UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Overview of artificial intelligence's use in marketing

Author: Schorochoff Dimitri
Supervisor: Schuiling Isabelle
Academic year 2023-2024
Dissertation for the
Master [60] in management
Daytime schedule

Abstract

Artificial intelligence is playing an increasingly important role in marketing. In this paper, we summarise the literature on the use of AI in marketing. We review some of the most widespread applications of AI to enhance marketing. We provide an intuitive explanation on how the power of AI is leveraged in each of these applications. We also go over some practical information including the most relevant data to collect and the metrics that can be used to evaluate the AI performances. We cover some real-life examples of companies that have used these applications. Finally, we propose a diagram representing all these applications, their sub-applications, and the interaction between them.

Acknowledgements

I would like to thank the Professor Isabelle Schuiling for providing me the opportunity to work on this subject.

I express my gratitude to Frank Schepper for attentively proofreading my work.

I also would like to thank my family for supporting in every circumstances.

Contents

Introduction **1**

1 Review of the main concepts and the literature **3**

1.1 Marketing 3

1.2 Artificial intelligence (AI) 4

1.3 Artificial intelligence in marketing 7

2 AI's fields of use in marketing **8**

2.1 Content creation (49.2%) 10

2.2 Content personalization (52.8%) 12

2.3 Optimizing marketing content and timing (36.6%) 18

2.4 Programmatic advertising and media buying (35%) 18

2.5 Data analytics (predictive analytics (32.9%)) 21

2.6 Customer segmentation (21.5%) 28

2.7 Targeting decisions (31.7%) 30

2.8 Marketing automation (28%) 31

Conclusion **34**

List of Figures

1 Companies' expenses in marketing, figure from [2] 1

1.1 AI branches based on different intelligence definitions 5

2.1 Summary diagram of the fields of use covered in this paper, their sub-fields, and their relationship 9

2.2 Non-exhaustive summary tabular. For each content category: how to use it, how good is a generative AI to create it, and some common GenAI names in the category 11

2.3 Performance evaluation steps of personalization 16

2.4 Programmatic advertising summary scheme 19

Introduction

Key marketing performance indicators have been followed every year since 2008 by the CMO (Chief Marketing Officer) survey, which is carried out by asking questions to executives of companies in the commercial sector in the United States. This survey is sponsored by prestigious organizations such as Deloitte LLP, Duke University's Fuqua School of Business and the American Marketing Association. It also gives us an insight into the importance of marketing in today's world, with marketing expenditure accounting for 10.6% of global company budgets in autumn 2023.

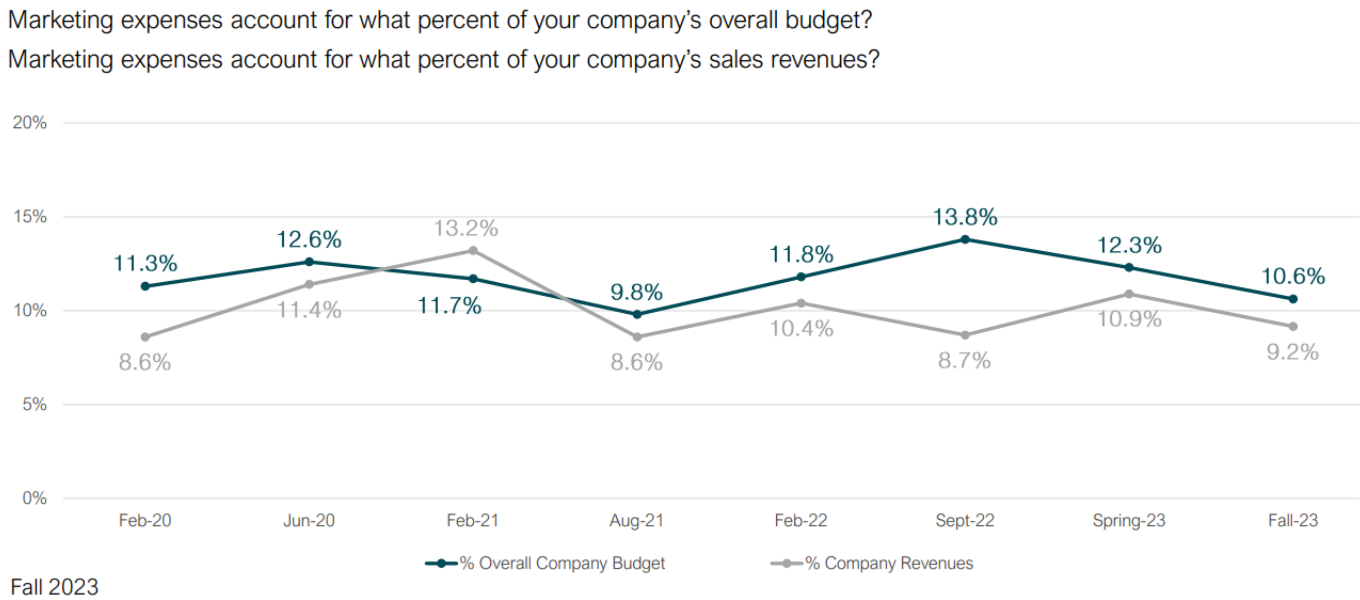

Figure 1: Companies' expenses in marketing, figure from [2]

Computer science plays an increasingly important role in marketing. Between September 2022 and September 2023, corporate spending on digital marketing increased by 7.9%, which is around three times more than the overall 2.6% increase in marketing expenditure. Computer science, and AI in particular, has become crucial to marketing in recent years. Indeed, 94.1% of companies using AI began using it three years ago or less, and 60.4% of companies began using it less than a year ago. In addition, the use of AI was found to be beneficial, with an estimated 6.2% improvement in sales and productivity, a 7% increase in customer satisfaction and a 7.2% reduction in marketing overheads[2].

Given all these numbers, we can safely say that marketing matters and that the impact of AI on it keeps increasing. It is no surprise that the subject is widely covered in the literature (see chapter 1). The articles covering the subject either review the most relevant applications of AI in marketing and briefly describe them, or articles choose a specific application of AI and make an in-depth explanation. We haven't any middle ground for readers interested in understanding of the state-of-the-art AI can be applied to marketing without spending countless hours reading specialized papers.

To fill this gap, we will review the fields utilizing AI that are the most used by companies[2]. For each field, we provide an easy-to-understand but pragmatic view of how AI can be applied to marketing. The aim is to provide a simple toolbox to understand the mechanism behind AI in marketing. Another goal is to give a basis to help more in-depth explorations of the applications of AI we covered. We still want to provide an overview of the current overall of AI in marketing hence we provide a diagram regrouping all the fields of use we covered and stating the links between them.

We will conduct a literature review and use the following methodology to systematically analyze each field of use: using the literature on the subject, we define each field, explain its purpose and derive the sub-fields related to it. Then we describe how AI is concretely leveraged to benefit it; this includes an identification of the type of AI used, a review of the central AI methods in the field, a citation of some of the most notorious software platform, and a description of two practical information to implement those methods: the type of data that are relevant and the metrics to evaluate the performances. Finally, we cover some concrete applications by companies.

This paper is divided into four chapters: (1) the current introduction; (2) a review of the concept of "marketing" and the concept of "artificial intelligence" to start the paper on a common basis, then a literature review of the use of AI in marketing; (3) a summary of the AI's field of use in marketing; and (4) the conclusion.

Chapter 1

Review of the main concepts and the literature

The topic of this paper involves the concept of "marketing" and the concept of "artificial intelligence", both of which involve considerable complexity. To avoid misunderstandings we will present the following brief considerations about them. Then we will review the literature about artificial intelligence in marketing.

1.1 Marketing

According to the American Marketing Association (AMA): "Marketing is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large."[4]

Marketing is a set of actions bringing added value to the targeted customers. However this isn't philanthropy, and in exchange for such benefits, the hope is to make a profit. Philip Kotler's definition of marketing states it more clearly: "The science and art of exploring, creating, and delivering value to satisfy the needs of a target market at a profit. [...]"[5]

Marketing isn't merely actions leading to sales, and the second part of his definition explains the importance of the analysis component. "[...] Marketing identifies unfulfilled needs and desires. It defines measures and quantifies the size of the identified market and the profit potential. It pinpoints which segments the company is capable of serving best and it designs and promotes the appropriate products and services."[5]

The marketing literature is quite voluminous and there is an abundance of good theories that help marketers to succeed. Here we will quickly go over the suggestions of J. Lambin and C. de Moerloose for the sake of giving an example of such a theory.

In their work [50], they split marketing into three parts.

1. **Understanding client's behavior**: we first try to understand the client's needs; then we make predictions on the client's cognitive, emotional, and behavioral responses to our potential actions.

2. **Strategy marketing**: we make a segmentation of customers to build a more accurate representation of them; then we make an analysis of the attractivity (i.e. how much each section is willing to buy), and competitivity (i.e. how can we build internal and external advantages against our competitors). Internal advantages are about spending less and less money to build the product, and external advantages are about making a better product that will be sold more or for a higher price; with all that in mind, we will choose a target and how we will position ourselves compared to our competitors.

3. **Operational marketing**: we address the topics of sales by making decisions on the 4Ps: (1) product and what are its exact details; (2) price (i.e. how much we will sell the product); (3) place (i.e. where we will sell the product); and (4) promotion (i.e. how we will advertise the product).

1.2 Artificial intelligence (AI)

The term "artificial intelligence" was introduced in 1956 during the Dartmouth Conference[92]. The subject is complex to define clearly and as of today, there still isn't any consensus on the definition of AI[131].

McKinsey defines AI as "a machine's ability to perform some cognitive functions we usually associate with human minds."[10] IBM has a similar definition: "Artificial intelligence, or AI, is a technology that enables computers and machines to simulate human intelligence and problem-solving capabilities."[19]. Both definitions involve the ability of a machine to copy human intelligence. Other definitions try to define "intelligence" and "artificial" separately but the terms remain ambiguous [57, 56].

We will base the rest of this paper on the definition provided by the most used AI textbook (adopted by over 1500 schools in the world): "Artificial Intelligence: A Modern

Approach"[106]. Next, we propose a summary of its introduction but we can only recommend reading the original paper.

From the definition of intelligence different branches of AI have emerged depending on its accorded meaning. There are mainly two dimensions of divergence leading to 4 main branches of AI. The first dimension is whether intelligence is an abstract concept, only concerned with doing "the right thing", or if it is being as close to human performance as possible. The second dimension is about the process itself: is it an internal thought process, the act of reasoning, or is it an external one, only based on actions and behaviors? What follows is a summary of those four branches.

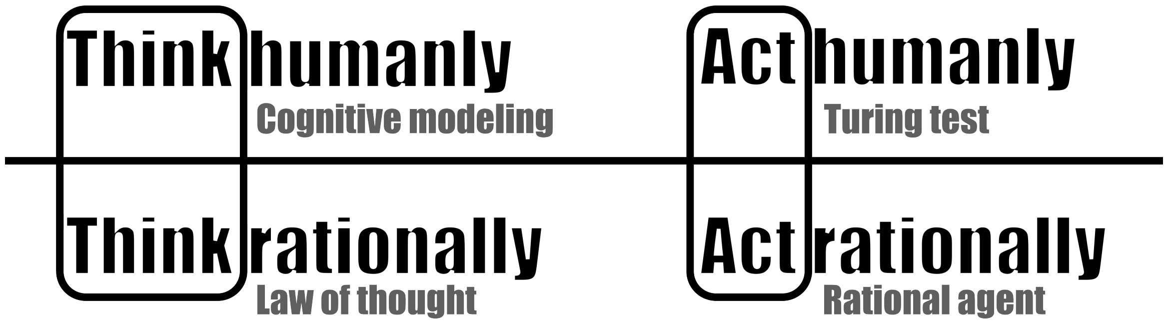

Figure 1.1: AI branches based on different intelligence definitions

- A full digital replica of a brain would allow AIs to think humanly. However, because of the current hardware limitations, a full replica is still out of reach. Instead, scientists try to mimic the human cognitive process by making machines that have the same input-output response to events as a human brain does. To this end, three methods are used to quantify human thought processes: introspection, psychological experiments, and brain imaging.

