Scholars Journal of Engineering and Technology

OPEN ACCESS

Abbreviated Key Title: Sch J Eng Tech
ISSN 2347-9523 (Print) | ISSN 2321-435X (Online)
Journal homepage: https://saspublishers.com

Harnessing Multi-Armed Bandits for Smarter Digital Marketing Decisions

Shashank Agarwal[1*], Gunjan Paliwal[2], Sumeer Basha Peta[1], Sriram Panyam[3], Praveen Gujar[4]

[1]Independent Researcher, Boston, MA, USA
[2]Independent Researcher, Seattle, WA, USA
[3]Chief Architect, Omlet, Sunnyvale, CA, USA
[4]Director, LinkedIn, Saratoga, CA, USA

DOI: https://doi.org/10.36347/sjet.2024.v12i10.002 | **Received:** 28.08.2024 | **Accepted:** 01.10.2024 | **Published:** 07.10.2024

***Corresponding author:** Shashank Agarwal
Independent Researcher, Boston, MA, USA

Abstract **Review Article**

In the current digital age, software programs and applications provide convenience and added value in a number of ways. Multi-armed Bandit algorithms (MAB) are a prime example of this; In capital markets, they assist in the adaptive design of trading strategies that adjust to market shifts and investor behavior. In e-commerce, MAB helps optimize product recommendations by learning customer preferences in real-time. In SaaS, MAB can optimize user experience by personalizing services or pricing models based on user behavior, continuously adjusting to maximize engagement or revenue. In cloud engineering MAB are indispensable for optimizing resource allocation based on user demand. The objective of this paper is to demonstrate the adaptability of MAB algorithms through an examination of their applications, specifically focusing on the Digital Marketing domain and the insight they offer for optimal decision-making. The article highlights the usefulness of major MAB algorithms in digital advertising, e-commerce content suggestion, SaaS and strategic pricing by carefully examining Thompson Sampling, UCB, Restless Bandit, and Structured Bandit as key types of MAB algorithms. The study emphasizes how these algorithms adjust to fluctuating conditions, balance a trade-off between exploration and exploitation, and eventually improve marketing strategies. This article aims to improve knowledge of MAB algorithms and encourage more research in this promising area by providing a detailed analysis of their applications.

Keywords: Multi-Armed Bandit, MAB Algorithms, Digital Marketing, E-Commerce, Content Recommendation, Saas, Digital Advertisement, Artificial Intelligence.

I. INTRODUCTION

In the contemporary digital environment, applications and programs add benefit and ease in a variety of ways, however many users are unaware of the core mechanics that underlie these functions. For example, complicated algorithms govern the interactive aspects of websites, an elementary but intricate component. Among these, Multi-armed Bandit (MAB) algorithms are notable for their ability to facilitate the decision-making process under precarious circumstances. MAB algorithms, a subclass of reinforcement learning, are driven by the challenge of choosing between several options having unknown rewards. In general, the MAB problem is about allocating limited resources optimally among varied options in order to yield maximum profits [1].

Multi-armed Bandit algorithms in capital markets help to design trading methods dynamically that can respond to market fluctuations and behavior of investors [2]. Marketing initiatives are sporadic, have similar characteristics, and must be assessed promptly [3]. MAB algorithms prove helpful for quick testing since they focus on interventions with the highest potential reward [4, 3]. They ideally tend to balance "exploration and exploitation" to minimize regret, incorporating client context to customize predictions [5, 6]. Thompson sampling serves as a prominent bandit method in which a campaign or initiative is chosen according to its probability of becoming optimal, based on past data. Fig.1 shows the comparison of the normalized performances of Thompson sampling over Upper Confidence Bound (UCB) MAB algorithms under various conditions.

