Information Systems Frontiers (2023) 25:1621–1638
https://doi.org/10.1007/s10796-022-10314-0

How to Maximize Clicks for Display Advertisement in Digital Marketing? A Reinforcement Learning Approach

Vinay Singh[1,2] · Brijesh Nanavati[3] · Arpan Kumar Kar[4] · Agam Gupta[4]

Accepted: 9 July 2022 / Published online: 21 July 2022

Abstract
One of the core challenges in digital marketing is that the business conditions continuously change, which impacts the reception of campaigns. A winning campaign strategy can become unfavored over time, while an old strategy can gain new traction. In data driven digital marketing and web analytics, A/B testing is the prevalent method of comparing digital campaigns, choosing the winning ad, and deciding targeting strategy. A/B testing is suitable when testing variations on similar solutions and having one or more metrics that are clear indicators of success or failure. However, when faced with a complex problem or working on future topics, A/B testing fails to deliver and achieving long-term impact from experimentation is demanding and resource intensive. This study proposes a reinforcement learning based model and demonstrates its application to digital marketing campaigns. We argue and validate with actual-world data that reinforcement learning can help overcome some of the critical challenges that A/B testing, and popular Machine Learning methods currently used in digital marketing campaigns face. We demonstrate the effectiveness of the proposed technique on real actual data for a digital marketing campaign collected from a firm.

Keywords Digital marketing · Computational advertising · Reinforcement learning · Upper confidence bound (UCB) algorithm · Big data analytics · Machine learning · Marketing analytics

1 Introduction

In the ever-changing landscape of Digital Marketing (DM), it can be hard to examine what works truly and, more importantly, what does not (Barone et al., 2007). As said by John Wanamaker – "Half the money I spend on advertising is wasted. The trouble is that I don't know which half."(Kohavi & Thomke, 2017). Every firm aspires and competes to be fully digital with limited marketing spend and stringent deadlines to realize the tangible dollar impact. Therefore, it is imperative to shift from a conventional test and target strategy to an algorithmic approach which is scalable and has higher reliability in terms of replicability of outcome (Boone & Roehm, 2002; Gordini & Veglio, 2017). Never have marketers had access to more customer data but leveraging upon this data is extremely challenging computationally. For example, user profiles can go beyond primary name and demographic data to include device preferences, social posts, browsing and content history, hobbies, interests, and much more. In theory, marketers should have a near-perfect understanding of their customer's needs, which customers to target, and the best ways to engage with them (Davenport et al., 2011). Artificial Intelligence (AI) is often used to tackle DM's complex data driven challenges in DM through different applications like chatbots and marketing automation (Erevelles et al., 2016; Kar & Kushwaha, 2021; Kushwaha & Kar, 2021). Reviews of AI applications in DM indicates that big data analytics is extensively used for optimizing marketing outcomes in DM projects (Kushwaha et al., 2021; Verma et al., 2021). Lately, the DM industry has been employing AI to boost customer engagement through

✉ Arpan Kumar Kar
arpan.kumar.kar@gmail.com

[1] BASF SE, Pfalzgrafenstraße 1, 67061 Ludwigshafen am Rhein, Germany

[2] University of Siegen, Adolf-Reichwein-Straße 2, 57076 Siegen, Germany

[3] Customer Credit Risk Management, BASF Services Europe GmbH, GBE/SO-AET-OBC3, Rotherstrasse 11, 10245 Berlin, Germany

[4] Department of Management Studies and School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India

personalization and marketing journeys (Choi et al., 2020; Goldfarb & Tucker, 2011). AI has been an essential tool for marketers to capture campaign data and turn it into experiences that maximize consumer happiness and company profits (Du et al., 2021; Netzer et al., 2008). For example, combining AI with natural language processing, firms increase brand engagement and content consumption via algorithms that automatically analyze a brand's content assets (Kushwaha & Kar, 2021). In the industry 4.0 era, all functional activities including marketing need to leverage AI for enhancing outcome quality and impacts (Huber et al., 2022).

However, when dealing with advertisement optimization. Digital marketers often resort to randomized trials, the default option to determine whether potential improvements of an alternative method (e.g., website design for a tech company or medication in clinical trials for pharmaceutical companies) are significant compared to a well-established default. It is often colloquially referred to as A/B testing or A/B/n testing for several alternatives in the applied domain. The standard practice is to divert a small amount of the traffic or patients to the alternative and control. If an option appears to be significantly better, it is implemented; otherwise, the default setting is maintained. For example, in Web Analytics and DM, A/B testing is the prevalent method of comparing digital campaigns, choosing the winning advertisement, and deciding targeting strategy (Gallo, 2017; Senz, 2021). Though the designing and implementation of A/B testing are simple, however this hypothesis based classical approach has few limitations, as highlighted multiple times in existing literature:

- Meager conversion rate: Digital advertisements have meager conversion rates. Our case study has a subtle 0.2% difference between the worst and best-performing advertisements. However, at scale, this difference can have a significant impact. The problem with A/B/n testing is that it's inefficient at finding these differences (Ascarza, 2021). It treats all the advertisements equally, and one may need to run each advertisements tens of thousands of times until it can be discovered at a reliable confidence level.
- No memory: Many people think that one can get away with a single A/B test. One should be continuously testing to optimize your marketing and advertising creativity for your audience. Knowledge gained is not linked, and every test is an independent test every time. A previous knowledge (prior probability) generally strengthens the confidence of posterior probability and hence the decision-making process (Bojinov et al., 2021).
- Time and resource-heavy: The traditional approaches are a two-step process. Through a test campaign, it explores the opportunities, and then based on the comparative analysis, it exploits the winning digital advertisement or creative to reap the $ benefit (Fabijan et al., 2018; Kohavi & Longbotham, 2017). As a result, it makes the whole test very slow and expensive.
- No Self Learning or guiding principle: There is no self-learning and feedback system to update outcomes continually. One cannot be sure; just because creative-X performed better over creative-Y one year ago does not mean it will still perform better now (Bojinov et al., 2021).
- A/B/n only works for specific goals: It is ideal if we want to solve one dilemma; for example, which product page gives me the best result? However, pure A/B/n testing won't provide those answers if the goal is less easy to measure.

Stemming from the limitations of A/B/n testing above as indicated in existing literature, in this study, we are guided by the following research question(s):

RQ1. How can we use AI algorithms to better optimize our digital marketing campaign?
RQ2: How can we use AI algorithms adapt to changing patterns of customer preferences to better optimize our digital marketing campaign?

This study is structured into 6 sections. Section 2 summarizes the relevant literature. In the third section, we describe our methodology to overcome the limitations of A/B/n testing. Section 4, describes the data used for our experiments, followed by results with randomized trials. In Section 5, we discuss our findings, directions for further research, and the limitations of our study. Finally, we conclude the study in Section 6.

2 Background Literature

Since we aim to use RL for optimizing DM campaigns, we will first analyses studies from DM literature on advertisements optimization and then review related literature from the RL and its use in Marketing.

2.1 The Role of AI in DM

DM conceptualizes marketing on electronic platforms through any technological device (American Marketing Association, 2021). With newspaper circulations in 2018 falling to their lowest level since 1940, the decline of the newspaper industry also increased the customer affinity for online advertisement and marketing. There are currently several literature reviews related to DM and its benefits for organization. For example, Lamberton & Stephen, 2016; and Luo et al., 2013; offer elaborated reviews of social media

literature in marketing. Recent works also investigate more specific aspects of digital and interactive marketing, for example, its use in Business-to-Business strategy (B2B) (Pandey et al., 2020), its application and relevance to marketing analytics (Iacobucci et al., 2019), how it relates to DM communication (Kim et al., 2021).

The future of DM also depends on the ability of marketing professionals in applying AI techniques to effectively implement DM strategies (Ruiz-Real et al., 2021). AI techniques can bring out hidden business intelligence from the given consumer data which streamlines complex marketing problems. It is found that 90% of the sales professional expected a substantial impact of AI on sales marketing (Nadkarni & Prügl, 2021). The use of AI has improved the quality of DM initiatives (Ruiz-Real et al., 2021). AI based marketing models can mine hidden signals from buying patterns of the targeted customers leading to promotions teams driving computational benefits to businesses (Saura, 2021) and consequently assure increased productivity and a better understanding of customers segments. Computational marketing strategies based on AI can streamline the market, optimizing both the business profit and satisfaction of user experience. In social media communications in DM, AI has been extensively used for preserving sanctity of communications and reducing spam (Aswani et al., 2018) and further AI applications for DM campaigns need to demonstrate responsibility surrounding ethical concerns (Liu et al., 2021). Despite enhanced DM strategies in place, their efficiencies are to be improved using contemporary technologies such as AI to understand emotions, behavior, respond to human customer's queries (S.-S. Chen et al., 2021; S.-Y. Chen et al., 2021), traceability of actions (Jain et al., 2021) and provide competitive edge (Miklosik et al., 2019).

One aspect of DM which may significantly benefit from the AI techniques is the design, delivery, and optimization of ads to prospective customers. Solutions for optimizing the delivery of ads have been proposed since the early days of digital advertising, for example (Karuga et al., 2001). More recently, AI based techniques have been used to remove ineffective ads (Wang & Hong, 2019), optimal delivery of banner ads (Obal & Lv, 2017), optimize exposure of ads (Stourm & Bax, 2017), drafting better text ads in the case of epidemic response by public health authorities (Youngmann et al., 2021). However, in most cases A/B split testing is the de facto for optimization and selection of ads/websites/landing pages that perform better (Javanmard & Montanari, 2018). Under this technique, one sends roughly 50% of the traffic to the control and 50% to variation or significant amount of data over which estimation could be done. The test is run until it is valid, and then the decision is made to implement the winning variation (Gallo, 2017). Post the selection of winner, all users are sent to the more successful version of the site (Gallo, 2017) entering a period of pure exploitation.