- We can test whether a machine acts humanly or not via the Turing test. A human interrogator will ask written questions. If the interrogator has no insight into whether the written responses come from a human or a machine (his guesses are wrong half of the time) then the machine passes the Turing test. Out of the attempt to succeed in this test, four fields emerged: natural language processing, knowledge representation, automated reasoning, and machine learning. The Total Turing test is an advanced version of the previous test in which physical interaction with the world is now required and thus two new fields are needed: computer vision and robotics. Some notable examples of this type of AI are the chatbots and robots such as Atlas.

- To think rationally is to apply the laws of logic whenever we have fully reliable information and the law of probabilities in other cases. The crucial part is to properly define those laws and then derive from these more and more advanced calculators.

- Acting rationally is by far the most widely spread branch of AI. A rational agent will try to get the best outcome possible based on a fixed objective. Such an objective can be winning a game of chess, maximizing money, or driving a car without accident. However, stating an objective that correctly defines what we want to achieve isn't an easy task. For example, computers quickly figured out that the best way to drive a car without accident is not to drive at all. Hence the paradigm is limited by how well we can define our goals [106].

In this paper, we will be concerned with AI which acts humanly to replace or reduce human labor, and AI which acts rationally to better achieve goals.

In rule-based systems, AI will strictly follow sets of rules. The AI doesn't learn and is entirely defined by its procedure. In production systems, these rules take the following form "IF a condition is met THEN do something". In logic programming, these rules formulate the problem, and the AI applies logical reasoning to simplify the problem until it becomes trivial.

Thanks to huge hardware improvements, one of the most outstanding fields of the last decades has been machine learning. Its purpose is to learn from past data and then apply that learning to unseen data to make predictions or decisions. There is a wide variety of algorithms in machine learning and they can range from the simple linear classifier to the complex deep artificial neural network (ANN). Deep learning is a subset of machine learning methods only involving these deep ANN.

Machine learning played a major role in the development of natural language processing, a field of AI concerned with the understanding, manipulation, and generation of linguistics. Especially, the recent addition of the transformer ANN architecture[127] which allowed the processing of texts in parallel while retaining consequent context information.

Generative AI is another field that greatly benefits from ANNs. Generating artificial content has been a concern since the 1960s but the first convincing results were made thanks to generative adversarial networks. [62] In short, two ANNs are confronted with each other. The first ANN wins if it succeeds in distinguishing real content from fake content generated by the other ANN. The transformer that we mentioned earlier is also at the root of a new wave of generative AI. Notably, it helped greatly improve large language models (LLMs). We will cover generative AI more in-depth in the following section (2.1).

We won't cover all the other fields of AI since they aren't used within the scope of marketing.

The concept of weak and strong AI was introduced by J. R. Searle in 1980. He defines what we have seen so far as weak AI because it is only a tool. The AI is trained to do a specific task or a range of tasks. In the future, the hope is to build a strong AI that is literally a mind, it would possess similar cognitive abilities as humans have. Such AI would be able to understand other cognitive processes, it would be self-aware, capable of making decisions and doing a variety of tasks. This concept of a strong AI is only theoretical, it might even be impossible to create such AI as of today, because of the restrictions of implied by the current state of hardware development. The final theoretical step would be a super AI that surpasses human intelligence in every way[111, 19].

1.3 Artificial intelligence in marketing

In this section, we cover the extant literature that also reviews the subject of AI in marketing. In his book, J. Sterne already covers in-depth artificial intelligence and especially machine learning for marketing[120]. M. H. Huang and R. T. Rust develop are three-stage framework for the inclusion of AI in strategic marketing: (1) mechanical AI for data collection and segmentation, (2) thinking AI for market analysis and targeting, and (3) feeling AI for customer understanding and positioning [72]. S. Chintalapati and S. Kumar Pandey categorize AI for marketing into 5 themes and 19 sub-themes; then they list and sort 170 use cases from the literature [42]. Several papers, list and explain AI applications in marketing[37, 53, 67]. Both S. Verma et al. and D. Schiessluse et al. use network analysis to offer a comprehensive and visual review of the extant literature on artificial intelligence in marketing[129, 108]. B. Vlačić proposes another review using content analysis and the HOMALS statistical procedure[17]. A. De Bruyn et al. propose recommendations to avoid pitfalls and seize opportunities when implementing AI in marketing, with a special focus on ANNs. V. Devant et al. lead a statistical analysis on the willingness to use applications of AI in marketing[51].

Chapter 2

AI's fields of use in marketing

This chapter contains the main substance of this paper: it contains an overview of the most common AI field of use (i.e. application). In order to have a reliable and up-to-date basis for this chapter, we will use the results for the fall 2023 CMO survey[2]; it provides a list of AI's field of use (i.e. how AI can be leveraged to improve marketing) and the corresponding percentage of use by US companies. Each section will named after one field of use and in parenthesis next to it, the corresponding use rate will be written according to this template: "Name of the field of use (use rate)". We will define each field of use and its purpose. We will also explain how it can be applied to marketing. We will see how it works in practice, what are the most prevalent algorithms and which metrics are used to evaluate the performance of those AIs. We will see which data are necessary and what are the most known software platforms that provide service for the field of use; then we will cover some concrete examples of its use by companies.

The structure of this work is summarised on the diagram in figure 2.1. It contains all fields of use we will cover and their related sub-fields. Their dependence relations is also indicated by arrows. For example, targeting decisions are often made based on a prior customer segmentation hence an arrow goes from targeting decision to customer segmentation. The fields of use are also placed in boxes indicating the main category of AI they leverage. However, this doesn't mean they use exclusively this type of AI (e.g. data analytics mainly use machine learning but can also specialise in deep learning or use methods from natural language processing).

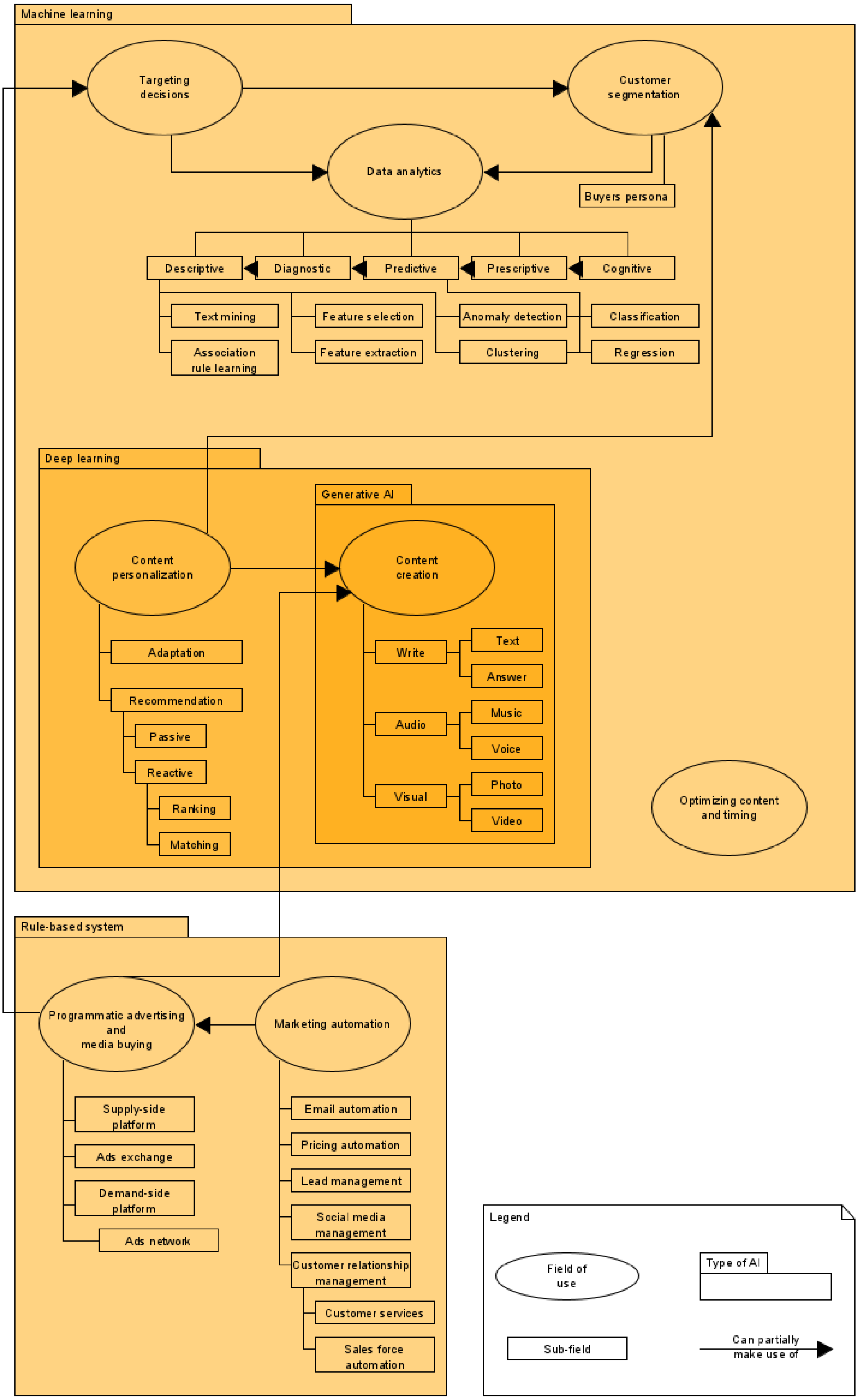

Figure 2.1: Summary diagram of the fields of use covered in this paper, their sub-fields, and their relationship

2.1 Content creation (49.2%)

Let's start by explaining why content creation is relevant to marketing. "Content marketing is a strategic marketing approach focused on creating and distributing valuable, relevant, and consistent content to attract and retain a clearly defined audience - and, ultimately, to drive profitable customer action." [11]. But what is content exactly ? In this paper we will use the Merriam-Webster dictionary's definition: content is "the principal substance (such as written matter, illustrations, or music) offered by a website".

Generative AIs were created in order to solve the exact problem of creating content and according to an April 2023 survey [103], 70% of CMOs from several sectors already use them (not exclusively for content creation though) while a remaining 19% are testing them. Moreover, 7% of the time spend in marketing activities is spend using generative AI[3]. Most of these AIs work in the following way.

First, they are trained to build their knowledge. They are fed vast amounts of data related to their purpose. Generative AIs perform better when trained on a specific type of content or even a specific subject. Some models even support the addition of new data to get specialized further.

Then the model can be used in a three-step process.

1. The user explains to the best of his capacity what he wants via a prompt. The prompt can take various forms such as an audio sample or an image but the most common one is a combination of text and parameters (e.g. Draw a square; then set the image's size to 256x256 pixels and heat to 8).
2. The trained model takes the prompt as an input and processes it. In most cases, the algorithm consists of a very complex artificial neural network.
3. After computation, the model returns an output to the user.

The quality of generated content is directly related to three factors: (1) the knowledge bases of the AI, (2) how well the user expresses what he wants, and (3) how well the algorithm digests this information. If we restrict the field to a specific type of content, the user will have an easier time expressing what he wants to achieve and the algorithm will perform better. It is no wonder that we have specialized generative AIs dedicated to each category within the content's definition. Following is a summary table of each of those categories.

Type	Marketing usages (see usage percentage by US companies [2])	Objective methods of evaluation	Generative AI names
Write	Blogs, websites, social media texts, product/service descriptions, new stories and copies (email, ad, sales, technical, packaging, logo)	MMLU[68](mix), Hellaswag[137](common sense), NaturalQuestions[86] (comprehension), GSM-8K[44](math), HumanEval[40](coding)	ChatGPT, Perplexity, Gemini, Mistral 7B (open source)[78]
Visual (image)	Blogs, websites, social media visuals, augmented/virtual reality and design (ad, packaging, logo, game)	FID[69], CLIP[100]	Stable Diffusion, Midjourney, DALL-E
Visual (video)		FVD[126]	Sora, Gen-1 Runway, Invideo
Audio (music)	Soundtracks	FAD[81]	AIVA, Soundful
Audio (voice)	Dubbing, chat bot voices		ElevenLabs, Speechify, Lovo.ai

Figure 2.2: Non-exhaustive summary tabular. For each content category: how to use it, how good is a generative AI to create it, and some common GenAI names in the category

In figure 2.2, we mention some objective methods used to evaluate the quality of generative AIs. Such methods are used very frequently, especially in the process of building and improving new generative AIs. We identified two other family of metrics with their own merits and flaws. Here is a rundown of each family of metrics.

