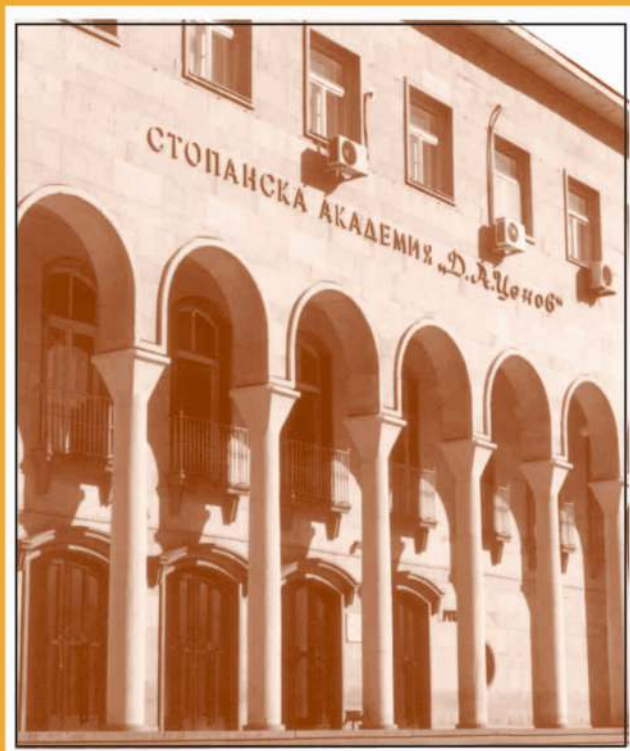




ИНСТИТУТ ЗА НАУЧНИ
ИЗСЛЕДВАНИЯ
ПРИ СТОПАНСКА АКАДЕМИЯ
„Д. А. ЦЕНОВ“ - СВИЩОВ

АЛМАНАХ

НАУЧНИ ИЗСЛЕДВАНИЯ

УСТОЙЧИВ
И ДИГИТАЛЕН ПРЕХОД –
ИНОВАЦИИ, ЗНАНИЕ
И ТРАНСФОРМАЦИЯ
НА ИКОНОМИКАТА

том 34, 2026 г.

Академично издателство „ЦЕНОВ“
Свищов - 2026 г.



ИНСТИТУТ ЗА НАУЧНИ ИЗСЛЕДВАНИЯ
СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ” – СВИЩОВ

АЛМАНАХ НАУЧНИ ИЗСЛЕДВАНИЯ

УСТОЙЧИВ И ДИГИТАЛЕН ПРЕХОД – ИНОВАЦИИ, ЗНАНИЕ И ТРАНСФОРМАЦИЯ НА ИКОНОМИКАТА

ТОМ 34
2026

АКАДЕМИЧНО ИЗДАТЕЛСТВО „ЦЕНОВ” – СВИЩОВ

Издава се със средства от целевата субсидия за научна дейност на СА „Д. А. Ценов”, съгласно Наредбата за условията и реда за оценката и планирането, разпределението и разходването на средствата от държавния бюджет за финансиране на присъщата на държавните висши училища научна или художествено творческа дейност.

РЕДАКЦИОНЕН СЪВЕТ:

Доц. д-р Галина Чиприянова	Главен редактор Стопанска академия „Д. А. Ценов” – Свищов
Проф. д-р Атанас Атанасов	Заместник-главен редактор Стопанска академия „Д. А. Ценов” – Свищов
Проф. д. ик. н. Виктор Чужиков	Киевски национален икономически университет, <i>Киев, Украйна</i>
Проф. д-р Николае Панеа	Университет в Крайова, <i>Крайова, Румъния</i>
Проф. д-р Стефан Симеонов	Стопанска академия „Д. А. Ценов” – Свищов
Проф. д-р Тадия Джукич	Университет Ниш, <i>Ниш, Сърбия</i>
Доц. д-р Анисоара Дуика	Университет Валахия, <i>Търговище, Румъния</i>
Доц. д-р Венцислав Василев	Стопанска академия „Д. А. Ценов” – Свищов
Доц. д-р Валентин Милинов	Стопанска академия „Д. А. Ценов” – Свищов
Доц. д-р Любомир Иванов	Стопанска академия „Д. А. Ценов” – Свищов
Доц. д-р Любка Илиева	Стопанска академия „Д. А. Ценов” – Свищов
Доц. д-р Елена Йорданова	Стопанска академия „Д. А. Ценов” – Свищов
Доц. д-р Зорница Ганчева	Стопанска академия „Д. А. Ценов” – Свищов
Д-р Рейчъл Маритц	Университет в Претория, <i>Претория, Южна Африка</i>

Анка Танева – стилев редактор

Ст. преп. д-р Радка Василева - стилев редактор на английски език

Живка Тананеева – технически секретар

© ИНСТИТУТ ЗА НАУЧНИ ИЗСЛЕДВАНИЯ

© СТОПАНСКА АКАДЕМИЯ „ДИМИТЪР А. ЦЕНОВ”

ISSN 1312-3815



INSTITUTE FOR SCIENTIFIC RESEARCH
D. A. TSENOV ACADEMY OF ECONOMICS – SVISHTOV

SCIENTIFIC RESEARCH ALMANAC

THE SUSTAINABLE AND DIGITAL TRANSITION – INNOVATION, KNOWLEDGE AND ECONOMIC TRANSFORMATION

**VOLUME 34
2026**

TSENOV ACADEMIC PUBLISHING HOUSE

This issue is funded by the state “Ordinance on the terms and procedure for the evaluation and planning, allocation and spending of the state budget funds in financing scientific or artistic activities, intrinsic to state higher schools” for the inherent to the "D. A. Tsenov" Academy of Economics scientific activity.

EDITORIAL BOARD

Assoc. Prof. Galina Chipriyanova, PhD	Editor-in-chief D. A. Tsenov Academy of Economics – Svishtov
Prof. Atanas Atanasov, PhD	Deputy editor-in-chief D. A. Tsenov Academy of Economics – Svishtov
Prof. Viktor Chuzhikov, DSc	Kyiv National Economic University, <i>Kyiv, Ukraine</i>
Prof. Nicolae Panea, PhD	University of Craiova, <i>Craiova, Romania</i>
Prof. Stefan Simeonov, PhD	D. A. Tsenov Academy of Economics – Svishtov
Prof. Tadija Djukic, PhD	University of Nis, <i>Nis, Serbia</i>
Assoc. Prof. Anisoara Duica, PhD	Valahia University of Targoviste, <i>Targoviste, Romania</i>
Assoc. Prof. Ventsislav Vasilev, PhD	D. A. Tsenov Academy of Economics – Svishtov
Assoc. Prof. Valentin Milinov, PhD	D. A. Tsenov Academy of Economics – Svishtov
Assoc. Prof. Lyubomir Ivanov, PhD	D. A. Tsenov Academy of Economics – Svishtov
Assoc. Prof. Lyubka Ilieva, PhD	D. A. Tsenov Academy of Economics – Svishtov
Assoc. Prof. Elena Yordanova, PhD	D. A. Tsenov Academy of Economics – Svishtov
Assoc. Prof. Zornitsa Gancheva, PhD	D. A. Tsenov Academy of Economics – Svishtov
Dr Rachel Maritz	University of Pretoria, <i>Pretoria, South Africa</i>

Anka Taneva – stylistic editor
Sen. Lect. Radka Vasileva – translator
Zhivka Tananeeva – technical secretary

© INSTITUTE FOR SCIENTIFIC RESEARCH

© D. A. TSENOV ACADEMY OF ECONOMICS – SVISHTOV

ISSN 1312-3815

СЪДЪРЖАНИЕ

Раздел I

КРЪГОВА И НИСКОВЪГЛЕРОДНА ИКОНОМИКА, УСТОЙЧИВО РАЗВИТИЕ И ИНОВАЦИИ

Елица Лазарова, Мариела Стоянова, Надежда Веселинова, Юлиян Господинов, Асен Божилов, Димчо Шопов Подходи и механизми за прилагане на отворени иновации в публичния сектор в България	7
Любчо Варамезов, Симеонка Петрова, Любомира Тодорова, Явор Станев, Любомир Николов, Мелинда Петкова Продуктови иновации и устойчиво развитие на европейските малки и средни предприятия в сектора на индустрията	48
Радосвета Кръстева-Христова, Диана Папраданова Интегрираното отчитане като инфраструктура за устойчиво управление: уроци от българската практика (2020–2025).....	85
Евгени Овчинников, Атанаска Решеткова Протеиновият преход в България: концептуални основи и емпирични измерения в контекста на устойчивите хранителни системи	126

Раздел II

ИЗКУСТВЕН ИНТЕЛЕКТ, ДИГИТАЛНА ТРАНСФОРМАЦИЯ И УПРАВЛЕНИЕ НА ЗНАНИЕТО

Тодор Кръстевич Предиктивно и каузално моделиране на мултитъч атрибуция в дигитална среда	159
Симеонка Петрова, Петранка Мидова, Иван Маринов, Зоя Иванова, Цветелина Кабакчиева, Елена Вълчева, Веселин Василев, Йоан Томов, Йордан Стефанов, Преслав Георгиев Въздействие на дигиталната трансформация на търговията върху моделите на потребителско поведение: емпирично изследване на онлайн покупките на български потребители	209

CONTENTS

Section I

CIRCULAR AND LOW-CARBON ECONOMY, SUSTAINABLE DEVELOPMENT AND INNOVATION

- Elitsa Lazarova, Mariela Stoyanova, Nadezhda Veselinova,
Yuliyana Gospodinov, Asen Bozhikov, Dimcho Shopov**
Approaches and mechanisms for implementing open innovation in the public sector
in Bulgaria 7
- Lyubcho Varametzov, Simeonka Petrova, Lyubomira Todorova, Yavor Stanev,
Lyubomir Nikolov, Melinda Petkova**
Product innovation and sustainable development of European small and medium-
sized enterprises in the industry sector 48
- Radosveta Krasteva-Hristova, Diana Papradanova**
Integrated reporting as an infrastructure for sustainable management: lessons from
Bulgarian practice (2020–2025) 85
- Evgeni Ovchinnikov, Atanaska Reshetkova**
The Protein Transition in Bulgaria: conceptual foundations and empirical dimensions
in the context of sustainable food systems 126

Section II

ARTIFICIAL INTELLIGENCE, DIGITAL TRANSFORMATION AND KNOWLEDGE MANAGEMENT

- Todor Krastevich**
Framework for predictive and causal modeling of multi-touch attribution
in a digital environment 159
- Simeonka Petrova, Petranka Midova, Ivan Marinov, Zoya Ivanova, Tsvetelina
Kabakchieva, Elena Valcheva, Vesselin Vassilev, Yoan Tomov, Yordan
Stefanov, Preslav Georgiev**
Impact of digital transformation of trade on consumer behavior patterns: an empirical
study of online purchases of Bulgarian consumers 209

СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ”

АЛМАНАХ НАУЧНИ ИЗСЛЕДВАНИЯ

ТОМ 34

**УСТОЙЧИВ И ДИГИТАЛЕН ПРЕХОД – ИНОВАЦИИ,
ЗНАНИЕ И ТРАНСФОРМАЦИЯ НА ИКОНОМИКАТА**

Даден за печат на 12.03.2026 г., излязъл от печат на 26.03.2026 г.

Поръчка № 18964, тираж: 50 бр.

Издателство и печат: Академично издателство „Ценов”

Свищов, ул. „Цанко Церковски“ 11А

ISSN 1312-3815

ПРЕДИКТИВНО И КАУЗАЛНО МОДЕЛИРАНЕ НА МУЛТИТЪЧ АТРИБУЦИЯ В ДИГИТАЛНА СРЕДА

Проф. д-р Тодор Кръстевич

Резюме

Във все по-сложната екосистема на дигиталния маркетинг определянето на инкременталния принос на различните маркетингови канали към конверсията на потребителите остава фундаментално нерешен проблем. Предмет на настоящото проучване е процесът на мултитъч атрибуция в дигитална среда, а негов обект е предсказуемият и каузален подход към оценката на приноса на каналите. Основната цел е да се разработи цялостна рамка за мултитъч атрибуция, която интегрира вероятностно моделиране (вериги на Марков), машинно обучение (XGBoost) и каузални методи (стойности на Шапли, контрафактуален анализ), за да се осигури по-точна и интерпретируема оценка на влиянието на каналите. Като част от проучването беше изградена и тествана хибридна моделна система върху реални данни за пътуването на потребителите, а нейната ефективност беше сравнена с тази на традиционните евристични модели. Резултатите показват значителни разлики от евристичните подходи и подчертават по-високата предсказуема точност на методите, базирани на машинно обучение (обща точност = 0,984; AUC = 0,846), докато прилагането на обясними AI техники увеличава прозрачността на атрибуцията. Теоретично, проучването комбинира каузални модели, усилващо обучение и интерпретируемост в моделирането на атрибуцията, свързвайки иконометричната теория с анализа на данни, а от гледна точка на управлението предлага възпроизводима рамка за оптимизация на ресурсите и вземане на решения на базата на данни.

Ключови думи: мултитъч атрибуция, каузално заключение, анализ на пътя на клиента, прогностично моделиране, оптимизация на маркетинга.

JEL: M31, M37, C81, C25.

FRAMEWORK FOR PREDICTIVE AND CAUSAL MODELING OF MULTI-TOUCH ATTRIBUTION IN A DIGITAL ENVIRONMENT

Prof. Todor Krastevich, PhD

Abstract

In the increasingly complex ecosystem of digital marketing, determining the incremental contribution of different marketing channels to consumer conversion remains a fundamentally unresolved challenge. The subject of this study is the process of multi-touch attribution in a digital environment, and its object is the predictive and causal approach to assessing the contribution of channels. The main goal is to develop a comprehensive framework for multi-

touch attribution that integrates probabilistic modeling (Markov chains), machine learning (XGBoost), and causal methods (Shapley values, counterfactual analysis) to provide a more accurate and interpretable assessment of channel influence. As part of the study, a hybrid model system was built and tested on real user journey data, and its effectiveness was compared to that of traditional heuristic models. The results show significant differences from heuristic approaches and highlight the superior predictive accuracy of machine learning-based methods (overall accuracy = 0.984; AUC = 0.846), while the application of explainable AI techniques increases the transparency of attribution. Theoretically, the study combines causal models, reinforcement learning, and interpretability in attribution modeling, linking econometric theory with data analysis, and in management terms, it offers a reproducible framework for resource optimization and data-driven decision-making.

Keywords: multi-touch attribution, causal inference, customer journey analysis, predictive modeling, marketing optimization.

JEL: M31, M37, C81, C25.

1. Въведение

Широкото разпространение на каналите за дигитален маркетинг коренно промени начина, по който потребителите взаимодействат с брендовете и рекламодателите оценяват ефективността на маркетинга. За разлика от линейните, едноканални пътища за покупка, които доминираха в традиционните маркетингови модели, съвременните потребителски пътувания се характеризират с динамични, многоканални и с различни устройства взаимодействия (Batra & Keller, 2016; Kannan et al., 2016). Преди да вземат решение за покупка, потребителите се сблъскват с редица дигитални точки на контакт – търсачки, дисплейни реклами, имейли, социални медии и препоръки. Тази сложна мрежа от взаимодействия поставя пред маркетингозите предизвикателството да определят точно стойността на всяка експозиция, която допринася за конверсията – процес, известен като маркетингова атрибуция (Anderl, Becker, et al., 2016a; H. (Alice) Li & Kannan, 2014).

Моделирането на атрибуцията (приписването) служи като аналитична основа за разбиране на инкременталното въздействие на всеки маркетингов канал върху резултатите от конверсията на клиентите. Въпреки технологичните постижения и наличието на подробни данни за кликовете, точното разпределение на заслугите за конверсията между различните точки на контакт остава трайно аналитично предизвикателство. Евристичните модели, като атрибуция на първия и последния контакт, са широко приети поради своята простота и интерпретируемост, но не отчитат нелинейните взаимодействия и последователните зависимости между маркетинговите канали (Berman, 2018). В повечето случаи тези модели приписват пълния принос само на един канал, без да отчитат междинните взаимодействия, които може да са повлияли на формирането на намерението за покупка и вземането на решение (Shao & Li, 2011a). В резултат на това маркетинговите бюджети,

базирани на евристични модели, често се разпределят неефективно, като усилват доминиращите канали за сметка на други, които играят решаваща поддържаща роля в пътя към конверсията.

Недостатъците на евристичните методи доведоха до разпространението на подходи за атрибуция, основани на данни, които се базират на вероятностни и машинно-обучаващи методологии. Вероятностните подходи – по-специално моделите на Марковски вериги – улавят взаимозависимостите между каналите и вероятностите за преход между точките на контакт (Anderl, Becker, et al., 2016a; Gagnius, 2017). В същото време прилагането на алгоритми за контролирано обучение, като логистична регресия, метод на случайната гора и градиентно усилване, позволява оценка на вероятността за конверсия въз основа на исторически закономерности в условията на многоканално въздействие (Chen & Guestrin, 2016; Dalessandro et al., 2012). Тези подходи надхвърлят описателните евристики и се насочват към предиктивни и предписващи анализи, които могат да моделират хетерогенни пътувания на клиенти във виртуална среда и да се адаптират динамично към променящото се поведение на потребителите.

