Параметр a рівномірно розподілений на відрізку $[0, 10]$, b на $[0, 0.1]$ та c на $[0, 5]$. Для кожної ознаки об'єкта параметри випадковим чином обираються спочатку для функції f_1 , потім для f_2 .

3.3 Перевірка роботи моделі за умови двох дій

Для оцінки роботи побудованої моделі вона порівнюється з методом з двома моделями на основі нейромережі та випадкових дерев. Для порівняння методів використовується метрика SN-UMG, коефіцієнт Qini та uplift крива. Для цього моделювання використовується дані з хакатону “MineThatData”.

Результати моделювання методом на основі випадкового лісу та марківського процесу прийняття рішень зображені на графіку на Рис.3.5

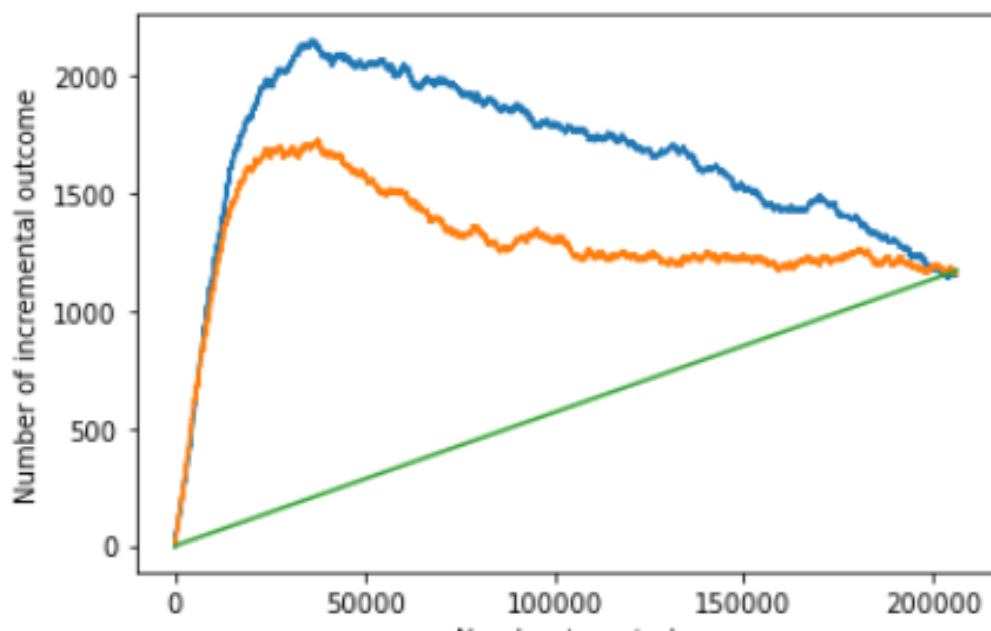


Рисунок 3.5 Порівняння моделей МППР та випадкового лісу. Синім позначено модель МППР, червоним випадкового лісу.

Метрики для цих двох моделей наведені у табл. 3.1:

Табл. 3.1 - Порівняння моделей МППР та випадкового лісу

	МППР	Випадковий ліс
QINI	0.068	0.045
SN-UMG	0.034	0.028

Результати моделювання методом на основі методу з двома нейромережами та марківського процесу прийняття рішень зображені на графіку на Рис.3.6

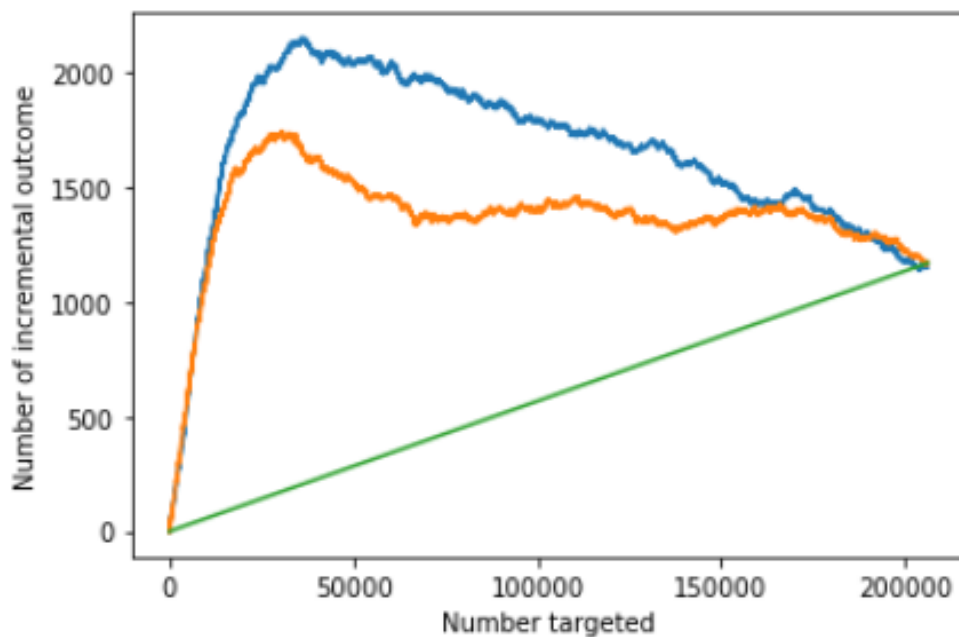


Рисунок 3.6 Порівняння моделей МППР та двох нейромереж. Синім позначено модель МППР, червоним двомодельного підходу.