- **Objective metrics** are consistent and easy to compute but are heavily dependent on what is set as "objectively" good and how well it can be measured. One of the most common methods is to give a higher score the closer a generated content is to an ideal content based on a distance function. Using an appropriate dataset, the process is repeated over a large number of sets of [prompts and the corresponding ideal content]. Finally, all the scores are averaged into a final score.
- **Subjective metrics** are best to fit human standards. To compute these, humans are asked to compare and/or score generated content based on a set of criteria and from these results, a ranking between the AIs can be established. The obvious drawback of such methods is the investment in human work time.
- **Meta metrics** are the same for every category of content and are very pragmatic since they try to evaluate the impact of the generative AI's usage on the company.

However, it can be hard to decorrelate the effect of using such AI with other strategic decisions. Since every company is different, it is also much harder to make comparisons between AIs. Some common meta metrics are return on investment, customer satisfaction, number of quality leads, and conversion rate.

We will end this section discussing the case of Coca-Cola, which leveraged the power of generative AI for a publicity campaign called "Create real magic". During March 2023, using a combination of GPT-4 and DALL-E, Coca-Cola proposed on a website [1] an easy-to-use toolkit to generate artwork that features Coca-cola assets. The motivation was that the best visuals were displayed on the billboards of cities such as London or New York.

2.2 Content personalization (52.8%)

According to McKinsey&Company, "personalization is when seller organizations use data to tailor messages to specific users' preferences." [13]. One common example is Netflix using customers' data in order to provide them with a personalized catalog of series and movies. The special experience that personalization offers, pleases most customers. In fact, 71% expect it and 76% are even frustrated if they don't find it [9]. It is then no surprise that good personalization yields good results for companies. It can raise revenue by 5 to 15%, increase ROI by 10 to 30% and reduce acquisition cost up to 50% [63].

The goal of personalization is to use data to deeply understand customers and then create personalized content for them, sometimes use content creation methods. Machine learning is a field of AI that perfectly fits this kind of application and it has plenty of methods to choose from. It is then advised to split a personalization problem into multiple personalization tasks (e.g. rather than personalizing a whole website at once, personalize the recommendations, messages, notifications, ads, search engine, etc. individually) as this offers more flexibility and there isn't an all case best algorithm. Instead, the performance greatly depends on the type of personalization task and the amount of available data. It's best to find the most suitable algorithm for each problem. [119]. To build an even better representation of the customers, tools such as segmentation can be used. Another approach, tedious but only reliant on segmentation, is to manually prepare customized content for each segment (see segmentation 2.6).

[1] https://createrealmagic.com/

The content personalization process can be divided into three successive steps: (1) data collection to obtain the information we will use to make the personalization, (2) learning models to leverage the power of AI to learn from the data and propose personalized content, (3) performance evaluation to see how well the learning model is doing and to compare them with each other.

Data collection

Data collection is the mandatory step to identify the client's behavior, needs, and preferences. We can distinguish three ways of obtaining new data. First-party data are collected by the company itself, second-party data are shared by partners, and third-party data are purchased by another company owning the data. We propose a division of customer data into two families.

The first family is about user-interaction data. They are specific to each user and can be further divided into two sub-families:

- **Transactional data** are all the information we can capture from transactions: the time, the place, the price, the people involved, the payment method, etc. Every time a transaction occurs, we can store related data.
- **Behavioral data** characterize all the other interactions between the user and the product/application. These data can be explicit or implicit: explicit when they result from a conscious user response (e.g. star rating, thumbs up, survey) and implicit when they are passively collected via web/application analytics such as cookies (e.g. number of clicks, watch duration, navigation path, search history).

Finally, the collection of user-interaction data might suffer from a cold start. Starting with few to no user-interaction data, the model can't be properly tuned and is therefore less attractive and less likely to collect new data. However, there are some solutions to reduce this effect. [64]

The second is about heterogeneous data which are general information without direct links with customers. They also can be split into two sub-families.

- **Contextual data** give insight into the circumstances of an event. This can be the place, weather condition, economic context, continuous/discrete time, etc. For each type of context, the collection methods will vary.

- **Demographic data** are statistics that describe the population according to different factors such as age, gender, income, family status, etc. They can be collected via a survey or in most cases are bought as third-party data.

Following data collection, data pre-processing is a crucial step in machine learning. However, we won't cover this aspect in this paper as it is too broad a topic, which strongly depends on the specifics of the problem and the algorithm used. Speaking of which, it seems that when using only user-interaction data, well-tuned simple models remain fairly competitive. However, when using more complex models (e.g. deep learning algorithms), the addition of heterogeneous data often increases the performance. [48, 119]

Learning model

Once the data are available, a variety of applications can be used to improve the user's experience and make the clients feel special. We sorted these applications according to two main types of personalization: adaptation applications modify existing content or fill templates to fit customer's tastes, while recommendation applications select and show the best content possible to improve sales and/or user experience.

The goal of content adaptation is to make content subjectively better by slightly tailoring it toward each client's tastes, only slightly, so as to avoid being too intrusive and to keep the initial intent intact. Modifications can be aesthetic improvements by finding the optimal layout, colors, images, videos, font, text size, etc, or they can be changes in the formulation of the text, adapting the language to the user, making description, email, chatbot, and notification more appealing. A common example is the addition of the customer's name in the message[119].

In recommendation, machine learning models are used to select the most adequate content for the customer. Here we have to distinguish passive recommendations that are done continuously from reactive recommendations that result from users' requests. When the client logs in and navigates through the website/application, passive recommendation algorithms put the spotlight on some products, they select which third-party advertising to display and make suggestions on the client's behavior (e.g. "Have you tried this new section?").

In reactive recommendations (i.e. search engines), the user makes a search to either fetch, find or explore: fetch when he has a specific item in mind (e.g. Lilium orientalis), find when he knows what he wants without being specific (e.g. flower) and explore when he wants to discover new content (e.g. gardening). Accordingly, he describes what he wants via a text prompt[2] then the search engine returns a list of items that best suits the user's request. In comparison with simple recommendations, we have extra information to work with but the user has a specific expectation that needs to be matched. In practice, the search engine can be a product in itself (e.g. Google) or it can help to navigate through a website or an application by providing better recommendations. When it comes to implementation, a common approach is to split the problem in two: a matching phase followed by a ranking phase. [87, 94, 32]

- **Matching phase**: one builds a smaller set of items that matches the user's prompt to improve speed performance during the ranking phase. This step is especially relevant for large databases of items where considering all items at once during ranking would take too much time. One also tries to understand the user's goal (fetch, find, explore) to adapt the filtering accordingly. Today, advanced matching algorithms will use behavioral data and semantic matching in addition to traditional keyword matching. In a basic keyword-matching algorithm, an item is added to the match set if it contains words that are in the prompt. However, this simple method leads to three issues. First, a lack of understanding of lexical variations such as hypernyms, synonyms, and antonyms. Second, a fragility to morphological variants (e.g. personalisation vs. personalization). Third, a sensitivity to typography mistakes[94]. Semantic matching solves these issues by trying to understand the meaning behind the prompt. It is done via a subfield of AI called natural language processing (NLP).

- **Ranking phase** one scores each item in the match set before ordering them. The ranking can be done in multiple stages. At each iteration, the top-ranked items from the last stage are re-ranked in the next stage until we are left with only the best items. The ranking of an item can be based on a great variety of criteria such as customer satisfaction, business metrics, the item's relevance, the degree of trust, and authority. The relevance criterion is almost always present and can be computed via the semantic similarity score of the item and behavioral data. One attributes higher rank to items that closely match the prompt semantically. Finally, an example of behavioral data usage is to promote items that were often selected and demote items that were ignored.

[2]Technically it could be any type of prompt such as image, video, and/or audio but in practice search engines use almost exclusively text for convenience reasons

Performance evaluation

Models are compared in order to pick the best model. We can identify three categories of metrics each corresponding to a step of the evaluation process. To have a better understanding, we will cover those categories backward.

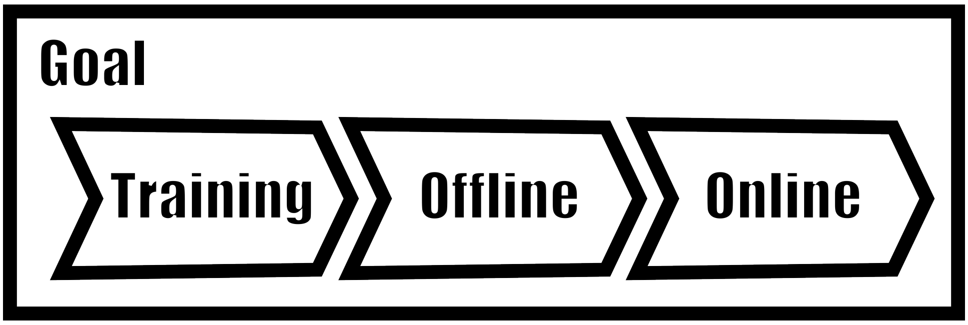

Figure 2.3: Performance evaluation steps of personalization

The **goal** is the frame that defines which metrics and criteria to use. We won't look at the same indicators when we want to optimize customer satisfaction as when we want to maximize sales.

Online metrics are measured on customers in real time. They are the most reliable set of measures because of their direct relationship with the user. However, they have a couple of flaws. To reach statistically significant conclusions, they require investment in time and resources because the flow of user feedback is limited. Also, since models are tested on real customers, there is an inherent risk of losing money if a model performs poorly.

In A/B testing, users are randomly exposed to (A) a control version of the content or (B) a test version. The test can also be run with any number of different content versions then users will be split evenly. For each version, users' responses are measured via metrics aligned with the tester's goals (e.g. retention rate, number of leads). Then a statistical analysis can determine whether one version gets significantly better responses than the others and which one it is.

Multi-armed bandits are more complex versions of A/B testing that dynamically adapt the distribution of users to improve benefits during the test[134]. Contextual bandits are similar reinforcement learning algorithms that take into account users' information to further improve the distribution. [88]

Interleaving methods [39, 79, 101] were designed specifically to evaluate recommendation ranking methods by using clickthrough data. Such methods are based on the observation that users typically look at the top k elements of a ranking and pick the most relevant option out of those; in web queries, k is around 10 [114]. Interleaving methods leverage this observation by making mixes of ranking A (control) and B (test). Those mixes are then tested by submitting those to the users. If significantly more users click on items from rank B than A, then the method that produced B is assumed to be the better one.

Offline metrics try to estimate model performance as well as online metrics but with one key difference: they aren't limited by the users' feedback rate. Instead, they use historical data. In practice, they are used to narrow down the number of models tested by online metrics.

Recommendation metrics make evaluation based on queries (i.e. sets of N ordered items returned by the model where N is a hyper-parameter to define). We will cover the most common metrics by gradually increasing the order of complexity. However, it is important to note that each metric can be used depending on the level of precision required.

The **recall** is a simple metric that measures the ratio between the number of relevant items in the query and the total number of relevant items. Other similar metrics, namely **precision** and **F-score**, don't take the order into account. This isn't the case for **MRR** which looks at the rank of the first relevant item of each query and then averages those ranks. A perfect score of 1 means that the first recommendation is always relevant. However, this method doesn't take into account the relevance of other items. **MAP** doesn't have this issue. For each query, this metric computes the average precision of relevant items. The final score is the mean of those average precision. Instead of a binary variable indicating whether the item is relevant or not, **NDCG**[76] can offer more flexibility by using relevance scores. Each rank has an associated penalty, a lower rank will have a higher penalty. The NDCG score is the normalized sum of the relevance scores of each query item divided by their respective penalty.

Training metrics, also called loss functions, are necessary to train some of the machine learning models (e.g. ANN). The metrics are split into two categories of tasks: classification and regression. We won't cover the details of machine learning as this is outside of the scope of this paper but two very common choices are cross-entropy for classification and RMSE for regression.