Въпреки своите предимства, съществуващите алгоритмични модели за атрибуция все още се сблъскват с концептуални и методологични предизвикателства. Първо, повечето от тях са чисто корелационни и им липсва причинно-следствена (каузална) интерпретативна рамка, която да прави разграничение между корелация и истински инкрементален ефект (Lewis & Wong, 2022). Второ, „черната кутия“ на много алгоритми за машинно обучение повдига въпроси за интерпретируемостта и отчетността, особено когато решенията включват автоматизирана оптимизация на бюджета или персонализация (Molnar, 2023). Трето, регулаторните промени – като Общия регламент за защита на данните (GDPR) на Европейския съюз и постепенното премахване на бисквитките на трети страни – допълнително усложняват достъпността на данните и проследяването на индивидуално ниво (Saura et al., 2024).

В отговор на тези предизвикателства настоящото проучване допринася за теоретичното и методологичното развитие на мултитъч атрибуцията чрез интегриране на предсказуеми, каузални и поддаващи се на интерпретиране парадигми за моделиране. По-конкретно то предлага хибридна аналитична рамка, която синтезира вероятно моделиране (вериги на Марков), прогнозиране на базата на машинно обучение (XGBoost) и техники за каузални заключения (стойности на Шапли, контрафактуално ML мислене), за да подобри точността и интерпретируемостта на атрибуцията. По този начин то позиционира мултитъч атрибуцията като едновременно икономически и изчислителен проблем, който изисква баланс между статистическа строгост, прозрачност и управленска релевантност.

Обект на изследването е мултитъч атрибуцията в дигиталния маркетинг (MTA) и по-специално – процесът на приписване на заслуги на

различни онлайн канали за постигнатата конверсия. Предмет на изследването е интегрираният подход за предиктивно и каузално моделиране на този процес.

Изследователската теза гласи, че точността, интерпретируемостта и управленската полезност на атрибуционните анализи могат значително да се повишат чрез съчетаване на вероятностни, машинно-обучителни и каузални методи в единна рамка, вместо използване на традиционни евристични модели. В съответствие с тази теза основната цел на проучването е да разработи цялостна аналитична рамка за мултитъч атрибуция, която комбинира горепосочените методи с цел по-надеждно измерване приноса на маркетинговите канали. За реализирането на така формулираната цел са изпълнени следните основни задачи: първо, извършване на критичен анализ на съществуващите модели и концептуални основи на маркетинговата атрибуция; второ, разработване на интегрирана предиктивно-каузална рамка за МТА, съчетаваща вероятностни модели, машинно обучение и каузални заключения; трето, емпирично тестване на предложената рамка върху набор от реални данни и сравнение на резултатите с традиционни атрибуционни модели; четвърто, анализ на получените резултати и формулиране на изводи за теоретичните приноси и практически насоки за маркетинговата дейност.

От теоретична гледна точка настоящото изследване допринася за развитието на маркетинговата аналитика по три важни начина. Първо, то операционализира концепцията за причинно-следствена атрибуция, съгласувайки прогностичното моделиране с принципите на контрафактуалния анализ. Второ, то поставя моделирането на атрибуцията в по-широкия контекст на анализа на потребителското пътуване и науката за поведенческите данни, като по този начин свързва потребителското поведение на микро ниво с разпределението на маркетинговите ресурси. Трето, то проучва потенциала на обучението чрез подсилване като динамичен механизъм за оптимизация на разпределението на бюджета между каналите – посока, която е успоредна на нововъзникващите проучвания в областта на автоматизацията на маркетинговите решения (Cai et al., 2023; Song et al., 2024).

Останалата част от студията е структурирана по следния начин. Раздел 2 предоставя разширена информация за контекста на изследването, като поставя мултитъч атрибуцията в рамките на еволюцията на маркетинговата аналитика и вземане на решения в дигитална среда. Раздел 3 предлага изчерпателен преглед на литературата, като акцентира върху последните постижения в областта на вероятностното, алгоритмичното и каузалното моделиране. Раздел 4 разработва теоретичната рамка за предиктивна и каузална атрибуция. Раздел 5 представя емпиричната методология и процедурата за моделиране, следван от раздел 6, в който се обсъждат резултатите, излагат се теоретичните импликации и ограничения, като се дават и практически препоръки за управлението на дигиталния маркетинг. Накрая, в раздел 7 се извеждат насоки за бъдещи изследвания и се призовава за

интегриране на интерпретируемостта и каузалността в маркетинговите системи за атрибуция от следващо поколение.

2. Обосновка на изследването

2.1. Теоретични основи на моделирането на атрибуцията

Моделирането на атрибуцията (приписването) представлява критична пресечна точка между маркетинговата теория, поведенческата икономика и науката за данните. Концептуално, атрибуцията може да се разбере като количествена оценка на причинно-следствения принос – доколко всеки маркетингов канал, експозиция или точка на контакт допринасят за крайната конверсия (Berman, 2018; H. (Alice) Li & Kannan, 2014). Този аналитичен процес има за цел да разкрие добавената стойност, създадена от различните взаимодействия по време на пътуването на клиента, като по този начин насочва оптималното разпределение на бюджета и стратегията за използване на маркетинговите канали.

Логиката, която стои в основата на моделите за атрибуция, произтича от две фундаментални теоретични традиции: теорията за комуникационния отговор и теорията за вземането на решения от потребителите. Ранните модели в рекламата и потребителското поведение, като модела на йерархията на ефектите (Lavidge & Steiner, 1961) и теоретичната рамка AIDA (Strong, 1925), концептуализираха процеса на покупка като линейно развитие през етапите на осведоменост, интерес, желание и действие. Макар тези идейни рамки да са били от съществено значение за ранните маркетингови идеи, техните линейни и едноканални предположения стават все по-неадекватни в днешната фрагментирана, многоканална дигитална среда (Court et al., 2009; Neslin et al., 2006).

Съвременната екосистема на дигиталния маркетинг, характеризираща се с интерактивност, персонализация и непрекъсната комуникация, налага преход от статични към динамични, базирани на данни атрибуционни парадигми. Тази трансформация е в съответствие с това, което Wedel и Kannan (2016) описват като „революция на големите данни“ в маркетинга, при която аналитичната прецизност се постига чрез високочестотни данни на индивидуално ниво и компютърно моделиране. В този контекст моделирането на атрибуцията се разви от описателно, евристично упражнение в предсказващо и каузално начинание, интегриращо концепции от иконометрията, стохастичните процеси и изкуствения интелект (de Haan et al., 2016a; Kannan & Li, 2017).

В основата си атрибуцията служи не само като проблем за измерване, но и като проблем за извод за поведението. Тя има за цел да оцени как потребителите реагират на многоканални стимули и как тези реакции взаи-

модействат във времето. Тази двойствена природа позиционира атрибуционното моделиране като обединяваща конструкция между моделирането на маркетинговия микс (МММ) и анализа на клиентите: докато МММ предоставя агрегирани, дългосрочни еластичности в различните медии, атрибуционните модели се фокусират върху секвенции от експозиции на микрониво, които водят до конверсии (Anderl & Weichert, 2016; Chatterjee et al., 2003). Интегрирането на двете перспективи е от решаващо значение за постигането на измерване на ефективността на цялата фуния на продажбите и за свързване на решенията за маркетингови инвестиции с поведенческите резултати.

2.2. Осмисляне на мултитъч атрибуцията в контекста на пътуването на клиента

За разлика от моделите с един източник, мултитъч атрибуцията (МТА) приема, че множество маркетингови канали съвместно допринасят за резултатите от конверсията. Всяко взаимодействие на потребителя с комуникацията на даден бранд, било то пряко или опосредствано, онлайн (като например кликуване върху реклама, посещение на уебсайт или отваряне на имейл) или офлайн, ще наричаме „точки на съприкосновение“ (Wiesel, Pauwels, & Arts, 2011). С разрастването на дигиталните екосистеми броят на точките на съприкосновение на потребител се е увеличил експоненциално, като взаимодействията обхващат органично търсене, платно търсене, социални медии, препратен трафик, дисплейни реклама и директни посещения на уебсайтове. В резултат на това пътят, който изминава клиентът, се формира като нелинейна, стохастична поредица от експозиции, които варират както между отделните лица, така и във времето (Anderl & Weichert, 2016; Bucklin & Sismeiro, 2009a).

Тази сложност поставя няколко аналитични предизвикателства. Първо, зависимостта от пътя предполага, че вероятността за конверсия зависи не само от най-скорошното взаимодействие, но и от последователността и честотата на предишните взаимодействия (Anderl, Becker, et al., 2016a; de Haan et al., 2016a; H. (Alice) Li & Kannan, 2014; Xu et al., 2014). Второ, взаимозависимостта на каналите предполага, че някои канали действат синергично (например социалните медии стимулират намерението за търсене), докато други могат да канибализират или да разводнят ефектите си (Kireyev et al., 2016). Трето, хетерогенността в поведението на клиентите, обусловена от демографски фактори, предпочитания и ситуационни контексти, изисква гъвкави модели, способни да улавят нелинейни ефекти и ефекти от взаимодействието (Danaher & Dagger, 2013; Rutz & Bucklin, 2011).

Стохастичната природа на пътя на клиентите привежда МТА в концептуално съответствие с Марковските процеси, които моделират преходите между състояния във времето (Gagnius, 2017). В маркетинговия смисъл всеки канал представлява „състояние“, а преходите на клиентите

между тези състояния могат да бъдат описани вероятно. Марковските МТА модели количествено измерват вероятността, премахването на даден канал да промени общата вероятност за конверсия – метод, понякога наричан моделиране на ефекта от премахването (Anderl, Becker, et al., 2016a). Този подход подобрява евристичните методи, като отчита косвените ефекти и зависимостите между каналите. Въпреки това, той все още предполага стационарен процес и често не успява да улови влиянието на контекстуалните фактори, сезонността и външните шокове (Kakalejčik et al., 2021).

2.3. От евристични модели към атрибуция, базирана на данни

Евристичните модели за атрибуция (приписване) продължават да доминират в практиката поради своята простота, лекота на интерпретация и съвместимост със съществуващите платформи за дигитална аналитика (например Google Analytics, Adobe Analytics). Сред най-популярните евристични модели се нареждат:

- Приписване на първото докосване/съприкосновение, което приписва пълна заслуга на първоначалното взаимодействие на потребителя с дадена точка на взаимодействие от конкретен канал;
- Приписване на последното докосване/съприкосновение, което приписва конверсията на последното взаимодействие преди покупката;
- Линеино приписване, което разпределя заслугата равномерно между всички точки на докосване; и
- Модели на времево затихване, които приписват по-голяма тежест на последните взаимодействия.

Макар тези практични правила да са удобни от оперативна гледна точка, те имат фундаментални теоретични и емпирични ограничения. Те предполагат адитивност и независимост между каналите, като пренебрегват динамичната и интерактивна природа на вземането на решения от страна на потребителите (Berman, 2018). Освен това те пренебрегват времевите зависимости, ефектите на пренасяне (carryover) и пределния принос на спомагателните канали (Shao & Li, 2011a). Емпиричните данни систематично показват, че евристичните модели дават изместени и пристрастни оценки за ефективността на каналите, което води до неефективно разпределение на маркетинговия бюджет и потенциално засилва съществуващите медийни силози¹ (Dalessandro et al., 2012; Tout, 2023).

Преходът към атрибуция, базирана на данни (АБД), произтича от признанието, че евристичните модели не успяват да уловят сложността на поведението и статистическите взаимоотношения между каналите. Моделите, базирани на данни, разчитат на исторически данни на ниво потребител,

¹ В този контекст, с понятието „медийни силози“ обозначаваме организационна и аналитична фрагментация, при която различните маркетингови канали се управляват, оценяват и оптимизират изолирано един от друг, без интегриран подход.

за да направят изводи за вероятностни или предсказуеми взаимоотношения между експозициите и резултатите (Abhishek et al., 2012a). Вероятностните подходи (напр. вериги на Марков, стойности на Шапли) и алгоритмите за машинно обучение (напр. логистична регресия, XGBoost, невронни мрежи) представляват двете доминиращи методологични направления в рамките на АБД. Всеки от тях предлага уникални предимства – вероятностните модели наблягат на интерпретируемостта и обективността, докато моделите за машинно обучение дават приоритет на прогностичната точност и адаптивността (Anderl, Schumann, et al., 2016; Berman, 2018; Chen & Guestrin, 2016).

Последните технологични постижения в облачната аналитика и инфраструктурата за големи данни (например Google BigQuery, Amazon Redshift, Snowflake) допълнително демократизираха достъпа до данни за кликанията на потребителско ниво, което позволява на организациите да прилагат тези модели в голям мащаб (Google Analytics, 2022). Нарастващата интеграция на програмни рамки с отворен код, като пакетите 'ChannelAttribution' и 'parsnip' на R (Altomare & Loris, 2023; Kuhn & Vaughan, 2024), позволява на изследователите и практиците да възпроизвеждат и валидират модели за атрибуция в различни набори от данни. Тази еволюция към отворени, възпроизводими анализи бележи парадигмена промяна от собственически модели, обвързани с конкретен доставчик, към прозрачни, управлявани от общността методологии.

2.4. Поведенческа интерпретация на ефектите на пренасяне и разпространение

Стохастичната природа на потребителското поведение в онлайн среда въвежда два важни поведенчески феномена: ефекти на пренасяне (ан. carryover) и ефекти на разпространение (ан. spillover).

- Ефектите на пренасяне се отнасят до ефектите на паметта в рамките на един канал, при които миналите експозиции в рамките на същия канал влияят върху вероятността за бъдеща конверсия (Naik et al., 1998).
- Ефектите на разпространение улавят влиянията между каналите, при които експозицията към един рекламен носител влияе върху реакцията към друг рекламен носител (Kireyev et al., 2016).

В онлайн средата тези ефекти се усилват от алгоритмите за персонализация и механизмите за пренасочване. Например излагането на реклама в социалните медии може да не доведе директно до конверсия, но може да увеличи вероятността последваща реклама в търсачката да предизвика покупка (Dinner et al., 2014). Такива взаимодействия поставят под въпрос идеята за независимия принос на каналите и налагат необходимостта от модели за атрибуция, способни да оценят синергиите между различните медии.

Последните емпирични проучвания показват, че игнорирането на тези ефекти води до систематично подценяване на каналите нагоре по веригата

(например дисплейни и социални) и надценяване на каналите надолу по веригата (например търсене, директни) (Abhishek et al., 2012a, 2012b; de Naan et al., 2016b). Тази поведенческа взаимозависимост подчертава необходимостта от модели, които третираят пътуването на клиента като взаимосвързана последователност, а не като набор от изолирани събития. Съответно, за да се уловят тези нелинейни зависимости, са се появили усъвършенствани модели като Марковски вериги от по-висок ред и модели на последователности, базирани на машинно обучение (например рекурентни невронни мрежи) (N. Li et al., 2018; Sun et al., 2020).

2.5. От описателна към предиктивна и каузална атрибуция

Еволюцията на моделирането на атрибуцията отразява по-широките тенденции в маркетинговата аналитика: преход от описателни към предсказващи и все по-често към каузални рамки за изводи (Kannan & Li, 2017; Lewis & Wong, 2022). Описателните модели отговарят на въпроса „Какво се случи?“, предсказващите модели оценяват „Какво е вероятно да се случи?“, а каузалните модели се опитват да обяснят „Защо и в каква степен?“. В тази траектория мултитъч атрибуцията е преминала от евристика, базирана на правила (описателна), към вероятностна оценка (предсказваща) и към каузална атрибуция – количествено измерване на инкременталния ефект на всеки канал при контрафактуални сценарии (Ben Mrad & Hnich, 2024; Sinha et al., 2022; Varian, 2016).

В нововъзникващата каузална парадигма атрибуцията вече не се разглежда единствено като проблем на класифициране или прогнозиране, а като задача за контрафактуално разсъждение. Централният въпрос става: какво би се случило, ако потребителят не беше изложен на въздействието на определен канал или поредица от канали? Това привежда атрибуцията в съответствие с теорията за потенциалните резултати и структурното каузално моделиране (Pearl, 2019). Съвременните рамки все по-често комбинират машинно обучение с каузално заключение, за да оценят ефектите от третирането в рамките на сложни процеси на генериране на данни (Athey & Imbens, 2019). В маркетинга този хибриден подход улеснява не само точното прогнозиране на вероятностите за конверсия, но и оценката на инкременталния лифт – каузалния принос на всеки канал или кампания към наблюдаваните резултати.

Този преход означава по-дълбока теоретична преориентация: от асоциативно измерване към каузално обяснение. Чрез включването на каузални изводи и обясняващи AI техники, предиктивните атрибуционни модели придобиват интерпретативна валидност, гарантирайки, че препоръките за оптимизация отразяват истинските поведенчески ефекти, а не фалшиви корелации. В резултат на това атрибуционното моделиране се превръща не само в инструмент за разпределяне на „заслуги“, но и в средство за научно разбиране на поведението на дигиталните потребители.

