Effect of A/B testing Considering Smart -AD

Rofiah Adeshina

INTRODUCTION

SmartAd is a mobile first advertiser agency that designs intuitive touch-enabled advertising. It provides brands with automated adverts based on the principle of voluntary participation. It is running a new Ad on a client website and wants to determine if the Ad increases brand awareness. To determine this, an A/B statistical testing will be used to compare two variants.

The goal of evaluating test is to determine brand awareness by using the results from a BIO (Brand Impact Optimizer) questionnaire served to two target groups to determine the success per target group. The target groups are: "Control group" which are exposed to the old Ad and "Exposed group" which are exposed to new Smart Ad. With this we can determine if the client company can adopt the new Ad fully as opposed the old one with a guaranteed increase in brand awareness.

Effect of A/B testing Considering Smart -AD

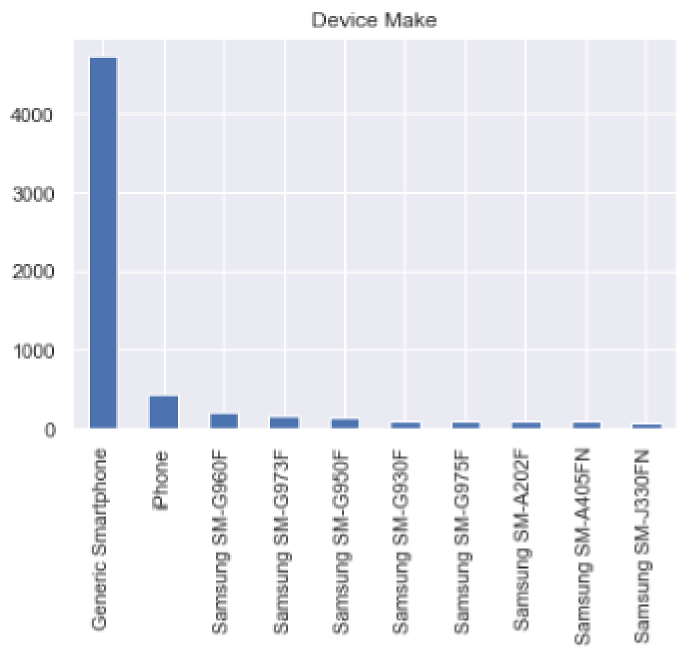
DATA EXPLORATION

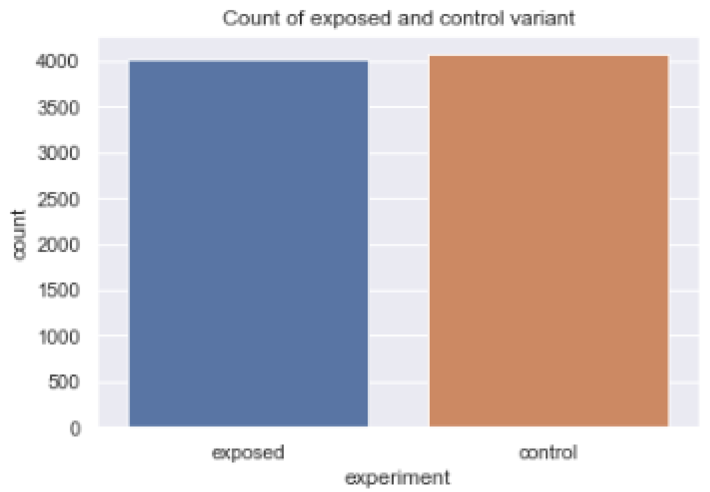
Data Exploration

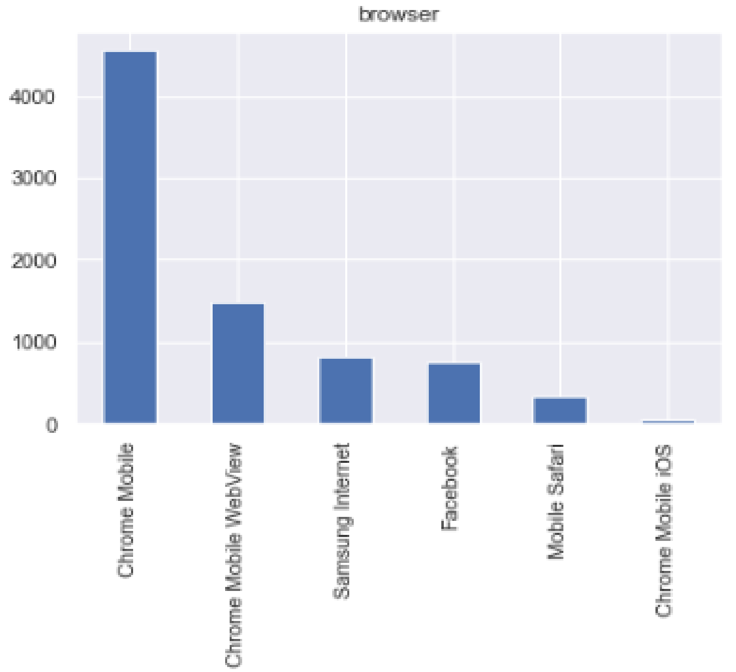
The data for this project is based on results from BIO questionnaire served defined variant groups. An overview of the dataset is shown below. The "auction_id" columns represent unique id of user who were presented the questionnaire, with the "yes" and "no" column representing their response. If the user doesn't give a reply then both columns a have zero value.

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
0	0008ef63-77a7-448b-bd1e-075f42c55e39	exposed	2020-07-10	8	Generic Smartphone	6	Chrome Mobile	0	0
1	000eabc5-17ce-4137-8efe-44734d914446	exposed	2020-07-07	10	Generic Smartphone	6	Chrome Mobile	0	0
2	0016d14a-ae18-4a02-a204-6ba53b52f2ed	exposed	2020-07-05	2	E5823	6	Chrome Mobile WebView	0	1
3	00187412-2932-4542-a8ef-3633901c98d9	control	2020-07-03	15	Samsung SM-A705FN	6	Facebook	0	0
4	001a7785-d3fe-4e11-a344-c8735acacc2c	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0

- The “experiment “ columns divides the data into variant groups of control and exposed. Exploring the data it was seen that 4006 users were exposed to the new SmartAd and 4071 to the old Ad.
- The “device_make” column represents the mobile phone users used to engage with website. From the visualization shown below most users accessing site use generic smartphones.
- Users also made use of various kinds of browsers to engage with the site, with “chrome mobile” browser being the most popular.







Effect of A/B testing Considering Smart -AD

METHODOLOGY

To determine which Ad leads to increase in awareness, I made use of A/B testing. A/B testing is the act of running a simultaneous experiment between two or more variants of a scenario to see which one performs the best. However, there are various approaches used in carrying out A/B testing. For this project, I explored various approaches as a means to determine their differences, similarities and the approach best fits a given scenario.

CLASSICAL A/B TESTING

This is also known as the traditional method of A/B testing. It compares only two variants of scenario (as in this case the control and experiment (or exposed) variant) in carrying out the test. Therefore, to carry out this analysis, the dataset was split into two variant groups and A/B testing was run simultaneously.

METRIC

To determining the success rate in both groups a metric of measure is defined based on the data at hand, available information and results to be accomplished. The metric I defined is success rate and it is given as:

$$\frac{Number\ of\ yes\ per\ variant\ group(success)}{Total\ responses\ received\ in\ group}$$

Control Group

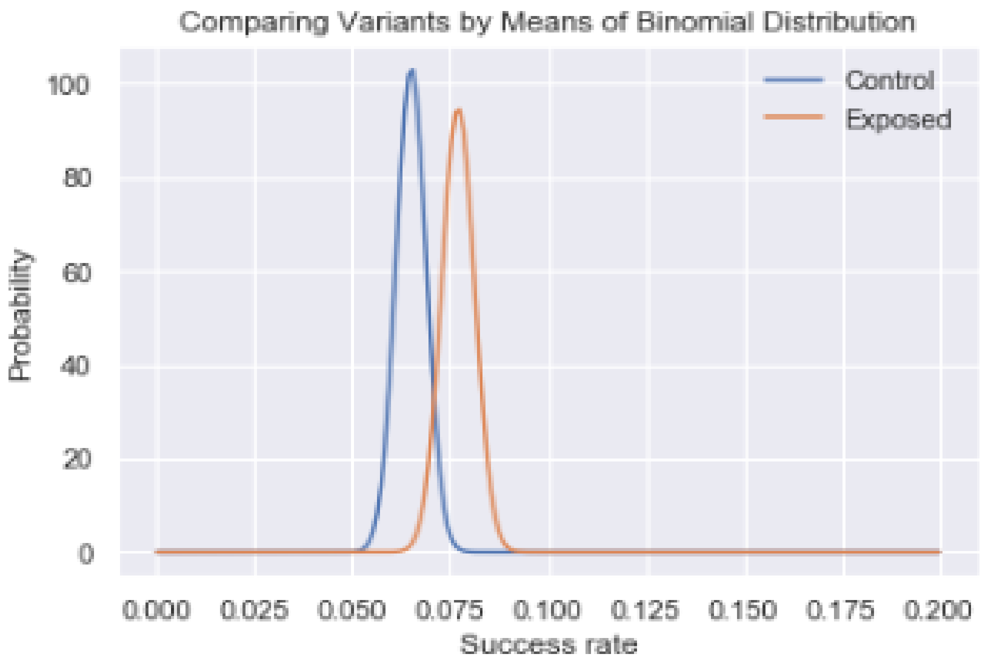
Population= 4071
Success rate = 0.065

Exposed Group

Population= 4006
Success rate = 0.077

- The population showed no significant difference between the variant groups, which indicated that the experiment won't be biased and is therefore reliable
- On computing the success rate, the exposed group showed a higher success rate. However, this was not enough to conclude that it will lead to an increase in brand awareness. Further analysis was done.

- Since a user could either respond “yes” or “no” , I can compared the two groups by plotting a binomial distribution. However, this was confusing as I was not able to see clearly the difference in the group. As we’re interested in the average conversion, or average time spent on the site, this averaging of an underlying distribution means our final estimate will be well approximated by a normal distribution.

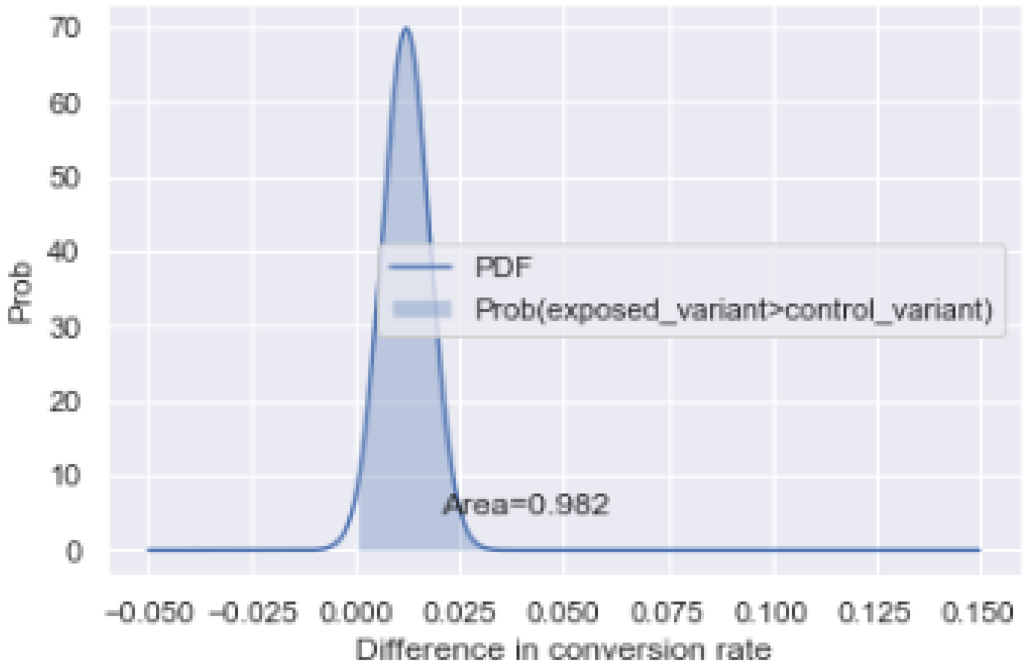


Here, the question is still the same: What is the chance that exposed group will have a higher success rate. Is it significant? To answer this I considered the sum or difference of normally distributed random numbers.

$$P(b-a) = \mathcal{N}(\mu_B - \mu_A, \sqrt{\sigma_A^2 + \sigma_B^2})$$

Hyphothesis

Null hyphothesis: Ad A and Ad B have no significant difference in brand awareness.