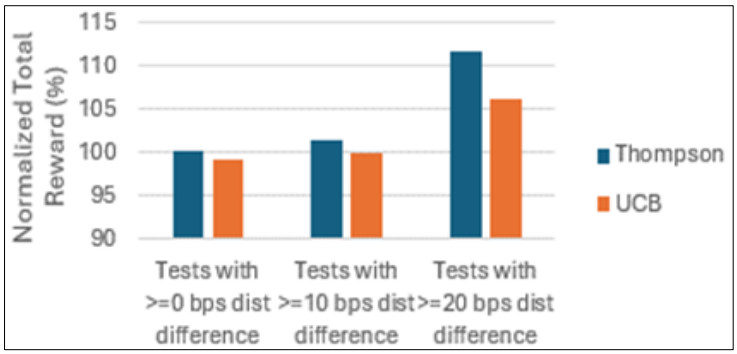

Fig. 1: Normalized Performances of Thompson sampling over Upper Confidence Bound (UCB) MAB algorithms under various conditions

Practically, this probability cannot be explicitly sampled. Rather, a single researcher selects the model's variables from their previous observations and selects a marketing approach that offers maximum reward [7]. While non-contextual bandit algorithms with undetected confounders have previously been constructed, the usual bandit setup implies unconfoundedness [8]. In order to eliminate bias from the estimates, inverse probability weighing as well as off-policy policy assessment techniques can be used [9, 10].

This paper specifically focuses on and reviews the applications of MAB algorithms in the digital marketing domain. The article is divided into three sections, the first one gives an overview of the fundamentals and basic types of MAB, the second highlights various applications and methods of distinct MAB algorithms, and finally, the third part provides suggestions for further research which reflects some of the research gaps that need to be fulfilled by the future researchers.

II. Fundamentals

Multi- armed bandit or MAB represent sequential tests with the objective of generating the greatest possible reward. A typical arrangement consists of "K" actions, or "arms." An arm is related to a quantity that is unknown and determines its "value". The aim is to select the arm that provides the most value while accumulating the greatest overall reward, as shown in Equation (1). The term "MAB or Multi-armed bandit" refers to a series of slot machines (also known as "one-armed bandits") with varying probabilities of reward [11]. The learner/experimenter's task is to select the slot machine that has the best possibility for a reward [12].

The learner/experimenter aims towards maximizing the cumulative reward across "n" rounds, i.e.

$$[\sum_{t=1}^{n} X_t = X_1 + X_2 + \cdots + X_n]$$

Where,
T = Each round
K = Possible arms/actions
X_t = Random reward

Alternatively, this goal can also be expressed as the reduction of regret, that can be described as follows: If the experimenter recognizes which arm is the "optimum" -$\sum_{t=1}^{n} X_t$, then regret = reward wasted by making improper choices = maximum cumulative reward in "n" rounds. Nevertheless, the environment conceals the reward of behaviors not chosen by the experimenter. Thus, the experimenter should gather information by frequently picking all actions. This is known as "exploration". Whenever the experimenter takes a sub-optimal/inappropriate action, it fails to achieve the cumulative reward, known as "exploitation". These two circumstances collectively form the "exploration-exploitation" issue. This problem is addressed by MAB algorithms [12].

III. Types of MAB Algorithms

A. UCB Algorithm

Operating on the concept of optimism in an environment of unpredictability, the UCB algorithm is renowned for its easy use and resilience. Considering a mix of projected rewards as well as a measure of uncertainty or variance within these estimations, it chooses arms. This strategy guarantees an appropriate balance between using "high-reward arms" and examining arms having considerable uncertainty. In stationary contexts, i.e. wherein the reward possibilities are constant across time, the effectiveness of the UCB method has been demonstrated [13].