2.1.1 Limitations of Current Approaches – A/B/n Testing

With A/B/n the objective is to allocate more traffic to a better performing digital asset (website/ad/landing page). However, A/B/n testing frameworks have the following three limitations. First, typically it splits the traffic uniformly over alternatives. Adaptive techniques should help to detect better options faster (Basse & Airoldi, 2018; Javanmard & Montanari, 2018) more so because the lack of sufficient evidence or a minor improvement of the metric may make it undesirable from a practical or financial perspective to replace the default. Second, companies often wish to monitor an ongoing A/B test continuously. Based on the performance, they may adjust their termination criteria as time progresses and possibly terminate earlier or later than initially intended. However, this practice may result in many more wrong conclusions if not adequately accounted for. It also results in, as one of the reasons, for the lack of reproducibility of marketing results (Basse & Airoldi, 2018; Javanmard & Montanari, 2018). Third, the opportunity cost for Short term campaigns and promotion is very high for the A/B test. For example, if you're running tests on an eCommerce site for Black Friday, an A/B test isn't that practical—you might only be confident in the result at the end of the day (Bojinov et al., 2021).

2.2 RL as a Substitute to A/B/n Testing

RL differs from the more widely studied problem of supervised learning in AI (Botvinick et al., 2019; Sutton & Barto, 2018). The most important difference is that there is no presentation of input/output pairs. Instead, after choosing an action, the agent is told the immediate reward and the subsequent state but is not told which action would have been in its best long-term interests (Jang et al., 2019). It is necessary for the agent to gather useful experience about the possible system states, actions, transitions, and rewards actively to act optimally. Another difference from supervised learning is that on-line performance is important: the evaluation of the system is often concurrent with learning.

Some aspects of RL are closely related to search and planning issues in AI (Gershman & Daw, 2017; Gupta et al., 2020) and therefore modes well with its use for optimizing DM campaigns. AI search algorithms generate a satisfactory trajectory through a graph of states. Planning operates similarly, but typically within a construct with more complexity than a graph, in which states are represented by compositions of logical expressions instead of atomic symbols. These AI algorithms are less general than the reinforcement-learning methods, in that they require a predefined model of state transitions, and with a few exceptions assume determinism. On the other hand, RL, at least in the kind of discrete cases

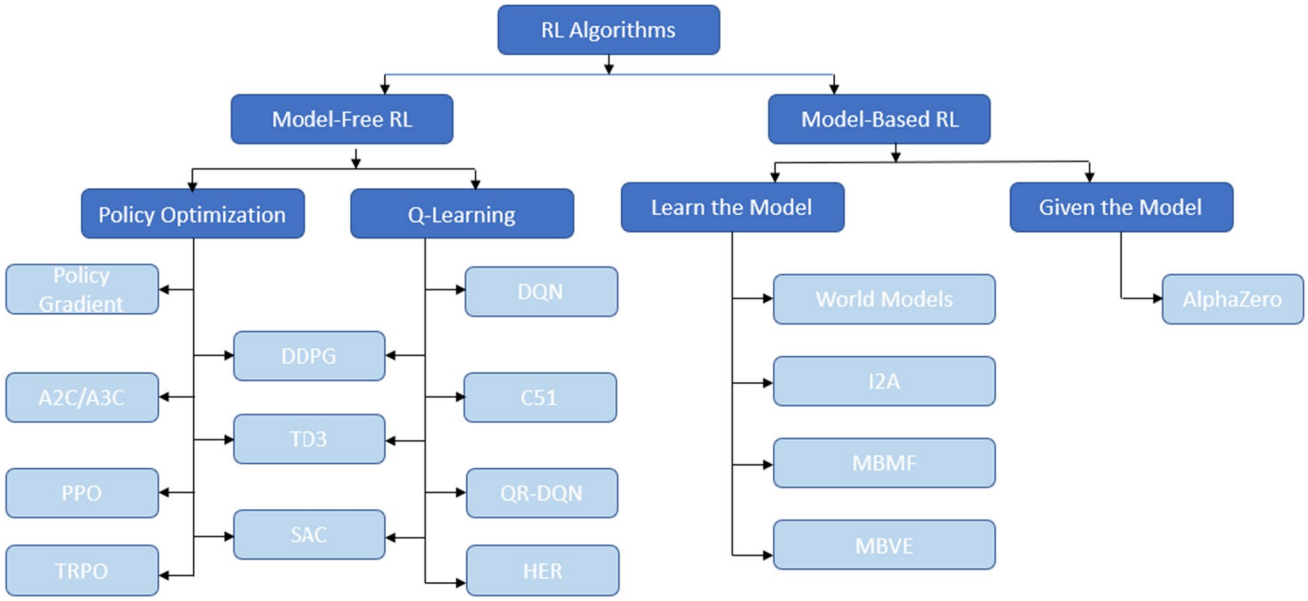

Fig. 1 Select taxonomy of dominant algorithms in Modern RL (combining Lillicrap et al., 2015; Mnih et al., 2016; Fujimoto et al., 2018; Haarnoja et al., 2018)

for which theory has been developed, assumes that the entire state space can be enumerated and stored in memory - an assumption to which conventional search algorithms are not tied (Sutton & Barto, 2018). Broadly there are two main branching points in an RL algorithm – ***whether the agent has access to (or learns) a model of the environment*** and second being ***what to learn***. Based on these two questions, the below Fig. 1 shows a taxonomy of algorithms in modern RL (Lillicrap et al., 2015; Mnih et al., 2016; Fujimoto et al., 2018; Haarnoja et al., 2018).

RL algorithms are generally classified into two categories: model-free RL (MFRL) and model based RL (MBRL). One way to understand the difference between the two is using the idea of Habit and Goal-directed behavior. While Habits are triggered by stimuli and then performed almost automatically, Goal-directed behavior is more purposeful, that is, controlled by knowledge of the value of goals (Sutton & Barto, 2018). In MFRL, the algorithm estimates the optimal policy without using or estimating the dynamics (transition and reward functions) of the environment whereas, as MBRL algorithm uses the transition function (and the reward function) in order to estimate the optimal policy.

We may broadly categorize Model-free algorithms into two. Value-based methods such as Q Learning and policy-based methods. While policy-based methods directly try to maximize the expected return by taking small steps in the direction of the policy gradient, Q learning approach tries to learn a Q-function that satisfies the Bellman (Optimality) Equation (Rathore et al., 2021). In addition, policies may be deterministic or stochastic. Deterministic policy maps state to action without uncertainty. Stochastic policy outputs a probability distribution over actions in each state. This process is called the Partially Observable Markov Decision Process (POMDP). Policy Optimization can be further studied under categories like Policy gradient (PG), Asynchronous Advantage Actor-Critic (A3C); Trust Region Policy Optimization (TRPO), and Proximal Policy Optimization (PPO), each differing based on the stability of the Actor training by limiting the policy update at each training step. Q-learning learns the action-value function: how good to act at a particular state. Q -learning can be further sub-divided (not limited to) into Deep Q-neural network (DQN), C51, Distributional Reinforcement Learning with Quantile Regression (QR-DQN), Hindsight Experience Replay (HER).

Model-based RL asks the questions of the form "*"what will happen if I do x?" to choose the best x"*.[1] Thus, model based RL tries to predict the environment to choose the optimal actions (Singh et al., 2022). They can be further subdivided into two approaches -Learn the model and Learn given the given model. A base policy is run to learn the model, like a random or any educated policy, while the trajectory is observed. Then, the model is fitted using the sampled data. However, these models may suffer from a bias problem (S.-S. Chen et al., 2021; S.-Y. Chen et al., 2021). They are further divided into world models,

[1] https://bair.berkeley.edu/blog/2019/12/12/mbpo/#fn:naming-conventions

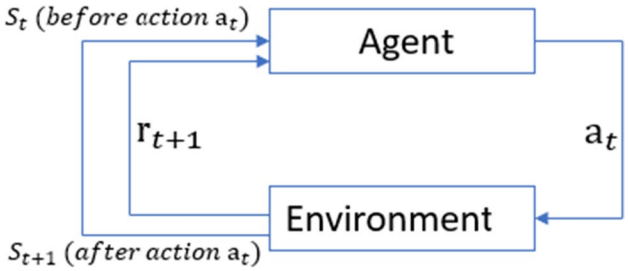

Fig. 2 The standard RL model. Where a_t is the action taken by the agent at time t, s_t is the state of the environment at time t, and r_{t+1} is the reward at time t + 1

Imagination-augmented agents (I2A), Model-Based Priors for Model-Free Reinforcement Learning (MBMF), and Model-Based Value Expansion (MBVE). These models bridge the gap between model based RL and model free RL algorithms.

In the RL approach, instead of two distinct periods of pure exploration and exploitation, the test is adaptive and simultaneously includes exploration and exploitation. The essence, the difference between the two approaches is how they deal with the explore-exploit dilemma. RL is the problem an agent faces that must learn behavior through trial-and-error interactions with a dynamic environment (Sutton & Barto, 2018) as shown in Fig. 2.

The agent is the learner who interacts with the environment, making decisions according to its observations made from it. The environment is every external condition that the agent cannot modify (Sutton & Barto, 2018). The task of the learning agent is to optimize a specific objective function, which is carried out using only information from the state of the environment, that is, without any external advisor (Merckling et al., 2022). The task of the learning agent can be accomplished by modeling the system as a Markov Decision Process (MDP). Table 1 defines the important terms associated with RL.

The long-term reward represented by the mathematical expression (1) is the objective function that the agent is interested in maximizing by taking the right actions (Sutton & Barto, 2018).

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{1}$$

Where γ is the discount-rate which ranges between 0 and 1. It determines the present value of future rewards?

The agent's goal is to maximize R_t by means of an optimal policy. *Action-value function* are used to estimate the expected $R_t$as the result of an action $a_t$from a state s_t represented by the formula (Vincent, 2018; Sutton & Barto, 2018):

$$Q^{\pi}(s,a) = E_{\Pi}\left[R_t \middle| s_t = s, a_t = a\right] \tag{2}$$

Where $Q^{\pi}(s,a)$ is the expected R_t at starting state s and action a following policy $\pi(s,a)$. The optimal policy can be represented now as (Sutton & Barto, 2018):

$$\pi^*(s,a) = \arg\max_{a \epsilon A} Q^{\pi}(s,a) \tag{3}$$

Where A is set of possible actions and Optimal policies also has the same optimal value function

$$Q^*(s,a) \; : \quad Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \tag{4}$$

One can note that the above eq. (4) is a deterministic policy where we have a given fixed state and only one action can be taken. However, one can have stochastically defined policies as well, where actions are taken according to their probabilities (Oh et al., 2017; Sutton & Barto, 2018). In both cases the optimal policy is always deterministic in single-agent MDPs and stationary in infinite horizon problems.