RESULT

After carrying out the analysis the following were obtained: z-score= 2.099 and p-value=0.018 (< 5%). Using frequentist approach, it can be said that given the null hypothesis is true (Ad A and Ad B have no significant difference in brand awareness). It is expected to get this result or a result more extreme only 0.018 * 100 = 1.8% (which is the value of our p) number of times. As this was a significant result (typically p < 5%), the null hypothesis is rejected, and evidence exists that the exposed group > control group. We can go ahead to say that the adverts that the advertising company runs resulted in a significant lift in brand awareness

Machine Learning A/B Testing

Another approach to A/B testing is the use of machine learning algorithms. In this analysis, I made use of three machine learning algorithms to carry out A/B testing; logistic regression, decision tree and XGBoost. As before, we want to determine which Ad leads to better awareness

Data Preprocessing

As is the case in all analysis, the data went through basic processing in order to get a good model. Irrelevant columns were removed and the data frame below was used in the machine learning analysis. The "yes" column served as the target variable and others as the features

	experiment	device_make	browser	yes
0	exposed	Generic Smartphone	Chrome Mobile	0
1	exposed	Generic Smartphone	Chrome Mobile	0
2	exposed	E5823	Chrome Mobile WebView	0
3	control	Samsung SM-A705FN	Facebook	0
4	control	Generic Smartphone	Chrome Mobile	0

All categorical columns were encoded into their numeric equivalent.

	exp	device_make	browser
0	1	46	2
1	1	46	2
2	1	29	3
3	0	137	6
4	0	46	2

After which the data was split into test and train in the ratio (90:10). Where 90% was used to train the model using 5-fold cross validation with the three algorithms and 10% was used in validating.

K-Fold Cross Validation

In order to ensure the accuracy of the model on every test set is as good as the accuracy it has obtained from the training set and generally ensure a good model, I made use of K-fold cross validation. In K-fold Cross-Validation, the training set is randomly split into K(here 5) subsets known as folds. Where K-1 folds are used to train the model and the other fold is used to test the model. This technique improves the high variance problem in a dataset as we randomly select the training and test folds.

Result
Logistic Regression

mean scores computed = 0.9281882288790161(92%)
Standard deviation of scores computed = 0.005083343959127255

The mean value for the accuracies computed after K-fold cross validation is 92% with a mean deviation 0.5%. That means our model is accurate for 92.5% or 91.5% of time.

loss function for logistic regression model is: 0.2487592975524973

Result
Decision Tree

Mean scores computed:0.9247489659964538
Standard deviation of scores computed :0.003928590791678719
The mean value for the accuracies computed after K-fold cross validation is 92% with a mean deviation 0.4%. That means our model is accurate for 92.3% or 91.7% of time.
The loss function for decision tree model is: 0.2487592975524973

Feature Importance
Experiment: 0.11711795763860647
Device_make: 0.671144265250116
Browser : 0.21173777711127753
From above, it seen that 'device_make' is a strong predictor driving the model. It therefore contributes more to the goal of gaining more "yes" results.

Result
XGBOOST

mean scores computed:0.9266748774768516
Standard deviation of scores computed : 0.004231396662205607
The mean value for the accuracies computed after K-fold cross validation is 92% with a mean deviation 0.5%. That means our model is accurate for 92.4% or 91.6% of time.
the loss function of XGBoost model is: 0.2632620950556105

Feature Importance
Experiment: 0.26487526
Device_make: 0.4610094
Browser : 'browser': 0.27411535
From above, it seen that 'device_make' is a strong predictor driving the model. It therefore contributes more to the goal of gaining more "yes" results.

Effect of A/B testing Considering Smart -AD

SUMMARY, CONCLUSION AND RECOMMENDATION

SUMMARY

Though classical A/B testing is a common approach, it poses the some issue: performing many tests, not necessarily concurrently, will multiply the probability of encountering a false positive; false positives increase if you stop a test when you see a positive result.

Machine learning A/B testing allows modelling of complex systems compared to the traditional classical A/b tests that only compares two 2 variables - an experiment/control to an outcome. The problem with comparing only two variants is that customer behavior is vastly more complex than 2 possible outcomes.

Machine learning tests gives feature importance which make it easy to know features that drive predictions in their order of importance and contribute to achieving the set out goal.

CONCLUSION

From the analysis I can recommend running SmartAd on the client website as brand awareness would be increased. Also, that the data points are enough to make a reasonable judgement. Further test can be run but it should noted that it will lead to more cost being incurred.

REFERENCE

- https://www.business-science.io/business/2019/03/11/ab-testing-machine-learning.html
- https://cosmiccoding.com.au/tutorials/ab_tests
- https://towardsdatascience.com/the-math-behind-a-b-testing-with-example-code-part-1-of-2-7be752e1d06f
- https://www.aionlinecourse.com/tutorial/machine-learning/k-fold-cross-validation

Thank You