We have seen the example of Netflix; it personalize is whole web page to user[119]. Amazon recommend items that will please the customers. Apple personalize the App store to the user; the search engine is adapted to the user's data, and the app includes tabs dedicated to recommendations[6].

2.3 Optimizing marketing content and timing (36.6%)

Search Engine Optimization (SEO) is the art of boosting a website's page ranking on search engines. It can be done through web analytics or by leveraging the power of ML and NLP. Some common names of SEO tools using AI are SurferSEO, Frase, and Semrush.

Optimizations are based on datasets containing search engine queries. Then information such as search volume, relevance, and competition are extracted to identify the most valuable structure and keywords. It is recommended to optimize according to both text-based and voice-based searches. Indeed, spoken queries tend to be longer and more conversational (i.e. include more stop words) than written ones[47]. Moreover, spoken queries tend to be associated with AI assistants such as Siri, Alexa, or Bixby which have their own search algorithm and often provide a single answer instead of a list of answers.

On-page optimization is concerned with improving the content and structure of a web page to achieve SEO. The most crucial step is to optimize the title tag and meta description of a page by making them more engaging and including the most relevant keywords in optimal densities. Then an analysis of every bit of text inside the page can be done in order to identify keyword optimization opportunities and adapt the text accordingly. Finally, images and videos can also be improved by modifying their title, descriptions, and captions.

2.4 Programmatic advertising and media buying (35%)

According to Amazon: "programmatic advertising refers to the practice of automating media buying and creating digital ads with the use of marketing technology. For an effective programmatic advertising strategy, use an automated workflow to effectively deliver ads to your audience" while "programmatic media buying uses an automated process to buy digital space for ads." [1].

The problem of advertising and media buying is the following. A publisher has available space on his website, radio, video, etc. Advertisers want to buy this space to display advertisement that promotes their brand. For online advertising, the process of making deals between publishers and advertisers can be automated (see figure 2.4).

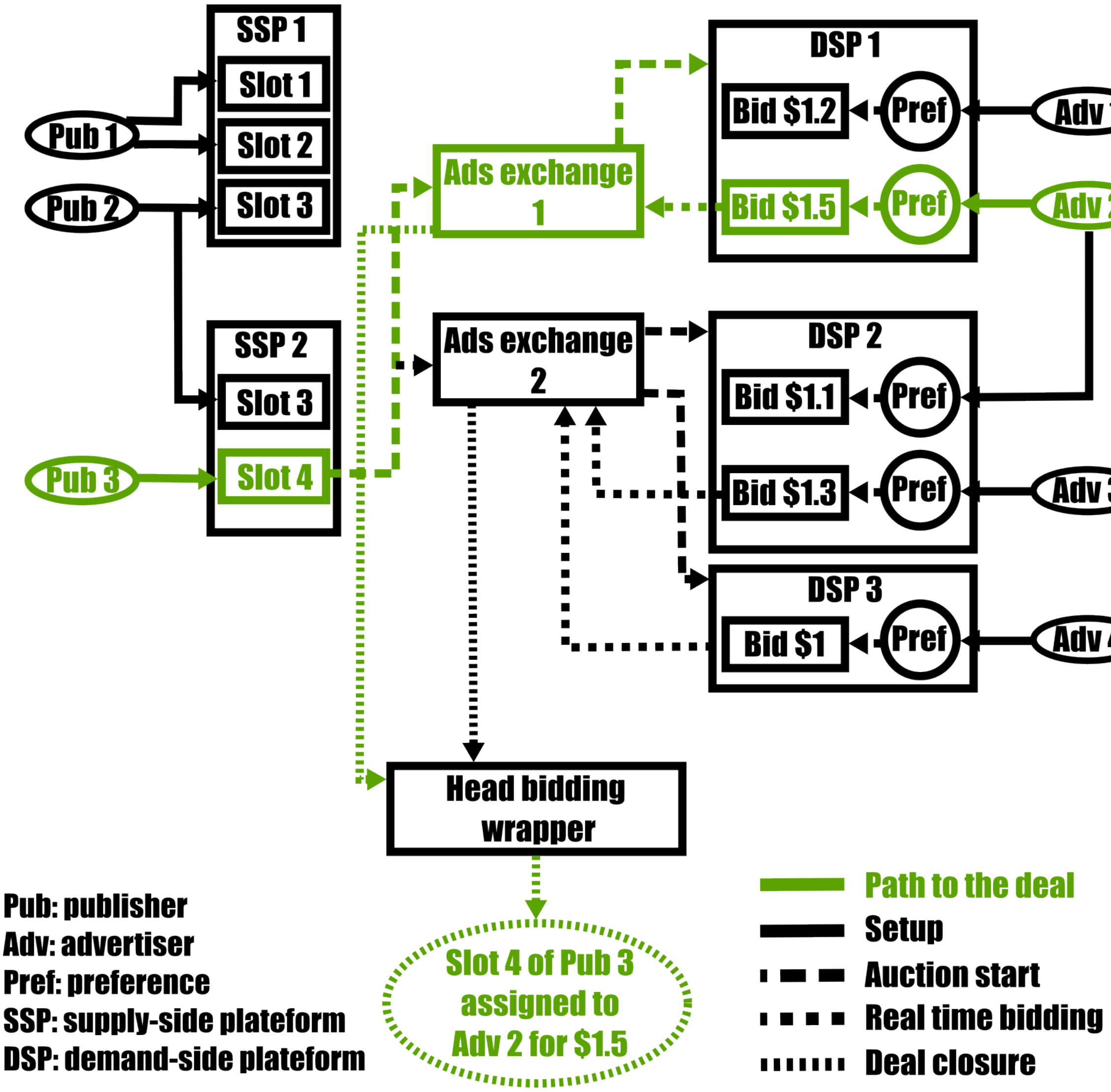

Figure 2.4: Programmatic advertising summary scheme

We can divide the process into four steps. The first one is the setup. The last three steps occur in a fraction of a seconds, during the loading of the page whenever a consumer enters a publisher's website. Step two is the beginning of the auction, step three is the real-time bidding part and finally, there is the deal closure.

1. On the one hand, publishers choose their supply-side platform (SSP) (e.g. OpenX, Google Ad Manager, PubMatic). On these software platforms, they select which

space of their website is available for ad auctions. Sometimes they even use the same slots on different SSPs to increase their odds of sale. On the other hand, advertisers pick their demand-side platform (DSP) (e.g. The Trade Desk, Amazon DSP, Google Display & Video 360). Based on consumers' information such as demographic and shopping patterns, advertisers state their targeted audiences (see section 2.7) and how much they are willing to pay for each of them.

2. When a user is about to enter the website, for every available slot, SSPs will automatically start an auction on dedicated digital marketplaces called Ads exchanges (e.g. Google Doubleclick, Bing Ads, Rubicon Project). The information is thus transmitted to DSPs.

3. DSPs analyze their subscribers' choices and see which one is willing to pay the most to get this specific container for that specific consumer. Then each DSP proceeds to bid on Ads exchanges.

4. The highest bid of every Ads exchange is then collected by a piece of JavaScript code called "head bidding wrapper" which determines the winner. The winning advertiser gets to print its ads on the publisher's website in exchange for the promised money.

Ad networks (e.g. Google AdSense, Mediavine, Monumetric) are platforms for publishers that manage the whole process internally. They sometimes have a premium deal with the publisher and/or advertiser and they place bids on the ads exchange. They can apply three different cost policies: (1) cost per mile (CPM), in which the publisher pays a fixed amount every 1000 impressions, (2) cost per click (CPC) in which case it pays for every click on the ad, and (3) cost per action (CPA) in which case it pays every time the ad leads to a sale.

The whole process of automation can be seen as a form of AI. It acts as an intermediary between publishers and advertisers and makes decisions based on their preferences. Moreover, AI, and more specifically machine learning, can be used to optimize decisions inside this framework, especially for advertisers. Data management platforms (DMP) are mines of data about the consumers that can be used to improve bidding decisions. This can be done by choosing better campaign durations, optimizing the audience targeting (see section 2.7), delivering the customers personalized ads, or finding which digital containers are the most cost-effective for specific users. Finally, historical data can be used to determine which media channels are the most relevant for the audience (e.g. Annika [24]).

A set of key performance indicators (KPI) can be used to evaluate how well a publicity campaign is doing. There are click-related indicators such as the number of clicks per ad, and the click-through rate (CTR), which is the ratio between the number of people clicking on the ads and the one seeing it. Click-through conversion (CTC) is similar to the CTR but instead takes the number of persons purchasing after clicking on the ads into account. There are also more global indicators like the number of ad impressions on the website or the cost per acquisition (i.e. the number of converted customers divided by the total cost of the campaign).

Social media (e.g. Meta, TikTok, Snapchat) almost all use programmatic advertising to sell their media slots. Often they even develop their own platform[3] to ease the process buying for advertisers. The advertising time on the start of Youtube videos are sold on auction using the programmatic advertising process.

2.5 Data analytics (predictive analytics (32.9%))

Data analytics is the process of analyzing raw data to uncover meaningful insights. Based on the level of insight gained, Z. Król and D. Karol state five complementary categories of analytics maturity [85]. The better the maturity, the greater the competitive advantage. We will cover each type of analytics by stating its purpose, explaining how to perform it, and giving typical questions it aims to answer.

1. **Descriptive analytics** are the processes that summarise vast amounts of raw data in order to learn "what happened". This is mainly done through data mining. We believe that the most relevant methods for this task are **feature selection** in order to only keep the most relevant features, **feature extraction** in order to obtain a limited set of mixtures of features that preserve most of the information, **anomaly detection** in order to spot uncommon behavior, **clustering** in order to identify groups of objects, **association rule learning** in order to identify relationship between variable, and **text mining** so as to find the most relevant information in texts. Finally, some examples of questions that descriptive analytics answer are "How did the sales go ?" or "What is the state of our marketing campaign ?".

2. **Diagnostic analytics** aims at understanding "why it happened". They rely on a good descriptive analytics basis and then require a more in-depth analysis, understanding, and interpretation of the results. While descriptive analytics only need a

[3]Meta: https://business.meta.com/
Titktok:https://www.tiktok.com/business/
Snapchat:https://forbusiness.snapchat.com/

snapshot of the subject, diagnostic analytics often require an evolution of it through time. Diagnostic analytics answer typical questions like "Why did this campaign outperform our previous ones ?" or "Why does this store make such poor sales ?"

3. **Predictive analytics** is about forecasting the future based on past information. They also require an understanding of the relation between variables through time in order to extrapolate this relationship. It can be divided into two categories: estimation which predicts numerical values and classification which predicts the class of objects[135]. There is a wide variety of machine learning techniques to make predictions. We can cite **regressions analysis**, in which we assume that our data follow a certain trend and make an estimation based on it. The most common regression model by far is linear regression. **Classification analysis** methods classify objects based on previous observations. **Anomaly detection** and **clustering** are descriptive analytics methods that can also be used to make classification once the models are trained. All these models help to answer questions such as "What amount of sales will we make next year ?" or "Is this new customer likely to come back?"

4. **Prescriptive analytics** make predictions and then use them to determine "how we should act" in order to reach the desired result. It often offers a panel of decisions to choose from, as well as their implications. One of the most common questions they try to answer is "What can we do to increase our sales ?".

5. **Cognitive analytics** are very similar to prescriptive analytics but include an automated decision-making process. For example, this can lead to a system that uses real-time data to manage the stocks in warehouses and that supply them when needed.

We will now briefly cover each method we cited earlier. The aim is to offer intuitive explanations of the techniques that constitute the data analytic landscape.

Dimensionality reduction can be divided into **feature selection** and **extraction**. It is useful when dealing with high-dimensional data as it allows for 2D or 3D graphic representation or a representation in the shape of a table with a limited number of columns. Outside the scope of data analytics, it is often used in machine learning during preprocessing, so as to enhance results and escape the curse of dimensionality [25, 128].

Feature selection aims at narrowing down the wide number of features. It is often divided into three categories: (1) **filter methods** progressively discard features that provide the least amount of variance or any other metrics quantifying information; (2) **wrapper methods** try different subsets of features and keep the subsets with the most information, while this method takes into account the interactions between variables, it is slower due to the large number of iteration it requires; (3) **embedded methods** try to combine the advantages of both previous methods by incorporating the feature selection process inside the algorithm that use it. One example of embedded methods is LASSO[124], which during the regression process can discard features[4].