3 Methodology

We now describe the methodology we have developed for solving the problem of DM campaigns. There are two main strategies for solving reinforcement-learning problems. The first is to search in the space of behaviors to find one that performs well in the environment. This approach has been taken by work in genetic algorithms and genetic programming, and some more novel search techniques (Schmidhuber, 2015). The second is to use statistical techniques and

Table 1 Terms and definitions

SN	Term	Definition
1	State (S_t):	A discrete set of environment states at time *t*.
2	Action (A_t):	One of the set of agent actions taken at time *t*.
3	Policy (π (*s, a*)):	Probability distribution over the actions in the state S.
4	Reward -Intermediate (r_{t+1}):	Value returned by the environment to the agent depending on the action at time $t + 1$.
5	Reward – Long term (R_t):	Sum of all the immediate rewards throughout a complete decision process.

dynamic programming methods to estimate the utility of taking actions in states of the world. In this paper we have taken the second set of techniques because they take advantages of the special structure of RL problems that are not available in optimization problems in general actions in states of the world (Bennett & Parrado-Hernández, 2006; Zhang et al., 2022). We first present the stationary UCB algorithm wherein the distribution of reward does not change in time followed by non-stationary case wherein the reward distribution of reward remains constant over epochs and changes at an unknown time. To handle impact of seasonality we also look at the sliding window UCB.

3.1 Upper-Confidence Bound (UCB) Algorithm – Stationary Case, the Distribution of Reward Does Not Change in Time

Exploring different actions after having already learned from the environment overtime, exploring different actions should be ideally limited to the best performing actions. The UCB algorithm enables this by not selecting a random action, but constantly changing the magnitude of exploration alongside learning more about the action results (Agrawal, 1995). In the beginning, UCB promotes exploration to know the best performing action and then continues selecting the best action. After choosing the best action several times, based on the confidence in known actions, UCB promotes exploration again. If a different action performs better than the previous best action, then the new best action is promoted (Liu et al., 2020). So UCB starts by exploring the actions that have been tried the least amount of times and then learns the best rewarding action to exploit it until the point at which other actions end up being not chosen for a long time (Auer et al., 2002).

For our current marketing problem, choosing an advertisement can be considered as performing an action. Let $Q_d(t)$ be the estimated value of advertisement d at time-step t. The advertisement chosen by UCB at time-step t will be:

$$D_t = \arg\max_d \left[Q_d(t) + u\sqrt{\left\{ \frac{\log t}{N_d(t)} \right\}} \right] \quad (5)$$

Where, $N_d(t)$ is the number of times advertisement d has been selected up till time step t.

The first part of the equation $Q_d(t)$ is essentially the exploitation term. UCB selects the best rewarding advertisement until uncertainty rises too high. Up till time step t, if the estimated reward of advertisement d is highest then it will be promoted. UCB equation conveniently allows us to understand and manipulate the level of exploration. The parameter u is the uncertainty level of rewards that allows control and manipulation of exploration. High uncertainty in the estimated reward of ads results in low confidence. This results in increasing quantum of exploration term. One more variable affects the exploration term – number of times an advertisement has been selected. If an advertisement is selected for a small number of times or never, then the exploration term is large. As we constantly perform more and more actions, we get a better understanding of the expected rewards from certain ads. This results in increasing value of $N_d(t)$ and thus the overall sum in the UCB equation decreases. So, it is less likely that this advertisement will be explored. As the number of actions becomes very large, UCB relies more and more on exploiting the best rewarding ad.

UCB methods are deterministic policies extended to a non-parametric context. They consist in playing during the *t-th* round the arm i that maximizes the upper bound of a confidence interval for expected reward $\mu(i)$, which is constructed from the past observed rewards. The most popular, called UCB-1,[2] relies on the upper-bound:

$$\bar{x}_t(i) + c_t(i)$$

Where $\bar{x}_t(i) = (N_t(i))^{-1} \sum_{s=1}^{t} x_s(i) \; \mathbb{1}_{\{l_s=i\}}$ denotes the empirical mean, and $c_t(i)$ is a padding function.

Algorithm 1 UCB-1 (Garivier and Moulines, 2011)

for t from 1 to **K**, play arm l_t = t.

for t from $K + 1$ to **T**, play arm

$$l_t = arg \max_{1 \le i \le k} \overline{X_t}(i) + c_t(i)$$

3.2 Upper-Confidence Bound (UCB) Algorithm – Non-Stationary Case, the Distribution of Reward Remains Constant over Epochs and Changes at Unknown Time

The stationary formulation of the Multi Armed Bandit Problem addresses the challenges of exploration versus exploitation intuitively (Auer et al., 2002; Garivier & Moulines, 2011). However, it may fail to model an evolving environment where the reward distributions change in time. Lai et al. (2010) discusses the cognitive medium radio access problem, where a multi-channel system wants to exploit an empty channel. To model such situations where the

[2] UCB-1 belongs to the family of "follow the perturbed leader" algorithms and has proven to retain the optimal logarithmic rate (but with suboptimal constant). A finite-time analysis of this algorithm has been given in Auer et al. (2002); Auer et al. (2002); Auer et al. (2002/03). Other types of padding functions are considered in Audibert et al. (2007).

distribution of regards may change in time, we would need to consider non-stationary MAB problems. The upper confidence bound policies have been known to be rate optimal for linear search space (Agrawal, 1995; Auer et al., 2002). However, it becomes challenging in situation where the distribution of rewards remains constant over epochs and change at unknown time instants. Therefore, an upper-bound for the expected regret by upper-bounding the expectation of the number of times a suboptimal arm is played would be required. Discounted UCB and Sliding window UCB are used to this cause.

In the marketing campaign, the distributions of the rewards undergo abrupt changes due to seasonal (special event) effects. We derive a lower bound for the regret of any policy and analyze two algorithms: The Discounted UCB (Upper Confidence Bound) proposed by Kocsis and Szepesvári (2006) and the Sliding Window UCB Proposed by Garivier and Moulines (2008). We find that they are almost rate-optimal, as their regret almost matches a lower bound.

3.2.1 Discounted Upper Confidence Bound (D-UCB)

In the family of UCB policies many researchers including Kleinberg et al. (2008) and Kocsis and Szepesvári (2006) have proposed adding discount factor $\gamma \in (0,1)$ to the policies. One may make a note that when $\gamma = 1$, discounted UCB will be UCB. To estimate the instantaneous expected reward, the D-UCB policy averages past rewards with a discount factor giving more weight to recent observations.

Algorithm 2 Discounted **UCB** (Garivier and Moulines, 2011)

for t from 1 to **K**, play arm l_t = t.

for t from $K + 1$ to **T**, play arm

$$l_t = \arg\max_{1 \le i \le k} \overline{X_t}(\square, i) + c_t(\square, i).$$

3.3 Sliding Window UCB

Garivier and Moulines (2011) proposes a more abrupt variant of UCB where averages are computed on a fixed-size horizon. At time t, instead of averaging the rewards over the whole past with a discount factor, sliding window UCB relies on a local empirical average of the observed rewards, using only the τ last plays. Specifically, this algorithm constructs an UCB for the instantaneous expected reward.[3]

[3] The focus of this paper is on an abruptly changing environment, but it is believed that the theoretical tools developed to handle the non-stationarity can be applied in different context.

Algorithm 3 Sliding-Window **UCB** (Garivier and Moulines, 2011)

for t from 1 to **K**, play arm l_t = t.

for t from $K + 1$ to **T**, play arm

$$l_t = \arg\max_{1 \le i \le k} \overline{X_t}(\top, i) + c_t(\top, i).$$

3.4 Dataset and UCB Implementation

To empirically show the effectiveness of RL in the DM, we conducted the experiment using UCB algorithm for digital campaign at a startup firm ***yourfirstad.com***. Yourfirstad designs and places ads to get traffic for advertisers. Traditionally they have been using A/B/n testing for optimizing display of ads. To compare the performance of the A/B/n testing approach with the proposed RL approach we compare the clicks the ads are able to get in the two cases. We compare the results of our model for an advertiser which had six different creatives rotated almost equally between January 2019 and May 2019 with approximately 10,000 impressions.

In the RL implementation in June 2019, different creatives were served each time a customer visited the webpage during a period of one month. We ran this experiment for a total of 10,000 rounds (to ensure comparability), that is, each time a customer connects to this web page, it counts as a round n. In each round, the version gets a reward, 1 if the customer clicks on the ad, 0 if it does not click. Further, to evaluate the UCM algorithm with sliding window and discounted rewards, we tested this algorithm on another 10,500 customers from 1st July 2019 – 31st July 2019. Figure 2 explicates the visual implementation of RL Algorithm.

1. At each round n, we have two values for a given advertisement

- $N_d(n)$ - the number of times advertisement d was selected up to round n.
- $R_d(n)$ - the sum of rewards for advertisement d up to round n

2. Then we compute the average reward for advertisement d as:

$$r_d(n) = R_d(n)/N_d(n) \quad (6)$$

Where d = one of the 6 ads; N = total number of iterations; n = Nth iteration at time t.

3. We also compute the upper bound as follows:

Table 2 Key characteristics of data and simulation

Parameter	A/B Testing	RL Implementation: Simple UCB	RL Implementation UCB with sliding window & discounted rewards
Number of Customers (Rounds)	10,000	10,000	10,500
Versions of Ads	6	6	6
Rewards (Positive/Negative)	NA	1/0	1/0
Number of Episodes[a]	NA	30	30
Date range for campaign	Jan 1st, 2019 to June 30th 2019	June 1, 2019 – June 30, 2019	July 1, 2019 – 31st July 2019
Sliding window size	NA	NA	100

[a]For the 30 episodes are associated with different learning rates for episodes. Starting from large values and then gradually decreasing the learning rate for more exploitation

$$U_d(n) = r_d(n) + E_d(n) \tag{7}$$

Where E constitutes the exploration term of the upper bound.

After every round, the upper bound is calculated for every ad. The advertisement with the highest upper bound is chosen. For the sliding window UCB simulations, N is restricted to the last w rounds or time steps. This directly affects the average reward r as well as the exploration term E. Consequently, this affects the upper bound U of every advertisement at every time step. Table 2 captures the key aspects of the RL-based implementation, and Fig. 3 represents the Schematic approach to RL algorithm implementation on the platform.