At time *t* for selecting arm *i*, UCB can be calculated as:

$$[UCB_i(t) = \mu_i^{(t)} + c\sqrt{\frac{2\ln t}{n_i(t)}}]$$

Where,
c = Determinant controlling the trade-off between exploration and exploitation
$\mu_i^{(t)}$ = Arm *i*'s estimated mean reward up to time *t*.
$n_i(t)$ = Number of times up to time t that arm *i* has been played.
Therefore, the arm k having the greatest UCB value can be determined by using Equation:

$$[k = \arg\max_i UCB_i(t)]$$

B. Thompson Sampling

Thompson Sampling is a Bayesian technique that chooses "arms" determined by their probability to produce the best choice. It keeps a posterior distribution to each arm's reward and makes decisions based on these distributions. The approach performs exceptionally well, notably in contexts with irregular reward distributions. Jin *et al.*, conducted a thorough review of Thompson Sampling methods to measure exponential family bandits, demonstrating their usefulness in a variety of scenarios [14].

The following is a revised posterior distribution for the success probability of a Bernoulli bandit (successful or unsuccessful outcomes):

$$\theta_i \sim \text{Beta}(\alpha_i + \text{Successes}_i, \beta_i + \text{Failures}_i)$$

Where,
$\alpha_i$and β_i= Beta distribution parameters for arm *i*

Therefore, the arm k having the greatest sampled value can be determined by selecting a sample from each arm's posterior distribution:
k= arg max θ_i
Where,
k = Arm having the maximum value
θ_i = Sample
arg max = Argument of the maximum.

C. Restless Bandit

Restless Bandits function according to the assumption that the distribution of reward for every action varies over time, but the transformation should take place independently of the experimenter's actions. This dynamic nature creates intricacy because identifying the right action at an earlier stage is now insufficient, demanding constant adaptation. This type of MAB algorithm frequently includes systems for detecting and responding to modifications in reward distributions, requiring a balance between short-term benefits and long-lasting adaptability [15].

The standard model for formulating restless bandit is a Partially Observable Markov Decision Process (POMDP), in which the player's actions have no influence on how each arm's state changes over time.

At time *t*, the arm *i*'s state, $s_i(t)$, changes in accordance with a transition matrix P:

$$s_i(t+1) \sim P\big(s_i(t), a_i(t)\big)$$

Now, the expected reward can be determined by the current state as follows:

$$r_i(t) = \boldsymbol{E}\big[R\big(s_i(t), a_i(t)\big)\big]$$

Where,
r = Expected reward
i = Arm (action)
t = time
s = State of arm
R = Reward function
E = Expectation operator
a = Action taken

D. Structured Bandit

There is a noticeable or implicit pattern dictating how rewards are distributed across various behaviors/actions in the structured bandit problem. Structured bandit issues, as opposed to standard MAB problems, are more linked with comprehending the underpinning reward pattern. This comprehension enables improved decision-making to optimize immediate rewards and interpret the reward process itself, thus enhancing performance over the long run [16].

The expected reward is stated as follows if the actions and rewards have a known linear relationship:
$r_i(t) = x_i^{\top}\ \theta$

Where,
θ = Parameter vector to be estimated
x_i = Arm *i*'s feature vector
Regularized least squares may be used to estimate the parameter vector θ as follow:

$$\widehat{\theta^{(t)}} = (X^{\top}X + \lambda I)^{-1}X^{\top}y$$

Where,
λ = Regularization parameter
X = Matrix of feature vectors
y = Vector of observed rewards
Now, the arm having the highest expected reward, based on the current estimate $\theta\hat{}(t)$ is chosen as:
$k = \arg\max_i x_i^{\top}\,\theta^{(t)}$

IV. APPLICATIONS IN DIGITAL MARKETING

E. Digital Advertising

MAB is mainly employed in online advertising to maximize ad delivery techniques which not only raise click-through rates but further boost the overall performance of the placement of advertisements, resulting in higher advertising income as well as improved user experiences [2].