Feature extraction, as opposed to feature selection, doesn't conserve the features untouched. Without this constraint, it can create more efficient low-dimensional spaces at the cost of interpretability: since it extracts new features out of the previous ones, it is often hard to know what these features represent. Following is a list of the most used techniques for feature extraction.

- **Principal component analysis (PCA)** [133] is by far the most notable technique. PCA uses linear combinations of features to build an arbitrary number of principal components (i.e. features) that capture the most variances out of the data, one by one. **Kernel PCA** [109] is a variant of PCA that handles non-linear data better by first mapping the data onto a higher dimensional space.

- **Linear discriminant analysis (LDA)** [58, 123] is similar to PCA as it also creates new features out of a linear combination of previous features. It allows for a clearer representation if we can sort our data into classes, as, instead of maximizing the variance, it maximizes the separability between those classes. It does so by minimizing the inner variance of each class and by maximizing their distance from a central point.

- **t-distributed Stochastic Neighbor Embedding (t-SNE)** [70] is one of the best algorithms to visualize high dimensional non-linear data. First, it builds a similarity matrix between each pair of points; then the similarities are normalized as probabilities. Such matrices are built both for the high dimensional space and the desired low dimensional space. Once the two matrices are built, the algorithm iteratively minimizes a distance, traditionally the Kull-Back Leibler distance, between them.

- **Auto-encoder** [84] is an hourglass shaped artificial neural network structure. It is composed of two parts: an encoder which compresses the input into a bottleneck of neurons and a decoder which decompresses it. The bottleneck is composed of a small number of neurons, which can be considered as features enabling a compressed representation of the input.

[4] For practical details: the method doesn't directly discard a feature but instead sets its coefficient to zero essentially removing its impact

Regression analysis and **classification** both belong to supervised learning, meaning that during the training phase, each data point has a label indicating its class for classification or its value for regression. After the training, models will make predictions on the labels of new data points.

Regression estimates the relationship between a dependent variable, the label, and a set of independent variables, the features. **Linear regression** chooses a parametric function, then fits its parameters to the historical data. Fitted functions can be lines, polynomials, logistic functions, or anything else, as long as the inputs are expressed as linear combinations. **Non-linear regression** doesn't have such restriction. Therefore, the choice of functions is increased but the fitting of the parameters is harder. Estimators determine the values of the parameters by minimizing a loss function. However, in most cases, there is no closed-form expression for non-linear models and optimization must be done by numerical optimization. For the linear model, the traditional choice is the "ordinary least square (OLS)" that minimizes the sum of the squared differences between the true value of data points and their estimated value. **Robust regression** as its name suggests uses loss functions that are more robust. For example, the Huber loss is less sensitive to outliers (i.e. points that differ significantly from the others). In **step-wise regression**, the choice of features is performed automatically. **Non-parametric regression** models use no parameters, **k-nearest neighbour (k-NN)** and **artificial neural networks (ANN)** are two of them.

With some minor changes, some models can be used for both **regression analysis** and **classification**.

- **Logistic regression** [26, 91] is one of the leading models to make binary predictions[135]. It is a statistical model that fits the parameters of a logistic function similarly to other regression algorithms. The values on the curve can be used for regression as such or as probabilities for binary classification. For classification, if the probability is above a fixed threshold, usually 50%, it is classified in the first category otherwise in the second. Logistic regression can be extended to multi-class classification via the **softmax regression** [77]. Softmax is also often used in the last layer of ANN to perform classification tasks.

- **k-NN** [46], as its name suggests, is a method that, according to an arbitrarily chosen number k and a distance metric, makes predictions based on the k closest data point to the new observation. For classification, the most common class among neighbors becomes the predicted label and for regression, the predicted value is an aggregation[5] of the neighbors' values.

- **ANN** [60] are extensions of the perceptron [93] algorithm that was used for binary classification. The networks consist of processing units called neurons regrouped into layers. Neurons are connected to each other, and a strong bond between two neurons is characterized by a weight of high value. Backpropagation is a trial-and-error learning process made to adjust those weights based on historical data. To make predictions, new observations are fed into an input layer; after computation on the hidden layers, the predictions come out on the output layer.

- **Decision trees** [99] predicts a category from a succession of tests defined by a tree-like structure. Compared to more opaque algorithms such as **ANN**, having a structure explaining the decision process helps with the interpretation. **Random forests** [71, 30] use multiple decision trees to make more accurate decisions. **Regression tree** is the regression variant of the algorithm. Instead of associating a class label, it associates the mean[6] of the values from the class.

Following are the most prominent algorithms for **classification** only.

- **Naïve bayes** is a statistical model. For each feature and for each class, it use past data to estimate the likelihood that the new observation belongs to that class if it has that feature value. Then it combine those probabilities to determine which class is the most likely for the new observation.

- **Support vector machines (SVM)**[45] fit a hyperplane to split the data and make binary decisions. In some cases, a mapping to a higher dimension space called feature space via the kernel trick is operated to obtain better results. Such methods are called **kernel SVM**[28].

Clustering is a family of non-supervised learning algorithms; hence they don't require any label during training. Its role is to group similar objects into clusters. Clusters can be defined based on distance, density, connectivity, and many other ways; hence there is a wide variety of clustering algorithms, of which we will present the main ones.

- **k-means** [90, 59] split the data into k clusters. The number k is either chosen arbitrarily or via heuristics (e.g. silhouette method [105]). The naive version of the algorithm can be summarised as follows. First, k centroids are placed randomly

[5] Most common choice is an average or a weighted sum

[6] Or any other aggregation method

in the data space. Then the following loop repeats until stabilization. The data points are assigned to the closest centroid. The centroids are moved to the center of their assigned data points. After stabilization, the final centroids and their associate data-point are the clusters.

- **Density-based spatial clustering of applications with noise (DBSCAN)** [107] determines clusters based on density. Clusters are made of core points and the points they can reach. A point is defined as core if it has enough points sufficiently close to it and a point is reachable if it is sufficiently close to a core point. The exact number of points and distances are meta-parameters that need to be set by the user.

- **Hierarchical clustering** [27] can be divided into two families: agglomerative and divisive. In agglomerative methods, there are as many clusters as observations at first; then, step by step, according to a linkage criterion, the most similar clusters are merged; the process continues until there remains only one cluster containing all the observations. Divisive methods work the other way around: they start with one cluster and divide it until each observation has its cluster. The main particularity of these algorithms is that each step gives a valid clustering. Thus one of the main difficulties is to choose the step that will give the best clustering. To help the decision process, results are usually presented as dendrograms.

Association rule learning is also a non-supervised learning method. The goal is to identify strong association rules between items for transactional data which consist of a list of transactions. A transaction can contain one or more items. These algorithms will efficiently[7] search the dataset to see if some items are frequently bought together. Such information can then be used for marketing activities such as item placement or promotions. We will cover the three main algorithms for this purpose:

- **Apriori**[20] is based on the observation that a set of items cannot be frequently bought if one of its subsets isn't frequently bought (e.g customers can't possibly buy milk and chocolate often if they don't even buy milk often). The algorithm proceeds level by level. On level x, it searches for frequent itemsets of x items. Then by using the observation we explained, it generates the candidates for the next iteration, keeping only itemsets of $x + 1$ items that can still be frequent.

- **Eclat**[136] uses a depth-first approach and a vertical representation of the data (i.e. transactions are associated with items instead of the other way around). By

[7] Since the number of possible item combinations is exponential in the number of items, testing every possibility often takes too long.

focusing its attention on subsets of items one by one, it allows for smaller representations of the dataset during the search process (e.g. when searching for frequent itemsets containing bread, any itemset that doesn't contain bread can momentarily be discarded).

- **Frequent pattern(FP) growth**[65] is an algorithm that uses another data representation: a tree structure that stores the association between items. The tree is then divided into a set of conditional trees each associated with a frequent pattern. Finally, the trees are mined separately to find every frequent pattern.

Anomaly detection methods are used to identify results out of the norm. They can be great for identifying potential opportunities or dangers, although these abnormal results are often due to an error during the data collection process. The methods are divided into three categories: supervised, unsupervised, and semi-supervised detection. Supervised means that during the training each data point has a label indicating whether or not it is an outlier. On the contrary, in unsupervised detection, there are no such labels. Semi-supervised is a mix of the two, only a portion of the data is labeled. Quite often, only the normal data are identified as such and there are no examples of outliers. Supervised detection can be seen as a binary classification problem; hence the classification methods we have seen earlier can be used for this purpose. Next, we describe some of the machine learning algorithms that can be used for unsupervised and semi-supervised outlier detection.

- **k-NN** [138] is one of the algorithms we encountered earlier. Exploiting the fact that outliers tend to be far away from other data points and computing the average distance[8] to their neighbor, we can assign an outlier score to each data-point.

- **Local outlier factor (LOF)** [31] is partially based on the K-NN algorithm. It is designed to identify local outliers more easily. Local outliers aren't unusual when considering the dataset as a whole but are when considering the neighborhood. LOF computes how easily a data point is reached compared to its k-nearest neighbor. If it is harder to reach it will be suspected to be an outlier.

- **Auto-encoder** [74] can also be exploited to detect anomaly. The process is based on the following assumption: since the ANN is mostly or entirely trained on normal data, it will make more errors when encoding and decoding outlier data. Hence, if the reconstruction error is above a defined threshold (i.e. the input and output of an auto-encoder differ too much) the data point is suspected to be an outlier.

[8]Various distance or even density measure can be used but the usual choice is the Euclidean distance

- **Isolation forests** [89] are based on the assumption that outliers are more easily separated from the rest via a succession of binary decisions. This is justified since outliers often have very different values for some of the features. After running the random forest algorithm usually used for classification, data points with a low depth in binary trees (i.e. reached within a few binary decisions) are suspected to be outliers.

- **One class SVM** is a variant of the classification algorithm SVM with only one class: the standard data points. Either a hyperplane [110] or a hypersphere [122] is fitted into the feature space to build a boundary for the class.

- **DBSCAN** [107] is one of the clustering methods we have seen previously. If a point is neither core nor reachable, it is by definition far from a dense region of data points; therefore, it can be suspected to be an outlier.

Text mining is the process of extracting information out of texts. Documents can be regrouped using clustering or put into categories (e.g sport, scientific, leisure) using classification. **Sentiment analysis** reduce large amount of texts to a digital value, usually between -1 and 1, stating the emotional tone of the message. Positive values are associated with positive feelings and negative values with negative feelings[36]. Generative AI (see section 2.1) can be used to summarise the text. NLP methods can be utilized to obtain deeper information on the meaning of the text[16].

IBM sells analytics services called Watson AI. It leverage machine learning and generative AI to provide descriptive and predictive analytics. Based on historical data provided by the customer it can for example provide summary spreadsheets, Q&A tools powered by generative AI or generic predictive models.

2.6 Customer segmentation (21.5%)

The segmentation-targeting-positioning (STP) framework is one of the core concepts in strategic marketing[83]. Segmentation[116] splits the market by classifying customers into groups with similar needs, called segments. The division can be done according to different variables, such as searched advantage, psychographic, behavioral, lifestyle, buy opportunity, and real-time data. A segmentation is efficient if it has these four characteristics: (1) it is realistic, (2) it has a sufficient earning potential, (3) the communication and distribution can be done selectively, (4) the demands differ from segments to segments but inside each of segments, people have similar demands[50].

Segmentation is all about predicting a client's category based on available information. Clustering and classification are machine learning techniques made for this purpose; we've briefly covered them in section 2.5 on data analytics. The Recency, Frequency, Monetary Value (RFM) [132] model has been widely used for segmentation[117]. The aim is to create a segmentation based on customers' purchase patterns; the three following RFM features are effective in characterizing the customer's behavior[15]: (1) recency is how recently the customer made its last purchase, (2) frequency measure how often he makes purchases, and (3) monetary value means how much the client is willing to spend on purchases. These features are usually simplified to take values from 1 to 5.

Many implementations use clustering approaches based on variations of standard algorithms such as k-means or hierarchical clustering[113] or DBSCAN[125]. Some of these variations make improvements based on the rough set theory [41], others progressively reduce the number of centroids to palliate the irregularities induced by their initial selection[52], others are specialized on specific subjects such as hotel customers [54].