4 Results

We first present results with randomized trials where no strategy is used, followed by the UCB algorithm. Then we look at the results with special treatment of the UCB algorithm in terms of sliding window and discounted rewards. We then demonstrate the influence of seasonality on advertisement selection and how our approach can still work by taking care of it. Finally, we explain how sliding window UCB performs better than UCB when looking at timestamp-based advertisement selections. Finally, we do a Comparison of methods performance across all datasets.

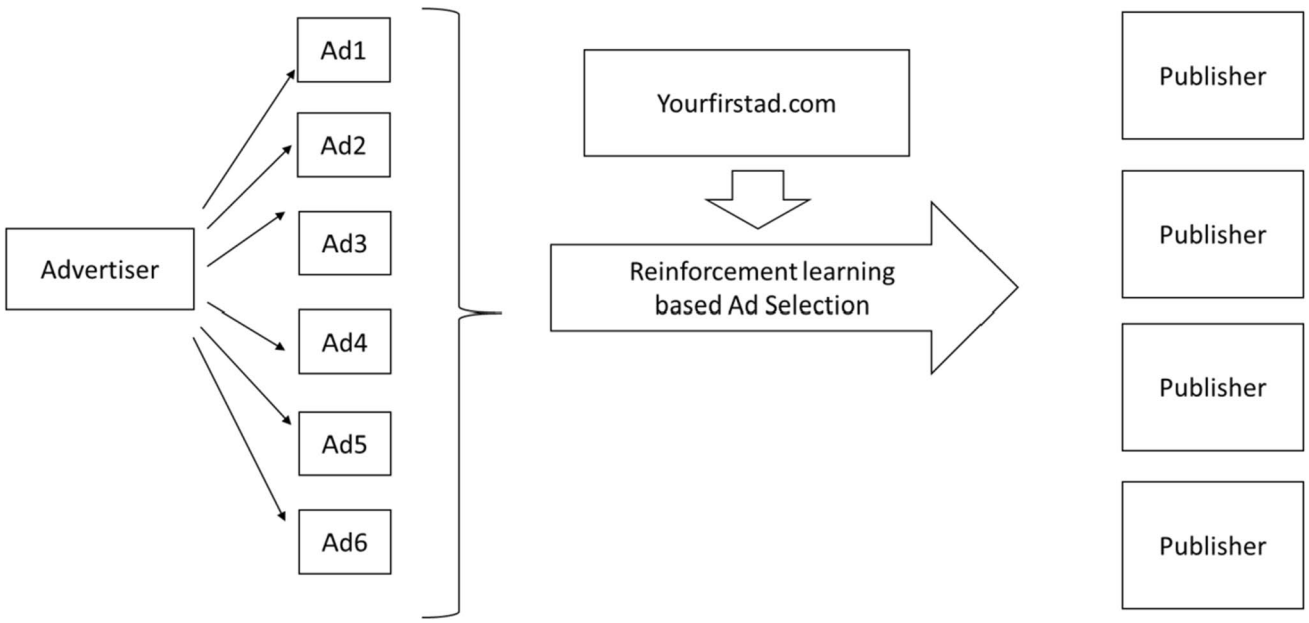

Fig. 3 Schematic representation of RL algorithm implementation on the platform

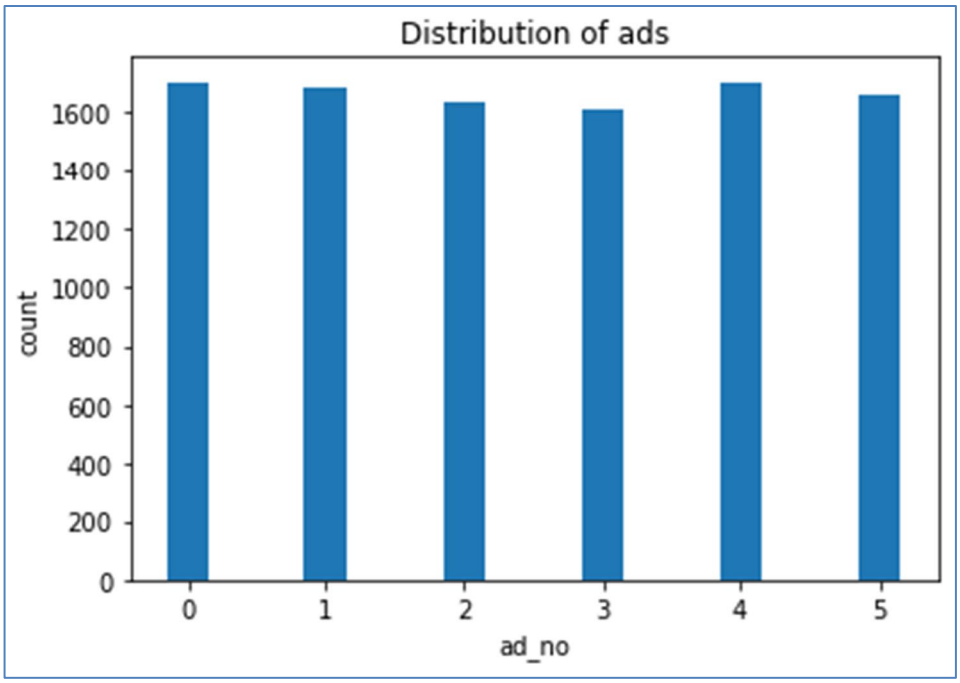

Fig. 4 Total number of clicks attained with randomized trial: 1288

4.1 Comparing RL Based AI Implementation (UCB Algorithm) and A/B/n Optimization

Figure 4 and Table 3 show the results of not using any strategy, thereby using randomized trials (A/B/n testing). All the ads were selected a relatively similar number of times. The variance in the count of selection of ads is low. However, the results do not help us in discovering the best ads. The standard deviation of the results is 43.34.

From the statistics, the best advertisement was automatically selected 16.6% of the time. Out of a possible pool of 6 ads, this percentage is not far from the mean. So, conducting a randomized trial would generate rewards equivalent to the sum of the weighted mean of the rewards from all six ads. Clearly, there is a significant amount of loss in not choosing the best rewarding advertisement more frequently. Even if we stop midway, and then run only the best performing advertisement (Ad 2) the total number of clicks (projected at the same click through rate) would be 1710 clicks.

Comparing the performance with UCB algorithm clearly prefers advertisement 2, as shown in Table 4 below. The total reward achieved with the UCB algorithm is 2028 compared to 1288 achieved without any strategy (Refer Table 3).

The best performing advertisement 2 accounts for 94.1% of comprehensive selections of all ads. Initial intuition suggests that advertisement 2 must have a higher mean of reward than any other ad which reflects in the distribution of ad exposures as seen in Fig. 5.

Further, it would be interesting to explore the impact of the uncertainty level coefficient on the number of clicks on campaigns. Table 5 illustrates the counts of selection of ads for different settings of the parameter – uncertainty level coefficient. The performance of the UCB algorithm was checked for several levels of uncertainty. As we increase

Table 3 Randomized Trials for first 5000 (Jan -May)

Ad	Count of selection	Percentage count	Reward generated by Ad (10,000 runs)	Reward for first 5000 runs
Ad 0	1674	16.7%	172	83
Ad 1	1636	16.4%	252	127
Ad 2	1661	16.6%	353	173
Ad 3	1687	16.9%	184	95
Ad 4	1734	17.3%	224	113
Ad 5	1608	16.1%	103	57
Total			1288	684

Table 4 Trials with UCB

	Count of selection	Percentage count	Reward generated by Ad
Ad 0	98	1.0%	9
Ad 1	190	1.9%	28
Ad 2	9409	94.1%	1962
Ad 3	122	1.2%	14
Ad 4	98	1.0%	9
Ad 5	83	0.8%	6
Total			2028

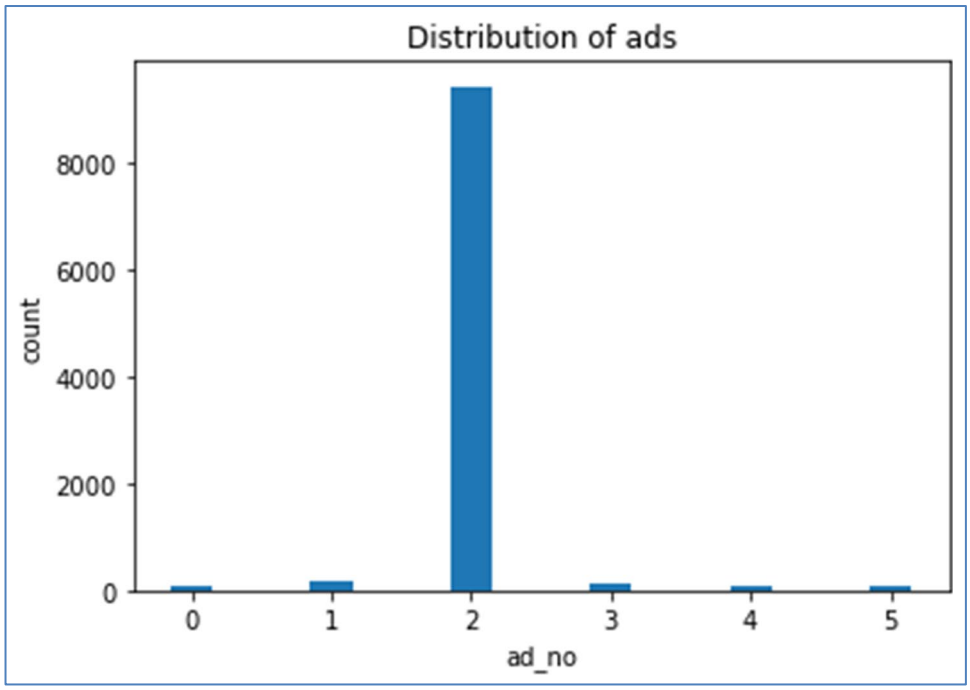

Fig. 5 Number of clicks attained with UCB algorithm: 2028

Table 5 Level of uncertainty vs Rewards for each Advertisement

Uncertainty Level	0.1		0.5		1		2		3	
Ad 0	8	0.1%	37	0.4%	98	1.0%	146	1.5%	228	2.3%
Ad 1	8	0.1%	175	1.8%	190	1.9%	359	3.6%	529	5.3%
Ad 2	8230	82.3%	9628	96.3%	9409	94.1%	8747	87.5%	8287	82.9%
Ad 3	1730	17.3%	32	0.3%	122	1.2%	296	3.0%	430	4.3%
Ad 4	8	0.1%	91	0.9%	98	1.0%	320	3.2%	326	3.3%
Ad 5	16	0.2%	37	0.4%	83	0.8%	132	1.3%	200	2.0%
Reward	1947		2056		2028		1964		1944	

the value of the uncertainty coefficient, the level of exploration increases. From the table, we see that the proportion selection of ads other than advertisement 2 increases with increasing uncertainty level above 1. Our experiment shows this trend continuing for higher values of uncertainty coefficient. As the level of exploration increases beyond the uncertainty coefficient of 1, advertisement 2 is selected fewer and fewer times. We also see the total reward also decreases after the uncertainty coefficient increases beyond 1.