In the year 2021, the costs of digital advertising in the United States hit 189 billion dollars, representing a remarkable 35 percent increase over the previous year [17]. This situation ignited the machine learning community's attention for a variety of factors. Since it is totally digital, the effect of decisions can be accurately measured, effectively closing the informational loop between action and reward. A large percentage (approximately 93%) of total expenditures goes to three internet advertising formats which are search engine ads (40%), display advertisements (30%), and online video advertisements (23%). Leading platforms of advertising employ the identical fundamental approach to determine which commercials are displayed to an internet audience: if a user is suitable to view an advertisement, compatible advertisers participate in automatically programmed auctions. For each advertisement, the advertiser must

therefore opt for a target (keywords and customer profiles), a bidding price for auctions, as well as a maximum budget per day [18]. The purpose of MAB testing in digital advertising is to optimize revenue prospects by selecting an everyday mix of budgets and biddings for the entire holdings, upon which the entire budget is determined. To accomplish this, the experimenter uses the whole mix of budgets and bids as a bandit arm [19].

Using these algorithms, advertising platforms have the potential to balance exploitation—using proven, successful ad strategies—and exploration—trying out new advertisements and niche markets. In particular, to determine the ads that are more likely to capture a particular user's attention, MAB examines an individual's click actions and interaction feedback [2].

In two recent papers [20], 21], the everyday budget/bid optimization problem is framed as a MAB problem. Whenever the learner performs a certain combination, the writers of [20], observe that in addition to the portfolio's daily total of clicks and conversions, it also records the individual totals for each ad group. To put it another way, although the range of probable actions is combinatorial, the issue is manageable because the learner may get trained to correlate the budget and bids of one advertising campaign with the projected number of views provided it receives stronger feedback when compared to pure bandit feedback.

As a variation on typical MAB, Tran-Thanh presented the "budget-limited bandit" framework [22], in which decisions (arm pulls) are limited by budget and expenses. For practical uses where allocation of resources is crucial, such as web-based advertising and networked wireless sensors, this method has substantial implications.

F. Content Recommendation in E-Commerce

E-commerce content recommendations have been revolutionized by MAB algorithms, an important shift from conventional A/B testing. To improve the likelihood of presenting particular products to potential consumers, e-commerce uses these algorithms to personalize content recommendations. Enhancing consumer engagement and driving revenues is the primary goal of offering the most pertinent services or products. A/B testing divides traffic equally among choices for a predetermined period of time; in contrast, MAB algorithms continuously track how consumers respond to different recommendations made and adapt their suggestion approaches instantly.

For instance, a study examines ways to use MAB algorithms within e-commerce stores to reduce purchasing barriers and increase consumer satisfaction [23]. The investigators studied MAB algorithms employing synthetic datasets that replicate non-stationary consumer choices, the study discovered that the algorithms may maximize recommendations in a versatile approach. The investigators then performed more than a thousand trials employing former A/B testing records from an existing online purchasing website. The findings suggest that higher variations in efficiency rates of competitive recommendations result in higher compounding rewards for Multi-armed Bandit algorithms.

To address the problem of delayed rewards in the field of e-commerce, the authors of a study [1], created a "batch-updated" MAB algorithm to maximize digital content recommendations. This algorithm outperforms previous approaches in dealing with fluctuating customer demands, preferences, and buying frictions. MAB algorithms' excellent results prove that they are capable of increasing user satisfaction, rates of click-through, as well as conversion rates, all of which are critical e-commerce key performance indicators [1].

Significantly, the rapidly evolving process of learning associated with content recommendation technology adjusts to varying consumer preferences and constantly offers novel content. The MAB method promotes continual learning via customer interactions, allowing services to adapt recommendations in real-time. In addition, viewers using content recommendation services exhibit extremely specific attributes; hence, utilizing such user attributes as context can considerably improve the total reward of recommendations. For example, empirical studies verified the efficacy of a contextual bandit strategy for improving social media streaming recommendations [24]. Content recommendation merges with studies of communication, especially within the context of marketing recommendations.

G. Strategic Optimization of Pricing Models

Multi-armed bandit algorithms represent a valuable asset for strategic pricing model optimization. Standard pricing methods frequently underperform in highly volatile markets because of changing consumer preferences and competition environments. Conversely, MAB algorithms present a robust remedy by analyzing customer reactions and responses to various pricing models instantly [7]. This functionality enables continuous evaluation of multiple price ranges, gaining insights from customer buying patterns, adjusting pricing tactics, and eventually digital marketing. MAB algorithms' learning ability is critical for comprehending flexibility in pricing as well as consumer behavior.

Likewise, their ability to balance exploration and exploitation promises avenues to uncover better solutions while increasing direct business revenue by applying the most economical pricing strategies. In highly competitive marketplaces, enterprises may utilize Multi-armed bandit models to constantly adjust price points as a defense against unforeseen circumstances, while obtaining an improved comprehension of

competitors' approach to pricing as well as trends in the marketplace. This method enables price structures that are responsive to customer demands and trends, increasing sales volume, revenues, and market penetration [1-10].

The study [25], provides an example of how MAB algorithms are used in strategic pricing. It develops an elementary pricing strategy based on risk and tiered strategies. The study stresses the efficiency of the "Sliding-Window Thompson Sampling" method with the loan pricing problem represented as a MAB problem in this scenario, as proven by the algorithms' performance results. The researchers of this study concluded by confirming the "Sliding-Window Thompson Sampling" algorithm's effective performance in unpredictable situations, as well as its resiliency against preconceptions or assumptions within non-stationary case design.

H. Financial Market Trading

In the last few years, the interaction of deep learning with financial modeling has shifted toward the sequential selection of portfolios. The most important part of optimizing overall rewards is to achieve a balance between exploring new opportunities and exploiting existing resources and assets. In a study [2], the authors generated an online portfolio selection algorithm implementing a MAB strategy that leverages the correlations between different options for investment. They constructed a profitable investment approach by building orthogonal portfolios of different assets while combining this strategy together with the "upper-confidence-bound" bandits' approach. In the meantime, it incorporated risk concerns into the traditional MAB framework and suggested a new algorithm that balanced both risks and returns through the combination of an intuitive risk mitigation technique with asset filtering on the basis of the financial industry structure.

I. Boosting user engagement in SaaS

SaaS platforms are dynamic in nature - user preferences and behaviors rapidly shift. Here MAB algorithms help optimize user experience and revenue by personalizing service offerings or pricing models based on real-time data. For example, SaaS platforms offering project management tools can employ MAB algorithms to personalize feature rollouts and pricing tiers. Using simulated data that reflects user behavior, the algorithm identifies features most likely to drive engagement for different user segments. As behavior changes, the algorithm adjusts feature recommendations, improving user satisfaction and reducing churn. A study applying MAB algorithms to SaaS platforms analyzed historical A/B testing data and demonstrated that MAB consistently optimized interactions, driving higher engagement [28]. Metrics like daily active users and feature adoption showed significant improvements because the algorithm adapted in real-time to user preferences, leading to better outcomes than traditional A/B tests.

V. POTENTIAL DIRECTIONS FOR FUTURE RESEARCH

A number of promising domains exist for potential future MAB algorithm research. Certain possible directions are as follows:

Integrating deep learning methods with Multi-armed bandit algorithms for handling multidimensional and intricate data spaces presents one such avenue. This type of analysis may aid in dealing with complex and nonlinear data patterns that standard MAB algorithms fail to catch.

The MAB algorithms also serve as an opportunity to investigate how numerous learners interact with a similar setting. Specifically, studies in this area shall offer insights into competitive marketplace conditions and collaborative filtering structures. The examination of MAB with restricted communications is a perfect framework for prospective future research in this area [26]. Furthermore, future research might concentrate on designing MAB algorithms that enhance their adaptability across other fields. This lessens the need for domain-specific tailoring.

Developing competent contextual bandit algorithms having immediate feedback adaption is essential in settings having high fluidity and dynamism where conditions may alter rapidly. Research in this arena may contribute to the usability of contextual Multi-armed bandit algorithms in areas such as financial markets and real-time bidding platforms. For example, the study "Risk-averse Contextual Multi-armed Bandit Problem with Linear Payoffs" demonstrated the possibilities of applying this kind of method. Prospective studies may implement hybrid learning models, which combine MAB algorithms with additional artificial intelligence approaches such as supervised learning, to create hybrid models that capitalize on each other's benefits [27].

Table I: Comparison of the different types and applications of MAB algorithms in Digital Marketing

Type of MAB Algorithms	Major Features	Strengths	Weaknesses
Thompson Sampling	Bayesian approach; employs posterior distributions for making decisions	High adaptability, Performs exceptionally well in fluctuating (non-stationary) environments	Needs careful tuning of previous distributions
UCB	Optimism in an environment of unpredictability, Easily achieves a balance between exploration and exploitation	Ease of use, Efficiency in stationary environments	Less efficient in fluctuating (non-stationary) settings
Restless Bandit	Distribution of reward for each action varies over time	Adaptability in dynamic environments	Complex, Increased computation
Structured Bandit	Follows implicit pattern of how rewards are distributed across various actions	Performance optimization on long-term basis	Needs understanding of reward structure

VI. CONCLUSION

This paper underscores the crucial significance and adaptability of Multi-armed Bandit (MAB) algorithms within the field of digital marketing. These applications were most notable when dealing with decision-making under unpredictability in digital marketing. Some of the sub-areas of digital marketing where MAB algorithms have been implemented are strategic pricing, e-commerce content recommendation, and digital advertising. MAB algorithms are essential for contemporary marketing approaches due to the fact that they balance exploration and exploitation so well to improve performance and can adapt to fluctuating environments. However, improved adaptability in non-stationary settings and the incorporation of deep learning methods are two areas that the paper highlights as possible avenues for improvement. Therefore, future studies ought to concentrate on filling these shortcomings so that MAB algorithms can be optimally utilized in digital marketing to the next level.

AUTHOR BIOGRAPHY

Shashank Agarwal:

Shashank Agarwal is a data science expert whose experience cuts across various areas in experimentation science, artificial intelligence, brand analytics, predictive modeling, launch strategy, and multi-channel marketing in several Fortune 500 companies such as CVS Health, AbbVie, and IQVIA. Additionally, he holds a Master of Science in Engineering Management from The Johns Hopkins University, USA.

Gunjan Paliwal:

Gunjan Paliwal, an alum of MIT, is a distinguished tech professional with over a decade of experience in product marketing and development at industry leaders like Newell Brands, Sears, Microsoft and Meta. Specializing in AdTech, e-commerce marketplaces, and ML/AI-based products.

Sumeer Peta:

Sumeer Peta is a technology expert who is currently serving as a Lead Architect at CVS Health. In his 15 year career, he has led significant software engineering and data science projects for multiple Fortune 500 companies like Adobe, AT&T, and Kaiser Permanente to name a few.

Sriram Panyam:

Sriram Panyam is an world renowned engineering leader with a track record for developing technical organizations within major tech firms like Google, LinkedIn, and Amazon. His expertise spans large-scale systems, cloud platforms, AI/ML and data analytics. As a strategic and highly technical leader, he has led programs impacting billions of users worldwide, fostered innovation cultures, and empowered deeply-technical engineering.

Praveen Gujar:

Praveen Gujar stands at the forefront of product innovation, boasting an illustrious career close to two decades and marked by the successful launch of cutting-edge Enterprise Data products in Digital AdTech. A stalwart in the tech industry, Praveen has left his mark on tech giants like LinkedIn, Amazon, and Twitter, spearheading initiatives that have catapulted these companies into multi-billion dollar echelons.