Based on our research, classification approaches appear to be less common. Most likely, this is because it can be hard to associate labels with individuals, a priori, which is mandatory for supervised learning. Nevertheless, some implementations use naïve Bayes, logistic regression, SVM, decision tree, and decision table algorithms for segmentation. [29]

An RFM model is weak in predicting the segment of a new customer without any associated purchasing data. In these cases, models that use more traditional variables such as demographics or geographical data are preferable. [112, 102, 118] Deep learning methods like deep belief networks [95] and self-organizing neural networks [130] are also used for the purpose of segmentation. ANN-based methods typically use much more features and perform dimensionality reductions such as PCA before the segmentation. Association rule learning (see section 2.5) can also be used to identify relevant segments and discard all segments that don't contain enough individuals by fine-tuning the frequency threshold[34].

Some of the software platform that provides services for segmantation are Twilio segment, Marketo or Klaviyo. One example of use is Spotify that segment their users based on their demographics and the music they listen.

Created in 1985 by Alan Cooper a buyer persona is a "detailed description of imaginary people constructed out of well-understood, highly specified data about real people” [98]. They can be used as stand-alone but also as an extension of a segmentation. Personas are used to transform customer-centric data into a more understandable representation. It provides better performance metrics such as speed of tasks, effectiveness of effort, sales, advertising engagement[23] by helping to focus all efforts in one direction[98] and helping to answer empathetic questions (e.g. Why do they use our services or products ? Which services are important to the customers ?) [35]. Machine learning techniques can be used to automatically generate personas from collected data (e.g.social media data [22]). An example of an entire process of generation is described by M. Koponen[82]. He proposes an approach in three steps: a pre-processing of the data, a segmentation using unsupervised learning, and a conversion of the clusters into buyers’ personas via vector trees. Generative AI (see section 2.1) can also help to fasten the human process of creating personas. [61] Some websites (e.g. DelveAI, Hubspot, Userforge) provide tools to automatically generate personas.

2.7 Targeting decisions (31.7%)

In the STP framework, targeting follows segmentation. Its purpose is to identify the segments that are the most profitable for the company. The boundary between humans and computers is shifting from operational to strategic. AI can be more rational and gas faster decision-making speed, especially for new coming data[121]; however, it is notoriously worse at evaluating subjective criteria. As of today, AI is rather a support than a decision maker to make strategic decisions [43]. The relation seems to be heading towards a symbiosis in which decisions are made using both the analytical processing power of computers and the more creative, intuitive, and holistic approach of humans[75].

Humans can choose to use descriptive analytics (see section 2.5) to make important targeting decisions. However, for repetitive operational decisions, a fully automated AI decision system can be beneficial. The data used for training can be demographic, geographical, or purchase data. A standard approach is also to run a pilot experiment to collect relevant data[115].

One of the main decision-making problems of targeting decisions is to distinguish profitable customers from non-profitable ones. One of the solutions is to estimate a so-called "cutoff response rate (CRR)", which is the ratio between the cost of contacting a customer and the net profit of a standard sale. A customer is classified as a target if the estimated probability that he ends up making a sale is higher than the CRR. Estimations are made using predictive analytics (see section 2.5). If they are correct, it is profitable to invest in contacting the client as the contact cost is lower than the average money financial gain per customer[135]. The problem can also be solved using binary classification algorithms to separate targets from non-targets.

An extension of this problem is to choose which promotion plan is most fitting for a given client. Machine learning algorithms can be used to predict the profit generated by different promotion policies. D. Simester and A. Timoshenko have compared the performance of different families of algorithms for this issue. Model-driven methods such as Lasso and finite mixture models were less robust but outperformed distance-driven methods and classification methods on ideal data[115].

For all of these methods, the typical performance metrics we have seen for data analytics are relevant to determining how well the models fit the data. Additionally, gains tables and gains charts can be used to give more information on the performance of different levels of investment. They plot the difference between the predicted and the real response rates of customers for different prospect rates[135].

X Corps. propose an integrated toolkit to manage the targeting of ad campaigns on social media. Here the AI supports human decisions by providing analytics and then automatically distributing content to targets. The targeting can be either based on information directly provided by the customer such as age, gender, or location, or based on variables provided by machine learning algorithms such as the interests or the similarity to followers of specified accounts. X Corps. also allows retargeting of previously exposed customers to expose them to follow-up content[14].

2.8 Marketing automation (28%)

Marketing automation streamlines and automates redundant marketing processes in order to reduce human errors and save time. Instead of executing those processes manually, more time can be allocated to make strategic decisions and planning. Different areas can be automated such as email marketing, pricing, social media management, lead management, customer relationship management, programmatic advertising (see 2.4) and data analytics (see section 2.5)[12].

Email automation sends emails based on defined triggers. This can be as simple as sending an automatic response whenever an email is received during vacations, but it can also involve much more complex triggers based on a tree of possibilities. This allows one to send mail personalized to clients based on their actions or absence of actions. The AI essentially follows a strict set of rules to make email management easier.

Pricing automation optimizes the price of services or goods based on their costs, the market conditions, and the customer's behavior, so as to maximize profits. Other variables such as shipping costs, taxes, discounts, or the type of relation (e.g. B2B, B2C) can be taken into account[7]. Prices are changed dynamically according to machine learning algorithms. For example, one approach for B2B is to model the salespersons' pricing policies and then predict the optimal price based on regression analysis [80].

Social media management (SMM) is concerned with the quality, timing, and targeting of social media posts. AI can automate the process of skimming through social media posts and then provides capabilities. A. Micu et al. divide them into three clusters: audience, image, and sentiment analysis (see section 2.5). Audience analysis helps predict how customers will react to posts. It can identify the current stage of the customer buying cycle, identify which timing or medium is best to reach the maximum audience, recommend key opportunities to engage with customers and help generate buyers' personas (see section 2.6). Image analysis uses deep learning computer vision algorithms to make personalized recommendations based on what they post, recognize brand-related images, and identify common places and times where consumption is more likely.[38, 97].

Lead management (LM) is the "process of capturing leads, tracking their activities and behavior, qualifying them, nurturing them to make them sales-ready, and then passing them on to the sales team."[66] J. D'Haen and D. Van den Poel propose a machine learning model to improve the quality of prospects to make them have an easier time converting into leads. The model uses different classification methods and adapts itself based on a feedback loop[55]. Another use case is the use of data mining to classify leads[18]. LM also helps allocate more resources to high-potential leads by making predictions of leads' likelihood of conversion. After comparisons of machine learning techniques for such predictions, random forest is one of the best-performing algorithms [21, 96].

"**Customer relationship management (CRM)** is the strategic process of selecting customers that a firm can most profitably serve and shaping interactions between a company and these customers. The ultimate goal is to optimize the current and future value of customers for the company."[104] CRM can use the previous areas we just mentioned and leverage machine learning to make adaptations and improve long-term relationships.

First-generation CRM can be divided into **sales force automation (SFA)** and **customer service support (CSS)**.

SFA is "the application of information technology to support the sales function."[33] It provides a framework to deal with customer-related data collection, storage, analysis, and distribution to salespeople and managers. Some of the tasks involved are summarizing the customer's profile, sending alerts whenever some variables reach dangerous thresholds, managing the subscription live cycle, or highlighting important deals[8].

CSS is the assistance provided to the customer before and after the sales. Its main purpose is to answer to customers' questions. Error diagnosis or decision support can be enhance by data mining that will extract the most relevant information out of the data record on the customer[73]. Generative AI (see section 2.1) can be used to provide real time conversation support via chatbots.

Customer service is the support that organizations offer to customers before and after purchasing a product or service. In customer service, the organization's representative values both potential and existing customers equally. Customer service representatives are the main line of contact between an organization and its customers, making CX a critical facet and the main priority of customer service teams.

Based on their area of expertise, models will use different types of data. Email automation, SMM, or CSS will use communication data including email, social media posts, or prompted questions. Price automation and SFA will use transactional data but also market data, competitor profiles, or pricing schedules. Lead management will use lead-specific data. But all areas almost always profile data on the customers.

Based on the purpose of marketing automation, one central KPI for all areas is the amount of time freed up for the company. Pricing automation is additionally concerned with profits, SMM with the number of viewers and subscribers, lead management with the rentability of prospects and lead conversion rates, SFA again with profits and the number of sales, and CSS with customer satisfaction.

Some of the top companies providing marketing automation services are Siebel, Oracle, Adobe, and Salesforce. F. Buttle et al. proposed a comparison of SFA software[33]. For practical examples, most large companies (e.g. Spotify) will use some degree of email automation to increase customer retention. Ryan uses dynamic pricing to adjust the price of their low-cost carriers to increase revenues[49].

Conclusion

In this paper, we made a literature review of some of the artificial intelligence applications in marketing that are the most used and summarise their interaction in a diagram. We defined each application and listed its sub-application; then we explained how AI could be applied to it in practice. We also covered some concrete examples of use by companies.

We started the review with content creation that use generative AI to create either written, audio, or visual content. Content personalization uses deep learning to either adapt its content to the customers or make recommendations. The recommendations can either react to the user's inputs like a web browser or make passive recommendations. We also covered the five different stages of data analytics. The most used are descriptive which summarises the data and predictive analytics which forecasts the information based on historical data. Both use a wide variety of machine learning techniques; we listed each of those; then we further explained the most prominent algorithms used by those techniques. Data analytics are essential to enable customer segmentation and targeting. Customer segmentation is most commonly performed using clustering algorithms and the Recency, Frequency, and Monetary value data model; but classification algorithms and other data models are also used. Marketers can make targeting decisions with the help of descriptive analytics or use some automatic models that exploit predictive analytics. Marketing automation uses AI to replace human labor, and we covered the diverse areas that can be automatized. We explained how programmatic advertising automatizes the process of buying and selling media slots for advertisement.

To go even further, another paper could cover all the applications that are less used such as augmented and virtual reality, voice search optimization, autonomous object/systems, facial recognition, or biometrics and chipping.

Overall we provided an intuitive but practical review of the mechanism behind AI's usage in marketing.

Bibliography

[1] "A beginner's guide to programmatic advertising." [Online]. Available: https://advertising.amazon.com/en-us/blog/programmatic-advertising

[2] "The CMO Survey Fall 2023." [Online]. Available: https://cmosurvey.org/results/fall-2023/

[3] "The CMO Survey Spring 2024." [Online]. Available: https://cmosurvey.org/results/spring-2024/

[4] "Definitions of Marketing." [Online]. Available: https://www.ama.org/the-definition-of-marketing-what-is-marketing/

[5] "Kotler Marketing Group, Inc." [Online]. Available: https://www.kotlermarketing.com/phil_questions.shtml

[6] "Legal - Personal Data Use for Personalized User Experiences - Apple." [Online]. Available: https://www.apple.com/legal/privacy/en-ww/personalized-user-experiences/

[7] "Pricing Automation." [Online]. Available: https://dealhub.io/glossary/pricing-automation/

[8] "Sales force automation (SFA) defined." [Online]. Available: https://www.oracle.com/be/cx/sales/sales-force-automation/what-is-sfa/

[9] "The value of getting personalization right—or wrong—is multiplying | McKinsey." [Online]. Available: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying

[10] "What is AI (artificial intelligence)? | McKinsey." [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai

[11] "What is Content Marketing?" [Online]. Available: https://contentmarketinginstitute.com/what-is-content-marketing/

[12] "What is marketing automation?" [Online]. Available: https://www.oracle.com/be/cx/marketing/automation/what-is-marketing-automation/

[13] "What is personalization? | McKinsey." [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-personalization

[14] "X targeting in campaign management." [Online]. Available: https://developer.x.com/en/docs/twitter-ads-api/campaign-management/overview/targeting

[15] "The new rules of marketing: how to use one-to-one relationship marketing to be the leader in your industry," in *Choice Reviews Online*, vol. 35, Dec. 1997, pp. 35–2212–35–2212. [Online]. Available: http://choicereviews.org/review/10.5860/CHOICE.35-2212

[16] "Chapter 4 - Applications and Use Cases for Text Mining," in *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, G. Miner, D. Delen, J. Elder, A. Fast, T. Hill, and R. A. Nisbet, Eds. Boston: Academic Press, Jan. 2012, pp. 53–72. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123869791000049

[17] "The evolving role of artificial intelligence in marketing: A review and research agenda," *Journal of Business Research*, vol. 128, pp. 187–203, May 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0148296321000643

[18] "Lead management optimization using data mining: A case in the telecommunications sector," *Computers & Industrial Engineering*, vol. 154, p. 107122, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835221000267

[19] "What is Artificial Intelligence (AI)? | IBM," Aug. 2023. [Online]. Available: https://www.ibm.com/topics/artificial-intelligence

[20] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB '94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Sep. 1994, pp. 487–499.