The results are more interesting when the value of the uncertainty coefficient is below 1. While moving from the uncertainty coefficient of 1 to 0.5, the exploration decreases further, resulting in a very high selection of advertisement 2. The highest reward with 2056 clicks was attained at an uncertainty coefficient of 0.5. For an uncertainty coefficient of 0.1, theoretically, the exploration level decreases. However, we also see that advertisement 3 was selected more than any other uncertainty level. It is due to variation in the expected value of reward from advertisement 2. Due to a deficient level of exploration, the UCB algorithm selects advertisement 3 many times because the expected reward from advertisement 3 was higher than that of advertisement 2 for a few observations. But since the exploration is low, the algorithm could not promote advertisement 2 over advertisement 3, which indicates the importance of exploitation – exploration trade-off. From our results, we clearly understand that the best total reward is achieved between the uncertainty level of 0.1 and 1.

4.2 Sliding Window UCB – Adapting RL Based AI Implementation to Changes in Consumer Patterns

The implementation of a non-stationary UCB algorithm was done to an extended dataset with 10,500 more-time steps. So, the final comprehensive dataset consisted of 20,500-time steps. For experimental purposes, we considered a sliding window size of 100-time steps. Therefore, the input

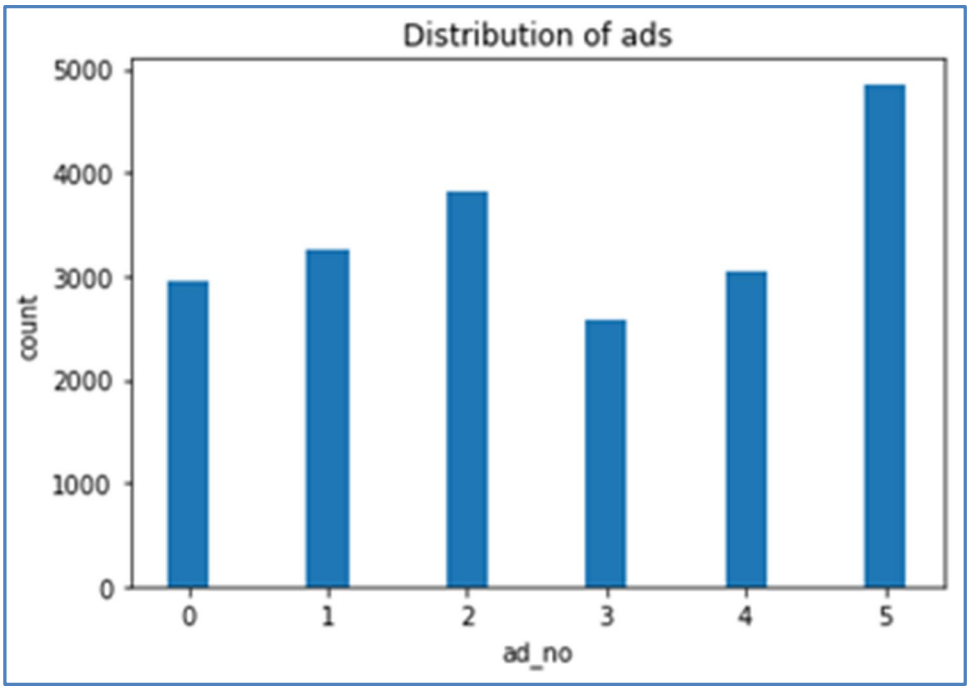

Fig. 6 Total number of click for non-stationary UCB

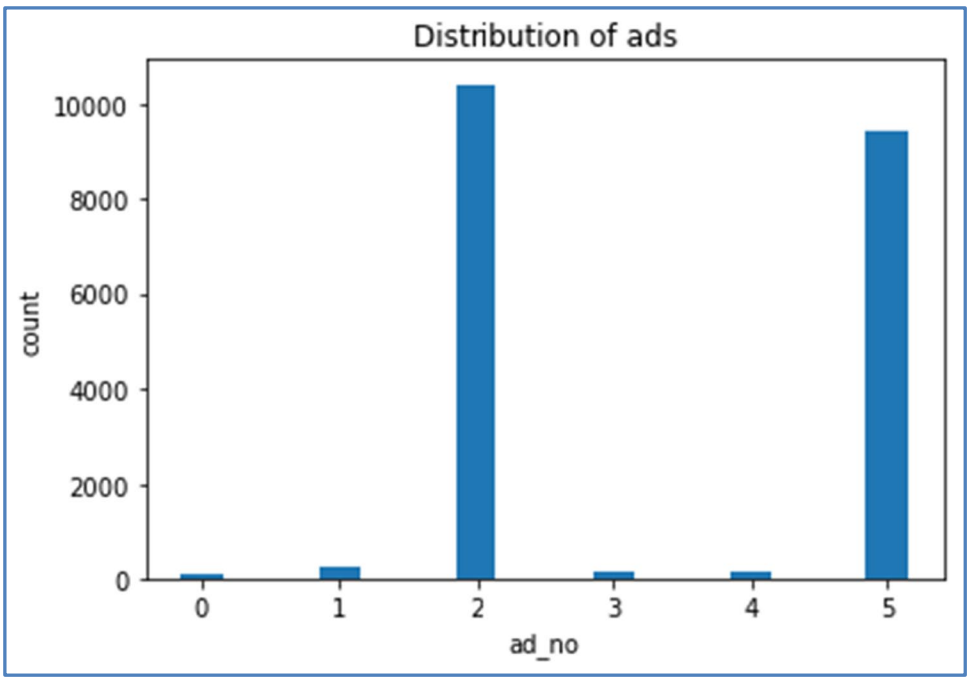

Fig. 7 Total number of click for non-stationary UCB

parameters will be based on the calculations of only 100 latest time steps of historical data at every time step. The chart below in Fig. 7 shows the result of using a non-stationary UCB algorithm. The total reward generated from this simulation was 2944. The uncertainty level coefficient was maintained at 1. Thus, we see a somewhat ambiguous distribution of advertisement selection. It is not clear whether the advertisement selection is optimum or not from the chart.

We ran another simulation without the sliding window but with the same uncertainty level coefficient to evaluate the performance. Figure 6 shows the result with the stationary UCB algorithm.

First, we see an evident result pointing us to the two best advertisements. Advertisement 2 and advertisement 5 are the best advertisements. We already knew advertisement 2 was the best performing advertisement from our first simulation

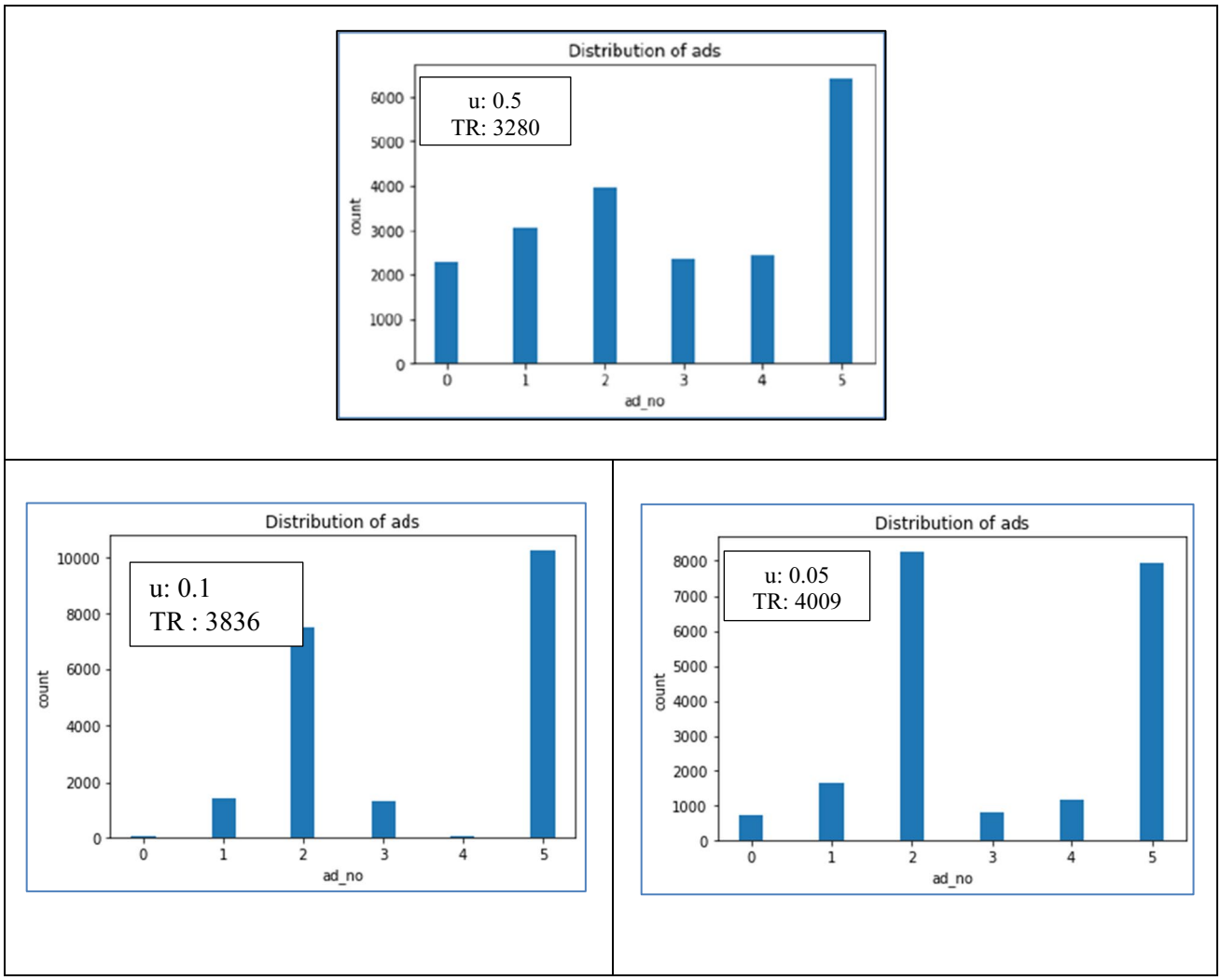

Fig. 8 Total number of clicks for sliding window UCB

runs for the first 10,000-time steps. So, advertisement 5 is the best for the next 10,500 steps. The total reward generated from this simulation was 4048. For the same uncertainty level coefficient (of value 1), the stationary UCB seems to outperform the non-stationary UCB algorithm with a sliding window of 100 (Fig. 7).