Метрики для цих двох моделей наведено у таблиці 3.2:

Табл. 3.2 - Порівняння моделей МППР та двомодельного підходу

	МППР	двомодельний підхід
QINI	0.068	0.0506
SN-UMG	0.034	0.0301

За результатами моделювання видно що запропонована модель значно краще за класичні підходи. Оскільки Qini крива підходить тільки для моделювання uplift з двома діями, в даному випадку порівнювали саме такі випадки.

3.4 Перевірка роботи моделі при моделюванні кількох дій

Для перевірки роботи моделі з кількома діями використовується метрика SN-UMG. Модель МППР порівнюється з методом з двома нейромережами. Моделювання проводиться на основі згенерованого набору даних, описаних в розділі 3.2. В даному наборі даних досить складна залежність між цільовою змінною та властивостями об'єктів та їх достатньо багато. Результати моделювання наведені у табл. 3.3:

Табл. 3.3 - Порівняння моделей МППР та двомодельного підходу

	МППР	двомодельний підхід
SN-UMG	0.1683	0.1106

За результатами видно, що і у випадку виконання кількох дій, запронована модель дає кращі результати.

3.5 Висновки до розділу

В даному розділі був описаний побудований програмний продукт для uplift моделювання за допомогою марківського процесу прийняття рішень. Також реалізовані методи на основі двох нейромереж та випадковий ліс.

Для оцінки роботи моделей розглянуто два набори даних - відкритий "MineThatData" та згенерований для оцінки роботи моделі з кількома діями. В розділі описано метод генерації та проведено загальний аналіз набору даних "MineThatData". Моделі порівнюються за допомогою метрик Qini та SN-UMG. Також, деякі результати моделювання зображені на графіках з Qini кривими.

За результати порівняння можна зробити висновок, що модель МППР набагато краще виконує uplift моделювання ніж більш класичні методи. Для більш детального аналізу моделювання uplift для кількох дій потрібно перевірити роботу методів на реальному наборі даних, проте навіть на використаному наборі даних модель значно краще працює ніж дві нейромережі.

РОЗДІЛ 4 РОЗРОБЛЕННЯ СТАРТАП – ПРОЕКТУ

В останні роки набув великої популярності такий вид малого підприємництва як стартап. Стартап проект – є комерційним проектом, який знаходиться в стані розробки, або нещодавно вийшов на ринок. Характерною особливістю стартапу, що відрізняє його від малого бізнесу, є оригінальність та інновації, він не може бути копією вже реалізованих ідей. При цьому проект не обов'язково повинен бути масштабного характеру, головне, щоб він був креативним, а його завдання – спрощувати людям будь-які дії в їх повсякденному житті.

Наразі, з появою Інтернету та сучасних технологій, стало простіше заходити на ринок, знаходити інвесторів та споживачів. З'явилося набагато більше можливостей для розвитку свого проекту за кордоном, ніж раніше. Проте розробка стартапу є досить ризикованим завданням. Не всім вдається довести свій стартап проект до ринкового впровадження. За статистикою успіху досягає лише 10 – 20% від усіх стартап проектів. Запуск стартапу передбачає цілий ряд обов'язкових дій, в межах яких визначають ринкові перспективи стартапу, графік розробки, принципи організації виробництва, заходи з залучення інвесторів та аналіз ризиків.

4.1 Опис ідеї стартап проекту

Назва стартап проекту: “Модель марківського процесу прийняття рішень для uplift моделювання”.

У табл. 4.1 подано зміст ідеї стартап проекту, можливі напрямки застосування та основні вигоди, що може отримати користувач товару. У табл. 4.2 визначені сильні, слабкі та нейтральні сторони проекту.

Таблиця 4.1 – Опис ідеї стартап – проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Модель марківського процесу прийняття рішень для uplift моделювання за допомогою даних користувача визначає оптимальні маркетингові дії	Маркетингові кампанії	Система дозволяє визначити групу людей на яких необхідно спрямувати маркетингові дії і таким чином зменшити витрати на маркетингову кампанію

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

	Технічно-економічні характеристики	Потенційні концепції конкуренти		
		Мій проект	Команда таргетологів	Проект sklift
1.	Ціна	Середня	Висока	Низька
2.	Функціонал	Широкий	Вузький	Середній

Отже, з табл. 4.2 можна визначити, що ціна та функціонал є сильними характеристиками для потенційного товару.