REFERENCES

1. Li, H. (2024). Expanding the Horizon: Diverse Applications and Insights from Multi-Armed Bandit Algorithms. *Highlights in Science, Engineering and Technology*, *94*, 147-155.
2. Wu, J. (2024). In-depth Exploration and Implementation of Multi-Armed Bandit Models Across Diverse Fields. *Highlights in Science, Engineering and Technology*, *94*, 201-205.
3. Hill, D. N., Nassif, H., Liu, Y., Iyer, A., & Vishwanathan, S. V. N. (2017, August). An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1813-1821).
4. Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-

armed bandit problems. *Foundations and Trends® in Machine Learning*, *5*(1), 1-122.
5. Agrawal, S., & Goyal, N. (2013, May). Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning* (pp. 127-135). PMLR.
6. Dani, V., Hayes, T. P., & Kakade, S. M. (2008, July). Stochastic Linear Optimization under Bandit Feedback. In *COLT* (Vol. 2, p. 3).
7. Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, *24*.
8. Bareinboim, E., Forney, A., & Pearl, J. (2015). Bandits with unobserved confounders: A causal approach. *Advances in Neural Information Processing Systems*, *28*.
9. Li, L., Chen, S., Kleban, J., & Gupta, A. (2015, May). Counterfactual estimation and optimization of click metrics in search engines: A case study. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 929-934).
10. Swaminathan, A., & Joachims, T. (2015, June). Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning* (pp. 814-823). PMLR.
11. Scott, S. L. (2015). Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, *31*(1), 37-45.
12. Genalti, G. (2020). A multi-armed bandit approach to dynamic pricing.
13. Han, Y. (2024). Comparative Evaluation, Challenges, and Diverse Applications of Multi-Armed Bandit Algorithms. *Highlights in Science, Engineering and Technology*, *94*, 206-210.
14. Jin, T., Xu, P., Xiao, X., & Anandkumar, A. (2022). Finite-time regret of thompson sampling algorithms for exponential family multi-armed bandits. *Advances in Neural Information Processing Systems*, *35*, 38475-38487.
15. Burtini, G., Loeppky, J., & Lawrence, R. (2015, April). Improving online marketing experiments with drifting multi-armed bandits. In *International Conference on Enterprise Information Systems* (Vol. 2, pp. 630-636). SCITEPRESS.
16. Zhang, Q. (2024). "Real-world applications of bandit algorithms: Insights and innovations," *Transactions on Computer Science and Intelligent Systems Research*, *5*, 753-758.
17. PwC, "IAB Internet advertising revenue report, Full year 2021 results," 2022.
18. PwC, "IAB Internet advertising revenue report, Full year 2022 results," 2023.
19. Gigli, M., & Stella, F. (2024). Multi-armed bandits for performance marketing. *International Journal of Data Science and Analytics*, 1-15.
20. Nuara, A., Trovò, F., Gatti, N., & Restelli, M. (2018). "A combinatorial-bandit algorithm for the online joint bid/budget optimization of pay-per-click advertising campaigns," in Proc. 32nd AAAI Conf. Artificial Intelligence.
21. Nuara, A., Trovò, F., Gatti, N., & Restelli, M. (2022). Online joint bid/daily budget optimization of internet advertising campaigns. *Artificial Intelligence*, *305*, 103663.
22. Tran-Thanh, L. (2012). *Budget-limited multi-armed bandits* (Doctoral dissertation, University of Southampton).
23. Kojima, M. (2022). Application of multi-armed bandits to model-assisted designs for dose-finding clinical trials. *arXiv preprint arXiv:2201.05268*.
24. Gisselbrecht, T., Lamprier, S., & Gallinari, P. (2016). Dynamic data capture from social media streams: A contextual bandit approach. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 10, No. 1, pp. 131-140).
25. Gan, M., & Kwon, O. C. (2022). A knowledge-enhanced contextual bandit approach for personalized recommendation in dynamic domains. *Knowledge-Based Systems*, *251*, 109158.
26. Lin, Y., Wang, Y., & Zhou, E. (2023). Risk-averse contextual multi-armed bandit problem with linear payoffs. *Journal of Systems Science and Systems Engineering*, *32*(3), 267-288.
27. Sawant, N., Namballa, C. B., Sadagopan, N., & Nassif, H. (2018). Contextual multi-armed bandits for causal marketing. *arXiv preprint arXiv:1810.01859*.
28. Kaukanen, M. (2020). Evaluating the impacts of machine learning to the future of A/B testing.