The charts in Fig. 8 show the results of simulation runs at different values of uncertainty level coefficient. We see clearer distributions as the uncertainty level decreases. The total rewards are also shown in every chart. Both the total reward and the advertisement selection distribution converges towards the results obtained from stationary UCB simulation at an uncertainty level of 1. For these non-stationary simulations, decreasing the uncertainty level means decreasing the impact of the exploration part of our mathematical equation on the upper bound at every time step. So, reducing exploration of the model by lowering the uncertainty value makes the model performance converge towards the stationary UCB results. This reiterates the argument that non-stationary UCB indirectly increases the exploration of the model.

4.2.1 Advertisement Selection by Time Step

To understand the effect of using a sliding window, we need to look deeper at advertisement selection. Figures 9 and 10 provide the complete picture of advertisement selection at every time step. Within the first ~2000-time steps (shaded in light yellow), we see a difference in advertisement selection between the two methods – UCB and SW-UCB.We see high exploration concentrated in the starting period. However, if we focus on the time steps from 10,000 to end (shaded in light red), we observe a significant difference in exploration without using a sliding window. With UCB, the only advertisement selected, after ~12,000-time steps, is advertisement 5. However, with SW-UCB, other advertisements are also selected after ~12,000-time steps. The UCB

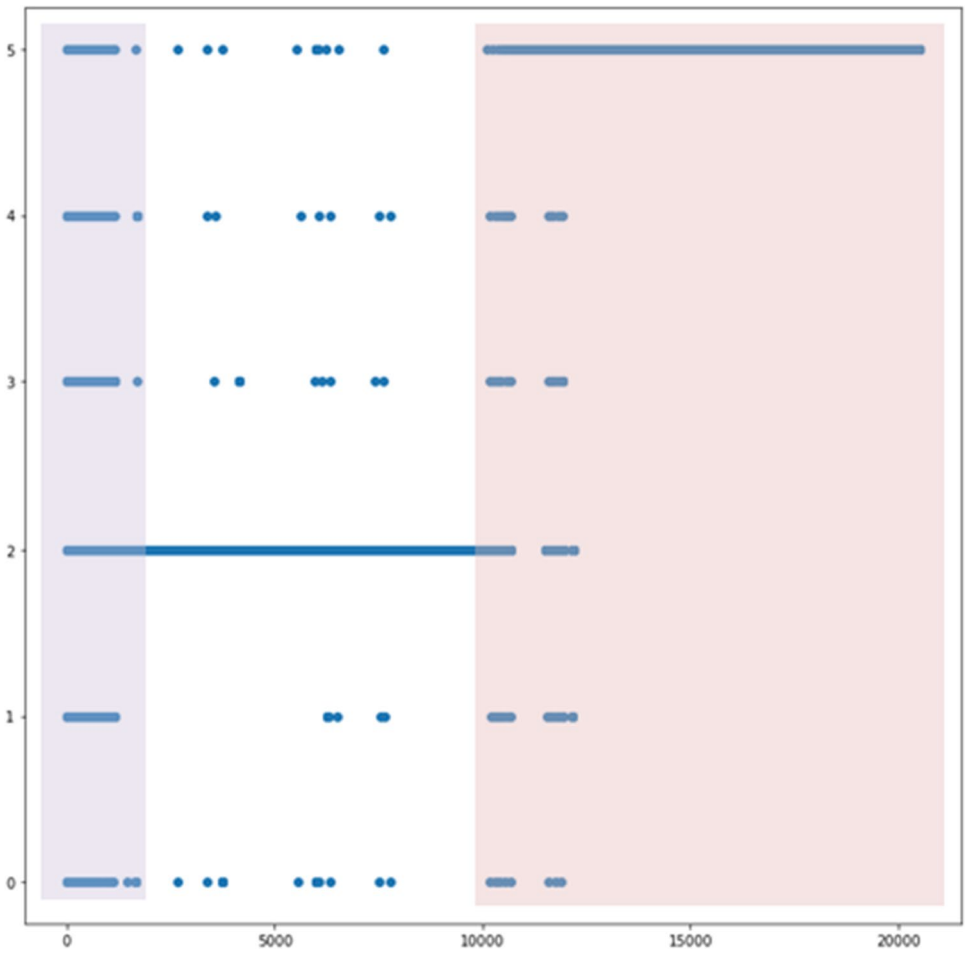

Fig. 9 Advertisement selection with time stamp using UCB

method stops exploring after ~12,000-time steps, whereas SW-UCB continues exploration.

The reason for such a distinguished advertisement selection among the two methods builds a strong case for SW-UCB. With the UCB method, the upper bound is highest for advertisement 5 after ~12,000-time steps because a large amount of history is being used to calculate upper bounds. In the case of SW-UCB, the upper bound is highest for advertisement 5 for most time steps in that period. But the critical point to note is that, because the short history of data is used for upper bound calculation at any time step, the upper bound of other advertisements exceeds the upper bound of advertisement 5 several times. This behavior is critical in real-life marketing use cases. For example, one advertisement could be the top performer for a long duration. But if certain events impact advertisement selection for the short term, then these effects will be lost within the UCB method. For example, let's take a prevalent aspect of the seasonality effect. Capturing the impact of festivals is one example. Sales of certain goods may outperform sales of all other goods throughout the year. But during a short-term festival period, some other goods may take over the highest-grossing role. SW-UCB helps in capturing such effects, unlike UCB.

By introducing a small sliding window, which allows exploiting only the latest small amount of historical data, we indirectly introduce more exploration. If advertisement 2 has been performing well for several time steps, the stationary UCB keeps selecting advertisement 2 while exploring only a few times throughout the long term. In the case of non-stationary UCB, if any other advertisement is selected 100 consecutive times, the exploration part of the mathematical formula reaches infinity for all the other five advertisements even if the uncertainty level coefficient is set to the same value in the stationary algorithm. Hence, we experimentally prove that using a sliding window non-stationary algorithm indirectly implies higher exploration. To test this finding further, we ran more non-stationary UCB while lowering the value of the uncertainty level coefficient.

5 Discussion

The use of AI for marketing decisions has been growing (S.-S. Chen et al., 2021; S.-Y. Chen et al., 2021; Miklosik et al., 2019) and is expected to revolutionize marketing (Huang & Rust, 2018; Huang & Rust, 2021; Rust, 2020) through better

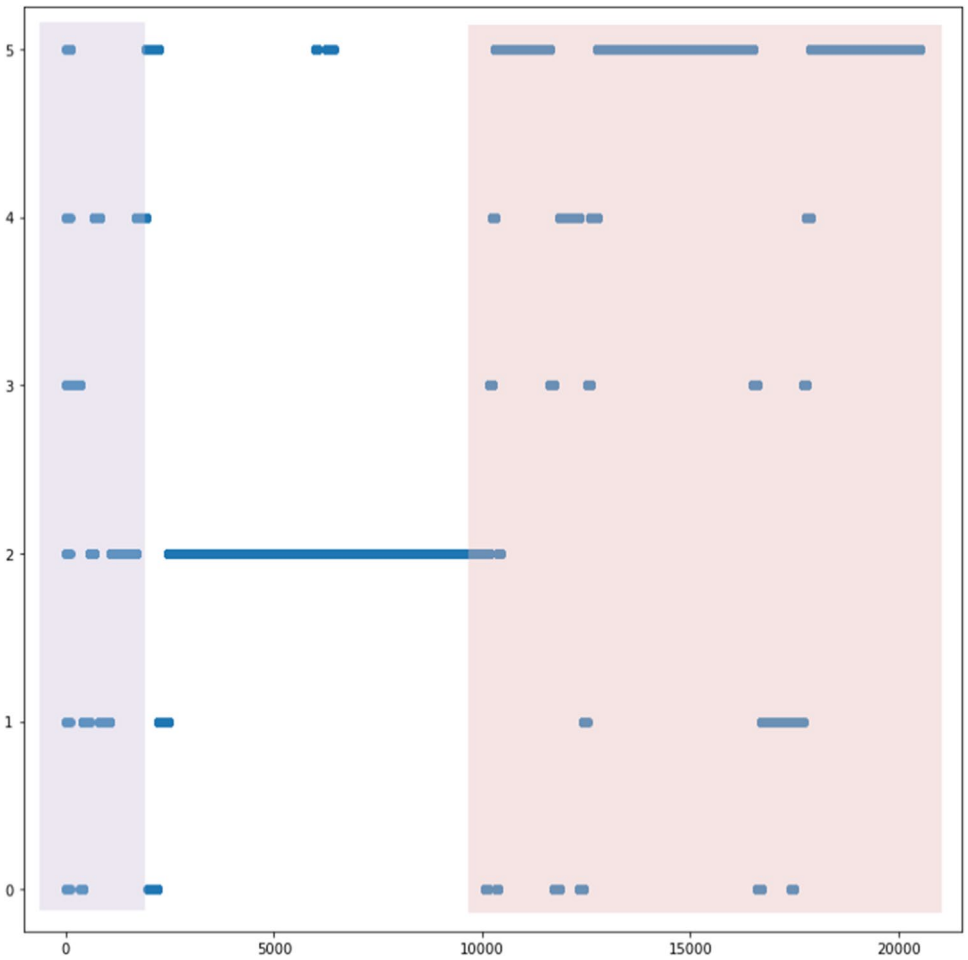

Fig. 10 Advertisement selection with time stamp using SW-UCB

personalization and targeting, real time optimization and automation, and better understanding of customer journeys (Ma & Sun, 2020). It is argued that AI has the potential to affect both, revenue and costs. While better marketing decisions like pricing promotions and recommendations are expected to increase revenue, AI based automation in customer service is expected to lower costs (Davenport et al., 2020). In our study, we explore an advertiser's problem of maximizing clicks on the advertisement. More specifically, we used RL to optimize this marketing decision. Within the domain of RL, researchers have explored how this approach to optimize marketing decisions (Schwartz et al., 2017) such as framing and framing dynamic pricing policy (Misra et al., 2019) but its applicability to ad optimization has been rather limited.