3. Преглед на специализираната литература

3.1. Еволюцията на мултитъч атрибуцията в маркетинговата аналитика

През последното десетилетие мултитъч атрибуцията (МТА) се превърна от оперативна аналитична техника в теоретично обоснована и изчислително сложна област в маркетинговата наука. По-ранните изследвания признават и подчертават ограниченията на детерминистичните и основани на правила подходи, като призовават за по-стриктни вероятностни и причинно-следствени основи (Li & Kannan, 2014; Shao & Li, 2011). В академичната литература от 2011 г. насам се разкрива прогресивна интеграция на иконометричната теория, вероятностното моделиране и машинното обучение при справянето с основното предизвикателство на атрибуцията: количествено измерване на инкременталния ефект на всеки канал в динамичен, мултиканален контекст.

Shao и Li (2011a) са сред първите, които предлагат рамка за мултитъч атрибуция, базирана на данни, основана на верижна логистична регресия. Тяхната работа показва, че атрибуцията може да се разглежда като проблем на контролираното машинно обучение, при който целта е да се предскаже вероятността за конверсия въз основа на последователности от експозиции. Този подход създава мост между иконометричното моделиране и предсказуемия анализ. Малко след това Abhishek, Fader и Hosanagar (2012a) разработват многоетапен атрибуционен модел, основан на теорията за потребителската фуния, като прилагат структурна иконометрия за моделиране на преходите между осведоменост, обмисляне и покупка. Тяхната аналитична рамка допринесе за по-доброто разбиране на последователното влияние между каналите, като илюстрира, че експозициите в медиите нагоре по веригата оказват забавени, но измерими ефекти върху конверсионните проценти надолу по веригата.

Успоредно с това Dalessandro et al. (2012) въвеждат причинно-следствено мотивирана атрибуция, като интегрират каузалното заключение в измерването на рекламата въз основа на кликания. Използвайки контрафактуално разсъждение и инкрементално моделиране, тяхното изследване установява, че атрибуцията трябва да оценява не само корелацията, но и причинно-следственото изменение на експозицията, предвиждайки по-късни развития в каузалното машинно обучение (Athey & Imbens, 2019). Тези ранни разработки (2011–2014) поставят основите на една промяна в цялата изследователска парадигма: от детерминистични модели, базирани на правила, към подходи, базирани на данни, вероятностни и в крайна сметка каузални, които по-добре съответстват на съвременните маркетингови данни.

3.2. Вероятностни подходи за моделиране на атрибуцията: вериги на Марков и модел на Шапли

Вероятностните модели за атрибуция се появяват като първото статистически значимо подобрене в сравнение с евристичните правила, като предлагат статистическа рамка за моделиране на пътя на клиентите като стохастични процеси. Моделите, основаващи се на Марковска верига, третираат всеки маркетингов канал като състояние в една преходна система, като оценяват вероятността, потребителят да премине от един канал към друг преди конверсията (Anderl, Becker, et al., 2016a; Gagnius, 2017). Тези модели могат количествено да измерят както вероятностите за преход, така и ефектите от премахването – как изключването на даден канал променя общата вероятност за конверсия. По този начин те предоставят по-нюансирана представа за взаимозависимостта между точките на контакт.

Anderl et al. (2016a) са първите, които формализират подхода на Марковската верига в маркетинговите проучвания, използвайки графични представяния на пътуването на клиентите. Техните заключения, публикувани в *International Journal of Research in Marketing (IJRM)*, показват, че вероятностните модели преразпределят атрибуционните кредити по-равномерно в сравнение с евристиките на последното съприкосновение, особено за спомагателни канали за дисплейна реклама и препращащ трафик. Kakalejćik et al. (2018, 2021) разширяват този подход, като използваха Марковски вериги от по-висок ред, за да уловят ефектите на паметта в последователните данни за експозицията. Такива модели отчитат по-дълги зависимости в пътуването на клиента, което води до по-реалистично разпределение на кредитите за многоетапни взаимодействия.

Същевременно изследователите възприемат и методи от теорията на игрите, като например стойността на Шапли (Shapley, 1953, 2009), за да разгледат справедливостта при разпределянето на заслугите на отделните канали. Атрибуцията, базирана на стойността на Шапли, изчислява средния пределен принос на всеки канал към всички възможни комбинации от канали, като по този начин гарантира математически обосновано разпределение на заслугите (Zhao et al., 2018). Berman (2018) в своя публикация в *Marketing Science* сравнява моделите на стойността на Шапли с традиционните евристични и Марковски методи, демонстрирайки тяхната теоретична съгласуваност с принципите на кооперативната теория на игрите. Въпреки това, независимо от своята интерпретативна елегантност, моделите на Шапли се сблъскват с практически ограничения по отношение на мащабируемостта – изчислителната сложност нараства експоненциално с броя на каналите, което ограничава тяхното използване в големи цифрови екосистеми (Tout, 2023). Последните постижения решават в известна степен проблема с мащабируемостта чрез използването на приблизителни методи на Шапли и Монте Карло симулации, които оценяват пределните приноси чрез извадки (Ribeiro et al., 2016). Предлагат се и хибридни модели „Марков-Шапли“,

които комбинират вероятностни преходи с механизми за осигуряване на справедливост от теорията на игрите, за да се постигне баланс между интерпретируемост и изчислителна осъществимост (Gaur et al., 2024; Singal et al., 2019).

Като цяло, вероятностните подходи представляват статистически значима крачка напред към атрибуцията, базирана на данни, като предлагат интерпретируемост, обективност и теоретична строгост. Въпреки това, те остават ограничени по отношение на предсказуемостта и адаптивността си, особено в динамични или многомерни маркетингови среди. Това ограничение проправи пътя за алгоритмични и базирани на машинно обучение подходи, които дават приоритет на точността на прогнозите и гъвкавостта на моделите.

3.3. Атрибуция, базирана на алгоритми и машинно обучение

Възходът на машинно обучение (ML) коренно промени ландшафта на атрибуцията, като въведе методи, способни да научават нелинейни, многомерни взаимовръзки в данните за пътуването на клиентите. Алгоритмичните модели за атрибуция третираат проблема като задача за класификация или регресия под наблюдение: при даден набор от точки на допир моделът предсказва вероятността за конверсия.

Логистичната регресия е първият алгоритмичен метод, адаптиран за МТА, който служи като основа за много комерсиални системи (Cramer, 2003b, 2003a; Shao & Li, 2011a). Въпреки че са интерпретируеми, логистичните модели предполагат линейност и независимост на предикторите – условия, които рядко се срещат в реалните данни за кликванията. За да преодолее тези ограничения, изследователите започнаха да прилагат ансамблови методи като случайни гори и машини за усилване на градиента (Chen & Guestrin, 2016). Тези алгоритми итеративно изграждат дървета на решения, за да уловят нелинейните взаимодействия между каналите, постигайки висока точност на прогнозите и устойчивост на шума.

Extreme Gradient Boosting (XGBoost), въведен от Chen и Guestrin (2016), се наложи като доминиращ алгоритъм в изследванията в областта на предиктивното атрибутиране. Tout (2023) демонстрира, че XGBoost систематично превъзхожда както евристичните, така и вероятностните базови линии, като предоставя по-точни прогнози за вероятността от конверсия и по-балансирано приписване на важността на каналите. Гъвкавостта на този алгоритъм позволява интеграция с техники за важност на признаците и SHAP (SHapley Additive exPlanations), което дава възможност за интерпретируеми прозрения в поведението на модела (Lundberg & Lee, 2017).

Интегрирането на обяснима изкуствена интелигентност (XAI) в моделирането на атрибуцията представлява важна стъпка към съгласуване на прогностичната ефективност с интерпретируемостта (Christoph Molnar, 2022; Ribeiro et al., 2016). Концептуалната рамка Local Interpretable Model-

agnostic Explanations (LIME) и методологията SHAP, предложена от Lundberg & Lee (2017), осигуряват прозрачност по отношение на това, как всяка входяща променлива допринася за прогнозираните резултати. В маркетинговия контекст това означава, че анализаторите могат да проследят как всяка точка на контакт – като платена реклама в търсачката или отварянето на имейл – влияе върху прогнозираната вероятност за конверсия за даден индивид или сегмент. Това е особено ценно за надеждното внедряване на изкуствен интелект в маркетинговия анализ, където лицата, вземащи решения, изискват обяснения за автоматизирания препоръки за разпределение на бюджета (Saura et al., 2024).

Последните проучвания разширяват тези алгоритмични подходи чрез използване на архитектури за дълбоко обучение. Li et al. (2018) предлагат модела Deep Neural Attention Multi-Touch Attribution (DNAMTA), който интегрира рекурентни невронни мрежи (RNN) с механизми за внимание², за да улови както последователните, така и контекстуалните зависимости в пътуванията на потребителите. Техните експерименти с реални данни за електронна търговия показват, че моделите, базирани на внимание, превъзхождат моделите на Марков и XGBoost по точност на прогнозите, като същевременно предлагат визуална интерпретируемост на важността на каналите. По подобен начин Sun, Nasraoui и Shafto (2020) въведоха подход за последователно обучение, който моделира многоетапни конверсии като временни зависимости. Тези постижения бележат преход към обучение за представяне в атрибуцията, при което моделите научават латентни вграждания на канали и потребители, което позволява обобщаване в различни кампании и контексти.

3.4. Каузална инференция и моделиране на въздействието върху атрибуцията

Въпреки че моделите за машинно обучение се отличават с висока точност на прогнозите, те често не предлагат причинно-следствена интерпретируемост. Предизвикателството се състои в разграничаването на истинските причинно-следствени ефекти от маркетинговите експозиции от

² Механизмите на вниманието (attention mechanisms) са ключова иновация в дълбокото обучение, която позволява на невронните мрежи да фокусират вниманието си върху най-релевантните части от входните данни при генерирането на резултат. В контекста на мултитъч атрибуция, когато анализираме потребителско пътуване с множество точки на съприкосновение (например: органично търсене → дисплей реклама → имейл → платено търсене → конверсия), механизмът на внимание динамично определя кои взаимодействия са най-важни за крайната конверсия, вместо да третира всички точки на контакт еднакво. Тъй като традиционните рекурентни невронни мрежи обработват последователности, като компресират цялата историческа информация в един скрит вектор на състояние, това създава проблем с дългите последователности – информацията от ранните точки на съприкосновение може да се "забрави". Механизмът на внимание решава този проблем.

мнимите корелации, обусловени от систематични грешки в таргетирането (Lewis & Wong, 2022). Този проблем е мотивирал вълна от изследвания в областта на каузалното и инкременталното моделиране на маркетинговата атрибуция (Athey & Imbens, 2019; Loureiro, 2023).

Изследванията в областта на причинно-следствени изводи, основани на каузалния модел на Rubin (1974) и структурното каузално моделиране (Pearl, 2019), позволяват оценка на ефектите от лечението при контрафактни условия. При атрибуцията това се превежда като оценка на инкременталното повишение от даден канал: колко допълнителна вероятност за конверсия възниква в резултат на експозицията, като всички останали фактори остават постоянни. Инкременталните модели, известни още като модели за въздействието, оценяват директно това каузално изменение чрез моделиране на диференциалната реакция между третираните и контролните групи (Haughton et al., 2025; Kane et al., 2014).

Последните публикации в реномирани научни списания като Information Systems Research и IJRM илюстрират сближаването между каузалната инференция и машинно обучение. Например Lewis (2022) предлага рамката Causal Multi-Touch Attribution (CMTA), която интегрира претеглен показател за склонност и дървовидните модели, за да отчете смесващите фактори в емпирично наблюдаваните рекламни данни. По подобен начин Ahrens et al. (2025) демонстрират потенциала на двойното машинно обучение (DML) за безпристрастна оценка на ефектите от въздействието в многомерни маркетингови масиви от данни. DML комбинира гъвкави ML модели (например, градиентно усилване, невронни мрежи) с иконометрична ортогонализация, за да раздели предсказуемите и причинните компоненти, постигайки както точност, така и интерпретируемост.

В нововъзникващата каузална парадигма атрибуцията вече не се разглежда единствено като проблем на класифициране или прогнозиране, а като задача за контрафактуално разсъждение. Централният въпрос става: какво би се случило, ако потребителят не беше изложен на определен канал или поредица от канали? Тази преориентация привежда атрибуцията в съответствие с рамката на потенциалните резултати и принципите на структурното каузално моделиране (Pearl, 2019). В рамките на тази парадигма последните изследвания са усъвършенствали методите за оценка на инкременталното или каузалното въздействие на маркетинговите експозиции, като използват както бейсиански структурни модели, така и каузални оценяващи, базирани на машинно обучение. Например Brodersen, Gallusser, Koehler, Remy и Scott (2015) предлагат бейсиански структурни модели на времеви редове за извеждане на причинно въздействие от наблюдателни маркетингови данни, докато Kelly, Vaver и Koehler (2018) формализират дигиталната атрибуция като проблем на причинно-следственото заключение, основано на потенциални резултати. Въз основа на тези разработки Du, Zhong, Nair, Cui и Shou (2019) въвеждат CausalMTA, подход с рекурентна невронна мрежа,

който оценява инкременталния принос на рекламните канали, като коригира смесващите се пристрастия. Тези постижения илюстрират как машинното обучение и причинно-следственото заключение могат да се комбинират, за да се оцени ефектът от лечението при сложни процеси на генериране на данни (Athey & Imbens, 2019). В маркетинговия контекст този хибриден подход улеснява не само точното прогнозиране на вероятностите за конверсия, но и надеждното оценяване на инкременталния ръст – истинския каузален принос на всеки канал или кампания към наблюдаваните резултати.

Като цяло тази тенденция в изследваната литература е белег за концептуално узряване на изследванията в областта на проучванията на маркетинговата атрибуция. Приписването вече не се разглежда само като задача за прогнозна класификация, а като проблем за каузална оценка, изискващ контрафактуално разсъждение. Интегрирането на иконометрични инструменти и инструменти за машинно обучение позволява по-точно представяне на основните поведенчески процеси, които определят реакцията на потребителите.

3.5. Хибридни подходи и техники за моделиране на атрибуцията

Изхождайки от вероятностни и каузални основи, най-новите проучвания изследват хибридни рамки, които комбинират множество методологични парадигми. Тези модели използват интерпретируемостта на вероятностните структури, гъвкавостта на машинно обучение и възможностите за оптимизация на обучението чрез подсилване (ан. reinforcement learning, RL).

Обучението чрез подсилване (RL) се е наложило като естествено продължение на динамичното разпределение на бюджета и адаптивната атрибуция (Song et al., 2024). Моделите на RL концептуализират маркетинговите взаимодействия като последователни процеси на вземане на решения, при които агентът (маркетологът) се научава да разпределя бюджета или импресиите между каналите, за да максимализира дългосрочните конверсии или стойността на клиента за целия му жизнен цикъл. От гледна точка на атрибуцията RL предоставя механизъм за оптимизиране на политиките чрез „научаване“ как постепенното излагане на всеки канал влияе на траекториите на конверсиите надолу по веригата.

Song et al. (2024) въвеждат рамка за обяснимо дълбоко обучение чрез подсилване (DRQN-Attention) за привличане на клиенти, демонстрирайки как агентите могат адаптивно да научават оптимални последователности от канали, като същевременно поддържат интерпретируемост чрез визуализация на вниманието. По подобен начин в последните изследвания са използвани алгоритми за многоръки бандити (ан. multi-armed bandit, MAB), за да се коригират динамично теглата на атрибуцията и разпределението на бюджета въз основа на обратна връзка в реално време за конверсията. Подобни аналитични рамки позволяват непрекъснато учене и адаптивна оптимизация на маркетинговите инвестиции, което представлява парадигмена промяна от

ретроспективно присвояване на заслуги към ориентирано към бъдещето и основано на данни вземане на решения (Avadhanula et al., 2021; Soboleva et al., 2025).

Хибридните модели интегрират каузални и основани на подсилване подходи за машинно обучение, за да се справят с компромиса между проучване и използване, присъщ на моделирането на атрибуцията. Например каузалното обучение чрез подсилване (CRL) въвежда каузални априори в RL средата, гарантирайки, че политиките за обучение се ръководят от каузални взаимоотношения, а не от чисто емпирични корелации (Zhang et al., 2020). Тези постижения като цяло сигнализират за преход към автономни маркетингови системи, способни на каузално обучение в реално време и адаптивно разпределение на ресурсите – еволюция, която е в съответствие с по-широката визия за управление на маркетинга, задвижвано от изкуствен интелект (Wedel & Kannan, 2016).

3.6. Нови перспективи и теоретична интеграция

Синтезът на очертаните изследователски направления подчертава продължаващата конвергенция към интегрирани, интерпретируеми и каузални рамки за мултитъч атрибуция. Съвременните проучвания все повече се застъпват за обясними, запазващи неприкосновеността на личния живот и каузални системи за изкуствен интелект, способни да работят при регулаторни ограничения, като същевременно запазват предсказуемата си сила (Christoph Molnar, 2022; Saura et al., 2024).

Това ново поколение изследвания в областта на МТА набляга на три ръководни принципа:

1. Интерпретируемост – гарантиране, че алгоритмичните модели предоставят разбираеми за човека обосновки за присвояване на заслуги на различните канали и точки на съприкосновение (Lundberg & Lee, 2017; Ribeiro et al., 2016).
2. Каузалност – разграничаване на истинското инкрементално въздействие от корелацията чрез контрафактичен анализ и експериментална валидация (Athey & Imbens, 2019).
3. Адаптивност – използване на алгоритми за обучение чрез подсилване и за динамична оптимизация на бюджета в условия на несигурност (Song et al., 2024).

Чрез комбинирането на тези принципи, изследваната област преминава от статични, ретроспективни системи за измерване към интелигентни, самооптимизиращи се екосистеми за атрибуция, които интегрират прогнозиране, обяснение и причинно-следствено мислене. Тази траектория подчертава нарастващото сближаване между маркетинговия анализ и науката за данните, като поставя атрибуцията на преден план в процеса на вземане на прогностични, предписващи и етични маркетингови решения.

4. Теоретична рамка на изследването

4.1. Основи на предиктивното и каузално моделиране на атрибуцията

Както се надяваме, става ясно от литературния обзор, че теоретичната основа за мултитъч атрибуцията се намира на пресечната точка между маркетинговата наука, анализа на поведенчески данни и иконометрията. Исторически погледнато, атрибуционните модели се развиха като инструменти за разлагане на ефектите от маркетинговите канали върху резултатите от конверсията. Ранните подходи, базирани на правила, предлагаха евристични приближения на влиянието на потребителите, но им липсваше теоретична основа в поведенческия отговор или статистическа значимост. Появата на вероятностни подходи и подходи, базирани на машинно обучение, предоставиха по-солидна аналитична основа, като се третираше пътуването на клиента като стохастичен и генериращ данни процес, а не като детерминирана последователност.

Основната теза при формулирането на рамката за прогностично и каузално приписване е, че пределният принос на даден канал може да се оцени като очакваната промяна във вероятността за конверсия, резултат от излагането на този канал, при условие че се вземат предвид предишното пътуване на клиента и контекстуалните фактори. Тази концептуализация е в съответствие както с вероятностното моделиране (марковска динамика на прехода), така и с каузалната инференция (потенциални резултати и контрафактуално разсъждение). Предсказуемият компонент оценява условните вероятности за конверсия; причинно-следственият компонент изолира инкременталния ръст, като сравнява наблюдавани и контрафактични сценарии на експозиция (Athey & Imbens, 2019; Lewis & Wong, 2022). В така предлаганата концептуална рамка се вметват три теоретични перспективи:

1. Вероятностно моделиране на атрибуцията с верига на Марков – моделиране на зависимостите между каналите и вероятностите за преход;
2. Предиктивно машинно обучение – оценяване на вероятността за конверсия чрез гъвкави, нелинейни взаимоотношения; и
3. Каузална инференция – установяване на посоката и величината на истинските инкрементални ефекти.

Следващите подраздели разглеждат подробно тези теоретични стълбове и тяхното интегриране в усъвършенствана аналитична архитектура за маркетингова атрибуция.

4.2. Вероятностно моделиране на приписването: пътуването на клиента като Марковски процес

Вероятностната атрибуция предоставя структурната основа на предлаганата аналитична рамка. При нея поведението на клиента се описва като

марковски процес, при който всеки маркетингов канал представлява дискретно състояние в рамките на пътуването на клиента, а преходите между състоянията C_t следват стохастичен процес, характеризиращ се с вероятности за преход $P(C_{t+1} | C_t)$. Тези вероятности определят възможността, даден потребител, изложен на канал C_t , да взаимодейства след това с канал C_{t+1} , преди да достигне крайното състояние – конверсия или изход/липса на конверсия (Gagnius, 2017).

Това представяне позволява оценка на ефектите от премахването, концепция, формализирана от Anderl et al. (2016a), измерваща спада в общата вероятност за конверсия, когато даден канал е хипотетично премахнат от структурата на пътя. Каналите с по-големи ефекти вследствие на премахването се считат за по-влиятелни при стимулирането на конверсиите. Марковските модели от по-висок ред разширяват тази хипотеза, като включват ефекти на паметта, улавяйки зависимости отвъд непосредствено предшестващото състояние (Kakalejčík et al., 2018).

Вероятностните модели формират първия теоретичен слой на предложената рамка, представяйки поведенческата структура на пътуванията на клиентите. Въпреки това, тяхната описателна сила трябва да бъде допълнена от предсказващи и каузални механизми, които могат да правят изводи за бъдещи резултати и да разграничават корелацията от причинно-следствената връзка.

4.3. Предиктивно моделиране: машинното обучение като механизъм за извличане на прозрения въз основа на данни

Вторият теоретичен стълб на концептуалната рамка е предиктивното (прогностично) моделиране чрез машинно обучение (ML). Атрибуцията, базирана на ML, разглежда връзката между експозициите на каналите и конверсията като проблем на контролирано обучение. Всяко наблюдение съответства на индивидуално пътуване, представено като вектор от взаимодействия между каналите и контекстуални характеристики. Задачата е да се оцени условната вероятност за конверсия $P(Y = 1 | X)$, където X обозначава вектора на характеристиките на точките на съприкосновение, честотите, времевите интервали и подреждането на каналите.

За разлика от традиционните иконометрични модели, ML алгоритмите като Extreme Gradient Boosting (XGBoost), Random Forests и Neural Networks не предполагат линейност на функционалната форма или независимост между предикторите (Chen & Guestrin, 2016). Те „научават“ сложните взаимодействия от висок ред директно от данните, което им позволява да приближат нелинейните повърхности на отговора, които характеризират вземането на решения от потребителите в реалния свят (Tout, 2023).

В тази рамка ML служи като предсказващ механизъм, преобразуващ исторически данни за експозицията в прогнози за вероятностите за конверсия на индивидуално ниво. Тези прогнози могат да бъдат агрегирани по

канални или последователности, за да се оцени очакваната възвръщаемост. Важно е, че ML осигурява гъвкавост за интегриране на хетерогенни характеристики, включително демографски данни за потребителите, контекст на сесията и тип устройство, като улавя латентни фактори, които влияят върху резултатите от атрибуцията (de Naan, 2022; Shao & Li, 2011b).

Въпреки това, „чистото“ прогностично моделиране страда от класическото ограничение на „черната кутия“: макар и точно, то често не успява да предостави интерпретируеми или каузални прозрения. За да се справи с този проблем, рамката въвежда причинно-следствен слой, основан на контрафактична инференция.

4.4. Каузална инференция в моделирането на атрибуцията

Каузалната инференция³ въвежда теоретична строгост, като премества въпроса за атрибуцията от „Кои канали корелират с конверсията?“ към „Кои канали причиняват конверсията и в каква степен?“ Това разграничение е от жизненоважно значение за вземането на управленски решения: моделите, базирани на корелация, могат да разпределят бюджета към високочестотни експозиции, които са просто свързани с конверсията, а не са каузално ефективни (Lewis, 2022).

Формално, причинно-следствената атрибуция има за цел да оцени средния ефект от въздействието (average treatment effect, ATE) или условния среден ефект от въздействието (CATE) на експозицията на даден канал T върху резултата от конверсията Y :

$$ATE = E[Y(1) - Y(0)],$$

където $Y(1)$ обозначава потенциалния резултат при експозиция, а $Y(0)$ – потенциалния резултат при липса на експозиция.

В контекста на атрибуцията обаче рядко е възможно да се провеждат рандомизирани експерименти. Затова изследователите използват квазиекспериментални и базирани на модели техники за причинно-следствени изводи. Най-широко използваните са:

- Претегляне и съпоставяне на склонността (ан. propensity score matching)⁴, което коригира смесването на фактори чрез балансиране на наблюдаваните ковариати (Athey & Imbens, 2019; Lewis & Wong, 2022);

³ Използваме понятието „инференция“ в смисъл на умозаключение, което се прави въз основа на дадени предпоставки или информация, в логиката или изкуствения интелект.

⁴ Propensity score (склонността) е условната вероятност, даден потребител да получи определено въздействие (treatment), базирана на наблюдавани ковариати: $e(X) = P(T = 1|X)$, където T е индикатор за въздействие, а X е вектор от наблюдавани характеристики.

- Двойно машинно обучение (DML), което използва ML за гъвкаво моделиране както на резултатите, така и на процесите на разпределение на въздействието, като осигурява ортогонализация на грешките в прогнозите (Ahrens et al., 2025), и
- Структурни каузални модели (SCM), които кодират каузални графове, представящи зависимостите между експозициите и резултатите (Pearl, 2019).

В рамките на МТА каузалната инференция може да бъде приложена чрез изчисляване на инкременталния ефект на всеки канал, изразяващ се в неговото пределно въздействие върху вероятността за конверсия при постоянни други експозиции. Възможен подход за реализация е използването на байсиански модел за причинно-следствена атрибуция, който интегрира потенциалните резултати с йерархични априорни разпределения, отразявайки както несигурността, така и хетерогенността в различните кампании. Това позволява на маркетолозите да оценят очаквания инкрементален принос на всеки канал както на индивидуално, така и на агрегирано ниво.

Възможен е и хибриден подход, комбиниращ причинно-следствена инференция с машинно обучение (понякога наричан каузално машинно обучение (CML), чрез който се подобрява както валидността, така и адаптивността на модела (Athey & Imbens, 2019). Например, моделите за градиентно усиляване могат да оценяват ефектите от въздействието непараметрично, като същевременно контролират смущаващите фактори, а обясненията, базирани на SHAP, могат да разкрият приноса на ниво характеристики към тези ефекти. Тази теоретична интеграция формира основата на каузалната предсказуема атрибуция, при която резултатите от модела представляват не само предсказания, но и каузални оценки⁵.

Концептуално, каузалното заключение закрепва рамката за атрибуция в контрафактичната логика: всяка оценена стойност на приноса представлява разликата между действителния резултат и това, което би се случило, ако експозицията не беше налице. Това привежда атрибуцията в съответствие с принципите на експерименталния дизайн и гарантира, че разпределението на бюджета отразява истинското създаване на добавена стойност.

4.5. Обясним изкуствен интелект и интерпретируемост на моделите

Нарастващата зависимост от сложни модели за машинно обучение пороци сериозни опасения относно прозрачността, отчетността и интерпретируемостта. В маркетинговата аналитика използването на „черни кутии“ без механизми за интерпретация подкопава доверието на мениджмънта и пречи на спазването на регулаторните изисквания (Christoph Molnar, 2022;

⁵ За детайли (*SHAP Latest Documentation*, v0.49.0, n.d.)
<https://shap.readthedocs.io/en/latest/index.html>, 30.10.2025 г.

Saura et al., 2024). В резултат на това обяснимата изкуствена интелигентност (ХАІ) се превърна в жизненоважен теоретичен и методологичен компонент на съвременните рамки за атрибуция.

Ролята на ХАІ в МТА е двойна. Първо, тя осигурява локална интерпретируемост, като обяснява индивидуалните прогнози чрез идентифициране на признаците (канални или взаимодействия), които оказват най-силно влияние върху вероятността за конверсия. Второ, тя осигурява глобална интерпретируемост, като обобщава взаимоотношенията в рамките на модела между експозициите и резултатите.

Ключовите ХАІ методи, прилагани в атрибуцията, включват:

- LIME (Local Interpretable Model-Agnostic Explanations), който приспособява прост интерпретируем модел локално около всяка прогноза (Ribeiro et al., 2016) (Ribeiro, Singh, & Guestrin, 2016);
- SHAP (SHapley Additive exPlanations), който приписва резултата от модела към всеки признак въз основа на теорията на кооперативните игри (Lundberg & Lee, 2017), и
- Механизми за внимание в невронните мрежи, които визуализират как моделът претегля последователните входни данни (N. Li et al., 2018).

SHAP стойностите (вчастност) предлагат теоретично съгласувано средство за интерпретируемост, което е в съответствие с кооперативните принципи, лежащи в основата на атрибуцията на стойността на Шапли. Всяка SHAP оценка представлява средния пределен принос на дадена характеристика към прогнозата на модела – по същество интерпретируемо разлагане на сложни нелинейни взаимоотношения.

От управленска гледна точка обяснимостта повишава увереността при вземането на решения и спазването на нормативните изисквания. Например, съгласно GDPR и нововъзникващите регулаторни рамки за управление на изкуствения интелект, организациите трябва да докажат, че автоматизираните маркетингови решения са обясними и недискриминационни. По този начин ХАІ преодолява различията между статистическото обучение и управленската отчетност, като гарантира, че високопроизводителните системи за прогнозиране остават прозрачни, подлежащи на одит и са съобразени с етичните стандарти (Saura et al., 2024).

От теоретична гледна точка ХАІ може да се разглежда като епистемологично допълнение към каузалната инференция. Докато каузалните модели обясняват защо настъпват промени, обяснимите модели разясняват как се генерират прогнозите. Интегрирането на тези парадигми гарантира, че системите за прогнозна атрибуция отговарят както на научните, така и на практическите критерии за интерпретируемост, което ги прави подходящи за използване в реалния свят на мениджмънта.

4.6. Обучение чрез подсилване и динамично оптимизиране на маркетинговия бюджет

Значителен напредък в теоретичното развитие на моделирането на атрибуцията е интегрирането на обучението чрез подсилване (ан. Reinforcement learning, RL) като динамична рамка за вземане на решения. За разлика от традиционните модели на атрибуция, които анализират исторически данни ретроспективно, RL моделите активно се учат от текущите взаимодействия, оптимизирайки бъдещите действия, за да максимализират дългосрочната възвръщаемост от маркетинга (Song et al., 2024).

RL концептуализира маркетинговата среда като Марковски процес на вземане на решения, състоящ се от състояния (контексти на клиенти), действия (експозиции на канали или разпределение на бюджети) и награди (конверсии или приходи). Агентът, представляващ маркетинговата система за вземане на решения, научава политика $\pi(a | s)$, която максимализира очакваната кумулативна награда $E(\sum_t \gamma^t R_t)$, където γ обозначава дисконтов фактор (Sutton & Barto, 2020)⁶.

Приложена към атрибуцията, RL предоставя рамка за динамична оптимизация на бюджета, в която системата непрекъснато коригира инвестициите в каналите въз основа на наблюдаваната обратна връзка за ефективността. С течение на времето моделът научава коя последователност от експозиции дава най-висока вероятност за конверсия за всеки сегмент клиенти. Song et al. (2024) демонстрират този принцип, използвайки дълбока рекурентна Q-мрежа с внимание (Deep Recurrent Q-Network with Attention, съкратено DRQN-Attention), което позволява обясними политики за вземане на решения, които визуализират кои минали взаимодействия са повлияли на всеки избор на действие.

Освен това каузалното обучение чрез подсилване (CRL) разширява тази парадигма, като включва каузални ограничения в процеса на обучение (Uehara et al., 2022; Zhang et al., 2020). CRL гарантира, че оптимизацията на политиката на агента се ръководи от каузални зависимости, а не от обикновени корелации, като по този начин предотвратява „хакерството на награди“ и подобрява генерализацията в различни контексти. Това е особено ценно в маркетинга, където фалшивите асоциации (например сезонност или кръстосано покритие на кампании) могат да заблудят чисто емпиричните RL системи.

⁶ Обучението чрез подсилване (RL) е изчислителна рамка за последователно вземане на решения. При него агентът взаимодейства с околната среда през дискретни времеви стъпки $t = 0, 1, 2, \dots$, като на всяка стъпка наблюдава състоянието на средата s_t , избира дадено действие a_t съобразно политиката $\pi(a_t | s_t)$. Околната среда реагира, като предоставя „награда“ R_t и преход към ново състояние s_{t+1} . Целта на агента е да научи оптимална политика π , която максимализира очакваната му кумулативна награда във времето.

Като цяло RL поставя атрибуцията в по-широката област на предписващата аналитика. Докато предиктивните и каузалните модели обясняват минали резултати, обучение чрез подсилване въздейства върху тях, затваряйки кръга между измерването и вземането на решения. С развитието на маркетинга към автоматизация и персонализация в реално време комбинацията от каузална инференция и обучение чрез подсилване представлява следващата еволюционна фаза в науката за атрибуцията, позволяваща автономни, самооптимизирани се маркетингови екосистеми (Wedel & Kannan, 2016).

4.7. Интегративен модел за предсказуема каузална атрибуция

Обединявайки коментираните парадигми, в настоящото проучване предлагаме концептуална теоретична рамка за предсказуема причинно-следствена атрибуция (ППСА), обединяваща вероятностни, предсказуеми, причинно-следствени и подсилващи слоеве на обучение в рамките на последователна аналитична йерархия, а именно:

- Вероятностният слой моделира пътуванията на клиентите като Марковски преходи между състояния, като количествено измерва структурните зависимости между каналите.
- Предсказващият слой използва машинно обучение, за да оцени вероятностите за конверсия въз основа на характеристики на експозицията и контекстуални данни.
- Каузалният слой изолира инкременталните ефекти на каналите чрез контрафактична инференция, като гарантира, че атрибуцията отразява истинското каузално въздействие.
- Подсилващият слой операционализира тези прозрения в предписващ контекст, като динамично коригира бюджетите, за да оптимизира бъдещите резултати.

Всеки от описаните слоеве се основава на предходния, като постепенно подобрява интерпретируемостта, точността и управленската полезност. Заедно те възплъщават цялостна парадигма за моделиране на атрибуцията от ново поколение – такава, която надхвърля описателната евристика и обхваща пълния аналитичен цикъл: наблюдение → прогнозиране → обяснение → оптимизация.

5. Методология и работна процедура

5.1. Преглед и обосновка на изследователския план

Емпиричният план на това проучване следва структуриран четириетапен аналитичен процес, който операционализира концептуалните слоеве на

рамката за предсказуема причинно-следствено атрибуция (ППСА), представена в раздел 4. Четирите етапа – подготовка на данни, евристично и вероятностно моделиране, предиктивно моделиране чрез машинно обучение и интерпретационен анализ – представляват последователен, но взаимосвързан методологичен конвейер от описателни към предсказуеми и каузални изводи.

Този план интегрира три допълващи се епистемологични логики:

1. Описателна и вероятностна логика, реализирана чрез евристични модели и модели на Марковска верига за количествено измерване на наблюдаваните взаимодействия между каналите;
2. Предиктивна логика, реализирана чрез алгоритми за машинно обучение, които оценяват вероятността за конверсия въз основа на поведенчески данни на индивидуално ниво; и
3. Каузална и интерпретативна логика, реализирана чрез техники за обяснимост на моделите, които осигуряват прозрачност и позволяват приписване на каузална заслуга както на агрегирано, така и на индивидуално ниво.

Всеки слой изпълнява отделна аналитична функция: евристичните и марковските модели улавят структурата на пътуванията на клиентите, слойът за машинно обучение улавя модели и нелинейни зависимости, а слойът за интерпретируемост предоставя каузални обяснения и управленски прозрения. Тази многостепенна методология е в съответствие с последните методологични препоръки в маркетинговата аналитика, които препоръчват хибридни, обясними и причинно-следствени дизайни (Athey & Imbens, 2019; Loureiro, 2023; Wedel & Kannan, 2016).

5.2. Източници на данни и структура на извадката

За да се демонстрира емпирично прилагането на тази рамка, бяха използвани публично достъпни данни за кликанията, извлечени от примерния набор от данни BigQuery за уеб имплементацията на Google Analytics 360 за електронна търговия (Google Analytics, 2022). Наборът от данни е съставен от записи на реалното поведение на клиенти в Google Merchandise Store⁷ (<https://shop.googlemerchandisestore.com>).

Този набор от данни беше избран целенасочено, тъй като отговаря на ключови методологични критерии за изследвания в областта на атрибуцията:

- Предоставя детайлна информация на ниво потребител, което позволява възстановяване на индивидуалните пътувания на клиентите, вместо да се разчита на обобщени данни;
- Улавя многоканални експозиции, включително органично търсене, платно търсене, препратки, дисплей и директни източници на трафик;

⁷ Хранилището за данни е достъпно тук: <https://bit.ly/3YsaAnV> (15.10.2025 г.)

- Съдържа показатели за конверсия и транзакции, което позволява както класификация (конверсия срещу неконверсия), така и регресионни анализи (прогнозиране на приходите).

Данните обхващат 847 224 сесии от 714 167 уникални клиенти, генериращи 11 282 конверсии. Тези стойности осигуряват достатъчна статистическа сила за многостепенно моделиране и позволяват надеждна кръстосана валидация в етапите на прогностично машинно обучение.

Таблица 1 съдържа извлечение от първите 10 реда на пълния работен набор от данни, съдържащ 847214 записа.

Таблица 1

Извлечение на първите 10 реда от работния набор от данни

Идентификатор на посетителя	Дата	Група канали	Източник на трафик	Общ брой транс-акции	Общ приход от трансакциите
33024620802809684	6.12.2016	Referral	(direct)	1	2115,80
1546626226233851002	2.9.2016	Referral	(direct)	2	2103,69
7284466025557220497	27.3.2017	Direct	(direct)	1	2092,81
7383823269227966775	19.10.2016	Referral	(direct)	1	2023,70
3751433429956400495	7.12.2016	Direct	(direct)	1	2007,60
1410336440849356236	16.11.2016	Referral	(direct)	1	1998,00
8740176326132717435	20.10.2016	Organic Search	google	3	1986,76
0110237584337011595	14.4.2017	Direct	(direct)	1	1960,55
1956307607572137989	21.11.2016	Referral	mail.google.com	1	1578,25
6913633170731491634	5.12.2016	Direct	(direct)	2	1940,13
# i 847,214 more rows	...	...	...	...	...

Източник на данни: BigQuery данни от електронна търговия за Google Analytics 360 (<https://bit.ly/4f3B5WR>)

5.3. Подготовка на данните и инженеринг на признаци

Първата стъпка от процеса традиционно включва операции по извличане, трансформиране и зареждане на първичните данни (ан. ETL, от Extract, Transform, Load) и инженеринг на функции – процеси, които са в основата както на вероятностния, така и на предиктивния етап на концептуалната рамка. Съгласно стандартните процедури за маркетингова аналитика (Wickham et al., 2023), необработените данни бяха извлечени, почистени и агрегирани, за да се конструират каналните пътеки на ниво посетител.

След предварителната обработка наборът от данни включва:

- Осем групи канали;
- 265 източника на трафик и
- 19 уникални комбинации от двойки канал-източник.

Тези комбинации формират многоизмерното пространство на признаците, използвано в етапа на моделиране. Таблица 2 представя избрани описателни статистически показатели, обобщаващи приноса на всяка група канали по честота на трансакциите и приходи.

Логиката на инженеринга на признаците произтича директно от предложената рамка ППСА в раздел 4. Всяка точка на съприкосновение се кодира като бинарна или честотна променлива, което позволява на модела да научи както ефектите от присъствието, така и интензивността на експозициите на каналите. Полученият набор от данни поддържа:

- Вероятностно моделиране чрез Марковски преходи (изисква последователни пътища);
- Предиктивно моделиране чрез контролирано обучение (изисква етикетирани резултати), и
- Интерпретативно моделиране чрез техники за обяснимост (изискващи представяне на ниво признаци).

Таблица 2

Основни показатели за групите канали по източници на трафик

Група канали	Източник на трафик	Честота на контакт	Общ брой трансакции	Общо приходи (\$)	Приходи от трансакция (\$)
Referral	other_source	27276	190	36003,00	189,49
Direct	(direct)	131037	2219	377477,00	170,11
Display	other_source	4935	132	19716,00	149,36
Referral	(direct)	61801	5348	613367,00	114,69
Paid Search	google	11592	248	27876,00	112,40
Organic Search	google	212125	2174	210189,00	96,68
Organic Search	(direct)	137501	1360	117125,00	86,12
Paid Search	(direct)	10913	231	19667,00	85,14
Display	google	432	18	1472,00	81,78
Referral	google	5333	5	380,00	76,00
Affiliates	other_source	15062	9	654,00	72,67
Social	other_source	12639	120	8042,00	67,02
Organic Search	other_source	6426	47	2631,00	55,98
Display	(direct)	73	2	77,30	38,65
Social	youtube	209971	11	355,00	32,27
(Other)	google	33	1	12,00	12,00
(Other)	(direct)	2	0	0,00	0,00
(Other)	other_source	72	0	0,00	0,00
Paid Search	other_source	1	0	0,00	0,00

За да се подобри възможността за генерализиране на резултатите, всички признаци бяха нормализирани, категориите променливи бяха кодирани с двоичен код с фиксирана дължина (one-hot encoding), а секвенциите на каналите бяха стандартизирани. ETL процесът гарантира, че структурата на данните отговаря на изискванията както на стохастичното моделиране (Марковски предположения), така и на машинно обучение (векторно представяне).

5.4. Процедура за моделиране: интегриране на евристична, вероятностна и алгоритмична атрибуция

Използваната аналитична процедура следва йерархична комбинация от методи, състояща се от четири интегрирани етапа на моделиране, съответстващи на описателните, вероятностните, предиктивните и интерпретативните компоненти на теоретичната рамка.

Етап 1: Евристични базови модели

Първият етап внедрява три широко използвани евристични модели за атрибуция:

- Атрибуция на първото съприкосновение (first-touch);
- Атрибуция на последното съприкосновение (last-touch) и
- Линейна атрибуция.

Тези модели предоставят базови референтни точки, спрямо които може да се оцени ефективността на подходите, основани на данни. Въпреки че теоретично са опростени, те остават стандартна практика в бранша и служат като полезни референтни точки за демонстриране на относителните предимства на вероятностните и алгоритмичните методи (Berman, 2018; Shao & Li, 2011b).

Етап 2: Вероятностно моделиране чрез вериги на Марков

За реализиране на вероятностен модел за описание и предсказване на приписването предлагаме използването на верижен модел на Марков от k -ти ред за оценяване на вероятностите за преход между каналите. Този модел „улавя“ зависимостите от пътя, простиращи се три точки назад – достатъчно, за да представи значима поведенческа непрекъснатост без прекомерна изчислителна тежест (Anderl, Becker, et al., 2016a; Kakalejčík et al., 2018). Тук изхождаме от хипотезата, че матрицата с преходните вероятности на Марков количествено измерва условната вероятност за преминаване от един канал към друг. На базата на тази матрица могат да се изчисляват ефектите от премахването, за да се оцени относителният принос на всеки канал към общите конверсии.

Този подход към вероятностно моделиране съответства на вероятностния слой на рамката ППСА, превръщайки структурните зависимости в интерпретируеми показатели за влиянието на каналите.

Етап 3: Предиктивно моделиране чрез Extreme Gradient Boosting (XGBoost)

Третият етап операционализира прогностичния слой на рамката чрез машинно обучение. Алгоритъмът XGBoost (Chen & Guestrin, 2016) беше избран и се препоръчва поради високата си производителност при многомерни, разреждени данни, типични за потоците от кликания.

На този етап целевата променлива е конверсията (бинарна: 1 = конверсия, 0 = без конверсия). Всеки входен признак представлява характеристики на експозицията на ниво канал, включително честота, последователност и взаимодействия между каналите. Инженерингът на признаците следва методологията на Tout (2023), което позволява на модела да научи сложни взаимоотношения между променливите.

За обучението на модела и настройката на хиперпараметрите може да се използва пакетът `parisip` за R (Kuhn & Voughan, 2024). Чрез кръстосана валидация бе осигурена генерализуемост, а като показатели за оценка на предиктивната способност на модела са използвани общата точност (overall accuracy), ROC-AUC и log-loss.

Както ще видим по-късно, моделът XGBoost постигна обща точност = 0,984 и AUC = 0,846, което потвърждава неговата надеждност при прогнозиране. Важно е да се отбележи, че включването на инструменти за интерпретируемост (виж етап 4) свързва прогнозните и причинно-следствените измерения на атрибуцията, превръщайки чисто алгоритмичния модел в прозрачен инструмент за подпомагане вземането на решения.

Етап 4: Обяснимост и причинно-следствена интерпретация

Четвъртият етап е свързан с използване на обяснима изкуствена интелигентност (XAI) за интерпретиране на предиктивния модел и свързването му с причинно-следствената логика на рамката. Използвайки пакета `lime` (Hvitfeldt et al., 2022) и SHAP анализ на стойностите, ние количествено измерваме локалното и глобалното влияние на всеки канал върху вероятността за конверсия.

На локално ниво LIME обяснява индивидуалните прогнози, като идентифицира кои конкретни точки на съприкосновение са оказали най-голямо влияние върху конверсията за всеки потребител. На глобално ниво SHAP разлага общите резултати от модела на принос на ниво канал, като ефективно превръща резултатите от машинно обучение в интерпретируеми каузални оценки.

Тази интеграция на прогностичното моделиране с причинно-следствената интерпретация операционализира каузалния и интерпретативния слой на ППСА. Подходът отговаря както на методологичната строгост, така и на управленската интерпретируемост, като свързва ефективността на машинно обучение с практически приложими прозрения.

5.5. Емпирично тестване на процедурата за моделиране

Пълният аналитичен работен процес е илюстриран на Фигура 1.

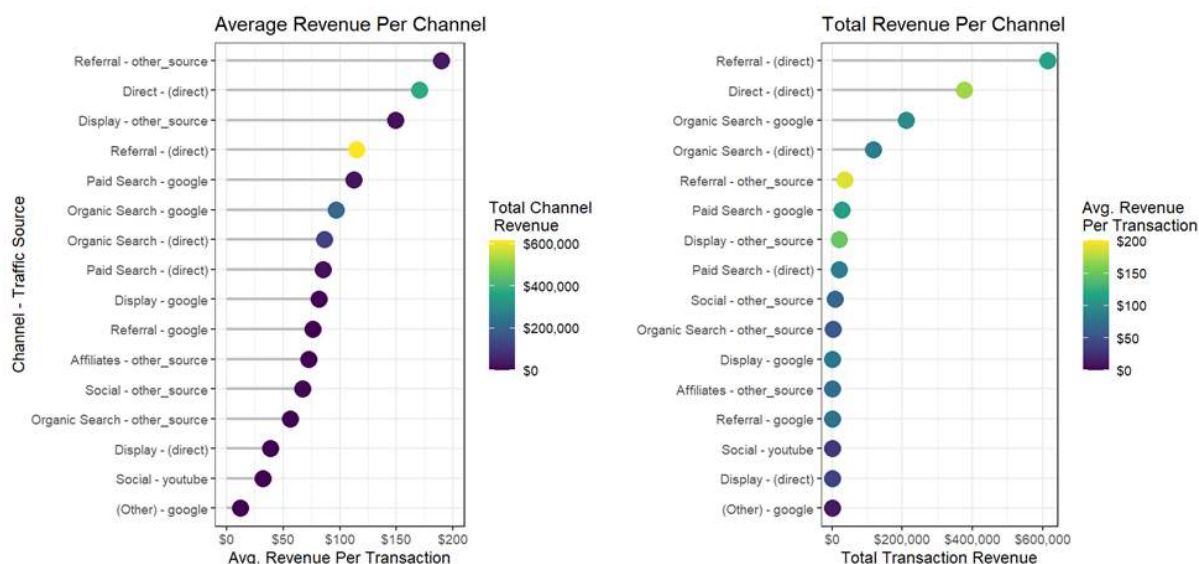


Фигура 1. Схематичен работен процес за атрибутивно моделиране

Процедурата се състои от четири стъпки. Тъй като първата е чисто техническа (вж. пълния програмен код в Приложение 1), тук ще коментираме резултатите от провеждането следващите три в сравнителен план.

5.5.1. Проучвателен анализ и визуализация на данните

Важна предварителна стъпка от работната процедура е проучването на структурата на набора от данни чрез анализ на честотата и приходите. Фигура 2 визуализира разпределението на средните и общите приходи по канали.



Фигура 2. Средни приходи на транзакция (вляво) и общи приходи от транзакции по канали (вдясно)

Както може да се наблюдава от Фигура 2, каналите за препращане (referral) за директно и органично търсене заедно генерират по-голямата част

от приходите от продажби. Тези описателни статистически данни обаче са недостатъчни за извеждане на заключения за причинно-следствената връзка. Следващите стъпки от моделирането решават този проблем, като количествено измерват инкременталния ефект на всеки канал чрез вероятностна и прогнозна оценка.

След това индивидуалните пътувания на клиентите се реконструират в последователни пътеки (вж. Таблица 3). Тези пътеки формират основния набор от данни за моделиране на Марковски вериги и XGBoost, което позволява оценка както на вероятностите за конверсия на ниво пътека, така и на ниво индивид.

Таблица 3

Извличане на първите 10 реда от работния набор от данни с поредици от канали

ID на клиент	Път на клиента	Общо конвер-сии	Без конвер-сии	С-т на конвер-сиите (\$)	Дължина (n) на пътя
1957...	Organic Search-google > Organic Search-google > Organic...	14	132	4256,00	146
3608...	Paid Search-google > Paid Search-google > Paid Search-g...	1	130	27,50	131
0720...	Social-other_source > Social-other_source > Social-othe...	0	119	0,00	119
0824...	Organic Search-google > Organic Search-google > Organic...	0	117	0,00	117
4038...	Organic Search-google > Organic Search-google > Paid Se...	0	110	0,00	110
1856...	Organic Search-google > Organic Search-google > Organic...	0	104	0,00	104
3269...	Direct-(direct) > Direct-(direct) > Direct-(direct) > D...	1	87	189,00	88
6018...	Direct-(direct) > Direct-(direct) > Direct-(direct) > D...	0	83	0,00	83
9801...	Organic Search-google > Organic Search-google > Organic...	0	83	0,00	83
3694...	Direct-(direct) > Organic Search-google > Organic Searc...	1	73	21,20	74
# i 714,157 more rows					

Тези пътеки формират основния набор от данни за моделиране на марковски и XGBoost модели, което позволява оценка на вероятностите за конверсия както на ниво пътека, така и на индивидуално ниво.

5.5.2. Сравнение на евристичните модели за атрибуция с Марковски модел

Като начало изпълнихме серия от евристични модели, включително последно съприкосновение, първо съприкосновение и линейно съприкосновение, които се базират на пътеките на организираните канали, илюстрирани в Таблица 3. Успоредно с тези три евристични модели, за сравнение беше използван Марков модел от трети ред за оценка на вероятностната атрибуция. По същество моделът на Марков е вероятностен модел, който представя зависимостите между точките на контакт във времето. Това е инструмент, който ни позволява да оценим вероятността, даден път да доведе до покупка. За обработката на данните бе използван пакетът `ChannelAttribution` (Altomare & Loris, 2023). Резултатите изчисленията са показани в Таблица 4.

Таблица 4

Обобщение на стойностите на процентното разпределение на заслугите на отделните канали, оценени от различните модели

Наименование на канала	Markov	First touch	Last touch	Linear model
Referral-(direct)	35,9%	40,0%	43,8%	42,3%
Organic Search-google	19,3%	17,2%	18,9%	18,2%
Direct-(direct)	18,5%	23,2%	16,5%	19,1%
Organic Search-(direct)	13,3%	11,1%	11,9%	11,6%
Paid Search-google	3,4%	2,1%	2,5%	2,3%
Paid Search-(direct)	2,9%	2,1%	2,1%	2,1%
Referral-other_source	2,6%	1,5%	1,6%	1,6%
Display-other_source	2,3%	1,2%	1,3%	1,3%
Social-other_source	0,9%	0,7%	0,7%	0,7%
Organic Search-other_source	0,3%	0,4%	0,4%	0,4%
Display-google	0,2%	0,2%	0,2%	0,2%
Social-youtube	0,1%	0,1%	0,1%	0,1%
Affiliates-other_source	0,1%	0,1%	0,1%	0,1%
Referral-google	0,1%	0,1%	0,0%	0,0%
Display-(direct)	0,0%	0,0%	0,0%	0,0%
(Other)-google	0,0%	0,0%	0,0%	0,0%
(Other)-other_source	0,0%	0,0%	0,0%	0,0%
Paid Search-other_source	0,0%	0,0%	0,0%	0,0%
(Other)-(direct)	0,0%	0,0%	0,0%	0,0%

Таблица 4 съдържа относителните дялове на приносите на отделните канали за наблюдаване на конверсия, получени от четирите сравнявани модели. Виждаме, че за някои канали се наблюдава съществена разлика в дела на конверсиите, изведени с трите тествани евристични модела. Например, при директния трафик, моделът „първо съприкосновение (first touch)“ показва значително по-високи проценти на конверсия в сравнение с другите два модела, докато при препращанията е обратното. В канала на препращанията (Referral) най-високият процент на конверсии се получава при правилото за последно докосване, докато в директния канал същият модел дава най-ниския процент на конверсии. В други канали обаче има почти равномерно разпределение на резултатите между трите модела (например органично търсене (директно)). Въпреки това, анализът на резултатите, получени от модела на Марков, показва, че частта от заслугите, която е статистически значима по отношение на недоминиращите канали, включваща органично търсене, платно търсене и дисплейни реклами, е значително по-голяма. От друга страна, приносът на доминиращия канал на препращане се оказва значително подценен.

Резултатите, представени в Таблица 4, подсказват, че евристичните модели прекомерно приписват заслуги на доминиращите канали (например препратки и директен трафик), докато моделът на Марков преразпределя

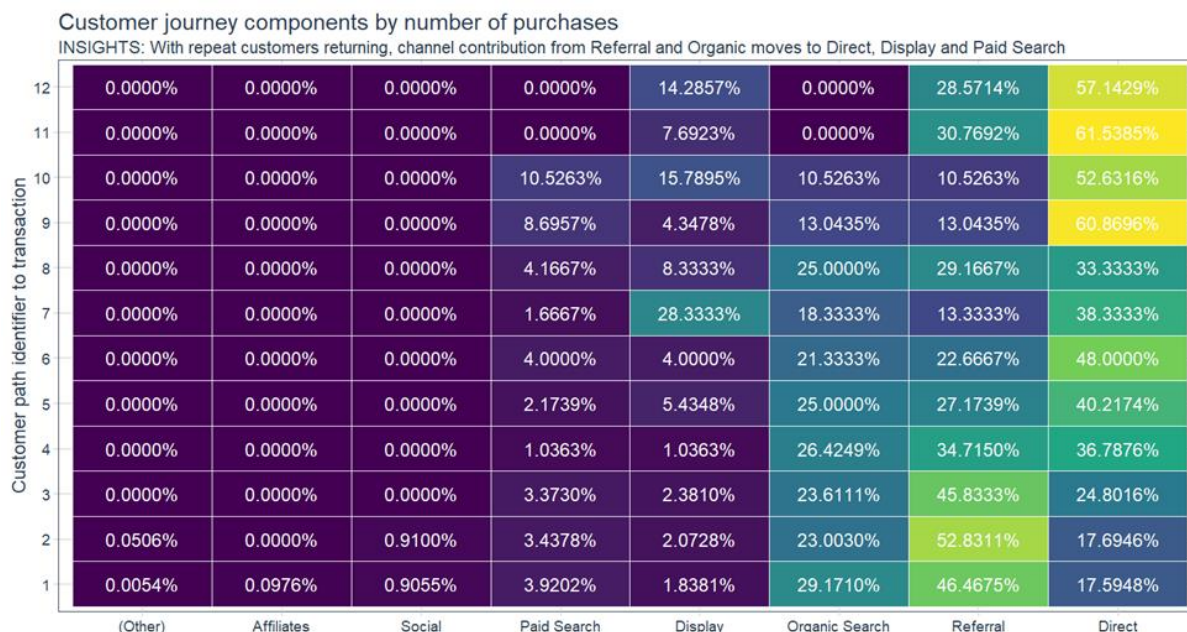
заслугите към спомагателни канали, като органично търсене и дисплей. Това емпирично потвърждава теоретичното твърдение, че вероятностните модели по-добре пресъздават зависимостите между каналите и избягват систематичните отклонения и субективизма на приписването, присъщо на евристиките, базирани на правила.

Получените резултати са в голяма степен сравними с тези, получени чрез тривиалния проучвателен анализ, илюстриран на Фигура 2.

Явното предимство на този подход е, че маркетинговият бюджет може да бъде разпределен между различните комуникационни канали, които генерират трафик към сайта на рекламодателя, като дяловете се използват директно за тази цел. Важно е да се отбележи, че последователността на каналите е изградена въз основа на цялостния път на клиента, който не винаги е линеен и може да съдържа множество конверсии и цикли. За да се отчете тази „скрита“ информация при анализа на поведението на клиентите, би било подходящо да се фрагментират индивидуалните пътеки в „сегмен“, завършващи с една конверсия. Предимството на подхода е, че ни позволява да защитим хипотезата, че индивидуалното поведение на посетителите може да се променя с времето и че следователно на всеки етап от продажбения процес може да доминира различен модел на атрибуция. Това налага необходимостта от допълнителни интензивни групови изчисления на етапа на подготовка на данните. Предимството обаче е възможността да се използват алгоритми за машинно обучение при прогнозиране на вероятността за конверсия на индивидуално ниво, като се използва трансакционен флаг 0/1 като целева променлива.

5.5.3. Моделиране на атрибуцията чрез класификационен XGBoost

Използваме пакета ``dplyr`` в R (Wickham et al., 2023), за да разделим пътя на клиента на отделни етапи (частични пътеки), завършващи с конверсия (маркирани с флаг 1). Това ни помага да проучим влиянието на конкретни канали (точки на контакт) върху повторните покупки и да проследим промените в поведението на купувачите по време на пътя на клиента. Полезен инструмент за визуализация при интерпретирането на резултатите е топлинната карта, илюстрирана на Фигура 3. Картата ясно показва, че с всяко повторно връщане на клиентите се наблюдава забележима промяна в каналите от препратки и органични резултати към директни, дисплейни и платени търсения. Така сравнително лесно може да се проследява динамиката на еволюцията на пътеките в съответствие с марковските предположения, обсъдени в раздел 4.2.



Фигура 3. Топлинна карта на пътуването на клиента

Тази визуализация потвърждава вероятностната теория, че поведението на клиентите показва стохастични преходи между каналите, което подкрепя уместността на модела на Марков за представяне на такива процеси.

Топлинната карта предлага полезна основа, върху която да се разгледа пътят на клиента на обобщено ниво. Въпреки това, ограничението на този подход е, че той не взема предвид вероятността за конверсия, а се фокусира върху броя на успешните конверсии. Наличността на неагрегирани данни за пътя на клиента, при които всеки идентификатор на пътя завършва с флаг, показващ конверсия или липса на конверсия, предоставя обаче допълнителна възможност за обучение на предиктивния класификационен модел. В нашия случай ние извършихме някои инженерни промени чрез дихотомизация на оригиналните променливи.

Целта на тази предварителна процедура е да се обучи предсказващ класификационен модел, използвайки подготвените данни. Моделът трябва да може да предскаже с висока степен на достоверност вероятността за конверсия на индивидуално ниво, използвайки точки на допир от различни канали като предиктори. За да постигнем това, използваме един от най-разпространените алгоритми за класификация в машинно обучение, а именно Extreme Gradient Boosting (Chen & Guestrin, 2016). За обучението на модела прилагаме пакета parsnip за R (Kuhn & Vaughan, 2024). След няколко процедури за настройка на хиперпараметрите на модела успяхме да обучим модел с приемлива прогнозна ефективност (обща точност = 0,984, AUC на ROC кривата = 0,846).

Таблица 5 съдържа извлечение от първите няколко реда на набора от данни с прогнозирана вероятност за конверсия (.pred_0, .pred_1). Например,

моделът оценява, че вторият път на посетителите има 93% вероятност за конверсия. Всъщност това е правилно прогнозиран случай (флаг за транзакция = 1). За разлика от това, пътят на първите посетители е неправилно предсказан от модела (.pred_1 = 0,938, флаг за транзакция = 0).

Таблица 5

Прогнозни резултати и предиктори в класификационния модел

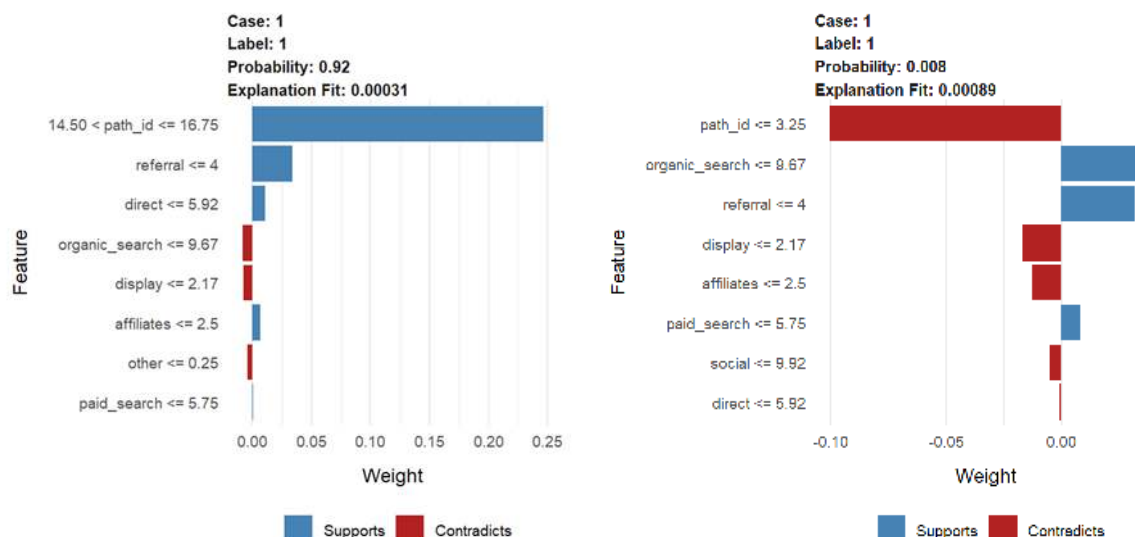
Visitor_path id	.pred_0	.pred_1	Transaction flag	Path id	Social	Affiliates	Direct	Display	Organic search	Paid search	Referral	Other
...100_15	0.0624	0.938	0	15	0	0	0	0	1	0	0	0
...100_12	0.0704	0.930	0	12	0	0	0	0	1	0	0	0
...100_13	0.0719	0.928	1	13	0	0	0	0	1	0	0	0
...100_10	0.0756	0.924	0	10	1	0	0	0	0	0	0	0
...158_12	0.0809	0.919	0	12	0	0	0	0	0	1	0	0
...932_15	0.0813	0.919	0	15	0	0	0	0	1	0	0	0
...386_13	0.0813	0.919	0	13	0	1	0	0	0	0	0	0
...386_16	0.0813	0.919	0	16	0	0	0	0	1	0	0	0
...386_17	0.0813	0.919	0	17	0	0	1	0	0	0	0	0
...386_25	0.0813	0.919	0	25	0	0	0	0	1	0	0	0

i 719,103 more rows

5.5.4. Каузална инференция на индивидуално ниво

За да предоставим прозрачно и разбираемо обяснение на решенията, генерирани от XGBoost, ние използвахме и обясним алгоритъм за машинно обучение (Ribeiro et al., 2016). Това ни даде възможност да обясним прогнозите на класификаторите от типа „черна кутия“. За реализацията му използвахме R порт на Python пакета `lime` (Hvitfeldt et al., 2022). Алгоритъмът позволява оценка на приноса на всеки от разглежданите канали на индивидуално ниво на клиента чрез LIME и SHAP методологии (Lundberg & Lee, 2017; *SHAP Latest Documentation*, v0.49.0, n.d.).

Два случая, избрани на случаен принцип, са представени на Фигура 4. В лявата половина на фигурата е показан клиент, чиято вероятност за конверсия е висока (~ 91,9%), както е посочено в лявата част. Дясната част показва случая на друг клиент, чиято вероятност за конверсия е много ниска (~0,8%). Ключовите предиктори, които са оказали най-силно влияние върху тази вероятност, са, че потребителят е използвал пътеки 15 и 16 и е взаимодействал с дисплейни реклами между 8 и 10 пъти. Обстоятелствата, показани в синьо на графиката, също имат положително влияние върху вероятността за конверсия. Неизползването на социалните медии като точка на контакт (по-малко от 9 пъти) обаче има отрицателно влияние върху вероятността за конверсия.



Фигура 4. Два случайно избрани клиента с висока (вляво) и ниска (вдясно) вероятност за конверсия

В случая с втория случайно избран потребител основният фактор, който оказва негативно влияние върху прогнозираната вероятност за конверсия, беше използването на пътеки 1, 2 и 3. От друга страна, използването на органично търсене по-малко от девет пъти и препратки по-малко от четири пъти оказва благоприятно влияние върху вероятността за покупка.

Локалните вектори на обяснение, генерирани от SHAP, показват, че органичното търсене, директното и платеното търсене имат най-висок положителен маргинален принос към вероятността за конверсия, което е в съответствие с глобалната важност на характеристиките, изведена от модела. Тези констатации емпирично потвърждават теоретичното очакване, че подпомагащите и директните канали играят различни, но допълващи се роли във формирането на конверсията (Anderl, Becker, et al., 2016a; Berman, 2018).

От методологична гледна точка тази интеграция на прогнозно моделиране и обяснимост представлява практическа реализация на каузалния атрибутивен слой в рамките на ППАС: прогнозираните вероятности служат като условни очаквания, докато SHAP стойностите приближават пределните ефекти от третирането.

Този анализ, макар и изключително интензивен от гледна точка на изчислителните усилия, може да бъде от голяма полза при вземането на решения за персонализирано и/или групово таргетиране. При планирането на онлайн маркетингови кампании с персонализирано таргетиране списъкът с прогнозираните вероятности за конверсия може да се използва на индивидуално ниво.

Емпирично тестваните процедури и коментираните резултати съответстват на четвъртата стъпка от процеса, илюстриран на Фигура 1, който съдържа описаните слоеве в теоретичната ППСА рамка. Методологичната

реализация и аналитичните цели, преследвани в тези слоеве, могат да се обобщят, както следва (Вж. Таблица 6):

Таблица 6

Методологичната реализация на ППСА модела

Слой от ППСА	Методология	Аналитични цели
Вероятностен слой	Модел, основан на марковски вериги	Количествено оценяване на вероятностите за преминаване през канала и ефектите от премахването
Предиктивен слой	XGBoost класификатор	Оценка на вероятността за конверсия за всеки възможен път на клиента
Каузален слой	LIME/SHAP анализ	Идентифициране на пределните приноси и причинно-следствената връзка
Подсилващ слой (разширение)*	Обучение по политики (проект)*	Динамично оптимизиране на разпределението на бюджета (проект)*

* Четвъртия слой е насока за бъдещо разширение

Тази интеграция гарантира, че емпиричният анализ е основан на теорията, като същевременно се запазват възпроизводимостта и изчислителната ефективност.

5.6. Валидност на методологията и съображения за надеждност на резултатите

Стабилността на предложената аналитична рамка за моделиране беше оценена по три измерения:

1. Прогностичната валидност, оценена чрез кръстосана валидация и тестове върху независима извадка, потвърди, че представянето на модела остава стабилно при случайни извадки.
2. Интерпретативната устойчивост, валидирана чрез сравнение на резултатите от LIME и SHAP, показва, че локалните и глобалните обяснения на предикторите са съгласувани.
3. Вероятностната съгласуваност, потвърдена чрез сравнение на оценките, базирани на Марковски преход, с важността на каналите, изведена от модела, показва, че разпределенията на атрибуциите са съответстващи по посока при всички методи.

Като цяло тези проверки за валидиране потвърждават методологичната обосновааност на предложения подход. Освен това отвореният код на всички компоненти (R-базиран работен поток) е в съответствие със стандартите за прозрачност и възпроизводимост, подчертавани в съвременните маркетингови аналитични изследвания (Wedel & Kannan, 2016).

Всичко това ни дава основание да вярваме, че представената тук методологична рамка развива изследванията в областта на мултитъч атрибуцията по три различни начина:

- Интегриране на теоретични и емпирични слоеве – свързване на вероятностни, предиктивни и причинно-следствени компоненти в единен работен процес;
- Операционализиране на обяснима изкуствена интелигентност в маркетинговия анализ – повишаване на прозрачността на резултатите от моделите; и
- Основа за динамична, базирана на подсилване оптимизация – позициониране на атрибуцията като част от по-широка адаптивна система за маркетингови решения.

Този дизайн не само свързва теоретичните и изчислителните перспективи, но и създава план за бъдещо емпирично възпроизвеждане в различни платформи и индустрии.

6. Резултати и дискусия

6.1. Преглед на емпиричните резултати

Емпиричните резултати, представени в предходната секция, предоставят оперативна валидация на рамката за предсказуема причинно-следствена атрибуция (ППСА). Сравнителната ефективност на евристичните, вероятностните и базирани на машинно обучение модели илюстрира как точността и интерпретируемостта на атрибуцията се развиват в различните аналитични парадигми.

В съответствие с теоретичната логика на предложената концептуална рамка ППСА резултатите показват, че:

1. Евристичните модели преувеличават значението на доминиращите канали и не успяват да уловят зависимостите между каналите;
2. Вероятностните модели (вериги на Марков) подобряват разпределението на заслугите чрез моделиране динамиката на прехода и ефектите от премахването; и
3. Моделите за машинно обучение (XGBoost) постигат по-висока точност на прогнозите и по-детайлна диференциация на каналите, особено когато се комбинират с обясними методи на изкуствен интелект, които свързват прогнозите с причинно-следственото тълкуване.

Това йерархично подобрение отразява прехода от описателна към предсказуема и каузална атрибуция, концептуализирана в раздел 4.7, където всеки слой добавя ново измерение на аналитична дълбочина, преминавайки от разпознаване на модели към причинно-следствено разбиране и в крайна сметка към предписваща оптимизация.

6.2. Интерпретация на резултатите от евристичната и вероятностната атрибуция

На Таблица 4 бяха представени процентните разпределения на заслугите, получени от четирите сравнявани модела: верига на Марков, първи контакт, последен контакт и линейна атрибуция. Резултатите потвърждават отдавнашните критики към евристичните подходи (Berman, 2018; Shao & Li, 2011b). Моделът на последния контакт, например, приписва непропорционално голяма заслуга на канала за препращане (43,8%), който обикновено представлява крайното трансакционно взаимодействие, а не инициатора на потребителското намерение. От друга страна, атрибуцията на първия контакт преувеличава значението на директния трафик (23,2%), което отразява пристрастие към първоначалния контакт. Линейният модел, който разпределя заслугите равномерно между всички точки на съприкосновение, избягва крайностите, но все пак не отчита последователните зависимости и ефектите на пренасяне (Anderl, Becker, et al., 2016b; Naik et al., 1998).

За разлика от това, моделът на Марков от трети ред дава по-балансирано разпределение, като присвоява статистически значими, но пропорционално намалени заслуги на препращащите и директните канали, докато повишава приноса на спомагателните канали, като органичното търсене и платеното търсене. Това преразпределение показва, че конверсията не е резултат от изолирани действия на каналите, а от последователни синергии – резултат, който е в съответствие с хипотезата за взаимозависимост, предложена в литературата за многоканален маркетинг (Bucklin & Sismeiro, 2009b; de Naan et al., 2016a).

Тези констатации потвърждават емпирично вероятностния слой на рамката ППСА, която моделира пътуванията на потребителите като стохастични преходи между каналите. Марковската структура улавя динамичното взаимодействие между точките на допир, приближавайки временните зависимости, които определят вероятността за конверсия. Това съответства на твърдението на Anderl et al. (2016a), че графичните модели за атрибуция предоставят по-реалистично представяне на поведението на клиента по време на пътуването му в сравнение с всякакви статични евристични модели.

От управленска гледна точка това означава, че прекомерното разчитане на атрибуцията на последното докосване може систематично да наруши разпределението на маркетинговите ресурси – прекомерно финансиране на каналите за сключване на сделки (например препоръки), докато недостатъчно се инвестира в каналите за повишаване на осведомеността или генериране на намерение (например дисплей, органично търсене). Чрез количествено измерване на вероятностите за преход и ефектите от премахването вероятностната атрибуция осигурява по-справедлива и емпирично обоснована основа за оптимизиране на бюджета.

6.3. Интерпретация на резултати от предиктивното моделиране

Резултатите от предиктивното моделиране с класификатора Extreme Gradient Boosting (XGBoost) бележат статистически значим прогрес в сравнение с евристичните и вероятностните подходи. Както е посочено в раздел 5.7, моделът XGBoost постигна обща точност от 0,984 и AUC (площ под ROC кривата) от 0,846, демонстрирайки изключителна предиктивна способност за класифициране на пътеки към конверсия и неконверсия.

Тези показатели за ефективност са в съответствие с констатациите на Tout (2023) и Berman (2018), които съобщават, че методите за градиентно усилване и ансамблови методи превъзхождат както логистичната регресия, така и вероятностните модели при многомерни атрибутивни данни. Повишената точност на прогнозите отразява способността на модела да улавя нелинейните взаимодействия между каналите и контекстуалните променливи – явления, които не могат да бъдат представени от адитивни или ограничени по отношение на паметта модели, като например веригите на Марков.

Фигура 4, която визуализира два произволно избрани случая на клиенти – единият с висока (~91,9%) и другият с ниска (~0,8%) прогнозна вероятност за конверсия – илюстрира интерпретативния потенциал на комбинирането на машинно обучение с методи на обяснима изкуствена интелигентност (XAI).

В случая с висока вероятност моделът идентифицира честите органични търсения и препратки като силни положителни предиктори в съответствие с тяхната подкрепяща и пряка роля в процеса на покупка. От друга страна, пътеките с ниска вероятност за конверсия са свързани с рядко или непоследователно използване на ключови канали. Използването на методите за интерпретируемост SHAP и LIME изяснява тези взаимоотношения, разлагайки сложните резултати от модела на ефекти на ниво канал.

Тези резултати операционализират предиктивния и каузалния слой на рамката PCAF. В предиктивно отношение моделът точно прогнозира вероятността за конверсия; в каузално отношение методите за атрибуция на признаците приближават пределния ефект на третирането на всеки канал. Чрез комбинирането на тези две инферентни перспективи проучването преодолява разминаването между точност и интерпретируемост – ключово предизвикателство в алгоритмичната маркетингова аналитика (Athey & Imbens, 2019; Christoph Molnar, 2022).

Освен това, интегрирането на обяснимостта превръща модела XGBoost от „черна кутия” в инструмент, който може да се използва за управление, като предоставя атрибуции на ниво канал, които отразяват причинно-следственото мислене. Например маркетингологът може да интерпретира положителна SHAP стойност за органично търсене като доказателство за инкрементален принос, а не просто като статистическа корелация. Това отразява теоретичния преход от предсказуема асоциация към причинно-

следствена атрибуция, както е дискутирано и от Loureiro et al. (2023) и Lewis (2022).

6.4. Сравнителен анализ: От евристична пристрастност към каузална интерпретируемост

Резултатите от различните етапи на моделиране ясно показват прогресивно намаляване на пристрастието при атрибуцията с увеличаване сложността на анализа – феномен, предсказан от теоретичната йерархия на ППСА модела, предложен в раздел 4.7. В Таблица 7 (която е концептуална и не замества съществуващите емпирични таблици) е обобщена емпирично наблюдаваната методологична прогресия. Всеки слой коригира ограниченията на предходния, като резултатът е модел, който не само предсказва точно, но и съгласува тежестта на атрибутите с причинно-следствените механизми.

Таблица 7

Концептуално обобщение на емпирично наблюдаваната методологична прогресия

Тип модел	Аналитич на основа	Логика на приписване	Склонност към систематични отклонения
Евристичен (Fisrt-touch, Last-touch, Linear)	Базиран на правила	С една точка или с еднаква тежест	Високо структурно отклонение/пристрастност
Вероятностен (Вериги на Марков)	Случаен преход	Зависим от последователността	Намалена пристрастност; липсва контекст
Предиктивен (XGBoost)	Машинно обучение	Нелинейни взаимодействия между признаците	Минимално; липсва причинно-следствена връзка
Каузален/интерпретативен (SHAP/LIME)	Теория на игрите	Контрафактуални оценки на въздействието/приносителите	Каузално последователен и интерпретируем

Емпирично тази йерархия се потвърждава от преразпределението на заслугите в модела на Марков (Таблица 4) и от резултатите за локална/глобална интерпретируемост в модела XGBoost (Таблица 5).

Когато SHAP стойностите се агрегират в целия набор от данни, те показват, че органичното търсене, директният трафик и платеното търсене са най-влиятелните предиктори за конверсия, което отразява както вероятностните, така и поведенческите констатации. Съответствието между ефектите на Марковското премахване и важността на признаците, базирани на SHAP, демонстрира емпиричната съгласуваност на рамката: и двата подхода конвергират към един и същ основен причинно-следствен модел, въпреки различаващата се аналитична логика.

Тези резултати подкрепят силно аргумента, изказан от Wedel и Kannan (2016), че интегрирането на прогностични и каузални изводи представлява бъдещето на маркетинговата аналитика. Те потвърждават, че атрибуционните модели, основани на машинно обучение, когато са съчетани с инструменти за обясняемост, могат да постигнат както висока прогностична ефективност, така и теоретична интерпретируемост – баланс, който традиционно се счита за недостижим в чисто алгоритмичните системи.

6.5. Дискусия: обвързване на емпиричните резултати с теорията и с резултати от предходни проучвания

(а) Съответствие с теоретичните очаквания и с резултатите от предходни изследвания

Емпиричните констатации са в тясна връзка с теоретичните очаквания, формулирани в рамките на ППАС. Преходът от евристично към машинно обучение отразява концептуалната еволюция от описателен към каузален анализ, описана в раздел 4.7. По конкретно:

- Верижният модел на Марков валидира вероятностния слой, потвърждавайки, че последователното моделиране улавя стохастичните зависимости в пътуванията на клиентите.
- Моделът XGBoost, допълнен с обясненията LIME и SHAP, операционализира предсказуемия и каузалния слой, демонстрирайки, че машинно обучение може да приближи както предсказанията, така и каузалните изводи при подходящи ограничения за интерпретируемост.
- Наблюдаваната съгласуваност между вероятностните резултати и резултатите от машинно обучение подкрепя интегративната теза на рамката ППАС – че хибридните модели могат да съчетаят реализма на поведението с предсказуемата прецизност.

Тези резултати са в съзвучие с рапортуваните от Anderl et al. (2016b), които твърдят, че вероятностните модели осигуряват интерпретируемост на поведението, но им липсва предсказуема мащабируемост, както и с тези на Loureiro et al. (2023), които подчертават необходимостта от каузална инференция за управленска надеждност. Чрез синтезиране на тези перспективи емпиричният анализ подкрепя теоретичното твърдение, че съвременното моделиране на атрибуцията трябва да интегрира предсказване, обяснение и каузалност в рамките на единна система.

Сравнителни данни от скорошни проучвания подкрепят надеждността на тези заключения. Например, Tout (2023), сравнявайки моделите XGBoost и Shapley Value в множество набори от данни и стига до заключението, че подходите за ансамблово машинно обучение не само постигат по-висока точност на прогнозите, но и са в тясна връзка с разпределението на кредитите според теорията на игрите. По подобен начин Mrad и Hnich (2024)

демонстрират, че комбинирането на вероятностно и алгоритмично моделиране води до 12–18% подобрение в ефективността на оптимизацията на бюджета в многоканалните кампании.

Нашите емпирични констатации потвърждават тези резултати: комбинацията от XGBoost–SHAP даде интерпретируеми, каузално последователни атрибуции, които преразпределят заслугите към недостатъчно представените канали. Освен това методологичната съгласуваност между вероятностите за преход на Марков и оценките за важност на SHAP емпирично потвърждават принципа на многослойна съгласуваност, постулиран в модела ППСА – че структурните, вероятностните и предсказващите показатели трябва да се сближават, когато моделът улавя истинския основен каузален процес.

(б) Интерпретируемост, обективност и етични аспекти

От етична и управленска гледна точка интегрирането на методи за обяснимост (LIME, SHAP) гарантира спазването на нововъзникващите стандарти за прозрачност на изкуствения интелект в маркетинговата аналитика (Saura et al., 2024). Обяснимата атрибуция намалява риска от алгоритмична непрозрачност, като позволява на одиторите и мениджърите да верифицират, че препоръките за бюджета са основани на данни и са безпристрастни и обективни. Сближаването между обясненията на SHAP и вероятностните атрибуции също така предполага, че интерпретируемите модели могат да постигнат обективност, без да жертват точността – все по-нарастваща загриженост в отговорното изследване на изкуствения интелект (Christoph Molnar, 2022).

(в) Ограничения

Въпреки силната съгласуваност между теорията и резултатите, трябва да се отбележат някои ограничения. Първо, наборът от данни представлява една единствена област (Google Merchandise Store) и може да не е приложим за всички отрасли с различни структури на потребителското пътуване. Второ, моделите на Марков и XGBoost са статични и не се адаптират динамично към променящите се във времето ефекти на кампаниите – ограничение, което бъдещите изследвания биха могли да преодолеят, като използват подсилващо обучение, както е обсъдено в раздел 4.6 (Song et al., 2024). Трето, причинно-следствената валидност остава по-скоро предположение, отколкото експериментално доказателство; бъдещите изследвания трябва да включват рандомизирани дизайни на експозиция или естествени експерименти, за да валидират оценените каузални ефекти (Lewis & Wong, 2022). Четвърто, компонентът на обучението чрез подсилване остава по-скоро концептуален, отколкото емпиричен. Прилагането на архитектури за каузално обучение чрез подсилване (CRL) с потоци от реални данни би осигурило солидно тестване на предписващите възможности на предложената аналитична рамка.

Въпреки това, в рамките на тези граници, емпиричните констатации подкрепят силно теоретичния аргумент, че многослойните, интерпретируеми и каузални кохерентни модели превъзхождат традиционните евристични методи както по отношение на аналитичната строгост, така и по отношение на приложимостта в управлението.

6.6. Изводи и препоръки за управлението на дигиталния маркетинг

Считаме, че практическите последствия от изложените констатации биха били съществени за лицата, вземащи маркетингови решения и стремящи се да оптимизират инвестициите в различни канали, в поне следните направления:

1. Разпределение на бюджета въз основа на данни – чрез количествено измерване на преките и спомагателните ефекти на каналите работният процес, основан на ППСА, подкрепя пропорционални инвестиционни стратегии, които отразяват реалните инкрементални приноси.
2. Прозрачност и отчетност – интерпретируемостта, базирана на SHAR, позволява на мениджърите да обяснят бюджетните решения на заинтересованите страни и да спазват регулаторните стандарти за алгоритмична прозрачност.
3. Динамична адаптивност – макар че в това проучване са приложени статични модели, демонстрираната интеграция на каузален и предиктивен анализ предоставя основа за адаптивни системи за подсилване, способни на оптимизация в реално време.
4. Стратегически анализ, основан на доказателства – резултатите показват, че маркетинговете трябва да преминат от показатели, ориентирани към каналите (например ROI на последния клик), към показатели, ориентирани към пътуването на клиента, основани на вероятностни и каузални изводи.

Като цяло тези импликации позиционират обяснимата атрибуция, базирана на машинно обучение, като крайъгълен камък на измерването на маркетинговата ефективност от следващо поколение.

7. Заключение и насоки за бъдещи изследвания

Това проучване допринася за теоретичното разбиране на мултитъч атрибуцията (MTA), като я позиционира в интегрирана, многослойна концептуална рамка, обединяваща вероятностни, предиктивни и каузални перспективи в една обща аналитична архитектура. Ключовото теоретично прозрение се състои в преосмислянето на атрибуцията не само като проблем на измерването, но и като проблем на поведенческата инференция и оптимизацията на решенията, в който всеки аналитичен слой допринася с отличителна епистемологична стойност.