4.2 Технологічний аудит ідеї проекту

За результатами аналізу табл. 4.3 можна зробити висновок про можливість технологічної реалізації проекту.

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1.	Модель марківського процесу прийняття рішень для uplift моделювання	Python, TensorFlow	Наявна	Доступна
2.	Модель марківського процесу прийняття рішень для uplift моделювання	C++	Наявна	Недоступна
Обрана технологія реалізації ідеї проекту: Статичне ребалансування, прогнозування на основі градієнтного бустингу (Python, TensorFlow)				

4.3 Аналіз ринкових можливостей запуску стартап проекту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту та ринкових загроз, які можуть перешкодити реалізації проекту дозволяє планувати напрями розвитку проекту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів конкурентів.

Попередню характеристику ринку можна побачити у табл.4.4.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап – проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1.	Кількість головних гравців, од	15
2.	Загальний обсяг продаж, грн/ум.од	100 000 ум.од.
3.	Динаміка ринку (якісна оцінка)	Зростає
4.	Наявність обмежень для входу (вказати характер обмежень)	Немає обмежень для входу
5.	Специфічні вимоги до стандартизації та сертифікації	Немає специфічних вимог до стандартизації та сертифікації
6.	Середня норма рентабельності в галузі (або по ринку), %	75 %

Середня норма рентабельності галузі вища за банківський відсоток на вкладення, що означає привабливість ринку для входження за попереднім оцінюванням. Визначимо потенційні групи клієнтів, їх характеристики, та сформуємо орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.5 – Характеристика потенційних клієнтів стартап – проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1.	Потреба визначати групи людей на яких потрібно направити маркетингові дії	Компанії, які мають велику аудиторію і є дуже популярні, одна хочуть ще збільшити обсяг продажу	Клієнту надається лише частина продукту, а саме модуль моделювання uplift	Простота у використанні, висока точність
2.	Потреба класифікації потенційної аудиторії	Компанії які тільки вийшли на ринок і мають потребу визначити основні групи потенційних клієнтів	Клієнт використовує всі можливості системи	Зручний інтерфейс, охоплення усіх можливих випадків, простота у використанні, висока точність

Проведемо аналіз ринкового середовища. Нижче наведені таблиці факторів, що сприяють ринковому впровадженню проекту та факторів, що йому перешкоджають (табл. 4.6 – 4.7).

Таблиця 4.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1.	Конкуренція	Збільшення функціоналу роботи sklift	Потрібно розробити вдалу маркетингову та рекламну кампанію

Продовження таблиці 4.6

2.	Зовнішній вплив на ринок	Вплив на успішність маркетингової зовнішніх факторів, таких як пандемії, війни та інше.	Вдосконалити або модифікувати алгоритми та аналітику платформи
3.	Зміна потреб користувача	Користувачам необхідні рішення з іншим функціоналом	Передбачення рішення з іншим функціоналом додавання можливості нового функціоналу
4.	Постачальники	Постачальники планують підвищити ціну на ліцензійне програмне забезпечення, необхідне для реалізації та підтримки проект	Залучити інвестиції з урахуванням змін вести перемовини з постачальниками з приводу фіксування ціни за умов довготривалої співпраці

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1.	Конкуренція	Збільшення функціоналу роботи sklift	Демпінгувати ціни на послуги. Побудова рекламних кампаній
2.	Зовнішній вплив на ринок	Розширення функціоналу	Адаптувати рішення під нові умови зовнішнього середовища
3.	Поява нових цільових груп	Потреба в аналогічному продукті в інших сферах діяльності	Адаптація продукту під нові сфери використання
4.	Постачальники	Поява нових постачальників, з більш вигідними пропозиціями ресурсів	Налагодити комунікацію з потенційними партнерами

Проведемо аналіз пропозиції: визначимо загальні риси конкуренції на ринку (табл. 4.8).

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Тип конкуренції – монополістична конкуренція	Ринок складається з відносно великої к-ті продавців, кожен з яких володіє невеликою (не нескінченно малою) часткою ринку; встановлюючи ціни, продавці намагаються виділитися за неціновими ознаками; товар продавця є недосконалим замінником товарів інших фірм; ринок не має бар'єрів входу/виходу	Для того, щоб відрізнитися від конкурентів чимось, окрім ціни, продавці розробляють різні пропозиції для різних сегментів споживачів і широко використовують практику використання марок, рекламу та методи особистих продажів. Також необхідною є підтримка якості продукту та постійні вдосконалення
За рівнем конкурентної боротьби – інтернаціональний рівень	Конкуренти діють на інтернаціональному просторі – в мережі Інтернет	Забезпечити можливість користування послугами (продуктом) незалежно від місцезнаходження. Підтримувати продукт на національному ринку
За галузевою ознакою – внутрішньогалузева	Має місце суперництво між окремими підприємцями і фірмами однієї галузі щодо одержання прибутку	Підвищення продуктивності праці, зменшення витрат виробництва, зниження індивідуальної цінності товару, розширення функціоналу продукту, вдосконалення для застосування у нових галузях