5.1 Contributions to Literature

The current study extends literature in line with the editorial directions for marketing management based on big data and ML methods (Chintagunta et al., 2016). In this study we propose a novel approach that addresses the above shortcomings of A/B or A/B/n testing and other prevailing supervised approaches and attempts to contribute to the digital marketing literature through practical applications of AI algorithms (Chiusano et al., 2021; Choi et al., 2020). To overcome the above drawbacks of existing classical and AI approaches, we present a novel application of the Upper Confidence Bound (RL algorithm) and apply it to actual data collected at the firm for a DM campaign. Our study shows that the Reinforcement Learning (RL) approach has the following two advantages over existing approaches. First, the existing supervised learning approaches predict a class and are trained on the class, while the reinforcement algorithm learns from the reward/punishment and updates itself over time. Second, RL has a temporal dimension that looks at the past and current state. This approach helps to minimize the impact of external factors as the algorithm is getting updated in real-time with dynamic attribute lists.

We benchmark the performance of the proposed AI based approach against the popular A/B/n testing approach and find that the rewards generated by RL are higher. If the A/B/n testing is performed a vast number of times, then there would be a significant loss of rewards in randomly

promoting advertisements. Effective manipulation of exploration and exploitation is the fundamental concept of RL and we find that high exploration, during the initial rounds, to learn reward expectations from every advertisement would result in a sub-optimal realization of rewards. However, after several epochs, the algorithm promotes the advertisements to provide better rewards, thereby increasing the dominance of exploitation. The strategy to promote the better rewarding advertisement, and at the same time, learning rewards from other advertisements with controlled exploration resulted in improved reward realization. Moreover, utilizing a tuning parameter of uncertainty level helps determine the level of exploration required for a specific marketing case. The best results could be achieved by analyzing the impact of the uncertainty coefficient on the total reward. Usage of the above approach by the marketer can help navigate seasonality variations in the performance of an advertisement creative. In addition, the inclusion of new creatives and testing them does not become a cumbersome process. In achieving the above objectives, we further contribute in empirical generalizations in marketing in the digital age whereby sensing using smart technologies and adaptive change in marketing strategies are needed in literature (Hanssens, 2018).

5.2 Practical Implications

Recent academic literature has suggested that although digital advertising markets are growing, market inefficiencies need further attention (Gordon et al., 2021). One such in efficiency concerns online ad measurement and it is suggested that the chasm between marketing practitioners and academicians is around the issues of endogeneity (Rutz & Watson, 2019). Despite the prevalence of field experiments in Marketing, often presented as gold standards to create causal insights (Johnson et al., 2017), problems concerning A/B testing remain. Feit and Berman (2019) in their research reframe A/B tests as tools to manage the trade-off between the opportunity cost of the test and the potential losses associated with deploying a suboptimal treatment to the entire population and propose an alternative that theoretically achieves the same performance as MAB implementation. They argue that MABs are hard to implement. Our study, shows that RL based MAB implementation can be achieved in practice. An RL approach has been adopted, which has been demonstrated to have better quality of outcome in terms of advertisement selection and advertisement distribution, concerning parameters like advertisement clicks.

Although the last decade was touted as the decade of 'data-rich' digital marketing, made available to marketers based on unprecedented data on firm and customer behavior (Sridhar & Fang, 2019), challenges with respect to privacy tensions as the product of firm–consumer interactions, facilitated by digital technologies (Quach et al., 2022) and the black boxed nature of AI algorithms such as neural networks (Rai, 2020) continue to linger. The inscrutability of such algorithms can affect users' trust in the system, especially in contexts where the consequences are significant, and lead to the rejection of the systems (Rai, 2020). Our study provides an instance of dynamic and inexpensive AI that used minimal user information to avoid privacy issues. At every epoch, the algorithm updates the information stored about every advertisement. Historical information on advertisement selection and reward at every epoch is not required to be stored and used subsequently. There are numerous real-world cases where a real-time update of an AI model is highly crucial. In the majority of those cases, either there is no possibility of a fast update of the results for real-time decision making, or the model is too computationally expensive to deploy in production realistically. Another advantage of this approach is that the optimization and performance improvement can happen without taking into consideration micro level data such as user profile or advertisement characteristics. Further since this algorithm does not use any personal identifiers, it is therefore less amenable to be adversely affected in the wake of GDPR laws. Further an environment sensitive learning model and measurement of campaign outcomes can help organizations justify the deployment of financial and human resources more prudently into marketing activities (Votto et al., 2021).

6 Conclusion

This paper provides a RL approach to the marketer's decision of optimizing and selecting advertisements. The proposed approach optimizes the delivery of most performing advertisements and is also tuned to handling seasonal variations or exploring newer creatives than the most frequently used A/B testing approach. Therefore, we feel the proposed approach is generic and would work for different categories of products or Industries. There are, however, some limitations to the study. Firstly, in this study we measure the performance of the advertisement only in terms of the number of clicks that the advertisement gets. While clicks are important more advanced measures like conversions or purchases may be used in future study as there may be trade-offs between getting clicks and conversions. Secondly, the results obtained through the simulations are benchmarked against the currently used approach of A/B testing employed in the Industry. One possibility of extending this work could be using the deep Q network (Arulkumaran et al., 2017) that takes advertisement characteristics, user demographics, and history as attributes to build a more advanced RL model. Such a model could enhance capabilities by marrying the exploitation –exploration trade off with the predictability feature captured in the form of ad characteristics such as

text, color, size and user characteristics such as age, gender, and affinities. Finally, further studies may compare the performance of the proposed approach to other approaches like boosting and decision trees, which may require substantially more data than the proposed approach.

Acknowledgements The authors gratefully acknowledge the team at YourFirstad for their insights and sharing of data for our study. This research paper is submitted as part of the efforts undertaken under a collaboration between the research teams of Indian Institute of Technology Delhi and BASF Germany, funded by BASF Germany.

Declarations

Conflict of Interest The authors report no conflict of interest in the study undertaken. All contributors have been listed as authors and all authors have contributed.

References

Agrawal, R. (1995). Sample mean based index policies by O (log n) regret for the multi-armed bandit problem. *Advances in Applied Probability, 27*(4), 1054–1078 {Sutton, 2018 #654}.

American Marketing Association (2021). *Digital marketing virtual conference*. https://www.ama.org/events/virtual-conference/2021-digital-marketing-virtual-conference/

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine, 34*(6), 26–38.

Ascarza, E. (2021). *When a/B testing Doesn't tell you the whole story*. Harvard Business Review.

Aswani, R., Kar, A. K., & Ilavarasan, P. V. (2018). Detection of spammers in twitter marketing: A hybrid approach using social media analytics and bio inspired computing. *Information Systems Frontiers, 20*(3), 515–530.

Audibert, J.-Y., Munos, R., & Szepesvári, C. (2007). Tuning bandit algorithms in stochastic environments. In M. Hutter, R. A. Servedio, & E. Takimoto (Eds.), *Algorithmic Learning Theory* (Vol. 4754, pp. 150–165). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75225-7_15

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning, 47*(2), 235–256.

Barone, M. J., et al. (2007). Consumer response to retailer use of cause-related marketing: Is more fit better? *Journal of Retailing, 83*(4), 437–445.

Basse, G. W., & Airoldi, E. M. (2018). Limitations of design-based causal inference and a/B testing under arbitrary and network interference. *Sociological Methodology, 48*(1), 136–151.

Bennett, K. P., & Parrado-Hernández, E. (2006). The interplay of optimization and machine learning research. *Journal of Machine Learning Research, 7*(46), 1265–1281.

Bojinov, I., Rambachan, A., & Shephard, N. (2021). Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics, 12*(4), 1171–1196.

Boone, D. S., & Roehm, M. (2002). Retail Segmentation Using Artificial Neural Networks. *International Journal of Research in Marketing, 19*(3), 287–301.

Botvinick, M., et al. (2019). Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences, 23*(5), 408–422.

Chen, S.-S., Chobey, B., & Singh, V. (2021a). A neural network based Price sensitive recommender model to predict customer choices based on Price effect. *Journal of Retailing and Consumer Services, 61*, 102573.

Chen, S.-Y., He, Q.-F., & Lai, C.-F. (2021b). Deep reinforcement learning-based robot exploration for constructing map of unknown environment. *Information Systems Frontiers*.

Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Marketing science and big data. *Marketing Science, 35*(3), 341–342.

Chiusano, S., Cerquitelli, T., Wrembel, R., & Quercia, D. (2021). Breakthroughs on cross-cutting data management, data analytics, and applied data science. *Information Systems Frontiers, 23*(1), 1–7.

Choi, H., Mela, C. F., Balseiro, S. R., & Leary, A. (2020). Online display advertising markets: A literature review and future directions. *Information Systems Research, 31*(2), 556–575.

Davenport, Thomas H., et al. (2011). Know what your customers want before they do. *Harvard Business Review*, Dec. 2011. https://hbr.org/2011/12/know-what-your-customers-want-before-they-do. Accessed 30 Apr 2021.

Davenport, T., Guha, A., Grewal, D., & Bresgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*, 24–42.

Du, R. Y., et al. (2021). Capturing Marketing Information to Fuel Growth. *Journal of Marketing, 85*(1), 163–183.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897–904.

Fabijan, A., Dmitriev, P., McFarland, C., Vermeer, L., Holmström Olsson, H., & Bosch, J. (2018). Experimentation growth: Evolving trustworthy a/B testing capabilities in online software companies. *Journal of Software: Evolution and Process, 30*(12), e2113.

Feit, E. M., & Berman, R. (2019). Test & roll: Profit-maximizing a/b tests. *Marketing Science, 38*(6), 1038–1058.