[21] J. Aissam, H. Mustapha, and A. Hasbaoui, "Artificial intelligence-based lead propensity prediction," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, p. 1281, Sep. 2023.

[22] J. An, H. Cho, H. Kwak, M. Hassen, and J. Jansen, "Towards Automatic Persona Generation Using Social Media," Aug. 2016, pp. 206–211.

[23] J. An, H. Kwak, S.-g. Jung, J. Salminen, and B. J. Jansen, "Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data," *Social Network Analysis and Mining*, vol. 8, no. 1, p. 54, Aug. 2018. [Online]. Available: https://doi.org/10.1007/s13278-018-0531-0

[24] M. Architects, "Blog-How Annika Buys Media For High-Performing TV Campaigns." [Online]. Available: https://www.marketingarchitects.com/blog/how-annika-buys-media-for-a-high-performing-campaign

[25] R. Bellman and R. E. Bellman, *Dynamic Programming.* Princeton University Press, 1957.

[26] J. Berkson, "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, vol. 39, no. 227, pp. 357–365, 1944. [Online]. Available: https://www.jstor.org/stable/2280041

[27] J. Bodlaj and V. Batagelj, "Hierarchical link clustering algorithm in networks," *Physical Review E*, vol. 91, no. 6, p. 062814, Jun. 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.91.062814

[28] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, ser. COLT '92. New York, NY, USA: Association for Computing Machinery, Jul. 1992, pp. 144–152. [Online]. Available: https://dl.acm.org/doi/10.1145/130385.130401

[29] D. Bratina and A. Faganel, "Using Supervised Machine Learning Methods for RFM Segmentation: A Casino Direct Marketing Communication Case," *Market-Tržište*, vol. 35, no. 1, pp. 7–22, Jun. 2023. [Online]. Available: https://hrcak.srce.hr/303766

[30] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[31] M. Breunig, P. Kröger, R. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers." vol. 29, Jun. 2000, pp. 93–104.

[32] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, Apr. 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016975529800110X

[33] F. Buttle, L. Ang, and R. Iriana, "Sales force automation: Review, critique, research agenda," *International Journal of Management Reviews*, vol. 8, pp. 213–231, Dec. 2006.

[34] M. Böttcher, M. Spott, D. Nauck, and R. Kruse, "Mining changing customer segments in dynamic markets," *Expert Systems with Applications*, vol. 36, no. 1, pp. 155–164, Jan. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417407004393

[35] L. Caballero, A. Moreno, and A. Seffah, "Persona as a Tool to Involving Human in Agile Methods: Contributions from HCI and Marketing," Sep. 2014, pp. 283–290.

[36] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, *A Practical Guide to Sentiment Analysis*, Jan. 2017, vol. 5.

[37] J. Cannella, "Artificial Intelligence in Marketing," May 2018. [Online]. Available: https://www.semanticscholar.org/paper/Artificial-Intelligence-in-Marketing-Cannella/0f45a67d79df3c8a489dd75c1c5406ae3df5e242

[38] A. Capatina, A. Micu, and A.-E. Micu, "Exploring Artificial Intelligence Techniques' Applicability in Social Media Marketing," Sep. 2018.

[39] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue, "Large-scale validation and analysis of interleaved search evaluation," *ACM Transactions on Information Systems*, vol. 30, no. 1, pp. 1–41, Feb. 2012. [Online]. Available: https://dl.acm.org/doi/10.1145/2094072.2094078

[40] M. Chen and E. al., "Evaluating Large Language Models Trained on Code," Jul. 2021. [Online]. Available: http://arxiv.org/abs/2107.03374

[41] C.-H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, Part 1, pp. 4176–4184, Apr. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417408002091

[42] S. Chintalapati and S. K. Pandey, "Artificial intelligence in marketing: A systematic literature review," *International Journal of Market Research*, vol. 64, no. 1, pp. 38–68, Jan. 2022. [Online]. Available: https://doi.org/10.1177/14707853211018428

[43] M. Claude and D. Combe, "THE ROLES OF ARTIFICIAL INTELLIGENCE AND HUMANS IN DECISION MAKING: TOWARDS AUGMENTED HUMANS? : A focus on knowledge-intensive

firms." 2018. [Online]. Available: https://www.semanticscholar.org/paper/THE-ROLES-OF-ARTIFICIAL-INTELLIGENCE-AND-HUMANS-IN-Claude-Combe/a259956bfaa73a06945e53824edc7bcd80f1e43d

[44] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano, C. Hesse, and J. Schulman, "Training Verifiers to Solve Math Word Problems," Nov. 2021. [Online]. Available: http://arxiv.org/abs/2110.14168

[45] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. [Online]. Available: https://doi.org/10.1007/BF00994018

[46] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967. [Online]. Available: https://ieeexplore.ieee.org/document/1053964

[47] F. Crestani and H. Du, "Written versus spoken queries: A qualitative and quantitative comparative analysis," *Journal of the American Society for Information Science and Technology*, vol. 57, no. 7, pp. 881–890, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20350

[48] M. F. Dacrema, P. Cremonesi, and D. Jannach, "Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches," in *Proceedings of the 13th ACM Conference on Recommender Systems*, Sep. 2019, pp. 101–109. [Online]. Available: http://arxiv.org/abs/1907.06902

[49] T. D'Alfonso, P. Malighetti, and R. Redondi, "The pricing strategy of Ryanair," *Airline Industry: Strategies, Operations and Safety*, pp. 119–141, Jan. 2011.

[50] C. de Moerloose and J. Lambin, *Marketing stratégique et opérationnel, 10ème édition.* Dunod, 2021.

[51] V. Devang, S. Chintan, T. Gunjan, and K. Rai, "Applications of Artificial Intelligence in Marketing," *Annals of Dunarea de Jos University of Galati. Fascicle I. Economics and Applied Informatics*, vol. 25, pp. 28–36, Apr. 2019.

[52] P. Dhandayudam and I. Krishnamurthi, "AN IMPROVED CLUSTERING ALGORITHM FOR CUSTOMER SEGMENTATION," *International Journal of Engineering Science and Technology*, vol. 4, Feb. 2012.

[53] E. P. K. Dr. N.Thilagavathy, "Artificial Intelligence on digital Marketing-An overview," *NVEO - NATURAL VOLATILES & ESSENTIAL OILS*

Journal / NVEO, pp. 9895–9908, Dec. 2021. [Online]. Available: https://www.nveo.org/index.php/journal/article/view/2772

[54] A. Dursun and M. Caber, "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis," *Tourism Management Perspectives*, vol. 18, pp. 153–160, Apr. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211973616300149

[55] J. D'Haen and D. Van den Poel, "Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework," *Industrial Marketing Management*, vol. 42, no. 4, pp. 544–551, May 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019850113000564

[56] W. Ertel, *Introduction to Artificial Intelligence.* Springer, Jan. 2018.

[57] J. H. Fetzer, "What is Artificial Intelligence?" in *Artificial Intelligence: Its Scope and Limits*, J. H. Fetzer, Ed. Dordrecht: Springer Netherlands, 1990, pp. 3–27. [Online]. Available: https://doi.org/10.1007/978-94-009-1900-6_1

[58] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x

[59] E. Forgy, "Cluster analysis of multivariate data : efficiency versus interpretability of classifications," *Biometrics*, 1965. [Online]. Available: https://www.semanticscholar.org/paper/Cluster-analysis-of-multivariate-data-%3A-efficiency-Forgy/5c4feeae0d911e30866b7149c1195cd8c007199b

[60] Y. Freund and R. E. Schapire, "Large Margin Classification Using the Perceptron Algorithm," *Machine Learning*, vol. 37, no. 3, pp. 277–296, Dec. 1999. [Online]. Available: https://doi.org/10.1023/A:1007662407062

[61] T. Goel, O. Shaer, C. Delcourt, Q. Gu, and A. Cooper, "Preparing Future Designers for Human-AI Collaboration in Persona Creation," in *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, ser. CHIWORK '23. New York, NY, USA: Association for Computing Machinery, Sep. 2023, pp. 1–14. [Online]. Available: https://dl.acm.org/doi/10.1145/3596671.3598574

[62] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," Jun. 2014. [Online]. Available: http://arxiv.org/abs/1406.2661

[63] B. Gregg, H. Kalaoui, J. Maynes, and G. Schuler, "Marketing's Holy Grail: Digital personalization at scale | McKinsey." [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/marketings-holy-grail-digital-personalization-at-scale

[64] P. Gupta, T. Dreossi, J. Bakus, Y.-H. Lin, and V. Salaka, "Treating Cold Start in Product Search by Priors," in *Companion Proceedings of the Web Conference 2020*. Taipei Taiwan: ACM, Apr. 2020, pp. 77–78. [Online]. Available: https://dl.acm.org/doi/10.1145/3366424.3382705

[65] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '00. New York, NY, USA: Association for Computing Machinery, May 2000, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/10.1145/342009.335372

[66] I. Hansen, J. Poulter, N. Elkin, and C. Ferguson, "Gartner Magic Quadrant for CRM Lead Management," Aug. 2020. [Online]. Available: https://www.gartner.com/en/documents/3989183

[67] A. Hassan, "The Usage of Artificial Intelligence in Digital Marketing: A Review," in *Applications of Artificial Intelligence in Business, Education and Healthcare*, A. Hamdan, A. E. Hassanien, R. Khamis, B. Alareeni, A. Razzaque, and B. Awwad, Eds. Cham: Springer International Publishing, 2021, pp. 357–383. [Online]. Available: https://doi.org/10.1007/978-3-030-72080-3_20

[68] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, "Measuring Massive Multitask Language Understanding," Jan. 2021. [Online]. Available: http://arxiv.org/abs/2009.03300

[69] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Jan. 2018. [Online]. Available: http://arxiv.org/abs/1706.08500

[70] G. E. Hinton and S. Roweis, "Stochastic Neighbor Embedding," in *Advances in Neural Information Processing Systems*, vol. 15. MIT Press, 2002. [Online]. Available: https://papers.nips.cc/paper_files/paper/2002/hash/6150ccc6069bea6b5716254057a194ef-Abstract.html

[71] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, Aug. 1995, pp. 278–282 vol.1. [Online]. Available: https://ieeexplore.ieee.org/document/598994

[72] M.-H. Huang and R. T. Rust, "A strategic framework for artificial intelligence in marketing," *Journal of the Academy of Marketing Science*, vol. 49, no. 1, pp. 30–50, Jan. 2021. [Online]. Available: https://doi.org/10.1007/s11747-020-00749-9

[73] S. C. Hui and G. Jha, "Data mining for customer service support," *Information & Management*, vol. 38, no. 1, pp. 1–13, Oct. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378720600000513

[74] N. Japkowicz, C. Myers, and M. Gluck, "A Novelty Detection Approach to Classification," Aug. 1995. [Online]. Available: https://www.semanticscholar.org/paper/A-Novelty-Detection-Approach-to-Classification-Japkowicz-Myers/22c0bf46726ad1c90b5fac7f16638c813e86c829

[75] M. H. Jarrahi, "Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making," *Business Horizons*, vol. 61, Jul. 2018.

[76] O. Jeunen, I. Potapov, and A. Ustimenko, "On (Normalised) Discounted Cumulative Gain as an Off-Policy Evaluation Metric for Top-n Recommendation," Nov. 2023. [Online]. Available: http://arxiv.org/abs/2307.15053

[77] S. Ji and Y. Xie, "Logistic Regression: From Binary to Multi-Class."