Продовження таблиці 4.8

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Конкуренція за видами товарів – товарно – видова	Це конкуренція між товарами одного виду – різні рекомендаційні системи виконують один і той же набір функцій	Використовувати цінові та нецінові методи конкурентної боротьби на ринку
За характером конкурентних переваг – нецінова	Проводиться головним чином за допомогою вдосконалення якості продукції, технології виробництва, інновацій та нанотехнологій, патентування, брендування, умов її продажу.	Випускати новий товар (послугу), які принципово відрізняються від своїх попередників або представляють модернізований варіант старої моделі
За інтенсивністю – не марочна	Роль торгової марки незначна, хоча самі марки можуть бути присутніми на ринку	Немає необхідності вкладати кошти у створення та розкрутку бренду, необхідно приділяти увагу якості продукту а не бренду компанії

Після аналізу конкуренції проведемо більш детальний аналіз умов конкуренції в галузі (за моделлю 5 сил М. Портера) (табл. 4.9).

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Клієнти	Товари – замітники
Складові аналізу	sklift	Маркетингові та рекламні компанії	Вплив покупців на ринкову ціну; Очікування щодо якості обслуговування	Поява відносно дешевих або безкоштовних замінників
Висновки	З боку прямих конкурентів очікується інтенсивна конкурентна боротьба.	- є можливість виходу на ринок, бар'єри відсутні; - потенційні конкуренти – великі компанії з потужними виробничими можливостями - загалом названі вище компанії мають декілька вже успішних реалізованих варіантів схожої продукції та готують інновації, тому потенційний ріст популярності продукції компаній невідомий	Клієнти диктують умови роботи на ринку	Є схожі товари, однак дана версія є унікальною в певному роді, саме таких безкоштовних аналогів наразі не існує

Отже, з огляду на конкурентну ситуацію у галузі, бачимо, що є можливість входження на ринок. Щоб бути конкурентноспроможним на ринку, проект має бути привабливим за ціною, якістю роботи та мати зручний інтерфейс. Зокрема проект дещо відрізняється від вже існуючих систем, а саме дає більшу точність та можливість обирати різні дії.

У табл. 4.10 наведено фактори конкурентноспроможності, сформовані на основі аналізу конкуренції та з урахуванням характеристик ідеї проекту, вимог споживачів до товару (табл. 4.5) та факторів маркетингового середовища (табл. 4.6 – 4.7)

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1.	Технологія	Рівень застосовуваних технологій відіграє значну роль в побудові ефективної системи uplift моделювання, а отже надання якісного та швидкого сервісу споживачу
2.	Ціна	Ціна має значення як для приватних особи-користувача онлайн-системи, так і для великих компаній, які зацікавлені у придбанні абонементу на користування системою uplift моделювання
3.	Маркетинг	- Частка ринку, займаного підприємством; - Престиж торгової марки (бренд); - Витрати на стимулювання збуту і їх ефективність
4.	Репутація	- Репутація (імідж) підприємства; - Якість обслуговування

Продовження таблиці 4.10

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
5.	Унікальність позиціонування	В умовах монополістичної конкуренції, коли фактор диференціації торгової марки є ключовим засобом ведення конкурентної боротьби, важливим є створення та підтримання унікального позиціонування, що створює певний захист від конкурентних зіткнень. Для поточного продукту наявна інновація, котра відрізняє його від вже існуючих продуктів на ринку

У табл. 4.11 наведено аналіз сильних та слабких сторін проекту, сформульованих за визначеними факторами конкурентоспроможності (табл. 4.10).

Серед основних компаній конкурентів можна виділити такі як:

1. Causalml.
2. Sklift.
3. Inweb.
4. NetRocket.

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін «Система побудови рекламних кампаній»

	Фактор конкурентоспроможності	Бали 1 – 20							
			-3	-2	-1	0	1	2	3
1.	Технологія	20		3/4			1	2	
2.	Ціна	10					1/2	3	4
3.	Маркетинг	9			1		2		3/4
4.	Репутація	8			1		2/4	3	
5.	Унікальність	19		3/4			1/2		

4.4 SWOT – аналіз

У табл 5.12 наведено SWOT – аналіз (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (табл. 4.12) на основі виділених ринкових загроз та можливостей, та сильних і слабких сторін (табл. 4.11).

Таблиця 4.12 – SWOT – аналіз стартап – проекту

Сильні сторони: розумна цінова політика, функціонал забезпечує рішення більшості задач клієнта	Слабкі сторони: відсутність співпраці з інноваційними центрами, велика ресурсоемність, можливі проблеми за наявності зовнішнього впливу
Можливості: впровадження інноваційних рішень, оптимізація роботи продукту	Загрози: неоптимізована робота продукту, узагальнені рішення, що непридатні для вирішення клієнтських задач

На основі SWOT аналізу визначимо альтернативи ринкової поведінки (перелік заходів) для виведення стартап проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок (табл. 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартап – проект

Альтернатива комплекс (орієнтовний заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
Впровадження існуючого товару на існуючому ринку	Середня	6-9 міс
Розвиток товару на існуючому ринку	Висока	3-4 міс
Розвиток нового ринку з існуючим товаром	Низька	12-16 міс

Продовження таблиці 4.13

Альтернатива комплекс (орієнтовний західів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
Впровадження нового товару і створення нового ринку	Низька	18-20 міс

Обрано альтернативу 2, для якої ймовірність отримання ресурсів є високою і строк реалізації відносно невеликий.

4.5 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів (табл. 4.14).

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

	Опис цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтони й попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу в сегмент
1.	Малий бізнес	Низька - для них можуть бути актуальні більш прості та дешевші рішення, які розбивають аудиторію на схильних купити та ні	Середній	Висока	Вхід в сегмент достатньо складний

Продовження таблиці 4.14

	Опис цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтоний попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу в сегмент
2.	Середній бізнес	Висока	Середній або високий	Середня	Вхід в сегмент достатньо складний
3.	Великий бізнес	Висока - це основна аудиторія, оскільки вони досить популярні і затрати на зайву рекламу можуть бути досить високі	Дуже високий	Низька	Середня
Які цільові групи обрано: 2,3					

У табл. 4.15 наведено базову стратегію розвитку.

Таблиця 4.15 – Визначення базової стратегії розвитку

Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Розробка та створення додаткових функціональних модулів	Таргетні пропозиції бізнесу, проведення презентації функціональних рішень на ярмарках та конференціях	Відсутність аналогічних до новостворених функціональних модулів у конкурентів	Розробка та удосконалення існуючих модулів на основі потреб ринку та інформації від клієнтів

У табл. 4.16 наведено стратегію конкурентної поведінки.

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
Так	Так	Стандартні функціональні модулі будуть виконувати схожі функції до конкурентів	функціональні інновації

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту (див. табл. 4.5), а також в залежності від обраної базової стратегії розвитку (табл. 4.15) та стратегії конкурентної поведінки (табл. 4.16) розробляється стратегія позиціонування (табл. 4.17), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.17 – Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
Висока якість прогнозування в клієнтській сфері застосування	Розробка та удосконалення існуючих модулів на основі потреб ринку та інформації від клієнтів	Спеціалізовані рішення, хмарні сервіси	Прогнозування, передбачення, аналіз

4.6 Розроблення маркетингової програми стартап – проекту

У табл. 4.18 наведено результати попереднього аналізу конкурентоспроможності товару.

Таблиця 4.18 – Визначення ключових переваг концепції потенційного товару

Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
Широкий функціонал	Вирішення задач	Забезпечує вирішення більшої кількості задач бізнесу
Спеціалізовані рішення	Вирішення задач	Забезпечує більш ефективне вирішення задач у звуженій сфері застосування
Технічні ресурси	Хмарні сервіси	Дозволяє користуватись рішенням за рахунок віддалених технічних потужностей

Надалі розробляється трирівнева маркетингова модель товару: уточнюється ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (табл. 4. 19).

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
Товар за задумом	Обробка, аналіз даних. Моделювання поведінки різних груп людей		
Товар у реальному виконанні	Швидкодія Ефективність Користувацький інтерфейс	Нм Нм Нм	Тх/Тл/Е Тх/Тл Е

Продовження таблиці 4.19

Рівні товару	Сутність та складові
	Якість: стандарти відповідні до законодавства. Створені функціональні скріпти.
	Пакування: Власний сайт
	Марка: MDPlift
Товар із підкріпленням	До продажу: застосунок для інтеграції в існуючі системи керування підприємством для моделювання дій та реакцій споживачів на основі великих масивів даних
	Після продажу: Швидкодія, ефективність, легкість у користуванні
Закритий код. Захищений від можливості декомпіляції.	

Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово – економічного аналізу проекту), яке передбачає аналіз ціни на товари – аналоги або товари субституту, а також аналіз рівня доходів цільової групи споживачів (табл. 4.20). Аналіз проводиться експертним методом.

Таблиця 4.20 – Визначення меж встановлення ціни

Рівень цін на товари замітники	Рівень цін на товари аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1500\$ міс	2000\$ міс	Рівень доходів підприємств надзвичайно високий	1000\$ міс

У таблиці 4.21 наведено формування системи збуту.

Таблиця 4.21 – Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Таргетні пропозиції для компаній	Презентації функціоналу	-	-

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позицінування, визначену специфіку поведінки клієнтів (табл. 4.22).

Таблиця 4.22 – Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Середній бізнес - оптимальні рішення за невисоку ціну	Таргетні дзвінки до клієнтів	Прогнозування покупок споживача	Короткий опис переваг продукту, заохочення дізнатись більше	Передбачення покупок споживачів

Продовження таблиці 4.22

Специфіка поведінки цільових клієнтів	Канали комунікації, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Великий бізнес – повноцінні рішення для покращення маркетингових кампаній	Таргетні дзвінки до клієнтів	Прогнозування покупок споживача	Донести інформацію про оптимальність рішення для бізнесу клієнта	Передбачення покупок споживачів

4.7 Висновки до розділу

Відповідно до вищенаведених результатів, можна стверджувати про наявність попиту на запропоновану систему. Варто зауважити, що на ринку присутня сильна конкуренція, але інноваційна складова продукту дозволяє суттєво збільшити конкурентоспроможність проекту. Подальша імплементація проекту доцільна за умови фокусування на середній та великий бізнес, оскільки саме цей ринок динамічно зростає і є достатньо рентабельним.

Сформовано ринкову (маркетингову) програму, що включає в себе концепції товару, збуту, просування та попередній аналіз можливостей ціноутворення, спирається на цінності та потреби потенційних клієнтів, конкурентні переваги ідеї, стан та динаміку ринкового середовища, в межах якого буде впроваджено проект, та відповідну обрану альтернативу ринкової поведінки.

ВИСНОВОК

В роботі запропоновано метод моделювання маркетингово впливу на різні групи споживачів. Розроблений метод базується на uplift методиці, яка дозволяє визначити чистий ефект від взаємодії зі споживачем. Задача uplift моделювання була сформульована як проблема марківського процесу прийняття рішень шляхом побудови еквівалентного марківського ланцюга.

Пояснено актуальність обраної теми визначено, що вирішення подібних задач необхідно компаніям на різних ступенях розвитку для ефективного застосування маркетингових інструментів. Проведено детальний аналіз класичних моделей, таких як двомодельний підхід, трансформація класів, дерева рішень, та метрик на базі існуючих методик. Зроблено висновок, про новизну поставленої задачі та необхідності різнобічних досліджень, оскільки існуючі методи не дають змоги досягти бажаного результату.

В другому розділі роботи розглянуто теоретичну базу марківських ланцюгів та марківського процесу прийняття рішень. Для застосування даного методу, потрібно переформулювати задачу uplift моделювання і вводити поняття стану, дій та матриці переходу. Було проведено порівняння задачі uplift моделювання та задачі “контекстних бандитів”. Зроблено висновок, що при певних умовах їх можна вважати еквівалентними. Оскільки метою роботи було побудова універсальної моделі, яка дозволяє розв’язувати широкий клас задач, визначення формулювання задачі uplift моделювання через “контекстних бандитів” не є раціональним. В другому розділі була запропонована метрика, для оцінки роботи моделей. Введена метрика ґрунтується на вже відомий метриці для uplift моделювання для кількох дій.

На базі проведених досліджень створено програму, яка будує марківський ланцюг та виконує uplift моделювання за допомогою марківського процесу прийняття рішень, дерева рішень і двох нейромереж. На основі отриманих результатів зроблено висновок, про обмеженість застосування класичних

моделей для uplift моделювання, в той час як запропонована модель марківського процесу прийняття рішень показує значно кращі результати.

Результатами роботи є:

- висновки, щодо актуальності застосування uplift моделювання для моделювання маркетингового впливу на різні групи людей;
- побудована програма для моделювання маркетингового впливу за допомогою МППР;
- порівняльні характеристики відомих моделей з розробленої в даній роботі моделі.

Подальший розвиток обраної теми: дослідження нових моделей, розширення функціоналу розробленої моделей, пошук нових областей застосування uplift моделювання.

ПЕРЕЛІК ПОСИЛАНЬ

1. Angrist J. D., Pischke J. S. Mostly harmless econometrics. *Princeton: Princeton university press*, 2008. 392 p.
2. Athey S., Imbens G. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*. 2016. Vol. 113, No. 27. P. 7353-7360.
3. Betlei A., Diemert E., Amini M. R. Uplift prediction with dependent feature representation in imbalanced treatment and control conditions. *International Conference on Neural Information Processing*. Springer, Cham, 2018. P. 47-57.
4. Devriendt F. et al. Learning to rank for uplift modeling. *IEEE Transactions on Knowledge and Data Engineering*. 2020. No. 7. P. 75.
5. Gutierrez P., Gérardy J. Y. Causal inference and uplift modelling: A review of the literature. *International Conference on Predictive Applications and APIs*. PMLR, 2017. P. 1-13.
6. Hansotia B., Rukstales B. Incremental value modeling. *Journal of Interactive Marketing*. 2002. Vol. 16, No. 3. P. 35-46.
7. Kolobov A. Planning with Markov decision processes: An AI perspective. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2012. Vol 6, No. 1. P. 1-210.
8. Lewis F. L., Liu D. (ed.). Reinforcement learning and approximate dynamic programming for feedback control. John Wiley & Sons, 2013. Vol. 17. URL: <https://ieeexplore.ieee.org/abstract/document/5227780>
9. Li Chencheng et al. Reinforcement learning for uplift modeling. arXiv preprint arXiv:1811.10158. 2018.
10. Michel R., Schnakenburg I., Von Martens T. Targeting uplift: An introduction to net scores. Switzerland, Cham: Springer Nature, 2019. 350 p.
11. Nassif H. et al. Uplift Modeling with ROC: An SRL Case Study. *CEUR Workshop Proceedings. ILP (Late Breaking Papers)*. 2013. P. 40-45 URL:

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.3042&rep=rep1&type=pdf>

12. Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983. Vol. 70, No. 1. P. 41–55 .
13. Pavel Surmenok. Contextual Bandits and Reinforcement Learning. 2017. URL: <https://towardsdatascience.com/contextual-bandits-and-reinforcement-learning-6bdfeaece72a>
14. Piotr Rzepakowski and Szymon Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*. 2012. Vol. 32, No. 2. P. 303—327.
15. Radcliffe N. Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal*. 2007. P. 14-21.
16. Rohan Jagtap. Understanding Markov Decision Process (MDP). 2020. URL: <https://towardsdatascience.com/understanding-the-markov-decision-process-mdp-8f838510f150>
17. Sanyam Kapoor. Policy Gradients in a Nutshell. Everything you need to know. 2018. URL: <https://towardsdatascience.com/policy-gradients-in-a-nutshell-8b72f9743c5d>
18. Smith-Miles K. et al. Towards objective measures of algorithm performance across instance space. *Computers & Operations Research*. 2014. Vol. 45. P. 12-24.
19. Steve Roberts. Multi-Armed Bandits: Part 1. Mathematical Framework and Terminology. 2020. URL: <https://towardsdatascience.com/multi-armed-bandits-part-1-b8d33ab80697>
20. Swaminathan A., Joachims T. The self-normalized estimator for counterfactual learning. in neural information processing systems. 2015. No. 28. P. 3231–3239.
21. Verbeke W., Baesens B., Bravo C. Profit driven business analytics: A practitioner's guide to transforming big data into added value. Canada, New Jersey: John Wiley & Sons, 2017. 381 p.

22. Zaniewicz Ł., Jaroszewicz S. Support vector machines for uplift modeling. 2013 *IEEE 13th International Conference on Data Mining Workshops*. IEEE, Dallas, TX, USA 7-10 December 2013. P. 131-138 .
23. Zhao Y., Fang X., Simchi-Levi D. Uplift modeling with multiple treatments and general response types. *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics. Houston, Texas, USA, April 27 – April 29, 2017. P. 588-596.

ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```

class PolicyGradient:
    def __init__(
        self,
        n_action,
        n_feature,
        hidden_layers=None,
        batch_size=512,
        learning_rate=0.01,
        reward_decay=1.0,
        output_graph=False,
        reuse=False
    ):
        self.n_action = n_action
        self.n_feature = n_feature
        self.lr = learning_rate
        self.gamma = reward_decay
        self.batch_size = batch_size

        self.ep_obs, self.ep_algo_action, self.ep_real_action, self.ep_rs = [], [], [], []
        # with tf.device('/device:GPU:0'):
        self._build_net()
        self.losses = []
        config = tf.ConfigProto()
        config.gpu_options.allow_growth = True
        self.sess = tf.Session(config=config)
        self.sess.run(tf.global_variables_initializer())

    def _build_net(self):

        with tf.name_scope('inputs'):
            self.tf_obs = tf.placeholder(
                tf.float32, [None, self.n_feature], name="observations")
            self.tf_acts = tf.placeholder(
                tf.int32, [None, ], name="actions_num")
            self.tf_vt = tf.placeholder(
                tf.float32, [None, ], name="actions_value")
            self.tf_lbs = tf.placeholder(
                tf.float32, [None, self.n_action], name="labels_value")

```

```

with tf.variable_scope("policy_net"):
    # self.model = self._build_shared_network(
    #     self.tf_obs)
    # all_act = self.model.get_layer("dense_3").output
    # fc1
    layer_1 = tf.layers.dense(
        inputs=self.tf_obs,
        units=512,
        activation=tf.nn.relu, # tanh activation
        kernel_initializer=tf.contrib.layers.xavier_initializer(),
        bias_initializer=tf.constant_initializer(0.1),
        kernel_regularizer=tf.contrib.layers.l2_regularizer(0.001),
        name='fc1'
    )
    # bn = tf.contrib.layers.batch_norm(layer_1)
    # layer_2 = tf.layers.dense(
    #     inputs=layer_1,
    #     units=64,
    #     activation=tf.nn.relu, # tanh activation
    #     kernel_initializer=tf.contrib.layers.xavier_initializer(),
    #     bias_initializer=tf.constant_initializer(0.1),
    #     kernel_regularizer=tf.contrib.layers.l2_regularizer(0.001),
    #     name='fc2'
    # )
    # bn = tf.contrib.layers.batch_norm(layer_1)
    layer_2 = tf.layers.dense(
        inputs=layer_1,
        units=64,
        activation=tf.nn.relu, # tanh activation
        kernel_initializer=tf.contrib.layers.xavier_initializer(),
        # kernel_initializer=tf.random_uniform_initializer(),
        bias_initializer=tf.constant_initializer(0.1),
        kernel_regularizer=tf.contrib.layers.l2_regularizer(0.001),
        name='fc2'
    )
    # fc2
    # dropout = tf.nn.dropout(layer_2, keep_prob=0.5)
    all_act = tf.layers.dense(
        inputs=layer_1,
        units=self.n_action,
        activation=None,

```