Fujimoto, S., Hoof, H., & Meger, D. (2018, July). Addressing function approximation error in actor-critic methods. In International Conference on Machine Learning (pp. 1587-1596). PMLR.

Gallo, A. (2017). A refresher on a/B testing. *Harvard Business Review*. https://hbr.org/2017/06/a-refresher-on-ab-testing. Accessed 30 Apr 2021.

Garivier, A., & Moulines, E. (2008). *On upper-confidence bound policies for non-stationary bandit problems* (arXiv:0805.3415). arXiv. http://arxiv.org/abs/0805.3415

Garivier, A., & Moulines, E. (2011). On upper confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory* (pp. 174–188). Springer.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology, 68*(1), 101–128.

Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science, 30*(3), 389–404.

Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management, 62*, 100–107.

Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing, 85*(1), 7–25.

Gupta, S., et al. (2020). Optimizing creative allocations in digital marketing. In *Advances in Computing and Data Sciences* (Vol. 1244, pp. 419–429). Springer Singapore.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018, July). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In International conference on machine learning (pp. 1861-e1870). PMLR.

Hanssens, D. M. (2018). The value of empirical generalizations in marketing. *Journal of the Academy of Marketing Science, 46*(1), 6–8.

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172.

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50.

Huber, R., Oberländer, A. M., Faisst, U., & Röglinger, M. (2022). Disentangling capabilities for industry 4.0-an information systems capability perspective. *Information Systems Frontiers*, 1–29. https://doi.org/10.1007/s10796-022-10260-x

Iacobucci, D., et al. (2019). The state of marketing analytics in research and practice. *Journal of Marketing Analytics, 7*(3), 152–181.

Jain, D., Dash, M. K., Kumar, A., & Luthra, S. (2021). How is blockchain used in marketing: A review and research agenda. *International Journal of Information Management Data Insights, 1*(2), 100044.

Jang, B., et al. (2019). Q-Learning Algorithms: A Comprehensive Classification and Applications. *IEEE Access, 7*, 133653–133667.

Javanmard, A., & Montanari, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *The Annals of Statistics, 46*(2), 526–554.

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research, 54*(6), 867–884.

Kar, A. K., & Kushwaha, A. K. (2021). Facilitators and Barriers of Artificial Intelligence Adoption in Business–Insights from Opinions Using Big Data Analytics. *Information Systems Frontiers, 1-24*. https://doi.org/10.1007/s10796-021-10219-4

Karuga, G. G., Khraban, A. M., Nair, S. K., & Rice, D. O. (2001). AdPalette: An algorithm for customizing online advertisements on the fly. *Decision Support Systems, 32*(2), 85–106.

Kim, J., Kang, S., & Lee, K. H. (2021). Evolution of digital marketing communication: Bibliometric analysis and network visualization from key articles. *Journal of Business Research, 130*, 552–563.

Kleinberg, R., Slivkins, A., & Upfal, E. (2008). Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing* (pp. 681-690).

Kocsis, L., & Szepesvári, C. (2006). Bandit based Monte-Carlo planning. In *European conference on machine learning* (pp. 282–293). Springer.

Kohavi, R., & Longbotham, R. (2017). Online controlled experiments and a/B testing. *Encyclopedia of machine learning and data mining, 7*(8), 922–929.

Kohavi, R., & Thomke, S. (2017). The surprising power of online experiments. *Harvard Business Review, 95*(5), 74–82.

Kushwaha, A. K., & Kar, A. K. (2021). MarkBot–a language model-driven chatbot for interactive marketing in post-modern world. *Information Systems Frontiers*,1–18. https://doi.org/10.1007/s10796-021-10184-y

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), 100017.

Lai, L., El Gamal, H., Jiang, H., & Poor, H. V. (2010). Cognitive medium access: Exploration, exploitation, and competition. *IEEE Transactions on Mobile Computing*, *10*(2), 239–253.

Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and Mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing, 80*(6), 146–172.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

Liu, X., Derakhshani, M., Lambotharan, S., & Van der Schaar, M. (2020). Risk-aware multi-armed bandits with refined upper confidence bounds. *IEEE Signal Processing Letters, 28*, 269–273.

Liu, R., Gupta, S., & Patel, P. (2021). The application of the principles of responsible AI on social media marketing for digital health. *Information Systems Frontiers*, 1–25. https://doi.org/10.1007/s10796-021-10191-z

Luo, L., et al. (2013). Marketing via Social Media: A Case Study. *Library Hi Tech, 31*(3), 455–466.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing, 37*(3), 481–504.

Merckling, A., et al. (2022). Exploratory state representation learning. *Frontiers in Robotics and AI, 9*, 1–16.

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access, 7*, 85705–85718.

Misra, K., Schwartz, E. M., & Abernethy, J. (2019). Dynamic online pricing with incomplete information using multiarmed bandit experiments. *Marketing Science, 38*(2), 226–252.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *In International conference on machine learning* (pp. 1928-1937). PMLR.

Nadkarni, S., & Prügl, R. (2021). Digital transformation: A review, synthesis and opportunities for future research. *Management Review Quarterly, 71*(2), 233–341.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science, 27*(2), 185–204.

Obal, M. W., & Lv, W. (2017). Improving banner ad strategies through predictive modeling. *Journal of Research in Interactive Marketing, 11*(2), 198–212.

Oh, Junhyuk, et al. (2017). 'Value prediction network'. Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., Neural Information Processing Systems.

Pandey, N., Nayal, P., & Rathore, A. S. (2020). Digital marketing for B2B organizations: Structured literature review and future research directions. *Journal of Business & Industrial Marketing, 35*(7), 1191–1204.

Quach, S., Thaichon, P., Martin, K. D., Weaven, S., & Palmatier, R. W. (2022). Digital technologies: Tensions in privacy and data. *Journal of the Academy of Marketing Science*, 1–25.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141.

Rathore, H., Sahay, S. K., Nikam, P., & Sewak, M. (2021). Robust android malware detection system against adversarial attacks using q-learning. *Information Systems Frontiers, 23*(4), 867–882.

Ruiz-Real, J. L., Uribe-Toril, J., Torres, J. A., & De Pablo, J. (2021). Artificial intelligence in business and economics research: Trends and future. *Journal of Business Economics and Management, 22*(1), 98–117.

Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing, 37*(1), 15–26.
Rutz, O. J., & Watson, G. F. (2019). Endogeneity and marketing strategy research: An overview. *Journal of the Academy of Marketing Science, 47*, 479–498.
Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge, 6*(2), 92–102.
Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117.
Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science, 36*(4), 500–522.
Senz, K. (2021). Is A/B testing effective? Evidence from 35,000 startups. Retrieved November 25, 2021, from https://hbswk.hbs.edu/item/is-ab-testing-effective-evidence-from-35000-startups. Accessed 30 Apr 2021.
Singh, V., Chen, S. S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. *International Journal of Information Management Data Insights*, *2*(2), 100094
Sridhar, S., & Fang, E. (2019). New vistas for marketing strategy: Digital, data-rich, and developing market (D3) environments. *Journal of the Academy of Marketing Science, 47*, 977–985.
Stourm, V., & Bax, E. (2017). Incorporating hidden costs of annoying ads in display auctions. *International Journal of Research in Marketing, 34*(3), 622–640.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second ed.). The MIT Press.
Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights, 1*(1), 100002.
Votto, A. M., Valecha, R., Najafirad, P., & Rao, H. R. (2021). Artificial intelligence in tactical human resource management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(2), 100047.
Wang, H., & Hong, M. (2019). Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach. *Electronic Commerce Research and Applications, 35*, 100852.
Youngmann, B., Yom-Tov, E., Gilad-Bachrach, R., & Karmon, D. (2021). Algorithmic copywriting: Automated generation of health-related advertisements to improve their performance. *Information Retrieval Journal, 24*(3), 205–239.
Zhang, Y., et al. (2022). A deep reinforcement learning based hyper-heuristic for combinatorial optimisation with uncertainties. *European Journal of Operational Research, 300*(2), 418–427.

Vinay Singh is a Data Science Manager at BASF SE, Germany and Research fellow at National Central University, Taiwan. He has over 15 years of industry experience in area of Intelligent systems and currently leads Industry 4.0 programs at BASF SE. His research interest is in business analytics and applied artificial Intelligence. Educationally he completed his Masters in Business Administration from Manheim Business School (Germany) and Doctorate from National Central University (Taiwan). He is also a visiting researcher in Siegen University, Germany. He actively engages in academic collaboration and research supervision in Manheim and Siegen with academia.

Brijesh Nanavati is a Global Expert of Analytics and Assessments in Customer Credit Risk Management at BASF Services Europe GmbH (Berlin). He is a Master of Economics and Management Science graduate of the Humboldt University of Berlin. His research interests include Business Analytics and Data Science within the domain of credit risk, payment behavior and e-commerce recommendation.

Arpan Kumar Kar holds a Chair Professorship in Indian Institute of Technology Delhi, India. He has a joint appointment in the Department of Management Studies and Yardi School of Information Systems. His research interests are in the domain of data science, digital transformation, internet ecosystems, social media and ICT-based public policy. He has authored over a 170 peer reviewed articles and edited 9 research monographs. He is the recipient of Research Excellence Award by Clarivate Analytics (Web of Science) for research impact between 2015 and 2020. He is the recipient of Basant Kumar Birla Distinguished Researcher Award for the highest count of ABDC A*/ABS 4/FT50 level publications in India between the period 2014 - 2019. He is the Editor in Chief of International Journal of Information Management Data Insights, published by Elsevier. He has undertaken over 40 research, consultancy and training projects from national and international organizations. He has received over 20 awards from reputed organizations for his research like IFIP, ACM, PMI, Elsevier, TCS, Ivey/Harvard, Birla, etc. Prior to joining IIT Delhi, he worked in IIM Rohtak, Cognizant Business Consulting and IBM India Research Laboratory.

Agam Gupta is Assistant Professor at Indian Institute of Technology Delhi, India. He is a fellow of the Indian Institute of Management Calcutta, and his areas of research include digital advertising, digitization, platform ecosystems, agent based simulations, and networks. His research has been published in top international journals such as Information Systems Journal, PlosOne, International Journal of Information Management, Computational Economics and Marketing Intelligence and Planning. He has undertaken sponsored research and advocacy projects from organizations like BASF and Ministry of Electronics and IT (CSC), Government of India.