[78] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, "Mistral 7B," Oct. 2023. [Online]. Available: http://arxiv.org/abs/2310.06825

[79] T. Joachims, "Optimizing search engines using clickthrough data," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '02. New York, NY, USA: Association for Computing Machinery, Jul. 2002, pp. 133–142. [Online]. Available: https://doi.org/10.1145/775047.775067

[80] Y. Karlinsky Shichor, "Automation, Decision Making and Business to Business Pricing," Ph.D. dissertation, Columbia University, 2018. [Online]. Available: https://doi.org/10.7916/D83X9Q5M

[81] K. Kilgour, M. Zuluaga, D. Roblek, and M. Sharifi, "Fr\'echet Audio Distance: A Metric for Evaluating Music Enhancement Algorithms," Jan. 2019. [Online]. Available: http://arxiv.org/abs/1812.08466

[82] M. Koponen, "Developing Marketing Personas with Machine Learning for Educational Program Finder," 2017. [Online]. Available: https://trepo.tuni.fi/handle/123456789/25419

[83] P. Kotler, “Reconceptualizing marketing: An interview with Philip Kotler,” *European Management Journal*, vol. 12, no. 4, pp. 353–361, Dec. 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0263237394900213

[84] M. A. Kramer, “Nonlinear principal component analysis using autoassociative neural networks,” *AIChE Journal*, vol. 37, no. 2, pp. 233–243, 1991. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aic.690370209

[85] K. Król and D. Zdonek, “Analytics Maturity Models: An Overview,” *Information*, vol. 11, p. 142, Mar. 2020.

[86] T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M.-W. Chang, A. M. Dai, J. Uszkoreit, Q. Le, and S. Petrov, “Natural Questions: A Benchmark for Question Answering Research,” *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 453–466, Aug. 2019. [Online]. Available: https://doi.org/10.1162/tacl_a_00276

[87] S. Lamkhede and S. Das, “Challenges in Search on Streaming Services: Netflix Case Study,” Mar. 2019. [Online]. Available: http://arxiv.org/abs/1903.04638

[88] J. Langford and T. Zhang, “The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information,” in *Advances in Neural Information Processing Systems*, vol. 20. Curran Associates, Inc., 2007. [Online]. Available: https://proceedings.neurips.cc/paper/2007/hash/4b04a686b0ad13dce35fa99fa4161c65-Abstract.html

[89] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 413–422. [Online]. Available: https://ieeexplore.ieee.org/document/4781136

[90] S. Lloyd, “Least squares quantization in PCM,” *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982. [Online]. Available: https://ieeexplore.ieee.org/document/1056489

[91] M. Maalouf, “Logistic regression in data analysis: An overview,” *International Journal of Data Analysis Techniques and Strategies*, vol. 3, pp. 281–299, Jul. 2011.

[92] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955,” *AI Magazine*, vol. 27, no. 4, pp. 12–12, Dec. 2006. [Online]. Available: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904

.

[93] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943. [Online]. Available: https://doi.org/10.1007/BF02478259

[94] P. Nigam, Y. Song, V. Mohan, V. Lakshman, Weitian, Ding, A. Shingavi, C. H. Teo, H. Gu, and B. Yin, "Semantic Product Search," Jul. 2019. [Online]. Available: http://arxiv.org/abs/1907.00937

[95] M. Nilashi, S. Samad, B. Minaei-Bidgoli, F. Ghabban, and E. Supriyanto, "Online Reviews Analysis for Customer Segmentation through Dimensionality Reduction and Deep Learning Techniques," *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8697–8709, Sep. 2021. [Online]. Available: https://doi.org/10.1007/s13369-021-05638-z

[96] R. Nygård and J. Mezei, "Automating Lead Scoring with Machine Learning: An Experimental Study," 2020. [Online]. Available: https://hdl.handle.net/10125/63916

[97] E. Perakakis, G. Mastorakis, and I. Kopanakis, "Social Media Monitoring: An Innovative Intelligent Approach," *Designs*, 2019. [Online]. Available: https://www.semanticscholar.org/paper/Social-Media-Monitoring%3A-An-Innovative-Intelligent-Perakakis-Mastorakis/a5c40679f59e9d61b3f5418bb575c3122c178f92

[98] J. Pruitt and T. Adlin, "The Persona Lifecycle," *The Persona Lifecycle*, Jan. 2006.

[99] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986. [Online]. Available: https://doi.org/10.1007/BF00116251

[100] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," Feb. 2021. [Online]. Available: http://arxiv.org/abs/2103.00020

[101] F. Radlinski, M. Kurup, and T. Joachims, "How does clickthrough data reflect retrieval quality?" in *Proceedings of the 17th ACM conference on Information and knowledge management*. Napa Valley California USA: ACM, Oct. 2008, pp. 43–52. [Online]. Available: https://dl.acm.org/doi/10.1145/1458082.1458092

[102] P. E. Ramírez-Correa, F. J. Rondán-Cataluña, J. Arenas-Gaitán, E. E. Grandón, J. L. Alfaro-Pérez, and M. Ramírez-Santana, "Segmentation of Older Adults in the Acceptance of Social Networking Sites Using Machine Learning," *Frontiers in*

Psychology, vol. 12, Aug. 2021. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.705715/full

[103] D. Ratajczak, M. Kropp, S. Palumbo, N. de Bellefonds, J. Apotheker, S. Willersdorf, and G. Paizanis, "How CMOs Are Succeeding with Generative AI," Jun. 2023. [Online]. Available: https://www.bcg.com/publications/2023/generative-ai-in-marketing

[104] W. Reinartz and V. Kumar, *Customer Relationship Management: Concept, Strategy, and Tools*, Oct. 2012.

[105] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

[106] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, 4th US ed.* Pearson, Apr. 2020. [Online]. Available: https://aima.cs.berkeley.edu/

[107] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 169–194, Jun. 1998. [Online]. Available: https://doi.org/10.1023/A:1009745219419

[108] D. Schiessl, H. B. A. Dias, and J. C. Korelo, "Artificial intelligence in marketing: a network analysis and future agenda," *Journal of Marketing Analytics*, vol. 10, no. 3, pp. 207–218, Sep. 2022. [Online]. Available: https://doi.org/10.1057/s41270-021-00143-6

[109] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998. [Online]. Available: https://doi.org/10.1162/089976698300017467

[110] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support Vector Method for Novelty Detection," in *Advances in Neural Information Processing Systems*, vol. 12. MIT Press, 1999. [Online]. Available: https://papers.nips.cc/paper_files/paper/1999/hash/8725fb777f25776ffa9076e44fcfd776-Abstract.html

[111] J. R. Searle, "Minds, brains, and programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–424, Sep. 1980. [Online]. Available: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A

[112] E. H. Sharaf Addin, N. Admodisastro, S. N. S. Mohd Ashri, A. Kamaruddin, and Y. C. Chong, “Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning,” *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2009223, Dec. 2022. [Online]. Available: https://doi.org/10.1080/08839514.2021.2009223

[113] S. H. Shihab, S. Afroge, and S. Z. Mishu, “RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study,” in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2019, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/8679376

[114] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, “Analysis of a very large web search engine query log,” *ACM SIGIR Forum*, vol. 33, no. 1, pp. 6–12, Sep. 1999. [Online]. Available: https://dl.acm.org/doi/10.1145/331403.331405

[115] D. Simester, A. Timoshenko, and S. I. Zoumpoulis, “Targeting Prospective Customers: Robustness of Machine-Learning Methods to Typical Data Challenges,” *Management Science*, Nov. 2019. [Online]. Available: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2019.3308

[116] W. R. Smith, “Product Differentiation and Market Segmentation as Alternative Marketing Strategies,” *Journal of Marketing*, vol. 21, no. 1, pp. 3–8, Jul. 1956. [Online]. Available: https://doi.org/10.1177/002224295602100102

[117] P. Spring, P. S. Leeflang, and T. Wansbeek, “The Combination Strategy to Optimal Target Selection and Offer Segmentation in Direct Mail,” *Journal of Market-Focused Management*, vol. 4, no. 3, pp. 187–203, Oct. 1999. [Online]. Available: https://doi.org/10.1023/A:1009899802421

[118] A. Srivihok and W. Yotsawat, “Market Segmentation of Inbound Business Tourists to Thailand by Binding of Unsupervised and Supervised Learning Techniques,” *Journal of Software*, vol. 9, May 2014.

[119] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico, “Deep Learning for Recommender Systems: A Netflix Case Study,” *AI Magazine*, vol. 42, no. 3, pp. 7–18, Nov. 2021. [Online]. Available: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/18140

[120] J. Sterne, *Artificial Intelligence for Marketing: Practical Applications*. John Wiley & Sons, Aug. 2017.

[121] M. Stone, E. Aravopoulou, Y. Ekinci, G. Evans, M. Hobbs, A. Labib, P. Laughlin, J. Machtynger, and L. Machtynger, “Artificial intelligence (AI) in strategic

marketing decision-making: a research agenda," *The Bottom Line*, vol. 33, no. 2, pp. 183–200, Jan. 2020. [Online]. Available: https://doi.org/10.1108/BL-03-2020-0022

[122] D. Tax and R. Duin, "Support Vector Data Description," *Machine Learning*, vol. 54, pp. 45–66, Jan. 2004.

[123] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *Ai Communications*, vol. 30, pp. 169–190,, May 2017.

[124] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: https://www.jstor.org/stable/2346178

[125] A. Ullah, M. I. Mohmand, H. Hussain, S. Johar, I. Khan, S. Ahmad, H. A. Mahmoud, and S. Huda, "Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation Using Recency, Frequency, Monetary, and Time," *Sensors (Basel, Switzerland)*, vol. 23, no. 6, p. 3180, Mar. 2023.

[126] T. Unterthiner, S. van Steenkiste, K. Kurach, R. Marinier, M. Michalski, and S. Gelly, "Towards Accurate Generative Models of Video: A New Metric & Challenges," Mar. 2019. [Online]. Available: http://arxiv.org/abs/1812.01717

[127] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Aug. 2023. [Online]. Available: http://arxiv.org/abs/1706.03762

[128] N. Venkat, *The Curse of Dimensionality: Inside Out*, Sep. 2018.

[129] S. Verma, R. Sharma, S. Deb, and D. Maitra, "Artificial intelligence in marketing: Systematic review and future research direction," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100002, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667096820300021

[130] C. Wang, "Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach," *Information Processing & Management*, vol. 59, no. 6, p. 103085, Nov. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457322001868

[131] P. Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence*, vol. 10, no. 2, pp. 1–37, Jan. 2019. [Online]. Available: https://sciendo.com/article/10.2478/jagi-2019-0002

[132] J.-T. Wei, S.-Y. Lin, and H.-H. Wu, "A review of the application of RFM model," *African Journal of Business Management December Special Review*, vol. 4, pp. 4199–4206, Jan. 2010.

[133] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, Aug. 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0169743987800849

[134] M. Woodroofe, "A One-Armed Bandit Problem with a Concomitant Variable," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 799–806, 1979. [Online]. Available: https://www.jstor.org/stable/2286402

[135] J. Zahavi, "Predictive Analytics for Targeting Decisions," in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, L. Rokach, O. Maimon, and E. Shmueli, Eds. Cham: Springer International Publishing, 2023, pp. 751–777. [Online]. Available: https://doi.org/10.1007/978-3-031-24628-9_33

[136] M. Zaki, "Scalable algorithms for association mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, May 2000. [Online]. Available: https://ieeexplore.ieee.org/document/846291

[137] R. Zellers, A. Holtzman, Y. Bisk, A. Farhadi, and Y. Choi, "HellaSwag: Can a Machine Really Finish Your Sentence?" May 2019. [Online]. Available: http://arxiv.org/abs/1905.07830

[138] M. Zhao, J. Chen, and Y. Li, "A Review of Anomaly Detection Techniques Based on Nearest Neighbor," Jan. 2018.

Abstract :

Artificial intelligence is playing an increasingly important role in marketing. In this paper, we summarise the literature on the use of AI in marketing. We review some of the most widespread applications of AI to enhance marketing. We provide an intuitive explanation on how the power of AI is leveraged in each of these applications. We also go over some practical information including the most relevant data to collect and the metrics that can be used to evaluate the AI performances. We cover some real-life examples of companies that have used these applications. Finally, we propose a diagram representing all these applications, their sub-applications, and the interaction between them.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm