Journal of Applied Data Sciences
Vol. 2, No. 2, May 2021, pp. 14-26

ISSN 2723-6471

Comparing Epsilon Greedy and Thompson Sampling model for Multi-Armed Bandit algorithm on Marketing Dataset

Izzatul Umami [1,*] & Lailia Rahmawati [2]

[a] Darul Ulum University, Indonesia
[1] izzatul.umami@undar.ac.id*; [2] lailia.rahmawati@undar.ac.id;
* corresponding author

(Received July 1, 2019 Revised October 21, 2019 Accepted October 29, 2019, Available online October 29, 2019)

Abstract

Background, A/B checking is a regular measure in many marketing procedures for ecommerce companies. Through well-designed A/B research, advertisers can gain insight about when and how marketing efforts can be maximized and active promotions driven. In practical terms, standard A/B experimentation makes less money relative to more advanced machine learning methods. *Purposes,* in order to examine the current A/B testing state, identify some popular machine learning algorithms (multi-arm bandits) which are used to optimize A/B testing, and then explain the output in some standard marketing cases of these algorithms. *Methodology,* In this study, the state of A/B testing have been addressed, some typical A/B learning algorithms (Multi-Arms Bandits) like Thompson Sampling, Epsilon Greedy and UCB-1 will be implemented and compared used to optimize A/B testing are described and comparable. As a result, UCB-1 and Thompson Sampling, be an exceptional winner to optimize payouts in this situation. Because it showed more effective results, without losing experimentation and statistical variations, to maximize total payouts. Based on its accuracy and strong tolerance to noise on the results, UCB-1 is the right option for MAB for a low base conversion, a limited impact size scenario.

Keywords: A/B, Epsilon Greedy, Thompson Sampling, E-Commerce, Machine Learning;

1. Introduction

The internet has changed our lives and our way of interacting with friends and doing our company. The Internet has also changed the activities of marketing, advertisement and promotion. The Internet's effect on brand value is also very strong. While many more people browse the Internet, the Internet has been utilized by strategists to build a strategic edge. By helping to create a reputation on the online platform, it shifted the competitive landscape. New businesses that start e-commerce Websites are trying to raise brand awareness pushing retailers to develop brand recognition across e-commerce platforms in the online marketplace. The plan allowed them to get current brand properties online and to replace those assets with competitive eCommerce programs to create new brands.

A/B checking is the standard phase in many marketing processes of e-commerce companies. Marketers may obtain information about when and how to optimize their marketing initiative and guide effective strategies through well-designed A/B measures. But in realistic terms, standard A/B research generates less money relative to more advanced approaches to machine learning[1]. We will examine the current A/B testing state, identify some popular machine learning algorithms (multi-arm bandits) which are used to optimize A/B testing, and then explain the output in some standard marketing cases of these algorithms[2-3].

2. Literature Review

In regular A/B tests we tried to calculate the possibility that one campaign variant was more successful than another while checking the probability that our indicators were incorrect - or we thought a winner was there when it was not there or what we overlooked when a winning variant was found. Testing of a/b would take into consideration two key values in order to perform precise measurements: predictive strength and statistical importance.

In reality, statistical strength is the likelihood that the experiment detects an effect when an effect happens (where one is greater than the other), and statistical value tests our faith that the effect we test exists[4]. However, analyzing these two concepts in terms of trust intervals will be far more intuitive [5].

Statistical research requires survey data to measure or decide a statistical population. In a concrete context of a two-sample comparison, the goal is to determine if in both sub-populations the average results of the several attributes obtained for individuals vary[6]. For example, in order to evaluate the null hypothesis of the mean results of men and women in a test being different, a sample is taken of men and women, experiments are performed, and the mean score of one category is correlated with another using statistical tests such as z-tested two samples [7]. The strength of the research is that the test is likely to show a statistically meaningful discrepancy between men and women depending on the real difference of size in the two populations [8].

We calculate the conversion rate for each variant in certain tests, but we know that the average conversion rate just estimates the "real" one.[9] We can be more or less optimistic with the estimated value based on the amount of measurements we are producing, and we can reflect this confidence with the interval in which true value is found. For example, if we claim 0.032 +/- 0.002 at 95% trust, we're 95% positive that the real rate of conversion is 0.030 and 0.034[10]. In reality, we would aim for trust periods which do not overlap, because the real conversion rate is more than 95%.

However, it can take some time to wait for different periods. In the realistic scenario, to identify 10 percent increases in click conversion (0.03 * 0.1= 0.003) with uniform statistical capacity and significance thresholds, a sample range of at least 106,000 cumulative contacts will be required for an average click rate of 0.03 (53,000 contacts per variant). It is possible to gather such several data from a few days to many months, depending on traffic, and whether we have further variations, lower conversion rates or lower impact sizes, the collection time could be even longer [11].

3. Methodology

The MAB algorithm can be viewed as an alternative A/B test which balances use and study during the study process [12]. The MAB approach uses the experiment findings to add more contacts to the low performance version, thus allocating less traffic to the poor performance variant [13]. Theoretically, a multi-strategy algorithm can yield higher results overall (and fewer regret), but enable data to be collected about how consumers engage with various campaign variations [14].

There are many MAB algorithms, each of which favour experimentation to a particular level [15]. Epsilon Selfish, Thompson Sampling and Upper Confidence Bound 1 (UCB-1) [16-17] are the most common three. In this article each of these MAB algorithms is addressed separately, accompanied by a contrast between their actions and the A/B test setup under various experimental conditions. In the next part, one of the simplified simulation settings that is more detailed below demonstrates the action of the algorithm: the simulation of two low-difference variants.

3.1. Epsilon Greedy

As the name says, Epsilon Greedy is the greedy of the three MAB algorithms. The constant β (value between 0 and 1) is chosen by the consumer before the experiment starts in the Epsilon Greedy experiment. The randomly chosen version is identified in time as contacts are allocated to various variants in the campaign. Another 1-μ time is chosen for the variant with the maximum known yield. The stronger μ, the more experimentation is assisted by this algorithm. μ is set to 0.1 for all of our instances.

The findings of the simulation are seen in Figure 1. The thinner traces are derived from the 10 simulations chosen randomly and the thicker streaks are an average of 1000. The left panel displays the distribution of contacts between two variants and, during the simulation, the right panel indicates the real exchange rate for each model.

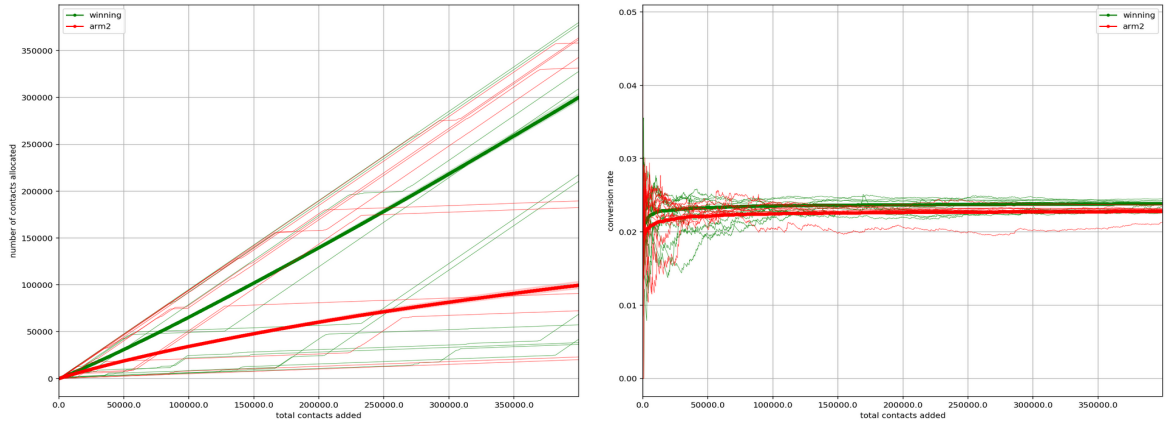


Fig. 1. The results of a two-variant simulation of Epsilon Greedy

3.2. Thompson Sampling

Thompson on the other side, sampling is a more principled method that in null situations will yield more balanced outcomes. For each vary, we establish a likelihood distribution (most frequently, for statistical purposes, a beta distribution) of the real success rate using the effects observed. The beta distribution is defined in probability theory and statistics as a family of continuous probability distributions at the interval [0,1] parameterised by the two positive form parameters α and β which appear as random variable exponents and regulate the form of the distribution. The Dirichlet distribution is considered a generalization to several variables. The beta distribution has been used to model the action of the random variables restricted in different scientific disciplines at intervals of finite duration. We sampled one potential success rate with any new communication from the beta distribution corresponding to each variant and allocated the best sample success rate to the variant. The more evidence we observe, the more trust we have in the true performance rate and the sample success rate would be similar to the real ratio as we gather more data.

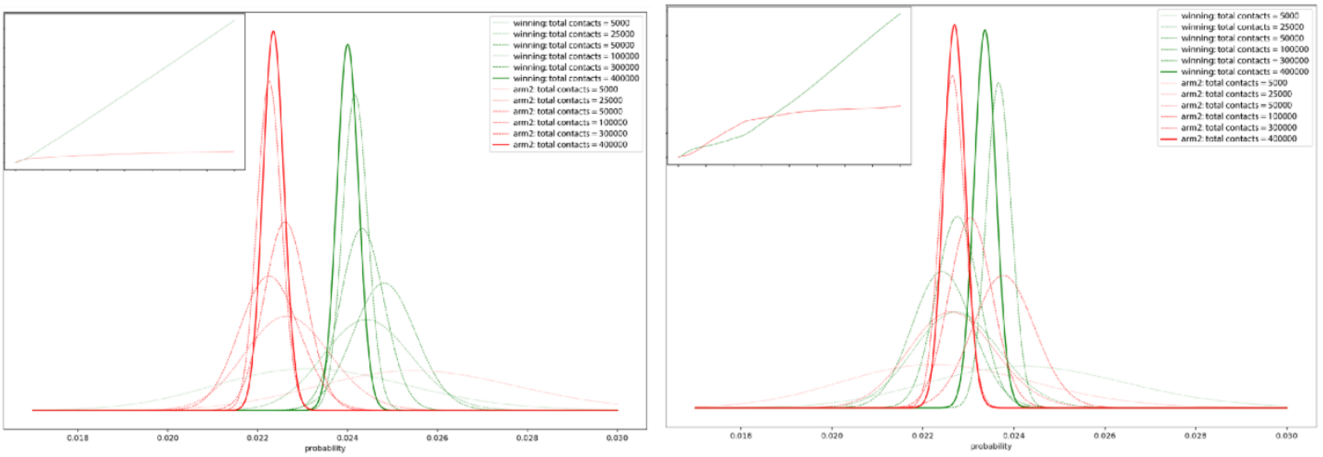


Fig. 2. Beta distribution probability density distributions for two separate simulations at different points of the experiment.

The parameters (α and β) of each beta distribution are default to 1 and 1 and result in the same large beta distribution for all versions, implying no prejudice (as we have no evidence to suggest which variant might win). The distribution is modified to provide the latest findings as we obtain further data (with an α = n transformed and β = n not converted) and the likelihood densities start to concentrate around the mean results. The further data points, the greater the likelihood density function (see Figure 2). If the behaviour of a variant has previously been firmly believed, we should adjust α and β to reflect the previous distribution before the experiment starts. Below (Figure 3)

ISSN 2723-6471

are the typical Thompson Sampling effects of our 2-variant simulation. Like Epsilon Greedy (Fig 1), individual Thompson Sampling simulations during the initial part of the experiment deviated slightly from their mean behaviour. But the participant simulation behaviour becomes more stable and comparable in a later stage of the experiment to the average outcome.

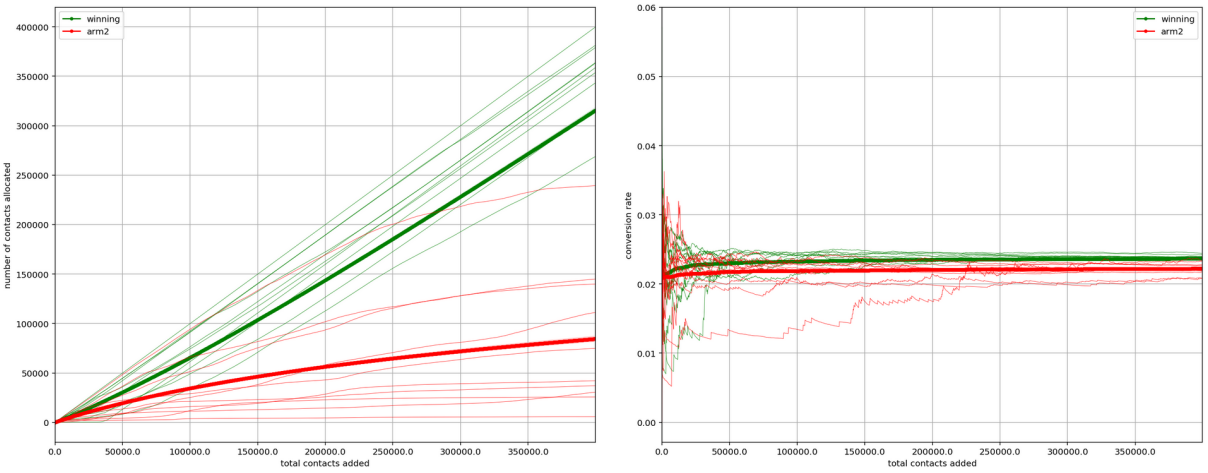


Fig. 3. A two-variant Thompson Sampling simulation yielded the following results.

3.2.1. UCB-1

For any version of the campaign, we can define the upper trust band (UCB), which is our highest estimation of the potential performance. The algorithm assigns the maximum UCB contacts to the version. The UCB is dependent on the average product of the variation and the amount of contacts assigned to the version with the equation:

$$\text{UCB} = \overline{x_j} + \sqrt{2\ log\ t/n_j},$$

where

$\overline{x_j}$ = the average payoff at the jth step

t = total number of contacts that have entered the experiment

n_j = total number of contacts allocated to a particular variant

Fig. 4. UCB Algorithm

If the equation indicates, the UCB score for a variant declines when the amount of contacts joining the variant rises, such that even though an average variant profits out, the UCB might be marginally less than the less explored variant. This allows the algorithm to combine the exploitation of unconventional alternatives with the use of winning variants.

The findings of a two-variant simulation (0.024 vs. 0.023 translation rate) using UCB-1 are again seen below (Figure 4). UCB-1 is much closer to the additive assignment (Figure 3) which is much more cautious than Epsilon Selfish (Figure 1). The algorithm was basically pure experimentation at the first step of the experiment (the first 30,000 data points), distributing exactly the same amount of contacts between the two weapons. Like the proportional allocation, the UCB-1 effects are reasonably stable and mimic the mean outcomes of individual simulations. That shows once again that the algorithm can combine discovery and usage in a single experiment, refine the assignment of contacts to discover the true winning variant and then use it.

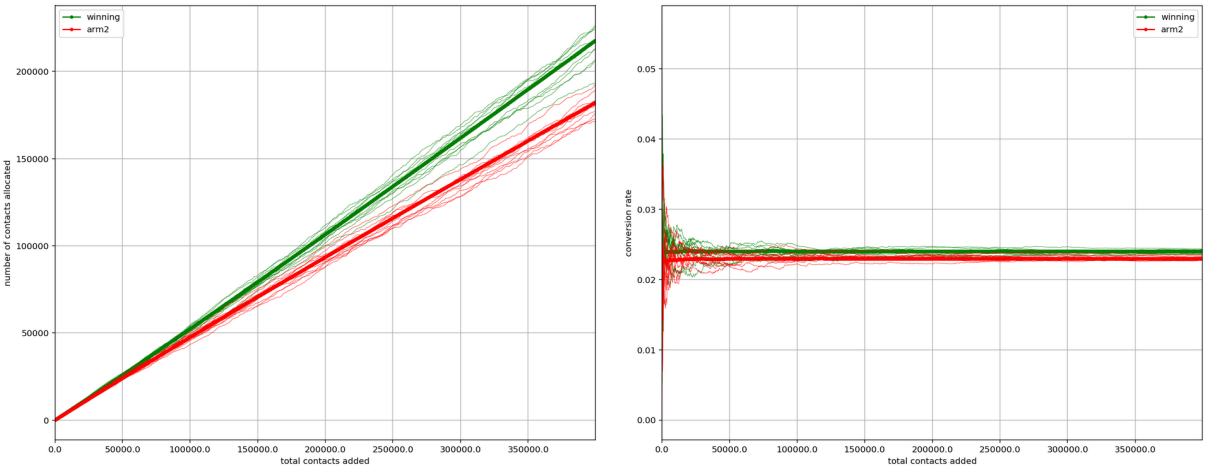


Fig. 4. A two-variant UCB-1 simulation produced the following findings.

3.3. Model Configuration

Now that we have an understanding of how each algorithm operates, our actions can be contrasted to the initial A / B test setup. We do this by preparing the following five simulations.

The first 3 simulations include five variants, each with a fixed conversion ratio below. For testing how this algorithm performs in practical scenarios, in each simulation we delegate the actual conversion rates for variants to fit what is usually used in email marketing campaigns. We also ran two iterations with two additional variations of the simulation conversion rate. Table 1 lists all conversion ratios used in the simulation.

Table. 1. Conversion rates used in simulations

Sim #1: Open Rate		Sim #2: High Difference Conversion Rate		Sim #3: Low Difference Conversion Rate		Sim #4: 2-variant High Diff Conversion		Sim #5: 2-variant Low Difference Conversion	
Winning:	0.24	Winning:	0.06	Winning:	0.024	Winning:	0.05	Winning:	0.025
Arm2:	0.22	Arm2:	0.04	Arm2:	0.022	Arm2:	0.01	Arm2:	0.020
Arm3:	0.21	Arm3:	0.01	Arm3:	0.021				
Arm4:	0.215	Arm4:	0.03	Arm4:	0.0215				
Arm5:	0.23	Arm5:	0.025	Arm5:	0.023				

We rendered an initial allocation phase for all simulations, with 200 contacts distributed equally between five variants (because there were no previous results). We also introduced another 200 contacts to the experiment for each allocation stage based on a particular MAB algorithm. We ran a total of 400 phases, repeating 1000 times for 2-variant simulations, open-speed simulation and high-difference transfer simulations. For the conversion simulation with low differences, we conducted 1000 steps and 2000 iterations to evaluate the action of the algorithm better. In both instances, our concept of payoff is binary conversion, and the conversion rate is what we equate with the variants. If we use other concepts of payment such as sales, the findings do not alter substantially. μ is set to 0.10 for all Epsilon Greedy simulations. We monitored the touch point, overall output and statistical significance for each

assignment phase (p-value of the 2-proportion z test for 2 variants, and the Chi-square contingency test for more than 3 variants).

4. Method Result

4.1. Contact Allocation

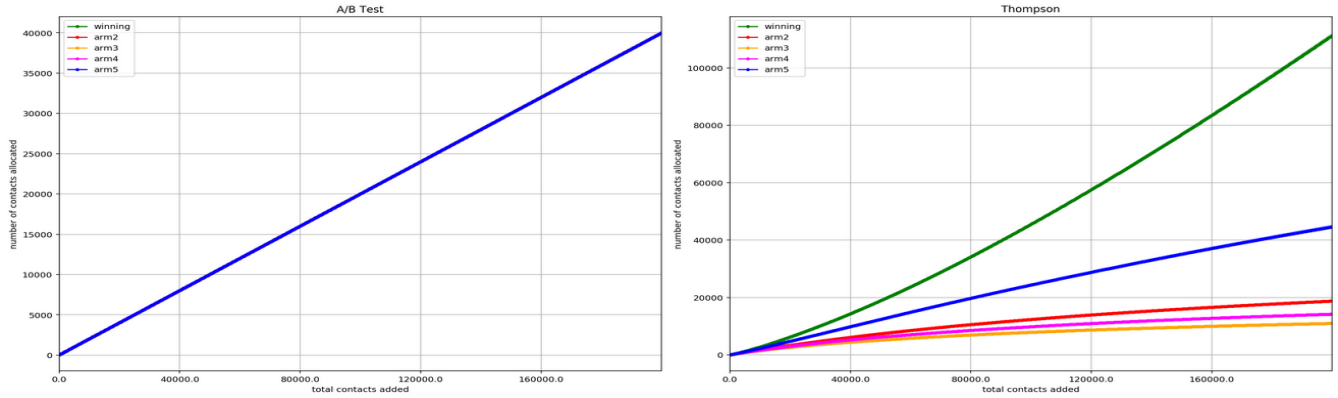


Fig. 5. A/B and Thompson Sampling testing

UCB-1 allocates the more restrictive contacts (most like the same allocation) for models of minimal pay differentials (Figure 5, Figure 6), as the variation in results is very tiny. Thompson Sampling and Epsilon Greedy is even more voracious, allocating twice as many contacts as the UCB-1. In the first third of the experiment, Thompson Sampling was significantly more conservative than Epsilon Greedy in low-conversion 5-variant simulation, but ultimately captured.

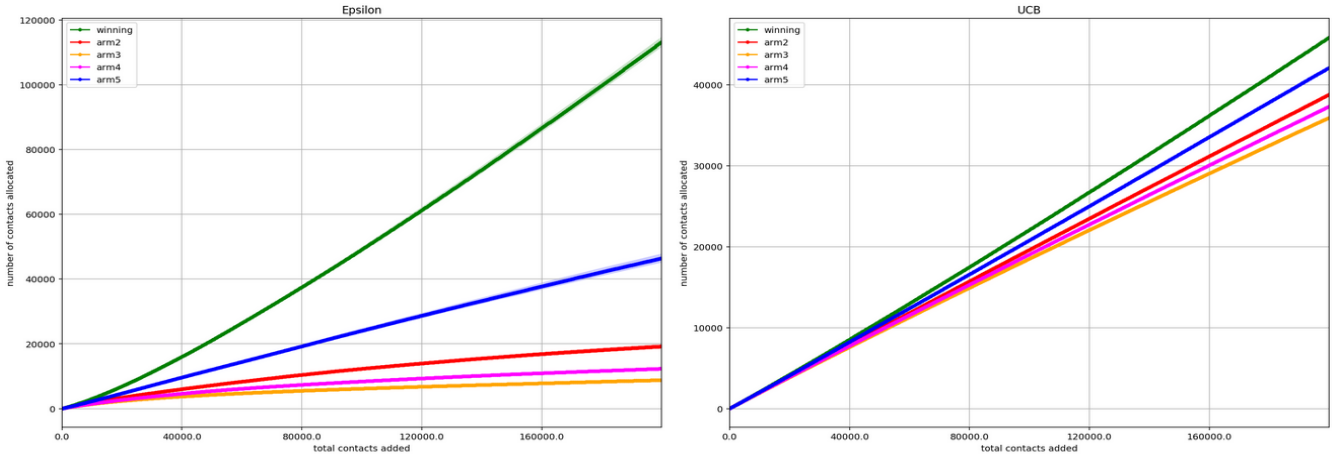


Fig. 6. Low-Difference Conversion Rate Simulation's average communication allocation

ISSN 2723-6471

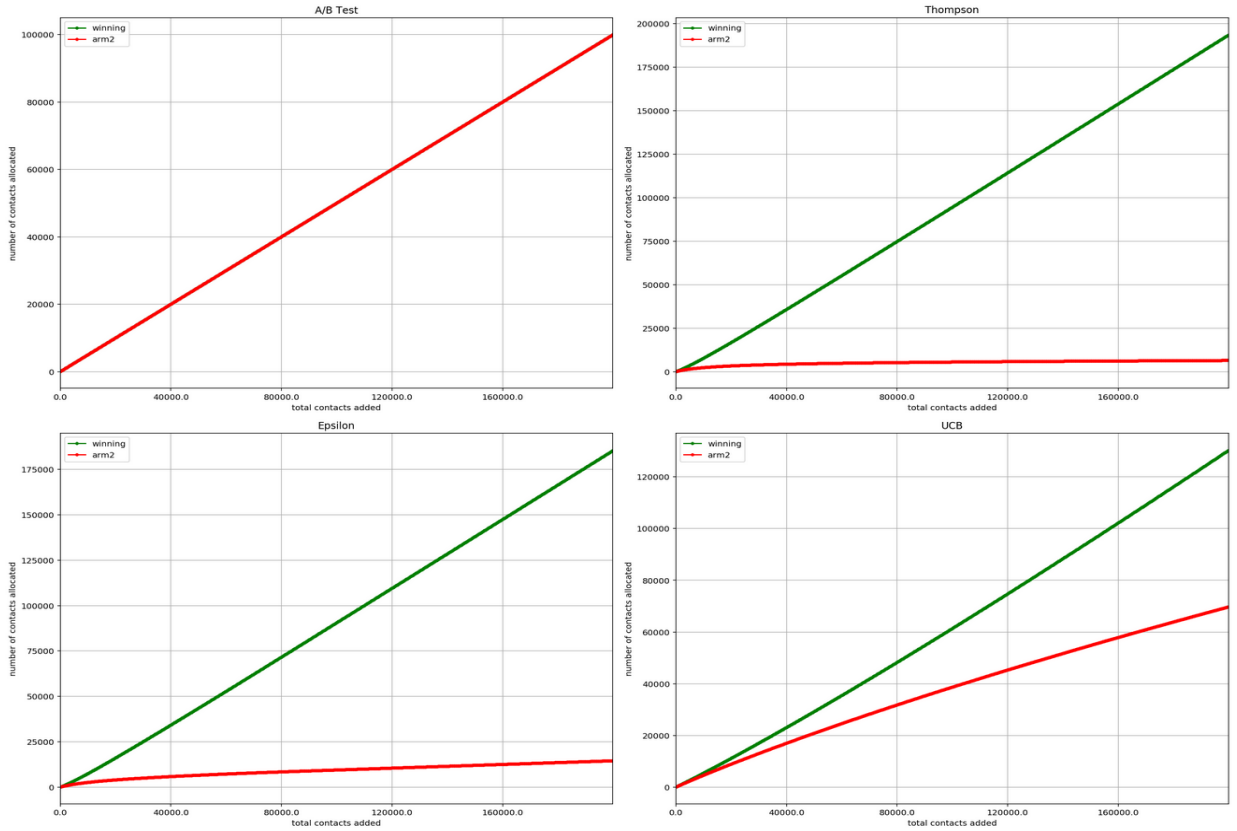


Fig. 7. 2-variant touch allocation average Simulation of Low-Differences

UCB-1 continues to be the most restrictive algorithm for simulating natural conversion speeds (Figure 7), and Thompson sampling the most voracious. Interestingly, in this simulation, the Thompson sample is much more voracious than the low-differences simulation. This illustrates Thompson Sampling's willingness to combine discovery and extraction and encourages this in situations of simple and readily observable pay variations. In one of the three cases, at the end of the simulation Thompson Sampling assigned about twice as many contacts to the winning version as UCB-1.

ISSN 2723-6471

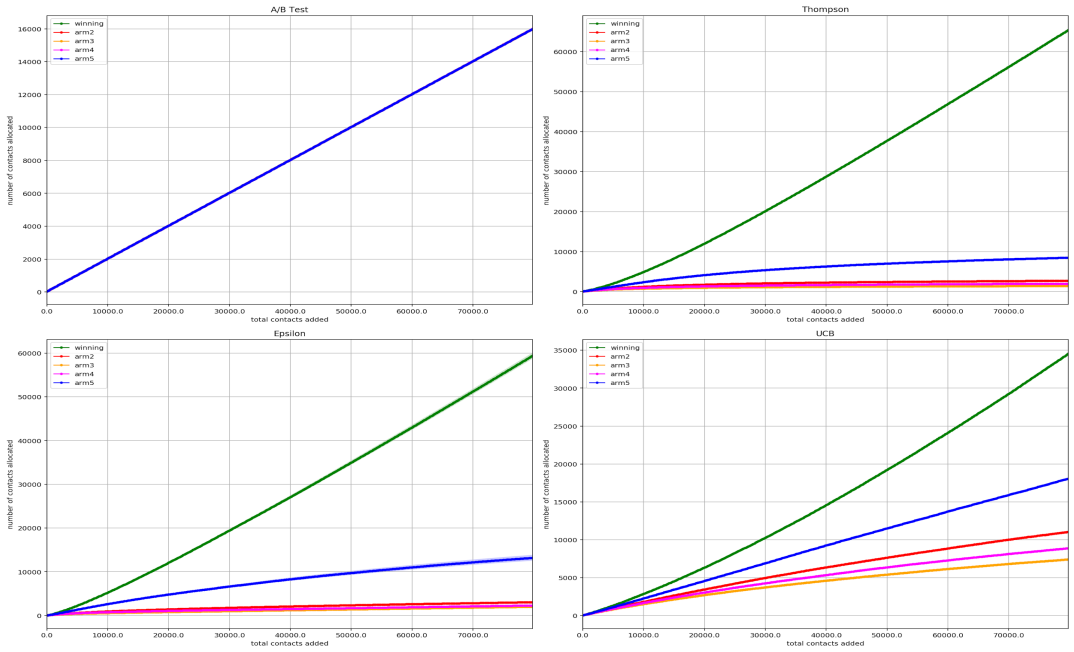


Fig. 8. Model Result after 1st training

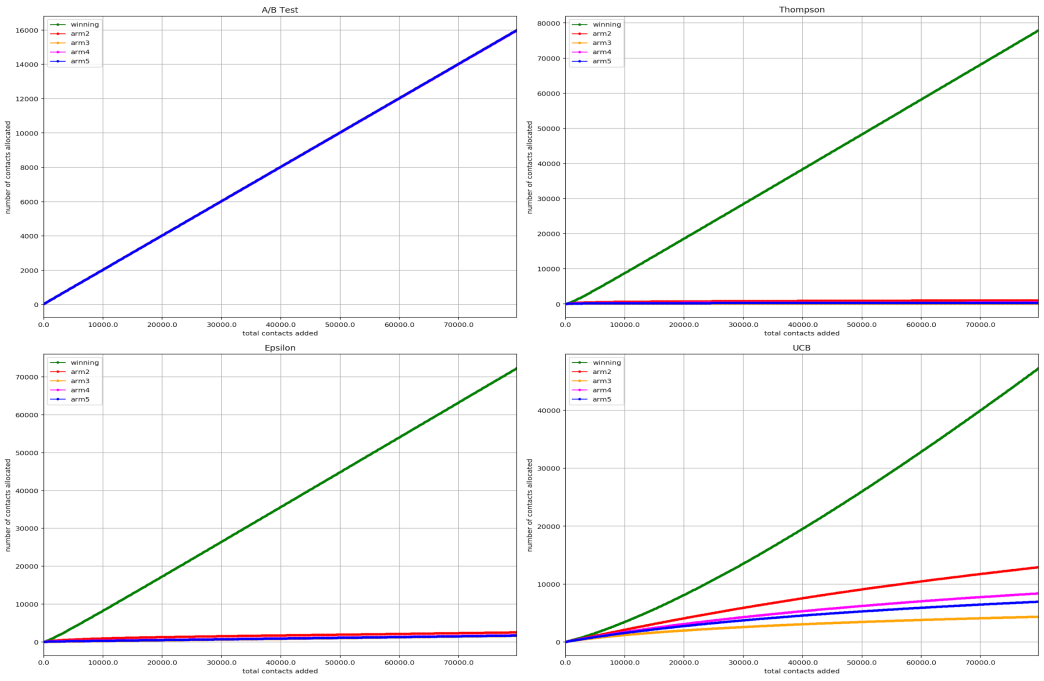


Fig. 9. Average contact allocations from high-difference conversion rate simulations

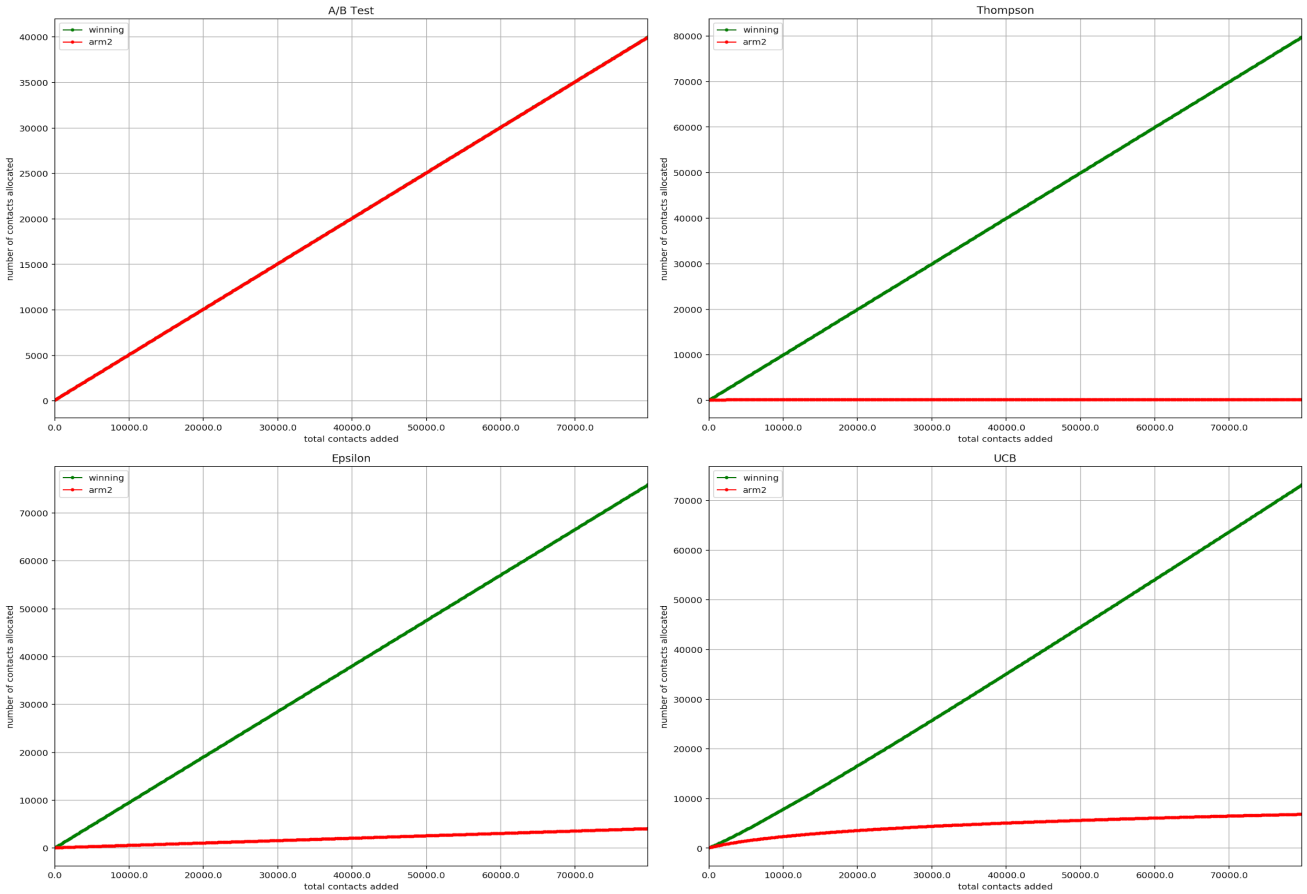
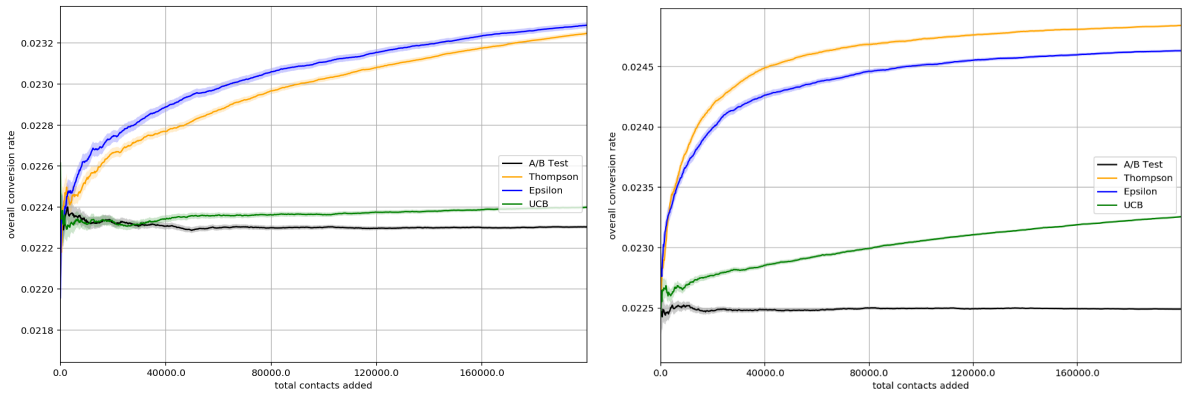


Fig. 10. Average interaction allocations from two-variant high-difference simulations

4.2. Overall Payout

As predicted, all three algorithms received higher average payouts for all simulations than the A/B test set-up, and higher output algorithms. Epsilon Greedy exceeds other algorithms in the hardest situation with a predefined μ0 (low conversion variant 5). In the other four cases, Epsilon Greedy got a higher payout in the beginning of the evaluation, but Thompson Sampling regularly overtook Epsilon Greedy, was greedy as further data points were obtained and had a higher payout. This illustrates again how Thompson Sampling encourages experimentation in the early stages of the experiment in the face of confusion, while also enabling profitable usage during later experimental phases.



ISSN 2723-6471

Figure. 11. Total payout for LDS simulations (left-hand) & Total payout rate for two variable LDS simulations (Right)

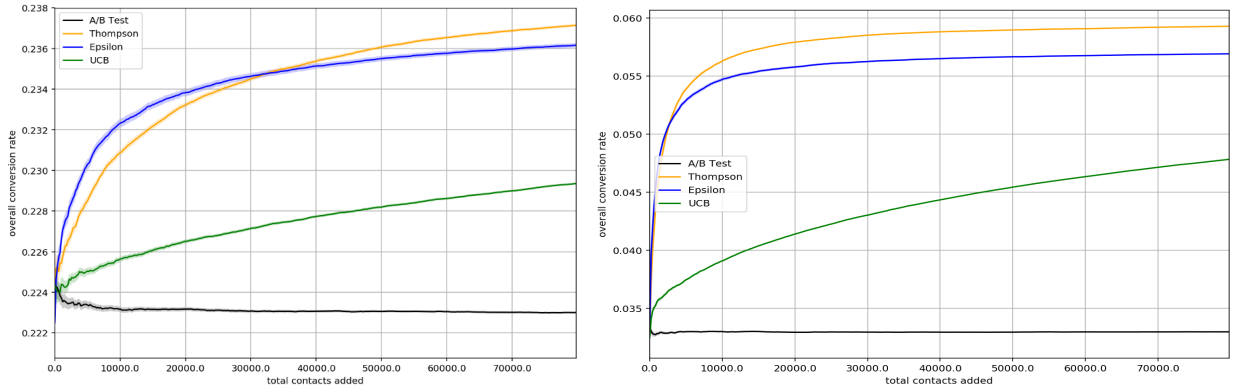


Figure. 12. Open-rate simulation overall payout (Left) & high-differential conversion simulation overall payout (Right)

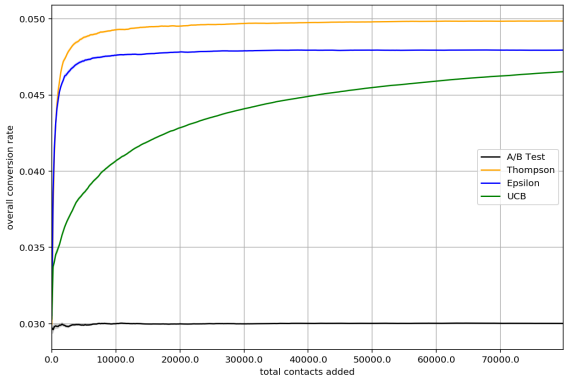


Figure. 13. Absolute payout for two variant simulation with high-difference

4.3. Statistical Significance

UCB-1 reached statistical significance ($p<0.05$) at the same period with all simulations (total number of contacts added), when the A/B test was finished, but there were more contacts assigned to the winning version and therefore many more data points for the losing variant. Interestingly, in the initial step of the trial, the Thompson sampling demonstrated a quicker reduction in p-values but slowed down and was eventually behind virtual UCB1 and A/B checking where the p-value was between 0.1 and 0.05. This may be attributed to the fact that the sampling of Thompson is progressively selfish with a different beta distribution (which usually results in lower p-values). In both instances, though, Epsilon greed suffers from inadequate evidence on lack of variance and is not statistically significant unless two times as many data points are found. For the simulation of low-difference transfer rates (Figure 15), also at the end of the trial, the Epsilon Greedy algorithm could not achieve statistical significance.

ISSN 2723-6471

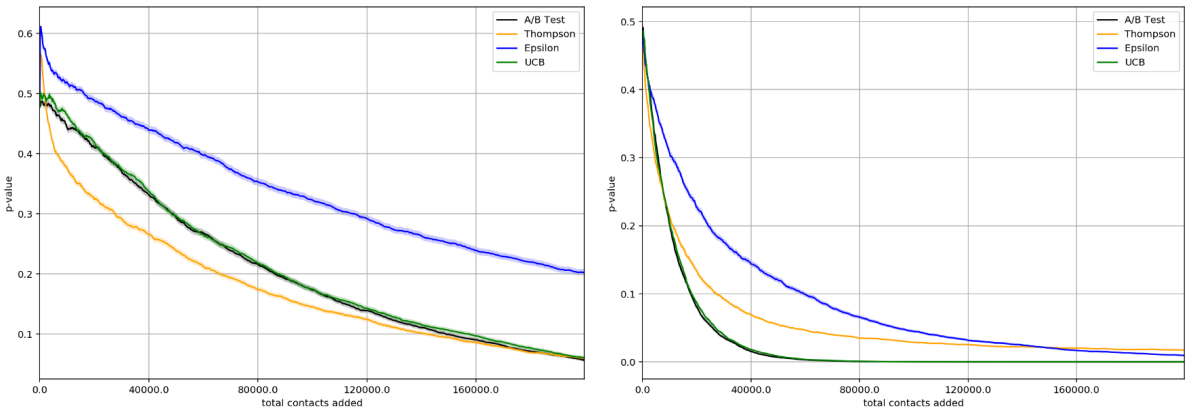


Figure. 14. Statistical value for low-difference transfer simulations (left) and low-difference 2-variant simulations (right) (Right)

Both algorithms gain statistical significance over the course of the experiment in high-difference simulations. Similar to the above findings, the UCB-1 and A/B experiments were $p<0.05$ with the same volume of data. The Thompson sampling was quite behind and Epsilon Greedy took much of the data to gain statistical significance.

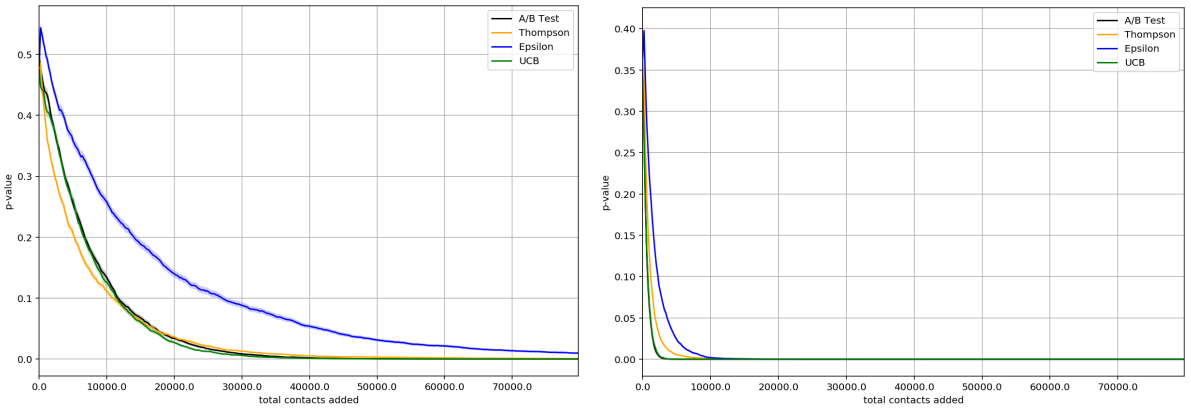


Figure. 15. Statistical value for open rate simulations (left) and high-difference transfer simulations (right)

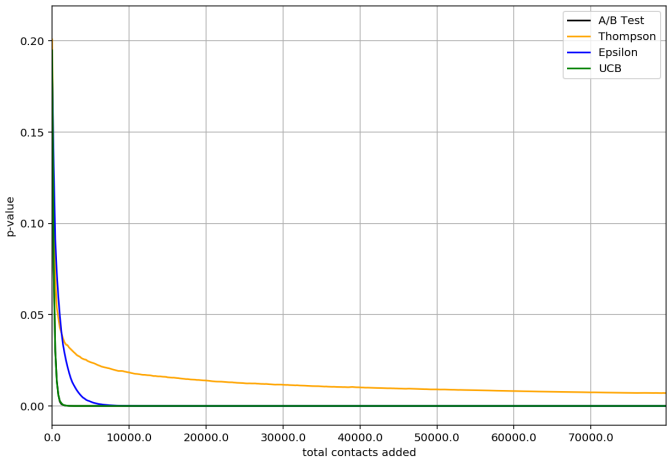


Figure. 16. Statistical significance for high-difference 2-variant simulations

5. Conclusion

With these simulation findings, we have shown that the MAB algorithm is often a more desirable option to A/B testing when randomized, regulated trials are needed. The preference of a specific algorithm relies on whether the consumer prioritizes benefit or data processing, size and length of the data. experience. experiment.

In rare cases where the variations between the variants under test are understood to be sufficiently high, each algorithm will display statistically meaningful differences between the variants at many data points. Thompson Sampling, or the gullible Epsilon Selfish algorithm, allows an impressive winner in optimizing payouts.

Thompson Sampling and UCB-1 have in all cases been willing, without losing experimentation and statistical variations, to maximize total payouts. UCB-1 results in assignments much like A/B experiments, whereas Thompson is better suited to improve total long-term outcomes. UCB-1 often worked more reliably in individual studies than Thompson Sampling, where the random sampling of the algorithm contributed to more noise. With the minor variant variation, a characteristic of the A/B test findings we have seen in the past, the UCB-1 appears to be very cautious in relation to Thompson Sampling and Epsilon Greedy (at μ=0.1). When μ is set too big, the Epsilon Greedy will grab most values in the best case, but the action of the algorithm still becomes pretty volatile.

Based on its accuracy and strong tolerance to noise on the results, UCB-1 is the right option for MAB for a low base conversion, a limited impact size scenario. In situations with a higher base rate or a higher anticipated impact size, thompson sampling might be a safer alternative, where the algorithm is robust.

References

[1] P. K. Andersen and R. D. Gill, “Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to The Annals of Statistics. ® www.jstor.org,” *Statistics (Ber).*, vol. 10, no. 3, pp. 1100–1120, 1982, doi: 10.1214/aos/1176348654.

[2] J. Vermorel and M. Mohri, “Multi-armed bandit algorithms and empirical evaluation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3720 LNAI, pp. 437–448, 2005, doi: 10.1007/11564096_42.

[3] V. Kuleshov and D. Precup, “Algorithms for multi-armed bandit problems,” vol. 1, pp. 1–32, 2014, [Online]. Available: http://arxiv.org/abs/1402.6028.

[4] T. De Feyter, “Modelling heterogeneity in manpower planning : dividing the personnel system into more homogeneous subgroups,” *Appl. Stoch. Model. Bus. Ind.*, no. March, pp. 321–334, 2006, doi: 10.1002/asmb.

[5] S. Bubeck and N. Cesa-Bianchi, “Regret analysis of stochastic and nonstochastic multi-armed bandit problems,” *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, 2012, doi: 10.1561/2200000024.

[6] M. N. Katehakis and A. F. Veinott, “Multi-Armed Bandit Problem: Decomposition and Computation.,” *Math. Oper. Res.*, vol. 12, no. 2, pp. 262–268, 1987, doi: 10.1287/moor.12.2.262.

[7] S. Agrawal and N. Goyal, “Analysis of thompson sampling for the multi-armed bandit problem,” *J. Mach. Learn. Res.*, vol. 23, pp. 1–26, 2012.

[8] A. Ö. Saritaç and C. Tekin, “Combinatorial multi-armed bandit problem with probabilistically triggered arms: A case with bounded regret,” *2017 IEEE Glob. Conf. Signal Inf. Process. Glob. 2017 - Proc.*, vol. 2018-January, pp. 111–115, 2018, doi: 10.1109/GlobalSIP.2017.8308614.

[9] T. Lu, D. Pál, and M. Pál, “Showing Relevant Ads via Lipschitz Context Multi-Armed Bandits,” *Thirteen. Int. Conf. Artif. Intell. Stat.*, vol. 9, pp. 485–492, 2010, [Online]. Available: http://proceedings.mlr.press/v9/lu10a/lu10a.pdf%0Ahttp://www.jmlr.org/proceedings/papers/v9/lu10a/lu10a.pdf%0Ahttp://www.ualberta.ca/~dpal/papers/clicks/lipschitz-clicks.pdf.

[10] K. Liu and Q. Zhao, “Distributed learning in multi-armed bandit with multiple players,” *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, 2010, doi: 10.1109/TSP.2010.2062509.

[11] N. Cesa-Bianchi, "Multi-armed Bandit Problem," *Encycl. Algorithms*, pp. 1–5, 2014, doi: 10.1007/978-3-642-27848-8_768-1.

[12] E. Even-Bar, S. Mannor, and Y. Mansour, "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems," *J. Mach. Learn. Res.*, vol. 7, pp. 1079–1105, 2006.

[13] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "Gambling in a rigged casino: the adversarial multi-armed bandit problem," *Annu. Symp. Found. Comput. Sci. - Proc.*, pp. 322–331, 1995, doi: 10.1109/sfcs.1995.492488.

[14] Henderi and T. Wahyuningsih, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijiis.v4i1.73.

[15] S. Bubeck, R. Munos, and G. Stoltz, "Pure exploration in multi-armed bandits problems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5809 LNAI, pp. 23–37, 2009, doi: 10.1007/978-3-642-04414-4_7.

[16] S. Pandey, D. Chakrabarti, and D. Agarwal, "Multi-armed bandit problems with dependent arms," *ACM Int. Conf. Proceeding Ser.*, vol. 227, pp. 721–728, 2007, doi: 10.1145/1273496.1273587.

[17] E. Kaufmann, O. Cappé, and A. Garivier, "On the complexity of best-arm identification in multi-armed bandit models," *J. Mach. Learn. Res.*, vol. 17, pp. 1–42, 2016.