Претенцията за теоретичен принос на това проучване се състои в неговата синтеза: то изгражда единен модел, който свързва поведенческата логика на потребителските пътувания, статистическата строгост на причинно-следствената инференция, изчислителната адаптивност на машинно обучение и етичната прозрачност, необходима в ерата на алгоритмичния маркетинг.

Важно заключение от достигнатите емпирични констатации се отнася и до ролята на обяснимата изкуствена интелигентност (XAI) в предефинирането на епистемологични граници на маркетинговата аналитика. Исторически погледнато, маркетинговите модели се ценяха заради тяхната интерпретируемост (например регресионни коефициенти, оценки на еластичността), докато съвременните ML модели често жертват прозрачността в името на предсказуемата ефективност. Моделът ППСА показва, че обяснимостта не е непременно противоположна на ефективността, а по-скоро може да бъде нейна интегративна компонента. Бъдещата теоретична работа може да формализира още повече връзката между обяснимостта и доверието, като проучи как прозрачността влияе върху възприеманата легитимност и приемането на AI системи в маркетинговия мениджмънт.

Въз основа на изложените теоретични и емпирични резултати се очертават няколко направления за изследвания, които могат да разширят и усъвършенстват предсказуемо-причинната парадигма на моделирането на атрибуцията:

- Изучаване на каузалните закономерности на атрибуцията чрез методи за дълбоко обучение: Бъдещите изследвания трябва да проучат архитектури за дълбоко машинно обучение, които могат едновременно да изучават представи и причинно-следствени структури на пътуванията на клиентите. Хибридни модели, комбиниращи вариационни автоенкодери, графични невронни мрежи и структурни причинно-следствени модели, биха могли да позволят мащабируема причинно-следствена атрибуция в големи, мултимодални набори от данни.
- Атрибуция между устройства и запазване на поверителността: С ограничаването на бисквитките на трети страни и засиления регулаторен контрол изследователите трябва да разработят модели за атрибуция, които да функционират ефективно при ограничения на поверителността. Диференциалната поверителност, федеративното обучение и генерирането на синтетични данни предлагат обещаващи пътища за постигане както на аналитична прецизност, така и на законово съответствие (Bonawitz et al., 2019; Dwork & Roth, 2014).
- Каузално обучение чрез подсилване (reinforcement learning) и динамично разпределение на бюджета: Интегрирането на каузалната инференция с обучение чрез подсилване (утвърждаване) може да доведе до модели, които непрекъснато изучават инкременталната ефективност на всеки канал в реално време. Бъдещите проучвания трябва да

експериментират с оценка извън политиката и методи за градиент на политиката, адаптирани за каузални ограничения, за да се гарантира стабилна и обяснима оптимизация. В това отношение вече има първоначални опити (Cai et al., 2023; Song et al., 2024).

- **Хибридни каузално-поведенчески модели за атрибуция:** Комбинирането на компютърно моделиране с поведенческа теория може да доведе до по-богати прозрения за вземането на решения от потребителите. Експериментални и симулационни дизайни биха могли да потвърдят дали машинно изчислените каузални атрибуции съответстват на действителните когнитивни и емоционални реакции към маркетинговите стимули.
- **Етичен одит и управление на атрибуционните системи:** Бъдещите емпирични изследвания трябва да проучат прилагането на етични одити за маркетинговите системи, задвижвани от изкуствен интелект, като оценят как механизмите за прозрачност, справедливост и отчетност влияят върху доверието на потребителите и резултатите на фирмите.
- **Интердисциплинарна интеграция:** предсказуемо-каузалната парадигма би следвало да насърчава сътрудничеството между различни дисциплини – икономика, компютърни науки, психология и етика – за изграждането на една наистина холистична наука за маркетинговата атрибуция. Такова интегриране не само би подобрило методологичната строгост, но и би засилило значимостта на областта при справянето с обществените и регулаторните предизвикателства, породени от икономиката, основана на данни.

В заключение, това проучване потвърждава, че бъдещето на мултитъч атрибуцията – и на маркетинговата аналитика като цяло – зависи от интегрирането на три основни принципа: причинност, прозрачност и отговорност. Като основава атрибуцията на поведенчески изводи, предиктивно моделиране и етично управление, проучването установява път към по-интелигентни, интерпретируеми и социално устойчиви маркетингови системи.

По този начин се отговаря на призива, отправен от изследователи и учени в областта на маркетинга (Loureiro, 2023; Saura et al., 2024; Wedel & Kannan, 2016): да се премине от евристиката към единна парадигма, в която данните, теорията и етиката се сливат в приложими знания. Рамката за предсказуема причинно-следствена атрибуция (ППСА) представлява както проявление на тази еволюция, така и основа за следващия етап – модел за маркетингова аналитика, който не само предсказва поведението, но и го обяснява и насочва по отговорен начин.

Приложение 1



Пълният авторов програмен код на R за възпроизвеждане на получените резултати е достъпен тук: https://data.eacademybg.com/MTA_code.R

Използвани източници

- Abhishek, V., Fader, P., & Hosanagar, K. (2012a). Media Exposure through the Funnel: A Model of Multi-Stage Attribution. In SSRN. <https://doi.org/10.2139/ssrn.2158421>
- Abhishek, V., Fader, P., & Hosanagar, K. (2012b). The Long Road to Online Conversion: A Model of Multi-Channel Attribution. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2158421>
- Ahrens, A., Chernozhukov, V., Hansen, C., Kozbur, D., Schaffer, M., & Wiemann, T. (2025). An Introduction to Double/Debiased Machine Learning.
- Altomare, D., & Loris, D. (2023). `ChannelAttribution`: Markov Model for Online Multi-Channel Attribution. CRAN. <https://channelattribution.io>
- Anderl, E., Becker, I., von Wangenheim, F., & Schumann, J. H. (2016a). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, 33(3), 457–474. <https://doi.org/10.1016/j.ijresmar.2016.03.001>
- Anderl, E., Schumann, J. H., & Kunz, W. (2016). Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys. *Journal of Retailing*, 92(2). <https://doi.org/10.1016/j.jretai.2015.10.001>
- Anderl, E., & Weichert, L. (2016). Attribution im Online-Marketing: datenbasierte Bewertung von Customer Journeys. *Marketing Review St. Gallen*, 33(2), 24–31.
- Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, 11(1), 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Avadhanula, V., Colini Baldeschi, R., Leonardi, S., Sankararaman, K. A., & Schrijvers, O. (2021). Stochastic bandits for multi-platform budget optimization in online advertising. *Proceedings of the Web Conference 2021*, 2805–2817. <https://doi.org/10.1145/3442381.3450074>
- Batra, R., & Keller, K. L. (2016). Integrating Marketing Communications: New Findings, New Lessons, and New Ideas. *Journal of Marketing*, 80(6), 122–145. <https://doi.org/10.1509/jm.15.0419>
- Ben Mrad, A., & Hnich, B. (2024). Intelligent attribution modeling for enhanced digital marketing performance. *Intelligent Systems with Applications*, 21. <https://doi.org/10.1016/j.iswa.2024.200337>
- Berman, R. (2018). Beyond the last touch: Attribution in online advertising. *Marketing Science*, 37(5). <https://doi.org/10.1287/mksc.2018.1104>

- Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1). <https://doi.org/10.1214/14-AOAS788>
- Bucklin, R. E., & Sismeiro, C. (2009b). Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing. *Journal of Interactive Marketing*, 23(1), 35–48. <https://doi.org/10.1016/j.intmar.2008.10.004>
- Cai, T., Jiang, J., Zhang, W., Zhou, S., Song, X., Yu, L., Gu, L., Zeng, X., Gu, J., & Zhang, G. (2023). Marketing Budget Allocation with Offline Constrained Deep Reinforcement Learning. *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 186–194. <https://doi.org/10.1145/3539597.3570486>
- Chatterjee, P., Hoffman, D. L., & Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. In *Marketing Science* (Vol. 22, Issue 4). <https://doi.org/10.1287/mksc.22.4.520.24906>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. <https://doi.org/10.1145/2939672.2939785>
- Christoph Molnar. (2022). *Interpretable Machine Learning*, Christoph Molnar. <https://christophm.github.io/interpretable-ml-book/>.
- Court, D., Elzinga, D., Mulder, S., & Vetvik, O. J. (2009). The consumer decision journey. *McKinsey Quarterly*, 3, 97–107.
- Cramer, J. S. (2003a). *Logit Models from Economics and Other Fields*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615412>
- Cramer, J. S. (2003b). The Origins of Logistic Regression. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.360300>
- Dalessandro, B., Stitelman, O., Perlich, C., & Provost, F. (2012). Causally motivated attribution for online advertising. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2351356.2351363>
- Danaher, P. J., & Dagger, T. S. (2013). Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. *Journal of Marketing Research*, 50(4). <https://doi.org/10.1509/jmr.12.0241>
- de Haan, E. (2022). Attribution Modeling. In *Handbook of Market Research* (pp. 1–65). Springer International Publishing. https://doi.org/10.1007/978-3-319-05542-8_39-1
- de Haan, E., Wiesel, T., & Pauwels, K. (2016a). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing*, 33(3), 491–507. <https://doi.org/10.1016/j.ijresmar.2015.12.001>
- Deshmukh, S. V., R, S., Vijay, S., Subramanian, J., & Agarwal, C. (2023). Counterfactual Explanation Policies in RL.
- Dinner, I. M., Heerde Van, H. J., & Neslin, S. A. (2014). Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising. *Journal of Marketing Research*, 51(5), 527–545. <https://doi.org/10.1509/jmr.11.0466>

- Du, R., Zhong, Y., Nair, H., Cui, B., & Shou, R. (2019). Causally Driven Incremental Multi Touch Attribution Using a Recurrent Neural Network.
- Gagniuc, P. A. (2017). Markov Chains. Wiley.
<https://doi.org/10.1002/9781119387596>
- Gaur, J., Bharti, K., & Bajaj, R. (2024). Maximizing marketing impact: heuristic vs ensemble models for attribution modeling. Global Knowledge, Memory and Communication. <https://doi.org/10.1108/GKMC-04-2023-0112>
- Google Analytics. (2022, September 20). BigQuery sample dataset for Google Analytics ecommerce web implementation. <https://bit.ly/3YsaAnV>
- Haughton, D., Haughton, J., & Lo, V. S. Y. (2025). Cause and Effect Business Analytics and Data Science. Chapman and Hall/CRC.
<https://doi.org/10.1201/9780429172588>
- Hvitfeldt, E., Pedersen, T. L., & Benesty, M. (2022). Package ‘lime’: Local Interpretable Model-Agnostic Explanations. CRAN. <https://lime.data-imaginist.com>
- Kakalejčák, L., Bucko, J., & Resende, P. A. A. (2021). Multichannel Marketing Attribution Using Markov Chains for E-Commerce. Statistika, 101(2).
- Kakalejčák, L., Bucko, J., Resende, P. A. A., & Ferencova, M. (2018). Multichannel Marketing Attribution Using Markov Chains. Journal of Applied Management and Investments, 7(1), 49–60.
- Kane, K., Lo, V. S. Y., & Zheng, J. (2014). Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods. Journal of Marketing Analytics, 2(4), 218–238.
<https://doi.org/10.1057/jma.2014.18>
- Kannan, P. K., & Li, H. “Alice.” (2017). Digital marketing: A framework, review and research agenda. International Journal of Research in Marketing, 34(1), 22–45. <https://doi.org/10.1016/j.ijresmar.2016.11.006>
- Kannan, P. K., Reinartz, W., & Verhoef, P. C. (2016). The path to purchase and attribution modeling: Introduction to special section. International Journal of Research in Marketing, 33(3), 449–456.
<https://doi.org/10.1016/j.ijresmar.2016.07.001>
- Kelly, J., Vaver, J., & Koehler, J. (2018). A Causal Framework for Digital Attribution. <https://storage.googleapis.com/gweb-research2023-media/pubtools/4415.pdf>
- Kireyev, P., Pauwels, K., & Gupta, S. (2016). Do display ads influence search? Attribution and dynamics in online advertising. International Journal of Research in Marketing, 33(3).
<https://doi.org/10.1016/j.ijresmar.2015.09.007>
- Kuhn, M., & Voughan, D. (2024). ‘parsnip’: A Common API to Modeling and Analysis Functions. CRAN. <https://parsnip.tidymodels.org>
- Lavidge, R. J., & Steiner, G. A. (1961). A Model for Predictive Measurements of Advertising Effectiveness. Journal of Marketing, 25(6), 59–62.
<https://doi.org/10.1177/002224296102500611>
- Lewis, R., & Wong, J. (2022). Incrementality Bidding and Attribution.

- Li, H. (Alice), & Kannan, P. K. (2014). Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment. *Journal of Marketing Research*, 51(1), 40–56. <https://doi.org/10.1509/jmr.13.0050>
- Li, N., Arava, S. K., Dong, C., Yan, Z., & Pani, A. (2018). Deep Neural Net with Attention for Multi-channel Multi-touch Attribution.
- Loureiro, S. M. C. (2023). Overview of the brand journey and opportunities for future studies. *Italian Journal of Marketing*, 2023(2), 179–206. <https://doi.org/10.1007/s43039-023-00069-0>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017-December.
- Naik, P. A., Mantrala, M. K., & Sawyer, A. G. (1998). Planning media schedules in the presence of dynamic advertising quality. *Marketing Science*, 17(3). <https://doi.org/10.1287/mksc.17.3.214>
- Neslin, S. A., Grewal, D., Leghorn, R., Shankar, V., Teerling, M. L., Thomas, J. S., & Verhoef, P. C. (2006). Challenges and Opportunities in Multichannel Customer Management. *Journal of Service Research*, 9(2), 95–112. <https://doi.org/10.1177/1094670506293559>
- Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54–60. <https://doi.org/10.1145/3241036>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. <https://doi.org/10.1037/h0037350>
- Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, 48(1). <https://doi.org/10.1509/jmkr.48.1.87>
- Saura, J. R., Škare, V., & Dosen, D. O. (2024). Is AI-based digital marketing ethical? Assessing a new data privacy paradox. *Journal of Innovation & Knowledge*, 9(4), 100597. <https://doi.org/10.1016/j.jik.2024.100597>
- Shao, X., & Li, L. (2011b). Data-driven multi-touch attribution models. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 258–264. <https://doi.org/10.1145/2020408.2020453>
- SHAP latest documentation, v0.49.0. (n.d.). Retrieved August 7, 2025, from https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html
- Shapley, L. S. (1953). A Value for n-person Games. *Contributions to the Theory of Games*. In *Contributions to the Theory of Games II*.

- Shapley, L. S. (2009). A value for n -person games . In *The Shapley Value*.
<https://doi.org/10.1017/cbo9780511528446.003>
- Singal, R., Besbes, O., Desir, A., Goyal, V., & Iyengar, G. (2019). Shapley Meets Uniform: An Axiomatic Framework for Attribution in Online Advertising. *The World Wide Web Conference*, 1713–1723.
<https://doi.org/10.1145/3308558.3313731>
- Sinha, R., Arbour, D., & Puli, A. M. (2022). Bayesian Modeling of Marketing Attribution.
- Soboleva, A., Pudovikov, A., Snetkov, R., Babenko, A., Samosvat, E., & Dorn, Y. (2025). Optimizing Online Advertising with Multi-Armed Bandits: Mitigating the Cold Start Problem under Auction Dynamics.
- Song, Y., Wang, W., & Yao, S. (2024). Customer Acquisition Via Explainable Deep Reinforcement Learning. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4802411>
- Strong, E. K. (1925). *Psychology in Selling and Advertising* (1st ed.). McGraw-Hill Book Company, Inc.
- Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. *PLOS ONE*, 15(8), e0235502.
<https://doi.org/10.1371/journal.pone.0235502>
- Sutton, R. S. ., & Barto, A. G. . (2020). *Reinforcement learning : an introduction*. The MIT Press.
- Tout, J. (2023). Comparing the Performance of XGBoost and Shapley Values to Last Touch Attribution [Master Thesis, Erasmus University Rotterdam].
<http://hdl.handle.net/2105/66661>
- Uehara, M., Shi, C., & Kallus, N. (2022). A Review of Off-Policy Evaluation in Reinforcement Learning.
- Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, 113(27), 7310–7315.
<https://doi.org/10.1073/pnas.1510479113>
- Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6).
<https://doi.org/10.1509/jm.15.0413>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). “dplyr”: A Grammar of Data Manipulation. CRAN. <https://dplyr.tidyverse.org>
- Xu, L., Duan, J. A., & Whinston, A. (2014). Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion. *Management Science*, 60(6), 1392–1412. <https://doi.org/10.1287/mnsc.2014.1952>
- Zhang, A., Lyle, C., Sodhani, S., Filos, A., Kwiatkowska, M., Pineau, J., Gal, Y., & Precup, D. (2020). Invariant causal prediction for block MDPs. 37th International Conference on Machine Learning, ICML 2020, Part F168147-15.
- Zhao, K., Mahboobi, S. H., & Bagheri, S. R. (2018). Shapley Value Methods for Attribution Modeling in Online Advertising.