```

        kernel_initializer=tf.random_normal_initializer(0, 1),
        # kernel_initializer=tf.contrib.layers.xavier_initializer(),
        bias_initializer=tf.constant_initializer(0.1),
        name='fc3'
    )
    # use softmax to convert to probability
    self.all_act_prob_soft = tf.nn.softmax(all_act, name='act_prob')
    self.all_act_prob = tf.clip_by_value(
        self.all_act_prob_soft, 1e-4, 1)

    self.entropy = - \
        tf.reduce_sum(self.all_act_prob *
            tf.log(self.all_act_prob), 1, name="entropy")
    self.entropy_mean = tf.reduce_mean(
        self.entropy, name="entropy_mean")

    # self.variance =
    self.prob_avg = tf.reduce_mean(self.all_act_prob_soft, axis=0)
    self.loss_avg = tf.abs(self.prob_avg[0] - 0.5)

with tf.name_scope('loss'):

    self.neg_log_prob = tf.nn.sparse_softmax_cross_entropy_with_logits(
        logits=all_act, labels=self.tf_acts)

    self.loss = tf.reduce_mean(
        self.neg_log_prob * self.tf_vt) + self.entropy_mean
        # + tf.losses.get_regularization_loss()
    self.train_op = tf.train.AdamOptimizer(
        self.lr).minimize(self.loss)

def choose_action(self, observation, mode='random', greedy=None):
    # print('observation', observation.shape)
    prob_weights = self.sess.run(self.all_act_prob_soft, feed_dict={
        self.tf_obs: observation})
    # print('prob_weights', prob_weights)
    actions = []
    probs = []
    for i, p in enumerate(prob_weights):
        # print('self.n_action', self.n_action, p)
        p /= np.sum(p)
        # print('p', p)

```

```

# try:
if mode == 'random':
    # prob_rand = np.ones(self.n_action).astype(np.float32) * 1. / self.n_action
    # if greedy is not None and np.random.rand() < greedy:
    #     action = np.random.choice(self.n_action)
    # else:
    # action = np.random.choice(self.n_action, p=p.ravel())
    action = np.random.choice(self.n_action, p=p)
    # p = prob_rand
elif mode == 'max':
    action = np.argmax(p)
else:
    print('choose_action mode is Wrong')
    exit()
actions.append(action) # select action w.r.t the actions prob
probs.append(p)
# except:
#     print(p)
return actions, probs

def store_transition(self, state, algo_action, real_action, reward):
    self.ep_obs.append(state)
    self.ep_algo_action.append(algo_action)
    self.ep_rs.append(reward)
    self.ep_real_action.append(real_action)

def learn(self, tran=None):
    if tran is not None:
        self.ep_obs, self.ep_rs, self.ep_algo_action, self.ep_real_action = tran.all_data()
        # self.ep_obs, self.ep_rs, self.ep_algo_action, self.ep_real_action =
        tran.sample(batch_size=self.batch_size)
        # print('ep_rs', self.ep_rs[0])
        discounted_ep_rs_norm = self._discount_and_norm_rewards()
        # print('discounted_ep_rs_norm', discounted_ep_rs_norm[:10])

        labels = np.ones((len(discounted_ep_rs_norm), self.n_action))
        _, loss, prob_avg, loss_avg = self.sess.run([self.train_op, self.loss, self.prob_avg, self.loss_avg],
        feed_dict={
            self.tf_obs: np.vstack(self.ep_obs), # shape=[None, n_obs]
            self.tf_acts: np.array(self.ep_algo_action), # shape=[None, ]
            self.tf_vt: discounted_ep_rs_norm, # shape=[None, ]
            self.tf_lbs: labels

```

```
})
```

```
    #print('prob_avg', prob_avg)
```

```
    #print('loss_avg', loss_avg)
```

```
    self.ep_obs, self.ep_algo_action, self.ep_real_action, self.ep_rs = [
```

```
    ], [], [], [] # empty episode data
```

```
    return discounted_ep_rs_norm
```

```
def _discount_and_norm_rewards(self, terminal=None):
```

```
    discounted_ep_rs = np.ones_like(self.ep_rs)
```

```
    for t in reversed(range(0, len(self.ep_rs))):
```

```
        discounted_ep_rs[t] = self.ep_rs[t]
```

```
    return discounted_ep_rs
```